Graduate Texts in Mathematics

Enrico Le Donne

Metric Lie Groups

Carnot-Carathéodory Spaces from the Homogeneous Viewpoint

Springer

Graduate Texts in Mathematics 306

Graduate Texts in Mathematics

Graduate Texts in Mathematics bridge the gap between passive study and creative understanding, offering graduate-level introductions to advanced topics in mathematics. The volumes are carefully written as teaching aids and highlight characteristic features of the theory. Although these books are frequently used as textbooks in graduate courses, they are also suitable for individual study.

Enrico Le Donne

Metric Lie Groups

Carnot-Carathéodory Spaces from the Homogeneous Viewpoint

Springer

Enrico Le Donne
Department of Mathematics
University of Fribourg
Fribourg, Switzerland

ISSN 0072-5285 ISSN 2197-5612 (electronic)
Graduate Texts in Mathematics
ISBN 978-3-031-98831-8 ISBN 978-3-031-98832-5 (eBook)
https://doi.org/10.1007/978-3-031-98832-5

This work was supported by European Research Council (713998)

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

To my students and my colleagues

Preface

The present book treats sub-Riemannian geometries and their generalizations, which go under the names of sub-Finsler geometries or Carnot-Carathéodory spaces. We will discuss these non-smooth geometries, focusing on the cases where there is the additional presence of a group structure. The techniques from Lie group theory will then be mixed with metric geometry, giving a new viewpoint.

This preface describes the origins of this book, outlines its purpose, and identifies its intended audience. The text draws heavily on the following references: [Pon66, War83, CG90, Bel96, Gro99, AFP00, BBI01, Hel01, Mon02, Kna02, HN12], as well as on various articles due to the author, his collaborators, and students, such as [AKL09, BL13, LD15, LD+16, LD17, LN21, Pil22, Pur23, Sop23, ALN23, Cow+24]. It incorporates insights from conversations with collaborators and mentors. The author would like to acknowledge, in particular, Bruce Kleiner, Urs Lang, Emmanuel Breuillard, Alessandro Ottazzi, Pierre Pansu, and Yves de Cornulier, who have provided invaluable guidance and support throughout this work.

This text has its origin in lecture notes for a course titled "Sub-Riemannian Geometry" taught at ETH Zürich during the fall of 2009 and later at the University of Jyväskylä in the spring of 2014. Additional sections were included after the author delivered courses on "Carnot Groups" at a summer school in Levico Terme (Trento, Italy) in 2015 and "Riemannian and Sub-Riemannian Geometry on Lie Groups" at the Neurogeometry summer school in Cortona (Italy) in 2017. The notes were further expanded for the course 'Sub-Riemannian Geometry' taught at the University of Fribourg (Switzerland) in spring 2021.

This book's primary audience consists of young researchers seeking an introduction to Lie groups equipped with sub-Riemannian metrics. It can serve as background reading material for a master's thesis or as an initial reference for those beginning a PhD program focusing on subjects at the crossroads of geometry, analysis, and group theory.

In contrast to other sources, such as [Mon02, BLU07, Jea14, Rif14, ABB20], this book employs the language and formalism of Lie groups and treats the general category of Carnot-Carathéodory spaces, considering norms that do not necessarily come from scalar products. In fact, one of the aims of this book is to demonstrate how sub-Finsler geometries manifest in other mathematical domains, including hyperbolic geometry and geometric group theory, through the perspective of Lie groups.

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

A special acknowledgment goes to Sebastiano Nicolussi Golo, who has been a personal trainer for bringing this book to completion. Many other students and colleagues read earlier versions of this text and gave input and feedback. The book has been a big team effort. Thanks to Gioacchino Antonelli, Ugo Boscain, Emmanuel Breuillard, Elia Bubani, Luca Capogna, Michael Cowling, Daniela Di Donato, Sylvester Eriksson-Bique, David Freeman, Patrick Ghanaat, Alessio Giorgi, Georg Grützner, Eero Hakavuori, Ville Kivioja, Terhi Moisala, Richard Montgomery, Luca Nalon, Runo-Mikael Ojala, Alessandro Ottazzi, Nicola Paddeu, Gabriel Pallier, Pierre Pansu, Alessandro Pilastro, Valto Purho, Cristian Sopio, Andrea Tettamanti, Francesca Tripaldi, Jeremy Tyson, and Davide Vittone.

Most of the research that led to this book was supported by the European Research Council under the ERC Starting Grant 713998 GeoMeG—*Geometry of Metric Groups*. This grant also funded the open-access publication of the book.

Fribourg, Switzerland Enrico Le Donne
February 2025

Competing Interests The author has no competing interests to declare that are relevant to the content of this manuscript.

Contents

Chapter 1
Introduction

This book is an exploration of Carnot-Carathéodory spaces through the perspective of Lie groups. It is intended to study these non-smooth geometries, focusing on the prototypical examples called Carnot groups. Carnot groups are a fundamental class of nilpotent Lie groups equipped with sub-Riemannian or, more generally, sub-Finsler structures. They play an essential role in various mathematical domains, including metric geometry and geometric group theory—as we shall see.

Carnot groups are particular examples of metric Lie groups, i.e., Lie groups equipped with left-invariant distances inducing the manifold topology. Moreover, metric Lie groups for which the distance is geodesic are precisely the sub-Finsler Lie groups. In several problems, it is more natural to consider the abstract setting of metric groups.

In this introductory chapter, we begin by providing a glimpse into the core concepts of Carnot-Carathéodory spaces: contact distributions and sub-Riemannian distances. The second section provides an outline of the book's content and structure, offering a roadmap for our readers. While experts in the field can directly proceed to the subsequent chapters, readers new to the topic will find guidance within these pages; see Fig. 1.1. In Sect. 1.3.1, we present a series of applications in mathematics, physics, and various other scientific disciplines where Carnot-Carathéodory spaces appear, underscoring the versatility and significance of these spaces in real-world scenarios.

1.1 What Sub-Riemannian Geometry Is

Sub-Riemannian geometry is also known as non-holonomic Riemannian geometry in Russia, and in France got the name Carnot geometry, or Carnot-Carathéodory, when it is appropriately generalized. Since the 1980s, it has been a full research domain, with motivations and ramifications in several parts of pure and applied

E. Le Donne, *Metric Lie Groups*, Graduate Texts in Mathematics 306,
https://doi.org/10.1007/978-3-031-98832-5_1

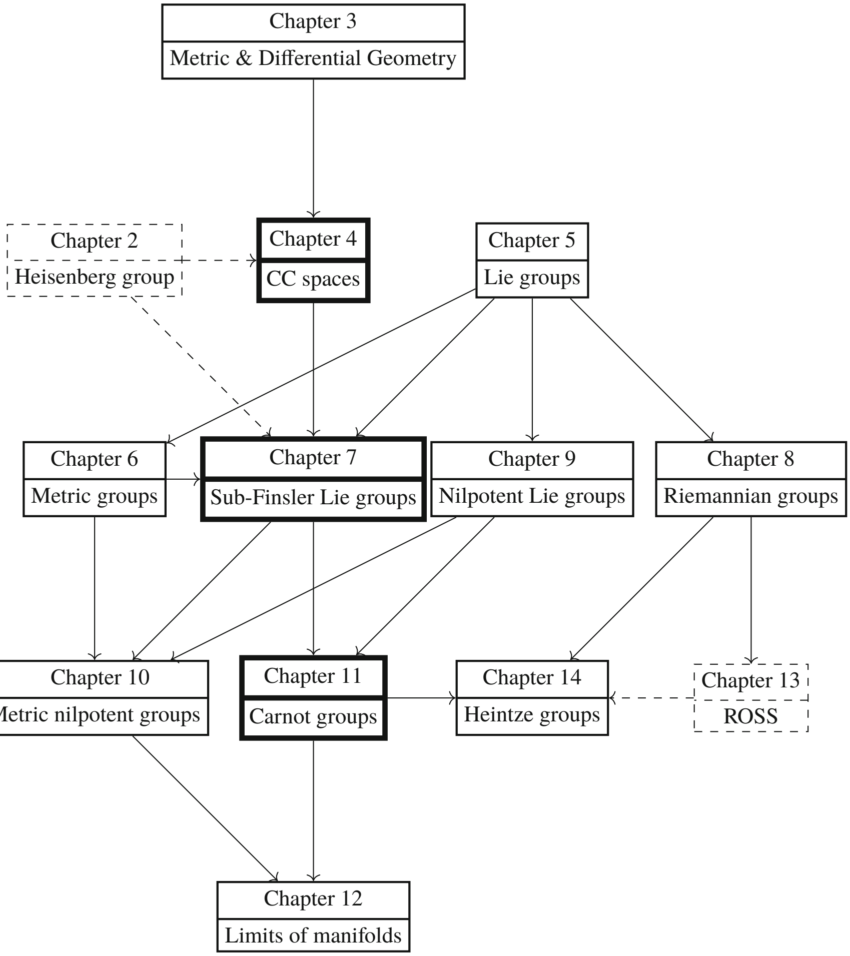


Fig. 1.1 Interdependence of chapters. Thick boxes represent the main chapters. Dashed boxes represent chapters devoted to examples

mathematics. However, historically, it was not clear that such theories were heading into the same notions. Thus, each source provided its own jargon in the field. Accordingly, some concepts have multiple terminologies: a contact structure is a particular distribution of hyper-planes in an odd-dimensional manifold, and the concept of Carnot-Carathéodory metric is a generalization of a sub-Riemannian distance.

Sub-Riemannian geometry is a generalization of Riemannian geometry. In addition to a Riemannian structure, in each sub-Riemannian manifold, there is a constraint on admissible velocities for curves. Geometrically, in Riemannian geometry, every smooth curve has locally finite length. In sub-Riemannian geometry, curves that fail to satisfy the constraint have infinite length.

Typical examples to keep in mind come from mechanics. Indeed, the state of a moving object is determined by its position in space and the velocities of its parts: the momenta. Thus, in the space 'positions and speeds' of configurations, the possible evolutions of the object should satisfy the fact that the derivatives of the first coordinates are equal to the second coordinates. For example, the movement of a single particle in space is described by a curve $t \mapsto (x(t), v(t))$ in $\mathbb{R}^3 \times \mathbb{R}^3$. However, not every curve in $\mathbb{R}^3 \times \mathbb{R}^3$ is possible. Some trajectories are not allowed by the dynamical constraint. As trivial examples, you cannot vary your speed without changing your position, or similarly, you cannot move into another place at speed zero!

The 3-dimensional (3D, for short) Heisenberg group equipped with its contact geometry is one of the most essential examples in sub-Riemannian geometry among those that actually are not Riemannian manifolds. Visualizing some of its features is relatively easy. As a set and topological space, the 3D Heisenberg group $\mathcal{H}$ is equivalent to $\mathbb{R}^3$. The constraint on curves in this space is determined by what is called a 'distribution of planes', or a 'rank-2 polarization'. Similar to how a smooth vector field X smoothly assigns a tangent vector X_p at each point p of a manifold, a distribution Δ of planes in $\mathbb{R}^3$ smoothly assigns to each point $p \in \mathbb{R}^3$ a plane Δ_p within the 3D tangent space at p. The curves that we call 'admissible' are those that are tangent to one such a distribution, in the sense that a smooth curve γ is *admissible with respect to a distribution* Δ if, for every t in the domain of the curve, the velocity vector $\dot{\gamma}(t)$ belongs to the plane $\Delta_{\gamma(t)}$. Refer to Fig. 1.2 for a visual representation of a distribution.

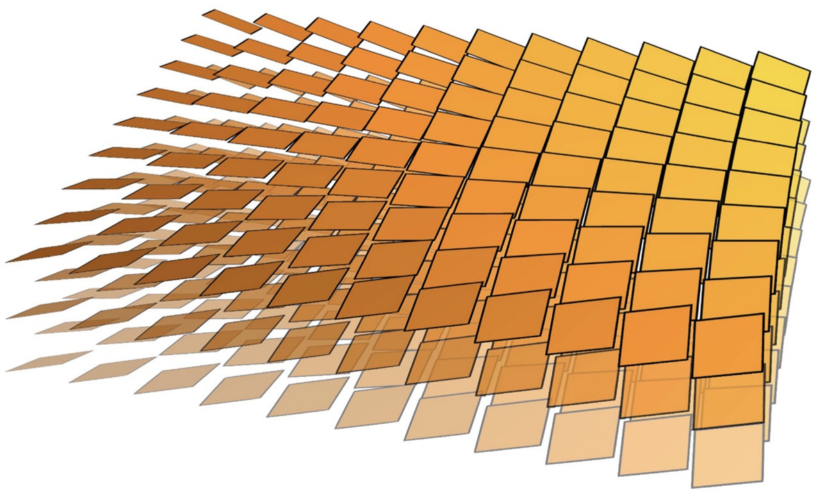


Fig. 1.2 A contact distribution on $\mathbb{R}^3$ is a polarization by planes

The particular feature of the Heisenberg group $\mathcal{H}$ is that it comes with a distribution that is curly enough in a way that each pair of points can be connected by at least one admissible curve. Therefore, one can define a finite-valued distance similarly to the Riemannian case: The distance between two points p and q in $\mathcal{H}$ is given by the infimum of the length of all admissible curves from p to q,

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

More generally, a sub-Riemannian manifold is a triple (M, g, Δ) consisting of a manifold M, a Riemannian tensor g inducing a length structure, and a distribution Δ of subspaces in the tangent bundle of M. These data define admissible curves and a distance as described in ($\star$). These geometries provide a broad generalization of Riemannian geometries in several aspects. Those sub-Riemannian spaces that are non-Riemannian, in the sense that we are dealing with a proper distribution, are rather different from Riemannian spaces: They exhibit fractal properties as their Hausdorff dimension exceeds the topological dimension. Additionally, there exist smooth curves with locally infinite length, as well as other smooth admissible curves that are isolated in the topological space of smooth admissible curves with the same endpoints. Consequently, sub-Riemannian geometry requires techniques that are different from those used in Riemannian geometry.

According to Mikhail Gromov, [Gro96], the concept of sub-Riemannian distance can be traced back to the ideas of Nicolas Carnot in 1824 and Constantin Carathéodory in 1909; the early references are [Car24, Car72, Car09]. For this reason, 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, metric geometry, group theory, and the analysis of hypoelliptic differential operators.

1.2 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 possibly non-Riemannian geometries manifest in other mathematical domains, such as metric geometry and geometric group theory, through Carnot groups, which are key examples to keep in mind. Carnot groups are a class of nilpotent Lie groups equipped with sub-Riemannian or, more generally, sub-Finsler structures. This text aims to explain the role of Carnot groups as asymptotic cones of finitely generated nilpotent groups. We also explore their presence as parabolic boundaries of rank-one symmetric spaces and their involvement as limits of Riemannian manifolds and tangents of Carnot-Carathéodory spaces.

In Chap. 2, we will focus on the plane distribution in the 3D Heisenberg group $\mathcal{H}$. We will consider the induced distance (⋆). Specifically, we will discuss the following facts:

1. This distance d turns the space $\mathbb{R}^3$ into a metric space. The topology associated with d is the standard topology of $\mathbb{R}^3$. In particular, nearby points can be connected by short admissible curves.
2. The distance between every two points is equal to the length of some admissible 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 bi-Lipschitz equivalent to any Riemannian manifold. This is because the Heisenberg geometry exhibits characteristics of fractal geometry. Indeed, the metric on this topologically 3-dimensional space has a metric dimension equal to 4, as determined by its Hausdorff 4-measure.
4. The geometry is homogeneous. Actually, it is invariant under a smooth group structure.

The general definition of a Carnot-Carathéodory space arises when we formally define the concept of a distribution being 'curly enough'. This property should ensure that every pair of points can be connected by an admissible curve. To explore this topic, we require basic notions of both differential geometry and metric geometry, which we review in Chap. 3.

Then, in Chap. 4, we delve into the sub-Finsler geometry of Carnot-Carathéodory spaces, focusing on their distributions, distances, and dimensions. A *distribution* on a manifold M refers to a sub-bundle 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. Constant-rank distributions are also referred to as *polarizations*. A distribution $\Delta \subseteq TM$ is said to be *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 $X_1, X_2, X_3, \dots$ are tangent to Δ, and $v \in T_pM$; and we use the standard notation $[X, Y]$ for the Lie bracket of two vector fields X and Y.

A *sub-Riemannian 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, the map 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* is defined by the same formula (⋆). More generally, if the length comes from a smoothly varying norm, then the distance is called *Carnot-Carathéodory metric* or *sub-Finsler*. Most of the previously mentioned results on the Heisenberg group will be valid for every Carnot-Carathéodory space.

The understanding of many Riemannian geometric properties comes from the fact that 'metric' tangent spaces T_pM, as $p \in M$, of each Riemannian manifold M are Euclidean spaces, and Euclidean geometry is sufficiently understood. Such a notion of tangent space is 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, as we will discuss in Sect. 4.1.5, for each 3-dimensional (non-Riemannian) sub-Riemannian manifold M, each T_pM is isomorphic to the Heisenberg group—another reason for it to be important. In general, alas, for every $n \geq 7$, the possible tangents of sub-Riemannian manifolds of topological dimension n can comprise infinitely many distinct spaces. Moreover, the tangents may differ from point to point for some given sub-Riemannian manifold. The good news is that, analogously to the Heisenberg structure having a group structure, the metric tangent of a Carnot-Carathéodory space M has a Lie group structure for most points $p \in M$, and at the remaining points, it is still a quotient of some Lie group modulo a closed subgroup. 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 called Carnot groups.

We shall review the necessary theory of Lie groups in Chap. 5. Chapter 6 introduces the general objects that combine a group structure with metric geometry: metric Lie groups and, more broadly, isometrically homogeneous spaces. In Chap. 7, we then discuss the geometry of sub-Riemannian Lie groups and, more generally, sub-Finsler Lie groups. This chapter is fundamental since we present Chow's theorem about the fact that Carnot-Carathéodory distances induce the manifold topologies, and we discuss the notion of endpoint map, which leads to a first-order study of geodesics in sub-Riemannian manifolds. Chapter 8 is a side excursion into the classical basic theory of Riemannian Lie groups. Before getting into Carnot groups, we review many properties of nilpotent Lie groups in some detail. Respectively, in Chap. 9, we discuss the classical differential geometry of nilpotent Lie groups, while Chap. 10 is devoted to their metric geometry. In Chap. 11, we finally define and study Carnot groups.

The aim of the rest of the book is twofold:

[Aim 1] We explore the role of Carnot groups with their Carnot-Carathéodory distances in other mathematical areas. Namely, they appear as

(A) limits of Riemannian manifolds, asymptotic cones of nilpotent groups, and tangents of Carnot-Carathéodory spaces; see Chap. 12.
(B) parabolic boundaries of rank-one symmetric spaces and other negatively curved spaces; see Chap. 14.

In harmonic analysis on stratified Lie groups, and more generally on graded groups, Carnot-Carathéodory distances enter the study of hypoelliptic differential operators. In complex analysis, Carnot-Carathéodory spaces occur as boundaries of strictly pseudo-convex complex domains. We do not treat these last two settings in this monograph but refer to the books [Ste93, Cap+07] as initial references.

As we will explain in Sect. 12.8, 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 such 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 Carnot-Carathéodory metric, and eventually gives an algebraic consequence: the group is (virtually) nilpotent.

There is a general natural explanation for why Carnot groups appear in the above situations. The reason is that Carnot groups are the analogs of finite-dimensional normed vector spaces in the non-commutative case. Indeed, Carnot groups admit the following axiomatic characterization:

Theorem A *The sub-Finsler Carnot groups are the only metric spaces that are locally compact, geodesic, isometrically homogeneous, and admit metric dilations.*

In Chap. 11, we will prove this result. We will also previously discuss the more general setting where the geodesic assumption is replaced with connectedness; see Sects. 6.5 and 10.2.

[Aim 2] With the use of Lie group theory, one may develop calculus, analysis, geometric measure theory, calculus of variations, and geometric analysis on Carnot groups and, more generally, on sub-Finsler Lie groups. In Chap. 11, we shall prove some crucial results in this regard.

This book's approach is to focus on studying specific examples of tangent spaces of sub-Riemannian manifolds (i.e., Carnot groups) to shed light on the general case of Carnot-Carathéodory spaces. The reason for this perspective is that Carnot groups provide tools for developing calculus in such settings, thanks to the availability of translations by group elements and the dilation property. The classical definition of the derivative of a real function $f : \mathbb{R} \to \mathbb{R}$ 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 derivative known as the *blow-up differential*, or *Pansu derivative*, named after Pierre Pansu, who made pioneering contributions to the field in the late 1980s; see [Pan89]. The main differentiability result, obtained by Pansu and then generalized in [MM95, Vod07], is the following:

Theorem B (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 Chap. 11 for Carnot groups. In fact, the theorem also holds for quasi-conformal maps between Carnot-Carathéodory spaces; see [Vod07]. The theory of quasi-conformal mappings has been used in [Pan89] to prove rigidity theorems on hyperbolic spaces over the division algebras of real, complex, or quaternionic numbers. Indeed, as we shall see in Chap. 14, the 'parabolic visual boundaries' of rank-one symmetric spaces are Carnot groups. More generally, all negatively curved homogeneous Riemannian manifolds have graded groups as boundaries. This last fact is mainly based on the work of Heintze. In Chap. 13, we discuss rank-one symmetric spaces, which can be seen as semi-direct products of Lie groups. One of the factors of the semi-direct product is a Carnot group of Heisenberg type, and the other one is the one-dimensional simply connected Lie group acting on the Carnot group by its dilations. In Chap. 14, we review the notion of boundaries of CAT(-1) spaces and their visual boundaries. First, we observe that the boundaries of rank-one symmetric spaces are the particular Carnot group of Heisenberg type. Second, we discuss the general viewpoint of Heintze groups and show that their boundaries are metric Lie groups admitting dilations.

1.3 Sub-Riemannian Geometries as Models

In this section, we highlight several contexts in which Carnot-Carathéodory metrics find applications: either as specific examples, generalizations, or tools. While this overview is not comprehensive, it provides specific references for readers interested in delving deeper into these areas.

1.3.1 Examples from Mathematics

1.3.1.1 Control Theory and Metric Geometry

Control theory is an interdisciplinary branch of engineering and mathematics that deals with the behavior of dynamical systems. The primary objective is to manipulate a differential system by choosing control variables to guide trajectories toward a desired state, possibly optimally. Sub-Riemannian geometry specifically focuses on control systems that are linear in the controls. Many theorems in sub-Riemannian geometry have broader validity in control theory. For instance, foundational sub-Riemannian theorems such as the ones by Chow, Pontryagin, and Goh have more general statements within the framework of geometric control theory. Readers interested in this perspective should refer to the book [AS04].

However, the distinctive characteristic of sub-Riemannian geometry is the presence of the induced metric, which transforms these manifolds into metric spaces. In this way, methods from metric geometry become applicable in the study of sub-

Riemannian geometry, and sub-Riemannian geometries, in turn, offer intriguing examples of metric spaces, sometimes exhibiting strange or even pathological behavior. This geometric perspective on sub-Riemannian geometry was pioneered by Gromov in his seminal work [Gro96].

1.3.1.2 Geometric Group Theory and Asymptotic Geometry

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 with polynomial growth. A group Γ, generated by a finite set S, is said to *grow polynomially* if the cardinality of the product set S^n is polynomially bounded in $n \in \mathbb{N}$. One verifies easily that this property does not depend on the choice of the finite generating set S. The asymptotic properties of these groups are intricately connected to the geometry of their asymptotic cones; see Fig. 1.3.

For instance, according to a theorem by Pansu, as presented in Sect. 12.8 of this book, each nilpotent group Γ generated by a finite set S has polynomial growth, and there exist constants $Q \in \mathbb{N}$ and $V > 0$ such that

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

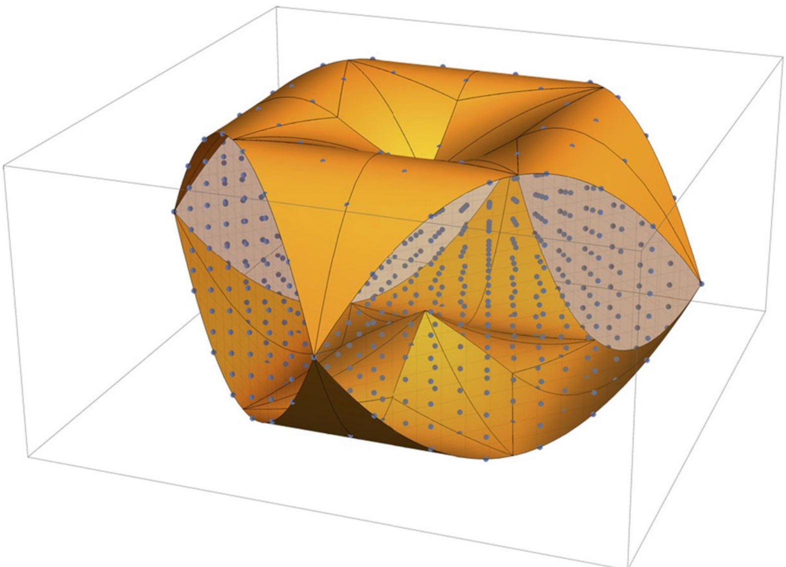


Fig. 1.3 A Carnot-Carathéodory ball, which is the limit of large balls for a word distance on a finitely generated nilpotent group. See Example 12.8.3

Here, the numbers Q and V possess clear geometric significance: Q is the Hausdorff dimension of the asymptotic cone of Γ, and V is the volume of the unit ball in the asymptotic cone. Furthermore, this asymptotic cone is an example of a Carnot group, equipped with a Carnot-Carathéodory distance. These are the spaces on which this book focuses. See Chaps. 11 and 12.

1.3.1.3 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, of great importance in analysis of several complex variables, are naturally equipped with visual distances that are of Carnot-Carathéodory type.

A domain in $\mathbb{C}^n$ is called *strictly pseudo-convex* if, near 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 [Gro96, Ma91, BB00, Pil22].

The main example that the reader should have in mind is the unit ball $\mathbb{B}$ in $\mathbb{C}^n$. On the one hand, it is an example of a strictly pseudo-convex domain. On the other hand, it is a rank-one symmetry space: Namely, $\mathbb{B}$ equipped with its Kobayashi-Carathéodory distance is the complex hyperbolic space. The real, complex, and quaternionic hyperbolic spaces, together with the octonionic plane, exhaust the list of the rank-one symmetry spaces of non-compact type. All of these spaces carry a natural metric on the boundary, which is a Carnot-Carathéodory distance. We present rank-one symmetry spaces from the Lie group viewpoint as semi-direct products in Chap. 13. Then, in Chap. 14, we discuss the visual boundaries of these spaces and, more generally, of all the negatively curved homogeneous spaces.

1.3.1.4 Analysis of Hypoelliptic Operators and Singular Integral Operators

The theory of partial differential equations (PDEs) deals with a wide variety of equations, each requiring different approaches. One of the most important operators is the Laplacian, which, in Euclidean space $\mathbb{R}^n$, is defined as:

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

This operator is called *hypoelliptic* because it has the property that

$$\Delta f \in C^{\infty} \qquad \Longrightarrow \qquad f \in C^{\infty}.$$

There is a connection between the Laplace 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, the Laplacian generalizes to the *sub-Laplacian*. An example is the operator on functions on $\mathbb{R}^3$ given by

$$f \longmapsto \Delta_{\text{sub}} f := X(Xf) + Y(Yf),$$

where

$$X := \frac{\partial}{\partial x_1} - x_2 \frac{\partial}{\partial x_3} \quad \text{and} \quad Y := \frac{\partial}{\partial x_2} + x_1 \frac{\partial}{\partial x_3}.$$

The sub-Laplacian takes into account the non-holonomic constraints imposed by the vector fields. In fact, it is intrinsically linked to the geometry of the sub-Riemannian metric obtained from the bracket-generating vector fields. Namely, certain regularity bounds for the sub-Laplacian, and many other subelliptic PDEs, can be obtained in terms of the Sobolev spaces with respect to the associated sub-Riemannian distances.

This relation is used to study diffusion processes, heat equations, singular integrals, and other differential equations on such manifolds or on spaces of homogeneous type. For more on the analysis of hypoelliptic operators, one can consult the following references: [Fol73, RS76, Goo76, NSW85, Cap97].

1.3.1.5 Classical Mechanics and Non-Holonomic Dynamical Systems

Sub-Riemannian geometry plays a significant role in classical mathematical mechanics, particularly when studying mechanical systems with constraints. These systems, referred to as *non-holonomic systems*, involve specific limitations on velocities or accelerations, constraining the possible motions of the system according to defined rules. Sub-Riemannian geometry offers a comprehensive framework for analyzing and understanding the geometric properties of these constrained mechanical systems. Some classical references on non-holonomic dynamical systems are [Tay34, Wag40, Gol51, Bro82]. A recent review is [BMB16]; see also [Blo03, ZB03].

1.3.1.6 Contact Geometry, Engel Structures, and Exterior Differential Systems

Contact structures, fundamental in various mathematical contexts, have their roots in Carathéodory's formalization of thermodynamics [Car09]. They are instrumental in understanding processes like the *Carnot cycle*, where work is transformed into heat due to the bracket-generating property of the distribution that is constraining the dynamics.

Contact manifolds are particular polarized manifolds that involve certain distributions on odd-dimensional manifolds. The distributions are hyperplane fields defined by certain differential one-forms, called *contact forms*. The Heisenberg group, as discussed in Chap. 2, is the standard example of a contact structure. Introductions to the subject can be found in [Gei08, Etn03].

In the realm of low-dimensional topology and geometry, equiregular bracket-generating distributions of dimension 2 on 4-dimensional manifolds are of particular interest. These are known as *Engel structures*; see Sect. 11.1.3 for the model space: the *Engel group*. For further insights into these structures and their relation to contact structures, refer to [Mon99].

The theoretical foundations of contact structures are rooted in *exterior differential systems*, systems of equations on a manifold defined by exterior differential forms. This theory, pioneered by Élie Cartan [Car45], serves as a general framework for understanding contact manifolds and many other polarized manifolds. For a comprehensive exploration of this subject, we refer to [Bry+91].

In some sense, sub-Riemannian geometry metrizes contact manifolds and, more generally, polarized manifolds, similar to how Riemannian geometry metrizes differentiable manifolds. In the presence of a distance function, problems concerning the contact equivalence of contact structures transform and extend to queries like bi-Lipschitz equivalence or quasi-conformal equivalence.

1.3.1.7 Riemannian Geometry

Sub-Riemannian geometry is a natural generalization of Riemannian geometry. Sub-Riemannian metrics often emerge as limit cases of Riemannian metrics.

A crucial instance of this phenomenon occurs in the asymptotic cones of Riemannian Lie groups with polynomial growth. Specifically, these Lie groups are of type (R), and their asymptotic cones are Carnot groups. Further details on these results will be explored in Sects. 6.4.1 and 12.8.

1.3.1.8 Univalent Function Theory

There is a remarkable recent application of sub-Riemannian geometry to univalent function theory. The following quick summary is based on the article [MPV07] and on kind help from Jeremy Tyson.

Classical univalent function theory considers the class $\mathcal{S}$ of injective analytic functions $f : \mathbb{D} \to \mathbb{C}$ defined on the unit disc $\mathbb{D}$ in the complex plane $\mathbb{C}$, normalized by $f(0) = 0$ and $f'(0) = 1$. A major subject of investigation is the so-called *coefficient body*:

$$M := \left\{ (c_k) \in \mathbb{C}^{\mathbb{N}} : f \in \mathcal{S},\, f(z) = z\left(1 + \sum_{k=1}^{\infty} c_k z^k\right) \right\}.$$

This set can be seen as the limit of its *finite-dimensional slices*:

$$M_n := \left\{ (c_1, \ldots, c_n) : f \in \mathcal{S},\, f|_{\mathbb{S}^1} \in C^{\infty}(\mathbb{S}^1),\, f(z) = z\left(1 + \sum_{k=1}^{\infty} c_k z^k\right) \right\}.$$

By a famous result by de Branges [dBr85] (formerly known as the Bieberbach conjecture), we have $|c_n| \leq n + 1$, for all n. This gives information on the size of M_n and M. However, there is no explicit description of M_n except for the trivial case $n = 2$ and the case $n = 3$; see [SS50].

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, called SLE, which stands for either the *stochastic Loewner equation* or for *Schramm–Loewner evolution*, a topic at the intersection of probability, complex analysis, stochastic PDE, and mathematical physics; see [Law04, Sch00, Sch07, Aba+10].

Irina Markina, Dmitri Prokhorov, and Alexander Vasilev used the Loewner flow on $\mathcal{S}$ to define a natural (partially integrable) Hamiltonian system on each set M_n. They found certain first integrals of the flow and calculated all the relevant commutators. Consequently, they constructed a complex sub-Riemannian structure on M_n that is naturally adapted to the underlying univalent function theory. In fact, the Loewner trajectories 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 include SLE as well as the classical (deterministic) Loewner equation. We refer to [Bra+14] for more on classical and stochastic Loewner-Kufarev equations.

1.3.1.9 Diffusion Processes and Financial Mathematics

Sub-Riemannian geometry has also found applications in the field of diffusion processes and stochastic calculus, particularly in the modeling and analysis of 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 the author by Josef Teichmann and by Andrea Pascucci; see, for example [FV10, CST23, Pas11, Pas05].

In mathematical finance, one often deals with controlled ordinary differential equations and with stochastic processes, in particular, with Ito diffusions. It is well known that Ito diffusions can be related to Riemannian and sub-Riemannian structures. When high-dimensional markets are modeled, it is quite natural to assume that the instantaneous covariance matrix is degenerate but that densities still exist, which leads to sub-Riemannian structures.

Another significant stream of literature involving sub-Riemannian structures is rough path theory and its applications, ranging from mathematical finance to machine learning. In its very foundations, free-nilpotent Lie groups play a role, for instance, when strategies are approximated by signatures, which are natural maps with values in nilpotent Lie groups.

To learn more about signature-kernel methods in machine learning, shuffle product, and SiG stochastic differential equations in mathematical finance, the reader might consult the following references: [Bay+23, CL16b, CO22, CGSF23, Cuc+22, FH20, FV10, HL10, KO19, LNP19, LNP20, Sal+21].

Sub-Riemannian geometry has been employed to model the evolution of asset prices and the dynamics of some financial markets. For example, in the Asian market, the price of the options depends on the evolution. Namely, asset-price equations are a class of path-dependent options characterized by a payoff that is a function of the history of the underlying asset price. The equations are strongly degenerate partial differential equations in three dimensions; see [AMP21, Kim09, BPV01].

In a simplified exposition of Asian option pricing, we may say that some variables are the integrals of other variables. Formally, we have a differential constraint implying that the dynamics are tangent to a proper subbundle of the manifold of the variables. The subbundle is bracket-generating. The equation that gives the dynamics is known as the Black–Scholes equation. It is an equation of Kolmogorov-Fokker-Planck type, which is a type of sub-Riemannian Laplacian with respect to Hörmander vector fields. When looking at the fundamental solution of these equations, the sub-Riemannian metric comes naturally.

The Black & Scholes model for the pricing problem goes back to the early 1970s; see the reprint [BS12]. In 1973, Myron Scholes, in close collaboration with Fischer Black, developed what we call nowadays the *Black-Scholes model*. For his contributions, he earned the 1997 Nobel Prize in economics. The model is generally used to determine the fair price or theoretical value of a call or a put option. In the case of path-dependent options, the model leads to Kolmogorov operators associated with some stochastic processes.

General references, also in connection with stochastic equations, are [Pas11, Chapter 9.5] and the review article [Pas05]. In addition, we suggest the following topics and references. For applications to Asian options, there is [BPV01], while for American options, there is [Pas08]. For path-dependent volatility models of asset prices, stochastic models of stock prices, originally underlying models, and the geometric Brownian motion with constant volatility, see [FP08, HR98]. For an introduction to the probabilistic theory of arbitrage pricing of financial derivatives, including stochastic optimal control theory, optimal stopping theory, and arbitrage theory in continuous time, we suggest [Bjö19], as many other books by Tomas

Björk. The paper [Pas05] contains a survey of results about partial differential equations of Kolmogorov type arising in physics and mathematical finance.

1.3.2 Examples from Physics

Sub-Riemannian geometry models various structures beyond pure mathematics, from finance to mechanics, from bio-medicine to quantum phases, from robots to falling cats! Lacking space as well as competence, in this section, we do not enter into details. Instead, we direct the interested reader to suitable references.

1.3.2.1 Geometry of Principal Bundles with Connections: Falling Cats

Theoretical physics describes most mechanical systems by configuration spaces. Gauge theory, also known as the geometry of principal bundles with connections, studies systems with physical symmetries, i.e., when there is a group G acting by isometries on the configuration space M. Most of the time, it is easier to understand the *dynamics up to isometries*. Namely, one first studies the system trajectories in the quotient space M/G. Subsequently, one has to study the 'lift' of the dynamics into the initial configuration space to obtain the trajectories in M. Such lifts will be subject to a sub-Riemannian restriction. This viewpoint is elaborated in [Mon02, Part II].

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

To explain the system, 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 large dimensions. The position of a cat is determined by its shape plus its orientation in space. Otherwise said, the group $G := \mathrm{Isom}(\mathbb{R}^3)$ of isometries of the Euclidean 3D space acts on the configuration space M, and the shape space is just the quotient of the action:

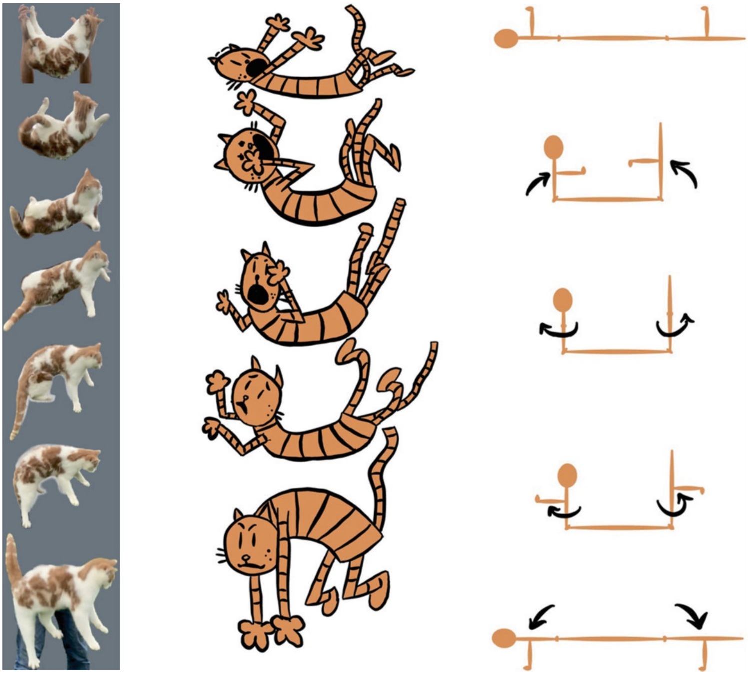
$$\pi : M \to M/G = S.$$

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

The key fact is that the cat has considerable 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

$$\pi \circ \tilde{\sigma} = \sigma.$$

Moreover, the lifted curve is unique: it has to satisfy the constraint given by what is called the *natural mechanical connection*. What the cat is demonstrating is that



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

Fig. 1.4 The cat spins itself around and rights itself

such a 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, resulting in a change in configuration: the legs initially point towards the sky, but in the end, they are toward the floor. We refer to [Mon02, Section 14.2] for an explanation of an engineers' model for the cat and to Fig. 1.4 for an action shot, a cartoon, and a schematic drawing of the drop.

1.3.2.2 From Mechanics: Parking Cars, Rolling Balls, Moving Robots, and Satellites

Sub-Riemannian geometry has been extensively used in 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.

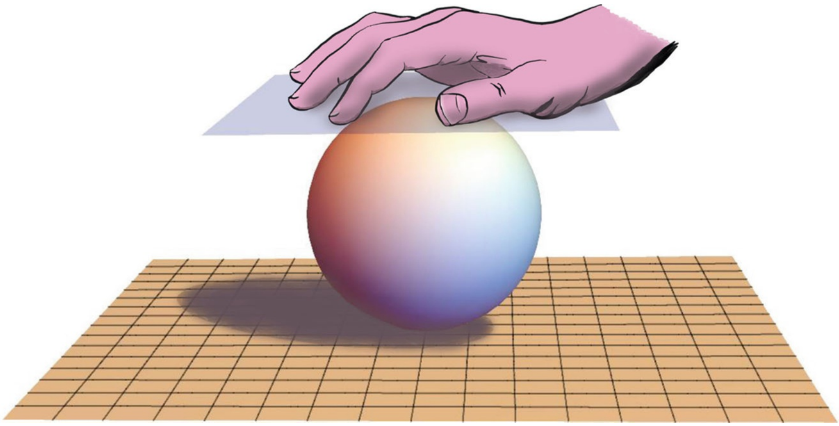


Fig. 1.5 A ball rolling on the plane without sliding, nor spinning

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 some fixed line. However, the driver has only two degrees of freedom: turning and rolling. Again, using non-trivial holonomy, we can move the car to any position we like. In this model, we only parametrize the back wheel. Taking into account also the front wheel, we get a 4D manifold with still only two controls.

Rolling a Ball on the Plane The position of a ball lying on a plane is described by five coordinates: two real numbers to characterize the point on the plane that touches the ball, another two spherical coordinates describing the point of the ball that touches the plane, and finally, an angular coordinate for spinning the ball around its vertical axis. When one moves the ball without sliding, there are only three admissible control movements: two for rolling the ball along a planar vector, and the third one for spinning it. One can also omit the possibility of spinning and have only a 2-dimensional space of control variables; see Fig. 1.5. In both cases, the ball can arrive at every position regardless of its initial position. This problem has been studied intensively. One is interested in minimizing the length of the curve traced on the plane. The locally optimal trajectories are known and are Gauss' elastic curves. However, it is not known up to what length each such curve is optimal. To more information see [Jur93, Sac10, Sac08, MB00].

In *robotics*, mechanisms are subject to constraints of movement that do not restrict the manifold of possible positions, like for the arm of a robot. We refer to [BJR98, Jea14, MLS94, GZ06] for the motion planning problem. The situation of *satellites* is similar. One should really think about a satellite as a falling cat: it should choose its strategy of modifying the shape properly to have the necessary change in configuration. Another similar example is the case of an astronaut in outer space. See [BP04, Tré12] for applications to aerospace engineering.

1.3.2.3 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.

The following application comes from conversations with Ugo Boscain and from reading his 'Habilitation à diriger des recherches', [Bos06].

Let $\mathtt{H}$ be a separable complex Hilbert space. Let us denote by $\mathbb{S}$ the unit sphere in $\mathtt{H}$. In quantum mechanics, the time evolution of a 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 $\mathtt{H}$ called the *free Hamiltonian*.

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

Assume now to act on the system with some external fields (e.g., an electromagnetic field) whose amplitude is represented by some functions $u_1, \ldots, u_m \in L^\infty(\mathbb{R}; \mathbb{R})$. 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 $u_1, \ldots, u_m$ 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); see [DP10, D'A08, WM10, DP10].

For a finite-dimensional quantum mechanical system, if n is the number of energy levels, then we have $\mathtt{H} = \mathbb{C}^n$ 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 can be formulated with matrix formalism. The solution is of the form

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

where $\mathrm{SU}(n)$ standardly denotes the special unitary group of degree n. The Schrödinger equation becomes $\frac{\mathrm{d}}{\mathrm{d}t} g(t) = -iH(t)g(t)$, 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 $\mathrm{SU}(n)$ 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:

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

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 are to be minimized are either the energy transferred by the controls to the system or the time of transfer.

1.3.2.4 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 interested reader should start with the introduction and Chapter 13 in [Mon02] and the references therein.

1.3.3 Scientific Applications

1.3.3.1 Neurobiology

Neurobiological research over the past few decades has enhanced our understanding of the functional mechanisms of the first layer of the visual cortex. This layer comprises various types of cells, including the so-called 'simple cells', which are sensitive to specific orientation-based brightness gradients. Recently, this cortical structure has been modeled using a sub-Riemannian manifold. The following summary of this sub-Riemannian application is based on [SCP08] and [Pet17], as well as discussions with S. Pauls and G. Citti.

The modeling space is $\mathbb{R}^2 \times \mathbb{S}^1$, where each point (x, y, θ) represents a cell associated with the point of retinal data $(x, y) \in \mathbb{R}^2$ and oriented according to the angle $\theta \in \mathbb{S}^1$. Namely, the vector $(\cos\theta, \sin\theta)$ indicates the direction of the maximum rate of change of brightness at the point (x, y) in the image perceived by the eye. In pictures with high contrast, this vector is normal to the image contours.

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

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

These vector fields are left-invariant under a Lie group structure. The Lie group is also known as *rototranslation group*. Researchers think that if a piece of the contour of a picture is missing to the eye vision (possibly by being 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 in the space of visual cells, and the brain constructs a sub-Riemannian geodesic between the endpoints of the missing data. For more about the neurogeometry of vision, we refer to the book [Pet17].

1.3.3.2 Image Processing and Computer Vision Tracking

Sub-Riemannian geometry has been applied in image processing and computer vision. 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.

Image processing has benefited from sub-Riemannian geometric methods with the use of geometric flows [Cit+16, Bek+18], hypoelliptic diffusion [Bos+14], and energy minimization [MAS13]. For anthropomorphic image reconstruction, see [Bos+12].

A very applied problem that is currently tackled with sub-Riemannian methods is the tracking of blood vessels on spherical images of the retina. Namely, given (several) images of eyes, we seek algorithms that recognize the blood vessels as curves that possibly pass one over the other. The problem can be solved via sub-Riemannian geodesics, see [Mas+17, Bek+17]. Using nilpotent approximations (namely, Carnot groups as studied in this book), researchers have obtained fast perceptual grouping of blood vessels in 2D and 3D, see [BCP18].

1.3.3.3 Neuroscience: Brain Modeling and Deep Learning

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 analyze brain connectivity and information processing.

The theory of non-isotropic propagation describes the brain propagation of signals that occur along the dense network of axons that constitutes the neural connectivity. Visual perception phenomena are expressed as anisotropic partial differential equations. The first differential models of the cortex go back to [Hof70, Mum94, PT99]. These geometric brain models have been developed in [CS06, SC15, Coc+15], where each family of cells is described via a Lie group with a sub-Riemannian metric. See also [DF10a, DF10b, Dui+23] for results in image analysis.

The geometric analysis of sub-Riemannian Lie groups led to group-invariant brain-inspired architectures of deep learning: group-equivariant convolutional neural networks. The considerable advances are in terms of parameters to be trained; see [CW16, Bek+18, SM14]. A more recent brain-inspired architecture is PDE-group equivariant convolution neural networks [Dui+21], which have been shown to have major advantages in terms of efficiency.

Chapter 2
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.

2.1 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 for being 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 is idealized as a straight line. Actually, whatever the shape of the coastline is, the leather string will take the shape of an arc of a circle.

E. Le Donne, *Metric Lie Groups*, Graduate Texts in Mathematics 306,
https://doi.org/10.1007/978-3-031-98832-5_2

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

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

where the latter is the expression in polar coordinates (r, θ), where $x = r\cos\theta$ and $y = r\sin\theta$. 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:

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

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 Fig. 2.1 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 $\int_\gamma \mathrm{d}s$, which expresses the length of the curve obtained by integrating it with respect to the element of arc length $\mathrm{d}s$.

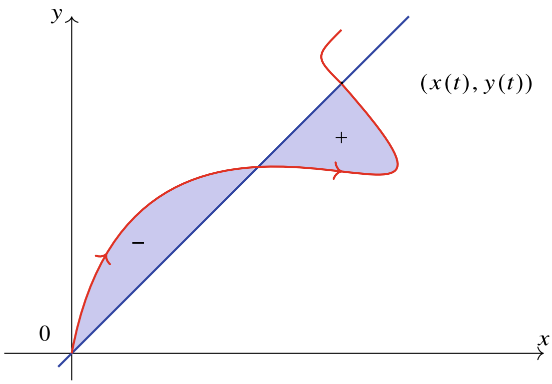


Fig. 2.1 The lift of the curve is performed by defining the third coordinate $z(t)$ as the oriented area of the region between the arc of the curve up to the point $(x(t), y(t))$ and the straight segment from $(0, 0)$ to $(x(t), y(t))$

2.2 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 $\sigma_{[0,t]}$, obtained by restricting σ to the interval $[0, t]$, and the segment connecting 0 to $(x(t), y(t))$, as in Fig. 2.1. Namely, we have

$$z(t) := \int_{\sigma_{[0,t]}} \alpha = \int_{\sigma_{[0,t]}} \frac{1}{2}(x\,\mathrm{d}y - y\,\mathrm{d}x) = \int_0^t \frac{1}{2}\,(x(s)\dot{y}(s) - y(s)\dot{x}(s))\,\mathrm{d}s. \tag{2.1}$$

Differentiating in t we get

$$\dot{z} = \frac{1}{2}(x\dot{y} - y\dot{x}). \tag{2.2}$$

Set $\xi = \xi_{(x,y,z)} := \mathrm{d}z - \frac{1}{2}(x\,\mathrm{d}y - y\,\mathrm{d}x)$. Consider a curve $\gamma = (\gamma_1, \gamma_2, \gamma_3) = (x, y, z) : [0, 1] \to \mathbb{R}^3$ starting at 0. Then, we have that such lifted curves satisfying (2.2) are exactly those satisfying $\xi_\gamma(\dot{\gamma}_1, \dot{\gamma}_2, \dot{\gamma}_3) \equiv 0$, which is $\dot{\gamma} \in \ker(\xi)$.

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

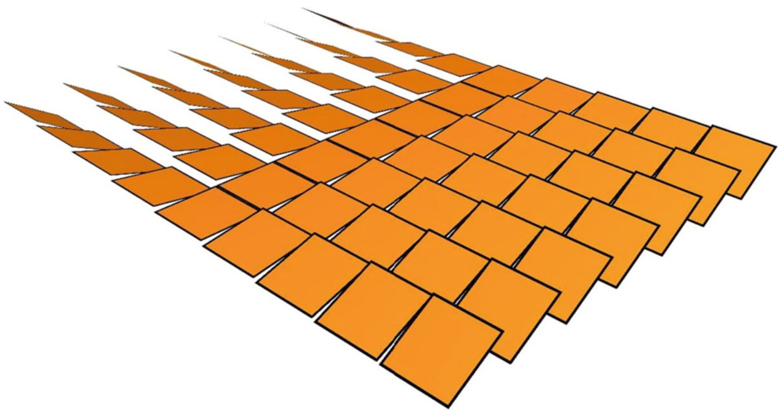


Fig. 2.2 Standard contact distribution on $\mathbb{R}^3$. We are drawing only the plane field at the points of height 0. One should understand that this distribution is invariant with respect to vertical translations, similar to Fig. 1.2

Definition 2.2.1 (Standard Contact Form) We refer to the differential one-form

$$\xi := \mathrm{d}z - \frac{1}{2}(x\,\mathrm{d}y - y\,\mathrm{d}x) = \mathrm{d}z - \frac{1}{2}r^2\,\mathrm{d}\theta \tag{2.3}$$

as the *standard contact form* in $\mathbb{R}^3$. More generally, a *contact form* on a $(2n+1)$-dimensional differentiable manifold M is a differential 1-form α, with the property that

$$\alpha \wedge (\mathrm{d}\alpha)^n \neq 0, \qquad \text{where} \qquad (\mathrm{d}\alpha)^n := \underbrace{\mathrm{d}\alpha \wedge \cdots \wedge \mathrm{d}\alpha}_{n}.$$

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*. See Fig. 2.2 for a visual representation.

As with every never-vanishing differential one-form on $\mathbb{R}^3$, the standard contact form (2.3) 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):

$$\Delta_{(x,y,z)} := \ker(\xi_{(x,y,z)}) = \left\{(v_1, v_2, v_3) \in \mathbb{R}^3 \;:\; v_3 = \frac{1}{2}(xv_2 - yv_1)\right\}. \tag{2.4}$$

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

$$\langle v, w\rangle := v_1w_1 + v_2w_2. \tag{2.5}$$

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-definite inner product. Alternatively, one can view this restriction as a restriction of a Riemannian tensor on $\mathbb{R}^3$, i.e., a positive-definite inner product on the entire tangent bundle of $\mathbb{R}^3$. In fact, we can consider the following frame[1] of $\mathbb{R}^3$:

$$\begin{cases} X := \frac{\partial}{\partial x} - \frac{1}{2}y\frac{\partial}{\partial z}, \\ Y := \frac{\partial}{\partial y} + \frac{1}{2}x\frac{\partial}{\partial z}, \\ Z := \frac{\partial}{\partial z}, \end{cases} \tag{2.6}$$

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

and declare it orthonormal. Let us verify that such a Riemannian metric gives the linear product (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

$$v = v_1 X + v_2 Y + \left(\frac{v_1}{2}y - \frac{v_2}{2}x + v_3\right) Z.$$

So, if $v \in \Delta_{(x,y,z)}$, we have $v = v_1 X + v_2 Y$ and thus (2.5) holds. We refer to Fig. 2.3 for a visual representation.

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, the tangent vector $\dot{\gamma}(t)$ must be in the plane $\Delta_{\gamma(t)}$, as defined in (2.4). For a given Legendrian curve γ, we define its length $L(\gamma)$ as the integral of the norm of $\dot{\gamma}$ using the scalar product (2.5). In simpler terms, the value $L(\gamma)$ is precisely 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 every pair of points p and q in $\mathbb{R}^3$, we define it as follows:

$$d_c(p, q) := \inf\{L(\gamma) \ : \ \gamma \text{ is a Legendrian curve between } p \text{ and } q\}. \tag{2.7}$$

An important fact to note is that because ξ was derived from Dido's problem, for every pair of points in $\mathbb{R}^3$, there are multiple Legendrian curves connecting them. In other words, it is always possible to find a Legendrian curve between any two points. Let's underline this crucial point:

A Crucial Fact Every pair of points in $\mathbb{R}^3$ is connected by a curve that is Legendrian with respect to the differential form ξ defined in (2.3).
In practice, to connect, for instance, $(0, 0, 0)$ to (x, y, z), it is enough to take a curve σ in $\mathbb{R}^2$ 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 connect $(0, 0, 0)$ to (x, y, z).

Furthermore, it's important to stress that the length of $\tilde{\sigma}$ is equal to the planar Euclidean length of σ. Consequently, there exists a correspondence between geodesics concerning the metric d_c (or, more accurately, curves that realize the infimum in (2.7)) and solutions to the *dual* Dido's isoperimetric problem: Fixed a value for the area, minimize the perimeter. Since it is relatively straightforward to find solutions to Dido's problem, we will be able to explicitly determine the geodesics of the metric space $(\mathbb{R}^3, d_c)$. We will delve into this topic further in Sect. 2.4.1.

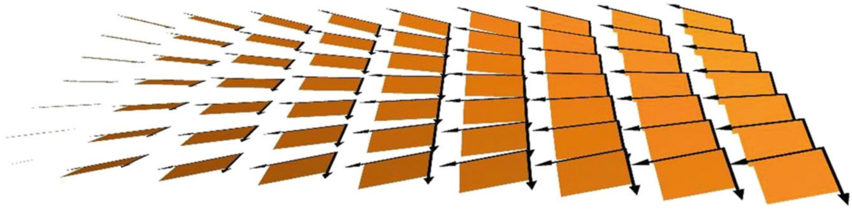


Fig. 2.3 The horizontal bundle spanned by the vector fields X and Y

2.3 The Heisenberg Group

2.3.1 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:

$$X(x, y, z) := \frac{\partial}{\partial x} - \frac{y}{2}\frac{\partial}{\partial z} = (1, 0, -\frac{y}{2}), \tag{2.8}$$
$$Y(x, y, z) := \frac{\partial}{\partial y} + \frac{x}{2}\frac{\partial}{\partial z} = (0, 1, \frac{x}{2});$$

at each point (x, y, z) we are considering $X(x, y, z)$ and $Y(x, y, z)$ as an orthonormal basis on their span $\Delta(z, y, z)$, as in Fig. 2.3; for each smooth curve $\gamma : [a, b] \to \mathbb{R}^3$ for which $\dot{\gamma}(t)$ is in $\Delta(\gamma(t))$ 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 $\int_a^b \sqrt{u_1(t)^2 + u_2(t)^2}\, dt$. Such a length structure defines the contact distance (2.7).

A crucial property of the contact geometry is that the space is metrically 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.

This group structure is named after Werner Heisenberg (1901–1976), a German theoretical physicist and pioneer of quantum mechanics. The Heisenberg group provides a geometric framework to represent Heisenberg's famous uncertainty principle. In our coordinates of $\mathbb{R}^3$, the group product of this structure is defined as follows:

$$(x, y, z) \cdot (x', y', z') := \left(x + x', y + y', z + z' + \frac{1}{2}(xy' - yx')\right). \tag{2.9}$$

One can easily check that (2.9) gives a group structure, and it turns $\mathbb{R}^3$ into a Lie group, i.e., multiplication and inversion are smooth maps. We will revisit the general theory of Lie groups in Chap. 5. We shall refer to the group $\mathbb{R}^3$ equipped with group law (2.9) 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 (2.6). Let us verify this claim for X. Fix a left translation f, say $f = L_{(s,t,u)}$, for $(s, t, u) \in \mathbb{R}^3$, i.e.,

$$\begin{aligned} f(x, y, z) &:= L_{(s,t,u)}(x, y, z) \\ &:= (s, t, u) \cdot (x, y, z) \\ &\overset{(2.9)}{=} \left(x + s, y + t, z + u + \frac{1}{2}(sy - tx) \right). \end{aligned} \tag{2.10}$$

The differential is

$$\mathrm{d}f = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -t/2 & s/2 & 1 \end{bmatrix}. \tag{2.11}$$

So, on the one hand, we have that $\mathrm{d}f(X)$ is given by:

$$\mathrm{d}f(X) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -t/2 & s/2 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ -y/2 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ -t/2 - y/2 \end{bmatrix} = \frac{\partial}{\partial x} + \left(-\frac{t}{2} - \frac{y}{2} \right) \frac{\partial}{\partial z}.$$

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

As a consequence of the fact that each left translation by the product (2.9) preserves the orthonormal frame $\{X, Y\}$ we deduce that each of these translations preserves the length of Legendrian curves and, consequently, preserves the contact distance as defined in (2.7).

The next proposition summarizes the above discussion.

Proposition 2.3.1 *The Heisenberg geometry is metrically homogeneous: the space has a Lie group structure in which each left translation acts as an isometry with respect to the contact distance d_c.*

The above model of the Heisenberg group offers 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:

$$t \in \mathbb{R} \longmapsto \gamma_v(t) = \exp(t(v_1, v_2, v_3)) = (tv_1, tv_2, tv_3), \qquad \text{for } (v_1, v_2, v_3) \in \mathbb{R}^3.$$

Furthermore, it is worth noting that all the lines through 0 in the xy-plane are curves that minimize the contact distance. This last statement is recommended as an exercise for novice readers.

2.3.2 The 3D Nilpotent Non-Abelian Matrix Group

The Heisenberg group can also be represented using matrices. It is a subgroup of the group of invertible matrices and can be defined as the set of 3×3 upper-triangular matrices equipped with the row-by-column matrix product:

$$\mathbb{G} := \left\{ \begin{bmatrix} 1 & a & c \\ 0 & 1 & b \\ 0 & 0 & 1 \end{bmatrix} \ : \ a, b, c \in \mathbb{R} \right\} < \mathrm{GL}(3, \mathbb{R}). \tag{2.12}$$

This matrix representation is particularly useful because, first, it simplifies the comprehension of the group structure. Second, it allows the Lie algebra associated with this group to be viewed as a matrix Lie algebra. Furthermore, the exponential of the Lie group corresponds to the classical matrix exponential.

The Lie algebra corresponding to this matrix group $\mathbb{G}$ is:

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

A basis of the Lie algebra is

$$X := \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \qquad Y := \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}, \qquad Z := \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}. \tag{2.13}$$

One parameter subgroups, as a, b, c vary in $\mathbb{R}$, are represented as:

$$\begin{aligned} t \in \mathbb{R} \longmapsto \gamma_{(a,b,c)}(t) &:= \exp\left(t \begin{bmatrix} 0 & a & c \\ 0 & 0 & b \\ 0 & 0 & 0 \end{bmatrix} \right) \\ &= I + t \begin{bmatrix} 0 & a & c \\ 0 & 0 & b \\ 0 & 0 & 0 \end{bmatrix} + \frac{t^2}{2!} \begin{bmatrix} 0 & a & c \\ 0 & 0 & b \\ 0 & 0 & 0 \end{bmatrix}^2 + \dots \end{aligned}$$

$$= I + t \begin{bmatrix} 0 & a & c \\ 0 & 0 & b \\ 0 & 0 & 0 \end{bmatrix} + \frac{t^2}{2} \begin{bmatrix} 0 & 0 & ab \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} + 0$$

$$= \begin{bmatrix} 1 & at & ct + abt^2/2 \\ 0 & 1 & bt \\ 0 & 0 & 1 \end{bmatrix}.$$

We claim that the map

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

is a Lie group isomorphism from the Lie group $\mathbb{R}^3$ with the product (2.9) to the Lie group $\mathbb{G}$ from (2.12) with the matrix product. Indeed, the map φ is a group homomorphism (a result of straightforward calculation), and its differential at the identity maps the left-invariant vector fields X, Y, Z from (2.6) to X, Y, Z from (2.13), respectively. In fact, the subsequent section will elaborate on this concept further.

2.3.3 Characterization of the Heisenberg Lie Algebra

The Lie algebra of the Heisenberg group is spanned by three vectors, denoted as X, Y, and Z, where the only non-trivial Lie bracket relation is given by $[X, Y] = Z$. Notably, for every three vectors X_1, X_2, and X_3 in this Lie algebra, the Lie bracket operation satisfies the property $[X_1, [X_2, X_3]] = 0$. This characteristic feature establishes the Heisenberg Lie algebra as a nilpotent Lie algebra of step 2. Namely, the second commutator vanishes, i.e., after two steps in the commutator hierarchy we reach zero. More generally, a Lie algebra is said to be *nilpotent* with *nilpotency step s* if, for every selection of more than s vectors within it, the iterated bracket of these vectors results in 0.

In every vector space, we can define the zero bracket operation, thus forming a Lie algebra. Such a Lie algebra is referred to as *commutative* and is nilpotent with a step of 1.

We assert that there are only two examples of 3D simply connected nilpotent Lie groups: the 3D vector space $(\mathbb{R}^3, +)$ and the Heisenberg group. To prove this, let $\mathfrak{g}$ be the Lie algebra of one such group. Since $\mathfrak{g}$ is nilpotent, we can take a non-zero element Z in the center of $\mathfrak{g}$. We extend Z to form a basis $\{X, Y, Z\}$ for $\mathfrak{g}$. Now, either X and Y commute, making the algebra commutative, or the vector $W := [X, Y]$ is not 0. In this second case, we write $W = aX + bY + cZ$, for some

$a, b, c \in \mathbb{R}$. Then, we have that $[W, Y] = aW$. Due to the nilpotency of $\mathfrak{g}$, we have $a = 0$. Similarly, $b = 0$. Thus, we infer that $c \neq 0$, and by replacing Z with cZ, we can define the algebra structure of $\mathfrak{g}$ by the relations:

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

We can conclude the proof by invoking the uniqueness of a simply connected Lie group with a fixed Lie algebra, as stated in Theorem 5.1.7.

2.4 The Sub-Riemannian Heisenberg Group

Our preferred model for the Heisenberg group is $\mathbb{R}^3$ with the product law (2.9), which we observed yield the following left-invariant vector fields: $\partial_x - \frac{y}{2}\partial_z$, $\partial_y + \frac{x}{2}\partial_z$, ∂_z. The reason why this model is advantageous lies in its canonical identification of the group with its Lie algebra. In other words, we are using exponential coordinates, a perspective that will be elucidated in Sect. 5.2.3. It's important to note that, given the uniqueness of the Heisenberg structure, all other models can be considered equivalent through a smooth group morphism.

In $\mathbb{R}^3$, we begin by selecting three vector fields, denoted as X, Y, and Z, that, as for (2.8), are linearly independent at every point and satisfy the following commutation relations:

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

It is a fact that the space can be endowed with a group law that renders them left-invariant.

We consider the subbundle Δ of the tangent bundle $T(\mathbb{R}^3)$ over $\mathbb{R}^3$ such that for every $p \in \mathbb{R}^3$

$$\Delta_p := \Delta \cap T_p\mathbb{R}^3 := \operatorname{span}\{X_p, Y_p\}.$$

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

$$d_{\mathrm{cc}}(p, q) := \inf\{L(\gamma) : \gamma \text{ is a horizontal curve from } p \text{ to } q\}. \tag{2.14}$$

Therefore, we have broadened the notion of Legendrian curve with that of horizontal curve, and the concept of a contact distance has been expanded to a Carnot-Carathéodory distance. This variation in terminology arises because sub-Riemannian geometry draws from various mathematical domains, resulting in multiple jargons.

We refer to the space $(\mathbb{R}^3, d_{cc})$ as *(a model for) the sub-Riemannian Heisenberg group*. In the rest of this section, we will exclusively operate within our preferred model: $\mathbb{R}^3$ with the product law (2.9) and the orthonormal frame (2.8).

2.4.1 Geodesics and Spheres in the Heisenberg Group

From Sects. 2.2 and 2.3, we can deduce the following properties of a curve $\gamma(t) = (x(t), y(t), z(t))$:

- The curve γ is horizontal, meaning $\dot{\gamma} \in \Delta$, if and only if

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

  which is equivalent to stating that $z(t)$ represents the area spanned by the curve $(x(\cdot), y(\cdot))$ up to the point t, as in Fig. 2.1.
- The condition $\dot{\gamma} \in \Delta$ holds if and only if $\dot{\gamma} = u_1 X + u_2 Y$, where $u_1 = \dot{x}$ and $u_2 = \dot{y}$. To see this, on the one hand, we have $\pi(\dot{\gamma}) = \pi((\dot{x}, \dot{y}, \dot{z})) = (\dot{x}, \dot{y})$ and, on the other hand, we have

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

- If $\dot{\gamma} \in \Delta$, then the length of γ is given by

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

Due to the preceding discussion, we can derive explicit formulas for the geodesics in the sub-Riemannian Heisenberg group. This is made possible by our understanding of the solutions to the isoperimetric problem, as detailed in Sect. 2.5.1. Specifically, we have discovered that given the way the geometry in the Heisenberg group has been constructed, the shortest curves concerning the length structure are the lifts of solutions to a variant of the isoperimetric problem. Namely, we search for the shortest curves in the plane that enclose a fixed area and join two specified points. Such curves turn out to be arcs of circles or straight-line segments. Consequently, the geodesics in the Heisenberg group are essentially the lifted versions of circles in the plane.

Remark 2.4.1 For a fixed endpoint $(x(1), y(1), z(1))$, every curve $(x(t), y(t))$ that encloses an area equal to $z(1)$, is such that $(x(0), y(0)) = (0, 0)$, and among such curves minimizes $\text{Length}_{\text{Eucl}}(x(\cdot), y(\cdot))$ is either a segment of a circle or a straight line in the plane.

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

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

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

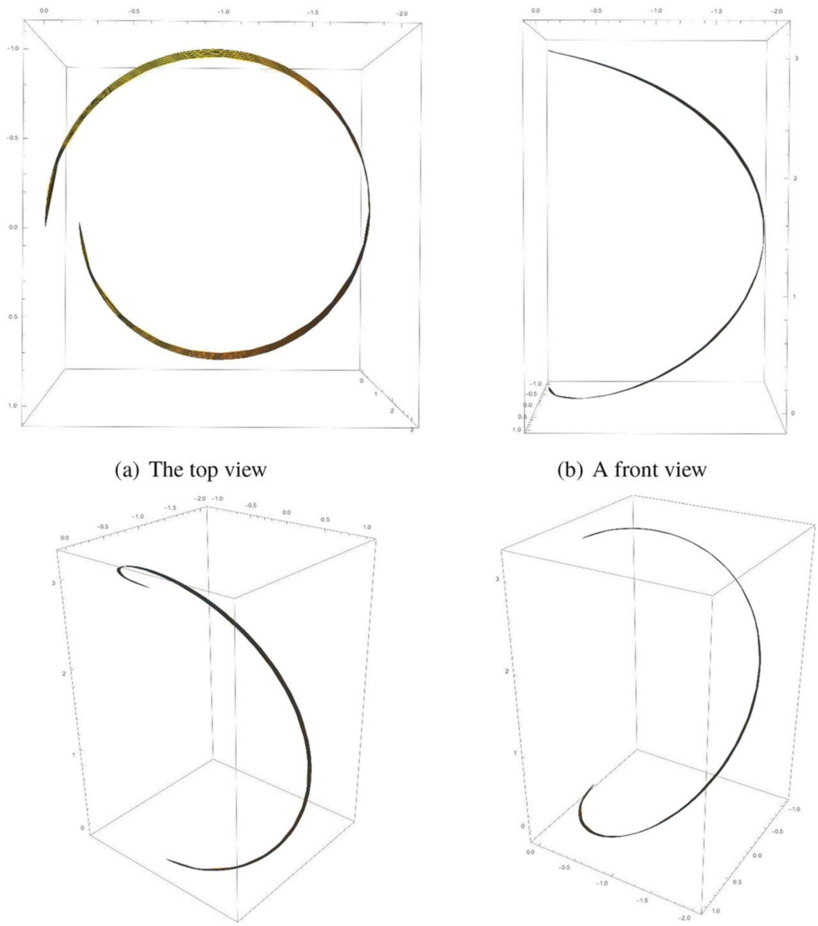
for $0 \leq t \leq \frac{2\pi}{|k|}$. Such a circle is parametrization by arc length and has its center on the x-axis. If $k > 0$, the center is on the negative x-axis; if $k < 0$, it is on the positive x-axis. Moreover, if $k > 0$, then the circle (x_0, y_0) encloses positive area; if $k < 0$ it encloses negative area. For $k = 0$, we can still consider the formula in the limit sense: the circles degenerate into the line $(0, t)$, defined for all $t \in \mathbb{R}$.

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

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

We can calculate the third coordinate of the lift of such circles via (2.1).

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



(a) The top view

(b) A front view

(c) A side view

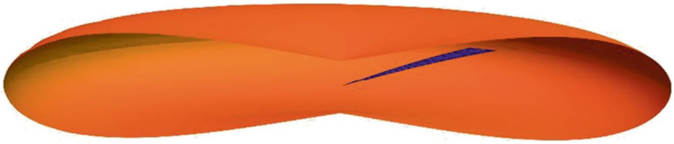
(d) A side view

Fig. 2.4 Various views of a geodesic with non-zero curvature in the sub-Riemannian Heisenberg geometry

We conclude that the length-minimizing curves originating from the point $(0, 0, 0)$ in $\mathbb{R}^3$ are smooth curves $\gamma = (\gamma_1, \gamma_2, \gamma_3)$ given by the equations:

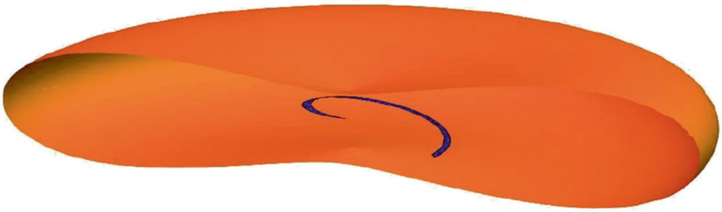
$$\begin{cases} \gamma_1(t) = \cos\theta \frac{\cos(kt)-1}{k} - \sin\theta \frac{\sin(kt)}{k} \\ \gamma_2(t) = \sin\theta \frac{\cos(kt)-1}{k} + \cos\theta \frac{\sin(kt)}{k} \\ \gamma_3(t) = \frac{kt-\sin(kt)}{2k^2} \end{cases}, \tag{2.15}$$

for some $\theta \in \mathbb{R}/2\pi\mathbb{Z}$ and $k \in \mathbb{R}$. We refer to Fig. 2.4 for a visual representation of a sub-Riemannian geodesic. Moreover, in Fig. 2.5, we have a picture of a unit sphere with several geodesics. For more info on spheres, see Figs. 2.7, 2.8, and Exercise 2.6.14.

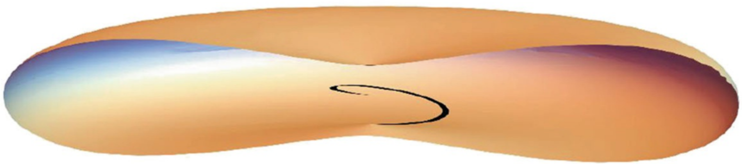


(a) A geodesic with zero curvature

(b) A geodesic with small curvature



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



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

Fig. 2.5 Geodesics within the unit sphere in the sub-Riemannian Heisenberg geometry

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

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

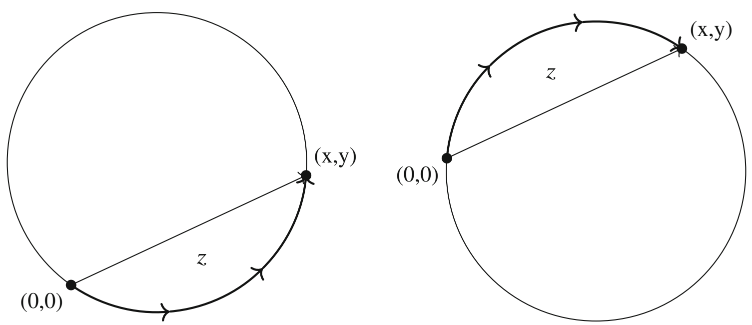
Indeed, lines passing through the origin in the xy-plane are geodesics.

We have identified *all* length-minimizing curves in the sub-Riemannian Heisenberg group. This characterization of the geodesics leads to several interesting conclusions:

1. If a point $(x, y, z) \in \mathbb{R}^3$ lies on the z-axis, i.e., $(x, y) = (0, 0)$, then there exist infinitely many length-minimizing curves between this point and the origin $(0, 0, 0)$. In fact, these curves form a one-parameter family and are the lifts of circles with area z, all containing the point $(0, 0)$.
2. If $(x, y) \neq (0, 0)$, then there exists a unique length-minimizing curve from (x, y, z) to $(0, 0, 0)$. This curve is the lift of a circular arc enclosing area z together with the segment connecting $(0, 0)$ to (x, y). Please refer to Fig. 2.6 for a visual representation.

Since the distance 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 and only if $\pi(p) = \pi(q)$, i.e., the points lie on the same vertical line. Conversely, if $\pi(p) \neq \pi(q)$, then there is only a single length-minimizing curve between them (Fig. 2.7).

We deduce that this sub-Riemannian geometry differs fundamentally from Riemannian geometry. Although all metric balls and metric spheres in the Heisenberg

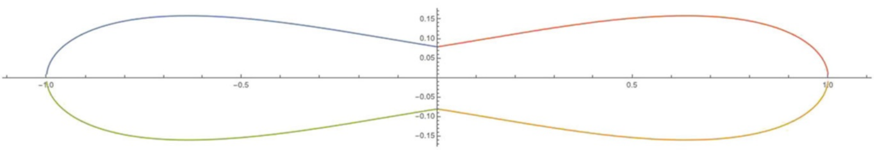


(a) Case of $z > 0$. (b) Case of $z < 0$.

Fig. 2.6 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 circle counterclockwise. If z is negative, it follows clockwise. In both cases, the area enclosed by the curve and the circle equals $|z|$

(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.

Fig. 2.7 Balls in the sub-Riemannian Heisenberg group are not smooth surfaces. At the two "poles" the sphere is not C^1; there is no cusp, but there is a corner. For a parametrization, see Exercise 2.6.14

group remain topological balls and spheres, respectively (see Exercise 2.6.14), this geometry is not Riemannian and, furthermore, it cannot be bi-Lipschitz equivalent to any Riemannian geometry, as it will be shown in Corollary 2.4.5.

2.4.2 Dilations on the Heisenberg Group

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

$$\begin{aligned} \delta_\lambda : \quad \mathbb{R}^3 &\longrightarrow \mathbb{R}^3 \\ (x, y, z) &\longmapsto (\lambda x, \lambda y, \lambda^2 z). \end{aligned} \tag{2.16}$$

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 (2.9).

Lemma 2.4.2 *The dilations* (2.16) *on the Heisenberg group satisfy the following properties: For all* $\lambda, \mu \in \mathbb{R}$ *and all* $p, q \in \mathbb{R}^3$*:*

2.4.2.i. $\delta_\lambda(p \cdot q) = \delta_\lambda(p) \cdot \delta_\lambda(q)$;
2.4.2.ii. $\delta_\lambda \circ \delta_\mu = \delta_{\lambda\mu}$;
2.4.2.iii. δ_λ *is a Lie group isomorphism with inverse* $\delta_{\frac{1}{\lambda}}$*, if* $\lambda \neq 0$;
2.4.2.iv. $d_{cc}(\delta_\lambda(p), \delta_\lambda(q)) = |\lambda| d_{cc}(p, q)$*, where* d_{cc} *is defined in* (2.14).

Proof of 2.4.2.i From the group law (2.9), we have:

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

□

Proof of 2.4.2.ii This is obvious from the definition (2.16):

$$\begin{aligned} (\delta_\lambda \circ \delta_\mu)(x, y, z) &= \delta_\lambda(\mu x, \mu y, \mu^2 z) \\ &= (\lambda\mu x, \lambda\mu y, \lambda^2\mu^2 z) \\ &= (\lambda\mu x, \lambda\mu y, (\lambda\mu)^2 z) \\ &= \delta_{\lambda\mu}(x, y, z). \end{aligned}$$

□

Proof of 2.4.2.iii From the previous points, we conclude that each δ_λ is a group homomorphism and $(\delta_\lambda)^{-1} = \delta_{\frac{1}{\lambda}}$, if $\lambda \neq 0$. Moreover, it is evident that each map is smooth. □

Proof of 2.4.2.iv 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. (This last statement is suggested as an exercise.) Hence, the map δ_λ preserves horizontal curves and multiplies their length by λ.

Method 2: By the left-invariance of d_{cc} and by (2.4.2.ii), 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

$$d_{cc}(\delta_\lambda(p), \mathbf{0}) = \lambda d_{cc}(p, \mathbf{0}). \tag{2.17}$$

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 (see Exercise 2.6.11) shows that $\delta_\lambda \circ \gamma$ is still of the same form, up to a linear reparametrization by λ. Hence, its length got multiplied by the factor λ.

Method 3: Reasoning as at the beginning of Method 2, proving (2.17) 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 $\begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix}$ multiplies length by λ and area by λ^2. Therefore, the curve $(\lambda x, \lambda y)$ spans areas that are λ^2 times the areas of (x, y) and has length λ times the length of (x, y). Thus the curve $(\lambda x, \lambda y, \lambda^2 z)$ is horizontal and has length $\lambda L(\gamma)$. Hence $d_{cc}(\delta_\lambda(p), \mathbf{0}) \le \lambda d_{cc}(p, \mathbf{0})$.
We conclude by arguing similarly with each curve σ joining $\delta_\lambda(p)$ to $\mathbf{0}$ and considering the curve $\delta_{\frac{1}{\lambda}} \circ \sigma$.

□

As a consequence of Lemma 2.4.2, we have the following properties for the metric balls in the Heisenberg group.

Corollary 2.4.3 *In the sub-Riemannian Heisenberg group, we have*

2.4.3.i. $B_{d_{cc}}(\mathbf{0}, r) = \delta_r(B_{d_{cc}}(\mathbf{0}, 1))$;
2.4.3.ii. $B_{d_{cc}}(p, r) = L_p(\delta_r(B_{d_{cc}}(\mathbf{0}, 1)))$,

for all points p and all radii r.

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

$$(x, y, z) \in B_{d_{cc}}((\mathbf{0}, 1) \Longleftrightarrow (rx, ry, r^2 z) \in B_{d_{cc}}(\mathbf{0}, r). \tag{2.18}$$

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 a 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 the growth of the balls in the Heisenberg geometry.

Corollary 2.4.4 *Let* vol *be the 3D Lebesgue volume in $\mathbb{R}^3$. The Heisenberg sub-Riemannian distance d_{cc} satisfies*

$$\mathrm{vol}(B_{d_{cc}}(p, r)) = r^4 \,\mathrm{vol}(B_{d_{cc}}(\mathbf{0}, 1)), \qquad \forall p \in \mathbb{R}^3,\ \forall r > 0. \tag{2.19}$$

Proof From (2.18), we know that $\mathrm{vol}(B(\mathbf{0}, r)) = r^4 \mathrm{vol}(B(\mathbf{0}, 1))$. Now, we can conclude the proof using the fact that left translations (2.10) in the Heisenberg group are isometries, and they preserve the volume. This last fact can be verified by noticing that the determinant of the differential of a left translation is 1, as seen in (2.11). Namely, every left translation L_p is such that $\mathrm{d}L_p = \begin{bmatrix} 1 & 0 & 0 \\ * & 1 & 0 \\ * & * & 1 \end{bmatrix}$, and thus $\mathtt{Jac}(L_p) = \det(\mathrm{d}L_p) = 1$. Notice that $\mathtt{Jac}(\delta_\lambda) = \det(\mathrm{d}\delta_\lambda) = \det \begin{bmatrix} \lambda & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & \lambda^2 \end{bmatrix} = \lambda^4$. Then

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

where we used Corollary 2.4.3. □

2.4.2.1 The Dimension of the Heisenberg Group

The following theorem will demonstrate that the sub-Riemannian Heisenberg group is not locally bi-Lipschitz equivalent to any Riemannian manifold. For the notions of bi-Lipschitz and Hausdorff dimension, please refer to Sect. 3.1.

Corollary 2.4.5 *The sub-Riemannian Heisenberg group endowed with the standard Carnot-Carathéodory distance has topological dimension equal to 3, but Hausdorff dimension equal to* 4. *In particular, locally, this metric space is not bi-Lipschitz equivalent to the Euclidean space.*

Proof The Carnot-Carathéodory distance makes the Heisenberg group homeomorphic to the vector space $\mathbb{R}^3$, which has topological dimension 3. From the general metric geometry theory, which we will review in Sect. 3.1.6, it is sufficient to prove the existence of 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:

$$k_1 \epsilon^{-4} < N_\epsilon < k_2 \epsilon^{-4}. \tag{2.20}$$

For the lower bound, consider $B_1, \ldots, B_{N_\epsilon}$ as such balls. Using (2.19), we have:

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

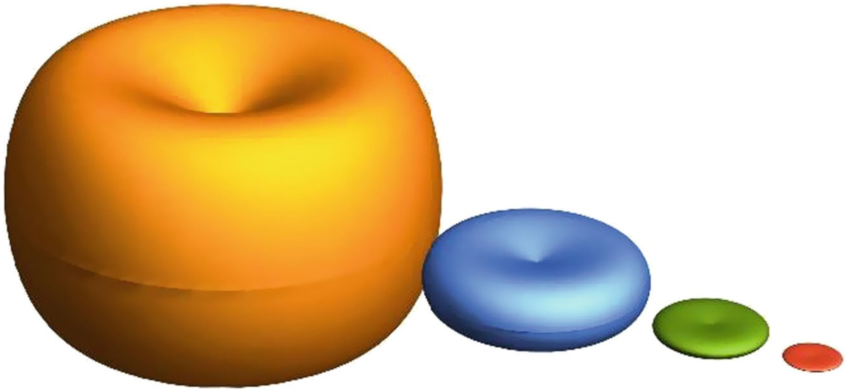


Fig. 2.8 Balls of different sizes in the Heisenberg geometry. All the balls are with the origin as the center. From the left, these are the balls of radius 2, 1, 1/2, 1/4

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 ϵ. 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 (2.19) we infer:

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

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

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

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

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

Hence, both bounds in (2.20) are proven.

The Hausdorff dimension is a bi-Lipschitz invariant. Therefore, the Heisenberg group is not bi-Lipschitz equivalent to the Euclidean 3-space. □

2.4.2.2 A Ball-Box Theorem

In this section, we provide an elementary explanation of why the balls in the sub-Riemannian Heisenberg geometry behave like boxes with anisotropic sides. Specifically, let us define:

$$\mathtt{Box}(r) := [-r, r] \times [-r, r] \times [-r^2, r^2] \subseteq \mathbb{R}^3. \tag{2.21}$$

Proposition 2.4.6 *In the sub-Riemannian Heisenberg group (in the standard coordinates as above), the balls at the origin satisfy*

$$\mathtt{Box}(c_1 r)) \subset B_{cc}(1, r) \subset \mathtt{Box}(c_2 r)), \tag{2.22}$$

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

Proof In the following argument, we do not aim to find the best possible choices for c_1 and c_2. Moreover, using the dilations δ_r from the previous section, one can prove the result for the unit ball and then dilate to obtain the general case. The existence of the two boxes (inside and outside) comes from the fact that the unit ball is a bounded open set. Nonetheless, we provide a direct proof without relying on 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 effectively $\|(x, y)\| < r$. Moreover, we claim that we have an upper bound on z as a function of r. Indeed, we should bound the oriented area enclosed by a curve of length r. We stress that the curve is not closed, and the area is a signed area. In other words, the coordinate $z(t)$ satisfies (2.2). Hence, for the considered curve (which we might think 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, to show that the r-ball contains a specific box, we claim that

$$\left[-\frac{r}{3}, \frac{r}{3}\right] \times \left[-\frac{r}{3}, \frac{r}{3}\right] \times \left[-\frac{r^2}{100}, \frac{r^2}{100}\right] \subset B_{cc}(1, r), \qquad \forall r > 0. \tag{2.23}$$

Indeed, take a point (x, y, z) such that $|x|, |y| \leq r/3$ and $|z| \leq r^2/100$. Then, construct the following planar curve: start from $(0, 0)$ and 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, so its lift is an admissible curve reaching

(x, y, z). The length of the curve is four times the side length of the square plus the length of the segment. The square has area at most $\frac{r^2}{100}$, so its side length is at most $\frac{r}{10}$. The segment has a length of at most $\frac{\sqrt{2}r}{3}$. From these bounds, we have $4\frac{r}{10} + \frac{\sqrt{2}r}{3} < r$. Therefore the point (x, y, z) is inside the r-ball, confirming (2.23). □

2.5 Supplementary Material

2.5.1 *Dido's Problem: A Proof of the Isoperimetric Problem*

To better understand how, in Sect. 2.4.1, we obtained formulas for the geodesics in the sub-Riemannian 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 [Mor88, Mag12].

We shall use the formalism of Calculus of Variations to prove that each of the shortest closed curves in the plane that encloses a fixed 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

$$\mathcal{A}(\sigma) := \text{Area enclosed by } \sigma = A_0, \text{ for some } A_0 > 0,$$

are circles. However, we shall assume that such a σ is a C^1 curve with Lipschitz derivative.

2.5.1.1 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 critical with respect to a variation h is expressed in Calculus of Variations as follows.

Definition 2.5.1 Given a Lipschitz curve $\sigma : [0, l] \to \mathbb{R}^2$, a *variation* on the interval $[0, l]$ is a smooth curve $h : [0, l] \to \mathbb{R}^2$ with $h(0) = h(l) = 0$, and we define the associated *variation of length* as

$$\delta\mathcal{L}(\sigma, h) := \frac{\mathrm{d}}{\mathrm{d}\epsilon}\mathcal{L}(\sigma + \epsilon h)\bigg|_{\epsilon=0}.$$

We say that a σ is *critical in the direction of* h if $\delta\mathcal{L}(\sigma, h) = 0$.

Let us calculate the variation $\delta\mathcal{L}$ of length in the case when σ is parametrized by arc length. So $|\dot{\sigma}| = 1$ and $l = \text{Length}(\sigma)$. The variation in this case is

$$\begin{aligned}
\delta\mathcal{L}(\sigma, h) &:= \frac{\mathrm{d}}{\mathrm{d}\epsilon}\mathcal{L}(\sigma + \epsilon h)\bigg|_{\epsilon=0} \\
&= \frac{\mathrm{d}}{\mathrm{d}\epsilon}\int_0^l \sqrt{\left(\dot{\sigma}_1(t) + \epsilon\dot{h}_1(t)\right)^2 + \left(\dot{\sigma}_2(t) + \epsilon\dot{h}_2(t)\right)^2}\,\mathrm{d}t\bigg|_{\epsilon=0} \\
&= \int_0^l \frac{\mathrm{d}}{\mathrm{d}\epsilon}\bigg(\dot{\sigma}_1(t)^2 + 2\epsilon\dot{\sigma}_1(t)\dot{h}_1(t) + \epsilon^2\dot{h}_1(t)^2 + \dot{\sigma}_2(t)^2 \\
&\qquad\qquad + 2\epsilon\dot{\sigma}_2(t)\dot{h}_2(t) + \epsilon^2\dot{h}_2(t)^2\bigg)^{1/2}\bigg|_{\epsilon=0}\,\mathrm{d}t \\
&= \int_0^l \frac{1}{2}\left(2\dot{\sigma}_1(t)\dot{h}_1(t) + 2\epsilon\dot{h}_1(t)^2 + 2\dot{\sigma}_2(t)\dot{h}_2(t) + 2\epsilon\dot{h}_2(t)^2\right) \\
&\qquad \cdot\bigg(\dot{\sigma}_1(t)^2 + 2\epsilon\dot{\sigma}_1(t)\dot{h}_1(t) + \epsilon^2\dot{h}_1(t)^2 + \dot{\sigma}_2(t)^2 \\
&\qquad\qquad + 2\epsilon\dot{\sigma}_2(t)\dot{h}_2(t) + \epsilon^2\dot{h}_2(t)^2\bigg)^{1/2}\bigg|_{\epsilon=0}\,\mathrm{d}t \\
&= \int_0^l \frac{\dot{\sigma}_1(t)\dot{h}_1(t) + \dot{\sigma}_2(t)\dot{h}_2(t)}{\sqrt{\dot{\sigma}_1(t)^2 + \dot{\sigma}_2(t)^2}}\,\mathrm{d}t \\
&= \int_0^l \frac{\langle\dot{\sigma}(t), \dot{h}(t)\rangle}{|\dot{\sigma}(t)|}\,\mathrm{d}t \\
&= \int_0^l \langle\dot{\sigma}(t), \dot{h}(t)\rangle\,\mathrm{d}t.
\end{aligned}$$

We conclude the following:

Lemma 2.5.2 *Let* $\sigma : [0, l] \to \mathbb{R}^2$ *be a planar curve parametrized by arc length. For every variation h, we have*

$$\delta\mathcal{L}(\sigma, h) = \int_0^l \langle \dot{\sigma}, \dot{h} \rangle \, \mathrm{d}t.$$

2.5.1.2 Area Functional and Its Variation

Because of Stokes' Theorem, the area enclosed by a Lipschitz curve σ can be computed 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, we use the *cross product* on $\mathbb{R}^2$, that is, the real number

$$v \times w := v_1 w_2 - w_1 v_2 = \left\langle \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} v_1 \\ v_2 \\ 0 \end{bmatrix} \times \begin{bmatrix} w_1 \\ w_2 \\ 0 \end{bmatrix} \right\rangle, \text{ for } v = (v_1, v_2), w = (w_1, w_2) \in \mathbb{R}^2.$$

We have linearity in v and w and $w \times v = -v \times w$. 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 on $[0, l]$. The new area is

$$\begin{aligned}
\mathcal{A}(\sigma + h) &= \frac{1}{2} \int_0^l (\sigma + h) \times (\dot{\sigma} + \dot{h}) \, \mathrm{d}t \\
&= \frac{1}{2} \int_0^l \sigma \times \dot{\sigma} + \sigma \times \dot{h} + h \times \dot{\sigma} + h \times \dot{h} \, \mathrm{d}t \\
&= \mathcal{A}(\sigma) + \frac{1}{2} \, h \times \sigma|_0^l + \frac{1}{2} \int_0^l -\dot{\sigma} \times h + h \times \dot{\sigma} + h \times \dot{h} \, \mathrm{d}t \\
&= \mathcal{A}(\sigma) + \int_0^l h \times \dot{\sigma} \, \mathrm{d}t + \frac{1}{2} \int_0^l h \times \dot{h} \, \mathrm{d}t.
\end{aligned}$$

We deduce that a variation h is area-preserving for a curve σ if and only if

$$\int_0^l h \times \dot{\sigma} + \frac{h \times \dot{h}}{2} \,\mathrm{d}t = 0.$$

We need a weaker notion than area preservation. We require $\mathcal{A}(\sigma + \epsilon h) = \mathcal{A}(\sigma) + o(\epsilon)$, as $\epsilon \to 0$.

Definition 2.5.3 We say that a variation h *infinitesimally preserves the area* of a curve σ if

$$\frac{\mathrm{d}}{\mathrm{d}\epsilon}\mathcal{A}(\sigma + \epsilon h)\bigg|_{\epsilon=0} = 0.$$

By the above calculation, a variation h infinitesimally preserves the area if and only if

$$0 = \frac{\mathrm{d}}{\mathrm{d}\epsilon}\int_0^l \epsilon h \times \dot{\sigma} + \frac{\epsilon h \times \epsilon \dot{h}}{2} \,\mathrm{d}t\bigg|_{\epsilon=0} = \int_0^l h \times \dot{\sigma} \,\mathrm{d}t.$$

Proposition 2.5.4 *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 constraint, then σ has zero first variation of length with respect to all infinitesimally 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 $a_\epsilon := \mathcal{A}(\sigma + \epsilon h)$, which we may assume positive. Since h infinitesimally preserves the area, we have $\frac{\mathrm{d}}{\mathrm{d}\epsilon}a_\epsilon\Big|_{\epsilon=0} = 0$. 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 of 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 Lemma 2.5.2. □

2.5.1.3 Conclusion

Proposition 2.5.5 *If σ is a $C^{1,1}$ closed curve in the plane that is one of the shortest among all Lipschitz curves that enclose the same 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 $\phi(0) = \phi(l) = 0$ 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). □

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 example, a complete proof, based on Poincaré-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 2.5.6 (Isoperimetric Solution) *If σ is an absolutely continuous closed curve in the plane that is one of the shortest among all absolutely continuous curves that enclose the same 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 2.5.7 (Dido's Solution) *Given two points p and q on the plane and a value $A > 0$, 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 an arc of a circle. Let γ be an arc of a circle joining the points and 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 2.5.6. □

2.6 Exercises

In this book, exercises are presented as statements, which the reader should prove. Several exercises contain hints or solutions, which the reader is encouraged not to read immediately. Most exercises do not require extensive calculations but rather a good understanding and application of the theory. Some harder exercises are marked with the symbol ☠.

Some of the exercises in this chapter require a basic understanding of differential geometry, ranging from calculus to Lie group theory. Novice readers could choose to skip some of these exercises and return to them after having read the relevant material in Chaps. 3 and 5.

Exercise 2.6.1 (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 a half disk.

Exercise 2.6.2 The differential forms $\mathrm{vol} := \mathrm{d}x \wedge \mathrm{d}y$ and $\alpha := \dfrac{1}{2}(x\mathrm{d}y - y\mathrm{d}x)$ in $\mathbb{R}^2$ have the following properties:

2.6.2.i. $\mathrm{d}(\alpha) = \mathrm{vol}$;

2.6.2.ii. In polar coordinate (r, θ), we have $\alpha = \dfrac{1}{2}r^2\,\mathrm{d}\theta$;

2.6.2.iii. If L is a line through the origin, then the line integral $\int_L \alpha$ is 0.

Exercise 2.6.3 Let $t \in [0, 1] \mapsto \sigma(t) = (x(t), y(t)) \in \mathbb{R}^2$ be a Lipschitz curve on the plane. Let $\sigma_{[0,t]}$ be the arc restricted to the interval $[0, t]$. Let $f : \mathbb{R}^2 \to \mathbb{R}$ be a smooth function. Then,

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\int_{\sigma_{[0,t]}} f(x, y)\,\mathrm{d}x\right) = f(x(t), y(t))\frac{\mathrm{d}x}{\mathrm{d}t}(t), \qquad \text{for almost every } t \in [0, 1],$$

where $\int_{\sigma_{[0,t]}} f(x, y)\,\mathrm{d}x$ denotes the line integral of a one-form along a curve.

Exercise 2.6.4 We have the relations $[X, Y] = Z$ and $[X, Z] = [Y, Z] = 0$ in the following cases: (a) for the vector fields in (2.6); (b) for the matrices (2.13).

Exercise 2.6.5 The inverse of an element (x, y, z), with respect to the group structure given by (2.9), is $(-x, -y, -z)$.

Exercise 2.6.6 In the group structure on $\mathbb{R}^3$ given by (2.9), the lines $t \in \mathbb{R} \longmapsto \gamma_v(t) = (tv_1, tv_2, tv_3)$ are one-parameter subgroups.

Exercise 2.6.7 The left translations and the right translations for the group structure given by (2.9) have a Jacobian determinant equal to 1. Consequently, they preserve the Lebesgue measure in $\mathbb{R}^3$.

Exercise 2.6.8 Let L be a line through 0 in the xy-plane of $\mathbb{R}^3$. Then, the curve L is a geodesic with respect to the contact distance d_c defined in (2.7).

Exercise 2.6.9 Consider the map

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

from $\mathbb{R}^3$ with the product (2.9) to the space of 3×3 upper-triangular matrices with the usual matrix product. Then

(a) the map is a Lie group isomorphism;
(b) the map sends the standard basis X, Y, and Z (defined in (2.6)) of the first Lie algebra to the standard basis X, Y, and Z (defined in (2.13)) of the second Lie algebra.

Exercise 2.6.10 On the vertical z-axis, the distance d_c defined in (2.7) is a multiple of the square root of the Euclidean one. Find this multiple.

Exercise 2.6.11 Each map δ_λ as in (2.16) sends geodesics of the Heisenberg group to reparametrized geodesics.
Solution. If $(\gamma_1, \gamma_2, \gamma_3)$ is a geodesic arc of length 1 starting from the origin, then it is of the form (2.15) for some $k \in \mathbb{R}$ with $2\pi/|k| \geq 1$, and the parameter t of (2.15) ranges from 0 to 1. Now, consider the curve $(r\gamma_1, r\gamma_2, r^2\gamma_3)$ which is

$$\begin{aligned}
&(r\gamma_1(t), r\gamma_2(t), r^2\gamma_3(t)) \\
&\qquad = \Bigg(\frac{\cos\theta(\cos(kt) - 1) - \sin\theta\sin(kt)}{k/r}, \\
&\qquad\qquad \frac{\sin\theta(\cos(kt) - 1) + \cos\theta\sin(kt)}{k/r}, \\
&\qquad\qquad\qquad \frac{kt - \sin(kt)}{2(k/r)^2} \Bigg), \qquad \text{for } t \in [0, 1].
\end{aligned}$$

This curve is a geodesic, albeit not parametrized by arc length but by a multiple of it, specifically r. Consequently, its length is r.

Exercise 2.6.12 We have Corollary 2.4.3, from Lemma 2.4.2.

Exercise 2.6.13 The map

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

given by

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

is a homeomorphism.

Exercise 2.6.14
(i) Let Φ be the map defined in Exercise 2.6.13. The unit ball in the Heisenberg geometry is given by

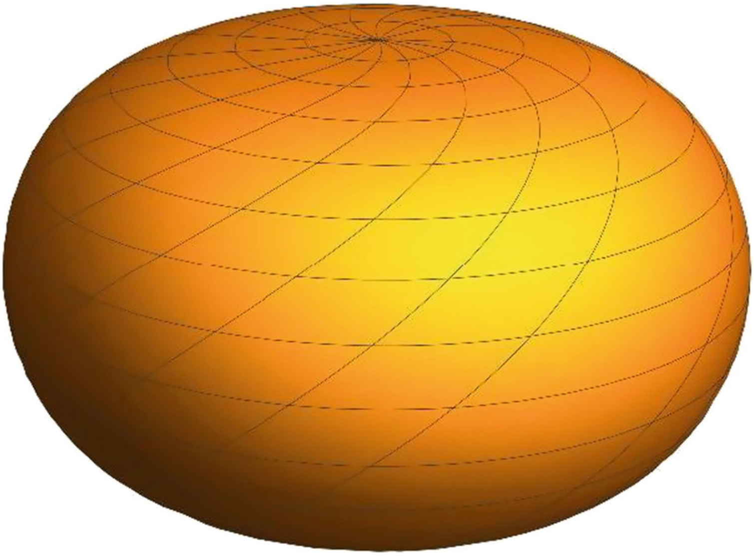
$$\begin{aligned} B(0, 1) &= \{\Phi(\theta, k, t) \ : \ \theta \in \mathbb{R}/2\pi\mathbb{Z}, k \in \mathbb{R}, t \in (0, 1)\} \\ &= \{\Phi(\theta, k, t) \ : \ \theta \in \mathbb{R}/2\pi\mathbb{Z}, k \in [-2\pi, 2\pi], t \in (0, 1)\}, \end{aligned}$$

and the unit sphere is

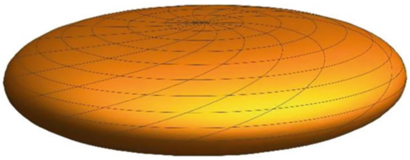
$$S(0, 1) = \{\Phi(\theta, k, 1) \ : \ \theta \in \mathbb{R}/2\pi\mathbb{Z}, k \in [-2\pi, 2\pi]\}.$$

(ii) All the metric balls and metric spheres in the sub-Riemannian Heisenberg group are topological balls and topological spheres, respectively; see Figs. 2.5, 2.7, 2.8.

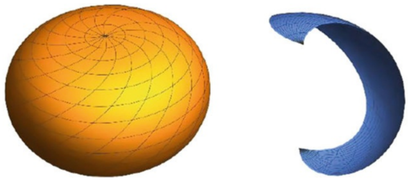
Exercise 2.6.15 (Pansu sphere) In the standard coordinates of the Heisenberg group, let S_P be the set of all the supports of the sub-Riemannian geodesics between $(0, 0, 1)$ and $(0, 0, -1)$, called *poles*; see Fig. 2.9. Then this set S_P, called *Pansu sphere*, is a surface that is C^2 diffeomorphic to the round sphere. Moreover, it is C^∞ away from the poles.



(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 its true axis.



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

Fig. 2.9 The (conjectured) isoperimetric sphere in the sub-Riemannian Heisenberg geometry

Chapter 3
A Review of Metric and Differential Geometry

Metric and differential geometry are the main tools for studying sub-Riemannian geometries. 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 is central and indispensable in sub-Riemannian geometry. It provides the mathematical framework for studying fundamental geometric objects such as tangent bundles and vector fields. It allows for analyzing 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.

3.1 Metric Geometry: Lengths, Geodesic Spaces, and Hausdorff Measures

3.1.1 Metric Spaces

Let M be a set. A function

$$d : M \times M \to [0, +\infty] \tag{3.1}$$

E. Le Donne, *Metric Lie Groups*, Graduate Texts in Mathematics 306,
https://doi.org/10.1007/978-3-031-98832-5_3

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

(3.1.i) *positiveness*: $d(x, y) = 0 \iff x = y$,
(3.1.ii) *symmetry*: $d(x, y) = d(y, x)$,
(3.1.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 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 Sect. 3.2.3.

Every metric space (M, d) has a natural topology which is generated by the *open balls*

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

We also consider distance functions that may have value ∞. However, on each connected component of the metric space, the distance is finite (see Exercise 3.4.1).

For subsets E, F of a metric space (M, d), the *distance* between the two subsets is

$$d(E, F) := \inf\{d(p, p') : p \in E, q \in F\}. \tag{3.2}$$

Whereas, the *diameter* of $E \subseteq M$ is defined as

$$\operatorname{diam}(E) := \sup\{d(p, p') : p, q \in E\}. \tag{3.3}$$

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, closed, 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 $\gamma(b) = q$.

3.1.2 Length of Curves in Metric Spaces

Definition 3.1.1 (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

$$\begin{aligned} L(\gamma) &:= \operatorname{Length}_d(\gamma) \\ &:= \sup\left\{\sum_{i=1}^{k} d(\gamma(t_{i-1}), \gamma(t_i)) \ : \ k \in \mathbb{N}, a = t_0 < t_1 < \cdots < t_k = b\right\}. \end{aligned} \tag{3.4}$$

A *rectifiable curve* is a curve with finite length. Verifying that the length does not depend on the parametrization is easy, as shown in Exercise 3.4.6. A curve $\gamma : [a, b] \to M$ is said to be *parametrized by arc length* if

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

Every rectifiable curve admits a reparametrization by arc length; see Exercise 3.4.7. With this parametrization, the curve is a 1-Lipschitz map; see Sect. 3.1.5 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 tuple $(t_0, t_1, \ldots, t_k) \in [a, b]^{k+1}$ with $k \in \mathbb{N}$ such that $a = t_0 \leq t_1 \leq \cdots \leq t_k = b$. We define

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

Hence, we have

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

Next, we recall the lower semicontinuity of length for sequences of curves that converge 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 3.1.2 (Semicontinuity of Length) *Let $\gamma, \gamma_1, \gamma_2, \ldots$ be curves in a metric space defined on the same interval. 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 3.4.9) and the general fact that the supremum of sequentially continuous functions is a sequentially lower semicontinuous function (see Exercise 3.4.10). A proof of the latter fact is given by a straightforward adaptation of the following argument.

Let $\mathcal{P}$ be a partition of $[a, b]$. Say $\mathcal{P} = (t_0, t_1, \ldots, t_k)$, 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 $i \in \{0, 1, \ldots, k\}$. By the 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 Fig. 3.1 for a visualization.

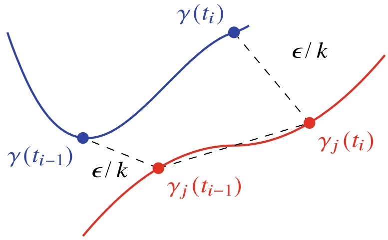


Fig. 3.1 The triangle inequality in the proof of Theorem 3.1.2 for bounding $d(\gamma(t_{i-1}), \gamma(t_i))$

Consequently, for all $j > N$, we have

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

Taking the liminf as j goes to ∞ and considering that ϵ is arbitrarily, we obtain

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

By taking the supremum over all partitions $\mathcal{P}$, we conclude the result. □

To show the existence of length-minimizing curves, we now recall a version of the Ascoli–Arzelà Compactness Theorem.

Theorem 3.1.3 (Ascoli–Arzelà) *In compact metric spaces, sequences of curves with uniformly bounded lengths contain subsequences that converge uniformly, up to reparameterization.*

Proof Let (M, d) be a compact metric space and $(\gamma_n)_{n\in\mathbb{N}}$ 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 3.4.7 and Exercise 3.4.8. 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 to 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 3.4.11. 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 $\mathcal{F} = \cup_{\lambda\in\Lambda}\mathcal{F}_\lambda$ and $\operatorname{diam}\mathcal{F}_\lambda \leq \epsilon$, for all $\lambda \in \Lambda$, where the diameter is defined by (3.3).

We start from 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 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, given such k_ε and h_ε, we define

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

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

$$\mathcal{F}_\lambda := \left\{\gamma \in \mathcal{F} : d(\gamma(x_i), p_{\lambda(i)}) < \varepsilon \quad \forall i \in \{1, \ldots, k_\varepsilon\}\right\},$$

which is the set of those curves for which the centers of the intervals get mapped into the balls according to λ. Clearly, we have $\mathcal{F} = \bigcup_{\lambda\in\Lambda} \mathcal{F}_\lambda$, for how we choose the points p_j. We 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 $\alpha, \beta \in \mathcal{F}_\lambda$. Therefore, the family $\mathcal{F}$ is precompact. □

The above argument actually proves various more general statements; see, for example, Exercise 3.4.12. The following result is an essential consequence of Ascoli–Arzelà Theorem 3.1.3.

Proposition 3.1.4 (Existence of Shortest Paths) *Let M be a compact metric space. For all $p, q \in M$ that a curve can join, there exists a curve γ from p to q such that*

$$L(\gamma) = \inf\{L(\sigma) : \sigma \textit{ curve from } p \textit{ to } q\}. \tag{3.5}$$

Proof Set L as the right-hand side of (3.5). If $L = \infty$, then there is nothing to prove since we can take any curve γ joining p to q. We next assume that $L < \infty$. Let γ_j curves from p to q with $L(\gamma_j) \to L$. By Ascoli–Arzelà Theorem 3.1.3, 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 3.1.2), we get $L(\gamma) \leq \liminf_{j\to\infty} L(\gamma_j) = L$. Hence, we conclude that $L(\gamma) = L$. □

3.1.3 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

$$d(p, q) = \inf\{\text{Length}_d(\gamma) \ : \ \gamma \text{ curve from } p \text{ to } q\}, \tag{3.6}$$

then (M, d) is called *length space* (or *path metric space*) and d is called an *intrinsic metric* (or *path distance*, or *length distance*). Notice that we have chosen 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 $d(p, q) = \text{Length}_d(\gamma)$, then (M, d) is called *geodesic space*, d is called a *geodesic metric*, and every such γ is called a *length-minimizing curve* joining p to q. Length-minimizing curves are also called *length minimizers*. Some authors use the term 'geodesic' to denote length minimizers or curves that are locally length-minimizing, in agreement with Riemannian geometry. In this text, we will call *geodesics* those length-minimizing curves that, in addition, are parametrized by arc length. In other words, a curve $\gamma : [a, b] \to M$ is a geodesic if

$$d(\gamma(s), \gamma(t)) = |s - t|, \qquad \forall s, t \in [a, b]. \tag{3.7}$$

Every geodesic space is a length space (Exercise 3.4.15). Not all length spaces are geodesic spaces; one reason can be a lack of completeness, as, for example, $\mathbb{R}^2\setminus\{(0, 0)\}$. As we will recall shortly, this is the only obstruction for locally compact spaces. Recall that a topological space is called *locally compact* if every point of the space has a local basis of compact neighborhoods. For other topological notions, we refer to any standard book in topology, such as [Mun75], and we suggest playing with the database of topological counterexamples: http://topology.jdabbs.com.

We prefer to assume the following stronger property, which gives compactness at every scale.

Definition 3.1.5 (Boundedly Compact) 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 $r > 0$. Equivalently, for every $p \in M$, the distance function $q \in M \mapsto d(p, q) \in \mathbb{R}$ is a proper function, in the sense that preimages of compact sets are compact.

Proposition 3.1.6 *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 to q with $L(\gamma) < d(p, q) + 1$. Notice that every other curve σ from p to q with $L(\sigma) \le L(\gamma)$ is inside $\overline{B}(p, d(p, q) + 1)$, which is compact by assumption. By Proposition 3.1.4, we have the existence of a shortest path and hence of a geodesic joining p to q since the distance is intrinsic. □

With a bit more topological arguments, one can prove the following stronger result. An explicit proof can be found in [BBI01, Proposition 2.5.22 and Theorem 2.5.23].

Theorem 3.1.7 (Hopf-Rinow-Cohn-Vossen) *If a length space (M, d) is complete and locally compact, then (M, d) is boundedly compact and, hence, a geodesic space.*

3.1.4 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 3.1.8 (Metric Derivative) Given a curve $\gamma\colon [a, b] \to M$ in 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 3.1.9 *For each Lipschitz curve $\gamma\colon [a, b] \to M$ in a metric space, we have:*

3.1.9.i. *The metric derivative $|\dot\gamma|$ exists almost everywhere;*
3.1.9.ii. $\mathrm{Length}_d(\gamma) = \int_a^b |\dot\gamma|(t)\,\mathrm{d}t$.

Proof For 3.1.9.i, we start by noticing that by the triangle inequality

$$|d(\gamma(s), y) - d(\gamma(t), y)| \le d(\gamma(s), \gamma(t)), \qquad \forall s, t \in [a, b],\ \forall y \in M, \tag{3.8}$$

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 (3.8) (and its equality when $x_n \to \gamma(t)$), we have

$$\sup_{n\in\mathbb{N}} |\varphi_n(s) - \varphi_n(t)| = d(\gamma(s), \gamma(t)). \tag{3.9}$$

Notice that each $\varphi_n : [a, b] \to \mathbb{R}$ is Lipschitz with the same Lipschitz constant as γ, and therefore differentiable almost everywhere and absolutely continuous, by the one-dimensional version of Rademacher Theorem; see [Fol99, Section 3.5]. Let

$$m(t) := \sup_n |\dot{\varphi}_n(t)|\,.$$

We claim that

$$|\dot{\gamma}|\,(t) = m(t), \qquad \text{for almost all } t. \tag{3.10}$$

For a first inequality, note that for each point t of differentiability for φ_n, we have from (3.9) that

$$|\dot{\varphi}_n|\,(t) \overset{\text{def}}{=} \lim_{h\to 0} \frac{|\varphi_n(t+h) - \varphi_n(t)|}{|h|} \overset{(3.9)}{\leq} \liminf_{h\to 0} \frac{d(\gamma(t+h), \gamma(t))}{|h|}.$$

Hence

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

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

$$\begin{aligned} d(\gamma(t), \gamma(s)) &\overset{(3.9)}{=} \sup_n |\varphi_n(t) - \varphi_n(s)| \\ &= \sup_n \left| \int_s^t \dot{\varphi}_n(\tau)\, \mathrm{d}\tau \right| \\ &\leq \sup_n \int_s^t |\dot{\varphi}_n(\tau)|\, \mathrm{d}\tau \\ &\leq \int_s^t m(\tau)\, \mathrm{d}\tau. \end{aligned} \tag{3.11}$$

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, [Fol99, p.98], at each Lebesgue point t for m we have that

$$\limsup_{h\to 0} \frac{d(\gamma(t+h), \gamma(t))}{|h|} \overset{(3.11)}{\leq} \limsup_{h\to 0} \left| \frac{1}{h} \int_t^{t+h} m(\tau)\, \mathrm{d}\tau \right| = m(t).$$

So (3.10) holds, and in particular $|\dot{\gamma}|$ exists almost everywhere. The first part of Theorem 3.1.9 is proven.

Regarding 3.1.9.ii, we first prove one inequality. For every partition $(t_0, t_1, \dots, t_k)$ of $[a, b]$, for some $k \in \mathbb{N}$, we have

$$\begin{aligned}\sum_{i=1}^{k} d(\gamma(t_{i-1}), \gamma(t_i)) &\overset{(3.11)}{\le} \sum_{i=1}^{k} \int_{t_{i-1}}^{t_i} m(\tau)\, \mathrm{d}\tau \\ &\overset{(3.10)}{=} \sum_{i=1}^{k} \int_{t_{i-1}}^{t_i} |\dot{\gamma}|(\tau)\, \mathrm{d}\tau \\ &= \int_a^b |\dot{\gamma}|(\tau)\, \mathrm{d}\tau.\end{aligned}$$

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

Regarding the other inequality, let $\varepsilon > 0$ and $n \ge 2$ such that $h := (b-a)/n \le \varepsilon$. We set $t_i := a + ih$ for $i \in \{0, 1, \dots, n\}$; so that we have $t_n = b$ and $b - \varepsilon < t_{n-1}$. Then

$$\begin{aligned}\int_a^{b-\varepsilon} d(\gamma(t), \gamma(t+h))\, \mathrm{d}t &\le \sum_{i=1}^{n-1} \int_{t_{i-1}}^{t_i} d(\gamma(t), \gamma(t+h))\, \mathrm{d}t \\ &= \int_0^h \sum_{i=1}^{n-1} d(\gamma(\tau + t_{i-1}), \gamma(\tau + t_i))\, \mathrm{d}\tau \\ &\le \int_0^h \mathrm{Length}(\gamma)\, \mathrm{d}\tau = h\, \mathrm{Length}(\gamma). \end{aligned} \tag{3.12}$$

Using Fatou's lemma [Fol99, p.52]:

$$\begin{aligned}\int_a^{b-\varepsilon} |\dot{\gamma}|(t)\, \mathrm{d}t &\overset{\text{def}}{=} \int_a^{b-\varepsilon} \liminf_{h\to 0^+} \frac{d(\gamma(t+h), \gamma(t))}{h}\, \mathrm{d}t \\ &\overset{\text{Fatou}}{\le} \liminf_{h\to 0^+} \frac{1}{h} \int_a^{b-\varepsilon} d(\gamma(t+h), \gamma(t))\, \mathrm{d}t \\ &\overset{(3.12)}{\le} \mathrm{Length}(\gamma).\end{aligned}$$

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

Example 3.1.10 (Finite-Dimensional Normed Spaces) Some first interesting examples of geodesic metric spaces are given by finite-dimensional normed spaces. Let $(V, \|\cdot\|)$ be a finite-dimensional normed space. Equip V 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 a Lipschitz curve (either with respect to the distance d or, equivalently, with respect to any Euclidean distance). Hence, since V has finite dimension, by Rademacher's Theorem, the curve γ is differentiable almost everywhere and absolutely continuous. For every point t of differentiability for γ, we have

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

where $|\dot{\gamma}|(t)$ is the metric derivative and $\gamma'(t)$ instead denotes the (classical) derivative. Consequently, from Theorem 3.1.9 we infer

$$\text{Length}_d(\gamma) = \int_a^b \left\|\gamma'(t)\right\| \, \mathrm{d}t. \tag{3.13}$$

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

$$\|p-q\| \overset{\text{def}}{=} d(p,q) \leq \text{Length}_d(\gamma) = \int_a^b \left\|\gamma'(t)\right\| \, \mathrm{d}t. \tag{3.14}$$

We stress that with the curve $t \in [0,1] \mapsto tp + (1-t)q$, we get equality in (3.14). In conclusion, we have proved that *every finite-dimensional normed space is a geodesic space with straight lines being length minimizing*.

3.1.4.1 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 than 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 in a metric space (M,d). Hence, by Theorem 3.1.9 its metric derivative $|\dot{\gamma}|$ exists almost everywhere. The *energy* of γ (with respect to the distance d) is defined as

$$\text{Energy}_d(\gamma) := \frac{1}{2}\int_a^b (|\dot{\gamma}|(t))^2 \, \mathrm{d}t \tag{3.15}$$

Contrary to 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 3.1.11 *Let $\gamma : [a, b] \to M$ be a Lipschitz curve in a metric space (M, d) and $p, q \in M$. Then, the energy satisfies the following properties:*

3.1.11.i.

$$\text{Length}_d(\gamma) \leq \sqrt{2\,(b-a)\,\text{Energy}_d(\gamma)}.$$

3.1.11.ii. *If γ is parametrized by a multiple of the arc length, then*

$$\text{Length}_d(\gamma) = \sqrt{2\,(b-a)\,\text{Energy}_d(\gamma)}.$$

3.1.11.iii.

$$\inf\left\{\text{Length}_d(\gamma) \;:\; \gamma \textit{ from } p \textit{ to } q\right\}$$
$$= \inf\left\{\sqrt{2 \cdot \text{Energy}_d(\gamma)} \;:\; \gamma \textit{ Lipschitz, on } [0,1] \textit{ from } p \textit{ to } q\right\}.$$

3.1.11.iv. *A curve defined on an interval I, parametrized by a multiple of arc length, and going from p to q is length-minimizing among curves from p to q if and only if it is energy-minimizing among curves defined on I going from p to q.*

Proof The statements are straightforward consequences of Jensen's Inequality or Cauchy-Schwarz inequality. □

Remark 3.1.12 More generally, one can minimize the p-energies. For $p \in [1, \infty)$, the *p-energy* is defined as the L^p-norm of the metric derivative, up to possibly a normalizing multiplicative constant. For $p = \infty$, the ∞-energy is the essential supremum of the metric derivative. The 1-energy is the length, by Theorem 3.1.9.ii. Because of Hölder's inequality, for $p \in (1, \infty]$, minimizing the p-energy among curves parametrized on $[0, 1]$, exactly determines the length-minimizing curves parametrized by a multiple of arc length.

3.1.5 Isometries, Lipschitz Maps, and Quasi-Isometries

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

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

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

Two functions α, β defined on the same set X are *bi-Lipschitz 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 bi-Lipschitz equivalence will be distances and measures. Notice that in particular, two distances d_1, d_2 on the same set M are *bi-Lipschitz equivalent* if and only if the identity map (M, d_1) to (M, d_2) is bi-Lipschitz.

The notion of bi-Lipschitz map can be further generalized to quasi-isometry. For the definition of the latter, we first introduce the notion of net. If $X, Y \subset M$ are subsets of a metric space M and $r > 0$, we say that X is an *r-net* for Y if

$$Y \subset B(X, r) := \{p \in M : d(p, X) < r\}.$$

Definition 3.1.13 (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 an (L, C)-*quasi-isometric embedding*, with $L \geq 1$ and $C \geq 0$, if

$$\frac{1}{L}\, d_2\big(f(x), f(y)\big) - C \leq d_1(x, y) \leq L\, d_2\big(f(x), f(y)\big) + C \quad \text{for all} \quad x, y \in M_1.$$

Moreover, an (L, C)-quasi-isometric embedding is called an (L, C)-*quasi-isometry* if $f(M_1)$ is a C-net for M_2. Two metric spaces M_1 and M_2 are called *quasi-isometric* if a quasi-isometry exists between them.

3.1.6 Hausdorff Measures and Dimension

A collection $\mathscr{F}$ of subsets 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 3.1.14 (Measure) A *measure* on a σ-algebra $\mathscr{F}$ is a function $\mu : \mathscr{F} \to [0, +\infty]$ such that

3.1.14.i. $\mu(\emptyset) = 0$;
3.1.14.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.,

$$\{A_n\}_{n\in\mathbb{N}} \subset \mathscr{F} \Longrightarrow \mu\left(\bigcup_{n\in\mathbb{N}} A_n\right) \leq \sum_{n=0}^{\infty} \mu(A_n), \tag{3.16}$$

see Exercise 3.4.19.

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 in a metric space M, then $\mu(B_M(p, r))$ is defined for all $p \in M$ and all $r > 0$.

For the following definition, we use the notion of diameter from (3.3).

Definition 3.1.15 (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

$$\mathcal{H}^Q_\delta(S) := \inf\left\{\sum_{i=1}^{\infty} (\operatorname{diam}(E_i))^Q : E_i \subseteq S,\ \operatorname{diam} E_i < \delta,\ S \subseteq \bigcup_{i=1}^{\infty} E_i\right\}, \tag{3.17}$$

with the 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, as explained in [Fol99], which gives a measure when restricted to the Borel σ-algebra.

Proposition 3.1.16 *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 3.1.17 (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 3.1.16. We notice that Lipschitz maps increase in a controlled way the Hausdorff measures; see Exercise 3.4.22. Consequently, the Hausdorff dimension is preserved by bi-Lipschitz homeomorphisms. Next, we observe that on metric spaces on which balls grow as powers of the radius, the Hausdorff dimension is precisely the exponent of these powers.

Theorem 3.1.18 *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*

$$\frac{1}{C} r^Q \leq \mu(B(p, r)) \leq C r^Q, \qquad \forall p \in M,\ \forall r \in (0, R]. \tag{3.18}$$

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

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

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

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

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 of 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 3.4.21.

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 (3.16) of μ, and the inclusions

$$\bigcup_i B(p_i, \operatorname{diam}(E_i)) \supset \bigcup_i E_i \supset B(p, R),$$

we have

$$\begin{aligned}
\mathcal{H}^Q_\delta(B(p,R)) &\geq \frac{1}{C}\sum_i \mu(B(p_i, \operatorname{diam} E_i)) - \epsilon \\
&\overset{(3.16)}{\geq} \frac{1}{C}\mu\left(\bigcup_i B\left(p_i, \operatorname{diam}(E_i)\right)\right) - \epsilon \\
&\geq \frac{1}{C}\mu(B(p,R)) - \epsilon \\
&\geq \frac{1}{C^2}R^Q - \epsilon.
\end{aligned}$$

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 (3.16) of the Hausdorff measure, also the last statement of the theorem follows.

□

Remark 3.1.19 The above proof actually shows that the Q-dimensional Hausdorff measure $\mathcal{H}^Q$ is bi-Lipschitz equivalent to the measure μ. In particular, the measure $\mathcal{H}^Q$ satisfies equation (3.18), with possibly some other choice for the constant C. We shall rephrase the last theorem using the following definition.

Definition 3.1.20 (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

$$\frac{1}{C}r^Q \leq \mu(B(p,r)) \leq Cr^Q, \qquad \forall p \in M, \ \forall r \in (0, R], \tag{3.19}$$

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

The following result is a consequence of Theorem 3.1.18.

Corollary 3.1.21 *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.*

We come back to considering curves. Using the Hausdorff measure, we rephrase the notion of length for injective curves.

Proposition 3.1.22 *If $\gamma : I \to M$ is an injective curve in a metric space M, then*

$$\mathcal{H}^1(\gamma(I)) = \mathrm{Length}(\gamma). \tag{3.20}$$

Proof We shall focus on the case when $\mathrm{Length}(\gamma) < \infty$ and leave the other case to the reader; see Exercise 3.4.24. Thus, we reparametrize $\gamma : [0, \ell] \to M$ by arc length. For proving (3.20), 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 $J_1, \dots, J_n$ of diameter less than δ. Since γ is parametrized by arc length, then it is 1-Lipschitz, and therefore, we have $\mathrm{diam}\, \gamma(J_j) < \delta$, for $j \in \{1, \dots, n\}$. Hence

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

where we have used in the first inequality that $(\gamma(J_j))_j$ is a admissible cover for (3.17) and in the second inequality that γ is 1-Lipschitz. Taking the limit for $\delta \to 0$, we infer the desired inequality in (3.20).

For the inequality $\geq$, we shall use the general bound

$$d(\gamma(s), \gamma(t)) \leq \mathcal{H}^1(\gamma([s, t]));$$

see Exercise 3.4.23. In fact, take a partition $t_0 < t_1 < \ldots < t_k$ of the interval I. Then we bound

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

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

3.1.7 Submetries

Definition 3.1.23 (Submetry) A map $\pi : X \to Y$ between metric spaces is a *submetry* if

$$\pi(\bar{B}(p, r)) = \bar{B}(\pi(p), r), \qquad \forall p \in X, \forall r > 0. \tag{3.21}$$

We stress that in (3.21), we consider closed balls. For boundedly compact metric spaces, it is equivalent to consider open balls; see Exercises 3.4.34 and 3.4.35. Also, notice that submetries are 1-Lipschitz; see Exercise 3.4.33. In addition, they are open maps. They are surjective if the distance function on the target is finite-valued.

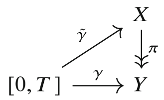


Fig. 3.2 In the presence of a submetry π, each geodesic γ in the target space can be lifted to a geodesic $\tilde{\gamma}$ in the source space. Whereas each geodesic $\tilde{\gamma}$ realizing the distance between fibers projects to a geodesic γ in the target

We can equivalently restate the condition of submetry with the notion of parallel sets, in the following sense: two subsets A and B of a metric space X are *parallel*, if for all $a \in A$ and all $b \in B$, there are $a' \in A$ and $b' \in B$ such that $d(A, B) = d(a, b') = d(a', b)$; see Exercise 3.4.38 for other viewpoints. Then, a map $\pi : X \to Y$ is a submetry if and only if its fibers are parallel and the distance in Y is exactly the distances of the fibers; see Exercises 3.4.37 and 3.4.39.

3.1.7.1 Lifting of Curves via Submetries

Submetries have the property that geodesics in the target space can be lifted to geodesics in the source space; see the diagram in Fig. 3.2. This lifting property will be important in Sect. 10.1.1 when we lift curves from the abelianization of nilpotent simply connected sub-Finsler Lie groups. 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.

In some settings, this lifting property is equivalent to the submetry property; see Proposition 3.1.28.

Proposition 3.1.24 *Let $\pi : X \to Y$ be a submetry between metric spaces. Assume X to be boundedly compact. Then, for every 1-Lipschitz curve $\gamma : [0, T] \to Y$ and every $x \in \pi^{-1}(\gamma(0))$, there is a 1-Lipschitz curve $\tilde{\gamma} : [0, T] \to X$ such that $\tilde{\gamma}(0) = x$ and $\pi \circ \tilde{\gamma} = \gamma$.*

Proof Fix $m \in \mathbb{N}$. We define

$$\tilde{\gamma}_m : [0, T] \cap \frac{1}{m}\mathbb{Z} \to X$$

as follows. Set $\tilde{\gamma}_m(0) = x$. By induction on $k \in \mathbb{Z} \cap [0, mT]$, assume that the value $\tilde{\gamma}_m\left(\frac{k}{m}\right)$ has been defined with $\pi\left(\tilde{\gamma}_m\left(\frac{k}{m}\right)\right) = \gamma\left(\frac{k}{m}\right)$. By the submetry assumption, we know that

$$\pi\left(\overline{B}\left(\tilde{\gamma}_m\left(\tfrac{k}{m}\right), \frac{1}{m}\right)\right) = \overline{B}\left(\gamma\left(\tfrac{k}{m}\right), \frac{1}{m}\right) \ni \gamma\left(\tfrac{k+1}{m}\right),$$

where the belonging follows from the fact that γ is 1-Lipschitz. We therefore define $\tilde{\gamma}_m\left(\frac{k+1}{m}\right)$ to be any point of $\overline{B}\left(\tilde{\gamma}_m\left(\frac{k}{m}\right), 1/m\right)$ that maps onto $\gamma\left(\frac{k+1}{m}\right)$ under π.

It follows that $\tilde{\gamma}_m$ is 1-Lipschitz for each $m \in \mathbb{N}$. By a standard Ascoli-Arzelà type argument, as in Theorem 3.1.3, a sub-sequence of $\{\tilde{\gamma}_m\}$ converges as $m \to \infty$ to a curve $\tilde{\gamma}$ as desired. □

As an immediate consequence, we can lift geodesics, which are isometric embeddings of intervals as in (3.7). As in the diagram of Fig. 3.2, every geodesic γ in the base is the projection of a geodesic $\tilde{\gamma}$ on the top.

Corollary 3.1.25 *Let $\pi : X \to Y$ be a submetry between metric spaces. Assume X is boundedly compact. Then for every geodesic $\gamma : [0, T] \to Y$ and every $x \in \pi^{-1}(\gamma(0))$ there exists a geodesic $\tilde{\gamma} : [0, T] \to X$ such that $\tilde{\gamma}(0) = x$ and $\pi \circ \tilde{\gamma} = \gamma$.*

Proof We apply Proposition 3.1.24 to such a γ in order to get some $\tilde{\gamma} : [0, T] \to X$ with the desired properties: to additionally see that it is a geodesic, we bound

$$|t - s| = d(\gamma(s), \gamma(t)) \leq d(\tilde{\gamma}(s), \tilde{\gamma}(t)) \leq |t - s|, \qquad s, t \in [0, T],$$

where we used that γ is a geodesic and that π and $\tilde{\gamma}$ are 1-Lipschitz. □

Remark 3.1.26 Proposition 3.1.24 holds in a bigger generality. In fact, the notion of submetry generalizes to Lipschitz quotients. One can use Lipschitz quotients to lift rectifiable curves, as above. This result was first proven in [Bat+99, Lemma 4.4] and [Joh+00, Lemma 2.2]. Though stated there for $\mathbb{R}^n$-targets, the proof works the same way in the setting above, as was observed in [DK18, Lemma 4.3] and [DK19, Lemma 3.3], where one can find a proof written in this generality. We also point out the article [DK20].

By considering the diagram in Fig. 3.2, one shows that each geodesic $\tilde{\gamma}$ realizing the distance between fibers projects to a geodesic γ in the target; see Exercise 3.4.40. A consequence of such a small argument is the following fact.

Proposition 3.1.27 *Let $\pi : X_1 \to X_2$ be a surjective submetry between metric spaces. If X_1 is a geodesic space, then so is X_2.*

Proof Take $p, q \in X_2$. Pick any $\tilde{p} \in \pi^{-1}(p)$, since π is surjective. Because π is a submetry, there is $\tilde{q} \in \pi^{-1}(q)$ such that $d(p, q) = d(\tilde{p}, \tilde{q})$. Since X_1 is a geodesic space, there is an isometric embedding $\tilde{\gamma} : [0, T] \to X_1$ with $\tilde{\gamma}(0) = \tilde{p}$, $\tilde{\gamma}(T) = \tilde{q}$, and $T = d(\tilde{p}, \tilde{q})$. Since both π and $\tilde{\gamma}$ are 1-Lipschitz, then so is $\gamma := \pi \circ \tilde{\gamma}$, which is a curve from p to q. We in fact have that γ is a geodesic, since $d(p, q) \leq L(\gamma) \leq T = L(\tilde{\gamma}) = d(\tilde{p}, \tilde{q}) = d(p, q)$. □

Proposition 3.1.28 *Let $\pi : X \to Y$ be a map between geodesic metric spaces. Assume that*

3.1.28.1. for every curve $\tilde{\gamma}$ in X we have $L(\pi \circ \tilde{\gamma}) \leq L(\tilde{\gamma})$ and
3.1.28.2. for every rectifiable curve $\gamma : [0, T] \to Y$ and every $\tilde{p} \in \pi^{-1}(\gamma(0))$ there is a curve $\tilde{\gamma} : [0, T] \to X$ such that

3.1.28.2.i. $\tilde{\gamma}(0) = \tilde{p}$,

3.1.28.2.ii. $\pi \circ \tilde{\gamma} = \gamma$, *and*
3.1.28.2.iii. $L(\tilde{\gamma}) = L(\gamma)$.

Then π is a submetry.

Proof We first check that π is 1-Lipschitz. Since the space X is assumed geodesic, for every $\tilde{p}_1$ and $\tilde{p}_2$ there is a curve $\tilde{\gamma}$ in X joining $\tilde{p}_1$ to $\tilde{p}_2$ such that $d(\tilde{p}_1, \tilde{p}_2) = L(\tilde{\gamma})$. Then, since $\pi \circ \tilde{\gamma}$ joins $\pi(\tilde{p}_1)$ to $\pi(\tilde{p}_2)$ and is shorter than $\tilde{\gamma}$, we have

$$d(\pi(\tilde{p}_1), \pi(\tilde{p}_2)) \leq L(\pi \circ \tilde{\gamma}) \overset{3.1.28.1}{\leq} L(\tilde{\gamma}) = d(\tilde{p}_1, \tilde{p}_2).$$

Thus, the map π is 1-Lipschitz, i.e., $\pi(\bar{B}(\tilde{p}, r)) \subseteq \bar{B}(\pi(\tilde{p}), r)$, for all $\tilde{p} \in X$, and all $r > 0$. To prove the opposite inclusion, take $q \in \bar{B}(\pi(\tilde{p}), r)$. Since the space Y is assumed geodesic, there is a curve γ in Y joining $\pi(\tilde{p})$ to q with $L(\gamma) \leq r$. By 3.1.28.2, there is a curve $\tilde{\gamma} : [0, T] \to X$ with the three properties 3.1.28.2.i–iii, with our choice of $\tilde{p}$. Then, the endpoint $\tilde{q}$ of $\tilde{\gamma}$ is such that $\pi(\tilde{q}) = q$ and

$$d(\tilde{p}, \tilde{q}) \leq L(\tilde{\gamma}) \overset{3.1.28.2.iii}{=} L(\gamma) \leq r.$$

Thus $\tilde{q} \in \bar{B}(\tilde{p}, r)$, and therefore $q \in \pi(\bar{B}(\tilde{p}, r))$. Since q was arbitrary in $\bar{B}(\pi(\tilde{p}), r)$ we infer $\bar{B}(\pi(\tilde{p}), r) \subseteq \pi(\bar{B}(\tilde{p}, r))$, for all $\tilde{p} \in X$, and all $r > 0$. Altogether, the map π satisfies (3.21). □

Proposition 3.1.29 *Let $\pi : M_1 \to M_2$ be a smooth submetry between smooth manifolds equipped with admissible distances. If each pair of points in M_1 can be joined by a C^k geodesic (or a piecewise C^k geodesic), then the same is valid on M_2.*

Proof Take $p, q \in M_2$. As in the previous proof, pick any $\tilde{p} \in \pi^{-1}(p)$ and $\tilde{q} \in \pi^{-1}(q)$ such that $d(p, q) = d(\tilde{p}, \tilde{q})$. Pick a 'good' geodesic $\tilde{\gamma} : [0, T] \to G$ from $\tilde{p}$ to $\tilde{q}$, which exists by assumption. Define $\gamma := \pi \circ \tilde{\gamma}$. We stress that γ connects p to q and has the same regularity as $\tilde{\gamma}$ since π is smooth. Moreover, since π is 1-Lipschitz, then γ has length at most $d(p, q)$. Therefore, it is a geodesic. □

3.2 Differential Geometry

3.2.1 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 nor the concept of a smooth map between manifolds, to which we refer any introductory book, such as [Lee13]. We denote by $C^\infty(M)$ the space of C^∞ functions from M to $\mathbb{R}$. We shall prefer the following viewpoint 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 3.2.1 (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

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

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

$$X(f)(p) = \sum_{j=1}^{n} X^j(p)\partial_j(f)(p), \qquad \forall f \in C^\infty(U), \forall p \in U.$$

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 space* $T_pM := \pi^{-1}(p)$ 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 Leibniz 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

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

gives a tangent vector at p. Hence, vector fields on M are sections of the tangent bundle TM, and moreover, one puts on TM the structure of a 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. For this reason, we write $\Gamma(TM)$ for $\mathrm{Vec}(M)$.

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 in M, then we have the formula

$$\mathrm{d}F_{\sigma(t)}(\sigma'(t)) = (F \circ \sigma)'(t), \tag{3.22}$$

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 from p at time t with respect to X*. Namely, we have

$$\begin{cases} \Phi_X^0(p) = p, \\ \dfrac{\mathrm{d}}{\mathrm{d}t}\Phi_X^t(p) = X_{\Phi_X^t(p)}. \end{cases} \tag{3.23}$$

One of the 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 that we consider.

Definition 3.2.2 (Lie Bracket) The *Lie bracket of vector fields* on a manifold M is the map

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

defined with any of the equivalent viewpoints a–d:

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

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

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

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

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

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

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

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

The Lie bracket induces on $\mathrm{Vec}(M)$ an infinite-dimensional Lie algebra structure; see Definition 5.1.2. Clearly, the push-forward via a diffeomorphism commutes with the Lie bracket operation; see Exercise 3.4.46. Here, 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,

$$(F_*X)f := [X(f \circ F)] \circ F^{-1}, \qquad \forall f \in C^\infty(N). \tag{3.24}$$

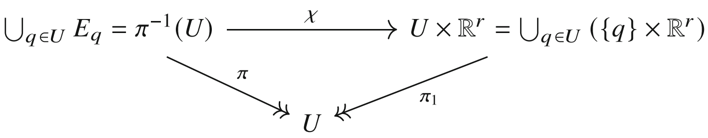
3.2.2 Vector Bundles

A simple example of a 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 a vector bundle of rank $\dim(M)$ over a manifold M is the tangent bundle TM of M. The abstract definition is the following.

Definition 3.2.3 (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:

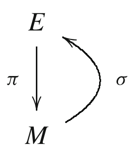
1. The fiber $E_p := \pi^{-1}(p)$ is equipped with the structure of a vector space of dimension r.
2. There is a neighborhood U of p in M and a diffeomorphism $\chi : \pi^{-1}(U) \to U \times \mathbb{R}^r$ such that

    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 map χ is called a *local trivialization*.

Definition 3.2.4 (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.



Definition 3.2.5 (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.

3.2.3 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 that

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

defines a smooth function from M to $\mathbb{R}$. This smooth assignment of an inner product ρ_p to each tangent space T_pM is called a *metric tensor*, or *Riemannian metric tensor*. A metric tensor will also be denoted by $\langle \cdot, \cdot \rangle$. Endowed with one such 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 $\partial_1, \dots, \partial_n$ from Definition 3.2.1, 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. Namely, on each tangent space, we have a norm but it is not necessarily induced by a scalar 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 3.2.9). 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 function is an example of a continuously varying norm.

Definition 3.2.6 A *continuously varying norm* on a differentiable manifold M is a continuous function from TM to $[0, \infty)$ usually denoted by $\| \cdot \|$

$$\| \cdot \| : TM \to [0, \infty), \qquad X \in TM \mapsto \|X\|,$$

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 $\forall X, Y \in T_p M$;
3. $\|X\| = 0 \Rightarrow X = 0$.

Definition 3.2.7 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 the sense of Definition 3.2.6. In this case, the function $\|\cdot\|$ is also called *Finsler structure*.

Example 3.2.8 There are at least two situations that we want the reader to keep in mind:

3.2.8.i. Every Riemannian manifold $(M, \langle\cdot,\cdot\rangle)$ naturally has the structure of a Finsler manifold.
3.2.8.ii. Every finite-dimensional normed vector space naturally has the structure of a Finsler manifold.

Remark 3.2.9 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 3.2.6. 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 3.2.6 is assumed only for $\lambda > 0$.

3.3 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 3.3.1 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 3.4.44).

As it is usual for other notions 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 3.4.43).

Definition 3.3.2 (Length of a Curve in a Finsler Manifold) Let $(M, \|\cdot\|)$ be a Finsler manifold in the sense of Definition 3.2.7. Let $\gamma : [a, b] \to M$ be an absolutely continuous curve. We define

$$\mathrm{Length}_{\|\cdot\|}(\gamma) := \int_a^b \|\dot{\gamma}(t)\| \, \mathrm{d}t. \tag{3.25}$$

We remark that the *Finsler length* (3.25) of an absolutely continuous curve is finite because the derivative is integrable by assumption and $\|\cdot\|$ is continuous.

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

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

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

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

$$d_{\|\cdot\|}(p, q) := \inf \mathrm{Length}_{\|\cdot\|}(\gamma), \qquad \forall p, q \in M, \tag{3.26}$$

where the infimum is taken among 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 $p = q$. To prove this property, we claim that, locally in a coordinate system, every Finsler structure (as every Riemannian structure) is bi-Lipschitz equivalent to the Euclidean structure, i.e., for some $c > 0$, we have

$$c^{-1}\|\cdot\| \le \|\cdot\|_{\mathbb{E}} \le c\|\cdot\|, \tag{3.27}$$

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 has a 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 $c^{-1} \le \|v\| \le c$. By homogeneity, we have (3.27) on K.

Consequently, based on (3.27), we can establish the bi-Lipschitz equivalence between distance functions. Specifically, we have proven that every two Finsler distance functions on the same manifold are bi-Lipschitz equivalent on compact sets. We summarize our findings in the following proposition.

Proposition 3.3.3 *On every Finsler manifold in local coordinates, on compact sets, the Finsler distance function is bi-Lipschitz equivalent to the Euclidean distance function. Consequently, on every compact set of every manifold, every Finsler structure is bi-Lipschitz 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 3.2.6, we associated a length structure as in (3.25) and a distance function as in (3.26). The distance function then induces another length structure, as in Definition 3.1.1. Next, we show that these two length structures coincide.

Proposition 3.3.4 *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 $\gamma : [a, b] \to M$ is an absolutely continuous curve, then*

$$\mathrm{Length}_{d_{\|\cdot\|}}(\gamma) = \mathrm{Length}_{\|\cdot\|}(\gamma). \tag{3.28}$$

Proof To prove the $\leq$ inequality in (3.28), notice that for all $t, s \in [a, b]$ we have

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

where the infimum is taken over all AC curves σ from $\gamma(s)$ to $\gamma(t)$. Using the definition of length, (3.4), and the additivity of integrals, we deduce that $\mathrm{Length}_{d_{\|\cdot\|}} \leq \mathrm{Length}_{\|\cdot\|}$.

Regarding the other inequality, we shall use the fact 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 $v \in T_x U \simeq \mathbb{R}^n$. 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

$$\frac{1}{K} F(q, v) \leq F(p, v) \leq K F(q, v), \qquad \forall q \in U_p, \forall v \in \mathbb{R}^n. \tag{3.29}$$

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

$$\gamma([a_{i-1}, a_i]) \subseteq U_{p_i}, \qquad \forall i \in \{1, \ldots, n\}. \tag{3.30}$$

Let us denote by d_i the distance induced by the (constant) norm $F(p_i, \cdot)$. Since we are in the case of a normed vector space (see Example 3.1.10), we have

$$\mathrm{Length}_{F(p_i,\cdot)} = \mathrm{Length}_{d_i}\,. \tag{3.31}$$

Moreover, as a consequence of (3.29), we have

$$d_i \le K d_{\|\cdot\|}. \tag{3.32}$$

Thus, using (3.29), (3.31), and (3.32), together with (3.30), we obtain that

$$\begin{aligned}
\mathrm{Length}_{\|\cdot\|}(\gamma) &\overset{\text{def}}{=} \int_a^b F(\gamma(t), \dot\gamma(t))\, \mathrm{d}t \\
&= \sum_{i=1}^n \int_{a_{i-1}}^{a_i} F(\gamma(t), \dot\gamma(t))\, \mathrm{d}t \\
&\overset{(3.29)}{\le} K \sum_{i=1}^n \int_{a_{i-1}}^{a_i} F(p_i, \dot\gamma(t))\, \mathrm{d}t \\
&\overset{(3.31)}{=} K \sum_{i=1}^n \mathrm{Length}_{d_i}(\gamma|_{[a_{i-1},a_i]}) \\
&\overset{(3.32)}{\le} K^2 \sum_{i=1}^n \mathrm{Length}_{d_{\|\cdot\|}}(\gamma|_{[a_{i-1},a_i]}) = K^2\, \mathrm{Length}_{d_{\|\cdot\|}}(\gamma).
\end{aligned}$$

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

Remark 3.3.5 Let $\gamma : [a, b] \to M$ be a curve on a manifold that 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:

3.3.5.i. If γ is AC, then $\mathrm{Length}_{\|\cdot\|}(\gamma) = \mathrm{Length}_{d_{\|\cdot\|}}(\gamma)$ and both these quantities are finite; see Proposition 3.3.4.

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

3.3.5.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, the curve γ is AC.

3.4 Exercises

Exercise 3.4.1 Let (M, d) be a metric space equipped with its natural topology.

(i) If M is connected, then d is finite.
(ii) In general, the function d is finite on each connected component of M.

Exercise 3.4.2 (Von Koch Showflake) Start with a curve that is the boundary of an equilateral triangle and change iteratively each line segment as follows to get a new curve: Remove the middle third of each segment and replace it with the outward pointing side of an equilateral triangle without the base, so that you get four line segments of equal length. Repeat now this procedure infinitely and add at each iteration smaller and smaller equilateral triangles on the previous line segments. The resulting limit curve K is called the *von Koch snowflake*; see Fig. 3.3 for a visual representation. Equip $K \subset \mathbb{R}^2$ with the distance function $d_{\mathbb{R}^2}$ of the plane $\mathbb{R}^2$. Then, the metric space K is bi-Lipschitz equivalent to the circle $\mathbb{S}^1 \subset \mathbb{R}^2$ equipped with the distance $(d_{\mathbb{R}^2})^{\log(3)/\log(4)}$.

Exercise 3.4.3 (Showflake of a Metric Space) Let (M, d) be a metric space and $\alpha \in (0, 1)$. Then, the pair (M, d^α) is a metric space, called the α-*snowflake* of (M, d). The Hausdorff dimension of (M, d^α) is α times the Hausdorff dimension of (M, d). The von Koch snowflake from Exercise 3.4.2 is the $\log(3)/\log(4)$-snowflake of the unit circle $\mathbb{S}^1$ and has dimension $\log(4)/\log(3) > 1$.

Exercise 3.4.4 The *mesh* of a partition $\mathcal{P} = (t_0, t_1, \dots, t_k)$ is defined as

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

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

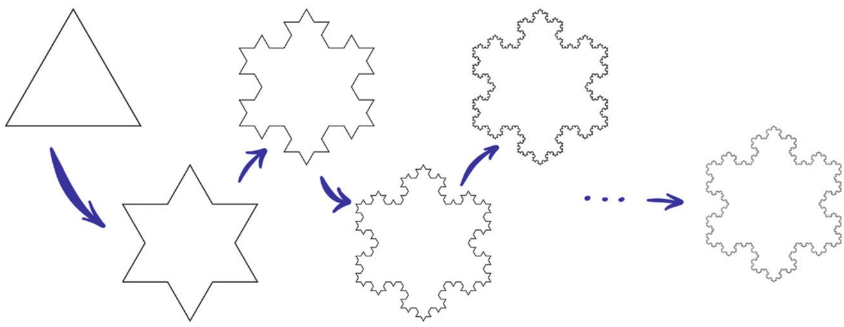


Fig. 3.3 Construction of the von Koch snowflake

Exercise 3.4.5 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 3.4.6 The length of a curve is independent of its parameterization. Namely, If $\gamma : I \to M$ is a curve in a metric space and $h : J \to I$ is a continuous weakly monotone surjection between intervals, then $L(\gamma) = L(\gamma \circ h)$. In this case, we say that h is a *change of parametrization* and that $\gamma \circ h$ is a *reparametrization* of γ.

Exercise 3.4.7 If $\gamma : [a, b] \to (M, d)$ is rectifiable, then it can be reparametrized by arc length.
Hint. Consider the change of parametrization given by $s \mapsto \text{Length}(\gamma|_{[a,s]})$.

Exercise 3.4.8 If $\gamma : [a, b] \to (M, d)$ is *parametrized with constant speed* s, with $s \in [0, \infty)$, i.e.,

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

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

Exercise 3.4.9 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 3.4.10 Let $f_n : X \to \mathbb{R}$ be a sequence of continuous functions on a topological space. Then, the function $\sup_n f_n$ is lower semicontinuous.
Hint. Adapt the proof of Theorem (3.1.2).

Exercise 3.4.11 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

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

Then, the pair $(\mathcal{F}, d_{\sup})$ is a complete metric space.

Exercise 3.4.12 (Ascoli–Arzelà) Let X and Y be metric spaces. Assume X is boundedly compact, and with a finite-valued distance function. If a family $\mathcal{F}$ of maps from X to Y is equicontinuous and, for each $x \in X$, the set $\{f(x) : f \in \mathcal{F}\}$ is precompact, then $\mathcal{F}$ is precompact in the uniform topology on compact sets.

Exercise 3.4.13 (The Space of L-Lipschitz Functions) Let X and Y be metric spaces, with X boundedly compact. Let $\mathtt{Lip}^L(X; Y)$ be the set of Lipschitz functions $X \to Y$ of Lipschitz constant at most L. Fix a base point $o \in X$. Consider the function

$$d_L(f, g) := \sup \left\{ \frac{d(f(x), g(x))}{n^2} : \ n \in \mathbb{N}, \ x \in B(o, n) \right\}.$$

3.4.13.i. The function d_L is a distance function on $\mathtt{Lip}^L(X; Y)$.
3.4.13.ii. Convergence with respect to d_L is equivalent to uniform convergence on compact sets.
3.4.13.iii. The space $(\mathtt{Lip}^L(X; Y), d_L)$ is separable.

