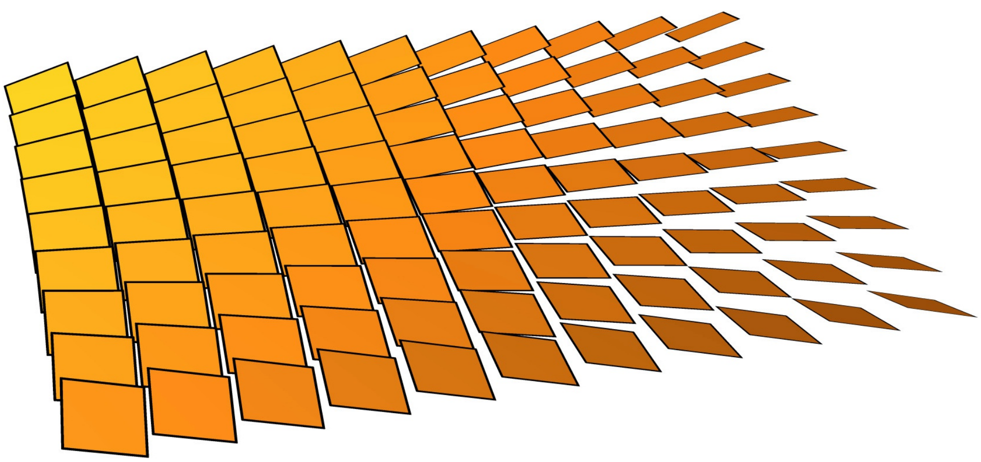
– book project –

Lecture notes on sub-Riemannian geometry

Carnot-Carathéodory spaces from the Lie group viewpoint



by Enrico Le Donne.

https://sites.google.com/view/enricoledonne/

Version of May 22, 2023

Contents

Chapter 0

An introduction

The asterisk * will denote incompleteness of the chapter or section.

0.1 About this text

This manuscript draws heavily on the following books and papers: [Pon66, War83, CG90, Bel96, Gro99, AFP00, BBI01, Hel01, Mon02, Kna02, HN12], as well as various articles authored by the writer and thesis of his students, such as [AKL09, BL13, LD15, LD17, LMO$^+$16, LN20, CKL$^+$17, **?**, **?**]. Additionally, the content of this manuscript incorporates insights from numerous conversations the author has had with his collaborators and mentors. The author would like to acknowledge, in the order of their significant contributions, B. Kleiner, U. Lang, E. Breuillard, A. Ottazzi, P. Pansu, and Y. Cornulier, who have provided invaluable guidance and support throughout this work.

This text was originally written for a course titled 'Sub-Riemannian Geometry' taught at ETH in Zürich (Switzerland) during Fall 2009 and later at the University of Jyväskylä (Finland) in Spring 2014. Subsequently, additional sections were included after the author delivered a course on 'Carnot Groups' at a summer school in Levico Terme (Trento, Italy) in 2015 and a course on 'Riemannian and Sub-Riemannian Geometry on Lie Groups' at the Neurogeometry summer school in Cortona (Italy) in 2017. These lecture notes were further expanded for the course 'Sub-Riemannian Geometry' taught at the University of Fribourg (Switzerland) in Spring 2021.

The primary audience for this book consists of young researchers who are seeking an introduction to the field of sub-Riemannian geometry. It can serve as reading material for a master's thesis or as an initial reference for those beginning a PhD program focusing on subjects that explore the interplay between geometry and analysis, along with group theory.

In contrast to other sources, such as [Mon02, ABB15, Jea14, Rif14, BLU07], this book employs the formalism of Lie groups. In fact, one of the aims of this book is to demonstrate how sub-Riemannian geometries manifest in other mathematical domains, including hyperbolic geometry and geometric group theory, through the lens of Lie groups.

Prerequisite topics, such as differential geometry, measure theory, and group theory, will be discussed within the main flow of the chapters. Given the positive feedback received from many students regarding this approach, the author has finally decided to publish this text.

0.2 What sub-Riemannian geometry is

Sub-Riemannian geometry is a generalization of Riemannian geometry. Roughly speaking, a sub-Riemannian manifold is a Riemannian manifold together with a constrain on admissible directions of movements. In Riemannian geometry every smoothly embedded curve has locally finite length. In sub-Riemannian geometry, if a curve fails to satisfy the obligation of the constrain, then it has infinite length.

One classical example one should carry in mind is coming from mechanics. Indeed, the stati of a moving object are enclosed by its position in space and the speeds of its parts: the momenta. Thus in the manifold 'positions times speeds' the possible evolutions of the object should satisfy the fact that the derivatives of the first coordinates are equal the second coordinates. In particular, some trajectories are not allowed. As trivial examples, you cannot vary your speed without changing your position or, similarly, you cannot move into another place at speed zero!

The 3D Heisenberg group is the most important sub-Riemannian geometry that is not in fact a Riemannian one. It is also not difficult to visualize some of its features. Topologically it is $\mathbb{R}^3$. The constrain on curves is given by what is called a 'distribution of planes'. Similarly as a smooth vector field smoothly assigns a tangent vector at each point of the manifold, a distribution of planes smoothly assigns to each point a plane inside the 3D tangent space at that point. The curves that we call 'admissible' are those curves that are tangent to one such a distribution. Refer to Figure 0.2 for a visual representation of a distribution.

The great feature of the Heisenberg group is that its distribution is curly enough in a way that each pair of point can be connected by at least one admissible curve. From this fact one can define a finite-valued distance similarly to the Riemannian case: the distance between two points p and q

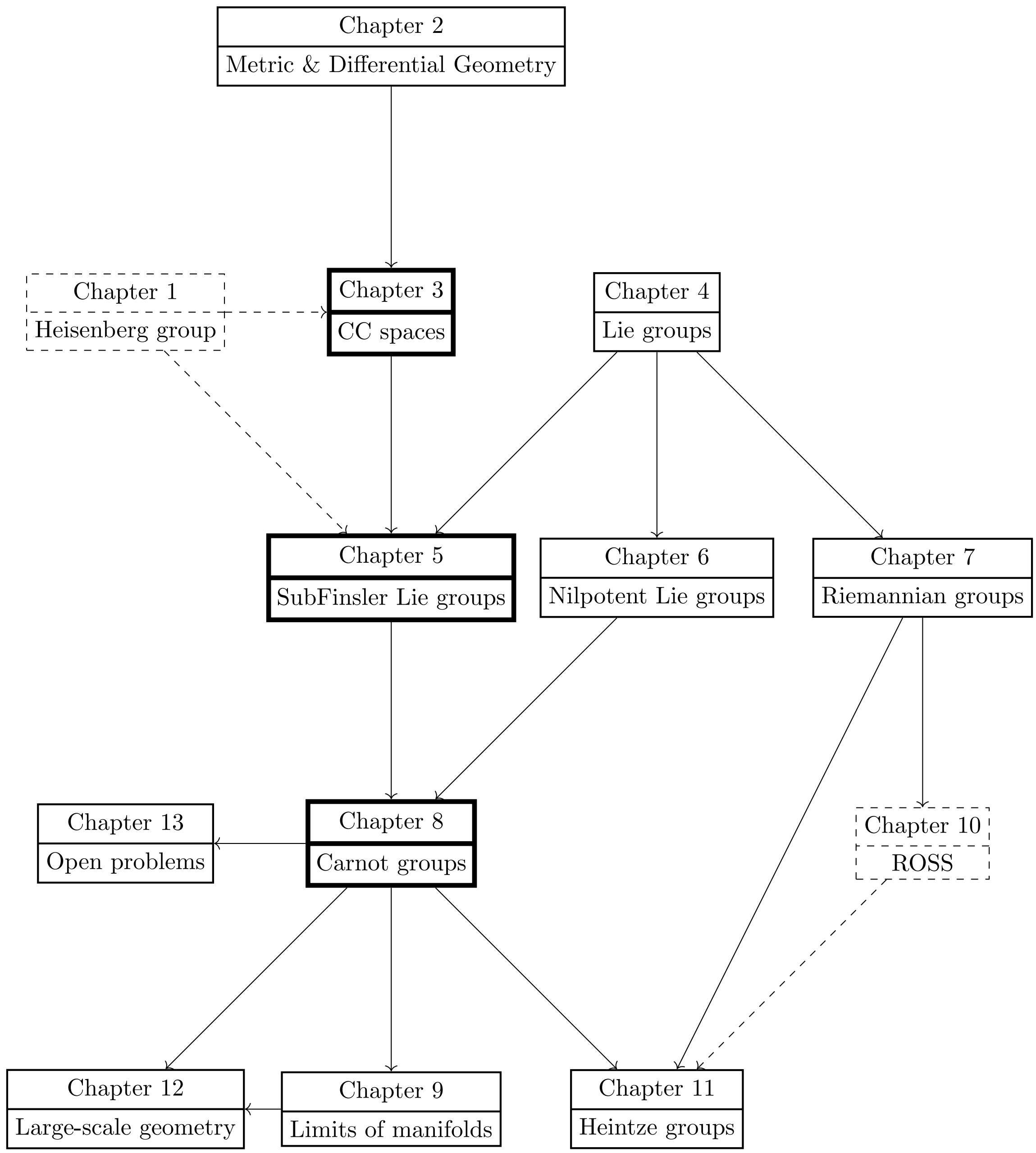


Figure 1: Book Content Diagram. Thick boxes represent the main chapters. Dashed boxes represent chapters devoted to examples.

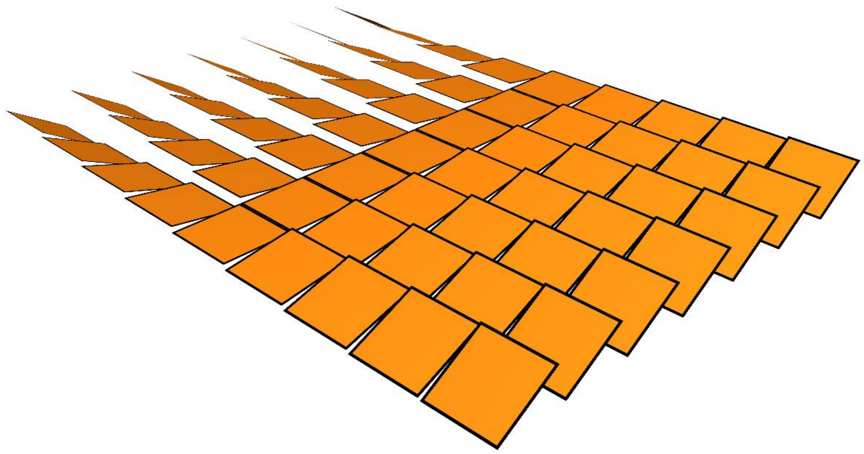


Figure 2: A contact distribution on $\mathbb{R}^3$

is given by the infimum of the length of all those admissible curves from p to q,

$$d(p,q) = \inf\{\text{Length}(\gamma) \;:\; \gamma \text{ admissible, from } p \text{ to } q\}. \tag{$\star$}$$

More generally, a sub-Riemannian manifold consists of a manifold, a length structure, and a subset of the tangent bundle. This data defines admissible curves and a distance as described in ($\star$). These geometries are distinct from Riemannian geometries in several aspects. Sub-Riemannian spaces are non-Riemannian, meaning they exhibit fractal properties as their Hausdorff dimension exceeds the topological dimension (assuming we are dealing with a proper subset of the subbundle). Additionally, there exist smooth curves with locally infinite length, as well as other smooth admissible curves that are isolated in the topology of smooth admissible curves with the same endpoints. Consequently, sub-Riemannian geometry requires different techniques than those used in Riemannian geometry.

The concept of sub-Riemannian distance can be traced back to the ideas of N. Carnot and C. Carathéodory, which is why sub-Riemannian manifolds are also referred to as Carnot-Carathéodory spaces. Since the 1980s, this geometry has emerged as a vibrant research field with applications and connections to various areas of pure and applied mathematics, including classical mechanics, control theory, geometry, group theory, and the analysis of hypoelliptic operators.

0.3 Content and structure of this text

We shall explore Carnot-Carathéodory spaces from the perspective of Lie groups. The main objective is to illustrate how these non-smooth geometries manifest in other mathematical domains, such as metric geometry and geometric group theory, through Carnot groups. Carnot groups play a fundamental role throughout the text and serve as a key example to keep in mind. They area class of nilpotent Lie groups equipped with sub-Riemannian and sub-Finsler structures. We will demonstrates the role of Carnot groups as asymptotic cones of finitely generated nilpotent groups. We also explores their presence as parabolic boundaries of rank-one symmetric spaces and their involvement as limits of Riemannian manifolds and tangents of sub-Riemannian manifolds.

In the opening Chapter 1, we will focus on the plane distribution in the 3D Heisenberg group. We will consider the induced distance ($\star$). Specifically, we will discuss the following facts:

1. This distance d turns the space $\mathbb{R}^3$ into a metric space with the same standard topology. In other words, nearby points can be connected by short admissible curves.

2. Between every two points, in fact there exists a geodesic curve. Namely, the distance between the two points is equal to the length of some curve connecting them. If a curve is admissible, its length is comparable to its Euclidean length. However, non-admissible curves have infinite length.

3. This metric space is distinct and different from Riemannian spaces. It is not biLipschitz equivalent to any Riemannian distance. This is because the Heisenberg geometry exhibits characteristics of fractal geometry. Indeed, the metric on this topologically 3-dimensional object has a metric dimension equal to 4, as determined by its Hausdorff measure.

The general definition of a Carnot-Carathéodory space arises when we formally define the concept of a distribution being "curly enough". This notion should ensure that every pair of points can be connected by an admissible curve. To delve deeper into this topic, we require a solid understanding of both Differential Geometry and Metric Geometry, which we review in Chapter 2.

Then, in Chapter 3, we delve into the sub-Riemannian geometry of Carnot-Carathéodory spaces, focusing on their distributions, distances, and dimensions. A *distribution* on M refers to a subbundle of the tangent bundle TM or, more generally, a subset of TM that, locally on the manifold, can be expressed as the span of a collection of vector fields. These distributions are also known as *polarizations*. A distribution $\Delta \subseteq TM$ is called *bracket generating* if, for every $p \in M$, the Lie algebra generated by the sections of Δ evaluated at p is the entire tangent space T_pM. In other words, a distribution Δ is bracket generating if every tangent vector $v \in TM$ can be represented as a linear combination of vectors of the following form: the evaluation at p of vector fields X_1, $[X_2, X_3]$, $[[X_4, [X_5, X_6]]]$, and so on, where all the vector fields $X_1, X_2, X_3, \ldots$ are tangent to Δ, and $v \in T_pM$.

A *subRiemannian manifold* is a triple (M, Δ, g), where M is a differentiable manifold, Δ is a bracket generating distribution, and g is a smooth section of positive-definite quadratic forms on Δ. In fact, g can be considered as the restriction to Δ of a Riemannian metric tensor on the manifold M. A curve γ on M is called *admissible*, or *horizontal*, with respect to Δ if it is absolutely continuous and $\dot{\gamma}(t) \in \Delta_{\gamma(t)}$ for almost every t. Then the *sub-Riemannian distance* (also known as *Carnot-Carathéodory metric*) is defined by the same formula $(\star)$. Most of the previously mentioned results on the Heisenberg group will be valid for every general sub-Riemannian distance.

The understanding of many Riemannian geometric properties come from the fact that 'metric' tangents of Riemannian manifolds are Euclidean spaces, and Euclidean geometry is enough understood. Such a notion of tangent is precisely defined in terms of limits of metric spaces, and we call them *tangent cones* or *metric tangents*. What are the metric tangents in sub-Riemannian geometry? The answer is not immediate. Under further assumptions of equiregularity, see Section 3.1.5, we will see that for 3-dimensional (non-Riemannian) sub-Riemannian manifolds we only have the Heisenberg group – another reason for it to be important. In general, alas, fixed a topological dimension greater or equal than 7, the possible tangents are infinitely many. It may not be the same one even for a given fixed sub-Riemannian manifold. The good news is that, analogously as the Heisenberg structure has a group structure, the metric tangent of a sub-Riemannian manifold has a Lie group structure at most points, and at every other point it is still a quotient of some Lie group. The metric tangent at 'regular' points has even more structure: it has a dilation property, and consequently it is a nilpotent Lie group. Such metric Lie groups are those called Carnot groups.

We shall review the needed theory of Lie groups in Chapter 4. In Chapter 5, we shall then see how the geometry of subRiemannian (or, more generally, subFinsler) Lie groups is manageable, using the left invariance of the structure. Before getting into Carnot groups, we shall see in Chapter 6 plenty of properties that one has on every nilpotent Lie group. In Chapter 7, we have a side discussion on the classical basic theory of Riemannian Lie groups, plus some other results on metric Lie groups. In Chapter 8, we finally define and study Carnot groups.

The aim of the book you are starting to read is twofold:

[1] We shall see that Carnot groups with their Carnot-Carathéodory distances appear in another mathematical areas. Namely, they appear as

(A) limits of Riemannian manifolds and tangents of subRiemannian manifolds, see Chapter 9.

(B) as asymptotic cones of finitely generated nilpotent groups, see Chapter 12.

(C) parabolic boundaries of rank-one symmetric spaces, see Chapter 11.

In harmonic analysis on stratified Lie groups, and more generally on graded groups, Carnot-Carathéodory distances appear in the study of hypoelliptic differential operators. In complex analysis, Carnot-Carathéodory spaces appear as boundaries of strictly pseudo-convex complex domains. We shall not treat these last two settings in this monograph, but we refer to the books [Ste93, CDPT07] as initial references.

[2] We shall see that with the use of Lie group theory on Carnot groups, or more generally on subFinsler Lie groups, one can perform calculus, analysis, Geometric Measure Theory, Calculus of Variations, and Geometric Analysis. In Chapter 8 we shall prove some important results. In Chapter 13 we discuss some open problems and recent developments.

Only after understanding the specific examples of tangent spaces of sub-Riemannian manifolds (i.e., Carnot groups), should one consider the general case of Carnot-Carathéodory spaces. The reason for this approach is that Carnot groups offer hope for understanding the theory of calculus in such settings, thanks to the availability of translations by group elements and the dilation property. It is worth noting how the classical definition of the derivative of a real function relies on addition, multiplication, and limits:

$$f'(x) = \lim_{h\to 0} \frac{f(x+h) - f(x)}{h}.$$

All these operations are present in Carnot groups, where addition is replaced with a possibly non-commutative group operation. Consequently, we can define a metric notion of the derivative known as the *blow-up differential*, or *Pansu derivative*, named after Pierre Pansu, who made pioneering contributions to the field (see [Pan89]). Let us state one of Pansu's theorems [Pan89], which has subsequently been expressed in this general form in [MM95, Vod07].

Theorem A (Pansu's Rademacher Theorem)**.** *Given a Lipschitz map between sub-Riemannian manifolds, at almost all points its blow-up differential exists, is a group homomorphism of the tangent* *cones, and is equivariant with respect to their dilations.*

This theorem will be proved in Chapter 8 in the setting of Carnot groups. In fact, the theorem holds also for quasi-conformal maps on Carnot-Carathéodory spaces, see [Vod07]. The theory of quasi-conformal mappings has been used to prove rigidity theorems on hyperbolic spaces over the division algebras of real, complex, or quaternionic numbers. Indeed, as we shall see in Chapter 11 the 'parabolic visual boundaries' of rank-one symmetric spaces are Carnot groups. More generally, all negatively curved homogenous Riemannian manifolds have graded groups as boundaries. This last fact is mostly based on the work of Heintze. In Chapter 10 we shall discuss rank-one symmetric spaces, which can be seen as semidirect products of Lie groups. One of the elements of the semidirect product is a Carnot group of Heisenberg type, the other element is the one-dimensional Lie group, which act on the Carnot group by its dilations. While in Chapter 11 we shall review the notion of boundary of CAT(-1) spaces, observe that the boundary of every rank-one symmetric space are the particular Carnot group of Heisenberg type, and discuss the general viewpoint of Heintze groups.

As we will explain in Chapter 12, Carnot groups with Carnot-Carathéodory distances, appear in Geometric Group Theory as asymptotic cones of nilpotent finitely generated groups, see [Gro96, Pan89]. Part of this text is devoted to the study of the coarse geometry of nilpotent groups. We will see how a geometric notion as the polynomial growth of balls in the Cayley graph of a discrete group relates with the geometry of the tangent cone at infinity of this graph, which in this case turns out to be a Carnot group endowed with a Finsler-Carnot-Carathéodory metric, and eventually gives an algebraic consequence: the group is (virtually) nilpotent.

The last part of this book will be focused on some current topics of Geometric Measure Theory in the setting of Carnot groups. Most of the presented results are proved in the case of nilpotent Lie groups endowed with their Carnot-Carathéodory metric, but might be valid for subRiemannian manifolds. We will discuss some regularity prolems, explaining how much is still unknown in the theory. In particular we shall focus on the following problems:

- Are sets that have finite perimeter rectifiable?
- How the theory of minimal surfaces differs from the Euclidean case?
- What is the regularity of geodesics?

The above questions have not complete answers yet. In fact they are leading most of the recent research in sub-Riemannian geometry.

0.4 Sub-Riemannian geometries as models*

[1] Sub-Riemannian geometry (also known as Carnot geometry in France, and non-holonomic Riemannian geometry in Russia) has been a full research domain from the 80's, with motivations and ramifications in several parts of pure and applied mathematics. However, historically it was not clear that such theories were heading into the same notions. Thus each source provided its own jargon to the field. The non-expert reader will soon realize that some concepts have multiple terminology: a contact structure is a particular distribution of hyper-plane in an odd-dimensional manifold and the concept of Carnot-Carathéodory metric is a generalization of a sub-Riemannian distance.

0.4.1 Examples from Mathematics

Control theory

Control theory is an interdisciplinary branch of engineering and mathematics that deals with the behavior of dynamical systems. The usual objective is to control a system, in the sense of finding, if possible, the trajectories to reach a desired state and do it in an optimal way. Sub-Riemannian geometry follows the same setting of considering systems that are controllable with optimal trajectories and study this spaces as metric spaces. Many of the theorems in sub-Riemannian geometry can be formulated and prove in the more general settings of control theory. For example, the sub-Riemannian theorems by Chow, Pontryagin, and Goh have more general statement in geometric control theory. The reader interested in this view point should consult the book [AS04].

Geometric Group Theory

Sub-Riemannian geometry has a significant presence in geometric group theory, which is the study of groups from a geometric perspective. Sub-Riemannian structures naturally arise as asymptotic cones of groups of polynomial growth. A group generated by a finite set S growths polynomially if the cardinality of the product set S^n is bounded polynomially in $n \in \mathbb{N}$. Asymptotic properties of these groups relates to the geometry of their asymptotic cones.

For example, by a theorem of Pansu that we will present in this book, if a nilpotent group Γ is generated by a finite set S and has polynomial growth, then there are constants $Q \in \mathbb{N}$ and $V > 0$

[1] The asterisk * will denote incompleteness of the chapter or section.

such that

$$\#(S^n)/n^Q \to V, \qquad \text{as } n \to \infty.$$

Here Q an V have a very clear geometric meaning: Q is the Hausdorff dimension of the asymptotic cone of Γ and V is the volume of the unit ball in the asymptotic cone. Moreover, this asymptotic cone is an example of a Carnot group, equipped with a Carnot-Carathéodory distance. These are the spaces on which this book will be focusing. See Chapters 8 and 12

Complex analysis and Cauchy-Riemann geometry

Sub-Riemannian geometry arises when studying the geometry of Cauchy-Riemann (CR) manifolds. Typical examples are domains in complex Euclidean space, $\mathbb{C}^n$. The boundaries of strictly pseudo-convex domains, are of great importance in complex analysis and several complex variables and they are naturally equipped with visual distances that of Carnot-Carathéodory.

A domain in $\mathbb{C}^n$ is called strictly pseudo-convex if, at every point on its boundary, there exists a defining function whose Levi form is positive definite. The Levi form encodes information about the local geometry of the boundary and is related to the complex Hessian of the defining function.

The boundaries of strictly pseudo-convex domains in $\mathbb{C}^n$ exhibit rich geometric and analytic properties. Understanding them is crucial in the study of several complex variables, where it serves as a foundation for topics like plurisubharmonic functions, the Cauchy-Riemann equations, and the theory of complex manifolds. In this book we will not discuss strictly pseudo-convex domains, but we refer to [].

The main example that the reader should have in mind is the unit ball in $\mathbb{C}^n$. On the one hand, it is an example of strictly pseudo-convex domain. On the other hand, it is an example of rank-one symmetry space. Namely, when equipped with its Carathéodory distance it is the complex hyperbolic space. Together with the real hyperbolic spaces and quaternionic hyperbolic space (and the octonionic plane), they for the rank-one symmetry spaces of non-compact type. All these spaces have a well defined metric on the boundary, which is a Carnot-Carathéodory distance. We will present rank-one symmetry spaces from the Lie group viewpoint as semidirect products in Chapter 10. Then in Chapter 11 we shall discuss the visual boundaries of them and of all the negatively curved homogeneous spaces.

Analysis of hypoelliptic operators

In the theory of partial differential equations (PDEs), there is a big variety of types of the equations, which therefore require different typologies of treatment. One of the most important operator is the Laplacian, which in Euclidean space $\mathbb{R}^n$ is defined as

$$f \mapsto \nabla f := \frac{\partial^2}{\partial x_1^2} f + \ldots + \frac{\partial^2}{\partial x_n^2} f.$$

This operator is said to be hypoelliptic because it has the property that

$$\nabla f \in C^\infty \Rightarrow f \in C^\infty.$$

There is a connection between the Laplacian operator in Euclidean space and the Euclidean distance Indeed, the fundamental solution to the Laplace equation, called the *Green's function* is written as a function of the Euclidean distance.

When considering differential equations defined by bracket-generating vector fields, such as the sub-Riemannian structure, the Laplacian generalizes to the sub-Laplacian. The sub-Laplacian takes into account the nonholonomic constraints imposed by the vector fields. In fact, it relates to the geometry of the sub-Riemannian metric obtained from the bracket-generating vector fields. This link is used to study diffusion processes, heat equations, and other differential equations on such manifolds. Initial references to know more about analysis of hypoelliptic operators are [Fol73, RS76, Cap97]

Classical mechanics

Sub-Riemannian geometry has a presence in classical mathematical mechanics in the context of studying mechanical systems with constraints. These systems, known as nonholonomic systems, involve constraints on the possible motions of the system, typically expressed as limitations on velocities or accelerations. Sub-Riemannian geometry provides a framework for analyzing and understanding the geometric properties of these constrained mechanical systems. By considering the sub-Riemannian structure associated with the constraints, researchers can develop techniques for modeling, controlling, and studying the dynamics of such systems in classical mechanics.

Symplectic and contact geometry

To Do

Riemannian geometry

Riemannian geometry (of which sub-Riemannian geometry constitutes a natural generalization, and where sub-Riemannian metrics may appear as limit cases)

Diffusion on manifolds

To Do

Univalent Function Theory

There is a very remarkable application of sub-Riemmanian geometry to Univalent Function Theory. The application is very recent and so not still well known, it is why we shall discuss in more length this application here. The following quick summary is based on the paper [MPV07] and on kind conversations with Jeremy Tyson.

Classical Univalent Function Theory considers the class $\mathcal{S}$ of analytic univalent functions f defined in the unit disc in the complex plane normalized by $f(0) = 0$ and $f'(0) = 1$. Little is known on descriptions and properties of the, so called, *coefficient body*

$$M := \{(a_k) \in \mathbb{C}^{\mathbb{N}} : (a_k) \text{ are the power series coefficients at } z = 0 \text{ for some } f \in \mathcal{S}\}$$

or its finite-dimensional slices

$$M_n := \{(a_2, a_3, \ldots, a_{n+1}) \quad : \quad (a_k) \text{ are the first } n \text{ (undetermined) power series coefficients} \\ \text{at } z = 0 \text{ for a function in } \mathcal{S}\}.$$

The Bieberbach Conjecture (proved by de Branges in 1984) says $|a_n| \le n$ for all n. This gives information on the size of M_n and M. There is no explicit description of M_n except for the cases $n = 2$ (trivial) and $n = 3$ (Schaeffer-Spencer, 1950).

One of the basic tools in the subject is the Loewner (or Loewner-Kufarev) parametric representation, which embeds each function $f \in \mathcal{S}$ into an ODE flow within the class $\mathcal{S}$. Loewner parametrizations were used by de Branges in his proof. Nowadays there is a stochastic version of the Loewner flow (SLE) which is a very hot topic at the intersection of probability, complex analysis, stochastic PDE, math physics, etc.

Anyways, what Markina-Prokhorov-Vasilev show is that one can use the Loewner flow on $\mathcal{S}$ to define a natural (partially integrable) Hamiltonian system on the coefficient bodies M_n. They find certain first integrals of the flow and calculate all the relevant commutators. From there they construct a *complex* sub-Riemannian structure on M_n which is naturally adapted to the underlying univalent function theory. In fact, the Loewner parametrices become horizontal curves with respect to this sub-Riemannian structure.

An interesting problem in the field is to extend Markina-Prokhorov-Vasilev's setup to cover SLE as well as the classical (deterministic) Loewner equation.

0.4.2 Examples from Physics

Sub-Riemannian geometry models various structures, from finance to mechanics, from bio-medicine to quantum phases, from robots to falling cats! We don't want to enter in the details first because of lack of time, second because of lack of competence. We will address the interested reader to other papers.

Geometry of principal bundles with connections

Theoretical physics defines most mechanical systems by a kinetic energy and a potential energy. Gauge theory also know as the geometry of principal bundles with connections studies systems with physical symmetries, i.e., when there is a group acting on the configuration space by isometries. Most of the times it will be easier to understand the dynamics up to isometries, successively one has to study the 'lift' of the dynamics into the initial configuration space. Such lifts will be subject to a sub-Riemannian restriction.

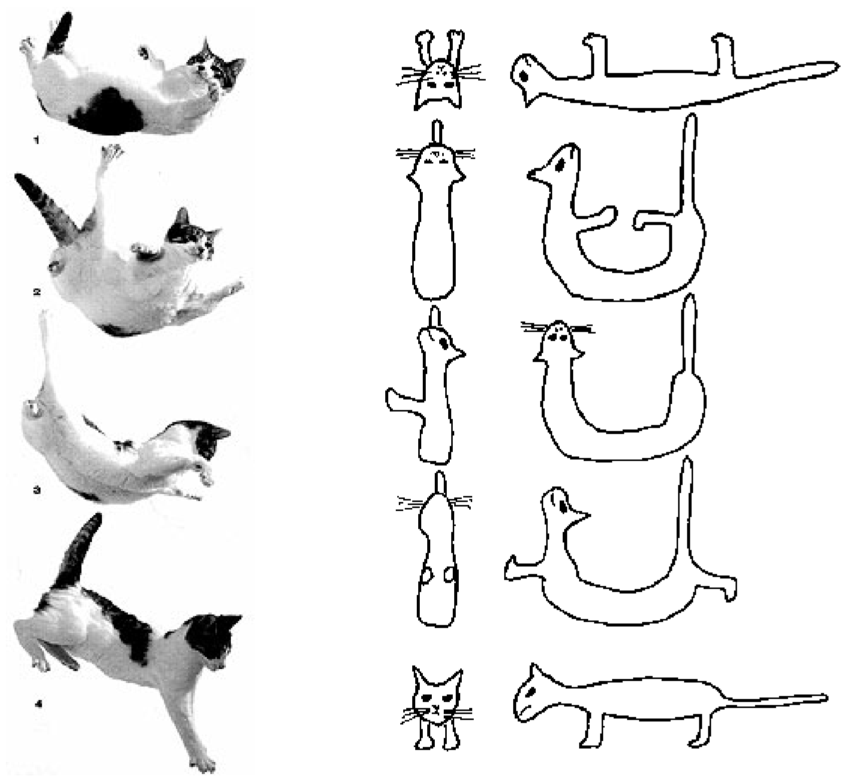
Falling cats

The formalism of principal bundles with connections is well presented by the example of the fall of a cat. A cat, dropped from upside down, will land on its self. The reason of this ability is the good flexibility of the cat in changing its shape.

Let us fix some formalism. Let M be the set of all the possible configurations in the 3D space of a given cat. Let S be the set of all the shapes that a cat can assume. Both M and S are manifolds of dimension quite huge. A position of a cat is just its shape plus its orientation in space. Otherwise said, the group of isometries $G := \mathrm{Isom}(\mathbb{R}^3)$ of the Euclidean 3D space acts on M and the shape space is just the quotient of the action:

$$\pi : M \to M/G = S.$$

In fancy words, M is a principal G-bundle.



(a) A photo. (b) A sketch.

Figure 3: The cat spins itself around and right itself.

The key fact is that the cat has complete freedom in deciding its shape $\sigma(t) \in S$ at each time t. However, during the fall, each strategy $\sigma(t)$ of changing shapes will give as a result a change in configurations $\tilde{\sigma}(t) \in M$. The curve $\tilde{\sigma}(t)$ satisfies

$$\pi(\tilde{\sigma}) = \sigma.$$

Moreover the lifted curve is unique: it has to satisfy the constrain given by the 'natural mechanical connection'. What the cat is proving is that such connection has non-trivial holonomy. In other words, the cat can choose to vary its shape from the standard normal shape into the same shape giving as a result a change in configuration: the legs were initially toward the sky, then they are toward the floor.

From mechanics: parking cars, rolling balls, moving robots, and satellites

Sub-Riemannian geometry has been extensively used in the field of mechanics and robotics. The study of sub-Riemannian structures provides a mathematical foundation for analyzing the motion planning and control of underactuated mechanical systems. These systems have fewer control inputs than the degrees of freedom, leading to nontrivial constraints on the achievable motions. By understanding the sub-Riemannian geometry associated with such systems, researchers can develop efficient control strategies for navigating robots in complex environments.

Parking a car or riding a bike. The configuration space is 3-dimensional: the position in the 2-dimensional street plus the angle with respect to a fixed line. However, the driver has only two degree of freedom: turning and pushing. Using again non-trivial holonomy we can move the car to every position we like.

Rolling a ball on the plane. A position of a ball lying on a plane requires five coordinates: two reals to characterize the point in the plane where the ball is touching it, another two coordinates to characterize the point of the ball which touches the plane, and the last one for spinning the ball around its vertical axis. When one rolls the ball without sliding, there are only three admissible control directions: two to choose a direction and then roll the ball and the third one for spinning it. Still, one can get to every position regardless of the initial position.

In *robotics* the mechanisms, as for example the arm of a robot, are subjected to constrain of movements but do not decrease the manifold of positions. Similar is the situation of *satellites*.

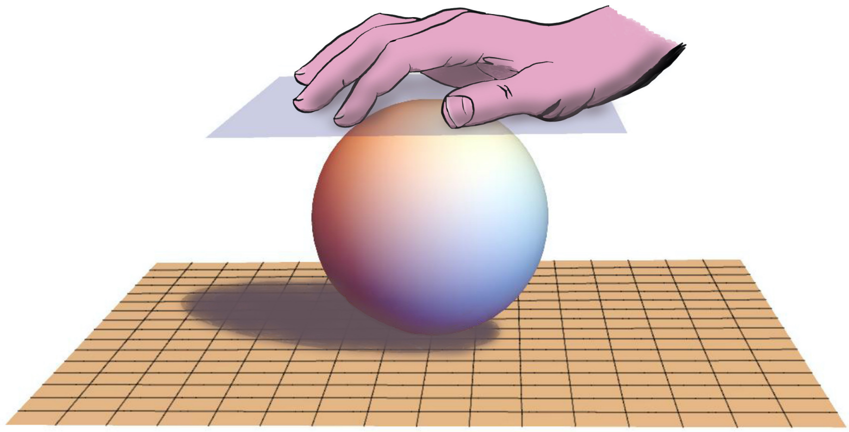


Figure 4: A ball rolling on the plane without sliding.

One should really think about a satellite as a falling cat: it should choose properly its strategy of modifying the shape to have the necessary change in configuration. Another similar example is the case of an astronaut in outer space.

Quantum Control and Quantum Information

Sub-Riemannian geometry has been utilized in the control and manipulation of quantum systems. Quantum control aims to steer quantum systems to desired states by applying suitable control fields. Sub-Riemannian structures naturally arise when considering the controllability of quantum systems subject to constraints on the available control resources. By applying techniques from sub-Riemannian geometry, researchers can design control protocols for quantum systems, which find applications in quantum computing, quantum information processing, and quantum sensing.

I became aware of the following application from a discussion with Ugo Boscain and reading his 'Habilitation à diriger des recherches'.

Let $\mathcal{H}$ be a complex separable Hilbert space. Let us denote by $\mathbb{S}$ the unit sphere in $\mathcal{H}$. In quantum mechanics the time evolution of quantum mechanical system (e.g., an atom, a molecule, or a system of particles with spin) is described by a map $\psi : \mathbb{R} \to \mathbb{S}$, called *wave function.* The vector $\psi(t)$ is called the *state of the system* at time t. The equation of evolution of the state is the so-called *Schrödinger equation.* If the system is isolated, the equation has the form:

$$i\frac{\mathrm{d}\psi}{\mathrm{d}t}(t) = H_0\psi(t),$$

where H_0 is a self-adjoint operator acting on $\mathcal{H}$ called free Hamiltonian.

For simplicity of notation, let us assume that the spectrum of H_0 is discrete and non-degenerate, with eigenvalues $E_1, E_2, \dots$ (called *energy levels*) and eigenvectors $\psi_1, \psi_2, \dots \in \mathbb{S}$.

Assume now to act on the system with some external fields (e.g an electromagnetic field) whose amplitude is represented by some functions $u_1, \dots, u_m \in L^\infty(\mathbb{R};\mathbb{R})$. In this case the Schrödinger equation becomes

$$i\frac{\mathrm{d}\psi}{\mathrm{d}t}(t) = H(t)\psi(t), \qquad \text{where } H(t) := H_0 + \sum_{j=1}^m u_j(t)H_j,$$

and H_j are self-adjoint operators representing the coupling between the system and the external fields. The time dependent operators $H(t)$ and $\sum_{j=1}^m u_j(t)H_j$ are called the *Hamiltonian* and the *control Hamiltonian*, respectively. The typical problem of quantum control is the so called *Population Transfer Problem*:

Assume that at time zero the system is in an eigenstate ϕ_j of the free Hamiltonian H_0. Design controls $u_1, \dots, u_m$ such that at a fixed time T the system is in another prescribed eigenstate ϕ_l of H_0.

Nowadays quantum control has many applications in chemical physics, in nuclear magnetic resonance (also in medicine) and it is central in the implementation of the so-called *quantum gates* (the basic blocks of a quantum computer), [].

For a finite-dimensional quantum mechanical system, if n is the number of energy levels, then we have $\mathcal{H} = \mathbb{C}^n$ and the state space $\mathbb{S}$ is the unit sphere $\mathbb{S}^{2n-1} \subset \mathbb{C}^n$. In this setting, problems of quantum mechanics (being multilinear) can be formulated with matrices. The solution is of the form

$$\psi(t) = g(t)\psi(0), \qquad \text{with } g(t) \in SU(n).$$

The Schrödinger equation becomes $\dfrac{\mathrm{d}}{\mathrm{d}t}g(t) = -iH(t)g(t)$, and now $-iH(t)$ is a skew trace-zero Hermitian matrix, i.e., belongs to the Lie algebra $\mathfrak{su}(n)$.

The controllability problem (i.e., proving that for every couple of points in $SU(n)$ one can find controls steering the system from one point to the other) is nowadays well understood. Indeed, the system is controllable if and only if the Hörmander's condition holds:

$$\mathrm{Lie}\{iH_0, iH_1, \dots, iH_m\} = \mathfrak{su}(n).$$

Once that controllability is proved one would like to steer the system, between two fixed points in the state space, in the most efficient way. In applications, typical costs that to be minimized are: either the energy transferred by the controls to the system or the time of transfer.

Quantum Berry's phases

Berry's phase is a phase factor that accumulates during the adiabatic evolution of quantum systems. It arises when a quantum system undergoes slow changes while staying in its instantaneous eigenstate. The connection with sub-Riemannian geometry arises in systems with degenerate energy levels, where the parameter space exhibits a geometric structure with the structure of a fiber bundle. The constraints imposed by the structure affect the system's evolution and lead to the accumulation of Berry's phase. The reader should start by reading the introduction in [Mon02], and the references therein.

0.4.3 Appearances in scientific applications

Neurobiology

I became aware of the following application from conversations with S. Pauls and G. Citti. A suggested-to-curious-readers paper is [SCP08].

Neuro-biologic research over the past few decades has greatly clarified the functional mechanisms of the first layer (V1) of the visual cortex. Such layer contains a variety of types of cells, including the so-called 'simple cells'. Researchers found that simple cells are sensitive to orientation specific brightness gradients.

Recently, this structure of the cortex has been modeled using a sub-Riemannian manifold. The space is $\mathbb{R}^2 \times \mathbb{S}^1$ where each point (x, y, θ) represents a column of cells associated to a point of retinal data $(x, y) \in \mathbb{R}^2$, all of which are attuned to the orientation given by the angle $\theta \in \mathbb{S}^1$. In other words, the vector $(\cos\theta, \sin\theta)$ is the direction of maximal rate of change of brightness at point (x, y) of the picture seen by the eye, such vector can be seen as the normal to the boundary of the picture.

The moral is that when the cortex cells are stimulated by an image, the border of the image gives a curve inside this 3D space. Such curves are restricted to be tangent to the distribution spanned by the vector fields

$$X_1 = \cos(\theta)\partial_x + \sin(\theta)\partial_y \quad \text{and} \quad X_2 = \partial_\theta.$$

Researchers think that, if a piece of the contour of a picture is missing to the eye vision (or maybe it is covered by an object), then the brain tends to 'complete' the curve by minimizing some kind of energy, in other words, there is some sub-Riemannian structure on the space of visual cells and the brain consider a sub-Riemannian geodesic between the endpoints of the missing data.

Applied Optimal Transport

Sub-Riemannian geometry has connections to optimal transport theory and mechanics. Optimal transport deals with finding the most efficient ways to transport mass from one configuration to another, subject to certain cost functions. Sub-Riemannian structures arise naturally in optimal transport problems when considering transport plans with constraints. The study of these geometric structures provides insights into the fundamental properties of optimal transport and has applications in diverse areas such as economics, fluid dynamics, and image registration.

Image Processing and Computer Vision

Sub-Riemannian geometry has been employed in image processing and computer vision applications. By representing images as curves in a suitable space, sub-Riemannian techniques allow for the extraction of intrinsic geometric features that are invariant under certain transformations. This enables robust object recognition, shape analysis, and image matching algorithms, which can be applied in fields like pattern recognition, medical imaging, and image-based navigation.

Neuroscience

Sub-Riemannian geometry has been applied to model and analyze the connectivity and activity patterns in the brain's neural networks. The brain's white matter, which consists of axonal bundles, can be viewed as a sub-Riemannian manifold, where the propagation of nerve impulses is subject to constraints imposed by the underlying anatomy. By studying the sub-Riemannian geometry of neural networks, researchers can gain insights into brain connectivity, information processing, and the relationship between structure and function.

Financial mathematics

Sub-Riemannian geometry has also found applications in the field of financial mathematics, particularly in the modeling and analysis of complex financial systems. Here are some ways in which sub-Riemannian geometry has been used in this context, as it has been in part explained to me by Josef Teichmann:

Sub-Riemannian geometry has been employed to model the evolution of asset prices and the dynamics of financial markets. It has been used to model and analyze the price dynamics of options in the presence of constraints, such as transaction costs and market frictions. It has been employed to model the dynamics of HFT systems, where trading strategies must account for constraints on transaction costs, market impact, and latency. It has also been used in risk management and regulatory compliance in the financial industry. Sub-Riemannian geometry has been applied to model and analyze the market microstructure, incorporating constraints such as bid-ask spreads, market impact, and trading volumes By considering the geometric properties and constraints present in financial markets, researchers can develop more accurate pricing models, risk management strategies, and trading algorithms that account for real-world complexities.

Chapter 1

The main example: the Heisenberg group

The sub-Riemannian Heisenberg group is the first prominent example of sub-Riemannian geometry that deviates from the Riemannian framework. Such a geometry is connected to the solution of the isoperimetric problem on the plane and has a formulation in terms of contact geometry.

In this chapter, we present the geometric models of the sub-Riemannian Heisenberg group and explore certain properties that will be further examined in Carnot groups. Given that the topological dimension of the Heisenberg group is 3, visualizing its sub-Riemannian geodesics and spheres becomes relatively simple.

1.1 An isoperimetric problem on the plane

The *isoperimetric problem* is a mathematical challenge where the goal is to find the maximum area among domains with a fixed length as perimeter. In our study, we will focus on a specific variation of the standard isoperimetric problem known as the problem of Dido.

Dido, as described in ancient Greek and Roman sources, is renowned as the founder and first queen of Carthage, located in modern-day Tunisia. Her story is famously depicted in the epic poem Aeneid by the Roman poet Virgil. According to this account, King Jarbas was convinced by Dido to grant her a parcel of land along the African coast for settlement. The condition set forth was that Queen Dido could claim as much land as she could enclose with a leather string, utilizing the coastline as part of the boundary. The optimal solution to maximize the area in this scenario involves a half-circle, assuming the coastline can be treated as a straight line.

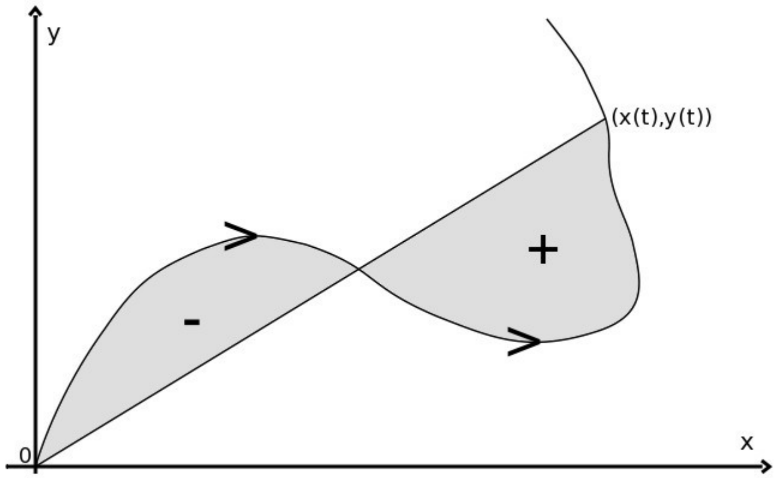


Figure 1.1: The lift of the curve is performed defining the third coordinate $z(t)$ as the oriented area of the region between the arc of the curve up to the point $(x(t), y(t))$ and the straight segment from $(0, 0)$ to $(x(t), y(t))$.

We next provide a mathematical model of such a problem. In $\mathbb{R}^2$ with coordinates (x, y), the area form is denoted by $\mathrm{vol} := \mathrm{d}x \wedge \mathrm{d}y$, which is the differential of the differential one-form

$$\alpha := \frac{1}{2}(x\,\mathrm{d}y - y\,\mathrm{d}x) = \frac{1}{2}r^2\,\mathrm{d}\theta.$$

By applying Stokes Theorem we deduce that if a closed, smooth, counterclockwise-oriented curve γ in $\mathbb{R}^2$ encloses a domain D_γ, then the area of D_γ is equivalent to the line integral of α along γ:

$$\mathrm{Area}(D_\gamma) := \iint_{D_\gamma} \mathrm{vol} = \int_\gamma \alpha.$$

Observe that at each point $(x, y) \in \mathbb{R}^2$, the vector (x, y) is in the kernel of α. Consequently, if L is a line passing through the origin, we have that $\int_L \alpha = 0$. This observation leads us to the conclusion that for a smooth curve γ, starting from the origin and not necessarily closed, the integral $\int_\gamma \alpha$ represents the signed area enclosed by γ and the line segment connecting the origin to the final point of γ. Refer to Figure 1.1 for a visual representation.

Therefore, Dido's problem can be reformulated as the task of maximizing the integral $\int_\gamma \alpha$ while fixing the integral $\int_\gamma \mathrm{d}s$. Here, $\int_\gamma \mathrm{d}s$ represents the length of the curve, obtained by integrating it with respect to the element of arc length $\mathrm{d}s$.

1.2 Contact-geometry formulation of the problem

One of the models of the Heisenberg geometry is constructed as follows, and it has the property that the projection $\pi : \mathbb{R}^3 \to \mathbb{R}^2$ onto the first two coordinates sends geodesics into the solutions of Dido's isoperimetric problem.

If we begin with a curve $\sigma(t) = (x(t), y(t))$ in $\mathbb{R}^2$, with $x(0) = y(0) = 0$, we can lift it to a curve in 3D space, where the third coordinate $z(t)$ is the signed area enclosed by the arc $\sigma_{[0,t]}$ and the segment connecting 0 to $(x(t), y(t))$. Please refer to Figure 1.1 for a visual representation. Namely, we have

$$z(t) := \int_{\sigma_{[0,t]}} \alpha = \int_{\sigma_{[0,t]}} \frac{1}{2}(x\,\mathrm{d}y - y\,\mathrm{d}x) = \int_0^t \frac{1}{2}\left(x(s)\dot{y}(s) - y(s)\dot{x}(s)\right)\,\mathrm{d}s. \tag{1.2.1}$$

Differentiating in t we get

$$\dot{z} = \frac{1}{2}(x\dot{y} - y\dot{x}). \tag{1.2.2}$$

Set $\xi := dz - \frac{1}{2}(x\,\mathrm{d}y - y\,\mathrm{d}x)$. Consider a curve $\gamma = (\gamma_1, \gamma_2, \gamma_3) : [0,1] \to \mathbb{R}^3$ starting at 0. Then we have that such lifted curves are exactly those satisfying $\dot{\gamma} \in \ker(\xi)$, i.e., $\xi((\dot{\gamma}_1, \dot{\gamma}_2, \dot{\gamma}_3)) \equiv 0$.

The differential one-form ξ can be written in cylindrical coordinates (r, θ, z) as $dz - \frac{1}{2}r^2 d\theta$.

Definition 1.2.3. We call the differential one-form

$$\xi := \mathrm{d}z - \frac{1}{2}(x\,\mathrm{d}y - y\,\mathrm{d}x) = \mathrm{d}z - \frac{1}{2}r^2\,\mathrm{d}\theta \tag{1.2.4}$$

the *standard contact form*[1] in $\mathbb{R}^3$.

As every never-vanishing differential one-form on $\mathbb{R}^3$, the standard contact form gives at each point $(x, y, z) \in \mathbb{R}^3$ a 2D kernel inside the tangent space $T_{(x,y,z)}\mathbb{R}^3 \cong \mathbb{R}^3$ at (x, y, z):

$$\Delta_{(x,y,z)} := \ker(\xi_{(x,y,z)}) = \left\{ (v_1, v_2, v_3) \in \mathbb{R}^3 \ : \ v_3 = \frac{1}{2}(xv_2 - yv_1) \right\}.$$

Geometrically, the set Δ is a field of 2D planes in the 3D space, also know as *distribution (of planes)*. Now, given vectors $v = (v_1, v_2, v_3)$ and $w = (w_1, w_2, w_3)$, consider the linear product given by

$$\langle v, w \rangle := v_1 w_1 + v_2 w_2. \tag{1.2.5}$$

[1]A *contact form* on a $(2n+1)$-dimensional differentiable manifold M is a differential 1-form α, with the property that

$$\alpha \wedge (\,\mathrm{d}\alpha)^n \neq 0,$$

with

$$(\,\mathrm{d}\alpha)^n = \underbrace{\mathrm{d}\alpha \wedge \cdots \wedge \mathrm{d}\alpha}_{n}.$$

Sometimes the contact forms $\mathrm{d}z - x\,\mathrm{d}y + y\,\mathrm{d}x = \mathrm{d}z - r^2\,\mathrm{d}\theta$ and $\mathrm{d}z + x\,\mathrm{d}y$ are also called standard.

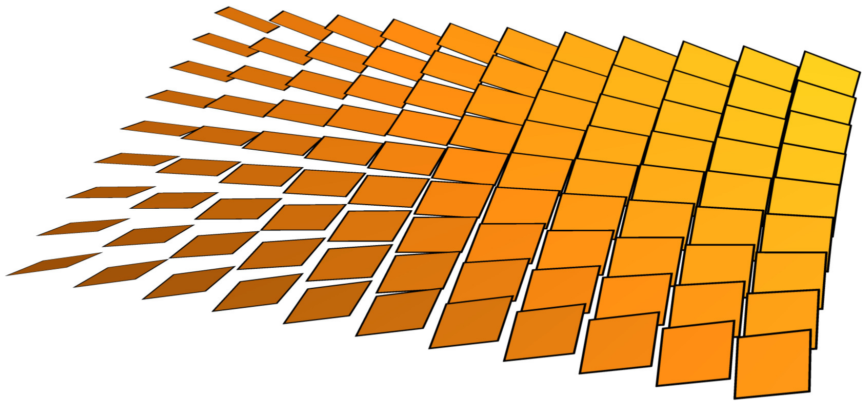


Figure 1.2: Standard contact distribution on $\mathbb{R}^3$.

Notice that, since each plane $\Delta_{(x,y,z)}$ never includes the z-axis, then the restriction of $\langle\cdot,\cdot\rangle$ on $\Delta_{(x,y,z)}$ is a positive-defined inner product. If one prefers, such restriction could be thought as a restriction of a Riemannian tensor on $\mathbb{R}^3$, i.e., a positive-defined inner product on the whole of the tangent bundle of $\mathbb{R}^3$. Indeed, we can fix the following frame[2] of $\mathbb{R}^3$:

$$\begin{cases} X & := & \frac{\partial}{\partial x} - \frac{1}{2}y\frac{\partial}{\partial z}, \\ Y & := & \frac{\partial}{\partial y} + \frac{1}{2}x\frac{\partial}{\partial z}, \\ Z & := & \frac{\partial}{\partial z}, \end{cases} \tag{1.2.6}$$

and declare it orthonormal. Let us verify that such a Riemannian metric gives the linear product (1.2.5) when restricted to the plane $\Delta_{(x,y,z)}$. Indeed, since $\frac{\partial}{\partial x} = X + \frac{1}{2}yZ$ and $\frac{\partial}{\partial y} = Y - \frac{1}{2}xZ$, then we can write

$$v = v_1X + v_2Y + (\frac{v_1}{2}y - \frac{v_2}{2}x + v_3)Z.$$

So, if $v \in \Delta_{(x,y,z)}$, we have $v = v_1X + v_2Y$ and thus (1.2.5) holds.

In contact geometry a curve γ is called *Legendrian* with respect to the differential 1-form ξ if $\xi(\dot\gamma) \equiv 0$. In other words, if the tangent vector $\dot\gamma(t)$ lies in the plane $\Delta_{\gamma(t)}$. Given a Legendrian curve γ, we define its length $L(\gamma)$ as the integral of the norm of $\dot\gamma$ with respect to the scalar product (1.2.5). In other words, $L(\gamma)$ is exactly the Euclidean length of the projection of γ onto the first two components of $\mathbb{R}^3$.

At this point we introduce a new distance on $\mathbb{R}^3$ to which we refer as the *contact distance*. For

[2] A frame is a set of vector fields on a differentiable manifold M that at each point $p \in M$ gives a basis of the tangent space T_pM.

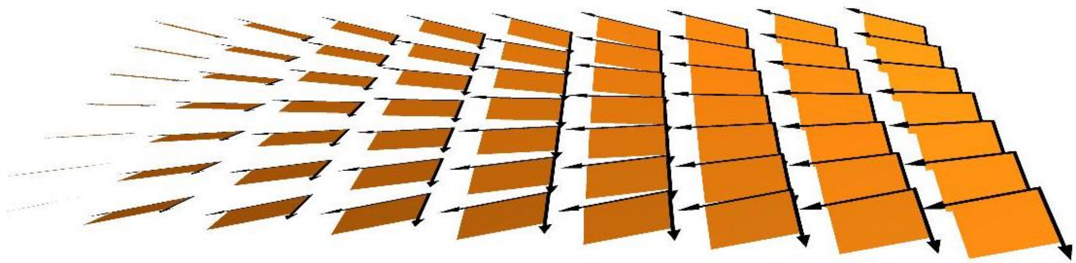


Figure 1.3: The horizontal bundle spanned by the vector fields X and Y.

every p and q in $\mathbb{R}^3$, we define

$$d_c(p,q) := \inf\{L(\gamma) \ : \ \gamma \text{ Legendrian between } p \text{ and } q\}. \tag{1.2.7}$$

The fact that ξ was obtained from the Dido's problem tells us that for every pair of points in $\mathbb{R}^3$ there are several Legendrian curves joining it:

A crucial fact: Every pair of points in $\mathbb{R}^3$ is connected by a curve that is Legendrian with respect to ξ.

Indeed, to connect say $(0,0,0)$ to (x,y,z), it is enough to take a curve σ on $\mathbb{R}^2$ from $(0,0)$ to (x,y) with the property that the signed area enclosed by σ and the segment from $(0,0)$ to (x,y) is exactly z. Then the lifted curve $\tilde{\sigma}$ will connected $(0,0,0)$ to (x,y,z).

Moreover we also know that the length of $\tilde{\sigma}$ equals the planar Euclidean length of σ. Therefore, there is a correspondence between geodesics with respect to the metric d_c (or, better, those curves realizing the infimum on (1.2.7)) and solutions of the 'dual' Dido's isoperimetric problem: fixed a value for the area, minimize the perimeter. Since it is easy to find solutions of Dido's problem we will be able to write explicitly the geodesics of the metric space $(\mathbb{R}^3, d_c)$. We will do this later in Section 1.4.1.

1.3 The Heisenberg group

1.3.1 Heisenberg-group invariance of the standard contact structure

At this point we have introduced a geometry, which we will call *contact geometry*. Specifically, we are considering the plane distribution that at every point $(x,y,z) \in \mathbb{R}^3$ is spanned by the vectors

$$\begin{aligned} X(x,y,z) &:= \frac{\partial}{\partial x} - \frac{y}{2}\frac{\partial}{\partial z} = (1,0,-\frac{y}{2}), \\ Y(x,y,z) &:= \frac{\partial}{\partial y} + \frac{x}{2}\frac{\partial}{\partial z} = (0,1,\frac{x}{2}); \end{aligned} \tag{1.3.1}$$

at each point (x, y, z) we are considering $X(x, y, z)$ and $Y(x, y, z)$ to be an orthonormal basis on their span $\Delta(z, y, z)$; for each smooth curve $\gamma : [a, b] \to \mathbb{R}^3$ for which $\dot\gamma(t)$ is in $\Delta(z, y, z)$ we define its length. Namely, if $u_1(t), u_2(t)$ are such that $\dot\gamma(t) = u_1(t)X_{\gamma(t)} + u_2(t)Y_{\gamma(t)}$, then the length of γ is defined as $\int_a^b \sqrt{u_1(t)^2 + u_2(t)^2}\,\mathrm{d}t$. Such a length structure defined the contact distance (1.2.7).

A crucial property of the contact geometry is that the space is isometrically homogeneous. In fact, the space $\mathbb{R}^3$ can be endowed with a group structure (different from the Euclidean one) in such a way that all of the above constructions are preserved by the action of the group onto itself.

Such a group structure is named after Werner Heisenberg. The group law of this structure is

$$(x, y, z) \cdot (x', y', z') := \left(x + x', y + y', z + z' + \frac{1}{2}(xy' - yx')\right). \tag{1.3.2}$$

One can easily check that (1.3.2) gives a group structure and it turns $\mathbb{R}^3$ into a Lie group, i.e., multiplication and inversion are smooth maps. We will go back to the general theory of Lie groups in Chapter 4. We shall refer to the group $\mathbb{R}^3$ equipped with group law (1.3.2) as the *Heisenberg group.*

We claim that the left translations preserve the distribution Δ and in fact preserve the orthonormal frame X, Y, Z defined by (1.2.6). Let us verify this claim for X. Fix a left translation f, say $f = L_{(s,t,u)}$, for $(s, t, u)\mathbb{R}^3$, i.e.,

$$f(x, y, z) := L_{(s,t,u)}(x, y, z) := (s, t, u) \cdot (x, y, z) \overset{(1.3.2)}{=} \left(x + s, y + t, z + u + \frac{1}{2}(sy - tx)\right). \tag{1.3.3}$$

The differential is

$$\mathrm{d}f = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -t/2 & s/2 & 1 \end{pmatrix}. \tag{1.3.4}$$

So on the one hand we have $\mathrm{d}fX = \frac{\partial}{\partial x} + \left(-\frac{t}{2} - \frac{y}{2}\right)\frac{\partial}{\partial z}$. On the other hand, we have $X \circ f = \frac{\partial}{\partial x} - \frac{1}{2}(t + y)\frac{\partial}{\partial z}$. Therefore $\mathrm{d}fX = X \circ f$, i.e., X is left-invariant. Analogously, $\mathrm{d}fY = \frac{\partial}{\partial y} + \frac{1}{2}(s + x)\frac{\partial}{\partial z} = Y \circ f$ and $\mathrm{d}fZ = \frac{\partial}{\partial z} = Z \circ f$.

As a consequence of the fact that each left translation by the product (1.3.2) preserves the orthonormal frame X, Y we deduce that each such a translation preserves the length of Legendrian curves and, consequently, preserves the contact distance as defined in (1.2.7).

The next proposition summarizes the above discussion.

Proposition 1.3.5. *The Heisenberg geometry is isometrically homogeneous: the space has a Lie group structure so that each left translation is an isometry with respect to the contact distance d_c.*

The aforementioned model of the Heisenberg group has the advantage that its one-dimensional subgroups are easily computable and visually understandable. Specifically, the one-parameter subgroups of this group structure correspond to the standard Euclidean lines passing through the origin.

$$\gamma_v(t) = \exp\left(t(v_1, v_2, v_3)\right) = (tv_1, tv_2, tv_3)\,.$$

In addition, we remark that all the lines through 0 in the xy-plane are curves that minimize the contact distance (Exercise).

1.3.2 The 3D nilpotent non-Abelian matrix group

The Heisenberg group has also a matrix model. It can be seen as a subgroup of the group of invertible matrices. The Heisenberg group is the group of 3×3 upper triangular matrices equipped with the usual matrix product:

$$\mathbb{G} = \left\{ \begin{pmatrix} 1 & a & c \\ 0 & 1 & b \\ 0 & 0 & 1 \end{pmatrix} \;:\; a, b, c \in \mathbb{R} \right\} < GL(3, \mathbb{R}). \tag{1.3.6}$$

Such a model is useful because (first, it is easy to remember the group structure! then) the Lie algebra can be also seen as a matrix group and the exponential of the Lie group is the classical exponential of matrices. Indeed, the Lie algebra is

$$\mathfrak{g} = \left\{ \begin{pmatrix} 0 & a & c \\ 0 & 0 & b \\ 0 & 0 & 0 \end{pmatrix} \;:\; a, b, c \in \mathbb{R} \right\}.$$

A basis of the Lie algebra is

$$X = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \qquad Y = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}, \qquad Z = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \tag{1.3.7}$$

One parameter subgroups are of the form:

$$\begin{aligned} \gamma_{(a,b,c)}(t) &:= \exp\left(t \begin{pmatrix} 0 & a & c \\ 0 & 0 & b \\ 0 & 0 & 0 \end{pmatrix} \right) \\ &= I + t \begin{pmatrix} 0 & a & c \\ 0 & 0 & b \\ 0 & 0 & 0 \end{pmatrix} + \frac{t^2}{2!} \begin{pmatrix} 0 & a & c \\ 0 & 0 & b \\ 0 & 0 & 0 \end{pmatrix}^2 + \dots \\ &= I + t \begin{pmatrix} 0 & a & c \\ 0 & 0 & b \\ 0 & 0 & 0 \end{pmatrix} + \frac{t^2}{2} \begin{pmatrix} 0 & 0 & ab \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} + 0 \\ &= \begin{pmatrix} 1 & at & ct + abt^2/2 \\ 0 & 1 & bt \\ 0 & 0 & 1 \end{pmatrix}. \end{aligned}$$

We claim that the map

$$\varphi : (x, y, z) \mapsto \begin{pmatrix} 1 & x & z + \frac{1}{2}xy \\ 0 & 1 & y \\ 0 & 0 & 1 \end{pmatrix}$$

is a Lie group isomorphism from the Lie group $\mathbb{R}^3$ with the product (1.3.2) to the Lie group $\mathbb{G}$ from (1.3.6) with the usual matrix product. Indeed, the map φ is a group homomorphism (straightforward calculation) and its differential at the identity send the left-invariant vector fields X, Y, Z from (1.2.6) to X, Y, Z from (1.3.7), respectively. In fact, in the next section we will see that more is true.

1.3.3 Characterization of the Heisenberg algebra

The Lie algebra of the Heisenberg group has the property that it is spanned by three vectors X, Y, Z whose only non-trivial Lie bracket relation is $[X, Y] = Z$. In particular, the Lie bracket of every three vectors X_1, X_2, X_3 in this Lie algebra have the property that $[X_1, [X_2, X_3]] = 0$. In other words, the Heisenberg group is a group of nilpotency step 2. Recall that a Lie algebra is nilpotent and its nilpotency step is s if, for all choice of more than s vectors in it, the iterated bracket of them is 0.

We claim that there are only two 3D simply-connected nilpotent Lie groups: the 3D vector group $(\mathbb{R}^3, +)$ and the Heisenberg group. Indeed, consider the Lie algebra $\mathfrak{g}$ of the group. Since $\mathfrak{g}$ is nilpotent, one can take a nonzero element Z in the center of $\mathfrak{g}$. Complete Z to a basis X, Y, Z of $\mathfrak{g}$. Now, either X and Y commute, and so the algebra is commutative, or $W := [X, Y] \neq 0$. write $W = aX + bY + cZ$. Then $[W, Y] = aW$ and so, since $\mathfrak{g}$ is nilpotent, we have $a = 0$. Analogously $b = 0$. Thus $c \neq 0$, and, replacing Z with cZ, we have that the algebra of $\mathfrak{g}$ is defined by the relations:

$$[X, Y] = Z \quad \text{and} \quad [X, Z] = [Y, Z] = 0.$$

We can conclude the proof recalling that there exists a unique simply-connected Lie group with a fixed Lie algebra (see Theorem 4.1.10)

1.4 The subRiemannian Heisenberg group

Our preferred model for the Heisenberg group is $\mathbb{R}^3$ with the product law (1.3.2), which we observed makes left invariant the following vector fields: $\partial_x - \frac{y}{2}\partial_z$, $\partial_y + \frac{x}{2}\partial_z$, ∂_z. The reason why this model is advantageous is that it canonically identifies the group with its Lie algebra (in other words, we are working in exponential coordinates– this viewpoint will be clarified in Section **??**). However, due to the uniqueness of the Heisenberg structure, all other models are equivalent via a smooth group morphism.

Consider three vector fields X, Y, Z on $\mathbb{R}^3$ that are linearly independent at every point and are such that

$$[X,Y]=Z \quad \text{and} \quad [X,Z]=[Y,Z]=0.$$

Then, (it is fact that) there is a group law that makes them left invariant.

We consider the subbundle $\Delta \subset T(\mathbb{R}^3)$ such that for all $p \in \mathbb{R}^3$

$$\Delta_p := \operatorname{span}\{X_p, Y_p\}.$$

A smooth (or, more generally, absolutely continuous) curve $\gamma : [0,1] \to \mathbb{R}^3$ such that $\dot\gamma \in \Delta$ is called *horizontal*. In this case, if we write $\dot\gamma(t) = u_1(t)X_{\gamma(t)} + u_2(t)Y_{\gamma(t)}$, for almost all $t \in [0,1]$ for some integrable functions u_1, u_2 on $[0,1]$, then the *length* of γ is defined as

$$L(\gamma) := \int \sqrt{u_1(t)^2 + u_2(t)^2}\,\mathrm{d}t.$$

We define the *Carnot-Carathéodory distance* between two points $p, q \in \mathbb{R}^3$ as

$$d_{CC}(p,q) := \inf\left\{L(\gamma) : \gamma \text{ horizontal from } p \text{ to } q\right\}.$$

Hence, we have generalized the term Legandrian as horizontal and the notion of contact distance as Carnot-Carathéodory distance. The reason is that since subRiemannian geometry came from different mathematical areas the jargon is multiple.

We say that the above $(\mathbb{R}^3, d_{CC})$ is *(a model for) the subRiemannian Heisenberg group*. In the rest of this section we will work in our favorite model: $\mathbb{R}^3$ with the product law (1.3.2) and orthonormal frame (1.3.1).

1.4.1 Geodesics and spheres in the Heisenberg group

From Section 1.2 and Section 1.3, we have that for a curve $\gamma(t) = (x(t), y(t), z(t))$ has the following properties.

- γ is horizontal (i.e., $\dot\gamma \in \Delta$) if and only if

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

  and this is equivalent to say that $z(t)$ is the area spanned by the curve $(x(\cdot), y(\cdot))$ until t.

- $\dot{\gamma} \in \Delta$ if and only if $\dot{\gamma} = u_1 X + u_2 Y$ where $u_1 = \dot{x}$ and $u_2 = \dot{y}$. Indeed, if $\dot{\gamma} = (\dot{x}, \dot{y}, \dot{z})$, then $\pi(\dot{\gamma}) = (\dot{x}, \dot{y})$ and

$$\pi(\dot{\gamma}) = \pi(u_1 X + u_2 Y) = u_1 \partial_x + u_2 \partial_y = (u_1, u_2).$$

- If $\dot{\gamma} \in \Delta$, then

$$L(\gamma) = \int \sqrt{\dot{x}^2 + \dot{y}^2} = \mathrm{Length}_{\mathrm{Eucl}}(\pi \circ \gamma).$$

Because of this previous discussion, we will obtain explicit formulae for the geodesics in the subRiemannian Heisenberg group by using the fact that we know the solutions of the isoperimetric problem (for which see Appendix A). In fact, we now know that for how the geometry in the Heisenberg group has been constructed, the shortest curves with respect to the length structure are the lifts of the solutions of a variant of the isoperimetric problem. Namely, we search for those shortest curves on the plane that enclose a fixed area and join two given points. Such curves are arc of circles or pieces of lined. Therefore, the geodesics in the Heisenberg group are lifts of circles.

Fact 1.4.1. *Fixed $(x(1), y(1), z(1))$, the curve $(x(t), y(t))$ that encloses area $z(1)$ and such that $(x(0), y(0)) = (0, 0)$ and minimizes $\mathrm{Length}_{\mathrm{Eucl}}(x(\cdot), y(\cdot))$ is a piece of a circle or of a line.*

Thus length-minimizing curves (from $(0, 0, 0)$) are lifts of circles if $z(1) \neq 0$ and straight lines if $z(1) = 0$.

We want to parametrize the curves that are solutions of Dido's problem. A circle of length $\frac{2\pi}{|k|}$, with $k \neq 0$, passing through $(0, 0)$ at time 0 is

$$(x_0(t), y_0(t)) = \left(\frac{\cos(kt) - 1}{k}, \frac{\sin(kt)}{k} \right)$$

for $0 \leq t \leq \frac{2\pi}{|k|}$. Such a circle is parametrization by arc length and has center on the x-axis, on the negative axis if $k > 0$ in the positive axis if $k < 0$.

Notice that if $k > 0$, then the circle (x_0, y_0) encloses positive area, if $k < 0$ it encloses negative area. For $k = 0$, we can still consider the formula in the limit sense: the circles degenerate to the line $(0, t)$, defined for all $t \in \mathbb{R}$.

We obtain every other circle by rotating by an angle $\theta \in \mathbb{R}/2\pi\mathbb{Z}$:

$$R_\theta(x_0(t), y_0(t)) := \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \cdot \begin{pmatrix} \frac{\cos(kt)-1}{k} \\ \frac{\sin(kt)}{k} \end{pmatrix} = \begin{pmatrix} \cos\theta \frac{\cos(kt)-1}{k} - \sin\theta \frac{\sin(kt)}{k} \\ \sin\theta \frac{\cos(kt)-1}{k} + \cos\theta \frac{\sin(kt)}{k} \end{pmatrix}$$

We can calculate the third coordinate as in (1.2.1).

$$\begin{aligned}
z(T) &= \int_0^T \frac{1}{2}(x\,\mathrm{d}y - y\,\mathrm{d}x) = \frac{1}{2}\int_0^T x\dot{y} - y\dot{x} \\
&= \frac{1}{2}\int_0^T \left(\cos\theta\frac{\cos(kt)-1}{k} - \sin\theta\frac{\sin(kt)}{k}\right)(-\sin\theta\sin(kt) + \cos\theta\cos(kt)) + \\
&\qquad - \left(\sin\theta\frac{\cos(kt)-1}{k} + \cos\theta\frac{\sin(kt)}{k}\right)(-\cos\theta\sin(kt) - \sin\theta\cos(kt))\,\mathrm{d}t \\
&= \frac{1}{2k}\int_0^T -\cos\theta(\cos(kt)-1)\sin\theta\sin(kt) + (\cos\theta)^2(\cos(kt)-1)\cos(kt) + \\
&\qquad +(\sin\theta)^2\left(\sin(kt)\right)^2 - \sin\theta\sin(kt)\cos\theta\cos(kt) + \\
&\qquad +\sin\theta(\cos(kt)-1)\cos\theta\sin(kt) + (\sin\theta)^2(\cos(kt)-1)\cos(kt) + \\
&\qquad +(\cos\theta)^2\left(\sin(kt)\right)^2 + \cos\theta\sin(kt)\sin\theta\cos(kt)\,\mathrm{d}t \\
&= \frac{1}{2k}\int_0^T (\cos(kt)-1)\cos(kt) + \left(\sin(kt)\right)^2\,\mathrm{d}t \\
&= \frac{1}{2k}\int_0^T 1 - \cos(kt)\,\mathrm{d}t = \frac{1}{2k^2}(Tk - \sin(kT)).
\end{aligned}$$

We conclude that length-minimizing curves starting from the origin $0 \in \mathbb{R}^3$ are smooth curves $\gamma = (\gamma_1, \gamma_2, \gamma_3)$ of the form

$$\begin{cases} \gamma_1(t) = \cos\theta\frac{\cos(kt)-1}{k} - \sin\theta\frac{\sin(kt)}{k} \\ \gamma_2(t) = \sin\theta\frac{\cos(kt)-1}{k} + \cos\theta\frac{\sin(kt)}{k} \\ \gamma_3(t) = \frac{kt - \sin(kt)}{2k^2} \end{cases} \tag{1.4.2}$$

for some $\theta \in \mathbb{R}/2\pi\mathbb{Z}$ and $k \in \mathbb{R}$.

Such curves are defined for $t \in [0, \frac{2\pi}{|k|}]$ and have length $\frac{2\pi}{|k|}$. When $k = 0$, these curve degenerate to lines:

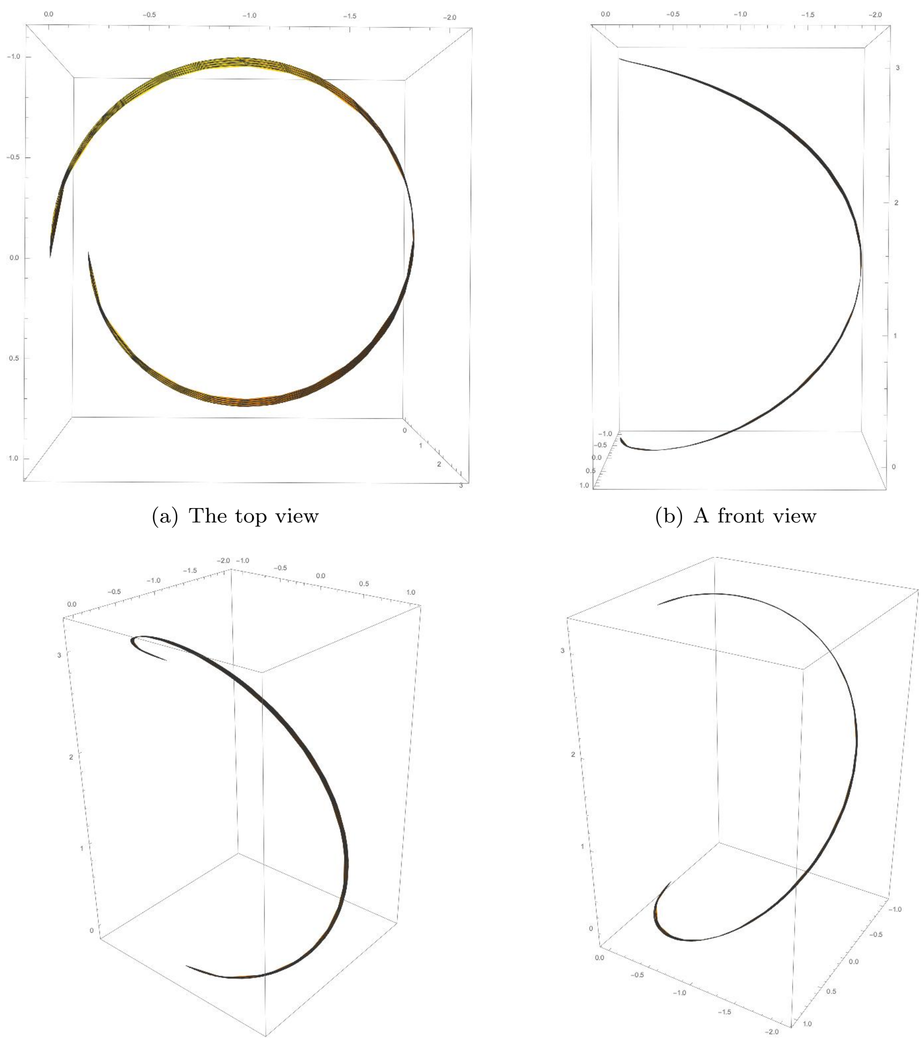
$$\begin{cases} \gamma_1(t) = -t\sin\theta \\ \gamma_2(t) = t\cos\theta \\ \gamma_3(t) = 0, \end{cases}$$

Indeed, lines through the origin in the xy-plane are geodesics.

We found *all* length-minimizing curves in the subRiemannian Heisenberg group. Some consequences of the above characterization of the geodesics are the following facts.

1. If a point $(x, y, z) \in \mathbb{R}^3$ is such that $(x, y) = (0, 0)$, i.e., it is on the z-axis, then there are infinitely many length-minimizing curves between it and the origin $(0, 0, 0)$. In fact, such curves are the one-parameter family of lifts of circles with area z containing $(0, 0)$.

2. If $(x, y) \neq (0, 0)$, then there is a unique length-minimizing curve from (x, y, z) to $(0, 0, 0)$. In fact the curve is the lift of a circular arc enclosing area z together with the segment from $(0, 0)$ to (x, y). Please refer to Figure 1.4.1 for a visual representation.



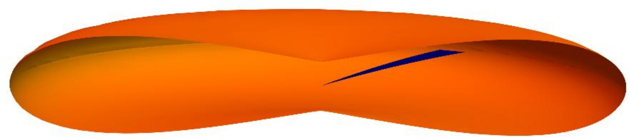
(a) The top view

(b) A front view

(c) A side view

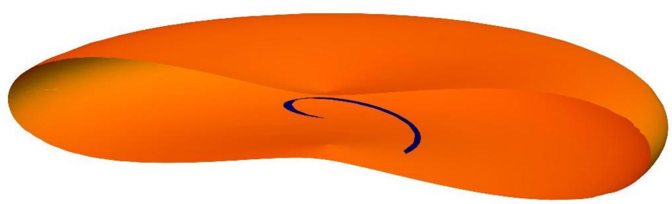
(d) A side view

Figure 1.4: A geodesic with non-zero curvature in the subRiemannian Heisenberg geometry

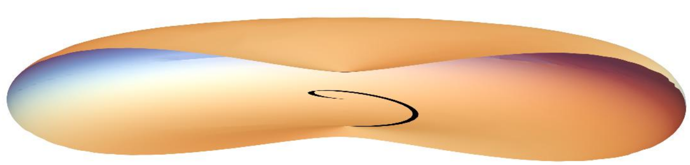


(a) A geodesic with zero curvature

(b) A geodesic with small curvature

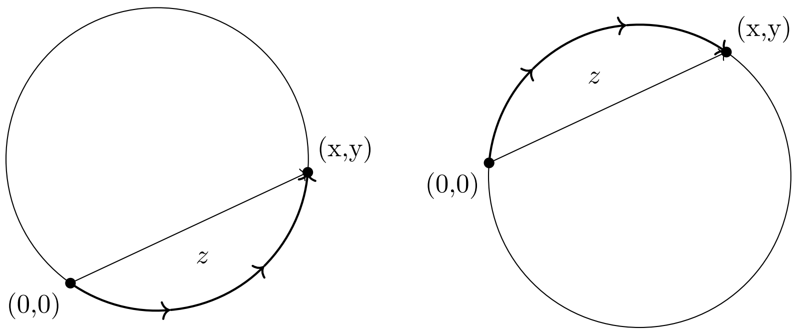


(c) A geodesic with some curvature less than $\frac{1}{2\pi}$.



(d) A geodesic with some curvature equal to $\frac{1}{2\pi}$. It joins points that can be connected with infinitely many geodesics.

Figure 1.5: Geodesics within the unit sphere in the subRiemannian Heisenberg geometry.



(a) Case of $z > 0$. (b) Case of $z < 0$.

Figure 1.6: Projections of minimizing curves from $(0,0,0)$ to (x,y,z) in the Heisenberg model. When the third coordinate z is positive, the curve follows a circe counterclockwise. If z is negative, it follows clockwise. In both cases, the area enclosed by the curve and the circle equals $|z|$.

Since d_{CC} is left-invariant and also $Z = \partial_z$ is left-invariant, we get that for all $p, q \in \mathbb{R}^3$ there exist infinitely many length-minimizing curves between p and q if $\pi(p) = \pi(q)$, i.e., p and q belong to the same vertical line. On the other hand, if $\pi(p) \neq \pi(q)$, then there is only one such a curve.

We deduce that this subRiemannian geometry is not a Riemannian geometry. However, on the one hand, we still have that all the metric balls and metric spheres in the Heisenberg group are topological balls and spheres, respectively, see Exercise 1.5.2. On the other hand, this geometry is not biLipschitz equivalent to any Riemannian geometry, see Corollary 1.4.10.

1.4.2 Dilations on the Heisenberg group

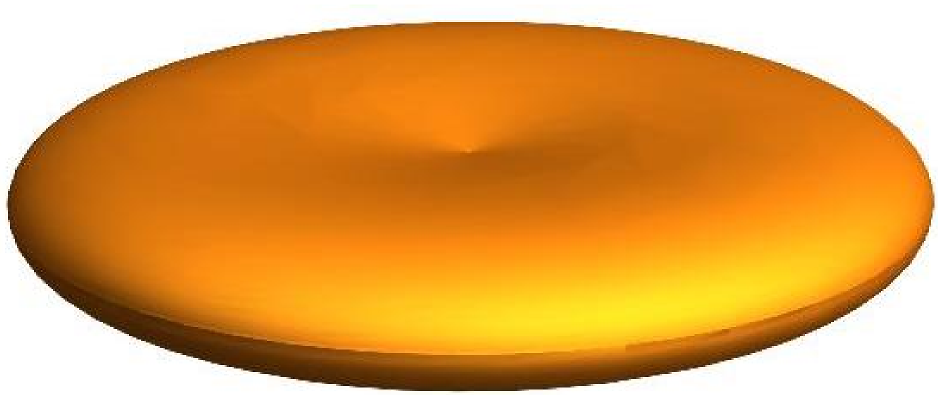
For all $\lambda \in \mathbb{R}$ we define the map

$$\begin{array}{cccc} \delta_\lambda : & \mathbb{R}^3 & \to & \mathbb{R}^3 \\ & (x,y,z) & \mapsto & (\lambda x, \lambda y, \lambda^2 z). \end{array} \tag{1.4.3}$$

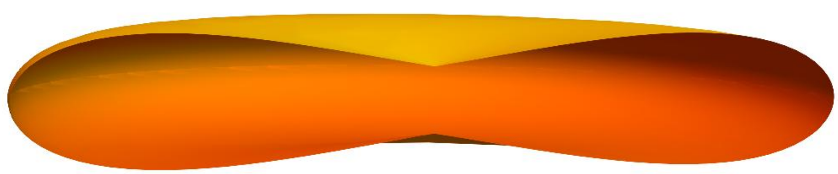
Notice the squared λ in the third component. For $\lambda = 0$ such a map is constantly equal to the origin $\mathbf{0} := (0,0,0)$, which is the identity element for the group law (1.3.2).

Lemma 1.4.4. *For all $\lambda, \mu \in \mathbb{R}$ and $p, q \in \mathbb{R}^3$*

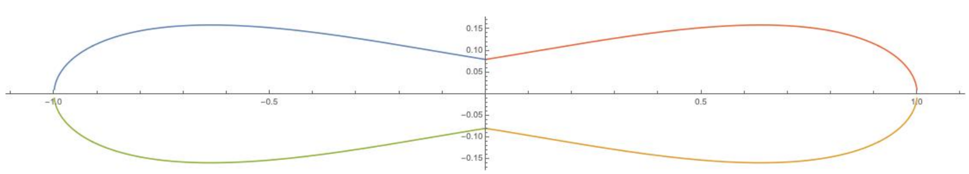
1. $\delta_\lambda(p \cdot q) = \delta_\lambda(p) \cdot \delta_\lambda(q)$;
2. $\delta_\lambda \circ \delta_\mu = \delta_{\lambda\mu}$
3. δ_λ *is a Lie group isomorphism, if* $\lambda \neq 0$;
4. $d_{CC}(\delta_\lambda(p), \delta_\lambda(q)) = |\lambda| d_{CC}(p,q)$.



(a) The unit sphere has a singularity at the intersection with the z-axis.



(b) The portion of the unit sphere in the half-space $\{y > 0\}$.



(c) A section of the sphere as intersection with the xz-plane.

Figure 1.7: Balls in the subRiemannian Heisenberg group are not smooth surfaces. At the two "poles" the sphere is not C^1, there is no cusp, there is a corner. For a parametrization, see Excercise 1.5.2.

Proof. 1. From the group law 1.3.2, we get

$$\delta_\lambda(p\cdot q)=\delta_\lambda\left(p_1+q_1,p_2+q_2,p_3+q_3+\frac{1}{2}(p_1q_2-p_2q_1)\right)=$$
$$=\left(\lambda p_1+\lambda q_1,\lambda p_2+\lambda q_2,\lambda^2p_3+\lambda^2q_3+\frac{1}{2}(\lambda p_1\lambda q_2-\lambda p_2\lambda q_1)\right)=$$
$$=(\lambda p_1,\lambda p_2,\lambda^2p_3)\cdot(\lambda q_1,\lambda q_2,\lambda^2q_3)=\delta_\lambda(p)\cdot\delta_\lambda(q).$$

2. This is obvious from the definition (1.4.3):
$(\delta_\lambda\circ\delta_\mu)(x,y,z)=\delta_\lambda(\mu x,\mu y,\mu^2z)=(\lambda\mu x,\lambda\mu y,\lambda^2\mu^2z)=(\lambda\mu x,\lambda\mu y,(\lambda\mu)^2z)=\delta_{\lambda\mu}(x,y,z).$

3. From the previous points we get that each δ_λ is a group homomorphism and $(\delta_\lambda)^{-1}=\delta_{\frac{1}{\lambda}}$, if $\lambda\neq 0$.

4. Regarding the last point, we shall give three methods of proof, for educational reasons.

Method 1 We claim that the map δ_λ is such that $(\delta_\lambda)_*X=\lambda X$ and $(\delta_\lambda)_*Y=\lambda Y$, where X,Y are the vector fields defining the subbundle Δ. (Check it!) Hence δ_λ preserves horizontal curves and multiplies their length by λ.

Method 2 By (ii) and invariance of d_{CC}, we have

$$d_{CC}(\delta_\lambda(p),\delta_\lambda(q))=d_{CC}((\delta_\lambda(p))^{-1}\cdot\delta_\lambda(q),\mathbf{0})=d_{CC}(\delta_\lambda(p^{-1}q),\mathbf{0}).$$

Hence it is enough to show that

$$d_{CC}(\delta_\lambda(p),\mathbf{0})=\lambda d_{CC}(p,\mathbf{0}).\tag{1.4.5}$$

Let γ be a length minimizing curve from $\mathbf{0}$ to an arbitrary p. Recall that we have an explicit formula for such curves. An easy calculation shows that $\delta_\lambda\circ\gamma$ is still of the same form [3] (up to a linear reparametrization by λ). Hence, its length got multiplied by λ.

Method 3 Reasoning as at the beginning of Method 2, proving (1.4.5) is enough. Take a horizontal curve $\gamma=(x,y,z)$ from $\mathbf{0}$ to p. Notice that the linear map of $\mathbb{R}^2$ represented by the matrix $\begin{pmatrix}\lambda&0\\0&\lambda\end{pmatrix}$ multiplies length by λ and area by λ^2. Therefore, the curve $(\lambda x,\lambda y)$ spans areas

[3] Indeed, if $(\gamma_1,\gamma_2,\gamma_3)$ is a geodesic arc of length 1 starting from the origin, then it is of the form (1.4.2) for some $k\in\mathbb{R}$ with $2\pi/|k|\geq 1$, and the time of the parametrization of (1.4.2) is $t\in[0,1]$. Now the curve $(r\gamma_1,r\gamma_2,r^2\gamma_3)$ is

$$\left(\frac{\cos\theta(\cos(kt)-1)-\sin\theta\sin(kt)}{k/r},\frac{\sin\theta(\cos(kt)-1)+\cos\theta\sin(kt)}{k/r},\frac{kt-\sin(kt)}{2(k/r)^2}\right),\quad\text{for } t\in[0,1],$$

which is a geodesic that is not parametrized by arc length, but by a multiple of it, namely r. Thus its length is r.

that are λ^2 times the areas of (x, y) and has length λ times the length of (x, y). Thus $(\lambda x, \lambda y, \lambda^2 z)$ is horizontal and has length $\lambda L(\gamma)$. Hence $d_{CC}(\delta_\lambda(p), \mathbf{0}) \leq \lambda d_{CC}(p, \mathbf{0})$.

We conclude by arguing similarly with each curve σ joining $\delta_\lambda(p)$ to $\mathbf{0}$ and considering the curve $\delta_{\frac{1}{\lambda}} \circ \sigma$. □

Corollary 1.4.6. *In the subRiemannian Heisenberg group we have*

1. $B_{d_{CC}}(\mathbf{0}, r) = \delta_r(B_{d_{CC}}(\mathbf{0}, 1))$;
2. $B_{d_{CC}}(p, r) = L_p(\delta_r(B_{d_{CC}}(\mathbf{0}, 1)))$,

where $\mathbf{0}$ *is the identity of the group.*

Proof. Easy exercise. □

In other words, we deduce that that if $B_{d_{CC}}((\mathbf{0}, r)$ is the ball of center $\mathbf{0}$ and radius r, then

$$(x, y, z) \in B_{d_{CC}}((\mathbf{0}, 1) \Longleftrightarrow (rx, ry, r^2 z) \in B_{d_{CC}}(\mathbf{0}, r). \tag{1.4.7}$$

Notice that we did not use the homogeneous dilation $\mathbf{v} \mapsto r\mathbf{v}$; the third coordinate has been multiplied by r^2. Thus, such map $(x, y, z) \mapsto (rx, ry, r^2 z)$ multiplies the volume by a factor of r^4, and not r^3 as the usual Euclidean dilation of factor r does!

We can now deduce how is the growth of the balls in the Heisenberg geometry.

Corollary 1.4.8. *Let* vol *be the 3D Lebesgue volume in* $\mathbb{R}^3$*. The Heisenberg subRiemannian distance* d_{CC} *satisfies*

$$\mathrm{vol}(B_{d_{CC}}(p, r)) = r^4 \, \mathrm{vol}(B_{d_{CC}}(\mathbf{0}, 1)) \qquad \forall p \in \mathbb{R}^3 \ \forall r > 0. \tag{1.4.9}$$

Proof. From (1.4.7) we know that $\mathrm{vol}(B(\mathbf{0}, r)) = r^4 \, \mathrm{vol}(B(\mathbf{0}, 1))$. Now we can conclude the proof using both the fact that left translations (1.3.3) in the Heisenberg group are isometries together with the fact that they preserve the volume. This last fact can be checked noticing that the determinant of the differential of a left translations is 1, see (1.3.4). Namely, every left translation L_p is such that $\mathrm{d}L_p = \begin{pmatrix} 1 & 0 & 0 \\ * & 1 & 0 \\ * & * & 1 \end{pmatrix}$ then $\mathtt{Jac}(L_p) = \det(\mathrm{d}L_p) = 1$. Notice that $\mathtt{Jac}(\delta_\lambda) = \det(\mathrm{d}\delta_\lambda) = \det \begin{pmatrix} \lambda & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & \lambda^2 \end{pmatrix} = \lambda^4$. Then

$$\mathrm{vol}(B(p, r)) = \mathrm{vol}(L_p(B(\mathbf{0}, r))) = \mathrm{vol}(B(\mathbf{0}, r)) = \mathrm{vol}(\delta_r(B(\mathbf{0}, 1))) = r^4 \, \mathrm{vol}(B(\mathbf{0}, 1).$$

□

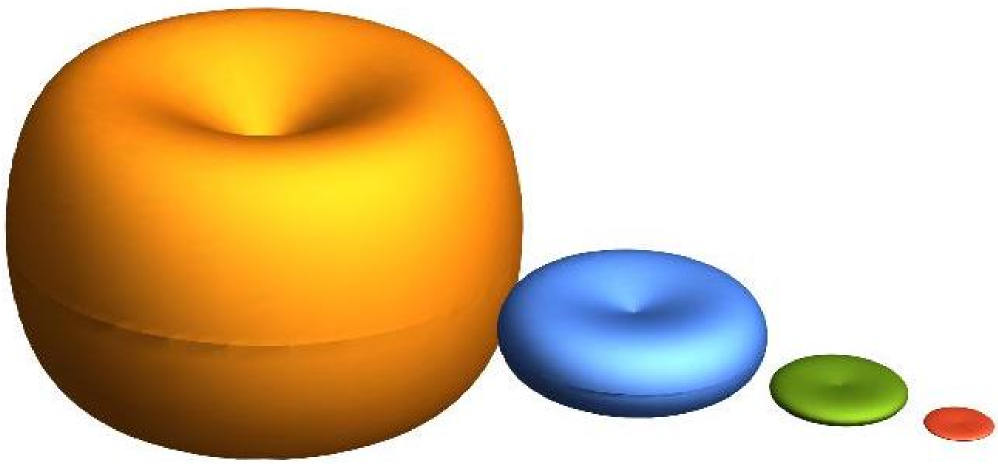


Figure 1.8: Balls of different sizes in the Heisenberg geometry. All the balls are with the origin as center. From the left, these are the balls of radius 2, 1, 1/2, 1/4.

The dimension of the Heisenberg group

The following theorem will tell us that the subRiemannian Heisenberg group is not locally biLipschitz equivalent to any Riemannian manifold. For the notion of biLipschitz and of Hausdorff dimension we refer to Section 2.1.

Corollary 1.4.10. *The Heisenberg group endowed with the standard Carnot-Carathéodory distance has Hausdorff dimension equal to* 4*. In particular, locally this metric space is not biLipschitz equivalent to Euclidean space.*

Proof. From the general Metric Geometry theory which we will revise in Section 2.1.6, it is enough to prove that there are positive constants k_1 and k_2 such that the minimal number N_ϵ of balls of radius ϵ, with $\epsilon \in (0,1)$, needed to cover the unit ball satisfies

$$k_1\epsilon^{-4} < N_\epsilon < k_2\epsilon^{-4}.$$

For the lower bound, let $B_1, \dots, B_{N_\epsilon}$ be such balls. Then, using (1.4.9)

$$\mathrm{vol}(B(0,1)) \leq \sum_{j=1}^{N_\epsilon} \mathrm{vol}(B_j) = N_\epsilon \epsilon^4 \,\mathrm{vol}(B(0,1)).$$

For the upper bound, let $x_1, \dots, x_N$ be a maximal set (which exists by Zorn's Lemma) of points in the unit ball such that the distance between each pair is at least $\epsilon/2$. Hence, the balls $B(x_1, \epsilon/2), \dots, B(x_N, \epsilon/2)$ are disjoint balls of radius $\epsilon/2$ contained in the ball of radius $1 + \epsilon/2$. Then from (1.4.9) we infer that

$$(1+\epsilon/2)^4 \operatorname{vol}(B(0,1)) = \operatorname{vol}(B(0,1+\epsilon/2)) \geq \sum_{j=1}^{N} \operatorname{vol}(B(x_j,\epsilon/2) = N\left(\frac{\epsilon}{2}\right)^4 \operatorname{vol}(B(0,1)).$$

Therefore, using that $\epsilon < 1$, we get that

$$6 > (1+\epsilon/2)^4 \geq N\frac{\epsilon^4}{16}.$$

Now, since the set $\{x_j\}_j$ is maximal, the balls $B(x_j,\epsilon)$, with have same centers but radius ϵ, make up a cover of the unit ball. Thus

$$N_\epsilon \leq N \leq 96\epsilon^{-4}. \qquad \square$$

A ball-box theorem

In this section we give an elementary explanation of why the balls in the subRiemannian Heisenberg geometry behave as boxes with inhomogeneous sides. Namely, let

$$\texttt{Box}(r) := [-r,r]\times[-r,r]\times[-r^2,r^2] \subseteq \mathbb{R}^3. \tag{1.4.11}$$

Proposition 1.4.12. *In the subRiemannian Heisenberg group (in the standard coordinates as above) the balls at the origin satisfy*

$$\texttt{Box}(c_1 r)) \subset B_{cc}(1,r) \subset \texttt{Box}(c_2 r)), \tag{1.4.13}$$

for some universal constants $c_1, c_2 > 0$ and for all $r > 0$.

Proof. In the following argument, we do not aim at the best possible choices for c_1, c_2. Moreover, using the dilations δ_r from the previous section, one can just prove the result for the unit ball and then dilate. The existence of the two boxes (inside and outside) come from the fact that the unit ball is an open bounded set. Nonetheless, we give next a direct proof without any use of the solution of the isoperimetric problem.

First, observe that for all $(x,y,z) \in B_{cc}(1,r)$ we have $|x|, |y| < r$ since the length of a horizontal curve is equal to its projection on the xy-plane, so actually $\|(x,y)\| < r$; and also we claim that we have a bound on z as a function of r. Indeed, we should bound the oriented area enclosed by a curve of length r. Now, we stress that the curve is not closed and the area is a signed area. In other words, the coordinate $z(t)$ satisfies (1.2.2). Hence, for the curve that we are considering (which we might think it is parametrized on the interval $[0,r]$ at unit speed, so that $\dot{y}, \dot{x} \leq 1$) we bound

$$|z(r)| = \left|\int_0^r \frac{1}{2}(x\dot{y} - y\dot{x})\right| \leq \int_0^r \frac{1}{2}(|x|\,|\dot{y}| + |y|\,|\dot{x}|) \leq \int_0^r \frac{1}{2}(r1 + r1) = r^2.$$

We then get

$$B_{cc}(1,r) \subset [-r,r] \times [-r,r] \times [-r^2,r^2], \qquad \forall r > 0.$$

Second, we want to show that the r-ball contains some box. We claim that

$$\left[-\frac{r}{3},\frac{r}{3}\right] \times \left[-\frac{r}{3},\frac{r}{3}\right] \times \left[-\frac{r^2}{100},\frac{r^2}{100}\right] \subset B_{cc}(1,r), \qquad \forall r > 0. \tag{1.4.14}$$

Indeed, take a point (x,y,z) such that $|x|,|y| \leq r/3$ and $|z| \leq r^2/100$. Then consider the following planar curve: starting from $(0,0)$ follow a square of area z (clockwise if $z < 0$, counterclockwise otherwise) then follow the segment from $(0,0)$ to (x,y). This curve encloses area z hence its lift is an admissible curve reaching (x,y,z). The length of the curve is 4 times the side length of the square plus the length of the segment. The square has area at most $r^2/100$ so its side length is at most $r/10$. The segment has length at most $\sqrt{2}r/3$. From these bounds we have $4\frac{r}{10} + \frac{\sqrt{2}r}{3} < r$. Therefore the point (x,y,z) is in the r-ball, so (1.4.14) is verified. □

1.5 Exercises

1. Prove Dido's solution: the maximal area enclosed by a curve of length l on the plane together with a fixed line is $\frac{l^2}{2\pi}$ and it is only obtained as an half disk.

2. Let $\mathrm{vol} = \mathrm{d}x \wedge \mathrm{d}y$ and $\alpha = \dfrac{1}{2}(x\,\mathrm{d}y - y\,\mathrm{d}x)$. Prove

   (a) $\mathrm{d}(\alpha) = \mathrm{vol}$;

   (b) in polar coordinate, we have $\alpha = \dfrac{1}{2}r^2\,\mathrm{d}\theta$;

   (c) if L is a line through the origin, then $\int_L \alpha = 0$.

3. Let σ be a Lipschitz curve on the plane. Let $\sigma_{[0,t]} = (x(t),y(t))$ be the arc up to time t. Let $f : \mathbb{R}^2 \to \mathbb{R}$ be a smooth function. Show that

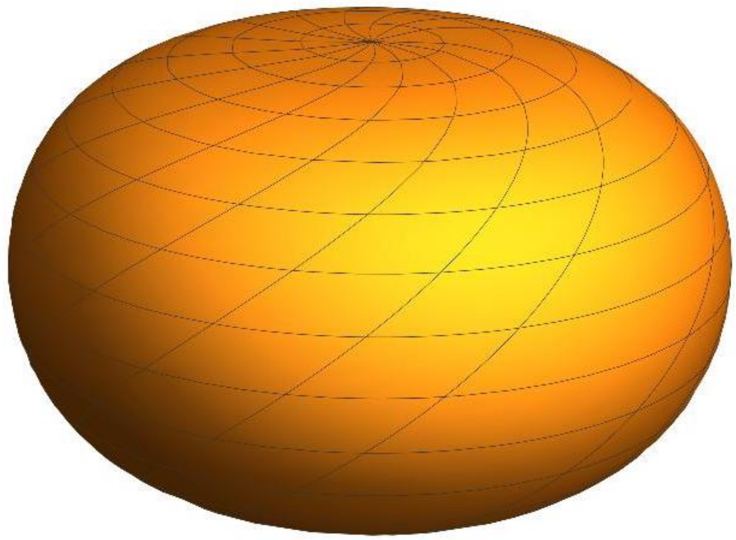
$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\int_{\sigma_{[0,t]}} f(x,y)dx\right) = f(x(t),y(t))\frac{\mathrm{d}x}{\mathrm{d}t}(t), \qquad \text{almost everywhere.}$$

4. Show the relations

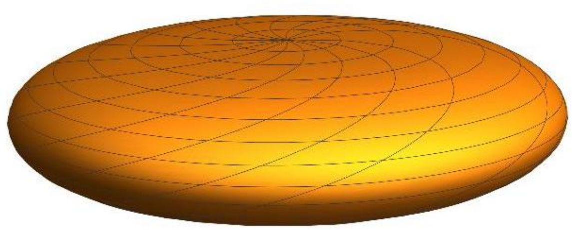
$$[X,Y] = Z \quad \text{ and } \quad [X,Z] = [Y,Z] = 0.$$

   in the following cases:

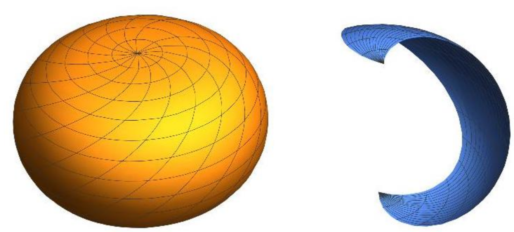
   (a) for the vector fields in (1.2.6),



(a) The so-called Pansu sphere is C^∞ outside of the poles, and C^2 around them. In the above picture the z-axis has been rescaled for aesthetics



(b) Another picture of the Pansu sphere with true axis.



(c) The Pansu sphere is obtained rotating a complete geodesic around the z-axis.

Figure 1.9: The (conjectured) isoperimetric sphere in the subRiemannian Heisenberg geometry

(b) for the matrices (1.3.7).

5. Calculate the inverse of an element (x, y, z) with respect to the group structure given by (1.3.2).

6. Consider the group structure on $\mathbb{R}^3$ given by (1.3.2). Prove that the lines

$$\gamma_v(t) = (tv_1, tv_2, tv_3)\,.$$

are one-parameter subgroups.

7. Let L be a line through 0 in the xy-plane of $\mathbb{R}^3$. Prove that L is a geodesic with respect to the contact distance distance d_c defined in (1.2.7).

8. Consider the map

$$\varphi : (x, y, z) \mapsto \begin{pmatrix} 1 & x & z + \frac{1}{2}xy \\ 0 & 1 & y \\ 0 & 0 & 1 \end{pmatrix}$$

from $\mathbb{R}^3$ with the product (1.3.2) to the space of 3×3 upper triangular matrices with the usual matrix product. Prove that

(a) the map is a Lie group isomorphism,

(b) the map sends the standard basis X, Y, and Z (defined in (1.2.6)) of the first Lie algebra to the standard basis X, Y, and Z (defined in (1.3.7)) of the second Lie algebra.

9. Prove that on the vertical z-axis the distance d_c defined in (1.2.7) is a multiple of the square root of the Euclidean one. Find this multiple.

Exercise 1.5.1. Denote by $C \subset \mathbb{R}^3$ the z-axis. The map

$$\Phi : \left\{(\theta, k, t) \;:\; \theta \in \mathbb{R}/2\pi\mathbb{Z},\; k \in \mathbb{R},\; t \in \left(0, \frac{2\pi}{|k|}\right)\right\} \to \mathbb{R}^3 \setminus C$$

given by

$$\Phi(\theta, k, t) = \left(\frac{\cos\theta(\cos(kt) - 1) - \sin\theta\sin(kt)}{k}, \frac{\sin\theta(\cos(kt) - 1) + \cos\theta\sin(kt)}{k}, \frac{kt - \sin(kt)}{2k^2}\right)$$

is a homeomorphism. ■

Exercise 1.5.2. (i) Let Φ be the map defined in Exercise 1.5.1. Prove that the unit ball in the Heisenberg geometry is given by

$$\begin{aligned} B(0,1) &= \{\Phi(\theta, k, t) \,:\, \theta \in \mathbb{R}/2\pi\mathbb{Z}, k \in \mathbb{R}, t \in (0,1)\} \\ &= \{\Phi(\theta, k, t) \,:\, \theta \in \mathbb{R}/2\pi\mathbb{Z}, k \in [-2\pi, 2\pi], t \in (0,1)\}, \end{aligned}$$

and the unit sphere is

$$S(0,1) = \{\Phi(\theta, k, 1) \,:\, \theta \in \mathbb{R}/2\pi\mathbb{Z}, k \in [-2\pi, 2\pi]\}.$$

(ii) Deduce that all the metric balls and metric spheres in the subRiemannian Heisenberg group are topological balls and spheres, respectively.

