# A Comprehensive Introduction to sub-Riemannian Geometry

*from Hamiltonian viewpoint*

ANDREI AGRACHEV

DAVIDE BARILARI

UGO BOSCAIN

February 6, 2019

Draft version

# Preface

This book presents material taught by the authors in graduated courses at Trieste (SISSA), Paris (Institut Henri Poincaré, Orsay, Paris Diderot), and several summer schools, in the period 2008 – 2018.

It contains material for an introductory course in sub-Remannan geometry at master or PhD level, as well as material for a more advanced course.

The book attempts to be as elementary as possible but, although the main concepts are recalled, it requires a certain ability in managing object in differential geometry (vector fields, differential forms, symplectic manifolds, etc.). We try to avoid as much as possible the use of functional analysis (some is required starting from Chapter 6).

We do not require any knowledge in Riemannian geometry. Actually from the book one can extract an introductory course in Riemannian geometry as a special case of sub-Riemannian one, starting from the geometry of surfaces in Chapter 1.

There are few other books of sub-Riemannian geometry available. Besides the pioneering book edited by A. Bellaïche and J.-J. Risler [BR96], a nowadays classical reference is the book of R. Montgomery [Mon02], that inspired several of our chapters. More recent books, written in a language similar to the one we use, are those of F. Jean [Jea14] and L. Rifford [Rif14]; see also the collection of lectures notes [BBS16a, BBS16b]. Other related books, although with a different approach, are the monographs [BLU07] and [CDPT07].

**Example of an introductory course of sub-Riemannan geometry.**
Chapters 2, 3 (without the appendices), 4, 7 (without 7.1), 9, 13, 21.

**Example of an advanced course of sub-Riemannan geometry.**
Chapters 2, 3 (with the appendices), 4, 6, 7 (together with 7.1), 8, 9, 10, 11, 12, 13, 14, 15, 17, 18, 19, 20, 21, the appendix by Zelenko.

**Example of a course of Riemannan geometry.**
Chapters 1, 2, 3 (without the appendices), 4, 5, 7, 8, 11, 14 (without 14.4 -14.5 -14.6), 15, 16, 21 (only 21.1).

*Acknowledgments.* A special thanks goes to Luca Rizzi and Mario Sigalotti, who contributed with several crucial corrections and suggestions on different topics of the book.

We are grateful to all the colleagues that helped us in the final reading of the manuscript, Ivan Beschastnyi, Francesco Boarotto, Daniele Cannarsa, Francesca Chittaro, Valentina Franceschi, Roberta Ghezzi, Frédéric Jean, Antonio Lerario, Paolo Mason, Eugenio Pozzoli, Dario Prandi, Ludovic Sacchelli, Yuri Sachkov, and Michele Stecconi.

We are also grateful to the colleagues that during the years gave us suggestions to improve both the content and the exposition: Carolina Biolo, Andrea Bonfiglioli, Jean-Baptiste Caillau, Jean-Paul Gauthier, Moussa Gaye, Velimir Jurdjevic, Andrea Mondino, Richard Montgomery, Roberto Monti, Elisa Paoli.

This project has been supported by the European Research Council: ERC StG 2009 "GeCoMethods", project number 239748 and ERC POC 2016 "ARTIV1", project number 727283. Also it was supported by the ANR project SRGI "Sub-Riemannian Geometry and Interactions", contract number ANR-15-CE40-0018.

# Contents

# Introduction

This book concerns a fresh development of the eternal idea of the distance as the length of a shortest path. In Euclidean geometry, shortest paths are segments of straight lines that satisfy all classical axioms. In the Riemannian world, Euclidean geometry is just one of a huge amount of possibilities. However, each of these possibilities is well approximated by Euclidean geometry at very small scale. In other words, Euclidean geometry is treated as geometry of initial velocities of the paths starting from a fixed point of the Riemannian space rather than the geometry of the space itself.

The Riemannian construction was based on the previous study of smooth surfaces in the Euclidean space undertaken by Gauss. The distance between two points on the surface is the length of a shortest path on the surface connecting the points. Initial velocities of smooth curves starting from a fixed point on the surface form a tangent plane to the surface, that is an Euclidean plane. Tangent planes at two different points are isometric, but neighborhoods of the points on the surface are not locally isometric in general; certainly not if the Gaussian curvature of the surface is different at the two points.

Riemann generalized Gauss' construction to higher dimensions and realized that it can be done in an intrinsic way; you do not need an ambient Euclidean space to measure the length of curves. Indeed, to measure the length of a curve it is sufficient to know the Euclidean length of its velocities. A Riemannian space is a smooth manifold whose tangent spaces are endowed with Euclidean structures; each tangent space is equipped with its own Euclidean structure that smoothly depends on the point where the tangent space is attached.

For a habitant sitting at a point of the Riemannian space, tangent vectors give directions where to move or, more generally, to send and receive information. He measures lengths of vectors, and angles between vectors attached at the same point, according to the Euclidean rules, and this is essentially all what he can do. It is important that our habitant can, in principle, completely recover the geometry of the space by performing these simple measurements along different curves.

In the sub-Riemannian space we cannot move, receive and send information in all directions. There are restrictions (imposed by the God, the moral imperative, the government, or simply a physical law). A sub-Riemannian space is a smooth manifold with a fixed admissible subspace in any tangent space where admissible subspaces are equipped with Euclidean structures. Admissible paths are those curves whose velocities are admissible. The distance between two points is the infimum of the length of admissible paths connecting the points. It is assumed that any pair of points in the same connected component of the manifold can be connected by at least an admissible path. The last assumption might look strange at a first glance, but it is not. The admissible subspace depends on the point where it is attached, and our assumption is satisfied for a more or less general smooth dependence on the point; better to say that it is not satisfied only for very special families of admissible subspaces.

Let us describe a simple model. Let our manifold be $\mathbb{R}^3$ with coordinates $x, y, z$. We consider the differential 1-form $\omega = -dz + \frac{1}{2}(xdy - ydx)$. Then $d\omega = dx \wedge dy$ is the pullback on $\mathbb{R}^3$ of the area form on the $xy$-plane. In this model the subspace of admissible velocities at the point $(x, y, z)$ is assumed to be the kernel of the form $\omega$. In other words, a curve $t \mapsto (x(t), y(t), z(t))$ is an admissible path if and only if $\dot{z}(t) = \frac{1}{2}(x(t)\dot{y}(t) - y(t)\dot{x}(t))$ or equivalently if

$$z(t) = z(0) + \frac{1}{2}\int_0^t \big(x(s)\dot{y}(s) - y(s)\dot{x}(s)\big)\, ds.$$

If $x(0) = y(0) = z(0) = 0$, then $z(t)$ is the signed area of the domain bounded by the curve and the segment connecting $(0, 0)$ with $(x(t), y(t))$.

In this geometry, the length of an admissible tangent vector $(\dot{x}, \dot{y}, \dot{z})$ is defined to be $(\dot{x}^2 + \dot{y}^2)^{\frac{1}{2}}$, that is the length of the projection of the vector to the $xy$-plane. By construction, the sub-Riemannian length of the admissible curve in $\mathbb{R}^3$ is equal to the Euclidean length of its projection to the plane.

In this geometry, to compute the shortest paths connecting the origin $(0, 0, 0)$ to a fixed point $(x_1, y_1, z_1)$ we are then reduced to solve the classical Dido isoperimetric problem: find a shortest planar curve among those connecting $(0, 0)$ with $(x_1, y_1)$ and such that the signed area of the domain bounded by the curve and the segment joining $(0, 0)$ and $(x_1, y_1)$ is equal to $z_1$ (see Figure 1).

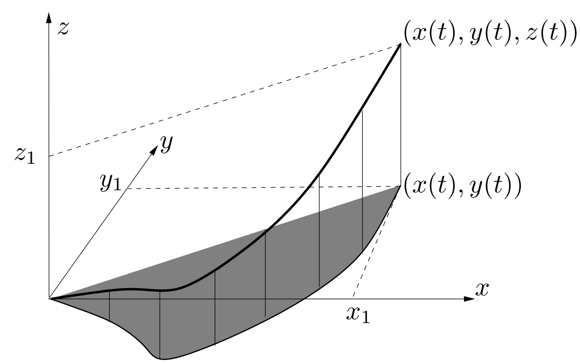



Figure 1: The Dido problem

Solutions of the Dido problem are arcs of circles and their lifts to $\mathbb{R}^3$ are spirals where $z(t)$ is the area of the piece of disc cut by the hord connecting $(0, 0)$ with $(x(t), y(t))$ (see Figure 2).

A piece of such a spiral is a shortest admissible path between its endpoints while the planar projection of this piece is an arc of the circle. The spiral ceases to be a shortest path when its planar projection starts to run the circle for the second time, i.e., when the spiral starts its second turn. Sub-Riemannian balls centered at the origin for this model look like apples with singularities at the poles (see Figure 3).

Singularities are points on the sphere connected with the center by more than one shortest path. The dilation $(x, y, z) \mapsto (rx, ry, r^2z)$ transforms the ball of radius 1 into the ball of radius $r$. In particular, arbitrary small balls have singularities. This is always the case when admissible subspaces are proper subspaces.

Another important symmetry connects balls with different centers. Indeed, the product operation

$$(x, y, z) \cdot (x', y', z') \doteq \left(x + x',\ y + y',\ z + z' + \frac{1}{2}(xy' - x'y)\right)$$

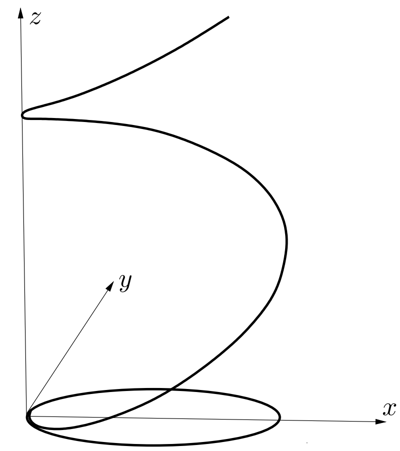



Figure 2: Solutions to the Dido problem

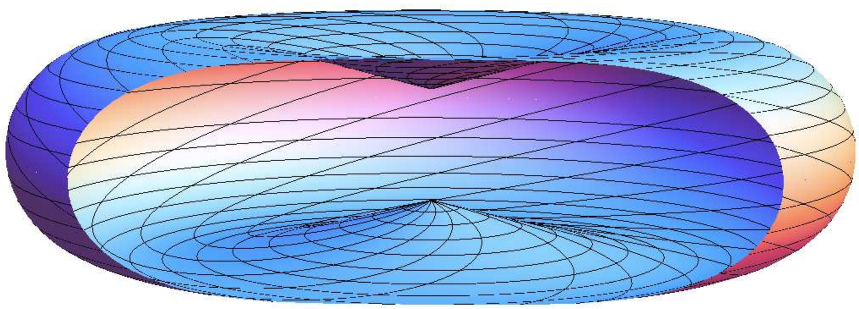


Figure 3: The Heisenberg sub-Riemannian sphere

turns $\mathbb{R}^3$ into a group, the *Heisenberg group*. The origin in $\mathbb{R}^3$ is the unit element of this group. It is easy to see that left-translations of the group transform admissible curves into admissible ones and preserve the sub-Riemannian length. Hence left translations transform balls in balls of the same radius. A detailed description of this example and other models of sub-Riemannian spaces is done in Sections 4.4.3, 7.5.1, 13.2.

Actually, even this simplest model tells us something about life in a sub-Riemannian space. Here we deal with planar curves but, in fact, operate in the three-dimensional space. Sub-Riemannian spaces always have a kind of hidden extra dimension. A good and not yet exploited source for mystic speculations but also for theoretical physicists who are always searching new crazy formalizations. In mechanics, this is a natural geometry for systems with nonholonomic constraints like skates, wheels, rolling balls, bearings etc. This kind of geometry could also serve to model social behavior that allows to increase the level of freedom without violation of a restrictive legal system.

Anyway, in this book we perform a purely mathematical study of sub-Riemannian spaces to provide an appropriate formalization ready for all potential applications. Riemannian spaces appear as a very special case. Of course, we are not the first to study the sub-Riemannian stuff. There is a broad literature even if there are not so many experts who could claim that sub-Riemannian geometry is his main field of expertise. Important motivations come from CR geometry, hyperbolic geometry, analysis of hypoelliptic operators, and some other domains. Our first motivation was control theory: length minimizing is a nice class of optimal control problems.

Indeed, one can find a control theory spirit in our treatment of the subject. First of all, we include admissible paths in admissible flows that are flows generated by vector fields whose values in all points belong to admissible subspaces. The passage from admissible subspaces attached at different points of the manifold to a globally defined space of admissible vector fields makes the structure more flexible and well-adapted to algebraic manipulations. We pick generators $f_1, \ldots, f_k$ of the space of admissible fields, and this allows us to describe all admissible paths as solutions to time-varying ordinary differential equations of the form: $\dot{q}(t) = \sum_{i=1}^{k} u_i(t) f_i(q(t))$. Different admissible paths correspond to the choice of different *control functions* $u_i(\cdot)$ and initial points $q(0)$ while the vector fields $f_i$ are fixed at the very beginning.

We also use a Hamiltonian approach supported by the Pontryagin maximum principle to characterize shortest paths. Few words about the Hamiltonian approach: sub-Riemannian geodesics are admissible paths whose sufficiently small pieces are length-minimizers, i. e. the length of such a piece is equal to the distance between its endpoints. In the Riemannian setting, any geodesic is uniquely determined by its velocity at the initial point $q$. In the general sub-Riemannian situation we have much more geodesics based at the the point $q$ than admissible velocities at $q$. Indeed, every point in a neighborhood of $q$ can be connected with $q$ by a length-minimizer, while the dimension of the admissible velocities subspace at $q$ is usually smaller than the dimension of the manifold.

What is a natural parametrization of the space of geodesics? To understand this question, we adapt a classical "trajectory – wave front" duality. Given a length-parameterized geodesic $t \mapsto \gamma(t)$, we expect that the values at a fixed time $t$ of geodesics starting at $\gamma(0)$ and close to $\gamma$ fill a piece of a smooth hypersurface (see Figure 4). For small $t$ this hypersurface is a piece of the sphere of radius $t$, while in general it is only a piece of the "wave front".

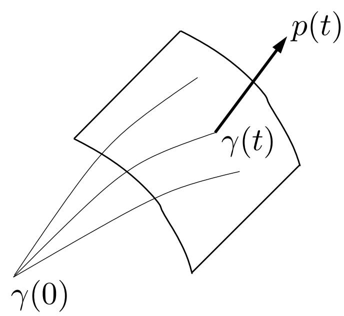



Figure 4: The "wave front" and the "impulse"

Moreover, we expect that $\dot{\gamma}(t)$ is transversal to this hypersurface. It is not always the case but this is true for a generic geodesic.

The "impulse" $p(t) \in T^*_{\gamma(t)}M$ is the covector orthogonal to the "wave front" and normalized by the condition $\langle p(t), \dot{\gamma}(t)\rangle = 1$. The curve $t \mapsto (p(t), \gamma(t))$ in the cotangent bundle $T^*M$ satisfies a Hamiltonian system. This is exactly what happens in rational mechanics or geometric optics.

The sub-Riemannian Hamiltonian $H : T^*M \to \mathbb{R}$ is defined by the formula $H(p, q) = \frac{1}{2}\langle p, v\rangle^2$, where $p \in T^*_qM$, and $v \in T_qM$ is an admissible velocity of length 1 that maximizes $\langle p, w\rangle$ among all admissible velocities $w$ of length one at $q \in M$.

Any smooth function on the cotangent bundle defines a Hamiltonian vector field and such a field generates a Hamiltonian flow. The Hamiltonian flow on $T^*M$ associated to $H$ is the *sub-Riemannian geodesic flow.* The Riemannian geodesic flow is just a special case.

As we mentioned, in general, the construction described above cannot be applied to all geodesics: the so-called abnormal geodesics are missed. An abnormal geodesic $\gamma(t)$ also possesses its "impulse" $p(t) \in T^*_{\gamma(t)}M$ but this impulse belongs to the orthogonal complement to the subspace of admissible velocities and does not satisfy the above Hamiltonian system. Geodesics that are trajectories of the geodesic flow are called *normal*. Actually, abnormal geodesics belong to the closure of the space of the normal ones, and elementary symplectic geometry provides a uniform characterization of the impulses for both classes of geodesics. Such a characterization is, in fact, a very special case of the Pontryagin maximum principle.

Recall that all velocities are admissible in the Riemannian case, and the Euclidean structure on the tangent bundle induces the identification of tangent vectors and covectors, i.e., of the velocities and impulses. We should however remember that this identification depends on the metric. One can think to a sub-Riemannian metric as the limit of a family of Riemannian metrics when the length of forbidden velocities tends to infinity, while the length of admissible velocities remains untouched. It is easy to see that the Riemannian Hamiltonians defined by such a family converge with all derivatives to the sub-Riemannian Hamiltonian. Hence the Riemannian geodesics with a prescribed initial impulse converge to the sub-Riemannian geodesic with the same initial impulse. On the other hand, we cannot expect any reasonable convergence for the family of Riemannian geodesics with a prescribed initial velocity: those with forbidden initial velocities disappear at the limit, while the number of geodesics with admissible initial velocities jumps to infinity.

## Outline of the book

We start in Chapter 1 from surfaces in $\mathbb{R}^3$ that is the beginning of everything in differential geometry, and also a starting point of the story told in this book. There are not yet Hamiltonians here, but a control flavor is already present. The presentation is elementary and self-contained. A student in applied mathematics or analysis who missed the geometry of surfaces at the university or simply is not satisfied by his understanding of these classical ideas, might find it useful to read just this chapter even if he does not plan to study the rest of the book.

In Chapter 2, we recall some basic properties of vector fields and vector bundles. Sub-Riemannian structures are defined in Chapter 3 where we also study three fundamental facts: the finiteness and the continuity of the sub-Riemannian distance, the existence of length-minimizers, and the infinitesimal characterization of geodesics. The first is the classical Rashevskii-Chow theorem, the second and the third one are simplified versions of the Filippov existence theorem and of the Pontryagin maximum principle.

In Chapter 4, we introduce the symplectic language. We define the geodesic Hamiltonian flow, we consider some interesting two- and three-dimensional problems, and we prove a general sufficient condition for length-minimality of normal trajectories. Chapter 5 is devoted to integrable Hamiltonian systems. We explain the construction of the action-angle coordinates and we describe classical examples of integrable geodesic flows, such as the geodesic flow on ellipsoids.

Chapters 1–5 form a first part of the book where we do not use any tool from functional analysis. In fact, even the knowledge of the Lebesgue integration and elementary real analysis are not essential with a unique exception of the existence theorem in Section 3.3. In all other parts of the text, the reader will nevertheless understand the content just replacing the terms "Lipschitz" and "absolutely continuous" with "piecewise $C^1$" and the term "measurable" with "piecewise continuous".

We start to use some basic functional analysis in Chapter 6. In this chapter, we give elements of an operator calculus that simplifies and clarifies calculations with non-stationary flows, their variations and compositions. In Chapter 7, we give a brief introduction to the Lie group theory. Lie groups are introduced as subgroups of the groups of diffeomorphisms of a manifold $M$ induced by a family of vector fields whose Lie algebra is finite dimensional. Then we study left-invariant sub-Riemannian structures and their geodesics.

In Chapter 8, we interpret the "impulses" as Lagrange multipliers for constrained optimization problems and apply this point of view to the sub-Riemannian case. We also introduce the sub-Riemannian exponential map and we study cut and conjugate points.

In Chapter 9, we consider two-dimensional sub-Riemannian metrics; such a metric coincides with a Riemannian one on an open and dense subset. We describe in details the model space of this geometry, known as the Grushin plane, and we discuss several properties in the generic case, among which a Gauss-Bonnet like theorem.

In Chapter 10, we construct the nonholonomic tangent space at a point $q$ of the manifold: a first quasi-homogeneous approximation of the space if you observe and exploit it from $q$ by means of admissible paths. In general, such a tangent space is a homogeneous space of a nilpotent Lie group equipped with an invariant vector distribution; its structure may depend on the point where the tangent space is attached. At generic points, this is a nilpotent Lie group endowed with a left-invariant vector distribution. The construction of the nonholonomic tangent space does not need a metric; if we take into account the metric, we obtain the Gromov–Hausdorff tangent to the sub-Riemannian metric space. Useful "ball-box" estimates of small balls follow automatically.

In Chapter 11, we study general analytic properties of the sub-Riemannian distance as a function of points of the manifold. It is shown that the distance is smooth on an open dense subset and is Lipschitz out of the points connected by abnormal length-minimizers. Moreover, if these bad points are absent, then almost every sphere is a Lipschitz submanifold.

In Chapter 12, we turn to abnormal geodesics, which provide the deepest singularities of the distance. Abnormal geodesics are critical points of the endpoint map defined on the space of admissible paths, and the main tool for their study is the Hessian of the endpoint map. This study permits to prove also that the cut locus from a point is adjacent to the point itself as soon as the structure is not Riemannian.

Chapter 13 is devoted to the explicit calculation of the sub-Riemannian optimal synthesis for model spaces. After a discussion on Carnot groups, we describe a technique based on the Hadamard theorem that permits, under certain assumptions, to compute the cut locus explicitly. We then apply this technique to several relevant examples.

This is the end of the second part of the book; next few chapters are devoted to the curvature and its applications. Let $\Phi^t : T^*M \to T^*M$, for $t \in \mathbb{R}$, be a sub-Riemannian geodesic flow. Submanifolds $\Phi^t(T_q^*M)$, $q \in M$, form a fibration of $T^*M$. Given $\lambda \in T^*M$, let $J_\lambda(t) \subset T_\lambda(T^*M)$ be the tangent space to the leaf of this fibration.

Recall that $\Phi^t$ is a Hamiltonian flow and $T_q^*M$ are Lagrangian submanifolds; hence the leaves of our fibrations are Lagrangian submanifolds and $J_\lambda(t)$ is a Lagrangian subspace of the symplectic space $T_\lambda(T^*M)$.

In other words, $J_\lambda(t)$ belongs to the Lagrangian Grassmannian of $T_\lambda(T^*M)$, and $t \mapsto J_\lambda(t)$ is a curve in the Lagrangian Grassmannian, a *Jacobi curve* of the sub-Riemannian structure. The curvature of the sub-Riemannian space at $\lambda$ is simply the "curvature" of this curve in the Lagrangian Grassmannian.

Chapter 14 is devoted to the elementary differential geometry of curves in the Lagrangian Grassmannian. In Chapter 15 we apply this geometry to Jacobi curves, that are curves in the Lagrange Grassmannian representing Jacobi fields.

The language of Jacobi curves is translated to the traditional language in the Riemannian case in Chapter 16. We recover the Levi-Civita connection and the Riemannian curvature and demonstrate their symplectic meaning. In Chapter 17, we explicitly compute the sub-Riemannian curvature for contact three-dimensional spaces and we show how the curvature invariants appear in the classification of sub-Riemannian left-invariant structures on 3D Lie groups. In Chapter 18, after a brief introduction on Poisson manifolds, we prove the integrability of the sub-Riemannian geodesic flow on 3D Lie groups. As a byproduct, we obtain a classification of coadjoint orbits on 3D Lie algebras. In the next Chapter 19 we study the small distance asymptotics of the exponential map for three-dimensional contact case and see how the structure of the conjugate locus is encoded in the curvature.

In Chapter 20 we address the problem of defining a canonical volume in sub-Riemannian geometry. First we introduce the Popp volume, that is a canonical volume that is smooth for equiregular sub-Riemannian manifold, and we study its basic properties. Then we define the Hausdorff volume and we study its density with respect to Popp's one.

In the last Chapter 21 we define the sub-Riemannian Laplace operator, and we study its properties (hypoellipticity, self-adjointness, etc.). We conclude with a discussion of the sub-Riemannian heat equation and an explicit formula for the heat kernel in the three-dimensional Heisenberg case.

The book is finished by an Appendix on the canonical frames for a wide class of curves in the Lagrangian Grassmannians, written by Igor Zelenko. This is a necessary background for a deeper systematic study of the curvature-type sub-Riemannian invariants, beyond the scope of this book.

We stop here this introduction into the "Comprehensive Introduction". We hope that the reader won't be bored; comments to the chapters contain references and suggestions for further reading.

# Chapter 1

# Geometry of surfaces in $\mathbb{R}^3$

In this preliminary chapter we study the geometry of smooth two dimensional surfaces in $\mathbb{R}^3$ as a "warm-up problem" and we recover some classical results.

In the fist part of the chapter we consider surfaces in $\mathbb{R}^3$ endowed with the standard Euclidean product, which we denote by $\langle \cdot \,|\, \cdot \rangle$. In the second part we study surfaces in the 3D pseudo-Euclidean space, that is $\mathbb{R}^3$ endowed with a sign-indefinite inner product, which we denote by $\langle \cdot \,|\, \cdot \rangle_h$

**Definition 1.1.** A *surface* of $\mathbb{R}^3$ is a subset $M \subset \mathbb{R}^3$ such that for every $q \in M$ there exists a neighborhood $U \subset \mathbb{R}^3$ of $q$ and a smooth function $a : U \to \mathbb{R}$ such that $U \cap M = a^{-1}(0)$ and $\nabla a \neq 0$ on $U \cap M$.

## 1.1 Geodesics and optimality

Let $M \subset \mathbb{R}^3$ be a surface and $\gamma : [0,T] \to M$ be a smooth curve in $M$. The *length* of $\gamma$ is defined as

$$\ell(\gamma) := \int_0^T \|\dot{\gamma}(t)\| dt. \tag{1.1}$$

where $\|v\| = \sqrt{\langle v \,|\, v \rangle}$ denotes the norm of a vector in $\mathbb{R}^3$.

Notice that the definition of length in (1.1) is invariant by reparametrizations of the curve. Indeed let $\varphi : [0,T'] \to [0,T]$ be a monotone smooth function. Define $\gamma_\varphi : [0,T'] \to M$ by $\gamma_\varphi := \gamma \circ \varphi$. Using the change of variables $t = \varphi(s)$, one gets

$$\ell(\gamma_\varphi) = \int_0^{T'} \|\dot{\gamma}_\varphi(s)\| ds = \int_0^{T'} \|\dot{\gamma}(\varphi(s))\| |\dot{\varphi}(s)| ds = \int_0^T \|\dot{\gamma}(t)\| dt = \ell(\gamma).$$

The definition of length can be extended to piecewise smooth curves on $M$, by adding the length of every smooth piece of $\gamma$.

When the curve $\gamma$ is parametrized in such a way that $\|\dot{\gamma}(t)\| \equiv c$ for some $c > 0$ we say that $\gamma$ has *constant speed.* If moreover $c = 1$ we say that $\gamma$ is *parametrized by arc length* (or *arc length parametrized*).

The *distance* between two points $p, q \in M$ is the infimum of length of curves that join $p$ to $q$

$$d(p,q) = \inf\{\ell(\gamma) \mid \gamma : [0,T] \to M \text{ piecewise smooth}, \gamma(0) = p, \gamma(T) = q\}. \tag{1.2}$$

Now we focus on *length-minimizers*, i.e., piece-wise smooth curves $\gamma : [0,T] \to M$ realizing the distance between their endpoints, i.e., satisfying $\ell(\gamma) = d(\gamma(0), \gamma(T))$.

**Exercise 1.2.** Prove that, if $\gamma : [0,T] \to M$ is a length-minimizer, then the curve $\gamma|_{[t_1,t_2]}$ is also a length-minimizer, for all $0 < t_1 < t_2 < T$.

The following proposition characterizes smooth minimizers. We prove later that all length-minimizers are smooth (cf. Corollary 1.12 and 1.13).

**Proposition 1.3.** *Let $\gamma : [0,T] \to M$ be a smooth minimizer parametrized by arc length. Then $\ddot{\gamma}(t) \perp T_{\gamma(t)}M$ for all $t \in [0,T]$.*

The proof of Proposition 1.3 requires some preliminary constructions. Fix a smooth length-minimizer parametrized by arc length $\gamma : [0,T] \to M$ and consider a smooth non-autonomous vector field that extends the tangent vector to $\gamma$ in a neighborhood $W$ of the graph of the curve $\{(t,\gamma(t)) \mid t \in [0,T]\} \subset \mathbb{R} \times M$, i.e., a smooth map $(t,q) \mapsto f_t(q) \in T_qM$ satisfying

$$f_t(\gamma(t)) = \dot{\gamma}(t), \qquad \text{and} \qquad \|f_t(q)\| = 1, \quad \forall\, (t,q) \in W.$$

Let now $(t,q) \mapsto g_t(q) \in T_qM$ be a smooth non-autonomous vector field such that $f_t(q)$ and $g_t(q)$ define a local orthonormal frame in the following sense:

$$\langle f_t(q) \,|\, g_t(q) \rangle = 0, \qquad \|g_t(q)\| = 1, \qquad \forall\, (t,q) \in W.$$

Piecewise smooth curves parametrized by arc length on $M$ are solutions of the following ordinary differential equation

$$\dot{x}(t) = \cos u(t) f_t(x(t)) + \sin u(t) g_t(x(t)), \tag{1.3}$$

for some initial condition $x(0) = q$ and some piecewise continuous function $u(t)$, which we call *control*. The curve $\gamma$ is the solution to (1.3) associated with the control $u(t) \equiv 0$ and initial condition $\gamma(0)$.

Let us consider now the family of controls

$$u_{\tau,s}(t) = \begin{cases} 0, & t < \tau \\ s, & t \geq \tau \end{cases} \qquad 0 \leq \tau \leq T, \quad s \in \mathbb{R} \tag{1.4}$$

and denote by $x_{\tau,s}(t)$ the solution of (1.3) that corresponds to the control $u_{\tau,s}(t)$ and with initial condition $x_{\tau,s}(0) = \gamma(0)$. We have the following two results.

**Lemma 1.4.** *For every $\tau_1, \tau_2, t \in [0,T]$ the following vectors are linearly dependent*

$$\left.\frac{\partial}{\partial s}\right|_{s=0} x_{\tau_1,s}(t), \qquad \left.\frac{\partial}{\partial s}\right|_{s=0} x_{\tau_2,s}(t). \tag{1.5}$$

*Proof.* Thanks to Exercice 1.2, it is not restrictive to assume $t = T$. Fix $0 \leq \tau_1 \leq \tau_2 \leq T$ and consider the family of controls

$$v_{h_1,h_2}(t) = \begin{cases} 0, & t \in [0,\tau_1[, \\ h_1, & t \in [\tau_1,\tau_2[, \\ h_1 + h_2, & t \in [\tau_2, T+\varepsilon[, \end{cases} \tag{1.6}$$

where $h_1, h_2$ belong to a neighborhood of 0 and $\varepsilon > 0$ is chosen small enough to guarantee the existence of the solutions of (1.3) associated with controls $v_{h_1,h_2}$. Denote by $t \mapsto \phi(t; h_1, h_2)$ the corresponding solution. Notice that this defines a map $\phi : U \subset \mathbb{R}^3 \to M$, well-defined and smooth on a neighborhood $U$ of $(T, 0, 0)$. Moreover, by construction

$$\frac{\partial \phi}{\partial h_i}\bigg|_{(T,0,0)} = \frac{\partial}{\partial s}\bigg|_{s=0} x_{\tau_i,s}(T), \qquad i = 1, 2.$$

By contradiction assume that the vectors in (1.5) are linearly independent. Then $\frac{\partial \phi}{\partial h}$ is invertible (recall that $\phi$ takes values in $M$) and the implicit function theorem applied to the map $(t, h_1, h_2) \mapsto \phi(t; h_1, h_2)$ at the point $(T, 0, 0)$ implies that there exists $\delta > 0$ such that

$$\forall\, t \in ]T - \delta, T + \delta[, \quad \exists\, h_1, h_2, \quad \text{s.t.} \quad \phi(t; h_1, h_2) = \gamma(T),$$

In particular there exists a curve with unit speed joining $\gamma(0)$ and $\gamma(T)$ in time $t < T$, which gives a contradiction, since $\gamma$ is a length-minimizer. □

**Lemma 1.5.** *For every $\tau, t \in [0, T]$ the following identity holds*

$$\left\langle \frac{\partial}{\partial s}\bigg|_{s=0} x_{\tau,s}(t) \;\bigg|\; \dot{\gamma}(t) \right\rangle = 0. \tag{1.7}$$

*In particular $\frac{\partial}{\partial s}\big|_{s=0} x_{\tau,s}(t)$ is parallel to $g_t(\gamma(t))$.*

*Proof.* If $t \le \tau$, then by construction (cf. (1.4)) the first vector is zero since there is no variation w.r.t. $s$ and the conclusion follows. Assume now that $t > \tau$. Let us write the Taylor expansion of $\psi(t) = \frac{\partial}{\partial s}\big|_{s=0} x_{\tau,s}(t)$ in a right neighborhood of $t = \tau$. For $t \ge \tau$, by (1.4), one has

$$\dot{x}_{\tau,s} = \cos(s) f_t(x_{\tau,s}) + \sin(s) g_t(x_{\tau,s}).$$

Hence

$$\psi(\tau) = \frac{\partial}{\partial s}\bigg|_{s=0} x_{\tau,s}(\tau) = 0, \qquad \dot{\psi}(\tau) = \frac{\partial}{\partial s}\bigg|_{s=0} \dot{x}_{\tau,s}(\tau) = g_\tau(x_{\tau,s}(\tau)).$$

Then, for $t \ge \tau$, we have

$$\psi(t) = (t - \tau) g_\tau(x_{\tau,s}(\tau)) + O((t - \tau)^2). \tag{1.8}$$

By Exercice 1.2, it is sufficient to prove the statement at $t = T$. Then taking $t = T$ in (1.8) and passing to the limit for $\tau \to T$ one gets

$$\frac{1}{T - \tau} \frac{\partial}{\partial s}\bigg|_{s=0} x_{\tau,s}(T) \underset{\tau \to T}{\longrightarrow} g_T(\gamma(T)).$$

Now, by Lemma 1.4 all vectors in the left hand side are parallel among them, hence they are parallel to $g_T(\gamma(T))$. The lemma is proved since $\dot{\gamma}(T) = f_T(\gamma(T))$ and $f_T$ and $g_T$ are orthogonal. □

We can now prove Proposition 1.3.

*Proof of Proposition 1.3.* Let $\gamma : [0, T] \to M$ be a smooth length-minimizer parametrized by arc length and consider the smooth non-autonomous vector fields $f_t(q)$ and $g_t(q)$ defining a local orthonormal frame as above. The claim $\ddot{\gamma}(t) \perp T_{\gamma(t)} M$ is equivalent to the following

$$\langle \ddot{\gamma}(t) \,|\, f_t(\gamma(t)) \rangle = \langle \ddot{\gamma}(t) \,|\, g_t(\gamma(t)) \rangle = 0. \tag{1.9}$$

Recall that $\langle \dot\gamma(t) \,|\, \dot\gamma(t) \rangle = 1$ for every $t \in [0,T]$. Differentiating this identity with respect to $t$, one gets for $t \in [0,T]$

$$0 = \frac{d}{dt} \langle \dot\gamma(t) \,|\, \dot\gamma(t) \rangle = 2 \langle \ddot\gamma(t) \,|\, \dot\gamma(t) \rangle \,.$$

This shows that $\ddot\gamma(t)$ is orthogonal to $f_t(\gamma(t))$ for every $t \in [0,T]$. Next, differentiating[1] (1.7) with respect to $t$, we have for $t \neq \tau$

$$\left\langle \frac{\partial}{\partial s}\Big|_{s=0} \dot x_{\tau,s}(t) \,\middle|\, \dot\gamma(t) \right\rangle + \left\langle \frac{\partial}{\partial s}\Big|_{s=0} x_{\tau,s}(t) \,\middle|\, \ddot\gamma(t) \right\rangle = 0. \tag{1.10}$$

Moreover, from the identity $\langle \dot x_{\tau,s}(t) \,|\, \dot x_{\tau,s}(t) \rangle = 1$ one gets

$$\left\langle \frac{\partial}{\partial s} \dot x_{\tau,s}(t) \,\middle|\, \dot x_{\tau,s}(t) \right\rangle = 0, \qquad \text{for } t \neq \tau. \tag{1.11}$$

Evaluating (1.11) at $s = 0$, using that $x_{\tau,0}(t) = \gamma(t)$, one has

$$\left\langle \frac{\partial}{\partial s}\Big|_{s=0} \dot x_{\tau,s}(t) \,\middle|\, \dot\gamma(t) \right\rangle = 0, \qquad \text{for } t \neq \tau. \tag{1.12}$$

Hence, combining (1.12) with (1.10), it follows that

$$\left\langle \frac{\partial}{\partial s}\Big|_{s=0} x_{\tau,s}(t) \,\middle|\, \ddot\gamma(t) \right\rangle = 0,$$

which holds for $t \neq \tau$ and then, by continuity, for every $t \in [0,T]$. Using that $\frac{\partial}{\partial s}\big|_{s=0} x_{\tau,s}(t)$ is parallel to $g_t(\gamma(t))$ (by Lemma 1.5), it follows that $\langle g_t(\gamma(t)) \,|\, \ddot\gamma(t) \rangle = 0$. □

**Definition 1.6.** A smooth curve $\gamma : [0,T] \to M$ parametrized with constant speed is called *geodesic* if it satisfies

$$\ddot\gamma(t) \perp T_{\gamma(t)}M, \qquad \forall\, t \in [0,T]. \tag{1.13}$$

Proposition 1.3 says that a smooth curve that minimizes the length is a geodesic.

Next we get an explicit characterization of geodesics when the manifold $M$ is *globally* defined as the zero level of a smooth function. In other words there exists a smooth function $a : \mathbb{R}^3 \to \mathbb{R}$ such that

$$M = a^{-1}(0), \qquad \text{and} \qquad \nabla a \neq 0 \text{ on } M. \tag{1.14}$$

*Remark* 1.7. Recall that for every $q \in M$ it holds $\nabla_q a \perp T_q M$. Indeed, for every $q \in M$ and $v \in T_q M$, let $\gamma : [0,T] \to M$ be a smooth curve on $M$ such that $\gamma(0) = q$ and $\dot\gamma(0) = v$. By definition of $M$ one has $a(\gamma(t)) = 0$. Differentiating this identity with respect to $t$ at $t = 0$ one gets $\langle \nabla_q a \,|\, v \rangle = 0$.

**Proposition 1.8.** *A smooth curve $\gamma : [0,T] \to M$ is a geodesic if and only if it satisfies, in matrix notation:*

$$\ddot\gamma(t) = -\frac{\dot\gamma(t)^T (\nabla^2_{\gamma(t)} a) \dot\gamma(t)}{\|\nabla_{\gamma(t)} a\|^2} \nabla_{\gamma(t)} a, \qquad \forall\, t \in [0,T], \tag{1.15}$$

*where $\nabla^2_{\gamma(t)} a$ is the Hessian matrix of $a$.*

[1] notice that $x_{\tau,s}$ is smooth on the set $[0,T] \setminus \{\tau\}$.

*Proof.* Differentiating the equality $\langle \nabla_{\gamma(t)} a \,|\, \dot\gamma(t) \rangle = 0$ with respect to $t$, we get (we use matrix notation):

$$\dot\gamma(t)^T (\nabla^2_{\gamma(t)} a) \dot\gamma(t) + \ddot\gamma(t)^T \nabla_{\gamma(t)} a = 0.$$

By definition of geodesic there exists a function $b(t)$ such that

$$\ddot\gamma(t) = b(t) \nabla_{\gamma(t)} a.$$

Combining the two previous formulas we get

$$\dot\gamma(t)^T (\nabla^2_{\gamma(t)} a) \dot\gamma(t) + b(t) \|\nabla_{\gamma(t)} a\|^2 = 0,$$

from which (1.15) easily follows. □

*Remark* 1.9. A surface in $\mathbb{R}^3$ (cf. Definition 1.1) is always locally defined as the zero set of a smooth function, hence the characterization given in Proposition 1.8 is still true on every open set $U$ where in (1.15) $a : U \subset \mathbb{R}^3 \to \mathbb{R}$ denotes a smooth function such that $M \cap U = a^{-1}(0)$ and $\nabla a \neq 0$.

### 1.1.1 Existence and minimizing properties of geodesics

As a direct consequence of Proposition 1.8 one gets the following existence and uniqueness theorem for geodesics.

**Corollary 1.10.** *Let $q \in M$ and $v \in T_q M$. There exists a unique geodesic $\gamma : [0, \varepsilon] \to M$, for $\varepsilon > 0$ small enough, such that $\gamma(0) = q$ and $\dot\gamma(0) = v$.*

*Proof.* By Proposition 1.8, geodesics satisfy a second order ODE, hence they are smooth curves, characterized by ther initial position and velocity. □

To end this section we show that small pieces of geodesics are always global minimizers.

**Theorem 1.11.** *Let $\gamma : [0, T] \to M$ be a geodesic. For every $\tau \in [0, T[$ there exists $\varepsilon > 0$ such that*

*(i) $\gamma|_{[\tau, \tau+\varepsilon]}$ is a minimizer, i.e., $d(\gamma(\tau), \gamma(\tau + \varepsilon)) = \ell(\gamma|_{[\tau,\tau+\varepsilon]})$,*

*(ii) $\gamma|_{[\tau, \tau+\varepsilon]}$ is the unique minimizer joining $\gamma(\tau)$ and $\gamma(\tau + \varepsilon)$ in the class of piecewise smooth curves, up to reparametrization.*

*Proof.* Without loss of generality let us assume that $\tau = 0$ and that $\gamma$ is arc length parametrized. Consider a arc length parametrized curve $\alpha$ on $M$ such that $\alpha(0) = \gamma(0)$ and $\dot\alpha(0) \perp \dot\gamma(0)$ and denote by $(t, s) \mapsto x_s(t)$ a smooth variation of geodesics such that $x_0(t) = \gamma(t)$ and (see also Figure 1.1)

$$x_s(0) = \alpha(s), \qquad \dot x_s(0) \perp \frac{\partial}{\partial s}\alpha(s). \tag{1.16}$$

The map $\psi : (t, s) \mapsto x_s(t)$ is smooth, and is a local diffeomorphism near $(0, 0)$. Indeed we compute the partial derivatives

$$\frac{\partial \psi}{\partial t}\Big|_{t=s=0} = \frac{\partial}{\partial t}\Big|_{t=0} x_0(t) = \dot\gamma(0), \qquad \frac{\partial \psi}{\partial s}\Big|_{t=s=0} = \frac{\partial}{\partial s}\Big|_{s=0} x_s(0) = \dot\alpha(0),$$

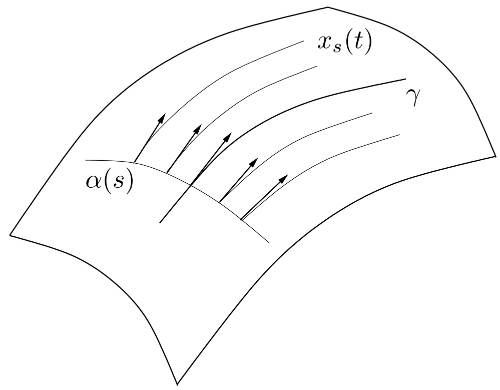



Figure 1.1: Proof of Theorem 1.11

and they are linearly independent. Thus $\psi$ maps a neighborhood $U$ of $(0,0)$ on a neighborhood $W$ of $\gamma(0)$. We now consider the function $\phi$ and the vector field $X$ defined on $W$ by

$$\phi : x_s(t) \mapsto t, \qquad X : x_s(t) \mapsto \dot{x}_s(t).$$

**Claim.** For every $q \in W$ it holds $\nabla_q \phi = X(q)$.

To prove the claim, we first show that the two vectors are parallel, and then that they actually coincide. To show that they are parallel, first notice that $\nabla\phi$ is orthogonal to its level set $\{t = \text{const}\}$, hence

$$\left\langle \nabla_{x_s(t)}\phi \,\middle|\, \frac{\partial}{\partial s} x_s(t) \right\rangle = 0, \qquad \forall\, (t,s) \in U. \tag{1.17}$$

Now, let us now consider the quantity, for fixed $s$

$$\zeta_s(t) := \left\langle \frac{\partial}{\partial s} x_s(t) \,\middle|\, \dot{x}_s(t) \right\rangle. \tag{1.18}$$

Let us prove that, for fixed $s$, one has $\dot{\zeta}_s(t) = 0$. Indeed computing the derivative with respect to $t$ of (1.18) one gets

$$\dot{\zeta}_s(t) = \left\langle \frac{\partial}{\partial s} \dot{x}_s(t) \,\middle|\, \dot{x}_s(t) \right\rangle + \left\langle \frac{\partial}{\partial s} x_s(t) \,\middle|\, \ddot{x}_s(t) \right\rangle \tag{1.19}$$

which is identically zero. Indeed the first term in (1.19) vanishes since $\dot{x}_s(t)$ has unit speed, while the second one vanishes thanks to the geodesic property (1.13). Hence, (1.18) is constant with respect to $t$ and $\zeta_s(t) = \zeta_s(0) = 0$ by the orthogonality assumption (1.16), for every $s$.

Combining (1.17) and (1.18) one gets that $\nabla\phi$ is parallel to $X$. Actually they coincide since

$$\langle \nabla\phi \,|\, X \rangle = \frac{d}{dt}\phi(x_s(t)) = 1,$$

which proves the claim. Now consider $\varepsilon > 0$ small enough such that $\gamma|_{[0,\varepsilon]}$ is contained in $W$ and take a piecewise smooth and length parametrized curve $c : [0,\varepsilon'] \to M$ contained in $W$ and joining $\gamma(0)$ to $\gamma(\varepsilon)$. Let us show that $\gamma$ is shorter than $c$. First notice that

$$\ell(\gamma|_{[0,\varepsilon]}) = \varepsilon = \phi(\gamma(\varepsilon)) = \phi(c(\varepsilon')).$$

Using that $\phi(c(0)) = \phi(\gamma(0)) = 0$ and that $\ell(c) = \varepsilon'$ we have that

$$\ell(\gamma|_{[0,\varepsilon]}) = \phi(c(\varepsilon')) - \phi(c(0)) = \int_0^{\varepsilon'} \frac{d}{dt}\phi(c(t))dt \tag{1.20}$$
$$= \int_0^{\varepsilon'} \langle \nabla\phi(c(t)) \,|\, \dot{c}(t) \rangle \, dt$$
$$= \int_0^{\varepsilon'} \langle X(c(t)) \,|\, \dot{c}(t) \rangle \, dt \leq \varepsilon' = \ell(c), \tag{1.21}$$

The last inequality follows from the Cauchy-Schwartz inequality. Indeed

$$\langle X(c(t)) \,|\, \dot{c}(t) \rangle \leq \|X(c(t))\| \|\dot{c}(t)\| = 1, \tag{1.22}$$

which holds at every smooth point of $c(t)$. In addition, equality in (1.22) holds if and only if $\dot{c}(t) = X(c(t))$ (at the smooth points of $c$). Hence we get that $\ell(c) = \ell(\gamma|_{[0,\varepsilon]})$ if and only if $c$ coincides with $\gamma|_{[0,\varepsilon]}$.

To show that there exists $\bar{\varepsilon} \leq \varepsilon$ such that $\gamma|_{[0,\bar{\varepsilon}]}$ is a global minimizer among all piecewise smooth curves joining $\gamma(0)$ to $\gamma(\bar{\varepsilon})$, it is enough to take $\bar{\varepsilon} < \mathrm{dist}(\gamma(0), \partial W)$. Indeed every curve that escapes from $W$ has length greater than $\bar{\varepsilon}$. $\square$

From Theorem 1.11 we obtain the following.

**Corollary 1.12.** *Any minimizer of the distance (in the class of piecewise smooth curves) is a geodesic, and hence smooth.*

### 1.1.2 Absolutely continuous curves

Notice that formula (1.1) defines the length of a curve even if $\gamma$ is only absolutely continuous, if one interprets the integral in the Lebesgue sense (recall that absolutely continuous curve are differentiable almost everywhere).

The proof of Theorem 1.11, and in particular estimates (1.20)-(1.21), can be extended to the class of absolutely continuous curves. This proves that small pieces of geodesics are minimizers also in the larger class of absolutely continuous curves on $M$. As a byproduct, we have the following corollary.

**Corollary 1.13.** *Any minimizer of the distance (in the class of absolutely continuous curves) is a geodesic, and hence smooth.*

## 1.2 Parallel transport

In this section we want to introduce the notion of parallel transport on a surface (along a curve), which let us to define its main geometric invariant: the Gaussian curvature.

**Definition 1.14.** Let $\gamma : [0, T] \to M$ be a smooth curve. A smooth curve of tangent vectors $\xi(t) \in T_{\gamma(t)}M$ is said to be *parallel* if $\dot{\xi}(t) \perp T_{\gamma(t)}M$.

This notion generalizes the notion of parallelism of vectors on the plane, where it is possible to canonically identify every tangent space to $M = \mathbb{R}^2$ with $\mathbb{R}^2$ itself.[2] In this case a smooth curve of tangent vectors $\xi(t) \in T_{\gamma(t)}M$ is parallel if and only if $\dot{\xi}(t) = 0$.

When $M$ is the zero level of a smooth function $a : \mathbb{R}^3 \to \mathbb{R}$ as in (1.14), we have the following description:

**Proposition 1.15.** *A smooth curve of tangent vectors $\xi(t)$ defined along $\gamma : [0,T] \to M$ is parallel if and only if it satisfies*

$$\dot{\xi}(t) = -\frac{\dot{\gamma}(t)^T(\nabla^2_{\gamma(t)}a)\xi(t)}{\|\nabla_{\gamma(t)}a\|^2}\nabla_{\gamma(t)}a, \qquad \forall\, t \in [0,T]. \tag{1.23}$$

*Proof.* As in Remark 1.7, $\xi(t) \in T_{\gamma(t)}M$ implies $\left\langle \nabla_{\gamma(t)}a, \xi(t)\right\rangle = 0$. Moreover, by assumption, $\dot{\xi}(t) = \alpha(t)\nabla_{\gamma(t)}a$ for some smooth function $\alpha$. With analogous computations as in the proof of Proposition 1.8 we get that

$$\dot{\gamma}(t)^T(\nabla^2_{\gamma(t)}a)\xi(t) + \alpha(t)\|\nabla_{\gamma(t)}a\|^2 = 0,$$

from which the statement follows. □

*Remark* 1.16. Notice that, since (1.23) is a first order linear ODE with respect to $\xi$, for a given curve $\gamma : [0,T] \to M$ and initial datum $v \in T_{\gamma(0)}M$, there is a unique parallel curve of tangent vectors $\xi(t) \in T_{\gamma(t)}M$ along $\gamma$ such that $\xi(0) = v$. Since (1.23) is a linear ODE, the operator that associates with every initial condition $\xi(0)$ the final vector $\xi(t)$ is a linear operator, which is called *parallel transport*.

Next we state a key property of the parallel transport.

**Proposition 1.17.** *The parallel transport preserves the inner product. In other words, if $\xi(t), \eta(t)$ are two parallel curves of tangent vectors along $\gamma$, then we have*

$$\frac{d}{dt}\langle \xi(t) \,|\, \eta(t)\rangle = 0, \qquad \forall\, t \in [0,T]. \tag{1.24}$$

*Proof.* From the fact that $\xi(t), \eta(t) \in T_{\gamma(t)}M$ and $\dot{\xi}(t), \dot{\eta}(t) \perp T_{\gamma(t)}M$ one immediately gets

$$\frac{d}{dt}\langle \xi(t) \,|\, \eta(t)\rangle = \langle \dot{\xi}(t) | \eta(t)\rangle + \langle \xi(t) \,|\, \dot{\eta}(t)\rangle = 0. \qquad \square$$

The notion of parallel transport just introduced, permits to give a new characterization of geodesics, cf. Propositions 1.8 and 1.15.

**Corollary 1.18.** *A smooth curve $\gamma : [0,T] \to M$ is a geodesic if and only if $\dot{\gamma}$ is parallel along $\gamma$.*

[2] The canonical isomorphism $\mathbb{R}^2 \simeq T_x\mathbb{R}^2$ is written explicitly as follows: $y \mapsto \frac{d}{dt}\big|_{t=0} x + ty$.

### 1.2.1 Parallel transport and Levi-Civita connection

**Definition 1.19.** An *orientation* of a surface $M$ is a smooth map $\nu : M \to \mathbb{R}^3$, defined globally on $M$, such that $\nu(q) \perp T_qM$ and $\|\nu(q)\| = 1$ for every $q \in M$. Notice that if $\nu$ is an orientation of $M$, then also $-\nu$ defines an orientation of $M$.

A surface $M$ is *oriented* if it is given (when it exists) an orientation. On an oriented surface $M$, an orthonormal frame $\{e_1, e_2\}$ of $T_qM$ is said *positively oriented* (resp. *negatively oriented*) if $e_1 \wedge e_2 = k\nu(q)$ with $k > 0$ (resp. $k < 0$).

In the following we assume that $M$ is an *oriented* surface.

**Definition 1.20.** The *spherical bundle* $SM$ on $M$ is the disjoint union of all unit tangent vectors to $M$:

$$SM = \bigsqcup_{q \in M} S_qM, \qquad S_qM = \{v \in T_qM, \|v\| = 1\}. \tag{1.25}$$

$SM$ can be endowed with the structure of smooth manifold of dimension 3, and more precisely of fiber bundle with base manifold $M$, typical fiber $S^1$, and canonical projection

$$\pi : SM \to M, \qquad \pi(v) = q \quad \text{if} \quad v \in T_qM.$$

*Remark* 1.21. Fix a positively oriented local orthonormal frame $\{e_1(q), e_2(q)\}$ on $M$. Since every vector in the fiber $S_qM$ has norm one, we can write every $v \in S_qM$ as $v = \cos(\theta)e_1(q) + \sin(\theta)e_2(q)$ for $\theta \in S^1$.

A choice of such an orthonormal frame induces then coordinates $(q, \theta)$ on $SM$. Notice that the choice of a different positively oriented local orthonormal frame $\{e_1'(q), e_2'(q)\}$ induces coordinates $(q', \theta')$ on $SM$ where $q' = q$ and $\theta' = \theta + \phi(q)$ for $\phi \in C^\infty(M)$.

The orientation of $M$ permits, once a unit tangent vector is given, to define a canonical orthonormal frame.

**Definition 1.22.** Let $\xi \in S_qM$. The *canonical orthonormal frame* associated with $\xi$ is the unique orthonormal frame $(\xi, \eta, \nu)$ of $\mathbb{R}^3$ where $(\xi, \eta)$ defines a positively oriented orthonormal frame on $T_qM$ and $\nu \perp T_qM$ is the unit vector defined by the orientation of $M$.

Let $t \mapsto \xi(t) \in S_{\gamma(t)}M$ be a smooth curve of unit tangent vectors along $\gamma : [0, T] \to M$. Define for every $t \in [0, T]$ the canonical orthonormal frame $(\xi(t), \eta(t), \nu(t))$ associated with $\xi(t)$ as above. Since $t \mapsto \xi(t)$ has constant speed, one has $\xi(t) \perp \dot{\xi}(t)$ for every $t$. Hence $\dot{\xi}$ has no component along $\xi$, and we can write

$$\dot{\xi}(t) = u_\xi(t)\eta(t) + v_\xi(t)\nu(t). \tag{1.26}$$

Next we introduce the Levi-Civita connection as a differential 1-form. A differential 1-form is a dual object of a vector field, and with every point associates a covector, that is a linear functional on the tangent space.

**Definition 1.23.** The *Levi-Civita connection* on $M$ is the 1-form $\omega \in \Lambda^1(SM)$ defined by

$$\omega_\xi : T_\xi SM \to \mathbb{R}, \qquad \omega_\xi(z) = u_z, \tag{1.27}$$

where $z = u_z\eta + v_z\nu$ and $(\xi, \eta, \nu)$ is the orthonormal frame defined above.

Notice that $\omega$ changes sign if we change the orientation of $M$.

**Lemma 1.24.** *A curve of unit tangent vectors $\xi(t)$ is parallel if and only if $\omega_{\xi(t)}(\dot{\xi}(t)) = 0$.*

*Proof.* By definition $\xi(t)$ is parallel if and only if $\dot{\xi}(t)$ is orthogonal to $T_{\gamma(t)}M$, i.e., collinear to $\nu(t)$. $\square$

In particular from the previous lemma it follows that a curve parametrized by arc length $\gamma : [0,T] \to M$ is a geodesic if and only if

$$\omega_{\dot{\gamma}(t)}(\ddot{\gamma}(t)) = 0, \qquad \forall\, t \in [0,T]. \tag{1.28}$$

**Proposition 1.25.** *The Levi-Civita connection $\omega \in \Lambda^1(SM)$ satisfies the following properties:*

*(i) there exist two smooth functions $a_1, a_2 : M \to \mathbb{R}$ such that*

$$\omega = d\theta + a_1(x_1,x_2)dx_1 + a_2(x_1,x_2)dx_2, \tag{1.29}$$

*where $(x_1,x_2,\theta)$ is a system of coordinates on $SM$ induced by the choice of coordinates on $M$ and a local orthonormal frame (cf. Remark 1.21).*

*(ii) $d\omega = \pi^*\Omega$, where $\Omega$ is a 2-form defined on $M$ and $\pi : SM \to M$ is the canonical projection.*

*Proof.* (i). Fix a system of coordinates $(x_1,x_2)$ on $M$ and a local orthonormal frame, in such a way that coordinates $(x_1,x_2,\theta)$ on $SM$ are defined as in Remark 1.21. Consider the vector field $\partial/\partial\theta$ on $SM$ (this vector field is independent on the choice of the coordinates). Let us prove that

$$\omega\left(\frac{\partial}{\partial\theta}\right) = 1.$$

Indeed consider a curve $t \mapsto \xi(t)$ of unit tangent vectors at a fixed point which describes a rotation in a single fibre. As a curve on $SM$, the velocity of this curve is exactly its orthogonal vector, i.e., $\dot{\xi}(t) = \eta(t)$ and the equality above follows from the definition of $\omega$. By construction, $\omega$ is invariant by rotations, hence the coefficients $a_i = \omega(\partial/\partial x_i)$ do not depend on the variable $\theta$.

(ii). A differential two-form on $SM$ is the pull-back of a two form on $M$ through the canonical projection $\pi : SM \to M$ if and only if, when written in coordinates, its expression depends only on coordinates on the base $M$. The claim then follows from the coordinate expression (1.29) since

$$d\omega = \left(-\frac{\partial a_1}{\partial x_2}(x_1,x_2) + \frac{\partial a_2}{\partial x_1}(x_1,x_2)\right) dx_1 \wedge dx_2. \tag{1.30}$$

$\square$

*Remark* 1.26. Notice that the functions $a_1, a_2$ in (1.29) are not invariant by change of coordinates on the fiber. Indeed fix a new angular coordinate $\theta' = \theta + \varphi(x_1,x_2)$ (induced by a different choice of orthonormal frame, cf. Remark 1.21). Then one gets $d\theta' = d\theta + (\partial_{x_1}\varphi)dx_1 + (\partial_{x_2}\varphi)dx_2$. Hence $\omega = d\theta' + a_1'(x_1,x_2)dx_1 + a_2'(x_1,x_2)dx_2$ where $a_i' = a_i + \partial_{x_i}\varphi$ for $i = 1,2$.

By definition $\omega$ is an intrinsic 1-form on $SM$. Its differential, by property (ii) of Proposition 1.25, is the pull-back of an intrinsic 2-form on $M$, that in general is not exact.

**Definition 1.27.** The area form $dV$ on a surface $M$ is the differential 2-form that on every tangent space to the manifold agrees with the volume induced by the inner product. In other words, for every positively oriented orthonormal frame $e_1, e_2$ of $T_qM$, one has $dV(e_1,e_2) = 1$.

Given a set $\Gamma \subset M$ its *area* is the quantity $|\Gamma| = \int_\Gamma dV$.

Since any 2-form on $M$ is proportional to the area form $dV$, it makes sense to give the following definition:

**Definition 1.28.** The *Gaussian curvature* of $M$ is the function $\kappa : M \to \mathbb{R}$ defined by the equality

$$\Omega = -\kappa dV. \tag{1.31}$$

Note that $\kappa$ does not depend on the orientation of $M$, since both $\Omega$ and $dV$ change sign if we reverse the orientation. Moreover the area 2-form $dV$ on the surface depends only on the metric structure on the surface.

## 1.3 Gauss-Bonnet theorems

In this section we will prove both the local and the global version of the Gauss-Bonnet theorem. A strong consequence of these results is the celebrated Gauss' Theorema Egregium which says that the Gaussian curvature of a surface is independent on its embedding in $\mathbb{R}^3$.

**Definition 1.29.** Let $\gamma : [0, T] \to M$ be a smooth curve parametrized by arc length. The *geodesic curvature* of $\gamma$ is defined as

$$\rho_\gamma(t) = \omega_{\dot\gamma(t)}(\ddot\gamma(t)). \tag{1.32}$$

Notice that if $\gamma$ is a geodesic, then $\rho_\gamma(t) = 0$ for every $t \in [0, T]$. The geodesic curvature measures how much a curve is far from being a geodesic.

*Remark* 1.30. The geodesic curvature changes sign if we move along the curve in the opposite direction. Moreover, if $M = \mathbb{R}^2$, it coincides with the usual notion of curvature of a planar curve.

### 1.3.1 Gauss-Bonnet theorem: local version

A regular polygon in $\mathbb{R}^2$ is a polygon that is equiangular and equilateral. We include disks among regular polygons (as a limit case, when the number of edges is infinite).

**Definition 1.31.** A *curvilinear polygon* $\Gamma$ on an oriented surface $M$ is the image of a regular polygon in $\mathbb{R}^2$ under a diffeomorphism. We assume that $\partial\Gamma$ is oriented consistently with the orientation of $M$.

Notice that a curvilinear polygon is always homeomorphic to a disk, and the case when $\partial\Gamma$ is smooth (and $\Gamma$ is diffeomorphic to the disk) is included in the definition.

In what follows, given a curvilinear polygon $\Gamma$ on an oriented surface $M$, we denote by (cf. also Figure 1.2)

- $\gamma_i : I_i \to M$, for $i = 1, \ldots, m$, the smooth curves parametrized by arc length, with orientation consistent with $\partial\Gamma$, such that $\partial\Gamma = \cup_{i=1}^m \gamma_i(I_i)$,
- $\alpha_i$, for $i = 1, \ldots, m$, the external angles at the points where $\partial\Gamma$ is not $C^1$.

**Theorem 1.32** (Gauss-Bonnet, local version)**.** *Let* $\Gamma$ *be a curvilinear polygon on an oriented surface* $M$*. Then we have*

$$\int_\Gamma \kappa dV + \sum_{i=1}^m \int_{I_i} \rho_{\gamma_i}(t)dt + \sum_{i=1}^m \alpha_i = 2\pi. \tag{1.33}$$

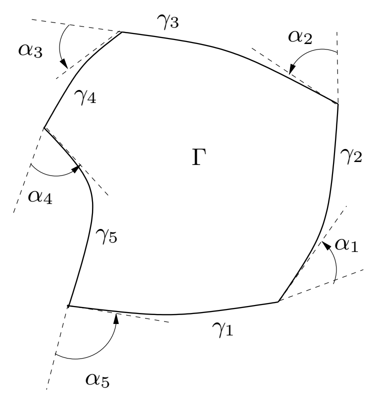



Figure 1.2: A curvilinear polygon

*Proof.* (i). Case $\partial\Gamma$ smooth.

In this case $\Gamma$ is the image of the unit (closed) ball $B_1$, centered in the origin of $\mathbb{R}^2$, under a diffeomorphism

$$F : B_1 \to M, \qquad \Gamma = F(B_1).$$

In what follows we denote by $\gamma : I \to M$ the curve such that $\gamma(I) = \partial\Gamma$. We consider on $B_1$ the vector field $V(x) = x_1\partial_{x_2} - x_2\partial_{x_1}$ which has an isolated zero at the origin and whose flow is a rotation around zero. Denote by $X := F_*V$ the induced vector field on $M$ with critical point $q_0 = F(0)$.

For $\varepsilon > 0$ small enough, we define (cf. Figure 1.3)

$$\Gamma_\varepsilon := \Gamma \setminus F(B_\varepsilon), \qquad \text{and} \qquad A_\varepsilon := \partial F(B_\varepsilon),$$

where $B_\varepsilon$ is the ball of radius $\varepsilon$ centered at zero in $\mathbb{R}^2$. We have $\partial\Gamma_\varepsilon = A_\varepsilon \cup \partial\Gamma$. Define the map

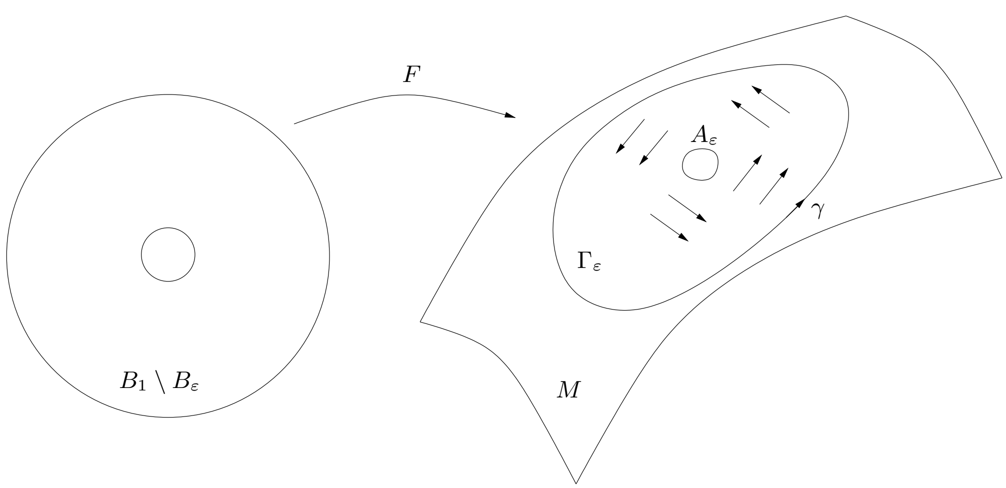



Figure 1.3: The map $F$

$$\phi : \Gamma_\varepsilon \to SM, \qquad \phi(q) = \frac{X(q)}{|X(q)|}.$$

First notice that

$$\int_{\phi(\Gamma_\varepsilon)} d\omega = \int_{\phi(\Gamma_\varepsilon)} \pi^*\Omega = \int_{\pi(\phi(\Gamma_\varepsilon))} \Omega = \int_{\Gamma_\varepsilon} \Omega, \tag{1.34}$$

where we used the fact that $\pi(\phi(\Gamma_\varepsilon)) = \Gamma_\varepsilon$. Then let us compute the integral of the curvature $\kappa$ on $\Gamma_\varepsilon$

$$\begin{aligned}
\int_{\Gamma_\varepsilon} \kappa dV &= -\int_{\Gamma_\varepsilon} \Omega = -\int_{\phi(\Gamma_\varepsilon)} d\omega, && \text{(by (1.34))} \\
&= -\int_{\partial\phi(\Gamma_\varepsilon)} \omega, && \text{(by Stokes Theorem)} \\
&= \int_{\phi(A_\varepsilon)} \omega - \int_{\phi(\partial\Gamma)} \omega, && (\text{since } \partial\phi(\Gamma_\varepsilon) = \phi(A_\varepsilon) \cup \phi(\partial\Gamma))
\end{aligned} \tag{1.35}$$

Notice that in the third equality we used the fact that the induced orientation on $\partial\phi(\Gamma_\varepsilon)$ gives opposite orientation on the two terms. Let us treat separately these two terms. The first one, by Proposition 1.25, can be written as

$$\int_{\phi(A_\varepsilon)} \omega = \int_{\phi(A_\varepsilon)} d\theta + \int_{\phi(A_\varepsilon)} a_1(x_1, x_2)dx_1 + a_2(x_1, x_2)dx_2. \tag{1.36}$$

The first element of (1.36) is equal to $2\pi$ since we integrate the 1-form $d\theta$ on a closed curve. The second element of (1.36), for $\varepsilon \to 0$, satisfies

$$\left| \int_{\phi(A_\varepsilon)} a_1(x_1, x_2)dx_1 + a_2(x_1, x_2)dx_2 \right| \leq C\ell(\phi(A_\varepsilon)) \to 0, \tag{1.37}$$

Indeed the functions $a_i$ are smooth (hence bounded on compact sets) and the length of $\phi(A_\varepsilon)$ goes to zero for $\varepsilon \to 0$.

Let us now consider the second term of (1.35). Since $\phi(\partial\Gamma)$ is parametrized by the curve $t \mapsto \dot\gamma(t)$ (as a curve on $SM$), we have

$$\int_{\phi(\partial\Gamma)} \omega = \int_I \omega_{\dot\gamma(t)}(\ddot\gamma(t))dt = \int_I \rho_\gamma(t)dt.$$

Concluding we have from (1.35)

$$\int_\Gamma \kappa dV = \lim_{\varepsilon\to 0} \int_{\Gamma_\varepsilon} \kappa dV = 2\pi - \int_I \rho_\gamma(t)dt,$$

that is (1.33) in the smooth case (i.e., when $\alpha_i = 0$ for all $i$).
(ii). Case $\partial\Gamma$ non smooth.

We reduce to the previous case by considering a sequence of polygons $\Gamma_n$ such that $\partial\Gamma_n$ is smooth and $\Gamma_n$ approximates $\Gamma$ in a "smooth" way. In particular, we assume that $\partial\Gamma_n$ coincides with $\partial\Gamma$ except in neighborhoods $U_i$, for $i = 1, \ldots, m$, of each point $q_i$ where $\partial\Gamma$ is not smooth, in such a way that the smooth curve $\sigma_i^{(n)}$ that parametrizes $(\partial\Gamma_n \setminus \partial\Gamma) \cap U_i$ satisfies $\ell(\sigma_i^n) \leq 1/n$.

If we apply the statement of the theorem for the smooth case to $\Gamma_n$ we have

$$\int_{\Gamma_n} \kappa dV + \int \rho_{\gamma^{(n)}}(t)dt = 2\pi,$$

where $\gamma^{(n)}$ is the curve that parametrizes $\partial\Gamma_n$. Since $\Gamma_n$ tends to $\Gamma$ as $n \to \infty$, then

$$\lim_{n\to\infty} \int_{\Gamma_n} \kappa dV = \int_{\Gamma} \kappa dV.$$

We are left to prove that

$$\lim_{n\to\infty} \int \rho_{\gamma^{(n)}}(t)dt = \sum_{i=1}^{m} \int_{I_i} \rho_{\gamma_i}(t)dt + \sum_{i=1}^{m} \alpha_i. \tag{1.38}$$

For every $n$, let us split the curve $\gamma^{(n)}$ as the union of the smooth curves $\sigma_i^{(n)}$ and $\gamma_i^{(n)}$. Then

$$\int \rho_{\gamma^{(n)}}(t)dt = \sum_{i=1}^{m} \int \rho_{\gamma_i^{(n)}}(t)dt + \sum_{i=1}^{m} \int \rho_{\sigma_i^{(n)}}(t)dt.$$

Since the curve $\gamma_i^{(n)}$ tends to $\gamma_i$ for $n \to \infty$ one has

$$\lim_{n\to\infty} \int \rho_{\gamma_i^{(n)}}(t)dt = \int \rho_{\gamma_i}(t)dt.$$

Moreover, with analogous computations of part (i) of the proof

$$\int \rho_{\sigma_i^{(n)}}(t)dt = \int_{\phi(\sigma_i^{(n)})} \omega = \int_{\phi(\sigma_i^{(n)})} d\theta + a_1(x_1,x_2)dx_1 + a_2(x_1,x_2)dx_2,$$

and one has, using that $\ell(\phi(\sigma_i^{(n)})) \to 0$

$$\int_{\phi(\sigma_i^{(n)})} d\theta \underset{n\to\infty}{\longrightarrow} \alpha_i, \qquad \int_{\phi(\sigma_i^{(n)})} a_1(x_1,x_2)dx_1 + a_2(x_1,x_2)dx_2 \underset{n\to\infty}{\longrightarrow} 0.$$

Then (1.38) follows. $\square$

An important corollary is obtained by applying the Gauss-Bonnet theorem to *geodesic triangles*. A geodesic triangle $T$ is a curvilinear polygon with $m = 3$ edges and such that every smooth piece of boundary $\gamma_i$ is a geodesic. For a geodesic triangle $T$ we denote by $A_i := \pi - \alpha_i$ its internal angles, for $i = 1, 2, 3$.

**Corollary 1.33.** *Let $T$ be a geodesic triangle and $A_i(T)$ its internal angles, for $i = 1, 2, 3$. Then*

$$\kappa(q) = \lim_{|T|\to 0} \frac{\sum_{i=1}^{3} A_i(T) - \pi}{|T|}. \tag{1.39}$$

*Proof.* Fix a geodesic triangle $T$. Using that the geodesic curvature of $\gamma_i$ vanishes, the local version of Gauss-Bonnet theorem (1.33) can be rewritten as

$$\sum_{i=1}^{3} A_i = \pi + \int_\Gamma \kappa dV. \tag{1.40}$$

Dividing for $|T|$ and passing to the limit for $|T| \to 0$ in the class of geodesic triangles containing $q$ one obtains

$$\kappa(q) = \lim_{|T|\to 0} \frac{1}{|T|} \int_T \kappa dV = \lim_{|T|\to 0} \frac{\sum_{i=1}^{3} A_i(T) - \pi}{|T|}. \tag{1.41}$$

□

### 1.3.2 Gauss-Bonnet theorem: global version

Now we state the global version of the Gauss-Bonnet theorem. In other words we want to generalize (1.33) to the case when $\Gamma$ is a region of $M$ not necessarily homeomorphic to the disk, see for instance Figure 1.4. As we will see that the result depends on the Euler characteristic $\chi(\Gamma)$ of this region.

In what follows, by a triangulation of $M$ we mean a decomposition of $M$ into curvilinear polygons (see Definition 1.31). Notice that every compact surface admits a triangulation.[3]

**Definition 1.34.** Let $M \subset \mathbb{R}^3$ be a compact oriented surface with piecewise smooth boundary $\partial M$. Consider a triangulation of $M$. We define the *Euler characteristic* of $M$ as

$$\chi(M) := n_2 - n_1 + n_0, \tag{1.42}$$

where $n_i$ is the number of $i$-dimensional faces in the triangulation.

The Euler characteristic can be defined for every region $\Gamma$ of $M$ in the same way. Here, by a *region* $\Gamma$ on a surface $M$, we mean a closed domain of the manifold with piecewise smooth boundary.

*Remark* 1.35. The Euler characteristic is well-defined. Indeed one can show that the quantity (1.42) is invariant for refinement of a triangulation, since at every step of the refinement the alternating sum does not change. Moreover, given two different triangulations of the same region, there always exists a triangulation that is a refinement of both of them. This shows that the quantity (1.42) is independent on the triangulation.

**Example 1.36.** For a compact connected orientable surface $M_g$ of genus $g$ (i.e., a surface that topologically is a sphere with $g$ handles) one has $\chi(M_g) = 2 - 2g$. For instance one has $\chi(S^2) = 2$, $\chi(\mathbb{T}^2) = 0$, where $\mathbb{T}^2$ is the torus. Notice also that $\chi(B_1) = 1$, where $B_1$ is the closed unit disk in $\mathbb{R}^2$.

Following the notation introduced in the previous section, for a given region $\Gamma$, we assume that $\partial\Gamma$ is oriented consistently with the orientation of $M$ and $\partial\Gamma = \cup_{i=1}^{m} \gamma_i(I_i)$ where $\gamma_i : I_i \to M$, for $i = 1, \ldots, m$, are smooth curves parametrized by arc length (with orientation consistent with $\partial\Gamma$). We denote by $\alpha_i$ the external angles at the points where $\partial\Gamma$ is not $C^1$ (see Figure 1.4).

[3]Formally, a *triangulation* of a topological space $M$ is a simplicial complex $K$, homeomorphic to $M$, together with a homeomorphism $h : K \to M$.

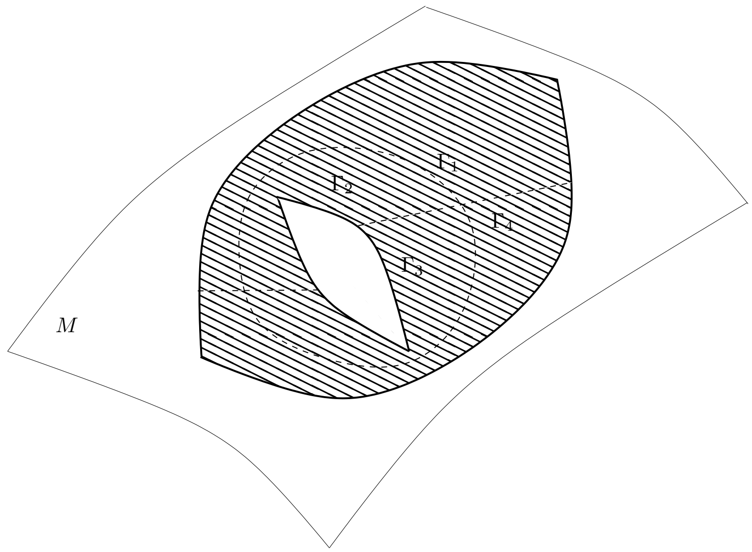



Figure 1.4: Gauss-Bonnet theorem

**Theorem 1.37** (Gauss-Bonnet, global version). *Let $\Gamma$ be a region of a surface on a compact oriented surface $M$. Then*

$$\int_\Gamma \kappa dV + \sum_{i=1}^m \int_{I_i} \rho_{\gamma_i}(t)dt + \sum_{i=1}^m \alpha_i = 2\pi\chi(\Gamma). \tag{1.43}$$

*Proof.* As in the proof of the local version of the Gauss-Bonnet theorem we consider two cases:
(i) Case $\partial\Gamma$ smooth (in particular $\alpha_i = 0$ for all $i$).
Consider a triangulation of $\Gamma$ and let $\{\Gamma_j, j = 1, \dots, n_2\}$ be the corresponding subdivision of $\Gamma$ in curvilinear polygons. We denote by $\{\gamma_k^{(j)}\}$ the smooth curves parametrized by arc length whose image are the edges of $\Gamma_j$ and by and $\theta_k^{(j)}$ the external angles of $\Gamma_j$. We assume that all orientations are chosen accordingly to the orientation of $M$. Applying Theorem 1.32 to every $\Gamma_j$ and summing w.r.t. $j$ we get

$$\sum_{j=1}^{n_2} \left( \int_{\Gamma_j} \kappa dV + \sum_k \int \rho_{\gamma_k^{(j)}}(t)dt + \sum_k \theta_k^{(j)} \right) = 2\pi n_2. \tag{1.44}$$

We have that

$$\sum_{j=1}^{n_2} \int_{\Gamma_j} \kappa dV = \int_\Gamma \kappa dV, \qquad \sum_{j,k} \int \rho_{\gamma_k^{(j)}}(t)dt = \sum_{i=1}^m \int \rho_{\gamma_i}(t)dt. \tag{1.45}$$

The second equality is a consequence of the fact that every edge of the decomposition that does not belong to $\partial\Gamma$ appears twice in the sum, with opposite sign. It remains to check that

$$\sum_{j,k} \theta_k^{(j)} = 2\pi(n_1 - n_0), \tag{1.46}$$

Let us denote by $N$ the total number of angles in the sum of the left hand side of (1.46). After reindexing we have to check that

$$\sum_{\nu=1}^{N} \theta_\nu = 2\pi(n_1 - n_0). \tag{1.47}$$

Denote by $n_0^\partial$ the number of vertexes that belong to $\partial\Gamma$ and with $n_0^I := n_0 - n_0^\partial$. Similarly we define $n_1^\partial$ and $n_1^I$. We have the following relations:

(i) $N = 2n_1^I + n_1^\partial$,

(ii) $n_0^\partial = n_1^\partial$,

Claim (i) follows from the fact that every curvilinear polygon with $n$ edges has $n$ angles, but the internal edges are counted twice since each of them appears in two polygons. Claim (ii) is a consequence of the fact that $\partial\Gamma$ is the union of closed curves. If we denote by $A_k := \pi - \theta_k$ the internal angles, we have

$$\sum_{\nu=1}^{N} \theta_\nu = N\pi - \sum_{\nu=1}^{N} A_\nu. \tag{1.48}$$

Moreover the sum of the internal angles is equal to $\pi$ for a boundary vertex, and to $2\pi$ for an internal one. Hence one gets

$$\sum_{\nu=1}^{N} A_\nu = 2\pi n_0^I + \pi n_0^\partial, \tag{1.49}$$

Combining (1.48), (1.49) and (i) one has

$$\sum_{i=1}^{\nu} \theta_\nu = (2n_1^I + n_1^\partial)\pi - (2n_0^I + n_0^\partial)\pi$$

Using (ii) one finally gets (1.47).

(ii) Case $\partial\Gamma$ non-smooth.

We consider a decomposition of $\Gamma$ into curvilinear polygons whose edges intersect the boundary in the smooth part (this is always possible). The proof is identical to the smooth case up to formula (1.45). Now, instead of (1.47), we have to check that

$$\sum_{\nu=1}^{N} \theta_\nu = \sum_{i=1}^{m} \alpha_i + 2\pi(n_1 - n_0), \tag{1.50}$$

Now (1.50) can be rewritten as

$$\sum_{\nu \notin \mathcal{A}} \theta_\nu = 2\pi(n_1 - n_0),$$

where $\mathcal{A}$ is the set of indices whose corresponding angles are non smooth points of $\partial\Gamma$.

Consider now a new region $\widetilde{\Gamma}$, obtained by smoothing the edges of $\Gamma$, together with the decomposition induced by $\Gamma$ (see Figure 1.4). Denote by $\widetilde{n}_1$ and $\widetilde{n}_0$ the number of edges and vertexes of the decomposition of $\widetilde{\Gamma}$. Notice that $\{\theta_\nu, \nu \notin \mathcal{A}\}$ is exactly the set of all angles of the decomposition of $\widetilde{\Gamma}$. Moreover $\widetilde{n}_1 - \widetilde{n}_0 = n_1 - n_0$, since $n_0 = \widetilde{n}_0 + m$ and $n_1 = \widetilde{n}_1 + m$, where $m$ is the number of non-smooth points. Hence, by part (i) of the proof:

$$\sum_{\nu \notin \mathcal{A}} \theta_\nu = 2\pi(\widetilde{n}_1 - \widetilde{n}_0) = 2\pi(n_1 - n_0). \qquad \square$$

**Corollary 1.38.** *Let $M$ be a compact oriented surface without boundary. Then*

$$\int_M \kappa dV = 2\pi\chi(M). \tag{1.51}$$

### 1.3.3 Consequences of the Gauss-Bonnet theorems

**Definition 1.39.** Let $M, M'$ be two surfaces in $\mathbb{R}^3$. A smooth map $\phi : \mathbb{R}^3 \to \mathbb{R}^3$ is called a *local isometry* between $M$ and $M'$ if $\phi(M) = M'$ and for every $q \in M$ it satisfies

$$\langle v \,|\, w \rangle = \langle D_q\phi(v) \,|\, D_q\phi(w) \rangle , \qquad \forall\, v, w \in T_qM. \tag{1.52}$$

If, moreover, the map $\phi$ is a bijection, then $\phi$ is called a *global isometry.*

Two surfaces $M$ and $M'$ are said to be *locally isometric* (resp. *globally isometric*) if there exists a local isometry (resp. global isometry) between $M$ and $M'$. Notice that the restriction $\phi$ of an isometry of $\mathbb{R}^3$ to a surface $M \subset \mathbb{R}^3$ always defines a global isometry between $M$ and $M' = \phi(M)$.

Formula (1.52) says that a local isometry between two surfaces $M$ and $M'$ preserves the angles between tangent vectors and, a fortiori, the length of curves and the distance between points.

By Corollary 1.33, thanks to the fact that the angles and the volumes are preserved by isometries, one obtains that the Gaussian curvature is invariant by local isometries, in the following sense.

**Theorem 1.40** (Gauss's Theorema Egregium)**.** *Let $\phi$ is a local isometry between $M$ and $M'$. Then for every $q \in M$ one has $\kappa(q) = \kappa'(\phi(q))$, where $\kappa$ (resp. $\kappa'$) is the Gaussian curvature of $M$ (resp. $M'$).*

The previous result says that the Gaussian curvature $\kappa$ depends only on the metric structure on $M$ and not on the specific fact that the surface is embedded in $\mathbb{R}^3$ with the induced inner product.

**Exercise 1.41.** (i). Prove that the Euclidean plane $\mathbb{R}^2$ has vanishing Gaussian curvature.

(ii). Prove that a surface $M$ is locally isometric to the Euclidean plane $\mathbb{R}^2$ if and only if there exists a coordinate system $(x_1, x_2)$ in a neighborhood of each point $q \in M$ such that the vectors $\partial_{x_1}$ and $\partial_{x_2}$ have unit length and are everywhere orthonormal.

**Corollary 1.42.** *Let $M$ be a surface and $q \in M$. If $\kappa(q) \neq 0$ then $M$ is not locally isometric to $\mathbb{R}^2$ in a neighborhood of $q$.*

As a converse of Corollary 1.42 we have the following.

**Theorem 1.43.** *Assume that $\kappa \equiv 0$ in a neighborhood $U$ of a point $q \in M$. Then $M$ is locally Euclidean (i.e., locally isometric to $\mathbb{R}^2$) on $U$.*

*Proof.* From our assumptions we have, in a neighborhood $U$ of $q$:

$$\Omega = \kappa dV = 0.$$

Hence $d\omega = \pi^*\Omega = 0$. From the explicit expression

$$\omega = d\theta + a_1(x_1, x_2)dx_1 + a_2(x_1, x_2)dx_2,$$

it follows that the 1-form $a_1dx_1 + a_2dx_2$ is locally exact, i.e., there exists a neighborhood $W$ of $q$, $W \subset U$, and a function $\phi : W \to \mathbb{R}$ such that $a_1(x_1, x_2)dx_1 + a_2(x_1, x_2)dx_2 = d\phi$. Hence

$$\omega = d(\theta + \phi(x_1, x_2)).$$

Thus we can define a new angular coordinate on $SM$ (cf. also Remark 1.26), which we still denote by $\theta$, in such a way that

$$\omega = d\theta. \tag{1.53}$$

Now, let $\gamma$ be an arc length parametrized geodesic, i.e., $\omega_{\dot\gamma(t)}(\ddot\gamma(t)) = 0$. Using the the angular coordinate $\theta$ just defined on the fibers of $SM$, the curve $t \mapsto \dot\gamma(t) \in S_{\gamma(t)}M$ is written as $t \mapsto \theta(t)$. Using (1.53), we have then

$$0 = \omega_{\dot\gamma(t)}(\ddot\gamma(t)) = d\theta(\ddot\gamma(t)) = \dot\theta(t).$$

In other words the angular coordinate along a geodesic $\gamma$ is constant.

We want to construct Cartesian coordinates in a neighborhood $U$ of $q$. Consider the two length parametrized geodesics $\gamma_1$ and $\gamma_2$ starting from $q$ and such that $\theta_1(0) = 0$, $\theta_2(0) = \pi/2$. Define them to be the $x_1$-axes and $x_2$-axes of our coordinate system, respectively.

Then, for each point $q' \in U$ consider the two geodesics starting from $q'$ and satisfying $\theta_1(0) = 0$ and $\theta_2(0) = \pi/2$. We assign coordinates $(x_1, x_2)$ to each point $q'$ in $U$ by considering the length parameter of the geodesic projection of $q'$ on $\gamma_1$ and $\gamma_2$ (See Figure 1.5). Notice that the family of geodesics constructed in this way, and parametrized by $q' \in U$, are mutually orthogonal at every point.

By construction, in this coordinate system the vectors $\partial_{x_1}$ and $\partial_{x_2}$ have length one (being the tangent vectors to length parametrized geodesics) and are everywhere mutually orthogonal. Hence the statement follows from part (ii) of Exercise 1.41. □

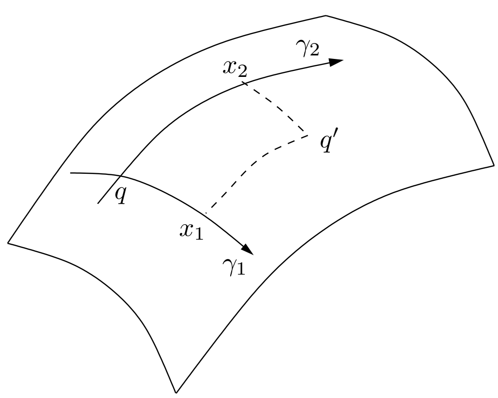



Figure 1.5: Proof of Theorem 1.43.

### 1.3.4 The Gauss map

We end this section with a geometric characterization of the Gaussian curvature of a manifold $M$, using the Gauss map. The Gauss map is a map from the surface $M$ to the unit sphere $S^2$ of $\mathbb{R}^3$.

**Definition 1.44.** Let $M$ be an oriented surface. We define the *Gauss map* associated with $M$ as

$$\mathcal{N} : M \to S^2, \qquad q \mapsto \nu_q, \tag{1.54}$$

where $\nu_q \in S^2 \subset \mathbb{R}^3$ denotes the external unit normal vector to $M$ at $q$.

Let us consider the differential of the Gauss map at the point $q$

$$D_q\mathcal{N} : T_qM \to T_{\mathcal{N}(q)}S^2.$$

Notice that a tangent vector to the sphere $S^2$ at $\mathcal{N}(q)$ is by construction orthogonal to $\mathcal{N}(q)$. Hence it is possible to identify $T_{\mathcal{N}(q)}S^2$ with $T_qM$ and think to the differential of the Gauss map $D_q\mathcal{N}$ as an endomorphism of $T_qM$.

**Theorem 1.45.** *Let $M$ be a surface of $\mathbb{R}^3$ with Gauss map $\mathcal{N}$ and Gaussian curvature $\kappa$. Then*

$$\kappa(q) = \det(D_q\mathcal{N}), \tag{1.55}$$

*where $D_q\mathcal{N}$ is interpreted as an endomorphism $T_qM$.*

We start by proving an important property of the Gauss map.

**Lemma 1.46.** *For every $q \in M$, the differential $D_q\mathcal{N}$ of the Gauss map is a symmetric operator, i.e., it satisfies*

$$\langle D_q\mathcal{N}(\xi) \,|\, \eta\rangle = \langle \xi \,|\, D_q\mathcal{N}(\eta)\rangle\,, \qquad \forall\, \xi, \eta \in T_qM. \tag{1.56}$$

*Proof.* The statement is local, hence it is not restrictive to assume that $M$ parametrized by a function $\phi : \mathbb{R}^2 \to M$. In this case $T_qM = \operatorname{Im} D_u\phi$, where $\phi(u) = q$. Let $v, w \in \mathbb{R}^2$ such that $\xi = D_u\phi(v)$ and $\eta = D_u\phi(w)$. Since $\mathcal{N}(q) \in T_qM^\perp$ we have

$$\langle \mathcal{N}(q) \,|\, \eta\rangle = \langle \mathcal{N}(q) \,|\, D_u\phi(w)\rangle = 0.$$

Differentiating the last identity in the direction of $\xi$, one gets

$$\langle D_q\mathcal{N}(\xi) \,|\, \eta\rangle + \left\langle \mathcal{N}(q) \,\middle|\, D_u^2\phi(v, w)\right\rangle = 0,$$

where $D_u^2\phi$ is the second differential. Exchanging the role of $v$ and $w$ in the previous argument, and using that $D_u^2\phi$ is a bilinear symmetric map, the identity (1.56) follows. □

The proof of Theorem 1.45 relies on the general Cartan's moving frame method, which is based on the following idea. Fix $\xi \in SM$, and denote by

$$(e_1(\xi), e_2(\xi), e_3(\xi)), \qquad e_i : SM \to \mathbb{R}^3, \tag{1.57}$$

the canonical orthonormal basis $(\xi, \eta, \nu)$ attached at $\xi$ and constructed as in Section 1.2. In particular $e_1(\xi) = \xi$ for every $\xi \in SM$. Then one computes the differentials of these maps (taking values in the ambient space $\mathbb{R}^3$) and writes them as linear combinations of the vectors $e_i$ as follows

$$d_\xi e_i(z) = \sum_{j=1}^{3} (\omega_\xi)_{ij}(z)\, e_j(\xi), z \in T_\xi SM. \tag{1.58}$$

where each coefficient $\omega_{ij} \in \Lambda^1 SM$ is a differential 1-form. Dropping $\xi$ and $z$ from the notation in (1.58), the previous identity is rewritten as

$$de_i = \sum_{j=1}^{3} \omega_{ij}\, e_j, \qquad \omega_{ij} \in \Lambda^1 SM. \tag{1.59}$$

Since for each $\xi$ the basis $(e_1(\xi), e_2(\xi), e_3(\xi))$ is orthonormal, the derivative of each vector is orthogonal to the vector itself. It follows that the matrix $\omega$ has vanishing coefficient on the principal diagonal, and is actually is skew-symmetric[4]. It follows that

$$\begin{array}{rlll} de_1 = & & \omega_{12}e_2 & +\omega_{13}e_3 \\ de_2 = & -\omega_{12}e_1 & & +\omega_{23}e_3 \\ de_3 = & -\omega_{13}e_1 & -\omega_{23}e_2 & \end{array} \tag{1.60}$$

**Lemma 1.47.** *We have the identity*

$$\omega_{13} \wedge \omega_{23} = d\omega_{12}. \tag{1.61}$$

*Proof.* Differentiating the first equation in (1.60) one gets, using that $d^2 = 0$,

$$\begin{aligned} 0 = d^2 e_1 &= d\omega_{12}e_2 + \omega_{12} \wedge de_2 + d\omega_{13}e_3 + \omega_{13} \wedge de_3 \\ &= (d\omega_{12} - \omega_{13} \wedge \omega_{23})e_2 + (d\omega_{13} - \omega_{12} \wedge \omega_{23})e_3, \end{aligned}$$

which implies in particular (1.61). □

*Remark* 1.48. By construction, the 1-form $\omega_{12}$ computes the coefficient of the derivative of the first vector of the orthonormal basis along the second one. It means that $\omega_{12} = \omega$, where $\omega$ is the Levi-Civita connection (cf. also Definition 1.54).

Before proving Theorem 1.45, let us recall the following linear algebra property.

**Exercise 1.49.** Let $V$ be a 2-dimensional Euclidean vector space and let $e_1, e_2$ be an orthonormal basis of $V$. Let $F : V \to V$ a linear map and write $F = F_1 e_1 + F_2 e_2$, where $F_i : V \to \mathbb{R}$ are linear functionals. Prove that $F_1 \wedge F_2 = (\det F)dV$, where $dV$ is the area form induced by the inner product.

*Proof of Theorem 1.45.* The statement can be rewritten as an identity between differential 2-forms on $M$ as follows

$$\kappa dV = \det(D_q\mathcal{N})dV, \tag{1.62}$$

where $dV$ denotes the area form on $M$. Applying the pullback $\pi^*$ of the canonical projection $\pi : SM \to M$ to both sides of 1.62 one gets

$$d\omega = \pi^*(\kappa dV) = \pi^*(\det(D_q\mathcal{N})dV) \tag{1.63}$$

Combining Lemma 1.47 and Remark 1.48, it is sufficient to prove the identity

$$\omega_{13} \wedge \omega_{23} = \pi^*(\det(D_q\mathcal{N})dV) = \det(D_{\pi(\xi)}\mathcal{N})\pi^* dV. \tag{1.64}$$

Since $e_3 = \mathcal{N} \circ \pi$, where $\pi : SM \to M$ is the canonical projection, one has

$$D_q\mathcal{N} \circ \pi_* = de_3 = -\omega_{13}e_1 - \omega_{23}e_2.$$

The identity (1.64) then follows by Exercice 1.49. □

[4]this can be seen as a consequence of the fact that $(e_1(\xi), e_2(\xi), e_3(\xi))$ defines an element of the Lie group $SO(3)$ hence $\omega$ belongs to its tangent space, that is the Lie algebra of skew-symmetric matrices $\mathfrak{so}(3)$.

**Further comments**

Lemma 1.46 allows us to define the *principal curvatures* of $M$ at the point $q$ as the two real eigenvalues $k_1(q), k_2(q)$ of the map $D_q\mathcal{N}$. In particular

$$\kappa(q) = k_1(q)k_2(q), \qquad q \in M.$$

The principal curvatures at $q$ can be geometrically interpreted as the maximum and the minimum of the curvature of the curves obtained by intersecting $M$ with planes passing through $q$ and orthogonal to $T_qM$.

Notice moreover that, using the Gauss-Bonnet theorem, one can relate then degree of the map $\mathcal{N}$ with the Euler characteristic of $M$ as follows

$$\deg \mathcal{N} := \frac{1}{|S^2|}\int_M (\det D_q\mathcal{N})dV = \frac{1}{4\pi}\int_M \kappa dV = \frac{1}{2}\chi(M).$$

where $|S^2|$ denotes the area of the unit sphere $S^2$ of $\mathbb{R}^3$.

## 1.4 Surfaces in $\mathbb{R}^3$ with the Minkowski inner product

The theory and the results obtained in this chapter can be adapted to the case when $M \subset \mathbb{R}^3$ is a surface in the Minkowski 3-space, that is $\mathbb{R}^3$ endowed with the hyperbolic (or pseudo-Euclidean) inner product

$$\langle q_1, q_2\rangle_h = x_1x_2 + y_1y_2 - z_1z_2. \tag{1.65}$$

Here $q_i = (x_i, y_i, z_i)$ for $i = 1, 2$, are two points in $\mathbb{R}^3$. When $\langle q, q\rangle_h \geq 0$, we denote by $\|q\|_h = \langle q, q\rangle_h^{1/2}$ the length of the vector induced by the inner product (1.65).

For the metric structure to be well-defined on $M$, we should require that the restriction of the inner product (1.65) to the tangent space to $M$ is positive definite at every point. Indeed, under this assumption, the inner product (1.65) can be used to define the length of a tangent vector to the surface (which is non-negative). Thus one can introduce the length of (piecewise) smooth curves on $M$ and its distance by the same formulas as in Section 1.1. These surfaces are also called *space-like* surfaces in the Minkowski space.

The structure of the inner product imposes some conditions on the structure of space-like surfaces, as the following exercice shows.

**Exercise 1.50.** Let $M$ be a space-like surface in $\mathbb{R}^3$ endowed with the inner product (1.65).

(i) Show that if $v \in T_qM$ is a non zero vector that is orthogonal to $T_qM$, then $\langle v, v\rangle_h < 0$.

(ii) Prove that, if $M$ is compact, then $\partial M \neq \emptyset$.

(iii) Show that the restriction to $M$ of the projection $\pi : \mathbb{R}^3 \to \mathbb{R}^2$ defined by $\pi(x, y, z) = (x, y)$ is a local diffeomorphism.

(iv) $M$ is locally a graph, i.e., for every point in $M$ there exists $U \subset \mathbb{R}^3$ such that $M \cap U = \{(x, y, z) \mid z = f(x, y)\}$, for a suitable smooth function $f$.

The results obtained in the previous sections for surfaces embedded in $\mathbb{R}^3$ can be recovered for space-like surfaces by simply adapting all formulas to their "hyperbolic" counterpart. For instance, *geodesics* are defined as curves of unit speed whose second derivative is orthogonal, with respect to $\langle\cdot\,|\,\cdot\rangle_h$, to the tangent space to $M$.

For a smooth function $a:\mathbb{R}^3\to\mathbb{R}$, its *hyperbolic gradient* $\nabla^h_q a$ is defined as

$$\nabla^h_q a = \left(\frac{\partial a}{\partial x}, \frac{\partial a}{\partial y}, -\frac{\partial a}{\partial z}\right).$$

If we assume that $M=a^{-1}(0)$ is a regular level set of a smooth function $a:\mathbb{R}^3\to\mathbb{R}$ and $\gamma(t)$ is a curve contained in $M$ (i.e., $a(\gamma(t))=0$), one has the identity

$$0 = \left\langle \nabla^h_{\gamma(t)} a \,\middle|\, \dot\gamma(t) \right\rangle_h.$$

The same computation shows that $\nabla^h_{\gamma(t)} a$ is orthogonal to the level sets of $a$, where orthogonal always means with respect to $\langle\cdot\,|\,\cdot\rangle_h$. In particular, if $M=a^{-1}(0)$ is space-like, one has $\langle\nabla_q a, \nabla_q a\rangle_h < 0$.

**Exercise 1.51.** Let $\gamma$ be a geodesic on $M=a^{-1}(0)$. Show that $\gamma$ satisfies the equation (in matrix notation)

$$\ddot\gamma(t) = -\frac{\dot\gamma(t)^T(\nabla^2_{\gamma(t)}a)\dot\gamma(t)}{\|\nabla^h_{\gamma(t)}a\|_h^2}\nabla^h_{\gamma(t)}a, \qquad \forall\, t\in[0,T]. \tag{1.66}$$

where $\nabla^2_{\gamma(t)}a$ is the (classical) matrix of second derivatives of $a$.[5]

Given a smooth curve $\gamma:[0,T]\to M$ on a surface $M$, a smooth curve of tangent vectors $\xi(t)\in T_{\gamma(t)}M$ is said to be *parallel* if $\dot\xi(t)\perp T_{\gamma(t)}M$, with respect to the hyperbolic inner product. It is then straightforward to check that, if $M$ is the zero level of a smooth function $a:\mathbb{R}^3\to\mathbb{R}$, then $\xi(t)$ is parallel along $\gamma$ if and only if it satisfies

$$\dot\xi(t) = -\frac{\dot\gamma(t)^T(\nabla^2_{\gamma(t)}a)\xi(t)}{\|\nabla^h_{\gamma(t)}a\|_h^2}\nabla^h_{\gamma(t)}a, \qquad \forall\, t\in[0,T]. \tag{1.67}$$

By definition a smooth curve $\gamma:[0,T]\to M$ is a geodesic if and only if $\dot\gamma$ is parallel along $\gamma$.

*Remark* 1.52. As for surfaces in the Euclidean space, given curve $\gamma:[0,T]\to M$ and initial datum $v\in T_{\gamma(0)}M$, there is a unique parallel curve of tangent vectors $\xi(t)\in T_{\gamma(t)}M$ along $\gamma$ such that $\xi(0)=v$. Moreover the operator $\xi(0)\mapsto\xi(t)$ is a linear operator, which the *parallel transport* of $v$ along $\gamma$.

**Exercise 1.53.** Show that if $\xi(t),\eta(t)$ are two parallel curves of tangent vectors along the same curve $\gamma$, then we have

$$\frac{d}{dt}\langle\xi(t)\,|\,\eta(t)\rangle_h = 0, \qquad \forall\, t\in[0,T]. \tag{1.68}$$

Assume that $M$ is oriented. Given an element $\xi\in S_qM$ we can complete it to an orthonormal frame $(\xi,\eta,\nu)$ of $\mathbb{R}^3$ in the following unique way:

[5] otherwise one can write the numerator of (1.66) as $\left\langle \nabla^{2,h}_{\gamma(t)}\dot\gamma(t) \,\middle|\, \dot\gamma(t)\right\rangle_h$, where $\nabla^{2,h}_{\gamma(t)}$ is the *hyperbolic Hessian.*

(i) $\eta \in T_q M$ is orthogonal to $\xi$ with respect to $\langle \cdot \mid \cdot \rangle_h$ and $(\xi, \eta)$ is positively oriented (w.r.t. the orientation of $M$),

(ii) $\nu \perp T_q M$ with respect to $\langle \cdot \mid \cdot \rangle_h$ and $(\xi, \eta, \nu)$ is positively oriented (w.r.t. the orientation of $\mathbb{R}^3$).

For a smooth curve of unit tangent vectors $\xi(t) \in S_{\gamma(t)} M$ along a curve $\gamma : [0, T] \to M$ we define $\eta(t), \nu(t) \in T_{\gamma(t)} M$ and we can write

$$\dot{\xi}(t) = u_\xi(t)\eta(t) + v_\xi(t)\nu(t).$$

**Definition 1.54.** The *hyperbolic Levi-Civita connection* on $M$ is the 1-form $\omega \in \Lambda^1(SM)$ defined by

$$\omega_\xi : T_\xi SM \to \mathbb{R}, \qquad \omega_\xi(z) = u_z, \tag{1.69}$$

where $z = u_z \eta + v_z \nu$ and $(\xi, \eta, \nu)$ is the orthonormal frame defined above.

It is again easy to check that a curve of unit tangent vectors $\xi(t)$ is parallel if and only if $\omega_{\xi(t)}(\dot{\xi}(t)) = 0$ and a curve parametrized by arc length $\gamma : [0, T] \to M$ is a geodesic if and only if

$$\omega_{\dot{\gamma}(t)}(\ddot{\gamma}(t)) = 0, \qquad \forall\, t \in [0, T]. \tag{1.70}$$

**Exercise 1.55.** Prove that the hyperbolic Levi Civita connection $\omega \in \Lambda^1(SM)$ satisfies:

(i) there exist two smooth functions $a_1, a_2 : M \to \mathbb{R}$ such that

$$\omega = d\theta + a_1(x_1, x_2)dx_1 + a_2(x_1, x_2)dx_2, \tag{1.71}$$

where $(x_1, x_2, \theta)$ is a system of coordinates on $SM$.

(ii) $d\omega = \pi^*\Omega$, where $\Omega$ is a 2-form defined on $M$ and $\pi : SM \to M$ is the canonical projection.

Again one can introduce the area form $dV$ on $M$ induced by the inner product and it makes sense to give the following definition:

**Definition 1.56.** The *Gaussian curvature* of a surface $M$ in the Minkowski 3-space is the function $\kappa : M \to \mathbb{R}$ defined by the equality

$$\Omega = -\kappa dV. \tag{1.72}$$

By reasoning as in the Euclidean case, one can define the geodesic curvature of a curve and prove the analogue of the Gauss-Bonnet theorem in this context. As a consequence one gets that the Gaussian curvature is again invariant under isometries of $M$ and hence is an intrinsic quantity that depends only on the metric properties of the surface and not on the fact that its metric is obtained as the restriction of some metric defined in the ambient space.

Finally one can define the hyperbolic Gauss map.

**Definition 1.57.** Let $M$ be an oriented surface. We define the *Gauss map*

$$\mathcal{N} : M \to H^2, \qquad q \mapsto \nu_q, \tag{1.73}$$

where $\nu_q \in H^2 \subset \mathbb{R}^3$ denotes the external unit normal vector to $M$ at $q$, with respect to the Minkowski inner product.

Let us now consider the differential of the Gauss map at the point $q$:

$$D_q\mathcal{N} : T_qM \to T_{\mathcal{N}(q)}H^2 \simeq T_qM$$

where an element tangent to the hyperbolic plane $H^2$ at $\mathcal{N}(q)$, being orthogonal to $\mathcal{N}(q)$, is identified with a tangent vector to $M$ at $q$.

**Theorem 1.58.** *The differential of the Gauss map $D_q\mathcal{N}$ is symmetric, and $\kappa(q) = \det(D_q\mathcal{N})$.*

## 1.5 Model spaces of constant curvature

In this section we briefly discuss surfaces embedded in $\mathbb{R}^3$ (with Euclidean or Minkowski inner product) that have constant Gaussian curvature, playing the role of model spaces. For each model we are interested in describing geodesics and, more generally, curves of constant geodesic curvature. These results will be useful in the study of sub-Riemannian model spaces in dimension three (cf. Chapter 7).

Assume that the surface $M$ has constant Gaussian curvature $\kappa \in \mathbb{R}$. We already know that $\kappa$ is a metric invariant of the surface, i.e., it does not depend on the embedding of the surface in $\mathbb{R}^3$. We will distinguish the following three cases:

(i) $\kappa = 0$: this is the flat model, corresponding to the Euclidean plane,

(ii) $\kappa > 0$: these corresponds to the case of the sphere,

(iii) $\kappa < 0$: these corresponds to the hyperbolic plane.

We will briefly discuss the case (i), since it is trivial, and study in some more detail the cases (ii) and (iii) of spherical and hyperbolic geometry.

### 1.5.1 Zero curvature: the Euclidean plane

The Euclidean plane can be realized as the surface of $\mathbb{R}^3$ defined by the zero level set of the function

$$a : \mathbb{R}^3 \to \mathbb{R}, \qquad a(x, y, z) = z.$$

It is an easy exercise, applying the results of the previous sections, to show that the Gaussian curvature of this surface is zero (the Gauss map is constant) and to characterize geodesics and curves with constant geodesic curvature.

**Exercise 1.59.** Prove that geodesics on the Euclidean plane are lines. Moreover, show that curves with constant geodesic curvature $c \neq 0$ are circles of radius $1/c$.

### 1.5.2 Positive curvature: the sphere

Let us consider the sphere $S^2_r$ of radius $r$ as the surface of $\mathbb{R}^3$ defined as the zero level set of the function

$$S^2_r = a^{-1}(0), \qquad a(x, y, z) = x^2 + y^2 + z^2 - r^2. \tag{1.74}$$

If we denote, as usual, with $\langle \cdot \,|\, \cdot \rangle$ the Euclidean inner product in $\mathbb{R}^3$, $S^2_r$ can be viewed also as the set of points $q = (x, y, z)$ whose Euclidean norm is constant

$$S^2_r = \{q \in \mathbb{R}^3 \mid \langle q \,|\, q \rangle = r^2\}.$$

The Gauss map associated with this surface can be easily computed and it is explicitly given by

$$\mathcal{N} : S^2_r \to S^2, \qquad \mathcal{N}(q) = \frac{1}{r} q, \tag{1.75}$$

It follows immediately by (1.75) that the Gaussian curvature of the sphere is $\kappa = 1/r^2$ at every point $q \in S^2_r$. Let us now recover the structure of geodesics and curves with constant geodesic curvature on the sphere.

**Proposition 1.60.** *Let $\gamma : [0, T] \to S^2_r$ be a curve with unit speed and constant geodesic curvature equal to $c \in \mathbb{R}$. Then, for every $w \in \mathbb{R}^3$, the function $\alpha(t) = \langle \dot{\gamma}(t) \,|\, w \rangle$ is a solution of the differential equation*

$$\ddot{\alpha}(t) + \left(c^2 + \frac{1}{r^2}\right) \alpha(t) = 0.$$

*Proof.* Differentiating twice the equality $a(\gamma(t)) = 0$, where $a$ is the function defined in (1.74), we get (in matrix notation):

$$\dot{\gamma}(t)^T (\nabla^2_{\gamma(t)} a) \dot{\gamma}(t) + \ddot{\gamma}(t)^T \nabla_{\gamma(t)} a = 0.$$

Moreover, since $\|\dot{\gamma}(t)\|$ is constant and $\gamma$ has constant geodesic curvature equal to $c$, there exists a function $b(t)$ such that

$$\ddot{\gamma}(t) = b(t) \nabla_{\gamma(t)} a + c\eta(t), \tag{1.76}$$

where $c$ is the geodesic curvature of the curve and $\eta(t) = \dot{\gamma}(t)^{\perp}$ is the vector orthogonal to $\dot{\gamma}(t)$ in $T_{\gamma(t)} S^2_r$ (defined in such a way that $\dot{\gamma}(t)$ and $\eta(t)$ is a positively oriented frame). Reasoning as in the proof of Proposition 1.8 and noticing that $\nabla_{\gamma(t)} a$ is proportional to the vector $\gamma(t)$, one can compute $b(t)$ and obtains that $\gamma$ satisfies the differential equation

$$\ddot{\gamma}(t) = -\frac{1}{r^2} \gamma(t) + c\eta(t). \tag{1.77}$$

**Claim.** We have $\dot{\eta}(t) = -c\dot{\gamma}(t)$.

The curve $\eta(t)$ has constant norm, hence $\dot{\eta}(t)$ is orthogonal to $\eta(t)$. Recall that the triple $(\gamma(t), \dot{\gamma}(t), \eta(t))$ defines an orthogonal frame at every point. Differentiating the identity $\langle \eta(t) \,|\, \gamma(t) \rangle = 0$ with respect to $t$ one has

$$0 = \langle \dot{\eta}(t) \,|\, \gamma(t) \rangle + \langle \eta(t) \,|\, \dot{\gamma}(t) \rangle = \langle \dot{\eta}(t) \,|\, \gamma(t) \rangle.$$

Hence $\dot{\eta}(t)$ has non vanishing component only along $\dot{\gamma}(t)$. Differentiating the identity $\langle \eta(t) \,|\, \dot{\gamma}(t) \rangle = 0$ one obtains

$$0 = \langle \dot{\eta}(t) \,|\, \dot{\gamma}(t) \rangle + \langle \eta(t) \,|\, \ddot{\gamma}(t) \rangle = \langle \dot{\eta}(t) \,|\, \dot{\gamma}(t) \rangle + c,$$

where we used (1.77). Hence $\dot{\eta}(t) = \langle \dot{\eta}(t) \,|\, \dot{\gamma}(t) \rangle \dot{\gamma}(t) = -c\dot{\gamma}(t)$, which proves the claim.

Next we compute the derivatives of the function $\alpha$ as follows

$$\dot{\alpha}(t) = \langle \ddot{\gamma}(t) \,|\, w \rangle = -\frac{1}{r^2} \langle \gamma(t) \,|\, w \rangle + c \langle \eta(t) \,|\, w \rangle. \tag{1.78}$$

Using the above claim, we have

$$\ddot{\alpha}(t) = -\frac{1}{r^2}\langle \dot{\gamma}(t) \,|\, w\rangle + c\,\langle \dot{\eta}(t) \,|\, w\rangle \tag{1.79}$$

$$= -\frac{1}{r^2}\langle \dot{\gamma}(t) \,|\, w\rangle - c^2\,\langle \dot{\gamma}(t) \,|\, w\rangle = -\left(\frac{1}{r^2} + c^2\right)\alpha(t). \tag{1.80}$$

which ends the proof of the Proposition 1.60. □

**Corollary 1.61.** *Constant geodesic curvature curves are contained in the intersection of $S_r^2$ with an affine plane of $\mathbb{R}^3$. In particular, geodesics are contained in the intersection of $S_r^2$ with planes passing through the origin, i.e., great circles.*

*Proof.* Let $\gamma$ be a curve with constant geodesic curvature. Without loss of generality we can assume that $\gamma$ is arc length parametrized. Let us fix a vector $w \in \mathbb{R}^3$ that is orthogonal to $\dot{\gamma}(0)$ and $\ddot{\gamma}(0)$. By Proposition 1.60, the function $\alpha(t) := \langle \dot{\gamma}(t) \,|\, w\rangle = 0$ is a solution of the Cauchy problem

$$\begin{cases} \ddot{\alpha}(t) + (\frac{1}{r^2} + c^2)\alpha(t) = 0 \\ \alpha(0) = \dot{\alpha}(0) = 0 \end{cases} \tag{1.81}$$

By uniqueness of the solution to the Cauchy problem (1.81), we have $\alpha(t) = 0$ for all $t$. This means that $\dot{\gamma}(t)$ is contained in a plane orthogonal to $w$.

If the curve is a geodesic, then $c = 0$ and the geodesic equation (1.77) is written as $\ddot{\gamma}(t) = -\frac{1}{r^2}\gamma(t)$. Then consider the function $\Gamma(t) := \langle \gamma(t) \,|\, w\rangle$, where $w$ is chosen as before. $\Gamma(t)$ is constant since $\dot{\Gamma}(t) = \alpha(t) = 0$. In fact $\Gamma(t)$ is identically zero since $\Gamma(0) = \langle \gamma(0) \,|\, w\rangle = -\frac{1}{r^2}\langle \ddot{\gamma}(0) \,|\, w\rangle = 0$, by the assumption on $w$. This proves that the curve $\gamma$ is contained in a plane passing through the origin. □

*Remark* 1.62. Curves with constant geodesic curvatures on the spheres are circles obtained as the intersection of the sphere with an affine plane. Moreover all these curves can be also characterized in the following two ways:

(i) curves that have constant distance from a geodesic (equidistant curves),

(ii) boundary of metric balls (spheres).

### 1.5.3 Negative curvature: the hyperbolic plane

The negative constant curvature model is the hyperbolic plane $H_r^2$ obtained as the surface of $\mathbb{R}^3$, endowed with the hyperbolic metric, defined as the zero level set of the function

$$a(x, y, z) = x^2 + y^2 - z^2 + r^2. \tag{1.82}$$

Indeed this surface is a two-fold hyperboloid, so we restrict our attention to the set of points $H_r^2 = a^{-1}(0) \cap \{z > 0\}$.

In analogy with the positive constant curvature model (which is the set of points in $\mathbb{R}^3$ whose euclidean norm is constant) the negative constant curvature can be seen as the set of points whose hyperbolic norm is constant in $\mathbb{R}^3$. In other words

$$H_r^2 = \{q = (x, y, z) \in \mathbb{R}^3 \mid \|q\|_h^2 = -r^2\} \cap \{z > 0\}.$$

The hyperbolic Gauss map associated with this surface can be easily computed since it is explicitly given by

$$\mathcal{N} : H^2_r \to H^2, \qquad \mathcal{N}(q) = \frac{1}{r}\nabla_q a, \tag{1.83}$$

**Exercise 1.63.** Prove that the Gaussian curvature of $H^2_r$ is $\kappa = -1/r^2$ at every point $q \in H^2_r$.

We can now discuss the structure of geodesics and curves with constant geodesic curvature on the hyperbolic space. We start with a result than can be proved in an analogous way to Proposition 1.60. The proof is left to the reader.

**Proposition 1.64.** *Let $\gamma : [0, T] \to H^2_r$ be a curve with unit speed and constant geodesic curvature equal to $c \in \mathbb{R}$. For every vector $w \in \mathbb{R}^3$, the function $\alpha(t) = \langle \dot\gamma(t) \,|\, w\rangle_h$ is a solution of the differential equation*

$$\ddot\alpha(t) + \left(c^2 - \frac{1}{r^2}\right)\alpha(t) = 0. \tag{1.84}$$

As for the sphere, this result implies immediately the following corollary.

**Corollary 1.65.** *Constant geodesic curvature curves on $H^2_r$ are contained in the intersection of $H^2_r$ with affine planes of $\mathbb{R}^3$. In particular, geodesics are contained in the intersection of $H^2_r$ with planes passing through the origin.*

**Exercise 1.66.** Prove Proposition 1.64 and Corollary 1.65.

Geodesics on $H^2_r$ are hyperbolas, obtained as intersections of the hyperboloid with plane passing through the origin. The classification of curves with constant geodesic curvature is in fact more rich. The sections of the hyperboloid with affine planes can have different shapes depending on the Euclidean orthogonal vector to the plane: they are circles when it has negative hyperbolic length, hyperbolas when it has positive hyperbolic length or parabolas when it has length zero (that is it belong to the $x^2 + y^2 - z^2 = 0$).

These distinctions reflects in the value of the geodesic curvature. Indeed, as the form of (1.84) also suggest, the value $c = \frac{1}{r}$ plays the role of threshold and we have the following situations:

(i) if $0 \leq |c| < 1/r$, then the curve is an hyperbola,

(ii) if $|c| = 1/r$, then the curve is a parabola,

(iii) if $|c| > 1/r$, then the curve is a circle.

This is not the only interesting feature of this classification. Indeed, following the description of Remark 1.62, curves of type (i) are equidistant curves, while curves of type (iii) are boundary of metric balls, (i.e., spheres) in the hyperbolic plane. Curves of type (ii) are also known as *horocycles*.

## 1.6 Bibliographical note

The material presented in this chapter is classical and covered by many textbook in basic differential geometry of curves and surfaces, as for instance in [dC76, Spi79, BG92, KÏ5]. Some results, such as those of Sections 1.1 and 1.2, are revisited in the spirit of geometric control theory, to serve as a model case study for the forthcoming chapters.

# Chapter 2

# Vector fields

In this chapter we collect some basic definitions of differential geometry, in order to recall some useful results and to fix the notation. We assume the reader to be familiar with the definitions of smooth manifold and smooth map between manifolds.

## 2.1 Differential equations on smooth manifolds

In what follows $I$ denotes an interval of $\mathbb{R}$ containing 0 in its interior.

### 2.1.1 Tangent vectors and vector fields

Let $M$ be a smooth $n$-dimensional manifold and $\gamma_1, \gamma_2 : I \to M$ two smooth curves based at $q = \gamma_1(0) = \gamma_2(0) \in M$. We say that $\gamma_1$ and $\gamma_2$ are *equivalent* if they have the same 1-st order Taylor polynomial in some (or, equivalently, in every) coordinate chart. This defines an equivalence relation on the space of smooth curves based at $q$.

**Definition 2.1.** Let $M$ be a smooth $n$-dimensional manifold and let $\gamma : I \to M$ be a smooth curve such that $\gamma(0) = q \in M$. Its *tangent vector* at $q = \gamma(0)$, denoted by

$$\left.\frac{d}{dt}\right|_{t=0} \gamma(t), \qquad \text{or} \qquad \dot{\gamma}(0), \tag{2.1}$$

is the equivalence class in the space of all smooth curves in $M$ such that $\gamma(0) = q$ (with respect to the equivalence relation defined above).

It is easy to check, using the chain rule, that this definition is well-posed (i.e., it does not depend on the representative curve).

**Definition 2.2.** Let $M$ be a smooth $n$-dimensional manifold. The *tangent space* to $M$ at a point $q \in M$ is the set

$$T_qM := \left\{ \left.\frac{d}{dt}\right|_{t=0} \gamma(t)\,,\ \gamma : I \to M \text{ smooth}, \gamma(0) = q \right\}.$$

It is a standard fact that $T_qM$ has a natural structure of $n$-dimensional vector space, where $n = \dim M$.

**Definition 2.3.** A smooth *vector field* on a smooth manifold $M$ is a smooth map

$$X : q \mapsto X(q) \in T_qM,$$

that associates to every point $q$ in $M$ a tangent vector at $q$. We denote by $\mathrm{Vec}(M)$ the set of smooth vector fields on $M$.

In coordinates we can write $X = \sum_{i=1}^n X^i(x)\frac{\partial}{\partial x_i}$, and the vector field is smooth if its components $X^i(x)$ are smooth functions. The value of a vector field $X$ at a point $q$ is denoted in what follows both with $X(q)$ and $X\big|_q$.

**Definition 2.4.** Let $M$ be a smooth manifold and $X \in \mathrm{Vec}(M)$. The equation

$$\dot{q} = X(q), \qquad q \in M, \tag{2.2}$$

is called an *ordinary differential equation* (or *ODE*) on $M$. A *solution* of (2.2) is a smooth curve $\gamma : J \to M$, where $J \subset \mathbb{R}$ is an open interval, such that

$$\dot{\gamma}(t) = X(\gamma(t)), \qquad \forall\, t \in J. \tag{2.3}$$

We also say that $\gamma$ is an *integral curve* of the vector field $X$.

A standard theorem on ODE ensures that, for every initial condition, there exists a unique integral curve of a smooth vector field, defined on some open interval.

**Theorem 2.5.** *Let $X \in \mathrm{Vec}(M)$ and consider the Cauchy problem*

$$\begin{cases} \dot{q}(t) = X(q(t)) \\ q(0) = q_0 \end{cases} \tag{2.4}$$

*For any point $q_0 \in M$ there exists $\delta > 0$ and a solution $\gamma : (-\delta, \delta) \to M$ of (2.4), denoted by $\gamma(t; q_0)$. Moreover the map $(t, q) \mapsto \gamma(t; q)$ is smooth on a neighborhood of $(0, q_0)$.*

The solution is unique in the following sense: if there exists two solutions $\gamma_1 : I_1 \to M$ and $\gamma_2 : I_2 \to M$ of (2.4) defined on two different intervals $I_1, I_2$ containing zero, then $\gamma_1(t) = \gamma_2(t)$ for every $t \in I_1 \cap I_2$. This permits to introduce the notion of *maximal solution* of (2.4), that is the unique solution of (2.4) that is not extendable to a larger interval $J$ containing $I$.

If the maximal solution of (2.4) is defined on a bounded interval $I = (a, b)$, then the solution leaves every compact $K$ of $M$ in a finite time $t_K < b$. We refer the reader, for instance, to [HS74] for a proof of classical results on ODE.

A vector field $X \in \mathrm{Vec}(M)$ is called *complete* if, for every $q_0 \in M$, the maximal solution $\gamma(t; q_0)$ of the equation (2.2) is defined on $I = \mathbb{R}$.

*Remark* 2.6. The classical theory of ODE ensures completeness of the vector field $X \in \mathrm{Vec}(M)$ in the following cases:

(i) $M$ is a compact manifold (or more generally $X$ has compact support in $M$),

(ii) $M = \mathbb{R}^n$ and $X$ has sub-linear growth at infinity, i.e., there exists $C_1, C_2 > 0$ such that

$$|X(x)| \leq C_1|x| + C_2, \qquad \forall\, x \in \mathbb{R}^n.$$

where $|\cdot|$ denotes the Euclidean norm in $\mathbb{R}^n$.

*Remark* 2.7. When we are interested in the behavior of the trajectories of a vector field $X \in \mathrm{Vec}(M)$ in a compact subset $K$ of $M$, the assumption of completeness is not restrictive.

Indeed consider an open neighborhood $O_K$ of a compact $K$ with compact closure $\overline{O}_K$ in $M$. There exists a smooth cut-off function $a : M \to \mathbb{R}$ that is identically 1 on $K$, and that vanishes out of $O_K$. Then the vector field $aX$ is complete, since it has compact support in $M$. Moreover, the vector fields $X$ and $aX$ coincide on $K$, hence their integral curves coincide on $K$ as well.

### 2.1.2 Flow of a vector field

Given a complete vector field $X \in \mathrm{Vec}(M)$ we can consider the family of maps

$$\phi_t : M \to M, \qquad \phi_t(q) = \gamma(t; q), \qquad t \in \mathbb{R}. \tag{2.5}$$

where $\gamma(t; q)$ is the integral curve of $X$ starting at $q$ when $t = 0$. By Theorem 2.5 it follows that the map

$$\phi : \mathbb{R} \times M \to M, \qquad \phi(t, q) = \phi_t(q),$$

is smooth in both variables and the family $\{\phi_t,\, t \in \mathbb{R}\}$ is a one parametric subgroup of $\mathrm{Diff}(M)$, namely, it satisfies the following identities:

$$\begin{aligned} &\phi_0 = \mathrm{Id}, \\ &\phi_t \circ \phi_s = \phi_s \circ \phi_t = \phi_{t+s}, \qquad &&\forall\, t, s \in \mathbb{R}, \\ &(\phi_t)^{-1} = \phi_{-t}, &&\forall\, t \in \mathbb{R}, \end{aligned} \tag{2.6}$$

Moreover, by construction, we have

$$\frac{\partial \phi_t(q)}{\partial t} = X(\phi_t(q)), \qquad \phi_0(q) = q, \quad \forall\, q \in M. \tag{2.7}$$

The family of maps $\phi_t$ defined by (2.5) is called the *flow* generated by $X$. For the flow $\phi_t$ of a vector field $X$ it is convenient to use the exponential notation $\phi_t := e^{tX}$, for every $t \in \mathbb{R}$. Using this notation, the group properties (2.6) take the form:

$$e^{0X} = \mathrm{Id}, \qquad e^{tX} \circ e^{sX} = e^{sX} \circ e^{tX} = e^{(t+s)X}, \qquad (e^{tX})^{-1} = e^{-tX}, \tag{2.8}$$

$$\frac{d}{dt} e^{tX}(q) = X(e^{tX}(q)), \qquad \forall\, q \in M. \tag{2.9}$$

*Remark* 2.8. When $X(x) = Ax$ is a linear vector field on $\mathbb{R}^n$, where $A$ is a $n \times n$ matrix, the corresponding flow $\phi_t$ is the matrix exponential $\phi_t(x) = e^{tA}x$.

### 2.1.3 Vector fields as operators on functions

A vector field $X \in \mathrm{Vec}(M)$ induces an action on the algebra $C^\infty(M)$ of the smooth functions on $M$, defined as follows:

$$X : C^\infty(M) \to C^\infty(M), \qquad a \mapsto Xa, \qquad a \in C^\infty(M), \tag{2.10}$$

where

$$(Xa)(q) = \left.\frac{d}{dt}\right|_{t=0} a(e^{tX}(q)), \qquad q \in M. \tag{2.11}$$

In other words $X$ differentiates the function $a$ along its integral curves.

*Remark* 2.9. Let us denote $a_t := a \circ e^{tX}$. The map $t \mapsto a_t$ is smooth and from (2.11) it immediately follows that $Xa$ represents the first order term in the expansion of $a_t$ when $t \to 0$: more precisely for $q \in M$ one has

$$a_t(q) = a(q) + t\,(Xa)(q) + O(t^2).$$

One can indeed show from the explicit expression that the remainder is uniform with respect to the point and write the identity

$$a_t = a + t\,Xa + O(t^2).$$

**Exercise 2.10.** Let $a \in C^\infty(M)$ and $X \in \mathrm{Vec}(M)$, and denote $a_t = a \circ e^{tX}$. Prove the following formulas

$$\frac{d}{dt} a_t = X a_t, \tag{2.12}$$

$$a_t = a + t\,Xa + \frac{t^2}{2!}X^2 a + \frac{t^3}{3!}X^3 a + \ldots + \frac{t^k}{k!}X^k a + O(t^{k+1}). \tag{2.13}$$

It is easy to see also that the following Leibniz rule is satisfied

$$X(ab) = (Xa)b + a(Xb), \qquad \forall\, a, b \in C^\infty(M), \tag{2.14}$$

This is equivalent to say that $X$, as an operator on smooth functions, is a *derivation* of the algebra of smooth functions $C^\infty(M)$.

*Remark* 2.11. Notice that, if $a \in C^\infty(M)$ and $X = \sum_{i=1}^n X_i(x)\frac{\partial}{\partial x_i}$ in some coordinate set, then $Xa = \sum_{i=1}^n X_i(x)\frac{\partial a}{\partial x_i}$. In particular, when $X$ is applied to the coordinate functions $a_i(x) = x_i$ for $i = 1, \ldots, n$, then $Xa_i = X_i$, which shows that a vector field is completely characterized by its action on functions.

**Exercise 2.12.** Let $a_1, \ldots, a_k \in C^\infty(M)$ and assume that $N = \{a_1 = \ldots = a_k = 0\} \subset M$ is a smooth submanifold. Show that $X \in \mathrm{Vec}(M)$ is tangent to $N$, i.e., $X(q) \in T_qN$ for all $q \in N$, if and only if $Xa_i(q) = 0$ for every $q \in N$ and $i = 1, \ldots, k$.

### 2.1.4 Nonautonomous vector fields

**Definition 2.13.** A *nonautonomous vector field* is family of vector fields $\{X_t\}_{t\in\mathbb{R}}$ such that the map $X(t,q) = X_t(q)$ satisfies the following properties

(C1) the map $t \mapsto X(t,q)$ is measurable, for every fixed $q \in M$,

(C2) the map $q \mapsto X(t,q)$ is smooth, for every fixed $t \in \mathbb{R}$,

(C3) for every system of coordinates defined in an open set $\Omega \subset M$ and every compact $K \subset \Omega$ and compact interval $I \subset \mathbb{R}$ there exists two functions $c(t), k(t)$ in $L^\infty(I)$ such that for all $(t,x), (t,y) \in I \times K$

$$\|X(t,x)\| \leq c(t), \qquad \|X(t,x) - X(t,y)\| \leq k(t)\|x - y\|.$$

Conditions (C1) and (C2) are equivalent to require that for every smooth function $a \in C^\infty(M)$ the scalar function $(t,q) \mapsto X_t a|_q$ defined on $\mathbb{R} \times M$ is measurable in $t$ and smooth in $q$.

*Remark* 2.14. In what follows we are mainly interested in nonautonomous vector fields of the following form

$$X_t(q) = \sum_{i=1}^{m} u_i(t) f_i(q), \tag{2.15}$$

where $u_i$ are $L^\infty$ functions and $f_i$ are smooth vector fields on $M$. For this class of nonautonomous vector fields, assumptions (C1)-(C2) are trivially satisfied. For what concerns (C3), thanks to the smoothness of $f_i$, for every compact set $K \subset \Omega$ we can find two positive constants $C_K, L_K$ such that for all $i = 1, \ldots, m$, and $j = 1, \ldots, n$, we have

$$\|f_i(x)\| \leq C_K, \qquad \left\| \frac{\partial f_i}{\partial x_j}(x) \right\| \leq L_K, \qquad \forall\, x \in K,$$

and one gets for all $(t, x), (t, y) \in I \times K$

$$\|X(t, x)\| \leq C_K \sum_{i=1}^{m} |u_i(t)|, \qquad \|X(t, x) - X(t, y)\| \leq L_K \sum_{i=1}^{m} |u_i(t)| \cdot \|x - y\|. \tag{2.16}$$

The existence and uniqueness of integral curves of a nonautonomous vector field is guaranteed by the following theorem (see [BP07]).

**Theorem 2.15** (Carathéodory theorem)**.** *Assume that the nonautonomous vector field* $\{X_t\}_{t\in\mathbb{R}}$ *satisfies (C1)-(C3). Then the Cauchy problem*

$$\begin{cases} \dot{q}(t) = X(t, q(t)) \\ q(t_0) = q_0 \end{cases} \tag{2.17}$$

*has a unique solution* $\gamma(t; t_0, q_0)$ *defined on an open interval* $I$ *containing* $t_0$ *such that* (2.17) *is satisfied for almost every* $t \in I$ *and* $\gamma(t_0; t_0, q_0) = q_0$*. Moreover the map* $(t, q_0) \mapsto \gamma(t; t_0, q_0)$ *is locally Lipschitz with respect to* $t$ *and smooth with respect to* $q_0$*.*

Let us assume now that the vector field $X_t$ is *complete*, i.e., for all $t_0 \in \mathbb{R}$ and $q_0 \in M$ the solution $\gamma(t; t_0, q_0)$ to (2.17) is defined on $I = \mathbb{R}$. Let us denote $P_{t_0,t}(q) = \gamma(t; t_0, q)$. The family of maps $\{P_{t,s}\}_{t,s\in\mathbb{R}}$ where $P_{t,s} : M \to M$ is the *(nonautonomous) flow* generated by $X_t$.

By definition, for every fixed $t_0 \in \mathbb{R}$, the nonautonomous flow $t \mapsto P_{t_0,t}$ associated to a nonautonomous vector field $X_t$ is locally Lipschitz and satisfies the equation for a.e. $t$

$$\frac{\partial}{\partial t} P_{t_0,t}(q) = X(t, P_{t_0,t}(q)), \qquad q \in M. \tag{2.18}$$

Moreover the following algebraic identities are satisfied by $\{P_{t,s}\}_{t,s\in\mathbb{R}}$

$$P_{t,t} = \mathrm{Id}, \tag{2.19}$$

$$P_{t_2,t_3} \circ P_{t_1,t_2} = P_{t_1,t_3}, \qquad \forall\, t_1, t_2, t_3 \in \mathbb{R}, \tag{2.20}$$

$$(P_{t_1,t_2})^{-1} = P_{t_2,t_1}, \qquad \forall\, t_1, t_2 \in \mathbb{R},$$

Conversely, with every family of smooth diffeomorphism $P_{t,s} : M \to M$ satisfying the relations (2.19)-(2.20), that is called a *flow* on $M$, one can associate its *infinitesimal generator* $X_t$ as follows:

$$X_t(q) = \left.\frac{d}{ds}\right|_{s=0} P_{t,t+s}(q), \qquad \forall\, q \in M. \tag{2.21}$$

Differentiating the equation (2.18) with respect to $q$ and assuming that one can exchange the time and spatial derivatives one obtains that the nonautonomous flow $t \mapsto P_{t_0,t}$ associated to a nonautonomous vector field $X_t$ satisfies for almost every $t$ the identity

$$\frac{\partial}{\partial t}\frac{\partial P_{t_0,t}}{\partial q}(q_0) = \frac{\partial X}{\partial q}(t, P_{t_0,t}(q_0))\frac{\partial P_{t_0,t}}{\partial q}(q_0). \tag{2.22}$$

The formal justification of the validity of (2.22) can be found in [BP07]. The following lemma characterizes flows whose infinitesimal generator is autonomous. Its proof is left as an exercise.

**Lemma 2.16.** *Let* $\{P_{t,s}\}_{t,s\in\mathbb{R}}$ *be a family of smooth diffeomorphisms satisfying* (2.19)-(2.20)*. Its infinitesimal generator is an autonomous vector field if and only if*

$$P_{0,t} \circ P_{0,s} = P_{0,t+s}, \qquad \forall\, t, s \in \mathbb{R}.$$

## 2.2 Differential of a smooth map

A smooth map between manifolds induces a map between the corresponding tangent spaces.

**Definition 2.17.** Let $\varphi : M \to N$ a smooth map between smooth manifolds and $q \in M$. The *differential* of $\varphi$ at the point $q$ is the linear map

$$\varphi_{*,q} : T_qM \to T_{\varphi(q)}N, \tag{2.23}$$

defined as follows:

$$\varphi_{*,q}(v) = \left.\frac{d}{dt}\right|_{t=0}\varphi(\gamma(t)), \qquad \text{if} \qquad v = \left.\frac{d}{dt}\right|_{t=0}\gamma(t), \quad q = \gamma(0).$$

It is easily checked that this definition depends only on the equivalence class of $\gamma$.

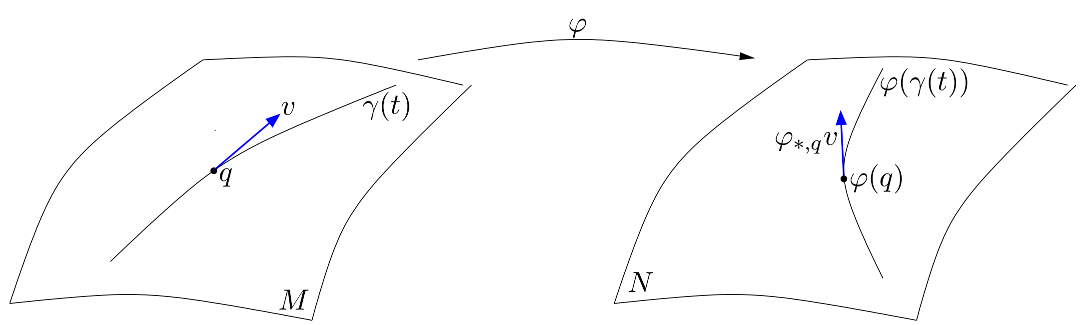



Figure 2.1: Differential of a map $\varphi : M \to N$

The differential $\varphi_{*,q}$ of a smooth map $\varphi : M \to N$, also called its *pushforward*, is sometimes denoted by the symbols $D_q\varphi$ or $d_q\varphi$ (see Figure 2.1).

**Exercise 2.18.** Let $\varphi : M \to N$, $\psi : N \to Q$ be smooth maps between manifolds. Prove that the differential of the composition $\psi \circ \varphi : M \to Q$ satisfies $(\psi \circ \varphi)_* = \psi_* \circ \varphi_*$.

As we said, a smooth map induces a transformation of tangent vectors. If we deal with diffeomorphisms, we can also pushforward a vector field.

**Definition 2.19.** Let $X \in \mathrm{Vec}(M)$ and $\varphi : M \to N$ be a diffeomorphism. The *pushforward* $\varphi_* X \in \mathrm{Vec}(N)$ is the vector field on $N$ defined by

$$(\varphi_* X)(\varphi(q)) := \varphi_*(X(q)), \qquad \forall\, q \in M. \tag{2.24}$$

When $P \in \mathrm{Diff}(M)$ is a diffeomorphism on $M$, we can rewrite the identity (2.24) as

$$(P_* X)(q) = P_*(X(P^{-1}(q))), \qquad \forall\, q \in M. \tag{2.25}$$

Notice that, in general, if $\varphi$ is a smooth map, the pushforward of a vector field is not well-defined.

*Remark* 2.20. From this definition it follows the useful formula for $X, Y \in \mathrm{Vec}(M)$

$$(e^{tX}_* Y)\big|_q = e^{tX}_* \big(Y\big|_{e^{-tX}(q)}\big) = \frac{d}{ds}\bigg|_{s=0} e^{tX} \circ e^{sY} \circ e^{-tX}(q).$$

If $P \in \mathrm{Diff}(M)$ and $X \in \mathrm{Vec}(M)$, then $P_* X$ is, by construction, the vector field whose integral curves are the image under $P$ of integral curves of $X$. The following lemma shows how it acts as operator on functions.

**Lemma 2.21.** *Let $P \in \mathrm{Diff}(M)$, $X \in \mathrm{Vec}(M)$ and $a \in C^\infty(M)$ then*

$$e^{tP_* X} = P \circ e^{tX} \circ P^{-1}, \tag{2.26}$$

$$(P_* X)a = (X(a \circ P)) \circ P^{-1}. \tag{2.27}$$

*Proof.* From the formula

$$\frac{d}{dt}\bigg|_{t=0} P \circ e^{tX} \circ P^{-1}(q) = P_*(X(P^{-1}(q))) = (P_* X)(q),$$

it follows that $t \mapsto P \circ e^{tX} \circ P^{-1}(q)$ is an integral curve of $P_* X$, from which (2.26) follows. To prove (2.27) let us compute

$$(P_* X)a\big|_q = \frac{d}{dt}\bigg|_{t=0} a(e^{tP_* X}(q)).$$

Using (2.26) this is equal to

$$\frac{d}{dt}\bigg|_{t=0} a(P(e^{tX}(P^{-1}(q))) = \frac{d}{dt}\bigg|_{t=0} (a \circ P)(e^{tX}(P^{-1}(q))) = (X(a \circ P)) \circ P^{-1}.$$

□

As a consequence of Lemma 2.21 one gets the following formula: for every $X, Y \in \mathrm{Vec}(M)$

$$(e^{tX}_* Y)a = Y(a \circ e^{tX}) \circ e^{-tX}. \tag{2.28}$$

## 2.3 Lie brackets

In this section we introduce a fundamental notion for sub-Riemannian geometry, the *Lie bracket* of two vector fields $X$ and $Y$. Geometrically it is defined as the infinitesimal version of the pushforward of the second vector field along the flow of the first one. As explained below, it measures how much $Y$ is modified by the flow of $X$.

**Definition 2.22.** Let $X, Y \in \mathrm{Vec}(M)$. We define their *Lie bracket* as the vector field

$$[X, Y] := \frac{\partial}{\partial t}\bigg|_{t=0} e_*^{-tX} Y. \tag{2.29}$$

*Remark* 2.23. The geometric meaning of the Lie bracket can be understood by writing explicitly

$$[X, Y]\big|_q = \frac{\partial}{\partial t}\bigg|_{t=0} e_*^{-tX} Y\big|_q = \frac{\partial}{\partial t}\bigg|_{t=0} e_*^{-tX}(Y\big|_{e^{tX}(q)}) = \frac{\partial}{\partial s \partial t}\bigg|_{t=s=0} e^{-tX} \circ e^{sY} \circ e^{tX}(q). \tag{2.30}$$

**Proposition 2.24.** *As derivations on functions, one has the identity*

$$[X, Y] = XY - YX. \tag{2.31}$$

*Proof.* By definition of Lie bracket we have $[X, Y]a = \frac{\partial}{\partial t}\big|_{t=0}(e_*^{-tX}Y)a$. Hence we have to compute the first order term in the expansion, with respect to $t$, of the map

$$t \mapsto (e_*^{-tX}Y)a.$$

Using formula (2.28) we have

$$(e_*^{-tX}Y)a = Y(a \circ e^{-tX}) \circ e^{tX}.$$

By Remark 2.9 we have $a \circ e^{-tX} = a - t\,Xa + O(t^2)$, hence

$$\begin{aligned}(e_*^{-tX}Y)a &= Y(a - t\,Xa + O(t^2)) \circ e^{tX} \\ &= (Ya - t\,YXa + O(t^2)) \circ e^{tX}.\end{aligned}$$

Denoting $b = Ya - t\,YXa + O(t^2)$, $b_t = b \circ e^{tX}$, and using again the expansion above we get

$$\begin{aligned}(e_*^{-tX}Y)a &= (Ya - t\,YXa + O(t^2)) + tX(Ya - t\,YXa + O(t^2)) + O(t^2) \\ &= Ya + t(XY - YX)a + O(t^2),\end{aligned}$$

that proves that the first order term with respect to $t$ in the expansion is $(XY - YX)a$. $\square$

Proposition 2.24 shows that $(\mathrm{Vec}(M), [\cdot,\cdot])$ is a Lie algebra.

**Exercise 2.25.** Prove the coordinate expression of the Lie bracket: let

$$X = \sum_{i=1}^n X_i \frac{\partial}{\partial x_i}, \qquad Y = \sum_{j=1}^n Y_j \frac{\partial}{\partial x_j},$$

be two smooth vector fields in $\mathbb{R}^n$. Show that

$$[X, Y] = \sum_{i,j=1}^n \left( X_i \frac{\partial Y_j}{\partial x_i} - Y_i \frac{\partial X_j}{\partial x_i} \right) \frac{\partial}{\partial x_j}.$$

Next we prove that every diffeomorphism induces a Lie algebra homomorphism on $\mathrm{Vec}(M)$.

**Proposition 2.26.** *Let $P \in \mathrm{Diff}(M)$. Then $P_*$ is a Lie algebra homomorphism of* $\mathrm{Vec}(M)$*, i.e.,*

$$P_*[X,Y] = [P_*X, P_*Y], \qquad \forall\, X,Y \in \mathrm{Vec}(M).$$

*Proof.* We show that the two terms are equal as derivations on functions. Let $a \in C^\infty(M)$, preliminarily we see, using (2.27), that

$$\begin{aligned} P_*X(P_*Ya) &= P_*X(Y(a\circ P)\circ P^{-1}) \\ &= X(Y(a\circ P)\circ P^{-1}\circ P)\circ P^{-1} \\ &= X(Y(a\circ P))\circ P^{-1}, \end{aligned}$$

and using twice this property and (2.31)

$$\begin{aligned} [P_*X, P_*Y]a &= P_*X(P_*Ya) - P_*Y(P_*Xa) \\ &= XY(a\circ P)\circ P^{-1} - YX(a\circ P)\circ P^{-1} \\ &= (XY - YX)(a\circ P)\circ P^{-1} \\ &= P_*[X,Y]a. \end{aligned}$$

□

To end this section, we show that the Lie bracket of two vector fields is zero (i.e., they commute as operator on functions) if and only if their flows commute.

**Proposition 2.27.** *Let $X, Y \in \mathrm{Vec}(M)$. The following properties are equivalent:*

*(i)* $[X,Y] = 0$,

*(ii)* $e^{tX}\circ e^{sY} = e^{sY}\circ e^{tX}, \quad \forall\, t,s \in \mathbb{R}$.

*Proof.* We start the proof with the following claim: $[X,Y]=0$ implies for every $t \in \mathbb{R}$

$$e_*^{-tX}Y = Y. \tag{2.32}$$

To prove (2.32) let us show that $[X,Y] = \frac{d}{dt}\big|_{t=0} e_*^{-tX}Y = 0$ implies that $\frac{d}{dt}e_*^{-tX}Y = 0$ for all $t\in\mathbb{R}$. Indeed we have

$$\begin{aligned} \frac{d}{dt}e_*^{-tX}Y &= \frac{d}{d\varepsilon}\bigg|_{\varepsilon=0} e_*^{-(t+\varepsilon)X}Y = \frac{d}{d\varepsilon}\bigg|_{\varepsilon=0} e_*^{-tX}e_*^{-\varepsilon X}Y \\ &= e_*^{-tX}\frac{d}{d\varepsilon}\bigg|_{\varepsilon=0} e_*^{-\varepsilon X}Y = e_*^{-tX}[X,Y] = 0, \end{aligned}$$

which proves (2.32) since the identity is true at $t=0$.

(i)⇒(ii). Fix $t\in\mathbb{R}$. Let us show that $\phi_s := e^{-tX}\circ e^{sY}\circ e^{tX}$ is the flow generated by $Y$. Indeed we have

$$\begin{aligned} \frac{\partial}{\partial s}\phi_s &= \frac{\partial}{\partial\varepsilon}\bigg|_{\varepsilon=0} e^{-tX}\circ e^{(s+\varepsilon)Y}\circ e^{tX} \\ &= \frac{\partial}{\partial\varepsilon}\bigg|_{\varepsilon=0} e^{-tX}\circ e^{\varepsilon Y}\circ e^{tX}\circ \underbrace{e^{-tX}\circ e^{sY}\circ e^{tX}}_{\phi_s} \\ &= e_*^{-tX}Y\circ\phi_s = Y\circ\phi_s, \end{aligned}$$

where in the last equality we used (2.32). Using uniqueness of the flow generated by a vector field we get

$$e^{-tX} \circ e^{sY} \circ e^{tX} = e^{sY}, \quad \forall\, t, s \in \mathbb{R},$$

which is equivalent to (ii).

(ii)⇒(i). For every function $a \in C^\infty$ we have

$$XYa = \frac{\partial^2}{\partial t \partial s}\Big|_{t=s=0} a \circ e^{sY} \circ e^{tX} = \frac{\partial^2}{\partial s \partial t}\Big|_{t=s=0} a \circ e^{tX} \circ e^{sY} = YXa.$$

Then (i) follows from (2.31). □

**Exercise 2.28.** Let $X, Y \in \mathrm{Vec}(M)$ and $q \in M$. Consider the curve on $M$

$$\gamma(t) = e^{-tY} \circ e^{-tX} \circ e^{tY} \circ e^{tX}(q).$$

Prove that the the curve $t \mapsto \gamma(\sqrt{t})$ is $C^1$ in a neighborhood of $t = 0$, and that its tangent vector at $t = 0$ is $[X, Y](q)$.

**Exercise 2.29.** Let $X, Y \in \mathrm{Vec}(M)$. Using the semigroup property of the flow, prove that

$$\frac{d}{dt} e_*^{-tX} Y = e_*^{-tX} [X, Y]. \tag{2.33}$$

Deduce the following formal series expansion (here $(\mathrm{ad}\, X)Y = [X, Y]$)

$$\begin{aligned} e_*^{-tX} Y &= \sum_{n=0}^{\infty} \frac{t^n}{n!} (\mathrm{ad}\, X)^n Y \qquad (2.34) \\ &= Y + t[X, Y] + \frac{t^2}{2}[X, [X, Y]] + \frac{t^3}{6}[X, [X, [X, Y]]] + \dots \end{aligned}$$

**Exercise 2.30.** Let $X, Y \in \mathrm{Vec}(M)$ and $a \in C^\infty(M)$. Prove the following Leibniz rule for the Lie bracket:

$$[X, aY] = a[X, Y] + (Xa)Y.$$

**Exercise 2.31.** Let $X, Y, Z \in \mathrm{Vec}(M)$. Prove that the Lie bracket satisfies the *Jacobi identity*:

$$[X, [Y, Z]] + [Y, [Z, X]] + [Z, [X, Y]] = 0. \tag{2.35}$$

Hint: Differentiate the identity $e_*^{tX}[Y, Z] = [e_*^{tX} Y, e_*^{tX} Z]$ with respect to $t$.

**Exercise 2.32.** Let $M$ be a smooth $n$-dimensional manifold and $X_1, \dots, X_n$ be linearly independent vector fields in a neighborhood of a point $q_0 \in M$. Prove that the map

$$\psi : \mathbb{R}^n \to M, \qquad \psi(t_1, \dots, t_n) = e^{t_1 X_1} \circ \dots \circ e^{t_n X_n}(q_0),$$

is a local diffeomorphism at 0. Moreover we have, denoting $t = (t_1, \dots, t_n)$,

$$\begin{aligned} \frac{\partial \psi}{\partial t_i}(t) &= e_*^{t_1 X_1} \circ \dots \circ e_*^{t_i X_i} \left( X_i \big|_{e^{t_{i+1} X_{i+1}} \circ \dots \circ e^{t_n X_n}(q_0)} \right) \qquad (2.36) \\ &= \left( e_*^{t_1 X_1} \cdots e_*^{t_i X_i} X_i \right) (\psi(t)). \qquad (2.37) \end{aligned}$$

Deduce that, when $[X_i, X_j] = 0$ for every $i, j = 1, \dots, n$, one has

$$\frac{\partial \psi}{\partial t_i}(t) = X_i(\psi(t)).$$

## 2.4 Frobenius theorem

In this section we prove Frobenius theorem about vector distributions.

**Definition 2.33.** Let $M$ be a smooth manifold. A *vector distribution* $D$ of rank $m$ on $M$ is a family of vector subspaces $D_q \subset T_qM$ where $\dim D_q = m$ for every $q$.

A vector distribution $D$ is said to be *smooth* if, for every point $q_0 \in M$, there exists a neighborhood $O_{q_0}$ of $q_0$ and a family of vector fields $X_1, \dots, X_m$ such that

$$D_q = \operatorname{span}\{X_1(q), \dots, X_m(q)\}, \qquad \forall\, q \in O_{q_0}. \tag{2.38}$$

**Definition 2.34.** A smooth vector distribution $D$ (of rank $m$) on $M$ is said to be *involutive* if there exists a local basis of vector fields $X_1, \dots, X_m$ satisfying (2.38) and smooth functions $a_{ij}^k$ on $M$ such that

$$[X_i, X_k] = \sum_{j=1}^{m} a_{ij}^k X_j, \qquad \forall\, i, k = 1, \dots, m. \tag{2.39}$$

**Exercise 2.35.** Prove that a smooth vector distribution $D$ is involutive if and only if for *every* local basis of vector fields $X_1, \dots, X_m$ satisfying (2.38) there exist smooth functions $a_{ij}^k$ such that (2.39) holds.

**Definition 2.36.** A smooth vector distribution $D$ on $M$ is said to be *flat* if for every point $q_0 \in M$ there exists a local diffeomorphism $\phi : O_{q_0} \to \mathbb{R}^n$ such that $\phi_{*,q}(D_q) = \mathbb{R}^m \times \{0\}$ for all $q \in O_{q_0}$.

**Theorem 2.37** (Frobenius Theorem)**.** *A smooth distribution is involutive if and only if it is flat.*

*Proof.* The statement is local, hence it is sufficient to prove the statement on a neighborhood of every point $q_0 \in M$.

(i). Assume first that the distribution is flat. Then there exists a diffeomorphism $\phi : O_{q_0} \to \mathbb{R}^n$ such that $D_q = \phi_{*,q}^{-1}(\mathbb{R}^m \times \{0\})$. It follows that for all $q \in O_{q_0}$ we have

$$D_q = \operatorname{span}\{X_1(q), \dots, X_m(q)\}, \qquad X_i(q) := \phi_{*,q}^{-1}\frac{\partial}{\partial x_i}.$$

and we have for $i, k = 1, \dots, m$,

$$[X_i, X_k] = \left[\phi_{*,q}^{-1}\frac{\partial}{\partial x_i}, \phi_{*,q}^{-1}\frac{\partial}{\partial x_k}\right] = \phi_{*,q}^{-1}\left[\frac{\partial}{\partial x_i}, \frac{\partial}{\partial x_i}\right] = 0.$$

(ii). Let us now prove that if $D$ is involutive then it is flat. As before, it is not restrictive to work on a neighborhood where

$$D_q = \operatorname{span}\{X_1(q), \dots, X_m(q)\}, \qquad \forall\, q \in O_{q_0}. \tag{2.40}$$

and (2.39) are satisfied. We first need a lemma.

**Lemma 2.38.** *For every $k = 1, \dots, m$, we have $e_*^{tX_k} D = D$.*

*Proof of Lemma 2.38.* Let us define the time dependent vector fields

$$Y_i^k(t) := e_*^{tX_k} X_i.$$

Using (2.39) and (2.33) we compute

$$\dot{Y}_i^k(t) = e_*^{tX_k}[X_i, X_k] = \sum_{j=1}^{m} e_*^{tX_k}\left(a_{ij}^k X_j\right) = \sum_{j=1}^{m} a_{ij}^k(t) Y_j^k(t),$$

where we set $a_{ij}^k(t) = a_{ij}^k \circ e^{-tX_k}$. Denote by $A^k(t) = (a_{ij}^k(t))_{i,j=1}^m$ and consider the unique solution $\Gamma^k(t) = (\gamma_{ij}^k(t))_{i,j=1}^m$ to the matrix Cauchy problem

$$\dot{\Gamma}^k(t) = A^k(t)\Gamma^k(t), \qquad \Gamma^k(0) = I. \tag{2.41}$$

Then we have

$$Y_i^k(t) = \sum_{j=1}^{m} \gamma_{ij}^k(t) Y_j^k(0),$$

that implies, for every $i, k = 1, \ldots, m$,

$$e_*^{tX_k} X_i = \sum_{j=1}^{m} \gamma_{ij}^k(t) X_j,$$

which proves the claim. □

We can now end the proof of Theorem 2.37. Complete the family $X_1, \ldots, X_m$ to a basis of the tangent space

$$T_q M = \operatorname{span}\{X_1(q), \ldots, X_m(q), Z_{m+1}(q), \ldots, Z_n(q)\},$$

in a neighborhood of $q_0$ and set $\psi : \mathbb{R}^n \to M$ defined by

$$\psi(t_1, \ldots, t_m, s_{m+1}, \ldots, s_n) = e^{t_1 X_1} \circ \ldots \circ e^{t_m X_m} \circ e^{s_{m+1} Z_{m+1}} \circ \ldots \circ e^{s_n Z_n}(q_0).$$

By construction $\psi$ is a local diffeomorphism at $(t, s) = (0, 0)$ and for $(t, s)$ close to $(0, 0)$ we have that (cf. Exercice 2.32)

$$\frac{\partial \psi}{\partial t_i}(t, s) = \left(e_*^{t_1 X_1} \ldots e_*^{t_i X_i} X_i\right)(\psi(t, s)),$$

for every $i = 1, \ldots, m$. These vectors are linearly independent and, thanks to Lemma 2.38, belong to $D$. Hence

$$D_q = \psi_* \operatorname{span}\left\{\frac{\partial}{\partial t_1}, \ldots, \frac{\partial}{\partial t_m}\right\}, \qquad q = \psi(t, s),$$

and the claim is proved. □

Reformulating Frobenius theorem in terms of submanifold, one immediately obtains the following corollary.

**Corollary 2.39.** *Let $D$ be an involutive distribution of rank $m$ on a smooth manifold $M$ of dimension $n \geq m$. Then, for every $q \in M$, there exists a (locally defined) submanifold $S$ of dimension $m$ passing through $q$ and that is tangent to $D$ at every point, i.e., $T_x S = D_x$ for every $x \in S$.*

### 2.4.1 An application of Frobenius theorem

Let $M$ and $N$ be two smooth manifolds. Given a vector field $X \in \mathrm{Vec}(M)$ and $Y \in \mathrm{Vec}(N)$ we define the vector field $X \times Y \in \mathrm{Vec}(M \times N)$ as the derivation

$$(X \times Y)a = Xa_y^1 + Ya_x^2,$$

where, given $a \in C^\infty(M \times N)$, we define $a_y^1 \in C^\infty(M)$ and $a_x^2 \in C^\infty(N)$ as follows

$$a_y^1(x) := a(x, y), \qquad a_x^2(y) := a(x, y), \qquad x \in M, y \in N.$$

Notice that, if we denote by $p_1 : M \times N \to M$ and $p_2 : M \times N \to N$ the two projections, we have

$$(p_1)_*(X \times Y) = X, \qquad (p_2)_*(X \times Y) = Y. \tag{2.42}$$

**Exercise 2.40.** Let $X_1, X_2 \in \mathrm{Vec}(M)$ and $Y_1, Y_2 \in \mathrm{Vec}(N)$. Prove that

$$[X_1 \times Y_1, X_2 \times Y_2] = [X_1, X_2] \times [Y_1, Y_2].$$

We can now prove the following result, which will be useful later for the classification of left-invariant structures on 3D Lie groups (cf. Section 17.5).

**Theorem 2.41.** *Let $M$ and $N$ be two $n$-dimensional smooth manifolds endowed with two local basis of vector fields $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_n$ around $x_0 \in M$ and $y_0 \in N$, respectively. Assume that there exists constants $c_{ij}^k \in \mathbb{R}$ such that*

$$[X_i, X_j] = \sum_{k=1}^n c_{ij}^k X_k, \qquad [Y_i, Y_j] = \sum_{k=1}^n c_{ij}^k Y_k.$$

*Then there exists a local diffeomorphism $\phi : O_{x_0} \to O_{y_0}$ such that $\phi_* X_i = Y_i$ for every $i = 1, \ldots, n$.*

*Proof.* Let us consider the family of vector fields $\{X_i \times Y_i\}_{i=1,\ldots,n}$ defined locally on $M \times N$. Thanks to Exercice 2.40 we have for $i, j = 1, \ldots, n$

$$[X_i \times Y_i, X_j \times Y_j] = [X_i, X_j] \times [Y_i, Y_j] = \sum_{k=1}^n c_{ij}^k X_k \times \sum_{k=1}^n c_{ij}^k Y_k = \sum_{k=1}^n c_{ij}^k (X_k \times Y_k).$$

It follows that the $n$-dimensional distribution $D = \mathrm{span}\{X_i \times Y_i\}_{i=1,\ldots,n}$ on the $2n$-dimensional smooth manifold $M \times N$ is involutive. Thanks to the Frobenius theorem (cf. Corollary 2.39), there exists a $n$-dimensional submanifold $S$ on $O_{x_0} \times O_{y_0}$ passing through $(x_0, y_0) \in M \times N$ and whose tangent space at every point coincides with $D$. It follows that $p_1|_S : S \to O_{x_0}$ and $p_2|_S : S \to O_{y_0}$ are two diffeomorphisms (thanks to (2.42) both $(p_i|_S)_*$ are isomorphisms) and $\phi := p_2 \circ (p_1|_S)^{-1} : O_{x_0} \to O_{y_0}$ is a diffeomorphism which satisfy $\phi(x_0) = y_0$ and $\phi_* X_i = Y_i$ for every $i = 1, \ldots, n$ by construction. $\square$

## 2.5 Cotangent space

In this section we introduce covectors, that are linear functionals on the tangent space. The space of all covectors at a point $q \in M$, called cotangent space is, in algebraic terms, simply the dual space to the tangent space.

**Definition 2.42.** Let $M$ be a $n$-dimensional smooth manifold. The *cotangent space* at a point $q \in M$ is the set

$$T_q^*M := (T_qM)^* = \{\lambda : T_qM \to \mathbb{R}, \lambda \text{ linear}\}.$$

If $\lambda \in T_q^*M$ and $v \in T_qM$, we will denote by $\langle \lambda, v \rangle := \lambda(v)$ the evaluation of the covector $\lambda$ on the vector $v$.

As we have seen, the differential of a smooth map yields a linear map between tangent spaces. The dual of the differential gives a linear map between cotangent spaces.

**Definition 2.43.** Let $\varphi : M \to N$ be a smooth map and $q \in M$. The *pullback* of $\varphi$ at point $\varphi(q)$, where $q \in M$, is the map

$$\varphi^* : T^*_{\varphi(q)}N \to T_q^*M, \qquad \lambda \mapsto \varphi^*\lambda,$$

defined by duality in the following way

$$\langle \varphi^*\lambda, v \rangle := \langle \lambda, \varphi_* v \rangle, \qquad \forall\, v \in T_qM,\ \forall\, \lambda \in T^*_{\varphi(q)}N.$$

**Example 2.44.** Let $a : M \to \mathbb{R}$ be a smooth function and $q \in M$. The differential $d_qa$ of the function $a$ at the point $q \in M$, defined through the formula

$$\langle d_qa, v \rangle := \frac{d}{dt}\bigg|_{t=0} a(\gamma(t)), \qquad v \in T_qM, \tag{2.43}$$

where $\gamma$ is any smooth curve such that $\gamma(0) = q$ and $\dot\gamma(0) = v$, is an element of $T_q^*M$. Indeed the right hand side of (2.43) is linear with respect to $v$.

**Definition 2.45.** A *differential 1-form* on a smooth manifold $M$ is a smooth map

$$\omega : q \mapsto \omega(q) \in T_q^*M,$$

that associates with every point $q$ in $M$ a cotangent vector at $q$. We denote by $\Lambda^1(M)$ the set of differential forms on $M$.

Since differential forms are dual objects to vector fields, it is well defined the action of $\omega \in \Lambda^1M$ on $X \in \mathrm{Vec}(M)$ pointwise, defining a function on $M$.

$$\langle \omega, X \rangle : q \mapsto \langle \omega(q), X(q) \rangle. \tag{2.44}$$

The differential form $\omega$ is *smooth* if and only if, for every smooth vector field $X \in \mathrm{Vec}(M)$, the function $\langle \omega, X \rangle \in C^\infty(M)$

**Definition 2.46.** Let $\varphi : M \to N$ be a smooth map and $a : N \to \mathbb{R}$ be a smooth function. The *pullback* $\varphi^*a$ is the smooth function on $M$ defined by

$$(\varphi^*a)(q) = a(\varphi(q)), \qquad q \in M.$$

In particular, if $\pi : T^*M \to M$ is the canonical projection and $a \in C^\infty(M)$, then

$$(\pi^*a)(\lambda) = a(\pi(\lambda)), \qquad \lambda \in T^*M,$$

which is constant on fibers.

## 2.6 Vector bundles

Heuristically, a smooth vector bundle on a smooth manifold $M$, is a smooth family of vector spaces parametrized by points in $M$.

**Definition 2.47.** Let $M$ be a $n$-dimensional manifold. A *smooth vector bundle of rank* $k$ over $M$ is a smooth manifold $E$ with a surjective smooth map $\pi : E \to M$ such that

(*i*) the set $E_q := \pi^{-1}(q)$, the *fiber* of $E$ at $q$, is a $k$-dimensional vector space,

(*ii*) for every $q \in M$ there exist a neighborhood $O_q$ of $q$ and a linear-on-fibers diffeomorphism (called *local trivialization*) $\psi : \pi^{-1}(O_q) \to O_q \times \mathbb{R}^k$ such that the following diagram commutes

$$\begin{tikzcd} \pi^{-1}(O_q) \arrow[r, "\psi"] \arrow[dr, "\pi"'] & O_q \times \mathbb{R}^k \arrow[d, "\pi_1"] \\ & O_q \end{tikzcd} \tag{2.45}$$

The space $E$ is called *total space* and $M$ is the *base* of the vector bundle. We will refer at $\pi$ as the *canonical projection* and rank $E$ will denote the rank of the bundle.

*Remark* 2.48. A vector bundle $E$, as a smooth manifold, has dimension

$$\dim E = \dim M + \text{rank } E = n + k.$$

In the case when there exists a global trivialization map (i.e., when one can choose a local trivialization with $O_q = M$ for all $q \in M$), then $E$ is diffeomorphic to $M \times \mathbb{R}^k$ and we say that $E$ is *trivializable*.

**Example 2.49.** For any smooth $n$-dimensional manifold $M$, the *tangent bundle* $TM$, defined as the disjoint union of the tangent spaces at all points of $M$,

$$TM = \bigcup_{q \in M} T_q M,$$

has a natural structure of $2n$-dimensional smooth manifold, equipped with the vector bundle structure (of rank $n$) induced by the canonical projection map

$$\pi : TM \to M, \qquad \pi(v) = q \quad \text{if} \quad v \in T_q M.$$

In the same way one can consider the *cotangent bundle* $T^*M$, defined as

$$T^*M = \bigcup_{q \in M} T_q^* M.$$

Again, it is a $2n$-dimensional manifold, and the canonical projection map

$$\pi : T^*M \to M, \qquad \pi(\lambda) = q \quad \text{if} \quad \lambda \in T_q^* M,$$

endows $T^*M$ with a structure of rank $n$ vector bundle.

Let $O \subset M$ be a coordinate neighborhood and denote by

$$\phi : O \to \mathbb{R}^n, \qquad \phi(q) = (x_1, \ldots, x_n),$$

a local coordinate system. The differentials of the coordinate functions

$$dx_i\big|_q, \qquad i = 1, \ldots, n, \qquad q \in O,$$

form a basis of the cotangent space $T_q^*M$. The dual basis in the tangent space $T_qM$ is defined by the vectors

$$\left.\frac{\partial}{\partial x_i}\right|_q \in T_qM, \qquad i = 1, \ldots, n, \qquad q \in O, \tag{2.46}$$

$$\left\langle dx_i, \frac{\partial}{\partial x_j} \right\rangle = \delta_{ij}, \qquad i, j = 1, \ldots, n. \tag{2.47}$$

Thus any tangent vector $v \in T_qM$ and any covector $\lambda \in T_q^*M$ can be decomposed in these basis

$$v = \sum_{i=1}^n v_i \left.\frac{\partial}{\partial x_i}\right|_q, \qquad\qquad \lambda = \sum_{i=1}^n p_i dx_i\big|_q,$$

and the maps

$$\psi : v \mapsto (x_1, \ldots, x_n, v_1, \ldots, v_n), \qquad \bar{\psi} : \lambda \mapsto (x_1, \ldots, x_n, p_1, \ldots, p_n), \tag{2.48}$$

define local coordinates on $TM$ and $T^*M$ respectively, which we call *canonical coordinates* induced by the coordinates $\psi$ on $M$.

**Definition 2.50.** A *morphism* $f : E \to E'$ between two vector bundles $E, E'$ on the base $M$ (also called a *bundle map*) is a smooth map such that the following diagram is commutative

$$\begin{tikzcd} E \arrow[r, "f"] \arrow[dr, "\pi"'] & E' \arrow[d, "\pi'"] \\ & M \end{tikzcd} \tag{2.49}$$

where $f$ is linear on fibers. Here $\pi$ and $\pi'$ denote the canonical projections.

**Definition 2.51.** Let $\pi : E \to M$ be a smooth vector bundle over $M$. A *local section* of $E$ is a smooth map[1] $\sigma : A \subset M \to E$ satisfying $\pi \circ \sigma = \mathrm{Id}_A$, where $A$ is an open set of $M$. In other words $\sigma(q)$ belongs to $E_q$ for each $q \in A$, smoothly with respect to $q$. If $\sigma$ is defined on all $M$ it is said to be a *global section*.

**Example 2.52.** Let $\pi : E \to M$ be a smooth vector bundle over $M$. The *zero section* of $E$ is the global section

$$\zeta : M \to E, \qquad \zeta(q) = 0 \in E_q, \qquad \forall\, q \in M.$$

We will denote by $M_0 := \zeta(M) \subset E$.

[1]here smooth means as a map between manifolds.

*Remark* 2.53. Notice that smooth vector fields and smooth differential forms are, by definition, sections of the vector bundles $TM$ and $T^*M$ respectively.

We end this section with some classical constructions on vector bundles.

**Definition 2.54.** Let $\varphi : M \to N$ be a smooth map between smooth manifolds and $E$ be a vector bundle on $N$, with fibers $\{E_{q'}, q' \in N\}$. The *induced bundle* (or *pullback bundle)* $\varphi^* E$ is a vector bundle on the base $M$ defined by

$$\varphi^* E := \{(q, v) \,|\, q \in M, v \in E_{\varphi(q)}\} \subset M \times E.$$

Notice that $\operatorname{rank} \varphi^* E = \operatorname{rank} E$, hence $\dim \varphi^* E = \dim M + \operatorname{rank} E$.

**Example 2.55.** (i). Let $M$ be a smooth manifold and $TM$ its tangent bundle, endowed with an Euclidean structure. The *spherical bundle* $SM$ is the vector subbundle of $TM$ defined as follows

$$SM = \bigcup_{q \in M} S_q M, \qquad S_q M = \{v \in T_q M \,|\, |v| = 1\}.$$

(ii). Let $E, E'$ be two vector bundles over a smooth manifold $M$. The *direct sum* $E \oplus E'$ is the vector bundle over $M$ defined by

$$(E \oplus E')_q := E_q \oplus E'_q.$$

## 2.7 Submersions and level sets of smooth maps

If $\varphi : M \to N$ is a smooth map, we define the *rank of* $\varphi$ at $q \in M$ to be the rank of the linear map $\varphi_{*,q} : T_q M \to T_{\varphi(q)} N$. It is of course just the rank of the matrix of partial derivatives of $\varphi$ in any coordinate chart, or the dimension of $\operatorname{im}(\varphi_{*,q}) \subset T_{\varphi(q)} N$. If $\varphi$ has the same rank $k$ at every point, we say $\varphi$ has *constant rank*, and write $\operatorname{rank} \varphi = k$.

An *immersion* is a smooth map $\varphi : M \to N$ with the property that $\varphi_*$ is injective at each point (or equivalently $\operatorname{rank} \varphi = \dim M$). Similarly, a *submersion* is a smooth map $\varphi : M \to N$ such that $\varphi_*$ is surjective at each point (equivalently, $\operatorname{rank} \varphi = \dim N$).

**Theorem 2.56** (Rank Theorem)**.** *Suppose $M$ and $N$ are smooth manifolds of dimensions $m$ and $n$, respectively, and $\varphi : M \to N$ is a smooth map with constant rank $k$ in a neighborhood of $q \in M$. Then there exist coordinates $(x_1, \ldots, x_m)$ centered at $q$ and $(y_1, \ldots, y_n)$ centered at $\varphi(q)$ in which $\varphi$ has the following coordinate representation:*

$$\varphi(x_1, \ldots, x_m) = (x_1, \ldots, x_k, 0, \ldots, 0). \tag{2.50}$$

*Remark* 2.57. The previous theorem can be rephrased in the following way. Let $\varphi : M \to N$ be a smooth map between two smooth manifolds. Then the following are equivalent:

(i) $\varphi$ has constant rank in a neighborhood of $q \in M$.

(ii) There exist coordinates near $q \in M$ and $\varphi(q) \in N$ in which the coordinate representation of $\varphi$ is linear.

In the case of a submersion, from Theorem 2.56 one can deduce the following result.

**Corollary 2.58.** *Assume $\varphi : M \to N$ is a smooth submersion at $q$. Then $\varphi$ admits a local right inverse at $\varphi(q)$. Moreover $\varphi$ is open at $q$. More precisely it exist $\varepsilon > 0$ and $C > 0$ such that*

$$B_{\varphi(q)}(C^{-1}r) \subset \varphi(B_q(r)), \qquad \forall\, r \in [0, \varepsilon), \tag{2.51}$$

*where the balls in* (2.51) *are considered with respect to some Euclidean norm in a coordinate chart.*

*Remark* 2.59. The constant $C$ appearing in (2.51) is related to the norm of the differential of the local right inverse, computed with respect to the chosen Euclidean norm in the coordinate chart. When $\varphi$ is a diffeomorphism, $C$ is a bound on the norm of the differential of the inverse of $\varphi$. This recover the classical quantitative statement of the inverse function theorem.

Using these results, one can give some general criteria for level sets of smooth maps (or smooth functions) to be submanifolds.

**Theorem 2.60** (Constant rank level set theorem)**.** *Let $M$ and $N$ be smooth manifolds, and let $\varphi : M \to N$ be a smooth map with constant rank $k$. Each level set $\varphi^{-1}(y)$, for $y \in N$ is a closed embedded submanifold of codimension $k$ in $M$.*

*Remark* 2.61. It is worth to specify the following two important sub-cases of Theorem 2.60:

(a) If $\varphi : M \to N$ is a submersion at every $q \in \varphi^{-1}(y)$ for some $y \in N$, then $\varphi^{-1}(y)$ is a closed embedded submanifold whose codimension is equal to the dimension of $N$.

(b) If $a : M \to \mathbb{R}$ is a smooth function such that $d_q a \neq 0$ for every $q \in a^{-1}(c)$, where $c \in \mathbb{R}$, then the level set $a^{-1}(c)$ is a smooth hypersurface of $M$

**Exercise 2.62.** Let $a : M \to \mathbb{R}$ be a smooth function. Assume that $c \in \mathbb{R}$ is a regular value of $a$, i.e., $d_q a \neq 0$ for every $q \in a^{-1}(c)$. Then $N_c = a^{-1}(c) = \{q \in M \,|\, a(q) = c\} \subset M$ is a smooth submanifold. Prove that for every $q \in N_c$

$$T_q N_c = \ker d_q a = \{v \in T_q M \mid \langle d_q a, v\rangle = 0\}.$$

## 2.8 Bibliographical note

The material presented in this chapter is classical and covered by many textbook in differential geometry, as for instance in [Boo86, Lee13, dC92, Spi79].

Theorem 2.15 is a well-known theorem in ODE. The statement presented here can be deduced from [BP07, Theorem 2.1.1, Exercice 2.4]. The functions $c(t), k(t)$ appearing in (C3) are assumed to be $L^\infty$, that is stronger than $L^1$ (on compact intervals). This stronger assumption imply that the solution is not only absolutely continuous with respect to $t$, but also locally Lipschitz (compare also with the discussion in Section 3.6).

# Chapter 3

# Sub-Riemannian structures

In this section we introduce the notion of sub-Riemannian structure: the definition given here is quite general, permitting to include all the classical notions appearing in the literature with different names such as constant-rank sub-Riemannian structure, rank-varying sub-Riemannian structure, almost-Riemannian structure etc. Riemannian manifolds appear as particular cases as well.

After having introduced the fundamental object of this book, the sub-Riemannian distance, we prove its finiteness and continuity, also known as Rashevskii-Chow theorem. Then we move to metric property of sub-Riemannian manifold as metric spaces, proving in particular existence of length-minimizers.

In the final part of the chapter we introduce Pontryagin extremals, which are curves in the cotangent space satisfying a first-order necessary condition for length-minimality.

## 3.1 Basic definitions

We start by introducing bracket-generating family of vector fields.

**Definition 3.1.** Let $M$ be a smooth manifold and let $\mathcal{F} \subset \mathrm{Vec}(M)$ be a family of smooth vector fields. The *Lie algebra generated* by $\mathcal{F}$ is the smallest sub-algebra of $\mathrm{Vec}(M)$ containing $\mathcal{F}$, namely

$$\mathrm{Lie}\,\mathcal{F} := \mathrm{span}\{[X_1, \ldots, [X_{j-1}, X_j]], X_i \in \mathcal{F}, j \in \mathbb{N}\}. \tag{3.1}$$

We will say that $\mathcal{F}$ is *bracket-generating* (or that satisfies the *Hörmander condition*) if

$$\mathrm{Lie}_q\mathcal{F} := \{X(q), X \in \mathrm{Lie}\,\mathcal{F}\} = T_qM, \quad \forall\, q \in M.$$

Moreover, for $s \in \mathbb{N}$, we define

$$\mathrm{Lie}^s\mathcal{F} := \mathrm{span}\{[X_1, \ldots, [X_{j-1}, X_j]], X_i \in \mathcal{F}, j \leq s\}. \tag{3.2}$$

We say that the family $\mathcal{F}$ has *step $s$ at $q$* if $s \in \mathbb{N}$ is the minimal integer satisfying

$$\mathrm{Lie}^s_q\mathcal{F} := \{X(q), X \in \mathrm{Lie}^s\mathcal{F}\} = T_qM,$$

Notice that, in general, the step may depend on the point on $M$ and $s = s(q)$ can be unbounded on $M$ even for bracket-generating families.

**Definition 3.2.** Let $M$ be a connected smooth manifold. A *sub-Riemannian structure* on $M$ is a pair $(\mathbf{U}, f)$ where:

(i) $\mathbf{U}$ is an Euclidean bundle with base $M$ and Euclidean fiber $U_q$, i.e., for every $q \in M$, $U_q$ is a vector space equipped with a scalar product $(\cdot\,|\,\cdot)_q$, smooth with respect to $q$. For $u \in U_q$ we denote the norm of $u$ as $|u|^2 = (u\,|\,u)_q$.

(ii) $f : \mathbf{U} \to TM$ is a smooth map that is a morphism of vector bundles and is fiber-wise *linear*. In particular the following diagram is commutative

$$\begin{tikzcd} \mathbf{U} \arrow[r, "f"] \arrow[dr, "\pi_{\mathbf{U}}"'] & TM \arrow[d, "\pi"] \\ & M \end{tikzcd} \tag{3.3}$$

where $\pi_{\mathbf{U}} : \mathbf{U} \to M$ and $\pi : TM \to M$ denote the canonical projections.

(iii) The *set of horizontal vector fields* $\mathcal{D} := \{f(\sigma)\,|\,\sigma : M \to \mathbf{U}$ smooth section$\}$, is a bracket-generating family of vector fields. We call *step of the sub-Riemannian structure at* $q$ the step of $\mathcal{D}$.

When the vector bundle $\mathbf{U}$ admits a global trivialization we say that $(\mathbf{U}, f)$ is a *free sub-Riemannian structure*.

A smooth manifold endowed with a sub-Riemannian structure (i.e., the triple $(M, \mathbf{U}, f)$) is called a *sub-Riemannian manifold*. When the map $f : \mathbf{U} \to TM$ is fiberwise surjective, $(M, \mathbf{U}, f)$ is called a *Riemannian manifold* (cf. Exercise 3.24).

**Definition 3.3.** Let $(M, \mathbf{U}, f)$ be a sub-Riemannian manifold. The *distribution* is the family of subspaces

$$\{\mathcal{D}_q\}_{q\in M}, \qquad \text{where} \qquad \mathcal{D}_q := f(U_q) \subset T_qM.$$

We call $m = \text{rank}(\mathbf{U})$ the *bundle rank* of the sub-Riemannian structure, and $r(q) := \dim \mathcal{D}_q$ the *rank* of the sub-Riemannian structure at $q \in M$. We say that the sub-Riemannian structure $(\mathbf{U}, f)$ on $M$ has *constant rank* if $r(q)$ is constant. Otherwise we say that the sub-Riemannian structure is *rank-varying*.

The set of horizontal vector fields $\mathcal{D} \subset \text{Vec}(M)$ has the structure of a finitely generated module over $C^\infty(M)$. The distribution at each point can be written in terms of horizontal vector fields as follows

$$\mathcal{D}_q = \{X(q)|\, X \in \mathcal{D}\}.$$

The rank of a sub-Riemannian structure $(M, \mathbf{U}, f)$ satisfies for every $q \in M$

$$r(q) \leq \min\{m, n\}, \qquad \text{where } m = \text{rank}\,\mathbf{U},\ n = \dim M. \tag{3.4}$$

In what follows we denote points in $\mathbf{U}$ as pairs $(q, u)$, where $q \in M$ is an element of the base and $u \in U_q$ is an element of the fiber. Following this notation we can write the value of $f$ at this point as

$$f(q, u) \qquad \text{or} \qquad f_u(q).$$

We prefer the second notation to stress that, for each $q \in M$, $f_u(q)$ is a vector in $T_qM$. .

**Definition 3.4.** A Lipschitz curve $\gamma : [0, T] \to M$ is said to be *admissible* (or *horizontal*) for a sub-Riemannian structure if there exists a measurable and essentially bounded function

$$u : t \in [0, T] \mapsto u(t) \in U_{\gamma(t)}, \tag{3.5}$$

called *control*, such that

$$\dot{\gamma}(t) = f(\gamma(t), u(t)), \qquad \text{for a.e. } t \in [0, T]. \tag{3.6}$$

In this case we say that $u(\cdot)$ is a *control corresponding* to $\gamma$. Notice that different controls may correspond to the same trajectory (see Figure 3.1).

Notice that a curve $\gamma : [0, T] \to M$ such that $\gamma(t) = \gamma(0)$ for every $t \in [0, T]$ is always admissible. In what follows we call such a curve *trivial trajectory*.

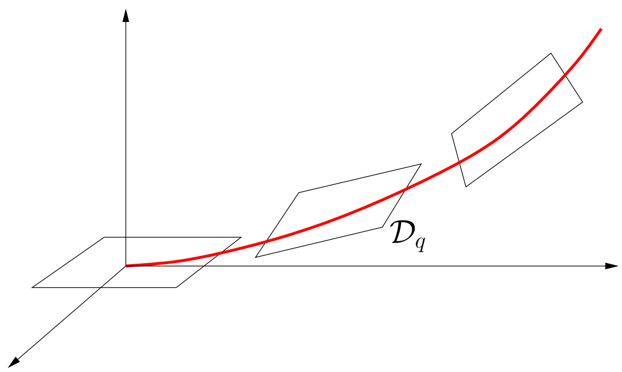



Figure 3.1: A horizontal curve

*Remark* 3.5. Once we have chosen a local trivialization $O_q \times \mathbb{R}^m$ for the vector bundle $\mathbf{U}$, where $O_q$ is a neighborhood of a point $q \in M$, we can choose a basis in the fibers and the map $f$ is written $f(q, u) = \sum_{i=1}^m u_i f_i(q)$, where $m$ is the rank of $\mathbf{U}$. In this trivialization, a Lipschitz curve $\gamma : [0, T] \to M$ is *admissible* if there exists $u = (u_1, \ldots, u_m) \in L^\infty([0, T], \mathbb{R}^m)$ such that

$$\dot{\gamma}(t) = \sum_{i=1}^m u_i(t) f_i(\gamma(t)), \qquad \text{for a.e. } t \in [0, T]. \tag{3.7}$$

Thanks to this local characterization and Theorem 2.15, for each initial condition $q \in M$ and $u \in L^\infty([0, T], \mathbb{R}^m)$ it follows that there exists an admissible curve $\gamma$, defined on a sufficiently small interval, such that $u$ is the control associated with $\gamma$ and $\gamma(0) = q$.

*Remark* 3.6. Notice that, for a curve to be admissible, it is not sufficient to satisfy $\dot{\gamma}(t) \in \mathcal{D}_{\gamma(t)}$ for almost every $t \in [0, T]$. Take for instance the two free sub-Riemannian structures on $\mathbb{R}^2$ having bundle rank two and defined by

$$f(x, y, u_1, u_2) = (x, y, u_1, u_2 x), \qquad f'(x, y, u_1, u_2) = (x, y, u_1, u_2 x^2). \tag{3.8}$$

and let $\mathcal{D}$ and $\mathcal{D}'$ the corresponding moduli of horizontal vector fields. It is easily seen that the curve $\gamma : [-1, 1] \to \mathbb{R}^2,\ \gamma(t) = (t, t^2)$ satisfies $\dot{\gamma}(t) \in \mathcal{D}_{\gamma(t)}$ and $\dot{\gamma}(t) \in \mathcal{D}'_{\gamma(t)}$ for every $t \in [-1, 1]$.

Moreover, $\gamma$ is admissible for $f$, since its corresponding control is $(u_1, u_2) = (1, 2)$ for a.e. $t \in [-1, 1]$, but it is not admissible for $f'$, since its corresponding control is uniquely determined as $(u_1(t), u_2(t)) = (1, 2/t)$ for a.e. $t \in [-1, 1]$, which is not essentially bounded (and even not integrable).

The example discussed in the previous remark shows that, for two different sub-Riemannian structures $(\mathbf{U}, f)$ and $(\mathbf{U}', f')$ on the same manifold $M$, one can have $\mathcal{D}_q = \mathcal{D}'_q$ for every $q \in M$, but $\mathcal{D} \neq \mathcal{D}'$.

**Exercise 3.7.** Prove that if the distribution has constant rank one has $\mathcal{D}_q = \mathcal{D}'_q$ for every $q \in M$ if and only if $\mathcal{D} = \mathcal{D}'$.

### 3.1.1 The minimal control and the length of an admissible curve

We start by defining the sub-Riemannian norm for vectors that belong to the distribution.

**Definition 3.8.** Let $v \in \mathcal{D}_q$. We define the *sub-Riemannian!norm* of $v$ as follows

$$\|v\| := \min\{|u|,\ u \in U_q \ \text{ s.t. } \ v = f(q, u)\}. \tag{3.9}$$

Notice that since $f$ is linear with respect to $u$, the minimum in (3.9) is always attained at a unique point. Indeed the condition $f(q, \cdot) = v$ defines an affine subspace of $U_q$ (which is nonempty since $v \in \mathcal{D}_q$) and the minimum in (3.9) is uniquely attained at the orthogonal projection of the origin onto this subspace (see Figure 3.2).

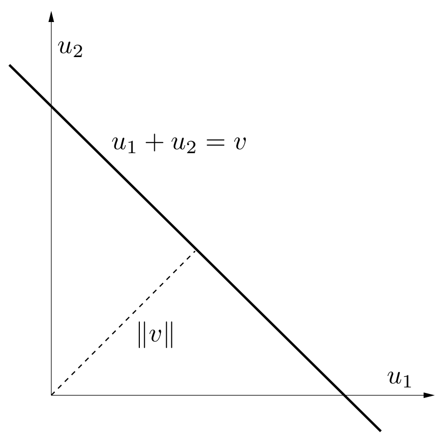



Figure 3.2: The norm of a vector $v$ for $f(x, u_1, u_2) = u_1 + u_2$

**Exercise 3.9.** Show that $\|\cdot\|$ is a norm in $\mathcal{D}_q$. Moreover prove that it satisfies the parallelogram law, i.e., it is induced by a scalar product $\langle \cdot \,|\, \cdot \rangle_q$ on $\mathcal{D}_q$, that can be recovered by the polarization identity

$$\langle v \,|\, w \rangle_q = \frac{1}{4}\|v + w\|^2 - \frac{1}{4}\|v - w\|^2, \qquad v, w \in \mathcal{D}_q. \tag{3.10}$$

**Exercise 3.10.** Let $u_1, \ldots, u_m \in U_q$ be an orthonormal basis for $U_q$. Define $v_i = f(q, u_i)$. Show that if $f(q, \cdot)$ is injective then $v_1, \ldots, v_m$ is an orthonormal basis for $\mathcal{D}_q$.

An admissible curve $\gamma : [0,T] \to M$ is Lipschitz, hence differentiable at almost every point. Hence it is well defined a.e. on $[0,T]$ the unique control $t \mapsto u^*(t)$ associated with $\gamma$ and realizing the minimum in (3.9).

**Definition 3.11.** Given an admissible curve $\gamma : [0,T] \to M$, we define at every differentiability point of $\gamma$

$$u^*(t) := \operatorname{argmin}\{|u|,\, u \in U_q \ \text{ s.t. } \ \dot\gamma(t) = f(\gamma(t), u)\}. \tag{3.11}$$

We say that the control $u^*$ is the *minimal control* associated with $\gamma$.

We stress that $u^*(t)$ is pointwise defined for a.e. $t \in [0,T]$. The proof of the following crucial Lemma is postponed to the Section 3.5.

**Lemma 3.12.** *Let $\gamma : [0,T] \to M$ be an admissible curve. Then its minimal control $u^*(\cdot)$ is measurable and essentially bounded on $[0,T]$.*

*Remark* 3.13. If the admissible curve $\gamma : [0,T] \to M$ is differentiable, its minimal control is defined everywhere on $[0,T]$. Nevertheless, it could be not continuous, in general. Indeed consider, as in Remark 3.6, the free sub-Riemannian structure on $\mathbb{R}^2$ given by

$$f(x, y, u_1, u_2) = (x, y, u_1, u_2 x), \tag{3.12}$$

and let $\gamma : [-1,1] \to \mathbb{R}^2$ defined by $\gamma(t) = (t, t^2)$. Its minimal control $u^*(t)$ satisfies $(u_1^*(t), u_2^*(t)) = (1,2)$ when $t \neq 0$, while $(u_1^*(0), u_2^*(0)) = (1,0)$, hence is not continuous.

Thanks to Lemma 3.12 we are allowed to introduce the following definition.

**Definition 3.14.** Let $\gamma : [0,T] \to M$ be an admissible curve. We define the *sub-Riemannian length* of $\gamma$ as

$$\ell(\gamma) := \int_0^T \|\dot\gamma(t)\| dt. \tag{3.13}$$

We say that $\gamma$ is *parametrized by arc length* (or *arc length parametrized* if $\|\dot\gamma(t)\| = 1$ for a.e. $t \in [0,T]$.

Formula (3.13) says that the length of an admissible curve is the integral of the norm of its minimal control.

$$\ell(\gamma) = \int_0^T |u^*(t)| dt. \tag{3.14}$$

In particular any admissible curve has finite length. For an arc length parametrized curve we have that $\ell(\gamma) = T$.

**Lemma 3.15.** *The length of an admissible curve is invariant by Lipschitz reparametrization.*

*Proof.* Let $\gamma : [0,T] \to M$ be an admissible curve and $\varphi : [0,T'] \to [0,T]$ a Lipschitz reparametrization, i.e., a Lipschitz and monotone surjective map. Consider the reparametrized curve

$$\gamma_\varphi : [0,T'] \to M, \qquad \gamma_\varphi := \gamma \circ \varphi.$$

First observe that $\gamma_\varphi$ is a composition of Lipschitz functions, hence Lipschitz. Moreover $\gamma_\varphi$ is admissible since, by the linearity of $f$, it has minimal control $(u^* \circ \varphi)\dot\varphi \in L^\infty$, where $u^*$ is the minimal control of $\gamma$. Using the change of variables $t = \varphi(s)$, one gets (for the validity of the chain rule one can see for instance [Rud87, Ch. 7])

$$\ell(\gamma_\varphi) = \int_0^{T'} \|\dot{\gamma}_\varphi(s)\| ds = \int_0^{T'} |u^*(\varphi(s))||\dot{\varphi}(s)| ds = \int_0^{T} |u^*(t)| dt = \int_0^{T} \|\dot{\gamma}(t)\| dt = \ell(\gamma). \quad (3.15)$$

$\square$

**Lemma 3.16.** *Every admissible curve of positive length is a Lipschitz reparametrization of an arc length parametrized admissible one.*

*Proof.* Let $\gamma : [0, T] \to M$ be an admissible curve with $\ell(\gamma) > 0$ and minimal control $u^*$. Consider the Lipschitz monotone function $\varphi : [0, T] \to [0, \ell(\gamma)]$ defined by

$$\varphi(t) := \int_0^t |u^*(\tau)| d\tau.$$

Notice that if $\varphi(t_1) = \varphi(t_2)$, the monotonicity of $\varphi$ ensures $\gamma(t_1) = \gamma(t_2)$. Hence we are allowed to define the curve $\zeta : [0, \ell(\gamma)] \to M$ by

$$\zeta(s) := \gamma(t), \qquad \text{if } s = \varphi(t) \text{ for some } t \in [0, T].$$

In other words, it holds $\gamma = \zeta \circ \varphi$. To show that $\zeta$ is Lipschitz let us first show that there exists a constant $C > 0$ such that, for every $t_0, t_1 \in [0, T]$ one has, in some local coordinates (where $|\cdot|$ denotes the Euclidean norm in coordinates)

$$|\gamma(t_1) - \gamma(t_0)| \leq C \int_{t_0}^{t_1} |u^*(\tau)| d\tau.$$

Indeed fix $K \subset M$ a compact set such that $\gamma([0, T]) \subset K$ and set $C := \max_{x \in K} \left( \sum_{i=1}^m |f_i(x)|^2 \right)^{1/2}$. Then

$$\begin{aligned}
|\gamma(t_1) - \gamma(t_0)| &\leq \int_{t_0}^{t_1} \sum_{i=1}^m |u_i^*(t) f_i(\gamma(t))| \, dt \\
&\leq \int_{t_0}^{t_1} \sqrt{\sum_{i=1}^m |u_i^*(t)|^2} \sqrt{\sum_{i=1}^m |f_i(\gamma(t))|^2 dt} \\
&\leq C \int_{t_0}^{t_1} |u^*(t)| dt,
\end{aligned}$$

Hence if $s_1 = \varphi(t_1)$ and $s_0 = \varphi(t_0)$ one has

$$|\zeta(s_1) - \zeta(s_0)| = |\gamma(t_1) - \gamma(t_0)| \leq C \int_{t_0}^{t_1} |u^*(\tau)| d\tau = C |s_1 - s_0|,$$

which proves that $\zeta$ is Lipschitz. It particular $\dot{\zeta}(s)$ exists for a.e. $s \in [0, \ell(\gamma)]$.

Next, we prove that $\zeta$ is admissible and its minimal control has norm one. Define for every $s$ such that $s = \varphi(t)$, $\dot{\varphi}(t)$ exists and $\dot{\varphi}(t) \neq 0$, the control

$$v(s) := \frac{u^*(t)}{\dot{\varphi}(t)} = \frac{u^*(t)}{|u^*(t)|}.$$

The control $v$ is defined for a.e. $s \in [0, \ell(\gamma)]$, thanks to Exercise 3.17. Moreover, by construction, $|v(s)| = 1$ for a.e. $s \in [0, \ell(\gamma)]$, and $v$ is the minimal control associated with $\zeta$. $\square$

**Exercise 3.17.** Let $\varphi : [0, T] \to \mathbb{R}$ be a Lipschitz and monotone function and define the set

$$C_\varphi = \{s \in \mathbb{R} \mid s = \varphi(t), \dot{\varphi}(t) \text{ exists}, \ \dot{\varphi}(t) = 0\}.$$

Prove that the Lebesgue measure of the set $C_\varphi$ is zero.

By the previous discussion, in what follows, it will be often convenient to assume that admissible curves are arc length parametrized (or parametrized such that $\|\dot{\gamma}(t)\|$ is constant).

### 3.1.2 Equivalence of sub-Riemannian structures

In this section we introduce the notion of equivalence for sub-Riemannian structures on the same base manifold $M$ and the notion of isometry between sub-Riemannian manifolds.

**Definition 3.18.** Let $(\mathbf{U}, f), (\mathbf{U}', f')$ be two sub-Riemannian structures on a smooth manifold $M$. They are said to be *equivalent as distributions* if

(i) there exist an Euclidean bundle $\mathbf{V}$ and two surjective vector bundle morphisms $p : \mathbf{V} \to \mathbf{U}$ and $p' : \mathbf{V} \to \mathbf{U}'$ such that the following diagram is commutative

$$\begin{tikzcd} & \mathbf{U} \arrow[dr, "f"] & \\ \mathbf{V} \arrow[ur, "p"] \arrow[dr, "p'"'] & & TM \\ & \mathbf{U}' \arrow[ur, "f'"'] & \end{tikzcd} \tag{3.16}$$

The structures $(\mathbf{U}, f), (\mathbf{U}', f')$ are said to be *equivalent as sub-Riemannian structures* (or simply *equivalent*) if (i) is satisfied and moreover:

(ii) the projections $p$, $p'$ are compatible with the scalar product, i.e., it holds

$$|u| = \min\{|v|, p(v) = u\}, \qquad \forall\, u \in \mathbf{U},$$
$$|u'| = \min\{|v|, p'(v) = u'\}, \qquad \forall\, u' \in \mathbf{U}'.$$

*Remark* 3.19. If $(\mathbf{U}, f)$ and $(\mathbf{U}', f')$ are equivalent as sub-Riemannian structures on $M$, then:

(a) the distributions $\mathcal{D}_q$ and $\mathcal{D}'_q$ defined by $f$ and $f'$ coincide, since $f(U_q) = f'(U'_q)$ for all $q \in M$.

(b) for each $w \in \mathcal{D}_q$ we have $\|w\| = \|w\|'$, where $\| \cdot \|$ and $\| \cdot \|'$ are the norms are induced by $(\mathbf{U}, f)$ and $(\mathbf{U}', f')$ respectively.

In particular the length of an admissible curve for two equivalent sub-Riemannian structures is the same.

**Exercise 3.20.** Prove that $(M, \mathbf{U}, f), (M, \mathbf{U}', f')$ are equivalent as distributions if and only if the two moduli of horizontal vector fields $\mathcal{D}$ and $\mathcal{D}'$ coincides.

**Definition 3.21.** Let $M$ be a sub-Riemannian manifold. We define the *minimal bundle rank* of $M$ as the infimum of rank of bundles that induce equivalent structures on $M$. Given $q \in M$ the *local minimal bundle rank* of $M$ at $q$ is the minimal bundle rank of the structure restricted on a sufficiently small neighborhood $O_q$ of $q$.

**Exercise 3.22.** Prove that the free sub-Riemannian structure on $\mathbb{R}^2$ defined by $f : \mathbb{R}^2 \times \mathbb{R}^3 \to T\mathbb{R}^2$ defined by

$$f(x, y, u_1, u_2, u_3) = (x, y, u_1, u_2 x + u_3 y)$$

has non-constant local minimal bundle rank.

For equivalence classes of sub-Riemannian structures we introduce the following definition.

**Definition 3.23.** Two equivalent classes of sub-Riemannian manifolds are said to be *isometric* if there exist two representatives $(M, \mathbf{U}, f)$ and $(M', \mathbf{U}', f')$, a diffeomorphism $\phi : M \to M'$ and an isomorphism of Euclidean bundles $\psi : \mathbf{U} \to \mathbf{U}'$ such that the following diagram is commutative

$$\begin{tikzcd} \mathbf{U} \arrow[r, "f"] \arrow[d, "\psi"'] & TM \arrow[d, "\phi_*"] \\ \mathbf{U}' \arrow[r, "f'"'] & TM' \end{tikzcd} \tag{3.17}$$

Here isomorphism of bundles is understood in the broad sense, i.e., $\psi$ is fiberwise linear but does not necessarily map a fiber into a fiber over the same point.

### 3.1.3 Examples

Our definition of sub-Riemannian manifold is quite general. In the following we list some classical geometric structures which are included in our setting.

1. **Riemannian structures**.
   Classically a Riemannian manifold is defined as a pair $(M, \langle \cdot | \cdot \rangle)$, where $M$ is a smooth manifold and $\langle \cdot | \cdot \rangle_q$ is a family of scalar product on $T_qM$, smoothly depending on $q \in M$. This definition is included in Definition 3.2 by taking $\mathbf{U} = TM$ endowed with the Euclidean structure induced by $\langle \cdot | \cdot \rangle$ and $f : TM \to TM$ the identity map.

   **Exercise 3.24.** Show that every Riemannian manifold in the sense of Definition 3.2 is indeed equivalent to a Riemannian structure in the classical sense above.

2. **Constant rank sub-Riemannian structures.**
   Classically a constant rank sub-Riemannian manifold is a triple $(M, D, \langle \cdot | \cdot \rangle)$, where $D$ is a vector subbundle of $TM$ and $\langle \cdot | \cdot \rangle_q$ is a family of scalar product on $D_q$, smoothly depending on $q \in M$. This definition is included in Definition 3.2 by taking $\mathbf{U} = D$, endowed with its Euclidean structure, and $f : D \hookrightarrow TM$ the canonical inclusion.

3. **Almost-Riemannian structures.**
   An *almost-Riemannian structure* on $M$ is a sub-Riemannian structure $(\mathbf{U}, f)$ on $M$ such that its local minimal bundle rank at every point is equal to the dimension of $M$.

4. **Free sub-Riemannian structures.**
   Let $\mathbf{U} = M \times \mathbb{R}^m$ be the trivial Euclidean bundle of rank $m$ on $M$. A point in $\mathbf{U}$ can be written as $(q, u)$, where $q \in M$ and $u = (u_1, \ldots, u_m) \in \mathbb{R}^m$.

   If we denote by $\{e_1, \ldots, e_m\}$ an orthonormal basis of $\mathbb{R}^m$, then we can define globally $m$ smooth vector fields on $M$ by $f_i(q) := f(q, e_i)$ for $i = 1, \ldots, m$. Then we have

$$f(q,u) = f\left(q, \sum_{i=1}^{m} u_i e_i\right) = \sum_{i=1}^{m} u_i f_i(q), \qquad q \in M. \tag{3.18}$$

   In this case, the problem of finding an admissible curve joining two fixed points $q_0, q_1 \in M$ and with minimal length is rewritten as the optimal control problem

$$\begin{cases} \dot\gamma(t) = \displaystyle\sum_{i=1}^{m} u_i(t) f_i(\gamma(t)) \\ \displaystyle\int_0^T |u(t)| dt \to \min \\ \gamma(0) = q_0, \quad \gamma(T) = q_1 \end{cases} \tag{3.19}$$

   For a free sub-Riemannian structure, the set of vector fields $f_1, \ldots, f_m$ build as above is called a *generating family*. Notice that, in general, a generating family is not orthonormal when $f$ is not injective.

5. **Surfaces in $\mathbb{R}^3$ as free sub-Riemannian structures**
   Due to topological constraints, in general it not possible to regard a surface of $\mathbb{R}^3$ (with the induced metric) as a free sub-Riemannian structure of bundle rank 2, i.e., defined by a pair of globally defined orthonormal vector fields. However, it is always possible to regard it as a free sub-Riemannian structure of bundle rank 3.

   Indeed, for an embedded surface $M$ in $\mathbb{R}^3$, consider the trivial Euclidean bundle $\mathbf{U} = M \times \mathbb{R}^3$, where points are denoted as usual $(q, u)$, with $u \in \mathbb{R}^3, q \in M$, and the map

$$f : \mathbf{U} \to TM, \qquad f(q,u) = \pi_q^\perp(u) \in T_qM. \tag{3.20}$$

   where $\pi_q^\perp : \mathbb{R}^3 \to T_qM$ is the orthogonal projection onto $T_qM \subset \mathbb{R}^3$.

   Notice that $f$ is a surjective bundle map and the set of vector fields $\{\pi_q^\perp(\partial_x), \pi_q^\perp(\partial_y), \pi_q^\perp(\partial_z)\}$ is a generating family for this structure.

**Exercise 3.25.** Show that $(\mathbf{U}, f)$ defined in (3.20) is equivalent to the Riemannian structure on $M$ induced by the embedding in $\mathbb{R}^3$.

### 3.1.4 Every sub-Riemannian structure is equivalent to a free one

The purpose of this section is to show that every sub-Riemannian structure $(\mathbf{U}, f)$ on $M$ is equivalent to a sub-Riemannian structure $(\mathbf{U}', f')$ where $\mathbf{U}'$ is a trivial bundle with sufficiently large bundle rank.

**Lemma 3.26.** *Let $M$ be a $n$-dimensional smooth manifold and $\pi : E \to M$ a smooth vector bundle of rank $m$. Then, there exists a vector bundle $\pi_0 : E_0 \to M$ with $\operatorname{rank} E_0 \leq 2n + m$ such that $E \oplus E_0$ is a trivial vector bundle.*

*Proof.* Remember that $E$, as a smooth manifold, has dimension

$$\dim\, E = \dim\, M + \operatorname{rank}\, E = n + m.$$

Consider the map $i : M \hookrightarrow E$ which embeds $M$ into the vector bundle $E$ as the zero section $M_0 = i(M)$. If we denote with $T_M E := i^*(TE)$ the pullback vector bundle, i.e., the restriction of $TE$ to the section $M_0$, we have the isomorphism (as vector bundles on $M$)

$$T_M E \simeq E \oplus TM. \tag{3.21}$$

Eq. (3.21) is a consequence of the fact that the tangent to every fibre $E_q$, being a vector space, is canonically isomorphic to its tangent space $T_q E_q$ so that

$$T_q E = T_q E_q \oplus T_q M \simeq E_q \oplus T_q M, \qquad \forall\, q \in M.$$

By Whitney theorem we have a (nonlinear on fibers, in general) immersion

$$\Psi : E \to \mathbb{R}^N, \qquad \Psi_* : T_M E \subset TE \hookrightarrow T\mathbb{R}^N,$$

for $N = 2(n+m)$, and $\Psi_*$ is injective as bundle map, i.e., $T_M E$ is a sub-bundle of $T\mathbb{R}^N \simeq \mathbb{R}^N \times \mathbb{R}^N$. Thus we can choose as a complement $E'$, the orthogonal bundle (on the base $M$) with respect to the Euclidean metric in $\mathbb{R}^N$, i.e.,

$$E' = \bigcup_{q \in M} E'_q, \qquad E'_q = (T_q E_q \oplus T_q M)^{\perp},$$

and considering $E_0 := T_M E \oplus E'$ we have that $E_0$ is trivial since its fibers are sum of orthogonal complements and by (3.21) the conclusion follows.

□

**Corollary 3.27.** *Every sub-Riemannian structure $(\mathbf{U}, f)$ on $M$ is equivalent to a sub-Riemannian structure $(\overline{\mathbf{U}}, \bar{f})$ where $\overline{\mathbf{U}}$ is a trivial bundle.*

*Proof.* By Lemma 3.26 there exists a vector bundle $\mathbf{U}'$ such that the direct sum $\overline{\mathbf{U}} := \mathbf{U} \oplus \mathbf{U}'$ is a trivial bundle. Endow $\mathbf{U}'$ with any metric structure $g'$. Define a metric on $\overline{\mathbf{U}}$ in such a way that $\bar{g}(u + u', v + v') = g(u, v) + g'(u', v')$ on each fiber $\bar{U}_q = U_q \oplus U'_q$. Notice that $U_q$ and $U'_q$ are orthogonal subspace of $\bar{U}_q$ with respect to $\bar{g}$.

Let us define the sub-Riemannian structure $(\overline{\mathbf{U}}, \bar{f})$ on $M$ by

$$\bar{f} : \overline{\mathbf{U}} \to TM, \qquad \bar{f} := f \circ p_1,$$

where $p_1 : \mathbf{U} \oplus \mathbf{U}' \to \mathbf{U}$ denotes the projection on the first factor. By construction, the diagram

$$\begin{tikzcd} & \overline{\mathbf{U}} \arrow[dr, "\bar{f}"] & \\ \mathbf{U} \oplus \mathbf{U}' \arrow[ur, "\mathrm{Id}"] \arrow[dr, "p_1"'] & & TM \\ & \mathbf{U} \arrow[ur, "f"'] & \end{tikzcd} \tag{3.22}$$

is commutative. Moreover condition (ii) of Definition 3.18 is satisfied since for every $\bar{u} = u + u'$, with $u \in U_q$ and $u' \in U_q'$, we have $|\bar{u}|^2 = |u|^2 + |u'|^2$, hence $|u| = \min\{|\bar{u}|, p_1(\bar{u}) = u\}$. □

Since every sub-Riemannian structure is equivalent to a free one, in what follows we can assume that there exists a global *generating family*, i.e., a family of $f_1, \ldots, f_m$ of vector fields globally defined on $M$ such that every admissible curve of the sub-Riemannian structure satisfies

$$\dot{\gamma}(t) = \sum_{i=1}^{m} u_i(t) f_i(\gamma(t)), \tag{3.23}$$

Moreover, by the classical Gram-Schmidt procedure, we can assume that $f_i$ are the image of an orthonormal frame defined on the fiber. (cf. Example 4 of Section 3.1.3)

Under these assumptions the length of an admissible curve $\gamma$ is given by

$$\ell(\gamma) = \int_0^T |u^*(t)| dt = \int_0^T \sqrt{\sum_{i=1}^{m} u_i^*(t)^2} dt,$$

where $u^*(t)$ is the minimal control associated with $\gamma$.

Notice that Corollary 3.27 implies that the modulus of horizontal vector fields $\mathcal{D}$ is globally generated by $f_1, \ldots, f_m$.

*Remark* 3.28. The curve $\gamma : [0, T] \to M$ defined by $\gamma(t) = e^{tf_i}$, integral curve of an element $f_i$ of a generating family $\mathcal{F} = \{f_1, \ldots, f_m\}$, is admissible and $\ell(\gamma) \le T$. If $\mathcal{F} = \{f_1, \ldots, f_m\}$ are linearly independent then they are an orthonormal frame and $\ell(\gamma) = T$.

**Exercise 3.29.** Consider a sub-Riemannian structure $(\mathbf{U}, f)$ over $M$. Let $m = \mathrm{rank}(\mathbf{U})$ and $h_{\max} = \max\{h(q) : q \in M\} \le m$ where $h(q)$ is the local minimal bundle rank at $q$. Prove that there exists a sub-Riemannian structure $(\overline{\mathbf{U}}, \overline{f})$ equivalent to $(\mathbf{U}, f)$ such that $\mathrm{rank}(\overline{\mathbf{U}}) = h_{\max}$.

## 3.2 Sub-Riemannian distance and Rashevskii-Chow theorem

In this section we introduce the sub-Riemannian distance and we prove the Rashevskii-Chow theorem.

Recall that, thanks to the results of Section 3.1.4, in what follows we can assume that the sub-Riemannian structure on $M$ is free, with generating family $\mathcal{F} = \{f_1, \ldots, f_m\}$. Notice that, by definition, of sub-Riemannian manifold, $M$ is assumed to be connected and $\mathcal{F}$ is assumed to be bracket-generating.

**Definition 3.30.** Let $M$ be a sub-Riemannian manifold and $q_0, q_1 \in M$. The *sub-Riemannian distance* (or *Carnot-Caratheodory distance*) between $q_0$ and $q_1$ is

$$d(q_0, q_1) = \inf\{\ell(\gamma) \mid \gamma : [0, T] \to M \text{ admissible, } \gamma(0) = q_0, \ \gamma(T) = q_1\}. \tag{3.24}$$

We now state the main result of this section.

**Theorem 3.31** (Rashevskii-Chow). *Let $M$ be a sub-Riemannian manifold. Then*

*(i)* $(M, d)$ *is a metric space,*

*(ii) the topology induced by* $(M, d)$ *is equivalent to the manifold topology.*

*In particular,* $d : M \times M \to \mathbb{R}$ *is continuous.*

One of the main consequences of this result is that, thanks to the bracket-generating condition, for every $q_0, q_1 \in M$, for every $q_0$ and $q_1$ in $M$ there always exists an admissible curve that joining them. Hence $d(q_0, q_1) < +\infty$.

In what follows $B(q, r)$ (sometimes denoted also $B_r(q)$) is the (open) sub-Riemannian ball of radius $r$ and center $q$

$$B(q, r) := \{q' \in M \mid d(q, q') < r\}.$$

The rest of this section is devoted to the proof of Theorem 3.31. To prove it, we have to show that $d$ is actually a distance, i.e.,

(a) $0 \leq d(q_0, q_1) < +\infty$ for all $q_0, q_1 \in M$,

(b) $d(q_0, q_1) = 0$ if and only if $q_0 = q_1$,

(c) $d(q_0, q_1) = d(q_1, q_0)$ and $d(q_0, q_2) \leq d(q_0, q_1) + d(q_1, q_2)$ for all $q_0, q_1, q_2 \in M$,

and the equivalence between the metric and the manifold topology: for every $q_0 \in M$ we have

(d) for every $\varepsilon > 0$ there exists a neighborhood $O_{q_0}$ of $q_0$ such that $O_{q_0} \subset B(q_0, \varepsilon)$,

(e) for every neighborhood $O_{q_0}$ of $q_0$ there exists $\delta > 0$ such that $B(q_0, \delta) \subset O_{q_0}$.

### 3.2.1 Proof of Rashevskii-Chow theorem

The symmetry of $d$ is a direct consequence of the fact that if $\gamma : [0, T] \to M$ is admissible, then the curve $\bar{\gamma} : [0, T] \to M$ defined by $\bar{\gamma}(t) = \gamma(T - t)$ is admissible and $\ell(\bar{\gamma}) = \ell(\gamma)$. The triangular inequality follows from the fact that, given two admissible curves $\gamma_1 : [0, T_1] \to M$ and $\gamma_2 : [0, T_2] \to M$ such that $\gamma_1(T_1) = \gamma_2(0)$, their concatenation

$$\gamma : [0, T_1 + T_2] \to M, \qquad \gamma(t) = \begin{cases} \gamma_1(t), & t \in [0, T_1], \\ \gamma_2(t - T_1), & t \in [T_1, T_1 + T_2]. \end{cases} \tag{3.25}$$

is still admissible. These two arguments prove item (c).

We divide the rest of the proof of the Theorem in the following steps.

S1. We prove that, for every $q_0 \in M$, there exists a neighborhood $O_{q_0}$ of $q_0$ such that $d(q_0, \cdot)$ is finite and continuous in $O_{q_0}$. This proves (d).

S2. We prove that $d$ is finite on $M \times M$. This proves (a).

S3. We prove (b) and (e).

To prove Step 1 we first need the following lemmas:

**Lemma 3.32.** *Let $N \subset M$ be a submanifold and $\mathcal{F} \subset \mathrm{Vec}(M)$ be a family of vector fields tangent to $N$, i.e., $X(q) \in T_qN$, for every $q \in N$ and $X \in \mathcal{F}$. Then for all $q \in N$ we have $\mathrm{Lie}_q\mathcal{F} \subset T_qN$. In particular $\dim \mathrm{Lie}_q\mathcal{F} \leq \dim N$.*

*Proof.* Let $X \in \mathcal{F}$. As a consequence of the local existence and uniqueness of the two Cauchy problems

$$\begin{cases} \dot{q} = X(q), & q \in M, \\ q(0) = q_0, & q_0 \in N. \end{cases} \qquad \text{and} \qquad \begin{cases} \dot{q} = X\big|_N(q), & q \in N, \\ q(0) = q_0, & q_0 \in N. \end{cases}$$

it follows that $e^{tX}(q) \in N$ for every $q \in N$ and $t$ small enough. This property, together with the definition of Lie bracket (see formula (2.30)) implies that, if $X, Y$ are tangent to $N$, the vector field $[X, Y]$ is tangent to $N$ as well. Iterating this argument we get that $\mathrm{Lie}_q\mathcal{F} \subset T_qN$ for every $q \in N$, from which the conclusion follows. □

**Lemma 3.33.** *Let $M$ be an $n$-dimensional sub-Riemannian manifold with generating family $\mathcal{F} = \{f_1, \ldots, f_m\}$. For every $q_0 \in M$ and every neighborhood $V$ of the origin in $\mathbb{R}^n$ there exist $\widehat{s} = (\widehat{s}_1, \ldots, \widehat{s}_n) \in V$, and a choice of $n$ vector fields $f_{i_1}, \ldots, f_{i_n} \in \mathcal{F}$, such that $\widehat{s}$ is a regular point of the map*

$$\psi : \mathbb{R}^n \to M, \qquad \psi(s_1, \ldots, s_n) = e^{s_n f_{i_n}} \circ \cdots \circ e^{s_1 f_{i_1}}(q_0).$$

*Remark* 3.34. Notice that, if $\mathcal{D}_{q_0} \neq T_{q_0}M$, then $\widehat{s} = 0$ cannot be a regular point of the map $\psi$. Indeed, for $s = 0$, the image of the differential of $\psi$ at 0 is $\mathrm{span}_{q_0}\{f_{i_j}, j = 1, \ldots, n\} \subset \mathcal{D}_{q_0}$ and the differential of $\psi$ cannot be surjective.

We stress that, in the choice of $f_{i_1}, \ldots, f_{i_n} \in \mathcal{F}$, a vector field can appear more than once, as for instance in the case $m < n$.

*Proof of Lemma 3.33.* We prove the lemma by steps.

1. There exists a vector field $f_{i_1} \in \mathcal{F}$ such that $f_{i_1}(q_0) \neq 0$, otherwise all vector fields in $\mathcal{F}$ vanish at $q_0$ and $\dim \mathrm{Lie}_{q_0}\mathcal{F} = 0$, which contradicts the bracket-generating condition. Then, for $|s|$ small enough, the map
$$\phi_1 : s_1 \mapsto e^{s_1 f_{i_1}}(q_0),$$
is a local diffeomorphism onto its image $\Sigma_1$. If $\dim M = 1$ the Lemma is proved.

2. Assume $\dim M \geq 2$. Then there exist $t_1^1 \in \mathbb{R}$, with $|t_1^1|$ small enough, and $f_{i_2} \in \mathcal{F}$ such that, if we denote by $q_1 = e^{t_1^1 f_{i_1}}(q_0)$, the vector $f_{i_2}(q_1)$ is not tangent to $\Sigma_1$. Otherwise, by Lemma 3.32, dim $\mathrm{Lie}_{q_0}\mathcal{F} = 1$, which contradicts the bracket-generating condition. Then the map
$$\phi_2 : (s_1, s_2) \mapsto e^{s_2 f_{i_2}} \circ e^{s_1 f_{i_1}}(q_0),$$
is a local diffeomorphism near $(t_1^1, 0)$ onto its image $\Sigma_2$. Indeed the vectors
$$\left.\frac{\partial \phi_2}{\partial s_1}\right|_{(t_1^1,0)} \in T_{q_1}\Sigma_1, \qquad \left.\frac{\partial \phi_2}{\partial s_2}\right|_{(t_1^1,0)} = f_{i_2}(q_1),$$
are linearly independent by construction. If $\dim M = 2$ the Lemma is proved.

3. Assume $\dim M \geq 3$. Then there exist $t_2^1, t_2^2$, with $|t_2^1 - t_1^1|$ and $|t_2^2|$ small enough, and $f_{i_3} \in \mathcal{F}$ such that, if $q_2 = e^{t_2^2 f_{i_2}} \circ e^{t_2^1 f_{i_1}}(q_0)$ we have that $f_{i_3}(q_2)$ is not tangent to $\Sigma_2$. Otherwise, by Lemma 3.32, $\dim \mathrm{Lie}_{q_1}\mathcal{D} = 2$, which contradicts the bracket-generating condition. Then the map

$$\phi_3 : (s_1, s_2, s_3) \mapsto e^{s_3 f_{i_3}} \circ e^{s_2 f_{i_2}} \circ e^{s_1 f_{i_1}}(q_0),$$

is a local diffeomorphism near $(t_2^1, t_2^2, 0)$. Indeed the vectors

$$\left.\frac{\partial \phi_3}{\partial s_1}\right|_{(t_2^1, t_2^2, 0)}, \left.\frac{\partial \phi_3}{\partial s_2}\right|_{(t_2^1, t_2^2, 0)} \in T_{q_2}\Sigma_2, \qquad \left.\frac{\partial \phi_3}{\partial s_3}\right|_{(t_2^1, t_2^2, 0)} = f_{i_3}(q_2),$$

are linearly independent since the last one is transversal to $T_{q_2}\Sigma_2$ by construction, while the first two are linearly independent since $\phi_3(s_1, s_2, 0) = \phi_2(s_1, s_2)$ and $\phi_2$ is a local diffeomorphisms at $(t_2^1, t_2^2)$ which is close to $(t_1^1, 0)$.

Repeating the same argument $n$ times (with $n = \dim M$), the lemma is proved. $\square$

*Proof of Step 1.* Thanks to Lemma 3.33 there exists a neighborhood $\widehat{V} \subset V$ of $\widehat{s}$ such that $\psi$ is a diffeomorphism from $\widehat{V}$ to $\psi(\widehat{V})$, see Figure 3.3. We stress that in general $q_0 = \psi(0)$ does not belong to $\psi(\widehat{V})$, cf. Remark 3.34.

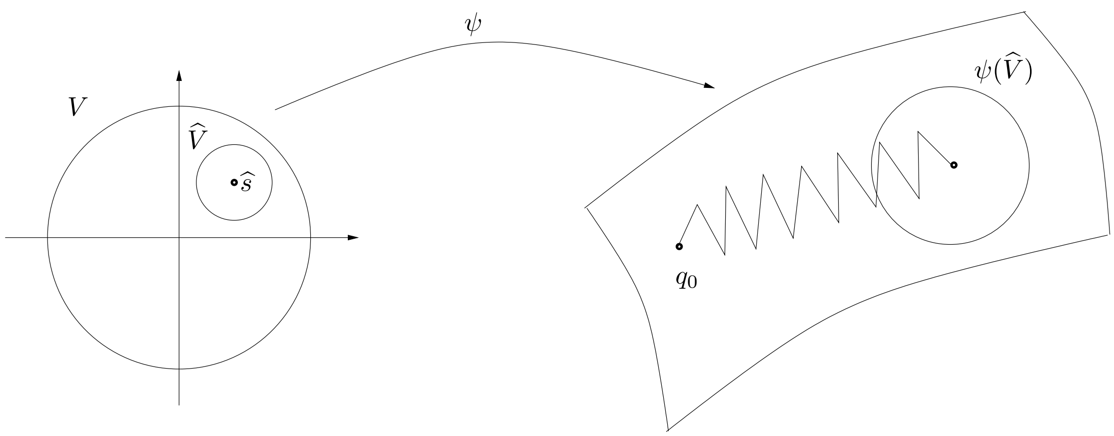



Figure 3.3: Proof of Lemma 3.33

To build a local diffeomorphism whose image contains $q_0$, we consider the map (here $\widehat{s} = (\widehat{s}_1, \ldots, \widehat{s}_n)$)

$$\widehat{\psi} : \mathbb{R}^n \to M, \qquad \widehat{\psi}(s_1, \ldots, s_n) = e^{-\widehat{s}_1 f_{i_1}} \circ \cdots \circ e^{-\widehat{s}_n f_{i_n}} \circ \psi(s_1, \ldots, s_n),$$

which has the following property: $\widehat{\psi}$ is a diffeomorphism from a neighborhood of $\widehat{s} \in V$, that we still denote $\widehat{V}$, to a neighborhood of $\widehat{\psi}(\widehat{s}) = q_0$.

Fix now $\varepsilon > 0$ and apply the construction above where $V$ is the neighborhood of the origin in $\mathbb{R}^n$ defined by $V = \{s \in \mathbb{R}^n \mid \sum_{i=1}^n |s_i| < \varepsilon\}$. Let us show that the claim of Step 1 holds with $O_{q_0} = \widehat{\psi}(\widehat{V})$. Indeed, for every $q \in \widehat{\psi}(\widehat{V})$, let $s = (s_1, \ldots, s_n)$ such that $q = \widehat{\psi}(s)$, and denote by $\gamma$ the admissible curve joining $q_0$ to $q$, built by $2n$-pieces, as in Figure 3.4.

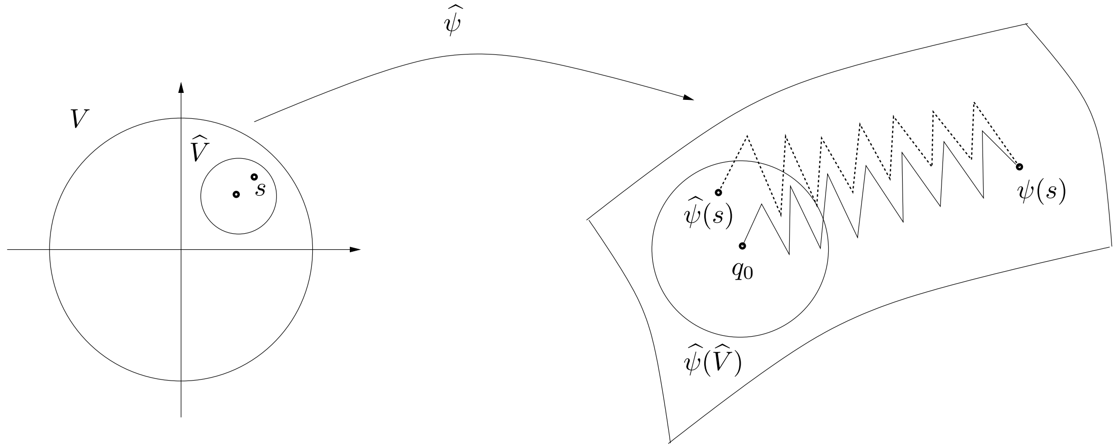



Figure 3.4: The map $\widehat{\psi}$

In other words $\gamma$ is the concatenation of integral curves of the vector fields $f_{i_j}$, i.e., admissible curves of the form $t \mapsto e^{tf_{i_j}}(q)$ defined on some interval $[0, T]$, whose length is less or equal than $T$ (cf. Remark 3.28). Since $s, \widehat{s} \in \widehat{V} \subset V$, it follows that:

$$d(q_0, q) \leq \ell(\gamma) \leq |s_1| + \ldots + |s_n| + |\widehat{s}_1| + \ldots + |\widehat{s}_n| < 2\varepsilon,$$

which ends the proof of Step 1, i.e., the finiteness and continuity of $d(q_0, \cdot)$ in $O_{q_0}$. $\square$

*Proof of Step 2.* To prove that $d$ is finite on $M \times M$ let us consider the equivalence classes of points in $M$ with respect to the relation

$$q_1 \sim q_2 \qquad \text{if} \qquad d(q_1, q_2) < +\infty. \tag{3.26}$$

From the triangular inequality and the proof of Step 1, it follows that each equivalence class is open. Moreover, by definition, the equivalence classes are disjoint and nonempty. Since $M$ is connected, it cannot be the union of open disjoint and nonempty subsets. It follows that there exists only one equivalence class. $\square$

*Remark* 3.35. Notice that from the triangular inequality one gets for $q_1, q_2, q_1', q_2'$ in $M$

$$|d(q_1', q_2') - d(q_1, q_2)| \leq d(q_1', q_1) + d(q_2', q_2),$$

hence the continuity of $d$ on $M \times M$ follows automatically.

**Lemma 3.36.** *Let $q_0 \in M$ and $K \subset M$ a compact set with $q_0 \in \operatorname{int} K$. Then there exists $\delta_K > 0$ such that every admissible curve $\gamma$ starting from $q_0$ and with $\ell(\gamma) \leq \delta_K$ is contained in $K$.*

*Proof.* Without loss of generality we can assume that $K$ is contained in a coordinate chart of $M$, where we denote by $|\cdot|$ the Euclidean norm in the coordinate chart. Let us define

$$C_K := \max_{x \in K} \left( \sum_{i=1}^{m} |f_i(x)|^2 \right)^{1/2}, \tag{3.27}$$

and fix $\delta_K > 0$ such that $\mathrm{dist}(q_0, \partial K) > C_K \delta_K$ (here dist denotes the Euclidean distance from a point to a set, in coordinates).

Let us show that for any admissible curve $\gamma : [0,T] \to M$ such that $\gamma(0) = q_0$ and $\ell(\gamma) \leq \delta_K$ we have $\gamma([0,T]) \subset K$. Indeed, if this is not true, there exists an admissible curve $\gamma : [0,T] \to M$ with $\ell(\gamma) \leq \delta_K$ and $t^* < T$, where $t^* := \sup\{t \in [0,T] \mid \gamma([0,t]) \subset K\}$. Then

$$
\begin{aligned}
|\gamma(t^*) - \gamma(0)| &\leq \int_0^{t^*} |\dot{\gamma}(t)| dt \leq \int_0^{t^*} \sum_{i=1}^m |u_i^*(t) f_i(\gamma(t))| \, dt && (3.28)\\
&\leq \int_0^{t^*} \sqrt{\sum_{i=0}^m |f_i(\gamma(t))|^2} \sqrt{\sum_{i=0}^m u_i^*(t)^2} \, dt && (3.29)\\
&\leq C_K \int_0^{t^*} \sqrt{\sum_{i=0}^m u_i^*(t)^2} \, dt \leq C_K \ell(\gamma) && (3.30)\\
&\leq C_K \delta_K < \mathrm{dist}(q_0, \partial K), && (3.31)
\end{aligned}
$$

which contradicts the fact that, at $t^*$, the curve $\gamma$ leaves the compact $K$. Thus $t^* = T$. □

*Proof of Step 3.* Let us prove that Lemma 3.36 implies property (b). Indeed the only nontrivial implication is that $d(q_0, q_1) > 0$ whenever $q_0 \neq q_1$. To prove this, fix a compact neighborhood $K$ of $q_0$ such that $q_1 \notin K$. By Lemma 3.36, each admissible curve joining $q_0$ and $q_1$ has length greater than $\delta_K$, hence $d(q_0, q_1) \geq \delta_K > 0$.

Let us now prove property (e). Fix $\varepsilon > 0$ and a compact neighborhood $K$ of $q_0$. Define $C_K$ and $\delta_K$ as in Lemma 3.36, and set $\delta := \min\{\delta_K, \varepsilon / C_K\}$. Let us show that $|q - q_0| < \varepsilon$ whenever $d(q_0, q) < \delta$, where again $|\cdot|$ is the Euclidean norm in a coordinate chart.

Consider a minimizing sequence $\gamma_n : [0,T] \to M$ of admissible trajectories joining $q_0$ and $q$ such that $\ell(\gamma_n) \to d(q_0, q)$ for $n \to \infty$. Without loss of generality, we can assume that $\ell(\gamma_n) \leq \delta$ for all $n$. By Lemma 3.36, $\gamma_n([0,T]) \subset K$ for all $n$.

We can repeat estimates (3.28)-(3.30) proving that $|q - q_0| = |\gamma_n(T) - \gamma_n(0)| \leq C_K \ell(\gamma_n)$ for all $n$. Passing to the limit for $n \to \infty$, one gets

$$
|q - q_0| \leq C_K d(q_0, q) \leq C_K \delta < \varepsilon. \tag{3.32}
$$

□

**Corollary 3.37.** *The metric space $(M, d)$ is locally compact, i.e., for any $q \in M$ there exists $\varepsilon > 0$ such that the closed sub-Riemannian ball $\overline{B}(q, r)$ is compact for all $0 \leq r \leq \varepsilon$.*

*Proof.* By the continuity of $d$, the set $\overline{B}(q, r) = \{d(q, \cdot) \leq r\}$ is closed for all $q \in M$ and $r \geq 0$. Moreover the sub-Riemannian metric $d$ induces the manifold topology on $M$. Hence, for radius small enough, the sub-Riemannian ball is bounded. Thus small sub-Riemannian balls are compact. □

### 3.2.2 Non bracket-generating structures

Sometimes can be useful to consider structures that satisfy only property (i) and (ii) of Definition 3.2, but that are not bracket-generating.

The typical example is the following: assume that the family of horizontal vector fields $\mathcal{D}$ satisfies

(i) $[\mathcal{D}, \mathcal{D}] \subset \mathcal{D}$,

(ii) $\dim \mathcal{D}_q$ does not depend on $q \in M$.

In this case the manifold $M$ is foliated by integral manifolds of the distribution (cf. Frobenius theorem), and each of them is endowed with a Riemannian structure. In this case the quantity

$$d(q_0, q_1) = \inf\{\ell(\gamma) \mid \gamma : [0, T] \to M \text{ admissible}, \ \gamma(0) = q_0, \ \gamma(T) = q_1\}. \tag{3.33}$$

does not define a metric, since (3.33) is infinite when $q_0$ and $q_1$ do not belong to the same leaf of the foliation.

On the other hand, observe that the bracket-generating condition is only a sufficient condition for the formula (3.33) defining a distance on the manifold $M$.

**Exercise 3.38.** Let $\phi : \mathbb{R} \to \mathbb{R}$ be the $C^\infty$ function defined by

$$\phi(x) = \begin{cases} 0, & x \leq 0 \\ e^{-1/x}, & x > 0 \end{cases}$$

Prove that the family of horizontal vector fields defined by the free sub-Riemannian structure

$$f : \mathbb{R}^2 \times \mathbb{R}^2 \to T\mathbb{R}^2, \qquad f(x, y, u_1, u_2) = (x, y, u_1, u_2\phi(x)).$$

is not bracket-generating. Prove that (3.33) still defines a distance in this case, but that $d$ is not continuous with respect to the Euclidean topology.

## 3.3 Existence of length-minimizers

In this section we want to discuss the existence of length-minimizers.

**Definition 3.39.** Let $\gamma : [0, T] \to M$ be an admissible curve. We say that $\gamma$ is a *length-minimizer* if it minimizes the length among admissible curves with same endpoints, i.e., $\ell(\gamma) = d(\gamma(0), \gamma(T))$.

*Remark* 3.40. Notice that the existence length-minimizers between two points is not guaranteed in general, as it happens for two points in $M = \mathbb{R}^2 \setminus \{0\}$ (endowed with the Euclidean distance) that are symmetric with respect to the origin. On the other hand, when length-minimizers exist between two fixed points, they may not be unique, as it happens for two antipodal points on the sphere $S^2$.

We now show a general semicontinuity property of the length functional.

**Theorem 3.41.** *Let $\gamma_n : [0, T] \to M$ be a sequence of admissible curves parametrized by arc length such that $\gamma_n \to \gamma$ uniformly on $[0, T]$ with $\liminf_{n\to\infty} \ell(\gamma_n) < +\infty$. Then $\gamma$ is admissible and*

$$\ell(\gamma) \leq \liminf_{n\to\infty} \ell(\gamma_n). \tag{3.34}$$

*Proof.* Let $L := \liminf_{n\to\infty} \ell(\gamma_n) < +\infty$ and choose a subsequence, still denoted $(\gamma_n)_{n\in\mathbb{N}}$, such that $\ell(\gamma_n) \to L$.

Fix $\delta > 0$. It is not restrictive to assume that, for $n$ large enough, $\ell(\gamma_n) \leq L + \delta$ and, by uniform convergence, that the image of $\gamma_n$ are all contained in a common compact set $K$.

Up to a common time rescaling, we can assume that all the curves are parametrized with constant speed on the interval $[0,1]$. Under this assumption, we have that $\dot{\gamma}_n(t) \in V_{\gamma_n(t)}$ for a.e. $t$, where

$$V_q = \{f_u(q), |u| \leq L + \delta\} \subset T_q M, \qquad f_u(q) = \sum_{i=1}^{m} u_i f_i(q).$$

Notice that $V_q$ is convex for every $q \in M$, thanks to the linearity of $f$ in $u$. Let us prove that $\gamma$ is admissible and satisfies $\ell(\gamma) \leq L + \delta$. Once this is done, since $\delta$ is arbitrary, this implies $\ell(\gamma) \leq L$, that is (3.34).

Writing in local coordinates, we have for every $\varepsilon > 0$

$$\frac{1}{\varepsilon}(\gamma_n(t+\varepsilon) - \gamma_n(t)) = \frac{1}{\varepsilon}\int_t^{t+\varepsilon} f_{u_n(\tau)}(\gamma_n(\tau))d\tau \in \operatorname{conv}\{V_{\gamma_n(\tau)} \mid \tau \in [t, t+\varepsilon]\}, \tag{3.35}$$

where $\operatorname{conv} S$ denotes the covex hull of a set $S$. Next we want to estimate the right hand side of (3.35) uniformly with respect to $n$. For $n \geq n_0$ sufficiently large, we have $|\gamma_n(t) - \gamma(t)| < \varepsilon$ (by uniform convergence) and an estimate similar to (3.30) gives for $\tau \in [t, t+\varepsilon]$

$$|\gamma_n(t) - \gamma_n(\tau)| \leq \int_t^{\tau} |\dot{\gamma}_n(s)| ds \leq C_K(L+\delta)|t-\tau| \leq C_K(L+\delta)\varepsilon, \tag{3.36}$$

where $C_K$ is the constant (3.27) defined by the compact $K$. Hence we deduce for every $\tau \in [t, t+\varepsilon]$ and every $n \geq n_0$

$$|\gamma_n(\tau) - \gamma(t)| \leq |\gamma_n(t) - \gamma_n(\tau)| + |\gamma_n(t) - \gamma(t)| \leq C'\varepsilon, \tag{3.37}$$

where $C'$ is independent on $n$ and $\varepsilon$. From the estimate (3.37) and the equivalence of the manifold and metric topology we have that, for all $\tau \in [t, t+\varepsilon]$ and $n \geq n_0$, $\gamma_n(\tau) \in B_{\gamma(t)}(r_\varepsilon)$, with $r_\varepsilon \to 0$ when $\varepsilon \to 0$. In particular

$$\operatorname{conv}\{V_{\gamma_n(\tau)} \mid \tau \in [t, t+\varepsilon]\} \subset \operatorname{conv}\{V_q \mid q \in B_{\gamma(t)}(r_\varepsilon)\}. \tag{3.38}$$

Using the inclusion (3.38) in (3.35) and passing to the limit for $n \to \infty$ we get finally to

$$\frac{1}{\varepsilon}(\gamma(t+\varepsilon) - \gamma(t)) \in \operatorname{conv}\{V_q,\, q \in B_{\gamma(t)}(r_\varepsilon)\}. \tag{3.39}$$

Passing to the limit for $n \to \infty$ in (3.36) one has that $\gamma$ is Lipschitz. Then for a.e. $t \in [0,1]$ the limit of the left hand side in (3.39) for $\varepsilon \to 0$ exists and gives $\dot{\gamma}(t) \in \operatorname{conv} V_{\gamma(t)} = V_{\gamma(t)}$. We can thus define the unique $u^*(t)$ satisfying $\dot{\gamma}(t) = f(\gamma(t), u^*(t))$ and $|u^*(t)| = \|\dot{\gamma}(t)\|$ for a.e. $t \in [0,1]$. Using the argument contained in Appendix 3.5 it follows that $u^*(t)$ is measurable on $[0,1]$. Moreover $|u^*(t)|$ is essentially bounded since, by construction, $|u^*(t)| \leq L + \delta$ for a.e. $t \in [0,1]$. Hence $\gamma$ is admissible and $\ell(\gamma) \leq L + \delta$ since $\gamma$ is defined on the interval $[0,1]$, which completes the proof. □

**Corollary 3.42.** *Let $\gamma_n : [0,T] \to M$ be a sequence of length-minimizers parametrized by arc length on $M$ such that $\gamma_n \to \gamma$ uniformly on $[0,T]$. Then $\gamma$ is a length-minimizer.*

*Proof.* Since $\gamma_n$ is a length-minimizer one has $\ell(\gamma_n) = d(\gamma_n(0), \gamma_n(T))$. By uniform convergence $\gamma_n(t) \to \gamma(t)$ for every $t \in [0,T]$ and, by continuity of the distance and semicontinuity of the length

$$\ell(\gamma) \leq \liminf_{n\to\infty} \ell(\gamma_n) = \liminf_{n\to\infty} d(\gamma_n(0), \gamma_n(T)) = d(\gamma(0), \gamma(T)),$$

that implies that $\ell(\gamma) = d(\gamma(0), \gamma(T))$, i.e., $\gamma$ is a length-minimizer. □

The semicontinuity of the length implies the existence of minimizers, under a natural compactness assumption on the space.

**Theorem 3.43** (Existence of minimizers)**.** *Let $M$ be a sub-Riemannian manifold and $q_0 \in M$. Assume that the ball $\overline{B}_{q_0}(r)$ is compact, for some $r > 0$. Then for all $q_1 \in B_{q_0}(r)$ there exists a length minimizer joining $q_0$ and $q_1$, i.e., we have*

$$d(q_0, q_1) = \min\{\ell(\gamma) \mid \gamma : [0, T] \to M \text{ admissible}\,, \gamma(0) = q_0, \gamma(T) = q_1\}.$$

*Proof.* Fix $q_1 \in B_{q_0}(r)$ and consider a minimizing sequence $\gamma_n : [0, 1] \to M$ of admissible trajectories, parametrized with constant speed, joining $q_0$ and $q_1$ and such that $\ell(\gamma_n) \to d(q_0, q_1)$.

Since $d(q_0, q_1) < r$, we have $\ell(\gamma_n) \leq r$ for all $n \geq n_0$ large enough, hence we can assume without loss of generality that the image of $\gamma_n$ is contained in the common compact $K = \overline{B}_{q_0}(r)$ for all $n$. In particular, the same argument leading to (3.36) shows that for all $n \geq n_0$

$$|\gamma_n(t) - \gamma_n(\tau)| \leq \int_\tau^t |\dot{\gamma}_n(s)|ds \leq C_K r|t - \tau|, \qquad \forall\, t, \tau \in [0, 1]. \tag{3.40}$$

where $C_K$ depends only on $K$. In other words, all trajectories in the sequence $\{\gamma_n\}_{n\in\mathbb{N}}$ are Lipschitz with the same Lipschitz constant. Thus the sequence is equicontinuous and uniformly bounded.

By the classical Ascoli-Arzelà Theorem there exist a subsequence of $\gamma_n$, which we still denote by the same symbol, and a Lipschitz curve $\gamma : [0, T] \to M$ such that $\gamma_n \to \gamma$ uniformly. By Theorem 3.41, the curve $\gamma$ satisfies $\ell(\gamma) \leq \liminf \ell(\gamma_n) = d(q_0, q_1)$, that implies $\ell(\gamma) = d(q_0, q_1)$. ☐

*Remark* 3.44. Assume that $\overline{B}(q, r_0)$ is compact for some $r_0 > 0$. Then for every $0 < r \leq r_0$ we have that $\overline{B}(q, r)$ is compact also, being a closed subset of a compact set $\overline{B}(q, r_0)$.

Combining Theorem 3.43 and Corollary 3.37 one gets the following corollary.

**Corollary 3.45.** *Let $q_0 \in M$. There exists $\varepsilon > 0$ such that for every $q_1 \in B_{q_0}(\varepsilon)$ there exists a minimizing curve joining $q_0$ and $q_1$.*

### 3.3.1 On the completeness of the sub-Riemannian distance

We provide here a characterization of metric completeness of a sub-Riemannian manifold. We start by proving a preliminary lemma. We recall that $B(x, r)$ (resp. $\overline{B}(x, r)$) denotes the open (resp. closed) ball of center $x \in M$ and radius $r > 0$ with respect to the sub-Riemannian distance.

**Lemma 3.46.** *Let $M$ be a sub-Riemannian manifold. For every $\varepsilon > 0$ and $x \in M$ we have*

$$B(x, r + \varepsilon) = \bigcup_{y\in B(x,r)} B(y, \varepsilon). \tag{3.41}$$

*Proof.* The inclusion $\supseteq$ is a direct consequence of the triangle inequality.

Let us prove the inclusion $\subseteq$. Fix $z \in B(x, r + \varepsilon) \setminus B(x, \varepsilon)$. Then there exists a length-parameterized curve $\gamma$ connecting $x$ with $z$ such that $\ell(\gamma) = t + \varepsilon$ where $0 \leq t < r$. Let $t' \in (t, r)$; then $\gamma(t') \in B(x, r)$ and $z \in B(\gamma(t'), \varepsilon)$. ☐

**Proposition 3.47.** *Let $M$ be a sub-Riemannian manifold. Then the three following properties are equivalent:*

*(i)* $(M, d)$ *is complete,*

*(ii)* $\overline{B}(x, r)$ *is compact for every* $x \in M$ *and* $r > 0$*,*

*(iii) there exists* $\varepsilon > 0$ *such that* $\overline{B}(x, \varepsilon)$ *is compact for every* $x \in M$*.*

*Proof.* (iii) implies (i). Let us prove that every Cauchy sequence $\{x_n\}_{n\in\mathbb{N}}$ in $M$ is convergent. Fix $\varepsilon > 0$ satisfying the assumption. Since $\{x_n\}_{n\in\mathbb{N}}$ is Cauchy there exists $N \in \mathbb{N}$ such that one has $d(x_n, x_m) < \varepsilon$ for all $n, m \geq N$.

In particular, by choosing $m = N$, for all $n \geq N$ one has that $x_n \in \overline{B}(x_N, \varepsilon)$, that is compact by assumption. Hence $\{x_n\}_{n\geq N}$ is a Cauchy sequence and admits a convergent subsequence, that implies that the whole sequence $\{x_n\}_{n\in\mathbb{N}}$ in $M$ is convergent.

(ii) implies (iii). This is evident.

(i) implies (ii). Assume now that $(M, d)$ is complete. Fix $x \in M$ and define

$$A := \{r > 0 \,|\, \overline{B}(x, r) \text{ is compact}\}, \qquad R := \sup A. \tag{3.42}$$

Since the topology of $(M, d)$ is locally compact then $A \neq \emptyset$ and $R > 0$. First we prove that $A$ is open and then we prove that $R = +\infty$. Notice in particular that this proves that $A = ]0, +\infty[$ since, by Remark 3.44, $r \in A$ implies $]0, r[ \subset A$.

(ii.a) It is enough to show that, if $r \in A$, then there exists $\delta > 0$ such that $r + \delta \in A$. For each $y \in B(x, r)$ there exists $\rho(y) < \varepsilon$ small enough such that $\overline{B}(y, \rho(y))$ is compact. We have

$$\overline{B}(x, r) \subset \bigcup_{y \in \overline{B}(x,r)} \overline{B}(y, \rho(y)).$$

By compactness of $\overline{B}(x, r)$ there exists a finite number of points $\{y_i\}_{i=1}^N$ in $\overline{B}(x, r)$ such that (denote $\rho_i := \rho(y_i)$)

$$\overline{B}(x, r) \subset \bigcup_{i=1}^N \overline{B}(y_i, \rho_i).$$

Moreover, since $\overline{B}(x, r+\delta)$ coincides with the set of points $\{y \in M \,|\, \mathrm{dist}(y, B(x, r)) \leq \delta\}$ by Lemma 3.46, there exists $\delta > 0$ such that

$$\overline{B}(x, r + \delta) \subset \bigcup_{i=1}^N \overline{B}(y_i, \rho_i).$$

This proves that $r + \delta \in A$, since a finite union of compact sets is compact.

(ii.b) Assume by contradiction that $R < +\infty$ and let us prove that $B := \overline{B}(x, R)$ is compact. Since $B$ is a closed set, it is enough to show that it is totally bounded, i.e., it admits a finite $\varepsilon$-net[1] for every $\varepsilon > 0$. Fix $\varepsilon > 0$ and consider an $(\varepsilon/3)$-net $S$ for the ball $B' = B(x, R - \varepsilon/3)$, that exists by compactness. By Lemma 3.46 one has for every $y \in B$ that $\mathrm{dist}(y, B') < \varepsilon/3$. Then it follows that

$$\mathrm{dist}(y, S) < \mathrm{dist}(y, B') + \varepsilon/3 < \varepsilon,$$

that is $S$ is an $\varepsilon$-net for $B$ and $B$ is compact.

This shows that if $R < +\infty$, then $R \in A$. Hence (ii.a) implies that $R + \delta \in A$ for some $\delta > 0$, contradicting the fact that $R$ is a sup. Hence $R = +\infty$. □

[1] A finite $\varepsilon$-net $S$ for a set $B$ in a metric space is a finite set of points $S = \{z_i\}_{i=1}^N$ such that for every $y \in B$ one has $\mathrm{dist}(y, S) < \varepsilon$ (or, equivalently, for every $y \in B$ there exists $i$ such that $d(y, z_i) < \varepsilon$).

*Remark* 3.48. Notice that we used that the distance is sub-Riemannian only to prove Lemma 3.46, which enters in the "(i) implies (ii)" part of the statement. Actually the same result holds true in the more general context of length metric space, see [BBI01, Ch. 2].

For the relation with geodesic completeness of the sub-Riemannian manifold, see Section 11.4.

Combining this result with Corollary 3.43, we obtain the following corollary.

**Corollary 3.49.** *Let $(M, d)$ be a complete sub-Riemannian manifold. Then for every $q_0, q_1 \in M$ there exists a length minimizer joining $q_0$ and $q_1$.*

### 3.3.2 Lipschitz curves with respect to $d$ vs admissible curves

The goal of this section is to prove that continous curves that are Lipschitz with respect to sub-Riemannian distance are exactly admissible curves.

**Proposition 3.50.** *Let $\gamma : [0, T] \to M$ be a continuous curve. Then $\gamma$ is Lipschitz with respect to the sub-Riemannian distance if and only if $\gamma$ is admissible.*

*Proof.* (i). Assume $\gamma$ is admissible and let $u$ be a control associated with $\gamma$. By definition $u$ is essentially bounded. Then for $t, s \in [0, T]$ with $t < s$ one has

$$d(\gamma(t), \gamma(s)) \leq \ell(\gamma|_{[t,s]}) \leq \int_s^t |u(\tau)| d\tau \leq C|t - s|,$$

for some constant $C > 0$. Then $\gamma$ is Lipschitz with respect to the sub-Riemannian distance.

(ii). Conversely assume that $\gamma$ is Lipschitz with respect to the sub-Riemannian distance, with Lipschitz constant $L > 0$, meaning that

$$d(\gamma(t), \gamma(s)) \leq L|t - s|, \qquad \forall\, t, s \in [0, T]. \tag{3.43}$$

Repeating arguments contained in the proof of Lemma 3.36 we have that for a compact neighborhood $K \subset M$ of $\gamma([0, T])$ there exists $C_K > 0$ such that

$$|\gamma(t) - \gamma(s)| \leq C_K d(\gamma(t), \gamma(s)), \tag{3.44}$$

for every $t, s$ close enough, where $|\cdot|$ denotes the Euclidean norm in coordinates. Combining (3.43) and (3.44) it follows that $\gamma$ is Lipschitz in charts and $\gamma$ is differentiable almost everywhere by Rademacher theorem.

Let us prove that $\gamma$ is admissile. Consider the partition $\sigma_n = \{t_{i,n}\}_{i=1}^{2^n}$ of the interval $[0, T]$ into $2^n$ intervals of length $T/2^n$, namely $t_{i,n} := iT/2^n$ for $i = 1, \ldots, 2^n$. By compactness of small balls and compactness of $[0, T]$, for $n$ large enough, there exists a length-minimizer joining $\gamma(t_{i,n})$ and $\gamma(t_{i+1,n})$ for $i = 1, \ldots, 2^n - 1$.

Denote by $\gamma_n$ the curve defined by the concatenation of length-minimizers joining $\gamma(t_{i,n})$ and $\gamma(t_{i+1,n})$ for $i = 1, \ldots, 2^n - 1$. Thanks to (3.43) we have the uniform bound on the length

$$\ell(\gamma_n) = \sum_{i=1}^{2^n} d(\gamma(t_{i,n}), \gamma(t_{i+1,n})) \leq \sum_{i=1}^{2^n} L|t_{i,n} - t_{i+1,n}| \leq \sum_{i=1}^{2^n} \frac{L}{2^n} T \leq LT. \tag{3.45}$$

Moreover, by construction, $\gamma_n$ converge uniformly to $\gamma$ when $n \to \infty$. By Theorem 3.41, the curve $\gamma$ is admissible and $\ell(\gamma) \leq L$. $\square$

**Exercise 3.51.** Let $\gamma : [0,T] \to M$ be an admissible curve. For every $t \in [0,T]$ let us define, whenever it exists, the limit

$$v_\gamma(t) := \lim_{\varepsilon \to 0} \frac{d(\gamma(t+\varepsilon), \gamma(t))}{|\varepsilon|}. \tag{3.46}$$

(i) Prove that $v_\gamma(t)$ exist for a.e. $t \in [0,T]$.