Solution of 3.4.13.ii. Let $\{f_k\}_k \subset \mathtt{Lip}^L(X; Y)$ and $f \in \mathtt{Lip}^L(X; Y)$. Suppose that $\lim_{k\to\infty} d_L(f_k, f) = 0$. If $E \subset X$ is compact, then there is $N \in \mathbb{N}$ such that $E \subset B(o, N)$. Since

$$\sup\{d(f_k(x), f(x)) : x \in B(o, N)\} \leq N^2 d_L(f_k, f) \to 0,$$

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 $\{o\}$ is compact, there is $C > 0$ such that $d(f_k(o), f(o)) \leq C$ for all $k \in \mathbb{N}$. Notice that, for all $n \in \mathbb{N}$ and $x \in B(o, n)$, we have

$$\frac{d(f_k(x), f(x))}{n^2} \leq \frac{d(f_k(x), f_k(o)) + d(f_k(o), f(o)) + d(f(o), f(x))}{n^2} \leq \frac{2L}{n} + \frac{C}{n^2}.$$

Therefore, there is $N \in \mathbb{N}$ such that $\frac{d(f_k(x), f(x))}{n^2} < \epsilon$ for all $n \geq N$ and $x \in B(o, n)$. Since X is boundedly compact, then $B(o, N)$ is precompact. Hence, there is $K \in \mathbb{N}$ be such that

$$\sup\{d(f_k(x), f(x)) : x \in B(o, N)\} \leq \epsilon, \qquad \forall k > K.$$

Then, for $k > K$, we have $d_L(f_k, f) \leq \epsilon$. We conclude that $\lim_{k\to\infty} d_L(f_k, f) = 0$.
Solution of 3.4.13.iii. Fixing $o' \in Y$, by Ascoli–Arzelà result (see Exercise 3.4.12), for every $n \in \mathbb{N}$ the set

$$\mathscr{K}(n) := \{f \in \mathtt{Lip}^L(X; Y) : f(o) \in \bar{B}(o', n)\}$$

is compact, hence separable. Since $\mathtt{Lip}^L(X; Y) = \bigcup_{n\in\mathbb{N}} \mathscr{K}(n)$ is a countable union of separable sets, then it is separable.

Exercise 3.4.14 Let $F : M_1 \to M_2$ be a map between two metric spaces that is K-Lipschitz. If γ is a curve in M_1, then $L(F \circ \gamma) \leq K \cdot L(\gamma)$.

Exercise 3.4.15 Every geodesic space is a length space—what is not automatic is that the distance is finite.

Exercise 3.4.16 Every complete locally compact length space is boundedly compact.
Hint. See [BBI01, Proposition 2.5.22].

Exercise 3.4.17 Let $L \geq 0$ and $F : M_1 \to M_2$ a map that is *locally L-Lipschitz*, i.e., for all $p \in M_1$ there is $r > 0$ such that $F_{B(p,r)}$ is L-Lipschitz. If M_1 is a length space, then F is L-Lipschitz.

Exercise 3.4.18 There is an example of a homeomorphism $F : M_1 \to M_2$ between 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 3.4.19 Each measure is countably subadditive as in (3.16).
Hint. Given countably many sets, split them into disjoint sets and apply 3.1.14.ii.

Exercise 3.4.20 Let $\gamma : [a, b] \to \mathbb{R}^n$ be absolutely continuous. The vector-valued function $\dot{\gamma}$ is unique up to a measure-zero subset of $[a, b]$.

Exercise 3.4.21 We have 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 3.4.22 If $F : M_1 \to M_2$ is an L-Lipschitz map between metric spaces, $Q \geq 0$, and $S \subset M_1$, then

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

Consequently, if $F : M_1 \to M_2$ is a bi-Lipschitz homeomorphism, then $\dim_H M_1 = \dim_H M_2$.

Exercise 3.4.23 (To be generalized in Exercise 3.4.25) For every continuous curve $\gamma : [a, b] \to M$ in 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 the 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 3.4.24 Complete the proof of Proposition 3.1.22 by showing that for every curve $\gamma : I \to M$ in a metric space if $\operatorname{Length}(\gamma) = \infty$, then $\mathcal{H}^1(\gamma(I)) = \infty$.
Hint. Use Exercise 3.4.23.

Exercise 3.4.25 For every connected subset X of a metric space, we have $\mathcal{H}^1(X) \geq \operatorname{diam}(X)$.

Exercise 3.4.26 (Maximal Separated Nets) Given a metric space M and $\delta > 0$, a set $E \subseteq M$ is called δ-*net* in M, or, simply, a *net*, if $M \subseteq B(E, \delta) := \{p \in M : d(p, E) < \delta\}$. While a set $E \subseteq M$ is called δ-*separated* if for all distinct $p, q \in E$, we have $d(p, q) \geq \delta$. Then, every separated set that is maximal (with respect to the inclusion) is a net. Moreover, for each $\delta > 0$, every compact metric space has a finite maximal δ-separated set.

Exercise 3.4.27 Let X and Y be metric spaces. Then, there exists a quasi-isometry from X to Y if and only if there is a net in X that is bi-Lipschitz homeomorphic to a net in Y.

Exercise 3.4.28 (Doubling Distance) A metric space M is called *doubling*, or, also, *metrically doubling*, if there is a constant $K \in \mathbb{N}$ such that, for all $p \in M$ and $r > 0$, there are $p_1, \dots, p_K$ such that $B(p, r) \subseteq \bigcup_{i=1}^{K} B(p_i, r/2)$. Then, a metric space M is doubling if and only if there is a constant $K \in \mathbb{N}$ such that, for every $d > 0$, every set in M of diameter d can be covered by K sets of diameter at most $d/2$.

Exercise 3.4.29 (Assouad Dimension) A metric space M is doubling if and only if there exist $C > 1$ and $\beta > 0$ such that, for every $p \in M$ and $r > 0$, every ϵr-separated set in the ball $B(p, r)$ in M has cardinality at most $C\epsilon^{-\beta}$. The infimum of such β's is called *Assouad dimension*.

Exercise 3.4.30 A metric space M is doubling if and only if there exist $C > 1$ and $\beta > 0$ such that, for every $d, \epsilon > 0$, every set in M of diameter d can be covered by at most $C\epsilon^{-\beta}$ sets of diameter at most ϵd.

Exercise 3.4.31 (Doubling Measure) A measure μ on a metric space M is called *doubling* if there is a constant $C > 1$ such that

$$0 < \mu(B(p, 2r)) \leq C\mu(B(p, r)) < \infty, \qquad \forall p \in M, \forall r > 0. \tag{3.33}$$

(i) If a metric space (M, d) admits a doubling measure μ, then the space is metrically doubling. Hence, we call the triple (M, d, μ) a *doubling metric measure space*.
(ii) On every metrically doubling metric space that is complete, there is a doubling measure. See [Hei01, Theorem 13.3].

Exercise 3.4.32 Every doubling measure μ in a metric space M is α-*homogeneous*, for some $\alpha > 0$ (and some $C > 0$), in the sense:

$$\frac{\mu(B(p, r))}{\mu(B(p, R))} \leq C\left(\frac{r}{R}\right)^{\alpha}, \qquad \forall p \in M, \forall 0 < r \leq R < \operatorname{diam}(M). \tag{3.34}$$

See [Hei01, Equation (4.16)].

Exercise 3.4.33 For every map $\pi : X \to Y$ between metric spaces, the following are equivalent:

i) $\pi(B(p, r)) \subseteq B(\pi(p), r), \qquad \forall p \in X, \forall r > 0;$
ii) $\pi(\bar{B}(p, r)) \subseteq \bar{B}(\pi(p), r), \qquad \forall p \in X, \forall r > 0;$
iii) π is 1-Lipschitz.

Exercise 3.4.34 If $\pi : X \to Y$ is a submetry as in Definition 3.1.23, then

$$\pi(B(p,r)) = B(\pi(p), r), \qquad \forall p \in X, \forall r > 0, \tag{3.35}$$

where now we are considering open balls.

Exercise 3.4.35 Let X and Y be metric spaces, with X assumed to be boundedly compact. Then $\pi : X \to Y$ is a submetry as in Definition 3.1.23 if and only if (3.35) holds. There are counterexamples when X is not boundedly compact; see Exercise 3.4.36.

Exercise 3.4.36 Consider the normed vector space $C^0([0,1];\mathbb{R})$ of continuous functions on the interval $[0,1]$ equipped with the norm given by the maximum of the absolute value. The map $f \in C^0([0,1];\mathbb{R}) \mapsto \int_0^{1/2} f(t)\,\mathrm{d}t - \int_{1/2}^1 f(t)\,\mathrm{d}t \in \mathbb{R}$ satisfies (3.35) but it is not a submetry.

Exercise 3.4.37 Let $\pi : X \to Y$ be a surjective map between metric spaces. Then π is a submetry if and only if for all $\hat{p}, \hat{q} \in Y$ and all $p \in \pi^{-1}(\hat{p})$, there exists $q \in \pi^{-1}(\hat{q})$ such that $d(p,q) = d(\hat{p},\hat{q}) = d(\pi^{-1}(\hat{p}), \pi^{-1}(\hat{q}))$.

Exercise 3.4.38 (Hausdorff Distance) Consider two subsets A and B of a metric space X. The *Hausdorff distance* between A and B is

$$d_H(A,B) := \max\left\{\sup_{a\in A}\{d(a,B)\}, \sup_{b\in B}\{d(A,b)\}\right\}.$$

If A and B are parallel, as in the sense defined at page 72, then $d(A,B) = d_H(A,B)$. If X is boundedly compact, A and B are closed, and $d(A,B) = d_H(A,B)$, then A and B are parallel.

Exercise 3.4.39 Let X be a metric space and Y be a nonempty set. Let $\pi : X \to Y$ be a surjective map. Assume that the fibers of π are parallel, i.e., for all $\hat{p} \in Y$, all $\hat{q} \in Y$, and all $p \in \pi^{-1}(\hat{p})$, one can find $q \in \pi^{-1}(\hat{q})$ such that $d(\pi^{-1}(\hat{p}), \pi^{-1}(\hat{q})) = d(p,q)$. Then

$$d(\hat{p},\hat{q}) := d(\pi^{-1}(\hat{p}), \pi^{-1}(\hat{q})), \qquad \forall \hat{p},\hat{q} \in Y,$$

defines a distance function on Y and π is a submetry from X onto Y.

Exercise 3.4.40 Let $\pi : X \to Y$ be a submetry between metric spaces. Let $p, q \in X$ be such that $d(p,q) = d(\pi^{-1}(\pi(p)), \pi^{-1}(\pi(q)))$. If $\tilde{\gamma}$ is a geodesic between p and q, then $\pi \circ \tilde{\gamma}$ is a geodesic between $\pi(p)$ and $\pi(q)$.
Hint. Check the proof of Proposition 3.1.27.

Exercise 3.4.41 If E is a vector bundle of rank r over a manifold M, then $\dim(E) = \dim(M) + r$.

Exercise 3.4.42 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.

Exercise 3.4.43 Let $\gamma : I \to M$ be a curve into a differentiable manifold. Then, the curve 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 3.4.44 Let $\gamma : I \to M$ be an absolutely continuous curve into a differentiable manifold. Let $\varphi_1, \varphi_2 : U \subset M \to \mathbb{R}^n$ be coordinate maps. Then, 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 sets.

Exercise 3.4.45 Every absolutely continuous curve in $\mathbb{R}^n$ can be re-parametrized to be a Lipschitz curve with respect to the Euclidean distance.

Exercise 3.4.46 The push-forward commutes with the Lie bracket: if $F : M \to N$ is a diffeomorphism of differentiable manifolds, then

$$[F_*X, F_*Y] = F_*[X, Y], \qquad \forall X, Y \in \Gamma(TM). \tag{3.36}$$

Exercise 3.4.47 If X and Y are vector fields tangent to a smooth submanifold $N \subseteq M$ of a differentiable manifold M, then also $[X, Y]$ is tangent to N.

Exercise 3.4.48 Let M be a differentiable manifold. For all $X, Y \in \mathrm{Vec}(M)$ and for all $f, g \in C^\infty(M)$

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

Exercise 3.4.49 Let X and Y be vector fields in a differentiable manifold M, and $p \in M$. Using the flows of the vector fields, define

$$\gamma(t) := \begin{cases} (\phi_{-Y}^{\sqrt{t}} \circ \phi_{-X}^{\sqrt{t}} \circ \phi_{Y}^{\sqrt{t}} \circ \phi_{X}^{\sqrt{t}})(p) & \text{for } t \geq 0, \\ (\phi_{Y}^{\sqrt{|t|}} \circ \phi_{-X}^{\sqrt{|t|}} \circ \phi_{-Y}^{\sqrt{|t|}} \circ \phi_{X}^{\sqrt{|t|}})(p) & \text{for } t < 0. \end{cases}$$

Then γ is a C^1 curve defined in a neighborhood of 0 and $\dot{\gamma}(0) = [X, Y]_p$.

Exercise 3.4.50 Let X be a (complete) vector field on a manifold M, with flow Φ_X^t. Then

$$Xf(\Phi_X^t(p)) = \frac{\mathrm{d}}{\mathrm{d}t} f(\Phi_X^t(p)), \qquad \forall f \in C^\infty(M), \forall t \in \mathbb{R}, \forall p \in M.$$

Chapter 4
General Theory of Carnot-Carathéodory Spaces

We have reached the point where we are ready to introduce the main object of our investigation: sub-Riemannian manifolds and, more generally, sub-Finsler 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 sub-Finsler 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 horizontal vector fields and their brackets generate all possible directions.

4.1 Definition of Carnot-Carathéodory Spaces

In this chapter, we denote by M a differentiable manifold, with its dimension primarily indicated as n. The tangent bundle of M, denoted as TM, 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 $\partial_{x_1}, \dots, \partial_{x_n}$. 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.

4.1.1 Bracket-Generating Distributions

Definition 4.1.1 (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

E. Le Donne, *Metric Lie Groups*, Graduate Texts in Mathematics 306,
https://doi.org/10.1007/978-3-031-98832-5_4

for every $\bar{p} \in M$ there exists smooth vector fields $X_1, \dots, X_m$ defined on some neighborhood U of $\bar{p}$ such that

$$\Delta_p := \Delta \cap T_p M = \operatorname{span}\{X_1(p), \dots, X_m(p)\}, \qquad \forall p \in U. \tag{4.1}$$

Distributions of tangent subspaces are also simply referred to as *distributions*. Furthermore, if 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*, and Δ is referred to as the *horizontal subbundle* of the polarized manifold.

Notice that each tangent subbundle is indeed a subbundle of the tangent bundle: A *subbundle* E of a vector bundle F (see Sect. 3.2.2) 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. In particular, a tangent subbundle of rank r on an n-manifold is a vector bundle with fibers of dimension r, and thus, it forms a manifold of 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 $f, g : \mathbb{R}^3 \to \mathbb{R}$ be smooth functions. Then, the two smooth vector fields

$$X_1(x, y, z) := \partial_x + f(x, y, z)\partial_z,$$
$$X_2(x, y, z) := \partial_y + g(x, y, z)\partial_z$$

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

$$\begin{aligned}\Delta_{(x,y,z)} &:= \{aX_1(x, y, z) + bX_2(x, y, z) : a, b \in \mathbb{R}\} \\ &= \{(a, b, af(x, y, z) + bg(x, y, z)) : a, b \in \mathbb{R}\}.\end{aligned}$$

Definition 4.1.2 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 by $\operatorname{Vec}(M)$ or $\Gamma(TM)$. In fact, 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 $\mathscr{F}_p := (\mathscr{F})_p := \{X_p : X \in \mathscr{F}\}$.

- 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 Sect. 3.2.1.

We specify that the set $\mathrm{Lie}(\mathscr{F})$ is the smallest subset of $\Gamma(TM)$ containing $\mathscr{F}$ and satisfying the property of being closed under both the Lie bracket and linear combinations, i.e.,

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

We are now prepared to introduce a criterion on a polarization Δ that allows us to connect points with curves tangent to Δ. The following condition (4.2) goes by many names, including *Hörmander's condition* or *Chow's condition*.

Definition 4.1.3 (Bracket Generating) A distribution Δ on a manifold M is *bracket generating* if

$$(\mathrm{Lie}(\Gamma(\Delta)))_p = T_pM, \qquad \forall p \in M. \tag{4.2}$$

Next, we clarify the meaning of a curve tangent to a distribution:

Definition 4.1.4 (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 3.3.1) 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 4.1.5 Special attention should be paid when verifying the condition (4.2) using a frame of vector fields. Indeed, let $X_1, \dots, X_m$ be vector fields spanning a distribution Δ on a manifold M, in the sense that (4.1) holds for all $p \in M$. On the one hand, if

$$(\mathrm{Lie}(\{X_1, \dots, X_m\}))_p = T_pM, \qquad \forall p \in M, \tag{4.3}$$

then Δ is bracket generating. On the other hand, the converse implication may not hold: For instance, consider a C^∞ function $\phi : \mathbb{R} \to \mathbb{R}$ such that $\phi(x) = 0$ if and only if $x = 0$, and $\frac{\mathrm{d}^k}{\mathrm{d}x^k}\phi(0) = 0$, for all $k \in \mathbb{N}$, as shown in Fig. 4.1 for an example. Consider on $\mathbb{R}^2$ with coordinates (x, y) the vector fields

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

Even though X, Y do not satisfy the bracket-generating condition (4.3), as demonstrated in Exercise 4.4.4, they define the same distribution as the bracket-generating frame $\partial_x, x\partial_y$; see Exercise 4.4.1.

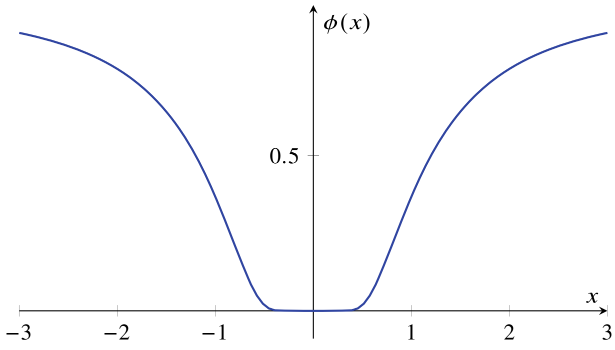


Fig. 4.1 Plot of the function $\phi(x) = \exp\left(-\frac{1}{x^2}\right)$, with $\phi(0) = 0$. This function is smooth everywhere and has all derivatives equal to zero at $x = 0$, but it is not identically zero. It is commonly used to define the 'Gaussian bump function'

4.1.2 Sub-Finsler Structures of Constant Rank

Definition 4.1.6 (Sub-Finsler and Sub-Riemannian Manifolds of Constant Rank) A *sub-Finsler manifold* is a triple $(M, \Delta, \|\cdot\|)$ where M is a connected manifold, $\|\cdot\|$ is a continuously varying norm (recall Definition 3.2.6), and Δ is a bracket-generating polarization on M, with the rank of Δ assumed to be constant. The pair $(\Delta, \|\cdot\|)$ is said to be a *sub-Finsler 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 a *sub-Riemannian manifold*.

We consider Riemannian and Finsler manifolds as particular cases of sub-Riemannian and sub-Finsler manifolds, respectively, which occur when Δ is the entire tangent bundle.

Since only the values of the restriction $\|\cdot\|_\Delta$ of $\|\cdot\|$ to Δ are important in what follows, we sometimes state that $(M, \Delta, \|\cdot\|_\Delta)$ is a sub-Finsler manifold with the sub-Finsler structure $(\Delta, \|\cdot\|_\Delta)$. Specifically, the length with respect to $\|\cdot\|$, as defined in (3.25), is considered only for curves that are horizontal with respect to Δ.

Definition 4.1.7 (CC-Distance) Given a sub-Finsler manifold $(M, \Delta, \|\cdot\|)$, the *Carnot-Carathéodory distance* between two points $p, q \in M$ is defined as:

$$d_{cc}(p,q) := \inf\left\{\text{Length}_{\|\cdot\|}(\gamma) : \gamma \text{ is a } \Delta\text{-horizontal curve from } p \text{ to } q\right\}. \quad (4.4)$$

If the infimum is achieved by a curve γ, that is, if $d_{cc}(p,q) = \text{Length}_{\|\cdot\|}(\gamma)$, then γ is *length minimizing* among the horizontal curves joining the two points p and q. The distance function d_{cc} is also called *Carnot-Carathéodory metric*.

For a sub-Finsler manifold $(M, \Delta, \|\cdot\|)$, we also consider the associated Finsler distance. If $d_F := d_{\|\cdot\|}$ denotes the Finsler distance associated with $(M, \|\cdot\|)$ as defined in (3.26), then it is evident that:

$$d_F(p, q) \le d_{\mathrm{cc}}(p, q), \qquad \forall p, q \in M, \tag{4.5}$$

because in the definition of d_{cc}, we infimize over a subset of the set used for d_F. It is important to note that the same length functional is used to define both distances.

We anticipate that the aforementioned function 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 4.1.8 (Chow; see Sect. 4.2.3) *If $(M, \Delta, \|\cdot\|)$ is a sub-Finsler manifold, then d_{cc} is finite and induces the manifold topology on M.*

Remark 4.1.9 (Terminology) The Carnot-Carathéodory distance is sometimes referred to as *CC-distance* or *sub-Finsler distance*. A sub-Finsler manifold equipped with its Carnot-Carathéodory distance is called *Carnot-Carathéodory space*. If $\|\cdot\|$ is the norm coming from a Riemannian metric, and hence $(M, \Delta, \|\cdot\|)$ is a sub-Riemannian manifold, then $(\Delta, \|\cdot\|)$ is called a *sub-Riemannian structure* and d_{cc} is called *sub-Riemannian distance*.

Some authors, ourselves included, refer to d_{cc} as a *Finsler-Carnot-Carathéodory distance*, or *FCC distance*, for short, to highlight that in the context d_{cc} might not necessarily be sub-Riemannian. Sub-Riemannian metrics have been discussed 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, though without a specific designation.

4.1.3 Control Theory Viewpoint

In control theory, the focus lies on systems of differential equations of the form:

$$\dot{\gamma} = \sum_{j=1}^{m} c_j(t) X_j(\gamma), \tag{4.6}$$

where $X_1, \ldots, X_m$ are predetermined vector fields on a manifold M, and $c_1, \ldots, c_m$ are variable L^1 functions defined on a bounded interval. These functions are called *control functions* or *controls*. Paths obtained by integrating (4.6) are termed *controlled paths*.

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

$$\Delta_p := \operatorname{span}_{\mathbb{R}} \{X_1(p), \dots, X_m(p)\}, \qquad \text{for } p \in M. \tag{4.7}$$

Conversely, every rank-m distribution Δ can, locally, be expressed as in (4.7). It is important to note that the adverb 'locally' is necessary due to global topological constraints, for instance, for the tangent bundle $\Delta = T(\mathbb{S}^2)$ of the 2D sphere $\mathbb{S}^2$.

However, in many systems of interest in control theory, the vector fields $X_1, \dots, X_m$ are not linearly independent at every point, and the distribution that they define does not have a constant rank. Nevertheless, a related distance can still be defined: for $p \in M$ and $v \in T_pM$, set

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

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

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

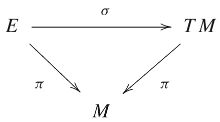
The *control distance associated with the system* $X_1, \dots, X_m$ is defined for every p and q in M as:

$$d(p,q) := \inf\left\{\int_0^1 g_p(\dot\gamma(t))^{1/2}\,\mathrm{d}t \;\middle|\; \gamma \text{ absolutely continuous path with } \gamma(0) = p, \gamma(1) = q\right\}. \tag{4.8}$$

4.1.4 The General Definition with Varying Rank

The Carnot-Carathéodory distance (4.4) and the control distance (4.8) fit into a broader context. Specifically, using the language of vector bundles, we can provide a more general definition. The point here is that the distribution on the manifold is obtained as the image of a bundle, and not only can the rank of the bundle be strictly larger than the distribution, but also the rank of the distribution can be different at different points of the manifold.

Definition 4.1.10 A *CC-bundle structure*, also called *possibly rank-varying sub-Finsler structure*, on a manifold M is a pair (σ, N) of functions $\sigma : E \to TM$ and $N : E \to \mathbb{R}$ where E is a vector bundle over M, N is a continuous function such that for all $p \in M$, the restriction of N to the fiber E_p is a symmetric norm (refer to Definition 3.2.6), and σ is a smooth map that is a morphism of vector bundles lifting the identity, i.e., the following diagram commutes:



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

For every such a CC-bundle structure (σ, N), we set

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

Analogously as before, the *sub-Finsler distance associated with the CC-bundle structure*, for every p and q in M, is defined as:

$$d(p,q) := \inf\left\{\int_0^1 \|\dot\gamma(t)\|\, \mathrm{d}t \,\middle|\, \gamma \text{ absolutely continuous path} \right.$$
$$\left. \text{with } \gamma(0) = p, \gamma(1) = q\right\}. \qquad (4.9)$$

One can verify that, in the case of the inclusion $\sigma : \Delta \hookrightarrow TM$ of a subbundle of the tangent bundle, one recovers the Carnot-Carathéodory distance (4.4). Similarly, for $E := M \times \mathbb{R}^m$ and $\sigma(p,u) := u_1X_1 + \cdots + u_mX_m$, one recovers the control distance (4.8); see also Exercise 4.4.25.

4.1.5 Equiregular Distributions

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

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

Then $\Delta^{[2]} := \bigcup_{p\in M} \Delta^{[2]}(p)$ is a subset of TM. In general, the subset $\Delta^{[2]}$ 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 4.1.11 (Non-Equiregular Distribution) In $\mathbb{R}^3$ the *Martinet distribution* is the subbundle $\Delta \subset T\mathbb{R}^3$ spanned by the vector fields:

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

By computing the Lie brackets, we find:

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

In this case, the subset $\Delta^{[2]}$ has fibers of varying dimension, since we have

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

Remark 4.1.12 If $X_1, \ldots, X_r$ is a frame for Δ, then the collection of vector fields

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

span $\Delta^{[2]}$ at every point. Indeed, if $X, Y \in \Gamma(\Delta)$, then $X = \sum_i a^i X_i$ and $Y = \sum_j b^j X_j$ for some smooth functions a^i, b^j. Thus, we have what we claimed:

$$[X, Y] = [a^i X_i, b^j X_j] \overset{\text{Ex.3.4.48}}{=} a^i b^j [X_i, X_j] + a^i (X_i b^j) X_j - b^j (X_j a^i) X_i.$$

Definition 4.1.13 ($\Delta^{[k]}$) Given a distribution $\Delta \subseteq TM$ on M, for each $k \in \mathbb{N}$ we shall define the subset $\Delta^{[k]} \subseteq TM$ by describing each of its fiber $\Delta^{[k]}(p) := \Delta^{[k]} \cap T_pM$ as p varies in M. The fiber $\Delta^{[k]}(p)$ is given by

$$\Delta^{[k]}(p) := \operatorname{span}\left\{[X_1, [X_2, \ldots, [X_{j-1}, X_j] \ldots]](p) : j \in \{1, \ldots, k\},\ X_1, \ldots, X_j \in \Gamma(\Delta)\right\}. \tag{4.10}$$

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

$$\Delta^{[k+1]}(p) = \Delta^{[k]}(p) + \operatorname{span}\{[X_1, [X_2, \ldots, [X_k, X_{k+1}] \ldots]](p) : X_1, \ldots, X_{k+1} \in \Gamma(\Delta)\}. \tag{4.11}$$

Definition 4.1.14 (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

$$q \longmapsto \dim \Delta^{[k]}(q) \tag{4.12}$$

is constant in a neighborhood of p.

Notice that the functions (4.12) is $\mathbb{N}$-valued. Hence, if it is locally constant, then it is constant on connected components.

Definition 4.1.15 (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 4.1.13, the *flag of subbundles* for Δ.

Remark 4.1.16 A distribution $\Delta \subset TM$ is equiregular if and only if, for all $k \in \mathbb{N}$, the set $\Delta^{[k]}$ is a subbundle.

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 4.1.17 (Equiregular Sub-Finsler Manifolds) A sub-Finsler manifold $(M, \Delta, \|\cdot\|)$ is called *equiregular* if Δ is equiregular.

4.2 Chow's Theorem and Existence of Geodesics

In this section, we explain how bracket-generating distributions allow the existence of horizontal curves connecting arbitrary points. Consequently, Carnot-Carathéodory distances on sub-Finsler manifolds are finite-valued, and do not alter the topology. Moreover, locally, sub-Finsler manifolds are geodesic spaces.

4.2.1 Local Transitivity and Sussmann's Orbit Theorem

In this section, we highlight the fact that, because in every sub-Finsler manifold, the distribution is assumed bracket generating, between every pair of points, there is at least a horizontal curve. The bracket-generating condition can be considered an infinitesimal form of transitivity. Chow's theorem states that this condition implies local transitivity:

Theorem 4.2.1 (Chow) *If a subbundle Δ of the tangent bundle of a manifold is bracket generating at some point p (i.e., (4.2) 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, nearby points in a sub-Finsler manifold can be joined by horizontal curves with small Finsler length. This is precisely what Theorem 4.1.8 asserts.

We first explain the validity of Theorem 4.2.1, taking for granted a theorem by Sussmann. We are omitting the proof of Sussmann's theorem, which is, in fact, the core of our first proof of Theorem 4.2.1, but it is well presented in [Bel96]. The reader can write a second complete proof of the above Theorem 4.2.1 by following the hints in Exercise 4.4.15. Later in the text, we will present a detailed proof of Theorem 4.1.8, a result of higher interest for us. Also, the simpler case of Carnot groups, discussed in Sect. 7.1.4, offers an elementary demonstration of Theorem 4.1.8.

Theorem 4.2.2 (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 by an absolutely continuous curve almost everywhere tangent to Δ. Then, the set Σ is an immersed submanifold of M.*

A first proof of Theorem 4.2.1, modulo Theorem 4.2.2 In the assumptions of Theorem 4.2.1, we employ Theorem 4.2.2. Given a vector field $X \in \Gamma(\Delta)$ and a point $q \in \Sigma$, the flow line $t \mapsto \Phi^t_X(q)$ is tangent to Δ and lies in Σ. Thus, the vector X_q is tangent to the submanifold Σ. Consequently, we have:

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

Since Σ is a submanifold, the family $\mathcal{F}$ is involutive on Σ, meaning $\mathrm{Lie}(\mathcal{F}) \subseteq \mathcal{F}$; see Exercise 3.4.47. Therefore, we deduce that $\mathrm{Lie}(\Gamma(\Delta)) \subseteq \mathcal{F}$. By the bracket-generating condition at p, we have

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

This implies $\dim M = \dim \Sigma$, and thus Σ is a neighborhood of p. □

4.2.2 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

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

Theorem 4.2.3 *Let M be a manifold of positive dimension, and $\mathscr{F}$ be a family of vector fields on 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 $\Phi^{<T}_{\mathscr{F}}(p)$ contains p in its interior.*

Proof Since $\dim(M) > 0$, 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-dimensional, the proof is concluded. If $\dim M > 1$, then there is $X_2 \in \mathscr{F}$ that is not tangent to M_1 (Otherwise, the family $\mathrm{Lie}(\mathscr{F})$ of vector fields would

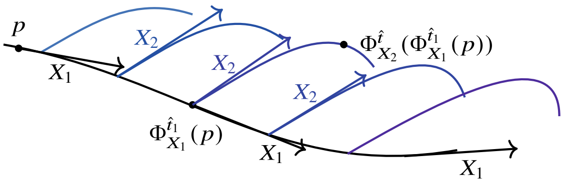


Fig. 4.2 Composition of two flows to construct a surface within the reachable set of a point p: first, flowing along the vector field X_1, followed by flowing along X_2, which is transverse to X_1

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 $t_1 < t_1 + \epsilon_2 < T$. We have obtained an embedded parametrized surface, as shown in Fig. 4.2.

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

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

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 exists 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 becomes an open subset of M. Take $X_1, \dots, X_k \in \mathscr{F}$ such that the previously defined points $F_k(t_1, \dots, t_k)$ cover a neighborhood of a point $q \in M$ when varying $t_j > 0$ with $\sum_j t_j < T$. Notice that if q is of the form $F_k(\bar{t}_1, \dots, \bar{t}_k)$, with $\bar{t}_j > 0$ and $\sum_j \bar{t}_j < T$, then the map

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

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 assumption. Therefore

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

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. □

With the use of Theorem 4.2.3, one can provide an alternative proof of Theorem 4.2.1; see Exercise 4.4.15.

4.2.3 The Metric Version of Chow's Theorem

We are now prepared to prove Theorem 4.1.8. Namely, we show that Carnot-Carathéodory distances induce the manifold topology.

Proof of Theorem 4.1.8 Let τ_M denote the manifold topology and τ_{CC} the topology induced by d_{cc}. We shall prove $\tau_{CC} = \tau_M$ by establishing the two containments.

Regarding the containment $\tau_{CC} \subset \tau_M$, let $U \in \tau_{CC}$ and $p \in U$. By definition of τ_{CC}, there exists $T > 0$ 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 Sect. 4.2.2, observe that

$$\Phi^{<T}_{\mathscr{F}}(p) \subset B_{d_{cc}}(p, T).$$

By Theorem 4.2.3, the point p is in the τ_M-interior of $\Phi^{<T}_{\mathscr{F}}(p)$. We deduce that p is also in the τ_M-interior of U.

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 that satisfies (4.5). Let $p \in U$. Then there is r such that $B_{d_F}(p, r) \subset U$ because Finsler distances induce the manifold topology. Since $d_F \leq d_{cc}$, then $B_{d_{cc}}(p, r) \subset B_{d_F}(p, r)$. Therefore p is also in the τ_{CC}-interior of U. □

4.2.4 Comparison of Length Structures

In certain situations, dealing with the case where Δ is a distribution with varying rank becomes more challenging. For instance, the following proposition remains valid when $(M, \Delta, \|\cdot\|)$ is a sub-Finsler manifold in the sense of Definition 4.1.10, as discussed in the work [ALN23]. However, it is crucial to note that altering Definition 4.1.6 by considering rank-varying distributions (instead of polarizations) with norms defined on the entire tangent bundle, as in Definition 4.1.6, may lead to the falsification of the conclusion of the following proposition. In fact, in that setting, there are examples of smooth curves parametrized by arc length that are nowhere tangent to the distribution, as illustrated in Example 4.4.3. Due to this issue, we confine our attention to sub-Finsler structures of constant rank, and we shall prove the following proposition exclusively in accordance with Definition 4.1.6.

Proposition 4.2.4 *Let $(M, \Delta, \|\cdot\|)$ be a (constant-rank) sub-Finsler manifold equipped with its Carnot-Carathéodory distance d_{cc}. Let $\gamma : [a, b] \to M$ be a curve.*

4.2.4.i. If $\mathrm{Length}_{d_{cc}}(\gamma) < \infty$, then the reparametrization by arc length of γ is Δ-horizontal.

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

Proof of 4.2.4.i Recall that in every metric space, every curve of finite length can be reparametrized by arc length; see Exercise 3.4.7. Hence, we consider γ to be parametrized by arc length with respect to the distance d_{cc}.

Let d_F be the Finsler distance for which we have (4.5). Recall that d_F is locally bi-Lipschitz equivalent to every Riemannian distance, as stated in Proposition 3.3.3. Since $d_F \le d_{cc}$, we have:

$$d_F(\gamma(s), \gamma(t)) \le d_{cc}(\gamma(s), \gamma(t)) \le \mathrm{Length}_{d_{cc}}(\gamma|_{[s,t]}) = |t - s|. \tag{4.13}$$

This implies that $\gamma : [a, b] \to (M, d_F)$ is 1-Lipschitz. Consequently, in local coordinates, the curve γ is (Euclidean) Lipschitz. By the 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 $t_0 = 0$, $\gamma(t_0) = 0 \in \mathbb{R}^n$, $\Delta_0 = \mathbb{R}^k \times \{0\}^{n-k}$, and $\dot\gamma(t_0) = e_n = (0, \dots, 0, 1)$. We then have:

$$\gamma_n(t) > t/2, \qquad \text{for small enough } t, \tag{4.14}$$

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:

$$p \in B_{d_F}(0, 2r_\epsilon),\ X \in \Delta_p,\ \|X\| \le 1 \implies |\langle \partial_n, X\rangle| < \epsilon, \tag{4.15}$$

where we use the Euclidean scalar product making $\partial_1, \dots, \partial_n$ orthonormal. Indeed, by contradiction, assume there exists $\epsilon > 0$ 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 (3.27) in some neighborhood of 0. Hence, eventually we have $\|X_j\|_{\mathbb{E}} \le c$. Therefore, being the sequence X_j in a compact set, it converges to some Y, up to a subsequence. Since Δ is assumed to be a polarization (hence a subbundle), it is a closed subset of TM; see Exercise 4.4.5. And since $p_j \to 0$, we have that $Y \in \Delta_0$. Thus we get

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

We inferred a contradiction, which gives the claim (4.15).

Let $\epsilon > 0$ and r_ϵ be chosen with the above property (4.15). 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 (4.14). 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 $b_\epsilon = \mathrm{Length}_{\|\cdot\|}(\sigma) \le$

$2d_{cc}(0, \gamma(r_\epsilon)) \leq 2r_\epsilon$, where in the last inequality we used (4.13). Hence, first we have

$$\frac{b_\epsilon}{r_\epsilon} \leq 2, \tag{4.16}$$

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 (4.15) 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:

$$0 < \frac{r_\epsilon}{2} \overset{(4.14)}{<} \gamma_n(r_\epsilon) = \sigma_n(b_\epsilon) = \int_0^{b_\epsilon} \dot{\sigma}_n(s)\, \mathrm{d}s \leq \int_0^{b_\epsilon} |\dot{\sigma}_n(s)|\, \mathrm{d}s \leq \epsilon b_\epsilon.$$

Thus we just obtained a bound that contradicts (4.16) 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. And the statement 4.2.4.i is proven. □

Proof of 4.2.4.ii 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 3.3.4, 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. □

With the previous Proposition 4.2.4, we deduce that in Carnot-Carathéodory spaces, the two a priori different length structures coincide. Hence, it is unambiguous what we mean when we refer to them as length spaces, a concept introduced in Sect. 3.1.3.

Corollary 4.2.5 *Carnot-Carathéodory spaces are length spaces.*

Proof Let $(M, \Delta, \|\cdot\|)$ be a sub-Finsler manifold with its Carnot-Carathéodory distance d_{cc}. We need to show that the distance d_{cc} is finite and equal to the infimum of the length of curves joining points, as in (3.6). Here, the length is calculated with respect to the distance d_{cc} itself. First, the distance is finite by Chow's Theorem 4.1.8. Second, Proposition 4.2.4 tells us that the curves with finite length with respect to d_{cc} are, up to reparametrizations, the Δ-horizontal curves.

Moreover, their length coincides with the integral of their velocity, as in (3.25). Therefore, we obtain (3.6) from the definition (4.4) for d_{cc}. □

4.2.5 Existence of Geodesics in CC Spaces

The Riemannian analog of the following theorem is due to Heinz Hopf (1894–1971) and Willi Rinow (1907–1979), who independently proved versions of it. The theorem guarantees the existence of geodesics between every two nearby points. Moreover, the two points can be chosen arbitrarily if the space is boundedly compact in the sense of Sect. 3.1.3.

Theorem 4.2.6 (Hopf-Rinow Theorem for CC Spaces) *Let M be a Carnot-Carathéodory space.*

1. *Every point in M has a neighborhood in which every two points can be joined by a curve that minimizes length with respect to the Carnot-Carathéodory distance.*
2. *If M is boundedly compact, then it is a geodesic space.*

Proof By Chow's Theorem 4.1.8, since M is connected and Δ is bracket generating, the distance function d_{cc} is finite and the topology of (M, d_{cc}) is locally compact. Moreover, the metric space (M, d_{cc}) is a length space, as stated in Corollary 4.2.5.

To find shortest paths, we shall use Proposition 3.1.4. This proposition ensures the existence of a shortest path between every two points that are sufficiently close. In fact, consider a point $p \in M$ and select $r > 0$ 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 by a length-minimizing curve. Indeed, by Proposition 3.1.4, 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 and p_2 can be connected to p via a curve of length strictly less than $r/2$, which therefore remains inside $\bar{B}(p, r)$. On the other hand, every competitor curve from p_1 to p_2 that leaves $\bar{B}(p, r)$ has a length of at least r because it starts in $B(p, r/2)$, exit $\bar{B}(p, r)$, and returns into $B(p, r/2)$. Therefore, the curve σ is a length-minimizing curve.

If, in addition, the metric space (M, d_{cc}) is boundedly compact, we can conclude by applying Proposition 3.1.6 □

4.3 Ball-Box Theorem and Hausdorff Dimension

Discussing the Hausdorff dimension is one of the most direct ways to demonstrate that Carnot-Carathéodory spaces are metrically equivalent to Riemannian spaces only when the polarization is the entire tangent bundle. This metric dimension can be calculated by observing that the metric balls, with respect to Carnot-

Carathéodory distances, do not necessarily behave like cubes where all the edges have comparable lengths; instead, they behave like boxes with edges of differing magnitudes. Such a statement is made precise by the so-called Ball-Box Theorem.

4.3.1 Ball-Box Theorem

Let $(M, \Delta, \|\cdot\|)$ be an equiregular sub-Finsler manifold of topological dimension n and step s. Consider the flag of subbundles:

$$\Delta = \Delta^{[1]} \subset \Delta^{[2]} \subset \cdots \subset \Delta^{[s]} = TM.$$

Since the upcoming considerations will be local in nature, we assume the existence of a frame $X_1, \ldots, X_n$ for TM and numbers $m_1, \ldots, m_s$ such that $X_1, \ldots, X_{m_k}$ form a frame for $\Delta^{[k]}$. In this case, we say that $X_1, \ldots, X_n$ is an *equiregular frame*. Equiregular frames are also known as *adapted frames*.

Notice that, for all $p \in M$,

$$m_j = \dim \Delta^{[j]}(p). \tag{4.17}$$

We also say that X_j has *degree* d_j if

$$X_j(p) \in \Delta^{[d_j]} \setminus \Delta^{[d_j-1]}, \qquad \forall p \in M, \tag{4.18}$$

i.e., $j \in \{m_{d_{j-1}} + 1, \ldots, m_{d_j}\}$. We might denote d_j by $\deg(X_j)$.

The plan is to parametrize the manifold M using the flow of linear sums of $X_1, \ldots, X_n$. To such vector fields, we associate an *exponential coordinate map* from a point $p \in M$ as

$$\Phi_p : \mathbb{R}^n \to M, \qquad (t_1, \ldots, t_n) \longmapsto \Phi^1_{t_1 X_1 + \cdots + t_n X_n}(p), \tag{4.19}$$

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 due to 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 $d_1, \ldots, d_n$ and radius $r > 0$ as

$$\mathsf{Box}(r) := \left\{ (t_1, \ldots, t_n) \in \mathbb{R}^n : |t_j| \leq r^{d_j} \right\}. \tag{4.20}$$

The following comparison theorem is due to several mathematicians, including Mitchell, Gershkovich, and Nagel-Stein-Wainger, as mentioned in [Gro99]. It is known as the Ball-Box Theorem because it establishes a comparison between the box $\text{Box}(r)$ in $\mathbb{R}^n$ and the ball $B(p, r')$ with respect to the Carnot-Carathéodory distance, providing a bi-Lipschitz relation between r and r'.

Theorem 4.3.1 (Ball-Box Theorem) *Let $(M, \Delta, \|\cdot\|)$ be a sub-Finsler manifold. Assume Δ is equiregular. Fix $\bar{p} \in M$ and an equiregular frame $X_1, \ldots, X_n$ in a neighborhood of $\bar{p}$ with degrees $d_1, \ldots, d_n$ and the corresponding boxes $\mathtt{Box}(\cdot)$. Then there exist a neighborhood U of $\bar{p}$ in M and constants $C > 1$ 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 proof of the Ball-Box Theorem in this general context of manifolds will not be presented here. Instead, it will be demonstrated later in the more straightforward scenario of Carnot groups and then, more generally, for sub-Finsler groups; see Theorems 11.2.3 and 12.5.3.

Remark 4.3.2 The Ball-Box Theorem 4.3.1 provides a quantitative version of Chow's theorems 4.2.1 and 4.1.8.

As of our current knowledge, there is no conclusive answer to the following natural question, except, perhaps, in the case of 2-step Carnot groups and contact 3-manifolds—further investigation is required [Bar00, BBL20].

Question 4.3.3 (Open!) Are all sufficiently small sub-Finsler balls and spheres homeomorphic to the usual Euclidean balls and spheres?

Here is a first very useful consequence of the Ball-Box Theorem 4.3.1.

Corollary 4.3.4 (Hölder Equivalence Between CC and Euclidean Metrics) *Locally, each equiregular sub-Finsler manifold is Hölder equivalent to a Riemannian manifold. Namely, if s is the step, then locally around every point, there exists $C > 1$ such that*

$$\frac{1}{C}(d_{cc})^s \leq d_{\text{Riem}} \leq C d_{cc}. \tag{4.21}$$

Proof Let $(M, \Delta, \|\cdot\|)$ be the sub-Finsler manifold. Let g be a Riemannian tensor with a norm smaller than $\|\cdot\|$ and denote by d_{Riem} the induced Riemannian distance. Recall that every other Riemannian distance is locally Lipschitz equivalent to d_{Riem}. Consider the identity map $\text{id} : M \to M$. Obviously, the map

$$\text{id} : (M, d_{cc}) \to (M, d_{\text{Riem}})$$

is 1-Lipschitz (and, therefore, Hölder); and hence we obtain the inequality on the right in (4.21).

For the other bound, we observe that the step s of Δ is equal to the maximum of the degree d_j of the vector fields of some equiregular frame $X_1, \ldots, X_n$, which we locally fix. For $r \in (0, 1)$, it is evident 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 4.3.1 and the fact that the exponential maps Φ_p are locally bi-Lipschitz maps (locally uniformly in p), as shown in Exercise 4.4.20, we obtain:

$$B_{d_{cc}}(p, Cr) \supseteq \Phi_p(\mathrm{Box}(r)) \supseteq \Phi_p(B_E(0, r^s)) \supseteq B_{d_{\mathrm{Riem}}}(p, C' r^s).$$

Hence, the map

$$\mathrm{id} : (M, d_{\mathrm{Riem}}) \to (M, d_{cc})$$

is $1/s$-Hölder on compact sets, establishing the inequality on the left of (4.21). □

4.3.2 Dimensions of CC Spaces

With the aid of the Ball-Box Theorem, we can finally calculate the Hausdorff dimensions of sub-Finsler manifolds.

Definition 4.3.5 (Homogeneous Dimension) If a distribution Δ on a manifold M is equiregular, we define its *homogeneous dimension* as the natural number

$$Q := Q_\Delta := \sum_{j=1}^{n} j \left(\dim \Delta^{[j]}(p) - \dim \Delta^{[j-1]}(p) \right), \tag{4.22}$$

which is independent of p as it varies in M.

In other words, in terms of the numbers $m_1, \ldots, m_s$ from (4.17), we express Q as:

$$Q = m_1 + 2(m_2 - m_1) + 3(m_3 - m_2) + \cdots + s(m_s - m_{s-1}). \tag{4.23}$$

It is important to note that the box defined in (4.20) satisfies

$$\mathcal{L}^n(\mathtt{Box}(r)) = r^Q,$$

where $\mathcal{L}^n$ is the Lebesgue measure in $\mathbb{R}^n$. In terms of the degrees of the vector fields of an adapted frame, as in (4.18), we also have

$$Q = \sum_{j=1}^{n} d_j. \tag{4.24}$$

Corollary 4.3.6 *If a sub-Finsler 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 bi-Lipschitz equivalent to each volume form. In particular, if $TM \neq \Delta$, the Hausdorff dimension is strictly greater than the topological dimension.*

Proof We arbitrarily choose a Riemannian structure. Since all volume forms are locally bi-Lipschitz equivalent, we assume that the volume form corresponds to the Riemannian volume form vol.

Using the notation of the Ball-Box Theorem 4.3.1, let k be the (locally uniform) bi-Lipschitz 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 4.4.20. 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 enough r,

$$\frac{1}{k^n}\mathcal{L}^n(\mathtt{Box}(r)) \leq \mathrm{vol}(\Phi_p(\mathtt{Box}(r))) \leq k^n\mathcal{L}^n(\mathtt{Box}(r)).$$

If Q is the homogeneous dimension, according to the Ball-Box theorem, for small r then have

$$\frac{1}{k^n C^Q} r^Q \leq \mathrm{vol}(B_{d_{cc}}(p, r)) \leq k^n C^Q r^Q.$$

By Theorem 3.1.18 and Remark 3.1.19, obtained in the section on Ahlfors regular measures, we can conclude. □

4.3.2.1 Dimensions of Submanifolds in CC Spaces

The question of computing the Hausdorff dimension and Hausdorff measure of submanifolds in sub-Finsler manifolds with respect to the Carnot-Carathéodory distance is a natural one. Gromov provided a general formula for the Hausdorff dimension of smooth submanifolds in equiregular Carnot-Carathéodory spaces in [Gro99, 0.6 B]. This formula was later demonstrated to coincide with the degree of the submanifold, introduced in [MV08], in [Mag10].

Theorem 4.3.7 ([Gro99, page 104]) *Let* $(M, \Delta, \|\cdot\|)$ *be a sub-Finsler manifold with an equiregular distribution* Δ *and Carnot-Carathéodory distance* d_{cc}*. Consider a smooth submanifold* $\Sigma \subset M$*. Then, the Hausdorff dimension of* (Σ, d_{cc}) *is*

$$\dim_H(\Sigma, d_{cc}) = \max\left\{\sum_{j=1}^{n} j \cdot \dim\left(T_p\Sigma \cap \Delta^{[j]}(p)\right) \Big/ \left(T_p\Sigma \cap \Delta^{[j-1]}(p)\right) \ : \ p \in \Sigma\right\}.$$

Nevertheless, questions concerning Hausdorff *measures* of smooth submanifolds remain unanswered. In [MV08], Magnani and Vittone derived an integral formula for the spherical Hausdorff measure of submanifolds in Carnot groups under a suitable 'negligibility condition'. This negligibility condition has been established in two-step groups, [Mag10] using covering arguments, and in the Engel group, using blow-up arguments [LM10]. However, the situation remains unclear in higher step groups and, more generally, in sub-Riemannian manifolds. For further details on this problem and its connections with existing literature, we direct the reader to Magnani's works [MV08, Mag08, Mag10, Mag19].

4.4 Exercises

Exercise 4.4.1 (Grushin Distribution) On $\mathbb{R}^2$ with coordinates (x, y), the vector fields

$$X = \partial_x \qquad \text{and} \qquad Y = x\partial_y, \tag{4.25}$$

satisfy the generating condition (4.3) and define a bracket-generating distribution whose rank is not constant.

Exercise 4.4.2 (Grushin Sub-Riemannian Plane) On $\mathbb{R}^2$ with coordinates (x, y), consider the vector fields X and Y as in (4.25). Define the Carnot-Carathéodory distance d_{cc} on $\mathbb{R}^2$ by setting X and Y orthonormal, as in Definition (4.9). For every $v \in \mathbb{R}$, $\lambda > 0$ and $p, q \in \mathbb{R}^2$, we have

$$d_{cc}(p + (0, v), q + (0, v)) = d_{cc}(p, q) \qquad \text{and} \qquad d_{cc}(\delta_\lambda p, \delta_\lambda q) = d_{cc}(p, q),$$

where $\delta_\lambda := \begin{pmatrix} \lambda & 0 \\ 0 & \lambda^2 \end{pmatrix}$, and

$$d_{cc}((0, y_1), (0, y_2)) = C\sqrt{|y_1 - y_2|}, \qquad \forall y_1, y_2 \in \mathbb{R},$$

where $C := d_{cc}((0,0),(0,1))$. In particular, the line $\{(0, y) : y \in \mathbb{R}\}$ is isometric to the 1/2-snowflake of the Euclidean line.

Exercise 4.4.3 On $\mathbb{R}^2$ with coordinates (x, y), consider the distribution Δ generated by the vector fields X and Y as in (4.25). Consider the continuously varying norm $\|\cdot\|$ given by the Euclidean norm for every tangent vector. Then, we have

(i) the Carnot-Carathéodory distance d_{cc} induced by $(\Delta, \|\cdot\|)$ is the Euclidean distance.
(ii) The curve $t \in \mathbb{R} \longmapsto (0, t)$ is parametrized by arc length, but at no point is it tangent to the distribution Δ.

Exercise 4.4.4 Let $\phi \in C^\infty(\mathbb{R})$ be such that $\frac{\mathrm{d}^k\phi}{\mathrm{d}x^k}(0) = 0$, for all $k \in \{0, 1, 2, \ldots\}$, and $\phi(x) \neq 0$, for all $x \neq 0$, as in Fig. 4.1. On $\mathbb{R}^2$ with coordinates (x, y), for the vector fields

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

we have

$$\mathrm{Lie}(\{X, Y\}) = \mathrm{span}_{\mathbb{R}} \left\{ \partial_x, \frac{\mathrm{d}^k\phi}{\mathrm{d}x^k}(x)\partial_y : k \in \{0, 1, 2, \ldots\} \right\}.$$

Consequently, the pair X and Y does not satisfy the generating condition (4.3) at $p = (0, 0)$. Still, the vector fields span the same distribution of Exercise 4.4.1

Exercise 4.4.5 Every subbundle of a vector bundle is a closed subset.

Exercise 4.4.6 Every Finsler distance on each manifold induces the manifold topology.

Exercise 4.4.7 Every two Finsler distances on a compact set are bi-Lipschitz equivalent.

Exercise 4.4.8 Carnot-Carathéodory distances, and in particular Riemannian and Finsler distances, are length distances.

Exercise 4.4.9 The Hausdorff dimension of a Riemannian n-manifold is n.

Exercise 4.4.10 If $\gamma : I \to (M, d_{cc})$ is a curve in a Carnot-Carathéodory space that is parametrized by arc length, then $\|\dot{\gamma}\| = 1$ a.e.

Exercise 4.4.11 If $\gamma : I \to (M, d_{cc})$ is a L-Lipschitz curve in a Carnot-Carathéodory space, then γ is a horizontal curve and $\|\dot{\gamma}\| \leq L$ a.e.

Exercise 4.4.12 Let $(M, \Delta, \|\cdot\|)$ be a sub-Finsler manifold. We denote by $\text{Length}_{d_{cc}}$ and $\text{Length}_{\|\cdot\|}$ the length with respect to the metric d_{cc} and the length with respect to the Finsler norm $\|\cdot\|$, respectively. Let γ be a horizontal curve. We have

$$\text{Length}_{\|\cdot\|}(\gamma) = \text{Length}_{d_{cc}}(\gamma).$$

Exercise 4.4.13 For every absolutely continuous curve γ in a sub-Finsler manifold, we have

$$\gamma \text{ is horizontal } \iff \text{Length}_{d_{cc}}(\gamma) < +\infty.$$

Exercise 4.4.14 Denote by $\Phi_{X_i}^{t_i}$ the flow at time i with respect to a vector field X_i. Calculate the differential of

$$(t_1, \dots, t_k) \longmapsto \Phi_{X_k}^{t_k} \circ \dots \circ \Phi_{X_1}^{t_1}(p).$$

Exercise 4.4.15 Theorem 4.2.3, together with the fact that the points where (4.2) holds form an open set, gives a (second) proof of Theorem 4.2.1.

Exercise 4.4.16 Recall that $\Gamma(\Delta)$ denotes the smooth sections of a subbundle Δ of a tangent bundle of a manifold M. The Hörmander's condition is equivalent to $\text{Lie}(\Gamma(\Delta)) = \text{Vec}(M)$.
Hint. Use Exercise 3.4.48.

Exercise 4.4.17 Let $\Delta^{[j]}(p)$ be the vector space defined in (4.10). The set $\Delta^{[j]}(p)$ can be equivalently 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 4.4.18 For the set $\Delta^{[j]}$ as defined in (4.10) we have the following properties:

(i) The set $\Delta^{[j]}$ might not be a subbundle of TM.
Hint. Try the distribution given by the frame $X_1 = \partial_1$, $X_2 = \partial_2 + x_1^2\partial_3$.
(ii) If $\Delta^{[j]}$ is a subbundle, then it makes sense to consider smooth sections $\Gamma(\Delta^{[j]})$ and

$$\Delta^{[j+1]}(p) = \Delta^{[j]}(p) + \mathbb{R}\text{- span}\left\{[X, Y](p) \ : \ X \in \Gamma(\Delta), Y \in \Gamma(\Delta^{[j]})\right\}.$$

Exercise 4.4.19 If $(M, \Delta, \|\cdot\|)$ is a sub-Finsler manifold with induced distance d_{cc}, then the metric space (M, d_{cc}) is homeomorphic to the manifold M via the identity map.

Exercise 4.4.20 The maps $\Phi_p : \mathbb{R}^n \to M$ from (4.19) are bi-Lipschitz in some neighborhood of 0 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-bi-Lipschitz homeomorphism between U equipped with the Euclidean distance and its image equipped with d_{Riem}.

Exercise 4.4.21 Each smooth surface in the Heisenberg group has Hausdorff dimension 3.
Hint. One may use Theorem 4.3.7 or give a direct proof considering vertical planes as preliminary cases.

Exercise 4.4.22 Give a proof of Theorem 4.3.7.

Exercise 4.4.23 Let $F : M_1 \to M_2$ be a smooth map between sub-Finsler manifolds $(M_1, \Delta^{M_1}, \|\cdot\|)$ and $(M_2, \Delta^{M_2}, \|\cdot\|)$. Let $L > 0$. Assume, for all $p \in M_1$, that $\mathrm{d}\pi(\Delta_p^{M_1}) \subseteq \Delta_{F(p)}^{M_2}$ and that

$$(\mathrm{d}\pi)_p : (\Delta_p, \|\cdot\|) \to (\Delta_{F(p)}, \|\cdot\|)$$

is L-Lipschitz. Then, the map $F : M_1 \to M_2$ is L-Lipschitz with respect to the respective sub-Finsler metrics.

Exercise 4.4.24 For every CC-bundle structure (σ, N) on a manifold M, as in Definition 4.1.10, given an absolutely continuous curve $\gamma : [0, 1] \to M$ with $\|\dot\gamma(t)\| < \infty$ for almost every $t \in [0, 1]$, then there exists a measurable map $u : [0, 1] \to E$ such that $\sigma(u(t)) = \dot\gamma(t)$ and $\|\dot\gamma(t)\| = N(u(t))$, for almost every $t \in [0, 1]$.

Exercise 4.4.25 As in Definition 4.1.10, let M be a smooth manifold and $f : M \times \mathbb{R}^m \to TM$ 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 $p \in M$. Then, the associated sub-Finsler distance (4.9) between p and q in M is

$$\begin{aligned} &d_{(f,N)}(p, q) \\ &:= \inf\left\{ \int_0^1 N(\gamma(s), u(s))\, \mathrm{d}s \;\middle|\; u \in L^\infty([0, 1]; \mathbb{R}^m),\ \begin{matrix} \gamma(0) = p \\ \gamma(1) = q \\ \dot\gamma(t) = f(\gamma(t), u(t)) \end{matrix} \right\}. \end{aligned} \tag{4.26}$$

Chapter 5
A Review of Lie Groups

In the following chapter, we will review the theory of Lie groups. This revision serves two purposes: First, Lie groups equipped with special sub-Finsler structures appear as tangent spaces of Carnot-Carathéodory spaces. Such Lie groups serve as infinitesimal models for sub-Riemannian manifolds, playing the same role as Euclidean vector spaces in Riemannian geometry. Second, sub-Finsler structures on Lie groups are highly interesting and arise in various contexts, including geometric group theory, harmonic analysis, hyperbolic geometry, and furthermore in stochastic processes and mechanics. They are, in a sense, easier to study than general Carnot-Carathéodory spaces.

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].

5.1 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* or *product law*, usually denoted by the symbol $\cdot$. The product is a function $(a, b) \in G \times G \longmapsto a \cdot b \in G$ that satisfies associativity, the existence of an identity element, and of an inversion map. The inversion map is denoted as $a \mapsto a^{-1}$. When it is clear, we may simply write ab for $a \cdot b$. The identity element of a group G is denoted by 1. If it is necessary to

E. Le Donne, *Metric Lie Groups*, Graduate Texts in Mathematics 306,
https://doi.org/10.1007/978-3-031-98832-5_5

emphasize that 1 is the identity element of the group G, it can be denoted by 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

$$\begin{aligned} L_g : \quad G &\longrightarrow G \\ h &\longmapsto gh. \end{aligned}$$

The *right translation* by g is the bijection

$$\begin{aligned} R_g : \quad G &\longrightarrow G \\ h &\longmapsto hg. \end{aligned}$$

The *conjugation* by g is the bijection

$$\begin{aligned} C_g : \quad G &\longrightarrow G \\ h &\longmapsto ghg^{-1}. \end{aligned}$$

We shall focus on Lie groups, which are differentiable manifolds with a smooth group operation. However, some of the remarks 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 form a special class of topological groups:

Definition 5.1.1 (Lie Group) A Lie group is a differentiable manifold (second countable, but not necessarily connected) together with a group structure such that both

$$\text{the product } \begin{aligned} G \times G &\to G \\ (x, y) &\mapsto x \cdot y \end{aligned} \qquad \text{and the inverse } \begin{aligned} G &\to G \\ g &\mapsto g^{-1} \end{aligned} \tag{5.1}$$

are C^∞ maps.

As in every manifold, the set $\mathrm{Vec}(G)$ of vector fields on G forms a Lie algebra. The general notion of Lie algebra is the following:

Definition 5.1.2 (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:

$[X, Y] = -[Y, X]$ (*anti-commutativity*),

$[[X, Y], Z] + [[Y, Z], X] + [[Z, X], Y] = 0$ (*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 than $\mathbb{R}$. However, in this text, we shall only consider those over the real numbers. The structure of a Lie algebra can be represented by expressing the

Lie bracket using a basis. Namely, if $\mathfrak{g}$ is a Lie algebra with bracket $[\cdot, \cdot]$ and $X_1, \ldots, X_n$ is an ordered basis of $\mathfrak{g}$ as vector space, then the *structural constants* of $\mathfrak{g}$ with respect to $X_1, \ldots, X_n$ are the real numbers $c_{ij}^k \in \mathbb{R}$ with $i, j, k \in \{1, \ldots, n\}$ such that

$$[X_i, X_j] = \sum_{k=1}^{n} c_{ij}^k X_k, \qquad \forall i, j \in \{1, \ldots, n\}. \tag{5.2}$$

The data $X_1, \ldots, X_n$ and $(c_{ij}^k)_{i,j,k \in \{1,\ldots,n\}}$ record the whole info about the Lie bracket; see Exercises 5.8.6, 5.8.7, and 5.8.8.

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 5.1.6.

The Lie algebra associated with a group G is isomorphic, as a vector space, to the tangent space $T_{1_G}G$ of 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 the space $\mathrm{Vec}(G)$ of smooth vector fields on G by suitably 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 $X \in \mathrm{Vec}(G)$ such that $(\mathrm{d}L_g)_h X_h = X_{L_g(h)}$ for all $g, h \in G$. Using (3.36) with $F = L_g$, it is easy to see that the class of left-invariant vector fields is closed under the Lie bracket; see Exercise 5.8.9. In other words, the set of left-invariant vector fields forms a Lie algebra.

Note that, after fixing a vector $v \in T_{1_G}G$, 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 $T_{1_G}G$, and proves that left-invariant vector fields form an n-dimensional subspace of $\mathrm{Vec}(G)$, where $n := \dim G$. We denote by $\mathfrak{g}$ the vector space $T_{1_G}G$ 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 summarize this definition:

Definition 5.1.3 (Lie Algebra of a Lie Group) Let G be a Lie group. The *Lie algebra* of G, denoted by $\mathrm{Lie}(G)$, has two realizations:

Interpretation 1: $\mathrm{Lie}(G)$ is the linear space $\mathrm{LIVF}(G)$ of left-invariant vector fields on G endowed with the bracket of vector fields.

Interpretation 2: $\mathrm{Lie}(G)$ 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 accustomed to placing function symbols to the left of variables, as in $x \mapsto f(x)$.

where $\tilde{X}, \tilde{Y}$ are the left-invariant vector fields such that $\tilde{X}_{1_G} = X$ and $\tilde{Y}_{1_G} = Y$, respectively. We shall use 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 a 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 5.8.28. A Lie subgroup $H < G$ is said to be a *closed Lie subgroup* if H is topologically closed within G. As a consequence, in this case, the inclusion $H \hookrightarrow G$ is an embedding; see Theorem 5.3.4. Closed Lie subgroups are also called *regular Lie subgroups*, especially if one wants to stress that it is an embedded submanifold (not just immersed).

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 it is an exercise to show that $\mathrm{Lie}(H)$ is canonically isomorphic to a subalgebra of $\mathrm{Lie}(G)$. Vice versa, every subalgebra comes from a Lie subgroup:

Theorem 5.1.4 (Existence of Subgroups; see Theorem 5.7.1) *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 $\mathfrak{h}$.*

A subspace $\mathfrak{i} \subseteq \mathfrak{g}$ is an *ideal* if $[\mathfrak{i}, \mathfrak{g}] \subseteq \mathfrak{i}$. A subgroup N of a group G is *normal*, if $C_g(N) = N$, for all $g \in G$. For ideals $\mathfrak{i} \subseteq \mathfrak{g}$ and normal subgroups $N \subseteq G$ we write $\mathfrak{i} \lhd \mathfrak{g}$ and $N \lhd G$, respectively. The two notions are connected: it is an exercise to show that the Lie algebra of a connected Lie subgroup is an ideal if and only if the subgroup is normal.

Next, we discuss the maps of the categories in which the objects are the Lie groups and the Lie algebras, respectively. A map $\varphi : G \to H$ between groups is a *group homomorphism*, or, simply, a *homomorphism* or a *morphism*, 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 called *Lie group homomorphism* if it is a smooth map. 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 a *Lie group automorphism*.

A map $\psi : \mathfrak{g} \to \mathfrak{h}$ between Lie algebras is called *Lie algebra homomorphism* if it is 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*).

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 element

$$\varphi_* := \mathrm{d}\varphi_{1_G} : T_{1_g} G \to T_{1_H} H \tag{5.3}$$

commutes with the Lie bracket operation; see Exercise 5.8.14. Namely, the map $\varphi_* : \mathrm{Lie}(G) \to \mathrm{Lie}(H)$ 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 is simply connected as topological space, then each Lie algebra homomorphism comes from a Lie group homomorphism:

Theorem 5.1.5 (Induced Lie Group Homomorphism; see Theorem 5.7.2) *Let G and H be Lie groups. Assume G is simply connected. For every 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 5.1.6 *If simply connected Lie groups G and H have isomorphic Lie algebras, then G and H are Lie group isomorphic.*

As a consequence of a theorem due to Ado, see [Jac79, page 199] and also Sect. 9.1.4, 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 5.1.7 *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, for stratified Lie algebras since the proof is much easier, and it is what is needed for the Lie groups of our interest: the Carnot groups. We refer to Sect. 9.2.4.

5.2 Exponential Map

Let M be a differentiable manifold. Consider a smooth vector field $X \in \mathrm{Vec}(M)$. 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* of X is defined as $\Phi^1_X(p) = \gamma(1)$, which gives the endpoint of the integral curve after one unit of time. It should be noted that generally, the exponential of X is defined only for X in some small neighborhood of zero in $\mathrm{Vec}(M)$ 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:

$$\exp\colon \mathfrak{g} \to G, \qquad X \mapsto \Phi_X^1(1_G).$$

Here, elements of the Lie algebra $\mathfrak{g}$ are identified with left-invariant vector fields, and thus we have $\mathfrak{g} \subset \mathrm{Vec}(G)$. It can be shown that for every $X \in \mathfrak{g}$, the ordinary differential equation $\dot{\gamma}(t) = X_{\gamma(t)}$ has a global solution and that this integral curve $\gamma(t)$ correspond to a Lie group homomorphism 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*.

5.2.1 One-Parameter Subgroups

Definition 5.2.1 (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 $\theta(\mathbb{R}) \subseteq G$ of one such 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 $\Phi_X^t(p)$ the *flow* of a vector field X at time t starting from a point p.

Proposition 5.2.2 *Let G be a Lie group and X be a left-invariant vector field on G.*
5.2.2.i. The flow line $t \longmapsto \Phi_X^t(1_G)$ of X from 1_G is a one-parameter subgroup.
5.2.2.ii. If $\theta : \mathbb{R} \to G$ is a one-parameter subgroup with $\dot{\theta}(0) = X_{1_G}$, then $\theta(t) = \Phi_X^t(1_G)$, for all $t \in \mathbb{R}$.

Proof of 5.2.2.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

$$\begin{aligned}
\dot{\gamma}(t) &= \frac{\mathrm{d}}{\mathrm{d}t}(\sigma(s) \cdot \sigma(t)) \\
&= \frac{\mathrm{d}}{\mathrm{d}t}(L_{\sigma(s)}(\sigma(t))) \\
&= (\mathrm{d}L_{\sigma(s)})_{\sigma(t)}\dot{\sigma}(t)
\end{aligned}$$

$$
\begin{aligned}
&= (\mathrm{d}L_{\sigma(s)})_{\sigma(t)} X_{\sigma(t)} \\
&= X_{\sigma(s)\cdot\sigma(t)} = X_{\gamma(t)}.
\end{aligned}
$$

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 $L_{\sigma(s)} \circ \gamma$, then σ is defined on the whole of $\mathbb{R}$. □

Proof of* 5.2.2*.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

$$
\begin{aligned}
\dot\theta(s) &= \left.\frac{\mathrm{d}}{\mathrm{d}t}\theta(s+t)\right|_{t=0} \\
&= \left.\frac{\mathrm{d}}{\mathrm{d}t}L_{\theta(s)}(\theta(t))\right|_{t=0} \\
&= \left(\mathrm{d}L_{\theta(s)}\right)_{\theta(0)} \dot\theta(0) \\
&= \left(\mathrm{d}L_{\theta(s)}\right)_1 X_1 = X_{\theta(s)}.
\end{aligned}
$$

So θ is the integral curve of X from 1_G. □

Remark 5.2.3 If θ is an OPS, then its image $\theta(\mathbb{R})$ 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.

5.2.2 Exponential Map

Definition 5.2.4 (Exponential Map) Let G be a Lie group and $\mathfrak{g}$ its Lie algebra, seen as the space of left-invariant vector fields. The *exponential map* is defined as

$$
\exp : \mathfrak{g} \to G, \quad X \in \mathfrak{g} \longmapsto \exp(X) := \Phi_X^1(1_G),
$$

i.e., $\exp(X)$ is the flow of X at time 1 starting from 1_G.

Remark 5.2.5 The exponential map may be different from the exponential map of Riemannian geometry. In Exercise 5.8.32, 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 admits Riemannian metrics that are invariant under left and right translations, and the Lie group exponential map coincides with the Riemannian exponential map of every one of these metrics; see Sect. 8.1.3.

One first key property of the exponential map is the following—see Exercise 5.8.17 for other properties.

Corollary 5.2.6 (of Proposition 5.2.2) *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.

Corollary 5.2.7 *Let X be a left-invariant vector field on a Lie group G. Then*

$$\Phi_X^t = R_{\exp(tX)}, \qquad \forall t \in \mathbb{R}. \tag{5.4}$$

Similarly, if we let $X^\dagger$ be the right-invariant vector field such that $(X^\dagger)_1 = X_1$, then we also have

$$\exp(tX) = \Phi_{X^\dagger}^t(1_G) \qquad \text{and} \qquad \Phi_{X^\dagger}^t = L_{\exp(tX)}, \quad \forall t \in \mathbb{R}. \tag{5.5}$$

From the fact that we explicitly know the above flows (5.4) and (5.5), we have many consequences; see Exercises 5.8.22, 5.8.23, and 5.8.24.

We summarize the following three interpretations of the exponential map:

$$\exp(X) = \begin{cases} \text{flow at time 1 of the left-invariant vector field } X, \\ \text{OPS at time 1 tangent to } X_{1_G} \text{ (or } X^\dagger_{1_G}\text{) at time 0,} \\ \text{flow at time 1 of the right-invariant vector field } X^\dagger. \end{cases}$$

The next result is a very important feature of the exponential map. It implies that exp gives a local parametrization of G near 1_G.

Proposition 5.2.8 *Let G be a Lie group with Lie algebra $\mathfrak{g}$. Then $\exp : \mathfrak{g} \to G$ is smooth and $(\mathrm{d}\exp)_0$ is the identity map:*

$$(\mathrm{d}\exp)_0 = \mathrm{id}_{\mathfrak{g}} : \mathfrak{g} = T_0\mathfrak{g} \to T_{1_G}G = \mathfrak{g}.$$

Consequently, the map $\exp$ is a diffeomorphism between some neighborhood of 0 in $\mathfrak{g}$ and some neighborhood of 1_G in G;

Proof For the smoothness of exp, we refer to Exercise 5.8.25. Regarding its differential, fix $X \in \mathfrak{g} = T_{1_G}G$. Let $\sigma : \mathbb{R} \to \mathfrak{g}$ be the curve $\sigma(t) := tX$ so that $\dot{\sigma}(0) = X$. Then

$$\begin{aligned}(\mathrm{d}\exp)_0(X) &= (\exp\circ\sigma)'(0) \\ &= \frac{\mathrm{d}}{\mathrm{d}t}\exp(tX)\bigg|_{t=0}\end{aligned}$$

$$
\begin{aligned}
&= \frac{\mathrm{d}}{\mathrm{d}t} \Phi^1_{t\tilde{X}}(1)\Big|_{t=0} \\
&= \frac{\mathrm{d}}{\mathrm{d}t} \Phi^t_{\tilde{X}}(1)\Big|_{t=0} \\
&= X,
\end{aligned}
$$

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:

Proposition 5.2.9 *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, as in* (5.3)*, then*

$$\exp \circ \varphi_* = \varphi \circ \exp,$$

i.e., the following diagram commutes:

$$
\begin{array}{ccc}
\mathrm{Lie}(G) & \xrightarrow{\varphi_*} & \mathrm{Lie}(H) \\
{\scriptstyle \exp}\downarrow & & \downarrow{\scriptstyle \exp} \\
G & \xrightarrow[\varphi]{} & H\,.
\end{array}
$$

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 every left-invariant vector field X and for every $t \in \mathbb{R}$

$$\sigma(t) := \varphi(\exp(tX)) = \exp(t\widetilde{(\mathrm{d}\varphi)_1 X_1}).$$

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:

$$
\begin{aligned}
\sigma(s)\sigma(t) &= \varphi(\exp(sX))\varphi(\exp(tX)) \\
&= \varphi\,(\exp(sX)\exp(tX)) \\
&= \varphi\,(\exp((s+t)X)) \\
&= \sigma(s+t),
\end{aligned}
$$

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

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t}\sigma(t)\bigg|_{t=0} &= \frac{\mathrm{d}}{\mathrm{d}t}\varphi(\exp(tX))\bigg|_{t=0} \\ &= (\mathrm{d}\varphi)_{\exp(0\cdot X)}\,\frac{\mathrm{d}}{\mathrm{d}t}\exp(tX)\bigg|_{t=0} \\ &= (\mathrm{d}\varphi)_1 X_1. \end{aligned}$$

□

5.2.3 Exponential Coordinates

We draw a first consequence of Proposition 5.2.8. Let $X_1, \ldots, X_n$ be a basis of the Lie algebra of a Lie group G. The map $\alpha : \mathbb{R}^n \to G$, defined as

$$(t_1, \ldots, t_n) \in \mathbb{R}^n \longmapsto \alpha(t_1, \ldots, t_n) := \exp(t_1 X_1 + \cdots + t_n X_n),$$

is a diffeomorphism between some neighborhood of $0 \in \mathbb{R}^n$ and some neighborhood of 1_G in G. Such a map is called the *exponential local coordinate map* (or *exponential local coordinates of the first kind*) with respect to $X_1, \ldots, X_n$.

The map $\beta : \mathbb{R}^n \to G$, defined as

$$(t_1, \ldots, t_n) \in \mathbb{R}^n \longmapsto \beta(t_1, \ldots, t_n) := \exp(t_1 X_1) \cdots \exp(t_n X_n),$$

is called *exponential local coordinates of the second kind* with respect to $X_1, \ldots, X_n$.

Intermediate examples can also be considered. For instance, given an integer $k \in \{1, \ldots, n-1\}$, one can define the map $\beta_k : \mathbb{R}^n \to G$ as

$$\beta_k(t_1, \ldots, t_n) := \exp(t_1 X_1 + \cdots + t_k X_k)\exp(t_{k+1} X_{k+1} + \cdots + t_n X_n),$$

which is called an *exponential local coordinates of mixed kind* with respect to $X_1, \ldots, X_n$.

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 an isomorphism. Indeed, for β we have

$$(\mathrm{d}\beta)_0(\partial_j|_0) \overset{\mathrm{def}}{=} \frac{\partial}{\partial t_j}\beta(t_1, \ldots, t_n)\bigg|_{(t_1,\ldots,t_n)=(0,\ldots,0)}$$

$$
\begin{aligned}
&= \frac{\mathrm{d}}{\mathrm{d}t_j} \beta(0, \ldots, 0, t_j, 0, \ldots, 0)\Big|_{t_j=0} \\
&= \frac{\mathrm{d}}{\mathrm{d}t_j} \exp(t_j X_j)\Big|_{t_j=0} \\
&= X_j.
\end{aligned}
$$

Warning: There are examples of groups for which α and β are not surjective.

5.3 From Continuity to Smoothness

5.3.1 Smoothness of Continuous Homomorphisms

This subsection aims to show that continuous homomorphisms between Lie groups are smooth. We begin with the 1-dimensional situation: the domain of the homomorphism is the additive group $\mathbb{R}$.

Theorem 5.3.1 *Let G be a Lie group. Every continuous group homomorphism $\theta : \mathbb{R} \to G$ is smooth.*

Proof Since $\theta(t+s) = \theta(t)\theta(s) = L_{\theta(t)}(\theta(s))$, it is enough to prove that θ is smooth at 0.

Let $U \subset \mathfrak{g}$ be an open set that is star-shaped with respect to 0 and such that $\exp|_U$ is a diffeomorphism between U and its image. Let

$$
U' := \tfrac{1}{2}U = \left\{ \tfrac{1}{2}X : X \in U \right\}.
$$

By continuity there is $t_1 > 0$ such that $\theta(t) \in \exp(U')$, for all $|t| \le t_1$. Let $X_1 \in U'$ be such that $\exp(X_1) = \theta(t_1)$. For $n \in \mathbb{N}$, let $X_n \in U'$ be such that $\exp(X_n) = \theta(t_1/n)$.

We claim that $nX_n \in U'$. Indeed, fix n and assume by induction that $jX_n \in U'$ for some $j < n$. On the one hand, we have $(j+1)X_n = \frac{j+1}{j} jX_n \in 2U' = U$, since $\frac{j+1}{j} < 2$. On the other hand, recalling Exercise 5.8.18, which we will use several times in this proof, we have

$$
\exp((j+1)X_n) = (\exp(X_n))^{j+1} = \left(\theta\left(\frac{t_1}{n}\right)\right)^{j+1} = \theta\left(\frac{j+1}{n} t_1\right),
$$

which is in $\exp(U')$ since $\frac{j+1}{n} \le 1$. Thus $(j+1)X_n \in U$ and $\exp((j+1)X_n) \in \exp(U')$. Since $\exp$ is injective on U, we get $(j+1)X_n \in U'$. By induction, we get the claim.

Now since

$$\exp(nX_n) = (\exp(X_n))^n = \left(\theta\left(\frac{t_1}{n}\right)\right)^n = \theta(t_1) = \exp(X_1)$$

and since exp is injective on U', we get $X_n = \frac{1}{n}X_1$.

Let $\frac{m}{n}$ be a rational number with $m \in \mathbb{Z}$ and $n > 0$. Then

$$\theta\left(\frac{m}{n}t_1\right) = \theta\left(\frac{t_1}{n}\right)^m = \exp(X_n)^m = \exp\left(\frac{X_1}{n}\right)^m = \exp\left(\frac{m}{n}X_1\right).$$

By continuity of θ (and density of rational numbers), we get

$$\theta(st_1) = \exp(sX_1), \qquad \forall s \in \mathbb{R}.$$

Hence, the homomorphism θ is the exponential curve

$$\theta(t) = \exp\left(\frac{t}{t_1}X_1\right),$$

and thus, it is smooth. □

Theorem 5.3.2 *Let G and H be Lie groups. Every continuous group homomorphism $\varphi : G \to H$ is smooth.*

Proof Consider $\beta : \mathbb{R}^n \to G$ the exponential local coordinates of the second kind with respect to a basis $X_1, \ldots, X_n$, i.e., $\beta(t_1, \ldots, t_n) := \exp(t_1 X_1) \cdots \exp(t_n X_n)$. For each $j \in \{1, \ldots, n\}$, the map $t \in \mathbb{R} \mapsto \varphi(\exp(tX_j))$ is a continuous homomorphism, so by the previous theorem (Theorem 5.3.1), it is smooth. Hence, the map $\varphi \circ \beta$ is smooth, being the product of smooth maps:

$$\begin{aligned}\varphi \circ \beta(t_1, \ldots, t_n) &= \varphi(\exp(t_1 X_1) \cdot \cdots \cdot \exp(t_n X_n)) \\ &= \varphi(\exp(t_1 X_1)) \cdot \cdots \cdot \varphi(\exp(t_n X_n)).\end{aligned}$$

Hence, in a neighborhood of 1_G the map $\varphi = (\varphi \circ \beta) \circ \beta^{-1}$ is smooth. Let $g \in G$ be an arbitrary point. Since $\varphi = L_{\varphi(g)} \circ \varphi \circ L_{g^{-1}}$ and φ is smooth at 1_G, then φ is smooth at g. □

Corollary 5.3.3 (Uniqueness of Lie Structures) *Each Lie group has only one differentiable structure of a Lie group with the same topology and group structure.*

Proof If the identity map between the group with two differentiable structures is a continuous homomorphism, then it is a diffeomorphism by Theorem 5.3.2. □

5.3.2 Closed Subgroups of Lie Groups

Theorem 5.3.4 (Closed Subgroups are Regular) *Let G be a Lie group and $H < G$ a subgroup. If H is a closed subset of G, then H is a regular Lie subgroup of G.*

Before the proof of the above theorem, we present a preparatory lemma that will be used twice in the proof of the theorem.

Lemma 5.3.5 *Let $H < G$ be closed. Let $X \in \mathrm{Lie}(G)$ and for each $j \in \mathbb{N}$ let $X_j \in \mathrm{Lie}(G)$ and $t_j > 0$. Assume $X_j \to X$ and $t_j \to 0$, as $j \to \infty$. If*

$$\exp(t_j X_j) \in H, \qquad \forall j \in \mathbb{N},$$

then

$$\exp(tX) \in H, \qquad \forall t \in \mathbb{R}.$$

Proof Fix $t \in \mathbb{R}$. For each $j \in \mathbb{N}$, take $m_j \in \left[\frac{t}{t_j} - 1, \frac{t}{t_j}\right] \cap \mathbb{Z}$. Hence, we have $m_j t_j \to t$ and $m_j t_j X_j \to tX$, as $j \to \infty$. Since H is closed,

$$\exp(tX) = \lim_{j\to\infty} \exp(m_j t_j X_j) = \lim_{j\to\infty} (\exp(t_j X_j))^{m_j} \in H,$$

where we used Exercise 5.8.18. □

Proof of Theorem 5.3.4 We need to show that H is an embedded submanifold. Let $\mathfrak{g} = T_{1_G} G$. Set

$$\mathfrak{h} := \{X \in \mathfrak{g} : \exists \sigma : \mathbb{R} \to H \text{ such that } \dot\sigma(0) = X\}.$$

The subset $\mathfrak{h}$ is a vector space: for all $X_1, X_2 \in \mathfrak{h}$ and all $\lambda_1, \lambda_2 \in \mathbb{R}$ we have $\lambda_1 X_1 + \lambda_2 X_2 \in \mathfrak{h}$, since if we have $\sigma_j : \mathbb{R} \to H$ with $\dot\sigma_j(0) = X_j$ for $j \in \{1, 2\}$, then $\sigma_1(\lambda_1 t) \cdot \sigma_2(\lambda_2 t) \in H$ and, recalling formula (5.27) of Exercise 5.8.11, we infer

$$\left.\frac{\mathrm{d}}{\mathrm{d}t}(\sigma_1(\lambda_1 t) \cdot \sigma_2(\lambda_2 t))\right|_{t=0} = \lambda_1 X_1 + \lambda_2 X_2.$$

We claim now that

$$\mathfrak{h} = \{X \in \mathfrak{g} : \exp(tX) \in H \ \forall t \in \mathbb{R}\}. \tag{5.6}$$

Indeed, it is obvious that $\mathfrak{h}$ contains the right-hand side. For the opposite inclusion, we shall use Lemma 5.3.5. If $X \in \mathfrak{h}$, i.e., $X = \dot\sigma(0)$ for some $\sigma : \mathbb{R} \to H$, we set $\tau(t) := \exp^{-1}(\sigma(t)) \in \mathfrak{g}$, which is defined when t is sufficiently small (recall Proposition 5.2.8). Therefore,

$$X = \dot{\sigma}(0) = \frac{\mathrm{d}}{\mathrm{d}t}\sigma(t)\bigg|_{t=0} = \frac{\mathrm{d}}{\mathrm{d}t}\exp(\tau(t))\bigg|_{t=0}$$
$$= (\mathrm{d}\exp)_{\tau(0)}\frac{\mathrm{d}}{\mathrm{d}t}\tau(t)|_{t=0} = \dot{\tau}(0) = \lim_{j\to\infty} j\tau\left(\frac{1}{j}\right).$$

Set $t_j := \frac{1}{j}$ and $X_j := j\tau(\frac{1}{j})$, so that $\exp(t_jX_j) = \exp(\tau(\frac{1}{j})) = \sigma(\frac{1}{j}) \in H$. By Lemma 5.3.5, we conclude that $\exp(tX) \in H$ for all $t \in \mathbb{R}$, i.e., the missing inclusion of the claim (5.6) is proved.

The idea is now to use the exponential map to obtain a coordinate map in a neighborhood of 1_G in H and then use left translations to get an atlas for H. Let $V \subset \mathfrak{g}$ be a vector subspace such that $\mathfrak{g} = \mathfrak{h} \oplus V$. We plan to show

$$\exists \Omega \subset V \text{ neighborhood of 0 in } V \text{ such that } \exp(\Omega) \cap H = \{1_G\}. \tag{5.7}$$

If this is not the case, there are $(Y_j)_{j\in\mathbb{N}} \subset V \setminus \{0\}$ with $Y_j \to 0$ and $\exp(Y_j) \in H$ for all $j \in \mathbb{N}$. Let $\|\cdot\|$ be any norm on $\mathfrak{g}$. Set $t_j := \|Y_j\|$ and $X_j := \frac{1}{t_j}Y_j$. So X_j are unit vectors in V and, up to passing to a subsequence, we have $X_j \to X$ for some $X \in V$. Notice that $t_j \to 0$ and $\exp(t_jX_j) = \exp(Y_j) \in H$ for all $j \in \mathbb{N}$. By Lemma 5.3.5 we get $\exp(tX) \in H$ for all $t \in \mathbb{R}$. So $X \in \mathfrak{h}$. Since $\|X\| = 1$, we get a contradiction:

$$0 \neq X \in V \cap \mathfrak{h} = \{0\}.$$

Hence, equation (5.7) is proven.

Let $\psi : \mathfrak{h} \times V \to G$ be defined by

$$\psi(X, Y) := \exp(X) \cdot \exp(Y), \qquad \forall X \in \mathfrak{h}, \forall Y \in V.$$

Then, the map ψ is a diffeomorphism between some neighborhood $\Omega_1 \times \Omega_2$ of $(0, 0)$ in $\mathfrak{h} \times V$ and some neighborhood Ω_3 of 1_G in G (since ψ is an exponential local coordinate map of a mixed kind, c.f. Sect. 5.2.3). We may assume that $\Omega = \Omega_2$, i.e., $\exp(\Omega_2) \cap H = \{1_G\}$.

We plan to show that

$$\Omega_3 \cap H = \exp(\Omega_1). \tag{5.8}$$

We have $\exp(\Omega_1) \subset \Omega_3$ by construction. Also $\exp(\Omega_1) \subset H$ follows from $\Omega_1 \subset \mathfrak{h}$ and the claim (5.6). Hence $\exp(\Omega_1) \subset \Omega_3 \cap H$.

Vice versa, let $h \in \Omega_3 \cap H$, then there are unique $X \in \Omega_1$ and $Y \in \Omega_2$ such that $h = \exp(X) \cdot \exp(Y)$. Thus $\exp(Y) = \exp(-X) \cdot h \in \exp(\Omega_2) \cap H = \{1_G\}$. Therefore $Y = 0$ and $h = \exp(X) \in \exp(\Omega_1)$.

Then $\varphi := (\psi|_{\Omega_1 \times \Omega_2})^{-1} : \Omega_3 \to \Omega_1 \times \Omega_2$ is a coordinate map for G that is centered at 1_G and is adapted to H, i.e., $\varphi|_H$ is a coordinate map for H into the vector space $\mathfrak{h} \subset \mathfrak{g}$.

To conclude, we consider the atlas $\{(L_h(\Omega_3), \varphi \circ L_{h^{-1}})\}_{h \in H}$. □

5.4 General Linear Groups, Their Lie Algebra, and Their Exponential Map

The General Linear Group of degree n, denoted as $\mathrm{GL}(n, \mathbb{K})$, consists of invertible $n \times n$ matrices over a given field $\mathbb{K}$. Its associated Lie algebra, denoted as $\mathfrak{gl}(n, \mathbb{K})$, 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 real vector spaces that are not explicitly identified with $\mathbb{R}^n$. Consequently, we consider general linear groups for vector spaces, i.e., sets of linear automorphisms. This abstraction enables us to consider linear structures like $\mathfrak{gl}(n, \mathbb{R})$ 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.

5.4.1 $\mathrm{GL}(V)$ *and* $\mathfrak{gl}(V)$

For $n \in \mathbb{N}$, we denote by $\mathrm{Mat}_{n \times n}(\mathbb{R})$ the space of $n \times n$ matrices with real entries.

The n-th *general linear group* is

$$\mathrm{GL}(n) := \mathrm{GL}(n, \mathbb{R}) := \{A \in \mathrm{Mat}_{n \times n}(\mathbb{R}) \mid \det A \neq 0\}.$$

This forms a group when equipped with the row-column matrix multiplication. Slightly more generally, if V is a (real) vector space, then

$$\mathrm{GL}(V) := \mathrm{Aut}(V) := \{A : V \to V \mid A \text{ is an invertible linear transformation}\}.$$

This is a group when equipped with the composition rule where the identity element is the identity transformation $\mathbb{I} : V \to V$. Noting that this product rule and the inversion rule are smooth, we infer that $\mathrm{GL}(n, \mathbb{R})$ 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 $n \in \mathbb{N}$, we define

$$\mathfrak{gl}(n) := \mathfrak{gl}(n, \mathbb{R}) := \mathrm{Mat}_{n\times n}(\mathbb{R}) := \{n \times n \text{ matrices with real entries}\}.$$

If V is a vector space, then

$$\mathfrak{gl}(V) := \mathrm{End}(V) := \{\text{ linear transformations from } V \text{ to } V\}.$$

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 consider the *commutator*:

$$[A, B] := AB - BA. \tag{5.9}$$

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 5.4.4.

5.4.2 Matrix Exponential

In this subsection, we recall the matrix exponential: the exponential of matrices. Since we shall consider linear endomorphisms of vector spaces, like, for example, the Lie algebra of a Lie group, we define the matrix exponential on the space $\mathfrak{gl}(V)$.

Definition 5.4.1 (Matrix Exponential) Let V be a finite-dimensional vector space. For each $A \in \mathfrak{gl}(V)$, define *the matrix exponential of* A as

$$e^A := \mathbb{I} + A + \frac{1}{2}A^2 + \frac{1}{3!}A^3 + \cdots = \sum_{k=0}^{\infty} \frac{1}{k!}A^k. \tag{5.10}$$

In fact, the series giving e^A is absolutely converging; see Exercise 5.8.41. Consequently, the function $A \mapsto e^A$ is smooth (in fact, analytic). Moreover, each e^A is invertible with inverse e^{-A}; see Exercises 5.8.42 and 5.8.43, so we have a map $A \in \mathfrak{gl}(V) \mapsto e^A \in \mathrm{GL}(V)$. In the following discussion, we use that $\mathrm{GL}(V) \subseteq \mathrm{End}(V)$ and that $\mathrm{End}(V)$ is a vector space. Therefore, tangent vectors of curves into $\mathrm{GL}(V)$ are represented by elements in $\mathrm{End}(V)$.

It is easy to see (and a proof is in Proposition 5.4.3) that for every linear map $A : V \to V$, the curve $t \mapsto e^{tA}$ satisfies

$$e^{tA}\Big|_{t=0} = \mathbb{I}, \qquad \frac{\mathrm{d}}{\mathrm{d}t}e^{tA}\Big|_{t=0} = A.$$

Moreover, the map ϕ^t that to every linear map $B : V \to V$ associates 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 $(Be^{sA})e^{tA} = Be^{(t+s)A}$, for every $A, B \in \mathfrak{gl}(V)$;
- ϕ^t is left-invariant, i.e., $\phi^t(MB) = M\phi^t(B)$ because $(MB)e^{tA} = M(Be^{tA})$, for every $A, B, M \in \mathfrak{gl}(V)$.

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 summarize, see also Proposition 5.4.3, the basic properties of the matrix exponential:

Proposition 5.4.2 (Matrix Exponential) *Let V be a finite-dimensional vector space.*

5.4.2.i. *The matrix exponential*

$$\exp : \mathfrak{gl}(V) \to \mathrm{GL}(V)$$
$$A \mapsto e^A$$

is an analytic map.

5.4.2.ii. *For every $A \in \mathfrak{gl}(V)$, the curve $t \mapsto e^{tA}$ is a one-parameter subgroup.*

5.4.2.iii. *For every $A \in \mathfrak{gl}(V)$, the map*

$$GL(V) \to T\,(GL(V))$$
$$B \mapsto BA,$$

defines a left-invariant vector field on $\mathrm{GL}(V)$ whose flow $\mathbb{R} \times \mathrm{GL}(V) \to \mathrm{GL}(V)$ is defined by $(t, B) \mapsto Be^{tA}$.

Rephrasing for the case when $V = \mathbb{R}^n$, for all $A \in \mathfrak{gl}(n, \mathbb{R}) \simeq T_{\mathbb{I}}GL(n, \mathbb{R})$, the unique left-invariant vector field on $\mathrm{GL}(n, \mathbb{R})$ that equals A at $\mathbb{I}$ is

$$B \in \mathrm{GL}(n, \mathbb{R}) \longmapsto BA \in \mathrm{Mat}_{n\times n}(\mathbb{R}).$$

In the next proposition, we spell out the argument that computes the derivative of the OPS $t \mapsto e^{tA}$. We shall refer to this proposition several times.

Proposition 5.4.3 (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 the one-parameter subgroup of $\mathrm{GL}(V)$ such that*

$$\frac{\mathrm{d}}{\mathrm{d}t} e^{tA} = Ae^{tA} \qquad \textit{and} \qquad \left.\frac{\mathrm{d}}{\mathrm{d}t} e^{tA}\right|_{t=0} = A.$$

Proof Recall 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

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}(e^{tA}) &= \frac{\mathrm{d}}{\mathrm{d}t}\left(\sum_{k=0}^{\infty} \frac{1}{k!}(tA)^k\right) \\
&= \sum_{k=0}^{\infty} \frac{1}{k!}\frac{\mathrm{d}}{\mathrm{d}t}(t^k A^k) \\
&= \sum_{k=1}^{\infty} \frac{1}{k!} k t^{k-1} A^k \\
&= A \sum_{k=1}^{\infty} \frac{1}{(k-1)!} t^{k-1} A^{k-1} \\
&= Ae^{tA},
\end{aligned}$$

where we could exchange the derivative and the summation because the series is absolutely convergent. □

5.4.3 Lie Algebras of General Linear Groups

This subsection aims to show that $\mathfrak{gl}(n, \mathbb{R})$, when equipped with (5.9), is the Lie algebra of the Lie group $\mathrm{GL}(n, \mathbb{R})$.

Proposition 5.4.4 *The Lie algebra of $\mathrm{GL}(V)$ is isomorphic to the Lie algebra $\mathfrak{gl}(V)$.*

Proof Recall from Proposition 5.4.2.iii that every element in $\mathfrak{gl}(V)$ induces a left-invariant vector field on $\mathrm{GL}(V)$ for which we have a formula for the flow. The key point of the proof is to show that for every $A, B \in \mathfrak{gl}(V)$, the vector field $M \in \mathrm{GL}(V) \mapsto M(AB - BA)$ is the Lie bracket between

$$M \mapsto MA \qquad \text{and} \qquad M \mapsto MB.$$

Thus, in terms of flows, by recalling Definition 3.2.2.d, we need to show that

$$\frac{\mathrm{d}}{\mathrm{d}t}\phi^t_{AB-BA}(M)\bigg|_{t=0} = \frac{\mathrm{d}}{\mathrm{d}t}(\phi_B^{-\sqrt{t}}\circ\phi_A^{-\sqrt{t}}\circ\phi_B^{\sqrt{t}}\circ\phi_A^{\sqrt{t}})(M)\bigg|_{t=0}. \tag{5.11}$$

On the one hand, the left-hand side is

$$\text{LHS of (5.11)} = \frac{\mathrm{d}}{\mathrm{d}t}Me^{t(AB-BA)}\bigg|_{t=0} = M(AB-BA).$$

On the other hand, recalling that $e^{\sqrt{t}A} = \mathbb{I} + \sqrt{t}A + \frac{tA^2}{2} + o(t)$, the right-hand side becomes

$$\begin{aligned}
\text{RHS of (5.11)} &= \frac{\mathrm{d}}{\mathrm{d}t}M(e^{\sqrt{t}A}e^{\sqrt{t}B}e^{-\sqrt{t}A}e^{-\sqrt{t}B})\bigg|_{t=0}\\
&= \frac{\mathrm{d}}{\mathrm{d}t}M(e^{\sqrt{t}A}e^{\sqrt{t}B}e^{-\sqrt{t}A}e^{-\sqrt{t}B})\bigg|_{t=0}\\
= \frac{\mathrm{d}}{\mathrm{d}t}M\bigg(\mathbb{I}+\sqrt{t}(A+B-A-B)\\
+t\left(\frac{A^2}{2}+\frac{B^2}{2}+\frac{A^2}{2}+\frac{B^2}{2}+AB-A^2-AB-BA-B^2+AB\right)+o(t)\bigg)\bigg|_{t=0}\\
&= M(AB-BA).
\end{aligned}$$

Hence, Eq. (5.11) holds, as desired. □

Proposition 5.4.3, together with Proposition 5.4.4, 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 5.4.5 (of Proposition 5.4.3) *For every finite-dimensional vector space V, the exponential map of the Lie group $\mathrm{GL}(V)$ is the matrix exponential $\exp : \mathfrak{gl}(V) \to \mathrm{GL}(V)$, $A \mapsto e^A$.*

5.5 Adjoint Representation

The adjoint representation, also known as the adjoint action, of a Lie group G provides a means of representing, possibly not injectively, 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 the group's action on itself through conjugation. Namely, we take the differentials. This natural representation captures the way the elements of the Lie group act on its Lie algebra, establishing a link between the group's abstract structure and the associated linear transformations.

5.5.1 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 Sect. 5.4.1.

Definition 5.5.1 (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 5.5.2 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, seeing $\mathfrak{gl}(\mathfrak{g})$ as a Lie algebra, the map $\mathrm{ad} : \mathfrak{g} \to \mathfrak{gl}(\mathfrak{g})$ is a Lie algebra homomorphism: for all $X, Y \in \mathfrak{g}$ and all $s, t \in \mathbb{R}$ we have

5.5.2.i. $\mathrm{ad}(sX + tY) = s\ \mathrm{ad}(X) + t\ \mathrm{ad}(Y)$,
5.5.2.ii. $\mathrm{ad}([X, Y]) = [\mathrm{ad}(X), \mathrm{ad}(Y)]$,

see Exercise 5.8.33 for the proof.

Definition 5.5.3 (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., $\mathrm{Ad}(g) : \mathfrak{g} \to \mathfrak{g}$ is the differential at the identity element of the conjugation $C_g : h \mapsto ghg^{-1}$. The map

$$\mathrm{Ad} : G \to \mathrm{GL}(\mathfrak{g})$$

is called *adjoint representation*.

Remark 5.5.4 The map Ad is indeed a representation, i.e., Ad is a group homomorphism into $\mathrm{GL}(\mathfrak{g})$:

$$\mathrm{Ad}(gh) = \mathrm{Ad}(g) \circ \mathrm{Ad}(h), \qquad \forall g, h \in G, \tag{5.12}$$

see Exercise 5.8.34 for the proof.

5.5.2 Properties and Formulas

Proposition 5.5.5 *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 with* Ad *is the adjoint map* $\mathrm{ad} : \mathfrak{g} \to \mathfrak{gl}(\mathfrak{g})$*, i.e.,*

$$(\mathrm{Ad})_* = \mathrm{ad}, \tag{5.13}$$

i.e., we have $(\mathrm{d}\,\mathrm{Ad})_{1_G}(X) = \mathrm{ad}(X)$, *for all* $X \in \mathfrak{g}$.

Proof Since $t \mapsto \exp(tX)$ is a curve in G that is tangent to X at 1_G, we have

$$(\mathrm{d}\,\mathrm{Ad})_{1_G}(X)(Y) = \frac{\mathrm{d}}{\mathrm{d}t} \mathrm{Ad}\,(\exp(tX))\,(Y)\Big|_{t=0}.$$

Here, we consider X, Y as elements 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

$$\begin{aligned}
\mathrm{Ad}(\exp(tX))(Y) &= \left(\mathrm{d}C_{\exp(tX)}\right)_{1_G}(Y) \\
&= \left(\mathrm{d}R_{\exp(-tX)}\right)_{\exp(tX)} (\mathrm{d}L_{\exp(tX)})_{1_G}(Y) \\
&= \left(\mathrm{d}R_{\exp(-tX)}\right)_{\exp(tX)} (\tilde{Y}_{\Phi^t_{\tilde{X}}(1_G)}) \\
&= \left(\mathrm{d}\Phi^{-t}_{\tilde{X}}\right)_{\Phi^t_{\tilde{X}}(1_G)} \tilde{Y}_{\Phi^t_{\tilde{X}}(1_G)},
\end{aligned}$$

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 Corollary 5.2.7. We get

$$\begin{aligned}
(\mathrm{d}\,\mathrm{Ad})_{1_G}(X)(Y) &= \frac{\mathrm{d}}{\mathrm{d}t}\left(\mathrm{d}\Phi^{-t}_{\tilde{X}}\right)_{\Phi^t_{\tilde{X}}(1_G)} \tilde{Y}_{\Phi^t_{\tilde{X}}(1_G)}\Big|_{t=0} \\
&\overset{\text{def. of}}{\underset{\text{Lie deriv.}}{=}} \left(\mathcal{L}_{\tilde{X}}(\tilde{Y})\right)_{1_G} \\
&\overset{3.2.2.c}{=} [\tilde{X}, \tilde{Y}]_{1_G} \\
&= [X, Y] \\
&= \mathrm{ad}_X(Y).
\end{aligned}$$

□

Recall that if φ is a Lie group homomorphism and φ_* is the Lie algebra homomorphism induced by φ, by Proposition 5.2.9 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.

$$\begin{array}{ccc} \mathfrak{g} & \xrightarrow{\varphi_*} & \mathfrak{h} \\ {\scriptstyle \exp}\downarrow & & \downarrow{\scriptstyle \exp} \\ G & \xrightarrow[\varphi]{} & H \end{array} \qquad \begin{array}{ccc} \mathfrak{g} & \xrightarrow{(C_g)_* = \mathrm{Ad}_g} & \mathfrak{g} \\ {\scriptstyle \exp}\downarrow & & \downarrow{\scriptstyle \exp} \\ G & \xrightarrow[C_g]{} & G \end{array} \qquad \begin{array}{ccc} \mathfrak{g} & \xrightarrow{(\mathrm{Ad})_* = \mathrm{ad}} & \mathfrak{gl}(\mathfrak{g}) \\ {\scriptstyle \exp}\downarrow & & \downarrow{\scriptstyle \exp = e^{\cdot}} \\ G & \xrightarrow[\mathrm{Ad}]{} & \mathrm{GL}(\mathfrak{g}) \end{array}$$

Formula 5.5.6 *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,

$$\exp(Y)\exp(X)\exp(-Y) = \exp(\mathrm{Ad}_{\exp(Y)} X), \qquad \forall X, Y \in \mathfrak{g}.$$

Formula 5.5.7 *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

$$\begin{aligned} e^{\mathrm{ad}_X}(Y) &= \sum_{k=0}^{\infty} \frac{(\mathrm{ad}_X)^k}{k!}(Y) \\ &= Y + [X, Y] + \frac{1}{2}[X, [X, Y]] + \frac{1}{3!}[X, [X, [X, Y]]] + \dots \end{aligned}$$

Formula 5.5.8 (See Exercises 5.8.36, 5.8.37, and 5.8.38) *Let V be a vector space. For all $X, Y \in \mathfrak{gl}(V)$ and $B \in \mathrm{GL}(V)$ we have*

5.5.8.i. $\mathrm{Ad}_B(X) = B \cdot X \cdot B^{-1}$;
5.5.8.ii. $e^{\mathrm{ad}_X} Y = e^X Y e^{-X}$;
5.5.8.iii. $e^{BXB^{-1}} = B e^X B^{-1}$.

Formula 5.5.9 (See Exercise 5.8.39) *For every left-invariant vector fields X, Y on a Lie group G, we have that $(\Phi_X^t)_* Y$ is a left-invariant vector field and*

$$(\Phi_X^t)_* Y = e^{-\mathrm{ad}(tX)} Y, \qquad \forall t \in \mathbb{R}.$$

5.6 Semi-Direct Products

In this section, our focus turns to the study of the semi-direct product of Lie groups. We begin by considering group actions by group automorphisms, as well as Lie actions by derivations. These actions allow the formation of semi-direct products of Lie groups and, in a parallel manner, semi-direct products of Lie algebras. A key result will emerge: the Lie algebra of a semi-direct product of Lie groups corresponds to a semi-direct product of the Lie algebras associated with the constituent groups.

Let G be a group and X a set. An *action* of G on X is a map $\Theta\colon G \times X \to X$ such that $\Theta(1_G, p) = p$ for all $p \in X$, and

$$\Theta(g_1, \Theta(g_2, p)) = \Theta(g_1 g_2, p), \qquad \forall g_1, g_2 \in G, \forall p \in X. \tag{5.14}$$

Actions, as we have just defined, are also called *left actions* or *group actions*. We write $G \curvearrowright X$ in case of an action, and we write $g.p$ for $\Theta(g, p)$. The action defines a group homomorphism $G \to \mathfrak{S}(X)$, $g \mapsto \Theta(g, \cdot)$, into the group $\mathfrak{S}(X)$ of permutations of X. Actions will be extensively considered again in Sect. 6.1.1.

5.6.1 Derivations and Actions by Automorphisms

Definition 5.6.1 (Group Action by Automorphisms) Let H and G be groups. An *action* of H *by automorphisms* of G is a group homomorphism $\theta : H \to \mathrm{Aut}(G)$. Equivalently, it is a map $\theta : H \times G \to G$, where for each $h \in H$, we define $\theta_h := \theta(h, \cdot)$, and the following hold:

$$\theta_h(g_1 g_2) = \theta_h(g_1)\theta_h(g_2), \qquad \forall g_1, g_2 \in G, \forall h \in H$$

and

$$\theta_{h_1 h_2} = \theta_{h_1} \circ \theta_{h_2}, \qquad \forall h_1, h_2 \in H. \tag{5.15}$$

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 smooth element $\varphi \in \mathrm{Aut}(G)$ has an associated Lie algebra automorphism φ_*. We introduce a notation for the *space of Lie group automorphisms* of a Lie group G:

$$\mathrm{Aut}_{\mathrm{Lie}}(G) := \mathrm{Aut}(G) \cap C^\infty(G; G) = \{\varphi : G \to G \text{ smooth automorphism}\}$$

and 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 5.6.2 (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 rule:

$$D([X, Y]) = [D(X), Y] + [X, D(Y)], \qquad \forall X, Y \in \mathfrak{g}. \tag{5.16}$$

Let $\mathrm{Der}(\mathfrak{g})$ be the *set of derivations* on $\mathfrak{g}$.

We state the relative results for reference (see Exercises 5.8.50–5.8.56):

Proposition 5.6.3 *Let G be a Lie group with Lie algebra $\mathfrak{g}$.*

5.6.3.i. The natural map

$$\begin{aligned} \mathrm{Aut}_{\mathrm{Lie}}(G) &\longrightarrow \mathrm{Aut}_{\mathrm{Lie}}(\mathfrak{g}) \\ \varphi &\longmapsto \varphi_* \end{aligned}$$

is a Lie group homomorphism, which is injective if G is connected. If G is simply connected, it is a Lie group isomorphism.

5.6.3.ii. $\mathrm{Aut}_{\mathrm{Lie}}(\mathfrak{g})$ *is a regular Lie subgroup of* $\mathrm{GL}(\mathfrak{g})$ *with*

$$\mathrm{Lie}(\mathrm{Aut}_{\mathrm{Lie}}(\mathfrak{g})) = \mathrm{Der}(\mathfrak{g}).$$

5.6.3.iii. The adjoint map ad *and the adjoint representation* Ad *satisfy*

$$\begin{array}{ccc} \mathfrak{g} & \xrightarrow{\ \mathrm{ad}\ } & \mathrm{Der}(\mathfrak{g}) \subset \mathfrak{gl}(\mathfrak{g}) \\ {\scriptstyle \exp}\downarrow & & \downarrow{\scriptstyle \exp} \\ G & \xrightarrow{\ \mathrm{Ad}\ } & \mathrm{Aut}_{\mathrm{Lie}}(\mathfrak{g}) \subset \mathrm{GL}(\mathfrak{g}). \end{array}$$

5.6.2 *Semi-Direct Products of Lie Algebras and Groups*

Definition 5.6.4 (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,

$$[(X, Y), (X', Y')] := ([X, X'] + \sigma(Y)(X') - \sigma(Y')(X), [Y, Y']), \quad \forall X, X' \in \mathfrak{g}, \forall Y, Y' \in \mathfrak{h}. \tag{5.17}$$

The resulting Lie algebra is the *semi-direct 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 to name it, we simply write $\mathfrak{g} \rtimes \mathfrak{h}$.

Remark 5.6.5 We have the following properties for the semi-direct product of Lie algebras.

5.6.5.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 $\mathfrak{g} \rtimes \mathfrak{h}$.
5.6.5.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 $\mathfrak{g} \times \mathfrak{h}$.
5.6.5.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.
5.6.5.iv. The map σ represents 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; see Exercise 5.8.52.

Definition 5.6.6 (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

$$(g_1, h_1) \cdot (g_2, h_2) := (g_1 \cdot \theta_{h_1}(g_2), h_1 h_2), \qquad \forall g_1, g_2 \in G, \forall h_1, h_2 \in H. \tag{5.18}$$

The resulting group is called 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 write θ explicitly.

Remark 5.6.7 Similarly to Remark 5.6.5, we have the following properties for the semi-direct product of groups.

5.6.7.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 $G \rtimes H$.
5.6.7.ii. If $\theta \equiv \mathrm{id}_G$, we call $G \rtimes H$ the *direct product* of G and H, and write it as $G \times H$.
5.6.7.iii. In $G \rtimes H$, the subset G is a normal subgroup. Hence, we have $G \lhd G \rtimes H$, explaining the choice of the symbol $\rtimes$.

5.6.7.iv. To understand and remember the product law, it is helpful to grasp its underlying reasoning.[2] Write the product of elements $g_1, g_2 \in G$ and $h_1, h_2 \in H$ as

$$g_1 h_1 g_2 h_2 = g_1 h_1 g_2 h_1^{-1} h_1 h_2 = g_1 C_{h_1}(g_2) h_1 h_2.$$

In other words, the map θ represents the conjugation in $G \rtimes H$ of H on G:

$$C_h(g) = \theta_h(g), \qquad \forall g \in G, \forall h \in H; \tag{5.19}$$

recall that indeed every C_g is a group automorphism; see Exercise 5.8.53.

5.6.7.v. The element (g, h) has inverse $(\theta_{h^{-1}}(g^{-1}), h^{-1})$.

5.6.3 Lie Algebras of Semi-Direct Products of Lie Groups

Proposition 5.6.8 *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.*

5.6.8.i. $G \rtimes_\theta H$ is a Lie group.

5.6.8.ii. The map $\tau : H \to \mathrm{Aut_{Lie}}(\mathfrak{g})$ defined by $\tau_h := (\theta_h)_*$, for $h \in H$, is a Lie group homomorphism.

5.6.8.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. Moreover, the group structure is smooth by construction. Thus, $G \rtimes_\theta H$ is a Lie group. Applying the chain rule to (5.15), we find that $\tau_{h_1 h_2} = \tau_{h_1} \circ \tau_{h_2}$ for all $h_1, h_2 \in H$. So also 5.6.8.ii is proved.

To prove 5.6.8.iii, we need to compute 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 induced morphisms; see Proposition 5.2.9. In fact, we have

$$\theta_h(\exp(X)) = \exp(\tau_h(X)), \qquad \forall X \in \mathfrak{g}, \forall h \in H \tag{5.20}$$

[2] A Finnish motto says: *sitä, minkä ymmärtää, ei tarvitse muistaa*. [what you understand, you don't need to remember.]

and

$$\tau(\exp(Y)) = \exp(\sigma(Y)), \qquad \forall Y \in \mathfrak{h}. \tag{5.21}$$

For all $X \in \mathfrak{g}$, $Y \in \mathfrak{h}$, and $t \in \mathbb{R}$, we have

$$\begin{aligned} \exp(t\,\mathrm{Ad}_{\exp(Y)}(X)) &= \exp(\mathrm{Ad}_{\exp(Y)}(tX)) \\ &\overset{\text{F. 5.5.6}}{=} C_{\exp(Y)}(\exp(tX)) \\ &\overset{(5.19)}{=} \theta_{\exp(Y)}(\exp(tX)) \\ &\overset{(5.20)}{=} \exp(\tau_{\exp(Y)}(tX)) \\ &= \exp(t\tau_{\exp(Y)}(X)), \end{aligned}$$

where we also used that both $\mathrm{Ad}_{\exp(Y)}$ and $\tau_{\exp(Y)}$ are linear. We got an identity between OPSs. 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,

$$e^{\mathrm{ad}_Y} \overset{(5.5.7)}{=} \mathrm{Ad}_{\exp(Y)} = \tau_{\exp(Y)} \overset{(5.21)}{=} e^{\sigma(Y)}.$$

Differentiating in Y we get $\mathrm{ad}_Y = \sigma(Y)$. □

Remark 5.6.9 Conversely, every semi-direct product of Lie algebras is the Lie algebra of a semi-direct product of Lie groups. Indeed, let $\mathfrak{g} \rtimes_\sigma \mathfrak{h}$ be a semi-direct product of Lie algebras, and let G and H be simply connected Lie groups with Lie algebras $\mathfrak{g}$ and $\mathfrak{h}$, respectively, whose existence is ensured by Theorem 5.1.7. From Theorem 5.1.5, 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 5.1.5, 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 $\mathrm{Lie}(G \rtimes_\theta H) = \mathfrak{g} \rtimes_\sigma \mathfrak{h}$.

5.7 From Algebras to Groups

In this section, we revisit the discussion from Sect. 5.1 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 group homomorphisms from Lie algebra homomorphisms, provided that the Lie group in the source is simply connected.

5.7.1 Existence of Subgroups

The next result shows the existence of Lie subgroups with given Lie subalgebras of Lie algebras of Lie groups. Together with Ado's theorem (see [Jac79, page 199] and Theorem 5.1.7), 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 5.7.1 (Existence of Subgroups) *Let G be a Lie group with Lie algebra $\mathfrak{g}$. For every Lie subalgebra $\mathfrak{h} \subset \mathfrak{g}$, there is a unique connected Lie subgroup H with Lie algebra $\mathfrak{h}$; in fact, H is the group generated by* $\exp(\mathfrak{h})$. *It may not be true that* $H = \exp(\mathfrak{h})$.

Proof This is a consequence of Frobenius's theorem; see for example [Lee13, page 496] or [AT11, Sec. 3.7]. We consider the subbundle $\Delta \subset TG$ defined by

$$\Delta_g := (\mathrm{d}L_g)_1(\mathfrak{h}), \qquad \forall g \in G.$$

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, we have $T_1 H = \mathfrak{h}$. 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_G and is tangent to Δ. By maximality, we have $h_1^{-1} h_2 \in L_{h_1^{-1}} H \subseteq H$.

Regarding uniqueness, if $\hat{H}$ is a connected subgroup with $\mathrm{Lie}(\hat{H}) = \mathfrak{h}$, since $\exp(\mathfrak{h})$ is an open neighborhood of 1 in $\hat{H}$, we have

$$\hat{H} = (\hat{H})^\circ = \langle \exp(\mathfrak{h}) \rangle,$$

where we used Exercise 5.8.3. The last assertion of the theorem comes from the fact that there are connected Lie groups with non-surjective exponential map; see Exercise 5.8.32. □

5.7.2 Existence of Group Homomorphisms

We shall show that every Lie algebra homomorphism between Lie algebras of Lie groups is induced by a Lie group homomorphism when 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.

We will use the following fact: when the Lie algebra homomorphism induced by a Lie group homomorphism is a bijection, then the group homomorphism is a covering map; see Exercise 5.8.31. For the basics of algebraic topology, such as the fact that a covering map onto a simply connected space is a homeomorphism, we refer to [Mun75].

Theorem 5.7.2 (Induced Lie Group Homomorphism) *Let G and H be Lie groups. Assume G is simply connected. For each 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 $\mathfrak{g} \times \mathfrak{h} = \mathrm{Lie}(G \times H)$: for $X, Y \in \mathfrak{g}$ we have

$$[(X, \psi(X)), (Y, \psi(Y))] = ([X, Y], [\psi(X), \psi(Y)]) = ([X, Y], \psi[X, Y]) .$$

By Theorem 5.7.1, there is a unique connected Lie subgroup $K \subset G \times H$ with $\mathrm{Lie}(K) = \mathfrak{k}$. 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

$$(\mathrm{d}\phi)_{(1_G,1_H)}(X, \psi(X)) = X, \quad \forall X \in \mathfrak{g}. \tag{5.22}$$

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 5.8.31, we deduce that $\phi : K \to G$ is a covering map. Since G is simply connected, the map ϕ is a bijection; see [Mun75, Theorem 54.4]. Hence, it is an isomorphism. Set $\varphi := \pi_2|_K \circ \phi^{-1} : G \to H$, which is a Lie group homomorphism. From (5.22) we also get that

$$\begin{aligned}(\mathrm{d}\varphi)_{1_G}(X) &= (\mathrm{d}\pi_2)_{(1_G,1_H)} \circ (\mathrm{d}\phi^{-1})_{1_G}(X) \\ &= (\mathrm{d}\pi_2)_{(1_G,1_H)}(X, \psi(X)) \\ &= \psi(X), \qquad\qquad \forall X \in \mathfrak{g},\end{aligned}$$

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 φ and $\tilde{\varphi}$ are group homomorphisms, we get that $\tilde{\varphi} = \varphi$. □

5.7.3 The Baker-Campbell-Dynkin-Hausdorff Formula

The punchline of this subsection is that near the identity element, the group product of every Lie group can be expressed in terms of operations on its Lie algebra. Various formulas have been found over the years. In particular, we recall chronologically the work of Campbell (1897–8), Baker (1901–5), Hausdorff (1906), and Dynkin (1947–49), see [Dyn49] and read more in [BF11]. The Baker-Campbell-Hausdorff formula links Lie groups to Lie algebras by expressing $\log(e^X e^Y)$ as an infinite sum in the iterated Lie brackets of elements X and Y in the Lie algebra. The logarithm is, by definition, the inverse of the exponential; in general, it is only locally defined in a neighborhood of the identity; see Proposition 5.2.8. However, for nilpotent simply connected Lie groups, the logarithm will be global by Theorem 9.4.6.

For matrix Lie groups, a Baker-Campbell-Hausdorff formula can be obtained using formal series as follows. Namely, for every $A, B \in \mathfrak{gl}(n)$ if A and B are enough near 0 so that the series converge, we write

$$\begin{aligned}
\log(e^A e^B) &= \log(\mathbb{I} + (e^A e^B - \mathbb{I})) \\
&= \sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k} (e^A e^B - \mathbb{I})^k, \\
&= \sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k} \left(\left(\sum_{i=0}^{\infty} \frac{A^i}{i!} \right) \left(\sum_{j=0}^{\infty} \frac{B^j}{j!} \right) - \mathbb{I} \right)^k \\
&= \sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k} \left(\sum_{\substack{i,j \in \mathbb{N}\cup\{0\} \\ (i,j)\neq(0,0)}} \frac{A^i B^j}{i!j!} \right)^k \\
&= \sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k} \sum_{\substack{i_\ell, j_\ell \in \mathbb{N}\cup\{0\} \\ (i_\ell, j_\ell)\neq(0,0)}} \frac{A^{i_1} B^{j_1} \cdots A^{i_k} B^{j_k}}{i_1! j_2! \cdots i_k! j_k!}.
\end{aligned} \tag{5.23}$$

We seek a formula that expresses the above series of products as a series of iterated adjoint maps, like

$$\begin{aligned}
&\left(\mathrm{ad}_X^{r_1} \circ \mathrm{ad}_Y^{s_1} \circ \ldots \circ \mathrm{ad}_X^{r_n}\right)(Y) \\
&= [\underbrace{X, [X, \ldots [X,}_{r_1} [\underbrace{Y, [Y, \ldots [Y,}_{s_1} \ldots [\underbrace{X, [X, \ldots [X,}_{r_n} Y] \ldots]].
\end{aligned}$$

We follow the original work of Dynkin [Dyn49]; see also the recent short review [Bos89]. We shall state the *Baker-Campbell-Hausdorff Formula* (*BCH Formula*, for short) using what is called the Dynkin product:

Definition 5.7.3 (Dynkin Product) For elements A and B in a Lie algebra, we define their *Dynkin product* as

$$A \star B := \sum_{k,m \geq 0,\, i_\ell + j_\ell \neq 0} \frac{(-1)^k}{(k+1)(j_1 + \cdots + j_k + 1)} \frac{\operatorname{ad}_A^{i_1} \operatorname{ad}_B^{j_1} \ldots \operatorname{ad}_A^{i_k} \operatorname{ad}_B^{j_k} \operatorname{ad}_A^{m}(B)}{i_1! \ldots i_k! j_1! \ldots j_k! m!},$$

if the series converges.

Proposition 5.7.4 (BCH Formula) *For every* $A, B \in \mathfrak{gl}(n)$ *with* $\|A\|, \|B\| < \frac{1}{2}\log(2 - \sqrt{2}/2)$ *we have*

$$A \star B = \log(e^A e^B). \tag{5.24}$$

A clear proof can be found in [HN12, Proposition 3.4.5]. Moreover, there are other equivalent ways of writing the Dynkin product:

$$X \star Y = \sum_{n>0} \frac{(-1)^{n-1}}{n} \sum_{\substack{r_i + s_i > 0 \\ 1 \leq i \leq n}} \frac{\left(\operatorname{ad}_X^{r_1} \circ \operatorname{ad}_Y^{s_1} \circ \operatorname{ad}_X^{r_2} \circ \operatorname{ad}_Y^{s_2} \ldots \circ \operatorname{ad}_X^{r_n} \circ \operatorname{ad}_Y^{s_n - 1}\right)(Y)}{r_1! s_1! \cdots r_n! s_n! \; \sum_{i=1}^n (r_i + s_i)}. \tag{5.25}$$

The first terms of the series are

$$\begin{aligned}
\log(\exp X \exp Y) = X + Y + \frac{1}{2}[X, Y] &+ \frac{1}{12}\Big([X, [X, Y]] + [Y, [Y, X]]\Big) \\
&- \frac{1}{24}[Y, [X, [X, Y]]] \\
&- \frac{1}{720}\Big([[[[X, Y], Y], Y], Y] + [[[[Y, X], X], X], X]\Big) \\
&+ \frac{1}{360}\Big([[[[X, Y], Y], Y], X] + [[[[Y, X], X], X], Y]\Big) \\
&+ \frac{1}{120}\Big([[[[Y, X], Y], X], Y] + [[[[X, Y], X], Y], X]\Big) + \cdots
\end{aligned}$$

Some consequences of the BCH formula are the following:

5.7.4.i. *First expansion of BCH*:

$$e^X e^Y = \exp\left(X + Y + \frac{1}{2}[X, Y] + o(\|X\| \cdot \|Y\|)\right), \text{ as } \|X\|, \|Y\| \to 0. \tag{5.26}$$

5.7.4.ii. *Trotter product formula*:

$$\lim_{k \to \infty} \left(e^{\frac{A}{k}} e^{\frac{B}{k}}\right)^k = e^{A+B}, \qquad \forall A, B \in \mathfrak{gl}(n).$$

5.7.4.iii. *Commutator formula*:

$$\lim_{k\to\infty}\left(e^{\frac{A}{k}}e^{\frac{B}{k}}e^{-\frac{A}{k}}e^{-\frac{B}{k}}\right)^{k^2}=e^{AB-BA},\qquad \forall A,B\in\mathfrak{gl}(n).$$

5.8 Exercises

Exercise 5.8.1 For every elements g, h in a group G we have

5.8.1.i. $L_h\circ L_g=L_{hg}$, 5.8.1.ii. $R_h\circ R_g=R_{gh}$,
5.8.1.iii. $L_h\circ R_g=R_g\circ L_h$, 5.8.1.iv. $(L_g)^{-1}=L_{g^{-1}}$,
5.8.1.iv. $(R_g)^{-1}=R_{g^{-1}}$, 5.8.1.vi. $C_{gh}=C_g\circ C_h$,

where L, R, and C denote the left translations, right translations, and conjugations.

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 5.8.2 Let G be a Lie group (or, more generally, a topological group). If $U\subset G$ is open, then U^2 is open.

Exercise 5.8.3 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, $V^{-1}=V$ and $1\in V$. 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 Exercise 5.8.2). 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 5.8.4 Let G be a Lie group.

5.8.4.i. If H is a subgroup of G that, topologically, is open, then it is closed;
5.8.4.ii. If H is a subgroup of G that has a nonempty interior, then it is open and closed.
5.8.4.iii. Let G° be the *identity component* of G, which is the connected component of G containing 1_G. We have that G° is an open, closed, normal subgroup of G. Its Lie algebra is the same as G;
5.8.4.iv. Every neighborhood $U\subseteq G$ of 1 in G generates G°, i.e., every element in G° is the product of finitely many elements in U.

Exercise 5.8.5 On topological groups, right translations and left translations are homeomorphisms. While, in Lie groups, they are smooth diffeomorphisms.

Exercise 5.8.6 The anti-commutativity and Jacobi identity for Lie algebras rephrase the fact that the structural constants c_{ij}^k as in (5.2) satisfy:

5.8.6.i. $0 = c_{ij}^k + c_{ji}^k$, $\forall i, j, k \in \{1, \dots, n\}$;

5.8.6.ii. $0 = \sum_{r=1}^n \left(c_{ij}^r c_{rk}^s + c_{jk}^r c_{ri}^s + c_{ki}^r c_{rj}^s \right)$, $\forall i, j, k, s \in \{1, \dots, n\}$.

Exercise 5.8.7 Let $c_{ij}^k \in \mathbb{R}$ satisfying 5.8.6.i-ii. Define $[\cdot, \cdot]$ by (5.2). Then $[\cdot, \cdot]$ uniquely extends into a Lie bracket turning $\operatorname{span}\{X_1, \dots, X_n\}$ into a Lie algebra.

Exercise 5.8.8 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 $X_1, \dots, X_n$, and let $\tilde{c}_{ij}^k$ be the structural constants of $\tilde{\mathfrak{g}}$ with respect to a basis $\tilde{X}_1, \dots, \tilde{X}_n$. If $c_{ij}^k = \tilde{c}_{ij}^k$, then the map $\psi : \mathfrak{g} \to \tilde{\mathfrak{g}}$ defined by $\psi(X_i) := \tilde{X}_i$ is a Lie algebra isomorphism.

Exercise 5.8.9 The Lie bracket of two left-invariant vector fields is left-invariant.
Solution. For left-invariant vector fields X, Y on a Lie group G and $g \in G$, we have $(L_g)_*[X, Y] = [(L_g)_* X, (L_g)_* Y] = [X, Y]$.

Exercise 5.8.10 (Right Translations of LIVFs) Let X be a left-invariant vector field on a Lie group G. Let R_g be the right translation by an element $g \in G$. Then, the vector field $(R_g)_* X$ is left-invariant.
Solution. Let $h \in G$. Then, using Exercise 5.8.1.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 5.8.11 (Derivative of Product of Curves) Let G be a Lie group. Let $\gamma, \sigma : \mathbb{R} \to G$ be two smooth curves into G. Consider the product of the two curves, i.e., the curve

$$t \longmapsto \gamma(t)\sigma(t)$$

and calculate the derivative of such a curve in terms of γ, σ, and their derivatives. In fact, a formula is

$$\frac{\mathrm{d}}{\mathrm{d}t}\gamma(t)\sigma(t) = (\mathrm{d}R_{\sigma(t)})_{\gamma(t)}\dot{\gamma}(t) + (\mathrm{d}L_{\gamma(t)})_{\sigma(t)}\dot{\sigma}(t). \tag{5.27}$$

Solution. Differentiating one variable at a time, we get

$$\begin{aligned}\left.\frac{\mathrm{d}}{\mathrm{d}t}\gamma(t)\sigma(t)\right|_{t=t_0} &= \left.\frac{\mathrm{d}}{\mathrm{d}t}\gamma(t)\sigma(t_0)\right|_{t=t_0} + \left.\frac{\mathrm{d}}{\mathrm{d}t}\gamma(t_0)\sigma(t)\right|_{t=t_0} \\ &= \left.\frac{\mathrm{d}}{\mathrm{d}t}(R_{\sigma(t_0)}\gamma(t))\right|_{t=t_0} + \left.\frac{\mathrm{d}}{\mathrm{d}t}(L_{\gamma(t_0)}\sigma(t))\right|_{t=t_0} \\ &= (\mathrm{d}R_{\sigma(t_0)})_{\gamma(t_0)}\dot{\gamma}(t_0) + (\mathrm{d}L_{\gamma(t_0)})_{\sigma(t_0)}\dot{\sigma}(t_0).\end{aligned}$$

Exercise 5.8.12 Let G be a Lie group. Let $\gamma : \mathbb{R} \to G$ be a smooth curve into G. Consider the curve

$$t \longmapsto \gamma(t)^{-1}$$

and calculate the derivative at an arbitrary t of such a curve in terms of γ and $\dot{\gamma}$. In fact, a formula is

$$\frac{\mathrm{d}}{\mathrm{d}t}(\gamma(t)^{-1}) = -(\mathrm{d}L_{\gamma(t)^{-1}})_{1_G}(\mathrm{d}R_{\gamma(t)^{-1}})_{\gamma(t)}\dot{\gamma}(t). \tag{5.28}$$

Solution. From the fact that $1 = \gamma(t)\gamma(t)^{-1}$, for all t, and formula (5.27), 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 5.8.13 Let $\varphi : G \to H$ be a group homomorphism.

5.8.13.i. We have $\varphi \circ L_g = L_{\varphi(g)} \circ \varphi$, for all $g \in G$;
5.8.13.ii. We have $\varphi \circ R_g = R_{\varphi(g)} \circ \varphi$, for all $g \in G$.

Exercise 5.8.14 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 $(\varphi_* X)_{1_H} = (\mathrm{d}\varphi)_{1_G}(X_{1_G})$.

5.8.14.i. The vector fields X and $\varphi_* X$ are φ-*related*, in the sense that $(\mathrm{d}\varphi)_g X_g = (\varphi_* X)_{\varphi(g)}$, for all $g \in G$.
5.8.14.ii. If $g, g' \in G$ are such that $\varphi(g) = \varphi(g')$, then $(\mathrm{d}\varphi)_g X_g = (\mathrm{d}\varphi)_{g'} X_{g'}$.
5.8.14.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.
5.8.14.iv. $\varphi_* : \mathrm{Lie}(G) \to \mathrm{Lie}(H)$ is a Lie algebra homomorphism.
5.8.14.v. $(\mathrm{d}\varphi)_{1_G} : (T_{1_G}, [\cdot,\cdot]) \to (T_{1_H} H, [\cdot,\cdot])$ is a Lie algebra homomorphism.

Hint. From Exercise 5.8.13.(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 5.8.15 Let $X^\dagger$ be a right-invariant vector field on G. A curve $\theta : \mathbb{R} \to G$ 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 5.8.16 For a vector X in the Lie algebra of a Lie group G, consider the map $\psi : \mathrm{Lie}(\mathbb{R}) \to \mathrm{Lie}(G)$, $t \mapsto tX$.

5.8.16.i. The map ψ is a Lie algebra homomorphism.
5.8.16.ii. There exists a one-parameter subgroup $\gamma : \mathbb{R} \to G$ with $\mathrm{d}\gamma = \psi$.
5.8.16.iii. We have $\dot\gamma(t) = X_{\gamma(t)}$.

Solution. Since $\mathbb{R}$ is simply connected, Theorem 5.7.2 asserts that there exists such a γ. Regarding (iii), we have

$$\begin{aligned}
\dot\gamma(t) &= \frac{\mathrm{d}}{\mathrm{d}h}\gamma(t+h)\Big|_{h=0} \\
&= \frac{\mathrm{d}}{\mathrm{d}h}\gamma(t)\gamma(h)\Big|_{h=0} \\
&= \frac{\mathrm{d}}{\mathrm{d}h}L_{\gamma(t)}(\gamma(h))\Big|_{h=0} \\
&= (L_{\gamma(t)})_*(\mathrm{d}\gamma)_0(\partial_t) \\
&= (L_{\gamma(t)})_*\psi(\partial_t)
\end{aligned}$$

$$= (L_{\gamma(t)})_* X$$
$$= X_{\gamma(t)}.$$

Exercise 5.8.17 Let G be a Lie group and X be a left-invariant vector field on G.

5.8.17.i. $\Phi^1_{tX} = \Phi^t_X$, for all $t \in \mathbb{R}$;
5.8.17.ii. $\exp(tX) = \Phi^t_X(1_G)$, for all $t \in \mathbb{R}$;
5.8.17.iii. $\exp(sX + tX) = \exp(sX)\exp(tX)$, for all $s, t \in \mathbb{R}$;
5.8.17.iv. $\exp(0) = 1_G$;
5.8.17.v. $\exp(-X) = (\exp(X))^{-1}$;
5.8.17.vi. $t \mapsto \exp(tX)$ is a one-parameter subgroup and an integral curve of X;
5.8.17.vii. Equations (5.4) and (5.5) holds true.

Exercise 5.8.18 Let G be a Lie group. For all $m \in \mathbb{Z}$ and $X \in \mathrm{Lie}(G)$ we have

$$\exp(mX) = (\exp(X))^m.$$

Pay attention that this is correct also when m is negative.

Exercise 5.8.19 Let X be a left-invariant vector field in a Lie group G. Then, we have

$$Xf(\exp(tX)) = \frac{\mathrm{d}}{\mathrm{d}t} f(\exp(tX)), \qquad \forall f \in C^\infty(G), \forall t \in \mathbb{R}.$$

Hint. See Exercise 3.4.50.

Exercise 5.8.20 (Lie Group Structures on Tangent Bundles of Lie Groups) Let G be a Lie group.

5.8.20.i. On TG, consider the following operation:

$$(g, X_g) * (h, Y_h) := d(R_h)_g(X_g) + d(L_g)_h(Y_h) \in T_{gh}G,$$

for all $g, h \in G$, $X_g \in T_gG$, and $Y_h \in T_hG$. Then, the pair $(TG, *)$ is a Lie group.

5.8.20.ii. On $\mathrm{Lie}(G) \times G$ consider the following operation:

$$(X, g) \bullet (Y, h) := (X + \mathrm{d}(C_g)_{1_G} Y, gh),$$

for all $g, h \in G$, $X, Y \in \mathrm{Lie}(G)$. Then, the pair $(\mathrm{Lie}(G) \times G, \bullet)$ is a Lie group.

5.8.20.ii. The two Lie groups $(TG, *)$ and $(\mathrm{Lie}(G) \times G, \bullet)$ are isomorphic.

Exercise 5.8.21 Let G be a Lie group. Let $\gamma : \mathbb{R} \to G$ be a smooth curve into G with $\gamma(0) = 1_G$ and $\dot\gamma(0) = X$. Then, we have

$$\lim_{k\to\infty} (\gamma(t/k))^k = \exp(tX), \qquad \forall t \in \mathbb{R}.$$

Solution. For t small enough, one can consider $\eta(t) := \exp^{-1}(\gamma(t))$. For fixed $t \in \mathbb{R}$, we have

$$\begin{aligned} t\dot{\eta}(0) &= t \lim_{h \to 0} \frac{\eta(h) - \eta(0)}{h} = t \lim_{k \to \infty} \frac{\eta(t/k) - \eta(0)}{t/k} \\ &= \lim_{k \to \infty} k\eta(t/k). \end{aligned}$$

Because of Proposition 5.2.8, we write

$$\dot{\eta}(0) = \frac{\mathrm{d}}{\mathrm{d}t} \exp^{-1}(\gamma(t))\Big|_{t=0} = (\mathrm{d}\exp^{-1})_0 \dot{\gamma}(0) = (\mathrm{d}\exp)_0^{-1} \dot{\gamma}(0) = \dot{\gamma}(0).$$

Thus, using the previous two formulas, we get

$$\begin{aligned} \exp(tX) &= \exp(t\dot{\gamma}(0)) = \exp(t\dot{\eta}(0)) = \exp(\lim_{k \to \infty} k\eta(t/k)) \\ &= \lim_{k \to \infty} \exp(k\eta(t/k)) = \lim_{k \to \infty} (\exp(\eta(t/k))^k = \lim_{k \to \infty} (\gamma(t/k))^k. \end{aligned}$$

Exercise 5.8.22 If X is a left-invariant vector field and Y is a right-invariant vector field, then $[X, Y] = 0$.

Exercise 5.8.23 If G is a commutative Lie group, then $\mathrm{Lie}(G)$ is a commutative Lie algebra.
Hint. Use Exercise 5.8.22.

Exercise 5.8.24 (What Happens if One Uses Right-Invariant Vector Fields as Lie Algebra) 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 $X \mapsto -X$.

Solution. 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

$$J_* \tilde{X} = -X^\dagger, \qquad \forall X \in T_{1_G} G. \tag{5.29}$$

Indeed, using Corollary 5.2.7, 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 (5.29). Therefore, the map $J_* : (T_{1_G} G, [\cdot, \cdot]) \to (T_{1_G} G, [\cdot, \cdot]_R)$ gives a Lie algebra isomorphism.

Exercise 5.8.25 Given a Lie group G with Lie algebra $\mathfrak{g}$, consider the vector field Y on the manifold $G \times \mathfrak{g}$ defined as follows. 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 have

$$\Phi_Y^t((g, X)) = (\Phi_X^t(g), X), \qquad \forall t \in \mathbb{R}, \forall g \in G, \forall X \in \mathfrak{g}.$$

Thus, the map $X \mapsto \exp(X)$ is smooth because it is the projection of the flow at time 1 of the smooth vector field Y.

Solution. We have $\frac{\mathrm{d}}{\mathrm{d}t}(\Phi_X^t(g), X) = (X_{\Phi_X^t(g)}, 0) = Y_{(\Phi_X^t(g), X)}$ and $\left(\Phi_X^t(g), X\right)\big|_{t=0}$ $= (g, X)$. We deduce that $\Phi_X^t(g)$, which is the first coordinate of the above flow, depends smoothly on the point (g, X) and so does $\exp(X) \stackrel{\text{def}}{=} \Phi_X^1(1_G)$.

Exercise 5.8.26 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. We have that $\varphi_1 = \varphi_2$. But, there are counterexamples when G is not connected.

Exercise 5.8.27 Proposition 5.2.9 implies that every Lie group homomorphism F has the following properties:

5.8.27.i. F_* is injective (resp. surjective) if and only if F is locally injective (resp. open) at 1_G.
5.8.27.ii. Given $g \in G$, F_* is injective (resp. surjective) if and only if F is locally injective (resp. open) at g.
5.8.27.iii. F_* is injective (resp. surjective) if and only if F is locally injective (resp. open).
5.8.27.iv. If F is bijective, then F^{-1} is smooth, hence F is a diffeomorphism.

Hint. Use Proposition 5.2.9.

Exercise 5.8.28 Every injective Lie group homomorphism is an immersion (i.e., the differential is injective).
Hint. Use Proposition 5.2.9.

Exercise 5.8.29 In the definition of Lie subgroup, one can replace the requirement that the inclusion is an immersion by requiring that it is continuous. Namely, a subgroup $H < G$ of a Lie group G is a Lie subgroup of G if H admits the structure of a Lie group such that the inclusion $H \hookrightarrow G$ is a continuous homomorphism.

Exercise 5.8.30 (Square Root of a Matrix) Let G be a Lie group with Lie algebra $\mathfrak{g}$ and identity component G°.

5.8.30.i. We have $\exp(\mathfrak{g}) \subset G^\circ$.
5.8.30.ii. For each $A \in \exp(\mathfrak{g})$ there exists $B \in G$ such that $B^2 = A$. (Every such B is called *a square root* of A).

Hint. If $A = \exp(X)$ take $B := \exp(\frac{1}{2}X)$.

Exercise 5.8.31 ☠ Let G, H be connected Lie groups, and $\varphi : G \to H$ a Lie group homomorphism. The following are equivalent:

5.8.31.i. The map φ is surjective and has a discrete kernel;
5.8.31.ii. The map φ is a covering map;
5.8.31.iii. The map φ_* is an isomorphism of Lie algebras;
5.8.31.iv. The map φ is a local diffeomorphism.

Hint. The proof can be found in [AT11, page 182, Proposizione 3.8.2].

Exercise 5.8.32 (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.

5.8.32.i. The Lie group $\mathrm{GL}^+(n, \mathbb{R})$ is open and connected, and it is the identity component of $\mathrm{GL}(n, \mathbb{R})$.
5.8.32.ii. Exercise 5.8.30 implies 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 $\begin{bmatrix} -1 & 0 \\ 0 & -2 \end{bmatrix}$ or $\begin{bmatrix} -1 & -1 \\ 0 & -1 \end{bmatrix} \in \mathrm{GL}^+(2, \mathbb{R})$.

Exercise 5.8.33 The linearity of the Lie bracket and the Jacobi Identity give (5.5.2).i and (5.5.2).ii.

Solution of (5.5.2).ii. For $X, Y, Z \in \mathfrak{g}$, we have $\operatorname{ad}([X, Y])(Z) = [[X, Y], Z] \overset{\text{Jacobi}}{=} [X, [Y, Z]] - [Y, [X, Z]] = \operatorname{ad}_X(\operatorname{ad}_Y(Z)) - \operatorname{ad}_Y(\operatorname{ad}_X(Z)) = [\operatorname{ad}_X, \operatorname{ad}_Y](Z)$.

Exercise 5.8.34 Exercise 5.8.1.vi implies (5.12).
Solution. Let $g, h \in G$ and differentiate at 1_G the identity $C_g \circ C_h = C_{gh}$, to get $\operatorname{Ad}(gh) = (\mathrm{d}C_{gh})_1 = (\mathrm{d}C_g)_1 \circ (\mathrm{d}C_h)_1 = \operatorname{Ad}(g) \circ \operatorname{Ad}(h)$.

Exercise 5.8.35 For all X, Y in the Lie algebra of a Lie group, we have

$$\exp(X)\exp(Y)\exp(-X) = \exp(e^{\operatorname{ad}_X} Y).$$

Solution. Using Formula 5.5.6 first and then Formula 5.5.7, we have $\exp(X)\exp(Y)\exp(-X) = \exp(\operatorname{Ad}_{\exp(X)} Y) = \exp(e^{\operatorname{ad}_X} Y)$.

Exercise 5.8.36 For every vector space V, $A \in \mathfrak{gl}(V)$, and $B \in \mathrm{GL}(V)$, we have $\operatorname{Ad}_B(A) = BAB^{-1}$.
Solution. $\operatorname{Ad}_B(A) \overset{\text{def}}{=} (\mathrm{d}C_B)_{\mathbb{I}} A = \frac{\mathrm{d}}{\mathrm{d}t} C_B(e^{tA})\Big|_{t=0} = \frac{\mathrm{d}}{\mathrm{d}t} B e^{tA} B^{-1}\Big|_{t=0} = \frac{\mathrm{d}}{\mathrm{d}t} e^{tBAB^{-1}}\Big|_{t=0} = BAB^{-1}$.

Exercise 5.8.37 For all $X, Y \in \mathfrak{gl}(V)$, we have $e^{\operatorname{ad}_X} Y = e^X Y e^{-X}$.
Solution. Using Formula 5.5.7 and then Exercise 5.8.36, we have $e^{\operatorname{ad}_X} Y = \operatorname{Ad}_{e^X} Y = e^X Y e^{-X}$.

Exercise 5.8.38 For all $A \in \mathfrak{gl}(V)$ and for all $B \in \mathrm{GL}(V)$, we have $e^{BAB^{-1}} = Be^A B^{-1}$.
Hint. Notice that $(BAB^{-1})^k = BA^k B^{-1}$, for $k \in \mathbb{N}$. Expand $e^{BAB^{-1}}$ in power series.

Exercise 5.8.39 Let X, Y be left-invariant vector fields on a Lie group G. For all $t \in \mathbb{R}$

$$(\Phi^t_X)_* Y = e^{-\operatorname{ad}(tX)} Y.$$

Solution.

$$\begin{aligned}
(\Phi^t_X)_*(Y) &= (R_{\exp(tX)})_* Y \\
&= (R_{\exp(tX)})_* (L_{\exp(-tX)})_* Y \\
&= (R_{\exp(tX)} \circ L_{\exp(-tX)})_* Y \\
&= (C_{\exp(-tX)})_* Y \\
&= \operatorname{Ad}_{\exp(-tX)} Y = e^{\operatorname{ad}(-tX)} Y.
\end{aligned}$$

Exercise 5.8.40 If γ is a curve into a Lie group, then

$$\frac{\mathrm{d}}{\mathrm{d}s}\operatorname{Ad}_{\gamma(s)} = \operatorname{Ad}_{\gamma(s)} \operatorname{ad}\big((\mathrm{d}L^{-1}_{\gamma(s)})_{\gamma(s)}(\frac{\mathrm{d}}{\mathrm{d}s}\gamma(s))\big).$$

Solution. Use twice that $\operatorname{Ad}_p \circ \operatorname{Ad}_q = \operatorname{Ad}_{pq}$ to obtain

$$\begin{aligned}
\partial_s \operatorname{Ad}_{\gamma(s)} &= \partial_\epsilon \operatorname{Ad}_{\gamma(s+\epsilon)} |_{\epsilon=0} \\
&= \partial_\epsilon \operatorname{Ad}_{\gamma(s)} \operatorname{Ad}_{\gamma(s)^{-1}} \operatorname{Ad}_{\gamma(s+\epsilon)} |_{\epsilon=0} \\
&= \operatorname{Ad}_{\gamma(s)} \partial_\epsilon \operatorname{Ad}_{\gamma(s)^{-1}\gamma(s+\epsilon)} |_{\epsilon=0} \\
&= \operatorname{Ad}_{\gamma(s)} \operatorname{ad}(\partial_\epsilon(\gamma(s)^{-1}\gamma(s+\epsilon))|_{\epsilon=0}) \\
&= \operatorname{Ad}_{\gamma(s)} \operatorname{ad}\big((\mathrm{d}L^{-1}_{\gamma(s)})_{\gamma(s)}(\partial_s\gamma(s))\big).
\end{aligned}$$

Exercise 5.8.41 For all $A \in \operatorname{Mat}_{n\times n}(\mathbb{R})$, entry by entry, the matrix exponential e^A is an absolutely convergent series.
Solution. For each $M \in \operatorname{Mat}_{n\times n}(\mathbb{R})$ set $\|M\| := \sup\{|Mv| : |v| \leq 1\}$, where $|\cdot|$ is the Euclidean norm in $\mathbb{R}^n$. Then

$$\left\| \sum_{k=N_1}^{N_2} \frac{1}{k!}A^k \right\| \leq \sum_{k=N_1}^{N_2} \frac{1}{k!}\|A^k\| \leq \sum_{k=N_1}^{N_2} \frac{1}{k!}\|A\|^k \overset{N_1,N_2\to\infty}{\longrightarrow} 0.$$

Exercise 5.8.42 Let $A, B \in \mathfrak{gl}(n\mathbb{R})$. 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 5.8.43 For every matrix A, the matrix e^A is invertible.
Solution: Use Exercise 5.8.42 and get $e^A e^{-A} = e^0 = I$.

Exercise 5.8.44 Calculate the exponential of the matrix $t\begin{bmatrix} 0 & a & c \\ 0 & 0 & b \\ 0 & 0 & 0 \end{bmatrix} \in \mathfrak{gl}(3, \mathbb{R})$.

Exercise 5.8.45 For each integer $n > 1$, there are $A, B \in \mathfrak{gl}(n, \mathbb{R})$ such that $e^{A+B} \neq e^A e^B \neq e^B e^A$. Compare with Exercise 5.8.42.

Exercise 5.8.46 Given a square matrix A and an invertible matrix B of the same size, we have $e^{BAB^{-1}} = Be^A B^{-1}$.