■

Chapter 2

A review of metric and differential geometry

The main tools used to study sub-Riemannian geometries are derived from Metric Geometry and Differential Geometry. Metric geometry provides the foundation for understanding distances, geodesics, and intrinsic geometric properties in sub-Riemannian manifolds, including the broader context of Carnot-Carathéodory spaces. Differential geometry plays a central and indispensable role in sub-Riemannian geometry, providing the mathematical framework for studying fundamental geometric objects such as tangent bundles and vector fields. It allows for the analysis of the geometric interpretation of the sub-Riemannian distance as the minimization of a cost functional. This geometric cost functional can be viewed in metric and differential geometry as a length functional defined on curves. To provide a clearer understanding of the setting and terminology, it is essential to have an overview of these key concepts. While there are several excellent books, such as [Fed69, Gro99, AFP00, Hei01, BBI01, AT04], that offer clear and detailed expositions of the material, this discussion aims to provide some insights for non-experts.

2.1 Metric geometry: lengths, geodesics spaces, and Hausdorff measures

2.1.1 Metric spaces

Let M be a set. A function

$$d : M \times M \to [0, +\infty]$$

is called a *distance function* (or just a *distance*, or a *metric*) on M if, for all $x, y, z \in M$, it satisfies

(i) *positiveness*: $d(x, y) = 0 \iff x = y$,

(ii) *symmetry*: $d(x,y) = d(y,x)$,

(iii) *triangle inequality*: $d(x,y) \leq d(x,z) + d(z,y)$.

The pair (M,d) is called *metric space*. If it is clear what metric we are considering or if we do not want to specify the notation for the distance, we shall write just M as an abbreviation for (M,d). We will use the term 'metric' as a synonym of distance function, and never as a shortening of 'Riemannian metric', which will be revised in Section 2.2.3.

A metric space has a natural topology which is generated by the *open balls*

$$B(p,r) := \{q \in M : d(p,q) < r\}, \qquad \forall p \in M, \forall r > 0.$$

In general, we also consider distance functions that may have value ∞. However, on each connected component of the metric space the distance is finite (see Exercise 2.4.1).

For a subset E of a metric space (M,d), the diameter of E is defined as

$$\operatorname{diam}(E) := \sup\{d(p,q) \, : \, p,q \in E\}. \tag{2.1.1}$$

A *curve* (or *path*, or *trajectory*) in a metric space M is a continuous map $\gamma : I \to M$, where $I \subset \mathbb{R}$ is an interval. The interval I may be open, close, half open, bounded or unbounded. When γ is injective, the map might be conflated with its image $\gamma(I)$. We will say that the curve $\gamma : [a,b] \to M$, with $a, b \in \mathbb{R}$, is a *curve from p to q* (or that *joins p to q*) if $\gamma(a) = p$ and $\gamma(b) = q$.

2.1.2 Length of curves in metric spaces

Definition 2.1.2 (Length of a curve)**.** Let M be a metric space with distance function d. The *length (with respect to d)* of a curve $\gamma : [a,b] \to M$ is

$$L(\gamma) := \mathrm{Length}_d(\gamma) := \sup\left\{\sum_{i=1}^{k} d(\gamma(t_{i-1}), \gamma(t_i)) \, : \, k \in \mathbb{N}, a = t_0 < t_1 < \cdots < t_k = b\right\}. \tag{2.1.3}$$

A *rectifiable curve* is a curve with finite length. One might easily check that the length does not depend on the parametrization, see Exercise 2.4.4. A curve $\gamma : [a,b] \to M$ is said to be *parametrized by arc length* if

$$\mathrm{Length}(\gamma|_{[t_1,t_2]}) = |t_2 - t_1|, \qquad \forall t_1, t_2 \in [a,b].$$

Every rectifiable curve admits a reparametrization by arc length, see Exercise 2.4.5, for which it is a 1-Lipschitz map, see Section 2.1.5 for the classical definition of Lipschitz map.

We shall rephrase the definition of length in terms of partitions. A *partition* $\mathcal{P}$ of an interval $[a,b]$ is a k-tuple $(t_1, t_2, \cdots, t_k) \in [a,b]^k$ with $k \in \mathbb{N}$ such that $a = t_1 \leq t_2 \leq \cdots \leq t_k = b$. We set

$$L(\gamma, \mathcal{P}) := \sum_{i=1}^{k-1} d(\gamma(t_{i+1}), \gamma(t_i)).$$

Hence, we have

$$L(\gamma) = \sup\{L(\gamma, \mathcal{P}) \; : \; \mathcal{P} \text{ partition of } [a,b]\}.$$

We next recall the lower semicontinuity of length for sequences of curves that are converging pointwise. A sequence of curves $\gamma_j : [a,b] \to M$ in a metric space M *converges pointwise* to a curve $\gamma : [a,b] \to M$ in the same metric space (note that all such curves have the same interval of definition), if, for all $t \in [a,b]$, we have $\gamma_j(t) \to \gamma(t)$. Furthermore, we say that γ_j *converges uniformly* to γ if $\sup_{t\in[a,b]} d(\gamma_j(t), \gamma(t)) \to 0$, as $j \to \infty$.

Theorem 2.1.4 (Semicontinuity of length)**.** *If $\gamma_j \to \gamma$ pointwise, then $L(\gamma) \leq \liminf_{j\to\infty} L(\gamma_j)$.*

Proof. We could make the result follow from the fact that for each partition $\mathcal{P}$ the function $\gamma \mapsto L(\gamma, \mathcal{P})$ is sequentially continuous (see Exercise 2.4.7) and the general fact that the supremum of sequentially continuous functions is a sequentially lower semicontinuous function (see Exercise 2.4.8). The argument for the proof of the latter fact is the straightforward adaptation of the following argument.

Let $\mathcal{P}$ be a partition of $[a,b]$. Say $\mathcal{P} = (t_1, t_2, \cdots, t_k)$, for some $k \in \mathbb{N}$. Let $\epsilon > 0$. Hence, there exists $N \in \mathbb{N}$ such that, for all $j > N$, we have $d(\gamma_j(t_i), \gamma(t_i)) < \epsilon/k$, for all $i \in \{1, \ldots, k\}$. So by triangle inequality, for all $j > N$, we have

$$d(\gamma(t_{i+1}), \gamma(t_i)) \leq d(\gamma_j(t_{i+1}), \gamma_j(t_i)) + 2\epsilon/k, \qquad \forall i \in \{1, \ldots, k\}.$$

See Figure 2.1 for a visualization. Thus, for all $j > N$, we have

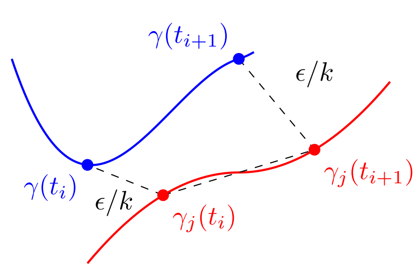


Figure 2.1: The triangle inequality in the proof of Theorem 2.1.4 for bounding $d(\gamma(t_{i+1}), \gamma(t_i))$.

$$L(\gamma, \mathcal{P}) \leq L(\gamma_j, \mathcal{P}) + 2(\epsilon/k) \cdot k \leq L(\gamma_j) + 2\epsilon.$$

Taking the liminf in $j \to \infty$ and considering that ϵ is arbitrarily, we get

$$L(\gamma, \mathcal{P}) \leq \liminf_{j \to \infty} L(\gamma_j).$$

Taking the supremum over all partitions $\mathcal{P}$ we get the result. □

For the purpose of showing the existence of length minimizing curves, we recall now Ascoli-Arzelà Compactness Theorem.

Theorem 2.1.5 (Ascoli-Arzelà)**.** *In compact metric spaces, sequences of curves with uniformly bounded lengths contain subsequences that, up to reparameterization, converge uniformly.*

Proof. Let (M, d) be a compact metric space and γ_n a sequence of curves in M with uniformly bounded length. Because of the bound on the lengths, the curves can be reparametrized with uniformly bounded constant speed to be curves $\gamma_n : [0,1] \to M$ that are uniformly Lipschitz, say L-Lipschitz, see Exercise (2.4.5) and Exercise (2.4.6). The key fact of the argument of Ascoli-Arzelà is that the family $\mathcal{F} = \{\gamma_n : n \in \mathbb{N}\}$ is equi-uniformly continuous (see later) and is equi-uniformly bounded (in our case this is trivial since M is bounded, being compact).

Our aim is show that $\mathcal{F}$ is precompact within the space $C^0([0,1]; M)$ equipped with the uniform convergence, which when considered with the sup-distance is a complete space, see Exercise 2.4.9. It is an exercise in topology [Mun75, Theorem 45.1] to show that in a complete metric space a subset is precompact if and only if it is totally bounded. Namely, by definition of totally bounded, we need to show that for all $\epsilon > 0$ there exists a finite set Λ and, for all $\lambda \in \Lambda$, there exists $\mathcal{F}_\lambda \subset \mathcal{F}$ such that $\mathcal{F} = \cup_\lambda \mathcal{F}_\lambda$ and $\operatorname{diam} \mathcal{F}_\lambda \leq \epsilon$, for all $\lambda \in \Lambda$, where the diameter is defined by (2.1.1).

We start from the the fact that, because of the uniform Lipschitz property, the family $\mathcal{F}$ is *equi-uniformly continuous*, i.e., for every $\varepsilon > 0$ there is $\delta > 0$ such that if $|s-t| < \delta$ then $d(\gamma(t), \gamma(s)) < \varepsilon$, for all $\gamma \in \mathcal{F}$. In fact, in our case it is enough to take $\delta := \varepsilon/L$. Given this $\delta = \delta_\varepsilon$, cover $[0,1]$ with k_ε intervals of radius δ and center $x_i \in [0,1]$, that is, $[0,1] \subset \bigcup_{i=1}^{k_\varepsilon} B(x_i, \delta)$. In addition, since M is compact, there exists $h_\varepsilon \in \mathbb{N}$ and points $p_1, \ldots, p_{h_\varepsilon} \in M$ such that

$$M \subset \bigcup_{i=1}^{h_\varepsilon} B(p_i, \varepsilon).$$

Next define

$$\Lambda := \{\lambda \colon \{1, \ldots, k_\varepsilon\} \to \{1, \ldots, h_\varepsilon\}\}$$

This set is finite, having $h_\varepsilon^{k_\varepsilon}$ elements. We will use it as index-set. For $\lambda \in \Lambda$, define

$$\mathcal{F}_\lambda := \{\gamma \in \mathcal{F} \,:\, |\gamma(x_i) - p_{\lambda(i)}| < \varepsilon \quad \forall i \in \{1, \dots, k_\varepsilon\}\},$$

which is the set of those curves for which the centers of the intervals get mapped into the balls according to λ. Clearly, $\mathcal{F} = \bigcup_{\lambda \in \Lambda} \mathcal{F}_\lambda$, for how we choosed the points p_j. We just need to bound the diameter of $\mathcal{F}_\lambda$. Pick $\alpha, \beta \in \mathcal{F}_\lambda$ and consider their distance, given by the sup-norm. For each $t \in [0,1]$ take i so that $t \in B(x_i, \delta)$. Then

$$\begin{aligned} d(\alpha(t), \beta(t)) &\leq d(\alpha(t), \alpha(x_i)) + d(\alpha(x_i), p_{\lambda(i)}) + d(p_{\lambda(i)}, \beta(x_i)) + d(\beta(x_i), \beta(t)) \\ &< 4\varepsilon, \end{aligned}$$

where we used the equi-uniform continuity of α, β and that $\alpha, \beta \in \mathcal{F}_\lambda$. ☐

Proposition 2.1.6 (Existence of shortest paths)**.** *Let M be a compact metric space. For all $p, q \in M$ there exists a curve γ from p to q such that*

$$L(\gamma) = \inf\{L(\sigma) \,:\, \sigma \textit{ curve from } p \textit{ to } q\}, \tag{2.1.7}$$

provided that the right-hand side of (2.1.7) *is finite.*

Proof. Set L to be the right-hand side of (2.1.7). We are assuming that $L < \infty$. Let γ_j curves from p to q with $L(\gamma_j) \to L$. By Ascoli-Arzelà Theorem 2.1.5, up to passing to a subsequence, we may assume that γ_j converges (uniformly and, hence, pointwise) to a curve γ joining p to q. By semicontinuity of length (Theorem 2.1.4), we get $L(\gamma) \leq \liminf_{j \to \infty} L(\gamma_j) = L$. Hence, we conclude that $L(\gamma) = L$. ☐

2.1.3 Length spaces, intrinsic metrics, and geodesic spaces

If a metric space (M, d) has the property that, for all $p, q \in M$, the value $d(p, q)$ is finite and

$$d(p, q) = \inf\{\mathrm{Length}_d(\gamma) \,:\, \gamma \text{ curve from } p \text{ to } q\},$$

then (M, d) is called *length space* (or *path metric space*) and d is called an *intrinsic metric*. Notice that we have made the choice to require intrinsic metrics to be finite, although this decision may not be shared by all authors in the field.

If a metric space (M, d) is such that, for all $p, q \in M$, there exists a curve γ from p to q with the property that $d(p, q) = \mathrm{Length}_d(\gamma)$, then (M, d) is called *geodesic space*, d is called a *geodesic metric*, and every such a γ is called a *length minimizing curve* joining p to q. Length minimizing curves are also called *length minimizers* or *geodesics*. Some authors use the term 'geodesic' to denote locally length minimizing curves, in agreement with Riemannian geometry.

Every geodesic space is a length space (Exercise 2.4.11). Not all length spaces are geodesic spaces, one reason can be lack of completeness, as for example $\mathbb{R}^2 \setminus \{(0,0)\}$. As we will recall shortly, for locally compact spaces this is the only obstruction.

A metric space is said to be *boundedly compact* (or *proper*) if its bounded subsets are precompact. Equivalently, a space is boundedly compact if its *closed balls*

$$\overline{B}(p,r) := \{q \in M : d(p,q) \leq r\}$$

are compact for all $p \in M$ and all $r > 0$.

Proposition 2.1.8. *Every boundedly compact length space is a geodesic space.*

Proof. Let (M,d) be a boundedly compact length space. Fix $p, q \in M$. Since the metric d is intrinsic, there is a curve γ from p and q with $L(\gamma) < d(p,q) + 1$. Notice that every other curve σ from p and q with $L(\sigma) \leq L(\gamma)$ is inside $\overline{B}(p, d(p,q)+1)$, which is compact. By Proposition 2.1.6, we have the existence of a shortest path and hence of a geodesic joining p to q, since the distance is intrinsic. □

With a little bit more of topological arguments, one can actually prove the following stronger result. An explicit proof can be found in [BBI01, Theorem 2.5.23].

Theorem 2.1.9 (Hopf-Rinow-Cohn-Vossen)**.** *If a length space (M,d) is complete and locally compact, then (M,d) is a geodesic space.*

2.1.4 Length as integral of metric derivative

Throughout the section, we will denote by d the distance function of a metric space $M = (M,d)$.

Definition 2.1.10 (Metric derivative)**.** Given a curve $\gamma\colon [a,b] \to M$ on a metric space, we define the *metric derivative* of γ at the point $t \in (a,b)$ as the limit

$$\lim_{h \to 0} \frac{d(\gamma(t+h), \gamma(t))}{|h|}$$

whenever it exists and, in this case, we denote it by $|\dot{\gamma}|(t)$.

The following is the main result in this subsection:

Theorem 2.1.11. *For each Lipschitz curve $\gamma\colon [a,b] \to M$ on a metric space, we have that*

(i) *the metric derivative $|\dot\gamma|$ exists almost everywhere;*

(ii) $\mathrm{Length}(\gamma) = \int_a^b |\dot\gamma|(t)\,\mathrm{d}t$.

Proof. For part (i), we start by noticing that by the triangle inequality

$$|d(\gamma(s), y) - d(\gamma(t), y)| \le d(\gamma(s), \gamma(t)), \qquad \forall s, t \in [a,b],\ \forall y \in M, \tag{2.1.12}$$

with equality if $y = \gamma(t)$. Fix a countable dense set $\{x_n\}_{n\in\mathbb{N}}$ in $\gamma([a,b])$ and define

$$\varphi_n(t) := d(\gamma(t), x_n).$$

Consequently, from (2.1.12) (and its equality when $x_n \to \gamma(t)$), we have

$$\sup_{n\in\mathbb{N}} |\varphi_n(s) - \varphi_n(t)| = d(\gamma(s), \gamma(t)). \tag{2.1.13}$$

Notice that each $\varphi_n : [a,b] \to \mathbb{R}$ is Lipschitz with same Lipschitz constant as γ, and therefore differentiable almost everywhere and absolutely continuous, by the one-dimensional version of Rademacher Theorem. Let

$$m(t) := \sup_n |\dot\varphi_n(t)|\,.$$

We claim that

$$|\dot\gamma|(t) = m(t), \qquad \text{for almost all } t. \tag{2.1.14}$$

For a first inequality, note that for each point t of differentiability for φ_n, we have from (2.1.13) that

$$|\dot\varphi_n|(t) \overset{\text{def}}{=} \lim_{h\to 0} \frac{|\varphi_n(t+h) - \varphi_n(t)|}{|h|} \overset{(2.1.13)}{\le} \liminf_{h\to 0} \frac{d(\gamma(t+h), \gamma(t))}{|h|}.$$

Hence

$$m(t) \le \liminf_{h\to 0} \frac{d(\gamma(t+h), \gamma(t))}{|h|}.$$

Regarding the other inequality, using the Fundamental Theorem of Calculus, we have for $s \le t$ that

$$\begin{aligned} d(\gamma(t), \gamma(s)) &\overset{(2.1.13)}{=} \sup_n |\varphi_n(t) - \varphi_n(s)| \\ &= \sup_n \left| \int_s^t \dot\varphi_n(\tau)\,\mathrm{d}\tau \right| \\ &\le \sup_n \int_s^t |\dot\varphi_n(\tau)|\,\mathrm{d}\tau \\ &\le \int_s^t m(\tau)\,\mathrm{d}\tau. \end{aligned} \tag{2.1.15}$$

Let us argue why the integral of m is finite. It is because the derivative of each φ_n is bounded from above by the Lipschitz constant of φ_n, which in turn is bounded from above by the one of γ. From Lebesgue's Differentiation Theorem, at each Lebesgue point t for m we have that

$$\limsup_{h\to 0} \frac{d(\gamma(t+h),\gamma(t))}{|h|} \overset{(2.1.15)}{\leq} \limsup_{h\to 0} \left| \frac{1}{h} \int_t^{t+h} m(\tau)\ \mathrm{d}\tau \right| = m(t).$$

So (2.1.14) holds, and in particular $|\dot\gamma|$ exists almost everywhere. The first part of Theorem 2.1.11 is proven.

Regarding the second claim of the theorem, we first prove one inequality. For every partition $(t_1, t_2, \cdots, t_k)$ of $[a,b]$, for some $k \in \mathbb{N}$, we have

$$\sum_{i=1}^{k-1} d(\gamma(t_{i+1}),\gamma(t_i)) \overset{(2.1.15)}{\leq} \sum_{i=1}^{k-1} \int_{t_i}^{t_{i+1}} m(\tau)\ \mathrm{d}\tau \overset{(2.1.14)}{=} \sum_{i=1}^{k-1} \int_{t_i}^{t_{i+1}} |\dot\gamma|(\tau)\ \mathrm{d}\tau = \int_a^b |\dot\gamma|(\tau)\ \mathrm{d}\tau.$$

Taking the supremum over all partitions gives $\mathrm{Length}(\gamma) \leq \int_a^b |\dot\gamma|(t)\ \mathrm{d}t$.

Regarding the other inequality, let $\varepsilon > 0$ and $n \geq 2$ such that $h := (b-a)/n \leq \varepsilon$. We set $t_i := a + ih$, so that $t_n = b$ and $b - \varepsilon < t_{n-1}$. Then

$$\begin{aligned} \int_a^{b-\varepsilon} d(\gamma(t),\gamma(t+h))\ \mathrm{d}t &\leq \sum_{i=1}^{n-1} \int_{t_{i-1}}^{t_i} d(\gamma(t),\gamma(t+h))\ \mathrm{d}t \\ &= \int_0^h \sum_{i=1}^{n-1} d(\gamma(\tau+t_{i-1}),\gamma(\tau+t_i))\ \mathrm{d}\tau \\ &\leq \int_0^h \mathrm{Length}(\gamma)\ \mathrm{d}\tau = h\,\mathrm{Length}(\gamma). \end{aligned} \tag{2.1.16}$$

Using Fatou's lemma:

$$\begin{aligned} \int_a^{b-\varepsilon} |\dot\gamma|(t)\ \mathrm{d}t &\overset{\text{def}}{=} \int_a^{b-\varepsilon} \liminf_{h\to 0} \frac{d(\gamma(t+h),\gamma(t))}{h}\ \mathrm{d}t \\ &\overset{\text{Fatou}}{\leq} \liminf_{h\to\infty} \frac{1}{h} \int_a^{b-\varepsilon} d(\gamma(t+h),\gamma(t))\ \mathrm{d}t \overset{(2.1.16)}{\leq} \mathrm{Length}(\gamma). \end{aligned}$$

Letting $\varepsilon \to 0^+$ gives the missing inequality. $\square$

Example 2.1.17. A first interesting example is given when the metric space (M,d) is a finite-dimensional normed space $(V, \|\cdot\|)$ with the metric d induced by $\|\cdot\|$, i.e., $d(p,q) := \|p-q\|$, for all $p, q \in V$. Let $\gamma : [a,b] \to V$ be an absolutely continuous curve. Up to reparametrizing, we assume that γ is a Lipschitz curve (either with respect to the distance d or with respect to any other Euclidean distance). Hence, by Rademacher Theorem, the curve γ is differentiable almost everywhere. For every point of differentiability t for γ, we have

$$\|\gamma'(t)\| = \left\| \lim_{h\to 0} \frac{\gamma(t+h)-\gamma(t)}{h} \right\| = \lim_{h\to 0} \frac{\|\gamma(t+h)-\gamma(t)\|}{|h|} \overset{\text{def}}{=} |\dot\gamma|(t),$$

where $|\dot{\gamma}|(t)$ is the metric derivative and $\gamma'(t)$ denotes the (classical) derivative. Consequently, from Theorem 2.1.11 we infer

$$\mathrm{Length}_d(\gamma) = \int_a^b \|\gamma'(t)\| \, \mathrm{d}t. \tag{2.1.18}$$

We deduce that for every two points $p, q \in V$ and every rectifiable curve γ between p and q we have

$$\|p - q\| \overset{\text{def}}{=} d(p, q) \leq \mathrm{Length}_d(\gamma) = \int_a^b \|\gamma'(t)\| \, \mathrm{d}t. \tag{2.1.19}$$

We stress that with the curve $t \in [0, 1] \mapsto tp + (1 - t)q$, we get equality in (2.1.19). In conclusion, we have proved that every finite-dimensional normed space is a geodesic space with straight lines being geodesics.

Energy functional

In geometric analysis, it is often more appropriate to consider the energy of curves rather than their length. The reason is that the energy functional often possesses better analytic and geometric properties compared to the length functional. It may be smoother and more amenable to analysis, allowing for the application of variational techniques and optimization methods.

Let $\gamma : [a, b] \to M$ be a Lipschitz curve on a metric space (M, d). Hence, by Theorem 2.1.11 its metric derivative $|\dot{\gamma}|$ exists almost everywhere. The *energy* of γ (with respect to the distance d) is defined as

$$\mathrm{Energy}_d(\gamma) := \frac{1}{2} \int_a^b (|\dot{\gamma}|(t))^2 \, \mathrm{d}t \tag{2.1.20}$$

On the contrary of length, energy depends on the parametrization of the curve. However, we shall now see that parametrizations with constant speed minimize the energy among all of the reparametrizations of the curve, and in that case the energy is a precise function of the length.

Proposition 2.1.21. *Let $\gamma : [a, b] \to M$ be a Lipschitz curve on a metric space (M, d) and $p, q \in M$. Then the energy satisfies the following properties:*

2.1.21.i. $\mathrm{Length}_d(\gamma) \leq \sqrt{2 \cdot \mathrm{Energy}_d(\gamma)}$.

2.1.21.ii. *If γ is parametrized by a multiple of the arc length, then* $\mathrm{Length}_d(\gamma) = \sqrt{2 \cdot \mathrm{Energy}_d(\gamma)}$.

2.1.21.iii. $\inf\{\mathrm{Length}_d(\gamma) \ : \ \gamma \textit{ from } p \textit{ to } q\} = \inf\{\sqrt{2 \cdot \mathrm{Energy}_d(\gamma)} \ : \ \gamma \textit{ Lipschitz, from } p \textit{ to } q\}$.

2.1.21.iv. *A curve parametrized by arc length from p to q is length-minimizing if and only if it is energy minimizing.*

Proof. It is an easy exercise using Jensen's inequality or Cauchy–Schwarz inequality. □

Remark 2.1.22. Actually, minimizing the length is equivalent to minimizing the *p-energy*, for each $p \in]1, \infty[$. For the p-energy one takes the p-power of the L^p-norm of the metric derivative, up to a constant. Again, the equivalence follows from Hölder's inequality.

2.1.5 Isometries and Lipschitz maps

Given two metric spaces (X, d_X) and (Y, d_Y), a map $f : X \to Y$ is called *Lipschitz* if there exists a real constant $K \geq 0$ such that

$$d_Y(f(x_1), f(x_2)) \leq K d_X(x_1, x_2), \qquad \forall x_1, x_2 \in X.$$

The value K (or many times the smallest value of such K's) is called a (or the) *Lipschitz constant* of the function f. A function is called *locally Lipschitz* if for every $x \in X$ there exists a neighborhood U of x such that f restricted to U is Lipschitz.

If there exists a $K \geq 1$ with

$$\frac{1}{K} d_X(x_1, x_2) \leq d_Y(f(x_1), f(x_2)) \leq K d_X(x_1, x_2), \qquad \forall x_1, x_2 \in X,$$

then f is called *biLipschitz embedding* (also written bi-Lipschitz or bilipschitz). Surjective biLipschitz embeddings are called *biLipschitz homeomorphisms* (or biLipschitz maps). BiLipschitz homeomorphisms are the isomorphisms in the category of Lipschitz maps. To be more explicit on the value of the constant K we would say that f is K-biLipschitz. BiLipschitz embeddings are injective and in fact embeddings, i.e., they are homeomorphisms onto their image. We call 1-biLipschitz homeomorphisms *isometries*; while 1-biLipschitz embeddings are *isometric embeddings*.

Two functions α, β defined on the same set X are *biLipschitz equivalent* if there exists $K > 1$ such that

$$\frac{1}{K} \alpha(x) \leq \beta(x) \leq K \alpha(x), \qquad \forall x \in X.$$

Two important examples of functions for which we will consider biLipschitz equivalence will be distances and measures. Notice that in particular, two distances d_1, d_2 on the same set M are *biLipschitz equivalent* if and only if the identity map (M, d_1) to (M, d_2) is biLipschitz.

2.1.6 Hausdorff measures and dimension

Recall that a collection $\mathscr{F}$ of subset of an arbitrary set X is called *σ-algebra* for X if

(i) $\emptyset, X \in \mathscr{F}$;

(ii) $A, B \in \mathscr{F} \Rightarrow A \setminus B \in \mathscr{F}$;

(iii) $\{A_n\}_{n\in\mathbb{N}} \subset \mathscr{F} \Rightarrow \bigcup_{n\in\mathbb{N}} A_n \in \mathscr{F}$.

If X is a topological space, the smallest σ-algebra containing the open sets is called *Borel σ-algebra*.

Definition 2.1.23 (Measure)**.** A *measure* on a σ-algebra $\mathscr{F}$ is a function $\mu : \mathscr{F} \to [0, +\infty]$ such that

2.1.23.i $\mu(\emptyset) = 0$;

2.1.23.ii $\{A_n\}_{n\in\mathbb{N}} \subset \mathscr{F}$, pairwise disjoint $\Rightarrow \mu(\bigcup_{n\in\mathbb{N}} A_n) = \sum_{n=0}^{\infty} \mu(A_n)$.

The latter condition is called *σ-additivity*.

Every measure has the property of being *countably subadditive* on arbitrary elements of $\mathscr{F}$, i.e., if $\{A_n\}_{n\in\mathbb{N}} \subset \mathscr{F}$ then (see Exercise 2.4.13)

$$\mu\left(\bigcup_{n\in\mathbb{N}} A_n\right) \leq \sum_{n=0}^{\infty} \mu(A_n).$$

A measure on a topological space is called a *Borel measure* if μ is defined on the Borel σ-algebra. Hence, if μ is a Borel measure on a metric space M, then $\mu(B_M(p, r))$ is defined for all $p \in M$ and all $r > 0$.

For next definition, we use the notion of diameter from (2.1.1).

Definition 2.1.24 (Hausdorff measures)**.** Let M be a metric space. Let $S \subset M$ be a subset, $Q \in [0, \infty)$, and $\delta > 0$. The *Q-dimensional Hausdorff δ-content* is defined as

$$\mathcal{H}^Q_\delta(S) := \inf\left\{\sum_{i=1}^{\infty} (\operatorname{diam}(E_i))^Q : E_i \subseteq S,\ \operatorname{diam} E_i < \delta,\ S \subseteq \bigcup_{i=1}^{\infty} E_i\right\}, \tag{2.1.25}$$

with convention that $0^0 = 1$. Notice that the function $\delta \mapsto \mathcal{H}^Q_\delta(S)$ is non-increasing. The *Q-dimensional Hausdorff measure* of S is defined as

$$\mathcal{H}^Q(S) := \sup_{\delta>0} \mathcal{H}^Q_\delta(S) = \lim_{\delta\to 0^+} \mathcal{H}^Q_\delta(S).$$

Each measure $\mathcal{H}^Q$ is an outer measure, see [Fol99], that restricted to the Borel σ-algebra gives a measure.

Exercise 2.1.26. If $F : M_1 \to M_2$ is an L-Lipschitz map, $Q \geq 0$ and $S \subset M_1$, then

$$\mathcal{H}^Q(F(S)) \leq L^Q \mathcal{H}^Q(S).$$

■

Proposition 2.1.27. *Let M be a metric space. Then there exists $Q_0 \in [0, +\infty]$ such that*

$$\mathcal{H}^Q(M) = 0 \quad \forall Q > Q_0 \quad \textit{and} \quad \mathcal{H}^Q(M) = \infty \quad \forall Q < Q_0.$$

Proof. Set

$$Q_0 := \inf\{Q \geq 0 : \mathcal{H}^Q(M) \neq \infty\}.$$

Hence $\mathcal{H}^Q(M) = \infty$ for all $Q < Q_0$.

If $Q_0 = \infty$, then there is nothing else to prove. If $Q_0 < \infty$, then take $Q > Q_0$. Then there is $Q' \in [Q_0, Q)$ with $\mathcal{H}^{Q'}(M) =: K < \infty$. Hence for all $\delta \in (0,1)$ we have $\mathcal{H}^{Q'}_\delta(M) \leq K$, i.e., there are $E_i \subset M$ with $M = \bigcup_i E_i$, $\operatorname{diam}(E_i) < \delta$ and $\sum_i \operatorname{diam}(E_i)^{Q'} < K + 1$. Notice that

$$\sum \operatorname{diam}(E_i)^Q \leq \delta^{Q-Q'} \sum_i \operatorname{diam}(E_i)^{Q'} < (K+1)\delta^{Q-Q'}.$$

Thus $\mathcal{H}^Q_\delta(M) \leq (K+1)\delta^{Q-Q'}$. Since $\delta^{Q-Q'} \to 0$ as $\delta \to 0^+$, we get $\mathcal{H}^Q(M) = 0$. □

Definition 2.1.28 (Hausdorff dimension)**.** The *Hausdorff dimension* of a metric space M is denoted by $\dim_H(M)$ and is equivalently defined as

$$\begin{aligned}
\dim_H(M) &:= \inf\{Q \geq 0 : \mathcal{H}^Q(M) = 0\} \\
&= \inf\{Q \geq 0 : \mathcal{H}^Q(M) \neq \infty\} \\
&= \sup(\{Q \geq 0 : \mathcal{H}^Q(M) = \infty\} \cup \{0\}).
\end{aligned}$$

The above definitions are equivalent because of Proposition 2.1.27.

Exercise 2.1.29. If $F : M_1 \to M_2$ is a biLipschitz homeomorphism, then $\dim_H M_1 = \dim_H M_2$. ■

Theorem 2.1.30. *Let M be a metric space and μ a Borel measure on M. Assume that there are $Q > 0$, $C > 1$, and $R > 0$ such that*

$$\frac{1}{C} r^Q \leq \mu(B(p,r)) \leq C r^Q, \qquad \forall p \in M, \ \forall r \in (0, R]. \tag{2.1.31}$$

Then for all $p \in M$

(i) $\mathcal{H}^Q(B(p,R)) \in (0,\infty)$,

(ii) $\dim_H B(p,R) = Q$,

and, if in addition M admits a countable cover of balls of radius R, then $\dim_H M = Q$.

Proof. Fix $p \in M$. We first show that $\mathcal{H}^Q(B(p,R)) < \infty$. Fix $r \in (0,R)$ and let $0 < \delta < R - r$. We claim that we can take a finite maximal family of points $p_1, \ldots, p_N \in B(p,r)$ such that $d(p_i,p_j) > \delta$ for all $i \neq j$. Indeed, such a finite set of points exists, because if $p_1, \ldots, p_k \in B(p,r)$ are such that $d(p_i,p_j) > \delta$, then the balls $B(p_i, \frac{\delta}{2})$ are disjoint and contained in $B(p,R)$, hence

$$k\frac{\delta^Q}{2^Q C} \;=\; \frac{1}{C}\sum_{i=1}^{k}\left(\frac{\delta}{2}\right)^Q \;\leq\; \sum_{i=1}^{k}\mu\left(B(p_i,\frac{\delta}{2})\right) \;=\; \mu\left(\bigcup_{i=1}^{k} B(p_i,\frac{\delta}{2})\right) \;\leq\; \mu(B(p,R)) \;\leq\; CR^Q.$$

Therefore the integer k has to be bounded and such a maximal set of points is finite.

Maximality implies that $B(p_1,\delta), \ldots, B(p_N,\delta)$ cover $B(p,r)$. Hence, we bound

$$\begin{aligned}
\mathcal{H}^Q_{2\delta}(B(p,r)) &\leq \sum_{j=1}^{N}(\operatorname{diam}(B(p_j,\delta)))^Q \\
&\leq N(2\delta)^Q = 4^Q C N \frac{1}{C}\left(\frac{\delta}{2}\right)^Q \\
&\leq 4^Q C \sum_{j=1}^{N}\mu\left(B(p_j,\frac{\delta}{2})\right) \\
&\leq 4^Q C\mu(B(p,R)),
\end{aligned}$$

where in the second inequality we used that the diameter of a ball is at most twice its radius. We stress that the last term is finite and independent on δ. Finally, for the ball of radius R we have $\mathcal{H}^Q(B(p,R)) = \mathcal{H}^Q(\bigcup_{r<R} B(p,r)) \leq 4^Q C\mu(B(p,R)) < \infty$, where we have used that the measure is continuous with respect to the increasing union of sets, see Exercise 2.4.18.

We then show that $\mathcal{H}^Q(B(p,R)) > 0$. Let $\delta \in (0,R)$. To bound from below the δ-Hausdorff content take $\epsilon > 0$ and countably many sets $E_1, E_2, \ldots \subset M$ such that $\operatorname{diam}(E_i) < \delta$, $B(p,R) \subset \bigcup_i E_i$ and

$$\mathcal{H}^Q_\delta(B(p,R)) \geq \sum_i (\operatorname{diam} E_i)^Q - \epsilon.$$

Such a cover exists because $\mathcal{H}^Q(B(p,R)) < \infty$. Take some $p_i \in E_i$, so $E_i \subset B(p_i, \operatorname{diam}(E_i))$ and

$$\mu(B(p_i, \operatorname{diam}(E_i))) \leq C \operatorname{diam}(E_i)^Q.$$

Thus, by the countably subadditivity of μ, we have, since $\bigcup_i B(p_i, \mathrm{diam}(E_i)) \supset \bigcup_i E_i \supset B(p,R)$,

$$\begin{aligned}
\mathcal{H}^Q_\delta(B(p,R)) &\geq \frac{1}{C}\sum_i \mu(B(p_i, \mathrm{diam}\, E_i)) - \epsilon \\
&\geq \frac{1}{C}\mu\left(\bigcup_i B\left(p_i, \mathrm{diam}(E_i)\right)\right) - \epsilon \\
&\geq \frac{1}{C}\mu(B(p,R)) - \epsilon \\
&\geq \frac{1}{C^2}R^Q - \epsilon
\end{aligned}$$

Since ϵ was arbitrary, we get that $\mathcal{H}^Q_\delta(B(p,R))$ is greater than a positive constant independent of δ.

So (i) is proved and (ii) is an immediate consequence. By countable subadditivity of the Hausdorff measure, also the last statement of the theorem follows. □

Remark 2.1.32. The above proof actually shows that the Q-dimensional Hausdorff measure $\mathcal{H}^Q$ is biLipschitz equivalent to the measure μ. In particular, the measure $\mathcal{H}^Q$ satisfies equation (2.1.31), with possibly some other choice for the constant C. We shall rephrase the last theorem using the following definition.

Definition 2.1.33 (Ahlfors regularity for measures)**.** A measure μ a on a metric space that is Borel and for which there are $Q \in (0,\infty)$, $C > 1$, and $R > 0$ such that

$$\frac{1}{C}r^Q \leq \mu(B(p,r)) \leq Cr^Q, \qquad \forall p \in M,\ \forall r \in (0,R], \tag{2.1.34}$$

is said to be *Ahlfors Q-regular up to scale R.*

Corollary 2.1.35. *If a metric space supports a measure that is Ahlfors Q-regular up to scale R, then the Q-dimensional Hausdorff measure $\mathcal{H}^Q$ of the metric space is Ahlfors Q-regular up to scale R, and the R-balls have Hausdorff dimension Q.*

Using the Hausdorff measure we rephrase the notion of length for (injective) curves.

Proposition 2.1.36. *If $\gamma : I \to M$ is an injective curve on a metric space M, then we have*

$$\mathcal{H}^1(\gamma(I)) = \mathrm{Length}(\gamma). \tag{2.1.37}$$

Proof. We shall focus on the case when $\mathrm{Length}(\gamma) < \infty$ and leave the other case as an exercise. Thus, we reparametrize $\gamma : [0,\ell] \to M$ by arc length. For proving (2.1.37), we shall consider one inequality at a time.

For the inequality $\leq$, for each $\delta > 0$ divide the interval $[0, \ell]$ into n disjoint intervals $J_1, \ldots, J_n$ of diameter less than δ. Since γ is parametrized by arc length, then it is 1-Lipschitz and therefore we have $\operatorname{diam} \gamma(J_j) < \delta$, for $j = 1, \ldots, n$. Hence

$$\begin{aligned} \mathcal{H}^1_\delta(\gamma([0,\ell])) &\leq \sum_{j=1}^{n} \operatorname{diam} \gamma(J_j) \\ &\leq \sum_{j=1}^{n} \operatorname{diam} J_j = \ell, \end{aligned}$$

where we have used in the first inequality that $(\gamma(J_j))_j$ is a admissible cover for (2.1.25) and in the second inequality that γ is 1-Lipschitz. Taking the limit for $\delta \to 0$, we infer the desired inequality in (2.1.37).

For the inequality $\geq$, we shall use Exercise 2.1.38. In fact, take a partition $t_0 < t_1 < \ldots < t_k$ of the interval I. Then we bound

$$\begin{aligned} \sum_{i=1}^{k} d(\gamma(t_{i-1}), \gamma(t_i)) &\leq \sum_{i=1}^{k} \mathcal{H}^1(\gamma([t_{i-1}, t_i])) \\ &\leq \mathcal{H}^1(\gamma(I)), \end{aligned}$$

where in the last inequality we have used that $\mathcal{H}^1$ is additive and that γ is injective. □

Exercise 2.1.38 (to be generalized in Exercise 2.1.40)**.** For every continuous curve $\gamma : [a,b] \to M$ on a metric space M, we have

$$\mathcal{H}^1(\gamma([a,b])) \geq d(\gamma(a), \gamma(b)).$$

Solution. Consider $\phi(x) := d(x, \gamma(a))$, which is 1-Lipschitz. Then, using that on $\mathbb{R}$ the measure $\mathcal{H}^1$ coincides with Lebesgue measure, bound $\mathcal{H}^1(\gamma([a,b])) \geq \mathcal{H}^1(\phi(\gamma([a,b]))) \geq \operatorname{diam}(\phi(\gamma([a,b]))) \geq d(\gamma(a), \gamma(b))$. ■

Exercise 2.1.39. Complete the proof of Proposition 2.1.36 by showing that for every curve $\gamma : I \to M$ on a metric space if $\operatorname{Length}(\gamma) = \infty$, then $\mathcal{H}^1(\gamma(I)) = \infty$. *Hint.* Use Exercise 2.1.38. ■

Exercise 2.1.40. For every connected subset X of a metric space, we have $\mathcal{H}^1(X) \geq \operatorname{diam}(X)$. ■

2.2 Differential geometry

2.2.1 Vector fields and Lie brackets

In this section, we will denote by M a smooth differentiable manifold. We will not review here the definition of a manifold and the concept of a smooth map between manifolds. We denote by $C^\infty(M)$

the space of C^∞ functions from M to $\mathbb{R}$. We shall prefer the following view point for the space of smooth vector fields on M: A linear function $X : C^\infty(M) \to C^\infty(M)$ is a *smooth vector field* on M if it satisfies the *Leibniz rule*:

$$X(fg) = X(f)g - fX(g), \qquad \forall f, g \in C^\infty(M).$$

We denote by $\mathrm{Vec}(M)$ or by $\Gamma(TM)$ the linear space of smooth vector fields; we will typically use the letter X, Y, Z to denote elements in $\mathrm{Vec}(M)$.

Definition 2.2.1 (Vector fields in charts)**.** Let $\varphi : U \to \mathbb{R}^n$ be a coordinate chart for an n-manifold M. For $j \in \{1, \dots, n\}$ we define the *j-th coordinate vector field* $\partial_j \in \mathrm{Vec}(U)$ by

$$\partial_j(f)(p) := \frac{\partial(f \circ \varphi^{-1})}{\partial x_j}(\varphi(p)) = \frac{\mathrm{d}}{\mathrm{d}t} f(\varphi^{-1}(\varphi(p) + te_j))\Big|_{t=0}, \quad \forall f \in C^\infty(U), \forall p \in U,$$

where e_j denotes the j-th element of the canonical basis of $\mathbb{R}^n$.

Given a chart (U, φ) for M, every vector field X on M restricted to U can be written using the coordinate vector fields as

$$X = \sum_{j=1}^n X^j \partial_j, \qquad \text{on } U,$$

for some smooth functions $X^j \in C^\infty(U)$. Namely, we have $X(f)(p) = \sum_{j=1}^n X^j(p)\partial_j(f)(p)$, for all $f \in C^\infty(U)$ and all $p \in U$.

To also consider tangent vectors and form the tangent bundle of M, we use the notion of germs of functions: For every $p \in M$, a *germ of C^∞ function* at p is the equivalence class of smooth functions from M to $\mathbb{R}$ with respect to the equivalence relation of being equal in some neighborhood of p. We denote by $C^\infty(p)$ the space of germs of C^∞ functions at p. The tangent bundle over M is a set, denoted by TM, together with a map $\pi : TM \to M$ called *(tangent bundle) projection map* with the following property: The *fiber* $T_pM := \pi^{-1}(p)$ of the tangent bundle TM is the linear space formed by all the derivations on the space $C^\infty(p)$. In other words, the elements of T_pM, called *tangent vectors* at p, are those $\mathbb{R}$-linear applications $v : C^\infty(p) \to \mathbb{R}$ that satisfy the Leibnitz rule: $v(fg) = v(f)g - fv(g)$, for all $f, g \in C^\infty(p)$. Therefore, if X is a vector field on M and p is in M, then X_p, defined as

$$X_p(u) := (X(f))(p), \qquad \forall f \in C^\infty(p),$$

gives a tangent vector at p. Hence, vector fields on M are sections of the tangent bundle TM, and moreover, one puts on TM a structure of manifold such that $X \in \mathrm{Vec}(M)$ if and only if $X : M \to T(M)$ is smooth and $\pi \circ X$ is the identity on M.

If $F : M \to N$ is a smooth map between smooth manifolds and $p \in M$, we shall denote by $\mathrm{d}F_p : T_pM \to T_{F(p)}N$ its differential, defined as follows. The pull back operator $u \mapsto F_p^*(u) := u \circ F$ maps $C^\infty(F(p))$ into $C^\infty(p)$; thus, for $v \in T_pM$ we have that

$$\mathrm{d}F_p(v)(f) := v(F_p^*(f)) = v(f \circ F), \qquad \forall f \in C^\infty(F(p)),$$

defines an element of $T_{F(p)}N$.

Every smooth curve $\sigma : I \to M$ gives a derivation at $\sigma(t)$ for each $t \in I$ by

$$\sigma'(t)(f) := \lim_{h\to 0} \frac{f(\sigma(t+h)) - f(\sigma(t))}{h}, \qquad \forall f \in C^\infty(\sigma(t)).$$

If $F : M \to N$ is smooth and σ is a smooth curve on M, then we have the formula

$$\mathrm{d}F_{\sigma(t)}(\sigma'(t)) = (F \circ \sigma)'(t), \tag{2.2.2}$$

where $\sigma'(t) \in T_{\sigma(t)}M$ and $(F \circ \sigma)'(t) \in T_{F(\sigma(t))}N$ are the tangent vectors along the two curves, in M and N, respectively. If $f \in C^\infty(M)$ and $p \in M$, identifying $T_{f(p)}\mathbb{R}$ with $\mathbb{R}$ itself, given $X \in \Gamma(TM)$, we have

$$\mathrm{d}f_p(X_p) = X_p(f).$$

For a vector field $X \in \Gamma(TM)$, a smooth curve $\sigma : (a, b) \to M$ is an *integral curve*, or a *flow line, of* X if

$$\sigma'(t) = X_{\sigma(t)}, \qquad \forall t \in (a, b).$$

For all $X \in \Gamma(TM)$ and all $p \in M$ there are $a < 0$, $b > 0$, and $\sigma : (a, b) \to M$ such that σ is an integral curve of X and $\sigma(0) = p$. Moreover such a σ is unique and has a unique maximal extension. We denote by $t \mapsto \Phi_X^t(p)$ the integral curve of X starting at p. We call $\Phi_X^t(p)$ the *flow at* p *at time* t *with respect to* X. Namely, we have

$$\begin{cases} \Phi_X^0(p) = p, \\ \dfrac{\mathrm{d}}{\mathrm{d}t}\Phi_X^t(p) = X_{\Phi_X^t(p)}. \end{cases} \tag{2.2.3}$$

One of the most important fundamental notions that we will utilize in our study is the Lie bracket of vector fields. The Lie bracket of vector fields has several equivalent definitions, and we will employ them all based on the viewpoint being considered.

Definition 2.2.4 (Lie bracket)**.** The *Lie bracket of vector fields* on a manifold M is the map

$$\begin{aligned}[\cdot,\cdot] : \mathrm{Vec}(M) \times \mathrm{Vec}(M) &\to \mathrm{Vec}(M) \\ (X, Y) &\mapsto [X, Y]\end{aligned}$$

defined with any of the equivalent viewpoints a–d:

2.2.4.a. Viewpoint of **derivations**: For $f \in C^\infty(M)$,

$$[X, Y](f) = X(Yf) - Y(Xf).$$

2.2.4.b. Viewpoint in **coordinates**: In local coordinates, if two vector fields are given by $X = \sum_{h=1}^n X^h \partial_h$ and $Y = \sum_{k=1}^n Y^k \partial_k$ for some smooth functions X^h and Y^k, then

$$[X, Y] = \sum_{h,k=1}^n \left(X^h \partial_h Y^k - Y^h \partial_h X^k\right) \partial_k.$$

2.2.4.c. Viewpoint of **Lie derivative**: For $p \in M$,

$$[X, Y]_p = \frac{\mathrm{d}}{\mathrm{d}t}\Big((\,\mathrm{d}\Phi_X^t)^{-1} Y_{\Phi_X^t(p)}\Big)\Big|_{t=0} =: (\mathcal{L}_X Y)_p \,.$$

2.2.4.d. Viewpoint of **commutation of flows**: For $p \in M$,

$$\begin{aligned} [X, Y]_p &= \frac{1}{2}\frac{\mathrm{d}^2}{\mathrm{d}t^2}\left(\Phi_Y^{-t} \circ \Phi_X^{-t} \circ \Phi_Y^{t} \circ \Phi_X^{t}\right)(p)\Big|_{t=0} \\ &= \frac{\mathrm{d}}{\mathrm{d}t}(\phi_Y^{-\sqrt{t}} \circ \phi_X^{-\sqrt{t}} \circ \phi_Y^{\sqrt{t}} \circ \phi_X^{\sqrt{t}})(p)\Big|_{t=0^+} . \end{aligned}$$

The Lie bracket induces on $\mathrm{Vec}(M)$ an infinite-dimensional Lie algebra structure, see Definition 4.1.3. Clearly, the push-forward via a diffeomorphism commutes with the Lie bracket operation, see Exercise 2.4.19, where if $F : M \to N$ is a diffeomorphism and $X \in \Gamma(TM)$, the push forward vector field $F_* X \in \Gamma(TN)$ is defined by the identity $(F_* X)_{F(p)} := \mathrm{d}F_p(X_p)$, for $p \in M$. Equivalently,

$$(F_* X) f := [X(f \circ F)] \circ F^{-1}, \qquad \forall f \in C^\infty(M). \tag{2.2.5}$$

2.2.2 Vector bundles

A simple example of vector bundle of rank r over a manifold M is the product space $M \times \mathbb{R}^r$ with the projection on the first component $\pi_1 : M \times \mathbb{R}^r \to M$. The next important example of vector bundle of rank $\dim(M)$ over a manifold M is the tangent bundle TM of M. Here, is the abstract definition.

Definition 2.2.6 (Vector bundle)**.** A *vector bundle of rank* r over a manifold M is a manifold E together with a smooth surjective map $\pi : E \to M$ such that, for all $p \in M$, the following properties hold:

1. The fiber $E_p := \pi^{-1}(p)$ has the structure of vector space of dimension r.
2. There is a neighborhood U of p in M and a diffeomorphism $\chi : \pi^{-1}(U) \to U \times \mathbb{R}^r$ such that
   (a) $\pi_1 \circ \chi = \pi$
   (b) for all $q \in U$, the restricted map $\chi|_{E_q} : E_q \to \{q\} \times \mathbb{R}^r$ is an isomorphism of vector spaces.

The space E is called *total space*, the manifold M is the *base*, the vector space E_p is the *fiber over* p and every such a map χ is called a *local trivialization*.

Exercise 2.2.7. Show that, if E is a vector bundle of rank r over a manifold M, then $\dim(E) = \dim(M) + r$. ■

Exercise 2.2.8. Show that if $\pi : E \to M$ is a vector bundle and $U \subset M$ is an open set, then $\pi|_{\pi^{-1}(U)} : \pi^{-1}(U) \to U$ is a vector bundle. ■

Definition 2.2.9 (Section)**.** A *section* of a vector bundle $\pi : E \to M$ is a smooth map $\sigma : M \to E$ such that $\pi \circ \sigma = \mathrm{Id}_M$. We will denote by $\Gamma(E)$ the set of all sections of E.

$$\begin{tikzcd} E \arrow[d, "\pi"'] \\ M \arrow[u, bend right, "\sigma"'] \end{tikzcd}$$

Definition 2.2.10 (Frames and local frames)**.** A *frame* of a bundle $\pi : E \to M$ is a set $\{X_1, \ldots, X_n\} \subset \Gamma(E)$ of sections on M such that, for all $p \in M$, the n-tuple $(X_1(p), \ldots, X_n(p))$ is a basis of the fiber E_p. A *local frame* for $\pi : E \to M$ at a point $p \in M$ is a frame for the bundle $\pi|_{\pi^{-1}(U)} : \pi^{-1}(U) \to U$ where U is some open neighborhood of p.

2.2.3 Riemannian and Finsler geometry

Let M be a differentiable manifold of dimension n. A *Riemannian metric* on M is a family of (positive-definite) inner products

$$\rho_p : T_pM \times T_pM \longrightarrow \mathbb{R}, \qquad p \in M,$$

such that, for all smooth vector fields X, Y on M, we have

$$p \mapsto \rho_p(X_p, Y_p)$$

defines a smooth function $M \to \mathbb{R}$. This smooth assignment of an inner product ρ_p to each tangent space T_pM is called a *metric tensor*. A metric tensor will also be denoted by $\langle \cdot, \cdot \rangle$. Endowed with one such a metric tensor, the pair $(M, \langle \cdot, \cdot \rangle)$ is called a *Riemannian manifold*.

Given a chart (U, φ) for the manifold M we have the coordinate vector fields $\partial_1, \ldots, \partial_n$ from Definition 2.2.1, and we consider the *components of the metric tensor relative to the coordinate system* as

$$\rho_{ij}(p) := \rho_p \left(\partial_i|_p , \partial_j|_p \right), \qquad \forall p \in U.$$

It is easy to verify that the functions $(\rho_{ij})_{ij}$ are smooth and contain all the information about ρ.

Finsler manifolds generalize Riemannian manifolds by no longer assuming that they are infinitesimally Euclidean in the sense that the norm on each tangent space is necessarily induced by an inner product. Two good references on Finsler geometry are [BCS00] and [AP94].

Classically a Finsler structure on a differentiable manifold M is given by a function $\|\cdot\| : TM \to \mathbb{R}$ that is smooth on the complement of the zero section of TM and such that the restriction of $\|\cdot\|$ to every tangent space T_pM is a (symmetric) norm (see Remark 2.2.14). We will consider a more general definition for Finsler structures: as regularity we only assume the continuity in the point and the convexity in the vector.

Every Riemannian manifold $(M, \langle\cdot,\cdot\rangle)$ has an associated function $TM \to [0, \infty)$, $X \mapsto \|X\| := \sqrt{\langle X, X\rangle}$. This is an example of a continuously varying norm.

Definition 2.2.11. A *continuously varying norm* on a differentiable manifold M is a continuous function from TM to $[0, \infty)$ usually denoted by $\| \cdot \|$ with the property that for all $p \in M$ the restriction of $\| \cdot \|$ to T_pM is a symmetric norm, i.e.,

1. $\|\lambda X\| = |\lambda|\|X\|$, $\forall X \in TM$, $\forall \lambda \in \mathbb{R}$;
2. $\|X + Y\| \leq \|X\| + \|Y\|$, $\forall p \in M$ and $\forall X, Y \in T_pM$;
3. $\|X\| = 0 \Rightarrow X = 0$.

Definition 2.2.12. In this text, we say that a *Finsler manifold* is a pair $(M, \| \cdot \|)$ where M is a differentiable manifold and $\| \cdot \|$ is a continuously varying norm on M. In this case, $\| \cdot \|$ is also called *Finsler structure*.

Example 2.2.13. There are at least two situations that we want the reader to keep in mind:

2.2.13.i Every Riemannian manifold $(M, \langle\cdot,\cdot\rangle)$ naturally has a structure of a Finsler manifold.

2.2.13.ii Every finite-dimensional normed vector space naturally has a structure of a Finsler manifold.

Remark 2.2.14. The notion of Finsler manifold is present in the literature with different meanings. On the one hand, the norm is classically required to be smooth (away from the zero section) and with a positive Hessian. Namely, some authors assume that norms for Finsler structures have strongly convex smooth unit spheres, while we do not in Definition 2.2.11. On the other hand, some authors considered other weak notions of norms. For example, they allow asymmetric norms, i.e., the first condition in Definition 2.2.11 is assumed only for $\lambda > 0$.

2.3 Length structures for Finsler manifolds

Connected Riemannian and Finsler manifolds carry the structure of length metric spaces. Let us recall the notion of absolutely continuous curve and its length with respect to a Finsler structure.

Definition 2.3.1. A curve $\gamma : [a, b] \to \mathbb{R}^n$ is *absolutely continuous* if there exists a Lebesgue integrable $\mathbb{R}^n$-valued function $g : [a, b] \to \mathbb{R}^n$ such that

$$\gamma(t) - \gamma(a) = \int_a^t g(s)\, \mathrm{d}s, \qquad \forall t \in [a, b].$$

The function g is sometimes denoted by $\dot\gamma$, however it is only defined almost everywhere with respect to the Lebesgue measure on $[a, b]$. A curve $\gamma : [a, b] \to M$ into a differentiable manifold is said *absolutely continuous* (or, AC, for short), if it is so when read in local coordinates, i.e., for all local coordinate map $\phi : U \to \mathbb{R}^n$ and for all $a', b' \in [a, b]$ such that $\gamma([a', b']) \subset U$, then $\phi \circ \gamma|_{[a',b']}$ is absolutely continuous. For every absolutely continuous curve $\gamma : [a, b] \to M$ one can also define a *derivative* $\dot\gamma : [a, b] \to TM$ using local coordinates, which is defined almost everywhere as a measurable map (see Exercise 2.4.16).

As usual in Differential Geometry, to check that a curve $\gamma : [a, b] \to TM$ is absolutely continuous it is sufficient that the image of the curve admits a covering of coordinate systems for M on which γ is absolutely continuous (see Exercise 2.4.15).

Definition 2.3.2 (Length of a curve in a Finsler manifold)**.** Let $(M, \|\cdot\|)$ be a Finsler manifold. Let $\gamma : [a, b] \to M$ be an absolutely continuous curve. We set

$$\mathrm{Length}_{\|\cdot\|}(\gamma) := \int_a^b \|\dot\gamma(t)\|\, \mathrm{d}t. \tag{2.3.3}$$

We remark that the *Finsler length* (2.3.3) of an absolutely continuous curve is finite.

The arc-length is independent of the chosen parametrization, as can be shown using the change-of-variables formula. In particular, a curve $\gamma : [a,b] \to M$ can be parametrized by its arc length, i.e., in such a way that

$$\mathrm{Length}_{\|\cdot\|}(\gamma|_{[t_1,t_2]}) = |t_2 - t_1|, \qquad \forall t_1, t_2 \text{ with } a \leq t_1 \leq t_2 \leq b.$$

A curve is parametrized by arc-length if and only if $\|\dot\gamma(t)\| = 1$, for almost all $t \in [a,b]$.

The *distance function* $d_{\|\cdot\|} : M \times M \to [0,+\infty)$ is defined by

$$d_{\|\cdot\|}(p,q) = \inf \mathrm{Length}_{\|\cdot\|}(\gamma), \tag{2.3.4}$$

where the infimum is taken over all absolutely continuous curves γ in M joining p to q.

The function $d_{\|\cdot\|}$ satisfies the properties of a distance function for a metric space. The only property that is not completely straightforward is that $d_{\|\cdot\|}(p,q) = 0$ implies $p = q$. For proving this property, we claim that locally in a coordinate system every Finsler structure (as every Riemannian structure) is biLipschitz equivalent to the Euclidean structure, i.e., for some $c > 0$, we have

$$c^{-1}\|\cdot\| \leq \|\cdot\|_{\mathbb{E}} \leq c\|\cdot\|, \tag{2.3.5}$$

where $\|\cdot\|_{\mathbb{E}}$ is the Euclidean norm. Indeed, let $U \subseteq \mathbb{R}^n$ be an open set parametrizing the manifold and fix a compact set $K \subseteq U$, which we think having nonempty interior. Consider $T^1K := \{(p,v) : p \in K, v \in T_pU, \|v\|_{\mathbb{E}} = 1\}$ the bundle of unit vectors on K. Notice that T^1K is compact. Hence, the continuous function $\|\cdot\|$ on T^1K admits maximum and minimum, moreover the minimum cannot be 0 since otherwise we would have a non-zero vector with norm 0. We deduce that there exists a constant $c > 0$ such that if $p \in K$ and $v \in T_pK$ is such that $\|v\|_{\mathbb{E}} = 1$ then $c^{-1} \leq \|v\| \leq c$. By homogeneity we have (2.3.5) on K.

Consequently, based on (2.3.5), we can establish the biLipschitz equivalence between distance functions. Specifically, we have proven that every two Finsler distance functions on the same manifold are biLipschitz equivalent on compact sets. We summarize our findings in the following proposition.

Proposition 2.3.6. *On every Finsler manifold in local coordinates, on compact sets, the Finsler distance function is biLipschitz equivalent to the Euclidean distance function. Consequently, on every compact set of every manifold, every Finsler structure is biLipschitz equivalent to every Riemannian structure. In particular, Finsler distance functions induce the same topology as the manifold topology.*

On each Finsler manifold to every continuously varying norm, as defined in Definition 2.2.11, we associated a length structure as in (2.3.3) and a distance function as in (2.3.4). The distance function then induces another length structure, as in Definition 2.1.2. We show next that the two length structures coincide.

Proposition 2.3.7. *Assume M is a differentiable manifold equipped with a continuously varying norm $\|\cdot\| : TM \to \mathbb{R}$ with induced length structure $\mathrm{Length}_{\|\cdot\|}$ and distance function $d_{\|\cdot\|}$. If $\gamma\colon [a,b] \to M$ is an absolutely continuous curve, then*

$$\mathrm{Length}_{d_{\|\cdot\|}}(\gamma) = \mathrm{Length}_{\|\cdot\|}(\gamma). \tag{2.3.8}$$

Proof. To prove the $\leq$ inequality in (2.3.8), notice that for all $t, s \in [a,b]$ we have

$$d_{\|\cdot\|}(\gamma(s),\gamma(t)) \overset{\text{def}}{=} \inf_{\sigma} \int_s^t \|\dot\sigma(\tau)\| \, \mathrm{d}\tau \leq \int_s^t \|\dot\gamma(\tau)\| \, \mathrm{d}\tau \overset{\text{def}}{=} \mathrm{Length}_{\|\cdot\|}(\gamma|_{[s,t]}),$$

where the infimum is taken over all AC curves σ from $\gamma(s)$ to $\gamma(t)$. Using the definition of length we deduce that $\mathrm{Length}_{d_{\|\cdot\|}} \leq \mathrm{Length}_{\|\cdot\|}$.

Regarding the other inequality, we shall use that the norm changes continuously. It is convenient to work in coordinates, and it is enough to prove our claim locally. Parametrizing M with an open subset U of $\mathbb{R}^n$ we write the norm as $\|v\|_x =: F(x,v)$, for $x \in U$ and $v \in T_xU \simeq \mathbb{R}^n$. Fix some $K > 1$. Since F is continuous and homogeneous in the second variable, then at each point $p \in U$ there exists a neighborhood U_p of p such that

$$\frac{1}{K}F(q,v) \leq F(p,v) \leq KF(q,v), \qquad \forall q \in U_p, \forall v \in \mathbb{R}^n. \tag{2.3.9}$$

We find a partition $a = a_0 < a_1 < \cdots < a_n = b$ and points $p_1, \ldots, p_n \in M$ such that

$$\gamma([a_{i-1}, a_i]) \subseteq U_{p_i}, \qquad \forall i = 1, \ldots, n. \tag{2.3.10}$$

Let us denote by d_i the distance induced by the (constant) norm $F(p_i,\cdot)$. Since then we are in the case of a normed vector space (see Example 2.1.17) we have

$$\mathrm{Length}_{F(p_i,\cdot)} = \mathrm{Length}_{d_i}. \tag{2.3.11}$$

Moreover, as a consequence of (2.3.9), we have

$$d_i \leq K d_{\|\cdot\|}. \tag{2.3.12}$$

Thus, using (2.3.9),(2.3.11), and (2.3.12), together with (2.3.10), we obtain that

$$\begin{aligned}
\mathrm{Length}_{\|\cdot\|}(\gamma) &\overset{\text{def}}{=} \int_a^b F(\gamma(t),\dot\gamma(t)) \\
&= \sum_{i=1}^n \int_{a_{i-1}}^{a_i} F(\gamma(t),\dot\gamma(t)) \\
&\overset{(2.3.9)}{\leq} K\sum_{i=1}^n \int_{a_{i-1}}^{a_i} F(p_i,\dot\gamma(t)) \\
&\overset{(2.3.11)}{=} K\sum_{i=1}^n \mathrm{Length}_{d_i}(\gamma|_{[a_{i-1},a_i]}) \\
&\overset{(2.3.12)}{\leq} K^2\sum_{i=1}^n \mathrm{Length}_{d_{\|\cdot\|}}(\gamma|_{[a_{i-1},a_i]}) = K^2\,\mathrm{Length}_{d_{\|\cdot\|}}(\gamma).
\end{aligned}$$

As K can be chosen arbitrarily close to 1, we also deduce that $\mathrm{Length}_{\|\cdot\|} \leq \mathrm{Length}_{d_{\|\cdot\|}}$. □

Remark 2.3.13. Let $\gamma : [a,b] \to M$ be a curve on a manifold, which is equipped with a continuously varying norm $\|\cdot\|$. With the following points, we shall clarify the relationship between absolute continuity (AC) and having finite length:

2.3.13.i. If γ is AC, then $\mathrm{Length}_{\|\cdot\|}(\gamma) = \mathrm{Length}_{d_{\|\cdot\|}}(\gamma)$ and both these quantities are finite, see Proposition 2.3.7.

2.3.13.ii. If γ is not AC, then $\mathrm{Length}_{\|\cdot\|}(\gamma)$ is not defined.

2.3.13.iii. If $\mathrm{Length}_{d_{\|\cdot\|}}(\gamma)$ is finite, then up to reparametrization γ is Lipschitz with respect to $d_{\|\cdot\|}$, and thus with respect to any euclidean distance, in coordinates. Therefore, by Rademacher Theorem γ is AC.

2.4 Exercises

Exercise 2.4.1. Let (M,d) be a metric space equipped with its natural topology.

(i) Show that if M is connected, then d is finite.

(ii) Show that in general d is finite on each connected component of M. ■

Exercise 2.4.2. The *mesh* of a partition $\mathcal{P} = (t_1,\dots,t_k)$ is defined as

$$\|\mathcal{P}\| := \max_{j=1,\dots,k-1} |t_{i+1} - t_i|.$$

Show that, if $\mathcal{P}_j$ are partitions such that $\|\mathcal{P}_j\| \to 0$ as $j \to \infty$, then $L(\gamma) = \lim_{j\to\infty} L(\gamma,\mathcal{P}_j)$. ■

Exercise 2.4.3. Show that, if $\mathcal{P}_1$ and $\mathcal{P}_2$ are partition of the same interval with $\mathcal{P}_1 \subset \mathcal{P}_2$, then $L(\gamma, \mathcal{P}_1) \leq L(\gamma, \mathcal{P}_2)$. ■

Exercise 2.4.4. Show that the length of a curve is independent on its parameterization. Namely, If $\gamma : I \to M$ is a curve in a metric space and $h : J \to I$ is a continuous monotone surjection between intervals, then $L(\gamma) = L(\gamma \circ h)$. ■

Exercise 2.4.5. If $\gamma : [a, b] \to (M, d)$ is rectifiable, then can be reparametrized by arc length. [Hint: consider the change of parametrization given by $s \mapsto \mathrm{Length}(\gamma|_{[a,s]})$.] ■

Exercise 2.4.6. If $\gamma : [a, b] \to (M, d)$ is *parametrized with constant speed* s, with $s \in [0, \infty)$, i.e.,

$$\mathrm{Length}(\gamma|_{[t_1,t_2]}) = s|t_2 - t_1|, \qquad \forall t_1, t_2 \in [a, b],$$

then $L(\gamma) = s|a - b|$ and γ is s-Lipschitz. ■

Exercise 2.4.7. Prove that for each partition $\mathcal{P}$, if a sequence $(\gamma_n)_{n\in\mathbb{N}}$ of curves pointwise converges to γ then $L(\gamma_n, \mathcal{P})$ converges to $L(\gamma, \mathcal{P})$. ■

Exercise 2.4.8. Let $f_n : X \to \mathbb{R}$ be a sequence of continuous functions on a topological space. Prove that the function $\sup_n f_n$ is lower semicontinuous. *Hint*: adapt the proof of Theorem (2.1.4). ■

Exercise 2.4.9. Let (M, d) be a complete metric space and let $\mathcal{F} := C^0(I; M)$ be the family of all curves from a fixed interval I into M. Endow $\mathcal{F}$ with the metric

$$d_{\sup}(\sigma, \gamma) = \sup_{t\in I}\{d_M(\sigma(t), \gamma(t))\}, \qquad \forall \sigma, \gamma \in \mathcal{F}.$$

Prove that $(\mathcal{F}, d_{\sup})$ is a complete metric space. ■

Exercise 2.4.10. Let $F : M_1 \to M_2$ a maps between two metric spaces that is K-Lipschitz. Show that if γ is a curve in M_1 then $L(F \circ \gamma) \leq K \cdot L(\gamma)$. ■

Exercise 2.4.11. Show that a geodesic space is a length space – what is not automatic is that the distance is finite. ■

Exercise 2.4.12. Find a homeomorphism $F : M_1 \to M_2$ between two metric spaces with the property that $L(F \circ \gamma) = L(\gamma)$ for all curves γ in M_1, but F is not an isometry. ■

Exercise 2.4.13. Show that each measure is countably subadditive.

Hint: Given countably many sets, split them into disjoint sets and apply 2.1.23. ■

Exercise 2.4.14. Let $\gamma : [a, b] \to \mathbb{R}^n$ be absolutely continuous. Show that $\dot{\gamma}$ is unique up to measure zero. ■

Exercise 2.4.15. Let $\gamma : I \to M$ be a curve. Show that γ is absolutely continuous if for all $t \in I$ there exist $\epsilon > 0$ and a local coordinate map $\varphi : U \subset M \to \mathbb{R}^n$ with $\gamma([t-\epsilon, t+\epsilon]) \subset U$ and such that $\varphi \circ \gamma|_{[t-\epsilon,t+\epsilon]}$ is absolutely continuous.

■

Exercise 2.4.16. Let $\gamma : I \to M$ be an absolutely continuous curve. Let $\varphi_1, \varphi_2 : U \subset M \to \mathbb{R}^n$ be two coordinate maps. Show that the derivative of $\varphi_1 \circ \gamma$ is related to the derivative of $\varphi_2 \circ \gamma$ by the differential of $\varphi_1 \circ \varphi_2^{-1}$ and hence one can define the derivative $\dot{\gamma}$ up to measure zero. ■

Exercise 2.4.17. Prove that every absolutely continuous curve in $\mathbb{R}^n$ can be re-parametrized to be a Lipschitz curve with respect to the Euclidean distance. ■

Exercise 2.4.18. Prove the *continuity from below* for measures, i.e., for every measure μ on a space X, if $E_1 \subseteq E_2 \subseteq \ldots \subseteq X$ are in the domain of μ then $\mu(\cup_{i=1}^{\infty} E_i) = \lim_{i\to\infty} \mu(E_i)$. ■

Exercise 2.4.19. The push-forward commutes with the Lie bracket, namely if $F : M \to N$ is a diffeomorphism of manifolds

$$[F_*X, F_*Y] = F_*[X, Y], \qquad \forall X,\, Y \in \Gamma(TM). \tag{2.4.20}$$

■

Exercise 2.4.21. If X and Y are vector fields tangent to a submanifold $N \subseteq M$, then also $[X, Y]$ is tangent to N. ■

Exercise 2.4.22. For all $X, Y \in \mathrm{Vec}(M)$ and for all $f, g \in C^\infty(M)$

$$[fX, gY] = fg[X, Y] + f(Xg)Y - g(Yf)X.$$

■

Chapter 3

General theory of Carnot-Carathéodory spaces

We have reached the point where we are ready to introduce the main object of our investigation: subRiemannian manifolds, and more generally, subFinsler manifolds, also known as Carnot-Carathéodory spaces. These spaces will be equipped with Carnot-Carathéodory distances. Our first significant result is the Chow-Rashevsky Theorem, which states that on every subFinsler manifold, the Carnot-Carathéodory distance induces the same topology as the manifold structure itself. It is important to emphasize that this result relies on the crucial assumption that the subbundle is bracket generating.

3.1 Definition of Carnot-Carathéodory spaces

In this chapter, we shall denote by M a differentiable manifold, whose dimension will mostly be denoted by n. Thus the tangent bundle of M is TM and is a $2n$-dimensional manifold with the following local parametrization: if $\varphi : U \subset \mathbb{R}^n \to M$ is a local parametrization for M, then it induces vector fields $\partial_{x_1}, \dots, \partial_{x_n}$ and the map $U \times \mathbb{R}^n \to TM$, $(x, v) \mapsto v_1 \partial_{x_1}|_{\varphi(x)} + \cdots + v_n \partial_{x_n}|_{\varphi(x)}$ is a local parametrization for TM. In other words, the vector fields $\partial_{x_1}, \dots, \partial_{x_n}$ form a local frame for TM.

3.1.1 Bracket-generating distributions

Definition 3.1.1 (Polarization, a.k.a. distribution or tangent subbundle)**.** A *distribution of tangent subspaces* on a manifold M is a subset $\Delta \subseteq TM$ such that for all $\bar{p} \in M$ there exists smooth vector fields $X_1, \dots, X_m$ on some neighborhood U of $\bar{p}$ such that

$$\Delta_p := \Delta \cap T_p M = \operatorname{span}\{X_1(p), \dots, X_m(p)\}, \qquad \forall p \in U. \tag{3.1.2}$$

Distributions of tangent subspaces are also simply called *distributions*. If moreover there exists $r \in \mathbb{N}$ such that $r = \dim \Delta_p$, for all $p \in M$, then we say that Δ has *constant rank* with *rank* equal to r. Distributions of rank r are also called *distributions of r-planes* or *r-plane fields*. Constant rank distributions are also called *polarizations* or *tangent subbundles*. The pair (M, Δ) of a manifold M and a polarization Δ is called *polarized manifold*.

Notice that each tangent subbundle is indeed a subbundle of the tangent bundle: A *subbundle* E of a vector bundle F (see Section 2.2.2) over a manifold M is a collection of linear subspaces E_p of the fibers F_p of F at each point p in M that forms a vector bundle in its own right. Moreover, a tangent subbundle of rank r on an n-manifold has dimension $n + r$.

Here is a simple example of a polarization on the 3-dimensional manifold $\mathbb{R}^3$, with coordinates x, y, z. Let $f, g : \mathbb{R} \to \mathbb{R}$ be smooth functions. Then the two smooth vector fields

$$X_1(x, y, z) \quad := \quad \partial_x + f(x, y, z)\partial_z, \tag{3.1.3}$$

$$X_2(x, y, z) \quad := \quad \partial_y + g(x, y, z)\partial_z \tag{3.1.4}$$

are linearly independent at every point (x, y, z) and define a rank-2 tangent subbundle Δ on $\mathbb{R}^3$ as

$$\Delta_{(x,y,z)} \quad := \quad \{aX_1(x, y, z) + bX_2(x, y, z) : a, b \in \mathbb{R}^2\} \tag{3.1.5}$$

$$= \quad \{(a, b, af(x, y, z) + bg(x, y, z)) : a, b \in \mathbb{R}^2\}. \tag{3.1.6}$$

Definition 3.1.7. Here is some notation and terminology that is commonly used for distributions and families of vector fields:

- The set of smooth vector fields on a manifold M is denoted with $\mathrm{Vec}(M)$ or $\Gamma(TM)$. Indeed, an element of $\Gamma(TM)$ is a smooth section $X : M \to TM$ of the bundle $TM \to M$.

- A vector field $X : M \to TM$ is said to be *tangent* to a distribution $\Delta \subseteq TM$ at a point $p \in M$ if $X(p) \in \Delta$.

- Given a distribution $\Delta \subset TM$, we denote by $\Gamma(\Delta)$ the set of smooth vector fields of M tangent to Δ at every point of M.

- Given a family $\mathscr{F} \subset \Gamma(TM)$ of vector fields on M and $p \in M$, we set $\mathscr{F}_p := \{X_p : X \in \mathscr{F}\}$.

- Given a family $\mathscr{F} \subset \Gamma(TM)$ of vector fields on M, we denote by $\mathrm{Lie}(\mathscr{F})$ the Lie algebra generated by $\mathscr{F}$ with respect to the Lie bracket of vector fields within $\Gamma(TM)$, see Section 2.2.1.

We spell out that the set $\mathrm{Lie}(\mathscr{F})$ is the smallest subset of $\Gamma(TM)$ with $\mathscr{F} \subset \mathrm{Lie}(\mathscr{F})$ and the property

$$X, Y \in \mathrm{Lie}(\mathscr{F}), a, b \in \mathbb{R} \implies [X, Y], aX + bY \in \mathrm{Lie}(\mathscr{F}).$$

We are ready to introduce the condition that will make us join points with curves tangent to a polarization Δ. The following condition (3.1.9) has many names. It is also called *Hörmander's condition* or *Chow's condition*.

Definition 3.1.8 (Bracket generating)**.** A distribution Δ on a manifold M is *bracket generating* if

$$(\mathrm{Lie}(\Gamma(\Delta)))_p = T_pM, \qquad \forall p \in M. \tag{3.1.9}$$

Let us clarify what is the meaning of a curve tangent to a distribution:

Definition 3.1.10 (Horizontal curve)**.** Given a polarized manifold (M, Δ) a curve $\gamma : [a, b] \to M$ is said to be Δ-*horizontal* if γ is absolutely continuous (see Definition 2.3.1) and $\dot\gamma(t) \in \Delta_{\gamma(t)}$ for almost every $t \in [a, b]$. Curves that are Δ-horizontal are also said to be *horizontal with respect to* Δ, or, simply, *horizontal* or *Legendrian*. The terms *admissible curve* and *controlled path* are also used to refer to such curves.

Remark 3.1.11. Let $X_1, \dots, X_m$ be vector fields spanning a distribution Δ on a manifold M, in the sense that (3.1.2) holds for all $p \in M$. On the one hand, if

$$(\mathrm{Lie}(\{X_1, \dots, X_m\}))_p = T_pM, \qquad \forall p \in M, \tag{3.1.12}$$

then Δ is bracket generating. On the other hand, the converse implication may not hold: For example, let $\phi : \mathbb{R} \to \mathbb{R}$ be a C^∞ function such that $\phi(0) = 0$ if and only if $x = 0$ and $\frac{\mathrm{d}^k}{\mathrm{d}x^k}\phi(0) = 0$, for all $k = 0, 1, 2, \dots$, see Figure 3.1 for an example. Consider on $\mathbb{R}^2$ with coordinates (x, y) the vector fields

$$X = \partial_x \qquad \text{and} \qquad Y = \phi(x)\partial_y.$$

Then X, Y do not satisfy (3.1.12), see Exercise 3.4.3. While they span the same distribution of the bracket-generating frame $\partial_x, x\partial_y$, see Exercise 3.4.1.

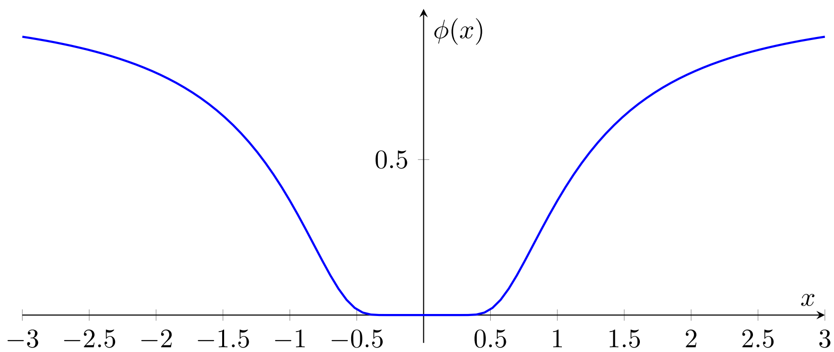


Figure 3.1: Plot of the function $\phi(x) = \exp\left(-\frac{1}{x^2}\right)$, with $\phi(0) = 0$. This function is smooth everywhere and has all derivatives equal to zero at $x = 0$, but it is not identically zero. It is commonly used to define the 'Gaussian bump function'.

3.1.2 SubFinsler structures of constant rank

Definition 3.1.13 (SubFinsler and subRiemannian manifolds of constant rank)**.** A *subFinsler manifold* is a triple $(M, \Delta, \|\cdot\|)$ where M is a connected manifold, $\|\cdot\|$ is a continuously varying norm (recall Definition 2.2.11), and Δ is a bracket-generating polarization on M, hence, the rank of Δ is here assumed constant. The pair $(\Delta, \|\cdot\|)$ is said to be a *subFinsler structure* on M. If the norm $\|\cdot\|$ is given by a Riemannian scalar product $\langle\cdot,\cdot\rangle$, then $(M, \Delta, \langle\cdot,\cdot\rangle)$ is called *subRiemannian manifold.*

We consider Riemannian and Finsler manifolds as particular cases of subRiemannian and subFinsler manifolds, respectively, which is the case when Δ is the whole tangent bundle.

Since in what follows only the values of the restriction $\|\cdot\||_\Delta$ of $\|\cdot\|$ to Δ will be important, we sometime say that $(M, \Delta, \|\cdot\||_\Delta)$ is a subFinsler manifold with subFinsler structure $(\Delta, \|\cdot\||_\Delta)$. In fact, we will consider the length with respect to $\|\cdot\|$, see (2.3.3), only for those curves that are horizontal with respect to Δ.

Definition 3.1.14 (CC-distance)**.** Given a subFinsler manifold $(M, \Delta, \|\cdot\|)$ the *Carnot-Carathéodory distance* between two points $p, q \in M$ is

$$d_{CC}(p,q) := \inf\left\{\mathrm{Length}_{\|\cdot\|}(\gamma) : \gamma \text{ is } \Delta\text{-horizontal curve from } p \text{ to } q\right\}. \tag{3.1.15}$$

If the infimum is realized by a curve γ, then γ is length minimizing among the horizontal curves joining the two points p, q, and in this case $d_{CC}(p,q) = \mathrm{Length}_{\|\cdot\|}(\gamma)$.

For us a subFinsler manifold $(M, \Delta, \|\cdot\|)$ is also equipped with a Finsler distance. If $d_F := d_{\|\cdot\|}$

is the Finsler distance associated to $(M, \|\cdot\|)$, see (2.3.4), then we obviously have

$$d_F(p,q) \leq d_{CC}(p,q), \qquad \forall p,q \in M, \tag{3.1.16}$$

because in the definition of d_{CC} we are infimizing over a subset of the set that we use for d_F. Notice that we use the same length structure for defining both distances.

We anticipate that the above d_{CC} is indeed a finite distance. In fact, because Δ is assumed bracket generating and M is assumed connected, we shall show the following result.

Theorem 3.1.17 (Chow, see Section 3.2.3)**.** *If $(M, \Delta, \|\cdot\|)$ is a subFinsler manifold, then d_{CC} is finite and induces the manifold topology on M.*

Remark 3.1.18 (Terminology)**.** The Carnot-Carathéodory distance is sometimes called *CC-distance* or *subFinsler distance*. A subFinsler manifold equipped with its Carnot-Carathéodory distance is called *Carnot-Carathéodory space*. If $\|\cdot\|$ is the norm coming from a Riemannian metric, i.e., $(M, \Delta, \|\cdot\|)$ is a subRiemannian manifold, then $(\Delta, \|\cdot\|)$ is called a *subRiemannian structure* and d_{CC} is called *subRiemannian distance*.

Some authors call d_{CC} a *Finsler-Carnot-Carathéodory distance* to emphasize that in their context d_{CC} might not necessarily be subRiemannian. Sub-Riemannian metrics appeared in the literature under a variety of names such as 'singular Riemannian metric' or 'non-holonomic Riemannian metric'. They were also considered in the theory of hypoelliptic PDEs, but without a specific name.

3.1.3 Control Theory viewpoint

In Control Theory one is interested in systems of differential equations of the form

$$\dot{\gamma} = \sum_{j=1}^{m} c_j(t) X_j(\gamma), \tag{3.1.19}$$

where $X_1, \ldots, X_m$ are given vector fields on a manifold M, and the $c_1, \ldots, c_m$ are variable L^1 functions on some bounded interval. These functions are called *control functions* or *controls*. Every path obtained integrating (3.1.19) is called a *controlled path*.

When the rank of the system of vector fields $X_1, \ldots, X_m$ is constant, controlled paths coincide with the absolutely continuous paths tangent to the distribution Δ generated by $X_1, \ldots, X_m$ as

$$\Delta_p := \operatorname{span}_{\mathbb{R}} \{X_1(p), \ldots, X_m(p)\}, \qquad \text{for } p \in M. \tag{3.1.20}$$

Conversely, every rank-m distribution Δ can, locally, be written as in (3.1.20). Observe that in the previous sentence, the adverb 'locally' is needed, for global topological reasons, as for example for the tangent bundle $\Delta = T(\mathbb{S}^2)$ of the 2D sphere $\mathbb{S}^2$.

However, for many systems of interest in Control Theory, the vector fields $X_1, \dots, X_m$ are not linearly independent at every point and/or the distribution that they define has not constant rank. Still, one can still define a related distance: for $p \in M$ and $v \in T_pM$, set

$$g_p(v) := \inf\{u_1^2 + \cdots + u_m^2 \mid u_1, \dots, u_m \in \mathbb{R}, u_1X_1(p) + \cdots + u_mX_m(p) = v\}.$$

We are using the notation that $\inf \emptyset = +\infty$. We then have that g_p is a positive-definite quadratic form on the subspace

$$\Delta_p := \operatorname{span}_{\mathbb{R}} \{X_1(p), \dots, X_m(p)\}.$$

The *control distance associated to the system* $X_1, \dots, X_m$ is defined as, for every p and q in M,

$$d(p,q) := \inf \left\{ \int_0^1 g_p(\dot\gamma(t))^{1/2} dt \;\middle|\; \gamma \text{ absolutely continuous path, } \gamma(0) = p, \gamma(1) = q \right\}. \tag{3.1.21}$$

3.1.4 The general definition with varying rank

The Finsler-Carnot-Carathéodory distance (3.1.15) and the control distance (3.1.21) fit in a general context. Namely, with the language of vector bundles we can give a broader definition.

Definition 3.1.22. A *(rank-varying) sub-Finsler structure* on a manifold M is a function $g : TM \to [0, \infty]$ obtained by the following construction: Let E be a vector bundle over M endowed with a function from E to $[0, \infty)$ denoted by $\|\cdot\|$ with the property that for all $p \in M$ the restriction of $\|\cdot\|$ to $\mathbb{E}_p$ is a symmetric norm, c.f. Definition 2.2.11. We consider a smooth map $\sigma : E \to TM$ that is a morphism of vector bundles lifting the identity, i.e. the following diagram commutes:

$$\begin{tikzcd} E \arrow[rr, "\sigma"] \arrow[dr, "\pi"'] & & TM \arrow[dl, "\pi"] \\ & M & \end{tikzcd}$$

and $\sigma|_{E_p}$ is a linear map from E_p to T_pM. We set

$$g_p(v) := \inf\{\|u\| : u \in E_p, \sigma(u) = v\}, \qquad \forall p \in M, \forall v \in T_pM.$$

Analogously as before, one defines the *sub-Finsler distance associated to the bundle morphism* σ, for every p and q in M, as

$$d(p,q) := \inf \left\{ \int_0^1 g_p(\dot\gamma(t))^{1/2} dt \;\middle|\; \gamma \text{ absolutely continuous path, } \gamma(0) = p, \gamma(1) = q \right\}.$$

One can check that, for the inclusion $\sigma : \Delta \hookrightarrow TM$ of a sub-bundle of the tangent bundle, one recovers the Finsler-Carnot-Carathéodory distance (3.1.15). For $E := M \times \mathbb{R}^m$ and $\sigma(p, u) := u_1X_1 + \cdots + u_mX_m$, one recovers the control distance (3.1.21).

3.1.5 Equiregular distributions

Let $\Delta \subset TM$ be a subbundle. For every $p \in M$ we define

$$\begin{aligned}
&\Delta^{[0]}(p) := \{0\} \subset T_pM \\
&\Delta^{[1]}(p) := \Delta_p \\
&\Delta^{[2]}(p) := \Delta^{[1]}(p) + \operatorname{span}\{[X, Y]_p : X, Y \in \Gamma(\Delta)\}.
\end{aligned}$$

Then $\Delta^{[2]} := \bigcup_{p\in M} \Delta^{[2]}(p)$ is a subset of TM. In general $\Delta^{[2]}$ may not be a subbundle since its rank may vary, i.e., the function $p \mapsto \dim \Delta^{[2]}(p)$ may not be constant.

Example 3.1.23 (Non-equiregular distribution)**.** In $\mathbb{R}^3$ the *Martinet distribution* is the subbundle $\Delta \subset T\mathbb{R}^3$ spanned by

$$\begin{aligned}
&X_1 = \partial_x + \frac{y^2}{2}\partial_z \\
&X_2 = \partial_y.
\end{aligned}$$

Notice that

$$X_3 := [X_2, X_1] = y\partial_z \qquad \text{and} \qquad X_4 := [X_2, X_3] = \partial_z.$$

Then

$$\Delta^{[2]}(p) = \begin{cases} T_p\mathbb{R}^3 & \text{if } p_2 \neq 0, \\ \Delta^{[1]}(p) & \text{if } p_2 = 0. \end{cases}$$

Remark 3.1.24. If $X_1, \ldots, X_r$ is a frame for Δ, then

$$\{X_1, \ldots, X_r\} \cup \{[X_i, X_j] : i, j = 1, \ldots, r\}$$

span $\Delta^{[2]}$ at every point. Indeed, if $X, Y \in \Gamma(\Delta)$, then $X = \sum_i a^iX_i$, $Y = \sum_j b^jX_j$ for some smooth functions a^i, b^j. We have

$$[X, Y] = [a^iX_i, b^jX_j] = a^ib^j[X_i, X_j] + a^i(X_ib^j)X_j - b^j(X_ja^i)X_i.$$

Definition 3.1.25 ($\Delta^{[k]}$)**.** Given a distribution $\Delta \subseteq TM$ on M, for each $k = 1, 2, \ldots$ we define the subset $\Delta^{[k]} \subseteq TM$ describing each of its fiber $\Delta^{[k]}(p)$ at t $p \in M$, i.e., the sets $\Delta^{[k]} \cap T_pM$. The fiber $\Delta^{[k]}(p)$ is given by

$$\Delta^{[k]}(p) := \operatorname{span}\{[X_1, [X_2, \ldots, [X_{j-1}, X_j] \ldots]](p) \,:\, j \in \{1, \ldots, k\},\ X_1, \ldots, X_j \in \Gamma(\Delta)\}. \tag{3.1.26}$$

The sets $\Delta^{[k]}(p)$ can also be defined inductively by $\Delta^{[1]} = \Delta$ and, for all $k \geq 2$,

$$\Delta^{[k+1]}(p) = \Delta^{[k]}(p) + \operatorname{span}\{[X_1, [X_2, \ldots, [X_k, X_{k+1}] \ldots]](p) \,:\, X_1, \ldots, X_{k+1} \in \Gamma(\Delta)\}\,. \tag{3.1.27}$$

Definition 3.1.28 (Regular point for Δ)**.** If Δ is a distribution on M and $p \in M$, we say that p is *regular* for Δ if for all $k \in \mathbb{N}$ the function

$$q \mapsto \dim \Delta^{[k]}(q) \tag{3.1.29}$$

is constant in a neighborhood of p.

Notice that the functions (3.1.29) is $\mathbb{N}$-valued. Hence, if it is locally constant, then it is constant on connected components.

Definition 3.1.30 (Equiregular distributions)**.** Let M be a manifold. A distribution $\Delta \subset TM$ is said to be *equiregular* if every $p \in M$ is regular for Δ. In this case we call $(\Delta^{[k]})_{k\in\mathbb{N}}$, as in Definition 3.1.25, the *flag of subbundles* for Δ.

Remark 3.1.31. A distribution $\Delta \subset TM$ is equiregular if and only if, for all $k \in \mathbb{N}$, the set $\Delta^{[k]}$ is a subbundle (Exercise).

Notice that if Δ is bracket generating and equiregular, then there is $s \in \mathbb{N}$ such that $\Delta^{[s]} = TM$. The minimal such an s is called *step* of Δ.

Definition 3.1.32 (Equiregular subFinsler manifolds)**.** A subFinsler manifold $(M, \Delta, \|\cdot\|)$ is called *equiregular* if Δ is equiregular.

3.2 Chow's theorem and existence of geodesics

3.2.1 Local transitivity and Sussmann's orbit theorem

We want to motivate now the fact that since in a subFinsler manifold the distribution is bracket generating, then the Carnot-Carathéodory distance is finite. The bracket-generating condition can be considered as an infinitesimal transitivity. Chow's theorem implies local transitivity:

Theorem 3.2.1 (Chow)**.** *If a subbundle Δ of the tangent bundle of a manifold is bracket generating at some point p (i.e.,* (3.1.9) *holds at p), then every point q that is sufficiently close to p can be joined to p by an absolutely continuous curve almost everywhere tangent to Δ.*

In fact, close points in a subFinsler manifold can be joined by horizontal curves that are short with respect to the Finsler length, i.e., Theorem 3.1.17 holds.

We first explain the validity of Theorem 3.2.1 taking for granted a theorem by Sussmann. We are omitting the proof of Sussmann's theorem which is in fact the core of Theorem 3.2.1, but it is well presented in [Bel96]. The reader can write a complete proof of the above Theorem 3.2.1 by following the hits in Exercise 3.2.4. Later in the notes we will give a detailed proof of the result that for us is of more interest: Theorem 3.1.17. Also, in the easier case of Carnot groups, see Section 8.2.1, Theorem 3.1.17 is an elementary fact.

Theorem 3.2.2 (Sussmann [Sus73, Ste74, Bel96])**.** *Let M be a manifold, $\Delta \subseteq TM$ a subbundle, and $p \in M$. Let $\Sigma \subset M$ be the set of points that can be joined to p with an absolutely continuous curve almost everywhere tangent to Δ. Then Σ is an immersed sub-manifold of M.*

A first proof of Theorem 3.2.1, modulo Theorem 3.2.2. In the assumptions of Theorem 3.2.1 we use Theorem 3.2.2. Given a vector field $X \in \Gamma(\Delta)$ and a point $q \in \Sigma$, the flow line $t \mapsto \Phi_X^t(q)$ is tangent to Δ, lies in Σ, and hence the vector X_q is tangent to the submanifold Σ. Therefore

$$\Gamma(\Delta) \subseteq \mathcal{F} := \{X \in \Gamma(TM) : X_q \in T\Sigma, \forall q \in \Sigma\}.$$

Being Σ a submanifold, the family $\mathcal{F}$ is involutive on Σ, i.e., $\mathrm{Lie}(\mathcal{F}|_\Sigma)_p = \mathrm{Lie}(\mathcal{F})_p = \mathcal{F}_p$, for all $p \in \Sigma$. Then $\mathrm{Lie}(\Gamma(\Delta)) \subseteq \mathcal{F}$. By the bracket-generating condition at p, we get

$$T_pM = \mathrm{Lie}(\Gamma(\Delta))_p \subseteq \mathcal{F}_p \subseteq T_p\Sigma.$$

From this we have $\dim M = \dim \Sigma$, and thus Σ is a neighborhood of p. □

3.2.2 Reachable sets of bracket-generating distributions

Let $\mathscr{F} \subset \mathrm{Vec}(M)$ be a family of smooth vector fields on a manifold M. Define the *reachable set for $\mathscr{F}$ from p at time less than T* as

$$\Phi_{\mathscr{F}}^{<T}(p) := \left\{ \Phi_{X_k}^{t_k} \circ \cdots \circ \Phi_{X_1}^{t_1}(p) : k \in \mathbb{N}, t_j > 0, \sum_{j=1}^{k} t_j < T, X_j \in \mathscr{F} \right\}.$$

Theorem 3.2.3. *Let $\mathscr{F}$ be a family of vector fields on a manifold M. If $-\mathscr{F} = \mathscr{F}$ and $(\mathrm{Lie}(\mathscr{F}))_p = T_pM$ for all $p \in M$, then for all $T > 0$ and for all $p \in M$, the set $\Phi_{\mathscr{F}}^{<T}(p)$ contains p in its interior.*

Proof. Unless $M = \{p\}$, there is $X_1 \in \mathscr{F}$ with $X_1(p) \neq 0$. Hence there is $\epsilon_1 \in (0, T)$ such that

$$M_1 := \{\Phi^t_{X_1}(p) : t \in (0, \epsilon_1)\}$$

is a 1-dimensional submanifold of M.

If M is 1-dimentional, the proof is concluded. If $\dim M > 1$, then there is $X_2 \in \mathscr{F}$ that is not tangent to M_1 (Otherwise $\mathrm{Lie}(\mathscr{F})$ would be tangent to M_1 and not bracket-generating on points of M_1). Let $\hat{t}_1 \in (0, \epsilon_1)$ such that

$$X_2(\Phi^{\hat{t}_1}_{X_1}(p)) \notin TM_1.$$

The map $(t_1, t_2) \mapsto \Phi^{t_2}_{X_2} \circ \Phi^{t_1}_{X_1}(p)$ has maximal rank (i.e., rank 2) at every point of the form $(\hat{t}_1, t_2)$ with t_2 sufficiently small, say $t_2 \in (0, \epsilon_2)$ with $t_1 < t_1 + \epsilon_2 < T$.

Proceeding in this way, for all k with $k \leq \dim(M)$, we obtain vector fields $X_1, \dots, X_k \in \mathscr{F}$ such that the map

$$F_k : (t_1, \dots, t_k) \mapsto \Phi^{t_k}_{X_k} \circ \dots \circ \Phi^{t_1}_{X_1}(p)$$

has maximal rank k at some point $(\hat{t}_1, \dots, \hat{t}_k)$ with $\hat{t}_j > 0$ and $\sum_j \hat{t}_j < T$. By the Constant-Rank Theorem, there is a neighborhood U_k of $(\hat{t}_1, \dots, \hat{t}_k)$ such that $M_k := F_k(U_k)$ is an embedded submanifold.

This procedure stops precisely when each element of $\mathscr{F}$ is tangent to M_k, i.e., when M_k is an open subset of M. Take $X_1, \dots, X_k \in \mathscr{F}$ such that the above defined $F_k(t_1, \dots, t_k)$ covers a neighborhood of a point $q \in M$ when $t_j > 0$, $\sum_j t_j < T$. Notice that if q is of the form $F_k(\bar{t}_1, \dots, \bar{t}_k)$, with $\bar{t}_j > 0$, $\sum_j \bar{t}_j < T$, then the map

$$q' \mapsto \Phi^{\bar{t}_1}_{-X_1} \circ \dots \circ \Phi^{\bar{t}_k}_{-X_k}(q')$$

is a diffeomorphism between some neighborhood of q and its image, which is a neighborhood of p. Notice that $-X_j \in -\mathscr{F} = \mathscr{F}$ by assumbtion. Therefore

$$(t_1, \dots, t_k) \mapsto \Phi^{\bar{t}_1}_{-X_1} \circ \dots \circ \Phi^{\bar{t}_k}_{-X_k} \circ \Phi^{t_k}_{X_k} \circ \dots \circ \Phi^{t_1}_{X_1}(p)$$

covers a neighborhood of p when $t_j > 0$ and $\sum_j t_j < T$. Thus $\Phi^{<2T}_{\mathscr{F}}(p)$ is a neighborhood of p. $\square$

Exercise 3.2.4. Use Theorem 3.2.3 and the fact that the points where (3.1.9) holds is open to give a proof of Theorem 3.2.1. $\blacksquare$

3.2.3 The metric version of Chow's theorem

We are now ready to prove Theorem 3.1.17. Namely we show that Carnot-Carathéodory distances induce the manifold topology.

Proof of Theorem 3.1.17. Let τ_M be the manifold topology and τ_{CC} the topology induced by d_{CC}.

Regarding the containment $\tau_{CC} \subset \tau_M$, let $U \in \tau_{CC}$ and $p \in U$. Then there is $T > 0$ such that $B_{d_{CC}}(p, T) \subset U$. Set

$$\mathscr{F} := \{X \in \Gamma(\Delta) : \|X(p)\| \leq 1 \ \forall p \in M\} \subset \mathrm{Vec}(M).$$

With the notation of Section 3.2.2, notice that

$$\Phi_{\mathscr{F}}^{\leq T}(p) \subset B_{d_{CC}}(p, T).$$

By Theorem 3.2.3, the point p is in the τ_M-interior of $\Phi_{\mathscr{F}}^{\leq T}(p)$. We deduce that p is in the τ_M-interior of U as well.

Regarding the containment $\tau_M \subset \tau_{CC}$, let $U \in \tau_M$. Together with the distance d_{CC} we have a Finsler distance d_F for which we have (3.1.16). Let $p \in U$. Then there is r such that $B_{d_F}(p, r) \subset U$. Since $d_F \leq d_{CC}$, then $B_{d_{CC}}(p, r) \subset B_{d_F}(p, r)$. Therefore p is in the τ_{CC}-interior of U as well. □

3.2.4 Comparison of length structures

In some situations, the case when Δ is a distribution of with varying rank is more difficult to treat. For example, next proposition is still valid when $(M, \Delta, \|\cdot\|)$ is a subFinsler manifold in the sense of Definition 3.1.22, see [**?**]. However, be aware that if one alters Definition 3.1.13 by considering rank-varying distributions (instead of polarizations) with norms defined on the whole tangent bundle, as in Definition 3.1.13, then the following proposition may be false. In fact, in that setting there are examples of smooth curves parametred by arc length that are nowhere tangent to the distribution, see Example 3.4.2. Because of this reason, we restrict to subFinsler structures of constant rank and we shall prove next proposition only according to Definition 3.1.13.

Proposition 3.2.5. *Let $(M, \Delta, \|\cdot\|)$ be a (constant-rank) subFinsler manifold equipped with its Carnot-Carathéodory distance d_{CC}. Let $\gamma : [a, b] \to M$ be a curve.*

1. *If $\mathrm{Length}_{d_{CC}}(\gamma) < \infty$, then the reparametrizion by arc length of γ is Δ-horizontal.*

2. *If γ is Δ-horizontal, then* $\mathrm{Length}_{d_{CC}}(\gamma) = \mathrm{Length}_{\|\cdot\|}(\gamma)$*; and γ is parametrized by arc length if and only if $\|\dot\gamma\| = 1$ almost everywhere.*

Proof. For part 1 recall that in every metric space every curve of finite length can be reparametrized by arc-length. (see Exercise 2.4.5). Hence, we consider that γ is parametrized by arc-length.

Let d_F be the Finsler distance for which we have (3.1.16), recall that d_F is locally biLipschitz equivalent to every Riemannian distance, see Proposition 2.3.6. Since $d_F \le d_{CC}$, we have

$$d_F(\gamma(s), \gamma(t)) \le d_{CC}(\gamma(s), \gamma(t)) \le \mathrm{Length}_{d_{CC}}(\gamma|_{[s,t]}) = |t - s|. \tag{3.2.6}$$

Thus $\gamma : [a, b] \to (M, d_F)$ is 1-Lipschitz, so in coordinates γ is (Euclidean) Lipschitz. By Rademacher Theorem, the curve γ is absolutely continuous and hence differentiable almost everywhere. Let $t_0 \in I$ be a point of differentiability for γ. We shall prove that $\dot\gamma(t_0) \in \Delta_{\gamma(t_0)}$.

Assume by contradiction that $\dot\gamma(t_0) \notin \Delta_{\gamma(t_0)}$. For simplicity we work in coordinates and assume $t_0 = 0$, $\gamma(t_0) = 0 \in \mathbb{R}^n$, $\Delta_0 = \mathbb{R}^k \times \{0\}^{n-k}$, $\dot\gamma(t_0) = e_n = (0, \ldots, 0, 1)$. We then have

$$\gamma_n(t) > t/2, \qquad \text{for } t \text{ small enough}, \tag{3.2.7}$$

where $\gamma_n(t)$ is the n-th component of γ.

We claim that for all $\epsilon > 0$ there exists $r_\epsilon > 0$ such that

$$p \in B_{d_F}(0, 2r_\epsilon),\ X \in \Delta_p,\ \|X\| \le 1 \implies |\langle \partial_n, X\rangle| < \epsilon, \tag{3.2.8}$$

where we use the Euclidean scalar product making ∂_i orthonormal. Indeed, by contradiction, there would exist $\epsilon > 0$ and sequences $(p_j)_j$ in M and $(X_j)_j$ in TM with $X_j \in \Delta_{p_j}$ such that $p_j \to 0$, $\|X_j\| \le 1$, and $|\langle \partial_n, X_j\rangle| \ge \epsilon$. Let $c > 0$ be a constant for which we have (2.3.5) in some neighbourhood of 0. Hence, eventually we have $\|X_j\|_{\mathbb{E}} \le c$. Therefore, being the sequence X_j in a compact set, up to subsequence, it converges to some Y. Since Δ is assumed to be a polarization (hence a subbundle), it is closed in TM, see Exercise 3.4.4, and since $p_j \to 0$ we have that $Y \in \Delta_0$ so

$$0 = |\langle \partial_n, Y\rangle| = \lim_{j\to\infty} |\langle \partial_n, X_j\rangle| \ge \epsilon > 0.$$

We inferred a contradiction, which gives the claim (3.2.8).

Let $\epsilon > 0$ and r_ϵ be with the above property (3.2.8). By definition of d_{CC}, we shall take a horizontal curve that almost realizes $d_{CC}(0, \gamma(r_\epsilon))$, which is not zero because of (3.2.7). In fact, there is a horizontal curve $\sigma : [0, b_\epsilon] \to M$ from 0 to $\gamma(r_\epsilon)$ such that $\|\dot\sigma\| = 1$ almost everywhere and

$b_\epsilon = \mathrm{Length}_{\|\cdot\|}(\sigma) \leq 2d_{CC}(0, \gamma(r_\epsilon)) \leq 2r_\epsilon$, where in the last inequality we used (3.2.6). Hence, first we have

$$\frac{b_\epsilon}{r_\epsilon} \leq 2, \tag{3.2.9}$$

second, we have that the image of σ is in $B_{d_F}(0, 2r_\epsilon)$. Consequently, because σ is horizontal and $\|\dot\sigma\| = 1$ almost everywhere, from (3.2.8) we have that $|\dot\sigma_n| < \epsilon$, where σ_n is the n-th component of σ, so $\dot\sigma_n = \langle \partial_n, \dot\sigma\rangle$. We then infer that

$$0 < \frac{r_\epsilon}{2} \overset{(3.2.7)}{<} \gamma_n(r_\epsilon) = \sigma_n(b_\epsilon) = \int_0^{b_\epsilon} \dot\sigma_n(s)\,\mathrm{d}\,s \leq \int_0^{b_\epsilon} |\dot\sigma_n(s)|\,\mathrm{d}\,s \leq \epsilon b_\epsilon.$$

Thus we just obtained a bound that gives a contradiction with (3.2.9) for small enough ϵ, since

$$\frac{b_\epsilon}{r_\epsilon} \geq \frac{1}{2\epsilon} \to \infty \quad \text{as } \epsilon \to 0.$$

We deduce that γ is horizontal.