(ii) Prove that $v_\gamma(t) = \|\dot\gamma(t)\| = |u^*(t)|$ for a.e. $t \in [0,T]$.

*Hint:* fix a dense set $\{x_n\}_{n\in\mathbb{N}}$ in $\gamma([0,T])$. Consider the functions $\varphi_n(t) = d(\gamma(t), x_n)$. Prove that $\varphi_n$ is Lipschitz for every $n$ and $v_\gamma(t) = \sup_n |\dot\varphi_n(t)|$ for a.e $t \in [0,T]$.

**Exercise 3.52.** Let $\gamma : [0,T] \to M$ be an admissible curve. Prove that

$$\ell(\gamma) = \sup\left\{ \sum_{i=1}^n d(\gamma(t_i), \gamma(t_{i-1})) : 0 = t_0 < t_1 < \ldots < t_{n-1} < t_n = T \right\}. \tag{3.47}$$

### 3.3.3 Lipschitz equivalence of sub-Riemannian distances

The goal of this section is to discuss a condition for two sub-Riemannian distances on the same manifold to be Lipschitz equivalent.

**Definition 3.53.** Let $(M, \mathbf{U}, f)$ and $(M, \mathbf{U}', f')$ be two sub-Riemannian structures on $M$. We say that the two structures are *Lipschitz equivalent* if the corresponding sub-Riemannian distance $d$ and $d'$ satisfy the following property: for every compact $K \subset M$ there exist constants $0 < c_1 < c_2$ such that

$$c_1 d(x,y) \le d'(x,y) \le c_2 d(x,y), \qquad \forall\, x, y \in K. \tag{3.48}$$

We have the following characterization of Lipschitz equivalent structures in terms of the distribution, in the case when the structure is *regular*, i.e., the distribution has constant rank.

**Proposition 3.54.** *Two complete and regular sub-Riemannian structures* $(\mathbf{U}, f)$ *and* $(\mathbf{U}', f')$ *on* $M$ *are Lipschitz equivalent if and only if they are equivalent as distributions.*

Recall that $(\mathbf{U}, f)$ and $(\mathbf{U}', f')$ are equivalent as distributions if and only if the corresponding moduli of horizontal vector fields $\mathcal{D}$ and $\mathcal{D}'$ coincide, cf. Exercice 3.20. For a regular distribution the modulus of horizontal vector fields corresponds to sections that are point-wise horizontal, cf. Exercice 3.7

$$\mathcal{D} = \{X \in \mathrm{Vec}(M) \mid X(q) \in \mathcal{D}_q, \forall\, q \in M\}. \tag{3.49}$$

*Proof.* Assume that $(\mathbf{U}, f)$ and $(\mathbf{U}', f')$ are equivalent as distribution. Thanks to Definition 3.18 this means we can assume that both structures are defined on the same vector bundle $\mathbf{V}$, endowed with different inner products $(\cdot\,|\,\cdot)$ and $(\cdot\,|\,\cdot)'$. Denote by $|v|$ and $|v|'$ the corresponding norm of an element $v \in V_q$.

Fix an arbitrary compact set $K \subset M$ and set $\overline{B} = \overline{B}(x_0, 3R)$, for some $x_0 \in K$ and $R := \max\{\mathrm{diam}\, K, \mathrm{diam}'\, K\}$. The set $\overline{B}$ is compact by completeness. Since any two norms on a finite-dimensional vector space are equivalent it follows that there exist constants $0 < c_1 < c_2$ such that

$$c_1 |v| \le |v|' \le c_2 |v|, \qquad \forall\, v \in V_q, q \in \overline{B}. \tag{3.50}$$

Then computing the length of an admissible curves $\gamma : [0, T] \to M$ contained in $\overline{B}$ we have

$$c_1\ell(\gamma) \leq \ell'(\gamma) \leq c_2\ell(\gamma). \tag{3.51}$$

Notice that if $x, y \in K$ then $\max\{d(x, y), d'(x, y)\} \leq R$ hence to compute distances it is not restrictive to consider only curves contained in $\overline{B}$. In particular if $\gamma$ is an admissible curve contained in $\overline{B}$ we have the two inequalities

$$c_1 d(x, y) \leq \ell'(\gamma), \qquad d'(x, y) \leq c_2\ell(\gamma). \tag{3.52}$$

Taking the infimum over all curves this implies

$$c_1 d(x, y) \leq d'(x, y) \leq c_2 d(x, y), \qquad \forall\, x, y \in K. \tag{3.53}$$

This proves that the two structures are Lipschitz equivalent.

Assume now that the two structures are Lipschitz equivalent, hence (3.48) holds. It follows that if $\gamma : [0, T] \to M$ is a Lipschitz curve for $d$ if and only if it is a Lipschitz curve for $d'$. Thanks to Proposition 3.50, the curve $\gamma : [0, T] \to M$ is admissible for $(\mathbf{U}, f)$ if and only if is admissible for $(\mathbf{U}', f')$. Moreover, given a horizontal vector field $X$ for $(\mathbf{U}, f)$, the integral curves of $X$ are admissible trajectories. Hence the integral curves of $X$ are admissible trajectories with respect to $(\mathbf{U}', f')$ as well. This implies that $X(q) \in \mathcal{D}'_q$ for every $q$ and thanks to (3.49) $X$ is horizontal for $(\mathbf{U}', f')$. □

### 3.3.4 Continuity of $d$ with respect to the sub-Riemannian structure

In this section, for $m \in \mathbb{N}$ we define the space $\mathcal{S}_m$ of free and *complete* sub-Riemannian structures of bundle rank $m$ defined by $f : \mathbb{R}^m \times M \to TM$.

The space $\mathcal{S}_m$ is naturally endowed with the $C^0$-topology as follows: embed $M$ into $\mathbb{R}^N$, for some $N \in \mathbb{N}$, thanks to Whitney theorem. Given $f, f' : \mathbb{R}^m \times M \to TM$, and $K \subset M$ compact, we define

$$\|f' - f\|_{0,K} = \sup\{|f'(q, v) - f(q, v)| : q \in K, \, |v| \leq 1\}.$$

The family of seminorms $\| \cdot \|_{0,K}$ induces a topology on $\mathcal{S}_m$ with a countable local basis of neighborhood as follows: take an increasing family of compact sets $\{K_n\}_{n \in \mathbb{N}}$ invading $M$, i.e., $K_n \subset K_{n+1} \subset M$ for every $n \in \mathbb{N}$ and $M = \cup_{n \in \mathbb{N}} K_n$.

For every $f \in \mathcal{S}_m$, a countable local basis of neighborhood of $f$ is given by

$$U_{f,n} := \left\{ f' \in \mathcal{S}_m : \|f' - f\|_{0,K_n} \leq \frac{1}{n} \right\}, \qquad n \in \mathbb{N}. \tag{3.54}$$

**Exercise 3.55.** (i) Prove that (3.54) defines a basis for a topology. (ii) Prove that this topology does not depend on the immersion of $M$ into $\mathbb{R}^N$.

Given $f \in \mathcal{S}_m$, we denote by $d_f$ the sub-Riemannian distance on $M$ associated with $f$.

**Theorem 3.56.** *Let $q_0, q_1 \in M$ and let $\mathcal{S}_m$ the space of free and complete sub-Riemannian structures of bundle rank $m$. The function*

$$\mathrm{dist}_{q_0,q_1} : \mathcal{S}_m \to \mathbb{R}, \qquad \mathrm{dist}_{q_0,q_1}(f) := d_f(q_0, q_1),$$

*is continuous with respect to the $C^0$-topology on $\mathcal{S}_m$.*

*Proof.* Let us prove separately the lower and the upper semi-continuity.
(i). Fix $f \in \mathcal{S}_m$ and $0 < r < d_f(q_0, q_1)$. To prove lower semi-continuity we show that there exist $\varepsilon > 0$ such that $r < d_{f'}(q_0, q_1)$ for any sub-Riemannian structure $f'$ with $\|f' - f\|_{0,K} < \varepsilon$ for a suitable choice of $K$.

Let $B_{q_0}(r)$ be the ball of radius $r$ and centered at $q_0$, with respect to the sub-Riemannian structure defined by $f$. By completeness, this is a precompact set and by construction we have $q_1 \notin B_{q_0}(r)$. Let $O \supset B_{q_0}(r)$ be an open neighbourhood of this ball in $M$ such that $q_1 \notin O$. To prove the claim it is sufficient to show that for $\varepsilon$ small enough the ball $B'_{q_0}(r)$ of radius $r$ and centered at $q_0$ defined by the sub-Riemannian structure $f'$ is also contained in $O$.

Let $K$ be a compact containing $O$ and let $a : M \to \mathbb{R}$ be a smooth cut-off function with compact support on $K$, satisfying $0 \leq a \leq 1$ and $a|_O \equiv 1$. By compactness, there exists $C > 0$ such that

$$|a(q')f(q', v) - a(q)f(q, v)| \leq C|q' - q|, \quad \forall q, q' \in M,\ |v| \leq 1. \tag{3.55}$$

Given $u \in L^\infty([0, 1]; \mathbb{R}^m)$ and $f' \in \mathcal{S}_m$, let us denote by $\gamma(t)$ (resp. $\gamma'(t)$) the solution of the equation $\dot{q} = a(q)f(q, u)$ (resp. $\dot{q} = a(q)f'(q, u)$) with initial condition $q(0) = q_0$. We set:

$$\delta_u(t) := |\gamma'(t) - \gamma(t)|.$$

Combining the definition of $\delta_u(t)$ and (3.55) one gets

$$\begin{aligned}
\delta_u(t) &\leq \int_0^t |a(\gamma(s))f(\gamma(s), u(s)) - a(\gamma'(s))f'(\gamma'(s), u(s))|ds \\
&\leq \int_0^t |a(\gamma(s))f(\gamma(s), u(s)) - a(\gamma'(s))f(\gamma'(s), u(s))|ds \\
&\qquad + \int_0^t |a(\gamma'(s))f(\gamma'(s), u(s)) - a(\gamma'(s))f'(\gamma'(s), u(s))|ds \\
&\leq C \int_0^t \delta_u(s)\, ds + \|f' - f\|_{0,K} \int_0^t |u(s)|\, ds, \quad 0 \leq t \leq 1.
\end{aligned}$$

where we used that $|a| \leq 1$ on $K$. By the Gronwall lemma, the previous inequality implies that for any sub-Riemannian structure $f'$ with $\|f' - f\|_{0,K} < \varepsilon$

$$\delta_u(t) \leq e^C \|f' - f\|_{0,K} \|u\|_{L^\infty} \leq \varepsilon e^C \|u\|_{L^\infty}.$$

Choosing $\varepsilon$ small enough we have that $\gamma'(t)$ belongs to $O$ for every control $u$ such that $\|u\|_{L^\infty} \leq r$. In particular, since $a = 1$ on $O$, we have $\gamma'(t)$ coincides with the solution of $\dot{q} = f'(q, u)$ for every $t \in [0, 1]$. Thus $B'_{q_0}(r) \subset O$, as claimed.

(ii). The upper semi-continuity is valid even without completeness of the sub-Riemannian structures. Fix $r > d_f(q_0, q_1)$ and let us show that $r > d_{f'}(q_0, q_1)$ for any sub-Riemannian structure $f'$ that is $C^0$-close to $f$.

Fix $u \in L^\infty([0, 1]; \mathbb{R}^m)$ such that $\gamma_f(1; u) = q_1$, with $\|u\|_{L^\infty} = r' < r$. Notice that $\|u\|_{L^1} \leq \|u\|_{L^\infty}$. Consider the local diffeomorphism (here, as usual, $n = \dim M$)

$$\widehat{\psi} : (s_1, \ldots, s_n) \mapsto e^{-\hat{s}_1 f_{i_1}} \circ \cdots \circ e^{-\hat{s}_n f_{i_n}} \circ e^{s_n f_{i_n}} \circ \cdots \circ e^{s_1 f_{i_1}}(q_1),$$

constructed as in the proof of the Rashevskii-Chow theorem, associated with the base point $q_1$ and defined for $|s| < \varepsilon$. Notice that $\widehat{s}$ can be chosen arbitrarily small. Fix $\varepsilon > 0$ small enough so that length of all admissible curves involved in the construction is smaller then $r - r'$.

Moreover, if $f'$ is $C^0$-close to $f$, then the map

$$\widehat{\psi}' : (s_1, \ldots, s_n) \mapsto e^{-\hat{s}_1 f'_{i_1}} \circ \cdots \circ e^{-\hat{s}_n f'_{i_n}} \circ e^{s_n f'_{i_n}} \circ \cdots \circ e^{s_1 f'_{i_1}} (\gamma_{f'}(1; u))$$

is uniformly close to $\widehat{\psi}$. The map $\widehat{\psi}'$ is a map that is $C^0$-close to a local diffeomorphism, hence its image contains the point $q_1$, as a consequence of Lemma 3.57 below. This implies that we can connect $q_0$ with $q_1$ by an admissible curve of the structure $f'$ that is shorter than $r$. □

In the next lemma we use the notation $B(0, r) = \{x \in \mathbb{R}^n \mid |x| \leq r\}$.

**Lemma 3.57.** *Let $F : \mathbb{R}^n \to \mathbb{R}^n$ be a continous map such that $F(x) = x + G(x)$, with $G$ continuous and $\|G\|_0 \leq \varepsilon$. Then the image of $F$ contains the ball $B(0, \varepsilon)$.*

*Proof.* Fix $y \in B(0, \varepsilon)$ and let us prove that there exists $x$ such that $F(x) = x + G(x) = y$. This is equivalent to prove that there exists $x \in \mathbb{R}^n$ such that $x = y - G(x)$, i.e., the map $\Phi : \mathbb{R}^n \to \mathbb{R}^n$ with $\Phi(x) = y - G(x)$ has a fixed point. But $\Phi$ is continuous and $\Phi(B(0, 2\varepsilon)) \subset B(0, 2\varepsilon)$ so, from the Brower fixed point theorem, it has a fixed point. □

As a consequence of Lemma 3.57 and the classical inverse function theorem, one obtains the following statement which completes the proof of Theorem 3.56.

**Lemma 3.58.** *Let $F : M \to N$ be a continuous map between smooth manifolds that is a small perturbation in the $C^0$-norm of a smooth map $\Phi$, where $\Phi$ is a local diffeomorphism at $q_0 \in M$. Then the image of $F$ contains $\Phi(q_0)$.*

## 3.4 Pontryagin extremals

In this section we want to give necessary conditions to characterize length-minimizer trajectories. To begin with, we would like to motivate our Hamiltonian approach that we develop in the sequel.

In classical Riemannian geometry length-minimizer trajectories satisfy a necessary condition given by a second order differential equation in $M$, which can be reduced to a first-order differential equation in $TM$. Hence the set of all length-minimizers is contained in the set of *extremals*, i.e., trajectories that satisfy the necessary condition, that are be parametrized by initial position and velocity.

In our setting (which includes Riemannian and sub-Riemannian geometry) we cannot use the initial velocity to parametrize length-minimizing trajectories. This can be easily understood by a dimensional argument. If the rank of the sub-Riemannian structure is smaller than the dimension of the manifold, the initial velocity $\dot{\gamma}(0)$ of an admissible curve $\gamma(t)$ starting from $q_0$, belongs to the proper subspace $\mathcal{D}_{q_0}$ of the tangent space $T_{q_0}M$. Hence the set of admissible velocities form a set whose dimension is smaller than the dimension of $M$, even if, by the Rashevskii-Chow theorem and Theorem 3.43, length-minimizer trajectories starting from a point $q_0$ cover a full neighborhood of $q_0$.

The right approach is to parametrize length-minimizers by their initial point and an initial *covector* $\lambda_0 \in T^*_{q_0}M$, which can be thought as the linear form annihilating the "front", i.e., the set

$F_\varepsilon = \{\gamma_{q_0}(\varepsilon) \mid \gamma_{q_0}$ is a length-minimizer starting from $q_0\}$ on the corresponding length-minimizer trajectory, for $\varepsilon \to 0$.

The next theorem gives the necessary condition satisfied by length-minimizers in sub-Riemannian geometry. Curves satisfying this condition are called *Pontryagin extremals.* The proof the following theorem is given in the next section.

**Theorem 3.59** (Characterization of Pontryagin extremals)**.** *Let $\gamma : [0,T] \to M$ be an admissible curve which is a length-minimizer, parametrized by constant speed. Let $\overline{u}(\cdot)$ be the corresponding minimal control, i.e., for a.e. $t \in [0,T]$*

$$\dot{\gamma}(t) = \sum_{i=1}^{m} \overline{u}_i(t) f_i(\gamma(t)), \qquad \ell(\gamma) = \int_0^T |\overline{u}(t)| dt = d(\gamma(0), \gamma(T)),$$

*with $|\overline{u}(t)|$ constant a.e. on $[0,T]$. Denote with $P_{0,t}$ the flow*[2] *of the nonautonomous vector field $f_{\overline{u}(t)} = \sum_{i=1}^{k} \overline{u}_i(t) f_i$. Then there exists $\lambda_0 \in T^*_{\gamma(0)}M$ such that defining*

$$\lambda(t) := (P_{0,t}^{-1})^* \lambda_0, \qquad \lambda(t) \in T^*_{\gamma(t)}M, \tag{3.56}$$

*we have that one of the following conditions is satisfied:*

*(N)* $\overline{u}_i(t) \equiv \langle \lambda(t), f_i(\gamma(t)) \rangle, \quad \forall\, i = 1, \dots, m,$

*(A)* $0 \equiv \langle \lambda(t), f_i(\gamma(t)) \rangle, \quad \forall\, i = 1, \dots, m.$

*Moreover in case (A) one has $\lambda_0 \neq 0$.*

**Definition 3.60.** Let $\gamma : [0,T] \to M$ be an admissible curve with minimal control $\overline{u} \in L^\infty([0,T], \mathbb{R}^m)$. Fix $\lambda_0 \in T^*_{\gamma(0)}M \setminus \{0\}$, and define $\lambda(t)$ by (3.56).

- If $\lambda(t)$ satisfies $(N)$ then it is called *normal extremal* (and $\gamma(t)$ a *normal extremal trajectory*).
- If $\lambda(t)$ satisfies $(A)$ then it is called *abnormal extremal* (and $\gamma(t)$ a *abnormal extremal trajectory*).

Observe that, by construction, the curve $\lambda(t)$ is Lipschitz continuous.

**Exercise 3.61.** Prove that condition (N) of Theorem 3.56 implies that the minimal control $\overline{u}(t)$ is smooth. In particular normal extremals are smooth.

For a given lift $\lambda(t)$, the two conditions (N) and (A) are mutually exclusive, unless $\overline{u}(t) = 0$ for a.e. $t \in [0,T]$, i.e., $\gamma$ is the trivial trajectory.

If the sub-Riemannian structure is not Riemannian at $q_0$, namely if

$$\mathcal{D}_{q_0} = \operatorname{span}_{q_0}\{f_1, \dots, f_m\} \neq T_{q_0}M,$$

then the trivial trajectory, corresponding to $\overline{u}(t) \equiv 0$, is always normal (with associated $\lambda_0 = 0$) and abnormal (with associated $\lambda_0 \in \mathcal{D}_{q_0}^\perp$).

Notice that even a nontrivial admissible trajectory $\gamma$ can be both normal and abnormal, since there may exist two different lifts $\lambda(t), \lambda'(t) \in T^*_{\gamma(t)}M$, such that $\lambda(t)$ satisfies $(N)$ and $\lambda'(t)$ satisfies $(A)$.

[2] $P_{0,t}(x)$ is defined for $t \in [0,T]$ and $x$ in a neighborhood of $\gamma(0)$.

*Remark* 3.62. In the Riemannian case there are no abnormal extremals. Indeed, since the map $f$ is fiberwise surjective, we can always find $m$ vector fields $f_1, \ldots, f_m$ on $M$ such that

$$\operatorname{span}_{q_0}\{f_1, \ldots, f_m\} = T_{q_0}M,$$

and $(A)$ would imply that $\langle \lambda_0, v\rangle = 0$, for all $v \in T_{q_0}M$, that gives the contradiction $\lambda_0 = 0$.

At this level it seems yet not obvious how to use Theorem 3.59 to find the explicit expression of extremals for a given sub-Riemannian structure. In the next chapter we provide another formulation of Theorem 3.59 which gives Pontryagin extremals as solutions of a Hamiltonian system.

The rest of this section is devoted to the proof of Theorem 3.59.

### 3.4.1 The energy functional

Let $\gamma : [0, T] \to M$ be an admissible curve. We define the *energy* functional $J$ on the space of Lipschitz curves on $M$ as follows

$$J(\gamma) = \frac{1}{2}\int_0^T \|\dot\gamma(t)\|^2 dt.$$

Notice that $J(\gamma) < +\infty$ for every admissible curve $\gamma$.

*Remark* 3.63. While $\ell$ is invariant by reparametrization (see Remark 3.15), $J$ is not. Indeed consider, for every $\alpha > 0$, the reparametrized curve

$$\gamma_\alpha : [0, T/\alpha] \to M, \qquad \gamma_\alpha(t) = \gamma(\alpha t).$$

Using that $\dot\gamma_\alpha(t) = \alpha\,\dot\gamma(\alpha t)$, we have

$$J(\gamma_\alpha) = \frac{1}{2}\int_0^{T/\alpha} \|\dot\gamma_\alpha(t)\|^2 dt = \frac{1}{2}\int_0^{T/\alpha} \alpha^2\|\dot\gamma(\alpha t)\|^2 dt = \alpha\, J(\gamma).$$

Thus, if the final time is not fixed, the infimum of $J$, among admissible curves joining two fixed points, is always zero.

The following lemma relates minimizers of $J$ with fixed final time with minimizers of $\ell$.

**Lemma 3.64.** *Fix $T > 0$ and let $\Omega_{q_0,q_1}$ be the set of admissible curves joining $q_0, q_1 \in M$. An admissible curve $\gamma : [0, T] \to M$ is a minimizer of $J$ on $\Omega_{q_0,q_1}$ if and only if it is a minimizer of $\ell$ on $\Omega_{q_0,q_1}$ and has constant speed.*

*Proof.* Applying the Cauchy-Schwarz inequality

$$\left(\int_0^T f(t)g(t)dt\right)^2 \le \int_0^T f(t)^2 dt \int_0^T g(t)^2 dt, \tag{3.57}$$

with $f(t) = \|\dot\gamma(t)\|$ and $g(t) = 1$ we get

$$\ell(\gamma)^2 \le 2J(\gamma)T. \tag{3.58}$$

Moreover in (3.57) equality holds if and only if $f$ is proportional to $g$, i.e., $\|\dot\gamma(t)\| = \text{const.}$ in (3.58). Since, by Lemma 3.16, every curve is a Lipschitz reparametrization of an arc length parametrized one, the minima of $J$ are attained at admissible curves with constant speed, and the statement follows.

$\square$

### 3.4.2 Proof of Theorem 3.59

By Lemma 3.64 we can assume that $\gamma$ is a minimizer of the functional $J$ among admissible curves joining $q_0 = \gamma(0)$ and $q_1 = \gamma(T)$ in fixed time $T > 0$. In particular, if we define the functional

$$\widetilde{J}(u(\cdot)) := \frac{1}{2}\int_0^T |u(t)|^2 dt, \tag{3.59}$$

on the space of controls $u(\cdot) \in L^\infty([0,T], \mathbb{R}^m)$, the minimal control $\overline{u}(\cdot)$ of $\gamma$ is a minimizer for the energy functional $\widetilde{J}$

$$\widetilde{J}(\overline{u}(\cdot)) \leq \widetilde{J}(u(\cdot)), \quad \forall\, u \in L^\infty([0,T], \mathbb{R}^m),$$

where trajectories corresponding to $u(\cdot)$ join $q_0, q_1 \in M$. In the following we denote the functional $\widetilde{J}$ by $J$.

Consider now a variation $u(\cdot) = \overline{u}(\cdot) + v(\cdot)$ of the control $\overline{u}(\cdot)$, and its associated trajectory $q(t)$, solution of the equation

$$\dot{q}(t) = f_{u(t)}(q(t)), \qquad q(0) = q_0, \tag{3.60}$$

Recall that $P_{0,t}$ denotes the local flow associated with the optimal control $\overline{u}(\cdot)$ and that $\gamma(t) = P_{0,t}(q_0)$ is the optimal admissible curve. We stress that in general, for $q$ different from $q_0$, the curve $t \mapsto P_{0,t}(q)$ is not optimal. Let us introduce the curve $x(t)$ defined by the identity

$$q(t) = P_{0,t}(x(t)). \tag{3.61}$$

In other words $x(t) = P_{0,t}^{-1}(q(t))$ is obtained by applying the inverse of the flow of $\overline{u}(\cdot)$ to the solution associated with the new control $u(\cdot)$ (see Figure 3.5). Notice that if $v(\cdot) = 0$, then $x(t) \equiv q_0$.

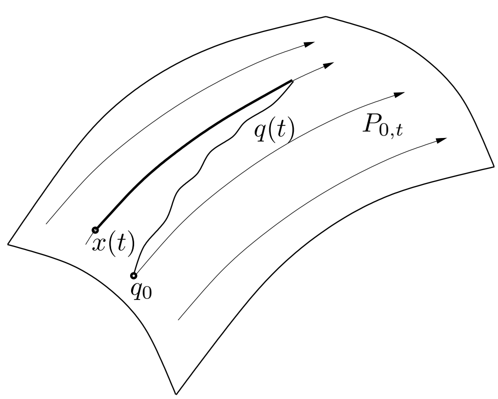



Figure 3.5: The trajectories $q(t)$, associated with $u(\cdot) = \overline{u}(\cdot) + v(\cdot)$, and the corresponding $x(t)$.

The next step is to write the ODE satisfied by $x(t)$. Differentiating (3.61) we get

$$\begin{aligned}\dot{q}(t) &= f_{\overline{u}(t)}(q(t)) + (P_{0,t})_*(\dot{x}(t)) && (3.62)\\ &= f_{\overline{u}(t)}(P_{0,t}(x(t))) + (P_{0,t})_*(\dot{x}(t)) && (3.63)\end{aligned}$$

and using that $\dot{q}(t) = f_{u(t)}(q(t)) = f_{u(t)}(P_{0,t}(x(t)))$ we can invert (3.63) with respect to $\dot{x}(t)$ and rewrite it as follows

$$\begin{aligned}
\dot{x}(t) &= (P_{0,t}^{-1})_* \left[(f_{u(t)} - f_{\overline{u}(t)})(P_{0,t}(x(t)))\right] \\
&= \left[(P_{0,t}^{-1})_*(f_{u(t)} - f_{\overline{u}(t)})\right](x(t)) \\
&= \left[(P_{0,t}^{-1})_*(f_{u(t)-\overline{u}(t)})\right](x(t)) \\
&= \left[(P_{0,t}^{-1})_* f_{v(t)}\right](x(t))
\end{aligned} \tag{3.64}$$

If we define the nonautonomous vector field $g_{v(t)}^t = (P_{0,t}^{-1})_* f_{v(t)}$ we finally obtain by (3.64) the following Cauchy problem for $x(t)$

$$\dot{x}(t) = g_{v(t)}^t(x(t)), \qquad x(0) = q_0. \tag{3.65}$$

Notice that the vector field $g_v^t$ is linear with respect to $v$, since $f_u$ is linear with respect to $u$. Now we fix the control $v(t)$ and consider the map

$$s \in \mathbb{R} \mapsto \begin{pmatrix} J(\overline{u} + sv) \\ x(T; \overline{u} + sv) \end{pmatrix} \in \mathbb{R} \times M$$

where $x(T; \overline{u} + sv)$ denote the solution at time $T$ of (3.65), starting from $q_0$, corresponding to control $\overline{u}(\cdot) + sv(\cdot)$, and $J(\overline{u} + sv)$ is the associated cost.

**Lemma 3.65.** *There exists $\bar{\lambda} \in (\mathbb{R} \oplus T_{q_0}M)^*$, with $\bar{\lambda} \neq 0$, such that for all $v \in L^\infty([0,T], \mathbb{R}^m)$*

$$\left\langle \bar{\lambda}, \left( \frac{\partial J(\overline{u} + sv)}{\partial s}\Big|_{s=0}, \frac{\partial x(T; \overline{u} + sv)}{\partial s}\Big|_{s=0} \right) \right\rangle = 0. \tag{3.66}$$

*Proof of Lemma 3.65.* We argue by contradiction: assume that (3.66) is not true, then there exist $v_0, \ldots, v_n \in L^\infty([0,T], \mathbb{R}^m)$ such that the vectors in $\mathbb{R} \oplus T_{q_0}M$

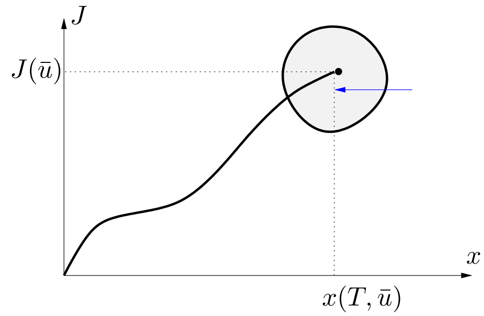
$$\begin{pmatrix} \dfrac{\partial J(\overline{u} + sv_0)}{\partial s}\Big|_{s=0} \\ \dfrac{\partial x(T; \overline{u} + sv_0)}{\partial s}\Big|_{s=0} \end{pmatrix}, \ \ldots \ , \begin{pmatrix} \dfrac{\partial J(\overline{u} + sv_n)}{\partial s}\Big|_{s=0} \\ \dfrac{\partial x(T; \overline{u} + sv_n)}{\partial s}\Big|_{s=0} \end{pmatrix} \tag{3.67}$$

are linearly independent. Let us then consider the map

$$\Phi : \mathbb{R}^{n+1} \to \mathbb{R} \times M, \qquad \Phi(s_0, \ldots, s_n) = \begin{pmatrix} J(\overline{u} + \sum_{i=0}^n s_i v_i) \\ x(T; \overline{u} + \sum_{i=0}^n s_i v_i) \end{pmatrix}. \tag{3.68}$$

By differentiability properties of solution of smooth ODEs with respect to parameters, the map (3.68) is smooth in a neighborhood of $s = 0$. Moreover, since the vectors (3.67) are the components of the differential of $\Phi$ and they are independent, then the inverse function theorem implies that $\Phi$ is a local diffeomorphism sending a neighborhood of $s = 0$ in $\mathbb{R}^{n+1}$ in a neighborhood of $(J(\overline{u}), q_0)$ in $\mathbb{R} \times M$. As a result we can find $v(\cdot) = \sum_i s_i v_i(\cdot)$ such that (see also Figure 3.4.2)

$$x(T; \overline{u} + v) = q_0, \qquad J(\overline{u} + v) < J(\overline{u}).$$




In other words the curve $t \mapsto q(t;\overline{u}+v)$ joins $q(0;\overline{u}+v) = q_0$ to

$$q(T;\overline{u}+v) = P_{0,T}(x(T;\overline{u}+v)) = P_{0,T}(q_0) = q_1,$$

with a cost smaller that the cost of $\gamma(t) = q(t;\overline{u})$, which is a contradiction $\square$

*Remark* 3.66. Notice that if $\bar{\lambda}$ satisfies (3.66), then for every $\alpha \in \mathbb{R}$, with $\alpha \neq 0$, $\alpha\bar{\lambda}$ satisfies (3.66) too. Thus we can normalize $\bar{\lambda}$ to be $(-1,\lambda_0)$ or $(0,\lambda_0)$, with $\lambda_0 \in T^*_{q_0}M$, and $\lambda_0 \neq 0$ in the second case (since $\bar{\lambda}$ is not zero).

Condition (3.66) implies that there exists $\lambda_0 \in T^*_{q_0}M$ such that one of the following identities is satisfied for all $v \in L^\infty([0,T],\mathbb{R}^m)$:

$$\frac{\partial J(\overline{u}+sv)}{\partial s}\Big|_{s=0} = \left\langle \lambda_0, \frac{\partial x(T;\overline{u}+sv)}{\partial s}\Big|_{s=0} \right\rangle, \tag{3.69}$$

$$0 = \left\langle \lambda_0, \frac{\partial x(T;\overline{u}+sv)}{\partial s}\Big|_{s=0} \right\rangle. \tag{3.70}$$

with $\lambda_0 \neq 0$ in the second case (cf. Remark 3.66). To end the proof we have to show that identities (3.69) and (3.70) are equivalent to conditions (N) and (A) of Theorem 3.59. Let us show that

$$\frac{\partial J(\overline{u}+sv)}{\partial s}\Big|_{s=0} = \int_0^T \sum_{i=1}^m \overline{u}_i(t)v_i(t)dt, \tag{3.71}$$

$$\frac{\partial x(T;\overline{u}+sv)}{\partial s}\Big|_{s=0} = \int_0^T g^t_{v(t)}(q_0)dt = \int_0^T \sum_{i=1}^m ((P^{-1}_{0,t})_*f_i)(q_0)v_i(t)dt. \tag{3.72}$$

The identity (3.71) follows from the definition of $J$

$$J(\overline{u}+sv) = \frac{1}{2}\int_0^T |\overline{u}+sv|^2 dt. \tag{3.73}$$

Eq. (3.72) can be proved in coordinates. Indeed by (3.65) and the linearity of $g_v$ with respect to $v$ we have

$$x(T;\overline{u}+sv) = q_0 + s\int_0^T g^t_{v(t)}(x(t;\overline{u}+sv))dt,$$

and differentiating with respect to $s$ at $s=0$ one gets (3.72).

Let us show that (3.69) is equivalent to $(N)$ of Theorem 3.59. Similarly, one gets that (3.70) is equivalent to $(A)$. Using (3.71) and (3.72), equation (3.69) is rewritten as

$$\begin{aligned}\int_0^T \sum_{i=1}^m \overline{u}_i(t)v_i(t)dt &= \int_0^T \sum_{i=1}^m \left\langle \lambda_0, ((P_{0,t}^{-1})_* f_i)(q_0) \right\rangle v_i(t)dt \\ &= \int_0^T \sum_{i=1}^m \langle \lambda(t), f_i(\gamma(t)) \rangle\, v_i(t)dt, \end{aligned} \tag{3.74}$$

where we used, for every $i = 1, \dots, m$, the identities

$$\left\langle \lambda_0, ((P_{0,t}^{-1})_* f_i)(q_0) \right\rangle = \left\langle \lambda_0, (P_{0,t}^{-1})_* f_i(\gamma(t)) \right\rangle = \left\langle (P_{0,t}^{-1})^* \lambda_0, f_i(\gamma(t)) \right\rangle = \langle \lambda(t), f_i(\gamma(t)) \rangle .$$

Since $v_i(\cdot) \in L^\infty([0,T], \mathbb{R}^m)$ are arbitrary, we get $\overline{u}_i(t) = \langle \lambda(t), f_i(\gamma(t)) \rangle$ for a.e. $t \in [0,T]$.

## 3.5 Appendix: Measurability of the minimal control

In this appendix we prove a technical lemma about measurability of solutions to a class of minimization problems. This lemma when specified to the sub-Riemannian context, implies that the minimal control associated with an admissible curve is measurable.

### 3.5.1 A measurability lemma

Let us fix an interval $I = [a,b] \subset \mathbb{R}$ and a compact set $U \subset \mathbb{R}^m$. Consider two functions $g : I \times U \to \mathbb{R}^n$, $v : I \to \mathbb{R}^n$ such that

(M1) $g(\cdot, u)$ is measurable in $t$ for every fixed $u \in U$,

(M2) $g(t, \cdot)$ is continuous in $u$ for every fixed $t \in I$,

(M3) $v(t)$ is measurable with respect to $t$.

Moreover we assume that

(M4) for every fixed $t \in I$, the problem $\min\{|u| : g(t,u) = v(t), u \in U\}$ has a unique solution.

Let us denote by $u^*(t)$ the solution of (M4) for a fixed $t \in I$.

**Lemma 3.67.** *Under assumptions (M1)-(M4), the function $t \mapsto |u^*(t)|$ is measurable on $I$.*

*Proof.* To prove the lemma we show that for every fixed $r > 0$ the set

$$A = \{t \in I : |u^*(t)| \leq r\},$$

is measurable in $\mathbb{R}$. By our assumptions

$$A = \{t \in I : \exists\, u \in U \text{ s.t. } |u| \leq r,\ g(t,u) = v(t)\}.$$

Let us fix $r > 0$ and a countable dense set $\{u_i\}_{i \in \mathbb{N}}$ in the ball of radius $r$ in $U$. Let show that

$$A = \bigcap_{n \in \mathbb{N}} A_n = \bigcap_{n \in \mathbb{N}} \underbrace{\bigcup_{i \in \mathbb{N}} A_{i,n}}_{:= A_n}, \tag{3.75}$$

where

$$A_{i,n} := \{t \in I : \, |g(t, u_i) - v(t)| < 1/n\}.$$

Notice that the set $A_{i,n}$ is measurable by construction and, if (3.75) is true, $A$ is also measurable.

$\subset$ inclusion. Let $t \in A$. This means that there exists $\bar{u} \in U$ such that $|\bar{u}| \leq r$ and $g(t, \bar{u}) = v(t)$. Since $g$ is continuous with respect to $u$ and $\{u_i\}_{i\in\mathbb{N}}$ is a dense, for each $n$ we can find $u_{i_n}$ such that $|g(t, u_{i_n}) - v(t)| < 1/n$, that is $t \in A_n$ for all $n$.

$\supset$ inclusion. Assume $t \in \bigcap_{n\in\mathbb{N}} A_n$. Then for every $n$ there exists $i_n$ such that the corresponding $u_{i_n}$ satisfies $|g(t, u_{i_n}) - v(t)| < 1/n$. From the sequence $u_{i_n}$, by compactness, it is possible to extract a convergent susequence $u_{i_n} \to \bar{u}$. By continuity of $g$ with respect to $u$ one easily gets that $g(t, \bar{u}) = v(t)$. That is $t \in A$. □

Next we exploit the fact that the scalar function $\varphi(t) := |u^*(t)|$ is measurable to show that the vector function $u^*(t)$ is measurable.

**Lemma 3.68.** *Under assumptions (M1)-(M4), the vector function $t \mapsto u^*(t)$ is measurable on $I$.*

*Proof.* In this proof we denote by $\varphi(t) := |u^*(t)|$. It is sufficient to prove that, for every closed ball $O$ in $\mathbb{R}^n$ the set

$$B := \{t \in I : u^*(t) \in O\}$$

is measurable. Since the minimum in (M4) is uniquely determined, this set is equal to

$$B = \{t \in I : \exists\, u \in O \text{ s.t. } |u| = \varphi(t),\, g(t, u) = v(t)\}.$$

Let us fix the ball $O$ and a countable dense set $\{u_i\}_{i\in\mathbb{N}}$ in $O$. Let show that

$$B = \bigcap_{n\in\mathbb{N}} B_n = \bigcap_{n\in\mathbb{N}} \underbrace{\bigcup_{i\in\mathbb{N}} B_{i,n}}_{:=B_n} \tag{3.76}$$

where

$$B_{i,n} := \{t \in I : \, |u_i| < \varphi(t) + 1/n, |g(t, u_i) - v(t)| < 1/n; \}$$

Notice that the set $B_{i,n}$ is measurable by construction and, if (3.76) is true, $B$ is also measurable.

$\subset$ inclusion. Let $t \in B$. This means that there exists $\bar{u} \in O$ such that $|\bar{u}| = \varphi(t)$ and $g(t, \bar{u}) = v(t)$. Since $g$ is continuous with respect to $u$ and $\{u_i\}_{i\in\mathbb{N}}$ is a dense in $O$, for each $n$ we can find $u_{i_n}$ such that $|g(t, u_{i_n}) - v(t)| < 1/n$ and $|u_{i_n}| < \varphi(t) + 1/n$, that is $t \in B_n$ for all $n$.

$\supset$ inclusion. Assume $t \in \bigcap_{n\in\mathbb{N}} B_n$. Then for every $n$ it is possible to find $i_n$ such that the corresponding $u_{i_n}$ satisfies $|g(t, u_{i_n}) - v(t)| < 1/n$ and $|u_{i_n}| < \varphi(t) + 1/n$. From the sequence $u_{i_n}$, by compactness of the closed ball $O$, it is possible to extract a convergent susequence $u_{i_n} \to \bar{u}$. By continuity of $f$ in $u$ one easily gets that $g(t, \bar{u}) = v(t)$. Moreover $|\bar{u}| \leq \varphi(t)$. Hence $|\bar{u}| = \varphi(t)$. That is $t \in B$. □

### 3.5.2 Proof of Lemma 3.12

Consider an admissible curve $\gamma : [0,T] \to M$. Since measurability is a local property it is not restrictive to assume $M = \mathbb{R}^n$. Moreover, by Lemma 3.16, we can assume that $\gamma$ is arc length parametrized so that its minimal control belong to the compact set $U = \{|u| \leq 1\}$. Define $g : [0,T] \times U \to \mathbb{R}^n$ and $v : [0,T] \to \mathbb{R}^n$ by

$$g(t,u) = f(\gamma(t), u), \qquad v(t) = \dot{\gamma}(t).$$

Assumptions (M1)-(M4) are satisfied. Indeed (M1)-(M3) follow from the fact that $g(t,u)$ is linear with respect to $u$ and measurable in $t$. Moreover (M4) is also satisfied by linearity with respect to $u$ of $f$. Applying Lemma 3.68 one gets that the minimal control $u^*(t)$ is measurable in $t$.

## 3.6 Appendix: Lipschitz vs absolutely continuous admissible curves

In these lecture notes sub-Riemannian geometry is developed in the framework of Lipschitz admissible curves (that correspond to the choice of $L^\infty$ controls). However, the theory can be equivalently developed in the framework of $W^{1,2}$ admissible curves (corresponding to $L^2$ controls) or in the framework of absolutely continuous admissible curves (corresponding to $L^1$ controls).

**Definition 3.69.** An absolutely continuous curve $\gamma : [0,T] \to M$ is said to be *AC-admissible* if there exists an $L^1$ function $u : t \in [0,T] \mapsto u(t) \in U_{\gamma(t)}$ such that $\dot{\gamma}(t) = f(\gamma(t), u(t))$, for a.e. $t \in [0,T]$. We define $W^{1,2}$-admissible curves similarly.

Being the set of absolutely continuous curve bigger than the set of Lipschitz ones, one could expect that the sub-Riemannian distance between two points is smaller when computed among all absolutely continuous admissible curves. However this is not the case thanks to the invariance by reparametrization. Indeed Lemmas 3.15 and 3.16 can be rewritten in the absolutely continuous framework in the following form.

**Lemma 3.70.** *The length of an AC-admissible curve is invariant by AC reparametrization.*

**Lemma 3.71.** *Any AC-admissible curve of positive length is a AC reparametrization of an arc length parametrized admissible one.*

The proof of Lemma 3.70 differs from the one of Lemma 3.15 only by the fact that, if $u^* \in L^1$ is the minimal control of $\gamma$ then $(u^* \circ \varphi)\dot{\varphi}$ is the minimal control associated with $\gamma \circ \varphi$. One can prove that $(u^* \circ \varphi)\dot{\varphi} \in L^1$, using the monotonicity of $\varphi$. Under these assumptions the change of variables formula (3.15) holds (see [Rud87, Ch. 7]).

The proof of Lemma 3.71 is unchanged. Notice that the statement of Exercise 3.17 remains true if we replace Lipschitz with absolutely continuous. We stress that the curve $\gamma$ built in the proof is Lipschitz (since it is arc length parametrized).

As a consequence of these results, if we define

$$d_{AC}(q_0, q_1) = \inf\{\ell(\gamma) \,|\, \gamma : [0,T] \to M \ AC\text{-admissible}, \ \gamma(0) = q_0, \ \gamma(T) = q_1\}, \tag{3.77}$$

we have the following proposition.

**Proposition 3.72.** $d_{AC}(q_0, q_1) = d(q_0, q_1)$

Since $L^\infty([0,T], \mathbb{R}^m) \subset L^2([0,T], \mathbb{R}^m) \subset L^1([0,T], \mathbb{R}^m)$, one can introduce $W^{1,2}$ admissible curves, i.e., those associated with $L^2$ controls, and obtains that $d_{W^{1,2}}(q_0, q_1) = d(q_0, q_1)$.

## 3.7 Bibliographical note

Sub-Riemannian manifolds have been introduced, even if with different terminology, in several contexts starting from the end of 60s, see for instance [JSC87, Hör67, Fol73, Hul76, Gav77] and [Jur97, Jur16, Pan89, GV88, Bro82, BR96, Bro84, VG87]. However, some pioneering ideas were already present in the work of Carathéodory [Car09] and Cartan [Car33]. The name sub-Riemannian geometry first appeared in [Str86].

Classical general references for sub-Riemannian geometry are [Mon02, Bel96, Mon96, Gro96, Sus96]. Some more recent monographs, written in a language similar to the one we use, are [Jea14, Rif14].

The definition of sub-Riemannian manifold using the language of bundles dates back to [Bel96, AG97]. For the original proof of the Raschevski-Chow theorem see [Ras38, Cho39]. The problem of the measurability of the minimal control can be seen as a problem of differential inclusion [BP07]. The proof of existence of sub-Riemannian length-minimizers presented here is an adaptation of the proof of Filippov theorem in optimal control. The fact that in sub-Riemannian geometry there exist strictly abnormal length-minimizers is due to Montgomery [Mon94, Mon02]. The fact that the theory can be equivalently developed for Lipschitz or absolutely continuous curves is well-known, a discussion can be found in [Bel96]. A sub-Riemannian manifold, from the metric viewpoint, is a length space. A link with this theory is provided by Exercices 3.51–3.52, see also [BBI01, Ch. 2].

The characterization of Pontryagin extremals given in Theorem 3.59 is a simplified version of the Pontryagin Maximum Principle [PBGM62]. The proof presented here is original and adapted to this setting. For more general versions of the Pontryagin Maximum Principle see [AS04, BP07]. The fact that every sub-Riemannian structure is equivalent to a free one (cf. Section 3.1.4) is a consequence of classical results on fiber bundles. A different proof in the case of classical (constant rank) distribution was also considered in [Rif14, Sus08].

# Chapter 4

# Pontryagin extremals: characterization and local minimality

This chapter is devoted to the study of geometric properties of Pontryagin extremals. To this purpose we first rewrite Theorem 3.59 in a more geometric setting, which permits to write a differential equation in $T^*M$ satisfied by Pontryagin extremals and to show that they do not depend on the choice of a generating family. Finally we prove that small pieces of normal extremal trajectories are length-minimizers.

To this aim, all along this chapter we develop the language of symplectic geometry, starting from the key concept of Poisson bracket.

## 4.1 Geometric characterization of Pontryagin extremals

Let $M$ be endowed with a sub-Riemannian structure with generating family $f_1, \dots, f_m$ and let $\gamma : [0,T] \to M$ be a length minimizer parametrized by constant speed, and associated with its minimal control $u(\cdot)$. In the previous chapter we proved that there exists $\lambda_0 \in T^*_{\gamma(0)}M$ such that defining

$$\lambda(t) = (P_{0,t}^{-1})^* \lambda_0, \qquad \lambda(t) \in T^*_{\gamma(t)}M, \tag{4.1}$$

one of the following conditions is satisfied:

(N) $u_i(t) \equiv \langle \lambda(t), f_i(\gamma(t)) \rangle, \quad \forall\, i = 1, \dots, m,$

(A) $0 \equiv \langle \lambda(t), f_i(\gamma(t)) \rangle, \quad \forall\, i = 1, \dots, m, \qquad \lambda_0 \neq 0.$

Here $P_{0,t}$ denotes the flow associated with the nonautonomous vector field $f_{u(t)} = \sum_{i=1}^m u_i(t) f_i$ and

$$(P_{0,t}^{-1})^* : T_q^*M \to T^*_{P_{0,t}(q)}M. \tag{4.2}$$

is the induced flow on the cotangent space.

The goal of this section is to characterize the curve (4.1) as the integral curve of a suitable (nonautonomous) vector field on $T^*M$. To this purpose, we start by showing that a vector field on $T^*M$ is completely characterized by its action on functions that are affine on fibers. To fix the ideas, we first focus on the case in which $P_{0,t} : M \to M$ is the flow associated with an autonomous vector field $X \in \mathrm{Vec}(M)$, namely $P_{0,t} = e^{tX}$.

### 4.1.1 Lifting a vector field from $M$ to $T^*M$

We start by some preliminary considerations on the algebraic structure of smooth functions on $T^*M$. As usual $\pi : T^*M \to M$ denotes the canonical projection.

The set $C^\infty(M)$ of smooth functions on $M$ is in one-to-one correspondence with the set $C^\infty(T^*M)$ of functions that are *constant on fibers* via the map $\alpha \mapsto \pi^*\alpha = \alpha \circ \pi$. In other words we have the isomorphism of algebras

$$C^\infty(M) \simeq C^\infty_{\mathrm{cst}}(T^*M) := \{\pi^*\alpha \,|\, \alpha \in C^\infty(M)\} \subset C^\infty(T^*M). \tag{4.3}$$

In what follows, with abuse of notation, we often identify the function $\pi^*\alpha \in C^\infty(T^*M)$ with the function $\alpha \in C^\infty(M)$.

In a similar way smooth vector fields on $M$ are in a one-to-one correspondence with smooth functions in $C^\infty(T^*M)$ that are *linear on fibers* via the map $Y \mapsto a_Y$, where $a_Y(\lambda) := \langle \lambda, Y(q)\rangle$ and $q = \pi(\lambda)$.

$$\mathrm{Vec}(M) \simeq C^\infty_{\mathrm{lin}}(T^*M) := \{a_Y \,|\, Y \in \mathrm{Vec}(M)\} \subset C^\infty(T^*M). \tag{4.4}$$

Notice that this is an isomorphism as modules over $C^\infty(M)$. Indeed, as $\mathrm{Vec}(M)$ is a module over $C^\infty(M)$, we have that $C^\infty_{\mathrm{lin}}(T^*M)$ is a module over $C^\infty(M)$ as well. For any $\alpha \in C^\infty(M)$ and $a_X \in C^\infty_{\mathrm{lin}}(T^*M)$ their product is defined as $\alpha a_X := (\pi^*\alpha)a_X = a_{\alpha X} \in C^\infty_{\mathrm{lin}}(T^*M)$.

**Definition 4.1.** We say that a function $a \in C^\infty(T^*M)$ is *affine on fibers* if there exist two functions $\alpha \in C^\infty_{\mathrm{cst}}(T^*M)$ and $a_X \in C^\infty_{\mathrm{lin}}(T^*M)$ such that $a = \alpha + a_X$. In other words

$$a(\lambda) = \alpha(q) + \langle \lambda, X(q)\rangle\,, \qquad q = \pi(\lambda).$$

We denote by $C^\infty_{\mathrm{aff}}(T^*M)$ the set of affine function on fibers.

*Remark* 4.2. Linear and affine functions on $T^*M$ are particularly important since they reflects the linear structure of the cotangent bundle. In particular every vector field on $T^*M$, as a derivation of $C^\infty(T^*M)$, is completely characterized by its action on affine functions,

Indeed for a vector field $V \in \mathrm{Vec}(T^*M)$ and $f \in C^\infty(T^*M)$, one has that

$$(Vf)(\lambda) = \left.\frac{d}{dt}\right|_{t=0} f(e^{tV}(\lambda)) = \langle d_\lambda f, V(\lambda)\rangle\,, \qquad \lambda \in T^*M. \tag{4.5}$$

which depends only on the differential of $f$ at the point $\lambda$. Hence, for each fixed $\lambda \in T^*M$, to compute (4.5) one can replace the function $f$ with any affine function whose differential at $\lambda$ coincide with $d_\lambda f$. Notice that such a function is not unique.

Let us now consider the infinitesimal generator of the flow $(P_{0,t}^{-1})^* = (e^{-tX})^*$. Since it satisfies the group law

$$(e^{-tX})^* \circ (e^{-sX})^* = (e^{-(t+s)X})^*, \qquad \forall\, t, s \in \mathbb{R},$$

by Lemma 2.16 its infinitesimal generator is an autonomous vector field $V_X$ on $T^*M$. In other words we have $(e^{-tX})^* = e^{tV_X}$ for all $t$.

Let us then compute the right hand side of (4.5) when $V = V_X$ and $f$ is either a function constant on fibers or a function linear on fibers.

The action of $V_X$ on functions that are constant on fibers coincides with the action of $X$ on functions on the base manifold, in the following sense: $V_X(\beta \circ \pi) = X\beta$ for every $\beta \in C^\infty(M)$. Indeed, for every $\lambda \in T^*M$, one has

$$\frac{d}{dt}\bigg|_{t=0} \beta \circ \pi((e^{-tX})^*\lambda)) = \frac{d}{dt}\bigg|_{t=0} \beta(e^{tX}(q)) = (X\beta)(q), \qquad q = \pi(\lambda). \tag{4.6}$$

For what concerns the action of $V_X$ on functions that are linear on fibers, of the form $a_Y(\lambda) = \langle \lambda, Y(q) \rangle$, we have for all $\lambda \in T^*M$

$$\begin{aligned}
\frac{d}{dt}\bigg|_{t=0} a_Y((e^{-tX})^*\lambda) &= \frac{d}{dt}\bigg|_{t=0} \big\langle (e^{-tX})^*\lambda, Y(e^{tX}(q)) \big\rangle \\
&= \frac{d}{dt}\bigg|_{t=0} \big\langle \lambda, (e_*^{-tX}Y)(q) \big\rangle = \langle \lambda, [X,Y](q) \rangle \\
&= a_{[X,Y]}(\lambda).
\end{aligned} \tag{4.7}$$

Hence, by linearity, one gets that the action of $V_X$ on functions of $C^\infty_{\text{aff}}(T^*M)$ is given by

$$V_X(\beta + a_Y) = X\beta + a_{[X,Y]}. \tag{4.8}$$

As explained in Remark 4.2, formula (4.8) characterizes completely the generator $V_X$ of $(P_{0,t}^{-1})^*$. To find its explicit form we introduce the notion of Poisson bracket.

### 4.1.2 The Poisson bracket

The purpose of this section is to introduce an operation $\{\cdot,\cdot\}$ on $C^\infty(T^*M)$, called *Poisson bracket*. First we introduce it on the set $C^\infty_{\text{lin}}(T^*M)$, where it is induced by the Lie bracket through the identification between $\text{Vec}(M)$ and $C^\infty_{\text{lin}}(T^*M)$. Then it is uniquely extended to $C^\infty_{\text{aff}}(T^*M)$, and then on $C^\infty(T^*M)$, by requiring it to be a derivation of the algebra $C^\infty(T^*M)$ in each argument.

More precisely, we start by the following definition.

**Definition 4.3.** Let $a_X, a_Y \in C^\infty_{\text{lin}}(T^*M)$ be two linear on fibers functions associated with vector fields $X, Y \in \text{Vec}(M)$. Their *Poisson bracket* is defined by

$$\{a_X, a_Y\} := a_{[X,Y]}, \tag{4.9}$$

where $a_{[X,Y]}$ is the function in $C^\infty_{\text{lin}}(T^*M)$ associated with the vector field $[X,Y]$.

*Remark* 4.4. Recall that the Lie bracket is a bilinear, skew-symmetric map defined on $\text{Vec}(M)$, that satisfies the Leibniz rule for $X, Y \in \text{Vec}(M)$:

$$[X, \alpha Y] = \alpha[X,Y] + (X\alpha)Y, \qquad \forall\, \alpha \in C^\infty(M). \tag{4.10}$$

As a consequence, the Poisson bracket is bilinear, skew-symmetric and satisfies the following relation

$$\{a_X, \alpha\, a_Y\} = \{a_X, a_{\alpha Y}\} = a_{[X,\alpha Y]} = \alpha\, a_{[X,Y]} + (X\alpha)\, a_Y, \qquad \forall\, \alpha \in C^\infty(M). \tag{4.11}$$

Notice that this expression makes sense since the product between $\alpha \in C^\infty_{\text{cst}}(T^*M)$ and $a_X \in C^\infty_{\text{lin}}(T^*M)$ belongs to $C^\infty_{\text{lin}}(T^*M)$, with $\alpha a_X = a_{\alpha X}$.

Next, we extend this definition on the whole $C^\infty(T^*M)$.

**Proposition 4.5.** *There exists a unique bilinear and skew-simmetric map*

$$\{\cdot,\cdot\} : C^\infty(T^*M) \times C^\infty(T^*M) \to C^\infty(T^*M),$$

*that extends* (4.9) *on* $C^\infty(T^*M)$*, and that is a derivation in each argument, i.e., it satisfies*

$$\{a, bc\} = \{a, b\}c + \{a, c\}b, \qquad \forall\, a, b, c \in C^\infty(T^*M). \tag{4.12}$$

*We call this operation the Poisson bracket on* $C^\infty(T^*M)$.

*Proof.* We start by proving that, as a consequence of the requirement that $\{\cdot,\cdot\}$ is a derivation in each argument, it is uniquely extended to $C^\infty_{\text{aff}}(T^*M)$.

By linearity and skew-symmetry we are reduced to compute Poisson brackets of kind $\{a_X, \alpha\}$ and $\{\alpha, \beta\}$, where $a_X \in C^\infty_{\text{lin}}(T^*M)$ and $\alpha, \beta \in C^\infty_{\text{cst}}(T^*M)$. Using that $a_{\alpha Y} = \alpha a_Y$ and (4.12) one gets

$$\begin{aligned}\{a_X, a_{\alpha Y}\} &= \{a_X, \alpha\, a_Y\} \\ &= \alpha\{a_X, a_Y\} + \{a_X, \alpha\}a_Y.\end{aligned} \tag{4.13}$$

Comparing (4.11) and (4.13) one gets

$$\{a_X, \alpha\} = X\alpha. \tag{4.14}$$

Next, using (4.12) and (4.14), one has

$$\{a_{\alpha Y}, \beta\} = \{\alpha\, a_Y, \beta\} = \alpha\{a_Y, \beta\} + \{\alpha, \beta\}a_Y \tag{4.15}$$
$$= \alpha Y\beta + \{\alpha, \beta\}a_Y. \tag{4.16}$$

Using again (4.14) one also has $\{a_{\alpha Y}, \beta\} = \alpha Y\beta$, hence $\{\alpha, \beta\} = 0$.

Combining the previous formulas one obtains the following expression for the Poisson bracket between two affine functions on $T^*M$

$$\{a_X + \alpha, a_Y + \beta\} := a_{[X,Y]} + X\beta - Y\alpha. \tag{4.17}$$

Notice that formula (4.17) involves only the first derivatives of $a_X + \alpha$ and $a_Y + \beta$. It follows that the Poisson bracket computed at a fixed $\lambda \in T^*M$ depends only on the differential of the two functions evaluated at $\lambda$.

Next we extend this definition to $C^\infty(T^*M)$ in such a way that it is a derivation. For $f, g \in C^\infty(T^*M)$ we define

$$\{f, g\}|_\lambda := \{a_{f,\lambda}, a_{g,\lambda}\}|_\lambda, \tag{4.18}$$

where $a_{f,\lambda}$ and $a_{g,\lambda}$ are two functions in $C^\infty_{\text{aff}}(T^*M)$ such that $d_\lambda f = d_\lambda(a_{f,\lambda})$ and $d_\lambda g = d_\lambda(a_{g,\lambda})$.

*Remark* 4.6. The definition (4.18) is well posed, since if we take two different affine functions $a_{f,\lambda}$ and $a'_{f,\lambda}$ their difference satisfy $d_\lambda(a_{f,\lambda} - a'_{f,\lambda}) = d_\lambda(a_{f,\lambda}) - d_\lambda(a'_{f,\lambda}) = 0$, hence by bilinearity of the Poisson bracket

$$\{a_{f,\lambda}, a_{g,\lambda}\}|_\lambda = \{a'_{f,\lambda}, a_{g,\lambda}\}|_\lambda.$$

Let us now compute the coordinate expression of the Poisson bracket. In canonical coordinates $(p, x)$ in $T^*M$, if

$$X = \sum_{i=1}^{n} X_i(x) \frac{\partial}{\partial x_i}, \quad Y = \sum_{i=1}^{n} Y_i(x) \frac{\partial}{\partial x_i},$$

we have

$$a_X(p, x) = \sum_{i=1}^{n} p_i X_i(x), \quad a_Y(p, x) = \sum_{i=1}^{n} p_i Y_i(x).$$

and, denoting $f = a_X + \alpha$, $g = a_Y + \beta$, we have

$$\begin{aligned}
\{f, g\} &= a_{[X,Y]} + X\beta - Y\alpha \\
&= \sum_{i,j=1}^{n} p_j \left( X_i \frac{\partial Y_j}{\partial x_i} - Y_i \frac{\partial X_j}{\partial x_i} \right) + X_i \frac{\partial \beta}{\partial x_i} - Y_i \frac{\partial \alpha}{\partial x_i} \\
&= \sum_{i,j=1}^{n} X_i \left( p_j \frac{\partial Y_j}{\partial x_i} + \frac{\partial \beta}{\partial x_i} \right) - Y_i \left( p_j \frac{\partial X_j}{\partial x_i} + \frac{\partial \alpha}{\partial x_i} \right) \\
&= \sum_{i=1}^{n} \frac{\partial f}{\partial p_i} \frac{\partial g}{\partial x_i} - \frac{\partial f}{\partial x_i} \frac{\partial g}{\partial p_i}.
\end{aligned}$$

From these computations we get the formula for Poisson brackets of two functions $a, b \in C^\infty(T^*M)$

$$\{a, b\} = \sum_{i=1}^{n} \frac{\partial a}{\partial p_i} \frac{\partial b}{\partial x_i} - \frac{\partial a}{\partial x_i} \frac{\partial b}{\partial p_i}, \qquad a, b \in C^\infty(T^*M). \tag{4.19}$$

The explicit formula (4.19) shows that the extension of the Poisson bracket to $C^\infty(T^*M)$ is still a derivation. □

*Remark* 4.7. As previously discussed, the value $\{a, b\}|_\lambda$ of the Poisson bracket at a point $\lambda \in T^*M$ depends only on $d_\lambda a$ and $d_\lambda b$. In particular, the Poisson bracket computed at the point $\lambda \in T^*M$ can be seen as a skew-symmetric and nondegenerate bilinear form

$$\{\cdot, \cdot\}_\lambda : T^*_\lambda(T^*M) \times T^*_\lambda(T^*M) \to \mathbb{R}.$$

**Exercise 4.8.** Let $h = (h_1, \ldots, h_k) : T^*M \to \mathbb{R}^k$, $g : T^*M \to \mathbb{R}$ and $\varphi : \mathbb{R}^k \to \mathbb{R}$ be smooth functions. Denote by $\varphi_h := \varphi \circ h$. Prove that

$$\{\varphi_h, g\} = \sum_{i=1}^{k} \frac{\partial \varphi}{\partial h_i} \{h_i, g\}. \tag{4.20}$$

### 4.1.3 Hamiltonian vector fields

By construction, the linear operator defined by

$$\vec{a} : C^\infty(T^*M) \to C^\infty(T^*M), \qquad \vec{a}(b) := \{a, b\}, \tag{4.21}$$

is a derivation of the algebra $C^\infty(T^*M)$, therefore can be identified with an element of $\mathrm{Vec}(T^*M)$.

**Definition 4.9.** The vector field $\vec{a}$ on $T^*M$ defined by (4.21) is called the *Hamiltonian vector field* associated with the smooth function $a \in C^\infty(T^*M)$.

From (4.19) we can easily write the coordinate expression of $\vec{a}$ for any arbitrary function $a \in C^\infty(T^*M)$

$$\vec{a} = \sum_{i=1}^{n} \frac{\partial a}{\partial p_i}\frac{\partial}{\partial x_i} - \frac{\partial a}{\partial x_i}\frac{\partial}{\partial p_i}. \tag{4.22}$$

Let us recall that the flow $(P_{0,t}^{-1})^*$ is generated by some autonomous vector field $V$ on $T^*M$. The following proposition gives the explicit form of $V$.

**Proposition 4.10.** *Let $X \in \mathrm{Vec}(M)$ be a complete vector field and let $P_{0,t} = e^{tX}$. The flow on $T^*M$ defined by $(P_{0,t}^{-1})^* = (e^{-tX})^*$ is generated by the Hamiltonian vector field $\vec{a}_X$, where $a_X(\lambda) = \langle \lambda, X(q)\rangle$ and $q = \pi(\lambda)$.*

*Proof.* To prove that the generator $V$ of $(P_{0,t}^{-1})^*$ coincides with the vector field $\vec{a}_X$ it is sufficient to show that their action on affine functions is the same. Indeed, by definition of Hamiltonian vector field, we have

$$\begin{aligned}\vec{a}_X(\alpha) &= \{a_X, \alpha\} = X\alpha,\\ \vec{a}_X(a_Y) &= \{a_X, a_Y\} = a_{[X,Y]}.\end{aligned}$$

Hence this action coincides with the action of $V$ as in (4.6) and (4.7). □

*Remark* 4.11. In coordinates $(p, x)$ if the vector field $X$ is written $X = \sum_{i=1}^n X_i \frac{\partial}{\partial x_i}$ then $a_X(p,x) = \sum_{i=1}^n p_i X_i$ and the Hamitonian vector field $\vec{a}_X$ is written as follows

$$\vec{a}_X = \sum_{i=1}^{n} X_i \frac{\partial}{\partial x_i} - \sum_{i,j=1}^{n} p_i \frac{\partial X_i}{\partial x_j}\frac{\partial}{\partial p_j}. \tag{4.23}$$

Notice that the projection of $\vec{a}_X$ onto $M$ coincides with $X$ itself, i.e., $\pi_*(\vec{a}_X) = X$.

This construction can be extended to the case of nonautonomous vector fields we consider (cf. also Definition 2.13).

**Proposition 4.12.** *Let $X_t = \sum_{i=1}^m u_i(t) X_i$ be a nonautonomous vector field, where $X_1, \dots, X_m$ are smooth vector fields and $u \in L^\infty([0,T], \mathbb{R}^m)$. Denote by $P_{0,t}$ the flow of $X_t$ on $M$. Then the nonautonomous vector field on $T^*M$*

$$V_t := \overrightarrow{a_{X_t}}, \qquad a_{X_t}(\lambda) = \langle \lambda, X_t(q)\rangle,$$

*is the generator of the flow $(P_{0,t}^{-1})^*$.*

The proof is based on the following idea: from the autonomous case one proves the identity between the flow generated by $V_t$ and $(P_{0,t}^{-1})^*$ for $X_t = \sum_{i=1}^m u_i(t) X_i$ with $u$ that is piecewise constant. Then one proves the continuity of both flows with respect to the control (for instance in the $L^1$ topology). Since one can approximate any $L^\infty$ control by piecewise constant ones, the required identity follows for any nonautonomous vector field of the form $X_t = \sum_{i=1}^m u_i(t) X_i$. These steps use the validity of the variation equation for nonautonomous vector fields, namely (2.22). The details are left to the reader.

## 4.2 The symplectic structure

In this section we introduce the symplectic structure of $T^*M$ following the classical construction. In subsection 4.2.1 we show that the symplectic form can be interpreted as the "dual" of the Poisson bracket, in a suitable sense.

**Definition 4.13.** The *tautological* (or *Liouville*) 1-form $s \in \Lambda^1(T^*M)$ is defined as follows:

$$s : \lambda \mapsto s_\lambda \in T^*_\lambda(T^*M), \qquad \langle s_\lambda, w \rangle := \langle \lambda, \pi_* w \rangle, \quad \forall\, \lambda \in T^*M,\, w \in T_\lambda(T^*M),$$

where $\pi : T^*M \to M$ denotes the canonical projection.

The name "tautological" comes from its expression in coordinates. Recall that, given a system of coordinates $x = (x_1, \ldots, x_n)$ on $M$, canonical coordinates $(p, x)$ on $T^*M$ are defined in such a way that every element $\lambda \in T^*M$ is written as follows

$$\lambda = \sum_{i=1}^n p_i dx_i.$$

For every $w \in T_\lambda(T^*M)$ we have the following

$$w = \sum_{i=1}^n \alpha_i \frac{\partial}{\partial p_i} + \beta_i \frac{\partial}{\partial x_i} \quad \Longrightarrow \quad \pi_* w = \sum_{i=1}^n \beta_i \frac{\partial}{\partial x_i},$$

hence we get

$$\langle s_\lambda, w \rangle = \langle \lambda, \pi_* w \rangle = \sum_{i=1}^n p_i \beta_i = \sum_{i=1}^n p_i \langle dx_i, w \rangle = \left\langle \sum_{i=1}^n p_i dx_i, w \right\rangle.$$

In other words the coordinate expression of the Liouville form $s$ at the point $\lambda$ coincides with the one of $\lambda$ itself, namely

$$s_\lambda = \sum_{i=1}^n p_i dx_i. \tag{4.24}$$

**Exercise 4.14.** Let $s \in \Lambda^1(T^*M)$ be the tautological form. Prove that

$$\omega^* s = \omega, \qquad \forall\, \omega \in \Lambda^1(M).$$

(Recall that a 1-form $\omega$ is a section of $T^*M$, i.e., a map $\omega : M \to T^*M$ such that $\pi \circ \omega = id_M$).

**Definition 4.15.** The differential of the tautological 1-form $\sigma := ds \in \Lambda^2(T^*M)$ is called the *canonical symplectic structure* on $T^*M$.

By construction $\sigma$ is a closed 2-form on $T^*M$. Moreover its expression in canonical coordinates $(p, x)$ shows immediately that is a nondegenerate two form. Indeed from (4.24)

$$\sigma = \sum_{i=1}^n dp_i \wedge dx_i. \tag{4.25}$$

*Remark* 4.16 (The symplectic form in non-canonical coordinates). Given a basis $\omega_1, \ldots, \omega_n$ of 1-forms on $M$, one can build coordinates on the fibers of $T^*M$ as follows.

Every $\lambda \in T^*M$ can be written uniquely as $\lambda = \sum_{i=1}^n h_i \omega_i$. Thus the functions $h_i$, for $i = 1, \ldots, n$, become coordinates on the fibers. Notice that these coordinates are not related to a specific choice of coordinates on the manifold, as the $p$ were. By definition, in these coordinates, we have

$$s = \sum_{i=1}^n h_i \omega_i, \qquad \sigma = ds = \sum_{i=1}^n dh_i \wedge \omega_i + h_i d\omega_i. \tag{4.26}$$

Notice that, with respect to (4.25) in the expression of $\sigma$ an extra term appears since, in general, the 1-forms $\omega_i$ are not closed.

### 4.2.1 Symplectic form vs Poisson bracket

Let $V$ be a finite dimensional vector space and let $V^*$ denote its dual (i.e., the space of linear forms on $V$). By classical linear algebra arguments one has the following identifications

$$\left\{\begin{matrix}\text{non degenerate}\\ \text{bilinear forms on } V\end{matrix}\right\} \simeq \left\{\begin{matrix}\text{linear invertible maps}\\ V \to V^*\end{matrix}\right\} \simeq \left\{\begin{matrix}\text{non degenerate}\\ \text{bilinear forms on } V^*\end{matrix}\right\}. \tag{4.27}$$

Indeed to every bilinear form $B : V \times V \to \mathbb{R}$ we can associate a linear map $L : V \to V^*$ defined by $L(v) = B(v, \cdot)$. On the other hand, given a linear map $L : V \to V^*$, we can associate with it a bilinear map $B : V \times V \to \mathbb{R}$ defined by $B(v, w) = \langle L(v), w\rangle$, where $\langle \cdot, \cdot \rangle$ denotes as usual the pairing between a vector space and its dual. Moreover $B$ is non-degenerate if and only if the map $B(v, \cdot)$ is an isomorphism for every $v \in V$, or equivalently, if and only if $L$ is invertible.

The previous argument shows how to identify a bilinear form on $B$ on $V$ with an invertible linear map $L$ from $V$ to $V^*$ (the same reasoning applied to the linear map $L^{-1}$, produces a bilinear map on $V^*$).

If, in the previous discussion, one choses as $V = T_\lambda(T^*M)$ and as $B$ the canonical symplectic form $\sigma$, one notices the map $w \mapsto \sigma_\lambda(\cdot, w)$ defines a linear isomorphism between the vector spaces $T_\lambda(T^*M)$ and $T_\lambda^*(T^*M)$ and $\vec{h}$ is the vector field canonically associated by the symplectic structure with differential $-dh$. For this reason $\vec{h}$ is also called *symplectic gradient* of $h$.

This is formalized in the following result, whose proof is left to the reader.

**Proposition 4.17.** *Let $h \in C^\infty(T^*M)$. The Hamiltonian vector field $\vec{h} \in Vec(T^*M)$ satisfies the following identity*

$$\sigma(\cdot, \vec{h}(\lambda)) = d_\lambda h, \qquad \forall\, \lambda \in T^*M. \tag{4.28}$$

*Moreover, for every $\lambda \in T^*M$, the bilinear form $\sigma_\lambda$ on $T_\lambda(T^*M)$ and $\{\cdot, \cdot\}_\lambda$ on $T^*_\lambda(T^*M)$ (cf. Remark 4.7) are dual under the identification* (4.27). *In particular*

$$\{a, b\} = \vec{a}(b) = \langle db, \vec{a}\rangle = \sigma(\vec{a}, \vec{b}), \qquad \forall\, a, b \in C^\infty(T^*M). \tag{4.29}$$

Thanks to formula (4.25) we have that in canonical coordinates $(p, x)$ the Hamiltonian vector field associated with $h$ is expressed as follows

$$\vec{h} = \sum_{i=1}^n \frac{\partial h}{\partial p_i}\frac{\partial}{\partial x_i} - \frac{\partial h}{\partial x_i}\frac{\partial}{\partial p_i},$$

and the *Hamiltonian system* $\dot{\lambda} = \vec{h}(\lambda)$ is rewritten as

$$\begin{cases} \dot{x}_i = \dfrac{\partial h}{\partial p_i} \\ \dot{p}_i = -\dfrac{\partial h}{\partial x_i} \end{cases}, \qquad i = 1, \dots, n.$$

We conclude this section with two classical but rather important results.

**Proposition 4.18.** *A function $a \in C^\infty(T^*M)$ is a constant of the motion of the Hamiltonian system associated with $h \in C^\infty(T^*M)$ if and only if $\{h, a\} = 0$.*

*Proof.* Let us consider a solution $\lambda(t) = e^{t\vec{h}}(\lambda_0)$ of the Hamiltonian system associated with $\vec{h}$, with $\lambda_0 \in T^*M$. From (4.29), we have the following formula for the derivative of the function $a$ along the solution

$$\frac{d}{dt} a(\lambda(t)) = \{h, a\}(\lambda(t)). \tag{4.30}$$

It is then easy to see that $\{h, a\} = 0$ if and only if the derivative of the function $a$ along the flow vanishes for all $t$, that means that the function $a$ is constant along $\lambda(t)$.

□

The skew-simmetry of the Poisson bracket immediately implies the following corollary.

**Corollary 4.19.** *A function $h \in C^\infty(T^*M)$ is a constant of the motion of the Hamiltonian system defined by $\vec{h}$.*

## 4.3 Characterization of normal and abnormal Pontryagin extremals

In this section we rewrite Theorem 3.59 using the symplectic language developed in the last section.

Given a sub-Riemannian structure on $M$ with generating family $\{f_1, \dots, f_m\}$, let us define the fiberwise linear functions on $T^*M$ associated with these vector fields

$$h_i : T^*M \to \mathbb{R}, \qquad h_i(\lambda) := \langle \lambda, f_i(q) \rangle, \quad i = 1, \dots, m.$$

Recall that the generating family contains the data of the scalar product on the distribution.

**Theorem 4.20** (Hamiltonian characterization of Pontryagin extremals)**.** *Let $\gamma : [0, T] \to M$ be an admissible curve which is a length-minimizer, parametrized by constant speed. Let $\overline{u}(\cdot)$ be the corresponding minimal control. Then there exists a Lipschitz curve $\lambda(t) \in T^*_{\gamma(t)}M$ such that*

$$\dot{\lambda}(t) = \sum_{i=1}^{m} \overline{u}_i(t) \vec{h}_i(\lambda(t)), \qquad a.e.\ t \in [0, T], \tag{4.31}$$

*and one of the following conditions is satisfied:*

*(N)* $h_i(\lambda(t)) \equiv \overline{u}_i(t), \quad i = 1, \dots, m, \ \forall\, t,$

*(A)* $h_i(\lambda(t)) \equiv 0, \qquad i = 1, \dots, m, \ \forall\, t.$

*Moreover in case (A) one has $\lambda(t) \neq 0$ for all $t \in [0,T]$.*

*Proof.* The statement is a rephrasing of Theorem 3.59, obtained by combining Proposition 4.10 and Proposition 4.12. $\square$

Notice that Theorem 4.20 says that normal and abnormal extremals appear as solution of an Hamiltonian system. Nevertheless, this Hamiltonian system is a priori nonautonomous and depends on the trajectory itself by the presence of the control $\overline{u}(t)$ associated with the extremal trajectory.

Moreover, the actual formulation of Theorem 4.20 for the necessary optimality condition still does not clarify if the extremals depend on the choice of a generating family $\{f_1, \dots, f_m\}$ of the sub-Riemannian structure. The rest of the section is devoted to the geometric intrinsic description of normal and abnormal extremals.

### 4.3.1 Normal extremals

In this section we show that normal extremals are characterized as solutions of a *smooth autonomous* Hamiltonian system on $T^*M$, where the Hamiltonian $H$ is a function that encodes all the informations on the sub-Riemannian structure.

**Definition 4.21.** Let $M$ be a sub-Riemannian manifold. The *sub-Riemannian Hamiltonian* is the function on $T^*M$ defined as follows

$$H : T^*M \to \mathbb{R}, \qquad H(\lambda) = \max_{u \in U_q} \left( \langle \lambda, f_u(q) \rangle - \frac{1}{2}|u|^2 \right), \quad q = \pi(\lambda). \tag{4.32}$$

**Proposition 4.22.** *The sub-Riemannian Hamiltonian $H$ is smooth and quadratic on fibers. Moreover, for every generating family $\{f_1, \dots, f_m\}$ of the sub-Riemannian structure, the sub-Riemannian Hamiltonian $H$ is written as follows*

$$H(\lambda) = \frac{1}{2} \sum_{i=1}^{m} \langle \lambda, f_i(q) \rangle^2, \qquad \lambda \in T_q^*M, \quad q = \pi(\lambda). \tag{4.33}$$

*Proof.* In terms of a generating family $\{f_1, \dots, f_m\}$, the sub-Riemannian Hamiltonian (4.32) is written as follows

$$H(\lambda) = \max_{u \in \mathbb{R}^m} \left( \sum_{i=1}^{m} u_i \langle \lambda, f_i(q) \rangle - \frac{1}{2} \sum_{i=1}^{m} u_i^2 \right). \tag{4.34}$$

Differentiating (4.34) with respect to $u_i$, one gets that the maximum in the right hand side is attained for $u_i = \langle \lambda, f_i(q) \rangle$, from which formula (4.33) follows. The fact that $H$ is smooth and quadratic on fibers then easily follows from (4.33). $\square$

**Exercise 4.23.** Prove that two equivalent sub-Riemannian structures $(\mathbf{U}, f)$ and $(\mathbf{U}', f')$ on a manifold $M$ define the same Hamiltonian.

**Exercise 4.24.** Consider the sub-Riemannian Hamiltonian $H : T^*M \to \mathbb{R}$. Denote by $H_q$ its restriction on fiber $T_q^*M$ and fix $\lambda \in T_q^*M$. The differential $d_\lambda H_q : T_q^*M \to \mathbb{R}$ is a linear form, hence it can be canonically identified with an element of $T_qM$.

(i) Prove that $d_\lambda H_q \in \mathcal{D}_x$ for all $\lambda \in T_q^*M$.

(ii) Prove that $\|d_\lambda H_q\|^2 = 2H(\lambda)$.

*Hint:* show that, if $f_1, \ldots, f_m$ is a generating family, then

$$d_\lambda H_q = \sum_{i=1}^{m} \langle \lambda, f_i(q) \rangle \, f_i(q).$$

**Theorem 4.25.** *A Lipschitz curve $\lambda : [0,T] \to T^*M$ is a normal extremal if and only if it is a solution of the Hamiltonian system*

$$\dot{\lambda}(t) = \vec{H}(\lambda(t)). \tag{4.35}$$

*Moreover, given a normal extremal, the corresponding normal extremal trajectory $\gamma(t) = \pi(\lambda(t))$ is smooth and has constant speed satisfying*

$$\frac{1}{2}\|\dot{\gamma}(t)\|^2 = H(\lambda(t)), \qquad \forall\, t \in [0,T].$$

*Proof.* Let $\{f_1, \ldots, f_m\}$ be a generating family and denote, as usual, $h_i(\lambda) = \langle \lambda, f_i(q) \rangle$ for $i = 1, \ldots, m$. Using the identity $\overrightarrow{h_i^2} = 2h_i\vec{h}_i$ (see (4.12)), it follows that

$$\vec{H} = \overrightarrow{\frac{1}{2}\sum_{i=1}^{m} h_i^2} = \sum_{i=1}^{m} h_i \vec{h}_i.$$

Let $\lambda(t)$ be a normal extremal. In particular $h_i(\lambda(t)) = \overline{u}_i(t)$ by condition (N) of Theorem 4.20, and one gets

$$\vec{H}(\lambda(t)) = \sum_{i=1}^{m} h_i(\lambda(t))\vec{h}_i(\lambda(t)) = \sum_{i=1}^{m} \overline{u}_i(t)\vec{h}_i(\lambda(t)).$$

On the other hand assume that $\lambda(t)$ satisfies $\dot{\lambda}(t) = \vec{H}(\lambda(t))$. In terms of a generating family this implies that

$$\dot{\lambda}(t) = \sum_{i=1}^{m} h_i(\lambda(t))\vec{h}_i(\lambda(t))$$

and for its projection $\gamma(t) = \pi(\lambda(t))$ one has

$$\dot{\gamma}(t) = \sum_{i=1}^{m} h_i(\lambda(t)) f_i(\gamma(t)) = \sum_{i=1}^{m} \langle \lambda(t), f_i(\gamma(t)) \rangle \, f_i(\gamma(t)), \tag{4.36}$$

since $f_i = \pi_* \vec{h}_i$. Hence $\overline{u}_i(t) = \langle \lambda(t), f_i(\gamma(t)) \rangle$ defines a control for the curve $\gamma$. This control is indeed the minimal one thanks to (4.34) (cf. also Exercice 4.24) and

$$\frac{1}{2}\|\dot{\gamma}(t)\|^2 = \frac{1}{2}\sum_{i=1}^{m} \overline{u}_i(t)^2 = \frac{1}{2}\sum_{i=1}^{m} \langle \lambda(t), f_i(\gamma(t)) \rangle^2 = H(\lambda(t)) \tag{4.37}$$

□

*Remark* 4.26. In canonical coordinates $\lambda = (p, x)$ in $T^*M$, $H$ is quadratic with respect to $p$ and

$$H(p, x) = \frac{1}{2}\sum_{i=1}^{m} \langle p, f_i(x)\rangle^2 .$$

The Hamiltonian system associated with $H$, in these coordinates, is written as follows

$$\begin{cases} \dot{x} = \dfrac{\partial H}{\partial p} = \sum_{i=1}^{m} \langle p, f_i(x)\rangle f_i(x) \\ \dot{p} = -\dfrac{\partial H}{\partial x} = -\sum_{i=1}^{m} \langle p, f_i(x)\rangle \langle p, D_x f_i(x)\rangle \end{cases} \tag{4.38}$$

From here it is easy to see that if $\lambda(t) = (p(t), x(t))$ is a solution of (4.38) then also the rescaled extremal $\alpha\lambda(\alpha t) = (\alpha\, p(\alpha t), x(\alpha t))$ is a solution of the same Hamiltonian system, for every $\alpha > 0$.

**Corollary 4.27.** *A normal extremal trajectory is parametrized by constant speed. In particular it is arc length parametrized if and only if its extremal lift is contained in the level set $H^{-1}(1/2)$.*

*Proof.* The fact that $H$ is constant along $\lambda(t)$, easily implies by (4.37) that $\|\dot{\gamma}(t)\|^2$ is constant. Moreover one easily gets that $\|\dot{\gamma}(t)\| = 1$ if and only if $H(\lambda(t)) = 1/2$.

Finally, by Remark 4.26, all normal extremal trajectories are reparametrization of arc length parametrized ones. □

*Remark* 4.28. Notice that from (4.36) it follows that if $\gamma(t)$ is a normal extremal trajectory associated with initial covector $\lambda_0 \in T^*_{q_0}M$ it follows that

$$\dot{\gamma}(0) = \sum_{i=1}^{m} \langle \lambda_0, f_i(q_0)\rangle f_i(q_0). \tag{4.39}$$

Let $\lambda(t)$ be a normal extremal such that $\lambda(0) = \lambda_0 \in T^*_{q_0}M$. The corresponding normal extremal trajectory $\gamma(t) = \pi(\lambda(t))$ can be written in the exponential notation

$$\gamma(t) = \pi \circ e^{t\vec{H}}(\lambda_0).$$

By Corollary 4.27, arc length parametrized normal extremal trajectories corresponds to the choice of $\lambda_0 \in H^{-1}(1/2)$.

We end this section by characterizing normal extremal trajectory as characteristic curves of the canonical symplectic form contained in the level sets of $H$.

**Definition 4.29.** Let $M$ be a smooth manifold and $\Omega \in \Lambda^2 M$ a 2-form. A Lipschitz curve $\gamma : [0, T] \to M$ is a *characteristic curve* for $\Omega$ if for almost every $t \in [0, T]$ it holds

$$\dot{\gamma}(t) \in \ker \Omega_{\gamma(t)}, \quad \text{(i.e., } \Omega_{\gamma(t)}(\dot{\gamma}(t), \cdot) = 0) \tag{4.40}$$

Notice that this notion is independent on the parametrization of the curve.

**Proposition 4.30.** *Let $H$ be the sub-Riemannian Hamiltonian and assume that $c > 0$ is a regular value of $H$. Then a Lipschitz curve on $H^{-1}(c)$ is a characteristic curve for $\sigma|_{H^{-1}(c)}$ if and only if it is the reparametrization of a normal extremal.*

*Proof.* Recall that if $c$ is a regular value of $H$, then the set $H^{-1}(c)$ is a smooth $(2n-1)$-dimensional manifold in $T^*M$. Notice that, thanks to the classical Sard Theorem, almost every $c > 0$ is a regular value for $H$.

For every $\lambda \in H^{-1}(c)$ let us denote by $E_\lambda = T_\lambda H^{-1}(c)$. Notice that, by construction, $E_\lambda$ is an hyperplane in $T_\lambda(T^*M)$ (i.e., $\dim E_\lambda = 2n-1$) and $d_\lambda H\big|_{E_\lambda} = 0$. The restriction $\sigma|_{H^{-1}(c)}$ is computed by $\sigma_\lambda|_{E_\lambda}$, for each $\lambda \in H^{-1}(c)$.

On one hand, $\ker \sigma_\lambda|_{E_\lambda}$ is non-trivial since the dimension of $E_\lambda$ is odd. On the other hand, the symplectic 2-form $\sigma$ is nondegenerate on $T^*M$. It follows that $\dim \ker \sigma_\lambda|_{E_\lambda} = 1$.

We are left to show that $\ker \sigma_\lambda|_{E_\lambda}$ is spanned by $\vec{H}(\lambda)$. Let $\ker \sigma_\lambda|_{E_\lambda} = \mathbb{R}\xi$, for some $\xi \in T_\lambda(T^*M)$. By construction, $E_\lambda$ coincides with the skew-orthogonal to $\xi$, namely

$$E_\lambda = \xi^\angle = \{w \in T_\lambda(T^*M) \mid \sigma_\lambda(\xi, w) = 0\}.$$

By definition of Hamiltonian vector field one has $\sigma(\cdot, \vec{H}) = dH$, hence for the restriction to $E_\lambda$ one gets

$$\sigma_\lambda(\cdot, \vec{H}(\lambda))\big|_{E_\lambda} = d_\lambda H\big|_{E_\lambda} = 0. \qquad \square$$

**Exercise 4.31.** Assume that two smooth Hamiltonians $h_1, h_2 : T^*M \to \mathbb{R}$ define the same level set, i.e., $E = \{h_1 = c_1\} = \{h_2 = c_2\}$ for some $c_1, c_2 \in \mathbb{R}$ regular values of $h_1, h_2$ respectively, then their Hamiltonian flows $\vec{h}_1, \vec{h}_2$ coincide on $E$, up to reparametrization.

**Exercise 4.32.** The goal of this exercice is to show that, given the sub-Riemannian Hamiltonian $H$, one can recover all the information about the sub-Riemannian structure.

(a) Prove that for a vector $v \in T_qM$ the two following properties are equivalent:

  (a.1) $v \in \mathcal{D}_q$ and $\|v\| \leq 1$,

  (a.2) $\frac{1}{2}|\langle \lambda, v\rangle|^2 \leq H(\lambda)$ for all $\lambda \in T_q^*M$.

(b) Show that the sub-Riemannian Hamiltonian can be written as follows

$$H(\lambda) = \frac{1}{2}\|\lambda\|^2, \qquad \|\lambda\| = \sup_{v \in \mathcal{D}_q, \|v\|=1} |\langle \lambda, v\rangle|. \tag{4.41}$$

### 4.3.2 Abnormal extremals

In this section we provide a geometric characterization of abnormal extremals. Even if for abnormal extremals it is not possible to determine a priori their regularity (which should be understood with respect to the length parametrization), we show that unparametrized abnormal trajectories can be characterized as characteristic curves of the symplectic form. This gives an unified point of view of both classes of extremals.

We recall that an abnormal extremal is a non-vanishing solution of the following equations

$$\dot\lambda(t) = \sum_{i=1}^m u_i(t)\vec{h}_i(\lambda(t)), \qquad h_i(\lambda(t)) = 0, \ i = 1, \dots, m,$$

where $\{f_1, \dots, f_m\}$ is a generating family for the sub-Riemannian structure and $h_1, \dots, h_m$ are the corresponding functions on $T^*M$ that are linear on fibers. In particular every abnormal extremal is contained in the set

$$H^{-1}(0) = \{\lambda \in T^*M \mid \langle \lambda, f_i(q)\rangle = 0, \, i = 1, \dots, m, \, q = \pi(\lambda)\}. \tag{4.42}$$

where $H$ denotes the sub-Riemannian Hamiltonian (4.33).

Notice that 0 is never a regular value of $H$. Nevertheless, the following regularity assumption on the distribution guarantees that $H^{-1}(0)$ is a smooth manifold.

**Definition 4.33.** Consider a sub-Riemannian structure on $M$ with generating family $\{f_1, \ldots, f_m\}$. We say that the sub-Riemannian structure is *regular* if there exists $r \in \mathbb{N}$ such that for every $q \in M$

$$\dim \mathcal{D}_q = \dim \operatorname{span}_q\{f_1, \ldots, f_m\} = r. \tag{4.43}$$

In this case we say that the structure is regular of rank $r$.

Notice that, thanks to the constant rank theorem (cf. Theorem 2.60), if the sub-Riemannian structure is *regular* of rank $r$, then the set $H^{-1}(0)$ defined by (4.42) is a smooth submanifold of $T^*M$ of codimension $r$.

**Proposition 4.34.** *Let $H$ be the sub-Riemannian Hamiltonian associated with a regular sub-Riemannian structure. Then a Lipschitz curve on $H^{-1}(0)$ is a characteristic curve for $\sigma|_{H^{-1}(0)}$ if and only if it is the reparametrization of an abnormal extremal.*

*Proof.* In this proof we denote for simplicity $N := H^{-1}(0) \subset T^*M$. For every $\lambda \in N$ we have the identity

$$\ker \sigma_\lambda|_N = T_\lambda N^\angle = \operatorname{span}\{\vec{h}_i(\lambda) \mid i = 1, \ldots, m\}. \tag{4.44}$$

Indeed, from the definition of $N$ and the regularity assumption, it follows that

$$\begin{aligned} T_\lambda N &= \{w \in T_\lambda(T^*M) \mid \langle d_\lambda h_i,\, w\rangle = 0,\, i = 1, \ldots, m\} \\ &= \{w \in T_\lambda(T^*M) \mid \sigma(w, \vec{h}_i(\lambda)) = 0,\, i = 1, \ldots, m\} \\ &= \operatorname{span}\{\vec{h}_i(\lambda) \mid i = 1, \ldots, m\}^\angle, \end{aligned}$$

and (4.44) follows by taking the skew-orthogonal on both sides. Thus $w \in T_\lambda H^{-1}(0)^\angle$ if and only if $w$ is a linear combination of the vectors $\vec{h}_i(\lambda)$. This implies that $\lambda(t)$ is a characteristic curve for $\sigma|_{H^{-1}(0)}$ if and only if it is a reparametrization of a curve satisfying

$$\dot{\lambda}(t) = \sum_{i=1}^{m} u_i(t)\vec{h}_i(\lambda(t)). \qquad \square \tag{4.45}$$

for some controls $u_i(\cdot)$, with $i = 1, \ldots, m$.

*Remark* 4.35. From Proposition 4.34 it follows that abnormal extremals *do not* depend on the sub-Riemannian metric, but only on the distribution. Indeed the set $H^{-1}(0)$ is characterized as the annihilator $\mathcal{D}^\perp$ of the distribution

$$H^{-1}(0) = \{\lambda \in T^*M \mid \langle \lambda, v\rangle = 0,\ \forall\, v \in \mathcal{D}_{\pi(\lambda)}\} = \mathcal{D}^\perp \subset T^*M.$$

Here the orthogonal is meant in the duality sense.

Under the regularity assumption (4.43), we can assume without loss of generality that $f_1, \ldots, f_m$ are linearly independent (hence $r = m$ in the regularity assumption) and select (at least locally) a basis of 1-forms $\omega_1, \ldots, \omega_m$ for the dual of the distribution

$$\mathcal{D}_q^\perp = \operatorname{span}\{\omega_i(q) \mid i = 1, \ldots, m\}. \tag{4.46}$$

Let us complete this set of 1-forms to a basis $\omega_1, \ldots, \omega_n$ of $T^*M$ and consider the induced coordinates $h_1, \ldots, h_n$ as defined in Remark 4.16. In these coordinates the restriction of the symplectic structure $\mathcal{D}^\perp$ to is expressed as follows

$$\sigma|_{\mathcal{D}^\perp} = d(s|_{\mathcal{D}^\perp}) = \sum_{i=1}^{m} dh_i \wedge \omega_i + h_i d\omega_i. \tag{4.47}$$

We stress that the restriction $\sigma|_{\mathcal{D}^\perp}$ can be written only in terms of the elements $\omega_1, \ldots, \omega_m$ (and not of a full basis of 1-forms) since the differential $d$ commutes with the restriction.

### 4.3.3 Codimension one and contact distributions

Let $M$ be a $n$-dimensional manifold endowed with a constant rank distribution $\mathcal{D}$ of codimension one, i.e., $\dim \mathcal{D}_q = n-1$ for every $q \in M$. In this case $\mathcal{D}$ and $\mathcal{D}^\perp$ are sub-bundles of $TM$ and $T^*M$ respectively and their dimensions, as smooth manifolds, are

$$\dim \mathcal{D} = \dim M + \operatorname{rank} \mathcal{D} = 2n-1,$$
$$\dim \mathcal{D}^\perp = \dim M + \operatorname{rank} \mathcal{D}^\perp = n+1.$$

Since the symplectic form $\sigma$ is skew-symmetric, a dimensional argument implies that for $n$ even, the restriction $\sigma|_{\mathcal{D}^\perp}$ has always a nontrivial kernel. Hence there always exist characteristic curves of $\sigma|_{\mathcal{D}^\perp}$, that correspond to reparametrized abnormal extremals by Proposition 4.34.

Let us consider in more detail the simplest case when the dimension is odd, namely $n = 3$. Assume that there exists a one-form $\omega \in \Lambda^1(M)$ such that $\mathcal{D} = \ker \omega$ (this is not restrictive for a local description) and consider a basis of one-forms $\omega_0, \omega_1, \omega_2$ such that $\omega_0 := \omega$ and the associated coordinates $h_0, h_1, h_2$ on the fibers (see Remark 4.16). By (4.47) one has

$$\sigma|_{\mathcal{D}^\perp} = dh_0 \wedge \omega + h_0\, d\omega, \tag{4.48}$$

Notice that, if the dimension of $M$ is odd, then $\mathcal{D}^\perp$ is even-dimensional. We recall the following criterion for a 2-form to be degenerate on an even-dimensional manifold.

**Lemma 4.36.** *Let $N$ be a smooth $2k$-dimensional manifold and $\Omega \in \Lambda^2 M$. Then $\Omega$ is non-degenerate on $N$ if and only if $\wedge^k \Omega \neq 0$.*[1]

In our specific case, from (4.48), one can easily compute (recall that $\mathcal{D}^\perp$ is 4-dimensional)

$$(\sigma \wedge \sigma)|_{\mathcal{D}^\perp} = 2h_0\, dh_0 \wedge \omega \wedge d\omega. \tag{4.49}$$

We introduce the following definition.

**Definition 4.37.** Let $M$ be a 3-dimensional manifold. We say that a constant rank distribution $\mathcal{D} = \ker \omega$ on $M$ of corank one is a *contact distribution* if $\omega \wedge d\omega \neq 0$.

Collecting the previous results we obtain the following.

[1] Here $\wedge^k \Omega = \underbrace{\Omega \wedge \ldots \wedge \Omega}_{k}$.

**Proposition 4.38.** *Let $M$ be a 3D sub-Riemannian manifold such that $\mathcal{D} = \ker \omega$. Then all nontrivial abnormal extremal trajectories are contained in the Martinet set*

$$\mathfrak{M} = \{q \in M | \, (\omega \wedge d\omega)|_q = 0\}.$$

*In particular, if the sub-Riemannian structure is contact, there is no nontrivial abnormal extremal trajectory.*

*Proof.* By Proposition 4.34 any abnormal extremal $\lambda(t)$ is a characteristic curve of $\sigma|_{\mathcal{D}^\perp}$. By Lemma 4.36 $\sigma|_{\mathcal{D}^\perp}$ is degenerate if and only if $\sigma \wedge \sigma|_{\mathcal{D}^\perp} = 0$.

Since the covector associated to an abnormal extremal is never vanishing, and $h_1 = h_2 = 0$ on $\mathcal{D}^\perp$, it follows that $h_0$ is never vanishing along abnormal extremals. Then, thanks to (4.49), $\sigma \wedge \sigma|_{\mathcal{D}^\perp}$ vanishes at a point $\lambda \in \mathcal{D}^\perp$ if and only if $\omega \wedge d\omega = 0$ at $q = \pi(\lambda)$ (notice that $\omega \wedge d\omega$ depend only on coordinates on the manifold while $h_0$ is a coordinate on the fibers).

This shows that, if $\gamma(t)$ is an abnormal trajectory and $\lambda(t)$ is the associated abnormal extremal, then $\lambda(t)$ is a characteristic curve of $\sigma|_{\mathcal{D}^\perp}$ if and only if $(\omega \wedge d\omega)|_{\gamma(t)} = 0$, that means that $\gamma(t)$ is contained in $\mathfrak{M}$.

If the distribution $\mathcal{D}$ is contact, then it follows directly from Definition 4.37 that the Martinet set $\mathfrak{M}$ is empty. $\square$

*Remark* 4.39. Let $M$ be three dimensional and fix a smooth volume form $dV$ on $M$. We can write $\omega \wedge d\omega = adV$ where $a \in C^\infty(M)$.

The Martinet set is $\mathfrak{M} = a^{-1}(0)$ and the distribution is *contact* if and only if the function $a$ is never vanishing. When 0 is a regular value of $a$, the Martinet set is a two-dimensional smooth manifold. Notice that this condition is satisfied for a generic choice of the (one form defining the) distribution.

Abnormal extremal trajectories then coincides with horizontal curves that are contained in the Martinet set.

When $\mathfrak{M}$ is a smooth surface, the intersection of the tangent bundle to the surface $\mathfrak{M}$ and the 2-dimensional distribution of admissible velocities defines, generically, a line field on $\mathfrak{M}$. Abnormal extremal trajectories coincide, up to a reparametrization, with the integral curves of this line field.

## 4.4 Examples

In this section we consider in detail some examples. First we focus on Riemannian structures on a 2-dimensional manifold. Then we consider sub-Riemannian structures associated with isoperimetric problems, containing the Heisenberg group $\mathbb{H}$ as a particular case.

### 4.4.1 2D Riemannian geometry

Let $M$ be a 2-dimensional Riemannian manifold and let $f_1, f_2 \in \text{Vec}(M)$ be a local orthonormal frame for the Riemannian structure. The problem of finding length-minimizers on $M$ could be described locally as the optimal control problem

$$\dot{q}(t) = u_1(t) f_1(q(t)) + u_2(t) f_2(q(t)),$$

where length and energy are expressed as

$$\ell(q(\cdot)) = \int_0^T \sqrt{u_1(t)^2 + u_2(t)^2}\, dt, \qquad J(q(\cdot)) = \frac{1}{2}\int_0^T \left(u_1(t)^2 + u_2(t)^2\right) dt.$$

Let us start by showing that there is no abnormal extremal. Indeed if $\lambda(t)$ is an abnormal extremal and $\gamma(t)$ is the associated abnormal trajectory we have

$$\langle \lambda(t), f_1(\gamma(t))\rangle = \langle \lambda(t), f_2(\gamma(t))\rangle = 0, \qquad \forall\, t \in [0,T], \tag{4.50}$$

that implies that $\lambda(t) = 0$ for all $t \in [0,T]$ since $\{f_1, f_2\}$ is a basis of the tangent space at every point. This is a contradiction since $\lambda(t) \neq 0$ by Theorem 3.59.

Normal extremal trajectories are projections of integral curves of the sub-Riemannian Hamiltonian in $T^*M$.

$$H(\lambda) = \frac{1}{2}(h_1(\lambda)^2 + h_2(\lambda)^2), \qquad h_i(\lambda) = \langle \lambda, f_i(q)\rangle\,, \quad i = 1,2.$$

Since the vector fields $f_1$ and $f_2$ are linearly independent, the functions $(h_1, h_2)$ defines a system of coordinates on fibers of $T^*M$. In what follows it is convenient to use $(q, h_1, h_2)$ as coordinates on $T^*M$ (notice that we are fixing a set of coordinates on the fibers but not on the base manifold).

Suppose now that $\lambda(t)$ is a normal extremal. Then $u_i(t) = h_i(\lambda(t))$ and the equation on the base manifold is

$$\dot q = h_1 f_1(q) + h_2 f_2(q). \tag{4.51}$$

For the equation on the fibers we have (remember that along solutions $\dot a = \{H, a\}$)

$$\begin{cases} \dot h_1 = \{H, h_1\} = -\{h_1, h_2\} h_2 \\ \dot h_2 = \{H, h_2\} = \{h_1, h_2\} h_1. \end{cases} \tag{4.52}$$

From these equations one can recover by a direct computation that $H$ is constant along solutions. Indeed

$$\dot H = h_1 \dot h_1 + h_2 \dot h_2 = 0.$$

If we require that normal extremals are parametrized by arclength one gets $u_1(t)^2 + u_2(t)^2 = 1$ for a.e. $t \in [0,T]$, and

$$H(\lambda(t)) = \frac{1}{2} \quad \Longleftrightarrow \quad h_1^2(\lambda(t)) + h_2^2(\lambda(t)) = 1.$$

It is then convenient to restrict to the spherical cotangent bundle $S^*M$ (cf. Example 2.55) defined by the equation $h_1^2 + h_2^2 = 1$. It is natural then to introduce an angular coordinate $\theta$ on each fiber by setting

$$h_1 = \cos\theta, \qquad h_2 = \sin\theta. \tag{4.53}$$

Let $c_1, c_2 \in C^\infty(M)$ be such that

$$[f_1, f_2] = c_1 f_1 + c_2 f_2. \tag{4.54}$$

Since $\{h_1, h_2\}(\lambda) = \langle \lambda, [f_1, f_2]\rangle$, we have $\{h_1, h_2\} = c_1 h_1 + c_2 h_2$ and equations (4.51) and (4.52) are rewritten in the $(q, \theta)$ coordinates

$$\begin{cases} \dot\theta = c_1(q)\cos\theta + c_2(q)\sin\theta \\ \dot q = \cos\theta f_1(q) + \sin\theta f_2(q) \end{cases} \tag{4.55}$$

In other words, an arc length parametrized curve on $M$ (namely, satisfying the second equation in (4.55)) is a normal extremal trajectory if and only if it satisfies the first equation of (4.55). Heuristically this suggests that the quantity

$$\dot{\theta} - c_1(q)\cos\theta - c_2(q)\sin\theta,$$

should be related to the geodesic curvature of the trajectory on $M$.

Let $\mu_1, \mu_2$ be the dual frame to $f_1, f_2$ (so that the Riemannian area form on $M$ writes as $dV = \mu_1 \wedge \mu_2$) and consider the Hamiltonian vector field in these coordinates

$$\vec{H} = \cos\theta f_1 + \sin\theta f_2 + (c_1\cos\theta + c_2\sin\theta)\partial_\theta. \tag{4.56}$$

The Levi-Civita connection on $M$ is expressed by some coefficients (cf. (1.29) in Chapter 1)

$$\omega = d\theta + a_1\mu_1 + a_2\mu_2,$$

where $a_i$ are suitable smooth functions on $M$. On the other hand, normal extremal trajectories are projections of integral curves of $\vec{H}$. Moreover

$$\langle \omega, \vec{H} \rangle = 0 \implies a_1 = -c_1, \quad a_2 = -c_2.$$

In particular if we apply $\omega = d\theta - c_1\mu_1 - c_2\mu_2$ to a generic curve on $S^*M$ (not necessarily a geodesic) satisfying

$$\dot{\lambda} = \cos\theta f_1 + \sin\theta f_2 + \dot{\theta}\,\partial_\theta,$$

which projects on $\gamma$ we find the geodesic curvature

$$\kappa_g(\gamma) = \dot{\theta} - c_1\cos\theta - c_2\sin\theta,$$

as we infer above. To end this section we prove a useful formula for the Gaussian curvature of $M$

**Proposition 4.40.** *The Gaussian curvature $\kappa$ of the Riemannian structure on $M$ defined by a local orthonormal frame $f_1, f_2$ is computed by*

$$\kappa = f_1(c_2) - f_2(c_1) - c_1^2 - c_2^2,$$

*where $c_1, c_2$ are the smooth functions satisfying $[f_1, f_2] = c_1 f_1 + c_2 f_2$.*

*Proof.* From (1.72) we have that $d\omega = -\kappa dV$ where $dV = \mu_1 \wedge \mu_2$ is the Riemannian volume form. On the other hand we have in this frame

$$dc_i = f_1(c_i)\mu_1 + f_2(c_i)\mu_2, \quad i = 1, 2.$$

We also can compute (cf. also Cartan's formula (4.84) which is proved later in this chapter)

$$d\mu_i = d\mu_i(f_1, f_2)\mu_1 \wedge \mu_2 = -c_i\mu_1 \wedge \mu_2, \quad i = 1, 2. \tag{4.57}$$

Then we can compute

$$\begin{aligned} d\omega &= d(-c_1\mu_1 - c_2\mu_2) \\ &= -dc_1 \wedge \mu_1 - dc_2 \wedge \mu_2 - c_1 d\mu_1 - c_2 d\mu_2 \\ &= -(f_1(c_2) - f_2(c_1) - c_1^2 - c_2^2)\mu_1 \wedge \mu_2. \end{aligned}$$

from which the claim follows. $\square$

*Remark* 4.41. Notice that the relations (4.57) imply that the one-form $\eta = -c_1\mu_1 - c_2\mu_2$ satisfies

$$d\mu_1 = \eta \wedge \mu_2, \qquad d\mu_2 = -\eta \wedge \mu_1. \tag{4.58}$$

In particular the geodesic curvature of a curve $\gamma$ parametrized by unit speed and satisfying

$$\dot{\gamma} = \cos\theta f_1 + \sin\theta f_2,$$

where $f_1, f_2$ is a local orthonormal frame, is given by $\kappa_g(\gamma) = \dot{\theta} + \eta(\dot{\gamma})$.

### 4.4.2 Isoperimetric problem

Let $M$ be a 2-dimensional orientable Riemannian manifold and denote by $\nu$ its Riemannian volume form. Fix a smooth one-form $A \in \Lambda^1 M$ and $c \in \mathbb{R}$.

**Problem 1.** Fix $c \in \mathbb{R}$ and $q_0, q_1 \in M$. Find, whenever it exists, the solution to

$$\min \left\{ \ell(\gamma) \mid \gamma(0) = q_0, \gamma(T) = q_1, \int_\gamma A = c \right\}. \tag{4.59}$$

*Remark* 4.42. Minimizers depend only on $dA$, i.e., if we add an exact term to $A$ we will find same minima for the problem (with a different value of $c$).

Problem 1 can be reformulated as a sub-Riemannian problem on the extended manifold

$$\overline{M} = M \times \mathbb{R},$$

in the sense that solutions of the problem (4.59) turns to be length-minimizers for a suitable sub-Riemannian structure on $\overline{M}$, that we are going to construct.

With every curve $\gamma$ on $M$ satisfying $\gamma(0) = q_0$ and $\gamma(T) = q_1$ we can associate the function

$$z(t) = \int_{\gamma|_{[0,t]}} A = \int_0^t A(\dot{\gamma}(s))ds.$$

The curve $\zeta(t) = (\gamma(t), z(t))$ defined on $\overline{M}$ satisfies $\omega(\dot{\zeta}(t)) = 0$ where $\omega = dz - A$ is a one form on $\overline{M}$, since

$$\omega(\dot{\zeta}(t)) = \dot{z}(t) - A(\dot{\gamma}(t)) = 0.$$

Equivalently, $\dot{\zeta}(t) \in \mathcal{D}_{\zeta(t)}$ where $\mathcal{D} = \ker \omega$. We define a metric on $\mathcal{D}$ by setting the norm of a vector $v \in \mathcal{D}$ as the Riemannian norm of its projection $\overline{\pi}_* v$ on $M$, where $\overline{\pi} : \overline{M} \to M$ is the projection on the first factor. This endows $\overline{M}$ with a sub-Riemannian structure.

If we fix a local orthonormal frame $f_1, f_2$ for $M$, the pair $(\gamma(t), z(t))$ satisfies

$$\begin{pmatrix} \dot{\gamma} \\ \dot{z} \end{pmatrix} = u_1 \begin{pmatrix} f_1 \\ \langle A, f_1 \rangle \end{pmatrix} + u_2 \begin{pmatrix} f_2 \\ \langle A, f_2 \rangle \end{pmatrix}. \tag{4.60}$$

Hence the two vector fields on $\overline{M}$

$$F_1 = f_1 + \langle A, f_1 \rangle \partial_z, \qquad F_2 = f_2 + \langle A, f_2 \rangle \partial_z, \tag{4.61}$$

constitute an orthonormal frame for the metric defined above on $\mathcal{D} = \mathrm{span}(F_1, F_2)$. Problem 1 is then equivalent to the following:

**Problem 2.** Fix $c \in \mathbb{R}$ and $q_0, q_1 \in M$. Find, whenever it exists, the solution to

$$\min \left\{ \ell(\zeta) \mid \zeta(0) = (q_0, 0), \zeta(T) = (q_1, c), \dot{\zeta}(t) \in \mathcal{D}_{\zeta(t)} \right\}. \tag{4.62}$$

Notice that, by construction, $\mathcal{D}$ is a distribution of constant rank (equal to 2) but is not necessarily bracket-generating.

Let us now compute normal and abnormal Pontryagin extremals associated with the sub-Riemannian structure just introduced on $\overline{M}$. In what follows we denote with $h_i(\lambda) = \langle \lambda, F_i(\zeta) \rangle$, for $i = 1, 2$, the Hamiltonians linear on fibers of $T^*\overline{M}$.

### Normal Pontryagin extremals

Equations of normal Pontryagin extremals are projections of integral curves of the sub-Riemannian Hamiltonian in $T^*\overline{M}$

$$H(\lambda) = \frac{1}{2}(h_1^2(\lambda) + h_2^2(\lambda)), \qquad h_i(\lambda) = \langle \lambda, F_i(\zeta) \rangle, \quad i = 1, 2.$$

Let us introduce $F_0 = \partial_z$ and $h_0(\lambda) = \langle \lambda, F_0(\zeta) \rangle$. Since $F_1, F_2$ and $F_0$ are linearly independent, then $(h_1, h_2, h_0)$ defines a system of coordinates on fibers of $T^*\overline{M}$, hence in what follows we use $(\zeta, h_1, h_2, h_0)$ as coordinates on $T^*\overline{M}$.

For a normal extremal we have $u_i(t) = h_i(\lambda(t))$ for $i = 1, 2$ and the equation on the base is

$$\dot{\zeta} = h_1 F_1(\zeta) + h_2 F_2(\zeta). \tag{4.63}$$

For the equation on the fibers we have (recall that $\dot{a} = \{H, a\}$)

$$\begin{cases} \dot{h}_1 = \{H, h_1\} = -\{h_1, h_2\} h_2 \\ \dot{h}_2 = \{H, h_2\} = \{h_1, h_2\} h_1. \\ \dot{h}_0 = \{H, h_0\} = 0 \end{cases} \tag{4.64}$$

Considering normal Pontryagin extremals parametrized by arclength is equivalent to restrict to the cylinder $H^{-1}(1/2) = \{h_1^2 + h_2^2 = 1\}$ of the cotangent bundle $T^*M$. Thus we can introduce the coordinate $\theta$ by setting

$$h_1 = \cos\theta, \qquad h_2 = \sin\theta.$$

Let $c_1, c_2 \in C^\infty(M)$ be such that

$$[f_1, f_2] = c_1 f_1 + c_2 f_2. \tag{4.65}$$

Then

$$\begin{aligned} [F_1, F_2] &= [f_1 + \langle A, f_1 \rangle \partial_z, f_2 + \langle A, f_2 \rangle \partial_z] \\ &= [f_1, f_2] + (f_1 \langle A, f_2 \rangle - f_2 \langle A, f_1 \rangle) \partial_z \\ \text{(by (4.65))-(4.61)} \qquad &= c_1 (F_1 - \langle A, f_1 \rangle \partial_z) + c_2 (F_2 - \langle A, f_2 \rangle \partial_z) + (f_1 \langle A, f_2 \rangle - f_2 \langle A, f_1 \rangle) \partial_z \\ &= c_1 F_1 + c_2 F_2 + dA(f_1, f_2) \partial_z. \end{aligned}$$

where in the last equality we use Cartan formula (cf. (4.84) for a proof). Let $\mu_1$, $\mu_2$ be the 1-forms on $M$ that are dual to $f_1$ and $f_2$. Then the Riemannian volume form on $M$ is written as $\nu = \mu_1 \wedge \mu_2$ and we can write $dA = b\mu_1 \wedge \mu_2$, for a suitable function $b \in C^\infty(M)$. It follows

$$[F_1, F_2] = c_1 F_1 + c_2 F_2 + b\partial_z,$$

and

$$\{h_1, h_2\} = \langle \lambda, [F_1, F_2] \rangle = c_1 h_1 + c_2 h_2 + b h_0. \tag{4.66}$$

With similar computations to the one performed in Section 4.4.1, we obtain the Hamiltonian system associated with $H$ in the $(\zeta, \theta, h_0)$ coordinates

$$\begin{cases} \dot{\zeta} = \cos\theta F_1(\zeta) + \sin\theta F_2(\zeta) \\ \dot{\theta} = c_1 \cos\theta + c_2 \sin\theta + b h_0 \\ \dot{h}_0 = 0 \end{cases} \tag{4.67}$$

Notice that the coefficients $c_1, c_1, b$ in the last formula depends only on $q$, where $\zeta = (q, z)$.

The projection $\gamma(t) = \overline{\pi}(\zeta(t))$ of a normal extremal path on $M$ (here $\overline{\pi} : \overline{M} \to M$), with geodesic curvature

$$\kappa_g(\gamma(t)) = \dot{\theta}(t) - c_1(\gamma(t)) \cos\theta(t) - c_2(\gamma(t)) \sin\theta(t), \tag{4.68}$$

then satisfies

$$\kappa_g(\gamma(t)) = b(\gamma(t)) h_0. \tag{4.69}$$

Namely, projections on $M$ of normal extremal paths are curves with geodesic curvature proportional to the function $b$ at every point. The case $b$ equal to constant is treated in the example of Section 4.4.3.

### Abnormal Pontryagin extremals

We give an explicit proof of the following fact, which follows also from the discussion of Section 4.3.3.

**Lemma 4.43.** *The projection on $M$ of non constant abnormal extremal trajectories on $\overline{M}$ is contained in the set $b^{-1}(0)$.*

*Proof.* Assume that $\lambda(t)$ is an abnormal extremal whose projection is a curve $\zeta(t) = \pi(\lambda(t))$ that is not reduced to a point. Then we have

$$h_1(\lambda(t)) = \langle \lambda(t), F_1(\zeta(t)) \rangle = 0, \qquad h_2(\lambda(t)) = \langle \lambda(t), F_2(\zeta(t)) \rangle = 0, \qquad \forall\, t \in [0, T], \tag{4.70}$$

We can differentiate the two equalities with respect to $t \in [0, T]$ and we get

$$\begin{aligned} \frac{d}{dt} h_1(\lambda(t)) &= u_2(t) \{h_1, h_2\}|_{\lambda(t)} = 0 \\ \frac{d}{dt} h_2(\lambda(t)) &= -u_1(t) \{h_1, h_2\}|_{\lambda(t)} = 0 \end{aligned}$$

Since the pair $(u_1(t), u_2(t)) \neq (0, 0)$ we have that $\{h_1, h_2\}|_{\lambda(t)} = 0$ that implies

$$0 = \langle \lambda(t), [F_1, F_2](\zeta(t)) \rangle = b(\gamma(t)) h_0, \tag{4.71}$$

where in the last equality $\gamma(t) = \overline{\pi}(\zeta(t))$ and we used (4.66) and the fact that $h_1(\lambda(t)) = h_2(\lambda(t)) = 0$. Recall that $h_0 \neq 0$ otherwise the covector is identically zero (that is not possible for abnormals), then $b(\gamma(t)) = 0$ for all $t \in [0, T]$, where $\gamma(t) = \overline{\pi}(\zeta(t))$. $\square$

The last result shows that abnormal extremal trajectories are forced to live in connected components of $b^{-1}(0)$.

**Exercise 4.44.** Prove that the set $b^{-1}(0)$ is *independent* on the Riemannian metric chosen on $M$ (and the corresponding sub-Riemannian metric defined on $\mathcal{D}$).

### 4.4.3 Heisenberg group

The *Heisenberg group* $\mathbb{H}$ is a basic example in sub-Riemannian geometry. It is the sub-Riemannian structure defined by the isoperimetric problem in $M = \mathbb{R}^2 = \{(x, y)\}$ endowed with its Euclidean scalar product and the 1-form (cf. previous section)

$$A = \frac{1}{2}(xdy - ydx).$$

Notice that $dA = dx \wedge dy$ defines the area form on $\mathbb{R}^2$, hence $b \equiv 1$ in this case. On the extended manifold $\overline{M} = \mathbb{R}^3 = \{(x, y, z)\}$ the one-form $\omega$ is written as

$$\omega = dz - \frac{1}{2}(xdy - ydx)$$

Following the notation of the previous paragraph we can choose as an orthonormal frame for $\mathbb{R}^2$ the frame $f_1 = \partial_x$ and $f_2 = \partial_y$. This induces the choice

$$F_1 = \partial_x - \frac{y}{2}\partial_z, \qquad F_2 = \partial_y + \frac{x}{2}\partial_z.$$

for the orthonormal frame on $\mathcal{D} = \ker \omega$. Notice that $[F_1, F_2] = \partial_z$, that implies that $\mathcal{D}$ is bracket-generating at every point. Defining $F_0 = \partial_z$ and $h_i = \langle \lambda, F_i(q) \rangle$ for $i = 0, 1, 2$, the Hamiltonians linear on fibers of $T^*\overline{M}$, we have

$$\{h_1, h_2\} = h_0,$$

hence the equation (4.67) for normal extremals become (here $q = (x, y, z)$)

$$\begin{cases} \dot{q} = \cos\theta F_1(q) + \sin\theta F_2(q) \\ \dot{\theta} = h_0 \\ \dot{h}_0 = 0 \end{cases} \tag{4.72}$$

It follows that the two last equations can be immediately solved

$$\begin{cases} \theta(t) = \theta_0 + h_0 t \\ h_0(t) = h_0 \end{cases} \tag{4.73}$$

Moreover

$$\begin{cases} h_1(t) = \cos(\theta_0 + h_0 t) \\ h_2(t) = \sin(\theta_0 + h_0 t) \end{cases} \tag{4.74}$$

From these formulas and the explicit expression of $F_1$ and $F_2$ it is immediate to recover the normal extremal trajectories starting from the origin ($x_0 = y_0 = z_0 = 0$) in the case $h_0 \neq 0$

$$x(t) = \frac{1}{h_0}(\sin(\theta_0 + h_0 t) - \sin(\theta_0)), \qquad y(t) = -\frac{1}{h_0}(\cos(\theta_0 + h_0 t) - \cos(\theta_0)), \tag{4.75}$$

and the vertical coordinate $z$ is computed as the integral

$$z(t) = \frac{1}{2}\int_0^t x(s)\dot{y}(s) - y(s)\dot{x}(s)ds = \frac{1}{2h_0^2}(h_0 t - \sin(h_0 t)).$$

When $h_0 = 0$ the curve is simply a straight line

$$x(t) = \cos(\theta_0)t, \qquad y(t) = \sin(\theta_0)t, \qquad z(t) = 0. \tag{4.76}$$

Notice that, as we know from the results of the previous subsection, normal extremal trajectories are curves whose projection on $\mathbb{R}^2 = \{(x, y)\}$ has constant geodesic curvature, i.e., straight lines or circles on $\mathbb{R}^2$ (that correspond to horizontal lines and helix on $\overline{M}$).

We remark finally that there is no non trivial abnormal extremal trajectory since $b = 1$.

*Remark* 4.45. This sub-Riemannian structure on $\mathbb{R}^3$ is called Heisenberg group since it can be seen as a left-invariant structure on a Lie group, see also Section 7.5.

## 4.5 Lie derivative

In this section we extend the notion of Lie derivative, already introduced for vector fields in Section 3.2, to differential forms. Recall that if $X, Y \in \mathrm{Vec}(M)$ are two vector fields, their Lie derivative is defined as

$$\mathcal{L}_X Y = [X, Y] = \left.\frac{d}{dt}\right|_{t=0} e_*^{-tX} Y.$$

If $P : M \to M$ is a diffeomorphism we can consider the pullback $P^* : T^*_{P(q)}M \to T^*_q M$ and extend its action to $k$-forms. For $\omega \in \Lambda^k M$, one defines $P^*\omega \in \Lambda^k M$ in the following way:

$$(P^*\omega)_q(\xi_1, \dots, \xi_k) := \omega_{P(q)}(P_*\xi_1, \dots, P_*\xi_k), \quad q \in M, \quad \xi_i \in T_q M. \tag{4.77}$$

It is easy to check that this operation is linear and satisfies the two following properties

$$P^*(\omega_1 \wedge \omega_2) = P^*\omega_1 \wedge P^*\omega_2, \tag{4.78}$$

$$P^* \circ d = d \circ P^*. \tag{4.79}$$

**Definition 4.46.** Let $X \in \mathrm{Vec}(M)$ and $\omega \in \Lambda^k M$, where $k \geq 0$. We define the *Lie derivative* of $\omega$ with respect to $X$ as the operator

$$\mathcal{L}_X : \Lambda^k M \to \Lambda^k M, \qquad \mathcal{L}_X \omega = \left.\frac{d}{dt}\right|_{t=0} (e^{tX})^*\omega. \tag{4.80}$$

For $k = 0$ this definition coincides with the Lie derivative of smooth functions, namely $\mathcal{L}_X f = Xf$, for $f \in C^\infty(M)$. From (4.78) and (4.79), one easily deduces the following properties of the Lie derivative:

(i) $\mathcal{L}_X(\omega_1 \wedge \omega_2) = (\mathcal{L}_X\omega_1) \wedge \omega_2 + \omega_1 \wedge (\mathcal{L}_X\omega_2)$,

(ii) $\mathcal{L}_X \circ d = d \circ \mathcal{L}_X$.

Property (i) can be also expressed by saying that $\mathcal{L}_X$ is a *derivation* of the exterior algebra of $k$-forms.

The Lie derivative of a $k$-form along a vector field defines a new $k$-form. Given a $k$-form and a vector field, one can also introduce their inner product, defining a $(k-1)$-form as follows.

**Definition 4.47.** Let $X \in \mathrm{Vec}(M)$ and $\omega \in \Lambda^k M$, with $k \geq 1$. We define the *inner product of $\omega$ and $X$* as the operator $i_X : \Lambda^k M \to \Lambda^{k-1} M$, such that

$$(i_X\omega)(Y_1, \ldots, Y_{k-1}) := \omega(X, Y_1, \ldots, Y_{k-1}), \quad Y_i \in \mathrm{Vec}(M). \tag{4.81}$$

The operator $i_X$ is an *anti-derivation*, in the following sense:

$$i_X(\omega_1 \wedge \omega_2) = (i_X\omega_1) \wedge \omega_2 + (-1)^{k_1}\omega_1 \wedge (i_X\omega_2), \quad \omega_i \in \Lambda^{k_i} M, \quad i = 1, 2. \tag{4.82}$$

We end this section proving two classical formulas, usually referred as Cartan's formulas.

**Proposition 4.48** (Cartan's formula)**.** *Let $X \in \mathrm{Vec}(M)$. The following identity holds true*

$$\mathcal{L}_X = i_X \circ d + d \circ i_X. \tag{4.83}$$

*Proof.* Set $D_X := i_X \circ d + d \circ i_X$. It is easy to check that $D_X$ is a derivation on the algebra of $k$-forms, since $i_X$ and $d$ are anti-derivations. Let us show that $D_X$ commutes with $d$. Indeed, using the fact that $d^2 = 0$, one gets

$$d \circ D_X = d \circ i_X \circ d = D_X \circ d.$$

Since any $k$-form can be expressed in coordinates as $\omega = \sum \omega_{i_1 \ldots i_k} dx_{i_1} \ldots dx_{i_k}$, it is sufficient to prove that $\mathcal{L}_X$ coincide with $D_X$ on functions. This last property is easily verified, since

$$D_X f = i_X(df) + \underbrace{d(i_X f)}_{=0} = \langle df, X \rangle = Xf = \mathcal{L}_X f.$$

□

**Corollary 4.49.** *Let $X, Y \in \mathrm{Vec}(M)$ and $\omega \in \Lambda^1 M$, then*

$$d\omega(X, Y) = X\langle \omega, Y \rangle - Y\langle \omega, X \rangle - \langle \omega, [X, Y] \rangle. \tag{4.84}$$

*Proof.* On one hand Definition 4.46 implies, by Leibniz rule

$$\begin{aligned} \langle \mathcal{L}_X\omega, Y \rangle_q &= \left.\frac{d}{dt}\right|_{t=0} \left\langle (e^{tX})^*\omega, Y \right\rangle_q \\ &= \left.\frac{d}{dt}\right|_{t=0} \left\langle \omega, e^{tX}_* Y \right\rangle_{e^{tX}(q)} \\ &= X\langle \omega, Y \rangle - \langle \omega, [X, Y] \rangle. \end{aligned}$$

On the other hand, Cartan's formula (4.83) gives

$$\begin{aligned} \langle \mathcal{L}_X\omega, Y \rangle &= \langle i_X(d\omega), Y \rangle + \langle d(i_X\omega), Y \rangle \\ &= d\omega(X, Y) + Y\langle \omega, X \rangle. \end{aligned}$$

Comparing the two identities one gets (4.84). □

**Exercise 4.50.** Prove the following Leibniz rule formula: for $X \in \mathrm{Vec}(M)$, $\omega \in \Lambda^k M$, and $f \in C^\infty(M)$

$$\mathcal{L}_{fX}\omega = f\mathcal{L}_X\omega + df \wedge i_X\omega \tag{4.85}$$

## 4.6 Symplectic manifolds

In this section we generalize some of the constructions we considered on the cotangent bundle $T^*M$ to the case of a general symplectic manifold.

**Definition 4.51.** A *symplectic manifold* $(N, \sigma)$ is a smooth manifold $N$ endowed with a closed, non degenerate 2-form $\sigma \in \Lambda^2(N)$. A *symplectomorphism* of $N$ is a diffeomorphism $\phi : N \to N$ such that $\phi^*\sigma = \sigma$.

Notice that, thanks to the non-degeneracy assumption on the symplectic form, a symplectic manifold $N$ is necessarily even-dimensional. We stress that, on a general symplectic manifold, the symplectic form $\sigma$ is not exact.

The symplectic structure on a symplectic manifold $N$ permits us to define the *Hamiltonian vector field* $\vec{h} \in \mathrm{Vec}(N)$ associated with a function $h \in C^\infty(N)$ by the formula $i_{\vec{h}}\sigma = -dh$, or equivalently $\sigma(\cdot, \vec{h}) = dh$.

**Proposition 4.52.** *A diffeomorphism $\phi : N \to N$ is a symplectomorphism if and only if for every $h \in C^\infty(N)$:*

$$(\phi_*^{-1})\vec{h} = \overrightarrow{h \circ \phi}. \tag{4.86}$$

*Proof.* Assume that $\phi$ is a symplectomorphism, namely $\phi^*\sigma = \sigma$. More precisely, this means that for every $\lambda \in N$ and every $v, w \in T_\lambda N$ one has

$$\sigma_\lambda(v, w) = (\phi^*\sigma)_\lambda(v, w) = \sigma_{\phi(\lambda)}(\phi_* v, \phi_* w),$$

where the second equality is the definition of $\phi^*\sigma$. If we apply the above equality at $w = \phi_*^{-1}\vec{h}$ one gets, for every $\lambda \in N$ and $v \in T_\lambda N$

$$\begin{aligned} \sigma_\lambda(v, \phi_*^{-1}\vec{h}) &= (\phi^*\sigma)_\lambda(v, \phi_*^{-1}\vec{h}) = \sigma_{\phi(\lambda)}(\phi_* v, \vec{h}) \\ &= \left\langle d_{\phi(\lambda)}h, \phi_* v \right\rangle = \left\langle \phi^* d_{\phi(\lambda)}h, v \right\rangle. \\ &= \langle d(h \circ \phi), v \rangle \end{aligned}$$

This shows that $\sigma_\lambda(\cdot, \phi_*^{-1}\vec{h}) = d(h \circ \phi)$, that is (4.86). The converse implication follows analogously. □

Next we want to characterize those vector fields whose flow generates a one-parametric family of symplectomorphisms.

**Lemma 4.53.** *Let $X \in \mathrm{Vec}(N)$ be a complete vector field on a symplectic manifold $(N, \sigma)$. The following properties are equivalent*

*(i)* $(e^{tX})^*\sigma = \sigma$ *for every* $t \in \mathbb{R}$,

*(ii)* $\mathcal{L}_X\sigma = 0$,

*(iii)* $i_X\sigma$ *is a closed 1-form on* $N$.

*Proof.* By the group property $e^{(t+s)X} = e^{tX} \circ e^{sX}$ one has the following identity for every $t \in \mathbb{R}$:

$$\frac{d}{dt}(e^{tX})^*\sigma = \frac{d}{ds}\bigg|_{s=0}(e^{tX})^*(e^{sX})^*\sigma = (e^{tX})^*\mathcal{L}_X\sigma.$$

This proves the equivalence between (i) and (ii), since the map $(e^{tX})^*$ is invertible for every $t \in \mathbb{R}$.

Recall now that the symplectic form $\sigma$ is, by definition, a closed form. Then $d\sigma = 0$ and Cartan's formula (4.83) reads as follows

$$\mathcal{L}_X\sigma = d(i_X\sigma) + i_X(d\sigma) = d(i_X\sigma),$$

which proves the the equivalence between (ii) and (iii). $\square$

**Corollary 4.54.** *The flow of a Hamiltonian vector field is a one-parameter family of symplectomorphisms.*

*Proof.* This is a direct consequence of the fact that, for a Hamitonian vector field $\vec{h}$, one has $i_{\vec{h}}\sigma = -dh$. Hence $i_{\vec{h}}\sigma$ is a closed form (actually exact) and property (iii) of Lemma 4.53 holds. $\square$

Notice that the converse of Corollary 4.54 is true when $N$ is simply connected, since in this case every closed form is exact.

**Definition 4.55.** Let $(N, \sigma)$ be a symplectic manifold and $a, b \in C^\infty(N)$. The Poisson bracket between $a$ and $b$ is defined as $\{a, b\} = \sigma(\vec{a}, \vec{b})$.

We end this section by collecting some properties of the Poisson bracket that follow from the previous results.

**Proposition 4.56.** *The Poisson bracket satisfies the identities*

*(i)* $\{a, b\} \circ \phi = \{a \circ \phi, b \circ \phi\}, \qquad \forall\, a, b \in C^\infty(N), \forall\, \phi \in \mathrm{Sympl}(N)$,

*(ii)* $\{a, \{b, c\}\} + \{c, \{a, b\}\} + \{b, \{c, a\}\} = 0, \qquad \forall\, a, b, c \in C^\infty(N)$.

*Proof.* Property (i) follows from (4.86). Property (ii) follows by considering $\phi = e^{t\vec{c}}$ in (i), for some $c \in C^\infty(N)$, and computing the derivative with respect to $t$ at $t = 0$. $\square$

**Corollary 4.57.** *For every* $a, b \in C^\infty(N)$ *we have*

$$\overrightarrow{\{a, b\}} = [\vec{a}, \vec{b}]. \tag{4.87}$$

*Proof.* Property (ii) of Proposition 4.56 can be rewritten, by skew-symmetry of the Poisson bracket, as follows

$$\{\{a, b\}, c\} = \{a, \{b, c\}\} - \{b, \{a, c\}\}. \tag{4.88}$$

Using that $\{a, b\} = \sigma(\vec{a}, \vec{b}) = \vec{a}b$ one rewrite (4.88) as

$$\overrightarrow{\{a, b\}}c = \vec{a}(\vec{b}c) - \vec{b}(\vec{a}c) = [\vec{a}, \vec{b}]c. \qquad \square$$

*Remark* 4.58. Property (ii) of Proposition 4.56 says that $\{a, \cdot\}$ is a derivation of the algebra $C^\infty(N)$. Moreover, the space $C^\infty(N)$ endowed with $\{\cdot, \cdot\}$ as a product is a Lie algebra isomorphic to a subalgebra of $\mathrm{Vec}(N)$. Indeed, by (4.87), the correspondence $a \mapsto \vec{a}$ is a Lie algebra homomorphism between $C^\infty(N)$ and $\mathrm{Vec}(N)$.

## 4.7 Local minimality of normal extremal trajectories

In this section we prove a fundamental result about local optimality of normal trajectories. More precisely we show that small pieces of a normal trajectory are length minimizers.

### 4.7.1 The Poincaré-Cartan one-form

Fix a smooth function $a \in C^\infty(M)$ and consider the smooth submanifold of $T^*M$ defined by the graph of its differential

$$\mathcal{L}_0 = \{d_q a \,|\, q \in M\} \subset T^*M. \tag{4.89}$$

Notice that the restriction of the canonical projection $\pi : T^*M \to M$ to $\mathcal{L}_0$ defines a diffeomorphism between $\mathcal{L}_0$ and $M$, hence $\dim \mathcal{L}_0 = n$. Assume that the Hamiltonian flow is complete and consider the image of $\mathcal{L}_0$ under the Hamiltonian flow

$$\mathcal{L}_t := e^{t\vec{H}}(\mathcal{L}_0), \qquad t \in [0, T]. \tag{4.90}$$

Define the $(n+1)$-dimensional manifold with boundary in $\mathbb{R} \times T^*M$ as follows

$$\begin{aligned} \mathcal{L} &= \{(t, \lambda) \in \mathbb{R} \times T^*M \,|\, \lambda \in \mathcal{L}_t,\, 0 \leq t \leq T\} && (4.91)\\ &= \{(t, e^{t\vec{H}}\lambda_0) \in \mathbb{R} \times T^*M \,|\, \lambda_0 \in \mathcal{L}_0,\, 0 \leq t \leq T\}. && (4.92) \end{aligned}$$

Finally, let us introduce the *Poincaré-Cartan* 1-form on $T^*M \times \mathbb{R}$ defined by

$$s - Hdt \in \Lambda^1(T^*M \times \mathbb{R})$$

where $s \in \Lambda^1(T^*M)$ denotes, as usual, the tautological 1-form of $T^*M$. We start by proving a preliminary lemma.

**Lemma 4.59.** $s|_{\mathcal{L}_0} = d(a \circ \pi)|_{\mathcal{L}_0}$

*Proof.* By definition of tautological 1-form $s_\lambda(w) = \langle \lambda, \pi_* w\rangle$, for every $w \in T_\lambda(T^*M)$. If $\lambda \in \mathcal{L}_0$ then $\lambda = d_q a$, where $q = \pi(\lambda)$. Hence for every $w \in T_\lambda(T^*M)$

$$s_\lambda(w) = \langle \lambda, \pi_* w\rangle = \langle d_q a, \pi_* w\rangle = \langle \pi^* d_q a, w\rangle = \langle d_q(a \circ \pi), w\rangle. \qquad \square$$

**Proposition 4.60.** *The 1-form $(s - Hdt)|_{\mathcal{L}}$ is exact.*

*Proof.* We divide the proof in two steps: (i) we show that the restriction of the Poincare-Cartan 1-form $(s - Hdt)|_{\mathcal{L}}$ is closed and (ii) that it is exact.

(i). To prove that the 1-form is closed we need to show that the differential

$$d(s - Hdt) = \sigma - dH \wedge dt, \tag{4.93}$$

vanishes when applied to every pair of tangent vectors to $\mathcal{L}$. Since, for each $t \in [0, T]$, the set $\mathcal{L}_t$ has codimension 1 in $\mathcal{L}$, there are only two possibilities for the choice of the two tangent vectors:

(a) both vectors are tangent to $\mathcal{L}_t$, for some $t \in [0, T]$.

(b) one vector is tangent to $\mathcal{L}_t$ while the second one is transversal.

Case (a). Since both tangent vectors are tangent to $\mathcal{L}_t$, it is enough to show that the restriction of the one form $\sigma - dH \wedge dt$ to $\mathcal{L}_t$ is zero. First let us notice that $dt$ vanishes when applied to tangent vectors to $\mathcal{L}_t$, thus $\sigma - dH \wedge dt|_{\mathcal{L}_t} = \sigma|_{\mathcal{L}_t}$. Moreover, since by definition $\mathcal{L}_t = e^{t\vec{H}}(\mathcal{L}_0)$ one has

$$\begin{aligned}\sigma|_{\mathcal{L}_t} &= \sigma|_{e^{t\vec{H}}(\mathcal{L}_0)} \\ &= (e^{t\vec{H}})^*\sigma|_{\mathcal{L}_0} = \sigma|_{\mathcal{L}_0} = ds|_{\mathcal{L}_0} = d^2(a\circ\pi)|_{\mathcal{L}_0} = 0.\end{aligned}$$

where in the last line we used Lemma 4.59 and the fact that $(e^{t\vec{H}})^*\sigma = \sigma$, since $e^{t\vec{H}}$ is an Hamiltonian flow and thus preserves the symplectic form.
Case (b). The manifold $\mathcal{L}$ is, by construction, the image of the smooth mapping

$$\Psi : [0,T]\times\mathcal{L}_0 \to [0,T]\times T^*M, \qquad \Psi(t,\lambda)\mapsto (t, e^{t\vec{H}}\lambda),$$

Thus a tangent vector to $\mathcal{L}$ that is transversal to $\mathcal{L}_t$ can be obtained by differentiating the map $\Psi$ with respect to $t$:

$$\frac{\partial\Psi}{\partial t}(t,\lambda) = \frac{\partial}{\partial t} + \vec{H}(\lambda) \in T_{(t,\lambda)}\mathcal{L}. \tag{4.94}$$

It is then sufficient to show that the vector (4.94) is in the kernel of the two form $\sigma - dH\wedge dt$. In other words we have to prove

$$i_{\partial_t+\vec{H}}(\sigma - dH\wedge dt) = 0. \tag{4.95}$$

The last equality is a consequence of the following identities

$$\begin{aligned}&i_{\vec{H}}\sigma = \sigma(\vec{H},\cdot) = -dH, \qquad i_{\partial_t}\sigma = 0,\\ &i_{\vec{H}}(dH\wedge dt) = \underbrace{(i_{\vec{H}}dH)}_{=0}\wedge dt - dH\wedge\underbrace{(i_{\vec{H}}dt)}_{=0} = 0,\\ &i_{\partial_t}(dH\wedge dt) = \underbrace{(i_{\partial_t}dH)}_{=0}\wedge dt - dH\wedge\underbrace{(i_{\partial_t}dt)}_{=1} = -dH.\end{aligned}$$

where we used that $i_{\vec{H}}dH = dH(\vec{H}) = \{H,H\} = 0$.

(ii). Next we show that the form $s - Hdt|_{\mathcal{L}}$ is exact. To this aim we have to prove that, for every closed curve $\Gamma$ in $\mathcal{L}$ one has

$$\int_\Gamma s - Hdt = 0. \tag{4.96}$$

Every curve $\Gamma$ in $\mathcal{L}$ can be written as follows

$$\Gamma : [0,T]\to\mathcal{L}, \quad \Gamma(s) = (t(s), e^{t(s)\vec{H}}\lambda(s)), \qquad \text{where } \lambda(s)\in\mathcal{L}_0.$$

Moreover, it is easy to see that the continuous map defined by

$$K : [0,1]\times\mathcal{L}\to\mathcal{L}, \qquad K(\tau,(t,e^{t\vec{H}}\lambda_0)) = (t\tau, e^{t\tau\vec{H}}\lambda_0)$$

defines an homotopy on $\mathcal{L}$ such that $K(1,(t,e^{t\vec{H}}\lambda_0)) = (t,e^{t\vec{H}}\lambda_0)$ and $K(0,(t,e^{t\vec{H}}\lambda_0)) = (0,\lambda_0)$. Then the curve $\Gamma$ is homotopic to the curve $\Gamma_0(s) = (0,\lambda(s))$. Since the 1-form $s - Hdt$ is closed, the integral is invariant under homotopy, namely

$$\int_\Gamma s - Hdt = \int_{\Gamma_0} s - Hdt.$$

Moreover, the integral over $\Gamma_0$ is computed as follows (recall that $\Gamma_0 \subset \mathcal{L}_0$ and $dt = 0$ on $\mathcal{L}_0$):

$$\int_{\Gamma_0} s - H dt = \int_{\Gamma_0} s = \int_{\Gamma_0} d(a \circ \pi) = 0,$$

where we used Lemma 4.59 and the fact that the integral of an exact form over a closed curve is zero. Then (4.96) follows. $\square$

### 4.7.2 Normal Pontryagin extremal trajectories are geodesics

Now we are ready to prove a sufficient condition that ensures the optimality of small arcs of normal Pontryagin extremal trajectories. As a corollary we will get that normal extremal trajectories are geodesics in the following sense.

**Definition 4.61.** An admissible trajectory $\gamma : [0, T] \to M$ is called a *geodesic* if it is parametrized by non-zero constant speed and for every $t \in [0, T]$ there exists a neighborhood $I$ of $t$ in $[0, T]$ such that $\ell(\gamma|_I)$ is equal to the distance between its end-points.

Let $f_1, \dots, f_m$ be a generating family for the sub-Riemannian structure. Recall that normal trajectories for the sub-Riemannian structure

$$\dot{q} = f_u(q) = \sum_{i=1}^{m} u_i f_i(q), \tag{4.97}$$

are projections of integral curves of the Hamiltonian vector field associated with the sub-Riemannian Hamiltonian

$$\dot{\lambda}(t) = \vec{H}(\lambda(t)), \qquad \text{(i.e., } \lambda(t) = e^{t\vec{H}}(\lambda_0)), \tag{4.98}$$

$$\gamma(t) = \pi(\lambda(t)), \qquad t \in [0, T]. \tag{4.99}$$

where

$$H(\lambda) = \max_{u \in U_q} \left\{ \langle \lambda, f_u(q) \rangle - \frac{1}{2}|u|^2 \right\} = \frac{1}{2} \sum_{i=1}^{m} \langle \lambda, f_i(q) \rangle^2 . \tag{4.100}$$

Recall that, given a smooth function $a \in C^\infty(M)$, we can consider the image of its differential $\mathcal{L}_0$ and its evolution $\mathcal{L}_t$ under the Hamiltonian flow associated with $H$ as in (4.89) and (4.90).

**Theorem 4.62.** *Assume that there exists $a \in C^\infty(M)$ such that the restriction of the projection $\pi|_{\mathcal{L}_t}$ is a diffeomorphism for every $t \in [0, T]$. Fix $\overline{\lambda}_0 \in \mathcal{L}_0$, then the normal extremal trajectory*

$$\overline{\gamma}(t) = \pi \circ e^{t\vec{H}}(\overline{\lambda}_0), \qquad t \in [0, T], \tag{4.101}$$

*is a strict length-minimizer among all admissible curves $\gamma$ with the same initial and final points (up to reparametrization).*

The previous Theorem is a consequence of the following more general argument.

**Proposition 4.63.** *Under the assumptions of Theorem 4.62, the curve $\overline{\gamma}$ is a strict minimizer of the energy functional with penalty defined by*

$$J_a(\gamma) := a(\gamma(0)) + J(\gamma),$$

*among all admissible curves $\gamma$ with the same final point.*

*Proof of Proposition 4.63.* Let $\gamma$ be an admissible trajectory, different from $\overline{\gamma}$, associated with the control $u(\cdot)$ and such that $\gamma(T) = \overline{\gamma}(T)$. We denote by $\overline{u}(\cdot)$ the control associated with the curve $\overline{\gamma}$ and by $\bar{\lambda}(\cdot)$ the corresponding covector, i.e., $\bar{\lambda}(t) = e^{t\vec{H}}(\overline{\lambda}_0)$.

By assumption, for every $t \in [0,T]$ the map $\pi|_{\mathcal{L}_t} : \mathcal{L}_t \to M$ is a local diffeomorphism, thus the trajectory $\gamma(t)$ can be uniquely lifted to a smooth curve $\lambda(t) \in \mathcal{L}_t$, for $t \in [0,T]$. Notice that the corresponding curves $\Gamma$ and $\overline{\Gamma}$ in $\mathcal{L}$ defined by

$$\Gamma(t) = (t, \lambda(t)), \qquad \overline{\Gamma}(t) = (t, \overline{\lambda}(t)) \tag{4.102}$$

have the same final conditions, since for $t = T$ they project to the same base point on $M$ and their lifts are uniquely determined by the diffeomorphism $\pi|_{\mathcal{L}_T}$.

Now, recall that, by definition of the sub-Riemannian Hamiltonian, we have

$$H(\lambda(t)) \geq \left\langle \lambda(t), f_{u(t)}(\gamma(t)) \right\rangle - \frac{1}{2}|u(t)|^2, \text{ for a.e. } t \in [0,T], \qquad \gamma(t) = \pi(\lambda(t)), \tag{4.103}$$

We claim that the inequality in (4.103) is strict on a set of positive measure in $[0,T]$, since $\gamma$ is different from $\overline{\gamma}$.

Indeed, assume by contradiction that in (4.103) we have equality for a.e. $t \in [0,T]$. Then by uniqueness of the maximum we have

$$u_i(t) = \langle \lambda(t), f_i(\gamma(t) \rangle =: h_i(\lambda(t)) \text{ for a.e. } t \in [0,T], \quad i = 1, \dots, m. \tag{4.104}$$

Let now $\lambda_0(t)$ be defined by $\lambda(t) = e^{t\vec{H}}(\lambda_0(t))$. Then $\gamma(t) = \pi \circ e^{t\vec{H}}(\lambda_0(t))$. Let us compute $\dot{\gamma}(t)$. On one side we have

$$\dot{\gamma}(t) = \pi_* \vec{H}(\lambda(t)) + \pi_* e^{t\vec{H}}_* \dot{\lambda}_0(t) = \sum_{i=1}^{m} h_i(\lambda(t)) f_i(\gamma(t)) + \pi_* e^{t\vec{H}}_* \dot{\lambda}_0(t).$$

On the other side $\dot{\gamma}(t) = \sum_i^m u_i(t) f_i(\gamma(t))$. Using (4.104) it follows $\dot{\lambda}_0 \equiv 0$ a.e. and, as a consequence, $\lambda_0(t) = \lambda_0(0) = \lambda_0(T)$ for every $t \in [0,T]$. It follows that $\lambda(\cdot)$ and $\bar{\lambda}(\cdot)$ are solutions of the same Hamiltonian system with the same final condition. Hence $\lambda(t) = \bar{\lambda}(t)$ for every $t \in [0,T]$ and $\gamma$ and $\overline{\gamma}$ coincide as well. Contradiction. The claim is proved.

Consider now a trajectory $\delta : [0,1] \to M$ on the manifold such that $\delta(0) = \overline{\gamma}(0)$ and $\delta(1) = \gamma(0)$. Let $\Delta(s) = (0, \eta(s))$ where $\eta(s)$ is the unique lift of $\delta(s)$ on $\mathcal{L}_0$. The concatenation of $\Gamma, \overline{\Gamma}$ and $\Delta$ form a closed curve in $T^*M \times \mathbb{R}$. By Lemma 4.60, the 1-form $s - Hdt$ is exact. Then the integral over a closed curve is zero and one gets

$$\int_{\overline{\Gamma}} s - Hdt = \int_{\Delta} s - Hdt + \int_{\Gamma} s - Hdt. \tag{4.105}$$

Let us compute each term in (4.105). First notice that, when restricted to $\mathcal{L}_0$, the 1-form $s - Hdt$ coincides with the differential of $a$, as shown in Lemma 4.59. Hence

$$\int_{\Delta} s - Hdt = \int_{\Delta} d(a \circ \pi) = a(\delta(1)) - a(\delta(0)) = a(\gamma(0)) - a(\overline{\gamma}(0)). \tag{4.106}$$

The other two terms can be computed as follows

$$
\begin{aligned}
\int_\Gamma s - H dt &= \int_0^T \langle \lambda(t), \dot\gamma(t)\rangle - H(\lambda(t))dt = \int_0^T \left\langle \lambda(t), f_{u(t)}(\gamma(t))\right\rangle - H(\lambda(t))dt \\
&< \int_0^T \left\langle \lambda(t), f_{u(t)}(\gamma(t))\right\rangle - \left(\left\langle \lambda(t), f_{u(t)}(\gamma(t))\right\rangle - \frac{1}{2}|u(t)|^2\right) dt \\
&= \frac{1}{2}\int_0^T |u(t)|^2 dt.
\end{aligned} \tag{4.107}
$$

where we used (4.103) with strict inequality in a set of positive measure. A similar computation for $\overline{\gamma}$, using the fact that now $H(\bar\lambda(t)) = \left\langle \bar\lambda(t), f_{\bar u(t)}(\bar\gamma(t))\right\rangle - \frac{1}{2}|\overline{u}(t)|^2$, for a.e. $t \in [0,T]$, gives

$$
\int_{\overline{\Gamma}} s - H dt = \frac{1}{2}\int_0^T |\overline{u}(t)|^2 dt, \tag{4.108}
$$

Combining (4.105) and the previous computations we have

$$
\frac{1}{2}\int_0^T |\overline{u}(t)|^2 dt < a(\gamma(0)) - a(\overline{\gamma}(0)) + \frac{1}{2}\int_0^T |u(t)|^2 dt, \tag{4.109}
$$

which is equivalent to

$$
J_a(\overline{\gamma}) = a(\overline{\gamma}(0)) + \frac{1}{2}\int_0^T |\overline{u}(t)|^2 dt < a(\gamma(0)) + \frac{1}{2}\int_0^T |u(t)|^2 dt = J_a(\gamma). \tag{4.110}
$$

□

As a corollary we state a local version of the same theorem, that can be proved by adapting the above technique.

**Corollary 4.64.** *Assume that there exists a smooth function $a \in C^\infty(M)$ and an open set $\Omega_0$ such that $\pi \circ e^{t\vec H} \circ da|_{\Omega_0}$ is a diffeomorphism onto its image (i.e., from $\Omega_0$ to $\Omega_t = \pi \circ e^{t\vec H} \circ da|_{\Omega_0}$) for every $t \in [0,T]$. Then if $\lambda_0 \in \pi^{-1}(\Omega_0) \cap \mathcal{L}_0$, we have that $\overline{\gamma}(t) = \pi \circ e^{t\vec H}(\lambda_0)$ is a strict length-minimizer (up to reparametrization) among all admissible trajectories $\gamma$ with the same boundary conditions and such that $\gamma(t) \in \Omega_t$ for all $t \in [0,T]$.*

We are in position to prove that small pieces of normal trajectories are global length-minimizers.

**Theorem 4.65.** *Let $\gamma : [0,T] \to M$ be a sub-Riemannian normal trajectory. Then for every $\tau \in [0,T[$ there exists $\varepsilon_0 > 0$ such that for $0 < \varepsilon < \varepsilon_0$*

*(i) $\gamma|_{[\tau,\tau+\varepsilon]}$ is a length-minimizer, i.e., $d(\gamma(\tau), \gamma(\tau+\varepsilon)) = \ell(\gamma|_{[\tau,\tau+\varepsilon]})$.*

*(ii) $\gamma|_{[\tau,\tau+\varepsilon]}$ is the unique length-minimizer joining $\gamma(\tau)$ and $\gamma(\tau+\varepsilon)$, up to reparametrization.*

*Proof.* Without loss of generality we can assume that the curve is parametrized by arc length and prove the theorem for $\tau = 0$. Let $\gamma(t)$ be a normal extremal trajectory, such that $\gamma(t) = \pi(e^{t\vec H}(\lambda_0))$, for $t \in [0,T]$. Consider a smooth function $a \in C^\infty(M)$ such that $d_q a = \lambda_0$ and let $\mathcal{L}_t$ be the family of submanifold of $T^*M$ associated with this function by (4.89) and (4.90). By construction, the extremal lift associated with $\gamma$ satisfies $\lambda(t) = e^{t\vec{H}}(\lambda_0) \in \mathcal{L}_t$ for all $t$.

Let $\Omega_0$ be a sufficiently small open set containing $\gamma([0,T])$. For $t = 0$, the map $\pi \circ e^{t\vec{H}} \circ da|_{\Omega_0}$ is a diffeomorphism from $\Omega_0$ to $\Omega_0$ since it is the identity. By smoothness there exists $t_0 \in ]0,T]$ such that $\pi \circ e^{t\vec{H}} \circ da|_{\Omega_0}$ is a diffeomorphism onto its image $\Omega_t$ for every $t \in [0, t_0]$.

By Corollary 4.64, $\gamma|_{[0,t_0]}$ is a strict length-minimizer among all admissible trajectories $\gamma'$ with same boundary conditions and such that $\gamma'(t) \in \Omega_t$ for all $t \in [0, t_0]$.

Let $t_1 \in ]0, t_0]$ be such that the intersection $\cap_{t\in[0,t_1]}\Omega_t$ still contains a compact neighborhood $K$ of $\gamma(0)$. Let $t_K \in ]0, t_1]$ be such that $\gamma|_{[0,t_K]} \subset \text{int}(K)$. Let us now denote $\delta_K > 0$ the constant defined in Lemma 3.36 such that every curve starting from $\gamma(0)$ and leaving $K$ is necessary longer than $\delta_K$. Then, defining $\varepsilon := \min\{\delta_K, t_K\}$, we have that the curve $\gamma|_{[0,\varepsilon]}$ is shorter than any other curve with the same boundary conditions. Thus $\gamma|_{[0,\varepsilon]}$ is a global minimizer. Moreover it is unique up to reparametrization by uniqueness of the solution of the Hamiltonian equations.

□

*Remark* 4.66. When $\mathcal{D}_{q_0} = T_{q_0}M$, as it is the case for a Riemannian structure, the level set of the Hamiltonian

$$\{H = 1/2\} = \{\lambda \in T^*_{q_0}M | H(\lambda) = 1/2\},$$

is diffeomorphic to an ellipsoid, hence compact. Under this assumption, for each $\lambda_0 \in \{H = 1/2\}$, the corresponding geodesic $\gamma(t) = \pi(e^{t\vec{H}}(\lambda_0))$ is optimal up to a time $\varepsilon = \varepsilon(\lambda_0)$. By compactness of the set $\{H = 1/2\}$, it follows that it is possible to find a common $\varepsilon > 0$ (depending only on $q_0$) such that each normal trajectory with base point $q_0$ is optimal on the interval $[0, \varepsilon]$.

It can be proved that this is false as soon as $\mathcal{D}_{q_0} \neq T_{q_0}M$. Indeed in this case, for every $\varepsilon > 0$ there exists a normal extremal path that lose optimality in time $\varepsilon$, see Theorem 12.17.

With essentially no modifications, the proof of Theorem 4.62 (and of Corollary 4.64) permits to obtain the following result. Details are left to the reader.

**Exercise 4.67.** Prove that given $T > 0$ and a normal extremal defined on the interval $[-T, T]$ by

$$\gamma : [-T, T] \to M, \qquad \gamma(t) = \pi \circ e^{t\vec{H}}(\lambda_0), \tag{4.111}$$

for some initial covector $\lambda_0 \in T^*_qM$, for every $\tau \in ]-T, T[$ there exists $\varepsilon_0 > 0$ such that for $0 < \varepsilon < \varepsilon_0$

(i) $\gamma|_{[\tau-\varepsilon,\tau+\varepsilon]}$ is a length-minimizer, i.e., $d(\gamma(\tau - \varepsilon), \gamma(\tau + \varepsilon)) = \ell(\gamma|_{[\tau-\varepsilon,\tau+\varepsilon]})$.

(ii) $\gamma|_{[\tau-\varepsilon,\tau+\varepsilon]}$ is the unique length-minimizer joining $\gamma(\tau-\varepsilon)$ and $\gamma(\tau+\varepsilon)$, up to reparametrization.

*Remark* 4.68. We stress that, thanks to the results of this section, nontrivial normal extremal trajectories are geodesics. Nontrivial abnormal extremal trajectories could be geodesics or not (for an example of abnormal extremal trajectory that is not a geodesic see Section 12.6.1).

Thanks to Exercice 4.67 we can prove the following statement

**Proposition 4.69.** *Let $M$ be a sub-Riemannian manifold. Then for every $q \in M$ there exists $r_0 > 0$ such that*

$$\text{diam}(B(q, r)) = 2r, \qquad \forall\, r \leq r_0. \tag{4.112}$$

*Here* $\text{diam}(A)$ *denotes the diameter of a set $A \subset M$ with respect to the sub-Riemannian distance $d$.*

*Proof.* Since $B(q,\varepsilon)$ is a ball in the metric space $(M,d)$ then by the triangular inequality we have $\mathrm{diam}(B(q,r)) \leq 2r$ for every $r > 0$. Let us show that the equality holds if $r$ is small enough.

Fix a normal extremal trajectory $\gamma(t) = \pi \circ e^{t\vec{H}}(\lambda_0)$, for some initial condition $\lambda_0 \in T_q^*M \cap H^{-1}(1/2)$, and defined on a interval $[-T,T]$. Thanks to claim (i) of Exercice 4.67 for $\tau = 0$, we have $d(\gamma(-\varepsilon),\gamma(\varepsilon)) = \ell(\gamma|_{[-\varepsilon,\varepsilon]}) = 2\varepsilon$ since the curve is parametrized by arc length (thanks to the choice $\lambda_0 \in H^{-1}(1/2)$). It follows that $\mathrm{diam}(B(q,\varepsilon)) \geq d(\gamma(-\varepsilon),\gamma(\varepsilon)) = 2\varepsilon$ hence $\mathrm{diam}(B(q,\varepsilon)) = 2\varepsilon$, for $\varepsilon \leq \varepsilon_0$. □

## 4.8 Bibliographical note

The Hamiltonian approach to sub-Riemannian geometry is nowadays classical. However the construction of the symplectic structure, obtained by extending the Poisson bracket from the space of affine functions, is not standard. The presentation given here is inspired by [Gam14].

The extension to nonautonomous case stated in Proposition 4.12 is based on the proof of the variation equation for nonautonomous ODE, the interested reader could find the details in [BP07].

Historically, in the setting of PDE, the sub-Riemannian distance (also called Carnot-Carathéodory distance) is introduced by means of sub-unit curves, see for instance [Gar16] and the discussion therein. The link between the two definitions is obtained through Exercice 4.32.

The proof that normal extremals are geodesics is an adaptation of a technique used to prove optimality taken from [AS04] for a more general class of problems. This is inspired by the idea of "fields of extremals" in classical Calculus of Variation.

# Chapter 5

# First integrals and integrable systems

In this chapter we present some applications of the Hamiltonian formalism developed in the Chapter 4. In particular we give a proof the well-known Arnold-Liouville's theorem and, as an application, we study the complete integrability of the geodesic flow on a special class of Riemannian manifolds.

More examples of completely integrable geodesic flows are presented in Chapters 7 and 13. A proof that all left-invariant sub-Riemannian geodesic flows on 3D Lie groups are completely integrable is given in Chapter 18.

## 5.1 Reduction of Hamiltonian systems with symmetries

Recall that a symplectic manifold $(N, \sigma)$ is a smooth manifold endowed with a closed non-degenerate two-form $\sigma$ (cf. Section 4.6). Fix a smooth Hamiltonian $h : N \to \mathbb{R}$.

**Definition 5.1.** A *first integral* for the Hamiltonian system defined by $h$ is any non-constant smooth function $g : N \to \mathbb{R}$ such that $\{h, g\} = 0$.

Recall that by definition $\{h, g\} = \vec{h}(g) = -\vec{g}(h)$, hence, if $g$ is a first integral for the Hamiltonian system defined by $h$, we have

$$\frac{d}{dt} g \circ e^{t\vec{h}} = \frac{d}{dt} h \circ e^{t\vec{g}} = 0. \tag{5.1}$$

namely, $g$ is preserved along the flow of $\vec{h}$, and viceversa.

We want to show that the existence of a first integral for the Hamiltonian flow generated by $h$ permits to define a reduction of the symplectic space and to reduce to $2n-2$ dimensions. The construction of the reduction is local, in general.

Fix a regular level set $N_{g,c} = \{x \in N \mid g(x) = c\}$ of the function $g$. This means that $d_x g \neq 0$ for every $x \in N_{g,c}$. Fix a point $x_0$ in the level set and a neighborhood $U$ of $x_0$ such that $\vec{g}(x) \neq 0$ for $x \in U$. Notice that this is possible since $d_{x_0} g = \sigma(\cdot, \vec{g}(x_0))$, and we have $d_{x_0} g \neq 0$ and $\sigma$ non-degenerate. By continuity this holds in a neighborhood $U$.

The set $N_{g,c}$ has the structure of smooth manifold of dimension $2n-1$. Being odd dimensional, the restriction of the symplectic form to the tangent space $T_x N_{g,c}$ is necessarily degenerate, and its kernel is one-dimensional. Indeed, following the same arguments as in the proof of Proposition 4.30, we have that

$$\ker \sigma|_{T_x N_{g,c}} = \vec{g}(x),$$

and integral curves of $\vec{g}$ are tangent to the level set $N_{g,c}$. This means that the flow of $\vec{g}$ is well defined on the level set. See Figure 5.1.

Next, consider on $U \cap N_{g,c}$ the equivalence relation $x_1 \sim x_2$ if there exists $s \in \mathbb{R}$ such that $x_2 = e^{s\vec{g}}(x_1)$. We define $\widetilde{N}$ as the set of orbits of the one parametric group $\{e^{s\vec{g}}\}_{s\in\mathbb{R}}$ contained in the fixed level set $N_{g,c}$ of $g$. Up to restricting $U$, the quotient has the structure of smooth manifold of dimension $2n-2$. To build a chart close to a point $[x_0] \in \widetilde{N}$ (with $x_0 \in N_{g,c}$) it is enough to find an hypersurface $N'_{g,c} \subset N_{g,c}$ passing through $x_0$ and transversal to the orbit itself, namely

$$T_{x_0}N_{g,c} = T_{x_0}N'_{g,c} \oplus \vec{g}(x_0),$$

Then local coordinates on $N'_{g,c}$, which has dimension $2n-2$, induces local coordinates on $\widetilde{N}$.

The construction of the above quotient is classical (see for instance [Arn89]). The restriction of the symplectic structure $\sigma$ to the quotient $\widetilde{N}$ is necessarily non-degenerate (using that $\sigma$ is non-degenerate on the whole space $N$), and gives to $\widetilde{N}$ the structure of symplectic manifold.

Coming back to the original Hamiltonian $h$ in involution with $g$, we have that $\vec{h}$ is indeed well defined on the quotient. Indeed since $\{h, g\} = 0$ we have (for every $t, s$ such that the left and right hand side are well-defined)

$$e^{s\vec{g}} \circ e^{t\vec{h}} = e^{t\vec{h}} \circ e^{s\vec{g}},$$

and $\vec{h}$ induces a well-defined Hamiltonian flow on $\widetilde{N}$. In particular every function $f$ on $N$ that commutes with $g$, thanks to (5.1), is constant along the trajectories of $\vec{g}$, hence defines a function on the quotient $\widetilde{N}$.

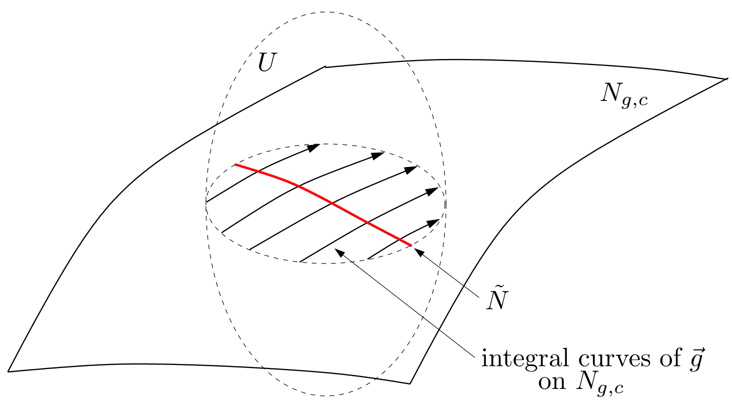



Figure 5.1: Symplectic reduction.

**Exercise 5.2.** Prove that given $f_1, f_2 \in C^\infty(N)$ such that $\{f_1, g\} = \{f_2, g\} = 0$, one has that $\{\{f_1, f_2\}, g\} = 0$. Deduce that the Poisson bracket defined on $N$ descends to a well-defined Poisson bracket defined on the quotient $\widetilde{N}$ where $C^\infty(\widetilde{N}) \simeq \{f \in C^\infty(N) \mid \{f, g\} = 0\}$.

We end this section by showing that the construction of the space of orbits of an (Hamiltonian) vector field is in general only local as the following classical example shows.

**Example 5.3.** Consider the torus[1] $T^2 \simeq [0,1]^2/_\sim$, endowed with the canonical symplectic structure $\sigma = dp \wedge dx$ and the Hamitonian $g(x,p) = -\alpha x + p$. The vector field $\vec{g}$ is written as follows

$$\vec{g}(x,y) = \frac{\partial g}{\partial p}\frac{\partial}{\partial x} - \frac{\partial g}{\partial x}\frac{\partial}{\partial p} = \frac{\partial}{\partial x} + \alpha\frac{\partial}{\partial p},$$

[1]with the equivalence relation $(x,0) \sim (x,1)$ and $(0,p) \sim (1,p)$.

whose trajectories are given by

$$x(t) = x_0 + t, \qquad p(t) = p_0 + \alpha t.$$

For $\alpha \in \mathbb{R} \setminus \mathbb{Q}$, then every trajectory is an immersed one-dimensional submanifold of $T^2$ that is dense in $T^2$. Hence the space of orbits (quotient with respect to the equivalence relation) has globally even no structure of topological manifold (the quotient topology is not Hausdorff).

The next subsection describes an explicit situation where the symplectic reduction is globally defined.

### 5.1.1 An example of symplectic reduction: the space of affine lines in $\mathbb{R}^n$

In this section we consider an important example of symplectic reduction, that is going to be used in what follows.

Let us consider the symplectic manifold $N = T^*\mathbb{R}^n = \mathbb{R}^n \times \mathbb{R}^n$ with coordinates $(p, x) \in \mathbb{R}^n \times \mathbb{R}^n$ and canonical symplectic form

$$\sigma = \sum_{i=1}^{n} dp_i \wedge dx_i.$$

Define the Hamiltonian $g : \mathbb{R}^{2n} \to \mathbb{R}$ given by

$$g(x, p) = \frac{1}{2}|p|^2.$$

We want to prove the following result.

**Proposition 5.4.** *For every $c > 0$ the level set $N_{g,c}$ of $g$ is globally diffeomorphic to $\mathbb{R}^n \times S^{n-1}$, and its symplectic reduction $\widetilde{N}$ is a smooth (symplectic) manifold of dimension $2n-2$ globally diffeomorphic to the space of oriented affine lines in $\mathbb{R}^n$.*

*Proof.* For every $c > 0$ then we have that the level set

$$N_{g,c} = \{(x, p) : g(x, p) = c\} = \{(x, p) : |p|^2 = 2c\},$$

is a smooth hypersurface of $\mathbb{R}^{2n}$ of dimension $2n - 1$, indeed globally diffeomorphic to $\mathbb{R}^n \times S^{n-1}$.

The Hamiltonian system for $\vec{g}$ is easily solved for every initial condition $(x(0), p(0)) = (x_0, p_0)$

$$\begin{cases} \dot{x} = \dfrac{\partial g}{\partial p}(x, p) = p \\ \dot{p} = -\dfrac{\partial g}{\partial x}(x, p) = 0 \end{cases} \qquad \Rightarrow \qquad \begin{cases} x(t) = x_0 + tp_0 \\ p(t) = p_0 \end{cases}, \tag{5.2}$$

and its flow is globally defined, described by a straight line contained in the space $N_{g,c}$ (notice that $c > 0$ implies $p_0 \neq 0$). Hence it is clear that the quotient $\widetilde{N}$ of $N_{g,c}$ with respect to orbits of the Hamiltonian vector field $\vec{g}$ is the space of oriented affine lines of $\mathbb{R}^n$ and is globally defined. The proof is completed by Proposition 5.6. $\square$

*Remark* 5.5. It is important to consider oriented affine lines. For instance, when $n = 1$, the space of orbits consists of two lines, being diffeomorphic to $\mathbb{R} \times S^0 = \mathbb{R} \times \{-1, 1\}$. These is the space of oriented lines on $\mathbb{R}$.

**Proposition 5.6.** *The set $A(n)$ of oriented affine lines in $\mathbb{R}^n$ has the structure of smooth (symplectic) manifold of dimension $2n-2$.*

*Proof.* We first fix some notation: denote by $H_i := \{x_i = 0\} \subset \mathbb{R}^n$ the $i$-th coordinate hyperplane and by $U_i^+ = S^{n-1} \cap \{x_i > 0\}$ and $U_i^- = S^{n-1} \cap \{x_i < 0\}$ the open sets of the standard covering of the sphere $S^{n-1}$, for every $i = 1, \ldots, n$.

We define an open cover on $A(n)$ in the following way: consider the open sets $W_i \subset A(n)$ of affine lines $L$ of $\mathbb{R}^n$ that are not parallel to the hyperplane $H_i$. Then for every oriented line $L \in W_i$ there exists a unique $\bar{x} \in H_i$ and $\bar{v} \in U_i^+ \cup U_i^-$ such that $L = \{\bar{x} + t\bar{v} \mid t \in \mathbb{R}\}$. We write accordingly $W_i = W_i^+ \cup W_i^-$. Then, for $i = 1, \ldots, n$, we define the coordinate charts

$$\phi_i^\pm : W_i^\pm \to H_i \times U_i^\pm, \qquad \phi_i^\pm(L) = (\bar{x}, \bar{v}).$$

Using the standard identification $H_i \simeq \mathbb{R}^{n-1}$ and the stereographic projection $W_i^\pm \simeq \mathbb{R}^{n-1}$, we build coordinate maps $\phi_i^\pm : W_i^\pm \to \mathbb{R}^{2n-2}$ for $i = 1, \ldots, n$. $\square$

**Exercise 5.7.** Check that $\{W_i\}_{i=1,\ldots,n}$ is an open cover of $A(n)$, and that the change of coordinates $\phi_i^\pm \circ (\phi_j^\pm)^{-1} : \mathbb{R}^{2n-2} \to \mathbb{R}^{2n-2}$ is smooth for every $i, j = 1, \ldots, n$.

## 5.2 Riemannian geodesic flow on hypersurfaces

The Riemannian geodesic flow on an hypersurface $M$ of $\mathbb{R}^n$ is an Hamiltonian flow on the $2n-2$ dimensional submanifold $T^*M$ of $T^*\mathbb{R}^n = \mathbb{R}^{2n}$. The goal of this section is to interpret this flow as the reduction (in the sense of the previous section 5.1.1) of the Hamiltonian flow of $\mathbb{R}^{2n}$ to the symplectic space of affine lines in $\mathbb{R}^n$.

### 5.2.1 Geodesics on hypersurfaces

Let us consider now a smooth function $a : \mathbb{R}^n \to \mathbb{R}$ and consider the family of hypersurfaces defined by the level sets of $a$

$$M_c := a^{-1}(c) \subset \mathbb{R}^n, \qquad c \text{ is a regular value of } a,$$

endowed with the Riemannian structure induced by the ambient space $\mathbb{R}^n$. Recall that, by classical Sard's Lemma for almost every $c \in \mathbb{R}$, $c$ is a regular value for $a$ (in particular, $M_c$ is a smooth submanifold of codimension one in $\mathbb{R}^n$).

An adaptation of the arguments of Proposition 1.3 in Chapter 1 to hypersurfaces, one can prove the following characterization of geodesics.

**Proposition 5.8.** *Let $M$ be a smooth hypersurface of $\mathbb{R}^n$ and let $\gamma : [0, T] \to M$ be a smooth length-minimizer parametrized by arc length. Then $\ddot{\gamma}(t) \perp T_{\gamma(t)}M$ for every $t \in [0, T]$.*

Notice that all length-minimizers are smooth thanks to the results of Chapter 4.

### 5.2.2 Riemannian geodesic flow and symplectic reduction

For a large class of functions $a$, we will find an Hamiltonian, defined on the ambient space $T^*\mathbb{R}^n$, whose (reparametrized) flow generates the geodesic flow when restricted to each level set $M_c$.

Consider the standard symplectic structure on $T^*\mathbb{R}^n$

$$T^*\mathbb{R}^n = \mathbb{R}^n \times \mathbb{R}^n = \{(x,p) \mid x, p \in \mathbb{R}^n\}, \qquad \sigma = \sum_{i=1}^n dp_i \wedge dx_i,$$

For $x, p \in \mathbb{R}^n$ we will denote by $x + \mathbb{R}p$ the line $\{x + tp \mid t \in \mathbb{R}\} \subset \mathbb{R}^n$ .

**Assumption.** We assume that the function $a : \mathbb{R}^n \to \mathbb{R}$ satisfies the following assumptions:

(A1) the restriction of $a : \mathbb{R}^n \to \mathbb{R}$ to *every* affine line is strictly convex,

(A2) $a(x) \to +\infty$ when $|x| \to +\infty$.

Under assumptions (A1)-(A2), the restriction of the function $a$ to each affine line in $\mathbb{R}^n$ always attains a minimum in a unique point, and we can define the Hamiltonian

$$h : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}, \qquad h(x,p) = \min_{t \in \mathbb{R}} a(x + tp). \tag{5.3}$$

By definition, the function $h$ is constant on every affine line in $\mathbb{R}^n$. If we define

$$g : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}, \qquad g(x,p) = \frac{1}{2}|p|^2. \tag{5.4}$$

this implies the following (cf. proof of Proposition 5.4).

**Lemma 5.9.** *The Hamiltonian $h$ is constant along the flow of $\vec{g}$, i.e., $\{h, g\} = 0$.*

We can then apply the symplectic reduction technique explained in Section 5.1: the flow of $\vec{h}$ induced a well defined flow on the reduced symplectic space of dimension $2n-2$ of affine lines in $\mathbb{R}^n$ (cf. Section 5.1.1). We want to interpret this flow of affine lines as a flow on the level set $M_c$ and to show that this is actually the Riemannian geodesic flow.

For every $x, p \in \mathbb{R}^n$ let us define the functions

$$s : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}, \qquad \xi : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n,$$

defined as follows

(a) $s(x,p)$ is the point at which the scalar function $t \mapsto a(x + tp)$ attains its minimum,

(b) $\xi(x,p) = x + s(x,p)p$.

Notice that, by construction, we have $h(x,p) = a(\xi(x,p))$ for every $x, p \in \mathbb{R}^n$.

The first observation is that the line $x + \mathbb{R}p$ is tangent at $\xi(x,p)$ to the level set $a^{-1}(c)$, with $c := a(\xi(x,p))$. Indeed combining (a) and (b) we have

$$\langle \nabla_\xi a \mid p \rangle = \left.\frac{d}{dt}\right|_{t=s(x,p)} a(x + tp) = 0, \tag{5.5}$$

where $\langle \cdot \mid \cdot \rangle$ denotes the scalar product in $\mathbb{R}^n$.

**Proposition 5.10.** *Let $(x(t), p(t))$, for $t \in [0, T]$, be a trajectory of the Hamiltonian vector field $\vec{h}$ associated with* (5.3)*. Then the curve*

$$t \mapsto \xi(t) := \xi(x(t), p(t)) \in \mathbb{R}^n, \tag{5.6}$$

*(i) is contained in a fixed level set $M_c = a^{-1}(c)$, for some $c \in \mathbb{R}$,*

*(ii) is a reparametrization of a geodesic on $M_c$.*

In other words, if we follow the motion of the affine lines $x(t) + \mathbb{R}p(t)$ along the flow $(x(t), p(t))$ of $\vec{h}$, then the family of lines stay tangent to a fixed level set and the point of tangency $\xi(t)$ describes a geodesic on it. See Figure 5.2.

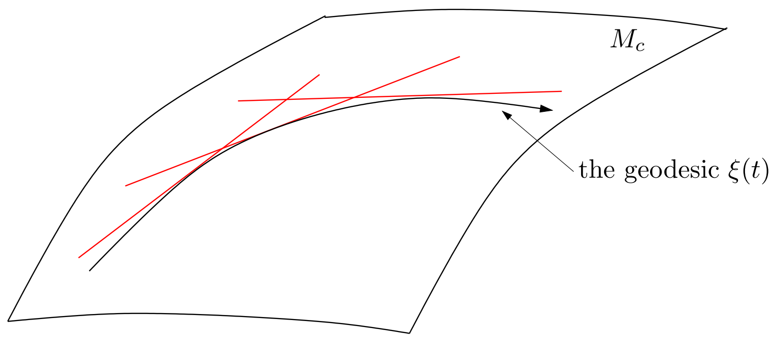



Figure 5.2: The line flow.

*Proof.* Property (i) is a simple consequence of Corollary 4.19, since every function is constant along the flow of its Hamiltonian vector field. Indeed by construction $h(x, p) = a(\xi(x, p))$ and, denoting by $(x(t), p(t))$ the Hamiltonian flow, one gets

$$a(\xi(t)) = a(\xi(x(t), p(t))) = h(x(t), p(t)) = \text{const},$$

i.e., the curve $\xi(t)$ is contained on a level set of $a$. Moreover by definition of $\xi(t)$ we have (cf. (5.5))

$$\left\langle \nabla_{\xi(t)} a \,\middle|\, p(t) \right\rangle = 0, \qquad \forall\, t. \tag{5.7}$$

The Hamiltonian system associated with $h$ reads

$$\begin{cases} \dot{x}(t) = s(t) \nabla_{\xi(t)} a \\ \dot{p}(t) = -\nabla_{\xi(t)} a \end{cases} \tag{5.8}$$

that immediately implies $\dot{x}(t) + s(t)\dot{p}(t) = 0$. Thus computing the derivative of $\xi(t) = x(t) + s(t)p(t)$ one gets

$$\dot{\xi}(t) = \dot{s}(t) p(t),$$

it follows that $\dot{\xi}(t)$ is parallel to $p(t)$. Notice that $s = s(t)$ is a well-defined parameter on the curve $\xi(t)$. Indeed computing the derivative with respect to $t$ in (5.7) we have that

$$\dot{s}(t) \left\langle \nabla^2_{\xi(t)} a \, p(t) \,\middle|\, p(t) \right\rangle - |\nabla_{\xi(t)} a|^2 = 0.$$

The strict convexity of $a$ implies $\left\langle \nabla^2_{\xi(t)} a\, p(t) \,\middle|\, p(t) \right\rangle \neq 0$ for every $t$. Hence for every $t \in [0, T]$ one has

$$\dot{s}(t) = \frac{|\nabla_{\xi(t)} a|^2}{\left\langle \nabla^2_{\xi(t)} a\, p(t) \,\middle|\, p(t) \right\rangle} \neq 0.$$

In particular $p(t)$ denotes the velocity of the curve $\xi(t)$, when reparametrized with the parameter $s = s(t)$, since $|p(t)| = 1$ implies $|\dot{\xi}(t)| = \dot{s}(t)$. With an abuse of notation we denote by $\xi(s)$ the curve $s \mapsto \xi(t^{-1}(s))$.

Finally, the second derivative of the reparametrized $\xi(s)$ is $\dot{p}(s)$ and, since $\dot{p}(s)$ is parallel to $\nabla_{\xi(s)} a = 0$ by (5.8), the second derivative $\ddot{\xi}(s)$ (i.e., the curve $\xi$ reparametrized by the arc length) is orthogonal to the level set, i.e., $s \mapsto \xi(s)$ is a geodesic on the level set.

□

*Remark* 5.11. Thus we can visualize the solutions of $\vec{h}$ as a motion of lines: the lines move in such a way to be tangent to one and the same geodesic. The tangency point $x$ on the line moves parallel to this line in this process. We will also refer to this flow as the "line flow" associated with $a$.

To end this section let us prove the following result, that will be used later in Section 5.6. Consider two functions $a, b : \mathbb{R}^n \to \mathbb{R}$ satisfying our assumptions (A1)-(A2). Following our notation, we set

$$\begin{aligned} h(x,p) &= a(\xi(x,p)), \qquad & \xi(x,p) &= x + s(x,p)p \\ g(x,p) &= b(\eta(x,p)), \qquad & \eta(x,p) &= x + \tau(x,p)p \end{aligned}$$

where $s(x,p)$ and $\tau(x,p)$ are defined as above, and $\xi$, $\eta$ denote the tangency point of the line $x + \mathbb{R}p$ with the level set of $a$ and $b$ respectively. The following proposition computes the Poisson bracket of these Hamiltonian functions

**Proposition 5.12.** *Under the previous assumptions*

$$\{h, g\} = (s - \tau) \langle \nabla_\xi a \,|\, \nabla_\eta b \rangle . \tag{5.9}$$

*Proof.* The coordinate expression of the Poisson bracket (4.19) can be rewritten as

$$\{h, g\} = \langle \nabla_p h \,|\, \nabla_x g \rangle - \langle \nabla_x h \,|\, \nabla_p g \rangle , \tag{5.10}$$

and using equation (5.8) for both $h$ and $g$ one gets

$$\{h, g\} = (s - \tau) \langle \nabla_\xi a \,|\, \nabla_\eta b \rangle . \tag{5.11}$$

□

## 5.3 Sub-Riemannian structures with symmetries

Let $M$ be a sub-Riemannian manifold and denote by $H$ the associated sub-Riemannian Hamiltonian (cf. Section 4.3.1).

**Definition 5.13.** Let $M$, $N$ be two sub-Riemannian structures, and let $x_0 \in M$, $y_0 \in N$. The two structures are said to be locally isometric if there exists a local diffeomorphism $\phi : O_{x_0} \to O_{y_0}$ such that $\phi(x_0) = y_0$ and such that $\phi_* : T_{x_0}M \to T_{y_0}N$ preserves the distribution and the inner product on it.

If in the previous definition the map $\phi$ can be chosen to be globally defined, then we say that $\phi$ is a *global isometry*.

**Definition 5.14.** Let $M$ be a sub-Riemannian manifold. We say that a complete smooth vector field $X \in \mathrm{Vec}(M)$ is a *Killing vector field* if it generates a one parametric flow of local isometries, i.e., $e^{tX} : M \to M$ is an isometry for all $t \in \mathbb{R}$.

For every $X \in \mathrm{Vec}(M)$, we can define the function $h_X \in C^\infty(T^*M)$ linear on fibers associated with $X$ by $h_X(\lambda) = \langle \lambda, X(q) \rangle$, where $q = \pi(\lambda)$.

The following lemma shows that $X$ is a Killing vector field if and only if $h_X$ commutes with the sub-Riemannian Hamiltonian $H$.

**Lemma 5.15.** *Let $M$ be a sub-Riemannian manifold and $H$ be the sub-Riemannian Hamiltonian. A vector field $X \in \mathrm{Vec}(M)$ is a Killing vector field if and only if $\{H, h_X\} = 0$.*

*Proof.* By definition a vector field $X$ generates isometries if and only if the differential of its flow $e^{tX}_* : T_qM \to T_{e^{tX}(q)}M$ preserves the sub-Riemannian distribution and the norm on it, i.e., $e^{tX}_* v \in \mathcal{D}_{e^{tX}(q)}$ for every $v \in \mathcal{D}_q$ and $\|e^{tX}_* v\| = \|v\|$. By definition of $H$, this is equivalent to the identity

$$H((e^{tX})^*\lambda) = H(\lambda), \qquad \forall\, \lambda \in T^*M. \tag{5.12}$$

Now, Proposition 4.10 implies that $(e^{-tX})^* = e^{t\vec{h}_X}$, where $h_X$ is the Hamiltonian linear on fibers related to $X$. Hence one obtains from (5.12) that $H \circ (e^{-t\vec{h}_X}) = H$. Differentiating with respect to $t$, one obtains $\vec{h}_X H = \{H, h_X\} = 0$. $\square$

In other words, with every 1-parametric group of isometries of $M$ we can associate an Hamiltonian in involution with $H$. Let us show two classical examples where we have a sub-Riemannian structure with symmetries.

**Example 5.16** (Revolution surfaces in $\mathbb{R}^3$)**.** Let $M$ be a 2-dimensional revolution surface in $\mathbb{R}^3$. Since rotations around the revolution axis are isometries, we have that the Hamiltonian $H_X$ associated with the generator $X$ of the rotations is a first integral of the geodesic flow.

**Example 5.17** (Isoperimetric sub-Riemannian problem)**.** Let us consider a sub-Riemannian structure associated with an isoperimetric problem defined on a 2-dimensional surface $M$ (see Section 4.4.2). The sub-Riemannian structure on $M \times \mathbb{R}$ is determined by the function $b \in C^\infty(M)$ satisfying $dA = bdV$, where $A \in \Lambda^1(M)$ is the 1-form defining the isoperimetric problem and $dV$ is the volume form on $M$.

We have that

(i) By construction the problem is invariant by translations along the $z$-axis, then the linear Hamiltonian $h_Z$ associated with the generator $Z$ of the translations is a first integral of the geodesic flow.

(ii) If $M$ is a revolution surface and $b$ is invariant by rotations around the revolution axis, then the linear Hamiltonian $h_X$ associated with the generator $X$ of the rotations is a first integral of the geodesic flow, as in the previous example.

## 5.4 Completely integrable systems

**Definition 5.18.** Let $M$ be an $n$-dimensional smooth manifold and let $h : T^*M \to \mathbb{R}^n$ be a smooth map defined by $h = (h_1, \ldots, h_n)$. We say that the map $h$ is *completely integrable* if

a) $\{h_i, h_j\} = 0$, for $i, j = 1, \ldots, n$.

b) the differentials $d_\lambda h_1, \ldots, d_\lambda h_n$ are independent on an open dense set of $T^*M$.

The same terminology applies to any of the Hamiltonian system defined by one of the Hamiltonian $h_i$, for $i = 1, \ldots, n$.

**Lemma 5.19.** *Assume that $h$ is completely integrable and let $c \in \mathbb{R}^n$ be a regular value of $h$. Then the set $h^{-1}(c)$ is a $n$-dimensional submanifold in $T^*M$ and we have*

$$T_\lambda h^{-1}(c) = \operatorname{span}\{\vec{h}_1(\lambda), \ldots, \vec{h}_n(\lambda)\}, \qquad \forall\, \lambda \in h^{-1}(c). \tag{5.13}$$

*Proof.* Since $c$ is a regular value of $h$, by Remark 2.62 the set $h^{-1}(c)$ is a submanifold of dimension $n$ in $T^*M$. In particular $\dim T_\lambda h^{-1}(c) = n$ for every $\lambda \in h^{-1}(c)$. Moreover, by Exercise 2.12, each vector field $\vec{h}_i$ is tangent to $h^{-1}(c)$, since $\vec{h}_i h_j = \{h_i, h_j\} = 0$ by assumption. To prove (5.13) it is then enough to show that these vector fields are linearly independent.

Since $c$ is a regular value of $h$, the differentials of the functions $h_i$ are linearly independent on $h^{-1}(c)$, namely

$$\dim \operatorname{span}\{d_\lambda h_1, \ldots, d_\lambda h_n\} = n, \quad \forall\, \lambda \in h^{-1}(c). \tag{5.14}$$

Moreover the symplectic form $\sigma$ on $T^*M$ induces for all $\lambda$ an isomorphism $T_\lambda(T^*M) \to T^*_\lambda(T^*M)$ defined by $w \mapsto \sigma_\lambda(\cdot, w)$. By nondegeneracy of the symplectic form, this implies that

$$\dim \operatorname{span}\{\vec{h}_1(\lambda), \ldots, \vec{h}_n(\lambda)\} = n, \quad \forall\, \lambda \in h^{-1}(c). \tag{5.15}$$

hence they form a basis for $T_\lambda h^{-1}(c)$. □

*Remark* 5.20. Notice that the symplectic form vanishes on $T_\lambda h^{-1}(c)$. Indeed this is a consequence of the fact that $\sigma(\vec{h}_i, \vec{h}_j) = \{h_i, h_j\} = 0$ for all $i, j = 1, \ldots, n$.

In what follows we denote by $N_c = h^{-1}(c)$ the level set of $h$. If $h^{-1}(c)$ is not connected, $N_c$ will denote a connected component of $h^{-1}(c)$.

**Proposition 5.21.** *Assume that $h$ is completely integrable and let $c \in \mathbb{R}^n$ be a regular value of $h$. Assume moreover that the vector fields $\vec{h}_i$ are complete and define the map*

$$\Psi : \mathbb{R}^n \to \operatorname{Diff}(N_c), \qquad \Psi(s_1, \ldots, s_n) := e^{s_1 \vec{h}_1} \circ \ldots \circ e^{s_n \vec{h}_n}\Big|_{N_c}. \tag{5.16}$$

*For every $\lambda \in N_c$, the map $\Psi_\lambda : \mathbb{R}^n \to N_c$ defined by $\Psi_\lambda(s) := \Psi(s)\lambda$ defines a transitive action of $\mathbb{R}^n$ onto $N_c$.*

*Proof.* The complete integrability assumption together with Corollary 4.57 implies that the flows of $\vec{h}_i$ and $\vec{h}_j$ commute for every $i, j = 1, \ldots, n$ since

$$[\vec{h}_i, \vec{h}_j] = \overrightarrow{\{h_i, h_j\}} = 0.$$

By Proposition 2.27, this is equivalent to

$$e^{t\vec{h}_i} \circ e^{\tau\vec{h}_j} = e^{\tau\vec{h}_j} \circ e^{t\vec{h}_i}, \qquad \forall\, t, \tau \in \mathbb{R}. \tag{5.17}$$

Since the vector fields are complete by assumption, we can compute for every $s, s' \in \mathbb{R}^n$

$$\begin{aligned}
\Psi(s+s') &= e^{(s_1+s_1')\vec{h}_1} \circ \ldots \circ e^{(s_n+s_n')\vec{h}_n} \\
&= e^{s_1\vec{h}_1} \circ e^{s_1'\vec{h}_1} \circ \ldots \circ e^{s_n\vec{h}_n} \circ e^{s_n'\vec{h}_n} \\
&= e^{s_1\vec{h}_1} \circ \ldots \circ e^{s_n\vec{h}_n} \circ e^{s_1'\vec{h}_1} \circ \ldots \circ e^{s_n'\vec{h}_n} \qquad\qquad \text{(by (5.17))} \\
&= \Psi(s) \circ \Psi(s'),
\end{aligned}$$

which proves that $\Psi$ is a group action. Denote, for every point $\lambda \in N_c$, its orbit under the group action, namely

$$\Omega_\lambda = \operatorname{im} \Psi_\lambda = \{\Psi_\lambda(s) \mid s \in \mathbb{R}^n\}.$$

Applying the Rashevski-Chow theorem at the family $\{\vec{h}_1, \ldots, \vec{h}_n\}$ of vector fields on $N_c$, and using the fact that $N_c$ is connected, it follows that $\Omega_\lambda = N_c$ for every $\lambda \in N_c$. $\square$

Notice that, for every $\lambda \in N_c$, the map $\Psi_\lambda$ is a smooth local diffeomorphism at every $s \in \mathbb{R}^n$. Indeed, using (5.17), one has (cf. also Exercice 2.32)

$$\frac{\partial \Psi_\lambda}{\partial s_i}(\Psi_\lambda(s)) = \vec{h}_i(\Psi_\lambda(s)), \quad i = 1, \ldots, n,$$

and the partial derivatives are linearly independent at each point of $N_c$. However, in general, the map $\Psi_\lambda$ is not a global diffeomorphism, since it is possibly non injective (as for instance in the case when $M$ is compact).

**Proposition 5.22.** *Assume that $h$ is completely integrable and let $c \in \mathbb{R}^n$ be a regular value of $h$ such that $N_c := h^{-1}(c)$ is connected. Then $N_c$ is diffeomorphic to $T^k \times \mathbb{R}^{n-k}$ for some $0 \leq k \leq n$, where $T^k$ denotes the $k$-dimensional torus. Fix coordinates $\theta \in T^k \times \mathbb{R}^{n-k}$, with $(\theta_1, \ldots, \theta_k) \in T^k$ and $(\theta_{k+1}, \ldots, \theta_n) \in \mathbb{R}^{n-k}$, then we have*

$$\vec{h}_i = \sum_{j=1}^{n} b_{ij}(c)\partial_{\theta_j}, \tag{5.18}$$

*for some constants $b_{ij}(c)$ that are independent on $\lambda \in N_c$.*

To prove Proposition 5.22, we need some preliminary considerations. Let us denote by $S_\lambda$ the stabiliser of the point $\lambda$, i.e., the set

$$S_\lambda = \{s \in \mathbb{R}^n \mid \Psi_\lambda(s) = \lambda\}.$$

**Exercise 5.23.** Prove that $S_\lambda$ is a discrete[2] subgroup of $\mathbb{R}^n$, independent on $\lambda \in N_c$.

We also need the following lemma.

[2]Recall that a subgroup $G$ of $\mathbb{R}^n$ is discrete if and only if for every $g \in G$ there exist an open set $U \subset \mathbb{R}^n$ containing $g$ and such that $U \cap G = \{g\}$.

**Lemma 5.24.** *Let $G$ be a non trivial discrete subgroup of $\mathbb{R}^n$. Then there exist $k \in \mathbb{N}$ with $1 \leq k \leq n$ and $v_1, \ldots, v_k \in \mathbb{R}^n$ such that*

$$G = \left\{ \sum_{i=1}^{k} m_i v_i,\ m_i \in \mathbb{Z} \right\}.$$

*Proof.* We prove the claim by induction on the dimension $n$ of the ambient space $\mathbb{R}^n$.

(i). Let $n = 1$. Since $G$ is a discrete subgroup of $\mathbb{R}$, then there exists an element $v_1 \neq 0$ closest to the origin $0 \in \mathbb{R}$. We claim that $G = \mathbb{Z}v_1 = \{mv_1,\ m \in \mathbb{Z}\}$. By contradiction assume that there exists an element $f \in G$ such that $mv_1 < f < (m+1)v_1$ for some $m \in \mathbb{Z}$. Then $\bar{f} := f - mv_1$ belong to $G$ and is closer to the origin with respect to $v_1$, that is a contradiction.

(ii). Assume the statement is true for $n-1$ and let us prove it for $n$. The discreteness of $G$ guarantees the existence of an element $v_1 \in G$, closest to the origin. Moreover one can prove that $G_1 := G \cap \mathbb{R}v_1$ is a subgroup and, as in part (i) of the proof, that

$$G_1 := G \cap \mathbb{R}v_1 = \mathbb{Z}v_1.$$

If $G = G_1$ then the theorem is proved with $k = 1$. Otherwise one can consider the quotient $G/G_1$.

**Exercise 5.25.** (i). Prove that there exists a nonzero element $v_2 \in G/G_1$ that minimize the distance to the line $\ell = \mathbb{R}v_1$ in $\mathbb{R}^n$.
(ii). Show that there exists a neighborhood of the line $\ell$ that does not contain elements of $G/G_1$.

By Exercise 5.25 the quotient group $G/G_1$ is a discrete subgroup in $\mathbb{R}^n/\ell \simeq \mathbb{R}^{n-1}$. Hence, by the induction step there exists $v_2, \ldots, v_k$ such that

$$G/G_1 = \left\{ \sum_{i=2}^{k} m_i v_i,\ m_i \in \mathbb{Z} \right\}. \qquad \square$$

*Proof of Proposition 5.22.* Let us consider the elements $v_1, \ldots, v_k \in \mathbb{R}^n$ generators of the stabiliser $S_\lambda$ (independent on $\lambda$) given by Lemma 5.24 and complete it to a global basis $v_1, \ldots, v_n$. Denote by $e_1, \ldots, e_n$ the canonical basis of $\mathbb{R}^n$ and by $B : \mathbb{R}^n \to \mathbb{R}^n$ an isomorphism such that $Be_i = v_i$ for $i = 1, \ldots, n$. Notice that $B$ does not depend on $\lambda \in N_c$ and is thus a function of $c$ only.

Then the map $B \circ \Psi_\lambda : \mathbb{R}^n \to N_c$ is a local diffeomorphism and, due to the fact that $S_\lambda$ is the stabiliser of $\Psi_\lambda$, descends to a well-defined map on the quotient

$$B \circ \Psi_\lambda : T^k \times \mathbb{R}^{n-k} \to N_c$$

that is a global diffeomorphism. Introduce the coordinates $(\theta_1, \ldots, \theta_n)$ in $\mathbb{R}^n$ induced by the choice of the basis $v_1, \ldots, v_n$.

Since $(\theta_1, \ldots, \theta_n)$ are obtained by $(s_1, \ldots, s_n)$ by a linear change of coordinates on each level set, the vector fields $\vec{h}_i$ are constant in the $s$ coordinates (indeed $\vec{h}_i = \partial_{s_i}$) we have and the basis $\partial_{\theta_1}, \ldots, \partial_{\theta_n}$ can be expressed as follows

$$\vec{h}_i = \partial_{s_i} = \sum_{j=1}^{n} b_{ij}(c) \partial_{\theta_j}, \tag{5.19}$$

where $b_{ij}$ are the coefficients of the operator $B$, depending only on $c$ (i.e., are constant on each level set $N_c$). $\square$

*Remark* 5.26. In general, the set $(c, \theta)$ does not define local coordinates on $T^*M$. *If we assume* that $(c, \theta)$ define a set of local coordinates, then the Hamiltonian system defined by $h_i$ takes the form (on the whole space $T^*M$)

$$\begin{cases} \dot{c} = 0 \\ \dot{\theta}_j = b_{ij}(c) \end{cases}, \qquad i = 1, \ldots, n. \tag{5.20}$$

Notice that, assuming that $(c, \theta)$ define local coordinates, the pair $(c, \theta')$ where $\theta'_i := \theta_i + \psi_i(c)$ for $i = 1, \ldots, n$, still defines a set of cylindirical coordinates on each level set. However, since $\dot{c} = 0$, we have $\dot{\theta}'_j = \dot{\theta}_j$. This means that the vector fields $\partial_{\theta_i}$ are well-defined, i.e., independent on this choice.

## 5.5 Arnold-Liouville theorem

In this section we consider in detail the case when the level sets of a completely integrable system defined by

$$h : T^*M \to \mathbb{R}^n, \qquad h = (h_1, \ldots, h_n),$$

are *compact*. More precisely we assume that for all $c \in \mathbb{R}^n$, $c$ is a regular values for $h$ and the level set $h^{-1}(c)$ is a smooth compact and connected manifold.

From Proposition 5.21 and the fact that $T^k \times \mathbb{R}^{n-k}$ is compact if and only if $k = n$ we have the following corollary.

**Corollary 5.27.** *If $N_c$ is compact, then $N_c \simeq T^n$.*

Fix $\lambda \in N_c$ and introduce the diffeomorphism

$$F_c : T^n \to N_c, \qquad F_c(\theta_1, \ldots, \theta_n) = \Psi_\lambda(\theta_1 + 2\pi\mathbb{Z}, \ldots, \theta_n + 2\pi\mathbb{Z}).$$

Next we want to analyze the dependence of this construction with respect to $c$. Fix $\bar{c} \in \mathbb{R}^n$ and consider a neighborhood $\mathcal{O}$ of the submanifold $N_{\bar{c}}$ in the cotangent space $T^*M$. Being $N_{\bar{c}}$ compact, in $\mathcal{O}$ we have a foliation of invariant tori $N_c$, for $c$ close to $\bar{c}$. In other words $(c_1, \ldots, c_n, \theta_1, \ldots, \theta_n)$ is a well defined coordinate set on $\mathcal{O}$.

**Theorem 5.28** (Arnold-Liouville)**.** *Let us consider the map $h : T^*M \to \mathbb{R}^n$ associated with a completely integrable system such that all $c \in \mathbb{R}^n$ are regular values for $h$ and every level set $N_c$ is compact and connected. Then for every $\bar{c} \in \mathbb{R}^n$ there exists a neighborhood $\mathcal{O}$ of $N_{\bar{c}}$ and a change of coordinates*

$$\Phi : h(\mathcal{O}) \to \mathbb{R}^n \times T^n, \qquad \Phi(c_1, \ldots, c_n, \theta_1, \ldots, \theta_n) \mapsto (I_1, \ldots, I_n, \varphi_1, \ldots, \varphi_n), \tag{5.21}$$

*such that*

*(i)* $I = \Phi \circ h$,

*(ii)* $\sigma = \sum_{j=1}^n dI_j \wedge d\varphi_j$.

**Definition 5.29.** The coordinates $(I, \varphi)$ defined in Theorem 5.28 are called *action-angle* coordinates.

*Proof of Theorem 5.28.* In this proof we will use the following notation: for $c = (c_1, \dots, c_n) \in \mathbb{R}^n$, $j = 1, \dots, n$ and $\varepsilon > 0$ we set

(a) $c^{j,\varepsilon} := (c_1, \dots, c_j + \varepsilon, \dots, c_n) \in \mathbb{R}^n$,

(b) $\gamma_i(c)$ as the closed curve in the torus $N_c$ parametrized by the $i$-th angular coordinate $\theta_i$, namely
$$\gamma_i(c) := \{F_c(\theta_1, \dots, \theta_i + \tau, \dots, \theta_n) \in N_c \mid \tau \in [0, 2\pi]\}.$$

(c) $C_i^{j,\varepsilon}$ denotes the cylinder defined by the union of curves $\gamma_i(c^{j,\tau})$, for $0 \le \tau \le \varepsilon$.

Let us first define the coordinates $I_i = I_i(c_1, \dots, c_n)$ by the formula
$$I_i(c) = \frac{1}{2\pi} \int_{\gamma_i(c)} s,$$
where $s$ is the tautological 1-form on $T^*M$. Being $\sigma|_{N_c} \equiv 0$, by Stokes Theorem the variable $I_i$ depends only on the homotopy class of $\gamma_i$.[3]

Let us compute the Jacobian of the change of variables.
$$\begin{aligned}
\frac{\partial I_i}{\partial c_j}(c) &= \frac{1}{2\pi} \frac{\partial}{\partial \varepsilon}\bigg|_{\varepsilon=0} \left( \int_{\gamma_i(c^{j,\varepsilon})} s - \int_{\gamma_i(c)} s \right) \\
&= \frac{1}{2\pi} \frac{\partial}{\partial \varepsilon}\bigg|_{\varepsilon=0} \int_{\partial C_i^{j,\varepsilon}} s \\
&= \frac{1}{2\pi} \frac{\partial}{\partial \varepsilon}\bigg|_{\varepsilon=0} \int_{C_i^{j,\varepsilon}} \sigma \qquad\qquad (\text{where } \sigma = ds) \\
&= \frac{1}{2\pi} \frac{\partial}{\partial \varepsilon}\bigg|_{\varepsilon=0} \int_{c_j}^{c_j+\varepsilon} \int_{\gamma_i(c^{j,\tau})} \sigma(\partial_{c_j}, \partial_{\theta_i}) d\theta_i d\tau \\
&= \frac{1}{2\pi} \int_{\gamma_i(c)} \sigma(\partial_{c_j}, \partial_{\theta_i}) d\theta_i.
\end{aligned}$$

Using that $\partial_{\theta_i} = \sum_{j=1}^n b^{ij}(c) \vec{h}_j$ (see (5.19)) (where $b^{ij}$ are the entries of the inverse matrix of $b_{ij}$) one gets
$$\sigma(\cdot, \partial_{\theta_i}) = \sum_{j=1}^n b^{ij}(c) dh_j. \tag{5.22}$$

Moreover $dh_i = dc_i$ since they define the same coordinate set. Hence
$$\begin{aligned}
\frac{\partial I_i}{\partial c_j}(c) &= \frac{1}{2\pi} \int_{\gamma_i(c)} \sum_{k=1}^n b^{ik}(c) \langle dc_k, \partial_{c_i} \rangle \, d\theta_i \\
&= \frac{1}{2\pi} \int_{\gamma_i(c)} b^{ij}(c) d\theta_i \\
&= b^{ij}(c).
\end{aligned}$$

[3]Hence, in principle, we are free to choose any basis $\gamma_1, \dots, \gamma_n$ for the fundamental group of $T^n$.

Combining the last identity with (5.22) one gets

$$\sigma(\cdot, \partial_{\theta_i}) = dI_i.$$

In particular this implies that the symplectic form has the following expression in the coordinates $(I, \theta)$

$$\sigma = \sum_{i,j=1}^{n} a_{ij}(I) dI_i \wedge dI_j + \sum_{i=1}^{n} dI_i \wedge d\theta_i, \tag{5.23}$$

where the smooth functions $a_{ij}$ depends only on the action variables, since the symplectic form $\sigma$ and the term $\sum_{i=1}^{n} dI_i \wedge d\theta_i$ are closed form. Moreover it is easy to see that the first term of (5.23) can be rewritten as

$$\sum_{i,j=1}^{n} a_{ij}(I) dI_i \wedge dI_j = d\left(\sum_{i=1}^{n} \beta_i(I)\right) \wedge dI_i,$$

and $\sigma$ can be rewritten as

$$\sigma = \sum_{i=1}^{n} dI_i \wedge d(\theta_i - \beta_i(I)).$$

The proof is completed by setting $\varphi_i := \theta_i - \beta_i(I)$. ☐

*Remark* 5.30. This proves that there exists a regular foliation of the phase space by invariant manifolds, that are actually tori, such that the Hamiltonian vector fields associated to the invariants of the foliation span the tangent distribution.

In other words, there exist special sets of canonical coordinates on the phase space such that the invariant tori are the level sets of the action variables. Moreover the angle variables are the natural periodic coordinates on the tori. The motion on the invariant tori, expressed in terms of these canonical coordinates, is linear in the angle variables. Indeed, since the $h_j$ are functions on $I$ variables only, we have

$$\vec{h}_j = \sum_{i=1}^{n} \frac{\partial h_j}{\partial I_i} \partial_{\varphi_i}.$$

In other words, the Hamiltonian system defined by $h_j$ in the angle-action coordinate $(I, \varphi)$ is written as follows

$$\dot{I}_i = -\frac{\partial h_j}{\partial \varphi_i} = 0, \qquad \dot{\varphi}_i = \frac{\partial h_j}{\partial I_i}. \tag{5.24}$$

This explains also why this property is called *complete integrability.* The Hamitonian equation in these coordinates can indeed be solved explicitly.

## 5.6 Geodesic flows on quadrics

In this chapter we prove that the geodesic flow on an ellipsoid is completely integrable. To do this, we specify the discussion of Section 5.2 to the case when the function $a$ is a quadratic polynomial, i.e., every level set of our function is a quadric in $\mathbb{R}^n$.

**Definition 5.31.** Let $A$ be an $n \times n$ non degenerate symmetrix matrix. The *quadric* $\mathcal{Q}$ associated with $A$ is the set

$$\mathcal{Q} = \{x \in \mathbb{R}^n,\ \langle A^{-1}x, x\rangle = 1\}. \tag{5.25}$$

For simplicity we deal with the case when $A$ has simple distinct eigenvalues. We denote by $\Lambda = \{\alpha_1, \ldots, \alpha_n\}$ the set of eigenvalues of $A$, ordered in such a way that $\alpha_1 < \ldots < \alpha_n$. Define, for every $\lambda \in \mathbb{R} \setminus \Lambda$,

$$a_\lambda(x) = \langle (A - \lambda I)^{-1} x, x \rangle, \qquad \mathcal{Q}_\lambda = \{x \in \mathbb{R}^n,\ a_\lambda(x) = 1\}.$$

We say that $\{\mathcal{Q}_\lambda\}_{\lambda \in \mathbb{R} \setminus \Lambda}$ is a family of *confocal quadrics*. Observe that, if $A = \mathrm{diag}(\alpha_1, \ldots, \alpha_n)$ is a diagonal matrix, then (5.25) reads

$$\mathcal{Q} = \{x \in \mathbb{R}^n,\ \sum_{i=1}^n \frac{x_i^2}{\alpha_i} = 1\},$$

and $\mathcal{Q}_\lambda$ represents the family of quadrics that are confocal to $\mathcal{Q}$

$$\mathcal{Q}_\lambda = \left\{ x \in \mathbb{R}^n,\ \sum_{i=1}^n \frac{x_i^2}{\alpha_i - \lambda} = 1 \right\}, \qquad \forall\, \lambda \in \mathbb{R} \setminus \Lambda,$$

Notice moreover that $\mathcal{Q}_\lambda$ is an ellipsoid only if $\lambda < \alpha_1$, while $\mathcal{Q}_\lambda = \emptyset$ when $\lambda > \alpha_n$.

**Note.** In what follows by a "generic" point $x$ for $A$ we mean a point $x$ that does not belong to any proper invariant subspace of $A$. In the diagonal case it is equivalent to say that $x = (x_1, \ldots, x_n)$, with $x_i \neq 0$ for every $i = 1, \ldots, n$.

**Exercise 5.32.** Let $A$ be an $n \times n$ non degenerate symmetrix matrix. Denote by $A_\lambda := (A - \lambda I)^{-1}$. Prove the two following formulas:

(i) $\frac{d}{d\lambda} A_\lambda = A_\lambda^2$,

(ii) $A_\lambda - A_\mu = (\mu - \lambda) A_\lambda A_\mu$.

**Lemma 5.33.** *Let $x \in \mathbb{R}^n$ be a generic point for $A$ and let $\{\mathcal{Q}_\lambda\}_{\lambda \in \mathbb{R} \setminus \Lambda}$ be the associated family of confocal quadrics. Then there exists exactly $n$ distinct real numbers $\lambda_1, \ldots, \lambda_n$ in $\mathbb{R} \setminus \Lambda$ such that $x \in \mathcal{Q}_{\lambda_i}$ for every $i = 1, \ldots, n$. Moreover the quadrics $\mathcal{Q}_{\lambda_i}$ are pairwise orthoghonal at the point $x$.*

*Proof.* For a fixed $x$, the function $\lambda \mapsto a_\lambda(x) = \langle A_\lambda x, x \rangle$ satisfies in $\mathbb{R} \setminus \Lambda$

$$\frac{\partial a_\lambda}{\partial \lambda}(x) = \left\langle A_\lambda^2 x, x \right\rangle = |A_\lambda x|^2 \geq 0,$$

as follows from part (i) of Exercise 5.32 and the fact that $A$ (hence $A_\lambda$) is self-adjoint. Thus $a_\lambda(x)$ is monotone increasing as a function of $\lambda$, and $a_\lambda(x)$ goes to zero for $|\lambda| \to \infty$. In particular it takes values from 0 to $+\infty$ on the interval $]-\infty, \alpha_1[$, while it takes values from $-\infty$ to $+\infty$ in each interval $]\alpha_i, \alpha_{i+1}[$ for $i = 1, \ldots, n-1$. Notice that it is negative on $]\alpha_n, +\infty[$. See also Figure 5.3.

This implies that, for a fixed $x$, there exist exactly $n$ values $\lambda_1, \ldots, \lambda_n$ such that $a_{\lambda_i}(x) = 1$ (i.e., $x \in \mathcal{Q}_{\lambda_i}$).

Next, using part (ii) of Exercise 5.32 (also known as *resolvent formula*) we can compute, for two distinct values $\lambda_i \neq \lambda_j$ and $x \in \mathcal{Q}_{\lambda_i} \cap \mathcal{Q}_{\lambda_j}$:

$$\begin{aligned}
\left\langle \nabla_x a_{\lambda_i}, \nabla_x a_{\lambda_j} \right\rangle &= 4 \left\langle A_{\lambda_i} x, A_{\lambda_j} x \right\rangle \\
&= 4 \left\langle A_{\lambda_i} A_{\lambda_j} x, x \right\rangle \\
&= \frac{4}{\lambda_j - \lambda_i} \left( \langle A_{\lambda_i} x, x \rangle - \left\langle A_{\lambda_j} x, x \right\rangle \right) = 0,
\end{aligned}$$

where again we used the fact that $A_\lambda$ is selfadjoint and $\langle A_\lambda x, x \rangle = 1$ for all $\lambda$. $\square$

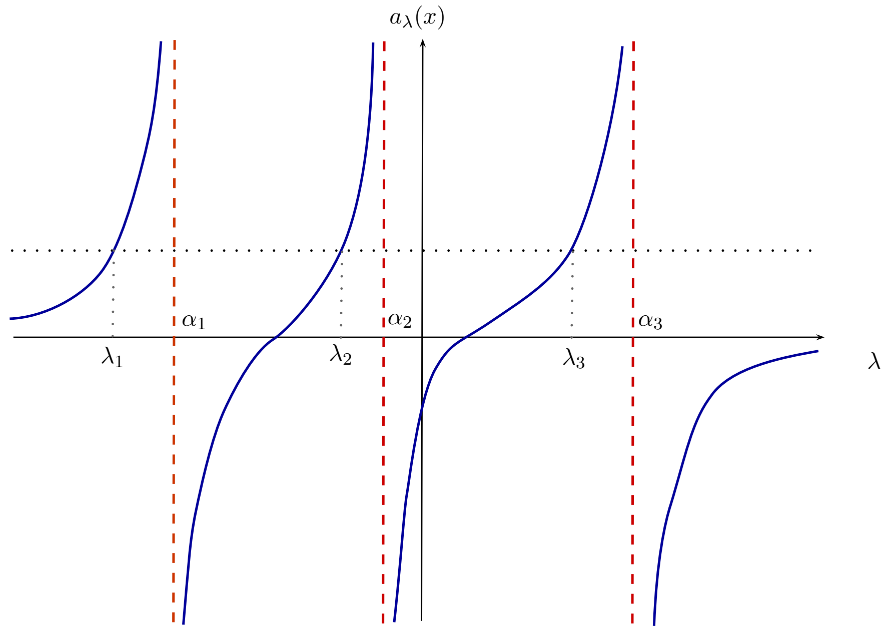



Figure 5.3: A qualitative picture of the function $\lambda \mapsto a_\lambda(x)$ for $n = 3$.

Next, we want to generalize the considerations of the previous sections to all quadrics associated to $\lambda \in \mathbb{R} \setminus \Lambda$, and not only to ellipsoids. Notice indeed that the family of Hamiltonians associated with the confocal quadrics

$$h_\lambda(x, p) = \min_t a_\lambda(x + tp) = a_\lambda(\xi_\lambda(x, p)), \tag{5.26}$$

are well-defined only if the corresponding quadric is an ellipsoid. Indeed in this case the minimum in (5.27) is attained at a unique point $\xi_\lambda(x, p)$, and the function $a_\lambda$ satisfies the assumptions (A1)-(A2) introduced in Section 5.2.2.

To generalize this, we define the Hamiltonian $h_\lambda$ as the value

$$h_\lambda(x, p) = a_\lambda(\xi_\lambda(x, p)), \tag{5.27}$$

of the function $a_\lambda$ at its *critical* point $\xi_\lambda(x, p)$ along the affine line $x + \mathbb{R}p$, still defining $h_\lambda$ as an Hamiltonian on the set of affine lines.

With this generalized definition, we prove an interesting "orthogonality" property of the family. We show that if two confocal quadrics are tangent to the same line, then their gradient are orthogonal at the tangency points.

**Proposition 5.34.** *Assume that two confocal quadrics $\mathcal{Q}_\lambda$ and $\mathcal{Q}_\mu$ are tangent to a given line, i.e., there exist $x, p \in \mathbb{R}^n$ such that*

$$a_\lambda(\xi_\lambda) = a_\mu(\xi_\mu), \qquad \text{where} \qquad \xi_\lambda = x + t_\lambda p, \quad \xi_\mu = x + t_\mu p.$$

*Then $\langle \nabla_{\xi_\lambda} a_\lambda, \nabla_{\xi_\mu} a_\mu \rangle = 0$. In particular $\{h_\lambda, h_\mu\} = 0$.*

*Proof.* The condition that the quadric $\mathcal{Q}_\lambda$ is tangent to the line $x + \mathbb{R}p$ at $\xi_\lambda$ is expressed by the following two equality

$$\langle A_\lambda \xi_\lambda, y\rangle = 0, \qquad \langle A_\lambda \xi_\lambda, \xi_\lambda\rangle = 1, \tag{5.28}$$

and an analogue relations is valid for $\mathcal{Q}_\mu$. Notice than from (5.28) one also gets $\langle A_\lambda \xi_\lambda, \xi_\mu\rangle = \langle A_\mu \xi_\mu, \xi_\lambda\rangle = 1$. Then, with the same computation as before using (5.32)

$$\begin{aligned}\left\langle \nabla_{\xi_\lambda} a_\lambda, \nabla_{\xi_\mu} a_\mu\right\rangle &= 4\left\langle A_\lambda \xi_\lambda, A_\mu \xi_\mu\right\rangle \\ &= 4\left\langle A_\lambda A_\mu \xi_\lambda, \xi_\mu\right\rangle \\ &= \frac{4}{\mu-\lambda}(\langle A_\lambda \xi_\lambda, \xi_\mu\rangle - \langle A_\mu \xi_\mu, \xi_\lambda\rangle) = 0.\end{aligned}$$

This implies also $\{h_\lambda, h_\mu\} = 0$, thanks to Proposition 5.12. □

**Proposition 5.35.** *A generic line in $\mathbb{R}^n$ is tangent to $n-1$ quadrics of a confocal family.*

*Proof.* Write $\mathbb{R}^n = L \oplus L^\perp$ where $L = x + \mathbb{R}p$ and $L^\perp$ is the orthogonal hyperplane (passing through $x$). Consider the orthogonal projection $\pi : \mathbb{R}^n \to L^\perp$ in the direction of $L$. The following exercise shows that the projection of a confocal family of quadrics in $\mathbb{R}^n$ is a confocal family of quadrics on $L^\perp$.

**Exercise 5.36.** (i). Show that the map $x \mapsto a^p_\lambda(x) := \langle A_\lambda(x + t_\lambda p), x + t_\lambda p\rangle$ is a quadratic form and that $p \in \ker a^p_\lambda$. In particular this implies that $a^p_\lambda$ is well defined on the quotient $\mathbb{R}^n/\mathbb{R}p$.
(ii). Prove that $\{a^p_\lambda\}_\lambda$ is a family of confocal quadrics on the factor space (in $n-1$ variables).

Applying then Lemma 5.33 to the family $\{a^p_\lambda\}_\lambda$ we get that, for a generic choice of $x$, there exists $n-1$ quadrics passing through the point on the plane where the line is projected, i.e., the line $x + \mathbb{R}p$ is tangent to $n-1$ confocal quadrics of the family $\{a_\lambda\}_\lambda$. □

*Remark* 5.37. Notice that this proves that every generic line in $\mathbb{R}^n$ is associated with an orthonormal frame of $\mathbb{R}^n$, being all the normal vectors to the $n-1$ quadrics given by Proposition 5.35 mutually orthogonal and orthogonal to the line itself.

**Theorem 5.38.** *The geodesic flow on an ellipsoid is completely integrable. Moreover, for a given geodesic on an ellipsoid, the set of lines are tangent to the geodesics ara tangent to the same set of confocal quadrics, i.e., the set is independent on the point on the geodesic.*

*Proof.* We want to show that the functions $\lambda_1(x,p), \dots, \lambda_{n-1}(x,p)$ (as functions defined on the set of oriented affine lines in $\mathbb{R}^n$) that assign to each line $x + \mathbb{R}p$ in $\mathbb{R}^n$ the $n-1$ values of $\lambda$ such that the line is tangent to $\mathcal{Q}_\lambda$ are independent and in involution.

First notice that each level set $\{\lambda_i(x,p) = c\}$ coincides with the level set $\{h_c = 1\}$. Hence, by Exercise 4.31, the two functions defines the same Hamiltonian flow on this level set (up to reparametrization). We are then reduced to prove that the functions $h_{c_1}, \dots, h_{c_{n-1}}$ are independent and in involution, which is a consequence of Proposition 5.34.

Since the lines that are tangent to a geodesic on the ellipsoid $\mathcal{Q}_\lambda$ form an integral curve of the Hamiltonian flow of the associated function $h_\lambda$, and all the Poisson brackets with the other Hamiltonians are zero, it follows that the line remains tangent to the same set of $n-1$ quadrics. □

## 5.7 Bibliographical note

The notion of complete integrability introduced here is the classical one given by Liouville and Arnold [Arn89]. Sometimes, complete integrability of a dynamical system is also referred to systems whose solution can be reduced to a sequence of quadratures. This means that, even if the solution is implicitly given by some inverse function or integrals, one does not need to solve any differential equation. Notice that by Theorem 5.28 complete integrability implies integrability by quadratures (see also Remark 5.30).

The complete integrability of the geodesic flow on the triaxial ellipsoid was established by Jacobi in 1838. Jacobi integrated the geodesic flow by separation of variables, see [Jac39]. The appropriate coordinates are called the elliptic coordinates, and this approach works in any dimension. Here we give a different derivation, essentially due to Moser [Mos80b], as an application of the theory developed in the first sections of the chapter. For further discussions on the geodesic flow on the ellipsoids or quadrics, one can see [Mos80a, Aud94, Knö80].

The theory of integrable systems has become an independent and extremely rich direction of research in dynamical systems. A discussion of all the aspects related to this beautiful theory goes beyond the scope of this book.

# Chapter 6

# Chronological calculus

In this chapter we develop a language, called *chronological calculus*, that will allow us to work in an efficient way with flows of nonautonomous vector fields. In fact, one of the main goals of the formalism, is to provide suitable tools to expand flows of nonautonomous vector fields in infinite Volterra series. These series, at first introduced as formal asymptotic expansions, are then shown to converge under certain assumptions.

The basic idea of chronological calculus is to replace a non-linear finite-dimensional object, the manifold $M$, with a linear infinite-dimensional one, the commutative algebra $C^\infty(M)$, cf. Section 6.2.

## 6.1 Motivation

Classical formulas from calculus that are valid in $\mathbb{R}^n$ are often no more meaningful on a smooth manifold, unless one consider them as written in coordinates.

Let us consider for instance a smooth curve $\gamma : [0,T] \to \mathbb{R}^n$. The fundamental theorem of calculus states that, for every $t \in [0,T]$, one has

$$\gamma(t) = \gamma(0) + \int_0^t \dot\gamma(s)\, ds. \tag{6.1}$$

Formula (6.1) has no meaning a priori if $\gamma$ takes values on a smooth manifold $M$. Indeed, if $\gamma : [0,T] \to M$, then $\dot\gamma(s) \in T_{\gamma(s)}M$ and one should integrate a family of tangent vectors belonging to different tangent spaces. Moreover, since $M$ has no affine space structure, one should define what is the sum of a point on $M$ with a tangent vector.

Saying that formula (6.1) is meaningful in coordinates means that, once we identify an open set $U$ on $M$ with $\mathbb{R}^n$ through a coordinate map $\phi : U \subset M \to \mathbb{R}^n$ (a set of $n$ independent scalar functions $\phi = (\phi_1, \ldots, \phi_n)$), we reduce (6.1) to $n$ scalar identities.

In fact, it is not necessary to choose a specific set of coordinate functions to let (6.1) have a meaning. The basic idea behind the formalism we introduce in this chapter is that formula (6.1) has a meaning along *any* scalar function, treating this function as the object where the formula is "evaluated".

More formally, let us fix a smooth curve $\gamma : [0,T] \to M$ and a smooth function $a : M \to \mathbb{R}$ and let us apply the fundamental theorem of calculus to the scalar function $a \circ \gamma : [0,T] \to \mathbb{R}$. We get, for every $t \in [0, T]$ the identity

$$a(\gamma(t)) = a(\gamma(0)) + \int_0^t \left\langle d_{\gamma(s)} a, \dot{\gamma}(s) \right\rangle ds. \tag{6.2}$$

Formula (6.2) is meaningful even if $\gamma$ takes values on a manifold since it is a scalar identity. The integrand is the duality product between $d_{\gamma(s)} a \in T^*_{\gamma(s)} M$ and $\dot{\gamma}(s) \in T_{\gamma(s)} M$.

If we think to a point on $M$ as acting on a function by evaluating the function at that point, and to a tangent vector as acting on a function by differentiating the function in the direction of the vector, then we can think to (6.2) as formula (6.1) when "evaluated at $a$", or at (6.2) as the coordinate version of (6.1). If we choose as $a$ the functions $\phi_i$ for $i = 1, \dots, n$ we are writing the coordinate version of the identity in the classical sense.

In what follows we develop in a formal way this flexible language that has the advantage of computing things "as in coordinates" keeping track of the geometric meaning of the object we are dealing with.

## 6.2 Duality

The set $C^\infty(M)$ of smooth functions on $M$ is an $\mathbb{R}$-algebra with the usual operation of pointwise addition and multiplication

$$\begin{aligned} (a+b)(q) &= a(q) + b(q), \\ (\lambda a)(q) &= \lambda\, a(q), \qquad\qquad a, b \in C^\infty(M),\ \lambda \in \mathbb{R}, q \in M. \\ (a \cdot b)(q) &= a(q) b(q). \end{aligned}$$

Any point $q \in M$ can be interpreted as the "evaluation" linear functional

$$\widehat{q} : C^\infty(M) \to \mathbb{R}, \qquad \widehat{q}(a) := a(q).$$

For every $q \in M$, the functional $\widehat{q}$ is a homomorphism of algebras, i.e., it satisfies

$$\widehat{q}(a \cdot b) = \widehat{q}(a) \widehat{q}(b).$$

A diffeomorphism $P \in \mathrm{Diff}(M)$ can be thought as the "change of variables" linear operator

$$\widehat{P} : C^\infty(M) \to C^\infty(M), \qquad \widehat{P}(a) := a \circ P,$$

which is an automorphism of the algebra $C^\infty(M)$.

*Remark* 6.1. One can prove that for every non-trivial homomorphism of algebras $\varphi : C^\infty(M) \to \mathbb{R}$ there exists $q \in M$ such that $\varphi = \widehat{q}$. Analogously, for every automorphism of algebras $\Phi : C^\infty(M) \to C^\infty(M)$, there exists a diffeomorphism $P \in \mathrm{Diff}(M)$ such that $\widehat{P} = \Phi$. A proof of these facts is contained in [AS04, Appendix A].

Next, we want to characterize tangent vectors as functionals on $C^\infty(M)$. As explained in Chapter 2, a tangent vector $v \in T_q M$ defines in a natural way the derivation in the direction of $v$, i.e., the functional

$$\widehat{v} : C^\infty(M) \to \mathbb{R}, \qquad \widehat{v}(a) = \langle d_q a, v \rangle,$$

which satisfies the Leibniz rule

$$\widehat{v}(a \cdot b) = \widehat{v}(a)b(q) + a(q)\widehat{v}(b), \qquad \forall\, a, b \in C^\infty(M). \tag{6.3}$$

If $v \in T_qM$ is the tangent vector of a curve $q(t)$ such that $q(0) = q$, it is also natural to check the identity as operators

$$\widehat{v} = \left.\frac{d}{dt}\right|_{t=0} \widehat{q}(t) : C^\infty(M) \to \mathbb{R}. \tag{6.4}$$

Indeed, it is sufficient to differentiate at $t = 0$ the following identity

$$\widehat{q}(t)(a \cdot b) = \widehat{q}(t)a \cdot \widehat{q}(t)b.$$

In the same spirit, a vector field $X \in \mathrm{Vec}(M)$ is characterized, as a derivation of $C^\infty(M)$ (cf. again the discussion in Chapter 2), as the infinitesimal version of a flow (i.e., family of diffeomorphisms smooth w.r.t $t$) $P_t \in \mathrm{Diff}(M)$. Indeed if we set

$$\widehat{X} = \left.\frac{d}{dt}\right|_{t=0} \widehat{P}_t : C^\infty(M) \to C^\infty(M),$$

we find that $\widehat{X}$ satisfies (see (2.14))

$$\widehat{X}(ab) = \widehat{X}(a)b + a\widehat{X}(b), \qquad \forall\, a, b \in C^\infty(M).$$

### 6.2.1 On the notation

In the following we will identify any object with its dual interpretation as operator on functions and stop to use a different notation for the same object when acting on the space of smooth functions.

If $P$ is a diffeomorphism on $M$ and $q$ is a point on $M$ the point $P(q)$ is simply represented by the usual composition $\widehat{q} \circ \widehat{P}$ of the corresponding linear operator.

Thus, when using the operator notation, composition works in the opposite side. To simplify the notation in what follows we will remove the "hat" identifying an object with its dual, but use the symbol $\odot$ to denote the composition of these object, so that $P(q)$ will be $q \odot P$.

Analogously, the composition $X \odot P$ of a vector field $X$ and a diffeomorphism $P$ will denote the linear operator $a \mapsto X(a \circ P)$.

## 6.3 Topology on the set of smooth functions

We introduce the standard topology on the space $C^\infty(M)$. Denote by $X_1, \ldots, X_r$ a family of globally defined vector fields such that

$$\mathrm{span}\{X_1, \ldots, X_r\}|_q = T_qM, \qquad \forall\, q \in M.$$

For $\alpha \in \mathbb{N}$ and $K \subset M$ compact, define the following seminorms of a function $f \in C^\infty(M)$

$$\|f\|_{\alpha,K} = \sup_{q \in K}\{|(X_{i_\ell} \odot \cdots \odot X_{i_1} f)(q)| : 1 \leq i_1, \ldots, i_\ell \leq r, 0 \leq \ell \leq \alpha\}. \tag{6.5}$$

The family of seminorms $\|\cdot\|_{\alpha,K}$ induces a topology on $C^\infty(M)$ with countable local basis of neighborhoods as follows: take an increasing family of compact sets $\{K_n\}_{n\in\mathbb{N}}$ invading $M$, i.e.,

$K_n \subset K_{n+1} \subset M$ for every $n \in \mathbb{N}$ and $M = \cup_{n\in\mathbb{N}} K_n$. For every $f \in C^\infty(M)$, a countable local basis of neighborhoods of $f$ is given by

$$U_{f,n} := \left\{ g \in C^\infty(M) : \|f - g\|_{n,K_n} \leq \frac{1}{n} \right\}, \qquad n \in \mathbb{N}. \tag{6.6}$$

**Exercise 6.2.** (i) Prove that (6.6) defines a basis for a topology. (ii) Prove that this topology does not depend neither on the family of vector fields $X_1, \ldots, X_r$ generating the tangent space to $M$ nor on the family of compact sets $\{K_n\}_{n\in\mathbb{N}}$ invading $M$.

This topology turns $C^\infty(M)$ into a Fréchet space, i.e., a complete, metrizable, locally convex topological vector space, see [Hir76, Chapter 2].

*Remark* 6.3. In differential topology this is also called weak topology on $C^\infty(M)$, in contrast with the *strong* (or *Whitney*) *topology* that can be defined on $C^\infty(M)$. The two topology coincide when the manifold $M$ is compact. For more details about different topologies on the spaces $C^k(M, N)$ of $C^k$ maps among two smooth manifolds $M$ and $N$ we refer to [Hir76, Chapter 2].

**Exercise 6.4.** Prove that, given a diffeomorphism $P \in \mathrm{Diff}(M)$ and $\alpha \in \mathbb{N}$ and a compact set $K \subset M$, there exists a constant $C_{\alpha,P,K} > 0$ such that for all $f \in C^\infty(M)$ one has

$$\|Pf\|_{\alpha,K} \leq C_{\alpha,P,K} \|f\|_{\alpha,P(K)}. \tag{6.7}$$

The previous exercice says that a diffeomorphism $P$, when interpreted as a linear operator on $C^\infty(M)$, is continuous in the Whitney topology. One can then define its seminorms

$$\|P\|_{\alpha,K} := \sup\{\|Pf\|_{\alpha,K} : \|f\|_{\alpha,P(K)} \leq 1\}.$$

Similarly, given a smooth vector field $X$ on $M$, one defines its seminorms by

$$\|X\|_{\alpha,K} := \sup\{\|Xf\|_{\alpha,K} : \|f\|_{\alpha+1,K} \leq 1\}.$$

### 6.3.1 Family of functionals and operators

Once the structure of a Fréchet space on $C^\infty(M)$ is given, one can define the regularity properties of family of functions in $C^\infty(M)$. In particular continuous and differentiable families of functions $t \mapsto a_t$ are defined in a standard way. Moreover, we say that the family $t \mapsto a_t \in C^\infty(M)$ defined on an interval $[t_0, t_1]$ is

- *measurable*, if the map $q \mapsto a_t(q)$ is measurable on $[t_0, t_1]$ for every $q \in M$.
- *locally integrable*, if for every $\alpha \in \mathbb{N}$ and $K \subset M$ compact one has

$$\int_{t_0}^{t_1} \|a_t\|_{\alpha,K} dt < +\infty.$$

- *absolutely continuous*, if there exists a locally integrable family of functions $b_t$ such that

$$a_t = a_{t_0} + \int_{t_0}^{t} b_s ds.$$

- *locally Lipschitz*, if for every $\alpha \in \mathbb{N}$ and $K \subset M$ compact there exists $C_{\alpha,K} > 0$ such that
$$\|a_t - a_s\|_{\alpha,K} \leq C_{\alpha,K}|t-s|.$$

Analogous regularity properties for a family of linear functionals (or linear operators) on $C^\infty(M)$ are then naturally defined in a weak sense: we say that a family of operators $t \mapsto A_t$ is continuous (differentiable, etc.) if the map $t \mapsto A_t a$ has the same property for every $a \in C^\infty(M)$.

Recall that a *nonautonomous vector field*, in the sense of Definition 2.13, is in particular a family of smooth vector fields $X_t$ that is measurable and locally bounded with respect to $t$, in the sense explained above. A *nonautonomous flow* is a family of smooth diffeomorphisms $P_t$ that is absolutely continuous with respect to $t$. The flow generated by a nonautonomous vector field is a particular case of nonautonomous flow. See also Section 2.1.4.

For any nonautonomous vector field $X_t$, the family of functions $t \mapsto X_t a$ is locally integrable for any $a \in C^\infty(M)$. Similarly, for any nonautonomous flow $P_t$ the family of functions $t \mapsto a \circ P_t$ is absolutely continuous for any $a \in C^\infty(M)$.

Integrals of measurable locally integrable families, and derivative of differentiable families are also defined in the weak sense: for instance, if $X_t$ denotes some locally integrable nonautonomous vector field we define
$$\int_0^t X_s\, ds : a \mapsto \int_0^t X_s a\, ds,$$
$$\frac{d}{dt}X_t : a \mapsto \frac{d}{dt}(X_t a).$$
One can show that if $A_t$ and $B_t$ are continuous families of operators on $C^\infty(M)$ which are differentiable at $t_0$, then the family $A_t \circ B_t$ is differentiable at $t_0$ and satisfies the Leibniz rule
$$\left.\frac{d}{dt}\right|_{t=t_0}(A_t \circ B_t) = \left(\left.\frac{d}{dt}\right|_{t=t_0} A_t\right) \circ B_{t_0} + A_{t_0} \circ \left(\left.\frac{d}{dt}\right|_{t=t_0} B_t\right). \tag{6.8}$$
The same result holds true for the composition of functionals with operators. For a proof of the last fact one can see [AS04, Chapter 2 and Appendix A].

## 6.4 Operator ODEs and Volterra expansions

Consider a nonautonomous vector field $X_t$ and the corresponding nonautonomous ODE on $M$
$$\frac{d}{dt}q(t) = X_t(q(t)). \tag{6.9}$$
Using the notation introduced in the previous section we can rewrite (6.9) in the following way
$$\frac{d}{dt}q(t) = q(t) \circ X_t. \tag{6.10}$$
Indeed assume that $q(t)$ satisfies (6.9) and let $a \in C^\infty(M)$. Using the "hat" notation of Section 6.2.1
$$\left(\frac{d}{dt}\widehat{q}(t)\right) a = \frac{d}{dt}\widehat{q}(t)a = \frac{d}{dt}a(q(t)) = \left\langle d_{q(t)}a, X_t(q(t))\right\rangle = (\widehat{X}_t a)(q(t)) = (\widehat{q}(t) \circ \widehat{X}_t)a. \tag{6.11}$$

which is then re-written in form (6.10) once we have removed the "hat" and replaced the composition sign by $\odot$, following Section 6.2.1. As discussed in Chapter 2, the solution to the nonautonomous ODE (6.9) defines a flow, i.e., a family of diffeomorphisms $P_{s,t}$. Reasoning as in (6.11) one immediately gets the following lemma.

**Lemma 6.5.** *The flow $P_{s,t}$ defined by* (6.9) *satisfies the operator differential equation*

$$\frac{d}{dt}P_{s,t} = P_{s,t} \odot X_t, \qquad P_{s,s} = \mathrm{Id}. \tag{6.12}$$

**Definition 6.6.** We call $P_{s,t}$ the *right chronological exponential* and use the notation

$$\overrightarrow{\exp} \int_s^t X_\tau d\tau := P_{s,t}. \tag{6.13}$$

### 6.4.1 Volterra expansion

In the following discussion we set for simplicity the initial time $s = 0$. In this case we use the short notation $P_t := P_{0,t}$. The operator differential equation (6.12) rewrites as

$$\begin{cases} \dot{P}_t = P_t \odot X_t \\ P_0 = \mathrm{Id} \end{cases} \tag{6.14}$$

and can be rewritten as an integral operator equation as follows

$$P_t = \mathrm{Id} + \int_0^t P_s \odot X_s ds. \tag{6.15}$$

Replacing iteratively $P_s$ in the right hand side of (6.15) with the equation (6.15) itself, we have

$$\begin{aligned} P_t &= \mathrm{Id} + \int_0^t \left( \mathrm{Id} + \int_0^{s_1} P_{s_2} \odot X_{s_2} ds_2 \right) \odot X_{s_1} ds_1 \\ &= \mathrm{Id} + \int_0^t X_s ds + \iint\limits_{0 \le s_2 \le s_1 \le t} P_{s_2} \odot X_{s_2} \odot X_{s_1} ds_1 ds_2 \\ &\vdots \\ &= \mathrm{Id} + \sum_{k=1}^{N-1} \int\limits_{0 \le s_k \le \ldots \le s_1 \le t} \cdots \int X_{s_k} \odot \cdots \odot X_{s_1} ds_1 \cdots ds_k + R_N, \end{aligned}$$

where the remainder term is defined as follows

$$R_N := \int\limits_{0 \le s_N \le \ldots \le s_1 \le t} \cdots \int P_{s_N} \odot X_{s_N} \odot \cdots \odot X_{s_1} ds_1 \cdots ds_N. \tag{6.16}$$

Formally, letting $N \to \infty$ and assuming that $R_N \to 0$, we can write the flow $P_t$ as the chronological series

$$P_t = \overrightarrow{\exp} \int_0^t X_s ds \simeq \mathrm{Id} + \sum_{k=1}^{\infty} \int\limits_{\Delta_k(t)} \cdots \int X_{s_k} \odot \cdots \odot X_{s_1} ds_1 \cdots ds_k, \tag{6.17}$$

where $\Delta_k(t) = \{(s_1, \ldots, s_k) \in \mathbb{R}^k : 0 \leq s_k \leq \ldots \leq s_1 \leq t\}$ denotes the $k$-dimensional symplex.

A discussion about an estimate of the remainder term (6.16) and the convergence of the series in the right hand side of (6.17) is contained in Section 6.6.

*Remark* 6.7. If we write expansion (6.17) when $X_t = X$ is an autonomous vector field, we find that the chronological exponential coincides with the exponential of the vector field

$$\overrightarrow{\exp}\int_0^t X ds \simeq \mathrm{Id} + \sum_{k=1}^{\infty} \int\cdots\int_{\Delta_k(t)} \underbrace{X \circ \cdots \circ X}_{k} ds_1 \cdots ds_k$$
$$= \sum_{k=0}^{\infty} \mathrm{vol}(\Delta_k(t)) X^k = \sum_{k=0}^{\infty} \frac{t^k}{k!} X^k = e^{tX},$$

since $\mathrm{vol}(\Delta_k(t)) = t^k/k!$. In the nonautonomous case for different times $X_{s_1}$ and $X_{s_2}$ might not commute, hence the order in which the vector fields appear in the composition is crucial.

**Exercise 6.8.** Prove that, in general, for a nonautonomous vector field $X_t$, one has

$$\overrightarrow{\exp}\int_0^t X_s ds \neq e^{\int_0^t X_s ds}. \tag{6.18}$$

Prove that, if $[X_t, X_\tau] = 0$ for all $t, \tau \in \mathbb{R}$, then the equality holds in (6.18).

**Proposition 6.9.** *Assume that $P_t$ satisfies* (6.14) *and consider the inverse flow $Q_t := (P_t)^{-1}$. Then $Q_t$ satisfies the Cauchy problem*

$$\begin{cases} \dot{Q}_t = -X_t \circ Q_t, \\ Q_0 = \mathrm{Id}. \end{cases} \tag{6.19}$$

*Proof.* By definition of inverse flow we have the identity for every $t \in \mathbb{R}$

$$P_t \circ Q_t = \mathrm{Id}. \tag{6.20}$$

In particular $Q_0 = \mathrm{Id}$. Moreover, differentiating (6.20)with respect to $t$ and using the Leibniz rule one obtains

$$\dot{P}_t \circ Q_t + P_t \circ \dot{Q}_t = 0. \tag{6.21}$$

Using (6.14) then we get

$$P_t \circ X_t \circ Q_t + P_t \circ \dot{Q}_t = 0. \tag{6.22}$$

Multiplying both sides by $Q_t$ on the left, one gets (6.19). $\square$

The solution to the problem (6.19) will be denoted by the *left chronological exponential*

$$Q_t := \overleftarrow{\exp}\int_0^t (-X_s) ds. \tag{6.23}$$

By an analogous reasoning to what we did for the right chronological exponential, we find the formal expansion

$$\overleftarrow{\exp}\int_0^t (-X_s) ds \simeq \mathrm{Id} + \sum_{k=1}^{\infty} \int\cdots\int_{0 \leq s_k \leq \ldots \leq s_1 \leq t} (-X_{s_1}) \circ \cdots \circ (-X_{s_k}) ds_1 \cdots ds_k. \tag{6.24}$$

*Remark* 6.10. The formal difference between right and left chronological exponential is in the order of composition. Notice that the arrow over the exponential says in which "direction" the time parameters are increasing in the chronological series (compare (6.12) and (6.24)), or, equivalently, in which "position" the vector field appears when differentiating the flow (compare (6.14) and (6.19)).

All the properties of the chronological exponential are summarized as follows

$$\frac{d}{dt}\overrightarrow{\exp}\int_0^t X_s ds = \overrightarrow{\exp}\int_0^t X_s ds \circ X_t, \tag{6.25}$$

$$\frac{d}{dt}\overleftarrow{\exp}\int_0^t X_s ds = X_t \circ \overleftarrow{\exp}\int_0^t X_s ds, \tag{6.26}$$

$$\left(\overrightarrow{\exp}\int_0^t X_s ds\right)^{-1} = \overleftarrow{\exp}\int_0^t (-X_s) ds. \tag{6.27}$$

### 6.4.2 Adjoint representation

Next, we study the action of diffeomorphisms on vectors and vector fields. Let $v \in T_qM$ and $P \in \mathrm{Diff}(M)$. We claim that, as functionals on $C^\infty(M)$, we have

$$P_* v = v \circ P.$$

Indeed consider a curve $q(t)$ such that $\dot{q}(0) = v$ and compute

$$(P_* v)a = \left.\frac{d}{dt}\right|_{t=0} a(P(q(t))) = \left(\left.\frac{d}{dt}\right|_{t=0} q(t)\right) \circ Pa = v \circ Pa.$$

Recall that, if $X \in \mathrm{Vec}(M)$ is a vector field we have $P_* X\big|_q = P_*(X\big|_{P^{-1}(q)})$. Hence, the expression for $P_* X$ as derivation of $C^\infty(M)$ is

$$P_* X = P^{-1} \circ X \circ P. \tag{6.28}$$

*Remark* 6.11. We can reinterpret the pushforward of a vector field in a totally algebraic way in the space of linear operator on $C^\infty(M)$. Indeed

$$P_* X = (\mathrm{Ad}\, P^{-1})X, \tag{6.29}$$

where

$$\mathrm{Ad}\, P : X \mapsto P \circ X \circ P^{-1}, \qquad \forall\, X \in \mathrm{Vec}(M), \tag{6.30}$$

is the adjoint action of $P$ on the space of vector fields. Notice that (6.30) is the differential of the conjugation $Q \mapsto P \circ Q \circ P^{-1}$, for $Q \in \mathrm{Diff}(M)$.

Assume now that $P_t = \overrightarrow{\exp}\int_0^t X_s ds$. We try to characterize the flow $\mathrm{Ad}\, P_t$ by looking for the ODE it satisfies. Applying it to a vector field $Y$ we have

$$\begin{aligned}
\left(\frac{d}{dt}\mathrm{Ad}\, P_t\right) Y &= \frac{d}{dt}(\mathrm{Ad}\, P_t)Y = \frac{d}{dt}(P_t \circ Y \circ P_t^{-1}) \\
&= P_t \circ X_t \circ Y \circ P_t^{-1} + P_t \circ Y \circ (-X_t) \circ P_t^{-1} \\
&= P_t \circ (X_t \circ Y - Y \circ X_t) \circ P_t^{-1} \\
&= (\mathrm{Ad}\, P_t)[X_t, Y] \\
&= (\mathrm{Ad}\, P_t)(\mathrm{ad}\, X_t)Y,
\end{aligned}$$

where

$$\operatorname{ad} X : Y \mapsto [X, Y],$$

is the adjoint action on the Lie algebra of vector fields. This proves that $\operatorname{Ad} P_t$ is a solution to the differential equation

$$\dot{A}_t = A_t \circ \operatorname{ad} X_t, \qquad A_0 = \mathrm{Id}.$$

Thus it can be expressed as a chronological exponential and we have the identity

$$\operatorname{Ad}\left(\overrightarrow{\exp}\int_0^t X_s ds\right) = \overrightarrow{\exp}\int_0^t \operatorname{ad} X_s ds. \tag{6.31}$$

Notice that combining (6.31) and (6.29) in the case of an autonomous vector field, one gets

$$e_*^{-tX} = e^{t \operatorname{ad} X}. \tag{6.32}$$

**Exercise 6.12.** Prove that, if $[X_t, Y] = 0$ for all $t$, then $(\operatorname{Ad} P_t)Y = Y$.

*Remark* 6.13. If $P_t = \overrightarrow{\exp}\int_0^t X_s ds$, we can write the following formal expansion

$$(\operatorname{Ad} P_t)Y \simeq Y + \sum_{k=1}^{\infty} \int \cdots \int_{0 \le s_k \le \ldots \le s_1 \le t} [X_{s_n}, \ldots, [X_{s_2}, [X_{s_1}, Y]] ds_1 \cdots ds_k, \tag{6.33}$$

The latter generalizes (2.34) for nonautonomous vector fields. Indeed, if $P_t = e^{tX}$ is the flow associated with an autonomous vector field $X$, one gets (2.34)

$$\begin{aligned}(\operatorname{Ad} e^{tX})Y = e_*^{-tX} Y &= Y + \sum_{k=1}^{\infty} \frac{t^k}{k!}[X, \ldots, [X, Y]] \\ &\simeq Y + t[X, Y] + \frac{t^2}{2}[X, [X, Y]] + o(t^2).\end{aligned}$$

**Exercise 6.14.** Prove the following, using the operator notation:

(a) Show that ad is the infinitesimal version of the operator Ad, i.e., if $P_t$ is the flow generated by the vector field $X \in \operatorname{Vec}(M)$ then

$$\operatorname{ad} X = \left.\frac{d}{dt}\right|_{t=0} \operatorname{Ad} P_t.$$

(b) Show that, if $P \in \operatorname{Diff}(M)$, then $P_*$ preserves Lie brackets, i.e., $P_*[X, Y] = [P_* X, P_* Y]$.

(c) Show that the Jacobi identity in $\operatorname{Vec}(M)$ is the infinitesimal version of the identity proved in question (b). (Hint: choose $P_t = e^{tZ}$ for $Z \in \operatorname{Vec}(M)$.)

**Exercise 6.15.** Prove the following change of variables formula for a nonautonomous flow:

$$P \circ \overrightarrow{\exp}\int_0^t X_s ds \circ P^{-1} = \overrightarrow{\exp}\int_0^t (\operatorname{Ad} P) X_s ds. \tag{6.34}$$

Notice that for an autonomous vector field this identity reduces to (2.26).

## 6.5 Variations formulae

Consider the following ODE

$$\dot{q} = X_t(q) + Y_t(q) \tag{6.35}$$

where $Y_t$ is seen as a perturbation term of the equation (6.9). We want to describe the solution to the perturbed equation (6.35) as the perturbation of the solution of the original one.

**Proposition 6.16.** *Let $X_t, Y_t$ be two nonautonomous vector fields. Then, denoting $P_t = \overrightarrow{\exp} \int_0^t X_s ds$, one has*

$$\overrightarrow{\exp} \int_0^t (X_s + Y_s) ds = \overrightarrow{\exp} \int_0^t \left( \overrightarrow{\exp} \int_0^s \operatorname{ad} X_\tau d\tau \right) Y_s ds \circ \overrightarrow{\exp} \int_0^t X_s ds \tag{6.36}$$

$$= \overrightarrow{\exp} \int_0^t (\operatorname{Ad} P_s) Y_s ds \circ P_t. \tag{6.37}$$

*Proof.* Our goal is to find a flow $R_t$ such that

$$Q_t := \overrightarrow{\exp} \int_0^t (X_s + Y_s) ds = R_t \circ P_t. \tag{6.38}$$

By definition of right chronological exponential we have

$$\dot{Q}_t = Q_t \circ (X_t + Y_t). \tag{6.39}$$

On the other hand, from (6.38), we also have

$$\begin{aligned} \dot{Q}_t &= \dot{R}_t \circ P_t + R_t \circ \dot{P}_t \\ &= \dot{R}_t \circ P_t + R_t \circ P_t \circ X_t \\ &= \dot{R}_t \circ P_t + Q_t \circ X_t. \end{aligned} \tag{6.40}$$

Comparing (6.39) and (6.40), one gets

$$Q_t \circ Y_t = \dot{R}_t \circ P_t,$$

and the ODE satisfied by $R_t$ is

$$\begin{aligned} \dot{R}_t &= Q_t \circ Y_t \circ P_t^{-1} \\ &= R_t \circ (\operatorname{Ad} P_t) Y_t. \end{aligned}$$

Since $R_0 = \mathrm{Id}$, this implies that $R_t$ is a chronological exponential and

$$\overrightarrow{\exp} \int_0^t (X_s + Y_s) ds = \overrightarrow{\exp} \int_0^t (\operatorname{Ad} P_s) Y_s ds \circ P_t,$$

which is (6.37). Plugging (6.31) in (6.37) one gets (6.36). □

**Exercise 6.17.** Prove the following versions of the variation formula:

(i). For every nonautonomous vector fields $X_t, Y_t$ on $M$

$$\overrightarrow{\exp} \int_0^t (X_s + Y_s) ds = \overrightarrow{\exp} \int_0^t X_s ds \circ \overrightarrow{\exp} \int_0^t \left( \overrightarrow{\exp} \int_t^s \operatorname{ad} X_\tau d\tau \right) Y_s ds. \tag{6.41}$$

(ii). For every autonomous vector fields $X, Y \in \mathrm{Vec}(M)$ prove that

$$e^{t(X+Y)} = \overrightarrow{\exp} \int_0^t e^{s\,\mathrm{ad}\,X} Y ds \circ e^{tX} = \overrightarrow{\exp} \int_0^t e_*^{-sX} Y ds \circ e^{tX} \tag{6.42}$$

$$= e^{tX} \circ \overrightarrow{\exp} \int_0^t e^{(s-t)\,\mathrm{ad}\,X} Y ds. \tag{6.43}$$

## 6.6 Appendix: Estimates and Volterra expansion

In this section we discuss the convergence of the Volterra expansion

$$\mathrm{Id} + \sum_{k=1}^{\infty} \int \cdots \int_{\Delta_k(t)} X_{s_k} \circ \cdots \circ X_{s_1} ds_1 \cdots ds_k \tag{6.44}$$

where $\Delta_k(t) = \{(s_1, \dots, s_k) \in \mathbb{R}^k | \, 0 \leq s_k \leq \ldots \leq s_1 \leq t\}$ denotes the $k$-dimensional symplex. Recall that if $X_s = X$ is autonomous then the series (6.44) reduces to

$$\sum_{k=0}^{\infty} \frac{t^k}{k!} X^k. \tag{6.45}$$

We prove the following result, saying that in general, if the vector field is not zero, the chronological exponential is never convergent on the whole space $C^\infty(M)$.