Exercise 5.8.47 (Trace as a Lie Algebra Homomorphism) The determinant function $\det : \mathrm{GL}(n, \mathbb{R}) \to (\mathbb{R}^*, \cdot)$ is a Lie group homomorphism, the trace function $\mathrm{tr} : \mathfrak{gl}(n, \mathbb{R}) \to (\mathbb{R}, +)$ is a Lie algebra homomorphism, and

$$\det(e^A) = e^{\mathrm{tr(A)}}.$$

Solution. Given a square matrix A, there is an invertible matrix B such that $\tilde{A} = BAB^{-1}$ is upper triangular, i.e., of the form

$$\tilde{A} = \begin{bmatrix} \alpha_1 & * & * & * \\ 0 & \alpha_2 & * & * \\ 0 & 0 & \ddots & \vdots \\ 0 & 0 & \dots & \alpha_n \end{bmatrix}.$$

For such matrices, we have

$$\tilde{A}^k = \begin{bmatrix} \alpha_1^k & * & * & * \\ 0 & \alpha_2^k & * & * \\ 0 & 0 & \ddots & \vdots \\ 0 & 0 & \dots & \alpha_n^k \end{bmatrix} \quad \text{and} \quad e^{\tilde{A}} = \begin{bmatrix} e^{\alpha_1} & * & * & * \\ 0 & e^{\alpha_2} & * & * \\ 0 & 0 & \ddots & \vdots \\ 0 & 0 & \dots & e^{\alpha_n} \end{bmatrix}.$$

Finally, using Formula 5.5.8.iii we conclude

$$\begin{aligned}
\det(e^A) &= \det(Be^A B^{-1}) \\
&= \det(e^{BAB^{-1}}) \\
&= \det(e^{\tilde{A}}) \\
&= e^{\alpha_1} \cdots e^{\alpha_n} \\
&= e^{\sum_{i=1}^n \alpha_i} \\
&= e^{\mathrm{tr}(\tilde{A})} \\
&= e^{\mathrm{tr(BAB^{-1})}} = e^{\mathrm{tr(A)}}.
\end{aligned}$$

Exercise 5.8.48 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 5.8.49 (Differential of Exponential Map) Let G be a Lie group and $X \in T_{1_G} G$. Then

$$(\mathrm{d}\exp)_X = (\mathrm{d}L_{\exp(X)})_{1_G} \circ \frac{\mathrm{Id} - e^{-\mathrm{ad}_X}}{\mathrm{ad}_X}.$$

Here, if $A \in \mathrm{End}(\mathfrak{g})$, we have $\frac{\mathrm{Id} - e^{-A}}{A} = \sum_{k=0}^{\infty} \frac{(-A)^k}{(k+1)!}$.

Exercise 5.8.50 One deduces 5.6.3.i from Exercise 5.8.14, the chain rule, and Theorem 5.1.5.

Exercise 5.8.51 The space $\mathrm{Der}(\mathfrak{g})$ is a Lie subalgebra of $\mathfrak{gl}(\mathfrak{g})$.
Solution. The set $\mathrm{Der}(\mathfrak{g})$ is a linear subspace of $\mathfrak{gl}(\mathfrak{g})$, because Eq. (5.16) is linear in D. 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

$$\begin{aligned}
[D, D']([X, Y]) &= (D \circ D')([X, Y]) - (D' \circ D)([X, Y]) \\
&= D([D'X, Y] + [X, D'Y]) - D'([DX, Y] + [X, DY]) \\
&= [DD'X, Y] + [D'(X), DY] + [DX, D'(Y)] + [X, DD'Y] \\
&\quad - \big([D'DX, Y] + [D(X), D'Y] + [D'X, DY]) + [X, D'DY])\big) \\
&= [(DD' - D'D)X, Y] + [X, (DD' - D'D)Y] \\
&= [[D, D']X, Y] + [X, [D, D']Y].
\end{aligned}$$

Exercise 5.8.52 Let $\mathfrak{g}$ be a Lie algebra. The map ad_X is a derivation on $\mathfrak{g}$ (because of Jacobi identity) and the map $X \mapsto \mathrm{ad}_X$ is a Lie algebra homomorphism of $\mathfrak{g}$ into $\mathrm{Der}(\mathfrak{g})$.

Exercise 5.8.53 Let G be a group, on which we denote by C_g the conjugation by an element $g \in G$. The map $g \mapsto C_g$ is a group action of G by automorphisms of G.

Exercise 5.8.54 Let G be a Lie group. For all $g \in G$ we have $\mathrm{Ad}_g \in \mathrm{Aut}_{\mathrm{Lie}}(\mathfrak{g})$ and the map $g \mapsto \mathrm{Ad}_g$ is a Lie group homomorphism of G into $\mathrm{Aut}_{\mathrm{Lie}}(\mathfrak{g})$.

Exercise 5.8.55 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 5.8.56 The previous exercises imply Proposition 5.6.3.iii.

Exercise 5.8.57 Let G be a Lie group with Lie algebra $\mathfrak{g}$. We denote by $Z(G) := \{g \in G : hg = gh, \forall h \in G\}$ the *center* of G and by $Z(\mathfrak{g}) := \{X \in \mathfrak{g} : [X, Y] = 0, \forall Y \in \mathfrak{g}\}$ the *center* of $\mathfrak{g}$. If G is connected, then $Z(\mathfrak{g}) = \mathrm{Lie}(Z(G))$.

Chapter 6
Metric Groups and Homogeneous Spaces

In differential geometry, the term *homogeneous space* refers to the quotient space of a Lie group modulo a closed subgroup, which results in a manifold with a smooth transitive action of the Lie group. Instead, in metric geometry and, more generally, in analysis on metric spaces, the term *homogeneous* often refers to functions that, when precomposed with dilations of the space, scale by certain constants; see Exercise 6.6.12.

At this point, it is necessary to distinguish between the terms. If a metric space admits a transitive action by isometries, we refer to it as an isometrically homogeneous space; see Sect. 6.2. If, instead, a space admits a self-map that non-trivially dilates the distance function, we refer to it as a self-similar space; see Sect. 6.5.

In this chapter, we will see that every isometrically homogeneous space with mild topological assumptions has the structure of a Lie homogeneous space (in the standard sense of Sect. 6.1). In Sect. 10.2, we will see that if, in addition, the space is self-similar, then it has the structure of a metric group in the sense of Sect. 6.3.

6.1 Lie Homogeneous Spaces

In this section, we review group actions and their quotients.

6.1.1 The General Viewpoint of Group Actions

The group of isometries of a metric space naturally acts on the space itself, and from this action, it inherits a topological structure. It is convenient to begin our discussions with the general notion of group action. Group actions on sets were

E. Le Donne, *Metric Lie Groups*, Graduate Texts in Mathematics 306,
https://doi.org/10.1007/978-3-031-98832-5_6

mentioned in (5.14). When, however, the sets and the groups have extra structures, it is natural to take these into account for the considered actions.

A *continuous action* of a topological group G on a topological space X, generically denoted by $G \curvearrowright X$, is a continuous map $G \times X \to X$, usually denoted by $(g, x) \mapsto g.x$, that obeys the *associativity law* and the *identity law*:

$$(gh).x = g.(h.x) \qquad \text{and} \qquad 1_G.x = x, \quad \forall g, h \in G, \forall x \in X.$$

Definition 6.1.1 (Special Types of Actions) Fix an action $G \curvearrowright X$ of a group on a set. The action is *transitive* if for every $x, y \in X$ there is $g \in G$ such that $g.x = y$. It is *faithful* if for every distinct elements $g, h \in G$ there is $x \in X$ such that $g.x \neq h.x$. Faithful actions are also known as *effective* actions. The action is *free* if for all $x \in X$ and all $g \in G \setminus \{1_G\}$ one has $g.x \neq x$. Next, assume that the action $G \times X \to X$ is a continuous action of a topological group on a topological space. The action is *proper* if the associated map

$$G \times X \to X \times X \qquad (g, p) \mapsto (g.p, p)$$

is a proper map (i.e., inverse images of compact sets are compact); see also Exercise 6.6.1 for alternative definitions. The action is *properly discontinuous* if for all $x, y \in X$ there exist a neighbourhood U_x of x and a neighbourhood U_y of y such that the set $\{g \in G \mid (g.U_x) \cap U_y \neq \emptyset\}$ is finite.

Given an action $G \curvearrowright X$ and an element $x \in X$, the *orbit* of x is defined as

$$G.x := \{g.x \mid g \in G\}.$$

Either two orbits $G.x$ and $G.x'$ coincide, or they are disjoint sets. Therefore, we have a well defined *quotient space* denoted by ${}_G\backslash^X$:

$${}_G\backslash^X := \{G.x \mid x \in X\}.$$

We may denote the map $x \in X \mapsto G.x \in {}_G\backslash^X$ by π and call it the *quotient map*. While, fixing a point $x_0 \in X$, we call the map $g \in G \mapsto g.x_0 \in X$ the *orbit map*.

Given an action $G \curvearrowright X$ and an element $x \in X$, the *stabilizer subgroup* of G at x, also called the *isotropy subgroup* at x, is the subgroup $G_x := \{g \in G : g.x = x\}$.

Example 6.1.2 There is a natural action of each group onto itself by left translations; see Exercise 6.6.4. More generally, every subgroup H of a group G acts on the whole group. Actually, there are two possible natural actions, as we now review. The map

$$\begin{aligned} H \times G &\to G \\ (h, g) &\mapsto hg \end{aligned}$$

is an action of H on G, called the *action by left translations* of H on G. Whereas, the map

$$
\begin{aligned}
H \times G &\to G \\
(h, g) &\mapsto gh^{-1}
\end{aligned}
$$

is an action of H on G, called the *action by right translations* of H on G. We observe that when $H \curvearrowright G$ by right translations, then ${}_{H}\backslash^{G} = G/H := \{gH \mid g \in G\}$. Moreover, the map $(g, g'H) \mapsto gg'H$ defines an action of G on G/H.

If $G \curvearrowright X$ is a continuous action of a topological group on a topological space, then there is a unique topology, which we call the *quotient topology*, on ${}_{G}\backslash^{X}$ that makes the quotient map continuous and open; see Exercise 6.6.2.

The following basic result gives a natural homeomorphism between orbits and groups modulo stabilizers. The general argument goes back to Arens, [Are46]; see also [DtK18, Lemma 5.38] for another version. In the following theorem, we shall begin to see the importance of dealing with spaces and groups that are locally compact (and are second-countable; see Exercise 6.6.15).

Theorem 6.1.3 *Let X be a locally compact Hausdorff space. Let G be a locally compact group with a countable basis. Let $G \curvearrowright X$ be a continuous action. Let $x \in X$. Then, the stabilizer subgroup G_x at x is closed. Moreover, if the action is transitive, then the map $gG_x \mapsto g \cdot x$ is a homeomorphism between G/G_x and X.*

Proof Since X is a Hausdorff space, the set $\{x\}$ is closed. Since the orbit map $g \mapsto g.x$ is continuous and G_x is the preimage of x under such a map, we deduce that G_x is closed in G. Next, we additionally assume that the action is transitive, so the orbit map is surjective. We then consider the following commutative diagram of surjective maps:

$$
\begin{array}{ccc}
G & \xrightarrow{\;g \mapsto g.x\;} & \\
\downarrow & & \searrow \\
G/G_x & \xrightarrow{\;gG_x \mapsto g.x\;} & X
\end{array}
$$

Notice that the bottom map is a bijection. Recalling that the quotient map $G \to G/G_x$ is open and continuous, we stress that it is enough to prove that the orbit map is open. Let V be an open subset of G and g a point in V. Since G is assumed locally compact, select a compact neighborhood U of 1_G in G such that $U = U^{-1}$ and $gU^2 \subseteq V$. Since G is assumed to have a countable basis, there exists a sequence $(g_n)_{n\in\mathbb{N}} \subseteq G$ such that $G = \bigcup_{n\in\mathbb{N}} g_n U$. The group action being transitive implies $X = \bigcup_{n\in\mathbb{N}} g_n U x$. Each summand is compact; hence, it is a closed subset of X. Recall that the Baire Category Theorem holds for locally compact Hausdorff spaces; see [NB11, Theorem 11.7.3, p.394] (or look at the general argument in [Fol99, p.161]). By such a theorem, some summand, and therefore $U.x$, contains an inner point $u.x$ with $u \in U$. Then x is an inner point of $u^{-1}U.x \subseteq U^2.x$ and consequently $g.x$ is an inner point of $V.x$. This shows that the orbit map is open. □

For Lie groups acting on manifolds, we now consider actions that take into account the differential structures of the group and the manifold.

Definition 6.1.4 (Lie Action) Let G be a Lie group and M a smooth manifold. A *Lie action* of G on M is an action $G \times M \to M$ that is a smooth map.

There is a natural Lie action of every Lie group onto itself by left translations; see Exercise 6.6.4. Likewise, every Lie subgroup H of a Lie group G gives Lie actions $H \curvearrowright G$, as in Example 6.1.2.

When a Lie action is proper and free, the quotient space is naturally a manifold; see Theorem 6.1.5. We shall not prove this theorem in this generality; we shall just refer to [Lee13, Theorem 21.10]. We shall prove the respective result in the case of actions of closed subgroups of Lie groups; see Theorem 6.1.6.

Theorem 6.1.5 (Quotient Manifold Theorem) *Let $G \curvearrowright M$ be a proper and free Lie action of a Lie group G on a manifold M. There exists a unique differentiable structure on ${}_{G}\backslash^{M}$ such that $M \to {}_{G}\backslash^{M}$ is a smooth map with surjective differential.*

6.1.2 Lie Coset Spaces

Let G be a Lie group and H a closed subgroup of G. We will consider manifold structures for the space

$$G/H := \{gH \mid g \in G\}\,.$$

When the set G/H is equipped with the differentiable structure from the following theorem, it is called a *homogeneous manifold*, or *Lie homogeneous manifold*, or *Lie coset space*.

Theorem 6.1.6 *Let G be a Lie group and $H < G$ a closed subgroup. Then, the topological space G/H admits the structure of a differentiable manifold. Moreover, there is a unique manifold structure for which the action $G \times G/H \to G/H$ is smooth.*

Sketch of the Proof Complete proofs can be found in [War83, Theorem 3.58] and in [Hel01, Chapter II, Theorem 4.2]. We begin by stressing that G/H is a second-countable Hausdorff space; see Exercises 6.6.10 and 6.6.11. We shall find an atlas for G/H by constructing *local cross sections* in G.

Recalling that H is a Lie group by Theorem 5.3.4, let $\mathfrak{g}$ and $\mathfrak{h}$ be the Lie algebras of G and H, respectively. Let $\mathfrak{m}$ be some vector subspace of $\mathfrak{g}$ that is in direct sum with $\mathfrak{h}$, that is, $\mathfrak{g} = \mathfrak{m} \oplus \mathfrak{h}$. Because of Proposition 5.2.8, see also [Hel01, Chapter II, Lemmas 2.4–2.5], there are open neighbourhoods $U_{\mathfrak{m}} \subset \mathfrak{m}$ of $0_{\mathfrak{m}}$ and $U_{\mathfrak{h}} \subset \mathfrak{h}$ of $0_{\mathfrak{h}}$

such that the map $\phi : \mathfrak{m} \oplus \mathfrak{h} \to G$, $\phi(x, y) := \exp(x)\exp(y)$, is a diffeomorphism between $U_{\mathfrak{m}} \times U_{\mathfrak{h}}$ and its image $U_G \subset G$, and such that $H \cap U_G = \exp(U_{\mathfrak{h}})$ (the latter claim uses the hypothesis that H is closed).

For each $g \in G$, define the map $\psi_g : U_{\mathfrak{m}} \to G/H$, $\psi_g(x) := \pi(g\exp(x))$, where $\pi(g) := gH$. One can show that, up to shrinking $U_{\mathfrak{m}}$, the map $\psi_e : U_{\mathfrak{m}} \to G/H$ is injective onto $\pi(U_G)$, and so ψ_g is a homeomorphism $U_{\mathfrak{m}} \to \pi(gU_G)$ for all $g \in G$. We have obtained the candidate atlas:

$$\{(\pi(gU_G), \psi_g^{-1})\}_{g \in G} \tag{6.1}$$

To check that each composition $\psi_{g_2}^{-1} \circ \psi_{g_1}$ is smooth on the appropriate domain, we locally write such a map as the composition $\pi_{\mathfrak{m}} \circ \phi^{-1} \circ R_h \circ L_{g_2^{-1}g_1} \circ \phi$, where $\pi_{\mathfrak{m}} : \mathfrak{m} \oplus \mathfrak{h} \to \mathfrak{m}$ is the projection modulo $\mathfrak{h}$, and $h \in H$ is a suitable element.

We stress that, with this differentiable structure (6.1), each map ψ_g^{-1} is a smooth section of the projection π. Consequently, the action $G \times G/H \to G/H$ is smooth. Also the uniqueness is an easy consequence. For these last details, we refer to [War83, page 122]. □

6.2 Isometrically Homogeneous Spaces

Definition 6.2.1 We say that a metric space M is *isometrically homogeneous* if its group of isometries acts on the space transitively. Explicitly, this means that, for every $p, q \in M$, there exists a distance-preserving homeomorphism $f : M \to M$ such that $f(p) = q$. An *isometrically homogeneous space* is a metric space that is isometrically homogeneous.

Examples of isometrically homogeneous spaces are provided by groups equipped with left-invariant distances, which we call *metric groups*. This whole book is devoted to studying various types of metric groups.

From the general viewpoint, we shall study isometrically homogeneous spaces with mild topological assumptions. One of them is local compactness. Other topological assumptions are connectedness and local connectedness; see [Mun75].

If M is a metric space, we consider its isometry group $\mathrm{Isom}(M)$, that is, the set of self-isometries of M equipped with the composition rule and the pointwise-convergence topology. We stress that in the isometry group, the compact-open topology and the pointwise-convergence topology coincide, and the topological group $\mathrm{Isom}(M)$ acts continuously on M; see Remark 6.2.6.

The main aim of this section is to show that isometry groups of isometrically homogeneous spaces (with mild topological assumptions) are Lie groups; see Theorem 6.2.10. Such a fact is a consequence of the solution of Hilbert's fifth problem by Montgomery–Zippin and Gleason, together with the observation that isometry groups are second-countable and locally compact. This latter property follows the Ascoli–Arzelà Theorem (Exercise 3.4.12). Regarding the axioms of countability, see Exercise 6.6.15.

6.2.1 Transitive Actions by Locally Compact Groups and Hilbert's Fifth Problem

We next state one of the results that follow from Gleason-Montgomery-Zippin's theory, as it is stated in Montgomery-Zippin's book [MZ74, Corollary on page 243, Section 6.3]: *If a (separable) locally compact group G satisfying Property A* (which we soon recall) *acts effectively and transitively on a locally compact, connected, (locally connected), and finite-dimensional space, then G is a Lie group.* We added the words in parentheses because we think the authors were implicitly assuming them. Moreover, in the language of Montgomery-Zippin [MZ74, Section 6.2], *Property A* means that for every neighborhood V of the identity element in G, there exists a compact subgroup K of G such that $K \subset V$ and G/K, equipped with the quotient topology, is a Lie group. In addition, we clarify that as dimension of a metric space, we are considering the topological dimension, also called *covering dimension*, as originally introduced by Lebesgue. For the definition of the dimension of a topological space, we refer to [Mun75, p. 305].

A clean argument for the above statement is not present in the literature. Since we want to present a proof of the Lie structure of locally compact groups transitively acting on 'nice' spaces, we shall then use a stronger and more established version of the solution of Hilbert's fifth problem, attributed to Gleason and Yamabe.

Theorem 6.2.2 (Gleason-Yamabe) *Let G be a first-countable locally compact group. Then, there exists an open subgroup $G' < G$ that is the inverse limit of a sequence of Lie groups.*

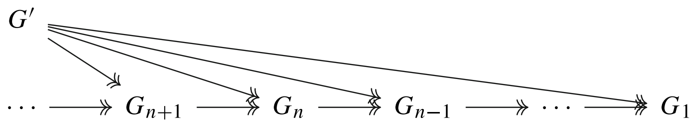
When using Theorem 6.2.2 in the proof of Theorem 6.2.3, we shall rephrase the concept of an inverse limit. We shall use the notion of inverse limit of topological groups and of topological spaces; as a reference, see [Tao14, Definition 4.2.5]. We shall not prove Theorem 6.2.2 but suggest to look at the book by Tao [Tao14, Chapter 6]. We stress that in [Tao14, paragraph below Theorem 6.0.11], there is the requirement of G being Hausdorff, but, for us, this is part of the definition of a topological group. Before drawing consequences of interest to isometrically homogeneous spaces in Theorem 6.2.10, we first give a proof of the following general result on continuous, faithful, and transitive actions.

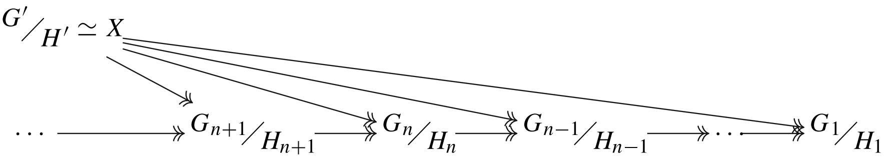
Theorem 6.2.3 (After Gleason-Montgomery-Yamabe-Zippin) *Let G be a topological group that is locally compact and second-countable. Let X be a Hausdorff topological space that is connected, locally connected, locally compact, and finite-dimensional. Assume that G acts on X continuously, faithfully, and transitively. Then G is a Lie group.*

Proof A proof in the case X is a manifold (and G is σ-compact, and not assumed second-countable) can be found in [Tao14, Section 6.4]. We present an argument following the same lines. Pick a point $x_0 \in X$ and let $H := G_{x_0}$ be the stabilizer of G at x_0. Observe that $H < G$ is a closed subgroup, that G/H is a Hausdorff topological space, and that the map $\phi : G/H \to X$, $\phi(gH) := gx_0$, is a homeomorphism (this is true because G is locally compact and second-countable and X is locally compact and Hausdorff, and the action is transitive; see Theorem 6.1.3.)

We apply the Gleason-Yamabe result, Theorem 6.2.2, recalling Exercise 6.6.15.i. Thus, there is a subgroup $G' < G$ and Lie groups G_n, for $n \in \mathbb{N}$, such that G' is open in G and $G' = \varprojlim G_n$. The fact that G' is the inverse limit of the sequence G_n in the category of topological groups rephrases as saying that there exist compact normal subgroups $K_n \lhd G'$ such that $G_n = G'/K_n$ is a Lie group, for each $n \in \mathbb{N}$, and $K_n \searrow \{1\}$ as $n \to \infty$ (i.e., $K_n \supset K_{n+1}$ and $\bigcap_{n\in\mathbb{N}} K_n = \{1\}$). By Exercise 6.6.17, since X is connected and G' open, then G' acts on X transitively.

Set H' to be the stabilizer of G' at the initial point x_0 and $H_n := H'/K_n < G'/K_n = G_n$. We have two inverse limits: one in the category of topological groups and one in the category of topological spaces. They are expressed by the following diagrams.





Notice that each H_n is a closed subgroup of the Lie group G_n. Hence, the quotient G_n/H_n is a smooth manifold. Moreover, the space X is the inverse limit of G_n/H_n in the category of topological spaces. In fact, the compact group K_n acts continuously on X with quotient map

$$\pi_n : X \to X_n := G_n/H_n.$$

For every $m \geq n$, the compact Lie group K_n/K_m smoothly acts on X_m with quotient map $\pi_{m,n} : X_m \to X_n$.

Since, by assumption, the dimension of X is finite, then the increasing sequence of dimensions $\dim(G_n/H_n)$ must stabilize; see [Tao14, Exercise 6.4.1]. Note that the structure group of the quotient

$$G_{n+1}/H_{n+1} \longrightarrow\!\!\!\!\!\rightarrow G_n/H_n \tag{6.2}$$

is K_n/K_{n+1}. By Exercise 6.6.5, since the dimensions stabilize, then eventually K_n/K_{n+1} is zero-dimensional and compact, and hence it is finite. Hence, this projection map is a covering map between manifolds, whose covering multiplicity is the cardinality $\#K_n/K_{n+1}$.

We claim that the projections in (6.2) eventually stabilize as homeomorphisms, which is equivalent to saying that eventually K_n/K_{n+1} is trivial. To show this, we shall use that $G'/H' \simeq X$ is assumed locally connected.

Let $\bar{n}$ be such that, for every $m \geq n \geq \bar{n}$, the group K_n/K_m is finite and thus the projections $\pi_{m,n} : X_m \to X_n$ are covering maps. Fix a simply connected neighbourhood $U_{\bar{n}}$ of $\pi_{\bar{n}}(x_0)$ in $X_{\bar{n}}$. For $m \geq \bar{n}$, define $U_m := \pi_{m,\bar{n}}^{-1}(U_{\bar{n}})$. Notice that U_m is a disjoint union of connected components, each homeomorphic to $U_{\bar{n}}$ via $\pi_{m,\bar{n}}$. Since X is locally connected, there is a connected neighbourhood $V \subset \pi_{\bar{n}}^{-1}(U_{\bar{n}})$ of x_0 in X. Consider the orbit map $\Phi : G' \to X$. Choose $\hat{n} \geq \bar{n}$ such that $K_{\hat{n}} \subset \Phi^{-1}(V) \subset G'$, so that $K_{\hat{n}}x_0 \subset V$.

We then check that $K_{\hat{n}} = K_m$ for every $m \geq \hat{n}$. To this aim, notice that the orbit $\Omega := (K_{\hat{n}}/K_m)(\pi_m(x_0))$ is a finite set with cardinality $\#(K_{\hat{n}}/K_m)$. On the one hand, every two points in Ω belong to different connected components of U_m. On the other hand, since $\Omega = \pi_m(K_{\hat{n}}x_0) \subset \pi_m(V)$ and since $\pi_m(V)$ is a connected subset of U_m, the set Ω is contained in one connected component of U_m. We conclude that $\#(K_{\hat{n}}/K_m) = 1$, i.e., $K_{\hat{n}} = K_m$.

Therefore, the quotient K_n/K_{n+1} is eventually trivial, which implies that the inverse limit $\varprojlim G_n$ stabilizes and G' is its stabilization. Thus, we deduce that G' is a Lie group, and G is a Lie group since G' is open in G. □

6.2.2 *Properties of Isometrically Homogeneous Spaces*

6.2.2.1 Two Preliminary Observations

We begin with a first topological property of locally compact isometrically homogeneous spaces: completeness.

Lemma 6.2.4 *Every locally compact isometrically homogeneous space is complete.*

Proof Pick a point $\bar{x}$ in a locally compact isometrically homogeneous space M. By local compactness, there is $\bar{r} > 0$ such that the closed ball $\bar{B}(\bar{x}, \bar{r})$ is compact. Now, take a Cauchy sequence $(x_n)_{n\in\mathbb{N}}$ in M. For $N \in \mathbb{N}$ large enough we have $d(x_n, x_m) < \bar{r}$, for all $n, m \geq N$. Thus the sequence $(x_n)_{n>N}$ is in the set $\bar{B}(x_N, \bar{r})$. This latter set is compact by isometric homogeneity. Thus $(x_n)_{n>N}$ has a convergent subsequence. □

The isometry group is equipped with the compact-open topology—review this notion in [Mun75, p.285]. It is a closed subset of the space of homeomorphisms; see Exercise 6.6.16. We shall see that it forms a locally compact group. However, this is not immediately clear from Ascoli–Arzelá theorem because it may happen that a locally compact isometrically homogeneous space M is not boundedly compact: even if small balls are compact, some large balls may not be compact. For example, this is the case for the distance $\min\{d_E, 1\}$ on $\mathbb{R}$, where d_E denotes the Euclidean distance. We shall solve this issue using the following result.

Lemma 6.2.5 (Reduction to Proper Distances) *If (M, d) is a connected locally compact isometrically homogeneous space, then there exists a finite-valued distance ρ inducing the same topology such that (M, ρ) is an isometrically homogeneous space that is boundedly compact and* $\mathrm{Isom}(M, d)$ *is a closed subgroup of* $\mathrm{Isom}(M, \rho)$.

Proof Fix a point $o \in M$. Since the metric is locally compact, there exists some $r_0 > 0$ such that the closed ball $\bar{B}_d(o, r_0)$ is compact. Notice that, by isometric homogeneity, every other r_0-ball $\bar{B}_d(p, r_0)$, with $p \in M$, is compact. Then, for each $p, q \in M$, we consider the value

$$\rho(p, q) := \inf\left\{ \sum_{i=1}^{k} d(p_{i-1}, p_i) \ : \right.$$
$$\left. k \in \mathbb{N},\ p_i \in M,\ p_0 = p,\ p_k = q,\ d(p_{i-1}, p_i) \le r_0/4 \right\}.$$

First, we observe that this is a distance function that gives the same topology, since for all $r \in (0, r_0/4)$, we have that $\rho(p, q) < r$ if and only if $d(p, q) < r$, as one can check by triangle inequality and considering the last condition defining ρ. Consequently, this distance must be finite-valued since, otherwise, the connected components of points at finite distance would disconnect the space, which is assumed connected. Second, because the construction is done intrinsically, we have $\mathrm{Isom}(M, d) \subseteq \mathrm{Isom}(M, \rho)$. Regarding its closedness, recall Exercise 6.6.16. Finally, we claim that for all $r > 0$, the set $\bar{B}_\rho(o, r)$ is compact. We prove by induction on $n \in \mathbb{N}$ that $\bar{B}_\rho(o, nr_0/8)$ is compact. The base of the induction is $n = 0$, and it is clear since the ball reduces to a point. Assume that $\bar{B}_\rho(o, nr_0/8)$ is compact. Hence, it is a totally bounded set: there exists a finite set $Y_n \subseteq \bar{B}_\rho(o, nr_0/8)$ such that

$$\bar{B}_\rho(o, nr_0/8) \subseteq \bigcup_{y \in Y_n} B_d(y, r_0/2). \tag{6.3}$$

We claim that

$$\bar{B}_\rho(o, (n+1)r_0/8) \subseteq \bigcup_{y \in Y_n} \bar{B}_d(y, r_0),$$

from which we will deduce that $\bar{B}_\rho(o, (n+1)r_0/8)$ is compact. To prove the claim, take $p \in \bar{B}_\rho(o, (n+1)r_0/8)$. By the definition of ρ, there exists $\bar{p} \in \bar{B}_\rho(o, nr_0/8)$ such that $\rho(\bar{p}, p) < r_0/2$ (in fact, one can take points $p_0 = p, p_1, \ldots, p_k = o$ such that $\sum_{i=1}^k d(p_{i-1}, p_i) < (n+2)r_0/8$ and $d(p_{i-1}, p_i) \leq r_0/4$; then a possible $\bar{p}$ is the first point among $p_0, p_1, \ldots, p_k$ such that $\rho(\bar{p}, o) \leq nr_0/8$). By (6.3), there is some $y \in Y_n$ such that $d(\bar{p}, y) < r_0/2$. Then, by triangle inequality, we infer $d(y, p) < r_0$. Thus, the claim is proved. □

6.2.2.2 Transitive Isometry Groups

We consider groups that act on metric spaces transitively and by isometries. In particular, we will clarify why the isometry group of a connected locally compact isometrically homogeneous space is locally compact.

Remark 6.2.6 We stress that for every metric space M, the compact-open topology, the topology of uniform convergence on compact sets, and the topology of pointwise convergence agree on the isometry group $\mathrm{Isom}(M)$ of M. Moreover, the group $\mathrm{Isom}(M, d)$ becomes a topological group continuously acting on M. For a reference to the fact that these topologies agree on $\mathrm{Isom}(M, d)$; see [CH16, Lemmas 5.B.1 and 5.B.2] or [Kel75, p.232, Theorem 15]. The fact that this structure makes the isometry group a topological group is well known; van Dantzig and van der Waerden [DW28] showed this in the case where M is connected, locally compact, and separable, and a general proof for metric spaces can be found in [CH16, Lemma 5.B.3]. The fact that the action is continuous is an immediate consequence of the definition of compact-open topology.

A well-known consequence of Ascoli–Arzelá argument, see for example [CH16, Lemma 5.B.4], is that if a metric space is boundedly compact, then the isometry group is locally compact; see also Exercise 6.6.18. We record this property for further reference.

Proposition 6.2.7 *Let M be a boundedly compact metric space. Then* $\mathrm{Isom}(M)$ *equipped with the equivalent topologies, compact-open or pointwise convergence, is a second-countable, σ-compact, locally compact group acting on M continuously and properly. For every $o \in M$ the stabilizer* $\mathrm{Stab}(o)$ *of* $\mathrm{Isom}(M)$ *at o is compact. If, in addition, the space M is isometrically homogeneous, the orbit map ϕ :* $\mathrm{Isom}(M) \to M$, *$\phi(f) := f(o)$, induces a homeomorphism between topological spaces*

$$\mathrm{Isom}(M)\big/\mathrm{Stab}(o) \longrightarrow M.$$

Proof Most of this standard fact is an exercise and can be mostly found in [CH16, Proposition 5.B.5]; see Remark 6.2.6. However, the result is just a consequence of the Ascoli-Arzelà argument (Exercise 3.4.12). The homeomorphism is the one discussed in Theorem 6.1.3. □

Remark 6.2.8 In the case M is a connected locally compact isometrically homogeneous space, then the conclusion of Proposition 6.2.7 is still valid. Indeed, if d denotes the distance function of M, then by Lemma 6.2.5 there is a proper distance ρ, i.e., (M, ρ) is boundedly compact, for which $\mathrm{Isom}(M, d)$ is a closed subgroup of $\mathrm{Isom}(M, \rho)$. Clearly, also the stabilizer of $\mathrm{Isom}(M, d)$ at a point $o \in M$ is a closed subgroup of the stabilizer at o of $\mathrm{Isom}(M, \rho)$. Because we have closed subgroups, we get the same conclusions about $\mathrm{Isom}(M, d)$ and its action from Proposition 6.2.7.

As in the proof of Theorem 6.2.3, it is essential to know when identity components of transitive groups still act transitively. This is the case for isometry groups of connected and locally compact spaces. The following result will be used a few times for the characterization of isometrically homogeneous spaces in Sect. 6.5.

Proposition 6.2.9 *Suppose that M is an isometrically homogeneous space and that M is connected and locally compact. Then, every open subgroup of* $\mathrm{Isom}(M)$ *acts transitively on M. If, in addition, the topological group* $\mathrm{Isom}(M)$ *has the structure of a Lie group, then the identity component* $\mathrm{Isom}(M)^\circ$ *of* $\mathrm{Isom}(M)$ *acts transitively on M.*

Proof We shall apply Theorem 6.1.3. Being a metric space, the space M is Hausdorff. Regarding $\mathrm{Isom}(M)$, we showed that it is a locally compact group and has a countable basis; see Remark 6.2.6 and Exercise 6.6.18.

Let H be an open subgroup of $\mathrm{Isom}(M)$. First, we claim that for every $q \in M$, the orbit $H \cdot q$ of q under H is open. Indeed, this is because, defining $G := \mathrm{Isom}(M)$ and G_p the stabilizer subgroup of G at p, the projection $G \to G/G_p$ is open and the orbit action $G/G_p \to M$ is a homeomorphism by Theorem 6.1.3. Next, fix a point $p \in M$, suppose by contradiction that $H \cdot p \neq M$. Hence,

$$M = (H \cdot p) \bigsqcup \left(\bigcup_{q \notin H \cdot p} H \cdot q \right)$$

is a disjoint union of two non-empty open sets of M. This contradicts the fact that M is connected. Therefore, the subgroup H acts transitively. For the last part of the statement, recall that connected components of manifolds are open subsets. □

We are ready to explain the Lie group structure of isometry groups of isometrically homogeneous spaces with mild topological assumptions.

Theorem 6.2.10 (After Gleason-Montgomery-Yamabe-Zippin) *Let M be a metric space that is connected, locally connected, locally compact, and has finite topological dimension. Assume that the isometry group* $\mathrm{Isom}(M)$ *of M acts transitively on M. Then* $\mathrm{Isom}(M)$ *has the structure of a Lie group with finitely many connected components, and M has the structure of a smooth manifold for which the action* $\mathrm{Isom}(M) \curvearrowright M$ *is smooth.*

Proof We shall apply Theorem 6.2.3. For another proof using the earlier works of Gleason, Montgomery, and Zippin, [MZ52, Gle52], see [MZ74] or [DtK18, Chapter 16]. By Proposition 6.2.7, together with Remark 6.2.8, the topological group $\mathrm{Isom}(M)$ is locally compact and second-countable, and it is acting continuously. Obviously, the group $\mathrm{Isom}(M)$ acts effectively. It acts transitively by assumption. Theorem 6.2.3 implies that $\mathrm{Isom}(M)$ is a Lie group, with finitely many components; see Exercise 6.6.19.

By Proposition 6.2.7, and Remark 6.2.8, each stabilizer $\mathrm{Stab}(o)$, with $o \in M$, is compact and hence a closed Lie subgroup. Consequently, by Theorem 6.1.3 the metric space M is homeomorphic to $\mathrm{Isom}(M)/\mathrm{Stab}(o)$, which is a manifold on which $\mathrm{Isom}(M)$ acts smoothly, by Theorem 6.1.6. □

6.3 Metric Groups

In this section, we focus on metric groups. Most of the discussion extends to quotients: isometrically homogeneous spaces. For this more general viewpoint, we refer to [LO16, Cow+24].

We generally refer to groups equipped with left-invariant distance functions as *metric groups*. With the term *metric Lie group*, we mean a Lie group equipped with a left-invariant distance function that induces the manifold topology. In general, when we have a topological space and we equip it with a distance function, we say that the distance function is *admissible* if it induces the topology of the space.

6.3.1 Smoothness of Isometries Between Metric Lie Groups

With the use of the results by Gleason-Montgomery-Yamabe-Zippin, Theorem 6.2.10, we deduce the differentiable regularity of isometries between metric Lie groups.

Theorem 6.3.1 *Isometries between metric Lie groups are smooth maps.*

Before giving the proof, we remark that in the Riemannian setting, the classical result of Myers and Steenrod gives smoothness of isometries; see [MS39], and more generally [CL16a] for sub-Riemannian manifolds. However, the following proof is different in spirit and, nonetheless, it will imply (see Theorem 8.2.1) that such metric isometries are Riemannian isometries for some Riemannian structures.

In the proof of Theorem 6.3.1, we will obtain a continuous isomorphism of Lie groups; hence, from what we saw in Sect. 5.3.1, this isomorphism is a smooth map. One can actually discuss analytic structures on Lie groups, and because of the BCH formula, every Lie group has a (unique) analytic structure, and in fact, we obtain that continuous group homomorphisms are analytic; see [Hel01, p. 117, Theorem 2.6]. Hence, one can improve Theorem 6.3.1 and prove the analyticity of isometries between metric Lie groups; see [KL17, Theorem 1.1] and also [LO16]. We shall mainly focus on the C^∞ regularity.

Proof of Theorem 6.3.1 Let $F\colon M_1 \to M_2$ be an isometry between metric Lie groups. Without loss of generality, we may assume that $F(1_{M_1}) = 1_{M_2}$ and that both M_1 and M_2 are connected since left translations are smooth isometries and connected components of identity elements are open. By Theorem 6.2.10, for $i \in \{1, 2\}$, the space $G_i := \mathrm{Isom}(M_i)$ is a Lie group smoothly acting on M_i; see Exercise 6.6.21. The conjugation map $\mathrm{C}_F\colon G_1 \to G_2$ defined as $I \mapsto F \circ I \circ F^{-1}$ is a group isomorphism that is continuous with respect to the point-wise convergence. Hence, the map C_F is smooth by Theorem 5.3.2.

Consider also the inclusion $\iota\colon M_1 \to G_1$, $m \mapsto L_m$, which is smooth being a continuous homomorphism, and the orbit map $\sigma\colon G_2 \to M_2$, $I \mapsto I(1_{M_2})$, which is smooth since the action is smooth. We deduce that $\sigma \circ \mathrm{C}_F \circ \iota$ is smooth. We claim that this map is F. Indeed, for every $m \in M_1$ it holds

$$(\sigma \circ \mathrm{C}_F \circ \iota)(m) = \sigma(F \circ L_m \circ F^{-1}) = (F \circ L_m \circ F^{-1})(1_{M_2}) = F(m).$$

□

Remark 6.3.2 Using the same techniques above, one can prove the following result, which is meant to summarize the smooth case and generalize it to the analytic category. Let M be a metric Lie group. Assume that M is connected. Consider M to be equipped with its unique analytic structure. Then, the isometry group $\mathrm{Isom}(M)$ has the structure of a Lie group (finite-dimensional and with finitely many connected components), which, equipped with its unique analytic structure, acts analytically on M. Moreover, the stabilizers of the action $\mathrm{Isom}(M) \curvearrowright M$ are compact analytic Lie subgroups. For more details, see [KL17].

We get another consequence of the solution of Hilbert's fifth problem on Lie homogeneous spaces. Let $M = G/H$ be a Lie coset space of a Lie group G modulo a compact subgroup H. Hence, the space M has the structure of an (analytic-) smooth manifold, as from Theorem 6.1.6. Assume that M is equipped with a G-invariant admissible distance function. From Theorem 6.2.10 with Exercise 6.6.21, it is immediate that the group of isometries $\mathrm{Iso}(M, d)$ of the manifold M is a Lie group acting transitively on M. Hence, the space M admits an analytic structure for which $\mathrm{Iso}(M, d)$ acts by analytic maps. One can prove that this analytic structure coincides with the initial one.

Theorem 6.3.3 ([LO16, Theorem 1.3]) *Let G/H be a homogeneous space of a Lie group G modulo a compact subgroup H. Assume that d is a G-invariant distance that induces the manifold topology. If $F : (G/H, d) \to (G/H, d)$ is an isometry, then F is analytic.*

6.3.2 Submetries Between Metric Groups

Distance functions on metric groups do not necessarily descend to quotients. In fact, let us fix a metric group (G, d) and a subgroup H of G. We stress that, by definition, the group G acts on itself isometrically by left translations. Naturally, the group H acts on G in two ways: by left translations or by right translations. Unless the group is normal, these actions have different orbits. The respective quotient spaces are

$$H\backslash G := \{Hg \;:\; g \in G\}$$

and

$$G/H := \{gH \;:\; g \in G\}.$$

On G/H there is still a transitive G-action: $G \curvearrowright G/H$ as

$$\bar{g}.(gH) := \bar{g}gH, \qquad \forall \bar{g}, g \in G.$$

However, there is no 'adapted' metric geometry on G/H, in the sense that there may be no admissible distance for which this action is by isometries; see Exercise 6.6.23.

Instead, regarding $H\backslash G$, the isometric action by left translations by elements of G does not pass to this quotient. But, there is a good candidate for a distance function:

$$d(Hg_1, Hg_2) := \inf\left\{d(h_1g_1, h_2g_2) \;:\; h_1, h_2 \in H\right\}, \qquad \forall g_1, g_2 \in G. \tag{6.4}$$

Note that this definition is in agreement with the notation that, for subsets A and B of a metric space, we set $d(A, B) := \inf\{d(p, q);\; p \in A,\; q \in B\}$.

The function (6.4) is symmetric and satisfies the triangle inequality (see Proposition 6.3.4); however, it may not be positively defined. In fact, in general situations, pathologies may be encountered. This is not the case if we assume that H is a closed subgroup, so that $H\backslash G$ is a Hausdorff space, see Exercise 6.6.3, and we assume that the right translations on G are continuous with respect to the topology induced by the distance on G, as in the case that G is a topological group equipped with an admissible distance function. In this setting, not only is it the case that (6.4) metrizes the quotient, but in addition, the quotient map sends balls to balls of the same radius.

There is an analog for metric spaces of the following result; see Exercise 3.4.39.

Proposition 6.3.4 *Let G be a topological group equipped with an admissible left-invariant distance function d. Let H be a closed subgroup of G. Then the function* (6.4) *satisfies*

$$d(Hg_1, Hg_2) = \inf\{d(g_1, hg_2) \;:\; h \in H\}, \qquad \forall g_1, g_2 \in G, \tag{6.5}$$

and is an admissible distance function on the quotient space ${}_H\backslash^G$ *for which the projection* $\pi : G \to {}_H\backslash^G$ *satisfies*

$$\pi(B(p,r)) = B(\pi(p), r), \qquad \forall p \in G, \forall r > 0. \tag{6.6}$$

If, in addition, H is boundedly compact, then π is a submetry in the sense of Definition 3.1.23.

Proof Because the distance function on G is left-invariant, then for all $h_1 \in H$ and all $g_1, g_2 \in G$ we have

$$d(g_1, Hg_2) \overset{\text{def}}{=} \inf\{d(g_1, hg_2) : h \in H\}$$
$$= \inf\{d(h_1 g_1, hg_2) : h \in H\} \overset{\text{def}}{=} d(h_1 g_1, Hg_2).$$

Taking the infimum over $h_1 \in H$ we infer

$$d(g_1, Hg_2) = d(Hg_1, Hg_2), \qquad \forall g_1, g_2 \in G. \tag{6.7}$$

We next check that it is a distance function. It is obviously symmetric and zero if $Hg_1 = Hg_2$. Whereas, assume $Hg_1 \neq Hg_2$, so $g_1 \notin Hg_2$. Because we are assuming that H is closed and that each right translation R_{g_2} is continuous, we have that Hg_2 is closed. Therefore, there exists $r > 0$ such that $B(g_1, r) \cap Hg_2 = \emptyset$. Hence, by (6.7) we infer $d(Hg_1, Hg_2) = d(g_1, Hg_2) \geq r > 0$. So the distance function is positively defined. Regarding the triangle inequality, for each $g_1, g_2, g_3 \in G$ and for each $\epsilon > 0$, let $h_\epsilon, h'_\epsilon \in H$ such that $d(g_1, h_\epsilon g_2) \leq d(Hg_1, Hg_2) + \epsilon$ and $d(h_\epsilon g_2, h_\epsilon h'_\epsilon g_3) = d(g_2, h'_\epsilon g_3) \leq d(Hg_2, Hg_3) + \epsilon$. Then

$$\begin{aligned} d(Hg_1, Hg_3) &\leq d(g_1, h_\epsilon h'_\epsilon g_3) \\ &\leq d(g_1, h_\epsilon g_2) + d(h_\epsilon g_2, h_\epsilon h'_\epsilon g_3) \\ &\leq d(Hg_1, Hg_2) + d(Hg_2, Hg_3) + 2\epsilon. \end{aligned}$$

By the arbitrariness of ϵ, we get the triangle inequality.

Before proving that this distance is admissible, we check that for all $p \in G$ and $r > 0$ the map π satisfies (6.6); and we show (3.21) if H is boundedly compact. For one inclusion, if $Hq \in \pi(\bar{B}(p,r)) = H\bar{B}(p,r)$ with $q \in \bar{B}(p,r)$, then $d(Hq, Hp) \leq d(q, p) \leq r$ and so $Hq \in \bar{B}(Hp, r) = \bar{B}(\pi(p), r)$. Similarly, $\pi(B(p,r)) \subseteq B(\pi(p), r)$; see also Exercise 3.4.33. Vice versa, assume $Hq \in B(Hp, r)$, i.e., $d(Hq, p) \overset{(6.7)}{=} d(Hq, Hp) < r$. Then, since this is an open condition, there exists $h \in H$ such that for $\bar{q} := hq$ we have $d(\bar{q}, p) < r$ and so $Hq = H\bar{q} = \pi(\bar{q}) \in \pi(B(p,r))$. So, we proved (6.6). Similarly, if H is boundedly compact, we take $Hq \in \bar{B}(Hp, r)$, so $d(Hq, p) \leq r$. Now, this latter distance has to be achieved on the compact set $Hq \cap \bar{B}(p,r)$, then there exists $h \in H$ such that for $\bar{q} := hq$ we have $d(\bar{q}, p) \leq r$ and so $Hq = H\bar{q} = \pi(\bar{q}) \in \pi(\bar{B}(p, r))$. So (3.21) is verified.

Finally, we show that the distance defines the quotient topology on ${}_H\backslash^G$. We stress that every submetry is an open map, is Lipschitz, and hence, is continuous. Since the quotient topology on G/N is the only topology for which the projection $\pi : G \to G/N$ is continuous and open, the considered distance is admissible. □

Clearly, if we consider actions of normal subgroups, then we have quotient spaces that are homogeneously metrized. Recall that $G/N = {}_N\backslash^G$ is a (Hausdorff) topological group when N is a closed normal subgroup of a topological group G. Therefore, the group G acts on the group G/N by isometries with respect to the distance (6.4). Thus, Proposition 6.3.4 gives the following consequence.

Corollary 6.3.5 *Let G be a topological group equipped with an admissible left-invariant distance function d. Let N be a normal, boundedly compact subgroup of G. Then the function*

$$d_{G/N}(g_1N, g_2N) := d(g_1N, g_2N) \stackrel{\text{def}}{=} \inf\{d(g_1n_1, g_2n_2) \, : \, n_1, n_2 \in N\} \tag{6.8}$$

is an admissible left-invariant distance function on the quotient group G/N for which the projection $\pi : G \to G/N$ becomes a submetry.

Remark 6.3.6 In the setting of infinite-dimensional spaces, distances between closed subsets may not be realized. Hence, already in the case of Banach spaces, there are examples of quotients by closed subspaces that do not give submetries as in (3.21) but they just verify (6.6); see Example 3.4.36.

6.3.3 Quasi-Isometric Equivalence of Geodesic Distances

In this section, we establish the result that geodesic left-invariant metrics on the same group G are quasi-isometrically equivalent, according to Definition 3.1.13. Actually, we can relax the assumption that the metrics are geodesic and instead require them to be quasi-geodesic.

Definition 6.3.7 (Quasi-Geodesic Space) A metric space M is said to be *quasi-geodesic* if there exist constants $C > 0$ and $L > 1$ such that every two points in M can be joined with a (L, C)-*quasi-arc* in the following sense: for all $x, x' \in M$, there exist $k \in \mathbb{N}$ and $x_0, x_1, \ldots, x_k \in M$ such that $x_0 = x$, $x_k = x'$, $d(x_{i-1}, x_i) \leq C$, for $i \in \{1, \ldots, k\}$, and $\sum_{i=1}^{k} d(x_{i-1}, x_i) \leq Ld(x, x') + C$. Equivalently, there exists an (L, C)-*quasi-arc* joining x and x' if and only if there is an (L, C)-quasi-isometric embedding of an interval into the metric space, with x and x' in its image. See Definition 3.1.13 for the notion of quasi-isometric embedding.

We stress 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; see Exercise 6.6.24.

Proposition 6.3.8 *Let d and d' be quasi-geodesic left-invariant distances on a group. Assume that d and d' have the same bounded subsets. Then there exist constants $c \geq 0$ and $L \geq 1$ such that $L^{-1}d - c \leq d' \leq Ld + c$.*

Proof Since d is quasi-geodesic, there are two constants $C_1 \geq 0$ and $L_1 \geq 1$ with the following property: Given a group element g, there are $g_1, \ldots, g_n$ such that $d(1, g_i) \leq C_1$, for all i, and $g = g_1 \cdot \ldots \cdot g_n$, while $\sum_1^n d(1, g_i) \leq L_1 d(1, g) + C_1$. Grouping some g_i's together if necessary, we may assume that $C_1/2 \leq d(1, g_i)$, for $i \in \{1, \ldots, n-1\}$, still having the weaker uniform condition $d(1, g_i) \leq 2C_1$, for $i \in \{1, \ldots, n\}$. We claim that

$$n \leq \frac{2L_1}{C_1} d(1, g) + 3. \tag{6.9}$$

Indeed, using the lower bound on each $d(1, g_i)$ we get

$$(n-1)C_1/2 \leq \sum_{i=1}^{n} d(1, g_i) \leq L_1 d(1, g) + C_1,$$

from which the claim (6.9) follows.

By our assumption on the distances, since the values $d(1, g_i)$ are uniformly bounded by $2C_1$, then also the values $d'(1, g_i)$ are uniformly bounded, say by some constant $\tilde{C} > 0$ independent of g. Then

$$d'(1, g) \leq \sum_{i=1}^{n} d'(1, g_i) \leq \tilde{C} n \overset{(6.9)}{\leq} L d(1, g) + c,$$

for $L := \dfrac{2\tilde{C} L_1}{C_1}$ and $c := 3\tilde{C}$. The proposition follows by left invariance and by exchanging the roles of d and d'. □

The above result will apply in particular to Finsler and sub-Finsler left-invariant metrics on Lie groups. In fact, every pair of left-invariant subFisler distances on the same Lie group are quasi-isometric equivalent. For example, the Riemannian Heisenberg group and the sub-Riemannian Heisenberg group are quasi-isometric. In fact, quasi-isometries can completely change the local geometry. We conclude the section with a question.

Problem 6.3.9 Are there translation-invariant quasi-geodesic metrics on $(\mathbb{R}^2, +)$ that are not quasi-isometric to a geodesic distance? Are there such metrics that, in addition, are boundedly compact?

6.4 Haar Measures and Polynomial Growth

Every Lie group, as every locally compact group, has a natural class of measures: the *Haar measures*. A measure μ on a group G is called *left-invariant*, if for every $g \in G$ and every set E on which μ is defined, we have that μ is defined on $L_g^{-1}(E)$ and

$$\left((L_g)_\# \mu\right)(E) := \mu\left(L_g^{-1}(E)\right) = \mu(E).$$

If G is equipped with a topology, then a Borel measure μ on G is said to be a *Radon measure* if it is finite on compact sets, outer regular on Borel sets, and inner regular on open sets; see [Fol99, page 212]. If G is a topological group, then a Borel measure μ on G is called a *left-Haar measure*, or simply a *Haar measure*, if it is left-invariant, Radon, and not the zero measure. Similarly, a *right-Haar measure* is a Radon non-zero measure that is right invariant.

On every locally compact group G, there are Haar measures; see [Fol99, Theorem 11.8]. Moreover, Haar measures are unique in the following sense: Every two left-Haar measures μ_1 and μ_2 on the same locally compact group differ by a constant multiplicative factor: $\mu_1 = c\mu_2$, for some $c > 0$. We refer to [Fol99, Theorem 11.9] for a proof.

Using the Haar measure, we can consider the asymptotic growth of locally compact groups:

Definition 6.4.1 (Polynomial Growth and Exponential Growth) A Lie group G is said to have *polynomial growth* if the Haar measures of powers of compact sets grow at most polynomially. Namely, writing μ to denote a Haar measure of G, for every compact set $U \subseteq G$ there are $C, Q > 0$ such that $\mu(U^n) \leq Cn^Q$, for all $n \in \mathbb{N}$. A Lie group G is said to have *exponential growth* if for every generating compact neighborhood $U \subseteq G$ of 1_G there are $k > 1$ and $C > 0$ such that $\mu(U^n) \geq Ck^n$, for all $n \in \mathbb{N}$.

We stress that in the case G is connected, to check whether it has polynomial growth, it is enough to check the property for some compact set U that is a neighborhood of the identity element; see Exercise 6.6.25. Similarly, a group has exponential growth if and only if for some compact neighborhood U of the identity element there is $k > 1$ such that $\mu(U^n) \geq k^n$, for all $n \in \mathbb{N}$.

6.4.1 Type (R) and Guivarc'h-Jenkins Theorem

In this section, we will present an algebraic characterization of groups with polynomial growth. We begin by introducing the algebraic property.

Definition 6.4.2 (Type (R)) A Lie algebra $\mathfrak{g}$ is said to be of *type (R)* if the eigenvalues of ad_X are purely imaginary for each $X \in \mathfrak{g}$. A Lie group G is said to be of *type (R)* if the eigenvalues of Ad_g are of absolute value 1 for each $g \in G$.

The two notions are related: A connected Lie group is of type (R) if and only if its Lie algebra is of type (R). A proof can be found in [Jen73b, Proposition 1.3'].

Guivarc'h and Jenkins independently proved the following equivalence:

Theorem 6.4.3 (Guivarc'h-Jenkins, [Gui73, Jen73b]) *A connected Lie group has polynomial growth if and only if it is of type (R).*

Sketch of the Proof If a group is not of type (R), then we will soon prove in Proposition 6.4.5 that the group has exponential growth.

Conversely, assume that the group is of type (R). The group is quasi-isometric to a solvable Lie group (This is because every Levi factor is compact, in this case. We refer to Exercise 10.5.31 for some review of Levi decomposition). Every solvable Lie group of type (R) can actually be made isometric to a nilpotent Lie group, called its *nilshadow*. For a study of nilshadows, we refer to [DER03, Bre14, Cow+24]. The punch line is that nilpotent Lie groups have polynomial growth because they are asymptotic to Carnot groups. Proving this latter statement will be the main aim of Sects. 12.4 and 12.8. □

Example 6.4.4 (Models of Groups Not of Type (R)) We present now some Lie groups, denoted by G_z, with $z \in \mathbb{C} \setminus i\mathbb{R}$, that are in some sense the reason for exponential growth. Namely, we will deduce that if a group has exponential growth, then it has one of such groups as a subgroup. Given a complex number z that is not purely imaginary, we distinguish two cases: either z is real or not.

If $z \in \mathbb{R} \setminus \{0\}$, then we consider the 2D Lie algebra that has a basis X, Y with operation $[X, Y] = zY$. Let G_z be the associated simply connected Lie group. Such a Lie group is not of type (R) and, when metrized with a left-invariant Riemannian metric, it has constant negative curvature. When properly renormalized, it is the hyperbolic plane, which has exponential growth.

If instead $z = a + bi$ with $a, b \neq 0$, then we consider the 3D Lie algebra that has a basis X_1, X_2, X_3 with operation

$$\begin{aligned}
[X_1, X_2] &= aX_2 + bX_3, \\
[X_1, X_3] &= aX_3 - bX_2, \\
[X_2, X_3] &= 0.
\end{aligned}$$

Let G_z be the associated simply connected Lie group. Such a Lie group is not of type (R) and it has exponential growth; see Exercise 6.6.27. These are examples of Heintze groups, as we will discuss in Chap. 14.

Proposition 6.4.5 *If G is not of type (R), then there is $z \in \mathbb{C}$ and a subgroup H of G such that H is isomorphic to G_z. Consequently, the group G has exponential growth.*

Proof We begin by searching for a subalgebra of $\mathfrak{g}$ that is isomorphic to the Lie algebra of some G_z. Let $X_1 \in \mathfrak{g}$ for which ad_{X_1} had some eigenvalue in $\mathbb{C} \setminus i\mathbb{R}$. We distinguish two cases: either there is a real eigenvalue, or not. If there is a real eigenvalue $z \in \mathbb{R} \setminus \{0\}$ with eigenvector X_2, then the span of X_1 and X_2 is the Lie algebra of G_z.

If, instead, there are no real eigenvalues for ad_{X_1}, then take an eigenvalue $z = a + bi$ with $a, b \neq 0$. Then there is X_2 and $X_3 \in \mathfrak{g}$ such that

$$\begin{aligned} [X_1, X_2] &= aX_2 + bX_3, \\ [X_1, X_3] &= aX_3 - bX_2 \end{aligned} .$$

Using these relations and the Jacobi identity, we have $\mathrm{ad}_{X_1}[X_2, X_3] = 2a[X_2, X_3]$. Since we are assuming that there are no real eigenvalues and $a \neq 0$, we deduce $[X_2, X_3] = 0$. The span of X_1, X_2, and X_3 gives the Lie algebra of G_z.

Once we have found a Lie algebra that is isomorphic to the Lie algebra of G_z's, we have a subgroup H of G with such a Lie algebra, recall Theorem 5.1.4. Because G_z is simply connected, the group H is a quotient of G_z; see Theorem 5.7.2 or Exercise 5.8.31. However, one can show that G_z has no normal discrete subgroups (see Exercise 6.6.28 and 6.6.27) Thus, the groups G_z and H are isomorphic.

We stress that H is necessarily closed in G. In fact, the group H is properly embedded in G because, due to the possible ad_X in G_z, when elements $h_n \in H$ diverge in H then Ad_{h_n} diverge as maps on the Lie algebra of H and then as a map on the Lie algebra of G. Hence, they cannot converge in G.

To obtain exponential growth of the volume, we will consider the growth of separated nets and rely on a general relation; see Exercise 6.6.29. We metrize G with a geodesic left-invariant distance, e.g., Riemannian, and consider on H the induced path distance. The intersection of the closed unit ball $\bar{B}_G(1, 1)$ and H is compact. Thus, for some $\delta > 0$ we have $\bar{B}_H(1, \delta) \supseteq \bar{B}_G(1, 1) \cap H$. Since G_z and H have exponential growth, we have that maximal δ-separated sets in the balls of radius R in H grow exponentially in R. Consequently, maximal 1-separated sets in the balls of radius R in G grow at least exponentially in R. Therefore, recalling Exercise 6.6.29, the group G has exponential growth. □

6.5 Isometrically Homogeneous Spaces with Dilations (First Part)

Next, we present the notion of dilation on a metric space. Fixed a scalar number λ, we will denote by δ_λ a dilation by factor λ. However, there is no such map for general metric spaces, and if it exists, it may not be unique. In some settings, like in vector spaces or, more generally, in Carnot groups, we have canonical maps given by the algebraic structure.

Let M be a metric space with distance d, and let $\lambda \in (0, +\infty)$. A *dilation of factor* λ on M is a surjective map $\delta_\lambda : M \to M$ such that

$$d(\delta_\lambda(p), \delta_\lambda(q)) = \lambda d(p, q), \qquad \forall p, q \in M.$$

We say that a metric space X is *self-similar* if it admits a dilation of factor $\lambda \in (0, +\infty) \setminus \{1\}$. Such dilations are also called *homotheties*.

Clearly, the dilations of factor 1 are precisely the isometries. Hence, isometrically homogeneous spaces have plenty of them.

In the next result, we show that, under some mild topological assumptions, if a metric space is isometrically homogeneous and self-similar, then it has the structure of a Lie coset space. After the revision of the theory of nilpotent Lie groups done in Chap. 9, we will reach a stronger conclusion: the space has the structure of a nilpotent metric Lie group for which the dilations fixing the identity element are group isomorphisms; see Theorem 10.2.1.

Theorem 6.5.1 (To be Improved in Theorem 10.2.1) *Let M be a metric space that is*

(1) *locally compact,* (2) *locally connected*
(3) *isometrically homogeneous,* (4) *self-similar.*

Then, M has the structure of a smooth manifold. In fact, there is a connected Lie group G, a compact subgroup S of G, and a G-invariant distance on the coset space G/S such that G/S is isometric to M.

Before proving Theorem 6.5.1, we provide reasons why the considered space has finite topological dimension and is connected. The aim is to use Gleason-Montgomery-Yamabe-Zippin's result, Theorem 6.2.10.

As a side remark, if a locally compact isometrically homogeneous space admits dilations of *every* factor, then it is connected; see Exercise 6.6.31. However, in Theorem 6.5.1, we assume only the presence of at least *one* dilation (together with all its iterations, of course).

Proposition 6.5.2 *Every locally compact, self-similar, isometrically homogeneous space is metrically doubling, and in particular, it has finite Hausdorff dimension and finite topological dimension.*

Proof We need to prove that for every such metric space X, there is $C > 0$ such that every ball of radius $r > 0$ in X can be covered with C balls of radius $r/2$. Since X is locally compact, it has a closed ball $\bar{B}(x_0, r_0)$ that is compact. Up to replacing the dilation by its inverse, we may assume that its factor λ is greater than 1. From Exercise 6.6.30, the metric space is boundedly compact. Hence, the closed ball $\bar{B}(x_0, \lambda r_0)$ can be covered with finitely many balls of radius $r_0/2$, say C many of them. Therefore, for every $s \in [1, \lambda]$ each ball $B(x_0, sr_0)$ can be covered with 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. Finally, recall that doubling metric spaces have finite Hausdorff dimension and, hence, finite topological dimension; see [Hei01, Theorem 8.14 and Exercise 10.16]. □

Proof of Theorem 6.5.1 Let $\delta_\lambda : M \to M$ be a dilation of factor $\lambda \in (0, +\infty) \setminus \{1\}$. Since δ_λ is surjective, it is a homeomorphism and its inverse $(\delta_\lambda)^{-1} : M \to M$ is a dilation of factor $\frac{1}{\lambda}$. Hence, up to replacing δ_λ with its inverse if necessary, we may assume $\lambda < 1$. Observe that the space is complete (see Lemma 6.2.4) and apply the Banach Fixed Point Theorem: The contraction δ_λ has a unique fixed point, which we denote by o. We consider the orbit map with respect to o, i.e., the map $f \in \mathrm{Isom}(M, d) \mapsto \pi(f) := f(o) \in M$.

Recall Proposition 6.2.7 regarding the fact that $\mathrm{Isom}(M, d)$ is a locally compact group acting continuously on M with compact stabilizers. Moreover, by Theorem 6.1.3, the quotient map π induces a homeomorphism between M and the manifold $\mathrm{Isom}(M, d)/S$, where $S := \mathrm{Stab}(o) = \{f \in \mathrm{Isom}(M, d) : f(o) = o\}$.

In order to apply Gleason-Montgomery-Yamabe-Zippin's Theorem 6.2.10 to (M, d), we need to observe that M has finite topological dimension and is connected; see Proposition 6.5.2 and Exercise 6.6.30, respectively. Hence, thanks to Theorem 6.2.10, $\mathrm{Isom}(M, d)$ and M have some differential structures such that $\mathrm{Isom}(M, d)$ is a Lie group and the action of $\mathrm{Isom}(M, d)$ on M is smooth. To conclude, we take $G := \mathrm{Isom}(M, d)^\circ$ to be the identity component of $\mathrm{Isom}(M, d)$ and the stabilizer S as compact subgroup. □

Suppose that G is a Lie group, and let $\mathrm{Der}(\mathfrak{g})$ be the space of derivations on its Lie algebra; see Definition 5.6.2. By Proposition 5.6.3.ii, every multiplicative one-parameter group $\mathbb{R}_+ \to \mathrm{Aut}(G)$, $\lambda \mapsto \delta_\lambda$, of Lie automorphisms is determined by some derivation $A \in \mathrm{Der}(\mathfrak{g})$ such that

$$(\delta_\lambda)_* = \lambda^A := e^{(\log \lambda) A}. \tag{6.10}$$

Such an A is the called the *infinitesimal generator* of $\lambda \mapsto \delta_\lambda$, and such $(\delta_\lambda)_\lambda$ form the *(multiplicative) one-parameter subgroup of automorphisms determined* by the derivation A.

Definition 6.5.3 (Homogeneous Metric Group) Let G be a Lie group and $A \in \mathrm{Der}(\mathfrak{g})$ the derivation on its Lie algebra determined by a one-parameter subgroup $(\delta_\lambda)_\lambda$ of automorphisms of G. Assume that d is a left-invariant distance function on G for which

$$d(\delta_\lambda x, \delta_\lambda y) = \lambda d(x, y) \qquad \forall x, y \in G,\ \forall \lambda > 0. \tag{6.11}$$

Then, we say that d is a *A-homogeneous* distance function on G and (G, d) is a *A-homogeneous metric group*.

The notion of a homogeneous metric group captures a deep interaction between algebraic and metric structures. These kinds of structures are indeed quite remarkable. What is particularly striking is their appearance in both geometric and analytic contexts. In Sect. 10.2, and later in Sect. 14.3, we will return to the discussion of homogeneous metric groups.

6.6 Exercises

Exercise 6.6.1 (Characterisations of Proper Actions) Let M be a manifold and G a Lie group acting continuously on M. The following are equivalent: (a) the action is proper; (b) if $(p_i)_{i\in\mathbb{N}}$ is a sequence in M and $(g_i)_{i\in\mathbb{N}}$ is a sequence in G such that both $(p_i)_{i\in\mathbb{N}}$ and $(g_i.p_i)_{i\in\mathbb{N}}$ converge, then a subsequence of $(g_i)_{i\in\mathbb{N}}$ converges; (c) for every compact subset $K \subseteq M$, the set $\{g \in G : (g.K) \cap K \neq \emptyset\}$ is compact.
Hint. A solution can be found in [Lee13, Proposition 21.5].

Exercise 6.6.2 For a continuous action $G \curvearrowright X$ of a topological group on a topological space, with quotient map π, declare a subset $U \subseteq G\backslash X$ to be open if and only if $\pi^{-1}(U)$ is open in X. This definition gives a topology on X, and it makes the map π continuous and open. Moreover, this is the only topology that makes π continuous and open.

Exercise 6.6.3 Let G be a topological group and $H < G$ a topological subgroup. The actions from Example 6.1.2 are continuous. If H is closed, then for each of these actions, we have that $H\backslash G$ is a Hausdorff topological space when equipped with the quotient topology.
Hint. Check Exercise 6.6.10, using that topological groups are T_3-regular.

Exercise 6.6.4 Given a group G, and denoting by L_g its left translation by g, we have that $g \mapsto L_g$ is an action of G on G. Moreover, if G is a Lie group, then this is a Lie action.

Exercise 6.6.5 If $G \curvearrowright M$ is a proper and free Lie action, then $\dim G\backslash M = \dim M - \dim G$.

Exercises 6.6.6–6.6.9 are extra exercises on Lie actions.

Exercise 6.6.6 Let Γ be a discrete Lie group, M a smooth manifold, and $\Gamma \curvearrowright M$ a Lie action that is proper and free. Then, there exists a differentiable structure on $\Gamma\backslash M$ such that $M \to \Gamma\backslash M$ is a smooth covering map.

Exercise 6.6.7 Let G be a Lie group and $\Gamma < G$ a closed subgroup. Then, the natural action $\Gamma \curvearrowright G$ by left-translations is proper and free.
Solution. The action by left-translations is free because $L_g(p) = p$ implies $g = 1_G$. To see that the action is proper, notice that the map $\Psi\colon G\times G \to G\times G$, $(g_1, g_2) \mapsto (g_1g_2, g_2)$, has the inverse $(g_1, g_2) \mapsto (g_1g_2^{-1}, g_2)$ and thus $\Psi \in \mathrm{Diffeo}(G \times G)$. The map that is claimed to be proper is $\tilde{\Theta} = \Psi|_{\Gamma\times G}$. Given a compact set $K \subset G \times G$, we have $\tilde{\Theta}^{-1}(K) = \Psi^{-1}(K) \cap (\Gamma \times G)$. Since Γ is closed in G, then $\tilde{\Theta}^{-1}(K)$ is an intersection of a compact set with a closed set.

Exercise 6.6.8 If G is a Lie group and $\Gamma < G$ a discrete subgroup, then there exists a differentiable structure on $\Gamma\backslash G$ such that the natural map $G \to \Gamma\backslash G$ is a smooth covering map.

Exercise 6.6.9 Let G be a Lie group and M a simply connected smooth manifold such that $\dim G \leq \dim M$. If there exists a transitive Lie action $G \curvearrowright M$, then M has a structure of a Lie group (i.e., M is diffeomorphic to a Lie group).

Exercise 6.6.10 If H is a closed subgroup of a Lie group G, then G/H is Hausdorff.
Solution. Consider the set

$$R := \{(h_1, h_2) \in G \times G \mid \text{ there exists } h \in H \text{ such that } h_1 = h_2 h\},$$

which is a closed subset of $G \times G$. So, if $h_1 H$ and $h_2 H$ are two distinct cosets, then (h_1, h_2) does not belong to R, thus one can find open sets $h_1 \in V$ and $h_2 \in W$ such that $(V \times W) \cap R = \emptyset$. Since the projection map π is an open map, the sets $\pi(V)$ and $\pi(W)$ are disjoint and open in G/H, containing $h_1 H$ and $h_2 H$, respectively.

Exercise 6.6.11 If H is a subgroup of a Lie group G, then G/H is second-countable.
Hint. Every countable basis for the topology of G projects to a countable basis for G/H.

Exercise 6.6.12 Let (X, d) be a metric space. Within this exercise, we say that the distance function d is a *homogeneous distance* if for every $\lambda > 0$ there is a bijection $\delta_\lambda : X \to X$ such that

$$d(\delta_\lambda(x), \delta_\lambda(x')) = \lambda d(x, x'), \qquad \forall x, x' \in X.$$

In this case, we also say that δ_λ is a *dilation* by λ. We have the following properties:

6.6.12.i. On $\mathbb{R}^n$, the Euclidean distance d_E is homogeneous;
6.6.12.ii. On $(\mathbb{R} \times \{0\}) \cup (\{0\} \times \mathbb{R})$ as subset of $\mathbb{R}^2$, the distance function d_E is homogeneous;
6.6.12.iii. On $\mathbb{S}^1$ as subset of $\mathbb{R}^2$, the distance function d_E is not homogeneous.

Question: Which of these spaces are isometrically homogeneous, according to Definition 6.2.1?

Exercise 6.6.13 If $G \curvearrowright X$ continuously, then the induced map $G \to \mathrm{Homeo}(X)$ is continuous with respect to the compact-open topology.

Exercise 6.6.14 If a non-necessarily Hausdorff topological group G acts on a Hausdorff space continuously and faithfully, then G is Hausdorff.
Hint. Use Exercise 6.6.13.

Exercise 6.6.15 Recall that a topological space X is called *first-countable* if its topology admits a countable local basis at every point. It is called *second-countable* if its topology admits a countable basis. While X is σ-*compact* if it is the countable union of compact sets.

6.6.15.i. Every second-countable space is first-countable.
6.6.15.ii. Every second-countable, locally compact space is σ-compact.
6.6.15.iii. There is a topological space that is σ-compact and locally compact, but it is not second-countable.
Hint. Consider the excluded-point topology on the real numbers.

Exercise 6.6.16 Pointwise limits of isometric embeddings are isometric embeddings.

Exercise 6.6.17 Let G be a locally compact group with a countable basis acting continuously and transitively on a connected locally compact Hausdorff space X. Then, every open subgroup of G acts transitively on X.
Hint. The proof is very similar to the one we provided for Proposition 6.2.9.

Exercise 6.6.18 Let M be a boundedly compact metric space. Equip the isometry group $\mathrm{Isom}(M)$ with the compact-open topology.

6.6.18.i. The action $\mathrm{Isom}(M) \curvearrowright M$ is continuous and proper.
6.6.18.ii. The topology on $\mathrm{Isom}(M)$ is second-countable and locally compact, and hence, σ-compact.
6.6.18.iii. Check that $\mathrm{Isom}(M)$ is a topological group.

Hint. One can find a proof in [CH16, Lemma 5.B.4]. Recall also Exercise 6.6.15.ii.

Exercise 6.6.19 Let $G \curvearrowright M$ be a Lie group action that is transitive, faithful, and with compact stabilizers. Then, the topological space G has finitely many connected components.
Hint. On the one hand, the quotient G/G° is discrete since connected components of manifolds are open. On the other hand, this space G/G° is compact because, fixing $p \in M$, for every sequence $g_n G^\circ$ there exists $g_n' \in g_n G^\circ$ such that $g_n'.p = p$. (Recall that G° acts transitively by Proposition 6.2.9). Thus, the sequence g_n' subconverges.

Exercise 6.6.20 Let M be a topological manifold. Let G be a locally compact group with a countable basis. Let $G \times M \to M$ be a continuous, effective, and transitive action of G on M. Then G is a Lie group, and M can be equipped with a differentiable structure so that G acts by diffeomorphisms.

Exercise 6.6.21 Let M be a connected metric Lie group. Then, the topological group $\mathrm{Isom}(M)$ has the structure of a Lie group smoothly acting on M.
Hint. In addition to Theorem 6.2.10 recall that Lie groups have unique smooth structures as in Corollary 5.3.3.

Exercise 6.6.22 Let M be a metric space. Let $G \curvearrowright M$ be a continuous action by isometries with closed orbits. Then, the quotient map $\pi : M \to G\backslash M$ becomes a submetry with respect to some unique distance on $G\backslash M$.

Hint. Check Exercise 3.4.39.

Exercise 6.6.23 Let $\mathcal{H}$ be the Heisenberg group with Lie algebra spanned by $X, Y, [X, Y]$. Let $L := \exp(\mathbb{R}X)$. Then $\mathcal{H}/L$ is a Lie homogeneous space on which $\mathcal{H}$ acts on the left. The stabilizer of this action is homeomorphic to L, which is not relatively compact inside the space of homeomorphisms of $\mathcal{H}/L$. In particular, this action cannot be isometric with respect to any admissible distance.

Exercise 6.6.24 Let X be a topological space equipped with a distance d that does not necessarily induce the same topology. Assume that d is locally bounded (in the sense that it is bounded on compact subsets of X) and proper (in the sense that the distance function from a point in X is a proper map). Then, the sets that are bounded with respect to d are exactly the precompact sets.

Exercise 6.6.25 Let G be a connected Lie group and let μ_G be a Haar measure on G. Let $U \subseteq G$ be a compact neighborhood of $1_G \in G$. If there is $C > 0$ such that $\mu_G(U^n) \le Cn^Q$, for all $n \in \mathbb{N}$, then G has polynomial growth.
Hint. By Exercise 5.8.3, for every other compact set $\tilde{U} \subseteq G$ there is $k \in \mathbb{N}$ such that $\tilde{U} \subseteq U^k$.

Exercise 6.6.26 Let M be a metric space as in Theorem 6.5.1. Let G be the identity component of the isometry group $\mathrm{Isom}(M, d)$, equipped with the topology of the uniform convergence on compact sets. Let $\pi : G \to M$ be the orbit map $f \mapsto f(o)$. Let μ_G be a Haar measure on G, and consider the push forward $\mu_M := \pi_*\mu_G$

6.6.26.i. The measure μ_M is a G-invariant Radon measure on M such that $\mu_M(E) = \mu_G(\pi^{-1}(E))$ for all open sets $E \subset M$.
6.6.26ii. The measure μ_M is a doubling measure (in the sense of Exercise 3.4.31) and there are $C > 0$ and $Q > 0$ such that for all $p \in M$ and all $r \ge 1$ one has $\mu_M(B_d(p, r)) \le Cr^Q$.
6.6.26.iii. Consequently, the topological group G has polynomial growth.

Hint. Recall Exercise 3.4.32. Consider the set $U := \pi^{-1}(B_d(o, 1))$. Then $\pi(U^n) \subset B_d(o, n)$ and $\mu_G(U^n) \le \mu_G(\pi^{-1}(\pi(U^n))) = \mu_M(\pi(U^n)) \le \mu_M(B_d(o, n)) \le Cn^Q$.

Exercise 6.6.27 For $z \in \mathbb{C} \setminus i\mathbb{R}$, let G_z be the Lie group from Example 6.4.4. Then, the center of G_z is trivial, and G_z has exponential growth.
Hint. See Exercise 8.3.10

Exercise 6.6.28 Normal discrete subgroups of connected topological groups are central.

Exercise 6.6.29 (Big Separated Net Gives Big Growth) Let G be a connected Lie group equipped with a left-invariant admissible geodesic distance. Let $\rho(n)$ be the maximal cardinality for a 2-separated set in the ball $B(1_G, n)$, for $n \in \mathbb{N}$. Then $\rho(n)\mu(B(1_G, 1)) \le \mu(B(1_G, 1)^n) \le \rho(n)\mu(B(1_G, 2))$, for all $n \in \mathbb{N}$.
Hint. The distance being geodesic implies $B(1_G, 1)^n = B(1_G, n)$. If two points have a distance of at least 2, their balls of radius 1 are disjoint.

Exercise 6.6.30 Let M be a metric space that admits a dilation $\delta_\lambda : M \to M$ of factor $\lambda > 1$ with a fixed point o. If M is locally connected (resp., locally simply connected; resp., locally compact; resp., locally contractible), then M is connected (resp., simply connected; resp., boundedly compact; resp., contractible).
Hint. Take a 'good' neighborhood U of o and $r > 0$ such that $B(o, r) \subset U$. Then, we have $M = \bigcup_{n=1}^{\infty} B(o, \lambda^n r) = \bigcup_{n=1}^{\infty} \delta_\lambda^n(B(o, r)) \subseteq \bigcup_{n=1}^{\infty} \delta_\lambda^n(U)$. Regarding contractibility, all homotopy groups are trivial, and one can use the Whitehead Theorem.

Exercise 6.6.31 Let (Y, d_Y) be a locally compact metric space and $y_0 \in Y$. Assume that, for all $\lambda > 0$, there is an isometry $f : (Y, d_Y) \to (Y, \lambda d_Y)$ with $f(y_0) = y_0$. Then Y is connected. In fact, every closed metric ball at y_0 is connected.
Solution. We begin with a technical claim:

$$\forall \epsilon > 0, \forall \bar{y} \in Y, \text{ with } \bar{y} \neq y_0 \quad \exists y' \in Y \text{ such that } \begin{cases} d(\bar{y}, y') < \epsilon \\ d(y', y_0) < d(\bar{y}, y_0). \end{cases} \tag{6.12}$$

For proving (6.12), for each $\delta \in (0, 1]$, we denote $\delta Y := (Y, \delta d_Y)$ and consider the set

$$\Omega_\delta := \{f(\bar{y}) \mid f : \delta Y \to Y \text{ isometry with } f(y_0) = y_0\}.$$

We have $\Omega_\delta \subseteq \{y \in Y : d(y_0, y) = \delta d(y_0, \bar{y})\}$. The assumption that $(\delta Y, y_0)$ is isometric to (Y, y_0) rephrases as $\Omega_\delta \neq \emptyset$. The space is boundedly compact by Exercise 6.6.30, so we use Ascoli–Arzelà's Theorem, Exercise 3.4.12. For $\delta \in (1/2, 1)$, pick $z_\delta \in \Omega_\delta$ and a respective δ-homothety f_δ. Using the Ascoli–Arzelà argument to the uniformly Lipschitz maps f_δ, we have that there exists a sequence $\delta_n \nearrow 1$ for which the maps f_{δ_n} converge uniformly on compact sets to an isometry f. In particular, $f(y_0) = y_0$ and $f_{\delta_n}(\bar{y}) \to f(\bar{y})$, as $n \to \infty$. Set $y_n := f^{-1}(f_{\delta_n}(\bar{y}))$. Observe that

$$d(\bar{y}, y_n) = d(f(\bar{y}), f_{\delta_n}(\bar{y})) \to 0, \qquad \text{as } n \to \infty,$$

and

$$\begin{aligned} d(y_0, y_n) &= d(f(y_0), f_{\delta_n}(\bar{y})) = d(y_0, f_{\delta_n}(\bar{y})) \\ &= d(f_{\delta_n}(y_0), f_{\delta_n}(\bar{y})) = \delta_n d(y_0, \bar{y}) < d(y_0, \bar{y}). \end{aligned}$$

Thus, for n large enough, one can take y' as y_n for obtaining (6.12).

To deduce Exercise 6.6.31, assume by contradiction that some closed ball $\bar{B}(y_0, r)$ is not connected. Thus, there are non-empty closed sets $K_1, K_2 \subseteq \bar{B}(y_0, r)$ such that $K_1 \cap K_2 = \emptyset$ and $K_1 \cup K_2 = \bar{B}(y_0, r)$. Say $y_0 \in K_1$. Since Y is boundedly compact (see Exercise 6.6.30), both K_1 and K_2 are compact. Consequently, first, the value $\epsilon := d(K_1, K_2)$ is strictly positive. Second, there exists $\bar{y} \in K_2$ such that $d(\bar{y}, y_0) = d(K_2, y_0)$. By (6.12), there is some $y' \in Y$ such that $d(\bar{y}, y') < \epsilon$ and $d(y', y_0) < d(\bar{y}, y_0)$. By the second inequality, we get that $y' \notin K_2$ and that $y' \in \bar{B}(y_0, r)$. By the first inequality, we get that $y' \notin K_1$. We contradicted the fact that $K_1 \cup K_2 = \bar{B}(y_0, r)$. Hence, the set $\bar{B}(y_0, r)$ is connected. By Exercise 6.6.30, the metric space Y is connected.

Chapter 7
Sub-Finsler Lie Groups

This chapter is fundamental; it is the core of this book. We consider sub-Finsler Lie groups and discuss their differential geometry and their metric geometry when equipped with Carnot-Carathéodory distances. In Sect. 7.1, we discuss the perspective of left-invariant sub-Finsler structures. We show how, in sub-Finsler geometry, quotients can be viewed as submetries, allowing for geodesic lifting. We discuss Chow's theorem and establish a weakened version of the Ball-Box Theorem, which will be further refined in Theorem 12.5.3. In Sect. 7.2, we discuss the endpoint map and its singular elements known as abnormal curves. In Sect. 7.3, in the framework of sub-Riemannian groups, we conduct a first-order analysis of geodesics, leading to the Pontryagin Maximum Principle. In the final Sect. 7.4, we demonstrate that geodesic homogeneous manifolds are Carnot-Carathéodory spaces.

7.1 Left-Invariant Sub-Finsler Structures on Lie Groups

A standard assumption in the geometry of Lie groups is that the objects under consideration, such as distributions and sub-Finsler structures, are left-invariant. As a result, every considered distribution is a polarization, meaning that it has a constant rank. Similarly, as for Lie algebras of Lie groups, we will have two interpretations for polarizations and two for continuously varying norms due to the left-invariance.

7.1.1 Left-Invariant Polarizations and Horizontal Curves

In this section, we set-wise interpret 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

E. Le Donne, *Metric Lie Groups*, Graduate Texts in Mathematics 306,
https://doi.org/10.1007/978-3-031-98832-5_7

algebra $\mathfrak{g} = \mathrm{Lie}(G)$. Let $\Delta \subset TG$ be a distribution, as in Definition 4.1.1. The distribution 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. Each $(\mathrm{d}L_g)_h$ is an isomorphism, and left translations act transitively; thus, the dimension of the subspace Δ_g is independent of $g \in G$. In other words, every left-invariant distribution is a polarization, following the terminology of Definition 4.1.1.

Every left-invariant distribution $\Delta \subset TG$ determines a vector subspace $V := \Delta_{1_G} \subseteq \mathfrak{g}$. Conversely, every vector subspace $V \subseteq \mathfrak{g}$ of $\mathfrak{g}$ determines a left-invariant distribution Δ, given by $\Delta_{1_G} := V$ and

$$\Delta_g := \left\{ v \in T_g G : (\mathrm{d}L_g)^{-1} v \in V \right\}, \qquad \forall g \in G. \tag{7.1}$$

Observe, indeed, that the set $\Delta \subseteq TG$ is left-invariant and a polarization, whose rank equals $\dim(V)$, for which there is a global frame; see Exercise 7.5.1. In essence, there is a one-to-one correspondence between vector subspaces of $\mathrm{Lie}(G)$ and left-invariant distributions on G.

Definition 7.1.1 (Polarized Group) Given a Lie group G and a vector subspace $V \subseteq \mathrm{Lie}(G)$, we say that the pair (G, V) forms a *polarized Lie group*, or simply, a *polarized group*, and we refer to the distribution Δ defined in (7.1) as the *induced (left-invariant) distribution*. We can also refer to V as a *polarization*, equating it with Δ. Similarly, the pair $(\mathrm{Lie}(G), V)$ is called a *polarized Lie algebra*.

Because every left-invariant distribution comes from a vector subspace of the Lie algebra, determining whether the distribution is bracket generating is straightforward. Indeed, we have two equivalent ways of calculating the iterated brackets and the flag of subbundles of Definition 4.1.13. It should not be surprising that each left-invariant distribution is equiregular, according to Definition 4.1.15, because the flag of subbundles preserves the symmetry of being left-invariant.

In the following, given two subsets U, V of a Lie algebra $\mathfrak{g}$, we use the notation

$$[U, V] := \mathrm{span}\big\{[u, v] : u \in U, v \in V\big\} \subseteq \mathfrak{g}.$$

Furthermore, for every vector subspace $V \subseteq \mathfrak{g}$, we iteratively define:

$$V^{[1]} := V, \qquad V^{[k]} := V^{[k-1]} + \left[V, V^{[k-1]}\right], \qquad \forall k = 2, 3, \ldots. \tag{7.2}$$

A first observation is that, for each $k \in \mathbb{N}$, the left-invariant distribution induced by $V^{[k]}$ is the k-th element $\Delta^{[k]}$ in the flag of subbundles associated with Δ as in Definition 4.1.13 (see Exercise 7.5.2). A second observation is that to determine 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, we have:

$$(\mathrm{Lie}(\Gamma(\Delta)))_1 = \bigcup_{k\in\mathbb{N}} V^{[k]} \qquad \text{and} \qquad (\mathrm{Lie}(\Gamma(\Delta)))_g = (\mathrm{d}L_g)_1(\mathrm{Lie}(\Gamma(\Delta)))_1.$$

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 $\{1, 2, \dots, n\}$, where $n := \dim(G)$. Actually, unless $n \leq 1$, we must have $\dim(V) > 1$ for Δ to be bracket generating. Thus, if Δ is not bracket generating then there exists $\bar{k} < n$ such that $V^{[\bar{k}]} = V^{[l]}$ for every $l \geq k$; see Exercise 7.5.3. We have proved the following result, which explains when a left-invariant distribution satisfies the bracket-generating condition, (4.2).

Proposition 7.1.2 (Criterion for Bracket Generation) *If (G, V) is a polarized Lie group of dimension $n \geq 2$, 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 $V^{[n-1]} \neq \mathfrak{g}$, in which case 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 TH. Here, $\Delta|_H := \{v \in \Delta \ : \ \pi(v) \in H\}$, where $\pi : TG \to G$ is the bundle projection.

Moreover, denoting by $\bar{k}$ the smallest integer for which $V^{[\bar{k}]} = V^{[\bar{k}+1]}$, if $V^{[\bar{k}]} \neq \mathfrak{g}$, then $\bar{k} < n - 1$.

We can further rephrase the notion of horizontal curve as from Definition 4.1.4. In a polarized group (G, V) with induced distribution Δ, an absolutely continuous curve $\gamma : I \to G$ defined on an interval I is Δ-horizontal if

$$\gamma'(t) := (\mathrm{d}L_{\gamma(t)})^{-1}\dot{\gamma}(t) \in V, \qquad \text{for almost every } t \in I. \tag{7.3}$$

Notice that we have just defined a V-valued curve $\gamma' : I \to V$ not to be confused with the derivative $\dot{\gamma} : I \to TG$, which is TG-valued. However, the curve γ', together with the initial point, maintains the whole information about γ, as we next see.

Proposition 7.1.3 (Integration of Tangent Vectors) *Let G be a Lie group, let $[a, b] \subseteq \mathbb{R}$ be an interval, and let $u : [a, b] \to \mathfrak{g}$ be integrable. Then, for every $p \in G$ there exists a unique absolutely continuous curve $\gamma : [a, b] \to G$ such that $\gamma(a) = p$ and $u = \gamma'$, where the latter is defined in (7.3).*

Proof We consider the ODE

$$\begin{cases} \dot{\gamma}(t) = (L_{\gamma(t)})_* u(t) \\ \gamma(a) = p. \end{cases} \tag{7.4}$$

The existence of an absolutely continuous solution to the ODE is a consequence of the general Carathéodory's theorem; see [O'R97, Theorem 3.4]. We stress that Lie groups equipped with Riemannian left-invariant metrics are complete; see Lemma 6.2.4. Consequently, short-time solutions stay in compact sets. Hence, we have global solutions; see [CL55, page 47, Chapter 2, Theorem 1.3]. 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 fact can be easily shown using Exercises 5.8.11 and 5.8.12. □

Definition 7.1.4 (Development of a Curve) Let $T \geq 0$ and $\gamma : [0, T] \to G$ be an absolutely continuous curve in a Lie group G. The *development* of γ is the curve $\sigma : [0, T] \to \mathfrak{g}$ defined as

$$\sigma(t) := \int_0^t \gamma'(s)\,\mathrm{d}s, \qquad \forall t \in [0, T],$$

where $\gamma' : [0, T] \to \mathfrak{g}$ is as in (7.3).

The above curve σ is valued in the vector space $\mathfrak{g}$, where we identify σ' with $\dot{\sigma}$. Thus, we have the identity

$$\sigma' = \gamma'.$$

We stress that if γ is Δ-horizontal, for some left-invariant polarization Δ, then σ is Δ_1-valued.

We also recall that by Proposition 7.1.3, every absolutely continuous curve $\sigma : [0, T] \to \mathfrak{g}$ is the development of some absolutely continuous curve $\gamma : [0, T] \to G$. We shall refer to this curve as the multiplicative integral.

Definition 7.1.5 (Multiplicative Integral) Let G be a Lie group with Lie algebra $\mathfrak{g}$, and $T > 0$. Let $\sigma : [0, T] \to \mathfrak{g}$ be an absolutely continuous curve. The absolutely continuous curve $\gamma : [0, T] \to G$ with $\gamma(0) = 1_G$ and $\sigma' = \gamma'$, which exists and is unique by Proposition 7.1.3, is called the *multiplicative integral* of σ.

Up to only considering curves starting at 1_G or $0_{\mathfrak{g}}$, respectively, the multiplicative integral is the inverse operation of the development of a curve, as of Definition 7.1.4.

7.1.2 Left-Invariant Norms and Distances on Lie Groups

The first aim of this subsection is to clarify that for Lie groups, left-invariant continuously varying norms are in one-to-one correspondence with symmetric norms on the Lie algebra. A function $N : TG \to \mathbb{R}$ on the tangent bundle of a Lie group G is said to be *left-invariant* if $N \circ \mathrm{d}L_g = N$ for all $g \in G$. In fact, via left translations, the tangent bundle TG is trivializable as $G \times \mathfrak{g}$ and, in the left-invariant case, every continuously varying norm $N : TG \simeq G \times \mathfrak{g} \to \mathbb{R}$, $(g, v) \mapsto N(g, v)$, will be constant with respect to the point g and will be a norm in the vector v.

One can check that if a function $\|\cdot\| : TG \to \mathbb{R}$ 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 3.2.6. 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 a symmetric norm on $\mathfrak{g}$, then

$$\|v\| := \|(\mathrm{d}L_{g^{-1}})_g v\|_{1_G}, \quad \forall g \in G, \ \forall v \in T_g G, \tag{7.5}$$

defines a left-invariant continuously varying norm (Exercise 7.5.4).

Definition 7.1.6 (Sub-Finsler Lie Group) A *sub-Finsler Lie group* is a triple $(G, V, \|\cdot\|)$ where G is a connected Lie group, V is a bracket-generating subspace of $T_{1_G}G$, and $\|\cdot\|$ is a symmetric norm on V. Every sub-Finsler Lie group is naturally seen as a Carnot-Carathéodory space where the distribution Δ is the induced distribution of the polarized group (G, V) as in (7.1) and $\|\cdot\|$ is extended on TG by first arbitrarily extending $\|\cdot\|$ as a norm on $T_{1_G}G$ and then as a left-invariant continuously varying norm by (7.5).

In the definition of sub-Finsler Lie groups, the left-invariant continuously varying norm restricts to Δ as a function $\Delta \to \mathbb{R}$. This function is the left-invariant extension of the norm on V. We still call it *continuously varying norm*.

Every sub-Finsler Lie group has an associated sub-Finsler metric, as in (4.4), which can be formulated using the above double viewpoint for left-invariant structures. Moreover, we recall Proposition 3.1.11 about the energy of curves. Thus, the Carnot-Carathéodory distance d_{cc} between two points $p, q \in G$ is

$$\begin{aligned} d_{cc}(p, q) &:= \inf \left\{ \mathrm{Length}_{\|\cdot\|}(\gamma) : \gamma \ \Delta\text{-horizontal curve from } p \text{ to } q \right\} \\ &= \inf \left\{ \int \|\gamma'\|_{1_G} : \gamma \text{ AC curve from } p \text{ to } q \text{ with } \gamma' \in V \right\} \\ &= \inf \left\{ \sqrt{\int_0^1 \|\gamma'\|_{1_G}^2} : \gamma \text{ AC curve on } [0,1] \text{ from } p \text{ to } q \text{ with } \gamma' \in V \right\}. \end{aligned}$$

In Definition 7.1.6, we made the choice of assuming that the polarization is bracket generating. Of course, one could also consider sub-Finsler metrics associated with non-bracket-generating polarizations. However, because of Proposition 7.1.2, if the polarization is not bracket generating, then one can just restrict to the Lie subgroup that it generates, as by Theorem 5.7.1.

The following are some basic metric properties of sub-Finsler Lie groups when they are seen as metric spaces with their Carnot-Carathéodory metrics.

Theorem 7.1.7 *Every sub-Finsler Lie group is a metric space that is*

7.1.7.i. *complete,*
7.1.7.ii. *geodesic,*
7.1.7.iii. *boundedly compact,*
7.1.7.iv. *isometrically homogeneous: the distance is left-invariant.*

Proof Let (G, d) be a sub-Finsler Lie group equipped with its CC distance. We begin by observing that (G, d) is isometrically homogeneous since the group of left translations acts transitively and by maps that preserve the family of horizontal curves and their length.

Then, since the distance induces the manifold topology, there exists $r > 0$ such that the closed metric ball $\bar{B}(1_G, r)$ is compact. By homogeneity, the metric space is locally compact and complete.

We conclude by invoking Proposition 3.1.6 or Theorem 3.1.7, or Theorem 4.2.6. □

Proposition 7.1.8 *If (G, V) is a polarized Lie group, then every two left-invariant CC distances induced by norms on V are bi-Lipschitz equivalent, globally.*

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 bi-Lipschitz equivalent to every other. This produces a bi-Lipschitz equivalence for CC distances. □

Because of Proposition 7.1.8, whenever our interest is in metric spaces up to bi-Lipschitz equivalence, we may assume that the norm $\|\cdot\|$ is coming from a scalar product.

7.1.3 Quotients and Submetries Between Sub-Finsler Lie Groups

Proposition 7.1.9 *Let $(G, \Delta, \|\cdot\|)$ be a sub-Finsler Lie group and $H < G$ a closed subgroup. Then, the distance function from (6.5) on the quotient manifold $H\backslash G := \{Hg : g \in G\}$, which makes the projection $\pi : G \to H\backslash G$ a submetry, is a sub-Finsler distance, where the distribution Δ on $H\backslash G$ is the only one such that*

$$\Delta_p = \mathrm{d}\pi(\Delta_{\tilde{p}}), \qquad \forall p \in H\backslash G, \forall \tilde{p} \in \pi^{-1}(p),$$

with norm

$$\begin{aligned} \|v\| &:= \inf\{||w|| : \tilde{p} \in \pi^{-1}(p), w \in T_{\tilde{p}}\Delta, \mathrm{d}\pi(w) = v\}, \qquad \forall v \in \Delta_p, \\ &= \inf\{||w|| : w \in T_{\tilde{p}}\Delta, \mathrm{d}\pi(w) = v\}, \qquad \forall \tilde{p} \in \pi^{-1}(p), \forall v \in \Delta_p. \quad (7.6) \end{aligned}$$

Proof We begin by stressing that the distribution is well defined, and the definition of the norm has the property (7.6) because the group H acts by isometries and transitively on fibers. Moreover, by construction we have that the map $(\mathrm{d}\pi)_{\tilde{p}} : (\Delta_{\tilde{p}}, \|\cdot\|) \to (\Delta_{\pi(\tilde{p})}, \|\cdot\|)$, for $\tilde{p} \in G$, is a submetry.

We denote by d_{cc} the sub-Finsler metric. Let d be the distance function on ${}_H\backslash^G$ given by (6.5), so that $\pi : (G, d_{cc}) \to \left({}_H\backslash^G, d\right)$ is a submetry, by Proposition 6.3.4 and the fact that G is boundedly compact by Theorem 7.1.7. The plan is to use Proposition 3.1.28 to show that also the map $\pi : (G, d_{cc}) \to \left({}_H\backslash^G, d_{cc}\right)$ is a submetry. Since there is a unique distance function on the target for which a map is a submetry, we would deduce that the distance functions d and d_{cc} on ${}_H\backslash^G$ coincide.

For the argument, we will make use of a measurable selection theorem, which we review at the end of this proof; see Theorem 7.1.10. Given $p \in {}_H\backslash^G$, if we take two preimages $\tilde{p}_1$ and $\tilde{p}_2 \in \pi^{-1}(p)$, then the maps $\pi \circ L_{\tilde{p}_1}$ and $\pi \circ L_{\tilde{p}_2}$ coincide, and we denote such a map by $\hat{\pi}_p$. Consider the multifunction Ψ taking each element $w \in \Delta_p$, with $p \in {}_H\backslash^G$, to the set

$$\Psi(w) := \{v \in \Delta_{1_G} : (\hat{\pi}_p)_*(v) = w,\ \|v\| = \|w\|\}.$$

Note that for every $w \in \Delta_p$, the set $\Psi(w)$ is a closed subset of Δ_{1_G}, which is nonempty because of the definition of the norm on ${}_H\backslash^G$. Let U be an open subset of Δ_{1_G}. We have

$$\begin{aligned}
&\left\{w \in \Delta \subseteq T({}_H\backslash^G) :\ \Psi(w) \cap U \neq \emptyset\right\} \\
&\quad = \left\{w :\ \{v \in \Delta_1 :\ (\hat{\pi}_p)_*(v) = w \in \Delta_p, p \in {}_H\backslash^G,\ \|v\| = \|w\|\} \cap U \neq \emptyset\right\} \\
&\quad = \left\{(\hat{\pi}_p)_*(v) \in \Delta_p :\ v \in U, p \in {}_H\backslash^G,\ \|v\| = \|(\hat{\pi}_p)_*(v)\|\right\} \\
&\quad = \pi_*\left(\left\{X_{\tilde{p}} \in \Delta^G :\ (L_{\tilde{p}})^* X_{\tilde{p}} \in U,\ \|X_{\tilde{p}}\| = \|\mathrm{d}\pi X_{\tilde{p}}\|\right\}\right).
\end{aligned}$$

First, notice that the set $\left\{X_{\tilde{p}} \in \Delta :\ (L_{\tilde{p}})^* X_{\tilde{p}} \in U\right\}$ is open in Δ. Second, notice that the set $\left\{X_{\tilde{p}} \in \Delta :\ \|X_{\tilde{p}}\| = \|\mathrm{d}\pi X_{\tilde{p}}\|\right\}$ is closed. Hence, they and their intersection are a countable union of compact sets. The image under π_*, which is a continuous map, is again a countable union of compact sets. Hence, it is a Borel set. Thus we can apply Theorem 7.1.10: there exists a Borel measurable selection $\psi : \Delta \subseteq T\left({}_H\backslash^G\right) \to \Delta_{1_G}$ of Ψ. We observe that the constructed selection map ψ has the following properties:

$$(\hat{\pi}_p)_*(\psi(w)) = w \quad \text{and} \quad \|\psi(w)\| = \|w\|, \qquad \forall p \in {}_H\backslash^G, \forall w \in \Delta_p. \tag{7.7}$$

To use Proposition 3.1.28, we consider a horizontal curve $\gamma : [0, T] \to {}_H\backslash^G$ and $\tilde{g} \in \pi^{-1}(\gamma(0))$ and we want to construct a curve γ with the properties 3.1.28.2.i-iii. Notice that $\psi \circ \dot{\gamma} : [0, T] \to \Delta_{1_G}$ is measurable. We consider the curve $\tilde{\gamma} : [0, T] \to G$ solution of

$$\begin{cases} \tilde{\gamma}(0) = \tilde{g} \\ \dot{\tilde{\gamma}}(t) = (\mathrm{d}L_{\tilde{\gamma}(t)})\psi(\dot{\gamma}(t)), \end{cases}$$

which exists by Proposition 7.1.3. The curve γ satisfies 3.1.28.2.i by construction. Regarding 3.1.28.2.ii, the two curves $\pi \circ \tilde{\gamma}$ and γ are equal because they have the same initial point and same derivatives:

$$\frac{\mathrm{d}}{\mathrm{d}t}(\pi \circ \tilde{\gamma}) = \pi_*(\mathrm{d}L_{\tilde{\gamma}(t)})\psi(\dot{\gamma}) = (\pi_{\gamma(t)})_*\psi(\dot{\gamma}) \overset{(7.7)}{=} \dot{\gamma}.$$

Regarding 3.1.28.2.iii, the two curves have the same length because they have the same speed:

$$\left\|\dot{\tilde{\gamma}}\right\| = \left\|(\mathrm{d}L_{\tilde{\gamma}(t)})\psi(\dot{\gamma})\right\| = \|\psi(\dot{\gamma})\| \overset{(7.7)}{=} \|\dot{\gamma}\|.$$

Finally, Condition 3.1.28.1 is satisfied since π is 1-Lipschitz; see Exercise 4.4.23. By Proposition 3.1.28, we deduce that the two distances d and d_{cc} on ${}_H\backslash^G$ are the same. □

In the above proof, we used the following general result by Kuratowski and Ryll-Nardzewski:

Theorem 7.1.10 ([KR65]) *Let X be a separable completely metrizable topological space, let $\mathcal{B}(X)$ be the Borel σ-algebra of X, let $(\Omega, \mathcal{F})$ be a measurable space, and Ψ a multifunction on Ω taking values in the set of nonempty closed subsets of X. If for every open subset U of X we have*

$$\{\omega : \Psi(\omega) \cap U \neq \emptyset\} \in \mathcal{F},$$

then Ψ has a selection that is $\mathcal{F}$-$\mathcal{B}(X)$-measurable.

Remark 7.1.11 When the distribution Δ of the sub-Finsler group G is such that $\Delta_{1_G} \oplus \mathrm{Lie}(H) = \mathrm{Lie}(G)$, then the existence of the selection in Theorem 7.1.10 is a triviality. In fact, the map $\mathrm{d}\pi|_{\Delta_p} : \Delta_p \to \Delta_{\pi(p)}$ is a linear isomorphism, hence we take $\psi := (\mathrm{d}\pi|_\Delta)^{-1}$. Moreover, in this case, the lifted curves $\tilde{\gamma}$ are unique, given the initial point.

Remark 7.1.12 Let H be a sub-Finsler Lie group and G a Lie group. Suppose that $\pi : G \to H$ is a surjective Lie homomorphism. Then, there is a sub-Finsler structure on G such that π is a submetry. In fact, if we take a bracket-generating polarization $V \subseteq T_1G$ such that $\pi_* V = \Delta_1^H$, e.g., $\pi_*^{-1}\Delta_1^H$, then there is a norm on V such that π is a submetry. The unit ball for such a norm can be taken as the intersection of the preimage of the unit ball on Δ_1^H and a strip that is transverse to the kernel of π.

Remark 7.1.13 Let $\pi : G \to H$ be a submetry and a Lie homomorphism between sub-Finsler Lie groups. Thus, by Proposition 3.1.29, geodesics in H can be lifted to geodesics in G. In particular, if all geodesics on G are smooth, then so are the geodesics in H.

7.1.3.1 Open Questions

The following are some unsolved questions on sub-Finsler Lie groups.

Question 7.1.14 Let G be a Lie group. If ρ is a left-invariant metric that is boundedly compact, locally bounded, and quasi-geodesic, is ρ at a bounded distance from a sub-Finsler left-invariant metric?

Question 7.1.15 Let $(G, \Delta, \|\cdot\|)$ be a sub-Finsler Lie group. If the unit ball of $(\Delta, \|\cdot\|)$ is polyhedral (i.e., it is the convex hull of finitely many points), then does there exist $K \in \mathbb{N}$ such that between every pair of points in G there exists a geodesic whose control is piecewise constant with at most K pieces?

Question 7.1.16 Let $(G, \Delta, \|\cdot\|)$ be a sub-Finsler Lie group. If the unit ball of $(\Delta, \|\cdot\|)$ is strictly convex, then does there exist, for every p and $q \in G$, a continuously differentiable geodesic from p to q?

Regarding Question 7.1.16, a very recent preprint, [Chi+25], exhibits an example that implies that there are sub-Riemannian Lie groups where some pairs of points cannot be joined by any geodesic of class C^3. Additionally, some preliminary work by Lev Lokutsievskiy casts doubt on a positive answer to Question 7.1.15.

7.1.4 A Direct, Effective Proof of Chow's Theorem

In this section, we will give an explicit construction of a horizontal path connecting an arbitrary point p in a bracket-generating polarized group (G, V) to the identity element 1_G. Moreover, when a norm is fixed on an s-step polarization, the path will have a length bounded by the $1/s$-power of some fixed Riemannian distance between p and 1_G. This is a weaker version of the Ball-Box Theorem: (4.21).

7.1.4.1 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 X and Y on a manifold M, 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 as in 3.2.2.d:

$$[X, Y]_p = \frac{1}{2}\frac{\mathrm{d}^2}{\mathrm{d}t^2}\left(\Phi_Y^{-t} \circ \Phi_X^{-t} \circ \Phi_Y^{t} \circ \Phi_X^{t}\right)(p)\bigg|_{t=0}, \qquad \forall p \in M.$$

In the above formula, we denote by Φ_X^t the flow of the vector field X at time t. By Corollary 5.2.6, for left-invariant vector fields in a Lie group G, we have

$$\Phi_X^t(p) = pe^{tX}, \qquad \forall p \in G, \forall t \in \mathbb{R}, \forall X \in \mathfrak{g},$$

where, for the sake of clarity, we also use the notation $e^X := \exp(X)$ for $X \in \mathfrak{g}$. Hence, for all X and Y in the Lie algebra $\mathfrak{g}$ of the Lie group G, we have

$$[X, Y]_{1_G} = \frac{1}{2}\frac{\mathrm{d}^2}{\mathrm{d}t^2} e^{tX}e^{tY}e^{-tX}e^{-tY}\bigg|_{t=0}, \qquad \forall X, Y \in \mathfrak{g}. \tag{7.8}$$

In fact, we can obtain a C^1 curve whose tangent at a point is $[X, Y]$. Indeed, as in the general case of manifolds seen in Exercise 3.4.49, we can consider the curve

$$\gamma(t) := \begin{cases} e^{\sqrt{t}X}e^{\sqrt{t}Y}e^{-\sqrt{t}X}e^{-\sqrt{t}Y} & \text{for } t \geq 0 \\ e^{\sqrt{|t|}X}e^{-\sqrt{|t|}Y}e^{-\sqrt{|t|}X}e^{\sqrt{|t|}Y} & \text{for } t < 0, \end{cases}$$

for which $\dot{\gamma}(0) = [X, Y]_{1_G}$.

Definition 7.1.17 (Maps P_t) Given a Lie group G, with Lie algebra $\mathfrak{g}$, $X, Y \in \mathfrak{g}$, and $t \in \mathbb{R}$, we define

$$P_t(X) := e^{tX} \quad \text{and} \quad P_t(X, Y) := e^{tX}e^{tY}e^{-tX}e^{-tY}.$$

By recurrence, for $k \geq 2$ and $X_1, \ldots, X_{k+1} \in \mathfrak{g}$, and $t \in \mathbb{R}$, we also define

$$P_t(X_1, \ldots, X_{k+1}) := P_t(X_1, \ldots, X_k)e^{tX_{k+1}}(P_t(X_1, \ldots, X_k))^{-1}e^{-tX_{k+1}}.$$

Remark 7.1.18 As an immediate consequence of (7.8) (or of BCH (5.26)), we have

$$P_t(X, Y) = \exp\left(t^2[X, Y] + o(t^2)\right), \qquad \text{as } t \to 0.$$

More generally, if we consider $X_1, \dots, X_d \in \mathfrak{g}$, with $d \in \mathbb{N}$, then we claim that

$$P_t(X_1, \dots, X_d) = \exp\left(t^d[X_1, \dots, X_d] + o(t^d)\right), \qquad \text{as } t \to 0, \tag{7.9}$$

where we use the notation

$$[X_1, X_2, X_3, \dots, X_d] := [\dots[[X_1, X_2], X_3], \dots, X_d].$$

Indeed, we proceed by induction. The statement is true for $d \in \{1, 2\}$. Assume it is true for an arbitrary d. Call $\omega(t)$ the $o(t^d)$ function such that $P_t(X_1, \dots, X_d) = e^{t^d[\cdots[[X_1,X_2],X_3],\dots,X_d]+\omega(t)}$. Then we have, by the BCH Formula (5.26),

$$\begin{aligned}
&P_t(X_1, \dots, X_{d+1})\\
&\quad= P_t(X_1, \dots, X_d)e^{tX_{d+1}}\left(P_t(X_1, \dots, X_d)\right)^{-1}e^{-tX_{d+1}}\\
&\quad= e^{t^d[\dots[X_1,X_2],\dots,X_d]+\omega(t)}e^{tX_{d+1}}\left(e^{t^d[\dots[X_2,X_1],\dots,X_d]+\omega(t)}\right)^{-1}e^{-tX_{d+1}}\\
&\quad= e^{t^d[\dots[X_1,X_2],\dots,X_d]+\omega(t)}e^{tX_{d+1}}e^{-t^d[\dots[X_2,X_1],\dots,X_d]-\omega(t)}e^{-tX_{d+1}}\\
&\overset{(5.26)}{=} e^{\left(tX_{d+1}+t^d[\dots[X_1,X_2],\dots,X_d]+\omega(t)+\frac{1}{2}t^{d+1}[\dots[X_1,X_2],\dots,X_{d+1}]+o(t^{d+1})\right)}\\
&\qquad\qquad \cdot e^{\left(-tX_{d+1}-t^d[\dots[X_1,X_2],\dots,X_d]-\omega(t)+\frac{1}{2}t^{d+1}[\dots[X_1,X_2],\dots,X_{d+1}]+o(t^{d+1})\right)}\\
&\overset{(5.26)}{=} e^{t^{d+1}[\dots[[X_1,X_2],X_3],\dots,X_{d+1}]+o(t^{d+1})}.
\end{aligned}$$

Each P_t is, in fact, a product of elements of the form $e^{\pm tX_i}$. We now form a map that will help in constructing horizontal paths.

Definition 7.1.19 (Map E) In the setting and notation of Definition 7.1.17, let $\{X_{j,k}\}_{j,k} \subset \mathfrak{g}$, with $j \in \{1, \dots, n\}$, $k \in \{1, \dots, d_j\}$, and $d_1, \dots, d_n \in \mathbb{N}$. For each j, using the maps P_t from above, we consider

$$\Phi^{(j)}(t) := \begin{cases} P_{t^{1/d_j}}(X_{j,1}, \dots, X_{j,d_j-1}, X_{j,d_j}) & \text{for } t \geq 0, \\ P_{|t|^{1/d_j}}(X_{j,1}, \dots, X_{j,d_j-1}, -X_{j,d_j}) & \text{for } t < 0. \end{cases}$$

We finally define the map $E : \mathbb{R}^n \to G$ associated with $\{X_{j,k}\}_{j,k}$ as

$$E(\mathbf{t}) := \prod_{j=1}^{n} \Phi^{(j)}(t_j) \quad \forall \mathbf{t} \in \mathbb{R}^n.$$

Each map $\Phi^{(j)}$ as above is C^1, see Exercise 7.5.13. Moreover, the constructed map E satisfies the following properties.

Proposition 7.1.20 *Let G be a Lie group of dimension $n \in \mathbb{N}$. Let $\{X_{j,k} | j \in \{1, \dots, n\},\ k \in \{1, \dots, d_j\}\}$ with $d_j \in \mathbb{N}$ be a family of vectors in* $\mathrm{Lie}(G)$. *Assume that* $([X_{j,1}, X_{j,2}, \dots, X_{j,d_j}])_{j=1}^n$ *span* $\mathrm{Lie}(G)$.

7.1.20.i. The map $E : \mathbb{R}^n \to G$ associated with $(X_{j,k})_{j,k}$ as in Definition 7.1.19 is a local C^1-diffeomorphism around 0.

7.1.20.ii. If, in addition, the vectors $(X_{j,k})_{j,k}$ are of unit length with respect to a sub-Finsler structure on G, then there exists C such that, for all $\mathbf{t} \in \mathbb{R}^n$,

$$d_{\mathrm{cc}}(1_G, E(\mathbf{t})) \le C \sum_{j=1} |t_j|^{1/d_j}. \tag{7.10}$$

Proof Regarding 7.1.20.i, we show that $(\mathrm{d}E)_0$ is non-singular. Let

$$X_j := [[\dots[[X_{j,1}, X_{j,2}], X_{j,3}], \dots], X_{j,d_j}], \qquad \text{for } j \in \{1, \dots n\}.$$

From how E has been defined and from (7.9), we have

$$\begin{aligned}
(\mathrm{d}E)_0 \partial_j &= \frac{\mathrm{d}}{\mathrm{d}t_j} E(\mathbf{t}) \Big|_{\mathbf{t}=0} \\
&\overset{\text{def}}{=} \frac{\mathrm{d}}{\mathrm{d}t_j} \Phi^{(j)}(t_j) \Big|_{t_j=0} \\
&\overset{\text{def}}{=} \frac{\mathrm{d}}{\mathrm{d}t} P_{t^{1/d_j}}(X_{j,1}, \dots, X_{j,d_j}) \Big|_{t=0^+} \\
&\overset{(7.9)}{=} \frac{\mathrm{d}}{\mathrm{d}t} e^{t[\dots[[X_{j,1}, X_{j,2}], X_{j,3}], \dots, X_{j,d_j}] + o(t)} \Big|_{t=0} \\
&= \frac{\mathrm{d}}{\mathrm{d}t} e^{tX_j + o(t)} \Big|_{t=0} \\
&= X_j.
\end{aligned}$$

In other words, the linear map $(\mathrm{d}E)_0$ sends the basis $\partial_1, \dots, \partial_n$ to the basis $X_1, \dots, X_n$. We conclude by the Inverse Function Theorem.

Regarding 7.1.20.ii, recall, by Corollary 5.2.6, that the flow lines of a left-invariant vector field X are the curves ge^{tX}, fixed $g \in G$ and varying $t \in \mathbb{R}$. Now, since P_t is a product of exponentials, then so is E. More explicitly, we have

$$E(\mathbf{t}) = \exp(\epsilon_1 t_{c_1}^{a_1} X_{b_1}) \cdots \exp(\epsilon_N t_{c_N}^{a_N} X_{b_N}), \qquad \forall \mathbf{t} \in \mathbb{R}^n,$$

for some $N \in \mathbb{N}$ and $\epsilon_i \in \{1, -1\}$, $a_i^{-1} \in \mathbb{N}$, $b_i \in \{(j,k) : j \in \{1, \dots n\}, k \in \{1, \dots d_j\}\}$, and $c_i \in \{1, \dots n\}$, for each $i \in \{1, \dots N\}$. Notice that, if $c_i = j$, then $a_i = 1/d_j$. Now it is enough to observe that, fixed k, the point

$$g := \exp(\epsilon_1 t_{c_1}^{a_1} X_{b_1}) \cdots \exp(\epsilon_{k-1} t_{c_{k-1}}^{a_{k-1}} X_{b_{k-1}})$$

can be connected to the point

$$\exp(\epsilon_1 t_{c_1}^{a_1} X_{b_1}) \cdots \exp(\epsilon_{k-1} t_{c_{k-1}}^{a_{k-1}} X_{b_{k-1}}) \exp(\epsilon_k t_{c_k}^{a_k} X_{b_k})$$

by the path

$$g \exp(\epsilon_k s X_{b_k}), \qquad \text{for } s \in [0, |t_{c_k}^{a_k}|],$$

which is tangent to $\pm X_{b_k}$, thus horizontal with length $|t_{c_k}|^{a_k} = |t_{c_k}|^{1/d_{c_k}}$. Thus the curve obtained by joining these paths connects 1_G to $E(\mathbf{t})$ with length at most $N \sum_{j=1}^{n} |t_j|^{1/d_j}$. □

As a consequence of Proposition 7.1.20, we obtain a quantitative bound for the Carnot-Carathéodory distance, which implies the Hölder equivalence between CC and Euclidean metrics; see (4.21). In Chap. 4, we explained this weaker bound as a consequence of the more precise bound given by the Ball-Box Theorem 4.3.1. Here, we do not prove that theorem in the setting of manifolds. However, first we will soon deduce (4.21) for Lie groups with the help of the previous Proposition 7.1.20. Later, in Theorems 11.2.3 and 12.5.3, we will prove the Ball-Box Theorem for Carnot groups and for sub-Finsler Lie groups, respectively.

Corollary 7.1.21 (Weak Ball-Box Theorem) *Let G be a sub-Finsler Lie group, equipped with the associated Carnot-Carathéodory metric d_{cc} and a Riemannian distance d_R. Then there exists a neighborhood U of 1_G and $C > 0$ such that*

$$\frac{1}{C} d_R \le d_{\mathrm{cc}} \le C d_R^{1/s} \qquad \text{on } U \times U, \tag{7.11}$$

where s is the step of the sub-Finsler structure.

Proof We may assume that d_R is left-invariant. The first inequality in (7.11) is a direct consequence of the definition of CC distance; see (4.5). For the second inequality, since the sub-Finsler structure is bracket generating, then there are vectors $([X_{j,1}, X_{j,2}, \dots, X_{j,d_j}])_{j=1}^{n}$ that span $\mathrm{Lie}(G)$ and are such that each $X_{j,k}$ is horizontal and of unit length. Then, we use the associated map E from Proposition 7.1.20.ii. Firstly, by Proposition 7.1.20.i, there are $R, C_1 > 0$ such that

$$\|\mathbf{t}\| \le C_1 d_R(1_G, E(\mathbf{t})), \qquad \forall \mathbf{t} \in \mathbb{R}^n, \ \text{ with } \|\mathbf{t}\| < R.$$

Secondly, by (7.10), there is $C_2 > 0$ such that, if $\|\mathbf{t}\| < R$, then

$$d_{cc}(1_G, E(\mathbf{t})) \leq C_2 \|\mathbf{t}\|^{1/s}.$$

Let $U' := E(\{\mathbf{t} \in \mathbb{R}^n : \|\mathbf{t}\| < R\}) \subset G$, which is a neighborhood of $1_G = E(\mathbf{0})$. If $p \in U'$, then there is $\mathbf{t} \in \mathbb{R}^n$ with $\|\mathbf{t}\| < R$ and $E(\mathbf{t}) = p$. So,

$$d_{cc}(1_G, p) \leq C_1 \|\mathbf{t}\|^{1/s} \leq C_1 C_2^{1/s} d_R(1_G, p)^{1/s}.$$

Finally, we take a neighborhood U of 1_G with $U^2 \subset U'$, to obtain (7.11) by left invariance of both d_{cc} and d_R. □

7.2 Endpoint Map on Polarized Groups

In this section, we begin with a parametrization of the space of those horizontal curves in a polarized group that start from the identity element. The parameterization of this space of curves leads to a Hilbert-space structure, providing a functional-analytic framework for investigating Carnot-Carathéodory spaces.

7.2.1 Endpoint Map

Let (G, V) be a polarized group. After fixing a basis $(e_1, \dots, e_r)$ for V we can identify V with $\mathbb{R}^r$, for $r := \dim V$. We equip $\mathbb{R}^r$ with the standard 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 element $u \in \Omega$, called *control*, the *trajectory associated* with u is the curve $\gamma_u : [0, 1] \to G$ that is the solution of the ODE

$$\begin{cases} \gamma(0) = 1_G, \\ \dot{\gamma}(t) = \left(\mathrm{d}L_{\gamma(t)}\right) u(t), \quad \text{for a.e. } t \in [0, 1]. \end{cases} \tag{7.12}$$

By Carathéodory Theorem on ODEs, see Proposition 7.1.3, the equation is well posed, and in this way, each $u \in \Omega$ induces a V-horizontal curve γ_u in G. Note that γ_u is absolutely continuous because $\dot{\gamma}$ is integrable by the Cauchy-Schwarz inequality. Conversely, every V-horizontal curve on $[0, 1]$ starting from 1_G is, up to reparametrization, of the form γ_u for some $u \in \Omega$. In fact, using the notation (7.3), if γ is horizontal and reparametrized by constant speed, 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-Finsler 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 *endpoint 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 (7.12).

7.2.2 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-Finsler geometry, by analyzing this differential, one can derive necessary conditions for length-minimizing curves; see Sect. 7.3.1.

We shall not discuss why the endpoint map is smooth, for which we refer to [Rif14]. We will directly calculate its (first) differential.

Proposition 7.2.1 *Let (G, V) be a polarized group with Ω as space of controls. For every $u \in \Omega$ the differential of* End *at u is given by*

$$\begin{aligned} \mathrm{d}\,\mathrm{End}_u : \Omega &\longrightarrow T_{\mathrm{End}(u)}(G) \\ v &\longmapsto \left(\mathrm{d}R_{\gamma_u(1)}\right)_{1_G} \int_0^1 \mathrm{Ad}_{\gamma_u(t)}(v(t))\,\mathrm{d}t, \end{aligned}$$

where $\mathrm{Ad}_g : \mathfrak{g} \to \mathfrak{g}$ is defined by $\mathrm{Ad}_g := (C_g)_*$ where $C_g(h) := ghg^{-1}$; see Sect. 5.5.1.

Proof We first discuss the proof for matrix groups, i.e., we assume $G \subset \mathrm{GL}(n, \mathbb{R})$, for some $n \in \mathbb{N}$, so that the Lie product can be interpreted as matrix multiplication, allowing us to work in 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 $\epsilon = 0$:

$$\sigma(t) := \lim_{\epsilon \to 0} \frac{1}{\epsilon}\Big(\gamma_{u+\epsilon v}(t) - \gamma_u(t)\Big).$$

It can be shown (see Exercise 7.5.17) that σ satisfies the following ODE (which is the derivation with respect to ϵ of (7.12) for $\gamma_{u+\epsilon v}$)

$$\frac{\mathrm{d}\sigma}{\mathrm{d}t} = \gamma(t) \cdot v(t) + \sigma(t) \cdot u(t).$$

Also the curve $t \mapsto \int_0^t \mathrm{Ad}_{\gamma(s)}(v(s))\,\mathrm{d}s \cdot \gamma(t)$ satisfies the above differential equation; see Exercise 7.5.18. Moreover, this curve has the same initial condition as σ, therefore, it is equal to σ.

To extend the proof to arbitrary Lie groups, one may use Ado's Theorem; see Exercise 7.5.22. References for the general case of manifolds are [AS13, Section 20.3], [Mon02, Section 5.2.2], or [ABB20, Section 8.1.1]. □

7.2.3 *Singular Curves*

In this subsection, we study those controls that are critical points for the endpoint map. The associated curves are called *singular curves* or *abnormal curves*. Studying these curves is one of the main difficulties in sub-Riemannian geometry. There are plenty of unsolved questions about them.

Fix a polarized group (G, V), with Ω as space of controls. If $u \in \Omega$ is a singular point for its endpoint map, then by definition, the linear map $\mathrm{dEnd}_u : \Omega \to T_{\mathrm{End}(u)}G$ is not surjective. Namely, the subspace $\mathrm{dEnd}_u(\Omega)$ is a proper subspace of $T_{\mathrm{End}(u)}G$. In this case, there is a nontrivial covector that annihilates its image, i.e., there exists an element ξ in the dual space $(T_{\gamma_u(1)}G)^*$ such that $\xi \neq 0$ and

$$\langle \xi | \mathrm{dEnd}_u(v) \rangle = 0, \quad \forall v \in \Omega. \tag{7.13}$$

Here, we denote by $\langle \xi | w \rangle := \xi(w)$ the evaluation of a dual element $\xi \in W^*$ at a vector $w \in W$, when W is a vector space.

By Proposition 7.2.1, Eq. (7.13) is equivalent to

$$0 = \xi\left(\mathrm{d}R_{\gamma_u(1)} \int_0^1 \mathrm{Ad}_{\gamma_u(t)}\, v(t)\,\mathrm{d}t\right) = \lambda\left(\int_0^1 \mathrm{Ad}_{\gamma_u(t)}\, v(t)\,\mathrm{d}t\right), \quad \forall v \in \Omega, \tag{7.14}$$

where $\lambda \in \mathfrak{g}^* \setminus \{0\}$ is defined as $\lambda := \xi \circ \mathrm{d}R_{\gamma_u(1)}$. Fixing $t \in [0, 1]$ and $\bar{v} \in V$, let $v \in \Omega$ diverge to the distribution $\bar{v}\delta_t$ given by the Dirac mass at t multiplied by $\bar{v}$, so to obtain $\lambda\left(\mathrm{Ad}_{\gamma_u(t)}\, \bar{v}\right) = 0$. Hence, the equation

$$\lambda\left(\mathrm{Ad}_{\gamma_u(t)}\, V\right) = \{0\}, \qquad \forall t \in [0, 1], \tag{7.15}$$

is equivalent to (7.14). If $e_1, \dots, e_r$ is a basis of V, then (7.15) rephrases as a linear system of equations: A horizontal curve γ is abnormal if and only if there exists $\lambda \in \mathfrak{g}^*$ such that $\lambda \neq 0$ and

$$\lambda\left(\mathrm{Ad}_{\gamma(t)}(e_i)\right) = 0, \qquad \forall i \in \{1, \dots, r\}, \forall t \in [0, 1]. \tag{7.16}$$

We call (7.16) the *abnormal equation*. We stress that if γ satisfies (7.16) and $g \in G$, then the curve $t \mapsto g\gamma(t)$ still satisfies (7.16) with λ replaced by $\lambda\,\mathrm{Ad}_g$. We still call all such curves *abnormal curves*, even if they do not start at 1_G. Thus, left translations of abnormal curves are abnormal curves. Moreover, every restriction of an abnormal curve is an abnormal curve.

When, as in the case above, we have $\gamma(0) = 1_G$, the last equation implies

$$\lambda(e_i) = 0, \qquad \forall i \in \{1, \dots, r\}. \tag{7.17}$$

Notice that, after we fix i and λ, the function $g \mapsto \lambda\left(\mathrm{Ad}_g(e_i)\right)$ is smooth and (7.16) says that $\gamma(t)$ 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, in exponential coordinates, the map Ad is polynomial; hence, these functions are polynomials.

Remark 7.2.2 If (G, V) is a polarized group with $V \neq \mathrm{Lie}(G)$, then constant curves are abnormal, and it may happen that the only abnormal curves are constant; see Exercise 7.5.26. Or it may happen that there are non-smooth examples; see Exercise 7.5.27. In Riemannian (and Finsler) geometry, i.e., if $V = \mathrm{Lie}(G)$, there are no abnormal curves. Indeed, there is no nonzero λ as in (7.17).

Definition 7.2.3 (Abnormal Set and Sard Property) The *abnormal set* of a polarized group (G, V) is the subset $\mathrm{Abn}(G, V) \subset G$ of all singular values of the endpoint map. Equivalently, the set $\mathrm{Abn}(G, V)$ is the union of all abnormal curves passing through the identity element 1_G. If the abnormal set has measure 0, then (G, V) is said to satisfy the *Sard Property*.

Establishing the Sard Property in the general setting of polarized manifolds remains one of the major open problems in sub-Riemannian geometry; see the questions in [Mon02, Sec. 10.2] and [Agr14, Problem III].

7.3 Extrema in Sub-Riemannian Groups

7.3.1 First-Order Necessary Conditions for Sub-Riemannian Minimizers

Let G be a Lie group, and let $V \subseteq \mathfrak{g}$ be a bracket-generating subspace of the Lie algebra $\mathfrak{g}$ of G. We shall consider sub-Riemannian structures on the polarized group (G, V). We refer to these structures as *sub-Riemannian groups*. Each left-invariant sub-Riemannian structure on (G, V) is completely determined by the choice of a scalar product on V and, hence, of an orthonormal basis $(e_1, \ldots, e_r)$ for V. We next study conditions for length-minimizing curves for this sub-Riemannian structure.

Recall from Proposition 3.1.11 that minimizing the length or minimizing the energy determines the same class of curves up to reparametrization with constant speed. And this is also why we can restrict to L^2 controls, i.e., elements in $\Omega := L^2([0, 1]; V)$. Actually, because of Remark 3.1.12 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 that we saw in (3.15) for metric spaces, and now it is equal to

$$\mathrm{Energy}_{d_{cc}}(\gamma_u) \overset{\text{def}}{=} \frac{1}{2}\int \|\dot{\gamma}_u\|^2 = \frac{1}{2}\int \|\gamma_u{}'\|_{1_G}^2 = \frac{1}{2}\|u\|^2, \qquad \forall u \in \Omega,$$

using notation (7.3).

Together with the endpoint map, we form the *extended endpoint 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 of curves between 1_G and p rephrases 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

$$\mathrm{End}(v) = \mathrm{End}(u) \implies \mathrm{Energy}(v) \geq \mathrm{Energy}(u).$$

Remark 7.3.1 We claim that 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 $(\mathrm{d}\widetilde{\mathrm{End}})_{u_0} : \Omega \to T_{\widetilde{\mathrm{End}}(u)}(G \times \mathbb{R})$ at u_0 were surjective, then we can take a vector subspace $W \subset \Omega$ for which $(\mathrm{d}\widetilde{\mathrm{End}})_{u_0}|_W : W \to T_{\widetilde{\mathrm{End}}(u)}(G \times \mathbb{R})$ 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 some neighborhood of u_0 within W and some 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 an expression for the differential of the extended endpoint map $\widetilde{\mathrm{End}}$. After Proposition 7.2.1 and the standard calculation of the differential of the energy, see Exercise 7.5.28, we obtain the differential of the extended endpoint map:

Corollary 7.3.2 *Let* $\widetilde{\mathrm{End}}$ *be the extended endpoint map of a polarized group* (G, V) *with* Ω *as space of controls. For every* $u \in \Omega$ *the differential of* $\widetilde{\mathrm{End}}$ *at* u *is*

$$\begin{aligned} \mathrm{d}\widetilde{\mathrm{End}}_u : \Omega &\longrightarrow T_{\widetilde{\mathrm{End}}(u)}(G \times \mathbb{R}) = T_{\mathrm{End}(u)} G \times \mathbb{R} \\ v &\longmapsto \left(\left(\mathrm{d}R_{\gamma_u(1)}\right)_{1_G} \int_0^1 \mathrm{Ad}_{\gamma_u(t)}(v(t))\, \mathrm{d}t, \quad \langle u, v\rangle \right). \end{aligned}$$

Assume now that, for some $u \in \Omega$, the curve γ_u is length minimizing and parametrized with constant speed, so it is energy minimizing. By Remark 7.3.1, we deduce that u is a critical point for $\widetilde{\mathrm{End}}$, i.e., the linear map $\mathrm{d}\widetilde{\mathrm{End}}_u : \Omega \to T_{\mathrm{End}(u)} G \times \mathbb{R}$ is not surjective. Since then $\mathrm{d}\widetilde{\mathrm{End}}_u(\Omega)$ 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 Corollary 7.3.2, this is equivalent to say that there exists $(\xi, \xi_0) \neq (0, 0)$ such that

$$\xi \left(\mathrm{d}R_{\gamma_u(1)} \int_0^1 \mathrm{Ad}_{\gamma_u(t)}\, v(t)\, \mathrm{d}t \right) + \xi_0 \langle u, v\rangle = 0, \qquad \forall v \in \Omega. \tag{7.18}$$

Since differentials of right translations give linear isomorphisms, Eq. (7.18) 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

$$\lambda \left(\int_0^1 \mathrm{Ad}_{\gamma_u(t)}\, v(t)\, \mathrm{d}t \right) = \xi_0 \langle u, v\rangle, \qquad \forall v \in \Omega. \tag{7.19}$$

We now consider two cases: either $\xi_0 \neq 0$ or $\xi_0 = 0$. The first case is called *normal*, and the second one is called *abnormal*. We stress that if the codimension of $\mathrm{d}\widetilde{\mathrm{End}}_u(\Omega)$ within $T_{\mathrm{End}(u)} G \times \mathbb{R}$ is strictly larger than 1, then there are linearly independent choices for (λ, ξ_0). Hence, some particular u may have a normal pair (λ, ξ_0) and a (different) abnormal pair (λ', ξ_0').

We separately consider the two cases:

Normal case: $\xi_0 \neq 0$. Firstly, we suppose that (λ, ξ_0) as in (7.19) is such that $\xi_0 \neq 0$. Up to multiplying the equation by a constant, we can assume that $\xi_0 = 1$:

$$\lambda \left(\int_0^1 \mathrm{Ad}_{\gamma(t)}\, v(t)\, \mathrm{d}t \right) = \langle u, v\rangle, \quad \forall v \in \Omega. \tag{7.20}$$

Fix a Lebesgue point t of u, and let v diverge to the distribution $e_i\delta_t$ given by the Dirac mass at t times a basis vector e_i of V. Formally, we have

$$\begin{aligned}\dot\gamma_u(t) &= \mathrm{d}L_{\gamma_u(t)}u(t) = \mathrm{d}L_{\gamma_u(t)}\sum_{i=1}^{r}\langle u, \delta_t e_i\rangle e_i\\ &\overset{(7.20)}{=} \mathrm{d}L_{\gamma_u(t)}\sum_{i=1}^{r}\left(\lambda\int_0^1 \mathrm{Ad}_{\gamma_u(s)}(\delta_t e_i)\,\mathrm{d}s\right)e_i\\ &= \sum_{i=1}^{r}\lambda\left(\mathrm{Ad}_{\gamma_u(t)}(e_i)\right)X_i(\gamma_u(t)),\end{aligned}$$

where in the last equality we have used the identity $X_i(g) = \left(\mathrm{d}L_g\right)e_i$. We therefore say that an absolutely continuous curve γ satisfies the *normal equation* (or the *sub-Riemannian geodesic equation*) with respect to the left-invariant orthonormal frame $X_1, \dots, X_n$, if there exists $\lambda \in \mathfrak{g}^*$ such that

$$\dot\gamma(t) = \sum_{i=1}^{r}\lambda\left(\mathrm{Ad}_{\gamma(t)}(e_i)\right)X_i(\gamma(t)), \qquad \text{for a.e. } t \in [0, 1]. \tag{7.21}$$

A solution to (7.21) is called *normal curve*. By a bootstrap argument using (7.21), we deduce that the horizontal curve γ and its control u are C^∞. In fact, they are analytic; see Exercise 7.5.30.

Writing $u = \sum_{i=1}^{r} u_i e_i$ and recalling (7.12), we get another version of the normal equation:

$$u_i(t) = \lambda\left(\mathrm{Ad}_{\gamma_u(t)}(e_i)\right), \qquad \text{for a.e. } t \in [0, 1] \text{ and } \forall i \in \{1, \dots, r\}. \tag{7.22}$$

In particular, since in our case $\gamma(0) = 1_G$, the last equation implies

$$u_i(0) = \lambda(e_i), \qquad \forall i \in \{1, \dots, r\}. \tag{7.23}$$

In other words, the 'horizontal' part of the covector λ is the velocity at the initial time.

Abnormal case: $\xi_0 = 0$. Secondly, we suppose that (λ, ξ_0) as in (7.19) is such that $\xi_0 = 0$. In other words, the curve γ_u is singular for the endpoint map, as studied in Sect. 7.2.3. These are the curves that we previously called abnormal. In (7.16), we saw that a horizontal curve is abnormal if and only if there exists $\lambda \in \mathfrak{g}^*$ such that $\lambda \neq 0$ and

$$\lambda\left(\mathrm{Ad}_{\gamma(t)}(v)\right) = 0, \qquad \forall t \in [0, 1], \forall v \in V. \tag{7.24}$$

In other words, the set $\left\{\mathrm{Ad}_{\gamma(t)} v \,:\, v \in V,\ t \in I\right\}$ of vectors in $\mathfrak{g}$ is contained in a proper subspace of $\mathfrak{g}$.

We summarise the obtained results with the following statement, which we attribute to Pontryagin because, in fact, he obtained a much more general result for control systems.

Theorem 7.3.3 (Pontryagin's Maximum Principle) *Let G be a sub-Riemannian group with polarization V. Let $\gamma : I \to G$ be an energy-minimizing curve with control u. Then, at least one of the following happens:*

(i) there is $\lambda \in \mathfrak{g}^*$ such that $\langle u(t), X\rangle = \lambda\left(\mathrm{Ad}_{\gamma(t)}(X)\right)$ for all $X \in V$ and all $t \in I$, or
(ii) $\mathrm{span}\left\{\mathrm{Ad}_{\gamma(t)} V \,:\, t \in I\right\} \neq \mathfrak{g}$.

Each curve satisfying any of the above conditions is called an *extremal curve* or an *extremal*, and the equations are called *extremal equations*.

By a calibration argument, one can prove that every normal curve is length minimizing on short enough intervals; see [Mon02, Section 1.9.3]. The converse is not true, and the first (surprising) example has been found by Montgomery [Mon94]. In other words, there are sub-Riemannian structures where it is possible to find energy-minimizing curves that are not normal, and so are abnormal. They are called *strictly abnormal geodesics*. Be aware that there are geodesics that are normal and abnormal, and there are abnormal curves that are not geodesics; see Exercise 7.5.27.

Not much is known about strictly abnormal geodesics. We close this discussion with a second-order theorem, which we do not prove here. We refer to [ABB20, Chapter 12].

Theorem 7.3.4 (Goh) *Let G be a sub-Riemannian group with polarization V. If $\gamma : I \to G$ is a strictly abnormal geodesic, then* $\mathrm{span}\left\{\mathrm{Ad}_{\gamma(t)}(V + [V, V]) \,:\, t \in I\right\} \neq \mathfrak{g}$.

We point out that the latter theorem is easy for rank-2 polarizations; see Exercise 7.5.33.

7.3.2 A Distinguished Class of Functions

Both in the normal and the abnormal equations, (7.22) and (7.16), the functions $t \mapsto \lambda\left(\mathrm{Ad}_{\gamma(t)}(e_i)\right)$ are considered. We naturally see these functions as the composition of the curve with one of the following functions.

For $\lambda \in \mathfrak{g}^*$ and $Y \in \mathfrak{g}$, define $P_Y^\lambda : G \to \mathbb{R}$ as

$$P_Y^\lambda(g) := \lambda\left(\mathrm{Ad}_g(Y)\right), \qquad \forall g \in G. \tag{7.25}$$

These maps are linear both in λ and in Y. Moreover, we will see that if G is nilpotent, they are polynomials in g when seen in exponential coordinates; see Sect. 9.4.4. A useful formula that these functions satisfy is the following:

$$XP_Y^\lambda = P_{[X,Y]}^\lambda, \qquad \forall X, Y \in \mathfrak{g}, \forall \lambda \in \mathfrak{g}^*. \tag{7.26}$$

Indeed, fixing in addition $g \in G$, we first notice that

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t} \mathrm{Ad}_{g\exp(tX)}\bigg|_{t=0} &= \frac{\mathrm{d}}{\mathrm{d}t} \mathrm{Ad}_g \mathrm{Ad}_{\exp(tX)}\bigg|_{t=0} \\ &= \mathrm{Ad}_g \frac{\mathrm{d}}{\mathrm{d}t} \mathrm{Ad}_{\exp(tX)}\bigg|_{t=0} \overset{(5.13)}{=} \mathrm{Ad}_g \circ \mathrm{ad}_X,\end{aligned}$$

see also Exercise 7.5.34. Then, we deduce

$$\begin{aligned}(XP_Y^\lambda)(g) &= \frac{\mathrm{d}}{\mathrm{d}t}\lambda\, \mathrm{Ad}_{g\exp(tX)}\, Y\bigg|_{t=0} \\ &= (\lambda \circ \mathrm{Ad}_g \circ \mathrm{ad}_X)Y \\ &= \lambda\, \mathrm{Ad}_g([X, Y]) = P_{[X,Y]}^\lambda(g).\end{aligned}$$

If we have a normal curve γ with covector λ, the **normal equation** (7.21) reads as

$$\begin{aligned}\gamma'(t) &= \sum_{i=1}^r \lambda\left(\mathrm{Ad}_{\gamma_u(t)}(e_i)\right) e_i \\ &= \sum_{i=1}^r P_{e_i}^\lambda(\gamma(t))e_i, \qquad \text{for a.e. } t \in [0, 1].\end{aligned} \tag{7.27}$$

We next deduce that normal curves are parametrized with constant speed: from (7.27) we have that the derivative of the squared speed is

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\|\dot\gamma(t)\|^2 &\overset{(7.27)}{=} \frac{\mathrm{d}}{\mathrm{d}t}\sum_{i=1}^r \left(P_{e_i}^\lambda(\gamma(t))\right)^2 \\ &= \sum_{i=1}^r 2P_{e_i}^\lambda(\gamma(t))\frac{\mathrm{d}}{\mathrm{d}t}P_{e_i}^\lambda(\gamma(t)) \\ &\overset{(7.26)}{=} \sum_{i=1}^r 2P_{e_i}^\lambda(\gamma(t))P_{[\gamma'(t),e_i]}^\lambda(\gamma(t))\end{aligned}$$

$$\overset{(7.27)}{=} \sum_{i=1}^{r} 2P^\lambda_{e_i}(\gamma(t))P^\lambda_{[\sum_{j=1}^r P^\lambda_{e_j}(\gamma(t))e_j, e_i]}(\gamma(t))$$

$$= \sum_{i,j=1}^{r} 2P^\lambda_{e_i}(\gamma(t))P^\lambda_{e_j}(\gamma(t))P^\lambda_{[e_j,e_i]}(\gamma(t)) = 0,$$

where at the end we used that the elements of this sum are antisymmetric in the indices, because $[e_j, e_i] = -[e_i, e_j]$. Hence, the speed of normal curves is constant.

In terms of the functions (7.25), the **abnormal equation** (7.16) for a curve γ reads as follows: there exists $\lambda \in \mathfrak{g}^*$ such that $\lambda \neq 0$ and

$$P^\lambda_{e_i} \circ \gamma \equiv 0, \qquad \forall i \in \{1, \dots, r\}. \tag{7.28}$$

7.3.3 Extremals in Groups with Rank-2 Polarizations

We rephrase the extremal equations in the case of polarized groups where the polarization has rank 2. We begin with the abnormal condition:

Proposition 7.3.5 *Let (G, V) be a sub-Riemannian group whose V is spanned by two vectors e_1, e_2. Let $\gamma : [0, 1] \to G$ be a horizontal curve. Then γ 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*

$$\lambda(\mathrm{Ad}_{\gamma(t)}([e_1, e_2])) = 0. \tag{7.29}$$

Proof We use the notation $e_{12} := [e_1, e_2]$. We write $u(t) := \gamma'(t) = u_1(t)e_1 + u_2(t)e_2$ and we have

$$[u(t), e_1] = -u_2 e_{12} \qquad \text{and} \qquad [u(t), e_2] = u_1 e_{12}. \tag{7.30}$$

Taking the derivative of the abnormal equations, see (7.53) in Exercise 7.5.35, we have that if γ_u is an abnormal normal curve with covector $\lambda \in \mathfrak{g}^*$, with $\lambda \neq 0$, then

$$u_2 \lambda(\mathrm{Ad}_{\gamma(t)}(e_{12})) = u_1 \lambda(\mathrm{Ad}_{\gamma(t)}(e_{12})) = 0. \tag{7.31}$$

In addition, notice that we may assume that γ_u is parametrized by arc length, so u has a constant nonzero norm almost everywhere. In particular, we have $(u_1, u_2) \neq (0, 0)$ almost everywhere. Therefore, we can conclude that for such an abnormal curve, we have (7.29).

Vice versa, assume γ satisfies (7.29) for some $\lambda \in \mathfrak{g}^*$ with $\lambda \neq 0$. Then, it clearly satisfies (7.31) and, since we have $r = 2$ and we have (7.30). Then, look at each function $\lambda\,\mathrm{Ad}_{\gamma(t)}(e_i)$, for $i = 1$ and 2. On the one hand, we have that its derivative is 0; see Exercise 7.5.34. On the other hand, if $\gamma(0) = 1_G$ and if λ satisfies (7.17), we have that the initial condition at time $t = 0$ for (7.16) is satisfied. Hence, such a curve is abnormal. □

We shall rephrase the normal equation in terms of a curvature condition for the development of the curve. See Definition 7.1.4 for the notion of development, and Exercise 7.5.36 for the oriented curvature.

Proposition 7.3.6 *Let (G, V) be a sub-Riemannian group whose V is spanned by two orthonormal vectors e_1, e_2. Let $\gamma : [0, 1] \to G$ be a horizontal curve with development σ. Assume σ is C^2, with never vanishing speed, and let κ be the oriented curvature of σ. Then γ is normal if and only if for some $\lambda \in \mathfrak{g}^*$ we have that γ satisfies $\gamma'(0) = \lambda(e_1)e_1 + \lambda(e_2)e_2$ and*

$$\kappa = \frac{1}{\|\gamma'\|}\lambda(\mathrm{Ad}_{\gamma(t)}([e_1, e_2])). \tag{7.32}$$

Proof Let $e_{12} := [e_1, e_2]$. We take the derivative of the normal equations; see Exercise 7.5.35 and its equation (7.54). By (7.30) we have that if $\gamma = \gamma_u$ is a normal curve with covector $\lambda \in \mathfrak{g}^*$ and control $u = \gamma'$, then we have (7.23) and

$$\begin{cases} \dot{u}_1 = -u_2\lambda(\mathrm{Ad}_{\gamma(t)}(e_{12})), \\ \dot{u}_2 = \ \ u_1\lambda(\mathrm{Ad}_{\gamma(t)}(e_{12})). \end{cases} \tag{7.33}$$

Let $\sigma : [0, 1] \to \mathbb{R}^2$ be the planar curve such that $\dot{\sigma} = u$. By Exercise 7.5.36 its curvature satisfies

$$\kappa = \frac{\sigma_1'\sigma_2'' - \sigma_2'\sigma_1''}{\|\sigma'\|^3} = \frac{u_1\dot{u}_2 - u_2\dot{u}_1}{\|u\|^3}$$
$$\overset{(7.33)}{=} \frac{u_1^2\lambda(\mathrm{Ad}_\gamma(e_{12})) + u_2^2\lambda(\mathrm{Ad}_\gamma(e_{12}))}{\|u\|^3} = \frac{1}{\|u\|}\lambda(\mathrm{Ad}_{\gamma(t)}(e_{12})).$$

We observe that the element $\mathrm{Ad}_{\gamma(t)}(e_{12})$ is in $[\mathfrak{g}, \mathfrak{g}]$, hence in the last equation we have lost the information of the value of λ on V. Still, normal curves need to satisfy (7.17).

Vice versa, we assume γ is a horizontal curve that for some $\lambda \in \mathfrak{g}^*$ satisfies (7.23) and (7.32). By bootstrapping (7.32), we see that γ and its control u are smooth. We consider the normal curve associated with λ starting at $\gamma(0)$, which we denote by γ^λ. We shall show that $\gamma = \gamma^\lambda$.

The reason is that both curves satisfy the ODE (7.32) with the same initial data. We stress that such an ODE is not of the simple type $u''(t) = F(t, u(t), u'(t))$, but it is of the type $F(t, u, u', u'') = 0$. However, by the implicit function theorem, one can write the ODE of the second type in the form of the first type, as long as $\partial_{u''}F \neq 0$; see Exercise 7.5.37. We then obtain an implicit function with a bound on the C^1 norm as long as the control u is in a compact set. □

7.3.4 Extrema in Two-Step Nilpotent Lie Groups

In this section, we study a particular class of Lie groups. We denote each such group by G_q, where q is a skew-symmetric bilinear form, which captures the Lie bracket. We will see later that these groups G_q are exactly the simply connected Lie groups that are nilpotent with nilpotency step 2; see Example 9.3.8.