Regarding part 2, let γ be a horizontal curve. On the one hand, since $d_F \leq d_{CC}$ and since $\mathrm{Length}_{\|\cdot\|} = \mathrm{Length}_{d_F}$ by Theorem 2.3.7, then $\mathrm{Length}_{\|\cdot\|} \leq \mathrm{Length}_{d_{CC}}$. On the other hand, since γ is horizontal,

$$\begin{aligned}\mathrm{Length}_{d_{CC}}(\gamma) &= \sup_{(t_1,\dots,t_k)} \sum_{i=1}^{k-1} d_{CC}(\gamma(t_{i+1}), \gamma(t_i)) \\ &\leq \sup_{(t_1,\dots,t_k)} \sum_{i=1}^{k-1} \mathrm{Length}_{\|\cdot\|}(\gamma|_{[t_i,t_{i+1}]}) \\ &= \mathrm{Length}_{\|\cdot\|}(\gamma),\end{aligned}$$

where the suprema are over all the partitions $(t_1, \dots, t_k)$ of the domain of γ. □

Corollary 3.2.10. *Carnot-Carathéodory spaces are length spaces.*

3.2.5 Existence of geodesics in CC spaces

Theorem 3.2.11 (Hopf-Rinow Theorem for CC spaces)**.** *Let M be a CC space.*

1. *Every point in M has a neighborhood in which every two points can be joined with a curve that is length minimizing with respect to the CC distance.*

2. *If M is boundedly compact, then it is a geodesic space.*

Proof. By Chow's theorem, since M is connected and Δ bracket-generating, the distance function d_{CC} is finite and the topology of (M, d_{CC}) is locally compact. Moreover, it is a length space, by Corollary 3.2.10.

To find shortest paths, we shall use Proposition 2.1.6. In fact, let $p \in M$ and take $r > 0$ small enough so that the closed ball $\bar{B}(p, r)$ is compact. We claim that every two points $p_1, p_2 \in B(p, r/2)$ can be joined with a length minimizing curve. Indeed, by Proposition 2.1.6, there exists a curve σ from p_1 to p_2 that is one of the shortest among the curves contained in $\bar{B}(p, r)$. On the one hand, notice that the length of σ is at most r, the reason being that each of p_1, p_2 can be connected to p via a curve of length strictly less than $r/2$, which therefore is in $\bar{B}(p, r)$. On the other hand, every other curve from p_1 to p_2 that leaves $\bar{B}(p, r)$ has length at least r, because it starts in $B(p, r/2)$ leaves $\bar{B}(p, r)$ and returns into $B(p, r/2)$. Therefore, the curve σ is a length minimizing curve.

If in addition (M, d_{CC}) is boundedly compact, we can conclude by Proposition 2.1.8 □

3.3 Ball-Box Theorem and Hausdorff dimension

3.3.1 Ball-Box Theorem

Let $(M, \Delta, \|\cdot\|)$ be an equiregular subFinsler manifold of topological dimension n. Let

$$\Delta = \Delta^{[1]} \subset \Delta^{[2]} \subset \cdots \subset \Delta^{[s]} = TM$$

be the flag of subbundles. Since the next considerations will be of local nature, we assume that there exists a frame $X_1, \dots, X_n$ for TM and there are $m_1, \dots, m_s$ such that $X_1, \dots, X_{m_k}$ is a frame for $\Delta^{[k]}$. In this case we say that $X_1, \dots, X_n$ is an *equiregular frame*. Equiregular frames are also called *adapted frames*.

Notice that, for all $p \in M$,

$$m_j = \dim \Delta^{[j]}(p). \tag{3.3.1}$$

We also say that X_j has *degree* d_j if, for all $p \in M$,

$$X_j(p) \in \Delta^{[d_j]} \setminus \Delta^{[d_j - 1]}, \tag{3.3.2}$$

i.e., $j \in \{m_{d_j - 1} + 1, \dots, m_{d_j}\}$. We might denote d_j by $\deg(X_j)$.

The plan is to parametrize the manifold M using the flow of linear sums of $X_1, \dots, X_n$. To such vector fields we associate an *exponential coordinate map* from a point $p \in M$ as

$$\Phi_p : \mathbb{R}^n \to M, \qquad (t_1, \dots, t_n) \mapsto \Phi^1_{t_1 X_1 + \cdots + t_n X_n}(p), \tag{3.3.3}$$

where $\Phi^1_X(p)$ is the flow of X at time 1 starting from p. Such a map might be defined only on a neighborhood of $0 \in \mathbb{R}^n$. However, for the sake of simplicity and for the fact that this is the case for Lie groups, we assume that Φ_p is globally defined.

We define the *box* with respect to the numbers $d_1, \dots, d_n$ and radius $r > 0$ as

$$\mathtt{Box}(r) := \left\{ (t_1, \dots, t_n) \in \mathbb{R}^n : |t_j| \le r^{d_j} \right\}. \tag{3.3.4}$$

The following comparison theorem is due to many people (Mitchell, Gershkovich, Nagel-Stein-Wainger, cf. [Gro99]) and it is called Ball-box Theorem since it compares the box $\mathrm{Box}(r)$ in $\mathbb{R}^n$ with the ball $B(p, r')$ with respect to the d_{CC} distance, with a biLipschitz relation between r and r'.

Theorem 3.3.5 (Ball-Box Theorem)**.** *Let $(M, \Delta, \|\cdot\|)$ be a subFinsler manifold. Assume Δ is equiregular. Fix $\bar{p} \in M$ and an equiregular frame $X_1, \dots, X_n$ in a neighborhood of $\bar{p}$ with degrees $d_1, \dots, d_n$ and related boxes $\mathtt{Box}(\cdot)$. Then there is a neighborhood U of $\bar{p}$ in M and there is $C > 1$ and $\rho > 0$ such that*

$$B_{d_{CC}}(p, r/C) \subset \Phi_p(\mathtt{Box}(r)) \subset B_{d_{CC}}(p, Cr), \qquad \forall p \in U, \forall r \in (0, \rho).$$

The Ball-Box Theorem will not be proved here in this generality. It will be proved later in the easier case of Carnot groups, see Theorem 8.2.8.

Remark 3.3.6. The Ball-Box Theorem 3.3.5 gives a quantitative version of Chow's theorems 3.2.1 and 3.1.17.

As far as we know, nothing is known regarding the following natural question. Except (maybe...) for contact 3-manifolds – TO BE CHECKED.

Question 3.3.7 (Open!)**.** *Are all sufficiently small sub-Finsler balls and spheres homeomorphic to the usual Euclidean balls and spheres?*

Here is a first consequence of the Ball-Box Theorem 3.3.5.

Corollary 3.3.8 (Hölder equivalence between CC and Euclidean metrics)**.** *Locally, each sub-Finsler manifold is Hölder equivalent to a Riemannian manifold. Namely, if s is the step, then locally near around every point there exists $C > 1$ such that*

$$\frac{1}{C}(d_{CC})^s \le d_{\mathrm{Riem}} \le C d_{CC}. \tag{3.3.9}$$

Proof. Let $(M, \Delta, \|\cdot\|)$ be the sub-Finsler manifold. Let g be a Riemannian tensor whose norm is smaller than $\|\cdot\|$ and denote by d_{Riem} the induced Riemannian distance.

Consider the identity map $\mathrm{id} : M \to M$. Obviously the map

$$\mathrm{id} : (M, d_{CC}) \to (M, d_{\text{Riem}})$$

is 1-Lipschitz, and so Hölder.

For the other direction, let $s := \max_j d_j$ the maximum of the degree d_j of the vector fields of some equiregular basis $\{X_j\}$, i.e, s is the step of Δ. Notice that, for $r \in (0, 1)$, one has that

$$B_E(0, r^s) \subset \prod_{j=1}^{n} [-r^s, r^s] \subset \mathrm{Box}(r),$$

where B_E denotes the Euclidean ball in $\mathbb{R}^n$. Therefore, using the second inclusion of the Ball-Box Theorem 3.3.5 and the fact that the exponential maps Φ_p are locally biLipschitz maps (locally uniformly in p), see Exercise 3.3.10, we get that

$$B_{d_{CC}}(p, Cr) \supseteq \Phi_p(\mathrm{Box}(r)) \supseteq \Phi_p(B_E(0, r^s)) \supseteq B_{d_{\text{Riem}}}(p, C' r^s).$$

Hence, the map

$$\mathrm{id} : (M, d_{\text{Riem}}) \to (M, d_{CC})$$

is $1/s$-Hölder on compact sets. □

Exercise 3.3.10. Show that the maps $\Phi_p : \mathbb{R}^n \to M$ from (3.3.3) are locally biLipschitz maps locally uniformly in p: Namely, fix a compact subset K of M and a Riemannian distance d_{Riem} on M, then there exists $C > 1$ and exists a neighborhood U of 0 in $\mathbb{R}^n$ such that of all $p \in K$ the map $\Phi_p|_U$ is a C-biLipschitz homeomorphism between U equipped with the Euclidean distance and its image equipped with d_{Riem}. ■

3.3.2 Dimensions of CC spaces

Definition 3.3.11 (Homogeneous dimension)**.** If a distribution Δ is equiregular, we define its *homogeneous dimension* as the natural number

$$Q := Q_\Delta := \sum_{j=1}^{n} j \left(\dim \Delta^{[j]}(p) - \dim \Delta^{[j-1]}(p) \right), \tag{3.3.12}$$

which is independent on p.

In other words, in terms of the numbers $m_1, \dots, m_s$ from 3.3.1, we write

$$Q = m_1 + 2(m_2 - m_1) + 3(m_3 - m_2) + \cdots + s(m_s - m_{s-1}). \tag{3.3.13}$$

Notice that the box defined in (3.3.4) satisfies

$$\mathcal{L}^n(\texttt{Box}(r)) = r^Q,$$

where $\mathcal{L}^n$ is the Lebesgue measure in $\mathbb{R}^n$. In terms of the degrees of the vector fields, as in (3.3.2), we also have

$$Q = \sum_{j=1}^{n} d_j. \tag{3.3.14}$$

Corollary 3.3.15. *If a subFinsler manifold $(M, \Delta, \|\cdot\|)$ has an equiregular distribution, then the Hausdorff dimension of (M, d_{CC}) equals the homogeneous dimension Q. Moreover, the Q-dimensional Hausdorff measure of (M, d_{CC}) is locally biLipschitz equivalent to each volume form.*

In particular, if $TM \neq \Delta$, the Hausdorff dimension is strictly greater than the topological dimension.

Proof. We auxiliarily fix a Riemannian structure. Since all the volume forms are locally biLipschitz equivalent, we assume that the volume form is the Riemannian volume form vol.

Using notation of the Ball-Box Theorem 3.3.5, let k be the (locally uniform) biLipschitz constant of the exponential map Φ_p with respect to the Riemannian distance on the n-manifold M and the Euclidean distance on $\mathbb{R}^n$, see Exercise 3.3.10. Since vol (resp., the Lebesgue measure $\mathcal{L}^n$) is the n-dimensional Hausdorff measure of the Riemannian manifold M (resp., of the Euclidean space $\mathbb{R}^n$), we have, for small r,

$$\frac{1}{k^n}\mathcal{L}^n(\texttt{Box}(r)) \leq \mathrm{vol}(\Phi_p(\texttt{Box}(r))) \leq k^n \mathcal{L}^n(\texttt{Box}(r)).$$

If Q is the homogeneous dimension, by the Ball-Box theorem then we get, for small r,

$$\frac{1}{k^n C^Q} r^Q \leq \mathrm{vol}(B_{d_{CC}}(p, r)) \leq k^n C^Q r^Q.$$

By Theorem 2.1.30 and Remark 2.1.32, we conclude. □

3.3.3 Dimensions of submanifolds in CC spaces

Computing the Hausdorff dimension and Hausdorff measure of submanifolds in sub-Finsler manifolds with respect to the Carnot-Carathéodory distance is a rather natural question. In 0.6 B of

[Gro99], Gromov has given a general formula for the Hausdorff dimension of smooth submanifolds in equiregular Carnot-Carathéodory spaces and in [Mag08a] it is shown that this formula coincides with the degree of the submanifold, recently introduced in [MV08].

Theorem 3.3.16 ([Gro99, page104])**.** *Let $(M, \Delta, \|\cdot\|)$ be a sub-Finsler manifold with an equiregular distribution Δ and Carnot-Carathéodory distance d_{CC}. Let $\Sigma \subset M$ a smooth sub-manifold. Then the Hausdorff dimension of (Σ, d_{CC}) is*

$$\dim_H(\Sigma, d_{CC}) = \max \left\{ \sum_{j=1}^{n} j \cdot \operatorname{rank} \left(T_p\Sigma \cap \Delta^{[j]}(p) \right) \Big/ \left(T_p\Sigma \cap \Delta^{[j-1]}(p) \right) \ : \ p \in \Sigma \right\}.$$

Nevertheless, the question regarding Hausdorff measures of smooth submanifolds has not yet an answer. In [MV08] Magnani and Vittone found an integral formula for the spherical Hausdorff measure of submanifolds in Carnot groups under a suitable ‘negligibility condition’. This negligibility condition has been recently obtained in all two step groups, [Mag08a] using standard covering arguments, and in the Engel group, using blow-up arguments [LM10]. However it is still open in higher step groups and in general sub-Riemannian manifolds. We address the reader to the work of Magnani [MV08, Mag08b, Mag08a] for more information on this problem and its connections with the literature.

3.4 Exercises

Exercise 3.4.1 (Grushin distribution)**.** Show that on $\mathbb{R}^2$ coordinates (x, y) the vector fields

$$X = \partial_x \qquad \text{and} \qquad Y = x\partial_y,$$

satisfy the generating condition (3.1.12) and define a bracket-generating distribution, whose rank is not constant. ■

Exercise 3.4.2. On $\mathbb{R}^2$ with coordinates (x, y), consider the distribution Δ generated by the vector fields

$$X = \partial_x \qquad \text{and} \qquad Y = x\partial_y.$$

Consider the continuously varying norm $\|\cdot\|$ given by the Euclidean norm for every tangent vector. Prove that

(i) the CC distance d_{CC} induced by $\Delta, \|\cdot\|$ is the Euclidean distance.

(ii) The curve $t \in \mathbb{R} \mapsto (0, t)$ is parametred by arc length but at no point it is tangent to the distribution Δ. ■

Exercise 3.4.3. Show that on $\mathbb{R}^2$ coordinates (x, y) for the vector fields

$$X = \partial_x \qquad \text{and} \qquad Y = \phi(x)\partial_y,$$

we have

$$\mathrm{Lie}(\{X, Y\}) = \mathrm{span}_{\mathbb{R}} \left\{ \partial_x, \frac{\mathrm{d}^k \phi}{\mathrm{d}x^k}(x)\partial_y : k = 0, 1, 2, \dots \right\}.$$

Consequently, X, Y do not satisfy the generating condition (3.1.12). Still they span the same distribution of Exercise 3.4.1 ■

Exercise 3.4.4. Show that every subbundle of a vector bundle is a closed subset. ■

Exercise 3.4.5. Show that a Finsler distance is a distance that induces the manifold topology. ■

Exercise 3.4.6. Show that two Finsler distances on a compact set are biLipschitz equivalent. ■

Exercise 3.4.7. Prove that Finsler-Carnot-Carathéodory distances, and in particular Riemannian and Finsler distances, are length distances. ■

Exercise 3.4.8. The Hausdorff dimension of a Riemannian n-manifold is n. ■

Exercise 3.4.9. If $\gamma : I \to (M, d_{CC})$ is parametrized by arc-length, then $\|\dot{\gamma}\| = 1$ a.e. ■

Exercise 3.4.10. Let $(M, \Delta, \|\cdot\|)$ be a sub-Finsler manifold. We denote by $\mathrm{Length}_{d_{CC}}$ and $\mathrm{Length}_{\|\cdot\|}$ respectively the length with respect to the metric d_{CC} and the length with respect to the Finsler norm $\|\cdot\|$. Let γ be a horizontal curve. Show that

$$\mathrm{Length}_{\|\cdot\|}(\gamma) = \mathrm{Length}_{d_{CC}}(\gamma).$$

■

Exercise 3.4.11. Let γ be an absolutely continuous curve in a sub-Finsler manifold. Prove that

$$\gamma \text{ is horizontal } \iff \mathrm{Length}_{d_{CC}}(\gamma) < +\infty.$$

■

Exercise 3.4.12. Denote by $\Phi^{t_i}_{X_i}$ the flow at time i with respect to a vector field X_i. Calculate the differential of

$$(t_1, \dots, t_k) \mapsto \Phi^{t_k}_{X_k} \circ \dots \circ \Phi^{t_1}_{X_1}(p).$$

■

Exercise 3.4.13. Let $\Delta^{[j]}(p)$ the vector space defined in (3.1.26). Prove that $\Delta^{[j]}(p)$ can be equivalently be defined as the subspace of T_pM spanned by all commutators of the X_i's of order $\leq j$ (including, of course, the X_i's). Namely, $X_i(p)$ has order 1; $[X_i, X_j](p)$ has order 2; $[X_i, [X_j, X_k]](p)$ has order 3; but those of order 4 are those in one of the two forms:

$$[X_i, [X_j, [X_k, X_l]]](p) \qquad \text{or} \qquad [[X_i, X_j], [X_k, X_l]](p).$$

■

Exercise 3.4.14. Let $\Delta^{[j]}(p)$ the vector space defined in (3.1.26).

1. Show that $\Delta^{[j]}$ might not be a sub-bundle of TM. [Hint: Try the distribution given by the frame $X_1 = \partial_1$, $X_2 = \partial_2 + x_1^2\partial_3$.]

2. Prove that, if $\Delta^{[j]}$ is a sub-bundle and so make sense to consider smooth sections $\Gamma(\Delta^{[j]})$ of the bundle $\Delta^{[j]}$, then

$$\Delta^{[j+1]}(p) = \Delta^{[j]}(p) + \mathbb{R}\text{-}\operatorname{span}\left\{[X,Y](p) \;:\; X \in \Gamma(\Delta), Y \in \Gamma(\Delta^{[j]})\right\}.$$

■

Exercise 3.4.15. Recall that $\Gamma(\Delta)$ denotes the smooth sections of the bundle Δ. Define $\operatorname{span}(\Delta) :=$ Lie-$\operatorname{span}\{\Gamma(\Delta)\}$. Show that the Hörmander's condition is equivalent to $\operatorname{span}(\Delta) = TM$. (What is not immediately obvious is that elements of the form $[[X_1, X_2], [X_3, X_4]]$, with $X_1, X_2, X_3, X_4 \in \Gamma(\Delta)$, are contained in some $\Delta^{[j]}(p)$.) ■

Exercise 3.4.16. Show that if $(M, \Delta, \|\cdot\|)$ is a subFinsler manifold with induced distance d_{CC}, then the metric space (M, d_{CC}) is homeomorphic to the manifold M via the identity map. ■

Exercise 3.4.17. Show, without using Theorem 3.3.16, that each smooth surface in the Heisenberg group has Haudorff dimension equal to 3. ■

Exercise 3.4.18. Give a proof of Theorem 3.3.16. ■

Chapter 4

A review of Lie groups

In the following chapter, we will review the theory of Lie groups. This revision serves two purposes: First, subriemannian structures on Lie groups are highly interesting and arise in various contexts, including mechanics. They are, in a sense, easier to study than general subriemannian manifolds. Second, we will explore the property that arbitrary subriemannian manifolds have tangent spaces that are special subRiemannian Lie groups.

The prerequisites for understanding Lie groups and Lie algebras primarily lie in the realm of differential geometry. The results presented in this chapter are classical and are based on the references: [War83, CG90, HN12].

4.1 Lie groups, Lie algebras, and their morphisms

In this section, we will review the following concepts: Lie group, Lie algebra, Lie algebra associated with a Lie group, Lie subgroup, Lie subalgebra, Lie group homomorphism, Lie algebra homomorphism, and Lie algebra homomorphism induced by a Lie group homomorphism. We will also state certain results regarding these objects, but the proofs will be deferred to later sections.

For clarity, we provide a reminder that a *group* is a set G equipped with a binary operation, referred to as its *product* or *group product* , denoted by the symbol $\cdot$. The product is a function $(a, b) \in G \times G \mapsto a \cdot b \in G$ that satisfies associativity, the existence of an identity element, and an inversion map. The inversion map is denoted as $a \mapsto a^{-1}$. The identity element of a group G is denoted by 1. If there is a need to emphasize that 1 is specifically the identity element of the group G, it can be denoted as 1_G. Other texts or references may use alternative symbols such as e or e_G.

Let G be a group and $g \in G$. The *left translation* by g is the bijection

$$\begin{array}{rccc} L_g: & G & \longrightarrow & G \\ & h & \mapsto & gh. \end{array}$$

The *right translation* by g is the bijection

$$\begin{array}{rccc} R_g: & G & \longrightarrow & G \\ & h & \mapsto & hg. \end{array}$$

The *conjugation* by g is the bijection

$$\begin{array}{rccc} C_g: & G & \longrightarrow & G \\ & h & \mapsto & ghg^{-1}. \end{array}$$

We shall focus on Lie groups, which are differentiable manifolds with a smooth group operation. However, some of the remarks we will hold in the general setting of topological groups: A *topological group* is a group together with a Hausdorff topology for which the group product and the inversion map are continuous. Lie groups are special topological groups:

Definition 4.1.1 (Lie group)**.** A Lie group is a differentiable manifold (second countable, but not necessarily connected) together with a group structure such that both

$$\text{the product} \quad \begin{array}{ccc} G \times G & \to & G \\ (x,y) & \mapsto & x \cdot y \end{array} \qquad \text{and the inverse} \quad \begin{array}{ccc} G & \to & G \\ g & \mapsto & g^{-1} \end{array} \tag{4.1.2}$$

are C^∞ maps.

As in every manifold, the set $\Gamma(TG)$ of vector fields on G forms a Lie algebra. The general notion of Lie algebra is the following:

Definition 4.1.3 (Lie algebra)**.** A *Lie algebra* $\mathfrak{g}$ *(over* $\mathbb{R}$*)* is a vector space (over $\mathbb{R}$) together with a bilinear operation $[\cdot,\cdot]: \mathfrak{g} \times \mathfrak{g} \to \mathfrak{g}$ called *Lie bracket*, such that for all $X, Y, Z \in \mathfrak{g}$ the following two properties hold:

$[X,Y] = -[Y,X]$ (called *anti-commutativity*),

$[[X,Y],Z] + [[Y,Z],X] + [[Z,X],Y] = 0$ (called *Jacobi identity*).

Lie algebras are usually denoted by gothic letters. The gothic letters for g, h, n, o, l, p, s are $\mathfrak{g}, \mathfrak{h}, \mathfrak{n}, \mathfrak{o}, \mathfrak{l}, \mathfrak{p}, \mathfrak{s}$. Lie algebras can also be considered on other fields. However, in this text we shall only consider those over the real numbers. The structure of a Lie algebra can be represented via expressing the Lie bracket using a basis. Namely, if $\mathfrak{g}$ is a Lie algebra with bracket $[\cdot,\cdot]$ and $X_1, \dots, X_n$

is an ordered basis of $\mathfrak{g}$ as vector space, then the *structural constants* of $\mathfrak{g}$ with respect to $X_1, \dots, X_n$ are the real numbers $c_{ij}^k \in \mathbb{R}$ with $i, j, k \in \{1, \dots, n\}$ such that

$$[X_i, X_j] = \sum_{k=1}^{n} c_{ij}^k X_k, \qquad \forall i, j \in \{1, \dots, n\}. \tag{4.1.4}$$

The data $X_1, \dots, X_n$ and $(c_{ij}^k)_{i,j,k=1}^k$ record the whole info about the Lie bracket, see Exercises 4.7.6, 4.7.8, and 4.7.9.

The importance of the concept of Lie algebra is that there is a special finite-dimensional Lie algebra intimately associated with each Lie group, and that properties of the Lie group are reflected in properties of its Lie algebra. We shall recall, for example, that simply connected Lie groups are completely determined (up to isomorphism) by their Lie algebras, see Corollary 4.1.9.

The Lie algebra associated to a group is isomorphic, as a vector space, to the tangent space $T_{1_G}G$ at the identity element 1_G. In order to define a Lie bracket structure, one identifies $T_{1_G}G$ as a subset of $\Gamma(TG)$, by extending each vector to a vector field. Forced to make a choice[1], we follow the majority of the literature focusing on the *left* invariant vector fields, i.e., the vector fields $X \in \Gamma(TG)$ such that $(L_g)_* X = X$, so that $(\mathrm{d}L_g)_x X_x = X_{L_g(x)}$ for all $x \in G$. Thanks to (2.4.20) with $F = L_g$, the class of left-invariant vector fields is easily seen to be closed under the Lie bracket, see Exercise 4.7.11. In other words, the set of left-invariant vector fields form a Lie algebra.

Note that, after fixing a vector $v \in T_{1_G}$, we can construct a left-invariant vector field X defining $X_g := (\mathrm{d}L_g)_{1_G}(v)$ for $g \in G$. This construction is a linear isomorphism between the set of all left-invariant vector fields and T_{1_G}, and proves that left-invariant vector fields form an n-dimensional subspace of $\Gamma(TG)$, where $n := \dim G$. We denote by $\mathfrak{g}$ the vector space T_{1_G} equipped with the Lie bracket coming from the identification with the left-invariant vector fields. Such a $\mathfrak{g}$ is called the *Lie algebra* of G and it is occasionally denoted by $\mathrm{Lie}(G)$. We next summarise this definition:

Definition 4.1.5 (Lie algebra of a Lie group)**.** Let G be a Lie group. The *Lie algebra* of G, denoted by Lie(G), has two realizations:

Interpretation 1: Lie(G) is the linear space $\mathrm{LIVF}(G)$ of left-invariant vector fields on G endowed with the bracket of vector fields.

Interpretation 2: Lie(G) is the tangent space $T_{1_G}G$ equipped with the bracket

$$[X, Y] := [\tilde{X}, \tilde{Y}]_{1_G}, \quad \forall X, Y \in T_{1_G}G,$$

[1] Actually, we prefer to consider left-actions by a group (on itself), because we think of groups as transformations, and we are nowadays used to put symbols of functions on the left of variables, like $f(x)$.

where $\tilde{X}, \tilde{Y}$ are the left-invariant vector fields such that $\tilde{X}_{1_G} = X$ and $\tilde{Y}_{1_G} = Y$. We shall use alternatively both points of view.

Let G be a Lie group and $H < G$ a subgroup. We say that H is a *Lie subgroup* of G if H admits the structure of Lie group such that the inclusion $H \hookrightarrow G$ is a smooth group homomorphism. It is a consequence that the inclusion is actually an immersion, see Exercise **??**. A Lie subgroup $H < G$ is said a *closed Lie subgroups* if H is topologically closed within G. It is a consequence that in this case the inclusion $H \hookrightarrow G$ is an embedding, see Exercise **??**. Closed Lie subgroups are also called *regular Lie subgroups.*

A *subalgebra* of a Lie algebra $\mathfrak{g}$ is a vector subspace $\mathfrak{h} \subset \mathfrak{g}$ that is closed under the Lie bracket operation of $\mathfrak{g}$. Hence, if H is a Lie subgroup of a Lie group G , then $\mathrm{Lie}(H)$ is canonically isomorphic to a subalgebra of $\mathrm{Lie}(G)$, (exercise). Viceversa, every subalgebra comes from a Lie subgroup:

Theorem 4.1.6 (Existence of subgroups, see Theorem 4.6.1)**.** *Let G be a Lie group. For every subalgebra $\mathfrak{h} \subset \mathrm{Lie}(G)$, there is a unique connected Lie subgroup H with Lie algebra $\mathfrak{h}$.*

We next discuss the maps of the categories in which the objects are the Lie groups and the Lie algebras, respectiverly. A map $\varphi : G \to H$ between groups is a *(group) homomorphism* if

$$\varphi(g_1 \cdot g_2) = \varphi(g_1) \cdot \varphi(g_2), \qquad \forall g_1, g_2 \in G.$$

If G, H are Lie groups, then a homomorphism $\varphi : G \to H$ is said a *Lie group homomorphism* if it is smooth. If in addition $H = G$, then φ is called *Lie group endomorphism.* A bijective Lie group homomorphism is called *Lie group isomorphism.* A bijective Lie group endomorphism is an *Lie group automorphism.*

A map $\psi : \mathfrak{g} \to \mathfrak{h}$ between Lie algebras is said a *Lie algebra homomorphism* if it is both linear and preserves brackets:

$$\psi([X, Y]) = [\psi(X), \psi(Y)], \qquad \forall X, Y \in \mathfrak{g}.$$

If in addition $\mathfrak{h} = \mathfrak{g}$, then ψ is called *Lie algebra endomorphism.* A bijective Lie algebra homomorphism (resp. endomorphism) is called *Lie algebra isomorphism* (resp. *automorphism*).

The first connection between Lie groups and their Lie algebras is that each Lie group homomorphism induces a Lie algebra homomorphism: if $\varphi : G \to H$ is a Lie group homomorphism, note that

$\varphi(1_G) = 1_H$, and one can easily show that the differential at the identity

$$\varphi_* := d\varphi_{1_G} : T_{1_g}G \to T_{1_H}H \tag{4.1.7}$$

preserves the Lie bracket operation, see Exercise 4.7.18. Namely, $\varphi_* : \mathrm{Lie}(G) \to \mathrm{Lie}(H)$ is a Lie algebra homomorphism, called the *Lie algebra homomorphism induced* by φ.

Vice versa, in the case when G is a Lie group that as a topological space is simply connected, then each Lie algebra homomorphism come from a Lie group homomorphism. Recall that a topological space X is called *simply connected* if it is path-connected and every loop in X is homotopic to a constant.

Theorem 4.1.8 (Induced Lie group homomorphism, see Theorem 4.6.4)**.** *Let G, H be Lie groups. Assume G simply connected. For all Lie algebra homomorphism $\psi : \mathrm{Lie}(G) \to \mathrm{Lie}(H)$, there exists a unique Lie group homomorphism $\varphi : G \to H$ with $\varphi_* = \psi$.*

Corollary 4.1.9. *If simply connected Lie groups G and H have isomorphic Lie algebras, then G and H are isomorphic.*

As a consequence of a theorem due to Ado, see [Jac79, page 199], for every Lie algebra $\mathfrak{g}$ there exists a simply connected Lie group G with Lie algebra $\mathfrak{g}$. We then have the following correspondence.

Theorem 4.1.10. *There is a one-to-one correspondence between isomorphism classes of Lie algebras and isomorphism classes of simply connected Lie groups.*

We shall only prove the above theorem, together with Ado's result, only for stratified algebras, since the proof is much easier and it is what is need for the Lie groups of our interest: the Carnot groups.

4.2 Exponential map

Let M be a differentiable manifold. Consider a smooth vector field $X \in \Gamma(TM)$. Given a point $p \in M$, there exists a unique curve $t \mapsto \gamma(t)$ satisfying $\gamma(0) = p$ and having a tangent vector $\dot\gamma(t) = X_{\gamma(t)}$. We refer to this curve as the integral curve of X passing through p. The exponential map associated with X is defined as $\Phi^1_X(p) = \gamma(1)$, which gives us the endpoint of the integral curve after a unit time parameterization. It should be noted that the exponential map is generally defined locally in X, meaning that it is only defined in a small neighborhood of zero in T_pM and maps it to a neighborhood of p in the manifold. This locality arises from the reliance on the theorem of existence and uniqueness of ordinary differential equations, which is itself local in nature.

In the theory of Lie groups, the exponential map is a map from the Lie algebra $\mathfrak{g}$ to the group G, denoted as

$$\exp\colon \mathfrak{g} \to G.$$

Here, elements of the Lie algebra $\mathfrak{g}$ are identified with left-invariant vector fields, and thus we have $\mathfrak{g} \subset \Gamma(TG)$. Therefore, we can apply the previous point of view of flows where p is taken to be the identity element 1_G of the group. Furthermore, it can be shown that for every $X \in \mathfrak{g}$, the ordinary differential equation $\dot{\gamma}(t) = X_{\gamma(t)}$ has global solutions. In fact, these integral curves $\gamma(t)$ correspond to group homomorphisms from the additive group $\mathbb{R}$ to the group G. Such homomorphisms from $\mathbb{R}$ to G are commonly referred to as *one-parameter subgroups*.

4.2.1 One-parameter subgroups

Definition 4.2.1 (One-parameter subgroup)**.** Let G be a Lie group. A Lie group homomorphism $\theta : \mathbb{R} \to G$ is called a *one-parameter subgroup* (OPS, for short). With abuse of terminology, sometimes we say that a one-parameter subgroup is the image $\theta(\mathbb{R}) \subset G$ of such a map.

Equivalently, $\theta : \mathbb{R} \to G$ is a one-parameter subgroup if and only if

i). θ is smooth,

ii). $\theta(0) = 1_G$,

iii). $\theta(t+s) = \theta(t) \cdot \theta(s)$, for all $s, t \in \mathbb{R}$.

We will soon see that the one-parameter subgroups are exactly the integral curves from the identity element of the left-invariant vector fields (and also of the right invariant vector fields). Recall that we denote by $\Phi^t_X(p)$ the *flow* of a vector field X at time t starting from a point p.

Proposition 4.2.2. *Let G be a Lie group and X be a left-invariant vector field on G.*

i). The flow line $t \mapsto \Phi^t_X(1_G)$ of X from 1_G is a one-parameter subgroup.

ii). If $\theta : \mathbb{R} \to G$ is a one-parameter subgroup with $\dot{\theta}(0) = X_{1_G}$, then $\theta(t) = \Phi^t_X(1_G)$, for all $t \in \mathbb{R}$.

Proof. i). Let $\sigma(t) = \Phi_X^t(1_G)$, which is defined for t in some maximal interval $(-\epsilon, \epsilon)$. Fix $s \in (-\epsilon, \epsilon)$ and consider $\gamma(t) := \sigma(s) \cdot \sigma(t)$. We claim that γ is the integral curve of X from $\sigma(s)$. Indeed, we have

$$\begin{aligned}\dot{\gamma}(t) &= \frac{\mathrm{d}}{\mathrm{d}t}(\sigma(s) \cdot \sigma(t)) \\ &= \frac{\mathrm{d}}{\mathrm{d}t}(L_{\sigma(s)}(\sigma(t))) \\ &= (\mathrm{d}L_{\sigma(s)})_{\sigma(t)}\sigma'(t) \\ &= (\mathrm{d}L_{\sigma(s)})_{\sigma(t)}X_{\sigma(t)} \\ &= X_{\sigma(s)\cdot\sigma(t)} \\ &= X_{\gamma(t)}.\end{aligned}$$

By uniqueness of integral curves, we have $\gamma(t) = \sigma(s+t)$ and so $\sigma(s+t) = \sigma(s) \cdot \sigma(t)$.

Moreover, since σ can be prolonged by $L_{\sigma(s)}\gamma$, then σ is defined on all $\mathbb{R}$.

ii). Being θ a one-parameter subgroup, we have $\theta(s+t) = \theta(s) \cdot \theta(t) = L_{\theta(s)}(\theta(t))$. Hence, since $\dot{\theta}(0) = X_1$, we have

$$\begin{aligned}\dot{\theta}(s) &= \frac{\mathrm{d}}{\mathrm{d}t}\theta(s+t)\Big|_{t=0} \\ &= \frac{\mathrm{d}}{\mathrm{d}t}L_{\theta(s)}(\theta(t))\Big|_{t=0} \\ &= \left(\mathrm{d}L_{\theta(s)}\right)_{\theta(0)}\dot{\theta}(0) \\ &= \left(\mathrm{d}L_{\theta(s)}\right)_1 X_1 \\ &= X_{\theta(s)}.\end{aligned}$$

So θ is the integral curve of X from 1_G. $\square$

Remark 4.2.3. If θ is a OPS, the $\theta(\mathbb{R})$ is a Lie subgroup. Indeed, if $\dot{\theta}(0) = 0$, then θ is constantly equal to 1_G, which is a Lie subgroup of dimension 0. If instead $\dot{\theta}(0) \neq 0$, then $\dot{\theta}(t) \neq 0$ for all $t \in \mathbb{R}$, and θ is an immersion. Hence $\theta(\mathbb{R}) \subset G$ is a Lie subgroup of dimension 1.

4.2.2 Exponential map

Definition 4.2.4 (Exponential map)**.** Let G be a Lie group and $\mathfrak{g}$ its Lie algebra, seen as left-invariant vector fields. The *exponential map*

$$\exp : \mathfrak{g} \to G$$

is defined as, for all $X \in \mathfrak{g}$,

$$\exp(X) := \Phi_X^1(1_G),$$

i.e., $\exp(X)$ is the flow of X at time 1 starting from 1_G.

Remark 4.2.5. The exponential map is in general different from the exponential map of Riemannian geometry. In Exercise 4.7.37 one can see that the exponential map of the Lie group $\mathrm{GL}^+(n, \mathbb{R})$ is not a Riemannian exponential for any Riemannian metric. However, if a Lie group is compact, then it has a Riemannian metric invariant under left and right translations, and the Lie group exponential map is the Riemannian exponential map of this Riemannian metric, see Section 7.1.3.

One first key property of the exponential map is the following (cf. Exercise 4.7.21).

Proposition 4.2.6. *For every left-invariant vector field X the curve $t \mapsto \exp(tX)$ is a one-parameter subgroup and an integral curve of X. For every $g \in G$, the curve $t \mapsto g\exp(tX) = L_g(\exp(tX))$ is the flow line of X starting at g.*

Consequently, first, we infer that left-invariant vector fields are complete. Second, we proved that the flows of *left* invariant vector fields are *right* translations, as we next express (see also Exercise 4.7.25).

Proposition 4.2.7. *Let X be a left-invariant vector field on a Lie group G. Then*

$$\Phi_X^t = R_{\exp(tX)}, \qquad \forall t \in \mathbb{R}. \tag{4.2.8}$$

Likewise, if we let $X^\dagger$ be the right-invariant vector field such that $(X^\dagger)_1 = X_1$, then we also have that

$$\exp(tX) = \Phi_{X^\dagger}^t(1_G) \qquad \text{and} \qquad \Phi_{X^\dagger}^t = L_{\exp(tX)}, \quad \forall t \in \mathbb{R}, \tag{4.2.9}$$

see Exercise 4.7.26. From the fact that we explicitly know the above flows (4.2.8) and (4.2.9), we have many consequences, see Exercises 4.7.27, 4.7.28, and 4.7.29.

We summarise the following three interpretations for the exponential map:

$$\exp(X) = \begin{cases} \text{flow at time 1 of the LIVF } X, \\ \text{OPS at time 1 tangent to } X_{1_G} \text{ (or } X^\dagger_{1_G}), \\ \text{flow at time 1 of the RIVF } X^\dagger. \end{cases}$$

Next is one very important feature of the exponential map. It implies that exp gives a local parametrization near 1_G.

Proposition 4.2.10. *Let G be a Lie group with Lie algebra $\mathfrak{g}$. Then* $\exp : \mathfrak{g} \to G$ *is smooth and* $(d\exp)_0$ *is the identity map,*

$$(d\exp)_0 = \mathrm{id}_{\mathfrak{g}} : \mathfrak{g} \to \mathfrak{g},$$

so exp *gives a diffeomorphism of a neighborhood of* 0 *in* $\mathfrak{g}$ *onto a neighborhood of* e *in* G*;*

Proof. For the smoothness of exp, we refer to Exercise 4.7.31. Regarding its differential, fix $X \in \mathfrak{g} = T_1G$. Let $\sigma : \mathbb{R} \to \mathfrak{g}$ be the curve $\sigma(t) := tX$ so that $\sigma'(0) = X$. Then

$$\begin{aligned}
(\mathrm{d}\exp)_0(X) &= (\exp\circ\sigma)'(0) \\
&= \frac{\mathrm{d}}{\mathrm{d}t}\exp(tX)\Big|_{t=0} \\
&= \frac{\mathrm{d}}{\mathrm{d}t}\Phi^1_{t\tilde{X}}(1)\Big|_{t=0} \\
&= \frac{\mathrm{d}}{\mathrm{d}t}\Phi^t_{\tilde{X}}(1)\Big|_{t=0} \\
&= X,
\end{aligned}$$

where $\tilde{X}$ is the left-invariant vector fields with $\tilde{X}_{1_G} = X$. The last part of the statement of the proposition is a consequence of the Inverse Function Theorem. ☐

The exponential map gives a first link between the Lie group level and the Lie algebra level with the following result.

Proposition 4.2.11. *Let $\varphi : G \to H$ be a Lie group homomorphism. If* $\varphi_* : \mathrm{Lie(G)} \to \mathrm{Lie(H)}$ *is the induced Lie algebra homomorphism (see* (4.1.7)*, then*

$$\exp\circ\varphi_* = \varphi\circ\exp,$$

i.e., the following diagram commutes.

$$\begin{tikzcd}
\mathrm{Lie(G)} \arrow[r, "\varphi_*"] \arrow[d, "\exp"'] & \mathrm{Lie(H)} \arrow[d, "\exp"] \\
G \arrow[r, "\varphi"'] & H
\end{tikzcd}$$

Proof. We need to show that for every left-invariant vector field X

$$\varphi(\exp(X)) = \exp(\widetilde{(\mathrm{d}\varphi)_1}X_1).$$

We plan to show that for all left-invariant vector field X and for all $t \in \mathbb{R}$

$$\sigma(t) := \varphi(\exp(tX)) = \exp(t\widetilde{(\mathrm{d}\varphi)_1}X_1).$$

Namely, we claim that the curve $t \mapsto \sigma(t)$ is the one-parameter subgroup in H generated by $(\,\mathrm{d}\varphi)_1 X_1$.

First, we check that σ is a one-parameter subgroup:

$$\begin{aligned}\sigma(s)\sigma(t) &= \varphi(\exp(sX))\varphi(\exp(tX)) \\ &= \varphi\left(\exp(sX)\exp(tX)\right) \\ &= \varphi\left(\exp((s+t)X)\right) \\ &= \sigma(s+t),\end{aligned}$$

where we used that φ is a homomorphism and that $t \mapsto \exp(tX)$ is a one-parameter subgroup.

Second, the derivative at 0 of σ is

$$\begin{aligned}\left.\frac{\mathrm{d}}{\mathrm{d}t}\sigma(t)\right|_{t=0} &= \left.\frac{\mathrm{d}}{\mathrm{d}t}\varphi(\exp(tX))\right|_{t=0} \\ &= (\,\mathrm{d}\varphi)_{\exp(0\cdot X)}\left.\frac{\mathrm{d}}{\mathrm{d}t}\exp(tX)\right|_{t=0} \\ &= (\,\mathrm{d}\varphi)_1 X_1.\end{aligned}$$

□

4.2.3 Exponential coordinates

Let $X_1, \dots, X_n$ be a basis of the Lie algebra of a Lie group G. The map $\alpha : \mathbb{R}^n \to G$,

$$\alpha(t_1, \dots, t_n) := \exp(t_1 X_1 + \cdots + t_n X_n)$$

is a diffeomorphism of a neighborhood of $0 \in \mathbb{R}^n$ with a neighborhood of 1_G in G, by Proposition 4.2.10. Such a map is called *exponential (local) coordinate map* (or *exponential local coordinates map of the first kind*) with respect to $X_1, \dots, X_n$.

The map $\beta : \mathbb{R}^n \to G$

$$\beta(t_1, \dots, t_n) := \exp(t_1 X_1) \cdots \exp(t_n X_n)$$

is called *exponential (local) coordinates map of the second kind* with respect to $X_1, \dots, X_n$.

One can consider intermediate examples. For example, given $k = 1, \dots, n-1$, one can let $\beta_k : \mathbb{R}^n \to G$ be the map

$$\beta_k(t_1, \dots, t_n) := \exp(t_1 X_1 + \cdots + t_k X_k)\exp(t_{k+1} X_{k+1} + \cdots + tt_n X_n),$$

which is called an *exponential (local) coordinates map of mixed kind* with respect to $X_1, \dots, X_n$.

Notice that β and β_k might depend on the ordering of the basis.

The maps β and β_k are indeed coordinate maps, since for them the differential at 0 is not singular. Indeed, for β we have

$$\begin{aligned}
(\mathrm{d}\beta)_0(\partial_j|_0) &\stackrel{\text{def}}{=} \frac{\delta}{\delta t_j}\beta(t_1,\ldots,t_n)\Big|_{(t_1,\ldots,t_n)=(0,\ldots,0)} \\
&= \frac{\mathrm{d}}{\mathrm{d}t_j}\beta(0,\ldots,0,t_j,0,\ldots,0)\Big|_{t_j=0} \\
&= \frac{\mathrm{d}}{\mathrm{d}t_j}\exp(t_jX_j)\Big|_{t_j=0} \\
&= X_j.
\end{aligned}$$

Warning: There are examples of groups for which α and β are not surjective.

4.3 General Linear Groups, its Lie algebra, and its exponential map

The General Linear Group, denoted as $\mathrm{GL}(n)$, consists of invertible $n\times n$ matrices over a given field. Its associated Lie algebra, denoted as $\mathfrak{gl}(n)$, consists of the set of all $n\times n$ matrices equipped with the commutator bracket operation. The exponential map, defined on the Lie algebra, provides a way to exponentiate matrices and obtain elements in the General Linear Group. It plays a crucial role in Lie theory and connects the algebraic structure of the Lie algebra with the geometric properties of the Lie group.

In our study, it is essential to work with finite-dimensional vector spaces that are not explicitly identified with $\mathbb{R}^n$. Consequently, we consider general linear groups over vector spaces, i.e., the set of its automorphisms. This abstraction enables us to consider structures like $\mathfrak{gl}(n)$ itself or, very importantly, the Lie algebra associated with a Lie group.

Throughout this chapter, all the vector spaces under consideration are defined over the field of real numbers. Similarly, the matrices we examine possess real coefficients. This choice facilitates the investigation of various phenomena within the framework of real analysis.

4.3.1 $\mathrm{GL}(V)$ and $\mathfrak{gl}(V)$

The n-th *general linear group* is

$$\mathrm{GL}(n,\mathbb{R}) := \{A:\ A \text{ is an } n\times n \text{ matrix with } \det A\neq 0\}.$$

This is a group when equipped with the row-column product of matrices. Slightly more generally, if V is a vector space, then

$$\mathrm{GL}(V) := \mathrm{Aut}(V) := \{A : \ A : V \to V \text{ is an invertible linear transformation}\}.$$

This is a group when equipped with the composition rule where the identity element is the identity transformation $\mathbb{I} : V \to V$. Because this product rule and the inversion rule are smooth (see Exercise **??**), then $\mathrm{GL}(n, \mathbb{R})$ and $\mathrm{GL}(V)$ are Lie groups, assuming that V is finite dimensional. Indeed, the Lie group $\mathrm{GL}(V)$ is Lie group isomorphic to $\mathrm{GL}(n, \mathbb{R})$ for $n := \dim(V)$.

For $n \in$, we define

$$\mathfrak{gl}(n, \mathbb{R}) := \mathrm{Mat}_{\mathrm{n}\times\mathrm{n}}(\mathbb{R}) := \{\text{all } n \times n \text{ matrices with real entries}\}.$$

If V is a vector space, then

$$\mathfrak{gl}(V) := \mathrm{End}(V) := \{\text{all linear transformations from } V \text{ to } V\ \}.$$

Clearly, we have $\mathrm{GL}(n, \mathbb{R}) = \mathrm{GL}(\mathbb{R}^n)$ and $\mathfrak{gl}(n, \mathbb{R}) = \mathfrak{gl}(\mathbb{R}^n)$.

For $A, B \in \mathfrak{gl}(n, \mathbb{R})$, with $n \in \mathbb{N}$, or, more generally, for $A, B \in \mathfrak{gl}(V)$ for a vector space V, we set

$$[A, B] := AB - BA.$$

Such an operation is a Lie bracket that makes $\mathfrak{gl}(V)$ into a Lie algebra. And, as the choice of name suggests, this Lie algebra is the Lie algebra of $\mathrm{GL}(V)$, see Proposition 4.3.5.

4.3.2 Matrix exponential

We next recall the matrix exponential: exponential of matrices. Since we shall consider linear endomorphisms of vector spaces, like for example of the Lie algebra of a Lie group, we define the matrix exponential on the space $\mathfrak{gl}(V)$.

Definition 4.3.1 (Matrix exponential)**.** Let V be a finite-dimensional vector space. For each $A \in \mathfrak{gl}(V)$, define *the matrix exponential of* A as

$$e^A := \mathbb{I} + A + \frac{1}{2}A^2 + \frac{1}{3!}A^3 + \cdots = \sum_{k=0}^{\infty} \frac{1}{k!}A^k. \tag{4.3.2}$$

In fact, the series giving e^A is absolutely converging, see Exercise 4.7.46. Consequently, the function $A \mapsto e^A$ is smooth (in fact, analytic). Moreover, each e^A is invertible with inverse e^{-A}, see

Exercise 4.7.47, so

$$e^A \in \mathrm{GL}(V).$$

It easy to see (c.f. Proposition 4.3.4) that for every $A \in \mathrm{Mat_{n\times n}}$ the curve $t \mapsto e^{tA}$ satisfies

$$e^{tA}\big|_{t=0} = \mathbb{I}, \qquad \frac{\mathrm{d}}{\mathrm{d}t} e^{tA}\bigg|_{t=0} = A.$$

Moreover, the map $\phi^t(B) := Be^{tA}$ satisfies the following properties:

- ϕ^t is a flow , i.e., $\phi^t \circ \phi^s = \phi^{t+s}$ because for every B we have that $(Be^{sA})e^{tA} = Be^{(t+s)A}$;
- ϕ^t is left invariant, i.e., $\phi^t(MB) = M\phi^t(B)$ because $(MB)e^{tA} = MBe^{tA}$.

Hence, this flow is the flow of its derivative at 0 :

$$\frac{\mathrm{d}}{\mathrm{d}t}\phi^t(B)\bigg|_{t=0} = \frac{\mathrm{d}}{\mathrm{d}t}Be^{tA}\bigg|_{t=0} = B\,\frac{\mathrm{d}}{\mathrm{d}t}e^{tA}\bigg|_{t=0} = BA.$$

We summarise, see also Proposition 4.3.4, the basic properties of the matrix exponential:

Proposition 4.3.3 (Matrix exponential)**.** *Let V be a finite-dimensional vector space.*

1. *The matrix exponential*
$$\exp : \mathfrak{gl}(V) \to \mathrm{GL}(V)$$
$$A \mapsto e^A,$$
is an analytic map.

2. *For every $A \in \mathfrak{gl}(V)$, the curve $t \mapsto e^{tA}$ is a one-parameter subgroup.*

3. *For every $A \in \mathfrak{gl}(V)$, the map*
$$GL(V) \to TGL(V)$$
$$B \mapsto BA,$$
defines a left-invariant vector field on $\mathrm{GL}(n)$ whose flow is $\mathbb{R} \times \mathrm{GL}(n) \to \mathrm{GL}(n)$ defined by $(t, B) \mapsto Be^{tA}$.

Rephrasing when $V = \mathbb{R}^n$, we have that for all $A \in \mathfrak{gl}(n) \simeq T_{\mathbb{I}}GL(n)$, the unique LIVF on $\mathrm{GL}(n)$ that equals A at $\mathbb{I}$ is

$$B \in \mathrm{GL}(n) \mapsto BA \in \mathrm{Mat_{n\times n}}(\mathbb{R}).$$

In the next proposition, we spell out the argument that shows what is the derivative of the OPS $t \mapsto e^{tA}$. We shall refer to this proposition sever times.

Proposition 4.3.4 (Derivative of e^{tA})**.** *For every finite-dimensional vector space V and every $A \in \mathrm{End}(V)$, the curve $t \mapsto e^{tA}$ is a one-parameter subgroup of* $\mathrm{GL}(V)$ *such that*

$$\frac{\mathrm{d}}{\mathrm{d}t} e^{tA} = A e^{tA}$$

and

$$\left.\frac{\mathrm{d}}{\mathrm{d}t} e^{tA}\right|_{t=0} = A.$$

Proof. Recall to notice that $A \mapsto e^A$ is smooth, that $e^{sA} \cdot e^{tA} = e^{(s+t)A}$, and that $e^0 = \mathbb{I}$. Therefore $t \mapsto e^{tA}$ is a one-parameter subgroup of $\mathrm{GL}(V)$. For the last two claims, we have

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}(e^{tA}) &= \frac{\mathrm{d}}{\mathrm{d}t}\left(\sum_{k=0}^{\infty} \frac{1}{k!}(tA)^k\right) \\
&= \sum_{k=0}^{\infty} \frac{1}{k!} \frac{\mathrm{d}}{\mathrm{d}t}(t^k A^k) \\
&= \sum_{k=1}^{\infty} \frac{1}{k!} k t^{k-1} A^k \\
&= A \sum_{k=1}^{\infty} \frac{1}{(k-1)!} t^{k-1} A^{k-1} \\
&= A e^{tA}.
\end{aligned}$$

□

4.3.3 Lie algebras of general linear groups

The key point of this section is to show that $\mathfrak{gl}(n) = \mathfrak{gl}(\mathbb{R}^n)$ is (isomorphic to) the Lie algebra of the Lie group $\mathrm{GL}(n) = \mathrm{GL}(\mathbb{R}^n)$.

Proposition 4.3.5. *The Lie algebra of* $\mathrm{GL}(V)$ *is isomorphic to Lie algebra* $\mathfrak{gl}(V)$*.*

Proof. The key point of the proof is to show that for every $A, B \in \mathfrak{gl}(V)$ the bracket between the vector fields

$$M \mapsto MA \qquad \text{and} \qquad M \mapsto MB$$

is $M \mapsto M(AB - BA)$.

Thus, in terms of flows we need to show that

$$\left.\frac{\mathrm{d}}{\mathrm{d}t} \phi^t_{AB-BA}(M)\right|_{t=0} = \left.\frac{\mathrm{d}}{\mathrm{d}t} (\phi_B^{-\sqrt{t}} \circ \phi_A^{-\sqrt{t}} \circ \phi_B^{\sqrt{t}} \circ \phi_A^{\sqrt{t}})(M)\right|_{t=0}. \tag{4.3.6}$$

We begin by considering left-hand side.

$$LHS := \frac{\mathrm{d}}{\mathrm{d}t} M \exp^{t(AB-BA)}\Big|_{t=0} = M(AB - BA).$$

On the other hand, recall that $\exp^{\sqrt{t}A} = \mathbb{I} + \sqrt{t}A + \frac{tA^2}{2} + o(t)$, the right-hand side becomes

$$\begin{aligned}
RHS := &\frac{\mathrm{d}}{\mathrm{d}t} M(\exp^{\sqrt{t}A} \circ \exp^{\sqrt{t}B} \circ \exp^{-\sqrt{t}A} \circ \exp^{\sqrt{t}B})\Big|_{t=0} \\
= &\frac{\mathrm{d}}{\mathrm{d}t} M(\exp^{\sqrt{t}A} \circ \exp^{\sqrt{t}B} \circ \exp^{-\sqrt{t}A} \circ \exp^{\sqrt{t}B})\Big|_{t=0} \\
= &\frac{\mathrm{d}}{\mathrm{d}t} M(\mathbb{I} + \sqrt{t}(A + B - A - B) \\
&+ t\left(\frac{A^2}{2} + \frac{B^2}{2} + \frac{A^2}{2} + \frac{B^2}{2} + AB - A^2 - AB - BA - B^2 + AB\right) + o(t))\Big|_{t=0} \\
= &M(AB - BA).
\end{aligned}$$

Hence (4.3.6) holds, as desired. □

Proposition 4.3.4, together with Proposition 4.3.5, therefore clarified that the exponential of $\mathrm{GL}(n, \mathbb{R})$ is the usual exponential of matrices $\exp : A \in \mathfrak{gl}(n, \mathbb{R}) \mapsto e^A \in \mathrm{GL}(n, \mathbb{R})$.

Corollary 4.3.7 (of Proposition 4.3.4)**.** *For every finite-dimensional vector space V, the exponential of the Lie group $\mathrm{GL}(V)$ is the matrix exponential $\exp : \mathfrak{gl}(V) \to \mathrm{GL}(V)$, $A \mapsto e^A$.*

4.4 Adjoint representation

The adjoint representation, also known as the adjoint action, of a Lie group G provides a means of representing the elements of the group as linear transformations of its Lie algebra, viewed as a vector space. Specifically, in the case of the general linear group $\mathrm{GL}(n, \mathbb{R})$, where the operations are linear, the adjoint representation corresponds to conjugation. To obtain the adjoint representation for a Lie group, we linearize (i.e., take the differential of) the group's action on itself through conjugation. This natural representation captures the way the elements of the Lie group act on its Lie algebra. It establishes a fundamental link between the group's abstract structure and the associated linear transformations, facilitating the exploration of geometric and algebraic properties within the realm of Lie theory.

4.4.1 Ad and ad

In this section we shall consider a Lie algebra $\mathfrak{g}$ as a vector space and then consider the spaces $\mathfrak{gl}(\mathfrak{g})$ and $\mathrm{GL}(\mathfrak{g})$, as in Section 4.3.1.

Definition 4.4.1 (Adjoint map)**.** Let $\mathfrak{g}$ be a Lie algebra. The *adjoint map* of $\mathfrak{g}$ is the linear map

$$\mathrm{ad} : \mathfrak{g} \to \mathfrak{gl}(\mathfrak{g})$$

given by

$$\mathrm{ad}(X)(Y) := \mathrm{ad}_X(Y) := [X, Y], \qquad \forall X, Y \in \mathfrak{g}.$$

Remark 4.4.2. The map $\mathrm{ad}_X : \mathfrak{g} \to \mathfrak{g}$ is indeed in $\mathfrak{gl}(\mathfrak{g})$, i.e., it is linear, not necessarily invertible. Moreover, seen $\mathfrak{gl}(\mathfrak{g})$ as a Lie algebra, ad is a Lie algebra homomorphism: for all $X, Y \in \mathfrak{g}$ and all $s, t \in \mathbb{R}$

4.4.2.i. $\mathrm{ad}(sX + tY) = s\ \mathrm{ad}(X) + t\ \mathrm{ad}(Y)$,

4.4.2.ii. $\mathrm{ad}([X, Y]) = [\mathrm{ad}(X), \mathrm{ad}(Y)]$,

cf. Exercise 4.7.38.

Definition 4.4.3 (Adjoint representation)**.** Let G be a Lie group with Lie algebra $\mathfrak{g}$. For $g \in G$ define

$$\mathrm{Ad}(g) := \mathrm{Ad}_g := (\,\mathrm{d}C_g)_{1_G},$$

i.e., $\mathrm{Ad}(g)$ is the differential at the identity of the conjugation $C_g : h \mapsto ghg^{-1}$. The map

$$\mathrm{Ad} : G \to \mathrm{GL}(\mathfrak{g})$$

is called *adjoint representation*.

Remark 4.4.4. The map Ad is indeed a representation, i.e., Ad is a group homomorphism into $\mathrm{GL}(\mathfrak{g})$:

$$\mathrm{Ad}(gh) = \mathrm{Ad}(g) \circ \mathrm{Ad}(h), \qquad \forall g, h \in G,$$

cf. Exercise 4.7.39.

4.4.2 Properties and formulas

Proposition 4.4.5. *Let G be a Lie group with Lie algebra $\mathfrak{g}$. The adjoint representation* $\mathrm{Ad} : G \to \mathrm{GL}(\mathfrak{g})$ *is a Lie group homomorphism and the Lie algebra homomorphism associated to* Ad *is the adjoint map* $\mathrm{ad} : \mathfrak{g} \to \mathfrak{gl}(\mathfrak{g})$*, i.e.,*

$$(\mathrm{Ad})_* = \mathrm{ad},$$

which is

$$(\,\mathrm{d}\,\mathrm{Ad})_{1_G}(X) = \mathrm{ad}(X), \qquad \forall X \in \mathfrak{g}.$$

Proof. Since $t \mapsto \exp(tX)$ is a curve in G that is tangent to X at 1_G, we have

$$(\mathrm{d}\,\mathrm{Ad})_{1_G}(X)(Y) = \left.\frac{\mathrm{d}}{\mathrm{d}t}\,\mathrm{Ad}\left(\exp(tX)\right)(Y)\right|_{t=0}.$$

Here X, Y are element in $T_{1_G}G$.

We denote by $\tilde{X}, \tilde{Y}$ the left-invariant vector fields such that $\tilde{X}_{1_G} = X$ and $\tilde{Y}_{1_G} = Y$. We have

$$\begin{aligned}
\mathrm{Ad}(\exp(tX))(Y) &= \left(\mathrm{d}C_{\exp(tX)}\right)_{1_G}(Y) \\
&= \left(\mathrm{d}R_{\exp(-tX)}\right)_{\exp(tX)}\left(\mathrm{d}L_{\exp(tX)}\right)_{1_G}(Y) \\
&= \left(\mathrm{d}R_{\exp(-tX)}\right)_{\exp(tX)}\left(\tilde{Y}_{\Phi^t_{\tilde{X}}(1_G)}\right) \\
&= \left(\mathrm{d}\Phi^{-t}_{\tilde{X}}\right)_{\Phi^t_{\tilde{X}}(1_G)}\tilde{Y}_{\Phi^t_{\tilde{X}}(1_G)},
\end{aligned}$$

where we used that the flow at time t of the left-invariant vector field $\tilde{X}$ is the right translation by $\exp(tX)$, see Proposition 4.2.7. We get

$$\begin{aligned}
(\mathrm{d}\,\mathrm{Ad})_{1_G}(X)(Y) &= \left.\frac{\mathrm{d}}{\mathrm{d}t}\left(\mathrm{d}\Phi^{-t}_{\tilde{X}}\right)_{\Phi^t_{\tilde{X}}(1_G)}\tilde{Y}_{\Phi^t_{\tilde{X}}(1_G)}\right|_{t=0} \\
&\underset{\text{Lie deriv.}}{\overset{\text{def. of}}{=}} \left(\mathrm{Lie}_{\tilde{X}}(\tilde{Y})\right)_{1_G} \\
&= [\tilde{X}, \tilde{Y}]_{1_G} \\
&= [X, Y] \\
&= \mathrm{ad}_X(Y).
\end{aligned}$$

$\square$

Recall that, if φ is a Lie group homomorphism and φ_* is the Lie algebra homomorphism induced by φ, by Proposition 4.2.11 we have the following first commutative diagram and for $\varphi = C_g$ (resp. $\varphi = \mathrm{Ad}$) we have the following second (resp. third) commutative diagram.

$$\begin{tikzcd}
\mathfrak{g} \arrow[r, "\varphi_*"] \arrow[d, "\exp"'] & \mathfrak{h} \arrow[d, "\exp"] \\
G \arrow[r, "\varphi"'] & H
\end{tikzcd}
\qquad
\begin{tikzcd}
\mathfrak{g} \arrow[r, "(C_g)_* = \mathrm{Ad}_g"] \arrow[d, "\exp"'] & \mathfrak{g} \arrow[d, "\exp"] \\
G \arrow[r, "C_g"'] & G
\end{tikzcd}
\qquad
\begin{tikzcd}
\mathfrak{g} \arrow[r, "(\mathrm{Ad})_* = \mathrm{ad}"] \arrow[d, "\exp"'] & \mathfrak{gl}(\mathfrak{g}) \arrow[d, "\exp = e^{\cdot}"] \\
G \arrow[r, "\mathrm{Ad}"'] & \mathrm{GL}(\mathfrak{g})
\end{tikzcd}$$

Formula 4.4.6. Since $\mathrm{Ad}_g = (C_g)_*$ by definition, we have

$$C_g(\exp(X)) = \exp(\mathrm{Ad}_g X), \qquad \forall X \in \mathfrak{g}, \forall g \in G.$$

Equivalently,

$$\exp(Y)(\exp(X)\exp(-Y) = \exp(\mathrm{Ad}_g X), \qquad \forall X, Y \in \mathfrak{g}.$$

Formula 4.4.7. Since $(\mathrm{Ad})_* = \mathrm{ad}$ by the previous proposition, we have

$$\mathrm{Ad}_{\exp(X)} = e^{\mathrm{ad}_X}.$$

In the above formula, ad_X is a linear transformation on $\mathfrak{g}$, i.e., an element of $\mathfrak{gl}(\mathfrak{g})$, which is the Lie algebra of $\mathrm{GL}(\mathfrak{g})$. We saw that $\exp : \mathfrak{gl}(\mathfrak{g}) \to \mathrm{GL}(\mathfrak{g})$ is given by the classical matrix exponential. Therefore

$$\begin{aligned} e^{\mathrm{ad}_X}(Y) &= \sum_{k=0}^{\infty} \frac{(\mathrm{ad}_X)^k}{k!}(Y) \\ &= Y + [X,Y] + \frac{1}{2}[X,[X,Y]] + \frac{1}{3!}[X,[X,[X,Y]]] + \ldots \end{aligned}$$

Formula 4.4.8. Let V be a vector space. For all $X, Y \in \mathfrak{gl}(V)$ and $B \in \mathrm{GL}(V)$ we have

4.4.8.i. $\mathrm{Ad}_B(X) = B \cdot X \cdot B^{-1}$.

4.4.8.ii. $e^{\mathrm{ad}_X} Y = e^X Y e^{-X}$.

4.4.8.iii. $e^{BXB^{-1}} = Be^X B^{-1}$.

cf. Exercises 4.7.41, 4.7.42, and 4.7.43.

Formula 4.4.9. Let X, Y left-invariant vector fields on a Lie group G. For all $t \in \mathbb{R}$

$$(\Phi_X^t)_* Y = e^{-\mathrm{ad}(tX)} Y,$$

cf. Exercise 4.7.44

4.5 Semi-direct products

In this section, our focus turns to the study of the semidirect product of Lie groups. We begin by considering group actions by group automorphisms, as well as Lie actions by derivations. These actions enable us to form semidirect products of Lie groups and, in a parallel manner, semidirect products of Lie algebras. A key result will emerges: the Lie algebra of a semidirect product of Lie groups corresponds to a semidirect product of the Lie algebras associated with the constituent groups.

4.5.1 Derivations and actions by automorphisms

Definition 4.5.1 (Group action by automorphisms)**.** Let H, G be groups. A *(group) action* of H *by automorphisms* of G is a group homomorphism $\theta : H \to \mathrm{Aut}(G)$. Equivalently, it is a map $\theta : H \times G \to G$ such that, for $\theta_h := \theta(h, \cdot)$, we have

$$\theta_h(g_1 g_2) = \theta_h(g_1)\theta_h(g_2), \qquad \forall g_1, g_2 \in G, \forall h \in H$$

and

$$\theta_{h_1 h_2} = \theta_{h_1} \circ \theta_{h_2}, \qquad \forall h_1, h_2 \in H. \tag{4.5.2}$$

In case G and H are Lie groups, we say that an action $\theta : H \to \mathrm{Aut}(G)$ is *smooth* if it is smooth as a map $\theta : H \times G \to G$.

Recall that every element $\varphi \in \mathrm{Aut}(G)$ has an associated Lie algebra automorphism φ_*. Let us introduce a notation for the *space of Lie algebra automorphisms* of a Lie algebra $\mathfrak{g}$:

$$\mathrm{Aut}_{\mathrm{Lie}}(\mathfrak{g}) := \{T \in \mathrm{GL}(\mathfrak{g}) \ : \ T[u, v] = [Tu, Tv], \forall u, v \in \mathfrak{g}\}.$$

We shall stress that $\mathrm{Aut}_{\mathrm{Lie}}(\mathfrak{g})$ is a closed Lie subgroup of $\mathrm{GL}(\mathfrak{g})$ with a Lie algebra that is a Lie subalgebra of $\mathfrak{gl}(\mathfrak{g})$. The elements of this Lie subalgebra are the, so called, derivations:

Definition 4.5.3 (Derivation on a Lie algebra)**.** Let $\mathfrak{g}$ be a Lie algebra. A *derivation* on $\mathfrak{g}$ is a linear map $D : \mathfrak{g} \to \mathfrak{g}$ that satisfies Leibniz's law:

$$D([X, Y]) = [D(X), Y] + [X, D(Y)], \qquad \forall X, Y \in \mathfrak{g}. \tag{4.5.4}$$

Let $\mathrm{Der}(\mathfrak{g})$ be the set of derivation on $\mathfrak{g}$.

We state the relative results for reference (cf. Exercises 4.7.54–4.7.59):

Proposition 4.5.5. *Let G be a Lie group with Lie algebra $\mathfrak{g}$.*

4.5.5.i. *The natural map*

$$\begin{aligned} \mathrm{Aut}(G) &\to \mathrm{Aut}_{\mathrm{Lie}}(\mathfrak{g}) \\ \varphi &\mapsto \varphi_* \end{aligned}$$

is an injective Lie group homomorphism and, if G is simply connected, it is a Lie group isomorphism.

4.5.5.ii. $\mathrm{Aut}_{\mathrm{Lie}}(\mathfrak{g})$ *is a regular Lie subgroup of* $\mathrm{GL}(\mathfrak{g})$ *with*

$$\mathrm{Lie}(\mathrm{Aut}_{\mathrm{Lie}}(\mathfrak{g}) = \mathrm{Der}(\mathfrak{g}).$$

4.5.5.iii. *The adjoint map* ad *and the adjoint representation* Ad *satisfy*

$$\begin{tikzcd} \mathfrak{g} \arrow[r, "\mathrm{ad}"] \arrow[d, "\exp"'] & \mathrm{Der}(\mathfrak{g}) \subset \mathfrak{gl}(\mathfrak{g}) \arrow[d, "\exp"] \\ G \arrow[r, "\mathrm{Ad}"] & \mathrm{Aut}_{\mathrm{Lie}}(\mathfrak{g}) \subset \mathrm{GL}(\mathfrak{g}) \end{tikzcd}$$

4.5.2 Semi-direct products of Lie algebras and groups

Definition 4.5.6 (Semi-direct product of Lie algebras)**.** Let $\mathfrak{g}$ and $\mathfrak{h}$ be Lie algebras, and let $\sigma : \mathfrak{h} \to \mathrm{Der}(\mathfrak{g})$ be a Lie algebra homomorphism into the space of derivations of $\mathfrak{g}$. On the direct sum $\mathfrak{g} \oplus \mathfrak{h}$ we consider the bracket that agrees with the brackets of $\mathfrak{g}$ and $\mathfrak{h}$, and additionally

$$[(0,Y),(X,0)] := \sigma(Y)(X), \qquad \forall X \in \mathfrak{g}, \forall Y \in \mathfrak{h}.$$

More explicitly,

$$[(X,Y),(X',Y')] := ([X,X'] + \sigma(Y)(X') - \sigma(Y')(X), [Y,Y']), \qquad \forall X, X' \in \mathfrak{g}, \forall Y, Y' \in \mathfrak{h}$$

The resulting Lie algebra is the *semidirect product* of $\mathfrak{g}$ and $\mathfrak{h}$ with respect to σ, and it is denoted by $\mathfrak{g} \rtimes_\sigma \mathfrak{h}$. When σ is understood, or there is no need of naming it, we simply write $\mathfrak{g} \rtimes \mathfrak{h}$.

Remark 4.5.7. We have the following properties for semi-direct product of Lie algebras.

4.5.7.i. We have that $\mathfrak{g} \rtimes \mathfrak{h}$ is a Lie algebra and the maps $X \in \mathfrak{g} \mapsto (X,0)$ and $Y \in \mathfrak{h} \mapsto (0,Y)$ give injective Lie algebra homomorphisms into $\mathfrak{g} \rtimes \mathfrak{h}$.

4.5.7.ii. If $\sigma \equiv 0$, we call $\mathfrak{g} \rtimes \mathfrak{h}$ the *direct product* of $\mathfrak{g}$ and $\mathfrak{h}$, and write it as $\mathfrak{g} \times \mathfrak{h}$.

4.5.7.iii. In $\mathfrak{g} \rtimes \mathfrak{h}$, the Lie subalgebra $\mathfrak{g}$ is an *ideal*, i.e., $[\mathfrak{g} \rtimes \mathfrak{h}, \mathfrak{g}] \subseteq \mathfrak{g}$. This is the reason for the choice of the symbol $\rtimes$ to resemble $\lhd$ and we write $\mathfrak{g} \lhd \mathfrak{g} \rtimes \mathfrak{h}$. If somewhere else you read $\mathfrak{g} \ltimes \mathfrak{h}$, then it means that in that setting it is $\mathfrak{h}$ that is an ideal and hence it is $\mathfrak{g}$ that is acting on $\mathfrak{h}$ by derivations.

4.5.7.iv. The map σ represent the adjoint map in $\mathfrak{g} \rtimes_\sigma \mathfrak{h}$ of $\mathfrak{h}$ on $\mathfrak{g}$:

$$\mathrm{ad}_Y(X) = \sigma(Y)(X), \qquad \forall X \in \mathfrak{g}, \forall Y \in \mathfrak{h};$$

recall that indeed every ad_Y is a derivation.

Definition 4.5.8 (Semi-direct product of groups)**.** Let G and H be groups, and $\theta : H \to \mathrm{Aut}(G)$ an action of H by automorphisms of G. On the set $\{(g,h) : g \in G,\ h \in H\}$ we put the product

$$(g_1,h_1)\cdot(g_2,h_2) = (g_1 \cdot \theta_{h_1}(g_2), h_1h_2), \qquad \forall g_1,g_2 \in G, \forall h_1,h_2 \in H. \tag{4.5.9}$$

The resulting group is the *semi-direct product* of G and H with respect to θ, and it is denoted by $G \rtimes_\theta H$, or simply $G \rtimes H$ if there is no need to explicitely write θ.

Remark 4.5.10. Similarly to Remark 4.5.7, we have the following properties for semi-direct product of groups.

4.5.10.i. We have that $G \rtimes H$ is a group and the maps $g \in G \mapsto (g, 1_H)$ and $h \in \mathfrak{h} \mapsto (1_G, h)$ give injective group homomorphisms into $G \rtimes H$.

4.5.10.ii. If $\sigma \equiv \mathrm{id}_G$, we call $G \rtimes H$ the *direct product* of G and H, and write it as $G \times H$.

4.5.10.iii. In $G \rtimes H$, the subgroup G is normal subgroup. Thus $G \lhd G \rtimes H$, which explain the symbol.

4.5.10.iv. To memorize the definition of the product law one need to understand its reason[2]. Write the product of elements $g_1, g_2 \in G$ and $h_1, h_2 \in H$ as

$$g_1h_1g_2h_2 = g_1h_1g_2h_1^{-1}h_1h_2 = g_1C_{h_1}(g_2)h_1h_2.$$

In other words The map θ represents the conjugation in $G \rtimes H$ of H on G:

$$C_h(g) = \theta_h(g), \qquad \forall g \in G, \forall h \in H; \tag{4.5.11}$$

recall that indeed every C_g is a group automorphism.

4.5.10.v. The element (g,h) has inverse $(\theta_{h^{-1}}(g^{-1}), h^{-1})$.

4.5.3 Lie algebras of semi-direct products of Lie groups

Proposition 4.5.12. *Let G and H be Lie groups with Lie algebras $\mathfrak{g}$ and $\mathfrak{h}$, respectively, and $\theta : H \to \mathrm{Aut}(G)$ be a smooth action.*

4.5.12.i. *$G \rtimes_\theta H$ is a Lie group.*

[2] A Finnish motto say 'what you understand, you don't need to remember'

4.5.12.ii. *The map $\tau : H \to \mathrm{Aut}_{\mathrm{Lie}}(\mathfrak{g})$ defined by $\tau_h := (\theta_h)_*$, for $h \in H$, is a Lie group homomorphism.*

4.5.12.iii. *For $\sigma := \tau_* : \mathfrak{h} \to \mathrm{Der}(\mathfrak{g})$, for the above τ, we have*

$$\mathrm{Lie}(G \rtimes_\theta H) = \mathfrak{g} \rtimes_\sigma \mathfrak{h}.$$

Proof. As a manifold, $G \rtimes_\theta H$ is the product of the manifolds G and H. And moreover, the group structure is smooth by construction. Thus $G \rtimes_\theta H$ is a Lie group.

By the chain rule applied to (4.5.2) we get $\tau_{h_1 h_2} = \tau_{h_1} \circ \tau_{h_2}$, for all $h_1, h_2 \in H$. So also 4.5.12.ii is proved.

For proving 4.5.12.iii we need to calculate the Lie bracket $\mathrm{ad}_Y(X) = [Y, X]$ where $X \in \mathfrak{g}$ and $Y \in \mathfrak{h}$. Hence, before calculating $\mathrm{ad}_Y(X)$ we calculate $\mathrm{Ad}_{\exp(Y)}(X)$ and before that $C_{\exp(Y)}(\exp(X))$. For doing the calculation we shall crucially use the relation between the exponential map and the induce morphisms, see Proposition 4.2.11. In fact, we have

$$\theta_h(\exp(X)) = \exp(\tau_h(X)), \qquad \forall X \in \mathfrak{g}, \forall h \in H \tag{4.5.13}$$

and

$$\tau(\exp(Y)) = \exp(\sigma(Y)),, \qquad \forall Y \in \mathfrak{h}. \tag{4.5.14}$$

For all $X \in \mathfrak{g}$, $Y \in \mathfrak{h}$, and $t \in \mathbb{R}$, we have

$$\begin{aligned}
\exp(t\,\mathrm{Ad}_{\exp(Y)}(X)) &= \exp(\mathrm{Ad}_{\exp(Y)}(tX)) \\
&\overset{4.4.6}{=} C_{\exp(Y)}(\exp(tX)) \\
&\overset{(4.5.11)}{=} \theta_{\exp(Y)}(\exp(tX)) \\
&\overset{(4.5.13)}{=} \exp(\tau_{\exp(Y)}(tX)) \\
&= \exp(t\tau_{\exp(Y)}(X)),
\end{aligned}$$

where we also used that both $\mathrm{Ad}_{\exp(Y)}$ and $\tau_{\exp(Y)}$ are linear. We got an identity between OPS. Therefore $\mathrm{Ad}_{\exp(Y)}(X) = \tau_{\exp(Y)}(X))$, for all $X \in \mathfrak{g}$, i.e., $\mathrm{Ad}_{\exp(Y)} = \tau_{\exp(Y)}$. Consequently, $e^{\mathrm{ad}_Y} \overset{(4.4.7)}{=} \mathrm{Ad}_{\exp(Y)} = \tau_{\exp(Y)} \overset{(4.5.14)}{=} e^{\sigma(Y)}$. Differentiating in Y we get $\mathrm{ad}_Y = \sigma(Y)$. □

Remark 4.5.15. Conversely, every semidirect product of Lie algebras is the Lie algebra of a semidirect product of Lie groups. Indeed, let $\mathfrak{g} \rtimes_\sigma \mathfrak{h}$ be a semidirect product of Lie algebras, and let G and H be simply connected Lie groups with Lie algebras $\mathfrak{g}$ and $\mathfrak{h}$, respectively, cf. Theorem 4.1.10. From

Theorem 4.1.8, since H is simply connected, there is a Lie group homomorphism $\tau : H \to \mathrm{Aut}_{\mathrm{Lie}}(\mathfrak{g})$ such that $\tau_* = \sigma$. Then, again from Theorem 4.1.8, since G is simply connected, for every $h \in H$ there is a Lie group automorphism $\theta_h : G \to G$ such that $(\theta_h)_* = \tau_h$. Such a map induces a smooth action $\theta : H \to \mathrm{Aut}(G)$. One can verify that $\mathrm{Lie}(G \rtimes_\theta H) = \mathfrak{g} \rtimes_\sigma \mathfrak{h}$.

4.6 From algebras to groups

In this section, we revisit the discussion from Section 4.1 regarding the relationship between objects at the level of the Lie algebra and their counterparts at the level of the Lie group. Two key examples illustrate this relationship: the correspondence between Lie subalgebras and Lie subgroups, and the induction of Lie algebra homomorphisms from Lie group morphisms, provided that the Lie group in the source is simply connected.

4.6.1 Existence of subgroups

The next result shows the existence of Lie subgroups with given Lie subalgebra of a Lie algebra of a Lie group. Together with Ado's theorem (see Section **??**), we will deduce that for every abstract Lie algebra (real and finite-dimensional) there exists at least one Lie group with this Lie algebra.

Theorem 4.6.1 (Existence of subgroups)**.** *Let G be a Lie group. For every subalgebra $\mathfrak{h} \subset T_{1_G}G$, there is a unique connected Lie subgroup H with Lie algebra $\mathfrak{h}$; in fact, H is the group generated by* $\exp(\mathfrak{h})$*.*

Remark 4.6.2. In general, it may not be true that $H = \exp(\mathfrak{h})$.

Proof. This is a consequence of Frobenius's theorem. We consider the subbundle $\Delta \subset TG$ defined by

$$\Delta_g := (\mathrm{d}L_g)_1(\mathfrak{h}).$$

Notice that Δ is left-invariant and involutive (since $\mathfrak{h}$ is closed under the bracket). Frobenius's theorem implies that there exists a maximal connected submanifold H of G such that $1_G \in H$ and $TH = \Delta|_H$. By construction, $T_1H = \mathfrak{h}$. We claim that since Δ is invariant under the maps $\{L_h\}_{h \in G}$, then H is a subgroup. Indeed, take $h_1, h_2 \in H$ and observe that $L_{h_1^{-1}}H$ contains $1 = 1_G = 1_H$ and is tangent to Δ. By maximality, $h_1^{-1}h_2 \in L_{h_1^{-1}}H \subseteq H$.

Regarding uniqueness, if $\hat{H}$ is a connected subgroup with $\mathrm{Lie}(\hat{\mathrm{H}}) = \mathfrak{h}$, since $\exp(\mathfrak{h})$ is an open neighborhood of e in $\hat{H}$, we have

$$\hat{H} = (\hat{H})^\circ = \langle \exp(\mathfrak{h}) \rangle,$$

where we used Proposition 4.7.3. □

4.6.2 Existence of group homomorphisms

We shall show that every Lie algebra homomorphism between Lie algebras of two Lie groups is induced by a Lie group homomorphism in case the source Lie group is simply connected. Moreover, this group homomorphism is unique. The existence fails in the case where the group is not simply connected. The uniqueness fails as long as the group is not connected.

The necessary requirements of simple connection are given by this following exercise. More material on covering maps is discussed in Chapter **??**.

Exercise 4.6.3. Let G, H be connected Lie groups, and $\varphi : G \to H$ a Lie group homomorphism. Show that the following are equivalent:

i). φ is surjective and has discrete kernel;

ii). φ is a covering map;

iii). φ_* is an isomorphism of Lie algebras;

iv). φ is a local diffeomorphism.

Hint. The proof can be found in the book by Abate-Tovena (page 182, Proposizione 3.8.2). ■

Theorem 4.6.4 (Induced Lie group homomorphism)**.** *Let G, H be Lie groups. Assume G simply connected. For all Lie algebra homomorphism $\psi : \mathrm{Lie(G)} \to \mathrm{Lie(H)}$, there exists a unique Lie group homomorphism $\varphi : G \to H$ with $\varphi_* = \psi$.*

Proof. Let $\mathrm{Lie(G)} = \mathfrak{g}$ and $\mathrm{Lie(H)} = \mathfrak{h}$. Since ψ is a homomorphism, its graph

$$\mathfrak{k} = \{(X, \psi(X)) : X \in \mathfrak{g}\} \subset \mathfrak{g} \times \mathfrak{h}$$

is a subalgebra of $\mathfrak{g} \times \mathfrak{h} = \mathrm{Lie(G \times H)}$:

$$[(X, \psi(X)), (Y, \psi(Y))] \quad = \quad ([X, Y], [\psi(X), \psi(Y)]) \quad = \quad ([X, Y], \psi[X, Y])\,.$$

By Theorem 4.6.1, there is a unique connected Lie subgroup $K \subset G \times H$ with Lie(K) $= \mathfrak{k}$. Let $\pi_1 : G \times H \to G$ and $\pi_2 : G \times H \to H$ be the projections, which are Lie group homomorphisms. Let

$$\phi := \pi_1|_K : K \to G.$$

We have that

$$(\mathrm{d}\phi)_{(1_G,1_H)}(X, \psi(X)) = X, \quad \forall X \in \mathfrak{g}. \tag{4.6.5}$$

In particular, we get that $(\mathrm{d}\phi)_{(1_G,1_H)} : \mathfrak{k} \to \mathfrak{g}$ is injective, and therefore it is an isomorphism (since $\dim \mathfrak{k} = \dim \mathfrak{g}$). By Exercise 4.6.3, we deduce that $\phi : K \to G$ is a covering map. Since G is simply connected, ϕ is an isomorphism thanks to Corollary **??**. Set $\varphi := \pi_2|_K \circ \phi^{-1} : G \to H$, which is a Lie group homomorphism. From (4.6.5) we also get that

$$(\mathrm{d}\varphi)_{1_G}(X) = (\mathrm{d}\pi_2)_{(1_G,1_H)} \circ (\mathrm{d}\phi^{-1})_{1_G}(X) = (\mathrm{d}\pi_2)_{(1_G,1_H)}(X, \psi(X)) = \psi(X),$$

that is $\varphi_* = \psi$.

Regarding the uniqueness, if $\tilde{\varphi}$ is another homomorphism such that $(\tilde{\varphi})_* = \psi$, we get that $\tilde{\varphi} \circ \exp = \exp \circ \psi = \varphi \circ \exp$. Since exp is invertible in a neighborhood U of the identity element, we have $\varphi|_U = \tilde{\varphi}|_U$. Since such a U generates G and since $\varphi, \tilde{\varphi}$ are group homomorphisms, we get that $\tilde{\varphi} = \varphi$. □

4.7 Exercises

Here are some more or less easy exercise on Lie groups, with some of their solutions.

Exercise 4.7.1. For all elements g, h in a group G we have

(i). $L_h \circ L_g = L_{hg}$,

(ii). $R_h \circ R_g = R_{gh}$,

(iii). $L_h \circ R_g = R_g \circ L_h$,

(iv). $(L_g)^{-1} = L_{g^{-1}}$,

(v). $(R_g)^{-1} = R_{g^{-1}}$,

(vi). $C_{gh} = C_g \circ C_h$.

■

For the next two exercises, for a subset U of a group and an integer $n \in \mathbb{N}$, set

$$U^n := \{g_1 \cdot \dots \cdot g_n : g_1, \dots, g_n \in U\}.$$

Exercise 4.7.2. Let G be a Lie group (or more generally a topological group). If $U \subset G$ is open, then U^2 is open. ∎

Exercise 4.7.3. Connected groups are generated by neighborhoods of the identity: Let G be a connected Lie group (or more generally a topological group) and $U \subset G$ an open subset with $1 \in U$. Then $G = \bigcup_{n=0}^{\infty} U^n$. In other words, G is the smallest group containing U.

Solution. Let $U^{-1} := \{g^{-1} : g \in U\}$ and $V := U \cap U^{-1}$. Then V is open, $V^{-1} = V$, $e \in V$. Let $H := \bigcup_{n=1}^{\infty} V^n \subset \bigcup_{n=1} U^n$. Observe that H contains V and is a union of the open sets V^n (see the Exercise 4.7.2). Moreover, H is closed under multiplication and inversion, since $V^n \cdot V^m \subset V^{n+m}$ and $V^{-n} \subset V^n$. In other words, H is an open subgroup of G.

Note that gH is open for all $g \in G$, so $\bigcup_{g \notin H} gH$ is an open set.

Since G is connected, $G = H \sqcup \bigcup_{g \notin H} gH$ and $H \neq \emptyset$, we conclude that $G = H$. ∎

Exercise 4.7.4. Let G be a Lie group. Show that

(i) if H is a subgroup of G that is (topologically) open, then it is closed;

(ii) every neighborhood $U \subseteq G$ of the identity element generates G°, i.e., every element in the identity component G° is the product of finitely many elements in U;

(iii) if H is a subgroup of G that has nonempty interior, then it is open and closed. ∎

Exercise 4.7.5. Argue that on a topological groups right translations and left translations are homeomorphisms. While in a Lie group, they are smooth diffeomorphisms. ∎

Exercise 4.7.6. Show that anti-commutativity and Jacobi identity imply that the structural constants c_{ij}^k of a Lie algebra satisfy:

$$\begin{aligned} c_{ij}^k + c_{ji}^k = 0, \qquad \forall i,j,k \in \{1,\dots,n\}; \\ \sum_{r=1}^n \left(c_{ij}^r c_{rk}^s + c_{jk}^r c_{ri}^s + c_{ki}^r c_{rj}^s\right) = 0, \quad \forall i,j,k,s \in \{1,\dots,n\}. \end{aligned} \tag{4.7.7}$$

∎

Exercise 4.7.8. Let $c_{ij}^k \in \mathbb{R}$ satisfying (4.7.7). Define $[\cdot,\cdot]$ by (4.1.4). Then $[\cdot,\cdot]$ uniquely extends into a Lie bracket turning $\operatorname{span}\{X_1,\dots,X_n\}$ into a Lie algebra. ∎

Exercise 4.7.9. Let $\mathfrak{g}$ and $\tilde{\mathfrak{g}}$ be Lie algebras of dimension n. Let c_{ij}^k be the structural constants of $\mathfrak{g}$ with respect to a basis $X_1, \ldots, X_n$, and let $\tilde{c}_{ij}^k$ be the structural constants of $\tilde{\mathfrak{g}}$ with respect to a basis $\tilde{X}_1, \ldots, \tilde{X}_n$. Show that if $c_{ij}^k = \tilde{c}_{ij}^k$, then the linear map $\psi : \mathfrak{g} \to \tilde{\mathfrak{g}}$ defined by $\psi(X_i) := \tilde{X}_i$ satisfies

$$\psi[X,Y] = [\psi X, \psi Y], \qquad \forall X, Y \in \mathfrak{g}. \tag{4.7.10}$$

Solution. For all $X, Y \in \mathfrak{g}$, write $X = \sum_{i=1}^n a_i X_i$, $Y = \sum_{j=1}^n b_j X_j$. Then

$$[\psi(X), \psi(Y)] = \sum_{ij} [a_i \psi(X_i), b_j \psi(X_j)] = \sum_{ij} a_i b_j [\tilde{X}_i, \tilde{X}_j] = \sum_{ijk} a_i b_j \tilde{c}_{ij}^k \tilde{X}_k$$

and

$$\psi[X,Y] = \sum_{ij} \psi[a_i X_i, b_j X_j] = \sum_{ij} a_i b_j \psi(\sum_k c_{ij}^k X_k) = \sum_{ijk} a_i b_j c_{ij}^k \tilde{X}_k.$$

Thus (4.7.10) holds. ■

Exercise 4.7.11. The space of LIVFs is closed under Lie bracket. In other words, the Lie bracket of two left-invariant vector fields is left invariant. ■

Proof. Let X, Y be left-invariant vector fields on a Lie group G and let $g \in G$. Then

$$(L_g)_*[X,Y] = [(L_g)_*X, (L_g)_*Y] = [X,Y]. \qquad \square$$

Exercise 4.7.12 (Right translation of LIVF)**.** Let X be a left invariant vector field on a Lie group G. Let R_g be the right translation by an element $g \in G$. Prove that $(R_g)_*X$ is a left-invariant vector field.

Solution. Let $h \in G$. Then, using Exercise 4.7.1.iii and that X is left invariant, we have

$$\begin{aligned}
\mathrm{d}L_h \circ ((R_g)_*X) &= \mathrm{d}L_h \circ \mathrm{d}R_g \circ X \circ R_g^{-1} \\
&= \mathrm{d}(L_h \circ R_g) \circ X \circ R_g^{-1} \\
&= \mathrm{d}(R_g \circ L_h) \circ X \circ R_g^{-1} \\
&= \mathrm{d}R_g \circ \mathrm{d}L_h \circ X \circ R_g^{-1} \\
&= \mathrm{d}R_g \circ X \circ L_h \circ R_g^{-1} \\
&= \mathrm{d}R_g \circ X \circ R_g^{-1} \circ L_h \\
&= (R_g)_*X \circ L_h.
\end{aligned}$$

■

Exercise 4.7.13 (Derivative of product of curves)**.** Let G be a Lie group. Let $\gamma : \mathbb{R} \to G$ and $\sigma : \mathbb{R} \to G$ be two smooth curves into G. Consider the product of the two curves, i.e., the curve

$$t \mapsto \gamma(t)\sigma(t)$$

and calculate the derivative of such a curve in terms of γ, σ, and their derivatives. In fact, the formula is

$$\frac{\mathrm{d}}{\mathrm{d}t}\gamma(t)\sigma(t) = (\,\mathrm{d}R_{\sigma(t)})_{\gamma(t)}\dot{\gamma}(t) + (\,\mathrm{d}L_{\gamma(t)})_{\sigma(t)}\dot{\sigma}(t). \tag{4.7.14}$$

Solution.

Derivating one variable at a time, we get

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\gamma(t)\sigma(t)\bigg|_{t=t_0} &= \frac{\mathrm{d}}{\mathrm{d}t}\gamma(t)\sigma(t_0)\bigg|_{t=t_0} + \frac{\mathrm{d}}{\mathrm{d}t}\gamma(t_0)\sigma(t)\bigg|_{t=t_0}\\ &= \frac{\mathrm{d}}{\mathrm{d}t}(R_{\sigma(t_0)}\gamma(t))\bigg|_{t=t_0} + \frac{\mathrm{d}}{\mathrm{d}t}(L_{\gamma(t_0)}\sigma(t))\bigg|_{t=t_0}\\ &= (\,\mathrm{d}R_{\sigma(t_0)})_{\gamma(t_0)}\dot{\gamma}(t_0) + (\,\mathrm{d}L_{\gamma(t_0)})_{\sigma(t_0)}\dot{\sigma}(t_0).\end{aligned}$$

■

Exercise 4.7.15. Let G be a Lie group. Let $\gamma : \mathbb{R} \to G$ be a smooth curve into G. Consider the curve

$$t \mapsto \gamma(t)^{-1}$$

and calculate the derivative at an arbitrary t of such a curve in terms of γ and $\dot{\gamma}$. In fact, the formula is

$$\frac{\mathrm{d}}{\mathrm{d}t}(\gamma(t)^{-1}) \quad = \quad -(\,\mathrm{d}L_{\gamma(t)^{-1}})_{1_G}(\,\mathrm{d}R_{\gamma(t)^{-1}})_{\gamma(t)}\dot{\gamma}(t). \tag{4.7.16}$$

Solution. From the fact that $e = \gamma(t)\gamma(t)^{-1}$, for all t, and formula (4.7.14), we have

$$0 = (\,\mathrm{d}R_{\gamma(t)^{-1}})_{\gamma(t)}\dot{\gamma}(t) + (\,\mathrm{d}L_{\gamma(t)})_{\gamma(t)^{-1}}\frac{\mathrm{d}}{\mathrm{d}t}(\gamma(t)^{-1}).$$

Thus

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}(\gamma(t)^{-1}) &= -\left((\,\mathrm{d}L_{\gamma(t)})_{\gamma(t)^{-1}}\right)^{-1}(\,\mathrm{d}R_{\gamma(t)^{-1}})_{\gamma(t)}\dot{\gamma}(t)\\ &= -(\,\mathrm{d}L_{\gamma(t)^{-1}})_{1_G}(\,\mathrm{d}R_{\gamma(t)^{-1}})_{\gamma(t)}\dot{\gamma}(t).\end{aligned}$$

■

Exercise 4.7.17. Let $\varphi : G \to H$ be a group homomorphism, then

(i) $\varphi \circ L_g = L_{\varphi(g)} \circ \varphi$, for all $g \in G$;

(ii) $\varphi \circ R_g = R_{\varphi(g)} \circ \varphi$, for all $g \in G$.

■

Exercise 4.7.18. Let $\varphi : G \to H$ be a Lie group homomorphism. Given a left-invariant vector field X on G, let $\varphi_* X$ be the left-invariant vector field on H for which $(\varphi_* X)_{1_H} = (\,\mathrm{d}\varphi)_{1_G}(X_{1_G})$.

(i) The vector fields X and $\varphi_* X$ are φ-related, i.e., $(\,\mathrm{d}\varphi)_g X_g = (\varphi_* X)_{\varphi(g)}$.

(ii) If $g, g' \in G$ are such that $\varphi(g) = \varphi(g')$ and X is a left-invariant vector field on G, then $(\,\mathrm{d}\varphi)_g X_g = (\,\mathrm{d}\varphi)_{g'} X_{g'}$.

(iii) For all $g \in G$, we have $(\,\mathrm{d}\varphi)_g(X_g) = (\,\mathrm{d}L_{\varphi(g)})_{1_H}(\,\mathrm{d}\varphi)_{1_G} X_{1_G}$. Hence, $\varphi_* X$ is the left-invariant extension of the (a-priori-not-well-defined) vector field on H given as the push forward of X via φ.

(iv) $\varphi_* : \mathrm{Lie(G)} \to \mathrm{Lie(H)}$ is a Lie algebra homomorphism.

(v) $(\,\mathrm{d}\varphi)_{1_G} : (T_{1_G}, [\cdot,\cdot]) \to (T_{1_H} H, [\cdot,\cdot])$ is a Lie algebra homomorphism.

Hints. From Exercise 4.7.17.(i), we have

$$\begin{aligned}
(\varphi_* X)_{\varphi(g)} &= (\,\mathrm{d}L_{\varphi(g)})_{1_H}(\,\mathrm{d}\varphi)_{1_G} X_{1_G} \\
&= (\,\mathrm{d}(L_{\varphi(g)} \circ \varphi))_{1_G} X_{1_G} \\
&= (\,\mathrm{d}(\varphi \circ L_g))_{1_G} X_{1_G} \\
&= (\,\mathrm{d}\varphi)_g(\,\mathrm{d}L_g)_{1_G} X_{1_G} \\
&= (\,\mathrm{d}\varphi)_g X_g.
\end{aligned}$$

For $X, Y \in \mathrm{Lie(G)}$, on the one hand $[X, Y] \in \mathrm{Lie(G)}$, on the other hand $[X, Y]$ and $[\varphi_* X, \varphi_* Y]$ are φ-related. Thus $\varphi_*[X, Y] = [\varphi_* X, \varphi_* Y]$.

■

Exercise 4.7.19. Let $X^\dagger$ be a right-invariant vector field on G. Show that θ is a one-parameter subgroup with $\dot\theta(0) = X^\dagger_{1_G}$ if and only if $\theta(t) = \Phi^t_{X^\dagger}(1_G)$, for all $t \in \mathbb{R}$. ■

Exercise 4.7.20. For a vector X in the Lie algebra of a Lie group G, define the map $\varphi : \mathrm{Lie}(\mathbb{R}) \to \mathrm{Lie}(G)$, as $t \mapsto tX$.

(i). show that φ is a Lie algebra homomorphism.

(ii) show that there exists a one-parameter subgroup $\gamma : \mathbb{R} \to G$ with $\mathrm{d}\,\gamma = \varphi$.

(iii) Show that $\dot{\gamma}(t) = X_{\gamma(t)}$.

Solution. Since $\mathbb{R}$ is simply connected, Theorem **??** asserts that there exists such a γ. Regarding (iii), we have

$$\begin{aligned}
\dot{\gamma}(t) &= \frac{\mathrm{d}}{\mathrm{d}h}\gamma(t+h)\Big|_{h=0} \\
&= \frac{\mathrm{d}}{\mathrm{d}h}\gamma(t)\gamma(h)\Big|_{h=0} \\
&= \frac{\mathrm{d}}{\mathrm{d}h}L_{\gamma(t)}(\gamma(h))\Big|_{h=0} \\
&= (L_{\gamma(t)})_*(\,\mathrm{d}\gamma)_0(\partial_t) \\
&= (L_{\gamma(t)})_*\varphi(\partial_t) \\
&= (L_{\gamma(t)})_*X \\
&= X_{\gamma(t)}.
\end{aligned}$$

■

Exercise 4.7.21. Let G be a Lie group and X be a left-invariant vector field on G.

i). $\Phi^1_{tX} = \Phi^t_X$, for all $t \in \mathbb{R}$;

ii). $\exp(tX) = \Phi^t_X(1_G)$, for all $t \in \mathbb{R}$;

iii). $\exp(sX + tX) = \exp(sX)\exp(tX)$, for all $s, t \in \mathbb{R}$;

iv). $\exp(0) = 1_G$;

v). $\exp(-X) = (\exp(X))^{-1}$;

vi). $t \mapsto \exp(tX)$ is a one-parameter subgroup and an integral curve of X.

■

Exercise 4.7.22. Let G be a Lie group. Show that for all $m \in \mathbb{Z}$ and $X \in \mathrm{Lie(G)}$ we have

$$\exp(mX) = (\exp(X))^m.$$

Pay attention that your proof is correct also when m is negative. ■

Exercise 4.7.23. Let X be a left-invariant vector field in a Lie group G. Show that for all function $f \in C^\infty(G)$ and all $t \in \mathbb{R}$

$$Xf(\exp(tX)) = \frac{\mathrm{d}}{\mathrm{d}t} f(\exp(tX)).$$

■

Exercise 4.7.24. Let G be a Lie group. Let $\gamma : \mathbb{R} \to G$ be a smooth curve into G with $\gamma(0) = 1_G$ and $\dot\gamma(0) = X$. Prove that for all $t \in \mathbb{R}$

$$\lim_{k\to\infty} (\gamma(t/k))^k = \exp(tX).$$

Solution. For t small enough, one can consider $\eta(t) := \exp^{-1}(\gamma(t))$. For fixed $t \in \mathbb{R}$ it holds

$$\begin{aligned} t\eta'(0) &= t \lim_{h\to 0} \frac{\eta(h) - \eta(0)}{h} = t \lim_{k\to\infty} \frac{\eta(t/k) - \eta(0)}{t/k} \\ &= \lim_{k\to\infty} k\eta(t/k). \end{aligned}$$

Because of (4.2.10), we write

$$\eta'(0) = \frac{\mathrm{d}}{\mathrm{d}t} \exp^{-1}(\gamma(t))\bigg|_{t=0} = (\,\mathrm{d}\exp^{-1})_0 \dot\gamma(0) = (\,\mathrm{d}\exp)_0^{-1} \dot\gamma(0) = \dot\gamma(0).$$

Thus, using the previous two formulas, we get

$$\begin{aligned} \exp(tX) &= \exp(t\dot\gamma(0)) = \exp(t\eta'(0)) = \exp(\lim_{k\to\infty} k\eta(t/k)) \\ &= \lim_{k\to\infty} \exp(k\eta(t/k)) = \lim_{k\to\infty} (\exp(\eta(t/k))^k = \lim_{k\to\infty} (\gamma(t/k))^k, \end{aligned}$$

where in the last equality the definition of η can be used since as in the limit for a fixed t we always are near enough the identity element. ■

Exercise 4.7.25. Let X be a left-invariant vector field on a Lie group G.

(i). $t \mapsto p \cdot \exp(tX)$ is the flow line of X starting from p;

(ii). The flow of X at time 1 is the right translation by $\exp(X)$, i.e., for all $p \in G$

$$\Phi_X^1(p) = p \cdot \exp(X);$$

(iii). For all $p \in G$ and $t \in \mathbb{R}$

$$\Phi_X^t(p) = R_{\exp(tX)}(p).$$

■

Proof. (ii) and (iii) immediately follow from (i).

Let $\sigma(t) := p\exp(tX)$. The curve σ starts at $\sigma(0) = p$. Its derivative at an arbitrary t is

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}\sigma(t) &= \frac{\mathrm{d}}{\mathrm{d}t}p\exp(tX) \\
&= \frac{\mathrm{d}}{\mathrm{d}t}L_p(\exp(tX)) \\
&= (\mathrm{d}L_p)_{\exp(tX)}\frac{\mathrm{d}}{\mathrm{d}t}\exp(tX) \\
&= (\mathrm{d}L_p)_{\exp(tX)}X_{\exp(tX)} \\
&= X_{p\exp(tX)} \\
&= X_{\sigma(t)},
\end{aligned}$$

where we used Proposition 4.7.21 and that X is left invariant. □

Exercise 4.7.26. Analogously to Exercise 4.7.25, show (4.2.9). Namely, let $X \in T_{1_G}G$ and let $X^\dagger$ be the right-invariant vector field such that $(X^\dagger)_{1_G} = X$. Then, recalling Exercise 4.7.19, show that for all $p \in G$ and all $t \in \mathbb{R}$

$$\Phi^t_{X^\dagger}(p) = L_{\exp(tX)}(p).$$

■

Exercise 4.7.27. Show that if X is a LIVF and Y is a RIVF, then $[X, Y] = 0$. ■

Exercise 4.7.28. Show that if G is a commutative Lie group, then $\mathrm{Lie}(G)$ is a commutative Lie algebra.

Hint: Make use of Exercise 4.7.27. ■

We show next what happens if we use right-invariant vector fields as Lie algebra.

Exercise 4.7.29. For $X, Y \in T_{1_G}G$. Let $\tilde{X}, \tilde{Y}$ be the left-invariant vector fields such that $\tilde{X}_{1_G} = X$ and $\tilde{Y}_{1_G} = Y$. Let $X^\dagger$ and $Y^\dagger$ be the right-invariant vector fields with $(X^\dagger)_{1_G} = X$ and $(Y^\dagger)_{1_G} = Y$.

(i). We have $[X^\dagger, Y^\dagger]_{1_G} = -[\tilde{X}, \tilde{Y}]_{1_G}$.

(ii). Setting $[X, Y]_R := [X^\dagger, Y^\dagger]_{1_G}$, the two Lie algebras $\mathfrak{g} = (T_{1_G}G, [\cdot,\cdot])$ and $(T_{1_G}G, [\cdot,\cdot]_R)$ are isomorphic Lie algebras via the map $X \mapsto -X$.

■

Proof. Consider the map $J : G \to G^\dagger$, $J(g) = g^{-1}$ from the group $G = (G, \cdot)$ to $G^\dagger = (G, *)$, where

$$g * h := h \cdot g, \qquad \forall g, h \in G.$$

Notice that $G^\dagger$ is a Lie group. Observe that J is a Lie group isomorphism:

$$J(g \cdot h) = (g \cdot h)^{-1} = h^{-1} \cdot g^{-1} = g^{-1} * h^{-1} = J(g) * J(h).$$

We claim that

$$J_* \tilde{X} = -X^\dagger, \qquad \forall X \in T_{1_G} G. \tag{4.7.30}$$

Indeed, using Proposition 4.2.7, for all $g \in G$ we have

$$\begin{aligned}
(\mathrm{d}J)_g \tilde{X}_g &= \frac{\mathrm{d}}{\mathrm{d}t} J(g \exp(tX))\Big|_{t=0} \\
&= \frac{\mathrm{d}}{\mathrm{d}t} \exp(-tX) \cdot g^{-1}\Big|_{t=0} \\
&= \frac{\mathrm{d}}{\mathrm{d}t} R_{g^{-1}}(\exp(-tX))\Big|_{t=0} \\
&= (\mathrm{d}R_{g^{-1}})_{1_G}(-X) \\
&= -(X^\dagger)_{g^{-1}} \\
&= -(X^\dagger)_{J(g)}.
\end{aligned}$$

This proves (4.7.30).

The proof of the proposition is thus complete, because $J_* : (T_{1_G} G, [\cdot,\cdot]) \to (T_{1_G} G, [\cdot,\cdot]_R)$ is the Lie algebra isomorphism we were looking for. □

Exercise 4.7.31. Show that $X \mapsto \Phi^1_X(1_G) =: \exp(X)$ is smooth.

Solution. We see the map exp as the projection of a flow at time 1 of a particular vector field on the manifold $G \times \mathfrak{g}$. Define $Y \in \Gamma(T(G \times \mathfrak{g}))$ as, for all $(g, X) \in \mathbb{G} \times \mathfrak{g}$,

$$Y_{(g,X)} := (X_g, 0) \in T_g G \times T_X \mathfrak{g} \simeq T_{(g,X)}(G \times \mathfrak{g}).$$

We claim that we have

$$\Phi^t_Y((g, X)) = (\Phi^t_X(g), X).$$

Indeed,

$$\frac{\mathrm{d}}{\mathrm{d}t}(\Phi^t_X(g), X) = (X_{\Phi^t_X(g)}, 0) = Y_{(\Phi^t_X(g),X)}$$

and $(\Phi^t_X(g), X)|_{t=0} = (g, X)$.