**Proposition 6.18.** *Let $X$ be a smooth vector field that is not identically zero. Then there exists $a \in C^\infty(M)$ such that the Volterra expansion*

$$\sum_{k=0}^{\infty} \frac{t^k}{k!} X^k a \tag{6.46}$$

*is not convergent at some point $q \in M$.*

*Proof.* Fix a point $q \in M$ such that $X(q) \neq 0$. By considering a smooth coordinate chart around $q$ such that $X$ is rectified in this chart, we are then reduced to prove the statement in the case when $M = \mathbb{R}^n$ and $X = \partial_{x_1}$.

Fix then an arbitrary real sequence $(c_n)_{n \in \mathbb{N}}$ and let $f : I \to \mathbb{R}$ be defined in a neighborhood $I$ of 0 such that $f^{(n)}(0) = c_n$, for every $n \in \mathbb{N}$. The existence of such a function is guaranteed by Lemma 6.19 below. Then define $a(x) = f(x_1)$, where $x = (x_1, x') \in \mathbb{R}^n$. In this case $X^k a(q) = \partial_{x_1}^k f(0) = c_k$ and

$$\sum_{k=0}^{\infty} \frac{t^k}{k!} X^k a|_q = \sum_{k=0}^{\infty} \frac{t^k}{k!} c_k \tag{6.47}$$

which is not convergent for a suitable choice of the sequence $(c_n)_{n \in \mathbb{N}}$. ☐

**Lemma 6.19** (Borel lemma)**.** *Let $(c_n)_{n \in \mathbb{N}}$ be a real sequence. Then there exist a $C^\infty$ function $f : I \to \mathbb{R}$ defined in a neighborhood $I$ of 0 such that $f^{(n)}(0) = c_n$ for every $n \in \mathbb{N}$.*

*Proof.* Fix a $C^\infty$ function $\phi : \mathbb{R} \to \mathbb{R}$ with compact support, such that $\phi(0) = 1$ and $\phi^{(j)}(0) = 0$ for every $j \geq 1$. Then set for $k \in \mathbb{N}$

$$g_k(x) := \frac{c_k}{k!} x^k \phi\left(\frac{x}{\varepsilon_k}\right). \tag{6.48}$$

where $(\varepsilon_k)_{k\in\mathbb{N}}$ will be fixed later. Notice that $g_k^{(j)}(0) = \delta_{jk} c_k$, where $\delta_{jk}$ is the Kronecker symbol, and $|g_k^{(j)}(x)| \leq C_{j,k}\varepsilon_k^{k-j}$ for every $x \in \mathbb{R}$ and some constant $C_{j,k} > 0$. Then choose $\varepsilon_k > 0$ in such a way that

$$|g_k^{(j)}(x)| \leq 2^{-j}, \qquad \forall\, j \leq k-1, \forall\, x \in \mathbb{R}, \tag{6.49}$$

and define the function

$$f(x) := \sum_{k=0}^{\infty} g_k(x). \tag{6.50}$$

The series (6.50) converges uniformly with all the derivatives by (6.49) and, by differentiating under the sum, one obtains

$$f^{(j)}(0) := \sum_{k=0}^{\infty} g_k^{(j)}(0) = a_j. \qquad \square$$

Even if, in general, the Volterra expansion is not convergent, it gives a good approximation of the chronological exponential. More precisely, if we denote by

$$S_N(t) := \mathrm{Id} + \sum_{k=1}^{N-1} \int \cdots \int_{\Delta_k(t)} X_{s_k} \circ \cdots \circ X_{s_1} ds_1 \cdots ds_k,$$

the $N$-th partial sum, we have the following estimate.

**Theorem 6.20.** *For every $t > 0$, $\alpha, N \in \mathbb{N}$, $K \subset M$ compact, we have*

$$\left\| \left( \overrightarrow{\exp} \int_0^t X_s ds - S_N(t) \right) a \right\|_{\alpha,K} \leq \frac{C}{N!} e^{C\int_0^t \|X_s\|_{\alpha,K'} ds} \left( \int_0^t \|X_s\|_{\alpha+N-1,K'} ds \right)^N \|a\|_{\alpha+N,K'}, \tag{6.51}$$

*for some $K'$ compact set containing $K$ and some constant $C = C_{\alpha,N,K'} > 0$.*

The proof of this result is postponed to Appendix 6.7. Let us specify this estimate for a nonautonomous vector field of the form

$$X_t = \sum_{i=1}^{m} u_i(t) X_i,$$

where $X_1, \ldots, X_m$ are smooth vector fields on $M$ and $u \in L^2([0,T], \mathbb{R}^m)$. In this case it is convenient to choose the seminorms $\|\cdot\|_{\alpha,K}$, defined in Section 6.3, in terms of a family $X_1, \ldots, X_r$ of smooth vector fields on $M$ which is a completion of $X_1, \ldots, X_m$.

**Theorem 6.21.** *For every $t > 0$, $\alpha, N \in \mathbb{N}$, $K \subset M$ compact, we have (denoting $\|u\|_{1,t} = \|u\|_{L^1([0,t],\mathbb{R}^m)}$)*

$$\left\| \left( \overrightarrow{\exp} \int_0^t \sum_{i=1}^{m} u_i(t) X_i - S_N(t) \right) a \right\|_{\alpha,K} \leq \frac{C}{N!} e^{C\|u\|_{1,t}} \|u\|_{1,t}^N \|a\|_{\alpha+N,K'}, \tag{6.52}$$

*for some $K'$ compact set containing $K$ and some constant $C = C_{\alpha,N,K'} > 0$.*

*Proof.* It follows from the previous theorem and from the fact that for a vector field of the form $X_t = \sum_{i=1}^m u_i(t)X_i$ we have the estimate

$$\int_0^t \|X_s\|_{\alpha,K'} ds \le \|u\|_{L^1([0,t],\mathbb{R}^m)}. \tag{6.53}$$

Indeed we have for every $f$ such that $\|f\|_{\alpha+1,K'} \le 1$ that

$$\left\| \sum_{i=1}^m u_i(s)X_i f \right\|_{\alpha,K'} \le \sup_{x\in K'} \left| X_{i_\ell} \circ \cdots \circ X_{i_1} \left( \sum_{i=1}^m u_i(s)X_i f \right) \right| \tag{6.54}$$

$$\le \sup_{x\in K'} \sum_{i=1}^m |u_i(s)||X_{i_\ell} \circ \cdots \circ X_{i_1} \circ X_i f| \le \sum_{i=1}^m |u_i(s)|. \tag{6.55}$$

□

To complete the discussion, let us describe a case when the Volterra expansion is actually convergent. One can prove the following result.

**Proposition 6.22.** *Let $X_t$ be a nonautonomous vector field, locally bounded with respect to $t \in I$. Assume that there exists a Banach space $(L, \|\cdot\|) \subset C^\infty(M)$ such that*

*(a) $X_t a \in L$ for all $a \in L$ and all $t \in I$,*

*(b) $\sup\{\|X_t a\| : a \in L, \|a\| \le 1, t \in I\} < \infty$.*

*Then the Volterra expansion (6.44) converges on $L$ for every $t \in I$.*

*Proof.* It is sufficient to bound the general term of the Volterra series with respect to the norm $\|\cdot\|$ of $L$ as follows

$$\left\| \int_{\Delta_k(t)} \cdots \int X_{s_k} \circ \cdots \circ X_{s_1} a \, ds_1 \cdots ds_k \right\| \le \int_{\Delta_k(t)} \cdots \int \|X_{s_k}\| \cdots \|X_{s_1}\| ds_1 \cdots ds_k \, \|a\| \tag{6.56}$$

$$= \frac{1}{k!} \left( \int_0^t \|X_s\| ds \right)^k \|a\|, \tag{6.57}$$

The norm of the $k$-th term of the Volterra series is bounded above by an exponential series, thus the Volterra expansion converges on $L$ uniformly. □

*Remark* 6.23. The assumption in Proposition 6.22 is satisfied in particular for every linear autonomous vector field $X$ on $M = \mathbb{R}^n$, by choosing as $L \subset C^\infty(\mathbb{R}^n)$ the subset of linear functions.

If the manifold $M$, the vector field $X_t$ and the function $a$ are real analytic, then it can be proved that the Volterra expansion is convergent for small time. For a precise statement see [AG78, Prop. 2.1].

## 6.7 Appendix: Remainder term of the Volterra expansion

In this Appendix we prove Theorem 6.20. We start with the following key result.

**Proposition 6.24.** *Let $X_t$ be a nonautonomous vector field and denote by $P_{t,s}$ its flow. Then for every $t > 0$, $\alpha \in \mathbb{N}$ and $K \subset M$ compact, there exists $K'$ compact containing $K$ and $C > 0$ such that*

$$\|P_{0,t}a\|_{\alpha,K} \leq Ce^{\int_0^t \|X_s\|_{\alpha,K'}ds}\|a\|_{\alpha,K'}. \tag{6.58}$$

*Proof.* Define the compact set

$$K_t := \bigcup_{s\in[0,t]} P_{0,s}(K),$$

and the real function

$$\beta(t) := \sup\left\{ \frac{\|P_{0,t}f\|_{\alpha,K}}{\|f\|_{\alpha+1,K_t}} \,\Big|\, f \in C^\infty(M), \|f\|_{\alpha+1,K_t} \neq 0 \right\}. \tag{6.59}$$

Notice that the function $\beta$ is finite (cf. Exercice 6.4 and notice that $\|f\|_{\alpha,K_t} \leq \|f\|_{\alpha+1,K_t}$). Moreover $\beta$ is measurable with respect to $t$ since the supremum in the right hand side can be taken over an arbitrary countable dense subset of $C^\infty(M)$. We have the following lemma, whose proof is postponed at the end of the proof of the proposition.

**Lemma 6.25.** *For every $t > 0$, $\alpha \in \mathbb{N}$ and $K \subset M$ compact, there exists $C > 0$ such that*

$$\|P_{0,t}f\|_{\alpha,K} \leq C\beta(t)\|f\|_{\alpha,K_t}, \qquad \forall\, f \in C^\infty(M). \tag{6.60}$$

Let us now consider the identity

$$P_{0,t}a = a + \int_0^t P_{0,s} \odot X_s a\, ds$$

which implies

$$\|P_{0,t}a\|_{\alpha,K} \leq \|a\|_{\alpha,K} + \int_0^t \|P_{0,s} \odot X_s a\|_{\alpha,K} ds.$$

Appying Lemma 6.25 with $f = X_s a$ we get

$$\begin{aligned}\|P_{0,t}a\|_{\alpha,K} &\leq \|a\|_{\alpha,K} + C\int_0^t \beta(s)\|X_s a\|_{\alpha,K_t} ds \\ &\leq \|a\|_{\alpha,K} + C\|a\|_{\alpha+1,K_t}\int_0^t \beta(s)\|X_s\|_{\alpha,K_t} ds,\end{aligned}$$

where we used that $K_s \subset K_t$ for $s \in [0,t]$, hence $\|\cdot\|_{\alpha,K_s} \leq \|\cdot\|_{\alpha,K_t}$. Dividing by $\|a\|_{\alpha+1,K_t}$ and using $\|a\|_{\alpha,K_t} \leq \|a\|_{\alpha+1,K_t}$ we get

$$\frac{\|P_{0,t}a\|_{\alpha,K}}{\|a\|_{\alpha+1,K_t}} \leq 1 + C\int_0^t \beta(s)\|X_s\|_{\alpha,K_t} ds.$$

From the definition of the function $\beta$, cf. (6.59), we have the inequality

$$\beta(t) \leq 1 + C\int_0^t \beta(s)\|X_s\|_{\alpha,K_t} ds. \tag{6.61}$$

Using Gronwall inequality, this implies

$$\beta(t) \leq e^{C \int_0^t \|X_s\|_{\alpha, K_t} ds}. \tag{6.62}$$

Then (6.58) follows combining the last inequality and (6.60) choosing $f$ equal to $a$ (for every compact set $K'$ containing $K_t$). $\square$

Now we complete the proof of the main result, namely Theorem 6.20. Recall that we can write

$$\overrightarrow{\exp} \int_0^t X_s ds - S_N(t) = \int \cdots \int_{0 \leq s_N \leq \ldots \leq s_1 \leq t} P_{0,s_N} \circ X_{s_N} \circ \cdots \circ X_{s_1} ds,$$

hence

$$\left\| \left( \overrightarrow{\exp} \int_0^t X_s ds - S_N(t) \right) a \right\|_{\alpha, K} \leq \int \cdots \int_{0 \leq s_N \leq \ldots \leq s_1 \leq t} \left\| P_{0,s_N} \circ X_{s_N} \circ \cdots \circ X_{s_1} a \right\|_{\alpha, K} ds.$$

Applying Proposition 6.24 to the function $X_{s_N} \circ \cdots \circ X_{s_1} a$ one obtains

$$\left\| \left( \overrightarrow{\exp} \int_0^t X_s ds - S_N(t) \right) a \right\|_{\alpha, K} \leq C e^{\int_0^t \|X_s\|_{\alpha, K} ds} \int \cdots \int_{0 \leq s_N \leq \ldots \leq s_1 \leq t} \left\| X_{s_N} \circ \cdots \circ X_{s_1} a \right\|_{\alpha, K'} ds, \tag{6.63}$$

for some compact $K'$ containing $K$. Now let us estimate the integral

$$\int \cdots \int_{0 \leq s_N \leq \ldots \leq s_1 \leq t} \left\| X_{s_N} \circ \cdots \circ X_{s_1} a \right\|_{\alpha, K'} ds \tag{6.64}$$

$$\leq \int \cdots \int_{0 \leq s_N \leq \ldots \leq s_1 \leq t} \|X_{s_N}\|_{\alpha, K'} \left\| X_{s_{N-1}} \right\|_{\alpha+1, K'} \cdots \circ \|X_{s_1}\|_{\alpha+N-1, K'} \|a\|_{\alpha+N, K'} ds \tag{6.65}$$

$$\leq \|a\|_{\alpha+N, K'} \int \cdots \int_{0 \leq s_N \leq \ldots \leq s_1 \leq t} \|X_{s_N}\|_{\alpha+N-1, K'} \left\| X_{s_{N-1}} \right\|_{\alpha+N-1, K'} \cdots \circ \|X_{s_1}\|_{\alpha+N-1, K'} ds \tag{6.66}$$

$$\leq \|a\|_{\alpha+N, K'} \frac{1}{N!} \left( \int_0^t \|X_s\|_{\alpha+N-1, K'} ds \right)^N. \tag{6.67}$$

Combining this inequality with (6.63), the proof is completed.

*Proof of Lemma 6.25.* By Whitney theorem it is not restrictive to assume that $M$ is a submanifold of $\mathbb{R}^n$ for some $n$. We still denote by $X_1, \ldots, X_r$ the vector fields (now defined on $\mathbb{R}^n$) spanning the tangent space to $M$.

Notice that if $\|f\|_{\alpha, K_t} = 0$ then also $\|P_{0,t} f\|_{\alpha, K} = 0$ and the identity is satisfied, hence we can assume $\|f\|_{\alpha, K_t} \neq 0$. Fix a point $q_0 \in K$ where the supremum in

$$\|P_{0,t} f\|_{\alpha, K} = \sup_{q \in K} \{ |(X_{i_\ell} \circ \cdots \circ X_{i_1} \circ P_{0,t} f)(q)| : 1 \leq i_1 \leq \ldots, i_\ell \leq r, 0 \leq \ell \leq \alpha \},$$

is attained (the existence of such a point is guaranteed by compactness of $K$). Let $p_f$ be the polynomial in $\mathbb{R}^n$ of degree $\leq \alpha$ that coincides with the Taylor polynomial of degree $\alpha$ of $f$ at $q_t = P_{0,t}(q_0)$. Let us prove that

$$\|p_f\|_{\alpha,\{q_t\}} \leq \|f\|_{\alpha,K_t}, \qquad \|P_{0,t}f\|_{\alpha,K} \leq \|P_{0,t}p_f\|_{\alpha,K}. \tag{6.68}$$

The first identity easily follows noticing that $\|p_f\|_{\alpha,\{q_t\}} = \|f\|_{\alpha,\{q_t\}} \leq \|f\|_{\alpha,K_t}$. To prove the second one, notice that, by construction, there exist $\ell \leq \alpha$ and $j_1, \ldots, j_\ell$ such that

$$\begin{aligned}\|P_{0,t}f\|_{\alpha,K} &= |(X_{j_\ell} \circ \cdots \circ X_{j_1} \circ P_{0,t}f)(q_0)| \\ &= |(X_{j_\ell} \circ \cdots \circ X_{j_1} \circ P_{0,t}p_f)(q_0)| \leq \|P_{0,t}p_f\|_{\alpha,K}.\end{aligned}$$

Notice that, on the vector space of polynomials in $\mathbb{R}^n$ of degree $\leq \alpha$, the two quantities $\|\cdot\|_{\alpha,K_t}$ and $\|\cdot\|_{\alpha,\{q_t\}}$ defines two norms. Since this vector space is finite-dimensional, every two norms are equivalent and there exist $C > 0$ such that

$$\|p_f\|_{\alpha,K_t} \leq C\|p_f\|_{\alpha,\{q_t\}}. \tag{6.69}$$

Combining (6.68) and (6.69) with $\|p_f\|_{\alpha,K_t} = \|p_f\|_{\alpha+1,K_t}$ (since $p_f$ is a polynomial of degree $\alpha$) and the definition of $\beta$, we have

$$\frac{\|P_{0,t}f\|_{\alpha,K}}{\|f\|_{\alpha,K_t}} \leq \frac{\|P_{0,t}p_f\|_{\alpha,K}}{\|p_f\|_{\alpha,\{q_t\}}} \leq C\frac{\|P_{0,t}p_f\|_{\alpha,K}}{\|p_f\|_{\alpha,K_t}} \leq C\frac{\|P_{0,t}p_f\|_{\alpha,K}}{\|p_f\|_{\alpha+1,K_t}} \leq C\beta(t).$$

□

## 6.8 Bibliographical note

The chronological calculus was originally conceveid by A. Agrachev and R. Gamkrelidze in the late 1970s, to investigate problems in optimization and control, in particular to extend Pontryagin Maximum Principle. This formalism aims to generalize well-known formulas for stationary flows to non-stationary ones, by overcoming the difficulty that the vector fields at different times do not commute.

The name "chronological calculus" appeared first [AG78, AG80], and its origin come from physics, where the term "chronological" is typically used in "nonstationary" situations. The presentation given here is very close to the one in [AS04, Chapter 2].

Chronological calculus have been successively and efficiently exploited to treat many problem of geometric control theory, as in [Agr96, Sar04] and [Kaw99, KS97, Kaw02], or to generalize results to non-smooth situations [RS07].

An extension to the case of infinite-dimensional manifolds have been recently proposed in [KL15]. A survey on this topic can be found in [Kaw12], as part of the encyclopedia [Mey12].

# Chapter 7

# Lie groups and left-invariant sub-Riemannian structures

In this chapter we study normal Pontryagin extremals on left-invariant sub-Riemannian structures on a Lie group $G$. Such structures provide most of the examples in which normal Pontryagin extremals can be computed explicitly in terms of elementary functions.

We introduce Lie groups by studying subgroups of the group of diffeomorphisms of a manifold $M$ induced by a family of vector fields whose Lie algebra is finite dimensional.

We then define left-invariant sub-Riemannian structures. Such structures have always constant rank and, if they are of rank $m$, they can be generated by exactly $m$ linearly independent vector fields defined globally. On these structures we have always global existence of length-minimizers.

We then discuss Hamiltonian systems on Lie groups with left-invariant Hamiltonians. They always admit a certain number of first integrals (cf. Chapter 18 for the complete integrability of left-invariant structures on three-dimensional Lie groups).

We study in details some classes of systems in which one can obtain the explicit expression of normal Pontryagin extremals.

## 7.1 Subgroups of Diff($M$) generated by a finite-dimensional Lie algebra of vector fields

Let $M$ be a connected smooth manifold of dimension $n$ and let $L \subset \mathrm{Vec}(M)$ be a finite-dimensional Lie algebra of vector fields of dimension $\dim L = \ell$. Assume that all elements of $L$ are complete vector fields. The set

$$\mathcal{G} := \{e^{X_1} \circ \ldots \circ e^{X_k} \mid k \in \mathbb{N}, X_1, \ldots, X_k \in L\} \subset \mathrm{Diff}(M), \tag{7.1}$$

has a natural structure of subgroup of the group of diffeomorphisms of $M$, where the group law is given by the composition. We want to prove the following result.

**Theorem 7.1.** *The group $\mathcal{G}$ can be endowed with a structure of connected smooth manifold of dimension $\ell = \dim L$. Moreover the group multiplication and the inversion are smooth with respect to the differentiable structure.*

To prove this theorem, we build the differentiable structure on $\mathcal{G}$ by explicitly defining charts. To this aim, for all $P \in \mathcal{G}$ let us consider the map

$$\Phi_P : L \to \mathcal{G}, \qquad \Phi_P(X) = P \circ e^X.$$

**Proposition 7.2.** *The following properties hold:*

*(i) there exists $U \subset L$ neighborhood of 0 such that $\Phi_P|_U$ is invertible on its image, for all $P \in \mathcal{G}$,*

*(ii) for every $W$ neighborhood of 0 in $U$, and for all $P' \in \Phi_P(W)$, there exists $V$ neighborhood of 0 in $U$ such that $\Phi_{P'}(V) \subset \Phi_P(W)$, for all $P \in \mathcal{G}$.*

Thanks to the previous result, one can introduce the following basis of neighborhoods [1] on $\mathcal{G}$:

$$\mathcal{B} = \{\Phi_P(W) \,|\, P \in \mathcal{G}, W \subset U, 0 \in W\}. \tag{7.2}$$

where $U$ is determined as in (i) of Proposition 7.2. Let now $B_1 = \Phi_{P_1}(W_1)$ and $B_2 = \Phi_{P_2}(W_2)$. Assume $B_1 \cap B_2 \neq \emptyset$. Then there exists $Q \in B_1 \cap B_2$ and, by part (ii) of Proposition 7.2, for $i = 1, 2$ there exists $V_i \subset U$ neighborhoods of the origin with $\Phi_Q(V_i) \subset \Phi_{P_i}(W_i)$. Set $B_3 = \Phi_Q(V_1 \cap V_2)$, then

$$B_3 = \Phi_Q(V_1 \cap V_2) \subset \Phi_Q(V_1) \cap \Phi_Q(V_2) = B_1 \cap B_2.$$

This proves that (7.2) satisfies the axioms of a basis for generates a unique topology on $\mathcal{G}$.

Once the topology generated by $\mathcal{B}$ is introduced the map $\Phi_P|_U$ is automatically an homeomorphism, and this proves that $\mathcal{G}$ is a topological group, i.e., a group that is also a topological manifold such that the multiplication and the inversion are continuous with respect to the topological structure. Indeed it can be shown that, if $\Phi_P(W) \cap \Phi_{P'}(W') \neq \emptyset$, then the change of chart

$$\Phi_P^{-1} \circ \Phi_{P'} : \Phi_{P'}^{-1}(\Phi_P(W) \cap \Phi_{P'}(W')) \to \Phi_P^{-1}(\Phi_P(W) \cap \Phi_{P'}(W'))$$

is smooth with respect to the smooth structure defined on the vector space $L$ (cf. Exercice 7.10 and Section 7.1.2). Hence $\mathcal{G}$ has the structure of smooth manifold.

The goal of the next few subsections is the proof of Proposition 7.2, which is based on a reduction to a finite dimensional setting, which we now explain.

### 7.1.1 A finite-dimensional approximation

The finite-dimensional reduction is based on the following idea: we replace elements of $G$, that are diffeomorphisms of $M$, with the evaluation of elements of $G$ on a special set of $\ell$ points, where $\ell$ is the dimension of the Lie algebra $L$.

To identify the "special set" of points where we shall evaluate diffeomorphisms, we first need a general lemma.

[1] Recall that a collection $\mathcal{B}$ of subsets of a set $X$ is a basis for a (unique) topology on $X$ if and only if

(a) $\cup_{B \in \mathcal{B}} = X$,

(b) for every $B_1, B_2 \in \mathcal{B}$ and $x \in B_1 \cap B_2$, there exists nonempty $B_3 \in \mathcal{B}$ such that $x \in B_3$ and $B_3 \subset B_1 \cap B_2$.

**Lemma 7.3.** *For every $k \in \mathbb{N}$ and $F_1, \ldots, F_k : \mathbb{R}^m \to \mathbb{R}^n$ a family of linearly independent functions, there exist $x_1, \ldots, x_k \in \mathbb{R}^m$ such that the vectors*

$$(F_i(x_1), F_i(x_2), \ldots, F_i(x_k)), \qquad i = 1, \ldots, k$$

*are linearly independent as elements of $\mathbb{R}^{n \times k}$.*

Here by linearly independent functions we mean that $F_1, \ldots, F_k$ are linearly independent as elements of the vector space of functions from $\mathbb{R}^m$ to $\mathbb{R}^n$.

*Proof.* We prove the statement by induction on $k$.

(i). Since $F_1$ is not the zero function then there exists $x_1 \in \mathbb{R}^m$ such that $F_1(x_1) \neq 0$.

(ii). Assume that the statement is true for every set of $k$ linearly independent functions and consider a family $F_1, \ldots, F_{k+1}$ of linearly independent functions. Let $x_1, \ldots, x_k$ be the set of points obtained by applying the inductive step to the family $F_1, \ldots, F_k$. If the claim is not true for $k+1$, it means that for every $\bar{x} \in \mathbb{R}^m$ there exists a non zero vector $(c_1(\bar{x}), \ldots, c_{k+1}(\bar{x}))$ such that

$$\sum_{i=1}^{k+1} c_i(\bar{x}) F_i(\bar{x}) = 0, \qquad \sum_{i=1}^{k+1} c_i(\bar{x}) F_i(x_j) = 0, \qquad j = 1, \ldots, k, \tag{7.3}$$

By definition of $x_1, \ldots, x_k$ we have that $c_{k+1}(\bar{x}) \neq 0$, otherwise we get a contradiction with the inductive assumption. Hence we can assume $c_{k+1}(\bar{x}) = -1$ and rewrite equations (7.3) as

$$\sum_{i=1}^{k} c_i(\bar{x}) F_i(x_j) = F_{k+1}(x_j), \qquad j = 1, \ldots, k, \tag{7.4}$$

$$\sum_{i=1}^{k} c_i(\bar{x}) F_i(\bar{x}) = F_{k+1}(\bar{x}), \tag{7.5}$$

Treating (7.4) as a linear equation in the variables $c_1, \ldots, c_k$, its matrix of coefficients has rank $k$ by assumption, hence its solution (that exists) is unique and independent on $\bar{x}$. Let us denote it by $(c_1, \ldots, c_k)$. Then (7.5) gives

$$\sum_{i=1}^{k} c_i F_i(\bar{x}) = F_{k+1}(\bar{x})$$

for every arbitrary $\bar{x} \in \mathbb{R}^m$, which is in contradiction with the fact that $F_1, \ldots, F_{k+1}$ is a linearly independent family of functions.

□

As an immediate consequence of the previous lemma one obtains the following property. Recall that given two smooth $n$-dimensional smooth manifolds $M, N$ we have $T(M \times N) = TM \times TN$ where the elements of $TM \times TN$ writes as products of vectors $v \times w$ (or vector fields $X \times Y$) as defined in Section 2.4.1.

**Proposition 7.4.** *Let $X_1, \ldots, X_\ell$ be a basis of $L$. Then there exists $q_1, \ldots, q_\ell \in M$ such that the vectors*

$$X_i(q_1) \times \ldots \times X_i(q_\ell), \qquad i = 1, \ldots, \ell,$$

*are linearly independent as elements of $T_{q_1}M \times \ldots \times T_{q_\ell}M$.*

In the rest of this section, the points $q_1, \ldots, q_\ell$ are determined as in Proposition 7.4. The following proposition defines the neighborhood $U$ that appears in the statement of Proposition 7.2.

**Proposition 7.5.** *There exists a neighborhood of the origin $U \subset L$ such that the map*

$$\phi : U \to M^\ell, \qquad \phi(X) = (e^X(q_1), \ldots, e^X(q_\ell)) \in M^\ell,$$

*is an immersion at the origin.*[2]

*Proof.* It is enough to show that the rank of $\phi_*$ is equal to $\ell$. Computing the partial derivatives at $0 \in L$ of $\phi$ in the directions $X_1, \ldots, X_\ell$ we have

$$\frac{\partial \phi}{\partial X_i}(0) = \frac{d}{dt}\bigg|_{t=0} (e^{tX_i}(q_1), \ldots, e^{tX_i}(q_\ell)) = X_i(q_1) \times \ldots \times X_i(q_\ell), \qquad i = 1, \ldots, \ell,$$

and these are linearly independent as elements of $T_{q_1}M \times \ldots \times T_{q_\ell}M$ by Lemma 7.4. □

We are now going to study $L$ seen as a Lie algebra of vector fields on $M^k$. Given $k \in \mathbb{N}$, we can give $\mathrm{Vec}(M^k) = \mathrm{Vec}(M)^k$ the structure of a Lie algebra as follows (cf. Exercice 2.40):

$$[X_1 \times \ldots \times X_k, Y_1 \times \ldots \times Y_k] = [X_1, Y_1] \times \ldots \times [X_k, Y_k].$$

**Lemma 7.6.** *For every $k \in \mathbb{N}$ the map $i : L \to \mathrm{Vec}(M)^k$ defined by $i(X) = X \times \ldots \times X$ defines an involutive distribution on $M^k$.*

*Proof.* It follows from the identity $[i(X), i(Y)] = i([X, Y])$, since

$$[X \times \ldots \times X, Y \times \ldots \times Y] = [X, Y] \times \ldots \times [X, Y].$$ □

**Lemma 7.7.** *If $P \in \mathcal{G}$ then $P_*L = L$.*

*Proof.* Let us first prove that $P_*L \subset L$ for every $P \in \mathcal{G}$. Since elements in $\mathcal{G}$ are written as

$$P = e^{X_1} \circ \ldots \circ e^{X_k}, \qquad X_j \in L$$

it is enough to show that for every $X, Y \in L$ we have that $e^X_* Y \in L$. By (6.32) we have the identity

$$e^X_* Y = e^{-\mathrm{ad}\, X} Y,$$

The Volterra exponential series of $-\mathrm{ad}\, X$ converges, since $L$ is a finite dimensional space. The $N$-th term of the sum

$$Y + \sum_{k=1}^{N} \frac{(-1)^k}{k!} (\mathrm{ad}\, X)^k Y,$$

belongs to $L$ for each $N \in \mathbb{N}$, since $L$ is a Lie algebra. Hence one can pass to the limit for $N \to \infty$ and $e^{-\mathrm{ad}\, X} Y \in L$. This proves that $P_*L \subset L$. Actually $P_*L = L$ since $P_*L$ is a Lie algebra and $\dim P_*L = \dim L$, since $P$ is a diffeomorphism. □

[2] here $M^\ell = \underbrace{M \times \ldots \times M}_{\ell \text{ times}}$.

For every $P \in \mathcal{G}$ we introduce

$$\phi_P : U \to M^\ell, \qquad \phi_P = P \circ \phi,$$

or, more explicitly

$$\phi_P(X) = (P \circ e^X(q_1), \ldots, P \circ e^X(q_\ell)), \qquad X \in U.$$

Thanks to Proposition 7.5 it follows that $\phi_P$ is an immersion at zero for all $P \in \mathcal{G}$, since it is a composition of an immersion with a diffeomorphism.

**Proposition 7.8.** *For all $P \in \mathcal{G}$ we have that $\phi_P(U)$ belongs to the integral manifold in $M^\ell$ of the foliation defined by L (seen as distribution in $\mathrm{Vec}(M)^\ell$) passing through the point $(P(q_1), \ldots, P(q_\ell)) \in M^\ell$. Moreover for every $P \in \mathcal{G}$, $\phi_P(U)$ belongs to the same leaf of the foliation.*

*Proof.* The Lie algebra $L$, seen as a distribution in $\mathrm{Vec}(M)^\ell$, is involutive. Thus it generates a foliation by Frobenius theorem. The leaf of the foliation passing through $(q_1, \ldots, q_\ell)$ (that has dimension $\ell$) has the expression

$$N = \{(\widehat{P}(q_1), \ldots, \widehat{P}(q_\ell)) \mid \widehat{P} = e^{X_1} \circ \ldots \circ e^{X_k}, k \in \mathbb{N}, X_1, \ldots, X_k \in L\},$$

while for each $P \in \mathcal{G}$,

$$\phi_P(U) = \{(P \circ e^X(q_1), \ldots, P \circ e^X(q_\ell)) \mid P \in \mathcal{G}, X \in U \subset L\},$$

hence for each $P \in \mathcal{G}$ we have that $\phi_P(U) \subset N$. The image $\phi_P(U)$ is an immersed submanifold of dimension $\ell$ that is tangent to $L$ thanks to Lemma 7.7, and passes through the point $\phi_P(0) = (P(q_1), \ldots, P(q_\ell)) \in M^\ell$.

□

*Remark* 7.9. The previous result implies that for every $(q_1', \ldots, q_\ell') \in \phi_P(U) \cap \phi_{P'}(U)$ there exists *uniques* $X, X' \in U$ such that

$$(P \circ e^X(q_1), \ldots, P \circ e^X(q_\ell)) = (P' \circ e^{X'}(q_1), \ldots, P' \circ e^{X'}(q_\ell)) = (q_1', \ldots, q_\ell'). \tag{7.6}$$

**Exercise 7.10.** Prove that the maps that associates $X \mapsto X'$ defined in (7.6) is smooth.

The identity (7.6) is saying that the two diffeomorphisms $P \circ e^X$ and $P' \circ e^{X'}$ coincide when evaluated on the set of points $\{q_1, \ldots, q_\ell\}$. The argument that is developed in the next section shows that indeed $P \circ e^X = P' \circ e^{X'}$ as diffeomorphisms.

### 7.1.2 Passage to infinite dimension

In what follows, to study elements of $\mathcal{G}$ as diffeomorphisms and not only as acting on a finite set of points, we use the following idea: we study diffeomorphisms on a set of $\ell + 1$ points, where the first one is "free".

Let $q \in M$. Let us introduce

$$\overline{\phi} : U \to M^{\ell+1}, \qquad \overline{\phi}(X) = (e^X(q), e^X(q_1), \ldots, e^X(q_\ell)) \in M^{\ell+1}.$$

Moreover, we define for every $P \in \mathcal{G}$

$$\overline{\phi}_P : U \to M^{\ell+1}, \qquad \overline{\phi}_P(X) = (P \circ e^X(q), P \circ e^X(q_1), \ldots, P \circ e^X(q_\ell)) \in M^{\ell+1}.$$

The following Proposition can be proved following the same arguments as the one of Proposition 7.8.

**Proposition 7.11.** *Let $q \in M$. For all $P \in \mathcal{G}$ we have that $\overline{\phi}_P(U)$ is an integral manifold of dimension $\ell$ in $M^{\ell+1}$ of a foliation defined by $L$ (seen as distribution in* $\mathrm{Vec}(M)^{\ell+1}$*) and passing through the point $(P(q), P(q_1), \ldots, P(q_\ell)) \in M^{\ell+1}$. Moreover, for every $P \in \mathcal{G}$, $\overline{\phi}_P(U)$ belong to the same leaf of the foliation.*

Notice that if $\pi : M^{\ell+1} \to M^\ell$ denotes the projection $\pi(q_0, q_1, \ldots, q_\ell) = (q_1, \ldots, q_\ell)$ that forgets about the first element we have $\phi = \pi \circ \overline{\phi}$ and $\phi_P = \pi \circ \overline{\phi}_P$. Notice that by construction

$$\pi : \overline{\phi}_P(U) \to \phi_P(U) \tag{7.7}$$

is a diffeomorphism for every choice of $P$ (in particular it is one-to-one).

### 7.1.3 Proof of Proposition 7.2

We can now complete the proof of the main result. Recall that $\Phi_P$ is the map defined on $L$ that takes values in the group $\mathcal{G}$, while $\phi_P$ is its $\ell$ finite-dimensional version taking values in $M^\ell$

$$\Phi_P(X) = P \circ e^X, \qquad \phi_P(X) = (P \circ e^X(q_1), \ldots, P \circ e^X(q_\ell)).$$

(i). It is enough to show that $\Phi_P$ is injective on its image. In other words we have to show that, if $P \circ e^X = P \circ e^Y$ for some $X, Y \in U$, then $X = Y$. The assumption implies that

$$\phi_P(X) = (P \circ e^X(q_1), \ldots, P \circ e^X(q_\ell)) = (P \circ e^Y(q_1), \ldots, P \circ e^Y(q_\ell)) = \phi_P(Y)$$

hence by invertibility of $\phi_P$ on $U$ we have that $X = Y$.

(ii). Let $W \subset U$, with $0 \in W$. Recall that, by construction, one has the following relation between $\Phi_P$ and its finite-dimensional representation $\phi_P$

$$\phi_P(W) = \{(Q(q_1), \ldots, Q(q_\ell)) : Q \in \Phi_P(W)\}, \qquad W \subset U.$$

For every $V \subset W$, with $0 \in V$, one has that $\phi_{P'}(V)$ and $\phi_P(W)$ are integral submanifold of $M^\ell$ belonging to the same leaf of the foliation, thanks to Proposition 7.8.

Since by assumption $P' \in \Phi_P(W)$, it follows that the intersection $\phi_{P'}(V) \cap \phi_P(W)$ is open and non empty in $M^\ell$ and contains the point $(P'(q_1), \ldots, P'(q_\ell))$. We can then choose $V$ small enough such that $\phi_{P'}(V) \subset \phi_P(W)$.

This inclusion of the finite-dimensional images implies the following: for every $X' \in V$ there exists a unique element $X \in W$ such that $P' \circ e^{X'} = P \circ e^X$ when evaluated on the special set of points, namely

$$(P \circ e^X(q_1), \ldots, P \circ e^X(q_\ell)) = (P' \circ e^{X'}(q_1), \ldots, P' \circ e^{X'}(q_\ell)). \tag{7.8}$$

To complete the proof it is enough to show that $P' \circ e^{X'} = P \circ e^X$ at every point.

To this aim fix an arbitrary $q \in M$ and let us consider the extended finite-dimensional maps $\overline{\phi}_P$ and $\overline{\phi}_{P'}$. Let us first prove that, for $V$ chosen as before, one has $\overline{\phi}_{P'}(V) \subset \overline{\phi}_P(W)$ (independently on $q$). Assume that $\overline{\phi}_P(W) \setminus \overline{\phi}_{P'}(V) \neq \emptyset$, then we have

$$\begin{aligned} \pi(\overline{\phi}_{P'}(V)) &= \pi(\overline{\phi}_{P'}(V) \cap \overline{\phi}_P(W)) \cup \pi(\overline{\phi}_P(U) \setminus \overline{\phi}_{P'}(V)) && (7.9) \\ &= \phi_{P'}(V) \cup \pi(\overline{\phi}_P(W) \setminus \overline{\phi}_{P'}(V)) && (7.10) \end{aligned}$$

This gives a contradiction since on one hand the left-hand is connected thanks to (7.7) (for $P = P'$), while on the other hand it is written as a union of nonempty disjoint sets.

This implies in particular: for every $X' \in V$ there exists a unique element $\widehat{X} \in W$ (a priori dependent on $q$) such that $P' \circ e^{X'} = P \circ e^{\widehat{X}}$ when evaluated at $\{q, q_1, \dots, q_\ell\}$, namely

$$(P \circ e^{\widehat{X}}(q), P \circ e^{\widehat{X}}(q_1), \dots, P \circ e^{\widehat{X}}(q_\ell)) = (P' \circ e^{X'}(q), P' \circ e^{X'}(q_1), \dots, P' \circ e^{X'}(q_\ell)). \tag{7.11}$$

Combining (7.8) with (7.11) one obtains

$$\phi_P(\widehat{X}) = (P \circ e^{\widehat{X}}(q_1), \dots, P \circ e^{\widehat{X}}(q_\ell)) = (P \circ e^{X}(q_1), \dots, P \circ e^{X}(q_\ell)) = \phi_P(X).$$

By invertibility of $\phi_P$ on $W \subset U$, it follows that $\widehat{X} = X$, independently on $q$. Thus by (7.11) and the arbitrarity of $q$ we have $P' \circ e^{X'}(q) = P \circ e^{X}(q)$ for every $q$, for every fixed $X' \in V$, as claimed.

## 7.2 Lie groups and Lie algebras

**Definition 7.12.** A *Lie group* is a group $G$ that has a structure of smooth manifold such that the group multiplication

$$G \times G \to G, \qquad (g, h) \mapsto gh$$

and inversion

$$G \to G, \qquad g \mapsto g^{-1}$$

are smooth with respect to the differentiable structure of $G$.

We denote by $L_g : G \to G$ and $R_g : G \to G$ the left and right translation respectively

$$L_g(h) = gh, \qquad R_g(h) = hg.$$

Notice that $L_g$ and $R_g$ are diffeomorphisms of $G$ for every $g \in G$. Moreover $L_g \circ R_{g'} = R_{g'} \circ L_g$ for every $g, g' \in G$.

**Definition 7.13.** A vector field $X$ on a Lie group $G$ is said to be *left-invariant* (resp. *right-invariant*) if it satisfies $(L_g)_* X = X$ (resp. $(R_g)_* X = X$) for every $g \in G$.

*Remark* 7.14. Every left-invariant vector field $X$ on a Lie group $G$ its uniquely identified with its value at the origin $\mathbb{1}$ of the Lie group. Indeed if $X$ is left-invariant, it satisfies the relation

$$X(g) = L_{g*} X(\mathbb{1}). \tag{7.12}$$

On the other hand a vector field defined by the formula $X(g) = L_{g*} v$ for some $v \in T_{\mathbb{1}} G$, is left-invariant.

Notice that left-invariant vector fields are always complete. Since the Lie bracket of two left-invariant vector fields is left-invariant, we can give the following definition.

**Definition 7.15.** The *Lie algebra associated with a Lie group* $G$ is the Lie algebra $\mathfrak{g}$ of the left-invariant vector fields on $G$, endowed with the Lie bracket.

By Remark 7.14 the Lie algebra $\mathfrak{g}$ associated with a Lie group $G$ is a finite dimensional Lie algebra, that is isomorphic to $T_{\mathbb{1}}G$ as vector space. Hence $\mathfrak{g}$ endows $T_{\mathbb{1}}G$ with the structure of Lie algebra. In particular $\dim \mathfrak{g} = \dim G$. Given a basis $e_1, \ldots, e_n$ of $T_{\mathbb{1}}G$ we will often consider the induced basis of $\mathfrak{g}$ given by

$$X_i(g) = (L_g)_* e_i, \qquad i = 1, \ldots, n.$$

When it is convenient we identify $\mathfrak{g}$ with $T_{\mathbb{1}}G$ and a left-invariant vector field $X$ with its value at the origin $X(\mathbb{1})$.

**Definition 7.16.** Given a Lie group $G$ and its Lie algebra $\mathfrak{g}$ the *group exponential map* is the map

$$\exp : T_{\mathbb{1}}G \to G, \qquad \exp(X) = e^X(\mathbb{1}). \tag{7.13}$$

It is important to remember that in general the exponential map (7.13) is not surjective.

If $G_1$ and $G_2$ are Lie groups, then a *Lie group homomorphism* $\phi : G_1 \to G_2$ is a smooth map such that $f(gh) = f(g)f(h)$ for every $g, h \in G_1$. Two Lie groups are said to be *isomorphic* if there exist a diffeomorphism $\phi : G_1 \to G_2$ that is also a Lie group homomorphism.

Two Lie groups $G_1$ and $G_2$ are said *locally isomorphic* if there exists neighborhoods $U \subset G_1$ and $V \subset G_2$ of the identity element and a diffeomorphism $f : U \to V$ such that $f(gh) = f(g)f(h)$ for every $g, h \in U$ such that $gh \in U$.

**Exercise 7.17** (Second theorem of Lie)**.** Let $G_i$ be a Lie group with Lie algebra $L_i$, for $i = 1, 2$. Prove that an isomorphism between Lie algebras $i : L_1 \to L_2$ induces a local isomorphism of groups.
(Hint: Prove that the set $(X, i(X))$ is a subalgebra $L$ of the Lie algebra of the product group $G_1 \times G_2$. Build the group $G \subset G_1 \times G_2$ associated with this subalgebra, and then show that the two projections $p_i : G_1 \times G_2 \to G_i$ define $p_2 \circ (p_1|_G)^{-1} : G_1 \to G_2$ a local isomorphism of groups.)

### 7.2.1 Lie groups as groups of diffeomorphisms

In Section 7.1 we have proved that given a manifold $M$ and a finite dimensional Lie algebra $L$ of vector fields, the subgroup of Diff$(M)$ generated by these vector fields has a structure of finite dimensional differentiable manifold for which the groups operations are smooth. We call such a subgroup $\mathcal{G}^{M,L}$. By Definition 7.12 we have

**Proposition 7.18.** *$\mathcal{G}^{M,L}$ is a connected Lie group.*

We now want to prove a converse statement for connected groups, i.e., every connected Lie group is isomorphic to a subgroup of the group of the diffeomorphisms of a manifold generated by a finite dimensional Lie algebra of vector fields. Indeed this is true with $M = G$ and $L$ being the Lie algebra of left-invariant vector fields on $G$. More precisely we have the following.

**Theorem 7.19.** *Let $G$ be a connected Lie group and $\mathfrak{g}$ the Lie algebra of its left-invariant vector fields. Then $G$ is isomorphic to $\mathcal{G}^{G,\mathfrak{g}}$.*

To prove Theorem 7.19, we give first the following definition.

**Definition 7.20.** Let $G$ be a Lie group and let us define the group of its right translations as $G_R = \{R_g \mid g \in G\}$. On $G_R$ we give consider the group structure given by the operation (notice the inverse order)

$$R_{g_1} \cdot R_{g_2} := R_{g_2} \circ R_{g_1}.$$

Then we need the following simple facts.

**Lemma 7.21.** *$G$ is isomorphic $G_R$.*

*Proof.* Clearly the map $\phi : g \to R_g$ is a diffeomorphism. That is a group homomorphism follows from the fact that $R_{g_1 g_2} h = h(g_1 g_2) = (R_{g_2} \circ R_{g_1})h$. Hence

$$\phi(g_1 g_2) = R_{g_1 g_2} = R_{g_2} \circ R_{g_1} = R_{g_1} \cdot R_{g_2}.$$

□

Similarly one obtains that a Lie group $G$ is isomorphic to the group $G_L = \{L_g \mid g \in G\}$ of left translations on $G$ endowed with the group law given by the standard composition.

**Lemma 7.22.** *The flow of a left-invariant vector field on a Lie group $G$ commutes with left translations.*

A similar statement holds for right-invariant vector fields.

*Proof.* If $\phi$ is a diffeomorphism and $X$ a vector field we have that (see Lemma 2.21)

$$e^{t\phi_* X} = \phi \circ e^{tX} \circ \phi^{-1}.$$

Composing on the right with $\phi$, we have

$$e^{t\phi_* X} \circ \phi = \phi \circ e^{tX}.$$

Now choosing $\phi = L_g$ for some $g$, taking $X$ a left-invariant vector field and using that $L_{g*}X = X$, we have that

$$e^{tL_{g*}X} \circ L_g = L_g \circ e^{tX} = L_g \circ e^{tL_{g*}X}.$$

The conclusion follows from the arbitrarity of $g$. □

**Lemma 7.23.** *Let $G$ be a Lie group. A diffeomorphism on $G$ is a right translation if and only if it commutes with all left translations.*

*Proof.* Let $P$ be the diffeomorphism. If $P$ is a right translation then it commutes with left translation since for every $g, h_1, h_2 \in G$, we have $L_{h_1} R_{h_2} g = h_1 g h_2 = R_{h_2} L_{h_1} g$. To prove the opposite, let us define $g = P(\mathbb{1})$. For every $h \in G$, we have

$$P(h) = P(L_h \mathbb{1}) = L_h P(\mathbb{1}) = L_h g = hg,$$

hence $P = R_g$. □

*Remark* 7.24. By Lemma 7.22 and Lemma 7.23 we have that the flow of a left-invariant vector field is a right translation.

*Proof of Theorem 7.19.* By Lemma 7.21, it remains to prove that $\mathcal{G}^{G,\mathfrak{g}}$ is isomorphic to $G_R$. Indeed we are going to prove that $\mathcal{G}^{G,\mathfrak{g}} = G_R$.

To prove that $\mathcal{G}^{G,\mathfrak{g}} \subseteq G_R$ observe that every element of $\mathcal{G}^{G,\mathfrak{g}}$ is a composition of the flow of left-invariant vector fields and hence it is a right translation.

To prove that $\mathcal{G}^{G,\mathfrak{g}} = G_R$, observe that by the argument above $\mathcal{G}^{G,\mathfrak{g}}$ is a subgroup of $G_R$. Moreover since $\dim(\mathcal{G}^{G,\mathfrak{g}}) = \dim(G_R)$. It follows that $\mathcal{G}^{G,\mathfrak{g}}$ contains an open neighborhood of the identity. The conclusion of the Theorem is then a consequence of the following Lemma.

**Lemma 7.25.** *Let $G$ be a connected Lie group. If $H$ is a subgroup of $G$ containing an open neighborhood of the identity then $H = G$.*

*Proof.* Since by hypothesis $H$ is nonempty and open it remains to prove that $H$ is closed.

To this purpose observe that if $g \in G \setminus H$, then $gH$ is disjoint from $H$ (otherwise there exists $u \in H$ such that $gu \in H$ which implies that $guu^{-1} = g \in H$). Hence

$$G \setminus H = \bigcup_{g \notin H} gH.$$

Since each set $gH$ is open, it follows that $G \setminus H$ is open and hence that $H$ is closed. □

□

### 7.2.2 Matrix Lie groups and the matrix notation

A very important example of Lie group is the group of all invertible $n \times n$ real matrices, with respect to the matrix multiplication

$$GL(n) = \{M \in \mathbb{R}^{n\times n} \mid \det(M) \neq 0\}.$$

Similarly one define

$$GL(n, \mathbb{C}) = \{M \in \mathbb{C}^{n\times n} \mid \det(M) \neq 0\}.$$

**Exercise 7.26.** Prove that $GL(n, \mathbb{C})$ is connected while $GL(n)$ is not. Prove that the Lie algebra of $GL(n)$ (resp. $GL(n, \mathbb{C})$) is $\mathfrak{gl}(n) = \{M \in \mathbb{R}^{n\times n}\}$ (resp. $\mathfrak{gl}(n, \mathbb{C}) = \{M \in \mathbb{C}^{n\times n}\}$).

**Definition 7.27.** A group of matrices is a subgroup of $GL(n)$ or of $GL(n, \mathbb{C})$.

*Remark* 7.28. The Lie algebra of a subgroup of $GL(n)$ (resp. $GL(n, \mathbb{C})$) is a sub-algebra of $\mathfrak{gl}(n)$ (resp. $\mathfrak{gl}(n, \mathbb{C})$).

Group of matrices that we are going to meet along the book are

- The special linear group

  $$SL(n) = \{M \in \mathbb{R}^{n\times n} \mid \det(M) = 1\},$$

  whose Lie algebra is $\mathfrak{sl}(n) = \{M \in \mathbb{R}^{n\times n} \mid \mathrm{trace}(M) = 0\}$.

- The orthogonal group and the special orthogonal group

  $$\begin{aligned} O(n) &= \{M \in \mathbb{R}^{n\times n} \mid MM^T = \mathbb{1}\}, \\ SO(n) &= \{M \in \mathbb{R}^{n\times n} \mid MM^T = \mathbb{1}, \det(M) = 1\}, \end{aligned} \tag{7.14}$$

  for both the Lie algebra is $\mathfrak{so}(n) = \{M \in \mathbb{R}^{n\times n} \mid M = -M^T\}$. $SO(n)$ is the connected component of $O(n)$ containing the identity.

- The special unitary group

  $$SU(n) = \{M \in \mathbb{C}^{n\times n} \mid MM^\dagger = \mathbb{1}, \det(M) = 1\},$$

  where $M^\dagger$ is the transpose of the complex conjugate of $M$. The Lie algebra of $SU(n)$ is $\mathfrak{su}(n) = \{M \in \mathbb{C}^{n\times n} \mid M = -M^\dagger, \mathrm{trace}(M) = 0\}$.

- The group of (positively oriented) Euclidean isometries of $\mathbb{R}^n$

$$SE(n) = \left\{ \left( \begin{array}{c|c} & c_1 \\ M & \vdots \\ & c_n \\ \hline 0 & 1 \end{array} \right) \mid M \in SO(n), \quad c_1, \ldots, c_n \in \mathbb{R} \right\}. \tag{7.15}$$

The name of this group comes from the fact that if we represent a point of $\mathbb{R}^n$ as a vector $(x_1, \ldots, x_n, 1)$ then the action of a matrix of $SE(n)$ produces a rotation and a translation. The Lie algebra of $SE(n)$ is

$$\mathfrak{se}(n) = \left\{ \left( \begin{array}{c|c} & b_1 \\ A & \vdots \\ & b_n \\ \hline 0 & 0 \end{array} \right) \mid A \in \mathfrak{so}(n), b_1, \ldots, b_n \in \mathbb{R} \right\}.$$

**Exercise 7.29.** Prove that $\mathfrak{so}(3)$ and $\mathfrak{su}(2)$ are isomorphic as Lie algebras.

**Lemma 7.30.** *On group of matrices a left-invariant vector field $X = L_{g*}A = gA$, $A \in T_{\mathbb{1}}G$.*

*Proof.* By using the expression in coordinates $L_g : h \mapsto \sum_k g_{ik}h_{kj}$ we have that

$$(L_{g*}A)_{ij} = \sum_{l,m,k} \frac{\partial(g_{ik}h_{kj})}{\partial_{h_{lm}}} A_{lm} = \sum_{l,m,k} g_{ik}\delta_{kl}\delta_{jm}A_{lm} = \sum_k g_{ik}A_{kj}$$

□

Similarly one obtains that for $R_{g*}A = Ag$ for every $A \in T_{\mathbb{1}}G$.

*Remark* 7.31. Notice that the for a left-invariant vector field on a group of matrix $X(g) = gA$, the integral curve of $X$ satisfying $g(0) = g_0$ is $g(t) = g_0e^{tA}$ where $e^{tA}$ is the standard matrix exponential. Hence the integral curve of a left-invariant vector field, at a given $t$, is a right translation. This is indeed a general fact as explained in the next section.

**Exercise 7.32.** (i). Let $X(g) = gA$ and $Y(g) = gB$ be two left-invariant vector on a group of matrices. Prove that

$$[X, Y](g) = g(AB - BA) = g[A, B].$$

(Hint: use the expression in coordinates: if $X_{ij} = \sum_k g_{ik}A_{kj}$ and $Y_{ij} = \sum_k g_{ik}B_{kj}$, then $[X, Y]_{ij} = \sum_{kl} \left( \frac{\partial Y_{ij}}{\partial g_{kl}} X_{kl} - \frac{\partial X_{ij}}{\partial g_{kl}} Y_{kl} \right)$.)

(ii). Prove that, given two right-invariant vector fields $X(g) = Ag$ and $Y(g) = Bg$, we have

$$[X, Y](g) = -[A, B]g.$$

**Notation.** For a left-invariant vector field on a group of matrices it is often convenient to abuse notation and set $X(g) = gX$. This formula clarifies the identification of $\mathfrak{g}$ with $T_{\mathbb{1}}G$. Here $X(\cdot) \in \mathfrak{g}$ and $X \in T_{\mathbb{1}}G$.

### On the matrix notation

Given a vector field $X$ on a manifold, one can consider (cf. Chapter 6 for the chronological notation)

- its integral curve on $M$, i.e., the solutions to $\dot{q} = X(q)$,
- the equation for the flow of $X$, i.e., $\dot{P}_t = P_t \circ X$.

Let us write these equations for a left-invariant vector field $X$ on a Lie group $G$,

$$\dot{g} = X(g),$$
$$\dot{P}_t = P_t \circ X.$$

These two equations are indeed the same equation because:

- the flow of a left-invariant vector field is a right translation (see Remark 7.24);
- an element $g$ of a Lie group $G$ can be interpreted both as a point on $G$ seen as a manifold or as a diffeomorphism over $G$, once that $G$ is identified with the group of right translations $G_R$.

This fact is particularly evident when written for left-invariant vector fields on group of matrices. In this case the two equations take exactly the same form

$$\dot{g} = gX$$
$$\dot{P}_t = P_t \circ X$$

In the following we take advantage of this fact to simplify the notation. We sometimes eliminate the use of the symbols $L_g$ and $L_{g*}$: we write a left-invariant vector field in the form $X(g) = gX$, thinking to $gX$ as the matrix product when we are working with Lie groups of matrices (and in this case we think to $X \in T_{\mathbb{1}}G$), or as the composition of the left translation $g$ with the left-invariant vector field $X$ otherwise (and in this case we think to $X \in \mathfrak{g}$).

### 7.2.3 Bi-invariant pseudo-metrics and Haar measures

In this section we assume that the Lie group $G$ is connected. Recall that a *pseudo-Riemannian metric* is a family of non-degenerate, symmetric bilinear forms on each tangent space, smoothly depending on the point.

Since a Lie group $G$ is a smooth manifold as well as a group, it is natural to introduce the class of pseudo-Riemannian metrics that respect the group structure of $G$.

**Definition 7.33.** Let $\langle \cdot \,|\, \cdot \rangle$ be a pseudo-Riemannian metric on $G$. It is said to be *left-invariant* if

$$\langle v \,|\, w \rangle = \langle L_{g*}v \,|\, L_{g*}w \rangle \,, \qquad \forall\, v, w \in T_{\mathbb{1}}G, g \in G.$$

Similarly, $\langle \cdot \,|\, \cdot \rangle$ is a *right-invariant metric* if

$$\langle v \,|\, w \rangle = \langle R_{g*}v \,|\, R_{g*}w \rangle \,, \qquad \forall\, v, w \in T_{\mathbb{1}}G, g \in G.$$

A *bi-invariant metric* is a pseudo-Riemannian metric that is at the same time left and right-invariant.

**Exercise 7.34.** Prove that for a bi-invariant pseudo-metric we have the following

$$\langle [X,Y] \,|\, Z\rangle = \langle X \,|\, [Y,Z]\rangle\,, \qquad \forall\, X,Y,Z \in \mathfrak{g}. \tag{7.16}$$

**Definition 7.35.** A Lie algebra $\mathfrak{g}$ is said to be compact if it admits a positive definite bi-invariant pseudo-metric (hence a bi-invariant Riemannian metric).

One can prove that the Lie algebra of a compact Lie group is compact in the sense above. See for instance [BR86]. Next we define the adjoint action of $G$ onto $\mathfrak{g}$.

**Definition 7.36.** For every $g \in G$, the conjugation $C_g : G \to G$, is the map

$$C_g = R_{g^{-1}} \circ L_g, \qquad C_g(h) = ghg^{-1}.$$

The *adjoint action* $\mathrm{Ad}_g : \mathfrak{g} \to \mathfrak{g}$ is defined as $\mathrm{Ad}_g = C_{g*}$, namely

$$\mathrm{Ad}_g(X) = R_{g^{-1}*} L_{g*} X = R_{g^{-1}*} X, \qquad X \in \mathfrak{g}.$$

In matrix notation

$$\mathrm{Ad}_g(X) = gXg^{-1}, \quad X \in T_{\mathbb{1}} G.$$

Recall that, given $x \in \mathfrak{g}$, its adjoint representation $\mathrm{ad}\, x : \mathfrak{g} \to \mathfrak{g}$ is given by $\mathrm{ad}\, x(y) = [x,y]$. Notice that the map $\mathrm{ad} : x \mapsto \mathrm{ad}\, x$, as a map from $\mathfrak{g}$ to the set of automorphisms of $\mathfrak{g}$, is the differential of the map $\mathrm{Ad} : g \mapsto \mathrm{Ad}_g$ from $G$ to the set of automorphisms of $\mathfrak{g}$.

**Definition 7.37.** The *Killing form* on a Lie algebra $\mathfrak{g}$ is the symmetric bilinear form

$$K : \mathfrak{g} \times \mathfrak{g} \to \mathbb{R}, \qquad K(x,y) = \mathrm{trace}(\mathrm{ad}\, x \circ \mathrm{ad}\, y). \tag{7.17}$$

The Killing form has the associativity property

$$K([x,y],z) = K(x,[y,z]). \tag{7.18}$$

**Definition 7.38.** A Lie algebra is said to be *semisimple* if the Killing form is non-degenerate.

**Exercise 7.39.** Prove that if $\mathfrak{g}$ is semisimple then $[\mathfrak{g},\mathfrak{g}] = \mathfrak{g}$. Show that the group of rototranslations of $\mathbb{R}^3$ satisfies $[\mathfrak{g},\mathfrak{g}] = \mathfrak{g}$, but is not semisimple.

**Definition 7.40.** An ideal of a Lie algebra $\mathfrak{l}$ is a subspace $\mathfrak{i}$ such that $[\mathfrak{l},\mathfrak{i}] \subset \mathfrak{i}$. Given a Lie algebra $\mathfrak{l}$, define the sequence of ideals $\mathfrak{l}^{(0)} = \mathfrak{l}$, $\mathfrak{l}^{(1)} = [\mathfrak{l},\mathfrak{l}^{(0)}]$, $\ldots$, $\mathfrak{l}^{(n+1)} = [\mathfrak{l},\mathfrak{l}^{(n)}]$. The Lie algebra $\mathfrak{l}$ is said to be *nilpotent* if there exists $n$ such that $\mathfrak{l}^{(n)} = 0$.

**Exercise 7.41.** Prove that the Killing form of a nilpotent Lie algebra is identically zero.

Nilpotent algebras are particular cases of solvable algebras defined as follow.

**Definition 7.42.** Given a Lie algebra $\mathfrak{l}$, define the sequence of ideals $\mathfrak{l}^0 = \mathfrak{l}$, $\mathfrak{l}^1 = [\mathfrak{l}^0,\mathfrak{l}^0]$, $\ldots$, $\mathfrak{l}^{n+1} = [\mathfrak{l}^n,\mathfrak{l}^n]$. The Lie algebra $\mathfrak{l}$ is said to be *solvable* if there exists $n$ such that $\mathfrak{l}^{(n)} = 0$.

**Definition 7.43.** Let $\omega$ be $n$-form, with $n = \dim G$. The form is said to be *left-invariant* (resp. *right-invariant*) if $L_g^* \omega = \omega$ (resp. $R_g^* \omega = \omega$) for every $g \in G$.

Notice that a left-invariant (resp. right-invariant) $n$-form on a Lie group $G$ is uniquely determined by its value at the identity. It follows that the set of left-invariant (resp. right-invariant) $n$-forms on a Lie group $G$ is a 1-dimensional space and that one can always find a never vanishing one. As a consequence, Lie groups are orientable and there exists a unique left-invariant (resp. right-invariant) $n$-form up to a non-zero normalization constant.

**Definition 7.44.** A non-vanishing left-invariant (resp. right-invariant) $n$-form is called a left Haar measure (resp. a right Haar measure) on $G$.

**Definition 7.45.** A Lie group $G$ is said to be *unimodular* if left-invariant measures are also right-invariant.

For unimodular groups one speaks of Haar measures (omitting left or right). Example of unimodular Lie groups are: abelian, semisimple, compact. On a compact Le group, one can normalize the Haar measure by requiring that the integral of $\omega$ over $G$ is equal to 1.

**Exercise 7.46.** Prove that nilpotent Lie groups are unimodular.

### 7.2.4 The Levi-Malcev decomposition

A very important result in the theory of Lie algebras (see for instance [Che55, Ch. 4, Sect. 4, Thm. 4]) states that every Lie algebra can be decomposed as

$$\mathfrak{g} = \mathfrak{r} \ni \mathfrak{s}, \tag{7.19}$$

where:

- $\mathfrak{r}$ is the so called *radical*, i.e., the maximal solvable ideal of $\mathfrak{g}$.
- $\mathfrak{s}$ is a semisimple sub-algebra.
- The symbol $\ni$ indicates the semidirect sum of two Lie algebras defined in the following way. Let $T$ and $M$ be two Lie algebras and $D$ the homomorphism of $M$ into the set of linear operators in the vector space $T$ such that every operator $D(X)$ is a derivation of $T$. The Lie algebra $T \ni M$ is the vector space $T \oplus M$ with a Lie algebra structure given by using the given Lie brackets of $T$ and $M$ in each subspace and for the Lie brackets between the two subspaces we set
$$[X, Y] = D(X)Y, \qquad X \in M, Y \in T.$$

**Exercise 7.47.** Prove that $T \ni M$ is a well defined Lie algebra.

### Product of Lie groups

Given two Lie groups $G_1$ and $G_2$ their direct product is the Lie group obtained by considering $G_1 \times G_2$ with the multiplication rule

$$(g_1, g_2), (h_1, h_2) \in G_1 \times G_2 \mapsto (g_1 h_1, g_2 h_2) \in G_1 \times G_2.$$

One immediately verify that if $\mathfrak{g}_1$ and $\mathfrak{g}_2$ are the Lie algebras of $G_1$ and $G_2$, the Lie algebra of $G_1 \times G_2$ is $\mathfrak{g}_1 \oplus \mathfrak{g}_2$. In $\mathfrak{g}_1 \oplus \mathfrak{g}_2$ we have that $[\mathfrak{g}_1, \mathfrak{g}_2] = 0$.

## 7.3 Trivialization of $TG$ and $T^*G$

**Lemma 7.48.** *The tangent bundle $TG$ of a Lie group $G$ is trivializable*

*Proof.* Recall that the tangent bundle $TM$ of a smooth manifold $M$ is trivializable if and only if there exists a basis of globally defined independent vector fields. In the case of the tangent bundle $TG$ of a Lie group $G$ we can build a global family of independent vector field by fixing a basis $e_1, \ldots, e_n$ of $T_{\mathbb{1}}G$ and considering the induced left-invariant vector fields given by

$$X_i(g) = (L_g)_* e_i, \qquad i = 1, \ldots, n,$$

that are linearly independent by construction. $\square$

We have then an isomorphism between $TG$ and $G \times T_{\mathbb{1}}G$. This isomorphism is given by $L_{g^{-1}*}$, that is acting in the following way

$$TG \ni (g, v) \mapsto (g, \nu) \in G \times T_{\mathbb{1}}G,$$

where $\nu = L_{g^{-1}*}v$.

Notice that given two left-invariant vector fields $X(g) = L_{g*}\nu$ and $Y(g) = L_{g*}\mu$ where $\nu, \mu \in T_{\mathbb{1}}G$, we have

$$[X, Y](g) = L_{g*}[\nu, \mu]$$

The isomorphism between $TG$ and $G \times T_{\mathbb{1}}G$ extends to the dual. Hence $T^*G$ is isomorphic to $G \times T_{\mathbb{1}}^*G$, the isomorphism being given by $L_g^*$, i.e.,

$$T^*G \ni (p, g) \mapsto (\xi, g) \in G \times T_{\mathbb{1}}^*G,$$

where $\xi = L_g^* p$.

Notice that without an additional notion of scalar product, the Lie algebra structure on $T_{\mathbb{1}}G$ induced by $\mathfrak{g}$ does not induce a Lie algebra structure on $T_{\mathbb{1}}^*G$.

In the following it is often convenient to make computations in $G \times T_{\mathbb{1}}G$ and $G \times T_{\mathbb{1}}^*G$ instead than $TG$ and $T^*G$. It is then useful to recall that if $v = L_{g*}\nu \in T_gG$ and $p = L_{g^{-1}}^*\xi \in T_gG$, then

$$\langle p, v \rangle_g = \langle \xi, \nu \rangle_{\mathbb{1}}.$$

## 7.4 Left-invariant sub-Riemannian structures

A left-invariant sub-Riemannian structure is a constant rank sub-Riemannian structure $(G, \mathcal{D}, \langle \cdot \,|\, \cdot \rangle)$ (cf. Section 3.1.3, Example 2) where

- $G$ is a connected Lie group of dimensione $n$;
- the distribution is left-invariant, i.e., $\mathcal{D}(g) = L_{g*}\mathbf{d}$, where $\mathbf{d}$ is a subspace of $T_{\mathbb{1}}^*G$. Moreover we assume that the distribution is bracket-generating or equivalently that the smallest Lie sub-algebra of $\mathfrak{g}$ containing $\mathcal{D}$ is $\mathfrak{g}$ itself;
- $\langle \cdot \,|\, \cdot \rangle$ is a scalar product on $\mathcal{D}(g)$ that is left-invariant, i.e., if $v = L_{g*}\nu$ and $w = L_{g*}\mu$ with $\nu, \mu \in \mathbf{d}$ we have $\langle v \,|\, w \rangle_g = \langle \nu \,|\, \mu \rangle_{\mathbb{1}}$.

*Remark* 7.49. Left-invariant sub-Riemannian structure are by construction free and constant rank. If $\mathcal{D}$ has dimension $m \leq n$ then the local minimum bundle rank is constantly equal to $m$ (cf. Definition 3.21).

Given a left-invariant sub-Riemannian structure we can always find $m$ linearly independent vectors $e_1, \ldots, e_m$ in $T_{\mathbb{1}}G$ such that

(i) $\mathcal{D}(g) = \{\sum_{i=1}^m u_i L_{g*} e_i \mid u_1, \ldots u_m \in \mathbb{R}\}$,

(ii) $\langle e_i \,|\, e_j \rangle_{\mathbb{1}} = \delta_{ij}$.

The problem of finding the shortest curve connecting two points $g_1, g_2 \in G$ can then be formulated as the optimal control problem

$$\begin{cases} \dot{\gamma}(t) = \sum_{i=1}^m u_i(t) L_{g*} e_i \\ \int_0^T \sqrt{\sum_{i=1}^m u_i(t)^2}\, dt \to \min \\ \gamma(0) = g_1, \quad \gamma(T) = g_2, \end{cases} \tag{7.20}$$

**Exercise 7.50.** (i). Prove that if $g \in G$ and $\gamma : [0,T] \to G$ is an horizontal curve, then the left-translated curve $\gamma_g := L_g \circ \gamma$ is also horizontal and $\ell(\gamma_g) = \ell(\gamma)$.

(ii). Prove that $d(L_g h_1, L_g h_2) = d(h_1, h_2)$ for every $g, h_1, h_2 \in G$. Deduce that for every $g, h \in G$ and $r > 0$ one has

$$L_g(B(h,r)) = B(gh, r),$$

where $B(g,r)$ is the sub-Riemannian ball centered at $g$ and of radius $r$.

### Existence of minimizers

Proposition 3.47 immediately implies the following.

**Corollary 7.51.** *Any left-invariant sub-Riemannian structure on a Lie group $G$ is complete.*

*Proof.* By Proposition 3.37 small balls are compact. Hence there exists $\varepsilon > 0$ such that the ball $\overline{B}(\mathbb{1}, \varepsilon)$ is compact, where $\mathbb{1}$ is the identity of $G$. By left-invariance (cf. Exercice 7.50) $\overline{B}(g, \varepsilon) = L_g(\overline{B}(\mathbb{1}, \varepsilon))$ is compact for every $g \in G$, independently on $\varepsilon$. By Proposition 3.47, the sub-Riemannian structure is complete. $\square$

## 7.5 Example: Carnot groups of step 2

The Heisenberg sub-Riemannian structure $\mathbb{H}$ that we studied in Section 4.4.3 as an isoperimetric problem is indeed a left-invariant sub-Riemannian structure on the group $G = \mathbb{R}^3$ endowed with the product

$$(x, y, z) \cdot (x', y', z') \doteq \left( x + x',\, y + y',\, z + z' + \frac{1}{2}(xy' - x'y) \right).$$

Such a group is called the Heisenberg group.

**Exercise 7.52.** Prove that the Lie algebra of the Heisenberg group can be written as $\mathfrak{g} = \mathfrak{g}_1 \oplus \mathfrak{g}_2$ where

$$\mathfrak{g}_1 = \operatorname{span}\left\{\partial_x - \frac{y}{2}\partial_z, \partial_y + \frac{x}{2}\partial_z\right\}, \quad \text{and} \quad \mathfrak{g}_2 = \operatorname{span}\{\partial_z\}.$$

Notice that we have the commutation relations $[\mathfrak{g}_1, \mathfrak{g}_1] = \mathfrak{g}_2$ and $[\mathfrak{g}_1, \mathfrak{g}_2] = 0$.

In this section we focus on *Carnot groups of step 2*, which are natural generalizations of the Heisenberg group, namely Lie groups $G$ on $\mathbb{R}^n$ such that its Lie algebra $\mathfrak{g}$ satisfies

$$\mathfrak{g} = \mathfrak{g}_1 \oplus \mathfrak{g}_2, \qquad [\mathfrak{g}_1, \mathfrak{g}_1] = \mathfrak{g}_2, \qquad [\mathfrak{g}_1, \mathfrak{g}_2] = [\mathfrak{g}_2, \mathfrak{g}_2] = 0. \tag{7.21}$$

$G$ is endowed by the left-invariant sub-Riemannian structure induced by the choice of a scalar product $\langle \cdot \mid \cdot \rangle$ on the distribution $\mathfrak{g}_1$, that is bracket-generating of step 2 thanks to (7.21). Notice that $\mathfrak{g}$ is a nilpotent Lie algebra and that we have the inequality

$$n \leq \frac{m(m+1)}{2}, \qquad m = \dim \mathfrak{g}_1,\ n = \dim \mathfrak{g}.$$

We say that $\mathfrak{g}$ is a *Carnot algebra of step 2*.

Let us now choose a basis of left-invariant vector fields (on $\mathbb{R}^n$) of $\mathfrak{g}$ such that

$$\mathfrak{g}_1 = \operatorname{span}\{X_1, \dots, X_m\}, \qquad \mathfrak{g}_2 = \operatorname{span}\{Y_1, \dots, Y_{n-m}\},$$

where $\{X_1, \dots, X_m\}$ define an orthonormal frame for $\langle \cdot \mid \cdot \rangle$ on the distribution $\mathfrak{g}_1$. Such a basis will be referred also as an *adapted basis*. We can write the commutation relations:

$$\begin{cases} [X_i, X_j] = \sum_{h=1}^{n-m} c_{ij}^h Y_h, & i, j = 1, \dots, m, \quad \text{where} \quad c_{ij}^h = -c_{ji}^h, \\ [X_i, Y_j] = [Y_j, Y_h] = 0, & i = 1, \dots, m, \qquad j, h = 1, \dots, n-m. \end{cases} \tag{7.22}$$

Define the the $n-m$ skew-symmetric matrices (of size $m$) $C_h = (c_{ij}^h)$, for $h = 1, \dots, n-m$. We stress that, thanks to left-invariance, the structure functions $c_{ij}^h$ are constant.

Given an adapted basis, we can associate with the family of matrices $\{C_1, \dots, C_{n-m}\}$ the subspace

$$\mathcal{C} = \operatorname{span}\{C_1, \dots, C_{n-m}\} \subset \mathfrak{so}(\mathfrak{g}_1), \tag{7.23}$$

of skew-symmetric operators on $\mathfrak{g}_1$ that are represented by linear combinations of this family of matrices.

**Proposition 7.53** (2-step Carnot algebras and subspaces of $\mathfrak{so}(\mathfrak{g}_1)$)**.** *For a given a 2-step Carnot algebra $\mathfrak{g}$, the subspace $\mathcal{C} \subset \mathfrak{so}(\mathfrak{g}_1)$ is independent on the choice of the adapted basis on $\mathfrak{g}$.*

*Proof.* Assume that we fix another adapted basis

$$\mathfrak{g}_1 = \operatorname{span}\{X_1', \dots, X_m'\}, \qquad \mathfrak{g}_2 = \operatorname{span}\{Y_1', \dots, Y_{n-m}'\}.$$

where $\{X_1', \dots, X_m'\}$ is orthonormal for the inner prodict. Then there exists $A = (a_{ij})$ an orthogonal matrix and $B = (b_{hl})$ an invertible matrix such that

$$X_i' = \sum_{j=1}^{m} a_{ij} X_j, \qquad Y_h' = \sum_{l=1}^{n-m} b_{hl} Y_l.$$

A direct computation shows that, denoting $B^{-1} = (b^{hl})$, we have

$$[X'_i, X'_j] = \sum_{h,l=1}^{m} a_{ih} a_{jl} [X_h, X_l] = \sum_{h,l=1}^{m} a_{ih} a_{jl} \sum_{r=1}^{n-m} c_{hl}^r Y_r \tag{7.24}$$

$$= \sum_{s=1}^{n-m} \left( \sum_{r=1}^{n-m} \sum_{h,l=1}^{m} a_{ih} a_{jl} c_{hl}^r b^{rs} \right) Y'_s \tag{7.25}$$

it follows that

$$C'_s = \sum_{h=1}^{n-m} b^{hs} (A C_h A^*) \tag{7.26}$$

Recall that two matrices $C$ and $C'$ represents the same element of $\mathfrak{so}(\mathfrak{g}_1)$ with respect to the two basis if and only if $C' = ACA^*$. Then formula (7.26) implies that elements of $\mathcal{C}'$ are written as linear combination of elements of $\mathcal{C}$ that represents the same linear operator, as claimed. $\square$

*Remark* 7.54. We have the following basis-independent interpretation of Proposition 7.53. The Lie bracket defines a well-defined skew-symmetric bilinear map

$$[\cdot,\cdot] : \mathfrak{g}_1 \times \mathfrak{g}_1 \to \mathfrak{g}_2.$$

If we compose this map with an element $\xi \in \mathfrak{g}_2^*$ we get a skew-symmetric bilinear form $[\cdot,\cdot]_\xi := \xi \circ [\cdot,\cdot] : \mathfrak{g}_1 \times \mathfrak{g}_1 \to \mathbb{R}$. For every $\xi \in \mathfrak{g}_2^*$ the map $[\cdot,\cdot]_\xi$ can be identified with an element of $\mathfrak{so}(\mathfrak{g}_1)$, thanks to the inner product on $\mathfrak{g}_1$. Hence with every Carnot algebra of step 2 we can associate a well-defined linear map

$$\Psi : \mathfrak{g}_2^* \to \mathfrak{so}(\mathfrak{g}_1)$$

The subspace $\mathcal{C}$ introduced in (7.23) coincides with $\operatorname{im}\Psi \subset \mathfrak{so}(\mathfrak{g}_1)$.

**Definition 7.55.** Two Carnot algebras $\mathfrak{g}$ and $\mathfrak{g}'$ are *isomorphic* if there exists a Lie algebra isomorphism $\phi : \mathfrak{g} \to \mathfrak{g}'$ such that $\phi|_{\mathfrak{g}_1} : \mathfrak{g}_1 \to \mathfrak{g}'_1$ preserves the scalar products, i.e.,

$$\langle \phi(v) \,|\, \phi(w) \rangle' = \langle v \,|\, w \rangle, \qquad \forall\, v, w \in \mathfrak{g}.$$

Following the same arguments one can prove the following result.

**Corollary 7.56.** *The set of equivalence classes of 2-step Carnot algebras (with respect to isomorphisms) on $\mathfrak{g} = \mathfrak{g}_1 \oplus \mathfrak{g}_2$ is in one-to-one correspondence with the set of subspaces of $\mathfrak{so}(\mathfrak{g}_1)$.*

### 7.5.1 Normal Pontryagin extremals for Carnot groups of step 2

Let us fix a 2-step Carnot group $G$ and let $\mathfrak{g}$ its associated Lie algebra.

A basis of a Lie algebra of vector fields on $\mathbb{R}^n = \mathbb{R}^m \oplus \mathbb{R}^{n-m}$ (using coordinates $g = (x, z) \in \mathbb{R}^m \oplus \mathbb{R}^{n-m}$) and satisfying (13.11) is given by

$$X_i = \frac{\partial}{\partial x_i} - \frac{1}{2} \sum_{j=1}^{m} \sum_{\ell=1}^{n-m} c_{ij}^{\ell} x_j \frac{\partial}{\partial z_\ell}, \qquad i = 1, \dots, m, \tag{7.27}$$

$$Z_\ell = \frac{\partial}{\partial z_\ell}, \qquad \ell = 1, \dots, n-m. \tag{7.28}$$

The group $G$ is $\mathbb{R}^n = \mathbb{R}^m \oplus \mathbb{R}^{n-m}$ endowed with the group law

$$(x, z) * (x', z') = \left(x + x', z + z' + \frac{1}{2} Cx \cdot x'\right)$$

where we denoted for the $(n-m)$-tuple $C = (C_1, \ldots, C_{n-m})$ of $m \times m$ matrices, the product

$$Cx \cdot x' = (C_1 x \cdot x', \ldots, C_{n-m} x \cdot x') \in \mathbb{R}^{n-m}.$$

and $x \cdot x'$ denotes the Euclidean inner product in $\mathbb{R}^m$. Let us introduce the following coordinates on $T^*G$

$$h_i(\lambda) = \langle \lambda, X_i(g) \rangle, \qquad w_\ell(\lambda) = \langle \lambda, Z_\ell(g) \rangle.$$

Since the vector fields $\{X_1, \ldots, X_m, Z_1, \ldots, Z_{n-m}\}$ are linearly independent, the functions $(h_i, w_\ell)$ defines a system of coordinates on fibers of $T^*G$. In what follows it is convenient to use $(x, z, h, w)$ as coordinates on $T^*G$.

Normal Pontryagin extremal trajectories are projections of integral curves of the sub-Riemannian Hamiltonian in $T^*G$

$$H = \frac{1}{2} \sum_{i=1}^m h_i^2$$

Suppose now that $\lambda(t) = (x(t), z(t), h(t), \omega(t))$ is a normal Pontryagin extremal. Then $u_i(t) = h_i(\lambda(t))$ and the equation on the base is

$$\dot{g} = \sum_{i=1}^m h_i X_i(g). \tag{7.29}$$

that rewrites as

$$\begin{cases} \dot{x}_i = h_i \\ \dot{z}_h = -\frac{1}{2} \sum_{i,j=1}^m c_{ij}^\ell h_i x_j \end{cases} \tag{7.30}$$

For the equations on the fiber we have (remember that along solutions $\dot{a} = \{H, a\}$)

$$\begin{cases} \dot{h}_i = \{H, h_i\} = -\sum_{j=1}^m \{h_i, h_j\} h_j = -\sum_{\ell=1}^{n-m} \sum_{j=1}^m c_{ij}^\ell h_j w_\ell \\ \dot{w}_\ell = \{H, w_\ell\} = 0. \end{cases} \tag{7.31}$$

From (7.31) we easily get that $\omega_h$ is constant and the vector $h = (h_1, \ldots, h_m) \in \mathbb{R}^m$ satisfies the linear equation

$$\dot{h} = -\Omega_w h, \qquad \Omega_w = \sum_{\ell=1}^{n-m} w_\ell C_\ell,$$

where we recall that the vector $w = (w_1, \ldots, w_{n-m})$ is constant. It follows that

$$h(t) = e^{-t\Omega_w} h(0),$$

and

$$x(t) = x(0) + \int_0^t e^{-s\Omega_w} h(0) ds.$$

Notice that the vertical coordinates $z$ can be always recovered, once $h(t)$ and $x(t)$ are computed, by a simple integration.

**Proposition 7.57.** *The projection $x(t)$ on the layer $\mathfrak{g}_1 \simeq \mathbb{R}^m$ of a Pontryagin extremal such that $x(0) = 0$ is the image of the origin through a one-parametric group of isometries of $\mathbb{R}^m$.*

The proof of the proposition uses the following observation.

**Exercise 7.58.** Recall that the group of (positively oriented) affine isometries on $\mathbb{R}^n$ can be identified with the matrix group $SE(n)$ given by the formula (7.15). The Lie algebra of $SE(n)$ is given by

$$\mathfrak{se}(n) = \left\{ \begin{pmatrix} A & b \\ 0 & 0 \end{pmatrix}, A \in \mathfrak{so}(n), b \in \mathbb{R}^n \right\}.$$

Prove the following formula for the exponential of an element of the Lie algebra

$$\exp\left( t \begin{pmatrix} A & b \\ 0 & 0 \end{pmatrix} \right) = \begin{pmatrix} e^{tA} & \int_0^t e^{sA} b ds \\ 0 & 1 \end{pmatrix}.$$

*Proof of Proposition 7.57.* The action of a 1-parametric group of isometries can be recovered by exponentiating an element of its Lie algebra (cf. Exercice 7.58). This reduces to compute the solution of the differential equation

$$\dot{x} = Ax + b$$

where $A$ is skew-symmetric and $b \in \mathbb{R}^m$. Its flow is given by

$$\phi_t(\bar{x}) = e^{tA}\bar{x} + \int_0^t e^{sA} b ds$$

and it is easy to see that the projection $x(t)$ on the layer $\mathfrak{g}_1 \simeq \mathbb{R}^m$ of a Pontryagin extremal satisfies this equation with $\bar{x} = x(0) = 0$, $A = -\Omega_w$ and $b = h(0)$. $\square$

### Heisenberg group

The simplest example of 2-step Carnot group is the Heisenberg group, whose Lie algebra $\mathfrak{g}$ has dimension 3. It can be realized in $\mathbb{R}^3$ by the left-invariant vector fields

$$X_1 = \frac{\partial}{\partial x_1} - \frac{1}{2}x_2\frac{\partial}{\partial z}, \qquad X_2 = \frac{\partial}{\partial x_2} + \frac{1}{2}x_1\frac{\partial}{\partial z}, \qquad Z = \frac{\partial}{\partial z},$$

satisfying the relation $[X_1, X_2] = Z$. In this case the set of matrices representing the Lie bracket is reduced to a single matrix $C$, namely

$$C = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix},$$

and the projection $x(t)$ on the layer $\mathfrak{g}_1 \simeq \mathbb{R}^k$ of a Pontryagin extremal starting from the origin satisfies the equation

$$x(t) = \int_0^t \exp\begin{pmatrix} 0 & -ws \\ ws & 0 \end{pmatrix} h(0) ds.$$

Computing

$$\int_0^t \exp\begin{pmatrix} 0 & -ws \\ ws & 0 \end{pmatrix} ds = \frac{1}{w}\begin{pmatrix} \sin(wt) & \cos(wt) - 1 \\ -\cos(wt) + 1 & \sin(wt) \end{pmatrix},$$

and choosing $h(0) = (-\sin\theta, \cos\theta) \in S^1$, we recovers the formulas already computed[3] in Section 4.4.3. We have

$$h(t) = \begin{pmatrix} \cos(wt) & -\sin(wt) \\ \sin(wt) & \cos(wt) \end{pmatrix} \begin{pmatrix} -\sin\theta \\ \cos\theta \end{pmatrix} = \begin{pmatrix} -\sin(wt+\theta) \\ \cos(wt+\theta) \end{pmatrix},$$

and

$$x(t) = \frac{1}{w} \begin{pmatrix} \sin(wt) & \cos(wt)-1 \\ -\cos(wt)+1 & \sin(wt) \end{pmatrix} \begin{pmatrix} -\sin\theta \\ \cos\theta \end{pmatrix} = \frac{1}{w} \begin{pmatrix} \cos(wt+\theta) - \cos\theta \\ \sin(wt+\theta) - \sin\theta \end{pmatrix}.$$

Notice that the $z$ component is recovered simply by integrating the last equation, that in this case gives

$$\dot{z} = \frac{1}{2}(-h_1 x_2 + h_2 x_1).$$

Integrating (and using $z(0) = 0$) one gets

$$\begin{aligned} z(t) &= \frac{1}{2w} \int_0^t \sin(ws+\theta)(\sin(ws+\theta) - \sin\theta) + \cos(ws+\theta)(\cos(ws+\theta) - \cos\theta) ds \\ &= \frac{1}{2w} \int_0^t 1 - \sin(ws+\theta)\sin\theta - \cos(ws+\theta)\cos\theta ds = \frac{1}{2w} \int_0^t 1 - \cos(ws) ds \\ &= \frac{1}{2w^2}(wt - \sin(wt)). \end{aligned}$$

Analogous computations are performed for higher-dimensional Heisenberg groups in Section 13.1. See Figure 7.1.

## 7.6 Left-invariant Hamiltonian systems on Lie groups

In this section we study Hamiltonian systems non necessarily coming from a sub-Riemnnian problem.

### 7.6.1 Vertical coordinates in $TG$ and $T^*G$

Thanks to the isomorphism between $TG$ and $G \times T_{\mathbb{1}}G$, a basis $\{e_1, \ldots, e_n\}$ of $T_{\mathbb{1}}G$ induces global coordinates on $TG$. Indeed a basis of $T_gG$ is $L_{g*}e_1, \ldots, L_{g*}e_n$ and every element $(v, g)$ of $TG$ can be written as

$$(v, g) = \left( \sum_{i=1}^n v_i L_{g*} e_i, g \right).$$

The coordinates $v_1, \ldots v_n$ are called the *vertical coordinates* in $TG$ and they are also coordinates in the vertical part of $G \times T_{\mathbb{1}}G$. Indeed if $(v, g) = (\sum_{i=1}^n v_i L_{g*} e_i, g) \in TG$, then the corresponding point in $G \times T_{\mathbb{1}}G$ is $(\xi, g) = (\sum_{i=1}^n v_i e_i, \xi)$ hence, in coordinates, both are represented by $(v_1, \ldots, v_n, g)$.

If $\{e_1^*, \ldots, e_n^*\}$ is the dual basis in $T_{\mathbb{1}}^*G$ to $\{e_1, \ldots, e_n\}$, i.e., $\langle e_i^*, e_j \rangle = \delta_{i,j}$, then every element $(p, g)$ of $T^*G$ can be written as

$$(p, g) = \left( \sum_{i=1}^n h_i L_{g^{-1}}^* e_i^*, g \right).$$

[3] notice that in Section 4.4.3 we have set $h(0) = (\cos\theta, \sin\theta)$. The final formulas are the same up to a rotation in the horizontal plane.

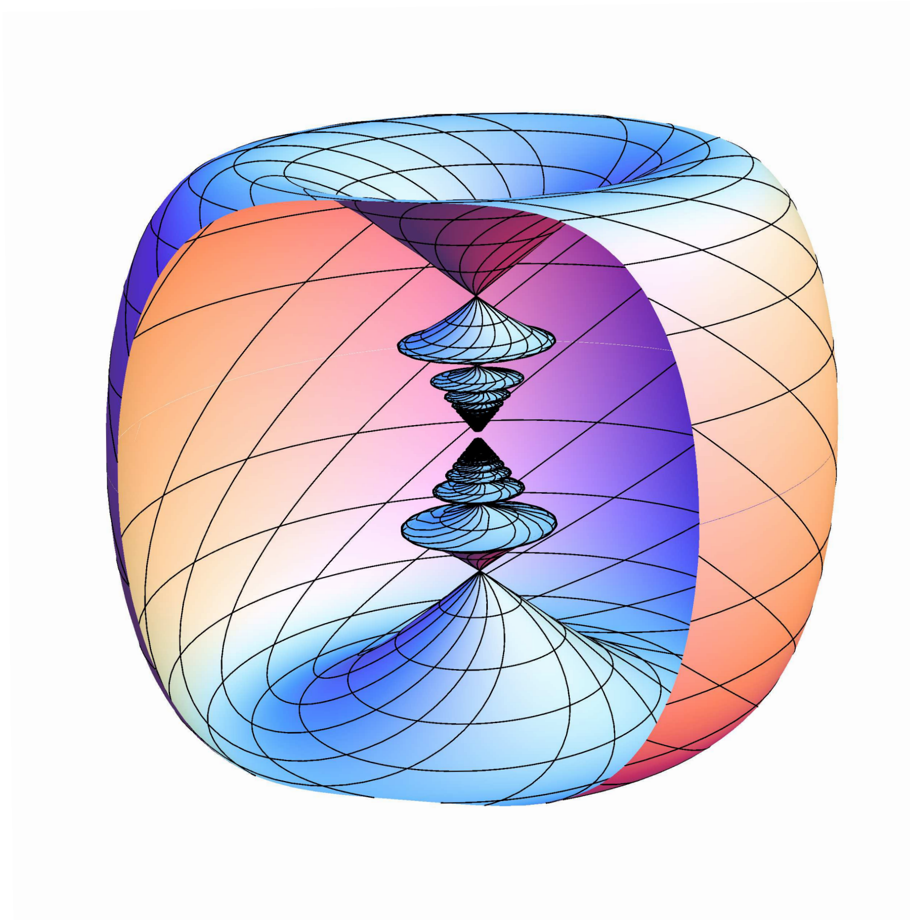


Figure 7.1: The set of end points of Pontryagin extremals of length 1 for the 3D Heisenberg group. Notice the singularities accumulating at the origin.

The coordinates $h_1, \dots h_n$ are called *vertical coordinates* in $T^*G$. For the same reason as above, in vertical coordinates $(h_1, \dots, h_n, g)$ represents both a point in $T^*G$ and the corresponding point in $G \times T^*_{\mathbb{1}}G$.

In other words, when using vertical coordinates it is not important to distinguish if we are working in $TG$ or $G \times T_{\mathbb{1}}G$ (the same holds for $T^*G$ or $G \times T^*_{\mathbb{1}}G$).

*Remark* 7.59. Notice that if $X_i(g) = L_{g*}e_i$ then

$$h_i(p, g) = \langle p, X_i(g) \rangle,$$

hence $h_i$ are the functions linear on fibers associated with $X_i$. Moreover if make the change of variable $(p, g) \to (\xi, g)$ where $p(\xi, g) = L^*_{g^{-1}}\xi$ where $\xi \in T^*_{\mathbb{1}}G$, we have that $h_i$ becomes independent from $g$. Indeed we can write

$$h_i(p(\xi, g), g) = \langle \xi, e_i \rangle_{\mathbb{1}}.$$

The vertical coordinates $h_1, \dots, h_n$ are functions on $T^*G$ hence we can compute their Poisson bracket (cf. Section 4.1.2)

$$\{h_i, h_j\} = \langle p, [X_i, X_j]\rangle_g = \langle \xi, [e_i, e_j]\rangle_{\mathbb{1}}. \tag{7.32}$$

*Remark* 7.60. Note that the vertical coordinates $h_i$ are not induced by a system of coordinates $x_1, \dots, x_n$ on the base $G$ (we have not fixed coordinates on $G$). If they were induced by coordinates on $G$, we would have obtained zero in the right-hand side of (7.32) since $[\partial_{x_i}, \partial_{x_j}] = 0$.

### 7.6.2 Left-invariant Hamiltonians

Consider a Hamiltonian function $H : T^*G \to \mathbb{R}$. Thanks to the isomorphism between $T^*G$ and $G \times T_{\mathbb{1}}G$ we can interpret it as a function on $G \times T_{\mathbb{1}}^*G$, i.e., we can define

$$\mathcal{H}(\xi, g) = H(g, L_{g^{-1}}^*\xi), \quad \mathcal{H} : G \times T_{\mathbb{1}}^*G \to \mathbb{R}.$$

We say that $H$ is left-invariant if $\mathcal{H}(\xi, g)$ is independent from $g$. For a left-invariant Hamiltonian we call the corresponding $\mathcal{H}$ the *trivialized Hamiltonian.*

Equivalently we can use the following definition

**Definition 7.61.** A Hamiltonian $H : T^*G \to \mathbb{R}$ is said to be left-invariant if there exists a function $\mathcal{H} : T_{\mathbb{1}}^*G \to \mathbb{R}$ such that

$$H(p, g) = \mathcal{H}(L_g^* p).$$

Hence a left-invariant Hamiltonian can be interpreted as a function on $T_{\mathbb{1}}^*G$.

**Example 7.62.** Given a set of left-invariant vector field $f_i(g) = L_{g*}w_i$, $w_i \in T_{\mathbb{1}}G$, $i = 1, \dots, m$, we have that $H(p, g) = \frac{1}{2}\sum_{i=1}^m \langle p, f_i(g)\rangle^2$ is a left-invariant Hamiltonian. Indeed

$$\mathcal{H}(\xi, g) = \frac{1}{2}\sum_{i=1}^m \langle L_{g^{-1}}^*\xi, L_{g*}w_i\rangle^2 = \frac{1}{2}\sum_{i=1}^m \langle \xi, w_i\rangle^2,$$

which is independent from $g$.

*Remark* 7.63. If we write $p = \sum_{j=1}^n h_j L_{g^{-1}}^* e_j^*$ then

$$H(\sum L_{g^{-1}}^* h_j e_j^*, g) = \mathcal{H}(L_g^* \sum h_j L_{g^{-1}}^* e_j^*) = \mathcal{H}(\sum h_j e_j^*).$$

In other words in vertical coordinates $h_1, \dots h_n$, we have for a left-invariant Hamiltonian

$$H(h_1, \dots, h_n, g) = \mathcal{H}(h_1, \dots, h_n).$$

and we can identify $H$ and $\mathcal{H}$.

*Remark* 7.64. In the context of Hamiltonian systems on Lie groups, it is convenient to avoid fixing coordinates on $G$ and use vertical coordinates on the fiber only. This permits to exploit better the trivialization of $T^*G$ in $G \times T_{\mathbb{1}}^*G$ and the left invariance of $H$. Since vertical coordinates $h_i$ do not come, in general, from coordinates on $G$, we do not have equations of the form $\dot{x}_i = \partial_{h_i}H$, $\dot{h}_i = -\partial_{x_i}H$ for a system of coordinates $x_1, \dots, x_n$ on $G$.

Consider a left-invariant Hamiltonian in vertical coordinates $H(h_1, \ldots, h_n, g)$. Let us write the vertical part of the Hamiltonian equations. We are going to see that this equation is particularly simple. We have

$$\dot{h}_i = \{H, h_i\}, \qquad i = 1, \ldots, n. \tag{7.33}$$

Using Exercice 4.8 we have for $i = 1, \ldots, n$,

$$\dot{h}_i = \sum_{j=1}^{n} \frac{\partial H}{\partial h_j} \{h_j, h_i\} = \sum_{j=1}^{n} \frac{\partial H}{\partial h_j} \langle \xi, [e_j, e_i] \rangle = \left\langle \xi, \left[ \sum_{j=1}^{n} \frac{\partial H}{\partial h_j} e_j, e_i \right] \right\rangle. \tag{7.34}$$

Notice that since $\mathcal{H}$ is a function on the linear space $T^*_{\mathbb{1}}G$, then $d\mathcal{H}(h_1, \ldots, h_n)$ is an element of $(T^*_{\mathbb{1}}G)^* = T_{\mathbb{1}}G$. If we write an element of $T^*_{\mathbb{1}}G$ as $h_1 e_1^* + \ldots + h_n e_n^*$, then an element of its tangent at $(h_1, \ldots, h_n)$ is written as $v_1 \partial_{h_1} + \ldots + v_n \partial_{h_n}$ with the identification $\partial_{h_i} = e_i^*$ due to the linear structure. An element of its cotangent space $(T^*_{\mathbb{1}}G)^*$ at $(h_1, \ldots, h_n)$ is then written as $\omega_1 dh_1 + \ldots + \omega_n dh_n$ with the identification $dh_i = (e_i^*)^* = e_i$ again due to the linear structure. Then

$$d\mathcal{H}(h_1, \ldots, h_n) = \sum_{j=1}^{n} \frac{\partial \mathcal{H}}{\partial h_j} dh_j = \sum_{j=1}^{n} \frac{\partial \mathcal{H}}{\partial h_j} e_j = \sum_{j=1}^{n} \frac{\partial H}{\partial h_j} e_j. \tag{7.35}$$

Hence the vertical part of the Hamiltonian equations can be written as

$$\begin{aligned} \dot{h}_i &= \langle \xi, [d\mathcal{H}, e_i] \rangle \\ &= \langle \xi, (\operatorname{ad} d\mathcal{H}) e_i \rangle \\ &= \langle (\operatorname{ad} d\mathcal{H})^* \xi, e_i \rangle \end{aligned} \tag{7.36}$$

or more compactly recalling that $\xi = \sum_{i=1}^{n} h_i e_i^*$,

$$\dot{\xi} = (\operatorname{ad} d\mathcal{H})^* \xi. \tag{7.37}$$

For what concerns the horizontal part, let $\beta \in \mathcal{C}^\infty(G)$, i.e., a function in $\mathcal{C}^\infty(T^*G)$ that is constant on fibers. For every curve $g(\cdot)$ solution of the horizontal part of the Hamiltonian system on $T^*G$ corresponding to $H$ we have

$$\frac{d}{dt} \beta(g(t)) = \{H, \beta\}_{(p(t), g(t))} = \sum_{j=1}^{n} \frac{\partial H}{\partial h_j} \{h_j, \beta\}_{(p(t), g(t))}.$$

Now recalling that (cf. (4.17)) $\{\langle p, X(g) \rangle + \alpha(g), \langle p, Y(g) \rangle + \beta(g)\} = \langle p, [X, Y](g) \rangle + X\beta(g) - Y\alpha(g)$ we have $\{h_j, \beta\} = \{\langle p, X_j \rangle, \beta\} = X_j \beta = (L_{g*} e_j) \beta$. Hence

$$\frac{d}{dt} \beta(g(t)) = \sum_{j=1}^{n} \frac{\partial H}{\partial h_j} (L_{g*} e_j) \beta \Bigg|_{g(t)} = \left( L_{g*} \sum_{j=1}^{n} \frac{\partial H}{\partial h_j} e_j \right) \beta \Bigg|_{g(t)} = \left( L_{g*} d\mathcal{H} \right) \beta \big|_{g(t)}.$$

Since the function $\beta$ is arbitrary we have $\dot{g} = L_{g*} d\mathcal{H}$.

We have then proved the following

**Proposition 7.65.** *Let $H$ be a left-invariant Hamiltonian on a Lie group $G$, i.e., $H(p,g) = \mathcal{H}(L_g^* p)$ where $(p,g) \in T^*G$ and $\mathcal{H}$ is a smooth function from $T_{\mathbb{1}}^*G$ to $\mathbb{R}$. Let $d\mathcal{H}$ be the differential of $\mathcal{H}$ seen as an element of $T_{\mathbb{1}}G$. Then the Hamiltonian equations $\frac{d}{dt}(p,g) = \vec{H}(p,g)$ are,*

$$\begin{cases} \dot{g} = L_{g*} d\mathcal{H} \\ \dot{\xi} = (\operatorname{ad} d\mathcal{H})^* \xi. \end{cases} \tag{7.38}$$

*Here $\xi \in T_{\mathbb{1}}^*G$ and $p(t) = L_{g^{-1}}^* \xi(t)$.*

Notice that the second equation is decoupled from the first since $d\mathcal{H}$ is a function of $\xi$ only (it does not involve $g$).

When the space is endowed with a bi-invariant pseudometric, then equation (7.37) can be written in a simpler form. Indeed in this case we can identify an element $\xi \in T_{\mathbb{1}}G$ with $M \in T_{\mathbb{1}}^*G$ by

$$\langle M \mid v \rangle = \langle \xi, v \rangle, \quad \forall v \in T_{\mathbb{1}}G. \tag{7.39}$$

Using (7.37) and (7.16), for every $v \in T_{\mathbb{1}}G$ let us compute

$$\begin{aligned} \left\langle \frac{dM}{dt} \,\middle|\, v \right\rangle &= \left\langle \frac{d\xi}{dt}, v \right\rangle = \langle (\operatorname{ad} d\mathcal{H})^* \xi, v \rangle = \langle \xi, (\operatorname{ad} d\mathcal{H}) v \rangle \\ &= \langle \xi, [d\mathcal{H}, v] \rangle = \langle M \mid [d\mathcal{H}, v] \rangle = \langle [M, d\mathcal{H}] \mid v \rangle . \end{aligned}$$

Hence the Hamiltonian equations for a left-invariant Hamiltonian, when we have a bi-invariant pseudometric, are:

$$\begin{cases} \dot{g} = L_{g*} d\mathcal{H} \\ \frac{dM}{dt} = [M, d\mathcal{H}]. \end{cases} \tag{7.40}$$

## 7.7 Normal extremals for left-invariant sub-Riemannian structures

Consider a left-invariant sub-Riemannian structure of rank $m$ (cf. (7.20)) for which an orthonormal frame is given by a set of left-invariant vector fields $X_i = L_{g*} e_i(g)$, $i = 1, \ldots, m$. The maximized Hamiltonian is

$$H(p,g) = \frac{1}{2} \sum_{i=1}^m \langle p, X_i(g) \rangle^2 = \frac{1}{2} \sum_{i=1}^m \langle p, L_{g*} e_i \rangle^2 ,$$

hence it is left-invariant (cf. Example 7.62). The corresponding trivialized Hamiltonian is

$$\mathcal{H}(\xi) = \frac{1}{2} \sum_{i=1}^m \langle \xi, e_i \rangle^2 .$$

Now $\langle \xi, e_i \rangle = h_i(p,g)$ hence in vertical coordinates we have

$$H(h_1, \ldots, h_m) = \frac{1}{2} \sum_{i=1}^m h_i^2.$$

### 7.7.1 Explicit expression of normal Pontryagin extremals in the $\mathbf{d} \oplus \mathbf{s}$ case

Explicit expressions of normal Pontryagin extremals can be obtained for left-invariant sub-Riemannain structures when

- a bi-invariant pseudo-metric $\langle \cdot \,|\, \cdot \rangle$ on $G$ is given;
- $T_{\mathbb{1}}G = \mathbf{d} \oplus \mathbf{s}$ where $\langle \cdot \,|\, \cdot \rangle|_{\mathbf{d}}$ is positive defined and $\mathbf{s}$ satisfies the following
  i) $\mathbf{s} := \mathbf{d}^{\perp}$ (where the orthogonality is taken with respect to $\langle \cdot \,|\, \cdot \rangle$);
  ii) $\mathbf{s}$ is a sub-algebra;
- The distribution is $\mathbf{d}$ and the metric is $\langle \cdot \,|\, \cdot \rangle|_{\mathbf{d}}$.

We say that such a sub-Riemannian structure is of type $\mathbf{d} \oplus \mathbf{s}$.

*Remark* 7.66. A classical example of such a $\mathbf{d} \oplus \mathbf{s}$ sub-Riemannian structure is provided by the group of matrices $SO(n)$ in which the distribution at the identity $\mathbf{d}$ is given by any codimension one subspace of $T_{\mathbb{1}}SO(n)$ and the norm of a vector in $\mathbf{d}$ is the square root of the sum of squares of its matrix elements.

**Exercise 7.67.** Prove that the distribution defined in Remark 7.66 is bracket-generating. Prove that the metric induced by the norm defined above is induced (up to a negative proportionality constant) by the Killing form.

Let us write an element of $v \in T_{\mathbb{1}}G$ as $v = x + y$ where $x \in \mathbf{d}$ and $y \in \mathbf{s}$. Set $m = \dim \mathbf{d}$. Let $e_1, \dots e_m$ be an orthonormal frame for the structure. In this case if $M = x + y$ is the element in $T_{\mathbb{1}}G$ corresponding to $\xi \in T_{\mathbb{1}}^*G$ via $\langle \cdot \,|\, \cdot \rangle$ (cf. (7.39)) we have

$$h_i = \langle \xi, e_i \rangle = \langle M \,|\, e_i \rangle = x_i.$$

Hence

$$\mathcal{H} = \frac{1}{2}\sum_{i=1}^{m} h_i^2 = \frac{1}{2}\sum_{i=1}^{m} x_i^2 = \frac{1}{2}\langle x \,|\, x \rangle = \frac{1}{2}\|x\|^2. \tag{7.41}$$

Notice that (cf. (7.35)) $d\mathcal{H} = \sum_{i=1}^n \frac{\partial \mathcal{H}}{\partial h_i} e_i = \sum_{i=1}^n \frac{\partial \mathcal{H}}{\partial x_i} e_i = \sum_{i=1}^n x_i e_i = x$. Hence the vertical part of the Hamiltonian equation $dM/dt = [M, d\mathcal{H}]$ become

$$\dot{x} + \dot{y} = [x + y, x] = [y, x]. \tag{7.42}$$

Now for every $v \in \mathbf{s}$ one has

$$\langle [y, x] \,|\, v \rangle = \langle x \,|\, [y, v] \rangle = 0,$$

where we have used equation (7.16) and, for the last equality, the fact that

- $[y, v] \in \mathbf{s}$ since $\mathbf{s}$ is a sub-algebra.
- $\mathbf{d}$ and $\mathbf{s}$ are orthogonal for $\langle \cdot \,|\, \cdot \rangle$.

We then conclude that $[y, x] \in \mathbf{d}$. Hence (7.42) become

$$\begin{aligned} \dot{x} &= [y, x] \\ \dot{y} &= 0 \end{aligned}$$

Hence all $y$ component are constant of the motion and we have (fixing $x(0) = x_0$ and $y(0) = y_0$)

$$\begin{aligned} y(t) &= y_0 \\ \dot{x} &= [y_0, x] = (\operatorname{ad} y_0)x \end{aligned}$$

The solution of the last equation is

$$x(t) = e^{t \operatorname{ad} y_0} x_0. \tag{7.43}$$

Then for the horizontal part we have

$$\dot{g} = L_{g*} d\mathcal{H} = L_{g*} x(t) = L_{g*} e^{t \operatorname{ad} y_0} x_0. \tag{7.44}$$

Using the variation formula for smooth vector fields (cf. (6.42)),

$$e^{t(Y+X)} = \overrightarrow{\exp} \int_0^t e^{s \operatorname{ad} Y} X ds \circ e^{tY}, \tag{7.45}$$

we have that the solution of (7.44) starting from $g_0$ and corresponding to $x_0, y_0$ is [4]

$$g(x_0, y_0; t) = g_0 e^{t(x_0 + y_0)} e^{-t y_0} \tag{7.46}$$

The parameterization by arclength is obtained requiring $H = 1/2$. From (7.41) at $t = 0$ we obtain that the normal Pontryagin extremals (7.46) are parametrized by arclength when $\langle x_0 \,|\, x_0 \rangle = \|x_0\|^2 = 1$.

The controls whose corresponding trajectories starting from $g_0$ are the normal Pontryagin extremals (7.46) are

$$u_i(t) = \langle p(t), X_i(g(t)) \rangle = h_i(p(t), g(t)) = x_i(t) = \left\langle e^{t \operatorname{ad} y_0} x_0 \,\middle|\, e_i \right\rangle, \quad i = 1, \ldots, m.$$

**Exercise 7.68.** Study abnormal extremals for this problem.

### 7.7.2 Example: The $\mathbf{d} \oplus \mathbf{s}$ problem on $SO(3)$

The Lie group $SO(3)$ is the group of special orthogonal $3 \times 3$ real matrices

$$SO(3) = \left\{ g \in \operatorname{Mat}(3, \mathbb{R}) \mid g g^T = \operatorname{Id}, \det(g) = 1 \right\}.$$

To compute its Lie algebra, let us compute its tangent space at the identity. Consider a smooth curve $g : [0, \varepsilon] \to SO(3)$, such that $g(0) = \mathbb{1}$. Computing the derivative in zero of both sides of the equation $g(t) g^T(t) = e$, we have $\dot{g}(0) g(0) + g(0) g^T(0) = 0$ from which we deduce $g(0) = -g^T(0)$.

[4] For a group of matrices: formula (7.43) reads as $e^{t y_0} x_0 e^{-t y_0}$, while (7.44) is $g e^{t y_0} x_0 e^{-t y_0}$.

Hence the Lie algebra of $SO(3)$ is the space of skew symmetric $3 \times 3$ real matrices and it is usually denoted by $\mathfrak{so}(3)$. In other words

$$\mathfrak{so}(3) = \left\{ \begin{pmatrix} 0 & -a & b \\ a & 0 & -c \\ -b & c & 0 \end{pmatrix} \in \mathrm{Mat}(3, \mathbb{R}) \right\}.$$

A basis of $\mathfrak{so}(3)$ is $\{e_1, e_2, e_3\}$ where

$$e_1 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}, \quad e_2 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix}, \quad e_3 = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

whose commutation relations are $[e_1, e_2] = e_3 \quad [e_2, e_3] = e_1 \quad [e_3, e_1] = e_2$. For $\mathfrak{so}(3)$ the Killing form is $K(X, Y) = \mathrm{trace}(XY)$ so, in particular, $K(e_i, e_j) = -2\delta_{ij}$. Hence

$$\langle \cdot \,|\, \cdot \rangle = -\frac{1}{2} K(\cdot, \cdot)$$

is a (positive definite) bi-invariant metric on $\mathfrak{so}(3)$. If we define

$$\mathbf{d} = \mathrm{span}\{e_1, e_2\}, \quad \mathbf{s} = \mathrm{span}\{e_3\}$$

and we provide $\mathbf{d}$ with the metric $\langle \cdot \,|\, \cdot \rangle |_{\mathbf{d}}$ we get a sub-Riemannian structre of type $\mathbf{d} \oplus \mathbf{s}$.

**Expression of normal Pontryagin extremals**

Let us write an initial covector $x_0 + y_0$ such that $\langle x_0 \,|\, x_0 \rangle = 1$ in the following form

$$x_0 + y_0 = \underbrace{\cos(\theta) e_1 + \sin(\theta) e_2}_{x_0} + \underbrace{c e_3}_{y_0}, \quad \theta \in S^1, \quad c \in \mathbb{R}.$$

Using formula (7.46), we have that the normal Pontryagin extremals starting from the identity are

$$g(\theta, c; t) := e^{(\cos(\theta) e_1 + \sin(\theta) e_2 + c e_3) t} e^{-c e_3 t} = \tag{7.47}$$

$$= \begin{pmatrix} K_1 \cos(ct) + K_2 \cos(2\theta + ct) + K_3 c \sin(ct) & K_1 \sin(ct) + K_2 \sin(2\theta + ct) - K_3 c \cos(ct) & K_4 \cos(\theta) + K_3 \sin(\theta) \\ -K_1 \sin(ct) + K_2 \sin(2\theta + ct) + K_3 c \cos(ct) & K_1 \cos(ct) - K_2 \cos(2\theta + ct) + K_3 c \sin(ct) & -K_3 \cos(\theta) + K_4 \sin(\theta) \\ K_4 \cos(\theta + ct) - K_3 \sin(\theta + ct) & K_3 \cos(\theta + ct) + K_4 \sin(\theta + ct) & \frac{\cos\left(\sqrt{1+c^2} t\right) + c^2}{1+c^2} \end{pmatrix}$$

with $K_1 = \frac{1 + \left(1 + 2c^2\right) \cos\left(\sqrt{1+c^2} t\right)}{2(1+c^2)}$, $K_2 = \frac{1 - \cos\left(\sqrt{1+c^2} t\right)}{2(1+c^2)}$, $K_3 = \frac{\sin\left(\sqrt{1+c^2} t\right)}{\sqrt{1+c^2}}$, $K_4 = \frac{c\left(1 - \cos\left(\sqrt{1+c^2} t\right)\right)}{1+c^2}$.

The end point of all normal Pontryagin extremals for $t = 1$ are plotted in Figure 7.2.

### 7.7.3 Further comments on the $\mathbf{d} \oplus \mathbf{s}$ problem: $SO(3)$ and $SO_+(2, 1)$

The group $SO(3)$ acts on the sphere $S^2$ by isometries (in fact, by definition). We claim that the induced action of $SO(3)$ on the spherical bundle $S\, S^2$ (see Definition 1.20) is a free transitive action. In other words, if $x_i \in S^2$, and $v_i \in T_{x_i} S^2$ with $|v_i| = 1$ for $i = 1, 2$, then there exists a unique $g \in SO(3)$ such that $g x_1 = x_2$, $g v_1 = v_2$. Indeed, $v$ is a tangent vector of length 1 at a point $x \in S^2$ if and only if $\{v, x\}$ is a couple of mutually orthogonal vectors of length 1 in $\mathbb{R}^3$. Obviously, such a

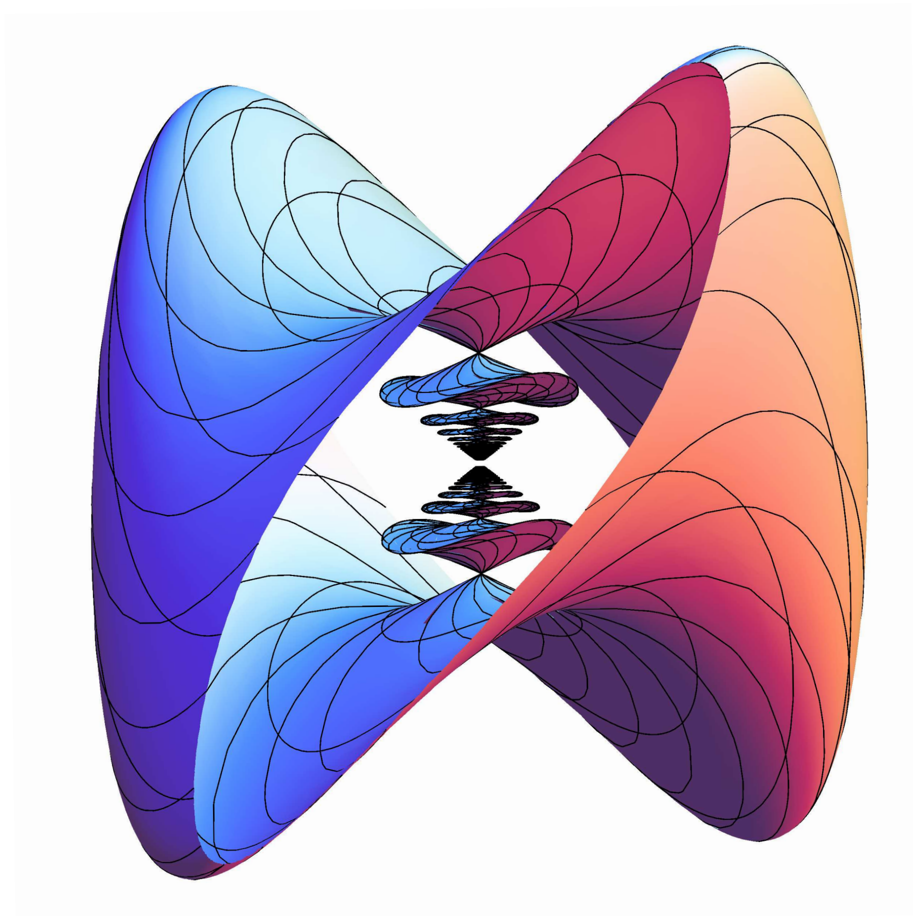


Figure 7.2: The set of end points of normal Pontryagin extremals of length 1 for the $\mathbf{d} \oplus \mathbf{s}$ sub-Riemannian problem on $SO(3)$. In the picture the $x$-axis is the element $(g)_{23}$, the $z$-axis is the element $(g)_{13}$, the $z$-axis is the element $(g)_{12}$. Notice the singularities accumulating at the origin. This picture looks very similar to the one of the Heisenberg group (cf. Figure 7.1). Indeed it is possible to prove (cf. Chapter 10) that the two pictures become more and more similar if one considers end points of normal Pontryagin extremals of length $r$ going to zero. For $r$ big enough, the two pictures become very different due to the different topology of $\mathbb{R}^3$ and $SO(3)$.

couple can be transformed to any other couple of this type by a unique orthogonal transformation of $\mathbb{R}^3$ preserving the orientation.

Let $g(t)$ be a geodesic for our sub-Riemannian structure on $SO(3)$. Then $g(t)\left(\begin{smallmatrix}0\\0\\1\end{smallmatrix}\right)$ is a circle, a curve of the constant geodesic curvature on the sphere. This is not occasional; if you think about it, you see that this sub-Riemannian problem is similar to isoperimetric problems studied in Section 4.4.2.

**Exercise 7.69.** Show that the differential of the map

$$SO(3) \to S\,S^2, \qquad g \mapsto \left(g\left(\begin{smallmatrix}0\\0\\1\end{smallmatrix}\right), g\left(\begin{smallmatrix}1\\0\\0\end{smallmatrix}\right)\right), \tag{7.48}$$

transforms the left-invariant distribution $\mathbf{d}$ into the kernel of the Levi-Civita connection (cf. Definition 1.54) on the spherical bundle $S\,S^2$.

Let $\omega$ be the Levi-Civita connection and $\pi : S\,S^2 \to S^2$ the standard projection; then $\pi_*\big|_{\ker \omega_\xi}$ is an isomorphism of $\ker \omega_\xi$ onto $T_{\pi(\xi)}S^2$, $\xi \in S\,S^2$. We can lift Riemannian structure on $S^2$ by this isomorphism and obtain a sub-Riemannian structure on $S\,S^2$. It is easy to see that the diffeomorphism described in the exercise induces an isometry of this sub-Riemannian structure and the "$\mathbf{d} \oplus \mathbf{s}$" structure on $SO(3)$.

Recall that an isoperimetric problem on a Riemannian surface $M$ is equivalent to a sub-Riemannian problem on the trivial bundle $\mathbb{R} \times M \to M$; the problem is defined by a non-vanishing differential 1-form $\omega$ on $\mathbb{R} \times M$, where $\omega$ is invariant under translations of $\mathbb{R}$ and $\ker \omega$ is transversal to the fibers (see Section 4.4.2). In this case, $d\omega$ is the pullback of a 2-form on $M$. Moreover, the 2-form is the product of the area form and a function $b$ on $M$, and normal geodesics are horizontal lifts to $\mathbb{R} \times M$ of the curves on $M$ whose geodesic curvature is proportional to $b$.

Of course, one gets the same characteristic of normal geodesic if we consider the bundle $S^1 \times M \to M$ instead of the bundle $\mathbb{R} \times M \to M$ and a non-vanishing form $\omega$ on $S^1 \times M$ that is invariant under translations in the group $S^1$ and whose kernel is transversal to the fibers. Moreover, we may equally consider an only locally trivial bundle $N \xrightarrow{S^1} M$ such that the group $S^1$ acts freely on $N$ and the orbits of this action are exactly the fibers of the bundle. Such a structure is called a *principal bundle with the structural group* $S^1$. An invariant under the action of $S^1$ non-vanishing 1-form on $N$ whose kernel is transversal to the fibers is called a *connection on the principal bundle*. The differential of the connection is the pullback of a 2-form on $M$ that is called the curvature of the connection.

Now consider the spherical bundle $SM \to M$ of a Riemannian surface. Rotations of the fibers with a constant velocity introduce a structure of the principal bundle on $SM$, and the Levi-Civita connection $\omega$ is a connection on this principal bundle. The curvature of the Levi-Civita connection equals the area form multiplied by the Gaussian curvature of the surface.

The sub-Riemannian structure defined by the Levi-Civita connection has a nice geometric interpretation: horizontal curves are parallel transports of tangent vectors along curves in $M$ and their length is just the length of these curves in $M$. Normal geodesics are parallel transports along the curves whose geodesic curvature is proportional to the Gaussian curvature. As we explained, in the case of $M = S^2$ we obtain an interpretation of the "$\mathbf{d} \oplus \mathbf{s}$" structure on $SO(3)$.

Group $SO(3)$ is the group of linear transformations of of $\mathbb{R}^3$ that preserve the orientation and Euclidean inner product. Similarly, we may consider the group $SO_+(2,1)$ of linear transformations that preserve the orientation, the Minkowski inner product $\langle \cdot \,|\, \cdot \rangle_h$ and, moreover, preserve the connected components of the hyperboloid defined by the equation $\langle q \,|\, q \rangle_h = -1$ (see Section 1.4). The matrices

$$f_1 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \quad f_2 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \quad f_3 = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} = e_3$$

form a basis of the Lie algebra of this group. This Lie algebra is denoted by $\mathfrak{so}(2,1)$ and it is isomorphic to $\mathfrak{sl}(2)$. We set $\langle X|Y \rangle = -\frac{1}{2}\mathrm{trace}(XY)$, a bi-invariant pseudo-metric on $\mathfrak{so}(2,1)$. If we define

$$\mathbf{d} = \mathrm{span}\{f_1, f_2\}, \quad \mathbf{s} = \mathrm{span}\{f_3\}$$

and we equip $\mathbf{d}$ with the metric $\langle\cdot|\cdot\rangle|_{\mathbf{d}}$ we obtain a sub-Riemannian structure of type $\mathbf{d}\oplus\mathbf{s}$.

The group $SO_+(2,1)$ acts on the surface

$$H^2 = \{(x,y,z) \in \mathbb{R}^3 : z^2 - x^2 - y^2 = 1,\ z > 0\},$$

in the Minkowski space by isometries (cf. Section 1.5.3). Moreover, the induced action of $SO_+(2,1)$ on the spherical bundle $SH^2$ is a free transitive action.

**Exercise 7.70.** Show that the differential of the map

$$SO_+(2,1) \to SH^2, \qquad g \mapsto \left(g\left(\begin{smallmatrix}0\\0\\1\end{smallmatrix}\right), g\left(\begin{smallmatrix}1\\0\\0\end{smallmatrix}\right)\right), \tag{7.49}$$

transforms the left-invariant distribution $\mathbf{d}$ into the kernel of the Levi-Civita connection on the spherical bundle $SH^2$.

The transformation (7.49) sends geodesics of the "$\mathbf{d}\oplus\mathbf{s}$" sub-Riemannian structure to the parallel transports along the curves of constant geodesic curvature in $H^2$. Recall that, when considered as Riemannian surface, $H^2$ has constant Gaussian curvature equal to $-1$, this is a model of the Lobachevsky hyperbolic plane.

The constructions described above have important multidimensional generalizations; some of them will be discussed later in this chapter.

### 7.7.4 Explicit expression of normal Pontryagin extremals in the $\mathbf{k}\oplus\mathbf{z}$ case

Another case in which one can get an explicit expression of normal Pontryagin extremals is when

- $G = G_{\mathbf{k}} \times G_{\mathbf{z}}$ where $G_{\mathbf{k}}$ has a compact algebra $\mathbf{k}$ and $G_{\mathbf{z}}$ is abelian. In other words the Lie algebra at the origin of $G$ can be written as $T_{\mathbb{1}}G = \mathbf{k}\oplus\mathbf{z}$ where $\mathbf{k}$ is a compact subalgebra and $\mathbf{z}$ is contained in the center of $T_{\mathbb{1}}G$, i.e., $[v,y] = 0$ for every $v \in T_{\mathbb{1}}G$ and $y \in \mathbf{z}$. In the following we write an element of $v \in T_{\mathbb{1}}G$ as $v = x + y$ where $x \in \mathbf{k}$ and $y \in \mathbf{z}$. Moreover we assume that a bi-invariant metric $\langle\cdot\,|\,\cdot\rangle_{\mathbf{k}}$ on $\mathbf{k}$ is given (this is always possible by definition of compact Lie algebra);

- we assume that the distribution (that we assume to be bracket-generating) projects well on $\mathbf{k}$, that is if $\pi : T_{\mathbb{1}}G \to \mathbf{k}$ is the canonical projection induced by the splitting, we have $\pi|_D$ is 1:1 over $\mathbf{k}$. Under this condition, there exists a linear operator $A : \mathbf{k} \to \mathbf{z}$ such that $\mathbf{d} = \{x + Ax \mid x \in \mathbf{k}\} \subset \mathbf{k}\oplus\mathbf{z} = T_{\mathbb{1}}G$.

- we assume that the metric on $\mathbf{d}$ is induced by the projection, i.e.,

  $$\langle w_1\,|\,w_2\rangle_{\mathbf{d}} = \langle \pi(w_1)\,|\,\pi(w_2)\rangle_{\mathbf{k}}, \qquad \text{for every } w_1, w_2 \in \mathbf{d},$$

  or equivalently that if $v_1, v_2 \in \mathbf{d}$, $v_1 = (x_1, Ax_1)$, $v_2 = (x_2, Ax_2)$ with $x_1, x_2 \in \mathbf{k}$, then

  $$\langle v_1\,|\,v_2\rangle_{\mathbf{d}} = \langle x_1\,|\,x_2\rangle_{\mathbf{k}}.$$

  See Figure 7.3.

Let us fix any scalar product on $\langle\cdot\,|\,\cdot\rangle_{\mathbf{z}}$ on $\mathbf{z}$ and define the scalar product $\langle\cdot\,|\,\cdot\rangle$ on $T_{\mathbb{1}}G$ by

$$\langle v_1\,|\,v_2\rangle = \langle x_1\,|\,x_2\rangle_{\mathbf{k}} + \langle y_1\,|\,y_2\rangle_{\mathbf{z}}, \quad \text{where } v_1 = x_1 + y_1, \quad v_2 = x_2 + y_2.$$

Notice that if $x \in \mathbf{k}$ and $y \in \mathbf{z}$ then $\langle x\,|\,y\rangle = 0$.

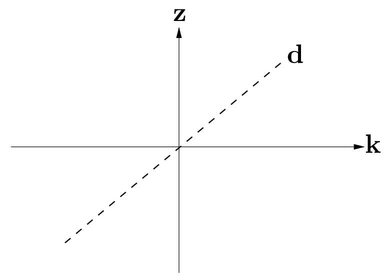



Figure 7.3: The $\mathbf{k} \oplus \mathbf{z}$ problem

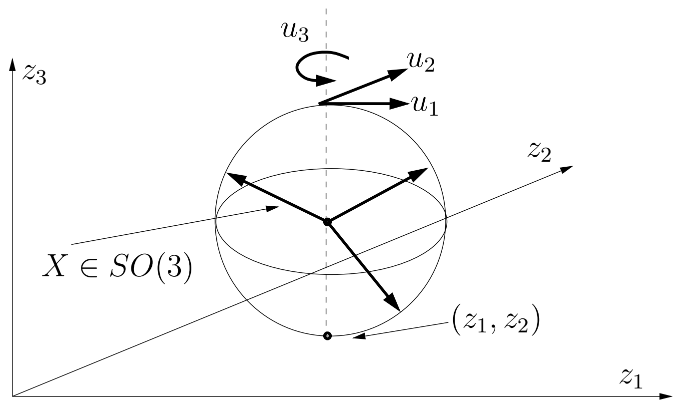



Figure 7.4: Rolling sphere with twisting.

**Exercise 7.71.** Prove that $\langle \cdot \,|\, \cdot \rangle$ is bi-invariant as a consequence of the bi-invariance of $\langle \cdot \,|\, \cdot \rangle_{\mathbf{k}}$ and of the fact that $\mathbf{z}$ is in the center of $T_{\mathbb{1}}G$.

The metric $\langle \cdot \,|\, \cdot \rangle_{T_{\mathbb{1}}G}$ is used to identify vectors and covectors, to use the simpler form (7.40) of the Hamiltonian equations for normal Pontryagin extremals. The resulting normal Pontryagin extremals will be independent on the choice of the scalar product $\langle \cdot \,|\, \cdot \rangle_{\mathbf{z}}$.

*Remark* 7.72. An example of such a structure is provided by the problem of rolling without slipping a sphere of radius 1 in $\mathbb{R}^3$ on a plane. Its state is described by a point in $\mathbb{R}^2$ giving the projection of its center on the plane and by an element of $SO(3)$ describing its orientation. Given an initial and final position in $SO(3) \times \mathbb{R}^2$ one would like to roll the sphere on the plane in such a way that the initial and final conditions are the given ones and $\int_0^T \sqrt{\sum_{i=1}^3 u_i(t)^2}\, dt$ is minimal, where $u_1$, $u_2$ and $u_3$ are the three controls corresponding to the rolling of the sphere along the two axes of the plane and to the twist. See Figure 7.4. Why this problem gives rise to a $\mathbf{k} \oplus \mathbf{z}$ sub-Riemannian structure is described in detail in the next section.

Let us write the maximized Hamiltonian. Let $e_1, \ldots, e_m$ be an orthonormal frame for $\mathbf{k}$. Then an orthonormal frame for $\mathbf{d}$ is $e_1 + Ae_1, \ldots, e_m + Ae_m$. We have

$$H(p, g) = \frac{1}{2} \sum_{i=1}^{m} \langle p, L_{g*}(e_i + Ae_i) \rangle^2 .$$

The corresponding trivialized Hamiltonian is

$$\mathcal{H}(\xi) = \frac{1}{2}\sum_{i=1}^{m} \langle \xi, (e_i + Ae_i)\rangle^2, \quad \xi \in T^*_{\mathbb{1}}G.$$

Now using the metric $\langle\cdot\,|\,\cdot\rangle_{T_{\mathbb{1}}G}$ we can identify $T_{\mathbb{1}}G$ with $T^*_{\mathbb{1}}G$ and write $\xi = x + y$. Then

$$\mathcal{H}(x,y) = \frac{1}{2}\sum_{i=1}^{m} \langle x + y\,|\,(e_i + Ae_i)\rangle^2_{T_{\mathbb{1}}G} = \frac{1}{2}\sum_{i=1}^{m}(\langle x\,|\,e_i\rangle + \langle y\,|\,Ae_i\rangle)^2. \tag{7.50}$$

Here we have used the the fact that $x, e_i \in \mathbf{k}$ and $y, Ae_i \in \mathbf{z}$ and we have used the orthogonality of $\mathbf{k}$ and $\mathbf{z}$ with respect to $\langle\cdot\,|\,\cdot\rangle$. Now $\langle y\,|\,Ae_i\rangle = \langle A^*y\,|\,e_i\rangle = \langle A^*y\,|\,e_i\rangle_{\mathbf{k}}$, where $A^*$ is the adjoint of $A$. Hence

$$\mathcal{H}(x,y) = \frac{1}{2}\sum_{i=1}^{m}(\langle x\,|\,e_i\rangle + \langle A^*y\,|\,e_i\rangle_{\mathbf{k}})^2 = \frac{1}{2}\|x + A^*y\|^2_{\mathbf{k}}. \tag{7.51}$$

The vertical part of the Hamiltonian equations are (cf. the second equation of (7.40) with $M$ replaced by x+y)

$$\dot{x} + \dot{y} = [x + y, d\mathcal{H}]. \tag{7.52}$$

The let us compute

$$d\mathcal{H} = \underbrace{x + A^*y}_{\in\mathbf{k}} + \underbrace{Ax + AA^*y}_{\in\mathbf{z}}$$

Now since $\mathbf{z}$ is in the center, the second part of $d\mathcal{H}$ disappear in the commutator in (7.52) and we get

$$\dot{x} + \dot{y} = [x + y, x + A^*y] = [x, A^*y],$$

from which we deduce

$$\begin{aligned}\dot{x} &= [x, A^*y],\\ \dot{y} &= 0.\end{aligned}$$

Hence all $y$ components are constant of the motion and we have

$$\begin{aligned} y(t) &= y_0\\ \dot{x} &= [x, A^*y_0] = -[A^*y_0, x] = -(\mathrm{ad}\,(A^*y_0))x\end{aligned}$$

The solution of the last equation is

$$x(t) = e^{-t\mathrm{ad}\,(A^*y_0)}x_0. \tag{7.53}$$

For the horizontal part of the Hamiltonian equations we have

$$\dot{g}(t) = L_{g(t)*}d\mathcal{H}(x(t), y(t)) = L_{g(t)*}(\underbrace{x(t) + A^*y_0}_{\in\mathbf{k}} + \underbrace{Ax(t) + AA^*y_0}_{\in\mathbf{z}}). \tag{7.54}$$

Using the fact that $G = G_{\mathbf{k}} \times G_{\mathbf{z}}$, it is convenient to write an element of $G$ as $g = (g_1, g_2)$ where $g_1 \in G_{\mathbf{k}}$ and $g_2 \in G_{\mathbf{z}}$. Then equation (7.54) splits in the following way

$$\dot{g}_1 = L_{g_1*}(x(t) + A^*y_0) \tag{7.55}$$

$$\dot{g}_2 = Ax(t) + AA^*y_0 \tag{7.56}$$

In the second equation we have used the fact that $L_{g_2*}(Ax(t) + AA^*y_0) = Ax(t) + AA^*y_0$, since we are in an Abelian group. Moreover if $g(0) = (g_{01}, g_{02})$, then for (7.55) and (7.55) we have the initial conditions $g_1(0) = g_{01}$ and $g_2(0) = g_{02}$.

Let us solve (7.55). Using (7.53) this equation is reduced to

$$\dot{g}_1 = L_{g_1*}(e^{-t\,\mathrm{ad}\,(A^*y_0)}x_0 + A^*y_0) = L_{g_1*}e^{-t\,\mathrm{ad}\,(A^*y_0)}(x_0 + A^*y_0), \tag{7.57}$$

where in the last formula we have used the fact that $e^{-t\,\mathrm{ad}\,(A^*y_0)}A^*y_0 = A^*y_0$. Using the variation formula (cf. (6.42)),

$$e^{t(Y+X)} = \overrightarrow{\exp}\int_0^t e^{s\,\mathrm{ad}\,Y}X ds \circ e^{tY}, \tag{7.58}$$

with $Y \to -A^*y_0$ and $X \to x_0 + A^*y_0$, we get

$$g_1(t) = g_{01}e^{t\,x_0}e^{t\,A^*y_0}. \tag{7.59}$$

For (7.56), using (7.53) and using the fact that $G_{\mathbf{z}}$ is Abelian, we have

$$g_2(t) = g_{02} + \int_0^t (Ax(s) + AA^*y_0)\; ds = g_{02} + \int_0^t \left(Ae^{-s\mathrm{ad}\,(A^*y_0)}x_0 + AA^*y_0\right)\; ds. \tag{7.60}$$

The parameterization by arclength is obtained requiring $H = \frac{1}{2}$. From (7.51) we obtain that the normal Pontryagin extremals are parametrized by arclength when $\langle x_0 + A^*y_0 \,|\, x_0 + A^*y_0\rangle = \|x_0 + A^*y_0\|^2 = 1$.

The controls corresponding to the normal Pontryagin extremals $(g_1(t), g_2(t))$ are (cf. Formula 7.50):

$$u_i(t) = \langle x(t) + y_0 \,|\, e_i + Ae_i\rangle = \langle x(t) \,|\, e_i\rangle + \langle y_0 \,|\, Ae_i\rangle \tag{7.61}$$

$$= \langle x(t) + A^*y_0 \,|\, e_i\rangle = \left\langle e^{-t\mathrm{ad}\,(A^*y_0)}x_0 + A^*y_0 \,\middle|\, e_i\right\rangle. \tag{7.62}$$

**Exercise 7.73.** Study abnormal extremals for this problem.

## 7.8 Rolling spheres

### 7.8.1 Rolling with twisting

Consider a sphere of radius 1 in $\mathbb{R}^3$ rolling on a plane without slipping. At every time the state of the system is described by a point on the plane (the projection of its center) and the orientation of the sphere.

We represent a point on the plane as $z = (z_1, z_2) \in \mathbb{R}^2$ and the orientation of the sphere by a point $X \in SO(3)$ representing the orientation of an orthonormal frame attached to the sphere with respect to the standard orthonormal frame in $\mathbb{R}^3$.

Let $\{e_1, e_2, e_3\}$ be the following basis of the Lie algebra $\mathfrak{so}(3)$ of $SO(3)$,

$$e_1 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}, \quad e_2 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix}, \quad e_3 = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \tag{7.63}$$

The condition that the sphere is rolling without slipping can be expressed by saying that the only admissible trajectories in $SO(3) \times \mathbb{R}^2$ are the horizontal trajectories of the following control system (here $u_i(\cdot) \in L^\infty([0,T],\mathbb{R})$, for $i = 1,2,3$).

$$\left\{ \begin{array}{l} \dot{z}_1 = u_1(t) \\ \dot{z}_2 = u_2(t) \\ \dot{X} = X(u_2(t)e_1 - u_1(t)e_2 + u_3(t)e_3). \end{array} \right. \tag{7.64}$$

The controls $u_1(\cdot)$ and $u_2(\cdot)$ correspond to the two rotations of the sphere that produce a movement in the plane, while the control $u_3(\cdot)$ correponds to a twist of the sphere (that produces no movement in the plane). See Figure 7.4. We would like to solve the following problem.

**P:** Given an initial and final position in $SO(3) \times \mathbb{R}^2$, roll the sphere on the plane in such a way that the initial and final conditions are the given ones and $\int_0^T \sqrt{\sum_{i=1}^3 u_i(t)^2}\, dt$ is minimal.

We have the following result.

**Proposition 7.74.** *The projection on the plane $(z_1, z_2)$ of normal Pontryagin extremals is (up to time reparameterization) the set of sinusoids on the plane:*

$$\left\{ \begin{pmatrix} z_{01} \\ z_{02} \end{pmatrix} + \begin{pmatrix} \cos(a_0) & -\sin(a_0) \\ \sin(a_0) & \cos(a_0) \end{pmatrix} \begin{pmatrix} f(\phi_0, b, r, t) \\ t \end{pmatrix} \mid a_0, \phi_0 \in S^1, \quad b, r \geq 0, \quad z_{01}, z_{02} \in \mathbb{R} \right\},$$

*where*

$$f(\phi_0, b, r, t) = \left\{ \begin{array}{ll} b \sin(rt + \phi_0) & \text{if } r > 0 \\ b\,t & \text{if } r = 0. \end{array} \right.$$

To prove Proposition 7.74 we first prove that the problem define a $\mathbf{k} \oplus \mathbf{z}$ sub-Riemannian structure and then we study its normal Pontryagin extremals.

**Claim.** The problem above is a problem of type $\mathbf{k} \oplus \mathbf{z}$.

To prove the claim let us set $G = SO(3) \times \mathbb{R}^2$. We have $T_{\mathbb{1}}G = \mathfrak{so}(3) \oplus \mathbb{R}^2$. Now let $f_1 = (1,0)^T$ and $f_2 = (0,1)^T$ be the generators of $\mathbb{R}^2$ and define

$$\mathbf{d} = \operatorname{span}\{f_1 - e_2, f_2 + e_1, e_3\} \subset \mathfrak{so}(3) \times \mathbb{R}^2.$$

Given a vector $v = u_1(f_1 - e_2) + u_2(f_2 + e_1) + u_3 e_3 \in \mathbf{d}$ we define its norm as $\|v\| = \sqrt{u_1^2 + u_2^2 + u_3^2}$. If $\pi : \mathfrak{so}(3) \times \mathbb{R}^2 \to \mathfrak{so}(3)$ is the canonical projection, this norm coincides with the norm of $\|\pi(v)\|_{\mathfrak{so}(3)}$, where $\|\cdot\|_{\mathfrak{so}(3)}$ is the standard norm for which $\{e_1, e_2, e_3\}$ is an orthonormal frame. This norm comes from a bi-invariant metric as explained in Section 7.7.2.

The corresponding sub-Riemannian problem is then

$$\dot{g} = g\big(u_1(t)(f_1 - e_2) + u_2(t)(f_2 + e_1) + u_3 e_3\big), \tag{7.65}$$

$$g(0) = g_0, \quad g(T) = g_1, \tag{7.66}$$

$$\int_0^T \sqrt{\sum_{i=1}^3 u_i(t)^2\, dt} \;\to \min, \tag{7.67}$$

where $g_0, g_1 \in SO(3) \times \mathbb{R}^2$. Writing elements in $SO(3) \times \mathbb{R}^2$ as pairs $g = (X, z)$, this problem become exactly (7.64).

If we define the linear application $A : \mathfrak{so}(3) \to \mathbb{R}^2$ via

$$Ae_1 = f_2, \quad Ae_2 = -f_1, \quad Ae_3 = 0,$$

we can write

$$\mathbf{d} = \{x + Ax \mid x \in \mathfrak{so}(3)\}.$$

*Remark* 7.75. Notice that if we write an element of $\mathfrak{so}(3)$ as $x_1e_1 + x_2e_2 + x_3e_3$ and an element of $\mathbb{R}^2$ as $y_1f_1 + y_2f_2$, we can think to $A$ and to its adjoint $A^*$ as to the rectangular matrices

$$A = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \quad A^* = \begin{pmatrix} 0 & 1 \\ -1 & 0 \\ 0 & 0 \end{pmatrix}.$$

Notice that $AA^* = \mathbb{1}_{2\times 2}$ while $A^*A \neq \mathbb{1}_{3\times 3}$. From the expression of $A^*$ we also get

$$A^*f_1 = -e_2, \quad A^*f_2 = e_1. \tag{7.68}$$

The problem $\mathbf{P}$ is then a $\mathbf{k}\oplus\mathbf{z}$ problem with $\mathbf{k} = \mathfrak{so}(3)$, $\mathbf{z} = \mathbb{R}^2$. Moreover $\mathbf{d}$, $A$ and the bi-invariant metric on $\mathbf{k}$, are defined as above.

### Normal Pontryagin extremals

Normal Pontryagin extremals are parametrized by arclength if we take $x_0 \in \mathfrak{so}(3)$ and $y_0 \in \mathbb{R}^2$ satisfying

$$\|x_0 + A^*y_0\| = 1. \tag{7.69}$$

Now writing $y_0 = y_{01}f_1 + y_{02}f_2$ and using (7.68) we have

$$A^*y_0 = A^*(y_{01}f_1 + y_{02}f_2) = \begin{pmatrix} 0 & 0 & -y_{01} \\ 0 & 0 & -y_{02} \\ y_{01} & y_{02} & 0 \end{pmatrix}.$$

Hence writing $x_0 = x_{01}e_1 + x_{02}e_2 + x_{03}e_3$, equation (7.69) become

$$\|(x_{01} + y_{02})e_1 + (x_{02} - y_{01})e_2 + x_{03}e_3\| = 1.$$

It is then convenient to parametrize normal Pontryagin extremals with

$$y_{01} \in \mathbb{R}, \quad y_{02} \in \mathbb{R}, \quad \theta \in [0, \pi], \quad \varphi \in [0, 2\pi], \tag{7.70}$$

taking

$$\begin{aligned} x_{01} &= -y_{02} + \cos(\theta)\cos(\varphi) \\ x_{02} &= y_{01} + \cos(\theta)\sin(\varphi) \\ x_{03} &= \sin(\theta) \end{aligned}$$

The **z** part of the normal Pontryagin extremal is given by the formula (7.60), with $g_2 \to (z_1, z_2)^T$, i.e.,

$$\begin{aligned}\begin{pmatrix} z_1(t) \\ z_2(t) \end{pmatrix} &= \begin{pmatrix} z_{01} \\ z_{02} \end{pmatrix} + \int_0^t \left( Ae^{-s\mathrm{ad}\,(A^* y_0)} x_0 + AA^* y_0 \right) ds. \\ &= \begin{pmatrix} z_{01} \\ z_{02} \end{pmatrix} + \int_0^t \left( Ae^{-s(A^* y_0)} x_0 e^{s(A^* y_0)} + \begin{pmatrix} y_{01} \\ y_{02} \end{pmatrix} \right) ds. \end{aligned} \tag{7.71}$$

If we fix $y_{01} = y_{02} = 0$, we get

$$\begin{aligned} z_1(t) &= z_{01} - t\cos(\theta)\sin(\varphi), \\ z_2(t) &= z_{02} + t\cos(\theta)\cos(\varphi). \end{aligned}$$

Otherwise if we set $y_{01} = r\cos(a)$ and $y_{02} = r\sin(a)$, we obtain for $r \neq 0$,

$$\begin{aligned} z_1(t) = z_{01} - &\frac{1}{r}(rt\cos^2(a)\cos(\theta)\sin(\varphi) + \sin(a)\cos(a)\cos(\theta)\cos(\varphi)(\sin(rt) - rt) + \\ &\sin(a)(\sin(a)\cos(\theta)\sin(\varphi)\sin(rt) + \sin(\theta) + \sin(\theta)(-\cos(rt)))), \\ z_2(t) = z_{02} + &\frac{1}{r}(\cos(\theta)\left(\cos(\varphi)\left(rt\sin^2(a) + \cos^2(a)\sin(rt)\right) + \sin(a)\cos(a)\sin(\varphi)(\sin(rt) - rt)\right) - \\ &\cos(a)\sin(\theta)(\cos(rt) - 1). \end{aligned}$$

that is a combination of sinus and cosinus. See Figure 7.5.

**Exercise 7.76.** Prove that each trajectory $(z_1(t), z_2(t))$ is a rototranslation of a sinusoid and that $\varphi$ determines its initial direction, $r$ its frequence, $\theta$ its amplitude and $a$ its rotation on the plane.

The **k** part of the normal Pontryagin extremal can be obtained with the formula

$$X(t) = e^{t\,x_0} e^{t\,A^* y_0}.$$

### 7.8.2 Rolling without twisting

We now consider a sphere rolling on a plane without slipping and without twisting. Similarly to what done in Section 7.8, the state space is the group $G = SO(3) \times \mathbb{R}^2$ whose Lie algebra is $T_{\mathbb{1}}G = \mathfrak{so}(3) \times \mathbb{R}^2$ and the distribution is still defined by equation (7.65) with the difference that now we have $u_3 \equiv 0$.

More precisely, the condition that the sphere is rolling without slipping and twisting can be expressed by saying that the only admissible trajectories in $SO(3)\times\mathbb{R}^2$ are the horizontal trajectories of the following control system

$$\dot{g} = g\big(u_1(t)(f_1 - e_2) + u_2(t)(f_2 + e_1)\big). \tag{7.72}$$

Here $f_1, f_2$ are the generators of $\mathbb{R}^2$ and $e_1, e_2, e_3$ are given by (7.63). The controls $u_1(\cdot)$ and $u_2(\cdot)$ belonging to $L^\infty([0,T], \mathbb{R})$ correspond to the rotations of the sphere along the $z_1$ and $z_2$ axis.

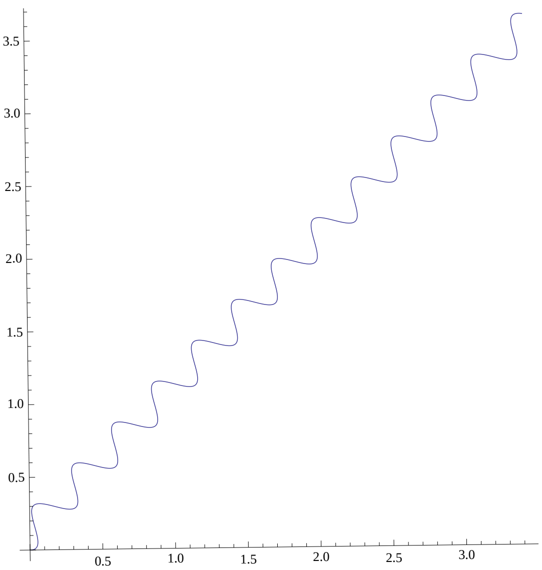



Figure 7.5: A Pontryagin extremals for the rolling sphere with twist

The commutators between $f_1, f_2, e_1, e_2, e_3$ are

$$\begin{aligned} [f_1, f_2] &= 0 \\ [f_i, e_j] &= 0, \quad i = 1, 2, \quad j = 1, 2, 3, \\ [e_1, e_2] = e_3, \quad [e_2, e_3] &= e_1, \quad [e_3, e_1] = e_2. \end{aligned} \tag{7.73}$$

We would like to solve the following problem.

**P:** Given an initial and final position in $SO(3) \times \mathbb{R}^2$, roll the sphere on the plane in such a way that the initial and final conditions are the given ones and $\int_0^T \sqrt{\sum_{i=1}^2 u_i(t)^2}\, dt$ is minimal.

*Remark* 7.77. Notice that solving problem **P** corresponds to find the shortest path on the plane such that the sphere rolling along that path goes from the prescribed initial condition to the prescribed final condition. See Figure (7.6).

Contrarily to what happens to the problem of rolling a sphere with twisting (Section 7.8.1), this time the problem is not of the form $\mathbf{k} \oplus \mathbf{z}$. Indeed the distribution is two dimensional and it is not projecting well on the compact sub-algebra $\mathfrak{so}(3)$. We are going to use the general equations.

Normal extremals are solutions of the Hamiltonian system associated with the following Hamiltonian

$$H(p, g) = \frac{1}{2} \left( \langle p, L_{g*}(f_1 - e_2) \rangle^2 + \langle p, L_{g*}(f_2 + e_1) \rangle^2 \right).$$

The trivialized Hamiltonian is

$$\mathcal{H}(\xi) = \frac{1}{2} \left( \langle \xi, (f_1 - e_2) \rangle^2 + \langle \xi, (f_2 + e_1) \rangle^2 \right), \quad \xi \in T_{\mathbb{1}}^* G.$$

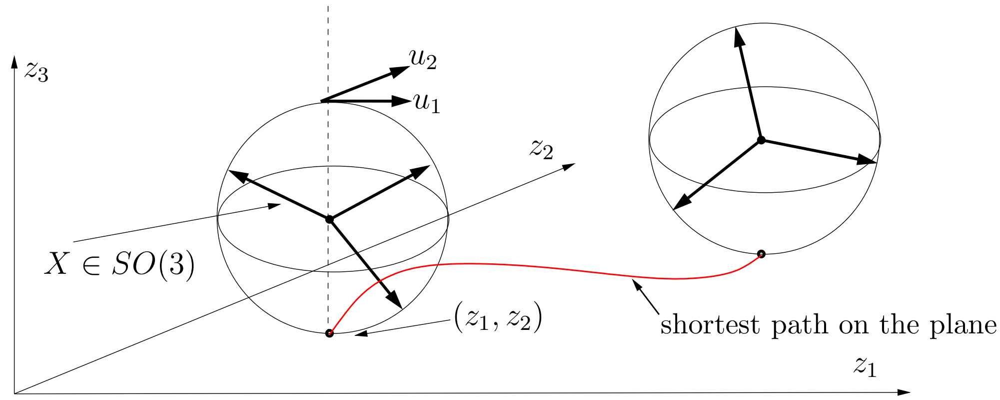



Figure 7.6: The sub-Riemannian problem of rolling a sphere without slipping and twisting.

It is convenient to use the following coordinates,

$$h_{f_1} = \langle \xi, f_i \rangle, \quad i = 1, 2, \quad h_{e_j} = \langle \xi, e_j \rangle, \quad j = 1, 2, 3.$$

Notice that, using (7.73) we have

$$\{h_{f_1}, h_{f_2}\} = \langle \xi, [f_1, f_2] \rangle = 0,$$
$$\{h_{f_i}, h_{e_j}\} = \langle \xi, [f_i, e_j] \rangle = 0, \quad i = 1, 2, \quad j = 1, 2, 3,$$
$$\{h_{e_1}, h_{e_2}\} = \langle \xi, [e_1, e_2] \rangle = \langle \xi, e_3 \rangle = h_{e_3}, \quad \{h_{e_2}, h_{e_3}\} = h_{e_1}, \quad \{h_{e_3}, h_{e_1}\} = h_{e_2}.$$

Then

$$\mathcal{H} = \frac{1}{2}\left((h_{f_1} - h_{e_2})^2 + (h_{f_2} + h_{e_1})^2\right).$$

The Hamiltonian equations are

$$\dot{h}_{f_i} = \{\mathcal{H}, h_{f_i}\}, \quad i = 1, 2, \quad \dot{h}_{e_j} = \{\mathcal{H}, h_{e_j}\}, \quad j = 1, 2, 3. \tag{7.74}$$

Let us start with the first one

$$\dot{h}_{f_1} = \{\mathcal{H}, h_{f_1}\} = \sum_{i=1}^{2} \frac{\partial \mathcal{H}}{\partial h_{f_i}} \{h_{f_i}, h_{f_1}\} + \sum_{i=1}^{3} \frac{\partial \mathcal{H}}{\partial h_{e_i}} \{h_{e_i}, h_{f_1}\} = 0,$$

where we have used that $h_{f_1}$ commutes (for the Poisson brackets) with everything. Similarly

$$\dot{h}_{f_2} = 0,$$
$$\dot{h}_{e_1} = (h_{f_1} - h_{e_2})h_{e_3},$$
$$\dot{h}_{e_2} = (h_{f_2} + h_{e_1})h_{e_3},$$
$$\dot{h}_{e_3} = -h_{f_1}h_{e_1} - h_{f_2}h_{e_2}.$$

Now if we consider normal Pontryagin extremals parametrized by arc length, i.e., if we work on the level set $\{\mathcal{H}=1/2\}\simeq S^1\times\mathbb{R}^3$, it is convenient to use the coordinates $r,\alpha,\theta,c$ defined by

$$\begin{aligned} h_{f_1} &= r\cos(\alpha)\\ h_{f_2} &= r\sin(\alpha)\\ h_{f_1}-h_{e_2} &= \cos(\theta+\alpha),\\ h_{f_2}+h_{e_1} &= \sin(\theta+\alpha),\\ h_{e_3} &= c. \end{aligned}$$

Normal Pontryagin extremals starting from a given initial condition, are parametrized by points in $\{\mathcal{H}=1/2\}$, i.e., by $\theta_0\in S^1$, $c_0\in\mathbb{R}$ and $(r_0,\alpha_0)$ parametrizing $\mathbb{R}^2$ in polar coordinates $(r_0\geq 0,\alpha\in S^1)$.

The Hamiltonian equations are then

$$\dot r = 0 \quad\Rightarrow\quad r=r_0, \tag{7.75}$$
$$\dot\alpha = 0 \quad\Rightarrow\quad \alpha=\alpha_0, \tag{7.76}$$
$$\dot\theta = c, \tag{7.77}$$
$$\dot c = -r_0\sin(\theta). \tag{7.78}$$

Once that equations (7.77) and (7.78) are solved in function of the initial conditions $(r_0,\theta_0,c_0)$, i.e., once that one gets $\theta(t;r_0,\theta_0,c_0)$, the controls are given by

$$\begin{aligned} u_1(t;r_0,\theta_0,c_0,\alpha_0) &= \langle\xi,f_1-e_2\rangle = h_{f_1}-h_{e_2} = \cos(\theta(t;r_0,\theta_0,c_0)+\alpha_0)\\ u_2(t;r_0,\theta_0,c_0,\alpha_0) &= \langle\xi,f_2+e_1\rangle = h_{f_2}+h_{e_1} = \sin(\theta(t;r_0,\theta_0,c_0)+\alpha_0). \end{aligned} \tag{7.79}$$

Once $u_1(\cdot)$ and $u_2(\cdot)$ are known, one can compute the corresponding trajectory by integrating (7.72). However here we are only interesting to the planar part of the normal Pontryagin extremals starting from $z_{01}$ and $z_{02}$, that is given by

$$z_1(t;\theta_0,c_0,\alpha_0) = z_{01}+\int_0^t u_1(s)ds = z_{01}+\int_0^t \cos(\theta(s;\theta_0,c_0)+\alpha_0)ds, \tag{7.80}$$
$$z_2(t;\theta_0,c_0,\alpha_0) = z_{02}+\int_0^t u_2(s)ds = z_{02}+\int_0^t \sin(\theta(s;\theta_0,c_0)+\alpha_0)ds. \tag{7.81}$$

In the following we refer to $(z_1(\cdot),z_2(\cdot))$ as the $z$-geodesics.

**Qualitative analysis of the trajectoris**

Equations (7.77) and (7.78) are the equation of a planar pendulum of mass 1, length 1, where $r_0$ represent the gravity. These equations admits an explicit solution in terms of elliptic functions. However their qualitative behaviour can be understood easily.

First notice that if we consider only $z$-geodesics starting from the origin and with $z_1'(0)=1$ and $z_2'(0)=0$, we can fix $z_{01}=z_{02}=0$, $\alpha_0=-\theta_0$. All other $z$-geodesics can be obtained by rototranslations of these ones.

Equation (7.77) and (7.78) admit a first integral that up to a constant is the energy of the pendulum:

$$H_{\mathrm{p}} = \frac{1}{2}c^2 - r_0\cos(\theta).$$

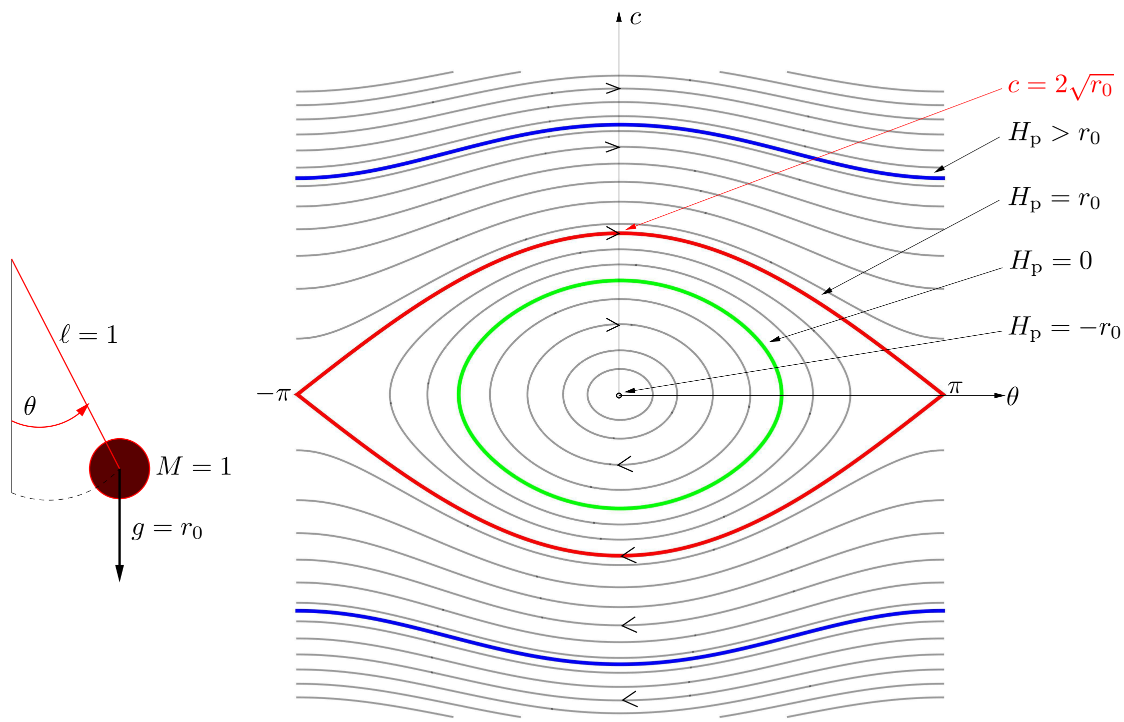



Figure 7.7: Level set of the pendulum for $r_0 \neq 0$. The vertical line $\theta = \pi$ is identified with the veritical line $\theta = -\pi$. We have also indicated the direction of parameterization that one gets from the equation $\dot{\theta} = c$. Notice that the only critical points are $(\theta, c) = (0, 0)$ (stable equilibrium) and $(\theta, c) = (\pi, 0)$ (unstable equilibrium).

Fixed $(r_0, c_0)$, one compute $H_{\mathrm{p}}$ and the corresponding trajectory in the $(\theta, c)$ plane should stay on this set.

Now let us compute the curvature of the $z$-geodesics. We have

$$K = \frac{z_1' z_2'' - z_2' z_1''}{((z_1')^2 + (z_2')^2)^{3/2}} = \theta'(t; r_0, \theta_0, c_0) = c(t; r_0, \theta_0, c_0).$$

Hence $c$ is precisely the curvature of the $z$-geodesic. Inflection points of $z$-geodesics corresponds to times in which $c$ changes sign.

**The case** $r_0 = 0$. In this case $\dot{c} = 0$ and $\theta(t) = \theta_0 + c_0 t$. The $z$-geodesic is a circle (if $c_0 \neq 0$) or a straight line (if $c_0 = 0$).

**The case** $r_0 > 0$. The level sets of $H_{\mathrm{p}}$ are shown in Figure 7.7. There are several types of trajectories:

- $H_{\mathrm{p}} > r_0$. In this case the pendulum is rotating and $\theta(\cdot)$ is monotonic increasing (no inflection points).
- $H_{\mathrm{p}} = r_0$. We have two cases:

– If $\theta_0 \neq \pm\pi$. The pendulum is on the separatrix. The $z$-geodesic has an inflection point at infinity.

– If $\theta_0 = \pm\pi$. The pendulum stays at the unstable equilibrium $(\theta, c) = (\pm\pi, 0)$. The $z$-geodesic is a straight line.

- $H_\mathrm{p} \in (-r_0, r_0)$. In this case the pendulum is oscillating and $\theta(\cdot)$ too. The $z$-geodesic present inflection points. Such $z$-geodesics are called "inflectional".

- $H_\mathrm{p} = -r_0$. The pendulum stays at the stable equilibrium $(\theta, c) = (0, 0)$. The $z$-geodesic is a straight line.

See Figure 7.8.

Evaluating when these normal Pontryagin extremals lose optimality is not an easy problem and it is outside the purpose of this book. See the bibliographical note.

**Exercise 7.78.** Find all abnormal extremals for this problem.

### 7.8.3 Euler's "cvrvae elasticae"

The $z$-geodesics for the rolling ball withouting twisting are called Euler's *cvrvae elasticae*, since they are obtained via (7.80) and (7.81) from the solution of equations (7.75), (7.76), (7.77), (7.78), that are the same equation that one gets while looking for the configurations of an elastic rod on the plane having a stationary point of elastic energy. See [Eul].

For convenience we re-write the equations here:

$$\dot{z}_1 = \cos(\theta + \alpha_0) \tag{7.82}$$
$$\dot{z}_2 = \sin(\theta + \alpha_0) \tag{7.83}$$
$$\dot{\theta} = c \tag{7.84}$$
$$\dot{c} = -r_0 \sin(\theta) \tag{7.85}$$

These equations contains several parameters: $r_0 > 0$, $\alpha_0$, and the initial conditions $\theta(0) = \theta_0$, $c(0) = c_0$, $z_1(0) = z_{01}$, $z_2(0) = z_{02}$, having the following meaning:

- $(z_{01}, z_{02})$ is the starting point of the curba elastica;
- $\theta_0 + \alpha_0$ is the starting angle of the curba elastica;
- $\theta_0$ gives the "starting point" of the solution of the pendulum that it is used in the interval $[0, T]$;
- $r_0$ and $c_0$ establish the gravity of the pendulum and the level of the Hamiltonian $H_\mathrm{p}$. This has consequences on the type of curba elastica (inflection, non inflectional etc,...) and on their "size" on the plane.

We have the following interesting characterization of cvrvae elasticae.

**Proposition 7.79.** *The set of cvrvae elasticae coincides with the set of planar curves parametrized by planar arclength for which the curvature is an affine function of the coordinates.*

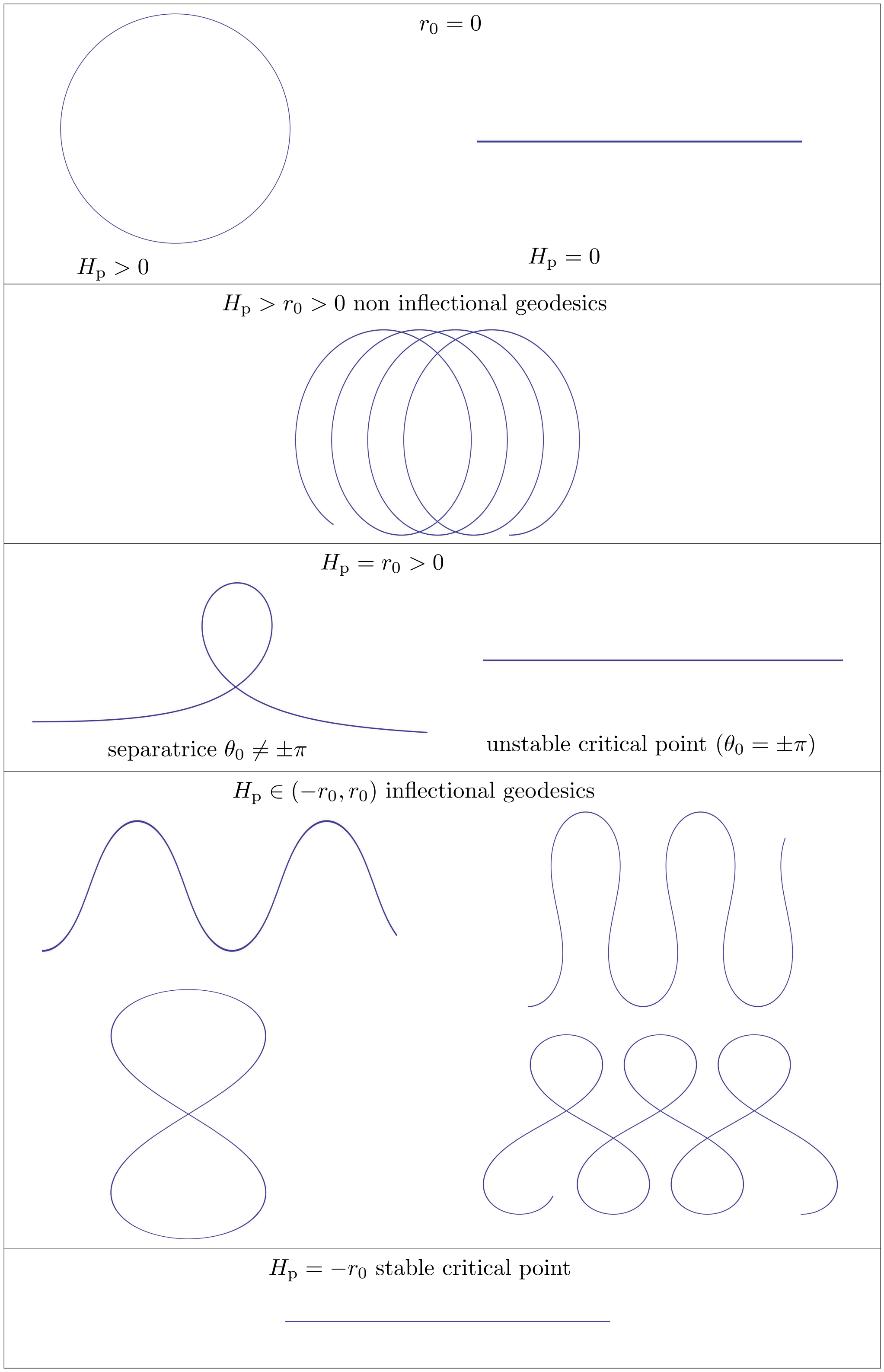



Figure 7.8: A picture of the $z$-geodesics. Notice the presence of a periodic trajectory.

*Proof.* Let us make the following change of coordinates $z_1, z_2 \to x_1, x_2$ where

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} \cos(\alpha_0) & \sin(\alpha_0) \\ -\sin(\alpha_0) & \cos(\alpha_0) \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \end{pmatrix}.$$

Then equations (7.82)–(7.85) become

$$\begin{aligned} \dot{x}_1 &= \cos(\theta), \\ \dot{x}_2 &= \sin(\theta), \\ \dot{\theta} &= c, \\ \dot{c} &= -r_0 \sin(\theta). \end{aligned}$$

Hence

$$\dot{c} = -r_0 \sin(\theta) = -r_0 \dot{x}_2.$$

Integrating we obtain

$$c(t) - c_0 = -r_0(x_2(t) - x_2(0)).$$

Hence

$$c(t) = c_0 - r_0(-\sin(\alpha_0)z_1 + \cos(\alpha_0)z_2) + r_0(-\sin(\alpha_0)z_{01} + \cos(\alpha_0)z_{02}) = a_0 + a_1 z_1 + a_2 z_2.$$

where

$$a_0 = c_0 + r_0(-\sin(\alpha_0)z_{01} + \cos(\alpha_0)z_{02}), \quad a_1 = r_0 \sin(\alpha_0), \quad a_2 = -r_0 \cos(\alpha_0).$$

One immediately verify that the Jacobian of the transformation $c_0, r_0, \alpha_0 \to a_0, a_1, a_2$ is equal to $r_0$. However this singularity is only due to the choice of polar coordinates. □

**Exercise 7.80.** Consider the Engel sub-Riemannian problem, i.e., the sub-Riemannian structure on $\mathbb{R}^4$ for which an orthonormal frame is given by the vector fields

$$X_1 = \partial_{x_1}, \quad X_2 = \partial_{x_2} - x_1 \partial_{x_3} + \frac{x_1^2}{2} \partial_{x_4}.$$

Prove that the Lie algebra generated by $X_1$ and $X_2$ is finite dimensional. Using Theorem 7.1 deduce that this problem define a sub-Riemannian structure on a Lie group. Find the group law. Study its normal Pontryagin extremals. Do the same for the Cartan sub-Riemannian problem, i.e., the sub-Riemannian structure on $\mathbb{R}^5$ for which an orthonormal frame is given by the vector fields

$$X_1 = \partial_{x_1}, \quad X_2 = \partial_{x_2} - x_1 \partial_{x_3} + \frac{x_1^2}{2} \partial_{x_4} + x_1 x_2 \partial_{x_5}.$$

### 7.8.4 Rolling spheres: further comments

A regular curve in the Euclidean plane is an elastica if and only if its curvature is an affine function of the coordinates. In other words, a plane curve is an elastica if and only if it is a normal Pontryagin extremal of a plane isoperimetric problem with an affine "magnetic field" (see Section 4.4.2).

One can realize that the rolling without slipping or twisting problem looks somehow similar to the isoperimetric one. The state space is $\mathbb{R} \times \mathbb{R}^2$ for the isoperimetric problem and is $SO(3) \times \mathbb{R}^2$ for the rolling problem. The horizontal distribution is a complement to the tangent space to $\mathbb{R} \times \cdot$

and is invariant under translations of the additive group $\mathbb{R}$ for the isoperimetric problem; it is a compliment to the tangent space to $SO(3) \times \cdot$ and is invariant under (left) translations of the group $SO(3)$. The sub-Riemannian length is induced by the Riemannian length in $\mathbb{R}^2$ for both problems. The general framework that contains both problems as well as the problems discussed in Section 7.7.4 is as follows.

Let $G$ be a Lie group. A *principal bundle with a structure group* $G$ is a locally trivial bundle $N \xrightarrow{G} M$ where the group $G$ acts freely on $N$ and the orbits of this action are exactly the fibers of the bundle. The typical example is the bundle of orthonormal frames on a Riemannian manifold and traditionally a right action of $G$ is considered. In the case of the bundle of oriented orthonormal frames on an $n$-dimensional Riemannian manifold the structure group is $SO(n)$; if $(v_1, \dots, v_n)$ is a frame and $A = \{a_{ij}\}_{i,j=1}^n \in SO(n)$, then the action is defined as

$$(v_1, \dots, v_n) \cdot A = \left( \sum_{i=1}^n a_{i1} v_i, \dots, \sum_{i=1}^n a_{in} v_i \right).$$

Let $\mathfrak{g}$ be the Lie algebra of the group $G$. A *connection on the principal bundle* $N \xrightarrow{G} M$ is a vector distribution on $N$ that is a complement to the tangent spaces to the fibers and is invariant under the action of $G$. Recall that right translations of the Lie group are generated by left-invariant vector fields; hence the tangent space to the fiber at any point is naturally identified with $\mathfrak{g}$. Let $D_q \subset T_q N,\ q \in N$ be a connection. We have $T_q N = \mathfrak{g} \oplus D_q$; a linear projection $\omega_q : T_q N \to \mathfrak{g}$ such that $\ker \omega_q = D_q$ defines a non-degenerate $G$-invariant $\mathfrak{g}$-valued vector differential form $\omega$ on $N$.

Of course, the construction can be inverted. According to another equivalent definition, a connection on the principal bundle is a non-degenerate $G$-invariant $\mathfrak{g}$-valued differential form. The kernel of such a form is the connection in the sense of the first definition.

Let $\pi : N \xrightarrow{G} M$ be the canonical projection to the base of the bundle and $\gamma : [0,1] \to M$ be a smooth curve. Given a point $q_0 \in \pi^{-1}(\gamma(0))$ there exists a unique horizontal lift $q_t$ of $\gamma(t)$ starting at $q_0$, i.e., $\dot{q}_t \in D_{q_t},\ 0 \le t \le 1$. The point $q_1 \in \pi^{-1}(\gamma(1))$ is called the parallel transport of $q_0$ along $\gamma$. The parallel transport commutes with the action of $G$; thus the transport of a point determines the transport of the whole fiber.

Assume that $M$ is equipped with a Riemannian structure. The length-minimization problem on the set of curves in $M$ that provide a parallel transport from $q_0$ to the given point $q_1$ is a *isoholonomic problem*. The two-dimensional isoperimetric problems, their modification considered in Section 7.7.4, and the rolling without slipping or twisting problem are just very special cases. Isoholonomic problems link sub-Riemannian geometry with numerous applications: dynamics of a particle in a gauge field, optimal shape transformation, and many others.

## 7.9 Bibliographical note

A basic result in Lie groups theory states that, given a Lie group $G$ with Lie algebra $\mathfrak{g}$, there is a one-to-one correspondence between subalgebras of $\mathfrak{g}$ and subgroups of $G$. Theorem 7.1 together with Proposition 7.18 is a generalization of the above result for finite-dimensional subgroups of the group of diffeomorphisms (which is not a Lie group). Such a result is certainly known, but we could not find a reference for that. The proof given here does not use functional analysis tools and it is original.

The geodesics for the Heisenberg group were computed first in [Gav77]. The picture of their end-points at time 1 were shown first in [Bro82] (optimal geodesics) and in [Nac82, GV88] (all geodesics). The geodesics for free Carnot groups of step 2 were computed first in [Gav77, Bro82].

The double exponential formula (7.46) for geodesics in the $\mathbf{d} \oplus \mathbf{s}$ problem was found independently by Agrachev [Agr95] and Brockett [Bro99] (this last paper is based on previous works [Bro73]). It was then intensively studied in [Jur99, Jur01, Jur16, BCG02a, BR08] and in [Mon02] (see p. 200). A nice geometrical interpretation of these formulas was given in [BZ15b, BZ15a, BZ16].

The formula for geodesics in the $\mathbf{k} \oplus \mathbf{z}$ case is original, although it was used implicitly in [Bes14].

The problem of rolling with twisting was studied by [Bes14]. The problem of rolling without twisting was formulated in [Ham83] and then studied in [AW86, Jur93]. See also [AS04, Jur97].

Euler's curvae elasicae were introduced by Euler in [Eul]. The local stability was studied first by Max Born in [Bor06]. Local and global optimality, in the context of optimal control was studied in a series of papers by Yuri Sachkov [Sac08b, Sac08a, SS14]. The fact that the set cvrvae elasticae coincides with the set of planar curves parametrized by planar arclength for which the curvature is an affine function of the coordinates was first noticed by Vladimir Zakalyukin (personal communication).

Left-invariant sub-Riemannian structures as defined in this chapter are indeed quite general thanks to the following result proved by Berestovskii [Ber88]: given a geodesic length space $X$ such that its isometry group acts transitively on $X$, then the associated distance is a sub-Finsler one. More precisely the metric space $X$ is an homogeneous space $G/H$ and there is a left-invariant distribution $\mathcal{D}$ and a left-invariant norm on $\mathcal{D}$ such that the distance on the metric space coincides with the sub-Finsler one.

Classical references for sub-Riemannian problems on Lie groups, or, more in general, for optimal control problems on Lie groups are [Jur97, Jur16, Blo15].

# Chapter 8

# End-point map and exponential map

In Chapter 4 we started to study necessary conditions for the length-minimality of a horizontal trajectory. First order conditions split candidates into two classes, namely normal and abnormal extremals. We proved that normal extremal trajectories are indeed geodesics, i.e., short arcs realize the sub-Riemannian distance.

In this chapter we go further and we study second order conditions. To this purpose, we introduce the end-point map $E_{q_0}$ that associates a control $u$ with the final point $E_{q_0}(u)$ of the admissible trajectory associated with $u$ and starting from $q_0$. Next, we consider the problem of minimizing the energy $J$ among horizontal curves joining two fixed points $q_0, q_1 \in M$, written as the constraint variational problem

$$\min J|_{E_{q_0}^{-1}(q_1)}, \qquad q_1 \in M. \tag{8.1}$$

It is then natural to introduce Lagrange multipliers. First order conditions recover Pontryagin extremals, while second order conditions give new information. This viewpoint permits to interpret candidate abnormal length-minimizers as critical points of the map $E_{q_0}$ defining the constraint.

Taking advantage of the invariance by reparametrization, when useful, we assume that all the horizontal trajectories are defined on the same interval $I = [0, 1]$. Also, since the energy of a horizontal curve coincides (up to a normalization factor) with the $L^2$-norm of the corresponding control, it is natural to take $L^2([0, 1], \mathbb{R}^m)$ as class of admissible controls (cf. the discussion in Section 3.6), which has a natural Hilbert space structure.

In this chapter we use some classical results of nonlinear analysis on Hilbert and Banach spaces. We refer the reader to the classical references [Die60] and [Car67].

## 8.1 The end-point map

In this chapter we always assume that the sub-Riemannian structure is free, i.e., $\mathbf{U} = M \times \mathbb{R}^m$ for some $m \in \mathbb{N}$. In the following $\{f_1, \dots, f_m\}$ denotes a global generating family (recall that every sub-Riemannian manifold $(M, \mathbf{U}, f)$ is equivalent to a free one, cf. Section 3.1.4).

Fix $q_0 \in M$. Recall that, for every control $u \in L^2([0, 1], \mathbb{R}^m)$, the corresponding trajectory $\gamma_u$ is the unique maximal solution of the Cauchy problem

$$\dot{\gamma}(t) = \sum_{i=1}^{m} u_i(t) f_i(\gamma(t)), \qquad \gamma(0) = q_0. \tag{8.2}$$

Let $\mathcal{U}_{q_0} \subset L^2([0,1], \mathbb{R}^m)$ the set of controls $u$ such that the corresponding trajectory $\gamma_u$ starting at $q_0$ is defined on the interval $[0,1]$.

**Exercise 8.1.** (i). Prove that $\mathcal{U}_{q_0}$ is an open subset of $L^2([0,1], \mathbb{R}^m)$.

(ii). Let $r_0 > 0$ such that the closed sub-Riemannian ball $\overline{B}_{q_0}(r_0)$ is compact (cf. Corollary 3.37), and denote by $\mathcal{B}_{L^2}(r_0)$ the ball of radius $r_0$ in $L^2$. Prove that $\mathcal{B}_{L^2}(r_0) \subset \mathcal{U}_{q_0}$.

**Definition 8.2.** Let $(M, \mathbf{U}, f)$ be a free sub-Riemannian manifold, and fix $q_0 \in M$. The *end-point map* based at $q_0$ is the map

$$E_{q_0} : \mathcal{U}_{q_0} \to M, \qquad E_{q_0}(u) = \gamma_u(1). \tag{8.3}$$

where $\gamma_u : [0,1] \to M$ is the unique solution of the Cauchy problem (8.2).

*Remark* 8.3. Similarly one can define the *end-point map at time* $t \in \mathbb{R}$ based at $q_0$, denoted by $E^t_{q_0} : \mathcal{U}^t_{q_0} \to M$ and defined by the identity $E^t_{q_0}(u) := \gamma_u(t)$ on the set $\mathcal{U}^t_{q_0}$ of controls $u$ for which the corresponding trajectory $\gamma_u$ is defined on $[0,t]$.

A first property of the end-point map, which is a consequence of the Chow-Raschevskii theorem, is its openness.

**Exercise 8.4.** Prove that $E_{q_0} : \mathcal{U}_{q_0} \to M$ is open at every $u \in \mathcal{U}_{q_0}$. Hint: combine the flow associated with the control $u$ with the map $\psi$ defined in the proof of Chow-Raschevskii theorem.

We now prove that the end-point map smooth in the Fréchet sense, and we compute explicitly its first differential.

**Proposition 8.5.** *The end-point map $E_{q_0}$ is smooth on $\mathcal{U}_{q_0}$. For every $u \in \mathcal{U}_{q_0}$ the differential $D_u E_{q_0} : L^2([0,1], \mathbb{R}^m) \to T_{\gamma_u(1)}M$ satisfies*

$$D_u E_{q_0}(v) = \int_0^1 (P^u_{t,1})_* f_{v(t)}\big|_{\gamma_u(1)} dt, \qquad v \in L^2([0,1], \mathbb{R}^m). \tag{8.4}$$

*Here $P^u_{t,s}$ denotes the flow generated by $u$ and $f_{v(t)} = \sum_{i=1}^m v_i(t) f_i(q)$.*

From the geometric viewpoint, the differential $D_u E_{q_0}(v)$ is the integral mean of the vector field $f_{v(t)}$ defined by $v$ along the trajectory $\gamma_u$, where all the vectors are pushed forward to the same tangent space $T_{\gamma_u(1)}M$ by $P^u_{t,1}$ (see Figure 8.1). It should be noted that, since $\mathcal{U}_{q_0}$ is an open set of the vector space $L^2([0,1], \mathbb{R}^m)$, the differential is defined on the tangent space $T_u\mathcal{U}_{q_0} \simeq L^2([0,1], \mathbb{R}^m)$.

### 8.1.1 Regularity of the end-point map: proof of Proposition 8.5.

The smoothness of the end-point map $E_{q_0} : \mathcal{U}_{q_0} \to M$ is equivalent to the smoothness of the map $a \circ E_{q_0} : \mathcal{U}_{q_0} \to \mathbb{R}$ for every smooth scalar function $a : M \to \mathbb{R}$. This is equivalent to prove that the end-point map is smooth in coordinates, adopting the viewpoint of chronological calculus.

Let $f_u(q) := \sum_{i=1}^m u_i f_i(q)$. The end-point map from $q_0$ can be rewritten as the right chronological exponential (cf. Chapter 6)

$$E_{q_0}(u) = q_0 \circ \overrightarrow{\exp} \int_0^1 f_{u(t)}\, dt. \tag{8.5}$$

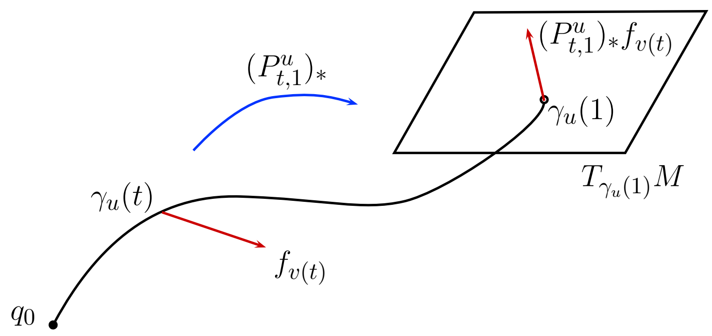



Figure 8.1: Differential of the end-point map.

We will show that for every control $u$ in the set $\mathcal{U}_{q_0}$ we can write a Taylor expansion for $E_{q_0}$ around $u$ and control the remainder at the corresponding order.

*Step 1.* Let us first compute the Taylor expansion of $E_{q_0}$ at the control $u = 0$. We omit the subscript $q_0$ and write

$$E(v(\cdot)) = \overrightarrow{\exp} \int_0^1 f_{v(t)}\, dt. \tag{8.6}$$

Using the Volterra series expansion (cf. Section 6.4), let us split it into the sum of the two terms

$$E(v(\cdot)) = S_N(v) + R_N(v), \tag{8.7}$$

where, for fixed $N \geq 1$

$$S_N(v) = \mathrm{Id} + \sum_{k=1}^{N-1} \int \cdots \int_{\Delta_k(1)} f_{v(s_k)} \circ \cdots \circ f_{v(s_1)} ds,$$

$$R_N(v) = \int \cdots \int_{\Delta_N(1)} P^v_{0,s_N} \circ f_{v(s_N)} \circ \cdots \circ f_{v(s_1)} ds.$$

By linearity of $f_v$ with respect to $v$, the integrand in the $k$-th term in the sum $S_N$ is $k$-linear as a function of $v(s_1), \dots, v(s_k)$. Moreover, applying Theorem 6.20 with $t = 1$, for every $\alpha \in \mathbb{N}$ and compact set $K \subset M$

$$\|R_N(v)a\|_{\alpha,K} \leq \frac{C}{N!} e^{C\|v\|_2} \|v\|_2^N \|a\|_{\alpha+N,K'}, \tag{8.8}$$

for some $K'$ compact set containing $K$ and some constant $C = C_{\alpha,N,K'} > 0$. We stress that the previous inequality holds (for suitable values of the constants) for every $N \in \mathbb{N}$. In the particular case when $N = 2$ it gives

$$\left\| \left( E(v(\cdot)) - \int_0^1 f_{v(t)} dt \right) a \right\|_{\alpha,K} \leq C e^{C\|v\|_2} \|v\|_2^2 \|a\|_{\alpha+1,K'}. \tag{8.9}$$

Since $a$ is arbitrary, choosing $\alpha = 0$ and a compact set $K$ containing the point $q_0$ one has, for $v$ sufficiently small

$$\left| E_{q_0}(v(\cdot)) - \int_0^1 f_{v(t)}(q_0) dt \right| \leq C e^{C\|v\|_2} \|v\|_2^2, \tag{8.10}$$

the inequality being meaningful in coordinates. Since the map $v \mapsto \int_0^1 f_{v(t)}(q_0)dt$ is linear and the right hand side is $o(\|v\|_2)$, the end-point map is differentiable at $u = 0$ and (8.4) holds.

Estimates proving higher order differentiability at $u = 0$ are proved in a similar way.

*Step 2.* To compute the Taylor expansion at an arbitrary $u \in \mathcal{U}_{q_0}$, let us consider the expansion in a neighborhood of $v = 0$ of the map

$$v \mapsto E_{q_0}(u+v) = q_0 \circ \overrightarrow{\exp} \int_0^1 f_{(u+v)(t)} dt.$$

Using the variation formula (6.36), one can write

$$\begin{aligned} \overrightarrow{\exp} \int_0^1 f_{(u+v)(t)} dt &= \overrightarrow{\exp} \int_0^1 f_{u(t)} + f_{v(t)} dt \\ &= \overrightarrow{\exp} \int_0^1 \left( \overrightarrow{\exp} \int_0^t \operatorname{ad} f_{u(s)} ds \right) f_{v(t)} dt \circ \overrightarrow{\exp} \int_0^1 f_{u(t)} dt \\ &= \overrightarrow{\exp} \int_0^1 (P^u_{0,t})^{-1}_* f_{v(t)} dt \circ P^u_{0,1}. \end{aligned} \tag{8.11}$$

We can then rewrite

$$E_{q_0}(u+v) = G^u_{q_0}(v) \circ P^u_{0,1}, \tag{8.12}$$

where $G^u_{q_0} : \mathcal{U}_{q_0} \to M$ is the map defined as follows (notice that $P^u_{0,1}$ is a fixed diffeomorphism that does not depend on $v$)

$$G^u_{q_0}(v) := q_0 \circ \overrightarrow{\exp} \int_0^1 (P^u_{0,t})^{-1}_* f_{v(t)} dt.$$

Then, the expansion of (8.12) at $v = 0$ is obtained by the Volterra expansion of the map $G^u_{q_0}$ with respect to $v$. Using the same computations and estimates as in Step 1, replacing $f_{v(t)}$ with $(P^u_{0,t})^{-1}_* f_{v(t)}$, one obtains

$$D_0 G^u_{q_0}(v) = q_0 \circ \int_0^1 (P^u_{0,t})^{-1}_* f_{v(t)} dt = \int_0^1 (P^u_{0,t})^{-1}_* f_{v(t)}(q_0) dt, \tag{8.13}$$

and, by composition (recall that $G \circ P = P \circ G$ in chronological notation)

$$\begin{aligned} D_u E_{q_0}(v) = (P^u_{0,1})_* \circ D_0 G^u_{q_0}(v) &= (P^u_{0,1})_* \int_0^1 (P^u_{0,t})^{-1}_* f_{v(t)}(q_0) dt \\ &= \int_0^1 (P^u_{t,1})_* f_{v(t)}(q_1) dt. \end{aligned}$$

where we denote $q_1 := E_{q_0}(u)$. By similar computations one obtains higher order Taylor polynomial and a control on the corresponding remainder.

*Remark* 8.6. Notice that the decomposition of the non-autonomous flow associated with $u + v$ into the flow associated with $u$ and a correction term obtained via the variation formula in (8.11) translates in "chronological terms" the change of variables argument used in the ODE proof of Proposition 3.59 (cf. Section 3.4.2).

## 8.2 Lagrange multipliers rule

Let $\mathcal{U}$ be an open set of an Hilbert space $\mathcal{H}$, and let $M$ be a smooth $n$-dimensional manifold. Consider two smooth maps

$$\varphi : \mathcal{U} \to \mathbb{R}, \qquad F : \mathcal{U} \to M. \tag{8.14}$$

In this section we discuss the Lagrange multipliers rule for the minimization of the function $\varphi$ under the constraint defined by $F$. More precisely, we want to write a necessary condition satisfied by the solutions of the problem

$$\min \varphi\big|_{F^{-1}(q)}, \qquad q \in M. \tag{8.15}$$

**Theorem 8.7.** *Assume $u \in \mathcal{U}$ is solution of the minimization problem* (8.15)*. Then there exists $(\lambda, \nu) \in T_q^*M \times \mathbb{R}$ such that $(\lambda, \nu) \neq (0, 0)$ and*

$$\lambda\, D_u F + \nu D_u \varphi = 0. \tag{8.16}$$

We explicitly remark that formula (8.16) means that for every $v \in \mathcal{H} = T_u\mathcal{U}$ one has

$$\langle \lambda, D_u F(v)\rangle + \nu D_u \varphi(v) = 0.$$

The compact notation in (8.16) will be used in the sequel, with analogous meaning.

*Proof.* Let us prove that if $u \in \mathcal{U}$ is solution of the minimization problem (8.15), then $u$ is a critical point for the extended map $\Psi : \mathcal{U} \to M \times \mathbb{R}$ defined by $\Psi(v) = (F(v), \varphi(v))$.

Indeed, if $u$ is not a critical point for $\Psi$, then $D_u\Psi$ is surjective (notice that target space is finite-dimensional). By (a corollary of) the inverse function theorem, this implies that $\Psi$ is locally open at $u$. In particular, for every neighborhood $V$ of $u$, there exists $v \in V$ such that $F(v) = F(u) = q$ and $\varphi(v) < \varphi(u)$, that contradicts that $u$ is a constrained minimum.

Hence $D_u\Psi = (D_u F, D_u\varphi)$ is not surjective and there exists a non zero covector $(\lambda, \nu) \in T_q^*M \times \mathbb{R}$ annihilating the image of $D_u\Psi$, i.e., such that $\lambda\, D_u F + \nu D_u\varphi = 0$. $\square$

## 8.3 Pontryagin extremals via Lagrange multipliers

Applying the previous result to the case when $F = E_{q_0}$ is the end-point map based at $q_0$ and $\varphi = J$ is the sub-Riemannian energy, one immediately obtains the following result.

**Corollary 8.8.** *Assume that a control $u \in \mathcal{U}$ is a solution of the minimization problem* (8.1)*, then there exists $(\lambda, \nu) \in T_q^*M \times \mathbb{R}$ such that $(\lambda, \nu) \neq (0, 0)$ and*

$$\lambda\, D_u E_{q_0} + \nu D_u J = 0. \tag{8.17}$$

Recall that, since $J(u) = \frac{1}{2}\|u\|^2_{L^2}$, then $D_u J(v) = (u, v)_{L^2}$ and, identifying $L^2([0, 1], \mathbb{R}^m)$ with its dual, we have $D_u J = u$.

We now prove that these necessary conditions are equivalent to those obtained in Chapter 4.

**Proposition 8.9.** *We have the following:*

*(N) $(u(t), \lambda(t))$ is a normal extremal if and only if there exists $\lambda_1 \in T_{q_1}^*M$, where $q_1 = E_{q_0}(u)$, such that $\lambda(t) = (P^u_{t,1})^*\lambda_1$ for all $t$, and $u$ satisfies* (8.17) *with $(\lambda, \nu) = (\lambda_1, -1)$, namely*

$$\lambda_1 D_u E_{q_0} = u. \tag{8.18}$$

*(A)* $(u(t), \lambda(t))$ *is an abnormal extremal if and only if there exists* $\lambda_1 \in T^*_{q_1}M$*, where* $q_1 = E_{q_0}(u)$*, such that* $\lambda(t) = (P^u_{t,1})^*\lambda_1$ *for all* $t$*, and* $u$ *satisfies* (8.17) *with* $(\lambda, \nu) = (\lambda_1, 0)$*, namely*

$$\lambda_1 D_u E_{q_0} = 0. \tag{8.19}$$

*Proof.* Let us prove (N). The proof of (A) is similar.

Recall that the pair $(u(t), \lambda(t))$ is a normal extremal if the curve $\lambda(t)$ satisfies $\lambda(t) = (P^u_{t,1})^*\lambda(1)$ (that is equivalent to say that $\lambda(t)$ is a solution of the Hamiltonian system, cf. Chapter 4) and $\langle \lambda(t), f_i(\gamma(t))\rangle = u_i(t)$ for every $i = 1, \ldots, m$, where $\gamma(t) = \pi(\lambda(t))$.

Assume that $u$ satisfies (8.18) for some $\lambda_1$, let us prove that the curve defined by $\lambda(t) := (P^u_{t,1})^*\lambda_1$ is a normal extremal. Condition (8.18) means that for every $v \in L^2([0, T], \mathbb{R}^m)$ we have

$$\langle \lambda_1, D_u E_{q_0}(v)\rangle = (u, v)_{L^2}. \tag{8.20}$$

Using (8.4), the left hand side is rewritten as follows

$$\begin{aligned}\langle \lambda_1, D_u E_{q_0}(v)\rangle &= \int_0^1 \big\langle \lambda_1, (P^u_{t,1})_* f_{v(t)}(q_1)\big\rangle\, dt = \int_0^1 \big\langle (P^u_{t,1})^*\lambda_1, f_{v(t)}(\gamma(t))\big\rangle\, dt \\ &= \int_0^1 \big\langle \lambda(t), f_{v(t)}(\gamma(t))\big\rangle\, dt = \int_0^1 \sum_{i=1}^m \langle \lambda(t), f_i(\gamma(t))\rangle\, v_i(t)dt,\end{aligned}$$

where we used that $\gamma(t) = (P^u_{t,1})^{-1}(q_1)$. Then (8.20) becomes

$$\int_0^1 \sum_{i=1}^m \langle \lambda(t), f_i(\gamma(t))\rangle\, v_i(t)dt = \int_0^1 \sum_{i=1}^m u_i(t)v_i(t)dt. \tag{8.21}$$

Since $v$ is arbitrary, this implies $\langle \lambda(t), f_i(\gamma(t))\rangle = u_i(t)$ for a.e. $t \in [0, 1]$ and every $i = 1, \ldots, m$. Following the same computations in the oppposite direction we have that if $(u(t), \lambda(t))$ is a normal extremal then the identity (8.18) is satisfied. $\square$

**Exercise 8.10.** Prove that if $\gamma$ is a length-minimizer associated with the minimal control $u$, which admits two different normal lifts, then it also admits an abnormal lift.

## 8.4 Critical points and second order conditions

In this section, we develop second order conditions for constrained critical points, when the constraint is defined by a submersion (at least locally). In the Section 8.5 we will apply this results in the sub-Riemannian case, obtaining second order conditions for normal extremals (that are not abnormal).

In what follows $\mathcal{H}$ denotes a separable Hilbert space. Recall that a smooth submanifold of $\mathcal{H}$ is a subset $\mathcal{V} \subset \mathcal{H}$ such that for every point $v \in \mathcal{V}$ there is an open neighborhood $\mathcal{Y}$ of $v$ in $\mathcal{H}$ and a smooth diffeomorphism $\phi : \mathcal{V} \to \mathcal{W}$ to an open subset $\mathcal{W} \subset \mathcal{H}$ such that $\phi(\mathcal{V} \cap \mathcal{Y}) = \mathcal{W} \cap \mathcal{U}$ for $\mathcal{U}$ a closed linear subspace of $\mathcal{H}$.

We now recall the implicit function theorem in this setting.

**Proposition 8.11** (Implicit function theorem)**.** *Let $M$ be a smooth manifold, $\mathcal{H}$ be an Hilbert space and let $F : \mathcal{H} \to M$ be a smooth map. Fix $q \in M$ and set $\mathcal{V}_q = F^{-1}(q)$. If the Fréchet differential $D_uF : \mathcal{H} \to T_qM$ is surjective for every $u \in \mathcal{V}_q$, then $\mathcal{V}_q$ is a smooth submanifold whose codimension is equal to the dimension of $M$. Moreover $T_u\mathcal{V}_q = \ker D_uF$.*

We now define critical points.

**Definition 8.12.** Let $\varphi : \mathcal{H} \to \mathbb{R}$ be a smooth function and $N \subset \mathcal{H}$ be a smooth submanifold. Then $u \in N$ is called a *critical point* of $\varphi\big|_N$ if $D_u\varphi\big|_{T_uN} = 0$.

We start with a geometric version of the Lagrange multipliers rule, which characterizes constrained critical points. This construction is then used later to develop a second order analysis.

**Proposition 8.13** (Lagrange multipliers rule)**.** *Let $\mathcal{U}$ be an open subset of $\mathcal{H}$ and assume that $u \in \mathcal{U}$ is a regular point of the smooth map $F : \mathcal{U} \to M$. Let $q = F(u)$. Then $u$ is a critical point of $\varphi\big|_{F^{-1}(q)}$ if and only if it exists $\lambda \in T_q^*M$ such that*

$$\lambda\, D_uF = D_u\varphi. \tag{8.22}$$

*Proof.* Recall that the differential of $F$ is the map

$$D_uF : T_u\mathcal{U} \to T_qM, \quad \text{where} \quad q = F(u).$$

Since $u$ is a regular point, $D_uF$ is surjective and, by implicit function theorem, the level set $\mathcal{V}_q := F^{-1}(q)$ is a smooth submanifold (of codimension $n = \dim M$), with $u \in \mathcal{V}_q$ and $T_u\mathcal{V}_q = \ker D_uF$. By definition, $u$ is a critical point of $\varphi\big|_{\mathcal{V}_q}$ if and only if $D_u\varphi\big|_{T_u\mathcal{V}_q} = D_u\varphi\big|_{\ker D_uF} = 0$, i.e.,

$$\ker D_uF \subset \ker D_u\varphi. \tag{8.23}$$

Using Exercice 8.14, (8.23) is equivalent to the existence a linear map $\lambda : T_qM \to \mathbb{R}$ (namely $\lambda \in T_q^*M$) that makes the following diagram commutative.

$$\begin{tikzcd} \mathcal{U} \simeq T_u\mathcal{U} \arrow[r, "D_uF"] \arrow[dr, "d_u\varphi"'] & T_qM \arrow[d, "\lambda"] \\ & \mathbb{R} \end{tikzcd} \tag{8.24}$$

□

**Exercise 8.14.** Let $V$ be a separable Hilbert spaces and $W$ be a finite-dimensional vector space. Let $G : V \to W$ and $\phi : V \to \mathbb{R}$ two linear maps such that $\ker G \subset \ker \phi$. Then show that there exists a linear map $\lambda : W \to \mathbb{R}$ such that $\lambda \circ G = \phi$.

Next we consider second order derivatives. Let $\mathcal{U}$ be an open set of an Hilbert space $\mathcal{H}$. Recall that, for a smooth function $\varphi : \mathcal{U} \to \mathbb{R}$, the *first* and *second differential* are defined in the following way, respectively

$$D_u\varphi(v) = \frac{d}{ds}\bigg|_{s=0} \varphi(u + sv), \qquad D_u^2\varphi(v, v) = \frac{d^2}{ds^2}\bigg|_{s=0} \varphi(u + sv). \tag{8.25}$$

For a smooth map $F : \mathcal{U} \to M$ whose range is a smooth finite-dimensional manifold $M$, the similar formulas

$$D_u F(v) = \frac{d}{ds}\bigg|_{s=0} F(u + sv), \qquad D_u^2 F(v, v) = \frac{d^2}{ds^2}\bigg|_{s=0} F(u + sv). \tag{8.26}$$

give a well-defined first differential $D_u F : \mathcal{H} \to T_{F(u)} M$, while the second differential $D_u^2 F$ depends on the choice of a set of coordinates on $M$, in general.

For a smooth function $\psi$ defined on a submanifold $\mathcal{V}$ of $\mathcal{H}$, we can no longer use the linear structure to define differentials as in (8.25)-(8.26). The first differential of a smooth function $\psi : \mathcal{V} \to \mathbb{R}$ at a point $u \in \mathcal{V}$ is well-defined as

$$D_u \psi : T_u \mathcal{V} \to \mathbb{R}, \qquad D_u \psi(v) = \frac{d}{ds}\bigg|_{s=0} \psi(w(s)),$$

where $w : (-\varepsilon, \varepsilon) \to \mathcal{V}$ is a curve that satisfies $w(0) = u,\ \dot{w}(0) = v$.

For the second differential things are more delicate. Indeed the similar formula

$$v \in T_u \mathcal{V} \mapsto \frac{d^2}{ds^2}\bigg|_{s=0} \psi(w(s)), \tag{8.27}$$

where $w : (-\varepsilon, \varepsilon) \to \mathcal{V}$ is a curve that satisfies $w(0) = u,\ \dot{w}(0) = v$, is well-defined (i.e., the right hand side depends only on $v$) only if $u$ is a critical point of $\psi$. If this is not the case, the quantity (8.27) depends also on the second derivative of $w$, as it is easily checked.

In conclusion, if $u$ is a critical point of $\psi : \mathcal{V} \to \mathbb{R}$ (i.e., $D_u \psi = 0$) the second order differential (8.27) is a well-defined quadratic form $T_u \mathcal{V}$, that is called the *Hessian* of $\psi$ at $u$:

$$\operatorname{Hess}_u \psi : T_u \mathcal{V} \to \mathbb{R}, \qquad v \mapsto \frac{d^2}{ds^2}\bigg|_{s=0} \psi(w(s)). \tag{8.28}$$

*Remark* 8.15. If $\psi = \varphi|_{\mathcal{V}}$ is the restriction on $\mathcal{V}$ of a function $\varphi : \mathcal{H} \to \mathbb{R}$ defined globally on $\mathcal{H}$, then $D_u \psi$ coincides with the restriction of the differential defined on the ambient space $\mathcal{H}$, namely $D_u \psi = D_u \phi|_{T_u \mathcal{V}}$.

On the other hand, the Hessian of $\psi = \varphi|_{\mathcal{V}}$ (defined at a critical point $u$) *does not* coincide, in general, with the restriction of the second differential of $\varphi$ to the tangent space $T_u \mathcal{V}$, as we now discuss.

Let $F : \mathcal{U} \to M$ and $\varphi : \mathcal{U} \to \mathbb{R}$ be smooth, and consider the restriction $\psi = \varphi\big|_{\mathcal{V}_q}$ where $\mathcal{V}_q = F^{-1}(q)$ is a smooth submanifold of $\mathcal{H}$. Using that $T_u F^{-1}(q) = \ker D_u F$, the Hessian (at a critical point $u$) is the well-defined quadratic form

$$\operatorname{Hess}_u \left( \varphi\big|_{F^{-1}(q)} \right) : \ker D_u F \to \mathbb{R}$$

that is computed in terms of the second differentials of $\varphi$ and $F$ (defined as in (8.25)-(8.26)) as follows.

**Proposition 8.16.** *Let $\varphi : \mathcal{U} \to \mathbb{R}$ and $F : \mathcal{U} \to M$ be smooth. Fix $q \in M$ and assume that $F$ is a submersion on $F^{-1}(q)$. Assume $u$ is a critical point for $\varphi\big|_{F^{-1}(q)}$. Then for all $v \in \ker D_u F$ we have*

$$\operatorname{Hess}_u \left( \varphi\big|_{F^{-1}(q)} \right)(v) = D_u^2 \varphi(v, v) - \lambda\, D_u^2 F(v, v), \tag{8.29}$$

*where $\lambda \in T_q^* M$ satisfies the identity $\lambda\, D_u F = D_u \varphi$.*

*Remark* 8.17. We stress that in (8.29), while the left hand side is a well-defined object, in the right hand side $D_u^2\varphi$ is well-defined thanks to the linear structure of $\mathcal{H}$, while $D_u^2 F$ needs also a choice of coordinates in the manifold $M$.

*Proof of Proposition 8.16.* By assumption $F^{-1}(q) \subset \mathcal{U}$ is a smooth submanifold in a Hilbert space. Fix $u \in F^{-1}(q)$ and consider a smooth path $w(s)$ in $\mathcal{U}$ such that $w(0) = u$ and $w(s) \in F^{-1}(q)$ for all $s$. Fixing some local coordinates in a neighborhood of $q$ in $M$, and differentiating twice with respect to $s$ the identity $F(w(s)) = q$, we have

$$D_u F(\dot{u}) = 0, \qquad D_u^2 F(\dot{u}, \dot{u}) + D_u F(\ddot{u}) = 0. \tag{8.30}$$

where we denoted by $\dot{u} = \dot{w}(0)$ and $\ddot{u} = \ddot{w}(0)$. Analogous computations for $\varphi$ gives

$$\begin{aligned}
\mathrm{Hess}_u \left( \varphi\big|_{F^{-1}(q)} \right)(\dot{u}) &= \left.\frac{d^2}{ds^2}\right|_{s=0} \varphi(w(s)) \\
&= D_u^2\varphi(\dot{u}, \dot{u}) + D_u\varphi(\ddot{u}) \\
&= D_u^2\varphi(\dot{u}, \dot{u}) + \lambda D_u F(\ddot{u}) \qquad \text{(by } \lambda\, D_u F = D_u\varphi) \\
&= D_u^2\varphi(\dot{u}, \dot{u}) - \lambda D_u^2 F(\dot{u}, \dot{u}) \qquad \text{(by (8.30))}
\end{aligned}$$

□

### 8.4.1 The manifold of Lagrange multipliers

As before, let us consider two smooth maps $\varphi : \mathcal{U} \to \mathbb{R}$ and $F : \mathcal{U} \to M$ defined on an open set $\mathcal{U}$ of a separable Hilbert space $\mathcal{H}$.

**Definition 8.18.** We say that a pair $(u, \lambda)$, with $u \in \mathcal{U}$ and $\lambda \in T^*M$, is a *Lagrange point* for the pair $(F, \varphi)$ if $\lambda \in T^*_{F(u)}M$ and $D_u\varphi = \lambda\, D_u F$. We denote the set of all Lagrange points by $C_{F,\varphi}$. More precisely

$$C_{F,\varphi} = \{(u, \lambda) \in \mathcal{U} \times T^*M \mid F(u) = \pi(\lambda),\ D_u\varphi = \lambda\, D_u F\}. \tag{8.31}$$

The set $C_{F,\varphi}$ is a well-defined subset of the vector bundle $F^*(T^*M)$ on $\mathcal{U}$, that we recall is defined as follows (cf. also Definition 2.54)

$$F^*(T^*M) = \{(u, \lambda) \in \mathcal{U} \times T^*M \mid F(u) = \pi(\lambda)\}. \tag{8.32}$$

Under the following regularity conditions on the pair $(F, \varphi)$, the set $C_{F,\varphi}$ is a smooth submanifold.

**Definition 8.19.** The pair $(F, \varphi)$ is said to be a *Morse pair* (or a *Morse problem*) if 0 is a regular value for the smooth map

$$\theta : F^*(T^*M) \to \mathcal{H}^* \simeq \mathcal{H}, \qquad \theta(u, \lambda) = D_u\varphi - \lambda\, D_u F. \tag{8.33}$$

where we are using the standard identification of $\mathcal{H}$ with its dual $\mathcal{H}^*$.

*Remark* 8.20. Notice that, if $M$ is a single point, then $F$ is the trivial map and with this definition we have that $(F, \varphi)$ is a Morse pair if and only if $\varphi$ is a Morse function. Indeed in this case $D_u F = 0$, and 0 is a regular value for $\theta$ if, by definition, the second differential $D_u^2\varphi$ is non-degenerate.

**Proposition 8.21.** *If $(F, \varphi)$ defines a Morse problem, then $C_{F,\varphi}$ is a smooth manifold in $F^*(T^*M)$. Moreover* $\dim C_{F,\varphi} = \dim M = n$.

Notice that $C_{F,\varphi} = \theta^{-1}(0)$ and, by definition of Morse pair, 0 is a regular value of $\theta$. The fact that $C_{F,\varphi}$ is a smooth manifold follows from the following version of the implicit function theorem.

**Lemma 8.22.** *Let $N$ be a smooth Hilbert manifold and $\mathcal{H}$ a Hilbert space. Consider a smooth map $f : N \to \mathcal{H}$ and assume that 0 is a regular value of $f$. Then $f^{-1}(0)$ is a smooth submanifold of $N$.*

If the dimension of $\mathcal{U}$, the target space of $\theta$, were finite, a simple dimensional argument would permit to compute the dimension of $C_{F,\varphi} = \theta^{-1}(0)$ (cf. Proposition 8.11). Indeed, since the differential of $\theta$ is surjective we would have that

$$\dim\, F^*(T^*M) - \dim\, C_{F,\varphi} = \dim \mathcal{U},$$

so we could compute the dimension of $C_{F,\varphi}$

$$\begin{aligned}\dim\, C_{F,\varphi} &= \dim\, F^*(T^*M) - \dim \mathcal{U} \\ &= (\dim \mathcal{U} + \operatorname{rank} T^*M) - \dim \mathcal{U} \\ &= \operatorname{rank} T^*M = n.\end{aligned}$$

However, in the case $\dim \mathcal{U} = +\infty$, the above argument is not valid, so we need the following explicit argument.

*Proof of Proposition 8.21.* To prove the statement, let us choose a set of coordinates $\lambda = (\xi, x)$ in $T^*M$. Then a triple $(u, \xi, x)$ in $F^*(T^*M)$ belongs to $C_{F,\varphi}$ if and only if it satisfies the following equations

$$\begin{cases} D_u\varphi - \xi D_u F = 0 \\ F(u) = x \end{cases} \tag{8.34}$$

where here $\xi$ is thought as a row vector (which we denote by $\mathbb{R}^{n*}$). To compute $\dim C_{F,\varphi}$, it will be enough to compute the dimension of its tangent space $T_{(u,\xi,x)}C_{F,\varphi}$ at a every point $(u, \xi, x)$. The tangent space $T_{(u,\xi,x)}C_{F,\varphi}$ is described in coordinates by the set of points $(u', \xi', x')$ satisfying the equations[1]

$$\begin{cases} D_u^2\varphi(u', \cdot) - \xi D_u^2 F(u', \cdot) - \xi' D_u F(\cdot) = 0 \\ D_u F(u') = x' \end{cases} \tag{8.35}$$

where here $D_u^2\varphi$ and $D_u^2 F$ are the bilinear maps associated with the quadratic forms defined in (8.25)-(8.26). Let us denote $Q : \mathcal{H} \to \mathcal{H}^* \simeq \mathcal{H}$ the linear map defined by

$$Q(u') = D_u^2\varphi(u', \cdot) - \xi D_u^2 F(u', \cdot). \tag{8.36}$$

Since $Q$ is defined by second derivatives of the maps $F$ and $\varphi$, which are smooth, it is a symmetric operator on the Hilbert space $\mathcal{H}$.

The definition of Morse problem is immediately rewritten as follows: the pair $(F, \varphi)$ defines a Morse problem if and only if the following map is surjective.

$$\Theta : \mathcal{H} \times \mathbb{R}^{n*} \to \mathcal{H}^* \simeq \mathcal{H}, \qquad \Theta(u', \xi') = Q(u') - B(\xi'). \tag{8.37}$$

[1]if a submanifold $C$ of a manifold $Z$ is described as the set $\{z \in Z \mid \Psi(z) = 0\}$, then its tangent space $T_zC$ at a point $z \in C$ is described by the linear equation $\{z' \in Z \mid D_z\Psi(z') = 0\}$.

where we denoted with $B : \mathbb{R}^{n*} \to \mathcal{H}^* \simeq \mathcal{H}$ the linear map

$$B(\xi') = \xi' D_u F(\cdot).$$

Notice that the first equation in (8.35) coincides with $\Theta = 0$. As a result, for each $(u', \xi') \in \ker \Theta$ there exists a unique $(u', \xi', x') \in T_{(u,\xi,x)} C_{F,\varphi}$ by setting $x' = D_u F(u')$. It follows that $\dim C_{F,\varphi} = \dim \ker \Theta$. Since $Q$ is self-adjoint, we have

$$\mathcal{H} = \ker Q \oplus \overline{\operatorname{im} Q}, \qquad \dim \ker Q = \operatorname{codim} \operatorname{im} Q.$$

Using that $\Theta$ is surjective and $\dim(\operatorname{im} B) \leq n$ we get that

$$\dim \ker Q = \operatorname{codim} \operatorname{im} Q \leq \dim \operatorname{im} B \leq n.$$

Then $\ker Q$ is finite dimensional (in particular $\operatorname{im} Q$ is closed in $\mathcal{H}$ and $\mathcal{H} = \ker Q \oplus \operatorname{im} Q$).

If we denote with $\pi_{\ker} : \mathcal{H} \to \ker Q$ and $\pi_{\operatorname{im}} : \mathcal{H} \to \operatorname{im} Q$ the orthogonal projection onto the two subspaces, it is easy to see that

$$\Theta(u', \xi') = 0 \quad \Longleftrightarrow \quad \begin{cases} \pi_{\ker} B(\xi') = 0, \\ \pi_{\operatorname{im}} B(\xi') = Q(u'). \end{cases}$$

Moreover $\pi_{\ker} B : \mathbb{R}^{n*} \to \ker Q$ is a surjective linear map between finite-dimensional spaces (the surjectivity is a consequence of the fact that $\Theta$ is surjective, and the symmetry of $Q$ as operator on $\mathcal{H}$). In particular we have $\dim \ker (\pi_{\ker} B) = n - \dim \ker Q$. Then we get the identity

$$\dim \ker \Theta = \dim \ker Q + \dim \ker (\pi_{\ker} B) = \dim \ker Q + (n - \dim \ker Q) = n,$$

since $\pi_{\ker} B : \mathbb{R}^n \to \ker Q$ is a surjective map. It follows that $\dim C_{F,\varphi} = n$.

□

As a consequence of Proposition 8.21, we have a convenient criterion to check whether a pair $(F, \varphi)$ defines a Morse problem.

**Lemma 8.23.** *The pair $(F, \varphi)$ defines a Morse problem if and only if for every $u \in \mathcal{U}$ we have*

*(i)* $\operatorname{im} Q$ *is closed,*

*(ii)* $\ker Q \cap \ker D_u F = \{0\}$*,*

*where $Q$ denotes the operator defined in* (8.36)*.*

*Proof.* Assume that $(F, \varphi)$ is a Morse problem. Then, arguing as in the proof of Proposition 8.21, $\operatorname{im} Q$ has finite codimension, hence is closed, and (i) is proved. Moreover, let $w \in \ker Q \cap \ker D_u F$. Then $Q(w) = 0$ and $D_u F(w) = 0$. Hence, for every $u', \xi'$ one has

$$(u', Q(w))_{\mathcal{H}} - \xi' D_u F(w) = 0, \qquad \forall\, (u', \xi').$$

Since $Q$ is self-adjoint, we can rewrite the previous identity as

$$(Q(u'), w)_{\mathcal{H}} - \xi' D_u F(w) = 0, \qquad \forall\, (u', \xi'),$$

This implies that $w \in \mathcal{U}$ is orthogonal to $\operatorname{im} \Theta$. Since $(F, \varphi)$ is a Morse problem, then the image of the differential of the map (8.37) is surjective, hence $w = 0$. This implies (ii). The converse implications are proved similarly. □

**Definition 8.24.** Let $M, N$ be $n$-dimensional manifolds. An immersion $F : N \to T^*M$ is said to be a *Lagrange immersion* if $F^*\sigma = 0$, where $\sigma$ denotes the standard symplectic form on $T^*M$.

Let us consider now the projection map $F_c : C_{F,\varphi} \longrightarrow T^*M$ defined by:

$$F_c(u, \lambda) = \lambda.$$

**Proposition 8.25.** *If the pair $(F, \varphi)$ defines a Morse problem, then $F_c$ is a Lagrange immersion.*

*Proof.* First prove that (i) $F_c$ is an immersion and then (ii) $F_c^*\sigma = 0$.

(i). Recall that $F_c : C_{F,\varphi} \to T^*M$ where

$$C_{F,\varphi} = \{(u, \xi, x) \,|\, \text{equations (8.34) hold}\}.$$

The differential $D_{(u,\lambda)}F_c : T_{(u,\lambda)}C_{F,\varphi} \to T_\lambda T^*M$ is defined by the linearization of equations (8.34), and in a coordinate set where $\lambda = (\xi, x)$ we have

$$T_{(u,\lambda)}C_{F,\varphi} = \{(u', \xi', x') \,|\, \text{equations (8.35) hold}\},$$

and $D_{(u,\lambda)}F_c(u', \xi', x') = (\xi', x')$. By (8.35), it easily seen that

$$D_{(u,\lambda)}F_c(u', \xi', x') = 0 \qquad \text{iff} \qquad Q(u') = D_uF(u') = 0.$$

Since $(F, \varphi)$ defines a Morse problem we have by Lemma 8.23 that $u' = 0$. This proves that the differential of $F_c$ is injective at every point, i.e., $F_c$ is an immersion.

(ii). Since $\sigma = ds$, where $s$ is the tautological form, and since the pullback commutes with the differential, we have $F_c^*\sigma = dF_c^*s$, it is sufficient to show that $F_c^*s$ is closed. Let us show the identity

$$F_c^*s = D(\varphi \circ \pi_{\mathcal{U}})\big|_{C_{F,\varphi}}.$$

where $\pi_{\mathcal{U}}$ denotes the canonical projection of $C_{F,\varphi}$ over $\mathcal{U}$. By definition of the map $F_c$ and $F^*(T^*M)$, the following diagram is commutative:

$$\begin{tikzcd} C_{F,\varphi} \arrow[r, "F_c"] \arrow[d, "\pi_{\mathcal{U}}"'] & T^*M \arrow[d, "\pi"] \\ \mathcal{U} \arrow[r, "F"'] & M \end{tikzcd} \tag{8.38}$$

Moreover, notice that if $\phi : M \to N$ is smooth and $\omega \in \Lambda^1(N)$, by definition of pull-back we have $(\phi^*\omega)_q = \omega_{\phi(q)} \circ D_q\phi$. Hence we have

$$\begin{aligned} (F_c^*s)_{(u,\lambda)} &= s_\lambda \circ D_{(u,\lambda)}F_c \\ &= \lambda \circ \pi_* \circ D_{(u,\lambda)}F_c && \text{(by } s_\lambda = \lambda \circ \pi_*) \\ &= \lambda \circ D_uF \circ \pi_{\mathcal{U}*} && \text{(by (8.38))} \\ &= D_u(\varphi \circ \pi_{\mathcal{U}}) && \text{(by (8.22), for } (u,\lambda) \in C_{F,\varphi}) \end{aligned}$$

□

**Definition 8.26.** The set $\mathcal{L}_{F,\varphi} \subset T^*M$ of *Lagrange multipliers* associated with the pair $(F,\varphi)$ is the image of $C_{F,\varphi}$ under the map $F_c$. More precisely

$$\mathcal{L}_{F,\varphi} = \{\lambda \in T^*M \mid \exists\, u \in \mathcal{U} : \lambda D_u F = D_u\varphi\}. \tag{8.39}$$

From Proposition 8.25 it follows that, if $\mathcal{L}_{F,\varphi}$ is a smooth submanifold, then it is a Lagrangian submanifold of $T^*M$, i.e., $\sigma|_{\mathcal{L}_{F,\varphi}} = 0$.

Collecting the results obtained above, we have the following proposition.

**Proposition 8.27.** *Let $(F,\varphi)$ be a Morse pair and assume $(u,\lambda)$ is a Lagrange point such that $u$ is a regular point for $F$. Let $q = F(u) = \pi(\lambda)$. Then the following properties are equivalent:*

*(i)* $\mathrm{Hess}_u\left(\varphi|_{F^{-1}(q)}\right)$ *is degenerate,*

*(ii)* $(u,\lambda)$ *is a critical point for the map* $\pi \circ F_c : C_{F,\varphi} \to M$*,*

*Moreover, if $\mathcal{L}_{F,\varphi}$ is a submanifold, then (i) and (ii) are equivalent to*

*(iii)* $\lambda$ *is a critical point for the map* $\pi|_{\mathcal{L}_{F,\varphi}} : \mathcal{L}_{F,\varphi} \to M$*.*

Notice that the map $\pi \circ F_c$ can be regarded as a restriction $F|_{C_{F,\varphi}}$ in the sense of diagram (8.38).

*Proof.* Recall that $F^{-1}(q)$ is a smooth submanifold and $T_u(F^{-1}(q)) = \ker D_u F$. From Proposition 8.16 that we have the following expression for the Hessian of the restriction

$$\mathrm{Hess}_u\left(\varphi|_{F^{-1}(q)}\right)(v) = (Q(v), v)_{\mathcal{H}}, \qquad \forall\, v \in \ker D_u F.$$

Here $Q$ is the linear operator defined in (8.36), and $(\cdot,\cdot)_{\mathcal{H}}$ denotes the inner product in $\mathcal{H}$. Assume that $\mathrm{Hess}_u\left(\varphi|_{F^{-1}(q)}\right)$ is degenerate, i.e., there exists a non zero $u' \in \ker D_u F$ such that

$$(Q(u'), v)_{\mathcal{H}} = 0, \qquad \forall\, v \in \ker D_u F.$$

This means that $Q(u')$ is orthogonal (with respect to the inner product in $\mathcal{H}$) to $\ker D_u F$. In other words $Q(u')$ is a linear combination of the rows of the Jacobian matrix of $F$, namely there exists a row vector $\xi'$ such that

$$(Q(u'), \cdot)_{\mathcal{H}} = \xi' D_u F(\cdot).$$

Hence $(u', \xi')$ is in the kernel of the map $\Theta$ defined in (8.37), hence the differential of the map $\pi \circ F_c : C_{F,\varphi} \to M$ is not surjective, and (i) implies (ii). The converse is proved similarly.

Notice finally that, if $\mathcal{L}_{F,\varphi}$ is a submanifold, then $\pi|_{\mathcal{L}_{F,\varphi}}$ is well-defined and (ii) is equivalent to (iii) since $F_c$ is an immersion. □

*Remark* 8.28. Notice that, even without requiring that $(F,\varphi)$ is a Morse problem, the previous arguments shows the inclusion

$$\ker Q \cap \ker D_u F \subset \ker \mathrm{Hess}_u(\varphi|_{F^{-1}(q)}). \tag{8.40}$$

Then, if one can prove that $\mathrm{Hess}_u(\varphi|_{F^{-1}(q)})$ is non-degenerate, one proves at the same time that the corresponding pair $(F,\varphi)$ is a Morse pair.

## 8.5 Sub-Riemannian case

In this section we want to specify the theory of Morse problems to the case of sub-Riemannian normal extremal. Hence, in the language of the previous section, we consider the pair $(F, \varphi)$ where $\varphi$ is the energy functional $J$ defined by $J(u) = \frac{1}{2}\int_0^1 |u(t)|^2 dt$ and $F$ is the end-point map $E_{q_0}$.

We already characterized critical points by means of Lagrange multipliers, now we want to consider second order informations. We start by computing the Hessian of the restriction $J\big|_{E_{q_0}^{-1}(q_1)}$. In what follows we assume $q_0$ to be fixed and we write $E = E_{q_0}$.

**Lemma 8.29.** *Let $q_1 \in M$ and $(u, \lambda)$ be a critical point of $J\big|_{E^{-1}(q_1)}$. Then for every $v \in \ker D_u F$*

$$\mathrm{Hess}_u \left( J\big|_{E^{-1}(q_1)} \right)(v) = \|v\|_{L^2}^2 - \left\langle \lambda, D_u^2 E(v,v) \right\rangle, \tag{8.41}$$

*where*

$$D_u^2 E(v,v) = 2 \iint\limits_{0 \le s \le t \le 1} [(P_{s,1}^u)_* f_{v(s)}, (P_{t,1}^u)_* f_{v(t)}](q_1)\, ds dt. \tag{8.42}$$

*and $P_{t,s}^u$ denotes the non-autonomous flow defined by the control $u$.*

*Proof.* By Proposition 8.16 we have

$$\mathrm{Hess}_u \left( J\big|_{E^{-1}(q_1)} \right)(v) = D_u^2 J(v,v) - \lambda D_u^2 E(v,v).$$

It is easy to compute derivatives of $J$. Indeed we can rewrite it as $J(u) = \frac{1}{2}(u,u)_{L^2}$, hence

$$D_u J(v) = (u,v)_{L^2}, \qquad D_u^2 J(v,v) = (v,v)_{L^2} = \|v\|_{L^2}^2, \qquad \forall\, v \in \ker D_u E.$$

It remains to compute the second derivative of the end-point map. Following the arguments contained in the proof of Proposition 8.5 and considering the Volterra expansion up to second order one gets

$$D_u^2 E(v,v) = 2\, q_1 \circ \iint\limits_{0 \le s \le t \le 1} (P_{s,1}^u)_* f_{v(s)} \circ (P_{t,1}^u)_* f_{v(t)} ds dt. \tag{8.43}$$

To end the proof we use the following lemma on chronological calculus, which we will use to symmetrize the right hand side of (8.43).

**Lemma 8.30.** *Let $X_t$ be a non-autonomous vector field on $M$. Then*

$$\iint\limits_{0 \le s \le t \le 1} X_s \circ X_t ds dt = \frac{1}{2} \int_0^1 X_s ds \circ \int_0^1 X_t dt + \frac{1}{2} \iint\limits_{0 \le s \le t \le 1} [X_s, X_t] ds dt. \tag{8.44}$$

*Proof of the Lemma.* We have

$$\begin{aligned}
2\iint\limits_{0\le s\le t\le 1} X_s\odot X_t dsdt &= \iint\limits_{0\le s\le t\le 1} X_s\odot X_t dsdt + \iint\limits_{0\le s\le t\le 1} X_s\odot X_t dsdt \\
&\quad - \iint\limits_{0\le s\le t\le 1} X_t\odot X_s dsdt + \iint\limits_{0\le s\le t\le 1} X_t\odot X_s dsdt \\
&= \iint\limits_{0\le s\le t\le 1} X_s\odot X_t dsdt + \iint\limits_{0\le s\le t\le 1} [X_s, X_t] dsdt + \iint\limits_{0\le s\le t\le 1} X_t\odot X_s dsdt \\
&= \int_0^1\int_0^1 X_s\odot X_t dsdt + \iint\limits_{0\le s\le t\le 1} [X_s, X_t] dsdt \\
&= \int_0^1 X_s ds\odot \int_0^1 X_t dt + \iint\limits_{0\le s\le t\le 1} [X_s, X_t] dsdt.
\end{aligned}$$

□

Using Lemma 8.30 we obtain from (8.43)

$$D_u^2E(v,v) = q_1\odot 2\iint\limits_{0\le s\le t\le 1} [(P^u_{s,1})_*f_{v(s)}, (P^u_{t,1})_*f_{v(t)}]dsdt \tag{8.45}$$

where we used that $q_1\odot\int_0^1 (P^u_{t,1})_*f_{v(t)}dt = 0$ since $v\in\ker D_uE$. □

**Proposition 8.31.** *The pair $(E, J)$ defined by the sub-Riemannian problem is a Morse pair.*

*Proof.* We use the characterization of Lemma 8.23. We have to show that

$$\text{im}\left(\text{Id}-\lambda D_u^2E\right)\ \text{is closed},\qquad \ker\left(\text{Id}-\lambda D_u^2E\right)\cap\ker\left(D_uE\right) = \{0\}. \tag{8.46}$$

Using the notation of Lemma 8.29, and defining $g_v^t := (P^u_{t,1})_*f_v$, we can write

$$D_uE(v) = q_1\odot\int_0^1 g^t_{v(t)}dt.$$

Fix any smooth function $a$ such that $d_{q_1}a = \lambda$. Then

$$\lambda D_u^2E(v,v) = 2q_1\odot\iint\limits_{0\le s\le t\le 1} g^s_{v(s)}\odot g^t_{v(t)}dsdt\odot a \tag{8.47}$$

$$= q_1\odot\iint\limits_{0\le s\le t\le 1} g^s_{v(s)}\odot g^t_{v(t)}dsdt\odot a + \iint\limits_{0\le t\le s\le 1} g^t_{v(t)}\odot g^s_{v(s)}dsdt\odot a \tag{8.48}$$

$$= q_1\odot\int_0^1\int_0^t g^s_{v(s)}\odot g^t_{v(t)}dsdt\odot a + \int_0^1\int_t^1 g^t_{v(t)}\odot g^s_{v(s)}dsdt\odot a \tag{8.49}$$

The kernel of the bilinear form is, by definition, the kernel of the symmetric linear operator associated to it through the scalar product, i.e., the unique symmetric operator $Q$ satisfying

$$\lambda D_u^2 E(v,v) = (Qv,v)_{L^2} = \int_0^1 (Qv)(t)v(t)dt.$$

Then it follows that $Q$ has the following expression

$$(Qv)(t) = \left( \int_0^t g_{v(s)}^s ds \odot g^t + g^t \odot \int_t^1 g_{v(s)}^s ds \right) \odot a \tag{8.50}$$

where $g^t$ denotes the vector $(g_1^t, \ldots, g_m^t)$ and we recall that for each $i = 1, \ldots, m$ the vector field $g_i^t$ is defined by $g_i^t = (P_{t,1})_* f_i$.

Since (8.50) is a compact integral operator, then $I - Q$ is Fredholm, and the closedness of $\operatorname{im}(I - Q)$ follows from the fact that it is of finite codimension. On the other hand, when we restrict to controls $v \in \ker D_u E$ we have the identity (cf. (8.4))

$$q_1 \odot \int_0^t g_{v(s)}^s ds = -q_1 \odot \int_t^1 g_{v(s)}^s ds.$$

Hence, $v$ belongs to the intersection in (8.46) if and only if it satisfies the integral equation

$$v(t) - \lambda \int_0^t \left[ g_{v(s)}^s, g^t \right] (q_1) ds = 0.$$

Thanks to Lemma 8.32, the unique solution to this equation in $L^2([0,T], \mathbb{R}^m)$ is $v = 0$. This proves that $\ker\left(\mathrm{Id} - \lambda D_u^2 E\right) \cap \ker\left(D_u E\right) = \{0\}$ and that the sub-Riemannian problem $(E, J)$ is a Morse pair. $\square$

**Lemma 8.32.** *Let $K(t,s)$ be a function in $L^\infty([0,1] \times [0,1], \mathbb{R}^m)$. Then the integral equation*

$$v(t) = \int_0^t K(t,s)v(s)ds, \tag{8.51}$$

*has the unique solution $v = 0$ in $L^2([0,1], \mathbb{R}^m)$.*

*Proof.* Let $v \in L^2([0,1], \mathbb{R}^m)$ be a solution to the equation

$$v(t) = \int_0^t K(t,s)v(s)ds, \tag{8.52}$$

Notice that $v$ is absolutely continuous. Denoting by $\|K\|_\infty$ the $L^\infty$ norm of $K$ we have

$$|v(t)| \leq \|K\|_\infty \int_0^t |v(s)|ds, \qquad t \in [0,1]. \tag{8.53}$$

Iterating this inequality

$$\begin{aligned} |v(t)| &\leq \|K\|_\infty^n \int_0^t \int_0^{s_1} \cdots \int_0^{s_{n-1}} |v(s_n)| ds_n \cdots ds_1 \\ &= \frac{\|K\|_\infty^n}{(n-1)!} \int_0^t (t - s_n)^{n-1} |v(s_n)| ds_n \leq \frac{\|K\|_\infty^n}{(n-1)!} \|v\|_{L^1}, \end{aligned}$$

where we used Fubini theorem and the fact that, given $0 \leq s_n \leq t \leq 1$, one has

$$\text{meas}\{(s_1, \dots, s_{n-1}) : s_n \leq s_{n-1} \leq \dots \leq s_1 \leq t\} = \frac{(t - s_n)^{n-1}}{(n-1)!}.$$

By integrating and using $\|v\|_{L^1} \leq \|v\|_{L^2}$, one obtains in particular the $L^2$ estimate

$$\|v\|_{L^2} \leq \frac{\|K\|_\infty^n}{(n-1)!} \|v\|_{L^2}, \tag{8.54}$$

which, for $n$ large enough, implies $v = 0$.

$\square$

Combining the last result with Proposition 8.25 we obtain the following corollary.

**Corollary 8.33.** *The manifold of Lagrange multilpliers of the sub-Riemannian problem* $(E, J)$

$$\mathcal{L}_{(E,J)} := \{\lambda_1 \in T^*M \,|\, \lambda_1 = e^{\vec{H}}(\lambda_0),\, \lambda_0 \in T^*_{q_0}M\}$$

*is a smooth $n$-dimensional submanifold of $T^*M$.*

## 8.6 Exponential map and Gauss' Lemma

A key object in sub-Riemannian geometry is the exponential map, that is the map that associates normal extremals with their initial covectors.

**Definition 8.34.** Let $q_0 \in M$. The *sub-Riemannian exponential map* (based at $q_0$) is the map

$$\exp_{q_0} : \mathscr{A}_{q_0} \subset T^*_{q_0}M \to M, \qquad \exp_{q_0}(\lambda_0) = \pi \circ e^{\vec{H}}(\lambda_0). \tag{8.55}$$

defined on the domain $\mathscr{A}_{q_0}$ of covectors such that the corresponding solution of the Hamiltonian system is defined on the interval $[0, 1]$. When there is no confusion on the base point, we might use the simplified notation exp.

The homogeneity of the sub-Riemannian Hamiltonian $H$ yields the following homogeneity property of the flow associated with $\vec{H}$.

**Lemma 8.35.** *Let $H$ be the sub-Riemannian Hamiltonian. Then, for every $\lambda \in T^*M$ we have*

$$e^{t\vec{H}}(\alpha\lambda) = \alpha e^{\alpha t\vec{H}}(\lambda), \tag{8.56}$$

*for any $\alpha > 0$ and $t > 0$ such that both sides of the identity are defined.*

*Proof.* By Remark 4.26 we know that if $\lambda(t) = e^{t\vec{H}}(\lambda_0)$ is a solution of the Hamiltonian system associated with $H$, then also $\lambda_\alpha(t) := \alpha\lambda(\alpha t)$ is a solution. The identity (8.56) follows from the uniqueness of the solution and the fact that $\lambda_\alpha(0) = \alpha\lambda(0)$. $\square$

The homogeneity property (8.56) permits to recover the whole extremal trajectory as the image of the ray joining 0 to $\lambda_0$ in the fiber $T^*_{q_0}M$.

**Corollary 8.36.** *Let $\lambda(t)$, for $t \in [0,T]$, be the normal extremal that satisfies the initial condition*

$$\lambda(0) = \lambda_0 \in T^*_{q_0}M.$$

*Then the normal extremal path $\gamma(t) = \pi(\lambda(t))$ satisfies*

$$\gamma(t) = \exp_{q_0}(t\lambda_0), \qquad t \in [0,T].$$

*Proof.* Using (8.56) we get

$$\exp_{q_0}(t\lambda_0) = \pi(e^{\vec{H}}(t\lambda_0)) = \pi(e^{t\vec{H}}(\lambda_0)) = \pi(\lambda(t)) = \gamma(t).$$

□

*Remark* 8.37 (Unit speed normal extremals). Thanks to the homogeneity property one can introduce the cylinder $\Lambda_{q_0}$ of normalized covectors

$$\Lambda_{q_0} = \{\lambda \in T^*_{q_0}M |\ H(\lambda) = 1/2\},$$

and consider the exponential map as follows (notice the two arguments)

$$\exp_{q_0} : \mathbb{R}^+ \times \Lambda_{q_0} \to M, \qquad \exp(t, \lambda_0) := \exp_{q_0}(t\lambda_0)$$

In other words one restricts to length parametrized extremal paths, considering the time as an extra variable. In what follows, with an abuse of notation, we set

$$\exp^t_{q_0}(\lambda_0) := \exp_{q_0}(t\lambda_0), \qquad \lambda_0 \in \Lambda_{q_0},$$

whenever the right hand side is defined.

**Proposition 8.38.** *If the metric space $(M,d)$ is complete, then $\mathscr{A}_{q_0} = T^*_{q_0}M$. Moreover, if there are no strictly abnormal length-minimizers, the exponential map $\exp_{q_0}$ is surjective.*

*Proof.* To prove that $\mathscr{A}_{q_0} = T^*_{q_0}M$, it is enough to show that any normal extremal $\lambda(t)$ starting from $\lambda_0 \in T^*_{q_0}M$ with $H(\lambda_0) = 1/2$ is defined for all $t \in \mathbb{R}$. Assume that the extremal $\lambda(t)$ is defined on $[0,T[$, and assume that it is not extendable to some interval $[0,T+\varepsilon[$. The projection $\gamma(t) = \pi(\lambda(t))$ defined on $[0,T[$ is a curve with unit speed, thus for any sequence $t_j \to T$ the sequence $(\gamma(t_j))_j$ is a Cauchy sequence on $M$ since

$$d(\gamma(t_i), \gamma(t_j)) \leq |t_i - t_j|.$$

The sequence $(\gamma(t_j))_j$ is then convergent to a point $q_1 \in M$ by completeness. Let us now consider coordinates around the point $q_1$ and show that, in coordinates $\lambda(t) = (p(t), x(t))$, the curve $p(t)$ is uniformly bounded. This contradicts the fact that $\lambda(t)$ is not extendable. By Hamilton equations (4.38)

$$\dot{p}(t) = -\frac{\partial H}{\partial x}(p(t), x(t)) = -\sum_{i=1}^{m} \langle p(t), f_i(\gamma(t))\rangle \langle p(t), D_x f_i(\gamma(t))\rangle .$$

Since $H(\lambda(t)) = \frac{1}{2}\sum_{i=1}^{m} \langle p(t), f_i(\gamma(t))\rangle^2 = 1/2$ then $|\langle p(t), f_i(\gamma(t))\rangle| \leq 1$ for every $i = 1, \ldots, m$. Moreover by smoothness of $f_i$, the derivatives $|D_x f_i| \leq C$ are locally bounded and one gets the inequality

$$|\dot{p}(t)| \leq C|p(t)|,$$

which by Gronwall's lemma implies that $|p(t)|$ is uniformly bounded on a bounded interval. The second part of the statement follows from the existence of length-minimizers on a complete sub-Riemannian manifold, cf. Proposition 3.47 and Corollary 3.49. □

**Corollary 8.39.** *If the metric space $(M, d)$ is complete, then every normal extremal trajectory is extendable on* $[0, +\infty[$.

Next, we discuss an elementary but important observation on the behavior of the exponential map in a neighborhood of zero.

**Proposition 8.40.** *The sub-Riemannian exponential map* $\exp_{q_0} : T^*_{q_0}M \to M$ *is a local diffeomorphism at* $0$ *if and only if* $\mathcal{D}_{q_0} = T_{q_0}M$. *More precisely* $\mathrm{im}\,(D_0\exp_{q_0}) = \mathcal{D}_{q_0}$.

*Proof.* Fix any element $\xi \in T^*_{q_0}M$. By definition of differential

$$D_0\exp_{q_0}(\xi) = \frac{d}{dt}\bigg|_{t=0} \exp_{q_0}(0 + t\xi) = \frac{d}{dt}\bigg|_{t=0} \gamma_\xi(t) = \dot{\gamma}_\xi(0). \tag{8.57}$$

where $\gamma_\xi$ is the horizontal curve associated with initial covector $\xi \in T^*_{q_0}M$. This proves that $\mathrm{im}\, D_0\exp_{q_0} = \mathcal{D}_{q_0}$. To prove the equality let us notice that from (4.39) one has

$$\dot{\gamma}_\xi(0) = \sum_{i=1}^{m} \langle \xi, f_i(q_0) \rangle \, f_i(q_0). \tag{8.58}$$

Since $\xi \in T^*_{q_0}M$ is arbitrary, the proof is completed. $\square$

*Remark* 8.41. In the Riemannian case $\exp_{q_0}$ gives local coordinates to $M$ around $q_0$, being a diffeomorphism of a small ball in $T^*_{q_0}M$ onto a small geodesic ball in $M$, where geodesics are images of straight lines in the cotangent space. Moreover there is a unique length-minimizer joining $q_0$ to every point of the (sufficiently small) ball and the distance from $q_0$ is a smooth function in a neighborhood of $q_0$ itself.

This is no more true as soon as $\mathcal{D}_{q_0} \neq T_{q_0}M$ and, as we will show in Corollary 11.6 and Theorem 12.17, singularities appear naturally.

We end this section with a Hamiltonian version of Gauss' Lemma. Recall that the sub-Riemannian Hamiltonian $H : T^*M \to \mathbb{R}$ is fiberwise-quadratic. We denote by the same symbol $H$ the symmetric bilinear form associated with it. If $f_1, \ldots, f_m$ is a generating family and $\lambda, \eta \in T^*_qM$ then we have

$$H(\lambda, \eta) = \frac{1}{2} \sum_{i=1}^{m} \langle \lambda, f_i(q) \rangle \langle \eta, f_i(q) \rangle \, . \tag{8.59}$$

**Proposition 8.42** (Cotangent Gauss' Lemma)**.** *Let $q_0 \in M$, $\lambda_0 \in T^*_{q_0}M$ and set $\lambda_1 := e^{\vec{H}}(\lambda_0)$. Then for every $w \in T^*_{q_0}M \simeq T_{\lambda_0}(T^*_{q_0}M)$ one has the identity*

$$\big\langle \lambda_1, D_{\lambda_0} \exp_{q_0}(w) \big\rangle = 2H(\lambda_0, w). \tag{8.60}$$

*Proof.* Let us consider a smooth variation $\eta^s \in T^*_{q_0}M$, for $s \in (-\varepsilon, \varepsilon)$, of initial covectors such that $\eta^0 = \lambda_0$ and $\frac{d}{ds}\big|_{s=0} \eta^s = w$.

For $t \in [0, 1]$, let $\eta^s(t) := e^{t\vec{H}}(\eta^s)$ and $\gamma^s(t) = \pi(\eta^s(t))$ be the corresponding trajectory. Define the family of controls $u^s(\cdot)$ satisfying for a.e. $t \in [0, 1]$

$$u^s_i(t) := \langle \eta^s(t), f_i(\gamma^s(t)) \rangle \, , \qquad i = 1, \ldots, m, \tag{8.61}$$

where $f_1, \dots, f_m$ denotes as usual a generating family. By definition (8.61) of $u^s$ we have $\exp_{q_0}(\eta^s) = E_{q_0}(u^s)$, hence we can compute

$$D_{\lambda_0} \exp_{q_0}(w) = \left.\frac{d}{ds}\right|_{s=0} \exp_{q_0}(\eta^s) = \left.\frac{d}{ds}\right|_{s=0} E_{q_0}(u^s) = D_u E_{q_0}(v), \tag{8.62}$$

where we denoted $v := \frac{d}{ds}\big|_{s=0} u^s$. We have

$$\left\langle \lambda_1, \left.\frac{d}{ds}\right|_{s=0} \exp_{q_0}(\eta^s) \right\rangle = \langle \lambda_1, D_u E_{q_0}(v) \rangle = (u, v)_{L^2}, \tag{8.63}$$

where the second identity follows from the condition (8.18). On the other hand recall that, thanks to Theorem 4.25, for every fixed $s \in [0, 1]$ the sum of the squares of the quantities in (8.61) is constant with respect to $t \in [0, 1]$, and

$$\frac{1}{2} \int_0^1 \sum_{i=1}^m u_i^s(t)^2 dt = \frac{1}{2} \sum_{i=1}^m \langle \eta^s(t), f_i(\gamma^s(t)) \rangle^2 = H(\eta^s(t)).$$

Differentiating the last identity at $s = 0$ one gets

$$(u, v)_{L^2} = \int_0^1 \sum_{i=1}^m u_i(t) v_i(t) dt = 2H \left( \eta^0(t), \left.\frac{d}{ds}\right|_{s=0} \eta^s(t) \right) = 2H(\lambda_0, w)$$

where the last equality follows from the fact that $t \mapsto \frac{\partial}{\partial s} H(\eta^s(t))$ is also constant in $t$ (hence we evaluate at $t = 0$). $\square$

An immediate corollary is that the final covector of a normal extremal trajectory annihilates the tangent space to the front (assuming the front is locally a smooth manifold). More precisely we have the following.

**Corollary 8.43.** *Fix $q_0 \in M$. Let $\lambda_0 \in \Lambda_{q_0}$ that is not a critical point for $\exp_{q_0}$. Let $U$ be a small neighborhood of $\lambda_0 \in \Lambda_{q_0}$ and set $\mathscr{F} := \exp_{q_0}(U)$. Then $\lambda_1 := e^{\vec{H}}(\lambda_0)$ annihilates the tangent space $T_q\mathscr{F}$ to $\mathscr{F}$ at $q := \exp_{q_0}(\lambda_0)$.*

**Exercise 8.44.** Deduce from Corollary 8.43 and the homogeneity property of the Hamiltonian that if $\lambda_0 \in \Lambda_{q_0}$ is not a critical point for $\exp_{q_0}^t$, then $\lambda_t := e^{t\vec{H}}(\lambda_0)$ annihilates the tangent space $T_{q_t}\mathscr{F}_t$ to $\mathscr{F}_t := \exp_{q_0}^t(U)$ at $q_t := \exp_{q_0}^t(\lambda_0)$.

## 8.7 Conjugate points

In this section we introduce conjugate points and we discuss a basic result on the structure of the set of conjugate points along an extremal trajectory. Recall that given $q_0 \in M$ we denote by $\Lambda_{q_0} = \{\lambda \in T_{q_0}^*M \mid H(\lambda) = 1/2\}$, where $H$ is the sub-Riemannian Hamiltonian.

**Definition 8.45.** Fix $q_0 \in M$. A point $q \in M$ is *conjugate* to $q_0$ if there exists $s > 0$ and $\lambda_0 \in \Lambda_{q_0}$ such that $q = \exp_{q_0}(s\lambda_0)$ and $s\lambda_0$ is a critical point of $\exp_{q_0}$.

In this case we say that $q$ is *conjugate to* $q_0$ along $\gamma(t) = \exp_{q_0}(t\lambda_0)$. Moreover we say that $q$ is the *first conjugate point* to $q_0$ along $\gamma(t) = \exp_{q_0}(t\lambda)$ if $q = \gamma(s)$ and $s = \inf\{\tau > 0 \mid \tau\lambda \text{ is a critical point of } \exp_{q_0}\}$.

We denote by $\mathrm{Con}_{q_0}$ the *conjugate locus* to $q_0$, that is the set of all first conjugate points to $q_0$ along some normal extremal trajectory starting from $q_0$.

*Remark* 8.46. Let $\gamma : [0,1] \to M$ be a normal extremal trajectory defined by $\gamma(t) = \exp_{q_0}(t\lambda_0)$. Notice that if $\gamma$ admits an abnormal lift, then $\gamma(1)$ is conjugate to $\gamma(0)$. Indeed by definition of abnormal, this means that the control $u$ associated with $\gamma$ is a critical point for $E_{q_0}$, i.e., the differential $D_u E_{q_0}$ is not surjective. Since, by definition of the exponential map, one has $\mathrm{im}\,(D_{\lambda_0}\exp_{q_0}) \subset \mathrm{im}\,(D_u E_{q_0})$, it follows that $D_{\lambda_0}\exp_{q_0}$ is not surjective as well.

Since the restriction of an abnormal extremal is still abnormal, Remark 8.46 implies that all points belonging to an abnormal segment are conjugate points. The following theorem discuss somehow a converse statement.