Definition 7.3.7 (The Lie Group G_q) Let V and W be finite-dimensional vector spaces and $q : V \times V \to W$ a skew-symmetric bilinear map. Equip $V \times W$ with the group law

$$(v_1, w_1) \cdot (v_2, w_2) := \left(v_1 + v_2,\ w_1 + w_2 + \frac{1}{2} q(v_1, v_2) \right), \qquad \forall v_1, v_2 \in V, \forall w_1, w_2 \in W,$$

whose identity element is $(0, 0)$ and the inverse of (v_1, w_1) is $(-v_1, -w_1)$. We denote the group $(V \times W, \cdot)$ with G_q. We equip the vector space $V \times W$ with the Lie bracket

$$[(v_1, w_1), (v_2, w_2)] := (0, q(v_1, v_2)).$$

Notice that $(V \times W, [\cdot, \cdot])$ is a Lie algebra and for every $x, y, z \in V \times W$, we have $[x, [y, z]] = 0$. Moreover, when equipped with the differential structure of vector space, we have that G_q is a Lie group, and its tangent at $(0, 0)$ is naturally identified with $V \times W$. The identity map $(V \times W, [\cdot, \cdot]) \to (V \times W, \cdot)$ is the exponential map; see Exercise 7.5.38.

For $(v_1, w_1), (v_2, w_2) \in V \times W$ we have

$$\begin{aligned} \mathrm{Ad}_{(v_1,w_1)}(v_2, w_2) &= \mathrm{Ad}_{\exp(v_1,w_1)}(v_2, w_2) \\ &= e^{\mathrm{ad}_{(v_1,w_1)}}(v_2, w_2) \\ &= (v_2, w_2) + [(v_1, w_1), (v_2, w_2)] \\ &= (v_2, w_2 + q(v_1, v_2)). \end{aligned} \tag{7.34}$$

7.3.4.1 Abnormal Curves in Two-step Nilpotent Groups

Proposition 7.3.8 *Let $G = G_q$ be a two-step nilpotent group equipped with a polarization. For each abnormal curve γ in G with $\gamma(0) = 1_G$, there exists a proper subgroup G' of G containing γ, in which γ is a non-abnormal horizontal curve, with respect to the induced polarization.*

Proof Let G' be the smallest subgroup of G containing γ, which is still of the form $G_{q'}$; see Exercise 7.5.40. Equip G' with the induced polarization Δ from G, which by the minimality of G' is bracket generating. Assume by contradiction that γ is abnormal in G'. By the abnormal equation (7.16), there exists $\lambda \in \mathrm{Lie}(G')^*$ such that $\lambda \neq 0$ and

$$0 = \lambda\left(\mathrm{Ad}_{\gamma(t)}(X)\right) \overset{(7.34)}{=} \lambda(X + [\gamma(t), X]) \overset{(7.17)}{=} \lambda([\gamma(t), X]), \qquad \forall X \in \Delta_1. \tag{7.35}$$

We consider the set

$$Z := \{(v, w) \in V \times W \; : \; \lambda([v, X]) = 0, \forall X \in \Delta_1\}.$$

Hence, the curve γ is contained in Z. Moreover, the set Z is a proper subgroup of G_q, because in the definition of Z the dependence on (v, w) is linear, $[\mathfrak{g}, \mathfrak{g}] \subseteq Z$, and λ is not zero on $[\Delta_1, \Delta_1]$. We contradicted the minimality of G'. Hence, the curve γ is not abnormal in G'. □

As a consequence, since every 2-step nilpotent group is a quotient of some G_q, we have that every geodesic curve in a group of nilpotency step 2 cannot be abnormal in the minimal group containing it. Since the curve is geodesic also within this subgroup, then, by PMP Theorem 7.3.3, the curve is normal in the subgroup. We deduced, without using Goh Theorem 7.3.4, the following consequence:

Corollary 7.3.9 *Let G be a two-step nilpotent group. Equip G with a left-invariant sub-Riemannian structure. Then, every geodesic curve in G is smooth.*

7.3.4.2 Normal Curves in Two-step Nilpotent Groups

Proposition 7.3.10 *Let $G = G_q = V \times W$ be a two-step nilpotent group as in Definition 7.3.7. Equip G with a left-invariant sub-Riemannian structure with polarization $\Delta \subset V \times W$ such that $V \subset \Delta$ and V is the orthogonal of $W \cap \Delta$ in Δ. Then, for every energy-minimizing curve $\gamma = (v, w) : [0, T] \to G$ with $\gamma(0) = (0, 0)$, there are $M \in \mathfrak{so}(V)$, $b \in \ker(M)$, $c \in M(V)$, and $\zeta \in W \cap \Delta$ such that*

$$\begin{cases} v(t) = c - e^{tM}c + tb, \\ w(t) = \frac{1}{2}\int_0^t [v(s), \dot{v}(s)]\,\mathrm{d}s + t\zeta, \end{cases} \qquad \forall t \in [0, T]. \tag{7.36}$$

Proof By Pontryagin's Theorem 7.3.3, the curve γ is normal or abnormal. By Goh Theorem 7.3.4, the curve γ is normal. Let $e_1, \ldots, e_r$ be an orthonormal basis of V and $f_1, \ldots, f_m$ an orthonormal basis of $W \cap \Delta$. The normal Eq. (7.27) says that there is $\lambda \in \mathfrak{g}^*$ such that

$$\gamma'(t) = \sum_{i=1}^{r} \lambda(\mathrm{Ad}_{\gamma(t)}\, e_i)e_i + \sum_{i=1}^{m} \lambda(\mathrm{Ad}_{\gamma(t)}\, f_i) f_i. \tag{7.37}$$

For $\gamma = (v, w)$ we decompose $\gamma'(t) = (\gamma'(t))_1 + (\gamma'(t))_2$ with $(\gamma'(t))_1 \in V$ and $(\gamma'(t))_2 \in W$. Then, we have

$$(\dot{v}, \dot{w}) = \dot{\gamma} = \mathrm{d}L_\gamma(\gamma') = \left((\gamma')_1, \frac{1}{2}[v, (\gamma')_1] + (\gamma')_2\right).$$

Since $\mathrm{Ad}_{(v,w)} X = X + [v, X]$, we obtain from (7.37) that the normal curve $\gamma = (v, w)$ satisfies the ODE

$$\begin{cases} \dot{v} &= \sum_{i=1}^r \lambda(e_i + [v, e_i])e_i \\ \dot{w} &= \frac{1}{2}[v, \dot{v}] + \sum_{i=1}^m \lambda(f_i) f_i. \end{cases}$$

We define $M : V \to V$ to be the linear map $Mx := \sum_{i=1}^r \lambda([x, e_i])e_i$ and $(\lambda_H, \zeta) \in V \times W$ to be the vectors $\lambda_H := \sum_{i=1}^r \lambda(e_i)e_i$ and $\zeta := \sum_{i=1}^m \lambda(f_i) f_i$. We can rewrite the ODE as

$$\begin{cases} \dot{v} &= Mv + \lambda_H \\ \dot{w} &= \frac{1}{2}[v, \dot{v}] + \zeta. \end{cases} \tag{7.38}$$

The linear transformation M is skew-symmetric, because the matrix representation of M in the orthonormal basis $e_1, \dots, e_n$ is $M_{ij} = \lambda([e_i, e_j])$. The solution of (7.38) with $v(0) = w(0) = 0$ is (7.36), where $\lambda_H = b - Mc$ is decomposed into $b \in \ker(M)$ and $c \in M(V)$. □

Corollary 7.3.11 *Let G be a simply connected sub-Riemannian group of nilpotency step 2 with polarization Δ. Every isometric embedding $\gamma : [0, \infty) \to G$ with $\gamma(0) = 1_G$ is the restriction of an OPS with $\dot{\gamma}(0) \perp (\Delta_1 \cap [\mathfrak{g}, \mathfrak{g}])$.*

Proof Let $G = G_q = V \times W$ be for some $q : V \times V \to W$ as in Definition 7.3.7, with $V \subset \Delta$ and $V \perp W \cap \Delta$. Since the bracket given by q is a bilinear map, there exists $K > 0$ such that

$$|[v_1, v_2]| \leq K|v_1||v_2| \quad \text{for every } v_1, v_2 \in \mathfrak{g}, \tag{7.39}$$

where $|\cdot|$ is a fixed auxiliary norm on $\mathfrak{g}$ that is equal to the sub-Riemannian norm on Δ_1.

Let $\gamma : [0, T] \to G$ be isometric. Hence, it is of the form $\gamma(t) = v(t) + w(t)$, with $v(t) \in V$ and $w(t) \in W$ for every $t \in [0, T]$, satisfying (7.36), for some $M \in \mathfrak{so}(V)$, $b \in \ker(M)$, $c \in M(V)$, and $\zeta \in W$. Since γ is parametrized by arc length, we have $|Mc|^2 + |b|^2 + |\zeta|^2 = 1$.

Since e^{TM} is an isometry and $c, e^{TM}c \in M(V) \subseteq b^\perp$, from (7.36) we get

$$d_{\mathrm{sR}}(0, v(T)) \overset{\text{Ex.7.5.39}}{=} |v(T)| \overset{(7.36)}{=} |c - e^{TM}c + Tb| \leq \sqrt{|2c|^2 + |Tb|^2}. \tag{7.40}$$

Next, we bound $w(T)$. By (7.36), we further have

$$
\begin{aligned}
|w(T)| &\overset{(7.36)}{\leq} \frac{1}{2}\left|\int_0^T [c - e^{tM}c + tb, -e^{tM}Mc + b]\,\mathrm{d}t\right| + T|\zeta| \\
&\leq \frac{1}{2}\left|\int_0^T [c - e^{tM}c, e^{tM}Mc]\,\mathrm{d}t\right| + \frac{1}{2}\left|\int_0^T [c - e^{tM}c, b]\,\mathrm{d}t\right| \\
&\qquad + \frac{1}{2}\left|\int_0^T [tb, e^{tM}Mc]\,\mathrm{d}t\right| + T|\zeta| \\
&\overset{(7.39)}{\leq} \frac{1}{2}\int_0^T K2|c||Mc|\,\mathrm{d}t + \frac{1}{2}\int_0^T 2|c||b|\,\mathrm{d}t \\
&\qquad + \frac{1}{2}\left|\int_0^T [b, te^{tM}Mc]\,\mathrm{d}t\right| + T|\zeta| \\
&\overset{(7.39)}{\leq} TK|c||Mc| + TK|c||b| + \frac{1}{2}\left|\left[b, \int_0^T te^{tM}Mc\,\mathrm{d}t\right]\right| + T|\zeta|,
\end{aligned}
$$

where we used the linearity of the Lie bracket. Focusing on the third term, we integrate by parts:

$$
\frac{1}{2}\left|\left[b, \int_0^T te^{tM}Mc\,\mathrm{d}t\right]\right| = \frac{1}{2}\left|\left[b, Te^{TM}c - \int_0^T e^{tM}c\,\mathrm{d}t\right]\right| \overset{(7.39)}{\leq} TK|c||b|.
$$

All together, we obtain a bound that is linear in T:

$$
|w(T)| \leq T(K|c|\,(|Mc| + 2|b|) + |\zeta|). \tag{7.41}
$$

Lie groups of step 2 satisfy a global upper bound given by the box distance; see Exercise 7.5.41. In particular, using (7.55), we deduce that for some constants C and $\tilde{C} = \tilde{C}(G, M, c, b)$ we have

$$
\begin{aligned}
T &= d_{\mathrm{sR}}(0, v(T) + w(T)) \\
&\leq d_{\mathrm{sR}}(0, v(T)) + d_{\mathrm{sR}}(0, w(T)) \\
&\overset{(7.40)\&(7.55)}{\leq} \sqrt{|Tb|^2 + |2c|^2} + C|w(T)|^{\frac{1}{2}} \\
&\overset{(7.41)}{\leq} \sqrt{|Tb|^2 + |2c|^2} + C\sqrt{T(K|c|\,(|Mc| + 2|b|) + |\zeta|)} \\
&\leq T|b| + \tilde{C}(\sqrt{T} + T^{-1}) + \tilde{C}.
\end{aligned}
$$

If $|b| < 1$, then the latter inequality cannot hold for arbitrarily large T. We conclude that, if $\gamma : [0, \infty) \to G$ is an isometric embedding, then $|b| = 1$ and thus $c = \zeta = 0$, i.e., the curve γ is a straight lines with directions orthogonal to $[\mathfrak{g}, \mathfrak{g}]$. □

7.4 Characterization of Geodesic Left-Invariant Distances

Berestovskii's work [Ber88, Ber89a, Ber89b] clarified what are the possible isometrically homogeneous distances on manifolds that, in addition, are geodesic distances: They are sub-Finsler metrics.

Theorem 7.4.1 (Berestovskii) *Let $M := G/H$ be the Lie coset space 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 sub-Finsler 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* (4.4).

We will only partially prove Theorem 7.4.1; one can check the original reference [Ber88, Theorem 2]. We will discuss how to give a constructive characterization of the sub-Finsler structure.

At the beginning of the proof of Theorem 7.4.1, there is a crucial rectifiability result. Namely, one proves that the geodesics and, more generally, the curves of finite length for the geodesic distance on G/H also have finite length for Riemannian distances. Consequently, such curves are absolutely continuous. Their derivatives will necessarily form the subbundle of the sub-Finsler structure; see (7.51).

Lemma 7.4.2 (Crucial Lemma for Berestovskii's Construction) *Let $M := G/H$ be the Lie coset space of a Lie group G modulo a closed subgroup H. Let d be an admissible geodesic distance on M that is G-invariant. Then, for every Riemannian G-invariant distance d_R on M there is a constant $C > 0$ such that*

$$d_R \leq Cd.$$

Proof For simplicity of exposition, we consider only the case when H is trivial and $M = G$. As a preparation, via Proposition 5.2.8, let $\delta > 0$ such that there exists a neighborhood of U of 0 in $\mathfrak{g}$ such that $\exp|_U : U \to B_{d_R}(1_G, \delta)$ is bi-Lipschitz for some constant $k > 1$, in particular,

$$\frac{1}{k}\|\log(p)\| \leq d_R(1_G, p) \leq k\|\log(p)\|, \qquad \forall p \in B_{d_R}(1_G, \delta). \tag{7.42}$$

Also, since the topologies of the two metrics are the same, we take $r > 0$ such that

$$\bar{B}_d(1_G, r) \subseteq B_{d_R}(1_G, \delta). \tag{7.43}$$

We prove the lemma by contradiction and assume that for each $n \in \mathbb{N}$, there exist $p_n, q_n \in M$ such that

$$d_R(p_n, q_n) \geq n\, d(p_n, q_n). \tag{7.44}$$

We shall show that we may suppose that $d_R(p_n, q_n)$ is bounded by the δ coming from (7.42). Indeed, for each $n \in N$, consider a curve γ from p_n to q_n that is a geodesic with respect to d. By continuity, take points $a_0 = p_n, a_1, a_2, \dots, a_N = q_n$, for some $N \in \mathbb{N}$, along γ such that

$$d_R(a_{j-1}, a_j) < \delta, \qquad \forall j \in \{1, \dots, N\}. \tag{7.45}$$

We claim that there exists an index $j \in \{1, \dots, N\}$ such that

$$d_R(a_{j-1}, a_j) \geq n\, d(a_{j-1}, a_j). \tag{7.46}$$

Indeed, if not, for every j, we would have $d_R(a_{j-1}, a_j) < n\, d(a_{j-1}, a_j)$; summing and using that the points a_j's are along a d-geodesic, we would have

$$\begin{aligned} d_R(p_n, q_n) &\leq \sum_{j=1}^{N} d_R(a_{j-1}, a_j) \\ &< n \sum_{j=1}^{N} d(a_{j-1}, a_j) \\ &= n\, d(p_n, q_n). \end{aligned}$$

We obtained a contradiction with (7.44) and, thus, the claim is proved, which means that by replacing p_n and q_n with the just found a_{j-1} and a_j, we have the bounds

$$d(p_n, q_n) \overset{(7.46)}{\leq} \frac{1}{n} d_R(p_n, q_n) \overset{(7.45)}{\leq} \frac{\delta}{n}, \qquad \forall n \in \mathbb{N}. \tag{7.47}$$

By homogeneity of both distances, it is possible to assume that q_n is always the point $1 := 1_G$. Equation (7.47) becomes

$$d(p_n, 1) \leq \frac{1}{n} d_R(p_n, 1) \leq \frac{\delta}{n} < \delta, \qquad \forall n \in \mathbb{N}. \tag{7.48}$$

Notice that, in particular, we have $d(p_n, 1) \to 0$, i.e., the sequence p_n converges to 1. Moreover, for the value r from (7.43), the distance between $B_d(1, r)$ and $M \setminus B_{d_R}(1, \delta)$ is positive. Therefore, for n large enough, then there exists $m_n \in \mathbb{N}$ such that

$$(p_n)^{m_n} = \exp(m_n \log(p_m)) \in B_{d_R}(1, \delta) \setminus \bar{B}_d(1, r). \tag{7.49}$$

We stress that $\log((p_n)^{m_n}) = m_n \log(p_n)$.

Next, using in addition the triangle inequality and the left invariance of the distance function, we bound:

$$
\begin{aligned}
0 < r &\overset{(7.49)}{<} d(p_n^{m_n}, 1) \\
&\leq \sum_{j=1}^{m_n} d(p_n^{j-1}, p_n^j) \\
&= m_n d(p_n, 1) \\
&\overset{(7.48)}{\leq} \frac{1}{n} m_n d_R(p_n, 1) \\
&\overset{(7.42)}{\leq} \frac{k m_n}{n} \|\log(p_n)\| \\
&= \frac{k}{n} \|m_n \log(p_n)\| \\
&= \frac{k}{n} \left\|\log(p_n^{m_n})\right\| \\
&\overset{(7.42)}{\leq} \frac{k^2}{n} d_R(p_n^{m_n}, 1) \overset{(7.49)}{\leq} \frac{k^2 \delta}{n} \overset{n\to\infty}{\longrightarrow} 0.
\end{aligned}
$$

This contradiction proves the lemma. □

7.4.1 Berestovskii's Construction

With the crucial help of Lemma 7.4.2, we will then explain the construction of the Berestovskii CC structure that appears in his Theorem 7.4.1.

Sketch of a Proof of Theorem 7.4.1 From Lemma 7.4.2, we know that some dilation of the distance d is greater than a Riemannian distance. So, rescaling the metric d_R, we assume to have $d_R \leq d$. We deduce that every curve $\gamma : I \to M$ that is L-Lipschitz with respect to d is also L-Lipschitz with respect to d_R, since

$$
d_R(\gamma(t_1), \gamma(t_2)) \leq d(\gamma(t_1), \gamma(t_2)) \leq L|t_1 - t_2|, \qquad \forall t_1, t_2 \in I. \tag{7.50}
$$

Consequently, by Rademacher's Theorem, we obtain that every d-rectifiable curve parametrized by a multiple of arc length is differentiable almost everywhere and absolutely continuous.

To prove Berestovskii's Theorem 7.4.1, we need to find a sub-bundle Δ of the tangent bundle of M and a norm on it. As a result of (7.50), it makes sense to look at the set of velocities of d-rectifiable curves. Denote by $\mathtt{Lip}([-1, 1]; M)$ the collection of curves from $[-1, 1]$ to M that are Lipschitz with respect to d. We

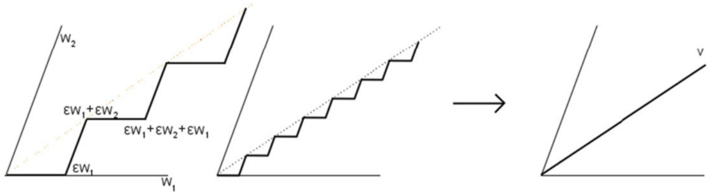


Fig. 7.1 The sequence of ϵ-zigzag curves converges (up to subsequences) to a curve whose tangent vector at the initial point is the average v of the two tangent vectors w_1 and w_2 of the two initial curves

define the *Berestovskii bundle* as the subset of TM given by

$$\begin{aligned} &\Delta := \cup_{p \in M} \Delta_p, \\ &\quad \text{where} \quad \Delta_p := \{\dot{\gamma}(0) \ : \ \gamma \in \mathtt{Lip}([-1,1]; M), \gamma(0) = p, \exists \dot{\gamma}(0)\}. \end{aligned} \tag{7.51}$$

One way to characterize the sub-Finsler norm that one needs to put on Δ is to describe the unit ball for the norm. One considers tangent vectors of curves that are Lipschitz with constant at most 1:

$$\|v\| \leq 1 \overset{\text{def}}{\iff} \exists \gamma : [-1, 1] \to M \text{ with } \gamma \text{ 1-Lipschitz and } \dot{\gamma}(0) = v.$$

The fact that every fiber of Berestovskii's bundle is a vector subspace can be proved by constructing limits of zigzag curves obtained via translations of curves. Figure 7.1 pictures the infinitesimal construction: we begin with two Lipschitz curves γ_1, γ_2, giving tangent vectors w_1 and w_2, respectively. Given $\epsilon > 0$ we follow γ_1 for time ϵ, then we follow a translation of γ_2 for time ϵ, then we follow a translation of γ_1 for time ϵ, etc. As $\epsilon \to 0$, the limit curve will be rectifiable and will have a tangent equal to $\frac{1}{2}(w_1 + w_2)$ at 0. Similarly, if γ_1 and γ_2 are 1-Lipschitz, then so are the zigzag curves and their limits. Hence, the defined norm is convex.

We stress that Berestovskii's distribution and norm are invariant under the (transitive) isometry group. In particular, the distribution is smooth and has a constant rank, and the norm is continuously varying. We obtain a Carnot-Carathéodory space with length structure given by (3.25) and distance d_{cc} given by (4.4).

To obtain Berestovskii's result, one argues separately for the two inequalities: $d_{cc} \leq d$ and $d \leq d_{cc}$. The first inequality is straightforward: Given $p, q \in M$, let γ be a d-geodesic from p to q that is parametrize by arc length with respect to d; so, for $\ell := d(p, q)$, the curve $\gamma : [0, \ell] \to M$ is 1-Lipschitz. By definition of the norm, we have $\|\dot{\gamma}(t)\| \leq 1$. Therefore, we have the bound:

$$d_{cc}(p, q) \overset{\text{def}}{\leq} \text{Length}_{\|\cdot\|}(\gamma) \overset{\text{def}}{=} \int_0^\ell \|\dot{\gamma}(t)\| \, dt \leq \int_0^\ell 1 \, dt = \ell = d(p, q).$$

The second inequality is more involved. Given a curve that is admissible for Δ, one has to construct a d-rectifiable curve with almost the same length. This is done similarly with the zigzag method. We again refer to [Ber88], but also suggest [LD11a, Section 5.2]. □

The argument that we overviewed is quite flexible and can be used to study Lie coset spaces equipped with distances that are bi-Lipschitz homogeneous.

Theorem 7.4.3 ([LD11a, Theorem 1.1]) *Let $M := G/H$ be a Lie coset space equipped with a geodesic admissible distance d. Suppose there exists a subgroup G' of G that acts transitively on M and that acts by maps that are locally bi-Lipschitz with respect to d. Then there exists a G'-invariant Carnot-Carathéodory structure on M whose distance is locally bi-Lipschitz equivalent to d.*

7.5 Exercises

Exercise 7.5.1 Let Δ be a left-invariant distribution on a Lie group. Then, there exists a global frame for Δ made of left-invariant vector fields.

Exercise 7.5.2 Let Δ be a left-invariant distribution on a Lie group G. For $k \in \mathbb{N}$, let $\Delta^{[k]}$ be the k-th element in the flag of subbundles associated with Δ as in Definition 4.1.13. Then, the sequence of subspaces $V^{[k]} := \Delta^{[k]}_{1_G}$, as $k \in \mathbb{N}$, satisfies (7.2).
Hint. Recall the two interpretations of the Lie bracket for Lie algebras.

Exercise 7.5.3 Let V be a subspace of a Lie algebra $\mathfrak{g}$, with $n := \dim \mathfrak{g}$. Consider the spaces $V^{[k]}$ from (7.2)

(i) If $\dim V \leq 1$, then $V^{[k]} = V$ for all $k \in \mathbb{N}$.
(ii) If V is bracket generating and $V \neq \mathfrak{g}$, then $\dim V \geq 2, n \geq 3$, and $V^{[n-1]} = \mathfrak{g}$.
(iii) If V is not bracket generating and $n \geq 3$, then $V^{[n-2]} = V^{[k]}$, for all $k \geq n-2$.

Exercise 7.5.4 Formula (7.5) defines a function that is left-invariant, continuous, and gives a continuously varying norm in the sense of Definition 3.2.6.

Exercise 7.5.5 Let $\varphi : G \to H$ be a Lie group homomorphism between sub-Finsler Lie groups $(G, \Delta^G, \|\cdot\|)$ and $(H, \Delta^H, \|\cdot\|)$. Assume $\varphi_*(\Delta^G_1) \supseteq \Delta^H_1$. Then $\varphi : G \to H$ is surjective.
Hint. The set $\varphi_*(\mathfrak{g})$ is a Lie algebra containing the generating sub-space Δ^H_1.

Exercise 7.5.6 Let $\varphi : G \to H$ be a Lie group homomorphism between sub-Finsler Lie groups $(G, \Delta^G, \|\cdot\|)$ and $(H, \Delta^H, \|\cdot\|)$. Assume $\varphi_*(\Delta^G_1) \subseteq \Delta^H_1$. Then $\varphi : G \to H$ is L-Lipschitz with respect to the respective sub-Finsler metrics, where L is the Lipschitz constant of the linear map

$$\varphi_*|_{\Delta^G_1} : (\Delta^G_1, \|\cdot\|) \to (\Delta^H_1, \|\cdot\|).$$

Solution. It is enough to observe that if $\gamma : [0, 1] \to G$ is a horizontal curve, then

$$\begin{aligned}
\mathrm{Length}(\varphi \circ \gamma) &= \int_0^1 \left\| \frac{\mathrm{d}}{\mathrm{d}t} (\varphi(\gamma(t))) \right\| \mathrm{d}t \\
&= \int_0^1 \left\| (L_{\varphi(\gamma(t))})^* \frac{\mathrm{d}}{\mathrm{d}t} (\varphi(\gamma(t))) \right\|_{\Delta_1^H} \mathrm{d}t \\
&= \int_0^1 \left\| \varphi_*(\gamma'(t)) \right\|_{\Delta_1^H} \mathrm{d}t \\
&\leq L \int_0^1 \left\| \gamma'(t) \right\|_{\Delta_1^G} \mathrm{d}t \\
&= L \int_0^1 \| \dot{\gamma}(t) \| \, \mathrm{d}t \\
&= L \, \mathrm{Length}(\gamma).
\end{aligned}$$

Exercise 7.5.7 Let $\pi : \mathfrak{g} \to \mathfrak{h}$ be a surjective Lie algebra homomorphism. If $V \subseteq \mathfrak{g}$ Lie generates $\mathfrak{g}$, then $\pi(V) \subseteq \mathfrak{h}$ Lie generates $\mathfrak{h}$.

Exercise 7.5.8 Let M be a sub-Finsler manifold. Let $G \curvearrowright M$ be a Lie group action that is proper, free, and by isometries. Then, there exists a sub-Finsler distance d on $G \backslash M$ that makes π a submetry.
Hint. Regarding the existence of a submetric distance, check Exercise 6.6.22. Regarding the fact that it is sub-Finsler, check Proposition 7.1.9.

Exercise 7.5.9 Let M_1 and M_2 be sub-Riemannian manifolds with horizontal frames $X_1, \dots, X_m$ and $Y_1, \dots, Y_m$, respectively. Let $\pi : M_1 \to M_2$ be smooth and surjective. If X_i is π-related to Y_i, for all $i \in \{1, \dots, m\}$, then π is a submetry.

Exercise 7.5.10 Let $k \in \mathbb{N}$. Let $\pi : M_1 \to M_2$ be a smooth map between sub-Finsler manifolds. Assume that M_1 and M_2 are boundedly compact and their distributions have rank k. Assume π surjective and such that

$$(\mathrm{d}\pi)_p : (\Delta_p, \|\cdot\|) \to (\Delta_{\pi(p)}, \|\cdot\|), \qquad \forall p \in M_1,$$

is a submetry, then π is a submetry. What if the ranks are different?

Exercise 7.5.11 For the standard basis in the Heisenberg group, the map E from Definition 7.1.19 is $E(\mathbf{t}) = e^{t_1 X} e^{t_2 Y} e^{\sqrt{t_3} X} e^{\sqrt{t_3} Y} e^{-\sqrt{t_3} X} e^{-\sqrt{t_3} Y}$.

Exercise 7.5.12 For the standard basis in the Engel group (see (11.14)), the map E from Definition 7.1.19 is

$$\begin{aligned}
E(\mathbf{t}) = e^{t_2 X} \, e^{t_2 Y} \, e^{\sqrt{t_3} X} e^{\sqrt{t_3} Y} e^{-\sqrt{t_3} X} e^{-\sqrt{t_3} Y} \\
e^{\sqrt[3]{t_4} X} \, e^{\sqrt[3]{t_4} Y} e^{-\sqrt[3]{t_4} X} e^{-\sqrt[3]{t_4} Y} e^{\sqrt[3]{t_4} X} e^{\sqrt[3]{t_4} Y} e^{\sqrt[3]{t_4} X} e^{-\sqrt[3]{t_4} Y} e^{-\sqrt[3]{t_4} X} \, e^{-\sqrt[3]{t_4} X}.
\end{aligned}$$

Exercise 7.5.13 Each map $\Phi^{(j)}$ from Definition 7.1.19 satisfies
$$\frac{\mathrm{d}}{\mathrm{d}t}\Phi^{(j)}(t)\bigg|_{t=0^+} = \frac{\mathrm{d}}{\mathrm{d}t}\Phi^{(j)}(t)\bigg|_{t=0^-}.$$

Exercise 7.5.14 (Chow's Theorem for sub-Finsler Groups) Each point $p \in G$ in every sub-Finsler group G can be joined to the identity element 1_G by a horizontal path. Moreover, the CC distance induces the manifold topology.
Hint. Use Corollary 7.1.21.

Exercise 7.5.15 (Grönwall Lemma—Classical Version) Let $f : [0, 1] \to \mathbb{R}$ be an absolutely continuous, $T > 0$, and $\alpha, \beta \in L^1([0; T])$ such that

$$f'(t) \leq \alpha(t) f(t) + \beta(t), \qquad \text{for a.e. } t \in [0, T].$$

Then, for every $t \in [0, T]$ the following holds:

$$f(t) \leq f(0) e^{\int_0^t \alpha(s)\,\mathrm{d}s} + e^{\int_0^t \alpha(s)\,\mathrm{d}s} \int_0^t \beta(s) e^{-\int_0^s \alpha(r)\,\mathrm{d}r}\,\mathrm{d}s.$$

Exercise 7.5.16 (Grönwall Lemma—Integral Version) Given $a < b$, let $\alpha, \beta, f : [a, b] \to \mathbb{R}$ be functions such that α and β are integrable, while f is continuous. Assume β is non-negative and

$$f(t) \leq \alpha(t) + \int_a^t \beta(s) f(s)\,\mathrm{d}s, \qquad \forall t \in [a, b].$$

Then,

$$f(t) \leq \alpha(t) + \int_a^t \alpha(s)\beta(s) e^{\int_s^t \beta(r)\,\mathrm{d}r}\,\mathrm{d}s, \qquad \forall t \in [a, b].$$

If, moreover, the function α is non-decreasing, then

$$f(t) \leq \alpha(t) e^{\int_a^t \beta(s)\,\mathrm{d}s}, \qquad \forall t \in [a, b].$$

Hint. Define $v(s) := \exp\left(-\int_a^s \beta(r)\,\mathrm{d}r\right) \int_a^s f(r)\beta(r)\,\mathrm{d}r$. Show that v is absolutely continuous and that $v'(s) \leq \exp\left(-\int_a^s \beta(r)\,\mathrm{d}r\right) \beta(s)\alpha(s)$.

Exercise 7.5.17 Let $u, v : [0, 1] \to \mathfrak{gl}(n)$ be measurable functions. For $\epsilon > 0$, denote by $\gamma_{u+\epsilon v} : [0, 1] \to \mathfrak{gl}(n)$ the solution of the Cauchy Problem

$$\begin{cases} \dot{\gamma}_{u+\epsilon v}(t) &= \gamma_{u+\epsilon v}(t) \cdot \big(u(t) + \epsilon\, v(t)\big), \qquad \forall t \in [0, 1], \\ \gamma_{u+\epsilon v}(0) &= \mathbb{I} \in \mathrm{GL}(n). \end{cases}$$

Let $\sigma : [0,1] \to \mathfrak{gl}(n)$ be the solution to

$$\begin{cases} \dot\sigma(t) &= \gamma_u(t)\cdot v(t) + \sigma(t)\cdot u(t) \\ \sigma(0) &= 0. \end{cases}$$

Then

$$\lim_{\epsilon\to 0} \frac{1}{\epsilon}\big(\gamma_{u+\epsilon v}(t) - \gamma_u(t)\big) = \sigma(t), \qquad \forall t \in [0,1].$$

Hint. Set $\eta_\epsilon := \frac{1}{\epsilon}(\gamma_{u+\epsilon v}(t) - \gamma_u(t))$ and then use Grönwall Lemma from Exercise 7.5.16 to show

$$\|\sigma(t) - \eta_\epsilon(t)\| \le \epsilon\alpha(t)e^{\int_a^t \beta(s)\,\mathrm{d}s},$$

where $\alpha(t) := \int_0^t \|\eta_\epsilon(s)\cdot v(s)\|\,\mathrm{d}s$ and $\beta(t) := \int_0^t \|u(s)\|\,\mathrm{d}s$.

Exercise 7.5.18 Let $\gamma : [0,1] \to \mathrm{GL}(n,\mathbb{R})$ an absolutely continuous curve with $u := \gamma' \in L^2([0,1];\mathfrak{g})$. For $v \in L^2([0,1];\mathfrak{g})$, consider $\bar\sigma(t) := \int_0^t \mathrm{Ad}_{\gamma(s)}(v(s))\,\mathrm{d}s\cdot\gamma(t)$. Then

$$\frac{\mathrm{d}\bar\sigma}{\mathrm{d}t}(t) = \gamma(t)\cdot v(t) + \bar\sigma(t)\cdot u(t), \qquad \text{for almost every } t \in [0,1].$$

Solution.

$$\begin{aligned} \frac{\mathrm{d}\bar\sigma}{\mathrm{d}t}(t) &= \mathrm{Ad}_{\gamma(t)}(v(t))\cdot\gamma(t) + \int_0^t \mathrm{Ad}_{\gamma(s)}(v(s))\,\mathrm{d}s\cdot\dot\gamma(t) \\ &\overset{5.5.8.\mathrm{i}}{=} \gamma(t)\cdot v(t)\cdot\gamma(t)^{-1}\cdot\gamma(t) + \int_0^t \mathrm{Ad}_{\gamma(s)}(v(s))\,\mathrm{d}s\cdot\gamma(t)\cdot u(t) \\ &= \gamma(t)\cdot v(t) + \bar\sigma(t)\cdot u(t). \end{aligned}$$

Exercise 7.5.19 Given a Lie group G, there is a matrix group H with the same universal covering group as G.
Hint. Use Ado's Theorem, or Birkhoff Theorem if G is nilpotent.

Exercise 7.5.20 Let $\pi : (\tilde G, \tilde V) \to (G, V)$ be a homomorphism of polarized Lie groups. Then

$$\pi \circ \mathrm{End}_{(\tilde G,\tilde V)} = \mathrm{End}_{(G,V)} \circ \pi_*,$$

where $\mathrm{End}_{(G,V)}$ denotes the endpoint map for the polarized group (G,V) and $\pi_* : L^2([0,1];\tilde V) \to L^2([0,1];V)$ is induced by the composition $\pi_*(u) := \pi_* \circ u$.
Hint. The curves $\pi(\gamma_u(t))$ and $\gamma_{\pi_* u}(t)$ satisfy the same ODE.

Exercise 7.5.21 Let $\pi : (\tilde{G}, \tilde{V}) \to (G, V)$ be a homomorphism between polarized Lie groups such that $\pi_* : \tilde{\mathfrak{g}} \to \mathfrak{g}$ is a Lie algebra isomorphism and $\pi_*(\tilde{V}) = V$. Then, Proposition 7.2.1 holds for $\tilde{G}$ if and only if it does for G.
Hint. Use Exercise 7.5.20.

Exercise 7.5.22 Exercise 7.5.21 gives a proof for Proposition 7.2.1, which, in this text, has been proved only for matrix groups.
Hint. Use Exercise 7.5.19.

Exercise 7.5.23 If (G_1, V_1) and (G_2, V_2) are polarized groups, then $(G_1 \times G_2, V_1 \times V_2)$ is a polarized group. Moreover, the polarization $V_1 \times V_2$ is bracket generating if and only if so are both V_1 and V_2.

Exercise 7.5.24 For polarized groups G_1 and G_2 consider the polarized group $G_1 \times G_2$ as in Exercise 7.5.23. A curve $\gamma = (\gamma_1, \gamma_2) : I \to G_1 \times G_2$ is abnormal if and only if at least one between γ_1 and γ_2 is abnormal.
Hint. The differential of the endpoint map splits into a block linear transformation.

Exercise 7.5.25 Let (G, V) be a polarized group with $V \neq \mathrm{Lie}(G)$. Then, the constant curve 1_G is abnormal in (G, V).

Exercise 7.5.26 In the sub-Riemannian Heisenberg group, abnormal curves are constant.

Exercise 7.5.27 Let H be the sub-Riemannian Heisenberg group. Then, every absolutely continuous curve $t \mapsto (1_H, x(t), y(t)) \in H \times \mathbb{R}^2$ is abnormal. There are injective non-smooth examples parametrized by arclength.
Hint. Use Exercises 7.5.24 and 7.5.25.

Exercise 7.5.28 On the Hilbert space $L^2([0, 1]; \mathbb{R}^n)$ with scalar product $\langle u, v\rangle := \int_0^1 u(t) \cdot v(t)\, \mathrm{d}t$, consider the energy function given by $u \mapsto \frac{1}{2}\|u\|^2 := \frac{1}{2}\langle u, u\rangle$. This function is smooth and its differential at $u \in L^2(0, 1)$ is $v \mapsto \langle u, v\rangle$.

Exercise 7.5.29 In Riemannian Lie groups, length minimizers are reparametrizations of normal curves.

Exercise 7.5.30 Equation (7.21) is an analytic ODE, and its solutions are analytic.
Hint. Solutions of analytic ODEs are analytic; see [BR89, pp. 121–128].

Exercise 7.5.31 In contact structures, as for example in every 3D sub-Riemannian Lie group, every abnormal curve is constant.
Hint. See Proposition 7.3.5.

Exercise 7.5.32 In every sub-Riemannian group with a polarization of step 2, every energy minimizer is normal.
Hint. Use Goh Theorem 7.3.4.

Exercise 7.5.33 (Goh Condition for Rank-2 Polarizations) Let G be a Lie group equipped with a rank-2 polarization Δ. Let $\gamma : [0, 1] \to G$ be a Δ-horizontal curve. Then, we have a stronger statement than Theorem 7.3.4:

$$\left\{\operatorname{Ad}_{\gamma(t)}([\Delta_1, \Delta_1]) : t \in [0, 1]\right\} \subseteq \operatorname{span}\left\{\operatorname{Ad}_{\gamma(t)}(\Delta_1) : t \in [0, 1]\right\}.$$

Exercise 7.5.34 Let $\gamma : I \to G$ be a curve in a Lie group. If $\lambda \in \mathfrak{g}^*$ and $Y \in \mathfrak{g}$, then

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\lambda \operatorname{Ad}_{\gamma(t)}(Y)\right) = \lambda \operatorname{Ad}_{\gamma(t)}[\gamma'(t), Y], \qquad \forall t \in I. \tag{7.52}$$

Solution.

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}\left(\lambda \operatorname{Ad}_{\gamma(t)}(Y)\right) &= \frac{\mathrm{d}}{\mathrm{d}s}\left(\lambda \operatorname{Ad}_{\gamma(t+s)}(Y)\right)\Big|_{s=0} \\
&= \frac{\mathrm{d}}{\mathrm{d}s}\left(\lambda \operatorname{Ad}_{\gamma(t)} \operatorname{Ad}_{\gamma(t)^{-1}\gamma(t+s)}(Y)\right)\Big|_{s=0} \\
&\overset{(5.13)}{=} \lambda \operatorname{Ad}_{\gamma(t)} \operatorname{ad}_{(\mathrm{d}L_{\gamma(t)})^{-1}\gamma(t)^{-1}\dot{\gamma}(t)}(Y) \\
&= \lambda \operatorname{Ad}_{\gamma(t)}[\gamma'(t), Y],
\end{aligned}$$

where we used that $\operatorname{Ad}_{gh} = \operatorname{Ad}_g \operatorname{Ad}_h$, that λ and Ad_g are linear.

Exercise 7.5.35 (First Derivative of the Extremal Equations) Let $\gamma : I \to G$ be a horizontal curve in a polarized group (G, V). Let $u := \gamma'$. If γ is an abnormal curve with covector $\lambda \in \mathfrak{g}^* \setminus \{0\}$, then

$$0 = \lambda \operatorname{Ad}_{\gamma(t)}[u(t), X], \qquad \forall t \in I, \forall X \in V. \tag{7.53}$$

If V is equipped with a scalar product, $e_1, \dots, e_r$ are an orthonormal basis, $u = u_1 e_1 + \dots + u_r e_r$, and γ is a normal curve with covector $\lambda \in \mathfrak{g}^*$, then

$$\dot{u}_i = \lambda \operatorname{Ad}_{\gamma(t)}[u(t), e_i], \qquad \forall t \in I, \forall i \in \{1, \dots, r\}. \tag{7.54}$$

Hint. Take the derivative of the abnormal equation (7.16) and the normal equation (7.21), using Exercise 7.5.34.

Exercise 7.5.36 Let $\sigma : [0, 1] \to \mathbb{R}^2$ be a planar curve with $u := \sigma'$ of class C^2 and with never-vanishing speed. The oriented curvature of σ is $\kappa = \frac{\sigma_1'\sigma_2'' - \sigma_2'\sigma_1''}{\|\sigma'\|^3} = \frac{u_1\dot{u}_2 - u_2\dot{u}_1}{\|u\|^3}$.

Hint. See [AT12, Equation (1.11)] for an introduction to the oriented curvature.

Exercise 7.5.37 The ODE (7.32) is of the form

$$\begin{cases} F(t,\gamma(t),\gamma'(t),\gamma''(t)) = 0, \\ \gamma(0) = 1_G, \\ \gamma'(0) = v_0, \end{cases}$$

where $v_0 = \lambda(e_1)e_1 + \lambda(e_2)e_2 \neq 0$ and $F : I \times G \times V \times V \to \mathbb{R}$ is

$$F(t,p,v,w) = \lambda(\mathrm{Ad}_p\, e_{12}) - \frac{v_1 w_2 - v_2 w_1}{\|v\|^2}.$$

The derivative of F in w is not zero at points (t,p,v,w) with $v \neq 0$. Therefore, for every $v_0 \in V \setminus \{0\}$ there are $\Omega \subset I \times G \times V$ neighborhood of $(0, 1_G, v)$, and a smooth function $\tilde{F} : \Omega \to V$ such that the ODE (7.32) can be written as

$$\begin{cases} \gamma''(t) = \tilde{F}(t,\gamma(t),\gamma'(t)), \\ \gamma(0) = 1_G, \\ \gamma'(0) = v_0. \end{cases}$$

We conclude that solutions to the ODE (7.32) are unique and smooth.

Exercise 7.5.38 On the group G_q as in Definition 7.3.7, for $(v_1,w_1),(v_2,w_2) \in V \times W$ we have that $t \mapsto (tv_1, tw_1)$ is an OPS and

$$\begin{aligned}[(v_1,w_1),(v_2,w_2)] &= \frac{1}{2}\frac{\mathrm{d}^2}{\mathrm{d}t^2}(tv_1,tw_1)(tv_2,tw_2)(-tv_1,-tw_1)(-tv_2,-tw_2)\Big|_{t=0} \\ &= (0, q(v_1,v_2)).\end{aligned}$$

Hint. Recall Definition 3.2.2.d.

Exercise 7.5.39 On the group G_q as in Definition 7.3.7, let $v \in V$ and consider a sub-Finsler norm on G_q such that $|v| = 1$. Then, the line $t \mapsto tv$ is a geodesic.
Hint. See Proposition 10.1.14.

Exercise 7.5.40 Every connected subgroup of a group G_q as in Definition 7.3.7, is of the form $G_{q'}$ for some skew-symmetric bilinear map $q' : V' \times V' \to W'$.

Exercise 7.5.41 Let G_q be a two-step nilpotent group as in Definition 7.3.7.

7.5.41.i. For $\lambda > 0$, the map $\delta_\lambda : (v,w) \in V \times W \mapsto (\lambda v, \lambda^2 w)$ is a group automorphism of $V \times W$.

7.5.41.ii. Fixing a norm on V, the sub-Finsler distance satisfies $d_{\mathrm{sF}}(\delta_\lambda(v+w),0) = \lambda d_{\mathrm{sF}}(v+w,0)$, for all $v \in V$ and $w \in W$.

7.5.41.iii. If d'_{sF} is another sub-Finsler distance with polarization Δ such that $V \subseteq \Delta_1$, then for some C we have $d'_{\mathrm{sF}} \leq C d_{\mathrm{sF}}$. In particular, after fixing a norm on W, we have

$$d'_{\mathrm{sF}}(0, w) \leq C\sqrt{w}, \qquad \forall w \in W, \tag{7.55}$$

for some $C > 0$.

Exercise 7.5.42 Let $k > 1$. Let d be an admissible distance on $\mathbb{R}^2$ that is geodesic and has the property that the translations are k-bi-Lipschitz maps. Then, the distance d is bi-Lipschitz equivalent to the Euclidean distance.
Hint. Start by following the argument in the proof of Lemma 7.4.2.

Exercise 7.5.43 (Geodesic Distances on $\mathbb{R}^2$) The only isometrically homogeneous geodesic metric spaces topologically equivalent to a plane are the Euclidean 2-space and the hyperbolic plane, equipped with left-invariant Finsler metrics.
Hint. Use Berestovskii Theorem 7.4.1, pass to a sub-Riemannian metric, and deduce that the curvature is constant.

Chapter 8
Riemannian Lie Groups

In this chapter, we discuss a classical family of sub-Finsler Lie groups: those where the polarization is the entire tangent bundle and the norm comes from a scalar product. We refer to them as *Riemannian Lie groups*, but a more precise name could be Lie groups equipped with left-invariant Riemannian structures.

The results that we present in Sect. 8.1 are classical, and a reference is [Mil76]. In Sect. 8.2, we discuss the fact that isometry groups of metric Lie groups are subgroups of isometry groups of Riemannian Lie groups. Here, the main source is [KL17]. Some other expository readings are [AT11, Pur23, Sop23].

8.1 Left-Invariant Riemannian Metrics

In this section, we extensively use various notations from Lie group theory, as we recalled in Chap. 5. For example, the maps L_h and R_h are the left translation and the right translation, respectively, by a group element h. We shall use Ad and ad from Sect. 5.5.1, and the notion of structural constants as in (5.2). We will also discuss notions from Riemannian geometry. In addition to what we presented in Sect. 3.2, we point out the books [Lee13, Lee18]. The notion of Riemannian metric, also called Riemannian metric tensor, is introduced in Sect. 3.2.3.

Definition 8.1.1 (Left-Invariant Riemannian Metric) A Riemannian metric $\langle\cdot,\cdot\rangle.$ on a Lie group G is said to be *left-invariant* if

$$\langle (L_h)_*v, (L_h)_*w\rangle_{hg} = \langle v, w\rangle_g, \qquad \forall g, h \in G, \forall v, w \in T_gG.$$

Similarly, we say that $\langle\cdot,\cdot\rangle.$ is *right-invariant* if

$$\langle (R_h)_*v, (R_h)_*w\rangle_{gh} = \langle v, w\rangle_g, \qquad \forall g, h \in G, \forall v, w \in T_gG.$$

E. Le Donne, *Metric Lie Groups*, Graduate Texts in Mathematics 306,
https://doi.org/10.1007/978-3-031-98832-5_8

It is *bi-invariant* if it is left-invariant and right-invariant.

We shall see various equivalent characterizations for those Riemannian metrics that are both left-invariant and right-invariant. One of them is the property 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 8.1.2 *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:*

8.1.2.i. $\langle\cdot,\cdot\rangle$ *is right invariant;*
8.1.2.ii. Ad_g *is an isometry, for all $g \in G$;*
8.1.2.iii. ad_X *is skew-adjoint, for all $X \in \mathfrak{g}$.*

The connectedness of G is required only for the implication $(iii) \Rightarrow (ii)$.

Proof $(i) \Leftrightarrow (ii)$: Since $\langle\cdot,\cdot\rangle$ is left-invariant, item (i) is equivalent to the conjugation $C_g = R_{g^{-1}} \circ L_g$ being isometries for all $g \in G$, which is further equivalent to $(\mathrm{Ad})_g = (\mathrm{d}C_g)_{1_G}$ being isometries for all g.

$(ii) \Rightarrow (iii)$: By Formula 5.5.7 we have $\mathrm{Ad}_{\exp(X)} = e^{\mathrm{ad}_X}$, so $e^{\mathrm{ad}(tX)} = e^{t\,\mathrm{ad}\,X}$ 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 5.4.3 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 $\frac{\mathrm{d}}{\mathrm{d}t}e^{tA} = Ae^{tA}$, again from Proposition 5.4.3. 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, the map $\mathrm{Ad}_{\exp(X)}$ is an isometry, for all $X \in \mathfrak{g}$.

So Ad_g is an isometry, for all g in a neighborhood U of 1_G in G. Since, when G is connected, every element g in G is a finite product $g = g_1 \cdot \cdots \cdot g_k$ of elements $g_1, \ldots, g_k \in U$ (see Exercise 5.8.3), then $\mathrm{Ad}_g = \mathrm{Ad}_{g_1} \circ \cdots \circ \mathrm{Ad}_{g_k}$ is an isometry.

We used the connectedness assumption of G only to prove the implication $(iii) \Rightarrow (ii)$. Without this assumption, there are counterexamples; see Exercise 8.3.3. □

8.1.1 Connections and Geodesics on Lie Groups

Recall from Sect. 3.2.1 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

$$\begin{array}{rcl} \nabla : \mathrm{Vec}(M) \times \mathrm{Vec}(M) & \to & \mathrm{Vec}(M) \\ (X, Y) & \mapsto & \nabla_X Y \end{array}$$

that is $C^\infty(M)$-linear in X, $\mathbb{R}$-linear in Y, and satisfies the *Leibniz rule*:

$$\nabla_X(fY) = (Xf)Y + f\nabla_X Y, \qquad \forall f \in C^\infty(M).$$

Definition 8.1.3 (Left-Invariant Linear Connection) Let G be a Lie group. A linear connection ∇ on G is *left-invariant* if

$$(L_g)_*\nabla_X Y = \nabla_{(L_g)_*X}(L_g)_*Y, \qquad \forall g \in G, \forall X, Y \in \mathrm{Vec}(G).$$

On every Riemannian manifold, there is a unique linear connection that is compatible with the metric and is torsion-free; for its construction and properties, see [Lee18, Chapter 5]. This connection is called *Levi-Civita connection*. The Levi-Civita connection on a Riemannian manifold M satisfies the *Koszul formula*: for all $X, Y, Z \in \mathrm{Vec}(M)$

$$\begin{aligned} &\langle \nabla_X Y, Z\rangle \\ &= \frac{1}{2}\Big(X\langle Y, Z\rangle + Y\langle Z, X\rangle - Z\langle X, Y\rangle + \langle [X, Y], Z\rangle + \langle [Z, Y], X\rangle + \langle [Z, X], Y\rangle\Big). \end{aligned} \tag{8.1}$$

Proposition 8.1.4 *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*

$$\alpha_\nabla(X, Y) = \nabla_X Y, \qquad \forall X, Y \in \mathfrak{g}.$$

Proof The result is clear once we notice that, fixing a frame $X_1, \ldots, X_n$ of left-invariant vector fields on G, we can write every arbitrary pair of vector fields on

G as $\sum_j a^j X_j$ and $\sum_i b^i X_i$, for some $a^j, b^j \in C^\infty(G)$. Then, the connection is determined:

$$\nabla_{a^j X_j}(b^i X_i) = a^j b^i \nabla_{X_j} X_i + a^j (X_j b^i) X_i = a^j b^i \alpha_\nabla(X_j, X_i) + a^j (X_j b^i) X_i.$$

□

Proposition 8.1.5 *Let G be a Lie group with a left-invariant linear connection ∇, and let X be a left-invariant vector field. The following are equivalent:*

8.1.5.i. $\nabla_X X = 0$ *(i.e., $\alpha_\nabla(X, X) = 0$);*
8.1.5.ii. *the one-parameter subgroup $t \mapsto \Phi_X^t(1_G)$ is a geodesic with respect to ∇.*

Proof The curve $t \mapsto \gamma(t) := \Phi_X^t(1_G)$ has derivative $\dot\gamma(t) = X_{\gamma(t)}$. By definition, the curve γ is a geodesic with respect to ∇ if and only if $\nabla_{\dot\gamma}\dot\gamma = 0$, 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 8.1.8.

Example 8.1.6 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 8.1.4, 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 the definition in (5.2)) with respect to the same frame, multiplied by c.

Lemma 8.1.7 *Let $\langle\cdot,\cdot\rangle$ be a left-invariant Riemannian metric on a Lie group G, and let ∇ be the associated Levi-Civita connection.*

8.1.7.i. *For all left-invariant vector fields X, Y, Z on G, we have*

$$\langle\nabla_X Y, Z\rangle = \frac{1}{2}\Big(\langle[X, Y], Z\rangle + \langle[Z, Y], X\rangle + \langle[Z, X], Y\rangle\Big). \tag{8.2}$$

8.1.7.ii. *If $X_1, \ldots, X_n$ are orthonormal left-invariant vector fields that form a basis of* $\mathrm{Lie}(G)$ *and α_{ij}^k are the corresponding structural constants, then*

$$\langle[X_i, X_j], X_k\rangle = \alpha_{ij}^k, \tag{8.3}$$

$$\langle\nabla_{X_i} X_j, X_k\rangle = \frac{1}{2}\left(\alpha_{ij}^k - \alpha_{jk}^i + \alpha_{ki}^j\right), \tag{8.4}$$

$$\nabla_{X_i} X_j = \frac{1}{2}\sum_{k=1}^n \left(\alpha_{ij}^k - \alpha_{jk}^i + \alpha_{ki}^j\right) X_k. \tag{8.5}$$

Proof Regarding 8.1.7.i, recall that the Levi-Civita connection satisfies Koszul Formula (8.1). Note that if X, Y are left-invariant vector fields, then their scalar product $\langle X, Y\rangle$ is constant along G. Hence, Koszul Formula simplifies to (8.2).

Regarding 8.1.7.ii, the structural constants are defined by the equation $[X_i, X_j] = \sum_k \alpha_{ij}^k X_k$. This implies $\langle [X_i, X_j], X_k\rangle = \sum_h \alpha_{ij}^h \langle X_h, X_k\rangle = \alpha_{ij}^k$, because the X_i's are orthonormal. From 8.1.7.i, the rest follows. □

Corollary 8.1.8 *Let G be a Lie group endowed with a left-invariant Riemannian metric. Then, the one-parameter subgroup in the direction of X is geodesic if and only if X is orthogonal to $[X, \mathfrak{g}]$.*

Proof It follows from (8.2) and Proposition 8.1.5. □

Theorem 8.1.9 *Let $\langle\cdot,\cdot\rangle$ be a left-invariant Riemannian metric on a Lie group G and let ∇ be the associated Levi-Civita connection. The following are equivalent:*

8.1.9.i. *The group exponential map* $\exp$ *coincides with the Riemannian exponential map* $\exp_{1_G}$*, i.e., the family of one-parameter subgroups is exactly the family of the geodesics from the identity element 1_G;*

8.1.9.ii. *If $X, Y \in \mathfrak{g}$, then $\alpha_\nabla(X, Y) = \frac{1}{2}[X, Y]$, i.e.,*

$$\nabla_X Y = \frac{1}{2}[X, Y], \qquad \forall X, Y \in \mathfrak{g};$$

8.1.9.iii. *The map* ad_Z *is skew-adjoint, for all $Z \in \mathfrak{g}$.*

Proof Note that the formula (8.2) in Lemma 8.1.7 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, the point (i), which by Proposition 8.1.5 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 8.1.10 Note that the equation $\tilde{\nabla}_X Y := \frac{1}{2}[X, Y]$, for $X, Y \in \mathfrak{g}$ always define a left-invariant connection $\tilde{\nabla}$ on G. Condition (ii) of Theorem 8.1.9 is that such a $\tilde{\nabla}$ is the Levi-Civita connection ∇.

8.1.2 Curvatures of Left-Invariant Metrics

For the following discussion, be aware that our convention for the *Riemannian curvature tensor* is

$$R(X, Y, \cdot) := \nabla_X \nabla_Y - \nabla_Y \nabla_X - \nabla_{[X,Y]}. \tag{8.6}$$

Also, recall that given two linearly independent tangent vectors X, Y at the same point of a Riemannian manifold, the *sectional curvature* of their spanned plane is

$$\mathrm{Sec}(X, Y) := \frac{\langle R(X, Y, Y), X\rangle}{\|X\|^2\|Y\|^2 - \langle X, Y\rangle^2}. \tag{8.7}$$

Proposition 8.1.11 *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 corresponding 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

$$\begin{aligned} \mathrm{Sec}(X_1, X_2) = \sum_{\ell=1}^{n} \Bigg[&-\frac{1}{2}\alpha_{12}^{\ell}\left(\alpha_{12}^{\ell} - \alpha_{2\ell}^{1} - \alpha_{\ell 1}^{2}\right) \\ &-\frac{1}{4}\left(\alpha_{12}^{\ell} - \alpha_{2\ell}^{1} + \alpha_{\ell 1}^{2}\right)\left(\alpha_{12}^{\ell} + \alpha_{2\ell}^{1} - \alpha_{\ell 1}^{2}\right) \\ &-\alpha_{\ell 1}^{1}\alpha_{\ell 2}^{2}\Bigg]. \end{aligned} \tag{8.8}$$

Proof Recall that we defined R by (8.6). So

$$\begin{aligned} R(X_i, X_j, X_k) &= \nabla_{X_i}\nabla_{X_j}X_k - \nabla_{X_j}\nabla_{X_i}X_k - \nabla_{\sum_\ell \alpha_{ij}^{\ell}X_\ell}X_k \\ &\overset{(8.5)}{=} \nabla_{X_i}\left(\frac{1}{2}\sum_{\ell}(\alpha_{jk}^{\ell} - \alpha_{k\ell}^{j} + \alpha_{\ell j}^{k})X_\ell\right) \end{aligned}$$

$$
\begin{aligned}
&- \nabla_{X_j}\left(\frac{1}{2}\sum_\ell (\alpha^\ell_{jk} - \alpha^i_{k\ell} + \alpha^k_{\ell i})X_\ell\right)\\
&\qquad - \sum_\ell \alpha^\ell_{ij}\frac{1}{2}\sum_h (\alpha^h_{\ell k} - \alpha^\ell_{kh} + \alpha^k_{h\ell})X_h\\
&= \frac{1}{2}\sum_\ell (\alpha^\ell_{jk} - \alpha^j_{k\ell} + \alpha^k_{\ell j})\frac{1}{2}\sum_h (\alpha^h_{i\ell} - \alpha^i_{\ell h} + \alpha^\ell_{hi})X_h\\
&\qquad - \frac{1}{2}\sum_\ell (\alpha^\ell_{ik} - \alpha^i_{k\ell} + \alpha^k_{\ell i})\frac{1}{2}\sum_h (\alpha^h_{j\ell} - \alpha^j_{\ell h} + \alpha^\ell_{hj})X_h\\
&\qquad - \sum_\ell \frac{1}{2}\alpha^\ell_{ij}\sum_h (\alpha^h_{\ell k} - \alpha^\ell_{kh} + \alpha^k_{h\ell})X_h\\
&= \sum_{\ell,h}\left[\frac{1}{4}(\alpha^\ell_{jk} - \alpha^j_{k\ell} + \alpha^k_{\ell j})(\alpha^h_{i\ell} - \alpha^i_{\ell h} + \alpha^\ell_{hi})\right.\\
&\qquad - \frac{1}{4}(\alpha^\ell_{ik} - \alpha^i_{k\ell} + \alpha^k_{\ell i})(\alpha^h_{j\ell} - \alpha^j_{\ell h} + \alpha^\ell_{hj})\\
&\qquad \left. - \frac{1}{2}\alpha^\ell_{ij}(\alpha^h_{\ell k} - \alpha^\ell_{kh} + \alpha^k_{h\ell})\right]X_h.
\end{aligned}
$$

Regarding the sectional curvature (8.7), 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

$$
\begin{aligned}
&\mathrm{Sec}(X_1, X_2)\\
&= \langle R(X_1, X_2, X_2), X_1\rangle\\
&= \sum_\ell \left[\frac{1}{4}(-\alpha^2_{2\ell} + \alpha^2_{\ell 2})(\alpha^1_{1\ell} - \alpha^1_{\ell 1}) + \right.\\
&\qquad - \frac{1}{4}(\alpha^\ell_{12} - \alpha^1_{2\ell} + \alpha^2_{\ell 1})(\alpha^1_{2\ell} - \alpha^2_{\ell 1} + \alpha^\ell_{12}) +\\
&\qquad \left. - \frac{1}{2}\alpha^\ell_{12}(\alpha^1_{\ell 2} - \alpha^\ell_{21} + \alpha^2_{1\ell})\right]\\
&= \sum_\ell \left[\frac{1}{4}\left(2\alpha^2_{\ell 2}(-2\alpha^1_{\ell 1})\right) + \right.\\
&\qquad - \frac{1}{4}(\alpha^\ell_{12} - \alpha^1_{2\ell} + \alpha^2_{\ell 1})(\alpha^\ell_{12} + \alpha^1_{2\ell} - \alpha^2_{\ell 1}) +
\end{aligned}
$$

$$- \frac{1}{2}\alpha^{\ell}_{12}(\alpha^{\ell}_{12} - \alpha^{1}_{2\ell} - \alpha^{2}_{\ell 1})\Bigg]$$

$$= \sum_{\ell} \Bigg[-\alpha^{1}_{\ell 1}\alpha^{2}_{\ell 2} - \frac{1}{4}(\alpha^{\ell}_{12} - \alpha^{1}_{2\ell} + \alpha^{2}_{\ell 1})(\alpha^{\ell}_{12} + \alpha^{1}_{2\ell} - \alpha^{2}_{\ell 1}) - \frac{1}{2}\alpha^{\ell}_{12}(\alpha^{\ell}_{12} - \alpha^{1}_{2\ell} - \alpha^{2}_{\ell 1}) \Bigg].$$

□

Lemma 8.1.12 *Let* $X \in \mathrm{Lie}(G)$. *If* ad_X *is skew-adjoint, then* $\mathrm{Sec}(X, Y) \geq 0$, *for each* $Y \in \mathrm{Lie}(G)$ *that is linearly independent from* X.

Proof Recall that ad_X being skew-adjoint means

$$\langle \mathrm{ad}_X Y, Z \rangle = -\langle Y, \mathrm{ad}_X Z \rangle, \qquad \forall Y, Z \in \mathrm{Lie}(G).$$

Let $X \in \mathrm{Lie}(G)$ be such that ad_X is skew-adjoint. Since $\mathrm{ad}_{\lambda X} = \lambda\, \mathrm{ad}_X$, we can assume that $\langle X, X \rangle = 1$. 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 $\mathrm{Lie}(G)$ with $X_1 = X$ and $X_2 = Y$, 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

$$\alpha^k_{1j} = \langle \mathrm{ad}_{X_1} X_j, X_k \rangle = -\langle X_j, \mathrm{ad}_{X_1} X_k \rangle = -\alpha^j_{1k}.$$

Thus, we have $\alpha^2_{\ell 1} = -\alpha^2_{1\ell} = \alpha^{\ell}_{12}$ and $\alpha^1_{\ell 1} = 0$. Therefore, formula (8.8) simplifies to

$$\begin{aligned}
\mathrm{Sec}(X_1, X_2) &= \sum_{\ell} -\frac{1}{2}\alpha^{\ell}_{12}\big(-\alpha^{1}_{2\ell}\big) - \frac{1}{4}\big(2\alpha^{\ell}_{12} - \alpha^{1}_{2\ell}\big)\big(\alpha^{1}_{2\ell}\big) \\
&= \sum_{\ell} \frac{1}{2}\alpha^{\ell}_{12}\alpha^{1}_{2\ell} - \frac{1}{2}\alpha^{\ell}_{12}\alpha^{1}_{2\ell} + \frac{1}{4}\big(\alpha^{1}_{2\ell}\big)^2 \\
&= \sum_{\ell} \frac{1}{4}\big(\alpha^{1}_{2\ell}\big)^2 \geq 0.
\end{aligned}$$

□

8.1.3 Bi-Invariant Metrics

Recall that Riemannian metrics of Lie groups that are left-invariant and right-invariant are said to be bi-invariant. We obtained characterizations for bi-invariance of metrics in Theorems 8.1.2 and 8.1.9.

Corollary 8.1.13 *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:*

8.1.13.i. $\langle\cdot,\cdot\rangle$ *is bi-invariant;*
8.1.13.ii. Ad_g *is an isometry, for all* $g \in G$*;*
8.1.13.iii. ad_X *is skew-adjoint, for all* $X \in \mathfrak{g}$*;*
8.1.13.iv. $\exp_{1_G} = \exp$*;*
8.1.13.v. $\nabla_X Y = \frac{1}{2}[X, Y]$*, for all* $X, Y \in \mathfrak{g}$*;*

Lemma 8.1.12 gives a non-trivial property of bi-invariant metrics:

Corollary 8.1.14 *If a connected Lie group is equipped with a bi-invariant metric, then all its sectional curvatures are nonnegative.*

Theorem 8.1.15 *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.,

$$\langle\langle X, Y\rangle\rangle = \int_G \langle \mathrm{Ad}_g(X), \mathrm{Ad}_g(Y)\rangle \,\mathrm{d}\,\mathrm{vol}(g), \qquad \forall X, Y \in T_{1_G}G.$$

To observe that $\langle\langle\cdot,\cdot\rangle\rangle$ is finite, we point out that $\mathrm{vol}(G) < \infty$, because G is compact, and the function $g \mapsto \langle \mathrm{Ad}_g(X), \mathrm{Ad}_g(Y)\rangle$ is bounded, because it is a continuous function on the compact group G. 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 obtain a left-invariant Riemannian metric for which Ad_g are isometries. Theorem 8.1.2 implies that this metric is right-invariant. □

Because for each bi-invariant Riemannian metric, the volume measure is bi-invariant, we obtain a first consequence:

Corollary 8.1.16 *Every compact Lie group can be equipped with a probability measure that is bi-invariant.*

Because for each bi-invariant Riemannian metric, the exponential map coincides with the Riemannian exponential (see (8.1.13).iv), and the Riemannian exponential on complete connected manifolds is surjective (see Hopf-Rinow Theorem 3.1.7), we obtain a second consequence:

Corollary 8.1.17 *On every compact connected Lie group G, the exponential map* $\exp : \mathrm{Lie}(G) \to G$ *is surjective.*

The next result is another characterization of groups admitting bi-invariant distance functions that are admissible, in the sense defined at page 172.

Theorem 8.1.18 *Let G be a connected Lie group. Then the following are equivalent:*

8.1.18.i. There exists an admissible bi-invariant distance function on G;
8.1.18.ii. The set Ad_G *is a compact subset of* $\mathrm{GL}(\mathfrak{g})$*;*
8.1.18.iii. There exists a bi-invariant Riemannian metric on G;
8.1.18.iv. The Lie group G is the direct product of a compact group and a vector group, that is, a group isomorphic to $(\mathbb{R}^n, +)$ *for some* $n \in \mathbb{N}$.

Sketch of the Proof If there exists an admissible bi-invariant distance function on G, then by Lemma 6.2.5, we may additionally assume that this distance function is boundedly compact. Thus, by Ascoli–Arzelá theorem, the space of isometries fixing 1_G is compact; see Proposition 6.2.7. Since the distance function is bi-invariant, the conjugation maps are isometries. Therefore, the set $\mathrm{Ad}_G \subseteq \mathrm{GL}(\mathfrak{g})$ of their differentials is compact. Alternatively, one can directly construct a bi-invariant Riemannian metric on G using Lemma 8.2.2.

If the set Ad_G is compact in the space of linear transformations of the Lie algebra $\mathfrak{g}$, then, one constructs a bi-invariant Riemannian metric on G, as in the proof of Theorem 8.1.15.

Clearly, every bi-invariant Riemannian metric on G gives an admissible bi-invariant distance function. Thus, the first three items are equivalent.

Regarding the last item, if the Lie group G is the product of a compact group and a commutative group, then, obviously, the set Ad_G is compact. The converse implication is more difficult: It is a result of Milnor that every connected Lie group that admits a bi-invariant Riemannian metric is the Cartesian product of a compact group and a commutative group; see [Mil76, Lemma 7.5]. □

From the topological viewpoint, every Lie group is homeomorphic to the product of a compact group and a vector space. This result is due to Iwasawa; see Exercise 8.3.13.

8.1.4 More Results on Curvature

In Milnor's article [Mil76], one can also find the following results on the curvature of left-invariant Riemannian metrics on Lie groups.

Theorem 8.1.19 (Milnor, [Mil76, Theorem 3.3]) *If* $\mathrm{Lie}(G)$ *is not commutative, then G admits a left-invariant metric of strictly negative* scalar *curvature.*

Theorem 8.1.20 (Milnor, [Mil76, 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 that the *Ricci curvature* in the direction of X is $\mathrm{Ric}(X) := \sum_i \mathrm{Sec}(X, X_i)$, where $X_1, \ldots, X_n$ is any orthonormal frame.

Theorem 8.1.21 (Milnor, [Mil76, Theorem 2.5]) *If there are non-zero* $X, Y, Z \in \mathrm{Lie}(G)$ *such that* $[X, Y] = Z$*, then there is a left-invariant metric on G such that* $\mathrm{Ric}(X) < 0 < \mathrm{Ric}(Z)$.

Theorem 8.1.22 (Milnor, [Mil76, Corollary 7.7]) *Every Lie group whose universal covering space is compact admits a bi-invariant metric of constant Ricci curvature* $+1$.

8.2 Isometries of Metric Groups as Riemannian Isometries

Recall from Sect. 6.3 that a metric Lie group is a Lie group equipped with a left-invariant distance function that induces the manifold topology. In this section, we prove that isometries between metric Lie groups are Riemannian isometries for some left-invariant Riemannian metric. If M is a Lie group and ρ is a left-invariant Riemannian metric tensor on M, then one has an induced Riemannian distance d_ρ and, by the theorem of Myers and Steenrod [MS39], the group $\mathrm{Isom}(M, d_\rho)$ of distance-preserving bijections coincides with the group $\mathrm{Isom}(M, \rho)$ of tensor-preserving diffeomorphisms. In the following, we shall write (M, ρ) to denote the metric Lie group (M, d_ρ).

Theorem 8.2.1 *If* (M_1, d_1) *and* (M_2, d_2) *are connected metric Lie groups, then there exists left-invariant Riemannian metrics* ρ_1 *and* ρ_2 *on* M_1 *and* M_2*, respectively, such that* $\mathrm{Isom}(M_i, d_i) \subseteq \mathrm{Isom}(M_i, \rho_i)$ *for* $i \in \{1, 2\}$ *and for every isometry* $F\colon (M_1, d_1) \to (M_2, d_2)$ *the map* $F\colon (M_1, \rho_1) \to (M_2, \rho_2)$ *is a Riemannian isometry.*

Before proving Theorem 8.2.1, we provide an auxiliary result. We first consider the case $(M_1, d_1) = (M_2, d_2)$.

Lemma 8.2.2 *If* (M, d) *is a connected metric Lie group, then there is a left-invariant Riemannian metric* ρ *such that* $\mathrm{Isom}(M, d) \subseteq \mathrm{Isom}(M, \rho)$.

Proof Let S be the stabilizer subgroup of $\mathrm{Isom}(M, d)$ at the identity element $1 = 1_M$ of M. By Proposition 6.2.7, together with Lemma 6.2.5, the topological group S is compact. Let μ_S be the probability Haar measure on S. Moreover, from the discussions in Sect. 6.3.1, we know that S acts smoothly on M.

Fix a scalar product $\langle\langle \cdot, \cdot \rangle\rangle$ on the tangent space $T_1 M$ at 1. Consider for $v, w \in T_1 M$ the value

$$\langle v, w \rangle := \int_S \langle\langle \mathrm{d}Fv, \mathrm{d}Fw \rangle\rangle \, \mathrm{d}\mu_S(F).$$

Then $\langle \cdot, \cdot \rangle$ defines an S-invariant scalar product on $T_1 M$, and one can take ρ as the left-invariant Riemannian metric that coincides with $\langle \cdot, \cdot \rangle$ at the identity. □

Proof of Theorem 8.2.1 By Lemma 8.2.2 let ρ_2 be a Riemannian metric on M_2 with

$$\mathrm{Isom}(M_2, d_2) \subseteq \mathrm{Isom}(M_2, \rho_2). \tag{8.9}$$

Fix an isometry $\tilde{F} : (M_1, d_1) \to (M_2, d_2)$; if there is none, then there is nothing else to prove. By Theorem 6.3.1, the map $\tilde{F}$ is smooth, and we may define a Riemannian metric on M_1 by $\rho_1 := \tilde{F}^* \rho_2$. There are two things to verify: a) $\mathrm{Isom}(M_1, d_1) \subseteq \mathrm{Isom}(M_1, \rho_1)$, which in particular implies that ρ_1 is left-invariant and b) every isometry $F : (M_1, d_1) \to (M_2, d_2)$ is an isometry of Riemannian manifolds.

Since by construction $\tilde{F}$ is also a Riemannian isometry, the map $I \mapsto \tilde{F} \circ I \circ \tilde{F}^{-1}$ is a bijection between $\mathrm{Isom}(M_1, d_1)$ and $\mathrm{Isom}(M_2, d_2)$ and between $\mathrm{Isom}(M_1, \rho_1)$ and $\mathrm{Isom}(M_2, \rho_2)$. Therefore the inclusion (8.9) implies the inclusion $\mathrm{Isom}(M_1, d_1) \subseteq \mathrm{Isom}(M_1, \rho_1)$.

From $F \circ \tilde{F}^{-1} \in \mathrm{Isom}(M_2, d_2) \subseteq \mathrm{Isom}(M_2, \rho_2)$ we obtain $(F \circ \tilde{F}^{-1})^* \rho_2 = \rho_2$. Consequently, we conclude $F^* \rho_2 = \tilde{F}^* (F \circ \tilde{F}^{-1})^* \rho_2 = \rho_1$. □

8.3 Exercises

Exercise 8.3.1 Let G be a Lie group equipped with a Riemannian metric tensor that induces a distance function d. Then, the Riemannian metric tensor is left-invariant if and only if left translations are isometries with respect to d.

Exercise 8.3.2 For each left-invariant distance function d on a group G, 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.

Exercise 8.3.3 Consider the Lie group $G := \mathbb{R} \rtimes_\theta \mathbb{Z}$ the semi-direct product of groups given by $\theta(m)(t) := 2^m t$ for $m \in \mathbb{Z}$ and $t \in \mathbb{R}$, as in Definition 5.6.6. Equip G with any Riemannian left-invariant metric. Then, every map ad_X is skew-adjoint, for all $X \in \mathrm{Lie}(G)$, but for some $g \in G$ the map Ad_g is not an isometry. Compare this example with Theorem 8.1.2.

Exercise 8.3.4 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.

Exercise 8.3.5 Levi-Civita connections of left-invariant Riemannian metrics on Lie groups are left-invariant.

Exercise 8.3.6 A linear connection ∇ on G is left-invariant if and only if the Christoffel symbols Γ_{ij}^k defined by $\nabla_{X_i} X_j = \sum_k \Gamma_{ij}^k X_k$, with respect to some/every frame of left-invariant vector fields $X_1, \ldots, X_n$, are constant functions on the group.

Exercise 8.3.7 Let G be a Riemannian Lie group, ∇ its Levi-Civita connection, and R its Riemannian curvature tensor. For each $X, Y \in \mathfrak{g}$, define $U(X, Y) := \frac{1}{2}(\nabla_X Y + \nabla_Y X)$, which is the symmetric part of ∇. Then, for all $X, Y, Z \in \mathfrak{g}$ one has

8.3.7.i. $\nabla_X Y = U(X, Y) + \frac{1}{2}[X, Y]$,
8.3.7.ii. $\langle U(X, Y), Z\rangle = \frac{1}{2}\langle X, [Z, Y]\rangle + \frac{1}{2}\langle Y, [Z, X]\rangle$
8.3.7.iii.

$$\begin{aligned}\langle R(X, Y)Y, X\rangle \\ &= \|U(X, Y)\|^2 - \langle U(X, X), U(Y, Y)\rangle - \frac{3}{4}\|[X, Y]\|^2 \\ &\quad - \frac{1}{2}\langle [X, [X, Y]], Y\rangle - \frac{1}{2}\langle [Y, [Y, X]], X\rangle. \qquad (8.10)\end{aligned}$$

Hint. Use Koszul's formula and the fact that the Levi-Civita connection has zero torsion.

Exercise 8.3.8 Every Lie group admits a left-invariant Riemannian metric and a left-invariant measure, as follows. Let $X_1, \ldots, X_n$ be left-invariant vector fields forming a basis of $\mathrm{Lie}(G)$. Consider the Riemannian metric that makes $X_1, \ldots, X_n$ orthonormal and the differential n-form vol for which $\mathrm{vol}(X_1, \ldots, X_n) \equiv 1$.

Exercise 8.3.9 If G is a compact Lie group, then it admits a left-invariant Riemannian volume form that gives a probability measure.

Exercise 8.3.10 Calculate the sectional curvatures of the Lie groups G_z from Example 6.4.4 when equipped with some left-invariant Riemannian metric.

Exercise 8.3.11 Let $\mathfrak{g}$ be the 4D Lie algebra with basis $X_1, \ldots, X_4$ and only nonzero structural constants, as in (5.2),

$$c_{12}^3 = 1, \quad c_{41}^1 = 1, \quad c_{42}^2 = 1, \quad c_{43}^3 = 2.$$

Consider a Lie group G with $\mathfrak{g}$ as Lie algebra, and consider the left-invariant Riemannian structure on G for which $X_1, \ldots, X_4$ are orthonormal. Verify the following values for the sectional curvature:

$$\begin{aligned}
\mathrm{Sec}(X_1, X_2) &= -7/4, & \mathrm{Sec}(X_1, X_4) &= -1,\\
\mathrm{Sec}(X_1, X_3) &= -7/4, & \mathrm{Sec}(X_2, X_4) &= -1,\\
\mathrm{Sec}(X_2, X_3) &= -7/4, & \mathrm{Sec}(X_3, X_4) &= -4.
\end{aligned}$$

Hint. Apply (8.8).

Exercise 8.3.12 Let $\mathfrak{h}$ be the Heisenberg Lie algebra with basis $X_1, X_2, X_3 := [X_1, X_2]$. Consider the left-invariant Riemannian structure on the Heisenberg group for which X_1, X_2, X_3 are orthonormal. Verify the following values for the sectional curvature:

$$\mathrm{Sec}(X_1, X_2) = -3/4,$$
$$\mathrm{Sec}(X_1, X_3) = 1/4,$$
$$\mathrm{Sec}(X_2, X_3) = 1/4.$$

Hint. In this basis, the only nontrivial structural constant is $c_{12}^3 = 1$. Apply (8.8).

Exercise 8.3.13 (Iwasawa's Theorem; see [Mil76, page 327]) ☠ Let G be a connected Lie group. Then:

8.3.13.i. Every compact subgroup is contained in a maximal compact subgroup H.
8.3.13.ii. Every maximal compact subgroup is a connected Lie subgroup.
8.3.13.iii. Every two maximal compact subgroups are conjugate.
8.3.13.iv. As a topological space, G is homeomorphic to the product of H and some Euclidean space $\mathbb{R}^m$.

Chapter 9
Nilpotent Lie Groups

Nilpotent Lie groups play a fundamental role in the study of geometric structures and differential equations. These groups are characterized by an algebraic property known as nilpotency. The nilpotency assumption, which also reflects a condition on the Lie algebra, imposes a certain level of commutativity for the group's structure, allowing us to explore the interplay between algebra and geometry. It paves the way for implications in geometric analysis, geometric group theory, harmonic analysis, and control theory, as well as number theory, dynamics, and representation theory. In this chapter, we delve into the world of nilpotent Lie groups.

A recent reference that deserves a 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 metric spaces of sub-Riemannian manifolds, similar to how Euclidean vector spaces serve as tangents to Riemannian manifolds. We will discover that, akin to vector spaces, these metric tangents possess nilpotency, simply connectedness, and dilation structures.

9.1 Nilpotent Lie Algebras

We begin this section by introducing the concept of a nilpotent Lie algebra. Nilpotent Lie algebras are those for which iterated brackets $[x_1, [x_2, [x_3, [\ldots]]]]$ of sufficiently large order vanish. We anticipate that for connected Lie groups, a Lie algebra is nilpotent if and only if the group is nilpotent as a group, according to

E. Le Donne, *Metric Lie Groups*, Graduate Texts in Mathematics 306,
https://doi.org/10.1007/978-3-031-98832-5_9

Definition 9.3.3. This correspondence between nilpotent Lie algebras and nilpotent Lie groups is a result that we will further explore in Sect. 9.4.2. Typical examples of nilpotent Lie algebras are Lie algebras of strictly upper-triangular matrices, where the diagonal elements are all zero. The first fundamental result that we present is Engel's Theorem, which translates nilpotency into a pointwise condition; see Sect. 9.1.3. We close the section with the statement of Birkhoff Embedding Theorem, which we will only prove for positively graded Lie algebras in Sect. 9.2.4. Birkhoff's theorem will tell us that every nilpotent Lie algebra is isomorphic to some subalgebra of some space of strictly upper-triangular matrices; see Corollary 9.1.20.

Definition 9.1.1 (Nilpotent Lie Algebra) Let $\mathfrak{g}$ be a Lie algebra. The elements of the *descending central series* of $\mathfrak{g}$, also called *lower central series* of $\mathfrak{g}$, are inductively defined by

$$\begin{aligned} C^1(\mathfrak{g}) := \mathfrak{g}^{(1)} := \mathfrak{g}, \qquad & C^2(\mathfrak{g}) := \mathfrak{g}^{(2)} := [\mathfrak{g},\mathfrak{g}] = [\mathfrak{g}, C^1(\mathfrak{g})], \\ & C^n(\mathfrak{g}) := \mathfrak{g}^{(n)} := [\mathfrak{g}, C^{n-1}(\mathfrak{g})], \qquad \forall n \in \mathbb{N}. \end{aligned}$$

Here, for $V, W \subseteq \mathfrak{g}$, we use the notation $[V, W] := \operatorname{span}\{[v, w] : v \in V, w \in W\}$. The space $C^2(\mathfrak{g})$ is called the *commutator subalgebra*. 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 *nilpotency step* or, simply, *step*) of $\mathfrak{g}$, and $\mathfrak{g}$ is said *d-step nilpotent*.

One can rephrase the definition by 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 every bracket of order s is.

Remark 9.1.2 Each $C^n(\mathfrak{g})$ is an ideal and actually

$$[C^n(\mathfrak{g}), C^n(\mathfrak{g})] \subseteq [C^n(\mathfrak{g}), \mathfrak{g}] \stackrel{\text{def}}{=} C^{n+1}(\mathfrak{g}) \subseteq C^n(\mathfrak{g}),$$

where the last inclusion can be shown by induction, noting that $C^2(\mathfrak{g}) \subseteq C^1(\mathfrak{g})$.

A nilpotent Lie algebra $\mathfrak{g}$ has always non-trivial center $Z(\mathfrak{g})$ by Proposition 9.1.3.iii. In fact, if $\mathfrak{g}$ is s-step nilpotent, the subalgebra $\mathfrak{g}^{(s)}$ is central. However, the center might be strictly larger than $\mathfrak{g}^{(s)}$; see Exercise 9.5.16.

Proposition 9.1.3 *Let $\mathfrak{g}$ be a Lie algebra.*

9.1.3.i. If $\mathfrak{g}$ is nilpotent, then subalgebras and homomorphic images of $\mathfrak{g}$ are nilpotent.
9.1.3.ii. If $\mathfrak{a} < Z(\mathfrak{g})$ and $\mathfrak{g}/\mathfrak{a}$ is nilpotent, then $\mathfrak{g}$ is nilpotent.
9.1.3.iii. If $\mathfrak{g} \neq \{0\}$ and $\mathfrak{g}$ is nilpotent of step s, then $\{0\} \neq C^s(\mathfrak{g}) \subseteq Z(\mathfrak{g})$.
9.1.3.iv. If $\mathfrak{g}$ is nilpotent of step s, then $\operatorname{ad}_x^s \equiv 0$, for every $x \in \mathfrak{g}$.
9.1.3.v. If $\mathfrak{i}$ is an ideal of $\mathfrak{g}$, then $C^n(\mathfrak{i})$ is an ideal of $\mathfrak{g}$, for every $n \in \mathbb{N}$.

Proof

(i) Assume $\mathfrak{g}$ nilpotent. Let $\mathfrak{h}$ be a subalgebra of $\mathfrak{g}$. Then $[\mathfrak{h}, \mathfrak{h}] \subset [\mathfrak{g}, \mathfrak{g}]$ and so, by induction, $C^n(\mathfrak{h}) \subset C^n(\mathfrak{g})$. Hence, the Lie algebra $\mathfrak{h}$ is nilpotent. Moreover, if we consider a Lie algebra homomorphism $\alpha : \mathfrak{g} \to \mathfrak{h}$, then $[\alpha(\mathfrak{g}), \alpha(\mathfrak{g})] = \alpha([\mathfrak{g}, \mathfrak{g}])$. By induction we have that

$$C^n(\alpha(\mathfrak{g})) = \alpha(C^n(\mathfrak{g})), \qquad \forall n \in \mathbb{N}. \tag{9.1}$$

Consequently, the image $\alpha(\mathfrak{g})$ is nilpotent, since so is $\mathfrak{g}$.

(ii) If $\mathfrak{g}/\mathfrak{a}$ is nilpotent, then by definition there is $n \in \mathbb{N}$ such that $C^n(\mathfrak{g}/\mathfrak{a}) = \{0\}$ in $\mathfrak{g}/\mathfrak{a}$, i.e., $C^n(\mathfrak{g}/\mathfrak{a}) = \mathfrak{a}/\mathfrak{a}$. Now we apply (9.1) with α the projection $\mathfrak{g} \to \mathfrak{g}/\mathfrak{a}$ and we deduce that $C^n(\mathfrak{g}) + \mathfrak{a} = C^n(\mathfrak{g}/\mathfrak{a}) = \mathfrak{a}$, i.e., $C^n(\mathfrak{g}) \subset \mathfrak{a} \subset Z(\mathfrak{g})$, where the last inclusion holds by assumption. This implies that $\mathfrak{g}$ is nilpotent:

$$C^{n+1}(\mathfrak{g}) \stackrel{\text{def}}{=} [\mathfrak{g}, C^n(\mathfrak{g})] \subset [\mathfrak{g}, Z(\mathfrak{g})] = \{0\}.$$

(iii) By hypothesis, the natural number s is such that $C^{s+1}(\mathfrak{g}) = \{0\}$ and $C^s(\mathfrak{g}) \neq \{0\}$. Then, $\{0\} \neq C^s(\mathfrak{g}) \subseteq Z(\mathfrak{g})$, since $[C^s(\mathfrak{g}), \mathfrak{g}] \stackrel{\text{def}}{=} C^{s+1}(\mathfrak{g}) = \{0\}$.

(iv) Since $\mathfrak{g}$ is s-step nilpotent we have that $C^{s+1}(\mathfrak{g}) = \{0\}$. So for every $x \in \mathfrak{g}$

$$(\mathrm{ad}(x))^s(\mathfrak{g}) = \underbrace{[x, [x, \dots [x}_{s \text{ times}}, \mathfrak{g}] \dots] \subseteq C^{s+1}(\mathfrak{g}) = \{0\}.$$

(v) It follows from the general easy fact that if $\mathfrak{i}$, $\mathfrak{j}$ are ideals of $\mathfrak{g}$, then so is $[\mathfrak{i}, \mathfrak{j}]$. □

Proposition 9.1.3.iv states that if $\mathfrak{g}$ is nilpotent, then each map $Y \in \mathfrak{g} \mapsto [X, Y] \in \mathfrak{g}$ is a nilpotent transformation in the sense of Definition 9.1.12.i. We shall see in Theorem 9.1.18 that the inverse implication holds true.

9.1.1 Examples of Nilpotent Lie Algebras

We present some examples of nilpotent Lie algebras.

Example 9.1.4 (Abelian Lie Algebras) *Commutative Lie algebras* are those for which $[X, Y] = [Y, X]$, for all elements X and Y. By anticommutativity, this is equivalent to $[\cdot, \cdot] \equiv 0$. We also refer to commutative Lie algebras as *abelian Lie algebras*. Consequently, a Lie algebra $\mathfrak{g}$ is commutative if and only if it is nilpotent with nilpotency step equal to 1 because $C^2(\mathfrak{g}) \stackrel{\text{def}}{=} [\mathfrak{g}, \mathfrak{g}] = \{0\}$.

Example 9.1.5 (Heisenberg Lie Algebra) The Heisenberg Lie algebra

$$\mathfrak{nil}_3 := \left\{ \begin{bmatrix} 0 & x & z \\ 0 & 0 & y \\ 0 & 0 & 0 \end{bmatrix} : x, y, z \in \mathbb{R} \right\} \subseteq \mathfrak{gl}(3)$$

is nilpotent of step 2 because for such a Lie algebra $\mathfrak{g} := \mathfrak{nil}_3$

$$C^2(\mathfrak{g}) = \left\{ \begin{bmatrix} 0 & 0 & z \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} : z \in \mathbb{R} \right\} \qquad \text{and} \qquad C^3(\mathfrak{g}) = \{\mathbf{0}\}.$$

Example 9.1.6 (Nilpotent Lie Algebras of Step 2) We revise Definition 7.3.7 of nilpotent Lie algebras of step 2. Let V, W be vector spaces and $q : V \times V \to W$ a 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 2-step Lie algebra unless q is identically equal to 0, in which case it is a commutative Lie algebra. Namely, we have $[[X, Y], Z] = 0$ for every $X, Y, Z \in V \times W$.
As a more specific example, for $n \in \mathbb{N}$, we consider

$$\begin{aligned} V &:= \Lambda^1(\mathbb{R}^n) = \{\text{1-forms on } \mathbb{R}^n\}, \\ W &:= \Lambda^2(\mathbb{R}^n) = \{\text{2-forms on } \mathbb{R}^n\}, \\ q(v_1, v_2) &:= v_1 \wedge v_2, \end{aligned} \tag{9.2}$$

where $\wedge$ is the wedge of 1-forms. Then $\Lambda^1(\mathbb{R}^n) \times \Lambda^2(\mathbb{R}^n)$ becomes a 2-step Lie algebra called the *free-nilpotent Lie algebra* of rank n and step 2. We will generalize this example in Example 9.1.8.

One common convention in describing Lie algebras—and one that we shall often use—is the following. Suppose that $\mathfrak{g} = \mathbb{R}\text{-}\operatorname{span}\{X_1, \ldots, X_n\}$. To describe the Lie algebra structure of $\mathfrak{g}$, it suffices to give $[X_i, X_j]$ for all $i < j$ in terms of $X_1, \ldots, X_n$. We can shorten this description considerably by giving only the non-zero brackets; all the others are assumed to be zero.

Example 9.1.7 (Filiform Algebras of the First Kind) The $(n + 1)$*-dimensional filiform algebra of the first kind* is the Lie algebra with basis $X, Y_1, Y_2, \ldots, Y_n$ and only non-trivial relations

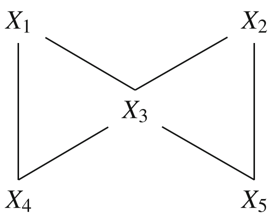
$$[X, Y_j] = Y_{j+1}, \qquad \text{for } j \in \{1, \ldots, n - 1\}.$$

It is an n-step nilpotent Lie algebra and can be realized as a matrix algebra by considering the matrices of the form:

$$\begin{bmatrix} 0 & x & 0 & \cdot & 0 & y_n \\ & \cdot & \ddots & \cdot & \cdot & \vdots \\ & & \cdot & \ddots & 0 & \vdots \\ & & & \cdot & x & y_2 \\ & & & & \cdot & y_1 \\ 0 & & & & & 0 \end{bmatrix}, \qquad \text{for } x, y_1, \ldots, y_n \in \mathbb{R}.$$

Example 9.1.8 (Free-Nilpotent Lie Algebras) The *free-nilpotent Lie algebra of rank n and step k*, denoted by $\mathfrak{f}_{n,k}$, is defined to be the quotient algebra $\mathfrak{f}_n/\mathfrak{f}_n^{(k+1)}$, where $\mathfrak{f}_n$ is the free Lie algebra on n generators. This quotient is a finite-dimensional Lie algebra. The (infinite-dimensional) Lie algebra $\mathfrak{f}_n$ can be explicitly constructed in the following way. One abstractly considers n-many letters $X_1, \ldots, X_n$, called *generators*; then performs abstract Lie brackets of at most k-many of them, in every order, e.g., $[X_1, [X_7, [X_5, X_3]]]$ but also $[[X_5, X_1], [X_7, X_3]]$; then one considers the span, imposing anti-commutativity and Jacobi identity. There is a web app for calculating bases of free Lie algebras available here: https://coropa.sourceforge.io/#cgi

For example, the Lie algebra of rank 2 and step 3 is given by the diagram



,

which has to be read as $[X_1, X_2] = X_3$, $[X_1, X_3] = X_4$, and $[X_3, X_2] = X_5$. Each bracket relation is expressed by a V-shape in the diagram, and it should be read as: the bracket between the *left* arm of the V and the *right* arm gives the vector at the bottom of the V. If the diagram should be read differently, we shall use another notation; see page 329.

Example 9.1.9 (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. We denote this important Lie algebra by

$$\mathfrak{nil}_n := \left\{ \begin{bmatrix} 0 & * & * \\ 0 & \ddots & * \\ 0 & 0 & 0 \end{bmatrix} \right\} \subset \mathfrak{gl}(n).$$

See the later Example 9.3.4 for the associated simply connected Lie group Nil_n. By a result of Birkhoff, we shall see that, up to isomorphism, the subalgebras of $\mathfrak{nil}_n$ are the most general examples of nilpotent Lie algebras; see Corollary 9.1.20. Moreover, by a result of Engel, if $\mathfrak{g}$ is a Lie subalgebra of $\mathfrak{gl}(n)$ made of nilpotent transformations (in the sense of Definition 9.1.12), then, up to a change of basis of $\mathbb{R}^n$, we have that $\mathfrak{g}$ is a subalgebra of $\mathfrak{nil}_n$; see Theorem 9.1.17.

Example 9.1.10 (Generalized Flag-Shifting Lie Algebras, $\mathfrak{g}_{\mathrm{nil}}(\mathcal{F})$) Here is a slight coordinate-free generalization of the previous example. Let V be a finite-dimensional vector space and let $\mathcal{F} = (V_0, \dots, V_m)$ be a flag for V, i.e.,

$$V_0 = \{0\} \subsetneq V_1 \subsetneq \dots \subsetneq V_m = V.$$

The set of flag-preserving transformations, denoted by $\mathfrak{g}(\mathcal{F})$, is not nilpotent; see Exercise 9.5.4. However, the set of *flag-shifting transformations*

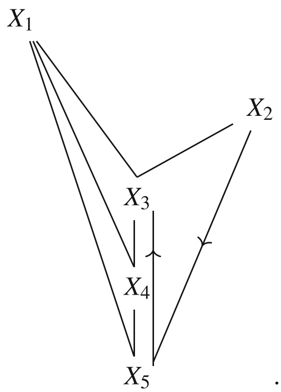
$$\mathfrak{g}_{\mathrm{nil}}(\mathcal{F}) := \left\{A \in \mathfrak{gl}(V) : A(V_k) \subseteq V_{k-1}, \forall k \in \{1, \dots, m\}\right\}$$

is nilpotent of step at most $m - 1$; see Exercise 9.5.5. Actually, the Lie algebra $\mathfrak{g}_{\mathrm{nil}}(\mathcal{F})$ can be seen as a subalgebra of $\mathfrak{nil}_n$, with $n := \dim V$.

Example 9.1.11 The following Lie algebra is denoted as $\mathfrak{n}_{5,1}$; see [LT22, page 162]. The non-trivial brackets in $\mathfrak{n}_{5,1}$ with respect to some basis $X_1, \dots, X_5$ are the following:

$$[X_1, X_2] = X_3\,, \quad [X_1, X_3] = X_4\,, \quad [X_1, X_4] = [X_2, X_3] = X_5\,.$$

This is a nilpotent Lie algebra of rank 2 and step 4. The Lie brackets can be pictured with the diagram:



.

9.1.2 Nilpotent and Unipotent Transformations

Definition 9.1.12 Let V be a vector space, on which we denote by $\mathbb{I}$ the identity map.

9.1.12.i. We say that $A \in \mathfrak{gl}(V)$ is a *nilpotent transformation* of V if there is $d \in \mathbb{N}$ such that $A^d \equiv 0$.
9.1.12.ii. We say that $B \in \mathfrak{gl}(V)$ is a *unipotent transformation* of V if $B - \mathbb{I}$ is nilpotent.

For the next result, recall that if $X \in \mathfrak{gl}(V)$ then the adjoint map for X, from Definition 5.5.1, is the element $\mathrm{ad}_X \in \mathfrak{gl}(\mathfrak{gl}(V))$ such that $\mathrm{ad}_X(Y) := [X, Y] = XY - YX$, for $y \in \mathfrak{gl}(V)$.

Proposition 9.1.13 *Let V be a finite-dimensional vector space. If $X \in \mathfrak{gl}(V)$ is a nilpotent transformation of V, then ad_X is a nilpotent transformation of $\mathfrak{gl}(V)$.*

Proof First, from the definition of ad, one sees by induction that there are constants $\{c_{k,j}\}_{j,k\in\mathbb{N}} \subset \mathbb{Z}$ such that

$$\mathrm{ad}_X^k Y = \sum_{j=0}^{k} c_{j,k} X^j Y X^{k-j}, \qquad \forall X, Y \in \mathfrak{gl}(V), \forall k \in \mathbb{N}. \tag{9.3}$$

Next, let $X \in \mathfrak{gl}(V)$ and $d \in \mathbb{N}$ such that $X^d = 0$. Notice that for all $j \in \{0, \ldots, 2d\}$ we have $j \geq d$ or $2d - j \geq d$. Therefore, for all $Y \in \mathfrak{gl}(V)$, if we write $\mathrm{ad}_X^{2d} Y$ as in (9.3), we see that $X^j Y X^{2d-j} = 0$ for all j, and thus $\mathrm{ad}_X^{2d} Y = 0$. □

The following proposition strengthens the result that nilpotent matrices only have 0 as an eigenvalue.

Proposition 9.1.14 *Let V be a finite-dimensional vector space. If $A \in \mathfrak{gl}(V)$ is a nilpotent transformation, then there is a basis of V in which A is strictly upper triangular.*

Proof Given $A \in \mathfrak{gl}(V)$, there is a basis of V such that the representation of A in this basis is in real Jordan form; see Exercise 9.5.9. Notice that A is nilpotent if and only if each block of the Jordan representation of A is nilpotent. Moreover, every such block is nilpotent if and only if its eigenvalue is zero. Therefore, the real Jordan form of A is upper triangular. □

9.1.3 Engel's Theorem

In this section, we prove Engel's Theorem. We begin by recalling the notion of representation of a Lie algebra, then we prove the statement for linear Lie algebras, and finally, the desired theorem.

Definition 9.1.15 A *representation* of a Lie algebra $\mathfrak{g}$ 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

$$[x, y]v = x(yv) - y(xv), \quad \forall x, y \in \mathfrak{g}, \forall v \in V.$$

The adjoint map $\mathrm{ad} : \mathfrak{g} \to \mathfrak{gl}(\mathfrak{g})$, which is a Lie algebra representation because of Jacobi identity, may not be injective. In fact, its kernel is exactly the center $Z(\mathfrak{g})$ of $\mathfrak{g}$ and thus $\mathrm{ad}(\mathfrak{g}) := \{\mathrm{ad}_x : x \in \mathfrak{g}\} \simeq \mathfrak{g}/Z(\mathfrak{g})$.

Remark 9.1.16 Another example of Lie algebra representation is given by subalgebras: if $\mathfrak{h}$ is a subalgebra of $\mathfrak{g}$, then we have the representation of $\mathfrak{h}$ on $\mathfrak{g}/\mathfrak{h}$

$$\mathrm{ad}_{\mathfrak{g}/\mathfrak{h}} : \mathfrak{h} \to \mathfrak{gl}(\mathfrak{g}/\mathfrak{h}),$$

defined by

$$\mathrm{ad}_{\mathfrak{g}/\mathfrak{h}}(h)(y + \mathfrak{h}) := [h, y] + \mathfrak{h}, \quad \forall h \in \mathfrak{h}, \forall y \in \mathfrak{g}. \tag{9.4}$$

Indeed, note that, if $h \in \mathfrak{h}$, then $[h, \mathfrak{h}] \subseteq \mathfrak{h}$, and so $[h, y+\mathfrak{h}]+\mathfrak{h} = [h, y]+\mathfrak{h}$, leading to a well-defined $\mathrm{ad}_{\mathfrak{g}/\mathfrak{h}}$. Moreover, we stress that if for $x \in \mathfrak{h}$ the transformation $\mathrm{ad}(x) : \mathfrak{g} \to \mathfrak{g}$ is nilpotent, then $\mathrm{ad}_{\mathfrak{g}/\mathfrak{h}}(x) : \mathfrak{g}/\mathfrak{h} \to \mathfrak{g}/\mathfrak{h}$ is nilpotent. Indeed, we have $(\mathrm{ad}_{\mathfrak{g}/\mathfrak{h}}(x))^k(y + \mathfrak{h}) = (\mathrm{ad}(x))^k(y) + \mathfrak{h}$.

9.1.3.1 Engel's Theorem on Linear Lie Algebras

In the next theorem, we see that every Lie algebra of nilpotent transformations is a subalgebra of a Lie algebra of the form $\mathfrak{g}_{\mathrm{nil}}(\mathcal{F})$ as defined in Example 9.1.10.

Theorem 9.1.17 (Engel's Theorem on linear Lie Algebras) *Let $V \neq \{0\}$ be a finite-dimensional vector space and $\mathfrak{g} \subseteq \mathfrak{gl}(V)$ a subalgebra. Assume that every $x \in \mathfrak{g}$ is a nilpotent transformation of V. Then, we have:*

9.1.17.i. There is $v_0 \in V \setminus \{0\}$ such that $\mathfrak{g}(v_0) = \{0\}$;
9.1.17.ii. There is a flag $\mathcal{F} = (V_0, \dots, V_n)$ for V with $\dim(V_k) = k$ and $\mathfrak{g} \subseteq \mathfrak{g}_{\mathrm{nil}}(\mathcal{F})$.

Consequently,

9.1.17.iii. There is a basis of V relative to which elements of $\mathfrak{g}$ are strictly upper-triangular matrices;

9.1.17.iv. The Lie algebra $\mathfrak{g}$ is nilpotent.

Proof We begin by proving 9.1.17.i. It is proved by induction on the dimension of $\mathfrak{g}$.

If $\dim \mathfrak{g} = 0$, then any $v_0 \in V \setminus \{0\}$ works. If $\dim \mathfrak{g} = 1$, then $\mathfrak{g} = \mathbb{R}x$ for some $x \in \mathfrak{gl}(V)$. Since x is nilpotent, it has 0 as its (only) eigenvalue (as shown in Proposition 9.1.14 by Jordan decomposition). Thus there is $v \neq 0$ such that $xv = 0$ and so $txv = 0$, for all $t \in \mathbb{R}$.

Next, assume that $\dim \mathfrak{g} > 1$, and assume the statement true for all Lie algebras of dimension strictly less than $\dim \mathfrak{g}$. Pick a subalgebra $\mathfrak{h} < \mathfrak{g}$ with $\mathfrak{h} \neq \mathfrak{g}$ of maximal dimension. Note that $\dim \mathfrak{h} \geq 1$ because 1-dimensional subalgebras always exist. By assumption, every $x \in \mathfrak{g} \subset \mathfrak{gl}(V)$ is nilpotent. Hence, the transformation $\mathrm{ad}(x) \in \mathfrak{gl}(\mathfrak{gl}(V))$ is nilpotent (see Proposition 9.1.13) and, by Remark 9.1.16, the transformation $\mathrm{ad}_{\mathfrak{g}/\mathfrak{h}}(x) \in \mathfrak{gl}(\mathfrak{g}/\mathfrak{h})$, as defined in (9.4), is nilpotent. Since $\dim(\mathrm{ad}_{\mathfrak{g}/\mathfrak{h}}(\mathfrak{h})) \leq \dim(\mathfrak{h}) < \dim(\mathfrak{g})$, we apply the inductive hypothesis 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, we have $[\mathfrak{h}, x_0] \subset \mathfrak{h}$. 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 consider the representation $\mathfrak{h} \to \mathfrak{gl}(V)$ given by the inclusion. Recall that since $\dim(\mathfrak{h}) < \dim(\mathfrak{g})$, by the inductive hypothesis there is $v \in V \setminus \{0\}$ such that $\mathfrak{h}(v) = \{0\}$ and so we can consider the nontrivial subspace

$$V_0 := \{v \in V : \mathfrak{h}v = \{0\}\}.$$

Note that

$$\mathfrak{g}(V_0) \subset V_0,$$

because 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 that $[x, y] \in [\mathfrak{g}, \mathfrak{h}] \subset \mathfrak{h}$ and that $\mathfrak{h}v = \{0\}$. Therefore, we can apply the base case of induction in dimension 1 to the representation $\mathbb{R}x_0 \to \mathfrak{gl}(V_0)$ defined as $tx_0 \mapsto (tx_0)_{|V_0}$: There is $v_0 \in V_0 \setminus \{0\}$ such that $x_0 v_0 = 0$ and thus, putting all together, we have that

$$\mathfrak{g}(v_0) = \mathfrak{h}v_0 + \mathbb{R}x_0 v_0 = \{0\}.$$

This proves 9.1.17.i.

Next, we show 9.1.17.ii. By 9.1.17.i, we choose $v_1 \in V \setminus \{0\}$ such that $\mathfrak{g}(v_1) = 0$ and we define $V_1 := \mathbb{R}v_1$. Then, the map $\alpha : \mathfrak{g} \to \mathfrak{gl}(V/V_1)$ defined by $x \in \mathfrak{g} \mapsto \alpha(x)(v + V_1) := x(v) + V_1$ is well defined and gives a representation of $\mathfrak{g}$ on V/V_1. Still, the set $\alpha(\mathfrak{g})$ consists of nilpotent transformations.

By induction on the dimension of V, it follows that V/V_1 possesses a complete flag $\mathcal{F}_1 := (W_0, \ldots, W_{n-1})$ with $\alpha(\mathfrak{g}) \subseteq \mathfrak{g}_{\mathrm{nil}}(\mathcal{F}_1)$. Then, the singleton $\{0\}$ together with the preimage of the flag $\mathcal{F}_1$ in V yields a complete flag $\mathcal{F}$ in V with $\mathfrak{g} \subseteq \mathfrak{g}_{\mathrm{nil}}(\mathcal{F})$. The consequences at the end of the statement of the theorem are immediate. □

9.1.3.2 Engel's Characterization Theorem for Nilpotent Lie Algebras

Theorem 9.1.18 (Engel's Theorem) *Let $\mathfrak{g}$ be a finite-dimensional Lie algebra. Then*

$$\mathfrak{g} \text{ is nilpotent} \quad \Longleftrightarrow \quad \mathrm{ad}_x \text{ is nilpotent, for every } x \in \mathfrak{g}.$$

Proof [$\Rightarrow$] The forward direction has already been proved in Proposition 9.1.3.iv.

[$\Leftarrow$] Regarding the apposite direction, considering $\mathrm{ad}(\mathfrak{g}) \subseteq \mathfrak{gl}(\mathfrak{g})$, we have that $\mathrm{ad}(\mathfrak{g}) \cong \mathfrak{g}/Z(\mathfrak{g})$. Moreover, the set $\mathrm{ad}(\mathfrak{g})$ is a Lie algebra of nilpotent transformations of $\mathfrak{g}$, by assumption. By Engel's Theorem on linear Lie algebras, Theorem 9.1.17, we have that $\mathrm{ad}(\mathfrak{g})$ is nilpotent. Finally, by Proposition 9.1.3.ii, we conclude that $\mathfrak{g}$ is nilpotent, as desired. □

In the above result, when using Proposition 9.1.3.ii, it is important that we are quotienting by a subspace that is central. The reader should be aware of the following fact: if $\mathfrak{a} \subseteq \mathfrak{g}$ is an ideal and both $\mathfrak{a}$ and $\mathfrak{g}/\mathfrak{a}$ are nilpotent, then $\mathfrak{g}$ may not be nilpotent; see Exercise 9.5.10.

9.1.4 The General Birkhoff-Embedding Theorem

Theorem 9.1.19 (Birkhoff-Embedding Theorem) *Let $\mathfrak{g}$ be a nilpotent finite-dimensional Lie algebra. Then there are a finite-dimensional vector space V and an injective homomorphism $\iota : \mathfrak{g} \to \mathfrak{gl}(V)$ such that for every $x \in \mathfrak{g}$ the transformation $\iota(x)$ is nilpotent.*

We will only prove this theorem in the particular case of positively graded Lie algebras, and, hence, for Carnot algebras; see Sect. 9.2.4. A proof of the general theorem can be found in [CG90, Theorem 1.1.11], but it relies on the construction of the universal enveloping algebra and the Poincaré-Birkhoff-Witt Theorem; see [Kna02, Chapter III]. There is a more general result due to Ado, which generalizes the previous Theorem 9.1.19, stating that every finite-dimensional Lie algebra has an injective finite-dimensional representation whose restriction to the maximal nilpotent ideal is nilpotent valued; see [Kna02, Theorem B.8, p.663].

We recall that we also have Engel's characterization of Lie algebras of nilpotent transformations: Theorem 9.1.17. Hence, Birkhoff's Theorem can be stated as follows:

Corollary 9.1.20 *For every nilpotent finite-dimensional Lie algebra $\mathfrak{g}$ there are $n \in \mathbb{N}$ and a subalgebra $\tilde{\mathfrak{g}}$ of the space $\mathfrak{nil}_n$ of strictly upper-triangular $n \times n$-matrices such that $\mathfrak{g}$ and $\tilde{\mathfrak{g}}$ are isomorphic.*

9.2 Gradings and Stratifications

In Chap. 11, we will see that Carnot groups, the tangents of Carnot-Carathéodory spaces, will have nilpotent 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

$$V_s \neq \{0\} \qquad \text{and} \qquad [V_1, V_j] = V_{j+1}, \quad \forall j \in \{1, \ldots, s\}, \tag{9.5}$$

where we set $V_{s+1} := \{0\}$. In Definition 9.2.4, we shall provide other alternative equivalent definitions. In fact, it is useful to see stratifications as special types of gradings. Hence, we discuss this broader concept next.

9.2.1 Graded Vector Spaces and Graded Lie Algebras

We begin with graded vector spaces.

Definition 9.2.1 (Grading for a Vector Space) Let A be a set (or, in many situations, a subset of an abelian group such as $\mathbb{Z}$ or $\mathbb{R}$) and let V be a vector space. A *linear grading* of V over A is a collection of vector subspaces $(V_a)_{a\in A}$ of V such that

$$V = \bigoplus_{a\in A} V_a.$$

This means that $V = \operatorname{span}\{V_a : a \in A\}$ and for every $a, a' \in A$ with $a \neq a'$ we have that $V_a \cap V_{a'} = \{0\}$. We refer to a linear grading of V also as a *grading of V as a vector space*, or as a *grading*, for short. 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 $A \subseteq \mathbb{R}$ and

$$V = V_{>0} := \bigoplus_{a>0} V_a,$$

then V is said to be *positively graded*. Given a grading $(V_a)_{a\in A}$ and $a \in A$, elements in V_a are said to have *degree* a. Each V_a is called a *layer*, or the *a-th layer*.

Next, we examine Lie algebras, for which we consider two more restrictive notions of linear gradings. We stress that a Lie algebra $\mathfrak{g}$ is a vector space with the additional structure of a Lie bracket.

Definition 9.2.2 (Compatible Linear Grading) Let $\mathfrak{g}$ be a Lie algebra. A *compatible linear grading* on $\mathfrak{g}$ is a linear decomposition of $\mathfrak{g}$ into vector subspaces $V_1, V_2, \ldots$ such that

$$\mathfrak{g} = \bigoplus_{i=1}^{\infty} V_i \quad \text{and} \quad \mathfrak{g}^{(i)} = \mathfrak{g}^{(i+1)} \oplus V_i, \quad \forall i \in \mathbb{N}, \tag{9.6}$$

where the $\mathfrak{g}^{(i)}$'s are the lower central series elements from Definition 9.1.1.

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 (see Exercise 9.5.32). Carnot algebras will have a stronger property.

Definition 9.2.3 (Lie Algebra Grading) Given an abelian group A (for instance $\mathbb{Z}$ or $\mathbb{R}$) and a Lie algebra $\mathfrak{g}$, a *grading over* A of $\mathfrak{g}$ *as a Lie algebra*, or a *Lie algebra A-grading*, is a linear grading $(V_a)_{a\in A}$ of $\mathfrak{g}$ as a vector space with the extra property that

$$[V_a, V_b] \subseteq V_{a+b}, \quad \forall a, b \in A.$$

9.2.2 Stratified Lie Algebras

We shall focus on a particular type of Lie algebra grading: stratifications. There are various equivalent definitions for them; see Remark 9.2.5.

Definition 9.2.4 (Stratification) If $\mathfrak{g}$ is a Lie algebra, a Lie algebra $\mathbb{Z}$-grading $(V_a)_{a\in\mathbb{Z}}$ of $\mathfrak{g}$ 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 a *Carnot algebra*.

Remark 9.2.5 (Equivalent Definitions) We rephrase the definition by 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

$$[V_1, V_j] = V_{j+1}, \quad \forall j \in \{1, \ldots, s\}, \text{ with } V_{s+1} := \{0\}.$$

The latter is the most common version of the definition of stratification.

Example 9.2.6 Every commutative Lie algebra $\mathfrak{g}$ admits a 1-step stratification with $V_1 = \mathfrak{g}$.

Example 9.2.7 Let $\mathfrak{g}$ be the Heisenberg Lie algebra spanned by X, Y, Z with relation $[X, Y] = Z$. Then the subspaces $V_1 := \text{span}\{X, Y\}$ and $V_2 := \text{span}\{Z\}$ form a 2-step stratification.

Remark 9.2.8 The following **non invertible** implications hold for finite-dimensional Lie algebras:

$$\text{Carnot} \quad \overset{\not\Leftarrow}{\Rightarrow} \quad \text{positively graded} \quad \overset{\not\Leftarrow}{\Rightarrow} \quad \text{nilpotent} \, .$$

For the forward implications, see Exercises 9.5.21 and 9.5.19. For the fact that reverse implications do not hold, see Exercises 9.5.24 and 9.5.26. In fact, more examples can be found in [Goo76, Hak+22, LT22].

9.2.2.1 Uniqueness of Stratifications

In the following proposition, we prove that every two stratifications on the same stratifiable Lie algebra differ by an automorphism.

Proposition 9.2.9 (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, we have the following properties:

9.2.9.i. *The steps coincide, i.e., $s = t$, and $\mathfrak{g}^{(k)} = V_k \oplus \cdots \oplus V_s = W_k \oplus \cdots \oplus W_s$, for all $k \in \{1, \ldots, s\}$;*

9.2.9.ii. *there is a Lie algebra automorphism $A : \mathfrak{g} \to \mathfrak{g}$ with $A(V_k) = W_k$, for all $k \in \{1, \ldots, s\}$.*

Proof The first point is simple; see Exercise 9.5.22. Then the quotient mappings $\pi_k : \mathfrak{g}^{(k)} \to \mathfrak{g}^{(k)}/\mathfrak{g}^{(k+1)}$ induce 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. Next, we need 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

$$A([a, b]) = \sum_{i=1}^{s}\sum_{j=1}^{s} A([a_i, b_j]) \quad \text{and} \quad [Aa, Ab] = \sum_{i=1}^{s}\sum_{j=1}^{s}[Aa_i, Ab_j].$$

Thus, we can just prove $A([a_i, b_j]) = [Aa_i, Ab_j]$ for $a_i \in V_i$ and $b_j \in V_j$. Notice that $A([a_i, b_j])$ and $[Aa_i, Ab_j]$ both belong to W_{i+j}. Therefore, writing $[a_i, b_j]$ as $[Aa_i, Ab_j] + ([a_i, b_j] - [Aa_i, Ab_j])$, we have that $A([a_i, b_j]) = [Aa_i, Ab_j]$ if and only if $[a_i, b_j] - [Aa_i, Ab_j] \in \mathfrak{g}^{(i+j+1)}$. In fact, we have

$$[a_i, b_j] - [Aa_i, Ab_j] = [a_i - Aa_i, b_j] - [Aa_i, Ab_j - b_j] \in \mathfrak{g}^{(i+j+1)},$$

because, on the 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 $[Aa_i, Ab_j - b_j] \in \mathfrak{g}^{(i+j+1)}$. We deduced that A is a Lie algebra homomorphism. □

9.2.2.2 Induced Grading on $\mathfrak{gl}(V)$

In the next definition, we consider subspaces of the linear transformations on a vector space equipped with a linear grading that will form a grading themselves.

Definition 9.2.10 ($\mathfrak{gl}(V)_a$ **and** $\mathfrak{gl}(V)_{>0}$) Let $(V_a)_{a\in A}$ be a linear grading of a vector space V over an abelian group A. For every $a \in A$ we define

$$\mathfrak{gl}(V)_a := \{M \in \mathfrak{gl}(V) \,:\, M(V_b) \subseteq V_{b+a}, \forall b \in A\},$$

and if A has an ordering (for instance, if $A \subseteq \mathbb{R}$) we define

$$\mathfrak{gl}(V)_{>0} := \bigoplus_{a>0} \mathfrak{gl}(V)_a.$$

We shall show that the collection $(\mathfrak{gl}(V)_a)_{a\in A}$ forms an A-grading of $\mathfrak{gl}(V)$ as a Lie algebra. Moreover, if V is a Carnot algebra, then $\mathfrak{gl}(V)_{>0}$ is a Carnot algebra.

Proposition 9.2.11 *Let V be a finite-dimensional vector space with a linear grading $(V_a)_{a\in A}$ over an abelian group A. Then, we have the following properties:*

9.2.11.i. $(\mathfrak{gl}(V)_a)_{a\in A}$ *is a Lie algebra grading of* $\mathfrak{gl}(V)$*;*
9.2.11.ii. *If A is a subgroup of $(\mathbb{R}, +)$, then $\mathfrak{gl}(V)_{>0}$ is a Lie subalgebra of nilpotent transformations;*

9.2.11.iii. If $A = (\mathbb{Z}, +)$ and there are $\bar{a}, \bar{b} \in A$ such that

$$V_m \neq \{0\} \quad \Leftrightarrow \quad m \in \mathbb{Z} \cap [\bar{a}, \bar{b}],$$

then $\mathfrak{gl}(V)_{>0}$ is a Carnot algebra.

Proof

(i) Fix $X_1, \ldots, X_n$ a basis of V adapted to the direct-sum decomposition $V = \bigoplus_{a \in A} V_a$. So for every $i \in \{1, \ldots, n\}$ there is $a_i \in A$ such that $X_i \in V_{a_i}$. For each $i, j \in \{1, \ldots, n\}$, let $E^i_j \in \mathfrak{gl}(V)$ be such that

$$E^i_j(X_i) := X_j \qquad \text{and} \qquad E^i_j(X_k) := 0, \quad \forall k \neq i.$$

Consequently, the elements $(E^i_j)_{i,j \in \{1,\ldots,n\}}$ 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 $a \neq a_i$ 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,

$$\bigoplus_{a \in A} \mathfrak{gl}(V)_a = \mathfrak{gl}(V).$$

Finally, notice that if $M_1 \in \mathfrak{gl}(V)_a$ and $M_2 \in \mathfrak{gl}(V)_b$ 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 completes the proof of 9.2.11.i.

(ii) Clearly, the subset $\mathfrak{gl}(V)_{>0}$ is a Lie subalgebra, see Exercise 9.5.29. Moreover, assuming $a_i \leq a_j$ for every $i \leq j$ and defining

$$W_0 := \{0\} \qquad \text{and} \qquad W_i := \operatorname{span}\{X_{n-i+1}, \ldots, X_n\}, \quad \forall i \in \{1, \ldots, n\},$$

the m-uple $\mathcal{F} := (W_0, \ldots, W_m)$ defines a flag for V such that $\mathfrak{gl}(V)_{>0} \subseteq \mathfrak{g}_{\mathrm{nil}}(\mathcal{F})$; see the last notation from Example 9.1.10.

(iii) We want to prove that if $V = V_{\bar{a}} \oplus \cdots \oplus V_{\bar{b}}$ with $V_j \neq \{0\}$ for every integer $j \in \{\bar{a}, \ldots, \bar{b}\}$, then $\mathfrak{gl}(V)_{>0}$ is generated by $\mathfrak{gl}(V)_1$. By induction, we shall prove that

$$[\mathfrak{gl}(V)_1, \mathfrak{gl}(V)_k] = \mathfrak{gl}(V)_{k+1}, \qquad \forall k \in \mathbb{N}. \tag{9.7}$$

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 $\bar{a} \leq a_i < a_j \leq \bar{b}$. Moreover, by assumption, there is a basis element X_ℓ with $a_\ell = a_i + 1$. 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, recalling that $a_i < a_\ell \leq a_j$, we have $E^i_\ell E^\ell_j = 0$ and $E^\ell_j E^i_\ell = E^i_j$. Since $E^i_\ell \in \mathfrak{gl}(V)_1$ and $E^\ell_j \in \mathfrak{gl}(V)_k$, we proved (9.7), as desired.

□

9.2.3 Dilation Structures

For vector spaces, gradings are in correspondence with dilation structures.

We stress that, in the presence of an $\mathbb{R}$-grading $(V_a)_{a\in\mathbb{R}}$ of a finite-dimensional vector space V, there is a finite subset $I \subset \mathbb{R}$ such that $V_i \neq \{0\}$ if and only if $i \in I$. Hence, every vector $v \in V$ can be written uniquely as $v = \sum_{i\in I} v_i$, with $v_i \in V_i$. With abuse of notation, we shall still write $v = \sum_{i\in\mathbb{R}} v_i$.

Definition 9.2.12 (Dilations on Graded Vector Spaces) Let $(V_a)_{a\in\mathbb{R}}$ be an $\mathbb{R}$-grading of a finite-dimensional vector space V. For every $\lambda > 0$, the *anisotropic dilation on V of factor λ relative to the grading* (or, simply, the *dilation* δ_λ, when the grading is understood) is the linear map $\delta_\lambda : V \to V$ such that

$$\delta_\lambda v = \lambda^a v, \qquad \forall a \in \mathbb{R}, \forall v \in V_a. \tag{9.8}$$

In addition, we notice that if the grading is a $\mathbb{Z}$-grading, then the above equation defines dilations δ_λ, also for $\lambda < 0$. If $V_{\leq 0}$ is trivial, then for $\lambda = 0$ we set $\delta_\lambda \equiv 0$. Therefore, if $V_{\leq 0} = \{0\}$ and the grading is a $\mathbb{Z}$-grading, the map $(\lambda, v) \in \mathbb{R} \times V \mapsto \delta_\lambda(v) \in V$ is continuous.

Proposition 9.2.13 *Given a Lie algebra grading $(V_a)_{a\in\mathbb{R}}$ of a Lie algebra $\mathfrak{g}$, consider the anisotropic dilations relative to the grading, as in* (9.8)*. Then, for $\lambda \in \mathbb{R}_{>0}$, the dilation $\delta_\lambda : \mathfrak{g} \to \mathfrak{g}$ is a Lie algebra automorphism. Moreover, the map $(\mathbb{R}_{>0}, \cdot) \to \mathrm{Aut}_{\mathrm{Lie}}(\mathfrak{g})$, $\lambda \mapsto \delta_\lambda$, is a one-parameter subgroup:*

$$\delta_\lambda \circ \delta_\mu = \delta_{\lambda\mu}, \qquad \forall \lambda, \mu \in \mathbb{R}. \tag{9.9}$$

Proof We need to show that the map is a linear bijection and

$$\delta_\lambda([X, Y]) = [\delta_\lambda X, \delta_\lambda Y], \qquad \forall X, Y \in \mathfrak{g}.$$

Take $X, Y \in \mathfrak{g}$ and decompose them as $X = \sum_{i\in\mathbb{R}} X_i$ and $Y = \sum_{i\in\mathbb{R}} Y_i$, 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, the map δ_λ is invertible with inverse $\delta_{1/\lambda}$. Equation (9.9) is trivial. □

In the presence of a $\mathbb{Z}$-grading, the map $(\mathbb{R} \setminus \{0\}, \cdot) \to \mathrm{Aut}_{\mathrm{Lie}}(\mathfrak{g})$, $\lambda \mapsto \delta_\lambda$ is a group homomorphism, also defined for negative λ's.

Vice versa, if we have a direct-sum decomposition $\mathfrak{g} = \bigoplus_{a\in\mathbb{R}} V_a$ and the map δ_λ as in (9.8) is a Lie algebra automorphism, then the decomposition is necessarily a Lie algebra grading. Notice that (9.8) can be rewritten as

$$\delta_\lambda = \exp(\log(\lambda)\alpha),$$

where α is the diagonal transformation multiplying by $a \in \mathbb{R}$ the space V_a. The assumption that the δ_λ's are Lie algebra automorphisms rephrases as α being a derivation, in the sense of Definition 5.6.2. More generally, every OPS of automorphisms gives a grading.

Proposition 9.2.14 *Given a Lie algebra $\mathfrak{g}$, let $(\mathbb{R}_{>0}, \cdot) \to \mathrm{Aut}_{\mathrm{Lie}}(\mathfrak{g})$, $\lambda \mapsto \delta_\lambda$, be a one-parameter subgroup of Lie algebra automorphisms. Then there exists $\alpha \in \mathrm{Der}(\mathfrak{g})$ such that $\delta_\lambda = \exp((\log(\lambda)\alpha)$, for all $\lambda \in \mathbb{R}_{>0}$, and the Lie algebra $\mathfrak{g}$ admits a Lie algebra grading where the layers are*

$$V_t := \mathfrak{g} \cap \bigoplus_{s\in\mathbb{R}} E^\alpha_{t+is}, \qquad \textit{for } t \in \mathbb{R},$$

where E^α_{t+is} is a generalized eigenspace of α corresponding to the eigenvector $t+is$; see (9.11).

We leave the proof as an exercise since we will actually prove a stronger statement in a few pages; see Proposition 9.2.19.

Later, when discussing measures on Carnot groups, we shall need the following result.

Lemma 9.2.15 *If $\mathfrak{g} = V_1 \oplus \cdots \oplus V_s$ is a stratified Lie algebra, then for every $\lambda > 0$ the anisotropic dilation $\delta_\lambda : \mathfrak{g} \to \mathfrak{g}$ of factor λ relative to the grading has determinant equal to λ^Q with*

$$Q := \sum_{j=1}^{s} j \cdot \dim(V_j).$$

Proof Fix a basis $X_1, \dots, X_n$ adapted to the stratification, i.e., for every i, there is a j such that $X_i \in V_j$. Then, in this basis, the map δ_λ is represented by the diagonal matrix with diagonal

$$\big(\underbrace{\lambda, \dots, \lambda}_{\dim V_1}, \underbrace{\lambda^2, \dots, \lambda^2}_{\dim V_2}, \dots, \underbrace{\lambda^s, \dots, \lambda^s}_{\dim V_s}\big).$$

Hence, the determinant is $\lambda^{\dim V^1} \cdot (\lambda^2)^{\dim V_2} \cdot \dots \cdot (\lambda^s)^{\dim V_s} = \lambda^Q$. □

9.2.3.1 Associated Carnot Algebra

To every nilpotent Lie algebra $\mathfrak{g}$, there is a canonical way to associate a stratified Lie algebra. Later in the book, in Sect. 12.4.1, we shall see that this associated Lie algebra also has a geometric meaning. We first define this associated Carnot algebra as an abstract direct sum of quotients. Later, we shall equivalently describe it via limiting dilations.

Definition 9.2.16 (Associated Carnot Algebra) Let $\mathfrak{g}$ be a Lie algebra that is nilpotent of step s. Let $\mathfrak{g}^{(i+1)} \stackrel{\text{def}}{=} [\mathfrak{g}, \mathfrak{g}^{(i)}]$ 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

$$\mathfrak{g}_\infty := \bigoplus_{i=1}^{s} \mathfrak{g}^{(i)} \big/ \mathfrak{g}^{(i+1)},$$

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 $\mathfrak{g}^{(i+j+1)}$, as

$$\left[\!\!\left[X + \mathfrak{g}^{(i+1)}, Y + \mathfrak{g}^{(j+1)}\right]\!\!\right]_\infty := [X, Y] + \mathfrak{g}^{(i+j+1)}.$$

The associated Carnot algebra is stratified by $\left(\mathfrak{g}^{(i)}/\mathfrak{g}^{(i+1)}\right)_{i=1}^{s}$. Fixing a compatible linear grading, one can setwise identify this new Lie algebra with the original one, and its Lie bracket can be equivalently defined by the following result.

Lemma 9.2.17 *Let* $(\mathfrak{g}, [\cdot, \cdot])$ *be a nilpotent Lie algebra. Consider the dilations* $(\delta_\lambda)_{\lambda>0}$ *relative to some compatible linear grading and define the map*

$$[\![X, Y]\!]_\infty := \lim_{\lambda \to +\infty} \delta_\lambda^{-1}[\delta_\lambda X, \delta_\lambda Y], \qquad \forall X, Y \in \mathfrak{g}. \tag{9.10}$$

Then $[\![\cdot, \cdot]\!]_\infty$ *defines a Lie bracket on* $\mathfrak{g}$, *such that*

$$[\![\delta_\lambda X, \delta_\lambda Y]\!]_\infty = \delta_\lambda [\![X, Y]\!]_\infty, \qquad \forall X, Y \in \mathfrak{g},\ \forall \lambda \in \mathbb{R},$$

and $(\mathfrak{g}, [\![\cdot, \cdot]\!]_\infty)$ *is isomorphic to the associated Carnot algebra of* $(\mathfrak{g}, [\cdot, \cdot])$.

Proof Let $(V_j)_{j=1}^\infty$ be a compatible linear grading of $(\mathfrak{g}, [\cdot, \cdot])$. Since the V_j's form a direct decomposition of $\mathfrak{g}$, it suffices to consider (9.10) for $X \in V_i$ and $Y \in V_j$, for some i, j. Because we are in the presence of a compatible linear grading, the element $[X, Y]$ belongs to $\bigoplus_{k=i+j}^\infty V_k$; see Exercise 9.5.32. Thus, we have that

$$[X, Y] = Z_{i+j} + Z_{i+j+1} + \ldots + Z_s,$$

for some vectors $Z_k \in V_k$. Hence, we compute

$$\begin{aligned}\delta_\lambda^{-1}[\delta_\lambda X, \delta_\lambda Y] &= \delta_\lambda^{-1}[\lambda^i X, \lambda^j Y]\\ &= \lambda^{i+j}\delta_\lambda^{-1}(Z_{i+j} + Z_{i+j+1} + \ldots + Z_s)\\ &= Z_{i+j} + \lambda^{-1} Z_{i+j+1} + \ldots + \lambda^{i+j-s} Z_s,\end{aligned}$$

which goes to Z_{i+j}, as $\lambda \to \infty$. The proof is concluded by observing that Z_{i+j} is a vector that represent $[X, Y]$ modulo $\mathfrak{g}^{(i+j+1)}$. □

9.2.3.2 Siebert Theorem

Lie groups whose Lie algebra admits a positive grading are precisely those Lie groups whose universal covering Lie group admits automorphisms that are topologically contractive, in a sense that we soon review. Such a characterization is due to Siebert, [Sie86]. We present here his result only focusing on Lie groups. More generally, Siebert showed that a connected locally compact group G admits a contractible automorphism if and only if G is a simply connected Lie group whose Lie algebra admits a positive grading, [Sie86, Corollary 2.4]. In addition, a locally compact group G admits a contractible automorphism if and only if G is topologically isomorphic to the direct product of two groups G_C and G_D admitting contractible automorphisms, where G_C is connected and G_D is totally disconnected, [Sie86, Proposition 4.2 and Corollary 4.3]. We shall not discuss these more comprehensive results here.

We use the term contractive in the following topological sense. A map $F : X \to X$ from a topological space into itself is said to be *contractive* if there exists $x_0 \in X$ such that for every $x \in X$ one has $\lim_{n\to\infty} F^n(x) = x_0$, where F^n denotes the n-times composition of the map F. We stress that if F is a Lie group automorphism $G \to G$ or a Lie algebra automorphism $\mathfrak{g} \to \mathfrak{g}$, then the (only possible choice for) x_0 is 1_G or $0 \in \mathfrak{g}$, respectively.

We saw that if a Lie algebra is graded, then the relative dilations from Definition 9.2.12 define maps δ_λ that are Lie algebra automorphisms, for all $\lambda > 0$; see Proposition 9.2.13. We stress that if $\lambda \in (0, 1)$ and the grading is a positive grading, then the map δ_λ is contractive; see Exercise 9.5.34. Siebert's theorem exactly states the inverse implication.

Theorem 9.2.18 (Siebert, [Sie86]) *For every simply connected Lie group G, the following are equivalent.*

9.2.18.i. $\mathrm{Lie}(G)$ *admits a positive grading;*
9.2.18.ii. G *admits a contractive Lie group automorphism;*
9.2.18.iii. $\mathrm{Lie}(G)$ *admits a contractive Lie algebra automorphism.*

Before the proof of Theorem 9.2.18, we discuss how automorphisms of Lie algebras induce gradings; see the following proposition. For defining the gradings, it is more convenient to pass to complexifications as we next review.

We denote by $V_{\mathbb{C}}$ the *complexification* of a real vector space V, so $V_{\mathbb{C}} := V \times V$ with complex scalar multiplication given by $i \cdot (X, Y) := (-Y, X)$, for $X, Y \in V$. It is a complex vector space with conjugation $(X, Y)^* := (X, -Y)$. If $\phi : V \to V$ is an $\mathbb{R}$-linear map, then its *complexification* is the $\mathbb{C}$-linear map $\phi_{\mathbb{C}} : V_{\mathbb{C}} \to V_{\mathbb{C}}$, $(X, Y) \mapsto \phi_{\mathbb{C}}(X, Y) := (\phi(X), \phi(Y))$. The *spectrum* of ϕ is defined by

$$\mathrm{Spec}(\phi) := \{\alpha \in \mathbb{C} : \det(\phi_{\mathbb{C}} - \alpha\mathbb{I}) = 0\},$$

where $\mathbb{I}$ is the identity map on $V_{\mathbb{C}}$, and the *generalized eigenspace* of ϕ corresponding to $\alpha \in \mathbb{C}$ by

$$E_\alpha^\phi := \left\{v \in V_{\mathbb{C}} : \exists n \in \mathbb{N} \quad (\phi_{\mathbb{C}} - \alpha\mathbb{I})^n v = 0\right\}. \tag{9.11}$$

By Jordan Theorem (see Exercise 9.5.9), we have $V_{\mathbb{C}} = \bigoplus_{\alpha\in\mathrm{Spec}(\phi)} E_\alpha^\phi$.

Proposition 9.2.19 *Let ϕ be an automorphism of a Lie algebra $\mathfrak{g}$. Fix $\lambda \in (0, +\infty) \setminus \{1\}$. For each $t \in \mathbb{R}$, define*

$$V_t := V_t(\lambda, \phi) := \mathfrak{g} \cap \left(\bigoplus \left\{E_\alpha^\phi : \alpha \in \mathbb{C}, |\alpha| = \lambda^t\right\}\right),$$

where the E_α^ϕ are defined in (9.11) for $V := \mathfrak{g}$. Then $\{V_t\}_{t\in\mathbb{R}}$ is a Lie algebra grading of $\mathfrak{g}$, with degrees in $\log_\lambda(|\mathrm{Spec}(\phi)|)$, i.e.,

$$\mathfrak{g} = \bigoplus_{t\in D} V_t, \qquad \textit{for } D := \left\{\log_\lambda(|\alpha|) : \alpha \in \mathrm{Spec}(\phi)\right\}. \tag{9.12}$$

Moreover,

$$|\det(\phi)| = \lambda^{\sum_{t\in\mathbb{R}} t\cdot\dim(V_t)}.$$

Proof We notice that if $\alpha \notin \mathrm{Spec}(\phi) =: \sigma(\phi)$, then $E_\alpha^\phi = \{0\}$. Since $\sigma(\phi)$ is a finite set, then only finitely many V_t's are not trivial. For $\alpha \in \sigma(\phi) \subset \mathbb{C}$, define $U_\alpha^\phi := (E_\alpha^\phi + E_{\bar\alpha}^\phi) \cap \mathfrak{g}$. We claim that

$$\mathfrak{g} = \bigoplus_{\alpha\in\sigma(\phi)} U_\alpha^\phi, \tag{9.13}$$

where the sum is direct up to the identification $U^\phi_\alpha = U^\phi_{\bar\alpha}$. Indeed, take $v \in \mathfrak{g}$ and write $v = \sum_\alpha v_\alpha$ with $v_\alpha \in E^\phi_\alpha$ for all α, which is possible by Jordan theorem. Since $v = v^*$, we have $v = \frac{1}{2}(v + v^*) = \frac{1}{2}\sum_\alpha (v_\alpha + v^*_\alpha)$, where $v_\alpha + v^*_\alpha \in U_\alpha$. So, we have $\mathfrak{g} = \sum_{\alpha\in\sigma(\phi)} U^\phi_\alpha$. Since $U^\phi_\alpha \cap U^\phi_\beta = \{0\}$ if $\alpha \notin \{\beta, \bar\beta\}$, the sum is direct. This proves claim (9.13).

Since ϕ is injective, then $U^\phi_0 = \{0\}$. Therefore, because of (9.13), we have that $\mathfrak{g} = \bigoplus_{t\in\mathbb{R}} V_t$.

Using Exercise 9.5.38, we have

$$[U^\phi_\alpha, U^\phi_\beta] \subset U^\phi_{\alpha\beta} \oplus U^\phi_{\bar\alpha\beta}, \qquad \forall \alpha, \beta \in \mathbb{C}. \tag{9.14}$$

If $X \in U^\phi_\alpha$ and $Y \in U^\phi_\beta$ with $|\alpha| = \lambda^t$ and $|\beta| = \lambda^s$, then $[X, Y] \in U^\phi_{\alpha\beta} \oplus U^\phi_{\bar\alpha\beta} \subset V_{t+s}$, because of (9.14) and $|\alpha\beta| = |\bar\alpha\beta| = \lambda^{s+t}$. Therefore, $[V_s, V_t] \subset V_{s+t}$ and $\{V_t\}_{t\in\mathbb{R}}$ is a real grading of $\mathfrak{g}$. Finally, if we set $\varepsilon_\alpha = 1$ if $\alpha \in \mathbb{R}$ and $\varepsilon_\alpha = 1/2$ if $\alpha \in \mathbb{C} \setminus \mathbb{R}$,

$$|\det(\phi)| = \left| \prod_{\alpha\in\sigma(\phi)} \alpha^{\dim_{\mathbb{C}}(E_\alpha)} \right| = \prod_{\alpha\in\sigma(\phi)} |\alpha|^{\varepsilon_\alpha \dim_{\mathbb{R}}(U_\alpha)} = \prod_{t\in\mathbb{R}} \lambda^{t\cdot\dim(V_t)}.$$

□

Proof of Theorem 9.2.18 The equivalence between 9.2.18.ii and 9.2.18.iii is trivial; see Exercise 9.5.36. The easy implication 9.2.18.i $\implies$ 9.2.18.iii is in Exercise 9.5.34.

Regarding 9.2.18.iii $\implies$ 9.2.18.i, if ϕ is an automorphism of $\mathfrak{g}$ that is contractive, then also $\phi_{\mathbb{C}} : \mathfrak{g}_{\mathbb{C}} \to \mathfrak{g}_{\mathbb{C}}$ is contractive. Therefore, the transformation $\phi_{\mathbb{C}}$ cannot have eigenvalues with norm ≥ 1. For $\lambda := 1/2$, we consider the grading $V_t := V_t(\lambda, \phi)$ of Proposition 9.2.19. We check that it is a positive grading. Indeed, for all $t \leq 0$ we have that for all $\alpha \in \mathbb{C}$ such that $|\alpha| = 1/2^t$ the set E^ϕ_α is trivial because $1/2^t \geq 1$. Thus $V_t = \{0\}$. □

We conclude the subsection with the following observation that tells us when an automorphism coming from a derivation is contractive.

Remark 9.2.20 Let A be a derivation of a Lie algebra $\mathfrak{g}$. For $\psi := \exp(A) \in \mathrm{Aut}_{\mathrm{Lie}}(\mathfrak{g})$ the following are equivalent:

9.2.20.i. ψ is contractive;
9.2.20.ii. $\mathrm{Spec}(\psi) \subseteq \{z \in \mathbb{C} : |z| < 1\}$;
9.2.20.iii. $\mathrm{Spec}(A) \subseteq \{z \in \mathbb{C} : \Re\mathfrak{e}(z) < 0\}$.

9.2.4 Birkhoff Theorem for Stratified Lie Algebras

We will present a proof of Birkhoff's Theorem for Carnot algebras, as Y. Cornulier explained it to the author. 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, containing a copy of $\mathfrak{g}$.

9.2.4.1 Induced Grading on $\mathfrak{g} \rtimes \mathbb{R}$

We begin very general: Let $\mathfrak{g}$ be a $\mathbb{Z}$-graded Lie algebra with grading $(V_m)_{m\in\mathbb{Z}}$. We consider the semidirect product $\mathfrak{g} \rtimes \mathbb{R}$ where $1 \in \mathbb{R}$ acts on $\mathfrak{g}$ as the derivation that multiplies by m the vectors in V_m. Namely, recalling the construction from (5.17), the Lie bracket on the semidirect product $\mathfrak{g} \rtimes \mathbb{R}$ is

$$[(X,s),(Y,t)] \overset{\text{def}}{=} \left([X,Y] + \sum_{m\in\mathbb{Z}} smY_m - \sum_{m\in\mathbb{Z}} tmX_m, 0\right), \quad \forall X, Y, \in \mathfrak{g}, \forall s, t \in \mathbb{R},$$

if $X = \sum_{m\in\mathbb{Z}} X_m$ and $Y = \sum_{m\in\mathbb{Z}} Y_m$ with $X_m, Y_m \in V_m$, for $m \in \mathbb{Z}$.

The Lie algebra $\mathfrak{g} \rtimes \mathbb{R}$ is $\mathbb{Z}$-graded by $(V'_m)_{m\in\mathbb{Z}}$ defined as

$$\begin{aligned} V'_0 &:= V_0 \times \{0\} \oplus \{0\} \times \mathbb{R}, \\ V'_m &:= V_m \times \{0\}, \qquad \forall m \neq 0; \end{aligned} \tag{9.15}$$

see Exercise 9.5.39.i. Moreover, if $\mathfrak{g}$ is a non-trivial Carnot algebra, then $\mathfrak{g} \rtimes \mathbb{R}$ has a trivial center; see Exercise 9.5.39.ii.

9.2.4.2 Proof of Birkhoff Theorem for Carnot Algebras

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 by (9.15); see Exercise 9.5.39. Note that $\mathfrak{g} \simeq \mathfrak{g} \times \{0\} \subset \mathfrak{g} \rtimes \mathbb{R}$, so $\operatorname{ad}(\mathfrak{g})$ can be seen as a subset of $\mathfrak{gl}(\mathfrak{g} \rtimes \mathbb{R})$. We next prove that such a map ad gives an injective representation of $\mathfrak{g}$ on the vector space $\mathfrak{g} \rtimes \mathbb{R}$.

Theorem 9.2.21 (Birkhoff-Embedding Theorem for Carnot Algebras) *Let $\mathfrak{g}$ be a Carnot algebra. Then $\operatorname{ad} : \mathfrak{g} \to \mathfrak{gl}(\mathfrak{g}\rtimes\mathbb{R})_{>0} \subset \mathfrak{gl}(\mathfrak{g}\rtimes\mathbb{R})$ is an injective Lie algebra homomorphism into the Carnot algebra $\mathfrak{gl}(\mathfrak{g} \rtimes \mathbb{R})_{>0}$ of nilpotent transformations.*

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; see Exercise 9.5.39.ii. Consequently, the map $\operatorname{ad} : \mathfrak{g} \to \mathfrak{gl}(\mathfrak{g} \rtimes \mathbb{R})$ is injective. Moreover, to see that the map is valued into $\mathfrak{gl}(\mathfrak{g} \rtimes \mathbb{R})_{>0}$, 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 $\operatorname{ad}_X(Y, s) = [(X, 0), (0, s)] = (-snX, 0) \in V_n$.

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 completed recalling Proposition 9.2.11. □

9.3 Nilpotent Lie Groups

Nilpotency for groups can be equivalently defined using either the lower central series or the upper central series, like for Lie algebras in Definition 9.1.1 and Exercise 9.5.3, respectively. For the central series in groups, recall the notation: For $a, b \in G$ we write $[a, b] := aba^{-1}b^{-1}$ and, instead, for subsets $A, B \subseteq G$ we define $[A, B]$ to be the subgroup of G generated by all those elements $[a, b]$ as a varies in A and b varies in B.

Definition 9.3.1 (Lower Central Series) Let G be a group. We iteratively define the elements of the *lower central series* $(C^i(G))_{i\in\mathbb{N}}$ of G, also called the *descending central series* of G, by

$$C^1(G) := G, \qquad \text{and} \qquad C^{i+1}(G) := [G, C^i(G)], \qquad \forall i \in \mathbb{N}.$$

Definition 9.3.2 (Upper Central Series) Let G be a group. We iteratively define the elements of the *upper central series* $(\zeta_i(G))_{i\in\mathbb{N}}$ of G by

$$\zeta_0(G) := \{1_G\} \quad \text{and} \quad \zeta_{i+1}(G) := \{g \in G : [g, G] \subseteq \zeta_i(G)\}, \qquad \forall i \in \mathbb{N}.$$

We make some simple observation about the elements $C^i(G)$ and $\zeta_i(G)$:

1. $\zeta_1(G) = Z(G)$ is the center of the group.
2. $C^1(G) = [G, G]$ is the commutator subgroup.
3. $\{1\} = \zeta_0(G) < \zeta_1(G) < \ldots < \zeta_{i-1}(G) < \zeta_i(G) < \ldots$.
4. $G = C^1(G) > C^2(G) > \ldots > C^{i-1}(G) > C^i(G) > \ldots$.

Definition 9.3.3 (Nilpotent Group) A group G is *nilpotent* if there exists d such that $C^{d+1}(G) = \{1\}$. The minimal d is called the *nilpotency step* of G. Equivalently, a group G is *nilpotent* if there exists d such that $\zeta_d = \{1\}$ and the minimal d is called the *nilpotency step* of G.

There will be an easy way to construct nilpotent Lie groups as subgroups of the matrix group Nil_n of upper triangular matrices, as presented in Example 9.3.4. Indeed, given any nilpotent Lie algebra $\mathfrak{n}$, by Birkhoff theorem, we can see it as a subalgebra of the Lie algebra $\mathfrak{nil}_n$ of Nil_n, for some $n \in \mathbb{N}$. By Theorem 5.1.4 there exists a unique connected Lie subgroup $N \subseteq \mathrm{Nil}_n$ (a priori, not closed) with $\mathrm{Lie}(N) = \mathfrak{n}$. We shall see that, actually, every such N is closed.

The main aim of this section is to show the following results:

(a) A connected Lie group is nilpotent if and only if its Lie algebra is nilpotent; see Sect. 9.4.2.
(b) Every nilpotent simply connected Lie group is isomorphic to a closed subgroup of Nil_n for some $n \in \mathbb{N}$; see Theorem 9.4.6.iii. In particular, each nilpotent simply connected Lie group is isomorphic to a matrix group, in fact to a closed subgroup of some general linear group. We will also prove that every connected subgroup of Nil_n is closed and simply connected; see Proposition 9.4.1. From these facts, we will get plenty of consequences.

9.3.1 *Examples of Nilpotent Lie Groups*

Example 9.3.4 (Upper-Triangular Unipotent Matrices) For each $n \in \mathbb{N}$, we consider the matrix group formed by all the matrices that have 1's along the diagonal and zero entries below the diagonal, e.g., for $n = 3$,

$$\mathrm{Nil}_n := \left\{ \begin{bmatrix} 1 & * & * \\ 0 & 1 & * \\ 0 & 0 & 1 \end{bmatrix} \right\} \subset \mathrm{GL}(n, \mathbb{R}).$$

The Lie group Nil_n is nilpotent of step $(n-1)$ and its Lie algebra is $\mathfrak{nil}_n$ as defined in Example 9.1.9.

Example 9.3.5 Every closed subgroup of Nil_n is a nilpotent Lie group. We shall see that every nilpotent simply connected Lie group is of this type; see Proposition 9.4.3.

Example 9.3.6 ($\mathrm{Nil}(\mathcal{F})$) Here is a slight generalization of the previous example. Let V be a finite-dimensional vector space and let $\mathcal{F} = (V_0, \ldots, V_n)$ be a flag for V. The Lie group

$$\mathrm{Nil}(\mathcal{F}) := \left\{ A \in \mathfrak{gl}(V) : (A - \mathbb{I})(V_k) \subseteq V_{k-1}, \forall k \in \{1, \ldots, n\} \right\}$$

is nilpotent of step $n-1$, with Lie algebra $\mathfrak{g}_{\mathrm{nil}}(\mathcal{F})$ as in Example 9.1.10.