We deduce that $\Phi^t_X(g)$, which is the first coordinate of the above flow, depends smoothly on the point (g, X) and so does $\Phi^1_X(1)$. ■

Exercise 4.7.32. Let $\varphi_1, \varphi_2 : G \to H$ be two Lie group homomorphisms such that the associated Lie algebra homomorphisms $(\varphi_1)_*, (\varphi_2)_*$ coincide. Assume that G is connected. Show that $\varphi_1 = \varphi_2$. Give a counterexample in the case when G is not connected. ■

Exercise 4.7.33. Use Proposition 4.2.11 to show that, if F is a Lie group homomorphism, then

i). F_* is injective (resp. surjective) if and only if F is locally injective (resp. open) at 1_G.

ii). Given $g \in G$, F_* is injective (resp. surjective) if and only if F is locally injective (resp. open) at g.

iii). F_* is injective (resp. surjective) if and only if F is locally injective (resp. open).

iv). If F is bijective, then F^{-1} is smooth, hence F is a diffeomorphism.

■

Exercise 4.7.34. Prove that a bijective Lie group homomorphism has continuous inverse. Hint: Use Proposition 4.2.11. ■

Exercise 4.7.35. Prove that an injective Lie group homomorphism is an immersion (i.e., the differential is injective). Hint: Use Proposition 4.2.11. ■

Exercise 4.7.36 (Square root of a matrix)**.** Let G be a Lie group with Lie algebra $\mathfrak{g}$ and identity component G°.

(i) Show that $\exp(\mathfrak{g}) \subset G^\circ$.

(ii) Show that for all $A \in \exp(\mathfrak{g})$ there exists $B \in G$ such that $B^2 = A$. (Every such a B is called *a square root* of A).

Hint. If $A = \exp(X)$ take $B := \exp(\frac{1}{2}X)$. ■

Exercise 4.7.37 (Non-surjective exponential)**.** Let $\mathrm{GL}^+(n, \mathbb{R})$ be the subgroup of $\mathrm{GL}(n, \mathbb{R})$ consisting of the matrices with positive determinant.

(i) Show that $\mathrm{GL}^+(n, \mathbb{R})$ is open and connected, and it is the identity component of $\mathrm{GL}(n, \mathbb{R})$.

(ii) Using Exercise 4.7.36, show that for some $n \in \mathbb{N}$ the map $\exp : \mathfrak{gl}(n, \mathbb{R}) \to \mathrm{GL}^+(n, \mathbb{R})$ is not surjective.

Hint: Try $\begin{pmatrix} -1 & 0 \\ 0 & -2 \end{pmatrix}$ or $\begin{pmatrix} -1 & -1 \\ 0 & -1 \end{pmatrix} \in \mathrm{GL}^+(2, \mathbb{R})$. ■

Exercise 4.7.38. Show (4.4.2).i and (4.4.2).ii.

Solution. Regarding the second identity: let $Z \in \mathfrak{g}$, then

$$\begin{aligned} \mathrm{ad}([X,Y])(Z) &= [[X,Y],Z] \\ &\overset{Jacobi}{=} [X,[Y,Z]] - [Y,[X,Z]] \\ &= \mathrm{ad}_X(\mathrm{ad}_Y(Z)) - \mathrm{ad}_Y(\mathrm{ad}_X(Z)) \\ &= [\mathrm{ad}_X, \mathrm{ad}_Y](Z), \end{aligned}$$

where we used the Jacobi Identity. ■

Exercise 4.7.39. Show that $\mathrm{Ad} : G \to \mathrm{GL}(\mathfrak{g})$ is a group homomorphism.

Solution. Let $g, h \in G$ and differentiate at 1_G the identity $C_g \circ C_h = C_{gh}$, to get

$$\mathrm{Ad}(gh) = (\,\mathrm{d}C_{gh})_1 = (\,\mathrm{d}C_g)_1 \circ (\,\mathrm{d}C_h)_1 = \mathrm{Ad}(g) \circ \mathrm{Ad}(h).$$

■

Exercise 4.7.40. For all X, Y in the Lie algebra of a Lie group,

$$\exp(X)\exp(Y)\exp(-X) = \exp(e^{\mathrm{ad}_X} Y).$$

Solution. Using Formula 4.4.6 first and then Formula 4.4.7, we have

$$\exp(X)\exp(Y)\exp(-X) = \exp(\mathrm{Ad}_{\exp(X)} Y) = \exp(e^{\mathrm{ad}_X} Y).$$

■

Exercise 4.7.41. Let V be a vector space. For all $A \in \mathfrak{gl}(V)$ and $B \in \mathrm{GL}(V)$

$$\mathrm{Ad}_B(A) = B \cdot A \cdot B^{-1}.$$

Solution.

$$\mathrm{Ad}_B(A) \overset{\mathrm{def}}{=} (\,\mathrm{d}C_B)_{\mathbb{I}} A = \frac{\mathrm{d}}{\mathrm{d}t} C_B(e^{tA})\bigg|_{t=0}$$

$$= \frac{\mathrm{d}}{\mathrm{d}t} B e^{tA} B^{-1}\bigg|_{t=0} = \frac{\mathrm{d}}{\mathrm{d}t} e^{tBAB^{-1}}\bigg|_{t=0} = BAB^{-1}.$$

■

Exercise 4.7.42. For all $X, Y \in \mathfrak{gl}(V)$

$$e^{\operatorname{ad}_X} Y = e^X Y e^{-X}.$$

Solution. Using Formula 4.4.7 first and then Exercise 4.7.41, we have

$$e^{\operatorname{ad}_X} Y = \operatorname{Ad}_{e^X} Y = e^X Y e^{-X}.$$

■

Exercise 4.7.43. For all $A \in \mathfrak{gl}(V)$ and for all $B \in \mathrm{GL}(V)$

$$e^{BAB^{-1}} = Be^A B^{-1}.$$

■

Exercise 4.7.44. Let X, Y left-invariant vector fields on a Lie group G. For all $t \in \mathbb{R}$

$$(\Phi_X^t)_* Y = e^{-\operatorname{ad}(tX)} Y.$$

Solution.

$$\begin{aligned}
(\Phi_X^t)_*(Y) &= (R_{\exp(tX)})_* Y \\
&= (R_{\exp(tX)})_* (L_{\exp(-tX)})_* Y \\
&= (R_{\exp(tX)} \circ L_{\exp(-tX)})_* Y \\
&= (C_{\exp(-tX)})_* Y \\
&= \operatorname{Ad}_{\exp(-tX)} Y \\
&= e^{\operatorname{ad}(-tX)} Y.
\end{aligned}$$

■

Exercise 4.7.45. Show that if γ is a curve into a Lie group, then

$$\frac{\mathrm{d}}{\mathrm{d}s} \operatorname{Ad}_{\gamma(s)} = \operatorname{Ad}_{\gamma(s)} \operatorname{ad}\big((\,\mathrm{d}L_{\gamma(s)}^{-1})_{\gamma(s)} (\frac{\mathrm{d}}{\mathrm{d}s}\gamma(s))\big)$$

Solution. Use twice that $\mathrm{Ad}_p \circ \mathrm{Ad}_q = \mathrm{Ad}_{pq}$ to obtain

$$\begin{aligned}
\partial_s \mathrm{Ad}_{\gamma(s)} &= \partial_\epsilon \mathrm{Ad}_{\gamma(s+\epsilon)} |_{\epsilon=0} \\
&= \partial_\epsilon \mathrm{Ad}_{\gamma(s)} \mathrm{Ad}_{\gamma(s)^{-1}} \mathrm{Ad}_{\gamma(s+\epsilon)} |_{\epsilon=0} \\
&= \mathrm{Ad}_{\gamma(s)} \partial_\epsilon \mathrm{Ad}_{\gamma(s)^{-1}\gamma(s+\epsilon)} |_{\epsilon=0} \\
&= \mathrm{Ad}_{\gamma(s)} \mathrm{ad}(\partial_\epsilon(\gamma(s)^{-1}\gamma(s+\epsilon))|_{\epsilon=0}) \\
&= \mathrm{Ad}_{\gamma(s)} \mathrm{ad}\big((\mathrm{d}L^{-1}_{\gamma(s)})_{\gamma(s)}(\partial_s\gamma(s))\big).
\end{aligned}$$

■

Exercise 4.7.46. For all $A \in \mathrm{Mat_{n\times n}}(\mathbb{R})$, entry by entry the matrix exponential e^A is an absolutely convergent series.

Solution. Indeed, for each $M \in \mathrm{Mat_{n\times n}}(\mathbb{R})$ set

$$\|M\| = \sup\{|Mv| : |v| \le 1\},$$

where $|\cdot|$ is the Euclidean norm in $\mathbb{R}^n$. Then

$$\left\| \sum_{k=N_1}^{N_2} \frac{1}{k!} A^k \right\| \le \sum_{k=N_1}^{N_2} \frac{1}{k!} \|A^k\| \le \sum_{k=N_1}^{N_2} \frac{1}{k!} \|A\|^k \overset{N_1,N_2\to\infty}{\longrightarrow} 0.$$

■

Exercise 4.7.47. Let $A, B \in \mathfrak{gl}(n\mathbb{R})$. Show that if $AB = BA$, then $e^{A+B} = e^A e^B = e^B e^A$.

Solution.

$$\begin{aligned}
e^A \cdot e^B &= \left(\sum_{k=0}^\infty \frac{1}{k!} A^k\right) \cdot \left(\sum_{l=0}^\infty \frac{1}{l!} B^l\right) \\
&= \sum_{k,l} \frac{1}{k!}\frac{1}{l!} A^k B^l \\
&= \sum_{m=0}^\infty \sum_{j=0}^m \frac{1}{j!(m-j)!} A^j B^{m-j} \\
&= \sum_{m=0}^\infty \frac{1}{m!} \sum_{j=0}^m \binom{m}{j} A^j B^{m-j} \overset{(AB=BA)}{=} \sum_{m=0}^\infty \frac{1}{m!} (A+B)^m \\
&= e^{A+B}.
\end{aligned}$$

■

Exercise 4.7.48. Show that, for every matrix A, the matrix e^A is invertible. Hint. Use Exercise 4.7.47 and get $e^A e^{-A} = e^0 = I$. ■

Exercise 4.7.49. Calculate the exponential of the matrix $t\begin{pmatrix} 0 & a & c \\ 0 & 0 & b \\ 0 & 0 & 0 \end{pmatrix}$. ■

Exercise 4.7.50. Let $A, B \in \mathfrak{gl}(n\mathbb{R})$. Find A, B such that $e^{A+B} \neq e^A e^B \neq e^B e^A$. Compare with Exercise 4.7.47. ■

Exercise 4.7.51. Given a matrix A and an invertible matrix B, show that

$$e^{BAB^{-1}} = Be^A B^{-1}.$$

Hint: Notice that $(BAB^{-1})^k = BA^k B^{-1}$, for all $k \in \mathbb{N}$. ■

Exercise 4.7.52. Show that the determinant function $\det : \mathrm{GL}(n, \mathbb{R}) \to (\mathbb{R}^*, \cdot)$ is a Lie group homomorphism, that the trace function $\mathrm{tr} : \mathfrak{gl}(n, \mathbb{R}) \to (\mathbb{R}, +)$ is a Lie algebra homomorphism, and that

$$\det(e^A) = e^{\mathrm{tr(A)}}.$$

Solution. Given a matrix A, there is an invertible matrix B such that $\tilde{A} = BAB^{-1}$ is upper triangular, i.e., of the form

$$\tilde{A} = \begin{pmatrix} \alpha_1 & * & * & * \\ 0 & \alpha_2 & * & * \\ 0 & 0 & \ddots & \vdots \\ 0 & 0 & \dots & \alpha_n \end{pmatrix}.$$

For such matrices we have

$$\tilde{A}^k = \begin{pmatrix} \alpha_1^k & * & * & * \\ 0 & \alpha_2^k & * & * \\ 0 & 0 & \ddots & \vdots \\ 0 & 0 & \dots & \alpha_n^k \end{pmatrix}$$

and therefore

$$e^{\tilde{A}} = \begin{pmatrix} e^{\alpha_1} & * & * & * \\ 0 & e^{\alpha_2} & * & * \\ 0 & 0 & \ddots & \vdots \\ 0 & 0 & \dots & e^{\alpha_n} \end{pmatrix}.$$

Finally, using Exercise 4.7.51 we conclude

$$\begin{aligned}
\det(e^A) &= \det(Be^AB^{-1}) \\
&= \det(e^{BAB^{-1}}) \\
&= \det(e^{\tilde{A}}) \\
&= e^{\alpha_1}\cdots e^{\alpha_n} \\
&= e^{\sum_{i=1}^n \alpha_i} \\
&= e^{\mathrm{tr}(\tilde{\mathrm{A}})} \\
&= e^{\mathrm{tr}(\mathrm{BAB}^{-1})} \\
&= e^{\mathrm{tr}(\mathrm{A})}.
\end{aligned}$$

■

Exercise 4.7.53. Show that for all $X, Y \in \mathfrak{gl}(n, \mathbb{R})$ the derivative of e^X in the direction Y has the formula:

$$\lim_{t\to 0} \frac{e^{X+tY} - e^X}{t} = \sum_{k=1}^{\infty} \frac{1}{k!} \sum_{i=1}^{k} X^{i-1} Y X^{k-i}.$$

■

Exercise 4.7.54. Deduce 4.5.5.i from Exercise 4.7.18, the chain rule, and Theorem 4.1.8. ■

Exercise 4.7.55. Show that the space $\mathrm{Der}(\mathfrak{g})$ is a Lie subalgebra of $\mathfrak{gl}(\mathfrak{g})$.

Solution. First $\mathrm{Der}(\mathfrak{g})$ is a linear subspace of $\mathfrak{gl}(\mathfrak{g})$, because equation (4.5.4) is linear in D. Second, it is easy to verify that if D, D' are derivations, then $[D, D'] := D \circ D' - D' \circ D$ is a derivation: for all $X, Y \in \mathfrak{g}$ we have

$$\begin{aligned}
[D, D']([X, Y]) &= (D \circ D')([X, Y]) - (D' \circ D)([X, Y]) \\
&= D([D'X, Y] + [X, D'Y]) - D'([DX, Y] + [X, DY]) \\
&= [DD'X, Y] + [D'(X), DY] + [DX, D'(Y)] + [X, DD'Y] \\
&\qquad - ([D'DX, Y] + [D(X), D'Y] + [D'X, DY]) + [X, D'DY])) \\
&= [(DD' - D'D)X, Y] + [X, (DD' - D'D)Y] \\
&= [[D, D']X, Y] + [X, [D, D']Y].
\end{aligned}$$

■

Exercise 4.7.56. Let $\mathfrak{g}$ be a Lie algebra.

(i) Show that ad_X is a derivation on $\mathfrak{g}$.

(ii) Show that $X \mapsto \mathrm{ad}_X$ is a Lie algebra homomorphism of $\mathfrak{g}$ into $\mathrm{Der}(\mathfrak{g})$. ■

Exercise 4.7.57. Let G be a Lie group.

(i) Show that for all $g \in G$ we have $\mathrm{Ad}_g \in \mathrm{Aut}_{\mathrm{Lie}}(\mathfrak{g})$.

(ii) Show that $g \mapsto \mathrm{Ad}_g$ is a Lie group homomorphism of G into $\mathrm{Aut}_{\mathrm{Lie}}(\mathfrak{g})$. ■

Exercise 4.7.58. Prove that the space $\mathrm{Aut}_{\mathrm{Lie}}(\mathfrak{g})$ is a closed Lie subgroup of $\mathrm{GL}(\mathfrak{g})$ whose Lie algebra is $\mathrm{Der}(\mathfrak{g})$. ■

Exercise 4.7.59. Deduce 4.5.5.iii from the previous exercises. ■

Chapter 5

SubFinsler Lie groups*

5.1 Left-invariant subFinsler structures on Lie groups

A standard assumption in the geometry of Lie groups is that the objects under consideration, such as distributions and subFinsler structures, are assumed to be left-invariant. As a result, every considered distributions is a polarization, meaning it has a constant rank. Similarly, for Lie algebras of Lie groups, we will have two interpretations for polarizations and and two for continuously varying norms, because of the left-invariance.

5.1.1 Left-invariant polarizations and horizontal curves

In this section we should interpret set-wise the Lie algebra of each Lie group G as the tangent space T_1G at the identity element $1 = 1_G$. Let G be a Lie group with Lie algebra $\mathfrak{g} = \mathrm{Lie}(G)$. Let $\Delta \subset TG$ be a distribution. Then Δ is said to be *left-invariant* if

$$\Delta_{gh} = (\mathrm{d}L_g)_h \Delta_h, \qquad \forall g, h \in G,$$

where, as in the previous chapter we denoted by L_g the left-translation by g. Because each $(\mathrm{d}L_g)_h$ is an isomorphism and left-translations act transitively, then the rank of the subspaces Δ_g is independent on $g \in G$. In other words, every left-invariant distribution is a polarization.

Moreover, every left-invariant distribution $\Delta \subset TG$ determines a vector subspaces $V := \Delta_{1_G} \subseteq \mathfrak{g}$. Vice versa, every vector subspace $V \subseteq \mathfrak{g}$ of $\mathfrak{g}$ determines a left-invariant distribution Δ, by $\Delta_{1_G} := V$ and

$$\Delta_g := \left\{ v \in T_gG : (\mathrm{d}L_g)^{-1} v \in V \right\}, \qquad \forall g \in G. \tag{5.1.1}$$

Observe that $\Delta \subseteq TG$ is indeed left-invariant and a polarization, whose rank equals $\dim(V)$. In essence, there is a one-to-one correspondence between vector subspaces of $\mathrm{Lie}(G)$ and left-invariant distributions on G.

Definition 5.1.2 (Polarized group)**.** Given a Lie groups G and a vector subspace $V \subseteq \mathrm{Lie}(G)$, we say that the pair (G, V) forms a *polarized group*, and we refer to the distribution Δ defined in (5.1.1) as the *induced (left-invariant) distribution*. We can also refer to V as a *polarization*, equating it with Δ.

Because every left-invariant distribution comes from a vector subspace of the Lie algebra it is easy to verify if the distribution is bracket generating. Indeed we have two (equivalent) way of calculating the iterated brackets and the the flag of subbundles 3.1.25. It should not be surprising that each left-invariant distribution is equiregular, cf. Definition 3.1.30, because the flag of subbundles should preserve the symmetry of being left-invariant.

In the following, given two subspaces U, V of a Lie algebra $\mathfrak{g}$, we denote by

$$[U, V] := \mathrm{span}\,\{[u, v] : u \in U, v \in V\} \subseteq \mathfrak{g}.$$

Moreover, for every vector subspace $V \subseteq \mathfrak{g}$ we iteratively define

$$V^{(1)} := \Delta, \qquad V^{(k)} := V^{(k-1)} + [V, V^{(k-1)}], \qquad \forall k = 2, 3, \dots.$$

A first observation is that, for each $k \in \mathbb{N}$ the left-invariant distribution induced by $V^{(k)}$ is the k-element in the flag of subbundles associated to Δ in Definition 3.1.25 (Exercise). A second observation is that one can the Lie algebra $\mathrm{Lie}(\Gamma(\Delta))$ generated by the sections of Δ one can use left-invariant frames, and hence look at the Lie algebra generated by $V := \Delta_{1_G}$ within $\mathfrak{g}$. In other words,

$$(\mathrm{Lie}(\Gamma(\Delta)))_1 = \cup_{k \in \mathbb{N}} V^{(k)} \qquad \text{and} \qquad (\mathrm{Lie}(\Gamma(\Delta)))_g = (\,\mathrm{d}L_g)_1 (\mathrm{Lie}(\Gamma(\Delta)))_1.$$

Moreover, we stress that the subspaces $V^{(k)}$ are nested and of integer dimension. Thus, the function $k \in \mathbb{N} \mapsto \dim(V^{(k)})$ is non-decreasing and it takes values in $\{1, 2, \dots, n\}$. Actually, unless $n \leq 1$, we have $\dim(V) > 1$. Thus, if Δ is not bracket generating then there exists $\bar{k} < n$ such that $V^{(\bar{k})} = V^{(l)}$ for every $l \geq k$. We have proved the next result, which explain when a left-invariant distribution is satisfies Chow's condition.

Proposition 5.1.3 (Criterion for bracket generation)**.** *If (G, V) is a polarized Lie group of dimension n, then we have the following dichotomy:*

(a) either $V^{(n-1)} = \mathfrak{g}$ and consequently the induced (left-invariant) distribution Δ is bracket generating with step less than n;

(b) or $V^{(n-1)} \neq \mathfrak{g}$, and in fact there exists a Lie subgroup $H < G$ with $\dim(H) < \dim(G)$ and the restriction $\Delta|_H$ is contained in TH and is bracket generating. Here $\Delta|_H := \{v \in \Delta \,:\, \pi(v) \in H\}$.

Moreover, if we denote by $\bar{k}$ the smallest integer for which $V^{(\bar{k})} = V^{(\bar{k}+1)}$, then either $V^{(\bar{k})} = \mathfrak{g}$ or $\bar{k} < n$.

We can further rephrase the notion of horizontal curve. In a polarized group (G, V) with induced distribution Δ An absolutely continous curve $\gamma\colon I \to G$ defined on an interval I is Δ-horizontal if

$$\gamma'(t) := (\,\mathrm{d}L_{\gamma(t)})^{-1}\dot{\gamma}(t) \in V, \qquad \text{for almost every } t \in I. \tag{5.1.4}$$

Notice that we just defined a V-valued curve $\gamma' : I \to V$ not to be confused with $\dot{\gamma} : I \to TG$, which is TG-valued. However, the curve γ', together with the initial point, maintains the whole information about γ, as we next see.

Proposition 5.1.5 (Integration of the tangent vector)**.** *Let G be a Lie group and let $\sigma : [a,b] \to \mathfrak{g}$ integrable defined on an interval $[a,b] \subseteq \mathbb{R}$. Then for every $p \in G$ there exists a unique $\gamma\colon [a,b] \to G$ absolutely continuous curve such that $\gamma(a) = p$ and $\sigma = \gamma'$, where the latter is defined in* (5.1.4).

Proof. We consider the ODE

$$\begin{cases} \dot{\gamma}(t) &=& (L_{\gamma(t)})_*\sigma(t) \\ \gamma(a) &=& p. \end{cases} \tag{5.1.6}$$

The existence of a solution of the ODE is a consequence of the general Carathéodory's theorem, cf.[CL55, page 43]. The uniqueness can be shown proving that, if $\gamma_1(t)$ and $\gamma_2(t)$ are two solutions, then

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\gamma_1(t)\gamma_2(t)^{-1}\right) \equiv 0.$$

This last this can be easily shown using Exercise 4.7.13 and Exercise 4.7.15. □

5.1.2 Left-invariant norms and distances on Lie groups

The first aim on this subsection is to clarify that left-invariant continuously varying norms on TG are in one to one correspondence with symmetric norms on $T_{1_G}G$. In fact, the tangent bundle TG is trivializable as $G \times \mathfrak{g}$ and, in the left-invariant case, a continuously varying norms $N : TG \simeq G \times \mathfrak{g} \to$

$\mathbb{R}$ will be smooth (actually constant in the trivialization) in the point $g \simeq (g, 0_{\mathfrak{g}})$ and will be a norm in the vector $v \simeq (1_G, v)$. Let us begin by clarifying the definition of left-invariance for the varying norm. A function $N : TG \to \mathbb{R}$ on the tangent bundle of a Lie group G is said *left-invariant* if $N \circ \mathrm{d}L_g = N$ for all $g \in G$.

One can check that, if $\|\cdot\| : TG \to \mathbb{R}$ is left-invariant and its restriction to $T_{1_G}G$ is a symmetric norm, then $\|\cdot\|$ is a continuously varying norm, in the sense of Definition 2.2.11. Moreover, every symmetric norm on $T_{1_G}G$ is the restriction to $T_{1_G}G$ of a unique left-invariant continuously varying norm $\|\cdot\|$. Indeed, if $\|\cdot\|_{1_G}$ is symmetric norm on $\mathfrak{g}$, then

$$\|v\| := \|(\mathrm{d}L_{g^{-1}})_g v\|_{1_G}, \quad \forall g \in G,\ \forall v \in T_gG, \tag{5.1.7}$$

defines a left-invariant continuously varying norm (Exercise).

Definition 5.1.8 (SubFinsler Lie group)**.** A *subFinsler Lie group* is a triple $(G, V, \|\cdot\|)$ where G is a Lie group, V is a bracket-generating subspace of $T_{1_G}G$, and $\|\cdot\|$ is a left-invariant continuously varying norm. Every subFinsler Lie group is naturally seen as a Carnot-Carathéodory space where the distribution Δ is the induced distribution from (5.1.1).

Every subFinsler Lie group has an associated subFinsler metric, which can be formulates Using the above double viewpoint for left-invariant structures. Moreover, we recall Proposition 2.1.21 about the energy of curves. Thus, the subFinsler distance d_{sF} between two points $p, q \in G$ is

$$\begin{aligned}
d_{sF}(p,q) &:= \inf\left\{\mathrm{Length}_{\|\cdot\|}(\gamma) : \gamma\ \Delta\text{-horizontal curve from } p \text{ to } p\right\}\\
&= \inf\left\{\int \|\gamma'\|_{1_G} : \gamma \text{ AC curve from } p \text{ to } p,\ \text{with } \gamma' \in V\right\}.\\
&= \inf\left\{\sqrt{2\int \|\gamma'\|_{1_G}^2} : \gamma \text{ AC curve from } p \text{ to } p,\ \text{with } \gamma' \in V\right\}.
\end{aligned}$$

In Definition 5.1.8 we made the choice of assuming that the polarization is bracket-generating. Of course, one could also consider subFinsler metrics associated to non-bracket-generating polarizations. However, because of Proposition 5.1.3, if the polarization is not bracket-generating, then one can just restrict to the Lie subgroup that it generates.

The following are some basic metric-geometry property of subFinsler Lie groups, when they are seen as metric spaces with their Carnot-Carathéodory metrics.

Theorem 5.1.9. *Every subFinsler Lie group is a metric space that is*

5.1.9.i. *complete,*

5.1.9.ii. *geodesic,*

5.1.9.iii. *boundedly compact,*

5.1.9.iv *isometrically homogeneous: the distance is left-invariant.*

Proposition 5.1.10. *If (G, V) is a polarized Lie group, then every two CC distances induced by left-invariant norms on the induced polarization are globally bi-Lipschitz.*

Proof. The notion of length of a horizontal curve γ (and hence the notion of the associated CC distance) depends on the norm $\|\cdot\|$ in the following way: $\mathrm{Length}_{\|\cdot\|}(\gamma) = \int \|\gamma'\|_{1_G}$ Since V is finite dimensional every choice of $\|\cdot\|_{1_G}$ is biLipschitz equivalent to any other. This produces a biLipschitz equivalence for CC distances. □

5.2 Endpoint map on polarized groups*

In this section, we begin with a parametrization of those horizontal curves in a polarized group that start from the identity element. The parametrization of these curves leads us to a Hilbert space structure, providing a powerful analytical framework for investigating the geometric intricacies of Carnot-Carathéodory spaces.

5.2.1 Endpoint map

Let (G, V) be a polarized group. After fixing a basis $(e_1, \ldots, e_r)$ for V we can identify V with $\mathbb{R}^r$, where we equip $\mathbb{R}^r$ with the Euclidean norm as an auxiliary tool to consider integrable functions. Namely, we consider $\Omega := L^2([0,1]; V) \cong L^2([0,1]; \mathbb{R}^r)$ and equip it with the L^2-norm

$$\|u\| := \left(\int_0^1 \sum_{i=1}^r u_i(t)^2 \, \mathrm{d}t \right)^{\frac{1}{2}}.$$

We refer to Ω as the *space of controls.*

For every $u \in \Omega$, let $\gamma_u : [0,1] \to G$ be the solution of the ODE

$$\begin{cases} \gamma(0) = 1_G, \\ \dot{\gamma}(t) = \left(\mathrm{d}L_{\gamma(t)}\right) u(t) \quad \text{for a.e. } t \in [0,1]. \end{cases} \tag{5.2.1}$$

By Carathéodory Theorem on ODEs, see Proposition 5.1.5, the equation is well posed and in this way each $u \in \Omega$ induces a V-horizontal curve γ_u on G. Every V-horizontal curve on $[0,1]$ starting from 1_G is of the form γ_u for some u. In fact, using the notation (5.1.4), if γ is horizontal, then $u := \gamma' \in \Omega$ and $\gamma = \gamma_u$. We call u the *control* of γ_u.

The endpoint map is a key concept in sub-Riemannian geometry, particularly from the perspective of control theory. It sends a control, and consequently the corresponding curve starting from a given base point, to the final point of the curve. This mapping allows for the analysis and optimization of trajectories.

The *end-point map* is defined as

$$\begin{aligned} \mathrm{End} : \Omega &\longrightarrow G \\ u &\longmapsto \mathrm{End}(u) := \gamma_u(1), \end{aligned}$$

where γ_u solves (5.2.1).

5.2.2 Differential of the endpoint map

The differential of the endpoint map allows for sensitivity analysis, which examines how small changes in the control or initial conditions affect the reachable points. In sub-Riemannian geometry, by analyzing this differential, one can derive necessary conditions for optimality.

We shall not show that the endpoint map is smooth. We will directly calculate its (first) differential.

Proposition 5.2.2. *For every $u \in \Omega$ the differential of* End *at u is*

$$\begin{aligned} \mathrm{d}\,\mathrm{End}_u : \Omega &\longrightarrow T_{\mathrm{End}(u)}(G) \\ v &\longmapsto \left(\mathrm{d}R_{\gamma_u(1)}\right)_{1_G} \int_0^1 \mathrm{Ad}_{\gamma_u(t)}(v(t))dt, \end{aligned}$$

where $\mathrm{Ad}_g : \mathfrak{g} \to \mathfrak{g}$ is defined by $\mathrm{Ad}_g = (C_g)_*$ where $C_g h = ghg^{-1}$.

Sketch of the proof. We sketch the proof for $G \subset \mathrm{GL}(n, \mathbb{R})$, where we can interpret the Lie product as a matrix product and work in the matrix coordinates. Let $\gamma_{u+\epsilon v}$ be the curve with the control $u+\epsilon v$ and $\sigma(t)$ be the derivative of $\gamma_{u+\epsilon v}(t)$ with respect to ϵ at $\epsilon = 0$. Then σ satisfies the following ODE (which is the derivation with respect to ϵ of (5.2.1) for $\gamma_{u+\epsilon v}$)

$$\frac{\mathrm{d}\sigma}{\mathrm{d}t} = \gamma(t) \cdot v(t) + \sigma(t) \cdot u(t).$$

Now it is easy to see that $t \mapsto \int_0^t \mathrm{Ad}_{\gamma(s)}(v(s))\,\mathrm{d}s \cdot \gamma(t)$ satisfies the above equation with the same initial condition as σ, hence is equal to σ. □

[...]

5.2.3 Singular curves

We study here those controls that are critical points for the endpoint map. The associated curves are called singular curves or abnormal curves.

[...]

If $u \in \Omega$ is a singular point for the endpoint map, then by definition $\mathrm{dEnd}_u : \Omega \to T_{\mathrm{End}(u)}G$ is not surjective. In this case, there is a nontrivial covector that annihilates its image, i.e, these exists $\xi \in T_{\gamma_u(1)}G$ such that $\xi \neq 0$ and

$$\langle \xi, \mathrm{dEnd}_u(v) \rangle = 0, \quad \forall v \in \Omega.$$

By Proposition 5.3.2, this is equivalent to say that

$$0 = \xi \left(\mathrm{d}R_{\gamma_u(1)} \int_0^1 \mathrm{Ad}_{\gamma_u(t)} v(t) \,\mathrm{d}t \right) = \lambda \left(\int_0^1 \mathrm{Ad}_{\gamma_u(t)} v(t) \,\mathrm{d}t \right) \quad \forall v \in \Omega, \tag{5.2.3}$$

where $\lambda \in \mathfrak{g}^*$ is defined as $\lambda := \xi \,\mathrm{d}R_{\gamma_u(1)}$. Choosing formally $v(t) = \delta_t$, i.e., letting v converge to the Dirac mass δ_t at t, we obtain that

$$\lambda \left(\mathrm{Ad}_{\gamma_u(t)} V \right) = \{0\}. \tag{5.2.4}$$

If $e_1, \ldots, e_r$ is a basis of V, then 5.2.4 rephrases as a linear system of equations: A horizontal curve is abnormal if and only if there exists $\lambda \in \mathfrak{g}$ such that $\lambda \neq 0$ and

$$\lambda \left(\mathrm{Ad}_{\gamma(t)}(e_i) \right) = 0, \qquad i = 1, \ldots, r. \tag{5.2.5}$$

In particular, since in our case $\gamma(0) = 1_G$, the last equation implies

$$\lambda(e_i) = 0, \qquad i = 1, \ldots, r. \tag{5.2.6}$$

Notice that, after we fix i and λ, the function $g \mapsto \lambda\left(\mathrm{Ad}_g(e_i)\right)$ is smooth and (5.2.5) says that $\gamma_u(t)$ lies in the zero level set of such a function. We shall notice that in nilpotent Lie groups (e.g., in Carnot groups) we have that, Ad is polynomial, hence these functions are polynomials, in exponential coordinates .

Remark 5.2.7. In Riemannian geometry there are no abnormal curves. Indeed, V is everything and so such a nonzero λ cannot exists.

5.3 Extrema in subRiemannian groups*

5.3.1 First order necessary conditions for subRiemannian minimizers

Let G be a Lie group, let $V \subseteq \mathfrak{g}$ be a subspace. We shall consider subRiemannian structures for the polarized group (G, V). A left-invariant subRiemannian structure is completely determined by the choice of an orthonormal basis $(e_1, \ldots, e_r)$ for V. We next study conditions for length-minimizing curves for this subRiemannian structure.

Recall from Proposition 2.1.21, that minimizing the length or the energy is the same. And this is also why we can restrict to control in L^2. Actually, because of Remark 2.1.22 one can also take controls in L^p with $p \in]1, \infty[$.

We consider the *energy function*

$$\begin{aligned} \mathrm{Energy} : \Omega &\longrightarrow \mathbb{R} \\ u &\longmapsto \mathrm{Energy}(u) := \frac{1}{2}\|u\|^2. \end{aligned}$$

This is the same functional we saw in (2.1.20) for metric spaces and now it satisfies

$$\mathrm{Energy}_{d_{cc}}(\gamma_u) = \frac{1}{2}\int \|\gamma_u{}'\|_{1_G}^2 = \frac{1}{2}\|u\|^2 \qquad , \forall u \in \Omega.$$

Together with the endpoint map we form the *extended end-point map*

$$\begin{aligned} \widetilde{\mathrm{End}} : \Omega &\longrightarrow G \times \mathbb{R} \\ u &\longmapsto \big(\mathrm{End}(u), \mathrm{Energy}(u)\big). \end{aligned}$$

Given a point $p \in G$ minimizing the energy between e and p rephrase as minimizing $\mathrm{Energy}(u)$ among all u for which $\gamma_u(1) = p$. We shall say that γ_u is a *minimizer for the energy*, or for short that u is a *minimizer*, if for all $v \in \Omega$ we have

$$\mathrm{End}(v) = \mathrm{End}(u) \implies E(v) \geq E(v).$$

Remark 5.3.1. If u_0 is a minimizer for the energy then $\widetilde{\mathrm{End}}$ cannot be open at any neighborhood of u_0 and therefore u_0 must be a singular point for $\widetilde{\mathrm{End}}$. Indeed, if there were a subset $U \subseteq \Omega$ for which $\widetilde{\mathrm{End}}(U)$ is a neighborhood of $\widetilde{\mathrm{End}}(u_0)$ within $G \times \mathbb{R}$, then we can find $\widetilde{u} \in U$ such that $\mathrm{End}(\widetilde{u}) = \mathrm{End}(u_0)$ and $\mathrm{Energy}(\widetilde{u}) < \mathrm{Energy}(u_0)$. This contradicts the minimality of u_0. Moreover, if the differential of $\mathrm{d}\widetilde{\mathrm{End}} : \Omega \to T_{\widetilde{\mathrm{End}}(u)}(G \times \mathbb{R})$ at u_0 were surjective, then we can take a vector subspace $W \subset \Omega$ for which $\mathrm{d}\widetilde{\mathrm{End}}|_W : W \to T_{\widetilde{\mathrm{End}}(u)}(G \times \mathbb{R})$ is an isomorphism. From the implicit function theorem, we conclude that the map $\widetilde{\mathrm{End}}|_W : W \to G \times \mathbb{R}$ gives a diffeomorphism between a neighborhood of u_0 within W and a neighborhood of $\widetilde{\mathrm{End}}(u_0)$ within $G\times\mathbb{R}$. Such a fact contradicts the property that $\widetilde{\mathrm{End}}$ cannot be open at u_0.

Because of this last remark, we need the differential of the extended endpoint map $\widetilde{\mathrm{End}}$. We recall Proposition 5.2.2 and the calculation of the differential of the energy **??**.

Proposition 5.3.2. *For every $u\in\Omega$ the differential of $\widetilde{\mathrm{End}}$ at u is*

$$\begin{aligned} \mathrm{d}\widetilde{\mathrm{End}}_u : \Omega &\longrightarrow T_{\widetilde{\mathrm{End}}(u)}(G\times\mathbb{R}) = T_{\mathrm{End}(u)}G\times\mathbb{R} \\ v &\longmapsto \left(\left(\mathrm{d}R_{\gamma_u(1)}\right)_{1_G}\int_0^1 \mathrm{Ad}_{\gamma_u(t)}(v(t))dt, \langle u,v\rangle\right). \end{aligned}$$

Assume now that γ_u is length minimizing for some $u\in\Omega$ that is energy minimizing. By Remark 5.3.1, we deduce that u is a critical point for $\widetilde{\mathrm{End}}$, that is $d\widetilde{\mathrm{End}}_u:\Omega\to T_{\mathrm{End}(u)}G\times\mathbb{R}$ is not surjective. Since then $d\widetilde{\mathrm{End}}_u(\Omega)$ is a strict subspace of $T_{\mathrm{End}(u)}G\times\mathbb{R}$, there exists $(\xi,\xi_0)\in \left(T_{\mathrm{End}(u)}G\right)^*\times\mathbb{R} = \left(T_{\mathrm{End}(u)}G\times\mathbb{R}\right)^*$ such that $(\xi,\xi_0)\neq(0,0)$ and

$$\langle(\xi,\xi_0),\, \mathrm{d}\widetilde{\mathrm{End}}_u(v)\rangle = 0, \quad \forall v\in\Omega.$$

By Proposition 5.3.2, this is equivalent to say that there exists $(\xi,\xi_0)\neq(0,0)$ such that

$$\xi\left(\mathrm{d}R_{\gamma_u(1)}\int_0^1 \mathrm{Ad}_{\gamma_u(t)}\, v(t)\,\mathrm{d}t\right) + \xi_0\langle u,v\rangle = 0 \quad \forall v\in\Omega. \tag{5.3.3}$$

Since differential of right translations are automorphisms, Equation (5.3.3) is true if and only if there exist $\lambda\in\mathfrak{g}^*$ and $\xi_0\in\mathbb{R}$ such that $(\lambda,\xi_0)\neq(0,0)$ and

$$\lambda\left(\int_0^1 \mathrm{Ad}_{\gamma_u(t)}\, v(t)\,\mathrm{d}t\right) = \xi_0\langle u,v\rangle, \quad \forall v\in\Omega. \tag{5.3.4}$$

We now consider two cases: either $\xi_0\neq0$ or $\xi_0=0$. The first case is called normal, the second one is called abnormal. We stress that in the case the codimension of $d\widetilde{\mathrm{End}}_u(\Omega)$ within $T_{\mathrm{End}(u)}G\times\mathbb{R}$ is strictly larger than 1, then there would be other choices for (λ,ξ_0). Hence, some particular u may have an normal pair (λ,ξ_0) and a (different) abnormal pair (λ',ξ_0').

Firstly, we suppose that (λ,ξ_0) as in (5.3.4) is such that $\xi_0\neq0$. Up to multiply the equation by a constant we can assume that $\xi_0=1$. Fix a Lebesgue point t of u, and let v converge to the Dirac mass at t. Formally, we have

$$\begin{aligned} \dot\gamma_u(t) &= \mathrm{d}L_{\gamma_u(t)}u(t) = \mathrm{d}L_{\gamma_u(t)}\sum_{i=1}^r \langle u,\delta_t e_i\rangle e_i \\ &\overset{(5.3.4)}{=} \mathrm{d}L_{\gamma_u(t)}\sum_{i=1}^r\left(\lambda\int_0^1 \mathrm{Ad}_{\gamma_u(s)}(\delta_t e_i)\,\mathrm{d}s\right)e_i \\ &= \sum_{i=1}^r \lambda\left(\mathrm{Ad}_{\gamma_u(t)}(e_i)\right)X_i(\gamma_u(t)), \end{aligned}$$

where in the last equality we have used the identity $X_i(g) = (\mathrm{d}L_g)\, e_i$. We therefore say that a curve γ satisfies the *normal equation* (or the *sub-Riemannian geodesic equation*) if there exists $\lambda \in \mathfrak{g}^*$ such that

$$\dot{\gamma}(t) = \sum_{i=1}^{r} \lambda\left(\mathrm{Ad}_{\gamma_u(t)}(e_i)\right) X_i(\gamma_u(t)), \qquad \text{for almost every } t \in [0,1]. \tag{5.3.5}$$

A solution to (5.3.5) is called *normal curve*. By a bootstrap argument using (5.3.5) we deduce that the horizontal curve γ and its control u are C^∞.

Recall that the curve γ_u is the solution of (5.2.1). Therefore, if we write $u = \sum_{i=1}^{r} u_i e_i$, another version of the normal equation is

$$u_i = \lambda\left(\mathrm{Ad}_{\gamma_u}(e_i)\right), \qquad \text{for almost every } t \in [0,1] \text{ and for every } i = 1, \dots, r. \tag{5.3.6}$$

In particular, since in our case $\gamma(0) = 1_G$, the last equation implies

$$u_i(0) = \lambda(e_i), \qquad i = 1, \dots, r. \tag{5.3.7}$$

Fact: every normal curve is locally length minimizing. The converse is not true.

Exercise 5.3.8. Prove that every solution of (5.3.5) is analytic and is parametrized by arclength. ■

Secondly, we suppose that (λ, ξ_0) as in (5.3.4) is such that $\xi_0 = 0$.

Exercise: In Riemannian Lie groups, all minimizers are normal.

In subRiemannian structures it is possible to find length-minimizing curves that are not normal, and so are abnormal.

Theorem 5.3.9. *In contact structures, as for example $SE(2)$, every abnormal curve is constant.*

In every subRiemannian manifold of step 2 every length minimizer is normal.

5.4 Geodesic left-invariant distances*

5.4.1 Quasi-isometric equivalence

In this section, we establish the result that geodesic left-invariant metrics on a group G are quasi-isometric. Actually, we can relax the assumption on the metrics being geodesic and instead require them to be quasi-geodesic.

A metric space (X, d) is said to be *quasi-geodesic* if there exist constants $C > 0$ and $L > 1$ such that every two points in X can be join with a (L, C)-quasi-arc. In other words, for all $x, x' \in X$, there exist $k \in \mathbb{N}$ and $x_0, x_1, \dots, x_k \in X$ such that $x_0 = x$, $x_k = x'$, $d(x_{i-1}, x_i) \leq C$, for $i = 1, \dots, k$, and $\sum_{i=1}^{k} d(x_{i-1}, x_i) \leq L d(x, x') + C$.

Also recall that if two distances on a locally compact group are locally bounded (i.e., bounded on compact sets) and proper (i.e., the distance from a point is a proper map), then they have the same bounded subsets.

Proposition 5.4.1. *Let d and d' be two quasi-geodesic left-invariant distances on a locally compact group. Assume that d and d' have the same bounded subsets. Then there exist constants $c > 0$ and $L > 1$ such that $L^{-1}d - c < d' < Ld + c$.*

Proof. Since d is quasi-geodesic, there are two constants $C_1 > 0$ and $L_1 \geq 1$ with the following property. Given any group element g, we may find $g_1, \dots, g_n$ such that $d(\mathrm{id}, g_i) \leq C_1$, for all i, and $g = g_1 \cdot \dots \cdot g_n$, while $\sum_1^n d(\mathrm{id}, g_i) \leq L_1 d(\mathrm{id}, g) + C_1$. Grouping some g_i's together if necessary, we may assume that $C_1/2 \leq d(\mathrm{id}, g_i)$, still having the weaker uniform condition $d(\mathrm{id}, g_i) \leq 2C_1$. Hence, $n \leq \frac{2L_1}{C_1} d(\mathrm{id}, g) + 2$. But then by our assumptions, $d'(\mathrm{id}, g_i)$ is uniformly bounded say by some constant $\tilde{C} > 0$. Then $d'(\mathrm{id}, g) \leq \sum_1^n d'(\mathrm{id}, g_i) \leq \tilde{C} n \leq L d(\mathrm{id}, g) + c$ for $L = \frac{2\tilde{C} L_1}{C_1}$ and $c = 2\tilde{C}$. The proposition follows by exchanging the roles of d and d' and using the assumed left invariance. □

The above applies in particular to Finsler and subFinsler left-invariant metrics on Lie groups.

Problem. Are there quasi-geodesic metrics on the plane that are not quasi-isometric to a geodesic distance?

5.5 Characterization of geodesic left-invariant distances*

5.5.1 Berestovskii's characterization

Berestovskii's work [Ber88, Theorem 2] clarified what are the possible isometrically homogeneous distances on manifolds that are also geodesic. They are subFinsler metrics.

Theorem 5.5.1 (Berestovskii)**.** *Let $M = G/H$ be the quotient of a Lie group G modulo a closed subgroup H. If M is metrized by a geodesic distance that is G-invariant, then the distance is a subFinsler metric, i.e., there is a G-invariant subbundle Δ on M and a G-invariant norm on Δ, such that the distance is given by the same formula* (3.1.15).

[...]

5.5.2 Geodesic distances on $\mathbb{R}^2$, $\mathbb{R}^3$, ...

[...]

Chapter 6

Nilpotent Lie groups*

The asterisk * will denote incompleteness of the chapter or section. Nilpotent Lie groups play a fundamental role in the study of geometric structures and differential equations. These groups are characterized by a powerful algebraic property known as nilpotency. The nilpotency assumption imposes a certain level of commutativity and control over the group's structure, leading to a wealth of remarkable properties and simplifications in various areas of mathematics. By restricting our attention to nilpotent Lie groups, we gain a deeper understanding of their geometry. The nilpotency condition, which also reflect to a condition on the Lie algebra, serves as a guiding principle that allows us to explore the interplay between algebra and geometry, paving the way for profound applications in Geometric Analysis, Geometric Group Theory, Harmonic Analysis, Control Theory, but also in Number Theory, Dynamics, Representation Theory, and ultimately in Physics. In this chapter, we delve into the fascinating world of nilpotent Lie groups, unraveling their unique features and uncovering the remarkable consequences of the nilpotency assumption.

A reference that deserves strong recommendation is [HN12]. Other valuable reading materials on this topic include [Rag72, Jac79, War83, CG90, Kna02] While our exposition may not be as comprehensive as those references, we will focus on the necessary concepts to understand Carnot groups, as well as other sub-Finsler Lie groups such as boundaries of Heintze groups, Malcev closures of finitely generated nilpotent groups, and their asymptotic cones.

Throughout our discussion, we will maintain a perspective rooted in differential geometry and linear algebra. It is worth noting that one of the compelling aspects of nilpotent Lie groups is their appearance as tangent spaces of sub-Riemannian manifolds, similar to how (Euclidean) vector groups serve as tangents to Riemannian manifolds. We will discover that, akin to vector groups, these tangents possess nilpotency, simply connectedness, and dilation structures.

6.1 Nilpotent Lie algebras

We begin this section by introducing the concept of nilpotent Lie algebra Nilpotent Lie algebras are those for which iterated brackets $[x_1, [x_2, [x_3, [, \dots]]]]]$ of sufficiently large oder vanish. We anticipate that for connected Lie groups, a Lie algebra is nilpotent if and only if the group is nilpotent as a group. This correspondence between nilpotent Lie algebras and nilpotent Lie groups is a result wee will be further explore, see **??**. Typical examples of nilpotent Lie algebras are Lie algebras of strictly upper triangular matrices, where the diagonal elements are all zero. A first result on nilpotent Lie algebras is Engel's Theorem, which is a translation of nilpotency into a pointwise condition.

Definition 6.1.1 (Nilpotent Lie algebra)**.** Let $\mathfrak{g}$ be a Lie algebra (over $\mathbb{R}$). The *descending (lower) central series* of $\mathfrak{g}$ is inductively defined by

$$C^1(\mathfrak{g}) := \mathfrak{g}^{(1)} := \mathfrak{g}, \qquad C^2(\mathfrak{g}) := \mathfrak{g}^{(2)} := [\mathfrak{g}, \mathfrak{g}] = [\mathfrak{g}, C^1(\mathfrak{g})],$$
$$C^n(\mathfrak{g}) := \mathfrak{g}^{(n)} := [\mathfrak{g}, C^{n-1}(\mathfrak{g})], \qquad \forall n \in \mathbb{N}.$$

Here, for $V, W \subseteq \mathfrak{g}$, we define $[V, W] := \operatorname{span}\{[v, w] \,:\, v \in V, w \in W\}$. The Lie algebra $\mathfrak{g}$ is said to be *nilpotent* if there is $d \in \mathbb{N}$ such that $C^{d+1}(\mathfrak{g}) = \{0\}$. If d is minimal with this property, then it is called *nilpotency degree* (or *step*) of $\mathfrak{g}$,, and $\mathfrak{g}$ is said *d-step nilpotent*

One can rephrase the definition saying that a Lie algebra $\mathfrak{g}$ is s-step nilpotent if and only if all brackets of at least $s+1$ elements of $\mathfrak{g}$ are 0 but not all brackets of order s are.

Remark 6.1.2. Each $C^n(\mathfrak{g})$ is an ideal and actually

$$[C^n(\mathfrak{g}), C^n(\mathfrak{g})] \subseteq [C^n(\mathfrak{g}), \mathfrak{g}] =: C^{n+1}(\mathfrak{g}) \subseteq C^n(\mathfrak{g}),$$

where the last inclusion holds by induction noticing that $C^2(\mathfrak{g}) \subseteq C^1(\mathfrak{g})$.

For finite-dimensional Lie algebras, the nilpotency of $\mathfrak{g}$ is equivalent to the vanishing of the ideal

$$C^\infty(\mathfrak{g}) := \bigcap_{n \in \mathbb{N}} C^n(\mathfrak{g}).$$

A nilpotent Lie algebra $\mathfrak{g}$ has always non-trivial center, see Exercise **??**.; in fact, if $\mathfrak{g}$ is s-step nilpotent, $\mathfrak{g}^{(s)}$ is central, i.e., it is contained in the center [1] of $\mathfrak{g}$. Be aware that the center might be strictly larger than $\mathfrak{g}^{(s)}$, see Exercise 6.6.16.

[1] Recall that the *center* of a Lie algebra $\mathfrak{g}$ is $\operatorname{Center}(\mathfrak{g}) := \{X \in \mathfrak{g} \,:\, [X, Y] = 0 \text{ for all } Y \in \mathfrak{g}\}$.

Proposition 6.1.3. *Let $\mathfrak{g}$ be a Lie algebra.*

i) *If $\mathfrak{g}$ is nilpotent, then all its subalgebras and all homomorphic images of $\mathfrak{g}$ are nilpotent.*

ii) *If $\mathfrak{a} < Z(\mathfrak{g}) := \{x \in \mathfrak{g} : \forall y \in \mathfrak{g}\ [x, y] = 0\}$ and if $\mathfrak{g}/\mathfrak{a}$ is nilpotent then so is $\mathfrak{g}$.*

iii) *If $\mathfrak{g} \neq \{0\}$ and $\mathfrak{g}$ is nilpotent, then $Z(\mathfrak{g}) \neq \{0\}$.*

iv) *If $\mathfrak{g}$ is nilpotent of step n, then $\mathrm{ad}(x)^n \equiv 0$ for every $x \in \mathfrak{g}$, i.e., the maps $\mathrm{ad}(x) : \mathfrak{g} \to \mathfrak{g}$ defined as $\mathrm{ad}(x)(y) = [x, y]$ are nilpotent as linear transformations.*

v) *If $\mathfrak{i} \trianglelefteq \mathfrak{g}$, then $C^n(i)$ are ideals of $\mathfrak{g}$, for every $n \in \mathbb{N}$.*

Proof. (i). Let $\mathfrak{h}$ be a subalgebra of $\mathfrak{g}$, i.e., $\mathfrak{h} < \mathfrak{g}$. Then $[\mathfrak{h}, \mathfrak{h}] \subset [\mathfrak{g}, \mathfrak{g}]$ and so, by induction, $C^n(\mathfrak{h}) \subset C^n(\mathfrak{g})$. Hence, $\mathfrak{h}$ is nilpotent if so is $\mathfrak{g}$. Moreover, if we consider $\alpha : \mathfrak{g} \to \mathfrak{h}$ be Lie algebra homomorphism, then by $[\alpha(\mathfrak{g}), \alpha(\mathfrak{g})] = \alpha([\mathfrak{g}, \mathfrak{g}])$ we have, by induction, that

$$C^n(\alpha(\mathfrak{g})) = \alpha(C^n(\mathfrak{g})), \tag{6.1.4}$$

for any $n \in \mathbb{N}$. Consequently, $\alpha(\mathfrak{g})$ is nilpotent if so is $\mathfrak{g}$.

(ii). If $\mathfrak{g}/\mathfrak{a}$ is nilpotent, by definition we know that there is $n \in \mathbb{N}$ such that $C^n(\mathfrak{g}/\mathfrak{a}) = \{0\}$ in $\mathfrak{g}/\mathfrak{a}$. Now apply (6.1.4) with α =projection then we deduce that $C^n(\mathfrak{g}) + \mathfrak{a} = C^n(\mathfrak{g}/\mathfrak{a}) = \{0\} + \mathfrak{g}/\mathfrak{a}$, i.e., $C^n(\mathfrak{g}) \subset \mathfrak{a} \subset Z(\mathfrak{g})$. This implies that $\mathfrak{g}$ is nilpotent, indeed

$$C^{n+1}(\mathfrak{g}) = [\mathfrak{g}, C^n(\mathfrak{g})] = [\mathfrak{g}, Z(\mathfrak{g})] = \{0\},$$

as desired.

(iii). By hypothesis, we know that there is $n \in \mathbb{N}$ such that $C^n(\mathfrak{g}) = \{0\}$ and $C^{n-1}(\mathfrak{g}) \neq \{0\}$. Then, $\{0\} \neq C^{n-1}(\mathfrak{g}) \subseteq Z(\mathfrak{g})$ since $[C^{n-1}(\mathfrak{g}), \mathfrak{g}] = C^n(\mathfrak{g}) = \{0\}$.

(iv). Since $\mathfrak{g}$ is n-step nilpotent we have that $C^{n+1}(\mathfrak{g}) = \{0\}$ and so for every $x \in \mathfrak{g}$

$$(\mathrm{ad}(x))^n(\mathfrak{g}) = [\underbrace{x, [x, \dots [x}_{n \text{ times}}, \mathfrak{g}], \dots] \subseteq C^{n+1}(\mathfrak{g}) = \{0\}.$$

(v). It follows from the general easy implication

$$\mathfrak{i}, \mathfrak{j} \trianglelefteq \mathfrak{g} \quad \Rightarrow \quad [\mathfrak{i}, \mathfrak{j}] \trianglelefteq \mathfrak{g}.$$

$\square$

6.1.1 Examples of nilpotent Lie algebras

We give some examples of nilpotent Lie algebras.

Example 6.1.5. Abelian Lie algebra are those for which $[\cdot,\cdot] \equiv 0$. Consequently, a Lie algebra is Abelian if and only if it is nilpotent with nilpotency step equal to 1, because $C^2(\mathfrak{g}) = [\mathfrak{g},\mathfrak{g}]$.

Example 6.1.6. The Heisenberg Lie algebra

$$\mathfrak{nil}_3 := \left\{ \begin{pmatrix} 0 & x & z \\ 0 & 0 & y \\ 0 & 0 & 0 \end{pmatrix} : x, y, z \in \mathbb{R} \right\} \subseteq \mathfrak{gl}(3)$$

is nilpotent of step 2, because

$$C^2(\mathfrak{g}) = \left\{ \begin{pmatrix} 0 & 0 & z \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} : z \in \mathbb{R} \right\} \qquad \text{and} \qquad C^2(\mathfrak{g}) = \mathbf{0}.$$

Example 6.1.7. Let V, W be vector spaces and $q : V \times V \to W$ skew-symmetric bilinear map then $[(v_1, w_1), (v_2, w_2)] := (0, q(v_1, v_2))$ is a Lie bracket on $V \times W$, and $V \times W$ becomes a step-2 Lie algebra. Namely, $[[x, y], z] = 0$ for every $x, y, z \in V \times W$.

For $n \in \mathbb{N}$ we consider

$$V := \Lambda^1(\mathbb{R}^n) = \{1 - \text{forms on } \mathbb{R}^n\},$$

$$W := \Lambda^2(\mathbb{R}^n) = \{2 - \text{forms on } \mathbb{R}^n\},$$

$$q(v_1, v_2) := v_1 \wedge v_2,$$

where $q(\cdot,\cdot)$ is the wedge of 1-forms. Then $\Lambda^1(\mathbb{R}^n) \times \Lambda^2(\mathbb{R}^n)$ becomes a step-2 Lie algebra called the *free-nilpotent Lie algebra* of rank n and step 2 (to be continued...).

One common convention in describing nilpotent Lie algebras - and one that we shall often use - is the following. Suppose that $\mathfrak{g} = \mathbb{R}\text{-span}\{X_1, \dots, X_n\}$. To describe the Lie algebra structure of $\mathfrak{g}$, it suffices to give $[X_i, X_j]$ for all $i < j$. We can shorten this description considerably by giving only the non-zero brackets; all others are assumed to be zero.

Example 6.1.8 (Heisenberg algebras)**.** The $(2n+1)$-dimensional Heisenberg algebra is the Lie algebra with basis $\{X_1, , \dots, X_n, Y_1, , \dots, Y_n, Z\}$, whose pairwise brackets are equal to zero expect for

$$[X_j, Y_j] = Z, \quad \text{for } j = 1, \dots, n.$$

It is a two-step nilpotent Lie algebra. One way to realize it as a matrix algebra is to consider $(n+2)\times(n+2)$ upper triangular matrices of the form

$$\begin{pmatrix} 0 & x_1 & \dots & x_n & z \\ \cdot & 0 & \cdot & 0 & y_1 \\ . & & . & . & \vdots \\ \cdot & & & 0 & y_n \\ 0 & \cdot & \cdot & \cdot & 0 \end{pmatrix}.$$

The Lie group associated is called the n'th Heisenberg group and as matrix group it is

$$G = \left\{ \begin{pmatrix} 1 & x_1 & \dots & x_n & z \\ \cdot & 1 & \cdot & 0 & y_1 \\ . & & \ddots & . & \vdots \\ \cdot & & & 1 & y_n \\ 0 & \cdot & \cdot & \cdot & 1 \end{pmatrix} : x_1,,\dots,x_n,y_1,,\dots,y_n,z\in\mathbb{R} \right\} \subset \mathrm{GL}(n+2,\mathbb{R}).$$

Example 6.1.9 (Filiform algebras of the first kind)**.** The $(n+1)$-dimensional filiform algebra of the first kind is the algebra spanned by $X, Y_1, Y_2, \dots, Y_n$, with only non-trivial relations

$$[X, Y_j] = Y_{j+1}, \qquad \text{for } j = 1, \dots, n-1.$$

It is an n-step nilpotent Lie algebra and can be realized as a matrix algebra considering the matrices of the form:

$$\begin{pmatrix} 0 & x & 0 & \cdot & 0 & y_n \\ & \cdot & \ddots & & & \vdots \\ & & \cdot & \ddots & & \vdots \\ & & & \cdot & x & y_2 \\ & & & & \cdot & y_1 \\ 0 & & & & & 0 \end{pmatrix}.$$

Example 6.1.10 (Strictly upper triangular matrix algebras)**.** The algebra of strictly upper triangular $n\times n$ matrices is an $(n-1)$-step nilpotent Lie algebra of dimension $n(n-1)/2$, and its center is one-dimensional. Namely, let

$$\mathfrak{g} = \mathfrak{n}_n := \left\{ \begin{pmatrix} 0 & * & * \\ 0 & \ddots & * \\ 0 & 0 & 0 \end{pmatrix} \right\} \subset \mathfrak{gl}(n,\mathbb{R})$$

and

$$G = N_n := \left\{ \begin{pmatrix} 1 & * & * \\ 0 & \ddots & * \\ 0 & 0 & 1 \end{pmatrix} \right\} \subset \mathrm{GL}(n,\mathbb{R})$$

So $\mathfrak{n}_n$ is the Lie algebra of N_n and is nilpotent of step $(n-1)$.

Example 6.1.11 (Free-nilpotent algebras)**.** The free nilpotent Lie algebra of step k and rank n (or on n generators) is defined to be the quotient algebra $\mathfrak{f}_n/\mathfrak{f}^{(k+1)}$, where $\mathfrak{f}_n$ is the free Lie algebra on n generators. It is not hard to see that it is finite-dimensional.

For example the Lie algebra of rank 2 and step 3 is given by the diagram

$$\begin{tikzcd} X \arrow[dr] \arrow[dd] & & Y \arrow[dl] \arrow[dd] \\ & Z \arrow[dl] \arrow[dr] & \\ U & & V, \end{tikzcd}$$

which has to be read as $[X, Y] = Z$, $[X, Z] = U$, and $[Z, Y] = V$.

Exercise 6.1.12. If $\mathfrak{g}$ a step-2 Lie algebra, then

$$x \cdot y = x + y + \frac{1}{2}[x, y],$$

defines a group structure on $\mathfrak{g}$. ■

Exercise 6.1.13. For $n = 2$, $\Lambda^1(\mathbb{R}^2) \times \Lambda^2(\mathbb{R}^2)$ gives the Heisenberg Lie algebra. ■

6.1.2 Nilpotent and unipotent transformations

Definition 6.1.14. Let V be a vector space. We say that

1. $A \in \mathfrak{gl}(V)$ is a *nilpotent transformation* if there is $d \in \mathbb{N}$ such that $A^d \equiv 0$.
2. $B \in \mathfrak{gl}(V)$ is a *unipotent transformation* if $B - \mathbb{I}$ is nilpotent.

Exercise 6.1.15. Let $A, B \in \mathfrak{gl}(V)$. Assume that A, B commute then

i. if $A, B \in \mathfrak{gl}(V)$ are nilpotent transformations, then $A + B$ is so too.

ii. if $A, B \in \mathfrak{gl}(V)$ are nilpotent transformations, then AB is so too.

■

Proposition 6.1.16. *Let V be a finite dimensional vector space. If $x \in \mathfrak{gl}(V)$ is a nilpotent transformation then ad_x is a nilpotent transformation. Here $\mathrm{ad}_x \in \mathfrak{gl}(\mathfrak{gl}(V))$ is defined as*

$$\begin{aligned} \mathrm{ad}_x : \mathfrak{gl}(V) &\to \mathfrak{gl}(V) \\ y &\mapsto \mathrm{ad}_x(y) := [x, y] = xy - yx. \end{aligned} \tag{6.1.17}$$

Proof. Let $d \in \mathbb{N}$ such that $x^d \equiv \mathbf{0}$ as elements in $\mathfrak{gl}(V)$, i.e., $x^d(v) = 0$ for every $v \in V$ as elements in V. We need to prove that there is $k \in \mathbb{N}$ such that

$$(ad_x)^k(y) = \mathbf{0}, \quad \forall y \in \mathfrak{gl}(V).$$

This is equivalent to ask that, for some $k \in \mathbb{N}$

$$(ad_x)^k(y)(v) = 0, \quad \forall y \in \mathfrak{gl}(V), \forall v \in V. \tag{6.1.18}$$

We begin noting that for $k = 1$ it would be $[x, y]v = 0$, i.e., $xyv - yxv = 0$. On the other hand, for $k = 2$ it would be $[x[x, y]] = 0$, i.e.,

$$0 = (x[x, y] - [x, y]x)v = (x(xy - yx) - (xy - yx)x)v.$$

Hence for $k = 2d$ it would have that $(ad_x)^k(y)(v)$ is the sum of terms of the form $x^\alpha y x^\beta v$ with either α or β larger or equal to d because of $\alpha + 1 + \beta = k + 1$. Finally, since $x^d w = 0$ for every $w \in V$ we have that $(ad_x)^{2d} = 0$, as desired. □

Proposition 6.1.19. *Let V be a finite dimensional vector space. If $A \in \mathfrak{gl}(V)$ is a nilpotent transformation then*

1. *there is a basis of V such that*
$$A = \begin{pmatrix} 0 & * & * \\ 0 & 0 & * \\ 0 & 0 & 0 \end{pmatrix}$$
is strictly upper triangular.

2. *0 is the only eigenvalue of A.*

Proof. (ii). If there are $\lambda \in \mathbb{R}$ and $v \neq 0$ such that $Av = \lambda v$, then $0 = A^d v = \lambda^d v$ with d nilpotency degree of A. Hence, $\lambda^d = 0$ and so $\lambda = 0$.

(i). We use real Jordan form. In general, put A as real Jordan form, i.e., the matrix with diagonal the matrices $J_1, \ldots, J_k$ where where each block J_i is a square matrix of the form

$$J_i = \begin{pmatrix} \gamma_1 & 1 & & & 0 \\ & \gamma_2 & 1 & & \\ & & \gamma_3 & 1 & \\ & & \cdots & & \\ 0 & & & \gamma_k & 1 \end{pmatrix}$$

□

6.1.3 Engel's Theorem

In this section we prove Engel's Theorem. Firstly, we explain what means a representation of Lie algebra, then we prove the statement for linear Lie algebra and finally, the desired theorem.

Definition 6.1.20. A representation of Lie algebra on a vector space V is a Lie algebra homomorphism from $\mathfrak{g}$ to $\mathfrak{gl}(V)$. Equivalently, it is a $\mathfrak{g}$-module structure on V, i.e., the map $\mathfrak{g} \times V \to V, (x, v) \mapsto xv$ is bilinear and it holds

$$[x, y]v = x(yv) - y(xv), \quad \forall x, y \in \mathfrak{g}, \forall v \in V.$$

Notice that

1. a representation of Lie algebra might not be injective.
2. ad is a representation because of Jacobi identity.
3. the kernel of ad representation is exactly the center $Z(\mathfrak{g})$ of $\mathfrak{g}$ and so $\mathrm{ad}(\mathfrak{g}) \simeq \mathfrak{g}/Z(\mathfrak{g})$.
4. if $\mathfrak{h} < \mathfrak{g}$ then we have a representation of $\mathfrak{h}$ on $\mathfrak{g}/\mathfrak{h}$ as

   $$\mathrm{ad}_{\mathfrak{g}/\mathfrak{h}} \to \mathfrak{gl}(\mathfrak{g}/\mathfrak{h}),$$

   defined as

   $$\mathrm{ad}_{\mathfrak{g}/\mathfrak{h}}(h)(y + \mathfrak{h}) := [h, y] + \mathfrak{h}, \quad \forall h \in \mathfrak{h}, \forall y \in \mathfrak{g}.$$

5. if $h \in \mathfrak{h}$, then $[h, \mathfrak{h}] \subseteq \mathfrak{h}$, and so $[h, y + \mathfrak{h}] + \mathfrak{h} = [h, y] + \mathfrak{h}$.

Remark 6.1.21. If $\mathrm{ad}(x) : \mathfrak{g} \to \mathfrak{g}$ is nilpotent transformation with $x \in \mathfrak{h} < \mathfrak{g}$ then $\mathrm{ad}_{\mathfrak{g}/\mathfrak{h}}(x) : \mathfrak{g}/\mathfrak{h} \to \mathfrak{g}/\mathfrak{h}$ is nilpotent. Indeed, $(\mathrm{ad}_{\mathfrak{g}/\mathfrak{h}}(x))^k(y + \mathfrak{h}) = (\mathrm{ad}(x))^k(y) + \mathfrak{h}$.

Engel's Theorem on linear Lie algebras

Theorem 6.1.22 (Engel's Theorem on linear Lie algebra)**.** *Let $V \neq \{0\}$ be a finite dimensional vector space and $\mathfrak{g} < \mathfrak{gl}(V)$. Assume that every $x \in \mathfrak{g}$ is a nilpotent transformation. Then*

1. *there is $v_0 \in V - \{0\}$ such that $\mathfrak{g}(v_0) = \{0\}$.*
2. *there is a flag $\mathcal{F} = (V_0, \dots, V_n)$ for V with $\dim(V_k) = k$ and $\mathfrak{g} = \mathfrak{g}_{nil}(\mathcal{F})$.*

Consequently,

- *there is a basis for V with respect to which elements of $\mathfrak{g}$ are strictly upper triangular matrices;*
- *$\mathfrak{g}$ is a nilpotent Lie algebra.*

Proof. We start proving part (1). It is proved by induction on dimension of $\mathfrak{g}$.

If $\dim\mathfrak{g} = 0$, then any $v_0 \in V \setminus \{0\}$ works.

If $\dim\mathfrak{g} = 1$, then any $\mathfrak{g} = \mathbb{R}x$ with $x \in \mathfrak{gl}(V)$. Because x is nilpotent, then it has 0 as (only) eigenvalue. (this is because of Jordan decomposition). Thus there is $v \neq 0$ such that $xv = 0$ and so $txv = 0$.

Assume that $\dim\mathfrak{g} > 1$. Pick a subalgebra $\mathfrak{h} < \mathfrak{g}$. with $\mathfrak{h} \neq \mathfrak{g}$ of maximal dimension. By assumption, every $x \in \mathfrak{g} \subset \mathfrak{gl}(V)$ is nilpotent. Hence, as previously seen $\mathrm{ad}(x) \in \mathfrak{gl}(V)$ is nilpotent (see Proposition 6.1.16) and as previously seen $\mathrm{ad}_{\mathfrak{g}/\mathfrak{h}}(x) \in \mathfrak{gl}(\mathfrak{g}/\mathfrak{h})$ is nilpotent.

Now, by $\dim(\mathrm{ad}_{\mathfrak{g}/\mathfrak{h}}(\mathfrak{h})) \leq \dim\mathfrak{h} < \dim\mathfrak{g}$, we apply the induction to the representation $\mathrm{ad}_{\mathfrak{g}/\mathfrak{h}} : \mathfrak{h} \to \mathfrak{gl}(\mathfrak{g}/\mathfrak{h})$. Then there is a non-zero element in $\mathfrak{g}/\mathfrak{h}$ say $x_0 + \mathfrak{h}$ with $x_0 \in \mathfrak{g}$ and $x_0 \notin \mathfrak{h}$ such that $\mathrm{ad}_{\mathfrak{g}/\mathfrak{h}}(\mathfrak{h})(x_0 + \mathfrak{h}) = \mathfrak{h}$. In other words, $[\mathfrak{h}, x_0] \subset \mathfrak{h}$. Hence, the vector space $\mathfrak{h} + \mathbb{R}x_0$ is a subalgebra of $\mathfrak{g}$. By maximality of $\mathfrak{h}$, we infer $\mathfrak{h} + \mathbb{R}x_0 = \mathfrak{g}$ and so $[\mathfrak{g}, \mathfrak{h}] \subset \mathfrak{h}$, i.e., $\mathfrak{h}$ is an ideal.

Next, we look at the representation $\mathfrak{h} \to \mathfrak{gl}(V)$. Recall that $\dim(\mathfrak{h}) < \dim(\mathfrak{g})$, by induction there is $v \in V - \{0\}$ such that $\mathfrak{h}(v) = \{0\}$ and so we can consider the non trivial subspace

$$V_0 := \{v \in V \,:\, \mathfrak{h}(v) = \{0\}\}.$$

We want to apply the base induction in dimension 1 to the representation $\mathbb{R}x_0 \to \mathfrak{gl}(V_0)$ defined as $tx_0 \mapsto (tx_0)_{|V_0}$ noting that

$$\mathfrak{g}(V_0) \subset V_0.$$

This follows from the simply fact that for every $x \in \mathfrak{g}, v \in V_0$ and every $y \in \mathfrak{h}$ we have that

$$yxv = xyv - [x, y]v \in x\mathfrak{h}v + \mathfrak{h}v = \{0\}.$$

Here we used $[x, y] \in [\mathfrak{g}, \mathfrak{h}] \subset \mathfrak{h}$ and $yv = 0$. Then by base induction in dimension 1, there is $v_0 \in V_0 - \{0\}$ such that $x_0v_0 = 0$ and so putting all together we have

$$\mathfrak{g}(v_0) = hv_0 + \mathbb{R}x_0v_0 = \{0\}.$$

This prove the part (1).

Now we show the point (2). By (1), take $v_1 \in V - \{0\}$ such that $\mathfrak{g}(v_1) = 0$ and we consider $V_1 := \mathbb{R}v_1$. Then, the map $\alpha : \mathfrak{g} \to \mathfrak{gl}(V/V_1)$ defined as $\alpha(x)(v + V_1) := x(v) + V_1$ is well defined and gives a representation of $\mathfrak{g}$ on V/V_1. Still $\alpha(\mathfrak{g})$ consists on nilpotent transformation. By induction on dimension of V, we have that V/V_1 posses a complete flag $\mathcal{F}_1 := (W_0, \ldots, W_{n-1})$ with $\alpha(\mathfrak{g}) \subseteq \mathfrak{g}_{nil}(\mathcal{F}_1)$. Then $\{0\}$ together with the preimage of the flag $\mathcal{F}_1$ in V is a complete flag $\mathcal{F}$ in V with $\mathfrak{g} \leq \mathfrak{g}_{nil}(\mathcal{F})$. The consequences at the end of the statement of theorem are immediate. □

Engel's characterization theorem for nilpotent Lie algebra

Theorem 6.1.23 (Engel's theorem). *Let $\mathfrak{g}$ be a finite dimensional Lie algebra. TFAE*

1. *$\mathfrak{g}$ is nilpotent;*

2. *for every $x \in \mathfrak{g}$ ad_x is nilpotent.*

Proof. 1. ⇒ 2. It has already been proved.

2. ⇒ 1. Considering $\mathrm{ad}(\mathfrak{g}) = \{\mathrm{ad}_x : x \in \mathfrak{g}\} \subseteq \mathfrak{gl}(\mathfrak{g})$, we have that $\mathrm{ad}(\mathfrak{g}) \cong \mathfrak{g}/Z(\mathfrak{g})$. On the other hand, $\mathrm{ad}(\mathfrak{g})$ is a Lie algebra of nilpotent transformations of $\mathfrak{g}$ (by assumption). By Engel's Theorem on linear Lie algebras, we have that $\mathrm{ad}(\mathfrak{g})$ is nilpotent. Finally, by a previous proposition (the second consequence) we get that $\mathfrak{g}$ is nilpotent, as desired. □

6.1.4 The general Birkhoff-Embedding Theorem

Theorem 6.1.24 (Birkhoff-Embedding Theorem). *Let $\mathfrak{g}$ be a nilpotent Lie algebra (over $\mathbb{R}$). Then there are a finite dimensional vector space V and an injective homeomorphism $i : \mathfrak{g} \to \mathfrak{gl}(V)$ such that $i(x)$ is a nilpotent transformation.*

We will prove this theorem in a special nilpotent Lie algebra, i.e., Carnot algebra. Before of this, we notice some basic facts:

1. This general theorem relies the construction of the universal enveloping algebra and the Poincaré-Birkhoff-Witt Theorem.

2. More generally, there is Ado's Theorem: every finite dimensional Lie algebras has an injective finite dimensional representation whose restriction to the maximal nilpotent ideal is nilpotent.

6.2 Gradings and stratifications

Carnot groups will have Lie algebras that have very special structures, called stratifications. For a Lie algebra $\mathfrak{g}$, an *s-step stratification of* $\mathfrak{g}$ is a direct sum decomposition

$$\mathfrak{g} = V_1 \oplus V_2 \oplus \cdots \oplus V_s$$

of $\mathfrak{g}$ with the property that

$$V_s \neq \{0\} \qquad \text{and } [V_1, V_j] = V_{j+1}, \quad \forall j = 1, \ldots, s, \tag{6.2.1}$$

where we set $V_{s+1} = \{0\}$. It is useful to see stratifications as a special type of gradings. Hence, we begin with this broader concept.

6.2.1 Graded vector spaces and graded Lie algebras

We start with graded vector spaces.

Definition 6.2.2 (Grading for a vector space)**.** Let A be an abelian group (for instance, $\mathbb{Z}$ or $\mathbb{R}$) and let V be a vector space. A *(linear) grading* of V over A is a collection of subspaces $(V_a)_{a\in A}$ of vector subspaces of V such that

$$V = \oplus_{a\in A} V_a.$$

This means that $V = \mathrm{span}\{V_a : a \in A\}$ and for every $a, a' \in A$ with $a \neq a'$ we have that $V_a \cap V_{a'} = \{0\}$.

When a grading of V over A is fixed, we shall say that V is an *A-graded vector space.* If the grading is such that $A < \mathbb{R}$ and

$$V_a \neq \{0\} \Rightarrow a > 0,$$

then V is said *positively graded.* Given a grading $(V_a)_{a\in A}$ and $a \in A$, elements in V_a are said to have *degree* a. Every V_a is called *layer.*

Next we consider Lie algebras, for which we consider a more restrictive notion of grading. We stress that a Lie algebra $\mathfrak{g}$ is a vector space with the additional structure of Lie bracket.

Definition 6.2.3 (Compatible linear grading)**.** A *compatible linear grading* on $\mathfrak{g}$ is a linear decomposition of $\mathfrak{g}$ in vector subspaces $V_1, V_2, \ldots$ such that

$$\mathfrak{g} = \bigoplus_{i=1}^{\infty} V_i \;\text{ and }\; \mathfrak{g}^{(i)} = \mathfrak{g}^{i+1} \oplus V_i, \quad \forall i = 1, 2, ... \tag{6.2.4}$$

A compatible linear grading is a particular $\mathbb{Z}$-grading with a mild interaction with the Lie algebra structure. Clearly, every nilpotent Lie algebra admits a compatible linear grading (Exercise). Carnot algebras have a stronger property:

Definition 6.2.5 (Lie algebra grading)**.** Forn an Abelian group A and a Lie algebra $\mathfrak{g}$, a *grading* of $\mathfrak{g}$ (*as a Lie algebra*) over A, or a *Lie algebra A-grading*, is a linear grading $(V_a)_{a\in A}$ of $\mathfrak{g}$ as vector space with the extra requirement that

$$[V_a, V_b] \subseteq V_{a+b}, \quad \forall a, b \in A.$$

6.2.2 Stratified Lie algebras

We shall focus on a very specific type of grading: stratifications. There are various definition for them, see **??**.

Definition 6.2.6 (Stratification of $\mathfrak{g}$)**.** If $\mathfrak{g}$ is a Lie algebra and $A < \mathbb{Z}$, a grading $(V_a)_{a\in A}$ of $\mathfrak{g}$ as a Lie algebra is called a *stratification* of $\mathfrak{g}$ if the smallest Lie subalgebra of $\mathfrak{g}$ containing V_1 is $\mathfrak{g}$. The maximal a for which $V_a \neq \{0\}$ is called the *step* of the stratification. A Lie algebra is *stratifiable* if it admits a stratification. When one fixes a stratification of a stratifiable Lie algebra $\mathfrak{g}$ we say that $\mathfrak{g}$ is *stratified*, or *Carnot algebra*.

We rephrase de definition saying that a stratification of a Lie algebra $\mathfrak{g}$ is a $\mathbb{Z}$-grading for which $\mathfrak{g}$ is Lie generated by the elements of degree 1. Equivalently, this means that there is a direct-sum decomposition $\mathfrak{g} = V_1 \oplus \cdots \oplus V_s$ for which

$$[V_1, V_j] = V_{j+1}, \quad \forall j = 1, \dots, s \text{ with } V_{s+1} = \{0\}.$$

Example 6.2.7. 6.2.7.i. An Abelian Lie algebra $\mathfrak{g}$ admits a 1-step stratification with $V_1 = \mathfrak{g}$.

6.2.7.ii. Let $\mathfrak{g}$ be the Heisenberg Lie algebra spanned by X, Y, Z with relation $[X, Y] = Z$. Then $V_1 := \mathrm{span}\{X, Y\}$ and $V_2 := \mathrm{span}\{Z\}$ form a 2-step stratification.

Exercise 6.2.8. Show that a stratification is completely determined by V_1. ■

Remark 6.2.9. The following **non invertible** implications hold for a Lie algebra:

$$\text{Carnot algebras} \quad \overset{\not\Leftarrow}{\Rightarrow} \quad \text{positively graded} \quad \overset{\not\Leftarrow}{\Rightarrow} \quad \text{nilpotent}.$$

Uniqueness of stratifications

In the next proposition we prove that every two stratifications on the same (stratifiable) Lie algebra differ by an automorphism.

Proposition 6.2.10 (Uniqueness of stratifications)**.** *Let $\mathfrak{g}$ be a stratifiable Lie algebra with two stratifications,*

$$V_1 \oplus \cdots \oplus V_s = \mathfrak{g} = W_1 \oplus \cdots \oplus W_t.$$

Then:

1. $s = t$,
2. $V_k \oplus \cdots \oplus V_s = W_k \oplus \cdots \oplus W_s$ *for all* k
3. *there is a Lie algebra automorphism* $A : \mathfrak{g} \to \mathfrak{g}$ *such that* $A(V_i) = W_i$ *for all* i.

Proof. The first two points are directly implied by lemma 6.6.9.

We have $\mathfrak{g}^{(k)} = V_k \oplus \cdots \oplus V_s = W_k \oplus \cdots \oplus W_s$. Then the quotient mappings $\pi_k : \mathfrak{g}^{(k)} \to \mathfrak{g}^{(k)}/\mathfrak{g}^{(k+1)}$ induces linear isomorphisms $\pi_k|_{V_k} : V_k \to \mathfrak{g}^{(k)}/\mathfrak{g}^{(k+1)}$ and $\pi_k|_{W_k} : W_k \to \mathfrak{g}^{(k)}/\mathfrak{g}^{(k+1)}$, by a dimension argument.

For $v \in V_k$ define $A(v) := (\pi_k|_{W_k})^{-1} \circ \pi_k|_{V_k}(v)$. Notice that for $v \in V_k$ and $w \in W_k$ we have

$$A(v) = w \quad \iff \quad v - w \in \mathfrak{g}^{(k+1)}.$$

Extend A to a linear map $A : \mathfrak{g} \to \mathfrak{g}$. This is clearly a linear isomorphism. We need now to show that A is a Lie algebra homomorphism, i.e., $[Aa, Ab] = A([a, b])$ for all $a, b \in \mathfrak{g}$.

Let $a = \sum_{i=1}^s a_i$ and $b = \sum_{i=1}^s b_i$ with $a_i, b_i \in V_i$. Then

$$A([a, b]) = \sum_{i=1}^s \sum_{j=1}^s A([a_i, b_j])$$
$$[Aa, Ab] = \sum_{i=1}^s \sum_{j=1}^s [Aa_i, Ab_j],$$

therefore we can just prove $A([a_i, b_j]) = [Aa_i, Ab_j]$ for $a_i \in V_i$ and $b_j \in W_j$.

Notice that $A([a_i, b_j])$ and $[Aa_i, Ab_j]$ both belong to W_{i+j}. Therefore we have $A([a_i, b_j]) = [Aa_i, Ab_j]$ if and only if $[a_i, b_j] - [Aa_i, Ab_j] \in \mathfrak{g}^{(i+j+1)}$. And in fact

$$[a_i, b_j] - [Aa_i, Ab_j] = [a_i - Aa_i, b_j] - [Aa_i, Ab_j - b_j] \in \mathfrak{g}^{(i+j+1)}$$

because, on one hand, $a_i - Aa_i \in \mathfrak{g}^{(i+1)}$ and $b_j \in W_j$, so $[a_i - Aa_i, b_j] \in \mathfrak{g}^{(i+j+1)}$, on the other hand, $Aa_i \in W_i$ and $Ab_j - b_j \in \mathfrak{g}^{(j+1)}$, so $[Aa_i, Ab_j - b_j] \in \mathfrak{g}^{(i+j+1)}$. This concludes the proof. $\square$

Induced grading on $\mathfrak{gl}(V)$

In the next definition we consider subspaces of the linear transformations on vector space equipped with a linear grading that will for a grading.

Definition 6.2.11 ($\mathfrak{gl}(V)_a$ and $\mathfrak{gl}(V)_{>0}$)**.** Let $(V_a)_{a\in A}$ be a grading of a vector space V over an abelian group A. For every $a \in A$ we define

$$\mathfrak{gl}(V)_a := \{M \in \mathfrak{gl}(V) \,:\, M(V_b) \subseteq V_{b+a}, \forall a, b \in A\},$$

and if A has an ordering (for instance, if $A < \mathbb{R}$) we define

$$\mathfrak{gl}(V)_{>0} := \oplus_{a>0}\, \mathfrak{gl}(V)_a.$$

The collection $(\mathfrak{gl}(V)_a)_{a\in A}$ form an A-grading of $\mathfrak{gl}(V)$ as a Lie algebra, Moreover, if $\mathfrak{g}$ is a Carnot algebra, then $\mathfrak{gl}(V)_{>0}$ is a Carnot algebra.

Proposition 6.2.12. *Let V be a finite dimensional vector space with a linear grading $(V_a)_{a\in A}$ over an abelian group A. Then,*

1. *$(\mathfrak{gl}(V)_a)_{a\in A}$ is an A-grading of $\mathfrak{gl}(V)$ as a Lie algebra.*
2. *If $A < \mathbb{R}$ then $\mathfrak{gl}(V)_{>0}$ is a Lie subalgebra of nilpotent transformation.*
3. *If $A = \mathbb{Z}$ and there is $\bar{a} \in A$ such that*

$$V_b \neq \{0\} \quad \Leftrightarrow \quad b \in \mathbb{Z} \cap [1, \bar{a}]$$

then $\mathfrak{gl}(V)_{>0}$ is a Carnot algebra.

Proof. (1). Fix $X_1, \ldots, X_n$ a basis of V adapted to the direct-sum decomposition $V = \oplus_{a\in A} V_a$. So for every $i = 1, \ldots, n$ there is $a_i \in A$ such that $X_i \in V_{a_i}$. For each $i, j \in \{1, \ldots, n\}$, let $E^i_j \in \mathfrak{gl}(V)$ such that

i) $E^i_j(X_i) = X_j$;

ii) $E^i_j(X_k) = 0$, for every $k \neq i$.

Consequently, the elements $(E^i_j)_{i,j=1,\ldots,n}$ form a basis of $\mathfrak{gl}(V)$. Moreover, every E^i_j is such that $E^i_j(V_{a_i}) \subseteq V_{a_j}$ and for every $a \neq a_j$ we have that $E^i_j(V_a) = \{0\} \subseteq V_{a+(a_j-a_i)}$, i.e., $E^i_j \in \mathfrak{gl}(V)_{a_j-a_i}$. Therefore,

$$\oplus_{a\in A}\, \mathfrak{gl}(V)_a = \mathfrak{gl}(V).$$

Finally, notice that if $M_1 \in \mathfrak{gl}(V)_a, M_2 \in \mathfrak{gl}(V)_b$ then

$$M_1 M_2, M_2 M_1 \in \mathfrak{gl}(V)_{a+b},$$

and so $[\mathfrak{gl}(V)_a, \mathfrak{gl}(V)_b] \subseteq \mathfrak{gl}(V)_{a+b}$. This complete the proof of point (1).

(2). Clearly, $\mathfrak{gl}(V)_{>0}$ is a Lie subalgebra. If $(V'_a)_{a\in\mathbb{R}}$ is a Lie algebra grading and $a, b > 0$ then $a + b > 0$ and so $[V'_a, V'_b] = V'_{a+b} \subseteq V'_{>0}$. Moreover, assuming $a_i \leq a_j$ for every $i \leq j$ and defining

$$W_i := \operatorname{span}\{X_{n-i+1}, \dots, X_n\},$$

then $(W_i)_{i=1,\dots,n}$ defines a flag $\mathcal{F}$ such that $\mathfrak{gl}(V)_{>0} \subseteq \mathfrak{gl}_{nil}(\mathcal{F})$, see **??**.

(3). We want to prove that if $V_1 \oplus \cdots \oplus V_s$ with $V_j \neq \{0\}$ for every $j = 1, \dots, s$ then $\mathfrak{gl}(V)_{>0}$ is generated by $\mathfrak{gl}(V)_1$.

By induction, we shall prove that

$$[\mathfrak{gl}(V)_1, \mathfrak{gl}(V)_k] = \mathfrak{gl}(V)_{k+1}, \tag{6.2.13}$$

for every $k \in \mathbb{Z}$ with $k \geq 1$. It is enough to prove that for every i, j such that $a_j - a_i = k + 1$ (i.e., $E^i_j \in \mathfrak{gl}(V)_{k+1}$) we get that

$$E^i_j \in [\mathfrak{gl}(V)_1, \mathfrak{gl}(V)_k].$$

Since $X_i \in V_{a_i}, X_j \in V_{a_j}$ and $a_i - a_j = k + 1 \neq 0$ we have that $1 \leq a_i < a_j \leq \bar{a}$. Moreover, by assumption, there is a basis element X_ℓ with $a_\ell = a_i + 1$. Then we claim that

$$-E^i_j = [E^i_\ell, E^\ell_j] = E^i_\ell E^\ell_j - E^\ell_j E^i_\ell.$$

Indeed, $E^i_\ell E^\ell_j = 0$ and $E^\ell_j E^i_\ell = E^i_j$, recall that $a_i < a_\ell \leq a_j$. Since $E^i_\ell \in \mathfrak{gl}(V)_1$ and $E^\ell_j \in \mathfrak{gl}(V)_k$ we proved (6.2.13), as desired. □

6.2.3 Dilation structures*

Definition 6.2.14 (Dilations on positively graded algebras)**.** Let $(V_a)_{a\in\mathbb{R}_{>0}}$ be a positive grading of a Lie algebra $\mathfrak{g}$. For every $\lambda \in \mathbb{R}$, the *(inhomogeneous) dilation on* $\mathfrak{g}$ *(relative to the grading) of factor* λ is the linear map $\delta_\lambda : \mathfrak{g} \to \mathfrak{g}$ such that

$$\delta_\lambda v = \lambda^j v, \qquad \forall v \in V_j.$$

Lemma 6.2.15. *For* $\lambda > 0$*, the dilation* $\delta_\lambda : \mathfrak{g} \to \mathfrak{g}$ *is a Lie algebra automorphism, i.e., it is a linear bijection and*

$$\delta_\lambda([X, Y]) = [\delta_\lambda X, \delta_\lambda Y], \qquad \forall X, Y \in \mathfrak{g}.$$

Proof. Take $X, Y \in \mathfrak{g}$ and decompose them as $X = \sum_{i=1}^s X_i$, $Y = \sum_{i=1}^s Y_i$, with $X_i, Y_i \in V_i$. Since $[X_i, Y_j] \in [V_i, V_j] \subset V_{i+j}$, we get

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

Moreover, δ_λ is invertible with inverse $\delta_{1/\lambda}$. □

The map $(\mathbb{R}_{>0}, \cdot) \to \mathrm{Aut}_{\mathrm{Lie}}(\mathfrak{g})$ is a one-parameter subgroup:

$$\delta_\lambda \circ \delta_\mu = \delta_{\lambda\mu}, \qquad \forall \lambda, \mu \in \mathbb{R}. \tag{6.2.16}$$

Lemma 6.2.17. *If $\mathfrak{g} = V_1 \oplus \cdots \oplus V_s$ is a stratified Lie algebra then the dilation $\delta_\lambda : \mathfrak{g} \to \mathfrak{g}$ has determinant equal to λ^Q with*

$$Q = \sum_{j=1}^s j \cdot \dim(V_j).$$

Proof. Fix a basis $X_1, \ldots, X_n$ adapted to the stratification, i.e., for every i there is a j such that $X_i \in V_j$. Then in this basis δ_λ is represented by the diagonal matrix with diagonal

$$(\underbrace{\lambda, \ldots, \lambda}_{\dim V_1}, \underbrace{\lambda^2, \ldots, \lambda^2}_{\dim V_2}, \ldots, \underbrace{\lambda^s, \ldots, \lambda^s}_{\dim V_s}).$$

Hence the determinant is $\lambda^{\dim V^1} \cdot (\lambda^2)^{\dim V_2} \cdot \cdots \cdot (\lambda^s)^{\dim V_s} = \lambda^Q$. □

Associated Carnot algebra

Definition 6.2.18 (Associated Carnot algebra)**.** Let $\mathfrak{g}$ be a Lie algebra that is nilpotent of step s. Let $\mathfrak{g}^{(i+1)} := [\mathfrak{g}; \mathfrak{g}^{(i)}]$ be the descending central series of $\mathfrak{g}$. The *associated Carnot algebra* of $\mathfrak{g}$ is the Lie algebra $\mathfrak{g}_\infty$ given by the direct sum decomposition

$$\mathfrak{g}_\infty := \bigoplus_{i=1}^s \mathfrak{g}^{(i)} / \mathfrak{g}^{(i+1)},$$

endowed with the unique Lie bracket $[\cdot, \cdot]_\infty$ that has the property that, if $x \in \mathfrak{g}^{(i)}$ and $y \in \mathfrak{g}^{(j)}$, the bracket is defined, modulo $\mathfrak{g}^{(i+j)}$, as

$$[\bar{x}, \bar{y}]_\infty = \overline{[x, y]}.$$

Lemma 6.2.19. *Let $\mathfrak{g} = (\mathfrak{g}, [\cdot, \cdot])$ be a nilpotent Lie algebra. Consider the dilations $(\delta_\lambda)_{\lambda > 0}$ relative to some compatible linear grading and define the map*

$$[X, Y]_\infty := \lim_{\lambda \to +\infty} \delta_\lambda^{-1} [\delta_\lambda X, \delta_\lambda Y], \qquad \forall X, Y \in \mathfrak{g}. \tag{6.2.20}$$

Then $[\cdot,\cdot]_\infty$ *defines a (possibly different) Lie bracket on* $\mathfrak{g}$*,*

$$[\delta_\lambda X, \delta_\lambda Y]_\infty = \delta_\lambda [X,Y]_\infty, \qquad \forall X, Y \in \mathfrak{g},$$

and $(\mathfrak{g}, [\cdot,\cdot]_\infty)$ *is isomorphic to the associated Carnot algebra of* $(\mathfrak{g}, [\cdot,\cdot])$*.*

Proof. Indeed, since the V_j's are a direct decomposition of $\mathfrak{g}$, it suffices to show (6.2.20) for $X \in V_i$ and $Y \in V_j$, for some i, j. In this case, we have that

$$[X,Y] = Z_{i+j} + Z_{i+j+1} + \ldots + Z_s,$$

for some vectors $Z_k \in V_k$. Hence

$$\begin{aligned}
\delta_\lambda^{-1}[\delta_\lambda X, \delta_\lambda Y] &= \delta_\lambda^{-1}[\lambda^i X, \lambda^j Y] \\
&= \lambda^{i+j}\delta_\lambda^{-1}(Z_{i+j} + Z_{i+j+1} + \ldots + Z_s) \\
&= Z_{i+j} + \lambda^{-1} Z_{i+j+1} + \ldots + \lambda^{i+j-s} Z_s,
\end{aligned}$$

which goes to Z_{i+j}, as $\lambda \to \infty$. The proof of (6.2.20) is concluded by observing that Z_{i+j} is a vector that represent $[X,Y]$ modulo $\mathfrak{g}^{(i+j+1)}$. □

Siebert theorem*

[...]

6.2.4 Birkhoff Theorem for stratified Lie algebras

We plan to prove Birkhoff Theorem for Carnot algebras. We begin with a Carnot algebra $\mathfrak{g}$. Then we perform a semidirect product $\mathfrak{g} \rtimes \mathbb{R}$, on which we naturally put a grading. Consequently, the Lie algebra $\mathfrak{gl}(\mathfrak{g} \rtimes \mathbb{R})_{>0}$ will be a Carnot algebra.

Induced grading on $\mathfrak{g} \rtimes \mathbb{R}$

More generally, let $\mathfrak{g}$ be a $\mathbb{Z}$-graded Lie algebra with grading $(V_m)_{m\in\mathbb{Z}}$. We consider semidirect product $\mathfrak{g} \rtimes \mathbb{R}$ where $1 \in \mathbb{R}$ acts on $\mathfrak{g}$ as the derivation multiplying by m the vectors in V_m. Namely, the Lie bracket on the semidirect product $\mathfrak{g} \rtimes \mathbb{R}$ is

$$[(X,s),(Y,t)] = \left([X,Y] + \sum_{m\in\mathbb{Z}} smY_m - \sum_{m\in\mathbb{Z}} tmX_m, 0\right), \qquad \forall X, Y, \in \mathfrak{g}, \forall s, t \in \mathbb{R}.$$

The Lie algebra $\mathfrak{g} \rtimes \mathbb{R}$ is $\mathbb{Z}$-graded by $(V'_m)_{m\in\mathbb{Z}}$ such that

$$V'_0 := V_0 \times \{0\} \oplus \{0\} \times \mathbb{R}, \tag{6.2.21}$$

$$V'_m := V_m \times \{0\}, \qquad \forall m \neq 0; \tag{6.2.22}$$

see Exercise ??

Moreover, if $\mathfrak{g}$ is a (non-trivial) Carnot algebra, then $\mathfrak{g} \rtimes \mathbb{R}$ has trivial center, see Exercise ??. Clearly, if $\mathfrak{g} = \{0\}$, then $\mathfrak{g} \rtimes \mathbb{R} = \mathbb{R}$, which is Abelian.

Exercise 6.2.23. Let $\mathfrak{g}$ be a $\mathbb{Z}$-graded Lie algebra.

6.2.23.i Show that the Lie algebra $\mathfrak{g} \rtimes \mathbb{R}$ is $\mathbb{Z}$-graded by (6.2.21).

6.2.23.ii Show that if $V_0 = \{0\}$ and $\mathfrak{g} \neq \{0\}$, then $\mathfrak{g} \rtimes \mathbb{R}$ has trivial center.

6.2.23.iii Show that if $\mathfrak{g}$ is Carnot, then the grading of $\mathfrak{g} \rtimes \mathbb{R}$ satisfies for some $\bar{a} \in A$

$$V_b \neq \{0\} \quad \Leftrightarrow \quad b \in \mathbb{Z} \cap [1, \bar{a}]$$

Solution. (1). It is enough to check that

$$[\{0\} \times \mathbb{R}, \{0\} \times \mathbb{R}] \subset V_0', \quad \text{and} \quad [V_n \times \{0\}, \{0\} \times \mathbb{R}] \subset V_n'.$$

In fact, $[(0,s),(0,t)] = (0,0) \in V_0'$ and $[(X,0),(0,t)] = (-tnX, 0) \in V_n'$ if $X \in V_1$.

(2). Let $(X,s) \in Z(\mathfrak{g} \rtimes \mathbb{R})$. Then $(0,0) = [(X,s),(0,1)] = (-\sum_{n\in\mathbb{Z}} nX_n, 0)$. Hence for every $n \neq 0$ we have that $X_n = 0$ and so $X = 0$ and $V_0 = \{0\}$.

Moreover, take $Y \in V_n - \{0\}$ for some $n \in \mathbb{Z} - \{0\}$, then $(0,0) = [(X,0),(Y,0)] = [(0,s),(Y,0)] = (snY, 0)$ and consequently $s = 0$.

(iii). From the definition of V_a', if the non-trivial layers of the gradings of $\mathfrak{g}$ are $V_1, \dots, V_s$, then the non-trivial layers of the gradings of $\mathfrak{g} \rtimes \mathbb{R}$ are $V_1', \dots, V_s'$. ∎

Proof of Birkhoff Theorem for Carnot algebras

We have that, when $\mathfrak{g}$ is a $\mathbb{Z}$-graded Lie algebra, then $\mathfrak{gl}(\mathfrak{g} \rtimes \mathbb{R})_{>0}$ is a Lie algebra of nilpotent transformations and $\mathfrak{g} \rtimes \mathbb{R}$ is graded as in the exercise above. Finally, note that $\mathfrak{g} \simeq \mathfrak{g} \times \{0\} \subset \mathfrak{g} \rtimes \mathbb{R}$.

Theorem 6.2.24 (Birkhoff-Embedding Theorem for Carnot algebras)**.** *Let* $\mathfrak{g}$ *be a Carnot algebra. Then* $\mathrm{ad} : \mathfrak{g} \to \mathfrak{gl}(\mathfrak{g} \rtimes \mathbb{R})_{>0} \subset \mathfrak{gl}(\mathfrak{g} \rtimes \mathbb{R})$ *is an injective Lie algebra homomorphism into the Carnot algebra* $\mathfrak{gl}(\mathfrak{g} \rtimes \mathbb{R})_{>0}$.

Proof. We can assume that $\mathfrak{g} \neq \{0\}$. Since $\mathfrak{g}$ is a Carnot algebra we have that $V_0 = \{0\}$ and so $\mathfrak{g} \rtimes \mathbb{R}$ has trivial center. Consequently, $\mathrm{ad} : \mathfrak{g} \to \mathfrak{gl}(\mathfrak{g} \rtimes \mathbb{R})$ (defined as $x \mapsto (Y \mapsto [X,Y])$) is injective. Moreover, take $n, m \in \mathbb{N}$, $X \in V_n(\mathfrak{g})$, and $(Y,s) \in V_m(\mathfrak{g} \rtimes \mathbb{R})$. On the one hand, if $m = 0$ so $Y = 0$, then $(Y,s) = [(X,0),(0,s)] = (-s+X,0) \in V_n$. Hence, ad_x increased the degree by n. On the other hand, if $m \neq 0$ so $s=0$, then $\mathrm{ad}_x(Y,s) = [(X,0,(Y,0))] = ([X,Y],0) \in V_{m+n}$ and consequently ad_x increased the degree by n. Thus $\mathrm{ad}_x \in \mathfrak{gl}(\mathfrak{g} \rtimes \mathbb{R})_{>0}$ and the proof is complete.

□

6.3 Nilpotent Lie groups*

6.3.1 Examples of nilpotent Lie groups

6.3.2 Exponential and logarithm function

6.3.3 BCH formula

6.3.4 Exponential and Malchev's coordinates

6.3.5 Lie groups with nilpotent Lie algebras

6.4 Structure of nilpotent Lie groups*

6.4.1 Structure of connected nilpotent Lie groups

6.4.2 Subgroups of simply connected nilpotent Lie groups

6.5 Extra*

The BCH formula

The Baker-Campbell-Hausdorff formula allows us to reconstruct every Lie group G locally, with its multiplication law, knowing only the structure of its Lie algebra $\mathfrak{g}$. The Baker-Campbell-Hausdorff formula links Lie groups to Lie algebras, by expressing the logarithm $\log(e^X e^Y)$ of the product of two Lie group elements as a Lie algebra element. The logarithm is by definition the inverse of the exponential, in general it is only locally defined in a neighborhood of the identity, thanks to Proposition 4.2.10. However, for simply connected nilpotent Lie groups logarithm will be global by Theorem 6.5.1.

The general *Baker-Campbell-Hausdorff formula* (BCH formula, for short) is given by:

$$\log(\exp X \exp Y) = \sum_{n>0} \frac{(-1)^{n-1}}{n} \sum_{\substack{r_i+s_i>0 \\ 1\leq i\leq n}} \frac{\left(\mathrm{ad}_X^{r_1} \circ \mathrm{ad}_Y^{s_1} \circ \mathrm{ad}_X^{r_2} \circ \mathrm{ad}_Y^{s_2} \ldots \circ \mathrm{ad}_X^{r_n} \circ \mathrm{ad}_Y^{s_n-1}\right)(Y)}{r_1! s_1! \cdots r_n! s_n! \ \sum_{i=1}^n (r_i+s_i)},$$

where $\operatorname{ad}_X Y = [X, Y]$, see (**??**). Thus

$$\left(\operatorname{ad}_X^{r_1} \circ \operatorname{ad}_Y^{s_1} \circ \operatorname{ad}_X^{r_2} \circ \operatorname{ad}_Y^{s_2} \ldots \circ \operatorname{ad}_X^{r_n} \circ \operatorname{ad}_Y^{s_n - 1}\right)(Y)$$

$$= [\underbrace{X, [X, \ldots [X,}_{r_1} \underbrace{[Y, [Y, \ldots [Y,}_{s_1} \ldots \underbrace{[X, [X, \ldots [X,}_{r_n} \underbrace{[Y, [Y, \ldots Y}_{s_n}]] \ldots]].$$

The first terms of the series should[2] be

$$\begin{aligned}
\log(\exp X \exp Y) &= X + Y + \frac{1}{2}[X,Y] + \frac{1}{12}[X,[X,Y]] - \frac{1}{12}[Y,[X,Y]] \\
&\quad - \frac{1}{24}[Y,[X,[X,Y]]] \\
&\quad - \frac{1}{720}([[[[X,Y],Y],Y],Y] + [[[[Y,X],X],X],X]) \\
&\quad + \frac{1}{360}([[[[X,Y],Y],Y],X] + [[[[Y,X],X],X],Y]) \\
&\quad + \frac{1}{120}([[[[Y,X],Y],X],Y] + [[[[X,Y],X],Y],X]) + \cdots
\end{aligned}$$

For matrix Lie groups the Baker-Campbell-Hausdorff formula can be obtained formally by solving for Z in $e^Z = e^X e^Y$, using that

$$\log(I + A) = \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n} A^n.$$

Indeed,

$$\begin{aligned}
Z &= \log(I + (e^X e^Y - I)) \\
&= \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n} (e^X e^Y - I)^n \\
&= \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n} \left(\sum_{p_i + q_i > 0, p_i, q_i \geq 0} \frac{X^{p_i} Y^{q_i}}{p_i! q_i!} \right)^n \\
&= \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n} \sum_{p_i + q_i > 0, p_i, q_i \geq 0} \frac{X^{p_1} Y^{q_1} \cdots X^{p_n} Y^{q_n}}{p_1! q_2! \cdots p_n! q_n!}.
\end{aligned}$$

One will get the BCH formula using that $\operatorname{ad}_A B = AB - BA$. Please, let me know if you find a clear and simple calculation of this ending.

Simply connected nilpotent Lie groups

Simply connected Lie groups are uniquely determined by their Lie algebras. Indeed, recall from Corollary 4.1.9 that if two simply connected Lie groups have isomorphic Lie algebras, then they are isomorphic. For nilpotent groups, the exponential map and the BCH formula provide a concrete identification. We will see how one can completely work on the Lie algebra using such coordinates.

[2]This calculations should be double checked!