**Theorem 8.47.** *Let $\gamma : [0,T] \to M$ be a normal extremal path. Assume that there exists $t_0 > 0$ such that $\gamma(t_0)$ is a limit of a decreasing (resp. increasing) sequence of points that are conjugate to $\gamma(t_0)$ along $\gamma$. Then there exists $\varepsilon > 0$ such that*

*(a) for all $\tau \in [t_0, t+\varepsilon]$ (resp. $[t_0-\varepsilon, t_0]$), the point $\gamma(\tau)$ is conjugate to $\gamma(t_0)$ along $\gamma$,*

*(b) $\gamma|_{[t_0,t_0+\varepsilon]}$ (resp. $\gamma|_{[t_0-\varepsilon,t_0]}$) is an abnormal extremal path.*

*Proof.* We shall consider only the case of a decreasing convergent sequence and leave to the reader to make necessary modifications in the case of an increasing one.

Let $(u(t), \lambda(t))$, for $0 \le t \le T$, be a normal extremal, where $\gamma(t) = \pi(\lambda(t))$ and

$$\dot\gamma(t) = f_u(t)(\gamma(t)) = \sum_{i=1}^m u_i(t) f_i(\gamma(t)). \tag{8.64}$$

We set $P_{0,t} := \overrightarrow{\exp}\int_0^t f_{u(\tau)}\,d\tau$, and we consider the maps

$$\mathcal{F}_t : U \subset T^*M \to M, \qquad \mathcal{F}_t(\lambda) = \pi \circ P^*_{0,t} \circ e^{\vec H}(t\lambda), \tag{8.65}$$

defined on a neighborhood $U$ of $\lambda_0 = \lambda(0)$ in $T^*_{q_0}M$, where $q_0 = \gamma(0)$. According to this construction, we have $\mathcal{F}_t(\lambda(t)) = q_0$ for all $t$. We claim that, given $t \in (0,T]$, $\gamma(t)$ is conjugate to $\gamma(0)$ along $\gamma$ if and only if $\lambda_0$ is a critical point of the map $\mathcal{F}_t$. Indeed, according to the definition, $\gamma(t)$ is conjugate to $\gamma(0)$ if and only if $t\lambda_0$ is a critical point of the map $\exp_{q_0} = \pi \circ e^{\vec H}|_{T^*_{q_0}M}$, i.e., if $T_{\lambda(t)} e^{\vec H}(T^*_{q_0}M) \cap T_{\lambda(t)}(T^*_{\gamma(t)}M) \neq 0$, where the diffeomorphism $P^*_{0,t}$ maps $T^*_{\gamma(t)}M$ into $T^*_{q_0}M$.

Recall that $(P^*_{0,t})^{-1} = \overrightarrow{\exp}\int_0^t \vec h_{u(t)}\,dt$, where $h_u(\lambda) = \langle \lambda, f_u\rangle$ (cf. Chapter 4). The variations formula (see Section 6.5) and formula (4.86) imply that the depending on $t \in [0,T]$ family of diffeomorphisms

$$\lambda \mapsto P^*_{0,t} \circ e^{\vec H}(t\lambda) = tP^*_{0,t} \circ e^{t\vec H}(\lambda), \quad \lambda \in T^*M,$$

is the Hamiltonian flow generated by the time-dependent Hamiltonian $g_t : T^*M \to \mathbb{R}$ defined by

$$g_t := (H - h_{u(t)}) \circ (P^*_{0,t})^{-1}.$$

Since $H$ is the maximized Hamiltonian (cf. (4.34)), then $g_t$ is everywhere non-negative and $g_t(\lambda_0) = 0$. It follows that $d_{\lambda_0} g_t = 0$ and $d^2_{\lambda_0} g_t$ is a non-negative quadratic form on the symplectic space $T_{\lambda_0}(T^*M)$. We introduce the following notations:

$$\Sigma := T_{\lambda_0}(T^*M), \qquad \Pi := T_{\lambda_0}(T^*_{q_0}M), \qquad Q_t := \frac{1}{2} d^2_{\lambda_0} g_t. \tag{8.66}$$

The linear Hamiltonian flow $\overrightarrow{\exp} \int_0^t \vec{Q}_\tau \, d\tau$ on $\Sigma$ is the linearization of the flow $\overrightarrow{\exp} \int_0^t \vec{g}_\tau \, d\tau$ at the equilibrium $\lambda_0$. Moreover, $\gamma(t)$ is conjugate to $\gamma(0)$ if and only if

$$\Pi \cap J_t \neq 0, \qquad \text{where} \qquad J_t := \overrightarrow{\exp} \int_0^t \vec{Q}_\tau \, d\tau(\Pi).$$

Recall that Lagrange subspaces of the $2n$-dimensional symplectic space $\Sigma$ are $n$-dimensional subspaces on which the symplectic form $\sigma$ vanishes identically. In particular, $\Pi$ is a Lagrange subspace. $J_t$ is also a Lagrange subspace because symplectic flows preserve the symplectic form. A Darboux basis for $\Sigma$ is a basis $e_1, \dots, e_n, f_1, \dots, f_n$ satisfying

$$\sigma(e_i, f_j) = \delta_{ij}, \quad \sigma(f_i, f_j) = \sigma(e_i, e_j) = 0, \qquad i, j = 1, \dots, n. \tag{8.67}$$

We need the following lemma:

**Lemma 8.48.** *Let $\Lambda_0, \Lambda_1$ be Lagrange subspaces of $\Sigma$, with $\dim(\Lambda_0 \cap \Lambda_1) = k$. Then there exist a Darboux basis $e_1, \dots, e_n, f_1, \dots, f_n$ in $\Sigma$ such that*

$$\Lambda_0 = \operatorname{span}\{e_1, \dots, e_n\}, \quad \Lambda_1 = \operatorname{span}\{e_1, \dots, e_k, e_{k+1} + f_{k+1}, \dots, e_n + f_n\}.$$

*Proof.* Consider any arbitrary basis $e_1, \dots, e_n$ of $\Lambda_0$ satisfying

$$\Lambda_0 \cap \Lambda_1 = \operatorname{span}\{e_1, \dots, e_k\}.$$

The nondegeneracy of $\sigma$ implies the existence of $f_1 \in \Sigma$ such that

$$\sigma(e_1, f_1) = 1, \qquad \sigma(e_2, f_1) = \cdots = \sigma(e_n, f_1) = 0.$$

Fix such an $f_1$, the nondegeneracy of $\sigma$ implies the existence of $f_2 \in \Sigma$ such that

$$\sigma(e_2, f_2) = 1, \qquad \sigma(f_1, f_2) = \sigma(e_1, f_2) = \sigma(e_3, f_2) = \cdots = \sigma(e_n, f_2) = 0.$$

Iterating one obtains linearly independent $f_1, \dots, f_k$ such that

$$\sigma(e_i, f_j) = \delta_{ij}, \quad \sigma(f_i, f_j) = \sigma(e_l, f_j) = 0, \qquad i, j = 1, \dots, k, \ l = k+1, \dots, n.$$

Let us introduce the space

$$\Gamma = \{v \in \Lambda_1 \mid \sigma(f_1, v) = \cdots = \sigma(f_k, v) = 0\}.$$

By construction $\Lambda_1 = \Gamma \oplus (\Lambda_0 \cap \Lambda_1)$. The linear map $\Psi : \Gamma \to \mathbb{R}^{n-k}$ defined by

$$\Psi(v) := (\sigma(e_{k+1}, v), \dots, \sigma(e_n, v)),$$

is invertible, hence there exist $v_{k+1}, \dots, v_n \in \Gamma$ such that $\sigma(e_i, v_j) = \delta_{ij}$, for $i, j = k+1, \dots, n$. Setting $f_i := v_i - e_i$, for $i = k+1, \dots, n$, one obtains the Darboux basis $e_1, \dots, e_n, f_1, \dots, f_n$. $\square$

We apply the Lemma 8.48 to the pair of Lagrange subspaces $\Pi$ and $J_{t_0}$, working in the coordinates $(p, x) \in \mathbb{R}^n \times \mathbb{R}^n$ induced by the Darboux basis. We have:

$$J_{t_0} = \{(p, x) \in \mathbb{R}^n \times \mathbb{R}^n \mid x = S_{t_0} p\},$$

where $S_{t_0} = \begin{pmatrix} 0_k & 0 \\ 0 & I_{n-k} \end{pmatrix}$ is a non-negative symmetric matrix.

The subspace of $\Sigma = \{(p, x) \in \mathbb{R}^n \times \mathbb{R}^n\}$ defined by the equation $\{x = 0\}$ is called vertical and the one defined by the equation $\{p = 0\}$ is called horizontal. Any $n$-dimensional subspace $\Lambda$ close to $J_{t_0}$ is transversal to the horizontal subspace, and can be presented in the form $\Lambda = \{(p, Ap) : p \in \mathbb{R}^n\}$ for some $n \times n$-matrix $A$. Moreover, $\Lambda$ is a Lagrange subspace if and only if $A$ is a symmetric matrix. Indeed,

$$\sigma((p_1, Ap_1), (p_2, Ap_2)) = p_1^* A p_2 - p_2^* A p_1 = p_1^*(A - A^*)p_2.$$

where $v^*$ denotes the transpose of a vector $v$ (and similarly for matrices). Let $J_t = \{(p, S_t p) \mid p \in \mathbb{R}^n\}$ for $t$ close to $t_0$; then $S_t$ is a symmetric matrix smoothly depending on $t$. Moreover, one has

$$\Pi \cap J_t = \{(p, 0) \in \mathbb{R}^n \times \mathbb{R}^n \mid S_t p = 0\}.$$

**Lemma 8.49.** *For every $p \in \mathbb{R}^n$ one has $p^* \dot{S}_t p \geq 0$.*

*Proof.* Recall that $Q_t$ is a non-negative quadratic form on $\Sigma$. We denote by the same symbol the matrix respresenting $Q_t$ in coordinates. Recall that $\vec{Q}_t$ is the vector field on $\Sigma \simeq \mathbb{R}^{2n}$ associated with $Q_t$. Let $t \mapsto \lambda_t$ be a solution of the equation $\dot{\lambda}_t = \vec{Q}_t \lambda_t$; then

$$\sigma(\lambda_t, \dot{\lambda}_t) = \sigma(\lambda_t, \vec{Q}_t \lambda_t) = 2\langle Q_t \lambda_t, \lambda_t \rangle \geq 0.$$

We apply this inequality to $\lambda_t = (p_t, S_t p_t)$ and obtain:

$$\sigma((p, S_t p), (\dot{p}, S_t \dot{p}) + (0, \dot{S}_t p)) = p^* \dot{S}_t p \geq 0.$$ □

**Lemma 8.50.** *If $S_{t_1} \bar{p} = 0$ for some $t_1 > t_0$ and $\bar{p} \in \mathbb{R}^n$, then $S_t \bar{p} = 0$, $\forall t \in [t_0, t_1]$.*

*Proof.* This statement is an easy corollary of Lemma 8.49. Indeed,

$$0 \leq \bar{p}^* S_{t_0} \bar{p} \leq \bar{p}^* S_t \bar{p} \leq \bar{p}^* S_{t_1} \bar{p} = 0.$$

Hence $\bar{p}^* S_t \bar{p} = 0$. Since $p \mapsto p^* S_t p$ is a non-negative quadratic form, we obtain that $S_t \bar{p} = 0$. □

Lemma 8.50 implies claim (a) of the theorem. Let us prove claim (b), whose proof is also based on Lemma 8.50.

The fiber $T_{q_0}^* M$ is a vector space, it is naturally identified with its tangent space $\Pi$, and the coordinates $p \in \mathbb{R}^n$ on $\Pi$ introduced above serve as coordinates on $T_{q_0}^* M$. The restriction of the Hamiltonian $g_t$ to $T_{q_0}^* M$ has the form:

$$g_t(p) = \frac{1}{2} \sum_{i=1}^{m} \langle p, ((P_{0,t}^{-1})_* f_i)(q_0) \rangle^2 - \langle p, ((P_{0,t}^{-1})_* f_{u(t)})(q_0) \rangle.$$

Hence, we have

$$\langle Q_t(p, 0), (p, 0) \rangle = \frac{1}{2} \sum_{i=1}^{m} \langle p, (P_{0,t*}^{-1} f_i)(q_0) \rangle^2. \tag{8.68}$$

Moreover, if $s \mapsto \lambda_s = (p_s, x_s)$ is a solution of the system $\dot{\lambda} = \vec{Q}_\tau \lambda$, and $x_t = 0$, then $\langle p, \dot{x}_t \rangle = \langle (p, 0), Q_t(p_t, 0) \rangle$, for all $p \in \mathbb{R}^n$. In particular, under conditions of Lemma 8.50, we get:

$$\langle (\bar{p}, 0), Q_t(\bar{p}_t, 0) \rangle = 0, \quad t \in [t_0, t_1],$$

and, according to the identity (8.68),

$$\langle \bar{p}, (P_{0,t*}^{-1} f_i)(q_0) \rangle = 0, \quad i = 1, \dots, k, \quad t \in [t_0, t_1].$$

Let $\eta(t) = (P_{0,t}^*)^{-1}(\bar{p}, q_0) \in T_{\gamma(t)}^* M$. We obtain that $(u(t), \eta(t))$ for $t \in [t_0, t_1]$ is an abnormal extremal, thanks to characterization of Proposition 8.9. $\square$

We deduce from Theorem 8.47 the following important corollary.

**Corollary 8.51.** *Let $\gamma : [0, 1] \to M$ be a normal extremal trajectory that does not contain abnormal segments. Define the set of conjugate times to zero*

$$\mathcal{T}_c := \{t > 0 \mid \gamma(t) \textit{ is conjugate to } \gamma(0)\}.$$

*Then the set $\mathcal{T}_c$ is discrete.*

## 8.8 Minimizing properties of extremal trajectories

In this section we study the relation between conjugate points and length-minimality properties of extremal trajectories. The space of horizontal trajectories on $M$ can be endowed with two different topologies:

- the $W^{1,2}$ topology, also called weak topology. This is the topology induced on the space of horizontal trajectories by the $L^2$ topology on the space of controls,
- the $C^0$ topology, also called strong topology. This is the usual uniform topology on the space of continuous curves on $M$.

The main result of this section is the following.

**Theorem 8.52.** *Let $\gamma : [0, 1] \to M$ be a normal extremal trajectory that does not contain abnormal segments. Then,*

*(i)* $t_c := \inf\{t > 0 \mid \gamma(t)$ *is conjugate to* $\gamma(0)\} > 0$.

*(ii) for every $\tau < t_c$ the curve $\gamma|_{[0,\tau]}$ is a local length-minimizer in the $W^{1,2}$ topology among horizontal trajectories with same endpoints.*

*(iii) for every $\tau > t_c$ the curve $\gamma|_{[0,\tau]}$ is not a length-minimizer.*

Claim (i) of Theorem 8.52 is a direct consequence of Corollary 8.51. Nevertheless we will give in this section an independent proof.

The proof of part (ii) and (iii) need some preliminary results. Some of these preliminary results hold true under weaker assumptions. For simplicity, in this section, we state them under the assumption that $\gamma$ does not contain abnormal segments. See Exercice 8.56 for more general assumptions.

The proof of Theorem 8.52 is then contained in Section 8.8.1. We conclude this discussion by stating explicitly the following consequence of Theorem 8.52.

**Corollary 8.53.** *Let $\gamma : [0, 1] \to M$ be a normal extremal trajectory that does not contain abnormal segments. Assume that the trajectory does not contain conjugate points. Then $\gamma$ is a local miminum for the length with respect to the $W^{1,2}$ topology on the space of admissible trajectories with the same endpoints.*

### 8.8.1 Local length-minimality in the $W^{1,2}$ topology. Proof of Theorem 8.52.

Let $s \in (0,1]$. Given a normal extremal trajectory $\gamma_u : [0,1] \to M$, let us denote by $u^s(t) := su(st)$ the reparametrized control associated with the reparametrized trajectory $\gamma^s(t) := \gamma_u(st)$, both defined for $t \in [0,1]$. Notice that if $\lambda \in T^*_{\gamma_u(1)}M$ is a Lagrange multiplier associated with $u$, then $\lambda^s = s(P^*_{s,1})\lambda \in T^*_{\gamma_u(s)}M$, is a Lagrange multiplier associated with $u^s$.

The first result concerns the characterisation of conjugate points through the second variation of the energy.

**Proposition 8.54.** *Let $s \in (0,1]$ and assume that the normal extremal trajectory $\gamma_u : [0,1] \to M$ contains no abnormal segments. Then $\gamma_u(s)$ is conjugate to $\gamma_u(0)$ along $\gamma_u$ if and only if $\mathrm{Hess}_{u^s}\left(J|_{E_{q_0}^{-1}(\gamma^s(1))}\right)$ is a degenerate quadratic form, where $q_0 = \gamma_u(0)$.*

*Proof.* Since the curve $\gamma_u$ contains no abnormal segments, the control $u^s(t) = su(st)$ is a regular point for the end-point map. Hence, thanks to Proposition 8.27 combined with Proposition 8.31 and Corollary 8.33, one has that $\gamma_u(s)$ is conjugate to $\gamma_u(0)$ if and only if $\lambda^s$ is a critical point of the exponential map, that is equivalent to the fact that $\mathrm{Hess}_{u^s}\left(J|_{E_{q_0}^{-1}(\gamma^s(1))}\right)$ is degenerate. $\square$

The following lemma, on the family of quadratic forms $s \mapsto \mathrm{Hess}_{u^s}\left(J|_{E_{q_0}^{-1}(\gamma^s(1))}\right)$, is crucial in what follows.

**Lemma 8.55.** *Assume that a normal extremal trajectory $\gamma_u : [0,1] \to M$ contains no abnormal segments. Define the function $\alpha : (0,1] \to \mathbb{R}$ as follows*

$$\alpha(s) := \inf\left\{\|v\|^2_{L^2} - \left\langle \lambda^s, D^2_{u^s}E_{q_0}(v,v)\right\rangle \mid \|v\|^2_{L^2} = 1,\ v \in \ker D_{u^s}E_{q_0}\right\}. \tag{8.69}$$

*Then $\alpha$ is continuous and has the following properties:*

*(a)* $\alpha(0) := \lim_{s\to 0}\alpha(s) = 1$*;*

*(b)* $\alpha(s) = 0$ *implies that* $\mathrm{Hess}_{u^s} J\big|_{E_{q_0}^{-1}(\gamma^s(1))}$ *is degenerate;*

*(c)* $\alpha$ *is monotone decreasing;*

*(d) if* $\alpha(\bar{s}) = 0$ *for some* $\bar{s} > 0$*, then* $\alpha(s) < 0$ *for* $s > \bar{s}$*.*

*Proof of Lemma 8.55.* We start with some preliminary observations. Notice that one can write

$$\|v\|^2_{L^2} - \lambda^s \circ D^2_{u^s}E_{q_0}(v,v) = \langle (I - Q_s)(v)|v\rangle_{L^2}, \tag{8.70}$$

where $Q_s : L^2([0,1],\mathbb{R}^m) \to L^2([0,1],\mathbb{R}^m)$ is a compact and symmetric operator.

Let us prove that the infimum in (8.69) is always attained. Indeed, fix $s \in (0,1]$ and let $v_n \in \ker D_{u^s}E_{q_0}$ be a sequence with $\|v_n\| = 1$ and such that $\|v_n\|^2_{L^2} - \langle Q_s(v_n)|v_n\rangle_{L^2} \to \alpha(s)$ for $n \to \infty$. Since the unit ball is weakly compact in $L^2$, up to extraction of a sub-sequence, we have that $v_n$ is weakly convergent to some $\bar{v}$, with $\|\bar{v}\|^2_{L^2} \le 1$. By compactness of $Q_s$, we deduce that $\langle Q_s(\bar{v})|\bar{v}\rangle_{L^2} = 1 - \alpha(s)$. Then

$$\|\bar{v}\|^2_{L^2} - \langle Q_s(\bar{v})|\bar{v}\rangle_{L^2} \le 1 - (1 - \alpha(s)) = \alpha(s). \tag{8.71}$$

Since $\alpha(s)$ is the infimum, it follows that $\|\bar{v}\|^2_{L^2} = 1$ and we have the equality in (8.71).

Observe now that since every restriction $\gamma|_{[0,s]}$ is not abnormal, the rank of $D_{u^s}E_x$ is maximal, equal to $n$, for all $s \in (0,1]$. Then, by Riesz representation Theorem, we find a continuous orthonormal basis $\{v_i^s\}_{i\in\mathbb{N}}$ for $\ker D_{u^s}E_x$, yielding a continuous one-parameter family of isometries $\phi_s : \ker D_{u^s}E_x \to \mathcal{H}$ on a fixed Hilbert space $\mathcal{H}$. Since also $s \mapsto Q_s$ is continuous (in the operator norm topology), we reduce (8.69) to

$$\alpha(s) = 1 - \sup\{\langle \phi_s \circ Q_s \circ \phi_s^{-1}(w)|w\rangle_{\mathcal{H}} \mid w \in \mathcal{H},\ \|w\|_{\mathcal{H}} = 1\}, \tag{8.72}$$

where the composition $\tilde{Q}_s := \phi_s \circ Q_s \circ \phi_s^{-1}$ is a continuous one-parameter family of symmetric and compact operators on a fixed Hilbert space $\mathcal{H}$. The supremum coincides with the largest eigenvalue of $\tilde{Q}_s$, which is well known to be continuous as a function of $s$ if $\tilde{Q}_s$ is (see [Kat95, V Thm. 4.10]). This proves that $\alpha$ is continuous.

Let us recall the formulas for the first and second differentials

$$D_{u^s}E_{q_0}(v) = \int_0^s (P_{t,1})_* f_{v(t)}|_{\gamma_u(s)} dt, \tag{8.73}$$

$$D^2_{u^s}E_{q_0}(v,v) = \iint\limits_{0\le\tau\le t\le s} [(P_{\tau,1})_* f_{v(\tau)}, (P_{t,1})_* f_{v(t)}]|_{\gamma_u(s)} d\tau dt. \tag{8.74}$$

Recalling that $u^s = su(s\cdot)$, by a change of variables one can see that

$$D_{u^s}E_{q_0}(v) = s\int_0^1 (P_{st,1})_* f_{v(st)}|_{\gamma_u(s)} dt, \tag{8.75}$$

$$D^2_{u^s}E_{q_0}(v,v) = s^2 \iint\limits_{0\le\tau\le t\le 1} [(P_{s\tau,1})_* f_{v(s\tau)}, (P_{st,1})_* f_{v(st)}]|_{\gamma_u(s)} d\tau dt. \tag{8.76}$$

Taking the limit $s \to 0$, one can show that $Q_s \to 0$, hence $\tilde{Q}_s \to 0$, proving (a).

To prove (b), notice that $\alpha(\bar{s}) = 0$ means that $I - Q_{\bar{s}} \geq 0$, and since the infimum is attained there exists $\bar{v}$ of norm one such that $\langle (I - Q_{\bar{s}})(\bar{v})|\bar{v}\rangle_{L^2} = 0$. Being $I - Q_{\bar{s}}$ a bounded, non-negative symmetric operator, and since $\bar{v} \neq 0$, this implies that $I - Q_{\bar{s}}$ is degenerate. [2]

To prove (c) let us fix $0 \le s \le s' \le 1$ and $v \in \ker D_{u^s}E_x$. Define

$$\widehat{v}(t) := \begin{cases} \sqrt{\dfrac{s'}{s}}\, v\left(\dfrac{s'}{s}t\right), & 0 \le t \le \dfrac{s}{s'}, \\ 0, & \dfrac{s}{s'} < t \le 1. \end{cases}$$

It follows that $\|\widehat{v}\|^2_{L^2} = \|v\|^2_{L^2}$, $\widehat{v} \in \ker D_{u^{s'}}E_x$, and $D^2_{u^s}E_x(v) = D^2_{u^{s'}}E_x(\widehat{v})$. As a consequence, $\alpha(s) \geq \alpha(s')$.

To prove (d), assume by contradiction that there exists $s_1 > \bar{s}$ such that $\alpha(s_1) = 0$. By monotonicity of point (c), $\alpha(s) = 0$ for every $\bar{s} \le s \le s_1$. This implies that every point in the image of $\gamma|_{[\bar{s},s_1]}$ is conjugate to $\gamma(0)$. Arguing as in the proof of Theorem 8.47, the segment $\gamma|_{[\bar{s},s_1]}$ is also abnormal, contradicting the assumption on $\gamma$. $\square$

[2] Let $V$ be a vector space and $\mathcal{Q} : V \times V \to \mathbb{R}$ be a quadratic form on $V$. Recall that $\mathcal{Q}$ is degenerate if there exists a non-zero $\bar{v} \in V$ such that $\mathcal{Q}(\bar{v}, \cdot) = 0$. Moreover a non negative quadratic form is degenerate if and only if there exists $\bar{v} \neq 0$ such that $\mathcal{Q}(\bar{v}, \bar{v}) = 0$.

*Proof of Theorem 8.52.* Thanks to Lemma 8.55 there exists $\varepsilon > 0$ such that $\alpha(s) > 0$ on the segment $[0, \varepsilon]$. This implies that this segment does not contain conjugate points thanks to Proposition 8.54. This proves claim (i).

To prove claim (ii) notice that if $\gamma|_{[0,s]}$ does not contain conjugate points, by Proposition 8.54 it follows that $\mathrm{Hess}_{u^s} J\big|_{E^{-1}(\gamma^s(1))}$ is non degenerate for every $s \in [0, \tau]$, hence $\mathrm{Hess}_{u^\tau} J\big|_{E^{-1}(\gamma^\tau(1))} > 0$ using items (b) and (c) of Lemma 8.55.

Let $\tau > t_c$ and assume by contradiction that the trajectory is a length-minimizer. Then, using the terminology of Lemma 8.55, one has $\alpha(t_c) = 0$ and $\alpha(\tau) < 0$ thanks to properties (c) and (d). This implies that the Hessian has a negative eigenvalue, hence we can find a variation joining the same end-points and shorter than the original geodesic, contradicting the minimality assumption. □

**Exercise 8.56.** Introduce the following definitions: a normal extremal trajectory $\gamma : [0, 1] \to M$ is said to be

- *left strongly normal*, if for every $s \in (0, 1]$ the curve $\gamma|_{[0,s]}$ does not admit abnormal lifts.
- *right strongly normal*, if for every $s \in [0, 1)$ the curve $\gamma|_{[s,1]}$ does not admit abnormal lifts.
- *strongly normal*, if $\gamma$ is both left and right strongly normal.

Prove that a normal extremal trajectory $\gamma : [0, 1] \to M$ does not contain abnormal segments if and only if $\gamma|_{[0,\tau]}$ is strongly normal for every $\tau \in [0, 1]$.

Prove that Theorem 8.52 claim (i)-(ii), Proposition 8.54 and claims (a)-(b)-(c) of Lemma 8.55 hold under the weaker assumption that the normal extremal trajectory $\gamma$ is left strongly normal.

### 8.8.2 Local length-minimality in the $C^0$ topology

In the previous section we proved, among other results, that a normal extremal trajectory that does not contain abnormal segments is a local miminum for the length with respect to the $W^{1,2}$ topology on the space of admissible trajectories with the same endpoints.

The goal of this section is to prove that the same conclusion holds true, but with respect to the uniform topology. The proof of this result, which is based upon the arguments of Theorem 4.62, requires a preliminary discussion on the free endpoint problem.

#### Free initial point problem

In all our previous discussions the initial point $q_0 \in M$ has always been fixed from the very beginning. Clearly, if the initial point $q_0$ is not fixed, and given a final point $q_1 \in M$, the minimization problem

$$\min_{q \in M, u \in E_q^{-1}(q_1)} J(u), \tag{8.77}$$

has only the trivial solution $(q, u) = (q_1, 0)$.

In this case, it is convenient to introduce a *penalty function* $a \in C^\infty(M)$, and consider the minimization problem

$$\min_{q \in M, u \in E_q^{-1}(q_1)} \left(J(u) + a(q)\right). \tag{8.78}$$

Let us introduce the *extendend end-point map*

$$\mathbb{E} : M \times \mathcal{U} \to M, \qquad (q, u) \mapsto E_q(u),$$

where $E_q$ is the end-point map based at $q$. It is not restrictive to assume that $\mathcal{U}$ is a fixed open set on the Hilber space $L^2([0,1], \mathbb{R}^m)$. Notice that $\mathbb{E}$ is a submersion at every point. First, notice that for every $q \in M$, one has $\mathbb{E}(q, 0) = q$. Moreover, denoting by $P^u_{t,s}$ the non-autonomous flow associated with $u$, one has

$$\mathbb{E}\big|_{\{q_0\}\times\mathcal{U}} = E_{q_0}, \qquad \mathbb{E}\big|_{M\times\{u\}} = P^u_{0,1}. \tag{8.79}$$

The minimization problem (8.78) is then rewritten as

$$\min_{\mathbb{E}^{-1}(q_1)} \varphi \tag{8.80}$$

where $\varphi : M \times \mathcal{U} \to \mathbb{R}$ is defined by $\varphi(q, u) := J(u) + a(q)$. This constrained minimization problem is of the type studied in Section 8.4, with $F = \mathbb{E}$.[3]

Notice that every level set $\mathbb{E}^{-1}(q_1)$ is regular since the map $\mathbb{E}$ is a submersion. The Lagrange multipliers rule (Proposition 8.13) is rewritten as follows: if the point $(q_0, u) \in M \times \mathcal{U}$ is a solution of (8.78) (or, equivalently, (8.80)), then there exists $\lambda_1 \in T^*M$ such that

$$\lambda_1 D_{(q_0,u)}\mathbb{E} = D_{(q_0,u)}(J + a). \tag{8.81}$$

The differentials $D_{(q_0,u)}\mathbb{E}$ and $D_{(q_0,u)}(J + a)$ are defined on the product space $T_{(q_0,u)}(M \times \mathcal{U}) \simeq T_{q_0}M \times L^2([0, T], \mathbb{R}^m)$. Thanks to the identities

$$D_{(q_0,u)}\mathbb{E} = (D_u E_{q_0}, (P^u_{0,1})_*), \qquad D_{(q_0,u)}(J + a) = (D_u J, d_{q_0} a),$$

equation (8.81) splits into the following system

$$\begin{cases} \lambda_1 D_u E_{q_0} = D_u J = u \\ \lambda_1 (P^u_{0,1})_* = d_{q_0} a \end{cases}$$

In other words, with every solution of the problem (8.80) we can associate a normal extremal

$$\lambda(t) = (P^{-1}_{0,t})^* \lambda_0,$$

where the initial condition is defined by the formula $\lambda_0 = d_{q_0} a$.

**Proposition 8.57.** *To every pair $(q_0, u) \in M \times \mathcal{U}$ that is a solution of the problem* (8.80) *we can associate an horizontal trajectory $\gamma_u(t)$ that is a normal extremal trajectory associated with initial covector $\lambda_0 = d_{q_0} a$, namely $\gamma(t) = \exp_{q_0}(t d_{q_0} a)$ for $t \in [0, 1]$.*

We end this subsection with an analogous statement for the free endpoint problem, where one does not restrict to a sublevel $E^{-1}(q_1)$, but considers a penalty in the functional at the end-point.

**Exercise 8.58.** Fix $q_0 \in M$ and $a \in C^\infty(M)$. Prove that with every solution $\bar{u} \in \mathcal{U}$ of the *free endpoint* problem

$$\min_{u\in\mathcal{U}} J(u) - a(E_{q_0}(u)), \tag{8.82}$$

we can associate a normal extremal trajectory whose final covector $\lambda_1 \in T^*_{F(\bar{u})}M$ satisfies

$$\lambda_1 D_{\bar{u}} E_{q_0} = \bar{u}, \qquad \lambda = d_{E_{q_0}(\bar{u})} a.$$

[3] To be precise, here the problem is defined on a Hilbert manifold and not on a subspace a Hilbert space, but since $M$ is finite dimensional, the theory applies with essentially no modifications.

**Local length-minimality in the $C^0$ topology.**

To prove the main result we first need the following lemma.

**Lemma 8.59.** *Let $\gamma : [0,1] \to M$ be a normal extremal trajectory that does not contain abnormal segments. If $\gamma$ does not contain points that are conjugate to $\gamma(0)$, then there exists $a \in C^\infty(M)$ such that for $s \in [0,1]$*

$$\lambda_0 = d_{q_0} a, \qquad \mathrm{Hess}_{(q_0,u^s)} \left( J + a \Big|_{\mathbb{E}^{-1}(\gamma(s))} \right) > 0$$

*where $u^s$ is the control associated with $\gamma^s(t) = \gamma(st)$, for $t \in [0,1]$.*

*Proof.* Recall that $\mathbb{E} : M \times \mathcal{U} \to M$ is the free end-point map $\mathbb{E}(q,u) = E_q(u)$, where $E_q$ is the end-point based at $q \in M$. We have

$$\ker(D_{(q_0,u)}\mathbb{E}) \subset T_{q_0}M \oplus L^2([0,1],\mathbb{R}^m).$$

In what follows we denote elements of $T_{q_0}M \oplus L^2([0,1],\mathbb{R}^m)$ by $(\xi, v)$ where $\xi \in T_{q_0}M$ and $v \in L^2([0,1],\mathbb{R}^m)$. Notice that we have the isomorphism

$$\ker(D_{(q_0,u)}\mathbb{E}) \cap (0 \oplus L^2([0,1],\mathbb{R}^m)) \simeq \ker(D_u E_{q_0}).$$

It follows that for $s \in (0,1]$

$$\mathrm{Hess}_{(q_0,u^s)} (J + a) \Big|_{0 \oplus \ker(D_{u^s} E_{q_0})} = \mathrm{Hess}_{u^s} \left( J \Big|_{E_{q_0}^{-1}(\gamma(s))} \right) > 0. \tag{8.83}$$

where the inequality follows since the curve $\gamma$ contains no conjugate points. We define the subspace (we omit the restriction in the Hessian to avoid heavy notations)

$$W_s := \{(\xi, v) \in \ker D_{(q_0,u^s)}\mathbb{E} \mid \mathrm{Hess}(J + a)\Big((\xi, v), 0 \oplus \ker D_{u^s} E_{q_0}\Big) = 0\}.$$

Notice that, by construction, $W_s$ depends on $a$ only through its first derivative and

$$\ker D_{(q_0,u^s)}\mathbb{E} = (0 \oplus \ker D_{u^s} F) \oplus W_s. \tag{8.84}$$

Moreover, it follows from (8.83) that, if there is some non-zero pair $(\xi, v) \in W_s$, then $\xi \neq 0$. Hence there exists a map $B_s : T_qM \to L^2([0,1],\mathbb{R}^m)$, which depends on $a$ only through its first derivative, such that

$$W_s = \{(\xi, B_s\xi) \mid \xi \in T_qM\}.$$

We can now show that the Hessian on the full space $\ker D_{(q_0,u^s)}\mathbb{E}$ in terms of the decomposition (8.84) as follows

$$\mathrm{Hess}_{(q_0,u^s)}(J + a)\Big((\xi, B_s\xi) + (0, v), (\xi, B_s\xi) + (0, v)\Big) = \tag{8.85}$$

$$= \mathrm{Hess}_{u^s} J(v, v) + \mathrm{Hess}_{(q_0,u^s)}(J + a)\Big((\xi, B_s\xi), (\xi, B_s\xi)\Big), \tag{8.86}$$

where we used that mixed terms give no contribution thanks to the definition of $W_s$. Notice that the first term in the sum is positive for every $s \in [0,1]$, thanks to (8.83), and is independent on $\xi$.

The second term can be computed in coordinates as follows

$$\mathrm{Hess}_{(q_0,u^s)}(J+a)\Big((\xi, B_s\xi),(\xi, B_s\xi)\Big) = D^2_{q_0}a(\xi,\xi) + 2J(B_s\xi, B_s\xi), \tag{8.87}$$

where we denote by $D^2_{q_0}a$ the second differential of $a$. Notice that, by construction, $J(B_s\xi, B_s\xi)$ is a quadratic form that does not depend on second derivatives of $a$, but only on first ones (through the map $B_s$). In particular, we can choose $a$ with second derivatives large enough in such a way that (8.87) is positive for all $s \in [0,1]$ and all $\xi$. This implies that (8.86) is the sum of two positive terms, hence positive.

□

We can now prove the main result.

**Proposition 8.60.** *Let $\gamma : [0,1] \to M$ be a normal extremal trajectory that does not contain abnormal segments. If $\gamma$ does not contain points that are conjugate to $\gamma(0)$, then $\gamma$ is a local miminum for the length with respect to the $C^0$ topology on the space of admissible trajectories with the same endpoints.*

*Proof.* Assume that $\gamma$ is associated with an initial covector $\lambda_0 \in T^*_qM$, i.e.,

$$\gamma(t) = \pi \circ e^{t\vec{H}}(\lambda_0), \qquad \lambda_0 \in T^*_qM.$$

Combining Lemma 8.59 and Remark 8.28, one obtains that $(\mathbb{E}, J+a)$ is a Morse problem (locally in a neighborhood of $(q_0,u^s)$, for $s \in [0,1]$). By the general argument of Section 8.4.1 we have that

$$\mathcal{L}_{(\mathbb{E},J+a)} = \{e^{\vec{H}}(d_qa) \mid \ q \in M\} \subset T^*M.$$

Moreover, for $s \in [0,1]$, then $s\lambda_0$ is a regular point of the map $\pi \circ e^{\vec{H}}|_{\mathcal{L}_0}$, where as usual $\mathcal{L}_0 = \{d_qa \mid q \in M\}$ denotes the graph of the differential of $a$. Using the homogeneity property (8.56) we can rephrase this property by saying that

$$\pi \circ e^{s\vec{H}}|_{\mathcal{L}_0} \ \text{ is an immersion at } \lambda_0, \quad \forall\, s \in [0,1],$$

In particular, being a map between manifolds of the same dimensions, it is a local diffeomorphism. Hence the assumptions of the local version of Theorem 4.62 are satisfied and $\gamma$ is local miminum for the length with respect to the $C^0$ topology on the space of admissible trajectories with the same endpoints. □

Combining the results obtained in the previous sections we have the following result.

**Theorem 8.61.** *Let $\gamma : [0,1] \to M$ be a normal extremal trajectory that does not contain abnormal segments.*

*(i) if $\gamma$ has no conjugate point to $\gamma(0)$ then its a local length-minimizer with respect to the $C^0$ topology on the space of admissible trajectories with the same endpoints,*

*(ii) if $\gamma$ has at least a conjugate point to $\gamma(0)$ then its not a local length-minimizer with respect to the $W^{1,2}$ topology on the space of admissible trajectories with the same endpoints.*

The first statement is Proposition 8.60. The second one follows from Lemma 8.55. Indeed if there exists $\bar{s} \in (0,1)$ such that $\gamma(\bar{s})$ is conjugate to $\gamma(0)$, then $\alpha(s) < 0$ for $s > \bar{s}$ from claim (d) in Lemma 8.55. This implies that the Hessian of the energy restricted to the level set defined by the end-point map has a negative eigenvalue, hence $\gamma|_{[0,s]}$ is not a local length-minimizer in the $W^{1,2}$ topology on the space of admissible trajectories with the same endpoints.

## 8.9 Compactness of length-minimizers

In this section we reinterpret in terms of the end-point map some results already obtained in Section 3.3, in order to prove compactness of length-minimizers. For simplicity of presentation we assume throughout this section that $M$ is complete with respect to the sub-Riemannian distance.

Fix a point $q_0 \in M$ and denote by $E_{q_0} : L^2([0,1],\mathbb{R}^m) \to M$ the end-point map. Notice that $E_{q_0}$ is globally defined thanks to the completeness assumption and Exercice 8.1.

Furthermore, by the reparametrization invariance of the length, we assume that trajectories are parametrized by constant speed on the interval $[0,1]$. In this case, if $\gamma_u$ is the horizontal curve corresponding to a control $u$, one has $\ell(\gamma_u) = \|u\|_{L^1} = \|u\|_{L^2}$, where

$$\|u\|_{L^1} = \int_0^1 |u(t)|dt, \qquad \|u\|_{L^2} = \left(\int_0^1 |u(t)|^2 dt\right)^{\frac{1}{2}},$$

and $|\cdot|$ denotes the standard norm on $\mathbb{R}^m$.

**Proposition 8.62.** *The end-point map $E_{q_0} : L^2([0,1],\mathbb{R}^m) \to M$ is weakly continuous, namely if $u_n \rightharpoonup u$ in the weak topology of $L^2$, then $E_{q_0}(u_n) \to E_{q_0}(u)$.*

*Proof.* First notice that, since $u_n \rightharpoonup u$ in the weak topology of $L^2$, then there exists $r_0 > 0$ such that $\|u_n\|_{L^2} \leq r_0$. Denote by $B$ the compact ball $\overline{B}_{q_0}(r_0)$. The unique solution $\gamma_n$ of the Cauchy problem

$$\dot\gamma(t) = f_{u_n(t)}(\gamma(t)), \qquad \gamma(0) = q_0$$

satisfies the integral identity

$$\gamma_n(t) = q_0 + \int_0^t f_{u_n(\tau)}(\gamma_n(\tau))d\tau, \tag{8.88}$$

Since $\|u_n\| \leq r_0$ for every $n$, all trajectories $\gamma_n$ are contained in the compact ball $B$, they are Lipschitzian with the same Lipchitz constant. In particular, thanks to the classical Ascoli-Arzelà theorem, the set $\{\gamma_n\}_{n\in\mathbb{N}}$ has compact closure in the space of continuous curves in $M$ with respect to the $C^0$ topology.

Then, by compactness, there exists a convergent subsequence (which we still denote $\gamma_n$) and a limit continuous curve $\gamma$ such that $\gamma_n \to \gamma$ uniformly. Let us show that $\gamma$ is the horizontal trajectory associated to $u$.

Since $u_n$ weakly converges to $u$ we have that $\int_0^t f_{u_n(\tau)}(\gamma_n(\tau))d\tau \to \int_0^t f_{u(\tau)}(\gamma(\tau))d\tau$, since this can be seen as a product between a strongly and a weakly convergent sequence.[4] Passing to the limit for $n \to \infty$ in (8.88), one finds that

$$\gamma(t) = q_0 + \int_0^t f_{u(\tau)}(\gamma(\tau))d\tau,$$

namely that $\gamma$ is the trajectory associated to $u$. This completes the proof. $\square$

*Remark* 8.63. Notice that in the proof one obtains the uniform convegence of trajectories and not only of their end-points.

The previous proposition given another proof of the existence of length-minimizers, cf. Theorem 3.43.

[4] writing the coordinate expression $\sum_{i=1}^m u_{n,i} f_i(\gamma_n(t))$.

**Corollary 8.64** (Existence of length-minimizers)**.** *Let $M$ be a complete sub-Riemannian manifold and $q_0 \in M$. For every $q \in M$ there exists $u \in L^2([0,1],\mathbb{R}^m)$ such that the corresponding horizontal trajectory $\gamma_u$ joins $q_0$ and $q$ and is a length-minimizer, i.e., $\ell(\gamma_u) = d(q_0, q)$.*

*Proof.* Consider a point $q$ in the compact ball $B$. Then take a minimizing sequence, i.e., a sequence $u_n$ such that $E_{q_0}(u_n) = q$ and $\|u_n\|_{L^2} \to d(q_0, q)$. The real sequence $(\|u_n\|_{L^2})_n$ is bounded, hence by weak compactness of balls in $L^2$ there exists a subsequence, that we still denote by the same symbol, such that $u_n \rightharpoonup u$ for some $u$. By Proposition 8.62, we have $E_{q_0}(u) = q$. Moreover the semicontinuity of the $L^2$ norm with respect to weak convergence proves that $u$ corresponds to a length-minimizer joining $q_0$ to $q$. Indeed

$$\|u\|_{L^2} \le \liminf_{n\to\infty} \|u_n\|_{L^2} = d(q_0, q). \qquad \square$$

**Definition 8.65.** A control $u$ is called a *minimizer* if it satisfies $\|u\|_{L^2} = d(q_0, E_{q_0}(u))$. We denote by $\mathcal{M}_{q_0} \subset L^2([0,1],\mathbb{R}^m)$ the set of all minimizing controls from $q_0$.

**Theorem 8.66** (Compactness of minimizers)**.** *Let $K \subset M$ be compact. The set of all minimal controls associated with trajectories reaching $K$*

$$\mathcal{M}_K = \{u \in \mathcal{M}_{q_0} \mid E_{q_0}(u) \in K\},$$

*is compact in the strong topology of $L^2$.*

*Proof.* Consider a sequence $(u_n)_{n\in\mathbb{N}}$ contained $\mathcal{M}_K$. Since $K$ is compact, the sequence of norms $(\|u_n\|_{L^2})_{n\in\mathbb{N}}$ is bounded. Since bounded sets in $L^2$ are weakly compact, up to extraction of a subsequence, we can assume that $u_n \rightharpoonup u$.

From Proposition 8.62 it follows that $E_{q_0}(u_n) \to E_{q_0}(u)$ in $M$ and the continuity of the sub-Riemannian distance implies that $d(q_0, E_{q_0}(u_n)) \to d(q_0, E_{q_0}(u))$. Moreover since $u_n \in \mathcal{M}$ we have that $\|u_n\| = d(q_0, E_{q_0}(u_n))$ and by weak semicontinuity of the $L^2$ norm we get

$$\|u\|_{L^2} \le \liminf_{n\to\infty} \|u_n\|_{L^2} = \liminf_{n\to\infty} d(q_0, E_{q_0}(u_n)) = d(q_0, E_{q_0}(u)). \tag{8.89}$$

Since by definition of distance $d(q_0, E_{q_0}(u)) \le \ell(\gamma_u) \le \|u\|_{L^2}$ we have that all inequalities are equalities in (8.89), hence $u$ is a minimizer and $\|u_n\|_{L^2} \to \|u\|_{L^2}$, which implies that $u_n \to u$ strongly in $L^2$. $\square$

Theorem 8.66 implies the following continuity property.

**Proposition 8.67.** *Let $M$ be a complete sub-Riemannian manifold and assume that $q \in M$ is reached by a unique length-minimizer starting from $q_0$ associated with $u$. If $u_n$ is any sequence of minimizing controls such that $E_{q_0}(u_n) \to q$, then $u_n \to u$ in the strong $L^2$ topology.*

*Proof.* Fix an arbitrary subsequence $u_{k_n}$ of the original sequence $u_n$. Consider the compact set $K := \{q\}$ in $M$. By construction $u_{k_n} \in \mathcal{M}_K$ for all $n \in \mathbb{N}$. Hence by Theorem 8.66 $u_{k_n}$ admit a convergent subsequence $u_{k_n} \to \widehat{u}$, for some control $\widehat{u} \in \mathcal{M}_K$, and the trajectory corresponding to $\widehat{u}$ is a length-minimizer joining $q_0$ to $q$. By uniqueness $\widehat{u} = u$.

This proves that every subsequence of $u_n$ admits a subsequence converging to the same element $u$. This implies that the whole sequence $u_n$ converges to $u$. $\square$

*Remark* 8.68. If $M$ is not complete, all the results of this section hold true by restricting the end-point map to a ball $\mathcal{B}_{L^2}(r_0) \subset L^2([0,1],\mathbb{R}^m)$, where $r_0 > 0$ is chosen in such a way that the sub-Riemannian ball $\overline{B}_{q_0}(r_0)$ is compact. See also Exercice 8.1.

## 8.10 Cut locus and global length-minimizers

In this section we discuss some global properties of length-minimizers. We assume throughout the section that $M$ is a complete sub-Riemannian manifold.

**Definition 8.69.** An admissible trajectory $\gamma : [0,T] \to M$ is called a *geodesic* if it is parametrized by constant speed and for every $t \in [0,T]$ there exists $\varepsilon > 0$ such that $\ell(\gamma|_{[t-\varepsilon,t+\varepsilon]})$ is equal to the distance between its end-points.

A geodesic $\gamma : [0,T] \to M$ is said to be *maximal* if it is not the restriction of a geodesic $\gamma' : [0,T'] \to M$ to a smaller interval, meaning that $\gamma = \gamma'|_{[0,T]}$. In this section a geodesic is always assumed to be maximal.

By Theorem 4.65, a normal extremal trajectory parametrized by unit speed is a geodesic. When $M$ is complete, every normal extremal is extendable to $[0,+\infty[$ thanks to Corollary 8.39.

**Exercise 8.70.** Let $\gamma$ be a geodesic. Introduce the set $A = \{t > 0 : \gamma|_{[0,t]}$ is length-minimizing$\}$. Prove that $A$ is an interval of the form $(0, t_*]$ or $(0, +\infty)$.

**Definition 8.71.** Let $\gamma$ be a geodesic. Define

$$t_*(\gamma) := \sup\{t > 0 : \gamma|_{[0,t]} \text{ is length-minimizing}\}.$$

If $t_*(\gamma) < +\infty$ we say that $\gamma(t_*)$ is the *cut point* to $\gamma(0)$ along $\gamma$. If $t_*(\gamma) = +\infty$ we say that $\gamma$ *has no cut point*. We denote by $\mathrm{Cut}_{q_0}$ the set of all cut points of geodesics starting from a point $q_0 \in M$.

The following is the fundamental property of cut locus along normal extremal trajectories.

**Theorem 8.72.** *Let $M$ be a complete sub-Riemannian manifold and $\gamma : [0,T] \to M$ be a normal extremal trajectory that does not contain abnormal segments.*

*Assume that $\gamma(t_0)$ is the cut point to $\gamma(0)$ along $\gamma$, for some $t_0 \in (0,T)$. Then*

*(a) either $\gamma(t_0)$ is the first conjugate point to $\gamma(0)$ along $\gamma$,*

*(b) or there exists a length-minimizer $\widehat{\gamma} \neq \gamma$ joining $\gamma(0)$ and $\gamma(t_0)$ with $\ell(\widehat{\gamma}) = \ell(\gamma|_{[0,t_0]})$.*

*Conversely, if there exists $t_0 \in (0,T)$ such that either (a) or (b) is satisfied, then there exists $t_* \in (0, t_0]$ such that $\gamma(t_*)$ is the cut point along $\gamma$.*

We stress that the two cases (a) and (b) are not mutually exclusive.

*Proof.* Assume first that $\gamma(t_0)$ is the cut point to $\gamma(0)$ along $\gamma$, and that (a) does not hold, i.e., the segment $[0, t_0]$ contains no conjugate points. Let us show that in this case (b) holds.

Fix a sequence $t_n \to t_0$ such that $t_n > t_0$ for all $n \in \mathbb{N}$. Since the manifold is complete, for every $n \in \mathbb{N}$ there exists a length-minimizer $\gamma_n$ joining $\gamma(0)$ to $\gamma(t_n)$, namely $\ell(\gamma_n) = d(\gamma(0), \gamma(t_n))$.

By compactness of length-minimizers there exists (up to extraction of a convergent subsequence) a limit length-minimizer $\widehat{\gamma}$ such that $\gamma_n \to \widehat{\gamma}$ uniformly, and the curve $\widehat{\gamma}$ joins $\gamma(0)$ and $\gamma(t_*)$. Moreover $\ell(\widehat{\gamma}|_{[0,t_*]}) = d(\gamma(0), \gamma(t_*)) = \ell(\gamma|_{[0,t_*]})$.

On the other hand, since the segment $\gamma|_{[0,t_*]}$ contains no conjugate points, the curve $\gamma|_{[0,t_*]}$ is a local length-minimizer in the $C^0$ topology. Thus $\widehat{\gamma}$ cannot be contained in a neighborhood of $\gamma$ and $\widehat{\gamma} \neq \gamma$, ending the proof.

Let us now prove the converse. Assume first that there exists $t_0 > 0$ such that (a) is satisfied and, by contradiction, that the cut time $t_*$ is strictly bigger than $t_0$. This implies that $\gamma|_{[0,t_*]}$ is a length-minimizer contradicting claim (ii) of Theorem 8.61.

Assume now that there exists $t_0 > 0$ such that assumption (b) is satisfied, namely there exists a length-minimizer $\widehat{\gamma} \neq \gamma$ such that $\widehat{\gamma}(t_0) = \gamma(t_0)$. From this it follows that the concatenation of the two curves $\widehat{\gamma}|_{[0,t_0]}$ and $\gamma|_{[t_0,T]}$ is also a length-minimizer, hence it satisfies the first-order necessary conditions. This defines two different normal lifts of the normal extremal trajectory $\gamma|_{[t_0,T]}$, hence $\gamma|_{[t_0,T]}$ would be an abnormal segment (cf. Exercise 8.10), contradicting our assumption on $\gamma$.

$\square$

**Theorem 8.73.** *Let $M$ be a complete sub-Riemannian manifold. Let $\gamma : [0,1] \to M$ be a normal extremal trajectory that does not contain abnormal segments. Assume that for some $t_0 \in (0,1)$*

*(i) $\gamma|_{[0,t_0]}$ is a length-minimizer,*

*(ii) there exists a neighborhood $U$ of $\gamma(t_0)$ such that every point of $U$ is reached by a unique length-minimizer from $\gamma(0)$, which is not abnormal.*

*Then $\gamma(t_0)$ is not conjugate to $\gamma(0)$. Moreover there exists $\varepsilon > 0$ such that $\gamma|_{[0,t_0+\varepsilon]}$ is a length-minimizer.*

*Proof.* It is enough to show that there exists $\varepsilon > 0$ such that the segment $[0, t_0+\varepsilon]$ does not contain conjugate points to $\gamma(0)$. Indeed this fact, together with assumptions (i) and (ii), imply that the cut time $t_*$ along $\gamma$ satisfies $t_* \geq t_0 + \varepsilon$ (cf. Theorem 8.72).

Fix a neighborhood $U$ of $\gamma(t_0)$ and, for each $q \in U$, let us denote by $\gamma^q$ (resp. $u^q$) the minimizing length-parametrized trajectory (resp. control) joining $\gamma(0)$ to $q$. Thanks to Proposition 8.67 the map $q \mapsto u^q$ is continuous in the strong topology in $L^2$.

Hence we can consider the family $\lambda_1^q$ of normal final covectors associated with $u^q$, i.e., satisfying the Lagrange multipliers rule

$$\lambda_1^q D_{u^q} F = u^q, \qquad \forall\, q \in U.$$

By the smoothness of the end-point map $E_{q_0}$, the map $q \mapsto D_{u^q}E_{q_0}$ is continuous. Moreover $D_{u^q}E_{q_0}$ is surjective for every $q$ since the normal extremal trajectory associated with $u^q$ is not abnormal. The adjoint map $(D_{u^q}F)^* : T_q^*M \to L^2([0,1],\mathbb{R}^m)$ is then injective and $\lambda_1^q$ is the unique solution to the linear equation $(D_{u^q}F)^*\xi = u^q$ (the uniqueness of covector is guaranteed since the trajectory is not abnormal by assumption (ii)). Since the coefficients of the linear equation are continuous with respect to $q$, this implies that the map $\Phi^1 : q \mapsto \lambda_1^q$ is continuous, as well as the map $\Phi^0 : q \mapsto \lambda_0^q$ that associates with every $q$ the initial covector $\lambda_0^q$ of the trajectory joining $q_0$ with $q$, since $\Phi^0(q) = (P_{0,1}^{u^q})^* \circ \Phi^1(q)$.

Moreover, by construction, we have $\exp_{q_0}(\Phi^0(q)) = q$ for every $q \in U$, i.e, $\Phi^0$ is a continuous right inverse of the exponential map $\exp_{q_0}$. Thus the map $\Phi^0$ is injective on $U$ and, by the Brouwer invariance of domain theorem, $\Phi^0$ is an open map and $\Phi^0 : U \to A := \Phi^0(U)$ is an homeomorphism, with $\lambda_0^{\gamma(t_0)} \in A \subset T_{q_0}^*M$.

Fix $\delta_0 > 0$ small enough such that $(1+\delta)\lambda_0^{\gamma(t_0)} \in A$ for $|\delta| < \delta_0$. By homogeneity $(1+\delta)\lambda_0^{\gamma(t_0)} = \lambda_0^{\gamma((1+\delta)t_0)}$. This means that the unique length-minimizer joining $q_0$ with $\gamma((1+\delta_0)t_0)$ is $\gamma|_{[0,(1+\delta_0)t_0]}$. Thus $\gamma$ deos not contain conjugate points in the segment $[0, t_0+\varepsilon]$ for every $\varepsilon < \delta_0 t_0$ (cf. Theorem 8.72). $\square$

We end this section by explicitly stating the converse of Theorem 8.73, in the case when the structure admits no abnormal length-minimizers.

**Corollary 8.74.** *Assume that the sub-Riemannian structure admits no abnormal length-minimizer. Let $\gamma : [0,1] \to M$ be a horizontal curve such that for some $t_0 \in (0,1)$*

*(i) $\gamma|_{[0,t_0]}$ is a length-minimizer,*

*(ii) $\gamma(t_0)$ is conjugate to $\gamma(0)$.*

*Then any neighborhood of $\gamma(t_0)$ contains a point reached from $\gamma(0)$ by at least two distinct length-minimizers.*

*Remark* 8.75. Thanks to Theorem 8.72, if a sub-Riemannian structure admits no abnormal length-minimizer, then points where geodesics lose global optimality can be of two types: (a) (first) conjugate points, or (b) points reached by two distinct length-minimizers.

Corollary 8.74 says that, if there are no abnormal length-minimizers, cut points of type (a) always appears as accumulation points of those of type (b). Hence to compute the cut locus is is enough to consider the closure of points reached by at least two length-minimizers.

To end this chapter we prove a regularity property for the cut time function when abnormal length-minimizers are absent. Notice that in this case all unit speed geodesic are parametrized through the initial covector, belonging to $H^{-1}(1/2) \subset T^*M$.

Given $\xi \in H^{-1}(1/2)$, we denote by $\gamma_\xi$ the corresponding geodesic and we define $c(\xi) := t_*(\gamma_\xi)$.

**Proposition 8.76.** *Let $M$ be a complete sub-Riemannian structure that admits no abnormal length-minimizer. Then the cut time function $c : H^{-1}(1/2) \to \mathbb{R}$ is continous.*

We stress that we regard here $H^{-1}(1/2)$ as a subset of $T^*M$ and we do not restrict here to a single fiber. Hence the continuity is both in the base point and the covector.

*Proof.* We prove separately (i) the upper semicontinuity and (ii) the lower semicontinuity.

(i) Fix $\xi \in H^{-1}(1/2)$, and let $(\xi_n)_n$ be a sequence in $H^{-1}(1/2)$ such that $\xi_n \to \xi$ for $n \to \infty$. In particular $q_n := \pi(\xi_n)$ tends to $q := \pi(\xi)$. We have to prove that

$$\limsup_{n\to+\infty} c(\xi_n) \leq c(\xi). \tag{8.90}$$

Denote by $c_n := c(\xi_n)$. Assume that $c_n$ is unbounded. Up to subsequences, we can assume $c_n \to +\infty$. Hence for every $T > 0$ we have $\lim_{j\to\infty} \gamma_{\xi_n}(T) = \gamma_\xi(T)$ and, by continuity of the sub-Riemannian distance,

$$d(q, \gamma_\xi(T)) = \lim_{j\to\infty} d(q_n, \gamma_{\xi_n}(T)) = T,$$

which proves $c(\xi) = +\infty$. If the left hand side of (8.90) is finite the it is not restrictive to assume that $c_n$ is convergent to some $c_* > 0$. Then for every $\varepsilon < c_*$ one has

$$d(q, \gamma_\xi(c_* - \varepsilon)) = \lim_{n\to\infty} d(q_n, \gamma_{\xi_n}(c_n - \varepsilon)) = \lim_{n\to\infty} c_n - \varepsilon = c_* - \varepsilon \tag{8.91}$$

This proves $c(\xi) \geq c_*$ and completes the proof of (i).

(ii) Let us now show that if $\xi_n \to \xi$ in $H^{-1}(1/2)$ for $n \to \infty$ we have

$$\liminf_{n\to+\infty} c(\xi_n) \geq c(\xi). \tag{8.92}$$

It is enough to consider the case when $c(\xi_n)$ is convergent to some $c_* > 0$. We want to show that the limit curve $\gamma_\xi$ is not minimizing after $\gamma_\xi(c_*)$. By passing to a subsequence, if necessary, we may assume that either (a) $\gamma_{\xi_n}(c_n)$ is conjugate to $q_n$ along $\gamma_{\xi_n}$ for all $n$, or (b) that for each $n$ there exists $\xi'_n \in H^{-1}(1/2)$, with $\pi(\xi'_n) = \pi(\xi_n)$ and $\xi'_n \neq \xi_n$, such that $\gamma_{\xi_n}(c_n) = \gamma_{\xi'_n}(c_n)$.

In case (a), by smoothness of the exponential map both in the base point and in the argument, the point $\gamma_\xi(c_*)$ is certainly conjugate to $q$ along $\gamma_\xi$, hence $c(\xi) \leq c_*$. In case (b), it is not restrictive to assume that $\xi'_n$ also converges to some $\xi'$. If $\xi \neq \xi'$, then there exists two different arclength parametrized length-minimizers (starting from $\pi(\xi) = \pi(\xi')$) reaching the same point in time $c_*$. If $\xi = \xi'$ then, again by smoothness of the exponential map in both argument, the exponential map based at $\pi(\xi)$ is not a local diffeomorphism, hence again $\gamma_\xi(c_*)$ is conjugate to $q$ along $\gamma_\xi$. In both cases we have $c(\xi) \leq c_*$. □

*Remark* 8.77. Notice that removing the assumptions of completeness and absence of abnormal length-minimizers, the part (i) of the proof of Proposition 8.76 still holds, hence the cut time function $c : H^{-1}(1/2) \to \mathbb{R}$ is always upper semicontinuous.

## 8.11 An example: the first conjugate locus on perturbed sphere

In this section we prove that a $C^\infty$ small perturbation of the standard metric on $S^2$ has a first conjugate locus with at least 4 cusps. See Figure 8.2. Recall that geodesics for the standard metric on $S^2$ are great circles, and the first conjugate locus from a point $q_0$ coincides with its antipodal point $\widehat{q}_0$. Indeed all geodesics starting from $q_0$ meet there and lose their optimality at $\widehat{q}_0$.

Denote $H_0$ the Hamiltonian associated with the standard metric on the sphere and let $H$ be an Hamiltonian associated with a Riemannian metric on $S^2$ such that $H$ is sufficiently close to $H_0$, with respect to the $C^\infty$ topology for smooth functions in $T^*M$.

Fix a point $q_0 \in S^2$. Normal extremal trajectories starting from $q_0$ and parametrized by length (with respect to the Hamiltonian $H$) can be parametrized by covectors $\lambda \in T^*_{q_0}M$ such that $H(\lambda) = 1/2$. The set $H^{-1}(1/2)$ is diffeomorphic to a circle $S^1$ and can be parametrized by an angle $\theta$. For a fixed initial condition $\lambda_0 = (q_0, \theta)$, where $q_0 \in M$ and $\theta \in S^1$ we write

$$\lambda(t) = e^{t\vec{H}}(\lambda_0) = (p(t,\theta), \gamma(t,\theta)),$$

and we denote by $\exp = \exp_{q_0}$ the exponential map based at $q_0$

$$\exp_{q_0}(t, \lambda_0) = \pi \circ e^{t\vec{H}}(\lambda_0) = \gamma(t,\theta).$$

For every initial condition $\theta \in S^1$ denote by $t_c(\theta)$ the first conjugate time along $\gamma(\cdot,\theta)$, i.e., $t_c(\theta) = \inf\{\tau > 0 \,|\, \gamma(\tau,\theta)$ is conjugate to $q_0$ along $\gamma(\cdot,\theta)\}$.

**Proposition 8.78.** *The first conjugate time $t_c(\theta)$ is characterized as follows*

$$t_c(\theta) = \inf\left\{ t > 0 \;\middle|\; \frac{\partial \exp}{\partial \theta}(t,\theta) = 0 \right\}. \tag{8.93}$$

*Proof.* First notice that $\frac{\partial \exp}{\partial t}(t,\theta) = \dot{\gamma}(t,\theta) \neq 0$. Hence conjugate points correspond to critical points of the exponential map, i.e., points $\exp(t,\theta)$ such that

$$\operatorname{rank}\left\{\frac{\partial \exp}{\partial t}(t,\theta), \frac{\partial \exp}{\partial \theta}(t,\theta)\right\} = 1. \tag{8.94}$$

Let us show that condition (8.94) occurs only if $\frac{\partial \exp}{\partial \theta}(t,\theta) = 0$. Indeed, by Proposition 8.42, one has that

$$\left\langle p(t,\theta), \frac{\partial \exp}{\partial t}(t,\theta)\right\rangle = 1, \qquad \left\langle p(t,\theta), \frac{\partial \exp}{\partial \theta}(t,\theta)\right\rangle = 0,$$

thus, whenever $\frac{\partial \exp}{\partial \theta}(t,\theta) \neq 0$, the two vectors appearing in (8.94) are always linearly independent. □

**Lemma 8.79.** *The function $\theta \mapsto t_c(\theta)$ is of class $C^1$.*

*Proof.* By Proposition 8.78, $t_c(\theta)$ is a solution to the equation (in the variable $t$)

$$\frac{\partial \exp}{\partial \theta}(t,\theta) = 0. \tag{8.95}$$

Let us first remark that, for the exponential map $\exp_0$ associated with the Hamitonian $H_0$ we have

$$\frac{\partial \exp_0}{\partial \theta}(t_c^0(\theta),\theta) = 0, \qquad \frac{\partial^2 \exp_0}{\partial t \partial \theta}(t_c^0(\theta),\theta) \neq 0, \tag{8.96}$$

where $t_c^0(\theta)$ is the first conjugate time with respect to the metric induced by $H_0$, as it is easily checked.

Since $H$ is close to $H_0$ in the $C^\infty$ topology, by continuity with respect to the data of solution of ODEs, we have that exp is close to $\exp_0$ in the $C^\infty$ topology too. Moreover the condition (8.96) ensures the existence of a solution $t_c(\theta)$ of (8.95) that is close to $t_c^0(\theta)$. Hence we have that

$$\frac{\partial^2 \exp}{\partial t \partial \theta}(t_c(\theta),\theta) \neq 0 \tag{8.97}$$

By the implicit function the function $\theta \mapsto t_c(\theta)$ is well-defined and of class $C^1$. □

Let us introduce the function $\beta : S^1 \to M$ defined by $\beta(\theta) = \exp(t_c(\theta),\theta)$. The first conjugate locus, by definition, is the image of the map $\beta$. The cuspidal point of the conjugate locus are by definition those points where the function $\theta \mapsto t_c'(\theta)$ change sign. By continuity (cf. proof of Lemma 8.79) the map $\beta$ takes value in a neighborhood of the point $\widehat{q}_0$ antipodal to $q_0$. Let us considr stereographic coordinates around this point and consider $\beta$ as a function from $S^1$ to $\mathbb{R}^2$. By the chain rule and (8.95), we have

$$\beta'(\theta) = t_c'(\theta)\frac{\partial \exp}{\partial t}(t_c(\theta),\theta) + \underbrace{\frac{\partial \exp}{\partial \theta}(t_c(\theta),\theta)}_{=0}. \tag{8.98}$$

Let us define $g, g_0 : S^1 \to \mathbb{R}^2$ by $g(\theta) := \frac{\partial \exp}{\partial t}(t_c(\theta),\theta)$ and $g_0(\theta) := \frac{\partial \exp_0}{\partial t}(t_c^0(\theta),\theta)$. The set

$$C_0 = \{\rho g_0(\theta) \,|\, \theta \in S^1, \rho \in [0,1]\},$$

is strictly convex, since

$$g_0(\theta) = \begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix}.$$

By assumption, the perturbation of the metric is small in the $C^\infty$-topology, hence

$$C = \{\rho g(\theta) \mid \theta \in S^1, \rho \in [0,1]\}, \tag{8.99}$$

remains strictly convex.

**Theorem 8.80.** *The conjugate locus of the perturbed sphere has at least 4 cuspidal points.*

*Proof.* Notice that the function $\theta \mapsto t_c'(\theta)$ can change sign only an even number of times on $S^1 = \mathbb{R}/2\pi\mathbb{Z}$. Moreover

$$\int_0^{2\pi} t_c'(\theta)d\theta = t_c(2\pi) - t_c(0) = 0. \tag{8.100}$$

A continuous function with zero integral mean on $[0, 2\pi]$, which is not identically zero, changes sign at least twice on the interval. Notice also that

$$\int_0^{2\pi} t_c'(\theta)g(\theta)d\theta = \int_0^{2\pi} \beta'(\theta)d\theta = \beta(2\pi) - \beta(0) = 0. \tag{8.101}$$

Let us now assume by contradiction that the function $\theta \mapsto t_c'(\theta)$ changes sign exactly twice at $\theta_1, \theta_2 \in S^1$. Then, by convexity of $C$, there exists a covector $\eta \in (\mathbb{R}^2)^*$ such that $\langle \eta, g(\theta_i)\rangle = 0$ for $i = 1, 2$ and such that $t_c'(\theta)\langle \eta, g(\theta)\rangle > 0$ if $\theta \neq \theta_i$ for $i = 1, 2$. This implies in particular

$$\left\langle \eta, \int_0^{2\pi} t_c'(\theta)g(\theta)d\theta \right\rangle = \int_0^{2\pi} t_c'(\theta)\langle \eta, g(\theta)\rangle\, d\theta \neq 0,$$

which contradicts (8.101). $\square$

*Remark* 8.81. A careful analysis of the proof shows that the statement remains true if one considers a small perturbation of the Hamiltonian (or equivalently, the metric) in the $C^4$ topology. Indeed the key point is that $g$ is close to $g_0$ in the $C^2$ topology, to preserve the convexity of the set $C$ defined by (8.99).

The same argument can be applied for every arbitrary small $C^\infty$ (and actually $C^4$) perturbation $H$ of the Riemannian Hamiltonian $H_0$ associated with the standard Riemannian structure on $S^2$, without requiring that $H$ comes from a Riemannian metric.

## 8.12 Bibliographical note

The study of Lagrange multipliers through the theory of Morse problems have been initiated in [AG98], see also the survey [Agr08]. The presentation given here is inspired by those references, where a more general setting is considered. Compactness of length-minimizers have been proved in [Agr98a].

The study of the cut and conjugate loci on surfaces is a classical topic in Riemannian geometry [Cha06, GHL90].

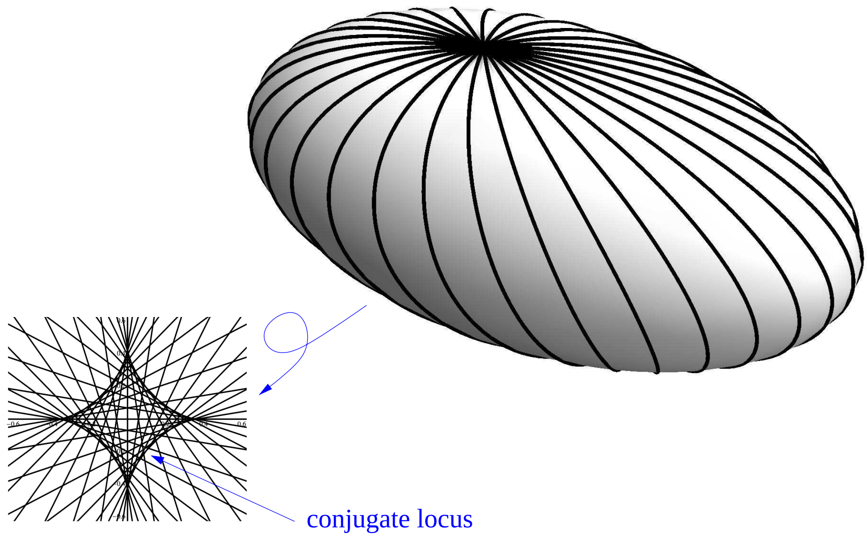



Figure 8.2: Perturbed sphere or ellipsoid

One of the first general results is due to S. B. Myers, who proved that the cut locus from a point in a two-dimensional, real analytic, complete Riemannian manifold form a one-dimensional graph [Mye36]. More recently the fact that every graph can be a cut locus have been proved in [IV15].

The fact that the conjugate locus on every ellipsoid has *exactly* four cusps has been proved only recently in [IK04], and it is known as the *Last geometric statement of Jacobi*.

In the sub-Riemannian case the cut locus and the conjugate locus have been widely studied in the literature. In this chapter we prove that, in absence of abnormal length-minimizers, the properties that hold in Riemannian geometry extends to the sub-Riemannian context.

# Chapter 9

# 2D Almost-Riemannian Structures

Almost-Riemannian structures are examples of sub-Riemannian structures such that the local minimum bundle rank (cf. Definition 3.21) is equal to the dimension of the manifold at each point (cf. Section 3.1.3). They are the prototype of rank-varying sub-Riemannian structures. In this chapter we study the 2-dimensional case, that is very simple since it is Riemannian almost everywhere (see Theorem 9.14), but presents already some interesting phenomena as for instance the presence of sets of finite diameter but infinite area and the presence of conjugate points even when the curvature is always negative (where it is defined). Also the Gauss-Bonnet theorem has a surprising form in this context.

## 9.1 Basic definitions and properties

Thanks to Exercise 3.29, given a structure having constant local minimum bundle rank $m$ one can find an equivalent one having bundle rank $m$. In dimension 2, due to the Lie bracket-generating assumption, also the opposite holds true in the following sense: a structure having bundle rank 2 has local minimal bundle rank 2. Hence we can define a 2D almost-Riemannian structure in the following simpler way.

**Definition 9.1.** Let $M$ be a 2-D connected smooth manifold. A 2D *almost-Riemannian structure* on $M$ is a pair $(\mathbf{U}, f)$ as follows:

- $\mathbf{U}$ is an Euclidean bundle over $M$ of rank 2. We denote each fiber by $U_q$, the scalar product on $U_q$ by $(\cdot\,|\,\cdot)_q$ and the norm of $u \in U_q$ as $|u| = \sqrt{(u\,|\,u)_q}$.
- $f : \mathbf{U} \to TM$ is a smooth map that is a morphism of vector bundles i.e., $f(U_q) \subseteq T_qM$ and $f$ is linear on fibers.
- $\mathcal{D} = \{f(\sigma) \mid \sigma : M \to \mathbf{U} \text{ smooth section}\}$, is a bracket-generating family of vector fields.

Let us recall some definitions and notations (already introduced for general sub-Riemannian structures).

- The *distribution* is defined as $\mathcal{D}(q) = \{X(q) \mid X \in \mathcal{D}\} = f(U_q) \subseteq T_qM$.
- The *step* of the structure at $q \in M$ is the minimal $s \in \mathbb{N}$, $s \geq 1$ such that $\mathcal{D}_s(q) = T_qM$, where $\mathcal{D}_1 := \mathcal{D}$, $\mathcal{D}_{i+1} := \mathcal{D}_i + [\mathcal{D}_1, \mathcal{D}_i]$, for $i \geq 1$.

- The *(almost-Riemannian) norm* of a vector $v \in \mathcal{D}_q$ is

$$\|v\| := \min\{|u|,\, u \in U_q \;\text{ s.t. }\; v = f(q,u)\}.$$

- An *admissible curve* is a Lipschitz curve $\gamma : [0,T] \to M$ such that there exists a measurable and essentially bounded function $u : [0,T] \ni t \mapsto u(t) \in U_{\gamma(t)}$, called *control function*, such that $\dot\gamma(t) = f(\gamma(t), u(t))$, for a.e. $t \in [0,T]$. Recall that there may be more than one control corresponding to the same admissible curve.

- The *minimal control* of an admissible curve $\gamma$ is

$$u^*(t) := \operatorname{argmin}\{|u|,\, u \in U_{\gamma(t)} \;\text{ s.t. }\; \dot\gamma(t) = f(\gamma(t), u)\}$$

  (for all $t$ differentiability point of $\gamma$). Recall that the minimal control is measurable (cf. Section 3.5).

- The *(almost-Riemannian) length* of an admissible curve $\gamma : [0,T] \to M$ is

$$\ell(\gamma) := \int_0^T \|\dot\gamma(t)\| dt = \int_0^T |u^*(t)| dt.$$

- The *(almost-Riemannian) distance* between two points $q_0, q_1 \in M$ is

$$d(q_0, q_1) = \inf\{\ell(\gamma) \,|\, \gamma : [0,T] \to M \text{ admissible},\; \gamma(0) = q_0,\; \gamma(T) = q_1\}. \tag{9.1}$$

Recall that thanks to the bracket-generating condition, the Chow-Rashevskii theorem (Theorem 3.31) guarantees that $(M,d)$ is a metric space and that the topology induced by $(M,d)$ is equivalent to the manifold topology.

In this chapter we use the terminology “orthonormal frame” in a slightly generalized sense.

**Definition 9.2.** An *orthonormal frame* for the 2D almost-Riemannian structure on $\Omega$ is the pair of vector fields $\{F_1, F_2\} := \{f \circ \sigma_1, f \circ \sigma_2\}$ where $\{\sigma_1, \sigma_2\}$ is an orthonormal frame for $(\cdot\,|\,\cdot)_q$ on a local trivialization $\Omega \times \mathbb{R}^2$ of $\mathbf{U}$.

On a local trivialization $\Omega \times \mathbb{R}^2$, the map $f$ can be written as $f(q,u) = u_1 F_1(q) + u_2 F_2(q)$. As usual, when this can be done globally (i.e., when $\mathbf{U}$ is the trivial bundle) we say that the 2D almost-Riemannian structure is *free*.

Notice that orthonormal frames in the sense of Definition 9.2 in the Riemannian sense out of the singular set (cf. Proposition 9.6).

In this chapter we do not work with an equivalent structure of larger bundle rank that is free (cf. Section 3.1.4). Technically such a structure fits Definition 3.21 (i.e., that local minimum bundle rank is equal to the dimension of the manifold at each point) but not Definition 9.1. This point of view permits to understand how global properties of $\mathbf{U}$ (as its orientability and its topology) are transferred in properties of the almost-Riemannian structure.

**Definition 9.3.** A 2D almost-Riemannian structure $(\mathbf{U}, f)$ over $M$ is said to be *orientable* if $\mathbf{U}$ is orientable. It is said to be *fully orientable* if both $\mathbf{U}$ and $M$ are orientable.

Observe that free 2D almost-Riemannian structures are always orientable.

Given an orientable 2D almost-Riemannian structure, if $\{F_1, F_2\}$ and $\{G_1, G_2\}$ are two positively oriented orthonormal frames defined respectively on two open subsets $\Omega$ and $\Xi$, then on $\Omega \cap \Xi$ there exists a smooth function $\theta : M \to S^1$ such that

$$\begin{pmatrix} G_1(q) \\ G_2(q) \end{pmatrix} = \begin{pmatrix} \cos(\theta(q)) & \sin(\theta(q)) \\ -\sin(\theta(q)) & \cos(\theta(q)) \end{pmatrix} \begin{pmatrix} F_1(q) \\ F_2(q) \end{pmatrix}.$$

As shown by the following examples, one can construct orientable 2D almost-Riemannian structures on non-orientable manifolds and viceversa.

**An orientable 2D almost-Riemannian structure on the Klein bottle.** Let $M$ be the Klein bottle seen as the square $[-\pi, \pi] \times [-\pi, \pi]$ with the identifications $(x_1, -\pi) \sim (x_1, \pi)$, $(-\pi, x_2) \sim (\pi, -x_2)$.

Let $\mathbf{U} = M \times \mathbb{R}^2$ with the standard Euclidean metric and consider the morphism of vector bundles given by

$$f : \mathbf{U} \to TM, \quad f(x_1, x_2, u_1, u_2) = \left(x_1, x_2, u_1, u_2 \sin\left(\frac{x_1}{2}\right)\right).$$

This structure is Lie bracket-generating and the two vector fields

$$F_1(x_1, x_2) = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad F_2(x_1, x_2) = \begin{pmatrix} 0 \\ \sin\left(\frac{x_1}{2}\right) \end{pmatrix},$$

which are well defined on $M$, provide a global orthonormal frame. This structure is orientable since $\mathbf{U}$ is trivial.

**Exercise 9.4.** Construct a non-orientable almost-Riemannian structure on the 2D-torus.

**Definition 9.5.** The *singular set $\mathcal{Z}$ of a 2D almost-Riemannian structure* $(\mathbf{U}, f)$ over $M$ is the set of points $q$ of $M$ such that $f$ is not fiberwise surjective, i.e., such that the rank of the distribution $r(q) := \dim(\mathcal{D}_q)$ is less than 2.

Notice if $q \in \mathcal{Z}$ then $r(q) = 1$. Indeed $r(q) = 0$ at some point $q$, then the structure is not bracket-generating at $q$.

Since outside the singular set $\mathcal{Z}$, $f$ is fiberwise surjective, the following property holds.

**Proposition 9.6.** *A 2D almost-Riemannian structure is a Riemannian structure on $M \setminus \mathcal{Z}$.*

A point $q \in M \setminus \mathcal{Z}$ is called a *Riemannian point*. On Riemannian points, the Riemannian metric is reconstructed with the polarization identity (see Exercice 3.9). It follows that if $\{F_1, F_2\}$ is a local orthonormal frame, $v = v_1 F_1(q) + v_2 F_2(q) \in T_q M$ and $w = w_1 F_1(q) + w_2 F_2(q) \in T_q M$ then the Riemannian metric at $q$ is given by

$$g_q(v, w) = v_1 w_1 + v_2 w_2.$$

By construction, at Riemannian points, $\{F_1, F_2\}$ is an orthonormal frame in the usual sense

$$g_q(F_i(q), F_j(q)) = \delta_{ij}, \quad i, j = 1, 2.$$

**Exercise 9.7.** Assume that in a local system of coordinates an orthonormal frame is given by

$$F_1 = \begin{pmatrix} F_1^1 \\ F_1^2 \end{pmatrix}, \quad F_2 = \begin{pmatrix} F_2^1 \\ F_2^2 \end{pmatrix} \text{ and let } F = (F_i^j)_{i,j=1,2} = \begin{pmatrix} F_1^1 & F_2^1 \\ F_1^2 & F_2^2 \end{pmatrix}.$$

Prove that at Riemannian points the Riemannian metric is represented by the matrix $g = {}^t(F^{-1})F^{-1}$.

The following proposition is very useful to study local properties of 2D almost-Riemannian structures.

**Proposition 9.8.** *Consider a 2D almost-Riemannian structure over $M$. For every point $q_0$ of $M$ there exists a neighborhood $\Omega$ of $q_0$, a local orthonormal frame $\{F_1, F_2\}$ defined in $\Omega$ and a system of coordinates $\{x_1, x_2\}$ in $\Omega$ such that $q_0 = (0,0)$, and $F_1$, $F_2$ can be written in these coordinates as:*

$$F_1(x_1,x_2) = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad F_2(x_1,x_2) = \begin{pmatrix} 0 \\ \mathfrak{f}(x_1,x_2) \end{pmatrix}, \tag{9.2}$$

*where $\mathfrak{f} : \Omega \to \mathbb{R}$ is a smooth function. Moreover*

*(i) the integral curves of $F_1$ are normal Pontryagin extremals;*

*(ii) let $s$ be the step of the structure at $q_0$. If $s = 1$ then $\mathfrak{f}(0,0) \neq 0$. If $s \geq 2$, we have $\mathfrak{f}(0,0) = 0$, $\partial_{x_1}^r \mathfrak{f}(0,0) = 0$ for $r = 1, 2, \dots, s-2$ and $\partial_{x_1}^{s-1}\mathfrak{f}(0,0) \neq 0$.*

*Remark* 9.9. Notice that using the system of coordinates and the orthonormal frame given by Proposition 9.8, we have that $\mathcal{Z} \cap \Omega = \{(x_1,x_2) \in \Omega \mid \mathfrak{f}(x_1,x_2) = 0\}$.

Before proving Proposition 9.8, let us prove the following Lemma.

**Lemma 9.10.** *Consider a 2D almost-Riemannian structure and let $W$ be a smooth embedded one-dimensional submanifold of $M$. Assume that $W$ is transversal to the distribution $\mathcal{D}$, i.e., $\mathcal{D}(q) + T_qW = T_qM$ for every $q \in W$. Then, for every $q \in W$ there exists an open neighborhood $\Omega$ of $q$ such that for every $\varepsilon > 0$ small enough, the set*

$$\{q' \in \Omega \mid d(q', W) = \varepsilon\}, \tag{9.3}$$

*is a smooth embedded one-dimensional submanifold of $\Omega$.*

*Proof.* Let $H : T^*M \to \mathbb{R}$ be the sub-Riemannian Hamiltonian (cf. (4.32)) and consider a smooth regular parametrization $\alpha \mapsto w(\alpha)$ of $W$. Let $\alpha \mapsto \lambda_0(\alpha) \in T^*_{w(\alpha)}M$ be a smooth map satisfying $H(\lambda_0(\alpha)) = 1/2$ and $\lambda_0(\alpha) \perp T_{w(\alpha)}W$.

Let $E(t,\alpha)$ be the solution at time $t$ of the Hamiltonian system with Hamiltonian $H$ and with initial condition $\lambda(0) = \lambda_0(\alpha)$. Fix $q \in W$ and define $\bar\alpha$ by $q = w(\bar\alpha)$. Now let us prove that $E(t,\alpha)$ is a local diffeomorphism around the point $(0,\bar\alpha)$. To this purpose let us show that the two vectors

$$v_1 = \frac{\partial E}{\partial \alpha}(0,\bar\alpha) \text{ and } v_2 = \frac{\partial E}{\partial t}(0,\bar\alpha) \tag{9.4}$$

are linearly independent. On one hand, since $v_1$ is equal to $\frac{dw}{d\alpha}(\bar\alpha)$, then it spans $T_qW$. On the other hand, being $H$ quadratic in $\lambda$,

$$\langle \lambda_0(\bar\alpha), v_2 \rangle = \langle \lambda_0(\bar\alpha), \frac{\partial H}{\partial \lambda}(\lambda_0(\bar\alpha)) \rangle = 2H(\lambda_0(\bar\alpha)) = 1. \tag{9.5}$$

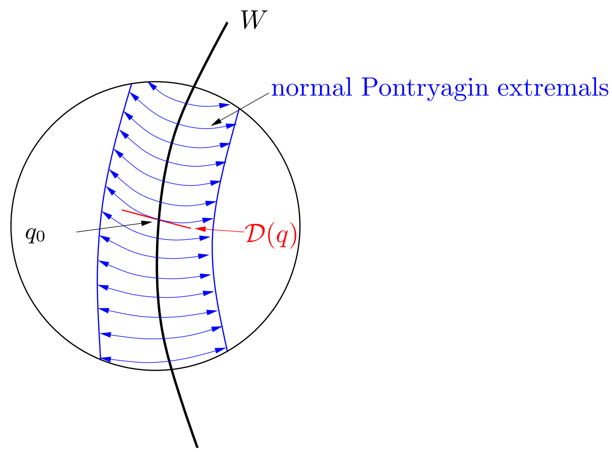



Figure 9.1: Normal Pontryagin extremals starting from the singular set

Thus $v_2$ does not belong to the orthogonal to $\lambda_0(\bar{\alpha})$, that is, to $T_qW$.

Therefore for a small enough neighborhood $\Omega$ of $q$, using the fact that small arcs of normal extremal paths are minimizers, we have that for $\varepsilon > 0$ small enough, the set $A = \{q' \in \Omega \mid d(q', W) = \varepsilon\}$ contains the intersection of $\Omega$ with the images of $E(\varepsilon, \cdot)$ and $E(-\varepsilon, \cdot)$. By possibly restricting $\Omega$, we are in the situation of Figure 9.1 and the set $A$ coincides with the intersection of $\Omega$ with the images of $E(\varepsilon, \cdot)$ and $E(-\varepsilon, \cdot)$. $\square$

*Remark* 9.11. Notice that in this proof we did not make any hypothesis on abnormal extremals. In Section 9.1.3 we are going to see that for 2D almost-Riemannian structures there are no nontrivial abnormal extremals.

*Proof of Proposition 9.8.* Following the notation of the proof of Lemma 9.10 let us parametrize $W$ in such a way that $q_0 = w(0)$. Take $(t, \alpha)$ as a system of coordinates on $\Omega$ and define the vector field $F_1$ by

$$F_1(t, \alpha) = \frac{\partial E(t, \alpha)}{\partial t}. \tag{9.6}$$

Notice that, by construction, for every point in $\Omega$ the vector $F_1$ belongs to the distribution and its almost-Riemannian norm is equal to 1. In the coordinates $(t, \alpha)$ we have $F_1 = (1, 0)$ and by construction its integral curves are normal Pontryagin extremals. Let $F_2$ be a vector field on $\Omega$ such that $\{F_1, F_2\}$ is an orthonormal frame for the 2D almost-Riemannian structure in $\Omega$.

We claim that the first component of $F_2$ is identically equal to zero. Indeed, were this not the case, the norm of $F_1$ would not be equal to one.

We are left to prove (ii). We have

$$F_3 := [F_1, F_2] = \begin{pmatrix} 0 \\ \partial_{x_1}\mathfrak{f}(x_1, x_2) \end{pmatrix} \tag{9.7}$$

Notice that the only iterated brackets between $F_1, F_2$ that could be different from zero are of the form

$$\underbrace{[F_1, \ldots, [F_1,}_{r \text{ times}} F_2]] = \begin{pmatrix} 0 \\ \partial^r_{x_1}\mathfrak{f}(x_1, x_2) \end{pmatrix}.$$

Hence if the structure has step 1 at $q_0 = (0,0)$ we have $\mathfrak{f}(0,0) \neq 0$. If the structure has step $s$, with $s \geq 2$ at $q_0 = (0,0)$ we have $\mathfrak{f}(0,0) = 0$, $\partial_{x_1}^r \mathfrak{f}(0,0) = 0$ for $r = 1, 2, \ldots, s-2$ and $\partial_{x_1}^{s-1} \mathfrak{f}(0,0) \neq 0$. $\square$

*Remark* 9.12. Notice that, in the coordinate system constructed in the proof of Lemma 9.10 and Proposition 9.8, the submanifold $W$ has equation $W = \{(x_1, x_2) \in \Omega \mid x_1 = 0\}$.

Proposition 9.8 is very useful to express the Riemannian quantities on $M \setminus \mathcal{Z}$. In fact one has

**Lemma 9.13.** *Assume that on an open set $\Omega \subset M$ an orthonormal frame for the 2D almost-Riemannian is given in the form (9.2). Then on $\Omega \cap (M \setminus \mathcal{Z})$ the Riemannian metric, the element of Riemannian area and the Gaussian curvatures are given by*

$$g_{(x_1,x_2)} = \begin{pmatrix} 1 & 0 \\ 0 & \frac{1}{\mathfrak{f}(x_1,x_2)^2} \end{pmatrix}, \tag{9.8}$$

$$dA_{(x_1,x_2)} = \frac{1}{|\mathfrak{f}(x_1,x_2)|} dx_1\, dx_2, \tag{9.9}$$

$$K(x_1,x_2) = \frac{\mathfrak{f}(x_1,x_2)\partial_{x_1}^2 \mathfrak{f}(x_1,x_2) - 2\left(\partial_{x_1}\mathfrak{f}(x_1,x_2)\right)^2}{\mathfrak{f}(x_1,x_2)^2}. \tag{9.10}$$

*Proof.* Formula (9.8) is a direct consequence of (9.1). Formula (9.9) comes from the definition of the Riemannian area $dA(F_1, F_2) = 1$ where $\{F_1, F_2\}$ is a local orthonormal frame. Formula (9.10) comes from the formula (see Corollary 4.40):

$$K(q) = -\alpha_1^2 - \alpha_2^2 + F_1\alpha_2 - F_2\alpha_1,$$

where $\alpha_1$ and $\alpha_2$ are the two functions defined by $[F_1, F_2] = \alpha_1 F_1 + \alpha_2 F_2$. $\square$

Hence in a 2D almost-Riemannian structure all Riemannian quantities bow up when approaching $\mathcal{Z}$.

### 9.1.1 How big is the singular set?

A natural question is how big could be the singular set. The answer is given by the following Lemma.

**Theorem 9.14.** *Let $\mu$ be a smooth area on $M$. Then $\mathcal{Z}$ has zero $\mu$-measure.*

A direct consequence is the following statement.

**Corollary 9.15.** *Consider a system of coordinates $(x_1, x_2)$ defined on an open set $\Omega$ and let $dx_1\, dx_2$ be the corresponding Lebesgue area. Then $\mathcal{Z} \cap \Omega$ has zero $dx_1\, dx_2$-measure.*

As a consequence of Theorem 9.14, since $\mathcal{Z}$ is closed, we have the following.

**Corollary 9.16.** *For a 2D almost Riemannian structure, the set of Riemannian points is open and dense.*

*Proof of Theorem 9.14.* Let us cover $M$ with a countable union of open coordinate neighborhoods $\{U_i\}_{i\in I}$ having the following properties:

- each $U_i$ it is the product of two non-empty intervals:

$$U_i = (a_i, b_i) \times (c_i, d_i),$$

- on $U_i$ we have an orthonormal frame of the form

$$F_1^i(x_1, x_2) = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad F_2^i(x_1, x_2) = \begin{pmatrix} 0 \\ \mathfrak{f}_i(x_1, x_2) \end{pmatrix}, \tag{9.11}$$

- on $U_i$ the maximal step of the structure is $s_i$.

Moreover on $U_i$ let us write $\mu(x_1, x_2) = h_i(x_1, x_2)dx_1dx_2$, where $h_i$ is a smooth and never vanishing function. Since the measure is countably additive, it is enough to show that the $\mu$-area of each of $\mathcal{Z} \cap U_i$ is zero. For simplicity of notation, we prove then the statement for one of the open set $U_i$ of the covering, that we denote $U$, and we remove the $i$ from the notation.

Let $\mathbf{1}_{\mathcal{Z}} : M \to \{0, 1\}$ be the characteristic function of $\mathcal{Z}$. Using Fubini's theorem the $\mu$-area of $\mathcal{Z} \cap U$ is

$$\int_{\mathcal{Z}\cap U} h(x_1, x_2)dx_1dx_2 = \int_U \mathbf{1}_{\mathcal{Z}}(x_1, x_2)\, h(x_1, x_2)dx_1dx_2 = \int_c^d \left( \int_a^b \mathbf{1}_{\mathcal{Z}}(x_1, x_2)h(x_1, x_2)dx_1 \right) dx_2.$$

We now prove that for every fixed $\bar{x}_2 \in (c, d)$, we have $\int_a^b \mathbf{1}_{\mathcal{Z}}(x_1, \bar{x}_2)h(x_1, \bar{x}_2)dx_1 = 0$. Recall that $\mathcal{Z} = \{(x_1, x_2) \mid \mathfrak{f}(x_1, x_2) = 0\}$. By possibly restricting $U$, we have that (ii) of Proposition 9.8 guarantees that for every $x_1 \in (a, b)$ there exists $r(x_1) \leq s - 1$ such that $\partial_{x_1}^r \mathfrak{f}(x_1, \bar{x}_2) \neq 0$. Hence $\mathfrak{f}(\cdot, \bar{x}_2)$ has only isolated zeros and $\int_a^b \mathbf{1}_{\mathcal{Z}}(x_1, \bar{x}_2)h(x_1, x_2)dx_1 = 0$. □

**Exercise 9.17.** Use the proof of Theorem 9.14 to show that the singular set is locally the countable union of zero- and one-dimensional submanifolds and hence that it is rectifiable.

*Remark* 9.18. Notice that we cannot use the Riemannian area $dA$ to measure $\mathcal{Z}$ since $dA$ is not defined on $\mathcal{Z}$.

### 9.1.2 Genuinely 2D almost-Riemannian structures have always infinite area

**Theorem 9.19.** *Let $U$ be an open set such that $U \cap \mathcal{Z} \neq \emptyset$. Then*

$$\int_{U\setminus\mathcal{Z}} dA = +\infty,$$

*where $dA$ is the Riemannian area on $U \setminus \mathcal{Z}$ associated with the almost-Riemannian structure.*

To prove Theorem 9.19, we need the following preliminary results.

**Lemma 9.20.** *Let $\Omega$ be an open neighborhood of the origin in $\mathbb{R}$ and let $f \in C^\infty(\Omega, \mathbb{R})$ such that $f(0) = 0$. Then there exists $g \in C^\infty(\Omega, \mathbb{R})$ such that $f(x) = x\, g(x)$. Moreover $f'(0) \neq 0$ if and only if $g(0) \neq 0$,*

*Proof.* For $x \in \Omega$, let us define $h : [0, 1] \to \mathbb{R}$ by $h(t) := f(tx)$. We have

$$f(x) = f(x) - f(0) = h(1) - h(0) = \int_0^1 h'(t)dt = \int_0^1 f'(tx)x\, dt = x \int_0^1 f'(tx)\, dt.$$

Since the function $g(x) := \int_0^1 f'(tx)\, dt$ is smooth, the result follows. □

The extension of Lemma 9.20 to dimension 2 is straightforward.

**Lemma 9.21.** *Let $\Omega \subset \mathbb{R}^2$ be an open neighorhood of the origin and let $f \in C^\infty(\Omega, \mathbb{R})$ such that $f(0, x_2) = 0$ for every $x_2$. Then there exists $g \in C^\infty(\Omega, \mathbb{R})$ such that $f(x_1, x_2) = x_1 g(x_1, x_2)$. Moreover $\partial_{x_1} f(0,0) \neq 0$ if and only if $g(0,0) \neq 0$.*

*Proof of Theorem 9.19.* Take a point $q_0 \in U \cap \mathcal{Z}$. Thanks to Proposition 9.8, we can find a neighborhood $\Omega \subset U$ of $q_0$ and a system of coordinates $(x_1, x_2)$ in $\Omega$ such that $q_0 = (0,0)$ and an orthonormal frame for the 2D almost-Riemannian structure can be written in $\Omega$ as:

$$F_1(x_1, x_2) = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad F_2(x_1, x_2) = \begin{pmatrix} 0 \\ \mathfrak{f}(x_1, x_2) \end{pmatrix}, \quad \mathfrak{f}(0,0) = 0.$$

We have that

$$\int_{U \setminus \mathcal{Z}} dA \geq \int_{\Omega \setminus \mathcal{Z}} dA = \int_{\Omega \setminus \mathcal{Z}} \frac{1}{|\mathfrak{f}(x_1, x_2)|} dx_1 dx_2 =: I_\Omega(\mathfrak{f}). \tag{9.12}$$

We prove next that $I_\Omega(\mathfrak{f})$ is infinity. Let $\nabla \mathfrak{f}$ be the Euclidean gradient of $\mathfrak{f}$. We have two cases.

**Case 1.** $\nabla \mathfrak{f}(0,0) \neq 0$. In this case, possibly restricting $\Omega$, we can assume that $\mathcal{Z}|_\Omega = \{x \in \Omega \mid \mathfrak{f}(x) = 0\}$ is a submanifold of $\Omega$.

Up to a change of coordinates we can assume that $\mathcal{Z}|_\Omega = \{(x_1, x_2) \in \Omega \mid x_1 = 0\}$. By Lemma 9.21, we can write in $\Omega$

$$\mathfrak{f}(x_1, x_2) = x_1\, g(x_1, x_2),$$

where $g(x_1, x_2)$ is a smooth function, that we can assume (possibly restricting $\Omega$) to satisfy $M_1 \leq |g(x_1, x_2)| \leq M_2$ for some $0 < M_1 < M_2$. It follows that

$$I_\Omega(\mathfrak{f}) = \int_{\Omega \setminus \mathcal{Z}} \frac{1}{|x_1|} \frac{1}{|g(x_1, x_2)|} dx_1 dx_2 \geq \frac{1}{M_2} \int_{\Omega \setminus \mathcal{Z}} \frac{1}{|x_1|} dx_1 dx_2 = +\infty. \tag{9.13}$$

**Case 2.** $\nabla \mathfrak{f}(0,0) = 0$. In this case the Taylor expansion of $\mathfrak{f}$ (in both variables) at $x_0$ is

$$\mathfrak{f}(x_1, x_2) = b\, x_1^2 + c\, x_2^2 + d\, x_1 x_2 + O(\|x\|^3).$$

Here $\|x\|$ is the Euclidean norm of $(x_1, x_2)$. Hence in $\Omega \setminus \{(0,0)\}$ we can write

$$\mathfrak{f}(x_1, x_2) = \|x\|^2\, g(x_1, x_2),$$

where $g(x_1, x_2) := b\frac{x_1^2}{\|x\|^2} + c\frac{x_2^2}{\|x\|^2} + d\frac{x_1 x_2}{\|x\|^2} + O(\|x\|)$ is a smooth function defined in $\Omega \setminus \{(0,0)\}$, that we can assume (possibly restricting $\Omega$) to satisfy $|g(x_1, x_2)| \leq M$ for some $M > 0$. It follows that

$$I_\Omega(\mathfrak{f}) = \int_{\Omega \setminus \mathcal{Z}} \frac{1}{\|x\|^2} \frac{1}{|g(x_1, x_2)|} dx_1 dx_2 \geq \frac{1}{M} \int_{\Omega \setminus \mathcal{Z}} \frac{1}{\|x\|^2} dx_1 dx_2 = \frac{1}{M} \int_{\Omega \setminus \{x_0\}} \frac{1}{\|x\|^2} dx_1 dx_2 = +\infty. \tag{9.14}$$

Here we have used that $\mathcal{Z}$ has zero Lebesgue measure.

$\square$

*Remark* 9.22. Notice that if in Theorem 9.19 we take $U$ compactly contained in $M$ then $\mathrm{diam}(U) < +\infty$ and $\int_{U \setminus \mathcal{Z}} dA = +\infty$.

### 9.1.3 Pontryagin extremals

Since 2D almost-Riemannian structures are particular cases of sub-Riemannian structures, there are two kind of candidate optimal trajectories: normal and abnormal extremal trajectories.

If we fix a local orthonormal frame $\{F_1, F_2\}$, in an open set $\Omega$ and an admissible trajectory $q(\cdot)$ taking values in $\Omega$, by definition we have (cf. Theorems 4.20 and 4.25).

- $q(\cdot)$ is normal extremal trajectory if there exists a Lipschitz covector $p(\cdot)$ such that $(q(\cdot), p(\cdot))$ is solution to the Hamiltonian system corresponding to

$$H(q,p) = \frac{1}{2}\big(\langle p, F_1(q)\rangle^2 + \langle p, F_2(q)\rangle^2\big).$$

*Remark* 9.23. Notice that for a system of coordinates and a choice of an orthonormal frame as those of Proposition 9.8, we have

$$H(x_1, x_2, p_1, p_2) = \frac{1}{2}\left(p_1^2 + p_2^2\, \mathfrak{f}(x_1, x_2)^2\right). \tag{9.15}$$

- $q(\cdot)$ is an abnormal extremal trajectory if there exists a never vanishing Lipschitz covector $p(\cdot)$ such that

$$\langle p(t), F_1(q(t))\rangle \equiv 0, \quad \langle p(t), F_2(q(t))\rangle \equiv 0.$$

Recall that nontrivial normal extremal trajectories are geodesics while nontrivial abnormal extremal trajectories could be geodesics or not (see Section 4.7.2). For 2D almost-Riemannian structures, the situation is particularly simple.

**Theorem 9.24.** *For a 2D almost-Riemannian structure, an extremal trajectory $\gamma$ admits an abnormal lift if and only if $\gamma$ is a constant curve contained in $\mathcal{Z}$.*

*Proof.* It is immediate to verify that if $\gamma$ is a constant curve contained in $\mathcal{Z}$, then $\gamma$ admits an abnormal lift.

Let $\gamma : [a, b] \to M$, $(a < b)$ be the projection of an abnormal extremal and let us prove that $\gamma$ is a constant curve contained in $\mathcal{Z}$.

Let us first prove that $\gamma([a, b]) \subset \mathcal{Z}$. By contradiction assume that there exists $\bar{t} \in ]a, b[$ such that $\gamma(\bar{t}) \notin \mathcal{Z}$. By continuity there exists a nontrivial interval $[c, d] \subset ]a, b[$ such that $\gamma([c, d]) \cap \mathcal{Z} = \emptyset$. Then $\gamma_{[c,d]}$ is a Riemannian extremal trajectory and hence cannot be abnormal. Recall that if an arc of an extremal trajectory is not abnormal, then the extremal trajectory is not abnormal too, hence it follows that $\gamma$ is not abnormal. This contradicts the hypothesis that $\gamma$ is the projection of an abnormal extremal.

Let us now prove that $\gamma$ is a constant curve. Let us fix a local system of coordinates and an orthonormal frame as in Proposition 9.8. If this is not possible globally on a neighborhood of $\gamma([a, b])$, one can repeat the proof chart by chart. Let us write in coordinates $\gamma(t) = (\gamma_1(t), \gamma_2(t))$. We have different cases.

- If $(\gamma_1(t), \gamma_2(t)) = (c_1, c_2)$ for every $t \in [a, b]$ we already know that $\gamma$ admits an abnormal lift.
- If $\gamma_1$ is not constant and $\gamma_2 = c$ in $[a, b]$, then $\dot{\gamma}_2 = 0$ in $[a, b]$ and $\mathcal{Z}$ contains a set of the type

$$\bar{\mathcal{Z}} = \{(x_1, c) \mid x_1 \in [x_1^A, x_1^B]\} \text{ with } x_1^A < x_1^B.$$

Since $\mathfrak{f}=0$ on $\mathcal{Z}$ and hence on $\bar{\mathcal{Z}}$, it follows that $\partial_{x_1}^r \mathfrak{f} \equiv 0$ on $\bar{\mathcal{Z}}$, for every $r=1,2,\ldots$. As in the proof of Theorem 9.14, it follows that all brackets between $F_1$ and $F_2$ are zero on $\bar{\mathcal{Z}}$ and that the bracket-generating condition is violated. Hence this case is not possible.

- There exists $\bar{t} \in ]a,b[$ such that $\dot{\gamma}_2(\bar{t})$ is defined and $\dot{\gamma}_2(\bar{t}) \neq 0$. Now since

$$\dot{\gamma}(\bar{t}) = \begin{pmatrix} v_1 \\ v_2 \mathfrak{f}(\gamma(\bar{t})) \end{pmatrix},$$

for some $v_1, v_2 \in \mathbb{R}$, we have $\mathfrak{f}(\gamma(\bar{t})) \neq 0$ and hence $\gamma(\bar{t}) \notin \mathcal{Z}$ violating the condition $\gamma([a,b]) \subset \mathcal{Z}$. Hence this case is not possible as well.

□

As a consequence we have

**Corollary 9.25.** *For 2D almost-Riemannian structures, the set of geodesics coincides with the set of non-trivial normal Pontryagin extremals.*

As a consequence of the fact that normal Pontryagin extremals are projections of solutions of a smooth Hamiltonian system and of Corollary 9.25, we have

**Proposition 9.26.** *In 2D almost-Riemannian geometry all geodesics are smooth.*

Notice, moreover, that since our structure is Riemannian on $M \setminus \mathcal{Z}$, we have that almost-Riemannian geodesics coincide with Riemannian geodesics on each connected component of $M \setminus \mathcal{Z}$. The only particular property of almost-Riemannian geodesics is that on the singular set their velocity is constrained to belong to the distribution (otherwise their length is infinite). All this is illustrated in the next section for the Grushin plane.

## 9.2 The Grushin plane

The Grushin plane is the simplest example of genuinely almost-Riemannian structure. It is the free almost-Riemannian structure on $\mathbb{R}^2$ for which a global orthonormal frame is given by

$$F_1(x_1,x_2) = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad F_2(x_1,x_2) = \begin{pmatrix} 0 \\ x_1 \end{pmatrix}.$$

In the sense of Definition 9.1, it can be seen as the pair $(\mathbf{U}, f)$ where $\mathbf{U} = \mathbb{R}^2 \times \mathbb{R}^2$ and $f(x_1,x_2,u_1,u_2) = (x_1,x_2,u_1,u_2x_1)$.

Here the singular set $\mathcal{Z}$ is the $x_2$-axis (see Figure 9.2) and on $\mathbb{R}^2 \setminus \mathcal{Z}$ the Riemannian metric, the Riemannian area and the Gaussian curvature are given respectively by:

$$g = \begin{pmatrix} 1 & 0 \\ 0 & \frac{1}{x_1^2} \end{pmatrix}, \quad dA = \frac{1}{|x_1|} dx_1\, dx_2, \quad K = -\frac{2}{x_1^2}. \tag{9.16}$$

From the expression of $dA$, it follows that the (almost-Riemannian) area of an open set intersecting $\mathcal{Z}$ is always infinite. This was prescribed in full generality by Theorem 9.19.

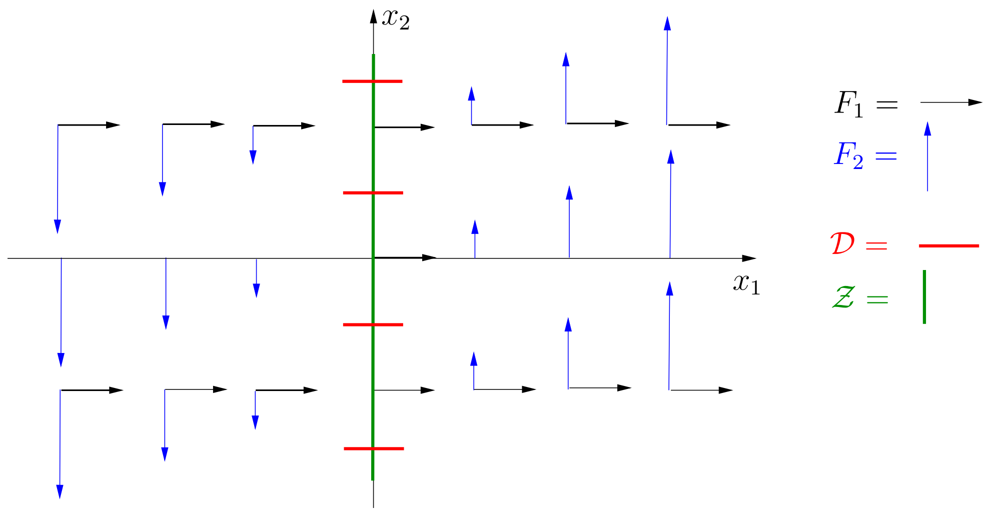



Figure 9.2: The Grushin plane

### 9.2.1 Geodesics on the Grushin plane

In this section we compute geodesics for the Grushin plane, with the purpose of stressing that they can cross the singular set with no singularities.

In this case the Hamiltonian (9.15) is given by

$$H(x_1, x_2, p_1, p_2) = \frac{1}{2}(p_1^2 + x_1^2 p_2^2) \tag{9.17}$$

and the corresponding Hamiltonian equations are:

$$\begin{aligned} \dot{x}_1 &= p_1, & \dot{p}_1 &= -x_1 p_2^2, \\ \dot{x}_2 &= x_1^2 p_2, & \dot{p}_2 &= 0. \end{aligned}$$

Arc length geodesics are projections on the $(x_1, x_2)$ plane of solutions of these equations, lying on the level set $H = 1/2$. We study arc length geodesics starting from: **i)** a point on $\mathcal{Z}$, e.g., $(0, 0)$; **ii)** a Riemannian point, e.g., $(-1, 0)$.

**Case** $(x_1(0), x_2(0)) = (0, 0)$
In this case the condition $H(x_1(0), x_2(0), p_1(0), p_2(0)) = 1/2$ implies that we have two families of arc length geodesics corresponding respectively to $p_1(0) = \pm 1$ and $p_2(0) =: a \in \mathbb{R}$. Their expression can be easily obtained and it is given by:

$$\begin{cases} x_1(t) = \pm t, & x_2(t) = 0, & \text{if } a = 0, \\ x_1(t) = \pm\dfrac{\sin(at)}{a}, & x_2(t) = \dfrac{2at - \sin(2at)}{4a^2}, & \text{if } a \neq 0. \end{cases} \tag{9.18}$$

Some geodesics are plotted in Figure 9.3 together with the "front" at time 1, i.e., the set of end-points of all arc length geodesics at time $t = 1$. Notice that all geodesics start horizontally. The particular form of the front shows the presence of a conjugate locus accumulating to the origin (compare also with the properties discussed in Chapter 19).

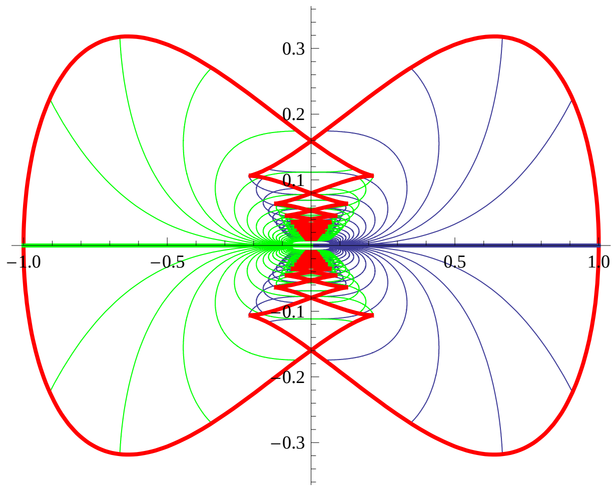



Figure 9.3: Arc length geodesics and the front for the Grushin plane, starting from the singular set.

**Case** $(x_1(0), x_2(0)) = (-1, 0)$
In this case the condition $H(x_1(0), x_2(0), p_1(0), p_2(0)) = 1/2$ becomes $p_1^2 + p_2^2 = 1$ and it is convenient to set $p_1 = \cos(\theta), p_2 = \sin(\theta)$, and $\theta \in S^1$. The expression of arc length geodesics is given by:

$$
\begin{cases}
x_1(t) = t - 1, \quad x_2(t) = 0, & \text{if } \theta = 0 \\[2ex]
x_1(t) = -t - 1, \quad x_2(t) = 0, & \text{if } \theta = \pi \\[2ex]
\left.\begin{array}{l} x_1(t) = -\dfrac{\sin(\theta - t\sin(\theta))}{\sin(\theta)}, \\[2ex] x_2(t) = \dfrac{2t - 2\cos(\theta) + \frac{\sin(2\theta - 2t\sin(\theta))}{\sin(\theta)}}{4\sin(\theta)} \end{array}\right\} & \text{if } \theta \notin \{0, \pi\}
\end{cases}
$$

Some arc length geodesics are plotted in Figure 9.4 together with the "front" at time $t = 4.8$. Notice that all geodesics pass horizontally through $\mathcal{Z}$, with no singularities. The particular form of the front shows the presence of a conjugate locus (compare again with Chapter 19). Geodesics may have conjugate times only after intersecting $\mathcal{Z}$. Before it is impossible since they are Riemannian and the curvature is negative, see Theorem 16.34.

The optimality of geodesics for the Grushin plane will be studied in Section 13.5.

## 9.3 Riemannian, Grushin and Martinet points

In 2D almost-Riemannian structures there are 3 kinds of important points: Riemannian, Grushin and Martinet points. As we are going to see in Section 9.4, these points are relevant in the following sense: if a system has only these type of points, then this remains true also after a small perturbation of it. Moreover arbitrarily close to any system there is a system where only these types points are present. Also we will see that Grushin points form 1D submanifolds of $M$ and Martinet points are isolated.

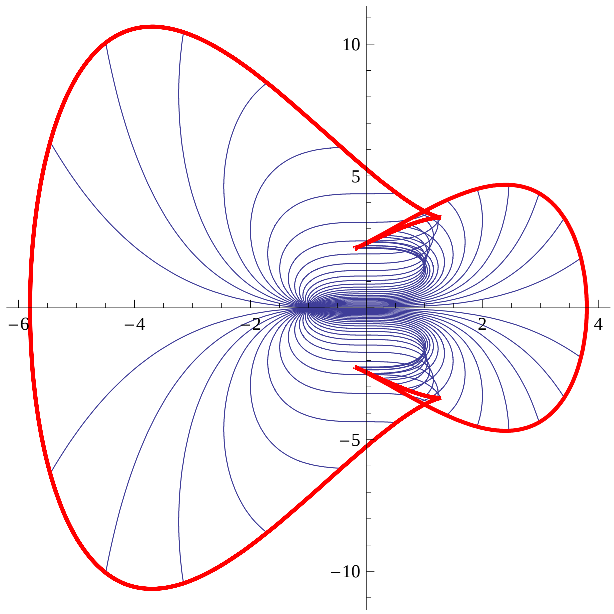



Figure 9.4: Arc length geodesics and the front for the Grushin plane, starting from a Riemannian point.

First we study under which conditions $\mathcal{Z}$ has the structure of a 1D submanifold of $M$. To this purpose we are going to study $\mathcal{Z}$ as the set of zeros of a function.

**Definition 9.27.** Let $\{F_1, F_2\}$ be a local orthonormal frame on an open set $\Omega$ and let $\omega$ be a volume form on $\Omega$. On $\Omega$ define the function $\Phi = \omega(F_1, F_2)$.

**Exercise 9.28.** Prove that $\Phi$ is invariant by a positively oriented change of orthonormal frame defined on the same open set $\Omega$.

Since a volume form can be globally defined when $M$ is orientable we have that $\Phi$ can be globally defined on fully orientable 2D almost-Riemannian structures (cf. Definition 9.3), just defining it as above on positively oriented orthonormal frames.

For structure that are not fully orientable, $\Phi$ can be defined only locally and up to a sign. (notice however that $|\Phi|$ is always well defined). This is what should be taken in mind every time that the function $\Phi$ appears in the following.

If in a system of coordinates $(x_1, x_2)$, we write

$$F_1 = \begin{pmatrix} F_1^1 \\ F_1^2 \end{pmatrix}, \quad F_2 = \begin{pmatrix} F_2^1 \\ F_2^2 \end{pmatrix}, \quad \omega(x_1, x_2) = h(x_1, x_2) dx_1 \wedge dx_2$$

then

$$\Phi(x_1, x_2) = h(x_1, x_2) \det \begin{pmatrix} F_1^1(x_1, x_2) & F_2^1(x_1, x_2) \\ F_1^2(x_1, x_2) & F_2^2(x_1, x_2) \end{pmatrix}.$$

*Remark* 9.29. For a system of coordinates and a choice of an orthonormal frame as those of Proposition 9.8, taking $\omega = dx_1 \wedge dx_2$, we have $\Phi(x_1, x_2) = \mathfrak{f}(x_1, x_2)$.

The function $\Phi$ permits to write

$$\mathcal{Z} = \{q \in M \mid \Phi(q) = 0\}.$$

We are now going to consider the following assumptions

$(\mathrm{H0})_{q_0}$: if $\Phi(q_0) = 0$, then $d\Phi(q_0) \neq 0$.

(H0): the condition $(\mathrm{H0})_{q_0}$ holds for every $q_0 \in M$.

**Exercise 9.30.** Prove that the above conditions are independent on the choice of the volume $\omega$.

By definition of submanifold, we have an immediate consequence.

**Proposition 9.31.** *Assume that* (H0) *holds. Then $\mathcal{Z}$ is a one-dimensional embedded submanifold of $M$.*

Recall that $\mathcal{D}_1 = \mathcal{D}$, $\mathcal{D}_{i+1} = \mathcal{D}_i + [\mathcal{D}_1, \mathcal{D}_i]$, for $i \geq 1$. We are now ready to define Riemannian, Grushin and Martinet points.

**Definition 9.32.** Consider a 2D almost Riemannian structure. Fix $q_0 \in M$.

- If $\mathcal{D}_1(q_0) = T_{q_0}M$ (equivalently if $q_0 \notin \mathcal{Z}$) we say that $q_0$ is a *Riemannian* point.
- If $\mathcal{D}_1(q_0) \neq T_{q_0}M$ (equivalently if $q_0 \in \mathcal{Z}$), then
  - if $\mathcal{D}_2(q_0) = T_qM$ we say that $q_0$ is a *Grushin* point.
  - if $\mathcal{D}_2(q_0) \neq T_qM$ and $(\mathrm{H0})_{q_0}$ holds we say that $q_0$ is a *Martinet* point.

*Remark* 9.33. Notice that at Riemannian, Grushin and Martinet points, the step of the structure is respectively $1, 2, s$ with $s \geq 3$.

Notice that at Riemannian points $(\mathrm{H0})_{q_0}$ is automatically satisfied. This is true also at Grushin points. Indeed using the normal form (9.2) and taking as volume form $dx \wedge dy$ one gets that $\Phi(x_1, x_2) = \mathfrak{f}(x_1, x_2)$. Hence the condition that the structure is step 2 at $q_0 = (0, 0)$ implies that $\partial_{x_1}\mathfrak{f}(0, 0) \neq 0$ hence $d\Phi(0, 0) \neq 0$. In other words we have the following.

**Proposition 9.34.** *Under* (H0)*, every point is either a Riemannian or a Grushin or a Martinet point.*

**Exercise 9.35.** By using the system of coordinate given by Proposition 9.8 prove the following:

(a) $q_0$ is a Grushin point if and only if $q_0 \in \mathcal{Z}$ and $L_v\Phi(q_0) \neq 0$ for $v \in \mathcal{D}(q)$, $\|v\| = 1$.

(b) $q_0$ is a Martinet point if and only if $q_0 \in \mathcal{Z}$, $d\Phi(q_0) \neq 0$, and for $v \in \mathcal{D}(q_0)$, $\|v\| = 1$, we have $L_v\Phi(q_0) = 0$.

The following proposition describes properties of Grushin and Martinet points (see Figure 9.5).

**Proposition 9.36.** *We have the following:*

*(i) $\mathcal{Z}$ is an embedded 1D submanifold of $M$ around Grushin or Martinet points;*

*(ii) if $q_0$ is a Grushin point then $\mathcal{D}(q_0)$ is transversal to $T_{q_0}\mathcal{Z}$;*

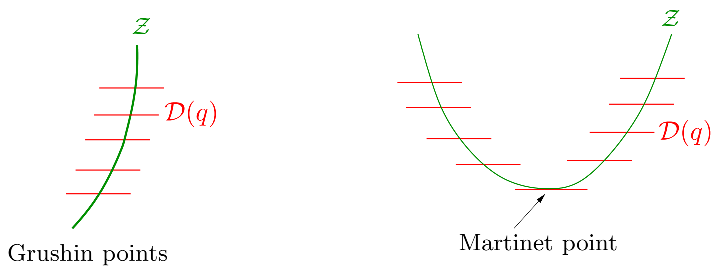



Figure 9.5: Grushin and Martinet points

*(iii) if $q_0$ is a Martinet point then $\mathcal{D}(q_0)$ coincides with $T_{q_0}\mathcal{Z}$;*

*Proof.* We use the system of coordinates and an orthonormal frame as those given by Proposition 9.8,

$$F_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad F_2 = \begin{pmatrix} 0 \\ \mathfrak{f} \end{pmatrix}.$$

If we take $\omega = dx \wedge dy$, we have $\Phi = \mathfrak{f}$, $d\Phi = (\partial_{x_1}\mathfrak{f}, \partial_{x_2}\mathfrak{f})$.

To prove (i), it is sufficient to notice that $d\Phi \neq 0$ at Grushin and Martinet points.

To prove (ii), notice that $\mathcal{D}(q_0) = \mathrm{span}(F_1(q_0)) = (1,0)$ while $T_{q_0}\mathcal{Z} = \mathrm{span}\{(-\partial_{x_2}\mathfrak{f}(q_0), \partial_{x_1}\mathfrak{f}(q_0))\}$ that are transversal since $\partial_{x_1}\mathfrak{f}(q_0) \neq 0$.

To prove (iii), notice that $\mathcal{D}(q_0) = \mathrm{span}(F_1(q_0)) = (1,0)$ while $T_{q_0}\mathcal{Z} = \mathrm{span}\{(-\partial_{x_2}\mathfrak{f}(q_0, 0)\}$ since the condition $\mathcal{D}_2(q_0) \neq T_{q_0}M$ implies $\partial_{x_1}\mathfrak{f}(q_0) = 0$.

$\square$

**Examples**

- All points on the $x_2$-axis for the Grushin plane are Grushin points.
- The origin of the following structure is the simplest example of Martinet point

$$F_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad F_2 = \begin{pmatrix} 0 \\ x_2 - x_1^2 \end{pmatrix}.$$

- The origin of the following example

$$F_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \text{ and } F_2 = \begin{pmatrix} 0 \\ x_2^2 - x_1^2 \end{pmatrix},$$

  is not a Martinet point since the condition $d\Phi(0,0) \neq 0$ is not satisfied. Outside the origin all points are either Riemannian or Grushin points, but at the origin $\mathcal{Z}$ is not a manifold.

- The $x_2$-axis of the following example

$$F_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \text{ and } F_2 = \begin{pmatrix} 0 \\ x_1^2 \end{pmatrix},$$

  is not made by Grushin points since $\mathcal{D}_2((0, x_2)) \neq T_{(0,x_2)}M$ and it is not made by Martinet points since $d\Phi(0, x_2) \neq 0$ is not satisfied (although in this case $\mathcal{Z}$ is a manifold). In this case $\mathcal{D}((0, x_2))$ is transversal to $\mathcal{Z}$.

### 9.3.1 Normal forms

**Theorem 9.37.** *Let $q_0$ be a Riemannian, Grushin or a Martinet point. There exists a neighborhood $\Omega$ of $q_0$ and a system of coordinates $(x_1, x_2)$ in $\Omega$ such that $q_0 = (0,0)$ an orthonormal frame for the 2D almost-Riemannian structure can be written in $\Omega$ as:*

*(NF1) if $q_0$ is a Riemannian point, then*

$$F_1(x_1, x_2) = (1, 0), \qquad F_2(x_1, x_2) = (0, e^{\phi(x_1,x_2)}),$$

*(NF2) if $q_0$ is a Grushin point, then*

$$F_1(x_1, x_2) = (1, 0), \qquad F_2(x_1, x_2) = (0, x_1 e^{\phi(x_1,x_2)}),$$

*(NF3) if $q_0$ is a Martinet point, then*

$$F_1(x_1, x_2) = (1, 0), \qquad F_2(x_1, x_2) = (0, (x_2 - x_1^{s-1}\psi(x_1))e^{\xi(x_1,x_2)}),$$

*where $\phi$, $\xi$ and $\psi$ are smooth real-valued functions such that $\phi(0, x_2) = 0$ and $\psi(0) \neq 0$. Moreover $s \geq 3$ is an integer, that is the step of the structure at the Martinet point.*

To prove Theorem 9.37 we need two preliminary results. The first is the following Lemma that can be obtained by induction from Lemma 9.20.

**Lemma 9.38.** *Let $\Omega$ be an open neighborhood of the origin in $\mathbb{R}$ and let $f \in C^\infty(\Omega, \mathbb{R})$. Assume there exists an integer $k$ such that $f(0) = f'(0) = \ldots = f^{(k-1)}(0) = 0$, and $f^{(k)}(0) \neq 0$. Then there exists $g \in C^\infty(\Omega, \mathbb{R})$ such that $f(x) = x^k g(x)$ with $g(0) \neq 0$.*

The second is the following Lemma that can be obtained from Lemma 9.21 with a change of variables.

**Lemma 9.39.** *Let $\Omega \subset \mathbb{R}^2$ be an open neighorhood of the origin and let $f \in C^\infty(\Omega, \mathbb{R})$ such that $f$ vanishes on the graph of a smooth function $x_1 = \Gamma(x_2)$ such that $\Gamma(0) = 0$. Then there exists $g \in C^\infty(\Omega, \mathbb{R})$ such that $f(x_1, x_2) = (x_1 - \Gamma(x_2))g(x_1, x_2)$. Moreover $\partial_{x_1} f(0,0) \neq 0$ if and only if $g(0,0) \neq 0$.*

*Remark* 9.40. Notice that if Lemma 9.39 applies, $\partial_{x_1} f(0,0) \neq 0$, and $f$ has no zeros out of the set $\{(x_1, x_2) \in \Omega \mid x_1 = \Gamma(x_2)\}$ then (up to restricting $\Omega$) $f$ can be written in the form $f(x_1, x_2) = \pm(x_1 - \Gamma(x_2))e^{\xi(x_1,x_2)}$, where $\xi$ is a smooth function defined on $\Omega$.

*Proof of Theorem 9.37.* By Proposition 9.8 we know that there exists a neighborhood $\Omega$ of $q_0$ and a system of coordinates $(x_1, x_2)$ in $\Omega$ such that $q_0 = (0,0)$ and an orthonormal frame for the 2D almost-Riemannian structure can be written in $\Omega$ as:

$$F_1(x_1, x_2) = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad F_2(x_1, x_2) = \begin{pmatrix} 0 \\ \mathfrak{f}(x_1, x_2) \end{pmatrix}, \tag{9.19}$$

where $\mathfrak{f} : \Omega \to \mathbb{R}$ is a smooth function.

Now if $q_0$ is a Riemannian point, we can assume (possibly restricting $\Omega$) $\mathfrak{f}(x_1, x_2) \neq 0$ for every $(x_1, x_2) \in \Omega$.

By applying a smooth coordinate transformation of the type $\bar{x}_1 = x_1$, $\bar{x}_2 = \nu(x_2)$ we get the new expressions for the vector fields

$$F_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad F_2 = \begin{pmatrix} 0 \\ \nu'(\nu^{-1}(\bar{x}_2))\mathfrak{f}(\bar{x}_1, \nu^{-1}(\bar{x}_2)) \end{pmatrix},$$

where $\nu'$ denotes the derivative of $\nu$. A normal form of type (NF1) is obtained choosing $\nu$ in such a way that $\nu'(x_2)\mathfrak{f}(0, x_2) = 1$ and setting $e^{\phi(\bar{x}_1,\bar{x}_2)} = \nu'(\nu^{-1}(\bar{x}_2))\mathfrak{f}(\bar{x}_1, \nu^{-1}(\bar{x}_2))$.

Now if $q_0$ is a Grushin point, we already know that $\mathcal{Z}$ is a 1D submanifold in a neighborhood of $q_0$ transversal to $\mathcal{D}(q_0) = \operatorname{span}\{(1, 0)\}$. If the normal form (9.19) is built using $\mathcal{Z}$ in place of $W$ (the 1D submanifold transversal to the distribution, cf. Lemma 9.10), we have that (possibly restricting $\Omega$), $\mathcal{Z} = \{(x_1, x_2) \in \Omega \mid x_1 = 0\}$. Hence the zeros of $\mathfrak{f}$ coincide with the $x_2$-axis. The condition that $q_0$ is a Grushin point implies that the step at $q_0$ is 2. Hence $\partial_{x_1}\mathfrak{f}(0, 0) \neq 0$. Using Lemma 9.21 (and possibly restricting $\Omega$), we have that $\mathfrak{f}$ admits a representation of the type $\mathfrak{f}(x_1, x_2) = \pm x_1 e^{\phi(x_1,x_2)}$, with $\phi$ smooth. The sign can be set to be $+$ since the vector fields of an orthonormal frame are defined up to their sign. Again, a change of coordinates $x_1 \to x_1$, $x_2 \to \nu(x_2)$ can be used in order to ensure that $\phi(0, x_2) = 0$. The normal form (NF2) is obtained.

If $q_0$ is a Martinet point we already know that $\mathcal{Z}$ is a manifold around $(0, 0)$ such that $T_{(0,0)}\mathcal{Z} = \operatorname{span}\{(1, 0)\}$. By possibly restricting $\Omega$, we can identify $\mathcal{Z}$ with the graph of a smooth function $x_2 = \Gamma(x_1)$. Using Lemma 9.39 and Remark 9.40 (with the change of notation $x_1 \leftrightarrow x_2$), $\mathfrak{f}(x_1, x_2)$ can be written in the form $(x_2 - \Gamma(x_1))e^{\xi(x_1,x_2)}$ with $\xi$ smooth. Denote by $s$ the step of the structure, and recall that at Martinet points $s \geq 3$. Then $f(0, 0) = 0$, $\partial_{x_1}\mathfrak{f}(0, 0) = 0, \ldots, \partial_{x_1}^{s-2}\mathfrak{f}(0, 0) = 0$, $\partial_{x_1}^{s-1}\mathfrak{f}(0, 0) \neq 0$. As a consequence $\Gamma(0) = 0, \partial_{x_1}\Gamma(0) = 0, \ldots, \partial_{x_1}^{s-2}\Gamma(0) = 0, \partial_{x_1}^{s-1}\Gamma(0) \neq 0$. Using Lemma 9.38, we can write $\Gamma(x_1) = x_1^{s-1}\psi(x_1)$ with $\psi(0) \neq 0$. The normal form (NF3) is obtained. □

We can now prove the following statement.

**Proposition 9.41.** *Martinet points are isolated.*

*Proof.* At Martinet points, we can use the normal form (NF3).

$$F_1(x_1, x_2) = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad F_2(x_1, x_2) = \begin{pmatrix} 0 \\ \mathfrak{f}(x_1, x_2) \end{pmatrix},$$
$$\text{where } \mathfrak{f} = (x_2 - x_1^{s-1}\psi(x_1))e^{\xi(x_1,x_2)}, \text{ with } s \geq 3 \text{ and } \psi(0) \neq 0.$$

In such normal form, the singular set is given by the equation $\mathcal{Z} = \{(x_1, x_2) \mid x_2 = x_1^{s-1}\psi(x_1)\}$. We are going to prove that there exists a neigborhood $U$ of $(0, 0)$ such that every point of $\bar{\mathcal{Z}} := U \cap \mathcal{Z} \setminus \{(0, 0)\}$ is a Grushin point, i.e., that on $\bar{\mathcal{Z}}$ we have that $F_3 := [F_1, F_2]$ is not parallel to $F_1$. We have

$$F_3 = (0, \partial_{x_1}\mathfrak{f}), \quad \text{where } \partial_{x_1}\mathfrak{f}(x_1, x_2) = e^{\xi}\big((x_2 - x_1^{s-1}\psi(x_1))\partial_{x_1}\xi - ((s-1)x_1^{s-2}\psi + x_1^{s-1}\psi')\big).$$

Let us see that this cannot be zero on $\bar{\mathcal{Z}}$ for $U$ sufficiently small. Indeed on $\bar{\mathcal{Z}}$ we have that

$$\partial_{x_1}\mathfrak{f}(x_1, x_2) = -e^{\xi}x_1^{s-2}\big((s-1)\psi + x_1\psi'\big),$$

that for $x_1 \neq 0$, sufficiently small, is not zero since $\psi(0) \neq 0$. □

## 9.4 Generic 2D almost-Riemannian structures

Recall hypothesis $(\mathrm{H0})_{q_0}$ and (H0):

$(\mathrm{H0})_{q_0}$: If $\Phi(q_0) = 0$ then $d\Phi(q_0) \neq 0$.

(H0): The condition $(\mathrm{H0})_{q_0}$ holds for every $q_0 \in M$.

Recall that hypotheses $(\mathrm{H0})_{q_0}$ and (H0) are independent from the volume form used to define the function $\Phi$. We have seen (cf. Proposition 9.34) that under hypothesis (H0) every point is either a Riemannian or a Grushin or a Martinet point. In this section we are going to prove that hypothesis (H0) holds for most of the systems. More precisely we are going to prove that hypothesis (H0) is generic in the following sense.

**Definition 9.42.** Fix a rank 2 Euclidean bundle $\mathbf{U}$ over a 2D *compact* manifold $M$. Let $\mathcal{F}$ be the set of all morphism of bundle from $\mathbf{U}$ to $TM$ such that $(\mathbf{U}, f)$, $f \in \mathcal{F}$ is a 2D almost-Riemannian structure. Endow $\mathcal{F}$ with the $C^1$ norm. We say that a subset $\mathcal{F}'$ of $\mathcal{F}$ is *g*eneric if it is open and dense in $\mathcal{F}$.

**Theorem 9.43.** *Under the same hypothesis of Definition 9.42, let $\mathcal{F}' \subset \mathcal{F}$ the subset of morphisms satisfying* (H0)*. Then $\mathcal{F}'$ is generic.*

*Remark* 9.44. In Theorem 9.43 we have assumed that $M$ is compact. A similar result holds also in the case in which $M$ is not compact. However, in the non compact case, one should use a suitable topology (Whitney's one) and one gets that $\mathcal{F}'$ is a countable union of open and dense subsets of $\mathcal{F}$. In this book we have decided not to enter inside transversality theory and we have provided a statement that can be proved easily via the Sard lemma.

*Remark* 9.45. As a consequence of Theorem 9.43, and of Proposition 9.34, a generic 2D almost-Riemannian structure has only Riemannian, Grushin and Martinet points.

### 9.4.1 Proof of the genericity result

Recall that, fixed $M$ and $\mathbf{U}$, any morphism any $f : \mathbf{U} \to TM$ such that $(\mathbf{U}, f)$, is a 2D almost-Riemannian structure can be seen locally as the data of two vector fields representing an orthonormal frame. Hence cover $M$ with a finite number of *compact* coordinate neighborhood $\mathcal{U}_i$, $i = 1 \dots N$, in such a way that an orthonormal frame for the almost-Riemannian structure in $\mathcal{U}_i$ is given by

$$F_i(x_1^i, x_2^i) = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad G_i(x_1^i, x_2^i) = \begin{pmatrix} 0 \\ \mathfrak{f}_i(x_1, x_2) \end{pmatrix}. \tag{9.20}$$

Let us consider the following hypothesis

$(\mathrm{H0})^i$: The condition $(\mathrm{H0})_{q_0}$ holds for every $q_0 \in \mathcal{U}_i$.

**Proposition 9.46.** *Let $\mathcal{F}_i$ be the subset of $\mathcal{F}$ satisfying* $(\mathrm{H0})^i$*. Then $\mathcal{F}_i$ is generic.*

Once Proposition 9.46 is proved, the conclusion of Theorem 9.43 follows immediately. Indeed $\mathcal{F}_i$ is open and dense in $\mathcal{F}$ and the open and dense set $\mathcal{F}' := \cap_{i=1}^N \mathcal{F}_i$ is made by systems satisfying (H0) in the whole $M$.

*Proof of Proposition 9.46.* Since the map that to $(F_i, G_i)$ associates $\Phi$ is continuous in the $C^1$ topology, a small perturbation of $(F_i, G_i)$ will induce a small perturbation of $\Phi$. Hence, fixed $q_0$, the condition $(\mathrm{H0})_{q_0}$ is open in the set of pair of vector fields defined in $\mathcal{U}_i$ for the $C^1 \times C^1$ topology. As a consequence of the compactness of $\mathcal{U}_i$, condition $(\mathrm{H0})^i$ is open as well.

We are now going to prove that $(\mathrm{H0})^i$ is dense. To this purpose we construct an arbitrarily small perturbation in the $C^1$ norm $(F_i^\varepsilon, G_i^\varepsilon)$ of $(F_i, G_i)$ for which $(\mathrm{H0})^i$ is satisfied.

**Lemma 9.47.** *There exists $C > 0$ such that for every $\varepsilon \in \mathbb{R}$, there exists a perturbation $(F_i^\varepsilon, G_i^\varepsilon)$ of $(F_i, G_i)$ such that $\|F_i^\varepsilon - F_i\|_{C^1} \leq C|\varepsilon|$, $\|G_i^\varepsilon - G_i\|_{C^1} \leq C|\varepsilon|$ and on $\mathcal{U}_i$ we have $\Phi_\varepsilon := \omega(F_i^\varepsilon, G_i^\varepsilon) = \Phi + \varepsilon$.*

Once Lemma 9.47 is proved, the density of $\mathcal{F}_i$ follows easily. Indeed let now apply the Sard Lemma to the $C^\infty$ function $\Phi$ in $\mathcal{U}_i$. We have that the set

$$\{c \in \mathbb{R} \text{ such that there exists } q \in \mathcal{U}_i \text{ such that } \Phi(q) = c \text{ and } d\Phi(q) = 0\}$$

has measure zero. As a consequence, since $\Phi_\varepsilon = \Phi + \varepsilon$, we have that the set

$$\{\varepsilon \in \mathbb{R} \text{ such that there exists } q \in \mathcal{U}_i \text{ such that } \Phi_\varepsilon(q) = 0 \text{ and } d\Phi_\varepsilon(q) = 0\}$$

has measure zero. It follows that, for almost every $\varepsilon$, condition $(\mathrm{H0})^i$ is realized for $(F_i^\varepsilon, G_i^\varepsilon)$. □

*Proof of Lemma 9.47.* In $\mathcal{U}_i$ let us write in coordinates

$$\omega = h_i(x_1^i, x_2^i)dx_1^i \wedge dx_2^i.$$

(Recall that $\omega$ is a volume form and hence $h_i$ is never vanishing.) Then

$$\Phi = \omega(F_i, G_i) = h_i(x_1^i, x_2^i)\mathfrak{f}_i(x_1^i, x_2^i).$$

Consider now a perturbation $G_i^\varepsilon$ of $G_i$ of the form

$$G_i^\varepsilon(x_1^i, x_2^i) = \begin{pmatrix} 0 \\ \mathfrak{f}_i(x_1^i, x_2^i) + \frac{\varepsilon}{h_i(x_1^i, x_2^i)} \end{pmatrix}. \tag{9.21}$$

and let us define $F_i^\varepsilon = F_i$. It follows that in $\mathcal{U}_i$,

$$\Phi_\varepsilon = \omega(F_i^\varepsilon, G_i^\varepsilon) = h_i(x_1^i, x_2^i)\left(\mathfrak{f}_i(x_1^i, x_2^i) + \frac{\varepsilon}{h_i(x_1^i, x_2^i)}\right) = h_i(x_1^i, x_2^i)\mathfrak{f}_i(x_1^i, x_2^i) + \varepsilon = \Phi + \varepsilon.$$

The Lemma is proved with $C = \|1/h_i\|_{C^1}$. □

## 9.5 A Gauss-Bonnet theorem

For a compact orientable 2D-Riemannian manifold, the classical Gauss-Bonnet theorem asserts that the integral of the curvature is a topological invariant that is the Euler characteristic of the manifold (see Section 1.3).

This theorem admits an interesting generalization in the context of 2D almost-Riemannian structures that are fully orientable. This generalization is nontrivial since one needs to integrate the Gaussian curvature (that in general is diverging while approaching to the singular set) on the manifold (that has always infinite volume).

This generalization holds under certain natural assumptions on the 2D almost-Riemannian structure, that we sum up in the following condition.

(HG) : The base manifold $M$ is compact. The 2D almost-Riemannian structure is fully orientable and every point of $\mathcal{Z}$ is a Grushin point.

The hypothesis that the structure is fully orientable is crucial and it is the almost-Riemannian version of the classical orientability hypothesis that one needs in Riemannian geometry. The hypothesis that every point is a Grushin point is a technical and can indeed be removed. A version of a Gauss-Bonnet theorem in presence of Martinet points holds true. However its formulation and proof is more complicated and outside the purpose of this book. See the Bibliographical note.

Notice that (HG) implies (H0). With an argument similar to the one of the beginning of Section 9.4.1, one gets the following.

**Theorem 9.48.** *Hypothesis* (HG) *is open in the set of morphisms $f : \mathbf{U} \to TM$ (such that $(\mathbf{U}, f)$ is a 2D almost-Riemannian structure) endowed with $C^1$ topology.*

Hypothesis (HG) is not dense. Indeed it is not hard to build Martinet points that do not disappear for small $C^1$ perturbations of the system.

It is important to notice that (HG) is not empty. Indeed we have

**Proposition 9.49.** *Every oriented compact surface can be endowed with a free almost-Riemannian structure such that every point of $\mathcal{Z}$ is a Grushin point.*

We are going to prove Proposition 9.49 in Section 9.5.4.

### 9.5.1 Integration of the curvature

**Definition 9.50.** Consider a fully orientable 2D almost-Riemannian structure $(\mathbf{U}, f)$ over $M$ and assume that (HG) holds.

Let $\nu$ a volume form for the Euclidean structure on $\mathbf{U}$, i.e., a never vanishing 2-form s.t. $\nu(\sigma_1, \sigma_2) = 1$ on every positively oriented local orthonormal frame for $(\cdot \mid \cdot)_q$. Let $\Xi$ be an orientation on $M$. We introduce the following objects:

- The *signed area form* $dA^s$ on $M$ as the two-form on $M \setminus \mathcal{Z}$ given by the pushforward of $\nu$ along $f$. Notice that the Riemannian area $dA$ on $M \setminus \mathcal{Z}$ is the density associated with the volume form $dA^s$.
- $M^+ = \{q \in M \setminus \mathcal{Z}$, s.t. the orientation given by $dA^s_q$ and $\Xi_q$ are the same $\}$.[1]
- $M^- = \{q \in M \setminus \mathcal{Z}$, s.t. the orientation given by $dA^s_q$ and $\Xi_q$ are opposite $\}$.

Notice that given a measurable function $h : \Omega \subset M^\pm \to \mathbb{R}$, we have

$$\int_\Omega h \, dA^s = \pm \int_\Omega h \, dA \quad \text{(if it exists).} \tag{9.22}$$

[1] i.e., $dA^s_q(F_1, F_2) = \alpha \Xi(F_1, F_2)$ with $\alpha > 0$

**Definition 9.51.** Under the same hypotheses of Definition 9.50, define

- $M_\varepsilon = \{q \in M \mid d(q, \mathcal{Z}) > \varepsilon\}$ where $d(\cdot,\cdot)$ is the 2D almost-Riemannian structure on $M$.
- $M_\varepsilon^\pm = M_\varepsilon \cap M^\pm$.
- Given a measurable function $h : M \setminus \mathcal{Z} \to \mathbb{R}$, we say that it is *AR-integrable* if

$$\lim_{\varepsilon \to 0} \int_{M_\varepsilon} h \; dA^s \tag{9.23}$$

  exists and is finite. In this case we denote such a limit by $\int h \, dA^s$.

*Remark* 9.52. Notice that (9.23) coincides with $\lim_{\varepsilon\to 0}\left(\int_{M_\varepsilon^+} h \; dA - \int_{M_\varepsilon^-} h \; dA\right)$.

### 9.5.2 The Euler number

We now define the Euler number of a fully orientable 2D almost-Riemannian structure on a compact manifold. It measure how the vector bundle $\mathbf{U}$ is far from the trivial one.

**Definition 9.53.** The *Euler number* a fully orientable 2D almost-Riemannian structure $(\mathbf{U}, f)$ on a compact manifold $M$ is the Euler number $e(\mathbf{U})$ of $\mathbf{U}$. It is the self-intersection number of $M$ in $\mathbf{U}$, where $M$ is identified with the zero section. To compute $e(\mathbf{U})$, consider a smooth section $\sigma : M \to \mathbf{U}$ transverse to the zero section. Then, by definition,

$$e(\mathbf{U}) = \sum_{p \mid \sigma(p)=0} i(p, \sigma),$$

where $i(p,\sigma) = 1$, respectively $-1$, if $d_p\sigma : T_pM \to T_{\sigma(p)}\mathbf{U}$ preserves, respectively reverses, the orientation. Notice that if we reverse the orientation on $M$ or on $\mathbf{U}$ then $e(\mathbf{U})$ changes sign. Hence, the Euler number of an orientable vector bundle $\mathbf{U}$ is defined up to a sign, depending on the orientations of both $\mathbf{U}$ and $M$.

*Remark* 9.54. Take a section $\sigma$ of $\mathbf{U}$ that has only isolated zeros, i.e., such that the set $\{p \mid \sigma(p) = 0\}$ is finite. Since $\mathbf{U}$ is endowed with a smooth scalar product $(\cdot \mid \cdot)_q$ we can define $\tilde\sigma : M \setminus \{p \mid \sigma(p) = 0\} \to S\mathbf{U}$ by $\tilde\sigma(q) = \frac{\sigma(q)}{\sqrt{(\sigma \mid \sigma)_q}}$ (here $S\mathbf{U}$ denotes the spherical bundle of $\mathbf{U}$). If $\sigma(p) = 0$, then $i(p, \tilde\sigma) = i(p, \sigma)$ is equal to the degree of the map $\partial B \to S^1$ that associates with each $q \in \partial B$ the value $\tilde\sigma(q)$, where $B$ is a neighborhood of $p$ diffeomorphic to an open ball in $\mathbb{R}^n$ that does not contain any other zero of $\sigma$.

Notice that if $i(p,\sigma) \neq 0$, the limit $\lim_{q\to p} \tilde\sigma(q)$ does not exist.

*Remark* 9.55. Notice that $\mathbf{U}$ is trivial if and only if $e(\mathbf{U}) = 0$.

*Remark* 9.56. Since reversing the orientation on $M$ also reverses the orientation of $TM$, the Euler number of $TM$ is defined unambiguously.

**Exercise 9.57.** Let $M$ be a compact orientable surface. Prove that the Euler number of $TM$ is equal to $\chi(M)$, the Euler characteristic of $M$ as defined in (1.42). Hint: take a surface of genus $g$ and construct a vector field for which it is easy to compute the sum of the indices of its zeros. A possibility is to project orthogonally on the surface a constant vector field in $\mathbb{R}^3$. Recall that the sum of the indices of the zeros a vector field on a surface is independent from the vector field itself.

*Remark* 9.58. Consider a 2D almost-Riemannian structure $(\mathbf{U}, f)$ on a 2D manifold $M$. Let $\sigma$ be a section of $\mathbf{U}$ and $\mathbf{z}_\sigma$ the set of its zeros. As in Remark 9.54, define on $M \setminus \mathbf{z}_\sigma$ the normalization $\tilde{\sigma}$ of $\sigma$ and let $\tilde{\sigma}^\perp$ (still defined on $M \setminus \mathbf{z}_\sigma$) its orthogonal with respect to $(\cdot \mid \cdot)_q$. Then the original structure is free when restricted to $M \setminus \mathbf{z}_\sigma$ and $\{\tilde{\sigma}, \tilde{\sigma}^\perp\}$ is a global orthonormal frame for $(\cdot \mid \cdot)_q$. The global orthonormal frame for the corresponding 2D almost-Riemannian structure is then $(f \circ \tilde{\sigma}, f \circ \tilde{\sigma}^\perp)$.

### 9.5.3 Gauss-Bonnet theorem

The main result of this section is the following.

**Theorem 9.59.** *Consider a 2D almost-Riemannian structure satisfying hypothesis* (HG)*. Let $dA^s$ be the signed area form and $K$ be the Riemannian curvature, both defined on $M \setminus \mathcal{Z}$. Then $K$ is AR-integrable and we have*

$$\int K \, dA^s = 2\pi e(\mathbf{U}),$$

*where $e(\mathbf{U})$ denotes the Euler number of the 2D almost-Riemannian structure. Moreover we have*

$$e(\mathbf{U}) = \chi(M^+) - \chi(M^-),$$

*where $\chi(M^\pm)$ denotes the Euler characteristic of $M^\pm$.*

Notice that in the Riemannian case $\int K \, dA^s$ is the standard integral of the Riemannian curvature and $e(\mathbf{U}) = \chi(M)$ since $\mathbf{U} = TM$. Hence Theorem 9.59 contains the classical Gauss-Bonnet theorem for compact surfaces (cf. Section 1.3).

In a sense, in Riemannian geometry the topology of the surface gives a constraint on the total curvature, while in 2D almost-Riemannian geometry such constraint is determined by the topology of the bundle $\mathbf{U}$.

For a free almost-Riemannian structure we have that $\mathbf{U}$ is a rank 2 trivial bundle over $M$. As a consequence we get that $\int K \, dA^s = 0$, generalizing the Riemannian Gauss-Bonnet theorem on the torus. We could interpret this result in the following way. Take a metric that is determined by a single pair of vector fields. In the Riemannian context $M$ is forced to be parallelizable (i.e., $M$ must be the torus). In the AR context, $M$ could be any compact orientable manifold, but the metric is forced to be singular somewhere. In any case, the integral of the curvature is zero.

#### Proof of Theorem 9.59

The proof is divided in two steps. First we prove that $\int K \, dA^s = \chi(M^+) - \chi(M^-)$. Then we prove that $e(\mathbf{U}) = \chi(M^+) - \chi(M^-)$.

#### Step 1

As a consequence of the compactness of $M$ and of Lemma 9.10 one has:

**Lemma 9.60.** *Assume that* (HG) *holds. Then the set $\mathcal{Z}$ is the union of finitely many curves diffeomorphic to $S^1$. Moreover, there exists $\varepsilon_0 > 0$ such that, for every $0 < \varepsilon < \varepsilon_0$, we have that $\partial M_\varepsilon$ is smooth and the set $M \setminus M_\varepsilon$ is diffeomorphic to $\mathcal{Z} \times [0, 1]$.*

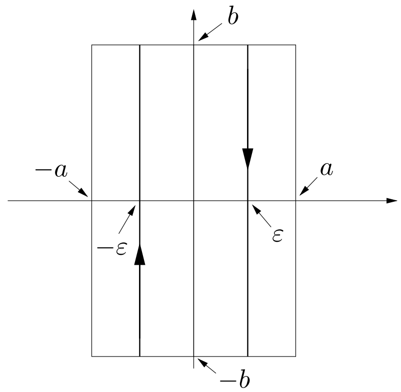



Figure 9.6: Proof of Step 1 of the Gauss-Bonnet theorem

Under (HG) the almost-Riemannian structure can be described, around each point of $\mathcal{Z}$, by a normal form of type (NF2).

Take $\varepsilon_0$ as in the statement of Lemma 9.60. For every $\varepsilon \in (0, \varepsilon_0)$, let $M_\varepsilon^\pm = M^\pm \cap M_\varepsilon$. By definition of $dA^s$ and $M^\pm$,

$$\int_{M_\varepsilon} K dA^s = \int_{M_\varepsilon^+} K dA - \int_{M_\varepsilon^-} K dA.$$

The classical Gauss-Bonnet theorem (see Theorem 1.37) asserts that for every compact oriented Riemannian manifold $(N, g)$ with smooth boundary $\partial N$, we have

$$\int_N K dA + \int_{\partial N} k_g ds = 2\pi\chi(N), \tag{9.24}$$

where $K$ is the curvature of $(N, g)$, $dA$ is the Riemannian density, $k_g$ is the geodesic curvature of $\partial N$ (whose orientation is induced by the one of $N$), and $ds$ is the length element.

Applying (9.24) to the Riemannian manifolds $(M_\varepsilon^+, g)$ and $(M_\varepsilon^-, g)$ (the smoothness of the boundary is guaranteed by Lemma 9.60), we have

$$\int_{M_\varepsilon} K dA^s = 2\pi(\chi(M_\varepsilon^+) - \chi(M_\varepsilon^-)) - \int_{\partial M_\varepsilon^+} k_g ds + \int_{\partial M_\varepsilon^-} k_g ds. \tag{9.25}$$

Thanks again to Lemma 9.60, $\chi(M_\varepsilon^\pm) = \chi(M^\pm)$. We are left to prove that

$$\lim_{\varepsilon \to 0} \left( \int_{\partial M_\varepsilon^+} k_g ds - \int_{\partial M_\varepsilon^-} k_g ds \right) = 0. \tag{9.26}$$

Fix $q \in \mathcal{Z}$ and a (NF2)-type local system of coordinates $(x_1, x_2)$ in a neighborhood $U_q$ of $q$. We can assume that $U_q$ is given, in the coordinates $(x_1, x_2)$, by a rectangle $[-a, a] \times [-b, b]$, $a, b > 0$. Assume that $\varepsilon < a < \varepsilon_0$. Notice that $\mathcal{Z} \cap U_q = \{0\} \times [-b, b]$ and $\partial M_\varepsilon \cap U_q = \{-\varepsilon, \varepsilon\} \times [-b, b]$.

We are going to prove that

$$\int_{\partial M_\varepsilon^+ \cap U_q} k_g ds - \int_{\partial M_\varepsilon^- \cap U_q} k_g ds = O(\varepsilon). \tag{9.27}$$

Then (9.26) follows from the compactness of $\mathcal{Z}$. (Indeed, $[-a,a]\times\{-b\}$ and $[-a,a]\times\{b\}$, the horizontal edges of $\partial U_q$, are the support of normal extremals trajectories minimizing the length from $\mathcal{Z}$. Therefore, $\mathcal{Z}$ can be covered by a finite number of neighborhoods of type $U_q$ whose pairwise intersections have empty interior.)

Without loss of generality, we can assume that $M^+\cap U_q=(0,a]\times[-b,b]$. Therefore, $M_\varepsilon^+$ induces on $\partial M_\varepsilon^+=\{\varepsilon\}\times[-b,b]$ a downward orientation (see Figure 9.6). First notice that the $k_g=0$ on the two segments $(0,\varepsilon]\times\{-b\}$ and $(0,\varepsilon]\times\{b\}$ (since they are the support of geodesics). The curve $s\mapsto c(s)=(\varepsilon,x_2(s))$ satisfying

$$\dot c(s)=-F_2(c(s))\,,\quad c(0)=(\varepsilon,0)\,,$$

is an oriented parametrization by arc length of $\partial M_\varepsilon^+$, making a constant angle with $F_1$. Let $\{\theta_1,\theta_2\}$ be the dual basis to $\{F_1,F_2\}$ on $U_q\cap M^+$, i.e., $\theta_1=dx_1$ and $\theta_2=x_1^{-1}e^{-\phi(x_1,x_2)}dx_2$. According to the results of Section 4.4.1 (cf. in particular Remark 4.41), the geodesic curvature of $\partial M_\varepsilon^+$ at $c(s)$ is equal to $\eta(\dot c(s))$, where $\eta\in\Lambda^1(U_q)$ is the unique one-form satisfying

$$d\theta_1=\eta\wedge\theta_2\,,\quad d\theta_2=-\eta\wedge\theta_1\,.$$

A simple computation shows that

$$\eta=\partial_{x_1}(x_1^{-1}e^{-\phi(x_1,x_2)})dx_2\,.$$

Thus (recall that $c(s)$ makes a constant angle with $F_1$ hence $\dot\theta=0$ in Remark 4.41)

$$k_g(c(s))=-\partial_{x_1}(x_1^{-1}e^{-\phi(c(s))})\,(dx_2(F_2))(c(s))=\frac{1}{\varepsilon}+\partial_{x_1}\phi(\varepsilon,x_2(s))\,.$$

Denote by $L_1$ and $L_2$ the almost-Riemannian lengths of the segments $\{\varepsilon\}\times[0,b]$ and $\{\varepsilon\}\times[-b,0]$, respectively. Then,

$$\begin{aligned}\int_{\partial M_\varepsilon^+\cap U_q}k_g ds &= \int_{-L_1}^{L_2}k_g(c(s))ds\\ &= \int_{-L_1}^{L_2}\left(\frac{1}{\varepsilon}+\partial_{x_1}\phi(\varepsilon,x_2(s))\right)ds\\ &= \int_{-b}^{b}\left(\frac{1}{\varepsilon}+\partial_{x_1}\phi(\varepsilon,x_2)\right)\frac{1}{\varepsilon e^{\phi(\varepsilon,x_2)}}dx_2\,,\end{aligned}$$

where the last equality is obtained taking $x_2=x_2(-s)$ as the new variable of integration.

With a similar reasoning on $\partial M_\varepsilon^-\cap U_q$, on which $M_\varepsilon^-$ induces the upward orientation. An orthonormal frame on $M^-\cap U_q$, oriented consistently with $M$, is given by $\{F_1,-F_2\}$, whose dual basis is $(\theta_1,-\theta_2)$. The same computations as above lead to

$$\int_{\partial M_\varepsilon^-\cap U_q}k_g ds=\int_{-b}^{b}\left(\frac{1}{\varepsilon}-\partial_{x_1}\phi(-\varepsilon,x_2)\right)\frac{1}{\varepsilon e^{\phi(-\varepsilon,x_2)}}dx_2\,.$$

Define

$$F(\varepsilon,x_2)=(1+\varepsilon\partial_{x_1}\phi(\varepsilon,x_2))e^{-\phi(\varepsilon,x_2)}.\tag{9.28}$$

Then
$$\int_{\partial M_\varepsilon^+\cap U_q} k_g ds - \int_{\partial M_\varepsilon^-\cap U_q} k_g ds = \frac{1}{\varepsilon^2}\int_{-b}^{b} (F(\varepsilon,x_2)-F(-\varepsilon,x_2))\, dx_2.$$
By Taylor expansion with respect to $\varepsilon$ we get
$$F(\varepsilon,x_2)-F(-\varepsilon,x_2) = 2\partial_\varepsilon F(0,x_2)\varepsilon + O(\varepsilon^3) = O(\varepsilon^3)$$
where the last equality follows from the relation $\partial_\varepsilon F(0,x_2)=0$ (see equation (9.28)). Therefore,
$$\int_{\partial M_\varepsilon^+\cap U_q} k_g ds - \int_{\partial M_\varepsilon^-\cap U_q} k_g ds = O(\varepsilon),$$
and (9.27) is proved.

**Step 2**

The idea of the proof is to find a section $\sigma$ of $S\mathbf{U}$ (the spherical bundle of $\mathbf{U}$) with isolated singularities $p_1,\ldots,p_m$ such that $\sum_{j=1}^m i(p_j,\sigma)=\chi(M^+)-\chi(M^-)$. In the sequel, we consider $\mathcal{Z}$ to be oriented with the orientation induced by $M^+$.

We start by defining $\sigma$ on a neighborhood of $\mathcal{Z}$. Let $W$ be a connected component of $\mathcal{Z}$. Since $M$ is oriented, there exists an open tubular neighborhood $\mathbf{W}$ of $W$ and a diffeomorphism $\Psi : S^1\times(-1,1)\to\mathbf{W}$ that preserves the orientation and $\Psi|_{S^1\times\{0\}}$ is an orientation-preserving diffeomorphism between $S^1$ and $W$. Remark that $f:\mathbf{U}|_{\mathbf{W}^+}\to T\mathbf{W}^+$ is an orientation-preserving isomorphism of vector bundles, while $f:\mathbf{U}|_{\mathbf{W}^-}\to T\mathbf{W}^-$ is an orientation-reversing isomorphism of vector bundles, where $\mathbf{W}^\pm=\mathbf{W}\cap M^\pm$. For every $s\neq 0$, lift the tangent vector to $\theta\mapsto\Psi(\theta,s)$ to $\mathbf{U}$ using $f^{-1}$, rotate it by the angle $\pi/2$ and normalize it: $\sigma$ is defined as this unit vector (belonging to $\mathbf{U}_{\Psi(\theta,s)}$) if $s>0$, its opposite if $s<0$. In other words, $\sigma:\mathbf{W}\setminus W\to S\mathbf{U}$ is given by
$$\sigma(q) = \mathrm{sign}(s)\frac{R_{\pi/2}f^{-1}(\frac{\partial\Psi}{\partial\theta}(\theta,s))}{\sqrt{\langle f^{-1}(\frac{\partial\Psi}{\partial\theta}(\theta,s)), f^{-1}(\frac{\partial\Psi}{\partial\theta}(\theta,s))\rangle}},\quad (\theta,s)=\Psi^{-1}(q), \tag{9.29}$$
where $R_{\pi/2}$ denotes the rotation (with respect to the Euclidean structure) in $\mathbf{U}$ by angle $\pi/2$ in the counterclockwise sense. The following lemma shows that $\sigma$ can be extended to a continuous section from $\mathbf{W}\setminus\mathcal{T}$ to $S\mathbf{U}$.

**Lemma 9.61.** *$\sigma$ can be continuously extended to every point $q\in W$.*

*Proof.* Let $q\in W$, $U$ be a neighborhood of $q$ in $M$ and $(x_1,x_2)$ be a system of coordinates on $U$ centered at $q$ such the almost-Riemannian structure has the form (NF2) (see Theorem 9.37 and recall that by hypotheses we have no Martinet points). Assume, moreover, that $U$ is a trivializing neighborhood of both $\mathbf{U}$ and $TM$ and the pair of vector fields $\{F_1,F_2\}$ is the image under $f$ of a positively-oriented local orthonormal frame of $\mathbf{U}$. Then $W\cap U=\{(x_1,x_2)\mid x_1=0\}$. Since $\frac{\partial\Psi}{\partial\theta}(\theta,0)$ is non-zero and tangent to $W$, $\frac{\partial\Psi}{\partial\theta}(\theta,0)$ is tangent to the $x_2$-axis. Hence, thanks to the Malgrange Preparation Theorem, there exist $h_2:\mathbb{R}\to\mathbb{R}$, $h_1,h_3:\mathbb{R}^2\to\mathbb{R}$ smooth functions such that $h_2(x_2)\neq 0$ for every $x_2\in\mathbb{R}$ and for $\Psi(\theta,s)\in U$
$$\frac{\partial\Psi}{\partial\theta}(\theta,s) = (x_1h_1(x_1,x_2), h_2(x_2)+x_1h_3(x_1,x_2)),$$

where $(x_1, x_2)$ are the coordinates of the point $\Psi(\theta, s)$. Let us compute $\sigma$ at a point $p \in (\mathbf{W} \cap U) \setminus W$. Since

$$x_1 \frac{\partial \Psi}{\partial \theta}(\theta, s) = x_1 h_1(x_1, x_2) F_1(x_1, x_2) + \frac{h_2(x_2) + x_1 h_3(x_1, x_2)}{x_1 e^{\phi(x_1, x_2)}} F_2(x_1, x_2),$$

then

$$f^{-1}\left(\frac{\partial \Psi}{\partial \theta}(\theta, s)\right) = x_1 h_1(x_1, x_2)\zeta(x_1, x_2) + \frac{h_2(x_2) + x_1 h_3(x_1, x_2)}{x_1 e^{\phi(x_1, x_2)}} \rho(x_1, x_2),$$

where $(\zeta, \rho)$ is the unique local orthonormal basis of $\mathbf{U}|_U$ such that $f \circ \zeta = F_1$ and $f \circ \rho = F_2$. Notice that $U \cap M^+ = \{(x_1, x_2) \mid x_1 > 0\}$ and $U \cap M^- = \{(x_1, x_2) \mid x_1 < 0\}$. Using formula (9.29), for $(x_1, x_2) = \Psi(\theta, s) \in U \setminus W$ one easily gets

$$\sigma(x_1, x_2) = \frac{\operatorname{sign}(x_1)}{l(x_1, x_2)} \left(-\frac{h_2(x_2) + x_1 h_3(x_1, x_2)}{x_1 e^{\phi(x_1, x_2)}} \zeta + x_1 h_1(x_1, x_2) \rho\right),$$

where

$$l(x_1, x_2) = \sqrt{x_1^2 h_1(x_1, x_2)^2 + \frac{(h_2(x_2) + x_1 h_3(x_1, x_2))^2}{x_1^2 e^{2\phi(x_1, x_2)}}}.$$

Since

$$\lim_{x_1 \to 0} \frac{\operatorname{sign}(x_1)(h_2(x_2) + x_1 h_3(x_1, x_2))}{l(x_1, x_2) x_1 e^{\phi(x_1, x_2)}} = \frac{h_2(x_2)}{|h_2(x_2)|} \quad \text{and} \quad \lim_{x_1 \to 0} \frac{\operatorname{sign}(x_1) x_1 h_1(x_1, x_2)}{l(x_1, x_2)} = 0,$$

$\sigma$ can be continuously extended to the set $\{x_1 = 0\} = W \cap U$. $\square$

Let $C(\mathcal{Z})$ denote the set of connected component of $\mathcal{Z}$. Let $\tilde{\mathcal{Z}} = \coprod_{W \in C(\mathcal{Z})} S^1$ and consider an orientation-preserving diffeomorphism $\Psi : \tilde{\mathcal{Z}} \times (-1, 1) \to \coprod_{W \in C(\mathcal{Z})} \mathbf{W}$ such that $\Psi|_{\tilde{\mathcal{Z}} \times \{0\}}$ is an orientation-preserving diffeomorphism onto $\mathcal{Z}$. Applying Lemma 9.61 to every $W \in C(\mathcal{Z})$ and reducing, if necessary, the cylinders $\mathbf{W}$, we can assume that on $\mathbf{U} = \coprod_{W \in C(\mathcal{Z})} \mathbf{W}$, $\sigma$ is continuous and has no singularities. Extend $\sigma$ to $M \setminus \mathbf{U}$. We can assume that the extended section has only isolated singularities $\{p_1, \ldots, p_k\} \in M \setminus \mathcal{Z}$. We are left to prove that

$$\sum_{j=1}^{k} i(p_j, \sigma) = \chi(M^+) - \chi(M^-). \tag{9.30}$$

To this aim, consider the vector field $F = f \circ \sigma$. Notice that by construction the set of singularities of $F$ is exactly $\{p_1, \ldots, p_k\}$. Let us compute the index of $F$ at a singularity $p \in \{p_1, \ldots, p_k\}$. Since $f : \mathbf{U}|_{M^+} \to TM^+$ preserves the orientation and $f : \mathbf{U}|_{M^-} \to TM^-$ reverses the orientation, it follows that $i(p, F) = \pm i(p, \sigma)$, if $p \in M^{\pm}$. Therefore,

$$\sum_{j=1}^{k} i(p_j, \sigma) = \sum_{j | p_j \in M^+} i(p_j, F) - \sum_{j | p_j \in M^-} i(p_j, F). \tag{9.31}$$

The theorem is proved if we show that

$$\sum_{j | p_j \in M^+} i(p_j, F) = \chi(M^+), \quad \sum_{j | p_j \in M^-} i(p_j, F) = \chi(M^-). \tag{9.32}$$

To deduce equation (9.32), define $N^+ = M^+ \setminus \Psi(\tilde{\mathcal{Z}} \times (0,1/2))$. Notice that, by construction, $\sigma|_{\Psi(\tilde{\mathcal{Z}}\times\{1/2\})}$ is non-singular, hence the same is true for $F|_{\Psi(\tilde{\mathcal{Z}}\times\{1/2\})}$. Moreover, the almost-Riemannian angle between $T_q(\partial N^+)$ and $\mathrm{span}(F(q))$ is constantly equal to $\pi/2$. Hence $F|_{\partial N^+}$ points towards $N^+$ and applying the Hopf's Index Formula to every connected component of $N^+$ we conclude that

$$\sum_{j|p_j\in M^+} i(p_j,F) = \sum_{j|p_j\in N^+} i(p_j,F) = \chi(N^+) = \chi(M^+).$$

Similarly, we find

$$\sum_{j|p_j\in M^-} i(p_j,F) = \chi(M^-).$$

■

**Example: the Grushin sphere**

The Grushin sphere is the free 2D almost Riemannian structure on the sphere $S^2 = \{y_1^2+y_2^2+y_3^2 = 1\}$ for which an orthonormal frame is given by two orthogonal rotations for instance

$$Y_1 = \begin{pmatrix} 0 \\ -y_3 \\ y_2 \end{pmatrix} \text{ (rotation along the } y_1\text{-axis)} \tag{9.33}$$

$$Y_2 = \begin{pmatrix} -y_3 \\ 0 \\ y_1 \end{pmatrix} \text{ (rotation along the } y_2\text{-axis)} \tag{9.34}$$

In this case $\mathcal{Z} = \{y_3 = 0,\ y_1^2 + y_2^2 = 1\}$. Passing in spherical coordinates

$$\begin{aligned} y_1 &= \cos(x)\cos(\phi) \\ y_2 &= \cos(x)\sin(\phi) \\ y_3 &= \sin(x) \end{aligned}$$

and letting

$$\begin{aligned} F_1 &= \cos(\phi - \pi/2)Y_1 + \sin(\phi-\pi/2)Y_2 \\ F_2 &= -\sin(\phi-\pi/2)Y_1 + \cos(\phi-\pi/2)Y_2 \end{aligned}$$

we get that an orthonormal frame in coordinates $(x,\phi)$ is given by

$$F_1 = \begin{pmatrix} 0 \\ \tan(x) \end{pmatrix}, \quad F_2 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

Notice that the singularity at $x = \pi/2$ is due to the spherical coordinates, whereas $\mathcal{Z} = \{x = 0\}$ and all points of $\mathcal{Z}$ are Grushin points. In this case we have

$$dA = \frac{1}{|\tan(x)|}dx\,d\phi, \quad dA^s = \frac{1}{\tan(x)}dx\wedge d\phi, \quad K = \frac{-2}{\sin(x)^2}$$

The loci $\mathcal{Z}$, $M^\pm$, are illustrated in Figure 9.7. In this case using the symmetries of the system one immediately verify that

$$\int K\,dA^s = 0,$$

in accordance with Theorem 9.59 and the fact that this almost-Riemannian structure is free.

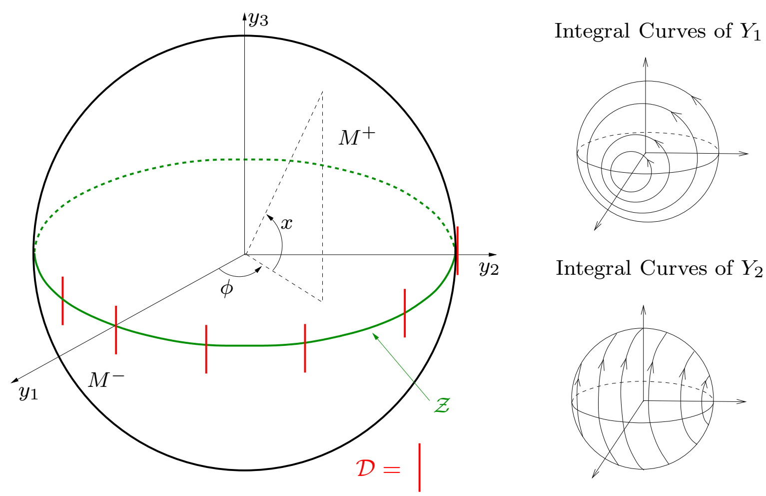



Figure 9.7: The Grushin sphere

### 9.5.4 Every compact orientable 2D manifold can be endowed with a free almost-Riemannian structure with only Riemannian and Grushin points

In this section we prove Proposition 9.49, by showing how to construct on every connected compact orientable two-dimensional manifold a free almost-Riemannian structure such that every point of $\mathcal{Z}$ is a Grushin point.

For the torus, an example of such structure is provided by the standard Riemannian one. The case of a connected sum of two tori can be treated by gluing together two copies of the pair of vector fields $F_1$ and $F_2$ represented in Figure 9.8A, which are defined on a torus with a hole cut out. In the figure the torus is represented as a square with the standard identifications on the boundary. The vector fields $F_1$ and $F_2$ are parallel on the boundary of the disk which has been cut out. Each vector field has exactly two zeros and the distribution spanned by $F_1$ and $F_2$ is transversal to the singular locus. Examples on the connected sum of three or more tori can be constructed similarly by induction. The resulting singular locus is represented in Figure 9.8B.

We are left to check the existence of a free almost-Riemannian structure with only Riemannian and Grushin points on the 2D sphere. But such a structure is provided by the Grushin sphere described above.

## 9.6 Bibliographical note

The first examples of almost-Riemannian structures have been discovered as generalized Riemannian structures underlying certain degenerate elliptic operators [Bao67, FL83, Gru70]. A systematic study of almost-Riemannian structures has been pursued in [ABS08, BCGS13, ABC$^+$10, BCG13, BCGJ11] (see also [Bel96, BP05]). In this series of papers:

- Riemannian points are called *ordinary points*;
- instead of Martinet points, *tangency points* are introduced. Comparing the normal form (NF3) and the normal form for tangency points (see (F3) in [ABS08, Theorem 1]), one sees

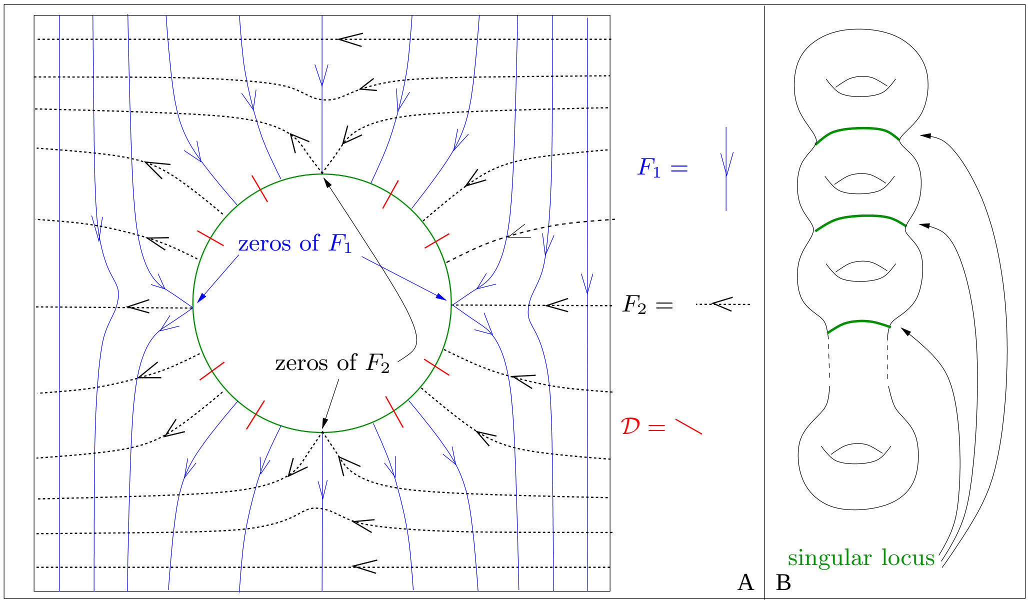



Figure 9.8: Construction of a free almost-Riemannian structure such that every point of $\mathcal{Z}$ is a Grushin point, on a compact orientable surface.

that tangency points are Martinet points of step exactly 3;

- the study of generic structures is more sophisticated than the one presented in this chapter (using Thom's transversality theorem instead than Sard's lemma). In particular, in [ABS08] it is proven that Grushin and tangency points do not disappear for small $C^1$-perturbations of the system;

- free almost-Riemannian structures are called *trivializable structures*.

A first version of a Gauss Bonnet theorem in the context of almost Riemannian structures was studied in [ABS08], while a version in presence of tangency points were studied in [ABC$^+$10].

Almost-Riemannian structures in dimension 3 were studied in [BCGM15].

The blow up of the area while approaching to the singular set and the Lebesgue and Hausdorff dimension of the singular set has been studied in a more general context in [GJ14, GJ15].

Almost-Riemannian structures appear also in applications: in space mechanics [BC14] and in problems of control of quantum mechanical systems [BCC05, BCG$^+$02b, BC14] (the Grushin sphere).

# Chapter 10

# Nonholonomic tangent space

In this chapter we introduce the notion of nonholomic tangent space, that can be regarded as the "principal part" of the structure defined on the manifold by the distribution in a neighborhood of a point. This notion is indeed *independent* on the inner product defined on the distribution.

When the distribution is endowed with an inner product, this process defines a metric tangent space (in the sense of Gromov) to the sub-Riemannian structure, that is itself a sub-Riemannian manifold. When the manifold is Riemannian one recovers on the tangent space the Euclidean structure induced by the Riemannian metric at the point. In the general case, the nonholonomic tangent space of a sub-Riemannian manifold at a point is endowed with a structure of Carnot group (cf. Section 10.1 for the definition) for an open dense subset of points on the manifold and homogeneous spaces of Carnot groups on the other points. In this sense Carnot groups play an analogous role of the Euclidean space in Riemannian geometry.

In this chapter we give an intrinsic construction of the nonholonomic tangent space through the theory of jets of curves and the notion of smooth admissible variation. We then prove the existence of privileged coordinates, i.e., special sets of coordinates where the nonholonomic tangent space writes conveniently to perform computations. Finally we provide both a geometric and an algebraic interpretation of this construction.

This chapter also contains some fundamental distance estimates, known in the literature as the Ball-Box theorem, and a classification of nonholonomic tangent space in low dimension.

## 10.1 Flag of the distribution and Carnot groups

Let us consider a distribution $\mathcal{D}$ associated with a structure $(M, \mathbf{U}, f)$, that is defined by a generating family $\{f_1, \dots, f_m\}$. If the distribution is bracket generating, we have a well-defined flag as follows.

**Definition 10.1.** Let us consider a bracket generating distribution $\mathcal{D}$ with generating family $f_1, \dots, f_m$ and fix $q \in M$. The *flag* of the sub-Riemannian structure at the point $q$ is the sequence of subspaces $\{\mathcal{D}_q^i\}_{i\in\mathbb{N}}$ of $T_qM$ defined by

$$\mathcal{D}_q^i := \operatorname{span}\{[f_{j_1}, \dots, [f_{j_{l-1}}, f_{j_l}]](q),\ \forall\, l \leq i\}. \tag{10.1}$$

Notice that $\mathcal{D}_q^1 = \mathcal{D}_q$ is the set of admissible directions. Moreover, by construction, $\mathcal{D}_q^i \subset \mathcal{D}_q^{i+1}$ for every $i \geq 1$.

The bracket generating assumption implies that for every $q \in M$ there exists a minimal integer $k = k(q)$ such that $\mathcal{D}_q^{k(q)} = T_qM$. The integer $k(q)$ is called the *step* of the sub-Riemannian structure at $q$.

The sub-Riemannian structure is said to be *equiregular* if for every $i \geq 1$ the integer $d_i(q) = \dim \mathcal{D}_q^i$ does not depend on $q$.

**Exercise 10.2.** (a). Prove that the filtration defined by the subspaces $\mathcal{D}_q^i$, for $i \geq 1$, and the value $k(q)$ depend only on the modulus of horizontal vector fields, i.e., are independent on the generating family (or equivalently, on the trivialization of the vector bundle $\mathbf{U}$ defining $\mathcal{D}$).

(b). Prove that the map $q \mapsto k(q)$ is upper semicontinuous.

Notice that in general we have only a well-defined filtration (i.e., sequence of increasing subspaces) of the tangent space

$$\mathcal{D}_q^1 \subset \mathcal{D}_q^2 \subset \ldots \subset \mathcal{D}_q^k = T_qM. \tag{10.2}$$

but no canonical gradation (i.e., splitting of the tangent space into sum of complementary subspaces) is defined. The object that is well defined is the graded vector space

$$\mathrm{gr}_q(\mathcal{D}) = \mathcal{D}_q \oplus \mathcal{D}_q^2/\mathcal{D}_q \oplus \ldots \oplus \mathcal{D}_q^m/\mathcal{D}_q^{m-1}. \tag{10.3}$$

If the sub-Riemannian structure is equiregular in a neighborhood of a point $q$, the graded vector space (10.3) inherits the structure of a Lie algebra from the Lie brackets of vector fields. This is by construction a homogeneous and stratified Lie algebra, whose corresponding Lie group is what is called a Carnot group.

Carnot groups play a crucial role in sub-Riemannian geometry: these are left-invariant sub-Riemannian structures arising as metric tangent space of equiregular sub-Riemannian manifolds.

**Definition 10.3** (Carnot Groups)**.** A *Carnot group* $G$ is a connected and simply connected Lie group whose Lie algebra $\mathfrak{g}$ admits a decomposition

$$\mathfrak{g} = \mathfrak{g}_1 \oplus \mathfrak{g}_2 \oplus \ldots \oplus \mathfrak{g}_r \tag{10.4}$$

satisfying the following properties

$$[\mathfrak{g}_1, \mathfrak{g}_i] = \mathfrak{g}_{i+1}, \qquad [\mathfrak{g}_1, \mathfrak{g}_r] = 0, \qquad i = 1, \ldots, r-1. \tag{10.5}$$

The smallest integer $r$ such that (10.4)-(10.5) are satisfied is called the *step* of the Carnot group.

When the first layer $\mathfrak{g}_1$ of the Lie algebra $\mathfrak{g}$ is endowed with an inner product, then $G$ is automatically endowed with a left-invariant sub-Riemannian structure (cf. Chapter 7), where the distribution is defined by left-invariant vector fields whose value at the identity belongs to $\mathfrak{g}_1$. This distribution is bracket generating thanks to (10.5).

Notice that Carnot groups of step 2 as defined in Section 7.5 are included in Definition 10.3.

*Remark* 10.4. Carnot groups are also known in the literature as *homogeneous and stratified Lie group.* Indeed the Lie agebra $\mathfrak{g}$ of a Carnot group $G$ admits the stratification (10.4) and thanks to the property (10.5) they posses a family $\{\delta_\alpha\}_{\alpha \in \mathbb{R}}$ of authomorphisms on $\mathfrak{g}$ (called *dilations*) defined by

$$\delta_\alpha(v) = \sum_{i=1}^r \alpha^i v_i, \qquad \text{if} \quad v = \sum_{i=1}^r v_i, \ v_i \in \mathfrak{g}_i.$$

When the structure is not equiregular in a neighborhood of a point $q$, then the graded vector space (10.3) *does not* inherits the structure of a Lie algebra, thus it is not possible to extract directly from (10.3) all the algebraic informations on the distribution. In this case the nonholonomic tangent space is a homogeneous space of a Carnot group, and its intrinsic construction is more sophisticated and requires the theory of jet spaces, that we not introduce.

## 10.2 Jet spaces

### 10.2.1 Jets of curves

In what follows, given a point $q \in M$, the symbol $\Omega_q$ denotes the set of smooth curves $\gamma$ on $M$ defined on some open interval $I$ containing 0 and based at $q$, that is $\gamma(0) = q$. In fact, we work with germs of smooth curves at 0 and sometimes it will be convenient to think to those curves $\gamma$ to be defined on $I = \mathbb{R}$.

Fix $q$ in $M$ and a curve $\gamma \in \Omega_q$. In every coordinate chart one can write the Taylor expansion

$$\gamma(t) = q + \dot{\gamma}(0)t + O(t^2). \tag{10.6}$$

The tangent vector $v \in T_qM$ to $\gamma$ at $t = 0$ is by definition the equivalence class of curves in $\Omega_q$ such that, in some coordinate chart, they have the same 1-st order Taylor polynomial. This requirement indeed implies that the same is true for every coordinate chart, by the chain rule.

In the same spirit one can consider, given a smooth curve $\gamma \in \Omega_q$, its $k$-th order Taylor polynomial at $q$

$$\gamma(t) = q + \dot{\gamma}(0)t + \ddot{\gamma}(0)\frac{t^2}{2} + \ldots + \gamma^{(k)}(0)\frac{t^k}{k!} + O(t^{k+1}), \tag{10.7}$$

and define analogously an equivalence class on higher order Taylor polynomial.

**Exercise 10.5.** Let $\gamma, \gamma' \in \Omega_q$. We say that $\gamma$ is equivalent up to order $k$ at $q$ to $\gamma'$, writing $\gamma \sim_{q,k} \gamma'$, if their Taylor polynomial at $q$ of order $k$ coincide in some coordinate chart. Prove that $\sim_{q,k}$ is a well-defined equivalence relation on the set of curves based at $q$.

**Definition 10.6.** Let $k > 0$ be an integer and $q \in M$. We define the set of *$k$-th jets of curves* at point $q \in M$ as the equivalence classes of $\Omega_q$ with respect to $\sim_{q,k}$. We denote by $J_q^k\gamma$ the equivalence class of a curve $\gamma$ and we set

$$J_q^kM := \{J_q^k\gamma \mid \gamma \in \Omega_q\}.$$

**Exercise 10.7.** Prove that $J_q^kM$ has a structure of smooth manifold of dimension $kn$, where $n = \dim M$. (*Hint*: use the coordinates representation (10.7) and the fact that the $k$-th order Taylor polynomial is characterized by the $n$-dimensional vectors $\gamma^{(i)}(0)$ for $i = 1, \ldots, k$.)

In the following we always assume that $q \in M$ is fixed and when working in a coordinate chart we always assume that $q = 0$. Identifying the jet of a curve $\gamma \in \Omega_q$, with its Taylor polynomial in some coordinate chart, we can write (recall that $\gamma(0) = q = 0$)

$$J_q^k\gamma = \sum_{i=1}^{k} \gamma^{(i)}(0)\frac{t^i}{i!}.$$

When $k=1$, we have easily from the definition that $J_q^1 M = T_q M$. To study more in detail the structure of jet space for $k \geq 2$, let us introduce the map $\Pi_{k-1}^k$ which "forgets" the $k$-th derivative

$$\Pi_{k-1}^k : J_q^k M \longrightarrow J_q^{k-1} M, \qquad \Pi_{k-1}^k \left( \sum_{i=1}^{k} \gamma^{(i)}(0) \frac{t^i}{i!} \right) := \sum_{i=1}^{k-1} \gamma^{(i)}(0) \frac{t^i}{i!}.$$

**Proposition 10.8.** *Let $k \geq 2$. Then $J_q^k M$ is an affine bundle over $J_q^{k-1} M$ with projection $\Pi_{k-1}^k$, whose fibers are affine spaces over $T_q M$.*

*Proof.* Fix an element $j \in J_q^{k-1} M$. The fiber $(\Pi_{k-1}^k)^{-1}(j)$ is the set of all $k^{th}$-jets with fixed $(k-1)^{th}$ jet equal to $j$. To show that it is an affine space over $T_q M$ it is enough to define the sum of a tangent vector and a $k^{th}$-jet, with $(k-1)^{th}$-jet fixed, in such a way that the resulting $k^{th}$-jet has the same $(k-1)^{th}$-jet.

Let $j = J_q^k \gamma$ be the $k^{th}$-jet of a smooth curve in $M$ and let $v \in T_q M$. Consider a smooth vector field $V \in \mathrm{Vec}(M)$ such that $V(q) = v$ and define the sum

$$J_q^k \gamma + v := J_q^k(\gamma^v), \qquad \gamma^v(t) = e^{t^k V}(\gamma(t)) \tag{10.8}$$

It is easy to see that, due to the presence of the factor $t^k$, the $(k-1)^{th}$ Taylor polynomial of $\gamma$ and $\gamma^v$ coincide. Indeed

$$J_q^k(e^{t^k V}(\gamma(t))) = J_q^k \gamma + t^k V(q)$$

Hence the sum (10.8) gives to $(\Pi_{k-1}^k)^{-1}(j)$ the structure of an affine space over $T_q M$. Notice that this definition does not depend on the representative curve $\gamma$ defining $j$. $\square$

Roughly speaking, the fact that $J_q^k M$ is an affine bundle (and not a vector bundle) is saying that one cannot complete in a canonical way a $(k-1)^{th}$-jet to a $k^{th}$-jet, i.e., we cannot fix an origin in the fibers. On the other hand there exists a sort of "global" origin on the space $J_q^k M$, that is the jet of the constant curve equal to $q$.

Now we introduce dilations on jet spaces, analogous to homotheties in Euclidean spaces. This is done via time rescaling.

**Definition 10.9.** Let $\alpha \in \mathbb{R}$ and define $\gamma_\alpha(t) := \gamma(\alpha t)$ for every $t$ such that the right hand side is defined. Define the *dilation* of factor $\alpha$ on $J_q^k M$ as

$$\delta_\alpha : J_q^k M \to J_q^k M, \qquad \delta_\alpha(J_q^k \gamma) = J_q^k(\gamma_\alpha).$$

One can check that this definition does not depend on the representative and, in coordinates, it is written as a *quasi-homogeneous* multiplication

$$\delta_\alpha \left( \sum_{i=1}^{k} t^i \xi_i \right) = \sum_{i=1}^{k} t^i \alpha^i \xi_i.$$

Next we extend the notion of jets also for vector fields. To start with we consider flows on the manifold. All flows we consider in what follows are a priori defined locally. To simplify the discussion, we work as if they are globally defined.

**Definition 10.10.** A *flow* on $M$ is a family of diffeomorphisms $P = \{P_t \in \mathrm{Diff}(M),\ t \in \mathbb{R}\}$ that is smooth with respect to $t$ and such that $P_0 = \mathrm{Id}$.

Notice that we do not require the family to be a one parametric group (i.e., the group law $P_t \circ P_s = P_{t+s}$ is not necessarily satisfied). Its infinitesimal generator is the nonautonomous vector field

$$X_t := \left.\frac{d}{d\varepsilon}\right|_{\varepsilon=0} P_{t+\varepsilon} \circ P_t^{-1}. \tag{10.9}$$

The set of all flows on $M$ is a group with the point-wise product, i.e., the product of the flows $P = \{P_t\}$ and $Q = \{Q_t\}$ is given by

$$(P \circ Q)_t := P_t \circ Q_t$$

The action of a flow (in the sense of Definition 10.10) on a smooth curve $\gamma$ is defined as

$$(P\gamma)(t) := P_t(\gamma(t)). \tag{10.10}$$

**Proposition 10.11.** *Let $P$ be a smooth flow on $M$. Then $P$ induces a well-defined map $P : J_q^k M \to J_q^k M$ defined as follows*

$$Pj := J_q^k(P\gamma), \qquad \textit{if} \quad j = J_q^k \gamma. \tag{10.11}$$

*Moreover $(P \circ Q)j = P(Qj)$ for every $j \in J_q^k M$*

*Proof.* Notice that, since $P_0 = \mathrm{Id}$, then $P\gamma \in \Omega_q$ for every $\gamma \in \Omega_q$. By the chain rule, $J_q^k(P\gamma)$ depends only on first $k$ derivatives of $\gamma$ at $q$, i.e., on $J_q^k\gamma$. Hence this action is well-behaved with respect to equivalence relations $\sim_{k,q}$. The last part of the statement is an easy check and is left to the reader. □

### 10.2.2 Jets of vector fields

As explained in the proof of Proposition 10.11, a flow on $M$ induces a diffeomeorphism in $\Omega_q$, and thus in the space of jets $J_q^k M$. In particular, given a vector field $V \in \mathrm{Vec}(M)$, the flow associated with $V$, i.e. the 1-parametric group $P_V = \{e^{tV}\}$, acts on curves

$$(P_V\gamma)(t) = e^{tV}(\gamma(t)),$$

and this action passes to the quotient on jets.

A vector field on a manifold is the infinitesimal generator of a family of diffeomorphism, hence an element of $\mathrm{Vec}(J_q^k M)$ is the infinitesimal generator of a family of diffeomorphism of $J_q^k M$.

A natural contstruction, given $V \in \mathrm{Vec}(M)$, is to consider the 1-parametric group of flows (indexed by $s$) defined by $P_V^s = \{e^{stV}\}$ and to define the $k$-th jet of the vector field as the infinitesimal generator of this family of diffeomorphism of $J_q^k M$.

**Definition 10.12.** For every $V \in \mathrm{Vec}(M)$, the vector field $J_q^k V \in \mathrm{Vec}(J_q^k M)$ is the smooth section $J_q^k V : J_q^k M \to T J_q^k M$ defined as follows

$$(J_q^k V)(J_q^k \gamma) := \left.\frac{\partial}{\partial s}\right|_{s=0} P_V^s(J_q^k\gamma) = \left.\frac{\partial}{\partial s}\right|_{s=0} J_q^k(e^{tsV}(\gamma(t))). \tag{10.12}$$

**Exercise 10.13** (1-jet of vector fields)**.** Prove that $J_q^1 M = T_q M$. Moreover, if $V \in \mathrm{Vec}(M)$ then $J_q^1 V = V(q)$ is the constant vector field on the vector space $T_q M$ defined by the value of $V$ at $q$.

**Exercise 10.14.** Prove the following formula for every $V \in \mathrm{Vec}(M)$

$$(J_q^k V)(J_q^k \gamma) = \sum_{i=1}^{k} \frac{t^i}{i!} \frac{d^i}{dt^i}\Big|_{t=0} (tV(\gamma(t))),$$

where $V$ is identified with a vector function $V : \mathbb{R}^n \to \mathbb{R}^n$ in coordinates.

To end this section we study the interplay between dilations and jets of vector fields. Since $\delta_\alpha$ is a map on $J_q^k M$ its differential $(\delta_\alpha)_*$ acts on elements of $\mathrm{Vec}(J_q^k M)$, and in particular on jets of vector fields on $M$. Surprisingly, its action on these particular vector fields is linear with respect to $\alpha$.

**Proposition 10.15.** *For every $\alpha \in \mathbb{R}$ and $V \in \mathrm{Vec}(M)$ one has*

$$(\delta_\alpha)_*(J_q^k V) = J_q^k(\alpha V) = \alpha\, J_q^k V.$$

*Proof.* By definition of the differential of a map (see also Chapter 2). we have

$$\begin{aligned}
((\delta_\alpha)_* J_q^k V))(J_q^k \gamma) &= \frac{\partial}{\partial s}\Big|_{s=0} J_q^k(\delta_\alpha\, e^{tsV}\, \delta_{1/\alpha}(\gamma(t))) \\
&= \frac{\partial}{\partial s}\Big|_{s=0} J_q^k(\delta_\alpha\, e^{tsV}(\gamma(t/\alpha))) \\
&= \frac{\partial}{\partial s}\Big|_{s=0} J_q^k(e^{\alpha tsV}(\gamma(t))) \\
&= J_q^k(\alpha V) = \alpha J_q^k V
\end{aligned}$$

□

## 10.3 Admissible variations and nonholonomic tangent space

The goal of this section is to define the appropriate notion of tangent vector, or more precisely to define the "tangent structure" to a distribution at a point.

As usual, we assume that the distribution $\mathcal{D}$ associated with a structure $(M, \mathbf{U}, f)$ is defined by a generating family $\{f_1, \dots, f_m\}$ and admissible curves on $M$ are maps $\gamma : [0, T] \to M$ such that there exists a control function $u \in L^\infty$ satisfying

$$\dot{\gamma}(t) = f_{u(t)}(\gamma(t)) = \sum_{i=1}^{m} u_i(t) f_i(\gamma(t)).$$

To build a notion of "tangent structure" as a first order approximation of the structure, thus encoding informations about all directions, we cannot restrict to study family of admissible curves, since these are all tangent to the distribution.

We shall reinterpret a "tangent vector" as the principal term of a "variation of a point". To give a precise meaning to this, we introduce the notion of smooth admissible variation.

### 10.3.1 Admissible variations

**Definition 10.16.** A curve $\gamma : [0,T] \to M$ in $\Omega_q$ is said a *smooth admissible variation* if there exists a family of controls $\{u(t,s)\}_{s\in[0,\tau]}$ such that

(i) $u(t,\cdot)$ is measurable and essentially bounded for all $t \in [0,T]$, uniformly in $s \in [0,\tau]$,

(ii) $u(\cdot,s)$ is smooth with bounded derivatives, for all $s \in [0,\tau]$, uniformly in $t \in [0,T]$,

(iii) $u(0,s) = 0$ for all $s \in [0,\tau]$,

(iv) $\gamma(t) = \overrightarrow{\exp} \int_0^\tau f_{u(t,s)}(q)ds$.

In other words $\gamma$ is a smooth admissible variation (or, shortly, admissible variation) if it can be parametrized as the final point of a smooth family of admissible curves.

*Remark* 10.17. Notice that from the property (iii) of the definition of admissible variation, we can rewrite $u(t,s) = t\bar{u}(t,s)$ for some suitable family of controls $\bar{u}(t,s)$ that are still smooth with respect to $t$ but do not necessarily satisfy $\bar{u}(0,s) = 0$.

The following example shows that admissible variations are *not* admissible curves, in general.

**Example 10.18.** Consider two vector fields $X, Y \in \mathrm{Vec}(M)$ and the curve

$$\gamma : [0,T] \to M, \qquad \gamma(t) = e^{-tY} \circ e^{-tX} \circ e^{tY} \circ e^{tX}(q).$$

If we set $f_u := u_1 X + u_2 Y$ and $u : [0,T] \times [0,4] \to \mathbb{R}^2$ defined by

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

It is easily seen that $\gamma$ is an admissible variation since

$$\gamma(t) = \overrightarrow{\exp} \int_0^4 f_{u(t,s)}(q)ds$$

and it admits the expansion in coordinates $\gamma(t) = q + t^2[X,Y](q) + o(t^2)$.

Iterating the previous construction one can actually build smooth admissible variations whose tangent vector at $t = 0$ is any element in $\mathcal{D}_q^i \backslash \mathcal{D}_q^{i-1}$ (cf. Lemmas 10.36-10.37 for a precise statement).

**Proposition 10.19.** *Equivalent distributions admits the same admissible variations. In particular the class of smooth admissible variation is independent on the inner product defined on the distribution.*

*Proof.* Recall that two distributions $\mathcal{D}, \mathcal{D}'$ are equivalent (see also Definitions 3.3 and 3.18) if and only if the corresponding modulus of horizontal vector fields are isomorphic where

$$\mathcal{D} = \mathrm{span}\{f(\sigma),\ \sigma \text{ smooth section of } \mathbf{U}\}.$$

It is not restrictive to assume that $\mathcal{D}$ and $\mathcal{D}'$ are finitely generated by $f_1, \ldots, f_m$ and $f_1', \ldots, f_{m'}'$ (we stress that a priori $m \neq m'$).

By definition, for any admissible variation $\gamma(t)$ there exists a family $q(t,s)$, for $s \in [0,\tau]$, such that $\gamma(t) = q(t,\tau)$ and $q(t,s)$ solves

$$\frac{\partial}{\partial s} q(t,s) = \sum_{i=1}^{m} u_i(t,s) f_i(q(t,s)), \qquad s \in [0,\tau], \tag{10.13}$$

Assume that $f_1', \ldots, f_{m'}'$ is another set of local generators of the modulus. Then there exist functions $a_{ij} \in C^\infty(M)$ for $i = 1, \ldots, m$ and $j = 1, \ldots, m'$, such that

$$f_i(q) = \sum_{j=1}^{m} a_{ij}(q) f_j'(q), \qquad \forall\, q \in M, \qquad \forall\, i = 1, \ldots, m. \tag{10.14}$$

Next we prove that there exist a family $\widetilde{u}(t,s)$ of controls such that $\gamma$ is an admissible variation for the frame $f_1', \ldots, f_{m'}'$. From (10.14) we get

$$\sum_{i=1}^{m} u_i(t,s) f_i(q) = \sum_{i=1}^{m} \sum_{j=1}^{m'} u_i(t,s) a_{ij}(q) f_j'(q). \tag{10.15}$$

Then we could define, through the solution $q(t,s)$ of (10.13), the new family of controls

$$u_j'(t,s) := \sum_{i=1}^{m} u_i(t,s) a_{ij}(q(t,s)), \qquad j = 1, \ldots, m',$$

and we see from identities above that

$$\frac{\partial}{\partial s} q(t,s) = \sum_{j=1}^{m'} u_j'(t,s) f_j'(q(t,s)), \qquad s \in [0,\tau]. \tag{10.16}$$

Since the role of $f_1, \ldots, f_m$ and $f_1', \ldots, f_{m'}'$ can be exchanged, this prove the equivalence. $\square$

**Assumption.** In what follows $\mathcal{D}$ denotes a distribution associated with the datum $(M, \mathbf{U}, f)$. Here the vector bundle $\mathbf{U}$ is not necessarily endowed with an Euclidean structure. We fix a point $q \in M$ and we assume that the distribution on $M$ is bracket generating of step $k$ at the point $q$.

**Definition 10.20.** Let $\mathcal{D}$ be a bracket generating distribution on $M$. The set of *admissible jets* is

$$J_q^f M := \{ J_q^k \gamma,\ \gamma \in \Omega_q \text{ is an admissible variation} \}$$

where $k$ is the step of the distribution at $q$, i.e., $\mathcal{D}_q^k = T_q M$.

### 10.3.2 Nonholonomic tangent space

To define the nonholonomic tangent space in a coordinate-free way, we need to introduce the group of flows of admissible variations.

**Definition 10.21.** Let $\mathcal{D}$ be a bracket generating distribution on $M$. The *group of flows of admissible variations* is

$$\mathcal{P}^f := \left\{ \overrightarrow{\exp} \int_0^\tau f_{u(t,s)} ds, \ u(t,s) \text{ smooth variation} \right\},$$

where the group structure on $\mathcal{P}^f$ is given by the following identity:

$$\overrightarrow{\exp} \int_0^{\tau_1} f_{u_1(t,s)} ds \circ \overrightarrow{\exp} \int_0^{\tau_2} f_{u_2(t,s)} ds = \overrightarrow{\exp} \int_0^{\tau_1+\tau_2} f_{v(t,s)} ds$$

where we set

$$v(t,s) := \begin{cases} u_1(t,s), & 0 \leq s \leq \tau_1, \\ u_2(t, s-\tau_1), & \tau_1 \leq s \leq \tau_1 + \tau_2. \end{cases}$$

*Remark* 10.22. Any admissible variation is given by $\gamma(t) = P_t(q)$ for some $P \in \mathcal{P}^f$, where we identify $q$ with the constant curve. Hence $J_q^f M$ is exactly the orbit of $q$ under the action of the group $\mathcal{P}^f$

$$J_q^f M = \{ J_q^k(P(q)) \mid P \in \mathcal{P}^f \}.$$

The nonholonomic tangent space will be defined as the quotient of $\mathcal{P}^f$ with respect to the action of the subgroup of "slow flows".

**Definition 10.23.** A smooth admissible variation $u(t,s)$ for $\mathcal{D}$ is said to be a *slow variation* if

$$u(0,s) = \frac{\partial u}{\partial t}(0,s) = 0, \qquad \forall\, s \in [0,\tau]. \tag{10.17}$$

A flow associated with a slow variation is said to be *purely slow*. The *subgroup of slow flows* $\mathcal{P}_0^f$ is the normal subgroup of $\mathcal{P}^f$ generated by flows associated with slow variations, namely

$$\mathcal{P}_0^f := \left\{ (P_t)^{-1} \circ Q_t \circ P_t \mid P \in \mathcal{P}^f,\, Q \text{ purely slow} \right\}. \tag{10.18}$$

*Remark* 10.24. By definition of slow variation and the linearity of $f$, a purely slow flow $Q_t$ is associated with a family of control that can be written in the form $u(t,s) = tv(t,s)$, where $v(0,s) = 0$ (cf. also Remark 10.17). Moreover we have

$$Q_t = \overrightarrow{\exp} \int_0^\tau f_{u(t,s)} ds = \overrightarrow{\exp} \int_0^\tau f_{tv(t,s)} ds = \overrightarrow{\exp} \int_0^\tau t f_{v(t,s)} ds.$$

Heuristically, a flow $Q_t$ is purely slow if the first nonzero jet $J_q^i \gamma$ of the trajectory $\gamma(t) = q \circ Q_t$ belongs to a subspace $\mathcal{D}_q^j$, with $j < i$. In particular $\dot\gamma(0) = 0$.

Being equivalent up to a slow flow is a well-defined equivalence relation on the space of jets.

**Exercise 10.25.** Let $j = J_q^k \gamma$ and $j' = J_q^k \gamma'$ for some $\gamma, \gamma' \in \Omega_q$. Prove that

$$J_q^k \gamma \sim J_q^k \gamma', \qquad \text{if} \qquad \gamma'(t) = P_t(\gamma(t)) \tag{10.19}$$

for some slow flow $P \in \mathcal{P}_0^f$ is a well defined equivalence relation on $J_q^f M$.

This permits us to introduce the main object of the section.

**Definition 10.26.** The *nonholonomic tangent space* $T_q^f M$ is defined as

$$T_q^f M := J_q^f M / \sim$$

where $\sim$ is the equivalence relation defined in (10.19).

Every horizontal vector field on the sub-Riemannian manifold induces a vector field on the noholonomic tangent space.

**Proposition 10.27.** *Let $\mathcal{D}$ be a bracket-generating distribution on $M$ of step $k$ at $q$ and $X$ be a horizontal vector field. Then the jet $J_q^k X$ is tangent to the submanifold $J_q^f M$. Moreover $J_q^k X$ induces a well defined vector field $\widehat{X}$ on the nonholonomic tangent space $T_q^f M$.*

*Proof.* By definition of $J_q^k X$, its action on a jet of an admissible variation $J_q^k \gamma$ is given by

$$(J_q^k X)(J_q^k \gamma) := \left.\frac{\partial}{\partial s}\right|_{s=0} P_X^s(J_q^k \gamma) = \left.\frac{\partial}{\partial s}\right|_{s=0} J_q^k(e^{tsX}(\gamma(t))). \tag{10.20}$$

It is easily seen that, if $\gamma(t)$ is an admissible variation, then for every $s$ the curve $t \mapsto e^{tsV}(\gamma(t))$ is an admissible variation as well, thus $J_q^k X$ is tangent to the submanifold $J_q^f M$.

To prove that the action is well defined on the quotient, assume that $\gamma(t) \sim \gamma'(t)$, i.e., $\gamma'(t) = \gamma(t) \circ Q_t$ for a slow flow $Q \in \mathcal{P}_0^f$. Then we compute, using chronological notation

$$\begin{aligned} \gamma'(t) \circ e^{stX} &= \gamma(t) \circ Q_t \circ e^{stX} \\ &= \gamma(t) \circ e^{stX} \circ e^{-stX} \circ Q_t \circ e^{stX} \\ &= (\gamma(t) \circ e^{stX}) \circ \widetilde{Q}_t^s \end{aligned}$$

where $\widetilde{Q}_t^s := e^{-tsX} \circ Q_t \circ e^{tsX}$ is a slow flow for every fixed $s$ and smooth with respect to $s$. This means that for every $s$ we have $e^{tsX}\gamma(t) \sim e^{tsX}\gamma'(t)$ through a slow flow $\widetilde{Q}_t^s$. Hence $J_q^k X$ defines a vector field $\widehat{X}$ on the quotient $T_q^f M$. □

## 10.4 Nonholonomic tangent space and privileged coordinates

In this section we discuss some special set of coordinates, called privileged, respecting the flag and in which we have an explicit and nice description of the nonholonomic tangent space $T_q^f M$.

### 10.4.1 Privileged coordinates

Consider non negative integers $n_1, \ldots, n_k$ such that $n = n_1 + \ldots + n_k$ and the splitting

$$\mathbb{R}^n = \mathbb{R}^{n_1} \oplus \ldots \oplus \mathbb{R}^{n_k}, \qquad x = (x_1, \ldots, x_k)$$

where $x_i = (x_i^1, \ldots, x_i^{n_i}) \in \mathbb{R}^{n_i}$ for $i = 1, \ldots, k$.

The space $\mathrm{Der}(\mathbb{R}^n)$ of all differential operators in $\mathbb{R}^n$ with smooth coefficients form an associative algebra with composition of operators as multiplication. The differential operators with polynomial coefficients form a subalgebra of this algebra with generators $1, x_i^j, \frac{\partial}{\partial x_i^j}$, where $i = 1, \dots, k;\ j = 1, \dots, n_i$. We define weights of generators as follows

$$\nu(1) := 0, \qquad \nu(x_i^j) := i, \qquad \nu\left(\frac{\partial}{\partial x_i^j}\right) := -\nu(x_i^j) = -i.$$

This defines by additivity the weight of any monomial

$$\nu\left(y_1 \cdots y_\alpha \frac{\partial^\beta}{\partial z_1 \cdots \partial z_\beta}\right) = \sum_{i=1}^{\alpha} \nu(y_i) - \sum_{j=1}^{\beta} \nu(z_j).$$

We say that a polynomial differential operator $D$ is *homogeneous* if it is a sum of monomial terms of the same weight. We stress that this definition depends on the coordinate set and the choice of the weights.

**Lemma 10.28.** *Let $D_1, D_2$ be two homogeneous differential operators. Then $D_1 \circ D_2$ is homogeneous and*

$$\nu(D_1 \circ D_2) = \nu(D_1) + \nu(D_2). \tag{10.21}$$

*Proof.* By linearity, it is sufficent to check formula (10.21) for monomials of the form

$$D_1 = \frac{\partial}{\partial x_{i_1}^{j_1}}, \qquad D_2 = x_{i_2}^{j_2}.$$

Then we have

$$D_1 \circ D_2 = \frac{\partial}{\partial x_{i_1}^{j_1}} \circ x_{i_2}^{j_2} = x_{i_2}^{j_2} \frac{\partial}{\partial x_{i_1}^{j_1}} + \frac{\partial x_{i_2}^{j_2}}{\partial x_{i_1}^{j_1}},$$

and formula (10.21) is easily checked in this case. □

A special case is when we consider first order differential operators, namely vector fields.

**Corollary 10.29.** *If $V_1, V_2 \in \mathrm{Vec}(\mathbb{R}^n)$ are homogeneous vector fields then $[V_1, V_2]$ is homogeneous and $\nu([V_1, V_2]) = \nu(V_1) + \nu(V_2)$.*

With these properties we can define a filtration in the space of all smooth differential operators Indeed we can write (in the multi-index notation)

$$D = \sum_\alpha \varphi_\alpha(x) \frac{\partial^{|\alpha|}}{\partial x^\alpha}.$$

Considering the Taylor expansion at 0 of every coefficient we can split $D$ as a sum of its homogeneous components

$$D \approx \sum_{i=-\infty}^{\infty} D^{(i)},$$

and define the filtration $\{\mathcal{F}^{(h)}\}_{h \in \mathbb{Z}}$ of $\mathrm{Der}(\mathbb{R}^n)$ as follows

$$\mathcal{F}^{(h)} := \{D \in \mathrm{Der}(\mathbb{R}^n) : D^{(i)} = 0, \forall\, i < h\}, \qquad h \in \mathbb{Z}.$$

It is easy to see that it is a decreasing filtration, i.e., $\mathcal{F}^{(h)} \subset \mathcal{F}^{(h-1)}$ for every $h \in \mathbb{Z}$. Moreover, if we restrict our attention to vector fields, we get

$$V \in \mathrm{Vec}(\mathbb{R}^n) \quad \Rightarrow \quad V^{(i)} = 0, \quad \forall\, i < -k.$$

Indeed every monomial of a $N^{th}$-order differential operator has weight not smaller than $-kN$. In other words we have

(i) $\mathrm{Vec}(\mathbb{R}^n) \subset \mathcal{F}^{(-k)}$,

(ii) $V \in \mathrm{Vec}(\mathbb{R}^n) \cap \mathcal{F}^{(0)}$ implies $V(0) = 0$.

In particular every vector field that does not vanish at the origin belongs at least to $\mathcal{F}^{(-1)}$. This motivates the following definition.

**Definition 10.30.** (i). A system of coordinates near the point $q$ is said *linearly adapted* to the flag $\mathcal{D}^1_q \subset \mathcal{D}^2_q \subset \ldots \subset \mathcal{D}^k_q$ if, in coordinates,

$$\mathcal{D}^i_q = \mathbb{R}^{n_1} \oplus \ldots \oplus \mathbb{R}^{n_i}, \qquad \forall\, i = 1, \ldots, k. \tag{10.22}$$

(ii). A system of coordinates near the point $q$ is said *privileged* if it is linearly adapted to the flag and $X \in \mathcal{F}^{(-1)}$ for every $X \in \mathcal{D}$.

Notice that condition (i) can always be satisfied after a suitable linear change of coordinates. Condition (ii) says that each horizontal vector field has no homogeneous component of degree less than $-1$.

**Example 10.31** (On privileged coordinates)**.** We discuss which coordinate systems are privileged in the case $k = 1, 2, 3$.

(i) For $k = 1$ all sets of coordinates are privileged. In fact $\nu(\partial_{x_i}) = -1$ for all $i$ easyly implies $\mathrm{Vec}(M) \subset \mathcal{F}^{(-1)}$.

(ii) For $k = 2$ all systems of coordinates that are linearly adapted to the flag are also privileged. Indeed, we have $\nu(\partial_{x_1^j}) = -1$ and $\nu(\partial_{x_2^j}) = -2$. Thus a vector field belonging to $\mathcal{F}^{(-2)} \setminus \mathcal{F}^{(-1)}$ contains a monomial vector field of the kind $\partial_{x_2^j}$, with constant coefficients. On the other hand a vector field $X \in \mathcal{D}$ cannot contain such a monomial since, by our assumption $X(0) \in \mathcal{D}^1_0 = \mathbb{R}^{n_1}$.

(iii) For $k = 3$, let us show an example of coordinates that are linearly adapted but not privileged. Consider the following set of vector fields in $\mathbb{R}^3 = \mathbb{R} \oplus \mathbb{R} \oplus \mathbb{R}$

$$X_1 = \partial_{x_1} + x_1\partial_{x_3}, \qquad X_2 = x_1\partial_{x_2}, \qquad X_3 = x_2\partial_{x_3}$$

and set $\nu(x_i) = i$ for $i = 1, 2, 3$. The nontrivial commutators between these vector fields are

$$[X_1, X_2] = \partial_{x_2}, \qquad [X_2, X_3] = x_1\partial_{x_3}, \qquad [[X_1, X_2], X_3] = \partial_{x_3}.$$

Then the flag (computed at $x = 0$) is given by

$$\mathcal{D}^1_0 = \mathrm{span}\{\partial_{x_1}\}, \qquad \mathcal{D}^2_0 = \mathrm{span}\{\partial_{x_1}, \partial_{x_2}\}, \qquad \mathcal{D}^3_0 = \mathrm{span}\{\partial_{x_1}, \partial_{x_2}, \partial_{x_3}\}.$$

These coordinates are then linearly adapted to the flag but they are not privileged since $\nu(x_1\partial_{x_3}) = -2$, thus $X_1 \in \mathcal{F}^{(-2)} \setminus \mathcal{F}^{(-1)}$.

The following theorem is the main result of this section and states the existence of privileged coordinates.

**Theorem 10.32.** *Let $\mathcal{D}$ be a bracket generating distribution on a smooth manifold $M$ and $q \in M$. There always exists a system of privileged coordinates around $q$.*

The proof of this theorem is postponed to Section 10.4.3.

### 10.4.2 Description of the nonholonomic tangent space in privileged coordinates

We showed in Proposition 10.27 that given a horizontal vector field $X$ it induces a well defined vector field $\widehat{X}$ on the nonholonomic tangent space $T_q^f M$ at $q \in M$. The goal of this section is to discuss the peculiar structure of the vector field $\widehat{X}$ in privileged coordinates and deduce the corresponding description of $T_q^f M$.

We start with a description of the space of jets $J_q^k M$ and the equivalence relation defining the nonholonomic tangent space $T_q^f M$.

**Theorem 10.33.** *Let $\mathcal{D}$ be a bracket generating distribution on a smooth manifold $M$ and $q \in M$. In privileged coordinates we have the following*

*(i)* $J_q^f M = \{\sum_{i=1}^k t^i \xi_i \mid \xi_i \in \mathcal{D}_q^i\}$ *and* $\dim J_q^f M = kn_1 + (k-1)n_2 + \ldots + n_k$.

*(ii) Let* $j_1, j_2 \in J_q^f M$. *Then* $j_1 \sim j_2$ *if and only if* $j_1 - j_2 = \sum_{i=1}^k t^i \eta_i$, *where* $\eta_i \in \mathcal{D}_q^{i-1}$.

*Proof of Theorem 10.33, Claim (i), part 1.* We start by proving the following inclusion

$$J_q^f M \subset \left\{ \sum_{i=1}^k t^i \xi_i \mid \xi_i \in \mathcal{D}_q^i \right\}. \tag{10.23}$$

For any smooth variation $\gamma(t) = q \circ \overrightarrow{\exp} \int_0^\tau f_{u(t,s)} ds$, we can write the Volterra expansion

$$\gamma(t) = q + \sum_{i=1}^k \int_{0 \le s_i \le \ldots \le s_1 \le \tau} \cdots \int q \circ f_{u(t,s_1)} \circ \ldots \circ f_{u(t,s_i)} \, ds_1 \ldots ds_i + O(t^{k+1}). \tag{10.24}$$

Let us write (cf. Remark 10.17) the controls $u(t, s_i) = t\bar{u}(t, s_i)$ for some suitable families $\bar{u}(t, s_i)$. Then (10.24) becomes, using the fact that $f$ is linear in $u$, as follows

$$\gamma(t) = q + \sum_{i=1}^k t^i \int_{0 \le s_i \le \ldots \le s_1 \le \tau} \cdots \int q \circ f_{\bar{u}(t,s_1)} \circ \ldots \circ f_{\bar{u}(t,s_i)} \, ds_1 \ldots ds_i + O(t^{k+1}). \tag{10.25}$$

By definition of privileged coordinates we have $f_{u(t,s_i)} \in \mathcal{F}^{(-1)}$ for each $i$, hence $f_{\bar{u}(t,s_i)} \in \mathcal{F}^{(-1)}$ and

$$f_{\bar{u}(t,s_1)} \circ \ldots \circ f_{\bar{u}(t,s_i)} \in \mathcal{F}^{(-j)} \tag{10.26}$$

Let us apply the differential operator (10.26) to a coordinate function $x_\alpha^\beta$, with $\alpha = 1, \ldots, k$ and $\beta = 1, \ldots, n_\alpha$. Since $\nu(x_\alpha^\beta) = \alpha$ we have

$$f_{\bar{u}(t,s_1)} \circ \ldots \circ f_{\bar{u}(t,s_i)} x_\alpha^\beta \in \mathcal{F}^{(-i+\alpha)} \tag{10.27}$$

Therefore, for every $\alpha > i$, this function has positive weight and vanishes when evaluated at $x = 0$.

In privileged coordinates satisfying (10.22), this says that, for every $i = 1, \ldots, k$, the sum in (10.24) up to the $i^{th}$-term contains only element in $\mathcal{D}_q^i$. $\square$

To prove the converse inclusion we have to show that, given arbitrary elements $\xi_i \in \mathcal{D}^i_q$ for $i = 1, \dots, k$, we can find a smooth variation that has these vectors as elements of its jet. The proof is constructive and we start with some preliminary lemmas.

**Lemma 10.34.** *Let $m, n$ be two integers. Assume that we have two flows such that, as operators*

$$P_t = \mathrm{Id} + Vt^n + O(t^{n+1}),$$
$$Q_t = \mathrm{Id} + Wt^m + O(t^{m+1}).$$

*Then* $P_t Q_t P_t^{-1} Q_t^{-1} = \mathrm{Id} + [V, W]t^{n+m} + O(t^{n+m+1})$.

*Proof.* Define $R(t, s) := P_t Q_s P_t^{-1} Q_s^{-1}$. We are interested in the expansion of $R(t, t)$ with respect to $t$. Since $P_0 = Q_0 = \mathrm{Id}$, we have $R(0, s) = R(t, 0) = \mathrm{Id}$, for every $t, s \in \mathbb{R}$. This implies that, when writing the Taylor expansion of $P_t Q_s P_t^{-1} Q_s^{-1}$, only mixed derivatives in $t$ and $s$ give contribution. Using that

$$P_t^{-1} = \mathrm{Id} - t^n V + O(t^{n+1}), \qquad Q_s^{-1} = \mathrm{Id} - s^m W + O(s^{m+1}).$$

one gets, denoting $\|(t, s)\| = \sqrt{t^2 + s^2}$,

$$\begin{aligned}(\mathrm{Id} + t^n V + O(t^{n+1}))(\mathrm{Id} + s^m W + O(s^{m+1}))(\mathrm{Id} - t^n V + O(t^{n+1}))(\mathrm{Id} - s^m W + O(s^{m+1})) = \\ = \mathrm{Id} + t^n s^m (VW - WV) + O(\|(t, s)\|^{n+m+1}) \\ = \mathrm{Id} + t^n s^m [V, W] + O(\|(t, s)\|^{n+m+1})\end{aligned}$$

and the lemma is proved. □

**Exercise 10.35.** Assume that the flow $P_t$ satisfies $P_t = \mathrm{Id} + Vt^n + O(t^{n+1})$. Show that the nonautonomous vector field $V_t$ associated with $P_t$ satisfies $V_t = nt^{n-1}V + O(t^n)$.

**Lemma 10.36.** *For all $i_1, \dots, i_h \in \{1, \dots, k\}$ and $l \geq h$, there exists an admissible variation $u(t, s)$, depending only on the Lie bracket structure, such that*

$$q \circ \overrightarrow{\exp} \int_0^\tau f_{u(t,s)} ds = q + t^l [f_{i_1}, \dots, [f_{i_{h-1}}, f_{i_h}]](q) + O(t^{l+1}). \tag{10.28}$$

*Proof.* The lemma is proved by induction on $h$.

(i) For all $i = 1, \dots, k$ and $l \geq 1$ there exists an admissible variation $u(t, s)$ such that

$$q \circ \overrightarrow{\exp} \int_0^\tau f_{u(t,s)} ds = q + t^l f_i(q) + O(t^{l+1}).$$

In fact, it is sufficient to take $u = (u_1, \dots, u_k)$ such that $u_i = t^l$ and $u_j = 0$ for all $j \neq i$.

(ii) For all $i, j \in \{1, \dots, k\}$ and $l \geq 2$, we have to show that there exists an admissible variation $u(t, s)$ such that

$$q \circ \overrightarrow{\exp} \int_0^\tau f_{u(t,s)} ds = q + t^l [f_i, f_j](q) + O(t^{l+1}).$$

In fact, it is sufficient to apply Lemma 10.34 where $P_t$ and $Q_t$ are the flows generated by the nonautonomous vector fields $V_t = t^{l-1} f_{i_1}$ and $W_t = t f_{i_2}$, respectively.

An iteration of this argument completes the proof. □

In other words we proved that every bracket monomial of degree $i$ can be presented as the $i$-th term of a jet of some admissible variation. Now we prove that we can do the same for any linear combination of such monomials (recall that $\mathcal{D}^i$ is the linear span of all $i$-th order brackets).

**Lemma 10.37.** *Let $\pi = \pi(f_1, \ldots, f_m)$ be a bracket polynomial of degree $\deg \pi \leq l$. There exists an admissible variation $u(t,s)$, depending only on the Lie bracket structure, such that*

$$q \circ \overrightarrow{\exp} \int_0^\tau f_{u(t,s)} ds = q + t^l \pi(f_1, \ldots, f_m)(q) + O(t^{l+1}). \tag{10.29}$$

*Proof.* Let $\pi(f_1, \ldots, f_m) = \sum_{j=1}^N V_j(f_1, \ldots, f_m)$ where $V_j$ are monomials. By our previous argument we can find $u^j(t,s)$, for $s \in [0, \tau_j]$ such that

$$q \circ \overrightarrow{\exp} \int_0^{\tau_j} f_{u^j(t,s)} ds = q + t^l V_j(f_1, \ldots, f_m)(q) + O(t^{l+1}).$$

Then (10.29) is obtained choosing as $u(t,s)$, where $s \in [0, \tau]$ and $\tau := \sum_{j=1}^N \tau_j$ the concatenation of controls defined as follows

$$u(t,s) = u^j\left(t, s - \sum_{i=1}^{j-1} \tau_i\right), \qquad \text{if} \quad \sum_{i=1}^{j-1} \tau_i \leq s < \sum_{i=1}^{j} \tau_i, \quad 1 \leq j \leq N,$$

where the sum is understood to be zero for $j = 1$. $\square$

**Exercise 10.38.** Complete the proof by showing that the flow associated with $u$ has as main term in the Taylor expansion $\sum_j V_j$ at order $l$. Then prove, by using a time rescaling argument, that also any monomial of type $\alpha V$ for $\alpha \in \mathbb{R}$ can be presented in this way.

We are now in position to complete the proof of Claim (i) of Theorem 10.33.

*Proof of Theorem 10.33, Claim (i), part 2.* We have to prove the remaining inclusion

$$\left\{ \sum_{i=1}^k t^i \xi_i \mid \xi_i \in \mathcal{D}_q^i \right\} \subset J_q^f M. \tag{10.30}$$

Let us consider a $k$-th jet $j = \sum_{i=1}^k t^i \xi_i$, with $\xi_i \in \mathcal{D}_q^i$. We prove the statement by steps: at $i$-th step we built an admissible variation whose $i$-th Taylor polynomial coincide with the one of $j$.

- Thanks to Lemma 10.37, there exists a smooth admissible variation $\gamma_1(t)$ such that

$$\gamma_1(t) = q \circ \overrightarrow{\exp} \int_0^\tau f_{u(t,s)} ds, \qquad \dot{\gamma}(t) = \xi_1$$

  Then we will have $\gamma_1(t) = t\xi_1 + t^2 \eta_2 + O(t^3)$ where $\eta_2 \in \mathcal{D}_q^2$ from the first part of the proof.

- Thanks to Lemma 10.37, there exists a smooth admissible variation $\widetilde{\gamma}_2(t)$ such that

$$\widetilde{\gamma}_2(t) = q \circ \overrightarrow{\exp} \int_0^\tau f_{v(t,s)} ds, \qquad \widetilde{\gamma}_2(t) = t^2(\xi_2 - \eta_2) + O(t^3)$$

Defining[1] the product $\gamma_2(t) := (\widetilde{\gamma}_2 * \gamma_1)(t)$ we have

$$\begin{aligned}\gamma_2(t) &= t\xi_1 + t^2\eta_2 + t^2(\xi_2 - \eta_2) + t^3\eta_3 + O(t^4)\\ &= t\xi_1 + t^2\xi_2 + t^3\eta_3 + O(t^4)\end{aligned}$$

where $\eta_3 \in \mathcal{D}_q^3$.

At every step we can correct the right term of the jet and after $k$ steps we have the inclusion.

□

*Proof of Theorem 10.33, Claim (ii).* We have to prove that

$$j \sim j' \iff j - j' = \sum_{i=1}^{k} t^i \eta_i, \qquad \eta_i \in \mathcal{D}_q^{i-1}.$$

($\Rightarrow$). Assume that $j \sim j'$, where $j = J_q^k\gamma = \sum t^i\xi_i$ and $j' = J_q^k\gamma' = \sum t^i\xi_i'$. Then $\gamma' = \gamma \circ Q_t$ for some slow flow $Q_t \in \mathcal{P}_0^f$ of the form

$$\begin{aligned}Q_t &= Q_t^1 \circ \cdots \circ Q_t^h,\\ Q_t^i &= P_t^i \circ \overrightarrow{\exp} \int_0^\tau f_{tv^i(t,s)} ds \circ (P_t^i)^{-1},\end{aligned}$$

for some $P^i \in \mathcal{P}^f$ and some admissible variations $v_i(t,s)$, for $i = 1, \dots, h$. It is sufficient to prove it for the case $h = 1$. By formula (6.34) we have that

$$Q_t = P_t \circ \overrightarrow{\exp} \int_0^\tau f_{tv(t,s)} ds \circ P_t^{-1} = \overrightarrow{\exp} \int_0^\tau (\operatorname{Ad} P_t) f_{tv(t,s)} ds,$$

then by linearity of $f$ we have

$$Q_t = \overrightarrow{\exp} \int_0^\tau t(\operatorname{Ad} P_t) f_{v(t,s)} ds.$$

Now recall that $P_t = \overrightarrow{\exp} \int_0^\tau f_{w(t,\theta)} d\theta$ for some admissible variation $w(t,\theta)$ and from (6.31) we get

$$Q_t = \overrightarrow{\exp} \int_0^\tau t \ \overrightarrow{\exp} \int_0^s \operatorname{ad} f_{w(t,\theta)} d\theta \ f_{v(t,s)} ds.$$

Finally, if $\gamma(t) = q \circ \overrightarrow{\exp} \int_0^\tau f_{u(t,s)} ds$ we can write

$$\gamma'(t) = q \circ \overrightarrow{\exp} \int_0^\tau f_{u(t,s)} ds \circ \overrightarrow{\exp} \int_0^\tau t \ \overrightarrow{\exp} \int_0^s \operatorname{ad} f_{w(t,\theta)} d\theta \ f_{v(t,s)} ds.$$

Expanding with respect to $t$ we have $Q_t \simeq (Id + t\sum t^i V_i) = Id + \sum t^{i+1} V_i$ where $V_i$ is a bracket polynomial of degree $\leq i$. Due to the presence of $t$ it is easy to see that in the expansion of $\gamma'$ we will find the same terms of $\gamma$ plus something that belong to $\mathcal{D}^{i-1}$.

[1] we define the product of two curves $\gamma(t) = q \circ P_t$ and $\gamma'(t) = q \circ P_t'$ as follows: $(\gamma' * \gamma)(t) := q \circ P_t \circ P_t'$.

($\Leftarrow$). Assume now that $j = J_q^k\gamma = \sum t^i\xi_i$ and $j' = J_q^k\gamma' = \sum t^i\xi_i'$, with

$$j - j' = \sum_{i=1}^{k} t^i\eta_i, \qquad \eta_i \in \mathcal{D}_q^{i-1}.$$

We need to find a slow flow $Q_t$ such that $\gamma' = \gamma \circ Q_t$. In other words it is sufficient to prove that we can realize with a slow flow every jet of type $\sum_{i=1}^k t^i\eta_i$, $\eta_i \in \mathcal{D}_q^{i-1}$. To this purpose one just adapts arguments from the proof of part $(i)$, using the following crucial observation, which given an adaptation of Lemma 10.34.

**Lemma 10.39.** *Let $P_t, Q_t$ be two flows with $P_t \in \mathcal{P}^f$ and $Q_t \in \mathcal{P}_0^f$ (or $P_t \in \mathcal{P}_0^f$ and $Q_t \in \mathcal{P}^f$). Then $P_tQ_tP_t^{-1}Q_t^{-1} \in \mathcal{P}_0^f$.*

*Proof.* If $Q_t \in \mathcal{P}_0^f$ then $Q_t^{-1} \in \mathcal{P}_0^f$. Moreover from the definition of $\mathcal{P}_0^f$ we have that $P_tQ_tP_t^{-1} \in \mathcal{P}_0^f$. Hence also their composition is in $\mathcal{P}_0^f$. □

□

We have the following corollary of Theorem 10.33, part (i).

**Corollary 10.40.** *In privileged coordinates $(x_1, \ldots, x_k)$ defined by the splitting $\mathbb{R}^n = \mathbb{R}^{n_1} \oplus \ldots \oplus \mathbb{R}^{n_k}$ we have*

$$J_q^f M = \left\{ \begin{pmatrix} tx_1 + O(t^2) \\ t^2x_2 + O(t^3) \\ \vdots \\ t^k x_k \end{pmatrix} : x_i \in \mathbb{R}^{n_i}, i = 1, \ldots, k \right\}. \qquad (10.31)$$

*Proof.* Indeed we know that $\mathcal{D}^i = \mathbb{R}^{n_1} \oplus \ldots \oplus \mathbb{R}^{n_i}$ and writing

$$\xi_i = x_{i,1} + \ldots + x_{i,i}, \qquad x_{i,j} \in \mathbb{R}^{n_j}$$

we have, expanding and collecting terms

$$\begin{aligned} \sum_{i=1}^{k} t^i\xi_i &= t\xi_1 + t^2\xi_2 + \ldots + t^k\xi_k \\ &= tx_{1,1} + t^2(x_{2,1} + x_{2,2}) + \ldots + t^k(x_{k,1} + \ldots + x_{k,k}) \\ &= (tx_{1,1} + t^2x_{2,1} + \ldots + t^kx_{k,1}, t^2x_{2,2} + \ldots + t^kx_{k,2}, t^kx_{k,k}) \end{aligned}$$

□

We can finally deduce the structure of the nonholonomic tangent space in privileged coordinates

**Theorem 10.41.** *The nonholonomic tangent space $T_q^f M$ is a smooth manifold and $\dim T_q^f M = \dim M$. In privileged coordinates we have*

$$T_q^f M = \left\{ \begin{pmatrix} tx_1 \\ t^2x_2 \\ \vdots \\ t^k x_k \end{pmatrix} : x_i \in \mathbb{R}^{n_i}, i = 1, \ldots, k \right\}, \qquad (10.32)$$

*and dilations $\{\delta_\alpha\}_{\alpha>0}$ acts on $T_q^f M$ in the following quasi-homogeneous way*

$$\delta_\alpha(tx_1, \dots, t^k x_k) = (\alpha t x_1, \dots, \alpha^k t^k x_k).$$

*Proof.* It follows directly from Corollary 10.40 that two elements $j$ and $j'$ can be written in coordinates as

$$\begin{aligned} j &= (tx_1 + O(t^2), t^2 x_2 + O(t^3), \dots, t^k x_k), \\ j' &= (ty_1 + O(t^2), t^2 y_2 + O(t^3), \dots, t^k y_k). \end{aligned}$$

Moreover, thanks to Theorem 10.33, claim (ii), we have that $j \sim j'$ if and only if $x_i = y_i$ for all $i = 1, \dots, k$. Notice finally that

$$\dim T_q^f M = \sum_{i=1}^{k(q)} n_i(q) = n = \dim M.$$

□

*Remark* 10.42. Notice that a polynomial differential operator homogeneous with respect to $\nu$ (i.e., whose monomials are all of same weight) is homogeneous with respect to dilations $\delta_t : \mathbb{R}^n \to \mathbb{R}^n$ defined by

$$\delta_t(x_1, \dots, x_k) = (tx_1, t^2 x_2, \dots, t^k x_k), \qquad t > 0. \tag{10.33}$$

In particular for a homogeneous vector field $X$ of weight $h$ it holds $\delta_{t*} X = t^{-h} X$.

Now we can improve Proposition 10.27 and see that actually the jet of a horizontal vector field is a vector field on the tangent space and belongs to $\mathcal{F}^{(-1)}$ (in privileged coordinates).

**Lemma 10.43.** *Fix a set of privileged coordinates. Let $V \in \mathcal{F}^{(-1)}$, then the vector field $\widehat{V} \in \mathrm{Vec}(T_q^f M)$ induced on the nonholonomic tangent space writes as follows*

$$V = \begin{pmatrix} v_1(x) \\ v_2(x) \\ \vdots \\ v_k(x) \end{pmatrix} \quad \Longrightarrow \quad \widehat{V} = \begin{pmatrix} \widehat{v}_1(x) \\ \widehat{v}_2(x) \\ \vdots \\ \widehat{v}_k(x) \end{pmatrix}, \tag{10.34}$$

*where $\widehat{v}_i$ is the homogeneous term of order $i-1$ of $v_i$.*

*Proof.* Let $V \in \mathcal{F}^{(-1)}$ and $\gamma(t)$ be an admissible variation. When expressed in coordinates we have

$$V = \begin{pmatrix} v_1(x) \\ v_2(x) \\ \vdots \\ v_k(x) \end{pmatrix}, \qquad \gamma(t) = \begin{pmatrix} tx_1 + O(t^2) \\ t^2 x_2 + O(t^3) \\ \vdots \\ t^k x_k, \end{pmatrix}.$$

Thanks to Exercise 10.14, the coordinate representation of $(J_q^k V)(J_q^k \gamma)$ is given as the $k$-th jet of $tV(\gamma(t))$. Hence we compute

$$(J_q^k V)(J_q^k \gamma) = \begin{pmatrix} tv_1(tx_1 + O(t^2), \dots, t^k x_k) \\ tv_2(tx_1 + O(t^2), \dots, t^k x_k) \\ \vdots \\ tv_k(tx_1 + O(t^2), \dots, t^k x_k) \end{pmatrix}. \tag{10.35}$$

Notice that $V \in \mathcal{F}^{(-1)}$ means exactly that decomposing $V$ in coordinates as follows

$$V = \sum_{i=1}^{k} v_i(x) \frac{\partial}{\partial x_i} = \sum_{i=1}^{k} \sum_{j=1}^{n_i} v_i^j(x) \frac{\partial}{\partial x_i^j},$$

every $v_i$ is a function of order $\geq i-1$, since $\nu(\partial/\partial x_i^j) = -i$. Let us denote with $\widehat{v}_i$ the homogeneous part of $v_i$ of order $i-1$. To compute the value of $\widehat{V}$ then we have to restrict its action on admissible variations from $T_q^f M$, then evaluate and neglect the higher order part (that corresponds to the projection on the factor space) in order to have

$$v_i(tx_1 + O(t^2), \dots, t^k x_k) = t^{i-1} \widehat{v}_i(x_1, \dots, x_k) + O(t^i),$$

and using identity 10.35 we have

$$(J_q^k V)\Big|_{T_q^f M} = \begin{pmatrix} tv_1(tx_1 + O(t^2), \dots, t^k x_k) \\ tv_2(tx_1 + O(t^2), \dots, t^k x_k) \\ \vdots \\ tv_k(tx_1 + O(t^2), \dots, t^k x_k) \end{pmatrix} = \begin{pmatrix} t\widehat{v}_1 + O(t^2) \\ t^2 \widehat{v}_2 + O(t^3) \\ \vdots \\ t^k \widehat{v}_k + O(t^{m+1}) \end{pmatrix}, \tag{10.36}$$

from which (10.34) follows. □

*Remark* 10.44. Notice that, since $\widehat{v}_i$ is a homogeneous function of weight $i-1$, it depends only on variables $x_1, \dots, x_{i-1}$ of weight smaller or equal than its weight. Hence $\widehat{V}$ has the following triangular form

$$\widehat{V}(x) = \begin{pmatrix} \widehat{v}_1 \\ \widehat{v}_2(x_1) \\ \vdots \\ \widehat{v}_k(x_1, \dots, x_{k-1}) \end{pmatrix}. \tag{10.37}$$

A triangular vector field of the kind (10.37) is complete and its flow can be easily computed by a step by step substitution.

### 10.4.3 Existence of privileged coordinates: proof of Theorem 10.32.

Fix a generating family $\{f_1, \dots, f_m\}$ of the distribution $\mathcal{D}$. Assume that $\mathcal{D}$ is bracket generating of step $k$ at the point $q$

$$\mathcal{D}_q^1 \subset \mathcal{D}_q^2 \subset \ldots \subset \mathcal{D}_q^k = T_q M. \tag{10.38}$$

Denote by $d_j := \dim \mathcal{D}_q^j$ the dimension of the elements of the flag, for $j = 1, \dots, k$.

**Definition 10.45.** A set $V_1, \dots, V_n$ of $n$ vector fields on $M$ is said to be a *privileged frame* for $\mathcal{D}$ at $q$ if it satisfies the following properties:

(a) $V_i = \pi_i(f_1, \dots, f_m)$, where $\pi_i$ is some bracket polynomial, for $i = 1, \dots, n$,

(b) $\deg \pi_i \leq j$ for every $i \leq d_j$,

(c) $\mathcal{D}_q^j = \operatorname{span}\{V_1(q), \dots, V_{d_j}(q)\}$, for $j = 1, \dots, k$.

A privileged frame can be constructed as follows: choose $V_1, \dots, V_{d_1}$ among the vector fields $\{f_1, \dots, f_m\}$ in such a way that $\mathcal{D}_q = \operatorname{span}\{V_1(q), \dots, V_{d_1}(q)\}$, then fix $V_{d_1+1}, \dots, V_{d_2}$ among the set $\{[f_i, f_j] : i, j = 1, \dots, m\}$ in such a way that $\mathcal{D}_q^2 = \operatorname{span}\{V_1(q), \dots, V_{d_2}(q)\}$, and so on.

*Remark* 10.46. Given a privileged frame $V_1, \dots, V_n$, one can introduce on $T_qM$ the weight on the coordinates $(y_1, \dots, y_n)$ induced by the flag. In other words we write every element $v$ in $T_qM$ along the basis $V_1(q), \dots, V_n(q)$ and set

$$v = (y_1, \dots, y_n) = \sum_{i=1}^{n} y_i V_i(q), \qquad \text{where} \quad \nu(y_i) = w_i := j \quad \text{if} \quad d_{j-1} < i \leq d_j.$$

Identifying $v \in T_qM$ with a constant vector field, it makes sense to consider the value of a polynomial bracket $X = \pi(f_1, \dots, f_m)$ at the point $q$ and consider its weight $\nu(X)$.

Privileged coordinates are then easily build in terms of a privileged frame.

**Theorem 10.47.** *Let $V_1, \dots, V_n$ be a privileged frame at $q$. Then the map*

$$\Psi : \mathbb{R}^n \to M, \qquad \Psi(s_1, \dots, s_n) = q \circ e^{s_1 V_1} \circ \dots \circ e^{s_n V_n}, \tag{10.39}$$

*is a local diffeomorphism at $s = 0$ and its inverse $\Psi^{-1}$ defines privileged coordinates around $q$.*

*Proof.* The map (20.46) is a local diffeomorphism at $s = 0$ since

$$\left.\frac{\partial \Psi}{\partial s_i}\right|_{s=0} = V_i(q), \qquad i = 1, \dots, n, \tag{10.40}$$

and these vectors are linearly independent by property (c) of privileged frame. To complete the proof we have to show that:

(i) $\Psi_*^{-1}(\mathcal{D}_q^j) = \operatorname{span}\left\{\dfrac{\partial}{\partial s_1}, \dots, \dfrac{\partial}{\partial s_{d_j}}\right\}$, for every $j = 1, \dots, k$,

(ii) $\Psi_*^{-1} f_i \in \mathcal{F}^{(-1)}$ for every $i = 1, \dots, m$.

Claim (i), that is $\Psi$ defines linearly adapted coordinates, easily follows from property (c) of privileged frame and (10.40). On the other hand, claim (ii) is not trivial since it requires the computation of the differential of $\Psi$ at every point, and not only at $s = 0$.

We prove the following preliminary result.

**Lemma 10.48.** *Let $X = \pi(f_1, \dots, f_m)(q) \in \operatorname{Vec}(T_qM)$ be a bracket polynomial with $\nu(X) \leq h$. Given a polynomial vector field on $T_qM$*

$$Y(y) := \sum y_{i_l} \cdots y_{i_1} (\operatorname{ad} V_{i_l} \circ \cdots \circ \operatorname{ad} V_{i_1} X)(q), \tag{10.41}$$

*there exists polynomials $p_i(y) \in \mathcal{F}^{(w_i - h)}$ for $i = 1, \dots, n$ such that*

$$Y(y) := \sum_{i=1}^{n} p_i(y) V_i(q).$$

We stress that the weight of the polynomial $p_i$ in the previous Lemma is *independent* on the degree of the polynomial vector field.

*Proof of Lemma 10.48.* It easily follows from definition of weights that

$$\operatorname{ad} V_{i_l} \circ \cdots \circ \operatorname{ad} V_{i_1}(X) \in \mathcal{F}^{(-w)}, \qquad w = \sum_{j=1}^{l} w_{i_j} + h.$$

By additivity, every term in the sum (10.41) belongs to $\mathcal{F}^{(-h)}$. Then if we rewrite the sum (10.41) in terms of the basis $V_i(q)$, for $i = 1, \ldots, n$ we have that every coefficient $p_i(y)$ must belong to $\mathcal{F}^{(w_i-h)}$, since $\nu(V_i(q)) = w_i$. □

The proof of existence of privileged coordinates is completed by the following proposition, applied in the particular case $h = 1$.

**Proposition 10.49.** *Let $X = \pi(f_1, \ldots, f_m)$ be a bracket polynomial with $\nu(X) \leq h$ and $\Psi$ be the map defined in* (20.46)*. Then $\Psi_*^{-1} X \in \mathcal{F}^{(-h)}$.*

*Proof.* Writing the vector field $\Psi_*^{-1} X$ in coordinates

$$\Psi_*^{-1} X = \sum_{i=1}^{n} a_i(s) \frac{\partial}{\partial s_i}, \tag{10.42}$$

the statement is proved if we show that $a_i \in \mathcal{F}^{(w_i-h)}$. We compute the differential of $\Psi$ (cf. also Exercice 2.32)

$$\begin{aligned}
\Psi_* \frac{\partial}{\partial s_i} &= \frac{\partial}{\partial \varepsilon}\bigg|_{\varepsilon=0} q \circ e^{s_1 V_1} \circ \cdots \circ e^{(s_i+\varepsilon)V_i} \circ \cdots \circ e^{s_n V_n} \\
&= q \circ e^{s_1 V_1} \circ \cdots \circ e^{s_i V_i} \circ V_i \circ e^{s_{i+1} V_{i+1}} \circ \cdots \circ e^{s_n V_n} \\
&= \underbrace{q \circ e^{s_1 V_1} \circ \cdots \circ e^{s_n V_n}}_{\Psi(s)} \circ e^{-s_n V_n} \circ \cdots \circ e^{-s_{i+1} V_{i+1}} \circ V_i \circ e^{s_{i+1} V_{i+1}} \circ \cdots \circ e^{s_n V_n}.
\end{aligned}$$

In geometric notation we can write

$$\Psi_* \frac{\partial}{\partial s_i} = e_*^{s_n V_n} \cdots e_*^{s_{i+1} V_{i+1}} V_i \Big|_{\Psi(s)}. \tag{10.43}$$

Remember that, as operator on functions, $e_*^{tY} = e^{-t \operatorname{ad} Y}$. This implies that in (10.43) we have a series of bracket polynomials. Applying $\Psi_*$ to (10.42) one gets

$$X\Big|_{\Psi(s)} = \sum_{i=1}^{n} a_i(s) e_*^{s_n V_n} \cdots e_*^{s_{i+1} V_{i+1}} V_i \Big|_{\Psi(s)}.$$

Now we apply $e_*^{-s_1 V_1} \cdots e_*^{-s_n V_n}$ to both sides to compute the vector field at the point $q$

$$e_*^{-s_1 V_1} \cdots e_*^{-s_n V_n} X\Big|_q = \sum_{i=1}^{n} a_i(s) e_*^{-s_1 V_1} \cdots e_*^{-s_{i-1} V_{i-1}} V_i \Big|_q. \tag{10.44}$$

Rewriting the last identity in the basis $V_1(q), \ldots, V_n(q)$ we have

$$\sum_{i=1}^{n} b_i(s) V_i(q) = \sum_{i,j=1}^{n} a_i(s)(V_i(q) + \varphi_{ij}(s) V_j(q)), \tag{10.45}$$

for some smooth functions $b_i, \varphi_{ij}$ such that $\varphi_{ij}(0) = 0$. Applying Lemma 10.48 to $X$ and $V_i$, for $i = 1, \ldots, n$, we have

$$b_i \in \mathcal{F}^{(w_i - h)}, \qquad \varphi_{ij} \in \mathcal{F}^{(w_j - w_i)}.$$

On the other hand we can rewrite relation between coefficients as follows

$$B(s) = A(s)(I + \Phi(s)),$$

where we denote $B(s) = (b_1(s), \ldots, b_n(s))$, $A(s) = (a_1(s), \ldots, a_n(s))$ and $\Phi(s) = (\varphi_{ij}(s))_{ij}$. Notice that $I + \Phi(s)$ is invertible. Thus we get

$$\begin{aligned} A(s) &= B(s)(I + \Phi(s))^{-1} \\ &= \sum_{p \geq 0} (-1)^p (B\Phi^p)(s), \end{aligned}$$

and we observe that

$$\begin{aligned} (B)_i &= b_i \in \mathcal{F}^{(w_i - h)}, \\ (B\Phi)_i &= \sum_{j=1}^{n} b_j \varphi_{ji} \in \mathcal{F}^{(w_j - h + w_i - w_j)} = \mathcal{F}^{(w_i - h)}. \end{aligned}$$

Iterating the argument it follows that $(B\Phi^p)_i \in \mathcal{F}^{(w_i - h)}$ for every $p \geq 0$. Hence $a_i \in \mathcal{F}^{(w_i - h)}$. □

□

*Remark* 10.50. The previous proof can be rewritten in purely algebraic way through chronological notation. In the above proof nothing changes if we consider some permutation $\sigma = (i_1, \ldots, i_n)$ of $(1, \ldots, n)$ and work with the map

$$\Psi_\sigma : (s_1, \ldots, s_n) \mapsto q \circ e^{s_{i_n} V_{i_n}} \circ \ldots \circ e^{s_{i_1} V_{i_1}}.$$

We stress that, even if we are allowed to switch the position of the vector fields in the composition, the coordinate $s_i$ has to correspond to the vector field $V_i$, for $i = 1, \ldots, n$.

We summarize the previous considerations in the next corollary.

**Corollary 10.51.** *Let $V_1, \ldots, V_n$ be a privileged frame at $q$ and $\sigma = (i_1, \ldots, i_n)$ a permutation of $\{1, \ldots, n\}$. Then the map*

$$\Psi_\sigma : \mathbb{R}^n \to M, \qquad \Psi_\sigma(s_1, \ldots, s_n) = q \circ e^{s_{i_n} V_{i_n}} \circ \ldots \circ e^{s_{i_1} V_{i_1}}, \tag{10.46}$$

*is a local diffeomorphism at $s = 0$ and its inverse $\Psi_\sigma^{-1}$ defines privileged coordinates around $q$.*

*Remark* 10.52. As a particular case of Corollary 10.51 we can consider the coordinate map

$$\Phi : (x_1, \ldots, x_n) \mapsto q \circ e^{x_n V_n} \circ \ldots \circ e^{x_1 V_1}.$$

Computing the differential $\Phi_*$ (cf. also Exercice 2.32) it is easy to see that for every $i = 1, \ldots, n$

$$\Phi_*^{-1} V_i \Big|_{x_1 = \cdots = x_{i-1} = 0} = \partial_{x_i}. \tag{10.47}$$

This implies in particular that for $i = 1, \dots, d_1$, we have in coordinates

$$V_i = \partial_{x_i} + \sum_{j \geq d_1} a_{ij}(x_1, \dots, x_{d_1})\partial_{x_j}, \tag{10.48}$$

for some functions $a_{ij}$ depending only on the coordinates of the first layer. Indeed the set of vector fields $\{V_i\}_{i=1,\dots,d_1}$ are chosen among $f_1, \dots, f_m$, (generating $\mathcal{D}_q$) and have weight $-1$.

**Exercise 10.53.** Let $V_1, \dots, V_n$ be a privileged frame at $q$. Prove that the map

$$\Psi_+ : \mathbb{R}^n \to M, \qquad \Psi_+(s_1, \dots, s_n) = q \circ e^{\sum_{i=1}^n s_i V_i}, \tag{10.49}$$

is a local diffeomorphism at $s = 0$ and its inverse $\Psi_+^{-1}$ defines privileged coordinates around $q$.

#### Equiregular case and uniform privileged coordinates

In a neighbourhood $O_q$ of a point $q \in M$ where the structure is equiregular, we can fix a frame $V_1, \dots, V_n$ which is privileged at every point in a neighborhood of $p$.