Example 9.3.7 (Heisenberg Groups) The $(2n+1)$-*dimensional Heisenberg Lie algebra* is the Lie algebra with basis $\{X_1, \ldots, X_n, Y_1, \ldots, Y_n, Z\}$, whose pairwise brackets are equal to zero except for

$$[X_j, Y_j] = Z, \qquad \text{for } j \in \{1, \ldots, n\}.$$

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

$$\begin{bmatrix} 0 & x_1 & \dots & x_n & z \\ \cdot & 0 & \cdot & 0 & y_1 \\ \cdot & & \cdot & \cdot & \vdots \\ \cdot & & & 0 & y_n \\ 0 & \cdot & \cdot & \cdot & 0 \end{bmatrix}, \qquad \text{for } x_1, \dots, x_n, y_1, \dots, y_n, z \in \mathbb{R}.$$

The simply connected Lie group associated with this Lie algebra is called the *n-th Heisenberg group*, and as a matrix group, it is

$$G = \left\{ \begin{bmatrix} 1 & x_1 & \dots & x_n & z \\ \cdot & 1 & \cdot & 0 & y_1 \\ \cdot & & \ddots & \cdot & \vdots \\ \cdot & & & 1 & y_n \\ 0 & \cdot & \cdot & \cdot & 1 \end{bmatrix} : x_1, \dots, x_n, y_1, \dots, y_n, z \in \mathbb{R} \right\} \subset \mathrm{GL}(n+2, \mathbb{R}).$$

Every Heisenberg group is nilpotent of step 2.

Example 9.3.8 If $\mathfrak{g}$ is a 2-step Lie algebra, then

$$(X, Y) \in \mathfrak{g} \times \mathfrak{g} \mapsto X \cdot Y := X + Y + \frac{1}{2}[X, Y],$$

defines a group structure on $\mathfrak{g}$. Such a Lie group is nilpotent of step 2. Hence, every nilpotent simply connected Lie group is isomorphic to a group G_q as in Definition 7.3.7.

9.3.2 The Exponential Function on Nilpotent Matrices

Nilpotent simply connected Lie groups have the feature that their exponential maps are global diffeomorphisms. Recall from Sect. 5.4.2 that for every finite-dimensional vector space V the exponential map $\exp : \mathfrak{gl}(V) \to \mathrm{GL}(V)$ is

$$A \mapsto e^A := \sum_{n=0}^{\infty} \frac{A^n}{n!}.$$

We shall first present a local inverse map around the identity transformation $\mathbb{I} = e^0$ in $\mathrm{GL}(V)$.

Definition 9.3.9 (Logarithm Function) Fix $n \in \mathbb{N}$. Using the operator norm on $\mathrm{GL}(\mathbb{R}^n)$, we define the set

$$B_{\mathbb{I}}(1) := \{M \in \mathrm{GL}(\mathbb{R}^n) \,:\, \|M - \mathbb{I}\| < 1\} \subseteq \mathrm{GL}(\mathbb{R}^n)$$

and the map $\log : B_{\mathbb{I}}(1) \subseteq \mathrm{GL}(\mathbb{R}^n) \to \mathfrak{gl}(\mathbb{R}^n)$ as

$$M \mapsto \log(M) := \sum_{k=1}^{\infty} \frac{(-1)^k}{k} (M - \mathbb{I})^k.$$

Remark 9.3.10 Since we have uniform convergence for all $x \in (-1, 1)$ of $\log(1 + x) = \sum_{k=1}^{\infty} \frac{(-1)^k}{k} x^k$, the map $\log$ is smooth on the open set $B_{\mathbb{I}}(1)$. Moreover, as a consequence of a formal series inversion, we have that the map $\log$ is the inverse of $\exp$ in small enough neighborhoods of $\mathbb{I}$ and 0, respectively; see Exercise 9.5.43.

In the next proposition, we shall consider the set $\mathcal{U}$ of unipotent matrices and the set $\mathcal{N}$ of nilpotent matrices. The set $\mathcal{U}$ is not a subgroup, nor is $\mathcal{N}$ a subalgebra; see Exercise 9.5.44. We shall then study how to equivalently express $\exp$ and $\log$ on these sets.

Proposition 9.3.11 *Fix $n \in \mathbb{N}$ and consider the following sets:*

$$\begin{aligned} \mathcal{U} &:= \{M \in \mathrm{GL}(\mathbb{R}^n) \,:\, (M - \mathbb{I})^n = 0\}, \\ \mathcal{N} &:= \{A \in \mathfrak{gl}(\mathbb{R}^n) \,:\, A^n = 0\}. \end{aligned} \tag{9.16}$$

The map $\exp_{|\mathcal{N}} : \mathcal{N} \to \mathcal{U}$ is equivalently defined as

$$A \mapsto e^A \stackrel{\text{def}}{=} \sum_{k=0}^{\infty} \frac{A^k}{k!} = \sum_{k=0}^{n} \frac{A^k}{k!},$$

and is a homeomorphism with inverse $\log_{|\mathcal{U}} : \mathcal{U} \to \mathcal{N}$, which is equivalently defined as

$$M \mapsto \log M \stackrel{\text{def}}{=} \sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k} (M - \mathbb{I})^k = \sum_{k=1}^{n} \frac{(-1)^{k+1}}{k} (M - \mathbb{I})^k.$$

We stress that in the second sums above, the sums over k stop at n.

Proof Observe that when restricting to $\mathcal{U}$, respectively to $\mathcal{N}$, the finite sums give the same maps. Moreover, since the second sums are finite, they define maps that are polynomial and globally defined on $\mathfrak{gl}(\mathbb{R}^n)$.

We check that $\exp_{|\mathcal{N}}$ is $\mathcal{U}$-valued: Take $A \in \mathcal{N}$, so $A^n = 0$. Observe that A and $\sum_{k=0}^{n} \frac{A^{k-1}}{k!}$ commute. Thus

$$(e^A - I)^n = \left(\sum_{k=1}^{n} \frac{A^k}{k!}\right)^n = A^n \left(\sum_{k=1}^{n} \frac{A^{k-1}}{k!}\right)^n = 0.$$

In a similar way we check that $\log_{|\mathcal{U}}$ is $\mathcal{N}$-valued: Take $M \in \mathcal{U}$ so $(M - \mathbb{I})^n = 0$ and so, because $M - \mathbb{I}$ commutes with its powers, we get

$$(\log M)^n = (M - \mathbb{I})^n \left(\sum_{k=1}^{n} \frac{(-1)^{k+1}}{k} (M - \mathbb{I})^{k-1}\right)^n = 0.$$

Finally, we check that $\exp \circ \log$ is the identity map. Fix $A \in \mathcal{N}$. Consider $t \in \mathbb{R} \mapsto \log_{|\mathcal{U}}(\exp_{|\mathcal{N}}(tA))$. Notice that $tA \in \mathcal{N}$ for every $t \in \mathbb{R}$ and, for all t small enough, the element tA is in a neighborhood of 0 where $\exp$ has $\log$ as inverse; see Exercise 9.5.43. Hence, for all t small enough, we get

$$\log(\exp(tA)) = tA.$$

Since both functions are polynomial in t, then they coincides for every t and in particular for $t = 1$.

Similarly, $\exp_{|\mathcal{N}}(\log_{|\mathcal{U}}(\mathbb{I} + tA))$ and $\mathbb{I} + tA$ are polynomials in t that coincide for all t small enough, hence they do at $t = 1$. □

We now consider $\mathfrak{nil}_n$ and Nil_n, from Examples 9.1.9 and 9.3.4, respectively. We see $\mathfrak{nil}_n$ as a subset of $\mathcal{N}$ and Nil_n as a subset of $\mathcal{U}$. The exponential map on $\mathfrak{nil}_n$ is given by the finite sum shown in Proposition 9.3.11.

Corollary 9.3.12 *The map* $\exp : \mathfrak{nil}_n \to \mathrm{Nil}_n$ *is a polynomial diffeomorphism with polynomial inverse* $\log_{|\mathrm{Nil}_n} : \mathrm{Nil}_n \to \mathfrak{nil}_n$.

9.4 Connected Nilpotent Lie Groups

9.4.1 Connected Lie Subgroups of Nil$_n$

We shall prove that every nilpotent simply connected Lie group is, in fact, a subgroup of some Nil_n; see Proposition 9.4.3. Before that, we study connected subgroups of Nil_n.

Proposition 9.4.1 *If G is a connected Lie subgroup of* Nil_n, *then* $\exp(\mathrm{Lie}(G)) = G$ *and in particular G is closed and simply connected.*

Proof Recall from Corollary 9.3.12 that on Nil_n and on its Lie algebra $\mathfrak{nil}_n$, the respective maps $\exp_{|\mathcal{N}}$ and $\log_{|\mathcal{U}}$ are polynomial and are inverse of each other. Let $\mathfrak{g} := \mathrm{Lie}(G) \subset \mathfrak{nil}_n$. Clearly, $\exp_{|\mathcal{N}}(\mathfrak{g}) = \exp(\mathfrak{g}) \subset G$. We next prove the other inclusion. Let U be a connected neighborhood of 0 in $\mathfrak{g}$ such that $\exp_{|U} : U \to \exp(U)$ is a diffeomorphism and $V := \exp(U)$ is a neighborhood of 1_G in G. From a general argument valid for topological groups, since G is connected, we have that $G = \bigcup_{m=1}^{\infty} V^m$; see Exercise 5.8.3. We shall prove by induction on $m \in \mathbb{N}$ that

$$V^m \subset \exp(\mathfrak{g}). \tag{9.17}$$

The base of induction is that $V^1 = V \overset{\text{def}}{=} \exp(U) \subset \exp(\mathfrak{g})$. Now assume that the claim (9.17) holds for $m \in \mathbb{N}$. Take $w \in V^{m+1} = V^m \cdot V$. Hence, there are $x \in \mathfrak{g}$ and $y \in U$ such that $w = \exp(x)\exp(y)$ and so using BCH formula (see Proposition 5.7.4) we have that $w = \exp(x + y + \frac{1}{2}[x, y] + \dots)$, where the sum is actually finite and it is by terms in $\mathfrak{g}$, then $w \in \exp(\mathfrak{g})$. Consequently, we infer that $\exp(\mathfrak{g}) = G$. The last assertion of the proposition is a consequence of the fact that $\mathrm{Lie}(G)$ is a vector subspace (hence closed and simply connected) and $\exp$ is a diffeomorphism on $\mathfrak{nil}_n$; see Corollary 9.3.12. □

9.4.2 Lie Groups with Nilpotent Lie Algebras

In this section, we explain why a connected Lie group is nilpotent if and only if its Lie algebra is nilpotent. What the reader should expect is that the lower central series are linked by $C^m(G) = \exp(C^m(\mathfrak{g}))$, for $m \in \mathbb{N}$, at least for (nilpotent) simply connected Lie groups.

9.4.2.1 Commutator Subgroups

We point out a general fact about Lie groups: For every connected Lie group with Lie algebra $\mathfrak{g}$, the commutator subgroup $[G, G]$ is a Lie subgroup of G whose Lie algebra is $[\mathfrak{g}, \mathfrak{g}]$. The hardest part in obtaining the mentioned fact is explaining why $[G, G]$ is a Lie subgroup. A warning is that $[G, G]$ may not be topologically closed. It is if G is simply connected. The reader can read about these facts in [Hoc65, page 138]. We will not further discuss this topic in this general context because when G is nilpotent and simply connected, we will easily have that $[G, G]$ is closed; see Exercise 9.5.53.

9.4.2.2 Lie Algebras of Nilpotent Lie Groups

It is easy to show that nilpotent Lie groups have nilpotent Lie algebras. Doing the increasingly difficult Exercises 9.5.50 and 9.5.51, the reader should be able to come up with an argument for the next proposition.

Proposition 9.4.2 *If G is a nilpotent Lie group of step s, then* $\mathrm{Lie}(G)$ *is a nilpotent Lie algebra of step s.*

9.4.2.3 Lie Groups with Nilpotent Lie Algebras

Clearly, unless the group is connected, we cannot expect that some information on the Lie algebra can give a global information on the group. For example, every finite group is a 0-dimensional Lie group whose Lie algebra is nilpotent (and abelian, trivially). Hence, there are Lie groups that are not nilpotent but have a nilpotent Lie algebra.

We shall prove that if G is a connected Lie group and its Lie algebra $\mathfrak{g}$ is nilpotent, then G is nilpotent. There are various ways to show the latter fact. A common point is to construct a nilpotent Lie group N with Lie algebra isomorphic to $\mathfrak{g}$.

One way is to use Birkhoff Embedding Theorem, Theorem 9.1.19:

Proposition 9.4.3 (After Birkhoff's Theorem) *For every nilpotent Lie algebra $\mathfrak{g}$, there exists $n \in \mathbb{N}$ and a nilpotent closed simply connected Lie subgroup $N \subseteq \mathrm{Nil}_n$ with Lie algebra isomorphic to $\mathfrak{g}$.*

Another way is to use the Baker-Campbell-Hausdorff-(Dynkin) formula of Proposition 5.7.4:

Proposition 9.4.4 (After BCH Formula) *For every nilpotent Lie algebra $\mathfrak{g}$, if we equip $\mathfrak{g}$ with the Dynkin product $\star$, then $(\mathfrak{g}, \star)$ is a nilpotent simply connected Lie group with Lie algebra isomorphic to $\mathfrak{g}$. Moreover, the exponential map $\exp : \mathfrak{g} \to (\mathfrak{g}, \star)$ is the identity map.*

We shall prove the above two propositions after the next corollary.

Corollary 9.4.5 (Consequence of Either of the Propositions) *Every connected Lie group with nilpotent Lie algebra is nilpotent.*

Proof of Corollary 9.4.5 Let G be a connected Lie group with nilpotent Lie algebra $\mathfrak{g}$. Either from Proposition 9.4.3 or from Proposition 9.4.4, let N be a nilpotent simply connected Lie group with Lie algebra isomorphic to $\mathfrak{g}$. Thus, the Lie group N and the universal cover $\tilde{G}$ of G are simply connected with isomorphic Lie algebras. Hence, by Corollary 5.1.6, they are isomorphic. In particular, since N is nilpotent, so is $\tilde{G}$. Consequently, being the quotient of a nilpotent group, also the group G is nilpotent, as desired. □

Proof of Proposition 9.4.3 By Birkhoff Theorem 9.1.19, there are $n \in \mathbb{N}$ and a subalgebra $\mathfrak{n} \subseteq \mathfrak{nil}_n$ that is isomorphic to $\mathfrak{g}$. By Theorem 5.7.1, let $N \subseteq \mathrm{Nil}_n$ be the connected Lie subgroup with Lie algebra $\mathfrak{n}$. Then, by Proposition 9.4.1, the Lie group N is closed and simply connected. Moreover, since Nil_n is nilpotent then so is N. □

Proof of Proposition 9.4.4 Being $\mathfrak{g}$ a nilpotent Lie algebra, the series in the Dynkin product is finite and, for every $x, y \in \mathfrak{g}$, the product $x \star y = \log_{|U}(\exp_{|N}(x)\exp_{|N}(y))$ is polynomial in x and y. Obviously, the manifold $(\mathfrak{g}, \star)$ is simply connected because it is isomorphic to a vector space. We shall check that $(\mathfrak{g}, \star)$ is a Lie group, it is nilpotent, its Lie algebra is $\mathfrak{g}$, and its exponential map is the identity. By Ado's theorem (see Theorem 5.1.7), there is[1] a group G with $\mathfrak{g}$ as Lie algebra. Moreover, by the BCH Formula, in a small enough neighborhood of 1_G, the group product of G is exactly the Dynkin product, up to the identification via the exponential map. In particular, since G is a group, then the product $\star$ satisfies the axioms of group products near $1_{(\mathfrak{g},\star)} = 0$ and also $x \star (-x) = 1_{(\mathfrak{g},\star)}$ is true near 0. These identities are polynomial maps, which are verified in a neighborhood of 0. By analytic continuation, they hold everywhere on $\mathfrak{g}$.

By the BCH formula, the exponential map $\exp : \mathfrak{g} \to G$ is a Lie group isomorphism between the Lie group $(\mathfrak{g}, \star)$ and the Lie group G. Thus, the Lie algebra of $(\mathfrak{g}, \star)$ is isomorphic to $\mathfrak{g}$.

We next explain why $(\mathfrak{g}, \star)$ is nilpotent. We stress that we do not know (yet) that G is nilpotent. The reader might find it helpful to know that there is a result by Lazard saying that every polynomial group product in $\mathbb{R}^n$ gives a nilpotent group; see [Laz55] or [Dek03]. Still, for the Dynkin product, this fact is easier to show because

$$x \star y - (x + y) \in [x, \mathfrak{g}] + [y, \mathfrak{g}] \subseteq [\mathfrak{g}, \mathfrak{g}], \qquad \forall x, y \in \mathfrak{g}.$$

Hence, we get

$$x \star y \star x^{-1} \star y^{-1} \in [\mathfrak{g}, \mathfrak{g}], \qquad \forall x, y \in \mathfrak{g}.$$

Since $\mathfrak{g}$ is a nilpotent Lie algebra, iterating this group product, we obtain nilpotency of the group $(\mathfrak{g}, \star)$.

The one-parameter subgroups for the group product $\star$ are the curves $t \in \mathbb{R} \mapsto tx$ for each $x \in \mathfrak{g}$, because $(sx) \star (tx) = (s + t)x$. So the identity $\mathfrak{g} \to \mathfrak{g}$ is the exponential map $\exp : \mathfrak{g} \simeq \mathrm{Lie}(\mathfrak{g}, \star) \to (\mathfrak{g}, \star)$. □

In the next section, we shall see that another consequence of Proposition 9.4.4 is that every nilpotent simply connected Lie group is isomorphic to a closed subgroup of Nil_n for some $n \in \mathbb{N}$.

[1] However, notice that for Corollay 9.4.5 we already know the existence of G.

9.4.3 Simply Connected Nilpotent Lie Groups

Simply connected Lie groups are uniquely determined by their Lie algebras. Indeed, recall from Corollary 5.1.6 that if two simply connected Lie groups have isomorphic Lie algebras, then they are isomorphic. For nilpotent Lie groups, either Birkhoff's embedding theorem or the exponential map and the BCH formula provide a concrete identification. Thus, we can work remaining at the level of the Lie algebra using such coordinates.

Theorem 9.4.6 *Every nilpotent simply connected Lie group G has the following properties:*

9.4.6.i. The exponential map $\exp : \mathrm{Lie}(G) \to G$ *is a diffeomorphism.*
9.4.6.ii. The Baker-Campbell-Hausdorff Formula (5.24) *holds globally.*
9.4.6.iii. There exists $n \in \mathbb{N}$ and a closed nilpotent simply connected Lie subgroup of Nil_n *isomorphic to G.*

Proof We begin by observing that 9.4.6.iii is a direct consequence of Theorems 9.1.19, 5.7.1, and Proposition 9.4.1, as we saw in the proof of Proposition 9.4.3.

In addition, on Nil_n we also have that the two functions $A \star B$ and $\log(e^A e^B)$ are analytic maps that coincide in a neighborhood of 0, recalling Proposition 5.7.4 and Corollary 9.3.12, and also Exercise 9.5.46. Then 9.4.6.ii holds for each Nil_n.

Once we know that the statements are valid for subgroups of Nil_n, they hold for arbitrary nilpotent simply connected Lie groups by Corollary 9.3.12. □

A Lie algebra is called *positively gradable* if it admits a positive grading, and *stratifiable* if it admits a stratification. A Lie algebra is referred to as $\mathbb{R}$-*graded* if it is equipped with an $\mathbb{R}$-grading, *positively graded* if it is equipped with a positive grading, and *stratified* if it is equipped with a stratification. Similarly, a Lie group is called *positively gradable*, *stratifiable*, $\mathbb{R}$-*graded*, *positively graded*, or *stratified* if it is simply connected and its Lie algebra is positively gradable, stratifiable, $\mathbb{R}$-graded, positively graded, or stratified, respectively.

If a Lie group G is $\mathbb{R}$-graded, then its Lie algebra $\mathfrak{g}$ has the dilations $\delta_\lambda : \mathfrak{g} \to \mathfrak{g}$ as in (9.8), and, because G is simply connected, by Theorem 5.1.5 we have that each Lie algebra endomorphism δ_λ induces a Lie group endomorphism of G whose induced Lie algebra homomorphism is δ_λ. Such a Lie group endomorphism is still denoted by $\delta_\lambda : G \to G$, so $(\delta_\lambda)_* = \delta_\lambda$. Moreover, from Theorem 5.2.9 we have

$$\delta_\lambda \circ \exp = \exp \circ \delta_\lambda. \tag{9.18}$$

When G, in addition, is simply connected and nilpotent, the map $\exp : \mathfrak{g} \to G$ is a diffeomorphism by Theorem 9.4.6.i. Therefore, every element $g \in G$ can be represented as $\exp(X)$ for some unique $X \in \mathfrak{g}$, and therefore uniquely written in the form

$$g = \exp\left(\sum_{a\in\mathbb{R}} v_a\right), \qquad v_a \in V_a,\ a \in \mathbb{R}.$$

This representation allows us to have the formula:

$$\delta_\lambda\left(\exp\left(\sum_{a\in\mathbb{R}} v_a\right)\right) = \exp\left(\sum_{a\in\mathbb{R}} \lambda^a v_a\right). \tag{9.19}$$

9.4.4 Canonical Coordinates

One important consequence of Theorem 9.4.6 is the existence of good coordinates on nilpotent simply connected Lie groups. For these groups, since the exponential map $\exp : \mathfrak{g} \to G$ is a diffeomorphism, we can use it to transfer coordinates from $\mathfrak{g}$ to the nilpotent simply connected Lie group G. Some authors use $\exp$ to identify $\mathfrak{g}$ with G. Then, the group multiplication can be calculated using the Baker-Campbell-Hausdorff formula.

We shall introduce different types of coordinate systems: exponential coordinates and Malcev coordinates. In these coordinates, the group operations will be polynomial, and the Lebesgue measure will be invariant both by left translations and by right translations. The following theorem summarises the content of this subsection.

Theorem 9.4.7 *In every nilpotent simply connected Lie group, exponential coordinates and Malcev coordinates are global coordinates. The group product in such coordinates is polynomial. Left translations and right translations have Jacobian one. Consequently, the Lebesgue measure in these coordinates is a bi-invariant Haar measure.*

We call *bi-invariant Haar measure* every left-Haar measure that is also a right-Haar measure, in the sense of Sect. 6.4. For the proof of the theorem, we refer to the end of this subsection.

9.4.4.1 Exponential Coordinates

Definition 9.4.8 (Exponential Coordinates: Canonical Coordinates of the 1st Kind) Let $(X_1, \dots, X_n)$ be an ordered basis for a nilpotent Lie algebra of a nilpotent simply connected Lie group G. The coordinates given by the map

$$\Phi : \mathbb{R}^n \longrightarrow G$$

$$\Phi(t_1, \dots, t_n) := \exp(t_1 X_1 + \dots + t_n X_n)$$

are called *exponential coordinates* with respect to $X_1, \dots, X_n$. The map Φ is called *exponential coordinate system*. Exponential coordinates are also known as *canonical coordinates of the first kind*.

With the choice of the basis and the use of exp, we are identifying $\mathbb{R}^n$ with $\mathrm{Lie}(G)$ and G. Moreover, the group product can be obtained via the BCH formula from Sect. 5.7.3:

$$(s_1, \dots, s_n) \star (t_1, \dots, t_n) = \log\left(\exp\left(\sum_{j=1}^n s_j X_j\right)\exp\left(\sum_{j=1}^n t_j X_j\right)\right).$$

Example 9.4.9 We consider the Heisenberg group, whose Lie algebra has a basis X_1, X_2, X_3 with only nontrivial relation $[X_1, X_2] = X_3$. The exponential coordinate system with respect to this basis is the map

$$(x_1, x_2, x_3) \overset{\Phi}{\longmapsto} \exp(x_1 X_1 + x_2 X_2 + x_3 X_3). \tag{9.20}$$

In these coordinates, the product can be expressed by the BCH formula:

$$\begin{aligned}
&\exp(x_1 X_1 + x_2 X_2 + x_3 X_3)\exp(x_1' X_1 + x_2' X_2 + x_3' X_3)\\
&\quad = \exp\Big(x_1 X_1 + x_2 X_2 + x_3 X_3 + x_1' X_1 + x_2' X_2 + x_3' X_3\\
&\qquad\qquad + \frac{1}{2}\big[x_1 X_1 + x_2 X_2 + x_3 X_3,\ x_1' X_1 + x_2' X_2 + x_3' X_3\big]\Big)\\
&\quad = \exp\left((x_1 + x_1')X_1 + (x_2 + x_2')X_2 + (x_3 + x_3')X_3 + \frac{1}{2}(x_1 x_2' - x_2 x_1')X_3\right),
\end{aligned}$$

where we used the Lie bracket relations of the Heisenberg Lie algebra. We conclude that the product, when read in these coordinates, is

$$(x_1, x_2, x_3) \star (x_1', x_2', x_3') = \left(x_1 + x_1', x_2 + x_2', x_3 + x_3' + \frac{1}{2}(x_1 x_2' - x_2 x_1')\right).$$

This is indeed the product structure (2.9) that we used to initially present the Heisenberg group in Chap. 2.

9.4.4.2 Malcev Bases

It is useful to choose bases that are better adapted to the Lie brackets. We shall then have other canonical coordinates in Definition 9.4.13.

Definition 9.4.10 (Malcev Basis for a Lie Algebra) Let $\mathfrak{g}$ be a Lie algebra. An ordered basis $(X_1, \dots, X_n)$ for $\mathfrak{g}$ is called a *strong Malcev basis*, or, simply, a *Malcev basis* if for every $k \in \{1, \dots, n\}$ the space

$$\mathfrak{g}_k := \mathrm{span}\{X_1, \dots, X_k\}$$

is an ideal of $\mathfrak{g}$, i.e., $[\mathfrak{g}, \mathfrak{g}_k] \subset \mathfrak{g}_k$.

We remark that there exist non-nilpotent Lie algebras (e.g., in 2D) that admit Malcev bases. We shall not discuss the notion of weak Malcev basis, for which we refer to [CG90, page 10].

The spaces $\mathfrak{g}_k$ coming from a Malcev basis satisfy the following stronger property.

Proposition 9.4.11 *If $\mathfrak{g}$ is a nilpotent Lie algebra and $(X_1, \dots, X_n)$ is a Malcev basis for $\mathfrak{g}$, then*

$$[\mathfrak{g}, \mathfrak{g}_k] \subseteq \mathfrak{g}_{k-1} \subseteq \mathfrak{g}_k, \qquad \textit{for } k \in \{1, \dots, n\}, \tag{9.21}$$

where $\mathfrak{g}_k := \mathrm{span}\{X_1, \dots, X_k\}$ and $\mathfrak{g}_0 := \{0\}$.

Proof Fix $k \in \{1, \dots, n\}$. 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 proposition were not true, then there would be some $Y \in \mathfrak{g}$ and $a_1, \dots, a_k \in \mathbb{R}$ with $a_k \neq 0$ such that

$$[Y, X_k] = a_k X_k + \sum_{i=1}^{k-1} a_i X_i.$$

Now we iterate bracketing by Y, i.e., we iterate the map $\mathrm{ad}_Y = [Y, \cdot]$. Thus, we get, for some $a_1^{(l)}, \dots, a_{k-1}^{(l)} \in \mathbb{R}$, the value

$$(\mathrm{ad}_Y^l)(X_k) = (a_k)^l X_k + \sum_{i=1}^{k-1} a_i^{(l)} X_i \in (a_k)^l X_k + \mathfrak{g}_{k-1}, \qquad \forall l \in \mathbb{N},$$

which is never zero and so contradicts the nilpotency of $\mathfrak{g}$. □

In the special class of Carnot groups, as considered in Chap. 11, the existence of Malcev bases will be a triviality; see Exercise 9.5.55. However, all nilpotent Lie algebras have Malcev bases. For additional information, we refer to [CG90, Theorem 1.1.13] and the notes therein.

Proposition 9.4.12 *In every nilpotent Lie algebra, Malcev bases exist.*

Proof Let $\mathfrak{g}$ be a nilpotent Lie algebra. For $i \in \mathbb{N}$, let $\zeta_i := \zeta_i(\mathfrak{g})$ be the i-th element of the upper central series of $\mathfrak{g}$, as in Exercise 9.5.3. For some $s \in \mathbb{N}$ we have

$$\{0\} \neq \zeta_1 = Z(\mathfrak{g}) \subsetneq \ldots \subsetneq \zeta_{s-1} \subsetneq \zeta_s = \mathfrak{g}.$$

To build a Malcev basis for $\mathfrak{g}$, we begin with a basis $X_1, \ldots, X_{n_1}$ of ζ_1. Next, we complete it to a basis $X_1, \ldots, X_{n_2}$ of ζ_2. Iterating, given a basis of ζ_{j-1}, we complete it to a basis $X_1, \ldots, X_{n_j}$ of ζ_j. Because of nilpotency, we obtain a basis of $\mathfrak{g}$ after s steps. In fact, the vectors $X_1, \ldots, X_{n_s}$ form a basis of $\mathfrak{g}$ such that for every $k \in \{1, \ldots, n_s\}$ there is $j_k \in \{1, \ldots, s\}$ such that $X_k \in \zeta_{j_k} \setminus \zeta_{j_k-1}$. For such a j_k, we have

$$\zeta_{j_k-1} \subseteq \mathfrak{g}_k := \operatorname{span}\{X_1, \ldots, X_k\} \subseteq \zeta_{j_k}.$$

Hence, we deduce $[\mathfrak{g}, \mathfrak{g}_k] \subset [\mathfrak{g}, \zeta_{j_k}] \overset{\text{def}}{=} \zeta_{j_k-1} \subset \mathfrak{g}_k$. □

9.4.4.3 Malcev Coordinates

Definition 9.4.13 (Malcev Coordinates: Canonical Coordinates of the 2nd Kind) Let G be a Lie group with Lie algebra $\mathfrak{g}$. Fix an ordered basis $(X_1, \ldots, X_n)$ of $\mathfrak{g}$. The coordinates given by the map

$$\Psi : (s_1, \ldots, s_n) \in \mathbb{R}^n \mapsto \exp(s_1 X_1) \cdots \exp(s_n X_n) \in G,$$

are called *exponential coordinates of second kind*, or *canonical coordinates of the second kind* with respect to $(X_1, \ldots, X_n)$. If $(X_1, \ldots, X_n)$ is a Malcev basis, then the defined coordinate system Ψ is called *Malcev coordinate system*.

We remark that by observing the injectivity of the differential at $\mathbf{0}$ of the map Ψ as in Definition 9.4.13, one can see that Ψ gives local coordinates around 1_G. We will soon see that it gives global coordinates when the basis is a Malcev basis and G is nilpotent and simply connected.

Example 9.4.14 (Malcev and Non-Malcev Coordinates) In the Heisenberg group, we have the exponential coordinates Φ, as in (9.20), with respect to the standard basis (X_1, X_2, X_3). Such a basis, in this order, is not a Malcev basis: we stress that X_3 is in the derived subalgebra, while X_1 is not. Instead, the ordered basis (X_3, X_2, X_1) is a Malcev basis. We may not have good properties if we consider canonical coordinate systems of the second kind with respect to bases that are not Malchev bases. For example, the triple $(X_1, X_2, X_1 + X_3)$ is not a Malcev basis, and if Ψ denotes the canonical coordinate system of the second kind with respect to this basis, then $\Phi^{-1} \circ \Psi$ is the map

$$(s_1, s_2, s_3) \in \mathbb{R}^3 \mapsto \left(s_1 + s_3,\ s_2,\ s_3 + \frac{s_1 s_2 - s_2 s_3}{2}\right) \in \mathbb{R}^3.$$

This map has a Jacobian determinant equal to $1 - s_2$. We should avoid canonical coordinates of the second kind with respect to non-Malcev bases.

If $(X_1, \ldots, X_n)$ is a Malcev basis for the Lie algebra of a nilpotent Lie group, we can consider both canonical coordinate systems; we have that the Malcev coordinates are related to the exponential coordinates by a polynomial diffeomorphism. Such a polynomial map has an upper triangular form, and, in particular, its Jacobian determinant is constantly equal to 1. The proof is an adaptation of [CG90, Proposition 1.2.7].

Proposition 9.4.15 (Change of Malcev Coordinates) *Let G be a nilpotent simply connected Lie group with Lie algebra $\mathfrak{g}$. Let $(X_1, \ldots, X_n)$ be a Malcev basis for $\mathfrak{g}$. Let $\Psi : \mathbb{R}^n \to G$ be the Malcev coordinate system and $\Phi : \mathbb{R}^n \longrightarrow G$ be the exponential coordinate system associated with the basis. Then*

9.4.15.i. *The map $P := \Phi^{-1} \circ \Psi$ from $\mathbb{R}^n$ to $\mathbb{R}^n$ is a polynomial diffeomorphism with polynomial inverse;*
9.4.15.ii. *Writing the components $P = (P_1, \ldots, P_n)$, then*

$$P_j(s) = s_j + \hat{P}_j(s_{j+1}, \ldots, s_n),$$

where $\hat{P}_j$ is a polynomial depending only on the last $n - j$ variables, for all $j \in \{1, \ldots, n\}$.

In other words, each Malcev coordinate system is a global diffeomorphism and, for some polynomials $P_1, \ldots, P_n$, we have the relation:

$$\exp(s_1 X_1) \cdots \exp(s_n X_n) = \exp(P_1(s) X_1 + \ldots + P_n(s) X_n).$$

Proof Recall that we are considering

$$\Psi(s_1, \ldots, s_n) \stackrel{\text{def}}{=} \exp(s_1 X_1) \cdots \exp(s_n X_n), \qquad \Phi(s_1, \ldots, s_n) \stackrel{\text{def}}{=} \exp\left(\sum_{i=1}^{n} s_i X_i\right).$$

Because of BCH Formula (Proposition 5.7.4), the map

$$P := \Phi^{-1} \circ \Psi : \mathbb{R}^n \to \mathbb{R}^n$$

is a polynomial map

$$s := (s_1, \ldots, s_n) \mapsto (P_1(s), \ldots, P_n(s)).$$

By induction on n, we shall prove

$$P_j(s) = s_j + Q_j(s_{j+1}, \ldots, s_n), \qquad \forall j \in \{1, \ldots, n\}.$$

When $n = 1$, it is trivial. For arbitrary n, we take the quotient $\pi : \mathfrak{g} \to \mathfrak{g}/\mathfrak{g}_1$ modulo $\mathfrak{g}_1 := \mathbb{R}X_1 \lhd \mathfrak{g}$. Then, for every $j \in \{2, \dots, n\}$, we set $\overline{X}_j := \pi(X_j)$, and we notice that $(\overline{X}_2, \dots, \overline{X}_n)$ is a Malcev basis for $\mathfrak{g}/\mathfrak{g}_1$. Now, in G we have the equation

$$\exp(s_1 X_1) \cdots \exp(s_n X_n) = \exp\left(\sum_{i=1}^{n} P_i(s) X_i\right),$$

which, after passing to the quotient $G/\exp(\mathbb{R}X_1)$, becomes

$$\exp(s_2 \overline{X}_2) \cdots \exp(s_n \overline{X}_n) = \exp\left(\sum_{i=2}^{n} P_i(s) \overline{X}_i\right).$$

Thus, by induction, we have

$$P_j(s) = s_j + Q_j(s_{j+1}, \dots, s_n), \qquad \forall j \in \{2, \dots, n\}.$$

Lifting to G we get

$$\exp(s_2 X_2) \cdots \exp(s_n X_n) = \exp\left(Q_1(s_2, \dots, s_n) X_1 + \sum_{i=2}^{n} P_i(s) X_i\right),$$

for some polynomial Q_1. Notice that X_1 is central, hence

$$\exp(X_1 + Y) = \exp(X_1)\exp(Y), \qquad \forall Y \in \mathfrak{g}.$$

Thus, from the previous equation, we get

$$\exp(s_1 X_1) \cdots \exp(s_n X_n) = \exp\left((s_1 + Q_1(s_2, \dots, s_n)) X_1 + \sum_{i=2}^{n} P_i(s) X_i\right).$$

Now, in order to finish, notice that given $t \in \mathbb{R}^n$, the equations

$$t_j = s_j + P_j(s_{j+1}, \dots, s_n), \qquad \forall j \in \{1, \dots, n\},$$

can be solved in s with a formula

$$s_k = t_k + \widetilde{Q}_k(t_{k+1}, \dots, t_n),$$

where $\widetilde{Q}_k$ are polynomials. □

Remark 9.4.16 If $(X_1, \ldots, X_n)$ is a Malcev basis of a nilpotent simply connected Lie group, then left and right cosets of $\exp(\mathfrak{g}_k)$ are affine planes (set-wise) in exponential coordinates of the second kind. This may not be true in exponential coordinates of the first kind.

9.4.4.4 Jacobian Determinants in Exponential Coordinates

The following result describes properties of a nilpotent Lie group law expressed in exponential coordinates with respect to some Malcev basis. The proof is an adaptation of [CG90, Proposition 1.2.9].

Proposition 9.4.17 (Group Law in Malcev Coordinates) *Let G be a nilpotent simply connected Lie group equipped with a Malcev basis. On G, consider exponential coordinates of the first or second kind associated with the basis. In this coordinate system, the group law has an upper triangular form:*

$$(s_1, \ldots, s_n) \cdot (t_1, \ldots, t_n) = s + t + \sum_{j=1}^{n} Q_j(s,t) e_j, \qquad \forall s, t \in \mathbb{R}^n,$$

where each Q_j is a polynomial that is not depending on $s_1, \ldots, s_{j-1}$ nor $t_1, \ldots, t_{j-1}$. In particular, 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 of the first kind and left translations. The case of right translations is similar. For Malcev coordinates, the result will also be true because, by Proposition 9.4.15, they differ from exponential coordinates by a polynomial diffeomorphism, which has an upper triangular form.

The proof is based on the BCH formula and (9.21). Let Φ be the exponential coordinate system, and L_g be the left translation by $g \in G$. We need to calculate the differential of $\Phi^{-1} \circ L_g \circ \Phi$. Thus, we consider the diagram

$$\begin{array}{ccc}
\mathbb{R}^n & \xrightarrow{\quad\Phi\quad} & G \\
(t_1, \ldots, t_n) & \stackrel{\Phi}{\longmapsto} & \exp(\sum_{j=1}^n t_j X_j) \\
& & \downarrow L_g \quad \big\downarrow L_g \\
\mathbb{R}^n & \xrightarrow{\quad\Phi\quad} & G \\
(s_1, \ldots, s_n) & \stackrel{\Phi}{\longmapsto} & \exp(\sum_{j=1}^n s_j X_j) = g \exp(\sum_{j=1}^n t_j X_j)
\end{array}$$

and we solve the dependence of the s_i's from the t_j's. Since by Proposition 9.4.15 Malcev coordinate systems are surjective, we can find $u_1, \ldots, u_n \in \mathbb{R}$ and write

$$g = \exp(u_1 X_1) \cdots \exp(u_n X_n).$$

It is enough to consider the case $g = \exp(u_k X_k)$ for an arbitrary $k \in \{1, \ldots, n\}$ and then conclude considering compositions. Thus, we need to consider the system

$$\exp\left(\sum_{j=1}^{n} s_j X_j\right) = \exp(u_k X_k) \exp\left(\sum_{j=1}^{n} t_j X_j\right).$$

By the BCH Formula (Proposition 5.7.4),

$$\sum_{j=1}^{n} s_j X_j = u_k X_k + \sum_{j=1}^{n} t_j X_j + \frac{1}{2}\left[u_k X_k, \sum_{j=1}^{n} t_j X_j\right] + \ldots.$$

Since we have chosen a Malcev basis, we have the property (9.21). Thus a bracket as $[X_k, X_j]$ is only a combination of $\{X_1, \ldots, X_{j-1}\}$. In other words, the function s_j is of the form t_j plus a polynomial that does not depend on the variables $t_1, \ldots, t_j$. Thus, the differential is of the form

$$\mathrm{d}(\Phi^{-1} \circ L_g \circ \Phi) = \begin{bmatrix} 1 & * & \ldots & * \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & * \\ 0 & \ldots & 0 & 1 \end{bmatrix} \in \mathrm{Nil}_n.$$

Thus, the Jacobian of a left translation in exponential coordinates with respect to a Malcev basis is 1 at every point. □

Definition 9.4.18 (Polynomial Coordinates) Let G be a Lie group and $X_1, \ldots, X_n$ a basis of its Lie algebra. If $P : \mathbb{R}^n \to \mathbb{R}^n$ is a diffeomorphism such that P and P^{-1} have polynomial components, then

$$(s_1, \ldots, s_n) \longmapsto \exp\big(P_1(s) X_1 + \cdots + P_n(s) X_n\big)$$

is called *polynomial coordinate system* for G.

Examples of polynomial coordinate maps are, obviously, exponential and, by Proposition 9.4.15, Malcev coordinate maps.

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 coordinates by a polynomial diffeomorphism with a polynomial inverse preserves those maps that preserve the Lebesgue measure.

Corollary 9.4.19 *In nilpotent simply connected Lie groups, left translations and right translations have Jacobian determinant 1 in polynomial coordinates.*

Proof If P is a polynomial map, then its Jacobian determinant $\mathtt{Jac}(P)$ is a polynomial. If P and P^{-1} are polynomial diffeomorphisms, then $1 = \mathtt{Jac}(\mathrm{Id}) = \mathtt{Jac}(P^{-1} \circ P) = (\mathtt{Jac}(P^{-1}) \circ P) \cdot \mathtt{Jac}(P)$.

Hence, the two maps $\mathtt{Jac}(P)$ and $\mathtt{Jac}(P^{-1}) \circ P$ are polynomials whose product is constant. Thus they are constant, and $\mathtt{Jac}(P^{-1})$ is constantly equal to $1/\mathtt{Jac}(P)$.

If Φ is an exponential coordinate map, then

$$\begin{aligned} &\mathtt{Jac}(P^{-1} \circ \Phi^{-1} \circ L_g \circ \Phi \circ P)_x = \\ &\qquad = \mathtt{Jac}(P^{-1})_{(\Phi^{-1} \circ L_g \circ \Phi \circ P)(x)} \cdot \mathtt{Jac}(\Phi^{-1} \circ L_g \circ \Phi)_{P(x)} \cdot \mathtt{Jac}(P)_x \\ &\qquad\qquad = \mathtt{Jac}(\Phi^{-1} \circ L_g \circ \Phi)_{P(x)} = 1, \end{aligned}$$

where we finally used that by Proposition 9.4.17 left-translations in exponential coordinates have Jacobian determinant equal to 1. Similarly, it holds for the right translations. □

Remark 9.4.20 If a map $F : \mathbb{R}^n \to \mathbb{R}^n$ has Jacobian 1, then it preserves the Lebesgue n-measure (because of the change-of-variables formula).

9.4.4.5 Measures in Nilpotent Lie Groups

On nilpotent Lie groups, we have an interpretation in coordinates for the Haar measure; we discussed such a notion in Sect. 6.4. An immediate consequence of Corollary 9.4.19 is the following result.

Corollary 9.4.21 *Let G be an n-dimensional nilpotent simply connected Lie group. Every polynomial coordinate system pushes forward the Lebesgue measure on $\mathbb{R}^n$ to a bi-invariant Haar measure on G.*

It is not always true that left-Haar measures are also right-Haar measures; groups with this property are called *unimodular*. However, in every nilpotent Lie group, Haar measures are both left and right-invariant. Corollary 9.4.21 shows such a property for nilpotent simply connected Lie groups.

To summarize, we provide the proof of Theorem 9.4.7.

Proof of Theorem 9.4.7 In every nilpotent simply connected Lie group, exponential coordinates are global coordinates by Theorem 9.4.6.i, and by the BCH formula, the group product in such coordinates is polynomial. In Proposition 9.4.17, we saw that, in exponential coordinates, left and right translations have Jacobian determinants equal to 1, and so they preserve the Lebesgue measure.

Regarding Malcev coordinates, in Proposition 9.4.15, we saw that they are polynomial coordinates in the sense of Definition 9.4.18; so in particular, they also are global coordinates, and still, the group product in such coordinates is polynomial.

In Corollary 9.4.19, we saw that, in polynomial coordinates, translations have Jacobian determinant equal to 1, and hence they preserve the Lebesgue measure, which is, therefore, a left-Haar measure and a right-Haar measure. □

9.4.5 Structure of Connected Nilpotent Lie Groups

In this section, we discuss connected nilpotent Lie groups that are not necessarily simply connected. Recall that nilpotent Lie algebras can be equipped with the Dynkin product, as in Definition 5.7.3, which is a polynomial group structure by Proposition 9.4.4.

Theorem 9.4.22 *Let G be a connected nilpotent Lie group with Lie algebra $\mathfrak{g}$. Equip $\mathfrak{g}$ with the Dynkin product $\star$. Then, we have the following properties:*

9.4.22.i. *The universal cover $\tilde{G}$ of G is isomorphic to $(\mathfrak{g}, \star)$, and we have Lie algebra isomorphisms $\mathfrak{g} \cong \mathrm{Lie}(\tilde{G}) \cong \mathrm{Lie}(\mathfrak{g}, \star)$. In fact, the exponential map $\exp : \mathfrak{g} \longrightarrow G$ is a covering (the universal covering) and $\exp : (\mathfrak{g}, \star) \longrightarrow G$ is a Lie group homomorphism. In particular, the map $\exp$ is surjective.*

9.4.22.ii. *The center $Z(G)$ equals $\exp(Z(\mathfrak{g}))$, and it is connected.*

9.4.22.iii. *There exists a discrete subgroup $\Gamma < Z(\mathfrak{g})$ such that*

$$G \cong (\mathfrak{g}, \star)/\,\Gamma.$$

In fact, the group Γ is the kernel of $\exp : (\mathfrak{g}, \star) \longrightarrow G$. In particular, the manifold G is diffeomorphic to the abelian Lie group $(\mathfrak{g}, +)/\,\Gamma$, as manifolds.

9.4.22.iv. *The set $\mathfrak{t} := \mathrm{span}(\Gamma) \subseteq Z(\mathfrak{g})$ is a central Lie subalgebra for which $T := \exp(\mathfrak{t})$ is a torus. Finally, the manifold G is diffeomorphic to the product manifold $(G/T) \times T$.*

Proof Let us prove the items.

(i) We saw in Proposition 9.4.4 that $(\mathfrak{g}, \star)$ is a simply connected Lie group with Lie algebra $\mathfrak{g}$. Using Theorem 5.7.2, let $\phi : (\mathfrak{g}, \star) \longrightarrow G$ be the unique Lie group homomorphism such that

$$\begin{array}{ccc} \mathfrak{g} = \mathrm{Lie}(\mathfrak{g}, \star) & \xrightarrow{\ \mathrm{id}\ } & \mathfrak{g} = \mathrm{Lie}(G) \\ \Big\downarrow{\scriptstyle \exp_{(\mathfrak{g},\star)}} & & \Big\downarrow{\scriptstyle \exp_G} \\ (\mathfrak{g}, \star) & \xrightarrow{\ \ \phi\ \ } & G \end{array}.$$

Since we have that $\exp_{(\mathfrak{g},\star)} = \mathrm{id}$, then $\phi = \exp_G$.

(ii) The fact that $\exp(Z(\mathfrak{g})) \subseteq Z(G)$ is a general fact for connected Lie groups (Exercise 9.5.49). Vice versa, if $g \in Z(G)$ since exp is surjective there is $X \in \mathfrak{g}$ such that $g = \exp(X)$. By the central property and Formula 5.5.7, it follows that $\mathrm{Id} = \mathrm{Ad}_g = e^{\mathrm{ad}_X}$. By Proposition 9.1.3.iv, the transformation ad_X is nilpotent. By Proposition 9.3.11, we conclude that $\mathrm{ad}_X = 0$, so $X \in Z(\mathfrak{g})$. Since $Z(\mathfrak{g})$ is a vector subspace, it is connected, and so is $\exp(Z(\mathfrak{g}))$.

(iii) Since $\exp : (\mathfrak{g}, \star) \longrightarrow G$ is a covering of Lie groups, then there exists $\Gamma < (\mathfrak{g}, \star)$ with $\Gamma \cong \pi_1(G)$, such that Γ is discrete and central, and

$$G \cong (\mathfrak{g}, \star)/\Gamma.$$

Indeed, we take $\Gamma := \exp^{-1}(1_G)$. We stress that normal discrete subgroups of connected groups are central (see Exercise 6.6.28). The quotient $(\mathfrak{g}, \star)/\Gamma$ is a Lie group isomorphic to G. However, for $x \in \mathfrak{g}$ and $\gamma \in \Gamma \subseteq Z(\mathfrak{g})$ we have

$$x \star \gamma = x + \gamma,$$

where the extra brackets in Dynkin product are zero since γ is central. Thus, we may identify the manifold $G \cong (\mathfrak{g}, \star)/\Gamma$ with the abelian Lie group $(\mathfrak{g}, +)/\Gamma$, only the group structures may be different.

(iv) We have that Γ is a discrete set that generates $\mathfrak{t} := \mathrm{span}(\Gamma)$. Thus $T := \mathfrak{t}/\Gamma \cong \exp(\mathfrak{t})$ is a torus. Moreover, the subgroup T is central (hence normal), so that $G/T \cong (\mathfrak{g}, \star)/\mathfrak{t}$ is also a Lie group, which is isomorphic to $(\mathfrak{g}/\mathfrak{t}, \star)$. Let $\mathfrak{m} \subseteq \mathfrak{g}$ be a subspace such that $\mathfrak{g} = \mathfrak{m} \oplus \mathfrak{t}$, then for $x \in \mathfrak{m}$ and $z \in \mathfrak{t} \subseteq Z(\mathfrak{g})$ we have $x \star z = x + z$. Thus, the map $\mathfrak{m} \times \mathfrak{t} \longrightarrow \mathfrak{g}$ given by $(x, z) \mapsto x \star z = x + z$ is a diffeomorphism. Hence

$$\begin{aligned} \mathfrak{m} \times \mathfrak{t}/\Gamma &\longrightarrow G \\ (x, z + \Gamma) &\longmapsto x \star (z + \Gamma) = x + z + \Gamma \end{aligned}$$

is a diffeomorphism. Since the quotient map $\mathfrak{g} \longrightarrow \mathfrak{g}/\mathfrak{t}$ induces a linear isomorphism $\mathfrak{m} \longrightarrow \mathfrak{g}/\mathfrak{t}$, we obtain diffeomorphisms

$$G \cong \mathfrak{m} \times \mathfrak{t}/\Gamma \cong G/T \times T.$$

In fact, we conclude that $G \cong (\mathfrak{g}, \star)/\Gamma \cong \mathfrak{m} \times \mathfrak{t}/\Gamma \cong \mathfrak{g}/\mathfrak{t} \times T \cong G/T \times T$. □

Corollary 9.4.23 *Every compact connected nilpotent Lie group is abelian.*

Proof From the previous result, Theorem 9.4.22, each connected nilpotent Lie group G is diffeomorphic to $\mathfrak{g}/\Gamma$ for some $\Gamma \subseteq Z(\mathfrak{g})$. If in addition $\mathfrak{g}/\Gamma$ is compact, then Γ must span $\mathfrak{g}$ and then $\mathfrak{g} \subseteq Z(\mathfrak{g})$. Hence, $\mathfrak{g}$ is abelian, and so is G, recall Exercise 9.5.49. □

Remark 9.4.24 The only abelian connected compact Lie groups are isomorphic as Lie groups to $\mathbb{R}^n/\mathbb{Z}^n =: \mathbb{T}^n$, for $n \in \mathbb{N}$.

9.4.6 Torsion in Nilpotent Lie Groups

Definition 9.4.25 (Torsion) Given a group G and an element $g \in G$, we say that g is a *torsion element* if it has finite order, say $k \in \mathbb{N}$, i.e., $g^k = 1_G$. (Notice that 1_G is a torsion element). The subset $\mathrm{Tor}(G)$ of G consisting of the torsion elements of G is called *torsion* of G. If $\mathrm{Tor}(G) = \{1_G\}$ we say that G is *torsion free* or that it has *no torsion*.

We warn that the set $\mathrm{Tor}(G)$ may not be a subgroup for general groups. The set $\mathrm{Tor}(G)$ is a subgroup when G is nilpotent. We close this section with a characterization: a connected nilpotent Lie group is simply connected if and only if it has no torsion.

Proposition 9.4.26

9.4.26.i. *Nilpotent simply connected Lie groups, and their subgroups, are torsion free.*
9.4.26.ii. *Every connected nilpotent Lie group that has no torsion is simply connected.*

Proof Regarding 9.4.26.i, recall, from Theorem 9.4.6.i, that in every nilpotent simply connected Lie group, the map exp is a diffeomorphism. Assume we have an element expressed in the form $\exp(X)$ and for some $m \in \mathbb{N}$ we have

$$\exp(0) = \exp(X)^m = \exp(mX). \tag{9.22}$$

Then $mX = 0$ and so $X = 0$. Regarding part 9.4.26.ii, by Theorem 9.4.22.iii we have that each connected nilpotent Lie group is of the form $G \cong (\mathfrak{g}, \star)/\Gamma$. Assume by contradiction that Γ is not trivial, and take $\gamma \in \Gamma$ with $\gamma \neq 1$. Thus there exists $X \in \mathfrak{g}$ such that $\gamma = \exp(X)$ and $X \neq 0$. By the fact that Γ is discrete, there exists $k \in \mathbb{N}$ such that $g_k := \exp(\frac{1}{k}X) \notin \Gamma$. Hence, the element g_k has finite order k in G, which is a contradiction. □

9.5 Exercises

Exercise 9.5.1 Let $\mathfrak{g}^{(i)}$ be the i-th element in the lower central series of a Lie algebra $\mathfrak{g}$.

(i) We have $\mathfrak{g}^{(i+1)} \subset \mathfrak{g}^{(i)}$ for all $i \in \mathbb{N}$.
(ii) If $\mathfrak{g}^{(i)} = \mathfrak{g}^{(i+1)}$ for some i, then for all $j > i$ we have $\mathfrak{g}^{(j)} = \mathfrak{g}^{(i)}$.

Exercise 9.5.2 For finite-dimensional Lie algebras, the nilpotency of $\mathfrak{g}$ is equivalent to the vanishing of the ideal $\mathfrak{g}^{(\infty)}(\mathfrak{g}) := \bigcap_{i\in\mathbb{N}} \mathfrak{g}^{(i)}(\mathfrak{g})$.

Exercise 9.5.3 (Upper Central Series of Lie Algebras) Let $\mathfrak{g}$ be a Lie algebra. One iteratively defines the elements of the *upper central series* $(\zeta_i(\mathfrak{g}))_{i\in\mathbb{N}}$ of $\mathfrak{g}$ by

$$\zeta_0(\mathfrak{g}) := \{0_{\mathfrak{g}}\} \quad \text{and} \quad \zeta_{i+1}(\mathfrak{g}) := \{X \in \mathfrak{g} : [X, \mathfrak{g}] \subseteq \zeta_i(v)\}\,.$$

The subset $\zeta_1(\mathfrak{g})$ is the center $Z(\mathfrak{g})$. The sequence $(\zeta_i(\mathfrak{g}))_{i\in\mathbb{N}}$ is weakly increasing, and $[\zeta_i, \mathfrak{g}] \subseteq \zeta_{i-1}$.

Exercise 9.5.4 Given a flag $\mathcal{F} = (V_0, \ldots, V_m)$ for a vector space V, the *set of flag-preserving transformations*

$$\mathfrak{g}(\mathcal{F}) := \left\{A \in \mathfrak{gl}(V) : A(V_k) \subseteq V_k, \forall k \in \{1, \ldots, m\}\right\}$$

is a Lie algebra that is not nilpotent as soon as $\dim V > 1$. Moreover, the subset $\mathfrak{g}_{\mathrm{nil}}(\mathcal{F})$, as defined in Example 9.1.10, is an ideal of $\mathfrak{g}(\mathcal{F})$.

Exercise 9.5.5 The set $\mathfrak{g}_{\mathrm{nil}}(\mathcal{F})$ of flag-shifting transformations for a flag $\mathcal{F} = (V_0, \ldots, V_m)$, as in Example 9.1.10, is a Lie algebra and

$$C^{\ell}(\mathfrak{g}_{\mathrm{nil}}(\mathcal{F}))V_k \subseteq V_{k-\ell}, \qquad \forall k, \ell \in \mathbb{N}.$$

Consequently, we have $C^m(\mathfrak{g}_{\mathrm{nil}}(\mathcal{F})) = \{0\}$, and $\mathfrak{g}_{\mathrm{nil}}(\mathcal{F})$ is nilpotent, with step at most $m - 1$.

Exercise 9.5.6 For $n = 2$, the Lie algebra $\Lambda^1(\mathbb{R}^2) \times \Lambda^2(\mathbb{R}^2)$ from (9.2) gives the first Heisenberg Lie algebra of Example 9.3.7.

Exercise 9.5.7 Every function defined on the set of generators of a free-nilpotent Lie algebra to a nilpotent Lie algebra of no larger step extends uniquely to a Lie algebra homomorphism.

Exercise 9.5.8 Let $A, B \in \mathfrak{gl}(V)$, for some vector space V. Assume that A and B commute. If A and B are nilpotent transformations, then so is $A + B$. If A and B are unipotent transformations, then so is AB.

Exercise 9.5.9 (Real Jordan Theorem) If $A \in \mathfrak{gl}(V)$ for some vector space V, then there is a basis of V such that the matrix representation of A in this basis is in *real Jordan form*, i.e., it is written in blocks

$$\begin{pmatrix} A_1 & 0 & \ldots \\ 0 & A_2 & \ldots \\ \vdots & \vdots & \ddots \end{pmatrix} \tag{9.23}$$

where each A_j is a matrix of one of the following four forms:

$$\begin{pmatrix} \alpha & 0 & & \dots & \\ 0 & \alpha & 0 & \dots & \\ 0 & 0 & \alpha & 0 & \dots \\ \vdots & \vdots & \vdots & \ddots & \end{pmatrix}, \qquad \begin{pmatrix} \alpha & 1 & 0 & \dots & \\ 0 & \alpha & 1 & 0 & \dots \\ 0 & 0 & \alpha & 1 & \dots \\ \vdots & \vdots & \vdots & \ddots & \ddots \end{pmatrix}, \tag{9.24}$$

$$\begin{pmatrix} \alpha & -\beta & 0 & 0 & \dots \\ \beta & \alpha & 0 & 0 & \dots \\ 0 & 0 & \alpha & -\beta & \dots \\ 0 & 0 & \beta & \alpha & \dots \\ & \vdots & & & \ddots \end{pmatrix}, \qquad \begin{pmatrix} \alpha & -\beta & 1 & 0 & 0 & 0 & \dots \\ \beta & \alpha & 0 & 1 & 0 & 0 & \dots \\ 0 & 0 & \alpha & -\beta & 1 & 0 & \dots \\ 0 & 0 & \beta & \alpha & 0 & 1 & \dots \\ & \vdots & & \vdots & & \ddots & \end{pmatrix}. \tag{9.25}$$

The blocks as in (9.24) correspond to a real eigenvalue α, while the blocks as in (9.25) correspond to a complex eigenvalue $\alpha + i\beta$. Consequently, the matrix A is nilpotent if and only if all blocks are upper triangular.

Exercise 9.5.10 Each of the following points implies that if $\mathfrak{a} \subseteq \mathfrak{g}$ is an ideal of a Lie algebra and both $\mathfrak{a}$ and $\mathfrak{g}/\mathfrak{a}$ are nilpotent, then $\mathfrak{g}$ may not be nilpotent.

9.5.10.i. Let $\mathfrak{g}$ be the Lie algebra spanned by two vectors X and Y with relation $[X, Y] = X$. Then, the subset $\mathfrak{h} := \mathbb{R}X$ is a commutative ideal, $\mathfrak{g}/\mathfrak{h}$ is commutative, but $\mathfrak{g}$ is not commutative, nor nilpotent.
9.5.10.ii. Let $\mathfrak{g}$ be a solvable Lie algebra (see Exercise 10.5.21 for the definition). Then, both $[\mathfrak{g}, \mathfrak{g}]$ and $\mathfrak{g}/[\mathfrak{g}, \mathfrak{g}]$ are nilpotent.

Exercise 9.5.11 Every stratification is completely determined by its first stratum V_1.

Exercise 9.5.12 Every compatible linear grading that is also a Lie algebra grading is a stratification.

Exercise 9.5.13 If a Lie algebra $\mathfrak{g}$ has an s-step stratification $\mathfrak{g} = V_1 \oplus \cdots \oplus V_s$, then

9.5.13.i. V_s is contained in the center of $\mathfrak{g}$;
9.5.13.ii. $V_k \oplus \cdots \oplus V_s$ is an ideal in $\mathfrak{g}$;
9.5.13.iii. $(V_k \oplus \cdots \oplus V_s)/(V_{k+1} \oplus \cdots \oplus V_s)$ is contained in the center of $(V_1 \oplus \cdots \oplus V_s)/(V_{k+1} \oplus \cdots \oplus V_s)$.
9.5.13.iv. $\mathfrak{g}/V_s$ is a stratifiable Lie algebra of step $s - 1$.

Exercise 9.5.14 Let $\mathfrak{g}$ be a Lie algebra equipped with a Lie algebra grading $(V_j)_{j\in\mathbb{R}}$. Assume that the elements of degree 1, namely V_1, generate $\mathfrak{g}$ as a Lie algebra, then $\mathfrak{g}$ is stratified by $V_1, \dots, V_s$.

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. Proceed by induction.

Exercise 9.5.15 Let $\mathfrak{h}$ be the Heisenberg Lie algebra generated by the vectors X, Y, and Z with only non-trivial relation $[X, Y] = Z$. Then, the Lie algebra $\mathfrak{h}$ can be stratified as $\mathfrak{h} = \mathrm{span}\{X, Y\} \oplus \mathrm{span}\{Z\}$, and this stratification has step 2.

Exercise 9.5.16 Let $\mathfrak{g} := \mathbb{R} \times \mathfrak{h}$ be the direct product of $\mathbb{R}$ with the (above) Heisenberg Lie algebra $\mathfrak{h}$. Then, the Lie algebra $\mathfrak{g}$ can be stratified as

$$\mathfrak{g} = (\mathbb{R} \times \mathrm{span}\{X, Y\}) \oplus (\{0\} \times \mathrm{span}\{Z\});$$

still, its center is $\mathbb{R} \times \mathrm{span}\{Z\}$, which is strictly bigger than V_2.

Exercise 9.5.17 (A Nontrivial Filiform Algebra) Consider the 6-dimensional Lie algebra $\mathfrak{g}$ given by $\mathrm{span}\{y_0, y_1, y_2, y_3, y_4, y_5\}$ with only non-zero brackets

$$\begin{aligned} [y_0, y_1] &= y_2, & [y_0, y_2] &= y_3, \\ [y_0, y_3] &= y_4, & [y_0, y_4] &= y_5, \\ [y_1, y_4] &= -y_5, & [y_2, y_3] &= y_5. \end{aligned}$$

The following facts hold true:

9.5.17.i. It is a Lie algebra, i.e., the Jacobi identity is satisfied.
9.5.17.ii. It admits a stratification.
9.5.17.iii. It is a filiform algebra (i.e., the dimensions of the subspaces of the stratification are the smallest possible, namely 2, 1, ..., 1).

Exercise 9.5.18 Every two-step nilpotent Lie algebra is stratifiable.

Exercise 9.5.19 (Positively Graded Algebras are Nilpotent) Every Lie algebra that is finite-dimensional and admits a positive grading is nilpotent.
Hint. Let $\mathfrak{g} = \bigoplus_{t>0} V_t = \bigoplus_{a\in[a,b]} V_t$, with $0 < a < b$. Then if $b/a < m \in \mathbb{N}$, then $C^m(\mathfrak{g}) = \{0\}$.

Exercise 9.5.20 For every Lie-algebra stratification $V_1 \oplus \cdots \oplus V_s$, we have $[V_2, V_2] \subseteq V_4$.
Solution. We have

$$\begin{aligned} [V_2, V_2] &= [[V_1, V_1], [V_1, V_1]] = \mathrm{span}\,\{[[X_1, X_2], [X_3, X_4]] : X_i \in V_1\} \subset \\ &\overset{\text{(Jacobi)}}{\subset} \mathrm{span}\,\{[X_1, [X_2, [X_3, X_4]]] : X_i \in V_1\} = [V_1, [V_1, [V_1, V_1]]] = V_4, \end{aligned}$$

where we used the properties of the stratification and Jacobi identity from Definition 5.1.2.

Exercise 9.5.21 (Stratifications are Positive Gradings) Let $\mathfrak{g} = V_1 \oplus \cdots \oplus V_s$ be a stratified Lie algebra. Then, setting $V_k := \{0\}$ for $k > s$, we have

$$[V_i, V_j] \subset V_{i+j}, \qquad \forall i, j \in \{1, \dots, s\}.$$

Solution. The proof is by induction on i. If $i = 1$ we know that $[V_1, V_j] \subset V_{j+1}$ for all j. Suppose that $[V_i, V_j] \subset V_{i+j}$ for all j and a fixed i. We shall check that $[V_{i+1}, V_j] \subset V_{i+1+j}$ for all j. Indeed, the space 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 $v_j \in V_j$ 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, we get $[[v_1, v_i], v_j] \in V_{i+1+j}$ and therefore $[V_{i+1}, V_j] \subset V_{i+1+j}$.

Exercise 9.5.22 (Elements of Lower Central Series in Terms of Stratifications) Let $\mathfrak{g}$ be a Lie algebra with an s-step stratification $\mathfrak{g} = V_1 \oplus \cdots \oplus V_s$. Then, we have

$$\mathfrak{g}^{(k)} = V_k \oplus \cdots \oplus V_s.$$

Solution. The proof is by induction. For $k = 1$, it 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 $\sum_{i=2}^{s}\sum_{j=k}^{s}[V_i, V_j] \subset \sum_{i=2}^{s}\sum_{j=k}^{s} V_{i+j} \subset V_{k+1} \oplus \cdots \oplus V_s$, by Exercise 9.5.21.

Exercise 9.5.23 If a Lie algebra $\mathfrak{g}$ admits an s-step stratification, then $\mathfrak{g}$ is s-step nilpotent.
Hint. Check Exercise 9.5.22.

Exercise 9.5.24 (Positively Gradable, Non-stratifiable Lie Algebra) The Lie algebra $\mathfrak{n}_{5,1}$ from Example 9.1.11 is positively gradable, yet not stratifiable.

Exercise 9.5.25 (Another Nilpotent Non-stratifiable Lie Algebra) Consider the 7-dimensional Lie algebra $\mathfrak{h}$ generated by $X_1, \ldots, X_7$ with only nontrivial brackets

$$\begin{aligned} &[X_1, X_2] = X_3\,, && [X_1, X_3] = 2X_4\,, && [X_1, X_4] = 3X_5\,,\\ &[X_2, X_3] = X_5\,, && [X_1, X_5] = 4X_6\,, && [X_2, X_4] = 2X_6\,,\\ &[X_1, X_6] = 5X_7\,, && [X_2, X_5] = 3X_7\,, && [X_3, X_4] = X_7\,. \end{aligned}$$

The following facts hold true:

9.5.25.i. $\mathfrak{g}$ is a Lie algebra
9.5.25.ii. $\mathfrak{g}$ is nilpotent
9.5.25.iii. $\mathfrak{g}$ admits a grading, with $V_i = \mathbb{R}X_i$.
9.5.25.iv. For every grading of $\mathfrak{g}$, the set V_1 of elements of degree 1 does not generate $\mathfrak{g}$.
9.5.25.v. $\mathfrak{g}$ does not admit any stratification.

Exercise 9.5.26 (Nilpotent Lie Algebras that are Not Positively Gradable) Consider the following 7-dimensional Lie algebras denoted by $\mathfrak{h}_{12457G}$ and $\mathfrak{h}_{12357B}$. They are generated by $X_1, \ldots, X_7$ with only nontrivial brackets

Relations for 12457G	*Relations for* 12357B
$[Y_1, Y_2] = Y_5$	$[Y_1, Y_6] = -Y_2$
$[Y_1, Y_3] = Y_6$	$[Y_1, Y_7] = Y_3 + Y_5$
$[Y_1, Y_4] = Y_6$	$[Y_2, Y_6] = -Y_3$
$[Y_1, Y_7] = -Y_2$	$[Y_2, Y_7] = Y_4$
$[Y_2, Y_3] = -Y_6$	$[Y_3, Y_6] = -Y_4$
$[Y_2, Y_7] = -Y_3$	$[Y_3, Y_7] = Y_5$
$[Y_3, Y_7] = -Y_4$	$[Y_4, Y_6] = -Y_5$
$[Y_5, Y_7] = -Y_6$	

The following facts hold true:

9.5.26.i. Both $\mathfrak{h}_{12457G}$ and $\mathfrak{h}_{12357B}$ are Lie algebras,
9.5.26.ii. They are nilpotent,
9.5.26.iii. ☠ None of them admit a positive grading.

Hint. Check [Hak+22].

Exercise 9.5.27 Fix a positive integer $n \geq 7$, and consider the n-dimensional Lie algebra $\mathfrak{h}$ generated by $X_1, \ldots, X_n$ with

$$[X_i, X_j] = \begin{cases} (j-i)X_{i+j}, & \text{if } i+j \leq n,\\ 0, & \text{otherwise.} \end{cases}$$

The following facts hold true:

9.5.27.i. It is a Lie algebra

9.5.27.ii. It is nilpotent
9.5.27.iii. It does not admit any stratification.

Exercise 9.5.28 (Suggested by E. Breuillard) Let $\mathfrak{g}$ be the quotient of the free-nilpotent Lie algebra of 3-step and generated by e_1, e_2, e_3 modulo the ideal generated by $[e_2, e_3]$, so we have the additional relation $[e_2, e_3] = 0$. The Lie algebra $\mathfrak{g}$ has dimension 10. Moreover, the following is a stratification of $\mathfrak{g}$:

$$\begin{aligned} V_1 &:= \operatorname{span}\{e_1, e_2, e_3\}, \\ V_2 &:= \operatorname{span}\{[e_1, e_2], \quad [e_1, e_3]\}, \\ V_3 &:= \operatorname{span}\{e_{112}, \quad e_{212}, \quad e_{312}, \quad e_{331}, \quad e_{113}\}, \end{aligned}$$

with the notation $e_{ijk} := [e_i, [e_j, e_k]]$. The subspace $V_1' := \operatorname{span}\{e_1, e_2 + [e_1, e_2], e_3\}$ is not the first layer of a stratification. Still, it is in direct sum with $[\mathfrak{g}, \mathfrak{g}]$.
Hint. The space $[V_1', V_1']$ is not in direct sum with $[\mathfrak{g}, [\mathfrak{g}, \mathfrak{g}]]$, 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 9.5.29 Let $(V_a)_{a\in\mathbb{R}}$ be a Lie algebra grading of a Lie algebra $\mathfrak{g}$. Then, the subspace $V_{>0} := \bigoplus_{a>0} V_a$ is a subalgebra. Moreover, if the grading is *nonnegative*, in the sense that $\mathfrak{g} = \bigoplus_{a\geq 0} V_a$, then $V_{>0}$ is an ideal.
Solution. If $a \geq 0$ and $b > 0$, then obviously $a + b > 0$, and so $[V_a, V_b] \subseteq V_{a+b} \subseteq V_{>0}$.

Exercise 9.5.30 Every one-parameter subgroup of Lie algebra automorphisms leads to a Lie algebra grading as in Proposition 9.2.14.
Hint. Use Propositions 5.6.3.ii and 9.2.19.

Exercise 9.5.31 Let $\mathfrak{g} = V_1 \oplus \cdots \oplus V_s$ be a stratified Lie algebra. For all $\lambda \geq 0$, let δ_λ be the dilation of factor λ as defined in Definition 9.2.12. Then, we have

$$\delta_\lambda \left(\sum_{i=1}^{s} v_i \right) = \sum_{i=1}^{s} \lambda^i v_i,$$

where $X = \sum_{i=1}^{s} v_i$ with $v_i \in V_i$, $1 \leq i \leq s$.

Exercise 9.5.32 (Nilpotency and Compatible Linear Gradings) Let $\mathfrak{g}$ be a finite-dimensional Lie algebra. We have that $\mathfrak{g}$ is nilpotent if and only if it admits a compatible linear grading. Moreover, if $\mathfrak{g} = \bigoplus_{i=1}^{\infty} V_i$ is a compatible linear grading, then

$$[V_i, V_j] \subseteq [\mathfrak{g}^{(i)}, \mathfrak{g}^{(j)}] \subseteq \mathfrak{g}^{(i+j)} = \bigoplus_{k=i+j}^{\infty} V_k.$$

Exercise 9.5.33 Every stratifiable Lie algebra is isomorphic to its associated Carnot algebra (as defined in Definition 9.2.16).

Exercise 9.5.34 If $\mathfrak{g}$ is a finite-dimensional positively graded Lie algebra, then for each $\lambda \in (0, 1)$ the dilation δ_λ is contractive. If $\mathfrak{g}$ is $\mathbb{N}$-graded, then the dilation δ_λ is contractive for $\lambda \in (-1, 1)$.
Hint. We have $\delta_\lambda^n(v) = \lambda^{jn} v \to 0$, as $n \to \infty$, for every v of degree j.

Exercise 9.5.35 Let $\delta : G \to G$ be an automorphism of a Lie group. Then, the sequence of maps $(\delta^n)_{n\in\mathbb{N}}$ pointwise converges to the constant map 1_G if and only if it converges uniformly on compact sets to 1_G.
Hint. Sequences of linear maps that converge on the elements of a basis converge uniformly on compact sets.

Exercise 9.5.36 Let G be a connected Lie group with Lie algebra $\mathfrak{g}$. Let $\delta : G \to G$ be a Lie group automorphism with induced Lie algebra automorphism $\delta_* : \mathfrak{g} \to \mathfrak{g}$. Then, the map δ is contractive if and only if so is δ_*. (Recall Proposition 5.2.9)

Exercise 9.5.37 If a Lie group G admits a contractive automorphism, then G is simply connected.
Hint. Manifolds are locally simply connected.

Exercise 9.5.38 If $\phi \in \mathrm{Aut}_{\mathbb{C}}(\mathfrak{g}_{\mathbb{C}})$ and $\alpha, \beta \in \mathbb{C}$, then for the E_α^ϕ's defined in (9.11) we have

$$[E_\alpha^\phi, E_\beta^\phi] \subset E_{\alpha\beta}^\phi, \qquad \forall \alpha, \beta \in \mathbb{C}.$$

Hint. The proof is elementary after one proves by induction on $n \in \mathbb{N}$ that

$$(\phi - \alpha\beta\mathbb{I})^n[v, w] = \sum_{\substack{j+k=n \\ j,k\geq 0}} \binom{n}{j} [\alpha^k(\phi - \alpha\mathbb{I})^j v, \phi^j(\phi - \beta\mathbb{I})^k w]$$

holds for all $v, w \in \mathfrak{g}_{\mathbb{C}}$, all $\alpha, \beta \in \mathbb{C}$ and all $n \in \mathbb{N}$. See also [Bou05, Ch.7, sec 1.4, Prop.12, p. 11].

Exercise 9.5.39 Let $\mathfrak{g}$ be a $\mathbb{Z}$-graded Lie algebra.

9.5.39.i. The Lie algebra $\mathfrak{g} \rtimes \mathbb{R}$ is $\mathbb{Z}$-graded by (9.15).
9.5.39.ii. If $V_0 = \{0\}$ and $\mathfrak{g} \neq \{0\}$, then $\mathfrak{g} \rtimes \mathbb{R}$ has trivial center.
9.5.39.iii. If $\mathfrak{g} = \{0\}$, then $\mathfrak{g} \rtimes \mathbb{R} = \mathbb{R}$, which is abelian.
9.5.39.iv. If $\mathfrak{g}$ is Carnot of step s, then the grading of $\mathfrak{g} \rtimes \mathbb{R}$ satisfies

$$V'_m \neq \{0\} \quad \Leftrightarrow \quad m \in \mathbb{Z} \cap [0, s].$$

Solution.

(i) It is enough to check that

$$[\{0\} \times \mathbb{R}, \{0\} \times \mathbb{R}] \subset V_0', \quad \text{and} \quad [V_m \times \{0\}, \{0\} \times \mathbb{R}] \subset V_m'.$$

In fact, one has $[(0, s), (0, t)] = (0, 0) \in V_0'$ and $[(X, 0), (0, t)] = (-tmX, 0) \in V_m'$ if $X \in V_m$.

(ii) Let $(X, s) \in Z(\mathfrak{g} \rtimes \mathbb{R})$. Write X as $\sum_{m\in\mathbb{Z}} X_m$, with $X_m \in V_m$. Then $(0, 0) = [(X, s), (0, 1)] = (-\sum_{m\in\mathbb{Z}} mX_m, 0)$. Hence, for every $m \neq 0$ we have that $X_m = 0$ and so $X = 0$ since $V_0 = \{0\}$.
Moreover, take $Y \in V_m \setminus \{0\}$ for some $m \in \mathbb{Z} \setminus \{0\}$, which exists since $\mathfrak{g}$ is not trivial. Then $(0, 0) = [(X, 0), (Y, 0)] = [(0, s), (Y, 0)] = (smY, 0)$ and consequently $s = 0$.

(iii) It trivially follows from the basic Definition 5.6.4.

(iv) From the definition of V_a', if the non-trivial layers of the gradings of $\mathfrak{g}$ are $V_1, \ldots, V_s$, then the non-trivial layers of the gradings of $\mathfrak{g} \rtimes \mathbb{R}$ are $V_0', V_1', \ldots, V_s'$.

Exercise 9.5.40 For each group G, the elements $\zeta_i := \zeta_i(G)$ of the upper central series are normal in G, satisfy $[\zeta_i, G] \subseteq \zeta_{i-1}$, and $\zeta_{i+1}/\zeta_i = Z(G/\zeta_i)$.

Exercise 9.5.41 The Lie group Nil_n is nilpotent and simply connected with Lie algebra $\mathfrak{nil}_n$.
Hint. Write the obvious diffeomorphism with $\mathbb{R}^{n(n-1)/2}$.

Exercise 9.5.42 The group Nil_3, as from Example 9.3.4, is the Heisenberg group, as defined in Sect. 2.3.

Exercise 9.5.43 (The Maps log and exp are the Inverse of Each Others, Locally)

For all $n \in \mathbb{N}$, we have the following properties:

9.5.43.i. For every $A \in \mathfrak{gl}(n)$ with $\|A\| < \log 2$ we have that $\log(e^A) = A$.
9.5.43.ii. For every $M \in \mathrm{GL}(\mathbb{R}^n)$ with $\|M - \mathbb{I}\| < 1$ we have that $e^{\log M} = M$.

Hint. Check [HN12, Proposition 3.3.2].

Exercise 9.5.44 For $\mathcal{N}$ as defined in (9.16), the set $\mathrm{span}(\mathcal{N})$ consist of the matrices in $\mathfrak{gl}(n)$ that are 0 in the diagonal. In particular, for all $n > 1$, some of them are not in $\mathcal{N}$.

Exercise 9.5.45 Let V be a finite-dimensional vector space, $v \in V$, and $A \in \mathfrak{gl}(V)$ a nilpotent transformation. We have that $Av = 0$ if and only if $e^A v = v$.
Solution. One direction is true even when A is not nilpotent. For the other one, if $(e^A - \mathbb{I})v = 0$ then $Av = \log e^A v = \sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k}(e^A - \mathbb{I})^k v = 0$.

Exercise 9.5.46 On every nilpotent Lie algebra the Dynkin product $(A, B) \in \mathfrak{g} \mapsto A \star B$ is a polynomial map.

Exercise 9.5.47 In Nil_n, the group product in exponential coordinates is polynomial: $(A, B) \mapsto \log(e^A e^B)$.

Exercise 9.5.48 Let $A, B \in \mathfrak{gl}(V)$ be nilpotent transformations of a vector space V. We have that $[A, B] = 0$ if and only if $e^A e^B = e^B e^A$.

Exercise 9.5.49 Let G be a Lie group with Lie algebra $\mathfrak{g}$. Then, the center $Z(G)$ is a closed Lie subgroup with $\mathrm{Lie}(Z(G)) = Z(\mathfrak{g})$. Consequently, if G is connected then G is commutative if and only if $\mathfrak{g}$ is commutative.

Exercise 9.5.50 If $[G, G] = \{1_G\}$, then $[\mathfrak{g}, \mathfrak{g}] = \{0\}$.

Exercise 9.5.51 If $[G, [G, G]] = \{1_G\}$, then $[\mathfrak{g}, [\mathfrak{g}, \mathfrak{g}]] = \{0\}$.
Hint. Use that $[e^{tX+o(t)}, e^{tY+o(t)}] = e^{t^2[X,Y]+o(t^2)}$, as $t \to 0$, for every $X, Y \in \mathfrak{g}$.

Exercise 9.5.52 Let G be a nilpotent simply connected Lie group with Lie algebra $\mathfrak{g}$. The two lower central series, from Definition 9.1.1 and Definition 9.3.1, satisfy

$$C^m(G) = \exp(C^m(\mathfrak{g})), \qquad \forall m \in \mathbb{N}.$$

We deduce that $\mathrm{step}(G) = \mathrm{step}(\mathfrak{g})$.

Exercise 9.5.53 Every connected Lie subgroup of every nilpotent simply connected Lie group is closed and simply connected.
Hint. Check Proposition 9.4.1 and Theorem 9.4.6.iii.

Exercise 9.5.54 Every nilpotent simply connected Lie group has an embedding as a closed subgroup of Nil_n, for some $n \in \mathbb{N}$.
Hint. Check Theorem 9.4.6.iii, Proposition 9.4.1, and Theorem 5.3.4.

Exercise 9.5.55 Every positively graded Lie algebra admits a Malcev basis. In fact, every ordered basis such that the reversed-ordered basis is adapted to the grading is a Malcev basis.
Solution. The solution is a variation of the one from Proposition 9.4.12. Let $\mathfrak{g} = \bigoplus_{a>0} V_a$, be a positively graded Lie algebra. Let $(X_1, \dots, X_n)$ be an ordered basis of $\mathfrak{g}$ such that the reversed-ordered basis $(X_n, \dots, X_1)$ is adapted to the direct sum. We claim that $(X_1, \dots, X_n)$ is a Malcev basis. Indeed, set $\mathfrak{g}_k := \mathrm{span}\{X_1, \dots, X_k\}$. Thus $X_k \in V_b$ for some $b > 0$, then

$$V_{>b} \subset \mathfrak{g}_k \subset V_{\geq b}.$$

Because we have a positive grading, we conclude $[\mathfrak{g}, \mathfrak{g}_k] \subset [\bigoplus_{a>0} V_a, V_{\geq b}] \subset V_{>b} \subset \mathfrak{g}_k$.

Exercise 9.5.56 For each prime number p, the group

$$\left\{ \begin{bmatrix} 1 & x & z \\ 0 & 1 & y \\ 0 & 0 & 1 \end{bmatrix} : x, y, z \in \mathbb{Z}/p\mathbb{Z} \right\}$$

is not a subgroup of a connected nilpotent Lie group.

Exercise 9.5.57 Nilpotent simply connected Lie groups do not have nontrivial compact subgroups.

Exercise 9.5.58 If N is a nilpotent connected Lie group and K is a compact subgroup of N, then K is central.
Hint. Take the universal cover, use Theorem 9.4.22, and Exercise 9.5.57.

Chapter 10
Metrics on Nilpotent Groups

In this chapter, we consider metrics on nilpotent Lie groups. In Sect. 10.1, we consider Carnot-Carathéodory metrics. In Sect. 10.2, we complete the description of isometrically homogeneous spaces with dilations, which we started in Sect. 6.5. In Sect. 10.3, we briefly mention results on the isometry group of nilpotent Lie groups. In Sect. 10.4, we discuss seminorms on nilpotent Lie groups and show some preliminary results useful for then obtaining Pansu Asymptotic Theorem in Sect. 12.4.

10.1 Sub-Finsler Nilpotent Groups

We will now turn our attention to nilpotent simply connected Lie groups equipped with sub-Finsler metrics. The notion of sub-Finsler Lie group was introduced in Chap. 7. In Chap. 9, we discussed the theory of nilpotent Lie groups.

Let G be a Lie group, which we will assume to be simply connected and nilpotent in this chapter. Let $V \subseteq T_1 G$ be a vector subspace of the tangent space at the identity element 1_G of G. Let Δ be the left-invariant distribution on G with $\Delta_1 = V$. We shall consider bracket-generating distributions. Fixing a norm $\|\cdot\|$ on V, we extend it to a left-invariant norm on Δ. Then, the triple $(G, \Delta, \|\cdot\|)$ is a sub-Finsler manifold, and we denote by d_{sF} its sub-Finsler metric, which is left-invariant by construction.

The main aim of this section is to show that nilpotent simply connected sub-Finsler groups have a special submetry map: the projection on the abelianization. Through this map, we will have a new viewpoint for developments and multiplicative integrals.

E. Le Donne, *Metric Lie Groups*, Graduate Texts in Mathematics 306,
https://doi.org/10.1007/978-3-031-98832-5_10

10.1.1 A Submetry Onto the Abelianization

10.1.1.1 The Abelianization of Nilpotent Simply Connected Lie Groups

Assume that G is a Lie group that is nilpotent and simply connected, whose Lie algebra is denoted by $\mathfrak{g}$. In this case, the subgroup $[G, G]$ is connected, normal, and closed; see Exercise 9.5.53. Thus, the quotient $G/[G, G]$ is a Lie group that is abelian and simply connected; see Exercise 10.5.1. Hence, it is a vector space, and it can be identified with its Lie algebra, which is naturally identified with $\mathfrak{g}/[\mathfrak{g}, \mathfrak{g}]$.

The group $\mathrm{Ab}(G) := G/[G, G]$ is called the *abelianization* of the group G, while $\mathfrak{g}/[\mathfrak{g}, \mathfrak{g}]$ is called the *abelianization* of the Lie algebra $\mathfrak{g}$. When G is a nilpotent simply connected Lie group, the Lie algebra of $\mathrm{Ab}(G)$ is isomorphic to $\mathfrak{g}/[\mathfrak{g}, \mathfrak{g}]$ via the diagram

$$\begin{array}{ccc} \mathfrak{g} & \xrightarrow{\exp} & G \\ \downarrow & & \downarrow \\ \mathrm{Ab}(G) \simeq \mathfrak{g}/[\mathfrak{g},\mathfrak{g}] & \xrightarrow{\exp} & G/[G,G] =: \mathrm{Ab}(G), \end{array}$$

which is commutative; see Exercise 10.5.2. In addition, recall from Theorem 9.4.6.i that the exponential map is a diffeomorphism. For this reason, we shall identify the two viewpoints for the abelianization and denote by π_{ab} either of the projections:

$$\pi_{\mathrm{ab}} : \mathfrak{g} \twoheadrightarrow \mathrm{Ab}(G) \quad \text{or} \quad \pi_{\mathrm{ab}} : G \twoheadrightarrow \mathrm{Ab}(G),$$

which we call *abelianization maps*.

10.1.1.2 A Sub-Finsler Submetry on the Abelianization

In addition to G being a nilpotent simply connected Lie group, we next assume that it is equipped with a left-invariant sub-Finsler structure $(\Delta, \|\cdot\|)$. Thus, the triple $(G, \Delta, \|\cdot\|)$ is a sub-Finsler Lie group, as in Sect. 7.1.2. We stress that $V := \Delta_{1_G}$ is assumed to be a bracket-generating subspace of $\mathfrak{g}$ and, if we project it via the abelianization map $\pi_{\mathrm{ab}} : \mathfrak{g} \to \mathrm{Ab}(G)$, we claim that

$$\pi_{\mathrm{ab}}(V) = \mathrm{Ab}(G).$$

Indeed, the set $\pi_{\mathrm{ab}}(V) \subseteq \mathfrak{g}/[\mathfrak{g}, \mathfrak{g}]$ is a bracket-generating subspace of an abelian Lie algebra, so it is necessarily equal to the whole Lie algebra, recall Exercise 7.5.7.

Definition 10.1.1 (Abelianization Norm) Let $(G, \Delta, \|\cdot\|)$ be a sub-Finsler nilpotent simply connected Lie group. Let $\|\cdot\|_{ab}$ be the norm on the abelianization $\mathrm{Ab}(G)$ whose unit ball at the origin is the image in $\mathrm{Ab}(G)$ under π_{ab} of the unit ball at the origin for $\|\cdot\|$ in Δ_{1_G}, that is,

$$\pi_{ab}\left(B_{(\Delta_{1_G},\|\cdot\|)}(0,1)\right) = B_{(\mathrm{Ab}(G),\|\cdot\|_{ab})}(0,1). \tag{10.1}$$

For another expression of the norm $\|\cdot\|_{ab}$ see Exercise 10.5.3. Notice that $\|\cdot\|_{ab}$ is actually the only norm that makes the projection $\pi_{ab}|_{\Delta_{1_G}} : \Delta_{1_G} \subseteq \mathfrak{g} \to \mathrm{Ab}(G)$ a submetry, in the sense of Definition 3.1.23, where on Δ_{1_G} we consider the distance induced by $\|\cdot\|$ and on $\mathrm{Ab}(G)$ the distance induced by $\|\cdot\|_{ab}$. We shall actually show that we obtained a submetry between sub-Finsler Lie groups.

Proposition 10.1.2 (π_{ab} is a Submetry) *Let G be a nilpotent simply connected Lie group metrized with a left-invariant sub-Finsler metric. Equip the abelianization* $\mathrm{Ab}(G)$ *with the norm* $\|\cdot\|_{ab}$ *from Definition 10.1.1. Then* $(\mathrm{Ab}(G), \|\cdot\|_{ab})$ *is a normed vector space and the abelianization map* $\pi_{ab} : G \to \mathrm{Ab}(G)$ *becomes a group homomorphism and a submetry.*

Proof Being the abelianization map the projection modulo the normal subgroup $[G, G]$, the map $\pi := \pi_{ab}$ is a group homomorphism; see also Exercise 10.5.4. As we explained at the beginning of the section, the group $\mathrm{Ab}(G)$ is a vector space equipped with the norm from Definition 10.1.1.

Regarding the fact that π is a submetry, this is a consequence of the more general Proposition 7.1.9. Since, actually, in this case, the proof is more elementary, we directly check that π satisfies the condition in Definition 3.1.23. Fix $r > 0$ and $g \in G$. Let $\bar{B}_{sF}(g, r)$ be the closed ball with respect to the sub-Finsler distance d_{sF} on G with center g and radius r. We need to show that $\pi\left(\bar{B}_{sF}(g, r)\right) = \bar{B}_{\|\cdot\|_{ab}}(\pi(g), r)$. From the left invariance of d_{sF}, we may assume that $g = 1_G$. By definition of $\|\cdot\|_{ab}$, we have $\|\pi(X)\|_{ab} \le \|X\|$ for every $X \in \Delta_{1_G}$; see Exercise 10.5.3. Integrating this inequality for the velocities along each horizontal path γ, it follows that

$$\left\|(\pi\circ\gamma)'\right\|_{ab} = \|\pi(\dot{\gamma})\|_{ab} = \left\|\pi(\gamma')\right\|_{ab} \le \left\|\gamma'\right\| = \|\dot{\gamma}\|,$$

where we use that π is a homomorphism and that the norms are left-invariant (and we also used the notation (7.3)). Therefore, the map π does not increase distances and we have one inclusion: $\pi(\bar{B}_{sF}(1_G, r)) \subset \bar{B}_{\|\cdot\|_{ab}}(0, r)$.

Regarding the other inclusion, take $Y \in \mathrm{Ab}(G)$ satisfying $\|Y\|_{ab} \le r$, then, by definition of the norm $\|\cdot\|_{ab}$, there exists $X \in \Delta_{1_G}$, such that $\pi(X) = Y$ and $\|X\| \le r$. The curve $t \in [0, 1] \mapsto \exp(tX)$ is a horizontal path connecting 1_G and $\exp(X)$ with length at most r. Hence, we have $d_{sF}(1_G, \exp(X)) \le r$. Finally $\pi(\exp(X)) = \pi(X) = Y$, so we have proved the opposite inclusion: $\bar{B}_{\|\cdot\|_{ab}}(0, r) \subset \pi(\bar{B}_{sF}(1_G, r))$. □

10.1.1.3 Lifts of Curves

Proposition 10.1.3 *Let G be a nilpotent simply connected Lie group metrized with a left-invariant sub-Finsler metric. Equip the abelianization* $\mathrm{Ab}(G)$ *with the norm* $\|\cdot\|_{\mathrm{ab}}$ *from Definition 10.1.1. Then for every $T > 0$ and every $w : [0, T] \to \mathrm{Ab}(G)$ measurable with $\|w\|_{\mathrm{ab}} = 1$ almost everywhere, there exists a horizontal curve $\gamma : [0, T] \to G$ with $\gamma(0) = 1_G$, $\pi_{\mathrm{ab}}(\dot\gamma) = w$, and $\|\dot\gamma\| = 1$, almost everywhere in* $[0, T]$.

Proof We get this existence result as a consequence of the fact that π is a submetry. Let $\sigma : [0, T] \to \mathrm{Ab}(G)$ be the absolutely continuous curve with $\sigma(0) = 0$ and $\sigma' = w$. Notice that since $\mathrm{Ab}(G)$ is a vector space, we just have $\sigma(t) := \int_0^t w(s)\,\mathrm{d}s$, and moreover the speed is $\|\dot\sigma\|_{\mathrm{ab}} = \|w\|_{\mathrm{ab}} = 1$, so the curve σ is 1-Lipschitz, recalling (3.14). The curve γ for which we are looking will be a lift of σ under the abelianization map $\pi := \pi_{\mathrm{ab}}$.

In Proposition 10.1.2, we saw that $\pi : G \to \mathrm{Ab}(G)$ is a submetry, in addition to a group homomorphism. Hence, via Proposition 3.1.24 we can lift σ. Namely, there is a 1-Lipschitz curve $\gamma : [0, T] \to G$ such that $\gamma(0) = 1_G$ and $\pi \circ \gamma = \sigma$.

First, we observe that $\pi(\dot\gamma) = \dot\sigma = \sigma' = w$, where we recall that we are identifying π with its differential, and using the notation (7.3). Second, we notice that, on the one hand, since γ is 1-Lipschitz, then $\|\dot\gamma\| \le 1$; see Exercise 4.4.11. On the other hand, since π is 1-Lipschitz (being a submetry), and $\|\pi(\dot\gamma)\|_{\mathrm{ab}} = \|\dot\sigma\|_{\mathrm{ab}} = 1$, then $\|\dot\gamma\| \ge 1$. Thus, we conclude that $\|\dot\gamma\| = 1$. □

The proof of Proposition 10.1.3 actually tells us that we can lift rectifiable curves from the abelianization to the group. We leave it to the reader to write the variation of the proof of the following consequence; see Exercise 10.5.5.

Corollary 10.1.4 *Let G be a nilpotent simply connected Lie group metrized with a left-invariant sub-Finsler metric. Equip the abelianization* $\mathrm{Ab}(G)$ *with the norm* $\|\cdot\|_{\mathrm{ab}}$ *from Definition 10.1.1. Then for every $T > 0$ and every AC curve $\sigma : [0, T] \to \mathrm{Ab}(G)$ with bounded speed, there exists a horizontal curve $\gamma : [0, T] \to G$ with $\pi_{\mathrm{ab}}(\gamma(0)) = \sigma(0)$, $\pi_{\mathrm{ab}} \circ \gamma = \sigma$, and $\|\dot\gamma\| = \|\dot\sigma\|_{\mathrm{ab}}$, almost everywhere in* $[0, T]$.

We picture Corollary 10.1.4 by the following diagram:

$$\begin{array}{ccc} & & G \\ & \overset{\gamma}{\nearrow} & \big\downarrow \pi_{\mathrm{ab}} \\ [0,T] & \xrightarrow{\ \sigma\ } & \mathrm{Ab}(G). \end{array}$$

The abelianization map behaves naturally under homomorphisms:

Lemma 10.1.5 *Let G and H be nilpotent simply connected Lie groups. Let $\varphi : G \to H$ be a Lie group homomorphism. There exists a unique linear homomorphism $\varphi_{\mathrm{ab}} : \mathrm{Ab}(G) \to \mathrm{Ab}(H)$ for which the following diagram commutes:*

$$\begin{array}{ccc} G & \xrightarrow{\varphi} & H \\ \downarrow{\scriptstyle \pi_{\mathrm{ab}}} & & \downarrow{\scriptstyle \pi_{\mathrm{ab}}} \\ \mathrm{Ab}(G) & \xrightarrow{\varphi_{\mathrm{ab}}} & \mathrm{Ab}(H). \end{array}$$

Moreover, if φ_{ab} is surjective, then so is φ.

Proof The first statement is completely algebraic. For the diagram to commute, we need to define

$$\varphi_{\mathrm{ab}}(g[G, G]) := \varphi(g)[H, H], \qquad \forall g \in G.$$

Since $\varphi([G, G]) \subseteq [H, H]$, then the map is well defined and is a homomorphism since $[G, G]$ and $[H, H]$ are normal subgroups.

If in addition φ_{ab} is surjective, then we have $\mathfrak{h} = \varphi_*(\mathfrak{g}) + [\mathfrak{h}, \mathfrak{h}]$. Hence, the Lie algebra $\varphi_*(\mathfrak{g})$ is bracket generating in $\mathfrak{h}$, by Exercise 10.5.6. Hence $\varphi_*(\mathfrak{g}) = \mathfrak{h}$, and hence $\varphi(G) = H$. □

10.1.2 A Special Sub-Finsler Geometry on Nilpotent Groups

In this subsection, we consider particular Carnot-Carathéodory metrics on nilpotent Lie groups: Given a nilpotent simply connected Lie group G with Lie algebra $\mathfrak{g}$, we consider left-invariant distributions Δ with the property that

$$\mathfrak{g} = \Delta_1 \oplus [\mathfrak{g}, \mathfrak{g}]. \tag{10.2}$$

This will be the case in the metric groups of our main interest: Carnot groups.

Such a Δ_1 is not unique, not even up to isomorphism; see Exercise 10.5.9. However, every Δ_1 satisfying (10.2) Lie generates the Lie algebra $\mathfrak{g}$, recall Exercise 10.5.6. Moreover, no smaller subspace would Lie generate; see Exercise 10.5.7. Thus, these distributions may be called *minimal bracket-generating polarizations*.

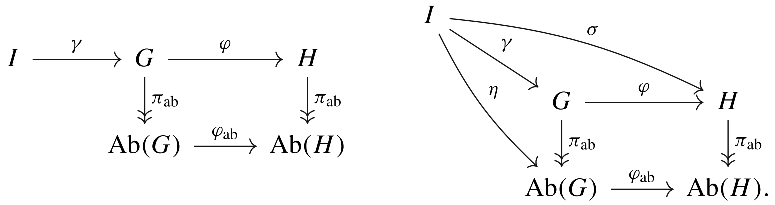
Next, we shall see how one can check via abelianizations whether a group homomorphism is Lipschitz. In connection with the next part 10.1.6.i, we also point out that there is a more general reason why the following map is Lipschitz; see Exercise 7.5.6.

Proposition 10.1.6 (Criterion for Lipschitz or Submetry Property) *Let G and H be nilpotent simply connected Lie groups equipped with left-invariant sub-Finsler* *structures:* $(G, \Delta^G, \|\cdot\|)$ *and* $(H, \Delta^H, \|\cdot\|)$. *Assume that* $\mathfrak{h} = \Delta_1^H \oplus [\mathfrak{h}, \mathfrak{h}]$. *Let* $\varphi : G \to H$ *be a Lie group homomorphism.*

10.1.6.i. *If* $\varphi_*(\Delta_1^G) \subseteq \Delta_1^H$, *then* φ *is Lipschitz with respect to the sub-Finsler distances. In fact, we have* $\mathtt{Lip}(\varphi) = \mathtt{Lip}(\varphi_{ab})$.

10.1.6.ii. *If* $\varphi_*(\Delta_1^G) \subseteq \Delta_1^H$ *and* $\varphi_{ab} : \mathrm{Ab}(G) \to \mathrm{Ab}(H)$ *is a submetry, then* φ *is a submetry.*

Proof Regarding 10.1.6.i, the fact that φ is Lipschitz is a more general fact; see Exercise 7.5.6. We prove that, in this case, the Lipschitz constant equals $\mathtt{Lip}(\varphi_{ab})$. Let γ be a Δ^G-horizontal curve, parametrized with speed 1. Because of the assumption, the curve $\varphi \circ \gamma$ is Δ^H-horizontal. Since $\pi := \pi_{ab} : G \to \mathrm{Ab}(G)$ is a submetry (and hence 1-Lipschitz), the curve $\pi \circ \gamma$ is 1-Lipschitz. By Lemma 10.1.5, the map φ_{ab} exists, and, being linear, it is Lipschitz. Let L be the Lipschitz constant of φ_{ab}. Then $\varphi_{ab} \circ \pi \circ \gamma$ is L-Lipschitz. The curve $\varphi \circ \gamma$ is a lift of $\varphi_{ab} \circ \pi \circ \gamma = \pi \circ \varphi \circ \gamma$, and, since Δ^H is in direct sum with $[\mathfrak{h}, \mathfrak{h}]$, this is the only horizontal lift; check the first commutative diagram (10.3). Hence, by Corollary 10.1.4, the curves $\varphi \circ \gamma$ and $\varphi_{ab} \circ \pi \circ \gamma$ have the same speeds so, in particular, we deduce that $\varphi \circ \gamma$ is L-Lipschitz. Since the distances are length distances, we infer that φ is L-Lipschitz.



(10.3)

Regarding 10.1.6.ii, the map φ is 1-Lipschitz from 10.1.6.i. To deduce that φ is a submetry, we can just show that for every absolutely continuous curve $\sigma : I \to H$ with $\|\dot{\sigma}\| \equiv 1$ there is an absolutely continuous curve $\gamma : I \to G$ with $\|\dot{\gamma}\| \equiv 1$ and $\varphi \circ \gamma = \sigma$. By Corollary 10.1.4, since Δ_1^H is in direct sum with $[\mathfrak{h}, \mathfrak{h}]$, the curve $\pi_{ab} \circ \sigma : I \to \mathrm{Ab}(H)$ is such that $\left\|(\pi_{ab} \circ \sigma)'\right\| \equiv 1$. Similarly, since φ_{ab} is a submetry by assumption, there is $\eta : I \to \mathrm{Ab}(G)$ such that $\varphi_{ab} \circ \eta = \pi_{ab} \circ \sigma$ and $\|\dot{\eta}\| \equiv 1$. Since $\pi_{ab} : G \to \mathrm{Ab}(G)$ is a submetry, there exists $\gamma : I \to G$ with $\|\dot{\gamma}\| \equiv 1$ and $\pi_{ab} \circ \gamma = \eta$. We conclude by checking that $\varphi \circ \gamma = \sigma$. In fact, the curves $\varphi \circ \gamma$ and σ are equal because they are the unique horizontal lift of $\pi_{ab} \circ \sigma$, because

$$\pi_{ab} \circ \varphi \circ \gamma = \varphi_{ab} \circ \pi_{ab} \circ \gamma = \varphi_{ab} \circ \eta = \pi_{ab} \circ \sigma.$$

In other words, we showed the commutativity of the diagrams in (10.3). □

10.1.2.1 Projection on the Good Polarization

Definition 10.1.7 (The Projection π_{Δ_1}) Under the assumption that the polarization Δ_1 is in direct sum with $[\mathfrak{g}, \mathfrak{g}]$, i.e., (10.2), let $\mathrm{proj}_{\Delta_1} : \mathfrak{g} \to \Delta_1$ be the projection onto Δ_1 with kernel $[\mathfrak{g}, \mathfrak{g}]$ and define

$$\begin{aligned} \pi_{\Delta_1} : G &\to \Delta_1 \\ g \mapsto \;& \pi_{\Delta_1}(p) := \mathrm{proj}_{\Delta_1}(\exp^{-1}(g)). \end{aligned} \tag{10.4}$$

Remark 10.1.8 We stress that, under the assumption (10.2), the quotient space $\mathfrak{g}/[\mathfrak{g}, \mathfrak{g}]$ can be identified with Δ_1, as vector spaces, and the projection $\mathfrak{g} \twoheadrightarrow \mathfrak{g}/[\mathfrak{g}, \mathfrak{g}]$ can be identified with $\mathfrak{g} \twoheadrightarrow \Delta_1$, modulo $[\mathfrak{g}, \mathfrak{g}]$. Hence, the map $\pi_{\Delta_1} : G \twoheadrightarrow \Delta_1$ is nothing else than the abelianization map $\pi_{\mathrm{ab}} : G \twoheadrightarrow \mathrm{Ab}(G)$, in the case (10.2).

Lemma 10.1.9 *Let G be a nilpotent simply connected Lie group and let Δ_1 be a polarization on G such that (10.2) holds. For the map π_{Δ_1} of Definition 10.1.7, we have:*

10.1.9.i. The map $\pi_{\Delta_1} : G \to (\Delta_1, +)$ is a Lie group homomorphism.
10.1.9.ii. The differential of π_{Δ_1} is the identity when restricted to Δ_1:

$$(\pi_{\Delta_1})_*|_{\Delta_1} = \mathrm{id}_{\Delta_1}.$$

Proof of (i) Because of Remark 10.1.8, this is just a restatement of Proposition 10.1.2. We spell out another proof that avoids identifications. We write π for π_{Δ_1} and proj for proj_{Δ_1}. By Theorem 9.4.6, since G is simply connected and nilpotent, for all p and $q \in G$, exist X and $Y \in \mathfrak{g}$ such that $\exp(X) = p$ and $\exp(Y) = q$. On the one hand, by the BCH formula and assumption (10.2), we have

$$\begin{aligned} \pi(p \cdot q) &\overset{(10.4)}{=} \mathrm{proj}(\exp^{-1}(pq)) = \mathrm{proj}\left(\exp^{-1}(\exp(X)\exp(Y))\right) \\ &\overset{(5.26)}{=} \mathrm{proj}\left(X + Y + \frac{1}{2}[X, Y] + \dots\right) \overset{(10.2)}{=} \mathrm{proj}(X + Y). \end{aligned}$$

On the other hand, we have

$$\begin{aligned} \pi(p) + \pi(q) &= \mathrm{proj}(\exp^{-1}(p)) + \mathrm{proj}(\exp^{-1}(q)) \\ &= \mathrm{proj}(\exp^{-1}(p) + \exp^{-1}(q)) = \mathrm{proj}(X + Y). \end{aligned}$$

□

Proof of (ii) By Proposition 5.2.8, we have

$$\begin{aligned} \mathrm{d}\pi|_{\Delta_1} &\overset{\text{def}}{=} (\mathrm{d}(\mathrm{proj} \circ \exp^{-1}))_{1_G}|_{\Delta_1} \\ &\overset{(5.2.8)}{=} (\mathrm{d}\,\mathrm{proj})_0|_{\Delta_1} \\ &= (\mathrm{d}\,\mathrm{id})_0|_{\Delta_1} = \mathrm{id}_{\Delta_1}\,. \end{aligned}$$

□

Here is another restatement of Proposition 10.1.2:

Corollary 10.1.10 *Let G be a nilpotent simply connected Lie group polarized by Δ so that $\Delta_1 \oplus [\mathfrak{g}, \mathfrak{g}] = \mathfrak{g}$. Fix a norm $\|\cdot\|$ on Δ_1. The map $\pi_{\Delta_1} : (G, d_{cc}) \to (\Delta_1, \|\cdot\|)$ is a submetry, where d_{cc} is the metric on the sub-Finsler Lie group $(G, \Delta, \|\cdot\|)$.*

The map π_{Δ_1} from (10.4) is useful since it gives a second link between the tangents of horizontal curves and vectors in Δ_1. Recall the notions of development and multiplicative integral from Definitions 7.1.4 and 7.1.5, respectively.

Proposition 10.1.11 (Development as Projection on Abelianization) *Let G be a nilpotent simply connected Lie group polarized by Δ so that $\Delta_1 \oplus [\mathfrak{g}, \mathfrak{g}] = \mathfrak{g}$. Fix $T > 0$. If $\gamma : [0, T] \to G$ is a Δ-horizontal curve, then the development of γ is $\pi_{\Delta_1} \circ \gamma$. Vice versa, if $\sigma : [0, T] \to \Delta_1$ is an absolutely continuous curve, then the multiplicative integral of σ is the only Δ-horizontal curve γ with $\sigma := \pi_{\Delta_1} \circ \gamma$.*

Proof We need to prove the formula

$$\gamma'(t) \overset{\text{def}}{=} (L_{\gamma(t)})_*^{-1}\dot{\gamma}(t) = \frac{\mathrm{d}}{\mathrm{d}t}\left(\pi_{\Delta_1} \circ \gamma\right)(t), \qquad \forall t \in [0, T], \tag{10.5}$$

as elements of Δ_1. Using Lemma 10.1.9, and that $\pi_{\Delta_1}(1_G) = 0$, we get

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t}\left(\pi_{\Delta_1} \circ \gamma\right)(t) &= \lim_{h\to 0} \frac{\pi_{\Delta_1}(\gamma(t+h)) - \pi_{\Delta_1}(\gamma(t))}{h} \\ &\overset{10.1.9.\mathrm{i}}{=} \lim_{h\to 0} \frac{\pi_{\Delta_1}(\gamma(t)^{-1}\gamma(t+h))}{h} \\ &= \lim_{h\to 0} \frac{\pi_{\Delta_1}(L_{\gamma(t)}^{-1}\gamma(t+h))}{h} \\ &= \lim_{h\to 0} \frac{\pi_{\Delta_1}(L_{\gamma(t)}^{-1}\gamma(t+h)) - \pi_{\Delta_1}(L_{\gamma(t)}^{-1}\gamma(t))}{h} \\ &= \frac{\mathrm{d}}{\mathrm{d}h}\Big((\pi_{\Delta_1} \circ L_{\gamma(t)}^{-1} \circ \gamma\Big)(t+h)\Big|_{h=0} \\ &= (\pi_{\Delta_1})_* \circ (L_{\gamma(t)}^{-1})_*\dot{\gamma}(t) \\ &\overset{10.1.9.\mathrm{ii}}{=} \mathrm{id}(\gamma'(t)) = \gamma'(t). \end{aligned}$$

The vice versa is obvious since multiplicative integrals are the inverse operation of developments. □

Corollary 10.1.12 *Let G be a nilpotent simply connected Lie group polarized by Δ so that $\Delta_1 \oplus [\mathfrak{g}, \mathfrak{g}] = \mathfrak{g}$. Fix a norm $\|\cdot\|$ on Δ_1. The length of the horizontal curves equals the length of their projections:*

$$\text{Length}(\gamma) = \text{Length}(\pi_{\Delta_1} \circ \gamma), \qquad \forall \gamma \text{ horizontal curve in } (G, \Delta),$$

where the first length is with respect to the sub-Finsler metric and the second one is in the normed space $(\Delta_1, \|\cdot\|)$.

Proof By Proposition 10.1.11 (see (10.5)), one has

$$\begin{aligned} \text{Length}(\pi \circ \gamma) &= \int_0^1 \left\| \frac{\mathrm{d}}{\mathrm{d}t} (\pi \circ \gamma)(t) \right\| \mathrm{d}t \\ &= \int_0^1 \left\| \gamma'(t) \right\| \mathrm{d}t \\ &= \int_0^1 \left\| (L_{\gamma(t)})_*^{-1} \dot{\gamma}(t) \right\| \mathrm{d}t \\ &= \int_0^1 \|\dot{\gamma}(t)\| \, \mathrm{d}t \\ &= \text{Length}(\gamma). \end{aligned}$$

□

10.1.2.2 Horizontal Lines as Geodesics

Definition 10.1.13 Let $(G, \Delta, \|\cdot\|)$ be a sub-Finsler Lie group. Let $X \in \Delta_1$. The curve $t \mapsto \exp(tX)$ is the one-parameter subgroup tangent to the vector X, and it is called the *horizontal line* in the direction of X.

On each sub-Finsler Lie group $(G, \Delta, \|\cdot\|)$, the curve $t \mapsto \gamma(t) := \exp(tX)$ is horizontal with respect to Δ, because, recalling Corollary 5.2.6, we have $\dot{\gamma}(t) = X_{\gamma(t)} = \mathrm{d}L_{\gamma(t)} X \in \Delta_{\gamma(t)}$. We claim that the length of $t \mapsto \gamma(t)$, for $t \in [0, T]$, with respect to the sub-Finsler structure of $(G, V, \|\cdot\|)$ is $T \|X\|$. Indeed, we have

$$\begin{aligned} \text{Length}(\gamma) &= \int_0^T \|\dot{\gamma}(t)\| \, \mathrm{d}t \\ &= \int_0^T \left\| X_{\gamma(t)} \right\| \mathrm{d}t \end{aligned}$$

$$= \int_0^T \left\| \left(L_{\gamma(t)}\right)_* X_1 \right\| \, \mathrm{d}t$$

$$= \int_0^T \|X\| \, \mathrm{d}t$$

$$= T \, \|X\| \, ,$$

where we used that both X and the norm are left-invariant. In fact, we get the formula

$$\text{Length}\left(\exp(tX)|_{t\in[a,b]}\right) = |b-a| \cdot \|X\| \qquad \forall a, b \in \mathbb{R}. \tag{10.6}$$

In Lie groups endowed with left-invariant Riemannian metrics, one-parameter subgroups may not be geodesics. For instance, in the semidirect product $\mathbb{R} \rtimes \mathbb{R}_+$, which is the hyperbolic half-plane, the distance between the identity element (0, 1) and $(t, 1)$ equals $2\operatorname{arcsinh}(t/2)$, for $t \in \mathbb{R}$; see Exercise 13.4.12.

Regarding nilpotent groups, in the Riemannian Heisenberg group with orthonormal frame $X, Y, Z = [X, Y]$, the one-parameter subgroup in the direction Z is a Riemannian geodesic, but not globally length-minimizing. Whereas, the one-parameter subgroup in the direction $X + Z$ is not even a Riemannian geodesic, recall Proposition 8.1.8.

Proposition 10.1.14 *Consider a nilpotent simply connected Lie group G endowed with a left-invariant sub-Finsler distance with respect to some distribution Δ such that $\Delta_1 \oplus [\mathfrak{g}, \mathfrak{g}] = \mathfrak{g}$. Then, one-parameter subgroups of horizontal vectors are length-minimizing.*

Proof Take $X \in \Delta_1$. We shall prove that the curve $t \in \mathbb{R} \mapsto \exp(tX)$ is an homothetic embedding:

$$d_{\mathrm{cc}}(\exp(aX), \exp(bX)) = |b-a| \cdot \|X\| \, , \qquad \forall a, b \in \mathbb{R}.$$

Recall from Corollary 10.1.10 that the map π_{Δ_1} is a submetry. We have the following bounds: for all real numbers $a < b$

$$\begin{aligned} d_{\mathrm{cc}}(\exp(aX), \exp(bX)) &\leq \text{Length}\left(\exp(tX)|_{t\in[a,b]}\right) \\ &\overset{(10.6)}{=} (b-a) \, \|X\| \\ &= \|(b-a)X\| \\ &= \left\|\pi_{\Delta_1}((b-a)X)\right\| \\ &\leq d_{\mathrm{cc}}(\exp(aX), \exp(bX)), \end{aligned}$$

where in the last inequality we used the 1-Lipschitz property of π_{Δ_1}. □

10.1.3 Lift of Regularity

With the notion of submetry, we can transport several regularity results about geodesics from a sub-Finsler Lie group to its quotients. Said with a different perspective, if a nilpotent sub-Finsler Lie group may admit pathological geodesics, then the free-nilpotent Lie group with the same rank and step admits the same pathologies. Next, we will discuss the property of joining points by smooth geodesics. For other regularity results, we point out to Exercises 10.5.12 and 10.5.13, and Proposition 3.1.29.

Proposition 10.1.15 *Let G and H be nilpotent simply connected Lie groups equipped with left-invariant sub-Finsler structures: $(G, \Delta^G, \|\cdot\|)$ and $(H, \Delta^H, \|\cdot\|)$ such that $\mathfrak{h} = \Delta_1^H \oplus [\mathfrak{h}, \mathfrak{h}]$. Let $\varphi : G \to H$ be a Lie group homomorphism. Assume that each pair of points in G can be joined by a smooth geodesic. If*

$$\varphi_*|_{\Delta_1^{(G)}} : \Delta_1^G \to \Delta_1^H$$

is a submetry of normed spaces, then each pair of points in H can be joined by a smooth geodesic.

Proof The proof is an immediate consequence of Propositions 10.1.6.ii. and 3.1.29. □

10.2 Isometrically Homogeneous Spaces with Dilations (Second Part)

This section is the continuation of Sect. 6.5, where we already considered isometrically homogeneous spaces with dilations and proved that they, at least, have a structure of Lie coset spaces. We will now see that, in fact, they have the structure of nilpotent positively gradable Lie groups. Moreover, with respect to this group structure, each dilation is an affine map: a composition of a left translation and a group homomorphism; see Sect. 10.3 for a discussion on affine maps. Such a group structure on the metric space comes from the nilradical of the isometry group. The *nilradical* is the largest nilpotent connected normal subgroup; see Exercise 10.5.18 for the definition of nilradical at the level of Lie algebra.

Theorem 10.2.1 *Let (M, d) be a metric space such that*

(1) (M, d) is locally compact;
(2) (M, d) is locally connected;
(3) the isometry group $\mathrm{Isom}(M, d)$ of (M, d) acts transitively on M;
(4) there is a dilation on (M, d) of factor in $(0, +\infty) \setminus \{1\}$.

Then there is a unique nilpotent Lie group N with a left-invariant distance d_N on N such that (N, d_N) is isometric to (M, d) and every dilation of (N, d_N) is an affine map of N. Moreover, this Lie group N is positively graded and isomorphic to the nilradical of $\mathrm{Isom}(M, d)$.

Proof Recall what, under the same assumptions, we already saw in the proof of Theorem 6.5.1: the identity component $G := \mathrm{Isom}(M, d)^\circ$ of the isometry group $\mathrm{Isom}(M, d)$ is a connected Lie group. We fixed a point $o \in M$ and a dilation $\delta_\lambda : M \to M$ of factor $\lambda \in (0, 1)$ fixing o. We defined $S := \mathrm{Stab}(o)$ to be the stabilizer of G at o, which is a compact Lie subgroup of G. Via the orbit map π, the Lie coset space G/S is homeomorphic to M.

Define for $f \in \mathrm{Isom}(M, d)$

$$Tf := \delta_\lambda \circ f \circ \delta_\lambda^{-1}.$$

We claim that the map T is valued into $\mathrm{Isom}(M, d)$, that $T : \mathrm{Isom}(M, d) \to \mathrm{Isom}(M, d)$ is a Lie group automorphism, and $T(S) = S$. Indeed, one easily sees that if f is an isometry of M, then $Tf : M \to M$ is an isometry as well, and that $T(f \circ g) = Tf \circ Tg$ for all $f, g \in \mathrm{Isom}(M, d)$. Moreover, T is continuous: if $\{f_k\}_{k\in\mathbb{N}}$ is a sequence converging in $\mathrm{Isom}(M, d)$ to f, i.e., $f_k \to f$ uniformly on compact sets, then Tf_k converges to Tf uniformly on compact sets as $k \to \infty$, because for every compact $K \subseteq M$

$$\sup\{d(Tf_k(p), Tf(p)) : p \in K\} = \lambda \sup\left\{d(f_k(p), f(p)) : p \in \delta_\lambda^{-1}(K)\right\}.$$

Since T is a continuous automorphism, then T is a Lie group automorphism; see Theorem 5.3.2. Finally, since $\delta_\lambda(o) = o$, then $T(f) \in S$, for all $f \in S$.

Let $\mathfrak{g} := \mathrm{Lie}(\mathrm{Isom}(M, d))$ be the Lie algebra of $\mathrm{Isom}(M, d)$ and $\mathfrak{s} := \mathrm{Lie}(S)$ the Lie algebra of the stabilizer S. We consider the bilinear form

$$B : \mathfrak{g} \times \mathfrak{g} \to \mathbb{R}, \qquad B(v, w) := \mathrm{Trace}(\mathrm{ad}(v)\,\mathrm{ad}(w)),$$

which is the bilinear symmetric form called the Killing form; see Exercise 10.5.16.

We claim that B restricted to $\mathfrak{s}$ is negative definite. For doing this, fix $v \in \mathfrak{s}$, for which we want to show that $B(v, v) < 0$ unless $v = 0$. Since Ad_S is compact in $\mathrm{GL}(\mathfrak{g})$, then the eigenvalues of ad_v are purely imaginary, see Exercise 10.5.15. So the eigenvalues of ad_v^2 are real and nonpositive. Consequently, because the trace is the sum of the eigenvalues, we have that $B(v, v) = \mathrm{Trace}(\mathrm{ad}_v^2) \leq 0$. It is left to prove that,

$$\text{if } B(v, v) = 0, \qquad \text{then } v = 0. \tag{10.7}$$

Assume that $\mathrm{Trace}(\mathrm{ad}_v^2) = B(v, v) = 0$, then since again the eigenvalues of ad_v^2 are nonpositive, we deduce that all the eigenvalues of ad_v^2, and hence the ones of ad_v, are zero. But the only diagonalizable transformation with all zero eigenvalues is the zero transformation (recall Exercise 10.5.15). We deduced that $\mathrm{ad}_v = 0$. Therefore, from Formula 5.5.7 we have $\mathrm{Ad}_{\exp(tv)} = e^{t\,\mathrm{ad}_v} = \mathrm{id}_{\mathfrak{g}}$ and $C_{\exp(tv)} = \mathrm{id}_G$, for all $t \in \mathbb{R}$, recalling that G is the identity component of $\mathrm{Isom}(M, d)$. Therefore, since G is connected, we have that $\exp(tv)$ is both in the center $Z(G)$ of G and also in S for all t. We claim that $Z(G) \cap S$ is trivial. Indeed, take $f \in Z(G) \cap S$ and $p \in M$. Since the action of G on the connected space M is transitive by Proposition 6.2.9, there is $g \in G$ such that $g(o) = p$. We deduce that $f(p) = f(g(o)) = g(f(o)) = g(o) = p$. Since this holds for every $p \in M$, we deduce that $f = \mathrm{id}_M$. The claim about the triviality of $Z(G) \cap S$ is proven. Consequently, $\exp(tv) = \mathrm{id}_M$ for all $t \in \mathbb{R}$. Hence $v = 0$. We proved (10.7), and, therefore, that B is negative definite on $\mathfrak{s}$.

We next consider the map

$$\phi : \mathfrak{g} \to \mathfrak{s}^*, \qquad \phi(v) := B(v, \cdot),$$

and define the orthogonal of $\mathfrak{s}$ with respect to the Killing form:

$$\mathfrak{n} := \mathrm{Ker}\,\phi \stackrel{\text{def}}{=} \{v \in \mathfrak{g} : B(v, w) = 0, \forall w \in \mathfrak{s}\}. \tag{10.8}$$

We claim that $\mathfrak{g} = \mathfrak{n} \oplus \mathfrak{s}$, as linear spaces. We begin by noting that $\mathfrak{n} \cap \mathfrak{s} = \{0\}$ by (10.7). To deduce that $\mathfrak{g} = \mathfrak{n} \oplus \mathfrak{s}$, we stress that $\mathfrak{n}$ is the kernel of the map ϕ and that ϕ restricted to $\mathfrak{s}$ is an isomorphism because B is negatively definite on $\mathfrak{s}$.

We next claim that $T_*(\mathfrak{n}) = \mathfrak{n}$. Indeed, since $T_* : \mathfrak{g} \to \mathfrak{g}$ is a Lie algebra automorphism, we have $B(T_*v, T_*w) = B(v, w)$ for all $v, w \in \mathfrak{g}$. Moreover, $T_*(\mathfrak{s}) = \mathfrak{s}$. Therefore, $T_*(\mathfrak{n}) = \mathfrak{n}$.

For $t \in \mathbb{R}$, let $V_t := V_t(e^{-1}, T_*)$ be the Lie algebra grading of $\mathfrak{g}$ induced by T_* as in Proposition 9.2.19. Since both $\mathfrak{s}$ and $\mathfrak{n}$ are invariant subspaces for T_*, we have $V_t = (V_t \cap \mathfrak{s}) \oplus (V_t \cap \mathfrak{n})$ for each $t \in \mathbb{R}$.

We claim that $\mathfrak{s} \subset V_0$. Indeed, we have seen that T_* is preserving B and that $-B$ is a scalar product on $\mathfrak{s}$. Therefore, the linear transformation $T_*|_{\mathfrak{s}}$ is orthogonal with respect to $-B$ and thus every eigenvalue α of $T_*|_{\mathfrak{s}}$ has norm 1. Thus, every element in $\mathfrak{s}$ has degree t with $t \stackrel{\text{def}}{=} -\log(|\alpha|) = -\log 1 = 0$; see (9.12).

Next, we claim that $\mathfrak{n} \subset V_{>0} := \bigoplus_{t>0} V_t$. Arguing by contradiction, suppose that there is an eigenvalue $\alpha \in \mathbb{C}$ such that $V_t \cap \mathfrak{n} \neq \{0\}$ with $t := -\log(|\alpha|) < 0$, i.e., with $|\alpha| \geq 1$. Let $\mathscr{B}$ be a basis of $\mathfrak{g}$ that represents T_* in Jordan form and let $|\cdot|$ be the Euclidean norm on $\mathfrak{g}$ so that $\mathscr{B}$ is an orthonormal basis. Let U be an open ball in $\mathfrak{g}$ centered at 0 such that the restriction of $\exp : \mathfrak{g} \to G$ to $2U$ is a diffeomorphism between $2U$ and its image and such that $\exp(\mathfrak{s} \cap U) = S \cap \exp(U)$. In a first case, if $\alpha \in \mathbb{R}$, then there is a nonzero $a \in \mathfrak{n} \cap U$ such that $T_*a = \alpha a$, while, in the second case, if $\alpha \notin \mathbb{R}$ there there are nonzero $a, b \in \mathfrak{n} \cap U$ such that $|a| = |b|$, $T_*a = \Re\mathfrak{e}(\alpha)a + \Im\mathfrak{m}(\alpha)b$ and $T_*b = -\Im\mathfrak{m}(\alpha)a + \Re\mathfrak{e}(\alpha)b$. Notice that, in both cases, we have that $|T_*^k a| = |\alpha|^k |a|$ for all $k \in \mathbb{N}$.

In a first case, if $|\alpha| = 1$, then there is a diverging sequence k_j of natural numbers with $\lim_{j\to\infty} T_*^{k_j} a = a \in U$. Hence

$$\begin{aligned}\exp(a)(o) &= \lim_{j\to\infty} \exp(T_*^{k_j} a)(o)\\ &= \lim_{j\to\infty} \delta_\lambda^{k_j} \circ \exp(a) \circ \delta_\lambda^{-k_j}(o)\\ &= \lim_{j\to\infty} \delta_\lambda^{k_j}(\exp(a)(o))\\ &= o,\end{aligned}$$

that is $\exp(a) \in S \cap \exp(U)$. Thanks to our conditions on U, we obtain $a \in \mathfrak{s} \cap \mathfrak{n} = \{0\}$, in a contradiction since $a \neq 0$.

In the second case, if $|\alpha| > 1$, then $|T_*^{-k} a| = |\alpha|^{-k}|a|$ is a vanishing sequence as $k \to \infty$. Therefore,

$$o = \mathrm{id}(o) = \lim_{k\to\infty} \exp(T_*^{-k} a)(o) = \lim_{k\to\infty} \delta_\lambda^{-k} \exp(a)\delta_\lambda^{k}(o) = \lim_{k\to\infty} \delta_\lambda^{-k} \exp(a)(o),$$

where $\lim_{k\to\infty} d(o, \delta_\lambda^{-k} \exp(a)(o)) = +\infty$, since δ_λ^{-1} is an expanding dilation. In both cases, we have a contradiction. We have thus proven our claim that $\mathfrak{n} \subset V_{>0}$.

We conclude that the grading is nonnegative, $\mathfrak{s} = V_0$, and $\mathfrak{n} = V_{>0}$. Consequently, the linear space $\mathfrak{n}$ is a nilpotent ideal in $\mathfrak{g}$; recall Exercises 9.5.19 and 9.5.29.

Therefore, by the definition of nilradical $\mathfrak{nil}(\mathfrak{g})$ of $\mathfrak{g}$, we have $\mathfrak{n} \subset \mathfrak{nil}(\mathfrak{g})$. We stress that the Killing form is negative definite on $\mathfrak{s}$, while it is zero on $\mathfrak{nil}(\mathfrak{g})$; see Exercise 10.5.20. We infer $\mathfrak{s} \cap \mathfrak{nil}(\mathfrak{g}) = \{0\}$ and hence $\mathfrak{n} = \mathfrak{nil}(\mathfrak{g})$.

Now that we know that $\mathfrak{n}$ is a Lie algebra, we can consider $N < \mathrm{Isom}(M, d)$ to be the connected Lie subgroup with Lie algebra $\mathfrak{n}$, by Theorem 5.1.4. We claim that the orbit map $\pi|_N : N \to M$ restricted to N is a homeomorphism. Indeed, we firstly recall that the identity component G of the isometry group acts transitively on the connected space M; see Proposition 6.2.9. Secondly, because $\mathfrak{g} = \mathfrak{n} \oplus \mathfrak{s}$, the group N acts transitively on M, i.e., $\pi(N) = M$. Thirdly, we have $N \cap S$ is discrete because $\mathfrak{n} \cap \mathfrak{s}$ is trivial. So, $\pi|_N : N \to M$ is a covering map. Since the manifold M is simply connected by Exercise 6.6.30, the map $\pi|_N$ is actually a homeomorphism. Therefore, N is simply connected, and it is a graded nilpotent Lie group.

Finally, we make N into a metric Lie group by pulling back the distance d from M to N via π:

$$d_N(f, g) := d(\pi(f), \pi(g)) = d(f(o), g(o)), \qquad \forall f, g \in N.$$

We stress that by the definitions, for every $f \in N$, we have $\pi(Tf) = Tf(o) = (\delta_\lambda \circ f \circ \delta_\lambda^{-1})(o) = (\delta_\lambda \circ f)(o) = \delta_\lambda(\pi(f))$. We conclude that $\pi|_N$ is an isometry between

(N, d_N) and (M, d) that relates the dilation δ_λ on M with the automorphism T on N.

For the uniqueness, suppose that M has another Lie group structure so that d is a left-invariant admissible distance, and δ_λ is a Lie automorphism. Then the set M^L of left translations is a subgroup of G whose Lie algebra $\mathfrak{m}$ is complementary to $\mathfrak{s}$, i.e., $\mathfrak{g} = \mathfrak{s} \oplus \mathfrak{m}$, and invariant under T_*, i.e., $T_*\mathfrak{m} = \mathfrak{m}$. Take $m \in \mathfrak{m}$ and decompose it as $m = m_\mathfrak{s} + m_\mathfrak{n}$ with $m_\mathfrak{s} \in \mathfrak{s}$ and $m_\mathfrak{n} \in \mathfrak{n}$. Then, since T_* is contractive on $\mathfrak{n}$, up to subsequence, we get

$$\mathfrak{m} \ni \lim_{k\to\infty} T_*^k m = \lim_{k\to\infty} T_*^k (m_\mathfrak{s} + m_\mathfrak{n}) = \lim_{k\to\infty} T_*^k m_\mathfrak{s} \in \mathfrak{s}.$$

Thus $\lim_{k\to\infty} T_*^k m_\mathfrak{s} = 0$ and, since T_* restricted to $\mathfrak{s}$ preserves the norm given by $-B$, we deduce that $m_\mathfrak{s} = 0$. We conclude that $\mathfrak{m} = \mathfrak{n} = \mathrm{nil}(\mathfrak{g})$, and thus the original group structure is equal to the one given by N. □

10.3 Affinity of Isometries of Nilpotent Lie Groups

We say that a map between groups is *affine* if it is the composition of a left translation and a group homomorphism. Equivalently, it is the composition of a group homomorphism and a left translation. For nilpotent groups, we have the following result on their isometries, which we will not prove but will discuss only partially.

Theorem 10.3.1 ([KL17, Theorem 1.2], after Wolf [Wol62, Wil82]) *Isometries between nilpotent connected metric Lie groups are affine.*

Here are some consequences:

[10.3.1.i] Two isometric nilpotent connected metric Lie groups are isomorphic;
[10.3.1.ii] Every dilation between nilpotent connected metric Lie groups is an affine map;
[10.3.1.iii] Given a connected metric Lie group N, its isometry group $\mathrm{Isom}(N)$, which always is a Lie group, is a semidirect product if N is nilpotent. Namely,

$$\mathrm{Isom}(N) = N \rtimes \mathrm{AutIsom}(N),$$

where N is seen inside $\mathrm{Isom}(N)$ as left translations and $\mathrm{AutIsom}(N)$ denotes the group of automorphisms of N that are isometries.

Moreover, with the above notation, we have

[10.3.1.iv] The group N is a maximal connected nilpotent subgroup of $\mathrm{Isom}(N)$, and the Lie algebra of N is the nilradical of the Lie algebra of $\mathrm{Isom}(N)$.

We will not prove Theorem 10.3.1 in this text. Historically, this has been proved in [KL17] using Theorem 8.2.1 and reducing it to the case of Riemannian manifolds where Wolf has proved the result in [Wol62]. In Wolf's proof, there is an unclear step, which has been later clarified by Wilson in [Wil82]. Nowadays, we have more direct proofs of Theorem 10.3.1. See, for example, [Cow+24].

Given a metric group (M, d), we denote by M^L the group of the left translations inside the isometry group $\mathrm{Isom}(M, d)$ and by $\mathrm{Stab}_1(\mathrm{Isom}(M, d))$ the stabilizer of the identity element $1 = 1_M$. We denote by $\mathrm{Aff}(M)$ the group of affine maps from M to M and by $\mathrm{Aut}(M)$ the group of automorphisms of M. Then the following properties are equivalent:

1. $M^L \lhd \mathrm{Isom}(M, d)$;
2. $\mathrm{Isom}(M, d) < \mathrm{Aff}(M)$;
3. $\mathrm{Stab}_1(\mathrm{Isom}(M, d)) < \mathrm{Aut}(M)$;
4. $\mathrm{Isom}(M, d) = M^L \rtimes \mathrm{Stab}_1(\mathrm{Isom}(M, d))$;
5. $\mathrm{Isom}(M, d) = M^L \rtimes (\mathrm{Isom}(M, d) \cap \mathrm{Aut}(M))$.

All of the properties hold when (M, d) is nilpotent and connected.

10.4 Guivarc'h Seminorms on Nilpotent Lie Groups

On nilpotent simply connected Lie groups, there are special (coarse) distance functions that are more linked to the Lie algebra nilpotency properties. They are called homogeneous quasi-norms since they behave well with respect to algebraic dilations. The most important examples of homogeneous quasi-norms are the following:

1. Guivarc'h seminorms, which are present in every nilpotent simply connected Lie group; see Lemma 10.4.2.
2. The distance $d(1, \cdot)$ from the identity element in every Carnot group; see Proposition 11.1.3.
3. The asymptotic distance $d_\infty(1_G, \cdot)$ from the identity element in every simply connected nilpotent sub-Finsler Lie group G, after the identification of G with its associated Carnot group G_∞, see Corollary 12.4.4.

Definition 10.4.1 Let G be a nilpotent simply connected Lie group. Fix on its Lie algebra $\mathfrak{g}$ a linear grading $(V_a)_{a\in\mathbb{R}}$, as in Definition 9.2.1 and let δ_λ be the inhomogeneous dilation on $\mathfrak{g}$ of factor λ relative to the grading, as in Definition 9.2.12. An *homogeneous quasi-norm* is a function $|\cdot| : G \to \mathbb{R}_{\geq 0}$ such that

10.4.1.i. it a continuous function,
10.4.1.ii. $|x| = 0$ if and only if $x = 1_G$, and
10.4.1.iii. $|\delta_\lambda(x)| = \lambda|x|$ for every $x \in G$ and every $\lambda \in \mathbb{R}_+$.

In particular, for each homogeneous quasi-norm, in exponential coordinates, we have that

$$|\lambda x| = \lambda^{1/a}|x|, \qquad \forall a \in \mathbb{R}, \forall x \in V_a.$$

As previously mentioned, important examples were constructed by Guivarc'h in the presence of compatible linear gradings as in Definition 9.2.2. We will present the proof taken from [Bre14, Lemma 2.5], which, however, is an alternative exposition of the original one [Gui73, Lemma II.1].

Lemma 10.4.2 (Guivarc'h) *Let G be a nilpotent simply connected Lie group. On its Lie algebra $\mathfrak{g}$ fix a compatible linear grading $\mathfrak{g} = V_1 \oplus \ldots \oplus V_s$, and denote by $((x)_1, \ldots, (x)_s)$ the decomposition of each $x \in \mathfrak{g}$ with respect to this grading. Then for every $\varepsilon > 0$ there exists a norm $\|\cdot\|$ on $\mathfrak{g}$ such that the function*

$$g \in G \longmapsto |g| := \max_{j\in\{1,\ldots,s\}} \left\|(\log g)_j\right\|^{1/j} \tag{10.9}$$

is a homogeneous quasi-norm that satisfies

$$|g \cdot h| \le |g| + |h| + \varepsilon, \quad \forall g, h \in G. \tag{10.10}$$

Proof We fix some scalar product on $\mathfrak{g}$ that makes the compatible linear grading orthogonal. We shall not change the induced norm $\|\cdot\|$ on V_1, while on every other V_j we will replace the norm with the norm $\lambda_j \|\cdot\|$ for some $\lambda_j \in \mathbb{R}_+$ so that (10.10) would hold. The λ_j's will be taken to be smaller and smaller as j increases.

We work in exponential coordinates. We set $|x|_\lambda := \max_j \left\|\lambda_j (x)_j\right\|^{1/j}$ for $\lambda \in \mathbb{R}^s_+$. We want that for every index $j \in \{1, \ldots, s\}$,

$$\lambda_j \left\|(x \star y)_j\right\| \le (|x|_\lambda + |y|_\lambda + \varepsilon)^j, \qquad \forall x, y \in \mathfrak{g}. \tag{10.11}$$

By BCH Formula, we have $(x \star y)_j = (x)_j + (y)_j + P_j(x, y)$, where P_j is a polynomial map into V_j depending only on $(x)_i$ and $(y)_i$ with $i \in \{1, \ldots, j-1\}$ such that

$$\left\|P_j(x, y)\right\| \le C_j \cdot \sum_{l,m\ge1,l+m\le j} M_{j-1}(x)^l M_{j-1}(y)^m,$$

where $M_k(x) := \max_{i\le k} \|(x)_i\|^{1/i}$ and $C_j > 0$ is a constant depending on P_j and on the norms $\|\cdot\|_i$'s. Since $\varepsilon > 0$, when expanding the right-hand side of (10.11) all terms of the form $|x|^l_\lambda |y|^m_\lambda$ with $l+m \le j$ appear with some positive coefficient, say $\varepsilon_{l,m}$. The terms $|x|^j_\lambda$ and $|y|^j_\lambda$ appear with coefficient 1 and cause no trouble since we always have $\lambda_j \left\|(x)_j\right\| \le |x|_\lambda^j$ and $\lambda_j \left\|(y)_j\right\| \le |y|_\lambda^j$. Therefore, for (10.11) to hold, it is sufficient that

$$\lambda_j C_j M_{j-1}(x)^l M_{j-1}(y)^m \le \varepsilon_{l,m} |x|_\lambda^l |y|_\lambda^m,$$

for all remaining l and m. However, clearly $M_k(x) \le \Lambda_k \cdot |x|_\lambda$ where $\Lambda_k := \max_{i \le k}\{1/\lambda_i^{1/i}\} \ge 1$. Hence, a sufficient condition for (10.11) to hold is

$$\lambda_j \le \left(C_j \Lambda_{j-1}^j\right)^{-1} \min_{l+m \le j} \varepsilon_{l,m}.$$

Since Λ_{j-1} depends only on $\lambda_1, \dots, \lambda_{j-1}$, such a set of conditions can be fulfilled by a suitable s-tuple λ. □

Every two homogeneous quasi-norms are bi-Lipschitz equivalent; see Exercise 10.5.24. Moreover, for every left-invariant geodesic distance, the distance from the identity element is coarsely bi-Lipschitz equivalent to each homogeneous quasi-norm, in the following sense.

Theorem 10.4.3 (Guivarc'h) *Let G be a nilpotent simply connected Lie group equipped with a left-invariant geodesic distance d. Let $|\cdot|$ be a homogeneous quasi-norm with respect to a compatible linear grading. Then there exists a constant $C > 1$ such that*

$$\frac{1}{C}|g| - C \le d(1, g) \le C|g| + C, \quad \forall g \in G. \tag{10.12}$$

Proof Recall that Guivarc'h seminorms, as defined in Lemma 10.4.2, are homogeneous quasi-norms. Then, since homogeneous quasi-norms (with respect to the same dilations) are bi-Lipschitz equivalent, see Exercise 10.5.24, it is enough to show (10.12) when $|\cdot|$ is some Guivarc'h seminorm, for some $\epsilon > 0$ as to satisfy (10.10).

We begin with the first inequality: $\frac{1}{C}|x| - C \le d(1, x)$. Let

$$C := \max\{|x| : d(1, x) \le 1\} + \varepsilon.$$

Take $x \in G$ with $L := d(1, x)$. Take a geodesic from 1 to x and subdivide it into $k := \lfloor L \rfloor + 1$ pieces $x_0 = 1, x_1, \dots, x_k = x$ with $d(x_{i-1}, x_i) \le 1$, for every $i = 1, 2, \dots, k$. Then we have

$$d(1, x_{i-1}^{-1} x_i) = d(x_{i-1}, x_i) \le 1, \quad \forall j \in \{1, 2, \dots, k\},$$

and thus, by definition of C, we have

$$|x_{i-1}^{-1} x_i| \le C - \varepsilon, \quad \forall j \in \{1, 2, \dots, k\}. \tag{10.13}$$

Hence, using that $|\cdot|$ comes from Lemma 10.4.2, we can bound

$$\begin{aligned}|x| &= |x_1 \cdot x_1^{-1}x_2 \cdot \dots \cdot x_{k-1}^{-1}x_k| \\ &\overset{(10.10)}{\leq} |x_1| + |x_1^{-1}x_2| + \dots + |x_{k-1}^{-1}x_k| + k\varepsilon \\ &\overset{(10.13)}{\leq} k(C-\varepsilon) + k\varepsilon \\ &= kC \leq C(L+1) = Cd(1,x) + C.\end{aligned}$$

We then prove the second inequality: $d(1,x) \leq C|x| + C$. The proof is by induction on the nilpotency step s of G. In the abelian case, it is clear. We assume the result is proved for groups up to step $s-1$, and we want to prove it for an s-step group G. We consider the quotient modulo the normal subgroup given by the last non-trivial element $C^s(G)$ of the lower central series. Note that ${}_{C^s(G)}\backslash^{G} = {}^{G}/_{C^s(G)}$ is a Lie group of step $s-1$ that is equipped with the distance (6.4), which in this case is left-invariant and geodesic since d is geodesic; see Proposition 3.1.27. Hence, the result is valid in ${}^{G}/_{C^s(G)}$. Namely, there exists $C > 0$ such that for every $x \in G$ we have

$$d(1, xC^s(G)) \leq C|xC^s(G)| + C.$$

It follows that, there exists $z \in C^s(G)$ such that

$$d(1, xz) \leq C|xC^s(G)| + C \leq C|x| + C, \tag{10.14}$$

where we used the fact that projections reduce Guivarc'h seminorms. On the one hand, by triangle inequality and left invariance, we have

$$\begin{aligned}d(1,x) &\leq d(1,xz) + d(xz,x) \\ &\overset{(10.14)}{\leq} C|x| + C + d(1,z).\end{aligned}$$

On the other hand, by Lemma 10.4.2, we also have

$$\begin{aligned}|z| &= |x^{-1} \cdot xz| \\ &\overset{(10.10)}{\leq} |x^{-1}| + |xz| + \varepsilon \\ &\leq |x| + Cd(1,xz) + C + \varepsilon \\ &\overset{(10.14)}{\leq} |x| + C(C|x| + C) + C + \varepsilon \\ &= (C^2+1)|x| + C^2 + C + \varepsilon,\end{aligned}$$

where in the second inequality, we used that in exponential coordinates, the inverse of x is $-x$, and the inequality proved in the first part of this proof. Therefore, we just need to bound $d(1, z)$ affinely in $|z|$, for $z \in C^s(G)$.

Let $c := \dim C^s(G)$ and $\Omega := B_d(1_G, 1)$. Then the set

$$\Omega' := \{e_1 \cdots e_c : \forall i = 1, 2, \ldots, c \ \exists\, u_{i1}, \ldots, u_{is} \in \Omega : e_i = [u_{i1}, \ldots, u_{is}]\}$$

is a neighborhood of 1 in $C^s(G)$; see Exercise 10.5.25. Then there exists $m \in \mathbb{N}$ such that $\log(\Omega')$ contains the $\frac{1}{m^s}$-ball with respect to the norm $\|\cdot\|$ on V_s giving the Guivarc'h seminorm. For every $z \in C^s(G)$, let $k \in \mathbb{N}$ be such that

$$k - 1 \leq |z| \stackrel{\text{def}}{=} \|z\|^{1/s} \leq k. \tag{10.15}$$

Consequently, without distinguishing the group with the Lie algebra via the exponential map, we have

$$\frac{1}{m^s k^s} z \in \Omega'.$$

Hence, we can write

$$\frac{1}{m^s k^s} z = e_1 \cdot \ldots \cdot e_c,$$

with $e_i = [u_{i1}, \ldots, u_{is}]$ and $u_{ij} \in B_d(1_G, 1)$, for every i, j. Then, using that each $[u_{i1}, \ldots, u_{is}]$ is a central element, we get

$$\begin{aligned} z &= \left(\frac{1}{m^s k^s} z\right)^{m^s k^s} \\ &= \left(\prod_{i=1}^{c} [u_{i1}, \ldots, u_{is}]\right)^{m^s k^s} \\ &= \prod_{i=1}^{c} [u_{i1}, \ldots, u_{is}]^{m^s k^s} \\ &= \prod_{i=1}^{c} [u_{i1}^{mk}, \ldots, u_{is}^{mk}], \end{aligned}$$

where in the last equation, we use this general property of nilpotent groups; see Exercise 10.5.26. Then, iteratively using the triangle inequality and the fact that u_{ij} are in the unit ball, we infer that for some constant $C_{c,s}$, we can bound

$$\begin{aligned} d(1,z) &= d\Big(1, \prod_{i=1}^{c} [u_{i1}^{mk}, \dots, u_{is}^{mk}]\Big) \\ &\leq C_{c,s} \max_{i,j} d(1, u_{ij}^{mk}) \\ &\leq C_{c,s}\, mk \max_{i,j} d(1, u_{ij}) \\ &\leq C_{c,s,m}\, k \overset{(10.15)}{\leq} C_{c,s,m}(|z|+1). \end{aligned}$$

□

10.5 Exercises

Exercise 10.5.1 Let H be a closed subgroup of a topological group G. Assume that G is simply connected and H is connected. Then, the topological space G/H is simply connected.
Hint. Every lift of a loop has extremal points joining elements in H.

Exercise 10.5.2 For every nilpotent simply connected Lie group G, one has

$$\exp(X + [\mathfrak{g}, \mathfrak{g}]) = \exp(X)\exp([\mathfrak{g}, \mathfrak{g}]) = \exp(X)[G, G], \qquad \forall X \in \mathfrak{g}.$$

Hint. Use Baker-Campbell-Hausdorff formula.

Exercise 10.5.3 Let $(G, \Delta, \|\cdot\|)$ be a sub-Finsler nilpotent simply connected Lie group, as in Definition 10.1.1. The norm given in (10.1) is

$$\|v\|_{\mathrm{ab}} = \min\{\|w\| : w \in \Delta_{1_G}, \pi_{\mathrm{ab}}(w) = v\}, \qquad \forall v \in \mathrm{Ab}(G).$$

Exercise 10.5.4 Let G be a simply connected and nilpotent Lie group. See $\mathrm{Ab}(G)$ as $\mathfrak{g}/[\mathfrak{g}, \mathfrak{g}]$ with abelianization map $\pi_{\mathrm{ab}} : \mathfrak{g} \to \mathrm{Ab}(G)$. Consider the map $\tilde{\pi} : G \to \mathrm{Ab}(G)$ given by $\tilde{\pi} := \pi_{\mathrm{ab}} \circ \log$. Then, we have

$$\tilde{\pi}(xy) = \tilde{\pi}(x) + \tilde{\pi}(y), \qquad \forall x, y \in G.$$

Hint. Baker-Campbell-Hausdorff formula gives $\exp^{-1}(xy) \equiv \exp^{-1}(x) + \exp^{-1}(y)$ modulo $[\mathfrak{g}, \mathfrak{g}]$.

Exercise 10.5.5 Using selection arguments, like Theorem 7.1.10 as used in Proposition 7.1.9, one proves Corollary 10.1.4.

Exercise 10.5.6 Let $\mathfrak{g}$ be a nilpotent Lie algebra and let Δ_1 be a subspace of $\mathfrak{g}$ such that

$$\mathfrak{g} = \Delta_1 + [\mathfrak{g}, \mathfrak{g}]. \tag{10.16}$$

For the lower central series $(\mathfrak{g}^{(i)})_{i\in\mathbb{N}}$ of $\mathfrak{g}$, we have

$$\mathfrak{g}^{(2)} = [\Delta_1, \Delta_1] + \mathfrak{g}^{(3)}$$

and, more generally, we have

$$\mathfrak{g}^{(i)} = [\Delta_1, [\Delta_1, [\ldots, [\Delta_1, \Delta_1]\ldots]] + \mathfrak{g}^{(i+1)},$$

where in the above brackets Δ_1 appears i-many times. Consequently, every Δ_1 with property (10.16) is Lie generating $\mathfrak{g}$.

Exercise 10.5.7 Let $\mathfrak{g}$ be a Lie algebra. Let $V \subset \mathfrak{g}$ be a subspace such that $V + [\mathfrak{g}, \mathfrak{g}] \neq \mathfrak{g}$. Then V is not bracket-generating $\mathfrak{g}$.
Hint. The Lie span of each subspace V is contained in $V + [\mathfrak{g}, \mathfrak{g}]$.