Theorem 6.5.1 ([CG90, Theorem 1.2.1])**.** *Let G be a connected, simply connected nilpotent Lie group.*

a *The exponential map* $\exp : \mathrm{Lie}(G) \to G$ *is an analytic diffeomorphism.*

b *The Baker-Campbell-Hausdorff Formula holds globally.*

Proof for the case of nilpotent matrix groups. If G is a matrix group of nilpotency step s, then for all $A \in G$

$$\exp(A) = e^A = \sum_{j=0}^{s} \frac{1}{j!} A^j.$$

So exp is a polynomial map.

Its (global) inverse is

$$\log(B) = \sum_{k=1}^{s} \frac{(-1)^{k+1}}{k} (B-I)^k.$$

Also the BCH series is finite and hence polynomial.

Since it coincide on an open neighborhood of 0 with the analytic function $\log(\exp(X)\exp(Y))$, it coincide globally. □

The following facts are consequences of Theorem 6.5.1 and its proof.

Fact 6.5.2. *Every Lie subgroup H of a connected, simply connected nilpotent Lie group G is closed and simply connected.*

Let N_n be the group whose Lie algebra are the strictly upper triangular matrices. Namely, N_n is the group of matrices that are upper triangular and have 1's in the diagonal.

Fact 6.5.3. *Every connected, simply connected nilpotent Lie group has a faithful embedding as a closed subgroup of N_n for some n.*

Good coordinates

One important application of Theorem 6.5.1 involves coordinates on G. Since exp is a diffeomorphism of $\mathfrak{g}$ onto G, we can use it to transfer coordinates from $\mathfrak{g}$ to G. Some authors use exp to identify $\mathfrak{g}$ with G. Then the group multiplication can be calculated by the Baker-Campbell-Hausdorff formula.

Definition 6.5.4 (Exponential coordinates: canonical coordinates of 1^{st} kind)**.** Let $\{X_1, \dots, X_n\}$ be a basis for a nilpotent Lie algebra of a simply connected nilpotent group G. The coordinates given by the map

$$\Phi : \mathbb{R}^n \longrightarrow G$$

$$\Phi(t_1, \ldots, t_n) := \exp(t_1 X_1 + \ldots + t_n X_n)$$

are called *exponential coordinates*. Exponential coordinates are also known as *canonical coordinates of the first kind.*

With exp we are identifying $\mathbb{R}^n$ with Lie(G) and G. Moreover, the group law can be obtained through the BCH formula

$$(s_1, \ldots, s_n) * (t_1, \ldots, t_n) = \log\left(\exp\left(\sum_{j=1}^n s_j X_j\right)\exp\left(\sum_{j=1}^n t_j X_j\right)\right)$$

Definition 6.5.5. Let $\mathfrak{g}$ be a nilpotent Lie algebra. An ordered basis $\{X_1, \ldots, X_n\}$ for $\mathfrak{g}$ is called (strong) Malcev basis if, for each $k \in \{1, \ldots, n\}$, the space

$$\operatorname{span}\{X_1, \ldots, X_k\}$$

is an ideal of $\mathfrak{g}$, i.e.

$$[\mathfrak{g}, \mathfrak{g}_k] \subset \mathfrak{g}_k.$$

In general, a subspace I of a Lie algebra $\mathfrak{g}$ is called an *ideal* of $\mathfrak{g}$ if $[\mathfrak{g}, I] \subseteq I$. By anticommutativity, there is no need of distinction between left and right ideals.

Fact 6.5.6. *In the special class of Carnot groups, see next chapter, the existence of Malcev basis will be a triviality. However, every nilpotent algebra has Malcev basis, see Theorem 1.1.13 in [CG90] and the notes following it.*

Lemma 6.5.7. *If $\{X_1, \ldots, X_n\}$ is a Malcev basis for a nilpotent Lie algebra $\mathfrak{g}$, then its ideals $\mathfrak{g}_k := \operatorname{span}\{X_1, \ldots, X_k\}$ are such that*

$$[\mathfrak{g}, \mathfrak{g}_k] \subseteq \mathfrak{g}_{k-1}. \tag{6.5.8}$$

Proof. By definition of Malcev basis, we have $[\mathfrak{g}, \mathfrak{g}_k] \subseteq \mathfrak{g}_k$ and also $[\mathfrak{g}, \mathfrak{g}_{k-1}] \subseteq \mathfrak{g}_{k-1}$. If the conclusion of the lemma were not true, then there would be some $j \in \{1, \ldots, n\}$ and $a_1, \ldots, a_k$ with $a_k \neq 0$ such that

$$[X_j, X_k] = a_k X_k + \sum_{i=1}^{k-1} a_i X_i.$$

Now we iterate bracketing by X_j, i.e., we iterate the map $\operatorname{ad}_{X_i} = [X_i, \cdot]$. Thus, we get, for some $a_1^{(l)}, \ldots, a_{k-1}^{(l)}$,

$$(\operatorname{ad}^l_{X_j})(X_k) = a^l X_k + \sum_{i=1}^{k-1} a_i^{(l)} X_i,$$

which is never zero and so contradicts the nilpotency of $\mathfrak{g}$. $\square$

Definition 6.5.9 (Malcev coordinates: canonical coordinates of the 2^{nd} kind)**.** Let $\{X_1,\dots,X_n\}$ be a (strong) Malcev basis for a nilpotent Lie algebra. Define the map

$$\Psi : \mathbb{R}^n \to G$$

$$\Psi(s) := \exp(s_1X_1)\cdots\exp(s_nX_n).$$

The coordinate system defined is called *strong Malcev coordinates* or also *canonical coordinates of the second kind*.

If $\{X_1,\dots,X_n\}$ is a Malcev basis for a nilpotent Lie algebra, we can consider both canonical coordinates; we have that the Malcev coordinates are related to the exponential coordinates by a polynomial diffeomorphism whose Jacobian determinant is constantly equal to 1.

Proposition 6.5.10 ([CG90, Proposition 1.2.7])**.** *Let $\{X_1,\dots,X_n\}$ be a Malcev basis for a nilpotent Lie algebra $\mathfrak{g}$. Let $\Psi : \mathbb{R}^n \to G$ the Malcev coordinate system and $\Phi : \mathbb{R}^n \longrightarrow M$ the exponential coordinate system associated to the basis. Then*

(i) *$\Psi(s) = \Phi(P(s))$ where $P : \mathbb{R}^n \to \mathbb{R}^n$ is a polynomial diffeomorphism with polynomial inverse.*

(ii) *writing $P = (P_1,\dots,P,n)$, then $P_j(s) = s_j + \hat{P}(s_{j+1},\dots,s_n)$.*

In other words, we have the relation:

$$\exp(s_1X_1)\cdots\exp(s_nX_n) = \exp(P_1(s)X_1 + \ldots + P_n(s)X_n).$$

Proposition 6.5.11 ([CG90, Proposition 1.2.9])**.** *Assume that G is equipped with either exponential or Malcev coordinates with respect to some basis. For each $g \in G$, the left translation L_g and the right translation R_g are maps whose Jacobian determinants are identically equal to* 1.

Proof. We prove the statement for exponential coordinates and left translations. The case of right translations is similar. For Malcev coordinates it will be true because of they differs from exponential coordinates by a polynomial diffeomorphism whose Jacobian determinant is constantly equal to 1, Proposition 6.5.10.

The proof is based on the BCH formula and (6.5.8). Indeed, we can assume that the basis $\{X_1,\dots,X_n\}$ is a Malcev basis, since linear changes of basis preserve Jacobians. So, let Φ the exponential coordinate system, and L_g the left translation by g. We need to calculate the Jacobian of $\Phi^{-1} \circ L_g \circ \Phi$. Thus we consider the diagram

$$\begin{array}{ccc}
\mathbb{R}^n & \xrightarrow{\quad\quad\Phi\quad\quad} & G \\
 & & \\
(t_1, \dots, t_n) & \longmapsto & \exp(\sum_j t_j X_j) \\
 & & \Big\downarrow {\scriptstyle L_g} \\
(s_1, \dots, s_n) & \longmapsto & g \exp(\sum_j t_j X_j) \\
 & & \\
\mathbb{R}^n & \xrightarrow{\quad\quad\Phi\quad\quad} & G,
\end{array}$$

and we solve the dependence of the s_i's from the t_j's. Since the Malcev coordinates are surjective we can find $u_1, \dots, u_n$ and write

$$g = \exp(u_1 X_1) \dots \exp(u_n X_n).$$

It is enough to consider the case $g = \exp(u_k X_k)$ and then conclude considering compositions. Thus we need to consider the system

$$\exp(\sum_j s_j X_j) = \exp(u_k X_k) \exp(\sum_j t_j X_j).$$

By the BCH formula,

$$\sum_j s_j X_j = u_k X_k + \sum_j t_j X_j + \frac{1}{2}[u_k X_k, \sum_j t_j X_j] + \dots.$$

Since we have chosen a Malcev basis we have the property (6.5.8). Thus a bracket as $[X_k, X_j]$ is only a combination of $\{X_1, \dots, X_{j-1}\}$. In other words, the function s_j is of the form t_j plus a polynomial that does not depend on the variables $t_1, \dots, t_j$. Thus the differential is of the form

$$d(\Phi^{-1} \circ L_g \circ \Phi) = \begin{pmatrix}
1 & * & \dots & \dots & \dots & \dots & * \\
0 & \ddots & \ddots & * & \dots & * & \vdots \\
\cdot & \cdot & 1 & * & \ddots & \vdots & \vdots \\
\cdot & \cdot & 0 & 1 & * & * & \vdots \\
\cdot & \cdot & \cdot & 0 & 1 & \ddots & \vdots \\
\cdot & \cdot & \cdot & \cdot & \cdot & \ddots & * \\
0 & \cdot & \cdot & \cdot & \cdot & 0 & 1
\end{pmatrix}.$$

Thus the Jacobian of a left translation in exponential coordinates with respect to a Malcev basis is 1 at every point. □

Definition 6.5.12. If $P : \mathbb{R}^n \to \mathbb{R}^n$ is a diffeomorphism such that P and P^{-1} have polynomial components, then

$$(s_1, \dots, s_n) \mapsto \exp(P_1(s)X_1 + \dots + P_n(s)X_n)$$

is called *polynomial coordinate map.*

Examples of polynomial coordinate maps are, obviously, exponential and, by Proposition 6.5.10, Malcev coordinate maps.

Exercise 6.5.13. Show that Malcev coordinates are polynomial coordinates. ■

The key observation is that the Jacobian of every polynomial diffeomorphism with polynomial inverse is a polynomial that is invertible inside the polynomial ring, so it is a constant. Thus, changing of coordinates by a polynomial diffeomorphism with polynomial inverse preserves Lebesgue measure preserving maps.

Corollary 6.5.14. *In polynomial coordinates, left translations have Jacobian 1.*

Proof. If P is a polynomial map, then $\mathtt{Jac}(P)$ is a polynomial. If P and P^{-1} are polynomial diffeomorphisms, then $1 = \mathtt{Jac}(\mathrm{Id}) = \mathtt{Jac}(P \circ P^1) = (\mathtt{Jac}(P) \circ P^{-1}) \cdot \mathtt{Jac}(P^{-1})$.

Hence, $\mathtt{Jac}(P)$ and $\mathtt{Jac}(P^{-1})$ are two polynomial whose product is constant. Thus they are constant.

If Φ is an exponential coordinate map, then

$$\begin{aligned}&\mathtt{Jac}(P^{-1} \circ \Phi^{-1} \circ L_g \Phi \circ P)_x = \\ &\qquad = \mathtt{Jac}(P^{-1})_{(\Phi^{-1} \circ L_g \circ P)(x)} \cdot \mathtt{Jac}(\Phi^{-1} \circ L_g \circ \Phi)_{\Phi(x)} \cdot \mathtt{Jac}(\Phi)_x = 1.\end{aligned}$$

□

Remark 6.5.15. If a map $F : \mathbb{R}^n \to \mathbb{R}^n$ has Jacobian 1, then it preserves the Lebesgue n-measure (because of change of variables formula).

Every Lie group, as every locally compact group, has a natural class of measures: the *Haar measures.* A Borel measure μ is called a left-Haar measure if it is left-invariant, i.e., if, for every left translation L_g,

$$((L_g)_\# \mu)(B) := \mu\left(L_g^{-1}(B)\right) = \mu(B), \qquad \text{for all Borel set } B.$$

Similarly, a *right-Haar measure* is a Borel measure that is right invariant. A Borel measure is called *Haar measure* if it is both right and left invariant.

Left-Haar measures, as right-Haar measures, are unique in the following sense.

Fact 6.5.16. *Left-Haar measures and right-Haar measures that are finite and not zero on compact sets with nonempty interior are unique up to multiplication by a constant.*

A consequence of the previous proposition and the last observation above is the following theorem.

Theorem 6.5.17 ([CG90, Theorem 1.2.10])**.** *Let G be an n-dimensional connected, simply connected, and nilpotent Lie group. Any polynomial coordinate map pushes forward the Lebesgue measure on $\mathbb{R}^n$ to a Haar measure on G.*

It is not always true that left-Haar measures are also right-Haar measures, groups with such property are called *unimodular*. However in every nilpotent Lie group Haar measure are both left and right-invariant. Theorem 6.5.17 shows such uniqueness for simply connected nilpotent Lie groups and it is suffices for our cases of interest.

Homogeneous manifolds

This part will probably will omitted in class.

Theorem 6.5.18 ([War83, Theorem 3.58])**.** *Let [...]*

Theorem 6.5.19 ([CG90, Theorem 1.2.12])**.** *Let [...]*

Theorem 6.5.20 ([CG90, Theorem 1.2.13])**.** *Let [...]*

read page 23 [CG90] remark 1 and 3.

6.6 Exercises

For the following exercises, recall that we denote by $\mathfrak{g}^{(k)}$ the k-th element in the lower central series of a Lie algebra $\mathfrak{g}$.

Exercise 6.6.1. Show that, if $\mathfrak{g}^{(i)} = \mathfrak{g}^{(i+1)}$ for some i, then for all $j > i$ $\mathfrak{g}^{(j)} = \mathfrak{g}^{(i)}$. ■

Exercise 6.6.2. Show that $\mathfrak{g}^{(i+1)} \subset \mathfrak{g}^{(i)}$ for all i. ■

Exercise 6.6.3. Show that N_3 is the Heisenberg group. ■

Exercise 6.6.4 (Positively graded algebras are nilpotent)**.** Prove that an algebra that admits a positive grading is nilpotent. ■

Exercise 6.6.5. Let G be a simply connected nilpotent Lie group and let V_1 be a sub-space such that

$$\mathfrak{g} = V_1 \oplus [\mathfrak{g}, \mathfrak{g}].$$

Denote by $\mathfrak{g}^{(i)}$ the i-th term in the lower (or descending) central series of $\mathfrak{g}$, Show first that

$$\mathfrak{g}^{(2)} = [V_1, V_1] + \mathfrak{g}^{(3)}.$$

Then, by induction, show

$$\mathfrak{g}^{(i)} = [V_1, [V_1, [\ldots, [V_1, V_1] \ldots]] + \mathfrak{g}^{(i+1)},$$

where in the above bracket there are i many V_1's. Finally deduce that such a V_1 generates the whole Lie algebra. ■

Exercise 6.6.6. Prove that every 2-step nilpotent Lie algebra is stratifiable. ■

Exercise 6.6.7. Show that for a stratification $[V_2, V_2] \subseteq V_4$.

Solution.

$$\begin{aligned}[V_2, V_2] = [[V_1, V_1], [V_1, V_1]] = \operatorname{span}\{[[X_1, X_2], [X_3, X_4]] : X_i \in V_1\} \subset \\ \overset{\text{(Jacobi)}}{\subset} \operatorname{span}\{[X_1, [X_2, [X_3, X_4]]] : X_i \in V_1\} = [V_1, [V_1, [V_1, V_1]]] = V_4\end{aligned}$$

■

Lemma 6.6.8 (Stratifications give gradings)**.** *Let $\mathfrak{g} = V_1 \oplus \cdots \oplus V_s$ be a stratified Lie algebra of step s, in the sense of* (**??**)*. Then*

$$[V_i, V_j] \subset V_{i+j},$$

for all $i, j = 1, \ldots, s$, where we set $V_k = \{0\}$ for $k > s$.

Proof. The proof is by induction on i. If $i = 1$ we already know that $[V_1, V_j] \subset V_{j+1}$ for all j. Now suppose that $[V_i, V_j] \subset V_{i+j}$ for all j and a fixed i. We shall show that this implies $[V_{i+1}, V_j] \subset V_{i+1+j}$ for all j. Indeed, V_{i+1} is generated by the elements $[v_1, v_i]$ where $v_1 \in V_1$ and $v_i \in V_i$, and for these elements we have for all $v_j \in V_j$ by the Jacobi identity

$$[[v_1, v_i], v_j] = -[[v_i, v_j], v_1] - [[v_j, v_1], v_i],$$

where $[v_i, v_j] \in V_{i+j}$ by the inductive hypothesis and so $-[[v_i, v_j], v_1] = [v_1, [v_i, v_j]] \in [V_1, V_{i+j}] = V_{i+1+j}$, and $-[[v_j, v_1], v_i] = [v_i, [v_j, v_1]] \in [V_i, V_{j+1}] \subset V_{i+1+j}$ by the inductive hypothesis again.

All in all, $[[v_1, v_i], v_j] \in V_{i+1+j}$ and therefore $[V_{i+1}, V_j] \subset V_{i+1+j}$. □

Exercise 6.6.9 (Elements of lower central series in terms of stratifications)**.** Let $\mathfrak{g}$ be a Lie algebra with a step s stratification $\mathfrak{g} = V_1 \oplus \cdots \oplus V_s$. Show that

$$\mathfrak{g}^{(k)} = V_k \oplus \cdots \oplus V_s.$$

In particular, $\mathfrak{g}$ is nilpotent of step s.

Solution. The proof is by induction. For $k = 1$ is trivial. Suppose it is true for k, then

$$\begin{aligned}\mathfrak{g}^{(k+1)} = [\mathfrak{g}, \mathfrak{g}^{(k)}] = [V_1 \oplus \cdots \oplus V_s, V_k \oplus \cdots \oplus V_s] = \\ = \sum_{i=1}^{s}\sum_{j=k}^{s}[V_i, V_j] = \sum_{j=k}^{s}[V_1, V_j] + \sum_{i=2}^{s}\sum_{j=k}^{s}[V_i, V_j] = \\ = V_{k+1} \oplus \cdots \oplus V_s + \sum_{i=2}^{s}\sum_{j=k}^{s}[V_i, V_j] = V_{k+1} \oplus \cdots \oplus V_s\end{aligned}$$

where $\sum_{i=2}^{s}\sum_{j=k}^{s}[V_i, V_j] \subset \sum_{i=2}^{s}\sum_{j=k}^{s} V_{i+j} \subset V_{k+1} \oplus \cdots \oplus V_s$. ■

Remark 6.6.10. Show that every stratifiable Lie algebra is nilpotent. In fact, if $\mathfrak{g}$ admits an s-step stratification, then $\mathfrak{g}$ is s-step nilpotent.

Exercise 6.6.11. [A nilpotent nonstratifiable algebra] Consider the 7-dimensional Lie algebra $\mathfrak{h}$ generated by $X_1, \dots, X_7$ with only nontrivial brackets

$$\begin{gathered}[X_1, X_2] = X_3\\ [X_1, X_3] = 2X_4\\ [X_1, X_4] = 3X_5\\ [X_2, X_3] = X_5\\ [X_1, X_5] = 4X_6\\ [X_2, X_4] = 2X_6\\ [X_1, X_6] = 5X_7\\ [X_2, X_5] = 3X_7\\ [X_3, X_4] = X_7\end{gathered}$$

Show the following facts:

1) it is a Lie algebra

2) it is nilpotent

3) it does not admit any stratification. ■

Exercise 6.6.12. Fix a positive integer $n \geq 7$, and consider the n-dimensional Lie algebra $\mathfrak{h}$ generated by $X_1, \dots, X_n$ with

$$[X_i, X_j] = \begin{cases} (j-i)X_{i+j}, & \text{if } i+j \leq n, \\ 0, & \text{otherwise.} \end{cases}$$

Show the following facts:

1) it is a Lie algebra

2) it is nilpotent

3) it does not admit any stratification.

■

Exercise 6.6.13. Let δ_λ be the dilation of factor λ as defined either at the group level in Definition 8.1.7 or at the algebra level in Definition 6.2.14. Show that $(\delta_\lambda)^{-1} = \delta_{1/\lambda}$. ■

Exercise 6.6.14. Let $\mathfrak{g} = V_1 \oplus \cdots \oplus V_s$ be a stratified algebra. For all $\lambda \geq 0$, let δ_λ be the dilation of factor λ as defined in Definition 6.2.14. Show that

$$\delta_\lambda \left(\sum_{i=1}^{s} v_i \right) := \sum_{i=1}^{s} \lambda^i v_i,$$

where $X = \sum_{i=1}^{s} v_i$ with $v_i \in V_i$, $1 \leq i \leq s$. ■

Exercise 6.6.15. Let $\mathfrak{h}$ be the Heisenberg Lie algebra generated by the vectors X, Y, and Z with only non-trivial relation $[X, Y] = Z$. Show that the decomposition

$$\mathfrak{h} = \operatorname{span}\{X, Y\} \oplus \operatorname{span}\{Z\}$$

is a step 2 stratification. ■

Exercise 6.6.16. Let $g := \mathbb{R} \times \mathfrak{h}$ be the (commutative) product of $\mathbb{R}$ with the (above) Heisenberg Lie algebra $\mathfrak{h}$. Show that

$$g = (\mathbb{R} \times \operatorname{span}\{X, Y\}) \oplus (\{0\} \times \operatorname{span}\{Z\})$$

is a step 2 stratification with center $\mathbb{R} \times \operatorname{span}\{Z\}$ which is stricly bigger than V_2.

■

Exercise 6.6.17. Show that if a Lie algebra $\mathfrak{g}$ has a step s stratification $\mathfrak{g} = V_1 \oplus \cdots \oplus V_s$, then

1. V_s is contained in the center of $\mathfrak{g}$;

2. $V_k \oplus \cdots \oplus V_s$ is normal in $\mathfrak{g}$; $(V_k \oplus \cdots \oplus V_s)/(V_{k+1} \oplus \cdots \oplus V_s)$ is contained in the center of $(V_1 \oplus \cdots \oplus V_s)/(V_{k+1} \oplus \cdots \oplus V_s)$.

■

Exercise 6.6.18. Show that a stratifiable algebra is isomorphic to its associated Carnot algebra (as defined in Definition 6.2.18) ■

Exercise 6.6.19 (Suggested by E. Breuillard)**.** Let $\mathfrak{g}$ be a Lie algebra that admits a grading. Assume that the elements of degree 1, namely V_1, generate $\mathfrak{g}$, as a Lie algebra, then $\mathfrak{g}$ is stratified by $V_1, \ldots, V_s$.

Hint: If the bracket of V_1 with itself were smaller than V_2, then V_1 would not generate, because the Lie subalgebra it generates will not contain all of V_2... ■

Exercise 6.6.20 (A graded nonstratifiable algebra)**.** Let $\mathfrak{g}$ be the algebra from Example 6.6.11. Show that

1) $\mathfrak{g}$ admits a grading. Hint: $V_i = \mathbb{R}X_i$.

2) For a given grading, the elements of degree 1, V_1, do not generate $\mathfrak{g}$.

3) $\mathfrak{g}$ does not admit any stratification. ■

Exercise 6.6.21 (A nontrivial filiform algebra)**.** Consider the 6-dimensional Lie algebra $\mathfrak{g}$ given by $\operatorname{span}\{y_0, y_1, y_2, y_3, y_4, y_5\}$ with only non-zero brackets

$$\begin{aligned}
[y_0, y_1] &= y_2, \\
[y_0, y_2] &= y_3, \\
[y_0, y_3] &= y_4, \\
[y_0, y_4] &= y_5 \\
[y_1, y_4] &= -y_5, \\
[y_2, y_3] &= y_5.
\end{aligned}$$

Show the following facts:

1) it is a Lie algebra, i.e., Jacobi identity is satisfied.

2) it admits a stratification.

3) it is a filiform algebra (i.e., the dimensions of the subspaces of the stratification are the smallest possible, namely $2, 1, \dots, 1$). ■

Exercise 6.6.22 (Suggested by E. Breuillard)**.** Let $\mathfrak{g}$ be the 3-step Lie algebra generated by e_1, e_2, e_3 and with the relation $[e_2, e_3] = 0$.

Show that $\mathfrak{g}$ is of dimension 10 and that the following is a stratification of $\mathfrak{g}$.

$$V_1 := \mathrm{span}\{e_1, e_2, e_3\}$$

$$V_2 := \mathrm{span}\{[e_1, e_2], \quad [e_1, e_3]\}$$

$$V_3 := \mathrm{span}\{[e_1, [e_1, e_2]], \quad [e_2, [e_1, e_2]], \quad [e_3, [e_1, e_2]], \quad [e_3, [e_3, e_1]], \quad [e_1, [e_1, e_3]]\}$$

Check that this satisfies the Jacobi identity and is thus a legitimate Lie algebra.

Now let $V_1' := \mathrm{span}\{e_1, e_2 + [e_1, e_2], e_3\}$. Clearly V_1' projects onto V_1 modulo $V_2 + V_3$ and has dimension 3, so it is in direct sum with $[\mathfrak{g}, \mathfrak{g}] = V_2 + V_3$. However $[V_1', V_1']$ is not in direct sum with $[g, [g, g]]$, because it contains $[e_3, e_2 + [e_1, e_2]] = [e_3, [e_1, e_2]]$, and in fact $V_2' := [V_1', V_1']$ has dimension 3, not 2. ■

Exercise 6.6.23. Every stratified Lie algebra admits a Malcev basis.

Solution. Let $\mathfrak{g} = V_1 \oplus \cdots \oplus V_s$ be a stratification of a Lie algebra $\mathfrak{g}$. Let $X_1, \dots, X_n$ be a basis of $\mathfrak{g}$ such that $X_n, \dots, X_1$ is adapted to the direct sum, i.e., there exist $n_1 > \cdots > n_s = 1$ such that $X_n, \dots, X_{n_1}$ is a basis of V_1, $X_{n_{j-1}-1}, \dots, X_{n_j}$ is a basis of V_j, for all $j = 2, \dots, s$.

We claim that $X_1, \dots, X_n$ is a Malcev bais. Indeed, set $\mathfrak{g}_k := \{X_1, \dots, X_k\}$. Thus $X_k \in V_j$, then

$$V_{j+1} \oplus \cdots \oplus V_s \subset \mathfrak{g}_k \subset V_j \oplus \cdots \oplus V_s$$

and

$$\begin{aligned}
[\mathfrak{g}, \mathfrak{g}_k] &\subset [V_1 \oplus \cdots \oplus V_s, V_j \oplus \cdots \oplus V_s] \\
&\subset V_{j+1} \oplus \cdots \oplus V_s \\
&\subset \mathfrak{g}_k
\end{aligned}$$

■

Chapter 7

Riemannian Lie groups*

7.1 Left-invariant Riemannian metrics

Definition 7.1.1 (Left-invariant Riemannian metric)**.** A Riemannian metric $\langle \cdot, \cdot \rangle_\cdot$ on a Lie group G is a *left-invariant Riemannian metric* if

$$\langle (L_h)_* v, (L_h)_* w \rangle_{hg} = \langle v, w \rangle_g, \qquad \forall g, h \in G, \forall v, w \in T_g G.$$

Similarly, we say that $\langle \cdot, \cdot \rangle_\cdot$ is *right-invariant* if

$$\langle (R_h)_* v, (R_h)_* w \rangle_{gh} = \langle v, w \rangle_g, \qquad \forall g, h \in G, \forall v, w \in T_g G.$$

Exercise 7.1.2. Show that a Riemannian metric on G is left invariant if and only if all left translations are isometries. ■

Exercise 7.1.3. Show that, given a group G and a left-invariant distance function d on G, then the following are equivalent: (i) d is right invariant; (ii) d is inversion invariant, that is, $d(x, y) = d(x^{-1}, y^{-1})$, for all x, y in the group; (iii) conjugations are isometries. ■

We shall see various equivalent characterizations for those Riemannian metrics that are both left invariant and right invariant. One of them is the fact that for all $Z \in \mathfrak{g}$, the adjoint map ad_Z is a *skew-adjoint* transformation of $(T_{1_G} G, \langle \cdot, \cdot \rangle_{1_G})$, i.e., it is antisymmetric in the sense that

$$\langle \mathrm{ad}_Z X, Y \rangle = -\langle X, \mathrm{ad}_Z Y \rangle, \qquad \forall X, Y, Z \in \mathfrak{g}.$$

Theorem 7.1.4. *Let G be a connected Lie group with Lie algebra $\mathfrak{g}$ and with a left-invariant Riemannian metric $\langle \cdot, \cdot \rangle$. The following are equivalent:*

(i). $\langle \cdot, \cdot \rangle$ is right invariant;

(ii). Ad_g *is an isometry, for all* $g \in G$*;*

(iii). ad_X *is skew-adjoint, for all* $X \in \mathfrak{g}$*.*

Remark 7.1.5. The connectedness of G is required only for $(iii) \Rightarrow (ii)$.

Proof. $(i) \Leftrightarrow (ii)$: Since $\langle \cdot, \cdot \rangle$ is left-invariant, then (i) is equivalent to $C_g = R_{g^{-1}} \circ L_g$ being isometries, for all $g \in G$, which is equivalent to $(\mathrm{Ad})_g = (\mathrm{d}C_g)_{1_G}$ being isometries, for all g.

$(ii) \Rightarrow (iii)$: Since $\mathrm{Ad}_{\exp(X)} = e^{\mathrm{ad}_X}$, then $e^{\mathrm{ad}(tX)}$ is an isometry, for all $t \in \mathbb{R}$ and all $X \in \mathfrak{g}$. Recall that $\frac{\mathrm{d}}{\mathrm{d}t} e^{tA}|_{t=0} = A$ from Proposition 4.3.4 and take the derivative at $t = 0$ of the identity

$$\langle e^{\mathrm{ad}(tX)} Y, e^{\mathrm{ad}(tX)} Z \rangle_{1_G} = \langle Y, Z \rangle_{1_G}, \qquad \forall X, Y, Z \in \mathfrak{g}.$$

We get

$$\langle \mathrm{ad}_X Y, Z \rangle_{1_G} + \langle Y, \mathrm{ad}_X Z \rangle_{1_G} = 0, \qquad \forall X, Y, Z \in \mathfrak{g}.$$

which is (iii).

$(iii) \Rightarrow (ii)$: Recall that $\frac{\mathrm{d}}{\mathrm{d}t} e^{tA} = A e^{tA}$ from Proposition 4.3.4. We calculate

$$\frac{\mathrm{d}}{\mathrm{d}t} \langle e^{\mathrm{ad}(tX)} Y, e^{\mathrm{ad}(tX)} Z \rangle_{1_G} = \langle \mathrm{ad}(X) e^{\mathrm{ad}(tX)} Y, e^{\mathrm{ad}(tX)} Z \rangle_{1_G} + \\ + \langle e^{\mathrm{ad}(tX)} Y, \mathrm{ad}(X) e^{\mathrm{ad}(tX)} Z \rangle_{1_G} \overset{(iii)}{=} 0.$$

Hence the function $t \mapsto \langle e^{\mathrm{ad}(tX)} Y, e^{\mathrm{ad}(tX)} Z \rangle_{1_G}$ is constant. Evaluating it at $t = 0$ and $t = 1$, we deduce that e^{ad_X} is an isometry. Hence $\mathrm{Ad}_{\exp(X)}$ is an isometry, for all $X \in \mathfrak{g}$.

So Ad_g is an isometry, for all g is a neighborhood U of e in G. Since, when G is connected, every element in G is a finite product $g = g_1 \cdot \cdots \cdot g_k$ of elements $g_1, \dots, g_k \in U$, then $\mathrm{Ad}_g = \mathrm{Ad}_{g_1} \circ \cdots \circ \mathrm{Ad}_{g_k}$ is an isometry. □

7.1.1 Connections and geodesics on Lie groups

Recall that, given a manifold M, we denote by $\mathrm{Vec}(M)$ the space of smooth vector fields on M.

A *linear connection* ∇ on a manifold M is a map

$$\begin{array}{rccc} \nabla : & \mathrm{Vec}(M) \times \mathrm{Vec}(M) & \to & \mathrm{Vec}(M) \\ & (X, Y) & \mapsto & \nabla_X Y \end{array}$$

that is $\mathscr{C}^\infty(M)$-linear in X, $\mathbb{R}$-linear in Y, and satisfies the *Leibniz rule*:

$$\nabla_X (fY) = (Xf)Y + f \nabla_X Y, \qquad \forall f \in \mathscr{C}^\infty(M).$$

Definition 7.1.6 (Left-invariant linear connection)**.** Let G be a Lie group. A linear connection ∇ on G is *left invariant* if

$$(L_g)_* \nabla_X Y = \nabla_{(L_g)_* X} Y, \qquad \forall g \in G, \forall X, Y \in \mathrm{Vec}(G).$$

Exercise 7.1.7. Show that a linear connection ∇ on G is left invariant if and only if for all left-invariant vector fields X, Y the vector field $\nabla_X Y$ is left invariant as well. ■

On every Rimannian manifold there is a unique affine connection that is compatible with the metric and is torsion-free. This connection is called *Levi-Civita connection.* The Levi-Civita connection on a Riemannian manifold M satisfies the *Koszul formula*: for all $X, Y, Z \in \mathrm{Vec}(M)$

$$\begin{aligned}\langle \nabla_X Y, Z\rangle = \frac{1}{2}\Big(& X\langle Y, Z\rangle + Y\langle Z, X\rangle - Z\langle X, Y\rangle + \\ & + \langle [X,Y], Z\rangle + \langle [Z,Y], X\rangle + \langle [Z,X], Y\rangle \Big)\end{aligned} \tag{7.1.8}$$

Exercise 7.1.9. Show that the Levi-Civita connection of a left-invariant Riemannian metric on G is left invariant. ■

Proposition 7.1.10. *Let G be a Lie group with Lie algebra $\mathfrak{g}$ of left-invariant vector fields. There is a one-to-one correspondence between the set of left-invariant linear connections ∇ on G and the set* $\mathrm{Mult}(\mathfrak{g}, \mathfrak{g}; \mathfrak{g})$ *of bilinear functions $\alpha : \mathfrak{g} \times \mathfrak{g} \to \mathfrak{g}$ given by*

$$\alpha_\nabla(X, Y) = \nabla_X Y, \qquad \forall X, Y \in \mathfrak{g}.$$

Notice that if $X_1, \dots, X_n$ is a frame of left-invariant vector fields, and $\sum_j a^j X_j, \sum_i b^i X_i \in \mathrm{Vec}(M)$, then

$$\nabla_{a^j X_j}(b^i X_i) = a^j b^i \nabla_{X_j} X_i + a^j (X_j b^i) X_i = a^j b^i \alpha_\nabla(X_j, X_i) + a^j (X_j b^i) X_i$$

Proposition 7.1.11. *Let G be a Lie group with a left-invariant linear connection ∇ and a left-invariant vector field X. The following are equivalent:*

(i). $\alpha_\nabla(X, X) = 0$ (i.e., $\nabla_X X = 0$);

(ii). the one-parameter subgroup $t \mapsto \Phi^t_X(1_G)$ is a geodesic with respect to ∇.

Proof. The curve $t \mapsto \gamma(t) := \Phi^t_X(1_G)$ has derivative $\gamma'(t) = X_{\gamma(t)}$. Hence γ is a geodesic with respect to ∇ if and only if (by definition) $\nabla_{\gamma'}\gamma' = 0$, thus if and only if $(\nabla_X X)_\gamma \equiv 0$. Since $\nabla_X X$ is a left-invariant vector field, then γ is a geodesic if and only if $(\nabla_X X)_{1_G} = 0$, if and only if $\nabla_X X = 0$. □

For another characterization of when a one-parameter subgroup is a Riemannian geodesic, see also Corollary 7.1.19

Exercise 7.1.12. Show that a linear connection ∇ on G is left invariant if and only if the Christoffel symbols with respect to any frame of left-invariant vector fields are constant functions on the group.

■

Example 7.1.13. Let G be a Lie group and $c \in \mathbb{R}$. Then the map

$$(X, Y) \mapsto c[X, Y]$$

is in $\mathrm{Mult}(\mathfrak{g}, \mathfrak{g}; \mathfrak{g})$. Hence, by Proposition 7.1.10, it induces a left-invariant linear connection on G. Notice that for this connection the Christoffel symbols Γ_{ij}^k with respect to a frame of left-invariant vector fields are precisely the structural constants c_{ij}^k (see Definition **??**) with respect to the same frame, multiplied by c.

Lemma 7.1.14. *Let $\langle \cdot, \cdot \rangle$ be a left-invariant Riemannian metric on a group G and let ∇ be the Levi-Civita connection associated.*

(i). For all left-invariant vector fields X, Y, Z on G, we have

$$\langle \nabla_X Y, Z \rangle = \frac{1}{2}\Big(\langle [X, Y], Z\rangle + \langle [Z, Y], X\rangle + \langle [Z, X], Y\rangle\Big). \tag{7.1.15}$$

(ii). If $X_1, \ldots, X_n$ are orthonormal left-invariant vector fields that form a basis of Lie(G) *and α_{ij}^k are the associated structural constants, then*

$$\langle [X_i, X_j], X_k\rangle = \alpha_{ij}^k, \tag{7.1.16}$$

$$\langle \nabla_{X_i} X_j, X_k\rangle = \frac{1}{2}\left(\alpha_{ij}^k - \alpha_{jk}^i + \alpha_{ki}^j\right), \tag{7.1.17}$$

$$\nabla_{X_i} X_j = \frac{1}{2}\sum_{k=1}^n \left(\alpha_{ij}^k - \alpha_{jk}^i + \alpha_{ki}^j\right) X_k. \tag{7.1.18}$$

Proof. (i) Recall that the Levi-Civita connection satisfies Koszul Formula (7.1.8). Note that if X, Y are left invariant vector fields, then $\langle X, Y\rangle$ is constant. Hence, Koszul Formula simplifies to (7.1.15).

(ii) Since $[X_i, X_j] = \sum_k \alpha_{ij}^k X_k$, so $\langle [X_i, X_j], X_k\rangle = \sum_h \alpha_{ij}^h \langle X_h, X_k\rangle = \alpha_{ij}^k$, because the X_i's are orthonormal. From part (i) the rest follows. □

Corollary 7.1.19. *Let G be a Lie group endowed with a left-invariant Riemannian metric. Then the one-parameter subgroups in the direction of X is geodesic if and only if X is orthogonal to $[X, \mathfrak{g}]$.*

Proof. It follows from (7.1.15) and Proposition 7.1.11. ☐

Theorem 7.1.20. *Let $\langle \cdot, \cdot \rangle$ be a left-invariant Riemannian metric on a Lie group G and ∇ be the Levi-Civita connection associated. The following are equivalent:*

(i). $\exp_{1_G} = \exp$*, i.e., the family of one-parameter subgroups is exactly the family of the geodesics from the identity element 1_G;*

(ii). If $X, Y \in \mathfrak{g}$, then $\alpha_\nabla(X, Y) = \frac{1}{2}[X, Y]$, i.e.,

$$\nabla_X Y = \frac{1}{2}[X, Y], \qquad \forall X, Y \in \mathfrak{g};$$

(iii). The map ad_Z *is skew-adjoint, for all $Z \in \mathfrak{g}$.*

Proof. Note that the formula (7.1.15) in Lemma 7.1.14 can be written also as

$$\begin{aligned}\langle \nabla_X Y, Z\rangle &= \frac{1}{2}\Big(\langle [X, Y], Z\rangle + \langle [Z, Y], X\rangle + \langle [Z, X], Y\rangle\Big)\\ &= \frac{1}{2}\Big(\langle [X, Y], Z\rangle + \langle \mathrm{ad}_Z Y, X\rangle + \langle \mathrm{ad}_Z X, Y\rangle\Big).\end{aligned}$$

The equivalence $(ii) \Leftrightarrow (iii)$ easily follows from the last equality. Moreover, for $X = Y$ we get

$$\langle \nabla_X X, Z\rangle = \langle \mathrm{ad}_Z X, X\rangle, \qquad \forall X, Z \in \mathfrak{g}.$$

Hence (i), which by Proposition 7.1.11 is equivalent to $\nabla_X X = 0$, for all $X \in \mathfrak{g}$, is also equivalent to $\langle \mathrm{ad}_Z X, X\rangle = 0$, for all $X, Z \in \mathfrak{g}$, which (by an easy computation) is equivalent to ad_Z being skew-adjoint. ☐

Remark 7.1.21. Note that the connection in the point (ii) of Theorem 7.1.20, i.e., $\nabla_X Y := \frac{1}{2}[X, Y]$, is always a left-invariant linear connection on G, but there we require it to be the Levi-Civita connection!

7.1.2 Curvatures of left-invariant metrics

In the following, be aware that our choice for the *Riemannian curvature tensor* is

$$R(X, Y, \cdot) := \nabla_X \nabla_Y - \nabla_Y \nabla_X - \nabla_{[X,Y]}. \tag{7.1.22}$$

Also, recall that the *sectional curvature* of two linearly independent tangent vectors X, Y at the same point is

$$\mathrm{Sec}(X,Y) := \frac{\langle R(X,Y,Y),X\rangle}{\|X\|^2\|Y\|^2 - \langle X,Y\rangle^2}. \tag{7.1.23}$$

Lemma 7.1.24. *Let G be a Lie group equipped with a left-invariant Riemannian metric. Let $X_1,\dots,X_n$ be orthonormal left-invariant vector fields that form a basis of* $\mathrm{Lie(G)}$ *and let α_{ij}^k be the associated structural constants. Then the Riemannian curvature tensor satisfies*

$$\begin{aligned} R(X_i,X_j,X_k) = \sum_{\ell,h=1}^n \Bigg[& \frac{1}{4}\left(\alpha_{jk}^\ell - \alpha_{k\ell}^j + \alpha_{\ell j}^k\right)\left(\alpha_{i\ell}^h - \alpha_{\ell h}^i + \alpha_{hi}^\ell\right) \\ & -\frac{1}{4}\left(\alpha_{ik}^\ell - \alpha_{k\ell}^i + \alpha_{\ell i}^k\right)\left(\alpha_{j\ell}^h - \alpha_{\ell h}^j + \alpha_{hj}^\ell\right) \\ & -\frac{1}{2}\alpha_{ij}^\ell\left(\alpha_{\ell k}^h - \alpha_{kh}^\ell + \alpha_{h\ell}^k\right)\Bigg] X_h. \end{aligned}$$

The sectional curvature satisfies

$$\begin{aligned} \mathrm{Sec}(X_1,X_2) = \sum_{\ell=1}^n \Bigg[& -\frac{1}{2}\alpha_{12}^\ell\left(\alpha_{12}^\ell - \alpha_{2\ell}^1 - \alpha_{\ell 1}^2\right) \\ & -\frac{1}{4}\left(\alpha_{12}^\ell - \alpha_{2\ell}^1 + \alpha_{\ell 1}^2\right)\left(\alpha_{12}^\ell + \alpha_{2\ell}^1 - \alpha_{\ell 1}^2\right) \\ & -\alpha_{\ell 1}^1\alpha_{\ell 2}^2\Bigg]. \end{aligned} \tag{7.1.25}$$

Proof. Recall that we defined R by (7.1.22). So

$$R(X_i,X_j,X_k) = \nabla_{X_i}\nabla_{X_j}X_k - \nabla_{X_j}\nabla_{X_i}X_k - \nabla_{\sum_\ell \alpha_{ij}^\ell X_\ell}X_k$$

$$
\begin{aligned}
&= \nabla_{X_i}\left(\frac{1}{2}\sum_\ell(\alpha^\ell_{jk}-\alpha^j_{k\ell}+\alpha^k_{\ell j})X_\ell\right)+\\
&\qquad -\nabla_{X_j}\left(\frac{1}{2}\sum_\ell(\alpha^\ell_{jk}-\alpha^i_{k\ell}+\alpha^k_{\ell i})X_\ell\right)+\\
&\qquad\qquad -\sum_\ell\alpha^\ell_{ij}\frac{1}{2}\sum_h(\alpha^h_{\ell k}-\alpha^\ell_{kh}+\alpha^k_{h\ell})X_h\\
&= \frac{1}{2}\sum_\ell(\alpha^\ell_{jk}-\alpha^j_{k\ell}+\alpha^k_{\ell j})\frac{1}{2}\sum_h(\alpha^h_{i\ell}-\alpha^i_{\ell h}+\alpha^\ell_{hi})X_h+\\
&\qquad -\frac{1}{2}\sum_\ell(\alpha^\ell_{ik}-\alpha^i_{k\ell}+\alpha^k_{\ell i})\frac{1}{2}\sum_h(\alpha^h_{j\ell}-\alpha^j_{\ell h}+\alpha^\ell_{hj})X_h+\\
&\qquad\qquad -\sum_\ell\frac{1}{2}\alpha^\ell_{ij}\sum_h(\alpha^h_{\ell k}-\alpha^\ell_{kh}+\alpha^k_{h\ell})X_h\\
&= \sum_{\ell,h}\left[\frac{1}{4}(\alpha^\ell_{jk}-\alpha^j_{k\ell}+\alpha^k_{\ell j})(\alpha^h_{i\ell}-\alpha^i_{\ell h}+\alpha^\ell_{hi})+\right.\\
&\qquad -\frac{1}{4}(\alpha^\ell_{ik}-\alpha^i_{k\ell}+\alpha^k_{\ell i})(\alpha^h_{j\ell}-\alpha^j_{\ell h}+\alpha^\ell_{hj})+\\
&\qquad\qquad \left.-\frac{1}{2}\alpha^\ell_{ij}(\alpha^h_{\ell k}-\alpha^\ell_{kh}+\alpha^k_{h\ell})\right]X_h.
\end{aligned}
$$

Regarding the sectional curvature (7.1.23), since X_1, X_2 are orthonormal, we have

$$\mathrm{Sec}(X_1,X_2)=\langle R(X_1,X_2,X_2),X_1\rangle.$$

So, using the above formula with $i=1$, $j=k=2$ and $h=1$, we have

$$\mathrm{Sec}(X_1,X_2)=\langle R(X_1,X_2,X_2),X_1\rangle$$

$$
\begin{aligned}
&= \sum_\ell\left[\frac{1}{4}(-\alpha^2_{2\ell}+\alpha^2_{\ell 2})(\alpha^1_{1\ell}-\alpha^1_{\ell 1})+\right.\\
&\qquad -\frac{1}{4}(\alpha^\ell_{12}-\alpha^1_{2\ell}+\alpha^2_{\ell 1})(\alpha^1_{2\ell}-\alpha^2_{\ell 1}+\alpha^\ell_{12})+\\
&\qquad\qquad \left.-\frac{1}{2}\alpha^\ell_{12}(\alpha^1_{\ell 2}-\alpha^\ell_{21}+\alpha^2_{1\ell})\right]
\end{aligned}
$$

$$= \sum_\ell \Big[\frac{1}{4}\left(2\alpha^2_{\ell 2}(-2\alpha^1_{\ell 1})\right) +$$

$$- \frac{1}{4}(\alpha^\ell_{12} - \alpha^1_{2\ell} + \alpha^2_{\ell 1})(\alpha^\ell_{12} + \alpha^1_{2\ell} - \alpha^2_{\ell 1}) +$$

$$- \frac{1}{2}\alpha^\ell_{12}(\alpha^\ell_{12} - \alpha^1_{2\ell} - \alpha^2_{\ell 1})\Big]$$

$$= \sum_\ell \Big[-\alpha^1_{\ell 1}\alpha^2_{\ell 2} - \frac{1}{4}(\alpha^\ell_{12} - \alpha^1_{2\ell} + \alpha^2_{\ell 1})(\alpha^\ell_{12} + \alpha^1_{2\ell} - \alpha^2_{\ell 1}) +$$

$$- \alpha^\ell_{12}(\alpha^\ell_{12} - \alpha^1_{2\ell} - \alpha^2_{\ell 1})\Big].$$

□

Lemma 7.1.26. *Let $X \in \mathrm{Lie(G)}$. If ad_X is skew-adjoint, then $\mathrm{Sec}(X,Y) \geq 0$ for all $Y \in \mathrm{Lie(G)}$.*

Proof. Recall that ad_X being skew adjoint means

$$\langle \mathrm{ad}_X Y, Z\rangle = -\langle Y, \mathrm{ad}_X Z\rangle.$$

Let $X \in \mathrm{Lie(G)}$ be such that ad_X is skew symmetric. Since $\mathrm{ad}_{\lambda X} = \lambda \,\mathrm{ad}_X$, we can suppose $\langle X, X\rangle = 1$. Let $Y \in \mathrm{Lie(G)}$ be such that $\langle X, Y\rangle = 0$ and $\langle Y, Y\rangle = 1$. Take an orthonormal basis $X_1, \dots, X_n$ of Lie(G) with $X_1 = X$, $X_2 = Y$ and let α^k_{ij} be the corresponding structural constants, i.e., $\mathrm{ad}_{X_i}(X_j) = [X_i, X_j] = \sum_k \alpha^k_{ij} X_k$. Then

$$\alpha^k_{1j} = \langle \mathrm{ad}_{X_1} X_j, X_k\rangle = -\langle X_j, \mathrm{ad}_{X_1} X_k\rangle = -\alpha^j_{1k}$$

The formula (7.1.25) simplifies to

$$\mathrm{Sec}(X_1, X_2) = -\frac{1}{4}(2\alpha^\ell_{12} - \alpha^1_{2\ell})(\alpha^1_{2\ell}) - \frac{1}{2}\alpha^\ell_{12}(-\alpha^1_{2\ell})$$
$$= \frac{1}{4}(\alpha^1_{2\ell})^2 - \frac{1}{2}\alpha^\ell_{12}\alpha^1_{2\ell} + \frac{1}{2}\alpha^\ell_{12}\alpha^1_{2\ell} = \frac{1}{4}(\alpha^1_{2\ell})^2 \geq 0.$$

□

7.1.3 Bi-invariant metrics

A Riemannian metric of a Lie group that is left invariant and right invariant is said to be *bi-invariant.*

For the aim of exposition, we recall the other characterizations that we obtained for the bi-invariance of metrics: Theorems 7.1.4 and 7.1.20.

Corollary 7.1.27. *Let G be a connected Lie group with Lie algebra $\mathfrak{g}$ and with a left-invariant Riemannian metric $\langle\cdot,\cdot\rangle$ with Levi-Civita connection ∇. The following are equivalent:*

(i). $\langle\cdot,\cdot\rangle$ *is bi-invariant;*

(ii). Ad_g *is an isometry, for all* $g \in G$*;*

(iii). ad_X *is skew-adjoint, for all* $X \in \mathfrak{g}$*;*

(iv). $\exp_{1_G} = \exp$*;*

(v). $\nabla_X Y = \frac{1}{2}[X,Y]$*, for all* $X, Y \in \mathfrak{g}$*;*

Lemma 7.1.26 gives a non-trivial property of bi-invariant metrics:

Corollary 7.1.28. *If a connected Lie group is equipped with a bi-invariant metric, then all its sectional curvature are nonnegative.*

Exercise 7.1.29. Show that every Lie group admits a left-invariant Riemannian metric and a left-invariant measure, as follows. Let $X_1, \dots, X_n$ be left-invariant vector fields forming a basis of Lie(G). Consider the Riemannian metric that makes $X_1, \dots, X_n$ orthonormal and the differential n-form vol for which $\mathrm{vol}(X_1, \dots, X_n) = 1$. ■

Exercise 7.1.30. Show that if G is a compact Lie group then it admits a left-invariant Riemannian volume form that gives a probability measure. ■

Theorem 7.1.31. *Every compact Lie group G admits a bi-invariant Riemannian metric.*

Proof. Let $\langle\cdot,\cdot\rangle$ be a scalar product on $T_{1_G}G$. Let vol be a left-invariant probability measure on G. Define a new product $\langle\!\langle\cdot,\cdot\rangle\!\rangle$ averaging $(\mathrm{Ad}_g)_*\langle\cdot,\cdot\rangle$ with vol, i.e.,

$$\langle\!\langle X, Y\rangle\!\rangle = \int_G \langle \mathrm{Ad}_g(X), \mathrm{Ad}_g(Y)\rangle \,\mathrm{d}\,\mathrm{vol}(g), \qquad \forall X, Y \in T_{1_G}G.$$

Notice that $\langle\!\langle\cdot,\cdot\rangle\!\rangle$ is finite since $\mathrm{vol}(G) < \infty$. Moreover, for all $g \in G$ the product $\langle\!\langle\cdot,\cdot\rangle\!\rangle$ is Ad_g invariant:

$$\begin{aligned} \langle\!\langle \mathrm{Ad}_g X, \mathrm{Ad}_g Y\rangle\!\rangle &= \int_G \langle \mathrm{Ad}_h \mathrm{Ad}_g X, \mathrm{Ad}_h \mathrm{Ad}_g Y\rangle \,\mathrm{d}\,\mathrm{vol}(h) \\ &= \int_G \langle \mathrm{Ad}_{g'} X, \mathrm{Ad}_{g'} Y\rangle \,\mathrm{d}\,\mathrm{vol}(g') \\ &= \langle\!\langle X, Y\rangle\!\rangle. \end{aligned}$$

Extending $\langle\!\langle\cdot,\cdot\rangle\!\rangle$ by left translations, we get a left-invariant Riemannian metric for which Ad_g are isometries. Theorem 7.1.4 implies that this metric is also right invariant. □

Corollary 7.1.32. *Every compact Lie group can be equipped with a probability measure that is bi-invariant.*

Corollary 7.1.33. *On every compact connected Lie group G the exponential map* $\exp : \mathrm{Lie(G)} \to \mathrm{G}$ *is surjective.*

Here is another characterization of groups that admit bi-invariant metrics, see [**?**].

Theorem 7.1.34. *Let G be a connected Lie group. Then the following are equivalent:*

1. *There exists an admissible bi-invariant distance on G;*

2. *There exists a bi-invariant Riemannian metric on G;*

3. *G is the direct product of a compact group and a vector group, that is, $\mathbb{R}^n$ for some $n \in \mathbb{N}$;*

4. *Ad_G is compact in the space of linear transformations of the Lie algebra $\mathfrak{g}$.*

The most difficult part of the above theorem is given by the following Milnor's result:

Lemma 7.1.35 (Milnor, Lemma 7.5)**.** *If a connected Lie group admits a bi-invariant metric then it is the cartesian product of a compact group and a commutative group.*

From the topological view point, we actually always have that a Lie group is the product of a compact group and a vector space. The following deep result is due to Iwasawa:

Theorem 7.1.36 (Iwasawa, see Milnor, page 327)**.** *Let G be connected Lie group, then:*

1. *Every compact subgroup is contained in a maximal compact subgroup H, which is necessarily a connected Lie group.*

2. *This maximal compact subgroup is unique up to conjugation.*

3. *As a topological space, G is homeomorphic with the product of H and some Euclidean space $\mathbb{R}^m$.*

7.1.4 Some other results on curvature

For the following results we refer to the paper of Milnor.

Theorem 7.1.37 (Milnor, Theorem 3.3)**.** *If* $\mathrm{Lie(G)}$ *is not commutative, then G admits a left-invariant metric of strictly negative* scalar *curvature.*

Theorem 7.1.38 (Milnor, Theorem 2.2)**.** *A connected Lie group admits a left-invariant metric with all* Ricci *curvatures strictly positive if and only if it is compact with finite fundamental group.*

Recall: $\mathrm{Ric}(X) := \sum_i \mathrm{Sec}(X, X_i)$, where $X_1, \dots, X_n$ is an orthonormal frame.

Theorem 7.1.39 (Milnor, Theorem 2.5)**.** *If there are $X, Y, Z \in \mathrm{Lie(G)}$ such that $[X, Y] = Z$, then there is a left-invariant metric on G such that $\mathrm{Ric}(X) < 0 < \mathrm{Ric}(Z)$.*

Theorem 7.1.40 (Milnor, Corollary 7.7)**.** *Every Lie group whose universal covering space is compact admits a bi-invariant metric of constant Ricci curvature $+1$.*

7.2 Isometries of metric groups*

[...]

Isometries of metric Lie groups are smooth.

Isometries of nilpotent Lie groups are affine

...

7.3 Rectifiable curves in sub-Finsler nilpotent groups*

7.3.1 A special sub-Finsler geometry on nilpotent groups

Let G be a simply connected nilpotent Lie group. Let $V_1 \subseteq T_e G$ be a sub-vector space. Let Δ be the left-invariant distribution with $\Delta_e = V_1$. Considering V_1 as a sub-space of the Lie algebra $\mathfrak{g}$ of G, assume that the algebra generated by V_1 is the whole of $\mathfrak{g}$. In other words, assume that Δ is bracket generating. Thus we have the flag of left-invariant bundles

$$\Delta = \Delta^{[1]} \subseteq \Delta^{[2]} = \Delta + [\Delta, \Delta] \subseteq \ldots \subseteq \Delta^{[s]} = TG.$$

There is a one-to-one correspondence between vectors in V_1 and vector fields in the intersection $\mathfrak{g} \cap \Gamma(\Delta)$ of the Lie algebra of G and the sections of Δ. We will confuse the two notions without problems.

Fix a norm $\|\cdot\|$ on V_1. It extends to a left-invariant norm on Δ. The triple $(G, \Delta, \|\cdot\|)$ is a sub-Finsler manifold.

In the sequel, whenever we speak of the FCC metric on the simply connected nilpotent Lie group G, we mean one that is associated to a norm $\|\cdot\|$ on a sub-space V_1 such that $\mathfrak{g} = V_1 \oplus [\mathfrak{g}, \mathfrak{g}]$ where $\mathfrak{g} = Lie(G)$.

One can easily check that every such a V_1 generates the Lie algebra, cf. Exercise 6.6.5.

Assume that G is a simply connected nilpotent Lie group with a left-invariant distribution Δ such that

$$\mathfrak{g} = \Delta_e \oplus [\mathfrak{g}, \mathfrak{g}], \tag{7.3.1}$$

as, for example, a Carnot group.

Question 7.3.2. *If G is a simply connected nilpotent Lie group and V_1 and W_1 are sub-spaces such that*

$$\mathfrak{g} = V_1 \oplus [\mathfrak{g}, \mathfrak{g}] = W_1 \oplus [\mathfrak{g}, \mathfrak{g}],$$

then, does exist a Lie algebra isomorphism $\phi : \mathfrak{g} \to \mathfrak{g}$ such that $\phi(V_1) = W_1$? The aswer should be no, however, see Exercise 6.6.5.

Definition 7.3.3 (The projection π_1)**.** Let $\mathrm{proj} : T_eG \to V_1 = \Delta_e$ be the projection onto V_1 with kernel $[\mathfrak{g}, \mathfrak{g}]$. Define

$$\begin{aligned} \pi_1 : \quad G &\to \quad V_1 \\ p &\mapsto \quad \pi_1(p) := \mathrm{proj}(\exp^{-1}(p)). \end{aligned} \tag{7.3.4}$$

Lemma 7.3.5. *The following properties hold:*

(i) *The map $\pi_1 : (G, \cdot) \to (V_1, +)$ is a group homomorphism.*

(ii) *The differential of π_1 is the identity when restricted to V_1:*

$$d\pi_1|_{V_1} = \mathrm{id}_{V_1}.$$

Proof of (i). By Theorem 6.5.1, since G is a simply connected and nilpotent, for all p and $q \in G$, exist X and $Y \in \mathfrak{g}$ such that $\exp(X) = p$ and $\exp(Y) = q$. Then, by BCH formula and assumption (7.3.1) we have

$$\begin{aligned} \pi_1(p \cdot q) &= \mathrm{proj}(\exp^{-1}(pq)) = \mathrm{proj}\left(\exp^{-1}(\exp(X)\exp(Y))\right) \\ &= \mathrm{proj}\left(X + Y + \frac{1}{2}[X, Y] + \dots\right) = \mathrm{proj}(X + Y). \end{aligned}$$

On the other hand,

$$\pi_1(p) + \pi_1(q) = \mathrm{proj}(\exp^{-1}(p)) + \mathrm{proj}(\exp^{-1}(q)) = \mathrm{proj}(\exp^{-1}(p) + \exp^{-1}(q)) = \mathrm{proj}(X + Y).$$

Proof of (ii). Since Theorem 4.2.6(iii), $d\pi_1|_{V_1} = d(\mathrm{proj}|_{V_1}) = d(\mathrm{id}|_{V_1}) = \mathrm{id}_{V_1}$. ☐

The "development" of a curve

The map π_1 is useful since it gives a second link between the tangents of a horizontal curves and vectors in V_1. Let $\gamma(t)$ be an absolute continuous curve with $\dot\gamma(t)$ horizontal, i.e., $\dot\gamma(t) \in \Delta_{\gamma(t)} \subseteq T_{\gamma(t)}G$ for almost every t. The vector $\dot\gamma(t)$ can be identified with a vector in V_1, so as a tangent vector at the identity. Namely, we consider $\gamma'(t) \in V_1 \subseteq T_eG$ as the vector

$$\gamma'(t) := (L_{\gamma(t)})_*^{-1}\dot\gamma(t).$$

We then have the following formula

$$\gamma'(t) = \frac{\mathrm{d}}{\mathrm{d}t}\left(\pi_1 \circ \gamma\right)(t) \tag{7.3.6}$$

Proof of Formula (7.3.6). Using Lemma 7.3.5, and that $\pi_1(1) = 0$, we get

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}\left(\pi_1 \circ \gamma\right)(t) &= \lim_{h\to 0} \frac{\pi_1(\gamma(t+h)) - \pi_1(\gamma(t))}{h} \\
&= \lim_{h\to 0} \frac{\pi_1(\gamma(t)^{-1}) + \pi_1(\gamma(t+h))}{h} \\
&= \lim_{h\to 0} \frac{\pi_1(\gamma(t)^{-1}\gamma(t+h))}{h} \\
&= \lim_{h\to 0} \frac{\pi_1(L^{-1}_{\gamma(t)}\gamma(t+h))}{h} \\
&= \lim_{h\to 0} \frac{\pi_1(L^{-1}_{\gamma(t)}\gamma(t+h)) - \pi_1(L^{-1}_{\gamma(t)}\gamma(t))}{h} \\
&= \frac{\mathrm{d}}{\mathrm{d}h}\left(\left(\pi_1 \circ L^{-1}_{\gamma(t)} \circ \gamma\right)(t+h)\right|_{h=0} \\
&= (\pi_1)_* \circ (L^{-1}_{\gamma(t)})_*\dot\gamma(t) \\
&= \mathrm{id}(\gamma'(t)) = \gamma'(t)
\end{aligned}$$

□

7.3.2 Horizontal lines as geodesics

Definition 7.3.7. Let $X \in V_1$. The curve $\gamma(t) := \exp(tX)$ is the one-parameter sub-group of the horizontal vector X, and it is called the horizontal line in the direction of X.

The curve $\gamma(t)$ is obviously horizontal with respect to Δ, since

$$\dot\gamma(t) = X_{\gamma(t)} \in \Delta_{\gamma(t)}.$$

The length of $\gamma(t)$, for $t \in [0,T]$, with respect to the CC metric of $(M, \Delta, \|\cdot\|)$ is $T\|X\|$. Indeed,

$$\begin{aligned}
\mathrm{Length}(\gamma) &= \int_0^T \|\dot\gamma(t)\| \, \mathrm{d}d \\
&= \int_0^T \left\|X_{\gamma(t)}\right\| \, \mathrm{d}d \\
&= \int_0^T \left\|\left(L_{\gamma(t)}\right)_* X_e\right\| \, \mathrm{d}d \\
&= \int_0^T \|X\| \, \mathrm{d}d \\
&= T\|X\|,
\end{aligned}$$

where we used that both X and the norm are left-invariant. Thus we get the formula

$$\mathrm{Length}\left(\exp(tX)|_{t\in[0,T]}\right) = T\|X\|. \tag{7.3.8}$$

In a Lie group endowed with a left-invariant Riemannian metric, the one-parameter subgroups are NOT always geodesics.

For instance in $SL(2,\mathbb{R})$ the upper triangular unipotent one parameter subgroup

$$\begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix}$$

is not a geodesic, because it's distance to Id is roughly $\log(t)$, not t.

In a non-compact simple Lie group only the one-parameter groups coming from the p part of the Cartan decomposition will be geodesics.

Also in the Heisenberg group, if you consider the vertical line, then it is a one-parameter group, but not a geodesic in Riemannian left-invariant metrics, recall Proposition 7.1.19.

Proposition 7.3.9. *Consider a nilpotent Lie group G endowed with a left-invariant sub-Finsler distance with respect to some distribution Δ such that*

$$\Delta \oplus [\mathfrak{g}, \mathfrak{g}] = \mathfrak{g}.$$

Then one-parameter subgroups of horizontal vectors are geodesics.

Proof. What we need to show is that

$$\|\pi_1(g)\| \le d_{CC}(e, g), \tag{7.3.10}$$

where d_{CC} is the Finsler-Carnot-Carathéodory distance and π_1 is the projection defined in (7.3.4).

Restricting (7.3.10) to g belonging to $\exp(V_1)$ will finish the proof because of calculation (7.3.8). Indeed, if now $X \in V_1$ then the curve $t \mapsto \exp(tX)$ is a geodesic since

$$d_{CC}(e, \exp(TX)) \leq \text{Length}\left(\exp(tX)|_{t\in[0,T]}\right) = T\,\|X\| = \|TX\| \leq d_{CC}(e, \exp(TX)).$$

Now inequality (7.3.10) is true because, by definition of the metric on G, there is a sequence of piece-wise linear (or piece-wise smooth) horizontal curves joining e and g whose length tends to $d_{CC}(e,g)$. But if $\gamma(t) : [0,1] \to G$ is such a curve, then, by Formula (7.3.6),

$$\begin{aligned}
\|\pi_1(g)\| &= \|\pi_1(\gamma(1)) - \pi_1(\gamma(0))\| \\
&= \left\|\int_0^1 \frac{\mathrm{d}}{\mathrm{d}t}\left(\pi_1 \circ \gamma\right)(t)dt\right\| \\
&\leq \int_0^1 \left\|\frac{\mathrm{d}}{\mathrm{d}t}\left(\pi_1 \circ \gamma\right)(t)\right\| dt \\
&= \int_0^1 \|\gamma'(t)\|\, dt \\
&= \int_0^1 \left\|(L_{\gamma(t)})_*^{-1}\dot\gamma(t)\right\| dt \\
&= \int_0^1 \|\dot\gamma(t)\|\, dt \\
&= \text{Length}(\gamma).
\end{aligned}$$

□

7.3.3 Lifts of curves

Lemma 7.3.11. *Let $\phi : \mathfrak{g} \to \mathfrak{h}$ a Lie algebra homomorphism.*

(i) *If ϕ has the property that $\phi(V_1^{(G)}) \subseteq V_1^{(H)}$, then*

$$\mathrm{proj}^H \circ \phi = \phi \circ \mathrm{proj}^G,$$

where $\mathrm{proj}^G : \mathfrak{g} \to \mathfrak{g}$ and $\mathrm{proj}^H : \mathfrak{h} \to \mathfrak{h}$ are the projections onto V_1^G and V_1^H respectively with kernels $[\mathfrak{g},\mathfrak{g}]$ and $[\mathfrak{h},\mathfrak{h}]$ respectively.

(ii) *If ϕ has the property that $\phi(V_1^{(G)}) \supseteq V_1^{(H)}$, then ϕ is surjective.*

(iii) *If $\varphi : G \to H$ is a group morphism such that*

$$\varphi_*|_{V_1^{(G)}} : V_1^{(G)} \to V_1^{(H)}$$

is an isometry of normed spaces, then $\varphi : G \to H$ is a 1-Lipschitz, with respect to the respective FCC metrics.

Proof of (i). If $X \in V_1^{(G)}$, then $(\phi \circ \mathrm{proj})(X) = \phi(X)$. Since by assumption we also have $\phi(X) \in V_1^{(H)}$, then $(\mathrm{proj} \circ \phi)(X) = \phi(X)$. So $\mathrm{proj} \circ \phi$ and $\phi \circ \mathrm{proj}$ are two homomorphisms that coincide on $V_1^{(G)}$. Since $V_1^{(G)}$ generates the algebra $\mathfrak{g}$, then the two homomorphisms are equal.

Proof of (ii). It is obvious since $\phi(\mathfrak{g})$ is a Lie algebra that contains the generating sub-space $V_1^{(H)}$.

Proof of (iii). It is enough to observe that if $\gamma : [0,1] \to G$ is a geodesic, then

$$\begin{aligned}
d(\gamma(0), \gamma(1)) &= \mathrm{Length}(\gamma) \\
&= \int_0^1 \|\dot\gamma(t)\| \,\mathrm{d}t \\
&= \int_0^1 \|\gamma'(t)\|_{V_1(G)} \,\mathrm{d}d \\
&= \int_0^1 \|\varphi_*(\gamma'(t))\|_{V_1(H)} \,\mathrm{d}t \\
&= \int_0^1 \left\| (L_{\varphi(\gamma(t))})^* \frac{\mathrm{d}}{\mathrm{d}t} (\varphi(\gamma(t))) \right\|_{V_1(H)} \mathrm{d}t \\
&= \int_0^1 \left\| \frac{\mathrm{d}}{\mathrm{d}t} (\varphi(\gamma(t))) \right\| \mathrm{d}t \\
&= \mathrm{Length}(\varphi \circ \gamma) \\
&\geq d(\varphi(\gamma(0)), \varphi(\gamma(1))).
\end{aligned}$$

□

Lemma 7.3.12. *The projection map $\pi_1 : G \to V_1$ has the following properties.*

(i) *For every Lipschitz curve σ in V_1 with $\sigma(0) = 0$, there exists a unique Lipschitz horizontal curve γ with $\pi_1(\gamma) = \sigma$ and $\gamma(0) = e$, and such a curve is the solution of the ODE*

$$\left\{ \begin{array}{lcl} \dot\gamma(t) &=& (L_{\gamma(t)})_* \dot\sigma(t) \\ \gamma(0) &=& e. \end{array} \right. \tag{7.3.13}$$

(ii) *The length of the horizontal curves equals the length of their projections:*

$$\mathrm{Length}(\gamma) = \mathrm{Length}(\pi_1 \circ \gamma),$$

for all horizontal curves γ, with $\gamma(0) = e$, where the first length is with respect to the FCC metric and the second one is in the normed space $(V_1, \|\cdot\|)$.

(iii) *If $\varphi : G \to H$ is a Lie group homomorphism with $\varphi_*(V_1^{(G)}) \subseteq V_1^{(H)}$, then*

$$\pi_1^{(H)} \circ \varphi \circ \gamma = \varphi_* \circ \pi_1^{(G)} \circ \gamma,$$

for all horizontal curves γ, with $\gamma(0) = e$.

Proof of (i). The existence of a solution of the ODE is a consequence of the general Carathéodory's theorem, cf. cite[page 43]Coddington-Levinson (1955). The uniqueness can be shown proving that, if $\gamma_1(t)$ and $\gamma_2(t)$ are two solutions, then

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\gamma_1(t)\gamma_2(t)^{-1}\right) \equiv 0.$$

Let $\gamma(t)$ be the solution of the ODE. Then

$$\gamma'(t) = (L_{\gamma(t)})^* \dot{\gamma}(t) = \frac{\mathrm{d}}{\mathrm{d}t}(\sigma(t)).$$

Since Formula (7.3.6), we have that $\pi_1 \circ \gamma$ and σ are two curves in V_1 with same starting point $\pi_1(\gamma(0)) = 0 = \sigma(0)$ and same derivative

$$\frac{\mathrm{d}}{\mathrm{d}t}(\pi_1 \circ \gamma) = \frac{\mathrm{d}}{\mathrm{d}t}\sigma.$$

Therefore $\pi_1 \circ \gamma = \sigma$.

Proof of (ii). By Formula (7.3.6), one has

$$\begin{aligned}
\mathrm{Length}(\pi_\circ\gamma) &= \int_0^1 \left\| \frac{\mathrm{d}}{\mathrm{d}t} (\pi_1 \circ \gamma)(t) \right\| dt \\
&= \int_0^1 \|\gamma'(t)\|\, dt \\
&= \int_0^1 \left\|(L_{\gamma(t)})_*^{-1} \dot{\gamma}(t)\right\| dt \\
&= \int_0^1 \|\dot{\gamma}(t)\|\, dt \\
&= \mathrm{Length}(\gamma).
\end{aligned}$$

Proof of (iii). By Theorem 4.2.11 and Lemma 7.3.11(i), one has

$$\begin{aligned}
\pi_1^{(H)} \circ \varphi \circ \gamma &= \mathrm{proj} \circ \exp^{-1} \circ \varphi \circ \gamma \\
&= \mathrm{proj} \circ \varphi_* \circ \exp^{-1} \circ \gamma \\
&= \varphi_* \circ \mathrm{proj} \circ \exp^{-1} \circ \gamma \\
&= \varphi_* \circ \pi_1^{(G)} \circ \gamma.
\end{aligned}$$

□

We have the following formula combining (i) and (iii):

$$(d\varphi)_{\gamma(t)}\dot\gamma(t) = \big((L_{\varphi(\gamma(t))})_* \circ \varphi_*\big)\left(\frac{\mathrm{d}}{\mathrm{d}t}(\pi\circ\gamma)(t)\right).$$

7.3.4 Curves in free nilpotent Lie groups

...

– put a discussion about free Carnot groups, i.e., free nilpotent Lie groups.

...

Proposition 7.3.14. *Assume that each pair of points in G can be joined by a smooth geodesic. If there is a homomorphism $\varphi : G \to H$ such that*

$$\varphi_*|_{V_1^{(G)}} : V_1^{(G)} \to V_1^{(H)}$$

is an isometry (or, more generally, a submetry) of normed spaces, then each pair of points in H can be joined by a smooth geodesic.

In the proposition, the word smooth can be replaced by C^k, C^ω, or piece-wise linear, since the good geodesics in H will be images under φ of good geodesics in G.

Proof. Pick a point $p \in H$ and a geodesic $\xi : [0,1] \to H$ connecting the identity element to p. Then Push the curve on $V_1^{(H)}$ and then back to $V_1^{(G)}$, i.e., consider the curve

$$(\varphi_*)^{-1} \circ \pi_1 \circ \xi.$$

By Lemma 7.3.12(i) consider a curve $\tilde\xi$ such that

$$\pi_1^{(G)} \circ \tilde\xi = (\varphi_*)^{-1} \circ \pi_1^{(H)} \circ \xi.$$

Note that $\varphi(\tilde\xi(1)) = p$. In fact $\varphi\circ\tilde\xi = \xi$. Indeed, $\varphi\circ\tilde\xi$ and ξ are the unique lift of $\pi_1^{(H)}\circ\xi$ under $\pi_1^{(H)}$, since

$$\pi_1^{(H)} \circ \varphi \circ \tilde\xi = \varphi_* \circ \pi_1^{(G)} \circ \tilde\xi = \varphi_* \circ (\varphi_*)^{-1} \circ \pi_1^{(H)} \circ \xi = \pi_1^{(H)} \circ \xi,$$

where we initially used Lemma 7.3.12(iii). Let $\tilde\gamma$ be a smooth geodesic joining e to $\tilde\xi(1)$. We claim that $\varphi\circ\tilde\gamma$ is the desired geodesic. Indeed, using, in order, that φ is 1-Lipschitz (cf. Lemma 7.3.11(iii)), Lemma 7.3.12(ii), the assumption $\varphi_*|_{V_1^{(G)}}$ isometry, and Lemma 7.3.12(ii) again, we get

$$\begin{aligned} L(\varphi \circ \tilde{\gamma}) &\leq L(\tilde{\gamma}) \\ &\leq L(\tilde{\xi}) \\ &= L(\pi_1^{(G)} \circ \tilde{\xi}) \\ &= L((\varphi_*)^{-1} \circ \pi_1^{(H)} \circ \xi) \\ &= L(\pi_1^{(H)} \circ \xi) \\ &= L(\xi) \\ &= d(e, p). \end{aligned}$$

□

Let G and H two nilpotent Lie groups, with horizontal layers $V_1^{(G)}$ and $V_1^{(H)}$, respectively. Consider a homomorphism

$$\varphi : G \to H,$$

such that

$$\varphi_*|_{V_1^{(G)}} : V_1^{(G)} \to V_1^{(H)}$$

and it is an isomorphism. Notice that such a φ is surjective.

Endow just H with a (left-invariant) FCC-metric with $V_1^{(H)}$ as horizontal bundle. In other words, we have fixed a norm $\|\cdot\|_H$ on $V_1^{(H)}$.

Considering that $\varphi_*|_{V_1^{(G)}}$ is an isomorphism, we might consider the following norm $\|\cdot\|_G$ on $V_1^{(G)}$:

$$\|v\|_G := \|\varphi_*(v)\|_H, \qquad \text{for } v \in V_1^{(G)}.$$

Such a norm induces a (left-invariant) FCC-metric with $V_1^{(G)}$ as horizontal bundle.

Then our surjective homomorphism $\varphi : G \to H$ becomes 1-Lipschitz (cf. Lemma 7.3.11(iii)).

Now suppose we know that the problem has a positive answer for G, i.e., that every point in G can be joined to the identity by a piece-wise linear geodesic.

Now pick a point in H and a geodesic connecting this point to the identity. This geodesic lifts to a rectifiable path in G with the same length (because of the choice of lifted FCC on G). Of course here I'm using the fact that since the homomorphism is surjective, the dimension of the abelianisation of M is at least that of N. [1]

[1] Explain more

Now observe that this new path on G is a geodesic, because otherwise there would be another path joining the endpoints of strictly smaller length ; however its projection to H will also be of strictly smaller length, because we said out projection map was 1-Lipschitz, thus contradicting that we had started with a geodesic in H.

Ok, so now we have this lifted path in G and we know it's a geodesic. By assumption we may now find another geodesic, piece-wise linear this time, joining the two points. Then its projection will also be piece-wise linear of course and it will again be a geodesic because once again the projection is 1-Lipschitz.

7.3.5 Open questions

Question 7.3.15. *If ρ is a FCC metric w.r.t. a polyhedral unit ball on G, then does there exist a constant K such that for every p and q there exists a geodesic for ρ joining p and q that has less than K breack points?*

Question 7.3.16. *Let G be a free nilpotent Lie group. If ρ is a FCC metric w.r.t. a strictly convex unit ball on G, then, for every p and q, does there exist a smooth geodesic for ρ joining p and q?*

Question 7.3.17. *Let G be a connected simply connected nilpotent Lie group. If ρ is a G-invariant metric which is coarsely geodesic, i.e.,*

$$d(x, y) \geq L(\gamma_{x,y}) + C.$$

Is ρ at bounded distance from a FCC metric.

7.4 Exercises*

[...]

Chapter 8

Carnot groups*

Tangent spaces of a sub-Riemannian manifold are themselves sub-Riemannian manifolds. They can be defined as metric spaces, using Gromov's definition of tangent spaces to a metric space, and they turn out to be sub-Riemannian manifolds. Moreover, they come with an algebraic structure: nilpotent Lie groups with dilations. In the classical, Riemannian case, they are indeed vector spaces, that is, Abelian groups with dilations. Actually, the above is true only for regular points. At singular points, instead of nilpotent Lie groups one gets quotient spaces G/H of such groups G.

8.1 Definition of Carnot groups*

Let G be a simply connected Lie group. Assume $\mathrm{Lie}(G) = V_1 \oplus \cdots \oplus V_s$ is a stratification. Fix a norm $\|\cdot\|$ on the vector space V_1. The vector space V_1 seen as a subset of T_{1_G} induces a left-invariant subbundle Δ of the tangent bundle TG:

$$\Delta_g := (L_g)_* V_1, \qquad \forall g \in G. \tag{8.1.1}$$

The norm on V_1 induces a norm on every Δ_g as

$$\|v\| := \|(L_g)^* v\|, \qquad \forall v \in \Delta_g, \quad \forall g \in G. \tag{8.1.2}$$

Remark 8.1.3. The triple $(G, \Delta, \|\cdot\|)$ is a subFinsler manifold. Indeed, to see that Δ is bracket generating, take $X \in V_j$ for an arbitrary j. Write X as $\sum_i [X_{i,1}, [X_{i,2}, \dots, X_{i,j}]]$ with $X_{i,k} \in V_1$. If $\tilde{X}_{i,k}$ are left-invariant vector fields extending $X_{i,k}$, then $\tilde{X}_{i,k} \in \Gamma(\Delta)$ and

$$\left(\sum_i [\tilde{X}_{i,1}, [\tilde{X}_{i,2}, \dots, \tilde{X}_{i,j}]] \right)_{1_G} = X.$$

Definition 8.1.4 (Stratified Lie group)**.** We say that a Lie group is *stratified* if it is simply connected and its Lie algebra is stratified.

Definition 8.1.5 (Carnot group)**.** Let G be a stratified group. Let V_1 be the first stratum of the stratification of $\mathrm{Lie}(G)$. Let Δ and $\|\cdot\|$ be defined by (8.1.1) and (8.1.2), respectively. Let d_{CC} be the Carnot-Carathéodory distance associated to Δ and $\|\cdot\|$. Both the subFinsler manifold $(G, \Delta, \|\cdot\|)$ and the metric space (G, d_{CC}) are called *Carnot groups*.

If $\mathrm{Lie}(G) = V_1 \oplus \cdots \oplus V_s$ is the stratification of the Lie algebra of a Carnot group G, then the topological dimension of G. is $n = \sum_i \dim V_i$ and the homogeneous dimension of the subFinsler manifold $(G, \Delta, \|\cdot\|)$ can be expressed as the value

$$Q := \sum_{i=1}^{s} i \dim V_i. \tag{8.1.6}$$

A Carnot group $(G, \Delta, \|\cdot\|)$ is indeed an equiregular Carnot-Carathéodory space. Indeed, one has that, for each j, $\Delta^{[j]}$ is the left-invariant subbundle for which

$$\Delta^{[j]}(1_G) = V_1 \oplus \cdots \oplus V_j.$$

One should observe that another choice of the norm would not change the biLipschitz equivalence class of the sub-Finsler manifold. Namely, if $\|\cdot\|_2$ is another left-invariant Finsler norm on G, then

$$\mathrm{id} : (\mathbb{G}, d_{CC,\|\cdot\|}) \to (\mathbb{G}, d_{CC,\|\cdot\|_2})$$

is globally biLipschitz. So as a consequence of our interest to metric spaces up to biLipschitz equivalence, we may assume that the norm $\|\cdot\|$ is coming from a scalar product $\langle\cdot|\cdot\rangle$.

In the definition of the Carnot-Carathéodory distance only the value of the scalar product on V_1, and not on all $\mathfrak{g}$, is important. Defining a scalar product on V_1 is equivalent to specifying an orthonormal basis of it. So, denoting by m the dimension of V_1, we fix an inner product in V_1 by fixing an orthonormal basis $X_1, \ldots, X_m$ of V_1. This basis of V_1 induces the Carnot-Carathéodory left-invariant distance d in $\mathbb{G}$, which we recall can be defined as follows:

$$d(x,y) := \inf \left\{ \int_0^1 \sqrt{\sum_{i=1}^{m} |a_i(t)|^2}\, dt : \ \gamma(0) = x, \ \ \gamma(1) = y \right\},$$

where the infimum is among all absolute continuous curves $\gamma : [0,1] \to \mathbb{G}$ such that $\dot\gamma(t) = \sum_1^m a_i(t)(X_i)_{\gamma(t)}$ for a.e. $t \in [0,1]$ (the so-called horizontal curves).

8.1.1 Dilations on Carnot groups

Definition 8.1.7 (Dilations on stratified groups)**.** Let G be a stratified group. Let $\delta_\lambda : \mathrm{Lie}(G) \to \mathrm{Lie}(G)$ be the dilation of factor λ associated to the stratification. Then the *dilation* $\delta_\lambda : G \to G$ of the group of factor λ is the only group automorphism such that $(\delta_\lambda)_* = \delta_\lambda$.

Such maps are also called the *intrinsic dilations* of the stratified group.

We have kept the same notation δ_λ for both dilations (in $\mathfrak{g}$ and in $\mathbb{G}$) because no ambiguity will arise since the two maps have different domains.

Remark 8.1.8. From Theorem **??** the above map is well defined since by assumption a stratified group is simply connected. Moreover, from Theorem 4.2.11 we have

$$\delta_\lambda \circ \exp = \exp \circ \delta_\lambda. \tag{8.1.9}$$

In fact, since stratified groups have nilpotent Lie algebras, the map $\exp : \mathfrak{g} \to \mathbb{G}$ is a diffeomorphism by Theorem 4.6.4, so every element $g \in \mathbb{G}$ can represented as $\exp(X)$ for some unique $X \in \mathfrak{g}$, and therefore uniquely written in the form

$$\exp\left(\sum_{i=1}^{s} v_i\right), \qquad v_i \in V_i,\ 1 \le i \le s. \tag{8.1.10}$$

This representation allows to have the formula:

$$\delta_\lambda\left(\exp\left(\sum_{i=1}^{s} v_i\right)\right) = \exp\left(\sum_{i=1}^{s} \lambda^i v_i\right). \tag{8.1.11}$$

We have the formula

$$\delta_\lambda \circ \delta_\eta = \delta_{\lambda\eta}. \tag{8.1.12}$$

We remark that one can use Baker-Campbell-Hausdorff formula to show that if one defines $\delta_\lambda : \mathbb{G} \to \mathbb{G}$ by (8.1.11), then such maps are group homomorphisms, i.e.,

$$\delta_\lambda(xy) = \delta_\lambda(x)\delta_\lambda(y) \qquad \forall x,\, y \in \mathbb{G}. \tag{8.1.13}$$

The fact that by definition we have $\delta_\lambda X = (\delta_\lambda)_* X$ says that

$$X(u \circ \delta_\lambda)(g) = (\delta_\lambda X)u(\delta_\lambda g) \qquad \forall X \in \mathrm{Lie}(G), \forall u \in C^\infty(G), \forall \lambda \ge 0, \forall g \in \mathbb{G}. \tag{8.1.14}$$

Such relation (8.1.14) between dilations in $\mathbb{G}$ and dilations in $\mathfrak{g}$ can also be directly shown using (8.1.9) as definition for the group dilation:

$$\begin{aligned} X(u \circ \delta_\lambda)(g) &= \frac{\mathrm{d}}{\mathrm{d}t} u \circ \delta_\lambda(g \exp(tX))\Big|_{t=0} = \frac{\mathrm{d}}{\mathrm{d}t} u(\delta_\lambda g \delta_\lambda \exp(tX))\Big|_{t=0} \\ &= \frac{\mathrm{d}}{\mathrm{d}t} u(\delta_\lambda g \exp(t\delta_\lambda X))\Big|_{t=0} = (\delta_\lambda X)u(\delta_\lambda g). \end{aligned}$$

Relations between dilations and CC distances

The Carnot-Carathéodory distance is well-behaved under the intrinsic dilations, in the sense that such dilations multiply distances of a constant factor.

Proposition 8.1.15. *If* (G, d_{CC}) *is a Carnot group with dilations* $(\delta_\lambda)_{\lambda\in\mathbb{R}}$*, then*

$$d_{CC}(\delta_\lambda p, \delta_\lambda q) = \lambda d_{CC}(p,q), \qquad \forall p,\, q \in \mathbb{G}, \forall \lambda \in \mathbb{R}. \tag{8.1.16}$$

Proof. Since $\delta_\lambda|_{V_1}$ is the multiplication by λ, we have that $\|\delta_\lambda v\| = \lambda\|v\|$, for all $v \in \Delta$. If γ in a horizontal curve from x to y, then $\delta_\lambda \circ \gamma$ is a curve going from $\delta_\lambda x$ to $\delta_\lambda y$ whose tangent vectors are, for almost all t,

$$(\delta_\lambda)_* \dot\gamma(t) = \delta_\lambda(\dot\gamma(t)) = \lambda\dot\gamma(t), \tag{8.1.17}$$

which are horizontal since $\dot\gamma(t)$ is horizontal. Moreover, from (8.1.17), the length of $\delta_\lambda \circ \gamma$ is λ times the length of γ, i.e., for all horizontal curve γ,

$$\mathrm{Length}_{\|\cdot\|}(\delta_\lambda \circ \gamma) = \lambda\, \mathrm{Length}_{\|\cdot\|}(\gamma).$$

Thus 8.1.16 has been shown. □

Exercise 8.1.18. Show that for all $p \in G$, for all $r > 0$

$$B_{d_{CC}}(p,r) = L_p(\delta_r(B_{d_{CC}}(1_G, 1))).$$

■

8.1.2 Good bases for Carnot groups

Let $\mathbb{G}$ be a Carnot group with stratification $\mathfrak{g} = V_1 \oplus \cdots \oplus V_s$. We want to construct a basis for $\mathfrak{g}$ that is structured with respect to the stratification, is a Malcev basis, and each element of the basis that is not in V_1, is the bracket of two vectors of such a basis.

Start by picking a basis $X_1, \dots, X_m$ of V_1. Then consider all brackets $[X_i, X_j]$, for $i, j = 1, \dots, m$. Since $[V_1, V_1] = V_2$, we can find among such brackets a basis for V_2, cf. Exercise 8.7.10. Pick one such a basis and call the elements $X_{m+1}, \dots, X_{m_2}$. Iterate the method: extract a basis $X_{m_2+1}, \dots, X_{m_3}$ of V_3 from the set $[X_i, X_j]$, for $i = 1, \dots, m, j = m+1, \dots, m_2$. And so on. In such a way we constructed a basis $X_1, \dots, X_n$ of $\mathfrak{g}$ and natural numbers $m_1 \dots, m_s$ such that

1. $X_{m_{j-1}+1}, \dots, X_{m_j}$ is a basis of V_j,

2. For every $i = m+1, \ldots, n$, there exist d_i, l_i, and k_i such that $X_i \in V_{d_i}$, $X_{l_i} \in V_1$, $X_{k_i} \in V_{d_i - 1}$, and

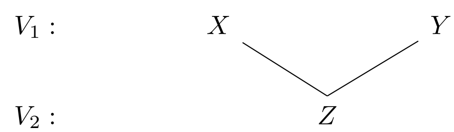
$$X_i = [X_{l_i}, X_{k_i}]. \tag{8.1.19}$$

3. The order-reversed basis $X_n, \ldots, X_1$ is a (strong) Malcev basis; in other words,

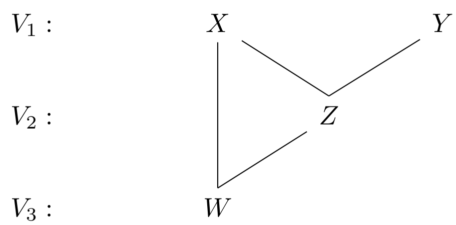
$$[\mathfrak{g}, \operatorname{span}\{X_k, \ldots, X_n\}] \subseteq \operatorname{span}\{X_{k+1}, \ldots, X_n\}.$$

We would suggest the terminology '*Carnot basis*' for a basis satisfying the above three conditions. The reader should notice that the above property 1 implies the property 3. See Exercise 8.7.11.

To describe a Carnot algebra we prefer to give a Carnot basis as a hierarchical diagram as

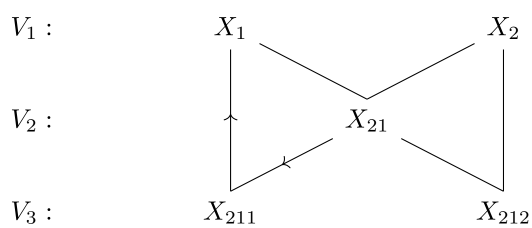


for the 3D Heisenberg algebra.



for the Engel algebra

(the step-3 filiform algebra).



for the Cartan algebra, in Hall's basis

(the the rank 2 and step 3 algebra).

The bracket relation expressed in the diagram should be read *from left to right*, unless there is an arrow *from right to left*.

The j-th line in the diagram list the vectors that span V_j. The black lines express the non-trivial brackets. However, one should notice that in the algebra structure might be more relations than just those in (8.1.19). (Give an example!)

8.1.3 Examples of Carnot groups

Engel in both coordinates

Cartan

Filiform groups

8.1.4 Use of dilations and canonical coordinates

Since every Carnot group is nilpotent and simply connected, the map $\exp \mathfrak{g} \to \mathbb{G}$ is a global diffeomorphism, cf. Theorem 6.5.1. Therefore the exponential coordinates are global (and one-to-one) coordinates. As a consequence, the dilations $\delta_\lambda : \mathbb{G} \to \mathbb{G}$ are well-defined. From them one has that such self-similar homomorphisms extend properties that hold in a neighborhood of the identity to the whole of $\mathbb{G}$. As an example, let us show the fact that Malcev coordinates maps are injective and surjective.

Proposition 8.1.20. *On every Carnot group, Malcev coordinates exist.*

Proof. The fact that a Malcev basis $X_1, \ldots, X_n$ exist was shown in the previous subsection. Now consider the coordinate map

$$\Psi : (s_1 \ldots, s_n) \to \exp(s_1 X_1) \cdots \exp(s_n X_n).$$

Obviously

$$(d\Psi)_0 \partial_j = \frac{\mathrm{d}}{\mathrm{d}s_j} \exp(s_j X_j)\Big|_{s_j=0} X_j,$$

so $(d\Psi)_0$ is an invertible $n \times n$ matrix. Thus Ψ is open at zero, i.e., $\Psi(\mathbb{R}^n)$ is a neighborhood of the identity e. Let us show that $\Psi(\mathbb{R}^n) = \mathbb{G}$. Take $p \in \mathbb{G}$. Then there exists some $\lambda \in \mathbb{R}$ and some $\mathbf{s} \in \mathbb{R}^n$ such that

$$\delta_\lambda^{-1}(p) = \Psi(\mathbf{s}).$$

Let $\tilde{\mathbf{s}} = \delta_\lambda(\mathbf{s})$. Then, since δ_λ on $\mathbb{G}$ is a group homomorphism, we have

$$\begin{aligned}
\Psi(\tilde{\mathbf{s}}) &= \exp(\delta_\lambda(s_1 X_1)) \cdots \exp(\delta_\lambda(s_n X_n)) \\
&= \delta_\lambda(\exp(s_1 X_1)) \cdots \delta_\lambda(\exp(s_n X_n)) \\
&= \delta_\lambda(\exp(s_1 X_1) \cdots \exp(s_n X_n)) \\
&= \delta_\lambda \Psi(\mathbf{s}) \\
&= p.
\end{aligned}$$

Let us show injectivity. Since $(d\Psi)_0$ is an invertible $n \times n$ matrix, then by the Inverse Function Theorem there is a neighborhood U on which Ψ is injective. Assume now that there are $\mathbf{s}_1, \mathbf{s}_2 \in \mathbb{R}^n$ such that

$$\Psi(\mathbf{s}_1) = \Psi(\mathbf{s}_2).$$

Then, for $\lambda \in \mathbb{R}$ small enough, we have $\delta_\lambda(\mathbf{s}_1), \delta_\lambda(\mathbf{s}_2) \in U$. By the above calculation, we have that

$$\Psi(\delta_\lambda(\mathbf{s}_1)) = \Psi(\delta_\lambda(\mathbf{s}_2)).$$

But, since Ψ is injective on U, we have $\delta_\lambda(\mathbf{s}_1) = \delta_\lambda(\mathbf{s}_2)$, and therefore $\mathbf{s}_1 = \mathbf{s}_2$. □

8.2 Chow's Theorem and Ball-Box Theorem*

8.2.1 A direct, effective proof of Chow's theorem

We will give now an explicit construction of an horizontal path connecting an arbitrary point p in a Carnot group to the origin 1. The reader should remind the elementary fact, cf. Proposition 4.2.6, that the curve pe^{tX} is the integral curve of X starting at p.

Brackets as products of exponentials

The philosophy behind the following discussion is that to go in a direction given as a bracket of two vector fields one can go along a not-necessarily-closed quadrilateral constructed using the flows of the two vector fields. We will give a generalization of the following formula with which the reader should be already familiar:

$$[X,Y] = \frac{\mathrm{d}^2}{2\mathrm{d}^2 t} e^{-tY} \circ e^{-tX} \circ e^{tY} \circ e^{tX}\Big|_{t=0} = \frac{\mathrm{d}}{\mathrm{d}t} e^{-\sqrt{t}Y} \circ e^{-\sqrt{t}X} \circ e^{\sqrt{t}Y} \circ e^{\sqrt{t}X}\Big|_{t=0}.$$

In the above formula, e^{tX} denotes the flow map of a general vector field on a manifold. So for left-invariant vector fields in a Lie group we have

$$e^{tX}(p) = pe^{tX}.$$

Thus the order might seems reversed.

For $X, Y \in \mathfrak{g}$ and $t \in \mathbb{R}$ define

$$P_t(X,Y) := e^{tX} e^{tY} e^{-tX} e^{-tY}.$$

Using twice the BCH formula one has that, for $t \to 0$,

$$P_t(X,Y) = e^{t^2[X,Y]+o(t^2)}.$$

Suppose we have defined by induction the function $P_t(X_1, \dots, X_k)$, for $k \geq 2$, define then

$$P_t(X_1, \dots, X_{k+1}) := P_t(X_1, \dots, X_k) e^{tX_{k+1}} (P_t(X_1, \dots, X_k))^{-1} e^{-tX_{k+1}}.$$

By induction we shall show that, as $t \to 0$,

$$P_t(X_1, \ldots, X_k) = e^{t^k [\ldots[[X_1,X_2],X_3],\ldots,X_k] + o(t^k)}. \tag{8.2.1}$$

The case $k = 2$ has been already mentioned above, and its proof is similar to the induction step. Assume it true for an arbitrary k. Call $\omega(t)$ the $o(t^k)$ function such that $P_t(X_1, \ldots, X_k) = e^{t^k [\ldots[[X_1,X_2],X_3],\ldots,X_k] + \omega(t)}$. Then we have, by the BCH formula,

$$\begin{aligned}
P_t(X_1, \ldots, X_{k+1}) &= P_t(X_1, \ldots, X_k) e^{tX_{k+1}} \left(P_t(X_1, \ldots, X_k)\right)^{-1} e^{-tX_{k+1}} \\
&= e^{t^k [\ldots[X_1,X_2],\ldots,X_k] + \omega(t)} e^{tX_{k+1}} \left(e^{t^k [\ldots[X_2,X_1],\ldots,X_k] + \omega(t)}\right)^{-1} e^{-tX_{k+1}} \\
&= e^{t^k [\ldots[X_1,X_2],\ldots,X_k] + \omega(t)} e^{tX_{k+1}} e^{-t^k [\ldots[X_2,X_1],\ldots,X_k] - \omega(t)} e^{-tX_{k+1}} \\
&= e^{\left(tX_{k+1} + t^k [\ldots[X_1,X_2],\ldots,X_k] + \omega(t) + \frac{1}{2} t^{k+1} [\ldots[X_1,X_2],\ldots,X_{k+1}] + o(t^{k+1})\right)} \cdot \\
&\qquad e^{\left(-tX_{k+1} - t^k [\ldots[X_1,X_2],\ldots,X_k] - \omega(t) + \frac{1}{2} t^{k+1} [\ldots[X_1,X_2],\ldots,X_{k+1}] + o(t^{k+1})\right)} \\
&= e^{t^{k+1} [\ldots[[X_1,X_2],X_3],\ldots,X_{k+1}] + o(t^{k+1})}.
\end{aligned}$$

One should note that each P_t is in fact a product of element of the form $e^{\pm tX_i}$. Thus the following properties are immediate:

$$P_{\lambda t}(X_1, \ldots, X_k) = P_t(\lambda X_1, \ldots, \lambda X_k), \tag{8.2.2}$$

$$\delta_\lambda P_t(X_1, \ldots, X_k) = P_t(\delta_\lambda X_1, \ldots, \delta_\lambda X_k). \tag{8.2.3}$$

We construct now a map that will help in constructing horizontal paths. Consider a Carnot basis $X_1, \ldots, X_n$, so in particular property (8.1.19) holds. Iterating such property, we have that each element X_j of the basis is such that

$$X_j = [\ldots[[X_{j,1}, X_{j,2}], X_{j,3}], \ldots, X_{j,d_j}],$$

where the basis elements $X_{j,1}, \ldots, X_{j,d_j}$ are in V_1, and d_j is such that $X_j \in V_{d_j}$, in other words, it is the degree of X_j.

For each j, we consider the expression

$$P^{(j)}(t) := P_t(X_{j,1}, \ldots, X_{j,d_j}).$$

In the following we will use the notation $t^\alpha = \mathrm{sgn}(t)|t|^\alpha$, so for example we have $\sqrt{-4} = -2$. We finally define the map

$$E(\mathbf{t}) := P^{(1)}(\sqrt[d_1]{t_1}) \cdots P^{(n)}(\sqrt[d_n]{t_n}).$$

E.g., for the standard basis in the Heisenberg group we get:

$$E(\mathbf{t}) = e^{t_1 X} e^{t_2 Y} e^{\sqrt{t_3} X} e^{\sqrt{t_3} Y} e^{-\sqrt{t_3} X} e^{-\sqrt{t_3} Y}.$$

For the standard basis in the Engel group we get:

$$\begin{aligned} E(\mathbf{t}) \quad = \quad & e^{t_2 X}\, e^{t_2 Y}\, e^{\sqrt{t_3} X} e^{\sqrt{t_3} Y} e^{-\sqrt{t_3} X} e^{-\sqrt{t_3} Y} \\ & e^{\sqrt[3]{t_4} X}\, e^{\sqrt[3]{t_4} Y} e^{-\sqrt[3]{t_4} X} e^{-\sqrt[3]{t_4} Y} e^{\sqrt[3]{t_4} X} e^{\sqrt[3]{t_4} Y} e^{\sqrt[3]{t_4} X} e^{-\sqrt[3]{t_4} Y} e^{-\sqrt[3]{t_4} X}\, e^{-\sqrt[3]{t_4} X}. \end{aligned}$$

We will show in order that such a map E satisfies the following three properties.

Proposition 8.2.4. *Let E be the map defined above.*

1. *$E : \mathbb{R}^n \to \mathbb{G}$ is open at $\mathbf{0}$.*

2. *E is surjective.*

3. *E gives a natural horizontal path from $\mathbf{0}$ to $E(\mathbf{t})$.*

The second property follows easily from the first one using dilations. The third is also very elementary since flows of left-invariant vector fields are right multiplications by exponentials. The first is a consequence of the interpretation of the bracket as product of exponential.

Proof of Property 1 of Proposition 8.2.4. We just need to show that $(dE)_0$ is a non-singular matrix. From how E has been defined and from (8.2.1), we have

$$\begin{aligned} (dE)_0 \partial_j \quad & = \quad \frac{\mathrm{d}}{\mathrm{d}t_j} E(\mathbf{t}) \bigg|_{\mathbf{t}=0} \\ & = \quad \frac{\mathrm{d}}{\mathrm{d}t_j} P^{(j)}(\sqrt[d_j]{t_j}) \bigg|_{t_j=0} \\ & = \quad \frac{\mathrm{d}}{\mathrm{d}t} P_{\sqrt[d_j]{t}}(X_{j,1}, \ldots, X_{j,d_j}) \bigg|_{t=0} \\ & = \quad \frac{\mathrm{d}}{\mathrm{d}t} e^{t[\ldots[[X_{j,1}, X_{j,2}], X_{j,3}], \ldots, X_{j,d_j}] + o(t)} \bigg|_{t=0} \\ & = \quad \frac{\mathrm{d}}{\mathrm{d}t} e^{t X_j + o(t)} \bigg|_{t=0} \\ & = \quad X_j. \end{aligned}$$

In other words, $(dE)_0$ sends the basis $\partial_1, \ldots, \partial_n$ to the basis $X_1, \ldots, X_n$. Property 1 follows from the Inverse Function Theorem.

Proof of Property 2 of Proposition 8.2.4. By Property 1, the set $E(\mathbb{R}^n)$ is a neighborhood of e. On the other hand for each fixed point $q \in \mathbb{G}$, the dilations δ_λ of the Carnot group have the property that $\lim_{\lambda\to 0} \delta_\lambda(q) = 1$. From these two facts we have that, for each $p \in \mathbb{G}$, there are $\lambda \in \mathbb{R}$ and $\mathbf{t} \in \mathbb{R}^n$ such that

$$\delta_\lambda(E(\mathbf{t})) = p.$$

Now, let $\tilde{\mathbf{t}} = \delta_\lambda(\mathbf{t})$, i.e., $\tilde{t}_j = \lambda^{d_j} t_j$. First by the properties (8.2.2) and (8.2.3) on P_t, and the fact that $X_{j,1}, \ldots, X_{j,d_j}$ are in V_1, one has

$$\begin{aligned}
P^{(j)}(\sqrt[d_j]{\tilde{t}_j}) &= P^{(j)}(\sqrt[d_j]{\lambda^{d_j} t_j}) \\
&= P^{(j)}(\lambda \sqrt[d_j]{t_j}) \\
&= P_{\lambda \sqrt[d_j]{t_j}}(X_{j,1}, \ldots, X_{j,d_j}) \\
&= P_{\sqrt[d_j]{t_j}}(\lambda X_{j,1}, \ldots, \lambda X_{j,d_j}) \\
&= P_{\sqrt[d_j]{t_j}}(\delta_\lambda(X_{j,1}), \ldots, \delta_\lambda(X_{j,d_j})) \\
&= \delta_\lambda \left(P_{\sqrt[d_j]{t_j}}(X_{j,1}, \ldots, X_{j,d_j}) \right) \\
&= \delta_\lambda P^{(j)}(\sqrt[d_j]{t_j}).
\end{aligned}$$

Then, since δ_λ on $\mathbb{G}$ is a group homomorphism, one get

$$\begin{aligned}
E(\tilde{\mathbf{t}}) &= P^{(1)}(\sqrt[d_1]{\tilde{t}_1}) \cdots P^{(n)}(\sqrt[d_n]{\tilde{t}_n}) \\
&= \delta_\lambda(P^{(1)}(\sqrt[d_1]{t_1})) \cdots \delta_\lambda(P^{(n)}(\sqrt[d_n]{t_n})) \\
&= \delta_\lambda \left(P^{(1)}(\sqrt[d_1]{t_1}) \cdots P^{(n)}(\sqrt[d_n]{t_n}) \right) \\
&= \delta_\lambda E(\mathbf{t}) \\
&= p.
\end{aligned}$$

Thus $E(\mathbb{R}^n)$ is in fact the whole of $\mathbb{G}$, i.e., E is surjective.