A continuous (actually smooth) system of privileged coordinates in a neighborhood $\Omega$ of a point $p \in M$ is given by the map

$$\overline{\Psi} : \Omega \times \mathbb{R}^n \to M, \qquad \Psi(q, s_1, \dots, s_n) = q \circ e^{s_1 V_1} \circ \dots \circ e^{s_n V_n}, \tag{10.50}$$

**Exercise 10.54.** Given a sub-Riemannian structure $\{f_1, \dots, f_m\}$ that is equiregular and fix a frame $V_1, \dots, V_n$ which is privileged at every point in a neighborhood of $p$ defining privileged coordinates as in (10.50). Denote by $\{\widehat{f}_1^q, \dots, \widehat{f}_m^q\}$ the nilpotent approximation at the point $q$. Prove that the family $\{\widehat{f}_1^q, \dots, \widehat{f}_m^q\}$ is smooth with respect to $q$, seen as a family of vector fields in $\mathbb{R}^n$.

Notice that the existence of a continuous system of privileged coordinates, in general, is not ensured in a neighborhood of a singular point.

### 10.4.4 Nonholonomic tangent spaces in low dimension

In Riemannian geometry the above procedure becomes very easy since when $k = 1$ we have that $J_q^1 M = T_q M$ and moreover every admissible variation is an admissible trajectory. This implies that if $(M, \mathbf{U}, f)$ is a Riemannian manifold and $X$ is a vector field on $M$, then the vector field $\widehat{X}$ induced on the tangent space $T_q^f M = T_q M$ is simply the constant vector field defined on $T_q M$ defined by the value of $X$ at $q$. Moreover, every local basis of the tangent space is a privileged frame and defines privileged coordinates

As soon as the structure is not Riemannian, the structure of the noholonomic tangent space can depend on the point $q$ and on the growth vector $(d_1, \dots, d_k)$ of the distribution $\mathcal{D}$ at $q$. Let us study the low dimensional cases.

If we consider regular sub-Riemannian distributions, namely when the dimension of $\mathcal{D}_q$ is constant with respect to $q$, then the simplest case is obtained in dimension $n = 3$ for a distribution of rank 2.

If the distribution is also equiregular, i.e, the dimension of all $\mathcal{D}_q^j$ is constant with respect to $q$, then the growth vector is necessarily $(2, 3)$ at every point. In this case the nonholonomic tangent space is unique and given by the Heisenberg group.

**Example 10.55** (Heisenberg group)**.** Assume $n = 3$ and that the growth vector is $(2, 3)$. Then we consider coordinates $(x_1, x_2, x_3)$ and weights $(w_1, w_2, w_3) = (1, 1, 2)$. Since we work locally around the point $q$, it is not restrictive to assume that $\mathcal{D}$ is locally generated by two vector fields $f_1, f_2$ and that we can choose as a privileged frame

$$V_1 = f_1, \qquad V_2 = f_2, \qquad V_3 = [f_1, f_2]. \tag{10.51}$$

Using privileged coordinates defined in Remark 10.52, we have that

$$V_1 = f_1 = \partial_{x_1}, \qquad V_2 = f_2 = \partial_{x_2} + \alpha x_1 \partial_{x_3}, \tag{10.52}$$

for some $\alpha \in \mathbb{R}$. On the other hand since

$$V_3 = [f_1, f_2] = \alpha \partial_{x_3} \tag{10.53}$$

and $V_3(0) = \partial_{x_3}$ from (10.47) we get $\alpha = 1$. This gives the following normal form for the generating family of the nonholonomic tangent space

$$f_1 = \partial_{x_1}, \qquad f_2 = \partial_{x_2} + x_1 \partial_{x_3}. \tag{10.54}$$

□

If we admit the regular distribution $\mathcal{D}$ of rank 2 in dimension $n = 3$ to be not equiregular, then the growth vector can be of the form $(2, \ldots, 2, 3)$ at some singular points. In the simplest case, for a growth vector $(2, 2, 3)$, the nonholonomic tangent space is the Martinet flat space.

**Example 10.56** (Martinet flat)**.** Assume $n = 3$ and that growth vector is $(2, 2, 3)$. This means that we have coordinates $(x_1, x_2, x_3)$ with corresponding weights $(w_1, w_2, w_3) = (1, 1, 3)$. Since we work locally around the point $q$, it is not restrictive to assume that $\mathcal{D}$ is locally generated by two vector fields $f_1, f_2$ and that we can choose as a privileged frame

$$V_1 = f_1, \qquad V_2 = f_2, \qquad V_3 = [f_1, [f_1, f_2]]. \tag{10.55}$$

Indeed if the three vector fields above are not linearly independent then we can choose $V_3 = [f_2, [f_2, f_1]]$ and we reduce to the previous case by switching the role of $f_1$ and $f_2$. Moreover denote $f_u := u_1 f_1 + u_2 f_2$ and consider the linear map

$$\varphi : \mathbb{R}^2 \to T_q M / \mathcal{D}_q, \qquad \varphi(u_1, u_2) := [f_u, [f_1, f_2]](q) \mod \mathcal{D}_q.$$

Since $\varphi$ is surjective (by bracket-generating assumption) and $\dim T_q M / \mathcal{D}_q = 1$, then $\ker \varphi$ is one dimensional. Thus, up to a rotation of constant angle of the generating family $f_1, f_2$ (which does not change the value $[f_1, f_2]$), we can assume that $f_2 \in \ker \varphi$. In particular this implies

$$[f_2, [f_1, f_2]] = 0. \tag{10.56}$$

Using privileged coordinates defined in Remark 10.52, we have that

$$V_1 = f_1 = \partial_{x_1}, \qquad V_2 = f_2 = \partial_{x_2} + x_1 a(x_1, x_2) \partial_{x_3}, \tag{10.57}$$

for some smooth function $a(x_1, x_2)$. Since $\nu(f_2) = -1$ then $a(x_1, x_2) = \alpha x_1 + \beta x_2$ for some $\alpha, \beta \in \mathbb{R}$ and we get the coordinate representation

$$f_1 = \partial_{x_1}, \qquad f_2 = \partial_{x_2} + (\alpha x_1^2 + \beta x_1 x_2) \partial_{x_3}. \tag{10.58}$$

Since $[f_1, [f_1, f_2]] = 2\alpha\partial_{x_3}$, the requirement $V_3|_{x=0} = \partial_{x_3}$ in (10.55) gives $\alpha = 1/2$. Moreover for this value o $\alpha$ we have $[f_2, [f_1, f_2]] = \beta\partial_{x_3}$ and the condition (10.56) gives $\beta = 0$. We have then the normal form for the generating family of the nonholonomic tangent space

$$f_1 = \partial_{x_1}, \qquad f_2 = \partial_{x_2} + \frac{1}{2}x_1^2\partial_{x_3}, \qquad f_3 = \partial_{x_3}. \tag{10.59}$$

□

If we consider non-regular distributions, then the simplest case is obtained as the nonholonomic tangent space to a distribution $\mathcal{D}$ in dimension $n = 2$ in some singular point. Analogously to the previous case the growth vector can be of the form $(1, \ldots, 1, 2)$ and the simplest case is obtained when the growth vector is $(1, 2)$. In this case the nonholonomic tangent space is the Grushin plane.

**Example 10.57** (Grushin plane)**.** Assume $n = 2$ and that the growth vector is $(1, 2)$. Then we consider coordinates $(x_1, x_2)$ and weights $(w_1, w_2) = (1, 2)$. Let $\{f_1, f_2\}$ be a generating family for $\mathcal{D}$. It is not restrictive to assume that

$$V_1 = f_1, \qquad V_2 = [f_1, f_2]$$

By properties of privileged coordinates defined in Remark 10.52, we have that

$$V_1 = f_1 = \partial_{x_1}, \qquad V_2 = [f_1, f_2] = \partial_{x_2}.$$

Moreover $f_2$ should be a vector field of weight $-1$ that vanishes at $x = 0$ so it is necessarily of the form

$$f_2 = \alpha x_1 \partial_{x_2},$$

for some $\alpha \in \mathbb{R}$. The condition $[f_1, f_2] = \partial_{x_2}$ gives $\alpha = 1$ and we obtain the normal form for the generating family of the nonholonomic tangent space

$$f_1 = \partial_{x_1}, \qquad f_2 = x_1\partial_{x_2}. \tag{10.60}$$

## 10.5 Metric meaning

In this section we study the interplay between the nonholonomic tangent space and the sub-Riemannian distance.

Given a sub-Riemannian structure $(M, \mathbf{U}, f)$, with $\dim M = n$, and let us denote by $\{f_1, \ldots, f_m\}$ a generating family and fix a point $q$ where the sub-Riemannian distribution is bracket generating of step $k$.

Once we fix a privileged coordinate chart in a neigborhood of $q$, we can treat the vector fields $\{f_1, \ldots, f_m\}$ as vector fields defined on (an open set of) $\mathbb{R}^n$, and introduce the corresponding family of dilations $\{\delta_\alpha\}_{\alpha>0}$ defined in (10.33).

Then one can consider the family $\{\widehat{f}_1, \ldots, \widehat{f}_m\}$ of vector fields defined by the nilpotent approximation of the generating family.

The next lemma explains in more geometric terms, once given a vector field $V$, in which sense the vector field $\widehat{V}$ defined on $T_q^f M$ is an approximation of $V$.

**Lemma 10.58.** *Let $V$ be a horizontal vector field on $M$ and let $\widehat{V}$ be its nilpotent approximation. In privileged coordinates around $q$ we have equality*

$$\varepsilon\delta_{\frac{1}{\varepsilon}*}V = \widehat{V} + \varepsilon W^\varepsilon, \tag{10.61}$$

*where $\{\delta_\alpha\}_{\alpha>0}$ denotes the family of dilations defined in* (10.33) *and $W^\varepsilon$ depends smoothly on the parameter $\varepsilon$. In particular $\widehat{V}$ is characterized as follows*

$$\widehat{V} = \lim_{\varepsilon\to 0} \varepsilon\delta_{\frac{1}{\varepsilon}*}V. \tag{10.62}$$

*Proof.* Recall that in privileged coordinates any horizontal vector fields $V$ belongs to $\mathcal{F}^{(-1)}$ and $\widehat{V}$ is its homogeneous part of degree $-1$. Let us write $V = \widehat{V} + W$ and apply the dilation $\delta_{\frac{1}{\varepsilon}*}$ to both sides of the equality. We have

$$\delta_{\frac{1}{\varepsilon}*}V = \delta_{\frac{1}{\varepsilon}*}\widehat{V} + \delta_{\frac{1}{\varepsilon}*}W = \frac{1}{\varepsilon}\widehat{V} + \delta_{\frac{1}{\varepsilon}*}W, \tag{10.63}$$

where we used the homogeneity of $\widehat{V}$ (cf. Remark 10.42). Noting that $W \in \mathcal{F}^{(0)}$, hence setting $W^\varepsilon := \varepsilon\delta_{\frac{1}{\varepsilon}*}W$ we have that $W^\varepsilon$ is smooth with respect to $\varepsilon$ and $\varepsilon W^\varepsilon \to 0$ for $\varepsilon \to 0$. $\square$

Geometrically this procedure means that if we consider a small neighborhood of the point $q$ and we make a nonisotropic dilation (with scaling related to the local structure of the Lie bracket) then $\widehat{V}$ catches the principal terms of $V$. This is a nonholonomic analogous of the linearization of a vector field in the Euclidean case.

### 10.5.1 Convergence of the sub-Riemannian distance and the Ball-Box theorem

Following the above construction, given a sub-Riemannian structure $(M, \mathbf{U}, f)$, with $\dim M = n$, and $\{f_1, \ldots, f_m\}$ as a generating family we can introduce the vector fields on a neigborhood of $q$ where we fixed privileged coordinates

$$f_i^\varepsilon := \varepsilon\delta_{\frac{1}{\varepsilon}*}f_i, \qquad i = 1, \ldots, m. \tag{10.64}$$

and define on $\mathbb{R}^n$ the following sub-Riemannian structures

a) the *$\varepsilon$-approximation* $f^\varepsilon$ whose generating family is $\{f_1^\varepsilon, \ldots, f_m^\varepsilon\}$, for every $\varepsilon > 0$,

b) the *nilpotent approximation* $\widehat{f}$ whose generating family is $\{\widehat{f}_1, \ldots, \widehat{f}_m\}$.

Thanks to Lemma 10.58 we have that $f_i^\varepsilon \to \widehat{f}_i$ for every $i = 1, \ldots, m$.

Moreover, from the definition (10.64) of the vector fields $f_i^\varepsilon$, it follows directly that the flag of the sub-Riemannian structure defined by $f^\varepsilon$ is the same as the one of the original one, since they are related by a change of coordinates. We have also the following basic observation on the structure of the Lie algebra generated by $\{\widehat{f}_1, \ldots, \widehat{f}_m\}$.

**Proposition 10.59.** *The Lie algebra* $\mathrm{Lie}\{\widehat{f}_1, \ldots, \widehat{f}_m\}$ *is a finite-dimensional nilpotent Lie algebra, which is bracket generating of step $k$, where $k$ is the nonholonomic degree of the sub-Riemannian structure at $q$.*

*Proof.* Consider privileged coordinates in a neighborhood of the point $q$. Then $\widehat{f}_i$ has weight $-1$ and is homogeneous with respect to the dilation $\{\delta_\alpha\}_{\alpha>0}$. Moreover, for any bracket monomial of length $j$ we have

$$\nu([\widehat{f}_{i_1}, \ldots, [\widehat{f}_{i_{j-1}}, \widehat{f}_{i_j}]]) = -j.$$

Since every vector field $V$ satisfies $\nu(V) \geq -k$, it follows that every bracket of length $j \geq k$ is necessarily zero. $\square$

Next, we investigate on the convergence of the corresponding sub-Riemannian distances. Let us start with the following identity, relating the original sub-Riemannian distance to the $\varepsilon$-approximating one. The proof is left as an exercice for the reader.

**Proposition 10.60.** *Let $d^\varepsilon$ and $d$ be the sub-Riemannian distances on $\mathbb{R}^n$ associated with the sub-Riemannian structures $f^\varepsilon$ and $f$, respectively. Then for every $x, y \in \mathbb{R}^n$ we have*

$$d^\varepsilon(x, y) = \frac{1}{\varepsilon} d(\delta_\varepsilon(x), \delta_\varepsilon(y)). \tag{10.65}$$

Proposition 10.60 is saying that $d^\varepsilon$ is $d$ when we "blow-up" the space near the point $q$ and rescale the distances. This relations rewrites as follows in terms of balls.

**Corollary 10.61.** *Let $B(x, r)$ (resp. $B^\varepsilon(x, r)$) be the sub-Riemannian ball with respect to the distance $d$ (resp. $d^\varepsilon$). Then for every $r > 0$ and $\varepsilon > 0$ one has*

$$\delta_\varepsilon(B^\varepsilon(x, r)) = B(\delta_\varepsilon x, \varepsilon r). \tag{10.66}$$

*In particular $\delta_\varepsilon(B^\varepsilon(0, 1)) = B(0, \varepsilon)$ for every $\varepsilon > 0$.*

**Exercise 10.62.** Prove Corollary 10.61.

The previous results relate the original distance $d$ with the approximating one $d^\varepsilon$. Next we move to the convergence of $d^\varepsilon$ for $\varepsilon \to 0$.

We start from an auxiliary proposition, studying the convergence of the end-point maps. Denote $E_x^\varepsilon$ and $\widehat{E}_x$ the end-point map of the approximating frame and of the nilpotent one based at a point $x \in \mathbb{R}^n$.

**Proposition 10.63.** *Let $x \in \mathbb{R}^n$. Then $E_x^\varepsilon \to \widehat{E}_x$ uniformly on balls in $L^2([0, 1], \mathbb{R}^k)$.*

*Proof.* Fix a control $u \in L^2([0, 1], \mathbb{R}^k)$ and consider the solution $x^\varepsilon(t)$ and $\widehat{x}(t)$ of the two systems

$$\dot{x} = \sum_{i=1}^m u_i(t) f_i^\varepsilon(x), \qquad \dot{x} = \sum_{i=1}^m u_i(t) \widehat{f}_i(x), \tag{10.67}$$

with fixed initial condition $x(0) = x \in \mathbb{R}^n$. Using Lemma 10.58, we write $f_i^\varepsilon = \widehat{f}_i + \varepsilon W_i^\varepsilon$ and the first equation in (10.67) becomes

$$\dot{x} = \sum_{i=1}^m u_i(t) \widehat{f}_i(x) + \varepsilon \sum_{i=1}^m u_i(t) W_i^\varepsilon(x). \tag{10.68}$$

In the right hand side the term

$$W_t^\varepsilon(x) := \varepsilon \sum_{i=1}^m u_i(t) W_i^\varepsilon(x), \tag{10.69}$$

is a non-autonomous vector field smoothly depending on the parameter $\varepsilon$. Moreover $W_t^\varepsilon(x) \to 0$ uniformly for $\varepsilon \to 0$. From classical result in ODE theory (continuity with respect to parameters) it follows that the solution $x^\varepsilon(t)$ converges uniformly on $[0,T]$ to the solution $\widehat{x}(t)$. In particular the final points converge. Notice that, since nilpotent vector fields are complete (cf. Remark 10.44), the solution $\widehat{x}(t)$ is defined for all $t \in \mathbb{R}$. $\square$

We notice that actually, thanks to the smoothness of the end-point map, the convergence in Proposition 10.63 holds in the $C^\infty$ sense.

We now prove a key uniform Hölder estimate (with respect to $\varepsilon$) for the approximating sub-Riemannian distance.

**Proposition 10.64.** *For every compact $K \subset \mathbb{R}^n$ there exists $\varepsilon_0, C > 0$, depending on $K$, such that*

$$d^\varepsilon(x,y) \le C|x-y|^{1/k}, \qquad \forall\, \varepsilon \in (0,\varepsilon_0),\ \forall\, x,y \in K. \tag{10.70}$$

*where $k$ is the degree of nonholonomy of the sub-Riemannian structure.*

*Proof.* Let $\widehat{V}_1, \dots, \widehat{V}_n$ be a privileged frame for the nilpotent system $\widehat{f}$ at the origin (cf. Definition 10.45), such that $\widehat{V}_i = \pi_i(\widehat{f}_1, \dots, \widehat{f}_m)$ for some bracket polynomials $\pi_i$, where $i = 1, \dots, n$. By construction we have

$$\widehat{V}_1(0) \wedge \ldots \wedge \widehat{V}_n(0) \neq 0. \tag{10.71}$$

By continuity, this implies that they are linearly independent also in a small neighborhood of the origin and, thanks to quasi-homogeneity, this implies

$$\widehat{V}_1(x) \wedge \ldots \wedge \widehat{V}_n(x) \neq 0, \qquad \forall\, x \in \mathbb{R}^n. \tag{10.72}$$

Let $V_i^\varepsilon := \pi_i(f_1^\varepsilon, \dots, f_k^\varepsilon)$ denote vector fields defined by the same bracket polynomials, written in terms of the vector fields of the approximating system. Fix a compact $K \subset \mathbb{R}^n$ and let $\varepsilon_0 = \varepsilon_0(K)$ be chosen such that

$$V_1^\varepsilon(x) \wedge \ldots \wedge V_n^\varepsilon(x) \neq 0, \qquad \forall\, x \in K, \forall\, \varepsilon \le \varepsilon_0. \tag{10.73}$$

Recall that by Lemma 10.37, given a bracket polynomial $\pi_i(g_1, \dots, g_k)$, with $\deg \pi_i = w_i$, there exists an admissible variation $u_i(t,s)$, depending only on $\pi_i$, such that

$$\overrightarrow{\exp} \int_0^1 g_{u_i(t,s)} ds = \mathrm{Id} + t^{w_i} \pi_i(g_1, \dots, g_k) + O(t^{w_i+1}).$$

If we apply this lemma for $g_i := f_i^\varepsilon$ we find $u_i(t,s)$ such that

$$\overrightarrow{\exp} \int_0^1 f^\varepsilon_{u_i(t,s)} ds = \mathrm{Id} + t^{w_i} V_i^\varepsilon + O(t^{w_i+1}), \qquad \forall\, \varepsilon > 0,$$

where we recall $w_i = \deg \pi_i$. Next we define the map for $\varepsilon > 0$

$$\Phi^\varepsilon(t_1, \dots, t_n, x) := x \circ \overrightarrow{\exp} \int_0^1 f^\varepsilon_{u_1(t_1^{1/w_1}, s)} ds \circ \ \ldots\ \circ \overrightarrow{\exp} \int_0^1 f^\varepsilon_{u_n(t_n^{1/w_n}, s)} ds. \tag{10.74}$$

Notice that we have the expansion

$$x \circ \overrightarrow{\exp} \int_0^1 f^\varepsilon_{u_i(t_i^{1/w_i},s)} ds = x + t_i V_i^\varepsilon(x) + O(t_i^{\frac{w_i+1}{w_i}}). \tag{10.75}$$

In particular (10.75) is a $C^1$ map in a neighborhood of $t = 0$ but, in general, it is not $C^2$ as soon as $w_i > 1$.

From this observation it follows that $\Phi^\varepsilon$ is $C^1$ as a function of $t$, being a composition of $C^1$ maps. Clearly $\Phi^\varepsilon$ is smooth as a function of $x$. Combining the contributions of (10.75) we obtain the expansion

$$\Phi^\varepsilon(x; t_1, \dots, t_n) = x + \sum_{i=1}^n t_i V_i^\varepsilon(x) + o(|t|), \tag{10.76}$$

This implies that the partial derivatives

$$\frac{\partial \Phi^\varepsilon}{\partial t_i}\Big|_{t=0} = V_i^\varepsilon(x), \tag{10.77}$$

are linearly independent at the origin thanks to (10.73) and $\Phi^\varepsilon$ is a local diffeomorphism at $t = (t_1, \dots, t_n) = 0$. Applying classical Implicit Function Theorem (see Corollary 2.58) we have that there exists a constant $c > 0$ satisfying

$$B(x, cr) \subset \Phi^\varepsilon(x; B(0, r)), \quad x \in K, \tag{10.78}$$

where here $B(x, r)$ denotes the ball in $\mathbb{R}^n$ and $c$ is independent of $x, \varepsilon$ and the parameter $r$ is small enough.

Let us denote now with $E_x$ the end-point map based at the point $x \in \mathbb{R}^n$ (with analogous meaning for $E^\varepsilon_x, \widehat{E}_x$), and with $\mathcal{B}$ the unit ball in $L_2([0,1], \mathbb{R}^m)$.

We claim that (10.78) implies that there exists a constant $c'$ such that for all $r > 0$ and $\varepsilon > 0$ small enough

$$B(x, c'r) \subset E^\varepsilon_x(r^{\frac{1}{k}} \mathcal{B}), \tag{10.79}$$

Since $t \mapsto u_i(t, \cdot)$ is a smooth map for every $i$, and $u_i(0, \cdot) = 0$ we have that there exist a constant $c_i$ such that

$$t \in B(0, r) \Rightarrow u_i(t, \cdot) \in c_i r \mathcal{B}, \tag{10.80}$$

$$\Rightarrow u_i(t^{1/w_i}, \cdot) \in c_i r^{1/w_i} \mathcal{B}, \tag{10.81}$$

for all $r > 0$ small enough.

For such values of $r > 0$ we have thanks to the inclusion (10.79) that for every $x, y \in K$ such that $|x - y| \leq cr$ then we have also $d^\varepsilon(x, y) \leq r^{1/k}$. Here we used the fact that $d^\varepsilon$ is the infimum of norm of $u$ such that $E^\varepsilon_x(u) = y$. From this it follows the inequality for every $x, y \in K$

$$d^\varepsilon(x, y) \leq c^{-\frac{1}{k}} |x - y|^{\frac{1}{k}}. \tag{10.82}$$

□

We are now ready to prove the main result of this section.

**Theorem 10.65.** *$d^\varepsilon \to \widehat{d}$ uniformly on compacts sets in $\mathbb{R}^n \times \mathbb{R}^n$.*

*Proof.* By Proposition 10.64 it is sufficient to prove the pointwise convergence

$$\lim_{\varepsilon \to 0^+} d^\varepsilon(x,y) = \widehat{d}(x,y). \tag{10.83}$$

for every fixed $x, y \in \mathbb{R}^n$. But (10.83) is a consequence of Theorem 3.56 and the fact that the vector fields $f_i^\varepsilon$ converge to $\widehat{f}_i$ thanks to Lemma 10.58. □

Combining Proposition 10.64 and Theorem 10.65 we obtain the following corollary.

**Corollary 10.66.** *For every compact $K \subset \mathbb{R}^n$ there exists $C > 0$, depending on $K$, such that*

$$\widehat{d}(x,y) \leq C|x-y|^{1/k}, \qquad \forall\, x, y \in K, \tag{10.84}$$

*where $k$ is the degree of nonholonomy of the sub-Riemannian structure.*

The uniform convergence given in Theorem 10.65 permits us to prove an important quantitative estimate on the shape of sub-Riemannian balls. Let us introduce the *box* $\mathrm{Box}(\varepsilon)$ of size $\varepsilon > 0$ defined, in privileged coordinates $x = (x_1, \ldots, x_k) \in \mathbb{R}^{n_1} \oplus \ldots \oplus \mathbb{R}^{n_k} = \mathbb{R}^n$, as follows

$$\mathrm{Box}(\varepsilon) := \{x \in \mathbb{R}^n : |x_i| \leq \varepsilon^i, i = 1, \ldots, k\}. \tag{10.85}$$

**Theorem 10.67** (Ball-Box Theorem)**.** *There exists constants $\varepsilon_0 > 0$, and $c_1, c_2 > 0$ such that*

$$c_1 \mathrm{Box}(\varepsilon) \subset B(x,\varepsilon) \subset c_2 \mathrm{Box}(\varepsilon), \qquad \forall\, \varepsilon \leq \varepsilon_0$$

*where $B(x,\varepsilon)$ is the sub-Riemannian ball in privileged coordinates.*

Notice that this statement is weaker with respect to Theorem 10.65.

*Proof.* We work in privileged coordinates $(x_1, \ldots, x_k) \in \mathbb{R}^{n_1} \oplus \ldots \oplus \mathbb{R}^{n_k} = \mathbb{R}^n$ where the base point is identified with the origin. Consider the unit ball $\widehat{B}(0,1)$ for the nilpotent approximation and fix two constants $c_1, c_2 > 0$ such that there exists a cube $[-c_1, c_1]^n \subset \widehat{B}(0,1) \subset [-c_2, c_2]^n$. Thanks to Theorem 10.65 there exists $\varepsilon_0 > 0$ such that for all $\varepsilon \leq \varepsilon_0$ we have

$$[-c_1, c_1]^n \subset B^\varepsilon(0,1) \subset [-c_2, c_2]^n,$$

where $B^\varepsilon(0,1)$ is the unit ball defined by the metric $d^\varepsilon$. Applying the dilation $\delta_\varepsilon$ to all sets we get that

$$\delta_\varepsilon[-c_1, c_1]^n \subset \delta_\varepsilon B^\varepsilon(0,1) \subset \delta_\varepsilon[-c_2, c_2]^n$$

but for $c > 0$ we have that $\delta_\varepsilon[-c, c]^n = c\mathrm{Box}(\varepsilon)$. Moreover by definition of $d^\varepsilon$ we have that $\delta_\varepsilon(B^\varepsilon(0,1)) = B(0,\varepsilon)$ (cf. also Corollary 10.61). □

## 10.6 Algebraic meaning

In this last section we discuss the algebraic structure induced on the nonholonomic tangent space and in particular how one can recover it in purely algebraic terms from the data of the vector fields.

Recall that given a generating family $\{f_1, \ldots, f_m\}$ for the sub-Riemannian structure and a point $q \in M$, there are well defined vector field $\{\widehat{f}_1, \ldots, \widehat{f}_m\}$ on the nilpotent tangent space $T_q^f M$.

The Lie algebra of vector fields $L := \mathrm{Lie}\{\widehat{f}_1, \dots, \widehat{f}_m\}$ is finite-dimensional and nilpotent thanks to Proposition 10.59. Denote by $\mathcal{G}$ the Lie group of associated flows (cf. Section 7.1)

$$\mathcal{G} = \{e^{t_1 \widehat{f}_{i_1}} \circ \dots \circ e^{t_j \widehat{f}_{i_j}} : t_i \in \mathbb{R}, j \in \mathbb{N}\}. \tag{10.86}$$

endowed with the product $\circ$. Notice that by construction $\mathcal{G}$ is connected and simply connected, whose Lie algebra $L$ nilpotent and stratified.

**Proposition 10.68.** *$\mathcal{G}$ is a Carnot group and* $\mathrm{Lie}(\mathcal{G}) = L$.

The group $\mathcal{G}$ naturally acts on $T_q^f M = J_q^k M/\sim$. Denote by $[j] \in J_q^k M/\sim$ the equivalence class of a jet $j = J_q^k \gamma \in J_q^k M$. The action of an generator of $\mathcal{G}$ on $T_q^f M$ is defined follows

$$e^{t\widehat{f}_i} \cdot [j] := [\gamma \circ e^{t\widehat{f}_i}], \qquad j = J_q^k \gamma \in J_q^k M. \tag{10.87}$$

Notice that this is a right action. Let us denote by $\mathcal{G}_0$ the isotropy sub-group of the trivial element of $T_q^f M$ under the action of $\mathcal{G}$.

Collecting the results proved in Section 10.4, and in particular Theorem 10.33, we have the following result

**Theorem 10.69.** *The nilpotent approximation $T_q^f M$ has the structure of a smooth manifold of dimension $\dim T_q^f M = \dim M$, diffeomorphic to the homogeneous space $\mathcal{G}/\mathcal{G}_0$ of the Carnot group $\mathcal{G}$ defined in* (10.86).

*Remark* 10.70. The diffeomorphism given by Theorem 10.69 was built explicitly thanks to privileged coordinates in in Section 10.4.

Notice that this could also be seen as a consequence of the theory of Lie groups. Indeed it is not difficult to see that in the proof of Theorem 10.33 we proved that the action of the Lie group $\mathcal{G}$ on $T_q^f M$ is transitive, hence $T_q^f M$ is diffeomorphic to the quotient of $\mathcal{G}$ with the isotropy group of the identity, that is $\mathcal{G}_0$. See for instance [Lee13].

Next we give a purely algebraic interpretation of this construction at the level of Lie algebras. Let us first recall some definitions.

**Definition 10.71.** The *free associative algebra* $\mathcal{A}_m$ (or $\mathcal{A}(x_1, \dots, x_m)$) generated by $x_1, \dots, x_m$ is the associative algebra of linear combinations of words of its generators, where the product of two element is defined by juxtaposition.

The *free Lie algebra* $\mathrm{Lie}_m$ or $\mathrm{Lie}\{x_1, \dots, x_m\}$ is the algebra of elements of $A_m$, where the product of two elements $x_i, x_j$ is defined by the commutator $[x_i, x_j] = x_i x_j - x_j x_i$.

The *free nilpotent Lie algebra of step $k$* on $m$ generators, denoted $\mathrm{Lie}_m^k$ or $\mathrm{Lie}^k\{x_1, \dots, x_m\}$, is the quotient $\mathrm{Lie}_m^k = \mathrm{Lie}_m/\mathcal{I}^{k+1}$ of the free Lie algebra $\mathrm{Lie}_m$ by the ideal $\mathcal{I}^{k+1}$ defined through the iterative formula

$$\mathcal{I}^1 = \mathrm{Lie}_m, \qquad \mathcal{I}^j = [\mathcal{I}^{j-1}, \mathrm{Lie}_m], \quad j > 1.$$

Let $\mathrm{Lie}_k\{x_1, \dots, x_m\}$ be the free Lie algebra nilpotent of step $k$ generated by the elements $x_1, \dots, x_m$. Notice that with every element $\pi \in \mathrm{Lie}_k\{x_1, \dots, x_m\}$ we can associate a vector field $\pi(X_1, \dots, X_m)$ (defined on $\mathbb{R}^n$) by replacing generators with vector fields $X_1, \dots, X_m$ .

**Definition 10.72.** Given a sub-Riemannian structure defined by the generating family $\{f_1, \dots, f_m\}$ that is bracket generating of step $k$ at a point $q$, we define the *core algebra*

$$C_q := \{\pi \in \mathrm{Lie}_k\{X_1, \dots, X_m\} \,|\, \pi(f_1, \dots, f_m)(q) \in \mathcal{D}_q^{\deg \pi - 1}\}. \tag{10.88}$$

**Exercise 10.73.** (i). Prove that $C_q$ is a subalgebra. (ii). Consider the subset

$$N_q := \{\pi \in \mathrm{Lie}_k\{X_1, \dots, X_m\} \,|\, \pi(f_1, \dots, f_m)(x) \in \mathcal{D}_x^{\deg \pi - 1}, \forall\, x \in O_q\}.$$

Prove that $N_q$ is an ideal contained in $C_q$.

Denote by $\mathcal{G}_m^k$ the connected and simply connected Lie group generated by the free nilpotent Lie algebra $\mathrm{Lie}_m^k$ and $\exp : \mathrm{Lie}_m^k \to \mathcal{G}_m^k$ its exponential map. Let $\mathcal{C}_q = \exp(C_q)$.

**Theorem 10.74.** *There exists a canonical isomorphism*

$$\phi : \mathcal{G}_m^k / \mathcal{C}_q \to T_q^f M.$$

*Its differential $\phi_*$ sends the generators $X_1, \dots, X_m$ to $\widehat{f}_1, \dots, \widehat{f}_m$.*

*Remark* 10.75. The core algebra can be rewritten in privileged coordinates in terms of the nilpotent approximation $\{\widehat{f}_1, \dots, \widehat{f}_m\}$ of the generators as follows:

$$C_q := \{\pi \in \mathrm{Lie}_k\{X_1, \dots, X_k\} \,|\, \pi(\widehat{f}_1, \dots, \widehat{f}_m)(0) = 0\}$$

**Exercise 10.76** (Grushin plane)**.** Let us analyze this algebraic construction in the case of the simplest non-holonomic tangent space arising as the tangent space to a non-regular structure in $\mathbb{R}^2$: the Grushin plane described in the Example 10.57.

We have shown that the nonholonomic tangent space has the following normal form

$$\widehat{f}_1 = \partial_{x_1}, \qquad \widehat{f}_2 = x_1 \partial_{x_2}. \tag{10.89}$$

In these coordinates indeed the two vector fields have weight one and are homogeneous with respect to the weights $\nu(x_1) = 1$ and $\nu(x_2) = 2$. In this case $m = k = 2$.

Since $[\widehat{f}_1, \widehat{f}_2] =: \widehat{f}_3 = \partial_{x_2}$ it is easy to see that

$$\mathrm{Lie}\{\widehat{f}_1, \widehat{f}_2\} = \mathrm{span}\{\widehat{f}_1, \widehat{f}_2, \widehat{f}_3\} \tag{10.90}$$

On the other hand the core algebra at the origin $C_0$ contains $\widehat{f}_2$ since it has weight one, but it vanishes at zero (it does not belong to $\mathcal{D}_0^1$). Hence $C_0 = \mathrm{span}\{\widehat{f}_2\}$.

### 10.6.1 Nonholonomic tangent space: the equiregular case

The last two statements concerns the case of an equiregular distribution. In this case one can show that the subgroup $\mathcal{G}_0$ of $\mathcal{G}$ is trivial.

**Proposition 10.77.** *Assume that the sub-Riemannian structure is equiregular, i.e., for every $i \geq 1$ the integer $d_i(q) = \dim \mathcal{D}_q^i$ does not depend on $q$. Then $C_q$ is an ideal. In particular $\mathcal{G}_0 = \{0\}$ and $T_q^f M$ is a Carnot group.*

*Proof.* To prove that the core subalgebra $C_q$ is an ideal, it is sufficient to prove that $X \in C_q$ implies $[f_i, X] \in C_q$ for every $i = 1, \dots, m$.

Thanks to the characterization (10.88), this is equivalent to prove the following claim: for every $X = \pi(f_1, \dots, f_m)$ bracket polynomial of degree $\deg \pi \le h$ such that $X(q) \in \mathcal{D}_q^{h-1}$, we have $[f_i, X](q) \in \mathcal{D}_q^h$ for every $i = 1, \dots, m$.

Since the structure has constant growth vector, we can consider a frame $V_1, \dots, V_n$ that is privileged at every point in neighborhood $O_q$ of $q$. In particular for every $x \in O_q$ we have

$$\mathcal{D}_x^i = \operatorname{span}\{V_1(x), \dots, V_{d_i}(x)\}. \tag{10.91}$$

Let $X = \pi(f_1, \dots, f_m)$ be a bracket polynomial of degree $\deg \pi \le h$. Then there exist smooth functions $a_j$ such that

$$X(x) = \sum_{j: w_j \le h} a_j(x) V_j(x), \qquad \forall\, x \in O_q. \tag{10.92}$$

Thanks to (10.91), $X(q) \in \mathcal{D}_q^{h-1}$ is equivalent to require that $a_j(q) = 0$ for every $j$ such that $w_j = h$. Let us compute

$$[f_i, X] = \left[ f_i, \sum_{w_j \le h} a_j V_j \right] = \sum_{w_j \le h} a_j [f_i, V_j] + f_i(a_j) V_j. \tag{10.93}$$

Evaluating (10.93) at the point $q$ and using that $a_j(q) = 0$ for every $j$ such that $w_j = h$, it follows that $[f_i, X](q) \in \mathcal{D}_q^h$ for every $i = 1, \dots, m$, that is our claim. $\square$

The next result explains how to find a generating family of the nilpotent sub-Riemannian structure on the Carnot group, once given a generating family of the original structure.

**Corollary 10.78.** *Assume that the sub-Riemannian structure is equiregular and $\{f_1, \dots, f_m\}$ is a generating family. Then $\widehat{f}_1, \dots, \widehat{f}_m$ are a generating family of left-invariant vector fields on $T_q^f M$.*

*Proof.* This is a consequence of the following two general facts: (i). Given a *right* action of a Lie group on a homogeneous space $G/H$, then a *left*-invariant vector field on $X$ induces a well-defined vector field $\pi_* X$ on $G/H$ through the projection $\pi : G \to G/H$. (ii). If the Lie subgroup $H$ is normal and $G/H$ is a Lie group, then $\pi_* X$ is also left-invariant. $\square$

**Exercise 10.79.** Prove the two statements contained in the proof of Corollary 10.78.

## 10.7 Carnot groups: normal forms in low dimension

In this section we provide normal forms for Carnot groups in dimension smaller or equal than 5. Recall that Carnot groups arise as nonholonomic tangent spaces to equiregular sub-Riemannian structures.

For an equiregular sub-Riemannian structure the integer $d_i = \dim \mathcal{D}_q^i$ are independent on $q$. Denote by $k$ the step of the sub-Riemannian structure, in particular $d_k = \dim M$. The sequence of integers $(d_1, \dots, d_k)$ is called *growth vector* of the sub-Riemannian structure.

**Exercise 10.80.** Assume that the structure is equiregular of step $k$. Prove that the sequence $(d_1, \dots, d_k)$ is strictly increasing. (*Hint*: prove that if $d_i = d_{i+1}$ for some $i < k$, then $d_i = d_k = \dim M$, contradicting the minimality of $k$.)

From Exercice 10.80 it easily follows that the possibilities for the growth vector in dimension smaller or equal than 5 are listed as follows:

- $(2,3)$, if $\dim(M) = 3$,
- $(2,3,4)$ and $(3,4)$, if $\dim(M) = 4$,
- $(2,3,4,5), (2,3,5), (3,4,5), (3,5)$ and $(4,5)$, if $\dim(M) = 5$.

The following theorem gives normal forms for Carnot groups of given growth vector in the previous list. In every case but the last one, the normal form is unique.

**Theorem 10.81.** *Let $(M, \mathbf{U}, f)$ be an equiregular sub-Riemannian manifold, with $\dim M \leq 5$. Its nonholonomic tangent space at a point is isomorphic to one of the following sub-Riemannian structures:*

- (Heisenberg). *If the growth vector is $(2,3)$, then the orthonormal frame can be chosen as*

$$\begin{aligned} f_1 &= \partial_{x_1}, \\ f_2 &= \partial_{x_2} + x_1 \partial_{x_3}. \end{aligned}$$

- (Engel). *If the growth vector is $(2,3,4)$, then the orthonormal frame can be chosen as*

$$\begin{aligned} f_1 &= \partial_{x_1}, \\ f_2 &= \partial_{x_2} + x_1 \partial_{x_3} + x_1 x_2 \partial_{x_4}. \end{aligned}$$

- (Quasi-Heisenberg). *If the growth vector is $(3,4)$, then the orthonormal frame can be chosen as*

$$\begin{aligned} f_1 &= \partial_{x_1}, \\ f_2 &= \partial_{x_2} + x_1 \partial_{x_4}, \\ f_3 &= \partial_{x_3}. \end{aligned}$$

- (Cartan rank 2). *If the growth vector is $(2,3,5)$, then the orthonormal frame can be chosen as*

$$\begin{aligned} f_1 &= \partial_{x_1}, \\ f_2 &= \partial_{x_2} + x_1 \partial_{x_3} + \frac{1}{2} x_1^2 \partial_{x_4} + x_1 x_2 \partial_{x_5}. \end{aligned}$$

- (Goursat rank 2). *If the growth vector is $(2,3,4,5)$, then the orthonormal frame can be chosen as*

$$\begin{aligned} f_1 &= \partial_{x_1}, \\ f_2 &= \partial_{x_2} + x_1 \partial_{x_3} + \frac{1}{2} x_1^2 \partial_{x_4} + \frac{1}{6} x_1^3 \partial_{x_5}. \end{aligned}$$

- (Cartan rank 3). *If the growth vector is* $(3,5)$, *then the orthonormal frame can be chosen as*

$$\begin{aligned}
f_1 &= \partial_{x_1} - \frac{1}{2}x_2\partial_{x_4},\\
f_2 &= \partial_{x_2} + \frac{1}{2}x_1\partial_{x_4} - \frac{1}{2}x_3\partial_{x_5},\\
f_3 &= \partial_{x_3} + \frac{1}{2}x_2\partial_{x_5}.
\end{aligned}$$

- (Goursat rank 3). *If the growth vector is* $(3,4,5)$, *then the orthonormal frame can be chosen as*

$$\begin{aligned}
f_1 &= \partial_{x_1} - \frac{1}{2}x_2\partial_{x_4} - \frac{1}{3}x_1x_2\partial_{x_5},\\
f_2 &= \partial_{x_2} + \frac{1}{2}x_1\partial_{x_4} + \frac{1}{3}x_1^2\partial_{x_5},\\
f_3 &= \partial_{x_3}.
\end{aligned}$$

- (Bi-Heisenberg). *If the growth vector is* $(4,5)$, *then there exists* $\alpha \in \mathbb{R}$ *such that the orthonormal frame can be chosen as*

$$\begin{aligned}
f_1 &= \partial_{x_1} - \frac{1}{2}x_2\partial_{x_5},\\
f_2 &= \partial_{x_2} + \frac{1}{2}x_1\partial_{x_5},\\
f_3 &= \partial_{x_3} - \frac{\alpha}{2}x_4\partial_{x_5},\\
f_4 &= \partial_{x_4} + \frac{\alpha}{2}x_3\partial_{x_5}.
\end{aligned}$$

*Proof.* Recall that given a basis $X_1,\dots,X_m$ of a Lie algebra $\mathfrak{g}$, the coefficients $c_{ij}^{\ell}$ satisfying $[X_i,X_j]=\sum_\ell c_{ij}^{\ell}X_\ell$ are called structural constant of $\mathfrak{g}$. To prove the theorem we will show that, for every choice of the growth vector, we can choose an orthonormal basis of the Lie algebra such that the structural constants are uniquely determined by the sub-Riemannian structure.

We give a sketch of the proof for the $(3,4,5)$, $(2,3,4,5)$ and $(4,5)$ cases. The other cases can be treated in a similar way. Since we deal with sub-Riemannian structures $(M,\mathbf{U},f)$ that are left-invariant on a nilpotent Lie group, we can identify the distribution $\mathcal{D}$ with its value at the identity of the group $\mathcal{D}_0$.

(a). Growth vector equal to $(3,4,5)$. Let $(M,\mathbf{U},f)$ be a nilpotent $(3,4,5)$ sub-Riemannian structure. Let $\{X_1,X_2,X_3\}$ be a basis for $\mathcal{D}_0$, as a vector subspace of the Lie algebra. By our assumption on the growth vector we know that

$$\dim \operatorname{span}\{[X_1,X_2],[X_1,X_3],[X_2,X_3]\}/\mathcal{D}_0 = 1. \tag{10.94}$$

In other words, we can define the skew-symmetric bilinear map

$$\Phi(\cdot,\cdot):\mathcal{D}_0\times\mathcal{D}_0\to T_0G/\mathcal{D}_0, \qquad \Phi(v,w)=[V,W](0) \mod \mathcal{D}_0 \tag{10.95}$$

where $V,W$ are smooth vector fields such that $V(0)=v$ and $W(0)=w$. The condition (10.94) implies that there exists a one dimensional subspace in the kernel of this map, namely a non-zero vector $v$ such that $\Phi(v,\cdot) = 0$. Let $f_3$ be a vector in $\ker \Phi \cap \mathcal{D}_0$ with norm one, and consider its orthogonal subspace $f_3^\perp \subset \mathcal{D}_0$ with respect to the inner product on the distribution $\mathcal{D}_0$. For every positively oriented orthonormal basis $\{X_1, X_2\}$ on $f_3^\perp$ it is easy to see that $f_4 := [X_1, X_2]$ is well defined, i.e., it does not depend on rotation of $X_1, X_2$ within $f_3^\perp$. Then, reasoning as in the proof of Example 10.56, we can choose a rotation of the original orthonormal frame, denoted $\{f_1, f_2\}$, such that $[f_2, f_4] = 0$. Defining $f_5 := [f_1, f_4]$, this gives a choice of a canonical basis $\{f_1, \dots, f_5\}$ for the Lie algebra where the only non trivial commutator relations are the following

$$[f_1, f_2] = f_4, \qquad\qquad [f_1, f_4] = f_5.$$

(b). Growth vector equal to $(2, 3, 4, 5)$. Let $(M, \mathbf{U}, f)$ be a nilpotent $(2, 3, 4, 5)$ sub-Riemannian structure. Consider any orthonormal basis $\{X_1, X_2\}$ for the two dimensional subspace $\mathcal{D}_0$. By our assumption on the growth vector we have that

$$\begin{gathered}\dim \operatorname{span}\{X_1, X_2, [X_1, X_2]\} = 3 \\ \dim \operatorname{span}\{X_1, X_2, [X_1, X_2], [X_1, [X_1, X_2]], [X_2, [X_1, X_2]]\} = 4.\end{gathered} \tag{10.96}$$

As in part (a) of the proof, it is easy to see that there exists a suitable rotation of $\{X_1, X_2\}$ on $\mathcal{D}_0$, which we denote $\{f_1, f_2\}$, such that $[f_2, [f_1, f_2]] = 0$. Using the Jacobi identity we get

$$[f_2, [f_1, [f_1, f_2]]] = -[f_1, [f_2, [f_1, f_2]] - [[f_1, f_2], [f_1, f_2]] = 0.$$

Then we set $f_3 := [f_1, f_2]$, $f_4 := [f_1, [f_1, f_2]]$ and $f_5 := [f_1, [f_1, [f_1, f_2]]]$. Relations (10.96) imply that these vectors are linearly independent. Hence we have a canonical basis for the Lie algebra, where the only nontrivial commutator relations are the folllowing:

$$[f_1, f_2] = f_3, \qquad [f_1, f_3] = f_4, \qquad [f_1, f_4] = f_5.$$

(c). Growth vector equal to $(4, 5)$. In this case let us consider again the map

$$\Phi(\cdot,\cdot) : \mathcal{D}_0 \times \mathcal{D}_0 \to T_0G/\mathcal{D}_0, \qquad \Phi(v, w) = [V, W](0) \mod \mathcal{D}_0 \tag{10.97}$$

where $V, W$ are smooth vector fields such that $V(0) = v$ and $W(0) = w$. Since $\dim T_0G/\mathcal{D}_0 = 1$, the map (10.97) is represented by a single $4 \times 4$ skew-simmetric matrix $L$. By skew-symmetricity its eigenvalues are purely imaginary $\pm i\alpha_1, \pm i\alpha_2$, one of which is different from zero since the structure is bracket generating. Up to relabelling indices, we can assume that $\alpha_1 \neq 0$. Then choose $f_1, f_2, f_3, f_4$ be a basis that puts the matrix $L$ in the normal form for skew-symmetric matrices

$$L = \begin{pmatrix} 0 & \alpha_1 & & \\ -\alpha_1 & 0 & & \\ & & 0 & \alpha_2 \\ & & -\alpha_2 & 0 \end{pmatrix}$$

Defining $f_5 := [f_1, f_2]$ and setting $\alpha := \alpha_2/\alpha_1$ we have that $[f_3, f_4] = \alpha f_5$. □

*Remark* 10.82. In the proof of Theorem 10.81 we showed that the structure of Lie brackets is uniquely determined by the choice of a suitable orthonormal frame (in the last example it is unique modulo a real parameter $\alpha$).

Of course the coordinate representation of the vector fields satisfying these structural equations is not unique (compare for instance the vector fields in the case of the Heisenberg group given here with respect to those used in the previous chapters). Nevertheless, all of them can be obtained from the one described here with a change of variable. We refer the reader to the Nagano principle [Nag66] for more details.

**Exercise 10.83.** Prove that in the three examples described in Section 10.4.4 there is a unique normal form for the generating family, even if the distribution is endowed with an inner product.

## 10.8 Bibliographical note

The nonholonomic tangent space appeared in the literature in different contexts, and with different names.

The first appearence is in the 70s, in relation to the study of hypoelliptic Hörmander operators. In this context, the main ideas behind the notion of nilpotent approximation and privileged coordinates were used, in the regular case, to approximate the differential operators by homogeneous invariant operators on nilpotent groups [FS74, Fol75]. A more general case has been treated by Rothschild and Stein [RS76]; their techniques involves "lifting" the differential operators to a (regular) higher-dimensional manifold and then approximating the lifted operators by operators on a group. These techniques have been also developed and refined by Goodman and Métivier in [Goo76, Ḿ76].

Some years later in [NSW85] the authors proved a first version of the Ball-Box theorem for metrics defined by vector fields.

A more geometric language have been developed later. A general notion of "metric tangent cone" has been introduced in the work of Gromov [Gro81] and then Mitchell in [Mit85] considered the sub-Riemannian case, stating that this tangent cone is, for equiregular structures, what in this Chapter is called a Carnot group. The arguments of Mitchell deeply relies on the previous works [Goo76, Ḿ76]. A definition of "tangent cone" for sub-Riemannian manifolds is also given in [MM95], and then refined in [MM00], to let the tangent cone at a point to be unique, up to isomorphism. A survey on this can be found on the celebrated paper of Bellaïche [Bel96].

Starting from the pioneering work of Pansu [Pan89], Lie groups equipped with a certain explicit left-invariant sub-Finsler metrics also appear in geometric group theory, as asymptotic cones of nilpotent finitely generated groups.

"Nilpotent approximation" played also a prominent role in Control Theory. In this context, a proof of existence of privileged coordinates is given in [AGS89, AS87, Bel96, BS90]. The book of F. Jean [Jea14] provides a modern approach to nonholonomic tangent space in the language of Control Theory and shows some of its interesting applications to motion planning.

The construction of nonholonomic tangent space presented here is inspired by the ideas developed in the paper [AM03]. It is more intrinsic than previous approaches: the nilpotent approximation of the distribution at a point depends only on the point and the distribution itself, it does not depend on the metric or local coordinates. We also recover the intrinsic meaning of the dilation: it is induced by the parameter rescaling of smooth curves.

# Chapter 11

# Regularity of the sub-Riemannian distance

In this chapter we investigate on the regularity properties of the sub-Riemannian distance from a fixed point. In particular, besides the Hölder continuity already discussed in Chapter 10, we prove that the sub-Riemannian distance is smooth on an open and dense subset of every compact ball.

On the other hand if one considers the squared sub-Riemannian distance from a fixed point, this function is never smooth in a neighborhood of the base point unless the structure is Riemannian. Moreover, as soon as the distribution is not full-dimensional at the point, every level set of the distance contains a non-differentiability point of the distance itself. This is an obstruction for sub-Riemannian spheres to be smooth hypersurfaces.

In absence of abnormal minimizers, one can show that the sub-Riemannian distance is locally Lipschitz outside the diagonal and, using a non-smooth version of the Sard Lemma, we show that almost every sphere is a Lipschitz submanifold.

## 11.1 Regularity of the sub-Riemannian squared distance

Let us consider a free sub-Riemannian structure $(M, \mathbf{U}, f)$ with generating family $f_1, \ldots, f_m$ and fix a point $q \in M$. Recall that the flag of the sub-Riemannian structure at the point $q$ is the sequence of increasing subspaces $\{\mathcal{D}^i_q\}_{i\in\mathbb{N}}$ defined by

$$\mathcal{D}^i_q := \operatorname{span}\{[f_{j_1}, \ldots, [f_{j_{l-1}}, f_{j_l}]](q),\ \forall\, l \leq i\}, \tag{11.1}$$

The step of the sub-Riemannian structure at $q$ is the minimal integer $k(q)$ such that $\mathcal{D}^{k(q)}_q = T_qM$.

In Chapter 10 we already proved that the sub-Riemannian distance is Hölder continuous, with Hölder exponent that is related to the step of the sub-Riemannian structure at the point. For the reader's convenience, we recall here the statement.

**Proposition 11.1.** *Let $M$ be a sub-Riemannian structure, fix $q \in M$ and denote by $k = k(q)$ is the step of the sub-Riemannian structure at $q$. There exists a neighborhood $O_q$ such that for every coordinate map $\phi : O_q \to \mathbb{R}^n$ there exists $C > 0$ such that for $q_0, q_1 \in O_q$ one has*

$$d(q_0, q_1) \leq C|\phi(q_0) - \phi(q_1)|^{1/k}.$$

We investigate next the differentiability properties of the distance function. More precisely we want to charachterize the set where the function $d$ is $C^\infty$.

Throughout this section, we fix a base point $q_0 \in M$ and we denote by $\mathsf{f}$ the squared sub-Riemannian distance from $q_0$

$$\mathsf{f} : M \to \mathbb{R}, \qquad \mathsf{f}(q) = \frac{1}{2} d^2(q_0, q). \tag{11.2}$$

The main result of this chapter is the following.

**Theorem 11.2.** *Assume that the closed ball $B := \overline{B}_{q_0}(r_0)$ is compact. Then the function $\mathsf{f}\big|_B : B \to \mathbb{R}$ is smooth on a open dense subset of $B$.*

For a complete sub-Riemannian structure all closed balls are compact (cf. Proposition 3.47). Hence we immediately obtain the following corollary.

**Corollary 11.3.** *Assume that $M$ is a complete sub-Riemannian manifold. Then $\mathsf{f}$ is smooth on an open and dense subset of $M$.*

Thanks to the existence theorem (Corollary 8.64), for each $q \in B$ there exists a length-minimizer joining $q_0$ and $q$. We start by looking for necessary conditions for $\mathsf{f}$ to be $C^\infty$ in a neighborhood of the point $q$.

**Proposition 11.4.** *Let $q \in B$ and assume that $\mathsf{f}$ is $C^\infty$ in a neighborhood of $q$. Then*

*(i) there exists a unique length minimizer $\gamma : [0, 1] \to M$ joining $q_0$ with $q$. Moreover $\gamma$ is not abnormal and $q$ is not conjugate to $q_0$ along $\gamma$.*

*(ii) $d_q \mathsf{f} = \lambda(1)$, where $\lambda : [0, 1] \to T^*M$ is the unique normal lift of $\gamma$.*

*Proof.* Denote by $E_{q_0}$ the end-point map based at $q_0$ and by $J$ the energy functional. Let us introduce the functional

$$\Psi : L^\infty([0, T], \mathbb{R}^m) \to \mathbb{R}, \qquad \Psi(v) = J(v) - \mathsf{f}(E_{q_0}(v)), \tag{11.3}$$

By the smoothness assumption on $\mathsf{f}$, the map $\Psi$ is smooth in a neighborhood of every control associated with a length-minimizer joining $q_0$ and $q$. Moreover $\Psi$ is non negative by construction.

Let $\gamma : [0, 1] \to M$ be any optimal trajectory associated with an optimal control $u$, joining $q_0$ and $q$. Then we have

$$0 = d_u \Psi = d_u J - d_q \mathsf{f} \circ D_u E_{q_0}. \tag{11.4}$$

Thus, $\gamma$ is a normal extremal trajectory, with Lagrange multiplier $\lambda_1 = d_q \mathsf{f}$. By Theorem 4.25, we can recover $\gamma$ by the formula $\gamma(t) = \pi \circ e^{(t-1)\vec{H}}(\lambda_1)$. Then, $\gamma$ is the unique minimizer of $J$ connecting its endpoints, and is normal. Morever $\lambda_1$ is also the final point of its normal lift, by construction.

Next we show that $\gamma$ is not abnormal and not conjugate. For $y$ in a neighborhood $O_q$ of $q$, let us consider the map

$$\Phi : O_q \mapsto T^*_{q_0} M, \qquad \Phi(y) = e^{-\vec{H}}(d_y \mathsf{f}). \tag{11.5}$$

Thanks to the smoothness assumption on $\mathsf{f}$, the map $\Phi$ is $C^\infty$. Moreover $\Phi$ is a right inverse for the exponential map, since for every $y \in O_q$ one has

$$\exp_{q_0}(\Phi(y)) = \pi \circ e^{\vec{H}}(e^{-\vec{H}}(d_y \mathsf{f})) = \pi(d_y \mathsf{f}) = y. \tag{11.6}$$

This implies that $q$ is a regular value for the exponential map and, a fortiori, $u$ is a regular point for the end-point map. This proves that $u$ corresponds to a trajectory that is at the same time strictly normal and not conjugate. $\square$

*Remark* 11.5. The conclusion of Proposition 11.4 holds, with the same proof, even if the function $\mathsf{f}$ is only of class $C^2$ in a neighborhood of $q$. Indeed in this case the map $\Phi$ introduced in the proof is of class $C^1$ and this is sufficient to conclude.

When $\mathsf{f}$ is only differentiable at $q$, one can still repeat the first part of the argument, proving that there exists a unique minimizer $\gamma : [0,1] \to M$ joining $q_0$ to $q$, that admits at least a normal lift.

Before going further in the study of the smoothness properties of the distance function, we are already able to prove an important corollary of this result.

For every $r > 0$, denote by $S_r := \mathsf{f}^{-1}(\frac{r^2}{2})$ the sub-Riemannian sphere of radius $r$ centered at $q_0$.

**Corollary 11.6.** *Assume that the closed ball $B := \overline{B}_{q_0}(r_0)$ is compact and that $\mathcal{D}_{q_0} \neq T_{q_0}M$. For every $r \leq r_0$, the sphere $S_r$ contains a point where the function $\mathsf{f}$ is not smooth.*

*Proof.* Since $r \leq r_0$, the sphere $S_r$ is non empty and contained in the compact ball $B$. Assume, by contradiction, that $\mathsf{f}$ is smooth at every point of $S_r$. Then $S_r$ is a level set defined by $\mathsf{f}$ and $d_q\mathsf{f} \neq 0$ for every $q \in S_r$ (indeed $d_q\mathsf{f}$ is the covector attached at the final point of a normal Pontryagin extremal, hence it is non vanishing, cf. Proposition 11.4). This implies that $S_r$ is a smooth submanifold of dimension $n-1$, without boundary. Moreover, being the level set of a continuous function, $S_r$ is closed, hence compact. Let us consider the map

$$\Phi : S_r \to T^*_{q_0}M, \qquad \Phi(q) = e^{-\vec{H}}(d_q\mathsf{f}), \tag{11.7}$$

By assumption $\mathsf{f}$ is smooth, hence $\Phi$ is a smooth right inverse of the exponential map (see also (11.6)). In particular the differential of $\Phi$ is injective at every point. Moreover by construction $H(\Phi(q)) = r^2/2$, for every $q \in S_r$ (cf. Theorem 4.25). It follows that $\Phi$ defines a smooth immersion

$$\Phi : S_r \to H^{-1}(r^2/2) \cap T^*_{q_0}M, \tag{11.8}$$

of the sub-Riemannian sphere $S_r$ into the set

$$C_r := H^{-1}(r^2/2) \cap T^*_{q_0}M = \left\{ \lambda \in T^*_{q_0}M \mid \frac{1}{2}\sum_{i=1}^{m} \langle \lambda, f_i(q_0)\rangle^2 = \frac{r^2}{2} \right\}. \tag{11.9}$$

where $f_1, \ldots, f_m$ is a generating family. Since we work at a fixed point, it is not restrictive to assume that $m = \dim \mathcal{D}_{q_0}$ is the rank of the structure at the point $q_0$. Notice that $C_r$ is a smooth connected $n-1$ dimensional submanifold of the fiber $T^*_{q_0}M$, indeed diffeomorphic to $S^{m-1} \times \mathbb{R}^{n-m}$. Since by assumption $m < n$, the manifold $C_r$ is not compact.

By continuity of $\Phi$, the image $\Phi(S_r)$ is closed in $C_r$. Moreover, since every immersion is a local submersion and $\dim S_r = \dim C_r$, the set $\Phi(S_r)$ is also open in $C_r$. Hence it is connected. Since $\Phi(S_r)$ has no boundary, it is a connected component of $C_r$, which proves $\Phi(S_r) = C_r$. This gives a contradiction since, by continuity, $\Phi(S_r)$ is compact, while $C_r$ is not. $\square$

Next we move to the proof of the main result, namely Theorem 11.2. Following Proposition 11.4, we introduce the following set.

**Definition 11.7.** Let $q_0 \in M$. The set of *smooth point* from $q_0$ is the subset $\Sigma \subset M$ of points $q \in M$ such that there exists a unique lenght-minimizer $\gamma$ joining $q_0$ to $q$, which is strictly normal, and such that $q$ is not conjugate to $q_0$ along $\gamma$.

If $\mathsf{f}$ is $C^\infty$ in a neighborhood of a point $q$, then $q \in \Sigma$ by Proposition 11.4 (cf. also Remark 11.5). Indeed, the converse also holds.

**Theorem 11.8.** *The set $\Sigma$ is open and dense in $B$. Moreover, $\mathsf{f}$ is $C^\infty$ on $\Sigma$.*

*Proof.* We split the proof into three parts: (a) the set $\Sigma$ is open, (b) the function $\mathsf{f}$ is $C^\infty$ on $\Sigma$, (c) the set $\Sigma$ is dense in $B$.

(a). To prove that $\Sigma$ is open we have to show that for every $q \in \Sigma$ there exists a neighborhood $O_q$ of $q$ such that $O_q \subset \Sigma$.

Let us start by proving the following claim: there exists a neighborhood of $q$ in $B$ such that every point in this neighborhood is reached by exactly one length-minimizer.

By contradiction, if this is not true, there exists a sequence $q_n$ of points in $B$ converging to $q$ admitting (at least) two length-minimizers $\gamma_n$ and $\gamma_n'$ joining $q_0$ and $q_n$. Let us denote by $u_n$ and $v_n$ the corresponding minimizing controls.

By Proposition 8.66, the set of controls associated with length-minimizers whose endpoint is in the compact ball $B$ is compact in the $L^2$ strong topology. Then, up to extraction of a subsequence, there exists two controls $u, v$ such that $u_n \to u$ and $v_n \to v$. Moreover, $u$ and $v$ are both associated with length-minimizers joining $q_0$ with $q$. Since, by assumption, there is a unique length-minimizer $\gamma$ joining $q_0$ with $q$, this implies $u = v$.

By smoothness of the end point map both $D_{u_n}E_{q_0}$ and $D_{v_n}E_{q_0}$ converge to $D_uE_{q_0}$. Moreover $D_uE_{q_0}$ has full rank (recall that $u$ is strictly normal, hence is not a critical point for $E_{q_0}$). This implies that, for $n$ big enough, both $D_{u_n}E_{q_0}$ and $D_{v_n}E_{q_0}$ are surjective and the corresponding controls $u_n$ and $v_n$ are strictly normal. Thus we can build the sequences $\lambda_1^n$ and $\xi_1^n$ of corresponding final covectors in $T^*_{q_n}M$ satisfying the identities

$$\lambda_1^n D_{u_n}E_{q_0} = u_n, \qquad \xi_1^n D_{v_n}E_{q_0} = v_n. \tag{11.10}$$

These relations can be rewritten in terms of the adjoint linear maps as follows

$$(D_{u_n}E_{q_0})^*\lambda_1^n = u_n, \qquad (D_{v_n}E_{q_0})^*\xi_1^n = v_n.$$

Since both $(D_{u_n}E_{q_0})^*$ and $(D_{v_n}E_{q_0})^*$ are a family of injective linear maps converging to $(D_uE_{q_0})^*$, with $u_n$ and $v_n$ converging to $u$, it follows that the corresponding (unique) solutions of the linear systems $\lambda_1^n$ and $\xi_1^n$ also converge to the solution of the limit problem $(D_uE_{q_0})^*\lambda_1 = u$, i.e, both sequences $\lambda_1^n$ and $\xi_1^n$ converge to the final covector $\lambda_1$ corresponding to $\gamma$. Composing with the flow defined by the corresponding controls we can deduce the convergence of the sequences $\lambda_0^n$ and $\xi_0^n$ of the initial covectors associated to $u_n$ and $v_n$ to the unique initial covector $\lambda_0$ corresponding to $\gamma$.

Finally, since $\lambda_0$ by assumption is a regular point of the exponential map, i.e., the unique minimizer $\gamma$ joining $q_0$ to $q$ is not conjugate, it follows that the exponential map is invertible in a neighborhood $V_{\lambda_0}$ of $\lambda_0$ onto its image $O_q := \exp(V_{\lambda_0})$, that is a neighborhood of $q$. This proves our initial claim.

In other words, we have proved that for every point $q' \in O_q$ there exists a unique minimizer joining $q_0$ to $q'$, whose initial covector $\lambda' \in V_\lambda$ is a regular point of the exponential map. This implies that every $q' \in O_q$ is a smooth point, and $\Sigma$ is open.

(b). Now we prove that $\mathsf{f}$ is smooth in a neighborhood of each point $q \in \Sigma$. From the part (a) of the proof it follows that if $q \in \Sigma$ there exists a neighborhood $V_{\lambda_0}$ of $\lambda_0$ and $O_q$ of $q$ such that $\exp|_{V_{\lambda_0}} : V_{\lambda_0} \to O_q$ is a smooth invertible map. Denote by $\Phi : O_q \to V_{\lambda_0}$ its smooth inverse. Since for every $q' \in O_q$ there is only one minimizer joining $q_0$ to $q'$ with initial covector $\Phi(q')$ it follows that,

$$\mathsf{f}(q') = \frac{1}{2} d^2(q_0, q') = H(\Phi(q')),$$

that is a composition of smooth functions, hence smooth.

(c). Our next goal is to show that $\Sigma$ is a dense set in $B$. We start by a preliminary definition.

**Definition 11.9.** A point $q \in B$ is said to be

(i) a *fair point* if there exists a unique minimizer joining $q_0$ to $q$, that is normal.

(ii) a *good point* if it is a fair point and the unique minimizer joining $q_0$ to $q$ is strictly normal.

We denote by $\Sigma_f$ and $\Sigma_g$ the set of fair and good points, respectively.

We stress that a fair point can be reached by a unique minimizer that is both normal and abnormal. From the definition it is immediate that $\Sigma \subset \Sigma_g \subset \Sigma_f$. The proof of (c) relies on the following four steps:

(c1) $\Sigma_f$ is a dense set in $B$,

(c2) $\Sigma_g$ is a dense set in $B$,

(c3) $\mathsf{f}$ is locally Lipschitz in a neighborhood of every point of $\Sigma_g$,

(c4) $\Sigma$ is a dense set in $B$.

(c1). Fix an open set $O \subset B$ and let us show that $\Sigma_f \cap O \neq \emptyset$. Consider a smooth function $a : O \to \mathbb{R}$ such that $a^{-1}([s, +\infty[)$ is compact for every $s \in \mathbb{R}$. Then consider the function

$$\psi : O \to \mathbb{R}, \qquad \psi(q) = \mathsf{f}(q) - a(q)$$

The function $\psi$ is continuous on $O$ and, since $\mathsf{f}$ is nonnegative, the set $\psi^{-1}(]-\infty, s[)$ are compact for every $s \in \mathbb{R}$ due to the assumption on $a$. It follows that $\psi$ attains its minimum at some point $q_1 \in O$. Let $u_1$ be a control associated with a length-minimizer $\gamma$ joining $q_0$ and $q_1 = E_{q_0}(u_1)$.

Since $J(u) \geq \mathsf{f}(E_{q_0}(u))$ for every control $u$, it is easy to see that the map

$$\Phi : L^\infty([0,1], \mathbb{R}^m) \to \mathbb{R}, \qquad \Phi(u) = J(u) - a(E_{q_0}(u))$$

attains a local minimum at $u_1$. In particular it holds

$$0 = D_{u_1}\Phi = u_1 - (d_{q_1}a) D_{u_1} E_{q_0}.$$

The last identity implies that $u_1$ is normal and $\lambda_1 = d_{q_1}a$ is the final covector associated with the trajectory. By Theorem 4.25, the corresponding trajectory $\gamma$ is uniquely recovered by the formula

$\gamma(t) = \pi \circ e^{(t-1)\vec{H}}(d_{q_1}a)$. In particular $\gamma$ is the unique minimizer joining $q_0$ to $q_1 \in O$, and is normal, namely $q_1 \in \Sigma_f \cap O$.

(c2). As in the proof of (c1), we shall prove that $\Sigma_g \cap O \neq \emptyset$ for any open $O \subset B$. By (c1) the set $\Sigma_f \cap O$ is nonempty. For any $q \in \Sigma_f \cap O$ we can define $\operatorname{rank} q := \operatorname{rank} D_u E_{q_0}$, where $u$ is the control associated to the unique minimizer $\gamma$ joining $q_0$ to $q$. To prove (c2) it is sufficient to prove that there exists a point $q' \in \Sigma_f \cap O$ such that $\operatorname{rank} q' = n$ (i.e., $D_{u'}E_{q_0}$ is surjective, where $u'$ is the control associated to the unique minimizer joining $q_0$ and $q'$). Assume by contradiction that

$$k_O := \max_{q \in \Sigma_f \cap O} \operatorname{rank} q < n,$$

and consider a point $\widehat{q}$ where the maximum is attained, i.e., such that $\operatorname{rank} \widehat{q} = k_O$.

We claim that all points of $\Sigma_f \cap O$ that are sufficiently close to $\widehat{q}$ have the same rank (we stress that the existence of points in $\Sigma_f \cap O$ arbitrary close to $\widehat{q}$ is also guaranteed by (c1)).

Assume that the claim is not true, i.e., there exists a sequence of points $q_n \in \Sigma_f \cap O$ such that $q_n \to \widehat{q}$ and $\operatorname{rank} q_n \leq k_O - 1$. Reasoning as in the proof of (a), using uniqueness and compactness of the minimizers, one can prove that the sequence of controls $u_n$ associated to the unique minimizers joining $q_0$ to $q_n$ satisfies $u_n \to \widehat{u}$ strongly in $L^2$, where $\widehat{u}$ is the control associated to the unique minimizer joining $q_0$ with $\widehat{q}$. By smoothness of the end-point map $E_{q_0}$ it follows that $D_{u_n}E_{q_0} \to D_{\widehat{u}}E_{q_0}$ which, by semicontinuity of the rank, implies the contradiction

$$\operatorname{rank} \widehat{q} = \operatorname{rank} D_{\widehat{u}}E_{q_0} \leq \liminf_{n \to \infty} \operatorname{rank} D_{u_n}E_{q_0} \leq k_O - 1.$$

Thus, without loss of generality, we can assume that $\operatorname{rank} q = k_O < n$ for every $q \in \Sigma_f \cap O$ (maybe by restricting our neighborhood $O$). We introduce the following set

$$\Pi_q = e^{-\vec{H}}\{\xi \in T_q^*M \mid \xi D_u E_{q_0} = \lambda_1 D_u E_{q_0}\} \subset T_{q_0}^*M.$$

The set $\Pi_q$ is the set of initial covector $\lambda_0 \in T_{q_0}^*M$ whose image via the exponential map is the point $q$.

**Lemma 11.10.** *$\Pi_q$ is an affine subset of $T_{q_0}^*M$ such that $\dim \Pi_q = n - k_O$. Moreover the map $q \mapsto \Pi_q$ is continuous.*

*Proof.* It is easy to check that the set $\widehat{\Pi}_q = \{\xi \in T_q^*M \mid \xi D_u E_{q_0} = \lambda_1 D_u E_{q_0}\}$ is an affine subspace of $T_{q_0}^*M$. Indeed $\xi \in \Pi_q$ if and only if $(D_u E_{q_0})^*(\xi - \lambda_1) = 0$, that is

$$\widehat{\Pi}_q = \{\xi \in T_q^*M \mid \xi D_u E_{q_0} = \lambda_1 D_u E_{q_0}\} = \lambda_1 + \ker (D_u E_{q_0})^*,$$

Moreover $\dim \ker (D_u E_{q_0})^* = n - \dim \operatorname{im} D_u E_{q_0} = n - k_O$. Since all elements $\xi \in \widehat{\Pi}_q$ are associated with the same control $u$, we have that $\Pi_q = e^{-\vec{H}}(\widehat{\Pi}_q) = P_{0,t}^*(\widehat{\Pi}_q)$, hence $\Pi_q$ is an affine subspace of $T_{q_0}^*M$.

Let us now show that the map $q \mapsto \Pi_q$ is continuous on $\Sigma_f \cap O$. Consider a sequence of points $q_n$ in $\Sigma_f \cap O$ such that $q_n \to q \in \Sigma_f \cap O$. Let $u_n$ (resp. $u$) be the unique control associated with the minimizing trajectory joining $q_0$ and $q_n$ (resp. $q$). By the uniqueness-compactness argument already used in the previous part of the proof we have that $u_n \to u$ strongly and moreover $D_{u_n}E_{q_0} \to D_u E_{q_0}$. Since rank $D_{u_n}E_{q_0}$ is constant, it follows that $\ker (D_{u_n}E_{q_0})^* \to \ker (D_u E_{q_0})^*$, as subspaces.

□

Consider now $A \subset T^*_{q_0}M$ a $k_O$-dimensional ball that contains $\lambda_0 = e^{-\vec{H}}(\lambda_1)$ and is transversal to $\Pi_q$. By continuity $A$ is transversal also to $\Pi_{q'}$, for $q' \in \Sigma_f \cap O$ close to $q$. In particular $\Pi_{q'} \cap A \neq \emptyset$.

Since $\exp(\Pi_q) = q$, this implies that $\Sigma_f \cap O \subset \exp(A)$. By (c1), $\Sigma_f \cap O$ is a dense set, hence $\exp(A)$ is also dense in $O$. On the other hand, since exp is a smooth map and $A$ is a compact ball of positive codimension ($k_O < n$), by Sard Lemma it follows that $\exp(A)$ is a closed dense set of $O$ that has measure zero, that is a contradiction.

*Remark* 11.11. If the structure is Riemannian, then $\Sigma_f = \Sigma_g$ since there are no abnormal extremal.

(c3) The proof of this claim relies on the following result, which is of independent interest.

**Theorem 11.12.** *Let $K \subset B$ a compact such that any length-minimizer connecting $q_0$ to $q \in K$ is strictly normal. Then $\mathsf{f} = \frac{1}{2}d^2(q_0, \cdot)$ is Lipschitz on $K$.*

*Proof of Theorem 11.12.* Let us first notice that, since $K$ is compact, it is sufficient to show that $\mathsf{f}$ is locally Lipschitz on $K$.

Fix a point $q \in K$ and some control $u$ associated with a minimizer joining $q_0$ and $q$ (it may be not unique). By our assumptions $D_u E_{q_0}$ is surjective, since $u$ is strictly normal. Thus, by inverse function theorem, there exist neighborhoods $\mathcal{V}$ of $u$ in $\mathcal{U}$ and $O_q$ of $q$ in $K$, together with a smooth map $\Phi : O_q \to \mathcal{V}$ that is a local right inverse for the end-point map, namey $E_{q_0}(\Phi(q')) = q'$ for all $q' \in O_q$ (cf. also Theorem 2.58).

Fix then local coordinates around $q$. Since $\Phi$ is smooth, there exists $R > 0$ and $C_0 > 0$ such that

$$B_q(C_0 r) \subset E_{q_0}(\mathcal{B}_u(r)), \qquad \forall\, 0 \leq r < R, \tag{11.11}$$

where $\mathcal{B}_u(r)$ is the ball of radius $r$ in $L^2$ and $B_q(r)$ is the ball of radius $r$ in coordinates on $M$. Let us also observe that, since $J$ is smooth on, there exists $C_1 > 0$ such that for every $u, u' \in \mathcal{B}_u(R)$ one has

$$J(u') - J(u) \leq C_1 \|u' - u\|_{L^2} \tag{11.12}$$

Pick then any point $q' \in K$ such that $|q' - q| = C_0 r$, with $0 \leq r \leq R$. By (11.11), there exists $u' \in \mathcal{B}_u(R)$ with $\|u' - u\|_{L^2} \leq r$ such that $E_{q_0}(u') = q'$. Using that $\mathsf{f}(q') \leq J(u')$ and $\mathsf{f}(q) = J(u)$, since $u$ is a minimizer, we have

$$\mathsf{f}(q') - \mathsf{f}(q) \leq J(u') - J(u) \leq C_1 \|u' - u\|_{L^2} \leq C' |q' - q|,$$

where $C' = C_1/C_0$. Notice that the above inequality is true for all $q'$ such that $|q' - q| \leq C_0 R$.

Since $K$ is compact, and the set of control $u$ associated with minimizers that reach the compact set $K$ is also compact, the constants $R > 0$ and $C_0, C_1$ can be chosen uniformly with respect to $q \in K$. Hence we can exchange the role of $q'$ and $q$ in the above reasoning and get

$$|\mathsf{f}(q') - \mathsf{f}(q)| \leq C' |q' - q|,$$

for every pair of points $q, q'$ such that $|q' - q| \leq C_0 R$.

$\square$

To end the proof of (c3) it is sufficient to show that if $q \in \Sigma_g$ there exists a (compact) neighborhood $O_q$ of $q$ such that every point in $O_q$ is reached by only strictly normal minimizers (we stress that no uniqueness is required here). By contradiction, assume that the claim is not true. Then there exists a sequence of points $q_n$ converging to $q$ and a choice of controls $u_n$, such that the corresponding minimizers are abnormal. By compacness of minimizers there exists $u$ such that $u_n \to u$ and by uniqueness of the limit $u$ is abnormal for the point $q$, that is a contradiction.

(c4). We have to prove that $\Sigma \cap O$ is non empty for every open neighborhood $O$ in $B$. By (c3) we can choose $q' \in \Sigma_g \cap O$ and fix $O' \subset O$ neighborhood of $q$ such that $\mathsf{f}$ is Lipschitz on $O'$. It is then sufficient to show that $\Sigma \cap O' \neq \emptyset$.

By Proposition 11.4 (see also Remark 11.5) every differentiability point of $\mathsf{f}$ is reached by a unique minimizer that is normal, hence is a fair point. Since we know that $\mathsf{f}$ is Lipschitz on $O'$, it follows by Rademacher Theorem that almost every point of $O'$ is fair, namely $\text{meas}(\Sigma_f \cap O') = \text{meas}(O')$.

Let us also notice that the set $\Sigma_f \cap O'$ of fair points of $O'$ is also contained in the image of the exponential map. Thanks to the Sard Lemma, the set of regular values of the exponential map in $O'$ is also a set of full measure in $O'$. Since by definition a point in $\Sigma_f$ that is a regular value for the exponential map is in $\Sigma$, this implies that $\text{meas}(\Sigma \cap O') = \text{meas}(\Sigma_f \cap O') = \text{meas}(O')$. This in particular proves that $\Sigma \cap O'$ is not empty. $\square$

As a corollary of this result we can prove that if there are no abnormal minimizers, then the set of smooth points has full measure

**Corollary 11.13.** *Assume that $M$ is a complete sub-Riemannian structure and that there are no abnormal minimizers. Then* $\text{meas}(M \setminus \Sigma) = 0$.

The conclusion of Corollary 11.13 remains true in absence of stricly abnormal minimizers, since in this case the squared sub-Riemannian distance is locally Lipschitz outside the diagonal, cf. Theorem 12.12.

We end this section by stating explicitly a result on the regularity of the squared distance on the diagonal, which is a direct consequence of the previous analysis.

**Proposition 11.14.** *Let $M$ be a sub-Riemannian structure. Then there exists a neighborhood $O_{q_0}$ of $q_0$ such that $\mathsf{f}$ is smooth on $O_{q_0}$ if and only if* $\dim \mathcal{D}_{q_0} = \dim M$.

## 11.2 Locally Lipschitz functions and maps

If $S$ is a subset of a vector space $V$, we denote by $\text{conv}(S)$ the *convex hull* of $S$, that is the smallest convex set of $V$ containing $S$. It is characterized as the set of $v \in V$ such that there exists a finite number of elements $v_0, \ldots, v_\ell \in S$ and real numbers $\lambda_0, \ldots, \lambda_\ell$ such that

$$v = \sum_{i=0}^{\ell} \lambda_i v_i, \qquad \lambda_i \geq 0, \quad \sum_{i=0}^{\ell} \lambda_i = 1.$$

Let $\varphi : M \to \mathbb{R}$ be a function defined on a smooth manifold $M$. We say that $\varphi$ is *locally Lipschitz* if $\varphi$ is locally Lipschitz in any coordinate chart, as a function defined on $\mathbb{R}^n$.

The classical Rademacher theorem implies that a locally Lipschitz function $\varphi : M \to \mathbb{R}$ is differentiable almost everywhere. Still one can introduce a weak notion of differential, that is defined at every point.

If $\varphi : M \to \mathbb{R}$ is locally Lipschitz, any point $q \in M$ is the limit of differentiability points. In what follows, whenever we write $d_q\varphi$, it is implicitly understood that $q \in M$ is a differentiability point of $\varphi$.

**Definition 11.15.** Let $\varphi : M \to \mathbb{R}$ be a locally Lipschitz function. The *(Clarke) generalized differential* of $\varphi$ at the point $q \in M$ is the set

$$\partial_q \varphi := \operatorname{conv}\{\xi \in T_q^* M \,|\, \xi = \lim_{q_n \to q} d_{q_n} \varphi\} \tag{11.13}$$

The set $\partial_q \varphi$ is a compact subset of the vector space $T_q^* M$. In fact, it is closed by definition, and bounded since the function is locally Lipschitz.

**Exercise 11.16.** (i). Show that the map $q \mapsto \partial_q \varphi$ is *upper semicontinuous*. In other words prove that for every sequence $q_n \to q$ in $M$, and every sequence $\xi_n \in \partial_{q_n} \varphi$ such that $\xi_n \to \xi$ in $T^* M$, one has $\xi \in \partial_q \varphi$.

(ii). We say that $q$ is a *regular point* for $\varphi$ if $0 \notin \partial_q \varphi$. Prove that the set of regular points for $\varphi$ is open in $M$.

From the very definition of generalized differential we have the following result.

**Lemma 11.17.** *Let $\varphi : M \to \mathbb{R}$ be a locally Lipschitz function and $q \in M$. The following statements are equivalent:*

*(i) $\partial_q \varphi = \{\xi\}$ is a singleton,*

*(ii) $d_q \varphi = \xi$ and the map $x \mapsto d_x \varphi$ is continuous at $q$ in the following sense: for every sequence of differentiability points $q_n \to q$ we have $d_{q_n} \varphi \to d_q \varphi$.*

*Remark* 11.18. Let $A$ be a subset of $\mathbb{R}^n$ of measure zero and, given $q \in \mathbb{R}^n$, consider the set $L_v = \{q + tv, t \geq 0\}$ of half-lines emanating from $q$ and parametrized by $v \in S^{n-1}$. It follows from Fubini's theorem that for almost every $v \in S^{n-1}$ the one-dimensional measure of the intersection $A \cap L_v$ is zero.

If we apply this fact to the case when $A$ is the set at which a locally Lipschitz function $\varphi : \mathbb{R}^n \to \mathbb{R}$ fails to be differentiable, we deduce that, given $q \in \mathbb{R}^n$, for almost every $v \in S^{n-1}$ the function $t \mapsto \varphi(q + tv)$ is differentiable a.e. on $[0, +\infty[$.

**Example 11.19.** Let $\varphi : \mathbb{R} \to \mathbb{R}$ defined by

(i) $\varphi(x) = |x|$. Then $\partial_0 \varphi = [-1, 1]$,

(ii) $\varphi(x) = x$, if $x < 0$ and $\varphi(x) = 2x$, if $x \geq 0$. In this case $\partial_0 \varphi = [1, 2]$.

In particular in the first example 0 is a minimum for $\varphi$ and $0 \in \partial_0 \varphi$. In the second case the function is locally invertible near the origin and $\partial_0 \varphi$ is separated from zero. In what follows we will prove that these fact corresponds to general results (cf. Proposition 11.23 and Theorem 11.27).

The following is a classical hyperplane separation theorem for closed convex sets in $\mathbb{R}^n$.

**Lemma 11.20.** *Let $K$ and $C$ be two disjoint, closed, convex sets in $\mathbb{R}^n$, and suppose that $K$ is compact. Then there exists $\varepsilon > 0$ and a vector $v \in S^{n-1}$ such that*

$$\langle x, v \rangle > \langle y, v \rangle + \varepsilon, \qquad \forall\, x \in K, \forall\, y \in C. \tag{11.14}$$

We also recall here another useful result from convex analysis.

**Lemma 11.21** (Carathéodory)**.** *Let $S \subset \mathbb{R}^n$ and $x \in \operatorname{conv}(S)$. Then there exist $x_0, \dots, x_n \in S$ such that $x \in \operatorname{conv}\{x_0, \dots, x_n\}$.*

The notion of generalized gradient permits to extend some classical properties of critical points of smooth functions to locally Lipschitz ones.

**Proposition 11.22.** *Let $\varphi : M \to \mathbb{R}$ be a locally Lipschitz function and $q$ be a local minimum for $\varphi$. Then $0 \in \partial_q \varphi$.*

*Proof.* Since the claim is a local property we can assume without loss of generality that $M = \mathbb{R}^n$. As usual we will identify vectors and covectors with elements of $\mathbb{R}^n$ and the duality covectors-vectors is given by the Euclidean scalar product, that we still denote $\langle \cdot, \cdot \rangle$.

Assume by contradiction that $0 \notin \partial_q \varphi$ and let us show that $q$ cannot be a minimum for $\varphi$. To this aim, we prove that there exists a direction $w$ in $S^{n-1}$ such that the scalar map $t \mapsto \varphi(q + tw)$ has no minimum at $t = 0$.

The set $\partial_q \varphi$ is a compact convex set that does not contain the origin, hence by Lemma 11.20, there exist $\varepsilon > 0$ and $v \in S^{n-1}$ such that

$$\langle \xi, v \rangle < -\varepsilon, \quad \forall\, \xi \in \partial_q \varphi.$$

By definition of generalized differential, one can find open neighborhoods $O_q$ of $q$ in $\mathbb{R}^n$ and $V_v$ of $v$ in $S^{n-1}$ such that for all differentiability point $q' \in O_q$ of $\varphi$ one has

$$\left\langle d_{q'} \varphi, v' \right\rangle \leq -\varepsilon/2, \qquad \forall\, v' \in V_v.$$

Fix $q' \in O_q$ where $\varphi$ is differentiable and a vector $w \in V_v$ such that the set of differentiable points of the restriction of $\varphi$ to the line $\{q + tw\}$ has full measure (cf. Remark 11.18). Then we can compute for $t > 0$

$$\varphi(q + tw) - \varphi(q) = \int_0^t \langle d_{q+sw} \varphi, w \rangle \, ds \leq -\varepsilon t/2.$$

Thus $\varphi$ cannot have a minimum at $q$. $\square$

The following proposition gives an estimate for the generalized differential of a special class of functions.

**Proposition 11.23.** *Let $\varphi_\omega : M \to \mathbb{R}$ be a family of $C^1$ functions, with $\omega \in \Omega$ a compact set. Assume that the following maps are continuous:*

$$(\omega, q) \mapsto \varphi_\omega(q), \qquad (\omega, q) \mapsto d_q \varphi_\omega \tag{11.15}$$

*Then the function $a(q) := \min_{\omega \in \Omega} \varphi_\omega(q)$ is locally Lipschitz on $M$ and*

$$\partial_q a \subset \operatorname{conv}\{d_q \varphi_\omega \mid \forall\, \omega \in \Omega \ s.t.\ \varphi_\omega(q) = a(q)\}. \tag{11.16}$$

*Proof.* As in the proof of Proposition 11.22 we can assume that $M = \mathbb{R}^n$. Notice that, if we denote by $\Omega_q = \{\omega \in \Omega \mid \varphi_\omega(q) = a(q)\}$ we have by compactness of $\Omega$ that $\Omega_q$ is non empy for every $q \in M$ and we can rewrite the claim as follows

$$\partial_q a \subset \operatorname{conv}\{d_q \varphi_\omega \mid \omega \in \Omega_q\}. \tag{11.17}$$

We divide the proof into two steps. In step (i) we prove that $a$ is locally Lipschitz and then in (ii) we show the estimate (11.17).

(i). Fix a compact $K \subset M$. Since every $\varphi_\omega$ is Lipschitz on $K$ and $\Omega$ is compact, thanks to the continuity of the map (11.15), there exists a common Lipschitz constant $C_K > 0$, i.e. the following inequality holds

$$\varphi_\omega(q) - \varphi_\omega(q') \leq C_K|q - q'|, \qquad \forall\, q, q' \in K, \quad \omega \in \Omega,$$

Clearly we have

$$\min_{\omega\in\Omega} \varphi_\omega(q) - \varphi_\omega(q') \leq C_K|q - q'|, \qquad \forall\, q, q' \in K, \quad \omega \in \Omega,$$

and since the last inequality holds for all $\omega \in \Omega$ we can pass to the min with respect to $\omega$ in the left hand side and

$$a(q) - a(q') \leq C_K|q - q'|, \qquad \forall\, q, q' \in K.$$

Since the constant $C_K$ depends only on the compact set $K$ we can exchange in the previous reasoning the role of $q$ and $q'$, that gives

$$|a(q) - a(q')| \leq C_K|q - q'|, \qquad \forall\, q, q' \in K.$$

(ii). Define $D_q := \mathrm{conv}\{d_q\varphi_\omega | \forall\, \omega \in \Omega_q\}$. Let us first prove prove that $d_q a \in D_q$ for every differentiability point $q$ of $a$.

Fix any $\xi \notin D_q$. By Lemma 11.20 applied to the pair $D_q$ and $\{\xi\}$, there exist $\varepsilon > 0$ and $v \in S^{n-1}$ such that

$$\langle d_q\varphi_\omega, v\rangle > \langle \xi, v\rangle + \varepsilon, \quad \forall\, \omega \in \Omega_q,$$

By continuity of the map $(\omega, q) \mapsto d_q\varphi_\omega$, there exists a neighborhood $O_q$ of $q$ and $V$ neighborhood of $\Omega_q$ such that

$$\left\langle d_{q'}\varphi_{\omega'}, v\right\rangle > \langle \xi, v\rangle + \varepsilon/2, \qquad \forall\, q' \in O_q, \quad \forall\, \omega' \in V,$$

An integration argument let us to prove that there exists $\delta > 0$ such that for $\omega \in V$

$$\frac{1}{t}(\varphi_\omega(q + tv) - \varphi_\omega(q)) > \langle \xi, v\rangle + \varepsilon/4, \qquad \forall\, 0 < t < \delta.$$

Clearly we have

$$\frac{1}{t}(\varphi_\omega(q + tv) - a(q)) \geq \langle \xi, v\rangle + \varepsilon/4, \qquad \forall\, 0 < t < \delta.$$

and since the minimum in $a(q + tv) = \min_{\omega\in\Omega} \varphi_\omega(q + tv)$ is attained for $\omega$ in $\Omega_{q+tv} \subset V$ for $t$ small enough, we can pass to the minimum w.r.t. $\omega \in V$ in the left hand side, proving that there exists $t_0 > 0$ such that

$$\frac{1}{t}(a(q + tv) - a(q)) \geq \langle \xi, v\rangle + \varepsilon/4, \qquad \forall\, 0 < t < t_0.$$

Passing to the limit for $t \to 0$ we get

$$\langle d_q a, v\rangle \geq \langle \xi, v\rangle + \varepsilon/4 \tag{11.18}$$

If $d_q a \notin D_q$ we can choose $\xi = d_q a$ in the above reasoning and (11.18) gives the contradiction $\langle d_q a, v\rangle \geq \langle d_q a, v\rangle + \varepsilon/4$. Hence $d_q a \in D$ for every differentiability point $q$ of $a$.

Now suppose that one has a sequence $q_n \to q$, where $q_n$ are differentiability points of $a$. Then $d_{q_n} a \in D_{q_n}$ for all $n$ from the first part of the proof. We want to show that, whenever the limit

$\xi = \lim_{n\to\infty} d_{q_n} a$ exists, then $\xi \in D_q$. This is a consequence of the fact that the map $(\omega, q) \mapsto d_q\varphi_\omega$ is continuous (in particular upper semicontinuous in the sense of Exercise 11.16) and the fact that $\Omega$ is compact. $\square$

**Exercise 11.24.** Complete the second part of the proof of Proposition 11.23. (Hint: use Carathéodory lemma.)

### 11.2.1 Locally Lipschitz map and Lipschitz submanifolds

As for scalar functions, a map $f : M \to N$ between smooth manifolds is said to be locally Lipschitz if for any coordinate chart in $M$ and $N$ the corresponding function from $\mathbb{R}^n$ to $\mathbb{R}^m$ is locally Lipschitz.

For a locally Lipschitz map between manifolds $f : M \to N$ the *(Clarke) generalized differential* is defined as follows

$$\partial_q f := \mathrm{conv}\{L \in \mathrm{Hom}(T_qM, T_{f(q)}N) |\, L = \lim_{q_n\to q} D_{q_n} f,\ q_n \text{ diff. point of } f\},$$

The following lemma shows how the standard chain rule extends to locally Lipschitz maps.

**Lemma 11.25.** *Let $M, N, W$ be smooth manifolds and $f : M \to N$ be a locally Lipschitz map.*

*(a) If $\phi : M \to M$ is a diffeomorphism and $q \in M$ we have*

$$\partial_q(f \circ \phi) = \partial_{\phi(q)} f \cdot D_q\phi. \tag{11.19}$$

*(b) If $\varphi : N \to W$ is a map of class $C^1$, and $q \in M$ we have*

$$\partial_q(\varphi \circ f) = D_{f(q)}\varphi \cdot \partial_q f. \tag{11.20}$$

*Moreover the generalized differential is* upper semicontinuous *as a set-valued function. Namely, for every neighborhood $\Omega \in \mathrm{Hom}(T_qM, T_{f(q)}N)$ of $\partial_q f$ there exists a neighborhood $O_q$ of $q$ such that $\partial_{q'} f \in \Omega$, for every $q' \in O_q$.*

*Sketch of the proof.* For a detailed proof of this result see [Cla90, Sect. 2.3]. Here we only give the main ideas.

Claim (a). Since $\phi$ is a diffeomorphism, it sends every differentiability point $q$ of $f \circ \phi$ to a differentiability point $\phi(q)$ for $f$. Then (11.19) is true at differentiability point and passing to the limit it is also valid for sub-differential (one proves both inclusions using $\phi$ and $\phi^{-1}$). Claim (b) can be proved along the same lines. The upper semicontinuity property can be proved by using the hyperplane separation theorem and the Carathéodory lemma. $\square$

**Definition 11.26.** Let $f : M \to N$ be a locally Lipschitz map. A point $q \in M$ is said *critical point* for $f$ if $\partial_q f$ contains a non-surjective map. If $q \in M$ is not a critical point, then it is said *regular point.*

Notice that by the semicontinuity property of Lemma 11.25, it follows that the set of regular points of a locally Lipschitz map $f$ is open.

**Theorem 11.27.** *Let $f : \mathbb{R}^n \to \mathbb{R}^n$ be a locally Lipschitz map and $q \in M$ be a regular point for $f$. Then there exist neighborhoods $O_q$ and $O_{f(q)}$ of $q$ and $f(q)$ respectively, and a locally Lipschitz map $g : O_{f(q)} \to \mathbb{R}^n$ such that $f \circ g = \mathrm{Id}_{O_{f(q)}}$ and $g \circ f|_{O_q} = \mathrm{Id}_{O_q}$.*

*Remark* 11.28. The classical $C^1$ version of the inverse function theorem can be proved from Theorem 11.27 and the chain rule (Lemma 11.25). Indeed Theorem 11.27 implies that there exists a locally Lipschitz inverse $g$ and using the chain rule it is easy to show that the sub-differential of $g$ contains only one element (this implies that it is differentiable at that point) and the differential of $g$ is the inverse of the differential of $f$.

Before proving Theorem 11.27 we need the following technical lemma.

**Lemma 11.29.** *Let $f : \mathbb{R}^n \to \mathbb{R}^n$ be a locally Lipschitz map and $q \in M$ be a regular point. Then there exist a neighborhood $O_q$ of $q$ and $\varepsilon > 0$ such that*

$$\forall\, v \in S^{n-1},\ \exists\, \xi_v \in S^{n-1} \qquad \text{s.t.} \qquad \langle \xi_v, \partial_x f(v)\rangle > \varepsilon, \quad \forall\, x \in O_q. \tag{11.21}$$

*Moreover $|f(x) - f(y)| \geq \varepsilon |x - y|$, for all $x, y \in O_q$.*

We stress that (11.21) means that the inequality $\langle \xi_v, L(v)\rangle > \varepsilon$ holds for every $x \in O_q$ and every element $L \in \partial_x f$.

*Proof.* Notice that, since $q$ is a regular point, the set $\partial_q f$ contains only invertible linear maps. For every $v \in S^{n-1}$, the set $\partial_q f(v)$ is compact and convex, and does not contain the zero linear map. By the hyperplane separation theorem we can find $\xi_v$ such that $\langle \xi_v, \partial_q f(v)\rangle > \varepsilon(v)$. The map $x \mapsto \partial_x f$ is upper semicontinuous, hence there exists a neighborhood $O_q$ of $q$ such that $\langle \xi_v, \partial_x f(v)\rangle > \varepsilon(v)$ for all $x \in O_q$. Since $S^{n-1}$ is compact, there exists a uniform $\varepsilon = \min\{\varepsilon(v), v \in S^{n-1}\}$ that satisfies (11.21).

To prove the second statement of the Lemma, write $y = x + sv$, where $s = |x - y|$ and $v \in S^{n-1}$. Consider a vector $v' \in S^{n-1}$ close to $v$ such that almost every point in the direction of $v'$ is a point of differentiability (cf. Remark 11.18), and set $y' = x + sv'$ and $\xi_{v'}$ the vector associated to $v'$ defined by (11.21). Then we can write

$$f(y') - f(x) = \int_0^s (D_{x+tv'} f) v' dt.$$

and we have the inequality

$$\begin{aligned} |f(y') - f(x)| &\geq \left\langle \xi_{v'}, f(y') - f(x) \right\rangle \\ &= \int_0^s \left\langle \xi_{v'}, (D_{x+tv'} f) v' \right\rangle dt \\ &\geq \varepsilon |y' - x| \end{aligned}$$

Since $\varepsilon$ does not depend on $v$, we can pass to the limit for $v' \to v$ in the above inequality (in particular $y' \to y$) and the Lemma is proved. $\square$

*Proof of Theorem 11.27.* The inequality proved in Lemma 11.29 implies that $f$ is injective in the neighborhood $O_q$ of the point $q$. If we show that $f(O_q)$ covers a neighborhood $O_{f(q)}$ of the point $f(q)$, then the inverse function $g : O_{f(q)} \to \mathbb{R}^n$ is well defined and locally Lipschitz.

Without loss of generality, up to restricting the neighborhood $O_q$, we can assume that every point in $O_q$ is regular for $f$ and moreover that the estimate of the Lemma 11.29 holds also on the topological boundary $\partial O_q$. Lemma 11.29 also implies that

$$\operatorname{dist}(f(q), \partial f(O_q)) \geq \varepsilon \operatorname{dist}(q, \partial O_q) > 0, \tag{11.22}$$

where $\operatorname{dist}(x, A) = \inf_{y \in A} |x - y|$ denotes the Euclidean distance from $x$ to the set $A$. Thanks to the continuity of $f$ and (11.22), there exists a neighborhood $W \subset f(O_q)$ of $f(q)$ such that $|y - f(q)| < \operatorname{dist}(y, \partial f(O_q))$, for every $y \in W$. Fix an arbitrary $\bar{y} \in W$ and let us show that the equation $f(x) = \bar{y}$ has a solution. Define the function

$$\psi : \overline{O_q} \to \mathbb{R}, \qquad \psi(x) = |f(x) - \bar{y}|^2. \tag{11.23}$$

By construction $\psi(q) < \psi(x)$, for all $x \in \partial O_q$, hence by continuity $\psi$ attains the minimum at some point $\bar{x} \in O_q$. By Proposition 11.22, we have $0 \in \partial_{\bar{x}}\psi$. Moreover, using the chain rule

$$\partial_{\bar{x}}\psi = (f(\bar{x}) - \bar{y})^T \cdot \partial_{\bar{x}} f. \tag{11.24}$$

Since $\bar{x}$ is a regular point of $f$, the linear map $\partial_{\bar{x}} f$ is invertible. Thus $0 \in \partial_{\bar{x}}\psi$ implies $f(\bar{x}) = \bar{y}$. $\square$

We say that $c \in \mathbb{R}$ is a regular value of a locally Lipschitz function $\varphi : M \to \mathbb{R}$ if $\varphi^{-1}(c) \neq \emptyset$ and every $x \in \varphi^{-1}(c)$ is a regular point.

**Corollary 11.30.** *Let $\varphi : M \to \mathbb{R}$ be locally Lipschitz and assume that $c \in \mathbb{R}$ is a regular value for $\varphi$. Then $\varphi^{-1}(c)$ is a Lipschitz submanifold of $M$ of codimension 1.*

*Proof.* We show that in any small neighborhood $O_x$ of every $x \in \varphi^{-1}(c)$ the set $O_x \cap \varphi^{-1}(c)$ can be described as the zero locus of a locally Lipschitz function. Since $\partial_x\varphi$ does not contain 0, by the hyperplane separation theorem there exists $v_1 \in S^{n-1}$, such that $\langle \partial_x\varphi, v_1 \rangle > 0$ for every $x$ in the compact neighborhood $O_x \cap \varphi^{-1}(y)$.

Let us complete $v_1$ to an orthonormal basis $\{v_1, v_2, \dots, v_n\}$ of $\mathbb{R}^n$ and consider the map

$$f : O_x \to \mathbb{R}^n, \qquad f(x') = \begin{pmatrix} \varphi(x') - c \\ \langle v_2, x' \rangle \\ \vdots \\ \langle v_n, x' \rangle \end{pmatrix}$$

By construction $f$ is locally Lipschitz and $x$ is a regular point of $f$. By Theorem 11.27 there exists a Lipschitz inverse $g$ of $f$. In particular the inverse map is a Lipschitz function that transforms the hyperplane $\{y_1 = 0\}$ into $\varphi^{-1}(c)$. Hence the level set $\varphi^{-1}(c)$ is a Lipschitz submanifold. $\square$

### 11.2.2 A non-smooth version of Sard Lemma

In this section we prove a Sard-type result for the special class of Lipschitz functions we considered in the previous section.

We first recall the statement of the classical Sard lemma. We denote by $C_f$ the critical point of a smooth map $f : M \to N$, i.e., the set of points $x$ in $M$ at which the differential of $f$ is not surjective.

**Theorem 11.31** (Sard lemma)**.** *Let $f : \mathbb{R}^n \to \mathbb{R}^m$ be a $C^k$ function, with $k \geq \max\{n - m + 1, 1\}$. Then the set $f(C_f)$ of critical values of $f$ has measure zero in $\mathbb{R}^m$.*

Notice that the classical Sard Lemma does not apply to $C^1$ functions $\varphi : \mathbb{R}^n \to \mathbb{R}$, whenever $n \geq 1$. The following version of Sard lemma is due to Rifford.

**Theorem 11.32.** *Let $M$ be a smooth manifold and $\varphi_\omega : M \to \mathbb{R}$ a family of smooth functions, with $\omega \in \Omega$. Assume that*

*(i)* $\Omega = \bigcup_{i \in \mathbb{N}} N_i$ *is the union of smooth submanifold, and is compact,*

*(ii)* *the maps $(\omega, q) \mapsto \varphi_\omega(q)$ and $(\omega, q) \mapsto d_q\varphi_\omega$ are continuous on $\Omega \times M$,*

*(iii)* *the maps $\psi_i : N_i \times M \to \mathbb{R}$, $(\omega, q) \mapsto \varphi_\omega(q)$ are smooth.*

*Then the set of critical values of the function $a(q) = \min\limits_{\omega \in \Omega} \varphi_\omega(q)$ has measure zero in $\mathbb{R}$.*

*Proof.* We are going to define a countable set of smooth functions $\Phi_\alpha$ indexed by $\alpha = (\alpha_0, \ldots, \alpha_n) \in \mathbb{N}^{n+1}$ (here $n = \dim M$), such that for every critical point $q$ of $a$ there exists a critical point $z_q$ of $\Phi_\alpha$, for some $\alpha$, such that $\Phi_\alpha(z_q) = a(q)$.

Denote by $\Lambda_n = \{(\lambda_0, \ldots, \lambda_n) \mid \lambda_i \geq 0, \sum_{i=0}^n \lambda_i = 1\}$. For every $\alpha = (\alpha_0, \ldots, \alpha_n) \in \mathbb{N}^{n+1}$ let us consider the map

$$\Phi_\alpha : N_{\alpha_0} \times \ldots \times N_{\alpha_n} \times \Lambda_n \times M \to \mathbb{R}$$

$$\Phi_\alpha(\omega_0, \ldots, \omega_n, \lambda_0, \ldots, \lambda_n, q) = \sum_{i=0}^n \lambda_i \varphi_{\omega_i}(q). \tag{11.25}$$

By computing partial derivatives, it is easy to see that a point $z = (\omega_0, \ldots, \omega_n, \lambda_0, \ldots, \lambda_n, q)$ is critical for $\Phi_\alpha$ if and only if it satisfies the following relations:

$$\begin{cases} \sum_{i=0}^n \lambda_i \dfrac{\partial \psi_{\alpha_i}}{\partial \omega}(\omega_i, q) = 0, \qquad i = 0, \ldots, n, \\ \sum_{i=0}^n \lambda_i d_q \varphi_{\omega_i} = 0, \qquad i = 0, \ldots, n, \\ \varphi_{\omega_0}(q) = \ldots = \varphi_{\omega_n}(q) \end{cases} \tag{11.26}$$

Recall that $\psi_i$ is the restriction of the map $(\omega, q) \mapsto \varphi_\omega(q)$ to $N_i \times M$.

Let us now show that every critical point $q$ of $a$ can be associated with a critical point $z_q$ of some $\Phi_\alpha$. By Proposition 11.23, the function $a$ is locally Lipschitz. Assume that $q$ is a critical point of $a$, then we have

$$0 \in \partial_q a \subset \operatorname{conv}\{d_q\varphi_\omega \mid \forall\, \omega \in \Omega \ s.t.\ \varphi_\omega(q) = a(q)\}.$$

By Carathéodory lemma there exist $n+1$ element $\bar\omega_0, \ldots, \bar\omega_n$ and $n+1$ scalars $\bar\lambda_0, \ldots, \bar\lambda_n$ such that $\bar\lambda_i \geq 0, \sum_{i=0}^n \bar\lambda_i = 1$ and

$$0 = \sum_{i=0}^n \bar\lambda_i d_q \varphi_{\bar\omega_i}, \qquad \varphi_{\bar\omega_i}(q) = a(q), \quad \forall\, i = 0, \ldots, n.$$

Moreover, let us choose for every $i = 0, \ldots, n$ an index $\bar{\alpha}_i \in \mathbb{N}$ such that $\bar{\omega}_i \in N_{\bar{\alpha}_i}$. Since $\varphi_{\bar{\omega}_i}(q) = a(q) = \min_\Omega \varphi_\omega(q)$, $\bar{\omega}_i$ is critical for the map $\psi_{\alpha_i}$, namely we have

$$\frac{\partial \psi_{\alpha_i}}{\partial \omega}(\bar{\omega}_i, q) = 0.$$

This implies that $z_q = (\bar{\omega}_0, \ldots, \bar{\omega}_n, \bar{\lambda}_0, \ldots, \bar{\lambda}_n, q)$ satisfies the relations (11.26) for the function $\Phi_{\bar{\alpha}}$, with $\bar{\alpha} = (\bar{\alpha}_0, \ldots, \bar{\alpha}_n)$. Moreover it is easy to check that $\Phi_{\bar{\alpha}}(z_q) = a(q)$ since

$$\Phi_{\bar{\alpha}}(z_q) = \sum_{i=0}^{n} \bar{\lambda}_i \varphi_{\bar{\omega}_i}(q) = \left(\sum_{i=0}^{n} \bar{\lambda}_i\right) a(q) = a(q).$$

Then if $C_a$ denotes the set of critical points of $a$ and $C_\alpha$ the set of critical points of $\Phi_\alpha$ we have

$$\operatorname{meas}(a(C_a)) \leq \operatorname{meas}\left(\bigcup_{\alpha \in \mathbb{N}^{n+1}} \Phi_\alpha(C_\alpha)\right) \leq \sum_{\alpha \in \mathbb{N}^{n+1}} \operatorname{meas}(\Phi_\alpha(C_\alpha)) = 0,$$

since $\operatorname{meas}(\Phi_\alpha(C_\alpha)) = 0$ for every fixed $\alpha \in \mathbb{N}^{n+1}$, by the classical Sard lemma for $C^\infty$ functions. $\square$

We want to apply the previous result in the case of functions that are infimum of smooth functions on level sets of a submersion.

**Theorem 11.33.** *Let $F : N \to M$ be a smooth map between finite dimensional manifolds and $\varphi : N \to \mathbb{R}$ be a smooth function. Assume that*

*(i) $F$ is a submersion*

*(ii) for all $q \in M$ the set $N_q = \{x \in N, \varphi(x) = \min_{y \in F^{-1}(q)} \varphi(y)\}$ is a non empty compact set.*

*Then the set of critical values of the function $a(q) = \min_{x \in F^{-1}(q)} \varphi(x)$ has measure zero in $\mathbb{R}$.*

*Proof.* Denote by $C_a$ the set of critical points of $a$ and $a(C_a)$ is the set of its critical values. Let us first show that for every point $q \in M$ there exist an open neighborhood $O_q$ of $q$ such that $\operatorname{meas}(a(C_a) \cap O_{q_n}) = 0$.

From assumption (i), it follows that for every $q \in M$ the set $F^{-1}(q)$ is a smooth submanifold in $N$. Let us now consider an auxiliary non negative function $\psi : N \to \mathbb{R}$ such that

(A0) $A_\alpha := \psi^{-1}([0, \alpha])$ is compact for every $\alpha > 0$.

and select moreover a constant $c > 0$ such that the following assumptions are satisfied:

(A1) $N_q \subset \operatorname{int} A_c$,

(A2) $c$ is a regular level of $\psi\big|_{F^{-1}(q)}$.

The existence of such a $c > 0$ is guaranteed by the fact that (A1) is satisfied for all $c$ big enough since $N_q$ is compact and $A_c$ contains any compact as $c \to +\infty$. Moreover, by classical Sard lemma (cf. Theorem 11.31), almost every $c$ is a regular value for the smooth function $\psi\big|_{F^{-1}(q)}$.

By continuity, there exists a neighborhood $O_q$ of the point $q$ such that assumptions (A0)-(A2) are satisfied for every $q' \in O_q$, for $c > 0$ and $\psi$ fixed. We observe that (A2) is equivalent to require that level set of $F$ are transversal to level of $\psi$. We can infer that $F^{-1}(O_q) \cap A_c$ is a smooth manifold with boundary that has the structure of locally trivial bundle. Maybe restricting the neighborhood of $q$ then we can assume

$$F^{-1}(q) \cap A_c = \Omega, \qquad F^{-1}(O_q) \cap A_c \simeq O_q \times \Omega,$$

where $\Omega$ is a smooth manifold with boundary. In this neighborhood we can split variables in $N$ as follows $x = (\omega, q)$ with $\omega \in \Omega$ and $q \in M$ and the restriction $a|_{O_q}$ is written as

$$a|_{O_q} : O_q \to \mathbb{R}, \qquad a(q) = \min_{\omega \in \Omega} \varphi(\omega, q).$$

Notice that $\Omega$ is compact and is the union of its interior and its boundary, which are smooth by assumptions (A0)-(A2). We can then apply the Theorem 11.32 to $a|_{O_q}$, that gives $\text{meas}(a(C_a \cap O_q) = 0$ for every $q \in M$.

We have built a covering of $M = \bigcup_{q \in M} O_q$. Since $M$ is a smooth manifold, from every covering it is possible to extract a countable covering, i.e., there exists a sequence $q_n$ of points in $M$ such that

$$M = \bigcup_{n \in \mathbb{N}} O_{q_n}$$

In particular this implies that

$$\text{meas}(a(C_a)) \leq \sum_{n \in \mathbb{N}} \text{meas}(a(C_a) \cap O_{q_n}) = 0$$

since $\text{meas}(a(C_a \cap O_q) = 0$ for every $q$.

$\square$

*Remark* 11.34. Notice that we do not assume that $N$ is compact. In that case the proof is easier since every submersion $F : N \to M$ with $N$ compact automatically endows $N$ with a locally trivial bundle structure.

## 11.3 Regularity of sub-Riemannian spheres

We end this chapter by applying the previous theory to get information about the regularity of sub-Riemannian spheres. Before proving the main result we need two lemmas.

**Lemma 11.35.** *Let $\mathcal{K} \subset T^*M \setminus H^{-1}(0)$ be a compact set of $T^*M$ such that all normal extremals associated with $\lambda_0 \in \mathcal{K}$ are not abnormal. Then there exists $\varepsilon = \varepsilon(\mathcal{K})$ such that $t\lambda_0$ is a regular point for $\exp_{\pi(\lambda_0)}$, for every $\lambda_0 \in \mathcal{K}$ and $0 < t \leq \varepsilon$.*

*Proof.* Let $\Pi : T^*M \setminus H^{-1}(0) \to H^{-1}(1/2)$ be the continuous map defined by

$$\Pi(\lambda) = \frac{\lambda}{\sqrt{2H(\lambda)}}.$$

The set $\mathcal{K}_1 := \Pi(\mathcal{K})$ is compact in $H^{-1}(1/2)$. Applying Proposition 8.76 to $\mathcal{K}_1$, there exists $\varepsilon' = \varepsilon'(\mathcal{K}_1)$ such that the cut time $t_*(\gamma_{\lambda_0}) \geq \varepsilon'$ for every normal trajectory $\gamma_{\lambda_0}$ associated with

$\lambda_0 \in \mathcal{K}_1$. Hence $t\lambda_0$ is a regular point for the $\exp_{\pi(\lambda_0)}$ for all $\lambda_0 \in \mathcal{K}_1$ and $0 < t \leq \varepsilon'$. Since $H$ is bounded on the compact set $\mathcal{K}$, by homogeneity of the exponential map (cf. Lemma 8.35) the existence of $\varepsilon$ follows. $\square$

We already proved that the set of controls associated to minimizers reaching a compact is compact in the $L^2$ topology. If there are no abnormal minimizer, the compactness transfers to the set of covectors parametrizing them.

**Lemma 11.36.** *Let $q_0 \in M$ and $K \subset M$ be a compact set such that every point of $K$ is reached from $q_0$ by only strictly normal minimizers. Define the set*

$$C = \{\lambda_0 \in T^*_{q_0}M |\ \lambda_0 \text{ minimizer},\ \exp_{q_0}(\lambda_0) \in K\}.$$

*Then $C$ is compact.*

*Proof.* The set $C$ is closed since if $\lambda_n \to \lambda_0$ with $\lambda_n \in C$ then we have a sequence $\gamma_n$ of corresponding minimizers converges uniformly, and the limit curve is necessary a minimizer associated with $\lambda_0$.

It is then enough to show that $C$ is bounded. Assume by contradiction that there exists a sequence $\lambda_n \in C$ of covectors (and the associate sequence of minimizing trajectories $\gamma_n$, associated with controls $u_n$) such that $|\lambda_n| \to +\infty$, where $|\cdot|$ is some norm in $T^*_{q_0}M$. Since these minimizers are normal they satisfy the relation

$$\lambda_n D_{u_n} E_{q_0} = u_n, \qquad \forall\, n \in \mathbb{N}. \tag{11.27}$$

and dividing by $|\lambda_n|$ one obtains the identity

$$\frac{\lambda_n}{|\lambda_n|} D_{u_n} E_{q_0} = \frac{u_n}{|\lambda_n|}, \qquad \forall\, n \in \mathbb{N}. \tag{11.28}$$

Using compactness of minimizers whose endpoints stay in a compact region (cf. Theorem 8.66), we can assume that $u_n \to u$. Morever the sequence $\lambda_n/|\lambda_n|$ is bounded and we can assume that $\lambda_n/|\lambda_n| \to \lambda$ for some final covector $\lambda$. Using that $D_{u_n}E_{q_0} \to D_u E_{q_0}$ and the fact that $|\lambda_n| \to +\infty$, passing to the limit for $n \to \infty$ in (11.28) we obtain $\lambda D_u E_{q_0} = 0$. This implies in particular that the minimizers $\gamma_n$ converge to a minimizer $\gamma$ (associated to $\lambda$) that is abnormal and reaches a point of $K$ that is a contradiction. $\square$

**Theorem 11.37.** *Let $M$ be a sub-Riemannian manifold, $q_0 \in M$ and $r_0 > 0$ such that every point different from $q_0$ in the compact ball $\overline{B}_{q_0}(r_0)$ is not reached by abnormal minimizers. Then the sphere $S_{q_0}(r)$ is a Lipschitz submanifold of $M$ for almost every $r \leq r_0$.*

*Proof.* Let us fix $\delta > 0$ and consider the annulus $A_\delta = B_{r_0}(q_0) \setminus B_\delta(q_0)$. Define the set

$$C_0 = \{\lambda_0 \in T^*_{q_0}M |\ \lambda_0 \text{ minimizer},\ \exp_{q_0}(\lambda_0) \in \overline{A}_\delta\}.$$

By Lemma 11.36 the set $C_0$ is compact.

For every $\lambda_0 \in T^*M$, let us consider the control $u_{\lambda_0}$ associated with $\gamma(t) = \exp_{q_0}(t\lambda_0)$ and denote by

$$\Phi_{\lambda_0} := (P_{0,1}^{-1})^* : T^*_{q_0}M \to T^*_{\exp_{q_0}(\lambda_0)}M,$$

the pullback of the corresponding flow associated with and $u_{\lambda_0}$, computed at $q_0$. Define the set

$$\mathcal{K} := \bigcup_{\lambda_0 \in C_0} \Phi_{\lambda_0}(C_0).$$

The set $\mathcal{K}$ is compact in $T^*M$. By applying Lemma 11.35 to $\mathcal{K}$, there exists $\varepsilon_0 > 0$ such that, defining

$$C_1 := C_0 \cap H^{-1}([0, \varepsilon_0]),$$

every $\lambda_1 \in C_1$ is a regular point for $\exp_{\pi(\lambda_1)}$. Notice that $C_1$ is also compact.

Let now $N_0$ and $N_1$ be small neighborhoods of $C_0$ and $C_1$ respectively. Define the following map

$$\Psi : N_0 \times N_1 \to M, \qquad \Psi(\lambda_0, \lambda_1) = \exp_{\exp_{q_0}(\lambda_0)}(\Phi_{\lambda_0}(\lambda_1)).$$

By construction, $\Psi$ is defined on a smooth manifold and is surjective on $A_\delta$. Moreover $\Psi$ is a submersion at every point of $A_\delta$. Indeed notice that for every fixed $\lambda_0 \in C_0$, defining $\Psi_{\lambda_0} : \lambda_1 \mapsto \Psi(\lambda_0, \lambda_1)$, then $D_0\Psi_{\lambda_0}$ is surjective.

We want to apply Theorem 11.33 to the submersion $\Psi$ and the scalar function

$$\mathcal{H} : N_0 \times N_1 \to \mathbb{R}, \qquad \mathcal{H}(\lambda_0, \lambda_1) = H(\lambda_0) + H(\lambda_1).$$

Let us show that the assumptions of Theorem 11.33 are satisfied. We have to show that the set

$$N_q = \{(\lambda_0, \lambda_1) \in N_0 \times N_1 \,|\, \mathcal{H}(\lambda_0, \lambda_1) = \min_{\Psi(\lambda_0, \lambda_1) = q} \mathcal{H}(\lambda_0, \lambda_1)\}, \qquad \forall\, q \in \overline{A}_\delta,$$

is non empty and compact. Let us first notice that

$$\Psi(\lambda_0, s\lambda_0) = \exp_{q_0}((1+s)\lambda_0), \qquad \mathcal{H}(\lambda_0, s\lambda_0) = (1+s^2)H(\lambda_0).$$

By definition of $N_0$, for each $q \in \overline{A}_\delta$ there exists $\bar{\lambda}_0 \in N_0$ such that $\exp_{q_0}(\bar{\lambda}_0) = q$ and such that the corresponding trajectory is a minimizer. Moreover we can always write this unique minimizer as the union of two minimizers. It follows that

$$\min_{\Psi(\lambda_0, \lambda_1) = q} \mathcal{H}(\lambda_0, \lambda_1) = \min_{\exp_{q_0}(\lambda_0) = q} H(\lambda_0) = \mathsf{f}(q), \qquad \forall\, q \in A_\delta.$$

This implies that $N_q$ is non empty for every $q$. Moreover one can show that $N_q$ is compact. By applying Theorem 11.33 one gets that the function

$$a(q) = \min_{\Psi(\lambda_0, \lambda_1) = q} \mathcal{H}(\lambda_0, \lambda_1) = \mathsf{f}(q),$$

is locally Lipschitz in $A_\delta$ and the set of its critical values has measure zero in $A_\delta$. Since $\delta > 0$ is arbitrary we let $\delta \to 0$ and we have that $\mathsf{f}$ is locally Lipschitz in $B_{q_0}(r_0) \setminus \{q_0\}$ and the set of its critical values has measure zero. In particular almost every $r \leq r_0$ is a regular value for $\mathsf{f}$. Then, applying Corollary 11.30, the sphere $\mathsf{f}^{-1}(r^2/2)$ is a Lipschitz submanifold for almost every $r \leq r_0$. $\square$

## 11.4 Geodesic completeness and Hopf-Rinow theorem

In this section we prove a sub-Riemannian version of the Hopf-Rinow theorem. Namely, in absence of abnormal minimizers, the geodesic completeness of $M$ implies the completeness of $M$ as a metric space.

**Theorem 11.38** (sub-Riemannian Hopf-Rinow)**.** *Let $M$ be a sub-Riemannian manifold that does not admit abnormal length minimizers. If there exists a point $x \in M$ such that the exponential map $\exp_x$ is defined on the whole $T_x^*M$, then $M$ is complete with respect to the sub-Riemannian distance.*

*Proof.* Given $x \in M$, let us define

$$A := \{r > 0 \mid \overline{B}(x,r) \text{ is compact }\}, \qquad R := \sup A. \tag{11.29}$$

Arguing as in the proof of Theorem 3.47, one can show that $A \neq \emptyset$ and that $A$ is open (by using the local compactness of the topology and repeating the proof of (ii.a)). Assume now by contradiction that $R < +\infty$ and let us show that $R \in A$. By openness of $A$ this will give a contradiction and $A = ]0, +\infty[$.

We are then reduced to show that $\overline{B}(x,R)$ is compact, i.e., every sequence $\{y_i\}$ in $\overline{B}(x,R)$ admits a convergent subsequence. Define $r_i := d(y_i, x)$. It is not restrictive to assume that $r_i \to R$ (if it is not the case, then sequence is contained in a compact ball, and the existence of a convergent subsequence is clear). Since the ball $\overline{B}(x,r_i)$ is compact, by Theorem 3.43 there exists a length minimizing trajectory $\gamma_i : [0, r_i] \to M$ joining $x$ and $y_i$, parametrized by unit speed. Each curve $\gamma_i$ is normal and parametrized by length: there exist $\lambda_i \in H^{-1}(1/2) \cap T_xM$ such that

$$\gamma_i(t) = \exp_x(t\lambda_i) = \pi \circ e^{t\vec{H}}(\lambda_i).$$

By assumption we can extend each trajectory to the common interval $[0, R]$. Notice that the image of each trajectory is contained in the compact set $\overline{B}(x,R)$. Since there is no abnormal minimizer, by Lemma 11.36 the sequence $\{\lambda_i\}$ is bounded in $T_x^*M$, thus there exists a subsequence $\lambda_{i_n}$ converging to $\lambda \in H^{-1}(1/2) \cap T_xM$. Then $r_{i_n}\lambda_{i_n} \to R\lambda$ and by continuity of $\exp_x$ we have that $\{y_i\}$ has a convergent subsequence

$$y_{i_n} = \gamma_{i_n}(r_{i_n}) = \exp_x(r_{i_n}\lambda_{i_n}) \to \exp_x(R\lambda) =: y.$$

This proves that an arbitrary Cauchy sequence in $\overline{B}(x,R)$ admits a convergent subsequence. □

As an immediate corollary we have the following version of geodesic completeness theorem.

**Corollary 11.39.** *Let $M$ be a sub-Riemannian manifold that does not admit abnormal length minimizers. If the vector field $\vec{H}$ is complete on $T^*M$, then $M$ is complete with respect to the sub-Riemannian distance.*

## 11.5 Bibliographical note

Corollary 11.6 is well-known to experts, but it is not easy to find a proof in the literature.

The proof of Theorem 11.8 is an adaptation of the arguments from [Agr09, RT05]. Similar arguments have been generalized later to affine optimal control problems with quadratic cost in [BB18].

The fact whether the set of smooth points has full measure (cf. Corollary 11.13) is not known in general, and it is indeed relate to one of the main conjecture of sub-Riemannian geometry: is the set of points reached by abormal trajectories (resp. abnormal minimizers ) starting from a fixed point a subset of measure zero? This is known as the *Sard conjecture* (resp. *minimizing Sard conjecture*), since it is related to a Sard-like property for the end-point map, see the questions in [Mon02, Sec. 10.2] and [Agr14, Problem III].

This question has been investigated in detail in the three-dimensional case, first in [ZZ95] and subsequently in [BdSR18]. In the three-dimensional case case it is easier to see that the set has measure zero, and the question can be refined by asking if the above set has Hausdorff 2-dimensional measure equal to zero. Another setting where this question has been widely investigated is that of Carnot groups, see for instance [LDMO$^+$16, OV17].

Besides the very general smoothness property, in absence of abnormal minimizers some finer properties have also been proved: the squared sub-Riemannian distance is semi-concave thanks to the results of [CR08], while the sub-analiticity (of the sub-Riemannian distance, balls and spheres) and its relation to absence of abnormals have been studied in [Agr98a, Jac99, AG01a, ABCK97], see also [AG01b] and [Tré00] for a survey and some extension to control affine systems.

A comprehensive introduction to generalized gradients for locally Lipschitz functions, we refer the reader to [Cla90]. In this chapter we develop a Sard-type result for a special class of Lipschitz functions that is taken from [Rif04]. This result is applied to show that, under the assumption that there are no abnormal minimizers, almost every sphere is a Lipschitz submanifold, proved by Rifford in [Rif06].

The result contained in the final part of the chapter about geodesic completeness is a counterpart of Proposition 8.38, and it is inspired by the general arguments of [BBI01]. A discussion on geodesic completeness in the literature is already present in [Str86, Str89].

# Chapter 12

# Abnormal extremals and second variation

This chapter is devoted to the analysis of the second order conditions for length-minimizers and of abnormal extremals. After a first result on the regularity of the squared sub-Riemannian distance from a fixed point, showing that Lipschitz regularity is lost at those points reached by only abnormal length-minimizers, we move to second order conditions, known as Goh and generalized Legendre conditions.

We then specify these results to the case of rank 2 distributions, where we analyse in detail a class of abnormal extremals, called *nice*. These are smooth abnormal extremals that are solution of an autonomous Hamiltonian system, which is different from the one defining normal extremals. We show in particular that short arcs of nice abnormal extremals trajectories are length-minimizers.

In the final part of the chapter we discuss the notion of conjugate points along abnormal extremals, and the equivalence, for smooth horizontal trajectories, of the notion of length-minimality with respect to the two natural topologies one can consider on the set of the horizontal curves joining two fixed points (namely the $C^0$ and the $W^{1,2}$ topology).

We end this chapter with a proof of the fact that corners, namely horizontal trajectories that are piecewise $C^\infty$, are not length-minimizers if the trajectory is not $C^1$.

## 12.1 Second variation

Let us introduce the notion of Hessian (and second derivative) for smooth maps between manifolds, cf. also Section 8.4. We start by recalling the definition of the second differential of a map defined on a linear space.

Let $F : V \to M$ be a smooth map from a linear space $V$ on a smooth manifold $M$. The first differential of $F$ at a point $x \in V$

$$D_xF : V \to T_{F(x)}M, \qquad D_xF(v) = \left.\frac{d}{dt}\right|_{t=0} F(x+tv), \quad v \in V,$$

is a well-defined linear map, independent on the linear structure on $V$. This is not true for the second differential. The second order derivative

$$D_x^2F(v) = \left.\frac{d^2}{dt^2}\right|_{t=0} F(x+tv) \tag{12.1}$$

has not an invariant meaning if $D_xF(v) \neq 0$. Indeed in this case the curve $\gamma : (-\varepsilon, \varepsilon) \to M$ defined by $\gamma(t) = F(x + tv)$ is a smooth curve in $M$ with nonzero tangent vector. By the inverse function theorem, there exist local coordinates on $M$ such that the curve $\gamma$ is a straight line. Hence the second derivative $D_x^2F(v)$ vanish in these coordinates.

In general, the linear structure on $V$ permits to define the second differential of $F$ as a quadratic map

$$D_x^2F : \ker D_xF \to T_{F(x)}M \tag{12.2}$$

in the sense that, for $v \in \ker D_xF$, the quantity in (12.3) depends only on $v$. On the other hand, the map (12.2) is not independent on the choice of the linear structure on $V$, and this construction cannot be used if the source of $F$ is a smooth manifold.

Assume now that $F : N \to M$ is a map between smooth manifolds. The first differential is the linear map between the tangent spaces

$$D_xF : T_xN \to T_{F(x)}M, \qquad x \in N.$$

while the definition of second order derivative should be modified using smooth curves with fixed tangent vector, belonging to the kernel of $D_xF$

$$D_x^2F(v) = \left.\frac{d^2}{dt^2}\right|_{t=0} F(\gamma(t)), \qquad \gamma(0) = x, \quad \dot\gamma(0) = v \in \ker D_xF, \tag{12.3}$$

A computation in coordinates gives

$$\left.\frac{d^2}{dt^2}\right|_{t=0} F(\gamma(t)) = \frac{d^2F}{dx^2}(\dot\gamma(0), \dot\gamma(0)) + \frac{dF}{dx}\ddot\gamma(0), \tag{12.4}$$

showing that the right hand side of (12.4) is defined only mod $\operatorname{im} D_xF$.

Thus is intrinsically defined only a certain part of the second differential, which is called the *Hessian of* $F$, i.e., the quadratic map

$$\operatorname{Hess}_x F : \ker D_xF \to T_{F(x)}M/\operatorname{im} D_xF$$

## 12.2 Abnormal extremals and regularity of the distance

One of the main results of the previous chapter states that the squared sub-Riemannian distance $\mathsf{f} = \frac{1}{2}d^2(q_0, \cdot)$ from a fixed point $q_0$ is smooth on an open and dense subset $\Sigma$ of every compact ball containing $q_0$ (cf. Theorem 11.2).

The characterization of the smooth set $\Sigma$ implies in particular that $\mathsf{f}$ is not smooth at those points $q$ when there exists abnormal length-minimizers joining $q_0$ and $q$.

If, moreover, we assume that all length-minimizers joining $q_0$ and $q$ are strictly abnormal, we can prove that the squared sub-Riemannian distance it is not even Lipschitz.

**Proposition 12.1.** *Let $M$ be a sub-Riemannian manifold. Let $q_0 \in M$ and $\mathsf{f} = \frac{1}{2}d^2(q_0, \cdot)$. Assume that there are no normal length-minimizers joining $q_0$ to $\bar q$. Then $\mathsf{f}$ is not locally Lipschitz in any neighborhood of $\bar q$, namely*

$$\lim_{\substack{q \to \bar q \\ q \in \Sigma}} |d_q\mathsf{f}| = +\infty. \tag{12.5}$$

*where $|\cdot|$ denotes an arbitrary norm of the fibers of $T^*M$.*

*Proof.* Fix a compact ball $B$ containing $q_0$ and consider a sequence of smooth points $q_n \in \Sigma \cap B$ such that $q_n \to \bar{q}$. Since $q_n$ are smooth points, for every $n$ there exists a unique minimizing control $u_n$ and a corresponding unique final covector $\lambda_n$ such that the following identity holds

$$\lambda_n D_{u_n} E_{q_0} = u_n, \qquad \lambda_n = d_{q_n}\mathsf{f}.$$

Assume, by contradiction, that $|d_{q_n}\mathsf{f}| \leq M$ for some $M > 0$. Then, by compactness of length-minimizers, we may assume that the sequence of controls is convergent to a some limit $u_n \to u$. Moreover, this implies that $\lambda_n \to \lambda$ for some $\lambda \in T^*M$ satisfying $\lambda D_u E_{q_0} = u$. This implies that the corresponding length-minimizer joins $q_0$ with $\bar{q}$. In other words, we have proved that there exists a normal length-minimizer joining $q_0$ with $\bar{q}$, which is a contradiction. □

Let us now consider the end-point map $E_{q_0} : \mathcal{U} \to M$, which we recall is defined on an open subset $\mathcal{U}$ of $L^2([0,1], \mathbb{R}^m)$.

Let $u \in \mathcal{U}$ be a critical point for $E_{q_0}$. Then we can associate with it the quadratic form

$$\mathrm{Hess}_u E_{q_0} : \ker D_u E_{q_0} \to \mathrm{Coker}\, D_u E_{q_0} = T_{E_{q_0}(u)} M / \mathrm{im}\, D_u E_{q_0}.$$

*Remark* 12.2. Recall that $\lambda\, D_u E_{q_0} = 0$ if and only if $\lambda \in (\mathrm{im}\, D_u E_{q_0})^{\perp}$. Then, for every abnormal extremal, there is a well-defined scalar quadratic form

$$\lambda\, \mathrm{Hess}_u E_{q_0} : \ker D_u E_{q_0} \to \mathbb{R}.$$

Notice that the dimension of the space $(\mathrm{im}\, D_u E_{q_0})^{\perp}$ of such covectors coincides with $\dim \mathrm{Coker}\, D_u E_{q_0}$.

**Definition 12.3.** Let $Q : V \to \mathbb{R}$ be a quadratic form defined on a vector space $V$. The *index* (or *negative index*) of $Q$ is the maximal dimension of a negative subspace of $Q$:

$$\mathrm{ind}^- Q = \sup\{\dim W \mid Q\big|_{W \setminus \{0\}} < 0\}. \tag{12.6}$$

Recall that in the finite-dimensional case, the negative index coincides with the number of negative eigenvalues in the diagonal form of $Q$.

The following notion of *index* of a map will be also useful:

**Definition 12.4.** Let $F : \mathcal{U} \to M$ be a map defined on an open subset of a Hilbert space $\mathcal{H}$, and let $u \in \mathcal{U}$ be a critical point for $F$. The *index* of $F$ at $u$ is

$$\mathrm{Ind}_u F = \min_{\substack{\lambda \in \mathrm{im}\, D_u F^{\perp} \\ \lambda \neq 0}} \left\{ \mathrm{ind}^- (\lambda\, \mathrm{Hess}_u F) - \mathrm{codim}\, \mathrm{im}\, D_u F \right\}.$$

*Remark* 12.5. If $\mathrm{codim}\, \mathrm{im}\, D_u F = 1$, then there exists a unique (up to scalar multiplication) non zero $\lambda \perp \mathrm{im}\, D_u F$, hence $\mathrm{Ind}_u F = \mathrm{ind}^- (\lambda\, \mathrm{Hess}_u F) - 1$.

Given a control $u$ associated with an abnormal extremal trajectory, the index $\mathrm{Ind}_u E_{q_0}$ is well-defined.

**Theorem 12.6.** *Let $u \in \mathcal{U}$ be a control associated with an abnormal extremal trajectory. If $\mathrm{Ind}_u E_{q_0} \geq 1$, then $u$ is not a strictly abnormal length-minimizer.*

We need the following technical lemma, which we state without proof (see [AS04, Lemma 20.8]).

**Lemma 12.7.** *Let $Q : \mathbb{R}^N \to \mathbb{R}^n$ be a vector valued quadratic form. Assume that* $\mathrm{Ind}_0 Q \geq 0$*. Then there exists a regular point $x \in \mathbb{R}^n$ of $Q$ such that $Q(x) = 0$.*

We introduce now the definition of solid map.

**Definition 12.8.** Let $\Phi : E \to \mathbb{R}^n$ be a smooth map defined on a linear space $E$ and $r > 0$. We say that $\Phi$ is *r-solid* at a point $x \in E$ if there exist a constant $C > 0$, $\bar{\varepsilon} > 0$ and a neighborhood $U$ of $x$ such that for all $\varepsilon < \bar{\varepsilon}$ there exists $\delta(\varepsilon) > 0$ satisfying

$$B_{\widehat{\Phi}(x)}(C\varepsilon^r) \subset \widehat{\Phi}(B_x(\varepsilon)), \tag{12.7}$$

for all maps $\widehat{\Phi} \in C^0(E, \mathbb{R}^n)$ such that $\|\widehat{\Phi} - \Phi\|_{C^0(U,\mathbb{R}^n)} < \delta$.

**Exercise 12.9.** Prove that if $x$ is a regular point of $\Phi : E \to \mathbb{R}^n$, then $\Phi$ is 1-solid at $x$. (Hint: use implicit function theorem to prove that $\Phi$ satisfies (12.7) and Brower fixed point theorem to show that the same holds for some small perturbation)

The following proposition relates index and solidness.

**Proposition 12.10.** *Let $\Phi : E \to \mathbb{R}^n$ be a smooth map defined on a linear space $E$ and $x \in E$. Assume that* $\mathrm{Ind}_x \Phi \geq 0$*. Then $\Phi$ is 2-solid at $x$.*

*Proof.* We can assume that $x = 0$ and that $\Phi(0) = 0$. We divide the proof in two steps: first we prove that there exists a finite dimensional subspace $E' \subset E$ such that the restriction $\Phi\big|_{E'}$ satisfies the assumptions of the theorem. Then we prove the proposition under the assumption that $\dim E < +\infty$.

(i). Denote $k := \dim \mathrm{Coker}\, D_0\Phi$ and consider the Hessian

$$\mathrm{Hess}_0\Phi : \ker D_0\Phi \to \mathrm{Coker}\, D_0\Phi.$$

We can rewrite the assumption on the index of $\Phi$ as follows

$$\mathrm{ind}^- \lambda\, \mathrm{Hess}_0\Phi \geq k, \qquad \forall\, \lambda \in (\mathrm{im}\, D_0\Phi)^\perp \setminus \{0\}. \tag{12.8}$$

Since the property (12.8) is invariant by multiplication of the covector by a positive scalar, it is sufficient to consider unit covectors[1]

$$\lambda \in S^{k-1} \simeq \{\lambda \in (\mathrm{im}\, D_0\Phi)^\perp, |\lambda| = 1\}.$$

By definition of index, for every $\lambda \in S^{k-1}$, there exists a subspace $E_\lambda \subset E$, $\dim E_\lambda = k$ such that

$$\lambda\, \mathrm{Hess}_u\Phi\big|_{E_\lambda \setminus \{0\}} < 0.$$

By the continuity of the form with respect to $\lambda$, there exists a neighborhood $O_\lambda$ of $\lambda$ such that $E_{\lambda'} = E_\lambda$ for every $\lambda' \in O_\lambda$. By compactness, we can choose a finite covering of $S^{k-1}$ made by open subsets

$$S^{k-1} = O_{\lambda_1} \cup \ldots \cup O_{\lambda_N}.$$

[1] this can be defined by fixing an auxiliary metric on the set of covectors, otherwise one can more formally consider the quotient of $(\mathrm{im}\, D_0\Phi)^\perp$ with respect to multiplication by a positive scalar.

Then it is sufficient to introduce the finite-dimensional subspace

$$E' = \bigoplus_{j=1}^{N} E_{\lambda_j}.$$

(ii). Assume $\dim E < +\infty$ and split

$$E = E_1 \oplus E_2 \qquad E_2 := \ker D_0\Phi.$$

The Hessian is a map

$$\mathrm{Hess}_0\Phi : E_2 \to \mathbb{R}^n / D_0\Phi(E_1).$$

According to Lemma 12.7, there exists $e_2 \in E_2$ a regular point of $\mathrm{Hess}_0\Phi$, such that

$$\mathrm{Hess}_0\Phi(e_2) = 0 \qquad \Longrightarrow \qquad D_0^2\Phi(e_2) = D_0\Phi(e_1), \quad \text{for some } e_1 \in E_1.$$

Define the map $Q : E \to \mathbb{R}^n$ by the formula

$$Q(v_1 + v_2) := D_0\Phi(v_1) + \frac{1}{2}D_0^2\Phi(v_2), \qquad v = v_1 + v_2 \in E = E_1 \oplus E_2,$$

and set $e := -e_1/2 + e_2$. From our assumptions it follows that $e$ is a regular point of $Q$ and $Q(e) = 0$. In particular there exists $c > 0$ such that

$$B_0(c) \subset Q(B_0(1)),$$

and the same property holds for some perturbation of the map $Q$ (see Exercice 12.9). Consider then the map

$$\Phi_\varepsilon : v_1 + v_2 \mapsto \frac{1}{\varepsilon^2}\Phi(\varepsilon^2 v_1 + \varepsilon v_2). \tag{12.9}$$

Using that $v_2 \in \ker D_0\Phi$, we compute the Taylor expansion with respect to $\varepsilon$

$$\Phi_\varepsilon(v_1 + v_2) = Q(v_1 + v_2) + O(\varepsilon). \tag{12.10}$$

It follows that, for small $\varepsilon > 0$, the image of $\Phi_\varepsilon$ contains a ball around 0, hence

$$B_{\Phi(0)}(c\varepsilon^2) \subset \Phi(B_0(\varepsilon)). \tag{12.11}$$

and the same estimate (12.11) holds for small perturbations of $\Phi$. $\square$

Actually from the proof one has the following statement, which is a more quantitative version of 2-solidness of $\Phi$.

**Lemma 12.11.** *Under the assumptions of the Theorem 12.10, there exists $C > 0$ such that for every $\varepsilon > 0$ small enough*

$$B_{\Phi(0)}(C\varepsilon^2) \subset \Phi(B_0'(\varepsilon^2) \times B_0''(\varepsilon)), \tag{12.12}$$

*where $B'$ and $B''$ denotes the balls in $E_1$ and $E_2$ respectively.*

The key point is that, in the subspace where the differential of $\Phi$ vanishes, the ball of radius $\varepsilon$ is mapped into a ball of radius $\varepsilon^2$, while the restriction on the other subspace "preserves" the order, as the estimates (12.9) and (12.10) show. Indeed

$$\begin{aligned} B_0(c) \subset \Phi_\varepsilon(B_0(1)) &\Leftrightarrow B_0(c\varepsilon^2) \subset \Phi(\varepsilon^2 v_1 + \varepsilon v_2),\ v_i \in B_0^i(1) \\ &\Leftrightarrow B_0(c\varepsilon^2) \subset \Phi(B_0'(\varepsilon^2) \times B_0''(\varepsilon)). \end{aligned}$$

*Proof of Theorem 12.6.* We prove that if $u$ is a control associated with a strictly abnormal extremal trajectory such that $\mathrm{Ind}_u E_{q_0} \geq 1$, then $u$ cannot be associated with a length-minimizer. In this proof we denote by $E = E_{q_0}$ the end-point based at $q_0$.

To prove the claim, we consider the "extended" end-point map

$$\Phi : \mathcal{U} \to \mathbb{R} \times M, \qquad \Phi(u) = \begin{pmatrix} J(u) \\ E(u) \end{pmatrix}.$$

and we show that $\Phi$ is locally open at $u$.

Recall that $D_u J = \lambda\, D_u E$, for some $\lambda \in T_{E(u)}M$, if and only if $D_u J\big|_{\ker D_u E} = 0$ (see also Proposition 8.13). Since $u$ is strictly abnormal, it follows that

$$D_u J\big|_{\ker D_u E} \neq 0. \tag{12.13}$$

Moreover from the definition of $\Phi$ and (12.13) one has

$$\ker D_u \Phi = \ker D_u J \cap \ker D_u E, \qquad \dim(\mathrm{im}\, D_u J) = 1.$$

Moreover, a covector $\bar\lambda = (\alpha, \lambda)$ in $\mathbb{R} \times T^*_{E(u)}M$ annihilates the image of $D_u\Phi$ if and only if $\alpha = 0$ and $\lambda \in (\mathrm{im}\, D_u E)^\perp$, indeed if

$$0 = \bar\lambda D_u \Phi = \alpha D_u J + \lambda D_u E,$$

with $\alpha \neq 0$, this would imply that $u$ is also normal. In other words we proved the equality

$$(\mathrm{im}\, D_u\Phi)^\perp = \{(0, \lambda) \in \mathbb{R} \times T^*_{E(u)}M \mid \lambda \in (\mathrm{im}\, D_u E)^\perp\}. \tag{12.14}$$

Combining (12.13) and (12.14) one obtains for every $\bar\lambda = (0, \lambda) \in (\mathrm{im}\, D_u\Phi)^\perp$

$$\bar\lambda\, \mathrm{Hess}_u \Phi = \lambda\, \mathrm{Hess}_u E\big|_{\ker D_u J \cap \ker D_u E}. \tag{12.15}$$

Moreover $\mathrm{codim}\,(\mathrm{im}\, D_u\Phi) = \mathrm{codim}\,(\mathrm{im}\, D_u E)$ since $\dim(\mathrm{im}\, D_u\Phi) = \dim(\mathrm{im}\, D_u E) + 1$ by (12.13) and $D_u\Phi$ takes values in $\mathbb{R} \times T_{E(u)}M$. Then for every $\bar\lambda = (0, \lambda) \in \mathrm{im}\, D_u\Phi^\perp$

$$\begin{aligned} \mathrm{ind}^-(\bar\lambda\, \mathrm{Hess}_u\Phi) - \mathrm{codim}\,(\mathrm{im}\, D_u\Phi) &= \mathrm{ind}^-(\lambda\, \mathrm{Hess}_u E\big|_{\ker D_u J \cap \ker D_u E}) - \mathrm{codim}\,(\mathrm{im}\, D_u E) \\ &\geq \mathrm{ind}^-(\lambda\, \mathrm{Hess}_u E) - 1 - \mathrm{codim}\,(\mathrm{im}\, D_u E). \end{aligned}$$

Passing to the infimum with respect to $\bar\lambda$, one gets

$$\mathrm{Ind}_u\Phi \geq \mathrm{Ind}_u E - 1 \geq 0.$$

By Proposition 12.10, this implies that $\Phi$ is locally open at $u$. Hence $u$ cannot be associated with a length-minimizer. □

Now we prove that the sub-Riemannian squared distance is locally Lipschitz in a neighborhood of a point if the abnormal length-minimizers reaching these points have index bigger than one.

**Theorem 12.12.** *Let $K \subset B_{q_0}(r_0)$ be a compact and assume that $\mathrm{Ind}_u E_{q_0} \geq 1$ for every control $u$ associated with abnormal length-minimizers such that $E_{q_0}(u) \in K$. Then $\mathsf{f} = \frac{1}{2}d^2(q_0, \cdot)$ is Lipschitz on $K$.*

*Proof.* Recall that if there are no abnormal minimizers reaching $K$, Theorem 11.37 ensures that $\mathsf{f}$ is Lipschitz on $K$. Then, using compactness of the set of length-minimizers reaching $K$, it is sufficient to prove the estimate in the neighborhood of a point $q = E_{q_0}(u)$, where $u$ is associated with an abnormal length-minimizer.

Since $\mathrm{Ind}_u E_{q_0} \geq 1$ by assumption, Theorem 12.6 implies that every abnormal minimizer $u$ is not strictly abnormal, i.e., has also a normal lift. We have

$$\mathrm{Hess}_u E_{q_0} : \ker D_u E_{q_0} \to \mathrm{Coker}\, D_u E_{q_0}, \qquad \text{with} \qquad \mathrm{Ind}_u E_{q_0} \geq 1.$$

and, since $u$ is also normal, it follows that $D_u J = \lambda D_u E_{q_0}$ for some $\lambda \in T^*_{E_{q_0}(u)}M$, hence $\ker D_u E_{q_0} \subset \ker D_u J$. The assumption of Lemma 12.11 are satisfied, hence splitting the the space of controls as

$$L^2([0,1], \mathbb{R}^m) = E_1 \oplus E_2, \qquad E_2 := \ker D_u E_{q_0},$$

we have that there exists $C_0 > 0$ and $R > 0$ such that for $0 \leq \varepsilon < R$ we have

$$B_q(C_0\varepsilon^2) \subset E_{q_0}(\mathcal{B}_\varepsilon), \qquad \mathcal{B}_\varepsilon := \mathcal{B}'_u(\varepsilon^2) \times \mathcal{B}''_u(\varepsilon), \qquad q = E_{q_0}(u), \tag{12.16}$$

where $\mathcal{B}'_u(r)$ and $\mathcal{B}''_u(r)$ are the ball of radius $r$ in $E_1$ and $E_2$ respectively, and $B_q(r)$ is the ball of radius $r$ in coordinates on $M$.

Let us also observe that, since $J$ is smooth on $\mathcal{B}'_u(\varepsilon^2) \times \mathcal{B}''_u(\varepsilon)$, with $D_u J = 0$ on $E_2$, by Taylor expansion we can find constants $C_1, C_2 > 0$ such that for every $u' = (u'_1, u'_2) \in \mathcal{B}_\varepsilon$ one has (we write $u = (u_1, u_2)$)

$$J(u') - J(u) \leq C_1 \|u'_1 - u_1\| + C_2 \|u'_2 - u_2\|^2$$

Pick then any point $q' \in K$ such that $|q' - q| = C_0\varepsilon^2$, with $0 \leq \varepsilon < R$. Then (12.16) implies that there exists $u' = (u'_1, u'_2) \in \mathcal{B}_\varepsilon$ such that $E_{q_0}(u') = q'$. Using that $\mathsf{f}(q') \leq J(u')$ and $\mathsf{f}(q) = J(u)$, since $u$ is a minimizer, we have

$$\mathsf{f}(q') - \mathsf{f}(q) \leq J(u') - J(u) \leq C_1 \|u'_1 - u_1\| + C_2 \|u'_2 - u_2\|^2 \tag{12.17}$$

$$\leq C\varepsilon^2 = C'|q' - q| \tag{12.18}$$

where we have set $C = \max\{C_1, C_2\}$ and $C' = C/C_0$.

Since $K$ is compact, and the set of control $u$ associated with minimizers that reach the compact set $K$ is also compact, the constants $R > 0$ and $C_0, C_1, C_2$ can be chosen uniformly with respect to $q \in K$. Hence we can exchange the role of $q'$ and $q$ in the above reasoning and get

$$|\mathsf{f}(q') - \mathsf{f}(q)| \leq C'|q' - q|,$$

for every pair of points $q, q'$ such that $|q' - q| \leq C_0 R^2$. ∎

## 12.3 Goh and generalized Legendre conditions

In this section we present some necessary conditions for the index of the quadratic form along an abnormal extremal to be finite.

**Theorem 12.13.** *Let $M$ be a sub-Riemannian manifold and $f_1, \ldots, f_m$ be a generating family. Let $u$ be an abnormal minimizer and let $\lambda_1 \in T^*_{E_{q_0}(u)}M$ satisfy $\lambda_1 D_u E_{q_0} = 0$. Assume that* $\mathrm{ind}^- \lambda_1 \mathrm{Hess}_u E_{q_0} < +\infty$. *Then the following conditions are satisfied:*

*(i)* $\langle \lambda(t), [f_i, f_j](\gamma(t)) \rangle \equiv 0, \quad$ *for* $t \in [0,1], \ \forall\, i, j = 1, \ldots, m,$ *(Goh condition)*

*(ii)* $\big\langle \lambda(t), [[f_{u(t)}, f_v], f_v](\gamma(t)) \big\rangle \geq 0, \quad$ *for a.e.* $t, \ \forall\, v \in \mathbb{R}^m,$ *(Generalized Legendre condition)*

*where $\lambda(t) = e^{(t-1)\vec{H}}(\lambda_1)$ for $t \in [0,1]$ and $\gamma(t) = \pi(\lambda(t))$ are respectively the extremal and the trajectory associated with the final covector $\lambda_1$.*

We will refer to abnormal minimizers satisfying $\mathrm{ind}^- \lambda_1 \mathrm{Hess}_u E_{q_0} < +\infty$ as abnormal minimizers with finite index.

*Remark* 12.14. Notice that, in the statement of the previous theorem, if $\lambda_1$ satisfies the assumption $\lambda_1 D_u F = 0$, then also $-\lambda_1$ satisfies the same assumption. Since $\mathrm{ind}^-(-\lambda_1 \mathrm{Hess}_u F) = \mathrm{ind}^+ \lambda_1 \mathrm{Hess}_u F$, this implies that the statement holds also under the assumption $\mathrm{ind}^+ \lambda_1 \mathrm{Hess}_u F < +\infty$. Indeed, as is proved in the proof, as soon as the Goh condition is not satisfied, both the positive and the negative index of this form are infinity.

Notice that these conditions are related to the properties of the distribution of the sub-Riemannian structure and not to the metric. Indeed recall that the extremal $\lambda(t)$ is an abnormal extremal if and only if it satisfies

$$
\begin{aligned}
&\dot{\lambda}(t) = \sum_{i=1}^{m} u_i(t) \vec{h}_i(\lambda(t)) \\
&\langle \lambda(t), f_i(\gamma(t)) \rangle = 0, \qquad \forall\, i = 1, \ldots, m.
\end{aligned}
$$

Notice that the first equation is satisfied for a.e. $t \in [0,1]$, while the second one is valid for every $t \in [0,1]$, since the functions $t \mapsto \langle \lambda(t), f_i(\gamma(t)) \rangle$ are absolutely continuous. Geometrically the solution $\lambda(t)$ to the Hamiltonian equation belongs to $\mathcal{D}^{\perp}_{\gamma(t)}$. The Goh condition is equivalent to require that $\lambda(t) \in (\mathcal{D}^2_{\gamma(t)})^{\perp}$ for every $t \in [0,1]$.

**Corollary 12.15.** *Let $M$ be a complete sub-Riemannian structure, bracket generating of step 2, i.e., $\mathcal{D}^2_q = T_q M$ for all $q \in M$. Then there are no strictly abnormal length-minimizers. In particular $\mathsf{f} = \frac{1}{2} d^2(q_0, \cdot)$ is locally Lipschitz on $M$.*

*Proof.* Since $\mathcal{D}^2_q = T_q M$ implies $(\mathcal{D}^2_{\gamma(t)})^{\perp} = 0$ for every $q \in M$, no abnormal extremal can satisfy the Goh condition. Hence by Theorem 12.13 it follows that $\mathrm{Ind}_u E_{q_0} = +\infty$ for any control $u$ associated with an abnormal length-minimizer (starting from $q_0$). In particular, from Theorem 12.6 it follows that the length-minimizer cannot be strictly abnormal. Hence $\mathsf{f}$ is locally Lipschitz on $M$ by Theorem 12.12. $\square$

*Remark* 12.16. Notice that $\mathsf{f}$ is locally Lipschitz on $M$ if and only if the sub-Riemannian structure is 2-generating. Indeed if the structure is not 2-generating at a point $q_0$, then from the Ball-Box Theorem (Theorem 10.67) it follows that the squared sub-Riemannian distance from $q_0$ is not locally Lipschitz in a neighborhood of the base point $q_0$.

Before going into the proof of the Goh conditions (Theorem 12.13) we discuss the folllowing result.

**Theorem 12.17.** *Let $M$ be a sub-Riemannian manifold and fix $q_0 \in M$ such that $\mathcal{D}_{q_0} \neq \mathcal{D}^2_{q_0}$. Then for every $\varepsilon > 0$ small enough, there exists a normal extremal trajectory $\gamma$ starting from $q_0$ such that $\ell(\gamma) = \varepsilon$ and $\gamma$ is not a length-minimizer.*

Let us first discuss the strategy of the proof: fix a non zero element $\xi \in \mathcal{D}^{\perp}_{q_0} \setminus (\mathcal{D}^2_{q_0})^{\perp}$, which is non empty by assumption. We want to build an abnormal minimizing trajectory that has $\xi$ as initial covector and that is the limit of a sequence of stricly normal lenth-minimizers. In this way this abnormal would have finite index (the abnormal quadratic form would be the limit of positive ones) and then by Goh condition $\xi$ annihilates $\mathcal{D}^2_{q_0}0$, which is a contradiction.

*Proof.* Assume by contradiction that there exists $T > 0$ such that all normal extremal paths $\gamma_\lambda$ associated with initial covector $\lambda \in H^{-1}(1/2) \cap T^*_{q_0}M$ and defined on the interval $[0, T]$ are length-minimizers. Since the restriction of a length-minimizer is still a length-minimizer, up to reducing $T$ we can assume that there exists a compact set $K$ such that $\{\gamma_\lambda(T) \mid \lambda \in H^{-1}(1/2)\} \subset K$.[2]

Fix an element $\xi \in \mathcal{D}^{\perp}_{q_0} \setminus (\mathcal{D}^2_{q_0})^{\perp}$, which is non empty by assumption. Given any $\lambda_0 \in H^{-1}(1/2) \cap T^*_{q_0}M$, consider the family of normal extremal paths (and corresponding normal trajectories)

$$\lambda_s(t) = e^{t\vec{H}}(\lambda_0 + s\xi), \qquad \gamma_s(t) = \pi(\lambda_s(t)), \qquad t \in [0, T].$$

and let $u_s$ be the control associated with $\gamma_s$, and defined on $[0, T]$. Due to Theorem 11.2, there exists a positive sequence $s_n \to +\infty$ such that $q_n := \gamma_{s_n}(T)$ is a smooth point for the squared distance from $q_0$, for every $n \in \mathbb{N}$. By compactness of minimizers reaching $K$, there exists a subsequence of $s_n$, that we still denote by the same symbol, and a minimizing control $\bar{u}$ such that $u_{s_n} \to \bar{u}$, when $n \to \infty$. In particular, thanks to the characterization of smooth points (cf. Chapter 11), $\gamma_{s_n}$ is a strictly normal length-minimizer for every $n \in \mathbb{N}$.

Denote $\Phi^n_t = P^{u_{s_n}}_{0,t}$ the non autonomous flow generated by the control $u_{s_n}$. The family $\lambda_{s_n}(t)$ satisfies

$$\lambda_{s_n}(t) = e^{t\vec{H}}(\lambda_0 + s_n\xi) = (\Phi^n_t)^*(\lambda_0 + s_n\xi).$$

Moreover, by continuity of the flow with respect to the convergence of controls, we have that $\Phi^n_t \to \Phi_t$ for $n \to \infty$, where $\Phi_t$ denotes the flow associated with the control $\bar{u}$. Hence we have that the rescaled family

$$\frac{1}{s_n}\lambda_{s_n}(t) = (\Phi^n_t)^* \left(\frac{1}{s_n}\lambda_0 + \xi\right)$$

converges for $n \to \infty$ to the limit extremal $\bar{\lambda}(t) = \Phi^*_t\xi$. Notice that $\bar{\lambda}(t)$ is, by construction, an abnormal extremal associated with the minimizing control $\bar{u}$, and with initial covector $\xi$.

[2]it is enough to fix an arbitrary compact $K$ with $q_0 \in \mathrm{int}(K)$ such that the corresponding $\delta_K$ defined by Lemma 3.36 is smaller than $T$.

The fact that $u_{s_n}$ is a strictly normal minimizer says that the Hessian of the energy $J$ restricted to the level set $E_{q_0}^{-1}(q_n)$ is non negative. Indeed, recall that

$$\mathrm{Hess}_u J|_{E_{q_0}^{-1}(q)} = I - \lambda_1 D_u^2 E_{q_0},$$

where $\lambda_1 \in T_{E_{q_0}}(u)M$ is the final covector of the extremal lift. In particular we have, for every $n \in \mathbb{N}$ and every control $v$, the following inequality

$$\|v\|^2 - \lambda_{s_n}(T) D_{u_{s_n}}^2 E_{q_0}(v,v) \geq 0.$$

This implies

$$\frac{1}{s_n}\|v\|^2 - \frac{1}{s_n}\lambda_{s_n}(T) D_{u_{s_n}}^2 E_{q_0}(v,v) \geq 0,$$

and passing to the limit for $n \to \infty$ (recall that $s_n \to +\infty$) one gets

$$-\bar{\lambda}(T) D_{\bar{u}}^2 E_{q_0}(v,v) \geq 0.$$

In particular one has that

$$\mathrm{ind}^+ \bar{\lambda}(T) \mathrm{Hess}_{\bar{u}} E_{q_0} = \mathrm{ind}^-(-\bar{\lambda}(T) D_{\bar{u}}^2 E_{q_0}) = 0.$$

Hence the abnormal extremal has finite (positive) index and we can apply Goh conditions (see Theorem 12.13 and Remark 12.14). Thus $\xi$ annihilates $\mathcal{D}_{q_0}^2$, which is a contradiction since $\xi \in \mathcal{D}_{q_0}^{\perp} \setminus (\mathcal{D}_{q_0}^2)^{\perp}$. □

*Remark* 12.18 (About the assumption of Theorem 12.17). Assume that the sub-Riemannian structure is bracket-generating and is not Riemannian in an open set $O \subset M$, i.e., $\mathcal{D}_{q_0} \neq T_{q_0}M$ for every $q \in O$. Then there exists a dense set $D \subset O$ such that $\mathcal{D}_{q_0} \neq \mathcal{D}_{q_0}^2$ for every $q \in D$.

Indeed assume that $\mathcal{D}_q = \mathcal{D}_q^2$ for all $q$ in an open set $A$, then it is easy to see that $\mathcal{D}_q^i = \mathcal{D}_q \neq T_qM$ for all $q \in A$ and $i \geq 1$. Hence the structure is not bracket-generating on $A$, which gives a contradiction.

### 12.3.1 Proof of Goh condition - (i) of Theorem 12.13

*Proof of Theorem 12.13.* Denote by $u$ the abnormal control and by $P_{0,t}^u = \overrightarrow{\exp}\int_0^t f_{u(s)} ds$ the nonautonomous flow generated by $u$. Following the argument used in the proof of Proposition 8.5 we can write the end-point map as the composition

$$E(u+v) = P_{0,1}^u(G(v)), \qquad D_uE = (P_{0,1}^u)_* \circ D_0G,$$

and reduced the problem to the expansion of $G$, which is easier. Indeed denoting $g_i^t := (P_{0,t}^u)_*^{-1} f_i$, the map $G$ can be interpreted as the end-point map for the system

$$\dot{q}(t) = g_{v(t)}^t(q(t)) = \sum_{i=1}^m v_i(t) g_i^t(q(t)),$$

and the Hessian of $E_{q_0}$ can be computed easily starting from the Hessian of $G$ at $v = 0$ (notice that $\ker D_u E_{q_0} = \ker D_0 G$)

$$\mathrm{Hess}_u E_{q_0} = (P_{0,1}^u)_* \mathrm{Hess}_0 G,$$

from which we get, using that $\lambda_0 = (P^u_{0,1})^* \lambda_1$,

$$\lambda_1 \mathrm{Hess}_u E_{q_0} = \lambda_1 (P^u_{0,1})_* \mathrm{Hess}_0 G = \lambda_0 \mathrm{Hess}_0 G.$$

Moreover computing

$$\begin{aligned} \langle \lambda(t), [f_i, f_j](\gamma(t)) \rangle &= \left\langle \lambda_0, (P^u_{0,t})^{-1}_* [f_i, f_j](\gamma(t)) \right\rangle \\ &= \left\langle \lambda_0, [g^t_i, g^t_j](\gamma(0)) \right\rangle, \end{aligned}$$

the Goh and generalized Legendre conditions can also be rewritten as

$$\left\langle \lambda_0, [g^t_i, g^t_j](\gamma(0)) \right\rangle \equiv 0, \qquad \text{for a.e. } t \in [0,1], \quad \forall\, i, j = 1, \ldots, m, \tag{G.1}$$

$$\langle \lambda_0, [[g^t_{u(t)}, g^t_i], g^t_i]](\gamma(0)) \rangle \geq 0, \qquad \text{for a.e. } t \in [0,1], \quad \forall\, i = 1, \ldots, m. \tag{L.1}$$

Now we want to compute the Hessian of the map $G$. Using the Volterra expansion computed in Chapter 6 we have

$$G(v(\cdot)) = q_0 \circ \left( \mathrm{Id} + \int_0^1 g^t_{v(t)} dt + \iint\limits_{0 \leq \tau \leq t \leq 1} g^\tau_{v(\tau)} \circ g^t_{v(t)} d\tau dt \right) + O(\|v\|^3),$$

where we used that $g^t_v$ is linear with respect to $v$ to estimate the remainder.

This expansion let us to recover immediately the linear part, i.e., the expressions for the first differential, which can be interpreted geometrically as the integral mean

$$D_0 G(v) = \int_0^1 g^t_{v(t)}(q_0) dt.$$

On the other hand the expression for the quadratic part, i.e., the second differential

$$D^2_0 G(v) = 2\, q_0 \circ \iint\limits_{0 \leq \tau \leq t \leq 1} g^\tau_{v(\tau)} \circ g^t_{v(t)} d\tau dt.$$

has not an immediate geometrical interpretation. Recall that the second differential $D^2_0 G$ is defined on the set

$$\ker D_0 G = \{ v \in L^2([0,1], \mathbb{R}^m) \mid \int_0^1 g^t_{v(t)}(q_0) dt = 0 \} \tag{12.19}$$

and, for such a $v$, $D^2_0 G(v)$ belong to the tangent space $T_{q_0} M$. Indeed, using Lemma 8.30, and that $v$ belong to the set (12.19), we can write the second derivative as

$$D^2_0 G(v) = \iint\limits_{0 \leq \tau \leq t \leq 1} [g^\tau_{v(\tau)}, g^t_{v(t)}](q_0) d\tau dt,$$

which shows that the second differential is computed by the integral mean of the commutator of the vector field $g^t_{v(t)}$ for different times. Now consider an element $\lambda_0 \in (\mathrm{im}\, D_0 G)^\perp$, i.e., that satisfies

$$\left\langle \lambda_0, g^t_v(q_0) \right\rangle = 0, \qquad \text{for a.e. } t \in [0,1], \forall\, v \in \mathbb{R}^m.$$

Then we can compute the Hessian

$$\lambda_1 \mathrm{Hess}_u E_{q_0}(v) = \lambda_0 \mathrm{Hess}_0 G(v) = \iint\limits_{0 \leq \tau \leq t \leq 1} \langle \lambda_0, [g^\tau_{v(\tau)}, g^t_{v(t)}](q_0) \rangle d\tau dt. \tag{12.20}$$

*Remark* 12.19. Denoting, for $\tau, t \in [0,1]$, by $K(\tau,t) : \mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R}$ the bilinear form

$$K(\tau,t)(v,w) = \left\langle \lambda_0, [g_v^\tau, g_w^t](q_0) \right\rangle,$$

the Goh and generalized Legendre conditions are rewritten respectively as follows:

$$K(t,t)(v,w) = 0, \qquad \forall\, v, w \in \mathbb{R}^m, \quad \text{for a.e. } t \in [0,1], \tag{$G.2$}$$

$$\left.\frac{\partial K}{\partial \tau}(\tau,t)\right|_{\tau=t}(v,v) \geq 0, \qquad \forall\, v \in \mathbb{R}^m, \quad \text{for a.e. } t \in [0,1]. \tag{$L.2$}$$

The first one easily follows from ($G.1$). Moreover, notice that $g_v^t = (P_{0,t}^u)_*^{-1} f_v$, hence the map $t \mapsto g_v^t$ is Lipschitz for every fixed $v$. By definition of $P_{0,t}^u = \overrightarrow{\exp} \int_0^t f_{u(t)} dt$, it follows that

$$\frac{\partial}{\partial t} g_v^t = [g_{u(t)}^t, g_v^t].$$

This shows that ($L.2$) is equivalent to ($L.1$).

Finally we want to express the Hessian of $G$ in Hamiltonian terms. To this end, we consider the family of functions on $T^*M$ which are linear on fibers, associated to the vector fields $g_v^t$:

$$h_v^t(\lambda) := \left\langle \lambda, g_v^t(q) \right\rangle, \qquad \lambda \in T^*M, \quad q = \pi(\lambda).$$

We define, for a fixed element $\lambda_0 \in (\operatorname{im} D_0 G)^\perp$:

$$\eta_v^t := \vec{h}_v^t(\lambda_0) \in T_{\lambda_0} T^*M. \tag{12.21}$$

Using the identities

$$\sigma_\lambda(\vec{h}_v^t, \vec{h}_w^t) = \{h_v^t, h_w^t\}(\lambda) = \left\langle \lambda, [g_v^t, g_w^t](q) \right\rangle, \qquad q = \pi(\lambda),$$

and computing at the point $\lambda_0 \in T_{q_0}^* M$ we find

$$\sigma_{\lambda_0}(\eta_v^t, \eta_w^t) = \left\langle \lambda_0, [g_v^t, g_w^t](q_0) \right\rangle.$$

Hence we get the final expression for the Hessian

$$\lambda_0 \mathrm{Hess}_0 G(v(\cdot)) = \iint\limits_{0 \leq \tau \leq t \leq 1} \sigma_{\lambda_0}(\eta_{v(\tau)}^\tau, \eta_{v(t)}^t) dt d\tau, \tag{12.22}$$

where the control $v \in \ker D_0 G$ satisfies the relation (notice that $\pi_* \eta_v^t = g_v^t(q_0)$)

$$\pi_* \int_0^1 \eta_{v(t)}^t dt = \int_0^1 \pi_* \eta_{v(t)}^t dt = 0.$$

Moreover the Goh and Legendre conditions are expressed in Hamiltonian terms as follows:

$$\sigma_{\lambda_0}(\eta_v^t, \eta_w^t) = 0, \qquad \forall\, v, w \in \mathbb{R}^m, \text{ for a.e. } t \in [0,1], \tag{$G.3$}$$

$$\sigma_{\lambda_0}(\dot{\eta}_v^t, \eta_v^t) \geq 0, \qquad \forall\, v \in \mathbb{R}^m, \text{ for a.e. } t \in [0,1]. \tag{$L.3$}$$

We are reduced to prove, under the assumption $\mathrm{ind}^{-}\lambda_0\mathrm{Hess}_0 G < +\infty$, that $(G.3)$ and $(L.3)$ hold. Actually we will prove that Goh and generalized Legendre conditions are necessary conditions for the restriction of the quadratic form to the subspace of controls in $\ker D_0G$ that are concentrated on small segments $[t, t+s]$.

In what follows we fix once for all $t \in [0,1[$. Consider an arbitrary vector control function $v : [0,1] \to \mathbb{R}^m$ with compact support in $[0,1]$ and build, for $s > 0$ small enough, the variation

$$v_s(\tau) = v\left(\frac{\tau - t}{s}\right), \qquad \operatorname{supp} v_s \subset [t, t+s]. \tag{12.23}$$

The idea is to apply the Hessian to this particular control functions and then compute the asymptotics for $s \to 0$.

Notice that the index of a quadratic form is finite if and only if the same holds for the restriction of the quadratic form to a subspace of finite codimension. Hence we can restrict to the subspace of controls of zero integral mean

$$E_s := \{v_s \in \ker D_0G \mid v_s \text{ defined by (12.23)}, \int_0^1 v(\tau)d\tau = 0\}.$$

Notice that this space depends on the choice of $s$, while $\operatorname{codim} E_s$ does not.

*Remark* 12.20. We will use the following identity (writing $\sigma$ for $\sigma_{\lambda_0}$), which holds for arbitrary control functions $v, w : [0,1] \to \mathbb{R}^m$

$$\iint\limits_{\alpha \le \tau \le t \le \beta} \sigma(\eta^{\tau}_{v(\tau)}, \eta^{t}_{w(t)})dtd\tau = \int_{\alpha}^{\beta} \sigma(\int_{\alpha}^{t} \eta^{\tau}_{v(\tau)}d\tau, \eta^{t}_{w(t)})dt = \int_{\alpha}^{\beta} \sigma(\eta^{\tau}_{v(\tau)}, \int_{\tau}^{\beta} \eta^{t}_{w(t)}dt)d\tau. \tag{12.24}$$

For the specific choice $w(t) = \int_0^t v(\tau)d\tau$ we have also the integration by parts formula

$$\int_{\alpha}^{\beta} \eta^{t}_{v(t)}dt = \eta^{\beta}_{w(\beta)} - \eta^{\alpha}_{w(\alpha)} - \int_{\alpha}^{\beta} \dot{\eta}^{t}_{w(t)}dt. \tag{12.25}$$

Combining (12.22) and (12.24), we rewrite the Hessian applied to $v_s$ as follows

$$\lambda_0\mathrm{Hess}_0 G(v_s(\cdot)) = \int_t^{t+s} \sigma(\int_t^{\tau} \eta^{\theta}_{v_s(\theta)}d\theta, \eta^{\tau}_{v_s(\tau)})d\tau. \tag{12.26}$$

Notice that the control $v_s$ is concentrated on the segment $[t, t+s]$, thus we have restricted the extrema of the integral.

The integration by parts formula (12.25), using our boundary conditions, gives

$$\int_t^{\tau} \eta^{\theta}_{v_s(\theta)}d\theta = \eta^{\tau}_{w_s(\tau)} - \int_t^{\tau} \dot{\eta}^{\theta}_{w_s(\theta)}d\theta, \tag{12.27}$$

where we defined

$$w_s(\theta) = \int_t^{\theta} v_s(\tau)d\tau, \qquad \theta \in [t, t+s].$$

Combining (12.26) and (12.27) one has

$$\lambda_0 \mathrm{Hess}_0 G(v_s(\cdot)) = \int_t^{t+s} \sigma(\eta^\tau_{w_s(\tau)}, \eta^\tau_{v_s(\tau)}) d\tau - \int_t^{t+s} \sigma(\int_t^\tau \dot\eta^\theta_{w_s(\theta)} d\theta, \eta^\tau_{v_s(\tau)}) d\tau$$
$$= \int_t^{t+s} \sigma(\eta^\tau_{w_s(\tau)}, \eta^\tau_{v_s(\tau)}) d\tau - \int_t^{t+s} \sigma(\dot\eta^\tau_{w_s(\tau)}, \int_\tau^{t+s} \eta^\theta_{v_s(\theta)} d\theta) d\tau, \tag{12.28}$$

where the second equality uses (12.24).

Next consider the second term in (12.28) and apply again the integration by parts formula (recall that $w_s(t+s) = 0$)

$$\int_t^{t+s} \sigma(\dot\eta^\tau_{w_s(\tau)}, \int_\tau^{t+s} \eta^\theta_{v_s(\theta)} d\theta) d\tau = - \int_t^{t+s} \sigma(\dot\eta^\tau_{w_s(\tau)}, \eta^\tau_{w_s(\tau)}) d\tau$$
$$- \int_t^{t+s} \sigma(\dot\eta^\tau_{w_s(\tau)}, \int_\tau^{t+s} \dot\eta^\theta_{w_s(\theta)} d\theta) d\tau.$$

Collecting together all these results one obtains

$$\lambda_0 \mathrm{Hess}_0 G(v_s(\cdot)) = \int_t^{t+s} \sigma(\eta^\tau_{w_s(\tau)}, \eta^\tau_{v_s(\tau)}) d\tau$$
$$+ \int_t^{t+s} \sigma(\dot\eta^\tau_{w_s(\tau)}, \eta^\tau_{w_s(\tau)}) d\tau$$
$$+ \int_t^{t+s} \sigma(\dot\eta^\tau_{w_s(\tau)}, \int_\tau^{t+s} \dot\eta^\theta_{w_s(\theta)} d\theta) d\tau.$$

This is indeed a homogeneous decomposition of $\lambda_0 \mathrm{Hess}_0 G(v_s(\cdot))$ with respect to $s$, in the following sense: since

$$w_s(\theta) = s\, w\left(\frac{\theta - t}{s}\right),$$

we can perform the change of variable

$$\zeta = \frac{\tau - t}{s}, \qquad \tau \in [t, t+s],$$

and obtain the following expression for the Hessian in terms of the original function $w$ defined on the interval $[0, 1]$:

$$\lambda_0 \mathrm{Hess}_0 G(v_s(\cdot)) = s^2 \int_0^1 \sigma(\eta^{t+s\theta}_{w(\theta)}, \eta^{t+s\theta}_{v(\theta)}) d\theta$$
$$+ s^3 \int_0^1 \sigma(\dot\eta^{t+s\theta}_{w(\theta)}, \eta^{t+s\theta}_{w(\theta)}) d\theta \tag{12.29}$$
$$+ s^4 \int_0^1 \sigma(\dot\eta^{t+s\theta}_{w(\theta)}, \int_\theta^1 \dot\eta^{t+s\zeta}_{w(\zeta)} d\zeta) d\theta.$$

We recall that here $v_s$ is defined through a control $v$ compactly supported in $[0, 1]$ by (12.23) and $w$ is the primitive of $v$, that is also compactly supported on $[0, 1]$. In particular we can write

$$\lambda_0 \mathrm{Hess}_0 G(v_s(\cdot)) = s^2 \int_0^1 \sigma(\eta^t_{w(\theta)}, \eta^t_{v(\theta)}) d\theta + O(s^3). \tag{12.30}$$

By assumption $\text{ind}^{-}\lambda_0\text{Hess}_0 G < +\infty$. This implies that the quadratic form given by its principal part

$$w(\cdot) \mapsto \int_0^1 \sigma(\eta^t_{w(\theta)}, \eta^t_{\dot{w}(\theta)})d\theta, \tag{12.31}$$

has also finite index. Indeed, assume that (12.31) has infinite negative index. Then by continuity every sufficiently small perturbation of (12.31) would have infinite index as well. Hence, for $s$ small enough, the quadratic form $\lambda_0\text{Hess}_0 G$ would also have infinite index, contradicting our assumption on (12.30).

To prove Goh condition, it is then sufficient to show that if (12.31) has finite index then the integrand is zero, which is guaranteed by the following

**Lemma 12.21.** *Let $A : \mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R}$ be a skew-symmetric bilinear form and define the qudratic form*

$$Q : \mathcal{U} \to \mathbb{R}, \qquad Q(w(\cdot)) = \int_0^1 A(w(t), \dot{w}(t))dt,$$

*where $\mathcal{U} := \{w(\cdot) \in \text{Lip}([0,1], \mathbb{R}^m) : w(0) = w(1) = 0\}$. Then $\text{ind}^{-}Q < +\infty$ if and only if $A = 0$.*

*Proof.* Clearly if $A = 0$, then $Q = 0$ and $\text{ind}^{-}Q = 0$. Assume then that $A \neq 0$ and let us prove that $\text{ind}^{-}Q = +\infty$. We divide the proof into steps

(i). The bilinear form $B : \mathcal{U} \times \mathcal{U} \to \mathbb{R}$ defined by

$$B(w_1(\cdot), w_2(\cdot)) = \int_0^1 A(w_1(t), \dot{w}_2(t))dt,$$

is symmetric. Indeed, integrating by parts and using the boundary conditions we get

$$\begin{aligned} B(w_1, w_2) &= \int_0^1 A(w_1(t), \dot{w}_2(t))dt \\ &= -\int_0^1 A(\dot{w}_1(t), w_2(t))dt \\ &= \int_0^1 A(w_2(t), \dot{w}_1(t))dt = B(w_2, w_1). \end{aligned}$$

(ii). $Q$ is not identically zero. Since $Q$ is the quadratic form associated to $B$ and from the polarization formula

$$B(w_1, w_2) = \frac{1}{4}(Q(w_1 + w_2) - Q(w_1 - w_2)),$$

it easily follows that $Q \equiv 0$ if and only if $B \equiv 0$. Then it is sufficient to prove that $B$ is not zero.

Since $A \neq 0$, there exists $x, y \in \mathbb{R}^m$ such that $A(x, y) \neq 0$, and consider a smooth non-constant function

$$\alpha : \mathbb{R} \to \mathbb{R}, \qquad \text{s.t.} \qquad \alpha(0) = \alpha(1) = \dot{\alpha}(0) = \dot{\alpha}(1) = 0.$$

Then $\dot{\alpha}(t)z, \alpha(t)z \in \mathcal{U}$ for every $z \in \mathbb{R}^m$ and we can compute

$$\begin{aligned} B(\dot{\alpha}(\cdot)x, \alpha(\cdot)y) &= \int_0^1 A(\dot{\alpha}(t)x, \dot{\alpha}(t)y)dt \\ &= A(x, y)\int_0^1 \dot{\alpha}(t)^2 dt \neq 0. \end{aligned}$$

(iii). $Q$ has the same number of positive and negative eigenvalues. Indeed it is easy to see that $Q$ satisfies the identity

$$Q(w(1-\cdot)) = -Q(w(\cdot)),$$

from which (iii) follows.

(iv). $Q$ is non vanishing on a infinite dimensional subspace. Consider some $w \in \mathcal{U}$ such that $Q(w) = \alpha \neq 0$. For every $x = (x_1, \ldots, x_N) \in \mathbb{R}^N$ one can build the function

$$w_x(t) = x_i\, w(Nt - i), \qquad t \in \left[\frac{i}{N}, \frac{i+1}{N}\right], \qquad i = 1, \ldots, N.$$

An easy computations shows that

$$Q(w_x) = \alpha \sum_{i=1}^{N} x_i^2.$$

In particular there exists a subspace of arbitrary large dimension where $Q$ has the same sign. □

### 12.3.2 Proof of generalized Legendre condition - (ii) of Theorem 12.13

Applying Lemma 12.21 for any $t$ we prove that the $s^2$ order term in (12.29) vanish and we get to

$$\begin{aligned}
\lambda_0 \mathrm{Hess}_0 G(v(\cdot)) &= s^3 \int_0^1 \sigma(\dot{\eta}^{t+s\theta}_{w(\theta)}, \eta^{t+s\theta}_{w(\theta)}) d\theta + O(s^4) \\
&= s^3 \int_0^1 \sigma(\dot{\eta}^{t+s\theta}_{w(\theta)}, \eta^{t}_{w(\theta)}) d\theta + O(s^4)
\end{aligned}$$

where the last equalily follows from the fact that $\eta^t_v$ is Lipschitz with respect to $t$ (see also (12.21)), i.e.,

$$\eta^{t+s\theta}_v = \eta^t_v + O(s).$$

*Remark* 12.22. Notice that the quantity $\dot{\eta}^t_v$ is only measurable bounded in $t$. On the other hand we have

$$\eta^t_v = \vec{h}^t_v(\lambda_0), \qquad h^t_v(\lambda) = \sum_{i=1}^{m} v_i \left\langle \lambda, (P^u_{0,t})^{-1}_* f_i \right\rangle.$$

hence the set of Lebesgue points of the control $u$ is contained in the set of Lebesgue points of $\dot{\eta}^t_v$, for every $v$ (here $u$ is the control where we compute the Hessian of the end-point map).

If $t$ is a Lebesgue point of $t \mapsto \dot{\eta}^t_v$, the quantity $\dot{\eta}^t_{w(\cdot)}$ is well-defined and we can write

$$\begin{aligned}
\lambda_0 \mathrm{Hess}_0 G(v(\cdot)) = s^3 &\int_0^1 \sigma(\dot{\eta}^t_{w(\theta)}, \eta^t_{w(\theta)}) d\theta \\
&- s^3 \left( \int_0^1 \sigma(\dot{\eta}^{t+s\theta}_{w(\theta)}, \eta^t_{w(\theta)}) - \sigma(\dot{\eta}^t_{w(\theta)}, \eta^t_{w(\theta)}) d\theta \right) + O(s^4).
\end{aligned}$$

Using the linearity of $\sigma$ and the boundedness of the vector fields we can estimate

$$\begin{aligned}
\left| \int_0^1 \sigma(\dot{\eta}^{t+s\theta}_{w(\theta)}, \eta^t_{w(\theta)}) - \sigma(\dot{\eta}^t_{w(\theta)}, \eta^t_{w(\theta)}) d\theta \right| &\leq C \int_0^1 |\dot{\eta}^{t+s\theta}_{w(\theta)} - \dot{\eta}^t_{w(\theta)}| d\theta \\
&\leq C \sup_{|v|\leq 1} \frac{1}{s} \int_0^s |\dot{\eta}^{t+\tau}_v - \dot{\eta}^t_v| d\tau \underset{s\to 0}{\longrightarrow} 0,
\end{aligned}$$

where the last term tends to zero by definition of Lebesgue point (this set is independent on $v$, cf. Remark 12.22). Hence we get

$$\lambda_0 \mathrm{Hess}_0 G(v(\cdot)) = s^3 \int_0^1 \sigma(\dot{\eta}^t_{w(\theta)}, \eta^t_{w(\theta)}) d\theta + o(s^3). \tag{12.32}$$

$\square$

To prove the generalized Legendre condition we have to prove that the integrand is a non negative quadratic form. This follows from the following lemma, which can be proved similarly to Lemma 12.21.

**Lemma 12.23.** *Let $Q : \mathbb{R}^m \to \mathbb{R}$ be a quadratic form on $\mathbb{R}^m$ and*

$$\mathcal{U} := \{w(\cdot) \in \mathrm{Lip}([0,1], \mathbb{R}^m) \mid w(0) = w(1) = 0\}.$$

*The quadratic form*

$$\mathcal{Q} : \mathcal{U} \to \mathbb{R}, \qquad \mathcal{Q}(w(\cdot)) = \int_0^1 Q(w(t)) dt,$$

*has finite index if and only if $Q$ is non negative.*

### 12.3.3 More on Goh and generalized Legendre conditions

If Goh condition is satisfied, the generalized Legendre condition can also be characterized as an intrinsic property of the module. Indeed one can see that the quadratic map defined on the fibers $U_q$ of the control bundle $\mathbf{U}$ defining the sub-Riemannian structure

$$U_{\gamma(t)} \to \mathbb{R}, \qquad v \mapsto \left\langle \lambda(t), [[f_{u(t)}, f_v], f_v](\gamma(t)) \right\rangle,$$

is well-defined and does not depend on the extension of $f_v$ to a vector field $f_{v(t)}$ on the vector bundle $\mathbf{U}$ (or, more concretely, on the choice of the generating family).

Let us introduce the notation $h_v(\lambda) = \langle \lambda, f_v(q) \rangle$. An abnormal extremal $\lambda(t)$ satisfies

$$h_v(\lambda(t)) \equiv 0, \qquad \forall\, v \in \mathbb{R}^m.$$

Recall that the Poisson bracket between linear functions on $T^*M$ is computed in terms of the Lie bracket

$$\{h_v, h_w\}(\lambda) = \langle \lambda, [f_v, f_w](q) \rangle,$$

We can reformulate Goh and generalized Legendre condition in Hamiltonian terms as follows.

**Theorem 12.24.** *Let $M$ be a sub-Riemannian manifold and let $\lambda(t)$ be an abnormal extremal with finite index, whose corresponding trajectory is a length-minimizer. Then the following conditions are satisfied:*

*(i)* $\{h_v, h_w\}(\lambda(t)) \equiv 0, \qquad for\ t \in [0,1], \forall\, v, w \in \mathbb{R}^m,$ *(Goh condition)*

*(ii)* $\{\{h_{u(t)}, h_v\}, h_v\}(\lambda(t)) \geq 0, \qquad a.e.\ t \in [0,1], \forall\, v \in \mathbb{R}^m.$ *(Generalized Legendre condition)*

**Exercise 12.25.** Use the Jacobi identity of the Poisson bracket to show that the bilinear form

$$B_t : \mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R}, \qquad B_t(v, w) := \{\{h_{u(t)}, h_v\}, h_w\}(\lambda), \tag{12.33}$$

is well-defined and symmetric.

The generalized Legendre condition in Theorem (12.24) says that the quadratic form associated to the bilinear form $B_t$ given in (12.33) is nonnegative.

Next we want to characterize in Hamiltonian terms the trajectories that satisfy these conditions. Let $\lambda(t)$ be an abnormal extremal and let $u(t)$ be the associated control. Then

$$\dot{\lambda}(t) = \vec{h}_{u(t)}(\lambda(t)), \qquad h_i(\lambda(t)) \equiv 0, \quad 0 \le t \le 1. \tag{12.34}$$

where $\vec{h}_{u(t)} = \sum_{i=1}^m u_i(t)\vec{h}_i(t)$. Let us denote the iterated Poisson brackets as follows

$$h_{i_1\dots i_k}(\lambda) = \{h_{i_1}, \dots, \{h_{i_{k-1}}, h_{i_k}\}\}(\lambda) \tag{12.35}$$
$$= \left\langle \lambda, [f_{i_1}, \dots, [f_{i_{k-1}}, f_{i_k}]](q) \right\rangle, \qquad q = \pi(\lambda) \tag{12.36}$$

Recall that for any smooth function $a : T^*M \to \mathbb{R}$

$$\frac{d}{dt} a(\lambda(t)) = \{h_{u(t)}, a\}(\lambda(t)) = \sum_{i=1}^m u_i(t)\{h_i, a\}(\lambda(t)).$$

**Proposition 12.26.** *Let $M$ be a sub-Riemannian structure and let $\lambda(t)$ be an abnormal extremal with finite index, whose corresponding trajectory is a length-minimizer associated with the control $u(t)$. Then the pair $(u(t), \lambda(t))$ satisfies the system*

$$\sum_{\ell=1}^m u_\ell(t) h_{\ell i j}(\lambda(t)) = 0, \tag{12.37}$$

*for a.e. $t \in [0, 1]$ and every $i, j = 1, \dots, m$.*

*Proof.* The trajectory satisfies the Goh conditions, hence for every $t \in [0, 1]$ and every $i, j = 1, \dots, m$, one has

$$h_{ji}(\lambda(t)) = 0. \tag{12.38}$$

If we differentiate equations (12.38) with respect to $t$, we find

$$\sum_{\ell=1}^m u_\ell(t) h_{\ell i j}(\lambda(t)) = 0. \tag{12.39}$$

for a.e. $t \in [0, 1]$ and every $i, j = 1, \dots, m$. □

*Remark* 12.27. Notice that, without requiring a priori that the abnormal extremal satisfies the Goh condition, differentiating the identities $h_i(\lambda(t)) = 0$ for every $i = 1, \dots, m$, one finds the set of $m(m-1)/2$ equations in the $m$ variables $u_1, \dots, u_m$.

$$\sum_{j=1}^m u_j(t) h_{ij}(\lambda(t)) = 0. \tag{12.40}$$

When $m$ is odd, one always has a non-constant solution of the system. When $m$ is even, this is possible only for those $\lambda$ such that $\det(h_{ij}(\lambda))_{i,j} \neq 0$. This is never the case when Goh conditions, are satisfied since the matrix is identically zero. Hence one cannot solve the the linear system (12.40) with respect to $u$, and has to differentiate once more to arrive to (12.39). About the last set of equations, notice that

(i) If $m = 2$, then (12.39) defines 1 equation in 2 variables, and we can recover the control $u_1, u_2$ up to a scalar multiplier, when at least one of the coefficients does not vanish. Since we can always deal with length-parametrized minimizers, this uniquely determine the control $u$.

(ii) If $m \geq 3$, then the system is always overdetermined (in the sense that the number of variables is bigger than the number of equations).

## 12.4 Rank 2 distributions and nice abnormal extremals

Consider a rank 2 distribution generated by a local generating family $f_1, f_2$ and let $h_1, h_2$ be the associated linear Hamiltonians. An abnormal extremal $\lambda(t)$ associated with a control $u(t)$ satisfies the system of equations

$$\begin{aligned} \dot{\lambda}(t) &= u_1(t)\vec{h}_1(\lambda(t)) + u_2(t)\vec{h}_2(\lambda(t)), \\ h_1(\lambda(t)) &= h_2(\lambda(t)) = 0. \end{aligned} \tag{12.41}$$

In this case the Goh condition is automatically satisfied by every abnormal extremal.

**Lemma 12.28.** *Every non-constant abnormal extremal on a rank 2 sub-Riemannian structure satisfies the Goh condition.*

*Proof.* Let $\lambda(t)$, for $t \in [0, 1]$, be an abnormal extremal on a rank 2 sub-Riemannian structure. Define the linear Hamiltonian associated with their Lie bracket $h_{12}(\lambda(t)) = \langle \lambda, [f_1, f_2](q) \rangle$. Notice that, in this special framework, the Goh condition $\lambda(t) \in (\mathcal{D}^2)^\perp$ is equivalent to

$$h_1(\lambda(t)) = h_2(\lambda(t)) = h_{12}(\lambda(t)) = 0, \qquad \text{for } t \in [0, 1]. \tag{12.42}$$

The first two identities are satisfied since $\lambda(t) \in \mathcal{D}^\perp$, hence

$$h_1(\lambda(t)) = h_2(\lambda(t)), \qquad \text{for } t \in [0, 1].$$

Differentiating these identities one gets (we omit $t$ in the notation for simplicity)

$$\begin{aligned} u_2\{h_2, h_1\} &= -u_2 h_{12}(\lambda) = 0, \\ u_1\{h_1, h_2\} &= u_1 h_{12}(\lambda) = 0. \end{aligned}$$

Since at least one among the controls $u_1$ and $u_2$ is not identically zero (the trajectory is non-constant), we have that $h_{12}(\lambda(t)) \equiv 0$ for every $t \in [0, 1]$, that is (12.42). ☐

In what follows we focus on a special class of abnormal extremals.

**Definition 12.29.** An abnormal extremal $\lambda(t)$ is called *nice* if, for every $t \in [0, 1]$, it satisfies

$$\lambda(t) \in (\mathcal{D}^2)^\perp \setminus (\mathcal{D}^3)^\perp.$$

*Remark* 12.30. Notice that, as soon as the distribution is bracket-generating and has rank 2 on a manifold with $\dim M > 3$, the set $(\mathcal{D}_q^2)^\perp \setminus (\mathcal{D}_q^3)^\perp$ is nonempty for an open dense set of points $q \in M$. Indeed assume that we have $\mathcal{D}_q^2 = \mathcal{D}_q^3$ for any $q$ in a open neighborhood $O_{q_0}$ of a point $q_0$ in $M$. Then it follows that

$$\mathcal{D}_{q_0}^2 = \mathcal{D}_{q_0}^i$$

for every $i > 1$. Hence structure cannot be bracket generating, since $\dim \mathcal{D}_{q_0}^i \leq 3 < \dim M$. The case $n = 3$ is discussed later, see 12.6.1.

**Theorem 12.31.** *Let $\lambda(t)$, for $t \in [0,1]$, be an abnormal extremal on a rank 2 sub-Riemannian structure. Then $\lambda(t)$ is a nice abnormal extremal if and only if it is a reparametrization of a solution of*

$$\dot{\lambda}(t) = \vec{H}_0(\lambda(t)), \qquad \lambda(0) \in (\mathcal{D}^2)^\perp \setminus (\mathcal{D}^3)^\perp, \tag{12.43}$$

*where $H_0 : T^*M \to \mathbb{R}$ is the smooth function defined by*

$$H_0 = h_{221}h_1 + h_{112}h_2. \tag{12.44}$$

*Proof.* Recall that, by Lemma 12.28, the Goh condition is automatically satisfied, hence every abnormal extremal satisfies $\lambda(t) \in (\mathcal{D}^2)^\perp$, hence

$$h_1(\lambda(t)) = h_2(\lambda(t)) = h_{12}(\lambda(t)) = 0, \qquad \text{for every } t \in [0,1]. \tag{12.45}$$

Notice moreover that, on the subset $\{h_1 = h_2 = 0\} \subset T^*M$, we have

$$\vec{H}_0 = h_{221}\vec{h}_1 + h_{112}\vec{h}_2. \tag{12.46}$$

Assume first that $\lambda(t)$ is a nice abnormal extremal. Differentiating twice the last equation in (12.45) one obtains the identity

$$u_1(t)h_{112}(\lambda(t)) = u_2(t)h_{221}(\lambda(t)). \tag{12.47}$$

If the abnormal is nice, then $(h_{112}(\lambda(t)), h_{221}(\lambda(t)) \neq (0,0)$, and we can uniquely recover the control $u = (u_1, u_2)$ up to a scalar as follows

$$u_1(t) = h_{221}(\lambda(t)), \qquad u_2(t) = h_{112}(\lambda(t)). \tag{12.48}$$

If we plug this control into the original equation, we find that $\lambda(t)$ is a solution of

$$\dot{\lambda} = h_{221}(\lambda)\vec{h}_1(\lambda) + h_{112}(\lambda)\vec{h}_2(\lambda). \tag{12.49}$$

Moreover, since $\lambda(t)$ is nice, we have $\lambda(0) \in (\mathcal{D}^2)^\perp \setminus (\mathcal{D}^3)^\perp$.

It remains to prove that every solution to the equation

$$\dot{\lambda}(t) = \vec{H}_0(\lambda(t)), \qquad \lambda_0 \in (\mathcal{D}^2)^\perp \setminus (\mathcal{D}^3)^\perp, \tag{12.50}$$

satisfies $\lambda(t) \in (\mathcal{D}^2)^\perp \setminus (\mathcal{D}^3)^\perp$ for every $t \in [0,1]$. First notice that a non-constant solution cannot intersect the set $(\mathcal{D}^3)^\perp$ since these are equilibrium points of the system (12.50) (at those points the Hamiltonian has a root of order two).

We are reduced to prove that $(\mathcal{D}^2)^\perp$ is an invariant subset for $\vec{H}_0$. We show that the three functions $h_1, h_2, h_{12}$ are constantly zero when computed on the extremal. Let us write the differential equations satisfied by these functions. Recall that, for any smooth function $a : T^*M \to \mathbb{R}$ and any solution of the Hamiltonian system $\lambda(t) = e^{t\vec{H}}\lambda_0$, we have $\dot{a} = \{H, a\}$. Hence we get

$$\begin{aligned}
\dot{h}_{12} &= \{h_{221}h_1 + h_{112}h_2, h_{12}\} \\
&= \{h_{221}, h_{12}\}h_1 + \{h_{112}, h_{12}\}h_2 + \underbrace{h_{112}h_{221} + h_{212}h_{112}}_{=0} \\
&= c_1h_1 + c_2h_2
\end{aligned}$$

for some smooth coefficients $c_1$ and $c_2$. Similarly, we see that there exist smooth functions $a_1, a_2, a_{12}$ and $b_1, b_2, b_{12}$ such that

$$\begin{cases}
\dot{h}_1 = a_1h_1 + a_2h_2 + a_{12}h_{12} \\
\dot{h}_2 = b_1h_1 + b_2h_2 + b_{12}h_{12} \\
\dot{h}_{12} = c_1h_1 + c_2h_2
\end{cases} \tag{12.51}$$

If we plug the solution $\lambda(t)$ into the equation of (12.50), i.e., if we consider it as a system of differential equations for the scalar functions $h_i(t) := h_i(\lambda(t))$, with variable coefficients $a_i(\lambda(t)), b_i(\lambda(t)), c_i(\lambda(t))$, we find that $h_1(t), h_2(t), h_{12}(t)$ satisfy a nonautonomous homogeneous linear system of differential equation with zero initial condition, since $\lambda_0 \in (\mathcal{D}^2)^\perp$, i.e.

$$h_1(\lambda_0) = h_2(\lambda_0) = h_{12}(\lambda_0) = 0. \tag{12.52}$$

Hence by uniqueness we have

$$h_1(\lambda(t)) = h_2(\lambda(t)) = h_{12}(\lambda(t)) = 0, \qquad \forall\, t \in [0, 1].$$

□

*Remark* 12.32. Notice that from the proof it follows that the control $u(t)$ associated to a nice abnormal extremal is smooth.

We also prove that nice abnormals satisfy the generalized Legendre condition. Recall that if $\lambda(t)$ is an abnormal extremal, then $-\lambda(t)$ is also an abnormal extremal.

**Lemma 12.33.** *Let $\lambda(t)$ be a nice abnormal extremal. Then $\lambda(t)$ or $-\lambda(t)$ satisfy the generalized Legendre condition.*

*Proof.* Let $u(t)$ be the control associated with the extremal $\lambda(t)$. It is sufficient to prove that the quadratic form

$$Q_t : \mathbb{R}^2 \to \mathbb{R}, \qquad Q_t(v) = \left\langle \lambda(t), [[f_{u(t)}, f_v], f_v] \right\rangle, \tag{12.53}$$

is non-negative. We already proved (cf. Exercice 12.25) that this is the quadratic form associated with the symmetric bilinear form

$$B_t : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}, \qquad B_t(v, w) = \left\langle \lambda(t), [[f_{u(t)}, f_v], f_w] \right\rangle. \tag{12.54}$$

From the explicit expression (12.54) it is easy to see that $u(t) \in \ker B_t$ for every $t \in [0, 1]$. Hence $Q_t$ is degenerate quadratic form, for every $t \in [0, 1]$. On the other hand, the quadratic form is not identically zero. Indeed in this case we have $\lambda(t) \in (\mathcal{D}^3)^\perp$, which is a contradiction.

Hence we have proved that the quadratic form has rank 1 and is semi-definite, we can then choose the sign in $\pm\lambda_0$ in such a way that (12.53) is positive at $t = 0$. Since the quadratic form is continuous with respect to $t$ (the control $u(t)$ is continuous, cf. Remark 12.32) and cannot vanish along the curve, then it is positive for all $t \in [0, 1]$. □

## 12.5 Minimality of nice abnormal in rank 2 structures

Up to now we proved that every nice abnormal extremal in a rank 2 sub-Riemannian structure automatically satisfies the necessary condition for optimality. Now we prove that actually they are strict local minimizers.

**Theorem 12.34.** *Let $\lambda(t)$ be a nice abnormal extremal, defined for $t \in [0,1]$, and let $\gamma(t)$ be the corresponding abnormal extremal trajectory. Then there exists $s > 0$ such that $\gamma|_{[0,s]}$ is a strict local length-minimizer in the the $W^{1,2}$ topology for horizontal trajectories joining the same endpoints.*

*Remark* 12.35. Recall that being a nice abnormal extremal trajectory is a property that is independent on the metric, and depends only on the distribution. In particular it turns out that the value of $s$ given in Theorem (12.34) is independent on the metric structure chosen on the distribution.

With similar arguments than those used in the proof of Theorem 4.65 (cf. Section 4.7) one can prove that, as soon as the metric is fixed, short arcs of nice abnormal are also global length-minimizers.

We now prove the following result, which should highlight the strategy we develop later to prove Theorem 12.34.

**Lemma 12.36.** *Let $\Phi : E \to \mathbb{R}^n$ be a smooth map defined on a Hilbert space $E$ such that $\Phi(0) = 0$, where $0$ is a critical point for $\Phi$*

$$\lambda\, D_0\Phi = 0, \qquad \lambda \in \mathbb{R}^{n*},\ \lambda \neq 0.$$

*Assume that $\lambda\,\mathrm{Hess}_0\Phi$ is a positive definite quadratic form. Then for every $v$ such that $\langle \lambda, v\rangle < 0$, there exists a neighborhood of zero $O \subset E$ such that*

$$\Phi(x) \notin \mathbb{R}^+ v, \qquad \forall\, x \in O, x \neq 0, \qquad \mathbb{R}^+ = \{\alpha \in \mathbb{R} \mid \alpha > 0\}.$$

*In particular the map $\Phi$ is not locally open, and $x = 0$ is an isolated point on its level set.*

*Proof.* In the first part of the proof we build some particular set of coordinates that simplifies the proof, exploiting the fact that the Hessian is well-defined independently on the coordinates.

Fix $v$ as in the statement and split the domain and the range of the map $\Phi$ as follows

$$E = E_1 \oplus E_2, \qquad E_2 = \ker D_0\Phi, \tag{12.55}$$

$$\mathbb{R}^n = \mathbb{R}^{k_1} \oplus \mathbb{R}^{k_2}, \qquad \mathbb{R}^{k_1} = \mathrm{im}\, D_0\Phi, \tag{12.56}$$

where we select the complement $\mathbb{R}^{k_2}$ in such a way that $v \in \mathbb{R}^{k_2}$ (notice that by our assumption $v \notin \mathbb{R}^{k_1}$). Accordingly to the notation introduced, let us write

$$\Phi(x_1, x_2) = (\Phi_1(x_1, x_2), \Phi_2(x_1, x_2)), \qquad x_i \in E_i,\ i = 1, 2.$$

Since $\Phi_1$ is a submersion by construction, the Implicit function theorem implies that by a smooth change of coordinates we can linearize $\Phi_1$ and assume that $\Phi$ has the form

$$\Phi(x_1, x_2) = (D_0\Phi(x_1), \Phi_2(x_1, x_2)),$$

since $x_2 \in E_2 = \ker D_0\Phi$. Notice that, by construction of the coordinate set, the function $x_2 \mapsto \Phi_2(0, x_2)$ coincides with the restriction of $\Phi$ to the kernel of its differential, modulo its image.

Hence for every scalar function $a : \mathbb{R}^{k_2} \to \mathbb{R}$ such that $d_0 a = \lambda$ we have the equality

$$\lambda \mathrm{Hess}_0 \Phi = \mathrm{Hess}_0(a \circ \Phi_2(0, \cdot)) > 0.$$

In particular the function $a \circ \Phi_2(0, y)$ is positive in a neighborhood of 0 and 0 is an isolated point in its level set. Assume now that $\Phi(x_1, x_2) = sv$ for some $s \geq 0$. Since $v \in \mathbb{R}^{k_2}$, by construction of our coordinates, it follows that

$$D_0\Phi(x_1) = 0 \implies x_1 = 0, \qquad \text{and} \qquad \Phi_2(0, x_2) = sv.$$

In particular we have

$$\left.\frac{d}{ds}\right|_{s=0} a(\Phi_2(0, x_2)) = \left.\frac{d}{ds}\right|_{s=0} a(sv) = \langle \lambda, v \rangle < 0 \quad \Rightarrow \quad a(sv) < 0 \quad \text{for} \quad s > 0,$$

which is a contradiction. $\square$

### 12.5.1 Proof of Theorem 12.34

Let $\lambda(t)$ be an abnormal extremal and let $\gamma(t)$ the be corresponding abnormal trajectory.

$$\dot{\gamma}(t) = u_1(t) f_1(\gamma(t)) + u_2(t) f_2(\gamma(t)), \qquad t \in [0, 1]. \tag{12.57}$$

In what follows we always assume that the support of the curve $\mathrm{supp}(\gamma) \doteq \{\gamma(t) \mid t \in [0, 1]\}$ is a smooth one-dimensional submanifold of $M$, with or without border. Then either the curve $\gamma$ has no self-intersection or $\mathrm{supp}(\gamma)$ is diffeomorfic to $S^1$. In both cases we can choose a basis $f_1, f_2$ in a neighborhood of $\mathrm{supp}(\gamma)$ in such a way that $\gamma$ is the integral curve of the vector field $f_1$ (in such a way that $\gamma$ is the solution of (12.57) with associated control $\overline{u} = (1, 0)$)

$$\dot{\gamma}(t) = f_1(\gamma(t)), \qquad t \in [0, 1].$$

Notice that a change of the frame on $M$ corresponds to a smooth change of coordinates on the end-point map. With analogous reasoning as in the previous section, we describe the end point map

$$E_{q_0} : \mathcal{U} \to M, \qquad E_{q_0}(u_1, u_2) = \gamma(1),$$

for $u = (u_1, u_2)$ in a neighborhood $\mathcal{U}$ of $\overline{u} = (1, 0)$, as the composition

$$E_{q_0} = e^{f_1} \circ G,$$

where $G$ is the end point map for the modified system

$$\dot{q} = (u_1 - 1) e_*^{-tf_1} f_1 + u_2 e_*^{-tf_1} f_2. \tag{12.58}$$

Since $e_*^{-tf_1} f_1 = f_1$, denoting $g_t := e_*^{-tf_1} f_2$ and defining the primitives

$$w(t) = \int_0^t (1 - u_1(\tau)) d\tau, \qquad v(t) = \int_0^t u_2(\tau) d\tau, \tag{12.59}$$

then $G$ is the end-point map for the system

$$\dot{q} = -\dot{w} f_1(q) + \dot{v} g_t(q).$$

Notice that $(\dot w, \dot v) = (0,0)$ means $u = \overline{u}$. The Hessian of $G$ is rewritten as a map of $\dot w, \dot v$ as follows

$$\lambda_0 \mathrm{Hess}_0 G(\dot w, \dot v) = \int_0^1 \langle \lambda_0, [\int_0^t -\dot w(\tau) f_1 + \dot v(\tau) g_\tau d\tau, -\dot w(t) f_1 + \dot v(t) g_t](q_0)\rangle dt. \tag{12.60}$$

Recall that

$$\begin{aligned} D_0 G(\dot w, \dot v) &= \int_0^1 -\dot w(t) f_1(q_0) + \dot v(t) g_t(q_0) dt \\ &= -w(1) f_1(q_0) + \int_0^1 \dot v(t) g_t(q_0) dt, \end{aligned}$$

and the condition $\lambda_0 \in (\operatorname{im} D_0 G)^\perp$ is rewritten as

$$\langle \lambda_0, f_1(q_0)\rangle = \langle \lambda_0, g_t(q_0)\rangle = 0, \qquad \forall\, t \in [0,1]. \tag{12.61}$$

Since equality (12.61) is valid for all $t \in [0,1]$ then by differentiating we get

$$\langle \lambda_0, \dot g_t(q_0)\rangle = \langle \lambda_0, [f_1, g_t](q_0)\rangle = 0, \qquad \forall\, t \in [0,1]. \tag{12.62}$$

Then we can rewrite our quadratic form only as a function of $\dot v$. Indeed plugging (12.61)-(12.62) in (12.60) one gets

$$\lambda_0 \mathrm{Hess}_0 G(\dot v) = \int_0^1 \langle \lambda_0, [\int_0^t \dot v(\tau) g_\tau d\tau, \dot v(t) g_t](q_0)\rangle dt, \tag{12.63}$$

where $v$ that satifies the extra condition

$$\int_0^1 \dot v(t) g_t(q_0) dt = w(1) f_1(q_0). \tag{12.64}$$

Now we rearrange these formulas, using integration by parts, rewriting the Hessian as a quadratic form on the space of primitives (which are in particular continuous functions)

$$v(t) = \int_0^t \dot v(\tau) d\tau.$$

Using the equality

$$\int_0^t \dot v(\tau) g_\tau d\tau = v(t) g_t - \int_0^t v(\tau) \dot g_\tau d\tau, \tag{12.65}$$

we have

$$\begin{aligned} \lambda_0 \mathrm{Hess}_0 G(\dot v) = &\int_0^1 \langle \lambda_0, [v(t) g_t, \dot v(t) g_t](q_0)\rangle dt \\ &- \int_0^1 \langle \lambda_0, [\int_0^t v(\tau) \dot g_\tau d\tau, \dot v(t) g_t](q_0)\rangle dt. \end{aligned}$$

The first term in the sum is zero since $[g_t, g_t] = 0$. Exchanging the order of integration in the second term

$$\int_0^1 \langle \lambda_0, [\int_0^t v(\tau) \dot g_\tau d\tau, \dot v(t) g_t](q_0)\rangle dt = \int_0^1 \langle \lambda_0, [v(t) \dot g_t, \int_t^1 \dot v(\tau) g_\tau d\tau](q_0)\rangle dt,$$

and then integrating by parts

$$\int_t^1 \dot{v}(\tau)g_\tau d\tau = v(1)g_1 - v(t)g_t - \int_t^1 v(\tau)\dot{g}_\tau d\tau,$$

we get to

$$\lambda_0 \mathrm{Hess}_0 G(\dot{v}) = \int_0^1 \langle \lambda_0, [\dot{g}_t, g_t](q_0)\rangle v(t)^2 dt \\ + \int_0^1 \langle \lambda_0, [\int_0^t v(\tau)\dot{g}_\tau, v(t)\dot{g}_t - v(1)g_1](q_0)\rangle dt. \tag{12.66}$$

The last expression can also be rewritten as follows

$$\lambda_0 \mathrm{Hess}_0 G(\dot{v}) = \int_0^1 \langle \lambda_0, [\dot{g}_t, g_t](q_0)\rangle v(t)^2 \, dt \\ + \int_0^1 \langle \lambda_0, [\int_1^t v(\tau)\dot{g}_\tau \, d\tau + v(1)g_1, v(t)\dot{g}_t](q_0) \, dt. \tag{12.67}$$

Integrating by parts the extra condition (12.64), we find

$$\int_0^1 v(t)\dot{g}_t(q_0)dt = -w(1)f_1(q_0) + v(1)g_1(q_0). \tag{12.68}$$

The vectors $f_1(q_1)$ and $f_2(q_1)$ are linearly independent, then the vectors

$$f_1(q_0) = e_*^{-f_1}(f_1(q_1)), \qquad \text{and} \qquad g_1(q_0) = e_*^{-f_1}(f_2(q_1)),$$

are linearly independent. From (12.68) and Cauchy-Scwartz inequality, it follows that for every pair $(w, v)$ in the kernel there exists $C > 0$ such that

$$|w(1)| \leq C\|v\|_{L^2}, \qquad |v(1)| \leq C\|v\|_{L^2}. \tag{12.69}$$

*Remark* 12.37. Notice that we cannot plug in the expression (12.68) directly into the formula of the Hessian since (12.68) is valid only at the point $q_0$, while in (12.66) we have to compute a Lie bracket.

**Theorem 12.38.** *Let $\gamma : [0,1] \to M$ be an abnormal trajectory and assume that the quadratic form* (12.66) *satisfies*

$$\lambda_0 \mathrm{Hess}_0 G(\dot{v}) \geq \alpha\|v\|_{L^2}^2, \tag{12.70}$$

*for some $\alpha > 0$. Then the curve is locally minimizer in the $L^2$ topology of controls.*

*Remark* 12.39. Notice that the estimate (12.70) depends only on $v$, while the map $G$ is a smooth map of $\dot{v}$ (and $\dot{w}$). Hence Lemma 12.36 does not apply directly to $G$.

Moreover, the statement of Lemma 12.36 violates for the end-point map, since the end-point map is always locally open thanks to the bracket generating condition (see Exercise 8.4). The final point of the trajectory is never isolated in the level set.

What we are going to use is a similar idea than one contained in the proof of this Lemma, to show that the statements holds for the restriction of the end-point map to a suitable subset of controls.

*Proof of Theorem 12.38.* In this proof we use all the notation introduced in Section 12.5.1. In particular $\lambda(t)$ is an abnormal extremal path in a rank 2 sub-Riemannian structure. We fix the sub-Riemannian structure on $M$ whose generating family $f_1, f_2$ is chosen in such a way that the corresponding abnormal extremal trajectory is associated with the control $\overline{u} = (1, 0)$.

Our goal is to prove, under the assumption 12.70, that there are no horizontal curves with controls close to $\overline{u}$ that are shorter than $\gamma$ and joining $q_0$ to $q_1 = \gamma(1)$. Recall that

$$\lambda_1 \mathrm{Hess}_{\overline{u}} E_{q_0} = \lambda_0 \mathrm{Hess}_0 G \tag{12.71}$$

We can moreover assume that the length of our reference trajectory is exactly 1 (we can always dilate all the distances on our manifold, and the local optimality of the curve is not affected). We split the proof into two steps: (a) it is sufficient to show the optimality in the subset of controls of constant modulus (namely, curves of constant speed) (b) proof of the minimality in the in the subset of controls of constant modulus.

(a) Let us consider the reparametrization map $\theta : L^2([0,1], \mathbb{R}^2) \to L^2([0,1], \mathbb{R}^2)$ that maps every control $v$ to the control $\theta(v)$ defined on $[0,1]$ with constant modulus (and associated with the reparametrized trajectory with constant speed, cf. Lemma 3.15 and 3.16). First notice that from Lemma 3.64

$$\|\theta(v)\|^2_{L^2} = 2J(\gamma_{\theta(v)}) = \ell(\gamma_{\theta(v)})^2 = \ell(\gamma_v)^2 \leq 2J(\gamma_v) = \|v\|^2_{L^2}. \tag{12.72}$$

Let us now prove that for $\overline{u} = (1, 0)$ (it is sufficient that $\bar{u}$ is constant)

$$\|\overline{u} - \theta(v)\|_{L^2} \leq \|\overline{u} - v\|^2_{L^2} \tag{12.73}$$

Indeed

$$\begin{aligned}
\|\overline{u} - \theta(v)\|^2_{L^2} &= \int_0^1 |\overline{u}(t) - \theta(v)(t)|^2 dt \\
&= \int_0^1 |\overline{u}(t)|^2 dt - 2\int_0^1 \langle \overline{u}(t), \theta(v)(t)\rangle_{\mathbb{R}^2}\, dt + \int_0^1 |\theta(v)(t)|^2 dt \\
&\leq \int_0^1 |\overline{u}(t)|^2 dt - 2\int_0^1 \langle \overline{u}(t), v(t)\rangle_{\mathbb{R}^2}\, dt + \int_0^1 |v(t)|^2 dt = \|\overline{u} - v\|^2_{L^2}
\end{aligned}$$

where in the last inequality we used (12.72) and the fact that $\overline{u}$ is constant to estimate the mixed term. The inequality (12.73) says that $\theta(\mathcal{B}_{\bar{u}}(r)) \subset \mathcal{B}_{\bar{u}}(r)$, where $\mathcal{B}_{\bar{u}}(r)$ is a ball centered in $\overline{u}$ in $L^2$, hence it is enough to prove minimality in the subset of controls of constant modulus.

(b) The set of curves of constant speed and length less or equal than 1 can be parametrized, using Lemma 3.16, by the set of controls

$$\{(u_1, u_2) |\, u_1^2 + u_2^2 \leq 1\}.$$

Following the notation (12.59), notice that

$$\{(u_1, u_2) |\, u_1^2 + u_2^2 \leq 1\} \subset \{(w, v) |\, \dot{w} \geq 0\}.$$

Let now $E = E_{q_0}$. We now fix a function $a \in C^\infty(M)$ and a smooth map $b : M \to \mathrm{Im}\, D_{\overline{u}}E$ such that

$$a(q_1) = 0, \quad d_{q_1} a = \lambda_1 \in (\mathrm{Im} D_{\overline{u}}E)^\perp, \qquad b(q_1) = 0, \quad D_{q_1} b\big|_{\mathrm{Im}\, D_{\overline{u}}E} = \mathrm{id}.$$

We want to show that

$$a \circ E\big|_D(\dot w, \dot v) = \lambda_1 \mathrm{Hess}_{\overline{u}} E(\dot w, \dot v) + R(\dot w, \dot v), \qquad \text{where} \qquad \frac{R(w,v)}{\|v\|^2_{L^2}} \underset{\|(\dot w, \dot v)\| \to 0}{\longrightarrow} 0, \tag{12.74}$$

in the domain

$$D = \{(\dot w, \dot v) \mid b(E(\dot w, \dot v)) = 0, \dot w \geq 0\}.$$

Notice that the at the denominator in (12.74) we have not $\|\dot v\|$ but $\|v\|$. This is necessary to apply inequality (12.70). Indeed if we prove (12.74) we have that the point $(\dot w, \dot v) = (0, 0)$ is locally optimal for $E$. This means that the curve $\gamma$, i.e., the curve associated with control $\overline{u} = (1, 0)$, is locally optimal.

Using the identity

$$\overrightarrow{\exp} \int_0^t \dot v(\tau) f_2 d\tau = e^{v(t) f_2},$$

and applying the variations formula (6.36) to the end-point map $E$ we get

$$\begin{aligned} E(\dot w, \dot v) &= q_0 \circ \overrightarrow{\exp} \int_0^1 (1 - \dot w(t)) f_1 + \dot v(t) f_2 \, dt \\ &= q_0 \circ \overrightarrow{\exp} \int_0^1 (1 - \dot w(t)) e_*^{-v(t) f_2} f_1 \, dt \circ e^{v(1) f_2} \end{aligned}$$

Hence we can express the end-point map as a smooth function of the pair $(\dot w, v)$.

Now, to compute (12.74), we can assume that the function $a$ is constant on the trajectories of $f_2$ (since we only fix its differential at one point) so that

$$e^{v(1) f_2} \circ a = a,$$

which simplifies our estimates:

$$a \circ E(\dot w, \dot v) = q_0 \circ \overrightarrow{\exp} \int_0^1 (1 - \dot w(t)) e_*^{-v(t) f_2} f_1 \, dt \, a.$$

Writing

$$(1 - \dot w(t)) e_*^{-v(t) f_2} f_1 = f_1 + X^0(v(t)) + \dot w(t) X^1(v(t)), \tag{12.75}$$

and using the variation formula (6.37), setting $Y_t^i = e_*^{(t-1) f_1} X^i$ for $i = 0, 1$, we get (recall that $q_1 = e^{f_1}(q_0)$)

$$a \circ E(\dot w, \dot v) = q_1 \circ \overrightarrow{\exp} \int_0^1 Y_t^0(v(t)) + \dot w(t) Y_t^1(v(t)) dt \, a, \qquad Y_t^0(0) = 0,$$

Expanding the chronological exponential we find that

(a) the zero order terms vanish since $Y_t^0(0) = 0$,

(b) all first order terms vanish since $\lambda_1 = d_{q_1} a$ is orthogonal to the subspace $\mathrm{Im}\, D_{\overline{u}} E$ spanned by the vectors $f_1(q_1)$ and $\left(e_*^{t f_1} f_2\right)(q_1)$.

(c) the second order terms are in the Hessian. Indeed, $b \circ E$ is a submersion and $\ker D_{\overline{u}}(b \circ E) = \ker D_{\overline{u}} E$. Intersection of $(b \circ E)^{-1}(0)$ with a small neighborhood of 0 is a smooth Hilbert submanifold of the domain of $E$ whose tangent space at $\overline{u}$ is equal to $\ker D_{\overline{u}} E$. Hence $D$ has a smooth local parameterization by the elements of $\ker D_{\overline{u}} E$ such that the linear part of the parameterization is just the identity map. Only this linear part appears in the second order terms of the expansion of $a \circ E\big|_D$ at 0.

The remainder depends on $v, \dot{w}$ and has the order at least 3. If we show that it can be estimated with $o(\|v\|^2)$ (where $o(\|v\|^2)$ have the same meaning as in (12.74)), then we can apply estimate (12.70) and conclude that the function $a \circ E\big|_D$ has constant sign for $\|\dot{v}\|$ and $\|\dot{w}\|$ small, i.e., the control $\overline{u}$ is optimal.

The detailed estimate is rather long, boring and not so instructive, it can be found in [AS95]. Here we explain only key estimates, about integral of monomials of order 3. We are going to prove that:

$$\int_0^1 \dot{w}(t) v^2(t) dt = o(\|v\|^2), \qquad \int_0^1 \dot{w}(t) \int_0^t \dot{w}(\tau) v(\tau) d\tau dt = o(\|v\|^2),$$

$$\int_0^1 \dot{w}(t) \int_0^t \dot{w}(\tau) \int_0^\tau \dot{w}(s) ds d\tau dt = o(\|v\|^2).$$

Using that $\dot{w} \geq 0$, which is the key assumption, and the fact that $(\dot{w}, \dot{v}) \in D$, which gives the estimates (12.69), we compute

$$\begin{aligned}
\left| \int_0^1 \dot{w}(t) v^2(t) dt \right| &\leq \int_0^1 |\dot{w}(t)| v^2(t) dt \\
&= \int_0^1 \dot{w}(t) v^2(t) dt \\
&= w(1) v^2(1) - \int_0^1 w(t) v(t) \dot{v}(t) dt \\
&\leq \|v\|^3 + \varepsilon \|v\|^2,
\end{aligned}$$

where the estimate for the second term follows from

$$\begin{aligned}
\left| \int_0^1 w(t) v(t) \dot{v}(t) dt \right| &\leq (\max |w(t)|) \left| \int_0^1 v(t) \dot{v}(t) dt \right| \\
&\leq w(1) \|v\| \|\dot{v}\| \\
&\leq C \|\dot{v}\| \|v\|^2.
\end{aligned}$$

The second integral can be rewritten

$$\int_0^1 \dot{w}(t) \int_0^t \dot{w}(\tau) v(\tau) d\tau dt = w(1) \int_0^1 \dot{w}(t) v(t) dt - \int_0^1 w(t) v(t) \dot{w}(t) dt$$

and then we estimate

$$\left| \int_0^1 \dot{w}(t) \int_0^t \dot{w}(\tau) v(\tau) d\tau dt \right| \leq 2|w(1)| \int_0^1 v(t)\dot{w}(t)dt$$
$$\leq C\|\dot{w}\|\|v\|^2.$$

Finally, the last integral is very easy to estimate using the equality

$$\int_0^1 \dot{w}(t) \int_0^t \dot{w}(\tau) \int_0^\tau \dot{w}(s) ds d\tau dt = \frac{1}{6} \int_0^1 \dot{w}(t)^3 dt$$
$$\leq C\|\dot{w}\|\|v\|^2.$$

Starting from these estimates it is easy to show that any mixed monomial of order greater that three and the whole remainder satisfy the desired estimate as well. □

Applying these results to a small piece of abnormal trajectory we can prove that small pieces of nice abnormals are minimizers

*Proof of Theorem 12.34* . If we apply the arguments above to a small piece $\gamma_s = \gamma|_{[0,s]}$ of the curve $\gamma$ it is easy to see that the Hessian rescales as follows,

$$\lambda_0 \mathrm{Hess}_0 G_s(\dot{v}) = \int_0^s \langle \lambda_0, [g_t, \dot{g}_t](q_0) \rangle v(t)^2 dt$$
$$+ \int_0^s \langle \lambda_0, [\int_0^t v(\tau)\dot{g}_\tau d\tau, v(t)\dot{g}_t - v(s)g_s](q_0) \rangle dt.$$

Since the generalized Legendre condition ensures that (see also Lemma 12.33)

$$\langle \lambda_0, [g_t, \dot{g}_t](q_0) \rangle \geq C > 0$$

(we know that the corresponding quadratic form is semidefinite and that $f_1$ is in the kernel), then the norm

$$\|v\|_g = \left( \int_0^s \langle \lambda_0, [g_t, \dot{g}_t](q_0) \rangle v(t)^2 dt \right)^{1/2} \tag{12.76}$$

is equivalent to the standard $L^2$-norm. Hence the Hessian can be rewritten as

$$\lambda_0 \mathrm{Hess}_0 G_s(\dot{v}) = \|v\|_g^2 + \langle Tv, v \rangle , \tag{12.77}$$

where $T$ is a compact operator in $L^2$ of the form

$$(Tv)(t) = \int_0^s K(t,\tau) v(\tau) d\tau$$

Since $\|T\|^2 = \|K\|_{L^2}^2 \to 0$ for $s \to 0$, it follows that the Hessian satisfy the estimate in the assumptions of Theorem 12.38, for $s$ small enough. □

## 12.6 Conjugate points along abnormals

In this section, we give an effective way to check the inequality (12.70), that implies local minimality of nice abnormal extremal trajectories according to Theorem 12.38.

Following the notation of the previous section, given a nice abnormal extremal we define the quadratic form $Q_1(v) := \lambda_0 \mathrm{Hess}_0 G(\dot{v})$. The quadratic form $Q_1$ is continuous in the topology defined by the norm $\|v\|_{L^2}$. The closure of the domain of $Q_1$ in this topology, is the space

$$D(Q_1) = \left\{ (v, v_1) \in L^2([0,1], \mathbb{R}^m) \times \mathbb{R} \mid \int_0^1 v(t)\dot{g}_t(q_0)\, dt + v_1 g_1(q_0) \in \mathrm{span}\{f_1(q_0)\} \right\}.$$

The extension of $Q_1$ to $D(Q_1)$ is denoted by the same symbol $Q_1$. We define

$$l(t) = \langle \lambda_0, [\dot{g}_t, g_t](q_0) \rangle, \quad X_t = v_1 g_1 + \int_1^t v(\tau)\dot{g}_\tau \, d\tau,$$

and we rewrite the form $Q_1$ in these more compact notations:

$$Q_1(v) = \int_0^1 l(t)v(t)^2\, dt + \int_0^1 \langle \lambda_0, [X_t, \dot{X}_t](q_0) \rangle\, dt.$$

Notice that

$$\dot{X}_t = v(t)\dot{g}_t, \quad X_1 \wedge g_1 = 0, \; X_0(q_0) \wedge f_1(q_0) = 0. \tag{12.78}$$

Moreover, we introduce the family of quadratic forms $Q_s$, for $0 < s \leq 1$, as follows

$$Q_s(v) := \int_0^s l(t)v(t)^2\, dt + \int_0^s \langle \lambda_0, [X_t, \dot{X}_t](q_0) \rangle\, dt.$$

Recall that $l(t)$ is a strictly positive continuous function. In particular the formula

$$\|v\|_l^2 := \int_0^1 l(t)v(t)^2\, dt, \tag{12.79}$$

is the square of a norm on $L^2([0,1], \mathbb{R}^m)$ that is equivalent to the standard $L^2$ norm. The next statement is proved by the same arguments as those used in Proposition 8.54 and Lemma 8.55. We leave details to the reader.

**Proposition 12.40.** *The form $Q_1$ is positive definite if and only if* $\ker Q_s = 0$ *for all* $s \in (0, 1]$.

**Definition 12.41.** A time moment $s \in (0, 1]$ is called *conjugate* to 0 for the abnormal extremal $\gamma$ if $\ker Q_s \neq 0$.

We are going to characterize conjugate times in terms of an appropriate "Jacobi equation".

Let $\xi_1 \in T_{\lambda_0}(T^*M)$ and $\zeta_t \in T_{\lambda_0}(T^*M)$ be the values at $\lambda_0$ of the Hamiltonian lifts of the vector fields $f_1$ and $g_t$. Recall that the Hamiltonian lift of a field $f \in \mathrm{Vec}(M)$ is the Hamiltonian vector field associated to the Hamiltonian function $\lambda \mapsto \langle \lambda, f(q) \rangle$, $\lambda \in T_q^*M$, $q \in M$. We have:

$$Q_s(v) = \int_0^s l(t)v(t)^2\, dt + \int_0^s \sigma(x(t), \dot{x}(t))\, dt,$$

$$\dot{x}(t) = v(t)\dot{\zeta}_t, \quad x(s) \wedge \zeta_s = 0,\ \pi_* x(0) \wedge \pi_* \xi_1 = 0,$$

where $\sigma$ is the standard symplectic product on $T_{\lambda_0}(T^*M)$ and $\pi : T^*M \to M$ is the standard projection. Moreover

$$l(t) = \sigma(\dot{\zeta}_t, \zeta_t), \quad 0 \le t \le 1. \tag{12.80}$$

Let $E = \operatorname{span}\{\xi_1, \zeta_t \mid 0 \le t \le 1\}$. We use only the restriction of $\sigma$ to $E$ in the expression of $Q_s$ and we are going to get rid of unnecessary variables. Namely, we set: $\Sigma := E/(\ker \sigma|_E)$.

**Lemma 12.42.** *Let* $d := \dim \operatorname{span}\{f_1(q_0), g_t(q_0) \mid 0 \le t \le 1\}$. *Then* $\dim \Sigma \le 2d - 2$.

*Proof.* The dimension of $\Sigma$ is equal to twice the codimension of a maximal isotropic subspace of $\sigma|_E$. We have: $\sigma(\xi_1, \zeta_t) = \langle \lambda_0, [f_1, g_t](q_0)] \rangle = 0$, $\forall t \in [0,1]$, hence $\xi_1 \in \ker \sigma|_E$. Moreover, $\pi_*(E) = \operatorname{span}\{f_1(q_0), g_t(q_0), 0 \le t \le 1\}$ and $E \cap \ker \pi_*$ is an isotropic subspace of $\sigma|_E$. □

We denote by $\underline{\zeta}_t \in \Sigma$ the projection of $\zeta_t$ to $\Sigma$ and by $\Pi \subset \Sigma$ the projection of $E \cap \ker \pi_*$. Note that the projection of $\xi_1$ to $\Sigma$ is 0. Moreover, equality (12.80) implies that $\underline{\zeta}_t \ne 0$, $\forall t \in [0,1]$. The final expression of $Q_s$ is as follows:

$$Q_s(v) = \int_0^s l(t) v(t)^2 \, dt + \int_0^s \sigma(x(t), \dot{x}(t)) \, dt,$$

$$\dot{x}(t) = v(t)\dot{\underline{\zeta}}_t, \quad x(s) \wedge \underline{\zeta}_s = 0,\ x(0) \in \Pi.$$

We have: $v \in \ker Q_s$ if and only if

$$\int_0^s \Big( l(t)v(t) + \sigma(x(t), \dot{\underline{\zeta}}_t) \Big) w(t) \, dt = 0,$$

for any $w(\cdot)$ such that

$$\int_0^s \dot{\underline{\zeta}}_t w(t) \, dt \in \Pi + \mathbb{R}\underline{\zeta}_s.$$

We obtain that $v \in \ker Q_s$ if and only if there exists $\nu \in \Pi^{\angle} \cap \underline{\zeta}_s^{\angle}$ such that

$$l(t)v(t) + \sigma(x(t), \dot{\underline{\zeta}}_t) = \sigma(\nu, \dot{\underline{\zeta}}_t), \quad 0 \le t \le s.$$

We set $y(t) = x(t) - \nu$ and obtain the following:

**Theorem 12.43.** *A time moment* $s \in (0,1]$ *is conjugate to* 0 *if and only if there exists a non-constant solution of the equation*

$$l(t)\dot{y} = \sigma(\dot{\underline{\zeta}}_t, y)\dot{\underline{\zeta}}_t, \tag{12.81}$$

*that satisfies the following boundary conditions:*

$$\exists\, \nu \in \Pi^{\angle} \cap \underline{\zeta}_s^{\angle} \quad \textit{such that} \quad (y(s) + \nu) \wedge \underline{\zeta}_s = 0,\ (y(0) + \nu) \in \Pi. \tag{12.82}$$

*Remark* 12.44. Notice that identity (12.80) implies that $y(t) = \underline{\zeta}_t$ for $t \in [0,1]$ is a solution to the equation (12.81). However this solution may violate the boundary conditions.

Let us consider the special case when $\dim \operatorname{span}\{f_1(q_0), g_t(q_0) \mid 0 \le t \le 1\} = 2$.

**Corollary 12.45.** *If* $\dim \operatorname{span}\{f_1(q_0), g_t(q_0), 0 \le t \le 1\} = 2$, *then the segment* $[0,1]$ *does not contain conjugate time moments and the assumption of Theorem 12.38 is satisfied.*

Indeed, under the assumption of the corollary, one has $\dim E = 2$ and $\dim \Pi = 1$. Hence $\Pi^\angle = \Pi, \underline{\zeta}_s^\angle = \mathbb{R}\underline{\zeta}_s$ and $\Pi^\angle \cap \underline{\zeta}_s^\angle = 0$. Then $\nu$ in the boundary conditions (12.82) must be 0 and $y(s) = c\underline{\zeta}_s$, where $c$ is a nonzero constant. Thus $y(t) = c\underline{\zeta}_t$ for $0 \leq t \leq 1$ and $y(0) = c\underline{\zeta}_0 \notin \Pi$, which proves the statement.

*Remark* 12.46. Notice that, denoting $d := \dim \operatorname{span}\{f_1(q_0), g_t(q_0) \mid 0 \leq t \leq 1\}$, then $2 \leq d < \dim M$, since the trajectory is abnormal. Hence the assumption of Corollary 12.45 is automatically satisfied for abnormal extremal in a 3-dimensional sub-Riemannian manifold.

### 12.6.1 Abnormals in dimension 3

In the three-dimensional case, thanks to Remark 12.46, we have the following result.

**Theorem 12.47.** *Let $M$ be a sub-Riemannian manifold, $\dim M = 3$, and let $\gamma : [0,1] \to M$ be a nice abnormal geodesic. Then $\gamma$ is a strict local length-minimizer in the the $W^{1,2}$ topology for horizontal trajectories joining the same endpoints.*

We recall that nice abnormals are integral curves of a smooth vector fields on $M$, hence they are smooth. In the particular case of the isoperimetric problem studied in Section 4.4.2, the characterization can be made more precise.

#### Nice abnormals for the isoperimetric problem on surfaces

Recall the isoperimetric problem: given two points $q_0, q_1$ on a 2-dimensional Riemannian manifold $M$, a 1-form $A \in \Lambda^1 M$ and $c \in \mathbb{R}$, we have to find (if it exists) the minimum:

$$\min\{\ell(\gamma) \mid \gamma(0) = q_0, \gamma(T) = q_1, \int_\gamma A = c\}. \tag{12.83}$$

As shown in Section 4.4.2, this problem can be reformulated as a sub-Riemannian problem on the extended manifold

$$\overline{M} = M \times \mathbb{R} = \{(q,z) \mid q \in M, z \in \mathbb{R}\},$$

where the sub-Riemannian structure is defined by the contact form

$$\mathcal{D} = \ker(dz - A),$$

and the sub-Riemannian length of a curve coincides with the Riemannian length of its projection on $M$. If we write $dA = b\, dV$, where $b$ is a smooth function and $dV$ denote the Riemannian volume on $M$, we have that the Martinet set is defined by the cilynder

$$\mathfrak{M} = \mathbb{R} \times b^{-1}(0).$$

**Corollary 12.48.** *Assume that the set $b^{-1}(0)$ is a regular level of $b$. Then the Martinet set $\mathfrak{M}$ is a smooth surface and all abnormal extremals are nice. Moreover the projection on $M$ of abnormal extremal trajectories on $\overline{M}$ are contained in the connected components of the set $b^{-1}(0)$.*

The projection $\pi : \overline{M} \to M$ satisfies $\pi_* \mathcal{D} = TM$, by construction. It follows that the distribution is always transversal to the Martinet set and all abnormal extremals are nice, since $\mathcal{D}_q^3 = T_q\overline{M}$ for all $(q,z) \in \overline{M}$.

Notice that one can recover the whole abnormal extremal integrating the 1-form $A$ to find the missing component. In other words the abnormal extremals are spirals on $\mathcal{M}$ with step equal to $\int_\Omega dV$, (if $dV$ is the volume form on $M$, it coincides with the area of the region $\Omega$ inside the curve defined on $M$ by the connected component of $b^{-1}(0)$).

Consider now a compact connected component of $b^{-1}(0)$; this is a smooth closed curve. Corollary 12.48 together with Theorem 12.38 implies that this closed curve passed once, twice, three times or an arbitrary number of times is a locally optimal solution of the isoperimetric problem (12.83). Moreover, this is true for any choice of the Riemannian metric on the surface $M$!

### A non-nice abnormal extremal

In this section we give an example of non nice (and indeed not smooth) abnormal extremal.

Consider the isoperimetric problem on $\mathbb{R}^2 = \{(x_1, x_2) \mid x_i \in \mathbb{R}\}$ defined by the 1-form $A$ such that

$$A = -\frac{x_2^2}{4} x_1 dx_1 + \frac{x_1^2}{4} x_2 dx_2, \qquad d\nu = x_1 x_2 dx_1 dx_2.$$

Here the function $b$ defined above is expressed as $b(x_1, x_2) = x_1 x_2$ and the set $b^{-1}(0)$ consists of the union of the two axes, with moreover $db|_0 = 0$. Notice that the level set is not smooth and Corollary 12.48 does not apply.

**Exercise 12.49.** Prove that the corresponding sub-Riemannian structure on $\mathbb{R}^3 = \{(x_1, x_2, z)\}$ is defined by the orthonormal frame

$$f_1 = \partial_{x_1} + x_1 \frac{x_2^2}{4} \partial_z, \qquad f_2 = \partial_{x_2} - x_2 \frac{x_1^2}{4} \partial_z.$$

Compute the step of the structure at every point and show in particular that it is equal to 4 at every point of the form $(0, 0, z)$.

Let us fix $\bar{x}_1, \bar{x}_2 > 0$ and consider the curve joining the points $(0, \bar{x}_2)$ and $(\bar{x}_1, 0)$ in $\mathbb{R}^2$ defined by the union of two segments contained in the coordinate axes:

$$\gamma : [-\bar{x}_2, \bar{x}_1] \to \mathbb{R}^2, \qquad \gamma(t) = \begin{cases} (0, -t), & t \in [-\bar{x}_2, 0], \\ (t, 0), & t \in [0, \bar{x}_1]. \end{cases}$$

The curve $\gamma$ is a projection of an abnormal extremal that is not nice, since it is not smooth.

**Proposition 12.50.** *The curve $\gamma$ is a projection of an abnormal extremal that is not a length-minimizer.*

*Proof of Proposition 12.50.* Let us build a family of admissible variations $\gamma_{\varepsilon,\delta}$ of the curve $\gamma$ defined as in Figure 12.1. Namely, $\gamma_{\varepsilon,\delta}$ we cut a corner of size $\varepsilon$ at the origin and we turn around a small circle of radius $\delta$ before reaching the same end-point of $\gamma$. Denoting by $D_\varepsilon$ and $D_\delta$ the two regions enclosed by the curve, it is easy to see that the isoperimetric condition rewrites as follows

$$0 = \int_{\gamma_{\varepsilon,\delta}} \nu = \int_{D_\varepsilon} d\nu - \int_{D_\delta} d\nu.$$

It is then easy using that $d\nu = x_1 x_2 dx_1 dx_2$ to show that there exists $c_1, c_2 > 0$ such that

$$\int_{D_\varepsilon} d\nu = c_1 \varepsilon^4, \qquad \int_{D_\delta} d\nu = c_2 \delta^3,$$

while

$$\ell(\gamma_{\varepsilon,\delta}) - \ell(\gamma) = 2\pi\delta - (2-\sqrt{2})\varepsilon. \tag{12.84}$$

Choosing $\varepsilon > 0$ in such a way that $c_1\varepsilon^4 = c_2\delta^3$, it is an easy exercise to show that the quantity (12.84) is negative for $\delta > 0$ small. $\square$

*Remark* 12.51. (i) If one considers some planar curve $\widetilde{\gamma}$ that is a projection of a normal extremal trajectory having the same end-points than $\gamma$ and contained in the set $\{(x_1, x_2) \in \mathbb{R}^2, x_1 > 0, x_2 > 0\}$, then $\widetilde{\gamma}$ must have self intersections. Indeed it is easy to see that, if it is not the case, then the isoperimetric condition

$$\int_{\widetilde{\gamma}} \nu = 0$$

cannot be satisfied.

(ii) It is still an open problem to determine which is the length-minimizer joining these two points. We know that it exists and it is a projection of a normal extremal (hence smooth) but for instance we do not know how many self-intersection it has.

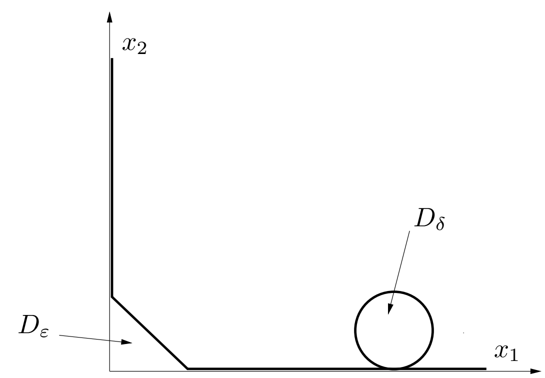



Figure 12.1: Non minimality of a non-nice abnormal extremal trajectory

**The general 3D case**

As we discussed in Section 4.3, ona general 3D sub-Riemannian manifold abnormal extremals are contained in the annichilator of the distribution $\mathcal{D}^\perp$. If $h_1, h_2$ are the Hamiltonians linear on fibers associated to the vector fields of a canonical frame $f_1, f_2$ for the sub-Riemannian structure, we can write

$$\mathcal{D}^\perp = \{\lambda \in T^*M,\ h_1(\lambda) = h_2(\lambda) = 0\},$$

By definition nice abnormal extremal trajectories, that are projections of nice abnormal extremals, live in the subset of $M$ where the step is equal to 3. So non-nice abnormal extremal trajectories necessary pass through a point where the step is at least 4 (cf. also the previous section and Exercice 12.49).

Recall that abnormal extremal trajectories are contained in the Martinet set, which is the set of points where the step is at least three

$$\mathfrak{M} = \{q \in M \mid \mathcal{D}_q^2 \neq T_q M\}.$$

Equivalently, denoting by $a$ the smooth function such that $d\nu = a dV$, where $dV$ is the volume form of the Riemannian manifold $M$, we have that $\mathfrak{M} = a^{-1}(0)$.

In what follows we always assume that $\mathfrak{M}$ defines a smooth submanifold in $M$ (of dimension 2)

**Exercise 12.52.** Assume that $\mathfrak{M}$ defines a smooth submanifold in $M$ (of dimension 2). Prove that $(\mathcal{D}_q^2)^\perp \neq (\mathcal{D}_q^3)^\perp$ if and ony if the Martinet set (which is a smooth surface) is transversal to the distribution at the point $q$.

Let us write the equations for abnormal extremals on $\mathfrak{M}$. We have to find controls $u_1, u_2$ such that the vector field $f_u = u_1 f_1 + u_2 f_2$ is tangent to the surface $\mathfrak{M}$, i.e., they have to satisfy

$$f_u(a) = u_1 f_1(a) + u_2 f_2(a) = 0.$$

It means that the vector field

$$V = (f_2 b) f_1 - (f_1 b) f_2. \tag{12.85}$$

is globally defined on $\mathfrak{M}$ and its equilibrium points are exactly the points where the Martinet set is tangent to the distribution (equivalently, the points where the step is larger than three)

At a points $q$ where the Martinet set is tangent to the distribution, it is well-defined the linearization $D_q V$ of the vector field $V$ and its trace, which is the divergence of $V$ satisfies $\operatorname{div} V = 0$.

**Exercise 12.53.** Prove that the divergence of the vector field defined in (12.85) satisfies $\operatorname{div} V = 0$.

Generically these critical points are isolated and the linearization $D_q V : T_q\mathfrak{M} \to T_q\mathfrak{M}$ is a well-defined operator which has two opposite non-zero real or imaginary eigenvalues.

(a) If the eigenvalues are imaginary, then the portrait of the dynamical system is a focus. One can prove that the horizontal trajectories on $\mathfrak{M}$ that spiral to the critical point have infinte length since $\operatorname{div} V(q) = 0$. The argument is based on the following exercice.

**Exercise 12.54.** Let $V$ be a smooth vector field on $\mathbb{R}^2$ such that $\operatorname{div} V(0) = 0$ such that $DV(0)$ has imaginary eigenvalues. Prove that the spirals through the origin have infinite Euclidean length.

(b) If the eigenvalues are real and opposite, then the portrait of the dynamical system is a saddle point. In this case one can always build a non-smooth and non-nice abnormal by considering a horizontal trajectory by joining the two separatrices.

**Exercise 12.55.** Prove that the sub-Riemannian structure described in the previous section (Exercise 12.49) is the nilpotent approximation of the generic case described in (b) above.

### 12.6.2 Higher dimension

Now consider another important special case that is typical if the dimension of the ambient manifold is greater than 3. Namely, assume that, for some $k \geq 2$, the vector fields

$$f_1,\ f_2,\ (\mathrm{ad} f_1) f_2,\ \ldots,\ (\mathrm{ad} f_1)^{k-1} f_2, \tag{12.86}$$

are linearly independent in any point of a neighborhood of our nice abnormal extremal trajectory $\gamma$, while $(\mathrm{ad} f_1)^k f_2$ is a linear combination of the vector fields (12.86) in any point of this neighborhood; in other words,

$$(\mathrm{ad} f_1)^k f_2 = \sum_{i=0}^{k-1} a_i (\mathrm{ad} f_1)^i f_2 + \alpha f_1,$$

where $a_i, \alpha$ are smooth functions. In this case, all solutions of the equation $\dot{q} = f_1(q)$ that are close to $\gamma$ are abnormal extremal trajectories.

A direct calculation based on the fact that $\langle \lambda(t), (\mathrm{ad} f_1^i) f_2)(\gamma(t) \rangle = 0,\ 0 \leq t \leq 1$, gives the identity:

$$\zeta_t^{(k)} = \sum_{i=0}^{k-1} a_i(\gamma(t)) \zeta^{(i)} + \alpha(\gamma(t)) \xi_1. \quad 0 \leq t \leq 1. \tag{12.87}$$

Identity (12.87) implies that $\dim E = k$ and $\Pi = 0$. The boundary conditions (12.82) take the form:

$$y(0) \in \underline{\zeta}_s^{\angle}, \quad (y(s) - y(0)) \wedge \underline{\zeta}_s = 0. \tag{12.88}$$

#### Engel-type distributions

The caracterization of conjugate points is especially simple and geometrically clear if the ambient manifold has dimension 4. Let $\mathcal{D}$ be a rank 2 equiregular distribution in a 4-dimensional manifold (Engel-type distributions). Notice that the equiregularity forces the growth vector associated with the distribution to be $(2, 3, 4)$.

Then abnormal extremal trajectories form a 1-dimensional foliation of the manifold and condition (12.86) is satisfied with $k = 2$. Moreover, $\dim E = 3,\ \dim \Sigma = 2$ and $\underline{\zeta}_s^{\angle} = \mathbb{R}\underline{\zeta}_s$. Recall that $y(t) = \underline{\zeta}_t$, for $0 \leq t \leq s$, is a solution to (12.81). Hence boundary conditions (12.88) are equivalent to the condition

$$\underline{\zeta}_s \wedge \underline{\zeta}_0 = 0. \tag{12.89}$$

It is easy to re-write relation (12.89) in an intrinsic way without the special notations we used to simplify calculations. We have the following characterization of conjugate times.

**Lemma 12.56.** *Let $\mathcal{D}$ be an Engel-type distribution. A time moment $t$ along an abnormal extremal trajectory $\gamma$ is conjugate to 0 if and only if*

$$e_*^{tf_1} \mathcal{D}_{\gamma(0)} = \mathcal{D}_{\gamma(t)}.$$

The flow $e^{tf_1}$ preserves $\mathcal{D}^2$ and $f_1$, but does not preserve $\mathcal{D}$. The plane $e_*^{tf_1}\mathcal{D}$ rotates around the line $\mathbb{R}f_1$ inside $\mathcal{D}^2$ with a nonvanishing angular velocity. A conjugate time is when the plane makes a complete revolution. Collecting all the information we obtain:

**Theorem 12.57.** *Let $\mathcal{D}$ be a Engel-type distribution, $f_1$ be a horizontal vector field such that $[f_1, \mathcal{D}^2] = \mathcal{D}^2$ and $\dot{\gamma} = f_1(\gamma)$. Then $\gamma$ is an abnormal extremal trajectory. Moreover*

*(i) if $e_*^{tf_1}\mathcal{D}_{\gamma(0)} \neq \mathcal{D}_{\gamma(t)}$, $\forall t \in (0,1]$, then $\gamma$ is a local length-minimizer for any sub-Riemannian structure on $\mathcal{D}$,*

*(ii) if $e_*^{tf_1}\mathcal{D}_{\gamma(0)} = \mathcal{D}_{\gamma(t)}$ for some $t \in (0,1)$ and $\gamma$ is not a normal geodesic, then $\gamma$ is not a local length-minimizer.*

## 12.7 Equivalence of local minimality with respect to $W^{1,2}$ and $C^0$ topology

In this section we prove that, under the assumption that our trajectory is smooth, it is equivalent to be locally optimal with respect to the $W^{1,2}$ topology or in the $C^0$ uniform topology for the trajectories.

Recall that a curve $\bar{\gamma}$ is called a $C^0$ *local length-minimizer* if $\ell(\bar{\gamma}) \leq \ell(\gamma)$ for every curve $\gamma$ that is $C^0$ close to $\gamma$ satisfying the same boundary conditions, while it is called a $W^{1,2}$ *local length-minimizer* if $\ell(\bar{\gamma}) \leq \ell(\gamma)$ for every curve $\gamma$ such that the control $u$ corresponding to $\gamma$ is close in the $L^2$ topology to the control $\bar{u}$ associated with $\bar{\gamma}$ and $\gamma$ satisfies the same boundary conditions.