Exercise 10.5.8 Let $\mathfrak{g}$ be a nilpotent Lie algebra and let Δ_1 be a subspace of $\mathfrak{g}$. Then, the set Δ_1 Lie generates $\mathfrak{g}$ if and only if $\mathfrak{g} = \Delta_1 + [\mathfrak{g}, \mathfrak{g}]$.

Exercise 10.5.9 There is a nilpotent simply connected Lie group G with sub-spaces Δ_1 and $\tilde{\Delta}_1$ such that

$$\mathfrak{g} = \Delta_1 \oplus [\mathfrak{g}, \mathfrak{g}] = \tilde{\Delta}_1 \oplus [\mathfrak{g}, \mathfrak{g}],$$

with the property that there is no Lie algebra isomorphism $\phi : \mathfrak{g} \to \mathfrak{g}$ such that $\phi(\Delta_1) = \tilde{\Delta}_1$. Compare it with Proposition 9.2.9.
Hint. Check Exercise 9.5.28.

Exercise 10.5.10 Let $\mathfrak{g}$ and $\mathfrak{h}$ be nilpotent Lie algebras. Let $\Delta_1^{\mathfrak{g}}$ (resp., $\Delta_1^{\mathfrak{h}}$) be a linear subspace of $\mathfrak{g}$ (resp., $\mathfrak{h}$) such that $\mathfrak{g} = \Delta_1^{\mathfrak{g}} \oplus [\mathfrak{g}, \mathfrak{g}]$ (resp., $\mathfrak{h} = \Delta_1^{\mathfrak{h}} \oplus [\mathfrak{h}, \mathfrak{h}]$). Let $\phi : \mathfrak{g} \to \mathfrak{h}$ be a Lie algebra homomorphism. If ϕ has the property that $\phi(\Delta_1^{\mathfrak{g}}) \subseteq \Delta_1^{\mathfrak{h}}$, then

$$\mathrm{proj}_{\Delta_1^{\mathfrak{h}}} \circ \phi = \phi \circ \mathrm{proj}_{\Delta_1^{\mathfrak{g}}}, \qquad \text{i.e.,} \qquad \begin{array}{ccc} \mathfrak{g} & \xrightarrow{\phi} & \mathfrak{h} \\ \downarrow{\scriptstyle \mathrm{proj}} & & \downarrow{\scriptstyle \mathrm{proj}} \\ \Delta_1^{\mathfrak{g}} & \xrightarrow{\phi} & \Delta_1^{\mathfrak{h}} \end{array},$$

where $\mathrm{proj}^{\mathfrak{g}} : \mathfrak{g} \to \mathfrak{g}$ and $\mathrm{proj}^{\mathfrak{h}} : \mathfrak{h} \to \mathfrak{h}$ are the projections onto $\Delta_1^{\mathfrak{g}}$ and $\Delta_1^{\mathfrak{h}}$ respectively with kernels $[\mathfrak{g}, \mathfrak{g}]$ and $[\mathfrak{h}, \mathfrak{h}]$ respectively.

Solution. If $X \in \Delta_1^{\mathfrak{g}}$, then $(\phi \circ \mathrm{proj})(X) = \phi(X)$. Since by assumption we also have $\phi(X) \in \Delta_1^{\mathfrak{h}}$, then $(\mathrm{proj} \circ \phi)(X) = \phi(X)$. So $\mathrm{proj} \circ \phi$ and $\phi \circ \mathrm{proj}$ are two homomorphisms that coincide on $\Delta_1^{\mathfrak{g}}$. Since $\Delta_1^{\mathfrak{g}}$ generates the algebra $\mathfrak{g}$, then the two homomorphisms are equal.

Exercise 10.5.11 Let G be a nilpotent simply connected Lie group polarized by Δ so that $\Delta_1 \oplus [\mathfrak{g}, \mathfrak{g}] = \mathfrak{g}$. The projection map $\pi := \pi_{\Delta_1}$ from (10.4) has the following properties: For every Lipschitz curve σ in Δ_1 with $\sigma(0) = 0$, there exists a unique Lipschitz horizontal curve γ with $\pi(\gamma) = \sigma$ and $\gamma(0) = 1_G$, and such a curve is the solution of the ODE

$$\begin{cases} \dot{\gamma}(t) = (L_{\gamma(t)})_* \dot{\sigma}(t) \\ \gamma(0) = 1_G. \end{cases} \tag{10.17}$$

Solution. We discussed the existence and the uniqueness of the ODE in the proof of Proposition 7.1.3. Let $\gamma(t)$ be the solution. Then

$$\gamma'(t) = (L_{\gamma(t)})^* \dot{\gamma}(t) = \frac{\mathrm{d}}{\mathrm{d}t}(\sigma(t)).$$

Because of Formula (10.5), we have that $\pi \circ \gamma$ and σ are two curves in Δ_1 with same starting point $\pi(\gamma(0)) = 0 = \sigma(0)$ and same derivative: $\frac{\mathrm{d}}{\mathrm{d}t}(\pi \circ \gamma) = \frac{\mathrm{d}}{\mathrm{d}t}\sigma$. Therefore $\pi \circ \gamma = \sigma$.

Exercise 10.5.12 Let G and H be sub-Finsler Lie groups and $\varphi : G \to H$ a Lie group homomorphism that is also a submetry. If geodesics in G are analytic, then so are those in H.

Hint. See Proposition 3.1.29.

Exercise 10.5.13 Let G be a nilpotent simply connected Lie group and let Δ_1 be a polarization on G such that (10.2) holds. If $\sigma : I \to \Delta_1$ is a curve of regularity C^k, for some $k \in \mathbb{N}$, then the multiplicative integral of σ is C^k.

Hint. Consider the free-nilpotent Lie group $\tilde{G}$ with same rank and step of G, so that there is a quotient map $\tilde{G} \to G$. In this situation, it is clear that the multiplicative integral of σ in $\tilde{G}$ is C^k because the system (7.4) can be integrated one stratum at a time. The multiplicative integral of σ in G is the projection of the one in $\tilde{G}$.

Exercise 10.5.14 Let G be a 2-step nilpotent simply connected group equipped with a polarization $\Delta \subset \mathfrak{g}$. Define $V_2 := [\mathfrak{g}, \mathfrak{g}]$ and $V_1 \subset \Delta$ as the orthogonal of $V_2 \cap \Delta$ in Δ. Then, the polarized group (G, Δ_1) is isomorphic to some G_q as in Definition 7.3.7, with $q : V_1 \times V_1 \to V_2$ being the Lie bracket.

Exercise 10.5.15 Given a Lie algebra $\mathfrak{g}$ and $v \in \mathfrak{g}$, the transformation ad_v is diagonalizable over $\mathbb{C}$ and its eigenvalues are purely imaginary if and only if $\{e^{t\,\mathrm{ad}_v} : t \in \mathbb{R}\}$ is precompact in $\mathrm{GL}(\mathfrak{g})$.

Exercise 10.5.16 (Killing Form) Let $\mathfrak{g}$ be a Lie algebra. The *Killing form* on $\mathfrak{g}$ is defined for all $X, Y \in \mathfrak{g}$ as

$$B(X, Y) := \mathrm{Trace}(\mathrm{ad}_X \mathrm{ad}_Y).$$

The Killing form B is bilinear, symmetric, and $B(\mathrm{ad}_X Y, Z) + B(Y, \mathrm{ad}_X Z) = 0$, for all $X, Y, Z \in \mathfrak{g}$ and invariant under automorphisms of the algebra $\mathfrak{g}$, that is, $B(\psi(X), \psi(Y)) = B(X, Y)$ for $\psi \in \mathrm{Aut}_{\mathrm{Lie}}(\mathfrak{g})$, and $X, Y \in \mathfrak{g}$.

Exercise 10.5.17 The Killing form of each nilpotent Lie algebra is identically zero—be aware that the inverse is not true already for some 3D Lie algebras.

Exercise 10.5.18 (Nilradical) The *nilradical* of a Lie algebra $\mathfrak{g}$, denoted by $\mathrm{nil}(\mathfrak{g})$, is defined as the largest nilpotent ideal of $\mathfrak{g}$. Then, the set $\mathrm{nil}(\mathfrak{g})$ can also be defined as the sum of all nilpotent ideals of $\mathfrak{g}$; see [HN12, Definition 5.2.10].

Exercise 10.5.19 Let X be an element in $\mathrm{nil}(\mathfrak{g})$ of a Lie algebra $\mathfrak{g}$. Then ad_X is a nilpotent transformation of $\mathfrak{g}$.

Exercise 10.5.20 Given a Lie algebra $\mathfrak{g}$, the Killing form restricted to the nilradical of $\mathfrak{g}$ is identically 0.
Hint. From Exercise 10.5.19, for $X \in \mathrm{nil}(\mathfrak{g})$, we have $B(X, X) = 0$.

Exercise 10.5.21 (Solvable Lie Algebra) Let $\mathfrak{g}$ be a Lie algebra. We have the following equivalent definitions:

(i) $\mathfrak{g}$ is *solvable*, in the sense that, defined the *derived series* of $\mathfrak{g}$ as $\mathfrak{g}^{(0)} := \mathfrak{g}$ and

$$\mathfrak{g}^{(n)} := [\mathfrak{g}^{(n-1)}, \mathfrak{g}^{(n-1)}], \quad \forall n \in \mathbb{N},$$

then the derived series terminates in the zero subalgebra.

(ii) $[\mathfrak{g}, \mathfrak{g}]$ is nilpotent.

(iii) The Killing form B of $\mathfrak{g}$ satisfies $B(X, Y) = 0$ for all $X \in \mathfrak{g}$ and $Y \in [\mathfrak{g}, \mathfrak{g}]$. This is *Cartan's criterion for solvability*.

Hint. Check [Kna02, page 31] and [Kna02, Proposition 1.39].

Exercise 10.5.22 Let M_1 and M_2 be groups. Suppose $F : M_1 \to M_2$ is a map such that $F \circ M_1^L \circ F^{-1} = M_2^L$, where M^L denote the group of left translations on a group M. Then F is affine.

Exercise 10.5.23 Let $|\cdot| : G \to \mathbb{R}$ be a homogeneous quasi-norm on a positively graded Lie group G. Then, the function $|\cdot| : G \to \mathbb{R}$ is proper, i.e., the pre-image of compact sets is compact. It induces the same topology, in the sense that $g \to \hat{g}$

in G if and only if $|\hat{g}g^{-1}| \to 0$. Moreover, there exists a constant $M \geq 1$ such that $|gh| \leq M(|g| + |h|)$, for $g, h \in G$.

Exercise 10.5.24 Let G be a nilpotent simply connected Lie group whose Lie algebra is equipped with a linear grading. Let $|\cdot|$ and $\|\cdot\|$ be homogeneous quasi-norms with respect to the grading. Then there exists a constant $C > 1$ such that

$$\frac{1}{C}|g| \leq \|g\| \leq C|g|, \qquad \forall g \in G. \tag{10.18}$$

Hint. For each inequality, take as constant the maximum of a quasi-norm on the unit ball of the other quasi-norm. Then, use homogeneity.

Exercise 10.5.25 For every neighborhood Ω of 1 in a nilpotent simply connected Lie group G, the set

$$\Omega' := \{e_1 \cdots e_c : \forall i = 1, 2, \ldots, c \ \exists\, u_{i1}, \ldots, u_{is} \in \Omega \colon e_i = [u_{i1}, \ldots, u_{is}]\}$$

is a neighborhood of 1 in $C^s(G)$, where $c := \dim C^s(G)$ and s is the nilpotency step of G.
Solution. We work in exponential coordinates, identifying the Lie algebra $\mathfrak{g}$ with G, and $C^s(\mathfrak{g})$ with $C^s(G)$. Being $C^s(\mathfrak{g})$ a c-dimensional vector space, for $i \in \{1, \ldots, c\}$, there are $g_{i1}, \ldots, g_{is} \in \mathfrak{g}$ such that $[g_{11}, \ldots, g_{1s}], \ldots, [g_{c1}, \ldots, g_{cs}]$ form a basis of $\mathfrak{g}^s$. Up to shrinking Ω, we assume that Ω is a convex and symmetric neighborhood of 0 in $\mathfrak{g}$. We choose $\varepsilon > 0$ sufficiently small so that $tg_{ij} \in \Omega$, for all $i \in \{1, \ldots, c\}$, $j \in \{1, \ldots, s\}$ and all $t \in (-\varepsilon, \varepsilon)$. Define $u_{ij} := \varepsilon g_{ij} \in \Omega$ and the map

$$\begin{aligned} \psi \colon [-1, 1]^c &\to \mathfrak{g}^s \\ (t_1, \ldots, t_c) &\mapsto \sum_{i=1}^{c} t_i [u_{i1}, \ldots, u_{is}]. \end{aligned}$$

The image of ψ is contained in Ω', since, assuming $t_1, \ldots, t_c > 0$ and leaving the general case to the reader, we have

$$\begin{aligned} \sum_{i=1}^{c} t_i [u_{i1}, \ldots, u_{is}] &= \sum_{i=1}^{c} [t_i^{1/s} u_{i1}, \ldots, t_i^{1/s} u_{is}] \\ &= [t_1^{1/s} u_{11}, \ldots, t_1^{1/s} u_{1s}] \star \cdots \star [t_c^{1/s} u_{c1}, \ldots, t_c^{1/s} u_{cs}] \in \Omega', \end{aligned}$$

where in the final inclusion we use the fact that $t_i^{1/s} u_{ij} \in \Omega$ for every i and j. In addition, the image of the differential of ψ at the identity is the span of $[u_{11}, \ldots, u_{1s}], \ldots, [u_{c1}, \ldots, u_{cs}]$, hence $(\mathrm{d}\psi)_0$ is surjective since these vectors form a basis of $\mathfrak{g}^s$. It follows that there exists an open neighborhood $U \subset [0, 1]^c$ of

0 such that $\psi(U) \subset \Omega'$ is an embedded c-dimensional submanifold of $\mathfrak{g}^s$, hence it is open.

Exercise 10.5.26 Let G be a nilpotent group of step s and $m_1, \dots, m_s \in \mathbb{N}$. Then

$$[x_1^{m_1}, \dots, x_s^{m_s}] = [x_1, \dots, x_s]^{m_1 \cdots m_s}, \qquad \forall x_1, \dots, x_s \in G.$$

Solution. We begin by showing that if $[x_1, x_2]$ commutes with x_1 and x_2, then for every integers m_1, m_2 we have

$$[x_1^{m_1}, x_2^{m_2}] = [x_1, x_2]^{m_1 m_2}. \tag{10.19}$$

Indeed, we have

$$\begin{aligned}
[x_1, x_2]^m &= x_1 x_2 x_1^{-1} x_2^{-1} [x_1, x_2]^{m-1} \\
&= x_1 [x_1, x_2] x_2 x_1^{-1} x_2^{-1} [x_1, x_2]^{m-2} \\
&= x_1 x_1 x_2 x_1^{-1} x_2^{-1} x_2 x_1^{-1} x_2^{-1} [x_1, x_2]^{m-2}. \\
&= x_1^2 x_2 x_1^{-2} x_2^{-1} [x_1, x_2]^{m-2} \\
&= x_1^2 [x_1, x_2] x_2 x_1^{-2} x_2^{-1} [x_1, x_2]^{m-3} \\
&= x_1^3 x_2 x_1^{-3} x_2^{-1} [x_1, x_2]^{m-3} \\
&= \dots \\
&= x_1^m x_2 x_1^{-m} x_2^{-1} = [x_1^m, x_2].
\end{aligned}$$

As a consequence, we obtain

$$[x_1^{m_1}, x_2^{m_2}] = [x_1, x_2^{m_2}]^{m_1} = [x_2^{m_2}, x_1]^{-m_1} = [x_2, x_1]^{-m_1 m_2} = [x_1, x_2]^{m_1 m_2}.$$

We solve the exercise by induction on the nilpotency step of G. In the abelian case, the result is trivial. Now assume that the result holds for groups up to step $s - 1$. Hence, the result is valid in ${G}/{C^s(G)}$ which has step $s - 1$. Thus, for every $x_2, \dots, x_s \in G$ and every $m_2, \dots, m_s \in \mathbb{N}$ we have

$$[x_2^{m_2} C^s(G), \dots, x_s^{m_s} C^s(G)] = [x_2 C^s(G), \dots, x_s C^s(G)]^{m_2 \cdots m_s}.$$

It follows that

$$[x_2^{m_2}, \dots, x_s^{m_s}] C^s(G) = [x_2, \dots, x_s]^{m_2 \cdots m_s} C^s(G),$$

and hence there exists a central element $z \in C^s(G)$ such that

$$[x_2^{m_2}, \dots, x_s^{m_s}] = [x_2, \dots, x_s]^{m_2 \cdots m_s} z. \tag{10.20}$$

Let $x_1 \in G$ and $m_1 \in \mathbb{N}$, then

$$\begin{aligned}[x_1^{m_1}, x_2^{m_2}, \ldots, x_s^{m_s}] &\overset{(10.20)}{=} [x_1^{m_1}, [x_2, \ldots, x_s]^{m_2 \cdots m_s} z] \\ &= \ [x_1^{m_1}, [x_2, \ldots, x_s]^{m_2 \cdots m_s}] \\ &\overset{(10.19)}{=} [x_1, x_2, \ldots, x_s]^{m_1 \cdots m_s},\end{aligned}$$

where in the second equality we use that for $x, y \in G$ and $z \in C^s(G)$ we have $[x, yz] = [x, y]$.

Extra Exercises:

Exercise 10.5.27 (Radical) Every Lie algebra $\mathfrak{g}$ has a maximal solvable ideal, and it is unique. This ideal is called the *radical* (also called *solvable radical* and denoted by $\mathrm{rad}(\mathfrak{g})$) of $\mathfrak{g}$.

Exercise 10.5.28 (Simple Lie Algebra) By definition, a *simple* Lie algebra is a non-abelian Lie algebra whose only ideals are 0 and itself. We have that the word 'non-abelian' can be replaced with 'dimension at least 2'.

Exercise 10.5.29 (Semisimple) A Lie algebra $\mathfrak{g}$ is called *semisimple* if $[\mathfrak{g}, \mathfrak{g}] = \mathfrak{g}$. We have that a Lie algebra is semisimple if and only if it is a direct sum of simple Lie algebras.

Exercise 10.5.30 (Criteria for Semi-Simplicity) Let $\mathfrak{g}$ be a Lie algebra. We have the following equivalent definitions:

(i) $\mathfrak{g}$ is semisimple, as in Exercise 10.5.29;
(ii) $\mathfrak{g}$ has no non-zero abelian ideals;
(iii) The solvable radical $\mathrm{rad}(\mathfrak{g})$ is zero;
(iv) ☠ The Killing form B of $\mathfrak{g}$ is non-degenerate. This is *Cartan's criterion for semi-simplicity*.

Hint. See [Kna02, Section I.7].

Exercise 10.5.31 (Levi Decomposition) ☠ Every Lie algebra $\mathfrak{g}$ is the semidirect product of a solvable ideal and a semisimple subalgebra. In fact, then there exists a subalgebra $\mathfrak{l}$, called a *Levi factor*, that is semisimple and for which $\mathfrak{g} = \mathrm{rad}(\mathfrak{g}) \rtimes \mathfrak{l}$. *Hint.* See [Kna02, Appendix B, Section 1].

Exercise 10.5.32 ☠ Given a Lie algebra $\mathfrak{g}$, consider its nilradical $\mathrm{nil}(\mathfrak{g})$ and its radical $\mathrm{rad}(\mathfrak{g})$. Then the Killing form B satisfies:

10.5.32.i. $B(\mathrm{rad}(\mathfrak{g}), [\mathfrak{g}, \mathfrak{g}]) = \{0\}$; (See [Bou98, I.5.5, p.48, Prop.5b])
10.5.32.ii. $B(\mathrm{nil}(\mathfrak{g}), \mathfrak{g}) = \{0\}$. (See [Bou98, I.4.4, p.42, Prop.6b])

Exercise 10.5.33 Let $\mathfrak{g}$ be a Lie algebra with radical $\mathrm{rad}(\mathfrak{g})$ and nilradical $\mathrm{nil}(\mathfrak{g})$. Then, we have that $[\mathfrak{g}, \mathrm{rad}(\mathfrak{g})] \subset \mathrm{nil}(\mathfrak{g})$, and consequently, every vector space V with $\mathrm{nil}(\mathfrak{g}) \subset V \subset \mathrm{rad}(\mathfrak{g})$ is an ideal of $\mathfrak{g}$.

Exercise 10.5.34 ☠ Let G be a connected semisimple Lie group. Then G is of type (R) if and only if it is compact.
Hint. Consult [DER03, Proposition II.4.8.III], or [Jen73b], [Jen73a], [Bre14].

Exercise 10.5.35 In the proof of Theorem 6.5.1, we have that the nilradical of the connected component of $\mathrm{Isom}(M, d)$ acts almost simply transitive on M, i.e., orbits are open immersions.

Chapter 11
Carnot Groups

Carnot groups are specific examples of Carnot-Carathéodory spaces. They are simply connected nilpotent Lie groups whose Lie algebras admit particular gradings: stratifications. The polarizations correspond to the first layers of stratifications with norms on them. Every such sub-Finsler Lie group is self-similar with respect to its Carnot-Carathéodory metric. Specifically, there is a natural family of metric dilations.

This chapter starts with Sect. 11.1, where we discuss the definition of Carnot groups, Carnot bases, and examples. In Sect. 11.2, we present some fundamental consequences of dilations, such as the presence of good coordinates, an easy proof of the Ball-Box Theorem, and properties of Haar measures.

Section 11.3 is dedicated to the Pansu-Rademacher Differentiation Theorem for Lipschitz maps between Carnot groups. In Sect. 11.4, we provide a complete characterization of the metric spaces that are isometric to Carnot groups, entirely in terms of metric geometry.

In Sect. 11.5, we explore extremal curves in Carnot groups. We outline some properties of abnormal curves and of normal geodesics. We also present a sublinear isometric property of projections of geodesics, which shows that geodesics cannot form corners. Additionally, we discuss some open problems.

Lastly, we include supplementary material: Sect. 11.7.1 on the Lie coset structure of self-similar sub-Finsler spaces, which are submetric images of Carnot groups; and Sect. 11.7.2 on isometries between open sets in Carnot groups.

11.1 Definition of Carnot Groups

Carnot groups are sub-Finsler Lie groups (as in Definition 7.1.6) such that the polarization is the first layer of a stratification (as in Definition 9.2.4). We spell out this definition to have a slightly more self-contained presentation.

E. Le Donne, *Metric Lie Groups*, Graduate Texts in Mathematics 306,
https://doi.org/10.1007/978-3-031-98832-5_11

Let G be a simply connected Lie group. Assume its Lie algebra $\mathfrak{g} := \mathrm{Lie}(G)$ admits a stratification $\mathfrak{g} = V_1 \oplus \cdots \oplus V_s$, i.e., $V_1, \ldots, V_s$ are vector subspaces in direct sum such that $V_{j+1} = [V_j, V_1]$ with $V_{s+1} := \{0\}$. We recall from Sect. 9.2.2 that each stratification gives a Lie algebra grading:

$$[V_i, V_j] \subseteq V_{i+j}, \qquad \forall i, j \in \mathbb{N},$$

where $V_j := \{0\}$, for $j > s$; see Exercises 9.5.21. Moreover, the layer V_1 of degree 1 is bracket generating; see Remark 9.2.5.

The vector space V_1 is called the *horizontal stratum* and is seen as a subset of the tangent space $T_{1_G}G$ of G at the identity element 1_G of G. As in (7.1), it induces a left-invariant subbundle Δ, called the *horizontal bundle*, of the tangent bundle TG:

$$\Delta_g := (L_g)_* V_1, \qquad \forall g \in G. \tag{11.1}$$

Fix a norm $\|\cdot\|$ on the vector space V_1. As in (7.5), the norm on V_1 induces a norm on every Δ_g as

$$\|v\| := \|(L_g)^* v\|, \qquad \forall v \in \Delta_g, \quad \forall g \in G. \tag{11.2}$$

Here, we use the notation $(L_g)_*$ for the pushforward by the differential of the left translation by g and $(L_g)^*$ for the pullback by this differential. The triple $(G, \Delta, \|\cdot\|)$ is a Carnot-Carathéodory space, and indeed, a sub-Finsler Lie group, which has an induced distance function as in Definition 4.1.7:

$$\begin{aligned} d_{cc}(p, q) &:= d_{V_1, \|\cdot\|}(p, q) \\ &:= \inf\left\{ \int \|\dot\gamma\| : \gamma \text{ AC curve from } p \text{ to } q, \text{ with } \dot\gamma \in \Delta \right\}, \quad \forall p, q \in G. \end{aligned} \tag{11.3}$$

Definition 11.1.1 (Carnot Group) Let G be a simply connected Lie group whose Lie algebra admits a stratification. Given the first stratum V_1 of a stratification of $\mathrm{Lie}(G)$ and a norm on it, let Δ and $\|\cdot\|$ be defined by (11.1) and (11.2), respectively. Let d_{cc} be the Carnot-Carathéodory distance associated with Δ and $\|\cdot\|$ as in (11.3). Both the sub-Finsler manifold $(G, \Delta, \|\cdot\|)$ and the metric space (G, d_{cc}) are called *Carnot groups*. In accordance with Definition 9.2.4, we call *Carnot algebras* the Lie algebras of Carnot groups.

Given a stratification $\mathrm{Lie}(G) = V_1 \oplus \cdots \oplus V_s$ of the Lie algebra of a Carnot group G, with $V_s \neq \{0\}$, the number s is called the *step* of the stratification of G, or simply, the *step* of G. The number $\dim V_1$ is called *rank* of G. The topological dimension of G is $n := \sum_i \dim V_i$, and the *homogeneous dimension* is

$$Q := \sum_{i=1}^{s} i \dim V_i. \tag{11.4}$$

Each Carnot group $(G, \Delta, \|\cdot\|)$ is indeed a sub-Finsler Lie group and an equiregular Carnot-Carathéodory space of step s. Indeed, one has that, for each $k \in \{1, \dots, s\}$, the subset $\Delta^{[k]}$ in the flag of subbundles for Δ as in Definition 4.1.15 is the left-invariant subbundle for which

$$\Delta^{[k]}(1_G) = V_1 \oplus \cdots \oplus V_k.$$

Because of Proposition 7.1.8, other choices of norms would not change the bi-Lipschitz equivalence class of the CC metric.

11.1.1 Dilations on Carnot Groups

As in every $\mathbb{R}$-graded Lie group, in Carnot groups we have a canonical one-parameter family of dilations on the Lie algebras and on the groups; recall the discussion on page 273:

Definition 11.1.2 (Dilations on Stratified Groups) Let G be a Carnot group. Let $\delta_\lambda : \mathrm{Lie}(G) \to \mathrm{Lie}(G)$ be the dilation of factor λ, with $\lambda \in \mathbb{R}$, associated with the stratification as in (9.8). Then the *dilation of factor* λ *on the group* G is the map $\delta_\lambda : G \to G$ that is the only group automorphism such that $(\delta_\lambda)_* = \delta_\lambda$. Such maps are also called the *intrinsic dilations* of the Carnot group or *Carnot dilations*.

The notation δ_λ is used consistently for both dilations on the Lie algebra and the group because no ambiguity will arise since the two maps have different domains.

On Carnot groups, the intrinsic dilations satisfy the following formulas:

$$\delta_\lambda \circ \exp = \exp \circ \delta_\lambda, \qquad \forall \lambda \in \mathbb{R}; \tag{11.5}$$

$$\delta_\lambda \circ \delta_\eta = \delta_{\lambda\eta}, \qquad \forall \lambda, \eta \in \mathbb{R}; \tag{11.6}$$

$$\delta_\lambda(xy) = \delta_\lambda(x)\delta_\lambda(y) \qquad \forall x,\ y \in G, \forall \lambda \in \mathbb{R}. \tag{11.7}$$

11.1.1.1 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 11.1.3 *If* (G, d_{cc}) *is a Carnot group with dilations* $(\delta_\lambda)_{\lambda\in\mathbb{R}}$*, then*

$$d_{cc}(\delta_\lambda p, \delta_\lambda q) = |\lambda|\, d_{cc}(p, q), \qquad \forall p,\ q \in G, \forall \lambda \in \mathbb{R}. \tag{11.8}$$

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,

$$(\delta_\lambda)_* \dot{\gamma}(t) = \delta_\lambda(\dot{\gamma}(t)) = \lambda \dot{\gamma}(t), \tag{11.9}$$

which are horizontal since $\dot{\gamma}(t)$ is horizontal. Moreover, from (11.9), the length of $\delta_\lambda \circ \gamma$ is $|\lambda|$ times the length of γ, i.e., for every horizontal curve γ,

$$\mathrm{Length}_{\|\cdot\|}(\delta_\lambda \circ \gamma) = |\lambda|\ \mathrm{Length}_{\|\cdot\|}(\gamma).$$

Thus, by (11.3) we get (11.8). □

11.1.2 Good Bases for Carnot Groups

Let 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 $X_1, \ldots, X_m$ of V_1. Then consider all brackets $[X_i, X_j]$, for $i, j \in \{1, \ldots, m\}$. Since $[V_1, V_1] = V_2$, we can find among such brackets a basis for V_2; see Exercise 11.8.8. Pick one such basis and call its elements $X_{m+1}, \ldots, X_{m_2}$. Iterate the method: extract a basis $X_{m_2+1}, \ldots, X_{m_3}$ of V_3 from the set $[X_i, X_j]$, for $i \in \{1, \ldots, m\}$, $j \in \{m+1, \ldots, m_2\}$. And so on. We have constructed a basis $X_1, \ldots, X_n$ of $\mathfrak{g}$ and natural numbers $m_1 \ldots, m_s$ such that

1. $X_{m_{j-1}+1}, \ldots, X_{m_j}$ is a basis of V_j,
2. For every $i \in \{m+1, \ldots, n\}$, 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

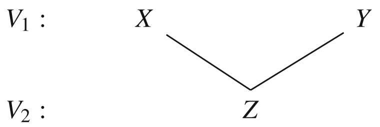
$$X_i = [X_{l_i}, X_{k_i}]. \tag{11.10}$$

3. The order-reversed basis $(X_n, \ldots, X_1)$ is a Malcev basis as in Definition 9.4.10; in other words,

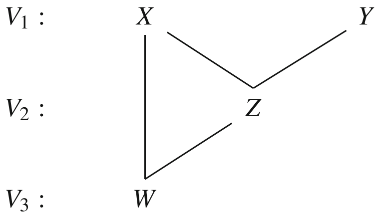
$$[\mathfrak{g}, \mathrm{span}\{X_k, \ldots, X_n\}] \subseteq \mathrm{span}\{X_{k+1}, \ldots, X_n\}, \qquad \forall k \in \{1, \ldots, n\}.$$

We suggest the terminology '*Carnot basis*' for a basis satisfying the above three conditions. The reader should notice that the above property 1 implies property 3. See Exercise 11.8.9.

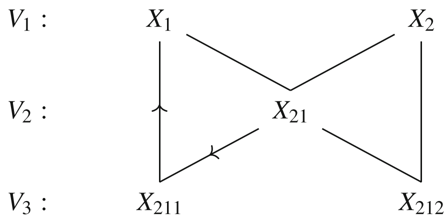
To describe a Carnot algebra, we prefer to give a Carnot basis with a hierarchical diagram as follows:



for the 3D Heisenberg Lie algebra;



for the Engel algebra (the 3-step filiform algebra);



for the Cartan Lie algebra, the free nilpotent Lie algebra of rank 2 and step 3.

As explained on page 247, in the diagrams, the black lines express the non-trivial brackets. We obtain the bracket relations by reading the arms *from left to right* unless there is an arrow, in which case they are read *from right to left*. The j-th line in the diagram lists the vectors that span the stratum V_j. In the Lie algebra structure, there might be more relations than just those in (11.10), as, for example, in the first quaternionic Heisenberg group; see page 333. In [LT22], there are several other uses of these diagrams to represent Carnot groups in low dimensions.

11.1.3 Examples of Carnot Groups and Carnot Algebras

Let G be a Carnot group with Lie algebra $\mathfrak{g}$. Given a basis $X_1, \ldots, X_n$ of $\mathfrak{g}$, we will use the identification

$$\begin{aligned} \mathbb{R}^n &\longleftrightarrow G \\ (x_1, \ldots, x_n) &\longmapsto \exp\left(\sum_{i=1}^{n} x_i X_i\right). \end{aligned} \tag{11.11}$$

This identification allows us to write the group product using the Dynkin product (5.24). In fact, for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, there exists a unique $\mathbf{z} \in \mathbb{R}^n$ such that

$$\exp\Big(\sum_{i=1}^{n} x_i X_i\Big) \exp\Big(\sum_{i=1}^{n} y_i X_i\Big) = \exp\Big(\sum_{i=1}^{n} z_i X_i\Big). \tag{11.12}$$

Via the identification (11.11), one can write the group law in (11.12) as

$$\mathbf{x} \star \mathbf{y} = \mathbf{z}. \tag{11.13}$$

Hence, we have a group law $\star$ on $\mathbb{R}^n$ that makes $(\mathbb{R}^n, \star)$ a simply connected Lie group with Lie algebra $\mathfrak{g}$, whose identity element is $\mathbf{0}$.

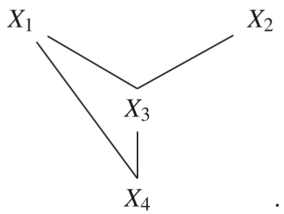
11.1.3.1 Carnot Groups in Dimension at Most 4

The Carnot groups of topological dimension between 1 and 3 are the commutative groups $\mathbb{R}$, $\mathbb{R}^2$, and $\mathbb{R}^3$, together with the Heisenberg group N_3.

In dimension 4, there are only the following nilpotent simply connected Lie groups, which are all stratifiable: $\mathbb{R}^4$, $\mathbb{R} \times N_3$, and $N_{4,2}$, where and $N_{4,2}$ is the *Engel group* as we now recall. The *Engel Lie algebra* is spanned by 4 vectors $X_1, \ldots, X_4$ with only non-trivial brackets

$$[X_1, X_2] = X_3\,,\ [X_1, X_3] = X_4\,. \tag{11.14}$$

This is a nilpotent Lie algebra of rank 2 and step 3 that is stratifiable. It is also known as the filiform Lie algebra of dimension 4; see Exercise 9.1.7. The Lie brackets can be pictured with the diagram:



The Lie group $N_{4,2}$ is the only simply connected Lie group with such a Lie algebra. The group law (11.12) of the Engel group in exponential coordinates is given by:

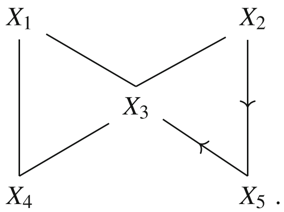
$$\mathbf{x}\star\mathbf{y} = \mathbf{z} \iff \begin{cases} z_1 = x_1 + y_1 \\ z_2 = x_2 + y_2 \\ z_3 = x_3 + y_3 + \frac{1}{2}(x_1 y_2 - x_2 y_1) \\ z_4 = x_4 + y_4 + \frac{1}{2}(x_1 y_3 - x_3 y_1) + \frac{1}{12}(x_1 - y_1)(x_1 y_2 - x_2 y_1) \end{cases}. \tag{11.15}$$

11.1.3.2 Cartan Group $\mathbb{F}_{2,3}$

Cartan Lie algebra is another name for the free-nilpotent Lie algebra of step 3 and 2 generators. The simply connected Lie group with this Lie algebra is known as the *Cartan group*, and it is sometimes denoted by $\mathbb{F}_{2,3}$ or by $N_{5,2,3}$ as in [Gon98]. With respect to some basis $X_1, \ldots, X_5$, the non-trivial brackets are the following:

$$[X_1, X_2] = X_3\,,\ [X_1, X_3] = X_4\,,\ [X_2, X_3] = X_5\,.$$

The Lie brackets can be pictured with the diagram:



The group law (11.12) of the Cartan group in exponential coordinates is given by:

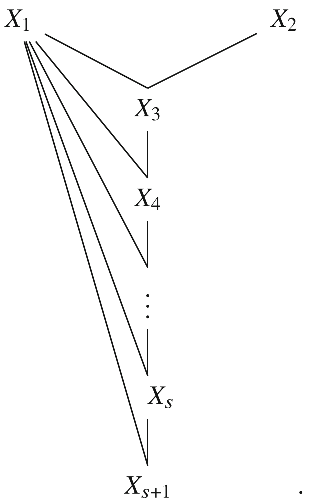
$$\mathbf{x}\star\mathbf{y} = \mathbf{z} \iff \begin{cases} z_1 = x_1 + y_1 \\ z_2 = x_2 + y_2 \\ z_3 = x_3 + y_3 + \frac{1}{2}(x_1y_2 - x_2y_1) \\ z_4 = x_4 + y_4 + \frac{1}{2}(x_1y_3 - x_3y_1) + \frac{1}{12}(x_1 - y_1)(x_1y_2 - x_2y_1) \\ z_5 = x_5 + y_5 + \frac{1}{2}(x_2y_3 - x_3y_2) + \frac{1}{12}(x_2 - y_2)(x_1y_2 - x_2y_1) \end{cases} . \tag{11.16}$$

11.1.3.3 Filiform Groups

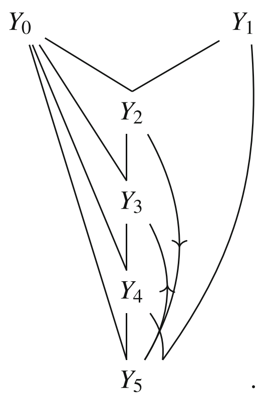
Filiform groups are those simply connected Lie groups whose Lie algebra is filiform, in the sense that the Lie algebra is dimensionwise the smallest among those Lie algebras with the same nilpotency step. We saw in Example 9.1.7 those of the first kind. With respect to some basis $X_1, \ldots, X_{s+1}$, with $s \in \mathbb{N}$, the non-trivial brackets of the $(s+1)$-dimensional example of the first kind are the following:

$$[X_1, X_i] = X_{i+1}\,, \qquad \text{for } i \in \{2, \ldots, s\}.$$

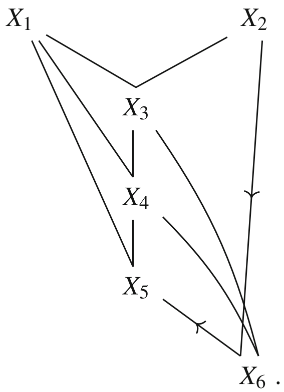
This is a nilpotent Lie algebra of rank 2 and step s that is stratifiable. The Lie brackets can be pictured with the diagram:



In [Ver70], M. Vergne classified all filiform Lie algebras. In addition to the ones of the first type, in each even dimension starting from dimension 6, there is exactly one more filiform Lie algebra, called *filiform Lie algebra of the second kind*. In dimension 6, we have the example discussed in Exercise 9.5.17. With respect to that basis $Y_0, Y_1, Y_2, Y_3, Y_4, Y_5$, the bracket diagram is the following:

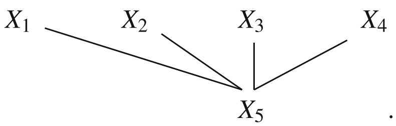


Another presentation (see $N_{6,2,2}$ in [Gon98]), in a basis $X_1, \ldots, X_6$ is given by the diagram:



11.1.3.4 The Second Heisenberg Group

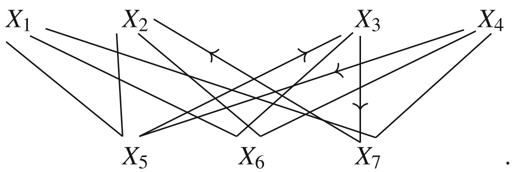
The Lie brackets of the second Heisenberg Lie algebra can be pictured with the diagram:



We will encounter this group again in Chap. 13 discussing complex hyperbolic spaces; see Exercise 13.4.13. It will be called the 2-nd $\mathbb{C}$-*Heisenberg group* and will be denoted by $\mathcal{N}_2^{\mathbb{C}}$.

11.1.3.5 The First Quaternionic Heisenberg Group

The Lie algebra of the first quaternionic Heisenberg group $\mathcal{N}_1^{\mathbb{H}}$ can be characterized as the only 7D Carnot algebra of rank 4 and step 2 where every element in the first stratum has maximal rank, i.e., for every nonzero X in the first stratum, the map ad_X is surjective. The Lie brackets can be pictured with the diagram:



11.1.3.6 Some Free-Carnot Groups: $\mathbb{F}_{n2}$, $\mathbb{F}_{24}$, $\mathbb{F}_{25}$, and $\mathbb{F}_{33}$

Free-nilpotent Lie algebras are stratifiable. Hence, the associated simply connected Lie groups are the free objects in the category of Carnot groups. Here are some examples:

The free-Carnot group $\mathbb{F}_{n2}$ of rank n and step 2 has been discussed in Example 9.1.6, recall also Example 9.3.8. Setwise $\mathbb{F}_{n2}$ is $\Lambda^1(\mathbb{R}^n) \oplus \Lambda^2(\mathbb{R}^n)$, while the group law is $x \cdot y = x + y + \frac{1}{2} x \wedge y$.

The Lie group $\mathbb{F}_{24}$ is the Carnot group whose Lie algebra is free-nilpotent with 2 generators and nilpotency step 4. It has dimension 8. The non-trivial brackets in some basis $X_1, \dots, X_8$ are:

$$[X_1, X_2] = X_3\,,\ [X_1, X_3] = X_4\,,\ [X_2, X_3] = X_5\,,$$
$$[X_1, X_4] = X_6\,,\ [X_1, X_5] = [X_2, X_4] = X_7\,,\ [X_2, X_5] = X_8\,.$$

The Lie group $\mathbb{F}_{25}$ is the Carnot group whose Lie algebra is free-nilpotent with 2 generators and nilpotency step 5. It has dimension 14. The non-trivial brackets in some basis $X_1, \dots, X_{14}$ are:

$$[X_1, X_2] = X_3\,,\ [X_1, X_3] = X_4\,,\ [X_2, X_3] = X_5\,,\ [X_1, X_4] = X_6\,,$$
$$[X_2, X_5] = X_8\,,\ [X_1, X_5] = [X_2, X_4] = X_7\,,\ [X_1, X_7] = X_{10} + X_{13}\,,$$
$$[X_1, X_6] = X_9\,,\ [X_1, X_8] = X_{11} + X_{14}\,,\ [X_2, X_6] = X_{10}\,,$$
$$[X_2, X_7] = X_{11}\,,\ [X_2, X_8] = X_{12}\,,\ [X_3, X_4] = X_{13}\,,\ [X_3, X_5] = X_{14}\,.$$

The Lie group $\mathbb{F}_{33}$ is the Carnot group whose Lie algebra is free-nilpotent with 3 generators and nilpotency step 3. It has dimension 14. The non-trivial brackets in some basis $X_1, \dots, X_{14}$ are:

$$[X_1, X_2] = X_4\,,\ [X_1, X_3] = X_5\,,\ [X_2, X_3] = X_6\,,\ [X_1, X_4] = X_7\,,$$
$$[X_1, X_5] = X_8\,,\ [X_1, X_6] = X_9\,,\ [X_2, X_4] = X_{10}\,,\ [X_2, X_6] = X_{11}\,,$$
$$[X_3, X_4] = X_{12}\,,\ [X_3, X_5] = X_{13}\,,\ [X_3, X_6] = X_{14}\,,\ [X_2, X_5] = X_9 + X_{12}\,.$$

Carnot groups are completely classified up to dimension 7. However, in dimension 7, there are infinitely many non-isomorphic classes. For a cornucopia of examples, we refer to [LT22].

11.2 Simple Consequences of Dilations

11.2.1 Canonical Coordinates

The next proposition holds in the general setting of nilpotent simply connected Lie groups; see Theorem 9.4.7. We will provide a simplified proof for Carnot groups.

Proposition 11.2.1 *Let G be a Carnot group with dilations $(\delta_\lambda)_{\lambda\in\mathbb{R}}$. Let $W_1 \oplus \cdots \oplus W_m = \mathrm{Lie}(G)$ be a direct-sum decomposition by dilation invariant subspaces. Let*

$$\begin{aligned} \Psi : \mathfrak{g} \simeq W_1 \times \cdots \times W_m &\to G, \\ (X_1, \ldots, X_m) &\mapsto \prod_{j=1}^{m} \exp(X_j). \end{aligned}$$

Then, the map $\Psi : \mathfrak{g} \to G$ gives a global coordinate system. In particular, in every Carnot group, Malcev coordinates and exponential coordinates exist globally.

Above, like in other places in the book, we used the notation

$$\prod_{j=1}^{m} g_j := g_1 \cdots g_m.$$

Proof of Proposition 11.2.1 The linear map $(\mathrm{d}\Psi)_0$ is an isomorphism, because

$$(\mathrm{d}\Psi)_0 X = X, \qquad \forall j \in \{1, \ldots, m\}, \forall X \in W_j.$$

Hence, the map $\Psi : \mathfrak{g} \to G$ is a diffeomorphism between some neighborhood of 0 in $\mathfrak{g}$ and some neighborhood of 1_G in G. By assumption, we have

$$\delta_\lambda W_j = W_j, \qquad \forall \lambda > 0, \forall j \in \{1, \ldots, m\}. \tag{11.17}$$

Consequently, we claim

$$\Psi \circ \delta_\lambda = \delta_\lambda \circ \Psi, \qquad \forall \lambda > 0. \tag{11.18}$$

Indeed, for all $(X_1, \cdots, X_m) \in W_1 \times \cdots \times W_m$, we have

$$\begin{aligned} \Psi(\delta_\lambda(X_1 + \cdots + X_m)) &= \Psi(\delta_\lambda X_1 + \cdots + \delta_\lambda X_m) \\ &\overset{(11.17)}{=} \prod_{j=1}^{m} \exp(\delta_\lambda X_j) \end{aligned}$$

$$
\begin{aligned}
&= \prod_{j=1}^{m} \delta_\lambda \exp(X_j) \\
&= \delta_\lambda \prod_{j=1}^{m} \exp(X_j) \\
&= \delta_\lambda \Psi(X_1 + \cdots + X_m).
\end{aligned}
$$

By (11.18), we deduce that Ψ is a global diffeomorphism. Regarding Malcev coordinates, recall that we proved the existence of Malcev bases in Sect. 11.1.2 (and more generally in Proposition 9.4.12). □

Group laws in exponential coordinates with respect to Carnot bases have triangular forms. This result follows from a more general result about nilpotent simply connected Lie groups. Indeed, since Carnot bases are order-reversed Malchev bases, we have the following consequence of Proposition 9.4.17. We stress that Proposition 9.4.17 was just a consequence of the BCH formula; thus, the reader should try to prove Proposition 11.2.2 directly as an exercise.

Proposition 11.2.2 *Let G be a Carnot group. Fix a Carnot basis $(X_1, \dots, X_n)$ for its Lie algebra. On G, consider exponential coordinates of the first or second kind associated with the basis. Then, in these coordinate systems, the product law has a lower triangular form:*

$$
(s_1, \dots, s_n) \cdot (t_1, \dots, t_n) = s + t + \sum_{j=1}^{n} Q_j(s,t) e_j, \qquad \forall s, t \in \mathbb{R}^n,
$$

where each Q_j is a polynomial that is not depending on $s_{j+1}, \dots, s_n$ nor on $t_{j+1}, \dots, t_n$.

Proof We have that the reversed-ordered basis $(X_n, \dots, X_1)$ is a Malcev basis. Hence, we invert the order of the coordinates: $(s_n, \dots, s_1)$. Then Proposition 9.4.17 gives the result, replacing upper triangular with lower triangular. □

11.2.2 A Proof of Ball-Box Theorem for Carnot Groups

Let G be a Carnot group with stratification $V_1, \dots, V_s$. Let $X_1, \dots, X_n$ be a basis of $\mathrm{Lie}(G)$ adapted to the stratification, hence, for all j let d_j such that $X_j \in V_{d_j}$. The *boxes* with respect to the numbers $(d_1, \dots, d_n)$ are

$$
\mathtt{Box}(r) := \left\{ (t_1, \dots, t_n) \in \mathbb{R}^n \ : \ |t_j| < r^{d_j} \right\}, \qquad \forall r \geq 0. \tag{11.19}
$$

The *dilations* in $\mathbb{R}^n$ with respect to the numbers $(d_1, \dots, d_n)$ are the maps $\delta_\lambda : \mathbb{R}^n \to \mathbb{R}^n$ defined as

$$\delta_\lambda(t_1, \dots, t_n) := (\lambda^{d_1} t_1, \dots, \lambda^{d_j} t_j, \dots, \lambda^{d_n} t_n), \qquad \forall (t_1, \dots, t_n) \in \mathbb{R}^n, \forall \lambda \in \mathbb{R}. \tag{11.20}$$

Let $\Phi : \mathbb{R}^n \to G$ be the exponential coordinate map with respect to the basis $X_1, \dots, X_n$, i.e., $\Phi(\mathbf{t}) := \exp(\sum_j t_j X_j)$. Given $p \in G$, the *exponential coordinate map from* p *with respect to* $X_1, \dots, X_n$ is $\Phi_p := L_p \circ \Phi$.

Theorem 11.2.3 (Ball-Box for Carnot Groups) *Let G be a Carnot group. Fix a basis adapted to the stratification. Then there is $C > 1$ such that*

$$B\big(p, r/C\big) \subset \Phi_p(\mathtt{Box}(r)) \subset B(p, rC), \qquad \forall p \in G, \forall r > 0, \tag{11.21}$$

where Φ_p is the exponential coordinate map from p with respect to the basis, and balls are with respect to the Carnot-Carathéodory distance.

Proof Each set $\Phi(\mathtt{Box}(r))$, with $r > 0$ is a bounded neighborhood of 1_G in G, recall Proposition 11.2.1. Let d_{cc} be the Carnot-Carathéodory distance function on the Carnot group G. By Chow Theorem 4.1.8, or alternatively see Corollary 7.1.21, the distance d_{cc} induces the manifold topology. Hence, there is $r_0, r_1 > 0$ such that

$$B(1_G, r_1) \subset \Phi(\mathtt{Box}(r_0)) \subset B(1_G, 1),$$

Recalling that $\delta_\lambda(B(1_G, r)) = B(1_G, \lambda r)$, by Proposition 11.1.3, and applying δ_λ, we get

$$B(1_G, \lambda r_1) \subset \delta_\lambda \Phi(\mathtt{Box}(r_0)) \subset B(1_G, \lambda).$$

Moreover, we have

$$\begin{aligned}
\delta_\lambda\left(\Phi(\mathtt{Box}(r_0))\right) &= \delta_\lambda\left(\Phi\left\{(t_1, \dots, t_n) : |t_j| < r_0^{d_j}\right\}\right)\\
&= \delta_\lambda\left\{\exp\left(\sum_j t_j X_j\right) : |t_j| < r_0^{d_j}\right\}\\
&= \left\{\exp\left(\delta_\lambda \sum_j t_j X_j\right) : |t_j| < r_0^{d_j}\right\}\\
&= \left\{\exp\left(\sum_j \lambda^{d_j} t_j X_j\right) : |t_j| < r_0^{d_j}\right\}
\end{aligned}$$

$$= \left\{ \exp\left(\sum_j s_j X_j \right) : |s_j| < \lambda^{d_j} r_0^{d_j} \right\}$$
$$= \Phi(\mathtt{Box}(\lambda r_0)).$$

Therefore, we deduce that

$$B(1_G, \lambda r_1) \subset \Phi(\mathtt{Box}(\lambda r_0)) \subset B(1_G, \lambda), \qquad \forall \lambda > 0. \tag{11.22}$$

Since d_{cc} is left-invariant, applying L_p to (11.22), we obtain (11.21) for all $p \in G$ and all $r > 0$. □

11.2.3 Canonical Measures

In each Carnot group, there are a few natural choices of measures: Haar, Hausdorff of top dimension, and Lebesgue measures in exponential coordinates. We introduced the Haar measures in Sect. 6.4, the Hausdorff measures in Sect. 3.1.6, and the exponential coordinates in Sect. 9.4.4. In this section, we will prove that these measures are the same up to a scalar factor and show some other properties.

Let G be a Carnot group. For every $k > 0$, let $\mathcal{H}^k$ be the k-dimensional Hausdorff measure. We can also consider the k-dimensional spherical Hausdorff measure $\mathcal{S}^k$; see [Mat95, page 75]. Because left translations are isometries, the measures $\mathcal{H}^k$ and $\mathcal{S}^k$ are left-invariant. We shall see that when k equals the homogeneous dimension Q, these measures are Radon measures, and therefore, they are Haar measures.

If we consider exponential coordinates $\exp : \mathbb{R}^n \to G$, linearly identifying the Lie algebra $\mathfrak{g}$ of G with $\mathbb{R}^n$ via the choice of a basis, then we can push the Lebesgue measure denoted by $\mathcal{L}^n$ from $\mathbb{R}^n$ to G. We shall see that this measure, which obviously is a Radon measure, is left-invariant and right-invariant. Hence, it is a bi-invariant Haar measure.

When we fix a Haar measure on G, we denote it by vol, or by vol_G if more than one group is considered. As discussed in Sect. 6.4, every pair of Haar measures differs by a multiplicative constant. Hence, the measures $\mathcal{H}^Q$, vol, and $\mathcal{L}^n$ are a multiple of each other.

Moreover, Carnot groups are unimodular in the sense that left-Haar measures are right-Haar measures, and vice versa. This latter fact holds for all nilpotent Lie groups; see Theorem 9.4.7.

Recall that if G is a Carnot group and $V_1, \dots, V_s$ is the stratification of its Lie algebra, in (11.4) we defined the homogeneous dimension of G as the integer number $Q := \sum_{j=1}^{s} j \cdot \dim V_j$.

Proposition 11.2.4 *Let G be a Carnot group of homogeneous dimension Q.*

11.2.4.i. If vol *denotes a Haar measure of G, then*

$$\mathrm{vol}(B(p,r)) = r^Q \,\mathrm{vol}(B(1_G,1)), \qquad \forall p \in G, \forall r > 0.$$

11.2.4.ii. Every Haar measure of G is Ahlfors Q-regular, the Hausdorff dimension of G is Q, and the Hausdorff Q-measure is a Haar measure.

11.2.4.iii. In exponential coordinates, the Lebesgue measure is the Hausdorff Q-measure up to a multiplication by a constant.

Proof By Theorem 9.4.7 (see also Proposition 9.4.17), in exponential coordinates, the Lebesgue measure $\mathcal{L}^n$ is both left and right-invariant. Moreover, every other Haar measure is a multiple of it; see [Fol99, Theorem 11.9]. In exponential coordinates, the anisotropic dilations δ_λ have Jacobian λ^Q, i.e., $\mathcal{L}^n(\delta_\lambda(E)) = \lambda^Q \cdot \mathcal{L}^n(E)$, for every measurable set $E \subseteq \mathbb{R}^n$. Hence, for all $p \in G$ and $\lambda \geq 0$, we have

$$\mathcal{L}^n(B(p,\lambda)) = \mathcal{L}^n(B(1_G,\lambda)) = \mathcal{L}^n(\delta_\lambda(B(1_G,1))) = \lambda^Q \mathcal{L}^n(B(1_G,1))$$

By Theorem 3.1.18, or Corollary 3.1.21, the metric measure space (G, vol) is Ahlfors regular of dimension Q. In particular, the Hausdorff Q-measure $\mathcal{H}^Q$ is Radon, and hence, it is a Haar measure. The last part of the proposition follows since both $\mathcal{L}^n$ and $\mathcal{H}^Q$ are Haar measures. □

11.3 Pansu-Rademacher Theorem

We would like to observe that the classical Rademacher Theorem states not only the almost-everywhere existence of a tangent map (called the differential) but also its realizability as a linear map, meaning as a group homomorphism that is compatible with the respective groups of dilations. Expressed in these terms, the theorem holds for general equiregular sub-Finsler manifolds as well; see [MM95]. This section aims to explain the content of such a differentiability result and to give a complete proof of it in the case of Carnot groups.

11.3.1 Pansu's Differentiability Theorem

We shall prove Pansu's version of the Rademacher Theorem.

Definition 11.3.1 (Pansu Differentiability) Let G and H be Carnot groups. We denote by δ_h the dilations of factor h in both of the groups. If $f : G \to H$ is a map, then its *Pansu differential* at a point $x \in G$ is, if it exists, 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 uniform convergence on compact sets. Moreover, we say that f is *Pansu differentiable* at x if Df_x exists and is a homogeneous group homomorphism, in the sense of Exercise 11.8.3.

For a map $f : G \to H$ between Carnot groups and $x, v \in G$, the value $Df(x; v) := \lim_{h\to 0^+} \delta_{1/h}(f(x)^{-1} f(x\delta_h(v)))$ may be called *partial Pansu derivative* of f at x along v. Notice that if $Df(x; v)$ exists, then $Df(x; \delta_\lambda v)$ exists for all $\lambda > 0$ and $Df(x; \delta_\lambda v) = \delta_\lambda Df(x; v)$. The value $Df(x; v)$ may exist for all $v \in G$, but the limit may not be uniform on compact sets. Moreover, even if the map Df_x exists, it may not be a group homomorphism from G to H.

Theorem 11.3.2 (Pansu's generalization of Rademacher Theorem) *Let $f : G \to H$ be a Lipschitz map between sub-Finsler Carnot groups. Then, for almost every $x \in G$, the map f is Pansu differentiable at x.*

11.3.1.1 Preliminaries to the Proof of Pansu's Theorem

In the proof of Theorem 11.3.2, we will only take for granted a few classical results to which we give hints for the proofs and references in the exercise section.

Theorem 11.3.3 (Rademacher Theorem in 1D; see [Fol99, Section 3.5]) *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 11.3.4 (Egorov Theorem for Metric Spaces; see Exercise 11.8.27) *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 $t \in \mathbb{R}_{>0}$. 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 11.3.5 (Consequence of Lebesgue Differentiation Theorem for Doubling Metric Spaces; see Exercise 11.8.28) *If (X, d, μ) is a doubling metric measure space and K is a measurable set in X, then μ-almost every point of K has density 1, that is,*

$$\lim_{r\to 0^+} \frac{\mu(K\cap B(x,r))}{\mu(B(x,r))} = 1, \qquad \text{for a.e. } x \in K.$$

11.3.1.2 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 11.3.6 (Case of Curves) *Let G be a Carnot group and $\gamma : [0,1] \to G$ a Lipschitz curve. Then γ is Pansu differentiable almost everywhere and for almost every $x \in [0,1]$ we have that for all $v \in \mathbb{R}$*

$$D\gamma(x; v) \stackrel{\text{def}}{=} \lim_{t\to 0} \delta_{1/t}\left(\gamma(x)^{-1}\gamma(x+tv)\right) = \exp\left(v(L_{\gamma(x)})^{*}\dot\gamma(x)\right).$$

Here are a 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 11.3.3. We also stress that Pansu's differentiability for curves is stronger than Euclidean differentiability. Namely, if we consider the curve in a rank-r Carnot group in exponential coordinates $\gamma(t) = (\gamma_1(t), \ldots, \gamma_n(t))$ with $\gamma(0) = \mathbf{0}$ and 0 is a point of Euclidean differentiability for γ, then $\dot\gamma(0) = \lim_{t\to 0}\gamma(t)/t = \lim_{t\to 0}(\gamma_1(t)/t, \ldots, \gamma_n(t)/t) = (h_1, \ldots, h_r, 0, \ldots, 0)$. However, we have to consider

$$\delta_{1/t}\gamma(t) = (\gamma_1(t)/t, \ldots, \gamma_n(t)/t^s)$$

and we need to prove that every 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 11.3.6 For simplicity, we take $v = 1$. We take a basis $X_1, \ldots, X_m$ of the first layer of the stratification of $\mathrm{Lie}(G)$. Let $h_1, \ldots, h_m \in L^\infty([0,1]; \mathbb{R})$ be such that

$$\dot\gamma(t) = \sum_{j=1}^{m} h_j(t) X_j(\gamma(t)), \qquad \text{for almost all } t \in [0,1]. \tag{11.23}$$

Assuming that γ is L-Lipschitz, we may take $|h_j(t)| \le L$, for all j and t. Let $x \in [0, 1]$ be both a point of Euclidean differentiability for γ and a Lebesgue point for each h_j, i.e.,

$$\frac{1}{|t-x|}\int_x^t |h_j(s) - h_j(t)|\,\mathrm{d}s \to 0, \qquad \text{as } t \to x, \quad \forall j \in \{1, \dots, m\}.$$

Up to replacing γ with the curve $t \mapsto \gamma(x)^{-1}\gamma(t+x)$, we may assume that $x = 0$ and $\gamma(x) = \mathbf{0}$.

We identify the group 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 $\sum_{j=1}^m h_j(0)X_j(0)$ since 0 is a Lebesgue point for each h_j.

Set $\eta_t(s) := \delta_{1/t}\gamma(t\,s)$, so each $\eta_t : [0, 1] \to G$ is a curve starting at 0 that is L-Lipschitz: for each $s, s' \in [0, 1]$ we have

$$d(\eta_t(s), \eta_t(s')) = d(\delta_{1/t}\gamma(t\,s), \delta_{1/t}\gamma(t\,s')) \le \frac{L}{t}|ts - ts'| = L|s - s'|.$$

Consequently, by Ascoli-Arzelà theorem, every sequence $(\eta_{t_k})_k$ has a uniformly converging subsequence. Moreover, we claim that we have the equality

$$\dot\eta_t(s) = \sum_{j=1}^m h_j(ts)X_j(\eta_t(s)). \tag{11.24}$$

Indeed,

$$\begin{aligned}
\dot\eta_t(s) &= \frac{\mathrm{d}}{\mathrm{d}s}\delta_{1/t}\gamma(ts) = (\mathrm{d}\delta_{1/t})_{\gamma(ts)}(t\dot\gamma(ts)) \\
&\overset{(11.23)}{=} t\sum_{j=1}^m h_j(ts)(\mathrm{d}\delta_{1/t})_{\gamma(ts)}(X_j(\gamma(ts)) \\
&\overset{\text{Ex. } 11.8.4}{=} t\sum_{j=1}^m h_j(ts)(\frac{1}{t}X_j(\delta_{1/t}\gamma(ts))) \overset{(11.24)}{=} \sum_{j=1}^m h_j(ts)X_j(\eta_t(s)),
\end{aligned}$$

which gives (11.24) from (11.23).

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, $\eta_t(1) = \delta_{1/t}\gamma(t) \to \dot\gamma(0)$. To prove the claim, we shall show that for each sequence $t_k \to 0$, there exists a subsequence t_{k_i} such that $\eta_{t_{k_i}} \to \eta_0$. Indeed, by Ascoli–Arzelà Theorem 3.1.3, there exists a subsequence t_{k_i} and there exists $\xi : [0, 1] \to G$ such that $\eta_{t_{k_i}} \to \xi$ uniformly. We want to show

$$\dot{\xi}(s) = \sum_{j=1}^{m} h_j(0) X_j(\xi(s)), \qquad \text{for almost every } s \in [0, 1].$$

Let σ be the curve such that $\sigma(0) = 0$ and $\dot{\sigma}(s) = \sum_{j=1}^{m} h_j(0) X_j(\xi(s))$. We integrate from 0 to an arbitrary $v \in (0, 1)$:

$$\begin{aligned}
\left|\sigma(v) - \eta_{t_{k_i}}(v)\right| &= \left|\int_0^v \dot{\sigma}(s)\,\mathrm{d}s - \int_0^v \dot{\eta}_{t_{k_i}}(s)\,\mathrm{d}s\right| \\
&= \left|\int_0^v \sum_{j=1}^{m} h_j(0) X_j(\xi(s))\,\mathrm{d}s - \int_0^v \sum_{j=1}^{m} h_j(t_{k_i}s) X_j(\eta_{t_{k_i}}(s))\,\mathrm{d}s\right| \\
&\leq \int_0^v \sum_{j=1}^{m} |h_j(0) - h_j(t_{k_i}s)| \left|X_j(\xi(s))\right| \mathrm{d}s + \\
&\qquad + \int_0^v \sum_{j=1}^{m} |h_j(t_{k_i}s)| \left|X_j(\xi(s)) - X_j(\eta_{t_{k_i}}(s))\right| \mathrm{d}s,
\end{aligned}$$

where we used (11.24). As $i \to \infty$, by continuity of X_j and boundedness of h_j, we have that the last summand goes to 0. Regarding the one before the last, we observe that

$$\begin{aligned}
\int_0^v \sum_{j=1}^{m} |h_j(0) - h_j(t_{k_i}s)|\,\mathrm{d}s &\leq \int_0^1 \sum_{j=1}^{m} |h_j(0) - h_j(t_{k_i}s)|\,\mathrm{d}s \\
&= \frac{1}{t} \int_0^t |h_j(0) - h_j(u)|\,\mathrm{d}u \to 0,
\end{aligned}$$

since 0 is a Lebesgue point. □

Proof of Theorem 11.3.2 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 a basis $X_1, \ldots X_m$ of V_1. For the entire proof, we take $j \in \{1, \ldots, m\}$ 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 11.3.6, for every $p \in G$ the maps $F \circ L_p|_{R_j}$ are almost everywhere differentable on R_j. By Fubini's Theorem (see Exercise 11.8.26), 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 be a Lipschitz constant for F. We shall consider the space $\mathrm{Lip}^L(R_j; H)$ of L-Lipschitz functions from R_j to H, with a separable distance that metrizes the uniform convergence on compact sets; see Exercise 3.4.13.

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 11.3.4 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

$$\begin{matrix} \{p_\epsilon\}_{\epsilon>0} \subset E_{\tau,r} \\ \lim_{\epsilon\to 0} p_\epsilon = p \in E_{\tau,r} \end{matrix} \quad \Rightarrow \quad \begin{matrix} \tilde{F}^j_{p_\epsilon,\epsilon} \to \tilde{F}^j_{p,0}, \text{ as } \epsilon \to 0, \\ \text{uniformly on compact sets of } R_j. \end{matrix} \tag{11.25}$$

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 claim that for all $v \in R_j$, and $\{p_\epsilon, q_\epsilon\}_{\epsilon>0} \subseteq E^\circ_{\tau,r}$

$$\begin{matrix} \lim_{\epsilon\to 0} \delta_{1/\epsilon}(p_\epsilon^{-1} q_\epsilon) = v \\ \lim_{\epsilon\to 0} p_\epsilon = p \end{matrix} \quad \Rightarrow \quad \lim_{\epsilon\to 0} \delta_{1/\epsilon}(F(p_\epsilon)^{-1} F(q_\epsilon)) = \tilde{F}^j_{p,0}(v). \tag{11.26}$$

Indeed, (11.26) is a consequence of $p_\epsilon \to p$ in $E_{\tau,r}$:

$$\begin{aligned} d(\delta_{1/\epsilon}(F(p_\epsilon)^{-1} F(q_\epsilon)), \tilde{F}^j_{p,0}(v)) &= d(F_{p_\epsilon,\epsilon}(\delta_{1/\epsilon}(p_\epsilon^{-1} q_\epsilon)), \tilde{F}^j_{p,0}(v)) \\ &\leq d(F_{p_\epsilon,\epsilon}(\delta_{1/\epsilon}(p_\epsilon^{-1} q_\epsilon)), F_{p_\epsilon,\epsilon}(v)) + d(F_{p_\epsilon,\epsilon}(v), \tilde{F}^j_{p,0}(v)) \\ &\leq L d(\delta_{1/\epsilon}(p_\epsilon^{-1} q_\epsilon), v) + d(\tilde{F}^j_{p_\epsilon,\epsilon}(v), \tilde{F}^j_{p,0}(v)) \to 0, \end{aligned}$$

where at the end we used the first assumption of (11.26) and (11.25).

Define

$$\mathscr{D}_p := \left\{ v \in G : \forall \{q_\epsilon\}_{\epsilon>0} \subseteq E^\circ_{\tau,r} \begin{array}{l} \text{if } \lim_{\epsilon\to 0} \delta_{1/\epsilon}(p^{-1} q_\epsilon) = v \\ \text{then } \delta_{1/\epsilon}(F(p)^{-1} F(q_\epsilon)) \text{ converges} \end{array} \right\}.$$

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

$$F_{p,0}(v) := \lim_{\epsilon\to 0} \delta_{1/\epsilon} \left(F(p)^{-1} F(q_\epsilon) \right).$$

We claim that

$$g \in \mathscr{D}_p, \; v \in R_j \Rightarrow gv \in \mathscr{D}_p, \tag{11.27}$$

with

$$F_{p,0}(gv) = F_{p,0}(g)\tilde{F}^j_{p,0}(v). \tag{11.28}$$

Indeed, let $\{q_\epsilon\}_{\epsilon>0} \subset E^\circ_{\tau,r}$ be such that $\lim_{\epsilon\to 0} \delta_{1/\epsilon}(p^{-1}q_\epsilon) = gv$; notice that since $p \in E^\circ_{\tau,r}$, then such a family q_ϵ exists. 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, by (11.26)

$$\begin{aligned}\lim_{\epsilon\to 0} \delta_{1/\epsilon}(F(p)^{-1}F(q_\epsilon)) &= \lim_{\epsilon\to 0} \delta_{1/\epsilon}(F(p)^{-1}F(p_\epsilon))\delta_{1/\epsilon}(F(p_\epsilon)^{-1}F(q_\epsilon)) \\ &\overset{(11.26)}{=} F_{p,0}(g)\tilde{F}^j_{p,0}(v).\end{aligned}$$

Next we observe the easy fact $1_G \in \mathscr{D}_p$, and therefore from (11.27) we infer

$$R_1, \ldots, R_m \subset \mathscr{D}_p. \tag{11.29}$$

From (11.29) and (11.27), together with the assumption that $R_1 \cup \ldots \cup R_m$ finitely generates G, we get that $\mathscr{D}_p = G$. From (11.28) the map $F_{p,0} : G \to H$ is a group homomorphism. 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

$$F_{p,0}(v) = \lim_{m\to\infty} F_{p,\epsilon_m}(v). \tag{11.30}$$

From (11.30), we conclude that every blowup of F at p, which exists by Ascoli–Arzelá theorem, coincides with the homomorphism $F_{p,0} : G \to H$. We have proved that F is Pansu differentiable at p. Since $\bigcup_{\tau,r>0} E^\circ_{\tau,r}$ has full measure in G, the map F is differentiable almost everywhere on G. □

11.3.1.3 Original Proof of Pansu's Theorem

In this subsection, we present the original proof by Pansu together with some extra explanation from Monti's thesis [Mon01]. For the proof, we introduce the *intrinsic difference quotients*:

$$R(x; v, t) := \delta_{1/t}\left(f(x)^{-1}f(x\delta_t v)\right),$$

so that $Df(x; v) := \lim_{t\to 0^+} R(x; v, t)$, when the limit exists.

We start with a preliminary result. It states that if we have partial derivatives in two directions, then we also have the partial derivative in the direction of the product of the two directions. Both the assumption and the conclusion are valid almost everywhere.

Proposition 11.3.7 *Let $f : G \to H$ be a Lipschitz map between sub-Finsler Carnot groups. Let $v, w \in G$. If $Df(x; v)$ and $Df(x; w)$ exist for almost every $x \in G$, then for almost every $x \in G$ we have that $Df(x; vw)$ exists and $Df(x; vw) = Df(x; v)Df(x; w)$.*

Proof Let $\Omega \subset G$ be open and with finite measure. Let $\eta > 0$. By Egorov's theorem for metric spaces (see Theorem 11.3.4), there exists a measurable subset $K = K(\eta) \subset \Omega$ 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 in x 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

$$R(x\delta_t v; w, t) \overset{t\to 0}{\longrightarrow} Df(x; w), \qquad \text{for almost every } x \in K. \tag{11.31}$$

Indeed, in this case, for $x \in K$, we have

$$\begin{aligned} R(x; vw, t) &= \delta_{1/t}\left(f(x)^{-1} f(x\delta_t(vw))\right) \\ &= \delta_{1/t}\left(f(x)^{-1} f(x\delta_t v)\right)\delta_{1/t}\left(f(x\delta_t v)^{-1} f(x\delta_t v\delta_t w)\right) \\ &= R(x; v, t)R(x\delta_t v; w, t) \overset{t\to 0}{\longrightarrow} Df(x; v)Df(x; w), \end{aligned}$$

where in the limit we use (11.31). Then, one concludes by taking the union of the sets $K = K(\eta)$ when $1/\eta$ varies in $\mathbb{N}$, which forms a full measure set.

For showing (11.31), take as x a point of density for K (recall that from Theorem 11.3.5, 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,

$$\frac{r_t^Q}{(2td(v,0))^Q} = \frac{|B_d(x\delta_t v, r_t)|}{|B_d(x, 2d(x, x\delta_t v))|} \leq \frac{|B_d(x, 2d(x, x\delta_t v)) \setminus K|}{|B_d(x, 2d(x, x\delta_t v))|} \to 0,$$

because x is a density point for K.

We now calculate the following quantity, defining the three points A_t, B_t, C_t in H:

$$\begin{aligned} &R(x\delta_t v; w, t) = \delta_{1/t}\left(f(x\delta_t v)^{-1} f(x\delta_t v\delta_t w)\right) \\ &= \underbrace{\delta_{\frac{1}{t}}\left(f(x\delta_t(v))^{-1} f(x_t)\right)}_{A_t} \underbrace{\delta_{\frac{1}{t}}\left(f(x_t)^{-1} f(x_t\delta_t w)\right)}_{B_t} \underbrace{\delta_{\frac{1}{t}}\left(f(x_t\delta_t w)^{-1} f(x\delta_t v\delta_t w)\right)}_{C_t}. \end{aligned}$$

First, we claim that the point A_t tends to the identity element 1_H, as $t \to 0$. Indeed, denoting by L a Lipschitz constant for f, we have

$$d(1_H, A_t) = \frac{1}{t} d(f(x_t), f(x\delta_t v)) \le \frac{L}{t} d(x_t, x\delta_t v) = Lr_t/t \to 0.$$

Second, we note that, since $x_t \in K$, $x_t \to x$, and on K the convergence of the intrinsic difference quotients is uniform, we have that $B_t = R(x_t; w, t) \to Df(x; w)$, as $t \to 0$. Third, we claim that $C_t \to 1_H$, as $t \to 0$. Indeed,

$$\begin{aligned} d(1_H, C_t) &= \frac{1}{t} d(f(x_t\delta_t w), f(x\delta_t v\delta_t w)) \\ &\le \frac{L}{t} d(x_t\delta_t w, x\delta_t v\delta_t w) \\ &= Ld(\delta_{1/t}(x_t)w, \delta_{1/t}(x\delta_t v)w) \to 0, \end{aligned}$$

where we used that $d(\delta_{1/t}(x_t), \delta_{1/t}(x\delta_t v)) = \dfrac{d(x_t, x\delta_t v)}{t} \to 0$, and that right translations are continuous. □

Another Proof of Theorem 11.3.2 Let $X_1, \dots, X_m$ be a basis of the first layer of the stratification of $\mathrm{Lie}(G)$.

We first claim that the set

$$E := \big\{p \in G : \\ Df(p; \exp(X_i)) \text{ and } Df(p; \exp(-X_i)) \text{ exist for all } i \in \{1, \dots, m\}\big\}$$

has full measure. Indeed, complete to a basis $X_1, \dots, X_n$ of $\mathrm{Lie}(G)$. For $j \in \{1, \dots, m\}$, define $\phi_j : \mathbb{R}^n \to G$ as $\phi_j(x_1, \dots, x_n) = \exp(\sum_{i \ne j} x_i X_i) \exp(x_j X_j)$. Then, the map ϕ_j is a diffeomorphism 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

$$\tilde{E}_j := \Big\{x \in \mathbb{R}^n : \\ t \mapsto f(\phi_j(x_1, \dots, x_{j-1}, t, x_{j+1}, \dots, x_n)) \text{ is Pansu differentiable at } t = x_j\Big\}.$$

By Fubini's Theorem for the Lebesgue measure and by Proposition 11.3.6, the set $\tilde{E}_j$ has full measure. Then, the set $E = \bigcap_{j=1}^m \phi_j(\tilde{E}_j)$ has full measure.

Let $S := \{v \in G : d(0, v) = 1\}$ be the unit sphere in G. For all $\ell \in \mathbb{N}$ there exist $v_1^\ell, \dots, v_{j_\ell}^\ell$ such that $S \subseteq \bigcup_{i=1}^{j_\ell} B_d(v_i^\ell, 1/\ell)$. We then claim that each set

$$E_\ell := \big\{p \in E : Df(p; v_i^\ell) \text{ exists for all } i \in \{1, \dots, j_\ell\}\big\}$$

has full measure. Indeed, since $\mathcal{G} := \left\{ \exp(\lambda X_i) : \lambda \in \mathbb{R}, i \in \{1, \dots, m\} \right\}$ generates G, then for all i and all ℓ there exists $w_1, \dots, w_k \in \mathcal{G}$ such that $v_i^\ell = w_1 \dots w_k$. Hence, from Proposition 11.3.7 for almost every $p \in G$ we have that $Df(p; v_i^\ell)$ exists. Thus, the set E_ℓ has full measure.

We finally claim that if $p \in \bigcap_{\ell \in \mathbb{N}} E_\ell$, then $R(p; v, t)$ converges uniformly in $v \in S$, as $t \to 0$. Indeed, we want to show that for all $\ell \in \mathbb{N}$ there exists $\delta > 0$ such that for all $s, t \in (0, \delta)$ and all $v \in S$

$$d(R(p; v, t), R(p; v, s)) \leq \frac{1 + 2L}{\ell},$$

where we are denoting by L a Lipschitz constant for f. Let $\ell \in \mathbb{N}$. Then there exists $\delta > 0$ such that for all $i \in \{1, \dots, i_\ell\}$ and all $s, t \in (0, \delta)$

$$d(R(p; v_i^\ell, t), R(p; v_i^\ell, s)) \leq \frac{1}{\ell}.$$

Let $v \in S$. Then there exists i such that $d(v, v_i^\ell) \leq \frac{1}{\ell}$. Then for all $s, t \in (0, \delta)$

$$\begin{aligned} d(R(p; v, t),& R(p; v, s)) \\ &\leq d(R(p; v, s), R(p; v_i^\ell, s)) \\ &\qquad + d(R(p; v_i^\ell, s), R(p; v_i^\ell, t)) \\ &\qquad\qquad + d(R(p; v_i^\ell, t), R(p; v, t)) \\ &\leq \frac{1}{s} d(f(p\delta_s v_i^\ell), f(p\delta_s v)) + \frac{1}{\ell} + \frac{1}{t} d(f(p\delta_t v_i^\ell), f(p\delta_t v)) \\ &\leq \frac{L}{s} s d(v_i^\ell, v) + \frac{1}{\ell} + \frac{L}{t} t d(v_i^\ell, v) \leq \frac{L + 1 + L}{\ell}. \end{aligned}$$

Since the intrinsic difference quotients converge uniformly on compact sets to a group homomorphism, we showed the Pansu differentiability. □

11.3.2 Applications to Non-Embeddability

It was observed by Semmes, [Sem96, Theorem 7.1], that Pansu's differentiation Theorem 11.3.2 implies that each Lipschitz embedding of the Heisenberg group with its CC distance into a Euclidean space cannot be bi-Lipschitz. The same holds for every non-commutative Carnot group.

Theorem 11.3.8 *There is no bi-Lipschitz embedding from an open nonempty set in a non-commutative Carnot group into any Euclidean space.*

Proof Let G be a non-commutative Carnot group. Suppose that such an embedding $f : U \subseteq G \to \mathbb{R}^n$ exists for some open set U and $n \in \mathbb{N}$. Pansu Rademacher Theorem 11.3.2 implies that there exists at least one point in U at which f is Pansu differentiable, and whose tangent map f_0 is a group homomorphism. The blowing-up procedure that is used to define the tangent map scales in a natural way: if f is L-bi-Lipschitz, then each rescaled f_λ is L-bi-Lipschitz. Therefore, also the tangent map f_0 is bi-Lipschitz. In particular, the map f_0 is injective. We now get a contradiction because f_0 is a group homomorphism between G and the abelian $\mathbb{R}^n$. However, every homomorphism from a Lie group into $\mathbb{R}^n$ must have a kernel that contains at least the commutator subgroup. Therefore, the subgroup $[G, G]$ is mapped to 0 via f_0, and hence f_0 cannot be injective. □

The same proof gives the following generalization.

Corollary 11.3.9 *Let G and H be Carnot groups. If no subgroup of H is isomorphic to G, then there is no bi-Lipschitz embedding of G in H.*

A challenging task is to characterize the Banach spaces into which Carnot groups can be bi-Lipschitz embedded. On the one hand, we have that Kuratowski embedding from [Kur35] (see also [Hei01, page 99]) implies that every separable metric space can be embedded isometrically into $\ell^\infty(\mathbb{N})$. But, on the other hand, we also have that every Carnot group that embeds bi-Lipschitz into $\ell^1(\mathbb{N})$ is commutative; see [EB+23].

11.4 A Metric Characterization of Carnot Groups

The purpose of this section is to give a more axiomatic presentation of Carnot groups from the viewpoint 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, in the sense discussed in Chap. 6. Such a result follows the spirit of Gromov's approach of 'seeing Carnot-Carathéodory spaces from within', [Gro96].

Theorem 11.4.1 *The sub-Finsler 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).*

We point out that each of the four conditions in Theorem 11.4.1 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 lines, like $\{(x, y) \in \mathbb{R}^2 : xy \geq 0\}$; every compact homogeneous space, such as the circle $\mathbb{S}^1$.

Theorem 11.4.1 provides a new equivalent definition of Carnot groups. Other papers focusing on metric characterizations of Carnot groups are [LD11b], [Bul11], [Fre13] (which is based on [LD11a]), and [BS14].

11.4.1 Proof of the Characterization

The proof of Theorem 11.4.1 is an easy consequence of three hard theorems: Montgomery-Zippin Theorem 6.2.10, Berestovskii Theorem 7.4.1, and Mitchell Theorem 12.1.8, which will be discussed in the next chapter.

Proof of Theorem 11.4.1 Let us verify that we can use Theorem 6.2.10. Obviously, each geodesic metric space is connected and locally connected. Regarding finite dimensionality, recall from Proposition 6.5.2 that every locally compact, self-similar, isometrically homogeneous space X is doubling and hence has finite topological dimension. Therefore, by Theorem 6.2.10, the isometry group G is a Lie group.

Since the distance is geodesic, Theorem 7.4.1 implies that our metric space is a sub-Finsler homogeneous manifold G/H. Since the sub-Finsler structure is G invariant, it is equiregular. Hence, on the one hand, because of Theorem 12.1.8, the tangents of our metric space are sub-Finsler 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 sub-Finsler Carnot group. □

11.5 Extremal Curves in Carnot Groups

In this section, we continue the discussion begun in Sect. 7.1.3.1. We review several problems concerning the regularity of geodesics in sub-Finsler Carnot groups, state some (optimistic) conjectures, and highlight a few partial results.

11.5.1 Expected Regularity of Sub-Riemannian Geodesics

11.5.1.1 Conjectures

Given two points in a sub-Finsler group, the existence of a geodesic between them is ensured by Ascoli–Arzelà Theorem; see Proposition 3.1.4. Being Lipschitz, we know that every such curve is differentiable almost everywhere.

It is expected that when the metric is sub-Riemannian, then every geodesic is C^1, or, in fact, C^∞. When instead the norm comes from a polytope, i.e., its unit ball is the convex hull of finitely many points, then we expect that 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.

Conjecture 11.5.1 (Regularity Conjecture for sub-Riemannian Manifolds) If $(M, \Delta, \langle\cdot,\cdot\rangle)$ is a sub-Riemannian manifold, then every geodesic for the CC-distance is C^2.

Conjecture 11.5.2 (Weak Regularity Conjecture for sub-Reimannian Carnot Groups) If G is a sub-Reimannian Carnot group, then each pair of points can be connected by a C^1 geodesic.

Conjecture 11.5.3 (Regularity Conjecture for sub-Finsler Carnot Groups) If $(G, V_1, \|\cdot\|_1)$ 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.

11.5.1.2 Various Partial Results

The following is a collection of some partial results from [LM08], [TY13, LD+16], [LPS24], and [Sus14], respectively:

Theorem 11.5.4 *Let G be a sub-Riemannian Carnot group and $\gamma : [0, 1] \to G$ be an energy minimizing curve.*

11.5.4.i. If G has rank 2 and step ≤ 4, then γ is C^∞.
11.5.4.ii. If G has step ≤ 3, then γ is C^∞.
11.5.4.iii. If $[G, G]$ is abelian, then γ is C^1.
11.5.4.iv. The curve γ is analytic on some open dense subset of its domain $[0, 1]$.

For sub-Riemannian manifolds, there are several other generic statements for which it is not clear if there is an analog for Carnot groups. For example, in [CJT06] the authors proved that for generic sub-Riemannian structures $(M, \Delta, \langle\cdot,\cdot\rangle)$ of rank at least 3, energy minimizers are C^∞. Here, generic means on an open dense set of structures, with respect to the Whitney C^∞ topology.

Regarding the case of polygonal norms, very little is known. The following result is a summary from [BL13, Bar+17, ALS19a, ALS19b]

Theorem 11.5.5 *Let G be either the Heisenberg group, the Engel group, or the Cartan group, with first stratum $V_1 \simeq \mathbb{R}^2$. Let $\|\cdot\|_1$ be the ℓ^1 norm on V_1. Consider the CC-distance of $(G, V_1, \|\cdot\|_1)$. Then, each pair of points in G can be joined with a geodesic that is made of, at most, a concatenation of* 14 *pieces of horizontal lines.*

11.5.2 Sublinear Isometric Property of Projections

In addition to the first-order analysis given by the PMP Theorem 7.3.3 (together with the second-order analysis of Goh's Theorem 7.3.4 and its generalizations [BMP20, BMS24]), there is another method to deduce some type of regularity for geodesics. The idea is to approximate geodesics in Carnot groups with lifts of geodesics in quotient groups. In brief, we have that for each geodesic, one of its quotients to a group of lower step is a geodesic up to a sublinear error. The following statement is the precise formulation; afterward, we will draw a list of consequences.

Theorem 11.5.6 ([HL23, Theorem 3.2]) *Let G be a Carnot group and $\gamma : [0,1] \to G$ be a geodesic. Let H be the smallest Carnot subgroup containing γ, and let s be the step of H. Then, there exists $C > 0$ such that*

$$|a-b| - C\,|a-b|^{\frac{s}{s-1}} \le d(\pi_s(\gamma(a)), \pi_s(\gamma(b))) \le |a-b|\,, \qquad \forall a, b \in [0,1], \tag{11.32}$$

where $\pi_s : H \to H/\exp(V_s(H))$ is the canonical submetry.

Proof Since π_s is a submetry, we have the second inequality of (11.32). To prove the first one, by the compactness of $[0,1]^2$, it is enough to show that for all $\bar{t} \in [0,1]$ there exists $C, \delta > 0$ such that (11.32) holds for all $a, b \in [\bar{t}-\delta, \bar{t}+\delta]$. Fix $\bar{t} \in [0,1]$. Write $V_j := V_j(H)$. Being H the minimal Carnot subgroup containing γ, there exist increasing $t_1, \dots, t_n \in [0,1] \setminus \{\bar{t}\}$ and $Y_1, \dots, Y_n \in V_1 \oplus \cdots \oplus V_{s-1}$ such that $\{\mathrm{Ad}_{\gamma(t_i)} Y_i\}_{i \in \{1,\dots,n\}}$ is a basis of $\mathfrak{h} := \mathrm{Lie}(H)$.

Define $\phi : \mathbb{R}^n \to \mathfrak{h}$ by setting

$$\phi(\mathbf{y}) := \log\left(\prod_{j=1}^{n} \mathrm{C}_{\gamma(t_j)}(\exp(y_j Y_j))\right), \qquad \forall \mathbf{y} \in \mathbb{R}^n. \tag{11.33}$$

Notice that, by Formula 5.5.6, we have

$$\mathrm{d}\phi_{\mathbf{0}}(e_j) = \frac{\mathrm{d}}{\mathrm{d}\epsilon} \log\left(\mathrm{C}_{\gamma(t_j)}(\exp(\epsilon Y_j))\right)\Big|_{\epsilon=0} = \mathrm{Ad}_{\gamma(t_j)} Y_j, \qquad \forall j \in \{1,\dots,n\}.$$

Therefore, the point $\mathbf{0} \in \mathbb{R}^n$ is a regular point for ϕ, since

$$\mathrm{span}\{\mathrm{Ad}_{\gamma(t_i)} Y_i\}_{i \in \{1,\dots,n\}} = \mathfrak{h}.$$

Hence, after fixing norms $\|\cdot\|$ on $\mathbb{R}^n$ and $\mathfrak{h}$, we apply the Inverse Mapping Theorem: there exist $\epsilon, C' > 0$ such that

$$Z \in \mathfrak{h}, \quad \|Z\| \leq \epsilon \implies \begin{array}{c} \exists\, \mathbf{y} \in \mathbb{R}^n \text{ such that} \\ \phi(\mathbf{y}) = Z \text{ and } \|\mathbf{y}\| \leq C'\|Z\|. \end{array} \tag{11.34}$$

Set $\delta := \min\left\{|t_1 - \bar{t}|, \dots, |t_n - \bar{t}|, \frac{\epsilon^{1/s}}{2C_{\mathrm{BB}}}\right\}$, where C_{BB} is the constant coming from the Ball-Box Theorem. Take $a, b \in (\bar{t} - \delta, \bar{t} + \delta)$. Consider a geodesic $\bar{\sigma} : [a, b] \to H/\exp(V_s)$ between $\pi_s(\gamma(a))$ and $\pi_s(\gamma(b))$ and let $\sigma : [a, b] \to H$ be the lift of $\bar{\sigma}$ from $\gamma(a)$. Set

$$Z := \log(\gamma(b)\sigma(b)^{-1}) \in V_s. \tag{11.35}$$

Notice, in particular, that Z is in the center of $\mathfrak{h}$. We have,

$$\|Z\|^{\frac{1}{s}} \leq C_{\mathrm{BB}} d(1, \exp(Z)) = C_{\mathrm{BB}} d(\sigma(b), \gamma(b)) \leq 2C_{\mathrm{BB}}|a - b|. \tag{11.36}$$

Since $|a - b| \leq \delta \leq \frac{\epsilon^{1/s}}{2C_{\mathrm{BB}}}$, by (11.36) we have $\|Z\| \leq \epsilon$. Thus by (11.34) there exists $\mathbf{y} \in \mathbb{R}^n$ with

$$\phi(\mathbf{y}) = Z, \tag{11.37}$$

$$\|\mathbf{y}\| \leq C'\|Z\|. \tag{11.38}$$

Let $k := \max\{j \in \{1, \dots, n\} \mid t_j < a\}$. By (11.37) we get

$$\begin{aligned} \gamma(1) &\overset{(11.37)}{=} \exp(\phi(\mathbf{y}))\exp(Z)^{-1}\gamma(1) \\ &\overset{(11.33),(11.35)}{=} \left(\prod_{j=1}^{n} \mathrm{C}_{\gamma(t_j)}(\exp(y_j Y_j))\right)\sigma(b)\gamma(b)^{-1}\gamma(1) \\ &= \left(\prod_{j=1}^{k} \mathrm{C}_{\gamma(t_j)}(\exp(y_j Y_j))\right)\gamma(a)\gamma(a)^{-1}\sigma(b)\gamma(b)^{-1} \\ &\qquad \left(\prod_{j=k}^{n} \mathrm{C}_{\gamma(t_j)}(\exp(y_j Y_j))\right)\gamma(1), \end{aligned}$$

where in the last equality we used that $\sigma(b)\gamma(b)^{-1}$ is in the center of H.

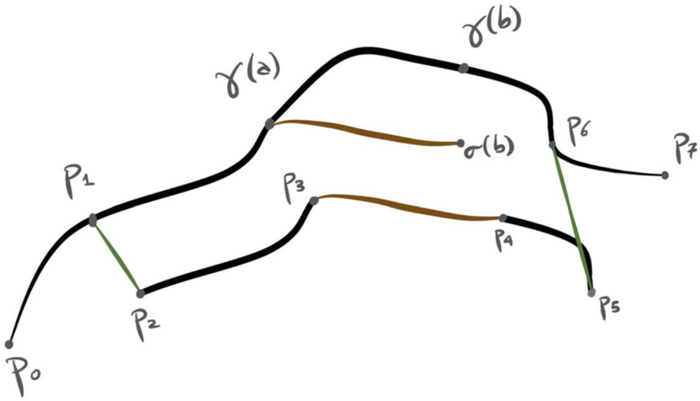


Fig. 11.1 A sketch of the points used in the triangle inequality in the proof of Theorem 11.5.6. For simplicity, we assume $\{t_1, \dots, t_n\} = \{t_1, t_2\}$, and $t_1 < a < b < t_2$. We consider the following points: $p_0 = \gamma(0)$; $p_1 = \gamma(t_1)$; $p_2 = p_1 \exp(y_1 Y_1)$; $p_3 = p_2 \gamma(t_1)^{-1}\gamma(a)$; $p_4 = p_3 \sigma(a)^{-1}\sigma(b)$; $p_5 = p_4 \gamma(b)^{-1}\gamma(t_2)$; $p_6 = p_5 \exp(y_2 Y_2) = \gamma(t_1)\exp(y_1 Y_1)\gamma(t_1)^{-1}\gamma(a)\sigma(a)^{-1}\sigma(b)\gamma(b)^{-1}\gamma(t_2)\exp(y_2 Y_2) = C_{\gamma(t_1)}(\exp(y_1 Y_1))\sigma(b)\gamma(b)^{-1}C_{\gamma(t_2)}(\exp(y_2 Y_2))\gamma(t_2) = \gamma(t_2)$; $p_7 = \gamma(1)$

Evaluating the distance from $\gamma(0)$, applying a triangular inequality (check Fig. 11.1) and using the fact that γ is a geodesic, we get:

$$
\begin{aligned}
1 &= d(\gamma(0), \gamma(1)) \\
&= d\Bigg(1, \gamma(0)^{-1}\Bigg(\prod_{j=1}^{k} \mathrm{C}_{\gamma(t_j)}(\exp(y_j Y_j))\Bigg)\gamma(a)\gamma(a)^{-1}\sigma(b)\gamma(b)^{-1} \\
&\qquad \Bigg(\prod_{j=k}^{n} \mathrm{C}_{\gamma(t_j)}(\exp(y_j Y_j))\Bigg)\gamma(1)\Bigg) \\
&\le \sum_{j=1}^{n} d(1, \exp(y_j Y_j)) + d(\gamma(0), \gamma(t_1)) \\
&\qquad + \sum_{j=1}^{k-1} d(\gamma(t_j), \gamma(t_{j+1})) + d(\gamma(t_k), \gamma(a)) \\
&\qquad + d(\gamma(a), \sigma(b)) + d(\gamma(b), \gamma(t_{k+1})) \\
&\qquad + \sum_{j=k+1}^{n-1} d(\gamma(t_j), \gamma(t_{j+1})) + d(\gamma(t_n), \gamma(1))
\end{aligned}
$$

$$
\begin{aligned}
&\overset{\text{geodesic}}{=} \sum_{j=1}^{n} d(1, \exp(y_j Y_j) + 1 - |b-a| + d(\sigma(a), \sigma(b)) \\
&= \sum_{j=1}^{n} d(1, \exp(y_j Y_j) + 1 - |b-a| + d(\pi_s(\gamma(a)), \pi_s(\gamma(b))) \\
&\overset{\text{BB}}{\leq} nC_{\text{BB}} \|\mathbf{y}\|^{\frac{1}{s-1}} + 1 - |b-a| + d(\pi_s(\gamma(a)), \pi_s(\gamma(b))) \\
&\overset{(11.38)}{\leq} C'' \|Z\|^{\frac{1}{s-1}} + 1 - |b-a| + d(\pi_s(\gamma(a)), \pi_s(\gamma(b))) \\
&\overset{(11.36)}{\leq} C|b-a|^{\frac{s}{s-1}} + 1 - |b-a| + d(\pi_s(\gamma(a)), \pi_s(\gamma(b))).
\end{aligned}
$$

□

Here are some important consequences of Theorem 11.5.6:

- The projection modulo the last layer $\exp(V_s)$ of every blowup of a geodesic in an s-step Carnot group is a geodesic.
- In every sub-Riemannian manifold of step s, blowing up a geodesic s times gives lines.
- At every point, among the blow-ups of a geodesic, there is a line.
- Corners cannot be geodesics.
- Every piecewise C^1 geodesic is C^1.

For more on the subject, we refer to [HL16, HL23, MPV18a, MPV18b].

11.6 Abnormal Curves in Carnot Groups

Non-constant abnormal curves are present in 2-step Carnot groups. For example, there is one in $\mathbb{R}$ times the Heisenberg group; recall Exercise 7.5.24. There are none exactly when the polarization is *strongly bracket generating*, in the sense that for every $X \in V_1 \setminus \{0\}$ we have $[X, V_1] = V_2$; see [LN18]. However, the abnormal curves starting from the identity element are confined in a subvariety, which has Haar measure 0. This general result is proved in [LD+16]. However, next in this section, we discuss the proof in the case of free groups.

11.6.1 Expected Sard Property for the Endpoint Map

Recall that the abnormal curves (or, better, their final points) are precisely the critical values of the endpoint map. Hence, in analogy with the classical Sard Theorem [Mil97, page 16], it is expected that they form a zero-measure set.

The query is still open, and there are several conjectures for possible results:

Conjecture 11.6.1 (Strong Sard Property) Let G be a nilpotent simply connected Lie group and $V \subseteq \mathrm{Lie}(G)$ a polarization. Then, the set of points in G that are connected from 1_G by some abnormal curve is contained in a proper real algebraic subvariety of G.

It is not clear whether the validity of the above conjecture is actually feasible. The next version is more likely to be true.

Conjecture 11.6.2 (Weak Sard Property) Let (G, V) be a sub-Riemannian Carnot group. Then, the set of points in G that are connected from 1_G by some abnormal energy minimizing curve is contained in a zero-measure set of G.

11.6.2 Extremals in Two-Step Free-Nilpotent Lie Groups

In two-step Carnot groups, the Strong Sard Conjecture holds. In fact, the subvariety containing the abnormal curves is explicit; see [BNV22]. We present it in the case of free groups, as in Sect. 11.1.3.6, together with the fact that geodesics are normal curves without using the Goh result.

Theorem 11.6.3 *Let $\mathbb{F}_{n2}$ be the free Carnot group of rank n and step 2, equipped with the standard sub-Riemannian Carnot metric. Then, we have the following two properties:*

(11.6.3)*.i. Every geodesic in $\mathbb{F}_{n2}$ is normal.*
(11.6.3)*.ii. The Strong Sard Property on $\mathbb{F}_{n2}$ holds: Seeing $\mathbb{F}_{n2}$ as $\mathbb{R}^n \times \Lambda^2(\mathbb{R}^n)$, abnormal curves starting from* $(0, 0)$ *are contained in the set*

$$\left\{(\theta, \omega) \in \mathbb{R}^n \times \bigwedge^2(\mathbb{R}^n) : \omega^{\frac{n}{2}} = 0\right\} \qquad \text{if } n \text{ is even, or}$$

$$\left\{(\theta, \omega) \in \mathbb{R}^n \times \bigwedge^2(\mathbb{R}^n) : \theta \wedge \omega^{\frac{n-1}{2}} = 0\right\} \qquad \text{if } n \text{ is odd.}$$

Proof Let $\gamma : [0, T] \to \mathbb{F}_{n2}$ be a geodesic. We also assume $\gamma(0) = 1_{\mathbb{F}_{n2}}$. By PMP Theorem 7.3.3, then it is a normal or an abnormal curve. Assume it is abnormal. Then, by Proposition 7.3.8, there exists a subgroup H containing γ and in which γ is not abnormal. There is a subspace $V \subset \mathbb{R}^n$ such that $H = V \times \Lambda^2(V)$. Notice that H admits a complementary normal subgroup N in $\mathbb{F}_{n2}$.

Back to the curve γ, it is also a geodesic within H. Since it is not abnormal, it is normal in H, say with covector $\lambda \in \mathrm{Lie}(H)^*$. Hence, the covector λ can be extended to be 0 in the Lie algebra of N. With this extension, the curve γ is normal in $\mathbb{F}_{n2}$. Thus, we proved (11.6.3).i.

Regarding (11.6.3).ii, let $\gamma = (\theta, \omega)$ be an abnormal curve in $\mathbb{R}^n \times \Lambda^2(\mathbb{R}^n)$. By the discussion above, the curve θ is contained in a proper subspace V of $\mathbb{R}^n$, and ω is contained in $\bigwedge^2 V$. If $(\theta, \omega) \in V \times \bigwedge^2 V$, then the degree of $\omega^{\frac{n}{2}}$ or $\theta \wedge \omega^{\frac{n-1}{2}}$ is $n > \dim(V)$. This implies that $\omega^{\frac{n}{2}} = 0$ or $\theta \wedge \omega^{\frac{n-1}{2}} = 0$. □

11.7 Supplementary Material

11.7.1 Self-Similar Sub-Finsler Spaces

In sub-Finsler geometry, self-similar spaces are well characterized. As differentiable manifolds, they have a Lie homogeneous structure of a Lie coset space as the quotient of a Carnot group modulo the action of a dilation-invariant subgroup via left multiplication. The Lie group metric quotient is the one we saw in Proposition 6.3.4 and then in Proposition 7.1.9. Hence, the well-defined action on the quotient is on the right, and it may not be by isometries. In fact, these quotients may not be isometrically homogeneous spaces. They are still called homogeneous because they admit dilations. To avoid this double use of the word homogeneity, since Sect. 6.5, we adopted the term self-similar spaces.

Definition 11.7.1 A *self-similar sub-Finsler space* (or, better, a *self-similar constant-rank sub-Finsler space*) is a sub-Finsler manifold obtained as the quotient space of a (left-invariant) sub-Finsler Carnot group with respect to the left-action of a dilation-invariant subgroup, and it is equipped with the quotient distribution and metric, as in Proposition 7.1.9. Namely, assume G is a Carnot group with distribution Δ and left-invariant norm $\|\cdot\|$ and $H < G$ is a closed dilation-invariant subgroup, for which $T_1 H \cap \Delta_1 = \{0\}$. On the quotient manifold ${}_H\backslash^G := \{Hg : g \in G\}$ we consider the sub-Finsler structure that makes the projection $\pi : G \to {}_H\backslash^G$ a submetry: we take $\Delta_{{}_H\backslash^G} := \pi_* \Delta$ as (constant-rank) distribution and we define the continuously varying norm on ${}_H\backslash^G$ setting for all $p \in {}_H\backslash^G$ and for all $v \in (\pi_*\Delta)_p \subseteq T_p({}_H\backslash^G)$

$$\|v\|_{{}_H\backslash^G} := \inf\left\{\|w\|_q : q \in \pi^{-1}(p), w \in T_q\Delta, \mathrm{d}\pi_q(w) = v\right\}. \tag{11.39}$$

We stress that, if G is a Carnot group with dilations $(\delta_\lambda)_{\lambda\in\mathbb{R}_+}$ then H is dilation-invariant if $\delta_\lambda(H) = H$, for all $\lambda \in \mathbb{R}_+$, and in this case the map

$$\delta_\lambda(Hg) := H\delta_\lambda(g), \qquad \forall g \in G, \forall\lambda \in \mathbb{R}_+ \tag{11.40}$$

is well defined. Metrically, each map δ_λ is a dilation of factor λ, in the sense of Sect. 6.5. We, therefore, clarified that, metrically, every self-similar sub-Finsler space is self-similar.

Vice versa, we claim that every sub-Finsler space that admits a non-trivial dilation is of the above form. Indeed, for every dilation of a factor different than 1, we can assume the factor is in the open interval (0, 1), up to replacing the map with its inverse. Hence, being a contraction, the map has a fixed point. Consequently, the space (pointed at this point) is isometric to a dilation of it. Therefore, the space is isometric to one of its metric tangent spaces. By the work of Bellaïche ([Bel96]; see also [ALN23] for the sub-Finsler case), we know that the metric tangents of (constant-rank) sub-Finsler manifolds are (constant-rank) self-similar spaces.

11.7.2 Local Isometries of Carnot Groups

From what we discussed in Theorem 10.2.1 and more generally in Theorem 10.3.1, isometries of sub-Finsler Carnot groups are affine maps:

Corollary 11.7.2 *Let G and H be sub-Finsler Carnot groups with normed first layers $(V_1(G), \|\cdot\|)$ and $(V_1(H), \|\cdot\|)$, respectively. A map $F : G \to H$ is an isometry with respect to the sub-Finsler metrics if and only if there is $g \in G$ and a Lie group automorphism $\phi : G \to H$ such that $\mathrm{d}\phi|_{V_1}$ is an isometry between $(V_1(G), \|\cdot\|)$ and $(V_1(H), \|\cdot\|)$ and $F = \phi \circ L_g$.*

In sub-Finsler Carnot groups, we additionally have that local isometries are restrictions of global isometries.

Theorem 11.7.3 ([LO16, Theorem 1.1]) *Let G_1 and G_2 be sub-Finsler Carnot groups, and open sets $\Omega_1 \subset G_1$ and $\Omega_2 \subset G_2$. If $F : \Omega_1 \to \Omega_2$ is an isometry, then F is the restriction to Ω_1 of an isometry $G_1 \to G_2$.*

Note that in the statement above, we require the domain Ω_1 to be open. The hypothesis that the set Ω_1 is open cannot be dropped. Unlike in the Euclidean space, Theorem 11.7.3 cannot be generalized to arbitrary subsets. We present a counterexample: take the sub-Riemannian Heisenberg group $(\mathbb{H}, d_{SR})$ in standard exponential coordinates (x, y, z) with respect to the basis of its Lie algebra given by vectors X, Y and $[X, Y] =: Z$. Consider the set E given by the xy plane together with the z axis:

$$E := \exp(\mathbb{R}X \oplus \mathbb{R}Y) \cup \exp(\mathbb{R}Z).$$

Then, the map $(x, y, z) \mapsto (x, y, -z)$ is an isometry of E into itself. However, this map is not the restriction of an isometry, and, actually, nor of a bi-Lipschitz map. Indeed, notice that every homeomorphism that extends F reverses the topological orientation, while the isometries (and the bi-Lipschitz homeomorphisms) of the Heisenberg group preserve the topological orientation; see Exercise 11.8.37.

11.8 Exercises

Exercise 11.8.1 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 11.8.2 Let δ_λ be the dilation of factor λ as defined either at the group level in Definition 11.1.2 or at the Lie-algebra level in Definition 9.2.12. Then, we have that $(\delta_\lambda)^{-1} = \delta_{1/\lambda}$.

Exercise 11.8.3 (Carnot Morphism) Let $\varphi : G \to H$ be a Lie group homomorphism between Carnot groups. Then, we have $\varphi_*(V_1(G)) \subseteq V_1(H)$ if and only if $\varphi \circ \delta_\lambda = \delta_\lambda \circ \varphi$, for all $\lambda \in \mathbb{R}$. In this case, we say that φ is a *homogeneous group homomorphism* or that it is a *Carnot morphism*.

Exercise 11.8.4 Let G be a Carnot group. For all $X \in \mathrm{Lie}(G)$, $u \in C^\infty(G)$, $\lambda \geq 0$, and $g \in G$ we have $X(u \circ \delta_\lambda)(g) = (\delta_\lambda X)u(\delta_\lambda g)$.

Exercise 11.8.5 Let G be a Carnot group. For all $p \in G$, for all $r > 0$

$$B_{d_{cc}}(p, r) = L_p(\delta_r(B_{d_{cc}}(1_G, 1))).$$

Exercise 11.8.6 Consider a Carnot basis $X_1, \ldots, X_n$ of a Carnot algebra. 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 $X_{j,1}, \ldots, X_{j,d_j}$ are basis elements 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.
Hint. Iterate (11.10).

Exercise 11.8.7 In every Carnot group, there is a Malcev basis.

Exercise 11.8.8 Let V and W be vector subspaces of a Lie algebra $\mathfrak{g}$ with $X_1, \ldots, X_l$ and $Y_1, \ldots, Y_m$ basis of V and W, respectively. Then, the vectors $[X_i, Y_j]$, for $i \in \{1, \ldots, l\}$, $j \in \{1, \ldots, m\}$ span the subspace $[V, W]$, thus one can extract a basis among such brackets.

Exercise 11.8.9 Let $\mathfrak{g} = V_1 \oplus \cdots \oplus V_s$ be a stratification of a Lie algebra. Assume that $X_{m_{j-1}+1}, \ldots, X_{m_j}$ is a basis of V_j, for all $j \in \{1, \ldots, s\}$, then the order-reversed basis $(X_n, \ldots, X_1)$ is a Malcev basis.

Exercise 11.8.10 For the Engel group, with group product (11.15), the differential at $\mathbf{0}$ of the left translation by $\mathbf{x}$ is

$$\mathrm{d}(L_\mathbf{x})_\mathbf{0} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ -\frac{x_2}{2} & \frac{x_1}{2} & 1 & 0 \\ -\frac{x_1 x_2}{12} - \frac{x_3}{2} & \frac{x_1^2}{12} & \frac{x_1}{2} & 1 \end{bmatrix}.$$

Exercise 11.8.11 For the Engel group, consider the exponential coordinates of the second kind with respect to the ordered basis $(X_1, \dots, X_4)$ with relations as in (11.14). In these coordinates, the left-invariant vector fields are

$$\begin{cases} \tilde{X}_1 = \partial_{x_1} \\ \tilde{X}_2 = \partial_{x_2} + x_1\,\partial_{x_3} + \frac{x_1^2}{2}\partial_{x_4} \\ \tilde{X}_3 = \partial_{x_3} + x_1\partial_{x_4} \\ \tilde{X}_4 = \partial_{x_4}\,. \end{cases}$$

Exercise 11.8.12 For the Cartan group, with group product (11.16), the induced left-invariant vector fields are:

$$\begin{cases} X_1 = \partial_{x_1} - \frac{x_2}{2}\partial_{x_3} - \left(\frac{x_3}{2} + \frac{x_1x_2}{12}\right)\partial_{x_4} - \frac{x_2^2}{12}\partial_{x_5} \\ X_2 = \partial_{x_2} + \frac{x_1}{2}\partial_{x_3} + \frac{x_1^2}{12}\partial_{x_4} - \left(\frac{x_3}{2} - \frac{x_1x_2}{12}\right)\partial_{x_5} \\ X_3 = \partial_{x_3} + \frac{x_1}{2}\partial_{x_4} + \frac{x_2}{2}\partial_{x_5} \\ X_4 = \partial_{x_4} \\ X_5 = \partial_{x_5}\,. \end{cases}$$

While the induced right-invariant vector fields are given by the columns of the following matrix:

$$\mathrm{d}(R_{\mathbf{x}})_{\mathbf{0}} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ \frac{x_2}{2} & -\frac{x_1}{2} & 1 & 0 & 0 \\ \frac{x_3}{2} - \frac{x_1x_2}{12} & \frac{x_1^2}{12} & -\frac{x_1}{2} & 1 & 0 \\ -\frac{x_2^2}{12} & \frac{x_3}{2} + \frac{x_1x_2}{12} & -\frac{x_2}{2} & 0 & 1 \end{bmatrix}.$$

Exercise 11.8.13 For the Cartan group, with group product (11.16), the left-invariant 1-forms dual to the standard basis are:

$$\begin{cases} \theta_1 = \mathrm{d}x_1 \\ \theta_2 = \mathrm{d}x_2 \\ \theta_3 = \mathrm{d}x_3 - \frac{x_1}{2}\mathrm{d}x_2 + \frac{x_2}{2}\mathrm{d}x_1 \\ \theta_4 = \mathrm{d}x_4 - \frac{x_1}{2}\mathrm{d}x_3 + \frac{x_1^2}{6}\mathrm{d}x_2 + \left(\frac{x_3}{2} - \frac{x_1x_2}{6}\right)\mathrm{d}x_1 \\ \theta_5 = \mathrm{d}x_5 - \frac{x_2}{2}\mathrm{d}x_3 + \left(\frac{x_3}{2} + \frac{x_1x_2}{6}\right)\mathrm{d}x_2 - \frac{x_2^2}{6}\mathrm{d}x_1\,. \end{cases}$$

Exercise 11.8.14 For $s \in \mathbb{N}$, the following vector fields in $\mathbb{R}^{s+1}$ generate a Lie algebra isomorphic to the s-step filiform Lie algebra of the first kind

$$\begin{cases} X_1 = \partial_{x_1} \\ X_2 = \partial_{x_2} + \sum_{j=1}^{s-1} x_1^j\,\partial_{x_{j+2}}\,. \end{cases}$$

Exercise 11.8.15 The proof of Proposition 11.2.1 extends to the case where G is a positively graded group.

Exercise 11.8.16 For the $\mathtt{Box}$ as in (11.19) and δ_λ as in (11.20), we have the property $\delta_\lambda(\mathtt{Box}(r)) = \mathtt{Box}(\lambda r)$, for all $r, \lambda > 0$

Exercise 11.8.17 On each Carnot group, the notion of homogeneous dimension from (11.4) agrees with the one on sub-Finsler manifolds in Definition 4.3.5.

Exercise 11.8.18 If $X_1, \dots, X_n$ is a Carnot basis of a Carnot group, vol is a Haar measure on it, and $\mathcal{L}^n$ is the Lebesgue measure on $\mathbb{R}^n$, then there is $c \in \mathbb{R}$ such that $\mathrm{vol}\big(\{\exp(\sum_{i=1}^n x_i X_i) : (x_1, \dots, x_n) \in A\}\big) = c\mathcal{L}^n(A)$, for all Borel sets $A \subseteq \mathbb{R}^n$.

Exercise 11.8.19 For all Borel subsets A of each Carnot group G equipped with a Haar measure vol, we have $\mathrm{vol}(\delta_\lambda(A)) = \lambda^Q \mathrm{vol}(A)$.

Exercise 11.8.20 (Homogeneous Lines are Snowflakes) Let G be a Carnot group with V_j the j-th stratum of its stratification. Let $X \in V_j$. Then, restricting the metric of G on $\exp(\mathbb{R}X)$ gives a metric space that is isometric to the $1/j$-snowflake of the Euclidean line.

Exercise 11.8.21 Let G be a Carnot group. For $f : \mathbb{R} \to G$ of class C^1, we have that f is Pansu differentiable if and only if it is a horizontal curve.

Exercise 11.8.22 Let G be a Carnot group. Every $f : G \to \mathbb{R}$ of class C^1 is Pansu differentiable.

Exercise 11.8.23 Let G, H be Carnot groups. A C^1 map $f : G \to H$ is Pansu differentiable at every point if and only if its (standard) differential preserves the horizontal bundle.

Exercise 11.8.24 (Suggested by S. Nicolussi Golo) Let G, H be Lie groups with stratified Lie algebras $\mathfrak{g} = \bigoplus_{j=1}^s V_j$ and $\mathfrak{h} = \bigoplus_{k=1}^t W_k$, respectively. Let $f : G \to H$ be a C^1-smooth map with $f(1_G) = 1_H$. We write the (standard) differential $\mathrm{d}f_{1_G} : \mathfrak{g} \to \mathfrak{h}$ of f at 1_G as

$$\mathrm{d}f_{1_G} = \begin{pmatrix} A_{W_1}^{V_1} & A_{W_1}^{V_2} & \dots & A_{W_1}^{V_s} \\ A_{W_2}^{V_1} & A_{W_2}^{V_2} & \dots & A_{W_2}^{V_s} \\ \vdots & & \ddots & \vdots \\ A_{W_t}^{V_1} & A_{W_t}^{V_2} & \dots & A_{W_t}^{V_s} \end{pmatrix},$$

where, for every j and k, $A_{W_k}^{V_j}$ is a linear map $V_j \to W_k$. Assume that f is Pansu differentiable at 1_G. Then $A_{W_k}^{V_j} = 0$ for all $k > j$, that is,

$$\mathrm{d}f_{1_G} = \begin{pmatrix} A_{W_1}^{V_1} & A_{W_1}^{V_2} & \dots & A_{W_1}^{V_s} \\ 0 & A_{W_2}^{V_2} & \dots & A_{W_2}^{V_s} \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \dots & A_{W_t}^{V_s} \end{pmatrix} \tag{11.41}$$

and the Pansu differential of f at 1_G is the linear map $\mathfrak{g} \to \mathfrak{h}$ given by

$$\begin{pmatrix} A_{W_1}^{V_1} & 0 & \dots & 0 \\ 0 & A_{W_2}^{V_2} & \dots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \dots & A_{W_t}^{V_s} \end{pmatrix}. \tag{11.42}$$

Solution. We identify both Lie groups with their respective Lie algebras via their exponential maps. Since f is Pansu differentiable at $1_G = 0$, then, for every $v \in \mathfrak{g}$, the limit $\lim_{\epsilon \to 0^+} \delta_{1/\epsilon} f(\delta_\epsilon v)$ exists, and it is a vector in $\mathfrak{h}$. Since f is C^1, then there exists a continuous function $\rho : \mathfrak{g} \to \mathfrak{h}$ such that

$$f(v) = \mathrm{d}f_{1_G} v + \rho(v), \qquad \text{with } \lim_{v \to 0} \frac{|\rho(v)|}{|v|} = 0,$$

where we denote by $|\cdot|$ some norms in both vector spaces $\mathfrak{g}$ and $\mathfrak{h}$. Write $\rho(v) = \sum_{k=1}^{t} \rho_k(v)$ for suitable $\rho_k : \mathfrak{g} \to W_k$. For every $j \in \{1, \dots, s\}$ and every (nonzero) $v_j \in V_j$, we have

$$\lim_{\epsilon \to 0^+} \frac{|\rho_k(\delta_\epsilon v_j)|}{\epsilon^j} = \lim_{\epsilon \to 0^+} \frac{|\rho_k(\epsilon^j v_j)|}{|\epsilon^j v_j|} |v_j| = 0. \tag{11.43}$$

For $k \in \{1, \dots, t\}$ and $v = v_\ell \in V_\ell$ for $\ell \in \{1, \dots, s\}$, we compute $[\delta_{1/\epsilon} f(\delta_\epsilon v)]_k = \sum_{j=1}^{s} \epsilon^{-k} A_k^j(\epsilon^j v_j) + \epsilon^{-k} \rho_k(\delta_\epsilon v) = \epsilon^{\ell-k} A_k^\ell(v_\ell) + \epsilon^{-k} \rho_k(\epsilon^\ell v_\ell)$.

If $\ell < k$, then we get

$$[\delta_{1/\epsilon} f(\delta_\epsilon v_\ell)]_k = \frac{1}{\epsilon^{k-\ell}} \left(A_k^\ell(v_\ell) + \frac{\rho_k(\epsilon^\ell v_\ell)}{\epsilon^\ell} \right).$$

Because the limit for $\epsilon \to 0^+$ exists and because of (11.43), we conclude that $A_k^\ell(v_\ell) = 0$. If instead $\ell \geq k$, then

$$\lim_{\epsilon \to 0} [\delta_{1/\epsilon} f(\delta_\epsilon v_\ell)]_k = \begin{cases} A_k^k(v_k) & \text{if } \ell = k, \\ 0 & \text{if } \ell > k. \end{cases}$$

This shows that the Pansu derivative of f at 1_G has the form (11.42).

Exercise 11.8.25 Let $\mathcal{H}$ be the Heisenberg group equipped with its Carnot structure. There are examples of C^1 curves $f : \mathbb{R} \to \mathcal{H}$ that at 0 are horizontal, but they are not Pansu differentiable at 0.

Exercise 11.8.26 (Fubini Theorem in Carnot Groups) Let G be a Carnot group and $X \in \mathfrak{g}$ horizontal. Take any $W \subseteq V_1$ transverse to X. Then, the set $H := \exp(W \oplus [\mathfrak{g}, \mathfrak{g}])$ is a normal subgroup and G is the semidirect product $G = H \rtimes$

$\exp(\mathbb{R}X)$. Consider exponential coordinates of a mixed form with respect to H and X: Namely, $(h,t) \in \mathrm{Lie}(H) \times \mathbb{R} \mapsto \exp(h)\exp(tX) \in G$. In these corrdinates, the flowlines of X are of the form $t \in \mathbb{R} \mapsto p + (0,t) \in \mathbb{R}^n$, where $n := \dim \mathfrak{g}$. The Lebesgue measure in $\mathbb{R}^n$ is a Haar measure vol for G. The Lebesgue $(n-1)$-measure in $\mathrm{Lie}(H)$ is a Haar measure for H. If $f : G \to \mathbb{R}$ is a measurable function, then

$$\int_G f(g)\,\mathrm{d}\,\mathrm{vol}(g) = \int_{\mathrm{Lie}(H)\times\mathbb{R}} f(\phi(p))\,\mathrm{d}\mathcal{L}^n(p)$$

$$= \int_{\mathrm{Lie}(H)} \int_{\mathbb{R}} f(\phi(q,t))\,\mathrm{d}t\,\mathrm{d}\mathcal{L}^{n-1}(q).$$

Exercise 11.8.27 Here is a scketch to prove Egorov's Theorem 11.3.4: For $k \in \mathbb{N}$ and $t \in (0,\infty)$, let

$$E_t(k) := \bigcup_{s\in(t,\infty)} \{x : |f_s(x) - f(x)| > k^{-1}\}.$$

Then, for fixed k, the set $E_t(k)$ decreases as t decreases, and $\mu\left(\bigcap_{t\in(0,\infty)} E_t(k)\right) = 0$, 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 $E = \bigcap_{k\in\mathbb{N}} E_{t_k}(k)$. 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 11.8.28 (Lebesgue Differentiation Theorem for Doubling Metric Spaces) If (X, d, μ) is a doubling metric measure 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. See [Hei01, Theorem 1.8].

Exercise 11.8.29 Every sub-Riemannian group of nilpotency step 2 is at a bounded distance from its asymptotic cone. That is, the two metric spaces are $(1, C)$-quasi-isometric for some $C > 1$.

Exercise 11.8.30 The geodesic lines in sub-Riemannian Carnot groups of step 2 are precisely the left translations of the horizontal one-parameter subgroups.
Hint. Recall Corollary 7.3.11

Exercise 11.8.31 Isometric embeddings of Carnot groups into sub-Riemannian Carnot groups of step 2 are affine.
Hint. Recall Exercise 11.8.30, or Corollary 7.3.11 for the isometric embeddings of $\mathbb{R}$. The extension to higher-dimensional domains can be found in [Hak20].

Exercise 11.8.32 Let $\widetilde{\mathrm{End}} : \Omega \longrightarrow G \times \mathbb{R}$ be the extended endpoint map on a sub-Riemannian group G. A control $u \in \Omega$ is *strictly abnormal* (i.e., abnormal, but not normal) if and only if $\{0\} \times \mathbb{R} \subseteq \mathrm{d}\widetilde{\mathrm{End}}_u(\Omega) \neq T_{\mathrm{End}(u)}G \times \mathbb{R}$.

Exercise 11.8.33 Let G be a free Carnot group and $H < G$ a Carnot subgroup. Let $\gamma \subset H$ be a normal curve. Then, γ is normal as a curve in G.
Hint. Find a normal subgroup N such that $G = N \rtimes H$. Extend the normal covector on $\mathfrak{h}$ to be 0 on $\mathfrak{n}$.

Exercise 11.8.34 Let G be a Carnot group of step at most 3. Then, all energy-minimizing curves in G are smooth.
Hint. By Goh Theorem 7.3.4, every strictly abnormal curve is contained in a proper subgroup. Hence, every energy-minimizing curve is contained in a subgroup within which it is normal.

Exercise 11.8.35 Let $\pi : G \to H$ be a Carnot morphism between Carnot groups. Let $\gamma : [0, 1] \to G$ be a strictly abnormal geodesic. Assume that $\pi \circ \gamma$ is energy minimizing. Then $\pi \circ \gamma$ is strictly abnormal.
Hint. Use Exercise 11.8.32.

Exercise 11.8.36 ☠ Let $\pi : G \to H$ be a Carnot morphism between Carnot groups. Let $\gamma : [0, 1] \to G$ be a normal geodesic. Assume that $\pi \circ \gamma$ is energy minimizing. Is then $\pi \circ \gamma$ normal?

Exercise 11.8.37 Every affine map of the first Heisenberg group has everywhere positive Jacobian determinant.

Chapter 12
Limits of CC Spaces

Sub-Riemannian and sub-Finsler Carnot groups emerge as limit metric spaces, both as distinguished asymptotic spaces and as tangent metric spaces. In this chapter, we begin by reviewing the notion of limits of metric spaces and consequently introduce the concept of asymptotic cones and metric tangent spaces; see Sect. 12.1. In Sect. 12.2, to study varying CC structure, we discuss some terminology and some preliminary results, which mostly are consequences of Grönwall Lemma. Section 12.3 is a collection of examples, which help the understanding of the more general statements and proofs contained in the subsequent chapters. In Sect. 12.4, we explain and prove Pansu's Theorem on the asymptotic geometry of nilpotent Lie groups. In Sect. 12.5, we discuss Mitchell's Theorem: we give a complete proof in the case of sub-Finsler Lie groups, showing that the tangent spaces are Carnot groups. We only mention the general result for sub-Finsler manifolds in Sect. 12.6, explaining what is the general strategy of proof for varying CC structures; see Sect. 12.7. We conclude the chapter with Sect. 12.8, where we discuss finitely-generated groups of polynomial growth and we mention the general strategy to prove a celebrated result of Gromov proving that these groups are virtually nilpotent and hence their asymptotic cones are sub-Finlser Carnot groups.

12.1 Limits of Metric Spaces

In most cases, the study of limits of CC metrics can be reduced, after some change of coordinates, to sequences of distances on some manifold that uniformly converge on compact sets. However, it can also be valuable to regard such convergence as a specific instance of Gromov-Hausdorff convergence.

E. Le Donne, *Metric Lie Groups*, Graduate Texts in Mathematics 306,
https://doi.org/10.1007/978-3-031-98832-5_12

12.1.1 A Topology on the Space of Metric Spaces

Recall the definition of quasi-isometric embedding from Definition 3.1.13. In the following, by the term *pointed metric space*, we mean a pair of a metric space together with a point in it. Also, a map $\phi : (X, x) \to (Y, y)$ between pointed metric spaces is a map $\phi : X \to Y$ such that $\phi(x) = y$.

Definition 12.1.1 (Hausdorff Approximating Sequence) Let (X_j, x_j) and (Y_j, y_j), for $j \in \mathbb{N}$, 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 a sequence $\epsilon_j > 0$ such that

1. $\epsilon_j \to 0$ as $j \to \infty$;
2. $\phi_j|_{B(x_j,R)}$ 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 12.1.2 We say that a sequence of pointed metric spaces (X_j, x_j) *converges* to a pointed metric space (Y, y), and that (Y, y) is the *limit* of the sequence if there exists a Hausdorff approximating sequence $\phi_j : (X_j, x_j) \to (Y, y)$.

Nowadays, this notion of convergence is called *Gromov-Hausdorff convergence*. It was originally introduced by Edwards in 1975, [Edw75], and then later independently by Gromov in 1981, [Gro81]. This is a generalization of the Hausdorff convergence from [Hau14]. The following criterion will be used to show the convergence of particular sequences of CC spaces.

Proposition 12.1.3 *Let d_j be a sequence of distance functions on a set X that converges to a distance function d_∞ uniformly on bounded sets with respect to d_∞. Let $x_0 \in X$. If*

$$\operatorname{diam}_{d_\infty}\Big(\bigcup_{j\in\mathbb{N}} B_{d_j}(x_0, R)\Big) < \infty, \quad \forall R > 0, \tag{12.1}$$

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 Fix $R, \delta > 0$. By (12.1) there is $R' > R$ such that $B_{d_j}(x_0, R) \subseteq B_{d_\infty}(x_0, R')$, for all $j \in \mathbb{N}$. Define

$$\bar\epsilon_j := \sup\{|d_j(x, y) - d_\infty(x, y)| : x, y \in B_{d_\infty}(x_0, R')\}.$$

Take $\epsilon_j := \begin{cases} \max\{R, \bar\epsilon_j\} & \text{if } \bar\epsilon_j \geq \delta \\ \bar\epsilon_j & \text{if } \bar\epsilon_j < \delta. \end{cases}$

By assumption of uniform convergence, we have $\epsilon_j \to 0$. To check that $\mathrm{id} : B_{d_j}(x_0, R) \to (X, d_\infty)$ is a $(1, \epsilon_j)$-quasi-isometric embedding, take $x, y \in$

$B_{d_j}(x_0, R) \subseteq B_{d_\infty}(x_0, R')$, then $|d_j(x, y) - d_\infty(x, y)| < \bar{\epsilon}_j \leq \epsilon_j$, by the definition of ϵ_j.

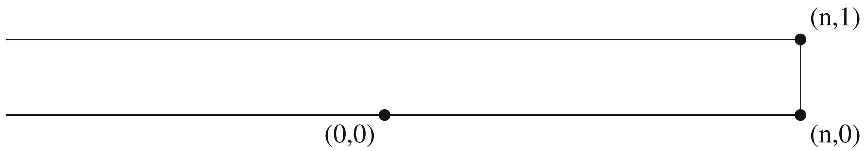
Regarding the fact that $B_{d_j}(x_0, R)$ is an ϵ_j-net for $B_{d_\infty}(x_0, R - \delta)$, we consider the two cases: Either $\bar{\epsilon}_j \geq \delta$ or not. In the first case, by definition, we have $\epsilon_j \geq R$, then, the ϵ_j-neighborhood, in the metric d_∞, of x_0 contains $B_{d_\infty}(x_0, R)$, which obviously contains $B_{d_\infty}(x_0, R - \delta)$. If, instead, we have $\bar{\epsilon}_j < \delta$, then for all $x \in B_{d_\infty}(x_0, R - \delta)$ we have

$$\begin{aligned} d_j(x, x_0) &\leq d_\infty(x, x_0) + \epsilon_j \\ &< R - \delta + \delta = R. \end{aligned}$$

So $B_{d_\infty}(x_0, R - \delta) \subseteq B_{d_j}(x_0, R)$. □

Example 12.1.4 The following example shows that condition (12.1) is necessary for the last proposition. For $n \in \mathbb{N}$ define $\gamma_n : \mathbb{R} \to \mathbb{R}^2$ by

$$\gamma_n(t) := \begin{cases} (t, 0) & t \leq n \\ (n, t - n) & n \leq t \leq n + 1 \\ (n - (t - n - 1), 1) & n + 1 \leq t \end{cases}$$



These mappings induce metrics d_n on $\mathbb{R}$ by

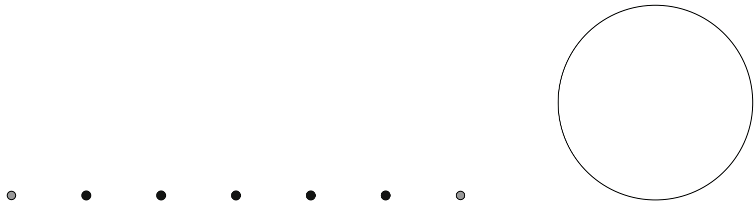
$$d_n(x, y) := d_{\mathbb{R}^2}(\gamma_n(x), \gamma_n(y)), \qquad \forall x, y \in \mathbb{R}.$$

Here, as $n \to \infty$, the sequence $d_n(x, y)$ converges 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. The sequence $(\mathbb{R}, d_n, 0)$ has a Gromov-Hausdorff limit, which however is isometric to $\left(\mathbb{R} \times \{0, 1\}, d_{\mathbb{R}^2}, (0, 0)\right)$.

12.1.2 Asymptotic Cones and Tangent Spaces

If $X = (X, d)$ is a metric space and $\lambda > 0$, we set $\lambda X := (X, \lambda d)$.

Definition 12.1.5 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



(a) Case of $X = \mathbb{Z}$. (b) Case of $X = \mathbb{S}^1$.

Fig. 12.1 The asymptotic cone of $\mathbb{Z}$ is $\mathbb{R}$, while at each point the tangent is a singleton. The asymptotic cone of $\mathbb{S}^1$ is a singleton, while at each point, the tangent is $\mathbb{R}$. (**a**) Case of $X = \mathbb{Z}$. (**b**) Case of $X = \mathbb{S}^1$

$(\lambda_j X, x) \to (Y, y)$, as $j \to \infty$. We say that (Y, y) is the *tangent metric 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 12.1.6 In general, asymptotic cones and tangent metric spaces may not exist. Within the space of boundedly compact metric spaces, limits are unique up to isometries. See Fig. 12.1 for some examples. The notion of asymptotic cone is independent of the base point x.

The following two theorems serve as the central focus of this chapter.

Theorem 12.1.7 (Pansu; [Pan83]; see Theorem 12.4.1) *Let G be a nilpotent simply connected Lie group equipped with a left-invariant sub-Finsler distance. Then, the asymptotic cone of G exists and is a sub-Finsler Carnot group.*

Theorem 12.1.8 (Mitchell; [Mit85]; see Theorems 12.5.1 and 12.6.3) *Let G be a sub-Finsler Lie group or, more generally, an equiregular sub-Finsler manifold, and $p \in G$. Then, the tangent metric space of G at p exists and is a sub-Finsler Carnot group.*

12.2 Varying CC Structures

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 3, while its asymptotic cone, the sub-Riemannian Heisenberg group, has a distribution of rank 2. To study the limits of CC spaces effectively, it is advantageous to adopt the perspective of sub-Finsler structures with possibly varying rank, as in Sect. 4.1.4.

12.2.1 Definition of Structures with Parameter

As in Definition 4.1.10, we use the language of bundles to have sub-Finsler structures where the rank could change. We consider families of CC-bundle structures that are continuously parameterized by a parameter.

Definition 12.2.1 (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 $f_\lambda := f(\lambda, \cdot, \cdot)$ is the bundle morphism and $N_\lambda := N(\lambda, \cdot, \cdot)$ is the continuously varying norm, as in Definition 4.1.10. We say that the family $\{(f_\lambda, N_\lambda)\}_{\lambda \in \Lambda}$ is a *smoothly varying CC-bundle structure* if

- $f \in C^\infty(\Lambda \times M \times \mathbb{R}^m)$;
- $N \in C^0(\Lambda \times M \times \mathbb{R}^m)$.

The above definition can be generalized to 'continuously varying Lipschitz-vector-fields structures', for which the results in this chapter have analogs; see Sect. 12.7 and [ALN23].

12.2.2 Divergence Bound By Grönwall Lemma

When comparing two sub-Finsler structures, as we will do when studying their convergence, we need to estimate the displacement of the endpoints of two curves in terms of the difference between their derivatives. These types of estimates are all consequences of the following general result, the so-called *Grönwall Lemma*.

Lemma 12.2.2 (Grönwall Lemma) *Let $\Omega \subset \mathbb{R}^n$ and $X, Y : \Omega \times [0, T] \to \mathbb{R}^n$. Let $\|\cdot\|$ be a Euclidean norm on $\mathbb{R}^n$. Suppose that there are $E, K > 0$ such that for all $p, q \in \Omega$ and all $t \in [0, T]$*

$$\|X(p,t) - Y(q,t)\| \le E + K\|p - q\|. \tag{12.2}$$

Let $\gamma, \eta : [0, T] \to \Omega$ be two absolutely continuous curves such that

$$\begin{cases} \gamma(0) = \eta(0), & \\ \dot\gamma(t) = X(\gamma(t), t), & \text{for almost every } t \in [0,T] \text{ and} \\ \dot\eta(t) = Y(\eta(t), t), & \text{for almost every } t \in [0,T]. \end{cases}$$

Then

$$\|\gamma(t) - \eta(t)\| \le E\frac{e^{Kt} - 1}{K}, \qquad \forall t \in [0, T]. \tag{12.3}$$

Proof Define $f(t) := \|\gamma(t) - \eta(t)\|$. Notice that $f : [0, T] \to \mathbb{R}$ is absolutely continuous and $f(0) = 0$. Moreover, for almost every $t \in [0, T]$ we have

$$\begin{aligned} 2f(t)f'(t) &= \frac{\mathrm{d}}{\mathrm{d}t}(f(t)^2) \\ &= 2\langle \gamma(t) - \eta(t), \dot\gamma(t) - \dot\eta(t)\rangle \\ &\le 2 \cdot \|\gamma(t) - \eta(t)\| \cdot \|\dot\gamma(t) - \dot\eta(t)\| \\ &= 2f(t) \cdot \|X(\gamma(t), t) - Y(\eta(t), t)\| \\ &\le 2f(t) \cdot (E + K\|\gamma(t) - \eta(t)\|) \\ &\le 2f(t) \cdot (E + Kf(t)). \end{aligned}$$

So, whenever $f(t) \neq 0$ we have

$$f'(t) \le E + Kf(t).$$

Let $g(t) := e^{-Kt}f(t)$. Then, whenever $g(t) \neq 0$, i.e., whenever $f(t) \neq 0$, we have

$$g'(t) = -Ke^{-Kt}f(t) + e^{-Kt}f'(t) \le -Ke^{-Kt}f(t) + e^{-Kt}(E + Kf(t)) = e^{-Kt}E.$$

We claim

$$g(t) \le \int_0^t e^{-Ks}E\,\mathrm{d}s, \text{ for almost every } t \in [0, T]. \tag{12.4}$$

Indeed, if $g(t) = 0$, then there is nothing to show because the right-hand side is nonnegative. If $g(t) > 0$ instead, since g is absolutely continuous, there is a maximal $\hat t < t$ such that $g(\hat t) = 0$, and we have

$$g(t) = g(\hat t) + \int_{\hat t}^t g'(s)\,\mathrm{d}s \le \int_{\hat t}^t e^{-Ks}E\,\mathrm{d}s \le \int_0^t e^{-Ks}E\,\mathrm{d}s,$$

and (12.4) is proved.

Finally, we obtain

$$e^{-Kt}f(t) = g(t) \le \int_0^t e^{-Ks}E\,\mathrm{d}s = \frac{E}{K}(1 - e^{-Kt}),$$

which gives (12.3). □

The following is a first application of the Grönwall Lemma that we will use in the next examples on Lie groups. We use the notation (7.3) for γ'.

Lemma 12.2.3 (Consequence of Grönwall Lemma) *Let G be a Lie group, $\|\cdot\|$ be a norm on $T_{1_G}G$, d a left-invariant Riemannian distance on G, and $\nu > 0$. Then there is $C > 0$ such that for all $\epsilon > 0$, for all absolutely continuous curves $\gamma, \sigma : [0, 1] \to G$ such that $\gamma(0) = \sigma(0)$, $\|\gamma'\|, \|\sigma'\| \leq \nu$ a.e., and $\|\gamma' - \sigma'\| < \epsilon$ a.e., then*

$$d(\gamma(1), \sigma(1)) \leq C\epsilon.$$

Proof Since d is left-invariant, we assume $\gamma(0) = \sigma(0) = 1_G$. Let U be a bounded neighborhood of 1_G that is a domain of exponential coordinates. We claim that there exists $\hat{C} > 0$ such that for all $\epsilon > 0$ and all absolutely continuous curves $\sigma, \gamma : [0, 1] \to U$ valued into U with $\|\gamma'(t)\|, \|\sigma'(t)\| \leq \nu$ and $\|\gamma'(t) - \sigma'(t)\| < \epsilon$ for a.e. $t \in [0, 1]$, we have

$$d(\gamma(1), \sigma(1)) \leq \hat{C}\epsilon. \tag{12.5}$$

We define the time-dependent vector fields X and Y on U by

$$X(g, t) := (L_g)_*\gamma'(t), \qquad Y(g, t) := (L_g)_*\sigma'(t).$$

Working in exponential coordinates, we will show an estimate as in (12.2). Since the map $(g, v) \mapsto (L_g)_*v$ is smooth, then it is Lipschitz on the bounded set $U \times B_{\|\cdot\|}(0, \nu)$. Hence there is $K > 0$ such that, for all $p, q \in U$ and for a.e. $t \in [0, 1]$,

$$\begin{aligned}\|X(p, t) - Y(q, t)\| &= \|(L_p)_*\gamma'(t) - (L_q)_*\sigma'(t)\| \\ &\leq \|(L_p)_*\gamma'(t) - (L_p)_*\sigma'(t)\| + \|(L_p)_*\sigma'(t) - (L_q)_*\sigma'(t)\| \\ &\leq K\|\gamma'(t) - \sigma'(t)\| + K\|p - q\| \leq K\epsilon + K\|p - q\|.\end{aligned}$$

So we have (12.2) with this K and $E := K\epsilon$. With these constants in (12.3), we conclude that $\|\gamma(1) - \sigma(1)\| \leq C\epsilon$ for some C. We finally obtain (12.5) from the fact that d and the norm $\|\cdot\|$ are bi-Lipschitz equivalent on U.

Next, since U is a neighborhood of 1_G, there is $\hat{\nu} > 0$ such that, if $\gamma : [0, 1] \to G$ is an absolutely continuous curve with $\gamma(0) = 1_G$ and $\|\gamma'(t)\| < \hat{\nu}$ for a.e. $t \in [0, 1]$, then $\gamma([0, 1]) \subset U$.

In order to prove the general statement of the lemma, for curves $\gamma, \sigma : [0, 1] \to G$ that are not necessarily valued into U, let $n \in \mathbb{N}$ be such that $(n - 1)\hat{\nu} < \nu \leq n\hat{\nu}$. For $j \in \{1, \dots, n\}$ define γ_j and $\sigma_j : [0, 1] \to G$ by

$$\gamma_j(t) := \gamma\left(\frac{j-1}{n}\right)^{-1} \gamma\left(\frac{j-1}{n} + \frac{t}{n}\right), \;\; \sigma_j(t) := \sigma\left(\frac{j-1}{n}\right)^{-1} \sigma\left(\frac{j-1}{n} + \frac{t}{n}\right).$$

Notice that

$$\gamma_j'(t) = \frac{1}{n}\gamma'\left(\frac{j-1}{n} + \frac{t}{n}\right),$$

and similarly for σ_j'. Therefore, $\|\gamma_j'\| < \nu/n \leq \hat{\nu}$, and $\|\sigma_j'\| < \nu/n \leq \hat{\nu}$, and $\|\gamma_j' - \sigma_j'\| < \epsilon/n$. It follows that the images of the curves γ_j and σ_j are in U. This shows, in particular, that the images of both γ and σ are in U^n because both curves are concatenations of left translations of the arcs γ_j and σ_j. Since U^n is bounded, there is L such that, for all $p \in U$ and all $x, y \in U^n$, we have $d(xp, yp) \leq Ld(x, y)$.

From the above claim, we have $d(\gamma(1/n), \sigma(1/n)) = d(\gamma_1(1), \sigma_1(1)) \leq \hat{C}\epsilon/n$. Next, assume that $j \in \{2, \ldots, n\}$ and that C_{j-1} is such that $d(\gamma((j-1)/n), \sigma((j-1)/n)) \leq C_{j-1}\epsilon$. Then,

$$\begin{aligned}
&d(\gamma(j/n), \sigma(j/n)) \\
&\quad\leq d\left(\gamma\left(\tfrac{j}{n}\right), \gamma\left(\tfrac{j-1}{n}\right)\sigma\left(\tfrac{j-1}{n}\right)^{-1}\sigma\left(\tfrac{j}{n}\right)\right) \\
&\qquad\qquad + d\left(\gamma\left(\tfrac{j-1}{n}\right)\sigma\left(\tfrac{j-1}{n}\right)^{-1}\sigma\left(\tfrac{j}{n}\right), \sigma\left(\tfrac{j}{n}\right)\right) \\
&\quad= d\left(\gamma_{j-1}(1), \sigma_{j-1}(1)\right) + d\left(\gamma\left(\tfrac{j-1}{n}\right)\sigma_{j-1}(1), \sigma\left(\tfrac{j-1}{n}\right)\sigma_{j-1}(1)\right) \\
&\quad\leq \hat{C}\epsilon + Ld\left(\gamma\left(\tfrac{j-1}{n}\right), \sigma\left(\tfrac{j-1}{n}\right)\right) \\
&\quad\leq (\hat{C} + LC_{j-1})\epsilon.
\end{aligned}$$

By iterating, we obtain that there is a positive number C that depends on ν but not on γ or σ such that $d(\gamma(1), \sigma(1)) \leq C\epsilon$. □

12.2.3 Analytically Varying Lie Structures

An example of varying CC structures is given by a normed vector space $(V, \|\cdot\|)$ and a one-parameter family of Lie group structures $\star_\lambda$ on V, or at least defined locally in a neighborhood of 0 in V. Each product gives rise to a left-invariant extension of the norm $\|\cdot\|$, seen initially as a norm on $T_{1_{\star_\lambda}} V \simeq V$. Instead of parametrizing the group structures, we prefer to parametrize the Lie bracket structures of the Lie algebras and consider the group products via the Dynkin formula; see Definition 5.7.3. We will assume that these structures depend analytically on the parameter and call them *analytically varying Lie-algebra structures*.

Lemma 12.2.4 *Let V be a vector space, $\Lambda \subset \mathbb{R}$ open, and $([\cdot,\cdot]_\lambda)_{\lambda\in\Lambda}$ a family of Lie brackets on V. Assume that the map $\lambda \mapsto [\cdot,\cdot]_\lambda$ is analytic. Denote by $\star_\lambda$ the Dynkin product associated with $[\cdot,\cdot]_\lambda$. Then, the maps*

$$(\lambda, x, y) \mapsto x \star_\lambda y \qquad \text{and} \qquad (\lambda, x, y) \mapsto \left.\frac{\mathrm{d}}{\mathrm{d}s} x \star_\lambda (sy)\right|_{s=0} \tag{12.6}$$

are analytic on some open subset of $\Lambda \times V \times V$ that contains $\Lambda \times \{0\} \times \{0\}$.

Moreover, if the Lie algebras $(V, [\cdot,\cdot]_\lambda)$ are nilpotent with uniform step and if $\lambda \mapsto [\cdot,\cdot]_\lambda$ is polynomial, then both of the above maps are polynomials from $\Lambda \times V \times V$ to V.

Proof Recall the alternative formula (5.25) for the Dynkin product: for X and Y in V,

$$X \star_\lambda Y = \sum_{n>0} \frac{(-1)^{n-1}}{n} \sum_{\substack{r_i+s_i>0\\1\le i\le n}} \frac{1}{r_1!s_1!\cdots r_n!s_n!\ \sum_{i=1}^n (r_i+s_i)} \cdot \\ \cdot \left((\mathrm{ad}_X^\lambda)^{r_1} \circ (\mathrm{ad}_Y^\lambda)^{s_1} \circ (\mathrm{ad}_X^\lambda)^{r_2} \circ (\mathrm{ad}_Y^\lambda)^{s_2} \ldots \circ (\mathrm{ad}_X^\lambda)^{r_n} \circ (\mathrm{ad}_Y^\lambda)^{s_n-1}\right)(Y),$$

where we denote $\mathrm{ad}_X^\lambda Y := [X, Y]_\lambda$. We define for $\alpha, \beta \in \mathfrak{gl}(V)$,

$$\mathscr{D}(\alpha,\beta) := \sum_{n>0} \frac{(-1)^{n-1}}{n} \sum_{\substack{r_i+s_i>0\\1\le i\le n}} \frac{\alpha^{r_1}\circ\beta^{s_1}\circ\alpha^{r_2}\circ\beta^{s_2}\ldots\circ\alpha^{r_n}\circ\beta^{s_n-1}}{r_1!s_1!\cdots r_n!s_n!\ \sum_{i=1}^n (r_i+s_i)}.$$

Using the fact that the series in (5.23) has a positive radius of convergence, we deduce that also the above series that defines $\mathscr{D}(\alpha,\beta)$ has a positive radius of convergence. Indeed, both series have the same monomials with different coefficients, and the coefficients for $\mathscr{D}(\alpha,\beta)$ are smaller. Therefore, $\mathscr{D}$ is a $\mathfrak{gl}(V)$-valued analytic map defined an open subset Ω of $\mathfrak{gl}(V) \times \mathfrak{gl}(V)$ that contains $\{0\} \times \{0\}$.

Since the function $(\lambda, X) \mapsto \mathrm{ad}_X^\lambda \in \mathfrak{gl}(V)$ is analytic, then the composition $(\lambda, X, Y) \mapsto \mathscr{D}(\mathrm{ad}_X^\lambda, \mathrm{ad}_Y^\lambda)$ and its derivatives are analytic on the open set

$$\left\{(\lambda, X, Y) \in \Lambda \times V \times V : (\lambda, \mathrm{ad}_X^\lambda, \mathrm{ad}_Y^\lambda) \in \Omega\right\},$$

which clearly contains $\Lambda \times \{0\} \times \{0\}$.

If all Lie-algebra structures $[\cdot,\cdot]_\lambda$ are nilpotent of step at most s, then $(\mathrm{ad}_X^\lambda)^s = 0$ for all $X \in V$ and $\lambda \in \Lambda$. Then, in the previous reasoning, we can truncate the series defining $\mathscr{D}$ to a finite sum. If also $(\lambda, X) \mapsto \mathrm{ad}_X^\lambda$ is polynomial, then the maps in (12.6) are polynomial. □

Remark 12.2.5 Let $(\star_\lambda)_{\lambda\in\Lambda}$ be the Dynkin product associated with an analytically varying Lie-algebra structure (as in Lemma 12.2.4) on a vector space V, equipped with a norm $\|\cdot\|$. For every fixed $\lambda \in \Lambda$ and $y \in V$, the map $p \in V \mapsto \frac{\mathrm{d}}{\mathrm{d}s} p \star_\lambda (sy)\Big|_{s=0}$ gives a vector field that is left-invariant with respect to $\star_\lambda$. Moreover, for $u : [0, 1] \to V$ and $\lambda \in \Lambda$, we define the time-dependent vector field

$$X^{u,\lambda}(t, p) := \frac{\mathrm{d}}{\mathrm{d}s} p \star_\lambda (su(t))\Big|_{s=0}. \tag{12.7}$$

As a consequence of Lemma 12.2.4 and Grönwall Lemma 12.2.2, for every compact sets $K \subseteq V$ and $\Lambda' \subset \Lambda$ and every positive constants C_1, C_2, and ℓ there exists C with the following property: Let $\epsilon > 0$, $u, v \in L^\infty([0, 1]; V)$ with $\|u-v\|_\infty < C_1\epsilon$, $\|u\|_\infty, \|v\|_\infty \leq \ell$ and $a, b \in \Lambda'$ with $|a - b| < C_2\epsilon$. Suppose α and $\beta : [0, 1] \to K$ are absolutely continuous integral curves of $X^{u,a}$ and $Y^{v,b}$, respectively, with $\alpha(0) = \beta(0)$. Then

$$\|\alpha(1) - \beta(1)\| \leq C\epsilon. \tag{12.8}$$

12.3 Examples of Limits

In this section, we present some examples that show the ideas for the more general strategy for proving Mitchell's and Pansu's theorems. The techniques are based on Grönwall estimates and on quantitative Chow's theorems, as in Sect. 7.1.4.