Proof of Property 3 of Proposition 8.2.4. Recall, cf. Proposition 4.2.6, that the flow lines of a left-invariant vector field X are the curves ge^{tX}, fixed $g \in \mathbb{G}$ and varying $t \in \mathbb{R}$. Now, since P_t is a product of exponentials, then E is too. More explicitly, fixed $\mathbf{t} \in \mathbb{R}^n$, we have

$$E(\mathbf{t}) = \exp(\xi_1 t_{\gamma_1}^{\alpha_1} X_{\beta_1}) \cdots \exp(\xi_N t_{\gamma_N}^{\alpha_N} X_{\beta_N}),$$

for $\xi_i \in \{1, -1\}$, $\alpha_i^{-1} \in \mathbb{N}$, $\beta_i \in \{1, \ldots, m\}$, $\gamma_i \in \{1, \ldots n\}$, and $N \in \mathbb{N}$. Now it is enough to observe that, fixed K, the point

$$g := \exp(\xi_1 t_{\gamma_1}^{\alpha_1} X_{\beta_1}) \cdots \exp(\xi_K t_{\gamma_K}^{\alpha_K} X_{\beta_K})$$

can be connected to the point

$$\exp(\xi_1 t_{\gamma_1}^{\alpha_1} X_{\beta_1}) \cdots \exp(\xi_K t_{\gamma_K}^{\alpha_K} X_{\beta_K}) \exp(\xi_{K+1} t_{\gamma_{K+1}}^{\alpha_{K+1}} X_{\beta_{K+1}})$$

by the path

$$g \exp(\xi_{K+1} s X_{\beta_{K+1}}), \quad \text{for } s \in [0, |t_{\gamma_{K+1}}^{\alpha_{K+1}}|],$$

which is tangent to $\pm X_{\beta_{K+1}}$, thus horizontal. □

Corollary 8.2.5 (Chow's theorem for Carnot groups)**.** *Any point $p \in \mathbb{G}$ in a Carnot group can be joined to the identity e by a horizontal path. Moreover, the CC-distance induces the manifold topology.*

Proof. Property 2 and 3 of Proposition 8.2.4 give the existence of a path from e to any given point p. Thus $d_{CC}(1_{\mathbb{G}}, p) < \infty$, for all $p \in \mathbb{G}$. By left invariance of d_{CC} we have $d_{CC}(p, q) < \infty$, for all $p, q \in \mathbb{G}$.

Since E is in fact open at $\mathbf{0}$, by Property 1 of Proposition 8.2.4, then points close to the origin can be connected to the origin by short horizontal curves. □

8.2.2 A proof of Ball-Box Theorem for Carnot groups

Let (G, d_{CC}) be a Carnot group and let $V_1, \ldots, V_s$ be a stratification of Lie(G),. Let $X_1, \ldots, X_n$ be a basis of Lie(G) adapted to the stratification i.e., for all j there exists d_j such that $X_j \in V_{d_j}$.

The the number d_j is called *degree* of X_j and it may be denoted by $\deg(X_j)$. The *box* with respect to the fixed basis $X_1, \ldots, X_n$ is defined as

$$\texttt{Box}(r) := \{(t_1, \ldots, t_n) \in \mathbb{R}^n \;:\; |t_j| < r^{d_j}\}$$

Let $\delta_\lambda : \mathbb{R}^n \to \mathbb{R}^n$ be the map

$$\delta_\lambda(t_1, \ldots, t_n) = (\lambda t_1, \ldots, \lambda^{d_j} t_j, \ldots, \lambda^s t_n).$$

Exercise 8.2.6. Show that for all $r, \lambda > 0$

$$\delta_\lambda(\texttt{Box}(r)) = \texttt{Box}(\lambda r).$$

■

Let $\Phi : \mathbb{R}^n \to G$ be the exponential coordinate map with respect to the basis $X_1, \ldots, X_n$, i.e., $\Phi(\mathbf{t}) = \exp(\sum_j t_j X_j)$. Then we have that $\Phi(\mathtt{Box}(1))$ is a bounded neighborhood of e in G. (Notice that this last fact holds since Φ is a diffeomorphism, however it is just a consequence of the fact that the differential at 0 of Φ is the identity and hence Φ is a local diffeomorphism in a neighborhood of the identity)

Let d_{CC} be the Carnot-Carathéodory distance of the Carnot group G. Since V_1 bracket generates Lie(G), by Chow Theorem 8.2.5 the distance d_{CC} induces the manifold topology. Hence, there is $C > 1$ such that

$$B(1_G, \frac{1}{C}) \subset \Phi(\mathtt{Box}(1)) \subset B(1_G, C),$$

where $B(1_G, r)$ is the CC-ball of center the identity element 1_G and radius r. Recalling that that $\delta_\lambda(B(1_G, r)) = B(1_G, \lambda r)$ and applying δ_λ, we get

$$B(1_G, \frac{\lambda}{C}) \subset \delta_\lambda \Phi(\mathtt{Box}(1)) \subset B(1_G, \lambda C),$$

where

$$\begin{aligned}
\delta_\lambda\left(\Phi(\mathtt{Box}(1))\right) &= \delta_\lambda\left(\Phi\left\{(t_1, \ldots, t_n) : |t_j| < 1\right\}\right) \\
&= \delta_\lambda \left\{ \exp\left(\sum_j t_j X_j\right) : |t_j| < 1 \right\} \\
&= \left\{ \exp\left(\delta_\lambda \sum_j t_j X_j\right) : |t_j| < 1 \right\} \\
&= \left\{ \exp\left(\sum_j \lambda^{d_j} t_j X_j\right) : |t_j| < 1 \right\} \\
&= \left\{ \exp\left(\sum_j s_j X_j\right) : |s_j| < \lambda^{d_j} \right\} \\
&= \Phi(\mathtt{Box}(\lambda)).
\end{aligned}$$

Therefore, we conclude that

$$B(1_G, \frac{\lambda}{C}) \subset \Phi(\mathtt{Box}(\lambda)) \subset B(1_G, \lambda C), \qquad \forall \lambda > 0. \tag{8.2.7}$$

Theorem 8.2.8 (Ball-Box for Carnot groups)**.** *Let G be a Carnot froup. Fix a basis adapted to the stratification $V_1, \ldots, V_s$. Then there is $C > 1$ such that for all $p \in G$ and all $r > 0$*

$$B(p, \frac{\lambda}{C}) \subset \Phi_p(\mathtt{Box}(\lambda)) \subset B(p, \lambda C), \tag{8.2.9}$$

where Φ_p is the exponential coordinate map from p with respect to the fixed basis.

Proof. By the definition of Φ_p we have

$$\Phi_p(\mathbf{t}) = p \exp(\sum t_j X_j) = L_p \Phi(\mathbf{t}).$$

Since d_{CC} is left-invariant, applying L_p to (8.2.7), we obtain (8.2.9) for all $p \in G$ and all $\lambda > 0$. □

8.3 Canonical measures*

We shall see that in a Carnot group there are few natural choices of measures: Haar, Hausdorff and Lebesgue measures. Up to a scalar fact they will be the same. In this section we se some of their properties.

Carnot groups are nilpotent and so unimodular, therefore right- and left-Haar measures coincide, up to constant multiples. We fix one of them and denote it by $\mathrm{vol}_{\mathbb{G}}$.

For every $k > 0$, the k-dimensional Hausdorff measure $\mathscr{H}^k$ and the k-dimensional spherical Hausdorff measure $\mathscr{S}^k$ are left-invariant.

We shall see that for $k = Q$ these measures are Radon measures, and therefore are Hausdorff measures, so a multiple of $\mathrm{vol}_{\mathbb{G}}$. We shall actually show that in exponential coordinates, all these measures are a constant multiple of the Lebesgue measure, which in $\mathbb{R}^n$ is denoted by $\mathcal{L}^n$.

Definition 8.3.1 (Homogeneous dimension for a Carnot group)**.** If G is a stratified group and $V_1, \ldots, V_s$ is the stratification of its Lie algebra, we call

$$Q := \sum_{j=1}^{s} j \cdot \dim V_j$$

the *homogeneous dimension* of G.

Exercise 8.3.2. Show that this notion of homogeneous dimension agrees with the one on subFinsler manifolds. ■

Proposition 8.3.3. *Let G be a Carnot group of homogeneous dimension Q.*

1. *If* vol *is a Haar measure of G, then*

$$\mathrm{vol}(B(p,r)) = r^Q \, \mathrm{vol}(B(1_G)).$$

2. *The Hausdorff dimension of G is Q.*

3. *In exponential coordinates, the Lebesgue measure is the Hausdorff Q-measure up to a multiplication by a constant.*

Proof. In exponential coordinates, the Lebesgue measure $\mathcal{L}^n$ is both left and right-invariant, soevery other Haar measure is a multiple of it. In exponential coordinate, the inhomogeneous dilations δ_λ have Jacobian λ^Q, i.e., $\lambda^Q \cdot \mathcal{L}^n(\mathtt{Box}(1)) = \mathcal{L}^n(\mathtt{Box}(\lambda))$. Hence

$$\mathcal{L}^n(B(p,\lambda)) = \mathcal{L}^n(B(1_G,\lambda)) = \mathcal{L}^n(\delta_\lambda(B(1_G,1))) = \lambda^Q \mathcal{L}^n(B(1_G,1))$$

By an early proposition, the Hausdorff dimension is Q. The last part follows since both $\mathcal{L}^n$ and the Hausdorff Q-measure are both Haar measures. □

Exercise 8.3.4. Show that, if X_i is a Carnot basis, then for some constant c we have

$$\mathrm{vol}_G\big(\{\exp(\sum_{i=1}^n x_i X_i) : \ (x_1,\dots,x_n) \in A\}\big) = c\mathcal{L}^n(A) \qquad \text{for all Borel sets } A \subseteq \mathbb{R}^n.$$

■

Exercise 8.3.5. Prove that

$$\mathrm{vol}_{\mathbb{G}}(\delta_\lambda(A)) = \lambda^Q \,\mathrm{vol}_{\mathbb{G}}(A) \tag{8.3.6}$$

for all Borel sets $A \subseteq \mathbb{G}$.

■

8.4 Geodesics in step-2 subRiemannian Carnot groups*

[...] Write a presentation of step-2 groups [...]

Explain geodesics in Riemannian/Carnot step-2 nilpotent Lie groups [...]

[...] Prepare to prove that the asymptotic cone is at bounded distance [...]

Proposition 8.4.1. *The only infinite geodesics in sub-Riemannian Carnot groups of step 2 are the horizontal lines.*

Proof. Let $\gamma : \mathbb{R} \to G$ be an infinite geodesic in a rank r step 2 Carnot group G. By lifting γ, we may assume that G is the free Carnot group of rank r and step 2.

In step 2 Carnot groups, every geodesic is normal, so γ satisfies **??** for some pair $(\lambda, 1)$. For normal geodesics, **??** can be rewritten as an ODE for γ by renormalizing so that $\xi = 1$. In step 2 Carnot groups, the ODE is affine, and in the specific case of a free Carnot group of step 2 we get the following form:

Decompose $\lambda = \lambda_H + \lambda_V \in V_1^* + V_2^*$ and fix an orthonormal basis of V_1. Then the horizontal projection $\pi \circ \gamma$ of the curve satisfies the ODE

$$\dot{x} = A_{\lambda_V} x + \lambda_H^*,$$

where $A_{\lambda_V} \in \mathfrak{so}(r)$ is a skew-symmetric matrix whose elements are (up to sign) the components of the vertical part λ_V, and $\lambda_H^* \in V_1$ is the dual of $\lambda_H \in V_1^*$ with respect to the sub-Riemannian inner product. By linearity we can translate the curve γ by some element $g \in G$ such that the projection $\pi(g \cdot \gamma) = \pi(g) + \pi \circ \gamma$ satisfies the ODE

$$\dot{x} = A_{\lambda_V} x + b_{\lambda_H}, \tag{8.4.2}$$

where $b_{\lambda_H} \in V_1$ is the projection of λ_H^* to the orthogonal complement of $\mathrm{Im}(A_{\lambda_V}) \subset V_1$. Furthermore, renormalizing the ξ component given by Lemma **??**, we see that the horizontal projection of a dilation $\gamma_h := \delta 1/h \circ (g \cdot \gamma) \circ \delta h$ satisfies a similar ODE, where λ is replaced by $\frac{1}{h}\delta h^* \lambda = \lambda_H + h\lambda_V$. Explicitly, since the matrix A_{λ_V} depends linearly on λ_V, we have

$$\dot{x} = A_{h\lambda_V} x + b_{\lambda_H} = hA_{\lambda_V} x + b_{\lambda_H}.$$

The solution of the above with the initial condition $x(0) = \pi \circ \gamma_h(0) = \frac{1}{h}\pi(g\gamma(0))$ is

$$x(t) = \frac{1}{h} e^{hA_{\lambda_V} t} \pi(g\gamma(0)) + b_{\lambda_H} t. \tag{8.4.3}$$

Consider any blowdown of the curve $g \cdot \gamma$, i.e., a limit $\sigma = \lim\limits_{j\to\infty} \gamma_{h_j}$ along some sequence $h_j \to 0$. By independence from the basepoint of a blowdown, σ is also a blowdown of γ for the same sequence h_j. Taking the limit of (8.4.3) as $h_j \to \infty$, we see that the limit curve is the line $\sigma(t) = b_{\lambda_H} t$.

Since γ is a geodesic, the ODE (8.4.2) implies that $\|A_{\lambda_V} x + b_{\lambda_H}\|^2 = 1$. On the other hand, the vector b_{λ_H} is by construction orthogonal to $\mathrm{Im}(A_{\lambda_V})$, so for any point $x = \pi(g) + \pi(\gamma(t))$, $t \in \mathbb{R}$, we have

$$1 = \|A_{\lambda_V} x + b_{\lambda_H}\|^2 = \|A_{\lambda_V} x\|^2 + \|b_{\lambda_H}\|^2.$$

That is, either $A_{\lambda_V} x = 0$ for all $x = \pi(g) + \pi(\gamma(t))$, in which case the ODE (8.4.2) implies that γ is a line, or $\|b_{\lambda_H}\| < 1$. But in the latter case we would have

$$\|\dot{\sigma}\| = \|b_{\lambda_H}\| < 1,$$

so the blowdown σ would not be parametrized with unit speed. This would contradict the assumption that γ is an infinite geodesic, so we see that γ must be a line. $\square$

Remark 8.4.4. Proposition 8.4.1 can be used to prove that in fact any isometric embedding of any Carnot group into any sub-Riemannian Carnot group of step 2 is affine. This follows by replicating the proof of [**?**, Theorem 1.1] that all infinite geodesics being lines is sufficient to conclude that arbitrary isometric embeddings from other Carnot groups are affine. Although the result of Balogh, Fässler, and Sobrino is stated in the setting of Heisenberg groups, their proof (with only superficial modifications) works also in the general setting of arbitrary step 2 Carnot groups.

8.5 Abnormal curves in Carnot groups*

8.5.1 A distinguished class of polynomials

For $\lambda \in \mathfrak{g}$ and $Y \in \mathfrak{g}$ define $P^\lambda_Y : G \to \mathbb{R}$ as

$$P^\lambda_Y(g) := \lambda\left(\mathrm{Ad}_g(Y)\right), \qquad \forall g \in G. \tag{8.5.1}$$

A useful formula that these polynomials satisfy is the following:

$$XP^\lambda_Y = P^\lambda_{[X,Y]}, \qquad \forall X, Y \in \mathfrak{g}, \forall \lambda \in \mathfrak{g}^*. \tag{8.5.2}$$

Indeed, [see notes at page 6 in file attached_image.0377_001.pdf]...

From (8.5.2), it is easy to deduce that normal equations are parametrized with constant speed. Indeed, [see notes at page 5 in file attached_image.0377_001.pdf]...

8.5.2 First derivative of the extremal equations

Both in (5.3.6) and in (5.2.5), the function $t \mapsto \lambda\left(\mathrm{Ad}_{\gamma(t)}(e_i)\right)$ is considered. Let us differentiate such a function from $[0,1]$ into $\mathbb{R}$.

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}\left(\lambda\,\mathrm{Ad}_{\gamma(t)}(e_i)\right) &= \frac{\mathrm{d}}{\mathrm{d}s}\left(\lambda\,\mathrm{Ad}_{\gamma(t+s)}(e_i)\right)\Big|_{s=0}\\
&= \frac{\mathrm{d}}{\mathrm{d}s}\left(\lambda\,\mathrm{Ad}_{\gamma(t)}\,\mathrm{Ad}_{\gamma(t)^{-1}\gamma(t+s)}(e_i)\right)\Big|_{s=0}\\
&= \lambda\,\mathrm{Ad}_{\gamma(t)}\,\mathrm{ad}_{\left(\mathrm{d}L_{\gamma(t)}\right)^{-1}\gamma(t)^{-1}\dot\gamma_u(t)}(e_i)\\
&= \lambda\,\mathrm{Ad}_{\gamma(t)}[u(t), e_i],
\end{aligned} \tag{8.5.3}$$

where we used that $\mathrm{Ad}_{gh} = \mathrm{Ad}_g\,\mathrm{Ad}_h$, that λ and Ad_g are linear, and finally we used 5.2.1. From this last calculation we draw two conclusions: If γ_u is a normal curve with covector $\lambda \in \mathfrak{g}^*$, then

$$\dot u_i = \lambda\,\mathrm{Ad}_{\gamma(t)}[u(t), e_i], \qquad i = 1, \dots, r. \tag{8.5.4}$$

If γ_u is an abnormal normal curve with covector $\lambda \in \mathfrak{g}^*$, with $\lambda \neq 0$, then

$$0 = \lambda \operatorname{Ad}_{\gamma(t)}[u(t), e_i], \qquad i = 1, \dots, r. \tag{8.5.5}$$

8.5.3 Sard property for step-2 Carnot groups

[...]

8.5.4 Extremals in rank-2 Carnot groups

Consider a horizontal curve $\gamma : [0,1] \to G$, where G is a Carnot groups of rank-2. Say that the horizontal layer is spanned by e_1 and e_2. We use the notation $e_{12} = [e_1, e_2]$.

Then $u(t) = u_1(t)e_1 + u_2(t)e_2$ and we have

$$[u(t), e_1] = -u_2 e_{12} \qquad \text{and } [u(t), e_i] = u_1 e_{12}. \tag{8.5.6}$$

From (8.5.5), we have that if γ_u is an abnormal normal curve with covector $\lambda \in \mathfrak{g}^*$, with $\lambda \neq 0$, then

$$u_2 \lambda(\operatorname{Ad}_{\gamma(t)}(e_{12})) = u_1 \lambda(\operatorname{Ad}_{\gamma(t)}(e_{12})) = 0. \tag{8.5.7}$$

In addition, notice that we may assume that γ_u is parametrized by arc length, so u has constant nonzero norm, almost everywhere. In particular $(u_1, u_2) \neq (0,0)$ almost surely. Therefore we con conclude that for such an abnormal curve we have

$$\lambda(\operatorname{Ad}_{\gamma(t)}(e_{12})) = 0. \tag{8.5.8}$$

Viceversa, assume γ is a horizontal curve in a rank-2 Carnot group and that γ satisfies (8.5.8) for some $\lambda \in \mathfrak{g}^*$ with $\lambda \neq 0$. Then it clearly satisfies (8.5.7) and, since we have $r = 2$ and we have (8.5.6), we also have (8.5.5). Then look at each function $\lambda \operatorname{Ad}_{\gamma(t)}(e_i)$, for $i = 1$ and 2. On the one hand, because of (8.5.3) we have that its derivative is 0. On the other hand, if $\gamma(0) = 1_G$ and if λ satisfies (5.2.6), we have that the initial condition at time $t = 0$ for (5.2.5) is satisfied. Hence such a curve is abnormal. Hence, (8.5.8) is equivalent to the abnormal equations, in rank 2. We summarize this last proof in the following statement.

Proposition 8.5.9. *In every Carnot group G whose horizontal layer is spanned by e_1, e_2, a horizontal curve $\gamma : [0,1] \to G$ with $\gamma(0) = 1_G$ is abnormal if and only if for some $\lambda \in \mathfrak{g}^*$ with $\lambda \neq 0$ and $\lambda(e_1) = \lambda(e_2) = 0$ it satisfies*

$$\lambda(\operatorname{Ad}_{\gamma(t)}([e_1, e_2])) = 0. \tag{8.5.10}$$

Whereas, from (8.5.4) we have that if γ_u is a normal curve with covector $\lambda \in \mathfrak{g}^*$, then

$$\begin{cases} \dot u_1 = -u_2 \lambda(\mathrm{Ad}_{\gamma(t)}(e_{12})), \\ \dot u_2 = u_1 \lambda(\mathrm{Ad}_{\gamma(t)}(e_{12})). \end{cases} \tag{8.5.11}$$

We shall rephrase such condition in terms of a curvature. Let $\sigma : [0,1] \to \mathbb{R}^2$ the planar curve such that $\dot\sigma = u$. Then its oriented curvature, see [AT12, Equation (1.11)], is $\kappa(t) = \frac{1}{\|\sigma'(t)\|^3}\det(\sigma'(t), \sigma''(t))$. Hence, from (8.5.11), if γ_u is a normal curve with covector $\lambda \in \mathfrak{g}^*$, then its curvature satisfies

$$\kappa = \frac{\sigma_1'\sigma_2'' - \sigma_2'\sigma_1''}{\|\sigma'\|^3} = \frac{u_1\dot u_2 - u_2\dot u_1}{\|u\|^3} \overset{(8.5.11)}{=} \frac{u_1^2\lambda(\mathrm{Ad}_\gamma(e_{12})) + u_2^2\lambda(\mathrm{Ad}_\gamma(e_{12}))}{\|u\|^3} = \frac{1}{\|u\|}\lambda(\mathrm{Ad}_{\gamma(t)}(e_{12})).$$

We observe that the element $\mathrm{Ad}_{\gamma(t)}(e_{12})$ is in $[\mathfrak{g},\mathfrak{g}]$, hence in the last equation we lost the information of the value of λ on V_1. Still, normal curves need to satisfy (5.2.6). Viceversa, let's assume γ is a horizontal curve in a rank-2 Carnot group and that for some $\lambda \in [\mathfrak{g},\mathfrak{g}]^*$ we have that γ satisfies

$$\kappa = \frac{1}{\|u\|}\lambda(\mathrm{Ad}_{\gamma(t)}(e_{12})). \tag{8.5.12}$$

First we observe that by bootstrapping (8.5.12) we have that γ and its control u are smooth. Then we can extend λ as an element of $\mathfrak{g}^*$ so that we also have (5.2.6). Now that we have $\lambda \in \mathfrak{g}^*$ we consider the normal curve (which is unique) associated to λ with $\gamma(0) = 1_G$, which we denote by γ^λ. We shall show that $\gamma = \gamma^\lambda$. The reason is that both curves satisfy the ODE (8.5.12) with same initial data. [EXPLAIN MORE]

Regular abnormal extremals*

We fix a (rank-2) Carnot group G whose horizontal layer is spanned by e_1, e_2.

Definition *to be verified*: A horizontal curve $\gamma : [0,1] \to G$, parameterized by arc length and with $\gamma(0) = 1_G$, is called a regular abnormal extremal if for some $\lambda \in \mathfrak{g}^*$ with $\lambda \neq 0$, $\lambda(e_1) = \lambda(e_2) = 0$, and

$$\lambda|_{V_3} \neq 0 \tag{8.5.13}$$

it satisfies

$$\lambda(\mathrm{Ad}_{\gamma(t)}([e_1, e_2])) = 0. \tag{8.5.14}$$

Dubbio: è la stessa cosa se invece di (8.5.13) chiediamo che

$$\lambda(\mathrm{Ad}_{\gamma(t)} V_3) \neq \{0\}, \qquad \forall t \in [0,1]$$

Liu and Sussmann [LS95, Theorem5] showed that regular abnormal extremals are length minimizers, in rank-2 Carnot groups (in rank-2 subRiem manifolds?).

8.6 Pansu-Rademacher Theorem*

We would like to observe that the classical Rademacher Theorem states not only the existence almost everywhere of a tangent map (called the differential), but also its realizability as a linear map, in other word, as a group homomorphism that is compatible with the respective groups of dilations. Stated in this terms, the theorem holds for general equiregular sub-Finsler manifolds as well, see Section **??**. The aim of this section is to explain the content of such a differentiability result and to give a complete proof of it in the case of Carnot groups

8.6.1 Pansu's theorem

We shall prove Pansu's version of Rademacher Theorem.

Definition 8.6.1 (Pansu differentiability)**.** Let $\mathbb{G}_1$ and $\mathbb{G}_2$ be Carnot groups. We denote by δ_h the dilations of factor h in both of the groups. If $f : \mathbb{G}_1 \to \mathbb{G}_2$ is a map, then its *Pansu differential* at a point $x \in \mathbb{G}_1$ is the limit

$$Df_x := \lim_{h \to 0^+} \delta_{1/h} \circ L^{-1}_{f(x)} \circ f \circ L_x \circ \delta_h,$$

where the limit is with respect to the convergence on compact sets. Moreover, we say that f is *Pansu differentiable* if Df_x exists and is a homogeneous group homomorphism.

The value $Df_x(v)$ may be called *partial Pansu derivative* of f at x along v. Notice that Df_x is a map from $\mathbb{G}_1$ to $\mathbb{G}_2$, which may not be continuous, even if it exists. Notice that if $Df(x; v)$ exists, then $Df(x; \delta_\lambda v)$ exists for all $\lambda > 0$ and $Df(x; \delta_\lambda v) = \delta_\lambda Df(x; v)$.

Theorem 8.6.2 (Pansu's generalization of Rademacher Theorem)**.** *Let $f : \mathbb{G} \to \bar{\mathbb{G}}$ be a Lipschitz map between sub-Finsler Carnot groups. Then for almost every $x \in \mathbb{G}$ the map f is Pansu differentiable at x.*

Preliminaries to the proof of Pansu's theorem

In the proof of the above theorem, we will only take for granted few classical results to which we give hints to the proofs and references in the exercise section.

Theorem 8.6.3 (Rademacher Theorem in 1D)**.** *If $\gamma : [0, 1] \to \mathbb{R}^n$ is Lipschitz with respect to the Euclidean distance on $\mathbb{R}^n$, then the derivative $\dot{\gamma}(t)$ exists for almost every t and*

$$\gamma(t) = \gamma(0) + \int_0^t \dot{\gamma}(s)\, \mathrm{d}\, s, \qquad \textit{for all } t \in [0, 1].$$

Theorem 8.6.4 (Egorov Theorem for metric spaces, see Exercise 8.7.13)**.** *Let (X, μ) be a measure space with $\mu(X) < \infty$ and let Y be a separable metric space. Let $(f_t)_{t>0}$ be a family of measurable functions from X to Y depending on a real parameter $t \in (0, \infty)$. Suppose that $(f_t)_t$ converges almost everywhere to some f, as $t \to 0$. Then for every $\eta > 0$, there exists a measurable subset $K \subset X$ such that the $\mu(\Omega \setminus K) < \eta$ and $(f_t)_t$ converges to f uniformly on K.*

Theorem 8.6.5 (Consequence of Lebesgue Differentiation Theorem for doubling metric spaces, see Exercise 8.7.14)**.** *If (X, d, μ) is a doubling measure metric space and K is a measurable set in X then μ-almost every point of K has density 1.*

A proof of Pansu's theorem

As in Pansu's original proof, we first deal with the case of curves. We shall prove that every Lipschitz curve into a Carnot group is Pansu differentiable almost everywhere.

Proposition 8.6.6 (Case of curves)**.** *Let $\gamma : [0,1] \to \mathbb{G}$ be a Lipschitz curve. Then γ is Pansu differentiable almost everywhere and for almost every $x \in [0,1]$ we have that for all $v \in \mathbb{R}$*

$$D\gamma(x; v) := \lim_{t\to 0} \delta_{1/t} \left(\gamma(x)^{-1}\gamma(x + tv)\right) = \exp\left(v(L_{\gamma(x)})^* \dot\gamma(x)\right).$$

Here are few remarks before the proof. First we notice that the above curve γ is in particular Euclidean Lipschitz, so the tangent vector $\dot\gamma(x)$ exists for almost every x by Theorem 8.6.3. We also stress that Pansu differentiability for curves is stronger than Euclidean differentiability. Namely, if we consider the curve in a rank-r Carnot group in exponential coordinates $\gamma(t) = (\gamma_1(t), \dots, \gamma_n(t))$ and x is a point of Euclidean differentiability (we may assume $x = 0$ and $\gamma(x) = 0$), then $\dot\gamma(0) = \lim \gamma(t)/t = (\gamma_1(t)/t, \dots, \gamma_n(t)/t) \to (h_1, \dots, h_r, 0, \dots, 0)$. However, we have to consider

$$\delta_{1/t}\gamma(t) = (\gamma_1(t)/t, \dots, \gamma_n(t)/t^s)$$

and we need to prove that each coordinate $\gamma_j(t)$, with j greater than the rank, in fact vanishes not just faster than t but faster than t to the power of the degree of the coordinate.

Proof of Proposition 8.6.6. For simplicity, we take $v = 1$. We take $X_1, \dots, X_r$ a basis of the first layer of the stratification of $\mathrm{Lie}(G)$. Let $h_1, \dots, h_r \in L^\infty([0,1]; \mathbb{R})$ be such that

$$\dot\gamma(t) = \sum_{j=1}^{r} h_j(t) X_j(\gamma(t)), \qquad \text{for almost all } t \in [0,1]. \tag{8.6.7}$$

Since γ is L-Lipschitz, we may take $|h_j(t)| \le L$, for all t. Let $x \in [0,1]$ be both a point of Euclidean differentiability for γ and a Lebesgue point for all h_j, i.e.,

$$\frac{1}{|t-x|} \int_x^t |h_j(s) - h_j(t)| \, \mathrm{d}s \to 0, \qquad \text{as } t \to x.$$

Up to replacing γ with the curve $t \mapsto \gamma(x)^{-1}\gamma(t+x)$ we may assume that $x = 0$ and $\gamma(x) = 0$.

We identify the group $\mathbb{G}$ with its Lie algebra via the exponential map. Our aim is now to show that

$$\lim_{t\to 0} \delta_{1/t}\gamma(t) = \dot\gamma(0),$$

where the latter equals $\sum_{j=1}^r h_j(0)X_j(0)$ since 0 is a Lebesgue point for all h_j.

Set $\eta_t(s) := \delta_{1/t}\gamma(t\,s)$, so each $\eta_t : [0,1] \to \mathbb{G}$ is a curve starting at 0 that is L-Lipschitz:

$$d(\eta_t(s), \eta_t(s')) = d(\delta_{1/t}\gamma(t\,s), \delta_{1/t}\gamma(t\,s')) \le \frac{L}{t}|ts - ts'| = L|s-s'|.$$

Consequently, every sequence $(\eta_{t_k})_k$ has a uniformly converging subsequence. Moreover, we claim we have the equality

$$\dot\eta_t(s) = \sum_{j=1}^r h_j(ts) X_j(\eta_t(s)). \tag{8.6.8}$$

Indeed,

$$\dot\eta_t(s) = \frac{\mathrm{d}}{\mathrm{d}s}\delta_{1/t}\gamma(t\,s) = (\delta_{1/t})_*(t\dot\gamma(t\,s)) = \dot\gamma(t\,s),$$

which gives (8.6.8) from (8.6.7).

We claim that η_t uniformly converges to η_0, as $t \to 0$, where $\eta_0(t) := t\dot\gamma(0)$. This claim will complete the proof since in particular, $\eta_t(1) = \delta_{1/t}\gamma(t) \to \dot\gamma(0)$. For proving the claim we shall show that for all sequences $t_k \to 0$ there exists a subsequence t_{k_i} such that $\eta_{t_{k_i}} \to \dot\gamma(0)$. Indeed, by Ascoli-Arzela, there exists a subsequence t_{k_i} and there exists $\xi : [0,1] \to \mathbb{G}$ such that $\eta_{t_{k_i}} \to \xi$ uniformly. We want to show that

$$\dot\xi(s) = \sum_{j=1}^r h_j(0) X_j(\xi(s)), \qquad \text{for almost every } s \in [0,1].$$

Let σ be the curve such that $\sigma(0) = 0$ and $\dot\sigma(s) = \sum_{j=1}^r h_j(0)X_j(\xi(s))$. Let us integrate from 0 to an arbitrary $v \in (0,1)$:

$$\begin{aligned}
\sigma(v) - \eta_{t_{k_i}}(v) &= \int_0^v \sum_{j=1}^r h_j(0) X_j(\xi(s))\,\mathrm{d}s - \int_0^v \dot\eta_{t_{k_i}}(s)\,\mathrm{d}s \\
&= \int_0^v \sum_{j=1}^r h_j(0) X_j(\xi(s))\,\mathrm{d}s - \int_0^v \sum_{j=1}^r h_j(t_{k_i}s) X_j(\eta_{t_{k_i}}(s))\,\mathrm{d}s \\
&\leq \int_0^v \sum_{j=1}^r |h_j(0) - h_j(t_{k_i}s)| X_j(\xi(s))\,\mathrm{d}s + \\
&\qquad + \int_0^v \sum_{j=1}^r |h_j(t_{k_i}s)||X_j(\xi(s)) - X_j(\eta_{t_{k_i}}(s))\,\mathrm{d}s,
\end{aligned}$$

where we used (8.6.8). As $i \to \infty$, by continuity of X_i we have that the last summand goes to 0. Regarding the one before the last, we observe that

$$\begin{aligned}
\int_0^v \sum_{j=1}^r |h_j(0) - h_j(t_{k_i}s)|\,\mathrm{d}s &\leq \int_0^1 \sum_{j=1}^r |h_j(0) - h_j(t_{k_i}s)|\,\mathrm{d}s \\
&= \frac{1}{t}\int_0^t |h_j(0) - h_j(u)|\,\mathrm{d}u \to 0,
\end{aligned}$$

since 0 was a Lebesgue point. □

Proof of Theorem 8.6.2. Let $F : G \to H$ be a Lipschitz map. Define

$$F_{p,\epsilon}(x) := \delta_{1/\epsilon}(F(p)^{-1}F(p\delta_\epsilon x)), \qquad \text{for } p, x \in G \text{ and } \epsilon > 0.$$

Fix $X_1, \dots X_m$ a basis of V_1. For the entire proof, $j \in \{1, \dots, m\}$ and $R_j := \exp(\mathbb{R}X_j)$.

Let $\tilde F^j_{p,\epsilon}$ be the restriction $F_{p,\epsilon}|_{R_j} : R_j \to H$. By Proposition 8.6.6, for every $p \in G$ the maps $F \circ L_p|_{R_j}$ are almost everywhere differentable on R_j. By Fubini's theorem, there is a subset $E \subset G$ of full measure such that, for all $p \in E$, the limit $\tilde F^j_{p,0} = \lim_{\epsilon\to 0^+} \tilde F^j_{p,\epsilon}$ exists and is a Lipschitz group homomorphism $R_j \to H$. The limit is uniform on compact subsets of R_j.

Let L is a Lipschitz constant of F. We shall consider the space $\mathrm{Lip}^L(R_j; H)$ of L-Lipschitz functions from R_j to H, with a separable distances that metrizes the uniform convergence on compact sets, see Exercise 8.7.15 .

We have $\tilde F^j_{p,\epsilon} \in \mathrm{Lip}^L(R_j; H)$ for every $p \in G$ and $\epsilon \geq 0$. We can apply Egorov Theorem 8.6.4 to the functions $p \in G \mapsto \tilde F^j_{p,\epsilon} \in \mathrm{Lip}^L(R_j; H)$. Therefore, for every $\tau, r > 0$ there exists a set $E_{\tau,r} \subset E \cap B(1_G, r)$ such that $|B(1_G, r) \setminus E_{\tau,r}| < \tau$ and

$$\begin{array}{l} \{p_\epsilon\}_\epsilon \subset E_{\tau,r} \\ \lim_{\epsilon\to 0} p_\epsilon = p \in E_{\tau,r} \end{array} \quad \Rightarrow \quad \begin{array}{l} \tilde F^j_{p_\epsilon,\epsilon} \to \tilde F^j_{p,0} \\ \text{uniformly on compact sets of } R_j. \end{array} \tag{8.6.9}$$

Finally, let $E^\circ_{\tau,r} \subset E_{\tau,r}$ be the set of density points of $E_{\tau,r}$. Since we are in a doubling metric space, $E^\circ_{\tau,r}$ has full measure within $E_{\tau,r}$.

For the next few paragraphs we fix $p \in E^\circ_{\tau,r}$. We notice that for all $v \in G$, since p is a point of density of $E^\circ_{\tau,r}$ there exists $q_\epsilon \in E^\circ_{\tau,r}$ such that $\lim_{\epsilon\to 0} \delta_{1/\epsilon}(p^{-1}q_\epsilon) = v$.

Then define

$$\mathscr{D}_p := \left\{ v \in G : \forall q_\epsilon \in E^\circ_{\tau,r} \quad \begin{array}{l} \text{if } \delta_{1/\epsilon}(p^{-1}q_\epsilon) \overset{\epsilon\to 0}{\to} v \\ \text{then } \delta_{1/\epsilon}(F(p)^{-1}F(q_\epsilon)) \text{ converges} \end{array} \right\}.$$

Therefore, for all $v \in \mathscr{D}_p$ there exists an element in H, which we denote by $F_{p,0}(v)$, such that if $q_\epsilon \in E^\circ_{\tau,r}$ are such that $\lim_{\epsilon\to 0} \delta_{1/\epsilon}(p^{-1}q_\epsilon) = v$, then

$$F_{p,0}(v) := \lim_{\epsilon\to 0} \delta_{1/\epsilon}(F(p)^{-1}F(q_\epsilon).$$

Notice that if $v \in \mathscr{D}_p$, then for every sequence $\epsilon_m \searrow 0$ such that F_{p,ϵ_m} converges uniformly, as $m \to \infty$, we have

$$F_{p,0}(v) = \lim_{m\to\infty} F_{p,\epsilon_m}(v). \tag{8.6.10}$$

We claim that for all $v \in R_j$, and $p_\epsilon, q_\epsilon \in E^\circ_{\tau,r}$

$$\begin{array}{l} \lim_{\epsilon\to 0} \delta_{1/\epsilon}(p_\epsilon^{-1}q_\epsilon) = v \\ \lim_{\epsilon\to 0} p_\epsilon = p \end{array} \quad \Rightarrow \quad \lim_{\epsilon\to 0} \delta_{1/\epsilon}(F(p_\epsilon)^{-1}F(q_\epsilon)) = \tilde{F}^j_{p,0}(v). \tag{8.6.11}$$

Indeed, (8.6.11) is a consequence of $p_\epsilon \to p$ in $E_{\tau,r}$:

$$\begin{aligned} d(\delta_{1/\epsilon}(F(p_\epsilon)^{-1}F(q_\epsilon)), \tilde{F}^j_{p,0}(v)) &= d(F_{p_\epsilon,\epsilon}(\delta_{1/\epsilon}(p_\epsilon^{-1}q_\epsilon)), \tilde{F}^j_{p,0}(v)) \\ &\le d(F_{p_\epsilon,\epsilon}(\delta_{1/\epsilon}(p_\epsilon^{-1}q_\epsilon)), F_{p_\epsilon,\epsilon}(v)) + d(F_{p_\epsilon,\epsilon}(v), \tilde{F}^j_{p,0}(v)) \\ &\le Ld(\delta_{1/\epsilon}(p_\epsilon^{-1}q_\epsilon), v) + d(\tilde{F}^j_{p_\epsilon,\epsilon}(v), \tilde{F}^j_{p,0}(v)) \to 0, \end{aligned}$$

where at the end we used the first assumption of (8.6.11) and (8.6.9).

Our next claim about $\mathscr{D}_p$ is

$$g \in \mathscr{D}_p,\ v \in R_j \Rightarrow gv \in \mathscr{D}_p, \tag{8.6.12}$$

and in fact

$$F_{p,0}(gv) = F_{p,0}(g)\tilde{F}^j_{p,0}(v). \tag{8.6.13}$$

To show these last two claims, let $\{q_\epsilon\}_\epsilon \subset E^\circ_{\tau,r}$ be such that $\lim_{\epsilon\to 0} \delta_{1/\epsilon}(p^{-1}q_\epsilon) = gv$. Since $p \in E^\circ_{\tau,r}$ then there is $\{p_\epsilon\}_\epsilon \subset E^\circ_{\tau,r}$ such that $\lim_{\epsilon\to 0} \delta_{1/\epsilon}(p_\epsilon^{-1}q_\epsilon) = v$. So,

$$\begin{aligned} \lim_{\epsilon\to 0} \delta_{1/\epsilon}(F(p)^{-1}F(q_\epsilon)) &= \lim_{\epsilon\to 0} \delta_{1/\epsilon}(F(p)^{-1}F(p_\epsilon))\delta_{1/\epsilon}(F(p_\epsilon)^{-1}F(q_\epsilon)) \\ &\overset{(8.6.11)}{=} F_{p,0}(g)\tilde{F}^j_{p,0}(v). \end{aligned}$$

Next we observe the easy fact $1_G \in \mathscr{D}_p$, and therefore from (8.6.12) we infer

$$R_1, \ldots, R_m \subset \mathscr{D}_p. \tag{8.6.14}$$

From (8.6.14) and (8.6.12), together with the assumption that $R_1 \cup \ldots \cup R_m$ finitely generates G we get that $\mathscr{D}_p = G$. From (8.6.10) and (8.6.13), we conclude that every blowup of F at p, which exists by Ascoli-Arzelá, coincides with the map $F_{0,p} : G \to H$ and moreover it is a group homomorphism.

Since $\bigcup_{\tau,r>0} E^\circ_{\tau,r}$ has full measure in G, the map F is differentiable almost every-where on G. □

Original proof of Pansu's theorem

We mostly shall follow the original proof by Pansu together with some extra explanation from Monti's thesis. For the proof, we introduce the *difference quotients*:

$$R(x; v, t) := \bar{\delta}_{1/t}\left(f(x)^{-1} f(x\delta_t v)\right),$$

so that $Df(x; v) := \lim_{t\to 0} R(x; v, t)$.

We start with a preliminary result. It states that if almost everywhere we have partial derivatives in two directions, then we also have it at the product of the directions.

Proposition 8.6.15. *Let $f : \mathbb{G} \to \bar{\mathbb{G}}$ be a Lipschitz map between sub-Finsler Carnot groups. If $Df(x; v)$ and $Df(x; w)$ exists for almost every $x \in \mathbb{G}$, then $Df(x; vw)$ exists for almost every $x \in \mathbb{G}$ and $Df(x; vw) = Df(x; v)Df(x; w)$.*

Proof of Proposition 8.6.15. Let $\Omega \subset \mathbb{G}$ open with finite measure. Let $\eta > 0$. By Egorov's theorem for metric spaces (see Theorem 8.6.4) there exists a measurable subset $K \subset \Omega$ such that the measure of $\Omega \setminus K$ is less than η and $R(x; w, t) \to Df(x; w)$, as $t \to 0$, uniformly on K. Moreover, since the measure is regular, we may assume that K is compact.

We claim that to conclude the proof it is enough to show

$$R(x\delta_t v; w, t) \to Df(x; w), \qquad \text{for almost every } x \in K. \tag{8.6.16}$$

Indeed, in this case, for $x \in K$, we have

$$\begin{aligned} R(x; vw, t) &= \bar{\delta}_{1/t}\left(f(x)^{-1} f(x\delta_t(vw))\right) \\ &= \bar{\delta}_{1/t}\left(f(x)^{-1} f(x\delta_t v)\right) \bar{\delta}_{1/t}\left(f(x\delta_t v)^{-1} f(x\delta_t v\delta_t w)\right) \\ &= R(x; v, t)R(x\delta_t v; w, t) \to Df(x; v)Df(x; w). \end{aligned}$$

Then on concludes taking the union of the sets $K = K(\eta)$ when η varies in $\mathbb{N}$, which form a full measure set.

For showing (8.6.16) take as x a point of density for K, recall that from Theorem 8.6.5 these points are of full measure in K. For $t > 0$, let $x_t \in K$ be one projection of $x\delta_t v$ on K, i.e., such that $d(x\delta_t v, x_t) = d(x\delta_t v, K) =: r_t$. Then $r_t \leq d(x\delta_t v, x) = td(v, 0)$. We claim that $r_t/t \to 0$. Indeed,

$$\frac{r_t^Q}{(2td(v,0))^Q} = \frac{|B_d(x\delta_t v, r_t)|}{|B_d(x, 2d(x, x\delta_t v))|} \leq \frac{|B_d(x, 2d(x, x\delta_t v)) \setminus K|}{|B_d(x, 2d(x, x\delta_t v))|} \to 0.$$

We now calculate

$$\begin{aligned} R(x\delta_t v, w, t) &= \bar{\delta}_{1/t}\left(f(x\delta_t v)^{-1} f(x\delta_t v\delta_t w)\right) \\ &= \underbrace{\bar{\delta}_{1/t}\left(f(x\delta_t(v))^{-1} f(x_t)\right)}_{A_t} \underbrace{\bar{\delta}_{1/t}\left(f(x_t)^{-1} f(x_t\delta_t w)\right)}_{B_t} \underbrace{\bar{\delta}_{1/t}\left(f(x_t\delta_t w)^{-1} f(x\delta_t v\delta_t w)\right)}_{C_t}. \end{aligned}$$

We claim that $A_t \to 0$ as $t \to 0$. Indeed,

$$\bar{d}(0, A_t) = \frac{1}{t}\bar{d}(f(x_t), f(x\delta_t v)) \leq \frac{L}{t} d(x_t, x\delta_t v) = Lr_t/t \to 0.$$

We then notice that, since $x_t \in K$, $x_t \to x$, and on K the convergence is uniform, we have that $B_t = R(x_t; w, t) \to Df(x; w)$ as $t \to 0$. We then claim that $C_t \to 0$ as $t \to 0$. Indeed,

$$\begin{aligned} \bar{d}(0, C_t) &= \frac{1}{t}\bar{d}(f(x_t\delta_t w), f(x\delta_t v\delta_t w)) \\ &\leq \frac{L}{t} d(x_t\delta_t w, x\delta_t v\delta_t w) \\ &= Ld(\delta_{1/t}(x_t)w, \delta_{1/t}(x\delta_t v)w) \to 0, \end{aligned}$$

where we used that that $d(\delta_{1/t}(x_t), \delta_{1/t}(x\delta_t v)) = \dfrac{d(x_t, x\delta_t v)}{t} \to 0.$ □

Another Proof of Theorem 8.6.2. Let $X_1, \ldots, X_r$ be a basis of the first layer of the stratification of $\mathrm{Lie}(G)$.

We first claim that the set $E := \{p \in \mathbb{G} : Df(p; \exp(X_i))$ and $Df(p; \exp(-X_i))$ exists for all $i\}$ has full measure. Indeed, complete to a basis $X_1, \ldots, X_n$ of $\mathrm{Lie}(G)$. For $j \in \{1, \ldots, r\}$, define $\phi_j : \mathbb{R}^n \to \mathbb{G}$ as $\phi_j(x_1, \ldots, x_n) = \exp(\sum_{i \neq j} x_i X_i)\exp(x_j X_j)$. Then ϕ_j is a diffeo and for all $x \in \mathbb{R}$ the curve $t \mapsto \phi_j(x + te_j)$ is the flowline of X_j starting at $\phi_j(x)$. Set

$$\tilde{E}_j := \{x \in \mathbb{R}^n : t \mapsto f(\phi_j(x_1, \ldots, x_{j-1}, t, x_{j+1}, \ldots, x_n)) \text{ is P-diff in } t = x_j\}.$$

By Fubini's theorem for Lebesgue measure and by Proposition 8.6.6, $\tilde{E}_j$ has full measure. Then $E = \cap_{j=1}^r \phi_j(\tilde{E}_j)$ has full measure.

Then let $S = \{v \in \mathbb{G} : d(0,v) = 1\}$ be the unit sphere in $\mathbb{G}$. For all $m \in \mathbb{N}$ there exists $v_1^m, \ldots, v_{j_m}^m$ such that $S \subseteq \cup_{i=1}^{j_m} B_d(v_i^m, 1/m)$. We then claim that each set

$$E_m := \{p \in E : Df(p; v_i^m) \text{ exists for all } i = 1, \ldots, j_m\}$$

has full measure. Indeed, since $\mathcal{G} := \{\exp(\lambda X_i) : \lambda \in \mathbb{R}, i = 1, \ldots, r\}$ generates $\mathbb{G}$, then for all i and all m there exists $w_1, \ldots, w_k \in \mathcal{G}$ such that $v_i^m = w_1 \ldots w_k$. Hence, from Proposition 8.6.15 for almost every $p \in \mathbb{G}$ we have that $Df(p; v_i^m)$ exists. Thus E_m has full measure.

We finally claim that if $p \in \cap_{m \in \mathbb{N}} E_m$, then $R(p; v, t)$ converges uniformly in $v \in S$, as $t \to 0$. Indeed, we want to show that for all $m \in \mathbb{N}$ there exists $\delta > 0$ such that for all $s, t \in (0, \delta)$ and all $v \in S$

$$\bar{d}(R(p; v, t), R(p; v, s)) \leq \frac{1 + 2L}{m}.$$

Let $m \in \mathbb{N}$. Then there exists $\delta > 0$ such that for all $i \in \{1, \ldots, i_m\}$ and all $s, t \in (0, \delta)$

$$\bar{d}(R(p; v_i^m, t), R(p; v_i^m, s)) \leq \frac{1}{m}.$$

Let $v \in S$. Then there exists i such that $d(v, v_i^m) \leq \dfrac{1}{m}$. Then for all $s, t \in (0, \delta)$

$$\begin{aligned}
\bar{d}(R(p; v, t), R(p; v, s)) &\leq \bar{d}(R(p; v, s), R(p; v_i^m, s),) + d(R(p; v_i^m, s), R(p; v_i^m, t)) + \bar{d}(R(p; v_i^m, t), R(p; v, t)) \\
&\leq \frac{1}{s} \bar{d}(f(p\delta_s v_i^m), f(p\delta_s v)) + \frac{1}{m} + \frac{1}{t} \bar{d}(f(p\delta_t v_i^m), f(p\delta_t v)) \\
&\leq \frac{L}{s} s d(v_i^m, v) + \frac{1}{m} + \frac{L}{t} t d(v_i^m, v) \leq \frac{L + 1 + L}{m}.
\end{aligned}$$

□

8.6.2 Applications to non-embeddability

It was observed by Semmes, [Sem96, Theorem 7.1], that Pansu's differentiation Theorem 9.4.9 implies that a Lipschitz embedding of the Heisenberg group with its CC distance into an Euclidean space, cannot be bi-Lipschitz.

Theorem 8.6.17. *There is no bi-Lipschitz embedding from an open set in a Heisenberg group to an Euclidean space $\mathbb{R}^n$.*

Proof. Suppose that such an embedding f exists. The Pansu Rademacher Theorem 9.4.9 would imply that there exists at least one point at which f is differentiable and whose tangent map is a group homomorphism. The blowing-up procedure used to define the tangent map scales in the natural way, i.e., if f is L-bi-Lipschitz, then each rescaled f_λ is L-bi-Lipschitz and so the tangent map is bilipschitz too. In particular, the tangent map is injective. We now get a contradiction, because we considered a tangent map which is a group homomorphism between tangents spaces which are the 3-dimensional Heisenberg group and the Abelian $\mathbb{R}^n$. However, every homomorphism from the Heisenberg group into $\mathbb{R}^n$ must have a kernel which is at least 1-dimensional (all commutators in the Heisenberg group must be mapped to 0 by the homomorphism) and hence cannot be injective. □

Corollary 8.6.18. *Let M_1 an M_2 be sub-Riemannian manifolds with tangents the Carnot groups $\mathbb{G}_1$, respectively $\mathbb{G}_2$. If no subgroup of $\mathbb{G}_2$ is isomorphic to $\mathbb{G}_1$ then there is no bi-Lipschitz embedding of M_1 in M_2.*

Corollary 8.6.19. *The Heisenberg group, or every other non-commutative Carnot group, is purely unrectifiable.*

A consequence of the proof of Theorem 8.6.17 is that each Lipschitz map from the Heisenberg group to an Euclidean space has to compress points in the direction of the center of the group.

Proposition 8.6.20 (Center collapse)**.** *If $U \subset H$ is an open subset, and $f : U \to \mathbb{R}^n$ is a Lipschitz map, then for almost every point $x \in H$, the map collapses in the direction of the center of H, i.e.,*

$$\lim_{g \to e} \frac{\|f(xg) - f(x)\|}{d(xg, x)} = 0\,, \qquad g \in \mathrm{Center}(H)\,. \tag{8.6.21}$$

This last theorem has been generalized by J. Cheeger and B. Kleiner to maps with values in the Banach space L^1. Such a result gave a proof of the following theorem which has been conjectured by J. Lee and A. Naor.

Theorem 8.6.22 (Lee-Naor-Cheeger-Kleiner)**.** *The Heisenberg group equipped with its CC metric does not admit a bi-Lipschitz embedding into L^1.*

This conjecture arose from the work of J. Lee and A. Naor, in which it is shown that the nonexistence of such an embedding provides a natural counter-example to the Goemans-Linial conjecture of theoretical computer science; S. Khot and N. Vishnoi gave a first such counterexample. Very roughly, the point is that in some instances, questions in algorithm design, such as the sparsest cut problem, could be solved if it were possible to embed a certain class of finite metric spaces (those with metrics of negative type) into ℓ^1 with universally bounded bi-Lipschitz distortion, i.e., distortion independent of the particular metric and the cardinality.

8.7 Exercises

Exercise 8.7.1. Show that if G is a Carnot group and Δ is the left-invariant distribution with $\Delta_{1_G} = V_1$, then $(\Delta^{[j]})_{1_G} = V_1 \oplus \cdots \oplus V_j$. ■

Exercise 8.7.2. Show that if G is a Carnot group and Δ is the left-invariant distribution with $\Delta_{1_G} = V_1$, then the three definitions (3.3.12), (3.3.13), (3.3.14), and (8.1.6) of Q coincide. ■

Exercise 8.7.3. Use the BCH formula to show (8.1.13). ■

Exercise 8.7.4. Use the definitions to prove (8.1.16). ■

Exercise 8.7.5. Show that in every Carnot groups there is a (strong) Malcev basis. ■

Exercise 8.7.6. Prove that, if M is a Riemannian manifold, then the Carnot group structure that every T_pM inherits is Abelian. ■

Exercise 8.7.7. Prove that, if M is a contact 3-manifold, then for every $p \in M$ the Carnot algebra structure that T_pM inherits is the Heisenberg algebra structure. ■

Exercise 8.7.8. Prove that, if G is a Carnot group, then for every $p \in G$ the Carnot algebra structure that T_pG inherits is the Lie algebra Lie(G) itself. ■

Exercise 8.7.9. Give an example of Lie group G with a left-invariant bracket-generating distribution such that Carnot group structure that $T_{1_G}G$ inherits is NOT isomorphic to the Lie algebra Lie(G). ■

Exercise 8.7.10. Let V and $W \subset \mathfrak{g}$ be two sub-vector spaces with $X_1, \ldots, X_l$ and $Y_1, \ldots, Y_m$ basis of V and W respectively. Then show that the vectors $[X_i, Y_j]$, for $i = 1, \ldots, l, j = 1, \ldots, m$ span $[V, W]$, thus one can extract a basis among such brackets. ■

Exercise 8.7.11. Let $\mathfrak{g} = V_1 \oplus \cdots \oplus V_s$ be a stratification of a Lie algebra. Assume that $X_{m_j+1}, \ldots, X_{m_j}$ is a basis of V_j, then show that the order-reversed basis $X_n, \ldots, X_1$ is a (strong) Malcev basis. ■

Exercise 8.7.12. Considering the horizontal path constructed in the proof of Property 3 in Proposition 8.2.4, give a lower bound on $d_{CC}(1_{\mathbb{G}}, E(\mathbf{t}))$. ■

Exercise 8.7.13. Fill in the details in the following argument to prove Theorem 8.6.4. Without loss of generality we may assume that $f_t \to f$ everywhere on X. For $k \in \mathbb{N}$ and $t \in (0, \infty)$, let

$$E_t(k) := \cup_{s \in (t,\infty)} \{x : |f_s(x) - f(x)| > k^{-1}\}.$$

Then, for fixed k, $E_t(k)$ decreases as t decreases, and $\cap_{t\in(0,\infty)} E_t(k) = 0$, so since $\mu(X) < \infty$ we conclude that $\mu(E_t(k)) \to 0$ as $t \to 0$. Given $\eta > 0$ and $k \in \mathbb{N}$, choose t_k so large that $\mu(E_{t_k}(k)) < \eta 2^{-k}$ and let $E = \cap_{k\in\mathbb{N}} E_{t_k}(k)$. Then $\mu(E) < \eta$, and we have $|f_t(x) - f(x)| < k^{-1}$ for $t \in (0, t_k)$ and $x \notin E$. Thus $(f_t)_t$ converges to f uniformly on $X \setminus E$. ■

Exercise 8.7.14 (Lebesgue Differentiation Theorem for doubling metric spaces)**.** If (X, d, μ) is a doubling measure metric space and $f \in L^1(X, \mu)$, then for μ-almost every $x \in X$ we have

$$\frac{1}{\mu(B(x,r))} \int_{B(x,r)} |f(y) - f(x)| \,\mathrm{d}\,\mu(y) \to 0, \qquad \text{as } r \to 0.$$

In particular, if $K \subseteq X$ is measurable, then μ-almost every point of K has density 1. ■

Exercise 8.7.15 (The space of L-Lipschitz functions)**.** Let G and H be Carnot groups. Let $\mathtt{Lip}^L(G; H)$ be the set of Lipschitz functions $G \to H$ of Lipschitz constant at most L. Consider the function

$$d_L(f, g) := \sup \left\{ \frac{d_H(f(x), g(x))}{n^2} : \ n \in \mathbb{N},\ x \in B(1_G, n) \right\}.$$

Show that

(i). The function d_L is a distance function on $\mathtt{Lip}^L(G; H)$.

(ii). Convergence with respect to d_L is equivalent to uniform convergence on compact sets.

(iii). The space $(\mathtt{Lip}^L(G; H), d_L)$ is separable.

Solution. (i). The axioms to check that d_L is a distance function are easy to verify.

(ii). Let $\{f_k\}_k \subset \mathtt{Lip}^L(G; H)$ and $f \in \mathtt{Lip}^L(G; H)$.

Suppose that $\lim_{k\to\infty} d_L(f_k, f) = 0$. If $E \subset G$ is compact, then there is $N \in \mathbb{N}$ such that $E \subset B(1_G, N)$. Since

$$\sup\{d_H(f_k(x), f(x)) : x \in B(1_G, N)\} \le N^2 d_L(f_k, f) \to 0,$$

then $f_k \to f$ uniformly on E. Since E is an arbitrary compact set, $f_k \to f$ uniformly on compact sets.

Suppose now that $f_k \to f$ uniformly on compact sets and let $\epsilon > 0$. Since $\{1_G\}$ is compact, there is $C > 0$ such that $d_H(f_k(1_G), f(1_G)) \le C$ for all $k \in \mathbb{N}$. Notice that, for all $n \in \mathbb{N}_{\ge 1}$ and $x \in B(1_G, n)$, we have

$$\begin{aligned} \frac{d_H(f_k(x), f(x))}{n^2} &\le \frac{d_H(f_k(x), f_k(1_G)) + d_H(f_k(1_G), f(1_G)) + d_H(f(1_G), f(x))}{n^2} \\ &\le \frac{2L}{n} + \frac{C}{n^2}. \end{aligned}$$

Therefore, there is $N \in \mathbb{N}$ such that $\frac{d_H(f_k(x),f(x))}{n^2} < \epsilon$ for all $n \geq N$ and $x \in B(1_G, n)$. Let $K \in \mathbb{N}$ be such that

$$\sup\{d_H(f_k(x), f(x)) : x \in B(1_G, N)\} \leq \epsilon$$

for all $k > K$. Then, for $k > K$, we have $d_L(f_k, f) \leq \epsilon$. We conclude that $\lim_{k\to\infty} d_L(f_k, f) = 0$.

(ii). The topology of uniform convergence on compact sets is equivalent to the compact-open topology. Moreover, by Ascoli-Arzelà, for every $n \in \mathbb{N}$ the set

$$\mathscr{K}(n) := \{f \in \mathtt{Lip}^L(G; H) : f(1_G) \in \bar{B}(1_H, n)\}$$

is compact, hence separable. Since $\mathtt{Lip}^L(G; H) = \bigcup_{n\in\mathbb{N}} \mathscr{K}(n)$ is a countable union of separable sets, then it is also separable. □

■

Chapter 9

Limits of Riemannian and subRiemannian manifolds*

9.1 Limits of metric spaces*

Sub-Riemannian Carnot groups emerge as limit metric spaces, both as distinguished asymptotic spaces and as tangent spaces. In most cases, after some change of coordinates, the study can be reduced to distances that uniformly converge on compact sets. However, it can also be valuable to regard such convergence as a specific instance of Gromov-Hausdorff convergence.

9.1.1 A topology on the space of metric spaces

Let X and Y be metric spaces, $L > 1$, and $C > 0$. A map $\phi : X \to Y$ is an (L, C)-*quasi-isometric embedding* if

$$\frac{1}{L}d(x,x') - C \leq d(\phi(x), \phi(x')) \leq Ld(x,x') + C, \qquad \forall x, x' \in X.$$

If $A, B \subset Y$ are subsets of a metric space Y and $\epsilon > 0$, we say that A is an ϵ-*net* for B if

$$B \subset \mathrm{Nbhd}_\epsilon^Y(A) := \{y \in Y : d(x, A) < \epsilon\}.$$

Definition 9.1.1 (Hausdorff approximating sequence)**.** Let $(X_j, x_j), (Y_j, y_j)$ be two sequences of pointed metric spaces. A sequence of maps $\phi_j : (X_j, x_j) \to (Y_j, y_j)$ is said to be *Hausdorff approximating* if for all $R > 0$ and all $\delta > 0$ there exists ϵ_j such that

1. $\epsilon_j \to 0$ as $j \to \infty$;
2. $\phi_j|_{B(x_i,R)}$ is a $(1, \epsilon_j)$-quasi isometric embedding;
3. $\phi_j(B(x_j, R))$ is an ϵ_j-net for $B(y_j, R - \delta)$.

Definition 9.1.2. We say that a sequence of pointed metric spaces (X_j, x_j) *converges* to a pointed metric space (Y, y) if there exists an Hausdorff approximating sequence $\phi_j : (X_j, x_j) \to (Y, y)$.

This notion of convergence was introduced by M. Gromov and it is also called *Gromov-Hausdorff convergence.*

Proposition 9.1.3. *Let d_j be a sequence of distances on a set X that converge to a distance d_∞ uniformly on bounded sets with respect to d_∞. Let $x_0 \in X$. If*

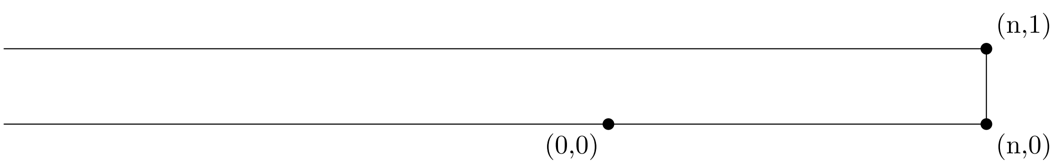
$$\operatorname{diam}_{d_\infty}\Big(\bigcup_{j\in\mathbb{N}} B_{d_j}(x_0, R)\Big) < \infty, \quad \forall R > 0, \tag{9.1.4}$$

then $\mathrm{id} : (X, d_j, x_0) \to (X, d_\infty, x_0)$ *is a Hausdorff approximating sequence and* (X, d_∞, x_0) *is the limit of* (X, d_j, x_0).

Proof. Exercise. □

Example 9.1.5. The following example shows that condition (9.1.4) is necessary in the last proposition. For $n \in \mathbb{N}$ define $\gamma_n : \mathbb{R} \to \mathbb{R}^2$ by

$$\gamma_n(t) := \begin{cases} (t, 0) & t \le n \\ (n, t - n) & n \le t \le n + 1 \\ (n - (t - n - 1), 1) & n + 1 \le t \end{cases}$$



These mappings induce metrics d_n on $\mathbb{R}$ by

$$d_n(x, y) := |\gamma_n(x) - \gamma_n(y)| \quad \forall x, y \in \mathbb{R}.$$

Here $d_n(x, y)$ converge to $d_\infty(x, y) := |x - y|$, for $x, y \in \mathbb{R}$. The convergence is uniform on compact sets, but not in the Gromov-Hausdorff sense.

9.1.2 Asymptotic cones and tangent spaces

If $X = (X, d)$ is a metric space and $\lambda > 0$, we set $\lambda X := (X, \lambda d)$.

Definition 9.1.6. Let X, Y be metric spaces, $x \in X$ and $y \in Y$. We say that (Y, y) is the *asymptotic cone* of X if for each infinitesimal sequence $\lambda_j \to 0$ we have $(\lambda_j X, x) \to (Y, y)$, as $j \to \infty$.

we say that (Y, y) is the *tangent space* of X at x if for each diverging sequence $\lambda_j \to \infty$, $(\lambda_j X, x) \to (Y, y)$, as $j \to \infty$.

Remark 9.1.7. The notion of asymptotic cone is independent from x.

Remark 9.1.8. In general, asymptotic cones and tangent spaces may not exists.

Remark 9.1.9. Within the space of boundedly compact metric spaces, limits are unique up to isometries.

The following two theorems serve as the central focus of this chapter.

Theorem 9.1.10. *Let G be a nilpotent Lie group equipped with a left-invariant subFinsler distance. Then the asymptotic cone of G exists and is a Carnot group.*

Theorem 9.1.11. *Let G be a subFinsler Lie group, or, more generally, an equiregular subFinsler manifold and $p \in G$. Then the tangent space of G at p exists and is a Carnot group.*

9.2 Limits of Carnot-Carathéodory distances*

When taking limits of sub-Riemannian structures, it is important to note that the rank of the distribution may change. This can be observed in the example of the Riemannian Heisenberg group, whose distribution has rank three, while its asymptotic cone, the sub-Riemannian Heisenberg group, has a distribution of rank two. To study limits of CC spaces effectively, it is advantageous to adopt the perspective of bundle structures. In Section 9.5 we will consider parameter family of those structures that we explored in Section 3.1.4.

However, we shall initially focus of subRiemannian structures (on Lie groups) that have the advantance that the Carnot-Carathéodory geometry is given by orthonormal frames. Clearly, at the limits the frames might degenerate and not be linearly independent anymore.

9.2.1 Dilations of CC structures

We begin by checking how one need to change a CC structure in order to multiply its distance by a factor. In fact, if (M, d) is a CC space, then $(M, \lambda d)$ is a CC space.

[...]

We start with a simple observation that shows to the inexpert reader how a subRiemannian Carnot group can appear as limit of Riemannian metrics on the same Lie group.

Lemma 9.2.1. *Let G be a stratified group, with $X_1, \ldots, X_n$ be a basis adapted to the stratification. Consider the Riemannian metric d_1 for which $X_1, \ldots, X_n$ are orthonormal. Then for all $\lambda > 0$ the metric space $(G, \lambda d_1)$ is isometric to (G, d_λ) via the map δ_λ, where d_λ is the Riemannian distance for which*

$$X_1, \ldots, \frac{\lambda^{deg(X_j)}}{\lambda} X_j, \ldots, \lambda^{s-1} X_n$$

are orthonormal.

Proof. The distance λd_1 associated to the Riemannian metric g_λ that makes $\frac{1}{\lambda} X_1, \ldots, \frac{1}{\lambda} X_n$ orthonormal. The map $\delta_\lambda : (G, \lambda d_1) \to (G, d_\lambda)$ is a Riemannian isometry since it sends the orthonormal vector $\frac{1}{\lambda} X_j$ to the orthonormal vectors $(\delta_\lambda)_* (\frac{1}{\lambda} X_j) = \frac{1}{\lambda} \lambda^{deg(X_j)} X_j$. □

[...]

9.2.2 Privileged coordinates

[...]

definition[Privileged coordinates] [from Jean pages 20-22-23]

Proposition 9.2.2. *In privileged coordinates, let $X_1, \ldots, X_m$ be an orthonormal frame for a sub-Riemannian manifold (M, d) and for all $\epsilon > 0$ set $X_j^{(\epsilon)} := \epsilon (\delta_\epsilon)^* (X_j)$, for $j = 1, \ldots, m$. Then the limit $\hat{X}_j := \lim_{\epsilon \to 0} X_j^{(\epsilon)}$ exists and $\hat{X}_1, \ldots, \hat{X}_m$ are bracket generating. Moreover, the map δ_ϵ gives an isometry between $(M, \epsilon d)$ and (M, d_ϵ) where d_ϵ has $X_1^{(\epsilon)}, \ldots, X_m^{(\epsilon)}$ as orthonormal frame.*

9.3 SubRiemannian Carnot group as Riemannian limits*

9.3.1 Limits of Riemannian manifolds

SubRiemannian manifolds appear as limiting objects of Riemannian manifolds. The following is an example of what one can prove with the techniques from this chapter, see Theorem 9.3.7.

Proposition 9.3.1. *Let M be a manifold, $\Delta \subset TM$ a bracket-generating subbundle. Let $(g_n)_{n \in \mathbb{N}}$ be a sequence of Riemannian metrics on M. Assume that the orthogonal to Δ is the same for each g_n, that*

$$g_n|_\Delta = g_1|_\Delta, \quad \forall n \in \mathbb{N},$$

and for all $X \notin \Delta$

$$g_n(X, X) \to +\infty, \quad as \quad n \to \infty.$$

Then for all $p, q \in M$

$$\lim_{n\to\infty} d_{g_n}(p,q) = d_{CC}(p,q)$$

where d_{CC} is the subRiemannian distance associated to Δ and $g_1|_\Delta$.

9.3.2 Preparatory example: The Riemannian Heisenberg group

Theorem 9.3.2. *Let X, Y, Z be a basis of the Lie algebra of the Heisenberg group G with only relation $[X,Y] = Z$. For all $n \in \mathbb{N}$, let d_n be the Riemannian distance for which $X, Y, \frac{1}{n}Z$ are orthonormal. Let d_{CC} be the subRiemannian distance for which X, Y are orthonormal.*

Then for all $R > 0$ there is a sequence $\epsilon_n \to 0$ as $n \to 0$ such that for all $p, q \in B_{CC}(1_G, R)$,

$$d_n(p,q) \le d_{CC}(p,q) \le d_n(p,q) + \epsilon_n.$$

In other words, $d_n \to d_\infty$ and the limit is uniform on compact sets.

Hence if d is the Riemannian distance for which X, Y, Z are orthonormal, then $(G, \frac{1}{n}d)$, which is isometric to (G, d_n) converge to (G, d_{CC}). In other words, the asymptotic cone of the Riemannian Heisenberg group is the subRiemannian Heisenberg group.

Exercise 9.3.3. Show that $(G, \frac{1}{n}d)$ and (G, d_n) are isometric. ■

Proof of Theorem 9.3.2. The fact that $d_n \le d_{CC}$ is clear, since every horizontal curve for d_{CC} has exactly the length with respect to d_n.

For the other inequality, take $p, q \in B_{CC}(1_G, R)$. Let $\gamma_n : [0,1] \to G$ be a curve from p to q that minimizes the length with respect to d_n. Decompose $\dot\gamma$ as

$$\dot\gamma(t) = a_1(t)X + a_2(t)Y + a_3(t)Z$$

with $a_3(t)$ not necessarily 0. Let $\sigma : [0,1] \to G$ be the curve such that $\sigma(0) = p$ and $\dot\sigma(t) = a_1(t)X + a_2(t)Y$. Let $\bar q := \sigma(1)$. Let $\eta : [0,1] \to G$ be the curve such that $\eta(0) = \bar q$ and $\dot\eta(t) = a_3(t)Z$.

We claim that

$$\eta(t) = (L_{\bar q} \circ L^{-1}_{\sigma(t)})(\gamma(t)), \qquad \forall t \in [0,1]. \tag{9.3.4}$$

Since

$$(L_{\bar q} \circ L^{-1}_{\sigma(0)})(\gamma(0)) = L_{\bar q} \circ L_p^{-1}(p) = \bar q = \eta(0),$$

it is enough to show that

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(L_{\bar q} \circ L^{-1}_{\sigma(t)} \circ \gamma(t))\right) = \dot\eta(t).$$

For doing this, lets consider exponential coordinate so

$$\dot{\gamma} = a_1 X + a_2 Y + a_3 Z = \Big(a_1, a_2, a_3 - \frac{\gamma_2}{2} a_1 + \frac{\gamma_1}{2} a_2 \Big)$$

and

$$\dot{\sigma} = \Big(a_1, a_2, -\frac{\sigma_2}{2} a_1 + \frac{\sigma_1}{2} a_2 \Big).$$

Thus $\gamma_1 = \sigma_1 = p_1 + \int_0^t a_1$ and $\gamma_2 = \sigma_2 = p_2 + \int_0^t a_2$.

$$\begin{aligned} \sigma(t)^{-1}\gamma(t) &= (\gamma_1 - \sigma_1, \gamma_2 - \sigma_2, \gamma_3 - \sigma_3 - \frac{1}{2}(\sigma_1\gamma_2 - \sigma_2\gamma_1) = \\ &= (0, 0, \gamma_3 - \sigma_3) \end{aligned}$$

Thus

$$\frac{\mathrm{d}}{\mathrm{d}t}\sigma(t)^{-1}\gamma(t) = (0, 0, \dot{\gamma}_3 - \dot{\sigma}_3) = a_3 Z.$$

The claim (9.3.4) is proved and, in particular, we have that

$$\eta(1) = \bar{q}\bar{q}^{-1} q = q.$$

We need to bound the length $L_{d_1}(\eta)$. Since X, Y, Z are orthogonal and $\|\frac{1}{n} Z\|_n = 1$, we have

$$\int_0^1 n \cdot |a_3| = \int_0^1 \|a_3 Z\|_n \leq \int_0^1 \|a_1 X + a_2 Y + a_3 Z\|_n = L_{d_n}(\gamma) = d_n(p,q) \leq d_{CC}(p,q) \leq 2R$$

Then

$$L_{d_1}(\eta) = \int_0^1 \|a_3 Z\|_1 = \int_0^1 |a_3| \leq \frac{2R}{n}$$

Thus, as $n \to \infty$, $d_1(\bar{q}, q)$ goes to 0 uniformly on $p, q \in B_{CC}(1_G, R)$. In fact, using the ball-box theorem,

$$d_{CC}(\bar{q}, q) \quad \leq \quad K d_1(\bar{q}, q)^{1/2} \quad \leq \quad (L_{d_1}(\eta))^{1/2} \quad \leq \quad K \left(\frac{2R}{n}\right)^{1/2} \quad = \quad O(\frac{1}{\sqrt{n}}).$$

Since $d_{CC}(p, \bar{q}) \leq L_{CC}(\sigma) \leq L_{d_n}(\gamma) = d_n(p,q)$, we conclude that

$$d_{CC}(p,q) \leq d_{CC}(p,\bar{q}) + d_{CC}(\bar{q}, q) \leq d_n(p,q) + O(\frac{1}{\sqrt{n}}).$$

□

9.3.3 Toward the general setting: Grönwall Lemma

For a general stratified group, the proof of the analogue result is slightly more involved since it may not be true that the analogue of the curve η ends at q.

However, we still have the property that, since γ and σ have very similar tangents, then their endpoints are close. The precise statement is the following, for which we use the notation that if ξ is a curve on a Lie group G and $\dot{\xi}(t)$ is it tangent vector at time t, which is a vector at $\xi(t)$, we denote by $\xi'(t) := (L_{\xi(t)})^* \dot{\xi}(t)$ its representative in the Lie algebra.

Lemma 9.3.5 (Grönwall Lemma)**.** *Let G be a Lie group, $\|\cdot\|$ a norm on T_{1_G}, d a Riemannian distance on G, $\nu > 0$. Then there is C such that for all $\epsilon > 0$, for all $\gamma, \sigma : [0,1] \to G$ absolutely continous curves such that $\gamma(0) = \sigma(0)$, $\|\gamma'\|, \|\sigma'\| \le \nu$ a.e., and $\|\gamma' - \sigma'\| < \epsilon$ a.e., then*

$$d(\gamma(1), \sigma(1)) \le C\epsilon.$$

Proof. Notice that the image of γ and σ are in a bounded set determined by d and ν. For simplicity, we assume that we are in exponential coordinates and that the distance d is given by the norm $\|\cdot\|$. Since the map $(g, v) \mapsto (L_g)_* v$ is smooth, then it is Lipschitz on bounded sets. Hence there is $K > 0$ such that

$$\begin{aligned} \|\dot\gamma - \dot\sigma\| = \|(L_\gamma)_*\gamma' - (L_\sigma)_*\sigma'\| \le \\ \le \|(L_\gamma)_*\gamma' - (L_\gamma)_*\sigma'\| + \|(L_\gamma)_*\sigma' - (L_\sigma)_*\sigma'\| \le \\ \le K \cdot \|\gamma' - \sigma'\| + K \cdot \|\gamma - \sigma\|. \end{aligned}$$

Set $f(t) := \|\gamma(t) - \sigma(t)\|^2$. Then, using that $2ab \le a^2 + b^2$, we get

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t} f = 2\langle \gamma - \sigma, \dot\gamma - \dot\sigma \rangle \le 2\|\gamma - \sigma\| \cdot \|\dot\gamma - \dot\sigma\| \le \\ \le 2K\|\gamma - \sigma\| \cdot \|\gamma' - \sigma'\| + 2K\|\gamma - \sigma\|^2 \le \\ \le K(\|\gamma - \sigma\|^2 + \|\gamma' - \sigma'\|^2) + 2K\|\gamma - \sigma\|^2 = 3Kf + K\epsilon^2 \end{aligned}$$

Then

$$\frac{\mathrm{d}}{\mathrm{d}t}(e^{-3Kt} f(t)) = -3Ke^{-3Kt} f'(t) = e^{-3kt}(f'(t) - 3Kf(t)) \le e^{-3Kt} K\epsilon^2$$

Therefore

$$\begin{aligned} e^{-3K} f(1) = e^{-3Kt} f(t)|_{t=0}^1 = \int_0^1 \frac{\mathrm{d}}{\mathrm{d}t}(e^{-3Kt} f(t))\,\mathrm{d}t \le \\ \le \int_0^1 e^{-3Kt} K\epsilon^2 \,\mathrm{d}t = \left.\frac{e^{-3Kt} K\epsilon^2}{-3K}\right|_{t=0}^1 = \frac{e^{-3K} K\epsilon^2}{-3K} - \frac{K\epsilon^2}{-3K} \end{aligned}$$

Thus

$$f(1) \leq e^{3K}(\frac{1}{3} - \frac{e^{-3K}}{3})\epsilon^2$$

and

$$\|\gamma(1) - \sigma(1)\| \leq \sqrt{\frac{e^{3K/2} - 1}{3}}\epsilon.$$

□

Exercise 9.3.6. Let d_1, d_2 be two left-invariant boundedly compact distances on a Lie group G inducing the manifold topology. Then the increasing function $\xi : (0, \infty) \to (0, \infty)$ defined by

$$\xi(r) = \operatorname{diam}_{d_1}\left(\overline{B}_{d_2}(1_G, r)\right)$$

is such that $\xi(r) \to 0$, as $r \to 0$, and $d_1(p, q) \leq \xi(d_2(p, q))$. ■

9.3.4 Asymptotic cones of Riemannian stratified groups

Theorem 9.3.7. *Let G be a Lie group and let $\Delta \subset TG$ be a bracket generating left-invariant distribution. Let $\Delta^\perp$ be a left-invariant distribution such that $\Delta_{1_G} \oplus \Delta^\perp_{1_G} = T_{1_G}G$. Let $(\langle\cdot,\cdot\rangle_n)_{n\in\mathbb{N}\cup\{0\}}$ be a sequence of left-invariant Riemannian metrics on G such that*

i. Δ_{1_G} is orthogonal to $\Delta^\perp_{1_G}$ with respect to every $\langle\cdot,\cdot\rangle_n$, for all $n \in \mathbb{N} \cup \{0\}$;

ii. for all $X \in \Delta$, we have $\|X\|_n = \|X\|_0$, for all $n \in \mathbb{N}$;

iii. for all $X \notin \Delta$, we have $\|X\|_n \to +\infty$, as $n \to \infty$.

Let d_{CC} be the subRiemannian distance associated to Δ and $\langle\cdot,\cdot\rangle_0$, and d_n the Riemannian distance associated to $\langle\cdot,\cdot\rangle_n$. Then d_n converges uniformly on compact sets to d_{CC}. In fact, for all $R > 0$ there exists an infinitesimal sequence ϵ_n such that

$$d_n(p, q) \leq d_{CC}(p, q) \leq d_n(p, q) + \epsilon_n, \qquad \forall p, q \in B_{CC}(1_G, R) \tag{9.3.8}$$

for all n large enough.

Proof. The left-hand side of (9.3.8) is obvious from (ii).

For the right-hand side, begin by noticing that the unit tangent bundle of $\Delta^\perp_{1_g}$ is compact. Consequently from (iii), if $K > 0$ then for all n large enough we have

$$K \cdot \|Z\|_0 \leq \|Z\|_n, \qquad \forall Z \in \Delta^\perp_{1_g}. \tag{9.3.9}$$

Take $R > 0$ and $p, q \in B_{CC}(1_G, R)$, so $p, q \in B_{d_n}(1_G, R)$. Let $\gamma = \gamma_n : [0, 1] \to G$ be a curve from p to q such that $L_{d_n}(\gamma) = d_n(p, q)$. Consequently, we have $\|\dot\gamma\|_n < 2R$. For all $t \in [0, 1]$ we decompose $\gamma' := (L_\gamma)^*\dot\gamma$ as $\gamma'(t) = X(t) + Z(t)$ with $X(t) \in \Delta_{1_G}$ and $Z(t) \in \Delta_{1_G}^\perp$.From (i) we know that $Z \perp X$ for every of the Riemannian metrics. Let $\sigma : [0, 1] \to G$ be a solution of $\sigma(0) = p$ and $\dot\sigma(t) = (L_\sigma)_* X(t)$. Then

$$K \cdot \|Z\|_0 \overset{(9.3.9)}{\leq} \|Z\|_n \overset{Z\perp X}{\leq} \|X + Z\|_n = \|\dot\gamma\|_n < 2R.$$

Let $\xi(t) := \mathrm{diam}_{d_{CC}}(\overline{B}_{d_0}(1_G, r))$ as in Exercise 9.3.6. Then we are going to use Lemma 9.3.5 since $\|\gamma'\|_0, \|\sigma'\|_0 \leq \|\gamma'\|_n < 2R$ and $\|\gamma' - \sigma'\|_0 = \|Z\|_0 < \frac{2R}{K}$ and get

$$\begin{aligned}
d_{CC}(p, q) &\leq d_{CC}(p, \sigma(1)) + d_{CC}(\sigma(1), \gamma(1)) \\
&\leq L_{CC}(\sigma) + \xi(d_0(\sigma(1), \gamma(1))) \\
&\leq L_{d_n}(\gamma) + \xi\left(C \cdot \frac{2R}{K}\right) \\
&= d_n(p, q) + o(1) \quad \text{as } n \to \infty.
\end{aligned}$$

□

Corollary 9.3.10. *Let G be a stratified group equipped with a Riemannian structure for which the stratification is orthogonal. Then the asymptotic cone of G is a Carnot group. In fact, if d is the Riemannian distance, then there exist Riemannian distances d_λ on G such that $d_\lambda \to d_{CC}$ uniformly on compact sets and $(G, \frac{1}{\lambda}d)$ is isometric to (G, d_λ), and (G, d_{CC}) is a Carnot group.*

9.3.5 Asymptotic cones of subFinsler groups

[...]

9.4 Tangent spaces*

Carnot groups are the tangents of subRiemannian manifolds at regular points. Such as result, originally attributed to Mitchell, is quite technical and involved, see [Bel96, Jea14]. We shall give a complete proof in a specific example, in which the reader can already observe the strategy. Later we shall give the proof of the general result, but without enter too much in the details of the argument.

9.4.1 Preparatory example: The subRiemannian rototranslation group

From the neurogeometry point of view, the most important subRiemannian manifold that is not a Carnot group is the rototranslation group. We begin by proving Mitchell's theorem for such a space.

Theorem 9.4.1. *The tangent space of the subRiemannian rototranslation group is the subRiemannian Heisenberg group.*

Quantitative Chow's theorem

The following proposition gives an explicit proof of Chow's theorem and Ball-Box theorem. Moreover, it gives a uniform estimate for sequences of structures. We denote by $p\exp(X) = \Phi_X^1(p)$ the flow at time 1 from p along X, and, for $t<0$, we denote by $\sqrt{t}$ the value $-\sqrt{-t}$.

Proposition 9.4.2. *Let X_λ, Y_λ be a pair of vector fields in $\mathbb{R}^3$ that depend smoothly on $\lambda \in [0,1]$. Assume $X_\lambda, Y_\lambda, [X_\lambda, Y_\lambda]$ is a frame of $\mathbb{R}^3$ for all λ. Consider the map (composition of flows)*

$$\Phi_\lambda^p(t_1,t_2,t_3) := p\exp(t_1X_\lambda)\exp(t_2Y_\lambda)\exp(\sqrt{t_3}X_\lambda)\exp(\sqrt{t_3}Y_\lambda)\exp(-\sqrt{t_3}X_\lambda)\exp(-\sqrt{t_3}Y_\lambda)$$

Then

1. *Φ_λ^p is smooth and $(\mathrm{d}\Phi_\lambda^p)_0$ has maximal rank.*
2. *The biLipschitz constant of $(\mathrm{d}\Phi_\lambda^p)_0$ is bounded when $\lambda \in [0,1]$ and p is in a compact set.*
3. *There exist $C>0$ and $R>0$ such that for all $\lambda \in [0,1]$ and for all $r \in (0,R)$, for all $p \in B_E(0,R)$*
$$\Phi_\lambda^p(B_E(0,Cr)) \supset B_E(p,r).$$
4. *If d_λ is the subRiemannian distance for which X_λ, Y_λ are orthonormal, then there are $C>0$ and $R>0$ such that for all $p,q \in B_E(0,R)$ and all $\lambda \in [0,1]$*
$$d_\lambda(p,q) \le C\sqrt{d_E(p,q)}.$$

Proof. One has that $(\partial_{t_1}\Phi_\lambda^p)(0) = X_\lambda(p)$, $(\partial_{t_2}\Phi_\lambda^p)(0) = Y_\lambda(p)$, and $(\partial_{t_3}\Phi_\lambda^p)(0) = [X_\lambda, Y_\lambda](p)$.

Hence, $(\mathrm{d}\Phi_\lambda^p)(0)$ has rank 3. Moreover, there is $C>0$ such that every nonzero vector $v \in \mathbb{R}^3$ is such that

$$\|(\mathrm{d}\Phi_\lambda^p)(x)(v)\| \ge C\|v\|$$

for x in a compact set.

By continuity in λ, we can take C uniform when $\lambda \in [0,1]$. In other words, $(\Phi_\lambda^p)^{-1}$ is C^{-1}-Lipschitz in a neighborhood of p for all $\lambda \in [0,1]$.

Part *(iii)* follows from the Inverse Mapping Theorem.

Regarding (iv), notice that

$$\begin{aligned} d_\lambda(p, \Phi^p_\lambda(t_1,t_2,t_3)) &\le |t_1| + |t_2| + 4\sqrt{|t_3|} \\ &\le K\sqrt{\|(t_1,t_2,t_3)\|_E} \end{aligned}$$

for some $K > 0$ and for all $t_1, t_2, t_3 \in (0,1)$.

Let R as in (iii), take $p, q \in B_E(0, \frac{R}{2})$ so for $r = d_E(p,q)$

$$q \in \overline{B}_E(p,r) \subset \Phi^p_\lambda(\overline{B}_E(0, Cr))$$

i.e., there are t_1, t_2, t_3 with $\|(t_1,t_2,t_3)\|_E < Cr$ such that $q = \Phi^p_\lambda(t_1,t_2,t_3)$. Hence,

$$d_\lambda(p,q) \le K\sqrt{\|(t_1,t_2,t_3)\|_E} \le K\sqrt{Cr} = K\sqrt{C}\sqrt{d_E(p,q)}$$

□

Proof of Theorem 9.4.1

An explicit restatement of Theorem 9.4.1 is the following.

Theorem 9.4.3. *In $\mathbb{R}^3$ with coordinates x, y, θ let*

$$\begin{aligned} X &= \cos\theta\partial_x + \sin\theta\partial_y & Y &= \partial_\theta \\ X_\infty &= \partial_x + \theta\partial_y & Y_\infty &= \partial_\theta \\ X_n &= \cos\frac{\theta}{n}\partial_x + n\sin\frac{\theta}{n}\partial_y & Y_n &= \partial_\theta \qquad \forall n \in \mathbb{N} \end{aligned}$$

Let d (resp. d_n, resp d_∞) be the subRiemannian distance for which X, Y (resp. X_n, Y_n, resp. X_∞, Y_∞) are orthonormal. Then

1. *$(\mathbb{R}^3, nd)$ is isometric to $(\mathbb{R}^3, d_n)$.*
2. *For all $R > 0$ there exists $\epsilon_n \to 0$ such that for all $p, q \in B_{d_\infty}(0,R)$*

$$|d_n(p,q) - d_\infty(p,q)| < \epsilon_n,$$

i.e., $d_n \to d_\infty$ uniformly on compact sets.

Proof. The distance nd is the subRiemannian distance associated to the orthonormal frame $\frac{1}{n}X, \frac{1}{n}Y$. Let

$$\delta_n : (x,y,\theta) \mapsto (nx, n^2y, x\theta).$$

Then

$$\begin{aligned} d\delta_n(\frac{1}{n}X) &= \cos\theta\partial_x + n\sin\theta\partial_y = X_n \circ \delta_n \\ d\delta_n(\frac{1}{n}Y) &= \cdots = Y_n \circ \delta_n. \end{aligned}$$

So δ_n is an isometry between $(\mathbb{R}^3, nd)$ and $(\mathbb{R}^3, d_n)$.

Take $p, q \in B_{d_\infty}(0, R)$. Let σ be a d_∞-geodesic from p to q, $\sigma : [0,1] \to \mathbb{R}^3$, $\|\dot\sigma\|_\infty < 2R$.

$$\dot\sigma = aX_\infty + bY_\infty$$

with $|a|, |b| < 2R$.

Let γ such that $\dot\gamma = aX_n + bY_n$. Then

$$\begin{aligned} |\dot\sigma - \dot\gamma| &\leq |a||X_\infty \circ \sigma - X_n \circ \gamma| + |b||Y_\infty \circ \sigma - Y_n \circ \gamma| \\ &\leq 2R(K|\sigma - \gamma| + \|X_\infty - X_n\|_{L^\infty(B_{d_\infty}(0,R))}) \\ &\leq 2RK|\sigma - \gamma| + 2RK\bar\epsilon_n \end{aligned}$$

where $\bar\epsilon_n = \sup_{B_{d_\infty}(0,R)} |X_n - X_\infty|$. Notice that $\bar\epsilon_n \to 0$, because $X_n \to X_\infty$ uniformly on compact sets.

From Grönwall Lemma (see TakeHome exam), we get

$$|\gamma(1) - \sigma(1)| = o(1)$$

Then, by Proposition 9.4.2

$$\begin{aligned} d_n(p,q) &\leq d_n(p, \gamma(1)) + d_n(\gamma(1), \sigma(1)) \\ &\leq L_{d_n}(\gamma) + C\sqrt{\gamma(1) - \sigma(1)} \\ &\leq L_{d_\infty}(\sigma) + o(1) \\ &= d_\infty(p,q) + o(1). \end{aligned}$$

In particular, $d_n(p,1) \leq 3R$ for n large enough.

Let γ be a d_n-geodesic from p to q, $\gamma : [0,1] \to \mathbb{R}^3$ with $\|\dot\gamma\|_n < 3R$.

$$\dot\gamma = aX_n + +bY_n,$$

with $|a|, |b| < 3R$. Let σ be such that $\dot\sigma = aX_\infty + bY_\infty$, then as before $|\gamma(1) - \sigma(1)| = o(1)$.

$$\begin{aligned} d_\infty(p,q) &\le d_\infty(p,\gamma(1)) + d_\infty(\gamma(1),\sigma(1)) \\ &\le L_{d_\infty}(\gamma) + C\sqrt{\gamma(1)-\sigma(1)} \\ &\le L_{d_n}(\sigma) + o(1) \\ &= d_n(p,q) + o(1). \end{aligned}$$

□

9.4.2 Nilpotentization

We explain now what is the Carnot group which appear as tangent to a given equi-regular distribution. Let Δ be a bracket-generating and equi-regular distribution in a manifold M, i.e.,

$$\Delta = \Delta^{[1]} \subset \Delta^{[2]} \subset \ldots \subset \Delta^{[s]} = TM$$

is a flag of sub-bundles of TM, where $\Delta^{[j+1]} = \Delta^{[j]} + [\Delta, \Delta^{[j]}]$. Note that in the last sum is not necessarily a direct sum. The simple but crucial fact is that

$$[\Delta^{[k]}, \Delta^{[l]}] \subseteq \Delta^{[k+l]}. \tag{9.4.4}$$

Equation (9.4.4) is obvious for $k = 1$ and can be proved by induction using Jacobi identity:

$$\begin{aligned} [\Delta^{[k+1]}, \Delta^{[l]}] &= \left[\Delta^{[k]} + [\Delta, \Delta^{[k]}], \Delta^{[l]}\right] \\ &= [\Delta^{[k]}, \Delta^{[l]}] + \left[[\Delta, \Delta^{[k]}], \Delta^{[l]}\right] \\ &\subseteq \Delta^{[k+l]} + \left[[\Delta^{[k]}, \Delta^{[l]}], \Delta\right] + \left[[\Delta^{[l]}, \Delta], \Delta^{[k]}\right] \\ &\subseteq \Delta^{[k+l]} + [\Delta^{[k+l]}, \Delta] + [\Delta^{[l+1]}, \Delta^{[k]}] \\ &\subseteq \Delta^{[k+l]} + \Delta^{[k+l+1]} + \Delta^{[k+l+1]} \\ &\subseteq \Delta^{[k+l+1]} \end{aligned}$$

Define $H_1 := \Delta$ and $H_j := \Delta^{[j]}/\Delta^{[j-1]}$, for $j = 2, \ldots, n$. Still H_j is a bundle over M, but not a sub-bundle of the tangent bundle TM. We obviously have the following isomorphism

$$TM \simeq H_1 \oplus H_2 \oplus \ldots \oplus H_s.$$

In this notes we also assume that the equi-regular distributions have the further property of having a global framing $X_1, \ldots, X_n$ of M such that, for some $m_1, \ldots, m_s$,

$$\Delta^{[j]}(p) = \mathbb{R}\text{-}\operatorname{span}\{X_1(p), \ldots, X_{m_j}(p)\}, \qquad \forall p \in M.$$

Fact 9.4.5. *For each point $p \in M$, the vector space T_pM inherits the structure of a Carnot group, with respect the stratification $H_j(p)$. Such Carnot group is sometimes called the* nilpotization *of T_pM with respect to Δ.*

The following proof is incomplete - a new proof will be given in the future - for now see [Bul02]. Let $V_j := H_j(p)$. Obviously T_pM and $V_1 \oplus \cdots \oplus V_s$ are isomorphic vector spaces. We need to define a Lie algebra product and then show that $[V_j, V_1] = V_{j+1}$. Take $x, y \in T_pM$, with $x \in V_j$ and $y \in V_l$. Since $V_j = H_j(p) = \Delta^{[j]}(p)/\Delta^{[j-1]}(p)$, we have that there exist $X \in \Delta^{[j]}$ and $Y \in \Delta^{[l]}$, such that

$$x = X(p) + \Delta^{[j-1]}(p) \qquad \text{and} \qquad y = Y(p) + \Delta^{[l-1]}(p).$$

We define, naturally,

$$[x, y] := [X, Y](p) + \Delta^{[j+l-1]}(p).$$

The definition is well posed because of (9.4.4): if $u \in \Delta^{[j-1]}$, then $[X + u, Y] = [X, Y] + [u, Y]$, with $[u, Y] \in [\Delta^{[j-1]}, \Delta^{[l]}] \subseteq \Delta^{[j+l-1]}$. Thus $[X + u, Y](p)$ and $[X, Y](p)$ are equal mod $\Delta^{[j+l-1]}(p)$. NEED TO SHOW INDEPENDENCE FROM THE REPRESENTATIVE X.

Again, if $y \in V_1$, from (9.4.4) we immediately have that $[x, y] \in \Delta^{[j+1]}(p)/\Delta^{[j]}(p) = V_{j+1}$. Thus $[V_j, V_1] \subseteq V_{j+1}$. To show the reverse inclusion, let $z \in V_{j+1}$. Consider a representative $Z \in \Delta^{[j+1]}$ such that $z = Z(p) + \Delta^{[j]}(p)$. By definition $\Delta^{[j+1]} = \Delta^{[j]} + [\Delta^{[j]}, \Delta]$, so there are $W \in \Delta^{[j]}$, $X_l \in \Delta^{[j]}$, and $Y_l \in \Delta$ such that $Z = W + \sum_l [X_l, Y_l]$. Take $x_l = X_l(p) \pmod{\Delta^{[j-1]}}$ and $y_l = Y_l(p)$. We have then

$$\begin{aligned}
\sum_l [x_l, y_l] &= \sum_l [X_l, Y_l](p) \qquad \pmod{\Delta^{[j]}(p)} \\
&= (Z - W)(p) \qquad \pmod{\Delta^{[j]}(p)} \\
&= Z(p) \qquad \pmod{\Delta^{[j]}(p)}.
\end{aligned}$$

Therefore we have shown that $[V_j, V_1] = V_{j+1}$. □

9.4.3 Mitchell's theorem on tangent cones

Given a metric space (X, d), one defines the dilated metric space $(X, \lambda d)$ dilated by a factor of $\lambda \in \mathbb{R}$ as the same set X endowed with the dilated distance $(\lambda d)(p, q) := \lambda d(p, q)$. Gromov has defined the notion of tangent space to a metric space as limit of such objects.

We say that a metric space (Z,ρ) is *a* tangent of (X,d) at the point $p \in X$ if there exists $\bar{p} \in Z$ and a sequence $\lambda_j \to \infty$ such that

$$\lim_j (X, p, \lambda_j d) = (Z, \bar{p}, \rho).$$

It signifies [1] that for each $r > 0$, there is a sequence of $\epsilon_j \to 0$ such that the ball of radius $r + \epsilon_j$ in $(X, \lambda_j d)$ about the base point p converges to the ball of radius r about $\bar{p}$. Namely, the infimum of the Gromov-Hausdorff distance between these compact abstract metric spaces approach 0 as $\lambda_j \to \infty$.

The Gromov-Hausdorff distance $\mathrm{GH}(B_1, B_2)$ between two compact metric spaces B_1 and B_2 is infimum $\inf_{\psi_1,\psi_2} \mathrm{H}(\psi_1 B_1, \psi_2 B_2)$ over all isometric embeddings ψ_1, ψ_2 of B_1 and B_2 into the same metric space C of the Hausdorff distance $\mathrm{H}(\psi_1 B_1, \psi_2 B_2)$ of the images as subset of C.

A distribution is said to be *equiregular* if, for each j, $\dim \Delta^{[j]}(p)$ is independent of the point p in M.

Theorem 9.4.6 (Mitchell)**.** *For an equiregular distribution Δ on M, the tangent cone of a sub-Riemannian manifold (M, d_{CC}) at $p \in M$ is isometric to (G, d_∞) where G is a Carnot group with a left-invariant Carnot-Carathéodory metric. In fact, the group G is the nilpotization of T_pM with respect to Δ.*

Remark 9.4.7. The simple fact that we would like the reader to observe is that the tangent cone of a Carnot group $\mathbb{G}$ is $\mathbb{G}$ itself. Indeed, dilations δ_λ provide isometries between (G, d_{CC}) and $(G, \lambda d_{CC})$.

Remark 9.4.8. Differently from the Riemannian case, it is NOT true that a sub-Riemannian manifold is locally biLipschitz equivalent to its tangent cone. It is however true for contact manifolds because of Darboux Theorem.

9.4.4 Margulis-Mostow's blow-up theorem

The Rademacher-type theorem for manifolds is attributed to Margulis and Mostow [MM95], who however, extended the proof by Pansu for the case of Carnot groups [Pan89].

[1] In the case when the metric space (X,d) is geodesic, the limit should be easier to understand. Look at [BBI01, page 272].

Theorem 9.4.9 (SubRiemannian Rademacher Theorem)**.** *At almost all points, the tangent map of a Lipschitz map between sub-Finsler equiregular manifolds exists, is unique, and is a group homomorphism of the tangent cones equivariant with respect to their dilations.*

Let us clarify what is the meaning of tangent map. Each map $f : (X, d) \to (X', d')$ induces a map $f_\lambda : (X, \lambda d) \to (X', \lambda d')$, for each $\lambda > 0$, which set-wise is the same map $f(x) = f_\lambda(x)$. Fix a point $x \in X$ and assume that (Z, ρ) and (Z', ρ') are tangent spaces respectively to (X, d) at x and to (X', d') at $f(x)$. One says that $\hat{f} : (Z, \rho) \to (Z', \rho')$ is a *tangent map* of f at x if, for some sequence $\lambda_j \to \infty$, f_{λ_j} converges to $\hat{f}$ in what sense?

Let us warn the reader about a possible confusion. Each sub-Riemannian manifold is in particular a differentiable manifold. However, the notion of the differential of a smooth map does not coincide with the tangent map which is defined in geometric terms. However, there is a link between the two tangent maps, see Exercise **??**.

9.5 Varying CC bundle structures*

Let M be a smooth manifold. Let

$$f : M \times \mathbb{R}^m \to TM$$

be a smooth M-bundle morphism. Let

$$N : M \times \mathbb{R}^m \to [0, +\infty)$$

be a continuous function such that $N(p, \cdot)$ is a norm for every $p \in M$.

The couple (f, N) induces a CC-structure as follows. For a fixed $o \in M$ and $u \in L^\infty([0, 1]; \mathbb{R}^m)$ we consider the following Cauchy problem

$$\begin{cases} \gamma'(t) & = f(\gamma(t), u(t)), \\ \gamma(0) & = o. \end{cases}$$

The solution of the previous problem will be denoted by $\gamma_{(o,f,u)}$. Hence one can define

$$d_{(f,N)}(p, q) := \inf \left\{ \int_0^1 N(\gamma(s), u(s))\, \mathrm{d}s : \gamma = \gamma_{(p,f,u)}, \gamma(1) = q \right\}. \tag{9.5.1}$$

Notice that the set in the infimum above could be empty. In that case $d_{(f,N)}(p, q) = +\infty$. Every couple (f, N) as above will be called a *CC-bundle structure*.

Definition 9.5.2 (Varying CC-bundle structure)**.** Let $\Lambda \subseteq \mathbb{R}$ be a set. Let M be a smooth manifold. Let $f : \Lambda \times M \times \mathbb{R}^m \to TM$ and $N : \Lambda \times M \times \mathbb{R}^m \to [0, +\infty)$ be maps such that for every $\lambda \in \Lambda$ we have that (f_λ, N_λ) is a CC-bundle structure, where $f_\lambda := f(\lambda,\cdot,\cdot)$ and $N_\lambda := N(\lambda,\cdot,\cdot)$. We say that the family $\{(f_\lambda, N_\lambda)\}_{\lambda\in\Lambda}$ is a *(smoothly) varying CC-bundle structure* if

- $f \in C^\infty(\Lambda \times M \times \mathbb{R}^m)$;
- $N \in C^0(\Lambda \times M \times \mathbb{R}^m)$;
- for every $(\lambda, v) \in \Lambda \times \mathbb{R}^m$ the vector field

$$M \ni p \mapsto f(\lambda, p, v) \in TM$$

  is smooth.

The above definition can be generalised to 'continuously varying Lipschitz-vector-fields structures', for which the results in this section have analogues, see [**?**].

We shall prove the following theorem.

Theorem 9.5.3. *Let $\Lambda \subseteq \mathbb{R}$, and let $\{(f_\lambda, N_\lambda)\}_{\lambda\in\Lambda}$ be a varying CC-bundle structure on a manifold M. Let $d_\lambda := d_{(f_\lambda,N_\lambda)}$ for every $\lambda \in \Lambda$, as in* (9.5.1)*. Let $\lambda_0 \in \Lambda$ be such that $f(\lambda_0, M \times \mathbb{R}^m)$ is a bracket-generating distribution and the metric space (M, d_{λ_0}) is boundedly compact. Then $d_\lambda \to d_{\lambda_0}$ uniformly on compact sets of M as $\lambda \to \lambda_0$.*

We give the proof of the previous theorem using the following crucial lemma.

Lemma 9.5.4 (Equicontinuity of the distances)**.** *In the same assumptions of Theorem 9.5.3, let $K \subseteq M$ be compact set and ρ Riemannian metric on M. Then there exists a neighborhood $I_{\lambda_0} \subseteq \Lambda$ of λ_0, and β homeomorphism of $[0, +\infty)$ such that*

$$d_\lambda(p, q) \le \beta(\rho(p, q)), \qquad \textit{for all } p, q \in K \textit{ and } \lambda \in I_{\lambda_0}.$$

Proof. Let us fix a Riemannian metric ρ on M^n, where n denotes the dimension of the manifold. Let us denote, for $\lambda \in \Lambda$ and $p \in M$,

$$X_i^\lambda(p) := f(\lambda, p, e_i),$$

where $\{e_1, \ldots, e_m\}$ is a standard basis of $\mathbb{R}^m$. Let us fix some $x \in M$ from now on. We know that $\{X_i^{\lambda_0}\}_{i=1}^m$ is a bracket-generating set of vector fields. Hence, for every $\eta > 0$, there exist $X_{i_1}^{\lambda_0}, \ldots, X_{i_n}^{\lambda_0}$, where $\{i_1, \ldots, i_n\} \subseteq \{1, \ldots, m\}$ may depend on η, such that the following holds.

There exists $\hat{t} := (\hat{t}_1, \dots, \hat{t}_n)$ with $|\hat{t}| < \eta$ such that the map

$$(t_1, \dots, t_n) \mapsto \Phi^{t_n}_{X^{\lambda_0}_{i_n}} \circ \dots \circ \Phi^{t_1}_{X^{\lambda_0}_{i_1}}(x),$$

has a regular point at $\hat{t}$. Notice now that, since the map f is smooth, then the map

$$\Psi_{(i_1,\dots,i_n)} : (\lambda, s_1, \dots, s_n, t_1, \dots, t_n) \mapsto \Phi^{s_1}_{X^{\lambda}_{i_1}} \circ \dots \Phi^{s_n}_{X^{\lambda}_{i_n}} \circ \Phi^{t_n}_{X^{\lambda}_{i_n}} \circ \dots \circ \Phi^{t_1}_{X^{\lambda}_{i_1}}(x) \tag{9.5.5}$$

is continuous and well defined on $I_{(i_1,\dots,i_n)} \times \overline{B}(0, \xi_{(i_1,\dots,i_n)})$, where $\overline{B}(0, \xi_{(i_1,\dots,i_n)})$ is a sufficiently small neighborhood of 0 in $\mathbb{R}^{2n}$, and $I_{(i_1,\dots,i_n)}$ is a sufficiently small compact neighborhood of λ_0. Let $I_{\lambda_0,x}$ be the intersection of $I_{(i_1,\dots,i_n)}$ over all the possible choices of $\{i_1, \dots, i_n\} \subseteq \{1, \dots, m\}$, and let $\overline{B}(0, \xi)$ be the intersection of $\overline{B}(0, \xi_{(i_1,\dots,i_n)})$ over all the possible choices of $\{i_1, \dots, i_n\} \subseteq \{1, \dots, m\}$. Let K be the union of $\Psi_{(i_1,\dots,i_n)}(I_{\lambda_0,x} \times \overline{B}(0, \xi))$, over all the possible choices of $\{i_1, \dots, i_n\} \subseteq \{1, \dots, m\}$. Hence, by continuity of N, there exists $L > 0$ such that

$$N(\lambda, p, v) \le L|v|, \quad \text{for all } \lambda \in I_{\lambda_0,x} \text{ and } p \in K. \tag{9.5.6}$$

Let us prove the following claim. We recall that $x \in M$ is fixed.