Any $C^0$ local length minimizer is automatically a $W^{1,2}$ local length minimizer. Indeed it is possible to show that for every $v, w$ in a neighborhood of a fixed control $u$ there exists a constant $C > 0$ such that

$$|\gamma_v(t) - \gamma_w(t)| \leq C\|v - w\|_{L^2}, \qquad \forall\, t \in [0, T],$$

where $\gamma_v$ and $\gamma_w$ are the trajectories associated to controls $v$,$w$ respectively.

**Theorem 12.58.** *Let $M$ be a sub-Riemannian structure that is the restriction to $\mathcal{D}$ of a Riemannian structure $(M, g)$. Assume $\bar{\gamma}$ is of class $C^\infty$ and has no self intersections. If $\bar{\gamma}$ is a (strict) local minimizer in the $W^{1,2}$ topology then $\bar{\gamma}$ is also a (strict) local minimizer in the $C^0$ topology.*

*Proof.* Since $\bar{\gamma}$ has no self intersections, we can look for a preferred system of coordinates on an open neighborhood $\Omega$ in $M$ of the set $V = \{\bar{\gamma}(t) : t \in [0,1]\}$. For every $\varepsilon > 0$, define the cylinder in $\mathbb{R}^n = \{(x, y) : x \in \mathbb{R}, y \in \mathbb{R}^{n-1}\}$ as follows

$$I_\varepsilon \times B_\varepsilon^{n-1} = \{(x, y) \in \mathbb{R}^n : x \in (-\varepsilon, 1+\varepsilon), y \in \mathbb{R}^{n-1}, |y| < \varepsilon\}, \tag{12.90}$$

We need the following technical lemma.

**Lemma 12.59.** *There exists $\varepsilon > 0$ and a coordinate map $\Phi : I_\varepsilon \times B_\varepsilon^{n-1} \to \Omega$ such that for all $t \in [0,1]$*

*(a) $\Phi(t, 0) = \bar{\gamma}(t)$,*

*(b) the Riemannian metric $\Phi^* g$ is the identity matrix at $(t, 0)$,i.e., along $\bar{\gamma}$.*

*Proof of the Lemma.* For every $\varepsilon > 0$ we can find coordinates in the cylinder $I_\varepsilon \times B_\varepsilon^{n-1}$ such that, in these coordinates, our curve $\bar{\gamma}$ is rectified $\bar{\gamma}(t) = (t, 0)$ and has length one.

Our normalization of the curve $\bar{\gamma}$ implies that for the matrix representing the Riemannian metric $\Phi^* g$ in these coordinates satisfies

$$\Phi^* g = \begin{pmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \end{pmatrix}, \qquad \text{with} \quad G_{11}(x,0) = 1,$$

where $G_{ij}$, for $i,j = 1,2$, are the blocks of $\Phi^* g$ corresponding to the splitting $\mathbb{R}^n = \mathbb{R} \times \mathbb{R}^{n-1}$ defined in (12.90). For every point $(x,0)$ let us consider the orthogonal complement $T(x,0)$ of the tangent vector $e_1 = \partial_x$ to $\bar{\gamma}$ with respect to $G$. It can be written as follows (here we denote by $\langle \cdot \,|\, \cdot \rangle$ the Euclidean product in $\mathbb{R}^n$)

$$T(x,0) = \left\{ (\langle v_x \,|\, y \rangle, y) \mid y \in \mathbb{R}^{n-1} \right\},$$

for some family of vectors $v_x \in \mathbb{R}^{n-1}$, depending smoothly with respect to $x$. Indeed it is easily checked that $v_x = -G_{21}^1(x,0)$, where $G_{21}^1$ denotes the first column of the $(n-1) \times (n-1)$ matrix $G_{21}$.

Let us consider now the smooth change of coordinates

$$\Psi : \mathbb{R}^n \to \mathbb{R}^n, \qquad \Psi(x,y) = (x - \langle v_x \,|\, y \rangle, y).$$

Fix $\varepsilon > 0$ small enough such that the restriction of $\Psi$ to $I_\varepsilon \times B_\varepsilon^{n-1}$ is invertible. Notice that this is possible since

$$\det D\Psi(x,y) = 1 - \langle \partial_x v_x \,|\, y \rangle .$$

It is not difficult to check that, in the new variables (that we still denote by the same symbol), one has

$$G(x,0) = \begin{pmatrix} 1 & 0 \\ 0 & M(x,0) \end{pmatrix},$$

where $M(x,0)$ is a positive definite matrix for all $x \in I_\varepsilon$. With a linear change of cooordinates in the $y$ space

$$(x,y) \mapsto (x, M(x,0)^{1/2} y),$$

we can finally normalize the matrix in such a way that $G(x,0) = \mathrm{Id}$ for all $x \in I_\varepsilon$. $\square$

We are now ready to prove the theorem. We check the equivalence between the two notions of local minimality in the coordinate set, denoted $(x,y)$, defined by the previous lemma. Notice that the notion of local minimality is independent on the coordinates.

Given an admissible curve $\gamma(t) = (x(t), y(t))$ contained in the cylinder $I_\varepsilon \times B_\varepsilon^{n-1}$ and satisfying $\gamma(0) = (0,0)$ and $\gamma(1) = (1,0)$ and denoting the reference trajectory $\bar{\gamma}(t) = (t,0)$ we have that

$$\begin{aligned} \|\gamma - \bar{\gamma}\|_{W^{1,2}}^2 &= \int_0^1 |\dot{x}(t) - 1|^2 + |\dot{y}(t)|^2 dt \\ &= \int_0^1 |\dot{x}(t)|^2 + |\dot{y}(t)|^2 dt - 2 \int_0^1 \dot{x}(t) dt + 1 \\ &= \int_0^1 |\dot{x}(t)|^2 + |\dot{y}(t)|^2 dt - 1, \end{aligned}$$

where we used that $x(0) = 0$ and $x(1) = 1$ since $\gamma$ satisfies the boundary conditions. If we denote by

$$J(\gamma) = \int_0^1 \langle G(\gamma(t))\dot{\gamma}(t), \dot{\gamma}(t)\rangle \, dt, \qquad J_e(\gamma) = \int_0^1 |\dot{x}(t)|^2 + |\dot{y}(t)|^2 dt \tag{12.91}$$

respectively the energy of $\gamma$ and the "Euclidean" energy, we have $\|\gamma - \bar{\gamma}\|^2_{W^{1,2}} = J_e(\gamma) - 1$ and the $W^{1,2}$-local minimality can be rewritten as follows:

($*$) there exists $\varepsilon > 0$ such that for every $\gamma$ admissible and $J_e(\gamma) \leq 1 + \varepsilon$ one has $J(\gamma) \geq 1$.

Next we build the following neighborhood of $\bar{\gamma}$: for every $\delta > 0$ define $\mathcal{A}_\delta$ as the set of admissible curves $\gamma(t) = (x(t), y(t))$ in $I_\varepsilon \times B^{n-1}_\varepsilon$ such that the dilated curve $\gamma_\delta(t) = (x(t), \frac{1}{\delta}y(t))$ is still contained in the cylinder. This implies that in particular that $\gamma$ is contained in $I_\varepsilon \times B^{n-1}_{\delta\varepsilon}$. Notice that $\mathcal{A}_\delta \subset \mathcal{A}_{\delta'}$ whenever $\delta < \delta'$. Moreover, every curve that is $\varepsilon\delta$ close to $\bar{\gamma}$ in the $C^0$-topology is contained in $\mathcal{A}_\delta$.

It is then sufficient to prove that, for $\delta > 0$ small enough, for every $\gamma \in \mathcal{A}_\delta$ one has $\ell(\gamma) \geq \ell(\bar{\gamma})$. Indeed it is enough to check that $J(\gamma) \geq J(\bar{\gamma})$. Let us consider two cases

(i) $\gamma \in \mathcal{A}_\delta$ and $J_e(\gamma) \leq 1 + \varepsilon$. In this case ($*$) implies that $J(\gamma) \geq 1$.

(ii) $\gamma \in \mathcal{A}_\delta$ and $J_e(\gamma) > 1 + \varepsilon$. In this case we have $G(x, 0) = \mathrm{Id}$ and, by smoothness of $G$, we can write for $(x, y) \in I_\varepsilon \times B^{n-1}_{\delta\varepsilon}$ and $\delta \to 0$

$$\langle G(x, y)v, v\rangle = (1 + O(\delta)) \langle v, v\rangle \,,$$

where $O(\delta)$ is uniform with respect to $(x, y)$. Since $\gamma \in \mathcal{A}_\delta$ implies that $\gamma$ is contained in $I_\varepsilon \times B^{n-1}_{\delta\varepsilon}$ we can deduce for $\delta \to 0$

$$J(\gamma) = J_e(\gamma)(1 + O(\delta)) \geq (1 + \varepsilon)(1 + O(\delta))$$

and one can choose $\bar{\delta} > 0$ small enough such that the last quantity is strictly bigger than one.

This proves that there exists $\bar{\delta} > 0$ such every admissible curve $\gamma \in \mathcal{A}_{\bar{\delta}}$ is longer than $\bar{\gamma}$. □

*Remark* 12.60. Notice that this result implies in particular Theorem 4.62, since normal extremals are always smooth. Nevertheless, the argument of Theorem 4.62 can be adapted for more general coercive functional (see [AS04]), while this proof use specific estimates that hold only for our explicit cost (i.e., the distance).

## 12.8 Non-minimality of corners

Is every sub-Riemannian length-minimizer smooth? We still do not know if this is always true. We proved that normal Pontryagin extremals are smooth, as well as nice abnormals. It is easy to construct non-smooth abnormal extremal trajectories, but all known examples are not length-minimizers. An example of non-smooth abnormal is given in Section 12.6.1. The trajectory is a local length-minimizer in the $L^\infty$-topology for controls but it is not a length-minimizer (and not a local length-minimizer in the $L^p$-topology, for every $p < \infty$).

The following important regularity result says that length-minimizers cannot have "corner" singularities. For simplicity, we state it for piecewise smooth horizontal curves. The full statement can be found in [HLD16].

**Theorem 12.61.** *Any piecewise smooth length-minimizer parametrized by arclength is of class $C^1$.*

*Proof.* Assume that $\gamma : [-T, T] \to M$ is a length-minimizer that is piecewise smooth but it is not of class $C^1$. It is not restrictive, by taking a suitable restriction of the curve, to assume that $\gamma$ is the concatenation of two smooth horizontal curves, i.e., there exist $\gamma_1, \gamma_2 : [0, T] \to M$ smooth horizontal curves parametrized by arclength such that

$$\gamma(t) = \begin{cases} \gamma_1(-t), & t \in [-T, 0] \\ \gamma_2(t), & t \in [0, T] \end{cases}, \qquad \dot{\gamma}_1(0) + \dot{\gamma}_2(0) \neq 0. \tag{12.92}$$

We have to prove that, for every $\varepsilon > 0$, the horizontal curve $\gamma|_{[-\varepsilon,\varepsilon]}$ is not a length-minimizer. Notice that this is equivalent to show that $d(\gamma_1(\varepsilon), \gamma_2(\varepsilon)) < 2\varepsilon$, where $d$ denotes the sub-Riemannian distance.

We split the proof into two parts: (a) first we consider the case of linearly independent $\dot{\gamma}_1(0)$ and $\dot{\gamma}_2(0)$, (b) we then explain the simpler case $\dot{\gamma}_1(0) = \dot{\gamma}_2(0)$, when the concatenation of the curves has a cusp. The proof of the main case (a) is divided in several steps.

(a.1) Let $f_1, f_2$ be two smooth and horizontal vector fields such that for $t \in [0, T]$ and $i = 1, 2$ one has

$$\dot{\gamma}_i(t) = f_i(\gamma_i(t)).$$

Assume by contradiction that $d(\gamma_1(t), \gamma_2(t)) = 2t$ for all sufficiently small $t > 0$. We are going to show that this assumption leads to a contradiction.

Let $q = \gamma_1(0) = \gamma_2(0)$ and fix a neighborhood $O_q$ together with a set of privileged coordinates. Let $\delta_\varepsilon : O_q \to O_q$, for $\varepsilon > 0$, be the associated dilation (see Chapter 10). We set

$$f_i^\varepsilon = \varepsilon \delta_{\frac{1}{\varepsilon}*} f_i, \quad i = 1, 2;$$

and by $d_\varepsilon$ the corresponding distance for every $q_1, q_2 \in O_q$ given by

$$d_\varepsilon(q_1, q_2) := \frac{1}{\varepsilon} d(\delta_\varepsilon(q_1), \delta_\varepsilon(q_2)).$$

Finally we set $\gamma_i^\varepsilon(t) = e^{tf_i^\varepsilon}$. It follows that $d_\varepsilon(\gamma_1^\varepsilon(t), \gamma_2^\varepsilon(t)) = 2t$ for every $\varepsilon > 0$. Moreover, thanks to the results of Section 10.5, $f_i^\varepsilon$ converges to $\widehat{f}_i$ in the $C^\infty$-topology and $d_\varepsilon$ uniformly converges to $\widehat{d}$ as $\varepsilon \to 0$, where the vector fields $\widehat{f}_1$ and $\widehat{f}_2$, are two of generators of the Carnot algebra acting on the nonholonomic tangent space at $q$ and $\widehat{d}(\cdot, \cdot)$ is the metric on the nonholonomic tangent space at $q$. We obtain that for all $t$ one has

$$\widehat{d}\left(e^{t\widehat{f}_1}(q), e^{t\widehat{f}_2}(q)\right) = 2t.$$

(a.2) The nonholonomic tangent space is a homogeneous space $G/H$ of a Carnot group $G$, and the distance $\widehat{d}(\widehat{q}_1, \widehat{q}_2)$ is, by definition, the minimum of the distances in the Carnot group $H$ between elements of the stable subgroups of the points $\widehat{q}_1, \widehat{q}_2$ for this action. We keep symbol $\widehat{d}$ for the distance in the Carnot group $G$. It follows that

$$\widehat{d}\left(e^{t\widehat{f}_1}, e^{t\widehat{f}_2}\right) = 2t.$$

The equality follows since the concatenation of the curves $\tau \to e^{(t-\tau)\hat{f}_1}$ and $\tau \to e^{\tau \hat{f}_2}$, $0 \leq \tau \leq t$, equals $2t$.

(a.3) The Carnot algebra may have more than two generators. Let us consider the subalgebra generated by $\widehat{f}_1, \widehat{f}_2$ and the correspondent Carnot subgroup. Given two points in the subgroup, the distance between the points in the subgroup is greater or equal than the distance in the ambient group.

(a.4) This is the key step of the proof and we would like to simplify the notations. Let $G$ be a Carnot group with a Carnot algebra $\mathfrak{g}$. We assume that $\mathfrak{g}$ is a Carnot algebra with step $k$ and two generators, namely

$$\mathfrak{g} = \mathfrak{g}_1 \oplus \cdots \oplus \mathfrak{g}_k, \quad \mathfrak{g} = \mathrm{Lie}\{\mathfrak{g}_1\}, \quad \mathfrak{g}_1 = \mathrm{span}\{x_1, x_2\}.$$

We also assume that $|x_1| = |x_2| = 1$, but $x_1$ might not be orthogonal to $x_2$. We denote the sub-Riemannian distance in $G$ by $d(\cdot, \cdot)$ (without "hat"). To prove the statement in the no-cusp case it is then sufficient to prove the next claim

**Proposition 12.62.** *The horizontal curve in $G$ defined by*

$$\widehat{\gamma}(t) = \begin{cases} e^{-tx_1}, & t \in [-1, 0] \\ e^{tx_2}, & t \in [0, 1] \end{cases} \tag{12.93}$$

*is not a length-minimizer, i.e., $d(e^{x_1}, e^{x_2}) < 2$.*

*Proof.* We prove this statement by induction on the step $k$ of $G$. For $k = 2$, $G$ is unique and isomorphic to the Heisenberg group, where we already computed all length-minimizers, and they are smooth.

Assume now that the statement is valid for every $(k-1)$-step Carnot group and let us prove it for every Carnot group of step $k$. Note that $\mathfrak{g}_k$ is contained in the center of $\mathfrak{g}$ and $e^{\mathfrak{g}_k}$ takes part of the center of $G$. Then $G/e^{\mathfrak{g}_k}$ is a Carnot group with a step $(k-1)$ Carnot algebra $\mathfrak{g}_1 \oplus \cdots \oplus \mathfrak{g}_{k-1}$. Moreover, the sub-Riemannian distance between two points in $G/e^{\mathfrak{g}_k}$ is by construction the minimum of the distances between the points of the correspondent residue classes. Taking into account the left-invariance of the distance, we can write:

$$d(e^{\mathfrak{g}_k} q_1, e^{\mathfrak{g}_k} q_2) = \min_{z \in \mathfrak{g}_k} d(e^z q_1, q_2).$$

Our induction assumption implies that there exists $z \in \mathfrak{g}_k$ and $\nu > 0$ such that

$$d(e^z e^{x_1}, e^{x_2}) = 2 - \nu,$$

Moreover, left-invariance of the distance implies that

$$d(e^z e^{x_1}, e^{x_2}) = d(1, e^{-x_1} e^{-z} e^{x_2}).$$

The trick is to show explicitly a competitor by playing with horizontal curves. We start by adding a short piece of the form $t \mapsto e^{-t\varepsilon^k z}$, $0 \le t \le 1$, as in Figure 12.2.

We claim that, for $\varepsilon > 0$ small enough, one has

$$d(e^{x_1}, e^{-\varepsilon^k z} e^{x_2}) \le 2 - \varepsilon\nu. \tag{12.94}$$

Indeed, again by left-invariance, we have

$$d(e^{x_1}, e^{-\varepsilon^k z} e^{x_2}) = d(1, e^{-x_1} e^{-\varepsilon^k z} e^{x_2}) \tag{12.95}$$

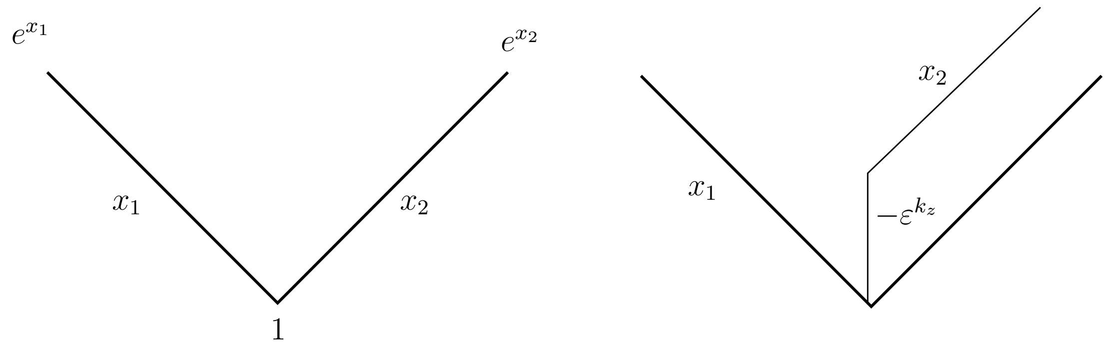



Figure 12.2: Adding one piece

and we can rewrite

$$e^{-x_1}e^{-\varepsilon^k z}e^{x_2} = e^{(\varepsilon-1)x_1}\left(e^{-\varepsilon x_1}e^{-\varepsilon^k z}e^{\varepsilon x_2}\right)e^{(1-\varepsilon)x_2}. \tag{12.96}$$

Notice now that

$$e^{-\varepsilon x_1}e^{-\varepsilon^k z}e^{\varepsilon x_2} = \delta_\varepsilon\left(e^{x_1}e^{-z}e^{x_2}\right), \tag{12.97}$$

where $\delta_\varepsilon$ is the dilation of the Carnot group of factor $\varepsilon > 0$. Moreover, $d(1, \delta_\varepsilon(q)) = \varepsilon d(1, q)$, for every $q \in G$. The triangle inequality for left-invariant metrics reads:

$$d(1, ab) \leq d(1, a) + d(1, b), \qquad a, b \in G. \tag{12.98}$$

Combining the previous identities from (12.95) to (12.98) one finally gets

$$\begin{aligned} d(1, e^{-x_1}e^{-z}e^{x_2}) &\leq d(1, e^{(\varepsilon-1)x_1}) + \varepsilon(2-\nu) + d(1, e^{(1-\varepsilon)x_2}) \\ &= (1-\varepsilon) + \varepsilon(2-\nu) + (1-\varepsilon) = 2 - \varepsilon\nu. \end{aligned}$$

which proves (12.94).

Next we would like to compensate the deviation of the end-point of the curve produced by the inserted piece $e^{-\varepsilon^k z}$. To this end, we insert some pieces of the form $e^{\varepsilon^k y_i}$, where $y_i \in \mathfrak{g}_{k-1}$. Each piece increases the final distance between end-points by $O(\varepsilon^{\frac{k}{k-1}})$ since $e^{\varepsilon^k y_i} = \delta_{\varepsilon^{\frac{k}{k-1}}}(e^{y_i})$. Hence the distance between the end-points of the resulting curve remains smaller than 2 if $\varepsilon$ is small enough, since $\frac{k}{k-1} > 1$.

It is actually sufficient to add three pieces as in Figure 12.3. More precisely, we want to find $y_1, y_2, y_3 \in \mathfrak{g}_{k-1}$ such that

$$e^{x_1}e^{\varepsilon^k y_1}e^{-x_1}e^{-\varepsilon^k z}e^{\frac{1}{2}x_2}e^{\varepsilon^k y_2}e^{\frac{1}{2}x_2}e^{\varepsilon^k y_3} = e^{x_2}, \tag{12.99}$$

for all $\varepsilon > 0$. To find a solution to equation (12.99) we use the fact that $e^{-\varepsilon^k z}$ commutes with all elements of the group and re-write (12.99) in the form:

$$\left(e^{x_1}e^{\varepsilon^k y_1}e^{-x_1}\right)\left(e^{\frac{1}{2}x_2}e^{\varepsilon^k y_2}e^{-\frac{1}{2}x_2}\right)\left(e^{x_2}e^{\varepsilon^k y_3}e^{-x_2}\right) = e^{\varepsilon^k z}. \tag{12.100}$$

Now we use a universal identity

$$e^x e^y e^{-x} = e^{(e^{\operatorname{ad} x} y)}.$$

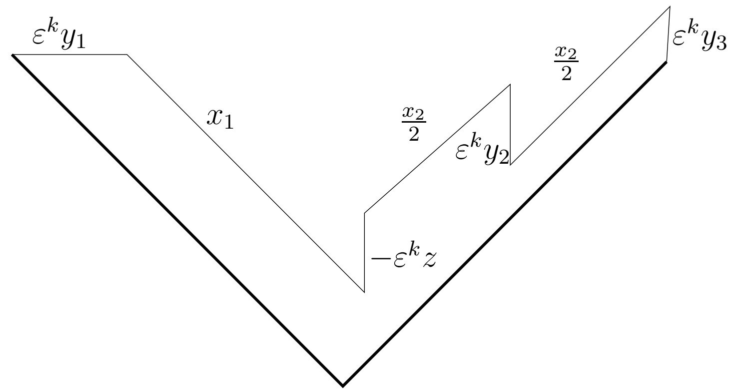



Figure 12.3: Adding more pieces

Moreover, since $\mathfrak{g}$ is a step $k$ nilpotent Lie algebra and $y_i \in \mathfrak{g}_{k-1}$, we obtain:

$$e^{\mathrm{ad} x_j} y_i = y_i + \frac{1}{2}[x_j, y_i], \qquad i = 1, 2, 3, \; j = 1, 2.$$

Notice that all elements in the set $\{y_i, [x_j, y_i] : i = 1, 2, 3, j = 1, 2\}$ are mutually commuting because $[y_i, y_j] \in \mathfrak{g}_{2k-2}$, and $\mathfrak{g}_{2k-2} = \{0\}$ since $k \geq 3$. It follows that the product of the exponentials is equal to the exponential of the sum. This permits finally to rewrite (12.101) as

$$e^{\varepsilon^k \left( \sum_{i=1}^{3} y_i + \frac{1}{2}[x_1, y_1] + \frac{1}{4}[x_2, y_2] + \frac{1}{2}[x_2, y_3] \right)} = e^{\varepsilon^k z}, \tag{12.101}$$

that is equivalent to the following system, obtained by separating the part in $\mathfrak{g}_{k-1}$ and the one in $\mathfrak{g}_k$.

$$\sum_{i=1}^{3} y_i = 0, \qquad [x_1, y_1] + \frac{1}{2}[x_2, y_2] + [x_2, y_3] = 2z.$$

Replacing $y_3 = -y_1 - y_2$ in the second equation, it is sufficient to find $y_1, y_2 \in \mathfrak{g}_{k-1}$ such that

$$[x_1 - x_2, y_1] - \frac{1}{2}[x_2, y_2] = 2z. \tag{12.102}$$

The existence of a pair $y_1, y_2$ satisfying (12.102) now follows from the fact that $z \in \mathfrak{g}_k$ and the relations

$$\mathfrak{g}_1 = \operatorname{span}\{x_1, x_2\} = \operatorname{span}\{x_1 - x_2, x_2\}, \qquad [\mathfrak{g}_1, \mathfrak{g}_{k-1}] = \mathfrak{g}_k. \qquad \square$$

(b). Now we prove Theorem 12.61 in the case of a cusp:

$$\gamma(t) = \begin{cases} \gamma_1(-t), & t \in [-T, 0] \\ \gamma_2(t), & t \in [0, T] \end{cases}, \qquad \dot{\gamma}_1(0) = \dot{\gamma}_2(0). \tag{12.103}$$

Under these assumptions, there exist a horizontal smooth vector field $f_1$ and smooth control $t \mapsto u(t)$ such that

$$\dot{\gamma}_1(t) = f_1(\gamma_1(t)), \quad \dot{\gamma}_2(t) = f_1(\gamma_2(t)) + t f_{u(t)}(\gamma_2(t)).$$

where $f_u = u_1 f_1 + u_2 f_2$. If the curve $\gamma$ is a length-minimizer, then $d(\gamma_1(t), \gamma_2(t)) = 2t$, for all $t > 0$. Applying the blow-up procedure and the lift to the Carnot group as in steps (a.1) and (a.2) of the proof in the no-cusp case, one obtains that

$$\widehat{d}\left(e^{t\widehat{f}_1}, \overrightarrow{\exp}\int_0^t \widehat{f}_1 + \tau \widehat{f}_{u(\tau)}\, d\tau\right) = 2t.$$

where $\widehat{d}$ is the left-invariant distance on the Carnot group. In particular

$$\widehat{d}\left(e^{t\widehat{f}_1}, \overrightarrow{\exp}\int_0^t \widehat{f}_1 + \tau \widehat{f}_{u(\tau)}\, d\tau\right) = \widehat{d}\left(1, e^{-t\widehat{f}_1}\overrightarrow{\exp}\int_0^t \widehat{f}_1 + \tau \widehat{f}_{u(\tau)}\, d\tau\right)$$

Moreover, setting $g_\tau^t := \tau e^{(t-\tau)\mathrm{ad}\widehat{f}_1}\widehat{f}_{u(\tau)}$, we have

$$e^{-t\hat{f}_1}\overrightarrow{\exp}\int_0^t \widehat{f}_1 + \tau \widehat{f}_{u(\tau)}\, d\tau = \overrightarrow{\exp}\int_0^t g_\tau^t\, d\tau,$$

according to the variations formula (see Proposition 6.16 and Exercice 6.17). Hence, combining the above computations, we have for all $t > 0$

$$\widehat{d}\left(1, \overrightarrow{\exp}\int_0^t g_\tau^t\, d\tau\right) = 2t. \tag{12.104}$$

If the Carnot group is of step $k$, then:

$$g_\tau^t = \sum_{i=0}^{k-1} \frac{\tau(t-\tau)^i}{i!}(\mathrm{ad}\widehat{f}_1)^i \widehat{f}_{u(\tau)}.$$

The $i$-th term of the sum defining $g_\tau^t$ belongs to the $(i+1)$-th stratum of the Carnot algebra, and has order $t^{i+1}$ for $t \to 0$ (notice that $0 \le \tau \le t$). Hence the $i$-th level component of $\overrightarrow{\exp}\int_0^t g_\tau^t\, d\tau$ in a privileged coordinates on the Carnot group has order $t^{i+1}$ as $t \to 0$. Indeed, this component is the value at $t$ of a solution of the ordinary differential equation starting from the origin and whose right-hand side has order $t^i$ as $t \to 0$.

The ball-box estimates imply that there exists a constant $C > 0$ such that

$$\widehat{d}\left(1, \overrightarrow{\exp}\int_0^t g_\tau^t\, d\tau\right) \le C t^{\frac{k}{k+1}}.$$

which contradicts (12.104) for $t$ small enough, since $\frac{k}{k+1} < 1$. □

## 12.9 Bibliographical note

The theory of the second variation and its relation to the study of abnormal extremal trajectories for control systems, including Goh and Legendre conditions, have been initiated in [Goh66, KKM67, Kre73, Kre77, AG76, Agr77]. The existence of strictly abnormal length-minimizers in sub-Riemannian geometry is due to Montgomery [Mon94], for a discussion see also [Mon02, Rif14].

The results contained in Theorems 12.6 and 12.12 have been proved in [AL09].

When the rank of the distribution is larger or equal than 3, Goh conditions are not satisfied for generic distributions as proved in [CJT06]. On the other hand, in the case of Carnot groups, for big codimension of the distribution, abnormal minimizers satisfying Goh conditions always appear [AG01a].

The definition of nice abnormals was introduced in [LS95]. The proof of the smoothness and minimality of nice abnormals is contained in [LS95] and [AS95]. The equivalence between $C^0$ and $W^{1,2}$ local length-minimality is discussed in [Agr98c].

The non-minimality of corners have been proved in [HLD16], refining a "cut-and-adjust" technique first used in [LM08] to exclude corners in some class of sub-Riemannian structures. More recently, some results have been obtained about the $C^1$ regularity of non-nice abnormal length-minimizers [Mon14, BFPR18, BCJ$^+$19].

The question whether all length-minimizers are smooth (or at least of class $C^1$) is still open.

# Chapter 13

# Some model spaces

In this chapter we are going to construct explicitly the full set of optimal arclength geodesics (the so-called *optimal synthesis*) starting from a point for certain relevant sub-Riemannian structures.

We start with a class of problems in which all computations can be done explicitly, namely Carnot groups of step 2. In this setting we give a general formula for Pontryagin extremals and we explicitly compute them in the case of multi-dimensional Heisenberg groups, together with the optimal synthesis. For free Carnot groups of step two we provide a description of the intersection of the cut locus with the vertical space and we give an explicit formula for the sub-Riemannian distance from the origin to those points.

Then we present a technique to identify the cut locus, that generalizes a classical technique used in Riemannian geometry due to Hadamard. We then apply in full detail this technique to compute the optimal synthesis for two cases: (i) the Grushin plane; (ii) the left-invariant sub-Riemannian structure on $SU(2)$ with the metric induced by the Killing form. The same technique can be applied to study $SO(3)$ and $SL(2)$ (again with the metric induced by the Killing form). These last two cases are left as exercises. The optimal synthesis for $SO(3)$ together with the one for $SO_+(2,1)$ is then obtained using an alternative (and more geometric) approach based on the Gauss-Bonnet Theorem.

We conclude by treating two relevant cases, namely the left-invariant sub-Riemannian structure on $SE(2)$ and the Martinet flat sub-Riemannian structure. For these cases we compute geodesics (that can be obtained explicitly in terms of elliptic functions) and we state the results concerning the cut locus. Their proof require an estimation of the conjugate locus that can be obtained via a fine analysis of properties of elliptic functions and it is outside the purpose of this book.

Let us recall the definition of cut time and cut locus.

**Definition 13.1.** Consider a sub-Riemannian manifold $M$ that is complete as metric space. Let $\gamma$ be an arclength maximal (i.e., non extendable) geodesic. The cut time along $\gamma$ is

$$t_{\text{cut}} := \sup\{t > 0 : \gamma|_{[0,t]} \text{ is length-minimizing}\}.$$

If $t_{\text{cut}} < +\infty$ we say that $\gamma(t_{\text{cut}})$ is the cut point of $\gamma(0)$ along $\gamma$. If $t_{\text{cut}} = +\infty$ we say that $\gamma$ has no cut point. We denote by $\text{Cut}_{q_0}$ the set of all cut points of geodesics starting from a point $q_0 \in M$.

*Remark* 13.2. Notice that with this definition, the starting point is never included in the cut locus.

**Definition 13.3.** Consider a sub-Riemannian manifold complete as metric space and fix a point $q_0 \in M$. The *optimal synthesis* from $q_0$ is the collection of all arclength geodesics starting from $q_0$ together with their cut time.

Given a sub-Riemannian manifold, constructing explicitly the optimal synthesis from a point $q_0$ is in general a very difficult problem. The main difficulties are the following:

(A) the integration of the Hamiltonian equations giving normal Pontryagin extremals. In most cases such equations are not integrable;

(B) the identification of abnormal extremals and the study of their optimality;

(C) the evaluation of the cut time for every Pontryagin extremal. Such problem is particularly difficult since in principle for every point of $M$ one should find all Pontryagin extremals reaching that point (and hence in particular one should be able to invert the exponential map) and then one should choose the one having the smaller cost (i.e., the smaller distance from $q_0$).

For the reasons explained above, only few optimal syntheses are known in sub-Riemannian geometry. Such examples all concern left-invariant sub-Riemannian structures on Lie groups or their projections to homogenous spaces.

## 13.1 Carnot groups of step 2

A Carnot group of step 2 is a Lie group structure $G$ on $\mathbb{R}^n$ such that its Lie algebra $\mathfrak{g}$ satisfies (cf. also Section 7.5)

$$\mathfrak{g} = \mathfrak{g}_1 \oplus \mathfrak{g}_2, \qquad [\mathfrak{g}_1, \mathfrak{g}_1] = \mathfrak{g}_2, \qquad [\mathfrak{g}_1, \mathfrak{g}_2] = [\mathfrak{g}_2, \mathfrak{g}_2] = 0. \tag{13.1}$$

The group $G$ is endowed by the left-invariant sub-Riemannian structure induced by the choice of a scalar product $\langle \cdot | \cdot \rangle$ on the distribution $\mathfrak{g}_1$, that is bracket-generating of step 2 thanks to (13.1).

Consider a basis of left-invariant vector fields (on $\mathbb{R}^n$) of $\mathfrak{g}$ such that

$$\mathfrak{g}_1 = \operatorname{span}\{X_1, \dots, X_m\}, \qquad \mathfrak{g}_2 = \operatorname{span}\{Z_1, \dots, Z_{n-m}\},$$

where $\{X_1, \dots, X_m\}$ define an orthonormal frame for $\langle \cdot | \cdot \rangle$ on the distribution $\mathfrak{g}_1$. Such a basis will be referred also as an *adapted basis*. We can write the commutation relations as follows

$$\begin{cases} [X_i, X_j] = \sum_{\ell=1}^{n-m} c_{ij}^\ell Z_\ell, \qquad i, j = 1, \dots, m, \quad \text{with} \quad c_{ij}^\ell = -c_{ji}^\ell, \\ [X_i, Z_j] = [Z_j, Z_\ell] = 0, \qquad i = 1, \dots, m, \qquad j, \ell = 1, \dots, n-m. \end{cases} \tag{13.2}$$

Given an adapted basis, we can introduce the family of skew-symmetric matrices $\{C_1, \dots, C_{n-m}\}$ encoding the structure constants of the Lie algebra, defined by $C_\ell = (c_{ij}^\ell)$, for $\ell = 1, \dots, n-m$, and the corresponding subspace of skew-symmetric operators on $\mathfrak{g}_1$ that are represented by linear combination of this family of matrices

$$\mathcal{C} := \operatorname{span}\{C_1, \dots, C_{n-m}\} \subset \mathfrak{so}(\mathfrak{g}_1) \tag{13.3}$$

We stress that, since the vector fields of the basis are left-invariant, then $c_{ij}^\ell$ are constant.

**Definition 13.4.** A Carnot algebra of step 2 is called *free* if $\mathcal{C} = \mathfrak{so}(\mathfrak{g}_1)$ and the matrices $C_\ell = (c_{ij}^\ell)$, for $\ell = 1, \dots, n-m$, define a basis of $\mathcal{C}$.

A representation of the Lie algebra defined above is given by the family of vector fields on $\mathbb{R}^n = \mathbb{R}^m \oplus \mathbb{R}^{n-m}$ (using coordinates $g = (x, z) \in \mathbb{R}^m \oplus \mathbb{R}^{n-m}$)

$$X_i = \frac{\partial}{\partial x_i} - \frac{1}{2} \sum_{j=1}^{m} \sum_{\ell=1}^{n-m} c_{ij}^\ell x_j \frac{\partial}{\partial z_\ell}, \qquad i = 1, \dots, m, \tag{13.4}$$

$$Z_\ell = \frac{\partial}{\partial z_\ell}, \qquad \ell = 1, \dots, n-m. \tag{13.5}$$

The group law on $G$, when identified with $\mathbb{R}^n = \mathbb{R}^m \oplus \mathbb{R}^{n-m}$, reads as follows

$$(x, z) * (x', z') = \left( x + x', z + z' + \frac{1}{2} C x \cdot x' \right),$$

where we denoted for the $(n-m)$-tuple $C = (C_1, \dots, C_{n-m})$ of $m \times m$ matrices, the product

$$Cx \cdot x' = (C_1 x \cdot x', \dots, C_{n-m} x \cdot x') \in \mathbb{R}^{n-m}.$$

and $a \cdot b$ denotes here the Euclidean inner product between two vectors $a, b \in \mathbb{R}^m$. The choice of the linearly independent vector fields $\{X_1, \dots, X_m, Z_1, \dots, Z_{n-m}\}$ induce corresponding coordinates on $T^*G$

$$h_i(\lambda) = \langle \lambda, X_i(g) \rangle, \qquad w_\ell(\lambda) = \langle \lambda, Z_\ell(g) \rangle.$$

The functions $\{h_i, w_\ell\}$ defines a system of global coordinates on the fibers of $T^*G$. In what follows it is convenient to use $(x, y, h, w)$ as global coordinates on the whole $T^*G$, identified with $\mathbb{R}^{2n}$.

Normal extremal trajectories are projections on $M$ of integral curves of the sub-Riemannian Hamiltonian in $T^*G$:

$$H = \frac{1}{2} \sum_{i=1}^{m} h_i^2. \tag{13.6}$$

Suppose now that $\lambda(t) = (x(t), z(t), h(t), w(t)) \in T^*G$ is a normal Pontryagin extremal. The equation $\dot{\lambda}(t) = \vec{H}(\lambda(t))$ is rewritten as follows

$$\begin{cases} \dot{x}_i = h_i \\ \dot{z}_\ell = -\frac{1}{2} \sum_{i,j=1}^{m} c_{ij}^\ell h_i x_j \end{cases} \qquad \begin{cases} \dot{h}_i = -\sum_{\ell=1}^{n-m} \sum_{j=1}^{m} c_{ij}^\ell h_j w_\ell \\ \dot{w}_\ell = 0 \end{cases} \tag{13.7}$$

where we used the relation $u_i(t) = h_i(\lambda(t))$ satisfied by normal extremals and the property $\dot{a} = \{H, a\}$ for the derivative of a smooth function $a$ along solutions of the Hamiltonian vector field $\vec{H}$, giving

$$\begin{cases} \dot{h}_i = \{H, h_i\} = -\sum_{j=1}^{m} \{h_i, h_j\} h_j = -\sum_{\ell=1}^{n-m} \sum_{j=1}^{m} c_{ij}^\ell h_j w_\ell \\ \dot{w}_\ell = \{H, w_\ell\} = 0. \end{cases} \tag{13.8}$$

Recall moreover that $H$ is constant along solutions, in particular $H = 1/2$ along extremals parametrized by arclength. From (13.8) we easily get that $w_\ell$ is constant for every $\ell = 1, \dots, n-m$, hence the first equation rewrites as an autonomous linear equation for $h = (h_1, \dots, h_m) \in \mathbb{R}^m$

$$\dot{h} = -\left( \sum_{\ell=1}^{n-m} w_\ell C_\ell \right) h,$$

It follows that

$$h(t) = e^{-t\Omega_w} h(0), \qquad \Omega_w := \sum_{\ell=1}^{n-m} w_\ell C_\ell. \tag{13.9}$$

From this expression one finds the $x$-component

$$x(t) = x(0) + \int_0^t e^{-s\Omega_w} h(0) ds.$$

Finally, injecting the above expression in the equation of $z$, one can recover the full normal extremal trajectory by integration.

## 13.2 Multi-dimensional Heisenberg groups

In this section we specify the previous analysis and provide explicit computation for the case of multidimensional Heisenberg groups. These are step-2 Carnot group structures on $\mathbb{R}^{2l+1}$ where

$$\mathfrak{g} = \mathfrak{g}_1 \oplus \mathfrak{g}_2, \qquad \dim \mathfrak{g}_1 = 2l, \qquad \dim \mathfrak{g}_2 = 1. \tag{13.10}$$

In particular the subspace $\mathcal{C}$ has dimension one and is spanned by a unique nonzero element in $\mathfrak{so}(\mathfrak{g}_1)$. Choosing a suitable basis

$$\mathfrak{g}_1 = \operatorname{span}\{X_1, \dots, X_{2l}\}, \qquad \mathfrak{g}_2 = \operatorname{span}\{Z\},$$

where $\{X_1, \dots, X_{2l}\}$ is chosen as an orthonormal basis for the scalar product $\langle \cdot \,|\, \cdot \rangle$ on the distribution $\mathfrak{g}_1$, we have that there exists a matrix $C = (c_{ij})$ satisfying

$$\begin{cases} \mathcal{D} = \operatorname{span}\{X_1, \dots, X_{2l}\}, \\ [X_i, X_j] = c_{ij} Z, \qquad i, j = 1, \dots, 2l, \quad \text{where} \quad c_{ij} = -c_{ji}, \\ [X_i, Z] = 0, \qquad i = 1, \dots, 2l. \end{cases} \tag{13.11}$$

Notice that this structure is free if and only if $l = 1$ and is contact if and only if $C$ is non-degenerate.

Recall that $C$ is a real skew-symmetric matrix, hence there exist $\alpha_1, \dots, \alpha_l \in \mathbb{R}$ such that

$$\operatorname{spec}(C) = \{\pm i\alpha_1, \dots, \pm i\alpha_l\}.$$

Up to an orthogonal transformation in the distribution, we can choose the orthonormal basis of $\mathfrak{g}_1$ in such a way that the matrix $C$ has the following (block-diagonal) canonical form for skew-symmetric matrices

$$C = \begin{pmatrix} A_1 & & 0 \\ & \ddots & \\ 0 & & A_l \end{pmatrix}, \qquad \text{where} \qquad A_i := \begin{pmatrix} 0 & \alpha_i \\ -\alpha_i & 0 \end{pmatrix}, \qquad \alpha_i \geq 0. \tag{13.12}$$

*Remark* 13.5. Notice that $\alpha_i > 0$ for at least one value of $i$, otherwise the matrix $C$ would be zero. In what follows we restrict our attention to the case when all coefficients $\alpha_i$ are strictly positive. This is equivalent to require that the structure is of contact type. This implies that there are no non trivial abnormal extremals (cf. Section 4.3.3).

According to this decomposition we denote by $\{X_1, \ldots, X_l, Y_1, \ldots, Y_l, Z\}$ the orthonormal basis of $\mathfrak{g}_1$, where the vector fields satisfy the relations

$$
\begin{cases}
\mathfrak{g}_1 = \operatorname{span}\{X_1, \ldots, X_l, Y_1, \ldots, Y_l\}, & \\
[X_i, Y_i] = \alpha_i Z, & i = 1, \ldots, l, \\
[X_i, Y_j] = 0, & i \neq j, \\
[X_i, Z] = [Y_i, Z] = 0, & i = 1, \ldots, l,
\end{cases}
\tag{13.13}
$$

Denoting points $q = (x, y, z) \in \mathbb{R}^{2l+1}$, the group law is written in coordinates as follows

$$
q \cdot q' = \left( x + x', y + y', z + z' + \frac{1}{2} \sum_{i=1}^{l} \alpha_i (x_i x_i' - y_i y_i') \right). \tag{13.14}
$$

Finally, from (13.14), we get the coordinate expression of the left-invariant vector fields of the Lie algebra, namely

$$
\begin{aligned}
X_i &= \partial_{x_i} - \frac{1}{2} \alpha_i y_i \partial_z, \qquad i = 1, \ldots, l, \\
Y_i &= \partial_{y_i} + \frac{1}{2} \alpha_i x_i \partial_z, \qquad i = 1, \ldots, l, \\
Z &= \partial_z.
\end{aligned}
\tag{13.15}
$$

where $x = (x_1, \ldots, x_l), y = (y_1, \ldots, y_l) \in \mathbb{R}^l$ and $z \in \mathbb{R}$.

### 13.2.1 Pontryagin extremals in the contact case

Next we compute the exponential map $\exp_{q_0}$ where $q_0$ is the origin. Thanks to left-invariance of the structure this permits to recover normal Pontryagin extremals starting from every point. With an abuse of notation, we define the Hamiltonians (linear on fibers)

$$
u_i(\lambda) = \langle \lambda, X_i(q) \rangle, \qquad v_i(\lambda) = \langle \lambda, Y_i(q) \rangle, \qquad w(\lambda) = \langle \lambda, Z(q) \rangle.
$$

Suppose now that $\lambda(t) = (x(t), y(t), z(t), u(t), v(t), w(t)) \in T^*G$ is a normal Pontryagin extremal. The equation $\dot{\lambda}(t) = \vec{H}(\lambda(t))$ is rewritten as follows

$$
\begin{cases}
\dot{x}_i = u_i \\
\dot{y}_i = v_i \\
\dot{z} = -\frac{1}{2} \sum_{i=1}^{l} \alpha_i (u_i y_i - v_i x_i)
\end{cases}
\qquad
\begin{cases}
\dot{u}_i = -\alpha_i w v_i \\
\dot{v}_i = \alpha_i w u_i \\
\dot{w} = 0
\end{cases}
\tag{13.16}
$$

*Remark* 13.6. Notice that from (13.16) it follows that the sub-Riemannian length of a geodesic coincide with the Euclidean length of its projection on the horizontal subspace $(x_1, \ldots, x_n, y_1, \ldots, y_n)$.

$$
\ell(\gamma) = \int_0^T \left( \sum_{i=1}^{l} (u_i^2(t) + v_i^2(t)) \right)^{1/2} dt.
$$

Now we solve (13.16) with initial conditions (corresponding to arclength parametrized trajectories starting from the origin)

$$(x^0, y^0, z^0) = (0, 0, 0), \tag{13.17}$$
$$(u^0, v^0, w^0) = (u_1^0, \ldots, u_l^0, v_1^0, \ldots, v_l^0, w^0) \in S^{2l-1} \times \mathbb{R}. \tag{13.18}$$

Notice that $w = w^0$ is constant along the trajectory. We consider separately the two cases:

(a). If $w \neq 0$, we have

$$\begin{aligned} u_i(t) &= u_i^0 \cos(\alpha_i w t) - v_i^0 \sin(\alpha_i w t), \\ v_i(t) &= u_i^0 \sin(\alpha_i w t) + v_i^0 \cos(\alpha_i w t), \\ w(t) &= w. \end{aligned} \tag{13.19}$$

From (13.16) one easily gets

$$\begin{aligned} x_i(t) &= \frac{1}{\alpha_i w}(u_i^0 \sin(\alpha_i w t) + v_i^0 \cos(\alpha_i w t) - v_i^0), \\ y_i(t) &= \frac{1}{\alpha_i w}(-u_i^0 \cos(\alpha_i w t) + v_i^0 \sin(\alpha_i w t) + u_i^0), \\ z(t) &= \frac{1}{2} \sum_{i=1}^{l} \alpha_i \frac{(u_i^0)^2 + (v_i^0)^2}{\alpha_i^2 w^2} (\alpha_i w t - \sin(\alpha_i w t)). \end{aligned} \tag{13.20}$$

(b). If $w = 0$, we find equations of horizontal straight lines in direction of the vector $(u^0, v^0)$:

$$x_i(t) = u_i^0 t, \qquad y_i(t) = v_i^0 t, \qquad z(t) = 0.$$

To recover symmetry properties of the exponential map it is useful to rewrite (13.20) in polar coordinates, using the following change of variables

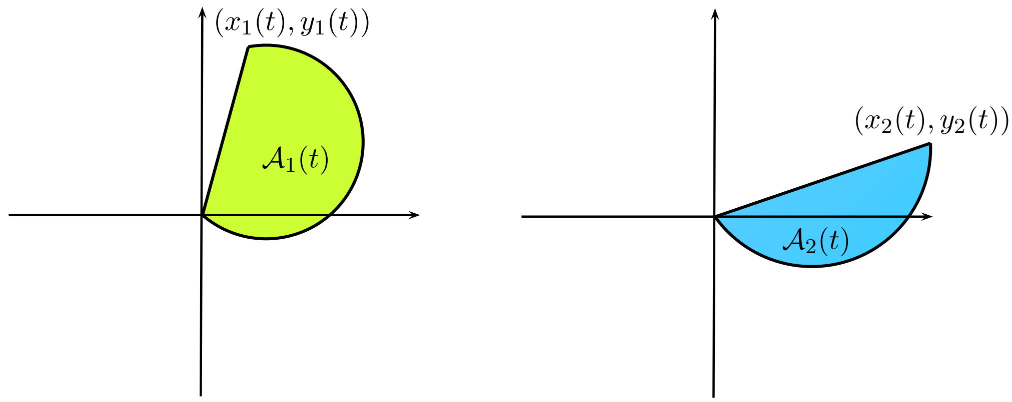
$$u_i^0 = -r_i \sin\theta_i, \qquad v_i^0 = r_i \cos\theta_i, \qquad i = 1, \ldots, l. \tag{13.21}$$

In these new coordinates (13.20) becomes (case $w \neq 0$)

$$\begin{aligned} x_i(t) &= \frac{r_i}{\alpha_i w}(\cos(\alpha_i w t + \theta_i) - \cos(\theta_i)), \\ y_i(t) &= \frac{r_i}{\alpha_i w}(\sin(\alpha_i w t + \theta_i) - \sin(\theta_i)), \\ z(t) &= \frac{1}{2} \sum_{i=1}^{l} \frac{r_i^2}{\alpha_i w^2} (\alpha_i w t - \sin(\alpha_i w t)), \end{aligned} \tag{13.22}$$

and the condition $(u^0, v^0) \in S^{2l-1}$ implies that $r = (r_1, \ldots, r_l) \in S^l$. This permits also to rewrite the $z$ component as follows

$$z(t) = \frac{1}{2w^2} \left( wt - \sum_{i=1}^{l} \frac{r_i^2}{\alpha_i} \sin(\alpha_i w t) \right). \tag{13.23}$$




$$z(t) = \alpha_1 \mathcal{A}_1(t) + \alpha_2 \mathcal{A}_2(t)$$

Figure 13.1: Projection of a non-horizontal geodesic: case $l = 2$ and $0 < \alpha_2 < \alpha_1$.

*Remark* 13.7. From equations (13.22) we easily see that the projection of a geodesic on every 2-plane $(x_i, y_i)$ is a circle, with radius $\rho_i$, center $c_i$, and period $T_i$, given by

$$\rho_i = \frac{r_i}{\alpha_i |w|} \qquad c_i = -\frac{r_i}{\alpha_i w}(\cos\theta_i, \sin\theta_i), \qquad T_i = \frac{2\pi}{\alpha_i |w|}, \qquad \forall\, i = 1, \dots, l \tag{13.24}$$

Moreover, generalizing the analogous property of the 3D Heisenberg group, from (13.16) one can see that the $z$ component of the geodesic at time $t$ is the weighted sum (with coefficients $\alpha_i$) of the areas $\mathcal{A}_i(t)$ of the circles spanned by the vectors $(x_i(t), y_i(t))$ in $\mathbb{R}^2$ (see Figure 13.1). More precisely we have the identities

$$z(t) = \sum_{i=1}^{l} \alpha_i \mathcal{A}_i(t), \qquad \mathcal{A}_i(t) := \frac{r_i^2}{2\alpha_i^2 w^2}(\alpha_i w t - \sin(\alpha_i w t)). \tag{13.25}$$

*Remark* 13.8. Prove the following simmetry identity for the exponential map on multi-dimensional Heisenberg groups: $\exp_0(t, r, \theta, -w) = \exp(-t, r, \theta + \pi, w)$.

### 13.2.2 Optimal synthesis

In this section we assume $\alpha_i > 0$ for every $i = 1, \dots, l$. In particular the structure is contact and there are no non trivial abnormal extremal (see Remark 13.5). It is then sufficient to compute the optimal synthesis to consider normal Pontryagin extremals.

We start the analysis of the optimal synthesis with the following general lemma.

**Lemma 13.9.** *Let $\gamma(t) = \exp_0(r, \theta, w)$ be an arclength parametrized normal trajectory starting from the origin. The cut time $t_*(\gamma)$ along $\gamma$ is equal to the first conjugate time and satisfies*

$$t_*(\gamma) = \frac{2\pi}{|w| \max_i \alpha_i}, \tag{13.26}$$

*with the understanding that $t_*(\gamma) = +\infty$, if $w = 0$.*

*Proof.* The case $w = 0$ is trivial. Indeed the geodesic is a straight line and, by Remark 13.6, the trajectory is optimal for all times hence $t_*(\gamma) = +\infty$. We can assume then $w \neq 0$. Moreover, thanks to Remark 13.8, and up to relabeling coordinates, it is not restrictive to assume that $w > 0$ and $\alpha_1 \geq \alpha_2 \geq \ldots \geq \alpha_l > 0$.

Since all $\alpha_i > 0$ are strictly positive, there are no abnormal minimizers. First we prove that at the point $\gamma(t_*)$ there is at least a one parametric family of trajectory reaching this point and with the same length. Thanks to Theorem 8.72, this will impy that the cut time is less or equal than $t_*(\gamma)$ given in (13.26). Then we prove that for every $t < t_c$ the restriction $\gamma|_{[0,t]}$ a is length-minimizer, proving that the formula given in (13.26) is the cut time.

(i). By our assumption, $\alpha_1 = \max_i \alpha_i$. From (13.22) it is easily seen that the projection on the $(x_1, y_1)$-plane of the trajectory $\gamma$ satisfies

$$x_1(t_*) = y_1(t_*) = 0.$$

Define the variation $\theta_\phi := (\theta_1 + \phi, \theta_2, \ldots, \theta_l)$ for $\phi \in [0, 2\pi]$, and consider the trajectories

$$\gamma_\phi(t) = \exp_0(t, r, \theta_\phi, w), \qquad \phi \in [0, 2\pi].$$

It is easily seen from equation (13.22) that all these curves have the same endpoints. Indeed neither $(x_i, y_i)$, for $i > 1$, nor $z$ depends on this variable. Then it follows that $t_*$ is a conjugate time.

(ii). Since $w > 0$, our geodesic is not contained in the hyperplane $\{z = 0\}$. Moreover, for every $i = 1, \ldots, l$, the projection of every non horizontal geodesic on on the plane $(x_i, y_i)$ is a circle. In particular, the distance from the origin of the projected curve is easily computed by

$$\eta_i(t) := \sqrt{x_i(t)^2 + y_i(t)^2} = \sin_c\left(\frac{\alpha_i w t}{2}\right) r_i t, \qquad \text{where} \qquad \sin_c(x) := \frac{\sin x}{x}.$$

Let now $t_0 < t_*$. We want to show that there is no length-parametrized geodesic starting from the origin $\widetilde{\gamma} \neq \gamma$ reaching the point $\gamma(t_0)$ in time $t_0$.

Assume by contradiction that there exists $\widetilde{\gamma}(t) = \exp_0(t, \widetilde{r}, \widetilde{\theta}, \widetilde{w})$ with $\widetilde{r} \in S^l$ such that $\gamma(t_0) = \widetilde{\gamma}(t_0)$. Then for every $i = 1, \ldots, l$ we have $\eta_i(t_0) = \widetilde{\eta}_i(t_0)$ which means

$$\sin_c\left(\frac{\alpha_i w t_0}{2}\right) r_i t_0 = \sin_c\left(\frac{\alpha_i \widetilde{w} t_0}{2}\right) \widetilde{r}_i t_0 \qquad i = 1, \ldots, l. \tag{13.27}$$

Notice that, once $\widetilde{w}$ is fixed, $\widetilde{r}_i$ are uniquely determined by (13.27) (here $t_0$ is fixed). Moreover, $\widetilde{\theta}_i$ also are uniquely determined (mod $2\pi$) by relations (13.24). Finally, from the assumption that $\widetilde{\gamma}$ reaches optimally the point $\widetilde{\gamma}(t_0)$ as well, it follows that

$$t_0 < t_*(\widetilde{\gamma}) = \frac{2\pi}{\alpha_1 \widetilde{w}} \qquad \Longrightarrow \qquad \frac{\alpha_i \widetilde{w} t_0}{2} < \pi \quad \forall\, i = 1, \ldots, l. \tag{13.28}$$

Assume $\widetilde{w} > w$. Since $\sin_c(x)$ is a strictly decreasing function on $[0, \pi]$, this implies $\widetilde{r}_i > r_i$ for every $i = 1, \ldots, l$. In particular

$$\sum_{i=1}^{l} \widetilde{r}_i^2 > \sum_{i=1}^{l} r_i^2 = 1$$

contradicting the fact that $\widetilde{r} \in S^l$. Then, since all $\alpha_i$ are positive there are no abnormal extremals, Theorem 8.72 and Corollary 8.74 permit to conclude that $\gamma(t_0)$ is not a cut point. The case $\widetilde{w} < w$ is analogous. $\square$

In the next proposition we compute the sub-Riemannian distance from the origin to a point contained in the vertical axis, which is always contained in the cut locus.

**Proposition 13.10.** *Let $(0, z) \in \mathbb{R}^{2l} \times \mathbb{R} \simeq \mathbb{R}^{2l+1}$, and let $\alpha_1, \alpha_2, \cdots, \alpha_l$ be the (possibly repeated) frequences of the Heisenberg sub-Riemannian structure. Then $(0, z) \in \mathrm{Cut}_0$ and*

$$d((0,0),(0,z))^2 = \frac{4\pi|z|}{\max_i \alpha_i}. \tag{13.29}$$

*Proof.* Without loss of generality we can assume $\alpha_1 \geq \alpha_2 \geq \cdots \geq \alpha_r > 0$. Consider the trajectory $\gamma(t) = \exp_0(r, \theta, w)$ with $r = (r_1, r_2) = (1, 0, \ldots, 0) \in S^l$ and $\theta = (\theta_1, \ldots, \theta_l)$, $w > 0$ arbitrary. Then by Lemma 13.9 the curve $\gamma|_{[0,t_*]}$ is a length-minimizer for $t_*$ given by (13.26). It follows that

$$d(\gamma(0), \gamma(t_*)) = t_*. \tag{13.30}$$

Thanks to (13.22) it follows easily that

$$x_1(t_*) = y_1(t_*) = x_2(t_*) = y_2(t_*) = 0, \qquad z(t_*) = \frac{\pi}{\alpha_1 w^2} = \frac{\alpha_1}{4\pi} t_*^2. \tag{13.31}$$

Plugging the last formula in (13.30) and writing $t_*$ as a function of $z$ one gets (13.29). □

The exact computation of the cut locus is possible thanks to the characterization of the cut time for every geodesic.

**Exercise 13.11.** Prove the folllowing facts

(a) Assume that $\alpha_1 = \ldots = \alpha_l$. Then $\mathrm{Cut}_0 = \{(0, z) \in \mathbb{R}^{2l+1} : z \in \mathbb{R} \setminus \{0\}\}$.

(b) Assume that $l = 2$ and $0 < \alpha_2 < \alpha_1$. Prove that

$$\mathrm{Cut}_0 = \{(0, 0, x_2, y_2, z) \in \mathbb{R}^5 : |z| \geq (x_2^2 + y_2^2)K(\alpha_1, \alpha_2),\ (x_2, y_2, z) \in \mathbb{R}^3 \setminus \{0\}\}, \tag{13.32}$$

where $K(\alpha_1, \alpha_2)$ is a positive constant satisfying $K(\alpha_1, \alpha_2) \to 0$ for $\alpha_2 \to 0$ and $K(\alpha_1, \alpha_2) \to +\infty$ for $\alpha_2 \to \alpha_1$.

(c) Assume that $l = 2$ and $0 = \alpha_2 < \alpha_1$. Compute $\mathrm{Cut}_0$.

Generalize the previous formulas to all other cases for $0 = \alpha_l \leq \ldots \leq \alpha_l$, and compute the dimension of $\mathrm{Cut}_0$ in terms of the frequences $\alpha_1, \alpha_2, \cdots, \alpha_l$.

## 13.3 Free Carnot groups of step 2

Recall from Definition 13.4 that a Carnot group of step 2 is *free* if the matrices $C_1, \ldots, C_{n-m}$ define a basis of the space of skew-symmetric matrices. In particular $n = m + \frac{m(m-1)}{2}$ and it is convenient to treat $\mathbb{R}^n = \mathbb{R}^m \oplus \mathbb{R}^{n-m}$ as the sum

$$\mathbb{R}^n = \mathbb{R}^m \oplus (\mathbb{R}^m \wedge \mathbb{R}^m).$$

In what follows we denote by $\mathbb{G}_m := \mathbb{R}^m \oplus \wedge^2\mathbb{R}^m$ the free Carnot groups of step 2 and we identify $\wedge^2\mathbb{R}^m$ with the vector space of skew-symmetric real matrices, that is $v \wedge w = vw^* - wv^*$ for $v, w \in \mathbb{R}^m$.

It is convenient to employ the following notation: we denote points $(x, Z) \in \mathbb{G}_m$, where $x \in \mathbb{R}^m$ and $Z$ is a skew-symmetric matrix. We fix the canonical basis $\{E_{\ell j}\}_{1\le \ell<j\le m}$ of $\mathfrak{so}(\mathbb{R}^m)$ and we write $Z = \sum_{\ell<j} Z_{\ell j} E_{\ell j}$.

As discussed in Section 13.1 we can choose a suitable basis in such a way that the sub-Riemannian structure is generated by the set of global orthonormal vector fields:

$$X_i := \partial_{x_i} - \frac{1}{2} \sum_{1\le \ell<j\le m} (e_i \wedge x)_{\ell j} \partial_{Z_{\ell j}}, \qquad i = 1, \dots, m, \tag{13.33}$$

where $\{e_1, \dots, e_m\}$ is the standard basis of $\mathbb{R}^m$. More precisely, the horizontal distribution is defined by $\mathcal{D} := \mathrm{span}\{X_1, \dots, X_m\}$ and the sub-Riemannian metric by $g(X_i, X_j) = \delta_{ij}$.

For all $i < j$, we have $[X_i, X_j] = \partial_{Z_{ij}}$. In particular, the vector fields (13.33) generate the free, nilpotent Lie algebra of step 2 with $m$ generators:

$$\mathfrak{g} = \mathfrak{g}_1 \oplus \mathfrak{g}_2, \qquad \text{where} \qquad \mathfrak{g}_1 = \mathrm{span}\{X_1, \dots, X_m\}, \quad \mathfrak{g}_2 = \mathrm{span}\{\partial_{Z_{ij}}\}_{i<j}. \tag{13.34}$$

The Lie group structure on $\mathbb{G}_m$ such that the vector fields $X_i$ are left-invariant is given by the polynomial product law

$$(x, Z) \star (x', Z') = \left( x + x', Z + Z' + \frac{1}{2} x \wedge x' \right). \tag{13.35}$$

Notice, moreover, that the matrices $C_1, \dots, C_{n-m}$ coincide in this case with the standard basis of $\mathfrak{so}(m)$. Hence the matrix $\Omega_w$ defined in (13.9) is simply an arbitrary skew-symmetric matrix and the $w$ component of the initial covector are coordinates on the space $\mathfrak{so}(m)$

$$\Omega_w = \sum_{1\le \ell<j\le m} w_{\ell j} C_{\ell j} = \sum_{1\le \ell<j\le m} w_{\ell j} E_{\ell j}.$$

For this reason in what follows we drop the $w$ from the notation and simply write $\Omega$ for $\Omega_w$.

**Example 13.12.** The case $m = 2$ is the well-known Heisenberg group. Indeed, we can identify $(x, Z) \in \mathbb{R}^2 \oplus \wedge^2\mathbb{R}^2$ with $(x, z) \in \mathbb{R}^2 \oplus \mathbb{R}$, so that the generating vector fields (13.33) read

$$X_1 = \partial_{x_1} - \frac{x_2}{2} \partial_z, \qquad X_2 = \partial_{x_2} + \frac{x_1}{2} \partial_z. \tag{13.36}$$

**Example 13.13.** The case $m = 3$ can be dealt with by identifying $(x, Z) \in \mathbb{R}^3 \oplus \wedge^2\mathbb{R}^3$ with $(x, t) \in \mathbb{R}^3 \oplus \mathbb{R}^3$. More precisely, any $3 \times 3$ skew-symmetric matrix can be written as $Z = v \wedge w$, and is identified with the cross product $z = v \times w$. Notice that $v \times w$ does not depend on the choice of the representatives $v, w$ such that $Z = v \wedge w$.

Under this identification, the tautological action of $Z$ on $\mathbb{R}^3$ reads

$$Zx = (v \wedge w)x = x \times (v \times w) = x \times z, \qquad \forall x \in \mathbb{R}^3, \tag{13.37}$$

and the generating vector fields (13.33) are

$$X_1 = \partial_{x_1} + \frac{x_3}{2}\partial_{z_2} - \frac{x_2}{2}\partial_{z_3}, \quad X_2 = \partial_{x_2} + \frac{x_1}{2}\partial_{z_3} - \frac{x_3}{2}\partial_{z_1}, \quad X_3 = \partial_{x_3} + \frac{x_2}{2}\partial_{z_1} - \frac{x_1}{2}\partial_{z_2}. \tag{13.38}$$

The goal of this section is to compute the intersection of the cut locus from the origin with the vertical space $V = \{(0, Z) \mid Z \in \wedge^2\mathbb{R}^m\}$. In particular we give the explicit formula of the distance from the origin to every point of $V$.

*Remark* 13.14. Since the sub-Riemannian structure has step 2, then there exists no strictly abnormal length-minimizer thanks to Corollary 12.15. In particular to compute the optimal synthesis and the sub-Riemannian distance it is sufficient to consider only normal Pontryagin extremals.

Suppose now that $\lambda(t) = (x(t), z(t), h(t), w(t)) \in T^*G$ is a normal Pontryagin extremal. Then, thanks to the previous analysis we have

$$h(t) = e^{-t\Omega}h(0), \qquad \Omega \in \mathfrak{so}(m).$$

From this expression one finds the $x$-component

$$x(t) = \int_0^t e^{-s\Omega}h(0)ds.$$

The vertical part of the horizontal trajectory can be recovered integrating the equation

$$\dot{Z}(t) = \frac{1}{2}x(t) \wedge h(t). \tag{13.39}$$

Then we the obtain the following formula (recall $Z(0) = 0$)

$$\begin{aligned} Z(t) &= \frac{1}{2}\int_0^1 \int_0^t e^{-s\Omega}h(0) \wedge e^{-t\Omega}h(0)ds\,dt, & (13.40)\\ &= \frac{1}{2}\int_0^1 \int_0^t (e^{-s\Omega}Pe^{t\Omega} - e^{-t\Omega}Pe^{-s\Omega})ds\,dt, & (13.41)\end{aligned}$$

where we denoted by $P$ the symmetric matrix $h(0)h(0)^*$.

For a fixed geodesic, there exists a good set of coordinates such that the matrix $\Omega$ is written in normal form. The main linear algebra ingredient is given by the following lemma.

**Lemma 13.15.** *Let $\Omega \in \mathfrak{so}(n)$, $x_0 \in \mathbb{R}^n$ and define the set*

$$\Theta := \{\Omega' \in \mathfrak{so}(n) \mid e^{t\Omega'}x_0 = e^{t\Omega}x_0, \text{ for all } t \geq 0\}.$$

*There exists $\overline{\Omega} \in \Theta$ with all nonzero eigenvalues that are simple and such that $\ker\overline{\Omega}$ has maximal dimension.*

*Proof.* Since $\Omega$ is skew-symmetric, there exist $\alpha_1, \ldots, \alpha_r$ such that $\mathrm{spec}(\Omega) = \{\pm i\alpha_1, \ldots, \pm i\alpha_r, 0\}$. Let us decompose $\mathbb{R}^n$ in real eigenspaces

$$\mathbb{R}^n = E_0 \oplus \bigoplus_{j=1}^r E_j, \qquad E_0 = \ker\Omega, \quad E_j = \ker(\Omega + i\alpha_j) \oplus \ker(\Omega - i\alpha_j),$$

and work in an adapted basis inducing coordinates adapted to the splitting. In this basis $\Omega$ has a block-diagonal form $\Omega = \mathrm{diag}\{\Omega_1, \ldots, \Omega_r, 0\}$ and we similarly decompose $x_0 = (x_{0,1}, \ldots, x_{0,r}, x_{0,0})$. Notice that, thanks to the block structure, we have $e^{t\Omega}x_0 = (e^{t\Omega_1}x_{0,1}, \ldots, e^{t\Omega_r}x_{0,r}, 0)$. The existence of the matrix $\overline{\Omega}$ is obtained by applying the following algorithm.

Assume that $x_{0,j} = 0$ for some $j > 0$. Then one can replace the corresponding block $\Omega_j$ with a zero block, without changing the value of $e^{t\Omega}x_0$.

Assume that $\dim E_j > 2$ for some $j > 0$, i.e., there exists a block with multiple eigenvalues. Then, thanks to Exercice 13.16, we have $\dim \operatorname{span}\{e^{t\Omega_j}x_{0,j} \mid t \in \mathbb{R}\} = \dim \operatorname{span}\{x_0, \Omega x_0\} = 2$, thus we can write

$$E_j = \operatorname{span}\{x_{0,j}, \Omega_j x_{0,j}\} \oplus \operatorname{span}\{x_{0,j}, \Omega_j x_{0,j}\}^{\perp}. \tag{13.42}$$

Choosing a basis in $E_j$ corresponding to the splitting (13.42), we can replace the block of $\Omega_j$ corresponding to $\operatorname{span}\{x_{0,j}, \Omega_j x_{0,j}\}^{\perp}$ with a zero block. The restriction of the matrix that one obtains to $E_j$ has $\pm i\alpha_j$ as simple eigenvalues, and a kernel of dimension $\dim(E_j) - 2$. □

**Exercise 13.16.** Let $\Omega \in \mathfrak{so}(n)$ and assume $\operatorname{spec}(\Omega) = \{\pm i\alpha\}$. Then for $x_0 \in \mathbb{R}^n$

$$\operatorname{span}\{e^{t\Omega}x_0 \mid t \in \mathbb{R}\} = \operatorname{span}\{x_0, \Omega x_0\}.$$

From the previous discussion it follows that, for a given geodesic, there exists a linear change of coordinates in the space such that the matrix $\Omega$ is presented as a block-diagonal matrix

$$\Omega = (\Omega_1, \dots, \Omega_\ell, \mathbb{O}),$$

where $\mathbb{O}$ is a block zero matrix and

$$\Omega_i = \begin{pmatrix} 0 & \alpha_i \\ -\alpha_i & 0 \end{pmatrix} = \alpha\mathbb{J},$$

where $\mathbb{J}$ denotes the $2 \times 2$ symplectic matrix $\mathbb{J} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$.

### 13.3.1 Intersection of the cut locus with the vertical subspace

First we prove that every vertical points in $\mathbb{G}_m$ is contained in the cut locus. More precisely we have the following.

**Lemma 13.17.** *The set of points $\{(0, Z) \in \mathbb{R}^m \oplus \wedge^2\mathbb{R}^m \mid Z \neq 0\}$ is contained in* $\operatorname{Cut}_0(\mathbb{G}_m)$.

*Proof.* Fix a point $(0, Z) \in \mathbb{G}_m$ with $Z \neq 0$. Thanks to Exercice13.18 there exists a non zero orthogonal matrix $M \in SO(m)$ such that $MZM^* = Z$ and $M$ equal to the identity on $\ker Z$. Let now $\gamma(t) = (x(t), Z(t))$ be a length-minimizer joining the origin to $(0, Z)$. The existence of such a geodesic is guaranteed by completeness of the sub-Riemannian structure. Let us show that there exist (at least) two length-minimizers reaching $(0, Z)$.

Consider the curve $\overline{\gamma}(t) = (Mx(t), MZ(t)M^*)$. Notice that $\overline{\gamma}(0) = (0, 0)$ and, by properties of $M$, one has $\overline{\gamma}(1) = (0, MZM^*) = (0, Z)$. Moreover $\ell(\gamma) = \ell(\overline{\gamma})$. Since $M \neq \mathbb{I}$ we have $\gamma \neq \overline{\gamma}$. Thus $\gamma$ and $\overline{\gamma}$ are two horizontal length-minimizers joining the same end-points. This proves the claim. □

**Exercise 13.18.** Let $Z \in \mathfrak{so}(m)$ be a non zero skew-symmetric matrix.

(a). Prove that there exists an orthogonal matrix $M \in SO(m), M \neq \mathbb{I}$, such that $MZM^* = Z$.

(b). Prove that the matrix $M$ can be chosen to be the identity on $\ker Z$.

(c). Show that the set of matrices satisfying properties (a) and (b) is a Lie group and compute its dimension.

Now we compute the sub-Riemannian distance from the origin of vertical points in $\mathbb{G}_m$.

**Proposition 13.19.** *Let $(0, Z) \in \mathbb{G}_m$, and let $\alpha_1 \geq \alpha_2 \geq \cdots \geq \alpha_r > 0$ be the (possibly repeated) absolute values of the non-zero eigenvalues of $Z$. Then,*

$$d((0,0),(0,Z))^2 = 4\pi \sum_{j=1}^{r} j\alpha_j. \tag{13.43}$$

*Proof.* Without loss of generality, Let $\gamma(t) = (x(t), Z(t))$ be a geodesic from the origin such that $x(1) = 0$ and $Z(1) = Z$, with $h(t) = e^{-\Omega t}h_0$, where we set $h_0 := h(0)$. By (13.40), we have

$$\int_0^1 e^{-t\Omega} h_0 \, dt = x(1) = 0. \tag{13.44}$$

Thus, the non-zero eigenvalues of $\Omega$ are of the form $\pm i2\pi\phi$, with $\phi \in \mathbb{N}$. By Lemma 13.15, and up to an orthogonal transformation, we may assume that $\Omega = (2\pi\phi_1\mathbb{J}, \ldots, 2\pi\phi_\ell\mathbb{J}, 0_{m-2\ell})$, with all simple eigenvalues, $2\ell = \mathrm{rank}\,(\Omega)$, and with distinct $\phi_i \in \mathbb{N}$. We split accordingly $h_0 = (h_{0,1}, \ldots, h_{0,\ell}, h_{0,0})$, with $h_{0,i} \in \mathbb{R}^2$ for $i = 1, \ldots, \ell$ and $h_{0,0} \in \mathbb{R}^{m-2\ell}$. Using the canonical form and the fact that $\phi \in \mathbb{N}$, it is not difficult to explicitly integrate the vertical part of the geodesic equations (13.40). We obtain

$$Z(1) = \left( \frac{|h_{0,1}|^2}{4\pi\phi_1}\mathbb{J}, \ldots, \frac{|h_{0,\ell}|^2}{4\pi\phi_\ell}\mathbb{J}, 0_{m-2\ell} \right). \tag{13.45}$$

Then $|h_{0,j}|^2 = 4\pi\phi_j\alpha_j$ for all $j = 1, \ldots, r$. The squared length of $\gamma$ is

$$\ell(\gamma)^2 = \left( \int_0^1 |u(t)| dt \right)^2 = |h_0|^2 = \sum_{j=1}^{r} |h_{0,j}|^2 = 4\pi \sum_{j=1}^{r} \phi_j \alpha_j. \tag{13.46}$$

The minimum of this quantity over all choice of $\phi_j \in \mathbb{N}$ and all distinct is obtained when $\phi_j = j$, for all $j = 1, \ldots, r$. □

### 13.3.2 The cut locus for the free step-two Carnot group of rank three

Lemma 13.17 shows that the cut locus $\mathrm{Cut}_0(\mathbb{G}_m)$ always contains the vertical points, i.e., those of the form $\{(0, Z) \in \mathbb{R}^m \oplus \wedge^2\mathbb{R}^m,\ Z \neq 0\}$.

In the case of the Heisenberg group, that is $\mathbb{G}_2$ in the notation of this chapter, the cut locus has been previously computed, and it indeed coincides with the vertical space

$$\mathrm{Cut}_0(\mathbb{G}_2) = \{(0, Z) \in \mathbb{R}^2 \oplus \wedge^2\mathbb{R}^2,\ Z \neq 0\}.$$

Nevertheless, in general, the cut locus is bigger than the vertical set. In this section, we give the explicit expression of the cut locus $\mathrm{Cut}_0(\mathbb{G}_m)$ when $m = 3$, also called free Carnot group of growth vector $(3, 6)$.

The detailed computations of the cut time and cut locus for the case $m = 3$ are much harder, and we do not give them here. We only state the following characterization.

**Theorem 13.20.** *The cut locus for the step 2 free Carnot groups of rank 3 is given by*

$$\mathrm{Cut}_0(\mathbb{G}_3) = \{(x, Z) \in \mathbb{R}^3 \oplus \wedge^2\mathbb{R}^3 \mid Z = v \wedge w \neq 0,\ Zx = 0\}. \tag{13.47}$$

A proof of this fact was first given in [Mya02] (cf. also [MM17]). The following (strict) inclusion holds for general $m > 3$

$$\{(x, Z) \in \mathbb{R}^m \oplus \wedge^2\mathbb{R}^m \mid Z = v \wedge w \neq 0,\ Zx = 0\} \subset \mathrm{Cut}_0(\mathbb{G}_m), \tag{13.48}$$

Indeed in [RS17] the authors prove that there exists a set of codimension 2 that is contained in the cut locus. It is still an open problem to characterize exactly the set $\mathrm{Cut}_0(\mathbb{G}_m)$ and we refer to [RS17] for more details.

## 13.4 An extended Hadamard technique to compute the cut locus

Let us consider a sub-Riemannian structure, complete as a metric space and fix $q_0 \in M$. Assume that we are able to solve the problems (A) and (B) stated after Definition 13.3. This usually is not so hard when one is considering left-invariant structures on Lie groups of small dimension. More precisely assume that:

- we are able to get the explicit expression of normal geodesics;
- we are able to prove that all strict abnormal extremals are not optimal.

Let $\exp_{q_0}(t, \theta)$ be the standard exponential map providing geodesic parametrized by arclength (here $\theta \in \Lambda_{q_0} = T^*_{q_0}M \cap H^{-1}(1/2)$). With a slight abuse of notation, let $\exp_{q_0}(\lambda)$ be the exponential map at time 1 (here $\lambda \in T^*_{q_0}M$). Notice that $\exp_{q_0}(t, \theta) = \exp_{q_0}(\lambda)$ with $\lambda = t\,\theta$.

A useful method to evaluate the cut time for every normal extremal consists in a suitable use of a classical result stating that if a smooth map between two connected manifolds of the same dimension is proper and its differential is nowhere singular then it is a covering.

**Definition 13.21.** A continuous map $f : M_1 \to M_2$ between smooth manifold is proper if $f^{-1}(K)$ is compact in $M_1$ for any $K$ compact in $M_2$.

To prove that a continuous map is proper it is sufficient to show that a sequence escaping out from any compact in $M_1$ escapes out from any compact in $M_2$. When $M_1$ and $M_2$ are subsets of two compact manifolds with the induced topologies, then to prove that $f$ is proper, it is sufficient to prove that $\partial M_1$ is mapped in $\partial M_2$ through $f$.

**Definition 13.22.** A continous (resp. smooth) map $f : M_1 \to M_2$ between connected smooth manifolds is a *continuous (resp. smooth) covering map* if for every $y \in M_2$, there exists an open neighborhood $V$ of $y$, such that $f^{-1}(V)$ is a union of disjoint open sets in $M_1$, each of which is mapped homeomorphically (resp. diffeomorphically) onto $V$.

We recall some important properties of covering maps:

P1: The preimage of a point is a discrete set, whose cardinality is independent from the point.

P2: Given a continuous curve $\gamma : [0, 1] \to M_2$ and a point $q_1$ in $M_1$ such that $f(q_1) = \gamma(0)$, then there exists a unique continuous curve $\Gamma_{q_1} : [0, 1] \to M_1$ such that $\Gamma_{q_1}(0) = q_1$ and $f(\Gamma_{q_1}) = \gamma$ (see Figure 13.2). The curve $\Gamma_{q_1}$ is called the lift of $\gamma$ (through $q_1$).

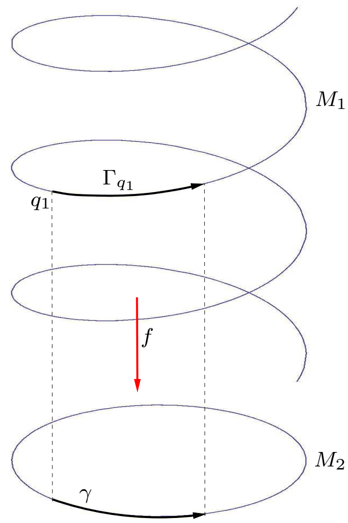



Figure 13.2: Uniqueness of the lift for a covering map.

P3: Consider two homotopic loops $\gamma, \gamma' : [0,1] \to M_2$ and a point $q_1$ in $M_1$ such that $f(q_1) = \gamma(0) = \gamma'(0)$. Let $\Gamma_{q_1}$ and $\Gamma'_{q_1}$ the corresponding lifts. Then the final points of $\Gamma_{q_1}$ and $\Gamma'_{q_1}$ are the same, namely $\Gamma_{q_1}(1) = \Gamma'_{q_1}(1)$.

**Theorem 13.23.** *Let $M_1$ and $M_2$ two smooth connected differentiable manifolds and $f : M_1 \to M_2$ be smooth. Assume that $f$ is proper and that its differential is nowhere singular. Then $f$ is a covering.*

*Proof.* We recall that any proper continuous map $f : M_1 \to M_2$ between smooth manifold is closed, i.e., $f(C)$ is closed in $M_2$ for every closed set $C \subset M_1$.

Since $f$ is a local diffeomorphism, it is open. Since $f$ is proper, it is closed. Hence $f(M_1)$ is open and closed in $M_2$ and, by connectedness, $f$ is surjective. Fix $y \in M_2$. Since $f$ is a local diffeomorphism, each point of $f^{-1}(y)$ has a neighborhood on which $f$ is injective, so $f^{-1}(y)$ is a discrete set. Since the singleton $\{y\}$ is compact and $f$ is proper, then $f^{-1}(y)$ is compact, hence finite. Set $f^{-1}(y) = \{x_1, \ldots, x_k\}$. Fix $U_i$ a neighborhood of $x_i$ where $f$ is a diffeomorphism. It is not restrictive to suppose that $U_i \cap U_j = \emptyset$ for $i \neq j$. Set $V = \cap_{i=1}^k f(U_i)$. Since each $f(U_i)$ is a neighborhood of $y$, $V$ is a neighborhood of $y$ also. By replacing $V$ with the connected component of $V \setminus f(M_1 \setminus \cup_i U_i)$ (which is open since $f$ is closed) containing $y$, we can moreover assume that $V$ is connected and $f^{-1}(V) \subset \cup_i U_i$. Hence if one set $\overline{U}_i := U_i \cap f^{-1}(V)$ one can check that $f^{-1}(V) = \cup_i \overline{U}_i$, disjoint union of its connected components, and that $f : \overline{U}_i \to V$ is a diffeomorphism, as desired.

□

Often one would like to prove that $f$ is indeed a diffeomorphism (at least this is what we will need later, with the exponential map playing the role of $f$). Once it is known that the map $f$ is a covering map, to show that it is injective one should prove that it is a 1-sheet covering, i.e., that the preimage of each point is a single point. The following corollary provides a criterium.

**Corollary 13.24** (of Theorem 13.23). *Under the assumptions of Theorem 13.23, if $M_2$ is simply connected, then $f : M_1 \to M_2$ is a diffeomorphism.*

*Proof.* It is enough to show that the map $f$ is injective. Let $x_1 \neq x_2$ in $M_1$ such that $f(x_1) = f(x_2)$. Take a continuous curve $\alpha : [0,1] \to M_1$ such that $\alpha(0) = x_1$ and $\alpha(1) = x_1$ homotopic to a point. Its image $\gamma := f \circ \alpha : [0,1] \to M_2$ is a closed loop in $M_2$ such that $\gamma(0) = \gamma(1) = y$. Since $M_2$ is simply connected there exists a continous map

$$\Gamma : [0,1] \times [0,1] \to M_2$$

such that $\Gamma(0,t) = y$ and $\Gamma(1,t) = \gamma(t)$. For $s$ sufficiently closed to 0 the curve $\gamma_s(t) = \Gamma(s,t)$ stays in the set $V$ where $f$ is a covering hence $f^{-1}(\gamma)$ is the union on $k$ closed loop and it should be homotopic to a point. This gives a contradiction. □

Another criterium, when $M_2$ is not simply connected, is given by the following result.

**Corollary 13.25** (of Theorem 13.23). *Under the assumptions of Theorem 13.23, if $M_2$ is homeomorphic to $S^1 \times N$, where $N$ is simply connected, and there exists a loop in $M_1$ whose image via $f$ is a loop in $M_2$ that is homotopic to $S^1 \times \{x\}$, for some $x \in N$, then $f : M_1 \to M_2$ is a global diffeomorphism.*

*Proof.* Assume by contradiction that the number of pre-images of a point is not one. Let $\Gamma : [0,1] \to M_1$ be a loop in $M_1$, where $q_1 = \Gamma(0)$, and let $\gamma = f(\Gamma)$ its image in $M_2$, that is homotopic to $S^1 \times \{x\}$, for some $x \in N$. Let $\overline{q}_1 \in f^{-1}(\gamma(0))$ with $q_1 \neq \overline{q}_1$. We refer to Figure 13.3.

Consider a continuous curve $\overline{\Gamma} : [0,1] \to M_1$ connecting $q_1$ and $\overline{q}_1$ (this is possible since $M_1$ is connected a manifold and hence path connected). Consider its image on $M_2$ that is $\overline{\gamma} := f(\overline{\Gamma})$. Since $M_2$ is homeomorphic to $S^1 \times N$, it is defined the winding number of a loop in $M_2$. In particular it is defined the winding number $w$ of $\overline{\gamma}$. Notice that $w$ is an integer, defined up to a sign up to a sign since there is no orientation.

If $w = 1$, i.e., $\overline{\gamma}$ is homotopic to $S^1 \times \{y\}$, for some $y \in N$, then $\overline{\gamma}$ is homotopic to $\gamma$ since $N$ is simply connected. Hence since $\Gamma(0) = \overline{\Gamma}(0) = q_1$ and thanks to property P3 we have that $\overline{\Gamma}(1) = \Gamma(1)$. As a consequence $q_1 = \overline{q}_1$.

If $w > 1$, then we consider the loop $\Gamma^n : [0,n] \to M_1$ obtained by concatenating $n$ times the curve $\Gamma$. Let us denote $\gamma^n$ its image on $M_2$. We have that $\overline{\gamma}$ is homotopic to $\gamma^n$. This implies $q_1 = \overline{q}_1$, as above.

If $w = 0$, i.e., if $\overline{\gamma}$ is contractible, we consider a contractible loop $\Gamma^0 : [0,1] \to M_1$ such that $\Gamma^0(0) = \Gamma^0(1) = q_1$. Let $\gamma^0$ be its image. Since a covering is a continuous map, the image of a contractible loop is a contractible loop. Hence $\gamma^0$ is contractible and we have that $\overline{\gamma}$ and $\gamma^0$ are homotopic. The same reasoning as before gives again $q_1 = \overline{q}_1$.

□

*Remark* 13.26. More in general, if $f : M_1 \to M_2$ is a covering map, then the induced map on the fundamental groups $[f] : \pi_1(M_1) \to \pi_1(M_2)$ is injective. Moreover the number of preimages of a point via $f$ coincides with the index of the subgroup $[f](\pi_1(M_1))$ in $\pi_1(M_2)$. If one assumes that $[f]$ is surjective, then it follows that the cardinality of preimages of points is one, hence $f$ is a global diffeomorphism. We refer to [Hat02, Prop. 1.31, 1.32] for a proof of these statements.

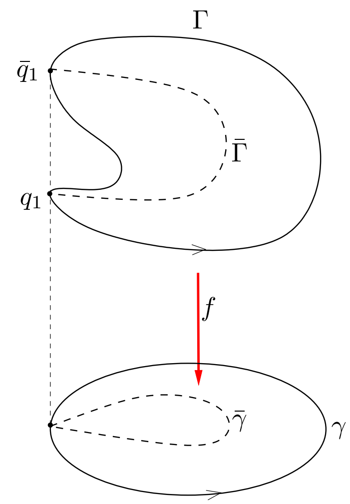



Figure 13.3: Proof of Corollary 13.25

In the special case of Corollary 13.25 one has $\pi_1(M_2) \simeq \mathbb{Z}$ and the existence of a loop satisfying the above assumption is equivalent to require that $[f]$ is surjective.

Finding the cut locus via Theorem 13.23 consists in the following steps. Notice that the method is slightly different if the structure is Riemannian at the starting point (i.e., if the rank of the sub-Riemannian structure at $q_0$ is equal to $\dim M = n$) or not. Recall that if the structure is Riemannian at $q_0$, then $\Lambda_{q_0}$ has the topology of $S^{n-1}$ while if the structure has rank $m < n$ at $q_0$ then $\Lambda_{q_0}$ has the topology of $S^{m-1} \times \mathbb{R}^{n-m}$.

**Step 1** Study the symmetries of the problem to identify points that are reached at the same time by more than one geodesic. This analysis has the purpose of having a guess about the cut locus and hence of the cut time for each geodesic.

Let us call the conjectured cut locus $\mathrm{Cut}^*_{q_0}$ and the conjectured cut times $t^*_{\mathrm{cut}}(\theta)$, for $\theta \in \Lambda_{q_0}$ (notice that it may happen that $t^*_{\mathrm{cut}}(\theta)$ is $+\infty$).

Notice that if $\mathrm{Cut}^*_{q_0}$ has a boundary then the points on the boundary are expected to be conjugate points (since the set $\mathrm{Cut}^*_{q_0}$ comes from the symmetries of the problem it is usually not difficult to verify that the points on his boundary are conjugate points). Conjugate points on the boundary of $\mathrm{Cut}^*_{q_0}$ must be included in $\mathrm{Cut}^*_{q_0}$.

We have two cases:

- If the structure is Riemannian at $q_0$ define $N_1 = \{t\,\theta \mid \theta \in \Lambda_{q_0}, t \in [0, t^*_{\mathrm{cut}}(\theta))\} \subset T^*_{q_0}M$. Notice that in this case $N_1$ is an open star-shaped set always covering a neighborhood of the origin in $T^*_{q_0}M$.
- If the structure is not Riemannian at $q_0$ define $N_1 = \{t\,\theta \mid \theta \in \Lambda_{q_0}, t \in (0, t^*_{\mathrm{cut}}(\theta))\}$; Notice that in this case $N_1$ is an open set that looks like a star-shaped set to which it was removed the starting point and the annihilator of the distribution.

Define $N_2 = \exp_{q_0}(N_1)$. Verify that $N_2 = M \setminus \mathrm{Cut}^*_{q_0}$. If this is not the case then the conjectured cut locus and cut times were wrong. Indeed, if there exists $q \in N_2 \setminus (M \setminus \mathrm{Cut}^*_{q_0})$, then $q$ is reached by a geodesic at its conjectured cut time and by another geodesic before its conjectured cut time. On the other side if there exists $q \in (M \setminus \mathrm{Cut}^*_{q_0}) \setminus N_2$ then $\exp_{q_0}|_{N_1}$ is not covering $M$ up to the conjectured cut locus. These facts are clarified by the following example.

**Example 13.27.** Consider the problem of finding the optimal synthesis starting from 0 for standard Riemannian metric on the circle $S^1 = [-\pi, \pi]/\sim$ where $\sim$ is the identification of $-\pi$ and $\pi$. We have only two geodesics parametrized by arclength: $q^+(t) = t$ and $q^-(t) = -t$. By symmetry the two geodesics meet at $t = 0, \pi, 2\pi, 3\pi, \dots$ etc. Assume that we make the (false) conjecture that the cut time is $t^*_{\mathrm{cut}} = 3\pi$ (instead than $t^*_{\mathrm{cut}} = \pi$). In this case **Step 1** fails because $N_2 = S^1 \neq S^1 \setminus \{\pi\} = S^1 \setminus \mathrm{Cut}^*_0$.

*Remark* 13.28. Notice that if the structure is Riemannian at $q_0$ and the conjectured cut locus is the right one, then $N_2$ is contractible (can be contracted to $q_0$ along the geodesics) and hence it is simply connected.

**Step 2** Prove that the differential of $\exp_{q_0}$ is invertible at every point in $N_1$ (i.e., there are no conjugate points in $N_2$ for $\exp|_{N_1}$). In the following, for simplicity, we assume that there are no non-trivial abnormal extremals. If there are non-strict abnormal extremals (that moreover are non trivial) then there are always conjugate points (cf. Remark 8.46). In this case one can apply the technique explained here to the larger subset of $N_1$ not containing points mapped to the support of the abnormal. In this way one can obtain the optimal synthesis outside the support of the abnormal and one should study the abnormal separately. See the bibliographical note for some references.

**Step 3** Prove that $\exp_{q_0}|_{N_1}$ is proper.

**Step 4 (R)** If the structure is Riemannian at $q_0$ and the conjectured cut locus is the right one, then $N_2$ should be simply connected (cf. Remark 13.28). After having verified that $N_2$ is simply connected, Corollary 13.24 (with $N_1, N_2, \exp_{q_0}$ playing the role of $M_1, M_2, f$) permits to conclude that $\exp_{q_0}|_{N_1}$ is a diffeomorfism and hence that the conjectured cut times and cut locus are the true ones.

**Step 4 (SR)** If the structure is not Riemannian at $q_0$, Theorem 13.23 permits to prove that $\exp_{q_0}|_{N_1}$ is a covering but one cannot conclude that $f$ is a diffeomorphism using Corollary 13.24 unless $N_2$ is simply connected. If $N_2$ is not simply connected, to conclude that $\exp_{q_0}|_{N_1}$ is a diffeomorphism one could for instance try to apply Corollary 13.25. Notice that if $n = 3$ and the structure is not Riemannian at $q_0$ then $N_2$ is never simply connected.

Writing $\gamma_\theta(\cdot) = \exp_{q_0}(\cdot, \theta)|_{[0, t^*_{\mathrm{cut}}(\theta)]}$ the optimal synthesis is then the collection of trajectories

$$\left\{ \gamma_\theta(\cdot) \mid \theta \in H^{-1}(1/2) \right\}.$$

*Remark* 13.29. The main difference between the case in which $q_0$ is a Riemannian point and when it is not, is that in the second case $q_0$ should be removed from $N_1$. This should be done to satisfy the hypothesis of Theorem 13.23 and in particular to guarantee that **i)** $N_1$ is a manifold **ii)** there are no conjugate points in $N_1$ (the starting point is always a conjugate point when the structure is not Riemannian at the starting point itself).

Notice that when $q_0$ is a Riemannian point, the starting point is not a conjugate point. Moreover $N_1$ is a manifold even without removing $q_0$. Thanks to the fact that in this case $N_1$ is star-shaped, it is enough to verify that $N_2$ is simply connected and one obtain directly that if there are no conjugate points in $N_2$ for $\exp|_{N_1}$ then $\exp|_{N_1}$ is a diffeomorphism.

We are now going to apply this technique to a structure that is Riemannian at the starting point and to a structure that is not Riemannian at the starting point.

## 13.5 The Grushin structure

The Grushin plane is the free almost-Riemannain structure on $\mathbb{R}^2$, with coordinates $(x, y)$, for which a global orthonormal frame is given by (cf. Section 9.2)

$$F_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \qquad F_2 = \begin{pmatrix} 0 \\ x \end{pmatrix}.$$

Such a structure is Riemannian out of the $y$ axis that is called the *singular set.* The only abnormal extremals are the trivial ones lying on the singularity. Indeed out of the singularity we are in the Riemannian setting and a curve whose support is entirely contained in the singular set is not admissible. We are then reduced to study normal Pontryagin extremals.

Writing $p = (p_1, p_2)$, the maximized Hamiltonian is given by

$$H(x, y, p_1, p_2) = \frac{1}{2}(\langle p, F_1\rangle^2 + \langle p, F_2\rangle^2) = \frac{1}{2}(p_1^2 + x^2 p_2^2), \tag{13.49}$$

and the corresponding Hamiltonian equations are:

$$\begin{aligned} \dot{x} &= p_1, & \dot{p}_1 &= -x\, p_2^2, \\ \dot{y} &= x^2 p_2, & \dot{p}_2 &= 0. \end{aligned}$$

Normal Pontryagin extremals parameterized by arclength are projections on the $(x, y)$ plane of solutions of these equations, lying on the level set $H = 1/2$.

### 13.5.1 Optimal synthesis starting from a Riemannian point

Let us construct the optimal synthesis starting from a point $(x_0, 0)$, $x_0 \neq 0$ (taking the second coordinate zero is not restrictive due to the invariance of the structure by $y$-translations). In this case the condition $H(x(0), y(0), p_1(0), p_2(0)) = 1/2$ becomes $p_1^2 + x_0^2 p_2^2 = 1$ and it is convenient to set $p_1 = \cos(\theta)$, $p_2 = \sin(\theta)/x_0$, for $\theta \in S^1$. The expression of the normal Pontryagin extremals parameterized by arclenght is $q(t, \theta) = \exp_{(x_0,0)}(t, \theta) = (x(t, \theta), y(t, \theta))$ where

$$\begin{cases} x(t, 0) = t + x_0, \quad y(t, 0) = 0, \\[2ex] y(t, \pi) = -t + x_0, \quad y(t, \pi) = 0, \\[2ex] \left.\begin{aligned} x(t, \theta) &= x_0 \dfrac{\sin(\theta + \frac{t \sin(\theta)}{x_0})}{\sin(\theta)}, \\ y(t, \theta) &= x_0 \dfrac{2t + 2x_0 \cos(\theta) - x_0 \frac{\sin\left(2\theta + 2\frac{t\sin(\theta)}{x_0}\right)}{\sin(\theta)}}{4 \sin(\theta)}, \end{aligned}\right\} \quad \text{if } \theta \notin \{0, \pi\}. \end{cases} \tag{13.50}$$

**Theorem 13.30.** *The cut time for the geodesic $q(\cdot,\theta)$ is*

$$t_{\mathrm{cut}}(\theta) = \left| x_0 \frac{\pi}{\sin(\theta)} \right|.$$

*For $\theta = 0$ or $\theta = \pi$ this formula should be interpreted in the sense that the corresponding geodesic $q(\cdot,0)$ and $q(\cdot,\pi)$ are optimal in $[0,\infty)$.*

*Let us fix $\theta \in (0,\pi)$ (being the case $\theta \in (\pi, 2\pi)$ symmetric). For $\theta \notin \pi/2$, the cut point $q(t_{\mathrm{cut}}(\theta),\theta)$ is reached exactly by two optimal geodesics. Namely the geodesics: $q(\cdot,\theta)$ and the geodesics $q(\cdot,\pi-\theta)$.*

*For $\theta = \pi/2$ the cut point $q(t_{\mathrm{cut}}(\theta),\theta)$ is reached exactly by one optimal geodesic for which $t_{\mathrm{cut}}(\theta)$ is also a conjugate point.*

By direct computation one gets

**Corollary 13.31.** *The cut locus starting from $(x_0, 0)$ is*

$$\mathrm{Cut}_{x_0} = \left\{ (-x_0, y) \in \mathbb{R}^2 \mid y \in (-\infty, -\frac{\pi}{2}x_0^2] \cup [\frac{\pi}{2}x_0^2, \infty) \right\}.$$

*the points $(-x_0, \pm\frac{\pi}{2}x_0^2)$ are also conjugate points.*

The optimal synthesis for Grushin plane with $x_0 = -1$ is depicted in Figure 13.4.

### Proof of Theorem 13.30

We are going to apply the extended Hadamard technique. Recall that in this case the starting point is Riemannian.

**Step 1: Construction of the conjectured cut locus and of the sets $N_1$ and $N_2$.**

By a direct computation one immediately obtains:

**Lemma 13.32.** *For $\theta \neq \{0,\pi\}$, we have*

$$q\left(\left|x_0 \frac{\pi}{\sin(\theta)}\right|, \theta\right) = q\left(\left|x_0 \frac{\pi}{\sin(\theta)}\right|, \pi-\theta\right) = \left(-x_0, \frac{\pi}{2} x_0^2 \frac{1}{\sin(\theta)^2}\right).$$

*Moreover the determinant of the differential of the exponential map is:*

$$D(t,\theta) := \begin{pmatrix} \partial_t x(t,\theta) & \partial_\theta x(t,\theta) \\ \partial_t y(t,\theta) & \partial_\theta y(t,\theta) \end{pmatrix} = \begin{cases} t^2 + \frac{t^3}{3x_0} + t x_0 & \text{if } \theta = 0, \\ -t^2 + \frac{t^3}{3x_0} + t x_0 & \text{if } \theta = \pi, \\ \dfrac{x_0 \left( x_0 \frac{\sin\left(\frac{t\sin(\theta)}{x_0}\right)}{\sin(\theta)} - t\cos(\theta)\cos\left(\theta + \frac{t\sin(\theta)}{x_0}\right)\right)}{\sin^2(\theta)}, & \text{if } \theta \notin \{0,\pi\}. \end{cases}$$

*In particular $D(|x_0\pi|, \pi/2) = 0$.*

We then conjecture that the cut time of the geodesic $q(t,\theta)$ is $t^*_{\mathrm{cut}}(\theta) = \left|x_0 \frac{\pi}{\sin(\theta)}\right|$ and that the cut locus is

$$\mathrm{Cut}^*_{x_0} = \{(-x_0, y) \in \mathbb{R}^2 \mid y \in (-\infty, -\frac{\pi}{2}x_0^2] \cup [\frac{\pi}{2}x_0^2, \infty)\}.$$

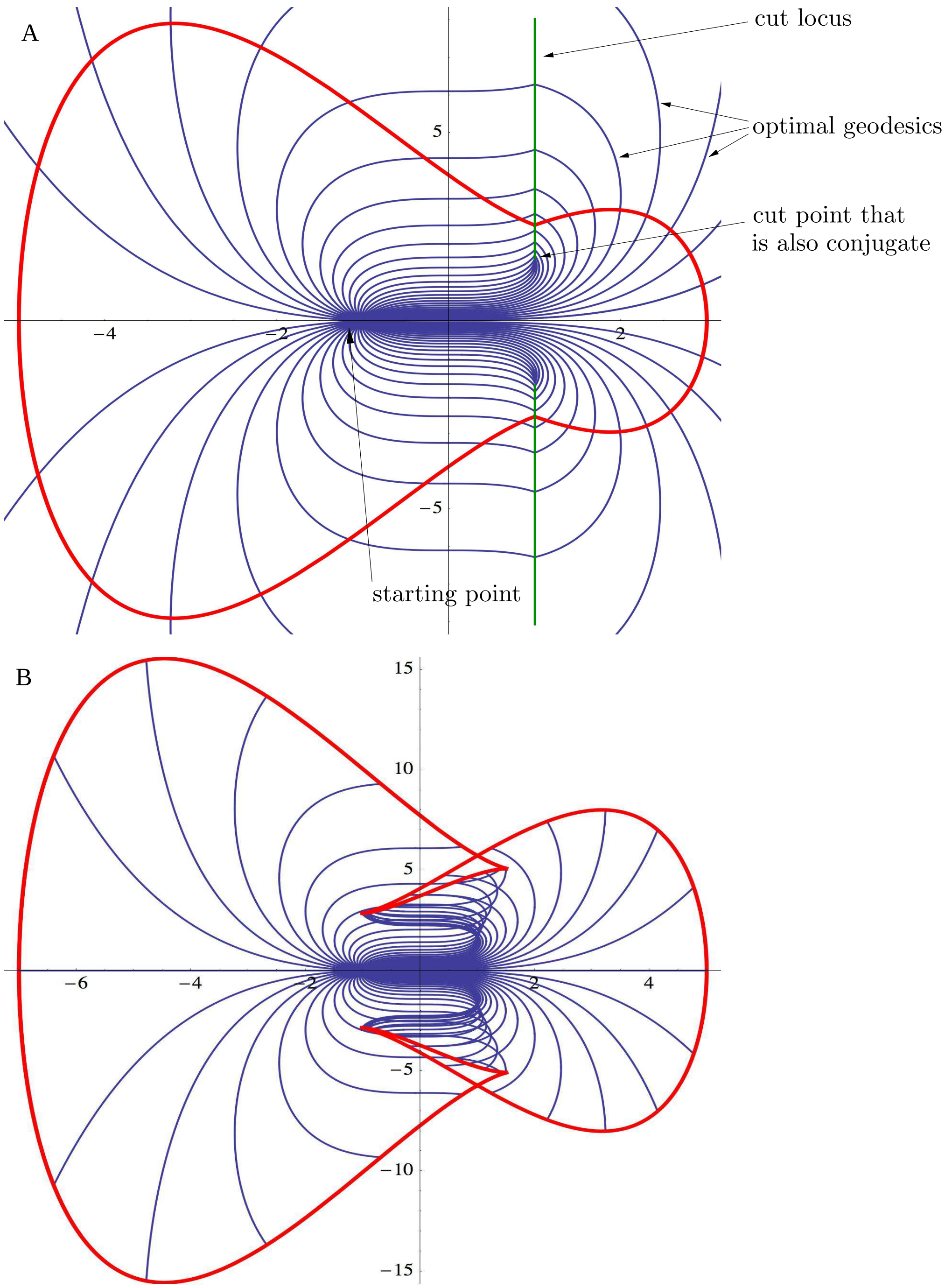



Figure 13.4: A: the optimal synthesis for the Grushin plane starting from the point $(-1, 0)$, together with the sub-Riemannian sphere of radius 4. B: all geodesics up to length 6 with the corresponding wave front.

We have then in polar coordinates

$$N_1 = \{(\rho,\theta) \mid \rho < \left| x_0 \frac{\pi}{\sin(\theta)} \right| \}.$$

In cartesian coordinates

$$N_1 = \{(p_1,p_2) \in T^*\mathbb{R}^2 : |p_2| < \pi\},$$
$$N_2 = \exp(N_1) = \{(x,y) \in \mathbb{R}^2 \mid (x,y) \notin \mathrm{Cut}^*_{x_0}\}.$$

**Step 2: Study of the conjugate points**

In this step we have to prove that there are no conjugate points in $N_1$. In other words we have to prove the following Lemma:

**Lemma 13.33.** *The geodesic $q(\cdot,\theta)$ has no conjugate points in $[0, t^*_{\mathrm{cut}}(\theta))$.*

*Proof.* Since the zeros of $D(\cdot,\theta)$ are not explicitly computable we proceed in the following way. By symmetry we can assume $x_0 > 0$ and $\theta \in [0,\pi]$. We have that

- $D(0,\theta) = 0$. Notice however that this does not mean that $t = 0$ is a conjugate time. Indeed in $x_0$ the structure is Riemannian and $D(0,\theta)$ vanishes only as a consequence of the choice of polar coordinates.
- $D(t^*_{\mathrm{cut}}(\theta),\theta) = \pi x_0^2 \frac{\cos^2\theta}{\sin^3\theta}$. This quantity is always larger than zero except for $\theta = \pi/2$ where it is zero.
- $\partial_t D(t,\theta) = \dfrac{(x_0 + t\cos\theta)\left(\sin(\theta + \frac{t\sin\theta}{x_0})\right)}{\sin\theta}$. Notice that this function is positive in $t = 0$. Let us study when this function is zero in the interval $(0, t^*_{\mathrm{cut}}(\theta))$. We have two type of zeros.
  - **Type one** when $x_0 + t\cos\theta = 0$, which means $t = -\frac{x_0}{\cos\theta}$. This value belongs to $(0, t^*_{\mathrm{cut}}(\theta))$ when $\theta \in (\bar\theta, \pi]$ where $\bar\theta = -\arctan(\pi) \simeq 1.88$. One immediately verifies that this zero corresponds to a minimum of $D(\cdot,\theta)$ and that the value of this minimum is positive.
  - **Type two** when $\theta + \frac{t\sin\theta}{x_0} = k\pi$ with $k = 0,1,2,\ldots$ which means $t = \frac{x_0}{\sin\theta}(k\pi - \theta)$. This value belongs to $(0, t^*_{\mathrm{cut}}(\theta))$ if and only if $k = 1$. One immediately verifies that this zero corresponds to a maximum of $D(\cdot,\theta)$ and that the value of this maximum is positive.

By this analysis it follows that $D(\cdot,\theta)$ is a function that is zero in zero; it has positive derivative in zero; it is positive at $t^*_{\mathrm{cut}}(\theta)$ (zero only when $\theta = \pi/2$); it has a maximum and a minimum (possible only a maximum) in which it is positive.

It follows that $D(\cdot,\theta)$ is never zero in $(0, t^*_{\mathrm{cut}}(\theta))$. Since $t = 0$ is not a conjugate point, it follows that there are no conjugate points in $[0, t^*_{\mathrm{cut}}(\theta))$.

□

**Step 3** We are now going to prove that the map $\exp : N_1 \to N_2$ is proper. But this is obvious since

- all points of the form $(p_1, \pm\pi)$ are mapped in points of $\mathrm{Cut}^*_{x_0}$;

- the image of any sequence in $N_1$ with $p_1 \to \infty$ (resp. $p_1 \to -\infty$) is mapped in a sequence tending to the point $(0, \infty)$ (resp. $(0, -\infty)$).

**Step 4 (R)** Since $N_2$ is simply connected, the application of Corollary 13.24 permits to conclude that exp is a diffeomorphism between $N_1$ to $N_2$. As a consequence the conjectured cut locus and cut times are the true ones.

### 13.5.2 Optimal synthesis starting from a singular point

Let us construct the optimal synthesis starting from a singular point. By invariance of the structure by $y$-translations we can assume that the starting point is the origin. In this case the condition $H(x(0), y(0), p_1(0), p_2(0)) = 1/2$ becomes $p_1^2 = 1$. We have then $p_1 = \pm 1$. Setting $p_2(0) = a$, the expression of the normal Pontryagin extremals parameterized by arclenght is $q^{\pm}(t,a) = (x^{\pm}(t,a), y(t,a))$ where

$$\begin{cases} x^{\pm}(t,0) = \pm t, \qquad y(t,0) = 0, \\ x^{\pm}(t,a) = \pm \dfrac{\sin(at)}{a}, \qquad y(t,a) = \dfrac{2at - \sin(2at)}{4a^2}, \qquad \text{for } a \neq 0. \end{cases} \tag{13.51}$$

**Theorem 13.34.** *The cut time for the geodesic $q^{\pm}(\cdot, a)$ is*

$$t_{\text{cut}}(a) = \frac{\pi}{|a|}.$$

*For $a = 0$ this formula should be interpreted in the sense that the corresponding geodesics $q^{\pm}(\cdot, 0)$ are optimal in $[0, +\infty)$. The cut locus is*

$$\text{Cut}_{(0,0)} = \{(0, y) \in \mathbb{R}^2 \mid y \neq 0\},$$

*and each point of the cut locus is reached exactly by two optimal geodesics.*

The optimal synthesis starting from the origin for Grushin plane is depicted in Figure 13.5.

#### Proof of Theorem 13.34

We give a proof of Theorem 13.34 by making a direct computation, without using the extended Hadamard technique. See also Exercise 13.35.

Due to the fact that the family of geodesics $\{q^-(\cdot, a)\}_{a \in \mathbb{R}}$ can be obtained from the family $\{q^+(\cdot, a)\}_{a \in \mathbb{R}}$ by reflection with respect to the $y$ axis, any geodesic starting from the origin has lost its optimality after intersection with the $y$ axis. From the expression of $x^{\pm}(t,a)$ one gets that for a given value of $a$, the first intersection with the $y$ axis occurs at time $t = \pi/|a|$.

Moreover the family $\{q^{\pm}(\cdot, a)\}_{a \in \mathbb{R}^+}$ can be obtained from the family $\{q^{\pm}(\cdot, a)\}_{a \in \mathbb{R}^-}$ by reflection with respect to the $x$ axis. Notice that the positive (resp. negative) part of the $x$ axis is the support of the geodesic $q^+(\cdot, 0)$ (resp. $q^-(\cdot, 0)$) and no other geodesic starting from the origin can intersect again the $x$ axis since $y(t,a)$ is monotone in $t$.

Then we can restrict ourself to the octant $x \geq 0$ $y \geq 0$ and we would like to prove the following:

**Claim.** For every $\bar{x} > 0$ and $\bar{y} \geq 0$ there exists a unique $a \geq 0$ and $t \in (0, \pi/a]$ such that

$$x^+(t,a) = \bar{x} \tag{13.52}$$

$$y(t,a) = \bar{y}. \tag{13.53}$$

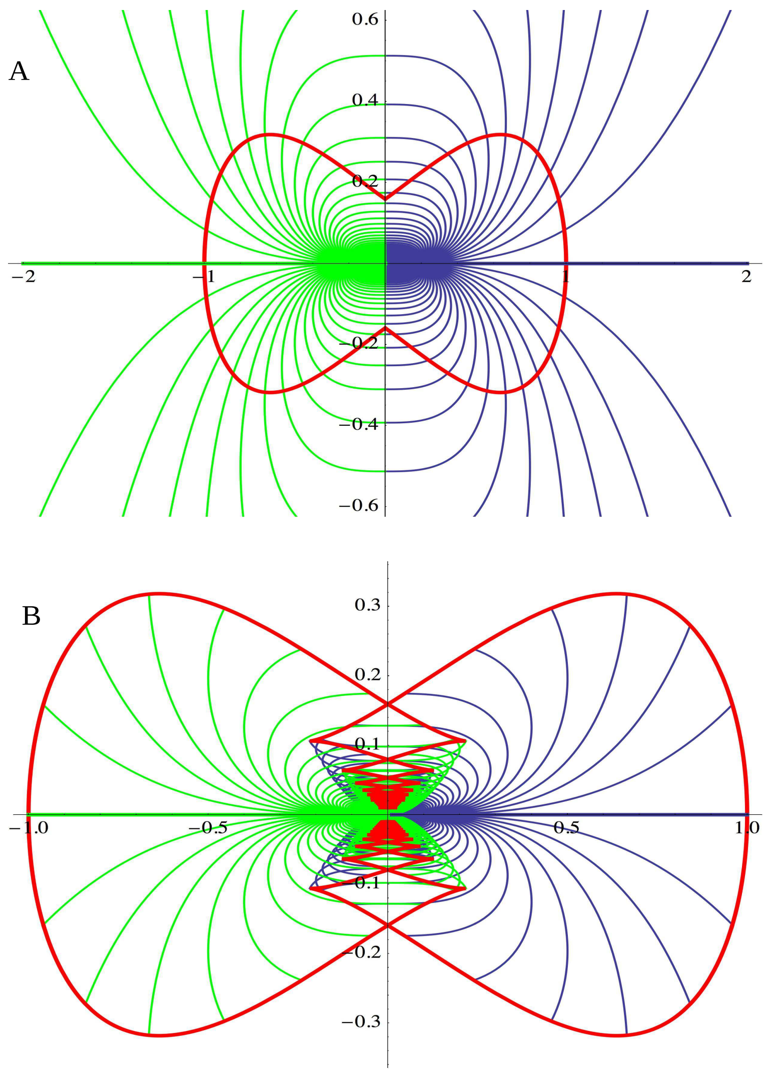



Figure 13.5: A: the optimal synthesis for the Grushin plane starting from the origin, together with the sub-Riemannian sphere for $t = 1$. B: all geodesics up to time 1 with the corresponding wave front.

*Proof of the Claim.* Fix $a$. Let us try to find $t(a)$ from equation (13.52). We have that such an equation has no solutions if $1/a < \bar{x}$ and has two (possibly coinciding) solutions if $1/a \geq \bar{x}$. Such solutions are

$$t_1(a) = \frac{\arcsin(a\bar{x})}{a},$$
$$t_2(a) = \frac{\pi - \arcsin(a\bar{x})}{a}.$$

Notice that $t_1(a) \leq t_2(a)$ and $t_1(a) = t_2(a)$ if and only if $1/a = \bar{x}$.

Let us compute $y(t_1(a), a)$ and $y(t_2(a), a)$. We have

$$y(t_1(a), a) = \frac{1}{4a^2}\big(2\arcsin(a\bar{x}) - \sin(2\arcsin(a\bar{x}))\big).$$

Using the formula $\sin(2\arcsin\xi) = 2\xi\sqrt{1-\xi^2}$, we have

$$y(t_1(a), a) = \frac{1}{4a^2}\big(2\arcsin(a\bar{x}) - 2a\bar{x}\sqrt{1-a^2\bar{x}^2}\big).$$

It is not difficult to check that such function is continuous and monotone increasing in the interval $a \in [0, \frac{1}{\bar{x}}]$. It takes all values from 0 to $\pi\bar{x}^2/4$. Similarly

$$y(t_2(a), a) = \frac{1}{4a^2}\big(2\pi - 2\arcsin(a\bar{x}) + 2a\bar{x}\sqrt{1-a^2\bar{x}^2}\big).$$

It is not difficult to check that such function is continuous and monotone decreasing in the interval $a \in [0, \frac{1}{\bar{x}}]$. It takes all values from $\pi\bar{x}^2/4$ to $+\infty$. The functions $y(t_1(a), a)$ and $y(t_2(a), a)$ are pictured in Figure 13.6.

Concluding, given $\bar{x}$ and $\bar{y}$, we have two cases.

- If $\bar{y} \leq \pi\bar{x}^2/4$ then it is in the image of $y(t_1(a), a)$. Since $y(t_1(a), a)$ is monotone, one can invert it and getting the required unique value of $a$. The corresponding value of $t$ is then obtained from $t_1(a)$.

- If $\bar{y} > \pi\bar{x}^2/4$ then it is in the image of $y(t_2(a), a)$. Since $y(t_2(a), a)$ is monotone, one can invert it and getting the required unique value of $a$. The corresponding value of $t$ is then obtained from $t_2(a)$.

**Exercise 13.35.** Prove Theorem 13.34 using the extended Hadamard technique. Notice that in this case $N_1$ is not connected, hence one should apply twice the technique to its connected components.

## 13.6 The standard sub-Riemannian structure on $SU(2)$

The Lie group $SU(2)$ is the group of unitary unimodular $2 \times 2$ complex matrices

$$SU(2) = \left\{ \begin{pmatrix} \alpha & \beta \\ -\overline{\beta} & \overline{\alpha} \end{pmatrix} \in \mathrm{Mat}(2, \mathbb{C}) \mid |\alpha|^2 + |\beta|^2 = 1 \right\}.$$

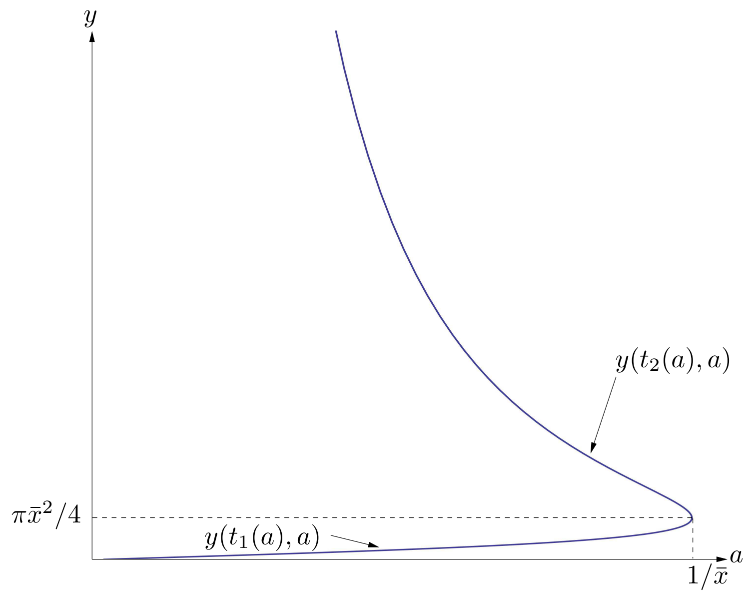



Figure 13.6: Proof of Theorem 13.34.

The Lie algebra of $SU(2)$ is the algebra of antihermitian traceless $2 \times 2$ complex matrices

$$\mathfrak{su}(2) = \left\{ \begin{pmatrix} i\alpha & \beta \\ -\overline{\beta} & -i\alpha \end{pmatrix} \in \mathrm{Mat}(2, \mathbb{C}) \mid \alpha \in \mathbb{R}, \beta \in \mathbb{C} \right\}.$$

A basis of $\mathfrak{su}(2)$ is $\{p_1, p_2, k\}$ where

$$p_1 = \frac{1}{2}\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \quad p_2 = \frac{1}{2}\begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix}, \quad k = \frac{1}{2}\begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}, \tag{13.54}$$

whose commutation relations are $[p_1, p_2] = k, \quad [p_2, k] = p_1, \quad [k, p_1] = p_2$.

For $\mathfrak{su}(2)$ we have $\mathrm{Kil}(X, Y) = 4\mathrm{Tr}(XY)$. In particular, $\mathrm{Kil}(p_i, p_j) = -2\delta_{ij}$, $\mathrm{Kil}(p_i, k) = 0$, $\mathrm{Kil}(k, k) = -2$. Hence

$$\langle \cdot \mid \cdot \rangle = -\frac{1}{2}\mathrm{Kil}(\cdot, \cdot)$$

is a positive definite bi-invariant metric on $\mathfrak{su}(2)$ (cf. Section 7.2.3).

If we define

$$\mathbf{d} = \mathrm{span}\{p_1, p_2\}, \quad \mathbf{s} = \mathrm{span}\{k\}$$

and we provide $\mathbf{d}$ with the metric $\langle \cdot \mid \cdot \rangle |_{\mathbf{d}}$ we get a sub-Riemannian structure of the type $\mathbf{d} \oplus \mathbf{s}$ (cf. 7.7.1).

Notice that since we are in dimension 3 and with one bracket one gets the Lie algebra $\mathfrak{su}(2)$, this problem is a contact sub-Riemannian problem and hence there are no non-trivial abnormal extremals.

*Remark* 13.36. Observe that all the $\mathbf{d} \oplus \mathbf{s}$ structures that one can define on $SU(2)$ are equivalent. For instance, one could set $\mathbf{d} = \mathrm{span}\,\{p_2, k\}$ and $\mathbf{s} = \mathrm{span}\,\{p_1\}$.

Recall that $SU(2) \simeq S^3 = \left\{ \begin{pmatrix} \alpha \\ \beta \end{pmatrix} \in \mathbb{C}^2 \mid |\alpha|^2 + |\beta|^2 = 1 \right\}$ via the map

$$\begin{array}{cccc} & SU(2) & \to & S^3 \\ \phi : & \begin{pmatrix} \alpha & \beta \\ -\overline{\beta} & \overline{\alpha} \end{pmatrix} & \mapsto & \begin{pmatrix} \alpha \\ \beta \end{pmatrix}. \end{array}$$

In the following we often write elements of $SU(2)$ as pairs of complex numbers.

Notice that in this representation the sub-group $e^{\mathbb{R}k}$ is

$$\left\{ \begin{pmatrix} \alpha \\ 0 \end{pmatrix} \mid |\alpha|^2 = 1 \right\}.$$

and the identity of the group is $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$.

**Expression of geodesics**

Let us write an initial covector in $\mathfrak{su}(2)$ as $x_0+y_0$, where $x_0 \in \mathbf{d}$ and $y_0 \in \mathbf{s}$. To parametrize geodesics by arclength, i.e., to be on the level set $\frac{1}{2}$ of the Hamiltonian, we have to require $\langle x_0 \,|\, x_0 \rangle = 1$. It is then convenient to write

$$x_0 + y_0 = \underbrace{\cos(\theta)p_1 + \sin(\theta)p_2}_{x_0} + \underbrace{ck}_{y_0}, \quad \theta \in S^1, \quad c \in \mathbb{R}.$$

Using formula (7.46), we have that the normal Pontryagin extremals starting from the identity are (here $\lambda = (\theta, c)$)

$$\exp_{\mathrm{Id}}(t, \lambda) = g(\theta, c; t) := e^{t(x_0+y_0)} e^{-ty_0} = e^{(\cos(\theta)p_1 + \sin(\theta)p_2 + ck)t} e^{-ckt} =$$

$$= \begin{pmatrix} \frac{c \sin(\frac{ct}{2}) \sin(\sqrt{1+c^2}\frac{t}{2})}{\sqrt{1+c^2}} + \cos(\frac{ct}{2}) \cos(\sqrt{1+c^2}\frac{t}{2}) + i \left( \frac{c \cos(\frac{ct}{2}) \sin(\sqrt{1+c^2}\frac{t}{2})}{\sqrt{1+c^2}} - \sin(\frac{ct}{2}) \cos(\sqrt{1+c^2}\frac{t}{2}) \right) \\ \frac{\sin(\sqrt{1+c^2}\frac{t}{2})}{\sqrt{1+c^2}} \left( \cos(\frac{ct}{2} + \theta) + i \sin(\frac{ct}{2} + \theta) \right) \end{pmatrix}.$$

*Remark* 13.37. We have the following cylindrical symmetry reflecting the invariance of the sub-Riemannan structure with respect to rotations along the $k$ axis.

$$g(\theta, c; t) = \begin{pmatrix} 1 & 0 \\ 0 & e^{i\theta} \end{pmatrix} g(0, c, t);$$

**Theorem 13.38.** *The cut time for the geodesic $g(\theta, c, t)$ coincides with its first conjugate time. It is independent from $\theta$ and it is given by the formula*

$$t_{\mathrm{cut}}(c) = \frac{2\pi}{\sqrt{1+c^2}}.$$

*Moreover $g(\theta, c; t_{\mathrm{cut}}(c))$ is independent from $\theta$. Hence each cut point is reached by an infinite number of geodesics (a one parameter family parameterized by $\theta$).*

Since the largest cut time is obtained for $c = 0$ we have

**Corollary 13.39.** *The diameter of $SU(2)$ with the standard sub-Riemannian structure is $2\pi$.*

By a direct computation one gets

**Corollary 13.40.** *The cut locus starting from the identity is*

$$\mathrm{Cut}_{\mathrm{id}} = e^{\mathbb{R}k} \setminus \{id\} = \left\{ \begin{pmatrix} \alpha \\ 0 \end{pmatrix} \mid |\alpha|^2 = 1, \alpha \neq 1 \right\}.$$

*Moreover each cut point is also a conjugate point.*

*Remark* 13.41. Notice that with our definition of cut locus, the starting point is never a cut point.

*Proof of Theorem 13.38.* We are going to apply the extended Hadamard technique.

**Step 1: Construction of the conjectured cut locus and of the sets $N_1$ and $N_2$.**

By a direct computation one immediately obtains:

**Lemma 13.42.** *For every $\theta_1, \theta_2 \in S^1$, we have*

$$g\left(\theta_1, c; \frac{2\pi}{\sqrt{1+c^2}}\right) = g\left(\theta_1, c; \frac{2\pi}{\sqrt{1+c^2}}\right) = \begin{pmatrix} -\cos\left(\frac{\pi c}{\sqrt{c^2+1}}\right) + i\sin\left(\frac{\pi c}{\sqrt{c^2+1}}\right) \\ 0 \end{pmatrix}$$

*Moreover the determinant of the differential of the exponential map is zero if and only if*

$$\sin\left(\sqrt{1+c^2}\frac{t}{2}\right)\left(2\sin\left(\sqrt{1+c^2}\frac{t}{2}\right) - \sqrt{1+c^2}t\cos\left(\sqrt{1+c^2}\frac{t}{2}\right)\right) = 0. \tag{13.55}$$

*In particular $\frac{2\pi}{\sqrt{1+c^2}}$ is a conjugate time for the geodesic $g(\theta, c; \cdot)$.*

We then conjecture that the cut time of the geodesic $g(\theta, c; \cdot)$ is $t^*_{\mathrm{cut}}(c) = \frac{2\pi}{\sqrt{1+c^2}}$ and that the cut locus is

$$\mathrm{Cut}^*_{\mathrm{id}} = e^{\mathbb{R}k} = \left\{ \begin{pmatrix} \alpha \\ 0 \end{pmatrix} \mid |\alpha|^2 = 1, \alpha \neq 1 \right\}.$$

We define

$$N_1 = \{ap_1 + bp_2 + ck \in \mathfrak{su}(2) \mid (a,b) \neq (0,0), |c| \leq \sqrt{2\pi - 1}\},$$

and

$$N_2 = \exp(N_1) = \{g \in SU(2) \mid g \notin \mathrm{Cut}^*_{\mathrm{Id}}\}.$$

**Step 2: Study of the conjugate points**

We are going to prove that the differential of the exponential map never vanishes in $N_1$ and hence that there are no conjugate points in $N_2$ for $\exp_{\mathrm{Id}}|_{N_1}$. Conjugate times are given by formula (13.55). The first term vanishes at times $\frac{2m\pi}{\sqrt{1+c^2}}$, where $m = 1, 2, \ldots$. The second term vanishes at times $\frac{2x_m}{\sqrt{1+c^2}}$ where $\{x_1, x_2, \ldots\}$ is the ordered set of the strictly positive solutions of $x = \tan(x)$. Since $x_1 \sim 4.49 > \pi$, the first positive time at which the geodesic $g(\theta, c; \cdot)$ is conjugate is $t^*_{\mathrm{cut}}(c)$, Hence the differential of the exponential map never vanishes in $N_1$.

**Step 3** We are now going to prove that the map $\exp : N_1 \to N_2$ is proper. But this is obvious since all points of $\partial N_1$ are mapped in points of $\partial N_2$.

**Step 4 (SR)** By Theorem 13.23 we know that $\exp : N_1 \to N_2$ is a covering. It remains to prove that it is a 1-covering. As already mentioned we cannot apply Corollary 13.24 since $N_2$ is not simply connected. Let us show that the hypotheses of Corollary 13.25 are verified. We have that $N_2$ is homeomorphic to $S^1 \times \mathbb{R}^2$. We are left to find a loop in $N_1$ that is mapped via the exponential map in a loop homotopic to $S^1$. Indeed as we know from Chapter 10, the nilpotent approximation of every 3D-contact structure is the Heisenberg group. For the Heisenberg group a loop $\ell_2$ winding once the cut locus is the image through the exponential map of a loop $\ell_1$.

Since for regular maps, the structure of the preimage of a set does not change for small perturbation of the map it follows that for $SU(2)$ a small loop winding $\mathrm{Cut}^*_{\mathrm{id}}$ is the image through the exponential map of a loop $\ell_1$. Then Corollary 13.25 permits to conclude that $\exp|_{N_1}$ is a diffeomorphism. As a consequence the conjectured cut locus and cut times are the true ones.

*Remark* 13.43. The argument above apply to any 3 dimensional structure that is genuinely sub-Riemannian at the starting point.

**Exercise 13.44.** Corollary 13.39 says that the diameter of $SU(2)$ for the standard sub-Riemannian structure is $2\pi$. Prove that the diameter of $SU(2)$ for the standard Riemannian structure (i.e., the structure for which $\{p_1, p_2, k\}$ is an orthonormal frame) is $2\pi$ as well.

A representation of the cut locus for $SU(2)$ is given in Figure 13.7.

**Exercise 13.45.** Consider the $\mathbf{d} \oplus \mathbf{s}$ sub-Riemannian structure on $SO(3)$ introduced in Section 7.7.2. By using the techniques presented in this chapter construct the optimal synthesis. Represent $SO(3)$ as a full three dimensional ball with opposite points on the boundary identified. Call this "boundary" $\mathbb{R}P^2$. Prove that the cut locus is the union of the subgroup $e^{\mathbb{R}e_3} = e^{\mathbf{s}}$ without the identity and $\mathbb{R}P^2$. Compute the diameter of $SO(3)$ for this structure. Compare it with the diameter of $SO(3)$ for the standard Riemannian structure (i.e., the structure for which $\{e_1, e_2, e_3\}$ is an orthonormal frame). An alternative technique to compute this optimal synthesis is provided in Section 13.7.

**Exercise 13.46.** Let $G = SL(2)$ and consider the left-invariant sub-Riemannian structure for which an orthonormal frame is given by

$$X_1(g) = L_{g*}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \qquad X_2(g) = L_{g*}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Prove that this structure is of type $\mathbf{d} \oplus \mathbf{s}$ for the metric induced by the Killing form. Construct the optimal synthesis starting from the identity.

## 13.7 Optimal synthesis on the groups $SO(3)$ and $SO_+(2,1)$.

In this section we find the time optimal synthesis for the structures on $SO(3)$ and $SO_+(2,1)$ introduced in Section 7.7.3. Here, instead of using the extended Hadamard technique, we use a more geometric approach using the Gauss-Bonnet theorem.

To describe these synthesis it is very convenient to use the interpretation of geodesics as parallel transports along curves of a constant geodesic curvature in the unit sphere $S^2$ and the Lobachevsky plane $H$ (see Section 7.7.3).

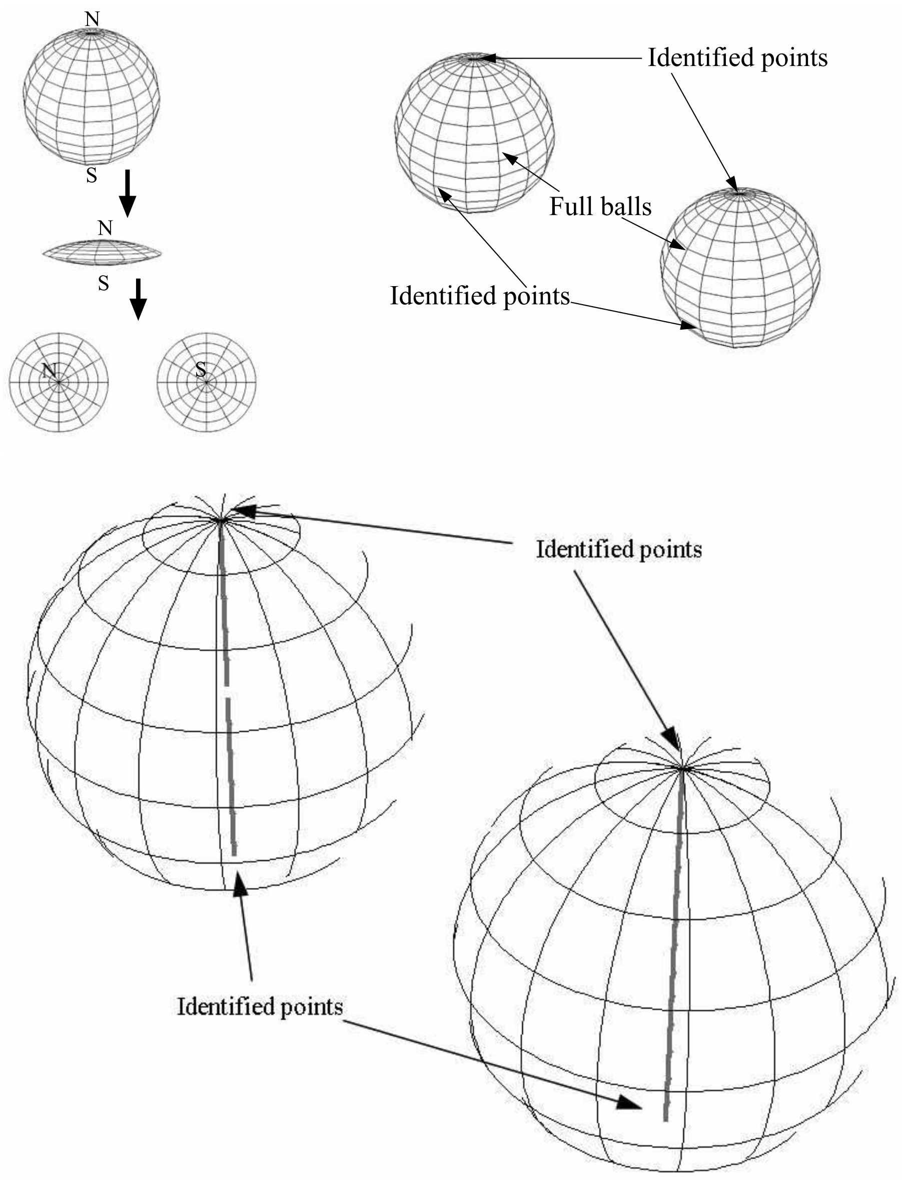



Figure 13.7: We recall a standard construction for representing $S^2$ in a two dimensional space and $S^3$ in a three dimensional one. Consider $S^2 \subset \mathbb{R}^3$ and flatten it on the equator plane, pushing the northern hemisphere down and the southern hemisphere up, getting two disks $D^2$ joined along their circular boundaries. The construction is drawn in the up-left side of the figure. Similarly, consider $S^3 \subset \mathbb{C}^2 \simeq \mathbb{R}^4$: it can be viewed as two balls joined along their boundaries. In this case the boundaries are two spheres $S^2$. A picture of $S^3$ is drawn in the up-right side of the figure. In this representation, the cut locus is given by the great circle passing through the identity, the north and the south pole (the identity should then be removed, cf. Remark 13.2).

According to the general scheme, we use nontrivial symmetries of the structure that preserve the endpoints of the geodesics in order to characterize the cut locus. In the cases under consideration, the sub-Riemannian space is identified with the spherical bundle of the surface. This allows us to give a nice and clear description of the cut locus in terms of natural symmetries of the surface. As we'll see, the Gauss-Bonnet formula plays a key role. Here we give a brief description of the cut locus; detailed proofs can be found in [BZ15b, BZ15a, BZ16] but we advise the reader to recover them by him/herself.

The projection of a geodesic to the surface is a curve of a constant geodesic curvature. First we describe symmetries of the surface that preserve endpoints of the curve. We use two essentially different types of symmetries. The first one concerns the case when the curve is closed, i.e., the initial point is equal to the final one. In this case, the initial and final velocities are also equal. The symmetries are just rotations of the surface around the initial point of the curve. We obtain a one-parametric family of symmetries where the angle of rotation is the parameter of the family.

The second type concerns any curve. If the endpoints of the curve are different then the symmetry is the reflection of the surface with respect to the geodesic (of the Riemannian surface) that contains both endpoints. If the endpoints are equal (the curve is closed) then the symmetry is the reflection of the surface with respect to the geodesic that is tangent to the curve at the initial point.

Now we turn to the parallel transport. Let $\gamma : [0, 1] \to M$ be a curve of constant geodesic curvature $\rho \in \mathbb{R}$ and the length $\ell > 0$. Let $v_0 \in S_{\gamma(0)}M$ and let $\theta_0$ be the angle between $\dot\gamma(0)$ and $v_0$. Then the parallel transport of $v_0$ along $\gamma$ is a vector $v_1 \in S_{\gamma(1)}M$ such that the angle between $\dot\gamma(1)$ and $v_1$ equals $\theta_0 + \rho\ell$.

A rotation around a point does not change neither the geodesic curvature nor the length of the curve; hence the parallel transport along the curve does not change as well. Let $\gamma(1) = \gamma(0)$ and $\Gamma \subset M$ be a compact domain such that $\gamma = \partial\Gamma$. The Gauss-Bonnet formula implies a relation:

$$\rho\ell = 2\pi \pm \mathrm{Area}(\Gamma).$$

Let $q \in M$; it follows that the rotation of the circle $S_qM$ on any angle can be realized as the parallel transport along a closed curve of constant geodesic curvature (recall that angles are defined modulo $2\pi$). We see that for any $v_0, v_1 \in S_qM$ there exists a one-parametric family of sub-Riemannian geodesics of the same length that connect $v_0$ with $v_1$.

Now we consider reflections. Let $\xi$ be the shortest path connecting $\gamma(1)$ with $\gamma(0)$ and $\phi$ be the angle between $\dot\gamma(0)$ and $\dot\xi(1)$. Then the angle between $\dot\gamma(1)$ and $\dot\xi(0)$ equals $-\phi$ (see Figure 13.8).

The reflection of $M$ with respect to the geodesic changes the sign of the geodesic curvature and the sign of $\phi$.

To compute the parallel transport along the curve $\gamma$ and along the reflected curve we choose the directions of $\dot\xi(1)$ and $\dot\xi(0)$ as the origins in the circles $S_{\gamma(0)}M$ and $S_{\gamma(1)}M$. Then the direction of $\dot\gamma(0)$ is $-\phi$ and the direction of $\dot\gamma(1)$ is $+\phi$. Hence the parallel transport of $\dot\xi(1)$ along $\gamma$ has the direction

$$\phi + \rho\ell + \phi = \rho\ell + 2\phi.$$

The parallel transport of the same vector along the reflected curve has the direction $-\rho\ell - 2\phi$. The parallel transports along the two curves coincide if and only if

$$2(\rho\ell + 2\phi) \equiv 0 \bmod 2\pi.$$

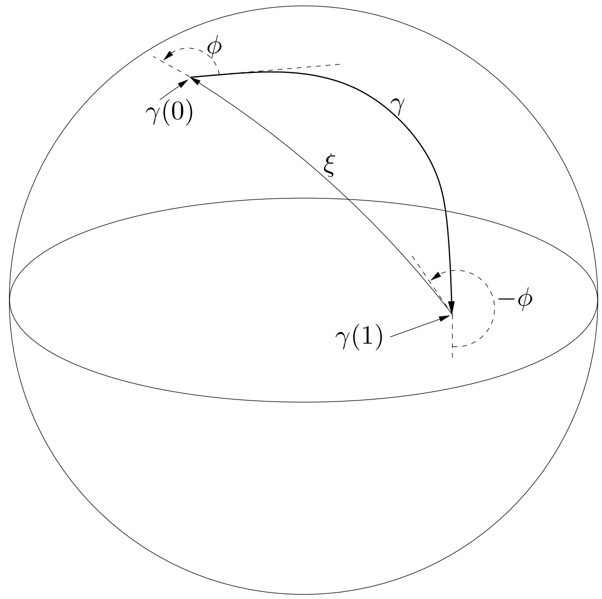



Figure 13.8: Construction of the optimal synthesis on $SO(3)$ and $SO_+(2,1)$. Definition of the angle $\phi$. (The picture refers to $SO(3)$)

Let us consider the curve $\bar{\gamma} = \gamma \cup \xi$ and the domain $\Gamma \subset M$ such that $\bar{\gamma} = \partial\Gamma$ (see the figure). The Gauss-Bonnet formula (1.33) applied to $\Gamma$ gives the relation:

$$\rho\ell + 2\phi \pm \text{Area}(\Gamma) = 2\pi.$$

If $M$ is the unit sphere, then $\rho\ell + 2\phi = 2\pi - \text{Area}(\Gamma)$. Hence $\rho\ell + 2\phi = \pi$ provides a natural candidate to the cut locus. If $M$ is the Lobachevsky plane, then $\rho\ell + 2\phi = 2\pi + \text{Area}(\Gamma)$ and a natural candidate to the cut locus is provided by $\rho\ell + 2\phi = 3\pi$. Both cases are characterized by the identity:

$$\text{Area}(\Gamma) = \pi.$$

We are now ready to describe the optimal synthesis. Let $M$ be either the unit sphere in the three-dimensional Euclidean space or the hyperbolic plane in the Minkowski space.

1. Geodesics are curves in $SM$ that are parallel transports along curves of a constant geodesic curvature in $M$, and curves of constant geodesic curvature are just the intersections of $M \subset \mathbb{R}^3$ with affine planes.

2. Let $t \mapsto \gamma(t)$ be a parameterized curve of constant geodesic curvature in $M$ and $\Gamma_t \subset M$ be the smaller domain between the two domains whose boundary is the concatenation of $\gamma|_{[0,t]}$ and the shortest path connecting $\gamma(t)$ with $\gamma(0)$. We assume that $\gamma$ is oriented in such a way that $\Gamma_t$ stays to the right from $\gamma$ (as in the figure). The cut time $t_\gamma$ for the parallel transport along $\gamma$:
$$t_\gamma = \min\{t > 0 : \gamma(t) = \gamma(0) \text{ or } \text{Area}(\Gamma_t) = \pi\}.$$

If $M = S^2$, then the maximal length until the cut point (the sub-Riemannian diameter of $SO(3)$) is equal to $\sqrt{3}\pi$ and is achieved when the equations $\gamma(t) = \gamma(0)$ and $\text{Area}(\Gamma_t) = \pi$ happen simultaneously. If $M = H$, then the surface is not compact and the diameter is equal to $+\infty$.

## 13.8 Synthesis for the group of Euclidean transformations of the plane $SE(2)$

The group of (positively oriented) Euclidean transformations of the plane is

$$SE(2) = \left\{ \left( \begin{array}{cc|c} \cos(\theta) & -\sin(\theta) & x_1 \\ \sin(\theta) & \cos(\theta) & x_2 \\ \hline 0 & 0 & 1 \end{array} \right), \quad \theta \in S^1, \quad x_1, x_2 \in \mathbb{R} \right\}.$$

The name of this group comes from the fact that if we represent a point of $\mathbb{R}^2$ as a column vector $(y_1, y_2, 1)^t$ then the action of a matrix of $SE(2)$ produces a rotation of angle $\theta$ and a translation of $(x_1, x_2)$ (cf. Section 7.2.2). The Lie algebra of $SE(2)$ is

$$\mathfrak{se}(2) = \operatorname{span}\{e_1, e_2, e_r\},$$

where

$$e_1 = \left( \begin{array}{cc|c} 0 & 0 & 1 \\ 0 & 0 & 0 \\ \hline 0 & 0 & 0 \end{array} \right), \quad e_2 = \left( \begin{array}{cc|c} 0 & 0 & 0 \\ 0 & 0 & 1 \\ \hline 0 & 0 & 0 \end{array} \right), \quad e_r = \left( \begin{array}{cc|c} 0 & -1 & 0 \\ 1 & 0 & 0 \\ \hline 0 & 0 & 0 \end{array} \right).$$

The commutation relations are:

$$[e_1, e_2] = 0, \quad [e_1, e_r] = -e_2, \quad [e_2, e_r] = e_1. \tag{13.56}$$

The sub-Riemannian problem on $SE(2)$ is obtained by declaring $\{e_1, e_r\}$ to be an orthonormal frame. In this way the sub-Riemannian problem can be written as (here $T > 0$ and $g_0, g_1$ are two fixed points in $SE(2)$),

$$\dot{g} = g(ue_1 + ve_r), \tag{13.57}$$

$$\int_0^T \sqrt{u(t)^2 + v(t)^2}\, dt, \to \min, \tag{13.58}$$

$$g(0) = g_0, \quad g(T) = g_1. \tag{13.59}$$

Notice that since we are in dimension 3 and with one bracket one gets the Lie algebra $\mathfrak{se}(2)$, this problem is a contact sub-Riemannian problem and hence there are no non-trivial abnormal extremals.

In coordinates $q = (x_1, x_2, \theta)$ this problem becomes

$$\dot{q} = uX_1(q) + vX_r(q), \tag{13.60}$$

$$\int_0^T \sqrt{u(t)^2 + v(t)^2}\, dt \to \min, \tag{13.61}$$

$$q(0) = q_0, \quad q(T) = q_1. \tag{13.62}$$

where

$$X_1 = \begin{pmatrix} \cos(\theta) \\ \sin(\theta) \\ 0 \end{pmatrix}, \quad X_r = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}. \tag{13.63}$$

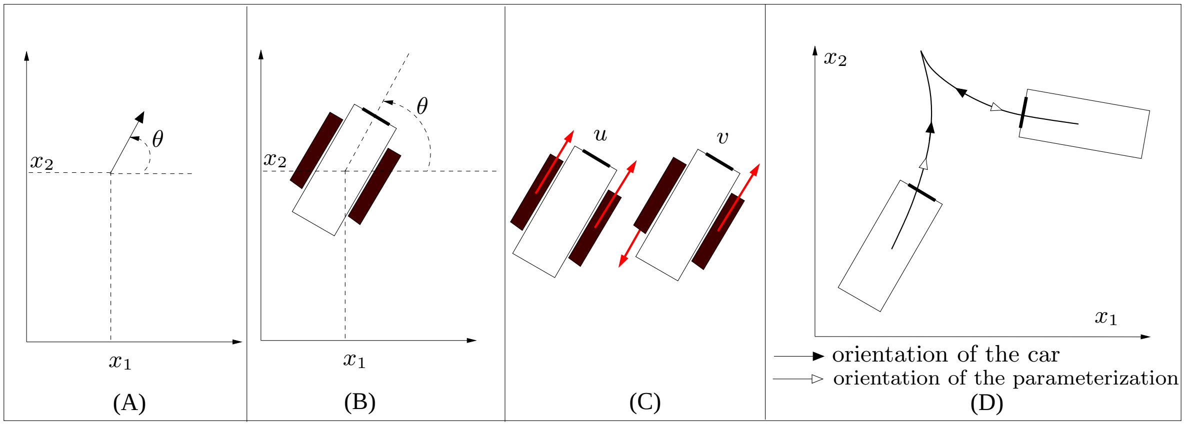



Figure 13.9: Mechanical interpretation of the problem on $SE(2)$.

Notice that if we define

$$-X_2 = [X_1, X_r] = \begin{pmatrix} \sin(\theta) \\ -\cos(\theta) \\ 0 \end{pmatrix},$$

the commutation relations are the same as (13.56) i.e., $[X_1, X_2] = 0$, $[X_1, X_r] = -X_2$ and $[X_2, X_r] = X_1$.

**Exercise 13.47.** Prove that every left-invariant sub-Riemannian structure on $SE(2)$ is isometric to the structure presented above, modulus a dilation in the $(x_1, x_2)$ plane.

### 13.8.1 Mechanical interpretation

Recall that a point $(x_1, x_2, \theta) \in SE(2)$ can be represented as a unit vector on the plane applied to the point $(x_1, x_2)$ with an angle $\theta$ with respect to the $x_1$ axis (see Figure 13.9 (A)). Then the optimal control problem (13.60)-(13.63) can be interpreted as the problem of controlling a car with two wheels on the plane. More precisely $x_1$ and $x_2$ are the coordinates of the center of the car, $\theta$ is the orientation of the car with respect to the $x_1$ direction (see Figure 13.9 (B)). The first control $u$ makes the two wheels rotating in the same directions and makes the car going forward with velocity $u$; the second control $v$ makes the two wheels rotating in opposite direction and makes the car rotating with angular velocity $v$ (see Figure 13.9 (C)). An admissible trajectory in $SE(2)$ can be represented as a planar trajectory with two type of arrows: an "empty" arrow giving the direction of the parameterization of the curve and a "bold" arrow indicating the orientation of the car (see Figure 13.9 (D)). Notice that in the drawn trajectory there is a cusp point where the car stops to go forward and starts to go backward. Indeed a smooth admissible trajectory in $SE(2)$ can have cusp points in this representation.

### 13.8.2 Geodesics

The maximized Hamiltonian for the problem (13.60), (13.61), (13.62), (13.63) is

$$H(p, q) = \frac{1}{2}\left(\langle p, X_1 \rangle^2 + \langle p, X_2 \rangle^2\right).$$

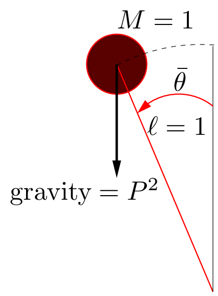



Figure 13.10: The inverted pendulum

Setting $p = (p_1, p_2, p_\theta)$, $p_1 = P\cos(p_a)$, $p_2 = P\sin(p_a)$ we have

$$H = \frac{1}{2}\left((p_1\cos\theta + p_2\sin\theta)^2 + p_\theta^2\right) = \frac{1}{2}\left(P^2\cos^2(\theta - p_a) + p_\theta^2\right).$$

The Hamiltonian equations are then

$$\begin{aligned}
\dot{x}_1 &= \frac{\partial H}{\partial p_1} = P\cos(\theta - p_a)\cos\theta, & \dot{p}_1 &= -\frac{\partial H}{\partial x_1} = 0,\\
\dot{x}_2 &= \frac{\partial H}{\partial p_2} = P\cos(\theta - p_a)\sin\theta, & \dot{p}_2 &= -\frac{\partial H}{\partial x_2} = 0,\\
\dot{\theta} &= \frac{\partial H}{\partial p_\theta} = p_\theta, & \dot{p}_\theta &= -\frac{\partial H}{\partial \theta} = \frac{1}{2}P^2\sin(2(\theta - p_a)).
\end{aligned}$$

Notice that this Hamiltonian system is completely integrable in the sense of Chapter 5, since we have three first integrals independent and in involution (i.e., $H, p_1, p_2$ or equivalently $H, P, \theta$). The last two equations give rise to

$$\ddot{\theta} = \frac{1}{2}P^2\sin(2(\theta - p_a)).$$

Now setting $\bar{\theta} = 2(\theta - p_a) \in 2S^1 = \mathbb{R}/(4\pi\mathbb{Z})$ that is the double covering of the standard circle $S^1 = \mathbb{R}/(2\pi\mathbb{Z})$, we get the equation

$$\ddot{\bar{\theta}} = P^2\sin\bar{\theta}. \tag{13.64}$$

This is the equation of a planar pendulum of mass 1, length 1, where $P^2$ represents the gravity (see Figure 13.10). In the following we will have to remember that $\dot{\bar{\theta}} = 2p_\theta$.

**Initial conditions.** By invariance by rototranslation we can assume $x_1(0) = 0$, $x_2(0) = 0$, $\theta(0) = 0$ which means $\bar{\theta}(0) = -2p_a$. Geodesics are then parameterized by $p_1$, $p_2$ (which are constants) and by $p_\theta(0)$ (or alternatively by $P, p_a, p_\theta(0)$). If we require that geodesics are parametrized by arclenght, we have $H(0) = \frac{1}{2}$ hence the initial covector belongs to the cylinder

$$p_1^2 + p_\theta(0)^2 = 1, \qquad \text{i.e.,} \qquad P^2\cos^2 p_a + p_\theta(0)^2 = 1.$$

Fixed an initial covector $p(0)$ on the cylinder $H(0) = 1/2$ one get $P, p_a, p_\theta(0)$. Then one has to consider the pendulum equation (13.64) with gravity $P^2$ and initial condition

$$\bar{\theta}(0) = -2p_a, \qquad \dot{\bar{\theta}}(0) = 2p_\theta(0).$$

Once the pendulum equation has been solved one obtains

$$\theta(t) = \frac{\bar{\theta}(t)}{2} + p_a \tag{13.65}$$

$$x_1(t) = \int_0^t \dot{x}_1(s)\,ds = P\int_0^t \cos(\theta(s) - p_a)\cos\theta(s)\,ds = P\int_0^t \cos\left(\frac{\bar{\theta}(s)}{2}\right)\cos\left(\frac{\bar{\theta}(t)}{2} + p_a\right)ds \tag{13.66}$$

$$x_2(t) = \int_0^t \dot{x}_2(s)\,ds = P\int_0^t \cos(\theta(s) - p_a)\sin\theta(s)\,ds = P\int_0^t \cos\left(\frac{\bar{\theta}(s)}{2}\right)\sin\left(\frac{\bar{\theta}(t)}{2} + p_a\right)ds \tag{13.67}$$

**Qualitative behaviour of the geodesics.**
Equation (13.64) admits an explicit solution in terms of elliptic functions. However the qualitative behaviour of the solutions can be understood without integrating it explicitly.

In particular this equation admits the first integral (the energy of the pendulum)

$$H_p = \frac{1}{2}\dot{\bar{\theta}}^2 + P^2\cos\bar{\theta}.$$

Notice that this constant of the motion is not independent from $H$. Indeed a simple computation gives:

$$H_p = 4H - P^2.$$

Since we are working on the level set $H = 1/2$, it will be much more convenient to work directly with $H$ that here we write in terms of the new variables

$$H = \frac{1}{2}\left(P^2\cos^2\left(\frac{\bar{\theta}}{2}\right) + p_\theta^2\right).$$

The level sets of $H$ are plotted in Figure 13.11. We are interested to the level set $H = 1/2$. Depending on the value of $(P, p_a, p_\theta(0))$ different types of the trajectories of the pendulum are possible. Notice that

- when $\bar{\theta}$ passes monotonically through $\pi$, then the projection on the $(x_1, x_2)$ plane of the geodesic has a cusp.
- Geodesics are parameterized by $(P, p_a, p_\theta(0)) \in H^{-1}(1/2)$. Changing $P$ correspond to change the gravity of the pendulum. This changes the period of the trajectories oscillating close the stable equilibrium and the time between two cusps. Notice that $P$ enters also in the equations for $x_1(t)$ and $x_2(t)$. Changing $p_a$ and $p_\theta(0)$ corresponds to change the starting point on the pendulum trajectory.

**Classification of normal Pontryagin extremals.**
We have the following type of trajectories (see Figure 13.12):

- Trajectories with $P > 0$ and corresponding to the rotating pendulum. In this case $\bar{\theta}(t)$ increases monotonically. Notice that the projection of the geodesics on the plane $(x_1, x_2)$ has a cusp each time that $\bar{\theta}$ passes through $\pi + 2k\pi$ with $k \in \mathbb{N}$.

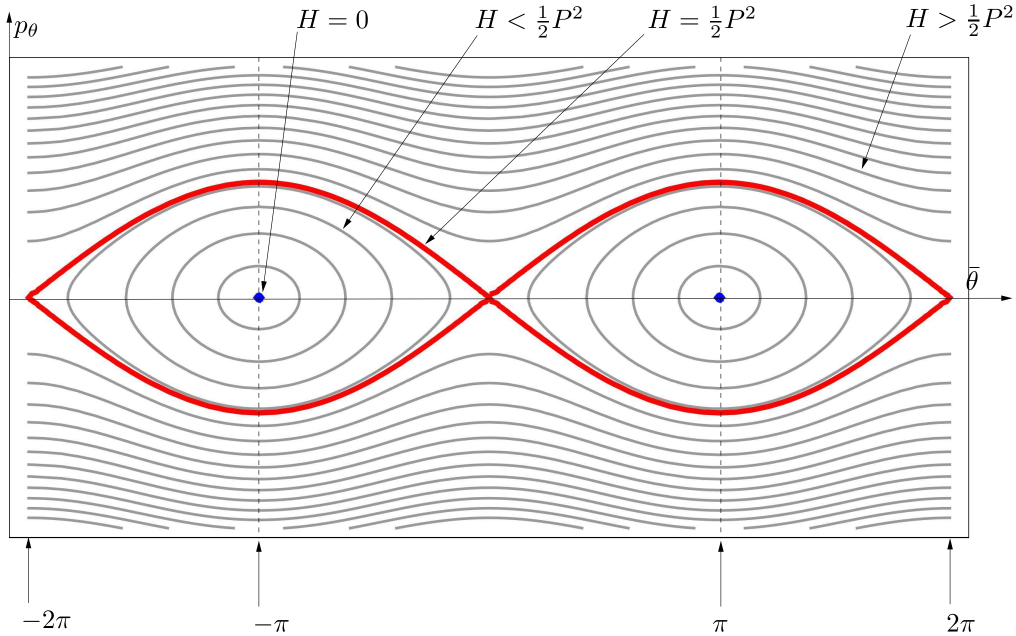



Figure 13.11: Trajectories of the inverted pendulum

- Trajectories with $P>0$ and corresponding to the oscillating pendulum. In this case $\bar{\theta}(t)$ is oscillating either around $\pi$ or around $-\pi$. Notice that the projection of the geodesics on the plane $(x_1, x_2)$ has a cusp each time that $\bar{\theta}$ passes through $\pi$ or $-\pi$. One can easily check that these trajectories have an inflection point between two cusps.

- Trajectories with $P>0$ and staying on the separatrix (but not on the unstable equilibria). The projection on the $(x_1, x_2)$ plane of these trajectories has at most one cusp.

- Trajectories with $P>0$ and staying on one of the unstable equilibria. In this case we have $p_\theta = 0$ and $p_a = 0$ (or $p_a = 2\pi$). As a consequence we have $\theta(t)=0$, $x_1(t) = \pm t$, $x_2(t) = 0$.

- Trajectories corresponding to $P=0$ in this case each level set of the pendulum is an horizontal line and equation (13.64) is reduced to $\ddot{\theta}(t) = 0$. then we have $\bar{\theta}(t) = -2p_a + 2p_\theta(0)t$, with $p_\theta(0) = \pm 1$. As a consequence we have $\theta(t) = \pm t$, $x_1(t) = 0$, $x_2(t) = 0$.

*Remark* 13.48. Notice that trajectoreis with $P>0$ and staying at one of the two stable equilibria have $H=0$ and they are abnormal extremals. For these trajectories $\bar{\theta} = \pm\pi$, $p_a = \mp\pi/2$. Hence $x_1(t) \equiv 0$, $x_2(t) \equiv 0$, $\theta(t) \equiv 0$. This is the trivial trajectory staying fixed at the identity.

**Optimality of geodesics**.
Let $q(\cdot) = (x_1(\cdot), x_2(\cdot), \theta(\cdot))$ defined on $[0,T]$ be a geodesic parameterized by arclength. Define the two mapping of geodesics

$$\mathbb{S} : q(\cdot) \mapsto q_{\mathbb{S}}(\cdot) \text{ and } \mathbb{T} : q(\cdot) \mapsto q_{\mathbb{T}}(\cdot)$$

in the following way. In the mechanical representation given above, consider the segment $\ell$ joining $(x_1(0), x_2(0))$ and $(x_1(T), x_2(T))$ and the line $\ell^\perp$ passing through the middle point of $\ell$ and orthogonal to $\ell$.

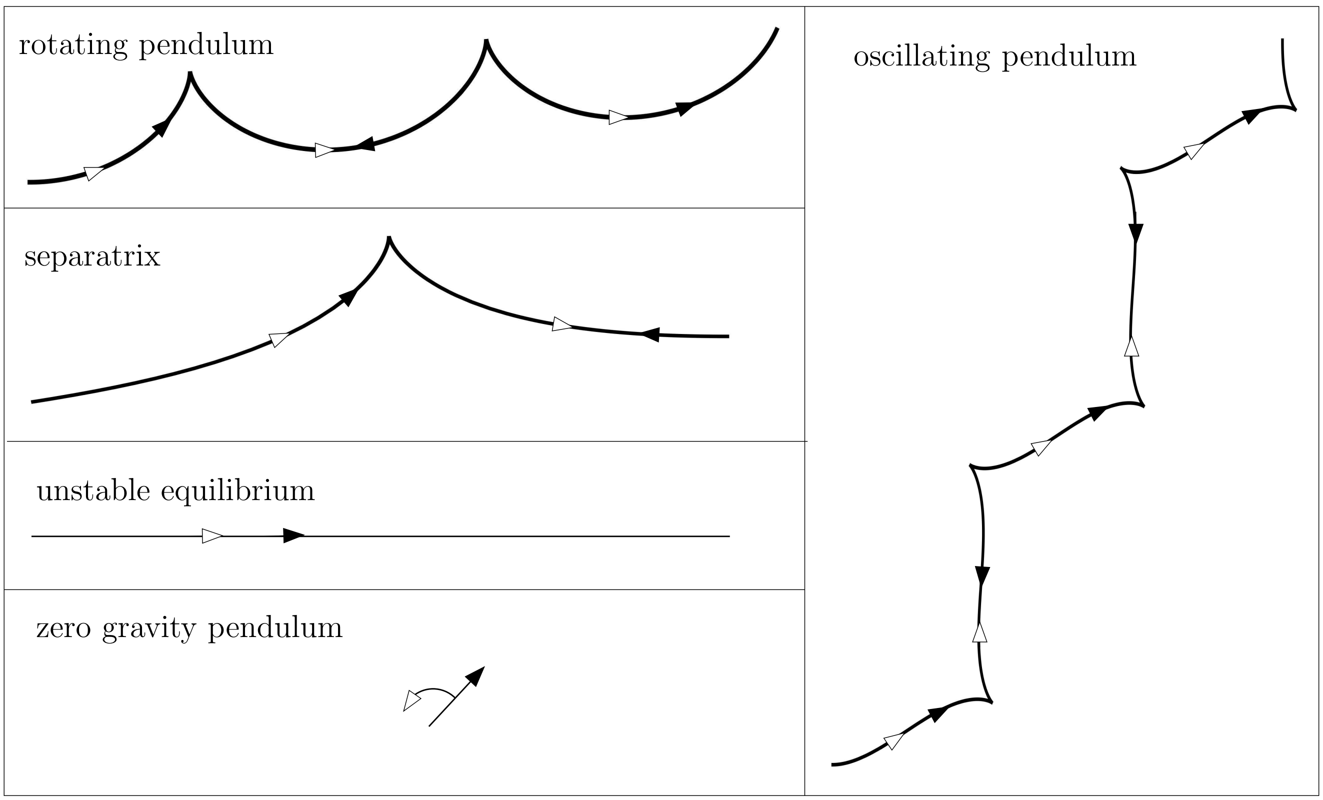



Figure 13.12: Geodesics for $SE(2)$

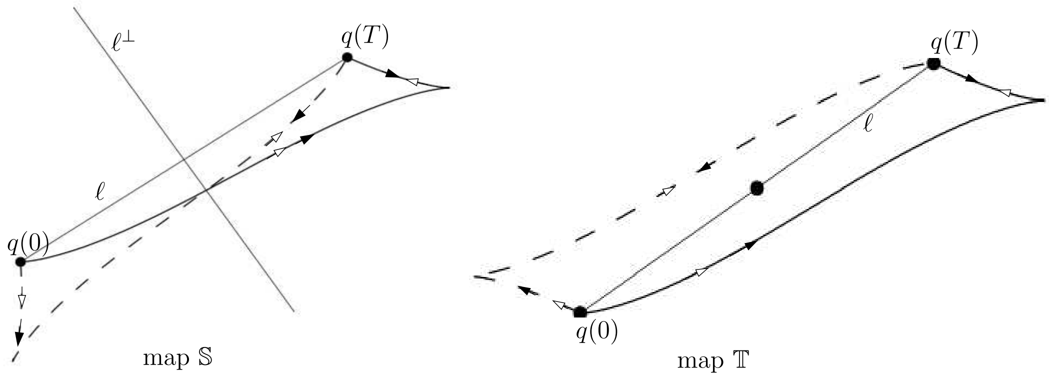



Figure 13.13: Maps $\mathbb{S}$ and $\mathbb{T}$. Courtesy of Y. Sachkov.

**Map** $\mathbb{S}$ the trajectory $q_{\mathbb{S}}(\cdot)$ is the trajectory obtained by considering the reflection of $q(\cdot)$ with respect to $\ell^\perp$.

**Map** $\mathbb{T}$ The trajectory $q_{\mathbb{T}}(\cdot)$ is the trajectory obtained by considering the reflection of $q(\cdot)$ with respect to the middle point of $\ell$.

In both cases the "bold arrows" should be reflected accordingly. The "empty arrows" giving the direction of the parameterization should be oriented in such a way that the initial (resp. final) point of $q_{\mathbb{S}}(\cdot)$ is $q(0)$ (resp. $q(T)$). The same holds for $q_{\mathbb{T}}(\cdot)$. See Figure 13.13.

*Remark* 13.49. Notice that if $q(\cdot)$ is defined in $[0,T]$ then in general $\mathbb{S}q(\cdot)$ is different from $\mathbb{S}\big(q(\cdot)|_{[0,t]}\big)$ for $t \in (0,T)$. The same applies to $\mathbb{T}q(\cdot)$.

**Definition 13.50.** Let $q(\cdot)$ defined on $[0,T]$ be a geodesic. We say that $q(T)$ is a Maxwell point corresponding to $\mathbb{S}$ (resp. $\mathbb{T}$) if $q(\cdot) \neq q_{\mathbb{S}}(\cdot)$ (resp. $q(\cdot) \neq q_{\mathbb{T}}(\cdot)$), $q(0) = q_{\mathbb{S}}(0)$ and $q(T) = q_{\mathbb{S}}(T)$ (resp. $q(T) = q_{\mathbb{T}}(T)$).

Examples of Maxwell points for $\mathbb{S}$ and $\mathbb{T}$ are shown at Figures 13.14. The following result, whose proof is out of the scope of this book, is due to Yuri Sachkov.

**Theorem 13.51.** *A geodesic $q(\cdot)$ on the interval $[0,T]$, is optimal if and only if each point $q(t)$, $t \in (0,T)$, is neither a Maxwell points corresponding to $\mathbb{S}$ or $\mathbb{T}$ for $q(\cdot)|_{[0,t]}$ nor the limit of a sequence of Maxwell points.*

The cut locus for the sub-Riemannian problem on $SE(2)$ has been computed by Y. Sachkov and it is pictured in Figure 13.15.

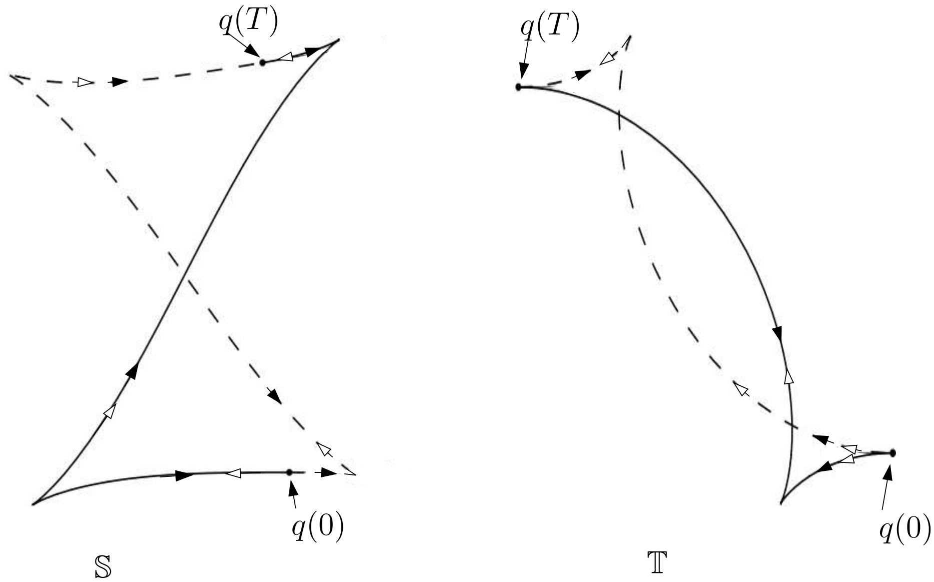



Figure 13.14: Cut loci corresponding to $\mathbb{S}$ and $\mathbb{T}$. Courtesy of Y. Sachkov.

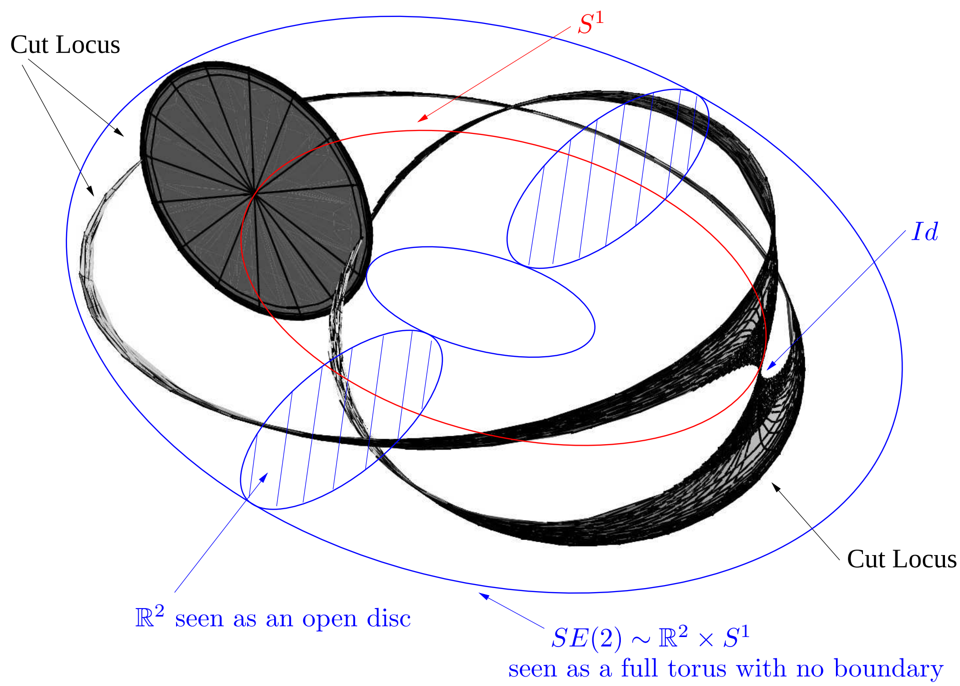



Figure 13.15: Cut locus (dark region) from the identity for the sub-Riemannian problem on $SE(2)$. Courtesy of Y. Sachkov. In this picture $SE(2)$ (that has the topology of $\mathbb{R}^2 \times S^1$) is represented as a solid torus without boundary given by $B_2 \times S^1$, where $B_2$ is the 2D disc without boundary.

## 13.9 The Martinet flat sub-Riemannian structure

Let us write a point of $\mathbb{R}^3$ as $(x, y, z)$. The Martinet flat sub-Riemannian structure is the structure in $\mathbb{R}^3$ for which an orthonormal frame is given by

$$X_1 = \begin{pmatrix} 1 \\ 0 \\ \frac{y^2}{2} \end{pmatrix}, \quad X_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}. \tag{13.68}$$

*Remark* 13.52. This problem can be formulated as an isoperimetric problem in the sense of Section 4.4.2. In this case the base manifold is given by the points $(x, y) \in \mathbb{R}^2$ and the 1-form $A$ defining the problem is $A = \frac{y^2}{2}dx$. In other words the trajectory realizing the sub-Riemannian distance for the Martinet sub-Riemannian structure between $(0, 0, 0)$ and $(x_1, y_1, z_1)$ is a curve $\gamma(t) = (x(t), y(t), z(t))$ defined in $[0, T]$ steering $(0, 0, 0)$ to $(x_1, y_1, z_1)$, for which

$$\int_\gamma A = \int_0^T A(\dot\gamma(t))dt = \int_0^T \frac{y(t)^2}{2}\dot x(t)dt = z_1,$$

and whose projection in the $(x, y)$-plane is a length-minimizer for the Euclidean distance.

This structure is bracket generating, but it is not equiregular. Indeed we have

$$X_3 := [X_1, X_2] = \begin{pmatrix} 0 \\ 0 \\ -y \end{pmatrix}, \qquad [X_3, X_2] = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}. \tag{13.69}$$

Hence the distribution is contact out of $\{y = 0\}$ and bracket-generating of step 3 at $\{y = 0\}$.

In the following two sections we are going to construct the Pontryagin extremals. We already know, by Section 4.4.2, that the support of abnormal extremals should be contained in the set $\{y = 0\}$. Such set is called the *Martinet set*, and in this case is a smooth surface. Let us use the notation $p = (p_x, p_y, p_z)$.

### 13.9.1 Abnormal extremals

For abnormal extremals we have for every $t$,

$$\begin{aligned} 0 &= \langle p(t), X_1(q(t)\rangle = p_x(t) + \frac{y(t)^2}{2}p_z(t), \\ 0 &= \langle p(t), X_2(q(t)\rangle = p_y(t). \end{aligned}$$

Differentiating with respect to $t$ we obtain for almost every $t$

$$\begin{aligned} 0 &= u_2(t)\langle p(t), [X_2, X_1](q(t))\rangle = -u_2(t)\langle p(t), X_3(q(t))\rangle = u_2(t)p_z(t)y(t), \\ 0 &= u_1(t)\langle p(t), [X_1, X_2](q(t))\rangle = u_1(t)\langle p(t), X_3(q(t))\rangle = -u_1(t)p_z(t)y(t). \end{aligned}$$

Hence if $\gamma : [a, b] \to \mathbb{R}^3$ is an abnormal extremal, either it is trivial (i.e., $\gamma(t) = \gamma(0)$ for every $t$) or we have

$$\langle p(t), X_3(q(t))\rangle = p_z(t)y(t) \equiv 0. \tag{13.70}$$

Since $(p_x(t), p_y(t), p_z(t))$ cannot vanish, we have that $\gamma$ is contained in the Martinet set i.e., $\gamma([a,b]) \subset \{y = 0\}$.

To obtain the controls corresponding to $\gamma$ let us differentiate once more (13.70). We have for almost every $t$

$$0 = u_1(t)\langle p(t), [X_1, X_3](q(t))\rangle + u_2(t)\langle p(t), [X_2, X_3](q(t))\rangle = -u_2(t)p_z(t),$$

where we used the fact that $[X_1, X_3] = 0$ when $y = 0$, and (13.69). Since again $(p_x(t), p_y(t), p_z(t))$ is never vanishing, we obtain

$$u_2(t) = 0 \qquad \text{for almost every } t.$$

Indeed we already knew this fact since the only way to stay on the Martinet set is to have $u_2(t) = 0$ almost everywhere. The value of $u_1$ is then obtained by requiring that $\gamma$ is parametrized by arlength, i.e., $|u_1(t)| = 1$ for almost every $t$. Notice that we have many of such trajectories: indeed the control $u_1$ can be any measurable function satisfying $|u_1(t)| = 1$ for almost every $t$. Such control can switch arbitrarily between 1 and $-1$. Because of Remark 13.52 only trajectories corresponding to a control that is almost everywhere constant are optimal. We then obtain the following.

**Proposition 13.53.** *Arclength parametrized trajectories admitting an abnormal lift are Lipschitz trajectories $\gamma : [a,b] \to \mathbb{R}^3$ lying on the Martinet set and corresponding to $u_2 \equiv 0$ almost everywhere. Among these trajectories, only those for which $u_1$ is constantly equal to $+1$ or $-1$ are optimal.*

### 13.9.2 Normal extremals

For normal extremals, the maximized Hamiltonian is given by

$$H(p,q) = \frac{1}{2}(h_1(p,q)^2 + h_2(p,q)^2),$$

where

$$h_1(p,q) = p_x + \frac{y^2}{2}p_z, \qquad h_2(p,q) = p_y.$$

The Hamiltonian equations are then

$$\dot{x} = \frac{\partial H}{\partial p_x} = h_1, \qquad \dot{p}_x = -\frac{\partial H}{\partial x} = 0, \tag{13.71}$$

$$\dot{y} = \frac{\partial H}{\partial p_y} = p_y, \qquad \dot{p}_y = -\frac{\partial H}{\partial y} = -h_1 y\, p_z, \tag{13.72}$$

$$\dot{z} = \frac{\partial H}{\partial p_z} = h_1\frac{y^2}{2}, \qquad \dot{p}_z = -\frac{\partial H}{\partial z} = 0. \tag{13.73}$$

Notice that this Hamiltonian system is completely integrable, since we have three first integrals independent and in involution (i.e., $H, p_x, p_z$).

From (13.73) we have that $p_z$ is constant. Let us set $p_z = a$. We can solve (13.71) and (13.72) since these equations are independent from $z$. Let us use as coordinates $(x, y, h_1, h_2)$. We have

$$\dot{x} = h_1, \qquad \dot{h}_1 = \dot{p}_x + y\underbrace{\dot{y}}_{p_y}a = a\,y\,h_2, \tag{13.74}$$

$$\dot{y} = p_y = h_2, \qquad \dot{h}_2 = \dot{p}_y = -a\,y\,h_1. \tag{13.75}$$

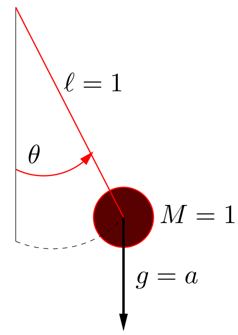



Figure 13.16: The pendulum for the Martinet distribution

Now if consider normal extremals parametrized by arclength, we have

$$\frac{1}{2} = H(q(t), p(t)) = h_1(t)^2 + h_2(t)^2.$$

It is then convenient to set

$$h_1(t) = \cos\theta(t), \qquad h_2(t) = \sin\theta(t).$$

The equations for $h_1$ and $h_2$ in (13.74) and (13.75) give then

$$\begin{aligned} -\sin(\theta)\dot{\theta} &= ay\sin(\theta), \\ \cos(\theta)\dot{\theta} &= -ay\cos(\theta), \end{aligned}$$

from which we have

$$\dot{\theta} = -ay. \tag{13.76}$$

This equation together with $\dot{y} = h_2 = \sin\theta$ (see the equation for $\dot{y}$ in (13.75)) gives

$$\ddot{\theta} = -a\sin\theta \tag{13.77}$$

We obtain again an equation for a pendulum of unit mass, unit length and gravity $a$. See Figure 13.16.

**Initial conditions**

We are going to consider normal Pontryagin extremals starting from the point $(x, y, z) = (0, 0, 0)$. Arclength geodesics are then parameterized by $\theta_0 := \theta(0)$ (giving $p_y(0)$ and $p_x$) and by $a$. Notice that from (13.76) we have that $\dot{\theta}(0) = 0$.

Once the pendulum equation has been solved, one gets

$$x(t) = \int_0^t \dot{x}(s)\,ds = \int_0^t \; h_1(q(s), p(s))\,ds = \int_0^t \cos\theta(s)\,ds, \tag{13.78}$$

$$y(t) = \int_0^t \dot{y}(s)\,ds = \int_0^t \; h_2(q(s), p(s))\,ds = \int_0^t \sin\theta(s)\,ds, \tag{13.79}$$

$$z(t) = \int_0^t \dot{z}(s)\,ds = \int_0^t \; h_1(q(s), p(s))\frac{y^2(s)}{2}\,ds = \int_0^t \cos(\theta(s))\frac{y^2(s)}{2}\,ds. \tag{13.80}$$

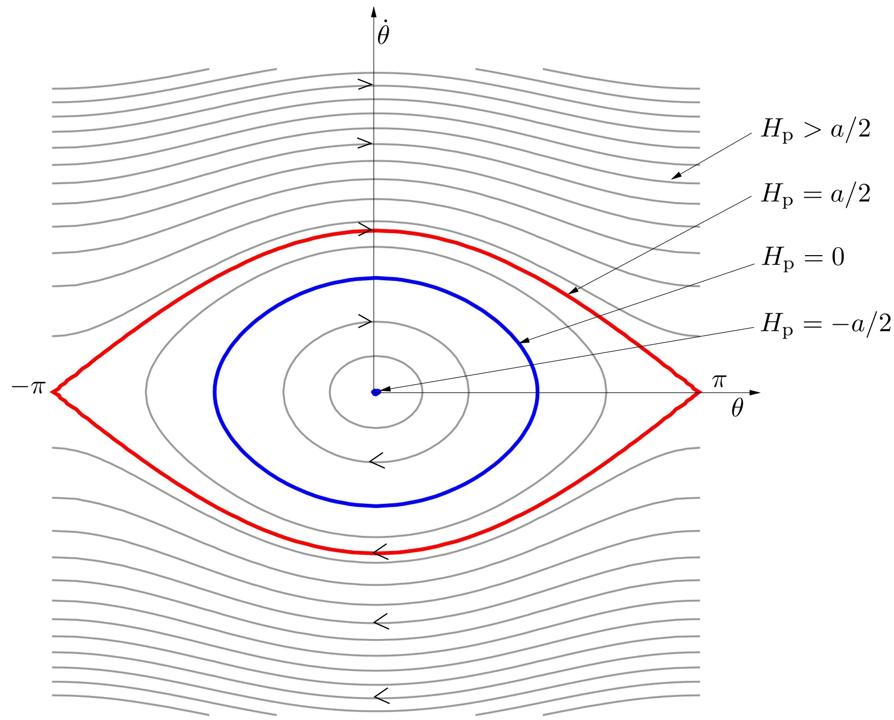



Figure 13.17: The phase portrait of the pendulum for the Martinet flat sub-Riemannian structure

The solution of the pendulum equation and the corresponding expressions for $x(t)$, $y(t)$ and $z(t)$ can be expressed in terms of elliptic functions. Here we are going to make a short qualitative analysis.

We already know that the pendulum equation admits the first integral

$$H_p(\theta, \dot\theta) = \frac{1}{2}\dot\theta^2 - a\cos(\theta).$$

Level sets of $H_p$ are plotted in Figure 13.17.

**Case** $a = 0$. In this case the level set of $H_p$ are horizontal lines. We have $\ddot\theta \equiv 0$ hence $\dot\theta(t) =$const. This constant is indeed zero since $\dot\theta(0) = 0$. Then $\theta(t) = \theta_0$. From (13.78)-(13.80) we have

$$x(t) = t\cos(\theta_0), \qquad y(t) = t\sin(\theta_0), \qquad z(t) = \cos(\theta_0)\sin^2(\theta_0)\frac{t^3}{6}.$$

For $\theta_0 \in \{0, \pi\}$ this trajectory is lying on the Martinet surface and it is both normal and abnormal.

**Case** $a \neq 0$ **and** $\theta_0 = 0$. This is the trajectory staying at the stable equilibrium of the pendulum. In this case we have $\theta(t) \equiv 0$ and

$$x(t) = t, \qquad y(t) = 0, \qquad z(t) = 0.$$

This trajectory is lying on the Martinet set and it is both normal and abnormal.

**Case** $a \neq 0$ **and** $\theta_0 = \pi$. This is the trajectory staying at the unstable equilibrium of the pendulum. In this case we have $\theta(t) \equiv \pi$ and

$$x(t) = -t, \qquad y(t) = 0, \qquad z(t) = 0.$$

As the previous one, this trajectory is lying on the Martinet set and it is both normal and abnormal. Notice that the heteroclinic orbit is not realized because of the initial condition $\dot{\theta}(0) = 0$.

Notice that all Pontryagin extremals studied up to now have a projection on the $(x, y)$ plane that is a straight line. Because of Remark 13.52 they are automatically optimal.

All other Pontryagin extremals are expressed in terms of Elliptic functions and are given by the Theorem below.

To this purpose let $\mathrm{sn}(\phi, m)$, $\mathrm{cn}(\phi, m)$, $\mathrm{dn}(\phi, m)$ be the standard Jacobi elliptic functions with parameter $m \in [0, 1]$ and recall the definition of:

- the *complete elliptic integral of the first kind*

$$K(m) := \int_0^{\pi/2} \left(1 - m \sin^2(\theta)\right)^{-\frac{1}{2}} d\theta.$$

- the Jacobi *epsilon function* [Law89, p. 62]

$$\mathrm{Eps}(\phi, m) := \int_0^{\phi} \mathrm{dn}^2(w, m)\, dw.$$

Let us define the following functions of $t, \theta_0, a$ (here we assume $a > 0$, $\theta_0 \in (0, \pi)$).

$$k = \sqrt{\frac{1 - \cos(\theta_0)}{2}}, \tag{13.81}$$

$$k' = \sqrt{\frac{1 + \cos(\theta_0)}{2}}, \tag{13.82}$$

$$u(t, k, a) = K(k^2) + t\sqrt{a}, \tag{13.83}$$

$$\Upsilon(t, k, a) = \mathrm{Eps}(u(t, k, a), k^2) - \mathrm{Eps}(K(k^2), k^2), \tag{13.84}$$

The following result, whose proof is out of the scope of this book, is due to Agrachev, Bonnard, Chyba, Kupka.

**Theorem 13.54.** *The normal geodesics starting from the origin for $\theta_0 \in (0, \pi)$ and $a > 0$ are given by:*

$$x(t) = -t + \frac{2}{\sqrt{a}} \Upsilon(t, k, a) \tag{13.85}$$

$$y(t) = -2\frac{k}{\sqrt{a}} \mathrm{cn}(u(t, k, a), k^2) \tag{13.86}$$

$$z(t) = \frac{2}{3a^{3/2}} \left[(2k^2 - 1)\Upsilon(t, k, a) + k'^2 t\sqrt{a} + 2k^2 \mathrm{sn}(u(t, k, a), k^2)\mathrm{cn}(u(t, k, a), k^2)\mathrm{dn}(u(t, k, a), k^2)\right] \tag{13.87}$$

*For negative values of $\theta_0$ and/or $a$, the formulas are obtained from the previous ones considering that a change in sign of $\theta_0$ produces a change of sign in the coordinate $y$ and a change of sign of $a$ produces a change of sign in the coordinates $x$ and $z$.*

*Remark* 13.55. These geodesics can be easily drawn using a commercial software having elliptic functions and integrals implemented, as for instance `Mathematica`. The Jacobi epsilon function can be written in terms of more common elliptic integrals using the formula (see for instance [Law89, p.63])

$$\mathrm{Eps}(\phi, m) = E(\mathrm{am}(\phi, m), m).$$

Here $E(\alpha, m) := \int_0^\alpha \left(1 - m \sin^2(\theta)\right)^{\frac{1}{2}} d\theta$, is the *elliptic integral of the second kind* and am is the *Jacobi amplitude* defined as the inverse of the elliptic integral of the first kind, i.e., if $\phi = F(\alpha, m) := \int_0^\alpha \left(1 - m \sin^2(\theta)\right)^{-\frac{1}{2}} d\theta$, then $\alpha = \mathrm{am}(\phi, m)$.

The optimality of these geodesics is not easy to be studied (the method presented at the beginning of the chapter does not apply directly because of the presence of abnormal minimizers, see also the Bibliographical note). However this study was completed in the '90s. And we have the following result.

**Theorem 13.56** (Agrachev, Bonnard, Chyba, Kupka)**.** *Normal Pontryagin extremals corresponding to $a = 0$ or to $\theta_0 = 0$ (i.e., those for which the projection on the $(x, y)$ plane is a straight line are optimal for every time. All other Pontryagin extremals are optimal up to their first intersection with the Martinet set $\{y = 0\}$. The cut time is given by the formula*

$$t_{\mathrm{cut}} = \begin{cases} 2\frac{K(k^2)}{\sqrt{a}}, & \text{for } a > 0, \\ 2\frac{K(k'^2)}{\sqrt{-a}}, & \text{for } a < 0. \end{cases}$$

The sub-Riemannian sphere of radius one is drawn in Figure 13.18. Its intersection with the Martinet surface (that is also the cut locus) is drown in Figure 13.19 A. In Figure 13.19 B it is pictured the point on the cylinder $H = 1/2$ that are mapped in the cut locus at $t = 1$ namely the points

$$a = (2K(k^2))^2 \qquad \text{and} \qquad a = -(2K(k'^2))^2.$$

Notice that, due to the presence of the abnormal trajectory, the cut locus is the image via the exponential map of an unbounded curve on the cylinder $\{H = 1/2\}$. Points on this curve having large values of $a$ correspond to the part of the sphere that become tangent to the abnormal as pictured.

## 13.10 Bibliographical note

The cut locus is a very old object in geometry. Its first appearance dates back to the famous paper by Poincaré [Poi05], where he uses the terminology "ligne de partage" for the cut locus on a surface. The first appearance of the term cut locus is probably in Whitehead [Whi35].

The literature about the cut locus (and conjugate locus) in Riemannian geometry is huge. For a comprehensive discussion, we refer the interested reader to [Ber03]: more precisely Sections 1.6.2 and Section 3.3 for cut locus of two-dimensional surfaces, and to Section 6.5 for cut locus in arbitrary dimension.

We focus here in the sub-Riemannian case and on the examples considered in this chapter. Explicit computations of Pontryagin extremals and of the cut locus for the Heisenberg group and its higher dimensional generalizations are well-known [Gav77, Bro82, Nac82, GV88, MPAM06,

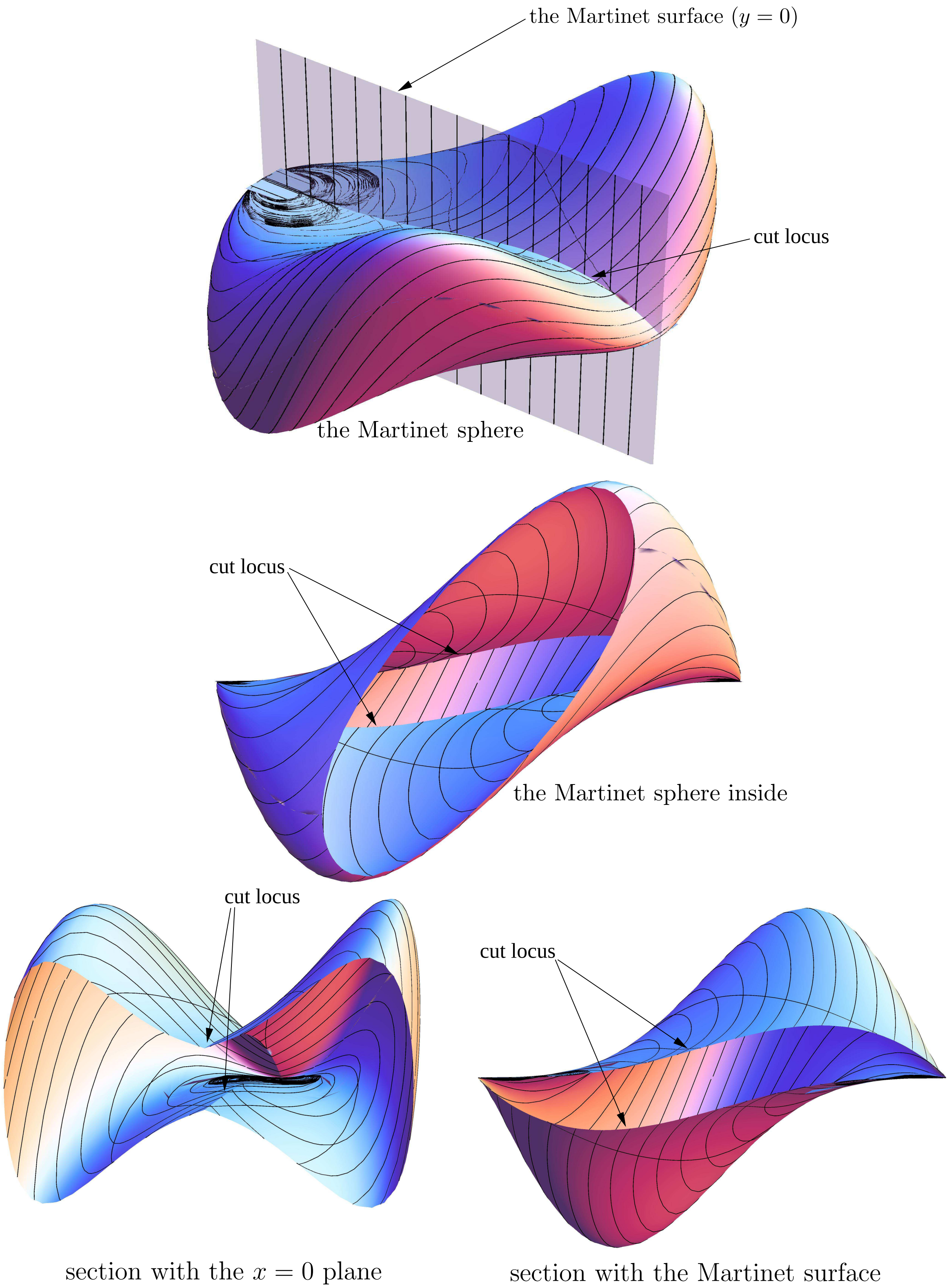



Figure 13.18: The Martinet sphere of radius one.

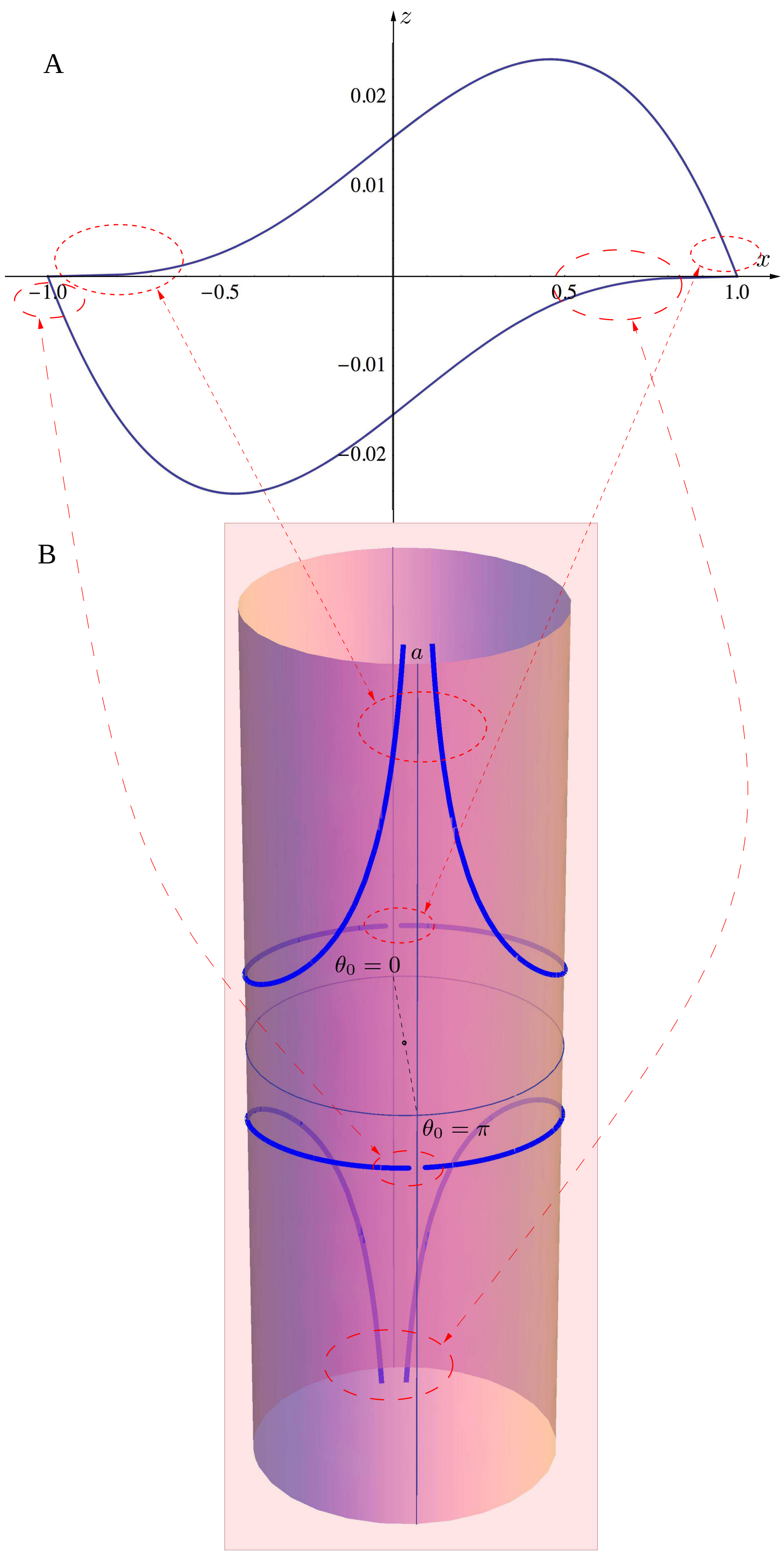



Figure 13.19: A: the intersection of the Martinet sphere for $t = 1$ with the Martinet surface, that is also the cut locus. B: the cut locus seen on the cotangent bundle on $H = \frac{1}{2}$.

ABB12, BBG12, BR18, BBNss, RS17]. It is still an open question to determine the cut locus of the step-two free Carnot group of rank $k \geq 4$. This problem was first studied in [Bro82]. In [RS17] the authors disprove the conjectures on the shape of the cut loci proposed in [Mya02, Mya06] and [MM17], by exhibiting sets of cut points which, for $k \geq 4$, are strictly larger than conjectured ones. For a detailed discussion about this conjecture and the comparison with previous literature we refer the reader to [RS17].

The technique explained in Section 13.4 to compute the cut locus is an extension of a classical technique due to Hadamard that was used in Riemannian geometry, in particular to study the optimal synthesis on surfaces with negative curvature (see [Had98]). Its sub-Riemannian variant was used to construct the optimal syntheses in several cases. See for instance [ABCK97, Mya02, Sac10, Sac11]. This technique cannot be adapted to structures containing strict abnormal minimizers since these trajectories are not seen from the exponential map. In principle one could apply the technique to normal Pontryagin extremals and then one could compare the length of normal and abnormal at points reached by both type of trajectories. However there are no known examples in which such an idea has been successfully employed. With some additional work, the extended Hadamard technique can be adapted to the presence of non-strict abnormal extremals. This program was successful for the construction of the optimal synthesis for the Martinet flat sub-Riemannian structure and in particular to prove Theorem 13.56. See [ABCK97].

The shape of the synthesis for the Grushin plane starting from a Riemannian point was drawn in [ABS08, BL13]. However, we present here for the first time computations in full detail. The optimal synthesis for $SU(2)$, $SO(3)$, $SL(2)$ were constructed in [BR08] but using a different technique. These optimal syntheses, together with the one for $SO_+(2,1)$, were also constructed in [BZ15a, BZ15b, BZ16] using the Gauss-Bonnet theorem. We follow this approach in Section 13.7.

The detailed analysis of geodesics for sub-Riemannian structure on $SE(2)$ was done by Yuri Sachkov in [MS10, Sac10, Sac11] that also proved Theorem 13.51 in full details.

The optimal synthesis for the Martinet flat sub-Riemannian structure was constructed in [ABCK97]. In the same paper one can also find the proof of Theorem 13.56. See also [BC03].

# Chapter 14

# Curves in the Lagrange Grassmannian

In this chapter we introduce the manifold of Lagrangian subspaces of a symplectic vector space. After a description of its geometric properties, we discuss how to define the curvature for regular curves in the Lagrange Grassmannian, that are curves with non-degenerate derivative. Then we discuss the non-regular case, where a reduction procedure let us to reduce to a regular curve in a reduced symplectic space.

The language developed in this chapter will be fundamental to encode in a single object, a curve in a space of Lagrangian subspaces, all information concerning Jacobi fields along sub-Riemannian geodesics, such as conjugate points and curvature, cf. Chapter 15. Indeed, in this chapter we introduce an "abstract" notion of conjugate point and curvature associated with a curve in the Lagrange Grassmannian and we will show that these notion recovers the classical ones when one considers curves of Jacobi fields associated with Riemannian geodesics.

## 14.1 The geometry of the Lagrange Grassmannian

In this section we recall some basic facts about Grassmannians of $k$-dimensional subspaces of an $n$-dimensional vector space. We then consider, for a vector space endowed with a symplectic structure, the submanifold of its Lagrangian subspaces.

**Definition 14.1.** Let $V$ be an $n$-dimensional vector space. The *Grassmannian of $k$-planes on $V$* is the set

$$G_k(V) := \{W \mid W \subset V \text{ is a subspace, } \dim(W) = k\}.$$

It is a standard fact that $G_k(V)$ is a compact manifold of dimension $k(n-k)$.

To fix a set of local coordinates on the Grassmannian $G_k(V)$, one proceeds as follow. Fix a subspace $Z \subset V$ such that $\dim Z = n-k$ and consider the set of all $k$-dimensional subspaces that are transversal to $Z$

$$Z^{\pitchfork} = \{W \in G_k(V) \mid W \cap Z = 0\}.$$

Clearly $Z^{\pitchfork}$ is an open subset of $G_k(V)$ and $G_k(V)$ is covered by such open subsets. To introduce coordinates on $Z^{\pitchfork}$ it is then sufficient to identify every $k$-dimensional subspace $U \subset V$ which is transversal to $Z$ with the graph of a linear map $A_U : W \to Z$. Once a basis on $W$ and $Z$ is chosen, this permits to identify $Z^{\pitchfork}$ with the space of $(n-k) \times k$ matrices.

Next we describe the tangent space to this manifold.

**Proposition 14.2.** *Let $W \in G_k(V)$. There exists a canonical isomorphism*

$$T_W G_k(V) \simeq \mathrm{Hom}(W, V/W).$$

*Proof.* Consider a smooth curve on $G_k(V)$ which starts from $W$, i.e., a smooth family of $k$-dimensional subspaces defined by a moving frame

$$W(t) = \mathrm{span}\{e_1(t), \dots, e_k(t)\}, \qquad W(0) = W.$$

We want to associate in a canonical way with the tangent vector $\dot{W}(0)$ a linear operator from $W$ to the quotient $V/W$. Fix $w \in W$ and consider any smooth extension $w(t) \in W(t)$, with $w(0) = w$. Then define the map

$$W \to V/W, \qquad w \mapsto \dot{w}(0) \ (\mathrm{mod}\ W). \tag{14.1}$$

We are left to prove that the map (14.1) is well-defined, i.e., independent on the choices of representatives. Indeed if we consider another extension $w_1(t)$ of $w$ satisfying $w_1(t) \in W(t)$ we can write

$$w_1(t) = w(t) + \sum_{i=1}^{k} \alpha_i(t) e_i(t),$$

for some smooth coefficients $\alpha_i(t)$ such that $\alpha_i(0) = 0$ for every $i$. It follows that

$$\dot{w}_1(t) = \dot{w}(t) + \sum_{i=1}^{k} \dot{\alpha}_i(t) e_i(t) + \sum_{i=1}^{k} \alpha_i(t) \dot{e}_i(t), \tag{14.2}$$

and evaluating (14.2) at $t = 0$ one has

$$\dot{w}_1(0) = \dot{w}(0) + \sum_{i=1}^{k} \dot{\alpha}_i(0) e_i(0).$$

This shows that $\dot{w}_1(0) = \dot{w}(0)\ (\mathrm{mod}\ W)$, hence the map (14.1) is well-defined. Similarly one can prove that the map does not depend on the moving frame defining $W(t)$.

Finally, it is easy to show that the map that associates the tangent vector to the curve $W(t)$ with the linear operator $W \to V/W$ is surjective, hence it is an isomorphism since the two spaces have the same dimension. $\square$

Let us now consider a symplectic vector space $(\Sigma, \sigma)$, i.e., a $2n$-dimensional vector space $\Sigma$ endowed with a non-degenerate symplectic form $\sigma \in \Lambda^2(\Sigma)$.

**Definition 14.3.** A vector subspace $\Pi \subset \Sigma$ of a symplectic space is called

(i) *symplectic* if $\sigma|_\Pi$ is nondegenerate,

(ii) *isotropic* if $\sigma|_\Pi \equiv 0$,

(iii) *Lagrangian* if $\sigma|_\Pi \equiv 0$ and $\dim \Pi = n$.

Notice that in general for every subspace $\Pi \subset \Sigma$, by nondegeneracy of the symplectic form $\sigma$, one has

$$\dim \Pi + \dim \Pi^\angle = \dim \Sigma, \tag{14.3}$$

where as usual we denote the symplectic orthogonal by $\Pi^\angle = \{z \in \Sigma \mid \sigma(z, w) = 0,\ \forall\, w \in \Pi\}$.

**Exercise 14.4.** Prove the following properties for a vector subspace $\Pi$ of a symplectic vector space $\Sigma$

(i) $\Pi$ is symplectic if and only if $\Pi \cap \Pi^{\angle} = \{0\}$,

(ii) $\Pi$ is isotropic if and only if $\Pi \subset \Pi^{\angle}$,

(iii) $\Pi$ is Lagrangian if and only if $\Pi = \Pi^{\angle}$.

**Exercise 14.5.** Let $(\Sigma, \sigma)$ be a symplectic space and $A, B \subset \Sigma$ be two subspaces. Prove the following identities: $(A+B)^{\angle} = A^{\angle} \cap B^{\angle}$ and $(A \cap B)^{\angle} = A^{\angle} + B^{\angle}$.

**Example 14.6.** Any symplectic vector space $\Sigma$ admits Lagrangian subspaces. Indeed let $\dim \Sigma = 2n$ and fix any non-zero element $e \neq 0$ in $\Sigma$. Let $e_1 := e$ and choose iteratively for $i = 2, \ldots, n$,

$$e_i \in \operatorname{span}\{e_1, \ldots, e_{i-1}\}^{\angle} \setminus \operatorname{span}\{e_1, \ldots, e_{i-1}\}. \tag{14.4}$$

Notice that the existence of an element satisfying (14.4) is possible by (14.3). Then the subspace $\Pi := \operatorname{span}\{e_1, \ldots, e_n\}$ is a Lagrangian subspace by construction.

**Lemma 14.7.** *Let $\Pi = \operatorname{span}\{e_1, \ldots, e_n\}$ be a Lagrangian subspace of $\Sigma$. Then there exists vectors $f_1, \ldots, f_n \in \Sigma$ such that*

*(i)* $\Sigma = \Pi \oplus \Delta, \quad$ *where* $\quad \Delta := \operatorname{span}\{f_1, \ldots, f_n\}$,

*(ii)* $\sigma(e_i, f_j) = \delta_{ij}, \quad \sigma(e_i, e_j) = \sigma(f_i, f_j) = 0, \qquad \forall\, i, j = 1, \ldots, n.$

*Proof.* We prove the statement by induction. By nondegeneracy of $\sigma$ there exists a non-zero $z \in \Sigma$ such that $\sigma(e_n, z) \neq 0$. Then the vector $f_n := z/\sigma(e_n, z)$ satisfies $\sigma(e_n, f_n) = 1$. This implies that $\sigma$ restricted to $\operatorname{span}\{e_n, f_n\}$ is nondegerate, hence by claim (i) of Exercise 14.4 one has

$$\operatorname{span}\{e_n, f_n\} \cap \operatorname{span}\{e_n, f_n\}^{\angle} = 0, \tag{14.5}$$

We can apply the induction step to the $2(n-1)$ subspace $\Sigma' := \operatorname{span}\{e_n, f_n\}^{\angle}$. Notice that (14.5) implies that $\sigma$ is non-degenerate also on $\Sigma'$. ☐

*Remark* 14.8. The complementary subspace $\Delta = \operatorname{span}\{f_1, \ldots, f_n\}$ defined in Lemma 14.7 is Lagrangian and transversal to $\Pi$, i.e., it holds

$$\Sigma = \Pi \oplus \Delta.$$

Considering coordinates induced from the basis chosen for this splitting, we can write $\Sigma = \mathbb{R}^n \oplus \mathbb{R}^n$ and any element of $\Sigma$ is written as $z = (p, x)$ where

$$z = \sum_{i=1}^{n} p_i e_i + x_i f_i, \qquad x = \big(x_1, \ldots, x_n\big)^T, \quad p = \big(p_1, \ldots, p_n\big)^T,$$

Using the canonical form of $\sigma$ on this basis (cf. Lemma 14.7) we have that in coordinates, if $z_1 = (p_1, x_1), z_2 = (p_2, x_2)$ the symplectic product is expressed as

$$\sigma(z_1, z_2) = p_1^T x_2 - p_2^T x_1. \tag{14.6}$$

*Remark* 14.9. A basis of a symplectic vector space $\Sigma$ satisfying conditions satisfying conditions (ii) of Lemma 14.7 is called a Darboux basis. The corresponding coordinates, in which the symplectic product takes the form (14.6), are called Darboux coordinates.

Lemma 14.7 can be interpreted geometrically as follows: the group of symplectomorphisms acts transitively on pairs of transversal Lagrangian subspaces. The next exercise, whose proof is an adaptation of the previous one, generalizes the previous property to the action of the group of symplectomorphisms on arbitrary pairs of subspaces. It shows that the unique invariant under this action is the dimension of their intersection.

**Exercise 14.10.** Let $\Lambda_1, \Lambda_2$ be two subspaces in a symplectic vector space $\Sigma$, and assume that $\dim \Lambda_1 \cap \Lambda_2 = k$. Show that there exists Darboux coordinates $(p, x)$ in $\Sigma$ such that

$$\Lambda_1 = \{(p, 0)\}, \qquad \Lambda_2 = \{((p_1, \ldots, p_k, 0, \ldots, 0), (0, \ldots, 0, x_{k+1}, \ldots, x_n)\}.$$

### 14.1.1 The Lagrange Grassmannian

**Definition 14.11.** The *Lagrange Grassmannian* $L(\Sigma)$ on a symplectic vector space $\Sigma$ is the set of its $n$-dimensional Lagrangian subspaces.

**Proposition 14.12.** *$L(\Sigma)$ is a compact submanifold of the Grassmannian $G_n(\Sigma)$. Moreover*

$$\dim L(\Sigma) = \frac{n(n+1)}{2}. \tag{14.7}$$

*Proof.* Recall that $G_n(\Sigma)$ is a $n^2$-dimensional compact manifold. Clearly $L(\Sigma) \subset G_n(\Sigma)$ as a subset. Consider the set of all Lagrangian subspaces that are transversal to a given one

$$\Delta^{\pitchfork} = \{\Lambda \in L(\Sigma) : \Lambda \cap \Delta = 0\}.$$

We have that $\Delta^{\pitchfork} \subset L(\Sigma)$ is an open subset. Since by Lemma 14.7 every Lagrangian subspace admits a Lagrangian complement

$$L(\Sigma) = \bigcup_{\Delta \in L(\Sigma)} \Delta^{\pitchfork}.$$

It is then sufficient to find some coordinates on these open subsets. The construction is similar to what one does for the general Grassmannian (see the discussion after Definition 14.1). Every $n$-dimensional subspace $\Lambda \subset \Sigma$ which is transversal to $\Delta$ is the graph of a linear map from $\Pi$ to $\Delta$. More precisely there exists a matrix $S_\Lambda$ such that

$$\Lambda \cap \Delta = 0 \Leftrightarrow \Lambda = \{(p^T, S_\Lambda p), p \in \mathbb{R}^n\}.$$

(Here we used the coordinates induced by the splitting $\Sigma = \Pi \oplus \Delta$.) Moreover it is easily seen that

$$\Lambda \in L(\Sigma) \Leftrightarrow S_\Lambda = (S_\Lambda)^T.$$

Indeed we have that $\Lambda \in L(\Sigma)$ if and only if $\sigma|_\Lambda = 0$ and using (14.6) this is rewritten as

$$\sigma((p_1^T, S_\Lambda p_1), (p_2^T, S_\Lambda p_2)) = p_1^T S_\Lambda p_2 - p_2^T S_\Lambda p_1 = 0,$$

which means exactly that $S_\Lambda$ is symmetric. Hence the open set of all subspaces that are transversal to $\Lambda$ is parametrized by the set of symmetric matrices, that gives coordinates in this open set. This also proves that the dimension of $L(\Sigma)$ coincides with the dimension of the space of symmetric matrices, from which (14.7) follows. Notice also that, being $L(\Sigma)$ a closed set in a compact manifold, it is compact. $\square$

Now we describe the tangent space to the Lagrange Grassmannian.

**Proposition 14.13.** *Let $\Lambda \in L(\Sigma)$. There exists a canonical isomorphism*

$$T_\Lambda L(\Sigma) \simeq Q(\Lambda),$$

*where $Q(\Lambda)$ denote the space of quadratic forms on $\Lambda$.*

*Proof.* Consider a smooth curve $\Lambda(t)$ in $L(\Sigma)$ such that $\Lambda(0) = \Lambda$ and denote by $\dot{\Lambda}(0) \in T_\Lambda L(\Sigma)$ its tangent vector. Consider a point $z \in \Lambda$ and fix a smooth extension $z(t) \in \Lambda(t)$ and denote with $\dot{z} := \dot{z}(0)$. We define the map

$$\underline{\dot{\Lambda}} : z \mapsto \sigma(z, \dot{z}), \tag{14.8}$$

We show that in coordinates $\underline{\dot{\Lambda}}$ is a well-defined quadratic map, independent on the extensions considered. Indeed

$$\Lambda(t) = \{(p^T, S_{\Lambda(t)}p), p \in \mathbb{R}^n\},$$

and the curve $z(t)$ can be written

$$z(t) = (p(t)^T, S_{\Lambda(t)}p(t)), \qquad z = z(0) = (p^T, S_\Lambda p),$$

for some curve $p(t)$ where $p = p(0)$. Differentiating the last identity we get

$$\dot{z}(t) = (\dot{p}(t)^T, \dot{S}_{\Lambda(t)}p(t) + S_{\Lambda(t)}\dot{p}(t)),$$

and evaluating at $t = 0$ (we simply omit $t$ when we evaluate at $t = 0$) we have

$$z = (p^T, S_\Lambda p), \qquad \dot{z} = (\dot{p}^T, \dot{S}_\Lambda p + S_\Lambda \dot{p}),$$

and finally get, using the simmetry of $S_\Lambda$, that

$$\begin{aligned} \sigma(z, \dot{z}) &= p^T(\dot{S}_\Lambda p + S_\Lambda \dot{p}) - \dot{p}^T S_\Lambda p \\ &= p^T \dot{S}_\Lambda p + p^T S_\Lambda \dot{p} - \dot{p}^T S_\Lambda p \\ &= p^T \dot{S}_\Lambda p. \end{aligned} \tag{14.9}$$

□

**Exercise 14.14.** Let $\Lambda(t) \in L(\Sigma)$ be such that $\Lambda = \Lambda(0)$ and $\sigma$ be the symplectic form. Prove that the map $B : \Lambda \times \Lambda \to \mathbb{R}$ defined by $B(z, w) = \sigma(z, \dot{w})$, where $\dot{w} = \dot{w}(0)$ is the tangent vector to a smooth extension $w(t) \in \Lambda(t)$ of $w$, is a symmetric bilinear map.

*Remark* 14.15. We have the following natural interpretation of Proposition 14.13: since $L(\Sigma)$ is a submanifold of the Grassmannian $G_n(\Sigma)$, its tangent space $T_\Lambda L(\Sigma)$ is naturally identified by the inclusion with a subspace of the Grassmannian

$$i : L(\Sigma) \hookrightarrow G_n(\Sigma), \qquad i_* : T_\Lambda L(\Sigma) \hookrightarrow T_\Lambda G_n(\Sigma) \simeq \mathrm{Hom}(\Lambda, \Sigma/\Lambda),$$

where the last isomorphism is given by Proposition 14.2. Being $\Lambda$ a Lagrangian subspace of $\Sigma$, the symplectic structure identifies in a canonical way the factor space $\Sigma/\Lambda$ with the dual space $\Lambda^*$ as follows

$$\Sigma/\Lambda \simeq \Lambda^*, \qquad \langle [z]_\Lambda, w\rangle = \sigma(z, w). \tag{14.10}$$

Hence the tangent space to the Lagrange Grassmannian consists of those linear maps in the space $\mathrm{Hom}(\Lambda, \Lambda^*)$ that are self-adjoint, which are naturally identified with quadratic forms on $\Lambda$ itself.[1]

[1] any quadratic form on a vector space $q \in Q(V)$ can be identified with a self-adjoint linear map $L : V \to V^*$, $L(v) = B(v, \cdot)$ where $B$ is the symmetric bilinear map such that $q(v) = B(v, v)$.

*Remark* 14.16. Given a curve $\Lambda(t)$ in $L(\Sigma)$, the above procedure associates to the tangent vector $\dot{\Lambda}(t)$ a family of quadratic forms $\underline{\dot{\Lambda}}(t)$, for every $t$.

We end this section by computing the tangent vector to a special class of curves that will play a major role in the sequel. More precisely the curve on $L(\Sigma)$ induced by the action on $\Lambda$ by the flow of the linear Hamiltonian vector field $\vec{h}$ associated with a quadratic form $h \in Q(\Sigma)$.

**Proposition 14.17.** *Let $\Lambda \in L(\Sigma)$ and $h \in Q(\Sigma)$. Define $\Lambda(t) = e^{t\vec{h}}(\Lambda)$. Then $\underline{\dot{\Lambda}} = 2h|_{\Lambda}$.*

*Proof.* Notice that since $h$ is a quadratic form on $\Sigma$ then $\vec{h}$ is a linear Hamiltonian vector field. This implies that $\Lambda(t)$ is a Lagrangian subspace, for every $t$. Consider $z \in \Lambda$ and the smooth extension $z(t) = e^{t\vec{h}}(z)$. Then $\dot{z}(t) = \vec{h}(z(t))$ and by definition of Hamiltonian vector field we find (we omit $t$ when evaluating at $t = 0$)

$$\begin{aligned}\sigma(z, \dot{z}) &= \sigma(z, \vec{h}(z)) \\ &= \langle d_z h, z\rangle \\ &= 2h(z),\end{aligned}$$

where in the last equality we used that $h$ is a quadratic form. □

## 14.2 Regular curves in Lagrange Grassmannian

The isomorphism between tangent vector to the Lagrange Grassmannian with quadratic forms makes sense to the following definition (we denote by $\underline{\dot{\Lambda}}$ the tangent vector to the curve at the point $\Lambda$ as a quadratic map)

**Definition 14.18.** Let $\Lambda(t) \in L(\Sigma)$ be a smooth curve in the Lagrange Grassmannian. We say that the curve is

(i) *monotone* increasing (resp. descreasing) if $\underline{\dot{\Lambda}}(t) \geq 0$ ($\underline{\dot{\Lambda}}(t) \leq 0$).

(ii) *strictly monotone* increasing (resp. decreasing) if the inequality in (i) is strict.

(iii) *regular* if its derivative $\underline{\dot{\Lambda}}(t)$ is a non-degenerate quadratic form.

*Remark* 14.19. Notice that if $\Lambda(t) = \{(p, S(t)p), p \in \mathbb{R}^n\}$ in some coordinate set, then it follows from the proof of Proposition 14.13 that the quadratic form $\underline{\dot{\Lambda}}(t)$ is represented by the matrix $\dot{S}_{\Lambda}(t)$ (see also (14.9)). In particular the curve is regular if and only if $\det \dot{S}_{\Lambda}(t) \neq 0$.

The main goal of this section is the construction of a canonical Lagrangian complement for a regular curve. More precisely we want to associate with a regular curve another curve $\Lambda^{\circ}(t)$ in the Lagrange Grassmannian such that $\Sigma = \Lambda(t) \oplus \Lambda^{\circ}(t)$.

Consider an arbitrary Lagrangian splitting $\Sigma = \Lambda(0) \oplus \Delta$ defined by a complement $\Delta$ to $\Lambda(0)$ (see Lemma 14.7) and fix coordinates in such a way that

$$\Sigma = \{(p, x),\ p, x \in \mathbb{R}^n\}, \qquad \Lambda(0) = \{(p, 0),\ p \in \mathbb{R}^n\}, \qquad \Delta = \{(0, x),\ x \in \mathbb{R}^n\}.$$

In these coordinates our regular curve is described by a one-parametric family of symmetric matrices $S(t)$

$$\Lambda(t) = \{(p, S(t)p),\ p \in \mathbb{R}^n\},$$

such that $S(0) = 0$ and $\dot{S}(0)$ is invertible. All Lagrangian complements to $\Lambda(0)$ are parametrized by a symmetrix matrix $B$ as follows

$$\Delta_B = \{(Bx, x), x \in \mathbb{R}^n\}, \qquad B = B^T.$$

The following lemma shows how the coordinate expression of our curve $\Lambda(t)$ changes in the new coordinate set defined by the splitting $\Sigma = \Lambda(0) \oplus \Delta_B$.

**Lemma 14.20.** *Let $S_B(t)$ be the one-parametric family of symmetric matrices defining $\Lambda(t)$ in coordinates with respect to the splitting $\Lambda(0) \oplus \Delta_B$. Then the following identity holds*

$$S_B(t) = (S(t)^{-1} - B)^{-1}. \tag{14.11}$$

*Proof.* It is easy to show that, if $(p, x)$ and $(p', x')$ denotes coordinates with respect to the splitting defined by the subspaces $\Delta$ and $\Delta_B$ we have

$$\begin{cases} p' = p - Bx \\ x' = x \end{cases} \tag{14.12}$$

The matrix $S_B(t)$ by definition is the matrix that satisfies the identity $x' = S_B(t)p'$. Using that $x = S(t)p$ by definition of $\Lambda(t)$, from (14.12) we find

$$x' = x = S(t)p = S(t)(p' + Bx'),$$

and with straightforward computations we finally get

$$S_B(t) = (I - S(t)B)^{-1}S(t) = (S(t)^{-1} - B)^{-1}.$$

□

From the previous lemma it follows that $\dot{S}_B(0) = \dot{S}(0)$, for every choice of $B$. Hence it is natural to look for a change of coordinates (i.e., a choice of the matrix $B$) that simplifies the second derivative of our curve.

**Corollary 14.21.** *There exists a unique symmetric matrix $B$ such that $\ddot{S}_B(0) = 0$.*

*Proof.* Recall that for a one-parametric family of matrices $X(t)$ we have

$$\frac{d}{dt}X(t)^{-1} = -X(t)^{-1}\dot{X}(t)X(t)^{-1}.$$

Applying twice this identity to (14.11) (we omit $t$ to denote the value at $t = 0$) we get

$$\begin{aligned} \left.\frac{d}{dt}\right|_{t=0} S_B(t) &= -(S^{-1} - B)^{-1}\left(\left.\frac{d}{dt}\right|_{t=0} S^{-1}(t)\right)(S^{-1} - B)^{-1} \\ &= (S^{-1} - B)^{-1}S^{-1}\dot{S}S^{-1}(S^{-1} - B)^{-1} \\ &= (I - SB)^{-1}\dot{S}(I - BS)^{-1}. \end{aligned}$$

Hence for the second derivative of $S_B(t)$ evaluated at $t = 0$ (remember that in our coordinates $S(0) = 0$), one gets

$$\ddot{S}_B = \ddot{S} + 2\dot{S}B\dot{S},$$

and using that $\dot{S}$ is non-degenerate, we can choose $B = -\frac{1}{2}\dot{S}^{-1}\ddot{S}\dot{S}^{-1}$. □

We set $\Lambda^\circ(0) := \Delta_B$, where $B$ is determined by Corollary 14.21. Notice that by construction $\Lambda^\circ(0)$ is a Lagrangian subspace and it is transversal to $\Lambda(0)$.

Then one can define $\Lambda^\circ(t)$ for every $t$ as follows: $\Lambda^\circ(t) := \Lambda_t^\circ(0)$ where the curve $\Lambda_t$ is defined by $\Lambda_t(s) = \Lambda(s+t)$.

**Definition 14.22.** Let $\Lambda(t)$ be a regular curve, the curve $\Lambda^\circ(t)$ defined by the condition above is called *derivative curve* of $\Lambda(t)$.

**Exercise 14.23.** Prove that, if $\Lambda(t) = \{(p, S(t)p),\, p \in \mathbb{R}^n\}$ (without the condition $S(0) = 0$), then the derivative curve $\Lambda^\circ(t) = \{(p, S^\circ(t)p),\, p \in \mathbb{R}^n\}$, satisfies

$$S^\circ(t) = B(t)^{-1} + S(t), \quad \text{where} \quad B(t) := -\frac{1}{2}\dot{S}(t)^{-1}\ddot{S}(t)\dot{S}(t)^{-1}, \tag{14.13}$$

provided $\Lambda^\circ(t)$ is transversal to the subspace $\Delta = \{(0,x), x \in \mathbb{R}^n\}$. (This condition is equivalent to the invertibility of $B(t)$.) Notice that if $S(0) = 0$ then $S^\circ(0) = B(0)^{-1}$.

*Remark* 14.24. The set $\Lambda^{tr}$ of all $n$-dimensional spaces transversal to a fixed subspace $\Lambda$ is an affine space over $\mathrm{Hom}(\Sigma/\Lambda, \Lambda)$. Indeed given two elements $\Delta_1, \Delta_2 \in \Lambda^{tr}$ we can associate with their difference the operator

$$\Delta_2 - \Delta_1 \mapsto A \in \mathrm{Hom}(\Sigma/\Lambda, \Lambda), \qquad A([z]_\Lambda) = z_2 - z_1 \in \Lambda, \tag{14.14}$$

where $z_i \in \Delta_i \cap [z]_\Lambda$ are uniquely identified.

If $\Lambda$ is Lagrangian, we have identification $\Sigma/\Lambda \simeq \Lambda^*$ given by the symplectic structure (see (14.10)) that $\Lambda^\pitchfork$, that coincide by definition with the intersection $\Lambda^{tr} \cap L(\Sigma)$ is an affine space over $\mathrm{Hom}^S(\Lambda^*, \Lambda)$, the space of selfadjoint maps between $\Lambda^*$ and $\Lambda$, that it isomorphic to $Q(\Lambda^*)$.

Notice that if we fix a distinguished complement of $\Lambda$, i.e., $\Sigma = \Lambda \oplus \Delta$, then we have also the identification $\Sigma/\Lambda \simeq \Delta$ and $\Lambda^\pitchfork \simeq Q(\Lambda^*) \simeq Q(\Delta)$.

**Exercise 14.25.** Prove that the operator $A$ defined by (14.14), in the case when $\Lambda$ is Lagrangian, is a self-adjoint operator.

*Remark* 14.26. Assume that the splitting $\Sigma = \Lambda \oplus \Delta$ is fixed. Then our curve $\Lambda(t)$ in $L(\Sigma)$, such that $\Lambda(0) = \Lambda$, is characterized by a family of symmetric matrices $S(t)$ satisfying $\Lambda(t) = \{(p, S(t)p), p \in \mathbb{R}^n\}$, with $S(0) = 0$.

By regularity of the curve, $\Lambda(t) \in \Lambda^\pitchfork$ for $t > 0$ small enough, hence we can consider its coordinate presentation in the affine space on the vector space of quadratic forms defined on $\Delta$ (see Remark 14.24) that is given by $S^{-1}(t)$.

The curve does not belong to the coordinate chart for $t = 0$ and is regular if and only if the coordinate presentation has a simple pole at $t = 0$. In this case we can write the Laurent expansion of this curve in the affine space

$$\begin{aligned} S(t)^{-1} &= \left(t\dot{S} + \frac{t^2}{2}\ddot{S} + O(t^3)\right)^{-1} \\ &= \frac{1}{t}\dot{S}^{-1}\left(I + \frac{t}{2}\ddot{S}\dot{S}^{-1} + O(t^2)\right)^{-1} \\ &= \frac{1}{t}\dot{S}^{-1} \underbrace{-\frac{1}{2}\dot{S}^{-1}\ddot{S}\dot{S}^{-1}}_{B} + O(t). \end{aligned}$$

In this expansion every term but the zero-order one (with respect to $t$) is an element of the vector space, while the zero-order term is a point, i.e., an element of the affine space. Then $\Lambda^\circ(0)$ is the free term in the Laurent expansion of $\Lambda(t)$ in the chart $\Lambda^\pitchfork$

It is then not occasional that the matrix $B$ coincides with the free term of this expansion. Indeed the formula (14.11) for the change of coordinates can be rewritten as follows

$$S_B(t)^{-1} = S^{-1}(t) - B, \tag{14.15}$$

and the choice of $B$ corresponds exactly to the choice of a coordinate set where the curve $\Lambda(t)$ has no free term in this expansion. Equivalently, a regular curve permits us to choose a privileged origin in the affine space of Lagrangian subspaces that are transversal to the curve itself.

## 14.3 Curvature of a regular curve

Now we want to define the curvature of a regular curve in the Lagrange Grassmannian. Let $\Lambda(t)$ be a regular curve and consider its derivative curve $\Lambda^\circ(t)$.

The tangent vectors to $\Lambda(t)$ and $\Lambda^\circ(t)$, as explained in Section 14.1, can be interpreted in a a canonical way as a quadratic form on the space $\Lambda(t)$ and $\Lambda^\circ(t)$, respectively

$$\underline{\dot{\Lambda}}(t) \in Q(\Lambda(t)), \qquad \underline{\dot{\Lambda}}^\circ(t) \in Q(\Lambda^\circ(t)).$$

Being $\Lambda^\circ(t)$ a canonical Lagrangian complement to $\Lambda(t)$ we have the identifications through the symplectic form[2]

$$\Lambda(t)^* \simeq \Lambda^\circ(t), \qquad \Lambda^\circ(t)^* \simeq \Lambda(t),$$

and the quadratic forms $\underline{\dot{\Lambda}}(t), \underline{\dot{\Lambda}}^\circ(t)$ can be treated as (self-adjoint) mappings:

$$\underline{\dot{\Lambda}}(t) : \Lambda(t) \to \Lambda^\circ(t), \qquad \underline{\dot{\Lambda}}^\circ(t) : \Lambda^\circ(t) \to \Lambda(t). \tag{14.16}$$

**Definition 14.27.** The operator $R_\Lambda(t) := \underline{\dot{\Lambda}}^\circ(t) \circ \underline{\dot{\Lambda}}(t) : \Lambda(t) \to \Lambda(t)$ is called the *curvature* operator of the regular curve $\Lambda(t)$.

*Remark* 14.28. In the monotonic increasing (resp. decreasing) case, when $\dot{\Lambda}(t)$ (resp. $-\dot{\Lambda}(t)$) defines a scalar product on $\Lambda(t)$, the operator $R(t)$ is, by definition, symmetric with respect to this scalar product. Moreover $R(t)$, as quadratic form, has the same signature and rank as $\underline{\dot{\Lambda}}^\circ(t)$ (resp. $-\underline{\dot{\Lambda}}^\circ(t)$).

**Definition 14.29.** Let $\Lambda_1, \Lambda_2$ be two transversal Lagrangian subspaces of $\Sigma$. We denote by

$$\pi_{\Lambda_1,\Lambda_2} : \Sigma \to \Lambda_2, \tag{14.17}$$

the projection on $\Lambda_2$ parallel to $\Lambda_1$, i.e., the linear operator such that

$$\pi_{\Lambda_1,\Lambda_2}|_{\Lambda_1} = 0 \qquad \pi_{\Lambda_1,\Lambda_2}|_{\Lambda_2} = \mathrm{Id}.$$

**Exercise 14.30.** Let $\Lambda_1$ and $\Lambda_2$ be two Lagrangian subspaces in $\Sigma$ and assume that, in some coordinate set, $\Lambda_i = \{(p, S_i p), p \in \mathbb{R}^n\}$ for $i = 1, 2$. Prove that $\Sigma = \Lambda_1 \oplus \Lambda_2$ if and only if $\ker(S_1 - S_2) = \{0\}$. In this case show that the following matrix expression for $\pi_{\Lambda_1,\Lambda_2}$:

$$\pi_{\Lambda_1,\Lambda_2} = \begin{pmatrix} S_{12}^{-1} S_1 & -S_{12}^{-1} \\ S_2 S_{12}^{-1} S_1 & -S_2 S_{12}^{-1} \end{pmatrix}, \qquad S_{12} := S_1 - S_2. \tag{14.18}$$

[2] if $\Sigma = \Lambda \oplus \Delta$ is a splitting of a vector space then $\Sigma/\Lambda \simeq \Delta$. If moreover the splitting is Lagrangian in a symplectic space, the symplectic form identifies $\Sigma/\Lambda \simeq \Lambda^*$, hence $\Lambda^* \simeq \Delta$.

From the definition of the derivative of our curve we can get the following geometric characterization of the curvature of a curve.

**Proposition 14.31.** *Let $\Lambda(t)$ a regular curve in $L(\Sigma)$ and $\Lambda^\circ(t)$ its derivative curve. Then*

$$\underline{\dot{\Lambda}}(t)(z_t) = \pi_{\Lambda(t),\Lambda^\circ(t)}(\dot{z}_t), \qquad \underline{\dot{\Lambda}}^\circ(t)(z_t) = -\pi_{\Lambda^\circ(t),\Lambda(t)}(\dot{z}_t).$$

*In particular the curvature is the composition $R_\Lambda(t) = -\pi_{\Lambda^\circ(t),\Lambda(t)} \circ \pi_{\Lambda(t),\Lambda^\circ(t)}$.*

*Proof.* Recall that, by definition, the linear operator $\underline{\dot{\Lambda}} : \Lambda \to \Sigma/\Lambda$ associated with the quadratic form is the map $z \mapsto \dot{z} \pmod{\Lambda}$. Hence to build the map $\Lambda \to \Lambda^\circ$ it is enough to compute the projection of $\dot{z}$ onto the complement $\Lambda^\circ$, that is exactly $\pi_{\Lambda,\Lambda^\circ}(\dot{z})$. Notice that the minus sign in equation (14.31) is a consequence of the skew symmetry of the symplectic product. More precisely, the sign in the identification $\Lambda^\circ \simeq \Lambda^*$ depends on the position of the argument. □

The curvature $R_\Lambda(t)$ of the curve $\Lambda(t)$ is a kind of relative velocity between the two curves $\Lambda(t)$ and $\Lambda^\circ(t)$. If $\Sigma$ has dimension two then a curve in the Lagrange Grassmannian is a curve of lines, i.e., a curve on $S^1$. In this case $R_\Lambda(t) > 0$ means that the curve $\Lambda(t)$ and its derivative $\Lambda^\circ(t)$ moves in the same direction.

Next we compute the expression of the curvature $R_\Lambda(t)$ in coordinates.

**Proposition 14.32.** *Assume that $\Lambda(t) = \{(p, S(t)p)\}$ is a regular curve in $L(\Sigma)$. Then we have the following coordinate expression for the curvature of $\Lambda$ (we omit $t$ in the formula)*

$$R_\Lambda = ((2\dot{S})^{-1}\ddot{S})^{\cdot} - ((2\dot{S})^{-1}\ddot{S})^2 \tag{14.19}$$

$$= \frac{1}{2}\dot{S}^{-1}\dddot{S} - \frac{3}{4}(\dot{S}^{-1}\ddot{S})^2. \tag{14.20}$$

*Proof.* Assume that both $\Lambda(t)$ and $\Lambda^\circ(t)$ are contained in the same coordinate chart with

$$\Lambda(t) = \{(p, S(t)p), p \in \mathbb{R}^n\}, \qquad \Lambda^\circ(t) = \{(p, S^\circ(t)p), p \in \mathbb{R}^n\}.$$

We start the proof by computing the expression of the linear operator associated with the derivative $\dot{\Lambda} : \Lambda \to \Lambda^\circ$ (we omit $t$ when we compute at $t = 0$). For each element $(p, Sp) \in \Lambda$ and any smooth extension $(p(t), S(t)p(t))$ one can apply the matrix representing the operator $\pi_{\Lambda,\Lambda^\circ}$ (see (14.18)) to the derivative at $t = 0$ and find

$$\pi_{\Lambda,\Lambda^\circ}(\dot{p}, \dot{S}p + S\dot{p}) = (p', S^\circ p'), \qquad p' = -(S - S^\circ)^{-1}\dot{S}p.$$

Exchanging the role of $\Lambda$ and $\Lambda^\circ$, and taking into account of the minus sign one finds that the coordinate representation of $R_\Lambda$ is given by

$$R_\Lambda = (S^\circ - S)^{-1}\dot{S}^\circ(S^\circ - S)^{-1}\dot{S}. \tag{14.21}$$

We prove formula (14.20) under the extra assumption that $S(0) = 0$. Notice that this is equivalent to the choice of a particular coordinate set in $L(\Sigma)$ and, being the expression of $R_\Lambda$ coordinate independent by construction, this is not restrictive.

Under this extra assumption, it follows from (14.13) that

$$\Lambda(t) = \{(p, S(t)p), p \in \mathbb{R}^n\}, \qquad \Lambda^\circ(t) = \{(p, S^\circ(t)p), p \in \mathbb{R}^n\},$$

where $S^\circ(t) = B(t)^{-1} + S(t)$ and we denote by $B(t) := -\frac{1}{2}\dot{S}(t)^{-1}\ddot{S}(t)\dot{S}(t)^{-1}$.

Hence we have, assuming $S(0) = 0$ and omitting $t$ when $t = 0$

$$\begin{aligned} R_\Lambda &= (S^\circ - S)^{-1}\dot{S}^\circ(S^\circ - S)^{-1}\dot{S} \\ &= B\frac{d}{dt}\bigg|_{t=0}\left(B(t)^{-1} + S(t)\right)B\dot{S} \\ &= (B\dot{S})^2 - \dot{B}\dot{S}. \end{aligned}$$

Plugging $B = -\frac{1}{2}\dot{S}^{-1}\ddot{S}\dot{S}^{-1}$ into the last formula, after some computations one gets to (14.20). $\square$

*Remark* 14.33. Consider a regular curve $\Lambda(t)$ that is monotone increasing. The formula for the curvature $R_\Lambda(t)$ of $\Lambda(t)$ takes a very simple form in a particular coordinate set given by the splitting $\Sigma = \Lambda(0) \oplus \Lambda^\circ(0)$, i.e., such that

$$\Lambda(0) = \{(p, 0), p \in \mathbb{R}^n\}, \qquad \Lambda^\circ(0) = \{(0, x), x \in \mathbb{R}^n\}.$$

Indeed using a symplectic change of coordinates in $\Sigma$ that preserves both $\Lambda$ and $\Lambda^\circ$ (i.e., of the kind $p' = Ap$ and $x' = (A^{-1})^*x$) we can choose the matrix $A$ in such a way that $\dot{S}(0) = I$. Moreover we know from Corollary 14.21 that the fact that $\Lambda^\circ = \{(0, x), x \in \mathbb{R}^n\}$ is equivalent to $\ddot{S}(0) = 0$. Hence one finds from (14.20) that

$$R_\Lambda = \frac{1}{2}\dddot{S}. \tag{14.22}$$

The curvature $R_\Lambda(t)$ represents a well-defined operator on $\Lambda(0)$, naturally endowed with the sign definite quadratic form $\dot{\Lambda}(0)$. Hence in these coordinates the eigenvalues of $\dddot{S}$ (and not only the trace and the determinant) are invariants of the curve.

**Exercise 14.34.** Let $f : \mathbb{R} \to \mathbb{R}$ be a smooth function such that $f' \neq 0$. The *Schwarzian derivative* of $f$ is defined as

$$\mathcal{S}f := \left(\frac{f''}{2f'}\right)' - \left(\frac{f''}{2f'}\right)^2. \tag{14.23}$$

Prove that $\mathcal{S}f = 0$ if and only if $f(t) = \dfrac{at+b}{ct+d}$ for some $a, b, c, d \in \mathbb{R}$.

*Remark* 14.35. The previous proposition says that the curvature $R$ is the matrix version of the Schwarzian derivative of the matrix valued function $S$, cf. (14.19) and (14.23).

**Example 14.36.** Let $\Sigma$ be a 2-dimensional symplectic space. In this case $L(\Sigma) \simeq \mathbb{P}^1(\mathbb{R})$ is the real projective line. Let us compute the curvature of a curve in $L(\Sigma)$ with constant (angular) velocity $\alpha > 0$. We have

$$\Lambda(t) = \{(p, S(t)p), p \in \mathbb{R}\}, \qquad S(t) = \tan(\alpha t) \in \mathbb{R}.$$

From the explicit expression it easy to find the relation

$$\dot{S}(t) = \alpha(1 + S^2(t)), \quad \Rightarrow \quad \frac{\ddot{S}(t)}{2\dot{S}(t)} = \alpha S(t),$$

from which one gets that $R_\Lambda(t) = \alpha\dot{S}(t) - \alpha^2 S^2(t) = \alpha^2$, i.e., the curve has constant curvature.

We end this section with a useful formula on the curvature of a reparametrized curve.

**Proposition 14.37.** *Let $\varphi : \mathbb{R} \to \mathbb{R}$ a diffeomorphism and define the curve $\Lambda_\varphi(t) := \Lambda(\varphi(t))$. Then*

$$R_{\Lambda_\varphi}(t) = \dot{\varphi}^2(t) R_\Lambda(\varphi(t)) + R_\varphi(t)\mathrm{Id}. \tag{14.24}$$

*Proof.* It is a simple check that the Schwarzian derivative of the composition of two function $f$ and $g$ satisfies

$$\mathcal{S}(f \circ g) = (\mathcal{S}f \circ g)(g')^2 + \mathcal{S}g.$$

Notice that $R_\varphi(t)$ makes sense as the curvature of the regular curve $\varphi : \mathbb{R} \to \mathbb{R} \subset \mathbb{P}^1$ in the Lagrange Grassmannian $L(\mathbb{R}^2)$. □

**Exercise 14.38.** (Another formula for the curvature). Let $\Lambda_0, \Lambda_1 \in L(\Sigma)$ be such that $\Sigma = \Lambda_0 \oplus \Lambda_1$ and fix two tangent vectors $\xi_0 \in T_{\Lambda_0}L(\Sigma)$ and $\xi_1 \in T_{\Lambda_1}L(\Sigma)$. As in (14.16) we can treat each tangent vector as a linear operator

$$\xi_0 : \Lambda_0 \to \Lambda_1, \qquad \xi_1 : \Lambda_1 \to \Lambda_0, \tag{14.25}$$

and define the *cross-ratio* $[\xi_1, \xi_0] = -\xi_1 \circ \xi_0$. If in some coordinates $\Lambda_i = \{(p, S_i p)\}$ for $i = 0, 1$ we have (here $\dot{S}_i$ denotes the matrix associated with $\xi_i$)

$$[\xi_1, \xi_0] = (S_1 - S_0)^{-1}\dot{S}_1(S_1 - S_0)^{-1}\dot{S}_0.$$

Let now $\Lambda(t)$ a regular curve in $L(\Sigma)$. By regularity $\Sigma = \Lambda(0) \oplus \Lambda(t)$ for all $t > 0$ small enough, hence the cross ratio

$$[\underline{\dot{\Lambda}}(t), \underline{\dot{\Lambda}}(0)] : \Lambda(0) \to \Lambda(0),$$

is well-defined. Prove the following expansion for $t \to 0$

$$[\underline{\dot{\Lambda}}(t), \underline{\dot{\Lambda}}(0)] \simeq \frac{1}{t^2}\mathrm{Id} + \frac{1}{3}R_\Lambda(0) + O(t). \tag{14.26}$$

## 14.4 Reduction of non-regular curves in Lagrange Grassmannian

In this section we want to extend the notion of curvature to non-regular curves. As we will see in the next chapter, it is always possible to associate with an extremal a family of Lagrangian subspaces in a symplectic space, i.e., a curve in a Lagrangian Grassmannian. This curve turns out to be regular if and only if the extremal is an extremal of a Riemannian structure. Hence, if we want to apply this theory for a genuine sub-Riemannian case we need some tools to deal with non-regular curves in the Lagrangian Grassmannian.

Let $(\Sigma, \sigma)$ be a symplectic vector space and $L(\Sigma)$ denote the Lagrange Grassmannian. We start by describing a natural submanifold of $L(\Sigma)$ associated with an isotropic subspace $\Gamma$ of $\Sigma$. This will allow us to define a reduction procedure for a non-regula curve.

Let $\Gamma$ be a $k$-dimensional isotropic subspace of $\Sigma$, i.e., $\sigma\big|_\Gamma = 0$. This means that $\Gamma \subset \Gamma^\angle$. In particular $\Gamma^\angle/\Gamma$ is a $2(n-k)$ dimensional symplectic space, where the symplectic structure is defined by the restriction of $\sigma$.

**Lemma 14.39.** *There is a natural identification of $L(\Gamma^\angle/\Gamma)$ as a submanifold of $L(\Sigma)$:*

$$L(\Gamma^\angle/\Gamma) \simeq \{\Lambda \in L(\Sigma), \Gamma \subset \Lambda\} \subset L(\Sigma). \tag{14.27}$$

*Moroever we have a natural projection*

$$\pi^\Gamma : L(\Sigma) \to L(\Gamma^\angle/\Gamma), \qquad \Lambda \mapsto \Lambda^\Gamma,$$

*where $\Lambda^\Gamma := (\Lambda \cap \Gamma^\angle) + \Gamma = (\Lambda + \Gamma) \cap \Gamma^\angle$.*

*Proof.* Assume that $\Lambda \in L(\Sigma)$ and $\Gamma \subset \Lambda$. Then, since $\Lambda$ is Lagrangian, $\Lambda = \Lambda^\angle \subset \Gamma^\angle$, hence the identification (14.27).

Assume now that $\Lambda \in L(\Gamma^\angle/\Gamma)$ and let us show that $\pi^\Gamma(\Lambda) = \Lambda$, i.e., $\pi^\Gamma$ is a projection. Indeed from the inclusions $\Gamma \subset \Lambda \subset \Gamma^\angle$ one has $\pi^\Gamma(\Lambda) = \Lambda^\Gamma = (\Lambda \cap \Gamma^\angle) + \Gamma = \Lambda + \Gamma = \Lambda$.

We are left to check that $\Lambda^\Gamma$ is Lagrangian, i.e., $(\Lambda^\Gamma)^\angle = \Lambda^\Gamma$.

$$\begin{aligned}(\Lambda^\Gamma)^\angle &= ((\Lambda \cap \Gamma^\angle) + \Gamma)^\angle \\ &= (\Lambda \cap \Gamma^\angle)^\angle \cap \Gamma^\angle \\ &= (\Lambda + \Gamma) \cap \Gamma^\angle = \Lambda^\Gamma,\end{aligned}$$

where we repeatedly used Exercise 14.5. Notice that the last identity $(\Lambda + \Gamma) \cap \Gamma^\angle = (\Lambda \cap \Gamma^\angle) + \Gamma$ is also a consequence of the same exercise. $\square$

*Remark* 14.40. Let $\Gamma^\pitchfork = \{\Lambda \in L(\Sigma) \mid \Lambda \cap \Gamma = \{0\}\}$. The restriction $\pi^\Gamma\big|_{\Gamma^\pitchfork}$ is smooth. Indeed one can show that $\pi^\Gamma$ is defined by a rational function, since it is expressed via the solution of a linear system.

The following example shows that the projection $\pi^\Gamma$ is not even continous when considered globally on $L(\Sigma)$.

**Example 14.41.** Consider the symplectic structure $\sigma$ on $\mathbb{R}^4$ and fix a Darboux basis $\{e_1, e_2, f_1, f_2\}$, i.e., $\sigma(e_i, f_j) = \delta_{ij}$. Let $\Gamma = \operatorname{span}\{e_1\}$ be a one dimensional isotropic subspace and define for $\varepsilon > 0$

$$\Lambda_\varepsilon = \operatorname{span}\{e_1 + \varepsilon f_2, e_2 + \varepsilon f_1\}.$$

It is easy to see that $\Lambda_\varepsilon$ is Lagrangian for every $\varepsilon > 0$ and that

$$\begin{aligned}\Lambda_\varepsilon^\Gamma &= \operatorname{span}\{e_1, f_2\}, \qquad \forall\, \varepsilon > 0, \\ \Lambda_0^\Gamma &= \operatorname{span}\{e_1, e_2\}.\end{aligned} \tag{14.28}$$

The case $\varepsilon = 0$ is trivial. Let $\varepsilon > 0$. Then $f_2 \in e_1^\angle$, that implies $e_1 + \varepsilon f_2 \in \Lambda_\varepsilon \cap \Gamma^\angle$, therefore $f_2 \in \Lambda_\varepsilon \cap \Gamma^\angle$. By definition of reduced curve $f_2 \in \Lambda_\varepsilon^\Gamma$ and (14.28) holds.

## 14.5 Ample curves

In this section we introduce ample curves.

**Definition 14.42.** Let $\Lambda(t) \in L(\Sigma)$ be a smooth curve in the Lagrange Grassmannian. The curve $\Lambda(t)$ is *ample* at $t = t_0$ if there exists $N \in \mathbb{N}$ such that

$$\Sigma = \operatorname{span}\{\lambda^{(i)}(t_0) \mid \lambda(t) \in \Lambda(t), \lambda(t) \text{ smooth}, 0 \leq i \leq N\}. \tag{14.29}$$

In other words we require that all derivatives up to order $N$ of all smooth sections of our curve in $L(\Sigma)$ span all the possible directions.

As usual, we can choose coordinates in such a way that, for some family of symmetric matrices $S(t)$, one has

$$\Sigma = \{(p, x) \mid p, x \in \mathbb{R}^n\}, \qquad \Lambda(t) = \{(p, S(t)p) \mid p \in \mathbb{R}^n\}.$$

**Exercise 14.43.** Assume that $\Lambda(t) = \{(p, S(t)p), p \in \mathbb{R}^n\}$ with $S(0) = 0$. Prove that the curve is ample at $t = 0$ if and only if there exists $N \in \mathbb{N}$ such that all the columns of the derivative of $S(t)$ up to order $N$ (and computed at $t = 0$) span a maximal subspace:

$$\operatorname{rank}\{\dot{S}(0), \ddot{S}(0), \dots, S^{(N)}(0)\} = n. \tag{14.30}$$

In particular, a curve $\Lambda(t)$ is regular at $t_0$ if and only if is ample at $t_0$ with $N = 1$.

An important property of ample and monotone curves is described in the following lemma.

**Lemma 14.44.** *Let $\Lambda(t) \in L(\Sigma)$ a monotone, ample curve at $t_0$. Then, there exists $\varepsilon > 0$ such that $\Lambda(t) \cap \Lambda(t_0) = \{0\}$ for $0 < |t - t_0| < \varepsilon$.*

*Proof.* Without loss of generality, assume $t_0 = 0$. Choose a Lagrangian splitting $\Sigma = \Lambda \oplus \Pi$, with $\Lambda = J(0)$. For $|t| < \varepsilon$, the curve is contained in the chart defined by such a splitting. In coordinates, $\Lambda(t) = \{(p, S(t)p) | \ p \in \mathbb{R}^n\}$, with $S(t)$ symmetric and $S(0) = 0$. The curve is monotone, then $\dot{S}(t)$ is a semidefinite symmetric matrix. It follows that $S(t)$ is semidefinite too.