12.3.1 The Sub-Riemannian Roto-Translation Group

From the neurogeometry point of view, the most important sub-Riemannian manifold that is not a Carnot group is the rototranslation group; see page 20 or [SCP08]. For this reason, we present a special case of the proof of Mitchell's Theorem 12.1.8 for such a space.

Theorem 12.3.1 (See Proposition 12.3.3) *The tangent metric space of the sub-Riemannian rototranslation group is the sub-Riemannian Heisenberg group.*

The statement will be restated and proved in Proposition 12.3.3. For the argument, we will use the following result, which implies Chow's theorem and the Ball-Box theorem for the sub-Riemannian roto-translation group. Moreover, it gives a uniform estimate for sequences of structures. We denote the Euclidean balls by B_E.

Proposition 12.3.2 (Uniform Chow Theorem) *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. For each $p \in \mathbb{R}^3$ and $\lambda \in [0,1]$, consider the map (composition of flows)*

$$(t_1, t_2, t_3) \in \mathbb{R}^3 \mapsto \Phi^p_\lambda(t_1, t_2, t_3) \in \mathbb{R}^3$$

defined as

$$\Phi^p_\lambda(t_1,t_2,t_3) := \begin{cases} (\Phi^{\sqrt{t_3}}_{-Y_\lambda} \circ \Phi^{\sqrt{t_3}}_{-X_\lambda} \circ \Phi^{\sqrt{t_3}}_{Y_\lambda} \circ \Phi^{\sqrt{t_3}}_{X_\lambda} \circ \Phi^{t_2}_{Y_\lambda} \circ \Phi^{t_1}_{X_\lambda})(p), & \text{for } t_3 \geq 0 \\ (\Phi^{\sqrt{|t_3|}}_{Y_\lambda} \circ \Phi^{\sqrt{|t_3|}}_{-X_\lambda} \circ \Phi^{\sqrt{|t_3|}}_{-Y_\lambda} \circ \Phi^{\sqrt{|t_3|}}_{X_\lambda} \circ \Phi^{t_2}_{Y_\lambda} \circ \Phi^{t_1}_{X_\lambda})(p), & \text{for } t_3 < 0. \end{cases}$$

12.3.2.i. *The map Φ^p_λ is smooth and $(\mathrm{d}\Phi^p_\lambda)_0$ has maximal rank and the bi-Lipschitz constant of $(\mathrm{d}\Phi^p_\lambda)_0$ is uniformly bounded for (λ, p) on compact sets of $[0,1] \times \mathbb{R}^3$.*

12.3.2.ii. *For every $R > 0$ there exists $C > 0$ such that*

$$\Phi^p_\lambda(B_E(0, Cr)) \supset B_E(p, r), \qquad \forall \lambda \in [0,1], \forall r \in (0,R), \forall p \in B_E(0,R).$$

12.3.2.iii. *If d_λ is the sub-Riemannian distance for which X_λ, Y_λ are orthonormal, then for every $R > 0$ there exists $C > 0$ such that*

$$d_\lambda(p,q) \leq C\sqrt{d_E(p,q)}, \qquad \forall p, q \in B_E(0,R), \forall \lambda \in [0,1].$$

Proof of 12.3.2.i We already saw the following strategy in Sect. 7.1.4. We have $(\partial_{t_1}\Phi^p_\lambda)(0) = X_\lambda(p)$, $(\partial_{t_2}\Phi^p_\lambda)(0) = Y_\lambda(p)$, and $(\partial_{t_3}\Phi^p_\lambda)(0) = [X_\lambda, Y_\lambda](p)$. Hence, the linear map $(\mathrm{d}\Phi^p_\lambda)(0) : \mathbb{R}^3 \to \mathbb{R}^3$ has rank 3. Moreover, for every compact set U there is $C > 0$ such that

$$\|(\mathrm{d}\Phi^p_\lambda)(x)(v)\| \geq C\|v\|, \qquad \forall x \in U, \forall v \in \mathbb{R}^3, \forall \lambda \in [0,1].$$

By continuity in λ, we can take C uniform when $\lambda \in [0,1]$. □

Proof of 12.3.2.ii It follows from 12.3.2.i and the Inverse Mapping Theorem. □

Proof of 12.3.2.iii Notice that for some $K > 0$

$$\begin{aligned} d_\lambda(p, \Phi^p_\lambda(t_1,t_2,t_3)) &\leq |t_1| + |t_2| + 4\sqrt{|t_3|} \\ &\leq K\sqrt{\|(t_1,t_2,t_3)\|_E}, \qquad \forall t_1, t_2, t_3 \in (0,1). \end{aligned}$$

Let R as in 12.3.2.ii, 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, $d_\lambda(p, q) \leq K\sqrt{\|(t_1, t_2, t_3)\|_E} \leq K\sqrt{Cr} = K\sqrt{C}\sqrt{d_E(p, q)}$. □

Because of Proposition 12.1.3, the next proposition implies Theorem 12.3.1.

Proposition 12.3.3 *In $\mathbb{R}^3$ with coordinates x, y, θ let*

$$\begin{aligned} X &:= \cos\theta\,\partial_x + \sin\theta\,\partial_y \quad , \quad Y := \partial_\theta, \\ X_\infty &:= \partial_x + \theta\,\partial_y \quad , \quad Y_\infty := \partial_\theta, \\ X_n &:= \cos\tfrac{\theta}{n}\partial_x + n\sin\tfrac{\theta}{n}\partial_y \quad , \quad Y_n := \partial_\theta, \qquad \forall n \in \mathbb{N}. \end{aligned}$$

Let d (resp. d_n, resp d_∞) be the sub-Riemannian distance for which X, Y (resp. X_n, Y_n, resp. X_∞, Y_∞) are orthonormal. Then

12.3.3.i. *$(\mathbb{R}^3, nd)$ is isometric to $(\mathbb{R}^3, d_n)$, for each $n \in \mathbb{N}$.*
12.3.3.ii *$d_n \to d_\infty$ uniformly on compact sets, as $n \to \infty$.*

Proof of 12.3.3.i The distance nd is the sub-Riemannian distance associated with the orthonormal frame $\frac{1}{n}X, \frac{1}{n}Y$. Let $\delta_n : (x, y, \theta) \longmapsto (nx, n^2 y, n\theta)$. Then $\mathrm{d}\delta_n(\frac{1}{n}X) = \cos\theta\,\partial_x + n\sin\theta\,\partial_y = X_n \circ \delta_n$ and $\mathrm{d}\delta_n(\frac{1}{n}Y) = Y_n \circ \delta_n$. So δ_n is an isometry between $(\mathbb{R}^3, nd)$ and $(\mathbb{R}^3, d_n)$. □

Proof of 12.3.3.ii Fix $R > 0$. Take $p, q \in B_{d_\infty}(0, R)$. Let $\sigma : [0, 1] \to \mathbb{R}^3$ be a d_∞-geodesic from p to q, parametrized by constant speed. Thus $\|\dot\sigma\|_\infty < 2R$ and for some measurable functions a and b bounded by $2R$ we have $\dot\sigma = aX_\infty + bY_\infty$. Let γ be a solution of $\dot\gamma = aX_n + bY_n$ with $\gamma(0) = p$. 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 12.2.2, we get $|\gamma(1) - \sigma(1)| = o(1)$. Then, by Proposition 12.3.2

$$\begin{aligned} d_n(p, q) &\leq d_n(p, \gamma(1)) + d_n(\gamma(1), \sigma(1)) \\ &\overset{(12.3.2.\mathrm{iii})}{\leq} L_{d_n}(\gamma) + C\sqrt{|\gamma(1) - \sigma(1)|} \\ &\leq L_{d_\infty}(\sigma) + o(1) \\ &= d_\infty(p, q) + o(1). \end{aligned}$$

In particular, we have $d_n(p, 1) \leq 3R$ for n large enough. One bound has been proved.

Regarding the other bound, we proceed similarly: Let γ be a d_n-geodesic from p to q, $\gamma : [0, 1] \to \mathbb{R}^3$ with $\dot{\gamma} = aX_n + bY_n$, for some measurable functions a and b bounded by $3R$. Let σ be such that $\dot{\sigma} = aX_\infty + bY_\infty$ and $\sigma(0) = p$. Then as before $|\gamma(1) - \sigma(1)| = o(1)$. We then bound

$$\begin{aligned} d_\infty(p, q) &\leq d_\infty(p, \sigma(1)) + d_\infty(\sigma(1), \gamma(1)) \\ &\leq L_{d_\infty}(\sigma) + C\sqrt{|\sigma(1) - \gamma(1)|} \\ &\leq L_{d_n}(\gamma) + o(1) \\ &= d_n(p, q) + o(1). \end{aligned}$$

□

12.3.2 Sub-Riemannian Carnot Groups as Asymptotic Spaces of Riemannian Groups

This section aims to present some instances in which sub-Riemannian manifolds appear as limiting objects of Riemannian manifolds. Examples are asymptotic cones, and a general result is Pansu's Theorem 12.1.7. We will also see that every sub-Riemannian manifold is the limit of some sequence of Riemannian manifolds; see Theorem 12.3.7.

12.3.2.1 The Riemannian Heisenberg Group

The following result should be seen as a specific example of more general statements. In Sect. 12.4, we will study asymptotic cones of nilpotent Lie groups.

Theorem 12.3.4 *The asymptotic cone of the Riemannian Heisenberg group is the sub-Riemannian Heisenberg group.*

We explain a stronger result that we will actually prove: Let X, Y, Z be a basis of the Lie algebra of the Heisenberg group $\mathcal{H}$ with only relation $[X, Y] = Z$. For $n \in \mathbb{N}$, let d_n be the Riemannian distance for which $X, Y, \frac{1}{n}Z$ form an orthonormal frame. Let d_{cc} be the sub-Riemannian distance for which X, Y are orthonormal. Then we will prove that for all $R > 0$ there is $C > 0$ such that

$$d_n(p, q) \leq d_{\mathrm{cc}}(p, q) \leq d_n(p, q) + C\frac{1}{\sqrt{n}}, \qquad \forall p, q \in B_{CC}(1_{\mathcal{H}}, R), \forall n \in \mathbb{N}. \tag{12.9}$$

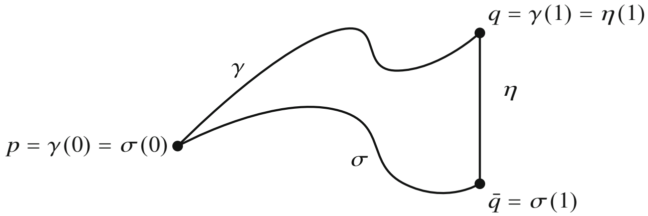


Fig. 12.2 The three curves γ, η, and σ considered in the proof of Theorem 12.3.4

In particular, we have $d_n \to d_\infty$ uniformly on compact sets. Consequently, by Proposition 12.1.3 if d is the Riemannian distance for which X, Y, Z are orthonormal, then the sequence $(\mathcal{H}, \frac{1}{n}d)$, which is isometric to $(\mathcal{H}, d_n)$, converges to $(\mathcal{H}, d_{cc})$.

Proof of Theorem 12.3.4 The fact that $d_n \leq d_{cc}$ is clear since every horizontal curve for d_{cc} has the same length with respect to d_n. For the other inequality, take $R > 0$ and $p, q \in B_{CC}(1_{\mathcal{H}}, R)$. Let $\gamma_n : [0, 1] \to \mathcal{H}$ 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 $\sigma : [0, 1] \to \mathcal{H}$ 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 $\eta : [0, 1] \to \mathcal{H}$ be the curve such that $\eta(0) = \bar{q}$ and $\dot{\eta}(t) = a_3(t)Z$. Refer to Fig. 12.2 for a visual representation.

We claim that

$$\eta(t) = (L_{\bar{q}} \circ L^{-1}_{\sigma(t)})(\gamma(t)), \qquad \forall t \in [0, 1]. \tag{12.10}$$

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, we consider exponential coordinates as in Sect. 2.3:

$$\dot{\gamma} = a_1X + a_2Y + a_3Z = \left(a_1, a_2, a_3 - \frac{\gamma_2}{2}a_1 + \frac{\gamma_1}{2}a_2\right)$$

and

$$\dot{\sigma} = \left(a_1, a_2, -\frac{\sigma_2}{2}a_1 + \frac{\sigma_1}{2}a_2\right).$$

Thus $\gamma_1(t) = \sigma_1(t) = p_1 + \int_0^t a_1$ and $\gamma_2(t) = \sigma_2(t) = p_2 + \int_0^t a_2$. We have

$$\sigma(t)^{-1}\gamma(t) \overset{(2.9)}{=} \left(\gamma_1 - \sigma_1, \gamma_2 - \sigma_2, \gamma_3 - \sigma_3 - \frac{1}{2}(\sigma_1\gamma_2 - \sigma_2\gamma_1)\right)$$
$$= (0, 0, \gamma_3 - \sigma_3).$$

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 (12.10) is proved, and, in particular, we have $\eta(1) = \bar q \bar q^{-1} q = q$.

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

$$\int_0^1 n \cdot |a_3| = \int_0^1 \|a_3 Z\|_n$$
$$\leq \int_0^1 \|a_1 X + a_2 Y + a_3 Z\|_n = L_{d_n}(\gamma) = d_n(p, q) \leq d_{\mathrm{cc}}(p, q) \leq 2R.$$

Then

$$L_{d_1}(\eta) = \int_0^1 \|a_3 Z\|_1 = \int_0^1 |a_3| \leq \frac{2R}{n}.$$

Using the Ball-Box Theorem for the sub-Riemannian Heisenberg group, proved in Proposition 2.4.6, we bound

$$d_{\mathrm{cc}}(\bar q, q) \overset{(4.21)}{\leq} K d_1(\bar q, q)^{1/2} \leq K(L_{d_1}(\eta))^{1/2} \leq K\left(\frac{2R}{n}\right)^{1/2} = C\frac{1}{\sqrt{n}},$$

for the suitable constant C. Since $d_{\mathrm{cc}}(p, \bar q) \leq L_{CC}(\sigma) \leq L_{d_n}(\gamma) = d_n(p, q)$, we conclude that

$$d_{\mathrm{cc}}(p, q) \leq d_{\mathrm{cc}}(p, \bar q) + d_{\mathrm{cc}}(\bar q, q) \leq d_n(p, q) + C\frac{1}{\sqrt{n}}.$$

We proved (12.9), and we conclude the theorem recalling Proposition 12.1.3. □

12.3.2.2 Asymptotic Cones of Some Riemannian Groups

Example 12.3.5 Let G be a Lie group and $\Delta \subset TG$ 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

12.3.5.i. Δ_{1_G} is orthogonal to $\Delta_{1_G}^{\perp}$ with respect to $\langle\cdot,\cdot\rangle_n$, for all $n\in\mathbb{N}\cup\{0\}$;
12.3.5.ii. $\|X\|_n = \|X\|_0$, for all $n\in\mathbb{N}$, for all $X\in\Delta$;
12.3.5.iii. for all $X\notin\Delta$, we have $\|X\|_n\to+\infty$, as $n\to\infty$.

Let d_{cc} be the sub-Riemannian distance associated with Δ and $\langle\cdot,\cdot\rangle_0$, and d_n the Riemannian distance associated with $\langle\cdot,\cdot\rangle_n$. 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

$$d_n(p,q)\le d_{cc}(p,q)\le d_n(p,q)+\epsilon_n, \qquad \forall p,q\in B_{CC}(1_G,R). \tag{12.11}$$

Proof The left-hand side of (12.11) is obvious from 12.3.5.ii. For the right-hand side, begin by noticing that the unit tangent bundle of $\Delta_{1_G}^{\perp}$ is compact. Consequently from 12.3.5.iii, there exists a diverging sequence K_n in $\mathbb{R}$ such that

$$K_n\cdot\|Z\|_0\le\|Z\|_n, \qquad \forall Z\in\Delta_{1_G}^{\perp}, \forall n\in\mathbb{N}. \tag{12.12}$$

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 constant-speed 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 12.3.5.i we know that $Z\perp X$ for each 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

$$K_n\cdot\|Z\|_0\overset{(12.12)}{\le}\|Z\|_n\overset{Z\perp X}{\le}\|X+Z\|_n=\|\dot\gamma\|_n<2R.$$

Let $\xi(r):=\operatorname{diam}_{d_{cc}}(\overline{B}_{d_0}(1_G,r))$ as in Exercise 12.9.1. Then we are going to use Grönwall Lemma 12.2.3 since $\|\gamma'\|_0,\|\sigma'\|_0\le\|\gamma'\|_n<2R$ and $\|\gamma'-\sigma'\|_0=\|Z\|_0<\frac{2R}{K}_n$ and get

$$\begin{aligned}
d_{cc}(p,q)&\le d_{cc}(p,\sigma(1))+d_{cc}(\sigma(1),\gamma(1))\\
&\le L_{CC}(\sigma)+\xi(d_0(\sigma(1),\gamma(1)))\\
&\le L_{d_n}(\gamma)+\xi\left(C\cdot\frac{2R}{K_n}\right)\\
&= d_n(p,q)+o(1) \quad ,\text{as } n\to\infty,
\end{aligned}$$

where we used that $K_n\to\infty$ and that $\xi(r)=o(1)$ by Exercise 12.9.1. □

The next statement is an easy consequence of what we showed in Example 12.3.5.

Corollary 12.3.6 *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 via $\delta_{1/\lambda}$ to (G, d_λ), and (G, d_{cc}) is a Carnot group.*

12.3.3 Sub-Riemannian Spaces as Monotone Limits of Riemannian Spaces

In this subsection, we show that every sub-Riemannian distance is a *monotone* limit of Riemannian distances. This fact has been well-known in nonholonomic geometry for the last 35 years, and it is probably due to V. Gershkovich.

Theorem 12.3.7 *Every sub-Riemannian distance is an increasing limit of Riemannian distances.*

The argument is easy and well-known when the sub-Riemannian distribution has constant rank. Here, we soon present a proof in the following general case: As in Definition 4.1.10, let E be a vector bundle over M endowed with a scalar product $\langle \cdot, \cdot \rangle$ and let $\sigma : E \to TM$ be a morphism of vector bundles. For each $p \in M$ and $v, v' \in T_pM$, set

$$\rho_p(v, v') := \inf\{\langle u, u' \rangle : u, u' \in E_p, \sigma(u) = v, \sigma(u') = v'\},$$
$$\rho_p(v) := \rho_p(v, v), \qquad \text{and} \qquad N_p(v) := \sqrt{\rho_p(v)}. \tag{12.13}$$

The sub-Riemannian distance d_ρ associated with ρ is given by (4.9). The only extra assumption on ρ is that the distance d_ρ is finite and induces the manifold topology, as for example, in Definition 4.1.6 by Theorem 4.1.8.

Proof of Theorem 12.3.7 Let ρ and $d := d_\rho$ as in (12.13) and (4.9). Notice that on each tangent space T_pM, with $p \in M$, the value of $\rho(\cdot, \cdot)$ is finite on some subspace, and on this it defines a scalar product. In particular, we have $\rho(v) = 0$ only if $v = 0$. We take some Riemannian tensor g_1 with the property that $g_1 \leq \rho$. Then, by recurrence, for each $m \in \mathbb{N}$, we consider g_m to be a (smooth) Riemannian tensor with the property that, for every $p \in M$ and $v, w \in T_pM$,

$$\max\{(g_{m-1})_p(v, w), \min\{(1 - 2^{-m})\rho_p(v, w), m(g_1)_p(v, w)\}\} \\ \leq (g_m)_p(v, w) \leq \rho_p(v, w).$$

Obviously, we have that

$$g_1 \le g_m \le g_{m+1} \le \rho.$$

Then, for every absolutely continuous path γ, we have that

$$L_{g_m}(\gamma) \le L_\rho(\gamma) := \int_0^1 \sqrt{\rho_{\gamma(t)}(\dot\gamma(t))}\,\mathrm{d}t.$$

Thus, for every p and q in M,

$$d_{g_m}(p,q) \le d_\rho(p,q),$$

and therefore

$$\lim_{m\to\infty} d_{g_m}(p,q) \le d_\rho(p,q).$$

Assume, by contradiction, that, for some p and q in M, we have that

$$\lim_{m\to\infty} d_{g_m}(p,q) < d_\rho(p,q).$$

Then there are curves γ_m from p to q such that

$$\lim_{m\to\infty} L_{g_m}(\gamma_m) < d_\rho(p,q).$$

Since

$$L_{g_1}(\gamma_m) \le L_{g_m}(\gamma_m),$$

we get a bound on the lengths $L_{g_1}(\gamma_m)$. Therefore, by Ascoli-Arzelà argument (Theorem 3.1.3), the sequence γ_m converges to a curve γ from p to q. We may assume that γ is parametrized by arc length with respect to the distance of g_1. Now, either $L_\rho(\gamma)$ is infinite, or it is finite. Namely, either there is a positive-measure set $A \subset [0, L_{g_1}(\gamma)]$ such that

$$\rho_{\gamma(t)}(\dot\gamma(t)) = \infty, \qquad \forall t \in A,$$

or, for almost every $t \in [0, L_{g_1}(\gamma)]$, the value $\rho_{\gamma(t)}(\dot\gamma(t))$ is finite. In the first case, for all $t \in A$,

$$(g_m)_{\gamma(t)}(\dot\gamma(t)) \ge m \cdot (g_1)_{\gamma(t)}(\dot\gamma(t)).$$

From this, we have that

$$L_{g_m}(\gamma) \geq m L_{g_1}(\gamma|_A) \to \infty, \qquad \text{as } m \to \infty.$$

We get a contradiction since by assumption $d_\rho(p, q) < \infty$. In the second case, for almost all t, for m big enough,

$$(1 - 2^{-m})\rho_{\gamma(t)}(\dot{\gamma}(t)) \leq (g_m)_{\gamma(t)}(\dot{\gamma}(t)) \leq \rho_{\gamma(t)}(\dot{\gamma}(t)).$$

From this, we have that

$$L_{g_m}(\gamma) \to L_\rho(\gamma), \qquad \text{as } m \to \infty.$$

We get a contradiction since we have that $d_\rho(p, q) \leq L_\rho(\gamma)$. □

12.4 Pansu Asymptotic Theorem

In this section, we explain and prove Pansu's Theorem 12.1.7. Let G be a Lie group with Lie algebra $\mathfrak{g}$. Let Δ be a left-invariant bracket-generating polarization corresponding to the subspace $\Delta_{1_G} \subset \mathfrak{g}$, as in (7.1). Fix a norm $\|\cdot\|$ on Δ_{1_G} and transfer it to a left-invariant norm on Δ as in (7.5). Let d_{sF} be the associated sub-Finsler metric on G as in (4.4). In addition, we assume that G is simply connected and nilpotent. Consequently, the Lie algebra $\mathfrak{g}$ is nilpotent, by Proposition 9.4.2.

12.4.1 Pansu Asymptotic Structure

Given the nilpotent Lie algebra $\mathfrak{g}$, we consider the associated Carnot algebra $\mathfrak{g}_\infty$ of G, as in Definition 9.2.16. This is a stratifiable algebra, and $\mathfrak{g}/[\mathfrak{g}, \mathfrak{g}]$ is the first stratum of a stratification. Consider G_∞ to be the nilpotent simply connected Lie group with Lie algebra $\mathfrak{g}_\infty$, by recalling the existence by Birkhoff Theorem; see Proposition 9.4.3. We shall put a Carnot metric on it. Consider the projection $\pi : \mathfrak{g} \to \mathfrak{g}/[\mathfrak{g}, \mathfrak{g}]$ modulo $[\mathfrak{g}, \mathfrak{g}]$. Since Δ_{1_G} generates the whole Lie algebra and since $\mathfrak{g}/[\mathfrak{g}, \mathfrak{g}]$ is an abelian Lie algebra, we have $\pi(\Delta_{1_G}) = \mathfrak{g}/[\mathfrak{g}, \mathfrak{g}]$. We define a norm $\|\cdot\|_\infty$, on $\mathfrak{g}/[\mathfrak{g}, \mathfrak{g}]$, which we name *Pansu limit norm*, by imposing $B_{\|\cdot\|_\infty}(0, 1) = \pi(B_{\|\cdot\|}(0, 1))$. Let d_∞ be the sub-Finsler distance associated with $(G_\infty, \mathfrak{g}/[\mathfrak{g}, \mathfrak{g}], \|\cdot\|_\infty)$. Thus (G_∞, d_∞) is a Carnot group. We call d_∞ the *Pansu limit metric* or the *asymptotic metric*.

Pansu Theorem 12.1.7 will follow from a more quantitative version as obtained in [BL13]:

Theorem 12.4.1 ([Pan83, BL13]; see Theorem 12.4.5) *Let (G, d) be a nilpotent simply connected Lie group equipped with a sub-Finsler left-invariant metric. Then, the associated Carnot group G_∞ equipped with the Pansu limit metric d_∞ is the asymptotic cone of (G, d) and there is a set-wise identification between G and G_∞ such that*

$$\left|d(p,q) - d_\infty(p,q)\right| = O\left(\max\{d(1,p), d(1,q)\}^{1-1/s}\right), \qquad \text{as } p, q \to \infty, \tag{12.14}$$

where s denotes the nilpotency step of G.

12.4.2 Structures of Contracted Metrics

12.4.2.1 Algebra-Group(s) Identification

There is a natural way to identify the underlying vector space of $\mathfrak{g}$ and the one of its associated Carnot algebra $\mathfrak{g}_\infty$. This identification depends on the choice of a compatible linear grading on $\mathfrak{g}$, as defined in Definition 9.2.2. The natural identification $\mathfrak{g} \simeq \mathfrak{g}_\infty$ has been discussed in Lemma 9.2.17. Moreover, via the exponential maps, which in this case are global diffeomorphisms, by Theorem 9.4.6, we also have identification at the group level:

$$G \underbrace{\simeq}_{\exp} \mathfrak{g} \overbrace{\simeq}^{\text{compatible grading}} \mathfrak{g}_\infty \underbrace{\simeq}_{\exp} G_\infty. \tag{12.15}$$

Hence, we may consider the Pansu limit metric d_∞ as a metric on G or on the vector space $\mathfrak{g}$.

12.4.2.2 A Family of Products and Metrics

Consider the dilations $(\delta_\epsilon)_{\epsilon\in\mathbb{R}}$ relative to the compatible linear grading as in (9.8), we stress that the family is defined also for nonpositive ϵ and it is polynomial in ϵ. Using these maps, we shall modify the Lie group structure of G and its metric into a one-parameter family. Recall that we are identifying G with $\mathfrak{g}$, as in (12.15).

We modify the Lie bracket of $\mathfrak{g}$ as

$$[X, Y]_\epsilon := \delta_\epsilon[\delta_\epsilon^{-1}X, \delta_\epsilon^{-1}Y], \qquad \forall X, Y \in \mathfrak{g}, \forall \epsilon \in \mathbb{R} \setminus \{0\}. \tag{12.16}$$

We stress that for $X, Y \in \mathfrak{g}$ the value $[X, Y]_\epsilon$, as defined for $\epsilon \neq 0$ is polynomially depending on ϵ and, by (9.10), it extends to $\epsilon = 0$ as

$$[X, Y]_0 := [\![X, Y]\!]_\infty \underset{(9.10)}{\overset{\text{def}}{=}} \lim_{\epsilon \to 0} [X, Y]_\epsilon, \tag{12.17}$$

where $[\![X, Y]\!]_\infty$ denotes the Lie bracket on the associated Carnot algebra $\mathfrak{g}_\infty$.

For $\epsilon \in \mathbb{R}$, we denote by $\star_\epsilon$ the Dynkin product associated with the nilpotent Lie bracket $[\cdot, \cdot]_\epsilon$; see Definition 5.7.3. The group $(\mathfrak{g}, \star_0)$ is then the Carnot group with Lie algebra $\mathfrak{g}_\infty$; see Proposition 9.4.4.

We also modify the metric $d =: \rho_1$ on G with the *contracted metrics* defined as

$$\rho_\epsilon(p, q) := \begin{cases} |\epsilon| d(\delta_\epsilon^{-1} p, \delta_\epsilon^{-1} q) & \forall \epsilon \in \mathbb{R} \setminus \{0\} \\ d_\infty(p, q) & \text{for } \epsilon = 0 \end{cases}, \qquad \forall p, q \in G, \tag{12.18}$$

where d_∞ is the Pansu limit metric from page 383. We point out that each distance ρ_ϵ is $\star_\epsilon$-left-invariant and, for $\epsilon \neq 0$, it is the sub-Finsler distance induced by the following distribution and norm:

$$\Delta_1^{(\epsilon)} := \delta_\epsilon(\Delta_1) \qquad \text{and} \qquad \|v\|^{(\epsilon)} := |\epsilon| \cdot \left\| \delta_\epsilon^{-1} v \right\|, \quad \forall v \in \Delta_1^{(\epsilon)}; \tag{12.19}$$

see Exercise 12.9.8. Therefore, for each $\epsilon \neq 0$, each map $\delta_\epsilon : (\mathfrak{g}, \star_1, |\epsilon| d) \to (\mathfrak{g}_\epsilon, \star_\epsilon, \rho_\epsilon)$ is a Lie-algebra isomorphism and an isometry.

Remark 12.4.2 Regarding the norm $\|\cdot\|^{(\epsilon)}$, we claim that, as soon as we fix a norm $\|\cdot\|_E$ on $\mathfrak{g}$, there is a positive number C such that

$$\|u - \pi(u)\|_E \leq C\epsilon \|u\|^{(\epsilon)}, \qquad \forall u \in \Delta_1^{(\epsilon)}, \forall \epsilon \in (0, 1). \tag{12.20}$$

Indeed, we assume, without loss of generality, that $\|\cdot\|_E$ is Euclidean, that makes orthogonal the layers $(V_k)_k$ of the compatible linear grading, and that $\|\cdot\|_E \leq \|\cdot\|_1$ on Δ_1. We write $u \in \Delta_1^{(\epsilon)}$ as $u = u_1 + \cdots + u_s$ with $u_k \in V_k$. Then, for each $k \in \{2, \ldots, s\}$ and $\epsilon \in (0, 1)$, we bound

$$\begin{aligned} \epsilon^{-1} \|u_k\|_E \leq \epsilon^{1-k} \|u_k\|_E &\leq \|u_1 + \epsilon^{-1} u_2 + \cdots + \epsilon^{1-s} u_s\|_E \\ &= \|\epsilon \delta_\epsilon^{-1} u\|_E \leq \|\epsilon \delta_\epsilon^{-1} u\|_1 \overset{\text{def}}{=} \|u\|^{(\epsilon)}. \end{aligned}$$

Therefore, we obtain (12.20):

$$\|u - \pi(u)\|_E \leq (\|u_2\|_E + \cdots + \|u_s\|_E) \leq (s-1)\epsilon \|u\|^{(\epsilon)}.$$

All these sub-Finsler structures, $(\Delta^{(\epsilon)}, \|\cdot\|^{(\epsilon)})$, make the abelianization map a submetry:

Proposition 12.4.3 *Under the identifications* $V_1 = G/[G, G] = G_\infty/[G_\infty, G_\infty]$*, the following maps are submetries:*

12.4.3.i. *The projection map* $\pi|_{\Delta_{1_G}} : (\Delta_{1_G}, \| \cdot \|) \to (V_1, \| \cdot \|_\infty)$*, modulo* $[\mathfrak{g}, \mathfrak{g}]$.
12.4.3.ii. *The projection map* $\pi : (G, d) \to (V_1, \| \cdot \|_\infty)$*, modulo* $[G, G]$.
12.4.3.iii. *The projection map* $\pi : (G_\infty, d_\infty) \to (V_1, \| \cdot \|_\infty)$*, modulo* $[G_\infty, G_\infty]$.
12.4.3.iv. *The projection map* $\pi : (\mathfrak{g}, \rho_\epsilon) \to (V_1, \|\cdot\|_\infty)$*, modulo* $[\mathfrak{g}, \mathfrak{g}]$.

Proof First, observe that the maps are well defined since the subspace V_1 is a complementary subspace of the commutator subalgebra, i.e., $\mathfrak{g} = [\mathfrak{g}, \mathfrak{g}] \oplus V_1$. These maps are group homomorphisms. The map 12.4.3.i is a submetry by the way the norm $\| \cdot \|_\infty$ has been defined. It is clear that the maps 12.4.3.ii and 12.4.3.iii are submetries, for example, as a consequence of Proposition 7.1.9. Regarding the map 12.4.3.iv, the reason is again that the norm in the target is the push forward of the norm in the domain:

$$
\begin{aligned}
\pi\left(B_{\|\cdot\|^{(\epsilon)}}(0, 1)\right) &= \pi\left(\delta_\epsilon B_{\|\cdot\|^{(1)}}(0, 1/\epsilon)\right) \\
&= \epsilon\pi\left(B_{\|\cdot\|^{(1)}}(0, 1/\epsilon)\right) \\
&= \epsilon B_{\|\cdot\|_\infty}(0, 1/\epsilon) \\
&= B_{\|\cdot\|_\infty}(0, 1),
\end{aligned}
$$

where we used the definition of $\|\cdot\|^{(\epsilon)}$ and of $\|\cdot\|_\infty$. □

12.4.2.3 Guivarc'h Apriori Bound

The distance function $d_\infty(1, \cdot)$ on G is a homogeneous quasi-norm with respect to the compatible linear grading, in the sense of Definition 10.4.1. Therefore, we have an immediate consequence of Guivarc'h Theorem 10.4.3:

Corollary 12.4.4 (Guivarc'h) *Let G be a nilpotent simply connected Lie group equipped with a left-invariant sub-Finsler metric d. Let d_∞ be the Pansu limit metric on G. Then there exists a constant $C > 1$ such that*

$$\frac{1}{C} d_\infty(1, x) - C \le d(1, x) \le C d_\infty(1, x) + C, \quad \forall x \in G. \tag{12.21}$$

in terms of the distances ρ_ϵ, as in (12.18), we have

$$\frac{1}{C}\rho_0(1, x) - C\epsilon \le \rho_\epsilon(1, x) \le C\rho_0(1, x) + C\epsilon, \quad \forall x \in G, \forall\epsilon \ge 0. \tag{12.22}$$

12.4.3 Proof of Pansu Asymptotic Theorem 12.1.7

We are ready to state and prove a quantitative version of Pansu Asymptotic Theorem 12.1.7, which implies Theorem 12.4.1.

Theorem 12.4.5 (Quantitative Pansu Theorem; [BL13]) *Let G be a simply connected s-step nilpotent Lie group equipped with a left-invariant sub-Finsler metric ρ_1. With respect to some compatible linear grading, consider the Pansu limit metric ρ_0 on G and the contracted metrics ρ_ϵ as in* (12.18)*. For every compact set $K \subseteq G$ there is a constant $C > 0$ such that*

$$|\rho_\epsilon(p,q) - \rho_0(p,q)| \leq C\epsilon^{1/s}, \qquad \forall p, q \in K, \forall \epsilon \in (0,1). \tag{12.23}$$

Proof Without loss of generality, because of dilations, we assume $K = B_{\rho_0}(1_G, 1)$. Fix $p, q \in K$. We begin with **a first inequality**: $\rho_0(p,q) \leq \rho_\epsilon(p,q) + C_1\epsilon^{1/s}$, for some constant $C_1 \geq 0$ independent on p and q. Let C be the Guivarc'h constant from Corollary 12.4.4. We have the uniform bound:

$$\begin{aligned}\rho_\epsilon(p,q) &\leq \rho_\epsilon(p, 1_G) + \rho_\epsilon(1_G, q)\\ &\overset{(12.22)}{\leq} C\rho_0(1_G, p) + C\epsilon + C\rho_0(1_G, q) + C\epsilon < 4C,\end{aligned} \tag{12.24}$$

where we used, in addition to the triangle inequality, the fact that p and q are in the unit ρ_0-ball. Let $\gamma_\epsilon : [0,1] \longrightarrow \mathfrak{g}$ be a ρ_ϵ-geodesic parametrized by constant ρ_ϵ-speed connecting p to q. We check that, independently on $\epsilon \in (0,1)$, the curve γ_ϵ is inside the bounded set $B_{\rho_0}(1_G, 6C^2 + C)$:

$$\begin{aligned}\rho_0(1_G, \gamma_\epsilon(t)) &\overset{(12.22)}{\leq} C\rho_\epsilon(1_G, \gamma_\epsilon(t)) + C\epsilon\\ &\leq C(\rho_\epsilon(1_G, p) + \rho_\epsilon(p, \gamma_\epsilon(t))) + C\epsilon\\ &\overset{(12.22)}{\leq} C(C\rho_0(1_G, p) + C\epsilon + \rho_\epsilon(p,q)) + C\epsilon\\ &\overset{(12.24)}{<} 6C^2 + C.\end{aligned}$$

Let $\gamma_0 \colon [0,1] \longrightarrow \mathfrak{g}$ be the left translation by p of the multiplicative integral in G_∞ of the curve $\pi \circ \gamma_\epsilon$, as in Proposition 10.1.11. Namely, if the curve γ_ϵ has the control $u \colon [0,1] \longrightarrow \mathfrak{g}$ for the group structure $\star_\epsilon$, then γ_0 is such that $\gamma_0(0) = p$ and it has the control $\pi \circ u$ for the group structure $\star_0$:

$$\begin{cases}\dot\gamma_\epsilon(t) = (L^\epsilon_{\gamma_\epsilon(t)})_* u(t) := \frac{\mathrm{d}}{\mathrm{d}s}\gamma_\epsilon(t) \star_\epsilon (su(t))\Big|_{s=0} & ,\\ \dot\gamma_0(t) = (L^0_{\gamma_0(t)})_* \pi(u(t)) := \frac{\mathrm{d}}{\mathrm{d}s}\gamma_0(t) \star_0 (s\pi(u(t)))\Big|_{s=0} & .\end{cases}$$

In particular, the curves γ_0 and $\pi \circ \gamma_\epsilon$ have the same length in their respective spaces: (G_∞, ρ_0) and $(V_1, ||\cdot||_\infty)$. Therefore, using in addition that $\pi : (\mathfrak{g}, \rho_\epsilon) \to (V_1, \|\cdot\|_\infty)$ is 1-Lipschitz by Proposition 12.4.3, we have

$$\begin{aligned} L_{\rho_0}(\gamma_0) &= L_{||\cdot||_\infty}(\pi \circ \gamma_\epsilon) \\ &\leq L_{\rho_\epsilon}(\gamma_\epsilon) \\ &= \rho_\epsilon(p, q) \\ &\overset{(12.24)}{\leq} 4C \end{aligned} \tag{12.25}$$

and also,

$$\|u(t)\|^{(\epsilon)} \overset{(12.24)}{\leq} 4C, \qquad \text{for a.e. } t \in [0, 1], \tag{12.26}$$

and

$$\begin{aligned} \rho_0(1_G, \gamma_0(t)) &\leq \rho_0(1_G, p) + \rho_0(p, \gamma_0(t)) \\ &\leq 1 + L_{\rho_0}(\gamma_0) \\ &\leq 1 + 4C. \end{aligned}$$

We have established that the curves γ_ϵ and γ_0 lie in the bounded set $B_{\rho_0}(1_G, 1 + 4C + 6C^2)$.

Moreover, fixed an Euclidean norm $\|\cdot\|_E$, from (12.20), together with (12.26), we obtain

$$\|u - \pi \circ u\|_E \leq (s - 1)4C\epsilon. \tag{12.27}$$

We stress that γ_ϵ is an integral curve of the vector field $X^{u,\epsilon}$ as defined in (12.7). While the curve γ_0 is an integral curve of $X^{\pi \circ u, 0}$. As a consequence of Grönwall Lemma and the smoothness of the Lie structures in ϵ, seen in Remark 12.2.5, together with (12.26) and (12.27), we obtain that $\|\gamma_0(1) - \gamma_\epsilon(1)\|_E \leq \hat{C}\epsilon$. Consequently, by the Ball-Box Theorem for Carnot groups (Theorem 11.2.3), we have for some constant $C' > 0$

$$\rho_0(\gamma_0(1), \gamma_\epsilon(1)) \leq C'\epsilon^{1/s}. \tag{12.28}$$

Finally, we obtain the first bound:

$$\begin{aligned} \rho_0(p, q) &\leq \rho_0(p, \gamma_0(1)) + \rho_0(\gamma_0(1), \gamma_\epsilon(1)) \\ &\overset{(12.28)}{\leq} L_{\rho_0}(\gamma_0) + C'\epsilon^{1/s} \\ &\overset{(12.25)}{\leq} \rho_\epsilon(p, q) + C'\epsilon^{1/s}. \end{aligned}$$

We consider the **second inequality**: $\rho_\epsilon(p,q) \leq \rho_0(p,q) + C_2\epsilon^{1/s}$, for some $C_2 \geq 0$. Let $\gamma_0 : [0,1] \longrightarrow \mathfrak{g}$ be a ρ_0-geodesic parametrized by constant ρ_0-speed connecting p to q.

Let $\pi \circ \gamma_0 : [0,1] \to (V_1, \|\cdot\|_\infty)$ be the development of γ_0 (recalling Proposition 10.1.11) and $\gamma_\epsilon : [0,1] \to G$ be a lift of $\pi \circ \gamma_0$ starting at p that has the same length, which exists since the map $\pi : (G, \rho_\epsilon) \to (V_1, \|\cdot\|_\infty)$ is a submetry; see Proposition 12.4.3 and Corollary 3.1.25. Namely, if the curve γ_0 has a control $v\colon [0,1] \longrightarrow V_1 \subseteq \mathfrak{g}$ with respect to the group structure $\star_0$, then, the control of γ_ϵ with respect to the group structure $\star_\epsilon$ is a measurable function $u\colon [0,1] \longrightarrow \Delta^{(\epsilon)} \subseteq \mathfrak{g}$ such that $\pi \circ u = v$ and $\|u\|^{(\epsilon)} = \|v\|$. In particular, the curves have the same length in their respective spaces: (G_∞, ρ_0) and (G, ρ_ϵ). Therefore, we have

$$L_{\rho_\epsilon}(\gamma_\epsilon) = L_{\rho_0}(\gamma_0) = \rho_0(p,q) \leq 2, \tag{12.29}$$

and in particular $\|u\|^{(\epsilon)} = \|\pi \circ u\| \leq 2$, almost everywhere. Consequently, from (12.20), we obtain

$$\|u - \pi \circ u\|_E = C'\epsilon, \tag{12.30}$$

for some positive number C' independent of ϵ, where $\|\cdot\|_E$ is a Euclidean norm that we fixed.

We claim that the two curves γ_0 and γ_ϵ stay within a bounded set independent of ϵ. Indeed, regarding the first curve, we obviously have that $\gamma_0(t) \in B_{\rho_0}(1_G, 2)$, since $p, q \in B_{\rho_0}(1_G, 1)$ and γ_0 is a geodesic. Regarding the second curve, we bound:

$$\begin{aligned}
\rho_0(1_G, \gamma_\epsilon(t)) &\overset{(12.22)}{\leq} C(\rho_\epsilon(1_G, \gamma_\epsilon(t)) + \epsilon) \\
&\leq C(\rho_\epsilon(1_G, p) + \rho_\epsilon(p, \gamma_\epsilon(t)) + \epsilon) \\
&\overset{(12.22)}{\leq} C(C\rho_0(1_G, p) + C\epsilon + L_{\rho_\epsilon}(\gamma_\epsilon) + \epsilon) \\
&\overset{(12.29)}{\leq} C(C + C + 2 + 1).
\end{aligned}$$

We proceed similarly to the first part of the proof: We notice that γ_ϵ is an integral curve of the vector field $X^{u,\epsilon}$ as defined in (12.7). While the curve γ_0 is an integral curve of $X^{\pi\circ u,0}$. Again, we have a bound between the two controls: (12.30). As a consequence of Grönwall Lemma and the smoothness of the Lie structures in ϵ, as seen in Remark 12.2.5, we obtain that $\|\gamma_0(1) - \gamma_\epsilon(1)\|_E \leq \hat{C}\epsilon$, for some constant $\hat{C}$ independent on $\epsilon \in (0,1)$. Consequently, since the maps (12.6) are Lipschitz on compact sets, by Lemma 12.2.4, and the curves stay in a bounded set, for some constant L, we obtain

$$\begin{aligned}
\|\gamma_0(1)^{-1} \star_\epsilon \gamma_\epsilon(1)\| &= \|\gamma_0(1)^{-1} \star_\epsilon \gamma_\epsilon(1) - \gamma_0(1)^{-1} \star_\epsilon \gamma_0(1)\| \\
&\leq L(\|\gamma_\epsilon(1) - \gamma_0(1)\|) \leq L(1+\hat{C})\epsilon.
\end{aligned}$$

Then, using the Ball-Box Theorem for Carnot groups (Theorem 11.2.3), we have for some constant $L' > 0$

$$\rho_0(1_G, \gamma_0(1)^{-1} \star_\epsilon \gamma_\epsilon(1)) \le L'\epsilon^{1/s}. \tag{12.31}$$

Consequently, for some constant $C'' > 0$, we bound the distance between the endpoints:

$$\begin{aligned} \rho_\epsilon(\gamma_0(1), \gamma_\epsilon(1)) &\le \rho_\epsilon(1_G, \gamma_0(1)^{-1} \star_\epsilon \gamma_\epsilon(1)) \\ &\overset{(12.22)}{\le} C\rho_0(1_G, \gamma_0(1)^{-1} \star_\epsilon \gamma_\epsilon(1)) + C\epsilon \\ &\overset{(12.31)}{\le} CL'\epsilon^{1/s} + C\epsilon \le C''\epsilon^{1/s}. \end{aligned} \tag{12.32}$$

Finally, we can conclude

$$\begin{aligned} \rho_\epsilon(p, q) &\le \rho_\epsilon(p, \gamma_\epsilon(1)) + \rho_\epsilon(\gamma_\epsilon(1), q) \\ &\le L_{\rho_\epsilon}(\gamma_\epsilon) + \rho_\epsilon(\gamma_\epsilon(1), \gamma_0(1)) \\ &\overset{(12.32)}{\le} L_{\rho_\epsilon}(\gamma_\epsilon) + C''\epsilon^{1/s} \\ &\overset{(12.29)}{=} \rho_0(p, q) + C''\epsilon^{1/s}. \end{aligned}$$

□

12.5 Mitchell Tangent Theorem

In this section, we explain and prove Mitchell's Theorem 12.1.8 for sub-Finsler Lie groups. Later, in Sect. 12.6, we will discuss the general case of manifolds.

12.5.1 Carnot Tangents of Sub-Finsler Lie Groups

Let G be a Lie group with Lie algebra $\mathfrak{g}$. Let $(\Delta, \|\cdot\|)$ be a bracket-generating left-invariant sub-Finsler structure on G, with distance d. Let $V_1 := \Delta_1$. The *osculating Carnot algebra* associated with the polarized Lie algebra $(\mathfrak{g}, V_1)$ is the Lie algebra $\mathfrak{g}_0$ given by the direct-sum decomposition

$$\mathfrak{g}_0 := \bigoplus_{i=1}^{\infty} V_1^{(i)}/V_1^{(i-1)}, \qquad \text{where } V_1^{(i)} := \sum_{k=0}^{i-1} \mathrm{ad}_{V_1}^k V_1 \subseteq \mathfrak{g}, \tag{12.33}$$

endowed with the unique Lie bracket $[\cdot, \cdot]_0$ that has the property that, if $X \in V_1^{(i)}$ and $Y \in V_1^{(j)}$, the bracket is defined, modulo $V_1^{(i+j-1)}$, as

$$\left[X + V_1^{(i-1)}, Y + V_1^{(j-1)}\right]_0 := [X, Y] + V_1^{(i+j-1)}, \tag{12.34}$$

which is well defined because $[V_1^{(i)}, V_1^{(j)}] \subseteq V_1^{(i+j)}$; see Exercise 12.9.17. The Lie algebra $\mathfrak{g}_0$ is also known as *nilpotentization* of $\mathfrak{g}$. But, be aware that if $\mathfrak{g}$ is nilpotent but not Carnot, then $\mathfrak{g}_0$ will not be isomorphic to $\mathfrak{g}$ for any choice of V_1. Recall the example in Exercises 9.5.24.

Consider the simply connected polarized Lie group (G_0, V_1) with Lie algebra $\mathfrak{g}_0$. We stress that G_0 is a stratified group with V_1 as the first stratum. In fact, the space V_1 is both a bracket-generating subspace of $\mathfrak{g}$ and of $\mathfrak{g}_0$. If $V_1 \subseteq \mathfrak{g}$ is equipped with a norm $\|\cdot\|$, then the pair $(V_1, \|\cdot\|)$ can be seen either as a sub-Finsler structure on G_0 and as a sub-Finsler structure on G. We consider G_0 equipped with the sub-Finsler distance d_0 induced by $(V_1, \|\cdot\|)$. The Carnot group (G_0, d_0) is called the *osculating Carnot group* of (G, d).

Mitchell Theorem 12.1.8 will follow from a more quantitative version, following the work of Bellaïche [Bel96, p.69].

Theorem 12.5.1 ([Mit85, Bel96]; see Theorem 12.5.5) *Let (G, d) be a Lie group equipped with a left-invariant sub-Finsler metric with respect to an s-step polarization. Then, its osculating Carnot group (G_0, d_0) is the tangent metric space of (G, d) at each point, and there is a local set-wise identification between G and G_0 such that*

$$\left|d(p, q) - d_0(p, q)\right| = O\left(\max\{d(1, p), d(1, q)\}^{1+1/s}\right), \qquad \text{as } p, q \to 1. \tag{12.35}$$

12.5.2 Structure of Dilated Metrics

There is a natural way to identify the underlying vector space of $\mathfrak{g}$ and the one of its nilpotentization $\mathfrak{g}_0$. This identification depends on the choice of a linear grading $\mathfrak{g} = V_1 \oplus \cdots \oplus V_s$ such that

$$V_j \oplus \sum_{i=0}^{j-2} \mathrm{ad}_{V_1}^i V_1 = \sum_{i=0}^{j-1} \mathrm{ad}_{V_1}^i V_1. \tag{12.36}$$

Linear grading with property (12.36) will be called *adapted linear grading*. Via the exponential maps, we also have local identifications near the identity elements:

$$G \overset{\text{loc}}{\simeq} \mathfrak{g} \simeq \mathfrak{g}_0 \simeq G_0. \tag{12.37}$$

Consider the one-parameter family of dilations $(\delta_\epsilon)_{\epsilon\in\mathbb{R}}$ relative to the linear grading as in (9.8), which is polynomial in ϵ. We modify the Lie bracket of $\mathfrak{g}$ as

$$[X, Y]_\epsilon := \delta_{1/\epsilon}[\delta_\epsilon X, \delta_\epsilon Y], \qquad \forall X, Y \in \mathfrak{g}, \forall\epsilon \in \mathbb{R} \setminus \{0\}. \tag{12.38}$$

The just-defined bracket (12.38) has a different formula than the one defined in the proof of Pansu's Theorem, as in (12.16). In fact, while in (12.17), we performed a large-scale asymptotic expansion, next, in (12.39), we will expand infinitesimally.

The bracket (12.38) satisfies the following formula:

$$\begin{aligned}[v, w]_\epsilon &= \left[\sum_{i=1}^{s} v_i, \sum_{j=1}^{s} w_j,\right]_\epsilon = \sum_{i,j} \delta_{1/\epsilon}\epsilon^{i+j}[v_i, w_j]\\ &= \sum_{i,j}\sum_{k\le i+j} \epsilon^{i+j-k}([v_i, w_j])_k,\end{aligned}$$

where we used the fact that if $v_i \in V_i$ and $w_j \in V_j$, then, thanks to (12.36), $[v_i, w_j] \in [V_1^{(i)}, V_1^{(j)}] \subseteq V_1^{(i+j)} = V_1 \oplus \cdots \oplus V_{i+j}$. It follows that the Lie brackets $[\cdot,\cdot]_\epsilon$ are polynomial and so they extend to $\epsilon = 0$:

$$[X, Y]_0 := \lim_{\epsilon\to 0}[X, Y]_\epsilon, \qquad \forall X, Y \in \mathfrak{g}. \tag{12.39}$$

Observe that the identification (12.37) gives a Lie algebra isomorphism between $(\mathfrak{g}, [\cdot,\cdot]_0)$ and $(\mathfrak{g}_0, [\cdot,\cdot]_0)$.

Let V be the vector space underlying $\mathfrak{g}$. By Lemma 12.2.4, the BCH formula gives a function $(\epsilon, x, y) \mapsto x \star_\epsilon y$ that is analytic on a neighborhood of $\mathbb{R}\times\{0\}\times\{0\}$ in $\mathbb{R} \times V \times V$. The group $(\mathfrak{g}, \star_0)$ is the Carnot group with Lie algebra $\mathfrak{g}_0$; see Proposition 9.4.4.

Let $r_0 > 0$ small enough so that the exponential map restricted to $B_0(0, r_0) \subset \mathfrak{g}$ is an analytic diffeomorphism between $B_0(0, r_0)$ and its image. Observe that for $\epsilon \in [-1, 1]$, the map δ_ϵ maps $B_0(0, r_0)$ into itself. Also, the map $\exp\circ\delta_\epsilon$ is the exponential map for the local Lie group $(\mathfrak{g}, \star_\epsilon)$ and $\star_\epsilon$ is well defined and analytic on $B_0(0, r_0) \times B_0(0, r_0)$.

We modify the metric $d =: d_1$ on G with the *dilated metrics* defined as

$$d_\epsilon(p, q) := \begin{cases} \frac{1}{|\epsilon|} d(\delta_\epsilon p, \delta_\epsilon q) & \forall\epsilon \in \mathbb{R} \setminus \{0\} \\ d_0(p, q) & \text{for } \epsilon = 0 \end{cases}, \qquad \forall p, q \in B_0(0, r_0). \tag{12.40}$$

We point out that each distance d_ϵ is $\star_\epsilon$-left-invariant and, for $\epsilon \neq 0$, it is locally the sub-Finsler distance induced by the polarization $(V_1, \|\cdot\|)$ on $(\mathfrak{g}, \star_\epsilon)$; see Exercise 12.9.8. For each $\epsilon \neq 0$, each map $\delta_\epsilon : (\mathfrak{g}_\epsilon, \star_\epsilon, d_\epsilon) \to (\mathfrak{g}, \star_1, \frac{1}{|\epsilon|}d)$ is a Lie-algebra isomorphism and an isometry.

Lemma 12.5.2 (Equicontinuity of Distances—1) *Fixed a Euclidean distance* $\|\cdot\|_E$ *on* $\mathfrak{g}$, *there exists a constant* C *such that, for every* $p \in B_0(0, r_0)$ *and every* $\epsilon \in [-1, 1]$, *one has*

$$d_\epsilon(1, p) \le C\|p\|_E^{1/s}, \tag{12.41}$$

and

$$d_\epsilon(1, p) \le C d_0(1, p)^{1/s}, \tag{12.42}$$

$$d_0(1, p) \le C d_\epsilon(1, p)^{1/s}. \tag{12.43}$$

Proof The weak Ball-Box Theorem 7.1.21 can be proved uniform in ϵ, as seen in Exercise 12.9.21. Hence, there are constants C_1, C_2 such that $d_\epsilon(1, p) \le C_1\|p\|^{1/s} \le C_2 d_{\epsilon'}(1, p)^{1/s}$, for all $\epsilon, \epsilon' \in [-1, 1]$ and all $p \in B_0, (0, r_0)$. The equations of the lemma are special cases. □

Given an adapted linear grading, as in (12.36), of a polarized Lie algebra $\mathfrak{g}$ and a norm $\|\cdot\|$ on $\mathfrak{g}$, we define the associated *boxes* as

$$\mathtt{Box}(r) := \left\{X_1 + \cdots + X_n \;:\; X_j \in V_j,\; \|X_j\| < r^j,\, j \in \{1, \ldots, s\}\right\}, \quad \forall r \ge 0. \tag{12.44}$$

Theorem 12.5.3 (Ball-Box Theorem for Lie Groups) *Let* G *be a Lie group equipped with a sub-Finsler metric* d_1 *with respect to an* s*-step polarization. With respect to some adapted linear grading, consider the osculating Carnot distance* d_0 *as in* (12.40), *and boxes as in* (12.44). *Then, there are a neighborhood* Ω *of* 1 *in* G *and a constant* $C > 0$ *such that*

$$\frac{1}{C} d_0(1, p) \le d_1(1, p) \le C d_0(1, p), \qquad \forall p \in \Omega. \tag{12.45}$$

In particular, there are constants $C > 0$ *and* $R > 0$ *such that*

$$\exp(\mathtt{Box}(r/C)) \subset B_{d_1}(1, r) \subset \exp(\mathtt{Box}(Cr)), \qquad \forall r \in [0, R]. \tag{12.46}$$

Proof Notice that, by the Ball-Box Theorem 11.2.3 for Carnot groups, the estimates (12.45) and (12.46) are equivalent to

$$\exists C\, \forall \epsilon \in [-1, 1] \qquad B_{d_0}(1, \epsilon/C) \subset B_{d_1}(1, \epsilon) \subset B_{d_0}(1, C\epsilon). \tag{12.47}$$

Let d_ϵ be the dilated metrics as in (12.40).

On the one hand, by (12.43), there is $C > 0$ such that, for all $\epsilon \in [-1, 1]$,

$$B_{d_\epsilon}(1, 1) \subset B_{d_0}(1, C).$$

Since $B_{d_\epsilon}(1,1) = \delta_{1/\epsilon}(B_{d_1}(1,\epsilon))$ and $B_{d_0}(1,C) = \delta_{1/\epsilon}(B_{d_0}(1,C\epsilon))$, we obtain

$$B_{d_1}(1,\epsilon) \subset B_{d_0}(1,C\epsilon),$$

which is one of the two containments in (12.47).

On the other hand, by (12.42), there is $C > 0$ such that, for all $\epsilon \in [-1,1]$,

$$B_{d_0}(1,1/C) \subset B_{d_\epsilon}(1,1).$$

As before, after applying the dilation δ_ϵ, we obtain

$$B_{d_0}(1,\epsilon/C) \subset B_{d_1}(1,\epsilon),$$

which is other containement in (12.47). □

Applying δ_ϵ to (12.45) we immediately obtain another consequence:

Corollary 12.5.4 (Equicontinuity of Distances—2) *Under the same assumptions of Theorem 12.5.3, there exists a constant C such that, for every $p \in B_0(1,r_0)$ and every $\epsilon \in [-1,1]$,*

$$\frac{1}{C} d_0(1,p) \leq d_\epsilon(1,p) \leq C d_0(1,p). \tag{12.48}$$

12.5.2.1 A Geodesically Linearly Connected Subset

For the proof of Mitchell's theorem, we shall need to consider a special set as follows. For the value r_0 fixed above and the constant $C > 1$ from Corollary 12.5.4, let $\Omega \subset \mathfrak{g}$ be the set

$$\Omega := B_0(0,r_1), \qquad \text{with } r_1 := \frac{r_0}{4C^2}. \tag{12.49}$$

We claim that Ω is d_ϵ-geodesically connected within $B_0(0,r_0)$, for all $\epsilon \in [-1,1]$, in the sense that for every pair $p,q \in \Omega$ there is a curve γ_ϵ from p to q valued in $B_0(0,r_0)$ such that $d_\epsilon(p,q) = L_{d_\epsilon}(\gamma_\epsilon)$. For $\epsilon \neq 0$, let $\tilde{\gamma}$ be a geodesic in (G,d) between $\delta_\epsilon(p)$ and $\delta_\epsilon(q)$. Let $\gamma := \delta_\epsilon^{-1} \circ \tilde{\gamma}$, which is a d_ϵ-geodesic between p and q. We claim that the curve γ is well defined and valued in $B_0(0,r_0)$. Indeed, as long as $\gamma(t)$ is in $B_0(0,r_0)$, then in fact it is in $B_0(0,3C^2 r_1)$, because we have

$$\begin{aligned} d_0(1,\gamma(t)) &\overset{(12.48)}{\leq} C d_\epsilon(1,\gamma(t)) \\ &\leq C(d_\epsilon(1,p) + L_{d_\epsilon}(\gamma)) \\ &= C(d_\epsilon(1,p) + d_\epsilon(p,q)) \end{aligned}$$

$$\begin{aligned} &\leq C(2d_\epsilon(1,p)+d_\epsilon(1,q)) \\ &\overset{(12.48)}{\leq} C^2(2d_0(1,p)+d_0(1,q)) \\ &< 3C^2 r_1 < r_0. \end{aligned}$$

Hence, the maximal time $t \in [0, 1]$ for which $\gamma(t) \in B_0(0, r_0)$ is $t = 1$.

For $\epsilon = 0$, let γ_0 be a d_0-geodesic in (G_0, d_0) between p and q. Then, similarly, as above, we bound

$$d_0(1, \gamma_0(t)) < 3r_1 < r_0.$$

Thus, the curve γ_0 is $B_0(0, r_0)$-valued.

12.5.3 Proof of Mitchell Theorem 12.1.8

We are ready to state and prove a quantitative version of Mitchell's Theorem 12.1.8, which also implies Theorem 12.5.1.

Theorem 12.5.5 (Quantitative Mitchell Theorem) *Let G be a Lie group equipped with a sub-Finsler metric d_1 with respect to an s-step polarization. With respect to some adapted linear grading, consider the osculating Carnot distance d_0 and the dilated metrics d_ϵ as in (12.40). Then, there are a neighborhood Ω of 1 in G and a constant $C > 0$ such that*

$$|d_\epsilon(p,q) - d_0(p,q)| \leq C\epsilon^{1/s}, \qquad \forall p, q \in \Omega, \forall \epsilon \in [0, 1]. \tag{12.50}$$

Proof We take as Ω the set constructed in (12.49), which we saw is d_ϵ-geodesically connected within $B_0(0, r_0)$, for all $\epsilon \in [-1, 1]$. Let $p, q \in \Omega$.

We begin with the **first inequality**: $d_0(p,q) \leq d_\epsilon(p,q) + C_1\epsilon^{1/s}$, for some constant $C_1 \geq 0$. Let $\gamma_\epsilon : [0, 1] \longrightarrow B_0(0, r_0)$ be a d_ϵ-geodesic parametrized by constant d_ϵ-speed connecting p to q. Let $u : [0, 1] \longrightarrow V_1$ be the control of γ_ϵ. Let $\gamma_0 : [0, 1] \longrightarrow \mathfrak{g}$ be the d_0-rectifiable curve starting at p with control u, with respect to $\star_0$. Namely, similarly, as in the proof of Pansu's theorem, we have

$$\begin{cases} \dot\gamma_\epsilon(t) = (L^\epsilon_{\gamma_\epsilon(t)})_* u(t) := \frac{\mathrm{d}}{\mathrm{d}s}\gamma_\epsilon(t) \star_\epsilon (su(t))\Big|_{s=0} & , \\ \dot\gamma_0(t) = (L^0_{\gamma_0(t)})_* u(t) := \frac{\mathrm{d}}{\mathrm{d}s}\gamma_0(t) \star_0 (su(t))\Big|_{s=0} & . \end{cases}$$

Notice that

$$\begin{aligned} L_{d_0}(\gamma_0) &= L_{d_\epsilon}(\gamma_\epsilon) \\ &= d_\epsilon(p,q) \end{aligned} \tag{12.51}$$

$$
\begin{aligned}
&\leq \quad d_\epsilon(0, p) + d_\epsilon(0, q) \\
&\overset{(12.48)}{\leq} \quad C(d_0(0, p) + d_0(0, q)) \\
&\leq \quad 2Cr_1
\end{aligned}
$$

and thus $d_0(0, \gamma_0(t)) \leq d_0(0, p) + L_{d_0}(\gamma_0) \leq (2C + 1)r_1 < r_0$, that is, $\gamma_0([0, 1]) \subset B_0(0, r_0)$.

By Remark 12.2.5, we obtain $\|\gamma_0(1) - \gamma_\epsilon(1)\|_E \leq C'\epsilon$. Consequently, by the Ball-Box Theorem for Carnot groups (Theorem 11.2.3) we have for some constant $C'' > 0$

$$
d_0(\gamma_0(1), \gamma_\epsilon(1)) \leq C''\epsilon^{1/s}. \tag{12.52}
$$

Finally, we obtain the first bound:

$$
\begin{aligned}
d_0(p, q) \quad &\leq \quad d_0(p, \gamma_0(1)) + d_0(\gamma_0(1), \gamma_\epsilon(1)) \\
&\overset{(12.52)}{\leq} \quad L_{d_0}(\gamma_0) + C''\epsilon^{1/s} \\
&\overset{(12.51)}{=} \quad d_\epsilon(p, q) + C''\epsilon^{1/s}.
\end{aligned}
$$

We consider the **second inequality**: $d_\epsilon(p, q) \leq d_0(p, q) + C_2\epsilon^{1/s}$, for some $C_2 \geq 0$. Let $\gamma_0 : [0, 1] \longrightarrow B_0(0, r_0)$ be a d_0-geodesic parametrized by constant d_0-speed connecting p to q. Let $u : [0, 1] \longrightarrow V_1$ be the control of γ_0, with respect to $\star_0$. Let $\gamma_\epsilon : [0, 1] \longrightarrow \mathfrak{g}$ be the d_ϵ-rectifiable curve starting at p with control u with respect to $\star_\epsilon$. The error between $\gamma_\epsilon(1)$ and q is estimated as

$$
\begin{aligned}
d_\epsilon(\gamma_\epsilon(1), q) \quad &= \quad d_\epsilon(1, (-\gamma_\epsilon(1)) \star_\epsilon q) \\
&\overset{(12.41)}{\leq} \quad C\|(-\gamma_\epsilon(1)) \star_\epsilon q\|_E^{1/s} \\
&= \quad C\|(-\gamma_\epsilon(1)) \star_\epsilon q - (-q) \star_\epsilon q\|_E^{1/s} \\
&\leq \quad C'\| - \gamma_\epsilon(1) + q\|_E^{1/s} \\
&\overset{\text{Rem. 12.2.5}}{\leq} \quad C''\epsilon^{1/s}.
\end{aligned}
$$

Notice also that $d_\epsilon(p, \gamma_\epsilon(1)) \leq L_{d_\epsilon}(\gamma_\epsilon) \leq L_{d_0}(\gamma_0) = d_0(p, q)$. Therefore, we obtain

$$
\begin{aligned}
d_\epsilon(p, q) &\leq d_\epsilon(p, \gamma_\epsilon(1)) + d_\epsilon(\gamma_\epsilon(1), q) \\
&\leq d_0(p, q) + C''\epsilon^{1/s}.
\end{aligned}
$$

□

12.6 Mitchell's Theorem for CC Spaces

Mitchell's result from the previous section holds more generally. In fact, Carnot groups are the tangent metric spaces of sub-Finsler manifolds at every regular point. Such a result, originally attributed to Mitchell, is quite technical and involved; a complete treatment is in [Bel96, Jea14]. In this section, we describe the Carnot group that appears as tangent at every such point.

12.6.1 Osculating Carnot Group, a.k.a. Nilpotentization

Let M be a manifold and let Δ be a bracket-generating distribution that is equiregular in the sense of Sect. 4.1.5. Denote by s the step of Δ. Let

$$\Delta = \Delta^{[1]} \subset \Delta^{[2]} \subset \ldots \subset \Delta^{[s]} = TM$$

be the flag of sub-bundles of TM as in Definition 4.1.13. The simple but crucial fact is that

$$[\Delta^{[k]}, \Delta^{[l]}] \subseteq \Delta^{[k+l]}. \tag{12.53}$$

Equation (12.53) is obvious for $k = 1$ and can be proved by induction using linearity and Jacobi identity:

$$\begin{aligned}
[\Delta^{[k+1]}, \Delta^{[l]}] &= \left[\Delta^{[k]} + [\Delta, \Delta^{[k]}], \Delta^{[l]}\right] \\
&= [\Delta^{[k]}, \Delta^{[l]}] + \left[[\Delta, \Delta^{[k]}], \Delta^{[l]}\right] \\
&\subseteq \Delta^{[k+l]} + \left[[\Delta^{[k]}, \Delta^{[l]}], \Delta\right] + \left[[\Delta^{[l]}, \Delta], \Delta^{[k]}\right] \\
&\subseteq \Delta^{[k+l]} + [\Delta^{[k+l]}, \Delta] + [\Delta^{[l+1]}, \Delta^{[k]}] \\
&\subseteq \Delta^{[k+l]} + \Delta^{[k+l+1]} + \Delta^{[k+l+1]} \\
&\subseteq \Delta^{[k+l+1]},
\end{aligned}$$

where we assumed that for given k it has been proved for all l and we showed it for the value $k + 1$ and every l. This last argument is similar to the one for Exercises 9.5.20 and 9.5.21.

Define $H_1 := \Delta$ and $H_j := \Delta^{[j]}/\Delta^{[j-1]}$, for $j = 2, \ldots, n$. Still, the space H_j is a bundle over M, but not a sub-bundle of the tangent bundle TM. We obviously have the isomorphism

$$TM \simeq H_1 \oplus H_2 \oplus \ldots \oplus H_s.$$

With the aim of defining a Lie group that we will denote by $T_p(M, \Delta)$, for $p \in M$, we set

$$\mathrm{Lie}(T_p(M, \Delta)) := \bigoplus_{j=1}^{s} \Delta^{[j]}(p) \big/ \Delta^{[j-1]}(p). \tag{12.54}$$

We equip this set with the Lie bracket that has the property that for all $x \in \Delta^{[i]}(p)$ and $y \in \Delta^{[j]}(p)$,

$$\left[x + \Delta^{[i-1]}(p), y + \Delta^{[j-1]}(p)\right] := [X, Y]_p + \Delta^{[i+j-1]}(p), \tag{12.55}$$

where $X \in \Gamma(\Delta^{[i]})$ with $X(p) = x$, and $Y \in \Gamma(\Delta^{[j]})$ with $Y(p) = y$. We show that the bracket (12.55) is well defined: by symmetry, we just show the independence on the representative of x. Let $\tilde{X} \in \Gamma(\Delta^{[i]})$ with $\tilde{X}(p) \in x + \Delta^{[i-1]}(p)$. Write $X - \tilde{X} = \sum_{\ell=1}^{k} a^\ell f_\ell$ with $f_1, \ldots, f_k$ a frame of $\Delta^{[i]}$ and $a^\ell(p) = 0$ when $f_\ell \notin \Gamma(\Delta^{[i-1]})$. Then

$$\begin{aligned}[X, Y] - [\tilde{X}, Y] = [X - \tilde{X}, Y] &= \sum_{\ell=1}^{k} [a^\ell f_\ell, Y] \\ &\overset{\text{Ex. 3.4.48}}{=} \sum_{\ell=1}^{k} a^\ell [f_\ell, Y] - \sum_{\ell=1}^{k} (Y a^\ell) f_\ell,\end{aligned}$$

which, when evaluated in p, gives

$$[X, Y]_p - [\tilde{X}, Y]_p \in [\Gamma(\Delta_p^{[i-1]}), \Gamma(\Delta^{[j]})] + \Delta^{[i]}(p) \subset \Delta^{[i+j-1]}(p).$$

Thus, the Lie bracket (12.55) is well defined.

Remark 12.6.1 The Lie algebra $\mathrm{Lie}(T_p(M, \Delta))$ is stratified by

$$\left(\Delta^{[j]}(p) \big/ \Delta^{[j-1]}(p) \right)_{j \in \{1, \ldots, s\}}.$$

The first layer is Δ_p. There is a Carnot group with this stratified algebra by Proposition 9.4.3.

Definition 12.6.2 (Osculating Carnot Group) When an equiregular polarized manifold (M, Δ) is equipped with a sub-Finsler norm $\| \cdot \|$, we consider the Carnot group $(T_p(M, \Delta), \Delta_p, \| \cdot \|_p)$ whose Lie algebra is $\mathrm{Lie}(T_p(M, \Delta))$ as in (12.54), and call it the *osculating Carnot group* of $(M, \Delta, \| \cdot \|)$ at p, and it is considered equipped with its Carnot-Carathéodory metric d_0.

12.6.2 Tangent Metric Spaces to Equiregular CC Spaces

Mitchell's theorem for equiregular sub-Finsler manifolds is then the following.

Theorem 12.6.3 (Mitchell) *Let $(M, \Delta, \|\cdot\|)$ be a sub-Finsler manifold, with Δ equiregular. Let $p \in M$. Then, the tangent metric space of (M, d) at p is isometric to the osculating Carnot group $(T_p(M, \Delta), d_0)$.*

We provided a full proof of this theorem for Lie groups; see Theorems 12.5.1 and 12.5.5. The general strategy is very similar. Again, one can see this result as a convergence of a one-parameter family of CC structures. We state the key result referring to [Jea14] for a proof. With the terminology of Bellaïche and Jean, the next proposition says that exponential local parametrizations are examples of privileged coordinates.

Proposition 12.6.4 *Let M be a manifold equipped with an equiregular bracket-generating distribution Δ. Fix $p \in M$. Let $X_1, \dots, X_n$ be a local frame of TM around p adapted to the flag $\Delta^{[1]}, \dots, \Delta^{[k]}, \dots$. Let $\Phi : U \subset \mathbb{R}^n \to M$ be the corresponding exponential local parametrization around p. For each $j \in \{1, \dots, n\}$ and $\epsilon \in (0, 1)$, set*

$$X_j^{(\epsilon)} := \epsilon^{\deg(j)} (\Phi \circ \delta_\epsilon)^* X_j,$$

where $\deg(j)$ is such that $X_j \in \Gamma(\Delta^{\deg(j)}) \setminus \Gamma(\Delta^{\deg(j)-1})$ and $\delta_\epsilon : \mathbb{R}^n \to \mathbb{R}^n$ is the map

$$\delta_\epsilon(x_1, \dots, x_n) := (\epsilon x_1, \dots, \epsilon^{\deg(j)} x_j, \dots).$$

Then, in the C^∞-topology, each vector field $X_j^{(\epsilon)}$ converges to $X_j^{(0)} \in \mathrm{Vec}(\mathbb{R}^n)$ such that $X_1^{(0)}, \dots, X_n^{(0)}$ form a frame of $T(\mathbb{R}^n)$ and generates a Lie algebra that is isomorphic to $\mathrm{Lie}(T_p(M, \Delta))$.

Remark 12.6.5 Differently from the Riemannian case, it is not true that sub-Riemannian manifolds are locally bi-Lipschitz equivalent to their tangent metric spaces, even in the case when the tangent metric space is the same Carnot group at every point; see [LOW14]. It is, however, true for contact manifolds because of the Darboux Theorem. See also [LY23] for a measure-theoretic parametrization.

12.7 A General Result on Convergence of CC Structures

Pansu's and Mitchell's theorems, together with the other examples that we have discussed, follow from a more general principle. In brief, when varying CC-bundle structures converge, the distances converge. The following theorem has been obtained in [ALN23, Appendix C].

Theorem 12.7.1 *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* (4.26)*. 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* $\lambda \to \lambda_0$*.*

The proof of the previous theorem stands on the following crucial lemma, which should be compared with (12.22) and (12.42).

Lemma 12.7.2 (Equicontinuity of the Distances) *In the same assumptions of Theorem 12.7.1, 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 a homeomorphism* β *of* $[0, +\infty)$ *such that*

$$d_\lambda(p, q) \le \beta(\rho(p, q)), \qquad \text{for all } p, q \in K \text{ and } \lambda \in I_{\lambda_0}.$$

Sketch of the Proof of Lemma 12.7.2 One begins by showing that for all $x \in M$ and all $\epsilon > 0$ there exist $\delta > 0$ and a neighborhood I of λ_0 such that

$$B_\rho(x, \delta) \subseteq B_{d_\lambda}(x, \epsilon), \qquad \forall\lambda \in I. \tag{12.56}$$

As in the proof of Chow's Theorem 4.2.1, we consider maps

$$F_\lambda : (t_1, \dots, t_N) \mapsto \Phi^{t_N}_{X^\lambda_N} \circ \cdots \circ \Phi^{t_1}_{X^\lambda_1}(x)$$

obtained as composition of flows starting at x of vector fields in f_λ. Since f_{λ_0} is bracket generating by assumption, from the proof of Chow's Theorem we have that $F_{\lambda_0}([0, \epsilon/N]^N)$ is a neighborhood of x. By smoothness of F_λ in λ, the same is true for λ in some neighborhood of λ_0. One infers (12.56) and concludes by compactness arguments. □

12.8 Finitely-Generated Groups of Polynomial Growth

Carnot-Carathéodory spaces appear when studying the asymptotic growth of finitely generated nilpotent groups, which essentially are the finitely generated groups of polynomial growth; see the theorems below.

A group Γ is *finitely generated* if there is a finite subset $S \subseteq \Gamma$, called *finite generating set* such that

$$\Gamma = \bigcup_{n\in\mathbb{N}} (S \cap S^{-1})^n.$$

In such a case, fixed a finite generating set $S \subset \Gamma$, then the distance function d_S on Γ that is defined by

$$\bar{B}_{d_S}(x, n) = x(S \cap S^{-1})^n, \qquad \forall x \in \Gamma, \forall n \in \mathbb{N}$$

is called the *word metric* with respect to S.

In addition, we say that a finitely generated group Γ has *polynomial growth* if for some (and consequently, for every) finite generating set S one has

$$\text{Card}\left((S \cap S^{-1})^n\right) \leq c_1 n^{c_2}, \qquad \forall n \in \mathbb{N},$$

for some constants $c_1, c_2 \in \mathbb{R}_+$, possibly depending on S. Here, the operator Card denotes the cardinality of sets.

Sub-Finsler geometries will appear in the following two results:

Theorem 12.8.1 (Wolf [Wol68], Bass [Bas72], Pansu, [Pan83]) *Let Γ be a nilpotent finitely generated group, with a finite generating set $S \subset \Gamma$ and word distance d_S.*

(12.8.1).*i. The asymptotic cone of (Γ, d_S) exists and is a sub-Finsler Carnot group, whose Hausdorff dimension is some $Q \in \mathbb{N}$ and the Q-Hausdorff measure of the unit ball is some $v \in \mathbb{R}_+$;*

(12.8.1).*ii. The group Γ has polynomial growth, and in fact one has*

$$\frac{\text{Card}\left((S \cap S^{-1})^n\right)}{v n^Q} \longrightarrow 1, \qquad \text{as } n \to \infty. \tag{12.57}$$

The following theorem is a reverse.

Theorem 12.8.2 (Gromov's Polynomial Growth, [Gro81]) *Let Γ be a finitely generated group. If Γ has polynomial growth, then there is a subgroup $\Gamma_{\text{nil}} \subset \Gamma$ that is finitely generated and nilpotent, and the quotient $\Gamma / \Gamma_{\text{nil}}$ is finite. In particular, the asymptotic cone of (Γ, d_S) is a sub-Finsler Carnot group for every finite generating set $S \subset \Gamma$ and word distance d_S.*

The two above theorems are summarized by the following sentence: A finitely generated group has polynomial growth if and only if it is virtually nilpotent. Actually, these are precisely the finitely generated groups that have Lie groups as asymptotic cones.

12.8.1 A Few Comments on Wolf-Bass-Pansu's Theorem

For the proof of Theorem 12.8.1, one can construct the asymptotic Carnot group explicitly. Starting with a nilpotent finitely generated group Γ, equipped with a word distance d_S, we describe a collection of metric groups that are quasi-isometric to (Γ, d_S), and actually, they are asymptotic to each other in the sense that, after a coarse identification, the ratio of the distances goes to 1 as the points go to infinity. One of these metric spaces will be a nilpotent simply connected sub-Finsler Lie group. Hence, we will conclude, with the help of Pansu's Theorem 12.4.1, that this Lie group and, hence, Γ are asymptotic to the associated Carnot group. Be aware that this final Carnot group may not be quasi-isometric to Γ.

After Γ, the next group in the collection that we consider is Γ modulo its torsion elements $\mathrm{Tor}(\Gamma)$, as we discussed in Sect. 9.4.6. One can prove that the set $\mathrm{Tor}(\Gamma)$ is a normal subgroup that is finite; see [Mac88, Theorem 9.17]. Let $\pi : \Gamma \to \Gamma/\mathrm{Tor}(\Gamma)$ be the quotient projection. The group $\pi(\Gamma) = \Gamma/\mathrm{Tor}(\Gamma)$ is nilpotent, torsion-free, and finitely generated by $\pi(S)$. Via the map π the metric spaces (Γ, d_S) and $(\pi(\Gamma), d_{\pi(S)})$ are $(1, C)$-quasi isometric.

By a fundamental result by Malcev [Mal51], called *Malcev completion*, there is a (unique) nilpotent simply connected Lie group G that admits a discrete cocompact subgroup $\tilde{\Gamma} \subset G$ isomorphic to $\Gamma/\mathrm{Tor}(\Gamma)$; see [Rag72, Theorem 2.18]. Here, cocompact means that the quotient $G/\tilde{\Gamma}$ is compact. Let us denote by $\phi : \Gamma/\mathrm{Tor}(\Gamma) \to \tilde{\Gamma}$ an isomorphism.

We equip the Lie group G with the following metric: Let $\tilde{S} \subset G$ be the finite set $\tilde{S} := \phi(\pi(S))$. Then, the *Stoll distance* between $g_1, g_2 \in G$ (*relative to* $\tilde{S}$) is

$$d_{\mathrm{Stoll}}(g_1, g_2) := \inf\Big\{ |t_1| + \ldots + |t_n| \ : \\ g_1^{-1} g_2 = s_1^{t_1} \cdot \ldots \cdot s_n^{t_n}, n \in \mathbb{N}, s_1, \ldots, s_n \in \tilde{S}, t_1, \ldots, t_n \in \mathbb{R} \Big\}.$$

This distance function on G is geodesic and left-invariant. Hence, it is a sub-Finsler metric by Berestovski's Theorem 7.4.1; see Exercise 12.9.28. The Stoll distance is named after M. Stoll since he proved that if G is two-step nilpotent, then on $\tilde{\Gamma}$, the difference between d_{Stoll} and the word metric $d_{\tilde{S}}$ is bounded; see [Sto98]. In groups of step at least 3, the same result may not be true; for a counterexample in Cartan's group, see [Bod23]. However, on $\tilde{\Gamma}$ the distances d_{Stoll} and $d_{\tilde{S}}$ are asymptotic, and also quantitatively, as shown in [BL13, Gia17]:

$$|d_{\tilde{S}}(p_1, p_2) - d_{\mathrm{Stoll}}(p_1, p_2)| \le C \left(d_{\tilde{S}}(p_1, p_2)\right)^{1-\alpha}, \qquad \forall p_1, p_2 \in \tilde{\Gamma},$$

for some $C = C(\tilde{S})$ and $\alpha = \alpha(G) > 0$.

After all the above steps, we obtained a quasi-isometry

$$\varphi : (\Gamma, d_S) \longrightarrow (G, d_{\mathrm{Stoll}}),$$

which is asymptotic:

$$|d_S(p_1, p_2) - d_{\text{Stoll}}(\varphi(p_1), \varphi(p_2))| \leq C\,(d_S(p_1, p_2))^{1-\alpha}, \quad \forall p_1, p_2 \in \Gamma, \tag{12.58}$$

for some $C = C(S, \phi)$ and $\alpha = \alpha(G) > 0$.

Let F be a compact fundamental domain for the action of Γ on G. Let vol be the Haar measure on G such that

$$\text{vol}(F) = \text{vol}(G/H) = 1.$$

With this choice of normalization for the Haar measure, we have

$$\text{vol}(S^n F) = \text{Card}(S^n), \qquad \forall n \in \mathbb{N}.$$

Let d_∞ be the Pansu limit metric associated with d_{Stoll}, as in Sect. 12.4.1. For the next observation, we use that F is bounded, the asymptotic property (12.58), and Theorem 12.4.1 applied to d_{Stoll}. There are constants $C, \alpha > 0$ such that

$$\begin{cases} S^n F \subseteq B_{d_\infty}(1, n + Cn^{1-\alpha}), \\ B_{d_\infty}(1, n) \subseteq S^{n+Cn^{1-\alpha}} F, \end{cases} \qquad \forall n \geq 1.$$

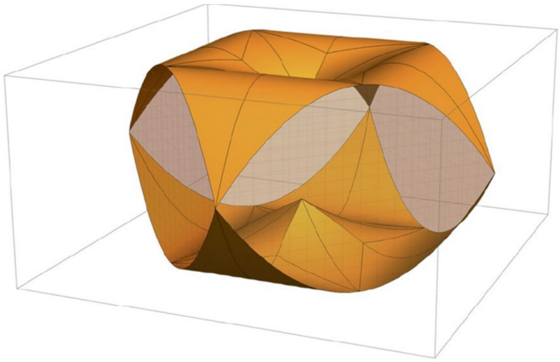
Recall the scaling property of d_∞ from Proposition 11.2.4. Consequently, we conclude the volume expansion:

$$\text{Card}(S^n) = \text{vol}(B_{d_\infty}(1, 1))n^Q + O(n^{Q-\alpha}), \qquad \text{as } n \to \infty.$$

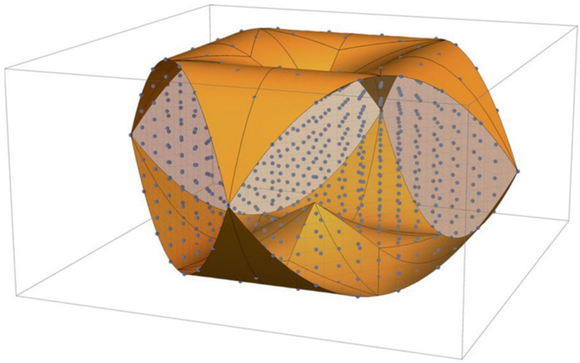
We actually obtained a more quantitative expansion than (12.57).

Example 12.8.3 (Standard Generating Set in the Heisenberg Group) The Heisenberg group in the standard exponential coordinates, as in Sect. 2.3, admits many discrete cocompact subgroups. In these coordinates, a standard one is $\Gamma := \mathbb{Z} \times \mathbb{Z} \times \frac{1}{2}\mathbb{Z}$, which is a nonabelian subgroup with respect to the product (2.9). The subset $S := \{(\pm 1, 0, 0), (0, \pm 1, 0)\} \subseteq \Gamma$ is a finite generating subset of Γ. Hence, it induces a word metric d_S and (Γ, d_S) has polynomial growth. From what we have seen just above, the asymptotic geometry of (Γ, d_S) is equal to the sub-Finsler geometry of the Heisenberg Lie group equipped with the ℓ_1-Carnot metric. Such a metric has the standard horizontal bundle with the norm given by the convex hull of $\log(S)$. For a visual representation of the ℓ_1-Carnot metric ball, we refer to Fig. 12.3.

Fig. 12.3 Balls for the case of the standard generating set in the Heisenberg group: $S := \{(\pm 1, 0, 0), (0, \pm 1, 0)\}$. The large balls with respect to the word distance, when rescaled by the intrinsic dilations, converge to the Carnot-Carathéodory ball. This geometry is called ℓ_1-*sub-Finsler*. (**a**) Unit ball for the Pansu limit. (**b**) S^7 rescaled by δ_7^{-1}



(a) Unit ball for the Pansu limit



(b) S^7 rescaled by δ_7^{-1}

12.8.2 A Comment on Gromov's Theorem

The full proof of Theorem 12.8.2 requires some algebra that goes beyond the tasks of this book. However, we can spend some words explaining why there is a Carnot group among the possible blowdown spaces of (Γ, d_S).

Because Γ is assumed to have polynomial growth, there is $K \in \mathbb{R}_+$ such that the metric space (Γ, d_S) is doubling with constant K. Consequently, for each $\epsilon > 0$ the metric space $(\Gamma, \epsilon d_S)$ is K-doubling, as well.

Let $\epsilon_j \searrow 0$ be an infinitesimal sequence. We consider the Gromov-Hausdorff convergence of boundedly compact pointed metric spaces. By a Gromov's argument, analogs to Ascoli-Arzelá results, up to passing to a subsequence, we may assume that the sequence of uniformly doubling pointed metric spaces $(\Gamma, \epsilon_j d_S, 1)$ converges to some pointed complete metric space (X, d_X, x_0), which still is K-doubling. In particular, the space X is locally compact and has finite topological dimension.

Here are some more properties of (X, d_X). As (Γ, d_S) is $(1, 1)$-quasi-geodesic (recall Definition 6.3.7), then, the metric space $(\Gamma, \epsilon d_S)$ is $(1, \epsilon)$-quasi-geodesic.

Hence, the limit (X, d_X) is geodesic. In particular, the space X is connected and locally connected.

Since each $(\Gamma, \epsilon d_S)$ is isometrically homogeneous, then so is (X, d_X). We checked all the conditions to apply Theorem 6.2.10, which was a consequence of the theory of Gleason-Montgomery-Yamabe-Zippin. Consequently, the metric space (X, d_X) has the structure of a Lie coset space. By Berestovskii's Theorem 7.4.1, the metric space (X, d_X) is a sub-Finsler manifold. Therefore, by Mitchell's Theorem 12.1.8, every tangent metric space of (X, d_X) is a sub-Finsler Carnot group G.

As a general fact, the blowup of a blowdown is a blowdown. In fact, consider $\lambda_j \to \infty$ diverging so slow that $\epsilon_j := \lambda_j \epsilon_j \searrow 0$. Then, the sequence $(\Gamma, \tilde{\epsilon}_j d_S, 1)$ converges to a tangent metric space of (X, d_X, x_0), hence, to G. Thus, we proved that one of the blowdowns of (Γ, d_S) is the sub-Finsler Carnot group G.

In Gromov's proof for Theorem 12.8.2, one considers various possible actions of Γ on the Carnot group G. We refer to [Gro81, Section 8] for the full argument. We also point out a book by Mann [Man12] and an article by Losert [Los87].

12.9 Exercises

Exercise 12.9.1 Let d_1, d_2 be two boundedly compact left-invariant distances on a Lie group G inducing the manifold topology. Then the increasing function $\xi : (0, \infty) \to (0, \infty)$ defined by

$$\xi(r) := \mathrm{diam}_{d_1}\left(\overline{B}_{d_2}(1_G, r)\right)$$

is such that $\xi(r) \to 0$, as $r \to 0$, and $d_1(p, q) \le \xi(d_2(p, q))$.

Exercise 12.9.2 Let $X^{u,\lambda}(t, p)$ as in (12.7). Let $K, \Lambda', C_1, C_2, \ell$ as in Remark 12.2.5. There exists $C > 0$ such that, for all $\epsilon > 0$, $u, v \in L^\infty([0, 1]; V)$ with $\|u - v\|_\infty < C_1\epsilon$, $\|u\|_\infty, \|v\|_\infty < \ell$, all $a, b \in \Lambda'$ with $|a - b| < C_2\epsilon$:

(12.9.2).i. $\|X^{u,a}(t, p) - X^{v,b}(t, p')\| \le C\|u\|_\infty + C\epsilon + C\|u\|_\infty\|p - p'\|$, for all $p, p' \in K$;

(12.9.2).ii. if α and $\beta : [0, 1] \to K$ are absolutely continuous integral curves of $X^{u,a}$ and $Y^{v,b}$, respectively, with $\alpha(0) = \beta(0)$, then $\|\alpha(1) - \beta(1)\| \le C\|u\|_\infty\epsilon$.

Exercise 12.9.3 Let $X, Y, Z := [X, Y]$ be the standard basis for the Lie algebra of the Heisenberg group H. For $n \in \mathbb{N}$, let d_n be the Riemannian distance for which $X, Y, \frac{1}{n}Z$ form an orthonormal frame. The metric spaces $(H, \frac{1}{n}d_1)$ and (H, d_n) are isometric.

Exercise 12.9.4 Let M be a manifold and $\Delta \subset TM$ 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 $g_n|_\Delta = g_1|_\Delta$, for all $n \in \mathbb{N}$,

and, for all $X \notin \Delta$, $g_n(X, X) \to +\infty$, as $n \to \infty$. For all $p, q \in M$, we have $\lim_{n\to\infty} d_{g_n}(p, q) = d_{cc}(p, q)$ where d_{cc} is the sub-Riemannian distance associated with $(\Delta, g_1|_\Delta)$.
Hint. See Example 12.3.5.

Exercise 12.9.5 If (M, d) is a CC space and $\lambda > 0$, then $(M, \lambda d)$ is a CC space.

Exercise 12.9.6 Every sub-Riemannian Carnot group is the limit of some (left-invariant) Riemannian metrics on the same Lie group.

Exercise 12.9.7 Let G be an s-step stratified group, with a basis $X_1, \dots, X_n$ adapted to the stratification. For each $\lambda > 0$, consider the dilations δ_λ and the Riemannian distance d_λ for which $X_1, \dots, \frac{\lambda^{\deg(X_j)}}{\lambda} X_j, \dots, \lambda^{s-1} X_n$ are orthonormal. Then, the distance λd_1 is associated with the Riemannian metric that makes $\frac{1}{\lambda} X_1, \dots, \frac{1}{\lambda} X_n$ orthonormal. Every map δ_λ sends each vector field $\frac{1}{\lambda} X_j$ to the vectors field $\frac{1}{\lambda}\lambda^{\deg(X_j)} X_j$. For all $\lambda > 0$ the metric space $(G, \lambda d_1)$ is isometric to (G, d_λ) via the map δ_λ.

Exercise 12.9.8 Let $F : M_1 \to M_2$ be a diffeomorphism between connected manifolds. Let $(\Delta, \|\cdot\|)$ be a sub-Finsler structure on M_1. Then the only metric on M_2 that makes F an isometry comes from the sub-Finsler structure where $\|v\| := \left\|dF^{-1} v\right\|$ for all $v \in \Delta_p := (dF)\Delta_{F^{-1}(p)}$ and all $p \in M_2$.

Exercise 12.9.9 In each s-step nilpotent simply connected Lie group, the Dynkin product for the BCH formula (5.24), has the form:

$$x_1 \star x_2 = \log(\exp(x_1) \cdot \exp(x_2)) = x_1 + x_2 + \sum_{k=2}^{s} \sum_{q \in \{1,2\}^k} b_{k,q} [x_{q_1}, \dots, x_{q_k}], \tag{12.59}$$

where $b_{k,q}$ are universal real constants, as in Definition 5.7.3, and here $[x_{q_1}, \dots, x_{q_k}]$ denotes the left-iterated bracket: $[x_{q_1}, \dots, x_{q_k}] := [x_{q_1}, [\dots, [x_{q_{k-1}}, x_{q_k}] \dots]]$.

Exercise 12.9.10 Suppose G is a nilpotent simply connected Lie group. Choose a basis $(e_1, \dots, e_n)$ of $\mathfrak{g}$ that is adapted to a compatible linear grading $\mathfrak{g} = \oplus_i V_i$. Given an index $i \in \{1, \dots, n\}$, let d_i be the degree of e_i, namely the integer such that $e_i \in V_{d_i}$. For every multi-index $\alpha = (\alpha_1, \dots, \alpha_n) \in \mathbb{N}^n$, we set $x^\alpha := x_1^{\alpha_1} \cdot \dots \cdot x_n^{\alpha_n}$ and $d_\alpha := d_1\alpha_1 + \dots + d_n\alpha_n$. Let $\star$ be the Dynkin product on $\mathfrak{g}$. Then, for some constants $C_{\alpha,\beta} \in \mathbb{R}$, we have

$$(x \star y)_i = x_i + y_i + \sum_{\{\alpha,\beta \mid d_\alpha \geq 1, d_\beta \geq 1, d_\alpha + d_\beta \leq d_i\}} C_{\alpha,\beta} x^\alpha y^\beta.$$

For the Dynkin product $\star_0$ associated with the asymptotic Lie bracket $[\![\cdot,\cdot]\!]_\infty$ of the associated Carnot algebra as in (12.17), we have

$$(x \star_0 y)_i = x_i + y_i + \sum_{\{\alpha,\beta \mid d_\alpha \geq 1, d_\beta \geq 1, d_\alpha + d_\beta = d_i\}} C_{\alpha,\beta} x^\alpha y^\beta,$$

where we have chopped the terms with $d_\alpha + d_\beta < d_i$.

Exercise 12.9.11 In the construction of the family of products on a nilpotent Lie algebra $\mathfrak{g}$ as in the proof of Pansu's Theorem, for $\epsilon > 0$, we denote by $\star_\epsilon$ the Dynkin product associated with the contracted brackets $[\cdot,\cdot]_\epsilon$ as in (12.16). We have

$$g_1 \star_\epsilon g_2 = \delta_\epsilon(\delta_\epsilon^{-1}(g_1) \star \delta_\epsilon^{-1}(g_2)), \qquad \forall g_1, g_2 \in \mathfrak{g}.$$

Exercise 12.9.12 Let G be a nilpotent simply connected sub-Finsler Lie group. Let d_∞ be its Pansu limit metric and ρ_ϵ the approximating distances, as in (12.18). Let $C > 0$ be the Guivarc'h constant from (12.22). Then, for all $R > 0$ and all $\epsilon \in (0, 1)$, we have $B_{\rho_\epsilon}(1, R) \subseteq B_{d_\infty}(1, C(R + 1))$.
Solution. For $p \in B_{\rho_\epsilon}(0, R)$, one has

$$d_\infty(1, p) = \epsilon d_\infty(1, \delta_\epsilon^{-1}(p)) \leq C\epsilon d(1, \delta_\epsilon^{-1}(p)) + C\epsilon \leq CR + C\epsilon.$$

Exercise 12.9.13 Theorem 12.4.5 implies Theorem 12.4.1, and, in particular,

$$\left| d(1, g) - d_\infty(1, g) \right| = O\left(d(1, g)^{1-1/s}\right), \qquad \text{as } g \to \infty.$$

Hint. As $d_\infty(1, g) \to \infty$, take $\epsilon := (d_\infty(1, g))^{-1}$ and $p := \delta_\epsilon(g)$.

Exercise 12.9.14 (Pansu Comparison Theorem) Let (G, d) be a nilpotent simply connected sub-Finsler Lie group. Let d_∞ be the Pansu limit metric on G. Then $\left|\frac{d(1,g)}{d_\infty(1,g)}\right| \to 1$, as $g \to \infty$.
Hint. Use Theorem 12.4.1

Exercise 12.9.15 Using Exercise 12.9.2 instead of Remark 12.2.5, one can upgrade Eq. (12.23) of the Quantitative Pansu Theorem to

$$-C\epsilon^{1/s}\rho_0(p, q)^{1/s} \leq \rho_0(p, q) - \rho_\epsilon(p, q) \leq C\epsilon^{1/s}\rho_\epsilon(p, q)^{1/s}.$$

Exercise 12.9.16 Let N be a simply connected nilpotent Lie group equipped with a Riemannian (or subFinsler) left-invariant metric. If N admits a quasi-isometric embedding into some Euclidean space, then N is a commutative group.
Hint. Pass to asymptotic cones and use Theorem 11.3.8.

Exercise 12.9.17 Given a polarized Lie group (G, V_1), with the notation from (12.33), we have $[V_1^{(i)}, V_1^{(j)}] \subseteq V_1^{(i+j)}$. Hence, the Lie bracket $[\cdot,\cdot]_0$, as in (12.34), is well defined on $\mathfrak{g}_0$ and (12.33) gives a stratification for $\mathfrak{g}_0$.

Exercise 12.9.18 Let (G, Δ) be a polarized Lie group. Choose a basis $(e_1, \ldots, e_n)$ of $\mathfrak{g}$ that is adapted to an adapted linear grading $\mathfrak{g} = \oplus_i V_i$ as in (12.36). Given an index $i \in \{1, \ldots, n\}$, let d_i be the degree of e_i, namely the integer such that $e_i \in V_{d_i}$. For every multi-index $\alpha = (\alpha_1, \ldots, \alpha_n) \in \mathbb{N}^n$, we set $x^\alpha := x_1^{\alpha_1} \cdot \ldots \cdot x_n^{\alpha_n}$ and $d_\alpha := d_1\alpha_1 + \ldots + d_n\alpha_n$. Let $\star$ be the Dynkin product on a neighborhood of 0 in $\mathfrak{g}$. Then, for some constants $C_{\alpha,\beta} \in \mathbb{R}$, we have

$$(x \star y)_i = x_i + y_i + \sum_{\{\alpha,\beta \mid d_\alpha \geq 1, d_\beta \geq 1, d_\alpha + d_\beta \geq d_i\}} C_{\alpha,\beta} x^\alpha y^\beta.$$

For the Dynkin product $\star_0$ associated with the Lie bracket $[\cdot,\cdot]_0$ of the osculating Carnot algebra as in (12.39), we have

$$(x \star_0 y)_i = x_i + y_i + \sum_{\{\alpha,\beta \mid d_\alpha \geq 1, d_\beta \geq 1, d_\alpha + d_\beta = d_i\}} C_{\alpha,\beta} x^\alpha y^\beta,$$

where we have chopped the terms with $d_\alpha + d_\beta > d_i$.

Exercise 12.9.19 In the construction of the family of products on a polarized Lie algebra $\mathfrak{g}$ as in the proof of Mitchell's Theorem, for $\epsilon > 0$, we denote by $\star_\epsilon$ the Dynkin product associated with the dilated brackets $[\cdot,\cdot]_\epsilon$ as in (12.38). We have

$$g_1 \star_\epsilon g_2 = \delta_\epsilon^{-1}(\delta_\epsilon(g_1) \star \delta_\epsilon(g_2)), \qquad \forall g_1, g_2 \in \mathfrak{g}.$$

Exercise 12.9.20 Theorem 12.5.5, implies Theorem 12.5.1, and, in particular,

$$\left| d(1,g) - d_0(1,g) \right| = O\left(d(1,g)^{1+1/s} \right), \qquad \text{as } g \to 1.$$

Hint. As $d_0(1,g) \to 0$, take $\epsilon := d_0(1,g)$ and $p := \delta_\epsilon^{-1}(g)$.

Exercise 12.9.21 Let $(G, \star_\epsilon)_{\epsilon \in [-1,1]}$ be a sequence of Lie group structures on a set G as in Sect. 12.2.3. Let $X^{1,\epsilon}, \ldots, X^{k,\epsilon}$ be left-invariant vector fields on $(G, \star_\epsilon)$ that depend smoothly on $\epsilon \in [-1, 1]$. Assume that $X^{1,\epsilon}, \ldots, X^{k,\epsilon}$ define a (bracket-generating) sub-Riemannian structure on G of step s. Let d_ϵ be the sub-Riemannian metric. Fix a Riemannian metric ρ on G. Then there exist a neighborhood U of 1 in G and a constant $C > 1$ such that

$$\frac{1}{C}\rho \leq d_\epsilon \leq C\rho^{1/s} \qquad \text{on } U, \forall \epsilon \in [-1, 1].$$

Hint. Define a map E_ϵ as in Proposition 7.1.20, which depends smoothly on ϵ. Consider the map $\mathbb{R}^n \times [-1, 1] \to G \times [-1, 1]$, $(\mathbf{t}, \epsilon) \mapsto (E_\epsilon(\mathbf{t}), \epsilon)$. Proceed as in Corollary 7.1.21.

Exercise 12.9.22 The tangent cone of each Carnot group G is G itself. In fact, dilations δ_λ provide isometries between (G, d_{cc}) and $(G, \lambda d_{cc})$.

Exercise 12.9.23 Using Exercise 12.9.2 instead of Remark 12.2.5, one can upgrade equation (12.50) of the Quantitative Mitchell Theorem to

$$-C\epsilon^{1/s} d_0(p, q)^{1/s} \leq d_0(p, q) - d_\epsilon(p, q) \leq C\epsilon^{1/s} d_\epsilon(p, q)^{1/s}.$$

Exercise 12.9.24 If M is a Riemannian manifold, then M is polarized by TM and, for every $p \in M$, the osculating Carnot group of M at p is commutative.

Exercise 12.9.25 If M is a contact 3-manifold, then for every $p \in M$ the osculating Carnot group of M at p is the Heisenberg group.

Exercise 12.9.26 If G is a Carnot group, then for every $p \in G$, the osculating Carnot group of G is G itself.

Exercise 12.9.27 There is a Lie group G with a left-invariant bracket-generating distribution such that the osculating Carnot group of G is not isomorphic to G.

Exercise 12.9.28 The Stoll metric d_{Stoll} relative to a finite generating set $S \subset G$ coincides with the left-invariant sub-Finsler metric induced by the norm whose unit ball is the convex hull of S in the Lie algebra $\text{Lie}(G)$ and with left-invariant distribution induced by the subspace spanned by S in $\text{Lie}(G)$.
Hint. Prove it directly or invoke Berestovski's Theorem 7.4.1.

Chapter 13
Rank-One Symmetric Spaces

In Chap. 14, we will show that to every Riemannian symmetric space, one can associate a 'visual boundary' that has the structure of a Carnot group. In this chapter, we review the classical notion of symmetric space. We will see that each rank-one symmetric space of noncompact type has the structure of a metric Lie group.

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 spaces. The *rank* of a symmetric space is the largest dimension of a flat subspace in M, where a *flat subspace of dimension* n in M is the image of a local isometry from Euclidean $\mathbb{R}^n$ to M. For example, spheres and hyperbolic spaces have rank 1, whereas Euclidean n-space has rank n. A symmetric space is said to be of *noncompact type* if it is not the product of two symmetric spaces one of which is either compact or Euclidean. Hence, we shall exclude the spheres and the Euclidean spaces. Symmetric spaces were first introduced and studied by Élie Cartan in 1926; see [Car26, Car27]. In particular, Cartan provided a complete description of these spaces by means of the classification of simple Lie algebras.

In this chapter, taking Cartan's description for granted, we first prove that every rank-one symmetric space of noncompact type admits a Lie group structure of a semi-direct product with a precise formula for the left-invariant distance. The fact that such spaces admit semi-direct-product structures has been known at least since Ernst Heintze's work in the 1970s; see [Hei74]. However, an explicit formula for the left-invariant distances is not found in the existing literature. To study these spaces, we will need the following result: Let M be a rank-one symmetric space of noncompact type, then M is one of the following spaces, which we call $\mathbb{A}$-*hyperbolic spaces* $\mathbb{A}\mathbf{H}^n$, 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 1950s; see Arthur Besse's 1978 book [Bes78,

E. Le Donne, *Metric Lie Groups*, Graduate Texts in Mathematics 306,
https://doi.org/10.1007/978-3-031-98832-5_13

Section 3.G] and see Ernst Heintze's 1974 paper [Hei74, Section 5] for a geometric proof.

We shall introduce $\mathbb{A}$-hyperbolic spaces as metric spaces, where $\mathbb{A}$ is the set of the real, complex, or quaternionic numbers, giving a unified treatment of the subject. We will comment on the octonionic case in Remark 13.0.2. For simplicity of notation, we shall restrict to the two-dimensional case, but all constructions generalize to higher dimensions without effort; see the exercise section. Following Felix Klein's construction, we shall describe the $\mathbb{A}$-*hyperbolic space* $\mathbb{A}\mathbf{H}^2$ as an open subset of the projectivization of the space $\mathbb{A}^{2,1}$ equipped with a Hermitian form of type (2, 1), as of Definition 13.1.4. We shall recall the distance function on $\mathbb{A}\mathbf{H}^2$, referring to Martin Bridson and André Häfliger's comprehensive book [BH99, Part II, Chapter 10].

To recall the Lie group structure on each $\mathbb{A}\mathbf{H}^2$, in Sect. 13.3, we revise the continuous Heisenberg group $\mathcal{N}$ modeled on $\mathbb{A}$ and its intrinsic dilations; we follow [Ste93, Chapter XII Section 1]. We will double check that, after the identification of $\mathbb{A}\mathbf{H}^2$ with $\mathcal{N} \rtimes \mathbb{R}_+$, the hyperbolic distance is invariant under left translations on $\mathcal{N} \rtimes \mathbb{R}_+$. We shall write explicitly the distance on the $\mathbb{A}$-hyperbolic space modeled as $\mathcal{N} \rtimes \mathbb{R}_+$ in terms of elementary functions of the coordinates. This whole chapter is devoted to the proof of the following summarizing result:

Theorem 13.0.1 *For every $\mathbb{A} \in \{\mathbb{R}, \mathbb{C}, \mathbb{H}\}$ the $\mathbb{A}$-hyperbolic space $\mathbb{A}\mathbf{H}^2$ is isometric to the manifold $\mathbb{A} \times \mathfrak{Im}(\mathbb{A}) \times \mathbb{R}_+$ equipped with the group product and the left-invariant distance d given for all $(\xi, \mathbf{v}, \lambda)$ and $(\xi', \mathbf{v}', \lambda') \in \mathbb{A} \times \mathfrak{Im}(\mathbb{A}) \times \mathbb{R}_+$ by*

$$(\xi, \mathbf{v}, \lambda) \cdot (\xi', \mathbf{v}', \lambda') = \left(\xi + \lambda\xi', \mathbf{v} + \lambda^2\mathbf{v}' + 2\mathfrak{Im}(\bar{\xi}\lambda\xi'), \lambda\lambda'\right),$$

and

$$d((\xi, \mathbf{v}, \lambda), (\xi', \mathbf{v}', \lambda')) = \operatorname{arccosh}\sqrt{\left(1 + \frac{|\xi - \xi'|^2 + |\lambda - \lambda'|^2}{2\lambda\lambda'}\right)^2 + \left(\frac{|\mathbf{v} + \bar{\xi}\xi' - \mathbf{v}' - \bar{\xi}'\xi|}{2\lambda\lambda'}\right)^2}.$$

The proof of Theorem 13.0.1 is spread in Theorems 13.2.2 and 13.3.3, and their preliminaries. See also the high-dimensional generalization in Exercise 13.4.22.

Remark 13.0.2 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 of defining a notion of vector space over the octonions. For the basic ideas on how to deal with this case and build the octonionic-hyperbolic plane, see [Pil22].

13.1 Preliminary Notions for Rank-One Symmetric Spaces

13.1.1 Euclidean Hurwitz Algebras

The algebraic structure needed to construct rank-one symmetric spaces is exactly given by the Euclidean Hurwitz algebras, also referred to as the *normed division algebras over* $\mathbb{R}$. Even if they seem very general structures, there are only 4 examples of them; see Theorem 13.1.2. Still, it is convenient to see them from a general viewpoint.

Definition 13.1.1 A *Euclidean Hurwitz algebra* $\mathbb{A}$ is a not necessarily associative division algebra over $\mathbb{R}$ with an identity 1, together with a multiplicative norm $|\cdot|$. In order to clarify, a *(not necessarily associative) algebra* $\mathbb{A}$ over $\mathbb{R}$ is a vector space over $\mathbb{R}$ that is equipped with a bilinear binary multiplication operation $\mathbb{A} \times \mathbb{A} \to \mathbb{A}$. Whereas, an algebra $\mathbb{A}$ is said to be a *division algebra* if

$$\xi, \zeta \in \mathbb{A},\ (\xi\zeta = 0) \implies (\xi = 0 \text{ or } \zeta = 0), \tag{13.1}$$

while a norm $|\cdot|$ on an algebra $\mathbb{A}$ is said to be *multiplicative* if

$$|\xi\zeta| = |\xi|\,|\zeta|, \qquad \forall \xi, \zeta \in \mathbb{A}. \tag{13.2}$$

Every Euclidean Hurwitz algebra admits an anti-involution, called *conjugation*. There is a general formula to define the conjugate in terms of the norm; see Exercise 13.4.4. In this sense, Hurwitz algebras may be thought of as a natural generalisation of the real numbers and the complex numbers. We denote the *conjugate of* ξ by $\bar{\xi}$.

It is a result of Hurwitz that the only Euclidean Hurwitz algebras are the following four examples:

1. The real numbers $\mathbb{R}$, with the absolute value as norm.
2. The complex numbers $\mathbb{C}$, with Euclidean norm given by the natural identification with $\mathbb{R}^2$.
3. The quaternionic numbers $\mathbb{H}$, with Euclidean norm given by the natural identification with $\mathbb{R}^4$.
4. The octonionic numbers $\mathbb{O}$, with Euclidean norm given by the natural identification with $\mathbb{R}^8$.

Theorem 13.1.2 (Hurwitz [Hur22]) *The only Euclidean Hurwitz algebras are* $\mathbb{R}$, $\mathbb{C}$, $\mathbb{H}$, *and* $\mathbb{O}$.

Given an element ξ in a Euclidean Hurwitz algebra we define its *real part* as $\mathfrak{Re}(\xi) := \frac{\xi+\bar{\xi}}{2}$ and its *imaginary part* as $\mathfrak{Im}(\xi) := \xi - \mathfrak{Re}(\xi) = \frac{\xi-\bar{\xi}}{2}$. Notice that we follow the convention that the imaginary part is an element of the set $\mathfrak{Im}(\mathbb{A})$ of *purely imaginary numbers*.

While we do not need to recall the real or complex numbers, we revise the quaternions and the octonions: The *quaternions*, or *quaternionic numbers*, are a 4-dimensional associative algebra over $\mathbb{R}$ with basis $\{1, i, j, k\}$, where 1 is the unit element and i, j, k follow the rules:

$$ij = k, \quad jk = i, \quad ki = j,$$

and

$$i^2 = j^2 = k^2 = -1.$$

The quaternions are typically denoted by $\mathbb{H}$ in honour of its discovery by W. R. Hamilton. On the quaternions, the *conjugate* of an element $\xi = \xi_0 + \xi_1 i + \xi_2 j + \xi_3 k$ is $\bar{\xi} = \xi_0 - \xi_1 i - \xi_2 j - \xi_3 k$, its real part is $\mathfrak{Re}(\xi) := \xi_0 \in \mathbb{R}$, while the *imaginary part* of ξ is $\mathfrak{Im}(\xi) := \xi_1 i + \xi_2 j + \xi_3 k \in \mathfrak{Im}(\mathbb{H}) \simeq \mathbb{R}^3$. On $\mathbb{H}$, we consider the Euclidean norm via the identification $\mathbb{H} \simeq \mathbb{R}^4$, which makes $\mathbb{H}$ into a Euclidean Hurwitz algebra; see Exercise 13.4.2.

The *octonions*, or *octonionic numbers*, are a nonassociative Euclidean Hurwitz algebra of dimension 8 over $\mathbb{R}$. This algebra is defined via the Cayley-Dickson construction [Dic19] as the space of pairs of quaternion numbers $\{(a, b) \mid a, b \in \mathbb{H}\}$, together with the multiplication

$$(a, b) \cdot (c, d) := (ac - \bar{d}b, da + b\bar{c}), \qquad \forall a, b, c, d \in \mathbb{H}.$$

Working with nonassociative algebras takes some time to get used to, and we do not want to discuss them in this text. Therefore, we will restrict our discussion from now on to the associative Euclidean Hurwitz algebras. But the reader should keep in mind that with a little more work, the story extends to the octonion setting as well; see [Pil22].

13.1.2 Hermitian Forms

Notice that every associative Euclidean Hurwitz algebra is a ring. Let $\mathbb{A}$ be either $\mathbb{R}$, $\mathbb{C}$, or $\mathbb{H}$. Consider the right $\mathbb{A}$-module $\mathbb{A}^3$: for $\mathbf{z} = (z_1, z_2, z_3) \in \mathbb{A}^3$ and $\lambda \in \mathbb{A}$, we shall consider $\mathbf{z}\lambda := (z_1\lambda, z_2\lambda, z_3\lambda)$.

Definition 13.1.3 A *Hermitian form* over the right $\mathbb{A}$-module $\mathbb{A}^3$ is a map

$$\langle \cdot, \cdot \rangle : \mathbb{A}^3 \times \mathbb{A}^3 \to \mathbb{A},$$

such that for all $\lambda, \mu \in \mathbb{A}$, and $\mathbf{z}, \mathbf{w}, \mathbf{z}', \mathbf{w}' \in \mathbb{A}^3$,

$$\langle \mathbf{z} + \mathbf{z}', \mathbf{w} + \mathbf{w}' \rangle = \langle \mathbf{z}, \mathbf{w} \rangle + \langle \mathbf{z}, \mathbf{w}' \rangle + \langle \mathbf{z}', \mathbf{w} \rangle + \langle \mathbf{z}', \mathbf{w}' \rangle,$$

$$\langle \mathbf{z}\lambda, \mathbf{w}\mu \rangle = \bar{\lambda} \langle \mathbf{z}, \mathbf{w} \rangle \mu, \tag{13.3}$$

and

$$\langle \mathbf{z}, \mathbf{w} \rangle = \overline{\langle \mathbf{w}, \mathbf{z} \rangle}. \tag{13.4}$$

A Hermitian form is *nondegenerate*, if

$$\left(\langle \mathbf{z}, \mathbf{y} \rangle = 0, \ \forall \mathbf{y} \in \mathbb{A}^3 \right) \Longrightarrow \mathbf{z} = 0.$$

The *orthogonal complement* of $\mathbf{z} \in \mathbb{A}^3$ is the set

$$\mathbf{z}^{\perp} := \{ \mathbf{y} \in \mathbb{A}^3 | \langle \mathbf{z}, \mathbf{y} \rangle = 0 \}.$$

If $\langle \cdot, \cdot \rangle$ is nondegenerate, then $\mathbf{v} \oplus \mathbf{v}^{\perp} = \mathbb{A}^3$. Every Hermitian form $\langle \cdot, \cdot \rangle$ over the right $\mathbb{A}$-module $\mathbb{A}^3$ is given by a 3×3 matrix

$$H = \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix}, \quad a, b, c, e, d, f, g, h, i \in \mathbb{A},$$

through $\langle \mathbf{z}, \mathbf{w} \rangle = \mathbf{z}^* H \mathbf{w}$. Here $\mathbf{z}^* = (\bar{z}_1, \bar{z}_2, \bar{z}_3)$ is the Hermitian conjugate of $\mathbf{z}$. Be aware that to use the matrix products, we should see $\mathbf{w}$ as a column vector. In order to give a Hermitian form, such a matrix H satisfies

$$\begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix} = \begin{pmatrix} \bar{a} & \bar{d} & \bar{g} \\ \bar{b} & \bar{e} & \bar{h} \\ \bar{c} & \bar{f} & \bar{i} \end{pmatrix}.$$

Equivalently said, a transformation is a Hermitian form if the matrix that represents it equals its Hermitian transpose, where given a matrix $M = (m_{ij})$ over a Euclidean Hurwitz algebra, the *Hermitian transpose* of M is $M^* := (\overline{m}_{ji})$.

The classical spectral theory over $\mathbb{C}$ extends to associative Hurwitz algebras; see, for example, [FP03]. In particular, the matrices representing Hermitian forms over $\mathbb{A}$ only have real eigenvalues.

13.1.3 Hermitian Forms of Signature (2, 1)

Definition 13.1.4 (The Space $\mathbb{A}^{2,1}$) A Hermitian form over $\mathbb{A}^3$ has *signature* (2, 1) if it has exactly two strictly positive eigenvalues and one strictly negative eigenvalue. To stress when $\mathbb{A}^3$ is considered equipped with a Hermitian form of signature (2, 1), we denote it by $\mathbb{A}^{2,1}$. Up to linear changes of variables, there is only one such space $\mathbb{A}^{2,1}$.

Lemma 13.1.5 *Let $\langle \cdot, \cdot \rangle$ be a Hermitian form of signature* (2, 1) *over the right $\mathbb{A}$-module $\mathbb{A}^3$. Let $\mathbf{z}$ be a vector in $\mathbb{A}^{2,1}$ with $\langle \mathbf{z}, \mathbf{z} \rangle < 0$, then $\langle \cdot, \cdot \rangle$ is positive definite on $\mathbf{z}^\perp$.*

Proof We begin by stressing that a Hermitian form of signature (2, 1) has a trivial kernel. Assume that there exists a vector $\mathbf{y} \in \mathbf{z}^\perp$ such that $\langle \mathbf{y}, \mathbf{y} \rangle < 0$. We claim that $\mathbf{z}$ and $\mathbf{y}$ are linearly independent over $\mathbb{A}$. Indeed, otherwise, noticing that $\mathbf{z}$ cannot be zero, we would have $\mathbf{y} = \mathbf{z}\lambda$ for some $\lambda \in \mathbb{A}$, but then

$$0 \overset{(\mathbf{y} \in \mathbf{z}^\perp)}{=} \langle \mathbf{z}, \mathbf{y} \rangle = \langle \mathbf{z}, \mathbf{z}\lambda \rangle \overset{(13.3)}{=} \langle \mathbf{z}, \mathbf{z} \rangle \lambda,$$

and, because we are in a division algebra, (13.1), either λ would be zero (but then this contradicts $\langle \mathbf{y}, \mathbf{y} \rangle < 0$) or $\langle \mathbf{z}, \mathbf{z} \rangle = 0$ (which contradicts the assumption $\langle \mathbf{z}, \mathbf{z} \rangle < 0$). Consequently, the space $\operatorname{span}(\mathbf{z}, \mathbf{y})$ is a two-dimensional subspace on which $\langle \cdot, \cdot \rangle$ is negative-definite. This is impossible since in $\mathbb{A}^{2,1}$, we only have one negative eigenvector by assumption. □

Lemma 13.1.6 (Reverse Schwarz Inequality) *Let $\langle \cdot, \cdot \rangle$ be a Hermitian form of signature* (2, 1) *over the right $\mathbb{A}$-module $\mathbb{A}^3$. For every $\mathbf{z}, \mathbf{w} \in \mathbb{A}^3$ with $\langle \mathbf{z}, \mathbf{z} \rangle < 0$ and $\langle \mathbf{w}, \mathbf{w} \rangle < 0$, we have*

$$\langle \mathbf{z}, \mathbf{w} \rangle \langle \mathbf{w}, \mathbf{z} \rangle \geq \langle \mathbf{z}, \mathbf{z} \rangle \langle \mathbf{w}, \mathbf{w} \rangle,$$

with equality if and only if $\mathbf{z}$ and $\mathbf{w}$ are linearly dependent over $\mathbb{A}$.

Proof Suppose that $\mathbf{z}$ and $\mathbf{w}$ are linearly independent. Since $\langle \mathbf{w}, \mathbf{w} \rangle < 0$ and $\langle \cdot, \cdot \rangle$ is positive definite on $\mathbf{z}^\perp$ by Lemma 13.1.5, we infer that $\mathbf{w} \notin \mathbf{z}^\perp$, i.e., $\langle \mathbf{z}, \mathbf{w} \rangle \neq 0$. Let $\lambda := -\langle \mathbf{z}, \mathbf{z} \rangle \langle \mathbf{z}, \mathbf{w} \rangle^{-1}$. Since $\mathbf{z}$ and $\mathbf{w}$ are linearly independent, $\mathbf{z} + \mathbf{w}\lambda \neq 0$ and a quick calculation shows that $\mathbf{z} + \mathbf{w}\lambda \in \mathbf{z}^\perp$. Again by Lemma 13.1.5, this implies that $\langle \mathbf{z} + \mathbf{w}\lambda, \mathbf{w}\lambda \rangle = \langle \mathbf{z} + \mathbf{w}\lambda, \mathbf{z} + \mathbf{w}\lambda \rangle > 0$. Expanding the left-hand side gives

$$-\langle \mathbf{z}, \mathbf{z} \rangle + \langle \mathbf{z}, \mathbf{z} \rangle^2 \langle \mathbf{w}, \mathbf{w} \rangle \langle \mathbf{w}, \mathbf{z} \rangle^{-1} \langle \mathbf{z}, \mathbf{w} \rangle^{-1} > 0.$$

Dividing by $\langle \mathbf{z}, \mathbf{z} \rangle < 0$ and rearranging, we obtain

$$\langle \mathbf{z}, \mathbf{w} \rangle \langle \mathbf{w}, \mathbf{z} \rangle > \langle \mathbf{z}, \mathbf{z} \rangle \langle \mathbf{w}, \mathbf{w} \rangle.$$

If $\mathbf{z}$ and $\mathbf{w}$ are linearly dependent, then there exists $\lambda \in \mathbb{A}$ such that $\mathbf{z} = \mathbf{w}\lambda$. Then the equality

$$\langle \mathbf{z}, \mathbf{w} \rangle \langle \mathbf{w}, \mathbf{z} \rangle = \langle \mathbf{z}, \mathbf{z} \rangle \langle \mathbf{w}, \mathbf{w} \rangle$$

follows straightforwardly from the definition of Hermitian forms. □

13.2 The $\mathbb{A}$-Hyperbolic Space $\mathbb{A}\mathbf{H}^2$

13.2.1 Definition and Properties

Let $\mathbb{A}$ be either $\mathbb{R}$, $\mathbb{C}$, or the quaternions $\mathbb{H}$. Let $\mathbb{A}^{2,1}$ be the right-module over $\mathbb{A}$ of $\mathbb{A}$-dimension 3, equipped with some Hermitian form $\langle \cdot, \cdot \rangle$ of signature $(2, 1)$. Define the subset $V_- := \{\mathbf{z} \in \mathbb{A}^{2,1} : \langle \mathbf{z}, \mathbf{z} \rangle < 0\}$. For every nonzero scalar λ, we have that $\mathbf{z}\lambda$ is in V_- if and only if so is $\mathbf{z}$. We may therefore speak of negative $\mathbb{A}$-lines in $\mathbb{A}^{2,1}$.

The *projective model of hyperbolic space* is the collection of negative lines in $\mathbb{A}^{2,1}$. Formally, we have a projective map $\mathbb{P} : \mathbb{A}^{2,1} \to \mathbb{P}(\mathbb{A}^{2,1})$, where $\mathbb{P}(\mathbb{A}^{2,1})$ is the projective space over $\mathbb{A}^{2,1}$, here

$$\mathbb{P}(\mathbf{z}) := \{\mathbf{z}\lambda : \lambda \in \mathbb{A} \setminus \{0\}\}, \qquad \text{for } \mathbf{z} \in \mathbb{A}^{2,1}.$$

Thus, we define the *hyperbolic space* $\mathbb{A}\mathbf{H}^2$ to be $\mathbb{P}(V_-)$. In symbols, we write:

$$\mathbb{A}\mathbf{H}^2 := \{\mathbb{P}(\mathbf{z}) \in \mathbb{P}(\mathbb{A}^{2,1}) : \langle \mathbf{z}, \mathbf{z} \rangle < 0\}. \tag{13.5}$$

We denote points in $\mathbb{A}\mathbf{H}^2$ simply by z or w. Following [BH99, Chapter II.10, p.302] for this projective model of hyperbolic space, the *Bergman metric* on $\mathbb{A}\mathbf{H}^2$ is given by the distance function d defined by the equation[1]

$$\left(\cosh\left(d(z, w)\right)\right)^2 = \frac{\langle \mathbf{z}, \mathbf{w} \rangle \langle \mathbf{w}, \mathbf{z} \rangle}{\langle \mathbf{z}, \mathbf{z} \rangle \langle \mathbf{w}, \mathbf{w} \rangle}, \tag{13.6}$$

where $\mathbf{z}$ and $\mathbf{w}$ are any vectors in $\mathbb{A}^{2,1}$ such that $\mathbb{P}(\mathbf{z}) = z$ and $\mathbb{P}(\mathbf{w}) = w$; here cosh denotes the hyperbolic cosine (13.13). Equivalently, using the inverse hyperbolic cosine, we write

$$d(z, w) := \operatorname{arccosh}\sqrt{\frac{\langle \mathbf{z}, \mathbf{w} \rangle \langle \mathbf{w}, \mathbf{z} \rangle}{\langle \mathbf{z}, \mathbf{z} \rangle \langle \mathbf{w}, \mathbf{w} \rangle}}$$

[1] We decided to follow the convention of not dividing by 2 in the definition of the distance, as instead it is done by Parker [Par03]. We instead follow, for example, Bridson-Haefliger in order to have a CAT(-1) space.

$$= \operatorname{arccosh}\sqrt{\frac{\langle \mathbf{z}, \mathbf{w}\rangle \overline{\langle \mathbf{z}, \mathbf{w}\rangle}}{\langle \mathbf{z}, \mathbf{z}\rangle \langle \mathbf{w}, \mathbf{w}\rangle}}$$

$$= \operatorname{arccosh}\left(\frac{|\langle \mathbf{z}, \mathbf{w}\rangle|}{\sqrt{\langle \mathbf{z}, \mathbf{z}\rangle \langle \mathbf{w}, \mathbf{w}\rangle}}\right).$$

Thanks to the Reverse Schwarz Inequality (see Lemma 13.1.6), we have

$$\frac{\langle \mathbf{z}, \mathbf{w}\rangle \langle \mathbf{w}, \mathbf{z}\rangle}{\langle \mathbf{z}, \mathbf{z}\rangle \langle \mathbf{w}, \mathbf{w}\rangle} \geq 1, \qquad \forall \mathbf{z}, \mathbf{w} \in V_-.$$

Hence, the equation makes sense: recall that $\cosh$ bijectively sends $[0, +\infty)$ onto $[1, +\infty)$.

We shall check that the formula does not depend on the representatives: If $\mathbf{z}'$ and $\mathbf{w}'$ are any other vectors in $\mathbb{A}^{2,1}$ such that $\mathbb{P}(\mathbf{z}') = z$ and $\mathbb{P}(\mathbf{w}') = w$, then there exist nonzero scalars λ and μ in $\mathbb{A}$ such that $\mathbf{z}' = \mathbf{z}\lambda$ and $\mathbf{w}' = \mathbf{w}\mu$. Thus

$$\begin{aligned} \frac{|\langle \mathbf{z}', \mathbf{w}'\rangle|^2}{\langle \mathbf{z}', \mathbf{z}'\rangle \langle \mathbf{w}', \mathbf{w}'\rangle} &= \frac{|\langle \mathbf{z}\lambda, \mathbf{w}\mu\rangle|^2}{\langle \mathbf{z}\lambda, \mathbf{z}\lambda\rangle \langle \mathbf{w}\mu, \mathbf{w}\mu\rangle} \\ &\overset{(13.3)}{=} \frac{|\bar{\lambda}\langle \mathbf{z}, \mathbf{w}\rangle \mu|^2}{|\lambda|^2 \langle \mathbf{z}, \mathbf{z}\rangle |\mu|^2 \langle \mathbf{w}, \mathbf{w}\rangle} \\ &\overset{(13.2)}{=} \frac{|\lambda|^2 \cdot |\langle \mathbf{z}, \mathbf{w}\rangle|^2 \cdot |\mu|^2}{|\lambda|^2 \cdot |\mu|^2 \cdot \langle \mathbf{z}, \mathbf{z}\rangle \langle \mathbf{w}, \mathbf{w}\rangle} = \frac{|\langle \mathbf{z}, \mathbf{w}\rangle|^2}{\langle \mathbf{z}, \mathbf{z}\rangle \langle \mathbf{w}, \mathbf{w}\rangle}. \end{aligned}$$

Here, we made explicit use of the fact that the norm on $\mathbb{A}$ is multiplicative and that the norms are real numbers (and hence central elements).

13.2.2 Models of $\mathbb{A}$-Hyperbolic Space

There are multiple models of real-hyperbolic space $\mathbb{R}\mathbf{H}^2$. The most common ones are:

1. the *Minkowski hyperboloid model* [BH99, p.18],
2. the *Poincaré ball model* [BH99, p.86],
3. the *Poincaré half-space model* [BH99, p.90],
4. the *Klein ball model* [BH99, p.83 and p.310],
5. and the *Siegel parabolic model* [BH99, p.310].

The Poincaré ball model and the Poincaré half-space model are *conformal*, in the sense that they are angle-preserving. The Klein model and the Siegel domain model are *projective* in the sense that hyperbolic straight lines and planes are the intersection of Euclidean straight lines and planes with the unit ball in $\mathbb{R}^2$ and the interior of the paraboloid $x_1 = \frac{x_2^2}{2}$, respectively. The Siegel domain model is not often mentioned in the real-hyperbolic case. It may be understood as the projective analog of the conformal Poincaré half-space model. The hyperboloid model, on the other hand, is particularly common in physics but neither conformal nor projective.

The projective models have straightforward generalizations to hyperbolic spaces modeled over each associative Hurwitz algebra.

Let $\mathbb{A}$ be either $\mathbb{R}$, $\mathbb{C}$, or the quaternions $\mathbb{H}$. Let $\mathbf{z}$ and $\mathbf{w}$ be vectors in $\mathbb{A}^{2,1}$. The *first Hermitian form* on $\mathbb{A}^{2,1}$ is defined as

$$\langle \mathbf{z}, \mathbf{w} \rangle_1 := \bar{z}_1 w_1 + \bar{z}_2 w_2 - \bar{z}_3 w_3.$$

It is associated with the Hermitian matrix

$$J_1 := \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix}.$$

Every representative $\mathbf{z}' \in \mathbb{A}^{2,1}$ of a point $z \in \mathbb{A}\mathbf{H}^2$ satisfies $\langle \mathbf{z}', \mathbf{z}' \rangle_1 < 0$, i.e., $0 > \bar{z}_1 z_1 + \bar{z}_2 z_2 - \bar{z}_3 z_3 = |z_1|^2 + |z_2|^2 - |z_3|^2$. Therefore, the third component of such a z is nonzero. Hence, there exists exactly one representative $\mathbf{z}$ of z of the form

$$\mathbf{z} = \begin{bmatrix} z_1 \\ z_2 \\ 1 \end{bmatrix} \simeq (z_1, z_2, 1), \qquad \text{for which } |z_1|^2 + |z_2|^2 < 1.$$

This gives an injection $\iota : \mathbb{A}\mathbf{H}^2 \hookrightarrow \mathbb{A}^{2,1}$ as a section of $\mathbb{P} : V_- \to \mathbb{A}\mathbf{H}^2$ and identifies the hyperbolic space $\mathbb{A}\mathbf{H}^2$ with Klein's model which setwise is the unit ball $\mathbb{B}$ in $\mathbb{A}^2$. We call such a $\mathbf{z}$ the *standard lift* of z.

The *second Hermitian form* on $\mathbb{A}^{2,1}$ is defined as

$$\langle \mathbf{z}, \mathbf{w} \rangle_2 := \bar{z}_1 w_3 + \bar{z}_2 w_2 + \bar{z}_3 w_1. \tag{13.7}$$

It is associated with the Hermitian matrix

$$J_2 := \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}.$$

Every representative $\mathbf{z}' \in \mathbb{A}^{2,1}$ of a point $z \in \mathbb{A}\mathbf{H}^2$ satisfies $\langle \mathbf{z}', \mathbf{z}' \rangle_2 < 0$, i.e., $0 > \bar{z}_1 z_3 + \bar{z}_2 z_2 + \bar{z}_3 z_1 = 2\mathfrak{Re}(\bar{z}_1 z_3) + |z_2|^2$. Therefore, in particular, the third component of such a z is nonzero. Hence, there exists exactly one representative $\mathbf{z}$ of z of the form

$$\mathbf{z} = \begin{bmatrix} z_1 \\ z_2 \\ 1 \end{bmatrix}.$$

This gives an injection $\iota : \mathbb{A}\mathbf{H}^2 \hookrightarrow \mathbb{A}^{2,1}$ as a section of $\mathbb{P} : V_- \to \mathbb{A}\mathbf{H}^2$ and identifies each point $z \in \mathbb{A}\mathbf{H}^2$ with the point $(z_1, z_2) \in \mathbb{A}^2$ such that $2\mathfrak{Re}(z_1) + |z_2|^2 < 0$. This domain

$$\mathbb{D} := \{(z_1, z_2) \in \mathbb{A}^2 : 2\mathfrak{Re}(z_1) + |z_2|^2 < 0\} \tag{13.8}$$

is called the *Siegel (parabolic) domain*.

A map that passes from one model of hyperbolic space to another is called a *Cayley transform*. For example, a Cayley transform interchanging the first and second Hermitian forms is given by

$$\frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 0 & 1 \\ 0 & \sqrt{2} & 0 \\ 1 & 0 & -1 \end{bmatrix}.$$

A Cayley transform is not unique since it may be pre- and post-composed by any unitary matrix preserving the relevant Hermitian form.

In the following we will denote by $\mathbf{z}$ both the standard lift $(z_1, z_2, 1) \in \mathbb{A}^{2,1}$ of $z \in \mathbb{A}\mathbf{H}^2$ and its identification with the point (z_1, z_2) in the Siegel domain $\mathbb{D}$.

13.2.3 A Foliation of the Siegel Domain

We consider the hyperbolic space $\mathbb{A}\mathbf{H}^2$, as defined in (13.5), as a subset of the projective space $\mathbb{P}(\mathbb{A}^{2,1})$. Within this larger subset, the hyperbolic space $\mathbb{A}\mathbf{H}^2$ has a topological boundary denoted by $\partial\mathbb{A}\mathbf{H}^2$.

When we consider $\mathbb{A}^{2,1}$ with respect to the second Hermitian form $\langle \cdot, \cdot \rangle_2$, from (13.7), we distinguish two types of points in $\partial\mathbb{A}\mathbf{H}^2$: we say that $z \in \partial\mathbb{A}\mathbf{H}^2$ is *finite*, if z admits a representative of the form $\mathbf{z} = (z_1, z_2, 1) \in \mathbb{A}^{2,1}$. Otherwise, we say that z is *not finite*.

We point out that there exists exactly one point on the boundary that is not finite, and it is represented by $(1, 0, 0)$. Instead, the finite points in $\partial\mathbb{A}\mathbf{H}^2$ are those that are represented by a point in the boundary of the Siegel domain $\mathbb{D}$ as a subset of $\mathbb{A}^2$. Namely, a point $z \in \mathbb{P}(\mathbb{A}^{2,1})$ is a finite boundary point for $\mathbb{A}\mathbf{H}^2$ if (and only if) it has a representative $\mathbf{z} = (z_1, z_2, 1) \in \mathbb{A}^{2,1}$ such that

$$z_1 + \bar{z}_1 + |z_2|^2 = 0.$$

For this reason, we define the *parabolic boundary* of the hyperbolic space $\mathbb{A}\mathbf{H}^2$ (modeled by the Siegel domain) as

$$\partial_{\mathrm{Par}}\mathbb{D} := \left\{(z_1, z_2) \in \mathbb{A}^2 : 2\Re\mathfrak{e}(z_1) + |z_2|^2 = 0\right\}.$$

We stress that the topological boundary of $\mathbb{A}\mathbf{H}^2$ is the one-point compactification of $\partial_{\mathrm{Par}}\mathbb{D}$. Setting $\xi := z_2/\sqrt{2}$, the condition for the parabolic boundary becomes $\Re\mathfrak{e}(z_1) = -|\xi|^2$. Hence, we may write $\mathbf{z} = (-|\xi|^2 - v, \sqrt{2}\xi)$ for a $v \in \Im\mathfrak{m}(\mathbb{A})$. This shows that the parabolic boundary can be identified with $\mathbb{A} \times \Im\mathfrak{m}(\mathbb{A})$. We shall extend this identification within the Siegel domain, seeing the boundary as a limiting leaf of a foliation of the domain.

Let $\lambda \in \mathbb{R}_+$ and consider each point $z \in \mathbb{A}\mathbf{H}^2$ for which the standard lift $\mathbf{z}$ satisfies $\langle \mathbf{z}, \mathbf{z}\rangle_2 = -2\lambda^2$. Define

$$H_\lambda := \left\{z \in \mathbb{A}\mathbf{H}^2 \mid \langle \mathbf{z}, \mathbf{z}\rangle_2 = -2\lambda^2\right\}.$$

The collection $\mathcal{F} = \{H_\lambda\}_{\lambda\in\mathbb{R}_+}$ is a foliation of $\mathbb{A}\mathbf{H}^2$. A point $z \in \mathbb{A}\mathbf{H}^2$ lies in H_λ if and only if its identification with $\mathbf{z} \in \mathbb{D}$, satisfies $2\Re\mathfrak{e}(z_1) = -|z_2|^2 - 2\lambda^2$. We write z_2 as $\sqrt{2}\xi$, with $\xi \in \mathbb{A}$, and we get that $z_1 = \Re\mathfrak{e}(z_1) + \Im\mathfrak{m}(z_1) = -|\xi|^2 - \lambda^2 - \mathbf{v}$ for $\mathbf{v} := \Im\mathfrak{m}(-z_1) \in \Im\mathfrak{m}(\mathbb{A})$. Thus $z \in \mathbb{D}$ corresponds to a point $(\xi, \mathbf{v}, \lambda) \in \mathbb{A} \times \Im\mathfrak{m}(\mathbb{A}) \times \mathbb{R}_+$, via $(z_1, z_2) = (-|\xi|^2 - \lambda^2 - \mathbf{v}, \sqrt{2}\xi)$. This construction identifies the Siegel domain with the set $\mathbb{A} \times \Im\mathfrak{m}(\mathbb{A}) \times \mathbb{R}_+$.

Theorem 13.2.1 (Horospherical Coordinates) *The Siegel domain* $\mathbb{D}$*,* (13.8)*, can be parametrized via the following diffeomorphism:*

$$\varphi : \mathbb{A} \times \Im\mathfrak{m}(\mathbb{A}) \times \mathbb{R}_+ \to \mathbb{D}, \quad (\xi, \mathbf{v}, \lambda) \overset{\varphi}{\longmapsto} \left(-|\xi|^2 - \lambda^2 - \mathbf{v},\ \sqrt{2}\xi\right). \tag{13.9}$$

Proof It follows from the above construction of φ, that the map

$$(z_1, z_2) \mapsto \left(\frac{z_2}{\sqrt{2}},\ -\Im\mathfrak{m}(z_1), \sqrt{-\frac{\langle \mathbf{z},\ \mathbf{z}\rangle_2}{2}}\right),$$

where $\mathbf{z} = (z_1, z_2, 1)$, is the inverse of φ. Both φ and φ^{-1} are smooth. Thus φ identifies $\mathbb{A} \times \Im\mathfrak{m}(\mathbb{A}) \times \mathbb{R}_+$ with $\mathbb{D}$ as manifolds. □

The map φ from Theorem 13.2.1 is called *horospherical coordinate map*. Next, we pull back the Bergman metric given by (13.6) via such a map. In other words, the next result expresses the hyperbolic distance in horospherical coordinates.

Theorem 13.2.2 (Hyperbolic Distance in Horospherical Coordinates) *The pullback $\varphi^* d$ of the Bergman metric d to the manifold $\mathbb{A} \times \Im\mathfrak{m}(\mathbb{A}) \times \mathbb{R}_+$ by the horospherical coordinate map φ satisfies*

$$\varphi^* d((\xi, \mathbf{v}, \lambda), (\xi', \mathbf{v}', \lambda')) = \operatorname{arccosh}\sqrt{\left(1 + \frac{|\xi - \xi'|^2 + |\lambda - \lambda'|^2}{2\lambda\lambda'}\right)^2 + \left(\frac{|\mathbf{v} - \mathbf{v}' + 2\Im\mathfrak{m}(\bar{\xi}\xi')|}{2\lambda\lambda'}\right)^2}, \tag{13.10}$$

for all $(\xi, \mathbf{v}, \lambda)$ *and* $(\xi', \mathbf{v}', \lambda') \in \mathbb{A} \times \Im\mathfrak{m}(\mathbb{A}) \times \mathbb{R}_+$.

Proof It follows from the definition of the Bergman metric (13.6) that

$$\begin{aligned}
&\left(\cosh\left(\varphi^* d((\xi, \mathbf{v}, \lambda), (\xi', \mathbf{v}', \lambda'))\right)\right)^2 \\
&\quad = \left(\cosh\left(d(\varphi(\xi, \mathbf{v}, \lambda), \varphi(\xi', \mathbf{v}', \lambda'))\right)\right)^2 \\
&\quad = \frac{|\langle \varphi(\xi, \mathbf{v}, \lambda), \varphi(\xi', \mathbf{v}', \lambda')\rangle_2|^2}{\langle \varphi(\xi, \mathbf{v}, \lambda), \varphi(\xi, \mathbf{v}, \lambda)\rangle_2 \langle \varphi(\xi', \mathbf{v}', \lambda'), \varphi(\xi', \mathbf{v}', \lambda')\rangle_2}.
\end{aligned}$$

We start by explicitly computing $\langle \varphi(\xi, \mathbf{v}, \lambda), \varphi(\xi', \mathbf{v}', \lambda')\rangle_2$: using the definition of φ from Theorem 13.2.1, the definition of the second Hermitian form (13.7), and simple properties of elements in $\mathbb{A}$ (see Exercise 13.4.4), we get

$$\begin{aligned}
&\langle \varphi(\xi, \mathbf{v}, \lambda), \varphi(\xi', \mathbf{v}', \lambda')\rangle_2 \\
&\quad = \langle (-|\xi|^2 - \lambda^2 - \mathbf{v}, \sqrt{2}\xi, 1), (-|\xi'|^2 - \lambda'^2 - \mathbf{v}', \sqrt{2}\xi', 1)\rangle_2 \\
&\quad = \left[-|\xi|^2 - \lambda^2 + \mathbf{v}, \sqrt{2}\bar{\xi}, 1\right] \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} -|\xi'|^2 - \lambda'^2 - \mathbf{v}' \\ \sqrt{2}\xi' \\ 1 \end{bmatrix} \\
&\quad = \left[1, \sqrt{2}\bar{\xi}, -|\xi|^2 - \lambda^2 + \mathbf{v}\right] \begin{bmatrix} -|\xi'|^2 - \lambda'^2 - \mathbf{v}' \\ \sqrt{2}\xi' \\ 1 \end{bmatrix} \\
&\quad = -|\xi'|^2 - \lambda'^2 - \mathbf{v}' + 2\bar{\xi}\xi' - |\xi|^2 - \lambda^2 + \mathbf{v} \\
&\quad = -\bar{\xi}\xi + \bar{\xi}\xi' + \bar{\xi}'\xi - \bar{\xi}'\xi' + \bar{\xi}\xi' - \bar{\xi}'\xi - (\lambda^2 + \lambda'^2) + (\mathbf{v} - \mathbf{v}') \\
&\quad = -|\xi - \xi'|^2 + 2\Im\mathfrak{m}(\bar{\xi}\xi') - (\lambda^2 + \lambda'^2) + (\mathbf{v} - \mathbf{v}').
\end{aligned}$$

From this last calculation, we first see that $\langle \varphi(\xi, \mathbf{v}, \lambda), \varphi(\xi, \mathbf{v}, \lambda)\rangle_2 = -2\lambda^2$ and $\langle \varphi(\xi', \mathbf{v}', \lambda'), \varphi(\xi', \mathbf{v}', \lambda')\rangle_2 = -2\lambda'^2$ as we expect for horospherical coordinates. Then, we see that the squared norm of $\langle \varphi(\xi, \mathbf{v}, \lambda), \varphi(\xi', \mathbf{v}', \lambda')\rangle_2$ has value

$$\left(|\xi - \xi'|^2 + \lambda^2 + \lambda'^2\right)^2 + |\mathbf{v} - \mathbf{v}' + 2\mathfrak{Im}(\bar{\xi}\xi')|^2$$
$$= \left(2\lambda\lambda' + |\xi - \xi'|^2 + |\lambda - \lambda'|^2\right)^2 + |\mathbf{v} - \mathbf{v}' + 2\mathfrak{Im}(\bar{\xi}\xi')|^2.$$

The claim of the theorem follows after normalization. See Exercise 13.4.10 for other formulations for the distance function. □

Compare the formula in Theorem 13.2.2 with the conformal Poincaré half-space model of real-hyperbolic space $\mathbb{R}\mathbf{H}^2$; see Exercise 13.4.12.

13.3 Group Structures and Their Relations

13.3.1 The (First) Heisenberg Group Over $\mathbb{A}$

The manifold $\mathbb{A} \times \mathfrak{Im}(\mathbb{A})$ can be given the following Lie group structure,

$$(\xi, \mathbf{v}) \cdot (\xi', \mathbf{v}') := \left(\xi + \xi', \mathbf{v} + \mathbf{v}' + 2\mathfrak{Im}(\bar{\xi}\xi')\right), \quad \forall (\xi, \mathbf{v}), (\xi', \mathbf{v}') \in \mathbb{A} \times \mathfrak{Im}(\mathbb{A}). \tag{13.11}$$

This construction turns $\mathbb{A} \times \mathfrak{Im}(\mathbb{A})$, which is the parabolic boundary of the Siegel domain, into a Lie group, known as the *Heisenberg group over* $\mathbb{A}$ which we denote by $\mathcal{N}$ or $\mathcal{N}_1^{\mathbb{A}}$.

Proposition 13.3.1 *Formula* (13.11) *is a group product on* $\mathbb{A} \times \mathfrak{Im}(\mathbb{A})$.

Proof The set is closed under the operation, and for every $(\xi, \mathbf{u}) \in \mathbb{A} \times \mathfrak{Im}(\mathbb{A})$, we have

$$(\xi, \mathbf{u}) \cdot (0, 0) = (0, 0) \cdot (\xi, \mathbf{u}) = (\xi, \mathbf{u}).$$

Thus, the point $(0, 0)$ is the identity element. Moreover, for every $(\zeta, \mathbf{v}) \in \mathbb{A} \times \mathfrak{Im}(\mathbb{A})$, we have

$$(\zeta, \mathbf{v}) \cdot (-\zeta, -\mathbf{v}) = (\zeta - \zeta, \mathbf{v} - \mathbf{v} + 2\mathfrak{Im}(-|\zeta|^2)) = (0, 0) = (-\zeta, -\mathbf{v}) \cdot (\zeta, \mathbf{v}).$$

Hence, the inverse element $(\zeta, \mathbf{v})^{-1}$ is given by $(-\zeta, -\mathbf{v})$. Finally, associativity holds as well, since

$$
\begin{aligned}
((\xi, \mathbf{u}) \cdot (\zeta, \mathbf{v})) \cdot (\eta, \mathbf{w}) &= (\xi + \zeta, \mathbf{u} + \mathbf{v} + 2\Im\mathfrak{m}(\bar{\xi}\zeta)) \cdot (\eta, \mathbf{w}) \\
&= (\xi + \zeta + \eta, \mathbf{u} + \mathbf{v} + 2\Im\mathfrak{m}(\bar{\xi}\zeta) + \mathbf{w} + 2\Im\mathfrak{m}((\overline{\xi + \zeta})\eta)) \\
&= (\xi + \zeta + \eta, \mathbf{u} + \mathbf{v} + \mathbf{w} + 2\Im\mathfrak{m}(\bar{\xi}\zeta + \bar{\xi}\eta + \bar{\zeta}\eta)) \\
&= (\xi + \zeta + \eta, \mathbf{u} + \mathbf{v} + \mathbf{w} + 2\Im\mathfrak{m}(\bar{\zeta}\eta) + 2\Im\mathfrak{m}(\bar{\xi}(\zeta + \eta))) \\
&= (\xi, \mathbf{u}) \cdot (\zeta + \eta, \mathbf{v} + \mathbf{w} + 2\Im\mathfrak{m}(\bar{\zeta}\eta)) \\
&= (\xi, \mathbf{u}) \cdot ((\zeta, \mathbf{v}) \cdot (\eta, \mathbf{w})).
\end{aligned}
$$

□

It is possible to define a Lie group structure on $\mathbb{A}^n \times \Im\mathfrak{m}(\mathbb{A})$ similarly to (13.11) and define the n-th Heisenberg group $\mathcal{N}_n^{\mathbb{A}}$ over $\mathbb{A}$; see Exercise 13.4.13.