Claim. For every $\varepsilon > 0$ there exists δ such that

$$B_\rho(x, \delta) \subseteq B_{d_\lambda}(x, \varepsilon), \quad \text{for all } \lambda \in I_{\lambda_0,x},$$

where $I_{\lambda_0,x}$ is defined above.

To prove the claim, take $\nu := \min\{\xi/4, \varepsilon/(4nL)\}$, where ξ is defined above. Hence there exists $\hat{t}$ with $|\hat{t}| < \nu$ and $\{i_1, \dots, i_n\} \subseteq \{1, \dots, m\}$ such that the map $\Psi_{(i_1,\dots,i_n)}(\lambda_0, -\hat{t}_1, \dots, -\hat{t}_n, t_1, \dots, t_n)$ is a diffeomorhpsim between a neighborhood $\hat{U}$ of $\hat{t}$ (that can be taken contained in $\overline{B}(0, 2\nu) \subseteq \mathbb{R}^n$) and a neighborhood of $x \in M$. By the continuity of the map $\Psi_{(i_1,\dots,i_n)}$ we get that the convergence

$$\Psi_{(i_1,\dots,i_n)}(\lambda, -\hat{t}, t_1, \dots, t_n) \to \Psi_{(i_1,\dots,i_n)}(\lambda_0, -\hat{t}, t_1, \dots, t_n), \quad \text{for } \lambda \to \lambda_0,$$

is uniform on $(t_1, \dots, t_n) \in \hat{U}$. Hence, applying **??** and **??**, we have that there exists $\delta > 0$ such that

$$B_\rho(x, \delta) \subseteq \Psi_{(i_1,\dots,i_n)}(\lambda, \hat{t}, \hat{U}), \quad \text{for all } \lambda \in I_{\lambda_0,x}.$$

Since $\hat{U} \subseteq \overline{B}(0, \xi/(2nL))$ we get that for every $(s_1, \dots, s_n) \in \hat{U}$ we have

$$|s_1| + \dots + |s_n| \le \varepsilon/(2L).$$

Moreover, also $|\hat{t}_1|+\dots|\hat{t}_n| \le \varepsilon/(2L)$, and then from the explicit expression (9.5.5) and the estimate (9.5.6), we get that the endpoint of the concatenation of the curves associated to $\Psi_{(i_1,\dots,i_n)}(\lambda,\hat{t},s_1,\dots,s_n)$ for every $(s_1,\dots,s_n)\in\hat{U}$ has length $\le\varepsilon$ for every $\lambda\in I_{\lambda_0,x}$. Hence

$$B_\rho(x,\delta)\subseteq B_{d_\lambda}(x,\varepsilon),\quad \text{for all } \lambda\in I_{\lambda_0,x},$$

which is the sought claim.

Now a routine compactness argument based on Claim 1. shows that, given a compact $K\subseteq M$, there exists a compact interval $I_{\lambda_0,K}\subseteq\Lambda$ of λ_0 such that for every $\varepsilon>0$ there exists δ such that

$$B_\rho(x,\delta)\subseteq B_{d_\lambda}(x,\varepsilon),\quad \text{for all } \lambda\in I_{\lambda_0,K}, \text{ for all } x\in K.$$

From the previous conclusion, the proof of the lemma follows. □

Proof of Theorem 9.5.3. We embed M isometrically into some $\mathbb{R}^N$, on which we denote with $|\cdot|$ the standard norm. Let us fix a compact set K and a Riemannian metric ρ on M. Notice that on every compact set of M, ρ and $|\cdot|$ are biLipschitz equivalent. Let us fix $0<\varepsilon<1$.

By continuity, there exists a constant $C>0$ such that $d_{\lambda_0}(p,q)\le C$ for every $p,q\in K$. Let $K':=\overline{B}_{\lambda_0}(K,C+1)$ the closed tubular neighborhood of K of radius $C+1$. Since (M,d_{λ_0}) is boundedly compact, we deduce that K' is compact.

Let β be the functions, and I_{λ_0} be the compact neighborhood of λ_0, associated to K' given from Lemma 9.5.4. Notice that for every $p,q\in K$ and for every $\lambda\in I_{\lambda_0}$ we have that

$$d_\lambda(p,q)\le\beta(|p-q|)\le\beta(\operatorname{diam}_{|\cdot|}K).$$

Since $N(\lambda,p,\cdot)$ is a norm for every $\lambda\in I_{\lambda_0}$ and every $p\in M$, and since N is continuous, we get that there exists a compact set $K''\subseteq\mathbb{R}^m$ such that

$$\text{if } N(\lambda,x,v)\le\beta(\operatorname{diam}_{|\cdot|}K)+1 \text{ for some } \lambda\in I_{\lambda_0} \text{ and } x\in K', \text{ then } v\in K''. \tag{9.5.7}$$

Moreover, by definition of varying CC-structures, we have that there exists $L>0$ such that for every $\lambda\in I_{\lambda_0}$ and $v\in K''$ the map

$$K'\ni p\mapsto f(\lambda,p,v),$$

is L-lipschitz.

Because of continuity of the functions N and f we get that there exist $0<\delta_2<\delta_1<\varepsilon$ such that $\overline{B}(\lambda_0,\delta_2)\subseteq I_{\lambda_0}$ and

$$|N(\lambda_0,x,v)-N(\lambda,y,v)|<\varepsilon,\quad \text{for all } \lambda\in\overline{B}(\lambda_0,\delta_2),\ x\in K',\ v\in K'',\ y\in\overline{B}_{|\cdot|}(x,\delta_1), \tag{9.5.8}$$

and

$$|f(\lambda_0, x, v) - f(\lambda, x, v)| < a, \quad \text{for all } \lambda \in \overline{B}(\lambda_0, \delta_2),\ x \in K',\ v \in K'', \tag{9.5.9}$$

where a is chosen such that $a\frac{e^L - 1}{L} < \delta_1$.

We claim that for every $\lambda \in \overline{B}(\lambda_0, \delta_2)$ and every $p, q \in K$, we have

$$d_{\lambda_0}(p, q) \leq d_\lambda(p, q) + 2\varepsilon + \beta(\varepsilon). \tag{9.5.10}$$

Indeed, fix p, q, λ as in the claim. Up to reparametrization, we can take a curve γ_λ connecting p and q such that $\gamma_\lambda' = f(\lambda, \gamma_\lambda, u_\lambda)$ and

$$N(\lambda, \gamma_\lambda(t), u_\lambda(t)) \leq d_\lambda(p, q) + \epsilon, \quad \text{for a.e. } t \in [0, 1]. \tag{9.5.11}$$

Let $B := \overline{B}_{\lambda_0}(p, d_{\lambda_0}(p, q))$. Notice that $B \subseteq K'$. Define

$$\bar{t} := \max\{t \in [0, 1] : \gamma_\lambda(s) \in B\ \forall s \in [0, t]\}.$$

Denote $q_\lambda' := \gamma_\lambda(\bar{t})$ and notice that $d_{\lambda_0}(p, q_\lambda') = d_{\lambda_0}(p, q)$. Moreover notice that $(\gamma_\lambda)|_{[p, q_\lambda']} \subseteq K'$. Take now $\gamma_{\lambda,0}$ such that $\gamma_{\lambda,0}' = f(\lambda_0, \gamma_{\lambda,0}, u_\lambda)$ and $\gamma_{\lambda,0}(0) = p$. Call $\overline{q}_\lambda := \gamma_{\lambda,0}(\bar{t})$.

We shall estimate $|\overline{q}_\lambda - q_\lambda'|$. From (9.5.11), (9.5.7), and the fact that $\gamma_\lambda([0, \bar{t}]) \in K'$ we get that $u_\lambda(t) \in K''$ for a.e. $t \in [0, \bar{t}]$. Hence we estimate, for every $x, y \in K'$ and a.e. $t \in [0, \bar{t}]$,

$$\begin{aligned} |f(\lambda, x, u_\lambda(t)) - f(\lambda_0, y, u_\lambda(t))| &\leq |f(\lambda, x, u_\lambda(t)) - f(\lambda_0, x, u_\lambda(t))| \\ &\quad + |f(\lambda_0, x, u_\lambda(t)) - f(\lambda_0, y, u_\lambda(t))| \\ &\leq a + L|x - y|. \end{aligned} \tag{9.5.12}$$

Hence Grönwall Lemma in **??** applied on K' directly implies that

$$|\gamma_\lambda(t) - \gamma_{\lambda,0}(t)| \leq a\frac{e^{Lt} - 1}{L} < \delta_1 < \varepsilon, \quad \text{for a.e. } t \in [0, \bar{t}], \tag{9.5.13}$$

and morevoer that $(\gamma_{\lambda,0})|_{[0,\bar{t}]} \subseteq K'$. Now let us conclude the estimate of the Claim 1. We have

$$\begin{aligned} d_{\lambda_0}(p, q) = d_{\lambda_0}(p, q_\lambda') &\leq d_{\lambda_0}(p, \overline{q}_\lambda) + d_{\lambda_0}(\overline{q}_\lambda, q_\lambda') \\ &\leq \int_0^{\bar{t}} N(\lambda_0, \gamma_{\lambda,0}(s), u_\lambda(s))\,\mathrm{d}s + \beta(|\overline{q}_\lambda - q_\lambda'|) \\ &\leq \int_0^{\bar{t}} N(\lambda, \gamma_\lambda(s), u_\lambda(s))\,\mathrm{d}s + \varepsilon + \beta(\varepsilon) \\ &\leq \int_0^{1} N(\lambda, \gamma_\lambda(s), u_\lambda(s))\,\mathrm{d}s + \varepsilon + \beta(\varepsilon) \\ &\leq d_\lambda(p, q) + \varepsilon + \beta(\varepsilon) + \epsilon, \end{aligned} \tag{9.5.14}$$

where we are using (9.5.13), (9.5.8), and (9.5.11).

We claim that for every $\lambda \in \overline{B}(\lambda_0, \delta_2)$ and every $p, q \in K$, we have

$$d_\lambda(p,q) \leq d_{\lambda_0}(p,q) + 2\varepsilon + \beta(\varepsilon). \tag{9.5.15}$$

Indeed, fix p, q, λ as in the claim. Up to reparametrization, we can take a curve γ connecting p and q such that $\gamma' = f(\lambda_0, \gamma, u)$ and

$$N(\lambda_0, \gamma(t), u(t)) \leq d_{\lambda_0}(p,q) + \epsilon, \quad \text{for a.e. } t \in [0,1]. \tag{9.5.16}$$

Notice that $\gamma \subseteq K'$. Take now γ_λ such that $\gamma_\lambda' = f(\lambda, \gamma_\lambda, u)$ and $\gamma_\lambda(0) = p$. Call $\overline{q}_\lambda := \gamma_\lambda(1)$.

We now want to estimate $|\overline{q}_\lambda - q|$. Arguing verbatim as before we obtain

$$|\gamma_\lambda(t) - \gamma(t)| \leq a\frac{e^{Lt}-1}{L} < \delta_1 < \varepsilon, \quad \text{for a.e. } t \in [0,1], \tag{9.5.17}$$

and moreover $\gamma_\lambda \subseteq K'$. Now let us conclude the estimate of the Claim 2. We have

$$\begin{aligned}
d_\lambda(p,q) &\leq d_\lambda(p, \overline{q}_\lambda) + d_\lambda(\overline{q}_\lambda, q) \\
&\leq \int_0^1 N(\lambda, \gamma_\lambda(s), u(s))\,\mathrm{d}s + \beta(|\overline{q}_\lambda - q|) \\
&\leq \int_0^1 N(\lambda_0, \gamma(s), u(s))\,\mathrm{d}s + \varepsilon + \beta(\varepsilon) \\
&\leq d_\lambda(p,q) + \varepsilon + \beta(\varepsilon) + \epsilon,
\end{aligned} \tag{9.5.18}$$

where we are using (9.5.17), (9.5.8), and (9.5.16).

From (9.5.10) and (9.5.15) jointly with the fact that $\beta(\varepsilon) \to 0$ as $\varepsilon \to 0$ we get the proof of the theorem. □

9.6 A metric characterization of Carnot groups*

The purpose of this section is to give a more axiomatic presentation of Carnot groups from the view point of Metric Geometry. In fact, we shall see that Carnot groups are the only locally compact and geodesic metric spaces that are isometrically homogeneous and self-similar. Such a result follows the spirit of Gromov's approach of 'seeing Carnot-Carathéodory spaces from within', [Gro96].

Let us recall and make explicit the above definitions. A topological space X is called *locally compact* if every point of the space has a compact neighborhood. A metric space is *geodesic* if, for all $p, q \in X$, there exists an isometric embedding $\iota : [0,T] \to X$ with $T \geq 0$ such that $\iota(0) = p$ and $\iota(T) = q$. We say that a metric space X is *isometrically homogeneous* if its group of isometries acts on the space transitively. Explicitly, this means that, for all $p, q \in X$, there exists a distance-preserving homeomorphism $f : X \to X$ such that $f(p) = q$. In this section, we say that a metric space X is *self-similar* if it admits a *dilation*, i.e., there exists $\lambda > 1$ and a homeomorphism $f : X \to X$ such that $d(f(p), f(q)) = \lambda d(p,q),$ for all $p, q \in X$.

Theorem 9.6.1. *The subFinsler Carnot groups are the only metric spaces that are*

1. *locally compact,*
2. *geodesic,*
3. *isometrically homogeneous, and*
4. *self-similar (i.e., admitting a dilation).*

Theorem 9.6.1 provides a new equivalent definition of Carnot groups. Obviously, (1) can be slightly strengthened assuming that the space is *boundedly compact* (the term *proper* is also used), i.e., closed balls are compact.

We point out that each of the four conditions in Theorem 9.6.1 is necessary for the validity of the result. Indeed, let us mention examples of spaces that satisfy three out of the four conditions but are not Carnot groups: every infinite-dimensional Banach space; every snowflake of a Carnot group, e.g., $(\mathbb{R}, \sqrt{\|\cdot\|})$; many cones such as the usual Euclidean cone of cone angle in $(0, 2\pi)$ or the union of two spaces such as $\{(x,y) \in \mathbb{R}^2 : xy \geq 0\}$; every compact homogeneous space such as $\mathbb{S}^1$.

Other papers focusing on metric characterizations of Carnot groups are [LD11b], [Bul11], [Fre12] (which is based on [LD11a]), and [BS14].

9.6.1 Proof of the characterization

The proof of Theorem 9.6.1 is an easy consequence of three hard theorems. We present now these theorems, before giving the proof.

The first theorem is well-known in the theory of locally compact groups. It is a consequence of a deep result of Dean Montgomery and Leo Zippin, [MZ52, Corollary on page 243, Section 6.3], together with the work [Gle52] of Andrew Gleason. An explicit proof can be found in Cornelia Drutu and Michael Kapovich's lecture notes, [DK11, Chapter 14].

Theorem 9.6.2 (Gleason-Montgomery-Zippin)**.** *Let X be a metric space that is connected, locally connected, locally compact and has finite topological dimension. Assume that the isometry group $Isom(X)$ of X acts transitively on X. Then $Isom(X)$ has the structure of a Lie group with finitely many connected components, and X has the structure of an analytic manifold.*

Notice that an isometrically homogeneous space that is locally compact is complete.

Successively, Berestovskii's work [Ber88, Theorem 2] clarified what are the possible isometrically homogeneous distances on manifolds that are also geodesic. They are subFinsler metrics.

Theorem 9.6.3 (Berestovskii)**.** *Under the same assumptions of Theorem 9.6.2, if in addition the distance is geodesic, then the distance is a subFinsler metric, i.e., the metric space X is a homogeneous Lie space G/H and there is a G-invariant subbundle Δ on the manifold G/H and a G-invariant norm on Δ, such that the distance is given by the same formula* (3.1.15).

Tangents, in the Gromov-Hausdorff sense, of subFinsler manifolds have been studied.

Theorem 9.6.4 (Mitchell)**.** *The metric tangents of an equiregular subFinsler manifold are subFinsler Carnot groups.*

Proof of Theorem 9.6.1. Let us verify that we can use Theorem 9.6.2. A geodesic metric space is obviously connected and locally connected. Regarding finite dimensionality, we claim that a locally compact, self-similar, isometrically homogeneous space X is doubling. Namely, there exists a constant $C > 0$ such that every ball of radius $r > 0$ in X can be covered with less than C balls of radius $r/2$. Since X is locally compact, there exists a ball $B(x_0, r_0)$ that is compact. Let $\lambda > 1$ be the factor of the dilation. Hence, the balls $B(x_0, sr_0)$ with $s \in [1, \lambda]$ form a compact family of compact balls. Hence, there exists a constant $C > 1$ such that each ball $B(x_0, sr_0)$ can be covered with less than C balls of radius $sr_0/2$. By self-similarity and homogeneity, every other ball can be covered with less than C balls of half radius. Doubling metric spaces have finite Hausdorff dimension and hence finite topological dimension. Therefore, by Theorem 9.6.2 the isometry group G is a Lie group.

Since the distance is geodesic, Theorem 9.6.3 implies that our metric space is a subFinsler homogeneous manifold G/H. Since the subFinsler structure is G invariant, in particular it is equiregular.

Hence, on the one hand, because of Theorem 9.6.4 the tangents of our metric space are subFinsler Carnot groups. On the other hand, the space admits a dilation, hence, iterating the dilation, we have that there exists a metric tangent of the metric space that is isometric to our original space. Then the space is a subFinsler Carnot group. □

9.6.2 Metric spaces with unique tangents

– tangents of tangents are tangents

– uniqueness of tangents implies group structure on tangents

Chapter 10

Rank-one symmetric spaces*

In the chapter after this one we will show that to every Riemannian symmetric space one can associate a 'visual boundary' that has a structure of Carnot group. In this chapter we review the classical notion of symmetric space.

A *Riemannian symmetric space* is a connected Riemannian manifold M where for each point $p \in M$ there exists an isometry σ_p of M such that $\sigma_p(p) = p$ and the differential of σ_p at p is the multiplication by -1. Simple examples of symmetric spaces are round spheres, Euclidean spaces and real hyperbolic. The *rank* of a symmetric space is the largest dimension of a flat subspace of M, where a *flat of dimension* n in M is a local isometry $\gamma : \mathbb{R}^n \to M$. For example, spheres and hyperbolic spaces have rank 1, whereas Euclidean n-space has rank n. A symmetric space is of *non-compact type* if it is not the product of two symmetric spaces one of which is either compact or Euclidean. Symmetric spaces were first introduced by Élie Cartan in 1926, see [Car26], [Car27]. In particular, he gave a complete description of these spaces by means of the classification of simple Lie algebras.

In this chapter we first prove that every rank-one symmetric space of non-compact type admits a group structure of a semidirect product with a precise formula for a left-invariant distance. The fact that such spaces admit semidirect-product structures has been known at least since Ernst Heintze's work in the 1970's, see [Hei74]. However, the formula for the left-invariant distances cannot be easily traced in literature. To study these spaces we will need the following result: Let M be a rank-one symmetric space of non-compact type, then M is one of the following spaces, which we call $\mathbb{K}$-*hyperbolic spaces* $\mathbb{K}\mathbf{H}^n$, with $n \in \mathbb{N}$: real hyperbolic n-space $\mathbb{R}\mathbf{H}^n$, complex hyperbolic n-space $\mathbb{C}\mathbf{H}^n$, quaternionic hyperbolic n-space $\mathbb{H}\mathbf{H}^n$ or the octonionic plane $\mathbb{O}\mathbf{H}^2$. The proof of such last fact was indicated by Cartan, but completely established in this form in the 1950's, see Arthur Besse's

1978 book [Bes78, Section 3.G] and see Heintze's 1974 paper [Hei74, Section 5] for a geometric proof.

We shall introduce $\mathbb{K}$-hyperbolic spaces as metric spaces. Initially, we restrict to the real, complex, and quaternionic case, which share a similar approach and we shall give a treatment as unified as possible. Following Felix Klein's construction, we shall describe the $\mathbb{K}$-*hyperbolic space* $\mathbb{K}\mathbf{H}^n$ *of dimension* n as an open subset of the projectivization of the space $\mathbb{K}^{n+1}$ equipped with a Hermitian form of type $(n, 1)$. We shall recall the distance function on $\mathbb{K}\mathbf{H}^n$, referring to Martin Bridson and André Häfliger's 1999 book [BH99, Part II, Chapter 10].

To recall the Lie group structure on each $\mathbb{K}\mathbf{H}^n$, we revise the continuous n-th Heisenberg group $G_{n,\mathbb{K}}$ modelled on $\mathbb{K}$ and its intrinsic dilations, see [Ste93, Chapter XII, Section 1] . We shall prove that the semidirect product of $G_{n,\mathbb{K}}$ with $\mathbb{R}$ acts simply transitively and by isometries on $\mathbb{K}\mathbf{H}^n$. We will double check that, after the identification of $\mathbb{K}\mathbf{H}^n$ with $G_{n,\mathbb{K}} \rtimes \mathbb{R}$, the hyperbolic distance is invariant under left translations on $G_{n,\mathbb{K}} \rtimes \mathbb{R}$. We shall write explicitly the distance on the $\mathbb{K}$-hyperbolic n-space modelled as $G_{n,\mathbb{K}} \rtimes \mathbb{R}$ in terms of elementary functions of the coordinates.

Theorem 10.0.1. *For every $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}, \mathbb{H}\}$ and every $n \in \mathbb{N} \setminus \{0\}$, the $\mathbb{K}$-hyperbolic n-space $\mathbb{K}\mathbf{H}^n$ is isometric to the manifold $\mathbb{K}^{n-1} \times \mathrm{Im}(\mathbb{K}) \times \mathbb{R}$ equipped with the multiplication law given by*

$$(u, s; a) \cdot (v, t; b) = (u + e^a v, s + e^{2a} t + \mathrm{Im}(u e^a \overline{v}); a + b)$$

and the left-invariant distance d such that

$$4\cosh^2 d(\mathbf{0}, (v, t; b)) = 4\cosh^2(b) + 2e^{-b}\cosh(b)|v|^2 + e^{-2b}\left(\frac{|v|^4}{4} + |t|^2\right).$$

There is a remaining case: the octonionic hyperbolic plane. It cannot be treated as described above due to the non-associativity of the octonions, and therefore the impossibility to define a notion of a vector space over the octonions. However, in the last section, we will give some basic ideas on how to deal with this case and build the octonionic hyperbolic plane.

10.1 Preliminary notions for rank-one symmetric spaces

10.1.1 Quaternionic numbers

In this section we introduce some notations. We denote by $\mathbb{R}, \mathbb{C}, \mathbb{H}$ the real, complex, and quaternion number sets, respectively. Throughout all the chapter $\mathbb{K}$ will denote one of the above number sets. We shall only recall the quaternions. The quaternions are a 4-dimensional algebra over $\mathbb{R}$ with basis

$\{1, i, j, k\}$, where 1 is central, and i, j, k follow the rules:

$$ij = k, \; jk = i, \; ki = j$$

and

$$i^2 = j^2 = k^2 = -1.$$

If $x \in \mathbb{K}$ we write $\overline{x}$ to denote the $\mathbb{K}$-conjugate of x. Conjugation on $\mathbb{R}$ is trivial. For quaternions, one defines the *conjugate* of $u = a+bi+cj+dk$ as $\overline{u} = a-bi-cj-dk$. We also recall how conjugation works with multiplication, that is, given $u, v \in \mathbb{K}$ it is true that $\overline{uv} = \overline{v}\,\overline{u}$. The *real part* of x is the number $\Re(u) = \frac{u+\overline{u}}{2}$. The *norm* of $|u|$ of $u \in \mathbb{K}$ is the non-negative real number $\sqrt{u\overline{u}}$.

We next recall the *imaginary part* of $u \in \mathbb{K}$ written as $\mathrm{Im}(u)$. If $u \in \mathbb{H}$ is written as $u = a + bi + cj + dk$, then

$$\mathrm{Im}(u) = \begin{pmatrix} b \\ c \\ d \end{pmatrix} \in \mathbb{R}^3$$

and $\mathrm{Im}_i(u)$ denotes the i-component of $\mathrm{Im}(u)$. The product on the quaternions is non commutative, so one must be careful while defining a vector space structure. We define the *left multiplication* and the *right multiplication* to be respectively λu and $u\lambda$ where $u \in \mathbb{K}^n$ and $\lambda \in \mathbb{K}$. For our purpose we say that $x, y \in \mathbb{K}^n$ are *linearly dependent* if there exists $\lambda \in \mathbb{K}$ such that $x = y\lambda$. Note that if $\mathbb{K} = \mathbb{R}$ or $\mathbb{C}$ then the definitions above are equivalent.

10.1.2 Hermitian forms

Let $\mathfrak{M}(\mathbb{K}, k, l)$ be the group of the $k \times l$ matrices over the number set $\mathbb{K}$. Let $A \in \mathfrak{M}(\mathbb{K}, k, l)$ be in the form $A = (a_{ij})$, the *Hermitian transpose* of A is $A^* \in \mathfrak{M}(\mathbb{K}, l, k)$ that satisfies $A^* = (\overline{a}_{ji})$. As with ordinary transpose operation for $\mathbb{C}$, the Hermitian transpose of a product is the product of the Hermitian transposes in the reverse order, that is $(AB)^* = B^*A^*$. A matrix $H \in \mathfrak{M}(\mathbb{K}, n) := \mathfrak{M}(\mathbb{K}, n, n)$ is said to be *Hermitian* if it equals its own Hermitian transpose, i.e., $H = H^*$. We claim that if H is Hermitian and μ is an eigenvalue of H with eigenvector $x \in \mathbb{K}^n$ then μ is real. In order to see this, observe that

$$x^*\mu x = x^*Hx = x^*H^*x = (Hx)^*x = (\mu x)^*x = x^*\overline{\mu}x.$$

Next by multiplying the RHS (Right-Hand Side) and the LHS (Left-Hand Side) on the left by x and on the right by x^* we obtain

$$xx^*\mu xx^* = xx^*\overline{\mu}xx^*.$$

Then we observe that xx^* is a row vector with real elements, therefore it commutes with μ and $\overline{\mu}$. We therefore infer that

$$\mu xx^*xx^* = \overline{\mu}xx^*xx^*$$

that is

$$\mu|x|^4 = \overline{\mu}|x|^4.$$

By the definition of eigenvector we know that x is not the zero vector and therefore $|x|^4 \neq 0$ leading to $\mu = \overline{\mu}$, that is, $\mu \in \mathbb{R}$.

To each Hermitian matrix $H \in \mathfrak{M}(\mathbb{K}, n)$ we associate a *Hermitian form* $\langle \cdot, \cdot \rangle_H : \mathbb{K}^n \times \mathbb{K}^n \to \mathbb{K}$ given by $\langle z, w \rangle_H = z^*Hw$. Hermitian forms are sesquilinear, that is they are conjugate linear in the first factor and linear in the second factor. In other words, for $z, z_1, z_2, w \in \mathbb{K}^n$ and $\lambda \in \mathbb{K}$, we have

$$\langle z_1 + z_2, w \rangle_H = (z_1 + z_2)^*Hw = z_1^*Hw + z_2^*Hw = \langle z_1, w \rangle_H + \langle z_2, w \rangle_H, \tag{10.1.1}$$

$$\langle z\lambda, w \rangle_H = (z\lambda)^*Hw = \overline{\lambda}z^*Hw = \overline{\lambda}\langle z, w \rangle_H, \tag{10.1.2}$$

$$\langle z, w\lambda \rangle_H = z^*Hw\lambda = \langle z, w \rangle_H\lambda, \tag{10.1.3}$$

$$\langle z, w \rangle_H = z^*Hw = z^*H^*w = (w^*Hz)^* = \overline{\langle w, z \rangle}_H. \tag{10.1.4}$$

The latter property leads to another observation: for every $z \in \mathbb{K}^n$ we have $\langle z, z \rangle_H \in \mathbb{R}$.

Let $\langle \cdot, \cdot \rangle_H$ be a Hermitian form associated to some Hermitian matrix H. Recalling that the eigenvalues of H are real, we say that

- $\langle \cdot, \cdot \rangle_H$ is *non-degenerate* if all the eigenvalues of H are non-zero;
- $\langle \cdot, \cdot \rangle_H$ is *positive definite* if all the eigenvalues of H are strictly positive;
- $\langle \cdot, \cdot \rangle_H$ is *negative definite* if all the eigenvalues of H are strictly negative;
- $\langle \cdot, \cdot \rangle_H$ is *indefinite* if some eigenvalues of H are positive and some negative.

We say that $\langle \cdot, \cdot \rangle_H$ has *signature* (p, q), if H has p strictly positive eigenvalues and q strictly negative eigenvalues, counted with multiplicity. We write $\mathbb{K}^{p,q}$ for $\mathbb{K}^{p+q}$ equipped with a non-degenerate Hermitian form of signature (p, q).

10.1.3 Hermitian forms of signature $(n,1)$

There are a lot of models for the $\mathbb{K}$-hyperbolic space $\mathbb{K}\mathbf{H}^n$. In this work we will focus on one particular models. Let $\langle x, y\rangle$ be the Hermitian form of signature $(n,1)$ on the space $\mathbb{K}^{n+1}$, given by

$$\langle x, y\rangle := -\overline{x}_1 y_{n+1} - \overline{x}_{n+1} y_1 + \sum_{\lambda=2}^{n} \overline{x}_\lambda y_\lambda, \tag{10.1.5}$$

where $x = (x_1, \ldots, x_{n+1})$ and $y = (y_1, \ldots, y_{n+1})$. We observe that this form is associated to the matrix

$$K := \begin{pmatrix} 0 & 0 & -1 \\ 0 & I_{n-1} & 0 \\ -1 & 0 & 0 \end{pmatrix}. \tag{10.1.6}$$

The matrix K is a Hermitian matrix that has n positive eigenvalues, and 1 negative eigenvalue. There are many more Hermitian forms that one could consider, which all lead to the construction of what is the same space up to isometries. The reason why we chose K as in (10.1.6) is to express some isometries of the hyperbolic space in a way that suits our needs.

Note that, thanks to the property (10.1.4) we have $\langle x, x\rangle = \overline{\langle x, x\rangle}$, i.e, $\langle x, x\rangle \in \mathbb{R}$. The *orthogonal complement* of $x \in \mathbb{K}^{n,1}$, denote by $x^\perp$, is $\{u \in \mathbb{K}^{n+1} | \langle x, u\rangle = 0\}$.

Lemma 10.1.7 (Reverse Schwartz Inequality). *If $\langle x, x\rangle < 0$ and $\langle y, y\rangle < 0$, then*

$$\langle x, y\rangle\langle y, x\rangle \geq \langle x, x\rangle\langle y, y\rangle$$

with equality if and only if x and y are linearly dependent over $\mathbb{K}$.

Proof. If x and y are linearly dependent then there exist $\lambda \in \mathbb{K}$ such that $x = y\lambda$. We rewrite the inequality as follow

$$\langle y\lambda, y\rangle\langle y, y\lambda\rangle \geq \langle y\lambda, y\lambda\rangle\langle y, y\rangle.$$

The LHS, thanks to the properties (10.1.2) and (10.1.3) of the Hermitian forms, is equivalent to

$$\overline{\lambda}\langle y, y\rangle\langle y, y\rangle\lambda = |\lambda|^2\langle y, y\rangle^2.$$

The RHS of the inequality consist of, thanks to the same properties,

$$\langle y\lambda, y\lambda\rangle\langle y, y\rangle = |\lambda|^2\langle y, y\rangle^2,$$

where λ commute with $\langle y, y\rangle$ due the fact that $\langle y, y\rangle$ is real. This prove the linearly dependent case.

Suppose now that x and y are linearly independent. The restriction of $\langle\cdot,\cdot\rangle$ to $x^\perp$ is positive definite and since $\langle y,y\rangle < 0$ we have $\langle x,y\rangle \neq 0$. Let $\lambda = -\langle x,x\rangle\langle x,y\rangle^{-1}$, then $x + y\lambda \in x^\perp$. Due the fact that x and y are linearly independent we have $x + y\lambda \neq 0$, therefore $\langle x+y\lambda, x+y\lambda\rangle = \langle x+y\lambda, y\lambda\rangle > 0$. By expanding this inequality we get

$$-\langle x,x\rangle + \langle x,x\rangle^2\langle y,y\rangle\langle y,x\rangle^{-1}\langle x,y\rangle^{-1} > 0.$$

After diving by $\langle x,x\rangle < 0$, this can be rearrange to give the inequality

$$\langle x,y\rangle\langle y,x\rangle \geq \langle x,x\rangle\langle y,y\rangle,$$

thus completing the proof. □

10.2 The $\mathbb{K}$-hyperbolic n-space $\mathbb{K}\mathbf{H}^n$

10.2.1 Definition and properties

Let $\mathbb{K}^{n,1}$ be equipped with the Hermitian form $\langle\cdot,\cdot\rangle$ defined in (10.1.5). We defined the *n-dimensional $\mathbb{K}$-projective space* $\mathbb{K}\mathbf{P}^n$, as the quotient of $\mathbb{K}^{n+1}\setminus\{0\}$ by the equivalence relation that identifies $x = (x_1,\dots,x_{n+1})$ with $x\lambda = (x_1\lambda,\dots,x_{n+1}\lambda)$ for all $\lambda \in \mathbb{K}\setminus\{0\}$. The class of x is denoted by $[x]$ and $[x_1,\dots,x_{n+1}]$ are called *homogeneous coordinates* for $[x]$. We finally give the definition of $\mathbb{K}\mathbf{H}^n$.

Definition 10.2.1. We define the $\mathbb{K}$-*hyperbolic n-space* as the set

$$\mathbb{K}\mathbf{H}^n := \{[x] \in \mathbb{K}\mathbf{H}^n : \langle x,x\rangle < 0\}$$

equipped with the distance d such that

$$\cosh^2 d([x],[y]) = \frac{\langle x,y\rangle\langle y,x\rangle}{\langle x,x\rangle\langle y,y\rangle}. \tag{10.2.2}$$

For the formula of the distance we give as reference [BH99, Part II, Chapter 10].

Namely, a point $[x]$ of the n-dimensional $\mathbb{K}$-projective space is in $\mathbb{K}\mathbf{H}^n$ if and only if

$$-\overline{x}_1x_{n+1} - \overline{x}_{n+1}x_1 + \sum_{\lambda=2}^{n}|x_\lambda|^2 < 0.$$

First of all we want to check that $\mathbb{K}\mathbf{H}^n$ is well defined, and that the distance formula does not depend on the representative chosen.

Proposition 10.2.3. *Let $x \in \mathbb{K}^{n,1}$, if $\langle x, x\rangle < 0$ then $\langle x\lambda, x\lambda\rangle < 0$ for every $\lambda \in \mathbb{K} \setminus \{0\}$. Furthermore, for every $[x], [y] \in \mathbb{K}\mathbf{H}^n$ the right hand side of* (10.2.2) *is bigger than 1 and for every $\lambda_1, \lambda_2 \in \mathbb{K} \setminus \{0\}$ is true that*

$$\frac{\langle x, y\rangle\langle y, x\rangle}{\langle x, x\rangle\langle y, y\rangle} = \frac{\langle x\lambda_1, y\lambda_2\rangle\langle y\lambda_2, x\lambda_1\rangle}{\langle x\lambda_1, x\lambda_1\rangle\langle y\lambda_2, y\lambda_2\rangle}.$$

Proof. Thanks to the properties (10.1.2) and (10.1.3) of Hermitian forms, we can write

$$\langle x\lambda, x\lambda\rangle = \overline{\lambda}\langle x, x\rangle\lambda = |\lambda|^2\langle x, x\rangle < 0.$$

For the second point, thanks to Lemma 10.1.7 we know that

$$\frac{\langle x, y\rangle\langle y, x\rangle}{\langle x, x\rangle\langle y, y\rangle} \geq 1.$$

Recalling the properties (10.1.4), (10.1.3) and (10.1.2) of the Hermitian form we obtain that

$$\langle x\lambda_1, y\lambda_2\rangle\langle y\lambda_2, x\lambda_1\rangle = |\langle x\lambda_1, y\lambda_2\rangle|^2 = |\lambda_1|^2|\langle x, y\rangle|^2|\lambda_2|^2,$$

and

$$\langle x\lambda_1, x\lambda_1\rangle\langle y\lambda_2, y\lambda_2\rangle = |\lambda_1|^2\langle x, x\rangle\langle y, y\rangle|\lambda_2|^2.$$

Therefore

$$\frac{\langle x\lambda_1, y\lambda_2\rangle\langle y\lambda_2, x\lambda_1\rangle}{\langle x\lambda_1, x\lambda_1\rangle\langle y\lambda_2, y\lambda_2\rangle} = \frac{|\lambda_1|^2|\langle x, y\rangle|^2|\lambda_2|^2}{|\lambda_1|^2\langle x, x\rangle\langle y, y\rangle|\lambda_2|^2} = \frac{\langle x, y\rangle\langle y, x\rangle}{\langle x, x\rangle\langle y, y\rangle}.$$

□

10.3 The $\mathbb{K}$-Heisenberg groups

Let $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}, \mathbb{H}\}$. We recall the definition of imaginary part: if $u \in \mathbb{R}$ or $u \in \mathbb{C}$ then $\mathrm{Im}(u) = \frac{u-\overline{u}}{2}$, while if $u \in \mathbb{H}$ then $u = a + bi + cj + dk$, with suitable $a, b, c, d \in \mathbb{R}$, and $\mathrm{Im}(u) = \begin{pmatrix} b \\ c \\ d \end{pmatrix} \in \mathbb{R}^3$.

Definition 10.3.1. The *n-th* $\mathbb{K}$*-Heisenberg group* $\mathbb{K}\mathcal{H}^n$, with $n \geq 1$, is the set

$$\mathbb{K}^{n-1} \times \mathrm{Im}(\mathbb{K}) = \{(u, s) | u \in \mathbb{K}^{n-1}, s \in \mathrm{Im}(\mathbb{K})\}$$

endowed with the multiplication law

$$(u, s)(v, t) = (u + v, s + t + \mathrm{Im}(u^t\overline{v})). \tag{10.3.2}$$

Proposition 10.3.3. *The set $\mathbb{K}^{n-1} \times \mathrm{Im}(\mathbb{K})$ is a group with the multiplication law given by* (10.3.2).

Proof. The set is clearly closed under such an operation. The identity element is given by $(0,0)$, as a matter of fact, for all $u \in \mathbb{K}^{n-1}$ and all $s \in \mathrm{Im}(\mathbb{K})$

$$(0,0)(u,s) = (0+u, 0+s+\mathrm{Im}(0\overline{u})) = (u,s),$$
$$(u,s)(0,0) = (u+0, s+0+\mathrm{Im}(u0)) = (u,s).$$

The inverse element $(u,s)^{-1}$ is given by $(-u,-s)$:

$$\begin{aligned}(u,s)(-u,-s) &= (u-u, s-s+\mathrm{Im}(-|u|^2)) = (0,0),\\ (-u,-s)(u,s) &= (-u+u, -s+s+\mathrm{Im}(-|u|^2)) = (0,0).\end{aligned}$$

And finally the associativity is given by

$$\begin{aligned}((u,s)(v,t))(w,r) &= (u+v, s+t+\mathrm{Im}(u^t\overline{v}))(w,r)\\ &= (u+v+w, s+t+\mathrm{Im}(u^t\overline{v})+r+\mathrm{Im}((u+v)^t\overline{w}))\\ &= (u+v+w, s+t+r+\mathrm{Im}(u^t\overline{v}+u^t\overline{w}+v^t\overline{w}))\\ &= (u+v+w, s+t+r+\mathrm{Im}(z^t(\overline{v+w}))+\mathrm{Im}(v^t\overline{w}))\\ &= (u,s)(v+w, t+r+\mathrm{Im}(v^t\overline{w})))\\ &= (u,s)((v,t)(w,r)).\end{aligned}$$

□

We define an *Heisenberg homothety of ratio a* on the group $\mathbb{K}\mathcal{H}^n$, by

$$\delta_a(u,s) = (au, a^2 s) \qquad \forall a \in \mathbb{R}\setminus\{0\}. \tag{10.3.4}$$

Proposition 10.3.5. *The Heisenberg homothety satisfies the following properties:*

1. $\delta_a((u,s)(v,t)) = (\delta_a(u,s))(\delta_a(v,t))$ *for all* $a \in \mathbb{R}\setminus\{0\}$*;*

2. $\delta_a^{-1} = \delta_{a^{-1}}$ *for all* $a \in \mathbb{R}\setminus\{0\}$*.*

Proof. All the equalities are simple application of the definition of δ_a or the multiplication law defined in (10.3.2). Regarding the first claim, we have:

$$\begin{aligned}\delta_a((u,s)(v,t)) &= \delta_a(u+v, s+t+\mathrm{Im}(u^t\overline{v}))\\ &= (au+av, a^2s+a^2t+a^2\mathrm{Im}(u^t\overline{v}))\\ &= (au+av, a^2s+a^2t+\mathrm{Im}(au^t\overline{av}))\\ &= (au, a^2s)(av, a^2t)\\ &= (\delta_a(u,s))(\delta_a(v,t)).\end{aligned}$$

Regarding the second claim, we have:

$$\begin{aligned}&\delta_a\delta_{a^{-1}}(u,s) = \delta_a(a^{-1}u, a^{-2}s) = (u,s),\\ &\delta_{a^{-1}}\delta_a(u,s) = \delta_{a^{-1}}(au, a^2s) = (u,s).\end{aligned}$$

□

We point out that if $\mathbb{K}=\mathbb{R}$ then $\mathbb{R}\mathcal{H}^n \cong \mathbb{R}^{n-1}$ as group, where $\mathbb{R}^{n-1}$ has the standard abelian group structure. As a matter of fact, following Definition 10.3.1, the n-th $\mathbb{R}$-Heisenberg group is the set $\mathbb{R}^{n-1}\times\{0\}$ endowed with the multiplication law $(u,0)(v,0)=(u+v,0)$ for all $u,v\in\mathbb{R}^{n-1}$. Eliminating the last coordinate, which is always 0, we obtain the group isomorphism.

10.4 Isometries of hyperbolic spaces

We start by recalling how we defined the Hermitian form $\langle\cdot,\cdot\rangle$ in $\mathbb{K}^{n+1}$. This Hermitian form is the form associated to the Hermitian matrix K, given by

$$K := \begin{pmatrix} 0 & 0 & -1 \\ 0 & I_{n-1} & 0 \\ -1 & 0 & 0 \end{pmatrix}, \tag{10.4.1}$$

so

$$\langle x,y\rangle := \langle x,y\rangle_K = x^*Ky.$$

Consider the group $\mathrm{GL}(n+1,\mathbb{K})$ that is the group of invertible $(n+1,n+1)$ matrices with coefficients in $\mathbb{K}$. There is a natural left action of $\mathrm{GL}(n+1,\mathbb{K})$ on $\mathbb{K}^{n,1}$ by $\mathbb{K}$-linear automorphism: the matrix $A=(a_{ij})$ sends $x=(x_1,\dots,x_{n+1})\in\mathbb{K}^{n,1}$ to

$$Ax := \begin{pmatrix} \sum_{j=1}^{n+1} a_{1j}x_j \\ \vdots \\ \sum_{j=1}^{n+1} a_{n+1j}x_j \end{pmatrix}. \tag{10.4.2}$$

Definition 10.4.3. Let $O_{\mathbb{K}}(n,1)$ denote the subgroup of $\mathrm{GL}(n+1,\mathbb{K})$ that preserves the form $\langle\cdot,\cdot\rangle$ induced by *(10.4.1)*, that is

$$O_{\mathbb{K}}(n,1) := \{A \in \mathrm{GL}(n+1,\mathbb{K}) \,:\, \langle Ax, Ay\rangle = \langle x, y\rangle \;\forall x, y \in \mathbb{K}^{n,1}\}.$$

We start by characterizing the elements of $O_{\mathbb{K}}(n,1)$.

Proposition 10.4.4. *$A \in O_{\mathbb{K}}(n,1) \Leftrightarrow A^*KA = K$, for the K in* (10.4.1)*.*

Proof. We firstly prove the left implication. Let $A \in O_{\mathbb{K}}(n,1)$. For every $x, y \in \mathbb{K}^{n,1}$ we have $\langle Ax, Ay\rangle = \langle x, y\rangle$. Therefore $\langle Ax, Ax\rangle = \langle x, x\rangle$ for all $x \in \mathbb{K}^{n,1}$, which can be written as

$$(Ax)^*KAx = x^*Kx \qquad \forall x \in \mathbb{K}^{n,1}.$$

This is equivalent to

$$x^*A^*KAx = x^*Kx \qquad \forall x \in \mathbb{K}^{n,1}.$$

By choosing x as the elements of the canonical base we can conclude that $A^*KA = K$.

To prove the other implication, let $A \in \mathrm{GL}(n+1,\mathbb{K})$ such that $A^*KA = K$ then

$$\langle Ax, Ay\rangle = (Ax)^*KAy = x^*A^*KAy = x^*Ky = \langle x, y\rangle. \qquad \square$$

Proposition 10.4.5. *The set $O_{\mathbb{K}}(n,1)$ with the matrix multiplication is a group.*

Proof. Let $A, B \in O_{\mathbb{K}}(n,1)$ then

$$(AB)^*KAB = B^*A^*KAB = B^*KB = K,$$

and

$$(A^{-1})^*KA^{-1} = (A^{-1})^*A^*KAA^{-1} = (AA^{-1})^*KAA^{-1} = K,$$

so $AB, A^{-1} \in O_{\mathbb{K}}(n,1)$ thanks to Proposition 10.4.4. The identity matrix obviously belong to $O_{\mathbb{K}}(n,1)$. The properties of the multiplication follow from the properties of the classic row-column multiplication between matrices. $\square$

We now note that there is an induced action of $\mathrm{GL}(n+1,\mathbb{K})$ on $\mathbb{K}\mathbf{P}^n$, and this is the action we shall focus on.

Lemma 10.4.6. *The induced action of $\mathrm{GL}(n+1,\mathbb{K})$ on $\mathbb{K}\mathbf{P}^n$ given by $A[x] = [Ax]$, for all $A \in \mathrm{GL}(n+1,\mathbb{K})$ is well defined.*

Proof. We need to prove that the induced action does not depend on a representative, that is $[A(x\lambda)] = [Ax]$ for every $A \in \mathrm{GL}(n+1,\mathbb{K})$, for every $[x] \in \mathbb{K}\mathbf{P}^n$ and for every $\lambda \in \mathbb{K} \setminus \{0\}$. Let $[x] = [x_1, \cdots, x_{n+1}]^t$ and let $A = (a_{ij})$ with $a_{ij} \in \mathbb{K}$ for all $i, j \in \{1, \cdots, n+1\}$. The fact follows from:

$$
\begin{aligned}
[A(x\lambda)] &= \left[A \begin{pmatrix} x_1\lambda \\ \vdots \\ x_{n+1}\lambda \end{pmatrix} \right] \\
&= \left[\begin{pmatrix} \sum_{j=1}^{n+1} a_{1j}x_j\lambda \\ \vdots \\ \sum_{j=1}^{n+1} a_{n+1j}x_j\lambda \end{pmatrix} \right] \\
&= \left[\begin{pmatrix} \sum_{j=1}^{n+1} a_{1j}x_j \\ \vdots \\ \sum_{j=1}^{n+1} a_{n+1j}x_j \end{pmatrix} \lambda \right] = [(Ax)\lambda] = [Ax],
\end{aligned}
$$

where the first equality follows from the hypothesis on $[x]$; the second, the third and the forth one follows from (10.4.2); and the last one holds thanks to the definition of $\mathbb{K}\mathbf{P}^{n+1}$. $\square$

Proposition 10.4.7. *The induced action of $O_{\mathbb{K}}(n,1)$ on $\mathbb{K}\mathbf{P}^n$ preserves the subset $\mathbb{K}\mathbf{H}^n$ and act by isometries on $\mathbb{K}\mathbf{H}^n$*

Proof. Let $[x] \in \mathbb{K}\mathbf{H}^n$ and $A \in O_{\mathbb{K}}(n,1)$, by definition A preserves the form $\langle \cdot, \cdot \rangle$ and therefore $\langle Ax, Ax \rangle = \langle x, x \rangle < 0$ so $[Ax] \in \mathbb{K}\mathbf{H}^n$. The fact that this action is by isometries follows directly by the definition (10.2.1) of the distance. $\square$

We shall focus on two particular subgroups of $O_{\mathbb{K}}(n,1)$:

Definition 10.4.8. We denote by A and N the following subsets of the group $O_{\mathbb{K}}(n,1)$:

- A denotes the 1-parameter set, formed by the elements
$$A(a) := \begin{pmatrix} e^a & 0 & 0 \\ 0 & I_{n-1} & 0 \\ 0 & 0 & e^{-a} \end{pmatrix}, \quad a \in \mathbb{R};$$
- N denotes the set of matrices of the form
$$\nu(M, M_{13}) := \begin{pmatrix} 1 & M & M_{13} \\ 0 & I_{n-1} & M^* \\ 0 & 0 & 1 \end{pmatrix}, \tag{10.4.9}$$
where M is a $(1, n-1)$-matrix with elements in $\mathbb{K}$ and M_{13} is in $\mathbb{K}$ and satisfies
$$|M|^2 = M_{13} + \overline{M}_{13}.$$

The following is a simple lemma that characterizes how the product works between elements of A and N.

Lemma 10.4.10. *For all $t \in \mathbb{R}$ and all $\nu(M, M_{13}) \in N$, we have*

$$\nu(M, M_{13})A(t) = A(t)\nu(e^{-t}M, e^{-2t}M_{13}),$$

and

$$A(t)\nu(M, M_{13}) = \nu(e^{t}M, e^{2t}M_{13})A(t).$$

Proof. Let $A(t) \in A$ and $\nu(M, M_{13}) \in N$, we compute the matrix product $\nu(M, M_{13})A(t)$.

$$\begin{aligned}\nu(M, M_{13})A(t) &= \begin{pmatrix}1 & M & M_{13}\\ 0 & I_{n-1} & M^*\\ 0 & 0 & 1\end{pmatrix}\begin{pmatrix}e^t & 0 & 0\\ 0 & I_{n-1} & 0\\ 0 & 0 & e^{-t}\end{pmatrix}\\ &= \begin{pmatrix}e^t & M & e^{-t}M_{13}\\ 0 & I_{n-1} & e^{-t}M^*\\ 0 & 0 & e^{-t}\end{pmatrix}\\ &= \begin{pmatrix}e^t & 0 & 0\\ 0 & I_{n-1} & 0\\ 0 & 0 & e^{-t}\end{pmatrix}\begin{pmatrix}1 & e^{-t}M & e^{-2t}M_{13}\\ 0 & I_{n-1} & e^{-t}M^*\\ 0 & 0 & 1\end{pmatrix}\\ &= A(t)\nu(e^{-t}M, e^{-2t}M_{13}).\end{aligned}$$

The other case follows from a change of variables: $N = e^{-t}M$ $N_{13} = e^{-2t}M_{13}$. Noting that both $\nu(e^{-t}M, e^{-2t}M_{13})$ and $\nu(e^{t}M, e^{2t}M_{13})$ satisfies the condition of being in N ends the proof. □

Theorem 10.4.11.

1. *A and N are subgroups of $O_{\mathbb{K}}(n, 1)$;*

2. *NA is a subgroup of $O_{\mathbb{K}}(n, 1)$;*

3. *N is normal in NA;*

4. *The group A is isomorphic to $\mathbb{R}$.*

Proof.

1. Firstly we prove that A and N are subset of $O_{\mathbb{K}}(n, 1)$. Thanks to Proposition 10.4.4 we only have to prove that given $\nu(M, M_{13}) \in N$ and $t \in \mathbb{R}$ is true that $A(t)^*KA(t) = K$ and

$\nu(M, M_{13})^* K \nu(M, M_{13}) = K$. For the first case we have:

$$\begin{aligned}A(t)^* K A(t) &= \begin{pmatrix} e^t & 0 & 0 \\ 0 & I_{n-1} & 0 \\ 0 & 0 & e^{-t} \end{pmatrix} \begin{pmatrix} 0 & 0 & -1 \\ 0 & I_{n-1} & 0 \\ -1 & 0 & 0 \end{pmatrix} \begin{pmatrix} e^t & 0 & 0 \\ 0 & I_{n-1} & 0 \\ 0 & 0 & e^{-t} \end{pmatrix} \\ &= \begin{pmatrix} 0 & 0 & -e^t \\ 0 & I_{n-1} & 0 \\ -e^{-t} & 0 & 0 \end{pmatrix} \begin{pmatrix} e^t & 0 & 0 \\ 0 & I_{n-1} & 0 \\ 0 & 0 & e^{-t} \end{pmatrix} \\ &= \begin{pmatrix} 0 & 0 & -1 \\ 0 & I_{n-1} & 0 \\ -1 & 0 & 0 \end{pmatrix} = K.\end{aligned}$$

For the second case we have:

$$\begin{aligned}&\nu(M, M_{13})^* K \nu(M, M_{13}) \\ &\quad= \begin{pmatrix} 1 & 0 & 0 \\ M^* & I_{n-1} & 0 \\ \overline{M}_{13} & M^* & 1 \end{pmatrix} \begin{pmatrix} 0 & 0 & -1 \\ 0 & I_{n-1} & 0 \\ -1 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & M & M_{13} \\ 0 & I_{n-1} & M^* \\ 0 & 0 & 1 \end{pmatrix} \\ &\quad= \begin{pmatrix} 0 & 0 & -1 \\ 0 & I_{n-1} & -M_{12}^* \\ -1 & M_{23}^* & -\overline{M}_{13} \end{pmatrix} \begin{pmatrix} 1 & M & M_{13} \\ 0 & I_{n-1} & M^* \\ 0 & 0 & 1 \end{pmatrix} \\ &\quad= \begin{pmatrix} 0 & 0 & -1 \\ 0 & I_{n-1} & M_{23} - M_{12}^* \\ -1 & -M_{12} + M_{23}^* & -M_{13} + |M_{23}|^2 - \overline{M}_{13} \end{pmatrix}.\end{aligned}$$

Thanks to the definition of N, it is true that

$$M_{23} = M_{12}^*,$$

and

$$|M_{23}|^2 = \overline{M}_{13} + M_{13},$$

so the last matrix is equal to K.

Now let $A(t), A(s) \in A$, a simple check shows that $A(t)A(s) = A(s+t)$ and $A(t)^{-1} = A(-t)$. To prove A is a subgroup of $O_{\mathbb{K}}(n,1)$ we only need to show that $A(t)A(s)^{-1} \in A$ for all $t, s \in \mathbb{R}$:

$$A(t)A(s)^{-1} = A(t)A(-s) = A(t-s) \in A.$$

In a similar way let $\nu(M, M_{13}), \nu(N, N_{13}) \in N$, a simple check shows that

$$\nu(M, M_{13})\nu(N, N_{13}) = \nu(M + N, N_{13} + MN^* + M_{13})$$

and

$$\nu(M, M_{13})^{-1} = \nu(-M, \overline{M}_{13}).$$

As above we next prove that given $\nu(M, M_{13}), \nu(N, N_{13}) \in N$ the product $\nu(M, M_{13})\nu(N, N_{13})^{-1}$ belongs to N:

$$\begin{aligned}
\nu(M, M_{13})\nu(N, N_{12})^{-1} &= \nu(M, M_{13})\nu(-N, \overline{N}_{13}) \\
&= \begin{pmatrix} 1 & M & M_{13} \\ 0 & I_{n-1} & M^* \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & -N & \overline{N}_{13} \\ 0 & I_{n-1} & -N^* \\ 0 & 0 & 1 \end{pmatrix} \\
&= \begin{pmatrix} 1 & M-N & \overline{N}_{13} - MN^* + M_{13} \\ 0 & I_{n-1} & M^* - N^* \\ 0 & 0 & 1 \end{pmatrix} \\
&= \nu(M - N, \overline{N}_{13} - MN^* + M_{13}).
\end{aligned}$$

The fact that $M - N$ and $\overline{N}_{13} - MN^* + M_{13}$ satisfy the needed condition ends the proof that both A and N are subgroups of $O_{\mathbb{K}}$.

2. Let $\nu(M, M_{13})A(t), \nu(N, N_{13})A(s) \in NA$. Thanks to Lemma 10.4.10

$$\begin{aligned}
\nu(M, M_{13})A(t)\nu(N, N_{13})A(s) &= \nu(M, M_{13})\nu(e^t N, e^{2t} N_{13})A(-t)A(s) \qquad (10.4.12) \\
&= \nu(M + e^t N, e^{2t}\overline{N}_{13} + e^t MN^* + M_{13})A(s-t).
\end{aligned}$$

We observe that $A(s-t) \in A$ and $\nu(M + e^t N, e^{2t}\overline{N}_{13} + e^t MN^* + M_{13}) \in N$ thus proving the closure of NA. The identity matrix $I = A(0)\nu(0, 0)$, where the zeros refers accordingly, is the neutral element. Given $\nu(M, M_{13})A(t) \in NA$, the calculation (10.4.12) shows also that the inverse element $(\nu(M, M_{13})A(t))^{-1}$ is $\nu(-e^{-t}M, -e^{2t}M_{13})A(-t)$.

3. Lemma 10.4.10 also let us prove this point, as a matter of fact

$$\begin{aligned}
\nu(M, M_{13})A(t)\nu(N, N_{13})(\nu(M, M_{13})A(t))^{-1} &= \nu(M, M_{13})A(t)\nu(N, N_{13})\nu(M_{-1}, (M_{-1})_{13})A(-t) \\
&= \nu(M, M_{13})\nu(P, P_{13})A(t)A(-t) \\
&= \nu(M, M_{13})\nu(P, P_{13}) \in N.
\end{aligned}$$

where $\nu(M_{-1}, (M_{-1})_{13})$ denotes the inverse of $\nu(M, M_{13})$ and $\nu(P, P_{13})$ is obtain from Lemma 10.4.10.

4. Let $\phi_A : A \to \mathbb{R}$ defined as follow

$$\phi_A : A \to \mathbb{R}$$

$$\phi_A(A(a)) \mapsto a.$$

Obviously $\phi_A(I) = 0$ and

$$\phi_A(A(a)A(b)) = \phi_A(A(ab)) = ab = \phi_A(A(a))\phi_A(A(b)).$$

The definition of A ensures that the homomorphism is bijective. The inverse homomorphism is given by

$$\begin{aligned} \phi_A^{-1} : \mathbb{R} &\to A \\ a &\mapsto A(a) \end{aligned}$$

□

10.5 Hyperbolic spaces as semidirect products

In this section we prove the following result:

Theorem 10.5.1. *For every $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}, \mathbb{H}\}$ and every $n \in \mathbb{N} \setminus \{0\}$, the $\mathbb{K}$-hyperbolic n-space $\mathbb{K}\mathbf{H}^n$ is isometric to the manifold $\mathbb{K}^{n-1} \times \mathrm{Im}(\mathbb{K}) \times \mathbb{R}$ equipped with the multiplication law given by*

$$(u, s; a) \cdot (v, t; b) = (u + e^a v, s + e^{2a} t + \mathrm{Im}(u^t e^a \overline{v}); a + b) \tag{10.5.2}$$

and the left-invariant distance d such that

$$4\cosh^2 d(\mathbf{0}, (v, t; b)) = 4\cosh^2(b) + 2e^{-b}\cosh(b)|v|^2 + e^{-2b}\left(\frac{|v|^4}{4} + |t|^2\right).$$

In order to prove this theorem we will discuss the real, complex and quaternion case separately. All three case follows the same structure of proof. First we characterize the groups N and A, previously discussed in Chapter 5. Then we prove that NA acts simply transitively on $\mathbb{K}\mathbf{H}^n$ thus obtaining the wanted identification. Then we express the distance on the group NA and check that it is left-invariant. While in the case of the Real Hyperbolic space, the proofs undergo some simplification, in the complex case and the quaternionic case the proofs are essentially the same. So in this presentation we will only deal with the real case and the quaternionic case, in which one must be a little more careful, and the complex case follows analogously.

10.5.1 The real case

Our aim is to prove that $\mathbb{R}\mathbf{H}^n$, with $n \geq 1$, admits a group structure for which the distance is left-invariant and such structure is $\mathbb{R}^{n-1} \rtimes \mathbb{R}$, where $\mathbb{R}$ acts on $\mathbb{R}^{n-1}$ by standard dilations. In this section we work with the group A and N as in Definition 10.4.8. We already know that A is isomorphic to $\mathbb{R}$ as proved in Lemma 10.4.11. Next we better characterize the group structure of $N = \{\nu(M, M_{13}) : M \in \mathfrak{M}(1, n-1, \mathbb{K}), M_{13} \in \mathbb{K}, |M|^2 = M_{13} + \overline{M}_{13}\}$, where ν is defined in (10.4.9).

Lemma 10.5.3. *When $\mathbb{K} = \mathbb{R}$ the group N is isomorphic to $\mathbb{R}^{n-1}$.*

Proof. Let $u \in \mathbb{R}^{n-1}$ and let $u := M^t$. From the relations in the definition of N, recalled above, follows that $M_{13} = \frac{|u|^2}{2}$, therefore we can write

$$N = \{h(u) : u \in \mathbb{R}^{n-1}\},$$

where

$$h(u) := \nu\left(u^t, \frac{|u|^2}{2}\right) = \begin{pmatrix} 1 & u^t & \frac{|u|^2}{2} \\ 0 & I_{n-1} & \overline{u} \\ 0 & 0 & 1 \end{pmatrix}. \tag{10.5.4}$$

We claim that the mapping $h : \mathbb{R}^{n-1} \to N$ is a group isomorphism. The injectivity follows from the fact that, if $u, v \in \mathbb{R}^{n-1}$ and $u \neq v$, then the first row of $h(u)$ differs from the first row of $h(v)$. The surjectivity follows from the observation at the beginning of the proof. The only thing left is prove is that $h(u)h(v) = h(u + v)$, which follows from the simple observation that $|u + v|^2 = |u|^2 + |v|^2 + 2u^t\overline{v}$. □

For the sequel we shall need the following identities.

Lemma 10.5.5. *Let $u, v \in \mathbb{R}^{n-1}$ and let $a, b \in \mathbb{R}$ then*

$$A(a)h(v) = h(e^a v)A(a),$$

$$h(u)A(a)h(v)A(b) = h(u + e^a v)A(a + b),$$

and

$$(h(u)A(a))^{-1} = h(-e^{-a}u)A(-a).$$

Proof. The first equality follows from Lemma 10.4.10 and Lemma 10.5.3. The second one follows from the calculation

$$h(u)A(a)h(v)A(b) = h(u)h(e^a v)A(a)A(b) = h(u + e^a v)A(a + b).$$

The last one follows from the properties of the matrix inverse and the first equality proved in this lemma. □

Theorem 10.5.6. *The group NA acts simply transitively on $\mathbb{R}\mathbf{H}^n$, that is, for every $x, y \in \mathbb{R}\mathbf{H}^n$ there exists a unique $g \in NA$ such that $g \cdot x = y$.*

Proof. We consider the point $o := [1, 0, \dots, 0, 1]^t \in \mathbb{R}\mathbf{H}^n$ and we want to show that given an arbitrary point $x \in \mathbb{R}\mathbf{H}^n$ with homogeneous coordinates $[x_1, \dots, x_n, 1]$, there exists a unique $(u; a) \in \mathbb{R}^{n-1} \times \mathbb{R}$ such that x is the image under $h(u)A(a)$ of the point o, where h is defined in (10.5.4) and A in Definition 10.4.8. We first compute $h(u)A(a)$

$$h(u)A(a) = \begin{pmatrix} e^a & u^t & e^{-a}\frac{|u|^2}{2} \\ 0 & I_{n-1} & e^{-a}\overline{u} \\ 0 & 0 & e^{-a} \end{pmatrix}.$$

Therefore

$$h(u)A(a) \cdot o = \begin{bmatrix} e^a + e^{-a}\frac{|u|^2}{2} \\ e^{-a}\overline{u} \\ e^{-a} \end{bmatrix} = \begin{bmatrix} e^{2a} + \frac{|u|^2}{2} \\ \overline{u} \\ 1 \end{bmatrix}.$$

So $h(u)A(a) \cdot o = x$ becomes the system of equations:

$$\begin{cases} x_1 = e^{2a} + \frac{|u|^2}{2} \\ X_2 = \overline{u} \end{cases},$$

where $X_2^t = (x_2, \dots, x_n)$. Thus X_2 uniquely determines $\overline{u}$ hence u. We know that if $[x] \in \mathbb{R}\mathbf{H}^n$ x must satisfies $\langle x, x\rangle < 0$, that is

$$-\overline{x}_1 x_{n+1} - \overline{x}_{n+1} x_1 + \sum_{i=2}^{n} \overline{x}_i x_i < 0,$$

which, in our case, is equivalent to

$$-x_1 - x_1 + \sum_{i=2}^{n} |x_i|^2 = |X_2|^2 - 2x_1 < 0 \Leftrightarrow |X_2|^2 < 2x_1.$$

We can conclude that

$$x_1 > \frac{|X_2|^2}{2} = \frac{|u|^2}{2}.$$

There therefore exists a unique $a \in \mathbb{R}$ such that $x_1 = e^{2a} + \frac{|u|^2}{2}$. Thus the above system of equation has a unique solution $(u; a) \in \mathbb{R}^{n-1} \times \mathbb{R}$.

Let now $x \in \mathbb{R}\mathbf{H}^n$ be an arbitrary element, we want to prove that given $y \in \mathbb{R}\mathbf{H}^n$ there exists a unique $(u; a) \in \mathbb{R}^{n-1} \times \mathbb{R}$ such that $h(u)A(a)x = y$. We now know that there exists a unique g_y and a unique $g_x \in NA$ such that $x = g_x \cdot o$ and $y = g_y \cdot o$. We claim that the wanted element is $g_y g_x^{-1}$. As a matter of fact

$$g_y g_x^{-1} \cdot x = g_y g_x^{-1} g_x \cdot o = g_y \cdot o = y.$$

Let $g_x = h(v)A(b)$ and $g_y = h(w)A(c)$. Then

$$\begin{aligned} g_y g_x^{-1} &= h(w)A(c)(h(v)A(b))^{-1} \\ &= h(w)A(c)h(-e^{-b}v)A(-b) \\ &= h(w - e^{c-b}v)A(c-b), \end{aligned}$$

where the equalities hold thanks to Lemma 10.5.5. From this calculation follow that $u = w - e^{c-b}v$ and $a = c - b$, thus the proof is completed. $\square$

From Theorem 10.5.6 we get the following consequence.

Corollary 10.5.7. *The mapping $(u, a) \mapsto h(u)A(a)\cdot o$ gives a smooth identification between $\mathbb{R}^{n-1}\times\mathbb{R}$ and $\mathbb{R}\mathbf{H}^n$, as manifolds.*

We have therefore obtained an identification between $\mathbb{R}\mathbf{H}^n$ and $\mathbb{R}^{n-1} \rtimes \mathbb{R}$ with the multiplication given by, thanks to Lemma 10.5.5,

$$(u; a)(v; b) = (u + e^a v; a + b). \tag{10.5.8}$$

We are left to find the distance on the group and prove the left-invariance.

Given the identifications $(u; a), (v; b) \in \mathbb{R}^{n-1} \times \mathbb{R}$ for $x, y \in \mathbb{R}\mathbf{H}^n$, respectively, we want to write the distance between x and y as a function of u, v, a, b. The hyperbolic distance $d(\cdot,\cdot)$, as in Definition 10.2.1, is written as a function of $\langle\cdot,\cdot\rangle$ so we need to express the Hermitian form in these new coordinates. The Hermitian form can be applied only on elements of $\mathbb{R}^{n+1}$, so we define $\hat{x} \in x$ such that $\hat{x} := h(u)A(a)\hat{o}$ where $\hat{o} = (1, 0, \ldots, 0, 1)^t \in \mathbb{R}^{n+1}$. We observe that we can make this calculation with a representative because, while the Hermitian form depends on the chosen representative, the distance does not, as proved in Proposition 10.2.3.

Lemma 10.5.9. *Given $x \in \mathbb{R}\mathbf{H}^n$, we have $\langle\hat{x}, \hat{x}\rangle = -2$.*

Proof. Thanks to the identification we know that there exists $(u, a) \in \mathbb{R}^{n-1} \times \mathbb{R}$ such that $x = h(u)A(a) \cdot o$. The lemma follows from the calculation

$$\langle\hat{x}, \hat{x}\rangle = \langle h(u)A(a)\hat{o}, h(u)A(a)\hat{o}\rangle = \langle\hat{o}, \hat{o}\rangle = -2.$$

Note that the second equality holds because $NA \subset O_{\mathbb{K}}(n, 1)$, as proved in Theorem 10.4.11. $\square$

To deal with the general case $|\langle\hat{x}, \hat{y}\rangle|^2$, we first look at the case when $\hat{x} = \hat{o}$:

Lemma 10.5.10. *Let $y \in \mathbb{R}\mathbf{H}^n$ be identified with $(v; b) \in \mathbb{R}^{n-1} \times \mathbb{R}$, we have*

$$|\langle \hat{o}, \hat{y} \rangle|^2 = 4\cosh^2(b) + 2e^{-b}\cosh(b)|v|^2 + e^{-2b}\frac{|v|^4}{4}.$$

Proof. The proof consists in a simple computation of $\langle o, y \rangle$: We have

$$\begin{aligned}
\langle \hat{o}, \hat{y} \rangle &= \langle \hat{o}, h(v)A(b)\hat{o} \rangle \\
&= \hat{o}^* K h(v) A(b) \hat{o} \\
&= \hat{o}^* K \begin{pmatrix} 1 & v^t & \frac{|v|^2}{2} \\ 0 & I_{n-1} & \overline{v} \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} e^b & 0 & 0 \\ 0 & I_{n-1} & 0 \\ 0 & 0 & e^{-b} \end{pmatrix} \hat{o} \\
&= \hat{o}^* K \begin{pmatrix} e^b + e^{-b}\frac{|v|^2}{2} \\ e^{-b}\overline{v} \\ e^{-b} \end{pmatrix} \\
&= \hat{o}^* \begin{pmatrix} -e^{-b} \\ e^{-b}\overline{v} \\ -e^b - e^{-b}\frac{|v|^2}{2} \end{pmatrix} = -e^{-b} - e^b - e^{-b}\frac{|v|^2}{2} \\
&= -2\cosh(b) - e^{-b}\frac{|v|^2}{2},
\end{aligned}$$

where by definition $\cosh b = \frac{e^b + e^{-b}}{2}$. Thus we conclude

$$|\langle \hat{o}, \hat{y} \rangle|^2 = 4\cosh^2(b) + 2\cosh(b)e^{-b}|v|^2 + e^{-2b}\frac{|v|^4}{4}. \qquad \square$$

Thanks to the previous lemma we can now deal with the general case:

Proposition 10.5.11. *Let $x, y \in \mathbb{R}\mathbf{H}^n$ be identified with $(u; a), (v; b) \in \mathbb{R}^{n-1} \times \mathbb{R}$ respectively, then*

$$|\langle \hat{x}, \hat{y} \rangle|^2 = 4\cosh^2(b-a) + 2e^{-a-b}\cosh(b-a)|v-u|^2 + e^{-2(a+b)}\frac{|v-u|^4}{4}.$$

Proof. We recall that $\langle \hat{x}, \hat{y} \rangle$ is the Hermitian form associated to the Hermitian matrix K defined in (10.1.6). That is $\langle \hat{x}, \hat{y} \rangle = \hat{x}^* K \hat{y}$. We compute

$$\begin{aligned}
\langle \hat{x}, \hat{y} \rangle &= \langle h(u)A(a)\hat{o}, h(v)A(b)\hat{o} \rangle \\
&= \hat{o}^t A(a) h(u)^* K h(v) A(b) \hat{o} \\
&= \hat{o}^t K A(-a) h(-u) h(v) A(b) \hat{o} \\
&= \hat{o}^t K A(-a) h(v-u) A(b) \hat{o} \\
&= \hat{o}^t K h(e^{-a}(v-u)) A(b-a) \hat{o} \\
&= \langle \hat{o}, h(e^{-a}(v-u)) A(b-a) \hat{o} \rangle,
\end{aligned}$$

where the equalities are obtained by the following reasoning: the first equality follows from the definition of Hermitian form. The second one is a consequence of the fact that $NA \subset O_{\mathbb{K}}(n,1)$ as proved in Theorem 10.4.11. The third one holds thanks to the isomorphism, proved in Theorem 10.5.3, between N and $\mathbb{R}^{n-1}$. The fourth one follows from Lemma 10.5.5 and the last one is simply the definition. Now using Lemma 10.5.10 we continue:

$$\begin{aligned}
|\langle \hat{x}, \hat{y}\rangle|^2 &= |\langle \hat{o}, h(e^{-a}(v-u))A(b-a)\hat{o}\rangle|^2 \\
&= 4\cosh^2(b-a) + 2e^{a-b}\cosh(b-a)|e^{-a}(v-u)|^2 + e^{-2(b-a)}\frac{|e^{-a}(v-u)|^4}{4} \\
&= 4\cosh^2(b-a) + 2e^{-a-b}\cosh(b-a)|v-u|^2 + e^{-2(a+b)}\frac{|v-u|^4}{4}. \qquad \square
\end{aligned}$$

We are now ready to write the function of the distance in these new coordinates:

Corollary 10.5.12. *After the identification of $\mathbb{R}\mathbf{H}^n$ with $\mathbb{R}^{n-1} \rtimes \mathbb{R}$ as done in Corollary 10.5.7, the distance on $\mathbb{R}^{n-1} \rtimes \mathbb{R}$ reads as*

$$\cosh^2 d((u;a),(v;b)) = \frac{4\cosh^2(b-a) + 2e^{-a-b}\cosh(b-a)|v-u|^2 + e^{-2(a+b)}\frac{|v-u|^4}{4}}{4}.$$

Such a distance is left-invariant with respect to the product structure of $\mathbb{R}^{n-1} \rtimes \mathbb{R}$, as in (10.5.2).

Proof. The distance function follows from the definition of $d(\cdot,\cdot)$ on $\mathbb{R}\mathbf{H}^n$, given in Definition 10.2.1, Lemma 10.5.10, and Proposition 10.5.11. We know that the distance is left-invariant because the multiplication in $\mathbb{R}^{n-1} \rtimes \mathbb{R}$ acts by isometries. A simple calculation can check this fact: let $(w;c) \in \mathbb{R}^{n-1} \rtimes \mathbb{R}$ then

$$(w;c)(u;a) = (w + e^c u; c+a)$$

and

$$(w;c)(v;b) = (w + e^c v; c+b).$$

Then we have the left-invariance:

$$\begin{aligned}
&\cosh^2(d((w;c)(u;a),(w;c)(v;b))) \\
&\quad= \frac{4\cosh^2(b-a) + 2e^{-a-b-2c}\cosh(b-a)e^{2c}|v-u|^2 + e^{-2(a+b+2c)}\frac{e^{4c}|v-u|^4}{4}}{4} \\
&\quad= \frac{4\cosh^2(b-a) + 2e^{-a-b}\cosh(b-a)|v-u|^2 + e^{-2(a+b)}\frac{|v-u|^4}{4}}{4} \\
&\quad= \cosh^2 d((u;a),(v;b)). \qquad \square
\end{aligned}$$

10.5.2 The quaternionic case

Our aim is to prove that the quaternionic n- hyperbolic space $\mathbb{H}\mathbf{H}^n$ admits a group structure for which the distance is left-invariant and such a structure is $\mathbb{H}\mathcal{H}^n \rtimes \mathbb{R}$, where $\mathbb{H}\mathcal{H}^n$ is the n-th $\mathbb{K}$-Heisenberg groups defined in Definition 10.3.1, and the action of $\mathbb{R}$ on $\mathbb{H}\mathcal{H}^n$ is by standard dilations. In this section we work with the group A and N as in Definition 10.4.8. We already know that A is isomorphic to $\mathbb{R}$ as proved in Lemma 10.4.11. Next we better characterize the group structure of $N = \{\nu(M, M_{13}) : M \in \mathrm{GL}(1, n-1, \mathbb{K}), M_{13} \in \mathbb{K}, |M|^2 = M_{13} + \overline{M}_{13}\}$, where ν is defined in (10.4.9).

Lemma 10.5.13. *When $\mathbb{K} = \mathbb{H}$ the group N is isomorphic to $\mathbb{H}\mathcal{H}^n$.*

Proof. To prove this isomorphism we need to write explicitly what form the matrices in the group N have. As we have recalled, the set N consists of the matrices of the form

$$\nu(M, M_{13}) = \begin{pmatrix} 1 & M & M_{13} \\ 0 & I_{n-1} & M^* \\ 0 & 0 & 1 \end{pmatrix},$$

where $M \in \mathcal{M}(1, n-1, \mathbb{H})$ and $M_{13} \in \mathbb{H}$ satisfy $|M|^2 = M_{13} + \overline{M}_{13}$. Let $u := M^t$. We note that

$$M_{13} + \overline{M}_{13} = 2\Re(M_{13}),$$

therefore

$$\Re(M_{13}) = \frac{|u|^2}{2}.$$

The matrices in N can be written in the form

$$h(u, s) := \begin{pmatrix} 1 & u^t & \frac{|u|^2}{2} + s_1 i + s_2 j + s_3 k \\ 0 & I_{n-1} & \overline{u} \\ 0 & 0 & 1 \end{pmatrix}. \tag{10.5.14}$$

So if $\mathbb{K} = \mathbb{H}$ the group N can be defined as follows:

$$N = \{h(u, s) : u \in \mathbb{H}^{n-1}, s \in \mathbb{R}^3\}.$$

We claim that the mapping $h : \mathbb{H}\mathcal{H}^n \to N$ defined as

$$h : \mathbb{H}\mathcal{H}^n \to N$$

$$(u, s) \mapsto h(u, s),$$

is a group isomorphism. It follows directly that $h(0, 0) = I$. We observe that for all $x, y \in \mathbb{H}^{n-1}$

$$|x + y|^2 = |x|^2 + |y|^2 + 2\Re(x^t \overline{y}). \tag{10.5.15}$$

By computing the product and using the relation (10.5.15) we obtain

$$\begin{aligned}
h(u,s)h(v,r) &= \begin{pmatrix} 1 & u^t & \frac{|u|^2}{2} + s_1 i + s_2 j + s_3 k \\ 0 & I_{n-1} & \overline{u} \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & v^t & \frac{|v|^2}{2} + r_1 i + r_2 j + r_3 k \\ 0 & I_{n-1} & \overline{v} \\ 0 & 0 & 1 \end{pmatrix} \\
&= \begin{pmatrix} 1 & u^t + v^t & \frac{|v|^2}{2} + u^t\overline{v} + \frac{|u|^2}{2} + (r_1+s_1)i + (r_2+s_2)j + (r_3+s_3)k \\ 0 & I_{n-1} & \overline{u}+\overline{v} \\ 0 & 0 & 1 \end{pmatrix} \\
&= \begin{pmatrix} 1 & u^t + v^t & \frac{|u+v|^2}{2} + \mathfrak{I}_1 i + \mathfrak{I}_2 j + \mathfrak{I}_3 k \\ 0 & I_{n-1} & u+v \\ 0 & 0 & 1 \end{pmatrix} \\
&= h(u+v, s+r+\mathrm{Im}(u^t\overline{v})),
\end{aligned}$$

where $\mathfrak{I}_\lambda := r_\lambda + s_\lambda + \mathrm{Im}_\lambda(u^t\overline{v})$ for $\lambda \in \{1,2,3\}$. This let us to prove that

$$h(u,s)h(v,r) = h(u+v, s+r+\mathrm{Im}(u^t\overline{v})) = h((u,s)(v,t)).$$

The bijectivity follows from the definition and the inverse of h is given by

$$h^{-1} : N \to \mathbb{H}\mathcal{H}^n$$
$$h(u,s) \mapsto (u,s).$$

□

For the sequel we shall need the following identities.

Lemma 10.5.16. *Let $u, v \in \mathbb{H}^{n-1}$, $s, t \in \mathbb{R}^3$ and $a, b \in \mathbb{R}$ then*

$$A(a)h(v,t) = h(e^a v, e^{2a} t)A(a),$$

$$h(u,s)A(a)h(v,t)A(b) = h(u + e^a v, s + e^{2a} t + \mathrm{Im}(u^t e^a \overline{v}))A(a+b),$$

and

$$(h(u,s)A(a))^{-1} = h(-e^{-a}u, -e^{-2a}s)A(-a).$$

Proof. Thanks to Lemma 10.4.10 and Lemma 10.5.13 we know that

$$A(a)h(v,t) = h(e^a v, e^{2a} t)A(a).$$

Therefore

$$\begin{aligned}
h(u,s)A(a)h(v,t)A(b) &= h(u,s)h(e^a v, e^{2a} t)A(a)A(b) \\
&= h(u + e^a v, s + e^{2a} t + \mathrm{Im}(u^t e^a \overline{v}))A(a+b),
\end{aligned}$$

where the last equality follows from Lemma 10.5.13. The formula for the inverse element follows from the properties of the matrix inverse and the first equality proved in this lemma. $\square$

We next prove that NA acts simply transitively on $\mathbb{H}\mathbf{H}^n$.

Theorem 10.5.17. *The group NA acts simply transitively on $\mathbb{H}\mathbf{H}^n$, that is, for every $x, y \in \mathbb{H}\mathbf{H}^n$ there exists a unique $g \in NA$ such that $g \cdot x = y$*

Proof. We consider the point $o := [1, 0, \ldots, 0, 1]^t \in \mathbb{H}\mathbf{H}^n$ and we want to show that given an arbitrary point $x \in \mathbb{H}\mathbf{H}^n$ with homogeneous coordinates $[x_1, \ldots, x_n, 1]$, there exists a unique $(u, s; a) \in \mathbb{H}^{n-1} \times \mathbb{R}^3 \times \mathbb{R}$ such that x is the image under $h(u, s)A(a)$ of the point o, where h is defined in (10.5.14) and A in Definition 10.4.8. Since

$$h(u,s)A(a) = \begin{pmatrix} e^a & u^t & e^{-a}\left(\frac{|u|^2}{2} + s_1 i + s_2 j + s_3 k\right) \\ 0 & I_{n-1} & e^{-a}\overline{u} \\ 0 & 0 & e^{-a} \end{pmatrix},$$

we have

$$\begin{aligned} h(u,s)A(a) \cdot o &= \begin{bmatrix} e^a + e^{-a}\left(\frac{|u|^2}{2} + s_1 i + s_2 j + s_3 k\right) \\ e^{-a}\overline{u} \\ e^{-a} \end{bmatrix} \\ &= \begin{bmatrix} e^{2a} + \frac{|u|^2}{2} + s_1 i + s_2 j + s_3 k \\ \overline{u} \\ 1 \end{bmatrix}. \end{aligned}$$

So $h(u, s)A(a) \cdot o = x$ becomes the system of equations:

$$\begin{cases} x_1 = e^{2a} + \frac{|u|^2}{2} + s_1 i + s_2 j + s_3 k \\ X_2 = \overline{u} \end{cases},$$

where $X_2^t = (x_2, \ldots, x_n)$. Thus X_2 uniquely determines $\overline{u}$, and hence u. We now observe that the number $x_1 \in \mathbb{H}$ is equal to $a_x + b_x i + c_x j + d_x k$ for suitable $a_x, b_x, c_x, d_x \in \mathbb{R}$. Noting that $e^{2a} + \frac{|u|^2}{2} \in \mathbb{R}$ leads the conclusion that $s_1 = b_x$, $s_2 = c_x$, $s_3 = d_x$. The fact that $[x] \in \mathbb{H}\mathbf{H}^n$ means x must satisfies $\langle x, x\rangle < 0$. We recall that $\langle \cdot, \cdot \rangle = \langle \cdot, \cdot \rangle_K$, where K is the matrix as in (10.1.6), so

$$-\overline{x}_1 x_{n+1} - \overline{x}_{n+1} x_1 + \sum_{i=2}^{n} \overline{x}_i x_i < 0$$

which, in our case, is equivalent to

$$-\overline{x}_1 - x_1 + \sum_{i=2}^{n} |x_i|^2 = |X_2|^2 - \overline{x}_1 - x_1 < 0 \Leftrightarrow |X_2|^2 < x_1 + \overline{x}_1 = 2\Re(x_1).$$

We now conclude that, if $B := \frac{|u|^2}{2} + s_1 i + s_2 j + s_3 k$, then

$$\Re(x_1) > \frac{|X_2|^2}{2} = \frac{1}{2}\left(B + \overline{B}\right) = \Re(B) = \frac{|u|^2}{2}.$$

Therefore, there exists a unique $a \in \mathbb{R}$ such that $\Re(x_1) = e^{2a} + \frac{|u|^2}{2}$. Thus the above system of equation has a unique solution $(u, s; a) \in \mathbb{H}^n \times \mathbb{R}^3 \times \mathbb{R}$.

Let now $x \in \mathbb{H}\mathbf{H}^n$ be an arbitrary element, we want to prove that given $y \in \mathbb{H}\mathbf{H}^n$ there exists a unique $(u, s; a) \in \mathbb{H}^{n-1} \times \mathbb{R}^3 \times \mathbb{R}$ such that $h(u, s)A(a)x = y$. We now know that there exist a unique g_y and a unique $g_x \in NA$ such that $x = g_x \cdot o$ and $y = g_y \cdot o$. We claim that the wanted element is $g_y g_x^{-1}$. As a matter of fact

$$g_y g_x^{-1} \cdot x = g_y g_x^{-1} g_x \cdot o = g_y \cdot o = y.$$

Let $g_x = h(v, t)A(b)$ and $g_y = h(w, r)A(c)$. Then

$$\begin{aligned} g_y g_x^{-1} &= h(w, r)A(c)(h(v, t)A(b))^{-1} \\ &= h(w, r)A(c)h(-e^{-b}v, -e^{-2b}t)A(-b) \\ &= h(w - e^{c-b}v, r - e^{2c-2b}t - \mathrm{Im}((we^{c-b} - e^{2c-2b}v)^t\overline{v}))A(c - b), \end{aligned}$$

where the equalities hold thanks to Lemma 10.5.16. From this calculation follow that $u = w - e^{c-b}v$, $s = r - e^{2c-2b}t - \mathrm{Im}((we^{c-b} - e^{2c-2b}v)^t\overline{v})$ and $a = c - b$, thus the proof is completed. □

From Theorem 10.5.17 we get the following consequence.

Corollary 10.5.18. *The mapping $(u, s; a) \mapsto h(u, s)A(a) \cdot o$ gives a smooth identification between $\mathbb{H}\mathcal{H}^n \times \mathbb{R}$ and $\mathbb{H}\mathbf{H}^n$, as manifolds.*

We have therefore obtain an identification between $\mathbb{H}\mathbf{H}^n$ and $\mathbb{H}\mathcal{H}^n \rtimes \mathbb{R}$ with the multiplication given by Lemma 10.5.16

$$(u, s; a)(v, t; b) = (u + e^a v, s + e^{2a}t + \mathrm{Im}(u^t e^a \overline{v}); a + b). \tag{10.5.19}$$

We are left to find the distance on the group and prove the left-invariance.

Given the identifications $(u, s; a), (v, t; b) \in \mathbb{H}\mathcal{H}^n \times \mathbb{R}$ of $x, y \in \mathbb{H}\mathbf{H}^n$, respectively, we want to write the distance between x and y as a function of u, v, s, t, a, b. The hyperbolic distance $d(\cdot, \cdot)$, as in Definition 10.2.1, is written as a function of $\langle \cdot, \cdot \rangle$ so we need to express the Hermitian form in these new coordinates. The Hermitian form can be applied only on elements of $\mathbb{H}^{n+1}$, so we define

$\hat{x} \in x$ such that $\hat{x} := h(u,s)A(a)\hat{o}$ where $\hat{o} = (1,0,\dots,0,1)^t \in \mathbb{H}^{n+1}$. We observe that we can make these calculations with a representative because, while the Hermitian form depends on the chosen representative, the distance does not, as proved in Proposition 10.2.3.

Lemma 10.5.20. *Given $x \in \mathbb{H}\mathbf{H}^n$ we have $\langle \hat{x}, \hat{x} \rangle = -2$.*

Proof. Given $x \in \mathbb{H}\mathbf{H}^n$ identified with $(u,s;a) \in \mathbb{H}\mathcal{H}^n \times \mathbb{R}$ we have

$$\langle \hat{x}, \hat{x} \rangle = \langle h(u,s)A(a)\hat{o}, h(u,s)A(a)\hat{o} \rangle = \langle \hat{o}, \hat{o} \rangle = -2. \qquad \square$$

To deal with the general case $|\langle \hat{x}, \hat{y} \rangle|^2$, we first look at the case when $\hat{x} = \hat{o}$:

Lemma 10.5.21. *Let $y \in \mathbb{H}\mathbf{H}^n$ identified with $(v,t;b) \in \mathbb{H}\mathcal{H}^n \times \mathbb{R}$ then*

$$|\langle \hat{o}, \hat{y} \rangle|^2 = 4\cosh^2(b) + 2e^{-b}\cosh(b)|v|^2 + e^{-2b}\left(\frac{|v|^4}{4} + |t|^2\right).$$

Proof. The proof consist in a simple computation of $\langle \hat{o}, \hat{y} \rangle$. We have

$$\begin{aligned}
\langle \hat{o}, \hat{y} \rangle = (\hat{o}, h(v,t)A(b)\hat{o}) &= \hat{o}^t K h(v,t)A(b)\hat{o} \\
&= \hat{o}^t K \begin{pmatrix} e^b + e^{-b}\left(\frac{|v|^2}{2} + t_1 i + t_2 j + t_3 k\right) \\ e^{-b}\overline{v} \\ e^{-b} \end{pmatrix} \\
&= \hat{o}^t \begin{pmatrix} -e^{-b} \\ e^{-b}\overline{v} \\ -\left(e^b + e^{-b}\left(\frac{|v|^2}{2} + t_1 i + t_2 j + t_3 k\right)\right) \end{pmatrix} \\
&= -e^{-b} - e^b - e^{-b}\left(\frac{|v|^2}{2} + t_1 i + t_2 j + t_3 k\right) \\
&= -2\cosh(b) - e^{-b}\left(\frac{|v|^2}{2} + t_1 i + t_2 j + t_3 k\right),
\end{aligned}$$

where by definition $\cosh b = \frac{e^b - e^{-b}}{2}$. By taking the squared norm of the number we conclude

$$|\langle \hat{o}, \hat{y} \rangle|^2 = 4\cosh^2(b) + 2e^{-b}\cosh(b)|v|^2 + e^{-2b}\left(\frac{|v|^4}{4} + |t|^2\right). \qquad \square$$

Thanks to the previous lemma we can now deal with the general case.

Proposition 10.5.22. *Let $x, y \in \mathbb{H}\mathbf{H}^n$ identified with $(u,s;a), (v,t;b) \in \mathbb{H}\mathcal{H}^n \times \mathbb{R}$, respectively, then*

$$|\langle \hat{x}, \hat{y} \rangle|^2 = 4\cosh^2(b-a) + 2e^{-a-b}\cosh(b-a)|v-u|^2 + e^{-2(a+b)}\left(\frac{|v-u|^4}{4} + |t - s - \operatorname{Im}(u\overline{v})|^2\right).$$

Proof. We recall that $\langle \cdot, \cdot \rangle$ is the Hermitian form associated to the Hermitian matrix K as defined in (10.1.6). That is $\langle \hat{x}, \hat{y} \rangle = \langle \hat{x}, \hat{y} \rangle_K$. We have the following calculations, which we will subsequently explain:

$$\begin{aligned}
\langle \hat{x}, \hat{y} \rangle &= \langle h(u,s)A(a)\hat{o}, h(v,t)A(b)\hat{o} \rangle \\
&= \hat{o}^t A(a) h(u,s)^* K h(v,t) A(b)\hat{o} \\
&= \hat{o}^t K A(a)^{-1} h(-u,-s) h(v,t) A(b)\hat{o} \\
&= \hat{o}^t K A(-a) h(w,r) A(b)\hat{o} \\
&= \hat{o}^t K h(\varepsilon_a^{-1}(w,r)) A(b-a)\hat{o} \\
&= \langle \hat{o}, h(\varepsilon_a^{-1}(w,r)) A(b-a)\hat{o} \rangle,
\end{aligned}$$

where $\mathbb{H}\mathcal{H}^n \ni (w,r) = (-u,-s)(v,t)$ and $\varepsilon_a^{-1} := \delta_{e^a}^{-1}$ is the Heisenberg homothety of ratio e^{-a}, as in (10.3.4). The equalities are obtained by the following reasoning: the first one follows from the identification in the hypothesis of the proposition. The second equality follows from the definition of the Hermitian form. The third one is a consequence of the fact that $NA \subset O_{\mathbb{H}}$, as proved in Theorem 10.4.11, and the characterization of the elements of $O_{\mathbb{H}}$, proved in Lemma 10.4.4. The fourth one holds thanks to the isomorphism between A and $\mathbb{R}$, Lemma 10.4.11, and the isomorphism between N and $\mathbb{H}\mathcal{H}^n$, proved in Lemma 10.5.13. The fifth one follows from Lemma 10.5.16 and the last one is simply the definition. Computing $\varepsilon_a^{-1}(w,r)$ we obtain

$$\varepsilon_a^{-1}(w,r) = \delta_{e^a}^{-1}(w,r) = \delta_{e^{-a}}(v-u, t-s+\mathrm{Im}(-u^t\overline{v})) = (e^{-a}(v-u), e^{-2a}(t-s-\mathrm{Im}(u^t\overline{v}))).$$

Now using Lemma 10.5.21, we continue.

$$\begin{aligned}
|\langle \hat{x}, \hat{y} \rangle|^2 &= |\langle \hat{o}, h(\varepsilon_a^{-1}(w,r)) A(b-a)\hat{o} \rangle|^2 \\
&= 4\cosh^2(b-a) + 2e^{a-b}\cosh(b-a)|e^{-a}(v-u)|^2 + \\
&\qquad + e^{-2(b-a)} \left(\frac{|e^{-a}(v-u)|^4}{4} + |e^{-2a}(t-s-\mathrm{Im}(u^t\overline{v}))|^2 \right) \\
&= 4\cosh^2(b-a) + 2e^{-a-b}\cosh(b-a)|v-u|^2 + \\
&\qquad + e^{-2(a+b)} \left(\frac{|v-u|^4}{4} + |t-s-\mathrm{Im}(u^t\overline{v})|^2 \right). \square
\end{aligned}$$

We are now ready to write the function of the distance in these new coordinates:

Corollary 10.5.23. *After the identification of $\mathbb{H}\mathbf{H}^n$ with $\mathbb{H}\mathcal{H}^n \rtimes \mathbb{R}^3$ as done in Corollary 10.5.18,* *the distance on* $\mathbb{H}\mathcal{H}^n \rtimes \mathbb{R}^3$ *reads as*

$$\begin{aligned}&\cosh^2 d((u,s;a),(v,t;b))\\ &\quad= \;4\cosh^2(b-a) + 2e^{-a-b}\cosh(b-a)|v-u|^2 + e^{-2(a+b)}\left(\frac{|v-u|^4}{4} + |t-s-\mathrm{Im}(u^t\overline{v})|^2\right).\end{aligned}$$

This distance is left-invariant with respect to the product structure of $\mathbb{H}\mathcal{H}^n \rtimes \mathbb{R}^3$*, as in* (10.5.2).

Proof. The formula of the distance follows from the definition of $d(\cdot,\cdot)$ on $\mathbb{H}\mathbf{H}^n$, Lemma 10.5.21, and Proposition 10.5.22. We know that the distance is left-invariant because the operation in $\mathbb{H}\mathcal{H}^n \rtimes \mathbb{R}$ acts by isometries. A simple check can prove this fact: let $(w,r;c) \in \mathbb{H}\mathcal{H}^n \rtimes \mathbb{R}$ then

$$(w,r;c)(u,s;a) = (w + e^c u, r + e^{2c}s + e^c\mathrm{Im}(w^t\overline{u}); c+a)$$

and

$$(w,r;c)(v,t;b) = (w + e^c v, r + e^{2c}t + e^c\mathrm{Im}(w^t\overline{v}); c+b).$$

Then we get the left-invariance:

$$\begin{aligned}
&4\cosh^2(d((w,r;c)(u,s;a),(w,r;c)(v,t;b)))\\
&= 4\cosh^2(b-a) + 2e^{-a-b-2c}\cosh(b-a)e^{2c}|v-u|^2+\\
&\qquad + e^{-2(a+b+2c)}\left(\frac{e^{4c}|v-u|^4}{4} + |e^{2c}(t-s) + e^c(\mathrm{Im}(w^t\overline{v}) - \mathrm{Im}(w^t\overline{u})) -\right.\\
&\qquad\qquad \left. -\mathrm{Im}(|w|^2 + e^c w^t\overline{v} + e^c u^t\overline{w} + e^{2c}u^t\overline{v})|^2\right)\\
&= 4\cosh^2(b-a) + 2e^{-a-b}\cosh(b-a)|v-u|^2 4+\\
&\qquad + e^{-2(a+b+2c)}\left(\frac{e^{4c}|v-u|^4}{4} + |e^{2c}(t-s) - e^c\mathrm{Im}(u^t\overline{w} + w^t\overline{u}) -\right.\\
&\qquad\qquad \left. -\mathrm{Im}(e^{2c}u^t\overline{v})|^2\right)\\
&= 4\cosh^2(b-a) + 2e^{-a-b}\cosh(b-a)|v-u|^2+\\
&\qquad + e^{-2(a+b+2c)}\left(\frac{e^{4c}|v-u|^4}{4} + |e^{2c}(t-s) - \mathrm{Im}(e^{2c}u^t\overline{v})|^2\right)\\
&= 4\cosh^2(b-a) + 2e^{-a-b}\cosh(b-a)|v-u|^2+\\
&\qquad + e^{-2(a+b)}\left(\frac{|v-u|^4}{4} + |t-s-\mathrm{Im}(u^t\overline{v})|^2\right)\\
&= \cosh^2 d((u,s;a),(v,t;b)).
\end{aligned}$$

Most of the equalities are simple calculation, the only thing to note is that for all $w,u \in \mathbb{H}^n$ is true that $w^t\overline{u} + u^t\overline{w} \in \mathbb{R}$. As a matter of fact $\overline{w^t\overline{u} + u^t\overline{w}} = u^t\overline{w} + w^t\overline{u}$, thanks to the properties of the conjugation described in the first chapter. $\square$

10.5.3 The complex case

Out for completeness we write the theorem for the complex case, that we remember follows the same exact reasoning of the quaternionic case.

Theorem 10.5.24. *The complex hyperbolic space $\mathbb{C}\mathbf{H}^n$ admits a group structure with a left invariant distance and such structure is $\mathbb{C}\mathcal{H}^n \rtimes \mathbb{R}$ with the operation given by*

$$(u, s; a)(v, t; b) = (u + e^a v, s + e^{2a} t + \mathrm{Im}(u^t e^a \overline{v}); a + b).$$

The distance on $\mathbb{C}\mathcal{H}^n \rtimes \mathbb{R}$ reads as

$$\begin{aligned} &4\cosh^2 d((u, s; a), (v, t; b)) \\ &\qquad = 4\cosh^2(b-a) + 2e^{-a-b}\cosh(b-a)|v-u|^2 + e^{-2(a+b)}\left(\frac{|v-u|^4}{4} + |t - s - \mathrm{Im}(u^t\overline{v})|^2\right). \end{aligned}$$

This distance is left-invariant with respect to the product structure of $\mathbb{C}\mathcal{H}^n \rtimes \mathbb{R}$.

Another way to prove Theorem 10.5.24 is to observe that $\mathbb{C}\mathbf{H}^n$ embeds isometrically into $\mathbb{H}\mathbf{H}^n$ by embedding a copy of $\mathbb{C}^{n+1}$ in $\mathbb{H}^{n+1}$.

10.6 The octonionic hyperbolic plane

One case still remains: the octonionic hyperbolic plane. In this chapter we explain why it can not be treated like the other cases, and provide some ideas on how to deal with it. We first start by introducing the Octonions.

The octonion number set $\mathbb{O}$ are the 8-dimensional algebra over $\mathbb{R}$ with base $\{i_j : j = 1, \dots, 7\}$, where

$$1 i_j = i_j 1 = i_j, \quad i_j^2 = -1, \quad i_j i_k = -i_k i_j \quad \forall j, k \in \{1, \dots, 7\},$$

and

$$i_j i_k = i_l$$

precisely when (j, k, l) is a cyclic permutation of one of the triples:

$$(1, 2, 4),\ (1, 3, 7),\ (1, 5, 6)\ (2, 3, 5),\ (2, 6, 7),\ (3, 4, 6),\ (4, 5, 7).$$

An octonion z has the form $z = z_0 + \sum_{j=1}^{7} z_j i_j$. The *conjugate* $\overline{z}$ of z is defined to be $z = z_0 - \sum_{j=1}^{7} z_j i_j$. Conjugation has the property that $\overline{zw} = \overline{w}\,\overline{z}$ for all $z, w \in \mathbb{O}$. In a similar way to the complex and quaternionic case one define the *real part* and the *imaginary part* as $\Re(z) = \frac{1}{2}(z + \overline{z})$

and $\mathrm{Im}(z) = \frac{1}{2}(z - \overline{z})$. The *norm* $|z|$ of an octonion is the non-negative real number defined by $|z| = \overline{z}z = z\overline{z} = \sum_{j=0}^{7} z_j^2$. It is easy to see that the product in $\mathbb{O}$ is not associative, for example

$$i_1((1 + i_4)i_3) = i_1(i_3 - i_6) = i_7 + i_5,$$

while

$$(i_1(1 + i_4))i_3 = (i_1 - i_2)i_3 = i_7 - i_5.$$

The lack of associativity make $\mathbb{O}$ lose the notion of a vector space. This is the point that does not let us to work with the Octonion as with the other numbers. While for the Quaternion we simply have to consider right and left multiplication as different things, there is no way to build a vector space on $\mathbb{O}$. The idea here is to use the fact that two generators subalgebras of $\mathbb{O}$ are associative, this result is due to Artin, see [Sch95, Section III.1].

Proposition 10.6.1. *For every octonions x and y the subalgebra with unit generated by x and y is associative. In particular, every product of octonions that may be written in terms of just two octonions is associative.*

Consider $z = (z_1, z_2)$ where $z_1, z_2 \in \mathbb{O}$, we define the *standard lift* of z as

$$\mathbf{z} = \begin{pmatrix} z_1 \\ z_2 \\ 1 \end{pmatrix}.$$

Suppose that λ is an octonion in the same associative subalgebra of $\mathbb{O}$ as z_1 and z_2, then we can let λ act on $\mathbf{z}$ by right multiplication:

$$\mathbf{z}\lambda = \begin{pmatrix} z_1\lambda \\ z_2\lambda \\ \lambda \end{pmatrix}.$$

We therefore define

$$\mathbb{O}_0^3 = \left\{ \mathbf{z} = \begin{pmatrix} z_1 \\ z_2 \\ z_3 \end{pmatrix} : z_1, z_2, z_3 \text{ all lie in some associative subalgebra of } \mathbb{O} \right\}.$$

We now work on $\mathbb{O}_0^3$ in a similar way we have done on $\mathbb{K}^{n+1}$.

Definition 10.6.2. We define an equivalence relation on $\mathbb{O}_0^3$ by $\mathbf{z} \sim \mathbf{w}$ if $\mathbf{w} = \mathbf{z}\lambda$ for some λ in an associative subalgebra of $\mathbb{O}$ containing the entries of $\mathbf{z}$. We denote the set of equivalence classes as $\mathbb{O}\mathbf{P}_0^2$.

Let H be a Hermitian matrix of signature $(2, 1)$, for example in the form of K, as (10.1.6). Given $\mathbf{z} \in \mathbb{O}_0^3$, we define $Z := \mathbf{z}\mathbf{z}^*H$. This is a 3×3 matrix whose entries lie in an associative subalgebra of $\mathbb{O}$.

Lemma 10.6.3. *Right multiplication of* $\mathbf{z}$ *by* λ *lying in the same associative subalgebra as the entries of* $\mathbf{z}$*, leads to multiplication of* Z *by* $|\lambda|^2$.

Proof. The proof is a simple calculation:

$$(\mathbf{z}\lambda)(\mathbf{z}\lambda)^* H = \mathbf{z}\lambda\lambda^*\mathbf{z}^* = |\lambda|^2 \mathbf{z}\mathbf{z}^* H. \qquad \square$$

We consider $M(\mathbb{O}, 3)$ to be the real vector space of 3×3 octonionic matrices. Let X^* the conjugate transpose of a matrix X in $M(\mathbb{O}, 3)$. We define

$$J = \big\{X \in M(\mathbb{O}, 3) : HX = X^* H\big\}.$$

Then J is closed under *Jordan multiplication*, that is

$$X * Y := \frac{1}{2}\left(XY + YX\right). \tag{10.6.4}$$

We call J the *Jordan algebra associated to* H. Real numbers act on $M(\mathbb{O}, 3)$ by multiplication of each entry of X. So we define an equivalence relation on J by $X \sim Y$ if and only if $Y = kX$ for some non-zero real number k. Then $\mathbb{J}\mathbf{P}$ is defined to be the set these of equivalence classes.
Let $\mathfrak{Z} : \mathbb{O}^3_0 \to M(\mathbb{O}, 3)$ such that $\mathfrak{Z}(\mathbf{z}) = Z$, then $\mathfrak{Z}(\mathbf{z}) \in J$ for all $\mathbf{z} \in \mathbb{O}^3_0$, as a matter of fact

$$H\mathfrak{Z}(\mathbf{z}) = HZ = H\mathbf{z}\mathbf{z}^* H = (\mathbf{z}\mathbf{z}^* H)^* H = Z^* H = \mathfrak{Z}(\mathbf{z})^* H.$$

Hence the map $\mathfrak{Z}$ defines an embedding $\mathbb{O}^3_0 \to J$. Moreover, the two projection maps are compatible and so there is a well defined map $\mathbb{O}\mathbf{P}^3_0 \to \mathbb{J}\mathbf{P}$, thanks to Lemma 10.6.3. The Hermitian form in this case is provided by $tr(Z) = tr(\mathbf{z}\mathbf{z}^* H)$, which is real thanks to the fact that

$$\overline{tr(Z)} = tr(\overline{Z}) = tr(\overline{\mathbf{z}\mathbf{z}^* H}) = tr(\mathbf{z}\mathbf{z}^* H^*) = tr(\mathbf{z}\mathbf{z}^* H) = tr(Z).$$

On $M(3, \mathbb{O})$ we define a bilinear form by

$$\langle X, Y\rangle := \Re(tr(X * Y)) = \frac{1}{2}\Re(tr(XY + YX)),$$

where $X * Y$ is defined in (10.6.4).