Suppose that, for some $t$, $\Lambda(t) \cap \Lambda(0) \neq \{0\}$ (assume $t > 0$). This means that $\exists\, v \in \mathbb{R}^n$ such that $S(t)v = 0$. Indeed also $v^*S(t)v = 0$. The function $\tau \mapsto v^*S(\tau)v$ is monotone, vanishing at $\tau = 0$ and $\tau = t$. Therefore $v^*S(\tau)v = 0$ for all $0 \leq \tau \leq t$. Being a semidefinite, symmetric matrix, $v^*S(\tau)v = 0$ if and only if $S(\tau)v = 0$. Therefore, we conclude that $v \in \ker S(\tau)$ for $0 \leq \tau \leq t$. This implies that, for any $i \in \mathbb{N}$, $v \in \ker S^{(i)}(0)$, which is a contradiction, since the curve is ample at 0. □

**Exercise 14.45.** Prove that a monotone curve $\Lambda(t)$ is ample at $t_0$ if and only if one of the following (equivalent) conditions is satisfied

(i) the family of matrices $S(t) - S(t_0)$ is nondegenerate for $t \neq t_0$ close enough, and the same remains true if we replace $S(t)$ by its $N$-th Taylor polynomial, for some $N$ in $\mathbb{N}$.

(ii) the map $t \mapsto \det(S(t) - S(t_0))$ has a finite order root at $t = t_0$.

Let us now consider a monotone curve on $L(\Sigma)$. Without loss of generality we can assume the curve to be non-decreasing, i.e., $\underline{\dot{\Lambda}}(t) \geq 0$. By monotonicity

$$\Lambda(0) \cap \Lambda(t) = \bigcap_{0 \leq \tau \leq t} \Lambda(\tau) =: \Upsilon_t$$

Clearly $\Upsilon_t$ is a decreasing family of subspaces, i.e., $\Upsilon_t \subseteq \Upsilon_\tau$ if $\tau \leq t$. Hence the family $\Upsilon_t$ for $t \to 0$ stabilizes and the limit subspace $\Upsilon$ is well-defined

$$\Upsilon := \lim_{t \to 0} \Upsilon_t.$$

The symplectic reduction of the curve by the isotropic subspace $\Upsilon$ defines a new curve $\widetilde{\Lambda}(t) := \Lambda(t)^{\Upsilon} \in L(\Upsilon^{\angle}/\Upsilon)$.

**Proposition 14.46.** *If $\Lambda(t)$ is analytic and monotone in $L(\Sigma)$, then $\widetilde{\Lambda}(t)$ is ample $L(\Upsilon^{\angle}/\Upsilon)$.*

*Proof.* By construction, in the reduced space $\Upsilon^{\angle}/\Upsilon$ we removed the intersection of $\Lambda(t)$ with $\Lambda(0)$. Hence for small $t$

$$\widetilde{\Lambda}(0) \cap \widetilde{\Lambda}(t) = \{0\}, \qquad \text{in} \quad L(\Upsilon^{\angle}/\Upsilon) \tag{14.31}$$

In particular, if $\widetilde{S}(t)$ denotes the symmetric matrix representing $\widetilde{\Lambda}(t)$ such that $\widetilde{S}(0)$ represents $\widetilde{\Lambda}(t_0)$, it follows that $\widetilde{S}(t)$ is non-degenerate for $0 < |t| < \varepsilon$. The analyticity of the curve guarantees that the Taylor polynomial (of a suitable order $N$) is also non-degenerate. □

## 14.6 From ample to regular

In this section we prove the main result of this chapter, i.e., that any ample monotone curve can be reduced to a regular one.

**Theorem 14.47.** *Let $\Lambda(t)$ be a smooth ample monotone curve and set $\Gamma := \ker \underline{\dot{\Lambda}}(0)$. Then the reduced curve $t \mapsto \Lambda^\Gamma(t)$ is a smooth regular curve. In particular $\underline{\dot{\Lambda}}^\Gamma(0) > 0$.*

Thanks to Theorem 14.47 the following definition is well-posed.

**Definition 14.48.** Let $\Lambda(t)$ be a smooth ample monotone curve and set $\Gamma := \ker \underline{\dot{\Lambda}}(0)$. Then the *principal curvature* of $\Lambda(t)$ is the cuvature operator associated with the regular curve $t \mapsto \Lambda^\Gamma(t)$.

Before proving Theorem 14.47, let us discuss two useful lemmas.

**Lemma 14.49.** *Let $v_1(t), \dots, v_k(t) \in \mathbb{R}^n$ and define $V(t)$ as the $n \times k$ matrix whose columns are the vectors $v_i(t)$. Define the matrix $S(t) := \int_0^t V(\tau)V(\tau)^* d\tau$. Then the following are equivalent:*

*(i) $S(t)$ is invertible (and positive definite),*

*(ii)* $\mathrm{span}\{v_i(\tau) | \, i = 1, \dots, k; \tau \in [0,t]\} = \mathbb{R}^n$.

*Proof.* Fix $t > 0$ and let us assume $S(t)$ is not invertible. Since $S(t)$ is non negative then there exists a nonzero $x \in \mathbb{R}^n$ such that $\langle S(t)x, x\rangle = 0$. On the other hand

$$\langle S(t)x, x\rangle = \int_0^t \langle V(\tau)V(\tau)^*x, x\rangle \, d\tau = \int_0^t \|V(\tau)^*x\|^2 d\tau.$$

This implies that $V(\tau)^*x = 0$ (or equivalently $x^*V(\tau) = 0$) for $\tau \in [0,t]$, i.e., the nonzero vector $x^*$ is orthogonal to $\mathrm{im}_{\tau\in[0,t]} V(\tau) = \mathrm{span}\{v_i(\tau) | \, i = 1, \dots, k, \, \tau \in [0,t]\} = \mathbb{R}^n$, that is a contradiction. The converse is similar. $\square$

**Lemma 14.50.** *Let $A, B$ be two positive and symmetric matrices such that $0 < A < B$. Then we have also $0 < B^{-1} < A^{-1}$.*

*Proof.* Assume first that $A$ and $B$ commute. Then $A$ and $B$ can be simultaneously diagonalized and the statement is trivial for diagonal matrices.

In the general case, since $A$ is symmetric and positive, we can consider its square root $A^{1/2}$, which is also symmetric and positive. We can write

$$0 < \langle Av, v\rangle < \langle Bv, v\rangle \, .$$

By setting $w = A^{1/2}v$ in the above inequality and using $\langle Av, v\rangle = \left\langle A^{1/2}v, A^{1/2}v\right\rangle$ one gets

$$0 < \langle w, w\rangle < \left\langle A^{-1/2}BA^{-1/2}w, w\right\rangle,$$

which is equivalent to $I < A^{-1/2}BA^{-1/2}$. Since the identity matrix commutes with every other matrix, we obtain

$$0 < A^{1/2}B^{-1}A^{1/2} = (A^{-1/2}BA^{-1/2})^{-1} < I,$$

which is equivalent to $0 < B^{-1} < A^{-1}$ reasoning as before. $\square$

*Proof of Theorem 14.47.* By assumption the curve $t \mapsto \Lambda(t)$ is ample, hence $\Lambda(t) \cap \Gamma = \{0\}$ and $t \mapsto \Lambda^\Gamma(t)$ is smooth for $t > 0$ small enough. We divide the proof into three parts: (i) we compute the coordinate presentation of the reduced curve. (ii) we show that the reduced curve, extended by continuity at $t = 0$, is smooth. (iii) we prove that the reduced curve is regular.

(i). Let us consider Darboux coordinates in the symplectic space $\Sigma$ such that

$$\Sigma = \{(p, x) : p, x \in \mathbb{R}^n\}, \qquad \Lambda(t) = \{(p, S(t)p) | \, p \in \mathbb{R}^n\}, \qquad S(0) = 0.$$

Morover we can assume also $\mathbb{R}^n = \mathbb{R}^k \oplus \mathbb{R}^{n-k}$, where $\Gamma = \{0\} \oplus \mathbb{R}^{n-k}$. According to this splitting we have the decomposition $p = (p_1, p_2)^T$ and $x = (x_1, x_2)^T$. The subspaces $\Gamma$ and $\Gamma^\angle$ are described by the equations

$$\Gamma = \{(p, x) : p_1 = 0, x = 0\}, \qquad \Gamma^\angle = \{(p, x) : x_2 = 0\}$$

and $(p_1, x_1)$ are natural coordinates for the reduced space $\Gamma^\angle/\Gamma$. Up to a symplectic change of coordinates preserving the splitting $\mathbb{R}^n = \mathbb{R}^k \oplus \mathbb{R}^{n-k}$ we can assume that

$$S(t) = \begin{pmatrix} S_{11}(t) & S_{12}(t) \\ S_{12}^*(t) & S_{22}(t) \end{pmatrix}, \qquad \text{with} \qquad \dot{S}(0) = \begin{pmatrix} \mathbb{I}_k & 0 \\ 0 & 0 \end{pmatrix}. \tag{14.32}$$

where $\mathbb{I}_k$ is the $k \times k$ identity matrix. Finally, the fact that $S$ is monotone and ample, implies that $S(t) > 0$ for each $t > 0$. It follows

$$S_{11}(t) > 0, \qquad S_{22}(t) > 0, \qquad \forall \, t > 0. \tag{14.33}$$

Then we can compute the coordinate expression of the reduced curve, i.e., the matrix $S^\Gamma(t)$ such that

$$\Lambda^\Gamma(t) = \{(p_1, S^\Gamma(t)p_1) \mid p_1 \in \mathbb{R}^k\}.$$

From the identity

$$\Lambda(t) \cap \Gamma^\angle = \{(p, S(t)p), S(t)p \in \mathbb{R}^k\} = \left\{ \left( S^{-1}(t) \begin{pmatrix} x_1 \\ 0 \end{pmatrix}, \begin{pmatrix} x_1 \\ 0 \end{pmatrix} \right), x_1 \in \mathbb{R}^k \right\}, \tag{14.34}$$

one gets the key relation $S^\Gamma(t)^{-1} = (S(t)^{-1})_{11}$.

Thus the matrix expression of the reduced curve $\Lambda^\Gamma(t)$ in $L(\Gamma^\angle/\Gamma)$ is simply recovered. We have

$$S(t)p = \begin{pmatrix} S_{11} & S_{12} \\ S_{12}^* & S_{22} \end{pmatrix} \begin{pmatrix} p_1 \\ p_2 \end{pmatrix} = \begin{pmatrix} S_{11}p_1 + S_{12}p_2 \\ S_{12}^*p_1 + S_{22}p_2 \end{pmatrix},$$

from which we get $S(t)p \in \mathbb{R}^k$ if and only if $S_{12}^*(t)p_1 + S_{22}(t)p_2 = 0$. Then

$$\begin{aligned} \Lambda^\Gamma(t) &= \{(p_1, S_{11}p_1 + S_{12}p_2) : S_{12}^*(t)p_1 + S_{22}(t)p_2 = 0\} \\ &= \{(p_1, (S_{11} - S_{12}S_{22}^{-1}S_{12}^*)p_1)\} \end{aligned}$$

that means

$$S^\Gamma = S_{11} - S_{12}S_{22}^{-1}S_{12}^*. \tag{14.35}$$

(ii). By the coordinate presentation of $S^\Gamma(t)$ the only term that can give rise to singularities is the inverse matrix $S_{22}^{-1}(t)$. In particular, since by assumption $t \mapsto \det S_{22}(t)$ has a finite order zero at $t = 0$, the a priori singularity can be only a finite order pole.

To prove that the curve is smooth it is enough to show that $S^\Gamma(t) \to 0$ for $t \to 0$, i.e., the curve remains bounded. This follows from the following

**Claim I.** As quadratic forms on $\mathbb{R}^k$, we have the inequalities $0 \le S^\Gamma(t) \le S_{11}(t)$.

Indeed $S(t)$ symmetric and positive one has that its inverse $S(t)^{-1}$ is symmetric and positive also. This implies that $S^\Gamma(t)^{-1} = (S(t)^{-1})_{11} > 0$ and so is $S^\Gamma(t)$. This proves the left inequality of the Claim I.

Moreover using (14.35) and the fact that $S_{22}$ is positive definite (and so $S_{22}^{-1}$) one gets

$$\left\langle (S_{11} - S^\Gamma)p_1, p_1 \right\rangle = \left\langle S_{12}S_{22}^{-1}S_{12}^* p_1, p_1 \right\rangle = \left\langle S_{22}^{-1}(S_{12}^* p_1), (S_{12}^* p_1) \right\rangle \ge 0.$$

Since $S(t) \to 0$ for $t \to 0$, clearly $S_{11}(t) \to 0$ when $t \to 0$, that proves that $S^\Gamma(t) \to 0$ also.

(iii). We are reduced to show that the derivative of $t \mapsto S^\Gamma(t)$ at 0 is non-degenerate matrix, which is equivalent to show that $t \mapsto S^\Gamma(t)^{-1}$ has a simple pole at $t = 0$.

We need the following lemma, whose proof is postponed at the end of the proof of Theorem 14.47.

**Lemma 14.51.** *Let $A(t)$ be a smooth family of symmetric nonnegative $n \times n$ matrices. If the condition* $\operatorname{rank}(A, \dot{A}, \dots, A^{(N)})|_{t=0} = n$ *is satisfied for some $N$, then there exists $\varepsilon_0 > 0$ such that $\varepsilon t A(0) < \int_0^t A(\tau)d\tau$ for all $\varepsilon < \varepsilon_0$ and $t > 0$ small enough.*

Applying the Lemma to the family $A(t) = \dot{S}(t)$ one obtains (see also (14.32))

$$\langle S(t)p, p \rangle > \varepsilon t |p_1|^2,$$

for all $0 < \varepsilon < \varepsilon_0$, any $p \in \mathbb{R}^n$ and any small time $t > 0$.

Now let $p_1 \in \mathbb{R}^k$ be arbitrary and extend it to a vector $p = (p_1, p_2) \in \mathbb{R}^n$ such that $(p, S(t)p) \in \Lambda(t) \cap \Gamma^\angle$ (it is enough to choose $p_2 = -S_{22}^{-1}S_{12}^* p_1$). This implies in particular that $S^\Gamma(t)p_1 = x_1$ and

$$\left\langle S^\Gamma(t)p_1, p_1 \right\rangle = \langle S(t)p, p \rangle \ge \varepsilon t |p_1|^2,$$

This identity can be rewritten as $S^\Gamma(t) > \varepsilon t\, \mathbb{I}_k > 0$ and implies by Lemma 14.50

$$0 < S^\Gamma(t)^{-1} < \frac{1}{\varepsilon t}\mathbb{I}_k,$$

which completes the proof. □

*Proof of Lemma 14.51.* We reduce the proof of the Lemma to the following statement:

**Claim II.** *There exists $\widehat{c}, \widehat{N} > 0$ such that for any sufficiently small $\varepsilon, t > 0$*

$$\det\left(\int_0^t A(\tau) - \varepsilon A(0)\, d\tau\right) > \widehat{c}\, t^{\widehat{N}}. \tag{14.36}$$

*Moreover $\widehat{c}, \widehat{N}$ depends only on the $2N$-th Taylor polynomial of $A(t)$.*

Indeed, assume Claim II is true and fix $t_0 > 0$ small enough. Since $A(t) \ge 0$, from (14.36) for $\varepsilon = 0$ we have that $\int_0^{t_0} A(\tau)d\tau > 0$ (a non-negative quadratic form with positive determinant is positive). Hence by continuity, for a fixed $t_0$, there exists $\varepsilon$ small enough such that $\int_0^{t_0} A(\tau) - \varepsilon A(0)\, d\tau > 0$. Assume now that the matrix $K(t) = \int_0^t A(\tau) - \varepsilon A(0)\, d\tau$ is not strictly positive for some $0 < t < t_0$, then $\det K(\tau) = 0$ for some $\tau \in [t, t_0]$, that is a contradiction.

We now prove Claim II. We may assume that $t \mapsto A(t)$ is analytic. Indeed, by smoothness of the determinant, the statement remains true if we substitute $A(t)$ by its Taylor polynomial of sufficiently big order.

An analytic one parameter family of symmetric matrices $t \mapsto A(t)$ can be simultaneously diagonalized (see [Kat95]), in the sense that there exists an analytic (with respect to $t$) family of vectors $v_i(t)$, with $i = 1, \ldots, n$, such that

$$\langle A(t)x, x \rangle = \sum_{i=1}^{n} \langle v_i(t), x \rangle^2 .$$

In other words $A(t) = V(t)V(t)^*$, where $V(t)$ is the $n \times n$ matrix whose columns are the vectors $v_i(t)$. (Notice that some of these vector can vanish at 0 or even vanish identically.)

Let us now consider the flag $E_1 \subset E_2 \subset \ldots \subset E_N = \mathbb{R}^n$ defined as follows

$$E_i = \operatorname{span}\{v_j^{(l)}, 1 \leq j \leq n, 0 \leq l \leq i\}.$$

Notice that this flag is finite by our assumption on the rank of the consecutive derivatives of $A(t)$ and $N$ is the same as in the statement of the Lemma. We then choose coordinates in $\mathbb{R}^n$ adapted to this flag (i.e., the spaces $E_i$ are coordinate subspaces) and define the following integers (here $e_1, \ldots, e_n$ is the standard basis of $\mathbb{R}^n$)

$$m_i = \min\{j : e_i \in E_j\}, \qquad i = 1, \ldots, n.$$

In other words, when written in this new coordinate set, $m_i$ is the order of the first nonzero term in the Taylor expansion of the $i$-th row of the matrix $V(t)$. Then we introduce a quasi-homogeneous family of matrices $\widehat{V}(t)$: the $i$-th row of $\widehat{V}(t)$ is the $m_i$-homogeneous part of the $i$-the row of $V(t)$. Then we define $\widehat{A}(t) := \widehat{V}(t)\widehat{V}(t)^*$. The columns of the matrix $\widehat{A}(t)$ satisfy the assumption of Lemma 14.49, then $\int_0^t \widehat{A}(\tau)d\tau > 0$ for every $t > 0$.

If we denote the entries $A(t) = \{a_{ij}(t)\}_{i,j=1}^n$ and $\widehat{A}(t) = \{\widehat{a}_{ij}(t)\}_{i,j=1}^n$ we obtain

$$\widehat{a}_{ij}(t) = \widehat{c}_{ij} t^{m_i+m_j}, \qquad a_{ij}(t) = \widehat{a}_{ij}(t) + O(t^{m_i+m_j+1}),$$

for suitable constants $\widehat{c}_{ij}$ (some of them may be zero).

Then we let $A^\varepsilon(t) := A(t) - \varepsilon A(0) = \{a_{ij}^\varepsilon(t)\}_{i,j=1}^n$. Of course $a_{ij}^\varepsilon(t) = c_{ij}^\varepsilon t^{m_i+m_j} + O(t^{m_i+m_j+1})$ where

$$c_{ij}^\varepsilon = \begin{cases} (1-\varepsilon)\widehat{c}_{ij}, & \text{if } m_i + m_j = 0, \\ \widehat{c}_{ij}, & \text{if } m_i + m_j > 0. \end{cases}$$

From the equality

$$\int_0^t a_{ij}^\varepsilon(\tau)d\tau = t^{m_i+m_j+1}\left(\frac{c_{ij}^\varepsilon}{m_i+m_j+1} + O(t)\right),$$

one gets

$$\det\left(\int_0^t A^\varepsilon(\tau)d\tau\right) = t^{n+2\sum_{i=1}^N m_i}\left(\det\left(\frac{c_{ij}^\varepsilon}{m_i+m_j+1}\right) + O(t)\right).$$

On the other hand

$$\det\left(\int_0^t \widehat{A}(\tau)d\tau\right) = t^{n+2\sum_{i=1}^N m_i}\left(\det\left(\frac{\widehat{c}_{ij}}{m_i+m_j+1}\right)+O(t)\right) > 0,$$

hence $\det\left(\frac{c^\varepsilon_{ij}}{m_i+m_j+1}\right) > 0$ for small $\varepsilon$. The proof is completed by setting

$$\widehat{c} := \det\left(\frac{\widehat{c}_{ij}}{m_i+m_j+1}\right), \qquad \widehat{N} := n+2\sum_{i=1}^N m_i.$$

□

## 14.7 Conjugate points in $L(\Sigma)$

In this section we introduce the notion of conjugate point for a curve in the Lagrange Grassmannian. In the next chapter we explain why this notion coincide with the one given for extremal paths in sub-Riemannian geometry.

**Definition 14.52.** Let $\Lambda(t)$ be a monotone curve in $L(\Sigma)$. We say that $\Lambda(t)$ *is conjugate to* $\Lambda(0)$ along $\Lambda(\cdot)$ if $\Lambda(t)\cap\Lambda(0)\neq\{0\}$.

Conjugate points along a curve are independent on the parametrization. When a parametrization is fixed, one can speak about conjugate times along the curve. We have the following proposition, cf. also Corollary 8.51.

**Proposition 14.53.** *Conjugate points on a monotone and ample curve in $L(\Sigma)$ are isolated.*

*Proof.* Without loss of generality we can assume that the curve is monotone increasing. Notice that from Lemma 14.44 we immediately get that the first conjugate time is separated from zero. To prove that the set of conjugate time is discrete, recall that thanks to Exercice 14.10 the unique invariant of two pair of Lagrangian subspaces is the dimension of their intersection.

Fix a time $\bar{t}$ that is conjugate to zero and a set of coordinates in the Lagrange Grassmannian such that $\Lambda(\bar{t}) = 0$ and $\Lambda(0) \leq 0$. Since the curve is monotone increasing and ample we have that $\Lambda(\tau) > 0$ for all $\tau > \bar{t}$, hence $\Lambda(\tau)\cap\Lambda(0) = 0$ for $\tau > \bar{t}$ small enough. To prove the analogue statement for $\tau < \bar{t}$ it is enough to choose the coordinate chart in such a way that $\Lambda(\bar{t}) = 0$ and $\Lambda(0) \geq 0$ and repeat the previous argument. □

The following results describe two general properties of conjugate points

**Theorem 14.54.** *Let $\Lambda(t), \Delta(t)$ two ample and monotone curves in $L(\Sigma)$ defined on $\mathbb{R}$ such that*

*(i)* $\Sigma = \Lambda(t)\oplus\Delta(t)$ *for every* $t \geq 0$,

*(ii)* $\dot{\Lambda}(t) \geq 0$, $\dot{\Delta}(t) \leq 0$, *as quadratic forms.*

*Then there exists no $\tau > 0$ such that $\Lambda(\tau)$ is conjugate to $\Lambda(0)$ along $\Lambda(\cdot)$. Moreover there exists* $\Lambda(\infty) := \lim_{t\to+\infty}\Lambda(t)$.

*Proof.* Fix coordinates induced by some Lagrangian splitting of $\Sigma$ in such a way that $S_{\Lambda(0)} = 0$ and $S_{\Delta(0)} = I$. The monotonicity assumption implies that $t \mapsto S_{\Lambda(t)}$ (resp. $t \mapsto S_{\Delta(t)}$) is a monotone increasing (resp. decreasing) curve in the space of symmetric matrices. Moreover the tranversality of $\Lambda(t)$ and $\Delta(t)$ implies that $S_\Delta(t) - S_{\Lambda(t)}$ is a non-degenerate matrix for all $t$. Hence

$$0 < S_{\Lambda(t)} < S_\Delta(t) < I, \qquad \text{for all } t > 0.$$

In particular $\Lambda(t)$ never leaves the coordinate neighborhood under consideration, the subspace $\Lambda(t)$ is always traversal to $\Lambda(0)$ for $t > 0$ and has a limit $\Lambda(\infty)$ whose coordinate representation is $S_\Lambda(\infty) = \lim_{t\to+\infty} S_\Lambda(t)$. □

**Theorem 14.55.** *Let $\Lambda_0(t), \Lambda_1(t)$ be two monotone and ample curves in $L(\Sigma)$ such that $\Lambda_0(0) = \Lambda_1(0) = \Lambda$. Assume that there exists a homotopy of curves $\Lambda_s(t)$ such that $\Lambda_s(0) = \Lambda$ for $s \in [0,1]$ and satisfying the following properties:*

*(i)* $\Lambda_0(\cdot)$ *contains no conjugate points to* $\Lambda$*,*

*(ii)* $\Lambda_s(\cdot)$ *is monotone and ample for every* $s \in [0,1]$*,*

*(iii)* $\Lambda_s(1) \cap \Lambda = \{0\}$ *for all* $s \in [0,1]$*.*

*Then the curve $t \mapsto \Lambda_1(t)$ contains no conjugate points to $\Lambda$.*

*Proof.* Let us consider the open chart $\Lambda^\pitchfork$ defined by all the Lagrangian subspaces traversal to $\Lambda$. The statement is equivalent to the fact that $\Lambda_1(t) \in \Lambda^\pitchfork$ for all $t > 0$. Let us fix coordinates induced by some Lagrangian splitting $\Sigma = \Lambda \oplus \Delta$ in such a way that $\Lambda = \{(p, 0), p \in \mathbb{R}^n\}$ and

$$\Lambda_s(t) = \{(B_s(t)x, x), x \in \mathbb{R}^n\},$$

for all $s$ and $t > 0$ (at least for $t$ small enough, indeed by ampleness $\Lambda_s(t) \in \Lambda^\pitchfork$ for $t$ small).

Recall that $B_s(t)$ is the inverse of a matrix $S_s(t)$ that is monotonically increasing and such that $S_s(t) \to 0$ for $t \to 0^+$.

It follows that $B_s(t)$ is a monotone decreasing family of symmetric matrices and, for every $x \in \mathbb{R}^n$, one has $x^T B_s(\tau)x \to +\infty$ when $\tau \to 0^+$ (this quantity goes monotonically to $\infty$ and since it is decreasing on $]0,1[$ it goes to $+\infty$)). Analogously, a necessary condition for $\Lambda_s(t)$ to be conjugate to $\Lambda$ is that there exists a nonzero $x$ such that $x^T B_s(\tau)x \to \infty$ for $\tau \to t$.

It is then enough to show that, for all $x \in \mathbb{R}^n$ the function $(t,s) \mapsto x^T B_s(t)x$ is bounded. Indeed by assumptions $t \mapsto x^T B_0(t)x$ and $t \mapsto x^T B_1(t)x$ are monotone decreasing and bounded up to $t = 1$. Hence the continuous family of values $M_s := x^T B_s(1)x$ is well-defined and bounded for all $s \in [0,1]$. The monotonicity implies that actually $x^T B_s(t)x > -\infty$ for all values of $t, s \in [0,1]$. (See also Figure 14.1).

□

## 14.8 Comparison theorems for regular curves

In this last section we prove two comparison theorems for regular monotone curves in the Lagrange Grassmannian.

**Proposition 14.56.** *Let $\Lambda(t)$ be a monotone and regular curve in the Lagrange Grassmannian such that $R_\Lambda(t) \leq 0$. Then $\Lambda(t)$ contains no conjugate points to $\Lambda(0)$ along $\Lambda(\cdot)$.*

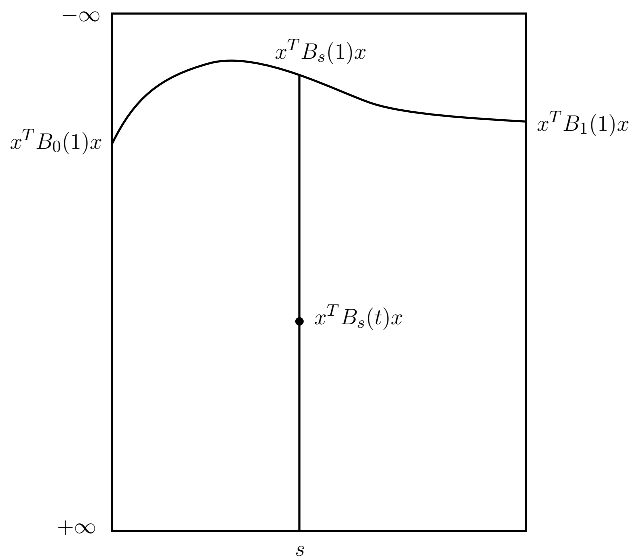



Figure 14.1: Proof of Theorem 14.55

*Proof.* Assume without loss of generality that $\Lambda(t)$ is monotone increasing. Then $R_\Lambda(t) \leq 0$ implies that $\Lambda^\circ(t)$ is monotone decreasing and applying Theorem 14.54 with $\Delta(t) = \Lambda^\circ(t)$ the statement follows. □

**Theorem 14.57.** *Let $\Lambda(t)$ be a monotone and regular curve in the Lagrange Grassmannian. Assume that there exists $k \geq 0$ such that*

*(i)* $R_\Lambda(t) \leq k\,\mathrm{Id}$ *for all $t \geq 0$. Then if $\Lambda(t_0)$ is conjugate to $\Lambda(0)$ along $\Lambda(\cdot)$, we have $t_0 \geq \frac{\pi}{\sqrt{k}}$.*

*(ii)* $\operatorname{trace} R_\Lambda(t) \geq nk$ *for all $t \geq 0$. Then for every $\tau \geq 0$ there exists $t_0 \in [\tau, \tau + \frac{\pi}{\sqrt{k}}]$ such that $\Lambda(t_0)$ is conjugate to $\Lambda(0)$ along $\Lambda(\cdot)$.*

We stress that assumption (i) means that all the eigenvalues of $R_\Lambda(t)$ are smaller or equal than $k$, while (ii) requires only that the average of the eigenvalues is bigger or equal than $k$.

*Remark* 14.58. Notice that the estimates of Theorem 14.57 are sharp, as it is immediately seen by considering the example of a 1-dimensional curve of constant velocity (cf. Example 14.36).

*Proof.* (i). Consider the real function

$$\varphi : \mathbb{R} \to ]0, \frac{\pi}{\sqrt{k}}[, \qquad \varphi(t) = \frac{1}{\sqrt{k}}(\arctan \sqrt{k}t + \frac{\pi}{2}).$$

Using that $\dot{\varphi}(t) = (1 + kt^2)^{-1}$ it is easy to show that the Schwarzian derivative of $\varphi$ is

$$R_\varphi(t) = -\frac{k}{(1 + kt^2)^2}.$$

Thus using $\varphi$ as a reparametrization we find, following the notation of Proposition 14.37 that

$$\begin{aligned} R_{\Lambda_\varphi}(t) &= \dot{\varphi}^2 R_\Lambda(\varphi(t)) + R_\varphi(t)\mathrm{Id} \\ &= \frac{1}{(1 + kt^2)^2}(R_\Lambda(\varphi(t)) - k\mathrm{Id}) \leq 0. \end{aligned}$$

By Corollary 14.56 the curve $\Lambda_\varphi := \Lambda \circ \varphi$ has no conjugate points on $\mathbb{R}$, i.e., $\Lambda$ has no conjugate points in the interval $]0, \frac{\pi}{\sqrt{k}}[$.

(ii). We prove the claim by showing that the curve $\Lambda(t)$, on every interval of length $\pi/\sqrt{k}$ has non trivial intersection with every subspace (hence in particular with $\Lambda(0)$). This is equivalent to prove that $\Lambda(t)$ is not contained in a single coordinate chart for a whole interval of length $\pi/\sqrt{k}$.

Assume by contradiction that $\Lambda(t)$ is contained in a single coordinate chart. Then there exists coordinates such that $\Lambda(t) = \{(p, S(t)p)\}$ and we can write the coordinate expression for the curvature:

$$R_\Lambda(t) = \dot{B}(t) - B(t)^2, \qquad \text{where } B(t) = (2\dot{S}(t))^{-1}\ddot{S}(t).$$

Let now $b(t) := \operatorname{trace} B(t)$. Computing the trace in both sides of the identity

$$\dot{B}(t) = B^2(t) + R_\Lambda(t),$$

one gets the scalar equation

$$\dot{b}(t) = \operatorname{trace}(B^2(t)) + \operatorname{trace} R_\Lambda(t). \tag{14.37}$$

**Lemma 14.59.** *For every $n \times n$ symmetric matrix $S$, the following inequality holds true*

$$\operatorname{trace}(S^2) \geq \frac{1}{n}(\operatorname{trace} S)^2. \tag{14.38}$$

*Proof.* For every symmetric matrix $S$ there exists a matrix $M$ such that $MSM^{-1} = D$ is diagonal. Since $\operatorname{trace}(MAM^{-1}) = \operatorname{trace}(A)$ for every matrix $A$, it is enough to prove the inequality (14.38) for a real diagonal matrix $D = \operatorname{diag}(\lambda_1, \dots, \lambda_n)$. In this case (14.38) reduces to the classical Cauchy-Schwartz inequality

$$\sum_{i=1}^n \lambda_i^2 \geq \frac{1}{n}\left(\sum_{i=1}^n \lambda_i\right)^2.$$

□

Applying Lemma 14.59 to (14.37) and using the assumption (ii), one gets

$$\dot{b}(t) \geq \frac{1}{n}b^2(t) + nk, \tag{14.39}$$

By standard results in ODE theory we have $b(t) \geq \varphi(t)$ , where $\varphi(t)$ is the solution of the differential equation

$$\dot{\varphi}(t) = \frac{1}{n}\varphi^2(t) + nk. \tag{14.40}$$

The unique solution of (14.40), with initial datum $\varphi(t_0) = 0$, is explicitly given by

$$\varphi(t) = n\sqrt{k}\tan(\sqrt{k}(t - t_0)).$$

This solution is defined on an interval of measure $\pi/\sqrt{k}$. Thus the inequality $b(t) \geq \varphi(t)$ completes the proof. □

## 14.9 Bibliographical note

The theory of invariants of curves in the Lagrange Grassmannian has been developed in the literature in relation to different questions, such as geodesic flows, control theory and projective differential geometry.

The fact that the Schwarzian derivative can be interpreted as the curvature of a curve on the projective line is already observed in [Fla70]. The theory of regular curves in the Lagrange Grassmannian has also been previously studied in [Ahd89], in relation to the Riemannian geodesic flow, in [Ovs89, Ovs93] in relation to Sturm systems, in [Zel00] in relation to Riccati equation, and, more recently, in [APD09] in relation to symplectic invariants. In these references, regular curves are called *fanning* curves.

The approach discussed here has been extensively developed in control theory starting from the papers [AG97, Agr98b, AG98] and then subsequently in [AZ02b, AZ02a]. A discussion of many of the results contained in this chapter can also be found in the lecture notes [Agr08], where the theory of Morse index is also developed. The proof of Theorem 14.47 given here is original; an alternative proof of the same result is also contained in [ABR18], where the relation between curvature of curves in Lagrange Grassmannian and curvature in sub-Riemannian geometry and more general control problem is revisited. The problem of finding a complete set of invariants for ample but non-regular curves in Lagrange Grassmannian have been developed in [Zel06, ZL07, ZL09] (cf. Appendix).

# Chapter 15

# Jacobi curves

Now we are ready to introduce the main object of this part of the book, i.e., the Jacobi curve associated with a normal extremal. Heuristically, we would like to extract geometric properties of the sub-Riemannian structure by studying the symplectic invariants of its geodesic flow, that is the flow of $\vec{H}$. The simplest idea is to look for invariants in its linearization.

As we explain in the next sections, this object is naturally related to geodesic variations, and generalizes the notion of Jacobi fields in Riemannian geometry to more general geometric structures.

In this chapter we consider a sub-Riemannian structure $(M, \mathbf{U}, f)$ on a smooth $n$-dimensional manifold $M$ and we denote as usual by $H : T^*M \to \mathbb{R}$ its sub-Riemannian Hamiltonian.

## 15.1 From Jacobi fields to Jacobi curves

Fix a covector $\lambda \in T^*M$, with $\pi(\lambda) = q$, and consider the normal extremal starting from $q$ and associated with $\lambda$, i.e.

$$\lambda(t) = e^{t\vec{H}}(\lambda), \qquad \gamma(t) = \pi(\lambda(t)). \qquad \text{(i.e. } \lambda(t) \in T^*_{\gamma(t)}M.\text{)}$$

For any $\xi \in T_\lambda(T^*M)$ we can define a vector field along the extremal $\lambda(t)$ as follows

$$X(t) := e^{t\vec{H}}_* \xi \in T_{\lambda(t)}(T^*M)$$

The set of vector fields obtained in this way is a $2n$-dimensional vector space which is the space of *Jacobi fields* along the extremal. For an Hamiltonian $H$ corresponding to a Riemannian structure, the projection $\pi_*$ gives an isomorphism between the space of Jacobi fields along the extremal and the classical space of Jacobi fields along the geodesic $\gamma(t) = \pi(\lambda(t))$.

Notice that this definition, equivalent to the standard one in Riemannian geometry, does not need curvature or connection, and can be extended naturally for any strongly normal sub-Riemannian geodesic.

In Riemannian geometry, the study of one half of this vector space, namely the subspace of classical Jacobi fields vanishing at zero, carries informations about conjugate points along the given geodesic. By the aforementioned isomorphism, this corresponds to the subspace of Jacobi fields along the extremal such that $\pi_* X(0) = 0$. This motivates the following construction: For any $\lambda \in T^*M$, we denote $\mathcal{V}_\lambda := \ker \pi_*|_\lambda$ the vertical subspace. We could study the whole family of (classical) Jacobi fields (vanishing at zero) by means of the family of subspaces along the extremal

$$L(t) := e_*^{t\vec{H}} \mathcal{V}_\lambda \subset T_{\lambda(t)}(T^*M).$$

Notice that actually, being $e_*^{t\vec{H}}$ a symplectic transformation and $\mathcal{V}_\lambda$ a Lagrangian subspace, the subspace $L(t)$ is a Lagrangian subspace of $T_{\lambda(t)}(T^*M)$.

### 15.1.1 Jacobi curves

The theory of curves in the Lagrange Grassmannian developed in Chapter 14 is an efficient tool to study family of Lagrangian subspaces contained in a single symplectic vector space. It is then convenient to modify the construction of the previous section in order to collect the informations about the linearization of the Hamiltonian flow into a family of Lagrangian subspaces at a fixed tangent space.

By definition, the pushforward of the flow of $\vec{H}$ maps the tangent space to $T^*M$ at the point $\lambda(t)$ back to the tangent space to $T^*M$ at $\lambda$:

$$e_*^{-t\vec{H}} : T_{\lambda(t)}(T^*M) \to T_\lambda(T^*M).$$

If we then restrict the action of the pushforward $e_*^{-t\vec{H}}$ to the vertical subspace at $\lambda(t)$, i.e. the tangent space $T_{\lambda(t)}(T^*_{\gamma(t)}M)$ at the point $\lambda(t)$ to the fiber $T^*_{\gamma(t)}M$, we define a one parameter family of $n$-dimensional subspaces in the $2n$-dimensional vector space $T_\lambda(T^*M)$. This family of subspaces is a curve in the Lagrangian Grassmannian $L(T_\lambda(T^*M))$.

**Notation.** In the following we use the notation $\mathcal{V}_\lambda := T_\lambda(T^*_qM)$ for the vertical subspace at the point $\lambda \in T^*M$, i.e. the tangent space at $\lambda$ to the fiber $T^*_qM$, where $q = \pi(\lambda)$. Being the tangent space to a vector space, sometimes it will be useful to identify the vertical space $\mathcal{V}_\lambda$ with the vector space itself, namely $\mathcal{V}_\lambda \simeq T^*_qM$.

**Definition 15.1.** Let $\lambda \in T^*M$. The *Jacobi curve* at the point $\lambda$ is defined as follows

$$J_\lambda(t) := e_*^{-t\vec{H}} \mathcal{V}_{\lambda(t)}, \tag{15.1}$$

where $\lambda(t) := e^{t\vec{H}}(\lambda)$ and $\gamma(t) = \pi(\lambda(t))$. Notice that $J_\lambda(t) \subset T_\lambda(T^*M)$ and $J_\lambda(0) = \mathcal{V}_\lambda = T_\lambda(T^*_qM)$ is vertical.

As discussed in Chapter 14, the tangent vector to a curve in the Lagrange Gassmannian can be interpreted as a quadratic form. In the case of a Jacobi curve $J_\lambda(t)$ its tangent vector is a quadratic form $\underline{\dot{J}}_\lambda(t) : J_\lambda(t) \to \mathbb{R}$.

**Proposition 15.2.** *The Jacobi curve $J_\lambda(t)$ satisfies the following properties:*

*(i)* $J_\lambda(t+s) = e_*^{-t\vec{H}} J_{\lambda(t)}(s)$, *for all* $t, s \geq 0$,

*(ii)* $\underline{\dot{J}}_\lambda(0) = -2H|_{T^*_qM}$ *as quadratic forms on* $\mathcal{V}_\lambda \simeq T^*_qM$.

*(iii)* $\operatorname{rank} \underline{\dot{J}}_\lambda(t) = \operatorname{rank} H|_{T^*_{\gamma(t)}M}$

*Proof.* Claim (i) is a consequence of the semigroup property of the family $\{e_*^{-t\vec{H}}\}_{t\geq 0}$.

To prove (ii), introduce canonical coordinates $(p, x)$ in the cotangent bundle. Fix $\xi \in \mathcal{V}_\lambda$. The smooth family of vectors defined by $\xi(t) = e_*^{-t\vec{H}}\xi$ (considering $\xi$ as a constant vertical vector field) is a smooth extension of $\xi$, i.e. it satisfies $\xi(0) = \xi$ and $\xi(t) \in J_\lambda(t)$. Therefore, by (14.8)

$$\underline{\dot{J}}_\lambda(0)\xi = \sigma(\xi, \dot{\xi}) = \sigma\left(\xi, \frac{d}{dt}\bigg|_{t=0} e_*^{-t\vec{H}}\xi\right) = \sigma(\xi, [\vec{H}, \xi]). \tag{15.2}$$

To compute the last quantity we use the following elementary, although very useful, property of the symplectic form $\sigma$.

**Lemma 15.3.** *Let $\xi \in \mathcal{V}_\lambda$ a vertical vector. Then, for any $\eta \in T_\lambda(T^*M)$*

$$\sigma(\xi, \eta) = \langle \xi, \pi_*\eta \rangle, \tag{15.3}$$

*where we used the canonical identification $\mathcal{V}_\lambda = T_q^*M$.*

*Proof.* In any Darboux basis induced by canonical local coordinates $(p, x)$ on $T^*M$, we have $\sigma = \sum_{i=1}^n dp_i \wedge dx_i$ and $\xi = \sum_{i=1}^n \xi^i \partial_{p_i}$. The result follows immediately. □

To complete the proof of point (ii) it is enough to compute in coordinates

$$\pi_*[\vec{H}, \xi] = \pi_*\left[\frac{\partial H}{\partial p}\frac{\partial}{\partial x} - \frac{\partial H}{\partial x}\frac{\partial}{\partial p}, \xi\frac{\partial}{\partial p}\right] = -\frac{\partial^2 H}{\partial p^2}\xi\frac{\partial}{\partial x},$$

Hence by Lemma 15.3 and the fact that $H$ is quadratic on fibers one gets

$$\sigma(\xi, [\vec{H}, \xi]) = -\left\langle \xi, \frac{\partial^2 H}{\partial p^2}\xi \right\rangle = -2H(\xi).$$

(iii). The statement for $t = 0$ is a direct consequence of (ii). Using property (i) it is easily seen that the quadratic forms associated with the derivatives at different times are related by the formula

$$\underline{\dot{J}}_\lambda(t) \circ e_*^{t\vec{H}} = \underline{\dot{J}}_{\lambda(t)}(0). \tag{15.4}$$

Since $e_*^{-t\vec{H}}$ is a symplectic transformation, it preserves the sign and the rank of the quadratic form.[1] □

*Remark* 15.4. Notice that claim (iii) of Proposition 15.2 implies that rank of the derivative of the Jacobi curve is equal to the rank of the sub-Riemannian structure. Hence the curve is regular if and only if it is associated with a Riemannian structure. In this case of course it is strictly monotone, namely $\underline{\dot{J}}_\lambda(t) < 0$ for all $t$.

**Corollary 15.5.** *The Jacobi curve $J_\lambda(t)$ associated with a sub-Riemannian extremal is monotone nonincreasing for every $\lambda \in T^*M$.*

[1] Notice that $\underline{\dot{J}}_\lambda(t)$, $\underline{\dot{J}}_{\lambda(t)}(0)$ are defined on $J_\lambda(t)$, $J_{\lambda(t)}(0)$ respectively, and $J_\lambda(t) = e_*^{-t\vec{H}} J_{\lambda(t)}(0)$.

## 15.2 Conjugate points and optimality

At this stage we have two possible definitions for conjugate points along normal geodesics. On one hand we have singular points of the exponential map along the extremal path, on the other hand we can consider conjugate points of the associated Jacobi curve. The next result show that actually the two definition coincide.

**Proposition 15.6.** *Let $\gamma(t) = \exp_q(t\lambda)$ be a normal geodesic starting from $q$ with initial covector $\lambda$. Denote by $J_\lambda(t)$ its Jacobi curve and fix $s > 0$. Then the following statement are equivalent:*

*(a) $\gamma(s)$ is conjugate to $\gamma(0)$ along $\gamma(\cdot)$,*

*(b) $J_\lambda(s)$ is conjugate to $J_\lambda(0)$ along $J_\lambda(\cdot)$.*

*Proof.* By Definition 8.45, $\gamma(s)$ is conjugate to $\gamma(0)$ if $s\lambda$ is a critical point of the exponential map $\exp_q$. This is equivalent to say that the differential of the map from $T_q^*M$ to $M$ defined by $\lambda \mapsto \pi \circ e^{s\vec{H}}(\lambda)$ is not surjective at the point $\lambda$, i.e. the image of the differential $e_*^{s\vec{H}}$ has a nontrivial intersection with the kernel of the projection $\pi_*$

$$e_*^{s\vec{H}} J_\lambda(0) \cap T_{\lambda(s)} T^*_{\gamma(s)} M \neq \{0\}. \tag{15.5}$$

Applying the linear invertible transformation $e_*^{-s\vec{H}}$ to both subspaces one gets that (15.5) is equivalent to

$$J_\lambda(0) \cap J_\lambda(s) \neq \{0\}$$

which means by definition that $J_\lambda(s)$ is conjugate to $J_\lambda(0)$. $\square$

The next result shows that, as soon as we have a segment of points that are conjugate to the initial one, the segment is also abnormal. The argument contained in this proof should be compared with the one given in the proof of Theorem 8.47.

**Theorem 15.7.** *Let $\gamma : [0, 1] \to M$ be a normal extremal path. Assume $\gamma|_{[t_0,t_1]}$ is a curve of conjugate points to $\gamma(0)$. Then the restriction $\gamma|_{[t_0,t_1]}$ is also abnormal.*

*Remark* 15.8. Recall that if a curve $\gamma : [0, T] \to M$ is a strictly normal trajectory, it can happen that a piece of it is abnormal as well. If the trajectory is strongly normal, then if $t_0, t_1$ satisfy the assumptions of Theorem 15.7 necessarily $t_0 > 0$.

*Proof.* Let us denote by $J_\lambda(t)$ the Jacobi curve associated with $\gamma(t)$. From Proposition 15.6 it follows that $J_\lambda(t) \cap J_\lambda(0) \neq \{0\}$ for each $t \in [t_0, t_1]$. We now show that actually this implies

$$J_\lambda(0) \cap \bigcap_{t \in [t_0,t_1]} J_\lambda(t) \neq \{0\}. \tag{15.6}$$

We can assume that the whole piece of the Jacobi curve $J_\lambda(t)$, with $t_0 \leq t \leq t_1$, is contained in a single coordinate chart. Otherwise we can cover $[t_0, t_1]$ with such intervals and repeat the argument on each of them. Let us fix coordinates given by a Lagrangian splitting in such a way that

$$J_\lambda(t) = \{(p, S(t)p), p \in \mathbb{R}^n\}, \qquad J_\lambda(0) = \{(p, 0), p \in \mathbb{R}^n\}$$

Moreover we can choose coordinates in such a way that $\dot S(t) \leq 0$ for every $t_0 \leq t \leq t_1$, i.e., is non positive and monotone decreasing. Recall that this is possible thanks to Exercice 14.10.

In particular $J_\lambda(t_1) \cap J_\lambda(0) \neq \{0\}$ if and only if there exists a vector $v$ such that $S(t_1)v = 0$. Since the map $t \mapsto v^T S(t) v$ is nonpositive and decreasing this means that $S(t)v = 0$ for all $t \in [t_0, t_1]$, thus

$$J_\lambda(0) \cap J_\lambda(t_1) \subset J_\lambda(0) \cap \bigcap_{t \in [t_0, t_1]} J_\lambda(t) \tag{15.7}$$

that implies that actually we have the equality in (15.7).

We are left to show that if a Jacobi curve $J_\lambda(t)$ is such that every $t$ is a conjugate point for $0 \leq t \leq \tau$, then the corresponding extremal is also abnormal. Indeed let us fix an element $\xi \neq 0$ such that

$$\xi \in \bigcap_{t \in [0, \tau]} J_\lambda(t)$$

which is non-empty by the above discussion. Then we consider the vertical vector field

$$\xi(t) = e_*^{t\vec{H}} \xi \in T_{\lambda(t)}(T^*_{\gamma(t)} M), \qquad 0 \leq t \leq \tau.$$

By construction, the derivative in time of the vector field $\xi$ is vertical, i.e., $\pi_*[\vec{H}, \xi](\lambda(t)) = 0$. Then the statement is proved by the following exercice.

**Exercise 15.9.** Define $\eta(t) = \xi(\lambda(t)) \in T^*_{\gamma(t)} M$ (by canonical identification $T_\lambda(T^*_q M) \simeq T^*_q M$). Show that the identity $\pi_*[\vec{H}, \xi](\lambda(t)) = 0$ rewrites in coordinates as follows

$$\sum_{i=1}^k h_i(\eta(t))^2 = 0, \qquad \dot\eta(t) = \sum_{i=1}^k h_i(\lambda(t)) \vec{h}_i(\eta(t)). \tag{15.8}$$

Exercise 15.9 indeed shows that $\eta(t)$ is a family of covectors associated with the extremal path corresponding to controls $u_i(t) = h_i(\lambda(t))$ and such that $h_i(\eta(t)) = 0$, that means that it is abnormal.

$\square$

**Corollary 15.10.** *Let $J_\lambda(t)$ be the Jacobi curve associated with $\lambda \in T^*M$ and $\gamma(t) = \pi(\lambda(t))$ the associated sub-Riemannian extremal path. Then $\gamma|_{[0,t_0]}$ is abnormal if and only if $J_\lambda(\tau) \cap J_\lambda(0) \neq \{0\}$ for all $0 \leq \tau \leq t_0$.*

## 15.3 Reduction of the Jacobi curves by homogeneity

The Jacobi curve at point $\lambda \in T^*M$ parametrizes all the possible geodesic variations of the geodesic associated with an initial covector $\lambda$. Since the variations in the direction of the motion are always trivial, i.e. the trajectory remains the same up to parametrizations, one can reduce the space of variation to an $(n-1)$-dimensional one.

This idea is formalized by considering a reduction of the Jacobi curve in a smaller symplectic space. As we show in the next section, this is a natural consequence of the homogeneity of the sub-Riemannian Hamiltonian.

*Remark* 15.11. This procedure was already exploited in Section 8.11, obtained by a direct argument via Proposition 8.42. Indeed one can recognize that the procedure that reduced the equation for conjugate points of one dimension corresponds exactly to the reduction by homogeneity of the Jacobi curve associated to the problem.

We start with a technical lemma, whose proof is left as an exercise.

**Lemma 15.12.** *Let $\Sigma = \Sigma_1 \oplus \Sigma_2$ be a splitting of the symplectic space, with $\sigma = \sigma_1 \oplus \sigma_2$. Let $\Lambda_i \in L(\Sigma_i)$ and define the curve $\Lambda(t) := \Lambda_1(t) \oplus \Lambda_2(t) \in L(\Sigma)$. Then one has the splittings:*

$$\underline{\dot{\Lambda}}(t) = \underline{\dot{\Lambda}}_1(t) \oplus \underline{\dot{\Lambda}}_2(t),$$
$$R_\Lambda(t) = R_{\Lambda_1}(t) \oplus R_{\Lambda_2}(t).$$

Consider now a Jacobi curve associated with $\lambda \in T^*M$:

$$J_\lambda(t) = e_*^{-t\vec{H}} \mathcal{V}_{\lambda(t)}, \qquad \mathcal{V}_\lambda = T_\lambda(T^*_{\pi(\lambda)}M).$$

Denote by $\delta_\alpha : T^*M \to T^*M$ the fiberwise dilation $\delta_\alpha(\lambda) = \alpha\lambda$, where $\alpha > 0$ .

**Definition 15.13.** The *Euler vector field* $\mathfrak{e} \in \mathrm{Vec}(T^*M)$ is the vertical vector field defined by

$$\mathfrak{e}(\lambda) = \left.\frac{d}{ds}\right|_{s=1} \delta_s(\lambda), \qquad \lambda \in T^*M.$$

It is easy to see that in canonical coordinates $(x,\xi)$ it satisfies $\mathfrak{e} = \sum_{i=1}^n \xi_i \frac{\partial}{\partial \xi_i}$ and the following identity holds

$$e^{t\mathfrak{e}}\lambda = e^t\lambda, \qquad \text{i.e. } e^{t\mathfrak{e}}(\xi, x) = (e^t\xi, x).$$

**Exercise 15.14.** Prove that the Euler vector field is characterized by the identity

$$i_{\mathfrak{e}}\,\sigma = s, \qquad\qquad s = \text{ Liouville 1-form in } T^*M.$$

**Lemma 15.15.** *We have the identity $e_*^{-t\vec{H}}\mathfrak{e} = \mathfrak{e} - t\vec{H}$. In particular $[\vec{H}, \mathfrak{e}] = -\vec{H}$.*

*Proof.* The homogeneity property (8.56) of the Hamiltonian can be rewritten as follows

$$e^{t\vec{H}}(\delta_s\lambda) = \delta_s(e^{st\vec{H}}(\lambda)), \qquad \forall\, s,t > 0.$$

Applying $\delta_{-s}$ to both sides and changing $t$ into $-t$ one gets the identity

$$\delta_{-s} \circ e^{-t\vec{H}} \circ \delta_s = e^{-st\vec{H}}. \tag{15.9}$$

Computing the $2^{nd}$ order mixed partial derivative at $(t,s) = (0,1)$ in (15.9) one gets, by (2.30), that $[\vec{H}, \mathfrak{e}] = -\vec{H}$. Thus, by (2.34) we have $e_*^{-t\vec{H}}\mathfrak{e} = \mathfrak{e} - t\vec{H}$, since every higher order commutator vanishes.

□

**Proposition 15.16.** *The subspace $\widetilde{\Sigma} = \mathrm{span}\{\mathfrak{e}, \vec{H}\}$ is invariant under the action of the Hamiltonian flow. Moreover $\{\mathfrak{e}, \vec{H}\}$ is a Darboux basis on $\widetilde{\Sigma} \cap H^{-1}(1/2)$.*

*Proof.* The fact that $\widetilde{\Sigma}$ is an invariant subspace is a consequence of the identities

$$e_*^{-t\vec{H}}\mathfrak{e} = \mathfrak{e} - t\vec{H}, \qquad e_*^{-t\vec{H}}\vec{H} = 0.$$

Moreover, on the level set $H^{-1}(1/2)$, we have by homogeneity of $H$ w.r.t. $p$:

$$\sigma(\mathfrak{e}, \vec{H}) = \mathfrak{e}(H) = \frac{d}{dt}\bigg|_{t=0} H(e^{t\mathfrak{e}}(p,x)) = p\frac{\partial H}{\partial p} = 2H = 1. \tag{15.10}$$

It follows that $\{\mathfrak{e}, \vec{H}\}$ is a Darboux basis for $\widetilde{\Sigma}$. $\square$

In particular we can consider the the symplectic splitting $\Sigma = \widetilde{\Sigma} \oplus \widetilde{\Sigma}^{\angle}$.

**Exercise 15.17.** Prove the following intrinsic characterization of the skew-orthogonal to $\widetilde{\Sigma}$:

$$\widetilde{\Sigma}^{\angle} = \{\xi \in T_\lambda^*(T^*M) : \langle d_\lambda H, \xi\rangle = \langle s_\lambda, \xi\rangle = 0\}.$$

The assumptions of Lemma 15.12 are satisfied.

**Definition 15.18.** The *reduced Jacobi curve* is defined as follows

$$\widehat{J}_\lambda(t) := J_\lambda(t) \cap \widetilde{\Sigma}^{\angle}. \tag{15.11}$$

Notice that, if we put $\widehat{\mathcal{V}}_\lambda := \mathcal{V}_\lambda \cap T_\lambda H^{-1}(1/2)$, we get

$$\widehat{J}_\lambda(0) = \widehat{\mathcal{V}}_\lambda, \qquad \widehat{J}_\lambda(t) = e_*^{-t\vec{H}}\widehat{\mathcal{V}}_\lambda.$$

Moreover we have the splitting

$$J_\lambda(t) = \widehat{J}_\lambda(t) \oplus \mathbb{R}(\mathfrak{e} - t\vec{H}).$$

We stress again that $\widehat{J}_\lambda(t)$ is a curve of $(n-1)$-dimensional Lagrangian subspaces in the $(2n-2)$-dimensional vector space $\widetilde{\Sigma}^{\angle}$.

**Exercise 15.19.** With the notation above

(i) Show that the curvature of the curve $J_\lambda(t) \cap \widetilde{\Sigma}$ in $L(\widetilde{\Sigma})$ is always zero.

(ii) Prove that $J_\lambda(0) \cap J_\lambda(s) \neq \{0\}$ if and only if $\widehat{J}_\lambda(0) \cap \widehat{J}_\lambda(s) \neq \{0\}$.

## 15.4 Bibliographical note

The theory of Jacobi curves have been developed in control theory starting from the papers [AG97, Agr98b]. For a recent survey on Jacobi curves and its relation to a canonical connection see [BR17]. Jacobi curves are particular curves in Lagrange Grassmannian; we refer to the Bibliographical note of Chapter 14 for further reading.

# Chapter 16

# Riemannian curvature

On a manifold, in general there is no canonical way to identify tangent spaces (or, more generally, fibers of a vector bundle) at different points. Thus, one has to expect that a notion of derivative for vector fields (or sections of a vector bundle), depends on a certain choice. The additional structure required to correctly define these notions is the one of connection.

In this chapter we introduce *Ehresmann connections*, with the associated notions of parallel transport and curvature. We then specify these notions in the case of a Riemannian manifold, where one can find a canonical connection associated with the metric structure, called Levi-Civita connection. We then explain how this connection is related to the theory of Jacobi curves developed in the previous chapter.

## 16.1 Ehresmann connection

Given a smooth fiber bundle $E$, with base $M$ and canonical projection $\pi : E \to M$, we denote by $E_q = \pi^{-1}(q)$ the fiber at the point $q \in M$. The *vertical distribution* is by definition the collection of subspaces in $TE$ that are tangent to the fibers

$$\mathcal{V} = \{\mathcal{V}_z\}_{z \in E}, \qquad \mathcal{V}_z := \ker \pi_*|_z = T_z E_{\pi(z)} \subset T_z E.$$

**Definition 16.1.** Let $E$ be a smooth fiber bundle. An *Ehresmann connection* on $E$ is a smooth vector distribution $\mathcal{H}$ in $E$ satisfying

$$\mathcal{H} = \{\mathcal{H}_z\}_{z \in E}, \qquad T_z E = \mathcal{V}_z \oplus \mathcal{H}_z.$$

Notice that $\mathcal{V}$, being the kernel of the pushforward $\pi_*$, is canonically associated with the fibre bundle. Defining a connection means exactly to define a canonical complement to this distribution. For this reason $\mathcal{H}$ is also called *horizontal distribution*.

**Definition 16.2.** Let $X \in \mathrm{Vec}(M)$. The *horizontal lift* of $X$ is the unique vector field $\nabla_X \in \mathrm{Vec}(E)$ such that

$$\nabla_X(z) \in \mathcal{H}_z, \qquad \pi_* \nabla_X = X, \qquad \forall\, z \in E. \tag{16.1}$$

The uniqueness follows from the fact that $\pi_{*,z} : T_z E \to T_{\pi(z)} M$ is an isomorphism when restricted to $\mathcal{H}_z$. Indeed $\pi_{*,z}$ is a surjective linear map with $\ker \pi_{*,z} = \mathcal{V}_z$.

In the following, with an abuse of terminology, we will identify the Ehresmann connection on $E$ with $\nabla$.

Given a smooth curve $\gamma : [0,T] \to M$ on the manifold $M$, the connection $\nabla$ let us to define a *parallel transport* along $\gamma$, i.e., a way to identify elements belonging to fibers of $E$ at different points of the curve. Let $X_t$ be a smooth nonautonomous vector field defined on a neighborhood of $\gamma$, that is an extension of the velocity vector field of the curve

$$\dot{\gamma}(t) = X_t(\gamma(t)), \qquad \forall\, t \in [0,T].$$

Notice that this is always possible with a (possibly non autonomous) vector field.

Then consider the nonautonomous vector field $\nabla_{X_t} \in \mathrm{Vec}(E)$ defined by its lift.

**Definition 16.3.** Let $\gamma : [0,T] \to M$ be a smooth curve and let $0 \leq t_0 < t_1 \leq T$. The *parallel transport* along $\gamma$ is the map $\Phi$ defined by the flow of $\nabla_{X_t}$

$$\Phi_{t_0,t_1} : E_{\gamma(t_0)} \to E_{\gamma(t_1)}, \qquad \Phi_{t_0,t_1} := \overrightarrow{\exp} \int_{t_0}^{t_1} \nabla_{X_s} ds \tag{16.2}$$

In full generality, one needs some extra assumption on the vector field to ensure that the flow (16.2) is well defined, even for small times. Indeed the time of existence of the solution of the corresponding ODE may depend on the initial point chosen the fiber. For instance if the fibers are compact, then the flow is complete and it is possible to define (16.2) for every $t_0 < t_1$.

**Exercise 16.4.** Show that the parallel transport map (16.2) is well-defined, i.e., it sends fibers to fibers and does not depend on the extension of the vector field $X_t$. (Hint: consider two extensions and use the existence and uniqueness of the flow.)

### 16.1.1 Curvature of an Ehresmann connection

Assume that $\pi : E \to M$ is a smooth fiber bundle and let $\nabla$ be an Ehresmann connection on $E$, defining the splitting $E = \mathcal{V} \oplus \mathcal{H}$. Given an element $z \in E$ we will also denote by $z_{hor}$ (resp. $z_{ver}$) its projection on the horizontal (resp. vertical) subspace at that point.

The commutator of two vertical vector field is always vertical. The curvature operator associated with the connection checks if the same holds true for two horizontal vector fields.

**Definition 16.5.** Let $E$ be a smooth fiber bundle and $\nabla$ a connection on $E$. Let $X, Y \in \mathrm{Vec}(M)$ and define

$$R(X,Y) := [\nabla_X, \nabla_Y]_{ver} \tag{16.3}$$

The operator $R$ is called the *curvature* of the connection $\nabla$.

Notice that, given a vector field on $E$, its horizontal part coincides, by definition, with the lift of its projection. In particular

$$[\nabla_X, \nabla_Y]_{hor} = \nabla_{[X,Y]}, \qquad (\text{i.e.,} \quad \pi_*[\nabla_X, \nabla_Y] = [X,Y]).$$

Hence $R(X,Y)$ computes the nontrivial part of the bracket between the lift of $X$ and $Y$. Moreover $R(X,Y) = 0$ for every $X, Y$ if and only if the horizontal distribution $\mathcal{H}$ is involutive.

Due to the previous identity, the curvature $R(X,Y)$ is also rewritten in the following more classical way

$$R(X,Y) = [\nabla_X, \nabla_Y] - \nabla_{[X,Y]} \tag{16.4}$$
$$= \nabla_X \nabla_Y - \nabla_Y \nabla_X - \nabla_{[X,Y]}. \tag{16.5}$$

Next we show that $R$ is a tensor, i.e., the value of $R(X,Y)$ at a point depends only on the value of $X$ and $Y$ at the point itself.

**Proposition 16.6.** *$R$ is skew-symmetric and $C^\infty(M)$-linear.*

*Proof.* The skew-symmetry is immediate from the formula (16.5). Also notice that, by definition of lift of a vector field, one has the following identities $\nabla_{aX} = a\nabla_X$ and $\nabla_X(\pi^*a) = Xa$ for every $a \in C^\infty(M)$. Notice that $\pi^*a := a \circ \pi$ is the function $a \in C^\infty(M)$ seen as an element of $C^\infty(E)$ that is constant on fibers. In what follows we simply write $a$ for $\pi^*a$, the meaning being clear by the context.

Next we prove that $R$ is $C^\infty(M)$-linear. By skew-symmetry, it is sufficient to prove that $R$ is linear in the first argument, namely that

$$R(aX,Y) = aR(X,Y), \qquad \text{where} \quad a \in C^\infty(M). \tag{16.6}$$

Applying the definition of $\nabla$ and the Leibniz rule for the Lie bracket one gets

$$\begin{aligned} R(aX,Y) &= [\nabla_{aX}, \nabla_Y] - \nabla_{[aX,Y]} \\ &= a[\nabla_X, \nabla_Y] - (\nabla_Y a)\nabla_X - \nabla_{a[X,Y]-(Ya)X} \\ &= a[\nabla_X, \nabla_Y] - (Ya)\nabla_X - a\nabla_{[X,Y]} + (Ya)\nabla_X \\ &= aR(X,Y). \end{aligned}$$

□

### 16.1.2 Linear Ehresmann connections

Assume now that $\pi : E \to M$ is a *vector* bundle on $M$ (i.e., each fiber $E_q = \pi^{-1}(q)$ has a structure of vector space). In this case it is natural to introduce the notion of *linear* Ehresmann connection $\nabla$ on $E$.

*Remark* 16.7. For a vector bundle $\pi : E \to M$, the base manifold $M$ can be considered immersed in $E$ as the zero section (see also Example 2.52). The "dual" version of this identification is the inclusion $i : C^\infty(M) \to C^\infty(E)$. Indeed any function in $C^\infty(M)$ can be considered as a functions in $C^\infty(E)$ which is constant on fibers, i.e. more precisely $a \in C^\infty(M) \mapsto \pi^*a = a \circ \pi \in C^\infty(E)$.

Following the notation introduced at the beginning of Chapter 4, the image of the inclusion $i : C^\infty(M) \to C^\infty(E)$ discussed in the previous remark, is denoted by $C^\infty_{\mathtt{cst}}(E)$.

**Exercise 16.8.** Let $V \in \mathrm{Vec}(E)$. Show that there exists $X \in \mathrm{Vec}(M)$ such that $V = \nabla_X$ if and only if $V$, seen as a differential operator on $C^\infty(E)$, it maps the subspace $C^\infty_{\mathtt{cst}}(M)$ into itself.

Analogously, given a vector bundle $\pi : E \to M$, we denote by $C^\infty_{\mathtt{lin}}(E)$ the set of smooth functions on $E$ that are linear on fibers. In light of the above discussion, it is natural to introduce the following definition.

**Definition 16.9.** A *linear connection* on a vector bundle $E$ on the base $M$ is an Ehresmann connection $\nabla$ such that the lift $\nabla_X$ of a vector field $X \in \mathrm{Vec}(M)$ satisfies the following property: for every $a \in C^\infty_{\mathrm{lin}}(E)$ it holds $\nabla_X a \in C^\infty_{\mathrm{lin}}(E)$.

Given a local basis of vector fields $X_1, \ldots, X_n$ on $M$ we can build the corresponding dual coordinates $(h_1, \ldots, h_n)$ on the fibers of $E$ by introducing the functions $h_i(z) = \langle z, X_i(q) \rangle$, where $q = \pi(z)$. In this way

$$E \simeq \{(q,h) \mid q \in M, h \in \mathbb{R}^n\},$$

and the tangent space to $E$ is split as $T_zE \simeq T_qM \oplus T_zE_q$. A connection on $E$ is determined by the lift of the vector fields $X_i$, for $i = 1, \ldots, n$, on the base manifold (recall that $\pi_* \nabla_{X_i} = X_i$)

$$\nabla_{X_i} = X_i + \sum_{j=1}^{n} a_{ij}(q,h)\partial_{h_j}, \qquad i = 1, \ldots, n, \tag{16.7}$$

where $a_{ij} \in C^\infty(E)$ are suitable smooth functions. Then $\nabla$ is linear if and only if for every $i, j$ the function $a_{ij}$ is linear, namely $a_{ij}(q,h) = \sum_{k=1}^n \Gamma_{ij}^k(q) h_k$ is linear with respect to $h$.

For a linear connection, the smooth functions $\Gamma_{ij}^k$ defined above are called the *Christoffel symbols* of the connection $\nabla$, associated with the frame $X_1, \ldots, X_n$ on $M$.

**Exercise 16.10.** Fix a frame $X_1, \ldots, X_n$ on $M$ and let $\Gamma_{ij}^k$ be the Christoffel symbols of a connection $\nabla$ associated with this frame. Let $\gamma$ be a smooth curve on the manifold such that $\dot\gamma(t) = \sum_{i=1}^n v_i(t) X_i(\gamma(t))$. Show that the differential equation $\dot\xi(t) = \nabla_{\dot\gamma(t)}\xi(t)$ for the parallel transport along $\gamma$, is written as

$$\dot\xi_j(t) = \sum_{i,k=1}^{n} \Gamma_{ij}^k(\gamma(t)) v_i(t) \xi_k(t)$$

where $(\xi_1, \ldots, \xi_n)$ denotes the vertical coordinates of $\xi$, namely $\xi_j(t) := h_j(\xi(t))$.

For a linear connection, the parallel transport is defined through a first order linear (nonautonomous) ODE. The existence of the flow is then guaranteed by classical results form ODE theory. Moreover, the map $\Phi_{t_0,t_1} : E_{\gamma(t_0)} \to E_{\gamma(t_1)}$ is a linear transformation between fibers.

### 16.1.3 Covariant derivative and torsion for linear connections

Once a linear connection on a linear vector bundle $E$ is given, we have a well-defined linear parallel transport map

$$\Phi_{t_0,t_1} := \overrightarrow{\exp} \int_{t_0}^{t_1} \nabla_{X_s} ds : E_{\gamma(t_0)} \to E_{\gamma(t_1)}, \qquad \text{for } 0 \leq t_0 < t_1 \leq T. \tag{16.8}$$

If we consider the dual map of the parallel transport one can naturally introduce a non autonomous linear flow on the dual bundle (notice the position of $t_0$ and $t_1$)

$$\Phi^*_{t_1,t_0} := \left( \overrightarrow{\exp} \int_{t_1}^{t_0} \nabla_{X_s} ds \right)^* : E^*_{\gamma(t_0)} \to E^*_{\gamma(t_1)}, \qquad \text{for } 0 \leq t_0 < t_1 \leq T. \tag{16.9}$$

The infinitesimal generator of this "adjoint" flow defines a linear parallel transport, hence a linear connection, on the dual bundle $E^*$ (that is the vector bundle $E^* \to M$ whose fibers are dual vector spaces to the fibers of $E \to M$).

In what follows we restrict our attention to the case of the vector bundle $E = T^*M$ and we assume that a linear connection $\nabla$ on $T^*M$ is given. Notice that, by the above discussion, all the constructions can be equivalently performed on the dual bundle $E^* = TM$.

For a vector field $Y \in \mathrm{Vec}(M)$ we denote with $Y^* \in C^\infty(T^*M)$ the function

$$Y^*(\lambda) = \langle \lambda, Y(q) \rangle, \quad q = \pi(\lambda),$$

namely the smooth function on $E$ associated with $Y$, and that is linear on fibers. This identification between vector fields on $M$ and linear functions on $T^*M$ permits us to define the *covariant derivative* of vector fields.

**Definition 16.11.** Let $X, Y \in \mathrm{Vec}(M)$. We define $\nabla_X Y = Z$ if $\nabla_X Y^* = Z^*$ with $Z \in \mathrm{Vec}(M)$.

Notice that the definition is well-posed since $\nabla$ is linear, hence $\nabla_X Y^*$ is a linear function and there exists $Z \in \mathrm{Vec}(M)$ such that $\nabla_X Y^* = Z^*$.[1]

**Lemma 16.12.** *Let $\{X_1, \dots, X_n\}$ be a local frame on $M$. Then $\nabla_{X_i} X_j = \Gamma_{ij}^k X_k$, where $\Gamma_{ij}^k$ are the Christoffel symbols of the connection $\nabla$ associated with the frame.*

*Proof.* Let us prove the statement in the coordinates dual to the reference frame. In these coordinates (see also (16.7)) the linear connection is specified by the lifts

$$\nabla_{X_i} = X_i + \Gamma_{ij}^k h_k \partial_{h_j}, \qquad \text{where} \qquad h_j(\lambda) = \langle \lambda, X_j \rangle. \tag{16.10}$$

Notice that $X_j^* = h_j$. Hence from (16.10) it follows $\nabla_{X_i} X_j^* = \Gamma_{ij}^k X_k^*$, and the lemma is proved. $\square$

We now introduce the torsion tensor of a linear connection on $T^*M$. As usual, $\sigma$ denotes the canonical symplectic structure on $T^*M$.

**Definition 16.13.** The *torsion* of a linear connection $\nabla$ is the map $T : \mathrm{Vec}(M) \times \mathrm{Vec}(M) \to \mathrm{Vec}(M)$ defined by the identity

$$T(X,Y)^* = \sigma(\nabla_X, \nabla_Y), \qquad \forall\, X, Y \in \mathrm{Vec}(M). \tag{16.11}$$

It is easy to check that $T$ is $C^\infty(M)$-linear, i.e., the value of $T(X,Y)$ at a point $q \in M$ depends only on the values of $X, Y$ at the point $q$ itself.

The torsion computes how much the horizontal distribution $\mathcal{H}$ is far from being Lagrangian. In particular $\mathcal{H}$ is Lagrangian if and only if $T$ is identically zero. The classical formula for the torsion tensor, in terms of covariant derivatives, is recovered in the following lemma.

**Lemma 16.14.** *The torsion tensor satisfies the following identity for every $X, Y \in \mathrm{Vec}(M)$*

$$T(X,Y) = \nabla_X Y - \nabla_Y X - [X,Y]. \tag{16.12}$$

[1]There is no confusion in the notation above since, by definition, $\nabla_X$ it is well defined when applied to smooth functions on $T^*M$. Whenever it is applied to a vector field, we follow the aforementioned convention.

*Proof.* The statement is equivalent to the identity $T(X,Y)^* = \nabla_X Y^* - \nabla_Y X^* - [X,Y]^*$. Notice that we can write $X^*(\lambda) = \langle \lambda, X\rangle = \langle s_\lambda, \nabla_X\rangle$, where $s \in \Lambda^1(T^*M)$ is the Liouville 1-form defined by $s_\lambda = \lambda \circ \pi_*$.

In particular $\sigma = ds$, and applying Cartan's formula (4.84)

$$\begin{aligned} T(X,Y)^* &= ds(\nabla_X, \nabla_Y) \\ &= \nabla_X \langle s, \nabla_Y\rangle - \nabla_Y \langle s, \nabla_X\rangle - \langle s, [\nabla_X, \nabla_Y]\rangle \\ &= \nabla_X \langle s, \nabla_Y\rangle - \nabla_Y \langle s, \nabla_X\rangle - \left\langle s, \nabla_{[X,Y]}\right\rangle \\ &= \nabla_X Y^* - \nabla_Y X^* - [X,Y]^*, \end{aligned}$$

where in the second equality we used that $\langle s, [\nabla_X, \nabla_Y]\rangle = \langle s, [\nabla_X, \nabla_Y]_{hor}\rangle = \left\langle s, \nabla_{[X,Y]}\right\rangle$ since the Liouville form by definition depends only on the horizontal part of the vector. □

**Exercise 16.15.** Show that a linear connection $\nabla$ on a vector bundle $E$ satisfies the Leibniz rule

$$\nabla_X(aY) = a\nabla_X Y + (Xa)Y, \qquad \text{for every } a \in C^\infty(M).$$

## 16.2 Riemannian connection

In this section we introduce the Levi-Civita connection on a Riemannian manifold $M$ by defining an Ehresmann connection on $T^*M$ through the Jacobi curve approach.

Recall that every Jacobi curve associated with a trajectory on a Riemannian manifold is regular thanks to Proposition 15.2. Moreover, as showed in Section 14.2, every regular curve in the Lagrangian Grassmannian admits a derivative curve, which defines a canonical complement to the curve itself. Hence we can introduce an Ehresmann connection on $T^*M$ by defining at every point $\lambda \in T^*M$ the canonical complement to the Jacobi curve defined at $\lambda$.

**Definition 16.16.** The *Levi-Civita connection* on $T^*M$ is the Ehresmann connection $\mathcal{H}$ is defined by

$$\mathcal{H}_\lambda = J^\circ_\lambda(0), \qquad \forall\, \lambda \in T^*M,$$

where $J_\lambda(t)$ denotes the Jacobi curve defined at $\lambda \in T^*M$ and $J^\circ_\lambda(t)$ denotes its derivative curve.

The next proposition describes the main properties of Levi-Civita connection as an Ehresmann connection on $T^*M$.

**Proposition 16.17.** *The Levi-Civita connection satisfies the following properties:*

*(i) it is a linear connection,*

*(ii) it has vanishing torsion,*

*(iii) it is metric-preserving, i.e., $\nabla_X H = 0$ for each vector field $X \in \mathrm{Vec}(M)$.*

*Proof.* (i). It is enough to prove that the connection $\mathcal{H}_\lambda$ is 1-homogeneous, namely

$$\mathcal{H}_{c\lambda} = \delta_{c*}\mathcal{H}_\lambda, \qquad \forall\, c > 0. \tag{16.13}$$

Indeed if (16.13) holds true, the functions $a_{ij} \in C^\infty(T^*M)$ defining the connection with respect to some frame (see (16.7)) are smooth and 1-homogeneous, hence linear (see also Exercise 16.20).

To prove (16.13), we need the following auxiliary result.

**Lemma 16.18.** *Let $J_\lambda(t)$ be the regular Jacobi curve associated to $\lambda \in T^*M$. Then we have for all $t \geq 0$:*

*(a)* $J_{c\lambda}(t) = \delta_{c*} J_\lambda(ct)$,

*(b)* $J^\circ_{c\lambda}(t) = \delta_{c*} J^\circ_\lambda(ct)$.

*Proof of Lemma 16.18.* (a). The differential of the dilation on the fibers $\delta_c : T^*M \to T^*M$ at a point $\lambda \in T^*M$ is a linear map satisfying the identity $\delta_{c*}(T_\lambda(T^*_qM)) = T_{c\lambda}(T^*_qM)$. Moreover the following identity holds

$$e^{t\vec{H}} \circ \delta_c = \delta_c \circ e^{ct\vec{H}}, \qquad \forall\, c > 0. \tag{16.14}$$

Hence one obtains

$$\begin{aligned} J_{c\lambda}(t) &= e^{-t\vec{H}}_*(T_{c\lambda}(T^*_qM)) \\ &= e^{-t\vec{H}}_* \circ \delta_{c*}(T_\lambda(T^*_qM)) \\ &= \delta_{c*} \circ e^{-ct\vec{H}}_*(T_\lambda(T^*_qM)) \\ &= \delta_{c*} J_\lambda(ct), \end{aligned}$$

where in the third equality we used the differential of (16.14). Notice that $\delta_{c*}$ does not preserve the symplectic structure but it preserves Lagrangian subspaces, since $\delta_c^*\sigma = c\sigma$.

To prove (b), recall that, given a regular curve $\Lambda(t)$ in the Lagrange Grassmannian, for $t \neq t_0$ small enough the subspace $\Lambda^\circ(t)$ is transversal to $\Lambda(t_0)$, hence belongs to the coordinate chart $\Lambda(t_0)^\pitchfork$. The space $\Lambda(t_0)^\pitchfork$ has the structure of affine space and $\Lambda^\circ(t_0)$ is, by construction, the free term (i.e., the zero-order term in the expansion with respect to $t - t_0$) in the Laurent expansion of the curve in this chart.

Notice that the map $\delta_{c*}$ sends $J_\lambda(ct_0)$ to $J_{c\lambda}(t_0)$ and the chart $J_\lambda(ct_0)^\pitchfork$ to $J_{c\lambda}(t_0)^\pitchfork$. Moreover, being a linear map between the tangent space to fibers, $\delta_{c*}$ is compatible with the affine structure on the charts. Finally, using (a), $\delta_{c*}$ sends the curve $t \mapsto J_\lambda(ct)$ to $t \mapsto J_{c\lambda}(t)$, and since the time reparametrization does not change the zero order term, it follows from the construction that $\delta_{c*}$ also sends $J^\circ_\lambda(ct)$ to $J^\circ_{c\lambda}(t)$, proving (b). $\square$

(ii). It is a direct consequence of the fact that $J^\circ_\lambda(0)$ is a Lagrangian subspace of $T_\lambda(T^*M)$ for every $\lambda \in T^*M$, hence the symplectic form vanishes when applied to two horizontal vectors.

(iii). Again, for every $X \in \mathrm{Vec}(M)$, both $\nabla_X$ and $\vec{H}$ are horizontal vector field. Since the horizontal space is Lagrangian

$$\nabla_X H = \sigma(\nabla_X, \vec{H}) = 0.$$

$\square$

**Exercise 16.19** (Proof of Lemma 16.18 in coordinates)**.** Let us fix a coordinate chart on the Lagrange Grassmannian in such a way that $J_\lambda(t) = \{(p, S_\lambda(t)p), p \in \mathbb{R}^n\}$. We follow the notations of Exercice 14.23.

(i) Show that claim (a) of Lemma 16.18 is rewritten as $S_{c\lambda}(t) = \frac{1}{c} S_\lambda(ct)$,

(ii) deduce the identities $B_{c\lambda}(t) = cB_\lambda(ct)$ and $S^\circ(t) = B^{-1}(t) + S(t)$, and prove claim (b).

**Exercise 16.20.** Let $f : \mathbb{R}^n \to \mathbb{R}$ be a smooth function that satisfies $f(\alpha x) = \alpha f(x)$ for every $x \in \mathbb{R}^n$ and $\alpha \geq 0$. Prove that $f$ is linear.

The following theorem says that a connection satisfying the three properties of Proposition 16.17 is unique. It also characterizes the Levi-Civita connection in terms of the structure constants of the Lie algebra defined by an orthonormal frame.

**Theorem 16.21.** *There is a unique Ehresmann connection $\nabla$ satisfying the properties (i), (ii), and (iii) of Proposition 16.17, that is the Levi-Civita connection. Given an orthonormal frame $X_1, \dots, X_n$, its Christoffel symbols are computed by*

$$\Gamma_{ij}^k = \frac{1}{2}(c_{ij}^k - c_{jk}^i + c_{ki}^j), \qquad i, j, k = 1, \dots, n, \tag{16.15}$$

*where $c_{ij}^k$ are smooth functions defined by the identities $[X_i, X_j] = \sum_{k=1}^n c_{ij}^k X_k$, for $i, j = 1, \dots, n$.*

*Proof.* Let $X_1, \dots, X_n$ be a local orthonormal frame for the Riemannian structure and let us consider coordinates $(q, h)$ in $T^*M$, where the fiberwise coordinates $h = (h_1, \dots, h_n)$ are dual to the orthonormal frame. From the linearity of the connection it follows that there exist smooth functions $\Gamma_{ij}^k : M \to \mathbb{R}$ such that

$$\nabla_{X_i} = X_i + \sum_{j=1}^n \Gamma_{ij}^k h_k \partial_{h_j}, \qquad i = 1, \dots, n. \tag{16.16}$$

In particular $\nabla_{X_i} X_j = \Gamma_{ij}^k X_k$. In these coordinates the Hamiltonian vector field associated with the Riemannian Hamiltonian $H = \frac{1}{2}\sum_{i=1}^n h_i^2$ reads

$$\vec{H} = \sum_{i=1}^n h_i \vec{h}_i = \sum_{i=1}^n h_i X_i + \sum_{i,j,k=1}^n c_{ij}^k h_i h_k \partial_{h_j}, \tag{16.17}$$

where we used the identity

$$\vec{h}_i = X_i + \sum_{j=1}^n \{h_i, h_j\} \partial_{h_j} = X_i + \sum_{j,k=1}^n c_{ij}^k h_k \partial_{h_j}. \tag{16.18}$$

Moreover the symplectic form $\sigma$ is written in terms of the basis of 1-forms $\nu_1, \dots, \nu_n$ that is dual to $X_1, \dots, X_n$ as follows

$$\sigma = \sum_{k=1}^n dh_k \wedge \nu_k - \sum_{i,j,k=1}^n c_{ij}^k h_k \nu_i \wedge \nu_k.$$

Since the horizontal space is Lagrangian, one has the relations

$$0 = \sigma(\nabla_{X_i}, \nabla_{X_j}) = \sum_{k=1}^n (\Gamma_{ij}^k - \Gamma_{ji}^k - c_{ij}^k) h_k, \qquad \forall\, i, j = 1, \dots, n,$$

hence $c_{ij}^k = \Gamma_{ij}^k - \Gamma_{ji}^k$ for all $i, j, k = 1, \dots, n$. Moreover the connection is metric, i.e., it satisfies

$$0 = \nabla_{X_i} H = \sum_{j,k=1}^n \Gamma_{ij}^k h_k h_j, \qquad \forall\, i = 1, \dots, n.$$

The last identity implies that $\Gamma_{ij}^k$ is skew-symmetric in the following sense: for every $i, j, k = 1, \ldots, n$ one has $\Gamma_{ij}^k = -\Gamma_{ik}^j$. Combining the two identities we obtained, one gets

$$\begin{aligned} c_{ij}^k - c_{jk}^i + c_{ki}^j &= (\Gamma_{ij}^k - \Gamma_{ji}^k) - (\Gamma_{jk}^i + \Gamma_{kj}^i) + (\Gamma_{ki}^j - \Gamma_{ik}^j) \\ &= \Gamma_{ij}^k - \Gamma_{ik}^j = 2\Gamma_{ij}^k. \end{aligned}$$

□

Recall that the vector field $\vec{H}$ is automatically horizontal for the connection. This is recovered also from the previous computations as follows.

**Corollary 16.22.** *We have* $\vec{H} = \sum_{i=1}^n h_i \nabla_{X_i}$.

*Proof.* Notice that using the identities for $i, j, k = 1, \ldots, n$,

$$c_{ij}^k = \Gamma_{ij}^k - \Gamma_{ji}^k, \qquad \Gamma_{ij}^k = -\Gamma_{ik}^j \tag{16.19}$$

one has, for $j = 1, \ldots, n$,

$$\sum_{i,k=1}^n c_{ij}^k h_i h_k = \sum_{i,k=1}^n \Gamma_{ij}^k h_i h_k.$$

Hence comparing (16.16) and (16.17), one obtains the claim. □

Let $X, Y, Z, W \in \mathrm{Vec}(M)$. We set $R(X,Y)Z = W$ if $R(X,Y)Z^* = W^*$.

**Proposition 16.23** (Bianchi identity)**.** *For every $X, Y, Z \in \mathrm{Vec}(M)$ the following identity holds*

$$R(X,Y)Z + R(Y,Z)X + R(Z,X)Y = 0. \tag{16.20}$$

*Proof.* We will show that (16.20) is a consequence of the Jacobi identity for vector fields (2.35). Using the fact that $\nabla$ is a torsion-free connection we can write

$$\begin{aligned} [X,[Y,Z]] &= \nabla_X[Y,Z] - \nabla_{[Y,Z]}X \\ &= \nabla_X\nabla_Y Z - \nabla_X\nabla_Z Y - \nabla_{[Y,Z]}X, \\ [Z,[X,Y]] &= \nabla_Z\nabla_X Y - \nabla_Z\nabla_Y X - \nabla_{[X,Y]}Z, \\ [Y,[Z,X]] &= \nabla_Y\nabla_Z X - \nabla_Y\nabla_X Z - \nabla_{[Z,X]}Y, \end{aligned}$$

Then, adding these identities and using (2.35), one gets

$$\begin{aligned} 0 &= [X,[Y,Z]] + [Y,[Z,X]] + [Z,[X,Y]] \\ &= \nabla_X\nabla_Y Z - \nabla_X\nabla_Z Y - \nabla_{[Y,Z]}X \\ &\quad + \nabla_Z\nabla_X Y - \nabla_Z\nabla_Y X - \nabla_{[X,Y]}Z \\ &\quad\quad + \nabla_Y\nabla_Z X - \nabla_Y\nabla_X Z - \nabla_{[Z,X]}Y \\ &= R(X,Y)Z + R(Y,Z)X + R(Z,X)Y. \end{aligned}$$

□

**Exercise 16.24** (second Bianchi identity)**.** Prove that for every $X, Y, Z, W \in \mathrm{Vec}(M)$ one has

$$(\nabla_X R)(Y, Z, W) + (\nabla_Y R)(Z, X, W) + (\nabla_Z R)(X, Y, W) = 0.$$

(Hint: Expand the identity $\nabla_{[X,[Y,Z]]+[Y,[Z,X]]+[Z,[X,Y]]} W = 0$.)

*Remark* 16.25*.* The relations (16.19) for the Christoffel symbols implies the following skew-symmetry property: for $X, Y, Z, W \in \mathrm{Vec}(M)$

$$\langle R(X, Y)Z \,|\, W\rangle = -\langle R(X, Y)W \,|\, Z\rangle ,$$

where $\langle \cdot \,|\, \cdot\rangle$ denotes the Riemannian inner product.

Let us introduce the notation

$$R(X, Y, Z, W) := \langle R(X, Y)Z \,|\, W\rangle .$$

Then, the first Bianchi identity (16.20) can be rewritten as follows: for $X, Y, Z, W \in \mathrm{Vec}(M)$ one has

$$R(X, Y, Z, W) + R(Z, X, Y, W) + R(Y, Z, X, W) = 0. \tag{16.21}$$

Moreover, the skew-symmetry properties of the curvature tensor discussed in Proposition 16.6 and Remark 16.25 can be rewritten as follows

$$R(X, Y, Z, W) = -R(Y, X, Z, W), \qquad R(X, Y, Z, W) = -R(X, Y, W, Z). \tag{16.22}$$

**Proposition 16.26.** *For every* $X, Y, Z, W \in \mathrm{Vec}(M)$ *we have* $R(X, Y, Z, W) = R(Z, W, X, Y)$*.*

*Proof.* Using (16.21) four times we can write the identities

$$\begin{aligned}
R(X, Y, Z, W) + R(Z, X, Y, W) + R(Y, Z, X, W) &= 0,\\
R(Y, Z, W, X) + R(W, Y, Z, X) + R(Z, W, Y, X) &= 0,\\
R(Z, W, X, Y) + R(X, Z, W, Y) + R(W, X, Z, Y) &= 0,\\
R(W, X, Y, Z) + R(Y, W, X, Z) + R(X, Y, W, Z) &= 0.
\end{aligned}$$

Summing these identities and using (16.22), one gets $R(X, Z, W, Y) = R(W, Y, X, Z)$. □

**Proposition 16.27.** *Assume that* $R(X, Y, X, W) = 0$ *for every* $X, Y, W \in \mathrm{Vec}(M)$*. Then*

$$R(X, Y, Z, W) = 0 \qquad \forall\, X, Y, Z, W \in \mathrm{Vec}(M).$$

*Proof.* By assumptions and the skew-symmetry properties (16.22) of the Riemann tensor we have that $R(X, Y, Z, W) = 0$ whenever any two of the vector fields coincide. In particular

$$0 = R(X, Y + W, Z, Y + W) = R(X, Y, Z, W) + R(X, W, Z, Y). \tag{16.23}$$

Notice that the two extra terms that should appear developing the left hand side vanish, by assumptions. Then (16.23) can be rewritten as

$$R(X, Y, Z, W) = R(Z, X, Y, W).$$

This means that the quantity $R(X, Y, Z, W)$ is invariant by cyclic permutations of $X, Y, Z$. But the cyclic sum of these terms is zero thanks to (16.21), hence $R(X, Y, Z, W) = 0$. □

From the properties of the Riemann curvature one obtains the following.

**Corollary 16.28.** *There is a well defined map*

$$\overline{R} : \wedge^2 T_q M \to \wedge^2 T_q M, \qquad \overline{R}(X \wedge Y) := R(X, Y).$$

*Moreover $\overline{R}$ is self-adjoint with respect to the scalar product on $\wedge^2 T_q M$ induced by the Riemannian scalar product, namely*

$$\left\langle \overline{R}(X \wedge Y) \mid Z \wedge W \right\rangle = \left\langle X \wedge Y \mid \overline{R}(Z \wedge W) \right\rangle.$$

## 16.3 Relation with Hamiltonian curvature

In this section we compute the curvature of the Jacobi curve associated with a Riemannian geodesic and we describe the relation with the Riemann curvature discussed in the previous section. As we show, the curvature associated with a geodesic can be interpreted as the sectional curvature operator in the direction of the geodesic.

**Definition 16.29.** The *Hamiltonian curvature* at $\lambda \in T^*M$ is the curvature of the Jacobi curve associated with $\lambda$ at $t = 0$, namely

$$\mathcal{R}_\lambda := \mathcal{R}_{J_\lambda(0)} : \mathcal{V}_\lambda \to \mathcal{V}_\lambda. \tag{16.24}$$

**Proposition 16.30.** *Let $\xi \in \mathcal{V}_\lambda$ and $V$ be a smooth vertical vector field extending $\xi$. Then*

$$\mathcal{R}_\lambda(\xi) = -[\vec{H}, [\vec{H}, V]_{hor}]_{ver}(\lambda). \tag{16.25}$$

*Proof.* This is a direct consequence of Proposition 14.31. In fact, recall that the curvature of the Jacobi curve is expressed through the composition

$$\mathcal{R}_\lambda = \underline{\dot{J}}^\circ_\lambda(0) \circ \underline{\dot{J}}_\lambda(0).$$

Moreover, being $J_\lambda(0) = \mathcal{V}_\lambda$ and $J^\circ_\lambda(0) = \mathcal{H}_\lambda$ we have that

$$\pi_{J(0)J^\circ(0)}(\xi) = \xi_{hor}, \qquad \pi_{J^\circ(0)J(0)}(\eta) = \eta_{ver}.$$

Extending the vectors in $J_\lambda(0)$ (resp. $J^\circ_\lambda(0)$) by applying the Hamiltonian vector field (recall the identities $J_\lambda(t) = e^{t\vec{H}}_* J_\lambda(0)$ and $J^\circ_\lambda(t) = e^{t\vec{H}}_* J^\circ_\lambda(0)$), one obtains the following formulas

$$\underline{\dot{J}}_\lambda(0)\xi = [\vec{H}, V]_{hor}, \qquad \underline{\dot{J}}^\circ_\lambda(0)\eta = -[\vec{H}, W]_{ver}$$

where $V$ is a vertical (resp. $W$ is a horizontal) extension of the vector $\xi \in \mathcal{V}_\lambda$ (resp. $\eta \in \mathcal{H}_\lambda$). $\square$

The following homogeneity property of the curvature operator is obtained by choosing $\varphi(t) = ct$, with $c > 0$, in Proposition 14.37. Recall that a rescaling of the covector corresponds to a reparametrization of the trajectory.

**Corollary 16.31.** *For every $c > 0$ we have $\mathcal{R}_{c\lambda} = c^2 \mathcal{R}_\lambda$.*

We end this section by relating the Hamiltonian curvature just introduced with Riemannian curvature. Let us denote by $\iota : TM \to T^*M$ the isomorphism defined by the Riemannian scalar product $\langle \cdot | \cdot \rangle$. In particular $\iota(v) = \lambda$ for $\lambda \in T_q^*M$ and $v \in T_qM$ if $\langle \lambda, w \rangle = \langle v | w \rangle$ for all $w \in T_qM$.

Let us denote $H_q := H|_{T_q^*M}$. Recall that for every $\lambda \in T_q^*M$ the differential $D_\lambda H_q$ is a linear functional on $T_q^*M$, that can be identified with a tangent vector. Hence the differential of $H_q$ can be interpreted as a linear map $DH_q : T_q^*M \to T_qM$. With this identification, the map $DH_q$ is the inverse of the isomorphism $\iota$, as stated in the following lemma.

**Lemma 16.32.** *Fix $q \in M$ and $\lambda \in T_q^*M$. We have the following identities:*

*(i)* $D_\lambda H_q = \iota^{-1}(\lambda)$.

*(ii)* $\vec{H}(\lambda) = \nabla_v$, *where* $v := \iota^{-1}(\lambda)$.

*Proof.* The first identity follows from the fact that the sub-Riemannian Hamiltonian $H$ is quadratic on fibers and $H(\lambda) = \frac{1}{2}\left\langle \lambda, \iota^{-1}(\lambda) \right\rangle$.

To prove the second one, since $\vec{H}$ is an horizontal vector field (cf. Corollary 16.22), it is sufficient to show the identity $\pi_*\vec{H}(\lambda) = v$. We have, for every $\xi \in T_\lambda(T_q^*M)$, that

$$\langle \xi, v \rangle = \left\langle \xi, \iota^{-1}(\lambda) \right\rangle = \langle D_\lambda H, \xi \rangle = \sigma(\xi, \vec{H}(\lambda)) = \left\langle \xi, \pi_*\vec{H}(\lambda) \right\rangle.$$

where we used claim (i), the definition of $\vec{H}$ and Lemma 15.3. Since $\xi$ is arbitrary, this implies $v = \pi_*\vec{H}(\lambda)$. $\square$

**Theorem 16.33.** *The following identity holds for every $X, Y \in T_qM$*

$$R(X,Y)X = \mathcal{R}_{\iota(X)}(\iota(Y)). \tag{16.26}$$

*Proof.* Thanks to (16.25), the right hand side satisfies

$$\mathcal{R}_{\iota(X)}(\iota(Y)) = -[\vec{H}, [\vec{H}, \iota(Y)]_{hor}]_{ver}(\iota(X)). \tag{16.27}$$

To compute $[\vec{H}, \iota(Y)]_{hor}$ we compute first its projection, that is $\pi_*[\vec{H}, \iota(Y)] = -Y$. This implies $[\vec{H}, \iota(Y)]_{hor} = -\nabla_Y$. Hence

$$-[\vec{H}, [\vec{H}, \iota(Y)]_{hor}]_{ver}(\iota(X)) = [\vec{H}, \nabla_Y]_{ver}(\iota(X)) = [\nabla_X, \nabla_Y]_{ver}(\iota(X)) = R(X,Y)X.$$

Here, in the second identity, we used that $[\vec{H}, \nabla_Y]_{ver}(\iota(X))$ depends only on the value of the vector fields at the point and that $\vec{H}(\iota(X)) = \nabla_X$, thanks to Lemma 16.32. Notice also that the last identity follows from the definition of $R(X,Y)$, cf. (16.3). $\square$

The previous result is saying that identifying tangent and cotangent vectors through the Riemannian product, the operator $\mathcal{R}_\lambda$ interpreted as a quadratic form on $T_qM$ is the sectional curvature operator along the tangent vector $v$ associated to $\lambda$.

## 16.4 Comparison theorems for conjugate points

In this section we specify the general results on conjugate points obtained in Section 14.8 to the case of a Riemannian manifold.

Recall that $R(X, Y, Z, W)$ is the full Riemann curvature tensor defined as follows

$$R(X, Y, Z, W) := \langle R(X, Y)Z \,|\, W \rangle .$$

Given two linearly independent vectors $X, Y$, we define the *sectional curvature*

$$\mathrm{Sec}(X, Y) := R(X, Y, Y, X)/(\|X\|^2\|Y\|^2 - \langle X \,|\, Y \rangle^2).$$

Notice that this quantity indeed depends only on the plane $\mathrm{span}\{X, Y\}$. We then define the *Ricci tensor*

$$\mathrm{Ric}(X) := \sum_{i=1}^{n} R(X, X_i, X_i, X),$$

where $X_1, \dots, X_n$ is an orthonormal basis for the Riemannian metric.

We have the following two comparison theorems, which are a direct consequence of Theorem 14.57 and the fact that the Jacobi curve associated with a Riemannian geodesic is regular.

**Theorem 16.34.** *Let $M$ be a complete $n$-dimensional Riemannian manifold. Assume that there exists $k \geq 0$ such that for every pair of unitary tangent vectors $X, Y$ one has*

$$\mathrm{Sec}(X, Y) \leq k.$$

*If a geodesic parametrized by length $\gamma : [0, T] \to M$ has a conjugate point, then $T \geq \pi/\sqrt{k}$. In particular, for $k = 0$, no geodesic has conjugate point.*

**Theorem 16.35.** *Let $M$ be a complete $n$-dimensional Riemannian manifold. Assume that there exists $k \geq 0$ such that for every unitary tangent vector $X$ one has*

$$\mathrm{Ric}(X, X) \geq nk, \tag{16.28}$$

*Then every geodesic parametrized by length $\gamma : [0, T] \to M$ such that $T \leq \pi/\sqrt{k}$ has at least a conjugate point. In particular $M$ is compact and* $\mathrm{diam}(M) \leq \pi/\sqrt{k}$.

*Remark* 16.36. Theorem 16.35 is known as Bonnet-Myers theorem. Actually the assumption (16.28) can be weakened as follows: for every unitary tangent vector $X$

$$\mathrm{Ric}(X, X) \geq (n-1)k, \tag{16.29}$$

since the curvature is always vanishing in the direction of the geodesic. This is tantamount to say that the operator $\mathcal{R}_\lambda$ associated with a Jacobi curve on a Riemannian manifold has a zero eigenvalue, thanks to the splitting due to homogeneity of Section 15.3 (cf. in particular with Exercice 15.19, claim (i)).

**Exercise 16.37** (Hadamard theorem)**.** Let $M$ be a complete Riemannian manifold, simply connected, and such that all sectional curvatures are negative. Prove that for every $q \in M$, the map $\exp_q : T_qM \to M$ is a global diffeomorphism. In particular, $M$ is diffeomorphic to $\mathbb{R}^n$.

*Hint*: The conjugate locus from $q$ is empty thanks to Theorem 16.34 for $k = 0$. Hence the exponential map is a local diffeomorphism at every point. If $\exp_q$ is proper, the one can conclude thanks to Corollary 13.24.

**Exercise 16.38.** Let $M$ be a Riemannian manifold and let us define $\mathrm{Ric}(\lambda)$ at a point $\lambda \in T^*M$ as the trace of the curvature operator $\mathcal{R}_\lambda$. If $X_1, \ldots, X_n$ is a local orthonormal frame and $\iota(v) = \lambda$. Prove that

$$\mathrm{Ric}(\lambda) := \mathrm{tr}(\mathcal{R}_\lambda) = \sum_{i=1}^{n} \sigma_\lambda([\vec{H}, \nabla_{X_i}], \nabla_{X_i}) = \sum_{i=1}^{n} \langle R(v, X_i)v | X_i \rangle = \mathrm{Ric}(v).$$

## 16.5 Locally flat spaces

In this section we want to show that the Riemannian curvature represents the obstruction for a Riemannian manifold to be locally Euclidean. We also show that the Riemann curvature tensor is completely recovered by the Hamiltonian curvature $\mathcal{R}_\lambda$.

**Theorem 16.39.** *A Riemannian manifold $M$ is locally isometric to $\mathbb{R}^n$ if and only if the Riemann curvature tensor vanishes at every point.*

*Proof.* If $M$ is locally isometric to $\mathbb{R}^n$, then its curvature tensor at every point is zero.

Then let us assume that the Riemann tensor $R$ vanishes identically and prove that $M$ is locally Euclidean. We will do that by showing that there exists coordinates such that the Hamiltonian is written as the Hamiltonian associated to the Euclidean structure of $\mathbb{R}^n$.

Since $R$ is identically zero the horizontal distribution (defined by the Levi Civita connection) is involutive. Hence, by Frobenius theorem, there exists a horizontal Lagrangian foliation of $T^*M$, i.e. for each $\lambda \in T^*M$, there exists a leaf $\mathfrak{L}_\lambda$ of the foliation passing through this point that is tangent to the horizontal space $\mathcal{H}_\lambda$. In particular each leaf is transverse to the fiber $T^*_qM$, where $q = \pi(\lambda)$.

Fix a point $q_0 \in M$ and a neighborhood $O_{q_0}$. We can assume that the curvature tensor $R$ is zero at every point of $O_{q_0}$. Define the map

$$\Psi : \pi^{-1}(O_{q_0}) \to T^*_{q_0}M, \qquad \lambda \in \pi^{-1}(O_{q_0}) \mapsto \mathfrak{L}_\lambda \cap T^*_{q_0}M$$

that assigns to each $\lambda$ the intersection of the leaf passing through this point and $T^*_{q_0}M$.

Notice that $\Psi$ is a fiberwise linear and orthogonal transformation, i.e., for every $q \in O_{q_0}$ we have that $\Psi|_{T^*_qM} : T^*_qM \to T^*_{q_0}M$ is linear and satisfies $H(\Psi(\lambda)) = H(\lambda)$. Indeed since the connection is linear then it easily follows that the map $\Psi$ is linear as well. Moreover the foliation is tangent to the horizontal space and $\vec{H}$ is horizontal, hence $H$ is constant on the leaves.

Fix now a basis $\{\nu_1, \ldots, \nu_n\}$ in $T^*_{q_0}M$ that is orthonormal (with respect to the dual metric). Being $\Psi$ linear on fibers, we can write

$$\Psi(\lambda) = \sum_{i=1}^{n} \psi_i(\lambda)\nu_i, \qquad \text{where} \qquad \psi_i(\lambda) = \langle \lambda, X_i(q) \rangle,$$

for a suitable basis of vector fields $X_1, \ldots, X_n$ on a neighborhood of $q_0$ (that we can assume to coincide with $O_{q_0}$, up to restricting it). Moreover $X_1, \ldots, X_n$ is an orthonormal basis since $\Psi$ is an orthogonal map.

We want to show that $[X_i, X_j] = 0$ for all $i, j = 1, \ldots, n$. This follows from the fact that the foliation is Lagrangian. Indeed in terms of this frame, we have the following expression for the tautological and the symplectic form

$$s = \sum_{i=1}^{n} \psi_i \, \nu_i, \qquad \sigma = ds = \sum_{i=1}^{n} d\psi_i \wedge \nu_i + \psi_i d\nu_i. \tag{16.30}$$

Since on each leaf the function $\psi_i$ is constant by construction (in particular $d\psi_i|_{\mathfrak{L}} = 0$), we have that $\sigma|_{\mathfrak{L}} = \sum_i \psi_i \, d\nu_i$. In particular each leaf is Lagrangian if and only if $d\nu_i = 0$ for $i = 1, \ldots, n$. Then, from the Cartan formula, one gets

$$0 = d\nu_i(X_j, X_k) = -\nu_i([X_j, X_k]), \qquad \forall\, i, j, k = 1, \ldots, n.$$

This proves that $[X_i, X_j] = 0$ for each $i, j = 1, \ldots, n$. Hence, there exists coordinates $x_1, \ldots, x_n$ on a possibly smaller neighborhood $O'_{q_0} \subset O_{q_0}$ such that $X_i = \partial/\partial x_i$ and in the corresponding dual coordinates $\{\psi_i\}_{i=1,\ldots,n}$ the Hamiltonian $H$ is written on $\pi^{-1}(O'_{q_0})$ as

$$H(\psi, x) = \frac{1}{2} \sum_{i=1}^{n} \langle \psi, X_i(x) \rangle^2 = \frac{1}{2} \sum_{i=1}^{n} \psi_i^2,$$

which completes the proof. □

To check if a manifold is locally Euclidean is indeed sufficient to compute the Hamiltonian curvature.

**Corollary 16.40.** *$M$ is flat if and only if $\mathcal{R}_\lambda = 0$ for every $\lambda \in T^*M$.*

*Proof.* Assume that $M$ is flat. Then $R$ is identically zero and a fortiori $\mathcal{R}_\lambda = 0$ for every $\lambda \in T^*M$, from (16.26). To prove the converse, recall that $\mathcal{R}_\lambda = 0$ for every $\lambda \in T^*M$ implies, by (16.26), that

$$R(X, Y, X, W) = 0, \qquad \forall\, X, Y, W \in \mathrm{Vec}(M).$$

Then the statement is a consequence of Proposition 16.27. □

**Exercise 16.41.** Prove that the Riemann tensor $R$ is completely determined by $\mathcal{R}$.

## 16.6 Curvature of 2D Riemannian manifolds

In this section we specify the link between Riemannian curvature and Hamiltonian curvature for two-dimensional Riemannian manifolds.

Let $M$ be a two-dimensional Riemannian manifold and let $X_1, X_2$ be an orthonormal frame satisfying

$$[X_1, X_2] = c_1 X_1 + c_2 X_2,$$

for some smooth functions $c_1, c_2$ on $M$.

On each fiber of the cotangent space we fix polar coordinates $(r, \theta)$, where $r^2 = 2H(\lambda)$ and $\theta$ is the angle coordinate in the fiber defined by the identity

$$\langle \lambda, X_1 \rangle = r \cos\theta, \qquad \langle \lambda, X_1 \rangle = r \sin\theta.$$

Denoting $\mu_1, \mu_2$ the dual basis of 1-forms dual to $X_1, X_2$, it follows from the results of Section 4.4.1 that the horizontal distribution defined by the Levi-Civita connection is given by

$$\mathcal{H} = \ker \omega, \qquad \omega = d\theta - c_1\mu_1 - c_2\mu_2.$$

In particular the horizontal lift of the basis $X_1, X_2$ is computed as follows

$$\nabla_{X_1} = X_1 + c_1\partial_\theta, \qquad \nabla_{X_2} = X_2 + c_2\partial_\theta.$$

Then one can compute the vertical vector field

$$\begin{aligned}
R(X_1, X_2) &= \nabla_{X_1}\nabla_{X_2} - \nabla_{X_2}\nabla_{X_1} - \nabla_{[X_1,X_2]} \\
&= (X_1 + c_1\partial_\theta)(X_2 + c_2\partial_\theta) - (X_2 + c_2\partial_\theta)(X_1 + c_1\partial_\theta) - c_1(X_1 + c_1\partial_\theta) - c_2(X_2 + c_2\partial_\theta) \\
&= (X_1(c_2) - X_2(c_1) - c_1^2 - c_2^2)\partial_\theta \\
&= \kappa\partial_\theta,
\end{aligned}$$

where $\kappa$ denotes the Gaussian curvature (cf. again Section 4.4.1). Notice that the Hamiltonian vector field $\vec{H}$ has the form

$$\begin{aligned}
\vec{H} &= h_1\nabla_{X_1} + h_2\nabla_{X_2} && (16.31) \\
&= \cos\theta X_1 + \sin\theta X_2 + (c_1\cos\theta + c_2\sin\theta)\partial_\theta, && (16.32)
\end{aligned}$$

hence one can equivalently compute the curvature $\mathcal{R}_\lambda$ through the formula (16.25). Indeed we have

$$[\vec{H}, \partial_\theta] = \sin\theta X_1 - \cos\theta X_2 + (c_1\sin\theta - c_2\cos\theta)\partial_\theta, \tag{16.33}$$

and

$$-[\vec{H}, [\vec{H}, \partial_\theta]] = (X_1(c_2) - X_2(c_1) - c_1^2 - c_2^2)\partial_\theta = \kappa\partial_\theta. \tag{16.34}$$

These two independent computations gives an explicit proof of the identity (16.26) of Theorem 16.33.

## 16.7 Bibliographical note

The material and the results contained in this chapter about Riemannian curvature are quite classical and present in most textbooks on the topic, such as [dC92, Cha06, GHL90, Lee97, Pet16, Boo86].

However our approach and, in particular, the symplectic interpretation of the Levi-Civita connection and Riemannian curvature are not classical. The presentation given here is partially inspired by [AG97, Agr98b].

# Chapter 17

# Curvature in 3D contact sub-Riemannian geometry

In this chapter we discuss the notion of curvature in sub-Riemannian geometry. We will consider in particular the case of three-dimensional contact manifolds. In Chapter 16 we showed that the Riemann curvature tensor (along the direction of geodesics) can be recovered from the curvature of the associated Jacobi curve, which is regular, cf. also Chapters 14-15.

When the structure is sub-Riemannian, the methods used in the previous chapter are no more available since Jacobi curve associated with a geodesic is not regular, and there is hence no canonical associated connection. Still, the Jacobi curve associated with a sub-Riemannian geodesic is monotone and ample (cf. Proposition 17.18), we can compute its *principal curvature*, that is the curvature of the reduced (regular) curve, as discussed in Chapter 14 (cf. in particular Definition 14.48).

The computation of the principal curvature in the 3D contact sub-Riemannian case reduces to two functional invariants, called $\chi$ and $\kappa$, which are constant for left-invariant structures on Lie groups. The final part of the chapter is then devoted to the classification of left-invariant structures on 3D Lie groups with respect to local isometries and dilations, in terms of the above invariants.

## 17.1 A worked-out example: the 2D Riemannian case

As a preliminary computation towards the sub-Riemannian case, in this section we compute the curvature of Jacobi curves associated with geodesics on two-dimensional Riemannian surfaces.

Let $M$ be a 2-dimensional surface endowed with a Riemannian metric and let $f_1, f_2 \in \mathrm{Vec}(M)$ be a local orthonormal frame. The Riemannian Hamiltonian $H$ is written as follows (we use canonical coordinates $\lambda = (p, x)$ on $T^*M$)

$$H(p,x) = \frac{1}{2}\left(\langle p, f_1(x)\rangle^2 + \langle p, f_2(x)\rangle^2\right). \tag{17.1}$$

Notice that, given $\lambda = (p, x) \in T^*M$, the symplectic vector space $\Sigma_\lambda = T_\lambda(T^*M)$ is 4-dimensional.

Recall that, being $M$ a 2-dimensional surface, the level set $H^{-1}(1/2) \cap T_q^*M$ is diffeomorphic to a circle. Hence, let us introduce the coordinate $\theta$ on the level $H^{-1}(1/2) \cap T_x^*M$ by setting

$$\langle p, f_1(x)\rangle = \cos\theta, \qquad \langle p, f_2(x)\rangle = \sin\theta.$$

This corresponds to set $\theta = 0$ in the direction of $f_1$. Denote by $\partial_\theta$ the rotation in the fiber of the unit tangent bundle and by $\mathfrak{e}$ the Euler vector field (cf. Definition 15.13). Denote finally by $\vec{H}' := [\partial_\theta, \vec{H}]$.

Notice that $\Sigma_\lambda = \mathcal{V}_\lambda \oplus \mathcal{H}_\lambda$ where $\mathcal{V}_\lambda = \operatorname{span}\{\mathfrak{e}, \partial_\theta\}$ and $\mathcal{H}_\lambda = \operatorname{span}\{\vec{H}, \vec{H}'\}$.

**Lemma 17.1.** *The vectors $\{\mathfrak{e}, \partial_\theta, \vec{H}, \vec{H}'\}$ at $\lambda$ form a Darboux basis for $\Sigma_\lambda$.*

*Proof.* We have to prove the following identities

$$\sigma(\partial_\theta, \mathfrak{e}) = 0, \qquad \sigma(\partial_\theta, \vec{H}) = 0, \qquad \sigma(\mathfrak{e}, \vec{H}) = 1, \tag{17.2}$$

$$\sigma(\partial_\theta, \vec{H}') = 1, \qquad \sigma(\mathfrak{e}, \vec{H}') = 0, \qquad \sigma(\vec{H}, \vec{H}') = 0. \tag{17.3}$$

Indeed, let us prove first (17.2). The first equality follows from the fact that both vectors belong to the vertical subspace, that is Lagrangian. The second one is a consequence of the fact that, by construction, $\partial_\theta$ is tangent to the level set of $H$, i.e., $\sigma(\partial_\theta, \vec{H}) = \partial_\theta(\vec{H}) = \langle dH, \partial_\theta \rangle = 0$. The last identity is (15.10).

As a preliminary step for the proof of (17.3) notice that, if $s = i_{\mathfrak{e}}\sigma$ denotes the tautological Liouville form, one has

$$\langle s, \vec{H} \rangle = 1, \qquad \langle s, \vec{H}' \rangle = 0. \tag{17.4}$$

These two identities follow from

$$\langle s, \vec{H} \rangle = \sigma(\mathfrak{e}, \vec{H}) = 1, \tag{17.5}$$

$$\langle s, \vec{H}' \rangle = \langle s, [\partial_\theta, \vec{H}] \rangle = ds(\vec{H}, \partial_\theta) = \sigma(\vec{H}, \partial_\theta) = 0, \tag{17.6}$$

where in the second line we used the Cartan formula (4.84) and the fact that $\partial_\theta$ is vertical.

Let us now prove (17.3). We have[1] $[\partial_\theta, \vec{H}'] = [\partial_\theta, [\partial_\theta, \vec{H}]] = -\vec{H}$, thus by Cartan formula and (17.4)

$$\sigma(\partial_\theta, \vec{H}') = ds(\partial_\theta, \vec{H}') = -\langle s, [\partial_\theta, \vec{H}'] \rangle = \langle s, \vec{H} \rangle = \sigma(\mathfrak{e}, \vec{H}) = 1.$$

Moreover by (17.4)

$$\sigma(\mathfrak{e}, \vec{H}') = \langle s, \vec{H}' \rangle = 0.$$

The last computation is similar. Let us write

$$\sigma(\vec{H}, \vec{H}') = \langle dH, \vec{H}' \rangle = \langle dH, [\partial_\theta, \vec{H}] \rangle,$$

and apply the Cartan formula to the last term (where $dH$ acts as a 1-form). Then

$$dH([\partial_\theta, \vec{H}]) = d^2H(\partial_\theta, \vec{H}) - \partial_\theta \langle dH, \vec{H} \rangle + \vec{H}\, \langle dH, \partial_\theta \rangle = 0,$$

since the three terms vanish.

$\square$

Now we compute the curvature of the Jacobi curve, reduced by homogeneity. Notice that by Lemma 17.1 we can remove the symplectic space spanned by $\{\mathfrak{e}, \vec{H}\}$ and, being $\{\mathfrak{e}, \vec{H}\}^\angle = \{\partial_\theta, \vec{H}'\}$, we have

$$\widehat{J}_\lambda(t) = \operatorname{span}\{e_*^{-t\vec{H}} \partial_\theta\}.$$

[1] this identity is easily checked, for instance, thanks to formulas in Section 16.6.

Then we define the generator of the Jacobi curve

$$V_t = e_*^{-t\vec{H}}\partial_\theta, \qquad \dot{V}_t = e_*^{-t\vec{H}}[\vec{H}, \partial_\theta] = -e_*^{-t\vec{H}}\vec{H}'.$$

Notice that

$$\sigma(V_t, \dot{V}_t) = -1, \qquad \text{for every } t \geq 0. \tag{17.7}$$

Indeed (17.7) is true for $t = 0$, and the equality is then valid for all $t$ since the transformation $e_*^{-t\vec{H}}$ is symplectic. To compute the curvature of the Jacobi curve let us write

$$V_t = \alpha(t)V_0 - \beta(t)\dot{V}_0. \tag{17.8}$$

We claim that the matrix $S(t)$ representing the 1-dimensional Jacobi curve (that actually is a scalar), is given in these coordinates by

$$S(t) = \frac{\beta(t)}{\alpha(t)} = \frac{\sigma(V_0, V_t)}{\sigma(\dot{V}_0, V_t)}.$$

Indeed, the identity

$$V_t = \alpha(t)V_0 - \beta(t)\dot{V}_0 = \alpha(t)\left(V_0 - \frac{\beta(t)}{\alpha(t)}\dot{V}_0\right), \tag{17.9}$$

tells us that the matrix representing the vector space spanned by $V_t$ is the graph of the linear map $V_0 \mapsto \frac{\beta(t)}{\alpha(t)}(-\dot{V}_0)$. Moreover, using that $V_0$ and $-\dot{V}_0$ form a Darboux basis, it is easy to compute

$$\sigma(V_0, V_t) = \alpha(t)\underbrace{\sigma(V_0, V_0)}_{=0} - \beta(t)\underbrace{\sigma(V_0, \dot{V}_0)}_{=-1} = \beta(t), \tag{17.10}$$

$$\sigma(\dot{V}_0, V_t) = \alpha(t)\underbrace{\sigma(\dot{V}_0, V_0)}_{=1} - \beta(t)\underbrace{\sigma(\dot{V}_0, \dot{V}_0)}_{=0} = \alpha(t). \tag{17.11}$$

Differentiating the identity (17.7) with respect to $t$ one gets the relations

$$\sigma(V_t, \ddot{V}_t) = 0, \qquad \sigma(V_t, \dddot{V}_t) = -\sigma(\dot{V}_t, \ddot{V}_t).$$

Notice that these quantities are constant with respect to $t$. Collecting the above results one can compute the asymptotic expansion of $S(t)$ with respect to $t$ as follows

$$S(t) = \frac{-t + \frac{t^3}{6}\sigma(V_0, \dddot{V}_0) + O(t^5)}{1 + \frac{t^2}{2}\sigma(\dot{V}_0, \ddot{V}_0) + O(t^4)} \tag{17.12}$$

$$= \left(-t + \frac{t^3}{6}\sigma(V_0, \dddot{V}_0) + O(t^5)\right)\left(1 - \frac{t^2}{2}\sigma(\dot{V}_0, \ddot{V}_0) + O(t^4)\right). \tag{17.13}$$

In particular one has

$$\dot{S}(0) = -1, \qquad \ddot{S}(0) = 0, \qquad \dddot{S}(0) = 2\sigma(\dot{V}_0, \ddot{V}_0).$$

Applying the formula for the curvature (14.20) at $t = 0$, the curvature $\mathcal{R}_\lambda$ at $\lambda \in T^*M$ is finally computed as follows:

$$\mathcal{R}_\lambda = \frac{1}{2}\frac{\dddot{S}(0)}{\dot{S}(0)} = \sigma_\lambda(\ddot{V}_0, \dot{V}_0). \tag{17.14}$$

Using that $V_t = e_*^{-t\vec{H}}\partial_\theta$ we can expand $V_t$ for $t \to 0$ as follows

$$V_t = \partial_\theta + t[\vec{H}, \partial_\theta] + \frac{t^2}{2}[\vec{H}, [\vec{H}, \partial_\theta]] + O(t^3).$$

Hence (17.14) is rewritten as

$$\mathcal{R}_\lambda = \sigma_\lambda([\vec{H}, [\vec{H}, \partial_\theta]], [\vec{H}, \partial_\theta]) \tag{17.15}$$

$$= \sigma_\lambda([\vec{H}, \vec{H}'], \vec{H}'). \tag{17.16}$$

Let us now compute explicitly (17.16) in terms of the structure functions of a local orthonormal frame $f_1, f_2$. Denote the Hamiltonians

$$h_i(p, x) = \langle p, f_i(x) \rangle, \qquad i = 1, 2.$$

Then the equations for Pontryagin extremals read

$$\begin{cases} \dot{x} = h_1 f_1(x) + h_2 f_2(x) \\ \dot{h}_1 = \{H, h_1\} = \{h_2, h_1\} h_2 \\ \dot{h}_2 = \{H, h_2\} = -\{h_2, h_1\} h_1 \end{cases} \tag{17.17}$$

Moreover $\{h_2, h_1\}(p, x) = \langle p, [f_2, f_1](x) \rangle$. Assume that

$$[f_1, f_2] = c_1 f_1 + c_2 f_2, \qquad c_i \in C^\infty(M).$$

Then we have

$$\{h_2, h_1\} = -c_1 h_1 - c_2 h_2.$$

If we restrict to the level set $H^{-1}(1/2)$ defined by the relations $h_1 = \cos\theta$ and $h_2 = \sin\theta$, then equations (17.17) become

$$\begin{cases} \dot{x} = \cos\theta f_1 + \sin\theta f_2 \\ \dot{\theta} = c_1 \cos\theta + c_2 \sin\theta \end{cases}$$

and it is easy to compute the following expressions[2]

$$\vec{H} = h_1 f_1 + h_2 f_2 + (c_1 h_1 + c_2 h_2)\partial_\theta,$$
$$\vec{H}' = -h_2 f_1 + h_1 f_2 + (-c_1 h_2 + c_2 h_1)\partial_\theta,$$
$$[\vec{H}, \vec{H}'] = (f_1 c_2 - f_2 c_1 - c_1^2 - c_2^2)\partial_\theta.$$

Recall that

$$\kappa = f_1 c_2 - f_2 c_1 - c_1^2 - c_2^2, \tag{17.18}$$

is the Gaussian curvature of the surface $M$ (see also Chapter 4). Since $\sigma(\partial_\theta, \vec{H}') = 1$, one gets

$$\mathcal{R}_\lambda = \sigma_\lambda([\vec{H}, \vec{H}'], \vec{H}') = \sigma_\lambda(\kappa\partial_\theta, \vec{H}') = \kappa. \tag{17.19}$$

Notice that $\kappa$ is a function defined on the base manifold $M$.

[2]Here, with a slight abuse of notation, we still use the notation $h_1, h_2$ as functions of $\theta$, satisfying $\partial_\theta h_1 = -h_2$, and $\partial_\theta h_2 = h_1$.

**Exercise 17.2.** In this exercise we recover the previous computations introducing dual coordinates to the frame. Let $\nu_1, \nu_2$ be the dual basis to $f_1, f_2$ and set

$$f_\theta := h_1 f_1 + h_2 f_2, \qquad \nu_\theta := h_1 \nu_1 + h_2 \nu_2.$$

Define the smooth function $b := c_1 h_1 + c_2 h_2$ on $T^*M$. In these notations

$$\vec{H} = f_\theta + b\partial_\theta, \qquad \vec{H}' = f_{\theta'} + b'\partial_\theta,$$

where $'$ denotes the derivative with respect to $\theta$ and $f_{\theta'} = f_\theta'$. Then, using that in these coordinates the tautological form is $s = \nu_\theta$, show that the symplectic form is written as

$$\sigma = ds = d\theta \wedge \nu_{\theta'} - b\, \nu_1 \wedge \nu_2,$$

where $\nu_{\theta'} := \nu_\theta'$, and compute the following expressions:

$$\begin{aligned} i_{\vec{H}'}\sigma &= (b' - b)\nu_{\theta'} - d\theta, \\ [\vec{H}, \vec{H}'] &= (f_\theta b' - f_{\theta'} b - b^2 - b'^2)\partial_\theta, \end{aligned}$$

Use the above to give an alternative derivation of (17.19).

## 17.2 3D contact sub-Riemannian manifolds

In this section we consider a sub-Riemannian manifold $M$ of dimension 3 whose distribution is defined as the kernel of a 1-form $\omega \in \Lambda^1(M)$, i.e., $\mathcal{D}_q = \ker \omega_q$ for all $q \in M$. Let us also fix a local orthonormal frame $f_1, f_2$ such that

$$\mathcal{D}_q = \ker \omega_q = \operatorname{span}\{f_1(q), f_2(q)\}.$$

The 1-form $\omega \in \Lambda^1(M)$ defines a *contact distribution* if, by definition, $\omega \wedge d\omega$ is never vanishing.

**Exercise 17.3.** Let $M$ be a smooth manifold of dimension 3, $\omega \in \Lambda^1 M$ and $\mathcal{D} = \ker \omega$. The following are equivalent:

*(i)* $\omega$ is a contact 1-form,

*(ii)* $d\omega\big|_{\mathcal{D}} \neq 0$,

*(iii)* $\forall\, f_1, f_2 \in \mathcal{D}$ linearly independent, then $[f_1, f_2] \notin \mathcal{D}$,

where we recall that $\mathcal{D}$ denotes the set of horizontal vector fields.

*Remark* 17.4. The contact form $\omega$ is defined up to a smooth function, i.e., if $\omega$ is a contact form, $a\omega$ is a contact form for every $a \in C^\infty(M)$, $a \neq 0$. This fact allow us to normalize the contact form by requiring that

$$d\omega\big|_{\mathcal{D}} = \nu_1 \wedge \nu_2, \qquad \text{(i.e., } d\omega(f_1, f_2) = 1\text{)},$$

where $\nu_1, \nu_2$ is the dual basis to $f_1, f_2$. This is equivalent to say that $d\omega$ is equal to the area form induced on the distribution by the sub-Riemannian scalar product.

**Definition 17.5.** The *Reeb vector field* of the contact structure is the unique vector field $f_0 \in \mathrm{Vec}(M)$ that satisfies

$$d\omega(f_0, \cdot) = 0, \qquad \omega(f_0) = 1.$$

In particular $f_0$ is transversal to the distribution and the triple $\{f_0, f_1, f_2\}$ defines a basis of $T_qM$ at every point $q \in M$. Notice that $\omega, \nu_1, \nu_2$ is the dual basis to this frame.

*Remark* 17.6. The flow generated by the Reeb vector field $e^{tf_0} : M \to M$ is a one-parameter group of diffeomorphisms satisfying $(e^{tf_0})^*\omega = \omega$. Indeed

$$\mathcal{L}_{f_0}\omega = d(i_{f_0}\omega) + i_{f_0}d\omega = 0,$$

since $i_{f_0}\omega = \omega(f_0) = 1$, and $i_{f_0}d\omega = d\omega(f_0, \cdot) = 0$.

In what follows, to simplify the notation, we will denote the contact form $\omega$ by $\nu_0$, as the dual element to the vector field $f_0$. We can write the structure equations of this basis of 1-forms

$$\begin{cases} d\nu_0 = \nu_1 \wedge \nu_2 \\ d\nu_1 = c^1_{01}\nu_0 \wedge \nu_1 + c^1_{02}\nu_0 \wedge \nu_2 + c^1_{12}\nu_1 \wedge \nu_2 \\ d\nu_2 = c^2_{01}\nu_0 \wedge \nu_1 + c^2_{02}\nu_0 \wedge \nu_2 + c^2_{12}\nu_1 \wedge \nu_2 \end{cases} \tag{17.20}$$

Here $c^k_{ij}$ are smooth functions defined on an open set $U$ of the manifold, for $i, j, k = 0, 1, 2$. For simplicity in what follows we work as if $U = M$, since all the construction is local. Recall that

$$d\nu_k = \sum_{i,j=0}^{2} c^k_{ij}\nu_i \wedge \nu_j \qquad \text{if and only if} \qquad [f_j, f_i] = \sum_{k=0}^{2} c^k_{ij} f_k.$$

Introduce the coordinates $(h_0, h_1, h_2)$ on each fiber of $T^*M$ induced by the dual frame, i.e.,

$$\lambda = h_0\nu_0 + h_1\nu_1 + h_2\nu_2$$

where $h_i(\lambda) = \langle \lambda, f_i(q) \rangle$ are the Hamiltonians linear on fibers associated with $f_i$, for $i = 0, 1, 2$. The sub-Riemannian Hamiltonian is written as

$$H = \frac{1}{2}(h_1^2 + h_2^2).$$

We now compute the Poisson bracket $\{H, h_0\}$, denoting with $\{H, h_0\}_q$ its restriction to the fiber $T^*_qM$.

**Proposition 17.7.** *The Poisson bracket $\{H, h_0\}_q$ is a quadratic form. Moreover we have*

$$\{H, h_0\} = c^1_{01}h_1^2 + (c^2_{01} + c^1_{02})h_1h_2 + c^2_{02}h_2^2, \tag{17.21}$$

$$c^1_{01} + c^2_{02} = 0. \tag{17.22}$$

Notice that $\mathcal{D}^\perp_q \subset \ker \{H, h_0\}_q$ and $\{H, h_0\}_q$ can be identified with a quadratic form on $T^*_qM/\mathcal{D}^\perp_q = \mathcal{D}^*_q$.

*Proof.* Using the equality $\{h_i, h_j\}(\lambda) = \langle \lambda, [f_i, f_j](q)\rangle$ we get

$$\begin{aligned}\{H, h_0\} &= \frac{1}{2}\{h_1^2 + h_2^2, h_0\} = h_1\{h_1, h_0\} + h_2\{h_2, h_0\}\\ &= h_1(c_{01}^1 h_1 + c_{01}^2 h_2) + h_2(c_{02}^1 h_1 + c_{02}^2 h_2)\\ &= c_{01}^1 h_1^2 + (c_{01}^2 + c_{02}^1) h_1 h_2 + c_{02}^2 h_2^2.\end{aligned}$$

Differentiating the first equation in (17.20) one gets:

$$\begin{aligned}0 = d^2\nu_0 &= d\nu_1 \wedge \nu_2 - \nu_1 \wedge d\nu_2\\ &= (c_{01}^1 \nu_0 \wedge \nu_1) \wedge \nu_2 - \nu_1 \wedge (c_{02}^2 \nu_0 \wedge \nu_2)\\ &= (c_{01}^1 + c_{02}^2)\nu_0 \wedge \nu_1 \wedge \nu_2,\end{aligned}$$

which proves (17.22). □

*Remark* 17.8. Being $\{H, h_0\}_q$ a quadratic form on the Euclidean plane $\mathcal{D}_q$ (using the canonical identification of the vector space $\mathcal{D}_q$ with its dual $\mathcal{D}_q^*$ given by the scalar product), it can be interpreted as a symmetric operator on the plane itself. In particular its determinant and its trace are well defined. From (17.22) we get

$$\operatorname{trace}\{H, h_0\}_q = c_{01}^1 + c_{02}^2 = 0.$$

This identity is a consequence of the fact that the flow defined by the normalized Reeb $f_0$ preserves not only the distribution but also the area form on it.

**Definition 17.9.** We define the *first functional invariant* $\chi : M \to \mathbb{R}$ as follows

$$\chi(q) = \sqrt{-\det\{H, h_0\}_q}. \tag{17.23}$$

Notice that the function $\chi$ measures an intrinsic quantity since both $H$ and $h_0$ are defined only by the sub-Riemannian structure and are independent by the choice of the orthonormal frame. Indeed the quantity $\{H, h_0\}$ computes the derivative of $H$ along the flow of $\vec{h}_0$, i.e., the obstruction to the flow of the Reeb field $f_0$ (which preserves the distribution and the volume form on it) to be metric-preserving. Notice that, by definition $\chi \geq 0$.

**Corollary 17.10.** *Assume that the vector field $f_0$ is complete. Then $\{e^{tf_0}\}_{t\in\mathbb{R}}$ is a one-parametric group of sub-Riemannian isometries if and only if $\chi \equiv 0$.*

**Definition 17.11.** We define the *second functional invariant* $\kappa : M \to \mathbb{R}$ as follows

$$\kappa = f_2 c_{12}^1 - f_1 c_{12}^2 - (c_{12}^1)^2 - (c_{12}^2)^2 + \frac{c_{01}^2 - c_{02}^1}{2}. \tag{17.24}$$

**Exercise 17.12.** Show that the expression (17.24) for $\kappa$ does not depend on the choice of an orthonormal frame $f_1, f_2$ for the sub-Riemannian structure.

To solve the previous exercice, one uses the following lemma, whose proof is also left as an exercice.

**Lemma 17.13.** *Let $f_1, f_2$ be a local orthonormal frame on $M$ and let $\theta \in C^\infty(M)$. Denote with $\widehat{f}_1, \widehat{f}_2$ the frame obtained from the previous one with a rotation by an angle $\theta$ and with $\widehat{c}_{ij}^k$ the structure functions of the rotated frame. Then we have:*

$$\begin{aligned}\widehat{c}_{12}^1 &= \cos\theta(c_{12}^1 - f_1(\theta)) - \sin\theta(c_{12}^2 - f_2(\theta)),\\ \widehat{c}_{12}^2 &= \sin\theta(c_{12}^1 - f_1(\theta)) + \cos\theta(c_{12}^2 - f_2(\theta)).\end{aligned}$$

## 17.3 Canonical frames

In the last section we introduced the quantities $\chi$ and $\kappa$ for a three-dimensional contact sub-Riemannian structure, that are smooth functions on the manifold.

In this section we select a canonical orthonormal frame for the sub-Riemannian structure. We study separately the two cases $\chi \neq 0$ and $\chi = 0$ and we start by rewriting and improving Proposition 17.7 when $\chi \neq 0$.

**Proposition 17.14.** *Let $M$ be a 3D contact sub-Riemannian manifold and assume that $\chi(q) \neq 0$, for some $q \in M$. Then there exists a local frame such that*

$$\{H, h_0\} = 2\chi h_1 h_2. \tag{17.25}$$

*In particular, for a left-invariant stucture on a Lie group, there exists a unique (up to a sign) canonical frame $\{f_0, f_1, f_2\}$ such that*

$$\begin{aligned} [f_1, f_0] &= c_{01}^2 f_2, \\ [f_2, f_0] &= c_{02}^1 f_1, \\ [f_2, f_1] &= c_{12}^1 f_1 + c_{12}^2 f_2 + f_0. \end{aligned} \tag{17.26}$$

*In this frame we have*

$$\chi = \frac{c_{01}^2 + c_{02}^1}{2}, \qquad \kappa = -(c_{12}^1)^2 - (c_{12}^2)^2 + \frac{c_{01}^2 - c_{02}^1}{2}. \tag{17.27}$$

*Proof.* From Proposition 17.7 we know that $\{h, h_0\}_q$ (the restriction of the Poisson bracket $\{h, h_0\}$ to the fiber $T_q^*M$) is a non degenerate symmetric operator with zero trace. Hence we have a well defined, up to a sign, orthonormal frame by setting $f_1, f_2$ as the orthonormal isotropic vectors of this operator (the normalized eigenvectors corresponding to the to real and opposite eigenvalues). We stress that $f_0$ depends only on the structure and not on the orthonormal frame on the distribution. It is easily seen that in both cases we obtain expressions (17.25)-(17.26), which in turns implies (17.27). □

*Remark* 17.15. Notice that, if we change sign to $f_1$ or $f_2$, then $c_{12}^2$ or $c_{12}^1$, respectively, change sign in (17.26), while $c_{02}^1$ and $c_{01}^2$ are unaffected. Hence equalities (17.27) do not depend on the orientation of the sub-Riemannian structure.

**Exercise 17.16.** Prove that if the 3D contact structure is left-invariant on a unimodular Lie group (cf. Definition 7.45), satisfying $\chi \neq 0$, then the canonical frame given by Proposition 17.14 satisfies (17.26) with $c_{12}^1 = c_{12}^2 = 0$. In particular we can write

$$\begin{aligned} [f_1, f_0] &= (\chi + \kappa) f_2, \\ [f_2, f_0] &= (\chi - \kappa) f_1, \\ [f_2, f_1] &= f_0. \end{aligned} \tag{17.28}$$

If $\chi = 0$ the above procedure cannot apply. Indeed both the trace and the determinant of the operator vanish, hence we have $\{H, h_0\}_q = 0$. From (17.21) we get the identities

$$c_{01}^1 = c_{02}^2 = 0, \quad c_{01}^2 + c_{02}^1 = 0. \tag{17.29}$$

so that the general formulas for the commutators

$$\begin{aligned}
[f_1, f_0] &= c_{01}^1 f_1 + c_{01}^2 f_2, \\
[f_2, f_0] &= c_{02}^1 f_1 + c_{02}^2 f_2, \\
[f_2, f_1] &= c_{12}^1 f_1 + c_{12}^2 f_2 + f_0.
\end{aligned} \tag{17.30}$$

simplifies into the following (we set $c := c_{01}^2$)

$$\begin{aligned}
[f_1, f_0] &= c f_2, \\
[f_2, f_0] &= -c f_1, \\
[f_2, f_1] &= c_{12}^1 f_1 + c_{12}^2 f_2 + f_0.
\end{aligned} \tag{17.31}$$

We want to show, with an explicit construction, that also in this case there exists a smooth orthonormal frame such that $\kappa$ is the only structure function appearing in (17.31).

Now we can prove the main result of this section.

**Proposition 17.17.** *Let $M$ be a 3D contact sub-Riemannian manifold such that $\chi = 0$. Then for every $q \in M$ there exists local orthonormal frame $\widehat{f}_1, \widehat{f}_2$ such that the following relations are satisfied:*

$$\begin{aligned}
[\widehat{f}_1, f_0] &= \kappa \widehat{f}_2, \\
[\widehat{f}_2, f_0] &= -\kappa \widehat{f}_1, \\
[\widehat{f}_2, \widehat{f}_1] &= f_0.
\end{aligned} \tag{17.32}$$

*Moreover, if $M$ is simply connected, then the frame satisfying* (17.32) *is global.*

*Proof.* Fix an orthonormal frame $f_1, f_2$ for the sub-Riemannian structure. Thanks to Lemma 17.13, the statement is equivalent to the following fact: there exists a smooth function $\theta : M \to \mathbb{R}$ such that

$$f_1(\theta) = c_{12}^1, \quad f_2(\theta) = c_{12}^2. \tag{17.33}$$

Indeed, (17.33) would imply $\widehat{c}_{12}^1 = \widehat{c}_{12}^2 = 0$ and $\kappa = c$. To prove the claim, let us introduce the simplified notations $\alpha_1 := c_{12}^1$ and $\alpha_2 := c_{12}^2$. Then

$$\kappa = f_2(\alpha_1) - f_1(\alpha_2) - (\alpha_1)^2 - (\alpha_2)^2 + c. \tag{17.34}$$

If $\{\nu_0, \nu_1, \nu_2\}$ denotes the basis of 1-forms dual to $\{f_0, f_1, f_2\}$, we have

$$d\theta = f_0(\theta)\nu_0 + f_1(\theta)\nu_1 + f_2(\theta)\nu_2.$$

From (17.31) we get:

$$\begin{aligned}
f_0(\theta) &= ([f_2, f_1] - \alpha_1 f_1 - \alpha_2 f_2)(\theta) \\
&= f_2(\alpha_1) - f_1(\alpha_2) - \alpha_1^2 - \alpha_2^2 \\
&= \kappa - c.
\end{aligned}$$

Suppose now that (17.33) are satisfied, we get

$$d\theta = (\kappa - c)\nu_0 + \alpha_1 \nu_1 + \alpha_2 \nu_2 =: \eta. \tag{17.35}$$

with the right hand side independent from $\theta$. To prove the statement, it is enough to show that $\eta$ is an exact 1-form. Actually, since the result is local (or the manifold is simply connected), it is sufficient to prove that $\eta$ is closed. If we denote $\nu_{ij} := \nu_i \wedge \nu_j$, the dual equations of (17.31) are:

$$\begin{aligned}
d\nu_0 &= \nu_{12},\\
d\nu_1 &= -c\nu_{02} + \alpha_1\nu_{12},\\
d\nu_2 &= c\nu_{01} - \alpha_2\nu_{12}.
\end{aligned}$$

Using that $d^2\nu_i = d(d\nu_i) = 0$ for $i = 0, 1, 2$, we get two nontrivial relations:

$$f_1(c) + c\alpha_2 + f_0(\alpha_1) = 0, \tag{17.36}$$

$$f_2(c) - c\alpha_1 + f_0(\alpha_2) = 0. \tag{17.37}$$

Collecting all these computations we have

$$\begin{aligned}
d\eta &= d(\kappa - c) \wedge \nu_0 + (\kappa - c)d\nu_0 + d\alpha_1 \wedge \nu_1 + \alpha_1 d\nu_1 + d\alpha_2 \wedge \nu_2 + \alpha_2 d\nu_2\\
&= -dc \wedge \nu_0 + (\kappa - c)\nu_{12} + f_0(\alpha_1)\nu_{01} - f_2(\alpha_1)\nu_{12} + \alpha_1(\alpha_1\nu_{12} - c\nu_{02})\\
&\quad + f_0(\alpha_2)\nu_{02} + f_1(\alpha_2)\nu_{12} + \alpha_2(c\nu_{01} - \alpha_2\nu_{12})\\
&= (f_0(\alpha_1) + \alpha_2 c + f_1(c))\nu_{01} + (f_0(\alpha_2) - \alpha_1 c + f_2(c))\nu_{02}\\
&\quad + (\kappa - c - f_2(\alpha_1) + f_1(\alpha_2) + \alpha_1^2 + \alpha_2^2)\nu_{12} = 0,
\end{aligned}$$

which proves $d\eta = 0$ thanks to (17.34) and (17.36)-(17.37). □

## 17.4 Curvature of a 3D contact structure

In this section we compute the sub-Riemannian curvature of a 3D contact structure with a technique similar to that used in Section 16.6 for the 2D Riemannian case. Let us consider the level set $\{H = 1/2\} = \{h_1^2 + h_2^2 = 1\}$ and define the coordinate $\theta$ in such a way that

$$h_1 = \cos\theta, \qquad h_2 = \sin\theta.$$

On the bundle $T^*M \cap H^{-1}(1/2)$ we introduce coordinates $(x, \theta, h_0)$. Notice that each fiber is topologically a cylinder $S^1 \times \mathbb{R}$.

The sub-Riemannian Hamiltonian equation written in these coordinates are

$$\begin{cases}
\dot{x} = h_1 f_1(x) + h_2 f_2(x)\\
\dot{h}_1 = \{H, h_1\} = \{h_2, h_1\}h_2\\
\dot{h}_2 = \{H, h_2\} = -\{h_2, h_1\}h_1\\
\dot{h}_0 = \{H, h_0\}
\end{cases} \tag{17.38}$$

Computing the Poisson bracket $\{h_2, h_1\} = h_0 + c_{12}^1 h_1 + c_{12}^2 h_2$, and introducing the two functions $a, b : T^*M \to \mathbb{R}$ given by

$$a = \{H, h_0\} = \sum_{i,j=1}^{2} c_{0i}^j h_i h_j, \qquad b := c_{12}^1 h_1 + c_{12}^2 h_2,$$

we can rewrite the system, when restricted to $H^{-1}(1/2)$, as follows

$$\begin{cases} \dot{x} = \cos\theta f_1 + \sin\theta f_2 \\ \dot{\theta} = -h_0 - b \\ \dot{h}_0 = a \end{cases} \tag{17.39}$$

Notice that, while $a$ is intrinsic, the function $b$ depends on the choice of the orthonormal frame.

In particular we have for the Hamiltonian vector field in the coordinates $(q, \theta, h_0)$ is written as

$$\vec{H} = h_1 f_1 + h_2 f_2 - (h_0 + b)\partial_\theta + a\partial_{h_0}, \tag{17.40}$$

$$[\partial_\theta, \vec{H}] = \vec{H}' = -h_2 f_1 + h_1 f_2 + a'\partial_{h_0} - b'\partial_\theta, \tag{17.41}$$

where $h_1, h_2$ are shorthands for $\cos\theta$ and $\sin\theta$, respectively, and we denoted by $'$ the derivative with respect to $\theta$, i.e., $h_1' = -h_2$ and $h_2' = h_1$.

Now consider the symplectic vector space $\Sigma_\lambda = T_\lambda(T^*M)$. The vertical subspace $\mathcal{V}_\lambda$ is generated by the vectors $\partial_\theta, \partial_{h_0}, \mathfrak{e}$. Hence the Jacobi curve is

$$J_\lambda(t) = \operatorname{span}\{e_*^{-t\vec{H}}\partial_\theta, e_*^{-t\vec{H}}\partial_{h_0}, e_*^{-t\vec{H}}\mathfrak{e}\}. \tag{17.42}$$

**Proposition 17.18.** *The Jacobi curve associated with a non-constant Pontryagin extremal on a sub-Riemannian 3D contact is monotone and ample.*

*Proof.* We fix the following basis on the symplectic vector space $\Sigma_\lambda = T_\lambda(T^*M)$

$$\Sigma_\lambda = \{\partial_\theta, \partial_{h_0}, \mathfrak{e}\} \oplus \{\vec{H}, \vec{H}', \vec{h}_0\}.$$

Notice that this global basis is not Darboux (compare with Lemma 17.1) as for instance

$$\sigma(\vec{H}, \vec{h}_0) = \{H, h_0\} = a,$$

which is in general a non-zero function. The fact that the Jacobi curve is monotone is general (Proposition 15.2). Let us prove that the curve is ample. To do this we have to compute derivatives of the vertical vector fields in (17.42). Recall that for a (vertical) vector field $W$ on $T^*M$ one has

$$\frac{d}{dt} e_*^{-t\vec{H}} W = e_*^{-t\vec{H}}[\vec{H}, W].$$

It is enough to prove that the curve is ample at $t = 0$ for every (nonzero) initial covector. We are then reduced to compute Lie brackets between $\vec{H}$ and the vertical fields. Recall that by Lemma 15.15

$$[\vec{H}, \mathfrak{e}] = -\vec{H}.$$

Moreover, by construction

$$[\vec{H}, \partial_\theta] = -\vec{H}'.$$

Hence it is enough to show that $[\vec{H}, [\vec{H}, \partial_\theta]]$ has a non-zero component along $\vec{h}_0$. A direct computation using (17.40)-(17.41) shows that

$$[\vec{H}, [\vec{H}, \partial_\theta]] = -[\vec{H}, \vec{H}'] = \vec{h}_0 \mod \{\partial_\theta, \partial_{h_0}, \mathfrak{e}\} \oplus \{\vec{H}, \vec{H}'\}.$$

which completes the proof. $\square$

We can then perform first the reduction of the Jacobi curve by homogeneity and split the symplectic space as follows (cf. also Section 15.3)

$$\Sigma_\lambda = \mathrm{span}\{\mathfrak{e}, \vec{H}\} \oplus \mathrm{span}\{\mathfrak{e}, \vec{H}\}^\angle.$$

The reduced Jacobi curve $\Lambda(t) := \widehat{J}_\lambda(t)$ in the 4-dimensional symplectic space

$$\Lambda(t) := e_*^{-t\vec{H}}\widehat{\mathcal{V}}_\lambda/\mathbb{R}\vec{H} = \mathrm{span}\{e_*^{-t\vec{H}}\partial_\theta, e_*^{-t\vec{H}}\partial_{h_0}\}/\mathbb{R}\vec{H},$$

where $\mathbb{R}v$ is a shorthand for $\mathrm{span}\{v\}$.

Next, we perform the second reduction of the Jacobi curve, the one related with the fact that the curve is non-regular. Indeed notice that the rank of $\widehat{J}_\lambda(t)$ is 1. To find the new reduced curve, we need to compute the kernel of the derivative of the curve at $t = 0$

$$\Gamma := \ker \underline{\dot{\Lambda}}(0).$$

From the definition of $\underline{\dot{\Lambda}} := \underline{\dot{\Lambda}}(0)$ it follows that

$$\underline{\dot{\Lambda}}(\partial_\theta) = \pi_*[\vec{H}, \partial_\theta] = h_2 f_1 - h_1 f_2,$$
$$\underline{\dot{\Lambda}}(\partial_{h_0}) = \pi_*[\vec{H}, \partial_{h_0}] = \pi_*(\partial_\theta) = 0.$$

Hence $\Gamma = \mathrm{span}\{\partial_{h_0}\}$ and $\Gamma^\angle$ is 3-dimensional in $\widehat{\mathcal{V}}_\lambda/\mathbb{R}\vec{H}$. We refer to the notation of Chapter 14 for the reduced curve.

**Proposition 17.19.** *We have the following characterizations:*

*(i)* $\Gamma^\angle = \mathrm{span}\{\partial_{h_0}, \partial_\theta, \vec{H}'\}$ *in* $\widehat{\mathcal{V}}_\lambda/\mathbb{R}\vec{H}$*,*

*(ii)* $\{\partial_\theta, \vec{H}'\}$ *is a Darboux basis for* $\Gamma^\angle/\Gamma$*.*

*Proof.* Since $\partial_{h_0}$ and $\partial_\theta$ are vertical, to prove (i) it is enough to show that $\vec{H}'$ is skew-orthogonal to $\partial_{h_0}$. It is easy to compute, by Cartan's formula

$$\sigma(\partial_{h_0}, \vec{H}') = \partial_{h_0}\langle s, \vec{H}'\rangle - \vec{H}'\,\langle s, \partial_{h_0}\rangle - \langle s, [\partial_{h_0}, \vec{H}']\rangle = 0,$$

since all the three terms in the right hand side vanish. Indeed $\langle s, \vec{H}'\rangle = \sigma(\vec{E}, \vec{H}') = 0$ and $\langle s, \partial_{h_0}\rangle = \langle s, [\partial_{h_0}, \vec{H}']\rangle = 0$ since $\partial_{h_0}$ and $[\partial_{h_0}, \vec{H}']$ are both vertical, as can be computed from (17.41).

To complete the proof of (ii) it is enough to notice that, using $[\partial_\theta, \vec{H}'] = -\vec{H}$, that

$$\sigma(\partial_\theta, \vec{H}') = \partial_\theta\langle s, \vec{H}'\rangle - \vec{H}'\,\langle s, \partial_\theta\rangle - \langle s, [\partial_\theta, \vec{H}']\rangle = \langle s, \vec{H}\rangle = 1. \qquad \square$$

Next we compute the curvature in terms of the Hamiltonian vector field and its commutators. For a vector field $W$ we use the notations

$$\dot{W} := [\vec{H}, W], \qquad W' := [\partial_\theta, W].$$

We stress that $\dot{W} = \frac{d}{dt}\big|_{t=0} e_*^{-t\vec{H}} W$.

**Lemma 17.20.** *Let us define* $V_t = e_*^{-t\vec{H}}\partial_{h_0}$*. We have the following identities*

$$V_0 = \partial_{h_0}, \qquad \dot{V}_0 = \partial_\theta, \qquad \ddot{V}_0 = -\vec{H}'. \tag{17.43}$$

*Moreover, for all* $t \geq 0$*, we have* *(i)* $\sigma(V_t, \dot{V}_t) = \sigma(V_t, \ddot{V}_t) = 0$,

*(ii)* $\sigma(V_t, V_t^{(3)}) = 1$, $\sigma(\dot{V}_t, V_t^{(3)}) = \sigma(V_t, V_t^{(4)}) = 0$,

*(iii)* $\sigma(\ddot{V}_t, V_t^{(3)}) = -\sigma(\dot{V}_t, V_t^{(4)}) = \sigma(V_t, V_t^{(5)})$.

*Proof.* Equation (17.43) follows from the computations performed in the proof of Proposition 17.18. The first equality of claim (i) is a consequence of the the fact that $\partial_\theta$ and $\partial_{h_0}$ are both vertical. Differentiating it with respect to $t$, one gets

$$0 = \sigma(\dot{V}_t, \dot{V}_t) + \sigma(V_t, \ddot{V}_t) = \sigma(V_t, \ddot{V}_t).$$

Differentiating again with respect to $t$ one gets

$$\sigma(\dot{V}_t, \ddot{V}_t) + \sigma(V_t, V_t^{(3)}) = 0,$$

Using that $\sigma(\dot{V}_0, \ddot{V}_0) = -\sigma(\partial_\theta, \vec{H}') = -1$ and the fact that $e_*^{-t\vec{H}}$ is a symplectic transformation (hence preserves the symplectic product) one gets $\sigma(\dot{V}_t, \ddot{V}_t) = -1$ for all $t$. This proves the first equality in (ii). With similar arguments one can show that $\sigma(\dot{V}_t, V_t^{(3)}) = \sigma(V_t, V_t^{(4)}) = 0$. Claim (iii) is proved similarly, evaluating derivatives of order 4. $\square$

The previous lemma shows that the only symplectic invariant one obtains up to derivatives of order five is the quantity

$$r_\lambda := \sigma_\lambda(\ddot{V}_t, V_t^{(3)}) = -\sigma_\lambda(\dot{V}_t, V_t^{(4)}) = \sigma_\lambda(V_t, V_t^{(5)}). \tag{17.44}$$

It turns out that this is the value of the sub-Riemannian curvature.

**Proposition 17.21.** *We have the identity*

$$\mathcal{R}_\lambda = -\frac{r_\lambda}{10} = -\frac{1}{10}\sigma_\lambda([\vec{H}, \vec{H}'], \vec{H}'). \tag{17.45}$$

*Proof.* The first identity in (17.45) follows from (17.44) and (17.43).

To prove the first identity in (17.45), we have to compute the Schwarzian derivative of the reduced curve, in the Darboux basis $\{\dot{V}_0, -\ddot{V}_0\}$ of the space $\Gamma^\angle/\Gamma$ (notice the minus sign, cf. also Proposition 17.19).

Recall that $\Lambda(t) = \text{span}\{V_t, \dot{V}_t\}$. To compute the 1-dimensional reduced curve $\Lambda^\Gamma(t)$ in the symplectic space $\Gamma^\angle/\Gamma$ we need to compute the intersection of $\Lambda(t)$ with $\Gamma^\angle$ (for all $t > 0$). In other words we look for $x(t)$ such that

$$\sigma(\dot{V}_t + x(t)V_t, V_0) = 0 \qquad \Longrightarrow \qquad x(t) = -\frac{\sigma(\dot{V}_t, V_0)}{\sigma(V_t, V_0)}. \tag{17.46}$$

We then write this vector as a linear combination of the Darboux basis (cf. (17.9) for the 2D Riemannian case)

$$\dot{V}_t + x(t)V_t = \alpha(t)\dot{V}_0 - \beta(t)\ddot{V}_0 + \xi(t)V_0. \tag{17.47}$$

To intepret it as a curve in the space $\Gamma/\Gamma^{\angle}$ we simply ignore the $V_0$ component. In the coordinates on $\Gamma^{\angle}/\Gamma$ given by the splitting $\dot{V}_0 \oplus \ddot{V}_0$ endowed with the Darboux basis $\{\dot{V}_0, -\ddot{V}_0\}$, the matrix $S(t)$, which is a scalar, representing the curve is

$$S(t) = \frac{\beta(t)}{\alpha(t)} \tag{17.48}$$

Notice that this is a one-dimensional non-degenerate curve. Moreover, using that $\{\dot{V}_0, -\ddot{V}_0\}$ is a Darboux basis, it is easy to compute (thanks to Lemma 17.20; compare also with (17.10)-(17.11))

$$\sigma(\dot{V}_0, \dot{V}_t + x(t)V_t) = \beta(t), \tag{17.49}$$

$$\sigma(\ddot{V}_0, \dot{V}_t + x(t)V_t) = \alpha(t). \tag{17.50}$$

Combining (17.49),(17.50) with (17.48) and (17.46) one gets

$$S(t) = \frac{\sigma(\dot{V}_t, \dot{V}_0)\sigma(V_t, V_0) - \sigma(V_t, \dot{V}_0)\sigma(\dot{V}_t, V_0)}{\sigma(\dot{V}_t, \ddot{V}_0)\sigma(V_t, V_0) - \sigma(V_t, \ddot{V}_0)\sigma(\dot{V}_t, V_0)} \tag{17.51}$$

After some long but straightforward computations, by Taylor expansion and using Lemma 17.20, one gets

$$S(t) = -\frac{t}{4} + \frac{t^3}{120}r + O(t^5). \tag{17.52}$$

Since $\ddot{S}(0) = 0$ the principal curvature $\mathcal{R}_\lambda$ (that is the curvature of the rediced regular curve) is computed by

$$\mathcal{R}_\lambda = \frac{\dddot{S}(0)}{2\dot{S}(0)} = -\frac{r_\lambda}{10}. \qquad \square$$

We end this section by computing the expression of the curvature in terms of the orthonormal frame for the distribution and the Reeb vector field. As usual we restrict to the level set $H^{-1}(1/2)$ where

$$h_1^2 + h_2^2 = 1, \qquad h_1 = \cos\theta, \qquad h_2 = \sin\theta.$$

In the following we use the notation

$$f_\theta = h_1 f_1 + h_2 f_2, \qquad \nu_\theta = h_1\nu_1 + h_2\nu_2.$$

If $h = (h_1, h_2) = (\cos\theta, \sin\theta)$ we denote by $h' = (-h_2, h_1) = (-\sin\theta, \cos\theta)$ its derivative with respect to $\theta$ and, more in general, we denote $F' := \partial_\theta F$ for a smooth function $F$ on $T^*M$.

To express the quantity $r = \sigma([\vec{H}, \vec{H}'], \vec{H}')$ we start by computing the complete expression of the commutator $[\vec{H}, \vec{H}']$. From (17.40) and (17.41) one gets

$$[\vec{H}, \vec{H}'] = -f_0 + h_0 f_\theta + (f_2 c_{12}^1 - f_1 c_{12}^2 - (h_0 + b)b - (b')^2 + a')\partial_\theta.$$

Next we write, following this notation, the symplectic form $\sigma = ds$. The Liouville form $s$ is expressed, in the dual basis $\nu_0, \nu_1, \nu_2$ to the basis of vector fields $f_1, f_2, f_0$ as follows

$$s = h_0\nu_0 + \nu_\theta.$$

Hence the symplectic form $\sigma$ is written as follows:

$$\sigma = dh_0 \wedge \nu_0 + h_0\, \nu_\theta \wedge \nu_{\theta'} + d\theta \wedge \nu_{\theta'} + d\nu_\theta,$$

where we used the shorthand $\nu_{\theta'} := \partial_\theta \nu_\theta$ and (17.20). Computing the symplectic product then one finds

$$10\mathcal{R}_\lambda = h_0^2 + \frac{3}{2}a' + \kappa,$$

where we recall that

$$\kappa = f_2 c_{12}^1 - f_1 c_{12}^2 - (c_{12}^1)^2 - (c_{12}^2)^2 + \frac{c_{01}^2 - c_{02}^1}{2}. \tag{17.53}$$

By homogeneity, the function $\mathcal{R}$ is defined on the whole $T^*M$, and not only for $\lambda \in H^{-1}(1/2)$. For every $\lambda = (h_0, h_1, h_2) \in T_x^*M$ we have

$$r_\lambda := 10\mathcal{R}_\lambda = h_0^2 + \frac{3}{2}a' + \kappa(h_1^2 + h_2^2). \tag{17.54}$$

### 17.4.1 Geometric interpretation

Let us consider the kernel of the restriction of sub-Riemannian Hamiltonian to the fiber $T_x^*M$

$$\ker H_x = \{\lambda \in T_x^*M \mid \langle \lambda, v\rangle = 0,\ \forall\, v \in \mathcal{D}_x\} = \mathcal{D}_x^\perp. \tag{17.55}$$

The restriction of $r_\lambda$ defined in (17.54) to the 1-dimensional subspace $\mathcal{D}_x^\perp$, for every $x \in M$, is the strictly positive quadratic form $r_\lambda|_{\mathcal{D}_x^\perp} = h_0^2$. Moreover it is equal to 1 when evaluated on the Reeb vector field. Hence $r_\lambda$ encodes both the contact form $\alpha$ and its normalization.

Let us consider the orthogonal complement $\mathcal{D}_x^\dagger$ of $\mathcal{D}_x^\perp$ in the fiber with respect to $r_\lambda$ (this is indeed isomorphic to the space of linear functionals defined on $\mathcal{D}_x$). This induces the well-defined splitting

$$T_x^*M = \mathcal{D}_x^\perp \oplus \mathcal{D}_x^\dagger = \operatorname{span}\{\nu_0\} \oplus \operatorname{span}\{\nu_1, \nu_2\}, \tag{17.56}$$

where $\nu_0 = \alpha$ and $\nu_1, \nu_2$ form a dual basis to $f_0, f_1, f_2$ (where $f_1, f_2$ is an isotropic frame in the sense of Proposition 17.14). Indeed the restriction of $r_\lambda$ to elements in $\mathcal{D}_x^\dagger$ is

$$r_\lambda|_{\mathcal{D}_x^\dagger} = (\kappa + 3\chi)h_1^2 + (\kappa - 3\chi)h_2^2. \tag{17.57}$$

By using the Euclidean metric induced by $H_x$ on $\mathcal{D}_x$, it can be identified with a symmetric operator. From this formulae it is easy to recover the two invariants $\chi, \kappa$

$$\operatorname{trace}\left(r_\lambda|_{\mathcal{D}_x^\dagger}\right) = 2\kappa, \qquad \operatorname{discr}\left(r_\lambda|_{\mathcal{D}_x^\dagger}\right) = 36\chi^2, \tag{17.58}$$

where the discriminant of an operator $Q$, defined on a two-dimensional space, is defined as the square of the difference of its eigenvalues, and is computed by the formula $\operatorname{discr}(Q) = \operatorname{trace}^2(Q) - 4\det(Q)$.

*Remark* 17.22. When $\chi = 0$, then the eigenvalues of $\mathcal{R}$ coincide. In this case the pushforward $(e^{tf_0})_*$ of the flow induced by the Reeb vector field preserves the metric on the distribution and it is possible to define locally the quotient of the manifold $M$ with respect to this action, i.e., the space of integral lines of $f_0$. The two dimensional surface defined by the quotient structure is a 2-dimensional manifold $N$ endowed with a well-defined Riemannian metric.

The sub-Riemannian structure on $M$ coincides with the isoperimetric Dido problem constructed on the surface $N$. The invariant $\kappa$ is constant along the orbits of the Reeb vector field, i.e., is a well-defined function on $N$, and it represents indeed its Riemannian curvature.

Indeed and it is easy to see that the identities

$$e_*^{tf_0} f_i = f_i, \quad i = 1, 2.$$

imply $[f_0, f_1] = [f_0, f_2] = 0$. Hence $c_{01}^2, c_{02}^1 = 0$ and the expression of $\kappa$ reduces to the Riemannian curvature of a surface whose orthonormal frame is $f_1, f_2$ (compare formulas (17.18) and (17.24)).

The Heisenberg case corresponds with the case when the surface $N$ has vanishing Gaussian curvature, i.e., is the Euclidean plane.

## 17.5 Local classification of 3D left-invariant structures

The goal of this section is to give a complete classification of left-invariant sub-Riemannian structures on 3D Lie groups up to local equivalence with respect to local isometries and dilations.

Recall that a *local isometry* between two sub-Riemannian structures is a local diffeomorphism preserving the distribution and the sub-Riemannian metric.

For simplicity, we restate here a definition in the context of 3D contact manifolds.

**Definition 17.23.** Let $M$, $N$ be two 3D contact sub-Riemannian structures, and let $x_0 \in M$, $y_0 \in N$. The two structures are said to be locally isometric if there exists a local diffeomorphism $\phi : O_{x_0} \to O_{y_0}$ such that $\phi(x_0) = y_0$ and such that $\phi_* : T_{x_0}M \to T_{y_0}N$ preserves the distribution and the inner product on it.

*Remark* 17.24. If $\phi : O_{x_0} \to O_{y_0}$ is a local isometry between two contact sub-Riemannian structures on $M$ and $N$, and if $f_1, f_2$ is a local orthonormal frame for $M$ on $O_{x_0}$ then

$$g_i := \phi_* f_i, \qquad \text{for } i = 1, 2. \tag{17.59}$$

defines a local orthonormal frame for $N$ on $O_{y_0}$.

A *dilation* is obtained by multiplying elements of an orthonormal frame by a common factor $\lambda > 0$. It corresponds to a multiplication of all distances in the manifold by a factor $\lambda^{-1}$.

Recall from Chapter 7 that a sub-Riemannian structure on a Lie group is said to be *left-invariant* if its distribution and the inner product are preserved by left translations on the group. A left-invariant distribution on a Lie group is uniquely determined by a two dimensional subspace of the Lie algebra of the group. In the 3D case, the distribution is bracket generating (and contact) if and only if the subspace is not a Lie subalgebra.

*Remark* 17.25. Notice that a sub-Riemannian local isometry $\phi : G \to G'$ between two left-invariant structures on Lie groups, induces a linear maps between Lie algebras $\phi_* : \mathfrak{g} \to \mathfrak{g}'$. In particular $\phi_*$ maps the distribution to the distribution and preserves the inner product on it.

Conversely, given a linear map $L : \mathfrak{g} \to \mathfrak{g}'$ which maps the distribution to the distribution and preserves the inner product on it, then there always exists a map $\phi : G \to G'$ defined locally in a neighborhood of the identities of the corresponding Lie groups, which is a local isometry and such that $\phi_* = L$.

This is a consequence of Theorem 2.41 in Section 2.4.1. We can then reduce the problem to the level of Lie algebras.

**Exercise 17.26.** Prove that for a left-invariant structure on a 3D Lie group, the two functional invariants $\chi$ and $\kappa$ are constant. Moreover prove that two left-invariant structures on 3D Lie groups which are locally isometric have the same values of $\chi$ and $\kappa$.

It follows then immediately from the definitions that two sub-Riemannian structures that have different invariants $\chi$ and $\kappa$ cannot be locally isometric.

**Proposition 17.27.** *Let $G, H$ be 3D Lie groups endowed with left-invariant sub-Riemannian structures. Then if $(\kappa_G, \chi_G) \neq (\kappa_H, \chi_H)$, then $G$ and $H$ are not locally isometric.*

Nevertheless the invariants $\kappa$ and $\chi$ turn out to be not complete, in the sense that there exists non-isometric left-invariant sub-Riemannian stuctures which has the same value of $\chi$ and $\kappa$.

**Exercise 17.28** (Homogeneity with respect to local dilations)**.** Prove that $\chi$ and $\kappa$ are homogeneous of degree 2 with respect to dilations.

More precisely, assume that the sub-Riemannian structure $(M, \mathcal{D}, \langle \cdot \,|\, \cdot \rangle)$ is locally defined by the orthonormal frame $f_1, f_2$, i.e.

$$\mathcal{D} = \operatorname{span}\{f_1, f_2\}, \qquad \langle f_i \,|\, f_j \rangle = \delta_{ij}.$$

Consider now the dilated structure defined by the orthonormal frame $\lambda f_1, \lambda f_2$

$$\mathcal{D} = \operatorname{span}\{f_1, f_2\}, \qquad \langle f_i \,|\, f_j \rangle = \frac{1}{\lambda^2}\delta_{ij}, \qquad \lambda > 0.$$

If $\chi, \kappa$ and $\widetilde{\chi}, \widetilde{\kappa}$ denote the invariants of the two structures respectively, prove that

$$\widetilde{\chi} = \lambda^2 \chi, \qquad \widetilde{\kappa} = \lambda^2 \kappa.$$

In the following we are interested in a classification up to local isometries and dilations, hence we can always suppose that the local invariants of our structure satisfy

$$\chi = \kappa = 0, \qquad \text{or} \qquad \chi^2 + \kappa^2 = 1. \tag{17.60}$$

A left-invariant sub-Riemannian structure on a 3D Lie group satisfying (17.60) is said to be *normalized*. The following result is the main result of this section.

**Theorem 17.29.** *Let $M$ be a left-invariant 3D sub-Riemannian structure on a Lie group and let $(\kappa, \chi)$ be its invariants.*

*(i) Assume that $\chi = \kappa = 0$. Then the structure is locally isometric to the Heisenberg group endowed with the standard sub-Riemannian structure,*

*(ii) Assume that $\chi^2 + \kappa^2 = 1$. Then there exist up to three non-isometric normalized sub-Riemannian structures with the same invariants; in particular there exists a unique normalized structure on a unimodular Lie group for the given value of the pair $(\kappa, \chi)$,*

*(iii) Assume $\chi \neq 0$ or $\chi = 0, \kappa \geq 0$. Then two structures are locally isometric if and only if their Lie algebras are isomorphic.*

### 17.5.1 A description of the classification

A more precise interpretation of Theorem 17.29 can be given in terms of Figure 17.1. To this aim, let us first recall the classification of 3D Lie algebras.

The classification of 3D Lie algebras (up to isomorphisms) is very classical and can be found, for instance, in [Jac62]. It can be expressed in terms of the dimension of the square $[\mathfrak{g}, \mathfrak{g}]$ of the Lie algebra. Every 3D Lie algebras is isomorphic to one of the following list.

- $\dim[\mathfrak{g}, \mathfrak{g}] = 0$
  - the 3D abelian Lie algebra (the Lie algebra of the abelian group $(\mathbb{R}^3, +)$).
- $\dim[\mathfrak{g}, \mathfrak{g}] = 1$
  - $\mathfrak{h}$, the nilpotent Lie algebra of the Heisenberg group $\mathbb{H}$,
  - $\mathfrak{a}^+(\mathbb{R}) \oplus \mathbb{R}$, where $\mathfrak{a}^+(\mathbb{R})$ is the solvable Lie algebra of the group $A^+(\mathbb{R})$ of orientation preserving affine maps on $\mathbb{R}$.
- $\dim[\mathfrak{g}, \mathfrak{g}] = 2$
  - $\mathfrak{se}(2)$ the solvable Lie algebras of the group $SE(2)$ of orientation preserving motions of the Euclidean plane,
  - $\mathfrak{sh}(2)$ the solvable Lie algebras of the group $SH(2)$ of orientation preserving motions of the hyperbolic plane.
  - the other algebras of this subclass can be interpreted as an operator acting on a 2-dimensional abelian algebra and splits into two sub-classes $\mathfrak{solv}^+$ and $\mathfrak{solv}^-$, depending on the sign of the determinant of the operator. See also Section 17.5.3.
- $\dim[\mathfrak{g}, \mathfrak{g}] = 3$
  - $\mathfrak{sl}(2)$, the simple Lie algebra of the group $SL(2)$,
  - $\mathfrak{su}(2)$, the simple Lie algebra of the group $SU(2)$.

To define a left-invariant structure on a 3D Lie group $G$, one needs first to fix a bracket-generating distribution, that is a 2-dimensional subspace of the Lie algebra $\mathfrak{g}$. Notice that the 3D abelian Lie algebra does not contain any bracket-generating 2-dimensional subspace.

In all other cases the classification of bracket-generating 2-dimensional subspaces up to automorphisms of the Lie algebra can be done explicitly as follows.

**Exercise 17.30.** (i). Prove that for every non-abelian Lie algebra listed above different from $\mathfrak{sl}(2)$, there exists a unique bracket-generating 2-dimensional subspace up to automorphisms of the Lie algebra.

A 2-dimensional subspace of $\mathfrak{sl}(2)$ is called *elliptic* (resp. *hyperbolic*) if the restriction of the Killing form (cf. Definition 7.37) on this subspace is sign-definite (resp. sign-indefinite).

(ii). Prove that for the Lie algebra $\mathfrak{sl}(2)$ there exists two equivalence classes, given by elliptic and hyperbolic 2-dimensional subspaces.

According to Exercice 17.30, we use the notation $\mathfrak{sl}_e(2)$ (resp. $\mathfrak{sl}_h(2)$) when an elliptic (resp. hyperbolic) distribution is fixed on $\mathfrak{sl}(2)$. Notice that in the case of $\mathfrak{sl}_e(2)$ one can define a natural inner product on the distribution by restricting the Killing form on it, since it is sign-definite.

Notice that a priori, even in the cases when the bracket-generating distribution is unique (i.e., there exists a unique equivalence class), one could define different inner products on it, giving structures with a priori different values of $\chi$ and $\kappa$.

As we explained, normalized structures with different $\chi$ and $\kappa$ are not locally isometric by Proposition 17.27. On the other hand, for a given pair $(\kappa, \chi)$ there exist up to three non-isometric normalized sub-Riemannian structures with the same invariants.

In Figure 17.1 a structure is identified by the point $(\kappa, \chi)$ that is either the origin (if $(\kappa, \chi) = (0, 0)$) or it belongs to the half unit circle $\{\kappa^2 + \chi^2 = 1, \chi \geq 0\}$. The three different arcs of the circle are not overlapped in Figure 17.1 to highlight that two distinct points represent non locally isometric structures.

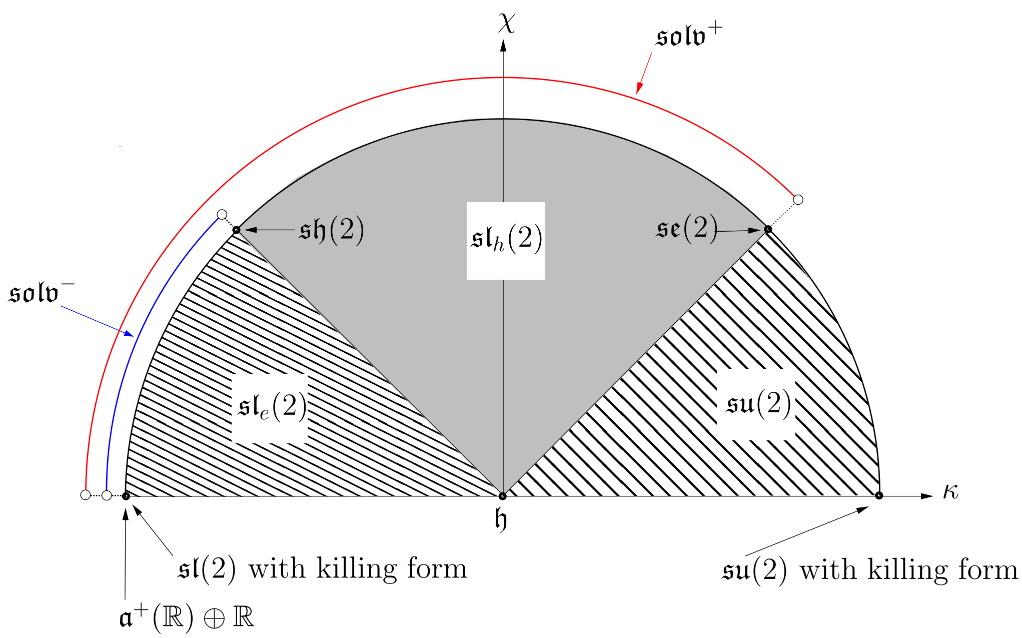



Figure 17.1: The classification of 3D left-invariant sub-Riemannian structures. Each point on the complete half-circle represents Lie algebras of unimodular Lie group ($\mathfrak{sl}(2), \mathfrak{sh}(2), \mathfrak{se}(2), \mathfrak{su}(2)$). Each point on the two sub-arcs of circles of $\mathfrak{solv}^+$ and $\mathfrak{solv}^-$ represents different sub-Riemannian structures on different solvable groups, that are not unimodular. The algebra $\mathfrak{a}^+(\mathbb{R}) \oplus \mathbb{R}$ is identified in the picture with $\mathfrak{sl}(2)$ with the Killing form since they are isometric (although not isomorphic). The algebras $\mathfrak{se}(2)$ (resp. $\mathfrak{sh}(2)$) can be seen as a limit of the cases $\mathfrak{solv}^+$ (resp. $\mathfrak{solv}^-$) for $\chi \to \kappa$ (resp. $\chi \to -\kappa$). When $\chi \to 0$ both $\mathfrak{solv}^+$ and $\mathfrak{solv}^-$ tend to $\mathfrak{a}^+(\mathbb{R}) \oplus \mathbb{R}$. Cf. the dotted lines.

To distinguish between structures with different $\chi$ and $\kappa$, we use the canonical orthonormal frames for the sub-Riemannian structure described in Section 17.3. In this way not only $\kappa$ and $\chi$, but all structure functions of the Lie algebra written with respect to this frame are invariant with respect to local isometries.

As a byproduct of the classification, for each distinct point $(\kappa, \chi)$ in the three arcs of the circles there exists a canonical choice of the sub-Riemannian structure on the corresponding Lie algebra (hence of an orthonormal frame on the distibution).

Some observations on Figure 17.1 are in order:

- each point on the complete half-circle represents Lie algebra of a unimodular Lie group. The algebras $\mathfrak{se}(2)$ and $\mathfrak{sh}(2)$ are solvable and for this reason they are connected (via a short dotted line) to the corresponding arcs of $\mathfrak{solv}^+$ and $\mathfrak{solv}^-$, since they arise as limiting cases of the latter. The algebra $\mathfrak{a}^+(\mathbb{R}) \oplus \mathbb{R}$ arises as a limit case as well.

- each of the three sub-arcs of circles of $\mathfrak{sl}_e(2)$, $\mathfrak{sl}_h(2)$, $\mathfrak{su}(2)$ represents different sub-Riemannian structures *on the same group.*

- each point on the two sub-arcs of circles of $\mathfrak{solv}^+$ and $\mathfrak{solv}^-$ represents different sub-Riemannian structures *on different groups.* These groups are not unimodular.

- the left-invariant sub-Riemannian structure $\mathfrak{a}^+(\mathbb{R})\oplus\mathbb{R}$ is isometric to the one on $\mathfrak{sl}_e(2)$ (where the inner product is given by the restriction to the Killing form on the distribution). Notice that these two Lie algebras are not isomorphic (cf. Section 17.5.2).

- the two sub-Riemannian structures on $\mathfrak{sl}_e(2)$ and $\mathfrak{su}(2)$ where the metric is defined by the restriction of the Killing form to the distribution are of type $\mathbf{d} \oplus \mathbf{s}$ (cf. Section 7.7.1 and Section 13.6).

#### A more intrinsic way to distinguish structures with the same invariants

Recall from the discussion in Section 17.4.1 that we have two quadratic forms on the cotangent space $H$ and $r$. For a given $x \in M$ we have $\ker H_x = \mathcal{D}_x^\perp$ and considering the orthogonal complement $\mathcal{D}_x^\dagger$ of $\mathcal{D}_x^\perp$ in the fiber with respect to $r_\lambda$ one has the well-defined splitting

$$T_x^*M = \mathcal{D}_x^\perp \oplus \mathcal{D}_x^\dagger, \tag{17.61}$$

and $H_x$ defines an Euclidean strtucture on $\mathcal{D}_x^\dagger$. The invariants $\kappa$ and $\chi$ are the rescaled trace and the discriminant of $r_\lambda|_{\mathcal{D}_x^\dagger}$ (cf. 17.57).

How to distinguish structures with the same values of $\kappa$ and $\chi$? Compute the Poisson bracket $\{H, r\}$. This is a cubic polynomial in $h_1, h_2, h_0$. If $\{H, r\}|_{\mathcal{D}_x^\dagger} = 0$, then the structure is unimodular. If not, the zeros of $\{H, r\}|_{\mathcal{D}_x^\dagger} = 0$ are exactly the eigenvectors of $r|_{\mathcal{D}_x^\dagger}$. One of the zeroes has multiplicity two, the other one is simple. If the eigenvector which has multiplicity two corresponds to the larger eigenvalue, then we are in the case $\mathfrak{solv}^+$, otherwise we are in the case $\mathfrak{solv}^-$.

### 17.5.2 A sub-Riemannian isometry between non isomorphic Lie groups

As a byproduct of Theorem 17.29 and Proposition 17.17, we get also a uniformization-like theorem for "constant curvature" manifolds in the sub-Riemannian setting:

**Corollary 17.31.** *Let $M$ be a complete simply connected 3D contact sub-Riemannian manifold. Assume that $\chi = 0$ and $\kappa$ is constant on $M$. Then $M$ is isometric to a left-invariant sub-Riemannian structure. More precisely:*

*(i) if $\kappa = 0$ it is isometric to the Heisenberg group $\mathbb{H}$,*

*(ii) if $\kappa = 1$ it is isometric to the group $SU(2)$ with Killing metric,*

*(iii) if $\kappa = -1$ it is isometric to the group $\widetilde{SL}(2)$ with elliptic type Killing metric,*

*where $\widetilde{SL}(2)$ is the universal cover of $SL(2)$.*

Another byproduct of the classification is the fact that there exist non isomorphic Lie groups with locally isometric sub-Riemannian structures. Indeed, as a consequence of Theorem 17.29, we get that there exists a unique normalized left-invariant structure defined on $A^+(\mathbb{R}) \oplus \mathbb{R}$ having $\chi = 0, \kappa = -1$. Thus $A^+(\mathbb{R}) \oplus \mathbb{R}$ is locally isometric to the group $SL(2)$ with elliptic type Killing metric by Corollary 17.31.

In Section 17.5.4, we explicitly compute the global sub-Riemannian isometry between $A^+(\mathbb{R})\oplus\mathbb{R}$ and the universal cover of $SL(2)$. We then show that this map is well defined on the quotient, giving a global isometry between the group $A^+(\mathbb{R}) \times S^1$ and the group $SL(2)$, endowed with the sub-Riemannian structure defined by the restriction of the Killing form on the elliptic distribution.

The group $A^+(\mathbb{R}) \oplus \mathbb{R}$ can be interpreted as the subgroup of the affine maps on the plane that acts as an orientation preserving affinity on one axis and as translations on the other one.

$$A^+(\mathbb{R}) \oplus \mathbb{R} := \left\{ \begin{pmatrix} a & 0 & b \\ 0 & 1 & c \\ 0 & 0 & 1 \end{pmatrix},\ a > 0, b, c \in \mathbb{R} \right\}.$$

Notice that we can recover the action as an affine map identifying $(x, y) \in \mathbb{R}^2$ with $(x, y, 1)^T$ and

$$\begin{pmatrix} a & 0 & b \\ 0 & 1 & c \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = \begin{pmatrix} ax + b \\ y + c \\ 1 \end{pmatrix}.$$

The standard left-invariant sub-Riemannian structure on $A^+(\mathbb{R})\oplus\mathbb{R}$ is defined by the orthonormal frame $\mathcal{D} = \text{span}\{f_1, f_2\}$, where $f_1 = e_2$ and $f_2 = e_1 + e_3$ defined in terms of the following basis of the Lie algebra of the group

$$e_1 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad e_2 = \begin{pmatrix} -1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad e_3 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix},$$

Notice that $[e_1, e_2] = e_1$.

The subgroup $A^+(\mathbb{R})$ is topologically homeomorphic to the half-plane $\{(a, b) \in \mathbb{R}^2, a > 0\}$ which can be described in polar coordinates as $\{(\rho, \theta) |\, \rho > 0, -\pi/2 < \theta < \pi/2\}$.

To get a global sub-Riemannian isometry we should rather consider the group $A^+(\mathbb{R}) \times S^1$.

**Theorem 17.32.** *The diffeomorphism $\Psi : A^+(\mathbb{R}) \times S^1 \longrightarrow SL(2)$ defined by*

$$\Psi(\rho, \theta, \varphi) = \frac{1}{\sqrt{\rho \cos\theta}} \begin{pmatrix} \cos\varphi & \sin\varphi \\ \rho\sin(\theta - \varphi) & \rho\cos(\theta - \varphi) \end{pmatrix}, \tag{17.62}$$

*where $(\rho, \theta) \in A^+(\mathbb{R})$ and $\varphi \in S^1$, is a global sub-Riemannian isometry.*

Using this global sub-Riemannian isometry as a change of coordinates one can a priori recover the geometry of the sub-Riemannian structure on the group $A^+(\mathbb{R})\times S^1$ (as for instance the explicit expression of the sub-Riemannian distance, the cut locus, etc.) starting from the corresponding properties of $SL(2)$.

### 17.5.3 Canonical frames and classification. Proof of Theorem 17.29

In this section $G$ denotes a 3D Lie group, with Lie algebra $\mathfrak{g}$, endowed with a left-invariant sub-Riemannian structure defined by the orthonormal frame $f_1, f_2$, i.e.,

$$\mathcal{D} = \operatorname{span}\{f_1, f_2\} \subset \mathfrak{g}, \qquad \operatorname{span}\{f_1, f_2, [f_1, f_2]\} = \mathfrak{g}.$$

Recall that, for a 3D left-invariant structure, to be bracket generating is equivalent to be contact, moreover the Reeb field $f_0$ is also a left-invariant vector field by construction.

Since $\chi \geq 0$ by definition, we study separately the two cases $\chi > 0$ and $\chi = 0$. For each case, we compute the canonical frame and we show that the structure is locally isometric to one of the list given in Theorem 17.29 and Figure 17.1.

**Case** $\chi > 0$

Let $G$ be a 3D Lie group with a left-invariant sub-Riemannian structure such that $\chi > 0$. From Proposition 17.14, we can assume that $\mathcal{D} = \operatorname{span}\{f_1, f_2\}$, where $f_1, f_2$ is the canonical frame of the structure (globally defined by left-invariance). From (17.26) we obtain the dual equations

$$\begin{aligned} d\nu_0 &= \nu_1 \wedge \nu_2, \\ d\nu_1 &= c_{02}^1 \nu_0 \wedge \nu_2 + c_{12}^1 \nu_1 \wedge \nu_2, \\ d\nu_2 &= c_{01}^2 \nu_0 \wedge \nu_1 + c_{12}^1 \nu_1 \wedge \nu_2. \end{aligned} \tag{17.63}$$

Using $d^2\nu_i = d(d\nu_i) = 0$ for $i = 1, 2, 0$, we obtain the structure equations

$$\begin{cases} c_{02}^1 c_{12}^2 = 0, \\ c_{01}^2 c_{12}^1 = 0. \end{cases} \tag{17.64}$$

We know that the structure functions associated with the canonical frame are invariant by local isometries (up to changing the signs of $c_{12}^1, c_{12}^2$, see Remark 17.15). Hence, every different choice of coefficients in (17.26) which satisfies also (17.64) belongs to a different class of non-isometric structures.

Taking into account that $\chi > 0$ implies that $c_{01}^2$ and $c_{02}^1$ cannot be both non positive (see (17.27)), we have the following cases:

(i) $c_{12}^1 = 0$ and $c_{12}^2 = 0$. In this first case we get

$$\begin{aligned} [f_1, f_0] &= c_{01}^2 f_2, \\ [f_2, f_0] &= c_{02}^1 f_1, \\ [f_2, f_1] &= f_0, \end{aligned}$$

and formulas (17.27) imply

$$\chi = \frac{c_{01}^2 + c_{02}^1}{2} > 0, \qquad \kappa = \frac{c_{01}^2 - c_{02}^1}{2}.$$

In addition, we find the relations between the invariants

$$\chi + \kappa = c_{01}^2, \qquad \chi - \kappa = c_{02}^1.$$

We have the following subcases:

(a) If $c_{02}^1 = 0$ we get the Lie algebra $\mathfrak{se}(2)$ of the group $SE(2)$ of the Euclidean isometries of $\mathbb{R}^2$, satisfying $\chi = \kappa$.

(b) If $c_{01}^2 = 0$ this can be realized as the Lie algebra $\mathfrak{sh}(2)$ of the group $SH(2)$ of the hyperbolic isometries of $\mathbb{R}^2$, satisfying $\chi = -\kappa$.

(c) If $c_{01}^2 > 0$ and $c_{02}^1 < 0$ this can be realized as the Lie algebra $\mathfrak{su}(2)$ and $\chi - \kappa < 0$.

(d) If $c_{01}^2 < 0$ and $c_{02}^1 > 0$ this can be realized as the Lie algebra $\mathfrak{sl}(2)$ with $\chi + \kappa < 0$.

(e) If $c_{01}^2 > 0$ and $c_{02}^1 > 0$ this can be realized as the Lie algebra $\mathfrak{sl}(2)$ with $\chi+\kappa > 0, \chi-\kappa > 0$.

(ii) $c_{02}^1 = 0$ and $c_{12}^1 = 0$. In this case we have

$$\begin{aligned} [f_1, f_0] &= c_{01}^2 f_2, \\ [f_2, f_0] &= 0, \\ [f_2, f_1] &= c_{12}^2 f_2 + f_0, \end{aligned} \tag{17.65}$$

and necessarily $c_{01}^2 \neq 0$. Moreover we get

$$\chi = \frac{c_{01}^2}{2} > 0, \qquad \kappa = -(c_{12}^2)^2 + \frac{c_{01}^2}{2},$$

from which it follows

$$\chi - \kappa \geq 0. \tag{17.66}$$

The Lie algebra $\mathfrak{g} = \operatorname{span}\{f_1, f_2, f_0\}$ defined by (17.65) satisfies $\dim[\mathfrak{g}, \mathfrak{g}] = 2$. Moreover, notice that $A = \operatorname{ad} f_1$ is a well defined operator acting on the subspace $\operatorname{span}\{f_0, f_2\}$, satisfying

$$\operatorname{trace} A = -c_{12}^2, \qquad \det A = c_{01}^2 > 0,$$

together with the following relation

$$2\frac{\operatorname{trace}^2 A}{\det A} = 1 - \frac{\kappa}{\chi}. \tag{17.67}$$

(iii) $c_{01}^2 = 0$ and $c_{12}^2 = 0$. In this last case we get

$$\begin{aligned} [f_1, f_0] &= 0, \\ [f_2, f_0] &= c_{02}^1 f_1, \\ [f_2, f_1] &= c_{12}^1 f_1 + f_0, \end{aligned} \tag{17.68}$$

and $c_{02}^1 \neq 0$. Moreover we get

$$\chi = \frac{c_{02}^1}{2} > 0, \qquad \kappa = -(c_{12}^1)^2 - \frac{c_{02}^1}{2},$$

from which it follows

$$\chi + \kappa \leq 0. \tag{17.69}$$

The Lie algebra $\mathfrak{g} = \text{span}\{f_1, f_2, f_0\}$ defined by (17.68) satisfies $\dim[\mathfrak{g}, \mathfrak{g}] = 2$. Moreover, notice that $A = \text{ad } f_2$ is a well-defined operator on the subspace $\text{span}\{f_0, f_1\}$, satisfying

$$\text{trace } A = c_{12}^1, \qquad \det A = -c_{02}^1 < 0,$$

together with the relation

$$2\frac{\text{trace}^2 A}{\det A} = 1 + \frac{\kappa}{\chi}. \tag{17.70}$$

Lie algebras of cases (ii) and (iii) are *solvable* algebras and we will denote respectively $\mathfrak{solv}^+$ and $\mathfrak{solv}^-$, respectively. Notice that the sign stands for the sign of the determinant of the operator $A$ it represents. Indeed (17.66) and (17.67) imply $\det A > 0$, while (17.69) and (17.70) imply $\det A < 0$.

In particular, formulas (17.67) and (17.70) permits to recover the ratio between the two invariants (hence to determine a unique normalized structure) only from intrinsic properties of the operator $A$ introduced above.

Notice finally that the limit case of $\mathfrak{solv}^+$ with $c_{12}^2 = 0$ recovers the normalized structure (i.a), while the limit case of $\mathfrak{solv}^-$ with $c_{12}^1 = 0$ recovers the normalized structure (i.b).

**Case** $\chi = 0$

A direct consequence of Proposition 17.17 for left-invariant structures is the following.

**Corollary 17.33.** *Let $G, H$ be Lie groups with left-invariant sub-Riemannian structures and assume $\chi_G = \chi_H = 0$. Then $G$ and $H$ are locally isometric if and only if $\kappa_G = \kappa_H$.*

Thanks to this result it is easy to complete our classification. Indeed it is sufficient to find all left-invariant structures such that $\chi = 0$ and to compare their second invariant $\kappa$.

An elementary but long calculation shows that among the Lie algebras listed in the classification, the only that admit left-invariant structures with $\chi = 0$ are:

- $\mathfrak{h}$ is the Lie algebra of the Heisenberg group; then $\kappa = 0$.

- $\mathfrak{su}(2)$ with the Killing inner product; then $\kappa > 0$.

- $\mathfrak{sl}_e(2)$ with the elliptic distribution and Killing inner product; then $\kappa < 0$.

- $\mathfrak{a}^+(\mathbb{R}) \oplus \mathbb{R}$; then $\kappa < 0$.

In particular this implies that there exists a global sub-Riemannian isometry between the simply connected groups $\widetilde{SL}(2)$ (the universal cover of $SL(2)$) and $A^+(\mathbb{R}) \oplus \mathbb{R}$. In Section 17.5.4 we construct explictly this isometry and we show that it descends to a global isometry between $SL(2)$ (with the elliptic distribution and Killing inner product on it) and $A^+(\mathbb{R}) \times S^1$.

The proof of Theorem 17.29 is now completed (cf. again Figure 17.1).

**Exercise 17.34.** Prove that for every left-invariant sub-Riemannian structures on $\mathbb{H}$ one has $\chi = \kappa = 0$. It follows that every left-invariant sub-Riemannian structure on $\mathbb{H}$ is globally isometric to the standard one.

### 17.5.4 An explicit isometry. Proof of Theorem 17.32

In this section we write explicitly the sub-Riemannian isometry between $SL(2)$ and $A^+(\mathbb{R}) \times S^1$. Consider the Lie algebra $\mathfrak{sl}(2) = \{A \in M_2(\mathbb{R}),\ \mathrm{trace}(A) = 0\} = \mathrm{span}\{g_1, g_2, g_3\}$, where

$$g_1 = \frac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad g_2 = \frac{1}{2}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad g_3 = \frac{1}{2}\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.$$

The sub-Riemannian structure on $SL(2)$ defined by the Killing form on the elliptic distribution is given by the orthonormal frame $\{g_1, g_2\}$ on the distribution

$$\mathcal{D}_{\mathfrak{sl}} = \mathrm{span}\{g_1, g_2\}, \tag{17.71}$$

and $g_0 := -g_3$ is the Reeb vector field. Notice that this frame is canonical (in the sense of Section 17.3), since equations (17.32) are satisfied. Indeed

$$[g_1, g_0] = -g_2 = \kappa g_2.$$

Recall that the universal cover of $SL(2)$, which we denote $\widetilde{SL}(2)$, is a simply connected Lie group with Lie algebra $\mathfrak{sl}(2)$. Hence (17.71) define a left-invariant structure also on the universal cover.

On the other hand we consider the following coordinates on the Lie group $A^+(\mathbb{R}) \oplus \mathbb{R}$, that are well-adapted for our further calculations

$$A^+(\mathbb{R}) \oplus \mathbb{R} := \left\{ \begin{pmatrix} -y & 0 & x \\ 0 & 1 & z \\ 0 & 0 & 1 \end{pmatrix}, \quad y < 0, x, z \in \mathbb{R} \right\}. \tag{17.72}$$

It is easy to see that, in these coordinates, the group law reads

$$(x, y, z)(x', y', z') = (x - yx', -yy', z + z'),$$

and its Lie algebra $\mathfrak{a}(\mathbb{R}) \oplus \mathbb{R}$ is generated by the vector fields

$$e_1 = -y\partial_x, \quad e_2 = -y\partial_y, \quad e_3 = \partial_z,$$

with the only nontrivial commutator relation $[e_1, e_2] = e_1$.

The left-invariant structure on $A^+(\mathbb{R}) \oplus \mathbb{R}$ is defined by the orthonormal frame

$$\begin{aligned} &\mathcal{D}_{\mathfrak{a}} = \mathrm{span}\{f_1, f_2\}, \\ &f_1 := e_2 = -y\partial_y, \\ &f_2 := e_1 + e_3 = -y\partial_x + \partial_z. \end{aligned} \tag{17.73}$$

With straightforward calculations we compute the Reeb vector field $f_0 = -e_3 = -\partial_z$.

This frame is not canonical since it does not satisfy equations (17.32). Hence we can apply Proposition 17.17 to find the canonical frame, that will be no more left-invariant. Following the notation of Proposition 17.17 we have

**Lemma 17.35.** *The canonical orthonormal frame on $A^+(\mathbb{R}) \oplus \mathbb{R}$ has the form:*

$$\begin{aligned} \widehat{f}_1 &= \quad y \sin z\, \partial_x - y \cos z\, \partial_y - \sin z\, \partial_z, \\ \widehat{f}_2 &= -y \cos z\, \partial_x - y \sin z\, \partial_y + \cos z\, \partial_z. \end{aligned} \tag{17.74}$$

*Proof.* It is enough to show that the rotation defined in the proof of Proposition 17.17 is $\theta(x,y,z) = z$. The dual basis to our frame $\{f_1, f_2, f_0\}$ is given by

$$\nu_1 = -\frac{1}{y}dy, \qquad \nu_2 = -\frac{1}{y}dx, \qquad \nu_0 = -\frac{1}{y}dx - dz.$$

Moreover we have $[f_1, f_0] = [f_2, f_0] = 0$ and $[f_2, f_1] = f_2 + f_0$ so that, in (17.35) we get $c = 0$, $\alpha_1 = 0$ and $\alpha_2 = 1$. Hence

$$d\theta = -\nu_0 + \nu_2 = dz.$$

$\square$

Now we have two canonical frames $\{\widehat{f}_1, \widehat{f}_2, f_0\}$ and $\{g_1, g_2, g_0\}$, whose Lie algebras satisfy the same commutation relations:

$$\begin{aligned} &[\widehat{f}_1, f_0] = -\widehat{f}_2, \qquad && [g_1, g_0] = -g_2, \\ &[\widehat{f}_2, f_0] = \widehat{f}_1, && [g_2, g_0] = g_1, \\ &[\widehat{f}_2, \widehat{f}_1] = f_0, && [g_2, g_1] = g_0. \end{aligned} \tag{17.75}$$

Let us consider the two maps

$$\widetilde{F} : \mathbb{R}^3 \to A^+(\mathbb{R}) \oplus \mathbb{R}, \qquad (t_1, t_2, t_0) \mapsto e^{t_0 f_0} \circ e^{t_2 \widehat{f}_2} \circ e^{t_1 \widehat{f}_1}(1_A), \tag{17.76}$$

$$\widetilde{G} : \mathbb{R}^3 \to SL(2), \qquad (t_1, t_2, t_0) \mapsto e^{t_0 g_0} \circ e^{t_2 g_2} \circ e^{t_1 g_1}(1_{SL}). \tag{17.77}$$

where we denote with $1_A$ and $1_{SL}$ identity element of $A^+(\mathbb{R}) \oplus \mathbb{R}$ and $\widetilde{SL}(2)$, respectively.

To simplify computation we introduce the rescaled maps

$$F(t) := \widetilde{F}(2t), \qquad G(t) := \widetilde{G}(2t), \qquad t = (t_1, t_2, t_0),$$

and solving the corresponding differential equations, we get from (17.76) the following expressions

$$F(t_1, t_2, t_0) = \left( 2e^{-2t_1} \frac{\tanh t_2}{1 + \tanh^2 t_2}, \ -e^{-2t_1} \frac{1 - \tanh^2 t_2}{1 + \tanh^2 t_2}, \ 2(\arctan(\tanh t_2) - t_0) \right). \tag{17.78}$$

The function $F$ is globally invertible on its image and its inverse

$$F^{-1}(x,y,z) = \left( -\frac{1}{2} \log \sqrt{x^2 + y^2}, \ \operatorname{arctanh}\left( \frac{y + \sqrt{x^2 + y^2}}{x} \right), \ \arctan\left( \frac{y + \sqrt{x^2 + y^2}}{x} \right) - \frac{z}{2} \right).$$

is defined for every $y < 0$ and for every $x$ (it is extended by continuity at $x = 0$).

On the other hand, the map (17.77) can be expressed as the product of exponential matrices as follows (recall that, since we consider left-invariant systems, we must multiply matrices from the right)

$$G(t_1, t_2, t_0) = \begin{pmatrix} e^{t_1} & 0 \\ 0 & e^{-t_2} \end{pmatrix} \begin{pmatrix} \cosh t_2 & \sinh t_2 \\ \sinh t_2 & \cosh t_2 \end{pmatrix} \begin{pmatrix} \cos t_0 & -\sin t_0 \\ \sin t_0 & \cos t_0 \end{pmatrix}. \tag{17.79}$$

To compute $G \circ F^{-1}$, we consider on the half-plane $\{(x,y), y<0\}$ the standard polar coordinates $(\rho,\theta)$, where $-\pi/2<\theta<\pi/2$ is the angle that the point $(x,y)$ defines with $y$-axis. Let us introduce the quantity

$$\xi(\theta) := \tan\frac{\theta}{2} = \begin{cases} \dfrac{y+\sqrt{x^2+y^2}}{x}, & \text{if} \quad x \neq 0, \\ 0, & \text{if} \quad x = 0. \end{cases}$$

Hence we can rewrite $F^{-1}$ as follows

$$F^{-1}(\rho,\theta,z) = \left(-\frac{1}{2}\log\rho,\ \operatorname{arctanh}\xi(\theta),\ \arctan\xi(\theta) - \frac{z}{2}\right),$$

and compute the composition $\Psi := G \circ F^{-1} : A^+(\mathbb{R}) \oplus \mathbb{R} \longrightarrow SL(2)$. Once we substitute these expressions in (17.79), the third factor is a rotation matrix by an angle $\arctan\xi(\theta) - z/2$. Splitting this matrix in two consecutive rotations and using standard trigonometric identities (we use the shorthand $\xi = \xi(\theta)$) we obtain:

$$\Psi(\rho,\theta,z) =$$

$$= \begin{pmatrix} \rho^{-1/2} & 0 \\ 0 & \rho^{1/2} \end{pmatrix} \begin{pmatrix} \dfrac{1}{\sqrt{1-\xi^2}} & \dfrac{\xi}{\sqrt{1-\xi^2}} \\ \dfrac{\xi}{\sqrt{1-\xi^2}} & \dfrac{1}{\sqrt{1-\xi^2}} \end{pmatrix} \begin{pmatrix} \dfrac{1}{\sqrt{1+\xi^2}} & -\dfrac{\xi}{\sqrt{1+\xi^2}} \\ \dfrac{\xi}{\sqrt{1+\xi^2}} & \dfrac{1}{\sqrt{1+\xi^2}} \end{pmatrix} \begin{pmatrix} \cos\dfrac{z}{2} & \sin\dfrac{z}{2} \\ -\sin\dfrac{z}{2} & \cos\dfrac{z}{2} \end{pmatrix}.$$

Then using identities: $\cos\theta = \dfrac{1-\xi^2}{1+\xi^2}$, $\sin\theta = \dfrac{2\xi}{1+\xi^2}$, we get

$$\Psi(\rho,\theta,z) = \begin{pmatrix} \rho^{-1/2} & 0 \\ 0 & \rho^{1/2} \end{pmatrix} \begin{pmatrix} \dfrac{1+\xi^2}{\sqrt{1-\xi^4}} & 0 \\ \dfrac{2\xi}{\sqrt{1-\xi^4}} & \dfrac{1-\xi^2}{\sqrt{1-\xi^4}} \end{pmatrix} \begin{pmatrix} \cos\dfrac{z}{2} & \sin\dfrac{z}{2} \\ -\sin\dfrac{z}{2} & \cos\dfrac{z}{2} \end{pmatrix}$$

$$= \sqrt{\frac{1+\xi^2}{1-\xi^2}} \begin{pmatrix} \rho^{-1/2} & 0 \\ 0 & \rho^{1/2} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ \dfrac{2\xi}{1+\xi^2} & \dfrac{1-\xi^2}{1+\xi^2} \end{pmatrix} \begin{pmatrix} \cos\dfrac{z}{2} & \sin\dfrac{z}{2} \\ -\sin\dfrac{z}{2} & \cos\dfrac{z}{2} \end{pmatrix}$$

$$= \frac{1}{\sqrt{\rho\cos\theta}} \begin{pmatrix} 1 & 0 \\ 0 & \rho \end{pmatrix} \begin{pmatrix} 1 & 0 \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} \cos\dfrac{z}{2} & \sin\dfrac{z}{2} \\ -\sin\dfrac{z}{2} & \cos\dfrac{z}{2} \end{pmatrix}$$

$$= \frac{1}{\sqrt{\rho\cos\theta}} \begin{pmatrix} \cos\dfrac{z}{2} & \sin\dfrac{z}{2} \\ \rho\sin\left(\theta - \dfrac{z}{2}\right) & \rho\cos\left(\theta - \dfrac{z}{2}\right) \end{pmatrix}.$$

**Lemma 17.36.** *The set $\Psi^{-1}(1_{SL})$ is a normal subgroup of $A^+(\mathbb{R}) \oplus \mathbb{R}$.*

*Proof.* It is easy to show that $\Psi^{-1}(1_{SL}) = \{F(0,0,2k\pi) \mid k \in \mathbb{Z}\}$. From (17.78) we see that $F(0,0,2k\pi) = (0,-1,-4k\pi)$ and (17.72) implies that $\Psi^{-1}(1_{SL})$ is a normal subgroup. Indeed it is enough to prove that $\Psi^{-1}(1_{SL})$ is a subgroup of the center, and this follows from the identity

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 4k\pi \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} -y & 0 & x \\ 0 & 1 & z \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} -y & 0 & x \\ 0 & 1 & z+4k\pi \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} -y & 0 & x \\ 0 & 1 & z \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 4k\pi \\ 0 & 0 & 1 \end{pmatrix}.$$

□

*Remark* 17.37. Using a topological argument it is possible to prove that actually $\Psi^{-1}(A)$ is a discrete countable set for every $A \in SL(2)$, and $\Psi$ is a representation of $A^+(\mathbb{R}) \oplus \mathbb{R}$ as universal cover of $SL(2)$.

By Lemma 17.36, the map $\Psi$ is a well-defined isomorphism between the quotient

$$\frac{A^+(\mathbb{R}) \oplus \mathbb{R}}{\Psi^{-1}(I)} \simeq A^+(\mathbb{R}) \times S^1,$$

and the group $SL(2)$, defined by restriction of $\Psi$ to $z \in [-2\pi, 2\pi]$. If we consider the new variable $\varphi := z/2$, defined on $[-\pi, \pi]$, we can finally write the global isometry as

$$\Psi(\rho,\theta,\varphi) = \frac{1}{\sqrt{\rho\cos\theta}} \begin{pmatrix} \cos\varphi & \sin\varphi \\ \rho\sin(\theta-\varphi) & \rho\cos(\theta-\varphi) \end{pmatrix}, \tag{17.80}$$

where $(\rho,\theta) \in A^+(\mathbb{R})$ and $\varphi \in S^1$.

Notice that, in the coordinates introduced above, we have $1_A = (1,0,0)$ and

$$\Psi(1_A) = \Psi(1,0,0) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = 1_{SL}.$$

On the other hand it is an easy observation that

**Lemma 17.38.** *$\Psi$ is not a group homomorphism*

Indeed one can easily check that on $A^+(\mathbb{R}) \oplus \mathbb{R}$ one has

$$\left(\frac{\sqrt{2}}{2}, \frac{\pi}{4}, \pi\right)\left(\frac{\sqrt{2}}{2}, -\frac{\pi}{4}, -\pi\right) = 1_A,$$

while from (17.80) one gets

$$\Psi\left(\frac{\sqrt{2}}{2}, \frac{\pi}{4}, \pi\right)\Psi\left(\frac{\sqrt{2}}{2}, -\frac{\pi}{4}, -\pi\right) = \begin{pmatrix} 2 & 0 \\ 1/2 & 1/2 \end{pmatrix} \neq 1_{SL}.$$

## 17.6 Appendix: Remarks on curvature coefficients

In this appendix we recall the relation of the geometric invariants $\chi$ and $\kappa$ defined in Section 17.2, with the invariants of a canonical connection that is possible to define on a 3D contact sub-Riemannian structures.

We extend the sub-Riemannian metric on $\mathcal{D}$ to a global Riemannian structure (that we denote with the symbol $g$) by promoting $X_0$ to an unit vector orthogonal to $\mathcal{D}$. We define the *contact endomorphism* $J : TM \to TM$ by:

$$g(X, JY) = d\omega(X, Y), \qquad \forall X, Y \in \Gamma(TM). \tag{17.81}$$

Clearly $J$ is skew-symmetric w.r.t. to $g$. In the 3-dimensional case, the previous condition forces $J^2 = -\mathbb{I}$ on $\mathcal{D}$ and $J(X_0) = 0$.

On a 3D contact sub-Riemannian manifold, it is possible to fix a canonical linear connection, which is different from the Levi-Civita connection associated with $g$ (it is not torsion-free). This connection is possible to define on a general sub-Riemannian contact manifold, and is called *Tanno* connection.

The conditions are written in terms of covariant derivative of tensors. Recall that, given a tensor $A$ of order $r$ and a linear connection $\nabla$, we define $\nabla A$ as a $r+1$ tensor defined by

$$\nabla A(X_1, \ldots, X_r, Y) := Y(T(X_1, \ldots, X_r)) - A(\nabla_Y X_1, \ldots, X_r) - \ldots - A(X_1, \ldots, \nabla_Y X_r).$$

where $X_1, \ldots, X_r$ and $Y$ are vector fields.

**Exercise 17.39** (On the Tanno connection)**.** Prove that, given a three-dimensional contact manifold with normalized contact form $\omega$, there exists a unique linear connection $\nabla$ such that for every $Z \in \Gamma(TM)$

(i) $\nabla\omega = 0$,

(ii) $\nabla X_0 = 0$,

(iii) $\nabla g = 0$,

and such that, denoting $T$ the torsion tensor associated with $\nabla$,

(iv) $T(X, Y) = d\omega(X, Y)X_0$ for any $X, Y \in \Gamma(\mathcal{D})$,

(v) $T(X_0, JX) = -JT(X_0, X)$ for any vector field $X \in \Gamma(TM)$,

Notice that, if $X$ is a horizontal vector field, $T(X_0, X)$ is horizontal as well. As a consequence, if we define $\tau(X) := T(X_0, X)$, $\tau$ is a symmetric endomorphism defined on the distribution, which satisfies $\tau \circ J + J \circ \tau = 0$, by property (v). Notice that $\mathrm{trace}(\tau) = 0$ and $\det(\tau) \le 0$. A standard computation gives the following result, whose proof is left as an exercice.

**Lemma 17.40.** *Let $R^\nabla$ be the curvature associated with the connection $\nabla$. Then*

$$\kappa = R^\nabla(X_1, X_2, X_2, X_1), \qquad \chi = \sqrt{-\det(\tau)}. \tag{17.82}$$

*where $X_1, X_2$ is a local orthonormal frame for the sub-Riemannian structure.*

## 17.7 Bibliographical note

The study of (complete sets of) metric invariants, connected with the problem of equivalence of 3D sub-Riemannian contact structures, has been considered in the literature in different context and with different languages, in particular in [Agr95, Agr96], in [Hug95] and [FG96].

In [Hug95] the authors introduces a generating set of invariants using the Cartan's moving frame method, while in [FG96] the authors introduce invariants associated with the canonical connection discussed in Section 17.6. These invariants recover $\kappa$ and $\chi$ respectively, up to normalization constants. For a more detailed discussion on the explicit relation we refer to [BBLss, Appendix A] (cf. also [ABR18, Section 7.5]).

The computation of the sub-Riemannian curvature on the three-dimensional case, following the definition given here but with a different notation, is done in [AL14]. The relation between local invariants and volume of small balls and small time heat kernel asymptotics on the diagonal have also been considered in [BBLss] and [Bar13], respectively. The reader is referred also to [ABR18] and [ABR17] for a discussion of the higher dimensional contact case.

The complete classification already appeared in Falbel and Gorodski [FG96], where the authors classify sub-Riemannian homogeneous spaces (i.e., sub-Riemannian structures which admits a transitive Lie group of isometries acting smoothly on the manifold) in dimension 3 and 4. The local classification of 3D left-invariant structures given here follows the approach given in [Agr95, AB12]. Three-dimensional sub-Riemannian symmetric spaces have been classified also in [Str86, Section 10].

A byproduct of the classification is an explicit sub-Riemannian isometry between two non-isomorphic Lie groups. The explicit expression given here is taken from [AB12] (cf. also [FG96, Remark 3.1]).

A more detailed discussion of the Nagano principle and its relation with the Orbit theorem can be found in the book [AS04], cf. also the papers [Nag66, Sus74, Sus83].

# Chapter 18

# Integrability of the sub-Riemannian geodesic flow on 3D Lie groups

In this chapter we show how to find certain first integrals for Hamiltonian systems on Lie groups, that are automatically in involution among them and with the Hamiltonian. We will study the so-called Casimir first integral and other first integrals whose presence is a consequence of the commutativity of left-invariant vector fields with right-invariant ones. This theory will be used to prove that the Hamiltonian system for normal Pontryagin extremals for rank-2 left-invariant sub-Riemannian structures on 3D Lie groups is completely integrable in the sense of Section 5.4.

## 18.1 Poisson manifolds and symplectic leaves

### Poisson manifolds

**Definition 18.1.** A *Poisson manifold* is a connected differentiable manifold $P$ endowed with an operation $\{\cdot,\cdot\} : \mathcal{C}^\infty(P) \times \mathcal{C}^\infty(P) \to \mathcal{C}^\infty(P)$ (called *Poisson bracket*) which satisfies the following properties:

- Bilinearity: $\{\cdot,\cdot\}$ is linear in both arguments;
- Antisymmetricity: $\{a, b\} = -\{b, a\}$;
- Leibniz rule: $\{ab, c\} = a\{b, c\} + \{a, c\}b$;
- Jacobi identity: $\{\{a, b\}, c\} + \{\{c, a\}, b\} + \{\{b, c\}, a\} = 0$.

The pair $(\mathcal{C}^\infty(P), \{\cdot,\cdot\})$ is called a *Poisson algebra*.

Given $a \in \mathcal{C}^\infty(P)$, as a consequence of the Leibniz rule in Definition 18.1, the operator $\{a, \cdot\}$ is a derivation and hence it defines a vector field. Hence we can give the following definition.

**Definition 18.2.** Given $a \in \mathcal{C}^\infty(P)$ we define the corresponding *Poisson vector field* as

$$\vec{a}(b) = \{a, b\}, \qquad \text{for every } b \in \mathcal{C}^\infty(P). \tag{18.1}$$

The set of all Poisson vector fields on a Poisson manifold is denoted $\mathscr{P}$.

In the previous definition we have used the same symbol as the one that we used for the Hamiltonian vector field since, as we will see later on, they coincide on symplectic manifolds.

## The Poisson bi-vector

The description of a Poisson manifold in terms of Poisson brackets is not always the most efficient one. There is an alternative description in terms of 2-vector fields, i.e., maps $\Lambda^1(P) \times \Lambda^1(P) \to \mathcal{C}^\infty(P)$. Actually, since $\{\cdot,\cdot\}$ is a derivation in each argument, it can be interpreted as a vector field both when acting on its first, or in its second argument. As a consequence the value of $\{a,b\}(q)$ depends only $da(q)$ and $db(q)$. We can then give the following definition.

**Definition 18.3.** Let $(P,\{\cdot,\cdot\})$ be a Poisson manifold. The operator $\Pi : \Lambda^1(P) \times \Lambda^1(P) \to \mathcal{C}^\infty(P)$ such that for every $a, b \in \mathcal{C}^\infty(P)$ we have $\Pi(da, db) = \{a,b\}$ is called *Poisson bi-vector.*

Notice that $\Pi$ is skew-symmetric and hence can be interpreted as a map $\Lambda^2(P) \to \mathcal{C}^\infty(P)$.

In coordinates $\{x_i\}$ we can write $\Pi(x) = \sum_{i,j} \Pi_{ij}(x)\partial_{x_i} \otimes \partial_{x_j}$. From the relation $\Pi(dx_i, dx_j) = \{x_i, x_j\}$ we get that $\Pi_{ij}$ is the antisymmetric matrix defined by $\Pi_{ij} = \{x_i, x_j\}$. It is then clear that the relation $\Pi(da, db) = \{a,b\}$ defines $\Pi$ on every element of $\Lambda^1(P) \times \Lambda^1(P)$ and not only on differential of functions.

*Remark* 18.4. Notice that, in general, given a bi-vector $\Upsilon(x) = \sum_{i,j} \Upsilon_{ij}\partial_{x_i} \otimes \partial_{x_j}$ with $(\Upsilon_{ij})$ skew-symmetrc, it does not satisfy the Jacobi identity and hence it does not induce a structure of Poisson manifold.

One immediately verifies the validity of following Proposition. Recall that the bi-vector $\Pi$ is non-degenerate if for every $q \in P$ we have the following: the relation $\Pi_q(\lambda_1, \lambda_2) = 0$ for every $\lambda_2 \in T_q^*P$ implies $\lambda_1 = 0$

**Proposition 18.5.** *Let $(P,\{\cdot,\cdot\})$ be a Poisson manifold and let $\Pi$ be the corresponding bi-vector. Assume that $\Pi$ is non-degenerate. Then the Poisson bracket is non-degenerate in the following sense*

$$\{a,b\} = 0 \textit{ for every } b \in \mathcal{C}^\infty(N), \textit{ implies } a = \textit{const.} \tag{18.2}$$

Notice, however, that (18.2) does not imply the non degeneracy of $\Pi$ (cf. Remark 18.13).

## Symplectic manifolds

Every symplectic manifold $(N,\sigma)$, and in particular every cotangent bundle $T^*M$ to a smooth manifold $M$, is a Poisson manifold with the usual Poisson bracket

$$\{a,b\} = \sigma(\vec{a}, \vec{b}). \tag{18.3}$$

Here $\vec{a}$ is the Hamiltonian vector field corresponding to the function $a$ via $\sigma$, i.e., $\sigma(\cdot, \vec{a}) = da$.

Let $(N,\sigma)$ be a symplectic manifold. The symplectic form $\sigma$, that is a map from $\mathrm{Vec}(N) \times \mathrm{Vec}(N) \to \mathcal{C}^\infty(N)$, can be regarded as a map from $\mathrm{Vec}(N) \to \Lambda^1(N)$, if we think to the application $X \mapsto \sigma(X, \cdot)$.

In a similar way, on a Poisson manifold $(P,\{\cdot,\cdot\})$, the bi-vector $\Pi$, that is a map from $\Lambda^1(P) \times \Lambda^1(P) \to \mathcal{C}^\infty(P)$, can be regarded as a map from $\Lambda^1(P) \to \mathrm{Vec}(P)$, if we think to the application $\omega \mapsto \Pi(\omega, \cdot)$.

It is then a simple exercise to prove the following.

**Proposition 18.6.** *Let $(N,\sigma)$ be a symplectic manifold and $\Pi$ the corresponding bi-vector. If we regard $\sigma$ as map from $\mathrm{Vec}(N)$ to $\Lambda^1(N)$ and $\Pi$ as map from $\Lambda^1(N)$ to $\mathrm{Vec}(N)$ then $\Pi = \sigma^{-1}$, i.e., in coordinates $(\Pi_{ij}) = (\sigma_{ij})^{-1}$.*

Hence for a symplectic manifold, as a consequence of the fact that $\sigma$ is non-degenerate, the bi-vector and the Poisson bracket are non-degenerate as well.

Notice that, although every symplectic manifold is a Poisson manifold, the opposite is not true. For instance, any manifold endowed with the operation $\{a, b\} = 0$ for every $a, b \in \mathcal{C}^\infty(P)$ is a Poisson manifold, but this bracket does not come from a symplectic structure through (18.3) since it is degenerate. Indeed the following holds.

**Proposition 18.7.** *Let $(P, \{\cdot,\cdot\})$ be a Poisson manifold and let $\Pi$ be the corresponding bi-vector. Then $(P, \{\cdot,\cdot\})$ is symplectic (i.e., there exists a symplectic form $\sigma$ on $P$ such that $\{a, b\} = \sigma(\vec{a}, \vec{b})$, for every $a, b \in \mathcal{C}^\infty(P)$) if and only if $\Pi$ is non-degenerate.*

*Proof.* The implication "$\Rightarrow$" is obvious. For the opposite implication one defines $\sigma = \Pi^{-1}$ and one verifies that the closure of $\sigma$ is guaranteed by the Jacobi identity in Definition 18.1. ☐

## Casimir functions

**Definition 18.8.** Let $(P, \{\cdot,\cdot\})$ be a Poisson manifold. A function $\mathscr{C} \in \mathcal{C}^\infty$ such that $\{\mathscr{C}, b\} = 0$ for every $b \in \mathcal{C}^\infty$ is called a *Casimir function.*

*Remark* 18.9. Notice that, if $\mathscr{C}$ is a Casimir function, then for every $c \in \mathbb{R}$, we have that $c\mathscr{C}$ is a Casimir function as well.

As a consequence of Definition 18.2 we have:

**Proposition 18.10.** *The Poisson vector field corresponding to a Casimir function is indentically zero.*

As a consequence of Proposition 18.7 we have the following.

**Proposition 18.11.** *On a symplectic manifold, the only Casimir functions are constant functions.*

*Remark* 18.12. Notice that on Poisson manifolds there may exist zero Poisson vector fields that correspond to non-constant functions (that are exactly those corresponding to non-constant Casimr functions). However this cannot occur if the manifold is symplectic.

*Remark* 18.13. Notice, however, that there exist Poisson manifolds that are not symplectic for which the only Casimir functions are the constant functions. Consider for instance the symplectic manifold $(\mathbb{R}^2, dx \wedge dy)$ and the corresponding Poisson bi-vector $\Pi = \partial_x \wedge \partial_y$. Let $a \in \mathcal{C}^\infty(\mathbb{R}^2)$ be a function that is zero in a single point. Then one immedately verfiy that $a\Pi$ is a bi-vector that gives to $\mathbb{R}^2$ the structure of Poisson manifold. However in such a Poisson manifolds the only Casimir functions are the constant functions.

As for symplectic manifolds we have (cf. Corollary 4.57),

**Proposition 18.14.** *For every $a, b \in \mathcal{C}^\infty(P)$ we have $[\vec{a}, \vec{b}] = \overrightarrow{\{a, b\}}$.*

In other words the Lie bracket of two Poisson vector fields is still a Poisson vector field. As a consequence

- the set of Poisson vector felds $\mathscr{P}$ is a Lie sub-algebra of $\mathrm{Vec}(P)$;
- the application $\mathcal{C}^\infty(P) \ni a \mapsto \vec{a} \in \mathscr{P}$ is an homomorphism from the Poisson algebra $(\mathcal{C}^\infty(P), \{\cdot,\cdot\})$ onto the Lie algebra $(\mathscr{P}, [\cdot,\cdot])$.

As for symplectic manifolds, given $a, b \in \mathcal{C}^\infty(P)$ we have

$$\frac{d}{dt} b(e^{t\vec{a}}(q)) = \vec{a}(b)|_{e^{t\vec{a}}(q))} = \{a, b\}(e^{t\vec{a}}(q)).$$

Hence the following holds.

**Proposition 18.15.** *Given $a, b \in \mathcal{C}^\infty(P)$, if $\{a, b\} = 0$ then $b$ is constant on the integral curves of $\vec{a}$ and $a$ is constant on the integral curves of $\vec{b}$. In particular a Casimir function is constant on the integral curves of every Poisson vector field.*

## Symplectic leaves

Consider the distribution (in general rank-varying) generated by the set $\mathscr{P}$ of all Poisson vector fields

$$\mathscr{P}_q = \operatorname{span}\{\vec{a}(q) \mid \vec{a} \in \mathscr{P}\}. \tag{18.4}$$

As a consequence of Proposition 18.14, we have that, if $\vec{a_1}, \vec{a_2} \in \mathscr{P}$, then $[\vec{a_1}, \vec{a_2}] \in \mathscr{P}$. We can rephrase this by saying that $\mathscr{P}_q$ is a (rank-varying) *involutive distribution.*

For every $q \in P$ consider the *orbit* associated with $\mathscr{P}$,

$$\mathcal{O}(q) = \{\bar{q} \in P \mid \exists\, \ell \in \mathbb{N}, t_1, \ldots, t_\ell \in \mathbb{R}, \vec{a}_1, \ldots, \vec{a}_\ell \in \mathscr{P} \text{ such that } \bar{q} = e^{t_\ell \vec{a}_\ell} \circ \ldots \circ e^{t_1 \vec{a}_1}(q)\}.$$

We have the following.

**Theorem 18.16.** *For every $q \in P$, $\mathcal{O}(q)$ is an immersed submanifold of $P$. Moreover for every $\bar{q} \in \mathcal{O}(q)$, we have $T_{\bar{q}}\mathcal{O}(q) = \mathscr{P}_{\bar{q}}$.*

*Proof.* If the distribution (18.4) is of constant rank then, being involutive, it satisfies the hypotheses of Frobenius Theorem 2.37 and the conclusion follows.

If the distribution (18.4) is rank-varying, similar arguments apply, but one should use a generalization of Frobenius Theorem for rank-varying involutive distribution. See for instance [AS04, p. 76]. □

**Theorem 18.17.** *For every $q \in P$ we have that $\mathcal{O}(q)$ is a symplectic manifold with (non-degenerate) Poisson bracket given by the restriction on $\mathcal{O}(q)$ of the Poisson bracket defined on $P$.*

*Proof.* Let us set $N = \mathcal{O}(q)$. It is enough to prove that for every $\bar{q} \in N$, the restriction of the Poisson bi-vector $\Pi_{\bar{q}}$ to $N$ is non-degenerate. Consider the following linear map

$$W : T^*_{\bar{q}}N \ni w \mapsto \vec{a}(\bar{q}) \in T_{\bar{q}}N,$$

where $a$ is any function in $\mathcal{C}^\infty(N)$ such that $d_{\bar{q}}a = w$. We have that $\mathscr{P}_{\bar{q}}$ is the image of $W$. Since $\mathscr{P}_{\bar{q}} = T_{\bar{q}}N$ we have that $W$ is surjective, hence $W$ has no kernel and $\vec{a}(\bar{q}) \neq 0$ for every $w \neq 0$.

Assume by contradiction that $\Pi_{\bar{q}}$ is degenerate, then there exists $w \in T^*_{\bar{q}}N$, $w \neq 0$, such that $\Pi_{\bar{q}}(w, \bar{w}) = 0$ for every $\bar{w} \in T^*_{\bar{q}}N$. Now if $a$ and $b$ belongs to $\mathcal{C}^\infty(N)$ and are such that $w = d_{\bar{q}}a$ and $\bar{w} = d_{\bar{q}}b$, we have

$$\vec{a}(b)_{\bar{q}} = \{a, b\}_{\bar{q}} = \Pi_{\bar{q}}(da, db) = \Pi_{\bar{q}}(w, \bar{w}) = 0.$$

Being $b$ arbitrary this would implies $\vec{a}(\bar{q}) = 0$. Contradiction. □

**Corollary 18.18.** *If $(P, \{\cdot,\cdot\})$ is symplectic, then for every $q \in P$ we have $\mathcal{O}(q) = P$.*

*Proof.* If $(P, \{\cdot,\cdot\})$ is symplectic, then for every $q \in P$ we have $\mathscr{P}_q = T_qP$. Hence the distribution has full rank and, as a consequence of the Raschevskii-Chow theorem, we have that $\mathcal{O}(q) = P$. ☐

*Remark* 18.19. $\mathcal{O}(q)$ could be zero-dimensional (as it happens for instance if the Poisson bracket is identically zero).

*Remark* 18.20. Each orbit is called a *symplectic leaf.* Symplectic leaves can have different dimensions, however since they are symplectic, they are even-dimensional. Casimir functions are constant on symplectic leaves.

## 18.2 Integrability of Hamiltonian systems on Lie groups

In the rest of this chapter we work on a Lie group $G$ and we denote by $\mathfrak{g}$ the Lie algebra of its left-invariant vector fields. We follow the notation of Chapter 7. We identify $T_{\mathbb{1}}G$ with $\mathfrak{g}$ and we denote $\mathfrak{g}^*$ the dual of $\mathfrak{g}$.

### 18.2.1 The Poisson manifold $\mathfrak{g}^*$

Recall that $\mathfrak{g}$ has a structure of Lie algebra, that is not possible to transfer canonically on $\mathfrak{g}^*$ without an additional structure. The space $\mathfrak{g}^*$ is the most important example of Poisson manifold. Indeed one immediately verify the validity of the following proposition.

**Proposition 18.21.** *For every $a, b \in \mathcal{C}^\infty(\mathfrak{g}^*)$, define the bracket $\{a, b\}_{\mathfrak{g}^*}(\xi) = \langle \xi, [da, db]\rangle$, $\xi \in \mathfrak{g}^*$. Then $(\mathfrak{g}^*, \{\cdot,\cdot\}_{\mathfrak{g}^*})$ is a Poisson manifold.*

Here $da, db \in \mathfrak{g}^{**}$ and hence can be identified with elements of $\mathfrak{g}$. The bracket $\{\cdot,\cdot\}_{\mathfrak{g}^*}(\xi)$ is called the *Lie-Poisson bracket.*

**Proposition 18.22.** *Let $H_1$ and $H_2$ be two left-invariant Hamiltonian on $T^*G$, i.e. $H_1(p, g) = \mathcal{H}_1(L_g^*p)$ and $H_2(p, g) = \mathcal{H}_2(L_g^*p)$, where $\mathcal{H}_1, \mathcal{H}_1$ are two functions $\mathfrak{g}^* \to \mathbb{R}$. Then $\{H_1, H_2\}$ is left-invariant and*

$$\{H_1, H_2\}|_{(p,g)} = \{\mathcal{H}_1, \mathcal{H}_2\}_{\mathfrak{g}^*}|_{(L_g^*p)}. \tag{18.5}$$

*Proof.* Let $h_1, \dots, h_n$ be the vertical coordinates induced by a basis $e_1^*, \dots, e_n^*$ of $\mathfrak{g}^*$. In matrix notation $h_i(p, g) = \langle p, g\, e_i\rangle$. Then

$$\begin{aligned}\{H_1, H_2\} &= \sum_{i,j=1}^n \frac{\partial H_1}{\partial h_i}\frac{\partial H_2}{\partial h_j}\{h_i, h_j\} = \sum_{i,j=1}^n \frac{\partial H_1}{\partial h_i}\frac{\partial H_2}{\partial h_j}\langle p, [g\, e_i, g\, e_j]\rangle \\ &= \sum_{i,j=1}^n \frac{\partial \mathcal{H}_1}{\partial h_i}\frac{\partial \mathcal{H}_2}{\partial h_j}\langle p, [g\, e_i, g\, e_j]\rangle. \end{aligned} \tag{18.6}$$

In the last formula we have used the fact that in vertical coordinates $H(g, h_1, \dots, h_n) = \mathcal{H}(h_1, \dots, h_n)$. Now the dependence on $g$ is either through $h_i$ that are left-invariant or through $\langle p, [g\, e_i, g\, e_j]\rangle$ that is left-invariant as well since $[g\,e_i, g\,e_j] = g[e_i, e_j]$. Hence $\{H_1, H_2\}$ is left-invariant. To prove Formula (18.5) it is then sufficient to prove it for $g = \mathbb{1}$. At the identity $(p, g) = (\xi, \mathbb{1})$, formula (18.6) becomes

$$\{H_1, H_2\}_{(\xi,\mathbb{1})} = \sum_{i,j=1}^{n} \frac{\partial H_1}{\partial h_i} \frac{\partial H_2}{\partial h_j} \langle \xi, [e_i, e_j] \rangle.$$

Now let us compute

$$\{\mathcal{H}_1, \mathcal{H}_2\}_{\mathfrak{g}^*}(\xi) = \langle \xi, [d\mathcal{H}_1, d\mathcal{H}_2] \rangle = \langle \xi, \left[ \sum_{i=1}^{n} \frac{\partial \mathcal{H}_1}{\partial h_i} e_i, \sum_{j=1}^{n} \frac{\partial \mathcal{H}_2}{\partial h_j} e_j \right] \rangle = \sum_{i,j=1}^{n} \frac{\partial \mathcal{H}_1}{\partial h_i} \frac{\partial \mathcal{H}_2}{\partial h_j} \langle \xi, [e_i, e_j] \rangle.$$

The conclusion follows. □

**Definition 18.23.** The symplectic leaves of the Poisson manifolds $(\mathfrak{g}^*, \{\cdot,\cdot\}_{\mathfrak{g}^*})$ are called *coadjoint orbits.*

**Exercise 18.24.** Prove that the coadjoint orbit of $G$ through a point $\xi \in \mathfrak{g}^*$ coincides with the set

$$\mathcal{O}^c(\xi) = \{\mathrm{Ad}^*_{g^{-1}}\xi,\ g \in G\}.$$

**Proposition 18.25.** *Let $\mathcal{H} \in \mathcal{C}^\infty(\mathfrak{g}^*)$ and let $\vec{\mathcal{H}}$ be the corresponding Poisson vector field. Then $\vec{\mathcal{H}}(\xi) = (\mathrm{ad}\, d\mathcal{H})^*\xi$.*

*Proof.* Let $a \in \mathcal{C}^\infty(\mathfrak{g}^*)$. By definition we have

$$\begin{aligned} < da, \vec{\mathcal{H}}(\xi) > &= \vec{\mathcal{H}}(a)|_\xi = \{\mathcal{H}, a\}_{\mathfrak{g}^*}(\xi) = \langle \xi, [d\mathcal{H}, da] \rangle = \langle \xi, (\mathrm{ad}\, d\mathcal{H}) da \rangle = \langle (\mathrm{ad}\, d\mathcal{H})^*\xi, da \rangle \\ &=< da, (\mathrm{ad}\, d\mathcal{H})^*\xi > . \end{aligned} \tag{18.7}$$

In this formula the symbol $\langle \cdot, \cdot \rangle$ is the duality product defined on $\mathfrak{g}^* \times \mathfrak{g}$, while the symbol $< \cdot, \cdot >$ is the duality product defined on $\mathfrak{g}^{**} \times \mathfrak{g}^*$. □

### 18.2.2 The Casimir first integral

Let us come back on the trivialized Hamiltonian equations associated with a left-invariant Hamiltonian $H(p, g) = \mathcal{H}(L_g^* p)$ (cf. Proposition 7.65).

$$\begin{cases} \dot{g} = L_{g*} d\mathcal{H} \\ \dot{\xi} = (\mathrm{ad}\, d\mathcal{H})^*\xi. \end{cases} \tag{18.8}$$

Here $g \in G$, $d\mathcal{H}$ is seen as an element of $\mathfrak{g}$, $\xi$ is the trivialized covector belonging to $\mathfrak{g}^*$ and $(\mathrm{ad}\, d\mathcal{H})^*$ is an operator from $\mathfrak{g}^*$ to itself.

Notice the second equation is decoupled from the first one since $\mathcal{H}$ is a function of $\xi$ only (although non-linear in general). In other words, writing explicitly the dependence on time we have

$$\begin{cases} \dot{g}(t) = L_{g*} d\mathcal{H}(\xi(t)) \\ \dot{\xi}(t) = \big(\mathrm{ad}\, d\mathcal{H}(\xi(t))\big)^*\xi(t). \end{cases} \tag{18.9}$$

Thanks to Proposition 18.25 the vector field $\big(\mathrm{ad}\, d\mathcal{H}(\xi(t))\big)^*\xi$ on $\mathfrak{g}^*$ is the Poisson vector field $\vec{\mathcal{H}}$ corresponding to the function $\mathcal{H}$. Hence we have

**Lemma 18.26.** *Let $(\xi(t), g(t))$ be a solution to (18.8) corresponding to an initial condition $(\xi(0), g(0))$. Then, for every $t$, $\xi(t)$ belong to the coadjoint orbit $\mathcal{O}(\xi(0))$.*

**Proposition 18.27.** *Let $H$ be a left-invariant Hamiltonian on a Lie group and $\mathscr{C} : \mathfrak{g}^* \to \mathbb{R}$ be a Casimir function. Define*

$$I_{\mathscr{C}}(p, g) = \mathscr{C}(L_g^* p).$$

*Then*

$$\{H, I_{\mathscr{C}}\} = 0.$$

*In particular, $I_{\mathscr{C}}$ is a constant along the integral curves of $\vec{H}$.*

*Proof.* We have $\{H, I_{\mathscr{C}}\}|_{(p(t),g(t))} = \{\mathcal{H}, \mathscr{C}\}_{\mathfrak{g}^*}|_{(L^*_{g(t)}p(t))} = 0$. □

When $\mathscr{C}$ is a non-constant Casimir function, the corresponding function $I_{\mathscr{C}}$ is called the *Casimir first integral* and it is constant along the integral curve of any left-invariant Hamiltonian.

*Remark* 18.28. The fact that when $\mathscr{C}$ is non constant, $I_{\mathscr{C}}$ is a first integral can also be seen in the following way. The trivialized Hamiltonian system corresponding to a left-invariant Hamiltonian $H$ is (18.9) with $\mathcal{H}$ defined by $H(p, g) = \mathcal{H}(L_g^* p)$ and $\xi = L_g^* p \in \mathfrak{g}^*$. Since $\xi(t)$ has support in a coadjoint orbit we have that $\mathscr{C}(\xi(t))$ is constant. As a consequence $I_{\mathscr{C}}(p(t), g(t)) = \mathscr{C}(\xi(t))$ is constant as well.

Notice that in vertical coordinates $h_1, \ldots, h_n$ we have that

$$I_{\mathscr{C}}(h_1, \ldots h_n) = \mathscr{C}(h_1, \ldots h_n).$$

### 18.2.3 First integrals associated with a right-invariant vector field

**Lemma 18.29.** *Let $X$ be a left-invariant vector field and $Y$ be a right-invariant one on a Lie group $G$. Then $[X, Y] = 0$.*

*Proof.* Recall that the flow induced by a left-invariant vector field is a right-translation, the flow induced by a right-invariant vector field is a left-translation and that the two flows commute (cf. Section 7.2.1). As a consequence the Lie brackets of the corresponding vector fields is identically zero. □

As a consequence we have

**Corollary 18.30.** *Let $H$ be a left-invariant Hamiltonian and $Y$ be a right-invariant vector field. Let $H_Y = \langle p, Y(g)\rangle$ be the corresponding right-invariant Hamiltonian linear on fibers. Then*

$$\{H, H_Y\} = 0.$$

*Proof.* In vertical coordinates $h_i(p, g) = \langle p, X_i(g)\rangle$, $i = 1, \ldots, n$, we have

$$\{H, H_Y\} = \sum_{i=1}^{n} \frac{\partial H}{\partial h_i}\{h_i, H_Y\} = \sum_{i=1}^{n} \frac{\partial H}{\partial h_i}\langle p, [X_i, Y](g)\rangle = 0$$

since $X_i$ is left-invariant and $Y$ is right-invariant. □

*Remark* 18.31. Notice that, given an Hamiltonian system on Lie group of dimension $n$, there are $n$ independent right-invariant vector fields $Y_1, \ldots, Y_n$. Hence $H_{Y_1} = \langle p, Y_1(g)\rangle, \ldots, H_{Y_n} = \langle p, Y_n(g)\rangle$ are $n$ first integrals in involution with the Hamiltonian. However we have that $\{H_{Y_i}, H_{Y_j}\} = 0$, for some $i, j$ if and only if $[Y_i, Y_j] = 0$. Hence the method of constructing first integrals in involutions via right-invariant vector fields provides more than one constant of the motion only if there is a commutative subalgebra in $\mathfrak{g}$, of dimension larger than one.

### 18.2.4 Complete integrability on Lie groups

Recall that an Hamiltonian system is completely integrable if one can find $n$ functions on $T^*G$ ($n = \dim(G)$), including the Hamiltonian, that are independent and in involution (cf. Section 5.4).

Given a Hamiltonian system on a Lie group associated with a left-invariant Hamiltonian $H$, above we have described a method to find two additional first integrals:

- $I_{\mathscr{C}}(p, g) = \mathscr{C}(L_g^* p)$, where $\mathscr{C}$ is a non-constant Casimir function on $\mathfrak{g}^*$;
- $H_Y(p, g) = \langle p, Y(g)\rangle$ where $Y(g)$ is any right-invariant vector field.

We have already seen that $H, H_Y, I_{\mathscr{C}}$ are automatically in involution because $I_{\mathscr{C}}$ is in involution with any left-invariant function (and in particular with $H$) and $H_Y$ (being right-invariant) is involution with any left-invariant function (and in particular with $H$ and $I_{\mathscr{C}}$). However, as we are going to see in the next section, there is no guarantee a priory that a non-constant Casimir function exists and that $H, H_Y, I_{\mathscr{C}}$ are independent. One should study case by case.

*Remark* 18.32. Notice that in general there may exist several independent Casimir functions. Moreover, as already remarked, if $\mathfrak{g}$ contains a commutative sub-algebra of dimension larger than one, there are more than one independent first integrals constructed with the corresponding right-invariant vector fields.

Notice that the exponential of an element belonging to the center of a Lie algebra $\mathfrak{g}$, is an element of the center of the group. As a consequence we have the following.

**Proposition 18.33.** *Let $W \neq 0$ be an element of the center of $\mathfrak{g}$ and $X$ be the corresponding left-invariant vector field. Then $X$ is also right-invariant. Moreover $\mathscr{C}(\xi) := \langle \xi, W\rangle$ is a (non-constant) Casimir function.*

An analogue statement holds for right-invariant vector fields. In other words, if $W \neq 0$ is in the center of $\mathfrak{g}$, and (in matrix notation) $X(g) = gW$ and $Y(g) = Wg$. Then $X = Y$.

In the next section we are going to study the case of 3-dimensional Lie groups.

## 18.3 Left-invariant Hamiltonian systems on 3D Lie groups

On a 3D Lie group $G$, we have that $\mathfrak{g}^*$ has dimension 3. Being $\mathfrak{g}^*$ odd dimensional, it cannot be symplectic. One could be tempted to say that in this case a non-constant Casimir function $\mathscr{C}$ (whose level sets are the symplectic leaves) always exists. Since one can always find a first integral $H_Y$ built with a right-invariant vector field $Y$, one could be tempted to say that any Hamiltonian system corresponding to a left-invariant Hamiltonian $H$, under the hypotheses that $H$, $I_{\mathscr{C}}(p, g)$ and $H_Y$ are independent, is completely integrable.

However, the situation is not so simple, since the existence of symplectic leaves does not guarantee the existence of a non-constant Casimir function $\mathscr{C}$.

Let us now study the existence of a non-constant Casimir function on a 3D Lie group. To this purpose it will be useful the classification studied in Section 17.5. Hence let us fix a rank-2 left-invariant sub-Riemannian structure for which an orthonormal frame is given by $X_1, X_2$ and $X_0$ is the corresponding Reeb field.

With such a structure one can associate the two invariants $\chi \geq 0$ and $\kappa \in \mathbb{R}$ (cf. Section 17.2). We will find Casimir functions that can be expressed in terms of these invariants (as consequence of the chosen frame), however they depend only on the Lie group structure.

It is then useful to recall that, besides the case $\kappa = \chi = 0$ (Heisenberg case), we can always normalize (up to a dilation of the structure) $\kappa^2 + \chi^2 = 1$. Moreover if we are interested only to the Lie group structure, in certain cases we can fix completely $\chi$ and $\kappa$ (cf. Section 17.5.1). This will be done in Section 18.3.2.

In the following we use the coordinates on $\mathfrak{g}^*$

$$h_i = \langle \xi, X_i \rangle, \quad X_i \in \mathfrak{g}, \quad \xi \in \mathfrak{g}^*, \quad i = 0, 1, 2,$$

and we refer to Sections 17.3, 17.4, and 17.5 for more details on the notation. Up to a rotation of $X_1$ and $X_2$, the following cases are possible for the structure of 3D-Lie algebras (we omit the abelian case for which the theory is obvious).

### Unimodular case with $\chi \neq 0$

Thanks to Exercice 17.16 we have the relations

$$[X_1, X_0] = (\chi + \kappa) X_2, \tag{18.10}$$
$$[X_2, X_0] = (\chi - \kappa) X_1, \tag{18.11}$$
$$[X_2, X_1] = X_0. \tag{18.12}$$

In this case we have the following.

**Proposition 18.34.** *In the unimodular case with $\chi \neq 0$, the function*

$$\mathscr{C}(h_0, h_1, h_2) = (\kappa - \chi) h_1^2 + (\kappa + \chi) h_2^2 + h_0^2,$$

*is a non-constant Casimir function.*

*Proof.* It is enough to verify that $\{\mathscr{C}, h_i\}_{\mathfrak{g}^*} = 0$, $i = 0, 1, 2$. Let us start with $i = 1$.

$$\begin{aligned}\{\mathscr{C}, h_1\} &= \sum_{i=0}^{2} \frac{\partial \mathscr{C}}{\partial h_i} \{h_i, h_1\}_{\mathfrak{g}^*} = (\kappa - \chi) 2h_1 \{h_1, h_1\}_{\mathfrak{g}^*} + (\kappa + \chi) 2h_2 \{h_2, h_1\}_{\mathfrak{g}^*} + 2h_0 \{h_0, h_1\}_{\mathfrak{g}^*} \\ &= (\kappa + \chi) 2h_2 h_0 + 2h_0(-\chi - \kappa) h_2 = 0. \end{aligned} \tag{18.13}$$

The other two relations are similar. ☐

*Remark* 18.35. Notice the presence of the invariants $\chi$ and $\kappa$ in $\mathscr{C}$ is due to the precise choice of the frame $X_1, X_2, X_0$. Indeed Casimir functions depends only on the group structure.

*Remark* 18.36. Notice that for $\kappa \neq \pm\chi$, the corresponding algebra is simple ($\mathfrak{su}(2)$ for $\kappa > \chi$ and $\mathfrak{sl}(2)$ for $\kappa < \chi$) and $\mathscr{C}$ is a weighted sum of $h_1^2$, $h_2^2$ and $h_0^2$. For $\kappa = \chi$ (resp. $\kappa = -\chi$) the corresponding algebra is $\mathfrak{se}(2)$ (resp. $\mathfrak{sh}(2)$) and $\mathscr{C}$ is a weighted sum of $h_2^2$ and $h_0^2$ (resp. $h_1^2$ and $h_0^2$) only.

### Unimodular case with $\chi = 0$

$$[X_1, X_0] = \kappa X_2, \tag{18.14}$$
$$[X_2, X_0] = -\kappa X_1, \tag{18.15}$$
$$[X_2, X_1] = X_0. \tag{18.16}$$

The same computation as (18.13) provides the following.

**Proposition 18.37.** *In the unimodular case with $\chi = 0$, the function*

$$\mathscr{C}(h_0, h_1, h_2) = \kappa(h_1^2 + h_2^2) + h_0^2,$$

*is a non-constant Casimir function.*

Notice that if $\kappa = 0$ we obtain the Heisenberg algebra and $\mathscr{C} = h_0^2$ implying that $h_0$ is a Casimir function as well. The same result can be obtained with Proposition 18.33, since $X_0$ belongs to the center of the algebra. If $\kappa \neq 0$ and we normalize $\kappa = \pm 1$, we obtain $\mathfrak{su}(2)$ and $\mathfrak{sl}(2)$ in their standard representations. It is then clear that $\mathscr{C}$ is proportional to the Killing form.

### Non-unimodular, $\chi \neq 0$, case 1 ($\mathfrak{solv}^+$)

Let us consider the commutation relations (with respect to the notation of Section 17.5.3, we have set $\alpha = c_{01}^2$ and $\beta = c_{12}^2$)

$$[X_1, X_0] = \alpha X_2, \tag{18.17}$$
$$[X_2, X_0] = 0, \tag{18.18}$$
$$[X_2, X_1] = \beta X_2 + X_0. \tag{18.19}$$

In this case $\alpha > 0$, $\beta \neq 0$ and $\chi = \frac{\alpha}{2} > 0$, $\kappa = -\beta^2 + \frac{\alpha}{2}$, $\chi - \kappa > 0$.

A Casimir function $\mathscr{C}$ should verify $\{\mathscr{C}, h_i\}_{\mathfrak{g}^*} = 0$, for $i = 0, 1, 2$. Hence we have

$$\{\mathscr{C}, h_0\} = \sum_{i=0}^{2} \frac{\partial \mathscr{C}}{\partial h_i}\{h_i, h_0\}_{\mathfrak{g}^*} = \frac{\partial \mathscr{C}}{\partial h_1}(\alpha h_2) = 0, \tag{18.20}$$

$$\{\mathscr{C}, h_1\} = \sum_{i=0}^{2} \frac{\partial \mathscr{C}}{\partial h_i}\{h_i, h_1\}_{\mathfrak{g}^*} = \frac{\partial \mathscr{C}}{\partial h_0}(-\alpha h_2) + \frac{\partial \mathscr{C}}{\partial h_2}(\beta h_2 + h_0) = 0, \tag{18.21}$$

$$\{\mathscr{C}, h_2\} = \sum_{i=0}^{2} \frac{\partial \mathscr{C}}{\partial h_i}\{h_i, h_2\}_{\mathfrak{g}^*} = \frac{\partial \mathscr{C}}{\partial h_1}(-\beta h_2 - h_0) = 0. \tag{18.22}$$

Equation (18.20) and (18.22) say that $\mathscr{C}$ is independent from $h_1$. As a consequence equation (18.21) gives

$$\left(\frac{\partial \mathscr{C}}{\partial h_0}, \frac{\partial \mathscr{C}}{\partial h_2}\right)(-\alpha h_2, \beta h_2 + h_0) = 0. \tag{18.23}$$

Now, in the plane $(h_0, h_2)$, define the linear vector field

$$V^+(h_0, h_2) := (-\alpha h_2, \beta h_2 + h_0) = \begin{pmatrix} 0 & -\alpha \\ 1 & \beta \end{pmatrix}\begin{pmatrix} h_0 \\ h_2 \end{pmatrix}. \tag{18.24}$$

Then equation (18.23) says that the gradient of $\mathscr{C}$ (with respect to $(h_0, h_2)$) is orthogonal to $V^+$. This implies that $\mathscr{C}$ is constant along the integral curves of $V^+$ (notice that $V^+$ is zero only at the origin). The eigenvalues of the matrix appearing in the definition of $V^+$ are

$$\frac{1}{2}(\beta \pm \sqrt{\beta^2 - 4\alpha}).$$

Now we have two cases.

A. If $\beta^2 - 4\alpha \geq 0$ (which corresponds to the case $\kappa + 7\chi \leq 0$) we have that the eigenvalues are either both positive or both negative, meaning that all integral curves of $V^+$ have the origin in their closure.

B. If $\beta^2 - 4\alpha < 0$ (which corresponds to the case $\kappa + 7\chi > 0$) we have that the eigenvalues are complex conjugate, meaning again that all integral curves of $V^+$ have the origin in their closure.