13.3.2 A Group Structure on $\mathbb{A} \times \Im\mathfrak{m}(\mathbb{A}) \times \mathbb{R}_+$

On the Heisenberg group $\mathcal{N} := (\mathbb{A} \times \Im\mathfrak{m}(\mathbb{A}), \cdot)$ with group product (13.11), we particularly identify a group homomorphism $\delta : \mathbb{R}_+ \to \mathrm{Aut}(\mathcal{N})$. This shall endow the set $\mathbb{A} \times \Im\mathfrak{m}(\mathbb{A}) \times \mathbb{R}_+$ with a group structure given by the semi-direct product $\mathcal{N} \rtimes_\delta \mathbb{R}_+$.

The *one-parameter subgroup of standard Heisenberg homotheties* is the homomorphism $\delta : \mathbb{R}_+ \to \mathrm{Aut}(\mathcal{N})$, where

$$\delta_\lambda(\xi, \mathbf{v}) := (\lambda\xi, \lambda^2\mathbf{v}), \qquad \forall\lambda \in \mathbb{R}_+, \forall(\xi, \mathbf{v}) \in \mathbb{A} \times \Im\mathfrak{m}(\mathbb{A}). \tag{13.12}$$

We call δ_λ the *Heisenberg homothety of ratio* λ.

Proposition 13.3.2 *The map* $\delta : \mathbb{R}_+ \to \mathrm{Aut}(\mathcal{N})$ *takes, indeed, values in* $\mathrm{Aut}(\mathcal{N})$ *and is a group homomorphism.*

Proof For every $\lambda \in \mathbb{R}_+$, we have that $\delta_\lambda \in \mathrm{Aut}(\mathcal{N})$, because δ_λ is bijective and for all $(\xi, \mathbf{u}), (\zeta, \mathbf{v}) \in \mathcal{N}$ we have

$$
\begin{aligned}
\delta_\lambda((\xi, \mathbf{u}) \cdot (\zeta, \mathbf{v})) &= \delta_\lambda((\xi + \zeta, \mathbf{u} + \mathbf{v} + 2\Im\mathfrak{m}(\bar{\xi}\zeta))) \\
&= (\lambda(\xi + \zeta), \lambda^2(\mathbf{u} + \mathbf{v} + 2\Im\mathfrak{m}(\bar{\xi}\zeta))) \\
&= (\lambda\xi + \lambda\zeta, \lambda^2\mathbf{u} + \lambda^2\mathbf{v} + 2\Im\mathfrak{m}(\lambda\bar{\xi}\lambda\zeta)) \\
&= \delta_\lambda(\xi, \mathbf{u}) \cdot \delta_\lambda(\zeta, \mathbf{v}).
\end{aligned}
$$

Furthermore, the map δ itself is a group homomorphism, since

$$\delta_{\lambda\lambda'} = \delta_\lambda \circ \delta_{\lambda'}, \qquad \forall\lambda, \lambda' \in \mathbb{R},$$

which is easy to check; see also Exercise 13.4.14. □

Thus, we perform the semi-direct product $\mathcal{N} \rtimes_\delta \mathbb{R}_+$; see Exercise 13.4.19.

13.3.3 Left Invariance of the Pullback of the Bergman Metric

From Theorem 13.2.1 we know that the hyperbolic space, in the form of the Siegel domain $\mathbb{D}$, can be modeled by $\mathbb{A} \times \Im\mathfrak{m}(\mathbb{A}) \times \mathbb{R}_+$. Next, we see this manifold as the Lie group $\mathcal{N} \rtimes_\delta \mathbb{R}_+$, where on $\mathcal{N} := \mathbb{A} \times \Im\mathfrak{m}(\mathbb{A})$ we put the group product (13.11) on which the group $\mathbb{R}_+$ acts by (13.12). On $\mathcal{N} \rtimes_\delta \mathbb{R}_+$, we put coordinates $(\xi, \mathbf{v}, \lambda) \in \mathbb{A} \times \Im\mathfrak{m}(\mathbb{A}) \times \mathbb{R}_+$. On this group, we consider the distance $\varphi^* d$ from Theorem 13.2.2.

Theorem 13.3.3 *The pullback $\varphi^* d$ of the Bergman metric d to the Lie group $\mathcal{N} \rtimes_\delta \mathbb{R}_+$ by the horospherical coordinate map φ is left-invariant.*

Proof In Theorem 13.2.2, we have shown that the distance $\varphi^* d$ between two points $(\xi, \mathbf{v}, \lambda)$ and $(\xi', \mathbf{v}', \lambda')$ in $\mathbb{A} \times \Im\mathfrak{m}(\mathbb{A}) \times \mathbb{R}_+ = \mathcal{N} \rtimes_\delta \mathbb{R}_+$ satisfies

$$\Big(\cosh\big(\varphi^* d((\xi', \mathbf{v}', \lambda'), (\xi'', \mathbf{v}'', \lambda''))\big)\Big)^2 = \frac{\big(|\xi' - \xi''|^2 + \lambda'^2 + \lambda''^2\big)^2 + |\mathbf{v}' - \mathbf{v}'' + 2\Im\mathfrak{m}(\bar{\xi}'\xi'')|^2}{4\lambda'^2\lambda''^2}.$$

Let $(\xi, \mathbf{v}, \lambda) \in \mathcal{N} \rtimes \mathbb{R}_+$. Using the basic definition of semi-direct product, as in Exercise 13.4.19, we get that the left translation is

$$\begin{aligned}
&\Big(\cosh\big(\varphi^* d((\xi, \mathbf{v}, \lambda) \cdot (\xi', \mathbf{v}', \lambda), (\xi, \mathbf{v}, \lambda) \cdot (\xi'', \mathbf{v}'', \lambda''))\big)\Big)^2 \\
&= \Bigg(\cosh\Bigg(\varphi^* d((\xi + \lambda\xi', \mathbf{v} + \lambda^2\mathbf{v}' + 2\Im\mathfrak{m}(\bar{\xi}\lambda\xi'), \lambda\lambda'), \\
&\qquad\qquad (\xi + \lambda\xi'', \mathbf{v} + \lambda^2\mathbf{v}'' + 2\Im\mathfrak{m}(\bar{\xi}\lambda\xi''), \lambda\lambda'')\Bigg)\Bigg)^2 \\
&= \frac{\big(|\xi + \lambda\xi' - (\xi + \lambda\xi'')|^2 + \lambda^2\lambda'^2 + \lambda^2\lambda''^2\big)^2}{4\lambda^2\lambda'^2\lambda^2\lambda''^2} \\
&\quad + \frac{1}{4\lambda^2\lambda'^2\lambda^2\lambda''^2}\Big|\mathbf{v} + \lambda^2\mathbf{v}' + 2\Im\mathfrak{m}(\bar{\xi}\lambda\xi') - (\mathbf{v} + \lambda^2\mathbf{v}'' + 2\Im\mathfrak{m}(\bar{\xi}\lambda\xi'')) \\
&\qquad + 2\Im\mathfrak{m}(\overline{(\xi + \lambda\xi')}(\xi + \lambda\xi''))\Big|^2
\end{aligned}$$

$$= \frac{\left(|\xi' - \xi'')|^2 + \lambda'^2 + \lambda''^2\right)^2}{4\lambda'^2\lambda''^2}$$

$$+ \frac{\left|\mathbf{v}' - \mathbf{v}'' + \frac{2}{\lambda^2}\Im\mathfrak{m}\Big(\bar{\xi}\lambda\xi' - \bar{\xi}\lambda\xi'' + \bar{\xi}\xi + \bar{\xi}\lambda\xi'' + \lambda\bar{\xi}'\xi + \lambda\bar{\xi}'\lambda\xi''\Big)\right|^2}{4\lambda'^2\lambda''^2}$$

$$= \frac{\left(|\xi' - \xi''|^2 + \lambda'^2 + \lambda''^2\right)^2 + |\mathbf{v}' - \mathbf{v}'' + 2\Im\mathfrak{m}(\bar{\xi}'\xi'')|^2}{4\lambda'^2\lambda''^2}$$

$$= \Big(\cosh\left(\varphi^* d((\xi', \mathbf{v}', \lambda'), (\xi'', \mathbf{v}'', \lambda''))\right)\Big)^2,$$

where we have used that $\bar{\xi}\lambda\xi' + \bar{\xi}\xi + \lambda\bar{\xi}'\xi = -|\xi|^2 + 2\Re\mathfrak{e}(\bar{\xi}\lambda\xi')$ has trivial imaginary part. □

13.4 Exercises

Exercise 13.4.1 Both the real numbers and the complex numbers, together with their standard norm, form a Euclidean Hurwitz algebra in the sense of Definition 13.1.1, which is associative and commutative.

Exercise 13.4.2 The quaternions, as defined at page 414, together with the norm $|z| = \sqrt{\bar{z}z}$ form a Euclidean Hurwitz algebra in the sense of Definition 13.1.1, which is associative but not commutative.

Exercise 13.4.3 In the quaternions, the conjugation satisfies anti-involution: $\overline{ab} = \bar{a}\bar{b}$, for all $a\ b \in \mathbb{H}$. Moreover, we have $\overline{a^{-1}} = (\overline{a})^{-1}$ and $|a|^2 = \Re\mathfrak{e}(a)^2 + |\Im\mathfrak{m}(a)|^2$, for all $a \in \mathbb{H}$.

Exercise 13.4.4 For every element ξ in a Euclidean Hurwitz algebra, we have $|\xi|^2 = \bar{\xi}\xi$, $2\Im\mathfrak{m}(\xi) = \xi - \bar{\xi}$, and $\bar{\xi} = -\xi + |\xi + 1|^2 - |\xi|^2 - 1$, where the latter may define the conjugate in terms of the norm.

Exercise 13.4.5 The Hermitian transpose of a product of two matrices over a Euclidean Hurwitz algebra is the product of the Hermitian transposes in the reverse order, that is, $(AB)^* = B^*A^*$.

Exercise 13.4.6 The eigenvalues of Hermitian transformations are real numbers.

Exercise 13.4.7 Because of (13.4), for every Hermitian form we have $\langle x, x\rangle \in \mathbb{R}$.

Exercise 13.4.8 From the definitions

$$\sinh x := \frac{e^x - e^{-x}}{2} \qquad \text{and} \qquad \cosh x := \frac{e^x + e^{-x}}{2}, \tag{13.13}$$

we obtain the general formula

$$\operatorname{arccosh}(x) = \log\left(x + \sqrt{x^2 - 1}\right), \qquad \forall x \geq 1, \tag{13.14}$$

and the general *half-argument formula* $\left(\sinh\left(\frac{x}{2}\right)\right)^2 = \frac{\cosh(x)-1}{2}$.

Exercise 13.4.9 (The $\mathbb{A}$-hyperbolic n-space $\mathbb{A}\mathbf{H}^n$) Let $\mathbb{A} \in \{\mathbb{R}, \mathbb{C}, \mathbb{H}\}$ and $n \in \mathbb{N}$. Let $\mathbb{P}(\mathbb{A}^{n+1})$ be the *n-dimensional* $\mathbb{A}$*-projective space*, defined as the quotient of $\mathbb{A}^{n+1} \setminus \{0\}$ by the equivalence relation that identifies $x = (x_1, \ldots, x_{n+1})$ with $x\lambda = (x_1\lambda, \ldots, x_{n+1}\lambda)$ for each $\lambda \in \mathbb{A} \setminus \{0\}$. The equivalence class of $x \in \mathbb{A}^{n+1} \setminus \{0\}$ is denoted by $\mathbb{P}(x)$. Fixing a Hermitian form $\langle\cdot,\cdot\rangle$ of signature $(n, 1)$, we define the $\mathbb{A}$*-hyperbolic n-space* as the set

$$\mathbb{A}\mathbf{H}^n := \{\mathbb{P}(x) \in \mathbb{P}(\mathbb{A}^{n+1}) : \langle x, x\rangle < 0\} \tag{13.15}$$

equipped with the distance d such that

$$d(\mathbb{P}(x), \mathbb{P}(y)) := \operatorname{arccosh}\sqrt{\frac{\langle x, y\rangle\langle y, x\rangle}{\langle x, x\rangle\langle y, y\rangle}}. \tag{13.16}$$

The set $\mathbb{A}\mathbf{H}^n$ is well defined, and the distance formula does not depend on the representative chosen.

Exercise 13.4.10 (Alternative Formulas for the Hyperbolic Distance in Horospherical Coordinates) Formula (13.10) has the equivalent formulations:

$$\operatorname{arccosh}\frac{\sqrt{\left(|\xi - \xi'|^2 + \lambda^2 + \lambda'^2\right)^2 + |\mathbf{v} - \mathbf{v}' + 2\Im\mathfrak{m}(\bar{\xi}\xi')|^2}}{2\lambda\lambda'}$$

and

$$\operatorname{arccosh}\sqrt{\left(1 + \frac{|\xi - \xi'|^2 + |\lambda - \lambda'|^2}{2\lambda\lambda'}\right)^2 + \left(\frac{|\mathbf{v} + \bar{\xi}\xi' - \mathbf{v}' - \bar{\xi}'\xi|}{2\lambda\lambda'}\right)^2},$$

from which one easily observes the symmetry.

Exercise 13.4.11 In the manifold $\mathbb{A} \times \Im\mathfrak{m}(\mathbb{A}) \times \mathbb{R}_+$, the distance (13.10) between $(0, 0, 1)$ and an arbitrary element $(\xi, \mathbf{v}, \lambda)$ is

$$\operatorname{arccosh} \frac{\sqrt{\left(|\xi|^2 + \lambda^2 + 1\right)^2 + |\mathbf{v}|^2}}{2\lambda}.$$

Exercise 13.4.12 Consider the Poincaré half-space model $\mathbb{R} \times \mathbb{R}_+$ of real-hyperbolic space $\mathbb{R}\mathbf{H}^2$. The distance satisfies:

$$\begin{aligned} d((\xi, \lambda), (\xi', \lambda')) &= \operatorname{arccosh}\left(\frac{|\xi - \xi'|^2 + \lambda^2 + \lambda'^2}{2\lambda\lambda'}\right) \\ &= 2\operatorname{arcsinh}\left(\frac{\|(\xi, \lambda) - (\xi', \lambda')\|}{2\sqrt{\lambda\lambda'}}\right), \quad \forall (\xi, \lambda), (\xi', \lambda') \in \mathbb{R} \times \mathbb{R}_+, \end{aligned}$$

where $\| \cdot \|$ is the standard Euclidean norm on $\mathbb{R}^2$. There are few other formulas in [Bea83, Theorem 7.2.1].
Hint. First, use Theorem 13.2.2 and the fact that the imaginary part in $\mathbb{R}$ is trivial. Then, recalling Exercise 13.4.8, use the general half-argument formula.

Exercise 13.4.13 (Heisenberg Groups Over $\mathbb{A}$) Let $\mathbb{A} \in \{\mathbb{R}, \mathbb{C}, \mathbb{H}\}$ and $n \in \mathbb{N}$. The n-th$\mathbb{A}$-*Heisenberg group* $\mathcal{N}_n^{\mathbb{A}}$ is the set $\mathbb{A}^n \times \Im\mathfrak{m}(\mathbb{A})$ endowed with the multiplication law

$$(\xi, \mathbf{v}) \cdot (\xi', \mathbf{v}') := \left(\xi + \xi', \mathbf{v} + \mathbf{v}' + 2\Im\mathfrak{m}(\bar{\xi} \cdot \xi')\right), \ \forall (\xi, \mathbf{v}), (\xi', \mathbf{v}') \in \mathbb{A}^n \times \Im\mathfrak{m}(\mathbb{A}), \tag{13.17}$$

where $\bar{\xi} \cdot \xi' = \bar{\xi}^t \xi'$ is the scalar product between the vector $\bar{\xi}$ and the vector ξ'. This indeed defines a Lie group structure, whose inverse satisfies $(\xi, s)^{-1} = (-\xi, -s)$, for all $(\xi, s) \in \mathcal{N}_n^{\mathbb{A}}$. The group $\mathcal{N}_n^{\mathbb{A}}$ is nilpotent.

Exercise 13.4.14 (Heisenberg Homotheties) Let $\mathbb{A} \in \{\mathbb{R}, \mathbb{C}, \mathbb{H}\}$ and $n \in \mathbb{N}$. On the group $\mathcal{N}_n^{\mathbb{A}}$ from Exercise 13.4.13, for each $\lambda \in \mathbb{R}$ define *Heisenberg homothety of ratio* λ by

$$\delta_\lambda(\xi, \mathbf{v}) := (\lambda\xi, \lambda^2 \mathbf{v}), \qquad \forall (\xi, \mathbf{v}) \in \mathbb{A}^n \times \Im\mathfrak{m}(\mathbb{A}). \tag{13.18}$$

The Heisenberg homothety satisfies the following properties:

13.4.14.i. $\delta_a((\xi, s)(v, t)) = \delta_a(\xi, s)\delta_a(v, t)$, for all $a \in \mathbb{R}$, $\xi, v \in \mathbb{A}^n$, and $s, t \in \Im\mathfrak{m}(\mathbb{A})$;
13.4.14.ii. $(\delta_a)^{-1} = \delta_{a^{-1}}$, for all $a \in \mathbb{R} \setminus \{0\}$;
13.4.14.iii. $\delta_{aa'} = \delta_a \circ \delta_{a'}$, for all $a, a \in \mathbb{R}$.

Exercise 13.4.15 Let $\mathbb{A} \in \{\mathbb{R}, \mathbb{C}, \mathbb{H}\}$ and $n \in \mathbb{N}$. On $\mathbb{A}^{n+1}$ consider the forms

$$\langle x, y \rangle := \langle x, y \rangle_K = x^* K y, \qquad \forall x, y \in \mathbb{A}^{n+1} \tag{13.19}$$

where K is either

$$J_1 := \begin{bmatrix} \mathbb{I}_n & 0 \\ 0 & -1 \end{bmatrix}, \text{ or } \begin{bmatrix} 0 & 0 & -1 \\ 0 & \mathbb{I}_{n-1} & 0 \\ -1 & 0 & 0 \end{bmatrix}, \text{ or } J_2 := \begin{bmatrix} 0 & 0 & 1 \\ 0 & \mathbb{I}_{n-1} & 0 \\ 1 & 0 & 0 \end{bmatrix}. \tag{13.20}$$

These forms are Hermitian forms of signature $(n, 1)$.

Exercise 13.4.16 (The Group $O_{\mathbb{A}}(n, 1)$) In the setting of Exercise 13.4.15, consider the group $\mathrm{GL}(n+1, \mathbb{A})$ of invertible $(n+1, n+1)$ matrices with coefficients in $\mathbb{A}$. Let $O_{\mathbb{A}}(n, 1) \subseteq \mathrm{GL}(n+1, \mathbb{A})$ be the set of matrices that preserve the form $\langle \cdot, \cdot \rangle$ given by (13.19), i.e.,

$$O_{\mathbb{A}}(n, 1) := \{A \in \mathrm{GL}(n+1, \mathbb{A}) \ : \ \langle Ax, Ay \rangle = \langle x, y \rangle, \ \forall x, y \in \mathbb{A}^{n,1}\}.$$

13.4.16.i. We have $A \in O_{\mathbb{A}}(n, 1) \Leftrightarrow A^* K A = K$, for the K in (13.20).
13.4.16.ii. The set $O_{\mathbb{A}}(n, 1)$ with the matrix multiplication is a Lie group.

Exercise 13.4.17 The induced action of $O_{\mathbb{A}}(n, 1)$ on the projective space $\mathbb{P}(\mathbb{A}^{n+1})$ preserves the subset $\mathbb{A}\mathbf{H}^n$, as defined in (13.15), and acts by isometries on $\mathbb{A}\mathbf{H}^n$.

Exercise 13.4.18 (Two Distinguished Subgroups: A and N) We denote by A and N the following subsets of the group $O_{\mathbb{A}}(n, 1)$:

- 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

$$\nu(M, M_{13}) := \begin{pmatrix} 1 & M & M_{13} \\ 0 & I_{n-1} & M^* \\ 0 & 0 & 1 \end{pmatrix}, \tag{13.21}$$

  where M is a $(1, n-1)$-matrix with elements in $\mathbb{A}$ and M_{13} is in $\mathbb{A}$ and satisfies

$$|M|^2 = M_{13} + \overline{M}_{13}.$$

13.4.18.i. 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

$$A(t)\nu(M, M_{13}) = \nu(e^t M, e^{2t} M_{13})A(t).$$

13.4.18.ii. The sets A, N, and NA are subgroups of $O_{\mathbb{A}}(n, 1)$;
13.4.18.iii. The group N is normal in NA;
13.4.18.iv. The group A is isomorphic to $\mathbb{R}$.
13.4.18.v. The group N is isomorphic to the Heisenberg group $\mathcal{N}_n^{\mathbb{A}}$, defined in Exercise 13.4.13.
13.4.18.vi. The group NA acts simply transitively on $\mathbb{A}\mathbf{H}^n$.

Exercise 13.4.19 If on $\mathcal{N} := \mathbb{A} \times \mathfrak{Im}(\mathbb{A})$ we consider the group product (13.11) and the action δ of $\mathbb{R}_+$ on $\mathcal{N}$ by (13.12), then the semi-direct product on $\mathcal{N} \rtimes_\delta \mathbb{R}_+$ is

$$(\xi, \mathbf{v}, \lambda) \cdot (\xi', \mathbf{v}', \lambda') = (\xi + \lambda\xi', \mathbf{v} + \lambda^2\mathbf{v}' + 2\mathfrak{Im}(\bar{\xi} \cdot \lambda\xi'), \lambda\lambda'),$$
$$\forall (\xi, \mathbf{v}, \lambda), (\xi, \mathbf{v}', \lambda') \in \mathbb{A} \times \mathfrak{Im}(\mathbb{A}) \times \mathbb{R}_+.$$

Hint. Use the definition (5.18).

Exercise 13.4.20 The Lie algebra of $\mathbb{C}\mathbf{H}^2$, seen as $\mathbb{C}^{n-1} \times i\mathbb{R} \times \mathbb{R}_+$ as in Exercise 13.4.19, is isomorphic to the one in Exercise 8.3.11.

Exercise 13.4.21 Every $\mathbb{A}$-hyperbolic n-space is a Riemannian manifold whose sectional curvature is at most -1.
Hint. For some evidence on the curvature, see Exercise 8.3.11.

Exercise 13.4.22 (Final Exercise) We have the following generalization of Theorem 13.0.1 to arbitrary dimensions. For $\mathbb{A} \in \{\mathbb{R}, \mathbb{C}, \mathbb{H}\}$ and $n \in \mathbb{N}$, the $\mathbb{A}$-hyperbolic n-space $\mathbb{A}\mathbf{H}^n$ is isometric to the manifold $\mathbb{A}^{n-1} \times \mathfrak{Im}(\mathbb{A}) \times \mathbb{R}_+$ equipped with the group product given by

$$(\xi, \mathbf{v}, \lambda) \cdot (\xi', \mathbf{v}', \lambda') = (\xi + \lambda\xi', \mathbf{v} + \lambda^2\mathbf{v}' + 2\mathfrak{Im}(\bar{\xi} \cdot \lambda\xi'), \lambda\lambda'),$$
$$\forall (\xi, \mathbf{v}, \lambda), (\xi, \mathbf{v}', \lambda') \in \mathbb{A}^{n-1} \times \mathfrak{Im}(\mathbb{A}) \times \mathbb{R}_+.$$

and, denoting by $\|\cdot\|$ the ℓ_2-norm on $\mathbb{A}^{n-1}$, the left-invariant distance d is such that

$$d(1, (\xi, \mathbf{v}, \lambda)) = \operatorname{arccosh} \frac{\sqrt{\left(\|\xi\|^2 + \lambda^2 + 1\right)^2 + |\mathbf{v}|^2}}{2\lambda},$$
$$\forall (\xi, \mathbf{v}, \lambda) \in \mathbb{A}^{n-1} \times \mathfrak{Im}(\mathbb{A}) \times \mathbb{R}_+.$$

Chapter 14
Heintze Groups and Their Visual Boundaries

In this chapter, we show that to every Riemannian symmetric space of rank one and noncompact type, one can associate a 'visual boundary' that has the structure of a Carnot group. Visual boundaries are associated with spaces with negative sectional curvature. In fact, every homogeneous negatively curved manifold has the structure of a semi-direct product of the form $N \rtimes \mathbb{R}$ for some positively graded nilpotent group N, which canonically represents the visual boundary. Hence, these boundaries are equipped with the structure of metric Lie groups that admit dilations.

14.1 CAT(-1) Spaces and Visual Boundaries

The $\mathbb{A}$-hyperbolic n-space $\mathbb{A}\mathbf{H}^n$ has been presented in Sect. 13.2 for $n = 2$ and in Exercise 13.4.9 for arbitrary $n \in \mathbb{N}$. This metric space can be seen as a Riemannian space whose sectional curvature is at most -1; see Exercise 13.4.21. From the metric viewpoint, we say that $\mathbb{A}\mathbf{H}^n$ is a CAT(-1) metric space in the sense of Definition 14.1.1. An explicit proof of this last statement can be found in [BH99, Part II, Chapter 10]. We shall proceed with applying the theory of CAT(-1) spaces. In particular, we shall recall the notion of visual boundary.

There are several equivalent definitions of CAT(-1). We recall just the most standard one and refer to [BH99] for more.

Definition 14.1.1 (CAT(-1) Space) A geodesic metric space M is said to be CAT(-1), or, *with generalized sectional curvature at most* -1, if the following

E. Le Donne, *Metric Lie Groups*, Graduate Texts in Mathematics 306,
https://doi.org/10.1007/978-3-031-98832-5_14

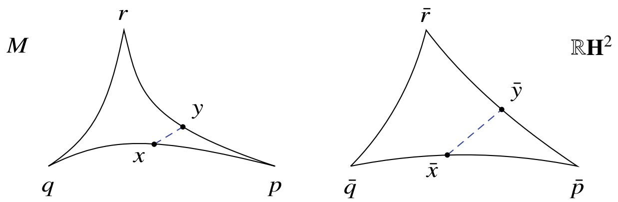


Fig. 14.1 Comparison of triangles for the CAT(-1) condition: The triangle on the left is a triangle in the CAT(-1) metric space M, and the triangle on the right is in the hyperbolic plane $\mathbb{R}\mathbf{H}^2$. All marked distances are pairwise equal except the dashed ones. The CAT(-1) condition requires that $d(x, y) \leq d(\bar{x}, \bar{y})$. One loosely says that the triangle on the left is *thiner* than its *comparison triangle*, i.e., the one on the right

comparison property holds (see Fig. 14.1): For all $p, q, r, x, y \in M$ and for all $\bar{p}, \bar{q}, \bar{r}, \bar{x}, \bar{y} \in \mathbb{R}\mathbf{H}^2$ such that

$$\begin{aligned} d(p, x) + d(x, q) = d(p, q), \quad & d(p, q) = d(\bar{p}, \bar{q}), \quad d(p, x) = d(\bar{p}, \bar{x}), \\ d(p, y) + d(y, r) = d(p, r), \quad & d(p, r) = d(\bar{p}, \bar{r}), \quad d(x, q) = d(\bar{x}, \bar{q}), \\ & d(r, q) = d(\bar{r}, \bar{q}), \quad d(p, y) = d(\bar{p}, \bar{y}), \\ & \qquad\qquad\qquad\quad d(y, r) = d(\bar{y}, \bar{r}), \end{aligned}$$

we have $d(x, y) \leq d(\bar{x}, \bar{y})$.

Definition 14.1.2 (Visual Boundary) Let M be a metric space. A *geodesic ray* in M is an isometric embedding $\gamma : [0, \infty) \to M$. Given two geodesic rays γ and σ in M, we write

$$\gamma \sim \sigma \overset{\text{def}}{\Longleftrightarrow} \sup\{d(\gamma(t), \sigma(t)) : t \in [0, \infty)\} \neq \infty,$$

and in this case, we say that γ and σ are *asymptotic*. Fix $o \in M$. The *visual boundary of M from o* is defined as the set of geodesic rays starting at o, up to asymptotic equivalence:

$$\partial_{\text{vis}} M := \partial_{\text{vis},o} M := \{\gamma : [0, \infty) \to M \ : \ \gamma \text{ geodesic ray }, \gamma(0) = o\}/\sim .$$

We refer to o as the *visual point* of the visual boundary.

For rank-one symmetric spaces, it is convenient to consider visual boundaries where the visual point is at infinity. In this case, we fix a point in the visual boundary, and we consider geodesic lines originating from that point at infinity.

Definition 14.1.3 (Parabolic Visual Boundary) Let M be a metric space. A *geodesic line* in M is an isometric embedding $\gamma : \mathbb{R} \to M$. Given two geodesic lines γ and σ in M (or two isometric embeddings $\gamma, \sigma : (-\infty, 0] \to M$), we say that

$$\gamma \text{ and } \sigma \text{ have the } \textit{same origin} \iff \lim_{t\to-\infty} d(\gamma(t), \sigma(t)) = 0, \tag{14.1}$$

while we say that they are *asymptotic* if the rays $\gamma|_{[0,\infty)}$ and $\sigma|_{[0,\infty)}$ are asymptotic. Fix a geodesic line $\omega : \mathbb{R} \to M$, or just an isometric embedding $\omega : (-\infty, 0] \to M$. The *parabolic visual boundary* of M *from* ω is defined as:

$$\partial_{\mathrm{Par}} M := \partial_{\mathrm{Par},\omega} M := \{\gamma : \text{ geodesic line in } M \text{ with the same origin as } \omega\}/\sim .$$

We refer to ω as the *visual point* of the parabolic visual boundary.

Visual boundaries can be naturally metrized, at least for CAT(-1) spaces. Something similar holds for Gromov hyperbolic spaces, which we will not discuss here; see [BS07, Proposition 3.3.3]. The construction of the parabolic boundary comes from the work of [HP97, p. 383]. Similar results on the metrization are also in the work of Hamenstädt in [Ham89], and for this reason, the visual distances, as we soon define, are also called *Hamenstädt distances*.

Definition 14.1.4 (Visual Distance) Let M be a metric space. Let γ and σ be geodesic lines in M. The *Gromov product* between γ and σ is defined as

$$(\gamma \mid \sigma)_\infty := \lim_{t\to+\infty} \left(t - \frac{1}{2} d(\gamma(t), \sigma(t))\right), \tag{14.2}$$

when the limit exists. In this case, the *visual distance* between γ and σ is defined as

$$\rho(\gamma, \sigma) := d_{\mathrm{vis}}(\gamma, \sigma) := e^{-(\gamma|\sigma)_\infty}. \tag{14.3}$$

The proof that when M is a CAT(-1) space, the limit in the Gromov product exists and that the visual distance as defined in (14.3) is a distance on $\partial_{\mathrm{Par},\omega} M$ will be discussed in the next subsection. The main reference for the following presentation is [Bou95].

14.1.1 Hamenstädt Distance

We adopt Hamenstädt's viewpoint of horospherical distances via Busemann functions associated with geodesic lines. The aim is to prove that visual distances on parabolic boundaries of CAT(-1) spaces satisfy the triangle inequality.

Let M be a metric space. Fix $\omega : \mathbb{R} \to M$ a geodesic line in M. The *Busemann function* associated with ω is the function $b_\omega : M \to \mathbb{R}$ defined as the existing limit

$$b_\omega(x) := \lim_{t\to-\infty} \big(d\big(\omega(t), x\big) + t\big), \qquad \forall x \in M; \tag{14.4}$$

see Exercise 14.4.1. In (14.4), the chosen signs and the fact that the limit is to $-\infty$ may not coincide with the choices of other authors. The viewpoint that we take here is that the origin of ω is where b_ω equals $-\infty$, and it is where all the geodesic lines come from when we see the visual boundary from ω. In fact, if γ is a geodesic line in M with the same origin as ω, then

$$\begin{aligned} b_\omega(\gamma(t)) &= \lim_{s\to-\infty} \big(d\big(\omega(s), \gamma(t)\big) + s\big) \\ &\simeq \lim_{s\to-\infty} \big(d(\gamma(s), \gamma(t)) \pm d(\omega(s), \gamma(s)) + s\big) \\ &= \lim_{s\to-\infty} \big(t - s \pm 0 + s\big) = t, \end{aligned} \tag{14.5}$$

where we used the symbol $\simeq$ to mean that we have upper and lower bounds according to the sign chosen as a consequence of the triangle inequality.

Set $(x, y)_b := \frac{1}{2}(b(x) + b(y) - d(x, y))$ to be the *Gromov product* between x and y with respect to a Busemann function $b := b_\omega$. The Gromov product is defined for $x, y \in M$, but will actually extend to $\partial_{\mathrm{Par},\omega} M$. We call $\rho_b(\cdot,\cdot) := \exp(-(\cdot,\cdot)_b)$ the *Hamenstädt function* with respect to b. Notice that we have generalized (14.2) since if γ and σ have the same origin as ω, then

$$(\,\gamma(t)\,|\,\sigma(t)\,)_b = \frac{1}{2}(b(\gamma(t)) + b(\sigma(t)) - d(\gamma(t), \sigma(t))) \overset{(14.5)}{=} t - \frac{1}{2} d(\gamma(t), \sigma(t)).$$

Proposition 14.1.5 (Bourdon) *Let M be a* CAT(-1) *space. Fix ω a geodesic line and b the Busemann function associated with ω. Then, the Hamenstädt function ρ_b is a distance function on $\partial_{\mathrm{Par},\omega} M$. In fact, if for every $t \in \mathbb{R}$ we consider the function*

$$\alpha_t : (x, y) \in M \times M \longmapsto \alpha_t(x, y) := \frac{2}{e^t} \sinh\left(\frac{d(x, y)}{2}\right),$$

then we have:

14.1.5.i *As t goes to ∞, the function α_t converges to ρ_b, in the sense that*

$$\lim_{t\to\infty} \alpha_t(\gamma(t), \sigma(t)) = \rho_b(\gamma, \sigma), \qquad \forall \gamma, \sigma \in \partial_{\mathrm{Par},\omega} M.$$

14.1.5.ii *The function α_t is a distance on the set $b^{-1}(t)$.*

Proof The following argument is an adaptation of [Bou95, Théorème 2.5.1]. Referring to Exercise 14.4.2, we begin by stressing that if $x, x' \in M$ and $z, z' \in \mathbb{C}$ are such that $\mathfrak{Im}(z) = \mathfrak{Im}(z') = ie^{-t}$ and $d(z, z') = d(x, x')$, then

$$\alpha_t(x, x') = |z - z'|. \tag{14.6}$$

Proof of 14.1.5.i. Take geodesic lines γ and γ' that have the same origin as ω. We spell out the definitions:

$$\begin{aligned}\alpha_t(\gamma(t), \gamma'(t)) &\stackrel{\text{def}}{=} 2e^{-t} \sinh\left(\frac{d(\gamma(t), \gamma'(t))}{2}\right)\\ &\stackrel{\text{def}}{=} 2e^{-t}\frac{1}{2}\left(e^{d(\gamma(t),\gamma'(t))/2} - e^{-d(\gamma(t),\gamma'(t))/2}\right),\end{aligned}$$

where we used the definition of α_t and the one of sinh. Then, since the function e^{-d} is bounded by 1 and $\lim_{t\to+\infty} e^{-t} = 0$, the above function has the same limit as

$$\begin{aligned}e^{-t}e^{d(\gamma(t),\gamma'(t))/2} &= \exp(-(t - d(\gamma(t), \gamma'(t))/2))\\ &\to \rho_b(\gamma, \gamma').\end{aligned}$$

Proof of 14.1.5.ii. We shall take $(0, e^{-t}) \in \mathbb{C}$ as comparison point for $\gamma(t)$, as $t \in \mathbb{R}$. Fix x, x', x'' points of the form $x = \gamma(t), x' = \gamma'(t), x'' = \gamma''(t)$ such that $\gamma, \gamma', \gamma''$ are geodesic lines with the same origin as ω. Take $z, z', z'' \in \mathbb{C}$ so that $\mathfrak{Im}(z) = \mathfrak{Im}(z') = \mathfrak{Im}(z'') = ie^{-t}$, $d(z, z') = d(x, x')$, $d(z', z'') = d(x', x'')$, and $\mathfrak{Re}(z) \leq \mathfrak{Re}(z') \leq \mathfrak{Re}(z'')$; see Fig. 14.2 for a visual representation. (N.B. the triangles z, z', ∞ and z', z'', ∞ are comparison triangles for x, x', ω and x', x'', ω, respectively. We say that these triangles are *ideal* since one of the points is at infinity)

Let η be the geodesic between z and z''. Let z^* be the point in η with $Re(z^*) = Re(z')$. Let x^* in M be on the geodesic γ' from ω to x' with $d(x', x^*) = d(z', z^*)$. Then, we bound using the CAT(− 1) comparison property, together with the fact that z^* is along the geodesic η and the triangle inequality:

$$\begin{aligned}d(z, z'') &\stackrel{\eta\text{ geodesic}}{=} d(z, z^*) + d(z^*, z'')\\ &\stackrel{\text{CAT}(-1)}{\geq} d(x, x^*) + d(x^*, x'')\\ &\stackrel{\text{triang. ineq.}}{\geq} d(x, x'').\end{aligned} \tag{14.7}$$

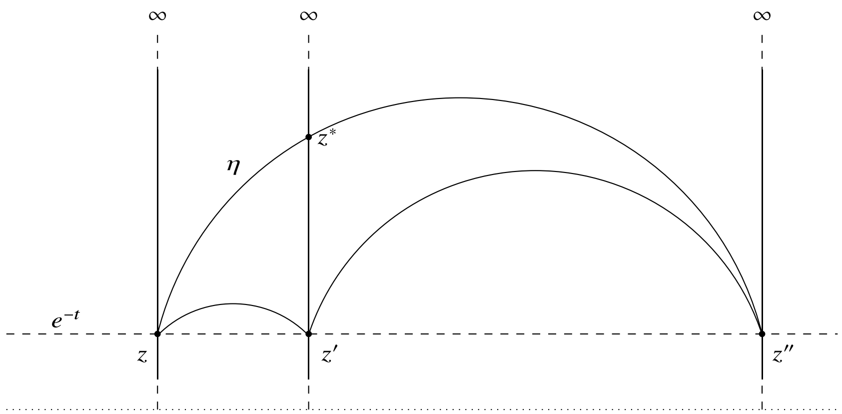


Fig. 14.2 The construction of two adjacent ideal triangles for the proof of Proposition 14.1.5.ii. The full curves are geodesic. The dashed horizontal line marks the height. The dotted horizontal line represents the parabolic visual boundary, which is the visual boundary minus a point, denoted by ∞

By monotonicity of sinh we deduce

$$\begin{aligned}\alpha_t(x, x'') &\overset{\text{def}}{=} \frac{2}{e^t}\sinh\left(\frac{d(x, x'')}{2}\right)\\ &\overset{(14.7)}{\leq} \frac{2}{e^t}\sinh\left(\frac{d(z, z'')}{2}\right)\\ &\overset{(14.6)}{=} |z - z''|\\ &= |z - z'| + |z' - z''|\\ &\overset{(14.6)}{=} \alpha_t(x, x') + \alpha_t(x', x''),\end{aligned}$$

where we used that by construction $\mathfrak{Re}(z')$ is between $\mathfrak{Re}(z)$ and $\mathfrak{Re}(z'')$. Hence, the function α_t is a distance function.

End of proof of Proposition 14.1.5 Take γ, γ', and $\gamma'' \in \partial_{\mathrm{Par},\omega} M$. Passing to the limit in the triangle inequality, given by 14.1.5.ii,

$$\alpha_t(\gamma(t), \gamma''(t)) \leq \alpha_t(\gamma(t), \gamma'(t)) + \alpha_t(\gamma'(t), \gamma''(t)),$$

we obtain, by 14.1.5.i,

$$\rho_b(\gamma, \gamma'') \leq \rho_b(\gamma, \gamma') + \rho_b(\gamma', \gamma'').$$

Then, also the function ρ_b satisfies the triangle inequality. □

14.2 The Visual Boundaries of $\mathbb{A}$-Hyperbolic Spaces

14.2.1 The Boundaries of $\mathbb{A}\mathbf{H}^n$

Let $\mathbb{A}$ be either $\mathbb{R}$, $\mathbb{C}$, or the quaternions $\mathbb{H}$. Let $n \in \mathbb{N}$. Thanks to the way we defined the $\mathbb{A}$-hyperbolic n-space $\mathbb{A}\mathbf{H}^n$, in Sect. 13.2 and then in Exercise 13.4.9, there is a natural way to represent its visual boundary. Denoting by $\mathbb{A}^{n,1}$ the space $\mathbb{A}^{n+1}$ equipped with a Hermitian form $\langle \cdot, \cdot \rangle$ of signature $(n, 1)$, the set of points in the $\mathbb{A}$-hyperbolic n-space is

$$\mathbb{A}\mathbf{H}^n = \{\mathbb{P}(x) \in \mathbb{P}(\mathbb{A}^{n,1}) \, : \, \langle x, x\rangle < 0\},$$

where $\mathbb{P}(\mathbb{A}^{n,1})$ is the $\mathbb{A}$-projective space of $\mathbb{A}^{n,1}$.

As in Sect. 13.2.2, we fix the Hermitian form (13.7) to have the Siegel domain $\mathbb{D}$, (13.8), with the correspondence of Theorem 13.2.1. The map φ generalizes to arbitrary n. We summarize this correspondence with the following diagram:

$$\mathbb{A}^{n-1} \times \mathfrak{Im}(\mathbb{A}) \times \mathbb{R}_+ \xrightarrow{(\varphi,1)} \mathbb{D} \times \{1\} \xrightarrow{\mathbb{P}} \mathbb{A}\mathbf{H}^n \subseteq \mathbb{P}(\mathbb{A}^{n,1}), \qquad \hat{\varphi} : \mathbb{A}^{n-1} \times \mathfrak{Im}(\mathbb{A}) \times \mathbb{R}_+ \to \mathbb{A}\mathbf{H}^n$$

defining the composition

$$\begin{aligned} \hat{\varphi}(\xi, \mathbf{v}, \lambda) \; &:= \; \mathbb{P}(\varphi(\xi, \mathbf{v}, \lambda), 1) \\ &\overset{(13.9)}{=} \mathbb{P}\left(-|\xi|^2 - \lambda^2 - \mathbf{v}, \, \sqrt{2}\xi, 1\right), \\ &\qquad\qquad \forall (\xi, \mathbf{v}, \lambda) \in \mathbb{A}^{n-1} \times \mathfrak{Im}(\mathbb{A}) \times \mathbb{R}_+. \end{aligned} \tag{14.8}$$

In $\mathbb{P}(\mathbb{A}^{n,1})$ we denote by ω the point $\omega := \mathbb{P}(1, \mathbf{0}, 0)$, where here with $\mathbf{0}$ we indicate the zero in $\mathbb{A}^{n-1}$. Obviously, the point ω belongs to the topological boundary of $\mathbb{A}\mathbf{H}^n$ within $\mathbb{P}(\mathbb{A}^{n,1})$. We shall see that there is a correspondence between the topological boundary of $\mathbb{A}\mathbf{H}^n$ and the visual boundary of $\mathbb{A}\mathbf{H}^n$. Moreover, with this identification, the parabolic boundary of $\mathbb{A}\mathbf{H}^n$ from (a geodesic line representing) ω is in correspondence with $\mathbb{A}^{n-1} \times \mathfrak{Im}(\mathbb{A}) \times \{0\}$. With this purpose, we consider the following family of curves.

For every $\xi \in \mathbb{A}^{n-1}$ and $\mathbf{v} \in \mathfrak{Im}(\mathbb{A})$, we consider the curve

$$\begin{aligned} \gamma_{(\xi,\mathbf{v})} : \mathbb{R} &\longrightarrow \mathbb{A}^{n-1} \times \mathfrak{Im}(\mathbb{A}) \times \mathbb{R}_+ \\ t &\longmapsto \gamma_{(\xi,\mathbf{v})}(t) := (\xi, \mathbf{v}, e^{-t}). \end{aligned} \tag{14.9}$$

See Fig. 14.3 for a visual representation.

Lemma 14.2.1 *The curves defined in* (14.9) *satisfy the following properties:*

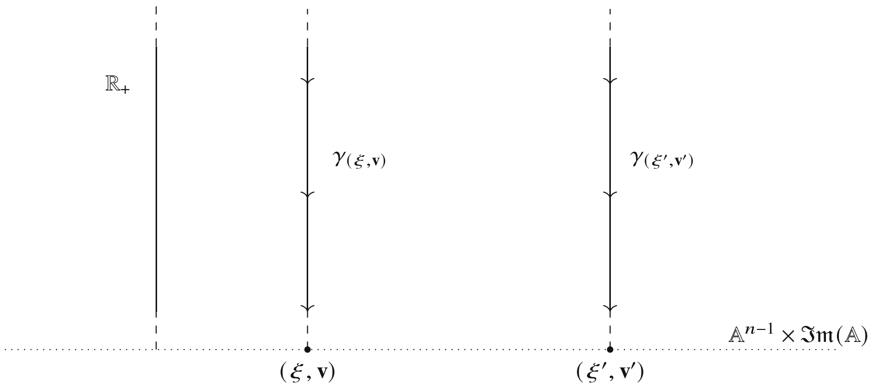


Fig. 14.3 The vertical geodesics in the $\mathbb{A}$-hyperbolic n-space $\mathbb{A}\mathbf{H}^n$ seen as $\mathbb{A}^{n-1} \times \mathfrak{Im}(\mathbb{A}) \times \mathbb{R}_+$ in horospherical coordinates. Every point $(\xi, \mathbf{v})$ in $\mathbb{A}^{n-1} \times \mathfrak{Im}(\mathbb{A})$ is associated with the geodesic line $\gamma_{(\xi,\mathbf{v})} := (\xi, \mathbf{v}, e^{-t})$, whose height decreases when time increases. The dotted horizontal line is $\mathbb{A}^{n-1} \times \mathfrak{Im}(\mathbb{A})$, which will represent the parabolic visual boundary

14.2.1.i. *each $\gamma_{(\xi,\mathbf{v})}$ is a geodesic;*
14.2.1.ii. *every two $\gamma_{(\xi,\mathbf{v})}$ and $\gamma_{(\xi',\mathbf{v}')}$ have the same origin;*
14.2.1.iii. *as $t \to -\infty$, the value $\hat{\varphi}(\gamma_{(\xi,\mathbf{v})}(t))$ tends to ω within $\mathbb{P}(\mathbb{A}^{n,1})$;*
14.2.1.iv. *as $t \to \infty$, the value $\gamma_{(\xi,\mathbf{v})}(t)$ tends to $(\xi, \mathbf{v}, 0)$ in the topology of $\mathbb{A}^{n-1} \times \mathfrak{Im}(\mathbb{A}) \times [0, +\infty)$.*

Proof Fix $(\xi, \mathbf{v}) \in \mathbb{A}^{n-1} \times \mathfrak{Im}(\mathbb{A})$. To show that $\gamma_{(\xi,\mathbf{v})}$ is a geodesic, we directly use the formula (13.10) for the distance, for $s, t \in \mathbb{R}$:

$$\begin{aligned}
\cosh d(\gamma_{(\xi,\mathbf{v})}(t), \gamma_{(\xi,\mathbf{v})}(s)) &\overset{(14.9)}{=} \cosh d((\xi, \mathbf{v}, e^{-t}), (\xi, \mathbf{v}, e^{-s})) \\
&\overset{(13.10)}{=} \sqrt{\left(1 + \frac{|e^{-t} - e^{-s}|^2}{2e^{-t-s}}\right)^2 + 0} \\
&= 1 + \frac{|e^{-t} - e^{-s}|^2}{2e^{-t-s}} \\
&= 1 + \frac{e^{t+s}}{2}(e^{-2t} - 2e^{-t-s} + e^{-2s}) \\
&= 1 + \frac{e^{s-t}}{2} - 1 + \frac{e^{t-s}}{2} \\
&= \frac{e^{s-t} + e^{t-s}}{2} \\
&\overset{(13.13)}{=} \cosh(s - t).
\end{aligned}$$

We deduce 14.2.1.i, thanks to the injectivity of cosh on the nonnegative real numbers.

Regarding 14.2.1.ii, according to the definition in (14.1) we need to prove that

$$\lim_{t\to-\infty} d(\gamma_{(\xi,\mathbf{v})}(t), \gamma_{(\xi',\mathbf{v}')}(t)) = 0, \qquad \forall(\xi,\mathbf{v}), (\xi',\mathbf{v}') \in \mathbb{A}^{n-1} \times \mathfrak{Im}(\mathbb{A}).$$

We calculate:

$$\begin{aligned} d(\gamma_{(\xi,\mathbf{v})}(t), \gamma_{(\xi',\mathbf{v}')}(t)) &\overset{(14.9)}{=} d((\xi, \mathbf{v}, e^{-t}), (\xi', \mathbf{v}', e^{-t})) \\ &\overset{(13.10)}{=} \operatorname{arccosh}\sqrt{\left(1 + \frac{|\xi-\xi'|^2}{2e^{-2t}}\right)^2 + \left(\frac{|\mathbf{v}-\mathbf{v}' + 2\mathfrak{Im}(\bar{\xi}\xi')|}{2e^{-2t}}\right)^2} \quad (14.10) \\ &\overset{t\to-\infty}{\longrightarrow} \operatorname{arccosh}(1) = 0. \end{aligned}$$

Regarding 14.2.1.iii, by the definition of $\hat{\varphi}$ in (14.8), we have, as $t \to -\infty$, that

$$\begin{aligned} \hat{\varphi}(\gamma_{(\xi,\mathbf{v})}(t)) &\overset{(14.9)}{=} \hat{\varphi}(\xi, \mathbf{v}, e^{-t}) \\ &\overset{(14.8)}{=} \mathbb{P}\left(-|\xi|^2 - e^{-2t} - \mathbf{v},\ \sqrt{2}\xi, 1\right) \\ &= \mathbb{P}\left(e^{2t}|\xi|^2 + 1 + e^{2t}\mathbf{v},\ -e^{2t}\sqrt{2}\xi, -e^{2t}\right) \longrightarrow \mathbb{P}(1, \mathbf{0}, 0) \overset{\text{def}}{=} \omega. \end{aligned}$$

Regarding 14.2.1.iv, we clearly have $\gamma_{(\xi,\mathbf{v})}(t) \overset{\text{def}}{=} (\xi, \mathbf{v}, e^{-t}) \to (\xi, \mathbf{v}, 0)$, as $t \to \infty$. □

Definition 14.2.2 Let $\mathbb{A} \in \{\mathbb{R}, \mathbb{C}, \mathbb{H}\}$ and $n \in \mathbb{N}$. We define the *boundary at infinity of* $\mathbb{A}\mathbf{H}^n$ as the set

$$\partial_\infty \mathbb{A}\mathbf{H}^n := \{\mathbb{P}(x) \in \mathbb{P}(\mathbb{A}^{n,1}) \ :\ \langle x, x\rangle = 0\}.$$

We arrived at the point where $\partial_\infty \mathbb{A}\mathbf{H}^n \setminus \{\omega\}$ can be identified with $\mathbb{A}^{n-1} \times \mathfrak{Im}(\mathbb{A}) \times \{0\}$, and hence with $\mathbb{A}^{n-1} \times \mathfrak{Im}(\mathbb{A})$. Recall that the latter may be endowed with the product from (13.17):

$$(\xi, \mathbf{v}; 0) \cdot (\xi', \mathbf{v}'; 0) := \left(\xi + \xi', \mathbf{v} + \mathbf{v}' + 2\mathfrak{Im}(\bar{\xi} \cdot \xi'); 0\right), \qquad \forall(\xi,\mathbf{v}), (\xi',\mathbf{v}') \in \mathbb{A}^{n-1} \times \mathfrak{Im}(\mathbb{A}),$$

which is actually isomorphic to $\mathbb{A}^{n-1} \times \mathfrak{Im}(\mathbb{A}) \times \{1\}$ inside $\mathbb{A}^{n-1} \times \mathfrak{Im}(\mathbb{A}) \times \mathbb{R}_+ = \mathcal{N}_n^{\mathbb{A}} \rtimes_\delta \mathbb{R}_+$, with its semi-direct product; see Exercise 13.4.19.

14.2.2 The Visual Distance on the Boundary of $\mathbb{A}\mathbf{H}^n$

Our next aim is to explicitly write the visual distance for the $\mathbb{A}$-hyperbolic n-space $\mathbb{A}\mathbf{H}^n$ in the new coordinates.

Definition 14.2.3 (Visual Distance on the n-th $\mathbb{A}$-Heisenberg Group) Let $\mathbb{A} \in \{\mathbb{R}, \mathbb{C}, \mathbb{H}\}$ and $n \in \mathbb{N}$. For every pair $(\xi, \mathbf{v})$ and $(\xi', \mathbf{v}') \in \mathbb{A}^{n-1} \times \mathfrak{Im}(\mathbb{A})$, we define the *visual distance* as

$$d\left((\xi, \mathbf{v}), (\xi', \mathbf{v}')\right) := d_{\text{vis}}\left(\gamma_{(\xi,\mathbf{v})}, \gamma_{(\xi',\mathbf{v}')}\right), \tag{14.11}$$

where $\gamma_{(\xi,\mathbf{v})}$ is defined in (14.9) and d_{vis} is defined in (14.3).

We finally prove that the parabolic visual boundaries of hyperbolic spaces are metric Lie groups:

Theorem 14.2.4 *Let $\mathbb{A} \in \{\mathbb{R}, \mathbb{C}, \mathbb{H}\}$ and $n \in \mathbb{N}$. On $\mathbb{A}^{n-1} \times \mathfrak{Im}(\mathbb{A})$, the visual distance reads as*

$$d\left((\xi, \mathbf{v}), (\xi', \mathbf{v}')\right) = \sqrt[4]{\|\xi - \xi'\|^4 + \left|\mathbf{v} - \mathbf{v}' + 2\mathfrak{Im}(\bar{\xi}\xi')\right|^2}, \quad \forall (\xi, \mathbf{v}), (\xi', \mathbf{v}') \in \mathbb{A}^{n-1} \times \mathfrak{Im}(\mathbb{A}), \tag{14.12}$$

it is left-invariant with respect to the group structure (13.17)*, and it is self-similar with respect to the dilations* (13.18).

Proof We start by substituting the definitions (14.11), (14.3), and (14.2):

$$\begin{aligned} d\left((\xi, \mathbf{v}), (\xi', \mathbf{v}')\right) &\overset{(14.11)}{=} d_{\text{vis}}\left(\gamma_{(\xi,\mathbf{v})}, \gamma_{(\xi',\mathbf{v}')}\right) \\ &\overset{(14.3)}{=} e^{-\left(\gamma_{(\xi,\mathbf{v})} | \gamma_{(\xi',\mathbf{v}')}\right)_\infty} \\ &\overset{(14.2)}{=} \lim_{t\to\infty} \exp\left(\frac{1}{2} d(\gamma_{(\xi,\mathbf{v})}(t), \gamma_{(\xi',\mathbf{v}')}(t)) - t\right). \end{aligned} \tag{14.13}$$

Therefore, we calculate the distance between the points along the curves, as defined in (14.9), as we did in (14.10) using (13.10):

$$\begin{aligned} &d(\gamma_{(\xi,\mathbf{v})}(t), \gamma_{(\xi',\mathbf{v}')}(t)) \\ &\overset{(14.10)}{=} \operatorname{arccosh}\sqrt{\left(1 + \frac{e^{2t}|\xi - \xi'|^2}{2}\right)^2 + \left(\frac{e^{2t}|\mathbf{v} - \mathbf{v}' + 2\mathfrak{Im}(\bar{\xi}\xi')|}{2}\right)^2} \\ &= \operatorname{arccosh}\left[\frac{e^{2t}}{2}\sqrt{\left(2e^{-2t} + |\xi - \xi'|^2\right)^2 + \left(|\mathbf{v} - \mathbf{v}' + 2\mathfrak{Im}(\bar{\xi}\xi')|\right)^2}\right]. \end{aligned}$$

We denote by $\beta(t)$ the value

$$\beta(t) := \left(2e^{-2t} + |\xi - \xi'|^2\right)^2 + \left(|\mathbf{v} - \mathbf{v}' + 2\Im\mathfrak{m}(\bar{\xi}\xi')|\right)^2, \tag{14.14}$$

so to continue, and using (13.14),

$$\begin{aligned} d(\gamma_{(\xi,\mathbf{v})}(t), \gamma_{(\xi',\mathbf{v}')}(t)) &= \operatorname{arccosh}\left(\frac{e^{2t}}{2}\sqrt{\beta(t)}\right) \\ &\overset{(13.14)}{=} \log\left(\frac{e^{2t}}{2}\sqrt{\beta(t)} + \sqrt{\frac{e^{4t}}{4}\beta(t) - 1}\right) \\ &= \log\left[e^{2t}\left(\frac{\sqrt{\beta(t)}}{2} + \sqrt{\frac{\beta(t)}{4} - e^{-4t}}\right)\right] \\ &= 2t + \log\left(\frac{\sqrt{\beta(t)}}{2} + \sqrt{\frac{\beta(t)}{4} - e^{-4t}}\right). \end{aligned}$$

Substituting the last equation in (14.13), we obtain

$$\begin{aligned} d\left((\xi, \mathbf{v}), (\xi', \mathbf{v}')\right) &= \lim_{t\to\infty} \exp\left(\frac{1}{2}\log\left(\frac{\sqrt{\beta(t)}}{2} + \sqrt{\frac{\beta(t)}{4} - e^{-4t}}\right)\right) \\ &= \lim_{t\to\infty} \sqrt{\frac{\sqrt{\beta(t)}}{2} + \sqrt{\frac{\beta(t)}{4} - e^{-4t}}} \\ &= \sqrt[4]{\beta(\infty)} \\ &= \sqrt[4]{\left(|\xi - \xi'|^2\right)^2 + \left(|\mathbf{v} - \mathbf{v}' + 2\Im\mathfrak{m}(\bar{\xi}\xi')|\right)^2}, \end{aligned}$$

where we took the limit of (14.14) as $t \to \infty$. Therefore, we obtained (14.12). The left-invariance and self-similarity are straightforward exercises. □

We refer to (14.12) as the *Cygan-Koranyi distance* on the n-th 𝔸-Heisenberg group.

Proposition 14.2.5 *The Cygan-Koranyi function on* $\mathbb{A}^{n-1} \times \Im\mathfrak{m}(\mathbb{A})$*, given by* (14.12)*, satisfies the triangle inequality; hence it is a distance function.*

Proof The Cygan-Koranyi function is the visual parabolic distance function of the CAT(− 1) space $\mathbb{A}\mathbf{H}^n$. Therefore, by Bourdon's Proposition 14.1.5, it is a distance function.

Still, we will give a more direct argument next. Because of left invariance, it is enough to prove the triangle inequality when the middle point is the identity element

$(0, \mathbf{0})$. Namely, we need to prove that for every $(\xi, \mathbf{v}), (\xi', \mathbf{v}') \in \mathbb{A}^{n-1} \times \Im\mathfrak{m}(\mathbb{A})$ we have

$$d\left((\xi, \mathbf{v}), (\xi', \mathbf{v}')\right) \leq d\left((\xi, \mathbf{v}), (0, \mathbf{0})\right) + d\left((0, \mathbf{0}), (\xi', \mathbf{v}')\right).$$

To do this, we rewrite the terms and bound using several times that the imaginary numbers are orthogonal to the real numbers:

$$\begin{aligned}
d\left((\xi, \mathbf{v}), (\xi', \mathbf{v}')\right)^2 &= \sqrt{\|\xi - \xi'\|^4 + \left|\mathbf{v} - \mathbf{v}' + 2\Im\mathfrak{m}(\bar{\xi}\xi')\right|^2} \\
&= \left|\left\|\xi - \xi'\right\|^2 - \left(\mathbf{v} - \mathbf{v}' + 2\Im\mathfrak{m}(\bar{\xi}\xi')\right)\right| \\
&= \left|\|\xi\|^2 - \bar{\xi}\xi' - \xi\bar{\xi}' + \left\|\xi'\right\|^2 - \mathbf{v} + \mathbf{v}' - \bar{\xi}\xi' + \xi\bar{\xi}'\right| \\
&= \left|\|\xi\|^2 - \mathbf{v} - 2\bar{\xi}\xi' + \left\|\xi'\right\|^2 + \mathbf{v}'\right| \\
&\leq \left|\|\xi\|^2 - \mathbf{v}\right| + \left|2\bar{\xi}\xi'\right| + \left|\left\|\xi'\right\|^2 + \mathbf{v}'\right| \\
&= \sqrt{\|\xi\|^4 + |\mathbf{v}|^2} + 2\,\|\xi\|\,\left\|\xi'\right\| + \sqrt{\|\xi'\|^4 + |\mathbf{v}'|^2} \\
&\leq \left(d\left((\xi, \mathbf{v}), (0, \mathbf{0})\right)\right)^2 + 2d\left((\xi, \mathbf{v}), (0, \mathbf{0})\right) d\left((0, \mathbf{0}), (\xi', \mathbf{v}')\right) \\
&\qquad + \left(d\left((0, \mathbf{0}), (\xi', \mathbf{v}')\right)\right)^2 \\
&= \left(d\left((\xi, \mathbf{v}), (0, \mathbf{0})\right) + d\left((0, \mathbf{0}), (\xi', \mathbf{v}')\right)\right)^2,
\end{aligned}$$

where in the second inequality, we used that the distance (14.27) is increasing in $|v|$. □

14.3 Heintze Groups

A well-known motivation for the study of self-similar metric Lie groups in Sects. 6.5 and 10.2 is their appearance as parabolic visual boundaries of negatively curved homogeneous Riemannian manifolds. More precisely, Heintze [Hei74] showed that every simply connected negatively curved homogeneous Riemannian manifold is isometric to a Riemannian Lie group G that is a semi-direct product $N \rtimes_\alpha \mathbb{R}_+$, where N is a simply connected nilpotent Lie group and at the Lie-algebra level, the element 1 in the Lie algebra $\mathbb{R}$ of $\mathbb{R}_+$ acts on $\mathfrak{n}$ by a derivation α whose eigenvalues have strictly positive real parts. The parabolic visual boundary of G may be identified with the Lie group N endowed with some α-homogeneous distance in the sense of Definition 6.5.3, which is the visual distance. It is important to remark that a quasi-isometric classification of Heintze groups, which at the moment is an unsolved problem, is equivalent to a quasi-symmetric classification of their parabolic boundaries, which in turn reduces to a bi-Lipschitz classification of self-similar metric Lie groups (see [Cor18, KLN22], and references therein).

14.3.1 Homogeneous Groups and Heintze Groups

Definition 14.3.1 A *homogeneous group* is a pair (N, α) where N is a simply connected Lie group and α is a derivation on the Lie algebra $\mathfrak{n}$ of N such that for each eigenvalue λ of α it holds $\mathfrak{Re}(\lambda) > 0$.

If (N, α) is a homogeneous group, then $e^{-\alpha}$ is a contractive automorphism of $\mathfrak{n}$, hence $\mathfrak{n}$ admits a positive grading; see Siebert's Theorem 9.2.18. In particular, the group N is nilpotent; see Exercise 9.5.19.

We already encountered homogeneous metric groups in Definition 6.5.3. However, in Chap. 6, we were considering them already metrized with a self-similar distance. One of the punchlines of this chapter is to see such homogeneous groups as metric groups appearing as visual boundaries of negatively curved metric groups. We begin by introducing this latter type of groups.

Definition 14.3.2 A Lie group G is called a *Heintze group* if it is of the form $N \rtimes_\alpha \mathbb{R}_+$, where N is a simply connected Lie group, and $\alpha : \mathfrak{n} \to \mathfrak{n}$ is a derivation whose eigenvalues have positive real parts. In other words, the pair (N, α) is a homogeneous group. Further, we say that a Heintze group $N \rtimes_\alpha \mathbb{R}_+$ (or, equivalently, the homogeneous group (N, α)) is

- *purely real*, if the eigenvalues of α are real numbers,
- *of Carnot type* if it is purely real, if α is diagonalizable over $\mathbb{R}$, and if the eigenspace corresponding to the smallest of the eigenvalues Lie generates $\mathfrak{n}$. Hence, the Lie group N is a stratifiable group, and some multiple of α is the derivation associated with the one-parameter subgroup of Carnot dilations from Definition 11.1.2.

Notice that for every Heintze group $G := N \rtimes_\alpha \mathbb{R}_+$, the nilpotent group N represents the commutator subgroup $[G, G]$. Hence, every Heintze group is solvable; see Exercise 10.5.21.

In this section, we discuss the following theorem from [Hei74].

Theorem 14.3.3 (Heintze) *Every connected isometrically homogeneous Riemannian manifold of negative sectional curvature is isometric to a Heintze group metrized with a left-invariant Riemannian metric.*

Vice versa, every Heintze group may be metrized with a left-invariant Riemannian metric of negative sectional curvature.

The work of Heintze relies on earlier results by Wolf and by Kobayashi; see [Wol64, Kob62, Hei74].

Theorem 14.3.4 (Heintze, after Wolf) *Every connected homogeneous Riemannian manifold of non-positive sectional curvature is isometric to a solvable Lie group metrized with a left-invariant Riemannian metric.*

Theorem 14.3.5 (Heintze) *Every connected solvable Lie group G that admits a left-invariant Riemannian metric of negative sectional curvature is a Heintze group.*

Obviously, the last two results give the first and harder part of Theorem 14.3.3. We will instead discuss the second part of the statement of Theorem 14.3.3.

14.3.1.1 A Discussion on Metrics on Some Heintze Groups

Not all left-invariant Riemannian metrics on Heintze groups are negatively curved. For example, on the Heintze group of Carnot type given by the semi-direct product between the Heisenberg group $\mathcal{N}_1^{\mathbb{C}}$ and $\mathbb{R}_+$, there are left-invariant Riemannian metrics with some planes of positive sectional curvature. We provide an example of one such metric in Exercise 14.4.7. In this example, the group is the same that represents the complex hyperbolic space $\mathbb{C}\mathbf{H}^2$ of complex dimension 2.

Example 14.3.6 (Heintze Groups of Carnot Type) Let N be a Carnot group. Let G be the Lie group obtained by the semi-direct product $\mathbb{R}_+ \ltimes N$ where $\mathbb{R}_+$ acts on N by Carnot dilations. We now show that on G, there is a left-invariant Riemannian metric with negative sectional curvature.

Proof We consider $\mathfrak{n} := \mathrm{Lie}(N)$ as a positively graded vector space, denoting it by V, with grading $V = \bigoplus_{j=1}^{s} V_j$ and associated dilations δ_ϵ, for $\epsilon \in \mathbb{R}$. Let $E_1, \ldots, E_n$ be a basis of V adapted to the direct-sum decomposition. For every $j \in \{1, \ldots, n\}$, let $w_j \in \{1, \ldots, s\}$ be such that $E_j \in V_{w_j}$. Let $\langle \cdot, \cdot \rangle$ be the scalar product that makes the basis $\{E_j\}_j$ orthonormal. Instead of changing the scalar product on $\mathfrak{n}$, we will change the Lie group structure on the vector space $V = \mathfrak{n}$, keeping the same scalar product but putting a different Lie bracket. We stress that the Lie brackets will be isomorphic to those of $\mathfrak{n}$ for every value of a parameter ϵ, except possibly for $\epsilon = 0$.

Let $[\cdot, \cdot]_1$ be Lie brackets on V seen as $\mathfrak{n}$. Let $\{\alpha_{ij}^k\}_{ijk}$ be the structure constants of $\mathfrak{n}$, i.e., $[E_i, E_j]_1 = \sum_{k=1}^{n} \alpha_{ij}^k E_k$, for all $i, j \in \{1, \ldots, n\}$. For $\epsilon \in [0, 1]$ define Lie brackets $[\cdot, \cdot]_\epsilon$ on V by

$$[X, Y]_\epsilon := \delta_\epsilon([X, Y]),$$

i.e., the structure constants are

$$[E_i, E_j]_\epsilon = \sum_{k=1}^{n} \epsilon^{w_k} \alpha_{ij}^k E_k, \qquad \forall i, j \in \{1, \ldots, n\}.$$

Here w_k denotes the weight of E_k.

Define $\mathfrak{n}_\epsilon$ as the vector space V endowed with the Lie bracket $[\cdot,\cdot]_\epsilon$. Whenever $\epsilon \neq 0$, the Lie algebra $\mathfrak{n}_\epsilon$ is isomorphic to $\mathfrak{n}_0$ via the Carnot dilations δ_ϵ. Notice that $[\cdot,\cdot]_\epsilon \to [\cdot,\cdot]_0 \equiv 0$ as $\epsilon \to 0$. In other words, the Lie algebra $\mathfrak{n}_0$ is commutative.

Define the Lie algebra $\mathfrak{g}_\epsilon := \mathbb{R} \ltimes \mathfrak{n}_\epsilon$, with the Lie brackets $[\cdot,\cdot]_\epsilon$ extended from V by

$$[E_0, E_j]_\epsilon := w_j E_j, \qquad \forall j \in \{1, \dots, s\},$$

where $E_0 := (1, \mathbf{0})$, where $\mathbf{0}$ is the zero in $V = \mathfrak{n}_\epsilon$. In total, the structural constants of $\mathfrak{g}_\epsilon$ with respect to the basis $E_0, E_1, \dots, E_n$ are

$$\begin{cases} \alpha_{ij}^k(\epsilon) = \epsilon^{w_k}\alpha_{ij}^k, & \forall i, j, k \neq 0, \\ \alpha_{ij}^0(\epsilon) = 0, & \forall i, j \in \{0, 1, \dots, n\}, \\ \alpha_{0j}^k(\epsilon) = w_j \delta_j^k, & \forall j, k \neq 0. \end{cases} \tag{14.15}$$

The map $(\mathrm{id}, \delta_\epsilon)$ is a Lie algebra isomomorphism between $\mathfrak{g}_1$ and $\mathfrak{g}_\epsilon$, for $\epsilon \neq 0$.

The sectional curvature, as in (8.7), is

$$\mathrm{Sec}_\epsilon(X, Y) := \frac{\langle R_\epsilon(X, Y, Y), X\rangle}{\|X\|^2\|Y\|^2 - \langle X, Y\rangle^2}, \tag{14.16}$$

where R_ϵ is the curvature tensor of $(\mathfrak{g}_\epsilon, \langle\cdot,\cdot\rangle)$. By Proposition 8.1.11, each value $\langle R_\epsilon(E_i, E_j, E_k), E_h\rangle$ can be expressed in terms of the structure constants of $\mathfrak{g}_\epsilon$, which depend on ϵ, with the formula:

$$\begin{aligned} &\langle R_\epsilon(E_i, E_j, E_k), E_h\rangle \\ &\quad = \sum_{\ell=0}^{n} \Bigg[\frac{1}{4}\left(\alpha_{jk}^\ell(\epsilon) - \alpha_{k\ell}^j(\epsilon) + \alpha_{\ell j}^k(\epsilon)\right)\left(\alpha_{i\ell}^h(\epsilon) - \alpha_{\ell h}^i(\epsilon) + \alpha_{hi}^\ell(\epsilon)\right) \\ &\qquad - \frac{1}{4}\left(\alpha_{ik}^\ell(\epsilon) - \alpha_{k\ell}^i(\epsilon) + \alpha_{\ell i}^k(\epsilon)\right)\left(\alpha_{j\ell}^h(\epsilon) - \alpha_{\ell h}^j(\epsilon) + \alpha_{hj}^\ell(\epsilon)\right) \\ &\qquad - \frac{1}{2}\alpha_{ij}^\ell(\epsilon)\left(\alpha_{\ell k}^h(\epsilon) - \alpha_{kh}^\ell(\epsilon) + \alpha_{h\ell}^k(\epsilon)\right)\Bigg]. \end{aligned} \tag{14.17}$$

We focus for the moment on the limit case of $\epsilon = 0$. In this case, the Lie algebra $\mathfrak{g}_0$ is the semi-direct product of $\mathbb{R}$ with an abelian Lie algebra $\mathbb{R}^n$ that is graded by the degrees $w_1, \dots, w_n$. We claim that $(\mathfrak{g}_0, \langle\cdot,\cdot\rangle)$ has negative sectional curvature, that is

$$\mathrm{Sec}_0(X, Y) < 0 \qquad \forall X, Y \in \mathbb{R} \times V \text{ linearly independent.} \tag{14.18}$$

In fact, we need to show that

$$\langle R_0(X, Y, Y), X\rangle \leq 0, \qquad \forall X, Y \in \mathfrak{g}, \tag{14.19}$$

with equality (if and) only if X and Y are linearly dependent. Notice that, writing X and Y in components, by linearity we have

$$\langle R_0(X, Y, Y), X\rangle = \sum_{i,j,k,h=0}^{n} X^i Y^j Y^k X^h \langle R_0(E_i, E_j, E_k), E_h\rangle.$$

So, we first distinguish three cases of indices in $\langle R_0(E_i, E_j, E_k), E_h\rangle$: the index 0 appears twice, it appears exactly once, or it does not appear. Because of the symmetries of the curvature tensor, the three cases cover all possibilities.

Case 1: Suppose $i = h = 0$ and $j, k \neq 0$. Then we claim that

$$\langle R_0(E_i, E_j, E_k), E_h\rangle = -w_j^2 \delta_{jk}.$$

Indeed, by the formula for the curvature tensor (14.17) in terms of the structural constants (14.15), we have:

$$\begin{aligned}
&\langle R_0(E_i, E_j, E_k),\ E_h\rangle = \langle R_0(E_0, E_j, E_k), E_0\rangle \\
&\overset{(14.17)}{=} \frac{1}{4}\left(\alpha^0_{jk}(0) - \alpha^j_{k0}(0) + \alpha^k_{0j}(0)\right)\left(\alpha^0_{00}(0) - \alpha^0_{00}(0) + \alpha^0_{00}(0)\right) \\
&\quad - \frac{1}{4}\left(\alpha^0_{0k}(0) - \alpha^0_{k0}(0) + \alpha^k_{00}(0)\right)\left(\alpha^0_{j0}(0) - \alpha^j_{00}(0) + \alpha^0_{0j}(0)\right) \\
&\quad - \frac{1}{2}\alpha^0_{0j}(0)\left(\alpha^0_{0k}(0) - \alpha^0_{k0}(0) + \alpha^k_{00}(0)\right) \\
&\quad + \sum_{\ell=1}^{n}\Bigg[\frac{1}{4}\left(\alpha^\ell_{jk}(0) - \alpha^j_{k\ell}(0) + \alpha^k_{\ell j}(0)\right)\left(\alpha^0_{0\ell}(0) - \alpha^0_{\ell 0}(0) + \alpha^\ell_{00}(0)\right) \\
&\quad - \frac{1}{4}\left(\alpha^\ell_{0k}(0) - \alpha^0_{k\ell}(0) + \alpha^k_{\ell 0}(0)\right)\left(\alpha^0_{j\ell}(0) - \alpha^j_{\ell 0}(0) + \alpha^\ell_{0j}(0)\right) \\
&\quad - \frac{1}{2}\alpha^\ell_{0j}(0)\left(\alpha^0_{\ell k}(0) - \alpha^\ell_{k0}(0) + \alpha^k_{0\ell}(0)\right)\Bigg]
\end{aligned}$$

$$\begin{aligned}
&\overset{(14.15)}{=} \sum_{\ell=1}^{n}\Bigg[-\frac{1}{4}\left(\alpha^\ell_{0k}(0) - \alpha^k_{0\ell}(0)\right)\left(\alpha^j_{0\ell}(0) + \alpha^\ell_{0j}(0)\right) \\
&\qquad - \frac{1}{2}\alpha^\ell_{0j}(0)\left(\alpha^\ell_{0k}(0) + \alpha^k_{0\ell}(0)\right)\Bigg]
\end{aligned}$$

$$\overset{(14.15)}{=} -\sum_{\ell=1}^{n}\frac{1}{2}w_j\delta_j^\ell\left(w_k\delta_k^\ell + w_k\delta_\ell^k\right)$$
$$= -w_j^2\delta_j^k.$$

Case 2: Suppose $i = 0$ and $j, k, h \neq 0$. Then, using (14.17), similarly to the previous case, one calculates:

$$\langle R_0(E_i, E_j, E_k), E_h\rangle = 0.$$

Case 3: Suppose $i, j, k, h \neq 0$. Then, using (14.17), similarly as in the first case, one calculates:

$$\langle R_0(E_i, E_j, E_k), E_h\rangle = w_i w_j(-\delta_{jk}\delta_{ih} + \delta_{ik}\delta_{jh}).$$

Using the symmetries of the curvature tensor, we see that in the general case, for $X, Y \in \mathfrak{g}$, we have

$$\begin{aligned}\langle R_0(X, Y, Y), X\rangle &= \sum_{i,j,k,h=0}^{n} X^i Y^j Y^k X^h \langle R_0(E_i, E_j, E_k), E_h\rangle \\ &= \sum_{j,h=1}^{n} X^0 Y^j Y^0 X^h \langle R_0(E_0, E_j, E_0), E_h\rangle \\ &\quad + \sum_{j,k=1}^{n} X^0 Y^j Y^k X^0 \langle R_0(E_0, E_j, E_k), E_0\rangle \\ &\quad + \sum_{i,h=1}^{n} X^i Y^0 Y^0 X^h \langle R_0(E_i, E_0, E_0), E_h\rangle \\ &\quad + \sum_{i,k=1}^{n} X^i Y^0 Y^k X^0 \langle R_0(E_i, E_0, E_k), E_0\rangle \\ &\quad + \sum_{i,j,k,h=1}^{n} X^i Y^j Y^k X^h \langle R_0(E_i, E_j, E_k), E_h\rangle\end{aligned}$$

$$\begin{aligned}&= \sum_{j,h=1}^{n} X^0 Y^j Y^0 X^h (w_j^2\delta_{jh}) \\ &\quad + \sum_{j,k=1}^{n} X^0 Y^j Y^k X^0 (-w_j^2\delta_{jk})\end{aligned}$$

$$
\begin{aligned}
&+\sum_{i,h=1}^{n} X^i Y^0 Y^0 X^h (-w_i^2 \delta_{ih}) \\
&+\sum_{i,k=1}^{n} X^i Y^0 Y^k X^0 (w_i^2 \delta_{ik}) \\
&+\sum_{i,j,k,h=1}^{n} X^i Y^j Y^k X^h w_i w_j (-\delta_{jk}\delta_{ih} + \delta_{ik}\delta_{jh})
\end{aligned}
$$

$$
\begin{aligned}
&= -\sum_{j=1}^{n} w_j^2 \left(X^0 Y^j - Y^0 X^j \right)^2 - \sum_{i,j=1}^{n} (X^i)^2 (Y^j)^2 w_i w_j + \sum_{i,j=1}^{n} X^i Y^j Y^i X^j w_i w_j \\
&= -w_j^2 \left(X^0 Y^j - Y^0 X^j \right)^2 - \|A\|^2 \cdot \|B\|^2 + \langle A, B \rangle^2,
\end{aligned}
$$

where

$$
A := \sum_{i=1}^{n} \sqrt{w_i} X^i E_i \quad \text{and} \quad B := \sum_{j=1}^{n} \sqrt{w_j} Y^j E_j .
$$

Notice that $-\|A\| \cdot \|B\| + \langle A, B \rangle \le 0$, with equality if and only if A and B are linearly dependent. We conclude that (14.19) holds for all $X, Y \in \mathfrak{g}_0$. If equality holds, i.e., if $\langle R_0(X, Y, Y), X \rangle = 0$, then A and B are linearly dependent, and thus so are $(X^1, \dots, X^n)$ and $(Y^1, \dots, Y^n)$, because $\{\sqrt{w_i} E_i\}_{i=1}^n$ is a basis of V. Up to exchanging X and Y, we can suppose $Y_j = r X_j$ for all j, with $r \in \mathbb{R}$. Moreover, we also have $(X^0 Y^j - Y^0 X^j)^2 = 0$ for every j, that is, $(X^0 r - Y^0)^2 (X^j)^2 = 0$. Therefore, either X and Y are multiples of E_0, or $Y = rX$: in both cases, they are linearly dependent. Thus, we obtained (14.18).

Next, we infer that there is $\epsilon_0 > 0$ such that

$$
\mathrm{Sec}_\epsilon(X, Y) < 0 \quad \forall \epsilon \in [0, \epsilon_0] \text{ and } \forall X, Y \in \mathbb{R} \times V \text{ linearly independent.} \tag{14.20}
$$

Indeed, thanks to (14.18) and the compactness of the Grassmannian of 2-planes in $\mathbb{R} \times V$, we can just show that $\mathrm{Sec}_\epsilon(X, Y) \to \mathrm{Sec}_0(X, Y)$ as $\epsilon \to 0$, locally uniformly in X, Y linearly independent. But this is clear from (14.16) and (14.17). □

14.3.2 Self-Similar Lie Groups as Parabolic Visual Boundaries

We now explain how the parabolic boundary of a Heintze group is identified with a self-similar metric Lie group.

Let $G := N \rtimes \mathbb{R}_+$ be a Heintze group equipped with a left-invariant Riemannian metric g such that the maximum of the sectional curvature is -1. Up to replacing the $\mathbb{R}_+$ factor with a one-parameter subgroup orthogonal to $\mathrm{Lie}(N)$, we may assume that $N \times \{1\}$ is orthogonal to $\{1_N\} \times \mathbb{R}_+$. Or, more precisely, we have that $T_{1_N} N \times \{0\}$ is perpendicular to $\{0\} \times \mathbb{R}$ in $T_{1_G} G$.

We begin by parametrizing the boundary of G. In the following proposition, we say that the Lie group $\mathbb{R}_+$ acts *expansively* on a Lie group N if every $\lambda \in (0, 1)$ acts as a contractive automorphism in the sense defined at page 261; see also Exercises 14.4.9 and 14.4.10.

Proposition 14.3.7 *Let $G := N \rtimes \mathbb{R}_+$ be a Heintze group, i.e., a semi-direct product of a Lie group N with the Lie group $\mathbb{R}_+$, such that $\mathbb{R}_+ \curvearrowright N$ expansively. Assume that G is equipped with a left-invariant Riemannian metric such that* $\mathrm{Sec} \leq -1$ *and such that $T_{1_N} N \times \{0\}$ is perpendicular to $\{0\} \times \mathbb{R}$ in $T_{1_G} G$. Then there exists $\mu > 0$ such that the map*

$$n \in N \longmapsto \big(t \in \mathbb{R} \mapsto \gamma_n(t) := (n, e^{-\mu t})\big) \tag{14.21}$$

gives a continuous parametrization of the parabolic visual boundary $\partial_{\mathrm{Par},\omega} G$ from ω where (Fig. 14.4)

$$t \in \mathbb{R} \mapsto \omega(t) := (1_N, e^{-\mu t}). \tag{14.22}$$

Proof We consider the projection $N \rtimes \mathbb{R}_+ \to N \rtimes \mathbb{R}_+/N \simeq \mathbb{R}_+$ modulo the normal subgroup N. Since $T_{1_N} N \times \{0\}$ is orthogonal to $\{0\} \times \mathbb{R}$, then this projection is a submetry; see Corollary 6.3.5. Therefore, every injective curve with support in $\{1_N\} \times \mathbb{R}_+$ is length minimizing between every two of its points. Therefore, the

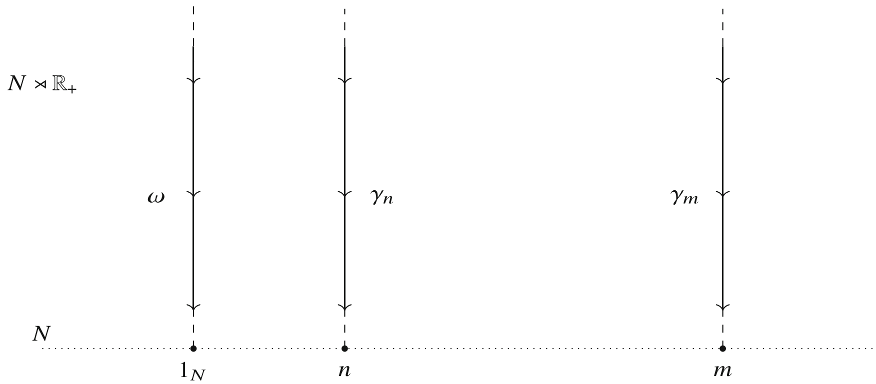


Fig. 14.4 The vertical geodesics in a Heitze group $N \rtimes \mathbb{R}_+$. Every point n in N is associated with the geodesic line $\gamma_n(t) := (n, e^{-\mu t})$, whose hight decreases when time increases. The dotted horizontal line represents the parabolic visual boundary N, where the visual point is the geodesic line $\omega(t) := (1_N, e^{-\mu t})$

one-parameter subgroup $s \in \mathbb{R} \mapsto (1_N, e^s)$ is length minimizing, and, because the metric is left-invariant, it is parametrized at a constant speed, possibly different from 1. We deduce that there exists $\mu \in \mathbb{R}^+$ such that ω as defined in (14.22) is an isometric embedding. By the left-invariance of the distance, we infer that every γ_n, as in (14.21), is a geodesic line.

We denote by $\lambda.n$ the action of $\lambda \in \mathbb{R}_+$ evaluated at $n \in N$. Using the left-invariance and the group laws of semi-direct products, we express the distance between points on two such curves γ_n and γ_m, for $n, m \in N$:

$$\begin{aligned} d(\gamma_n(t), \gamma_m(t)) &\overset{(14.21)}{=} d\left((n, e^{-\mu t}), (m, e^{-\mu t})\right) \\ &= d\left((m, e^{-\mu t})^{-1} \cdot (n, e^{-\mu t}), 1_G\right) \\ &\overset{\text{Rem. 5.6.7.v}}{=} d\left((e^{\mu t}.(m^{-1}), e^{\mu t})(n, e^{-\mu t}), 1_G\right) \\ &\overset{(5.18)}{=} d\left((e^{\mu t}.(m^{-1}) e^{\mu t}.(n), 1_{\mathbb{R}}), 1_G\right) \\ &= d\left((e^{\mu t}.(m^{-1}n), 1_{\mathbb{R}}), 1_G\right). \end{aligned} \tag{14.23}$$

To explain that every γ_n is an element in $\partial_{\mathrm{Par},\omega} G$, we observe that it has the same origin as ω:

$$\begin{aligned} d(\gamma_n(t), \omega(t)) &= d\left(\gamma_n(t), \gamma_{1_N}(t)\right) \\ &\overset{(14.23)}{=} d\left((e^{\mu t}.(n), 1_{\mathbb{R}}), 1_G\right) \\ &\xrightarrow{\text{Ex. 14.4.9}} d((1_N, 1_{\mathbb{R}}), 1_G) = 0, \qquad \text{as } t \to -\infty. \end{aligned}$$

We claim that every two curves γ_n and γ_m with $n, m \in N$ distinct, define two different points in $\partial_{\mathrm{Par},\omega} G$, i.e., they are not asymptotic. In fact, we show that

$$\lim_{t \to +\infty} d(\gamma_n(t), \gamma_m(t)) = \infty, \qquad \forall n, m \in N, n \neq m. \tag{14.24}$$

Indeed, since $m^{-1}n \neq 1_N$ and $\mathbb{R}_+ \curvearrowright N$ is expansive, then, as $t \to +\infty$, then $e^{\mu t}.(m^{-1}n)$ leaves every compact set in N. Consequently, the curve $(e^{\mu t}.(m^{-1}n), 1_{\mathbb{R}})$ leaves every compact set in G, and so it diverges. Therefore, we infer

$$d(\gamma_n(t), \gamma_m(t)) \overset{(14.23)}{=} d\left((e^{\mu t}.(m^{-1}n), 1_{\mathbb{R}}), 1_G\right) \to \infty, \qquad \text{as } t \to +\infty.$$

Thus, we proved (14.24).

We next show that the correspondence $n \in N \mapsto \gamma_n$ is surjective. Consider $\gamma \in \partial_{\mathrm{Par},\omega} G$. Take $n \in N$ and $t_0 \in \mathbb{R}_+$ such that $\gamma(0) = (n, e^{t_0}) = \gamma_n(t_0)$. We shall show that $\gamma = \gamma_n$. Because both γ and γ_n have the same origin as ω, then $\lim_{t\to-\infty} d(\gamma(t), \gamma_n(t)) = 0$. Therefore, for very large $T > 0$ the triangle $\gamma(0)$, $\gamma(-T)$, and $\gamma_n(-T)$ has side lengths that are T, $T + t_0$ and an infinitesimal length. Because the metric space G is CAT(-1) and T may be taken arbitrarily large, these triangles are degenerate, i.e., we have $\gamma|_{(-\infty,0]} = \gamma_n|_{(-\infty,0]}$. Since Riemannian geodesics do not branch, we conclude $\gamma = \gamma_n$.

We check that the map in (14.21) is continuous, where on $\partial_{\mathrm{Par},\omega} G$ we consider the topology given by the visual distance from Definition 14.1.4. Namely, we check that if $m \to n$ in N, then $d_{\mathrm{vis}}(\gamma_m, \gamma_n) \to 0$. For this purpose, for $n, m \in N$ distinct, let $t^* \in \mathbb{R}$ such that

$$d(\gamma_n(t^*), \gamma_m(t^*)) \overset{(14.23)}{=} d\left((e^{\mu t^*}.(m^{-1}n), 1_{\mathbb{R}}), 1_G\right) = 1.$$

We point out that, as $m \to n$, we have $t^* \to +\infty$. By triangle inequality, we bound

$$\begin{aligned} d(\gamma_m(t), \gamma_n(t)) &\le d(\gamma_m(t), \gamma_m(t^*)) + d(\gamma_m(t^*), \gamma_n(t^*)) + d(\gamma_n(t^*), \gamma_n(t)) \\ &= 2(t - t^*) + 1 \end{aligned}$$

and, consequently,

$$\begin{aligned} d_{\mathrm{vis}}(\gamma_m, \gamma_n) &\overset{\text{Def. 14.1.4}}{=} \exp\left(-\tfrac{1}{2} \lim_{t\to+\infty} (2t - d(\gamma_m(t), \gamma_n(t)))\right) \\ &\le \quad \exp\left(-t^* + \tfrac{1}{2}\right) \to 0, \text{ as } m \to n. \end{aligned}$$

Thus, the parametrization is continuous—it is actually a homeomorphism, but we shall motivate this fact later. □

Using the correspondence above, we view d_{vis} as a distance function on N.

Definition 14.3.8 (Induced Hamenstädt Distance Function) Let $N \rtimes \mathbb{R}_+$ be a Heintze group equipped with a left-invariant Riemannian metric such that Sec ≤ -1. Up to changing the $\mathbb{R}_+$ factor, assume that $T_{1_N}N \times \{0\}$ is perpendicular to $\{0\} \times \mathbb{R}$. The *induced Hamenstädt distance function* on N is the function

$$\hat{d}(n, m) := d_{\mathrm{vis}}(\gamma_n, \gamma_m), \qquad \forall n, m \in N, \tag{14.25}$$

where γ_n is the geodesic from (14.21).

Theorem 14.3.9 *Given a Heintze group $N \rtimes \mathbb{R}_+$, the induced Hamenstädt distance function on N is admissible, left-invariant, and self-similar. In fact, the group homomorphisms*

$$\delta_\lambda(n) := \lambda^\mu.n, \qquad \forall \lambda \in \mathbb{R}_+, \forall n \in N, \tag{14.26}$$

are dilations for the distance, where μ is the normalizing factor from Proposition 14.3.7.

Proof By (14.23), it is immediate that the visual distance is left-invariant on N. For $s \in \mathbb{R}$, consider the action of $e^{\mu s}$ on N. For $n, m \in N$, the visual distance changes as a translation in time for the geodesics as follows:

$$\begin{aligned}
\hat{d}(e^{\mu s}.n, e^{\mu s}.m) &\overset{(14.25)}{=} d_{\text{vis}}(\gamma_{e^{\mu s}.n}, \gamma_{e^{\mu s}.m}) \\
&\overset{\text{D. } 14.1.4}{=} \exp\left(-\lim_{t\to+\infty}\left(t - \tfrac{1}{2}d(\gamma_{e^{\mu s}.n}(t), \gamma_{e^{\mu s}.m}(t))\right)\right) \\
&\overset{(14.23)}{=} \exp\left(-\lim_{t\to+\infty}\left(t - \tfrac{1}{2}d((e^{\mu t}.((e^{\mu s}.m)^{-1}(e^{\mu s}.n)), 1_{\mathbb{R}}), 1_G)\right)\right) \\
&= \exp\left(s - \lim_{t\to+\infty}\left(t + s - \tfrac{1}{2}d((e^{\mu(t+s)}.(m^{-1}n), 1_{\mathbb{R}}), 1_G)\right)\right) \\
&= e^s \exp\left(-\lim_{t'\to+\infty}\left(t' - \tfrac{1}{2}d((e^{\mu t'}.(m^{-1}n), 1_{\mathbb{R}}), 1_G)\right)\right) \\
&\overset{(14.23)}{=} e^s \exp\left(-\lim_{t'\to+\infty}\left(t' - \tfrac{1}{2}d(\gamma_n(t'), \gamma_m(t'))\right)\right) \\
&\overset{\text{D. } 14.1.4}{=} e^s d_{\text{vis}}(\gamma_n, \gamma_m) \\
&\overset{(14.25)}{=} e^s \hat{d}(n, m).
\end{aligned}$$

Setting $\lambda := e^s$ so (14.26) gives $\delta_\lambda(n) = e^{\mu s}.n$, for $n \in N$, we deduce

$$\hat{d}(\delta_\lambda(n), \delta_\lambda(m)) = \lambda\hat{d}(n, m), \qquad \forall\lambda \in \mathbb{R}_+, \forall n, m \in N.$$

Regarding admissibility, it is a general property of left-invariant metrics on Lie groups admitting a one-parameter family $(\delta_\lambda)_\lambda$ of automorphisms each of which is a dilation of factor λ; see Exercise 14.4.11. However, we already proved in Proposition 14.3.7 that the parametrization is continuous. Hence, the function $\hat{d}$ is continuous with respect to the topology. Consequently, using the dilations, it is easier to conclude that it induces the manifold topology, recall Exercise 10.5.23. □

We conclude with a summary of this chapter. Every Riemannian homogeneous negatively curved manifold has a notion of parabolic visual boundary, which has the structure of a metric Lie group admitting dilations. All Carnot groups appear in this way, as boundaries of Heintze groups of Carnot type. From the Lie group viewpoint, metric Lie groups admitting dilations are precisely the nilpotent positively graded Lie groups, as discussed in Theorem 10.2.1.

14.4 Exercise

Exercise 14.4.1 Busemann functions associated with geodesic lines are well-defined (i.e., the limit exists) and are 1-Lipschitz.
Hint. The limit is monotone.

Exercise 14.4.2 In the upper-halfspace model of the hyperbolic plane for every positive t and D, we have $d\left((0, e^{-t}), (\frac{2}{e^t}\sinh\left(\frac{D}{2}\right), e^{-t})\right) = D$.
Hint. Use Exercise 13.4.12.

Exercise 14.4.3 The visual distance defined as in (14.3) is left-invariant when $\partial_\infty \mathbb{A}\mathbf{H}^n$ is identified with $\mathbb{A} \times \Im\mathfrak{m}(\mathbb{A})$ with multiplication law (13.17).
Hint. Either use the left-invariance proved in Theorem 13.3.3 or use formula (14.12).

Exercise 14.4.4 The Cygan-Koranyi distance on $\mathbb{A}^{n-1} \times \Im\mathfrak{m}(\mathbb{A})$, given by (14.12), between the point $(0, \mathbf{0})$ and a point $(\xi, \mathbf{v}) \in \mathbb{A}^{n-1} \times \Im\mathfrak{m}(\mathbb{A})$ is

$$d((0, \mathbf{0}), (\xi, \mathbf{v})) = \sqrt[4]{\|\xi\|^4 + |\mathbf{v}|^2}. \tag{14.27}$$

This function is called *Cygan-Koranyi gauge*.

Exercise 14.4.5 For $\mathbb{A} \in \{\mathbb{R}, \mathbb{C}, \mathbb{H}\}$ and $n \in \mathbb{N}$, equip $\mathbb{A}^{n-1} \times \Im\mathfrak{m}((\mathbb{A}) \times \mathbb{R}_+$ with the group product given in Exercise 13.4.22. This Lie group is a Heintze group.

Exercise 14.4.6 Every group G_z, with $z \in \mathbb{C} \setminus i\mathbb{R}$, from Example 6.4.4 is a Heintze group.

Exercise 14.4.7 Let $\mathfrak{g}$ be the 4D Lie algebra with basis $X_1, \dots, X_4$ and only nontrivial structural constants, as in (5.2),

$$c_{12}^3 = 4, \quad c_{41}^1 = 1, \quad c_{42}^2 = 1, \quad c_{43}^3 = 2.$$

This Lie algebra is isomorphic to the one in Exercise 8.3.11. Consider a Lie group with such a $\mathfrak{g}$ as Lie algebra, and on it, consider the left-invariant Riemannian structure for which $X_1, \dots, X_4$ form an orthonormal basis. We have the sectional curvature $\mathrm{Sec}(X_1, X_3) = 2$.
Hint. Apply (8.8).

Exercise 14.4.8 Let N be a simply connected Lie group whose Lie algebra is positively graded. Let G be the Lie group obtained by the semi-direct product $\mathbb{R}_+ \ltimes N$ where $\mathbb{R}_+$ acts on N by the dilations induced by the grading. On G, there is a left-invariant Riemannian metric with negative sectional curvature.
Hint. Adapt the proof of Example 14.3.6.

Exercise 14.4.9 An action $\mathbb{R}_+ \curvearrowright N$ by automorphisms is expansive if and only if $\lim_{t\to 0} t.n = 1_N$, for all $n \in N$.

Exercise 14.4.10 If an action $\mathbb{R}_+ \curvearrowright N$ is given by a derivation A, then the action is expansive if and only if the eigenvalues of A have strictly positive real parts. Equivalently, this is exactly when $N \rtimes_A \mathbb{R}_+$ is a Heintze group.

Exercise 14.4.11 ☠ Let G be a Lie group. Let $\lambda \in (0, \infty) \mapsto \delta_\lambda \in \mathrm{Aut}(G)$ be a one-parameter group of Lie group automorphisms. If there exists a left-invariant distance d on G with $d(\delta_\lambda x, \delta_\lambda y) = \lambda d(x, y)$, for all $x, y \in G$ and $\lambda > 0$, then d is admissible and G is connected.
Hint. Check [LN21, Theorem 1.1].

References

[Aba+10] M. Abate, F. Bracci, M.D. Contreras, S. Díaz-Madrigal, The evolution of Loewner's differential equations. Eur. Math. Soc. Newsl. **78**, 31–38 (2010)

[ABB20] A. Agrachev, D. Barilari, U. Boscain, A comprehensive introduction to sub-Riemannian geometry, in *Cambridge Studies in Advanced Mathematics*, vol. 181 (2020). From the Hamiltonian viewpoint, With an appendix by Igor Zelenko, pp. xviii+745

[AFP00] L. Ambrosio, N. Fusco, D. Pallara. *Functions of Bounded Variation and Free Discontinuity Problems*. Oxford Mathematical Monographs (The Clarendon Press Oxford University Press, New York, 2000), pp. xviii+434

[Agr14] A.A. Agrachev, Some open problems, in *Geometric Control Theory and sub-Riemannian Geometry*, vol. 5. Springer INdAM Series (Springer, Cham, 2014), pp. 1–13

[AKL09] L. Ambrosio, B. Kleiner, E. Le Donne, Rectifiability of sets of finite perimeter in Carnot groups: existence of a tangent hyperplane. J. Geom. Anal. **19**(3), 509–540 (2009)

[ALN23] G. Antonelli, E. Le Donne, S. Nicolussi Golo, Lipschitz Carnot-Carathéodory Structures and their Limits. J. Dyn. Control Syst. **29**(3), 805–854 (2023)

[ALS19a] A.A. Ardentov, E. Le Donne, Y.L. Sachkov, A sub-Finsler problem on the Cartan group. Tr. Mat. Inst. Steklova **304**, 49–67 (2019). Optimal′noe Upravlenie i Differentsial′nye Uravneniya

[ALS19b] A.A. Ardentov, E. Le Donne, Y.L. Sachkov, Sub-Finsler geodesics on the Cartan group. Regul. Chaotic Dyn. **24**(1), 36–60 (2019)

[AMP21] F. Anceschi, S. Muzzioli, S. Polidoro, Existence of a fundamental solution of partial differential equations associated to Asian options. Nonlinear Anal. Real World Appl. **62**, Paper No. 103373, 29 (2021)

[AP94] M. Abate, G. Patrizio, *Finsler Metrics—a Global Approach*, vol. 1591. Lecture Notes in Mathematics. With applications to geometric function theory (Springer-Verlag, Berlin, 1994), pp. x+180

[Are46] R. Arens, Topologies for homeomorphism groups. Am. J. Math. **68**, 593–610 (1946)

[AS04] A.A. Agrachev, Y.L. Sachkov. *Control Theory from the Geometric Viewpoint*, vol. 87. Encyclopaedia of Mathematical Sciences. Control Theory and Optimization, II. (Springer, Berlin, 2004), pp. xiv+412

[AS13] A.A. Agrachev,Y. Sachkov, *Control Theory from the Geometric Viewpoint*, vol. 87 (Springer Science & Business Media, Berlin, 2013)

[AT04] L. Ambrosio, P. Tilli, *Topics on Analysis in Metric Spaces*, vol. 25. Oxford Lecture Series in Mathematics and its Applications (Oxford University Press, Oxford, 2004), pp. viii+133

E. Le Donne, *Metric Lie Groups*, Graduate Texts in Mathematics 306,
https://doi.org/10.1007/978-3-031-98832-5

[AT11] M. Abate, F. Tovena, *Geometria differenziale*, vol. 54. Unitext. La Matematica per il 3+2 (Springer, Milan, 2011), pp. xiv+465

[AT12] M. Abate, F. Tovena. *Curves and Surfaces*, vol. 55. Unitext. Translated from the 2006 Italian original by D.A. Gewurz (Springer, Milan, 2012), pp. xiv+390

[Bar00] Y. Baryshnikov, On small Carnot-Carathéodory spheres. Geom. Funct. Anal. **10**(2), 259–265 (2000)

[Bar+17] D. Barilari, U. Boscain, E. Le Donne, M. Sigalotti, Sub-Finsler structures from the time-optimal control viewpoint for some nilpotent distributions. J. Dyn. Control Syst. **23**(3), 547–575 (2017)

[Bas72] H. Bass, The degree of polynomial growth of finitely generated nilpotent groups. Proc. London Math. Soc. (3) **25**, 603–614 (1972)

[Bat+99] S. Bates, W.B. Johnson, J. Lindenstrauss, D. Preiss, G. Schechtman, Affine approximation of Lipschitz functions and nonlinear quotients. Geom. Funct. Anal. **9**(6), 1092–1127 (1999)

[Bay+23] C. Bayer, P.P. Hager, S. Riedel, J. Schoenmakers, Optimal stopping with signatures. Ann. Appl. Probab. **33**(1), 238–273 (2023)

[BB00] Z.M. Balogh, M. Bonk, Gromov hyperbolicity and the Kobayashi metric on strictly pseudoconvex domains. Comment. Math. Helv. **75**(3), 504–533 (2000)

[BBI01] D. Burago, Y. Burago, S. Ivanov. *A Course in Metric Geometry*, vol. 33. Graduate Studies in Mathematics (American Mathematical Society, Providence, RI, 2001), pp. xiv+415

[BBL20] D. Barilari, I. Beschastnyi, A. Lerario, Volume of small balls and sub-Riemannian curvature in 3D contact manifolds. J. Symplectic Geom. **18**(2), 355–384 (2020)

[BCP18] E.J. Bekkers, D. Chen, J.M. Portegies, Nilpotent approximations of sub-Riemannian distances for fast perceptual grouping of blood vessels in 2D and 3D. J. Math. Imaging Vision **60**(6), 882–899 (2018)

[BCS00] D. Bao, S.-S. Chern, Z. Shen. *An introduction to Riemann-Finsler geometry*, vol. 200. Graduate Texts in Mathematics (Springer, New York, 2000), pp. xx+431

[Bea83] A.F. Beardon. *The Geometry of Discrete Groups*, vol. 91. Graduate Texts in Mathematics (Springer, New York, 1983), pp. xii+337

[Bek+17] E.J. Bekkers, R. Duits, A. Mashtakov, Y. Sachkov, Vessel tracking via sub-Riemannian geodesics on the projective line bundle, in *Geometric Science of Information*, vol. 10589. Lecture Notes in Computer Science (Springer, Cham, 2017), pp. 773–781

[Bek+18] E.J. Bekkers, M.W. Lafarge, M. Veta, K.A.J. Eppenhof, J.P.W. Pluim, R. Duits, Roto-translation covariant convolutional networks for medical image analysis, in *Medical Image Computing and Computer Assisted Intervention—MICCAI 2018*, ed. by A.F. Frangi, J.A. Schnabel, C. Davatzikos, C. Alberola-López, G. Fichtinger (Springer International Publishing, Cham, 2018), pp. 440–448

[Bel96] A. Bellaïche, The tangent space in sub-Riemannian geometry, in *Sub-Riemannian Geometry*, vol. 144. Progress in Mathematics (Birkhäuser, Basel, 1996), pp. 1–78

[Ber88] V.N. Berestovskiĭ, Homogeneous manifolds with an intrinsic metric. I. Sibirsk. Mat. Zh. **29**(6), 17–29 (1988)

[Ber89a] V.N. Berestovskiĭ, Homogeneous manifolds with an intrinsic metric. II. Sibirsk. Mat. Zh. **30**(2), 14–28, 225 (1989)

[Ber89b] V.N. Berestovskiĭ, The structure of locally compact homogeneous spaces with an intrinsic metric. Sibirsk. Mat. Zh. **30**(1), 23–34 (1989)

[Bes78] A.L. Besse. *Manifolds All of Whose Geodesics are Closed*, vol. 93. Ergebnisse der Mathematik und ihrer Grenzgebiete [Results in Mathematics and Related Areas]. With appendices by D.B.A. Epstein, J.-P. Bourguignon, L. Bérard-Bergery, M. Berger, J.L. Kazdan (Springer, Berlin, 1978), pp. ix+262

[BF11] A. Bonfiglioli, R. Fulci, *Topics in Noncommutative Algebra. The Theorem of Campbell, Baker, Hausdorff and Dynkin*, vol. 2034. English. Lecture Notes in Mathematics (Springer, Berlin, 2011)

[BH99] M.R. Bridson and A. Haefliger. *Metric spaces of non-positive curvature*. Vol. 319. Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]. Springer-Verlag, Berlin, 1999, pp. xxii+643.

[Bjö19] T. Björk. *Arbitrage Theory in Continuous Time*. Oxford University Press, 2019, p. 592.

[BJR98] A. Bellaïche, F. Jean, J.-J. Risler, Geometry of nonholonomic systems. In: *Robot motion planning and control*. Vol. 229. Lect. Notes Control Inf. Sci. Springer, London, 1998, pp. 55–91.

[BL13] E. Breuillard and E. Le Donne, On the rate of convergence to the asymptotic cone for nilpotent groups and subFinsler geometry. Proc. Natl. Acad. Sci. USA 110.48 (2013), pp. 19220–19226.

[Blo03] A.M. Bloch. *Nonholonomic mechanics and control*. Vol. 24. Interdisciplinary Applied Mathematics. With the collaboration of J. Baillieul, P. Crouch and J. Marsden, With scientific input from P.S. Krishnaprasad, R.M. Murray and D. Zenkov, Systems and Control. Springer-Verlag, New York, 2003, pp. xx+483.

[BLU07] A. Bonfiglioli, E. Lanconelli, F. Uguzzoni. *Stratified Lie groups and potential theory for their sub-Laplacians*. Springer Monographs in Mathematics. Springer, Berlin, 2007, pp. xxvi+800.

[BMB16] A.V. Borisov, I.S. Mamaev, I.A. Bizyaev, Historical and critical review of the development of nonholonomic mechanics: the classical period. Regul. Chaotic Dyn. 21.4 (2016), pp. 455–476.

[BMP20] F. Boarotto, R. Monti, F. Palmurella, Third order open mapping theorems and applications to the end-point map. English. Nonlinearity 33.9 (2020), pp. 4539–4567.

[BMS24] F. Boarotto, R. Monti, A. Socionovo, Higher order Goh conditions for singular extremals of corank 1. English. Arch. Ration. Mech. Anal. 248.2 (2024). Id/No 23, p. 46.

[BNV22] F. Boarotto, L. Nalon, D. Vittone, The Sard problem in step 2 and in filiform Carnot groups. English. ESAIM, Control Optim. Calc. Var. 28 (2022). Id/No 75, p. 20.

[Bod23] C. Bodart, *Intermediate geodesic growth in virtually nilpotent groups*. Groups Geom. Dyn. (2025), published online first

[Bos06] U. Boscain. *Motion Planning and Optimal Control for Quantum Mechanical Systems*. Habilitation à diriger des recherches. Université de Bourgogne, 2006.

[Bos+12] U. Boscain, J. Duplaix, J.-P. Gauthier, F. Rossi, Anthropomorphic image reconstruction via hypoelliptic diffusion. SIAM J. Control Optim. 50.3 (2012), pp. 1309–1336.

[Bos+14] U. Boscain, R.A. Chertovskih, J.P. Gauthier, A.O. Remizov, Hypoelliptic diffusion and human vision: a semidiscrete new twist. SIAM J. Imaging Sci. 7.2 (2014), pp. 669–695.

[Bos89] A. Bose, Dynkin's method of computing the terms of the Baker-Campbell-Hausdorff series. J. Math. Phys. 30.9 (1989), pp. 2035–2037.

[Bou05] N. Bourbaki. *Lie groups and Lie algebras. Chapters 7–9*. Elements of Mathematics (Berlin). Translated from the 1975 and 1982 French originals by Andrew Pressley. Springer-Verlag, Berlin, 2005, pp. xii+434.

[Bou95] M. Bourdon, Structure conforme au bord et flot géodésique d'un CAT 1—espace. Enseign. Math. (2) 41.1-2 (1995), pp. 63–102.

[Bou98] N. Bourbaki. *Lie groups and Lie algebras. Chapters 1–3*. Elements of Mathematics (Berlin). Translated from the French, Reprint of the 1989 English translation. Springer-Verlag, Berlin, 1998, pp. xviii+450.

[BP04] U. Boscain and B. Piccoli. *Optimal syntheses for control systems on 2-D manifolds*. Vol. 43. Mathématiques & Applications (Berlin) [Mathematics & Applications]. Springer-Verlag, Berlin, 2004, pp. xiv+261.

[BPV01] E. Barucci, S. Polidoro, V. Vespri, Some results on partial differential equations and Asian options. Math. Models Methods Appl. Sci. 11.3 (2001), pp. 475–497.

[BR89] G. Birkhoff and G.-C. Rota. *Ordinary differential equations*. Fourth. John Wiley & Sons, Inc., New York, 1989, pp. xii+399.

[Bra+14] F. Bracci, M.D. Contreras, S. Díaz-Madrigal, A. Vasil'ev, Classical and stochastic Löwner-Kufarev equations. In: *Harmonic and complex analysis and its applications*. Trends Math. Birkhäuser/Springer, Cham, 2014, pp. 39–134.

[Bre14] E. Breuillard, Geometry of locally compact groups of polynomial growth and shape of large balls. Groups Geom. Dyn. 8.3 (2014), pp. 669–732.

[Bro82] R.W. Brockett, Control theory and singular Riemannian geometry. In: *New directions in applied mathematics (Cleveland, Ohio, 1980)*. Springer, New York-Berlin, 1982, pp. 11–27.

[Bry+91] R.L. Bryant, S.S. Chern, R.B. Gardner, H.L. Goldschmidt, P.A. Griffiths. *Exterior differential systems*. Vol. 18. Mathematical Sciences Research Institute Publications. Springer-Verlag, New York, 1991, pp. viii+475.

[BS07] S. Buyalo and V. Schroeder. *Elements of asymptotic geometry*. EMS Monographs in Mathematics. European Mathematical Society (EMS), Zürich, 2007, pp. xii+200.

[BS12] F. Black and M. Scholes, The pricing of options and corporate liabilities [reprint of J. Polit. Econ. **81** (1973), no. 3, 637–654]. In: *Financial risk measurement and management*. Vol. 267. Internat. Lib. Crit. Writ. Econ. Edward Elgar, Cheltenham, 2012, pp. 100–117.

[BS14] S. Buyalo and V. Schroeder, Möbius characterization of the boundary at infinity of rank one symmetric spaces. Geom. Dedicata 172 (2014), pp. 1–45.

[Bul11] M. Buliga, A characterization of sub-Riemannian spaces as length dilation structures constructed via coherent projections. Commun. Math. Anal. 11.2 (2011), pp. 70–111.

[Cap+07] L. Capogna, D. Danielli, S.D. Pauls, J.T. Tyson. *An introduction to the Heisenberg group and the sub-Riemannian isoperimetric problem*. Vol. 259. Progress in Mathematics. Basel: Birkhäuser Verlag, 2007, pp. xvi+223.

[Cap97] L. Capogna, Regularity of quasi-linear equations in the Heisenberg group. Comm. Pure Appl. Math. 50.9 (1997), pp. 867–889.

[Car09] C. Carathéodory, Untersuchungen über die Grundlagen der Thermodynamik. Math. Ann. 67.3 (1909), pp. 355–386.

[Car24] S. Carnot, Réflexions sur la puissance motrice du feu et sur les machines propres à développer cette puissance. French. In: *Parigi, Mallet-Bachelier* (1824).

[Car26] E. Cartan, Sur une classe remarquable d'espaces de Riemann. Bull. Soc. Math. France 54 (1926), pp. 214–264.

[Car27] E. Cartan, Sur une classe remarquable d'espaces de Riemann. II. Bull. Soc. Math. France 55 (1927), pp. 114–134.

[Car45] E. Cartan. *Les systèmes différentiels extérieurs et leurs applications géométriques*. Vol. No. 994. Actualités Scientifiques et Industrielles [Current Scientific and Industrial Topics]. Hermann & Cie, Paris, 1945, p. 214.

[Car72] S. Carnot, Remarks on the motor force of fire and machines ables to develp this power. French. Ann. de l'Éc. Norm. (2) 1 (1872), pp. 393–458.

[CG90] L.J. Corwin and F.P. Greenleaf. *Representations of nilpotent Lie groups and their applications. Part I*. Vol. 18. Cambridge Studies in Advanced Mathematics. Basic theory and examples. Cambridge: Cambridge University Press, 1990, pp. viii+269.

[CGSF23] C. Cuchiero, G. Gazzani, S. Svaluto-Ferro, Signature-based models: theory and calibration. SIAM J. Financial Math. 14.3 (2023), pp. 910–957.

[CH16] Y. Cornulier and P. de la Harpe. *Metric geometry of locally compact groups*. Vol. 25. EMS Tracts in Mathematics. European Mathematical Society (EMS), Zürich, 2016, pp. viii+235.

[Chi+25] Y. Chitour, F. Jean, R. Monti, L. Rifford, L. Sacchelli, M. Sigalotti, A. Socionovo. *Not all sub-Riemannian minimizing geodesics are smooth*. Preprint, arXiv:2501.18920 [math.DG] (2025). 2025.

[Cit+16] G. Citti, B. Franceschiello, G. Sanguinetti, A. Sarti, Sub-Riemannian mean curvature flow for image processing. SIAM J. Imaging Sci. 9.1 (2016), pp. 212–237.

[CJT06] Y. Chitour, F. Jean, E. Trélat, Genericity results for singular curves. J. Differential Geom. 73.1 (2006), pp. 45–73.

[CL16a] L. Capogna and E. Le Donne, Smoothness of subRiemannian isometries. Amer. J. Math. 138.5 (2016), pp. 1439–1454.

[CL16b] I. Chevyrev and T. Lyons, Characteristic functions of measures on geometric rough paths. Ann. Probab. 44.6 (2016), pp. 4049–4082.

[CL55] E.A. Coddington and N. Levinson. *Theory of ordinary differential equations*. McGraw-Hill Book Co., Inc., New York-Toronto-London, 1955, pp. xii+429.

[CO22] I. Chevyrev and H. Oberhauser, Signature moments to characterize laws of stochastic processes. J. Mach. Learn. Res. 23 (2022), Paper No. [176], 42.

[Coc+15] G. Cocci, D. Barbieri, G. Citti, A. Sarti, Cortical spatiotemporal dimensionality reduction for visual grouping. Neural Comput. 27.6 (2015), pp. 1252–1293.

[Cor18] Y. d. Cornulier, On the quasi-isometric classification of locally compact groups. In: *New directions in locally compact groups*. Vol. 447. London Math. Soc. Lecture Note Ser. Cambridge Univ. Press, Cambridge, 2018, pp. 275–342.

[Cow+24] M. Cowling, V. Kivioja, E. Le Donne, S. Nicolussi Golo, A. Ottazzi, From homogeneous metric spaces to Lie groups. C.R. Math. Acad. Sci. Paris 362 (2024), pp. 943–1014.

[CS06] G. Citti and A. Sarti, A cortical based model of perceptual completion in the roto-translation space. J. Math. Imaging Vision 24.3 (2006), pp. 307–326.

[CST23] C. Cuchiero, P. Schmocker, J. Teichmann, Global universal approximation of functional input maps on weighted spaces. In: *Preprint, arXiv:2306.03303* (2023).

[Cuc+22] C. Cuchiero, L. Gonon, L. Grigoryeva, J.-P. Ortega, J. Teichmann, Discrete-time signatures and randomness in reservoir computing. In: *IEEE Trans. Neural Netw. Learn. Syst.* 33.11 (2022), pp. 6321–6330.

[CW16] T. Cohen and M. Welling, Group Equivariant Convolutional Networks. In: *Proceedings of The 33rd International Conference on Machine Learning*. Ed. by M.F. Balcan and K.Q. Weinberger. Vol. 48. Proceedings of Machine Learning Research. New York, New York, USA: PMLR, 2016, pp. 2990–2999.

[D'A08] D. D'Alessandro. *Introduction to quantum control and dynamics*. Chapman & Hall/CRC Applied Mathematics and Nonlinear Science Series. Chapman & Hall/CRC, Boca Raton, FL, 2008, pp. xiv+343.

[dBr85] L. de Branges, A proof of the Bieberbach conjecture. Acta Math. 154.1-2 (1985), pp. 137–152.

[Dek03] K. Dekimpe, On polynomial products in nilpotent and solvable Lie groups. Proc. Amer. Math. Soc. 131.3 (2003), pp. 973–978.

[DER03] N. Dungey, A.F.M. ter Elst, D.W. Robinson. *Analysis on Lie groups with polynomial growth*. Vol. 214. Progress in Mathematics. Birkhäuser Boston, Inc., Boston, MA, 2003, pp. viii+312.

[DF10a] R. Duits and E. Franken, Left-invariant parabolic evolutions on SE 2 and contour enhancement via invertible orientation scores Part I: linear left-invariant diffusion equations on SE(2). Quart. Appl. Math. 68.2 (2010), pp. 255–292.

[DF10b] R. Duits and E. Franken, Left-invariant parabolic evolutions on SE(2) and contour enhancement via invertible orientation scores Part II: non-linear left-invariant diffusions on invertible orientation scores. Quart. Appl. Math. 68.2 (2010), pp. 293–331.

[Dic19] L.E. Dickson, On quaternions and their generalization and the history of the eight square theorem. Ann. of Math. (2) **20**(3), 155–171 (1919)

[DK18] G.C. David, K. Kinneberg, Rigidity for convex-cocompact actions on rank-one symmetric spaces. Geom. Topol. **22**(5), 2757–2790 (2018)

[DK19] G.C. David, B. Kleiner, Rectifiability of planes and Alberti representations. Ann. Sci. Norm. Super. Pisa Cl. Sci. (5) **19**(2), 723–756 (2019)

[DK20] G.C. David, K. Kinneberg, Lipschitz and bi-Lipschitz maps from PI spaces to Carnot groups. Indiana Univ. Math. J. **69**(5), 1685–1731 (2020)

[DP10] D. Dong, I.R. Petersen, Quantum control theory and applications: a survey. IET Control Theory Appl. **4**(12), 2651–2671 (2010)

[DtK18] C. Druţu, M. Kapovich, Geometric group theory, in *American Mathematical Society Colloquium Publications*, vol. 63 (2018). With an appendix by Bogdan Nica, pp. xx+819

[Dui+21] R. Duits, B. Smets, E. Bekkers, J. Portegies, Equivariant deep learning via morphological and linear scale space PDEs on the space of positions and orientations, in *Scale Space and Variational Methods in Computer Vision*, ed. by A. Elmoataz, J. Fadili, Y. Quéau, J. Rabin, L. Simon (Springer International Publishing, Cham, 2021), pp. 27–39

[Dui+23] R. Duits, B.M.N. Smets, A.J. Wemmenhove, J.W. Portegies, E.J. Bekkers, Recent geometric flows in multi-orientation image processing via a Cartan connection, in *Handbook of Mathematical Models and Algorithms in Computer Vision and Imaging: Mathematical Imaging and Vision* (Springer International Publishing, Cham, 2023), pp. 1525–1583

[DW28] D.v. Dantzig, B.L.v.d. Waerden, Über metrisch homogene räume. Abh. Math. Sem. Univ. Hamburg **6**(1), 367–376 (1928)

[Dyn49] E.B. Dynkin, Über die Darstellung der Reihe log $(e^x e^y)$ von nicht vertauschbaren x und y durch Kommutatoren. Russian. Mat. Sb., Nov. Ser. **25**, 155–162 (1949)

[EB+23] S. Eriksson-Bique, C. Gartland, E. Le Donne, L. Naples, S. Nicolussi Golo, Nilpotent groups and bi-Lipschitz embeddings into L^1. English. Int. Math. Res. Not. **2023**(12), 10759–10797 (2023)

[Edw75] D.A. Edwards, The structure of superspace, in *Studies in topology (Proceedings of the Conference, University of North Carolina, Charlotte, N.C., 1974; dedicated to Mathematics of the Polish Academy of Sciences)* (Academic Press, New York, 1975), pp. 121–133

[Etn03] J.B. Etnyre, Introductory lectures on contact geometry, in *Topology and Geometry of Manifolds (Athens, GA, 2001)*, vol. 71. Proceedings of Symposia in Pure Mathematics (American Mathematical Society, Providence, RI, 2003), pp. 81–107

[Fed69] H. Federer. *Geometric Measure Theory*. Die Grundlehren der mathematischen Wissenschaften, Band 153 (Springer-Verlag New York Inc., New York, 1969), pp. xiv+676

[FH20] P.K. Friz, M. Hairer. *A Course on Rough Paths*. Second. Universitext. With an introduction to regularity structures (Springer, Cham, [2020] ©2020), pp. xvi+346

[Fol73] G.B. Folland, A fundamental solution for a subelliptic operator. Bull. Amer. Math. Soc. **79**, 373–376 (1973)

[Fol99] G.B. Folland. *Real analysis*. Second. Pure and Applied Mathematics (New York). Modern techniques and their applications, A Wiley-Interscience Publication. (Wiley, New York, 1999), pp. xvi+386

[FP03] D.R. Farenick, B.A.F. Pidkowich, The spectral theorem in quaternions. Linear Algebra Appl. **371**, 75–102 (2003)

[FP08] P. Foschi and A. Pascucci, Path dependent volatility. Decis. Econ. Finance **31**(1), 13–32 (2008)

[Fre13] D.M. Freeman, Transitive bi-Lipschitz group actions and bi-Lipschitz parametrizations. Indiana Univ. Math. J. **62**(1), 311–331 (2013)

[FV10] P.K. Friz, N.B. Victoir. *Multidimensional Stochastic Processes as Rough Paths*, vol. 120. Cambridge Studies in Advanced Mathematics. Theory and applications (Cambridge University Press, Cambridge, 2010), pp. xiv+656

[Gei08] H. Geiges. *An Introduction to Contact Topology*, vol. 109. Cambridge Studies in Advanced Mathematics. (Cambridge University Press, Cambridge, 2008), pp. xvi+440

[Gia17] V. Gianella, On the asymptotics of the growth of nilpotent groups, in Thesis, ETH (2017)

[Gle52] A.M. Gleason, Groups without small subgroups. Ann. of Math. (2) **56**, 193–212 (1952)

[Gol51] H. Goldstein. *Classical Mechanics* (Addison-Wesley Press, Inc., Cambridge, MA, 1951), pp. xii+399

[Gon98] M.-P. Gong. *Classification of Nilpotent Lie Algebras of Dimension 7 (Over Algebraically Closed Fields and R)*. Thesis (Ph.D.)–University of Waterloo (Canada) (ProQuest LLC, Ann Arbor, MI, 1998), p. 165

[Goo76] R.W. Goodman. *Nilpotent Lie Groups: Structure and Applications to Analysis*. Lecture Notes in Mathematics, vol. 562 (Springer, Berlin, 1976), pp. x+210

[Gro81] M. Gromov, Groups of polynomial growth and expanding maps. Appendix by Jacques Tits. English. Publ. Math., Inst. Hautes Étud. Sci. **53**, 53–78 (1981)

[Gro96] M. Gromov, Carnot-Carathéodory spaces seen from within, in *Sub-Riemannian Geometry*, vol. 144. Progress in Mathematics (Birkhäuser, Basel, 1996), pp. 79–323

[Gro99] M. Gromov. *Metric Structures for Riemannian and non-Riemannian Spaces*, vol. 152. Progress in Mathematics. Based on the 1981 French original, With appendices by M. Katz, P. Pansu, S. Semmes (Birkhäuser Boston Inc., Boston, MA, 1999), pp. xx+585

[Gui73] Y. Guivarc'h, Croissance polynomiale et périodes des fonctions harmoniques. Bull. Soc. Math. France **101**, 333–379 (1973)

[GZ06] J.-P. Gauthier, V. Zakalyukin, On the motion planning problem, complexity, entropy, nonholonomic interpolation. J. Dyn. Control Syst. **12**(3), 371–404 (2006)

[Hak20] E. Hakavuori, Infinite geodesics and isometric embeddings in Carnot groups of step 2. SIAM J. Control Optim. **58**(1), 447–461 (2020)

[Hak+22] E. Hakavuori, V. Kivioja, T. Moisala, F. Tripaldi, Gradings for nilpotent Lie algebras. J. Lie Theory **32**(2), 383–412 (2022)

[Ham89] U. Hamenstädt, A new description of the Bowen-Margulis measure. Ergodic Theory Dynam. Systems **9**(3), 455–464 (1989)

[Hau14] F. Hausdorff. *Grundzüge der Mengenlehre.* German. (Veit & Company, Leipzig, 1914)

[Hei01] J. Heinonen. *Lectures on Analysis on Metric Spaces*. Universitext (Springer, New York, 2001), pp. x+140

[Hei74] E. Heintze, On homogeneous manifolds of negative curvature. Math. Ann. **211**, 23–34 (1974)

[Hel01] S. Helgason. *Differential Geometry, Lie Groups, and Symmetric Spaces*, vol. 34. Graduate Studies in Mathematics. Corrected reprint of the 1978 original (American Mathematical Society, Providence, RI, 2001), pp. xxvi+641

[HL10] B. Hambly, T. Lyons, Uniqueness for the signature of a path of bounded variation and the reduced path group. Ann. of Math. (2) **171**(1), 109–167 (2010)

[HL16] E. Hakavuori, E. Le Donne, Non-minimality of corners in subriemannian geometry. Invent. Math. **206**(3), 693–704 (2016)

[HL23] E. Hakavuori, E. Le Donne, Blowups and blowdowns of geodesics in Carnot groups. J. Differential Geom. **123**(2), 267–310 (2023)

[HN12] J. Hilgert, K.-H. Neeb, *Structure and Geometry of Lie Groups*. Springer Monographs in Mathematics (Springer, New York, 2012), pp. x+744

[Hoc65] G. Hochschild. *The Structure of Lie Groups* (Holden-Day, Inc., San Francisco-London-Amsterdam, 1965), pp. ix+230

[Hof70] W.C. Hoffman, Higher visual perception as prolongation of the basic Lie transformation group. Math. Biosci. **6**, 437–471 (1970)

[HP97] S. Hersonsky, F. Paulin, On the rigidity of discrete isometry groups of negatively curved spaces. Comment. Math. Helv. **72**(3), 349–388 (1997)

[HR98] D.G. Hobson, L.C.G. Rogers, Complete models with stochastic volatility. Math. Finance **8**(1), 27–48 (1998)

[Hur22] A. Hurwitz, Über die Komposition der quadratischen Formen. Math. Ann. **88**(1-2), 1–25 (1922)

[Jac79] N. Jacobson. *Lie Algebras*. Republication of the 1962 original. (Dover Publications Inc., New York, 1979), pp. ix+331

[Jea14] F. Jean. *Control of Nonholonomic Systems: From sub-Riemannian Geometry to Motion Planning*. Springer Briefs in Mathematics (Springer, Cham, 2014), pp. x+104

[Jen73a] J.W. Jenkins, A characterization of growth in locally compact groups. Bull. Am. Math. Soc. **79**, 103–106 (1973)

[Jen73b] J.W. Jenkins, Growth of connected locally compact groups. J. Funct. Anal. **12**, 113–127 (1973)

[Joh+00] W.B. Johnson, J. Lindenstrauss, D. Preiss, G. Schechtman, Uniform quotient mappings of the plane. Michigan Math. J. **47**(1), 15–31 (2000)

[Jur93] V. Jurdjevic, The geometry of the plate-ball problem. Arch. Rational Mech. Anal. **124**(4), 305–328

[Kel75] J.L. Kelley, *General Topology*, vol. 27. Graduate Texts in Mathematics. Reprint of the 1955 edition [Van Nostrand, Toronto, Ont.] (Springer, New York, 1975), pp. xiv+298

[Kim09] S. Kim, On a degenerate parabolic equation arising in pricing of Asian options. J. Math. Anal. Appl. **351**(1), 326–333 (2009)

[KL17] V. Kivioja, E. Le Donne, Isometries of nilpotent metric groups. J. Éc. polytech. Math. **4**, 473–482 (2017)

[KLN22] V. Kivioja, E. Le Donne, S. Nicolussi Golo, Metric equivalences of Heintze groups and applications to classifications in low dimension. Illinois J. Math. **66**(1), 91–121 (2022)

[Kna02] A.W. Knapp. *Lie Groups Beyond an Introduction*. Second, vol. 140. Progress in Mathematics (Birkhäuser Boston Inc., Boston, MA, 2002), pp. xviii+812

[KO19] F.J. Király, H. Oberhauser, Kernels for sequentially ordered data. J. Mach. Learn. Res. **20**, Paper No. 31, 45 (2019)

[Kob62] S. Kobayashi, Homogeneous Riemannian manifolds of negative curvature. Tohoku Math. J. (2) **14**, 413–415 (1962)

[KR65] K. Kuratowski, C. Ryll-Nardzewski, A general theorem on selectors. Bull. Acad. Polon. Sci. Sér. Sci. Math. Astronom. Phys. **13**, 397–403 (1965)

[Kur35] C. Kuratowski, Quelques problèmes concernant les espaces métriques non-séparables. French. Fundam. Math. **25**, 534–545 (1935)

[Law04] G.F. Lawler, An introduction to the stochastic Loewner evolution, in *Random Walks and Geometry* (Walter de Gruyter, Berlin, 2004), pp. 261–293

[Laz55] M. Lazard, Sur le nilpotence de certains groupes algébriques. French. C. R. Acad. Sci., Paris **241**, 1687–1689 (1955)

[LD11a] E. Le Donne, Geodesic manifolds with a transitive subset of smooth biLipschitz maps. Groups Geom. Dyn. **5**(3), 567–602 (2011)

[LD11b] E. Le Donne, Metric spaces with unique tangents. Ann. Acad. Sci. Fenn. Math. **36**(2), 683–694 (2011)

[LD15] E. Le Donne, A metric characterization of Carnot groups. Proc. Am. Math. Soc. **143**(2), 845–849 (2015)

[LD17] E. Le Donne, A primer on Carnot groups: homogenous groups, Carnot-Carathéodory spaces, and regularity of their isometries. Anal. Geom. Metr. Spaces **5**(1), 116–137 (2017)

[LD+16] E. Le Donne, R. Montgomery, A. Ottazzi, P. Pansu, D. Vittone, Sard property for the endpoint map on some Carnot groups. Ann. Inst. H. Poincaré Anal. Non Linéaire **33**(6), 1639–1666 (2016)

[Lee13] J.M. Lee. *Introduction to Smooth Manifolds*. Second, vol. 218. Graduate Texts in Mathematics (Springer, New York, 2013), pp. xvi+708

[Lee18] J.M. Lee. *Introduction to Riemannian manifolds*. English, 2nd edn, vol. 176. Graduate Texts in Mathematics (Springer, Cham, 2018)

[LM08] G.P. Leonardi, R. Monti, End-point equations and regularity of sub-Riemannian geodesics. Geom. Funct. Anal. **18**(2), 552–582 (2008)

[LM10] E. Le Donne, V. Magnani, Measure of submanifolds in the Engel group. Rev. Mat. Iberoam. **26**(1), 333–346 (2010)

[LN18] E. Le Donne, S. Nicolussi Golo, Regularity properties of spheres in homogeneous groups. Trans. Am. Math. Soc. **370**(3), 2057–2084 (2018)

[LN21] E. Le Donne, S. Nicolussi Golo, Metric Lie groups admitting dilations. Ark. Mat. **59**(1), 125–163 (2021)

[LNP19] T. Lyons, S. Nejad, I. Perez Arribas, Numerical method for model-free pricing of exotic derivatives in discrete time using rough path signatures. Appl. Math. Finance **26**(6), 583–597 (2019)

[LNP20] T. Lyons, S. Nejad, I. Perez Arribas, Non-parametric pricing and hedging of exotic derivatives. Appl. Math. Finance **27**(6), 457–494 (2020)

[LO16] E. Le Donne, A. Ottazzi, Isometries of Carnot Groups and Sub-Finsler Homogeneous Manifolds. J. Geom. Anal. **26**(1), 330–345 (2016)

[Los87] V. Losert, On the structure of groups with polynomial growth. English. Math. Z. **195**, 109–117 (1987)
[LOW14] E. Le Donne, A. Ottazzi, B. Warhurst, Ultrarigid tangents of sub-Riemannian nilpotent groups. Ann. Inst. Fourier (Grenoble) **64**(6), 2265–2282 (2014)
[LPS24] E. Le Donne, N. Paddeu, A. Socionovo, *Metabelian Distributions and sub-Riemannian Geodesics* (2024). arXiv: 2405. 14997 [math.DG]
[LT22] E. Le Donne, F. Tripaldi, A cornucopia of Carnot groups in low dimensions. Anal. Geom. Metr. Spaces **10**(1), 155–289 (2022)
[LY23] E. Le Donne, R. Young, Carnot rectifiability of sub-Riemannian manifolds with constant tangent. Ann. Sc. Norm. Super. Pisa Cl. Sci. (5) **24**(1), 71–96 (2023)
[Ma91] D. Ma, Boundary behavior of invariant metrics and volume forms on strongly pseudo-convex domains. Duke Math. J. **63**(3), 673–697 (1991)
[Mac88] I.D. Macdonald, *The Theory of Groups*. Reprint of the 1968 original (Robert E. Krieger Publishing Co. Inc., Malabar, FL, 1988), pp. viii+254
[Mag08] V. Magnani, Non-horizontal submanifolds and coarea formula. J. Anal. Math. **106**, 95–127 (2008)
[Mag10] V. Magnani, Blow-up estimates at horizontal points and applications. English. J. Geom. Anal. **20**(3), 705–722 (2010)
[Mag12] F. Maggi, *Sets of Finite Perimeter and Geometric Variational Problems. An Introduction to Geometric Measure Theory*. English, vol. 135. Cambridge Studies in Advanced Mathematics (Cambridge University Press, Cambridge, 2012)
[Mag19] V. Magnani, Towards a theory of area in homogeneous groups. Calc. Var. Partial Differential Equations **58**(3), Paper No. 91, 39 (2019)
[Mal51] A.I. Malcev, On a class of homogeneous spaces. Am. Math. Soc. Translation **1951**(39), 33 (1951)
[Man12] A. Mann. *How Groups Grow.* English. London Mathematical Society Lecture Note Series, vol. 395 (Cambridge University Press, Cambridge, 2012)
[MAS13] A.P. Mashtakov, A.A. Ardentov, Y.L. Sachkov, Parallel algorithm and software for image inpainting via sub-Riemannian minimizers on the group of rototranslations. Numer. Math. Theory Methods Appl. **6**(1), 95–115 (2013)
[Mas+17] A.P. Mashtakov, R. Daĭts, Y.L. Sachkov, E. Bekkers, I.Y. Beschastnyĭ, Sub-Riemannian geodesics on the group SO 3 in a problem on tracking blood vessels on spherical images of the retina. Dokl. Akad. Nauk **473**(5), 521–524 (2017)
[Mat95] P. Mattila. *Geometry of Sets and Measures in Euclidean Spaces*, vol. 44. Cambridge Studies in Advanced Mathematics. Fractals and rectifiability. (Cambridge University Press, Cambridge, 1995), pp. xii+343
[MB00] A. Marigo, A. Bicchi, Rolling bodies with regular surface: controllability theory and applications. IEEE Trans. Automat. Control **45**(9), 1586–1599 (2000)
[Mil76] J. Milnor, Curvatures of left invariant metrics on Lie groups. Advances in Math. **21**(3), 293–329 (1976)
[Mil97] J.M. Milnor. *Topology from the Differentiable Viewpoint. Based on Notes by David W. Weaver.* English. Revised 2nd ed. (Princeton University Press, Princeton, NJ:, 1997)
[Mit85] J. Mitchell, On Carnot-Carathéodory metrics. J. Differential Geom. **21**(1), 35–45 (1985)
[MLS94] R.N. Murray, Z.X. Li, S.S. Sastry. *A Mathematical Introduction to Robotic Manipulation* (CRC Press, Boca Raton, FL, 1994), pp. xx+456
[MM95] G.A. Margulis, G.D. Mostow, The differential of a quasi-conformal mapping of a Carnot-Carathéodory space. Geom. Funct. Anal. **5**(2), 402–433 (1995)
[Mon01] R. Monti, Distances, boundaries and surface measures in Carnot-Caratheodory spaces, in *Phd Thesis, University of Trento* (2001)
[Mon02] R. Montgomery. *A Tour of Subriemannian Geometries, their Geodesics and Applications*, vol. 91. Mathematical Surveys and Monographs (American Mathematical Society, Providence, RI, 2002), pp. xx+259
[Mon94] R. Montgomery, Abnormal minimizers. SIAM J. Control Optim. **32**(6), 1605–1620 (1994)

[Mon99] R. Montgomery, Engel deformations and contact structures, in *Northern California Symplectic Geometry Seminar*, vol. 196. American Mathematical Society Translations, Series 2 (American Mathematical Society, Providence, RI, 1999), pp. 103–117

[Mor88] F. Morgan. *Geometric Measure Theory. A Beginner's Guide*. English (Academic Press, Inc., Boston, MA, 1988)

[MPV07] I. Markina, D. Prokhorov, A. Vasil′ev, Sub-Riemannian geometry of the coefficients of univalent functions. J. Funct. Anal. **245**(2), 475–492 (2007)

[MPV18a] R. Monti, A. Pigati, D. Vittone, Existence of tangent lines to Carnot-Carathéodory geodesics. English. Calc. Var. Partial Differ. Equ. **57**(3), Id/No 75, pp. 1–18 (2018)

[MPV18b] R. Monti, A. Pigati, D. Vittone, On tangent cones to length minimizers in Carnot-Carathéodory spaces. English. SIAM J. Control Optim. **56**(5), 3351–3369 (2018)

[MS39] S.B. Myers, N.E. Steenrod, The group of isometries of a Riemannian manifold. Ann. of Math. (2) **40**(2), 400–416 (1939)

[Mum94] D. Mumford, Elastica and computer vision, in *Algebraic Geometry and Its Applications (West Lafayette, IN, 1990)* (Springer, New York, 1994), pp. 491–506

[Mun75] J.R. Munkres. *Topology: A First Course* (Prentice-Hall Inc., Englewood Cliffs, N.J., 1975), pp. xvi+413

[MV08] V. Magnani, D. Vittone, An intrinsic measure for submanifolds in stratified groups. J. Reine Angew. Math. **619**, 203–232 (2008)

[MZ52] D. Montgomery, L. Zippin, Small subgroups of finite-dimensional groups. Ann. of Math. (2) **56**, 213–241 (1952)

[MZ74] D. Montgomery, L. Zippin. *Topological Transformation Groups*. Reprint of the 1955 original (Robert E. Krieger Publishing Co., Huntington, N.Y., 1974), pp. xi+289

[NB11] L. Narici, E. Beckenstein. *Topological Vector Spaces*. Second, vol. 296. Pure and Applied Mathematics (Boca Raton) (CRC Press, Boca Raton, FL, 2011), pp. xviii+610

[NSW85] A. Nagel, E.M. Stein, S. Wainger, Balls and metrics defined by vector fields. I. Basic properties. Acta Math. **155**(1-2), pp. 103–147 (1985)

[O'R97] D. O'Regan, *Existence Theory for Nonlinear Ordinary Differential Equations*, vol. 398. Mathematics and its Applications (Kluwer Academic Publishers Group, Dordrecht, 1997), pp. vi+196

[Oss78] R. Osserman, The isoperimetric inequality. Bull. Am. Math. Soc. **84**(6), 1182–1238 (1978)

[Pan83] P. Pansu, Croissance des boules et des géodésiques fermées dans les nilvariétés. Ergodic Theory Dynam. Systems **3**(3), 415–445 (1983)

[Pan89] P. Pansu, Métriques de Carnot-Carathéodory et quasiisométries des espaces symétriques de rang un. Ann. of Math. (2) **129**(1), 1–60 (1989)

[Par03] J.R. Parker, *Notes on Complex Hyperbolic Geometry* (University of Durham, Durham, 2003). maths.dur.ac.uk/users/j.r.parker/img/NCHG.pdf

[Pas05] A. Pascucci, Kolmogorov equations in physics and in finance, in *Elliptic and Parabolic Problems*, vol. 63. Progress in Nonlinear Differential Equations and their Applications (Birkhäuser, Basel, 2005), pp. 353–364

[Pas08] A. Pascucci, Obstacle problem and optimal stopping with applications in finance, in *"Bruno Pini" Mathematical Analysis Seminar, University of Bologna Department of Mathematics: Academic Year 2006/2007 (Italian)* (Tecnoprint, Bologna, 2008), pp. 47–65

[Pas11] A. Pascucci. *PDE and Martingale Methods in Option Pricing*, vol. 2. Bocconi & Springer Series (Springer, Milan; Bocconi University Press, Milan, 2011), pp. xviii+719

[Pet17] J. Petitot. *Elements of Neurogeometry*. Lecture Notes in Morphogenesis. Functional architectures of vision, Translated from the 2008 French edition [MR3077550] by S. Lyle (Springer, Cham, 2017), pp. xv+379

[Pil22] A. Pilastro, Sub-Riemannian structures of strictly pseudoconvex quaternionic domains, in *Tesi di laurea Magistrale, University of Pisa* (2022)

[Pon66] L.S. Pontryagin, *Topological Groups*. Translated from the second Russian edition by A. Brown (Gordon and Breach Science Publishers, Inc., New York, 1966), pp. xv+543

[PT99] J. Petitot, Y. Tondut, Vers une neurogéométrie. Fibrations corticales, structures de contact et contours subjectifs modaux. Math. Inform. Sci. Humaines **145**, 5–101 (1999)

[Pur23] V. Purho, Invariant Metrics on Lie Groups, in *Master thesis, University of Jyväskylä* (2023)

[Rag72] M.S. Raghunathan. *Discrete Subgroups of Lie Groups*. Ergebnisse der Mathematik und ihrer Grenzgebiete, Band 68 (Springer, New York, 1972), pp. ix+227

[Rif14] L. Rifford. *Sub-Riemannian Geometry and Optimal Transport*. Springer Briefs in Mathematics (Springer, Cham, 2014), pp. viii+140

[RS76] L.P. Rothschild, E.M. Stein, Hypoelliptic differential operators and nilpotent groups. Acta Math. **137**(3-4), 247–320 (1976)

[Sac08] Y.L. Sachkov, Maxwell strata in the Euler elastic problem. J. Dyn. Control Syst. **14**(2), 169–234 (2008)

[Sac10] Y.L. Sachkov, Maxwell strata and symmetries in the problem of optimal rolling of a sphere over a plane. Sb. Math. **201**(7-8), 1029–1051 (2010)

[Sal+21] C. Salvi, T. Cass, J. Foster, T. Lyons, W. Yang, The signature kernel is the solution of a Goursat PDE. SIAM J. Math. Data Sci. **3**(3), 873–899 (2021)

[SC15] A. Sarti, G. Citti, The constitution of visual perceptual units in the functional architecture of V1. J. Comput. Neurosci. **38**(2), 285–300 (2015)

[Sch00] O. Schramm, Scaling limits of loop-erased random walks and uniform spanning trees. Israel J. Math. **118**, 221–288 (2000)

[Sch07] O. Schramm, Conformally invariant scaling limits: an overview and a collection of problems, in *International Congress of Mathematicians*, vol. I (European Mathematical Society, Zürich, 2007), pp. 513–543

[SCP08] A. Sarti, G. Citti, J. Petitot, The symplectic structure of the primary visual cortex. Biol. Cybernet. **98**(1), 33–48 (2008)

[Sem96] S. Semmes, On the nonexistence of bi-Lipschitz parameterizations and geometric problems about A_∞-weights. Rev. Mat. Iberoamericana **12**(2), 337–410 (1996)

[Sie86] E. Siebert, Contractive automorphisms on locally compact groups. Math. Z. **191**(1), 73–90 (1986)

[SM14] L. Sifre, S. Mallat, Rigid-Motion Scattering for Texture Classification, in *Preprint, arXiv* abs/1403.1687 (2014)

[Sop23] C. Sopio, Milnor Theorems for sub-Finsler Lie Groups, in *Master thesis, University of Pisa* (2023)

[SS50] A.C. Schaeffer, D.C. Spencer. *Coefficient Regions for Schlicht Functions*, vol. 35. American Mathematical Society Colloquium Publications. With a Chapter on the Region of the Derivative of a Schlicht Function by Arthur Grad (American Mathematical Society, New York, 1950), pp. xv+311

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

[Ste93] E.M. Stein, *Harmonic Analysis: Real-Variable Methods, Orthogonality, and Oscillatory Integrals*, vol. 43. Princeton Mathematical Series. With the assistance of Timothy S. Murphy, Monographs in Harmonic Analysis, III (Princeton University Press, Princeton, NJ, 1993), pp. xiv+695

[Sto98] M. Stoll, On the asymptotics of the growth of 2-step nilpotent groups. J. London Math. Soc. (2) **58**(1), 38–48 (1998)

[Sus14] H.J. Sussmann, A regularity theorem for minimizers of real-analytic subriemannian metrics, in *2014 IEEE 53rd Annual Conference on Decision and Control (CDC)* (2014), pp. 4801–4806

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

[Tao14] T. Tao. *Hilbert's Fifth Problem and Related Topics*, vol. 153. Graduate Studies in Mathematics (American Mathematical Society, Providence, RI, 2014), pp. xiv+338

[Tay34] A.E. Taylor, On integral invariants of non-holonomic dynamical systems. Bull. Am. Math. Soc. **40**(10), 735–742 (1934)

[Tré12] E. Trélat, Optimal control and applications to aerospace: some results and challenges. J. Optim. Theory Appl. **154**(3), 713–758 (2012)

[TY13] K. Tan, X. Yang, Subriemannian geodesics of Carnot groups of step 3. ESAIM Control Optim. Calc. Var. **19**(1), 274–287 (2013)

[Ver70] M. Vergne, Cohomologie des algèbres de Lie nilpotentes. Application à l'étude de la variété des algèbres de Lie nilpotentes. Bull. Soc. Math. France **98**, 81–116 (1970)

[Vod07] S.K. Vodopyanov, Geometry of Carnot-Carathéodory spaces and differentiability of mappings, in *The Interaction of Analysis and Geometry*, vol. 424 (Contemporary Mathematics American Mathematical Society Providence, RI, 2007), pp. 247–301

[Wag40] V. Wagner, Differential geometry of non-linear non-holonomic manifolds in the three-dimensional Euclidean space. Rec. Math. [Mat. Sbornik] N.S. **8**(50), 3–39 (1940)

[War83] F.W. Warner. *Foundations of Differentiable Manifolds and Lie Groups*, vol. 94. Graduate Texts in Mathematics. Corrected reprint of the 1971 edition (Springer, New York, 1983), pp. ix+272

[Wil82] E.N. Wilson, Isometry groups on homogeneous nilmanifolds. Geom. Dedicata **12**(3), 337–346 (1982)

[WM10] H.M. Wiseman, G.J. Milburn. *Quantum Measurement and Control* (Cambridge University Press, Cambridge, 2010), pp. xvi+460

[Wol62] J.A. Wolf, On locally symmetric spaces of non-negative curvature and certain other locally homogeneous spaces. Comment. Math. Helv. **37**, 266–295 (1962/1963)

[Wol64] J.A. Wolf, Homogeneity and bounded isometries in manifolds of negative curvature. Illinois J. Math. **8**, 14–18 (1964)

[Wol68] J.A. Wolf, Growth of finitely generated solvable groups and curvature of Riemanniann manifolds. J. Differential Geometry **2**, 421–446 (1968)

[ZB03] D.V. Zenkov, A.M. Bloch, Invariant measures of nonholonomic flows with internal degrees of freedom. Nonlinearity **16**(5), 1793–1807 (2003)

Index

E. Le Donne, *Metric Lie Groups*, Graduate Texts in Mathematics 306,
https://doi.org/10.1007/978-3-031-98832-5