We can finally give the definition of the octonionic hyperbolic plane:

Definition 10.6.5. Let $V_- := \{\mathbf{z} \in \mathbb{O}^3_0 : tr(\mathbf{z}\mathbf{z}^* H) < 0\}$ and let $V_-\mathbf{P}$ be its projectivization as in Definition 10.6.2. We define the *octonionic hyperbolic plane* $\mathbb{O}\mathbf{H}^2$ to be the set $V_-\mathbf{P}$ endowed with the distance

$$\cosh^2\left(\frac{d([\mathbf{z}], [\mathbf{w}])}{2}\right) = \frac{\langle Z, W\rangle}{tr(Z)tr(W)}.$$

The octonionic hyperbolic plane $\mathbb{O}\mathbf{H}^2$ and its distance are well defined thanks to Lemma 10.6.3. We leave the study of the group structure of $\mathbb{O}\mathbf{H}^2$ for the reader.

Chapter 11

Heintze groups and their visual boundaries*

In this chapter we show that to every Riemannian symmetric space one can associate a 'visual boundary' that has a structure of Carnot group. Visual boundaries are associated to spaces with negative curvature. We have that every homogeneous negatively curved manifold has the structure of semidirect product of the for $N \rtimes \mathbb{R}$ with a graded nilpotent group N that canonically represent the visual boundary.

11.1 CAT(-1) spaces and visual boundary*

[...] As we shall see, the $\mathbb{K}$-hyperbolic n-space $\mathbb{K}\mathbf{H}^n$ has sectional curvature less or equal than -1. From the pure metric view point one says that it is a CAT(-1) metric space. The definition of $CAT(-1)$, together with an explicit proof of this last statement, can be found in [BH99, Part II, Chapter 10].

...

Definition 11.1.1. Let $\xi_\infty, \eta_\infty \in \partial_\infty \mathbb{K}\mathbf{H}^n$, the *Gromov product* of ξ_∞, η_∞ with respect to ω and o is defined as follows:

$$(\xi_\infty, \eta_\infty)_{(\omega,o)} := \frac{1}{2} \lim_{t\to+\infty} \left(2t - d(\xi_t, \eta_t)\right).$$

The *visual distance* on $\partial_\infty \mathbb{K}\mathbf{H}^n \setminus \{\omega\}$ can be defined as

$$d_{vis}(\xi_\infty, \eta_\infty) := e^{-(\xi_\infty, \eta_\infty)_{(\omega,o)}}. \tag{11.1.2}$$

The proof that the visual distance as defined in (11.1.2) is a distance won't be discuss here. However, this fact follows from a more general theorem by Bourdon [Bou95] about the conditions in which the visual distance is a distance. This theorem can be applied on CAT(-1) spaces, and therefore can be applied to $\mathbb{K}$-hyperbolic n-space.

11.2 The visual distance for $\mathbb{K}$-hyperbolic spaces

Thanks to the way we defined the $\mathbb{K}$-hyperbolic n-space $\mathbb{K}\mathbf{H}^n$, in Definition 10.2.1, there is a natural way to realize its boundary. We recall that setwise

$$\mathbb{K}\mathbf{H}^n = \{[x] \in \mathbb{K}\mathbf{P}^n | \langle x, x\rangle < 0\},$$

where $\mathbb{K}\mathbf{P}^n$ is the n-dimensional $\mathbb{K}$-projective space and $\langle \cdot, \cdot \rangle$ is the Hermitian form associated to the Hermitian matrix K given by (10.1.6).

Definition 11.2.1. We define the *boundary at infinity of* $\mathbb{K}\mathbf{H}^n$ as the set

$$\partial_\infty \mathbb{K}\mathbf{H}^n := \{[x] \in \mathbb{K}\mathbf{P}^n | \langle x, x\rangle = 0\}.$$

Let $\omega, o \in \mathbb{K}\mathbf{P}^n$ where $\omega := [1, 0, \dots, 0]^t$ and $o := [1, 0, \dots, 0, 1]^t$. A simple calculation checks that $\omega \in \partial_\infty \mathbb{K}\mathbf{H}^n$ while $o \in \mathbb{K}\mathbf{H}^n$. We recall the subgroups A and N of $O_{\mathbb{K}}(n, 1)$, that is, the group of isometries that preserve the Hermitian form $\langle \cdot, \cdot \rangle$, which were defined in Definition 10.4.8. The set A is the group of matrices in the form

$$A(a) = \begin{pmatrix} e^a & 0 & 0 \\ 0 & I_{n-1} & 0 \\ 0 & 0 & e^{-a} \end{pmatrix},$$

while N is the group of matrices in the form

$$\begin{pmatrix} 1 & M & M_{13} \\ 0 & I_{n-1} & M^* \\ 0 & 0 & 1 \end{pmatrix},$$

where $M \in \mathbb{K}^{n-1}$ and $M_{13} \in \mathbb{R}$ satisfy $M_{13} + \overline{M}_{13} = |M|^2$. We also remind that N was proved to be isomorphic to the n-th $\mathbb{K}$-Heisenberg group $\mathbb{K}\mathcal{H}^n$, Theorem 10.5.3 and Theorem 10.5.13, so we write $h(u, s)$ with $u \in \mathbb{K}^{n-1}$ and $s \in \mathrm{Im}(\mathbb{K})$ to denote the elements of N, as in (10.5.14).

For every $\xi_0 \in N \cdot o$ we consider the map

$$\begin{aligned} \xi : \mathbb{R} &\to \mathbb{K}\mathbf{H}^n \\ t &\mapsto \xi_t = h(u, s)A(-t) \cdot o, \end{aligned} \tag{11.2.2}$$

where $u \in \mathbb{K}^{n-1}$ and $s \in \mathrm{Im}(\mathbb{K})$ such that $h(u, s)$ is the unique element of N that satisfies $\xi_0 = h(u, s) \cdot o$. The following lemma allows to extend ξ to $\mathbb{R} \cup \{-\infty, +\infty\}$.

Lemma 11.2.3. *For all $\xi_0 \in N \cdot o$ the following facts are true for the curve* (11.2.2)*:*

1. *ξ is a geodesic;*
2. *$\lim_{t\to-\infty} \xi_t = \omega$;*
3. *$\xi_\infty := \lim_{t\to+\infty} \xi_t \in \partial_\infty \mathbb{K}\mathbf{H}^n \setminus \{\omega\}$.*

Before starting the proof we want to point out that these are limits taken with respect to the topology on $\mathbb{K}\mathbf{P}^n$, which is the quotient topology from $\mathbb{K}^{n+1}$.

Proof. To prove that ξ is a geodesic we need to prove that $d(\xi_{t_1}, \xi_{t_2}) = |t_1 - t_2|$ for all $t_1, t_2 \in \mathbb{R}$, where d is the distance in Definition 10.2.1. By definition $\xi_{t_1} = h(u,s)A(-t_1)\cdot o$ and $\xi_{t_2} = h(u,s)A(-t_2)\cdot o$. Let $\hat{o} = (1, 0, \dots, 0, 1)^t \in \mathbb{K}^{n+1}$ then, thanks to Theorem 10.5.1, we get

$$\begin{aligned}
\cosh^2(d(\xi_{t_1}, \xi_{t_2})) &= \cosh^2(d((u,s;-t_1)\hat{o}, (u,s;-t_2)\hat{o})) \\
&= \cosh^2(d((0,0;-t_1)\hat{o}, (0,0;-t_2)\hat{o})) \\
&= \cosh^2(t_1 - t_2),
\end{aligned}$$

where, in particular, the first equality follows from the identification between $\mathbb{K}\mathbf{H}^n$ and $\mathbb{K}\mathcal{H}^n \rtimes \mathbb{R}$; the second equality from the left-invariance of d and the last one follows from the explicit formula of d on $\mathbb{K}\mathcal{H}^n \rtimes \mathbb{R}$. We obtain, thanks to the injectivity of cosh on the nonnegative real numbers, that $d(\xi_{t_1}, \xi_{t_2}) = |t_2 - t_1|$.

To prove the second and the third point we suppose $\mathbb{K} = \mathbb{H}$ so

$$h(u,s) = \begin{pmatrix} 1 & u^t & \frac{|u|^2}{2} + s_1 i + s_2 j + s_3 k \\ 0 & I_{n-1} & \overline{u} \\ 0 & 0 & 1 \end{pmatrix}.$$

If $\mathbb{K} = \mathbb{R}$ or $\mathbb{K} = \mathbb{C}$ one can simply consider $s_1 = s_2 = s_3 = 0$ or $s_2 = s_3 = 0$, respectively. A simple calculation proves that

$$\xi_t = h(u,s)A(-t)\cdot o = \begin{bmatrix} e^{-t} + e^t(\frac{|u|^2}{2} + s_1 i + s_2 j + s_3 k) \\ e^t \overline{u} \\ e^t \end{bmatrix}.$$

So, on one side, we get

$$\begin{aligned}
\lim_{t\to-\infty} \xi_t &= \lim_{t\to-\infty} \begin{bmatrix} e^{-t} + e^t(\frac{|u|^2}{2} + s_1 i + s_2 j + s_3 k) \\ e^t \overline{u} \\ e^t \end{bmatrix} \\
&= \lim_{t\to-\infty} \begin{bmatrix} 1 + e^{2t}(\frac{|u|^2}{2} + s_1 i + s_2 j + s_3 k) \\ e^{2t} \overline{u} \\ e^{2t} \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \omega.
\end{aligned}$$

On the other side, we get

$$\lim_{t\to+\infty}\xi_t = \lim_{t\to+\infty}\begin{bmatrix} e^{-t}+e^t(\frac{|u|^2}{2}+s_1i+s_2j+s_3k) \\ e^t\overline{u} \\ e^t \end{bmatrix}$$
$$= \lim_{t\to+\infty}\begin{bmatrix} e^{-2t}+\frac{|u|^2}{2}+s_1i+s_2j+s_3k \\ \overline{u} \\ 1 \end{bmatrix}$$
$$= \begin{bmatrix} \frac{|u|^2}{2}+s_1i+s_2j+s_3k \\ \overline{u} \\ 1 \end{bmatrix} = \xi_\infty.$$

It follows that $\xi_\infty \neq \omega$ for all ξ, and simple calculations can prove that $\xi_\infty \in \partial_\infty \mathbb{K}\mathbf{H}^n$ thus concluding the proof. □

Theorem 11.2.4. *The set $N\cdot o$ can be identified with $\partial_\infty\mathbb{K}\mathbf{H}^n \setminus \{\omega\}$.*

Proof. Let $\phi : N\cdot o \to \partial_\infty\mathbb{K}\mathbf{H}^n \setminus \{\omega\}$ such that $\phi(\xi_0)=\xi_\infty$ for all $\xi_0 \in N\cdot o$, where ξ_∞ is defined as in Lemma 11.2.3. The injectivity of ϕ follows from the proof of Lemma 11.2.3, while the surjetivity follows from simple calculations. □

We now know that $\partial_\infty\mathbb{K}\mathbf{H}^n \setminus \{\omega\}$ can be identified with $\mathbb{K}^{n-1}\times \mathrm{Im}(\mathbb{K})\times\{0\}$ endowed with the product

$$(u_1,s_1;0)(u_2,s_2;0) = (u_1+u_2, s_1+s_2+\mathrm{Im}(u_1^t\overline{u}_2);0).$$

Our next aim is to explicitly write the visual distance for the $\mathbb{K}$-hyperbolic n-space in the new coordinates given by Theorem 11.2.4.

Lemma 11.2.5. *The visual distance defined as in* (11.1.2) *is left-invariant, when $\partial_\infty\mathbb{K}\mathbf{H}^n$ is identified with $\mathbb{K}\times\mathrm{Im}(\mathbb{K})\times\{0\}$.*

Proof. Let $(u_1,s_1;0),(u_2,s_2;0)$,$(u_3,s_3;0) \in \mathbb{K}^{n-1}\times\mathrm{Im}(\mathbb{K})\times\{0\}$ the identification of $\xi_\infty,\eta_\infty,\mu_\infty \in \partial_\infty\mathbb{K}\mathbf{H}^n \setminus \{\omega\}$, respectively. We can therefore identify, thanks to Theorem 10.5.1, the points $\xi_t,\eta_t,\mu_t \in \mathbb{K}\mathbf{H}^n$, respectively, with $(u_1,s_1;-t)$,$(u_2,s_2;-t)$,$(u_3,s_3;-t) \in \mathbb{K}^{n-1}\times\mathrm{Im}(\mathbb{K})\times\mathbb{R}$. Thanks to Theorem 10.5.1 we also know that d is left-invariant so

$$\begin{aligned}
&((u_3,s_3;0)(u_1,s_1;0),(u_3,s_3;0)(u_2,s_2;0))_{(\omega,o)}\\
&\qquad= \frac{1}{2}\lim_{t\to+\infty}(2t - d((u_3,s_3;0)(u_1,s_1;-t),(u_3,s_3;0)(u_2,s_2;-t)))\\
&\qquad= \frac{1}{2}\lim_{t\to+\infty}(2t - d((u_1,s_1;-t),(u_2,s_2;-t)))\\
&\qquad= ((u_1,s_1;0),(u_2,s_2;0))_{(\omega,o)},
\end{aligned}$$

where the first and last equalities are definitions and the last one follows from the fact that d is left-invariant. We have therefore obtained

$$d_{vis}((u_3, s_3; 0)(u_1, s_1; 0), (u_3, s_3; 0)(u_2, s_2; 0)) = d_{vis}((u_1, s_1; 0), (u_2, s_2; 0)). \qquad \square$$

Theorem 11.2.6. *The visual distance on $\mathbb{K}^{n-1} \times \mathrm{Im}(\mathbb{K}) \times \{0\}$ reads as*

$$d_{vis}(\mathbf{0}, (u, s; 0)) = \sqrt[4]{\frac{|u|^4}{4} + |s|^2}.$$

Proof. We need to calculate $(\mathbf{0}, (u, s; 0))_{(\omega,o)}$. By definition

$$(\mathbf{0}, (u, s; 0))_{(\omega,o)} = \frac{1}{2} \lim_{t\to+\infty} (2t - d((0, 0; -t), (u, s; -t))).$$

Thanks to Theorem 10.5.1 we know that

$$\begin{aligned} d((0, 0; -t), (u, s; -t)) &= \operatorname{arccosh} \sqrt{e^{2t} \frac{|u|^2}{2} + e^{4t} \left(\frac{|u|^4}{16} + \frac{|s|^2}{4} \right)} \\ &= \operatorname{arccosh} \left(e^{2t} \sqrt{\beta(u, s, t)} \right), \end{aligned}$$

where

$$\beta(u, s, t) := \sqrt{e^{-2t} \frac{|u|^2}{2} + \left(\frac{|u|^4}{16} + \frac{|s|^2}{4} \right)}.$$

We recall that

$$x \geq 1 \implies \operatorname{arccosh} x = \ln \left(x^2 + \sqrt{x^2 - 1} \right),$$

so

$$\begin{aligned} d((0, 0; -t), (u, s; -t)) &= \operatorname{arccosh} \left(e^{2t} \sqrt{\beta(u, s, t)} \right) \\ &= \ln \left(e^{2t} \beta(u, s, t) + \sqrt{\left(e^{2t} \sqrt{\beta(u, s, t)} \right)^2 - 1} \right) \\ &= \ln \left(e^{2t} \beta(u, s, t) + \sqrt{e^{4t} \beta(u, s, t)^2 - 1} \right) \\ &= \ln \left(e^{2t} \left(\beta(u, s, t) + \sqrt{\beta(u, s, t)^2 - \frac{1}{e^{4t}}} \right) \right) \\ &= 2t + \ln \left(\beta(u, s, t) + \sqrt{\beta(u, s, t)^2 - \frac{1}{e^{4t}}} \right), \end{aligned}$$

where the second equality follows from the definition of arccosh, the last one follows from the properties of logarithms, and the other ones are simple calculations. We can now write the Gromov product as follows:

$$
\begin{aligned}
(\mathbf{0},(u,s;0))_{(\omega,o)} &= \frac{1}{2}\lim_{t\to+\infty}\left(2t - d((0,0;-t),(u,s;-t))\right)\\
&= \frac{1}{2}\lim_{t\to+\infty}\left(2t-2t-\ln\left(\beta(u,s,t)+\sqrt{\beta(u,s,t)^2-\frac{1}{e^{4t}}}\right)\right)\\
&= \frac{1}{2}\lim_{t\to+\infty}\left(-\ln\left(\beta(u,s,t)+\sqrt{\beta(u,s,t)^2-\frac{1}{e^{4t}}}\right)\right).
\end{aligned}
$$

By observing that

$$\lim_{t\to+\infty}\beta(u,s,t)=\sqrt{\frac{|u|^4}{16}+\frac{|s|^2}{4}},$$

we obtain

$$
\begin{aligned}
(\mathbf{0},(u,s;0))_{(\omega,o)} &= \frac{1}{2}\lim_{t\to+\infty}\left(-\ln\left(\beta(u,s,t)+\sqrt{\beta(u,s,t)^2-\frac{1}{e^{4t}}}\right)\right)\\
&= -\frac{1}{2}\ln\left(2\sqrt{\frac{|u|^4}{16}+\frac{|s|^2}{4}}\right) = -\ln\left(\sqrt[4]{\frac{|u|^4}{4}+|s|^2}\right).
\end{aligned}
$$

Thanks to the definition (11.1.2) of d_{vis} we can conclude

$$d_{vis}(\mathbf{0},(u,s;0)=e^{-(\mathbf{0},(u,s;0))_{(\omega,o)}}=\sqrt[4]{\frac{|u|^4}{4}+|s|^2}. \qquad \square$$

11.3 Heintze groups*

Chapter 12

Large-scale geometry of nilpotent groups*

12.1 Elements of Geometric Group Theory*

A *discrete* group Γ is a topological group that as topological space is discrete.

A set S inside a group Γ is said to be *generating* if there is no proper subgroup of Γ containing S. In other words, every element in the group Γ can be written as a finite product of elements in S. If one interprets the elements in S as words of an alphabet, then one can use the expression: 'each element in Γ is represented by a word with letters in S'.

A group is said to be *finitely generated* if it admits a finite generating set.

After having fixed such a set S, one can construct a geometric graph related to the group Γ.

Definition 12.1.1 (Cayley graph)**.** Let Γ be a discrete group and let S be a generating set. The *(colored and directed) Cayley graph* $\mathcal{G} = \mathcal{G}(\Gamma, S)$ is the colored directed graph constructed as follows: The vertex set $\mathrm{Vertex}(\mathcal{G})$ of $\mathcal{G}$ is identified with Γ. Each generator s of S determines a color c_s and the directed edges of color c_s consists of the pairs of the form (g, gs), with $g \in \Gamma$.

Geometric Group Theory mostly studies finitely generated groups considering the large scale geometry (or coarse geometry) of the Cayley graph. In such case, the set S is usually assumed to be finite, symmetric, i.e., $S = S^{-1}$, and not containing the identity element of the group. In this case, the (uncolored) Cayley graph is an ordinary graph: its edges are not oriented and it does not contain loops.

Definition 12.1.2 (Word metric)**.** Let Γ be a discrete group and let S be a generating set. For every two elements g and $h \in \Gamma$, their *word distance* with respect to S, is denoted by $d_S(g, h)$ and is defined as the minimum number of elements (=letters) in S whose product (=word) equals $g^{-1}h$. Analogously, the *word metric* d_S on the whole Cayley graph $\mathcal{G}(\Gamma, S)$ is the length metric that gives length 1 to each edge of $\mathcal{G}(\Gamma, S)$. We have then an isometry between (Γ, d_S) and the vertex set of the graph $(\mathrm{Vertex}\,(\mathcal{G}(\Gamma, S)), d_S)$

The group Γ acts naturally on its Cayley graph $\mathcal{G}(\Gamma, S)$ sending the vertex h to the vertex gh, for each fixed $g \in \Gamma$. One can easily check that such left translations preserve the graph structure of $\mathcal{G}$.

Proposition 12.1.3 (Isometry of the left action)**.** *The left translation of a group* Γ *are isometries with respect to the word metric. Analogously, the left translations induce an isometric action of the group* Γ *on the metric space* $(\mathcal{G}(\Gamma, S), d_S)$*, and such action is transitive on the vertex set.*

The word metric on a group Γ is not unique, because different symmetric generating sets give different word metrics. However, finitely generated word metrics are unique up to biLipschitz equivalence.

Proposition 12.1.4 (Bilipschitz invariants of a group)**.** *If* S *and* S' *are two symmetric, finite generating sets for* Γ *with corresponding word metrics* d_S *and* $d_{S'}$*, then there is a constant* K *such that the identity map from* (Γ, d_S) *to* $(\Gamma, d_{S'})$ *is a* K*-biLipschitz map. In fact,* K *is just the maximum of the* d_S *word norms of elements of* S' *and the* $d_{S'}$ *word norms of elements of* S*.*

Definition 12.1.5 (Quasi-isometry)**.** Suppose (M_1, d_1) and (M_2, d_2) are metric spaces, and $f : M_1 \to M_2$ is a function (not necessarily continuous). Then f is called a (A, L)-*quasi-isometric embedding*, with $L \geq 1$ and $A \geq 0$, if

$$\frac{1}{L}\, d_2(f(x), f(y)) - A \leq d_1(x, y) \leq L\, d_2(f(x), f(y)) + A \quad \text{for all} \quad x, y \in M_1.$$

Moreover, a quasi-isometric embedding is called a *quasi-isometry* if there exists a constant $C \geq 0$ such that to every $u \in M_2$ there exists $x \in M_1$ with

$$d_2(u, f(x)) \leq C.$$

The spaces M_2 and M_2 are called *quasi-isometric* if there exists a quasi-isometry between them.

Theorem 12.1.6 (Fundamental observation of Geometric Group Theory)**.** *Let* X *be a metric space which is geodesic and proper, let* Γ *be a group acting on* X *by isometries. Assume that the action is proper and the quotient space* X/Γ *is compact.*

Then the group Γ *is finitely generated and quasi-isometric to* X.

More precisely, for every $x_0 \in X$, *the orbit mapping*

$$\Gamma \to X$$

$$\gamma \mapsto \gamma(x_0)$$

is a quasi-isometry.

Such fact was known in the 50's. A proof can be essentially re-contructed from [Lemma 2]Milnor. A detailed proof is in [Theorem 23]delaharpe.

From the above fundamental observation we deduce that Geometric Group Theory links the study of fundamental groups of compact manifolds and their Riemannian universal covers. Namely, let M be a compact differentiable manifold. Let $\pi_1(M)$ the fundamental group of M. By the above observation, such discrete group is finitely generated. We endow the group with a word metric. Fix now a Riemannian metric g on M. Then there is a unique Riemannian metric $\tilde{g}$ on the universal cover $\tilde{M}$ of M such that the universal projection

$$(\tilde{M}, \tilde{g}) \twoheadrightarrow (M, g)$$

is a local isometry. We refer to such a $\tilde{g}$ as the lifted Riemannian metric. The crucial result is that the coarse geometry of $\tilde{M}$ is the same that the coarse geometry of $\pi_1(M)$. A prove of the following proposition can be found in the lecture notes of M. Kapovich on GGT, use his Lemma 1.31.

Proposition 12.1.7. *Assume* M *is a Riemannian manifold that is compact.*

(i) *The fundamental group* $\pi_1(M)$ *is finitely generated.*

(ii) *The universal cover* $\tilde{M}$, *endowed with the lifted Riemannian distance, is quasi-isometric to* $\pi_1(M)$, *endowed with any word metric.*

Proposition 12.1.8. *Assume* G *is a finitely generated group and* $H < G$ *a subgroup.*

(i) *If* H *has finite index in* G, *then* G *and* H *are quasi-isometric.*

(ii) *If* H *is a finite group and it is normal in* G, *then* G *and* G/H *are quasi-isometric.*

Definition 12.1.9. We say that a group G is *virtually nilpotent* if there exists a sub-group $H < G$ of finite index in G that is nilpotent.

12.2 Growth rates of balls*

The bilipschitz equivalence of word metrics implies in turn that the growth rate of a finitely generated group is a well-defined isomorphism invariant of the group Γ, independent of the choice of a finite generating set S. This implies in turn that various properties of growth, such as polynomial growth, the degree of polynomial growth, and exponential growth, are isomorphism invariants of groups.

Given a finitely generated group Γ, we fix a finite symmetric generating set S. For each $R > 0$, let $B_S(1_\Gamma, R)$ be the metric ball in Γ with respect the distance d_S with center the origin 1_Γ and radius R. We then denote by $\#(B_S(1_\Gamma, R))$ the cardinality of the finite set $B_S(1_\Gamma, R)$.

Definition 12.2.1. The *growth rate* of a finitely generated group Γ is the growth rate of the function $R \mapsto \#(B_S(1_\Gamma, R))$.

12.2.1 Invariance of the growth rate

Proposition 12.2.2. *If two metric spaces are quasi isometric, then they have the same growth rate.*

Corollary 12.2.3. *Assume M is a Riemannian manifold that is compact. Then the grow rate of the group $\pi_1(M)$ is the same as the grow rate of the volume function on the universal cover of M.*

Namely, consider the Riemmanian structure on $\tilde{M}$ lifted from the structure on M. Let $\tilde{B}(p, r)$ be the metric ball in $\tilde{M}$. Let $\mathrm{vol}_{\tilde{M}}$ be the Riemmanian volume form on $\tilde{M}$. Then the above corollary states that there exist constants k, c such that, for all $R > 1$, one has the bounds

$$k^{-1}\ \#(B_S(1, c^{-1}R)) \leq \mathrm{vol}_{\tilde{M}}(\tilde{B}(p, R)) \leq k\ \#(B_S(1, cR)).$$

Now, if a group Γ is virtually nilpotent, then by definition it has a nilpotent sub-group Γ' of finite index. Then Γ and Γ' are quasi-isometric and thus have the same growth rate. We will describe the fact that the groups that are virtually nilpotent are exactly those that have a polynomial growth rate.

12.2.2 Polynomial growth and virtual nilpotency

Definition 12.2.4 (Polynomial growth)**.** A discrete group Γ is said to have *polynomial growth* if, for some (and thus for any) generating set S, there exist $C > 0$ and $k > 0$ such that for every integer $R \geq 1$

$$\#(B_S(1_\Gamma, R)) \leq C \cdot R^k.$$

Another choice for S would only change the constant C, but not the polynomial nature of the bound, because of Proposition 12.2.2. Actually one only requires that the growth of the balls are bounded by a polynomial function. However, a result of Pansu states that, in fact, the above equation can be improved saying that there exists $c(S) > 0$ and an integer $d(\Gamma) \geq 0$ depending on Γ only such that the following holds:

$$\#(B_S(1_\Gamma, R)) = c(\Gamma)R^{d(\Gamma)} + o(R^{d(\Gamma)}), \qquad \text{as } R \to \infty.$$

The condition of polynomial growth can be further weakened, cf. [vdDW84, Kle10].

A result of J. Wolf is that a group has polynomial growth if it is nilpotent. A deep result of Gromov is the equivalence of polynomial growth and virtual nilpotency.

Theorem 12.2.5 (Gromov's polynomial growth). *A finitely generated group has polynomial growth rate if and only if it is virtually nilpotent.*

The original proof in [Gro81] is based on Gleason-Montgomery-Zippin-Zippin-Yamabe structure theory of locally compact groups. A new short proof has been given by Kleiner in [Kle10].

A non trivial consequence of Gromov's theorem is that if a group has polynomial growth then the exponent of the growth rate is an integer. The plan of this chapter is to give an exposition of how sub-Riemannian geometry plays a role in the polynomial growth theorem and observe that such integer exponent is in fact the Hausdorff dimension of a Carnot group associated to the finitely generated group.

12.3 Asymptotic cone*

Theorem 12.3.1 (Wolf-Bass-Gromov-Pansu). *The degree of growth of a finitely generated group Γ of polynomial growth is an integer and equals the Hausdorff dimension of the Carnot group that is the asymptotic cone of Γ.*

[1]

The asymptotic cone, also known as the tangent cone at infinity, is similar to the tangent cone (at a point), except that instead of performing a blow-up procedure, we 'blow down'.

Definition 12.3.2 (Asymptotic cone). An asymptotic cone of a metric space (X, d) is a metric space (Z, ρ) with the property that there is $\bar{x} \in Z$ and, for each $j \in \mathbb{N}$, there are $x_j \in X$ and $\epsilon_j > 0$,

[1] Put here the Witt's formula. See also M. Hall.

with $\epsilon_j \to 0$ as $j \to \infty$, such that for each $R > 0$ there are $\delta_j \geq 0$, with $\delta_j \to 0$ as $j \to \infty$, with the property that

$$\mathrm{GH}\lim_{j\to\infty} B^{(X,\epsilon_j d)}(x_j, R+\delta_j) = B^{(Z,\rho)}(\bar{x}, R),$$

i.e., the sequence of balls in X with respect to the 'compressed' metric $\epsilon_j d$ with centers x_j and radii $R+\delta_j$ converges, in the Gromov Hausdorff sense, to the ball in (Z,ρ) with center $\bar{x}$ and radius R.

Proposition 12.3.3. *Two quasi-isometric spaces have the same class of asymptotic cones.*

Theorem 12.3.4 (Pansu [Pan83a])**.** *The asymptotic cone of a nilpotent Lie group G, endowed with a left-invariant geodesic distance, is a Carnot group G_∞ endowed with a left-invariant sub-Finsler structure. The Hausdorff dimension of G_∞ is the exponent of the growth rate of Γ.*

12.4 Malcev closure*

We shall explain now the connection between polynomial growth and sub-Riemannian geometry. We shall see how a nilpotent finitely generated discrete group is coarsely equivalent to a sub-Finsler Lie group. First we need to understand how such a discrete group is coarsely seen as a Lie group. Malcev Theorem 12.4.4 is the core of the argument.

Briefly, a lattice is a discrete subgroup with finite covolume. Here is the formal definition:

Definition 12.4.1 (Lattice)**.** Let G be a locally compact topological group. A subgroup $\Gamma < G$ is a *lattice* if it is discrete (as topological subspace) and has the property that on the quotient space G/Γ there is a finite G-invariant[2] measure.

Proposition 12.4.2. *Let G be a Lie group endowed with a left-invariant Riemannian metric. Let Γ be a lattice in G. Then the quotient G/Γ is in fact compact and thus Γ is quasi-isometric to G.*

Theorem 12.4.3 ([Rag72, Theorem 2.18])**.** *A group Γ is isomorphic to a lattice in a simply connected nilpotent Lie group if and only if*

1. *Γ is finitely generated,*

2. *Γ is nilpotent, and*

3. *Γ has no torsion.*

[2]Recall that the quotient on G/Γ is on the right, so G acts naturally on the left.

Corollary 12.4.4 (Malcev Theorem [Mal51]). *If Γ is a finitely generated group which is nilpotent and has no torsion then it is isomorphic to a discrete cocompact subgroup of a simply connected nilpotent Lie group G.*

Some useful facts:

1. Every subgroup of a nilpotent group is nilpotent. (easy!)

2. Every subgroup of a finitely generated nilpotent group is finitely generated, cf. [Theorem 9.16]Macdonalds-theory of groups or [Rag72, Theorem 2.7].

3. Every nilpotent group generated by finitely many elements of finite order is finite, cf. [Theorem 9.17]Macdonalds.

These facts implies the following:

Lemma 12.4.5 (on torsion of finitely generated nilpotent groups). *The elements of finite order in a nilpotent group G form a normal sub-group $Tor(G)$, called the torsion sub-group of G. If G is finitely generated, $Tor(G)$ is finite. The quotient $G/\mathrm{Tor}(G)$ is torsion-free, that is, its only element of finite order is the identity.*

Proposition 12.4.6. *Let Γ be a finitely generated discrete group Γ of polynomial growth, then Γ is quasi-isometric to a connected, simply connected, and nilpotent Lie group G.*

If a group Γ has polynomial growth, then, by Gromov Theorem 12.2.5, there is a subgroup $\Gamma_1 < \Gamma$ that is nilpotent and $[\Gamma, \Gamma_1] < \infty$. Let $\mathrm{Tor}(\Gamma_1)$ be the torsion of Γ_1, which is a finite and normal subgroup, by Lemma 12.4.5. Define $\Gamma_2 := \Gamma_1/\mathrm{Tor}(\Gamma_1)$. Then Γ_2 is nilpotent and has no torsion, thus, by Malcev Theorem 12.4.4, there is a connected, simply connected, and nilpotent Lie group G and a discrete cocompact subgroup $\Gamma' < G$, such that Γ_2 is isomorphic to Γ'.

The groups Γ, Γ_1, Γ_2, Γ', and G are quasi-isometric.

12.5 The limit CC metric*

Let Γ' be a discrete cocompact sub-group in a connected, simply connected, and nilpotent Lie group G. Let G_∞ be the unique connected, simply connected Lie group whose Lie algebra is the graded algebra $\mathfrak{g}_\infty$ of $\mathfrak{g}$.

Let $\|\cdot\| := d_S(1_\Gamma, \cdot)$ be a 'norm' on Γ' induced by a finite generating set S. We shall describe the CC metric induced on the Carnot group G_∞.

Consider the two sets:

$$A := \Gamma'/[\Gamma', \Gamma'] \qquad \text{and} \qquad B := G/[G, G].$$

Both A and B are Abelian groups. Moreover, B is a (finite-dimensional) vector space.

A is a subgroup of B. (?!?)

$\|\cdot\|$ induces a norm on A. (?!?)

One defines

$$\|a\|_\infty := \lim_{k\to\infty} \frac{1}{k} \|ka\| .$$

Such norm extends to B. (?!?)

Recall that, as in every Carnot group, we have that $V_1 \simeq \mathfrak{g}/[\mathfrak{g}, \mathfrak{g}]$. Thus we consider the projection

$$\pi : G_\infty \twoheadrightarrow G_\infty/[G_\infty, G_\infty].$$

Therefore we can transport the norm on V_1, using the isomorphism between V_1 and $B := G/[G, G]$. (?!?)

12.6 Proof of Pansu Asymptotic Theorem*

[...]

Chapter 13

Open problems in Geometry and Analysis on Carnot groups*

13.1 Regularity problems*

We will discuss the following issues:

- smoothness of geodesic curves;
- smoothness of metric spheres;
- smoothness (and existence) of minimal surfaces;
- smoothness (and existence) of solution of the isoperimetric problem.

Comments regarding geodesics

1. The existence is ensured by Ascoli-Arzelà Theorem, as a priori just Lipschitz curves, so differentiable almost everywhere.

2. People expect that when $(M, \Delta, \langle \cdot, \cdot \rangle)$ is a sub-Riemannian manifold, then every geodesic is C^1, or, in fact, C^∞. The question is still open.

3. People expect that when $\|\cdot\|$ is a norm coming from a polytope, i.e., the unit ball of $\|\cdot\|$ is the convex hull of finitely many points, then there exists a constant $N \in \mathbb{N}$ such that each pair of points can be connected with a geodesic made of N smooth pieces. The question is still open.

4. The query cannot be solved using the standard arguments from geometric analysis (e.g., Calculus of Variation or differential geometry) as in Riemannian geometry.

Comments regarding metric spheres

1. In Carnot groups, metric spheres are topological spheres. (In general, the conjecture is that small metric spheres are topological spheres.)

2. In the Heisenberg geometry, spheres are not smooth at the pole. See the picture of the section of the ball.

3. The expectation is that small metric spheres (at least in Carnot groups) should be piecewise smooth.

4. The regularity of geodesics is linked (at least philosophically) to the regularity of metric spheres.

Comments regarding minimal surfaces and isoperimetric solutions

1. They do exists in an extended sense.

2. Regularity is a tricky issue.

13.1.1 Common general philosophical strategy for regularity

Step 1 Consider the geometric objects as special elements inside a wider class of analytical objects.

Step 2 Prove that such analytical objects are in fact 'rectifiable', e.g., 'piece-wise Lipschitz'. (Here there will be an issue since Carnot groups are purely unrectifiable.)

Step 3 Rectifiability should be first improved as low (e.g., C^1) regularity, for example in the case of minimal objects.

Step 4 Minimal C^1 (or C^2) objects are in fact C^∞, or even analytic.

13.2 Generalized hyper-surfaces: sets with finite perimeter*

Both metric spheres and $(n-1)$-dimensional minimal surfaces inside an n-dimensional Carnot group have codimension 1. We can see them as boundary of an n-dimensional domain Ω. We then think about studying Ω instead $\partial\Omega$. The idea is to consider the characteristic function χ_Ω of Ω:

$$\chi_\Omega(x) = 1 \text{ if } x \in \Omega, \chi_\Omega(x) = 0 \text{ if } x \notin \Omega.$$

We consider the wide class of all measurable sets Ω, in other words, we have $\chi_\Omega \in L^1_{loc}$.

Which are the good χ_Ω? Clearly, even a request of continuity is too strong. The feeling is that if Ω is a hyper-space, then χ_Ω should be good. As an toy example, let us consider the I-don't-remember-the-name function, i.e., $\chi_{\mathbb{R}_{>0}}$. A nice property of such a function is that its derivative exists in the generalized sense, it is the delta measure δ_0.

We arrive at the conclusion that "our good sets are those whose characteristic functions have measures as generalized derivatives." We should explain in the following what is this generalized derivative.

13.2.1 A review of divergence and distributions

Let M be a smooth differentiable manifold with topological dimension n, endowed with an n-differential volume form vol_M. For example, vol_M could be a Riemannian volume form; however, eventually, M will be a Lie group $\mathbb{G}$, and vol_M a right Haar measure.

We use the volume form to define the divergence as follows:

Definition 13.2.1. For every vector field $X \in \Gamma(M)$ define the function $\operatorname{div} X : M \to \mathbb{R}$ implicitly as

$$\int_M Xu \, d\,\mathrm{vol}_M = -\int_M u \operatorname{div} X \, d\,\mathrm{vol}_M \qquad \forall u \in C_c^\infty(M). \tag{13.2.2}$$

We say that X is divergence-free if $\operatorname{div} X \equiv 0$.

For example the vector fields $\dfrac{\partial}{\partial_j}$ in $\mathbb{R}^n$ are divergence-free, because of the Fundamental Theorem of Calculus and the fact that the test functions have compact support.

When (M, g) is a Riemannian manifold and vol_M is the volume form induced by g, then an explicit expression of this differential operator can be obtained in terms of the components of X, and (13.2.2) corresponds to the divergence theorem on manifolds. We won't need either a Riemannian structure or an explicit expression of $\operatorname{div} X$ in the sequel, and for this reason we have chosen a definition based on (13.2.2): this emphasizes the dependence of $\operatorname{div} X$ on vol_M only.

Note that by Leibniz rule $X(uv) = uXv + vXu$, integrating over the manifold when X is a divergence-free vector field, one obtains

$$\int_M uXv \, d\,\mathrm{vol}_M = -\int_M vXu \, d\,\mathrm{vol}_M \qquad \forall u, v \in C_c^\infty(M). \tag{13.2.3}$$

This last identity motivates the following classical definition.

Definition 13.2.4 (X-distributional derivative)**.** Let $u \in L^1_{\text{loc}}(M)$ and let $X \in \Gamma(TM)$ be divergence-free. The generalized derivative of u in the direction of X is the operator $Xu \in (C^\infty_c(M))^*$ defined as

$$\langle Xu, v\rangle := -\int_M uXv \, d\,\mathrm{vol}_M, \qquad v \in C^\infty_c(M).$$

If $f \in L^1_{\text{loc}}(M)$, we write $Xu = f$ if $\langle Xu, v\rangle = \int_M vf \, d\,\mathrm{vol}_M$ for all $v \in C^\infty_c(M)$. Analogously, if μ is a Radon measure in M, we write $Xu = \mu$ if $\langle Xu, v\rangle = \int_M v\, d\mu$ for all $v \in C^\infty_c(M)$.

Since X is divergence-free and so (13.2.3) holds (it is still valid when $u \in C^1(M)$), the distributional definition of Xu is equivalent to the classical one whenever $u \in C^1(M)$.

Proposition 13.2.5. (i) *Let $\mathbb{G}$ be a nilpotent Lie group, and let $\mathrm{vol}_{\mathbb{G}}$ be a right Haar measure. Then each left invariant vector field is divergence-free.*

(ii) *More generally, for each manifold M, each volume form vol_M), and each $X \in \Gamma(M)$, one has that if the flows of X are vol_M-preserving, then $\operatorname{div} X \equiv 0$.*

Proof. The first assertion is consequence of the second one, since, as we saw, flows of left invariant vector fields are right translations, $g \mapsto ge^{tX}$. Regarding (ii), let $\Phi^t_X(\cdot)$ be the flow of X at time t. Thus we know that, for every t, we have

$$\left(\Phi^t_X\right)_\# \mathrm{vol}_M = \mathrm{vol}_M .$$

Therefore, for every test function u, $\int_M u \circ \Phi^t_X \, d\,\mathrm{vol}_M = \int_M u \, d\,\mathrm{vol}_M$. Such independence of t implies that

$$\begin{aligned}
-\int_M u \operatorname{div} X \, d\,\mathrm{vol}_M &= \int_M Xu \, d\,\mathrm{vol}_M \\
&= \int_M (Xu) \circ \Phi^t_X \, d\,\mathrm{vol}_M \\
&= \int_M \frac{\mathrm{d}}{\mathrm{d}t} \left(u \circ \Phi^t_X\right) \, d\,\mathrm{vol}_M \\
&= \frac{\mathrm{d}}{\mathrm{d}t} \int_M u \circ \Phi^t_X \, d\,\mathrm{vol}_M \\
&= 0.
\end{aligned}$$

Therefore $\int_M u \operatorname{div} X \, d\,\mathrm{vol}_M = 0$ for all $u \in C^1_c(M)$, and X is divergence-free. $\square$

One can prove the inverse implication: the flows are vol_M-measure preserving if $\operatorname{div} X$ is equal to 0, cf. the proof of Theorem 2.12 in [AKL09].

13.2.2 Caccioppoli sets: sets of locally finite perimeter

Definition 13.2.6 (Sets of locally finite perimeter)**.** A Borel set E in a Carnot group, with stratification $\mathfrak{g} = V_1 \oplus \cdots \oplus V_s$, is said a *Caccioppoli set* or to have *locally finite perimeter* if, for every left invariant horizontal vector field $X \in V_1$, the distribution $X\chi_E$ is a Radon measure.

Now that we generalized the object of study, we should first understand how to obtain back our hyper-surfaces.

Pick $X_1, \ldots, X_m$ a basis of V_1. We form the $\mathbb{R}^m$-valued Radon measure

$$D\chi_E := (X_1\chi_E, \ldots, X_m\chi_E), \tag{13.2.7}$$

and call it the *perimeter vector measure*. One can write

$$D\chi_E = \nu_E |D\chi_E|,$$

where $|D\chi_E|$ is the (positive) measure given by the variation of $D\chi_E$: if A is a Borel set, then

$$|D\chi_E|(A) = \sup_{\pi} \sum_{B\in\pi} \|D\chi_E(B)\|,$$

where the supremum is taken over all partitions π of A into a finite number of disjoint measurable subsets. And ν_E is the vector measurable function obtained as

$$\nu_E(x) := \lim_{r\downarrow 0} \frac{D\chi_E(B_r(x))}{|D\chi_E|(B_r(x))},$$

which exists $|D\chi_E|$-almost everywhere.

The terminology is that $|D\chi_E|$ is the *perimeter measure*, and ν_E is the *normal* of the set. Finally,

$$\mathrm{Per}(E) := |D\chi_E|(\mathbb{G}) \tag{13.2.8}$$

is the *perimeter* of E. More generally, if Ω is a Borel set, then $\mathrm{Per}(E, \Omega) := |D\chi_E|(\Omega)$ is the perimeter of E inside Ω.

All such objects depend on the choice of $X_1, \ldots, X_m$. The choice of such a basis is in correspondence to the choice of a sub-Riemannian metric on the Carnot group $\mathbb{G}$, for which $X_1, \ldots, X_m$ is an orthonormal basis.

Definition 13.2.9 (De Giorgi's reduced boundary)**.** Let $E \subseteq \mathbb{G}$ be a set of locally finite perimeter. Define the reduced boundary $\mathscr{F}E$ as the set of points $x \in \operatorname{supp} |D\chi_E|$ where:

(i) the limit defining ν_E exists and

(ii) $|\nu_E(x)| = 1$.

E.g., the reduced boundary of a square on the (Euclidean) plane is formed by its four edges with the four vertices removed.

Why it is better to consider such sets? Because in such class minima always exist.

Theorem 13.2.10 (Compactness [GN96] + Lower semicontinuity for BV functions [FSS96])**.** *Let $\mathbb{G}$ be a Carnot group and let E_j be a sequence of locally finite perimeter sets such that their perimeters in some Borel set Ω converge to a value $c \in \mathbb{R}$, i.e.,*

$$|D\chi_{E_j}|(\Omega) \to c.$$

Then there exists a locally finite perimeter set F such that, up to passing to a subsequence,

1. $\chi_{E_j} \to \chi_F$ *in* $L^1_{loc}(\Omega)$ *and*

2. $|D\chi_F|(\Omega) \leq c$.

13.2.3 Notions of rectificability

In general metric spaces the classical definition of 'good' surfaces goes back at least to Federer (see [Fed69, 3.2.14]). The 'good' surfaces are those that are images of open subsets in Euclidean spaces via Lipschitz maps.

However, there is a problematic fact: in the Heisenberg group there are no Lipschitz embedding of an open set $U \subset \mathbb{R}^2$ into the group. Indeed, differentiability theorems implies that the Heisenberg group is 2-purely unrectifiable, cf. [AK00, Theorem 7.2]. This means that each Lipschitz map $f : U \subset \mathbb{R}^2 \to \mathbb{G}$ is such that $\mathcal{H}^2(f(U)) = 0$. Roughly speaking, since the 3D Heisenberg group has Hausdorff dimension equal to 4, then the metric dimension of a hyper-surface is espected to be $4 - 1 = 3$. But the image by a Lipschitz map of a 2-dimensional Euclidean set has Hausdorff dimension no greater than 2.

There is a second notion (cf. [FSS03, FSS01]) of good surfaces which is only valid for hyper-surfaces: being (locally) the zero set of a 'intrinsically' C^1 real-valued function with non-vanishing gradient:

Definition 13.2.11 ($\mathbb{G}$-regular functions and hyper-surfaces)**.** Let $\mathbb{G}$ be a Carnot group with V_1 as horizontal layer. Let U be an open subset of $\mathbb{G}$ and $f : U \to \mathbb{R}$. We say that f belongs to $C^1_{\mathbb{G}}(U)$

if f and X_f are continuous functions in U, for all $X \in V_1$. We say that $S \subset \mathbb{G}$ is a $\mathbb{G}$-*regular hyper-surface* if for every $p \in S$ there is an neighborhood U of p in $\mathbb{G}$ and there is $f \in C^1_{\mathbb{G}}(U)$ with $(Xf)(q) \neq 0$, for all $q \in U$ and all $X \in V_1 \setminus \{0\}$, such that

$$S \cap U = f^{-1}(0).$$

Notice that if f is in C^1 then it is clearly in $C^1_{\mathbb{G}}$. However, the hyper-surface $f^{-1}(0)$ is $\mathbb{G}$-regular only if ∇f is never orthogonal to V_1.

Definition 13.2.12 ($\mathbb{G}$-rectifiable hyper-surface)**.** Let $\mathbb{G}$ be a Carnot group of Hausdorff dimension Q. A set $\Sigma \subset \mathbb{G}$ is said ($(Q-1)$-dimensional) $\mathbb{G}$-*rectifiable* if there exist a countable collection of $\mathbb{G}$-regular hyper-surfaces S_j such that

$$\mathcal{H}^{Q-1}_{cc}(\Sigma \setminus \cup_j S_j) = 0.$$

The following theorem is due to De Giorgi in the Euclidean setting and to Franchi, Serapioni, and Serra Cassano in Carnot groups of step 2, cf. [DG54, DG55, FSS03, FSS01].

Theorem 13.2.13 (Structure of finite perimeter sets)**.** *Let $\mathbb{G}$ be either the Euclidean space or a step-2 Carnot group. If E has locally finite perimeter, then its reduced boundary $\mathscr{F}E$ is $\mathbb{G}$-rectifiable.*

Question 13.2.14. *Is the above theorem true in Carnot groups of arbitrarily step?*

A partial answer to the above question has been obtained in [AKL09].

13.2.4 Notions of surface measures

We reach the conclusion that the problem of studying hyper-surfaces can be rephrased as the study of characteristic functions χ_E, focusing on their perimeter measures $|D\chi_E|$ and their reduce boundaries $\mathscr{F}E$. The reason for doing so is that perimeters have properties of compactness and lower semicontinuity, cf. Theorem 13.2.10.

For hyper-surfaces then we have that there are two natural notions of measures: $\mathcal{H}^{Q-1}_{cc}$ restricted to the hyper-surface or the perimeter of one of the side domains determined by the hyper-surface. People expect that the two notions should be related. For doing so, one should first prove rectifiability of reduced boundaries, cf. Question 13.2.14.

However, if $S = f^{-1}(0)$ is given as level set of a C^1 function f, the two measures are equal. Indeed, let $E = f^{-1}((-\infty, 0))$, so $\partial E = S$. Then

$$\begin{aligned} \operatorname{Per}(E) &= \\ &= \ldots \\ &= \mathcal{H}^{Q-1}_{cc} \llcorner_{\partial E} \end{aligned}$$

13.3 Partial regularity results and open questions*

13.3.1 Results on geodesics

The following theorem can be found in [Str86], however, in that paper the claim was wrongly stated in more generality. In fact, the proof was valid only for step-2 distributions. The paper has been corrected in [Str89].

Theorem 13.3.1 (Strichartz [Str89])**.** *If $(M, \Delta, \langle \cdot, \cdot \rangle)$ is a sub-Riemannian manifold of step-2, then each geodesic for the CC-distance is C^∞.*

The following theorem is proved more generally in [LM08], however the assumptions of step ≤ 4 and rank 2 are deeply used.

Theorem 13.3.2 (Leonardi-Monti)**.** *[LM08] If $\mathbb{G}$ is a Carnot group of step ≤ 4 and with 2-dimensional horizontal layer V_1, then each geodesic for the CC-distance is C^∞.*

The next result is proved in these notes.

Proposition 13.3.3. *Let G be a connected, simply connected, and nilpotent Lie group. Let $\Delta \subset \mathfrak{g}$ be a left-invariant sub-bundle such that*

$$\Delta \oplus [\mathfrak{g}, \mathfrak{g}] = \mathfrak{g}.$$

E.g., G could be a Carnot group. Then, if X is a left-invariant vector field in Δ, then $t \mapsto e^{tX}$ is a (smooth) geodesic with respect to the CC-distance of $(G, \Delta, \|\cdot\|)$, for every left-invariant norm $\|\cdot\|$.

The following theorem should be found in [Bre14].

Theorem 13.3.4 (Breuillard 2007)**.** *Let $\mathbb{G}$ be the 3D Heisenberg group. Let $\|\cdot\|_1$ be the ℓ^1 norm on V_1. Then the geodesics with respect to the CC-distance of $(G, V_1, \|\cdot\|_1)$ are made of at most 4 pieces of horizontal lines, i.e., each geodesic is the concatenation of at most 4 curves of the form $t \mapsto ge^{tX}$, with $g \in \mathbb{G}$ and $X \in V_1$.*

Conjecture 13.3.5 (Regularity conjecture for sub-Reimannian manifolds). *If $(M, \Delta, \langle \cdot, \cdot \rangle)$ is a sub-Riemannian manifold, then each geodesic for the CC-distance is C^∞.*

Conjecture 13.3.6 (Weak regularity conjecture for sub-Reimannian Carnot groups). *If $\mathbb{G}$ is a Carnot group, then each pair of points can be connected by a C^1 geodesic.*

Conjecture 13.3.7 (Regularity Conjecture for sub-Finsler Carnot groups). *If $(\mathbb{G}, V_1, \|\cdot\|_1)$ is a Carnot group where $\|\cdot\|_1$ is the ℓ^1 norm, then there exists a constant K such that each pair of points can be connected by a geodesic that is the concatenation of at most K horizontal lines.*

There are several statements that are true but for possibly a measure-zero collection of distributions. Compare the following result with Theorem 13.3.12.

Theorem 13.3.8 (Chitour-Jean-Trélat [CJT06]). *For generic sub-Riemannian structures $(M, \Delta, \langle \cdot, \cdot \rangle)$ of rank greater than or equal to 3, i.e., $\dim \Delta_p \geq 3$, for all $p \in M$, all geodesics for the CC-distance are C^∞.*

13.3.2 Results on metric spheres

Proposition 13.3.9. *If $\mathbb{G}$ is the 3D Heisenberg group, then each metric sphere $\partial B(1_{\mathbb{G}}, r)$, $r > 0$, is an (Euclidean) Lipschitz manifolds, and there are two points p_N and p_S (the two poles) such that $\partial B(1_{\mathbb{G}}, r) \setminus \{p_N, p_S\}$ is a C^∞ manifold.*

In the Carnot group setting, one can uses the dilations and the standard proof of the fact that open sets that are star-shaped are topological balls, to prove that metric balls in Carnot groups are topological balls. Moreover, the spheres can be written as graphs using 'inhomogeneous' spherical coordinates with respect to the dilations. Since metric spheres in CC-metrics are closed, one get the following result.

Proposition 13.3.10. *If $\mathbb{G}$ is a Carnot group, then each metric sphere $\partial B(1_{\mathbb{G}}, r)$, $r > 0$, is topologically a sphere.*

The following theorem should be found in [Bre14].

Theorem 13.3.11 (Breuillard 2007). *Let $\mathbb{G}$ be the 3D Heisenberg group. Let $\|\cdot\|_1$ be the ℓ^1 norm on V_1. Then the metric spheres of the sub-Finsler geometry of $(G, V_1, \|\cdot\|_1)$ are piece-wise analytical sub-variety.*

The work of Agracev and Gauthier [AG01] gives an piece-wise analytic answer in generic cases:

Theorem 13.3.12 (Agrachev-Gauthier)**.** *Generically, small balls in a sub-Riemannian manifold* $(M, \Delta, \langle\cdot,\cdot\rangle)$ *are sub-analytic if the rank of the distribution is* ≥ 3*.*

Conjecture 13.3.13. *If* $(M, \Delta, \|\cdot\|)$ *is a sub-Finsler manifold, then small metric spheres are piecewise smooth.*

Proposition 13.3.14. *Metric balls in Carnot groups are sets of finite perimeter and metric spheres are* $\mathbb{G}$*-rectifiable hyper-surfaces.*

13.3.3 Results on the isoperimetric problem

In studying minimal problems for hyper-surfaces inside a Carnot group $\mathbb{G}$ of Hausdorff dimension Q, it is more convenient to minimize the intrinsic perimeter of a class of sets $E \subset \mathbb{G}$ than the $(Q-1)$-dimensional Hausdorff measure of their boundaries.

Theorem 13.3.15 (Existence of isoperimetric sets)**.** *In every Carnot group, there exist solutions of the isoperimetric problem, i.e., sets minimizing the intrinsic perimeter among all measurable sets with prescribed volume measure.*

The above theorem is due, in the Carnot group setting to Leonardi and Rigot in [LR03], and it has been then generalized by Danielli, Garofalo, and Nhieu.

Proposition 13.3.16. *Metric spheres* $\partial B(1, r)$*,* $r > 0$*, in the Heisenberg group are not solutions of the isoperimetric problem.*

In [Pan82, Pan83b], Pierre Pansu draw attention on a class of sets which are called today *Pansu spheres*. Denote by $\mathbb{S}_\lambda$ the compact embedded surface of revolution, which is homeomorphic to a sphere, obtained considering a geodesic between two points in the center of the group at distance π/λ and rotating such a curve around the center. Every left translation of an $\mathbb{S}_\lambda$ is called a Pansu sphere.

Ritoré and Rosales arrived at a characterization of complete, oriented, connected C^2 immersed volume preserving area-stationary surfaces in the 3D Heisenberg group [RR08, Theorems 6.1, 6.8, 6.11], which led to a proof of the Pansu conjecture (cf. [Pan83b, page 172]) for the isoperimetric profile of the Heisenberg group in the C^2-smooth category [RR08, Theorem 7.2].

Theorem 13.3.17 (Ritoré and Rosales [RR08])**.** *In the 3D Heisenberg group,* C^2 *isoperimetric sets are Pansu spheres.*

Theorem 13.3.18 (Monti-Rickly [MR09]). *(Euclidean) convex isoperimetric sets are Pansu spheres.*

13.3.4 Results on minimal surfaces

Let S be a hyper-surface inside a Carnot group $\mathbb{G}$ of Hausdorff dimension Q. The first two natural surface measures on S are the $(Q-1)$-Hausdorff measure $\mathcal{H}^{Q-1}_{cc} \llcorner_S$ or the perimeter measure of one of the side regions determined by S, i.e., $\mathrm{Per}(E)$ with $\partial E = S$, where the perimeter has been defined in (13.2.8). The perimeter measure $\mathrm{Per}(E)$ has a better behavior and, at least when ∂E is a C^2 hyper-surface, it coincides with $\mathcal{H}^{Q-1}_{cc} \llcorner_{\partial E}$

Let us clearify now the terminology of 'minimal surface'.

Definition 13.3.19. If $\Sigma \subset \mathbb{G}$ is such that for all Σ' such that there exists $R > 0$ such that

[...] then we say that Σ is *globally area-minimizing*

Definition 13.3.20 (...). then we say that Σ is *(locally) area-minimizing*

Definition 13.3.21 (...).

$$-\nabla_{\mathbb{G}} \cdot \frac{\nabla_{\mathbb{G}} F}{|\nabla_{\mathbb{G}} F|} \equiv 0, \qquad \text{where } \nabla_{\mathbb{G}} f = (X_1 f, \ldots, X_m f), \tag{13.3.22}$$

then we say that Σ has *zero mean curvature* or that it is a *solution of the minimal surface equation.*

Definition 13.3.23 (...). then we say that Σ is *area-stationary.*

With the term 'minimal surface' authors can reefer to any of the 4 above definitions.

Theorem 13.3.24 (Existence of area-minimizing sets [GN96]). *In sub-Riemannian manifolds, area-minimizing sets exist.*

Explicitly, let Ω be a bounded open set in a Carnot group $\mathbb{G}$. Let L be a locally finite perimeter set. Then the above theorem guarantees the existence of a locally finite perimeter set E such that

i) $(E \Delta L) \setminus \Omega = \emptyset$, and

ii) $(F \Delta L) \setminus \Omega = \emptyset \implies Per(E \cap \Omega) \leq Per(F \cap \Omega)$.

In other words, the (reduced) boundary of E is the area minimizing (generalized) hyper-surface inside Ω with boundary data L outside Ω.

Cheng, Hwang and Yang [CHY07] have studied the weak solutions of the minimal surface equation for intrinsic graphs in the Heisenberg group and have proven existence and uniqueness results.

Fact: The minimal surface equation is a sub-elliptic PDE: a priori, neither existence, not uniqueness, nor regularity can be deduced.

Theorem 13.3.25 (Non-uniqueness of minimal surfaces [Pau04]). *There are loops in the Heisenberg group that admit more than one filling by zero-mean curvature disks.*

N.B. This happens in the Euclidean case too.

The main difference between Euclidean and sub-Riemannian geometry is the existence of low-regular minimal surfaces.

Theorem 13.3.26 (Existence of low-regular area minimizing surfaces [Rit09, CHY07, Pau04]). *There are area-minimizing surfaces in the 3D Heisenberg group that are not C^2.*

This is due to the fact that not all area-minimizing surfaces have zero-mean curvature. On the other hand, there are examples of zero-mean curvature surfaces that are not area-minimizing, cf. [DGN08]. Does this happen in Euclidean geometry?

Moreover, the condition of having zero mean curvature is not enough to guarantee that a given surface of class C^2 is area-stationary [RR08].

Theorem 13.3.27 (Regularity of zero mean curvature surfaces [Pau06, CHY09, CCM08]). *Let S be a surface in the Heisenberg group that is either C^1 or a Lipschitz intrinsic graphs. If S have zero mean curvature (in an extended sense), then it is smooth.*

Theorem 13.3.28 (Bernstein problem). *In the Euclidean 3D space, every entire minimizing graph $\{(x, y, f(x, y) : x, y \in \mathbb{R}\}$ is a plane.*

One would expect that such a fact would be true for every n-dimensional graph in $\mathbb{R}^{n+1}$, but Bombieri, De Giorgi and Giusti established the surprising result that the Bernstein property fails if $n \leq 8$.

Theorem 13.3.29 (Counterexample in $\mathbb{R}^9$, [BDGG69]). *If $n \leq 8$ there exist complete minimal graphs in $\mathbb{R}^{n+1}$ that are not hyper-planes: For $m \geq 4$, a Simons cone, i.e., the set $E \subset \mathbb{R}^4$ defined by ${x_1}^2 + {x_2}^2 + \cdots + {x_m}^2 = x_{m+1}^2 + x_{m+2}^2 + \cdots + {x_{2m}}^2$ is a minimal surface.*

Theorem 13.3.30 (Counterexample in Heisenberg-Garofalo and Pauls). *Let $G \sim \mathbb{R}^3$ be the Heisenberg group. The real analytic surface*

$$S = \{(x, y, t) \in G | y = -x \tan(\tanh(t))\},$$

is an entire graph with zero mean curvature.

13.3.5 More results on regularity

The work of Agracev and Gauthier [AG01] gives an analytic answer in generic cases:

Theorem 13.3.31 (Agrachev-Gauthier). *Generically, the germ at a point q_0 of the function $q \mapsto \rho(q) \overset{\text{def}}{\to}= \operatorname{dist}(q, q_0)$ is subanalytic if the dimension n of the manifold and the dimension k of the distribution satisfy $n \leq (k-1)k + 1$.*

Theorem 13.3.32 (Agrachev-Gauthier). *Generically (and, in fact, on the complement of a set of distributions of infinite codimension), small balls $\{q \colon \rho(q) \leq r\}$ are subanalytic if $k \geq 3$.*

Theorem 13.3.33 (Agrachev-Gauthier). *Generically, the germ of ρ at q_0 is not subanalytic if $n \geq (k-1)\left(\frac{k^2}{3} + \frac{5k}{6} + 1\right)$.*

(Monti, 2000, 2003), (Leonardi-Masnou, 2005): There is no direct counterpart of the Brunn-Minkowski inequality in Euclidean space

(Ritor´ -Rosales, 2005), (Danielli-Garofalo-Nhieu, 2006): The sets bounded by S_λ are isoperimetric regions in restricted classes of sets (C^2 rotationally symmetric and C^1 unions of two graphs over a ball in the xy-plane $t = 0$ divided by $t = 0$ into two regions of equal volumes)

Bonk-Capogna: flow by mean curvature of a C^2 convex surface which is the union of two radial graphs, converges to S_λ

13.4 Translations and flows*

Given $X \in \Gamma(TM)$ we can consider the associated flow, i.e., the solution $\Phi_X : M \times \mathbb{R} \to M$ of the following ODE

$$\begin{cases} \dfrac{\mathrm{d}}{\mathrm{d}t}\Phi_X(p,t) &= X_{\Phi_X(p,t)} \\ \Phi_X(p,0) &= p. \end{cases} \tag{13.4.1}$$

Notice that the smoothness of X ensures uniqueness, and therefore the semigroup property

$$\Phi_X(x, t+s) = \Phi_X(\Phi_X(x,t), s) \qquad \forall t,\ s \in \mathbb{R},\ \forall x \in M \tag{13.4.2}$$

but not global existence; it is guaranteed, however, for left-invariant vector fields in Lie groups. We obviously have

$$\frac{\mathrm{d}}{\mathrm{d}t}(u \circ \Phi_X)(p,t) = (Xu)(\Phi_X(p,t)) \qquad \forall u \in C^1(M). \tag{13.4.3}$$

An obvious consequence of this identity is that, for a C^1 function u, $Xu = 0$ implies that u is constant along the flow, i.e., $u \circ \Phi_X(\cdot, t) = u$ for all $t \in \mathbb{R}$. A similar statement holds even for distributional derivatives along vector fields: for simplicity let us state and prove this result for divergence-free vector fields only.

Theorem 13.4.4. *Let $u \in L^1_{\mathrm{loc}}(M)$ be satisfying $Xu = 0$ in the sense of distributions. Then, for all $t \in \mathbb{R}$, $u = u \circ \Phi_X(\cdot, t)$ vol_M-a.e. in M.*

Proof. Let $g \in C^1_c(M)$; we need to show that the map $t \mapsto \int_M gu \circ \Phi_X(\cdot, t)\, d\,\mathrm{vol}_M$ is independent of t. Indeed, the semigroup property (13.4.2), and the fact that X is divergence-free yield

$$\begin{aligned}
& \int_M gu \circ \Phi_X(\cdot, t+s)\, d\,\mathrm{vol}_M - \int_M gu \circ \Phi_X(\cdot, t)\, d\,\mathrm{vol}_M \\
= & \int_M ug \circ \Phi_X(\cdot, -t-s)\, d\,\mathrm{vol}_M - \int_M ug \circ \Phi_X(\cdot, -t)\, d\,\mathrm{vol}_M \\
= & \int_M ug \circ \Phi_X(\Phi_X(\cdot, -s), -t)\, d\,\mathrm{vol}_M - \int_M ug \circ \Phi_X(\cdot, -t)\, d\,\mathrm{vol}_M \\
= & -s \int_M uX(g \circ \Phi_X(\cdot, -t))\, d\,\mathrm{vol}_M + o(s) = o(s).
\end{aligned}$$

□

Remark 13.4.5. We notice also that the flow is vol_M-measure preserving (i.e. $\mathrm{vol}_M(\Phi_X(\cdot, t)^{-1}(A)) = \mathrm{vol}_M(A)$ for all Borel sets $A \subseteq M$ and $t \in \mathbb{R}$) if and only if $\operatorname{div} X$ is equal to 0. Indeed, if $f \in C^1_c(M)$, the measure preserving property gives that $\int_M f(\Phi_X(x,t))\, d\,\mathrm{vol}_M(x)$ is independent of t. A time differentiation and (13.4.3) then give

$$0 = \int_M \frac{\mathrm{d}}{\mathrm{d}t} f(\Phi_X(x,t))\ \mathrm{d}\,\mathrm{vol}_M(x) = \int_M Xf(\Phi_X(x,t))\ \mathrm{d}\,\mathrm{vol}_M(x) = \int_M Xf(y)\ \mathrm{d}\,\mathrm{vol}_M(y).$$

Therefore $\int_M f \operatorname{div} X\, d\,\mathrm{vol}_M = 0$ for all $f \in C^1_c(M)$, and X is divergence-free. The proof of the converse implication is similar, and analogous to the one of Theorem 13.4.4.

Let $\mathbb{G}$ be a Lie group with Lie algebra $\mathfrak{g}$. We shall also consider as volume form $\mathrm{vol}_{\mathbb{G}}$ a right-invariant Haar measure.

Let $X \in \mathfrak{g}$ and let us denote, as usual in the theory, by $\exp(tX)$ the flow of X at time t starting from e (that is, $\exp(tX) := \Phi_X(1_{\mathbb{G}}, t) = \Phi_{tX}(1_{\mathbb{G}}, 1)$); then, the curve $g\exp(tX)$ is the flow starting at g: indeed, since X is left-invariant, setting for simplicity $\gamma(t) := \exp(tX)$ and $\gamma_g(t) := g\gamma(t)$, we have

$$\frac{\mathrm{d}}{\mathrm{d}t}\gamma_g(t) = \frac{\mathrm{d}}{\mathrm{d}t}(L_g(\gamma(t))) = (\,\mathrm{d}L_g)_{\gamma(t)} \frac{\mathrm{d}}{\mathrm{d}t}\gamma(t) = (\,\mathrm{d}L_g)_{\gamma(t)} X = X_{\gamma_g(t)}.$$

This implies that $\Phi_X(\cdot, t) = R_{\exp(tX)}$ and so the flow preserves the right Haar measure, and the left translation preserves the flow lines. By Remark 13.4.5 it follows that all $X \in \mathfrak{g}$ are divergence-free, and Theorem 13.4.4 gives

$$f \circ R_{\exp(tX)} = f \quad \forall t \in \mathbb{R} \qquad \Longleftrightarrow \qquad Xf = 0 \tag{13.4.6}$$

whenever $f \in L^1_{\mathrm{loc}}(\mathbb{G})$.

13.4.1 X-derivative of nice functions and domains

If u is a C^1 function in $\mathbb{R}^n$, then Xu can be calculated as the scalar product between X and the gradient of u:

$$Xu = \langle X, \nabla u \rangle. \tag{13.4.7}$$

Assume that $E \subset \mathbb{R}^n$ is locally the sub-level set of the C^1 function f and that $X \in \Gamma(T\mathbb{R}^n)$ is divergence-free. Then, for every $v \in C^\infty_c(\mathbb{R}^n)$ we can apply the Gauss–Green formula to the vector field vX, whose divergence is Xv, to obtain

$$\int_E Xv \, \mathrm{d}x = \int_{\partial E} \langle vX, \nu_E^{eu} \rangle \, \mathrm{d}\mathscr{H}^{n-1},$$

where ν_E^{eu} is the unit (Euclidean) outer normal to E. This proves that

$$X\chi_E = -\langle X, \nu_E^{eu} \rangle \mathscr{H}^{n-1} \llcorner_{\partial E}.$$

However, we have an explicit formula for the unit (Euclidean) outer normal to E, it is $\nu_E^{eu}(x) = \nabla f(x)/|\nabla f(x)|$, so, by (13.4.7),

$$\begin{aligned} \langle X, \nu_E^{eu} \rangle &= \langle X, \frac{\nabla f}{|\nabla f|} \rangle \\ &= \frac{\langle X, \nabla f \rangle}{|\nabla f|} = \frac{Xf}{|\nabla f|}. \end{aligned}$$

Thus

$$X\chi_E = -\frac{Xf}{|\nabla f|} \mathscr{H}^{n-1} \llcorner_{\partial E}. \tag{13.4.8}$$

Appendix A

Dido's problem

For a better understanding of how in Section 1.4.1 we obtained formulas for the geodesics in the subRiemannian Heisenberg group, we discuss in this section the solutions of the isoperimetric problem. We then solve Dido's problem. The proof will be done under the nontrivial assumption that the minimizers of the problems are curves that are smooth enough. For the general case, we refer the reader to [].

A.1 A proof of the isoperimetric problem*

We shall use the formalism of Calculus of Variations for proving that each of the shortest closed curves in the plane that encloses a fix amount of area is a circle. We will not need to show any preliminary on the curve such as the fact that it is locally a graph or that the enclosed domain is convex. We prove that the only critical points of the variational integral functional

$$\mathcal{L}(\sigma) := \text{Length}(\sigma),$$

subjected to the bond

$$\mathcal{A}(\sigma) := \text{Area enclosed by } \sigma = A_0, \text{ for some } A_0,$$

are circles. However, we shall assume that such a σ is a C^1 curve with Lipschitz derivative.

A.1.1 Variation of length

A necessary condition for σ being a critical point, is the vanishing of the first variation of $\mathcal{L}$.

Let $\sigma : [0, l] \to \mathbb{R}^2$ be a Lipschitz curve with coordinates (σ_1, σ_2). Its length is given by

$$\mathcal{L}(\sigma) = \int_0^l \sqrt{\dot\sigma_1^2(t) + \dot\sigma_2^2(t)}\, \mathrm{d}\, t.$$

The fact that σ is a critical point with respect to a variation h is expressed in Calculus of Variations by the equation

$$\delta\mathcal{L}(\sigma, h) = 0.$$

More explicitly, h is a curve $h : [0, l] \to \mathbb{R}^2$ with $h(0) = h(l) = 0$ and

$$\delta\mathcal{L}(\sigma, h) := \left.\frac{\mathrm{d}}{\mathrm{d}\epsilon}\mathcal{L}(\sigma + \epsilon h)\right|_{\epsilon=0} = 0.$$

Let us calculate such variation $\delta\mathcal{L}$ in the case when σ is parametrized by arc length. So $|\dot{\sigma}| = 1$ and $l = \mathrm{Length}(\sigma)$. The variation in this case is

$$\begin{aligned}
\delta\mathcal{L}(\sigma, h) &:= \left.\frac{\mathrm{d}}{\mathrm{d}\epsilon}\mathcal{L}(\sigma + \epsilon h)\right|_{\epsilon=0} \\
&= \left.\frac{\mathrm{d}}{\mathrm{d}\epsilon}\int_0^l \sqrt{\Big(\dot{\sigma}_1(t) + \epsilon\dot{h}_1(t)\Big)^2 + \Big(\dot{\sigma}_2(t) + \epsilon\dot{h}_2(t)\Big)^2}\,\mathrm{d}\,t\right|_{\epsilon=0} \\
&= \int_0^l \left.\frac{\mathrm{d}}{\mathrm{d}\epsilon}\sqrt{\dot{\sigma}_1(t)^2 + 2\epsilon\dot{\sigma}_1(t)\dot{h}_1(t) + \epsilon^2\dot{h}_1(t)^2 + \dot{\sigma}_2(t)^2 + 2\epsilon\dot{\sigma}_2(t)\dot{h}_2(t) + \epsilon^2\dot{h}_2(t)^2}\right|_{\epsilon=0}\mathrm{d}\,t \\
&= \int_0^l \left.\frac{2\dot{\sigma}_1(t)\dot{h}_1(t) + 2\epsilon\dot{h}_1(t)^2 + 2\dot{\sigma}_2(t)\dot{h}_2(t) + 2\epsilon\dot{h}_2(t)^2}{2\sqrt{\dot{\sigma}_1(t)^2 + 2\epsilon\dot{\sigma}_1(t)\dot{h}_1(t) + \epsilon^2\dot{h}_1(t)^2 + \dot{\sigma}_2(t)^2 + 2\epsilon\dot{\sigma}_2(t)\dot{h}_2(t) + \epsilon^2\dot{h}_2(t)^2}}\right|_{\epsilon=0}\mathrm{d}\,t \\
&= \int_0^l \frac{\dot{\sigma}_1(t)\dot{h}_1(t) + \dot{\sigma}_2(t)\dot{h}_2(t)}{\sqrt{\dot{\sigma}_1(t)^2 + \dot{\sigma}_2(t)^2}}\,\mathrm{d}\,t \\
&= \int_0^l \frac{\langle\dot{\sigma}(t), \dot{h}(t)\rangle}{|\dot{\sigma}(t)|}\,\mathrm{d}\,t \\
&= \int_0^l \langle\dot{\sigma}(t), \dot{h}(t)\rangle\,\mathrm{d}\,t.
\end{aligned}$$

We conclude the following:

Lemma A.1.1. *A planar curve σ, parametrized by unit speed, is a critical point of the length functional with respect to a variation h if and only if*

$$\int_0^l \langle\dot{\sigma}, \dot{h}\rangle\,\mathrm{d}\,t = 0.$$

A.1.2 Area functional and its variation

The area enclosed by a Lipschitz curve σ can be computed (because of Stokes' Theorem) by the formula

$$\mathcal{A}(\sigma) = \frac{1}{2}\int_0^l \sigma_1(t)\dot{\sigma}_2(t) - \sigma_2(t)\dot{\sigma}_1(t)\,\mathrm{d}\,t.$$

For convenience of notation let us define the 'cross product' on $\mathbb{R}^2$ as the real number

$$v \times w := v_1 w_2 - w_1 v_2 = \left\langle \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} v_1 \\ v_2 \\ 0 \end{pmatrix} \times \begin{pmatrix} w_1 \\ w_2 \\ 0 \end{pmatrix} \right\rangle, \text{ for } v = (v_1, v_2), w = (w_1, w_2) \in \mathbb{R}^2.$$

Obviously we have linearity in v and w and $w \times v = -v \times w$. Thus the area enclosed by σ is

$$\mathcal{A}(\sigma) = \frac{1}{2} \int_0^l \sigma \times \dot{\sigma} \, \mathrm{d}\, t.$$

Let h be a variation. The new area would be

$$\begin{aligned} \mathcal{A}(\sigma + h) &= \frac{1}{2} \int_0^l (\sigma + h) \times (\dot{\sigma} + \dot{h}) \, \mathrm{d}\, t \\ &= \frac{1}{2} \int_0^l \sigma \times \dot{\sigma} + \sigma \times \dot{h} + h \times \dot{\sigma} + h \times \dot{h} \, \mathrm{d}\, t \\ &= \mathcal{A}(\sigma) + \frac{1}{2} \, h \times \sigma|_0^l + \frac{1}{2} \int_0^l -\dot{\sigma} \times h + h \times \dot{\sigma} + h \times \dot{h} \, \mathrm{d}\, t \\ &= \mathcal{A}(\sigma) + \int_0^l h \times \dot{\sigma} \, \mathrm{d}\, t + \frac{1}{2} \int_0^l h \times \dot{h} \, \mathrm{d}\, t. \end{aligned}$$

We conclude the following:

Lemma A.1.2. *A variation h of a curve σ is area-preserving if and only if*

$$\int_0^l h \times \dot{\sigma} + \frac{h \times \dot{h}}{2} \, \mathrm{d}\, t = 0.$$

Definition A.1.3. We say that a variation h of a curve σ *tangentially preserves* the area if

$$\mathcal{A}(\sigma + \epsilon h) = \mathcal{A}(\sigma) + o(\epsilon).$$

In other words, h tangentially preserves the area if

$$\frac{\mathrm{d}}{\mathrm{d}\epsilon} \mathcal{A}(\sigma + \epsilon h) \Big|_{\epsilon = 0} = 0.$$

Thus, by the above calculation, such an h satisfies

$$0 = \frac{\mathrm{d}}{\mathrm{d}\epsilon} \int_0^l \epsilon h \times \dot{\sigma} + \frac{\epsilon h \times \epsilon \dot{h}}{2} \, \mathrm{d}\, t \Big|_{\epsilon = 0} = \int_0^l h \times \dot{\sigma} \, \mathrm{d}\, t.$$

Proposition A.1.4. *Let $\sigma : [0, l] \to \mathbb{R}^2$ be a curve parametrized by arc length. If σ is a critical curve for the length functional under an area constrain, then σ has zero first variation of length with respect to all tangentially area-preserving variations. In particular,*

$$\int_0^l \langle \dot{\sigma} | \dot{h} \rangle \, \mathrm{d}\, t = 0,$$

for all $h : [0, l] \to \mathbb{R}^2$ with $h(0) = h(l) = 0$ and

$$\int_0^l h \times \dot{\sigma} \, \mathrm{d}\, t = 0.$$

Proof. Set $a_\epsilon := \mathcal{A}(\sigma + \epsilon h)$, hence $\left.\frac{\mathrm{d}}{\mathrm{d}\epsilon} a_\epsilon\right|_{\epsilon=0} = 0$. Consider the curves

$$\sigma_\epsilon := \sqrt{\frac{a_0}{a_\epsilon}}(\sigma + \epsilon h).$$

Then $\sigma_0 = \sigma$ and the area enclosed by σ_ϵ is independent on ϵ. Since σ is critical for the length functional under the area constraint, we have that $\left.\frac{\mathrm{d}}{\mathrm{d}\epsilon}\mathcal{L}(\sigma_\epsilon)\right|_{\epsilon=0} = 0$. Therefore,

$$\begin{aligned}
0 &= \left.\frac{\mathrm{d}}{\mathrm{d}\epsilon}\mathcal{L}(\sigma_\epsilon)\right|_{\epsilon=0} \\
&= \left.\frac{\mathrm{d}}{\mathrm{d}\epsilon}\sqrt{\frac{a_0}{a_\epsilon}}\mathcal{L}(\sigma + \epsilon h)\right|_{\epsilon=0} \\
&= \left.\frac{\mathrm{d}}{\mathrm{d}\epsilon}\sqrt{\frac{a_0}{a_\epsilon}}\right|_{\epsilon=0}\mathcal{L}(\sigma) + \sqrt{\frac{a_0}{a_0}}\left.\frac{\mathrm{d}}{\mathrm{d}\epsilon}\mathcal{L}(\sigma + \epsilon h)\right|_{\epsilon=0} \\
&= -\frac{1}{2}\sqrt{a_0}a_\epsilon^{-3/2}\left.\frac{\mathrm{d}}{\mathrm{d}\epsilon}a_\epsilon\right|_{\epsilon=0}\mathcal{L}(\sigma) + 1\cdot\delta\mathcal{L}(\sigma, h) \\
&= 0 + \int_0^l \langle\dot\sigma, \dot h\rangle\,\mathrm{d}\,t,
\end{aligned}$$

where we used the calculation to get to Lemma A.1.1. □

A.1.3 Conclusion

Proposition A.1.5. *If σ is a $C^{1,1}$ closed curve in the plane that is one of the shortest among all Lipschitz curves that enclose a fixed amount of area, then σ is a circle.*

Proof. Assume, without loss of generality that σ has unit speed. Let $\phi : [0, l] \to \mathbb{R}$ be a C^∞ function with $\phi(0) = \phi(1) = 0$ and $\int_0^l \phi(t)\,\mathrm{d}\,t = 0$. Take $h(t) = \phi(t)(\dot\sigma_2(t), -\dot\sigma_1(t))$, which, since σ is $C^{1,1}$, is Lipschitz. Such an h is an admissible variation since clearly $h(0) = h(l) = 0$ and also

$$\begin{aligned}
\int_0^l h\times\dot\sigma\ \mathrm{d}\,t &= \int_0^l \phi(t)\,(\dot\sigma_2(t), -\dot\sigma_1(t))\times(\dot\sigma_1(t), \dot\sigma_2(t))\ \mathrm{d}\,t \\
&= \int_0^l \phi(t)(\dot\sigma_2(t)^2 + \dot\sigma_1(t)^2)\ \mathrm{d}\,t \\
&= \int_0^l \phi(t)|\dot\sigma|^2\ \mathrm{d}\,t \\
&= \int_0^l \phi(t)\cdot 1\ \mathrm{d}\,t \\
&= \int_0^l \phi(t)\,\mathrm{d}\,t = 0.
\end{aligned}$$

Then, since $\dot h(t) = \dot\phi(t)(\dot\sigma_2(t), -\dot\sigma_1(t)) + \phi(t)(\ddot\sigma_2(t), -\ddot\sigma_1(t))$, the vanishing of the first variation of length becomes

$$
\begin{aligned}
0 &= \int_0^l \langle \dot\sigma, \dot h\rangle \,\mathrm{d}\, t \\
&= \int_0^l \Big\langle (\dot\sigma_1, \dot\sigma_2), \dot\phi(t)(\dot\sigma_2(t), -\dot\sigma_1(t)) + \phi(t)(\ddot\sigma_2(t), -\ddot\sigma_1(t))\Big\rangle \,\mathrm{d}\, t \\
&= \int_0^l \dot\phi(t)\,\langle (\dot\sigma_1, \dot\sigma_2), (\dot\sigma_2(t), -\dot\sigma_1(t))\rangle + \phi(t)\,\langle (\dot\sigma_1, \dot\sigma_2), (\ddot\sigma_2(t), -\ddot\sigma_1(t))\rangle \,\mathrm{d}\, t \\
&= \int_0^l \dot\phi(t)(\dot\sigma_1(t)\dot\sigma_2(t) - \dot\sigma_2(t)\dot\sigma_1(t)) + \phi(t)(\dot\sigma_1(t)\ddot\sigma_2(t) - \dot\sigma_2(t)\ddot\sigma_1(t)) \,\mathrm{d}\, t \\
&= \int_0^l \phi(t)(\dot\sigma_1(t)\ddot\sigma_2(t) - \dot\sigma_2(t)\ddot\sigma_1(t)) \,\mathrm{d}\, t.
\end{aligned}
$$

the conclusion is that the function $\kappa(t) := \dot\sigma_1(t)\ddot\sigma_2(t) - \dot\sigma_2(t)\ddot\sigma_1(t)$, which is in fact the curvature of the curve σ, is such that

$$\int_0^l \phi(t)\kappa(t)\,\mathrm{d}\, t = 0 \text{ for all } \phi \in C^\infty([0,l]) \text{ such that } \phi(0) = \phi(1) \text{ and } \int_0^l \phi(t)\,\mathrm{d}\, t = 0.$$

By the (second) Fundamental Lemma of Calculus of Variations (due to DuBois and Reymond) we deduce that κ is constant. The only planar curves of constant curvature are circles (and lines). $\square$

The assumption that the curve is $C^{1,1}$ can be dropped, but the proof of the result would not be as brief. We refer to other texts for the more general result. For examples, a complete proof, based on Poincare-Wirtinger inequality, can be found in [Oss78, pp. 1183-1185]. The following general statement of the isoperimetric solution is for curves that are absolutely continuous.

Theorem A.1.6 (Isoperimetric solution)**.** *If σ is a closed absolutely continuous curve in the plane that is one of the shortest among all absolutely continuous curves that enclose a fixed amount of area, then σ is a parametrization of a circle.*

From the solution of the isoperimetric problem, Dido's problem has an immediate solution.

Theorem A.1.7 (Dido's solution)**.** *Given two points p and q on the plane and a number $A \in \mathbb{R}$, the shortest curve from p to q that, together with the segment from p to q encloses area A is an arc of a circle.*

Proof. Assume by contradiction that there is a shortest curve σ that is not as arc of a circle. Let γ be the arc of circle enclosing area A. (Notice that such an arc is unique). Let $\hat\gamma$ be the circle of which γ is an arc. Let $\tilde\gamma$ be the complementary arc of γ, i.e., γ followed by $\tilde\gamma$ is $\hat\gamma$. Observe that the curve $\hat\sigma$ obtained following $\tilde\gamma$ after σ is such that

$$\mathcal{A}(\hat\sigma) = \mathcal{A}(\hat\gamma) \qquad \text{and} \qquad \mathcal{L}(\hat\sigma) < \mathcal{L}(\hat\gamma).$$

Hence we get a contradiction with Theorem A.1.6. $\square$

Appendix B

Parking motorbikes and cars*

We shall explains how to formulate and solve the car parking problem with the use of sub-Riemannian geometry.

The state space of a car on a parking lot is modelled by the manifold $\mathbb{R} \times \mathbb{S}_{\times}\mathbb{S}$. Namely, we are in $\mathbb{R}_{(x_1,x_2)} \times \mathbb{S}^1_\alpha \times \mathbb{S}^1_\theta$, where (x_1, x_2) is the location of the middle point of the front axle, α is the angle that the orientation of car makes with the chosen x-axis and θ is the orientation of the steering wheel. Assume the car is moving with constant speed v. To be precise, this speed is the speed of the point (x_1, x_2). This speed can be divided to components v_t and v_f as in the picture below

We look for the admissible direction of movement within the state space of the car. It is clear that one of the horizontal vectors is ∂_θ, since we are totally free to change this parameter by turning the steering wheel. It remains the question, what kind of a path does the car make in the state space if we move forward with constant speed v. The remaining horizontal vector is the tangent vector of this path.

The necessary physical observation is, that if we consider ourselves in the coordinate frame where the middle point of the *rear* axle is fixed (let this have the distance ℓ from the front axle), the only movement of the car is getting rotated, and this rotation is due to the component v_t. Quantitatively, the angular velocity satisfies $\omega\ell = v_t$, so

$$\dot{\alpha} = \frac{v_t}{\ell} = \frac{\|v\| \sin\theta}{\ell}$$

The movement of the car happens to the direction of v, so we get

$$(\dot{x}_1, \dot{x}_2) = \|v\| \cos(\alpha + \theta)\partial_{x_1} + \|v\| \sin(\alpha + \theta)\partial_{x_2}$$

Thus the other horizontal direction on the tangent space is

$$\cos(\alpha+\theta)\partial_{x_1} + \sin(\alpha+\theta)\partial_{x_2} + \frac{\sin\theta}{\ell}\partial_\alpha.$$

——

Rotational movement: $\dot\theta_1 = 1$ and $\dot x = \dot y = \dot\theta_2 = 0$, therefore

$$X = \partial_{\theta_1} = (0,0,1,0).$$

Forward movement: $(\dot x, \dot y) = (\cos\theta_1, \sin\theta_1)$, $\dot\theta_1 = 0$, $\dot\theta_2 = \sin(\theta_1 - \theta_2)$, therefore

$$Y = \cos\theta_1\partial_x + \sin\theta_1\partial_y + \sin(\theta_1-\theta_2)\partial_{\theta_2}$$

We want to show that the system of the car is controllable, i.e., the subbundle spanned by X and Y s bracket generating.

$$[X,Y] = ... = (-\sin\theta_1, \cos\theta_1, 0, \cos(\theta_1-\theta_2)).$$

$$[[X,Y],Y] = \cdots = (0,0,0,-1).$$

The vector fields $X, Y, [X,Y], [[X,Y],Y]$ span the tangent space at every point since ...[1]

[1] to be finished

Bibliography

[ABB15] Andrei Agrachev, Davide Barilari, and Ugo Boscain. Introduction to Riemannian and Sub-Riemannian geometry. *Manuscript*, 2015.

[AFP00] Luigi Ambrosio, Nicola Fusco, and Diego Pallara. *Functions of bounded variation and free discontinuity problems*. Oxford Mathematical Monographs. The Clarendon Press Oxford University Press, New York, 2000.

[AG01] Andrei Agrachev and Jean-Paul Gauthier. On the subanalyticity of Carnot-Caratheodory distances. *Ann. Inst. H. Poincaré Anal. Non Linéaire*, 18(3):359–382, 2001.

[AK00] Luigi Ambrosio and Bernd Kirchheim. Rectifiable sets in metric and Banach spaces. *Math. Ann.*, 318(3):527–555, 2000.

[AKL09] Luigi Ambrosio, Bruce Kleiner, and Enrico Le Donne. Rectifiability of sets of finite perimeter in Carnot groups: existence of a tangent hyperplane. *J. Geom. Anal.*, 19(3):509–540, 2009.

[AP94] Marco Abate and Giorgio Patrizio. *Finsler metrics—a global approach*, volume 1591 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 1994. With applications to geometric function theory.

[AS04] Andrei A. Agrachev and Yuri L. Sachkov. *Control theory from the geometric viewpoint*, volume 87 of *Encyclopaedia of Mathematical Sciences*. Springer-Verlag, Berlin, 2004. Control Theory and Optimization, II.

[AT04] Luigi Ambrosio and Paolo Tilli. *Topics on analysis in metric spaces*, volume 25 of *Oxford Lecture Series in Mathematics and its Applications*. Oxford University Press, Oxford, 2004.

[AT12] Marco Abate and Francesca Tovena. *Curves and surfaces*, volume 55 of *Unitext*. Springer, Milan, 2012. Translated from the 2006 Italian original by Daniele A. Gewurz.

[BBI01] Dmitriĭ Burago, Yuriĭ Burago, and Sergeiĭ Ivanov. *A course in metric geometry*, volume 33 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2001.

[BCS00] D. Bao, S.-S. Chern, and Z. Shen. *An introduction to Riemann-Finsler geometry*, volume 200 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 2000.

[BDGG69] E. Bombieri, E. De Giorgi, and E. Giusti. Minimal cones and the Bernstein problem. *Invent. Math.*, 7:243–268, 1969.

[Bel96] André Bellaïche. The tangent space in sub-Riemannian geometry. In *Sub-Riemannian geometry*, volume 144 of *Progr. Math.*, pages 1–78. Birkhäuser, Basel, 1996.

[Ber88] Valeriĭ N. Berestovskiĭ. Homogeneous manifolds with an intrinsic metric. I. *Sibirsk. Mat. Zh.*, 29(6):17–29, 1988.

[Bes78] Arthur L. Besse. *Manifolds all of whose geodesics are closed*, volume 93 of *Ergebnisse der Mathematik und ihrer Grenzgebiete [Results in Mathematics and Related Areas]*. Springer-Verlag, Berlin-New York, 1978. With appendices by D. B. A. Epstein, J.-P. Bourguignon, L. Bérard-Bergery, M. Berger and J. L. Kazdan.

[BH99] Martin R. Bridson and A. Haefliger. *Metric spaces of non-positive curvature*, volume 319 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 1999.

[BL13] Emmanuel Breuillard and Enrico Le Donne. On the rate of convergence to the asymptotic cone for nilpotent groups and subFinsler geometry. *Proc. Natl. Acad. Sci. USA*, 110(48):19220–19226, 2013.

[BLU07] A. Bonfiglioli, E. Lanconelli, and F. Uguzzoni. *Stratified Lie groups and potential theory for their sub-Laplacians*. Springer Monographs in Mathematics. Springer, Berlin, 2007.

[Bou95] M. Bourdon. Structure conforme au bord et flot géodésique d'un CAT(-1)-espace. *Enseign. Math. (2)*, 41(1-2):63–102, 1995.

[Bre14] Emmanuel Breuillard. Geometry of locally compact groups of polynomial growth and shape of large balls. *Groups Geom. Dyn.*, 8(3):669–732, 2014.

[BS14] Sergei Buyalo and Viktor Schroeder. Möbius characterization of the boundary at infinity of rank one symmetric spaces. *Geom. Dedicata*, 172:1–45, 2014.

[Bul02] Marius Buliga. Sub-Riemannian geometry and Lie groups. Part I. *Seminar Notes, DMA-EPFL, Preprint on Arxiv*, 2002.

[Bul11] Marius Buliga. A characterization of sub-Riemannian spaces as length dilation structures constructed via coherent projections. *Commun. Math. Anal.*, 11(2):70–111, 2011.

[Cap97] Luca Capogna. Regularity of quasi-linear equations in the Heisenberg group. *Comm. Pure Appl. Math.*, 50(9):867–889, 1997.

[Car26] E. Cartan. Sur une classe remarquable d'espaces de Riemann. *Bull. Soc. Math. France*, 54:214–264, 1926.

[Car27] E. Cartan. Sur une classe remarquable d'espaces de Riemann. II. *Bull. Soc. Math. France*, 55:114–134, 1927.

[CCM08] Luca Capogna, Giovanna Citti, and Maria Manfredini. Regularity of minimal surfaces in the one-dimensional Heisenberg group. In *"Bruno Pini" Mathematical Analysis Seminar, University of Bologna Department of Mathematics: Academic Year 2006/2007 (Italian)*, pages 147–162. Tecnoprint, Bologna, 2008.

[CDPT07] Luca Capogna, Donatella Danielli, Scott D. Pauls, and Jeremy T. Tyson. *An introduction to the Heisenberg group and the sub-Riemannian isoperimetric problem*, volume 259 of *Progress in Mathematics*. Birkhäuser Verlag, Basel, 2007.

[CG90] Lawrence J. Corwin and Frederick P. Greenleaf. *Representations of nilpotent Lie groups and their applications. Part I*, volume 18 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 1990. Basic theory and examples.

[CHY07] Jih-Hsin Cheng, Jenn-Fang Hwang, and Paul Yang. Existence and uniqueness for p-area minimizers in the Heisenberg group. *Math. Ann.*, 337(2):253–293, 2007.

[CHY09] Jih-Hsin Cheng, Jenn-Fang Hwang, and Paul Yang. Regularity of C^1 smooth surfaces with prescribed p-mean curvature in the Heisenberg group. *Math. Ann.*, 344(1):1–35, 2009.

[CJT06] Y. Chitour, F. Jean, and E. Trélat. Genericity results for singular curves. *J. Differential Geom.*, 73(1):45–73, 2006.

[CKL+17] Michael Cowling, Ville Kivioja, Enrico Le Donne, Sebastiano Nicolussi Golo, and Alessandro Ottazzi. From homogeneous metric spaces to Lie groups. *ArXiv e-prints*, 2017.

[CL55] Earl A. Coddington and Norman Levinson. *Theory of ordinary differential equations.* McGraw-Hill Book Co., Inc., New York-Toronto-London, 1955.

[DG54] Ennio De Giorgi. Su una teoria generale della misura $(r-1)$-dimensionale in uno spazio ad r dimensioni. *Ann. Mat. Pura Appl. (4)*, 36:191–213, 1954.

[DG55] Ennio De Giorgi. Nuovi teoremi relativi alle misure $(r-1)$-dimensionali in uno spazio ad r dimensioni. *Ricerche Mat.*, 4:95–113, 1955.

[DGN08] D. Danielli, N. Garofalo, and D. M. Nhieu. A notable family of entire intrinsic minimal graphs in the Heisenberg group which are not perimeter minimizing. *Amer. J. Math.*, 130(2):317–339, 2008.

[DK11] Cornelia Drutu and Michael Kapovich. Lectures on geometric group theory. *Manuscript*, 2011.

[Fed69] Herbert Federer. *Geometric measure theory.* Die Grundlehren der mathematischen Wissenschaften, Band 153. Springer-Verlag New York Inc., New York, 1969.

[Fol73] G. B. Folland. A fundamental solution for a subelliptic operator. *Bull. Amer. Math. Soc.*, 79:373–376, 1973.

[Fol99] Gerald B. Folland. *Real analysis.* Pure and Applied Mathematics (New York). John Wiley & Sons, Inc., New York, second edition, 1999. Modern techniques and their applications, A Wiley-Interscience Publication.

[Fre12] David M. Freeman. Transitive bi-Lipschitz group actions and bi-Lipschitz parameterizations. *To appear in the Indiana University Mathematics Journal*, 2012.

[FSS96] Bruno Franchi, Raul Serapioni, and Francesco Serra Cassano. Meyers-Serrin type theorems and relaxation of variational integrals depending on vector fields. *Houston J. Math.*, 22(4):859–890, 1996.

[FSS01] Bruno Franchi, Raul Serapioni, and Francesco Serra Cassano. Rectifiability and perimeter in the Heisenberg group. *Math. Ann.*, 321(3):479–531, 2001.

[FSS03] Bruno Franchi, Raul Serapioni, and Francesco Serra Cassano. On the structure of finite perimeter sets in step 2 Carnot groups. *J. Geom. Anal.*, 13(3):421–466, 2003.

[Gle52] Andrew M. Gleason. Groups without small subgroups. *Ann. of Math. (2)*, 56:193–212, 1952.

[GN96] Nicola Garofalo and Duy-Minh Nhieu. Isoperimetric and Sobolev inequalities for Carnot-Carathéodory spaces and the existence of minimal surfaces. *Comm. Pure Appl. Math.*, 49(10):1081–1144, 1996.

[Gro81] Mikhael Gromov. Groups of polynomial growth and expanding maps. *Inst. Hautes Études Sci. Publ. Math.*, (53):53–73, 1981.

[Gro96] Mikhail Gromov. Carnot-Carathéodory spaces seen from within. In *Sub-Riemannian geometry*, volume 144 of *Progr. Math.*, pages 79–323. Birkhäuser, Basel, 1996.

[Gro99] Mikhail Gromov. *Metric structures for Riemannian and non-Riemannian spaces*, volume 152 of *Progress in Mathematics*. Birkhäuser Boston Inc., Boston, MA, 1999. Based on the 1981 French original, With appendices by M. Katz, P. Pansu and S. Semmes.

[Hei74] E. Heintze. On homogeneous manifolds of negative curvature. *Math. Ann.*, 211:23–34, 1974.

[Hei01] Juha Heinonen. *Lectures on analysis on metric spaces*. Universitext. Springer-Verlag, NY, 2001.

[Hel01] Sigurdur Helgason. *Differential geometry, Lie groups, and symmetric spaces*, volume 34 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2001. Corrected reprint of the 1978 original.

[HN12] Joachim Hilgert and Karl-Hermann Neeb. *Structure and geometry of Lie groups*. Springer Monographs in Mathematics. Springer, New York, 2012.

[Jac79] Nathan Jacobson. *Lie algebras*. Dover Publications Inc., New York, 1979. Republication of the 1962 original.

[Jea14] Frédéric Jean. *Control of nonholonomic systems: from sub-Riemannian geometry to motion planning*. Springer Briefs in Mathematics. Springer, Cham, 2014.

[Kle10] Bruce Kleiner. A new proof of Gromov's theorem on groups of polynomial growth. *J. Amer. Math. Soc.*, 23(3):815–829, 2010.

[Kna02] Anthony W. Knapp. *Lie groups beyond an introduction*, volume 140 of *Progress in Mathematics*. Birkhäuser Boston Inc., Boston, MA, second edition, 2002.

[LD11a] Enrico Le Donne. Geodesic manifolds with a transitive subset of smooth biLipschitz maps. *Groups Geom. Dyn.*, 5(3):567–602, 2011.

[LD11b] Enrico Le Donne. Metric spaces with unique tangents. *Ann. Acad. Sci. Fenn. Math.*, 36(2):683–694, 2011.

[LD15] Enrico Le Donne. A metric characterization of Carnot groups. *Proc. Amer. Math. Soc.*, 143(2):845–849, 2015.

[LD17] Enrico Le Donne. A primer on Carnot groups: homogenous groups, Carnot-carathéodory spaces, and regularity of their isometries. *Analysis and Geometry in Metric Spaces*, 5(1):116–137, 2017.

[LM08] Gian Paolo Leonardi and Roberto Monti. End-point equations and regularity of sub-Riemannian geodesics. *Geom. Funct. Anal.*, 18(2):552–582, 2008.

[LM10] Enrico Le Donne and Valentino Magnani. Measure of submanifolds in the Engel group. *Rev. Mat. Iberoam.*, 26(1):333–346, 2010.

[LMO+16] Enrico Le Donne, Richard Montgomery, Alessandro Ottazzi, Pierre Pansu, and Davide Vittone. Sard property for the endpoint map on some Carnot groups. *Ann. Inst. H. Poincaré Anal. Non Linéaire*, 33(6):1639–1666, 2016.

[LN20] Enrico Le Donne and Sebastiano Nicolussi Golo. Metric Lie groups admitting dilations. *To appear in Arkiv för Matematik*, 2020.

[LR03] G. P. Leonardi and S. Rigot. Isoperimetric sets on Carnot groups. *Houston J. Math.*, 29(3):609–637 (electronic), 2003.

[LS95] Wensheng Liu and Héctor J. Sussman. Shortest paths for sub-Riemannian metrics on rank-two distributions. *Mem. Amer. Math. Soc.*, 118(564):x+104, 1995.

[Mag08a] Valentino Magnani. Blow-up estimates at horizontal points and applications. *Preprint*, 2008.

[Mag08b] Valentino Magnani. Non-horizontal submanifolds and coarea formula. *J. Anal. Math.*, 106:95–127, 2008.

[Mal51] A. I. Malcev. On a class of homogeneous spaces. *Amer. Math. Soc. Translation*, 1951(39):33, 1951.

[MM95] Gregori A. Margulis and George D. Mostow. The differential of a quasi-conformal mapping of a Carnot-Carathéodory space. *Geom. Funct. Anal.*, 5(2):402–433, 1995.

[Mon02] Richard Montgomery. *A tour of subriemannian geometries, their geodesics and applications*, volume 91 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2002.

[MPV07] Irina Markina, Dmitri Prokhorov, and Alexander Vasil′ev. Sub-Riemannian geometry of the coefficients of univalent functions. *J. Funct. Anal.*, 245(2):475–492, 2007.

[MR09] Roberto Monti and Matthieu Rickly. Convex isoperimetric sets in the Heisenberg group. *Ann. Sc. Norm. Super. Pisa Cl. Sci. (5)*, 8(2):391–415, 2009.

[Mun75] James R. Munkres. *Topology: a first course*. Prentice-Hall Inc., Englewood Cliffs, N.J., 1975.

[MV08] Valentino Magnani and Davide Vittone. An intrinsic measure for submanifolds in stratified groups. *J. Reine Angew. Math.*, 619:203–232, 2008.

[MZ52] Deane Montgomery and Leo Zippin. Small subgroups of finite-dimensional groups. *Ann. of Math. (2)*, 56:213–241, 1952.

[Oss78] Robert Osserman. The isoperimetric inequality. *Bull. Amer. Math. Soc.*, 84(6):1182–1238, 1978.

[Pan82] Pierre Pansu. Une inégalité isopérimétrique sur le groupe de Heisenberg. *C. R. Acad. Sci. Paris Sér. I Math.*, 295(2):127–130, 1982.

[Pan83a] Pierre Pansu. Croissance des boules et des géodésiques fermées dans les nilvariétés. *Ergodic Theory Dynam. Systems*, 3(3):415–445, 1983.

[Pan83b] Pierre Pansu. An isoperimetric inequality on the Heisenberg group. *Rend. Sem. Mat. Univ. Politec. Torino*, (Special Issue):159–174 (1984), 1983. Conference on differential geometry on homogeneous spaces (Turin, 1983).

[Pan89] Pierre Pansu. Métriques de Carnot-Carathéodory et quasiisométries des espaces symétriques de rang un. *Ann. of Math. (2)*, 129(1):1–60, 1989.

[Pau04] Scott D. Pauls. Minimal surfaces in the Heisenberg group. *Geom. Dedicata*, 104:201–231, 2004.

[Pau06] Scott D. Pauls. H-minimal graphs of low regularity in $\mathbb{H}^1$. *Comment. Math. Helv.*, 81(2):337–381, 2006.

[Pon66] L. S. Pontryagin. *Topological groups*. Translated from the second Russian edition by Arlen Brown. Gordon and Breach Science Publishers, Inc., New York-London-Paris, 1966.

[Rag72] M. S. Raghunathan. *Discrete subgroups of Lie groups*. Springer-Verlag, New York, 1972. Ergebnisse der Mathematik und ihrer Grenzgebiete, Band 68.

[Rif14] Ludovic Rifford. *Sub-Riemannian geometry and optimal transport*. Springer Briefs in Mathematics. Springer, Cham, 2014.

[Rit09] Manuel Ritoré. Examples of area-minimizing surfaces in the sub-Riemannian Heisenberg group $\mathbb{H}^1$ with low regularity. *Calc. Var. Partial Differential Equations*, 34(2):179–192, 2009.

[RR08] Manuel Ritoré and César Rosales. Area-stationary surfaces in the Heisenberg group $\mathbb{H}^1$. *Adv. Math.*, 219(2):633–671, 2008.

[RS76] Linda Preiss Rothschild and E. M. Stein. Hypoelliptic differential operators and nilpotent groups. *Acta Math.*, 137(3-4):247–320, 1976.

[Sch95] Richard D. Schafer. *An introduction to nonassociative algebras*. Dover Publications, Inc., New York, 1995. Corrected reprint of the 1966 original.

[SCP08] Alessandro Sarti, Giovanna Citti, and Jean Petitot. The symplectic structure of the primary visual cortex. *Biol. Cybernet.*, 98(1):33–48, 2008.

[Sem96] Stephen Semmes. On the nonexistence of bi-Lipschitz parameterizations and geometric problems about A_∞-weights. *Rev. Mat. Iberoamericana*, 12(2):337–410, 1996.

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

[Ste93] Elias M. Stein. *Harmonic analysis: real-variable methods, orthogonality, and oscillatory integrals*, volume 43 of *Princeton Mathematical Series*. Princeton University Press, Princeton, NJ, 1993. With the assistance of Timothy S. Murphy, Monographs in Harmonic Analysis, III.

[Str86] Robert S. Strichartz. Sub-Riemannian geometry. *J. Differential Geom.*, 24(2):221–263, 1986.

[Str89] Robert S. Strichartz. Corrections to: "Sub-Riemannian geometry" [J. Differential Geom. **24** (1986), no. 2, 221–263; *J. Differential Geom.*, 30(2):595–596, 1989.

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

[vdDW84] L. van den Dries and A. J. Wilkie. Gromov's theorem on groups of polynomial growth and elementary logic. *J. Algebra*, 89(2):349–374, 1984.

[Vod07] S. K. Vodopyanov. Geometry of Carnot-Carathéodory spaces and differentiability of mappings. In *The interaction of analysis and geometry*, volume 424 of *Contemp. Math.*, pages 247–301. Amer. Math. Soc., Providence, RI, 2007.

[War83] Frank W. Warner. *Foundations of differentiable manifolds and Lie groups*, volume 94 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1983. Corrected reprint of the 1971 edition.

Index