Since $\mathscr{C}$ is continuous, it is constant on the integral curves of $V^+$ and all such curves have the origin in their closure, it follows that $\mathscr{C}$ is a *constant* Casimir function. Hence we have

**Proposition 18.38.** *In the* $\mathfrak{solv}^+$ *case all Casimir functions are constant.*

### Non-unimodular, $\chi \neq 0$, case 2 ($\mathfrak{solv}^-$)

Let us consider the commutation relations (with respect to the notation of Section 17.5.3, we have set $\alpha = c^1_{02}$ and $\beta = c^1_{12}$)

$$[X_1, X_0] = 0, \tag{18.25}$$
$$[X_2, X_0] = \alpha X_1, \tag{18.26}$$
$$[X_2, X_1] = \beta X_1 + X_0. \tag{18.27}$$

In this case $\alpha > 0$, $\beta \neq 0$ and $\chi = \frac{\alpha}{2} > 0$, $\kappa = -\beta^2 - \frac{\alpha}{2}$, $\chi + \kappa < 0$. A similar computation as the one above permits to conclude that in this case $\mathscr{C}$ is independent from $h_2$. Moreover the gradient of $\mathscr{C}$ (in the plane $(h_0, h_1)$) is orthogonal to the vector

$$V^-(h_0, h_1) := \begin{pmatrix} 0 & -\alpha \\ -1 & -\beta \end{pmatrix} \begin{pmatrix} h_0 \\ h_1 \end{pmatrix}. \tag{18.28}$$

As before, this implies that $\mathscr{C}$ is constant along the integral curves of $V^-$. The eigenvalues of the matrix appearing in the definition of $V^-$ are

$$\lambda_\pm = \frac{1}{2}(-\beta \pm \sqrt{\beta^2 + 4\alpha}).$$

In this case, the origin is a saddle point for $V^-$. Then we have the following

**Proposition 18.39.** *In the* $\mathfrak{solv}^-$ *case define* $r = -\lambda_+/\lambda_- = \frac{-\beta + \sqrt{\beta^2+4\alpha}}{\beta+\sqrt{\beta^2+4\alpha}}$. *If* $r$ *is rational then a non-constant Casimir function exists in the analytic category. If* $r$ *is irrational then a non-constant Casimir function exists in the* $\mathcal{C}^\infty$ *category (and not in the analytic one).*

*Remark* 18.40. Notice that the existence or not of a non-constant Casimir function in the analytic category depends on the invariants $\chi$ and $\kappa$. However this is a property that depends only on the Lie group structure since in the $\mathfrak{solv}^-$ case different Lie algebras correspond to different values of $\chi$ and $\kappa$ and viceversa.

**Non-unimodular, $\chi = 0$, case ($\mathfrak{a}^+(\mathbb{R}) \oplus \mathbb{R}$)**

$$[X_1, X_0] = 0, \tag{18.29}$$
$$[X_2, X_0] = 0, \tag{18.30}$$
$$[X_2, X_1] = X_2 + X_0. \tag{18.31}$$

In this case $\chi = 0$ and $\kappa = -1$. Since $X_0$ belongs to the center of the algebra, from Proposition 18.33 follows

**Proposition 18.41.** *In the $\mathfrak{a}^+(\mathbb{R}) \oplus \mathbb{R}$, the function*

$$\mathscr{C}(h_0, h_1, h_2) = h_0$$

*is a non-constant Casimir function.*

Collecting all the results of this section, we have the following.

**Proposition 18.42.** *A 3D Lie group admits a non-constant Casimir function if and only if it is not in the $SOLV^+$ case.*

### 18.3.1 Rank 2 sub-Riemannian structures on 3D Lie groups

Let us now study the case of rank-2 left-invariant sub-Riemannian structure on 3D-Lie groups. In this case the Hamiltonian for normal extremals has the form

$$H = \frac{1}{2}(\langle p, X_1\rangle^2 + (\langle p, X_2\rangle^2),$$

for some non-commuting left-invariant vector fields $X_1$ and $X_2$.

We can now study, case by case, the independence of the various first integrals. In the following we call $Y_0, Y_1, Y_2$ the right-invariant vector fields corresponding to the left-invariant vector fields $X_0, X_1, X_2$ (i.e., in matrix notation if $X_i = gW_i$, $i = 0, 1, 2$ then $Y_i = W_i g$, $i = 0, 1, 2$). Recall that in vertical coordinates $I_{\mathscr{C}}(h_1, h_2, h_3) = \mathscr{C}(h_1, h_2, h_3)$.

- Unimodular case with $\chi \neq 0$. In this case we have the first integrals $H$, $I_{\mathscr{C}}$, $H_Y$ where $Y$ is any right-invariant vector field. Let us use $(h_0, h_1, h_2)$ as coordinates. $H$ is independent from $I_{\mathscr{C}}$ since $I_{\mathscr{C}}$ depends on $h_0$ while $H$ does not. $H_Y$ is independent from $H$ and $I_{\mathscr{C}}$ since it depends only on $\langle p, Y(g)\rangle$ while $H$ and $I_{\mathscr{C}}$ are quadratic in at least two of the $h_i$ ($i = 0, 1, 2$). The system is completely integrable.

- Unimodular case with $\chi = 0$. If $\kappa \neq 0$ then the situation is the same as the previous one. If $\kappa = 0$ (Heisenberg case) then $I_{\mathscr{C}} = h_0^2$. Moreover since in this case $X_0$ and (for instance) $X_1$ form a commutative sub-algebra, the linear Hamiltonians built with the corresponding right-invarant vector fields $Y_0$ and $Y_1$ are two first integrals. We found 4 first integral in involution: $H$, $I_{\mathscr{C}}$, $H_{Y_0}$ and $H_{Y_1}$. We have that $H_{Y_0}$ and $I_{\mathscr{C}}$ are dependent since $H_{Y_0} = H_{X_0} = h_0$ ($X_0$ belongs to the center of the group, cf. Proposition 18.33) and $I_{\mathscr{C}} = h_0^2$. However $H_{Y_1}$ is independent from $H_{Y_0}$ since it corresponds to another independent right-invariant vector field. Finally $H$ is independent from $H_{Y_0}$ (resp. $H_{Y_1}$) since it is quadratic in $h_1$ and $h_2$ while $H_{Y_0}$ (resp. $H_{Y_1}$) depends only on $\langle p, Y_0(g)\rangle$ (resp. $\langle p, Y_1(g)\rangle$). The system is completely integrable.

- Non-unimodular case with $\chi \neq 0$ ($\mathfrak{solv}^+$). In this case there are no non-constant Casimir functions. However since there is a commutative subalgebra (generated by $X_0$ and $X_2$) we have the first integrals $H_{Y_0}$ and $H_{Y_2}$. They are independent since they correspond to independent vector fields. As before, $H$ is independent from them since it is quadratic in $h_1$ and $h_2$. The system is completely integrable.

- Non-unimodular case with $\chi \neq 0$ ($\mathfrak{solv}^-$). In this case a non-constant Casimir function exists (in the analytic or $\mathcal{C}^\infty$ category depending on the value of the parameters). However we obtain complete integrability even without using this first integral. As in the previous case, we have a commutative subalgebra (generated by $X_0$ and $X_1$) and hence we have the first integrals $H_{Y_0}$ and $H_{Y_1}$. They are independent since they correspond to independent vector fields. As before, $H$ is independent from them since it is quadratic in $h_1$ and $h_2$. The system is completely integrable.

- $\mathfrak{a}^+(\mathbb{R}) \oplus \mathbb{R}$. Since $X_0$ belongs to the center of the algebra, $I_{\mathscr{C}} = h_0$ is a non-constant Casimir function. Moreover, since $X_0$ and (for instance) $X_1$ form a commutative sub-algebra, another independent first integral is $H_{Y_1}$. Finally the 3 first integrals that are independent and in involution are $H$, $H_{Y_0}$ and $H_{Y_1}$. The system is completely integrable.

We have then proved the following

**Theorem 18.43.** *The Hamiltonian system associated with a rank-2 left-invariant sub-Riemannian structure on a 3D Lie group is completely integrable.*

Notice that out of the rank-2 sub-Riemannian case, the first integrals could be different with respect to those identified in this section. Consider for instance the Hamiltonian system that one obtains looking for Riemannian geodesics on $SU(2)$ with the metric given by the opposite of the Killing form. In that case $I_{\mathscr{C}}$ and $H$ are proportional. However, besides $H$, one can take as independent first integrals $H_Y$ and $H_X$ where $X$ is a left-invariant vector field and $Y$ is a right-invariant vector field (not corresponding to $X$). This is possible since in this case $H$ is bi-invariant.

**Exercise 18.44.** Prove that the Hamiltonian system associated with a left-invariant Riemannian structure on a 3D Lie group is completely integrable.

**Exercise 18.45.** Find the explicit expression of the geodesics on $SU(2)$ for the Riemannan metric given by the opposite of the Killing form (cf. Secton 7.7.1).

**Exercise 18.46.** Consider the matrix realization of $SE(2)$ that we used in Section 13.8[1]

$$SE(2) = \left\{ \left( \begin{array}{cc|c} \cos(\theta) & -\sin(\theta) & x_1 \\ \sin(\theta) & \cos(\theta) & x_2 \\ \hline 0 & 0 & 1 \end{array} \right), \quad \theta \in S^1, \quad x_1, x_2 \in \mathbb{R} \right\},$$
$$\mathfrak{se}(2) = \operatorname{span}\{X_1, X_2, X_0\},$$

where

$$X_1 = \left( \begin{array}{cc|c} 0 & -1 & 0 \\ 1 & 0 & 0 \\ \hline 0 & 0 & 0 \end{array} \right), \quad X_2 = \left( \begin{array}{cc|c} 0 & 0 & 1 \\ 0 & 0 & 0 \\ \hline 0 & 0 & 0 \end{array} \right), \quad X_0 = \left( \begin{array}{cc|c} 0 & 0 & 0 \\ 0 & 0 & -1 \\ \hline 0 & 0 & 0 \end{array} \right).$$

[1] with respect to Section 13.8 here we have the change of notation $e_r \to X_1$, $e_1 \to X_2$, $e_2 \to -X_0$.

Let $p = (p_1, p_2, p_\theta)$ be the dual variable to $(x_1, x_2, \theta)$. Let $g \in SE(2)$. Call $\bar{X}_1$, $\bar{X}_2$, $\bar{X}_0$ the vector realizations of $gX_1$, $gX_2$, $gX_0$ (cf. Section 13.8) and $\bar{Y}_1$, $\bar{Y}_2$, $\bar{Y}_0$ the vector realizations of $X_1 g, X_2 g, X_0 g$. Prove that in these coordinates $I_{\mathscr{C}} = p_1^2 + p_2^2$ is a Casimir function and that $H_{Y_2} := \langle p, \bar{Y}_2 \rangle = p_1$ and $H_{Y_0} := \langle p, \bar{Y}_0 \rangle = -p_2$. Notice that we found 4 first integrals in involution: $H = \frac{1}{2}(\langle p, \bar{X}_1 \rangle^2 + \langle p, \bar{X}_2 \rangle^2)$, $I_{\mathscr{C}}$, $H_{Y_2}$, $H_{Y_0}$ (the presence of $H_{Y_2}$, $H_{Y_0}$ is due to the presence of a commutative sub-algebra). However $I_{\mathscr{C}}$ depends on $H_{Y_2}$ and $H_{Y_0}$.

### 18.3.2 Classification of symplectic leaves on 3D Lie groups

In the previous section we have studied Casimir functions on 3D Lie groups. We know that symplectic leaves are contained in the level set of Casimir functions. However they can be smaller. In this section we complete the analysis by computing explicitly all symplectic leaves.

To this purpose we are going to construct explicitly the set $\mathscr{P}$ of all Poisson vector fields. For the moment consider a general Lie group $G$ of dimension $n$ with Lie algebra $\mathfrak{g}$. Let $\{X_0, \ldots, X_{n-1}\}$ be a basis of $\mathfrak{g}$. On $\mathfrak{g}^*$ let us use coordinates $h_i = \langle \xi, X_i \rangle$, $i = 0, \ldots, n-1$. In these coordinates $dh_i = X_i$. We have

$$\mathscr{P} = \{\vec{a} \in \mathrm{Vec}(\mathfrak{g}^*) \mid \vec{a}(\cdot) = \langle \xi, [da, d(\cdot)] \rangle \text{ for some function } a \in \mathcal{C}^\infty(\mathfrak{g}^*)\}.$$

In the formula above $da$ belongs to $\mathfrak{g}^{**}$ and hence it is identified with an elements of $\mathfrak{g}$.

Let now $a, b \in \mathcal{C}^\infty(\mathfrak{g}^*)$ and let us compute $\vec{a}(b)$. Since $da = \sum_{i=0}^{n-1} (\partial_{h_i} a) X_i$, $db = \sum_{i=0}^{n-1} (\partial_{h_i} b) X_i$, we have that

$$\vec{a}(b) = \langle \xi, [da, db] \rangle = \sum_{i,j=0}^{n-1} (\partial_{h_i} a)\, (\partial_{h_j} b)\, \langle \xi, [X_i, X_j] \rangle. \tag{18.32}$$

Notice that if we set $[X_i, X_j] = \sum_k c_{ji}^k X_k$ then $\vec{a}(b) = \sum_{i,j,k=0}^{n-1} (\partial_{h_i} a)\, (\partial_{h_j} b)\, c_{ji}^k\, h_k$ and

$$\mathscr{P} = \left\{ \sum_{i,j,k=0}^{n-1} (\partial_{h_i} a)\, c_{ji}^k\, h_k\, \partial_{h_j} \mid a \in \mathcal{C}^\infty(\mathfrak{g}^*) \right\}.$$

Notice that all vector fields of $\mathscr{P}$ are zero at the origin of $\mathfrak{g}^*$. Hence we have

**Proposition 18.47.** *For any Lie group $G$ with Lie algebra $\mathfrak{g}$, the origin in $\mathfrak{g}^*$ is a symplectic leaf.*

Let us now come back to the 3D case. From (18.32) we have

$$\begin{aligned}\vec{a}(b) = \langle \xi, [da, db] \rangle = &\big(\partial_{h_1} a \langle \xi, [X_1, X_0] \rangle + \partial_{h_2} a \langle \xi, [X_2, X_0] \rangle\big) \partial_{h_0} b + \\ &+ \big(\partial_{h_0} a \langle \xi, [X_0, X_1] \rangle + \partial_{h_2} a \langle \xi, [X_2, X_1] \rangle\big) \partial_{h_1} b + \\ &+ \big(\partial_{h_0} a \langle \xi, [X_0, X_2] \rangle + \partial_{h_1} a \langle \xi, [X_1, X_2] \rangle\big) \partial_{h_2} b.\end{aligned} \tag{18.33}$$

Let us now study the different cases. Since here we are interested to the symplectic leaves, that depend only on the group structure, we normalize as much as possible the invariant $\chi$ and $\kappa$.

**Symplectic leaves on $\mathfrak{su}(2)$**

On $\mathfrak{su}(2)$ a non-constant Casimir function is given by (see Proposition 18.37, normalizing $\kappa = 1$)

$$\mathscr{C} = h_1^2 + h_2^2 + h_0^2.$$

The commutation relations are (see (18.14),(18.15),(18.16) with $\kappa = 1$)

$$[X_1, X_0] = X_2, \quad [X_2, X_0] = -X_1, \quad [X_2, X_1] = X_0.$$

Hence from (18.33) we have

$$\vec{a} = \big(\partial_{h_1} a\, h_2 - \partial_{h_2} a\, h_1\big)\partial_{h_0} + \big(-\partial_{h_0} a\, h_2 + \partial_{h_2} a\, h_0\big)\partial_{h_1} + \big(\partial_{h_0} a\, h_1 - \partial_{h_1} a\, h_0\big)\partial_{h_2}, \tag{18.34}$$

and

$$\mathscr{P} = \{\vec{a} \text{ of the form (18.34)} \mid a \in \mathcal{C}^\infty(\mathfrak{g}^*)\}$$

Taking $a = h_0$, then $a = h_1$, and finally $a = h_2$ we obtain that $\mathscr{P}$ contains the 3 vector fields

$$\begin{aligned} F_0 &= -h_2\partial_{h_1} + h_1\partial_{h_2}, \\ F_1 &= h_2\partial_{h_0} - h_0\partial_{h_2}, \\ F_2 &= -h_1\partial_{h_0} + h_0\partial_{h_1}. \end{aligned}$$

We already know that these vector fields are tangent to the level sets of $\mathscr{C} = h_1^2 + h_2^2 + h_0^2$. One immediately verify that they generate the tangent space of each level set. Hence we obtain the following.

**Proposition 18.48.** *On $\mathfrak{su}(2)$ the symplectic leaves are the level sets of the function $h_1^2 + h_2^2 + h_0^2$.*

Notice that the symplectic leaf corresponding to the zero level is zero-dimensional while all other symplectic leaves are 2-dimensional. The symplectic leaves for $\mathfrak{su}(2)$ are plotted in Figure 18.1.

**Symplectic leaves on $\mathfrak{sl}(2)$**

On $\mathfrak{sl}(2)$ a non-constant Casimir function is given by (see Proposition 18.37, normalizing $\kappa = -1$)

$$\mathscr{C} = -h_1^2 - h_2^2 + h_0^2.$$

The commutation relations are (see (18.14),(18.15),(18.16) with $\kappa = -1$)

$$[X_1, X_0] = -X_2, \quad [X_2, X_0] = X_1, \quad [X_2, X_1] = X_0.$$

Hence from (18.33) we have

$$\vec{a} = \big(-\partial_{h_1} a\, h_2 + \partial_{h_2} a\, h_1\big)\partial_{h_0} + \big(\partial_{h_0} a\, h_2 + \partial_{h_2} a\, h_0\big)\partial_{h_1} + \big(-\partial_{h_0} a\, h_1 - \partial_{h_1} a\, h_0\big)\partial_{h_2}. \tag{18.35}$$

Taking $a = h_0$, then $a = h_1$, and finally $a = h_2$ we obtain that $\mathscr{P}$ contains the 3 vector fields

$$\begin{aligned} F_0 &= h_2\partial_{h_1} - h_1\partial_{h_2}, \\ F_1 &= -h_2\partial_{h_0} - h_0\partial_{h_2}, \\ F_2 &= h_1\partial_{h_0} + h_0\partial_{h_1}. \end{aligned}$$

As in the previous case, we already know that these vector fields are tangent to the level sets of $\mathscr{C} = -h_1^2 - h_2^2 + h_0^2$. One immediately verify that they generate the tangent space of on each level different from zero. The zero level is a double cone $C$, whose vertex is the origin. We already know that the origin is a symplectic leaf (since all vector fields of $\mathscr{P}$ are zero there). One immediately verify that the three vector fields generate the tangent space of the two connected components of $C \setminus \{0\}$. Hence we obtain the following.

**Proposition 18.49.** *On $\mathfrak{sl}(2)$ the symplectic leaves are*

$$\begin{aligned} S_c &= \{(h_0, h_1, h_2) \mid -h_1^2 - h_2^2 + h_0^2 = c\}, \quad c \neq 0, \\ S_0^+ &= \{(h_0, h_1, h_2) \mid -h_1^2 - h_2^2 + h_0^2 = 0, \quad h_0 > 0\}, \\ S_0^- &= \{(h_0, h_1, h_2) \mid -h_1^2 - h_2^2 + h_0^2 = 0, \quad h_0 < 0\}, \\ S_0^0 &= \{0\}. \end{aligned}$$

Notice that the symplectic leaf $S_0^0$ is zero-dimensional while the others are 2-dimensional. The symplectic leaves for $\mathfrak{sl}(2)$ are shown in Figure 18.1.

### Symplectic leaves on the Heisenberg algebra

On the Heisenberg algebra, a non-constant Casimir function is given by (see Proposition 18.37, with $\kappa = 0$ and extracting the square root)

$$\mathscr{C} = h_0.$$

The commutation relations are (see (18.14),(18.15),(18.16) with $\kappa = 0$)

$$[X_1, X_0] = 0, \quad [X_2, X_0] = 0, \quad [X_2, X_1] = X_0.$$

Hence from (18.33) we have

$$\vec{a} = \partial_{h_2} a \, h_0 \, \partial_{h_1} - \partial_{h_1} a \, h_0 \, \partial_{h_2}. \tag{18.36}$$

Taking $a = h_1$ and then $a = h_2$ we obtain that $\mathscr{P}$ contains the 2 vector fields

$$\begin{aligned} F_1 &= -h_0 \partial_{h_2}, \\ F_2 &= h_0 \partial_{h_1}. \end{aligned}$$

We already know that these vector fields are tangent to the level sets of $\mathscr{C} = h_0$. These vector fields generate the tangent space of each level different from zero. On the zero level, all vector fields of the form (18.36) are zero. Hence we obtain the following.

**Proposition 18.50.** *On the Heisenberg algebra, the symplectic leaves are*

$$\begin{aligned} S_c &= \{(h_0, h_1, h_2) \mid h_0 = c\}, \quad c \neq 0, \\ S_{c_1, c_2} &= \{(h_0, h_1, h_2) \mid h_1 = c_1, \quad h_2 = c_2, \quad h_0 = 0\}, \quad c_1, c_2 \in \mathbb{R}. \end{aligned}$$

Notice that the symplectic leaves $S_{c_1,c_2}$ are zero-dimensional while the symplectic leaves $S_c$ are two-dimensional.

The symplectic leaves for the Heisenberg algebra are plotted in Figure 18.1.

### Symplectic leaves on $\mathfrak{se}(2)$

On $\mathfrak{se}(2)$ a non-constant Casimir function is given by (see Proposition 18.34, with $\chi = \kappa$ and normalizing $\kappa = 1/2$)

$$\mathscr{C} = h_2^2 + h_0^2.$$

The commutation relations are (see (18.10),(18.11),(18.12) with $\kappa = \chi = 1/2$)

$$[X_1, X_0] = X_2, \quad [X_2, X_0] = 0, \quad [X_2, X_1] = X_0.$$

Hence from (18.33) we have

$$\vec{a} = \big(\partial_{h_1} a\, h_2\big)\partial_{h_0} + \big(-\partial_{h_0} a\, h_2 + \partial_{h_2} a\, h_0\big)\partial_{h_1} + \big(-\partial_{h_1} a\, h_0\big)\partial_{h_2}. \tag{18.37}$$

Taking $a = h_0$, then $a = h_1$, and finally $a = h_2$ we obtain that $\mathscr{P}$ contains the 3 vector fields

$$\begin{aligned} F_0 &= -h_2 \partial_{h_1} \\ F_1 &= h_2 \partial_{h_0} - h_0 \partial_{h_2}, \\ F_2 &= h_0 \partial_{h_1}. \end{aligned}$$

We already know that these vector fields are tangent to the level sets of $\mathscr{C} = h_2^2 + h_0^2$. One immediately verify that they generate the tangent space each level set different from zero. The zero level is the $h_1$ axis and since (18.37) is zero on this axis, all its points are zero-dimensional symplectic leaves. Hence we obtain the following.

**Proposition 18.51.** *On $\mathfrak{se}(2)$ the symplectic leaves are*

$$\begin{aligned} S_c &= \{(h_0, h_1, h_2) \mid h_2^2 + h_0^2 = c\}, \quad c > 0, \\ S_{c_0} &= \{(h_0, h_1, h_2) \mid h_1 = c_0, \quad h_2 = h_0 = 0\}, \quad c_0 \in \mathbb{R}. \end{aligned}$$

Notice that the symplectic leaves $S_{c_0}$, are zero-dimensional while the the symplectic leaves $S_c$ are two-dimensional. The symplectic leaves for $\mathfrak{se}(2)$ are plotted in Figure 18.1.

### Symplectic leaves on $\mathfrak{sh}(2)$

On $\mathfrak{sh}(2)$ a non-constant Casimir function is given by (see Proposition 18.34, with $\kappa = -\chi$ and normalizing $\kappa = -1/2$)

$$\mathscr{C} = -h_1^2 + h_0^2.$$

The commutation relations are (see (18.10),(18.11),(18.12) with $\kappa = -1/2$, $\chi = 1/2$)

$$[X_1, X_0] = 0, \quad [X_2, X_0] = X_1, \quad [X_2, X_1] = X_0.$$

Hence from (18.33) we have

$$\vec{a} = \big(\partial_{h_2} a\, h_1\big)\partial_{h_0} + \big(\partial_{h_2} a\, h_0\big)\partial_{h_1} + \big(-\partial_{h_0} a\, h_1 - \partial_{h_1} a\, h_0\big)\partial_{h_2}. \tag{18.38}$$

Taking $a = h_0$, then $a = h_1$, and finally $a = h_2$ we obtain that $\mathscr{P}$ contains the 3 vector fields

$$\begin{aligned} F_0 &= -h_1 \partial_{h_2}, \\ F_1 &= -h_0 \partial_{h_2}, \\ F_2 &= h_1 \partial_{h_0} + h_0 \partial_{h_1}. \end{aligned}$$

We already know that these vector fields are tangent to the level sets of $\mathscr{C} = -h_1^2 + h_0^2$. One immediately verify that they generate the tangent space to each level set different from zero. The zero level is the $h_2$ axis and since (18.38) is zero on this axis, all its points are zero-dimensional symplectic leaves. Hence we obtain the following.

**Proposition 18.52.** *On $\mathfrak{sh}(2)$ the symplectic leaves are*

$$\begin{aligned} S_c &= \{(h_0, h_1, h_2) \mid -h_1^2 + h_0^2 = c\}, \quad c > 0, \\ S_{c_0} &= \{(h_0, h_1, h_2) \mid h_2 = c_0, \quad h_1 = h_0 = 0\}, \quad c_0 \in \mathbb{R}. \end{aligned}$$

The symplectic leaves $S_{c_0}$, are zero-dimensional while the the symplectic leaves $S_c$ are two-dimensional. See Figure 18.1.

### Symplectic leaves on $\mathfrak{a}^+(\mathbb{R}) \oplus \mathbb{R}$

On $\mathfrak{a}^+(\mathbb{R}) \oplus \mathbb{R}$ a non-constant Casimir function is given by (see Proposition 18.41)

$$\mathscr{C} = h_0.$$

The commutation relations are

$$[X_1, X_0] = 0, \quad [X_2, X_0] = 0, \quad [X_2, X_1] = X_2 + X_0.$$

Hence from (18.33) we have

$$\vec{a} = \partial_{h_2} a \,(h_2 + h_0)\partial_{h_1} - \partial_{h_1} a \,(h_2 + h_0)\partial_{h_2}. \tag{18.39}$$

Taking $a = h_1$, then $a = h_2$, we obtain that $\mathscr{P}$ contains the 2 vector fields

$$\begin{aligned} F_1 &= (h_2 + h_0)\partial_{h_1}, \\ F_2 &= -(h_2 + h_0)\partial_{h_2}. \end{aligned}$$

These vector fields are tangent to the level sets of $\mathscr{C} = h_0$. All points of the plane $\{h_2 + h_0 = 0\}$ are zero-dimensional symplectic leaves. This plane cuts in two connected components every level set of $\mathscr{C}$. On each of these components, $F_1$ and $F_2$ generate the tangent space. Hence we have

**Proposition 18.53.** *On $\mathfrak{a}^+(\mathbb{R}) \oplus \mathbb{R}$ the symplectic leaves are*

$$\begin{aligned} S_c^+ &= \{(h_0, h_1, h_2) \mid h_0 = c, h_2 + h_0 > 0\}, \quad c \in \mathbb{R}, \\ S_c^- &= \{(h_0, h_1, h_2) \mid h_0 = c, h_2 + h_0 < 0\}, \quad c \in \mathbb{R}, \\ S_{c_1,c_2} &= \{(h_0, h_1, h_2) \mid h_1 = c_1, h_2 = c_2, \quad h_2 + h_0 = 0\}, \quad c_1, c_2 \in \mathbb{R}. \end{aligned} \tag{18.40}$$

Notice that the symplectic leaves $S_{c_1,c_2}$, are zero-dimensional while the others are 2-dimensional. Figure 18.2.

### Symplectic leaves on $\mathfrak{solv}^+$

On $\mathfrak{solv}^+$ a non-constant Casimir function does not exist (see Propositions 18.38 and 18.39). The commutation relations are

$$[X_1, X_0] = \alpha X_2, \quad [X_2, X_0] = 0, \quad [X_2, X_1] = \beta X_2 + X_0.$$

with $\alpha > 0$, $\beta \neq 0$. From (18.33) we have

$$\vec{a} = \big(\partial_{h_1} a\, \alpha h_2\big)\partial_{h_0} + \big( - \partial_{h_0} a\, \alpha h_2 + \partial_{h_2} a\, (\beta h_2 + h_0)\big)\partial_{h_1} + \big( - \partial_{h_1} a\, (\beta h_2 + h_0)\big)\partial_{h_2}, \qquad (18.41)$$

and taking $a = h_0$, then $a = -h_1$, and finally $a = h_2$ we obtain that $\mathscr{P}$ contains the 3 vector fields

$$\begin{aligned} F_0 &= -\alpha\, h_2 \partial_{h_1}, \\ F_1 &= \alpha\, h_2 \partial_{h_0} - (\beta h_2 + h_0)\partial_{h_2}, \\ F_2 &= (\beta h_2 + h_0)\partial_{h_1}. \end{aligned}$$

The integral curves of the linear vector $F_1$ in the plane $(h_0, h_2)$ has been already studied above, since $F_1 = V^+$ (cf. (18.24)). There are two cases: case A (i.e., $\beta^2 - 4\alpha \geq 0$) for which the phase portrait of $V^+$ has a node and case B (i.e., $\beta^2 - 4\alpha < 0$) for which the phase portrait of $V^+$ has a focus. In both cases all integral curves have the origin in their closure. In the plane $(h_0, h_2)$ let us call $\{\Gamma^+_\theta\}_{\theta \in S^1}$ the set of all integral curves of $V^+$ . For instance we can take

$$\Gamma^+_\theta = \{e^{tV^+}(\cos(\theta), \sin\theta), \quad t \in \mathbb{R}\}.$$

The other two vector fields are directed along $\partial_{h_1}$ and they are never both zero excepted on the $h_1$ axis. On the $h_1$ axis all vector fields of the type (18.41) are vanishing. Hence one obtain the following

**Proposition 18.54.** *On $\mathfrak{solv}^+$ the symplectic leaves are*

$$\begin{aligned} S_{c_0} &= \{(h_0, h_1, h_2) \mid h_0 = 0, \quad h_2 = 0, \quad h_1 = c_0\}, \quad c_0 \in \mathbb{R}, \\ S_\theta &= \{(h_0, h_1, h_2) \mid (h_0, h_2) \in \Gamma^+_\theta\}, \quad \theta \in S^1. \end{aligned}$$

Notice that the symplectic leaves corresponding to $S_{c_0}$ are zero-dimensional while all others are 2-dimensional. The symplectic leaves for $\mathfrak{solv}^+$ are plotted in 18.2.

### Symplectic leaves on $\mathfrak{solv}^-$

On $\mathfrak{solv}^-$ the commutation relations are

$$[X_1, X_0] = 0, \quad [X_2, X_0] = \alpha X_1, \quad [X_2, X_1] = \beta X_1 + X_0.$$

with $\alpha > 0$, $\beta \neq 0$. From (18.33) we have

$$\vec{a} = \big(\partial_{h_2} a\, \alpha h_1\big)\partial_{h_0} + \big(\partial_{h_2} a\, (\beta h_1 + h_0)\big)\partial_{h_1} + \big( - \partial_{h_0} a\, \alpha h_1 - \partial_{h_1} a\, (\beta h_1 + h_0)\big)\partial_{h_2}, \qquad (18.42)$$

and taking $a = h_0$, then $a = h_1$, and finally $a = -h_2$ we obtain that $\mathscr{P}$ contains the 3 vector fields

$$\begin{aligned} F_0 &= -\alpha\, h_1 \partial_{h_2}, \\ F_1 &= -(\beta h_1 + h_0)\partial_{h_2}, \\ F_2 &= -\alpha\, h_1 \partial_{h_0} - (\beta h_1 + h_0)\partial_{h_1}. \end{aligned}$$

The integral curves of the linear vector $F_2$ in the plane $(h_0, h_1)$ has been already studied above, since $F_2 = V^-$ (cf. (18.28)). The linear vector field $V^-$ has a saddle at the origin.

In the plane $(h_0, h_1)$ let us call $\{\Gamma_\theta^-\}_{\theta\in\Pi}$ the set of all integral curves of $V^-$ (here each curve is seen as a subset of $\mathbb{R}^2$). The other two vector fields are directed along $\partial_{h_2}$ and they are never both zero excepted on the $h_2$ axis. On the $h_2$ axis all vector fields of the type (18.42) are vanishing. Hence one obtain the following.

**Proposition 18.55.** *On $\mathfrak{solv}^-$ the symplectic leaves are*

$$S_{c_0} = \{(h_0, h_1, h_2) \mid h_0 = 0, \quad h_1 = 0, \quad h_2 = c_0\}, \quad c_0 \in \mathbb{R},$$
$$S_\theta = \{(h_0, h_1, h_2) \mid (h_0, h_1) \in \Gamma_\theta^-\}, \quad \theta \in \Pi.$$

Notice that the symplectic leaves corresponding to $S_{c_0}$ are zero-dimensional while all others are 2-dimensional. The symplectic leaves for $\mathfrak{solv}^-$ are plotted in Figure 18.2.

## 18.4 Bibliographical note

The integrability of the Hamiltonian systems associated with rank-2 sub-Riemannian structures on 3D Lie groups is "mathematical folklore". A statement and a sketch of the proof can be found in [Jur99]. See also [MS15] for the integrability and superintegrability in the unimodular case, together with a list of references where certain cases have been solved explicitly. In the more general context of optimal control problems on Lie groups, many interesting discussions can be found in [Jur97, Jur16]. Theorem 18.16, guaranteeing that the orbits of the set of Poisson vector fields are immersed sub-manifolds, can be found in any book of symplectic geometry. However it is one of most compelling application of Sussmann's orbit theorem [Sus73]. See also [AS04].

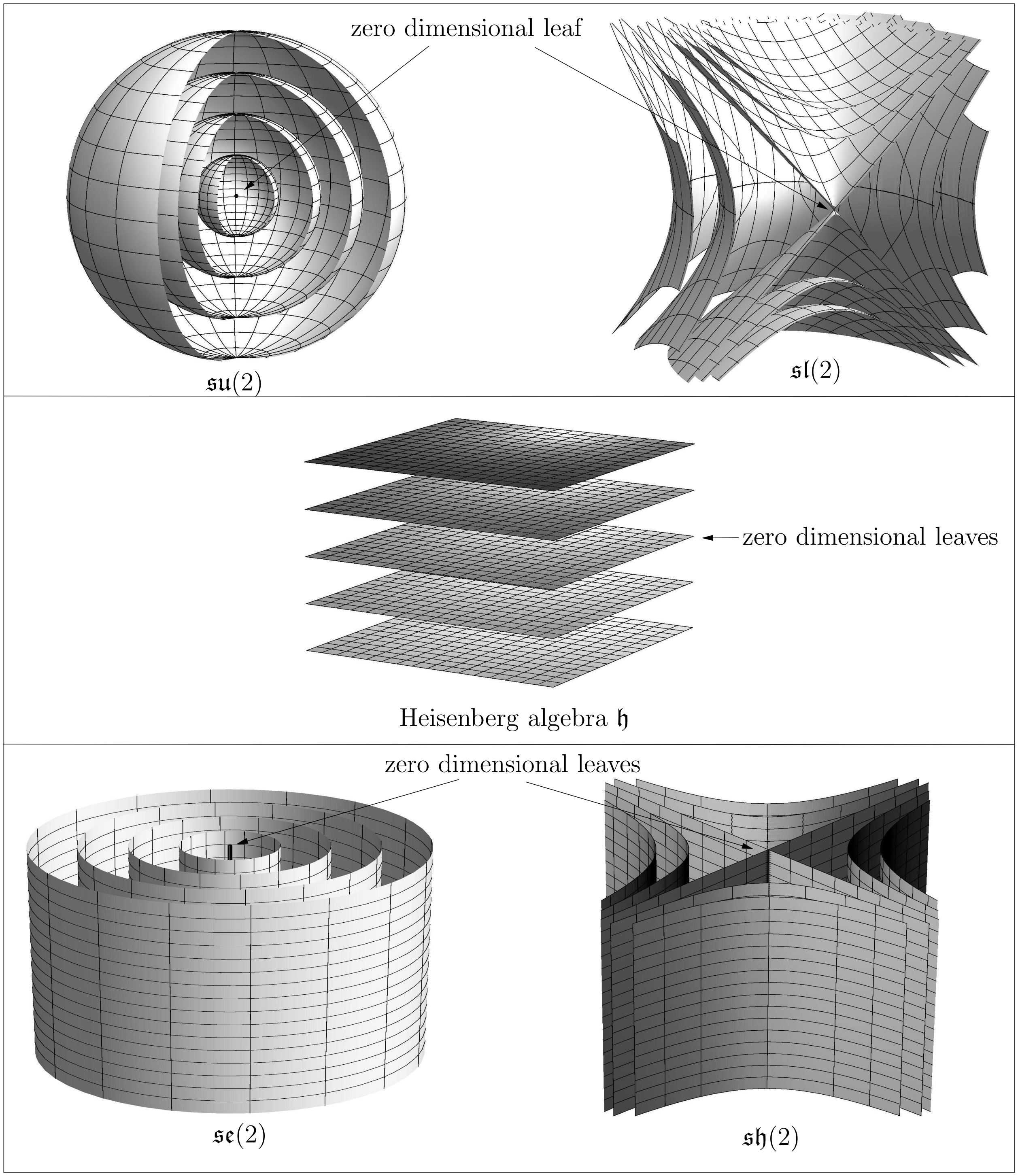



Figure 18.1: Symplectic leaves for 3D Lie algebras. Unimodular cases. In all these cases a non-constant Casimir function exists.

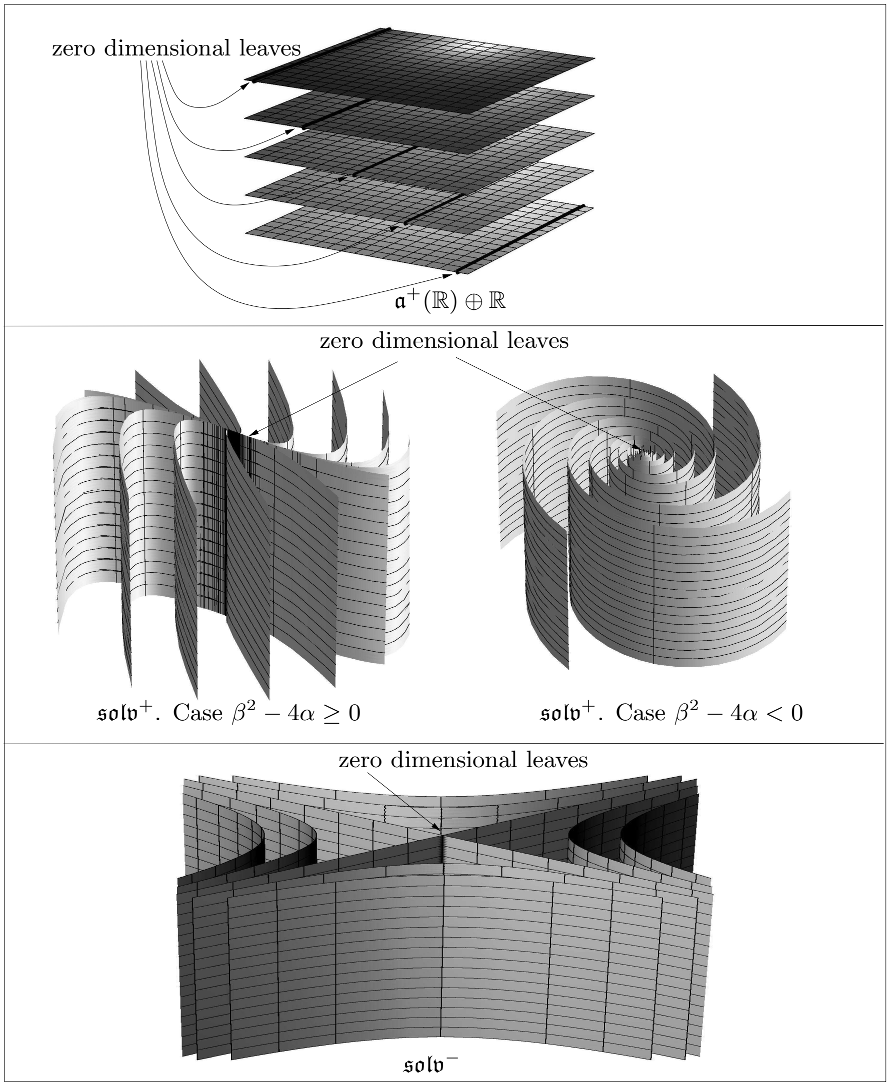



Figure 18.2: Symplectic leaves for 3D Lie algebras. Non-unimodular cases. Notice that in the $\mathfrak{solv}^+$ case, the fact that all two-dimensional symplectic leaves have the same line in their closure (of zero-dimensional symplectic leaves) implies the non existence of non-constant Casimir functions. In the $\mathfrak{solv}^-$ case there exists a non-constant Casimir function (sometimes analytic and sometimes only $\mathcal{C}^\infty$, depending on the eigenvalues).

# Chapter 19

# Asymptotic expansion of the 3D contact exponential map

In this chapter we compute the small time asymptotics of the exponential map in the three-dimensional contact case. We show how the structure of the cut and the conjugate locus is encoded in it, and we express in terms of the curvature invariants.

## 19.1 The exponential map

Let $M$ be a contact sub-Riemannian structure on a 3D manifold, let us fix a local orthonormal frame $f_1, f_2$ for the sub-Riemannian structure, and let $f_0$ be the Reeb vector field associated to the normalized contact form. We refer the reader to Section 17.4 for the corresponding definitions.

We fix the coordinates $(h_0, h_1, h_2)$ dual to the local frame $f_0, f_1, f_2$ (recall that then $\nu_0$ coincides with the normalized contact form). The Hamiltonian vector field $\vec{H}$ associated with the sub-Riemannian Hamiltonian $H$ is then written as follows:

$$\vec{H} = \cos\theta f_1 + \sin\theta f_2 - (h_0 + b)\partial_\theta + a\partial_{h_0}. \tag{19.1}$$

where we have restricted our attention to the level set $H^{-1}(1/2)$, and the coordinate $\theta$ is defined in such a way that $h_1 = \cos\theta$ and $h_2 = \sin\theta$. Recall that the functions $a$ and $b$ are defined as follows

$$a = \{H, h_0\} = \sum_{i,j=1}^{2} c_{0i}^{j} h_i h_j, \qquad b := c_{12}^{1} h_1 + c_{12}^{2} h_2,$$

where $c_{ij}^k$ are structure functions of the Lie algebra defined by the orthonormal frame and the Reeb vector field

$$[f_j, f_i] = \sum_{k=0}^{2} c_{ij}^{k} f_k, \qquad i, j = 0, 1, 2.$$

In the sequel it will be convenient to introduce the coordinate $\rho := -h_0$ for the function that is linear on fibers of $T^*M$ associated with the opposite of the Reeb vector field. The Hamiltonian system (19.1) on the level set $H^{-1}(1/2)$ is then rewritten in the coordinates $(q,\theta,\rho)$ as follows:

$$\begin{cases} \dot{q} = \cos\theta f_1 + \sin\theta f_2 \\ \dot{\theta} = \rho - b \\ \dot{\rho} = -a \end{cases} \tag{19.2}$$

The exponential map based at $q_0 \in M$ is the map that with each time $t > 0$ and every initial covector $(\theta_0,\rho_0) \in T^*_{q_0}M \cap H^{-1}(1/2)$ associates the first component of the solution at time $t$ of the system (19.2), denoted by $\exp_{q_0}(t,\theta_0,\rho_0)$ (or simply $\exp(t,\theta_0,\rho_0)$ when there is no confusion about the base point).

Conjugate points are critical points of the exponential map, i.e., the set of initial covectors such that the differential of the exponential map is not surjective. Hence conjugate points are solutions to the equation

$$\det\left(\frac{\partial \exp}{\partial\theta_0}, \frac{\partial \exp}{\partial\rho_0}, \frac{\partial \exp}{\partial t}\right) = 0. \tag{19.3}$$

The variation of the exponential map along time is always nonzero and independent with respect to variations of the covectors in the set $H^{-1}(1/2)$ (see also Section 8.11 and Proposition 8.42). Hence (19.3) is equivalent to

$$\frac{\partial \exp}{\partial\theta_0} \wedge \frac{\partial \exp}{\partial\rho_0} = 0, \tag{19.4}$$

meaning that the two vectors are linearly dependent.

### 19.1.1 The nilpotent case

We start by studying the exponential map on the Heisenberg group. This corresponds to the case when the functions $a$ and $b$ in (19.2) vanish identically, namely the system

$$\begin{cases} \dot{q} = \cos\theta f_1 + \sin\theta f_2 \\ \dot{\theta} = \rho \\ \dot{\rho} = 0 \end{cases} \tag{19.5}$$

Let us first recover, in this notation, the conjugate locus in the case of the Heisenberg group. Let us denote coordinates on the manifold $\mathbb{R}^3$ as follows

$$q = (x,z), \qquad x = (x_1,x_2) \in \mathbb{R}^2, z \in \mathbb{R}. \tag{19.6}$$

Notice moreover that in this case the Reeb vector field is proportional to $\partial_z$ and its dual coordinate $\rho$ is constant along trajectories. There are two possible cases:

(i). $\rho = 0$. Then the solution is a straight line contained in the plane $z = 0$ and is optimal for all time.

(ii). $\rho \neq 0$. In this case we claim that the equation (19.4) is equivalent to the following

$$\frac{\partial x}{\partial\theta_0} \wedge \frac{\partial x}{\partial\rho_0} = 0. \tag{19.7}$$

In fact, by Gauss' Lemma (Proposition 8.42), the covector $p = (p_x, \rho)$ at the final point annihilates the image of the differential of the exponential map restricted to the level set, i.e.

$$0 = \left\langle p, \frac{\partial \exp}{\partial \theta_0} \right\rangle = \left\langle p_x, \frac{\partial x}{\partial \theta_0} \right\rangle + \rho \frac{\partial z}{\partial \theta_0}, \tag{19.8}$$

$$0 = \left\langle p, \frac{\partial \exp}{\partial \rho_0} \right\rangle = \left\langle p_x, \frac{\partial x}{\partial \rho_0} \right\rangle + \rho \frac{\partial z}{\partial \rho_0}, \tag{19.9}$$

and since $\rho \neq 0$ it follows that among the three vectors

$$\begin{pmatrix} \dfrac{\partial x_1}{\partial \theta_0} \\ \dfrac{\partial x_1}{\partial \rho_0} \end{pmatrix} \qquad \begin{pmatrix} \dfrac{\partial x_2}{\partial \theta_0} \\ \dfrac{\partial x_2}{\partial \rho_0} \end{pmatrix} \qquad \begin{pmatrix} \dfrac{\partial z}{\partial \theta_0} \\ \dfrac{\partial z}{\partial \rho_0} \end{pmatrix} \tag{19.10}$$

the third one is always a linear combination of the first two.

**Proposition 19.1.** *The first conjugate time is* $t_c(\theta_0, \rho_0) = 2\pi/|\rho_0|$.

*Proof.* In the standard coordinates $(x_1, x_2, z)$ the two vector fields $f_1$ and $f_2$ defining the orthonormal frame are

$$f_1 = \partial_{x_1} - \frac{x_2}{2}\partial_z, \qquad f_2 = \partial_{x_2} + \frac{x_1}{2}\partial_z.$$

Thus, the first two coordinates of the horizontal part of the Hamiltonian system satisfy

$$\begin{cases} \dot{x}_1 = \cos\theta \\ \dot{x}_2 = \sin\theta \end{cases} \tag{19.11}$$

It is then easy to integrate the $x$-part of the exponential map being $\theta(t) = \theta_0 + \rho t$ (recall that $\rho \equiv \rho_0$ and, without loss of generality we can assume $\rho > 0$)

$$x(t; \theta_0, \rho_0) = \int_0^t \begin{pmatrix} \cos(\theta_0 + \rho s) \\ \sin(\theta_0 + \rho s) \end{pmatrix} ds = \int_{\theta_0}^{\theta_0 + t} \begin{pmatrix} \cos \rho s \\ \sin \rho s \end{pmatrix} ds \tag{19.12}$$

Due to the symmetry with respect to rotations around the $z$ axis, the determinant of the Jacobian map will not depend on $\theta_0$. Hence to compute the determinant of the Jacobian it is enough to compute partial derivatives at $\theta_0 = 0$

$$\frac{\partial x}{\partial \theta_0} = \begin{pmatrix} \cos \rho t - 1 \\ \sin \rho t \end{pmatrix}$$

$$\frac{\partial x}{\partial \rho_0} = -\frac{1}{\rho^2} \begin{pmatrix} \sin \rho t \\ 1 - \cos \rho t \end{pmatrix} + \frac{t}{\rho} \begin{pmatrix} \cos \rho t \\ \sin \rho t \end{pmatrix}$$

and denoting by $\tau := \rho t$ one can compute

$$\begin{aligned} \frac{\partial x}{\partial \theta_0} \wedge \frac{\partial x}{\partial \rho_0} &= \frac{1}{\rho^2} \det \begin{pmatrix} \cos\tau - 1 & \tau\cos\tau - \sin\tau \\ \sin\tau & -1 + \tau\sin\tau + \cos\tau \end{pmatrix}, \\ &= \frac{1}{\rho^2}(\tau \sin\tau + 2\cos\tau - 2). \end{aligned}$$

The fact that $t_c = 2\pi/|\rho|$ follows from the fact that $2\pi$ is the first positive root of the equation $\tau\sin\tau + 2\cos\tau - 2 = 0$. Notice that it is indeed a simple root. $\square$

## 19.2 General case: second order asymptotic expansion

Let us consider the Hamiltonian system for the general 3D contact case

$$\begin{cases} \dot{q} = f_\theta := \cos\theta f_1 + \sin\theta f_2 \\ \dot{\theta} = \rho - b \\ \dot{\rho} = -a \end{cases} \tag{19.13}$$

We are going to study the asymptotic expansion for our system for the initial parameter $\rho_0 \to \pm\infty$. To this aim, it is convenient to introduce the change of variables $r := 1/\rho$ and denote by $\nu := r(0) = 1/\rho_0$ its initial value. Notice that $\rho$ is no more constant in the general case and $\rho_0 \to \infty$ implies $\nu \to 0$.

The main result of this section says that the conjugate time for the perturbed system is a perturbation of the conjugate time of the nilpotent case, where the perturbation has no term of order 2.

**Proposition 19.2.** *The conjugate time $t_c(\theta_0, \nu)$ is a smooth function of the parameter $\nu$ for $\nu > 0$. Moreover for $\nu \to 0$*

$$t_c(\theta_0, \nu) = 2\pi|\nu| + O(|\nu|^3).$$

### 19.2.1 Proof of Proposition 19.2: second order asymptotics

To prove Proposition 19.2, let us introduce a new time variable $\tau$ such that $\frac{dt}{d\tau} = r$. If we denote by $\dot{F}$ the derivative of a function $F$ with respect to the new time $\tau$, the system (19.13) is rewritten in the new coordinate system $(q, \theta, r)$ (where we recall $r = 1/\rho$), as follows

$$\begin{cases} \dot{q} = r f_\theta \\ \dot{\theta} = 1 - rb \\ \dot{r} = r^3 a \\ \dot{t} = r \end{cases} \tag{19.14}$$

To compute the asymptotics of the conjugate time, it is also convenient to consider a system of coordinates, depending on a parameter $\varepsilon$, corresponding to the quasi-homogeneous blow up of the sub-Riemannian structure at $q_0$ and converging to the nilpotent approximation.

More precisely, first we consider the equation $\dot{q} = r f_\theta$ in any linearly adapted coordinates on $M$. These coordinates are automatically privileged coordinates, see Sec. 10.4.1 and Example 10.31. Next, we apply a quasi-homogeneous blow-up $\Phi_\varepsilon = \delta_{1/\varepsilon}$ defined via dilations, see Sec. 10.4.1 and Remark 10.42. This permits to introduce a new system of privileged coordinates on M depending on a parameter $\varepsilon > 0$. After the blow-up we get $(\Phi_\varepsilon)_* f_\theta = \frac{1}{\varepsilon} f_\theta^\varepsilon$ where

$$f_\theta^\varepsilon = \widehat{f}_\theta + \varepsilon f_\theta^{(0)} + \varepsilon^2 f_\theta^{(1)} + \dots$$

and $\widehat{f}_\theta$ is the Heisenberg nilpotent approximation.

Accordingly to this change of coordinates we have the equalities

$$(\Phi_\varepsilon)_* f_i = \frac{1}{\varepsilon} f_i^\varepsilon, \qquad (\Phi_\varepsilon)_* f_0 = \frac{1}{\varepsilon^2} f_0^\varepsilon,$$

where $f_0^\varepsilon$ is the Reeb vector field defined by the orthonormal frame $f_1^\varepsilon, f_2^\varepsilon$. Moreover denoting by $a^\varepsilon, b^\varepsilon$ the function $a$ and $b$ written in the new coordinate system, we have

$$b = \frac{1}{\varepsilon} b^\varepsilon, \qquad a = \frac{1}{\varepsilon^2} a^\varepsilon.$$

Let us now define, for fixed $\varepsilon$, the variable $w$ such that $r = \varepsilon w$.

**Proposition 19.3.** *The system* (19.14) *is rewritten in these variables as follows*

$$\begin{cases} \dot{q} = w f_\theta^\varepsilon \\ \dot{\theta} = 1 - w b^\varepsilon \\ \dot{w} = w^3 a^\varepsilon \\ \dot{t} = \varepsilon w \end{cases} \tag{19.15}$$

Notice that the dynamical system is written in a coordinate system that depends on $\varepsilon$. Moreover the initial asymptotic for $\rho_0 \to \infty$, corresponding to $r(0) \to 0$, is now reduced to fix an initial value $w(0) = 1$ and send $\varepsilon \to 0$.

Consider some linearly adapted coordinates $(x, z)$, with $x \in \mathbb{R}^2$ and $z \in \mathbb{R}$ (cf. Definition 10.30). If we denote by $q^\varepsilon = (x^\varepsilon, z^\varepsilon)$ the solution of the horizontal part of the $\varepsilon$-system (19.15), conjugate points are solutions of the equation

$$\left. \frac{\partial q^\varepsilon}{\partial \theta_0} \wedge \frac{\partial q^\varepsilon}{\partial w_0} \right|_{w_0 = 1} = 0.$$

As in Section 19.1.1, one can check that this condition is equivalent to

$$\left. \frac{\partial x^\varepsilon}{\partial \theta_0} \wedge \frac{\partial x^\varepsilon}{\partial w_0} \right|_{w_0 = 1} = 0.$$

Notice that the original parameters $(t, \theta_0, \rho_0)$ parametrizing the trajectories in the exponential map correspond to a conjugate point if the corresponding parameters $(\tau, \theta_0, \varepsilon)$ satisfy

$$\varphi(\tau, \varepsilon, \theta_0) := \left. \frac{\partial x^\varepsilon}{\partial \theta_0} \wedge \frac{\partial x^\varepsilon}{\partial w_0} \right|_{w_0 = 1} = 0. \tag{19.16}$$

For $\varepsilon = 0$, i.e., for the nilpotent approximation, the first conjugate time is $\tau_c = 2\pi$, and moreover it is a simple root. Thus one gets

$$\varphi(2\pi, 0, \theta_0) = 0, \qquad \frac{\partial \varphi}{\partial \tau}(2\pi, 0, \theta_0) \neq 0. \tag{19.17}$$

Hence the implicit function theorem guarantees that there exists a smooth function $\tau_c(\varepsilon, \theta_0)$ such that $\tau_c(0, \theta_0) = 2\pi$ and

$$\varphi(\tau_c(\varepsilon, \theta_0), \varepsilon, \theta_0) = 0. \tag{19.18}$$

In other words $\tau_c(\varepsilon, \theta_0)$ computes the conjugate time $\tau$ associated with parameters $\varepsilon, \theta_0$. By smoothness of $\tau_c$ one immediately has the expansion for $\varepsilon \to 0$

$$\tau_c(\varepsilon, \theta_0) = 2\pi + O(\varepsilon).$$

Now the statement of the proposition is rewritten in terms of the function $\tau_c$ as follows

$$\tau_c(\varepsilon, \theta_0) = 2\pi + O(\varepsilon^2). \tag{19.19}$$

Differentiating the identity (19.18) with respect to $\varepsilon$ one has

$$\frac{\partial \varphi}{\partial \tau}\frac{\partial \tau_c}{\partial \varepsilon} + \frac{\partial \varphi}{\partial \varepsilon} = 0,$$

hence, thanks to (19.17), the expansion (19.19) holds if and only if $\dfrac{\partial \varphi}{\partial \varepsilon}(2\pi, 0, \theta_0) = 0$.

Moreover differentiating the expression (19.16) with respect to $\varepsilon$ one has

$$\frac{\partial \varphi}{\partial \varepsilon}(2\pi, 0, \theta_0) = \left.\frac{\partial^2 x^\varepsilon}{\partial \varepsilon \partial \theta_0} \wedge \frac{\partial x^\varepsilon}{\partial w_0} - \frac{\partial^2 x^\varepsilon}{\partial \varepsilon \partial w_0} \wedge \frac{\partial x^\varepsilon}{\partial \theta_0}\right|_{w_0=1, \varepsilon=0, \tau=2\pi}$$

The second term vanishes since at $\varepsilon = 0$ is the Heisenberg case, whose horizontal part at $\tau = 2\pi$ does not depend on $\theta_0$. Hence it is sufficient to prove that

$$\left.\frac{\partial^2 x^\varepsilon}{\partial \varepsilon \partial \theta_0}\right|_{\varepsilon=0, \tau=2\pi} = 0. \tag{19.20}$$

which is a consequence of the following lemma.

**Lemma 19.4.** *The quantity* $\left.\dfrac{\partial x^\varepsilon}{\partial \varepsilon}\right|_{\varepsilon=0, \tau=2\pi}$ *does not depend on* $\theta_0$.

*Proof.* To prove the statement it will be enough to find the first order expansion in $\varepsilon$ of the solution of the system (19.15).

Recall that when $\varepsilon = 0$ the system corresponds to the Heisenberg case, i.e., we have $a^\varepsilon|_{\varepsilon=0} = 0$ and $b^\varepsilon|_{\varepsilon=0} = 0$. This gives the expansion of $w$ (recall that $w(0) = w_0 = 1$)

$$w(t) = w(0) + \int_0^t a^\varepsilon(\tau) w^3(\tau) d\tau \quad \Rightarrow \quad w(t) = 1 + \varepsilon \alpha(t) + O(\varepsilon^2). \tag{19.21}$$

Notice that $a^\varepsilon$, and thus $\alpha = \frac{\partial a^\varepsilon}{\partial \varepsilon}\big|_{\varepsilon=0}$, is a homogeneous polynomial in $\cos\theta$ and $\sin\theta$ (cf. also Section 17.4).

Analogously we have $b^\varepsilon = \varepsilon \langle \beta, u \rangle + O(\varepsilon^2)$, where $\langle \beta, u \rangle = \beta_1 u_1 + \beta_2 u_2$ and $\beta$ denotes the (constant) coefficient of weight zero in the expansion of $b^\varepsilon$ with respect to $\varepsilon$. Denoting $u(\theta) = (\cos\theta, \sin\theta)$, the equation for $\theta$ then is reduced to

$$\dot{\theta} = 1 - \varepsilon \langle \beta, u(\theta) \rangle + O(\varepsilon^2), \qquad \theta(0) = \theta_0.$$

This equation can be integrated and one gets

$$\left.\frac{\partial \theta}{\partial \varepsilon}\right|_{\varepsilon=0} = -\int_0^t \langle \beta, u(\theta(\tau)) \rangle \, d\tau = \left\langle \beta, u'(\theta_0 + t) - u'(\theta_0) \right\rangle, \tag{19.22}$$

where $u'(\theta) = (-\sin\theta, \cos\theta)$.

Next we are going to use (19.21) and (19.22) to compute the derivative of $x^\varepsilon$ wrt $\varepsilon$. The equation for the horizontal part of (19.15) can be expanded in $\varepsilon$ as follows

$$\dot{x} = u(\theta) + \varepsilon(\alpha u(\theta) + f_\theta^{(0)}(x)) + O(\varepsilon^2)$$

where the first term is Heisenberg, and $f_\theta^{(0)}$ is the term of weight zero of $f_\theta$, which is linear with respect to $x_1$ and $x_2$ (recall that this is the zero order part of the vector field $f_u$ along $\partial_x$, hence only $x$ variables appear and have order 1). To compute the derivative of the solution with respect to the parameter $\varepsilon$ we use the following general fact.

**Lemma 19.5.** *Let $\phi(\varepsilon, t)$ denote the solution of the differential equation $\dot{y} = F(\varepsilon, y)$ with fixed initial condition $y(0) = y_0$. Then the derivative $\frac{\partial \phi}{\partial \varepsilon}$ satisfies the following linear ODE*

$$\frac{d}{dt}\frac{\partial \phi}{\partial \varepsilon}(\varepsilon, t) = \frac{\partial F}{\partial y}(\varepsilon, \phi(\varepsilon, t))\frac{\partial \phi}{\partial \varepsilon}(\varepsilon, t) + \frac{\partial F}{\partial \varepsilon}(\varepsilon, \phi(\varepsilon, t)).$$

We apply the above lemma when $y = (x, \theta)$ and $F = (F^x, F^\theta)$ and we compute at $\varepsilon = 0$. In particular we need the solution of the original system at $\varepsilon = 0$

$$\phi(0, t) = (\bar{x}(t), \bar{\theta}(t)), \qquad \bar{\theta}(t) = \theta_0 + t, \qquad \bar{x}(t) = u'(\theta_0) - u'(\theta_0 + t).$$

Then by Lemma 19.5 we have

$$\frac{d}{dt}\frac{\partial x}{\partial \varepsilon} = \frac{\partial F^x}{\partial x}\frac{\partial x}{\partial \varepsilon} + \frac{\partial F^x}{\partial \theta}\frac{\partial \theta}{\partial \varepsilon} + \frac{\partial F^x}{\partial \varepsilon}$$

Computing the derivatives at $\varepsilon = 0$ gives

$$\left.\frac{\partial F^x}{\partial x}\right|_{\varepsilon=0} = 0, \qquad \left.\frac{\partial F^x}{\partial \theta}\right|_{\varepsilon=0} = u'(\bar{\theta}(t)), \qquad \left.\frac{\partial F^x}{\partial \varepsilon}\right|_{\varepsilon=0} = \alpha(t)u(\bar{\theta}(t)) + f_{\bar{\theta}(t)}^{(0)}(\bar{x}(t))$$

and we obtain the equation for the derivative $\frac{\partial x}{\partial \varepsilon}$ (recall that $\bar{\theta}(t) = \theta_0 + t$)

$$\left.\frac{d}{dt}\frac{\partial x}{\partial \varepsilon}\right|_{\varepsilon=0} = u'(\theta_0 + t)\left.\frac{\partial \theta}{\partial \varepsilon}\right|_{\varepsilon=0} + \alpha(t)u(\theta_0 + t) + f_{\theta_0+t}^{(0)}(u'(\theta_0) - u'(\theta_0 + t)).$$

If we introduce the new variable $s = \theta_0 + t$ we can rewrite the last equation in the more compact form

$$\left.\frac{d}{ds}\frac{\partial x}{\partial \varepsilon}\right|_{\varepsilon=0} = u'(s)\frac{\partial \theta}{\partial \varepsilon} + \alpha(s - \theta_0)u(s) + f_s^{(0)}(u'(\theta_0) - u'(s)).$$

Integrating and using (19.22) one has

$$\left.\frac{\partial x}{\partial \varepsilon}\right|_{\varepsilon=0,\tau=2\pi} = \int_{\theta_0}^{\theta_0+2\pi} \big\langle \beta, u'(s) - u'(\theta_0)\big\rangle\, u'(s)ds + \int_{\theta_0}^{\theta_0+2\pi} \alpha(s - \theta_0)u(s)ds$$
$$+ \int_{\theta_0}^{\theta_0+2\pi} f_s^{(0)}(u'(\theta_0) - u'(s))ds.$$

One can see that the last expression does not depend on $\theta_0$, since we compute integrals of a periodic functions over its period. This finishes the proof of Lemma 19.4, hence the proof of the Proposition 19.2.

## 19.3 General case: higher order asymptotic expansion

Next we continue our analysis about the structure of the conjugate locus for a 3D contact structure by studying higher order asymptotics. In this section we determine the coefficient of order 3 in the asymptotic expansion of the conjugate locus. Namely we have the following result, whose proof is postponed to Section 19.3.1.

**Theorem 19.6.** *In a system of local coordinates around $q_0 \in M$ one has the expansion*

$$\mathrm{Con}_{q_0}(\theta_0, \nu) = q_0 \pm \pi f_0 |\nu|^2 \pm \pi (a' f_{\theta_0} - a f_{\theta_0'}) |\nu|^3 + O(|\nu|^4), \qquad \nu \to 0^{\pm}. \tag{19.23}$$

*If we choose coordinates such that $a = 2\chi h_1 h_2$ one gets*

$$\mathrm{Con}_{q_0}(\theta_0, \nu) = q_0 \pm \pi f_0 |\nu|^2 \pm 2\pi \chi(q_0)(\cos^3\theta f_2 - \sin^3\theta f_1)|\nu|^3 + O(|\nu|^4), \qquad \nu \to 0^{\pm}. \tag{19.24}$$

*Moreover for the conjugate length we have the expansion*

$$\ell_c(\theta_0, \nu) = 2\pi|\nu| - \pi\kappa|\nu|^3 + O(|\nu|^4), \qquad \nu \to 0^{\pm}. \tag{19.25}$$

Analogous formulas can be obtained for the asymptotics of the cut locus at a point $q_0$ where the invariant $\chi$ is non vanishing.

**Theorem 19.7.** *Assume $\chi(q_0) \neq 0$. In a system of local coordinates around $q_0 \in M$ such that $a = 2\chi u_1 u_2$ one gets*

$$\mathrm{Cut}_{q_0}(\theta, \nu) = q_0 \pm \pi\nu^2 f_0(q_0) \pm 2\pi\chi(q_0)\cos\theta f_1(q_0)\nu^3 + O(\nu^4), \qquad \nu \to 0^{\pm}$$

*Moreover the cut length satisfies*

$$\ell_{cut}(\theta, \nu) = 2\pi|\nu| - \pi(\kappa + 2\chi\sin^2\theta)|\nu|^3 + O(\nu^4), \qquad \nu \to 0^{\pm} \tag{19.26}$$

We can collect the information given by the asymptotics of the conjugate and the cut loci in Figure 19.1.

All geometrical information about the structure of these sets is encoded in a pair of quadratic forms defined on the fiber at the base point $q_0$, namely the curvature $\mathcal{R}$ and the sub-Riemannian Hamiltonian $H$.

Recall that the sub-Riemannian Hamiltonian encodes the information about the distribution and about the metric defined on it (see Exercise 4.32).

Let us consider the kernel of the sub-Riemannian Hamiltonian

$$\ker H = \{\lambda \in T_q^*M : \langle \lambda, v\rangle = 0,\ \forall\, v \in \mathcal{D}_q\} = \mathcal{D}_q^{\perp}. \tag{19.27}$$

The restriction of $\mathcal{R}$ to the 1-dimensional subspace $\mathcal{D}_q^{\perp}$ for every $q \in M$, is a strictly positive quadratic form. Moreover it is equal to $1/10$ when evaluated on the Reeb vector field. Hence the curvature $\mathcal{R}$ encodes both the contact form $\omega$ and its normalization.

If we denote by $\mathcal{D}_q^*$ the orthogonal complement[1] of $\mathcal{D}_q^{\perp}$ in the fiber with respect to $\mathcal{R}$, we have that $\mathcal{R}$ is a quadratic form on $\mathcal{D}_q^*$ and, by using the Euclidean metric defined by $H$ on $\mathcal{D}_q$, we can identify it with a symmetric operator.

[1]this is indeed isomorphic to the space of linear functionals defined on $\mathcal{D}_q$.

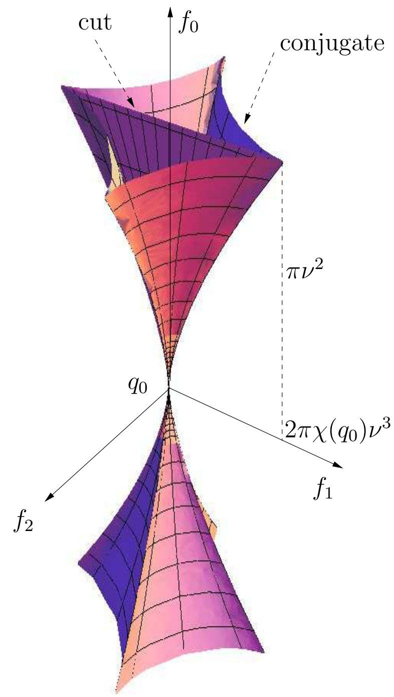



Figure 19.1: Asymptotic structure of cut and conjugate locus

As we explained in the previous chapter, at each $q_0$ where $\chi(q_0) \neq 0$ there always exists a frame such that

$$\{H, h_0\} = 2\chi h_1 h_2$$

and in this frame we can express the restriction of $\mathcal{R}$ to $\mathcal{D}_q^*$ (corresponding to the set $\{h_0 = 0\}$) on this subspace as follows (see Section 17.4)

$$10\mathcal{R} = (\kappa + 3\chi)h_1^2 + (\kappa - 3\chi)h_2^2.$$

From this formula it is easy to recover the two invariants $\chi, \kappa$ considering

$$\operatorname{trace}(10\mathcal{R}\big|_{h_0=0}) = 2\kappa, \qquad \operatorname{discr}(10\mathcal{R}\big|_{h_0=0}) = 36\chi^2,$$

where the discriminant of an operator $Q$, defined on a two-dimensional space, is defined as the square of the difference of its eigenvalues, and can be compute by the formula $\operatorname{discr}(Q) = \operatorname{trace}^2(Q) - 4\det(Q)$.

The cubic term of the asymptotic expansion of the conjugate locus parametrizes an astroid. The cuspidal directions of the astroid are given by the eigenvectors of $\mathcal{R}$, and the cut locus intersects the conjugate locus exactly at the cuspidal points, in the direction of the eigenvector of $\mathcal{R}$ corresponding to the larger eigenvalue.

Finally the "size" of the cut locus increases for bigger values of $\chi$, while $\kappa$ is involved in the length of curves arriving at cut/conjugate locus.

*Remark* 19.8. The expression of the conjugate locus (resp. cut locus) given in Theorem 19.6 (resp. Theorem 19.7) gives the truncation up to order three of the asymptotics of the conjugate locus (resp. cut locus) of the exponential map. It is indeed possible to show that this is actually the *exact* cut locus corresponding to the *truncated* exponential map at order three. For a discussion on this point we refer to Section 19.3.4.

### 19.3.1 Proof of Theorem 19.6: asymptotics of the exponential map

The proof of Theorem 19.6 requires a careful analysis of the asymptotic of the exponential map. Let us consider again our Hamiltonian system in the form (19.14)

$$\begin{cases} \dot{q} = r f_\theta \\ \dot{\theta} = 1 - rb \\ \dot{r} = r^3 a \\ \dot{t} = r \end{cases} \tag{19.28}$$

where we recall that equations are written with respect to the time $\tau$. In particular, since we restrict on the level set $H^{-1}(1/2)$, the trajectories are parametrized by length and the time $t$ coincides with the length of the curve. Thus in what follows we replace the variable $t$ by $\ell$.

Next, we consider a last change of the time variable. Namely we parametrize trajectories by the coordinate $\theta$. In other words we rewrite again the equations in such a way that $\dot{\theta} = 1$ and the dot will denote derivative with respect to $\theta$. The equations are rewritten in the following form:

$$\begin{cases} \dot{q} = \dfrac{r}{1 - rb} f_\theta \\ \dot{\theta} = 1 \\ \dot{r} = \dfrac{r^3}{1 - rb} a \\ \dot{\ell} = \dfrac{r}{1 - rb} \end{cases} \tag{19.29}$$

where we recall that $f_\theta = \cos\theta f_1 + \sin\theta f_2$ and we denote $\nu = r(0)$. Moreover we define $F(t;\theta_0,\nu) := q(t+\theta_0;\theta_0,\nu)$, where $q(\theta_0;\theta_0,\nu) = q_0$. This means that the curve that corresponds to initial parameter $\theta_0$ start from $q_0$ at time equal to $\theta_0$.

Notice that in (19.29) we can solve the equation for $r = r(\tau)$ and substitute it in the first equation. In this way we can write the trajectory as an integral curve of the nonautonomous vector field

$$F(t;\theta_0,\nu) = q_0 \circ Q_t^{\theta_0,\nu}, \qquad Q_t^{\theta_0,\nu} = \overrightarrow{\exp} \int_{\theta_0}^{\theta_0+t} \frac{r(\tau)}{1 - r(\tau)b(\tau)} f_\tau d\tau.$$

To simplify the notation in what follows we denote the flow $Q_t^{\theta_0,\nu}$ simply by $Q_t$ and by $V_t$ the non-autonomous vector field defined by this flow

$$Q_t = \overrightarrow{\exp} \int_{\theta_0}^{\theta_0+t} V_\tau d\tau, \qquad V_\tau := \frac{r(\tau)}{1 - r(\tau)b(\tau)} f_\tau. \tag{19.30}$$

We start by analyzing the asymptotics of the end-point map after time $t = 2\pi$.

**Lemma 19.9.** $F(2\pi;\theta_0,\nu) = q_0 - \pi f_0(q_0)\nu^2 + O(\nu^3)$

*Proof.* From (19.29), recalling that $r(0) = \nu$, it is easy to see that $r$ satisfies the identity

$$r(t) = \nu + \widetilde{r}(t)\nu^3 = \nu + O(\nu^3)$$

for some smooth function $\widetilde{r}(t)$. Thus, to find the second order term in $\nu$ of the endpoint map $F(2\pi; \theta, \nu)$, we can then assume that $r$ is constantly equal to $\nu = r(0)$.

Using the Volterra expansion (cf. (6.17))

$$\overrightarrow{\exp} \int_{\theta_0}^{\theta_0+2\pi} V_\tau d\tau = \left( \mathrm{Id} + \int_{\theta_0}^{\theta_0+2\pi} V_\tau d\tau + \iint\limits_{\theta_0 \le \tau_2 \le \tau_1 \le \theta_0+2\pi} V_{\tau_2} \odot V_{\tau_1} d\tau_1 d\tau_2 + \dots \right) \tag{19.31}$$

and substituting $r(\tau) \equiv \nu$ we have the following expansion for the first term in (19.31):

$$\begin{aligned}
\int_{\theta_0}^{\theta_0+2\pi} V_\tau d\tau = \int_{\theta_0}^{\theta_0+2\pi} \frac{\nu}{1 - \nu b(\tau)} f_\tau d\tau &= \int_{\theta_0}^{\theta_0+2\pi} \nu(1 + \nu b(\tau) + O(\nu^2)) f_\tau \, d\tau, \\
&= \nu \int_{\theta_0}^{\theta_0+2\pi} f_\tau d\tau + \nu^2 \int_{\theta_0}^{\theta_0+2\pi} b(\tau) f_\tau d\tau + O(\nu^3) \\
&= \nu^2 \int_{\theta_0}^{\theta_0+2\pi} b(\tau) f_\tau d\tau + O(\nu^3)
\end{aligned}$$

Notice that the first order term in $\nu$ vanishes since we integrate over a period and $\int_{\theta_0}^{\theta_0+2\pi} f_\tau d\tau = 0$. The second term in (19.31) can be rewritten using Lemma 8.30

$$\begin{aligned}
\iint\limits_{\theta_0 \le \tau_2 \le \tau_1 \le \theta_0+2\pi} V_{\tau_2} \odot V_{\tau_1} d\tau_1 d\tau_2 &= \frac{1}{2} \int_{\theta_0}^{\theta_0+2\pi} V_\tau d\tau \odot \int_{\theta_0}^{\theta_0+2\pi} V_\tau d\tau + \iint\limits_{\theta_0 \le \tau_2 \le \tau_1 \le \theta_0+2\pi} [V_{\tau_2}, V_{\tau_1}] d\tau_1 d\tau_2 \\
&= \frac{\nu^2}{2} \left( \int_{\theta_0}^{\theta_0+2\pi} f_\tau d\tau \odot \int_{\theta_0}^{\theta_0+2\pi} f_\tau d\tau + \iint\limits_{\theta_0 \le \tau_2 \le \tau_1 \le \theta_0+2\pi} [f_{\tau_2}, f_{\tau_1}] d\tau_1 d\tau_2 \right) \\
&= \frac{\nu^2}{2} \iint\limits_{\theta_0 \le \tau_2 \le \tau_1 \le \theta_0+2\pi} [f_{\tau_2}, f_{\tau_1}] d\tau_1 d\tau_2
\end{aligned}$$

where we used again $\int_{\theta_0}^{\theta_0+2\pi} f_\tau d\tau = 0$. Notice that higher order terms in the Volterra expansions are $O(\nu^3)$. Collecting together the two expansions and recalling that

$$[f_2, f_1] = f_0 + \alpha_1 f_1 + \alpha_2 f_2$$

one easily obtains

$$\begin{aligned}
F(2\pi; \theta_0, \nu) &= q_0 + \nu^2 \left( \int_{\theta_0}^{\theta_0+2\pi} b(t) f_t \, dt + \frac{1}{2} \left[ \int_{\theta_0}^{t} f_\tau d\tau, f_t \right] dt \right) + O(\nu^3) \\
&= q_0 - \pi \nu^2 f_0(q_0) + O(\nu^3)
\end{aligned} \tag{19.32}$$

Notice that the factor $\pi$ in (19.32) comes out from the evaluation of integrals of kind $\int_{\theta_0}^{\theta_0+2\pi} \cos^2 \tau d\tau$ and $\int_{\theta_0}^{\theta_0+2\pi} \sin^2 \tau d\tau$. □

Next we prove a symmetry of the exponential map

**Lemma 19.10.** $F(t;\theta_0,\nu) = F(t;\theta_0+\pi,-\nu)$

*Proof.* It is a direct consequence of our geodesic equation. Recall that $F(t;\theta_0,\nu) = q(t+\theta_0;\theta_0,\nu)$, is the solution of the system, with initial condition $q(\theta_0;\theta_0,\nu) = q_0$.

Applying the transformation $t \mapsto t+\pi$ and $\nu \to -\nu$ we see that the right hand side of $\dot{q}$ in (19.29) is preserved while the right hand side of $\dot{r}$ changes sign (we use that $u_i(t+\pi) = -u_i(t)$, hence $a(t+\pi) = a(t)$ and $b(t+\pi) = -b(t)$). Then, if $(q(t), r(t))$ is a solution of the system, the pair $(q(t+\pi), -r(t+\pi))$ is also a solution. The lemma follows. □

The symmetry property just proved permits to characterize all odd terms in the expansion in $\nu$ of the exponential map at $t = 2\pi$.

**Corollary 19.11.** *Consider the formal expansion*

$$F(2\pi;\theta,\nu) \simeq \sum_{n=0}^{\infty} q_n(\theta)\nu^n.$$

*We have the following identities*

*(i)* $q_n(\theta+\pi) = (-1)^n q_n(\theta)$,

*(ii)* $q_{2n+1}(\theta) = -\dfrac{1}{2}\displaystyle\int_\theta^{\theta+\pi} \frac{dq_{2n+1}}{d\theta}(\tau)d\tau.$

*Proof.* This is an immediate consequence of Lemma 19.10 and the identity

$$2q_{2n+1}(\theta) = q_{2n+1}(\theta) - q_{2n+1}(\theta+\pi) = -\int_\theta^{\theta+\pi} \frac{dq_{2n+1}}{d\theta}(\tau)d\tau.$$

□

We already computed the terms $q_1(\theta)$ and $q_2(\theta)$. To find $q_3(\theta)$ we start by computing the derivative of the map $F$ with respect to $\theta$.

**Lemma 19.12.** $\dfrac{\partial F}{\partial \theta_0}(2\pi;\theta_0,\nu) = -\pi[f_0, f_{\theta_0}]_{q_0}\nu^3 + O(\nu^4)$

*Proof.* We stress that, since we are now interested to third order term in $\nu$, we can no more assume that $r(\tau)$ is constant. Differentiating (3.74) with respect to $\theta$ gives two terms as follows:

$$\begin{aligned}\frac{\partial F}{\partial \theta_0} &= \frac{\partial}{\partial \theta_0}(q_0 \circ Q_t) = q_0 \circ \frac{\partial}{\partial \theta_0}\left(\overrightarrow{\exp}\int_\theta^{\theta+2\pi} V_\tau d\tau\right) \\ &= q_0 \circ (Q_{2\pi} \circ V_{\theta_0+2\pi} - V_{\theta_0} \circ Q_{2\pi}) \qquad (19.33)\end{aligned}$$

Next let us rewrite

$$\begin{aligned}Q_{2\pi} \circ V_{\theta_0+2\pi} &= Q_{2\pi} \circ V_{\theta_0+2\pi} \circ Q_{2\pi}^{-1} \circ Q_{2\pi} \\ &= \operatorname{Ad} Q_{2\pi} \circ V_{\theta_0+2\pi}\end{aligned}$$

so that (19.33) can be rewritten as

$$\frac{\partial F}{\partial \theta_0} = q_0 \odot (\operatorname{Ad} Q_{2\pi} \odot V_{\theta_0+2\pi} - V_{\theta_0}) \odot Q_{2\pi} \tag{19.34}$$

Thanks to Lemma 19.9 we can write

$$Q_{2\pi} = \mathrm{Id} - \pi\nu^2 f_0 + O(\nu^3) \tag{19.35}$$

that implies the following asymptotics for the action of its adjoint by (6.31)

$$\operatorname{Ad} Q_{2\pi} = \mathrm{Id} - \pi\nu^2 \operatorname{ad} f_0 + O(\nu^3)$$

We are left to compute the asymptotic expansion of (19.34). To this goal, recall that $r = r(\tau)$ satisfies

$$\dot{r} = \frac{r^3}{1-rb} a = r^3 a + O(r^4)$$

hence we can compute its term of order 3 with respect to $\nu$

$$r(t) = \nu + \nu^3 \int_{\theta_0}^{t} a(\tau) d\tau + O(\nu^4) \tag{19.36}$$

This in particular implies that $r(\theta_0 + 2\pi) = \nu + O(\nu^4)$ since $\int_{\theta_0}^{\theta_0+2\pi} a(t)dt = 0$ (this follows from the fact that $\int_{\theta_0}^{\theta_0+2\pi} \cos^2\theta d\theta = \int_{\theta_0}^{\theta_0+2\pi} \sin^2\theta d\theta$ and the fact that $a$ has zero trace).

This allows us to replace $r(\cdot)$ with $\nu$ in the term $V_{\theta_0+2\pi}$ since $r(\theta + 2\pi) = \nu + O(\nu^4)$. Moreover using that $b(\theta_0 + 2\pi) = b(\theta_0)$ and $f_{\theta_0+2\pi} = f_{\theta_0}$ we get

$$\begin{aligned} \operatorname{Ad} Q_{2\pi} \odot V_{\theta_0+2\pi} - V_{\theta_0} &= (\mathrm{Id} - \pi\nu^2 \operatorname{ad} f_0 + O(\nu^3)) \left(\frac{\nu}{1-\nu b} f_{\theta_0}\right) - \left(\frac{\nu}{1-\nu b} f_{\theta_0}\right) + O(\nu^4) \\ &= -\pi\nu^2 \operatorname{ad} f_0(\nu f_{\theta_0}) + O(\nu^4), \end{aligned} \tag{19.37}$$

and finally plugging (19.35) and (19.37) into (19.34) one obtains

$$\begin{aligned} \frac{\partial F}{\partial \theta} &= q_0 \odot \left(-\pi\nu^2 \operatorname{ad} f_0(\nu f_{\theta_0}) + O(\nu^4)\right) \odot (\mathrm{Id} + O(\nu)) \\ &= q_0 \odot (-\pi\nu^3 [f_0, f_{\theta_0}] + O(\nu^4)). \end{aligned}$$

□

### 19.3.2 Asymptotics of the conjugate locus

In this section we finally prove Theorem 19.6, by computing the expansion of the conjugate time $t_c(\theta_0, \nu)$. We know from Proposition 19.2 that

$$\tau_c(\theta_0, \nu) = 2\pi + \nu^2 s(\theta_0) + O(\nu^3).$$

Since $\tau_c(\theta_0, \nu)$ is a conjugate point, the function $s = s(\theta_0)$ is characterized as the solution of the equation

$$\left.\frac{\partial F}{\partial s} \wedge \frac{\partial F}{\partial \theta} \wedge \frac{\partial F}{\partial \nu}\right|_{(2\pi+\nu^2 s, \theta, \nu)} = 0, \tag{19.38}$$

where $s$ is considered as a parameter. Notice that the derivative with respect to $s$ is computed by

$$\frac{\partial F}{\partial s} = \frac{\partial F}{\partial t}\frac{\partial t}{\partial s} = (\nu f_\theta + O(\nu^2))\nu^2 = \nu^3 f_\theta + O(\nu^4).$$

Moreover, from the expansion of $F$ with respect to $\nu$ one has

$$\frac{\partial F}{\partial \nu} = -2\pi\nu f_0 + O(\nu^2).$$

Thus

$$F(2\pi + \nu^2 s; \theta, \nu) = F(2\pi, \theta, \nu) + \nu^3 s f_\theta + O(\nu^4),$$

and differentiation with respect to $\theta_0$ together with Lemma 19.12 gives

$$\frac{\partial F}{\partial \theta}(2\pi + \nu^2 s; \theta, \nu) = \nu^3(\pi[f_\theta, f_0] + s f_{\theta'}) + O(\nu^4),$$

where as usual $f_{\theta'}$ denotes the derivative with respect to $\theta$.

Then, collecting together all these computations, the equation for conjugate points (19.38) can be rewritten as

$$f_\theta \wedge (s f_{\theta'} + \pi[f_\theta, f_0]) \wedge f_0 = O(\nu). \tag{19.39}$$

Since $f_\theta, f_{\theta'}$ are an orthonormal frame on $\mathcal{D}$ and $f_0$ is transversal to the distribution, (19.39) is equivalent to

$$f_\theta \wedge (s f_{\theta'} + \pi[f_\theta, f_0]) = O(\nu),$$

that implies

$$s(\theta) = \pi \left\langle [f_0, f_\theta] \,|\, f_{\theta'} \right\rangle + O(\nu),$$

where $\langle \cdot \,|\, \cdot \rangle$ denotes the the scalar product on the distribution. Hence

$$t_c(\theta, \nu) = 2\pi + \pi\nu^2 \left\langle [f_0, f_\theta] \,|\, f_{\theta'} \right\rangle_{q_0} + O(\nu^3).$$

To find the expression of conjugate locus, we evaluate the exponential map at time $t_c(\theta, \nu)$.

We first consider the asymptotic of the conjugate locus. Using again that the first order term with respect to $\nu$ of $\partial_t F$ is $\nu f_\theta$ we have

$$F(2\pi + \nu^2 s(\theta_0), \theta_0, \nu) = F(2\pi; \theta_0, \nu) + \nu^3 s(\theta_0) f_{\theta_0} + O(\nu^4).$$

Hence, by Corollary 19.11 and Lemma 19.9 one gets

$$\mathrm{Con}_{q_0}(\theta_0, \nu) = q_0 - \pi\nu^2 f_0(q_0) - \frac{\nu^3}{2}\int_{\theta_0}^{\theta_0+\pi} \frac{dq_3}{d\tau} d\tau + \nu^3 s(\theta_0) f_{\theta_0} + O(\nu^4).$$

Moreover, since

$$\frac{\partial F}{\partial \theta_0}(2\pi, \nu, \theta_0) = \nu^3 [f_{\theta_0}, f_0] + O(\nu^4),$$

we have by definition that $q_3(\theta) = [f_\theta, f_0]$ and

$$\begin{aligned} \mathrm{Con}_{q_0}(\theta_0, \nu) &= q_0 - \nu^2 f_0(q_0) - \frac{\nu^3}{2}\int_{\theta_0}^{\theta_0+\pi} \pi[f_{\theta_0}, f_0] d\tau + \nu^3 s(\theta_0) f_{\theta_0} \\ &= q_0 - \nu^2 f_0(q_0) - \frac{\nu^3}{2}\int_{\theta_0}^{\theta_0+\pi} \pi[f_{\theta_0}, f_0] + s'(t) f_{\theta_0} + s(t) f_{\theta_0'} dt, \end{aligned} \tag{19.40}$$

where the last identify follows by writing $f_{\theta''} = -f_\theta$ and integrating by parts. Using that

$$\begin{aligned} s(\theta) &= \pi \langle [f_0, f_\theta] \,|\, f_{\theta'} \rangle \\ s'(\theta) &= \pi \langle [f_0, f_{\theta'}] \,|\, f_{\theta'} \rangle - \pi \langle [f_0, f_\theta] \,|\, f_\theta \rangle = 2\pi a, \end{aligned}$$

we can rewrite (19.40) as follows

$$\begin{aligned} \pi[f_{\theta_0}, f_0] + s'(t) f_{\theta_0} + s(t) f_{\theta_0'} &= \pi[f_{\theta_0}, f_0] + 2\pi a f_{\theta_0} + \pi \left\langle [f_0, f_{\theta_0}] \,\middle|\, f_{\theta_0'} \right\rangle f_{\theta_0'} \\ &= \pi \langle [f_{\theta_0}, f_0] \,|\, f_{\theta_0} \rangle f_{\theta_0} + 2\pi a f_{\theta_0} \\ &= 3\pi a f_{\theta_0}. \end{aligned}$$

Finally

$$\begin{aligned} \mathrm{Con}_{q_0}(\theta_0, \nu) &= q_0 - \pi\nu^2 f_0(q_0) - \frac{3\nu^3}{2}\pi \int_{\theta_0}^{\theta_0+\pi} a(\tau) f_\tau d\tau + O(\nu^4) \\ &= q_0 - \pi\nu^2 f_0(q_0) + \nu^3 \pi (a' f_{\theta_0} - a f_{\theta_0'}) + O(\nu^4). \end{aligned}$$

### 19.3.3 Asymptotics of the conjugate length

Similarly, we consider conjugate length. Recall that

$$\ell_c(\theta_0, \nu) = \int_{\theta_0}^{\theta_0 + t_c(\theta_0,\nu)} \frac{r(t)}{1 - r(t) Q_t^{\theta_0,\nu} b(t)} dt,$$

where we replaced $b(t)$ by its value along the flow $Q_t^{\theta_0,\nu} b(t)$.

As a first step, notice that we can reduce to an integral over a period, up to higher order terms with respect to $\nu$. Namely

$$\ell_c(\theta_0, \nu) = \int_{\theta_0}^{\theta_0 + 2\pi} \frac{r(t)}{1 - r(t) Q_t^{\theta_0,\nu} b(t)} dt + \nu^3 s(\theta_0) + O(\nu^4). \tag{19.41}$$

Indeed $t_c(\theta_0, \nu) = 2\pi + \nu^2 s(\theta) + O(\nu^3)$ and the first order term w.r.t. $\nu$ in the integrand is exactly $\nu$ by (19.36). In what follows we use again the notation $Q_t := Q_t^{\theta_0,\nu}$, and we compute the expansion in $\nu$ of the integral appearing in (19.41).

First notice that

$$\frac{r(t)}{1 - r(t) Q_t b(t)} = r(t) \left(1 + r(t) Q_t b(t) + r^2(t) [Q_t b(t) \odot Q_t b(t)] + O(r(t)^3)\right).$$

Using that $r(t) = \nu + O(\nu^3)$ and $Q_t b(t) = b(t) + O(\nu)$ we have that

$$\frac{r(t)}{1 - r(t) Q_t b(t)} = r(t) + r^2(t) Q_t b(t) + r^3(t) b(t)^2 + O(\nu^4).$$

Now each addend of the sum expands as follows

$$r(t) = \nu + \nu^3 \int_0^t a(t)dt + O(\nu^4) \tag{19.42}$$

$$r^2(t)Q_t(\nu)b(t) = (\nu^2 + O(\nu^4))\left(\mathrm{Id} + \nu \int_0^t f_\tau d\tau + O(\nu)\right) b(t) \tag{19.43}$$

$$= \nu^2 b(t) + \nu^3 \int_0^t f_\tau d\tau b(t) + O(\nu^4) \tag{19.44}$$

$$r^3(t)b(t)^2 = \nu^3 b(t)^2 + O(\nu^4). \tag{19.45}$$

Integrating the sum over the interval $[\theta_0, \theta_0 + 2\pi]$ and considering terms only up to $O(\nu^4)$ we have

$$\ell_c(\theta_0, \nu) = 2\pi\nu + \left(\int_{\theta_0}^{\theta_0+2\pi} \left[\int_0^t a(\tau)d\tau + \int_0^t f_\tau d\tau\right] b(t) + b^2(t)dt\right)\nu^3 + O(\nu^4),$$

where the coefficient in $\nu^2$ vanishes since $\int_{\theta_0}^{\theta_0+2\pi} b(\tau)d\tau = 0$. A straightforward computation of the integrals ends the proof of the theorem. □

### 19.3.4 Stability of the conjugate locus

In this section we want to prove that the third order Taylor polynomial of the exponential map corresponds to a stable map in the sense of singularity theory. More precisely it can be treated as a one parameter family of maps between 2-dimensional manifolds that has only singular points of "cusp" and "fold" type. As a consequence the original exponential map can be treated as a perturbation of the (truncated) stable one.

The classic Whitney theorem on the stability of maps between 2-dimensional manifolds then implies that the structure of their singularity will be the same, and actually the singular set of the perturbed one is the image under an homeomorphism of the singular set of the truncated map.

Fix some local coordinates $(x_0, x_1, x_2)$ around the point $q_0$ such that

$$q_0 = (0,0,0), \qquad f_i(q_0) = \partial_{x_i}, \quad \forall\, i = 0,1,2.$$

**Lemma 19.13.** *In these coordinates we have*

$$\begin{aligned}\frac{1}{\pi}F(2\pi + \pi\eta^2\tau, \theta, \nu) &= (x_0(\tau,\theta,\nu), x_1(\tau,\theta,\nu), x_2(\tau,\theta,\nu)) \\ &= (-\nu^2, (\tau - c_{02}^1)\cos(\theta)\nu^3, (\tau + c_{01}^2)\sin(\theta)\nu^3) + O(\nu^4).\end{aligned} \tag{19.46}$$

Let us define the new variable $\zeta = \sqrt{-x_0(\tau,\theta,\nu)} = \sqrt{\nu^2 + O(\nu^4)} = \nu + O(\nu^3)$ and apply the smooth change of variables $(\tau,\theta,\nu) \mapsto (\tau,\theta,\zeta)$. The map (19.46) is rewritten as follows

$$\frac{1}{\pi}F(2\pi + \pi\eta^2\tau, \theta, \nu) = (-\zeta^2, (\tau - c_{02}^1)\cos(\theta)\zeta^3 + O(\zeta^4), (\tau + c_{01}^2)\sin(\theta)\zeta^3 + O(\zeta^4)). \tag{19.47}$$

Notice that the first coordinate function of this map is constant in the new variables, when $\zeta$ is constant. The map (19.47) can be interpreted as a family of maps, parametrized by $\zeta$, depending on two variables

$$\frac{1}{\pi}F(2\pi + \pi\eta^2\tau, \theta, \nu) = (-\zeta^2, \zeta^3\Phi_\zeta(\tau,\theta)), \tag{19.48}$$

where we have defined

$$\Phi_\zeta(\tau,\theta) = ((\tau - c^1_{02})\cos(\theta), (\tau + c^2_{01})\sin(\theta)) + O(\zeta). \tag{19.49}$$

The critical set of the map $\Phi_0(\tau,\theta)$ is a smooth closed curve in $\mathbb{R}\times S^1$ defined by the equation

$$\tau = c^1_{02}\sin^2(\theta) - c^2_{01}\cos^2(\theta). \tag{19.50}$$

The critical values of this map, that is the image under the map $\Phi_0$ of the set defined by (19.50), is the astroid

$$\mathcal{A}_0 = \{2\chi(-\sin^3(\theta),\cos^3(\theta)), \theta \in S^1\}. \tag{19.51}$$

The restriction to $\Phi_0$ to the set $\mathcal{A}_0$ is a one-to-one map. Moreover every critical point of $\Phi_0$ is a fold or a cusp. This implies that $\Phi_0$ is a Whitney map. Hence it is stable, in the sense of Thom-Mather theory, see [Whi55, GG73].

In other words, for any compact $K \subset \mathbb{R}\times S^1$ big enough, there exists $\varepsilon > 0$ such that for all $\zeta \in ]0,\varepsilon[$, the map $\Phi_\zeta|_K$ is equivalent to $\Phi_0|_K$, under a smooth family of change of coordinates in the source and in the image. Moreover, this family can be chosen to be smooth with respect to the parameter $\zeta$.

Collecting these results, we have proved that the shape of the conjugate locus described in Figure 19.1 obtained via third order approximation of the end-point map is indeed a picture of the *true* shape.

**Theorem 19.14.** *Suppose $M$ is a 3D contact sub-Riemannian structure and $\chi(q_0) \neq 0$. Then there exists $\varepsilon > 0$ such that for every closed ball $B = B(q_0, r)$ with $r \leq \varepsilon$ there exists an open set $U \subset B \setminus \{q_0\}$ and a diffeomorphism $\Psi : U \to \mathbb{R}^3 \times \{\pm 1\}$ such that $B \cap \mathrm{Con}_{q_0} \subset U$ and*

$$\Psi(B \cap \mathrm{Con}_{q_0}) = \{(\zeta^2, \cos^3(\theta)\zeta^3, -\sin^3(\theta)\zeta^3) : \zeta > 0, \theta \in S^1\} \times \{\pm 1\}.$$

*In particular, each of the two connected components of $B \cap Con_{q_0}$ contains 4 cuspidal edges.*

A similar statement concerning the stability of the cut locus can be found in [Agr96].

## 19.4 Bibliographical note

The asymptotics of the exponential map for the 3D sub-Riemannian contact structures and the corresponding study of cut and conjugate locus have been first studied, independently, in [Agr96] and [EAGK96]. Other relevant references on the same topic are also [ACGZ00, ACEAG98, AG99].

The material contained in this chapter is based essentialy on the paper [Agr96].

# Chapter 20

# The volume in sub-Riemannian geometry

In this chapter we investigate the notion of instrinsic volume in sub-Riemannian geometry in the case of *equiregular* structures.

In particular we consider two constructions: the first one is the Popp volume that is a smooth volume which is canonically associated with the sub-Riemannian structure, and it is a natural generalization of the Riemannian one. The second one is the Hausdorff volume, a notion that is well defined on every metric space.

On a Riemannian manifold these two notions coincide. The goal of this chapter is to introduce these two volumes, prove their basic properties and show that in general these two volumes may be different in sub-Riemannian geometry.

## 20.1 Equiregular sub-Riemannian manifolds

We denote by $\mathcal{D}_q \subset T_qM$ the fiber of $\mathcal{D}$ over $q$. Recall that, for every $i \geq 1$, define recursively the submodules $\mathcal{D}^i$ of $\mathrm{Vec}(M)$ by

$$\mathcal{D}^1 = \mathcal{D}, \quad \mathcal{D}^{i+1} = \mathcal{D}^i + [\mathcal{D}, \mathcal{D}^i].$$

and we set $\mathcal{D}^i_q = \{X(q) \mid X \in \mathcal{D}^i\}$. The bracket-generating assumption implies that for every $q \in M$ there exists an integer $k(q)$, the *non-holonomy degree at* $q$, such that

$$\{0\} \subset \mathcal{D}^1_q \subset \cdots \subset \mathcal{D}^{k(q)}_q = T_qM. \tag{20.1}$$

The sequence of subspaces (20.1) is called the *flag of* $\mathcal{D}$ *at* $q$. Set $d_i(q) = \dim \mathcal{D}^i_q$ for $i \geq 0$ with the understaing that $d_0(q) = 0$. We say that $M$ is *equiregular* if $d_i(q)$ is constant as $q$ varies in $M$, for every $i \geq 1$.

For an equiregular structure the modulus $\mathcal{D}$ can be identified with the sections satisfying

$$\{X \in \mathrm{Vec}(M) \mid X(q) \in \mathcal{D}_q, \forall\, q \in M\}.$$

and we define the homogeneous dimension

$$Q := \sum_{i=1}^{k} i(d_i - d_{i-1}), \tag{20.2}$$

*Remark* 20.1. Notice that if we consider the dilations $\delta_\alpha(x_1, \dots, x_r)$ we have $\det \delta_{\alpha *} = \alpha^Q$.

## 20.2 The Popp volume

We introduce the nilpotentization of the distribution at the point $q$, which is fundamental for the definition of Popp's volume.

**Definition 20.2.** Let $\mathcal{D}$ be an equiregular distribution of step $m$. The *nilpotentization of* $\mathcal{D}$ at the point $q \in M$ is the graded vector space

$$\mathrm{gr}_q(\mathcal{D}) = \mathcal{D}_q \oplus \mathcal{D}_q^2/\mathcal{D}_q \oplus \ldots \oplus \mathcal{D}_q^m/\mathcal{D}_q^{m-1}.$$

The vector space $\mathrm{gr}_q(\mathcal{D})$ can be endowed with a Lie algebra structure, which respects the grading. Then, there is a unique connected, simply connected group, $\mathrm{Gr}_q(\mathcal{D})$, such that its Lie algebra is $\mathrm{gr}_q(\mathcal{D})$. The global, left-invariant vector fields obtained by the group action on any orthonormal basis of $\mathcal{D}_q \subset \mathrm{gr}_q(\mathcal{D})$ define a sub-Riemannian structure on $\mathrm{Gr}_q(\mathcal{D})$, which is called the *nilpotent approximation* of the sub-Riemannian structure at the point $q$.

In what follows, we provide the definition of Popp's volume. Our presentation follows closely the one that can be found in [BR13]. (See also [Mon02]). The definition rests on the following lemmas.

**Lemma 20.3.** *Let $E$ be an inner product space and $V$ be a vector space. Let $\pi : E \to V$ be a surjective linear map. Then $\pi$ induces an inner product on $V$ such that the norm of $v \in V$ is*

$$\|v\|_V = \min\{\|e\|_E \ s.t.\ \pi(e) = v\}\,. \tag{20.3}$$

*Proof.* It is easy to check that Eq. (20.3) defines a norm on $V$. Moreover, since $\|\cdot\|_E$ is induced by an inner product, i.e., it satisfies the parallelogram identity, it follows that $\|\cdot\|_V$ satisfies the parallelogram identity too. Notice that this is equivalent to consider the inner product on $V$ defined by the linear isomorphism $\pi : (\ker \pi)^\perp \to V$. Indeed the norm of $v \in V$ is the norm of the shortest element $e \in \pi^{-1}(v)$. $\square$

**Lemma 20.4.** *Let $E$ be a vector space of dimension $n$ with a flag of linear subspaces $\{0\} = F^0 \subset F^1 \subset F^2 \subset \ldots \subset F^m = E$. Let $\mathrm{gr}(F) = F^1 \oplus F^2/F^1 \oplus \ldots \oplus F^m/F^{m-1}$ be the associated graded vector space. Then there is a canonical isomorphism $\theta : \wedge^n E \to \wedge^n \mathrm{gr}(F)$.*

*Proof.* We only give a sketch of the proof. For $0 \leq i \leq m$, let $k_i := \dim F^i$. Let $X_1, \dots, X_n$ be a adapted basis for $E$, i.e., $X_1, \dots, X_{k_i}$ is a basis for $F^i$. We define the linear map $\widehat{\theta} : E \to \mathrm{gr}(F)$ which, for $0 \leq j \leq m-1$, takes $X_{k_j+1}, \dots, X_{k_{j+1}}$ to the corresponding equivalence class in $F^{j+1}/F^j$. This map is indeed a non-canonical isomorphism, which depends on the choice of the adapted basis. In turn, $\widehat{\theta}$ induces a map $\theta : \wedge^n E \to \wedge^n \mathrm{gr}(F)$, which sends $X_1 \wedge \ldots \wedge X_n$ to $\widehat{\theta}(X_1) \wedge \ldots \wedge \widehat{\theta}(X_n)$. It is a standard check that $\theta$ does not depend on the choice of the adapted basis. $\square$

The idea behind Popp's volume is to define an inner product on each $\mathcal{D}_q^i/\mathcal{D}_q^{i-1}$ which, in turn, induces an inner product on the orthogonal direct sum $\mathrm{gr}_q(\mathcal{D})$. The latter has a natural volume form, which is the canonical volume of an inner product space obtained by wedging the elements an orthonormal dual basis. Then, we employ Lemma 20.4 to define an element of $(\wedge^n T_q M)^* \simeq \wedge^n T_q^* M$, which is Popp's volume form computed at $q$.

Fix $q \in M$. Then, let $v, w \in \mathcal{D}_q$, and let $V, W$ be any horizontal extensions of $v, w$. Namely, $V, W \in \Gamma(\mathcal{D})$ and $V(q) = v$, $W(q) = w$. The linear map $\pi : \mathcal{D}_q \otimes \mathcal{D}_q \to \mathcal{D}_q^2/\mathcal{D}_q$

$$\pi(v \otimes w) := [V, W]_q \mod \mathcal{D}_q \,, \tag{20.4}$$

is well defined, and does not depend on the choice the horizontal extensions. Indeed let $\widetilde{V}$ and $\widetilde{W}$ be two different horizontal extensions of $v$ and $w$ respectively. Then, in terms of a local frame $X_1, \ldots, X_k$ of $\mathcal{D}$

$$\widetilde{V} = V + \sum_{i=1}^{k} f_i X_i \,, \qquad \widetilde{W} = W + \sum_{i=1}^{k} g_i X_i \,, \tag{20.5}$$

where, for $1 \leq i \leq k$, $f_i, g_i \in C^\infty(M)$ and $f_i(q) = g_i(q) = 0$. Therefore

$$[\widetilde{V}, \widetilde{W}] = [V, W] + \sum_{i=1}^{k} (V(g_i) - W(f_i)) X_i + \sum_{i,j=1}^{k} f_i g_j [X_i, X_j] \,. \tag{20.6}$$

Thus, evaluating at $q$, $[\widetilde{V}, \widetilde{W}]_q = [V, W]_q \mod \mathcal{D}_q$, as claimed. Similarly, let $1 \leq i \leq m$. The linear maps $\pi_i : \otimes^i \mathcal{D}_q \to \mathcal{D}_q^i/\mathcal{D}_q^{i-1}$

$$\pi_i(v_1 \otimes \cdots \otimes v_i) = [V_1, [V_2, \ldots, [V_{i-1}, V_i]]]_q \mod \mathcal{D}_q^{i-1} \,, \tag{20.7}$$

are well defined and do not depend on the choice of the horizontal extensions $V_1, \ldots, V_i$ of $v_1, \ldots, v_i$.

By the bracket-generating condition, $\pi_i$ are surjective and, by Lemma 20.3, they induce an inner product space structure on $\mathcal{D}_q^i/\mathcal{D}_q^{i-1}$. Therefore, the nilpotentization of the distribution at $q$, namely

$$\mathrm{gr}_q(\mathcal{D}) = \mathcal{D}_q \oplus \mathcal{D}_q^2/\mathcal{D}_q \oplus \ldots \oplus \mathcal{D}_q^m/\mathcal{D}_q^{m-1} \,, \tag{20.8}$$

is an inner product space, as the orthogonal direct sum of a finite number of inner product spaces. As such, it is endowed with a canonical volume (defined up to a sign) $\omega_q \in \wedge^n \mathrm{gr}_q(\mathcal{D})^*$, which is the volume form obtained by wedging the elements of an orthonormal dual basis.

Finally, Popp's volume (computed at the point $q$) is obtained by transporting the volume of $\mathrm{gr}_q(\mathcal{D})$ to $T_qM$ through the map $\theta_q : \wedge^n T_qM \to \wedge^n \mathrm{gr}_q(\mathcal{D})$ defined in Lemma 20.4. Namely

$$\mathcal{P}_q = \theta_q^*(\omega_q) = \omega_q \circ \theta_q \,, \tag{20.9}$$

where $\theta_q^*$ denotes the dual map and we employ the canonical identification $(\wedge^n T_qM)^* \simeq \wedge^n T_q^*M$. Eq. (20.9) is defined only in the domain of the chosen local frame. Since $M$ is orientable, with a standard argument, these $n$-forms can be glued together to obtain Popp's volume $\mathcal{P} \in \Omega^n(M)$. The smoothness of $\mathcal{P}$ follows directly from Theorem 20.6.

*Remark* 20.5. The definition of Popp's volume can be restated as follows. Let $(M, \mathcal{D})$ be an oriented sub-Riemannian manifold. Popp's volume is the unique volume $\mathcal{P}$ such that, for all $q \in M$, the following diagram is commutative:

$$\begin{array}{ccc}
(M, \mathcal{D}) & \xrightarrow{\mathcal{P}} & (\wedge^n T_qM)^* \\
{\scriptstyle \mathrm{gr}_q}\downarrow & & \downarrow{\scriptstyle \theta_q^*} \\
\mathrm{gr}_q(\mathcal{D}) & \xrightarrow[\omega]{} & (\wedge^n \mathrm{gr}_q(\mathcal{D}))^*
\end{array}$$

where $\omega$ associates the inner product space $\mathrm{gr}_q(\mathcal{D})$ with its canonical volume $\omega_q$, and $\theta_q^*$ is the dual of the map defined in Lemma 20.4.

## 20.3 A formula for Popp volume in terms of adapted frames

In this section we prove an explicit formula for the Popp volume.

We say that a local frame $X_1, \dots, X_n$ is adapted if $X_1, \dots, X_{k_i}$ is a local frame for $\mathcal{D}^i$, where $k_i := \dim \mathcal{D}^i$, and $X_1, \dots, X_k$ are orthonormal. It is useful to define the functions $c_{ij}^l \in C^\infty(M)$ by

$$[X_i, X_j] = \sum_{l=1}^{n} c_{ij}^l X_l \,. \tag{20.10}$$

With a standard abuse of notation we call them *structure constants*. For $j = 2, \dots, m$ we define the *adapted structure constants* $b_{i_1 \dots i_j}^l \in C^\infty(M)$ as follows:

$$[X_{i_1}, [X_{i_2}, \dots, [X_{i_{j-1}}, X_{i_j}]]] = \sum_{l=k_{j-1}+1}^{k_j} b_{i_1 i_2 \dots i_j}^l X_l \mod \mathcal{D}^{j-1} \,, \tag{20.11}$$

where $1 \leq i_1, \dots, i_j \leq k$. These are a generalization of the $c_{ij}^l$, with an important difference: the structure constants of Eq. (20.10) are obtained by considering the Lie bracket of all the fields of the local frame, namely $1 \leq i, j, l \leq n$. On the other hand, the adapted structure constants of Eq. (20.11) are obtained by taking the iterated Lie brackets of the first $k$ elements of the adapted frame only (i.e., the local orthonormal frame for $\mathcal{D}$), and considering the appropriate equivalence class. For $j = 2$, the adapted structure constants can be directly compared to the standard ones. Namely $b_{ij}^l = c_{ij}^l$ when both are defined, that is for $1 \leq i, j \leq k$, $l \geq k+1$.

Then, we define the $k_j - k_{j-1}$ dimensional square matrix $B_j$ as follows:

$$[B_j]^{hl} = \sum_{i_1, i_2, \dots, i_j = 1}^{k} b_{i_1 i_2 \dots i_j}^h b_{i_1 i_2 \dots i_j}^l \,, \qquad j = 1, \dots, m \,, \tag{20.12}$$

with the understanding that $B_1$ is the $k \times k$ identity matrix. It turns out that each $B_j$ is positive definite.

**Theorem 20.6.** *Let $X_1, \dots, X_n$ be a local adapted frame, and let $\nu^1, \dots, \nu^n$ be the dual frame. Then Popp's volume $\mathcal{P}$ satisfies*

$$\mathcal{P} = \frac{1}{\sqrt{\prod_j \det B_j}} \nu^1 \wedge \dots \wedge \nu^n \,, \tag{20.13}$$

*where $B_j$ is defined by* (20.12) *in terms of the adapted structure constants* (20.11).

To clarify the geometric meaning of Eq. (20.13), let us consider more closely the case $m = 2$. If $\mathcal{D}$ is a step 2 distribution, we can build a local adapted frame $\{X_1, \dots, X_k, X_{k+1}, \dots, X_n\}$ by completing any local orthonormal frame $\{X_1, \dots, X_k\}$ of the distribution to a local frame of the whole tangent bundle. Even though it may not be evident, it turns out that $B_2^{-1}(q)$ is the Gram matrix of the vectors $X_{k+1}, \dots, X_n$, seen as elements of $T_qM/\mathcal{D}_q$. The latter has a natural structure of inner product space, induced by the surjective linear map $[\,,\,] : \mathcal{D}_q \otimes \mathcal{D}_q \to T_qM/\mathcal{D}_q$ (see Lemma 20.3). Therefore, the function appearing at the beginning of Eq. (20.13) is the volume of the parallelotope whose edges are $X_1, \dots, X_n$, seen as elements of the orthogonal direct sum $\mathrm{gr}_q(\mathcal{D}) = \mathcal{D}_q \oplus T_qM/\mathcal{D}_q$.

## Proof of Theorem 20.6

We are now ready to prove Theorem 20.6. For convenience, we first prove it for a distribution of step $m = 2$. Then, we discuss the general case. In the following subsections, everything is understood to be computed at a fixed point $q \in M$. Namely, by $\mathrm{gr}(\mathcal{D})$ we mean the nilpotentization of $\mathcal{D}$ at the point $q$, and by $\mathcal{D}^i$ we mean the fibre $\mathcal{D}^i_q$ of the appropriate higher order distribution.

### Step 2 distribution

If $\mathcal{D}$ is a step 2 distribution, then $\mathcal{D}^2 = TM$. The growth vector is $\mathcal{G} = (k, n)$. We choose $n - k$ independent vector fields $\{Y_l\}_{l=k+1}^n$ such that $X_1, \dots, X_k, Y_{k+1}, \dots, Y_n$ is a local adapted frame for $TM$. Then

$$[X_i, X_j] = \sum_{l=k+1}^{n} b_{ij}^l Y_l \mod \mathcal{D}. \tag{20.14}$$

For each $l = k+1, \dots, n$, we can think to $b_{ij}^l$ as the components of an Euclidean vector in $\mathbb{R}^{k^2}$, which we denote by the symbol $b^l$. According to the general construction of Popp's volume, we need first to compute the inner product on the orthogonal direct sum $\mathrm{gr}(\mathcal{D}) = \mathcal{D} \oplus \mathcal{D}^2/\mathcal{D}$. By Lemma 20.3, the norm on $\mathcal{D}^2/\mathcal{D}$ is induced by the linear map $\pi : \otimes^2\mathcal{D} \to \mathcal{D}^2/\mathcal{D}$

$$\pi(X_i \otimes X_j) = [X_i, X_j] \mod \mathcal{D}. \tag{20.15}$$

The vector space $\otimes^2\mathcal{D}$ inherits an inner product from the one on $\mathcal{D}$, namely $\forall X, Y, Z, W \in \mathcal{D}$, $\langle X \otimes Y, Z \otimes W \rangle = \langle X, Z \rangle \langle Y, W \rangle$. $\pi$ is surjective, then we identify the range $\mathcal{D}^2/\mathcal{D}$ with $\ker \pi^\perp \subset \otimes^2\mathcal{D}$, and define an inner product on $\mathcal{D}^2/\mathcal{D}$ by this identification. In order to compute explicitly the norm on $\mathcal{D}^2/\mathcal{D}$ (and then, by polarization, the inner product), let $Y \in \mathcal{D}^2/\mathcal{D}$. Then

$$\|\mathcal{D}^2/\mathcal{D}\|_Y = \min\{\|Z\|_{\otimes^2\mathcal{D}} \text{ s.t. } \pi(Z) = Y\}. \tag{20.16}$$

Let $Y = \sum_{l=k+1}^n c^l Y_l$ and $Z = \sum_{i,j=1}^k a_{ij} X_i \otimes X_j \in \otimes^2\mathcal{D}$. We can think to $a_{ij}$ as the components of a vector $a \in \mathbb{R}^{k^2}$. Then, Eq. (20.16) writes

$$\|Y\|_{\mathcal{D}^2/\mathcal{D}} = \min\{|a| \text{ s.t. } a \cdot b^l = c^l,\ l = k+1, \dots, n\}, \tag{20.17}$$

where $|a|$ is the Euclidean norm of $a$, and the dot denotes the Euclidean inner product. Indeed, $\|Y\|_{\mathcal{D}^2/\mathcal{D}}$ is the Euclidean distance of the origin from the affine subspace of $\mathbb{R}^{k^2}$ defined by the equations $a \cdot b^l = c^l$ for $l = k+1, \dots, n$. In order to find an explicit expression for $\|Y\|^2_{\mathcal{D}^2/\mathcal{D}}$ in terms of the $b^l$, we employ the Lagrange multipliers technique. Then, we look for extremals of

$$L(a, b^{k+1}, \dots, b^n, \lambda_{k+1}, \dots, \lambda_n) = |a|^2 - 2\sum_{l=k+1}^{n} \lambda_l (a \cdot b^l - c^l). \tag{20.18}$$

We obtain the following system

$$\begin{cases} \displaystyle\sum_{l=k+1}^{n} \lambda_l \cdot b^l - a = 0, \\ \displaystyle\sum_{l=k+1}^{n} \lambda_l b^l \cdot b^r = c^r, \qquad r = k+1, \dots, n. \end{cases} \tag{20.19}$$

Let us define the $n-k$ square matrix $B$, with components $B^{hl} = b^h \cdot b^l$. $B$ is a Gram matrix, which is positive definite iff the $b^l$ are $n-k$ linearly independent vectors. These vectors are exactly the rows of the representative matrix of the linear map $\pi : \otimes^2 \mathcal{D} \to \mathcal{D}^2/\mathcal{D}$, which has rank $n-k$. Therefore $B$ is symmetric and positive definite, hence invertible. It is now easy to write the solution of system (20.19) by employing the matrix $B^{-1}$, which has components $B^{-1}_{hl}$. Indeed a straightforward computation leads to

$$\| \sum_{s=k+1}^{n} c^s Y_s \|^2_{\mathcal{D}^2/\mathcal{D}} = \sum_{h,l=k+1}^{n} c^h B^{-1}_{hl} c^l \,. \tag{20.20}$$

By polarization, the inner product on $\mathcal{D}^2/\mathcal{D}$ is defined, in the basis $Y_l$, by

$$\langle Y_l, Y_h \rangle_{\mathcal{D}^2/\mathcal{D}} = B^{-1}_{lh} \,. \tag{20.21}$$

Observe that $B^{-1}$ is the Gram matrix of the vectors $Y_{k+1}, \ldots, Y_n$ seen as elements of $\mathcal{D}^2/\mathcal{D}$. Then, by the definition of Popp's volume, if $\nu^1, \ldots, \nu^k, \eta^{k+1}, \ldots, \eta^n$ is the dual basis associated with $X_1, \ldots, X_k, Y_{k+1}, \ldots, Y_n$, the following formula holds true

$$\mathcal{P} = \frac{1}{\sqrt{\det B}} \, \nu^1 \wedge \cdots \wedge \nu^k \wedge \eta^{k+1} \wedge \cdots \wedge \eta^n \,. \tag{20.22}$$

**General case**

In the general case, the procedure above can be carried out with no difficulty. Let $X_1, \ldots, X_n$ be a local adapted frame for the flag $\mathcal{D}^0 \subset \mathcal{D} \subset \mathcal{D}^2 \subset \cdots \subset \mathcal{D}^m$. As usual $k_i = \dim(\mathcal{D}^i)$. For $j = 2, \ldots, m$ we define the adapted structure constants $b^l_{i_1 \ldots i_j} \in C^\infty(M)$ by

$$[X_{i_1}, [X_{i_2}, \ldots, [X_{i_{j-1}}, X_{i_j}]]] = \sum_{l=k_{j-1}+1}^{k_j} b^l_{i_1 i_2 \ldots i_j} X_l \mod \mathcal{D}^{j-1} \,, \tag{20.23}$$

where $1 \leq i_1, \ldots, i_j \leq k$. Again, $b^l_{i_1 \ldots i_j}$ can be seen as the components of a vector $b^l \in \mathbb{R}^{k^j}$.

Recall that for each $j$ we defined the surjective linear map $\pi_j : \otimes^j \mathcal{D} \to \mathcal{D}^j/\mathcal{D}^{j-1}$

$$\pi_j(X_{i_1} \otimes X_{i_2} \otimes \cdots \otimes X_{i_j}) = [X_{i_1}, [X_{i_2}, \ldots, [X_{i_{j-1}}, X_{i_j}]]] \mod \mathcal{D}^{j-1} \,. \tag{20.24}$$

Then, we compute the norm of an element of $\mathcal{D}^j/\mathcal{D}^{j-1}$ exactly as in the previous case. It is convenient to define, for each $1 \leq j \leq m$, the $k_j - k_{j-1}$ dimensional square matrix $B_j$, of components

$$[B_j]^{hl} = \sum_{i_1, i_2, \ldots, i_j = 1}^{k} b^h_{i_1 i_2 \ldots i_j} b^l_{i_1 i_2 \ldots i_j} \,. \tag{20.25}$$

with the understanding that $B_1$ is the $k \times k$ identity matrix. Each one of these matrices is symmetric and positive definite, hence invertible, due to the surjectivity of $\pi_j$. The same computation of the previous case, applied to each $\mathcal{D}^j/\mathcal{D}^{j-1}$ shows that the matrices $B_j^{-1}$ are precisely the Gram matrices of the vectors $X_{k_{j-1}+1}, \ldots, X_{k_j} \in \mathcal{D}^j/\mathcal{D}^{j-1}$, in other words

$$\langle X_{k_{j-1}+l}, X_{k_{j-1}+h} \rangle_{\mathcal{D}^j/\mathcal{D}^{j-1}} = B^{-1}_{lh} \,. \tag{20.26}$$

Therefore, if $\nu^1, \ldots, \nu^n$ is the dual frame associated with $X_1, \ldots, X_n$, Popp's volume is

$$\mathcal{P} = \frac{1}{\sqrt{\prod_{j=1}^m \det B_j}} \, \nu^1 \wedge \ldots \wedge \nu^n \,. \tag{20.27}$$

## 20.4 Popp volume and smooth isometries

In the last part of the paper we discuss the conditions under which a local isometry preserves Popp's volume. In the Riemannian setting, an isometry is a diffeomorphism such that its differential is an isometry for the Riemannian metric. The concept is easily generalized to the sub-Riemannian case.

**Definition 20.7.** A diffeomorphism $\varphi : M \to M$ is a *isometry* if its differential $\varphi_* : TM \to TM$ preserves the sub-Riemannian structure $(\mathcal{D}, \langle \cdot \,|\, \cdot \rangle)$, namely

(i) $\varphi_*(\mathcal{D}_q) = \mathcal{D}_{\varphi(q)}$ for all $q \in M$,

(ii) $\langle \varphi_* X \,|\, \varphi_* Y \rangle_{\varphi(q)} = \langle X \,|\, Y \rangle_q$ for all $q \in M$, $X, Y \in \mathcal{D}_q$.

Similarly, one defines a local isometry when $\varphi$ is a local diffeomorhism in a neighborhood $O_q$ of a point $q \in M$ and conditions (i) and (ii) are satisfied on $O_q$.

*Remark* 20.8. Condition $(i)$, which is trivially satisfied in the Riemannian case, is necessary to define isometries in the sub-Riemannian case. Actually, it also implies that all the higher order distributions are preserved by $\varphi_*$, i.e., $\varphi_*(\mathcal{D}^i_q) = \mathcal{D}^i_{\varphi(q)}$, for $1 \leq i \leq m$.

**Definition 20.9.** Let $M$ be a manifold equipped with a volume form $\mu \in \Omega^n(M)$. We say that a (local) diffeomorphism $\varphi : M \to M$ is a *(local) volume preserving transformation* if $\varphi^* \mu = \mu$.

In the Riemannian case, local isometries are also volume preserving transformations for the Riemannian volume. Then, it is natural to ask whether this is true also in the sub-Riemannian setting, for some choice of the volume. The next proposition states that the answer is positive if we choose Popp's volume.

**Proposition 20.10.** *Sub-Riemannian (local) isometries are volume preserving transformations for Popp's volume.*

Proposition 20.10 may be false for volumes different than Popp's one. We have the following.

**Proposition 20.11.** *Let* $\mathrm{Iso}(M)$ *be the group of isometries of the sub-Riemannian manifold* $M$*. If* $\mathrm{Iso}(M)$ *acts transitively on* $M$*, then Popp's volume is the unique volume (up to multiplication by scalar constant) such that Proposition 20.10 holds true.*

Recall that when $M$ be a Lie group, a sub-Riemannian structure $(M, \mathcal{D}, \langle \cdot \,|\, \cdot \rangle)$ is said to be left invariant if $\forall g \in M$, the left action $L_g : M \to M$ is an isometry.

As a trivial consequence of Proposition 20.10 we have the following

**Corollary 20.12.** *Let* $(M, \mathcal{D}, \langle \cdot \,|\, \cdot \rangle)$ *be a left-invariant sub-Riemannian structure. Then Popp's volume is left invariant, i.e.,* $L_g^* \mathcal{P} = \mathcal{P}$ *for every* $g \in M$*.*

The rest of this section is devoted to the proof of Propositions 20.10 and 20.11.

## Proof of Proposition 20.10

Let $\varphi \in \mathrm{Iso}(M)$ be a (local) isometry, and $1 \leq i \leq m$. The differential $\varphi_*$ induces a linear map

$$\overline{\varphi}_* : \otimes^i \mathcal{D}_q \to \otimes^i \mathcal{D}_{\varphi(q)} . \tag{20.28}$$

Moreover $\varphi_*$ preserves the flag $\mathcal{D} \subset \ldots \subset \mathcal{D}^m$. Therefore, it induces a linear map

$$\widehat{\varphi}_* : \mathcal{D}^i_q/\mathcal{D}^{i-1}_q \to \mathcal{D}^i_{\varphi(q)}/\mathcal{D}^{i-1}_{\varphi(q)} . \tag{20.29}$$

The key to the proof of Proposition 20.10 is the following lemma.

**Lemma 20.13.** *$\overline{\varphi}_*$ and $\widehat{\varphi}_*$ are isometries of inner product spaces.*

*Proof.* The proof for $\overline{\varphi}_*$ is trivial. The proof for $\widehat{\varphi}_*$ is as follows. Remember that the inner product on $\mathcal{D}^i/\mathcal{D}^{i-1}$ is induced by the surjective maps $\pi_i : \otimes^i \mathcal{D} \to \mathcal{D}^i/\mathcal{D}^{i-1}$ defined by Eq. (20.7). Namely, let $Y \in \mathcal{D}^i_q/\mathcal{D}^{i-1}_q$. Then

$$\|Y\|_{\mathcal{D}^i_q/\mathcal{D}^{i-1}_q} = \min\{\|Z\|_{\otimes^i \mathcal{D}_q} \text{ s.t. } \pi_i(Z) = Y\} . \tag{20.30}$$

As a consequence of the properties of the Lie brackets, $\pi_i \circ \overline{\varphi}_* = \widehat{\varphi}_* \circ \pi_i$. Therefore

$$\|Y\|_{\mathcal{D}^i_q/\mathcal{D}^{i-1}_q} = \min\{\|\overline{\varphi}_* Z\|_{\otimes^i \mathcal{D}_{\varphi(q)}} \text{ s.t. } \pi_i(\overline{\varphi}_* Z) = \widehat{\varphi}_* Y\} = \|\widehat{\varphi}_* Y\|_{\mathcal{D}^i_{\varphi(q)}/\mathcal{D}^{i-1}_{\varphi(q)}} . \tag{20.31}$$

By polarization, $\widehat{\varphi}_*$ is an isometry. $\square$

Since $\mathrm{gr}_q(\mathcal{D}) = \oplus_{i=1}^m \mathcal{D}^i_q/\mathcal{D}^{i-1}_q$ is an orthogonal direct sum, $\widehat{\varphi}_* : \mathrm{gr}_q(\mathcal{D}) \to \mathrm{gr}_{\varphi(q)}(\mathcal{D})$ is also an isometry of inner product spaces.

Finally, Popp's volume is the canonical volume of $\mathrm{gr}_q(\mathcal{D})$ when the latter is identified with $T_qM$ through any choice of a local adapted frame. Since $\varphi_*$ is equal to $\widehat{\varphi}_*$ under such an identification, and the latter is an isometry of inner product spaces, the result follows. $\square$

## Proof of Proposition 20.11

Let $\mu$ be a volume form such that $\varphi^*\mu = \mu$ for any isometry $\varphi \in \mathrm{Iso}(M)$. There exists $f \in C^\infty(M)$, $f \neq 0$ such that $\mathcal{P} = f\mu$. It follows that, for any $\varphi \in \mathrm{Iso}(M)$

$$f\mu = \mathcal{P} = \varphi^*\mathcal{P} = (f \circ \varphi)\,\varphi^*\mu = (f \circ \varphi)\,\mu , \tag{20.32}$$

where we used the $\mathrm{Iso}(M)$-invariance of Popp's volume. Then also $f$ is $\mathrm{Iso}(M)$-invariant, namely $\varphi^* f = f$ for any $\varphi \in \mathrm{Iso}(M)$. By hypothesis, the action of $\mathrm{Iso}(M)$ is transitive, then $f$ is constant. $\square$

# 20.5 Hausdorff dimension and Hausdorff volume

Let $(M, d)$ be a metric space. We denote by $\operatorname{diam} S$ the diameter of a set $S \subset M$, by $B(q, r)$ the open ball $\{q \in M \mid d(p, q) < r\}$, and by $\overline{B}(q, r)$ the closure of $B(q, r)$. Let $\alpha \geq 0$ be a real number. For every set $A \subset M$, the *$\alpha$-dimensional Hausdorff measure* $\mathcal{H}^\alpha$ of $A$ is defined as $\mathcal{H}^\alpha(A) = \lim_{\delta \to 0^+} \mathcal{H}^\alpha_\delta(A)$, where

$$\mathcal{H}^\alpha_\delta(A) = \inf \left\{ \sum_{i=1}^{\infty} (\operatorname{diam} S_i)^\alpha \;:\; A \subset \bigcup_{i=1}^{\infty} S_i,\ S_i \text{ closed set},\ \operatorname{diam} S_i \leq \delta \right\}, \tag{20.33}$$

and the *$\alpha$-dimensional spherical Hausdorff measure* is defined as $\mathcal{S}^\alpha(A) = \lim_{\delta \to 0^+} \mathcal{S}^\alpha_\delta(A)$, where

$$\mathcal{S}^\alpha_\delta(A) = \inf \left\{ \sum_{i=1}^{\infty} (\operatorname{diam} S_i)^\alpha \;:\; A \subset \bigcup_{i=1}^{\infty} S_i,\ S_i \text{ ball},\ \operatorname{diam} S_i \leq \delta \right\}. \tag{20.34}$$

For every set $A \subset M$, the non-negative number

$$\dim_H(A) := \sup\{\alpha \geq 0 \mid \mathcal{H}^\alpha(A) = +\infty\} = \inf\{\alpha \geq 0 \mid \mathcal{H}^\alpha(A) = 0\}, \tag{20.35}$$

is well-defined and is called the *Hausdorff dimension* of $A$.

**Exercise 20.14.** (i) Prove the equality of the two definitions of Hausdorff dimension in (20.35).
(ii) Prove that for every subset $S \subset N$ we have the inequalities

$$\mathcal{H}^\alpha(S) \leq \mathcal{S}^\alpha(S) \leq 2^\alpha \mathcal{H}^\alpha(S). \tag{20.36}$$

(iii) Deduce that in (20.35) one can replace $\mathcal{H}^\alpha$ with $\mathcal{S}^\alpha$.

If $(M, d)$ is a metric space, the *Hausdorff volume* (resp. *spherical Hausdorff volume*) on $M$ is the $D$-dimensional Hausdorff measure $\mathcal{H}^D$ (resp. $\mathcal{S}^D$), where $D = \dim_H(M)$. Notice that given $A \subset M$, its Hausdorff volume (resp. spherical Hausdorff volume) may be 0, positive, or $+\infty$.

## 20.6 Hausdorff volume on sub-Riemannian manifolds

A sub-Riemannian manifold $(M, \mathbf{U}, f)$, can be seen as a metric space $(M, d)$ endowed with the sub-Riemannian distance. The following questions are then natural in view of the previous discussion:

(a) What is the Hausdorff dimension $D := \dim_H(M)$ of $(M, d)$?

(b) Is $\mathcal{H}^D(B(p, r))$ (or, equivalently, $\mathcal{S}^D(B(p, r))$) finite?

(c) What is the Radon–Nikodym derivative of $\mathcal{H}^D$ (or $\mathcal{S}^D$) with respect to a smooth volume $\mu$ ?

Notice that, while the answer to question (b) is independent on the choice among $\mathcal{H}^D$ or $\mathcal{S}^D$, in question (c) the density depends on which Hausdorff volume we work with.

We will answer to these questions under the assumptions that the sub-Riemannian manifold is equiregular, and when we choose the spherical Hausdorff volume.

**Definition 20.15.** Let $M$ be a $n$-dimensional smooth manifold, which is connected and orientable. By a *smooth volume* on $M$ we mean a measure $\mu$ on $M$ associated with a smooth non-vanishing and positively oriented $n$-form $\omega \in \Lambda^n M$, i.e., for every subset $A \subset M$ we set

$$\mu(A) = \int_A \omega. \tag{20.37}$$

In a coordinate system $(x_1, \dots, x_n)$ on $M$ a smooth volume is locally written as

$$\omega = g(x) dx_1 \wedge \dots \wedge dx_n,$$

where $g : M \to \mathbb{R}$ is some smooth and strictly positive function. In this case (20.37) simply means

$$\mu(A) = \int_A g(x) dx_1 \cdots dx_n.$$

The Popp volume $\mathcal{P}$ defined in the first part of the chapter, when the sub-Riemannian structure is equiregular, is a smooth volume in the sense of Definition 20.15.

*Remark* 20.16. Orientation here is only a technical condition to have a globally defined non degenerate $n$-form. All the results in what follows can be stated without the orientability assumption, replacing differential forms with smooth densities. We refer the interested reader to [Lee13, Chapter 16].

### 20.6.1 Hausdorff dimension

Given an equiregular sub-Riemannian manifold (see Section 20.1) we set $d_i = \dim \mathcal{D}^i_q$ for $i \geq 0$ (where $d_0 = 0$) and we define the homogeneous dimension

$$Q := \sum_{i=1}^{k} i(d_i - d_{i-1}), \tag{20.38}$$

The following two lemmas are crucial in the sequel. The first one is a uniform volume estimate and is a consequence of the Ball-box theorem for equiregular manifold.

**Lemma 20.17.** *Let $M$ be an equiregular sub-Riemannian manifold and let $\mu$ be a smooth volume. For every compact $K \subset M$ there exist $\varepsilon_0 > 0$ and $0 < c_1 < c_2$ such that*

$$c_1 \varepsilon^Q \leq \mu(B(q, \varepsilon)) \leq c_2 \varepsilon^Q. \tag{20.39}$$

*for every point $q \in K$ and every $\varepsilon < \varepsilon_0$.*

*Proof.* Fix a point $q \in M$ and consider privileged coordinates around this point. By the Ball-Box theorem (Theorem 10.67) there exist constants $c_1', c_2' > 0$ such that

$$c_1' \mathrm{Box}(\varepsilon) \subset B(q, \varepsilon) \subset c_2' \mathrm{Box}(\varepsilon) \tag{20.40}$$

for $\varepsilon < \varepsilon_0$ small enough (both the constants and $\varepsilon_0$ depending on $q$). Recall that

$$\mathrm{Box}(\varepsilon) = \{x \in \mathbb{R}^n : |x_i| \leq \varepsilon^i, i = 1, \dots, k\}. \tag{20.41}$$

in privileged coordinates $x = (x_1, \dots, x_k) \in \mathbb{R}^{n_1} \oplus \ldots \oplus \mathbb{R}^{n_k} = \mathbb{R}^n$ (cf. Chapter 10). It follows that

$$\mu(c_1' \mathrm{Box}(\varepsilon)) \leq \mu(B(q, \varepsilon)) \leq \mu(c_2' \mathrm{Box}(\varepsilon)). \tag{20.42}$$

In privileged coordinates $\mu = g(x) dx$ where $g$ is a smooth function in $\mathbb{R}^n$ and $dx$ is the Lebesgue volume. The volume of $\mathrm{Box}(\varepsilon)$ with respect to the Lebesgue volume is $\varepsilon^Q$. Since $g$ is smooth, it is uniformly bounded from above and below on compact sets, hence there exist two constants $c_1, c_2$ such that

$$c_1\varepsilon^Q \leq \mu(c_1'\mathrm{Box}(\varepsilon)), \qquad \text{and} \qquad \mu(c_2'\mathrm{Box}(\varepsilon)) \leq c_2\varepsilon^Q. \tag{20.43}$$

for $\varepsilon > 0$ small enough. Since the structure is equiregular these estimate can be made uniform for points $q$ in a compact $K$ (more precisely, the constants $c_i'(q)$ are bounded from above and from below). $\square$

The second lemma permits to pass from a uniform estimate of the measure of balls in term of diameters to an estimate with respect to spherical Hausdorff measure.

**Lemma 20.18.** *Let $\mu$ be a smooth volume and fix $q \in M$. Assume there exists $\varepsilon_0 > 0$ and $0 < c_1 < c_2$ such that, for every point $p \in B(q, \varepsilon_0)$ and every $\varepsilon < \varepsilon_0$, there holds*

$$c_1 \operatorname{diam}(B(p,\varepsilon))^Q \leq \mu(B(p,\varepsilon)) \leq c_2 \operatorname{diam}(B(p,\varepsilon))^Q. \tag{20.44}$$

*Then, for every $\varepsilon < \varepsilon_0$,*

$$c_1\mathcal{S}^Q(B(q,\varepsilon)) \leq \mu(B(q,\varepsilon)) \leq c_2\mathcal{S}^Q(B(q,\varepsilon)) \tag{20.45}$$

*Proof.* We prove separately the two inequalities.
(i). Let $\bigcup_i B(p_i, r_i)$ be an arbitrary covering of $B(q,\varepsilon)$ with balls of radius smaller than $\delta < \varepsilon_0$. If $\delta$ is small enough, every $p_i$ belongs to $B(q, \varepsilon_0)$ and, using (20.44), there holds

$$\mu(B(q,\varepsilon)) \leq \sum_i \mu(B(p_i, r_i)) \leq c_2 \sum_i \operatorname{diam}(B(p_i, r_i))^Q.$$

Hence, passing to the infimum on such coverings we have

$$\mu(B(q,\varepsilon)) \leq c_2\mathcal{S}^Q(B(q,\varepsilon)).$$

(ii). For the other inequality, fix any $\eta > 0$ and $0 < \delta < \varepsilon_0$, and let $\bigcup_i B(p_i, r_i)$ be a covering of $B(q,\varepsilon)$ such that $p_i \in B(q,\varepsilon)$, $r_i < \delta$, and

$$\sum_i \mu(B(p_i, r_i)) \leq \mu(B(q,\varepsilon)) + \eta.$$

Such a covering exists due to the Vitali covering lemma. Using as above (20.44), we obtain

$$c_1 \sum_i \operatorname{diam}(B(p_i, r_i))^Q \leq \sum_i \mu(B(p_i, r_i)) \leq \mu(B(q,\varepsilon)) + \eta.$$

Since $\mathcal{S}_\delta^Q$ is an infimum on coverings we have

$$c_1\mathcal{S}_\delta^Q(B(q,\varepsilon)) \leq \mu(B(q,\varepsilon)) + \eta.$$

Letting $\eta$ and $\delta$ tend to 0, one gets the conclusion. $\square$

Combining these two lemmas we easily obtain the following result.

**Theorem 20.19.** *Let $M$ be an equiregular sub-Riemannian manifold and let $\mu$ be a smooth volume. Then* *(i)* $\mathcal{S}^Q(B(q,\varepsilon)) < +\infty$ *for* $\varepsilon < \varepsilon_0 = \varepsilon_0(q)$,

*(ii)* $\dim_H M = Q$,

*(iii)* $\mathcal{S}^Q$ *is a Radon measure on* $M$.

*Proof.* (i). Let us prove that small balls have finite $\mathcal{S}^Q$ volume, i.e., for every $p \in M$ there exists $\varepsilon > 0$ such that $\mathcal{S}^Q(B(p,\varepsilon)) < +\infty$ for $\varepsilon$ small enough. Thanks to Proposition 4.69 we have that $\operatorname{diam}(B(q,\varepsilon)) = 2\varepsilon$ for $\varepsilon$ small enough, combined with Lemma 20.17 we have that the assumptions of Lemma 20.18 are satisfied. It follows from estimate (20.45) that $\mathcal{S}^Q(B(p,\varepsilon)) < +\infty$.

(ii). Fix a ball $B(p,\varepsilon)$ such that $\mathcal{S}^Q(B(q,\varepsilon)) < +\infty$, thanks to part (i). Then $\mathcal{S}^{Q+\eta}(M) \geq \mathcal{S}^{Q+\eta}(B(p,\varepsilon)) = +\infty$ hence $\dim_H M \leq Q$. Analogously let us write $M = \cup_i B(q_i,\varepsilon_i)$ where every $\mathcal{S}^Q(B(q_i,\varepsilon_i)) < +\infty$. Then $\mathcal{S}^{Q+\eta}(M) \leq \sum_i \mathcal{S}^{Q+\eta}(B(q_i,\varepsilon_i)) = 0$ and $\dim_H M \geq Q$.

(iii). Since $\mathcal{S}^Q$ is locally finite by (i), then it is finite on compact sets by classical covering arguments. □

### 20.6.2 On the metric convergence

Recall that, given an equiregular sub-Riemannian structure, we can fix a basis of vector fields $V_1, \ldots, V_n$ on the tangent space which is privileged at every point in a neighborhood of a fixed point $q \in M$ (cf. Section 10.4.3).

A continuous (actually smooth) system of privileged coordinates in a neighborhood $\Omega$ of a point $q \in M$ is given by the map

$$\overline{\Psi} : \Omega \times \mathbb{R}^n \to M, \qquad \Psi(p, s_1, \ldots, s_n) = p \circ e^{s_1 V_1} \circ \ldots \circ e^{s_n V_n}, \tag{20.46}$$

We call this a set of uniform privileged coordinates around the point $q \in M$.

From the equiregularity assumption and convergence of the metrics we have the following uniform estimate.

**Lemma 20.20.** *Let* $M$ *be an equiregular sub-Riemannian manifold and fix uniform privileged coordinates around* $q \in M$. *For every* $\varepsilon > 0$ *small enough there exists* $r = r(\varepsilon)$ *such that*

$$\widehat{B}_q(\varepsilon(1-r)) \subset B(q,\varepsilon) \subset \widehat{B}_q(\varepsilon(1+r)), \tag{20.47}$$

*with* $r(\varepsilon) \to 0$ *when* $\varepsilon \to 0$. *Moreover* $r(\varepsilon)$ *can be chosen uniformly in a neighborhood of* $q$.

*Proof.* Fix a set of privileged coordinates and denote by $d^\varepsilon$ the sub-Riemannian distance defined by the $\varepsilon$-approximations of the vector fields from a generating frame at a point $q$. Thanks to Theorem 10.65 we have that $d^\varepsilon \to \widehat{d}$ on compact sets. This implies that there exists $\varepsilon_0$ such that for every $\varepsilon < \varepsilon_0$ we have the inclusions

$$\widehat{B}_q(1-r) \subset B^\varepsilon(q,1) \subset \widehat{B}_q(1+r), \tag{20.48}$$

for some $r = r(\varepsilon) > 0$, depending on $\varepsilon$, and such that $r(\varepsilon) \to 0$ when $\varepsilon \to 0$. Applying the dilation $\delta_\varepsilon$ to (20.48), one gets (20.47) (recall that $\delta_\varepsilon(B^\varepsilon(q,1)) = B(q,\varepsilon)$ thanks to 10.61). □

### 20.6.3 Induced volumes and estimates

**Definition 20.21.** Let $M$ be an equiregular sub-Riemannian manifold. If $\mu$ a smooth volume on $M$, associated to $\omega$, we define the *induced volume* $\widehat{\mu}_q$ at the point $q$ as the left-invariant volume on the nilpotent tangent space at $q$ canonically associated with $\omega_\mu(q) \in \wedge^n(T_q^*M)$ (cf. isomorphism given by the Lemma 20.4).

We prove now the main estimate, that is a refinement of Lemma 20.17.

**Proposition 20.22.** *Let $\mu$ be a smooth volume and $\widehat{\mu}_q$ the induced volume at the nilpotent approximation at $q$. For $\varepsilon \to 0$*

$$\mu(B(q,\varepsilon)) = \varepsilon^Q \widehat{\mu}_q(\widehat{B}_q) + o(\varepsilon^Q), \tag{20.49}$$

*where $o(\varepsilon^Q)$ is uniform as $q$ varies in $M$ and $\widehat{B}_q$ is the ball centered at $0$ of radius $1$ in the nilpotent approximation at $q$ of the sub-Riemannian manifold.*

*Proof.* The result is of local nature. Fix a set of privileged coordinates $(x_1, \dots, x_n)$ where $q = 0$, $\mu = g(x)dx$ and $\widehat{\mu}_q = g(0)dx$. Then

$$\mu(B(q,\varepsilon)) = \int_{B(q,\varepsilon)} g(x)dx = \int_{B(q,\varepsilon)} (g(0) + o(1))dx = \widehat{\mu}_q(B(q,\varepsilon)) + o(\varepsilon^Q), \tag{20.50}$$

where we used that the measure of $B(q,\varepsilon)$ is $O(\varepsilon^Q)$ (cf. Lemma 20.17). Then using the homogeneity of $\widehat{\mu}_q$ we get

$$\widehat{\mu}_q(B(q,\varepsilon)) = \varepsilon^Q \widehat{\mu}_q(\delta_{1/\varepsilon} B(q,\varepsilon)). \tag{20.51}$$

To conclude, we use the following fact, stated in privileged coordinates.

**Lemma 20.23.** *We have that $\widehat{\mu}_q(\delta_{1/\varepsilon} B(q,\varepsilon)) \to \widehat{\mu}_q(\widehat{B}_q)$ for $\varepsilon \to 0$.*

*Proof.* For $\varepsilon > 0$ small enough we have applying $\delta_{1/\varepsilon}$ to (20.47)

$$\widehat{B}_q(1-r) \subset \delta_{1/\varepsilon} B(q,\varepsilon) \subset \widehat{B}_q(1+r). \tag{20.52}$$

Hence

$$\begin{aligned} \widehat{\mu}_q(\delta_{1/\varepsilon} B(q,\varepsilon)) - \widehat{\mu}_q(\widehat{B}_q) &\leq \widehat{\mu}_q(\widehat{B}_q(1+r)) - \widehat{\mu}_q(\widehat{B}_q) \\ &\leq \left((1+r)^Q - 1\right) \widehat{\mu}_q(\widehat{B}_q). \end{aligned}$$

Analogously one proves the inequality

$$\widehat{\mu}_q(\delta_{1/\varepsilon} B(q,\varepsilon)) - \widehat{\mu}_q(\widehat{B}_q) \geq \left((1-r)^Q - 1\right) \widehat{\mu}_q(\widehat{B}_q).$$

Using that $r(\varepsilon) \to 0$ for $\varepsilon \to 0$ (uniformly in $q$) one gets the claim. $\square$

We now go back to the proof of the Proposition 20.22. Applying Lemma 20.23 to (20.51) we get

$$\widehat{\mu}_q(B(q,\varepsilon)) = \varepsilon^Q \widehat{\mu}_q(\delta_{1/\varepsilon} B(q,\varepsilon)) = \varepsilon^Q \widehat{\mu}_q(\widehat{B}_q) + o(\varepsilon^Q) \tag{20.53}$$

Combining (20.50) and (20.53) we have (20.49). $\square$

**Proposition 20.24.** *Assume $M$ is equiregular endowed with a smooth volume $\mu$. The function $\widehat{\mu} : M \to \mathbb{R}$ defined by $\widehat{\mu} : q \mapsto \widehat{\mu}_q(\widehat{B}_q)$ is continous.*

To prove this Proposition we use an analogue of Lemma 20.20 for a family of nilpotent approximation depending on the point.

**Lemma 20.25.** *Let $\{\widehat{d}_q\}_{q\in M}$ be the family of nilpontent approximations on an equiregular sub-Riemannian manifold. There exists $r = r(p, q)$ such that*

$$\widehat{B}_q(1-r) \subset \widehat{B}_p(1) \subset \widehat{B}_q(1+r), \tag{20.54}$$

*where $r(p, q) \to 0$ when $d(p, q) \to 0$. Moreover the estimate (20.54) is uniform for $p, q$ in a compact set.*

*Proof of Proposition 20.24.* Thanks to the previous discussion, the measure $\widehat{\mu}_q$ can be represented in a set of uniform privlieged coordinates as

$$\widehat{\mu}_q(A) = \int_A \widehat{g}(q, x) dx_1 \cdots dx_n,$$

where $\widehat{g}$ is a smooth function of the two parameters. We have to estimate the difference

$$|\widehat{\mu}_p(\widehat{B}_p) - \widehat{\mu}_q(\widehat{B}_q)| \le |\widehat{\mu}_p(\widehat{B}_p) - \widehat{\mu}_q(\widehat{B}_p)| + |\widehat{\mu}_q(\widehat{B}_q) - \widehat{\mu}_q(\widehat{B}_p)|,$$

and show that the right hans side tends to zero when $p \to q$.

For the first term, using that $\widehat{g}$ is smooth in $q$ uniformly in $x$ we get from the mean value theorem[1]

$$\begin{aligned} |\widehat{\mu}_p(\widehat{B}_p) - \widehat{\mu}_q(\widehat{B}_p)| &\le \int_{\widehat{B}_p} |\widehat{g}(p, x) - \widehat{g}(q, x)|\, dx_1 \cdots dx_n, \\ &\le d(p, q) \int_{\widehat{B}_p} |\widehat{G}(x)| dx_1 \cdots dx_n, \end{aligned}$$

for some smooth function $\widehat{G}(x)$. Fix a compact $K$ such that $\widehat{B}_p \subset K$ for all $p$ in a neighborhood of $q$. Hence there exists $C > 0$ such that for $p, q$ close enough

$$|\widehat{\mu}_p(\widehat{B}_p) - \widehat{\mu}_q(\widehat{B}_p)| \le C d(p, q),$$

For the second one

$$|\widehat{\mu}_q(\widehat{B}_q) - \widehat{\mu}_q(\widehat{B}_p)| \le \widehat{\mu}_q(\widehat{B}_p \ominus \widehat{B}_q),$$

where $\widehat{B}_p \ominus \widehat{B}_q := (\widehat{B}_p \setminus \widehat{B}_q) \cup (\widehat{B}_q \setminus \widehat{B}_p)$ is the symmetric set difference. Thanks to Lemma 20.25

$$\begin{aligned} \widehat{\mu}_q(\widehat{B}_p \ominus \widehat{B}_q) &\le \widehat{\mu}_q(\widehat{B}_q(1+r) \setminus \widehat{B}_q) \cup \widehat{\mu}_q(\widehat{B}_q \setminus \widehat{B}_q(1-r)) \\ &\le \widehat{\mu}_q(\widehat{B}_q(1+r) \setminus \widehat{B}_q(1-r)) \\ &\le [(1+r)^Q - (1-r)^Q]\widehat{\mu}_q(\widehat{B}_q). \end{aligned}$$

Since $(1+r)^Q - (1-r)^Q = O(r)$ and $r(p, q) \to 0$ when $d(p, q) \to 0$, the proof is completed. $\square$

[1]in coordinates

$$|\widehat{g}(p, x) - \widehat{g}(q, x)| \le \sup_{t\in[0,1]} \left| \frac{\partial \widehat{g}}{\partial q}(tp + (1-t)q, x) \right| |p - q| \le |\widehat{G}(x)| d(p, q).$$

## 20.7 Density of the spherical Hausdorff volume with respect to a smooth volume

We state now the main result of this section. Recall that $\mu$ is a fixed smooth volume on $M$.

**Theorem 20.26.** *The measure $\mathcal{S}^Q$ is absolutely continuous with respect to $\mu$. The Radon–Nikodym derivative is given by*

$$\frac{d\mu}{d\mathcal{S}^Q}(q) = \frac{\widehat{\mu}_q(\widehat{B}_q)}{2^Q} \tag{20.55}$$

*is a continuous function on $M$. In particular, if $A$ is a Borel set in $M$, we have the formula*

$$\mu(A) = \frac{1}{2^Q}\int_A \widehat{\mu}_q(\widehat{B}_q) d\mathcal{S}^Q. \tag{20.56}$$

Theorem 20.26 is a consequence of the differentiation theorem for Radon measures in metric spaces [Sim83, Thm. 4.7] (see also [Fed69]). This result guarantees that the Radon–Nikodym derivative can be computed as follows:

$$\frac{d\mu}{d\mathcal{S}^Q}(q) = \lim_{\varepsilon\to 0}\frac{\mu(B(q,\varepsilon))}{\mathcal{S}^Q(B(q,\varepsilon))}. \tag{20.57}$$

The proof is then completed by the following result.

**Proposition 20.27.** *Assume $M$ is equiregular. Then, for every smooth volume $\mu$ on $M$,*

$$\lim_{\varepsilon\to 0}\frac{\mu(B(q,\varepsilon))}{\mathcal{S}^Q(B(q,\varepsilon))} = \frac{\widehat{\mu}_q(\widehat{B}_q)}{2^Q}, \qquad \forall\, q\in M. \tag{20.58}$$

*Proof of Proposition 20.27.* Fix $q \in M$. For every $p$ in a neighborhood of $q$ we have the following expansions

$$\mu(B(p,\varepsilon)) = \varepsilon^Q\widehat{\mu}_p(\widehat{B}_p) + o(\varepsilon^Q), \tag{20.59}$$
$$\operatorname{diam}(B(p,\varepsilon)) = 2\varepsilon + o(\varepsilon). \tag{20.60}$$

Thanks to the equiregularity of $M$, these expansions (i.e., the quantities $o(\varepsilon^Q)$ and $o(\varepsilon)$) are uniform for $p$ in a neighborhood of $q$ (cf. the uniform estimates (20.47)). Moreover $p\mapsto \widehat{\mu}_p(\widehat{B}_p)$ is continuous in a neighborhood of $q\in M$, thanks to Proposition 20.24.

Therefore we can write in a neighborhood of $q$

$$\frac{\mu(B(p,\varepsilon))}{\operatorname{diam}(B(p,\varepsilon))^Q} = \frac{\varepsilon^Q\widehat{\mu}_p(\widehat{B}_p)+o(\varepsilon^Q)}{2^Q\varepsilon^Q+o(\varepsilon^Q)} = \frac{\widehat{\mu}_p(\widehat{B}_p)}{2^Q}(1+o(1)),$$

where $o(1)$ is uniform in $p$. Thus for every $\eta > 0$ we can find $0 < \varepsilon_0 < \eta$ such that for every $p \in B(q,\varepsilon_0)$ and every $\varepsilon < \varepsilon_0$ we have

$$\frac{\widehat{\mu}_q(\widehat{B}_q)}{2^Q} - \eta \le \frac{\mu(B(p,\varepsilon))}{\operatorname{diam}(B(p,\varepsilon))^Q} \le \frac{\widehat{\mu}_q(\widehat{B}_q)}{2^Q} + \eta.$$

Applying Lemma 20.18 one obtains that for every $\eta > 0$ there exists $\varepsilon > 0$ such that for $\varepsilon < \varepsilon_0$ we have

$$\frac{\widehat{\mu}_q(\widehat{B}_q)}{2^Q} - \eta \le \frac{\mu(B(q,\varepsilon))}{\mathcal{S}^Q(B(q,\varepsilon))} \le \frac{\widehat{\mu}_q(\widehat{B}_q)}{2^Q} + \eta,$$

that is the definition of the limit (20.58).

□

**Exercise 20.28.** Prove that if $(G, d)$ is a Carnot group and $Q$ is its Hausdorff dimension, then $\mathcal{S}^Q(B(x,r)) = 2^Q r^Q$. This says that formula (20.58) can be reinterpreted as

$$\lim_{\varepsilon\to 0} \frac{\mu(B(q,\varepsilon))}{\mathcal{S}^Q(B(q,\varepsilon))} = \frac{\widehat{\mu}_q(\widehat{B}_q)}{\widehat{\mathcal{S}}^Q(\widehat{B}_q)}, \tag{20.61}$$

where $\widehat{\mathcal{S}}^Q$ is the spherical Hausdorff volume on the nonholonomic tangent space.

## 20.8 Bibliographical note

The problem to define a canonical volume on a sub-Riemannian manifold was first pointed out by Brockett in his seminal paper [Bro82], motivated by the construction of a Laplace operator on a 3D sub-Riemannian manifold canonically associated with the metric structure, analogous to the Laplace-Beltrami operator on a Riemannian manifold.

Montgomery addressed the problem in the general case in [Mon02, Chapter 10]. A first natural volume is Popp's volume, first defined by Octavian Popp but appeared in the literature only in [Mon02]. A second natural volume Hausdorff volume. These two measure are mutually absolutely continuous and the question is whether the density is smooth or not.

A first answer to this question was given in [ABB12], where also the formula for the density of the spherical Hausdorff measure with respect to a smooth first appeared. The proof given here for the formula of the density is essentially taken from [GJ14], where the authors extends the study to non-equiregular situations (see also [GJ15]). Indeed the key argument in the proof of the density formula in [ABB12] contains a gap, in the sense that the proof is complete only for centered[2] Hausdorff measure. We refer, for instance, to the paper [FSSC15] for a discussion on centered Hausdorff measure and its comparison to spherical one. The formula for the Hausdorff dimension of sub-Riemannian manifolds first appeared in [Mit85].

The smoothness of the density of the spherical Hausdorff measure with respect to a smooth one have also been studiend in corank 2 sub-Riemannian manifolds in [BBG12, BG13].

Further discussions on Popp's volume can be found in [ABGR09] and [BR13].

[2]in the definition of centered Hausdorff measure one requires that the center of the balls belong to the set one is measuring.

# Chapter 21

# The sub-Riemannian heat equation

In this chapter we derive the sub-Riemannian heat equation and its relation to the notion of intrinsic volume in sub-Riemannian geometry. We then discuss (without proofs) the well-posedness of the Cauchy problem, the smoothness of its solution and the relation with the bracket-generating condition (Hörmander theorem). In the last part of the chapter we present en elementary method to compute the fundamental solution of the heat equation on the Heisenberg group (the celebrated Gaveau-Hulanicki formula) and we briefly discuss the relation between the small-time heat kernel asymptotics and the sub-Riemannian distance.

## 21.1 The heat equation

To write the heat equation in a general sub-Riemannian manifold, let us start by writing it in the Riemannian context and let us see which mathematical structures are missing in the sub-Riemannian one.

### 21.1.1 The heat equation in the Riemannian context

Let $(M, g)$ be an oriented[1] Riemannian manifold of dimension $n$ and let $\mathcal{R}$ be the Riemannian volume form defined by

$$\mathcal{R}(X_1, \ldots, X_n) = 1, \text{ where } \{X_1, \ldots, X_n\} \text{ is a local orthonormal frame.}$$

In coordinates if $g$ is represented by a matrix $(g_{ij})$, we have

$$\mathcal{R} = \sqrt{\det(g_{ij})}\, dx_1 \wedge \ldots \wedge dx_n.$$

Let $\phi$ be a quantity (depending on the position $q$ and on the time $t$) subjects to a diffusion process. For example it may represent the temperature of a body, the concentration of a chemical product, the noise etc. Let $\mathbf{F}$ be a time dependent vector field representing the *flux* of the quantity $\phi$, i.e., how much of $\phi$ is flowing through the unity of surface in unitary time.

Our purpose is to get a partial differential equation describing the evolution of $\phi$. The Riemannian heat equation is obtained by postulating the following two facts:

[1] we work an oriented manifold for simplicity of presentation. In the non-orientable case, a never vanishing globally defined $n$ form does not exist. However one can repeat the same arguments using densities. See for instance [Tay96, Section 2.2].

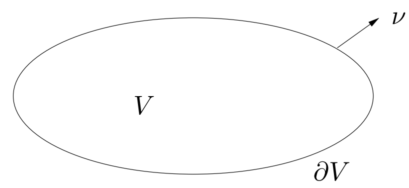



Figure 21.1: Heat conservation in $V$.

(R1) the flux is proportional to minus the gradient of $\phi$, i.e., normalizing the proportionality constant to one, we assume that

$$\mathbf{F} = -\mathrm{grad}(\phi); \tag{21.1}$$

(R2) the quantity $\phi$ satisfies a conservation law, i.e., for every bounded open set $V$ having a smooth boundary $\partial V$ we have the following: the rate of decreasing of $\phi$ inside $V$ is equal to the rate of flowing of $\phi$ via $\mathbf{F}$, out of $V$, through $\partial V$. See Figure 21.1. In formulas this is written as

$$-\frac{d}{dt}\int_V \phi\,\mathcal{R} = \int_{\partial V} \mathbf{F}\cdot\nu\,\mathrm{dS}. \tag{21.2}$$

Here $\nu$ is the external (Riemannian) normal to $\partial V$ and dS is the element of area induced by $\mathcal{R}$ on $M$, thanks to the Riemannian structure, i.e., $\mathrm{dS} = \mathcal{R}(\nu,\cdot)$. The symbol $\mathbf{F}\cdot\nu$ is a notation for $g_q(\mathbf{F}(q,t),\nu(q))$.

Applying the Riemannian divergence theorem to (21.2) and using (21.1) we have then

$$-\frac{d}{dt}\int_V \phi\,\mathcal{R} = \int_{\partial V} \mathbf{F}\cdot\nu\,\mathrm{dS} = \int_V \mathrm{div}_{\mathcal{R}}(\mathbf{F})\,\mathcal{R} = -\int_V \mathrm{div}_{\mathcal{R}}(\mathrm{grad}(\phi))\,\mathcal{R}.$$

By the arbitrarity of $V$ and defining the Riemannian Laplacian (usually called the Laplace-Beltrami operator) as

$$\triangle\phi = \mathrm{div}_{\mathcal{R}}(\mathrm{grad}(\phi)), \tag{21.3}$$

we get the heat equation

$$\frac{\partial}{\partial t}\phi(q,t) = \triangle\phi(q,t).$$

**Useful expressions for the Riemannian Laplacian**

In this section we get some useful expressions for $\triangle$. To this purpose we have to recall what are grad and $\mathrm{div}_{\mathcal{R}}$ in formula (21.3).

We recall that the gradient of a smooth function $\varphi : M \to \mathbb{R}$ is a vector field pointing in the direction of the greatest rate of increase of $\varphi$ and its magnitude is the derivative of $\varphi$ in that direction. In formulas, it is the unique vector field $\mathrm{grad}(\varphi)$ satisfying for every $q \in M$,

$$g_q(\mathrm{grad}(\varphi), v) = d\varphi(v), \quad \text{for every } v \in T_qM. \tag{21.4}$$

In coordinates, if $g$ is represented by a matrix $(g_{ij})$, and calling $(g^{ij})$ its inverse, we have

$$\mathrm{grad}(\varphi)^i = \sum_{j=1}^n g^{ij}\partial_j\varphi. \tag{21.5}$$

If $\{X_1, \dots, X_n\}$ is a local orthonormal frame for $g$, we have the useful formula

$$\mathrm{grad}(\varphi) = \sum_{i=1}^{n} X_i(\varphi) X_i. \tag{21.6}$$

**Exercise 21.1.** Prove that if the Riemannian metric is defined globally via a generating family $\{X_1, \dots, X_m\}$ with $m \geq n$, in the sense of Chapter 3, then $\mathrm{grad}(\varphi) = \sum_{i=1}^{m} X_i(\varphi) X_i$.

Recall that the divergence of a smooth vector field $X$ says how much the volume is increasing or decreasing along the flow of $X$. It is defined in the following way. The Lie derivative in the direction of $X$ of the volume form is still a $n$-form and hence point-wise proportional to the volume form itself. The function of proportionality is by definition is the divergence of $X$. In formulas,

$$\mathcal{L}_X \mathcal{R} = \mathrm{div}_{\mathcal{R}}(X) \mathcal{R}.$$

Now, using $d\mathcal{R} = 0$ and the Cartan formula (4.83), we have that $\mathcal{L}_X \mathcal{R} = i_X d\mathcal{R} + d(i_X \mathcal{R}) = d(i_X \mathcal{R})$. Hence the divergence of a vector field $X$ can be defined by

$$d(i_X \mathcal{R}) = \mathrm{div}_{\mathcal{R}}(X) \mathcal{R}. \tag{21.7}$$

In coordinates, if $\mathcal{R} = h(x) dx^1 \wedge \dots dx^n$ we have

$$\mathrm{div}_{\mathcal{R}}(X) = \frac{1}{h(x)} \sum_{i=1}^{n} \partial_i (h(x) X^i). \tag{21.8}$$

*Remark* 21.2. Notice that to define the divergence of a vector field it is not necessary a Riemannian structure, but only a volume form (i.e., a smooth $n$-form globally defined).

If we put together formula (21.5) and formula (21.8), with $X = \mathrm{grad}(\varphi)$ we get the well-known expression for the Laplace Beltrami operator,

$$\triangle(\varphi) = \mathrm{div}_{\mathcal{R}}(\mathrm{grad}(\varphi)) = \frac{1}{h(x)} \sum_{i,j=1}^{n} \partial_i (h(x) g^{ij} \partial_j \varphi). \tag{21.9}$$

Combining formula 21.6 with the property $\mathrm{div}(aX) = a\,\mathrm{div}(X) + X(a)$ where $X$ is a vector field and $a$ is a function, we get

$$\triangle(\varphi) = \sum_{i=1}^{n} \big( X_i^2 \varphi + \mathrm{div}_{\mathcal{R}}(X_i) X_i(\varphi) \big) \tag{21.10}$$

where $\{X_1, \dots X_n\}$ is a local orthonormal frame. Similarly, defining the Riemannian structure via a generating family $\{X_1, \dots X_m\}$, for $m \geq n$, we get

$$\triangle(\varphi) = \sum_{i=1}^{m} \big( X_i^2 \varphi + \mathrm{div}_{\mathcal{R}}(X_i) X_i(\varphi) \big). \tag{21.11}$$

*Remark* 21.3. Notice that one could consider a diffusion process on a Riemannian manifold measuring the gradient with the Riemannian structure and the volume with a volume form $\omega$ different from $\mathcal{R}$. In this case one would get a heat equation of the form (one can do this explicitly by using Lemma 21.4 below).

$$\frac{\partial}{\partial t} \phi(q,t) = \triangle \phi(q,t), \text{ where } \triangle \phi = \mathrm{div}_{\omega}(\mathrm{grad}(\phi)).$$

From Formula 21.10 or (21.11) one gets that the choice of the volume form does not affect the second order terms, but only the first order ones.

### 21.1.2 The heat equation in the sub-Riemannian context

Let $M$ be a sub-Riemannian manifold of dimension $n$. To write a heat-like equation in the sub-Riemannian context we follow what we did in the Riemannian case. However many ingredients are missing and we have to follow a different argument to derive the heat equation. We denote by $\phi$ the quantity that is subject to the diffusion process, and we postulate that:

(SR1) the heat flows in the direction where $\phi$ is increasing more, but only among horizontal directions;

(SR2) the quantity $\phi$ satisfies a conservation law, i.e., for every bounded open set $V$ having a smooth and orientable boundary $\partial V$, the rate of decreasing of $\phi$ inside $V$ is equal to the rate of flowing of $\phi$, out of $V$, through $\partial V$.

For (SR1) we need:

A. a notion of horizontal gradient;

for (SR2) we need:

B. a way of computing the volume;

C. a way to express the conservation law without using the Riemannian normal $\nu$ to $\partial V$, the scalar product between $\nu$ and the flux and the Riemannian divergence theorem.

Let us now discuss A, B, and C.

#### A. The horizontal gradient

In sub-Riemannian geometry the gradient of a smooth function $\varphi : M \to \mathbb{R}$ is a horizontal vector field (called horizontal gradient) pointing in the horizontal direction of the greatest rate of increase of $\varphi$ and its magnitude is the derivative of $\varphi$ in that direction. In formulas it is the unique vector field $\mathrm{grad}_H(\varphi)$ satisfying for every $q \in M$,

$$\langle \mathrm{grad}_H(\varphi) \,|\, v \rangle_q = d\varphi(v), \quad \text{for every } v \in \mathcal{D}_q M. \tag{21.12}$$

Here $\langle \cdot \,|\, \cdot \rangle_q$ is the scalar product induced by the sub-Riemannian structure on $\mathcal{D}_q$ (see Exercise 3.9). If $\{X_1, \dots, X_m\}$ is a generating family then

$$\mathrm{grad}_H(\varphi) = \sum_{i=1}^{m} X_i(\varphi) X_i.$$

#### B. Measuring the volume

As in the Riemannian case, let us assume for simplicity that $M$ is oriented. The construction of a canonical volume form in sub-Riemannian geometry (i.e., a volume form obtained using only the sub-Riemannian structure) have been discussed in Chapter 20. We have seen that, in the equiregular case, a canonical construction exists and the volume form obtained in that way is called Popp's volume. However other constructions are possible. Being $(M, d)$ a metric space one can for instance use the Hausdorff volume or the spherical Hausdorff volume. In certain cases, these different constructions give rise to the same volume form (up to a multiplicative constant). In others cases, different constructions give rise to a different volume form. Here, we are not going to discuss the details of this problem. Let us just recall that the three situations that one can encounter are (we refer to the bibliographical note for some references):

- rank-varying or non-equiregular cases. In this case a construction of a canonical smooth volume form is not known.

- equiregular cases for which the nilpotent approximation at different points are isometric. In this case Popp's volume is in a sense the only canonical volume (up to a multiplicative constant) that one can build;[2]

- equiregular cases for which nilpotent approximations at different points are not isometric. In this case one can build an infinite number of canonical volumes and Popp's volume is only one of the possible constructions.

For left-invariant sub-Riemannian structures on Lie groups, the nilpotent approximations at different points are isometric and we are in the second case. For these structures Popp's volume is a left-invariant volume form and hence it coincides (up to a multiplicative constant) with the left *Haar measure* on the group that is a canonical volume that can be built on any Lie group.

Due to these difficulties, in the following we assume that a volume form $\omega$ (i.e., a smooth $n$-form globally defined) is assigned independently of the sub-Riemannian structure.

### C. Conservation laws without a Riemannian structure

The next step is to express the conservation of the heat without a Riemannian structure. This can be done thanks to the following lemma, whose proof is left as an exercise.

**Lemma 21.4.** *Let $M$ be a smooth manifold provided with a smooth volume form $\omega$. Let $\Omega$ be an embedded bounded sub-manifold (possible with boundary) of codimension 1. Let $F(q,t)$ be a time-dependent complete vector field (we assume smoothness jointly in $q$ and $t$) and $P_{0,t}$ be the corresponding flow. Consider the cylinder defined by the images of $\Omega$ translated by the flow of $F$ for times between $0$ and $t$ (see Figure 21.2):*

$$\Pi_F(t,\Omega) = \{P_{0,t}(\Omega) \mid s \in [0,t]\}.$$

*Then*

$$\frac{d}{dt}\bigg|_{t=0} \int_{\Pi_F(t,\Omega)} \omega = \int_\Omega i_{F(q,0)}\, \omega(q).$$

### The heat equation

The postulate (SR1) consist then in declaring that the heat is flowing via a flux $\mathbf{F}$ given by

$$\mathbf{F} = -\mathrm{grad}_H(\phi).$$

[2]roughly speaking Popp's volume is the unique volume form (up to a multiplicative constant) that at every point $q$ depends only on the nilpotent approximation of the sub-Riemannian structure at the point $q$.

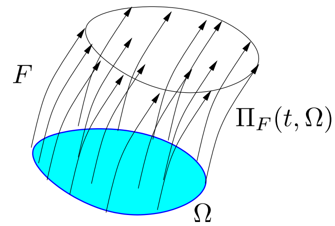



Figure 21.2: The cylinder $\Pi_F(t,\Omega)$.

The postulate (SR2) is then written as

$$-\frac{d}{dt}\int_V \phi\,\omega = \frac{d}{dt}\int_{\Pi_{\mathbf{F}(t,\partial V)}} \omega = \int_{\partial V} i_{\mathbf{F}}\,\omega,$$

where in the last equality we have used the Lemma 21.4.

Now, using Stokes theorem, the definition of divergence (21.7) and the identity $\mathbf{F} = -\mathrm{grad}_H \phi$ we have

$$\int_{\partial V} i_{\mathbf{F}}\,\omega = \int_V d(i_{\mathbf{F}}\,\omega) = \int_V \mathrm{div}_\omega(\mathbf{F})\,\omega = -\int_V \mathrm{div}(\mathrm{grad}_H(\phi))\,\omega.$$

**Definition 21.5.** Let $M$ be a sub-Riemannian manifolds and let $\omega$ be a volume on $M$. The operator $\triangle_H \phi = \mathrm{div}_\omega(\mathrm{grad}_H(\phi))$ is called the *sub-Riemannian Laplacian.*

By the arbitrarity of $V$ we get the *sub-Riemannian heat equation*

$$\frac{\partial}{\partial t}\phi(q,t) = \triangle_H \phi(q,t).$$

### 21.1.3 The Hörmander theorem and the existence of the heat kernel

The expression of the sub-Riemannian Laplacian does not change if we multiply the volume by a (non zero) constant. In the equiregular case and when the nilpotent approximation of the sub-Riemannian structure does not depend on the point, the sub-Riemannian Laplacian computed with respect to the Popp volume is called the *intrinsic sub-Laplacian.* $\triangle_{intr}\phi = \mathrm{div}_{\mathcal{P}}(\mathrm{grad}_H(\phi))$.

The same computation of the Riemannian case provides the following expression for the sub-Riemannian Laplacian,

$$\triangle_H(\phi) = \sum_{i=1}^m \left(X_i^2\phi + \mathrm{div}_\omega(X_i)X_i(\phi)\right) \quad \text{where } \{X_1,\dots X_m\}\text{, is a generating family.} \tag{21.13}$$

In the Riemannian case, the operator $\triangle_H$ is elliptic, i.e., in coordinates it has the expression

$$\triangle_H = \sum_{i,j=0}^n a_{ij}(x)\partial_i\partial_j + \text{first order terms},$$

where the matrix $(a_{ij})$ is symmetric and positive definite for every $x$.

In the sub-Riemannian (and not-Riemannian) case, $\triangle_H$ it is not elliptic since the matrix $(a_{ij})$ can have several zero eigenvalues. However, a theorem of Hörmander says that, thanks to the bracket-generating condition, $\triangle_H$ is hypoelliptic.

We have the following.

**Theorem 21.6** (Hörmander). *Let $\{Y_0, Y_1 \dots Y_k\}$ be a family of bracket-generating vector fields on a smooth manifold $M$. Then the operator $L = Y_0 + \sum_{i=1}^{k} Y_i^2$ is hypoelliptic which means that if $\varphi$ is a distribution defined on an open set $\Omega \subset M$, such that $L\varphi$ is $C^\infty$, then $\varphi$ is $C^\infty$ in $\Omega$.*

Notice that:

- Elliptic operators with $C^\infty$ coefficients are hypoelliptic.
- The heat operator $\triangle - \partial_t$, where $\triangle$ is the Laplace-Beltrami operator on a Riemannian manifold $M$, is not elliptic, since the matrix of coefficients of the second order derivatives in $\mathbb{R} \times M$ has one zero eigenvalue (the one corresponding to $t$). However it is hypoelliptic since if $\{X_1 \dots X_n\}$ is an orthonormal frame, then $Y_0 = \sum_{i=1}^{n} \mathrm{div}_{\mathcal{R}}(X_i) X_i(\phi) - \partial_t$ and $Y_1 := X_1, \dots, Y_n := X_n$ are bracket-generating in $\mathbb{R} \times M$.
- The sub-Riemannian heat operator $\triangle_H - \partial_t$ is hypoelliptic since, if $\{X_1 \dots X_m\}$ is a generating family, then $Y_0 = \sum_{i=1}^{m} \mathrm{div}_{\omega}(X_i) X_i(\phi) - \partial_t$ and $Y_1 := X_1, \dots, Y_m := X_m$ are bracket-generating in $\mathbb{R} \times M$. (The hypoellipticity of $\triangle_H$ alone is consequence of the fact that $\{X_1, \dots, X_m\}$ are bracket-generating on $M$.)

One of the most important consequences of the Hörmander theorem is that the heat evolution smooths out immediately every initial condition. Indeed if one can guarantee that a solution of $(\triangle_H - \partial_t)\varphi = 0$ exists in distributional sense in an open set $\Omega$ of $\mathbb{R} \times M$, then, being $0 \in C^\infty$, it follows that $\varphi$ is $C^\infty$ in $\Omega$.

A standard result for the existence of a solution in $L^2(M, \omega)$ is given by the following theorem.[3]

**Theorem 21.7.** *Let $M$ be a smooth manifold and $\omega$ be a volume on $M$. If $\triangle$ is a non-negative and essentially self-adjoint operator on $L^2(M, \omega)$, then, there exists a unique solution to the Cauchy problem*

$$\begin{cases} (\partial_t - \triangle)\phi = 0 \\ \phi(q, 0) = \phi_0(q) \in L^2(M, \omega), \end{cases} \tag{21.14}$$

*on $[0, +\infty[ \times M$. Moreover for each $t \in [0, +\infty[$ this solution belongs to $L^2(M, \omega)$.*

It is immediate to prove that $\triangle_H$ is non-negative and symmetric on $L^2(M, \omega)$. If in addition one can prove that $\triangle_H$ is essentially self-adjoint, then thanks to the Hörmander theorem, one has that the solution of (21.14) is indeed $C^\infty$ in $]0, +\infty[ \times M$.

The discussion of the theory of self-adjoint operators is out of the purpose of this book. However the essential self-adjointness of $\triangle_H$ is guaranteed by the completeness of the sub-Riemannian manifold as metric space.

**Theorem 21.8.** *Consider a sub-Riemannian manifold that is complete as metric space. Let $\omega$ be a volume on $M$. Then $\triangle_H$ defined on $\mathcal{C}_c^\infty(M)$ is essentially self-adjoint in $L^2(M, \omega)$.*

Typical cases in which the sub-Riemannian manifold is complete are left-invariant structure on Lie groups, sub-Riemannian manifold obtained as restriction of complete Riemannian manifolds, sub-Riemannian structures defined in $\mathbb{R}^n$ having as generating family a set of sub-linear vector fields.

[3] By $L^2(M, \omega)$ we mean functions from $M$ to $\mathbb{R}$ which are square integrable with respect to the volume $\omega$.

When the manifold is not complete as metric space (as for instance the standard Euclidean structure on the unitary disc in $\mathbb{R}^2$), then to study the Cauchy problem (21.14) one need to specify more the problem (e.g., boundary conditions).

As a consequence of the hypoellipticity of $\triangle_H - \partial_t$, of Therem 21.7 and of Theorem 21.8, we have

**Corollary 21.9.** *Consider a sub-Riemannian manifold that is complete as metric space. Let $\omega$ be a volume on $M$. There exists a unique solution to the Cauchy problem (21.14), that is $C^\infty$ in $]0,+\infty[\times M$.*

Under the hypothesis of completeness of the manifold one can also guarantee the existence of a convolution kernel.

**Theorem 21.10.** *Consider a sub-Riemannian manifold that is complete as metric space. Let $\omega$ be a volume on $M$. Then the unique solution to the Cauchy problem* (21.14) *on* $]0,+\infty[\times M$ *can be written as*

$$\phi(q,t) = \int_M \phi_0(\bar{q}) K_t(q,\bar{q})\, \omega(\bar{q})$$

*where $K_t(q,\bar{q})$ is a positive function defined on* $]0,+\infty[\times M\times M$ *which is smooth, symmetric for the exchange of $q$ and $\bar{q}$ and such that for every fixed* $(t,q) \in ]0,+\infty[\times M$*, we have $K_t(q,\cdot) \in L^2(M,\omega)$.*

The function $K_t(q,\bar{q})$ is called the *kernel* of the heat equation.

### 21.1.4 The heat equation in the non bracket-generating case

If the sub-Riemannian structure is not bracket-generating, then the operator $\triangle_H$ can be defined as above, but in general it is not hypoelliptic and the heat evolution does not smooth the initial condition.

Consider for example the non bracket-generating sub-Riemannian structure on $\mathbb{R}^3$ for which an orthonormal frame is given by $\{\partial_x, \partial_y\}$ (here we are calling $(x,y,z)$ the points of $\mathbb{R}^3$). Take as volume the Lebesgue volume on $\mathbb{R}^3$. Then $\triangle_H = \partial_x^2 + \partial_y^2$ on $\mathbb{R}^3$. This operator is not obtained from bracket-generating vector fields. Consider the corresponding heat operator $\triangle_H - \partial_t$ on $[0,+\infty[\times\mathbb{R}^3$. Since the $z$ direction is not appearing in this operator, any discontinuity in the $z$ variable is not smoothed by the evolution. For instance if $\psi(x,y,t)$ is a solution of the heat equation $\triangle_H - \partial_t = 0$ on $[0,+\infty[\times\mathbb{R}^2$, then $\psi(x,y,t)\theta(z)$ is a solution of the heat equation on $[0,+\infty[\times\mathbb{R}^3$, where $\theta$ is the Heaviside function.

## 21.2 The heat-kernel on the Heisenberg group

In this section we construct the heat kernel on the Heisenberg sub-Riemannian structure. To this purpose it is convenient to see this structure as a left-invariant structure on a matrix representation of the Heisenberg group. This point of view is useful because permits to fully exploit the left-invariance of the structure (construction of a canonical volume form, looking for a special form of the heat kernel that behave well for left-translations etc.).

### 21.2.1 The Heisenberg group as a group of matrices

The Heisenberg group $\mathbb{H}^1$ can be seen as the 3-dimensional group of matrices

$$\mathbb{H}^1 = \left\{ \begin{pmatrix} 1 & x & z+\frac{1}{2}xy \\ 0 & 1 & y \\ 0 & 0 & 1 \end{pmatrix} \mid x,y,z \in \mathbb{R} \right\},$$

endowed with the standard matrix product. $\mathbb{H}^1$ is indeed $\mathbb{R}^3$, endowed with the group law

$$(x_1,y_1,z_1)\cdot(x_2,y_2,z_2) = \left(x_1+x_2, y_1+y_2, z_1+z_2+\frac{1}{2}(x_1y_2-x_2y_1)\right). \tag{21.15}$$

This group law comes from the matrix product after the identification

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

The identity of the group is the element $(0,0,0)$ and the inverse element is given by the formula

$$(x,y,z)^{-1} = (-x,-y,-z).$$

A basis of the Lie algebra $\mathfrak{h}$ of $\mathbb{H}$ is $\{p_1,p_2,k\}$, where

$$p_1 = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad p_2 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}, \quad k = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \tag{21.16}$$

We have following commutation rules: $[p_1,p_2]=k$, $[p_1,k]=[p_2,k]=0$, hence $\mathbb{H}$ is a 2-step nilpotent Carnot group.

*Remark* 21.11. Notice that if one writes an element of the algebra as $xp_1+yp_2+zk$, one has that

$$\exp(xp_1+yp_2+zk) = \begin{pmatrix} 1 & x & z+\frac{1}{2}xy \\ 0 & 1 & y \\ 0 & 0 & 1 \end{pmatrix}. \tag{21.17}$$

Hence the coordinates $(x,y,z)$ are the coordinates on the Lie algebra related to the basis $\{p_1,p_2,k\}$, transported on the group via the exponential map. These are called coordinates of the *first type*. As we will see later, coordinates $(x,y,w)$, where $w = z+\frac{1}{2}xy$, are also useful.

The standard sub-Riemannian structure on $\mathbb{H}$ is the one having as generating family:

$$X_1(g) = gp_1, \qquad X_2(g) = gp_2.$$

With a straightforward computation one gets the following coordinate expression for the generating family:

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

that we already met several times in the previous chapters.

Let $L_g$ (resp. $R_g$) be the left (resp. right) translation on $\mathbb{H}$:

$$L_g(h) = gh, \qquad R_g(h) = hg, \qquad g,h \in \mathbb{H}.$$

**Exercise 21.12.** Prove that, up to a multiplicative constant, there exists a unique 3-form $dh_L$ on $\mathbb{H}$ which is left-invariant, i.e., such that $L_g^* dh_L = dh_L$. Show that $dh_L$ coincides (up to a multiplicative constant) with the Lebesgue measure $dx \wedge dy \wedge dz$. Prove the same for a right-invariant 3-form $dh_R$.

The left- and right-invariant forms built in the exercise above are the *left* and *right Haar* measures on $\mathbb{H}$. Since they coincide up to a multiplicative constant, the Heisenberg group is said to be *unimodular*. In the following we normalise the left and right Haar measures on the sub-Riemannian structure in such a way that

$$dh_L(X_1, X_2, [X_1, X_2]) = dh_R(X_1, X_2, [X_1, X_2]) = 1. \tag{21.18}$$

The 3-form obtained in this way on $\mathbb{H}$ coincides with the Lebesgue measure on $\mathbb{R}^3$ and in the following we call it simply the *Haar measure*

$$dh = dx \wedge dy \wedge dz. \tag{21.19}$$

As already remarked above, since we are on a Lie group this 3-form, this also coincides (up to a multiplicative constant) with Popp's measure.

**Exercise 21.13.** Prove that $dh$ coincides with Popp's measure.

**Exercise 21.14.** Prove that the two conditions (21.18) are invariant by change of the orthonormal frame.

### 21.2.2 The heat equation on the Heisenberg group

Given a volume form $\omega$ on $\mathbb{R}^3$, the sub-Riemannian Laplacian for the Heisenberg sub-Riemannian structure is given by the formula,

$$\triangle_H(\phi) = \big(X_1^2 + X_2^2 + \operatorname{div}_\omega(X_1)X_1 + \operatorname{div}_\omega(X_2)X_2\big)\phi. \tag{21.20}$$

If we take as volume the Haar volume $dh$, and using the fact that $X_1$ and $X_2$ are divergence free with respect to $dh$, we get for the sub-Riemannian Laplacian

$$\triangle_H(\phi) = (X_1)^2 + (X_2)^2 = (\partial_x - \frac{y}{2}\partial_z)^2 + (\partial_y + \frac{x}{2}\partial_z)^2. \tag{21.21}$$

The heat equation on the Heisenberg group is then

$$\partial_t \phi(x, y, z, t) = \triangle_H(\phi) = \big((\partial_x - \frac{y}{2}\partial_z)^2 + (\partial_y + \frac{x}{2}\partial_z)^2\big)\phi(x, y, z, t)$$

For this equation, we are looking for the heat kernel, namely a function $K_t(q, \bar{q})$ such that the solution to the Cauchy problem

$$\begin{cases} (\triangle_H - \partial_t)\phi = 0 \\ \phi(q, 0) = \phi_0(q) \in L^2(\mathbb{R}^3, dh) \end{cases} \tag{21.22}$$

can be expressed as

$$\phi(q, t) = \int_{\mathbb{R}^3} K_t(q, \bar{q})\phi_0(\bar{q})dh(\bar{q}), \qquad t > 0. \tag{21.23}$$

The existence of a heat kernel that is smooth, positive and symmetric is guaranteed by Theorem 21.8 since the Heisenberg group (as sub-Riemannian structure) is complete. Its explicit expression (as a matter of fact in a form of an inverse Fourier transform) is given by the following Theorem.

**Theorem 21.15.** *The heat kernel for the heat equation for the standard sub-Riemannian structure on the Heisenberg group, i.e., for the equation in* $\mathbb{R}^3$

$$\partial_t\phi(x,y,z,t) = \big((\partial_x - \frac{y}{2}\partial_z)^2 + (\partial_y + \frac{x}{2}\partial_z)^2\big)\phi(x,y,z,t),$$

*is given by the formula (here* $q = (x, y, z)$ *and "$\cdot$" is the group law* (21.15)*)*

$$K_t(q,\bar{q}) = P_t(q^{-1}\cdot\bar{q}),$$

*where*

$$P_t(x,y,z) = \frac{1}{(2\pi t)^2}\int_{\mathbb{R}} \frac{2\tau}{\sinh(2\tau)} \exp\left(-\frac{\tau(x^2+y^2)}{2t\tanh(2\tau)}\right)\cos\left(2\frac{z\tau}{t}\right)d\tau, \quad t>0. \tag{21.24}$$

Formula 21.24 is called the Gaveau-Hulanicki fundamential solution for the Heisenberg group. Notice that $P_t(q) = K_t(q, 0)$ hence it represents the evolution at time $t$ of an initial condition that at time zero is concentrated in the origin (a Dirac delta $\delta_0$).

$$P_t(q) = K_t(q,0) = \int_{\mathbb{R}^3} K_t(q,\bar{q})\delta_0(\bar{q})dh(\bar{q}).$$

### 21.2.3 Construction of the Gaveau-Hulanicki fundamental solution

The construction of the Gaveau-Hulanicki fundamental solution on the Heisenberg group was an important achievement of the end of the seventies (see the bibliographical note). Here we propose an elementary direct method divided in the following steps:

**STEP 1.** We look for a special form for $K_t(q, \bar{q})$ using the group law.

**STEP 2.** We make a change of variables in such a way that the coefficients of the heat equation depend only on one variable instead of two.

**STEP 3.** By using the Fourier transform in two variables, we transform the heat equation (that was a PDE in three spatial variables plus the time) in a heat equation with an harmonic potential in one variable plus the time.

**STEP 4.** We find the kernel for the heat equation with the harmonic potential, thanks to the Mehler formula for Hermite polynomials.

**STEP 5.** We come back to the original variables.

Let us make these steps one by one.

**STEP 1.** Due to invariance under the group law, we expect that $K_t(q, \bar{q}) = K_t(h\cdot q, h\cdot\bar{q})$ for every $h \in \mathbb{H}$. Taking $h = q^{-1}$ we then look for a heat kernel having the property $K_t(q, \bar{q}) = K_t(0, q^{-1}\bar{q})$. Hence setting $q = (x, y, z)$ and $\bar{q} = (\bar{x}, \bar{y}, \bar{z})$ we can write

$$K_t(q,\bar{q}) = P_t(q^{-1}\cdot\bar{q}) = P_t(\bar{x}-x, \bar{y}-y, \bar{z}-z) = P_t(x-\bar{x}, y-\bar{y}, z-\bar{z}), \tag{21.25}$$

for a suitable function $P_t(\cdot)$ called the *fundamental solution.* In the last equality we have used the symmetry of the heat kernel.

**STEP 2.** Let us make the change the variable $z \to w$, where

$$w = z + \frac{1}{2}xy,$$

(cf. Remark 21.11). In the new variables we have that the Haar measure is $dh = dx \wedge dy \wedge dw$. The generating family and the sub-Riemannian Laplacian becomes

$$X_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = \partial_x, \tag{21.26}$$

$$X_2 = \begin{pmatrix} 0 \\ 1 \\ x \end{pmatrix} = \partial_y + x\partial_w, \tag{21.27}$$

$$\triangle_H(\phi) = (X_1)^2 + (X_2)^2 = \partial_x^2 + (\partial_y + x\partial_w)^2. \tag{21.28}$$

The new coordinates are very useful since now the coefficients of the different terms in $\triangle_H$ depend only on one variable. We are then looking for the solution to the Cauchy problem

$$\begin{cases} \partial_t\varphi(x,y,w,t) = \triangle_H(\varphi(x,y,w,t)) = \big(\partial_x^2 + (\partial_y + x\partial_w)^2\big)\varphi(x,y,w,t) \\ \varphi(x,y,w,0) = \varphi_0(x,y,w) \in L^2(\mathbb{R}^3, dh) \end{cases} \tag{21.29}$$

where $\varphi(x,y,w,t) = \phi(x,y,w - \frac{1}{2}xy, t)$.

**STEP 3.** By making the Fourier transform in $y$ and $w$, we have $\partial_y \to i\mu$, $\partial_w \to i\nu$ and the Cauchy problem become

$$\begin{cases} \partial_t\hat{\varphi}(x,\mu,\nu,t) = \big(\partial_x^2 - (\mu + \nu x)^2\big)\hat{\varphi}(x,\mu,\nu,t) \\ \hat{\varphi}(x,\mu,\nu,0) = \hat{\varphi}_0(x,\mu,\nu). \end{cases} \tag{21.30}$$

By making the change of variable $x \to \theta$, where $\mu + \nu x = \nu\theta$, i.e., $\theta = x + \frac{\mu}{\nu}$ we get:

$$\begin{cases} \partial_t\bar{\varphi}^{\mu,\nu}(\theta,t) = \big(\partial_\theta^2 - \nu^2\theta^2\big)\bar{\varphi}^{\mu,\nu}(\theta,t) \\ \bar{\varphi}^{\mu,\nu}(\theta,0) = \bar{\varphi}_0^{\mu,\nu}(\theta), \end{cases} \tag{21.31}$$

where we set $\bar{\varphi}^{\mu,\nu}(\theta,t) := \hat{\varphi}(\theta - \frac{\mu}{\nu}, \mu, \nu, t)$, and $\bar{\varphi}_0^{\mu,\nu}(\theta) = \hat{\varphi}_0(\theta - \frac{\mu}{\nu}, \mu, \nu)$.

**STEP 4.**. We have the following

**Theorem 21.16.** *The solution of the Cauchy problem for the evolution of the heat in an harmonic potential, i.e.*

$$\begin{cases} \partial_t\psi(\theta,t) = \big(\partial_\theta^2 - \nu^2\theta^2\big)\psi(\theta,t) \\ \psi(\theta,0) = \psi_0(\theta) \in L^2(\mathbb{R}, d\theta) \end{cases} \tag{21.32}$$

*can be written in the form of a convolution kernel*

$$\psi(\theta,t) = \int_{\mathbb{R}} Q_t^\nu(\theta,\bar{\theta})\psi_0(\bar{\theta})d\bar{\theta}, \quad t > 0. \tag{21.33}$$

*where*

$$Q_t^\nu(\theta,\bar{\theta}) := \sqrt{\frac{\nu}{2\pi\sinh(2\nu t)}}\exp\left(-\frac{1}{2}\frac{\nu\cosh(2\nu t)}{\sinh(2\nu t)}(\theta^2 + \bar{\theta}^2) + \frac{\nu\theta\bar{\theta}}{\sinh(2\nu t)}\right), \tag{21.34}$$

*is the so-called Mehler Kernel.*

*Remark* 21.17. In the case $\nu = 0$ we interpret $Q_t^0(\theta, \bar{\theta})$ as

$$\lim_{\nu \to 0} Q_t^\nu(\theta, \bar{\theta}) = \frac{1}{\sqrt{4\pi t}} \exp\left(-\frac{(\theta - \bar{\theta})^2}{4t}\right). \tag{21.35}$$

*Proof.* For $\nu = 0$, equation (21.32) is the standard heat equation on $\mathbb{R}$ and the heat kernel is given by formula (21.35) (see for instance [Eva98]). In the rest of the proof we assume $\nu > 0$. The eigenvalues and the eigenfunctions of the operator $\partial_\theta^2 - \nu^2\theta^2$ on $\mathbb{R}$ are (see for instance [CTDL92])

$$\begin{aligned} E_j &= -2\nu(j + 1/2), \\ \Phi_j^\nu(\theta) &= \frac{1}{\sqrt{2^j j!}} \left(\frac{\nu}{\pi}\right)^{\frac{1}{4}} \exp(-\frac{\nu\theta^2}{2}) H_j(\sqrt{\nu}\,\theta), \end{aligned}$$

where $H_j$ are the Hermite polynomials $H_j(\theta) = (-1)^j \exp(\theta^2) \frac{d^j}{d\theta^j} \exp(-\theta^2)$. Since the operator $\partial_\theta^2 - \nu^2\theta^2$ is essentially self adjoint in $L^2(\mathbb{R})$, we have that $\{\Phi_j^\nu\}_{j\in\mathbb{N}}$ is an orthonormal frame of $L^2(\mathbb{R})$ and we can write

$$\psi(\theta, t) = \sum_{j\in\mathbb{N}} C_j(t)\Phi_j^\nu(\theta).$$

Using equation (21.32), we obtain that

$$C_j(t) = C_j(0)\exp(tE_j),$$

where $C_j(0) = \int_{\mathbb{R}} \Phi_j^\nu(\bar{\theta})\psi_0(\bar{\theta})\, d\bar{\theta}$. Hence

$$\psi(\theta, t) = \int_{\mathbb{R}} Q_t^\nu(\theta, \bar{\theta})\psi_0(\bar{\theta})\, d\bar{\theta},$$

where

$$Q_t^\nu(\theta, \bar{\theta}) = \sum_{j\in\mathbb{N}} \Phi_j^\nu(\theta)\Phi_j^\nu(\bar{\theta})\exp(tE_j).$$

After some algebraic manipulations and using the Mehler formula for Hermite polynomials (see for instance [CTDL92])

$$\sum_{j\in\mathbb{N}} \frac{H_j(\xi)H_j(\bar{\xi})}{2^j j!}(w)^j = (1 - w^2)^{-\frac{1}{2}} \exp\left(\frac{2\xi\bar{\xi}w - (\xi^2 + \bar{\xi}^2)w^2}{1 - w^2}\right), \quad \forall\, w \in (-1, 1),$$

with $\xi = \sqrt{\nu}\theta$, $\bar{\xi} = \sqrt{\nu}\bar{\theta}$, $w = \exp(-2\nu t)$, one gets formula (21.34). In the case $\nu < 0$ we get the same result. $\square$

Using Theorem 21.16 we can write the solution to (21.31) as

$$\bar{\varphi}^{\mu,\nu}(\theta, t) = \int_{\mathbb{R}} Q_t^\nu(\theta, \bar{\theta})\bar{\varphi}_0^{\mu,\nu}(\bar{\theta})d\bar{\theta}.$$

**An alternative construction of Mehler's kernel**

The construction of the Mehler kernel given above makes use of the Mehler formula that could appear a bit mysterious. In the following we give an alternative elementary construction based on a certain natural *ansatz*.

First observe that the following properties must hold:

P1. $Q^\nu_t(\theta,\bar\theta) = Q^\nu_t(-\theta,-\bar\theta)$. This because the of invariance of harmonic potential for $\theta \to -\theta$.

P2. $Q^\nu_t(\theta,\bar\theta) = Q^\nu_t(\bar\theta,\theta)$. This because the operator $\partial^2_\theta - \nu^2\theta^2$ is essentially self-adjoint and hence symmetric. As a consequence the operator $e^{t(\partial^2_\theta-\nu^2\theta^2)}$ is symmetric as well. It follows that $Q^\nu_t(\theta,\bar\theta)$ should be symmetric for the exchange of its arguments.

P3. $\lim_{t\to 0} Q^\nu_t(\theta,\bar\theta) \to \delta_{\bar\theta}(\theta)$. This because for $t \to 0$ formula (21.33) should reproduce the initial condition $\psi_0(\theta)$.

Now, if we want to find the solution of (21.32) in the form (21.33) we could make the following probabilistic interpretation. Fixed $t>0$, $Q^\nu_t(\theta,\bar\theta)$ is the density of probability of finding in $\theta$, a random particle with a quadratic rate of killing (that at time zero was in $\bar\theta$). Densities of probability are usually Gaussian. We then make the following ansatz:

A1 For fixed $t$ and $\bar\theta$, $Q^\nu_t(\theta,\bar\theta)$ is a Gaussian.

We then look for $Q^\nu_t(\theta,\bar\theta)$ in the following form

$$Q^\nu_t(\theta,\bar\theta) = \xi(t)e^{v(t)(\theta^2+\bar\theta^2)+w(t)\theta\bar\theta}. \tag{21.36}$$

Here for simplicity of notation we omit the dependence on $\nu$ of the functions $\xi, v, w$. We also assume $\nu > 0$, the opposite case being similar. The dependence on $\theta$ and $\bar\theta$ has been chosen to be the exponent of a quadratic form (as a consequence of ansatz A1) that has the required symmetries (as a consequence of properties P1 and P2). The time dependence of the functions $\xi, v, w$ will be obtained by equation (21.32) and by ansatz P3.

If we plug (21.36) in (21.32) we obtain

$$\begin{aligned}\dot\xi e^{v(\theta^2+\bar\theta^2)+w\theta\bar\theta} \quad &+ \quad \xi e^{v(\theta^2+\bar\theta^2)+w\theta\bar\theta}(\dot v(\theta^2+\bar\theta^2)+\dot w\theta\bar\theta) = \\ &\xi e^{v(\theta^2+\bar\theta^2)+w\theta\bar\theta}2v + \xi e^{v(\theta^2+\bar\theta^2)+w\theta\bar\theta}(2v\theta+w\bar\theta)^2 - \nu^2\theta^2\xi e^{v(\theta^2+\bar\theta^2)+w\theta\bar\theta}.\end{aligned}$$

Dividing by $\xi e^{v(\theta^2+\bar\theta^2)+w\theta\bar\theta}$ we get

$$\dot\xi/\xi + \dot v\theta^2 + \dot v\bar\theta^2 + \dot w\theta\bar\theta = 2v + (2v\theta+w\bar\theta)^2 - \nu^2\theta^2.$$

Separating the terms in $\theta^2, \bar\theta^2, \theta\bar\theta$ and those independent from $\theta$ and $\bar\theta$ we get

$$\dot v = 4v^2 - \nu^2 \tag{21.37}$$

$$\dot v = w^2 \tag{21.38}$$

$$\dot w = 4vw \tag{21.39}$$

$$\frac{\dot\xi}{\xi} = 2v \tag{21.40}$$

Let us use property P3 to fix the initial conditons for these differental equations. We must have $\lim_{t\to 0} Q_t^\nu(0,0) = +\infty$. Hence $\lim_{t\to 0} \xi(t) = +\infty$. For $\theta \neq 0$ we must have $\lim_{t\to 0} Q_t^\nu(\theta,0) = 0$. Hence $\lim_{t\to +\infty} v(t) = -\infty$. Finally $Q_t^\nu(\theta,\theta) = \xi(t)e^{t\theta^2(2v(t)+w(t))}$ should tend to infinity for $t$ that tends to zero. Now since $\xi(t)$ tends to infinity, $v(t)$ tends to minus infinity and $\xi(t)e^{t\theta^2 v(t)}$ tends to zero, we must have $\lim_{t\to 0} w(t) = +\infty$.

Equation (21.37) with initial condition $v(0) = -\infty$ has solution

$$v(t) = -\frac{\nu}{2}\coth(2\nu t). \tag{21.41}$$

From (21.37) and (21.38) we have $w(t) = \pm\sqrt{4v(t)^2 - \nu^2} = \pm\frac{\nu}{\sinh(2\nu t)}$. Here we should choose the sign + because of the condition $\lim_{t\to 0} w(t) = +\infty$. Finally

$$w(t) = \frac{\nu}{\sinh(2\nu t)}. \tag{21.42}$$

Equation (21.39) is automatically satisfied. Equation (21.40) with initial condition $\xi(t_0) = \xi_0$ has as solution

$$\xi(t;t_0,\xi_0) = \xi_0\sqrt{\frac{\sinh(2\nu t_0)}{\sinh(2\nu t)}}. \tag{21.43}$$

Thanks to property P3, we should have

$$\begin{aligned}1 &= \lim_{t\to 0}\int_{-\infty}^{+\infty} Q^\nu(\theta,0)d\theta = \lim_{t\to 0}\xi(t;t_0,\xi_0)\int_{-\infty}^{+\infty} e^{v(t)\theta^2} = \lim_{t\to 0}\xi(t;t_0,\xi_0)\sqrt{\frac{2\pi}{\nu\coth(2\nu t)}}\\ &= \lim_{t\to 0}\xi_0\sqrt{\frac{2\pi}{\nu}\frac{\sinh(2\nu t_0)}{\cosh(2\nu t)}} = \xi_0\sqrt{\frac{2\pi\sinh(2\nu t_0)}{\nu}}.\end{aligned}$$

Hence $\xi_0 = \sqrt{\frac{\nu}{2\pi}\frac{1}{\sinh 2\nu t_0}}$. Plugging this in (21.43) we get

$$\xi(t) = \sqrt{\frac{\nu}{2\pi}\frac{1}{\sinh(2\nu t)}}. \tag{21.44}$$

From (21.41), (21.42), (21.42) and (21.36) we get (21.34).

**STEP 5.** We now come back to the original variables step by step. We have

$$\hat{\varphi}(x,\mu,\nu,t) = \bar{\varphi}^{\mu,\nu}(x+\frac{\mu}{\nu},t) = \int_{\mathbb{R}} Q_t^\nu(x+\frac{\mu}{\nu},\bar{\theta})\bar{\varphi}_0^{\mu,\nu}(\bar{\theta})d\bar{\theta} = \int_{\mathbb{R}} Q_t^\nu(x+\frac{\mu}{\nu},\bar{x}+\frac{\mu}{\nu})\hat{\varphi}_0(\bar{x},\mu,\nu)d\bar{x}.$$

In the last equality we made the change of variable $\bar{\theta}\to\bar{x}$ with $\bar{\theta} = \bar{x}+\frac{\mu}{\nu}$ and we used the fact that $\hat{\varphi}_0^{\mu,\nu}(\bar{x}+\frac{\mu}{\nu}) = \hat{\varphi}_0(\bar{x},\mu,\nu)$.

Now, using the fact that $\hat{\varphi}_0(\bar{x},\mu,\nu)$ is the Fourier transform of the initial condition, i.e.

$$\hat{\varphi}_0(\bar{x},\mu,\nu) = \int_{\mathbb{R}}\int_{\mathbb{R}} \varphi_0(\bar{x},\bar{y},\bar{w})e^{-i\mu\bar{y}}e^{-i\nu\bar{w}}d\bar{y}\,d\bar{w},$$

and making the inverse Fourier transform we get

$$\varphi(x,y,w,t) = \frac{1}{(2\pi)^2}\int_{\mathbb{R}}\int_{\mathbb{R}} \hat{\varphi}(x,\mu,\nu,t)e^{i\mu y}e^{i\nu w}d\mu\, d\nu$$
$$= \int_{\mathbb{R}^3}\left(\frac{1}{(2\pi)^2}\int_{\mathbb{R}}\int_{\mathbb{R}} Q_t^{\nu}(x+\frac{\mu}{\nu},\bar{x}+\frac{\mu}{\nu})e^{i\mu(y-\bar{y})}e^{i\nu(w-\bar{w})}d\mu\, d\nu\right)\varphi_0(\bar{x},\bar{y},\bar{w})d\bar{x}\, d\bar{y}\, d\bar{w}.$$

Coming back to the variable $x, y, z$, we have

$$\phi(x,y,z,t) = \varphi(x,y,z+\frac{1}{2}xy,t) = \int_{\mathbb{R}^3} K_t(x,y,z,\bar{x},\bar{y},\bar{z})\phi_0(\bar{x},\bar{y},\bar{z})d\bar{x}\, d\bar{y}\, d\bar{z}.$$

where

$$K_t(x,y,z,\bar{x},\bar{y},\bar{z}) = \frac{1}{(2\pi)^2}\int_{\mathbb{R}}\int_{\mathbb{R}} Q_t^{\nu}(x+\frac{\mu}{\nu},\bar{x}+\frac{\mu}{\nu})e^{i\mu(y-\bar{y})}e^{i\nu(z-\bar{z}+\frac{1}{2}(xy-\bar{x}\bar{y}))}d\mu\, d\nu.$$

We have then (cf. (21.25))

$$P_t(x,y,z) = K_t(x,y,z;0,0,0) = \frac{1}{(2\pi)^2}\int_{\mathbb{R}}\int_{\mathbb{R}} Q_t^{\nu}(x+\frac{\mu}{\nu},\frac{\mu}{\nu})e^{i\mu y}e^{i\nu(z+\frac{1}{2}xy)}d\mu\, d\nu.$$

To simplify this formula and in particular to get rid of one of the two integrals let us set

$$A(\nu,t) = \sqrt{\frac{\nu}{2\pi\sinh(2\nu t)}},$$

and let us write explicitly from (21.34)

$$Q_t^{\nu}(x+\frac{\mu}{\nu},\frac{\mu}{\nu}) = A(\nu,t)\exp\left(-\frac{\nu}{2\tanh(2\nu t)}\left(\left(x+\frac{\mu}{\nu}\right)^2+\frac{\mu^2}{\nu^2}\right)+\frac{\nu\left(x+\frac{\mu}{\nu}\right)\frac{\mu}{\nu}}{\sinh(2\nu t)}\right)$$
$$= A(\nu,t)\exp\left(-\frac{\nu}{2\tanh(2\nu t)}x^2+(\mu\nu x+\mu^2)\alpha(\nu,t)\right),$$

where

$$\alpha(\nu,t) = \frac{1}{\nu}\left(\frac{1}{\sinh(2\nu t)}-\frac{1}{\tanh(2\nu t)}\right) = \frac{1}{\nu}\left(\frac{1-\cosh(2\nu t)}{\sinh(2\nu t)}\right) = -\frac{1}{\nu}\tanh(\nu t) < 0, \quad \forall t>0 \text{ and } \nu\in\mathbb{R}.$$

If we notice that $\mu\nu x+\mu^2 = \left(\mu+\frac{\nu}{2}x\right)^2-\frac{\nu^2}{4}x^2$, we can rewrite

$$Q_t^{\nu}(x+\frac{\mu}{\nu},\frac{\mu}{\nu}) = A(\nu,t)\exp\left(-\left(\frac{\nu}{2\tanh(2\nu t)}+\frac{\nu^2\alpha(\nu,t)}{4}\right)x^2\right)\exp\left(\alpha(\nu,t)\left(\mu+\frac{\nu}{2}x\right)^2\right).$$

Since

$$-\left(\frac{\nu}{2\tanh(2\nu t)}+\frac{\nu^2\alpha(\nu,t)}{4}\right) = -\frac{\nu}{4}\frac{1}{\tanh(\nu t)},$$

we have then

$$P_t(x,y,z) = \frac{1}{(2\pi)^2}\int_{\mathbb{R}}\int_{\mathbb{R}} A(\nu,t)\exp\left(-\frac{\nu}{4}\frac{1}{\tanh(\nu t)}x^2\right)\exp\left(\alpha(\nu,t)\left(\mu+\frac{\nu}{2}x\right)^2\right)e^{i\mu y}e^{i\nu(z+\frac{1}{2}xy)}d\mu\, d\nu.$$

Let us make the change of variable $\mu \to \omega = \mu + \frac{\nu}{2}x$ implying that $d\omega = d\mu$. We have

$$\begin{aligned} P_t(x,y,z) &= \frac{1}{(2\pi)^2} \int_{\mathbb{R}} \int_{\mathbb{R}} A(\nu,t) \exp\left(-\frac{\nu}{4}\frac{1}{\tanh(\nu t)}x^2\right) \exp\left(\alpha(\nu,t)\omega^2\right) e^{i(\omega - \frac{\nu}{2}x)y} e^{i\nu(z+\frac{1}{2}xy)} d\omega\, d\nu \\ &= \frac{1}{(2\pi)^2} \int_{\mathbb{R}} \int_{\mathbb{R}} A(\nu,t) \exp\left(-\frac{\nu}{4}\frac{1}{\tanh(\nu t)}x^2\right) e^{i\nu z} \underbrace{\exp\left(\alpha(\nu,t)\omega^2\right) e^{i\omega y}}_{\mathtt{T}_0} d\omega\, d\nu. \end{aligned}$$

Now the variable $\omega$ appear only in the term in $\mathtt{T}_0$. The integral in $d\omega$ can then be computed explicitly. Indeed being $\alpha(\nu,t) < 0$ we have that

$$\int_{\mathbb{R}} \exp\left(\alpha(\nu,t)\omega^2\right) e^{i\omega y} d\omega = \sqrt{\frac{\pi}{-\alpha(\nu,t)}} \exp\left(\frac{y^2}{4\alpha(\nu,t)}\right).$$

Hence

$$P_t(x,y,z) = \frac{1}{(2\pi)^2} \int_{\mathbb{R}} \underbrace{\sqrt{\frac{\pi}{-\alpha(\nu,t)}}}_{\mathtt{T}_1} \overbrace{\exp\left(\frac{y^2}{4\alpha(\nu,t)}\right)}^{\mathtt{T}_2} \underbrace{A(\nu,t)}_{\mathtt{T}_3} \overbrace{\exp\left(-\frac{\nu}{4}\frac{1}{\tanh(\nu t)}x^2\right)}^{\mathtt{T}_4} e^{i\nu z}\, d\nu.$$

Let us now compute

$$\begin{aligned} \mathtt{T}_1 \times \mathtt{T}_3 &= \sqrt{\frac{\pi}{-\alpha(\nu,t)}} A(\nu,t) = \sqrt{\frac{\nu\pi}{\tanh(\nu t)}} \sqrt{\frac{\nu}{2\pi \sinh(2\nu t)}} = \frac{\nu}{2\sinh(\nu t)} \\ \mathtt{T}_2 \times \mathtt{T}_4 &= \exp\left(\frac{y^2}{4(-)\frac{1}{\nu}\tanh(\nu t)}\right) \exp\left(-\frac{\nu}{4}\frac{1}{\tanh(\nu t)}x^2\right) = \exp\left(-\frac{\nu}{4}\frac{1}{\tanh(\nu t)}(x^2+y^2)\right) \end{aligned}$$

Hence

$$P_t(x,y,z) = \frac{1}{(2\pi)^2} \int_{\mathbb{R}} \frac{\nu}{2\sinh(\nu t)} \exp\left(-\frac{\nu}{4}\frac{1}{\tanh(\nu t)}(x^2+y^2)\right) e^{i\nu z}\, d\nu.$$

Finally we make the change of variables $\nu \to \tau = \frac{\nu t}{2}$ implying $d\nu = \frac{2}{t}d\tau$ and we get

$$P_t(x,y,z) = \frac{1}{(2\pi)^2} \int_{\mathbb{R}} \frac{\frac{2}{t}\tau}{2\sinh(2\tau)} \exp\left(-\frac{\frac{2}{t}\tau}{4}\frac{1}{\tanh(2\tau)}(x^2+y^2)\right) e^{i\frac{2}{t}\tau z} \frac{2}{t} d\tau.$$

Now, being the integrand an even function of $\tau$, we can replace $e^{i\frac{2}{t}\tau z}$ with $\cos(\frac{2}{t}\tau z)$ and we get

$$P_t(x,y,z) = \frac{1}{(2\pi t)^2} \int_{\mathbb{R}} \frac{2\tau}{\sinh(2\tau)} \exp\left(-\frac{\tau(x^2+y^2)}{2t\tanh(2\tau)}\right) \cos(2\frac{z\tau}{t}) d\tau. \tag{21.45}$$

**Exercise 21.18.** With the same technique explained above, find the heat kernel for the heat equation on the Grushin plane where the Laplacian is calculated with respect to Euclidean volume.

### 21.2.4 Small-time asymptotics for the Gaveau-Hulanicki fundamental solution

The integral representation (21.24) can be computed explicitly on the origin and on the $z$ axis. Let $q_0 = (0,0,0)$ and $q_z = (0,0,z)$. We have

$$K_t(q_0, q_0) = P_t(0,0,0) = \frac{1}{16t^2} \tag{21.46}$$

$$K_t(q_0, q_z) = P_t(0,0,z) = \frac{1}{8t^2\left(1+\cosh\left(\frac{\pi z}{t}\right)\right)} = \frac{1}{4t^2}\exp\left(-\frac{d^2(q_0,q_z)}{4t}\right) f_z(t) \tag{21.47}$$

In the last equality we have used the fact that for the Heisenberg group $d(q_0, q_z) = \sqrt{4\pi|z|}$. Here

$$f_z(t) := \frac{e^{\frac{2\pi z}{t}}}{\left(e^{\frac{\pi z}{t}}+1\right)^2}$$

is a function that for $z \neq 0$ is smooth as function of $t$ and satisfies $f_z(0) = 1$. A more detailed analysis (cf. also the bibliographical note) permits to get for every fixed $q = (x, y, z)$ such that $x^2 + y^2 \neq 0$

$$K_t(q_0, q) = P_t(x, y, z) = \frac{C + O(t)}{t^{3/2}} \exp\left(-\frac{d^2(q_0, q)}{4t}\right). \tag{21.48}$$

Notice that the asymptotics (21.46), (21.47), (21.48) are deeply different with respect to those in the Euclidean case. Indeed the heat kernel for the standard heat equation in $\mathbb{R}^n$ is given by the formula

$$K_t(q_0, q) = \frac{1}{(4\pi t)^{n/2}} \exp\left(-\frac{d_E(q_0, q)^2}{4t}\right). \tag{21.49}$$

Here $q_0, q \in \mathbb{R}^n$ and $d_E$ is the standard Euclidean norm. Comparing (21.49) with (21.46), (21.47), (21.48), one has the impression that the heat diffusion on the Heisenberg group at the origin and on the points on the $z$ axis, is similar to the one in an Euclidean space of dimension 4 (i.e., beside constants it has an asymptotics of the type $\frac{1}{t^2}\exp(-\frac{d^2(q_0,q)}{4t})$ for $t \to 0$). While on all the other points it is similar to the one in an Euclidean space of dimension 3, (i.e., beside constants it has an asymptotics of the type $\frac{1}{t^{(3/2)}}\exp(-\frac{d^2(q_0,q)}{4t})$ for $t \to 0$). Indeed the difference of asymptotics between the Heisenberg and the Euclidean case at the origin is related to the fact that the Hausdorff dimension of the Heisenberg group is 4, while its topological dimension is 3 (See Chapter 20). While the difference of asymptotics on the $z$ axis (without the origin) is related to the fact that these are points reached by a one parameter family of optimal geodesics starting from the origin and hence they are at the same time cut and conjugate points. For more details see the bibliographical note.

It is interesting to remark that on a Riemannian manifold of dimension $n$ the asymptotics are similar to the Euclidean ones for points close enough. Indeed for every $q$ close enough to $q_0$ we have $K_t(q_0, q) = \frac{C+O(t)}{(4\pi t)^{n/2}} \exp\left(-\frac{d^2(q_0,q)}{4t}\right)$ for some $C = C(q_0, q) > 0$ depending on the point and $C(q_0, q_0) = 1$. However if $q$ is a point that is in the cut locus from $q_0$ (situation that never occurs when $q$ is close enough to $q_0$) then $K_t(q_0, q) = \frac{C+O(t)}{t^m} \exp\left(-\frac{d^2(q_0,q)}{4t}\right)$, where $C > 0$ and $m \geq n/2$ are constants whose values depend on the structure of optimal geodesics starting from $q_0$ and arriving in a neighborhood of $q$.

## 21.3 Bibliographical note

The problem of existence of an intrinsic volume in sub-Riemannian geometry and hence of a Laplacian was first formulated by Brockett in [Bro82]. The problem was then studied by Strichartz in [Str86], Montgomery in [Mon02] who introduced the Popp measure, and in [ABGR09]. Concerning the uniqueness of an intrinsic volume see [ABB12, BNR17].

For the heat equation in Riemannian geometry, we refer to [Ros97] and references therein. For an elementary introduction in $\mathbb{R}^n$ we refer to the book of Evans [Eva98].

Theorem 21.8 has been proved in [Str86, Str89]. This result has been first proved in the Riemannian context in [Gaf54, Gaf55]. In [Str86, Str89] one finds also the proof of Theorem 21.10. For the proof of Theorem 21.7, see for instance [FOT94]. Hörmander theorem was proved in [Hör67]. Today there are alternative proofs based on stochastic analysis. See for instance [Hai11, CF10, CHLT15]. For a nice discussion concerning the Hörmander theorem see [Bra14].

The fundamental solution of the heat equation on the Heisenberg group (cf. Theorem 21.15) was obtained by Gaveau using Hamilton-Jacobi theory [Gav77] and by Hulanicki using non-commutative Fourier analysis [Hul76]. For this second method applied on other 3-dimensional Lee groups see also [ABGR09, BB09, Bon12]. The elementary method presented here, that uses the standard Fourier transform after a change of coordinates that make the sub-Laplacian depending only on one variable, is original.

The small time heat kernel estimates for the Heisenberg group (21.46), (21.47), (21.48) have been obtained in [Gav77]. For more general sub-Riemannian structures, small time heat kernel estimates on the diagonal (i.e., for $P_t(q,q)$) and their relation with the Hausdorff dimension were studied in [BA89, Ĺ92, BAL91a, BAL91b], see also [Bar13]. Small time heat kernel estimates out of the diagonal (i.e., for $P_t(q,q')$ with $q \neq q'$) and their relation with the sub-Riemannian distance were studied in [BA88] (out of the cut locus) and in [BBCN17, BBN12, BBNss] on the cut locus, adapting a technique due to Molchanov [Mol75].

# Appendix A

# Geometry of parametrized curves in Lagrangian Grassmannians (by Igor Zelenko)

The aim of this Appendix is to describe how to construct canonical bundles of moving frames and differential invariants for parametrized curves in Lagrangian Grassmannians, at least in the monotonic case. Such curves appear as Jacobi curves of sub-Riemannian extremals. Originally this construction was done in [ZL07, ZL09], where it uses the specifics of Lagrangian Grassmannian. In later works [DZ12, DZ13] a much more general theory for construction of canonical bundles of moving frames for parametrized or unparametrized curves in the so-called generalized flag varieties was developed, so that the problem which is discussed here can be considered as a particular case of this general theory. Although this was briefly discussed at the very end of [DZ12], the application of the theory of [DZ12, DZ13] to obtain the results of [ZL07, ZL09] were never written in detail and this is our goal here. We believe that this exposition gives a more conceptual point of view on the original results of [ZL07, ZL09] and especially clarifies the origin of the normalization conditions of the canonical bundles of moving frames there, which in fact boil down to a choice of a complement to a certain subspace of the symplectic Lie algebra. I would like to thank David Sykes and Chengbo Li for editing most of this text.

## A.1 Preliminaries

### Basics on moving frames and structure functions

Throughout this appendix $I$ denotes an interval of $\mathbb{R}$ and the parameter $t$ takes values in $I$. We first start with more elementary and naive point of view on moving frames. For a moment, by a moving frame in $\mathbb{R}^n$ we mean an $n$-tuple $E(t) = \big(e_1(t), \dots e_n(t)\big)$ of vectors such that $E(t)$ constitute a basis of $\mathbb{R}^n$ for every $t$ and it smoothly depends on $t \in I$. We can regard $E(t)$ as an $n \times n$ matrix with $i$th column equal to the column vector $e_i(t)$, or an element of the Lie group $\mathrm{GL}_n$. So, the moving frame $E(t)$ can be seen as a smooth curve in this Lie group.

The velocity $e_j'(t)$ of the $j$th vector $e_j(t)$ of the moving frame $E(t)$ can be decomposed into the linear combination with respect to the basis $E(t)$, i.e., there exist scalars $r_{ij}(t)$ such that

$$e'_j(t) = \sum_{i=1}^{n} r_{ij}(t)e_i(t), \quad 1 \le j \le n, \ t \in I \tag{A.1}$$

Let $R(t)$ be the $n \times n$ matrix with the $ij$th entry $r_{ij}(t)$. Then (A.1) is equivalent to

$$E'(t) = E(t)R(t). \tag{A.2}$$

The equation (A.2) is called the *structure equation of the moving frame* $E(t)$ and the matrix function $R(t)$, is called the *structure function of the moving frame* $E(t)$.

*Remark* A.1. Recall that the (left) Maurer-Cartan form $\Omega$ on a Lie group $G$ with the Lie algebra $\mathfrak{g}$ is the $\mathfrak{g}$-valued one-form such that for any $X \in T_aG, a \in G$

$$\Omega_a(X) := d(L_a^{-1})X, \tag{A.3}$$

where $L_a$ denotes the left translation by an element $a$ in $G$. For a matrix Lie group $\Omega_a(X) := a^{-1}X$, where in the right-hand side the matrix multiplication is used. Note that (A.2) can be written as

$$R(t) = E(t)^{-1}E'(t), \tag{A.4}$$

which is equivalent to

$$R(t) = \Omega_{E(t)}\big(E'(t)\big), \tag{A.5}$$

where $\Omega$ is the Maurer-Cartan form of $GL_n$, i.e., the structure function of the frame $E(t)$ is equal to the value of the Maurer-Cartan form at the velocity to the frame.

Now let $G$ be a Lie subgroup of $\mathrm{GL}_n$ with Lie algebra $\mathfrak{g}$. We will say that a moving frame $E(t)$ is *$G$-valued* if $E(t)$, considered as an $n \times n$- matrix, belongs to $G$. From (A.5) it follows that the structure function of a $G$-valued frame takes value in the Lie algebra $\mathfrak{g}$.

Two $G$-valued moving frames $E(t)$ and $\widetilde{E}(t)$ are called equivalent with respect to $G$, or *$G$-equivalent*, if there exists $A \in G$ such that $\tilde{e}_j(t) = Ae_j(t)$ for any $1 \le j \le n$ and $t \in I$ or, equivalently, in the matrix form

$$\widetilde{E}(t) = AE(t), \quad \forall t \in I. \tag{A.6}$$

The following simple lemma is fundamental for the applications of moving frames in geometry of curves:

**Lemma A.2.** *1. Two $G$-valued moving frames $E(t)$ and $\widetilde{E}(t)$ with the structure functions $R(t)$ and $\widetilde{R}(t)$, respectively are $G$-equivalent if and only if $R(t) \equiv \widetilde{R}(t)$ on $I$.*

*2. Given any function $R : I \to \mathfrak{g}$ there exists the unique, up to the action of $G$, $G$-valued moving frame with the structure function $R(t)$.*

*Proof.* The "only if" part of the first statement of the lemma is trivial, because if $\widetilde{E}(t) = AE(t)$, then

$$\widetilde{R}(t) = \big(\widetilde{E}(t)\big)^{-1}\widetilde{E}'(t) = \big(E(t)^{-1}A^{-1}\big)\big(AE'(t)\big) = E(t)^{-1}E'(t) = R(t).$$

For the "if" part take $t_0 \in I$ and let $A := \widetilde{E}(t_0)E(t_0)^{-1}$. Then clearly

$$\widetilde{E}(t_0) = AE(t_0). \tag{A.7}$$

Further, by the same arguments as in the previous part, the moving frame $AE(t)$ has the same structure function as $E(t)$ and therefore, by our assumptions, as $\widetilde{E}(t)$. In other words, the frames $AE(T)$ and $\widetilde{E}(t)$ satisfy the same system of linear ODEs. Since by (A.7) they meet the same initial conditions at $t_0$, we have (A.6) by the uniqueness theorem for linear ODEs.

The existence claim of the second statement of the lemma follows from the existence theorem for linear systems of ODEs, while the uniqueness part follows from the "if" part of the first statement.

□

### Applications to geometry of curves in Euclidean space

The previous lemma is the basis for application of moving frames and construction of the complete system of invariants for various types of curves with respect to the action of various groups. Perhaps, every Differential Geometry student quickly encounters the Frenet-Serret moving frame in the study of curves in Euclidean space up to a rigid motion[1]. Recall its construction: Assume that a curve $\gamma(t)$ in $\mathbb{R}^n$ is parametrized by an arc length and for simplicity [2] satisfies the following regularity assumption:

$$\operatorname{span}\{\gamma'(t), \dots \gamma^{(n)}(t)\} = \mathbb{R}^n \tag{A.8}$$

i.e., $\Big(\gamma'(t), \dots \gamma^{(n)}(t)\Big)$ is a moving frame. The Frenet-Serret moving frame is obtained from this frame by the Gram-Schmidt orthogonalization procedure. This is $O_n$-valued (or orthonormal) moving frame[3] uniquely assigned to each curve with the properties above and two curves can be transformed one to another by a rigid motion if and only of their Frenet-Serret frames are $O_n$-equivalent, which by Lemma A.2 is equivalent to the fact that the structure functions of their Frenet-Serret frames coincide. By constructions, these structure functions are $\mathfrak{so}_n$-valued (i.e., skew-symmetric) such that the following condition holds:

***The normalization condition for the Frenet-Serret frame:*** *The only possible nonzero entries below the diagonal are $(j+1, j)$-entry with $1 \leq j \leq n-1$.* These $n-1$ entries completely determine the structure function by skew-symmetricity and therefore, again by Lemma A.2, constitute constitute its complete system of invariants. The $(2,1)$-entry classically called the curvature of the curve, the $(3,2)$-entry is classically called the torsion, at least for $n = 3$, and for higher dimensions all other non-zero entries are called higher order curvatures.

Already in this classical example one encounters, at least implicitly, the notion of the curve of osculating flags, which will be important in the sequel. Recall that a *flag* in a vector space $V$ or a *filtration* of $V$ is a collection of nested subspaces of $V$. The flag is called *complete*, if each dimension between 0 and $\dim V$ appears exactly once in the collection of dimensions of the subspaces of the flag. To a curve $\gamma$, satisfying regularity assumption (A.8), one can assign the following curve of complete flags in $\mathbb{R}^n$:

$$0 \subset \operatorname{span}\{\gamma'(t)\} \subset \operatorname{span}\{\gamma'(t), \gamma''(t)\} \subset \ldots \subset \operatorname{span}\{\gamma^{(j)}(t)\}_{j=1}^{i} \subset \ldots \subset \operatorname{span}\{\gamma^{(j)}(t)\}_{j=1}^{n} = \mathbb{R}^n, \tag{A.9}$$

[1]Here by the rigid motion we mean the map $x \mapsto a + Ux$, where $U \in O_n$, while often one assumes that $U \in SO_n$

[2]Often, especially in the case of a modification of the problem mentioned in the previous footnote,one makes a weaker assumption that dim $\operatorname{span}\{\gamma'(t), \dots \gamma^{(n-1)}(t)\} = n-1$

[3]In the case of the weaker assumption of the previous footnote related to the problem mentioned in the first footnote one uses the Gram-Schmidt orthogonalization for the tuple of vectors $\{\gamma'(t), \dots \gamma^{(n-1)}(t)\}$ to construct $n-1$ unit and pairwise orthogonal vectors and then completes this to $\mathrm{SO}_n$-valued frame.

which is called the *curve of osculating flags*, associated with the curve $\gamma$. A moving frame $(v_1(t), \dots, v_n(t)$ is called *adapted* to the curve of osculating flags (A.9), if $\operatorname{span}\{v_j(t)\}_{j=1}^i = \operatorname{span}\{\gamma^{(j)}(t)\}_{j=1}^i$ for all $1 \le i \le n$. The Gram-Schmidt orthogonalization procedure is nothing but the procedure of construction of the orthonormal frame adapted to the curve of flag (A.9) and such that the $i$th vector of the frame points toward the same half-space of the $i$th subspaces of the flag (A.9) with respect to the $(i-1)$th subspace of this flag as the vector $\gamma^{(i)}(t)$. In this case there is exactly one such adapted frame.

In this example, one can see another important point for our exposition. The Frenet-Serret frame can be described in terms of its structure function without referring to the Gram-Schmidt orthogonalization: The Frenet-Serret frame of a curve $\gamma$ parametrized by an arc length is the only orthonormal frame such that the very first vector is $\gamma'(t)$ and the structure function satisfies the *normalization conditions* above with all nonzero entries being positive.

### More general point of view: homogeneous spaces and moving frames as lifts to the group, defining the equivalence

Equivalence problem for curves in $\mathbb{R}^n$ up to a rigid motion is a particular case of equivalence problem for curves in a homogeneous space. In more detail, given a Lie group $G$ and its closed subgroup $G^0$, the space $G/G^0$ of the left cosets of $G^0$ is a smooth manifold with the natural transitive action of $G$ induced by the left translation on $G$. The space $G/G^0$ is called a *homogeneous space* of the group $G$. Two curves in $G/G^0$ are called equivalent if there exist and element of $G$ sending one curve to another. In the case of curves in $\mathbb{R}^n$ up to a rigid motion, $G$ is the group of rigid motion, denoted by $\mathrm{AO}_n$ and $G^0$ can be taken as its subgroup preserving the origin of $\mathbb{R}^n$, i.e., the group of orthogonal transformations $O_n$, so that $\mathbb{R}^n \cong \mathrm{AO}_n/O_n$.

*Grassmannians* and, more generally, *flag varieties* provide another class of examples of homogeneous spaces. A flag variety is a set of flags of a vector space $V$ with fixed dimensions of subspaces in these flags. Fix one of the flags and let $G^0$ be the subgroup of $\mathrm{GL}(V)$ preserving this flag. Then the flag variety can be identified with $GL(V)/G^0$. A Grassmannian corresponds to a flag variety with flags consisting of a one subspace of a fixed dimension.

If $V$ has some additional structure, then one can distinguish special flags and consider proper subgroups of $GL(V)$ as the group $G$. For example, let $V$ be a $2m$-dimensional vector space ended with a symplectic, i.e., a non-degenerate skew-symmetric, form $\sigma$. Given a subspace $\Lambda$ of $V$ let $\Lambda^\angle$ be the *skew-orthogonal complement* of $\Lambda$ with respect to $\sigma$,

$$\Lambda^\angle = \{v \in V : \sigma(v, z) = 0 \ \forall z \in \Lambda\}$$

Then one can distinguish the following classes of subspaces of $V$: a subspace $\Lambda$ of $V$ is called *isotropic*, if $\Lambda \subset \Lambda^\angle$ or, equivalently. $\sigma|_\Lambda = 0$, a subspace $\Lambda$ is called *coisotropic*, if $\Lambda^\angle \subset \Lambda$, or equivalently, if $\Lambda^\angle$ is isotropic, and it is called *Lagrangian* if it is both isotropic and coisotropic, i.e., $\Lambda = \Lambda^\angle$. Since from nodegenericty of $\sigma$ subspaces $\Lambda$ and $\Lambda^\angle$ have complementary dimensions to $2m$, the dimension of isotropic spaces is not greater than $m$ and of the coisotropic subspaces is not smaller than $m$. Hence the dimension of Lagrangian subspaces is equal to $m$. The set $L(V)$ of all Lagrangian subspaces is called the *Lagrangian Grassmannian*. Let $\mathrm{Sp}(V)$ be the group of all symplectic transformations, i.e., of all $A \in \mathrm{GL}(V)$ preserving the form $\sigma$, $\sigma(Av, Aw) = \sigma(v, w)$. Since $\mathrm{Sp}(V)$ acts transitively on $L(W)$, the latter can be identified with $\mathrm{SP}(V)/G^0$, where $G^0$ is the subgroup of $\mathrm{Sp}(V)$ preserving one chosen Lagrangian subspace $\Lambda$.

Initially we defined a moving frame as a one-parametric family of bases in $\mathbb{R}^n$. However the most important thing in the equivalence problem was the structure function of the frame and in order to define it we essentially used the corresponding one-parametric family (i.e., a curve) of matrices. This motivates the following slightly more abstract definition of a moving frame: a *moving frame in a vector space* $W$ is a curve $E(t)$ of bijective linear operators on $W$, i.e., a curve in $GL(W)$. In the same way if $G$ is a Lie subgroup of $GL(W)$, then the $G$-valued moving frame is a curve in $G$. The structure function of $E(t)$ is defined by equation (A.4).

In order to relate this point of view to the previously defined notion of a moving frames in $\mathbb{R}^n$ it is just enough to choose some basis in $W$. This identifies $W$ with $\mathbb{R}^n$ and the operators of the frame with the matrices with respect to this basis. The one-parametric family of bases in $W$ can be obtained by taking the images of the chosen basis under the operators of the moving frame.

Since the structure functions of the moving frame are defined via the Maurer-Cartan, which is defined on any Lie group, we can go further and give the following definition of a moving frame in a homogeneous space of an abstract Lie group without any relation to the initial naive notion of the moving frame as a curve of bases:

**Definition A.3.** A moving frame over a curve $\gamma$ in a homogeneous space $G/G^0$ is a *smooth lift* of $\gamma$ to the Lie group $G$, i.e., a smooth curve $\Gamma$ in $G$ such that $\pi\big(\Gamma(t)\big) = \gamma(t)$ for every $t$, where $\pi : G \to G/G^0$ is the canonical projection. The structure function of the moving frame $\Gamma(t)$ is the $\mathfrak{g}$-valued function

$$C_\Gamma(t) := \Omega_{\Gamma(t)}(\Gamma'(t)) \tag{A.10}$$

where $\Omega$ is the left Maurer -Cartan form on the Lie group $G$.

This definition can be related to the previous one if one choose a faithful representation of $G$.

Note that in the case of curves in a flag variety $G/G^0$ the set of moving frames over this curve can be naturally related to the set of moving frames adapted to this curve of flags in the sense of the previous subsection.

For completeness, let us adjust this more general point of view on moving frames for curves in homogeneous spaces to the case of curves in Euclidean space. Here a standard representation of the affine group, although one can easily manage without any representation. For this identify $\mathbb{R}^n$ with the affine subspace of $\mathbb{R}^{n+1}$ by identifying a point $(x_1, \dots, x_n) \in \mathbb{R}^n$ with the point $(1, x_1, \dots x_n) \in \mathbb{R}^{n+1}$ and the group of rigid motions $\mathrm{AO}_n$ with the subgroup of $\mathrm{GL}_{n+1}$, consisting of the matrices of the form

$$\left(\begin{array}{c|c} 1 & 0 \\ \hline & \\ a & U \end{array}\right), \tag{A.11}$$

where $a \in \mathbb{R}^n$ and $U \in \mathrm{O}_n$. Here the matrix (A.11) corresponds to the rigid motion sending $x \in \mathbb{R}^n$ to $a + Ux$, because this matrix sends the vector $\begin{pmatrix} 1 \\ x \end{pmatrix}$, $x \in \mathbb{R}^n$ to $\begin{pmatrix} 1 \\ a + Ux \end{pmatrix}$. To a curve $\gamma$ we assign the moving frame, given by the following curve $\Gamma(t)$ in $\mathrm{AO}_n \subset \mathrm{GL}_{n+1}$

$$\Gamma(t) = \begin{pmatrix} 1 & 0 & \dots & 0 \\ \gamma(t) & e_1(t) & \dots & e_n(t) \end{pmatrix}, \tag{A.12}$$

where $\big(e_1(t), \dots, e_n(t)\big)$ is the Frenet-Serret frame of $\gamma(t)$. Since by constructions $\gamma'(t) = e_1(t)$, the structure function of this frame is

$$\left(\begin{array}{c|c} 0 & 0 \ \dots \ 0 \\ \hline \begin{matrix} 1 \\ 0 \\ \vdots \\ 0 \end{matrix} & R(t) \end{array}\right),$$

where $R(t)$ is the structure function of the Frenet-Serret frame of $\gamma$. In contrast to the Frenet-Serret frame, the frame (A.12) takes values in the entire group of rigid motions that defines the considered equivalence relation. Hence, equivalence of curves with the same structure functions of their $\mathrm{AO}_n$-valued frames, as given in (A.12), is obtained immediately from Lemma A.2, while using the Frenet-Serret frame, Lemma A.2 gives that the curve of velocities are equivalent by an orthogonal transformation and the equivalence of the original curves up to a rigid motion is obtained after integration only. Although technically these two arguments are equally elementary, the use of frames, which take values in the entire group defining the equivalence relation (or, equivalently, in the entire group of the homogeneous space) has an obvious conceptual advantage.

The moving frame $\Gamma(t)$ from (A.12) can be seen as the *canonical lift* of the original curve $\gamma$ from $\mathbb{R}^n = \mathrm{AO}_n/O_n$ to the group $\mathrm{AO}_n$. In view of Lemma A.2, the construction of such a canonical lift or a bundle of such lifts is the main idea for solving such type of equivalence problems. Canonical means that two curves $\gamma(t)$ and $\widetilde{\gamma}(t)$ in $G/G^0$ are equivalent via $g \in G$, i.e., $g.\gamma(t) = \widetilde{\gamma}(t)$ if and only if $g$ sends any canonical lift of $\gamma(t)$ to a canonical lift of $\tilde{\gamma}(t)$.

### Some general ideas on canonical bundles of moving frames: symmetries and normalization conditions

A way to choose the canonical moving frame or, equivalently, the canonical lift to the group $G$, is to specify certain restrictions, called *normalization conditions*, on its structure function In the case of curves in Euclidean space the necessity of such specification does not really emerge, because it follows automatically from the condition on the moving frame to be adapted to the osculating flag. Besides, in this case such a frame is unique.

For curves in general homogeneous spaces $G/G^0$ one cannot expect that there exists a unique canonical lift to the corresponding Lie group. The reason for this is that a curve $\gamma(t)$ in $G/G^0$ may have a nontrivial *non-effective* symmetry, i.e., an element $s$ of $G$, which is not the identity, but preserves each point of $\gamma$, $s.\gamma(t) = \gamma(t)$ for every $t$. The group of non-effective symmetries will be denoted by $\mathrm{Sym}_\gamma^{\mathrm{ne}}$. This group of symmetries is relevant to geometry of parametrized curves. In the case of unparametrized curves they should be replaced by a larger subgroup of $G$ consisting of symmetries preserving a distinguished point of $\gamma$ only and not necessary other points of $\gamma$.

If $s \in \mathrm{Sym}_\gamma^{\mathrm{ne}}$ and $\Gamma(t)$ is a lift of $\gamma(t)$ then $s\Gamma(t)$ is a lift of $\gamma(t)$ as well and there is not any preference of choosing $\Gamma(t)$ over $s\Gamma(t)$ and vice versa. In particular, they have the same structure functions. In other words the group $\mathrm{Sym}_\gamma^{\mathrm{ne}}$ encodes the minimal possible freedom of choice of a canonical frame for the curve $\gamma(t)$, so that if a normalization condition on the structure function is chosen, the set of all lifts satisfying this condition forms a fiber bundle over the curve $\gamma$ with the fibers of dimension not smaller than $\dim \mathrm{Sym}_\gamma^{\mathrm{ne}}$ and which is foliated by the lifts. This bundle will be referred as the *canonical bundle* corresponding to the chosen normalization conditions.

The reason why we do not insist that such bundle will have fibers of dimension exactly equal to $\dim \mathrm{Sym}^{\mathrm{ne}}{}_{\gamma}$ is that different curves in $G/G^0$ may have non-effective symmetry groups of different dimension. For example, generic curves may not have any nontrivial non-effective symmetry. So, if we insist to assign to each curve the bundle of moving frames of the minimal possible dimension, we may have too much branching in this construction with different normalization conditions for each branch. Instead, it is preferable to choose the widest possible classes of curves so that within each class the uniform normalization conditions for moving frames are used and so that for some distinguished curves in this class the symmetry group has the maximal possible dimension within the class, i.e., for these curves the canonical bundle is of the smallest possible dimension. The natural candidates for such curves are orbits of one- parametric subgroups of $G$ and in the case of Grassmannians and flag varieties these one-parametric subgroups are generated by nilpotent elements of the Lie algebra of $G$.

## A.2 Algebraic theory of curves in Grassmannians and flag varieties

It turns out that for curves in Grassmannians and, more generally, in flag varieties all main steps in the construction of canonical moving frames, including the choice of the class of curves, the description of maximal group of symmetries and of the normalization conditions can be maid purely algebraically. In this section we we will describe this algebraic theory.

### Tangent spaces to Grassmannians

Tangent spaces to Grassmannians and Lagrangian Grassmannians were already discussed in Chapter 14, see Propositions 14.2 and 14.13. Here for uniformity of presentation of the appendix we discuss this topic from the point of view of homogeneous spaces. The tangent space to a Lie group $G$ at a point $a$ can be identified with its Lie algebra $\mathfrak{g}$ via the Maurer-Cartan form as in (A.3). This immediately implies that the tangent space to a homogeneous space $G/G^0$ at a point $o$ can be identified with the quotient space $\mathfrak{g}/\mathfrak{g}^0$,

$$T_o(G/G^0) \cong \mathfrak{g}/\mathfrak{g}^0 \tag{A.13}$$

where $\mathfrak{g}$ and $\mathfrak{g}^0$ are the Lie algebras of $G$ and $G^0$, respectively.

Consider the case of the Grassmannian $\mathrm{Gr}_k(V)$ of $k$-dimensional subspaces in a vector space $V$. Fix a point $\Lambda \in \mathrm{Gr}_k(V)$. As already mentioned in subsection A.1, $\mathrm{Gr}_k(V)$ can be identified with $GL(V)/G^0$, where $G^0$ is the subgroup of $GL(V)$, preserving the subspace $\Lambda$. The Lie algebra $\mathfrak{gl}(V)$ of $\mathrm{GL}(V)$ is the algebra of all endomorphisms of $V$ and the Lie algebra $\mathfrak{g}^0$ of $G^0$ is the algebra of all endomorphisms of $V$, preserving $\Lambda$. It is easy to see that the quotient space $\mathfrak{gl}(V)/\mathfrak{g}^0$ can be canonically identified with the space $\mathrm{Hom}(\Lambda, V/\Lambda)$. Indeed, let $p : V \to V/\Lambda$ be the canonical projection to the quotient space. The assignment

$$A \in \mathfrak{gl}(V) \mapsto (p \circ A)|_{\Lambda} \in \mathrm{Hom}(\Lambda, V/\Lambda) \tag{A.14}$$

maps the endomorphisms from the same coset in $\mathfrak{gl}(V)/\mathfrak{g}^0$ to the same element of $\mathrm{Hom}(\Lambda, V/\Lambda)$ and is onto. Therefore by (A.13)

$$T_{\Lambda}\mathrm{Gr}_k(V) \cong \mathrm{Hom}(\Lambda, V/\Lambda). \tag{A.15}$$

This identification can be described also as follows: Take $Y \in T_\Lambda \mathrm{Gr}_k(V)$ and let $\Lambda(t)$ be a curve in $\mathrm{Gr}_k(V)$ such that $\Lambda(0) = \Lambda$ and $\Lambda'(0) = Y$. Given $l \in \Lambda$ take a smooth curve of vectors $\ell(t)$ satisfying the following two properties:

1. $\ell(0) = l$,
2. $\ell(t) \in \Lambda(t)$ for every $t$ close to 0.

**Exercise A.4.** Show that the coset of $\ell'(t)$ in $V/\Lambda$ is independent of the choice of the curve $\ell$ satisfying the properties (1) and (2) above.

Based on the previous exercise to $Y$ we can assign the element of $\mathrm{Hom}\big(\Lambda(t), W/\Lambda(t)\big)$ that sends $l \in \Lambda$ to the coset of $\ell'(0)$ in $V/\Lambda$, where the curve $\ell$ satisfies properties (1) and (2) above.

Now assume that $V$ is $2m$-dimensional and is endowed with a symplectic form $\sigma$. Describe the identification analogous to (A.15) for the Lagrange Grassmannian $L(V)$. Take $\Lambda \in L(V)$. Since $L(V) \subset \mathrm{Gr}_m(V)$, the space $T_\Lambda L(V)$ can be identified with a subspace of $\mathrm{Hom}(\Lambda, V/\Lambda)$. To describe this subspace, first note that $\sigma$ defines the identification of $V/\Lambda$ with the dual space $\Lambda^*$: the assignment

$$v \in V \mapsto (i_v\sigma)|_\Lambda \tag{A.16}$$

maps the elements from the same coset of $V/\Lambda$ to the same element of $\Lambda^*$ and is onto. Hence, it defines the required identification. Here $i_v\sigma$ defines the interior product of the vector $v$ and the form $\sigma$, that is $i_v\sigma(w) = \sigma(v, w)$. Second, since $L(W) \cong \mathrm{Sp}(V)/G^0$, where $G^0$ is the subgroup of $\mathrm{Sp}(V)$ preserving the space $\Lambda$, by (A.13) the space $T_\Lambda L(V)$ can be identified with $\mathfrak{sp}(V)/\mathfrak{h}$, where $\mathfrak{sp}(V)$ is the Lie algebra of $\mathrm{Sp}(V)$ and it consists of $A \in \mathfrak{gl}(V)$ such that

$$\sigma(Av, w) = \sigma(Aw, v), \tag{A.17}$$

i.e., the bilinear form $\sigma(A\cdot, \cdot)$ is symmetric, and $\mathfrak{g}^0$ is the Lie algebra of $G^0$. So, in (A.14) we have to take $A \in \mathfrak{sp}(V)$. From (A.16) and (A.17) it follows that $(p \circ A)|_\Lambda$ considered as a map from $\Lambda$ to $\Lambda^*$ is self-adjoint. Besides, any self-adjoint map from $\Lambda$ to $\Lambda^*$ can be obtained in this way from some $A \in \mathfrak{sp}(V)$. The space of self-adjoint maps from $\Lambda$ to $\Lambda^*$ can be identified with the space $\mathrm{Quad}(\Lambda)$ on $\Lambda$,

$$T_\Lambda L(V) \cong \mathrm{Quad}(\Lambda). \tag{A.18}$$

Similarly to the case of the Grassmannian, if $Y \in T_\Lambda L(V)$, $\Lambda(t)$ is a curve in $L(V)$ such that $\Lambda(0) = \Lambda$ and $\Lambda'(0) = Y$, and a smooth curve of vectors $\ell(t)$ satisfies the properties (1) and (2) above, then the quadratic form on $\Lambda$ corresponding to $Y$ is the form sending $l$ to $\sigma(\ell'(0), l)$.

We say that a curve $\Lambda(t)$ in Lagrange Grassmannians is *monotonically nondecreasing* if it velocity $\dfrac{d}{dt}\Lambda(t)$ is non-negative quadratic form for every $t$.

### Osculating flags and symbols of curves in Grassmannians

The goal of this subsection is to distinguish the classes of curves in Grassmannians for which the uniform construction of canonical moving frames can be made. For this, first to a curve $\Lambda(t)$ in the Grassmannian $\mathrm{Gr}_k(V)$ we assign a special curve of flags, called the *curve of osculating flags*. Denote by $C(\Lambda)$ the canonical bundle over the curve $\Lambda$: The fiber of $C(\Lambda)$ over the point $\Lambda(t)$ is the vector space $\Lambda(t)$. Let $\Gamma(\Lambda)$ be the space of all sections of $C(\Lambda)$ . Set $\Lambda^{(0)}(t) := \Lambda(t)$ and define inductively

$$\Lambda^{(-j)}(t) = \text{span}\left\{\frac{d^k}{dt^k}\ell(t) : \ell \in \Gamma(\Lambda), 0 \le k \le j\right\}$$

for $j \ge 0$. The space $\Lambda^{(-j)}(t)$, $j > 0$, is called the *jth extension* or the *jth osculating subspace* of the curve $\Lambda$ at point $t$. The usage of negative indices here is in fact natural because, as we will see later, it is in accordance with the order of invariants of the curve (i.e., the order of jet of a curve on which an invariant depends) and also with the natural filtration of the algebra of infinitesimal symmetries, given by stabilizers of jets of a curve of each order (and indexed by this jet order).

Further, given a subspace $L$ in $V$ denote by $L^\perp$ the annihilator of $L$ in the dual space $V^*$:

$$L^\perp = \{p \in V^* : p(v) = 0, \forall\, v \in L\}.$$

Set

$$\Lambda^{(j)}(t) = \left(\left(\Lambda(t)^\perp\right)^{(-j)}\right)^\perp, \quad j \ge 0. \tag{A.19}$$

The subspace $\Lambda^{(j)}(t)$, $j > 0$, is called the *jth contraction* of the curve $\Lambda$ at point $t$. Clearly, $\Lambda^{(j)}(t) \subseteq \Lambda^{(j-1)}(t)$. The flag (the filtration)

$$\dots \Lambda^{(2)}(t) \subseteq \Lambda^{(1)}(t) \subseteq \Lambda^{(0)} \subseteq \Lambda^{(-1)}(t) \subseteq \Lambda^{(-2)}(t) \subseteq \dots \tag{A.20}$$

is called the *osculating flag (filtration) of the curve* $\Lambda$ *at point* $t$. By construction, the curves in Grassmannians (Lagrangian Grassmannian) are $\mathrm{GL}(V)$-equivalent ($\mathrm{Sp}(V)$-equivalent) if and only if the curves of their osculating flags are $\mathrm{GL}(V)$-equivalent ($\mathrm{Sp}(V)$ equivalent).

A flag $\{X^j\}_{j\in\mathbb{Z}}$ in a symplectic space $X$ with $X^j \subset X^{j-1}$ will be called *symplectic* if all subspaces $X^j$ with $j > 0$ are coisotropic and $X^{-j} = (X^j)^\angle$. Then all subspaces $X^j$ with $j < 0$ are isotropic. If $\Lambda(t)$ is a curve in a Lagrangian Grassmannian, then the flag (A.20) is symplectic. Indeed, as $\Lambda(t) \subset \Lambda^{(j)}(t)$ for $j < 0$ and $\Lambda(t)$ is Lagrangian, we have that

$$\Lambda^{(-j)} = \Lambda^{(j)}(t)^\angle \subset \Lambda(t) \subset \Lambda^{(j)}(t),$$

which implies that the subspaces $\Lambda(t) \subset \Lambda^{(j)}(t)$ are isotropic. Further, from (A.16) and (A.19) it follows that

$$\Lambda^{(j)}(t) = \left(\Lambda^{(-j)}(t)\right)^\angle. \tag{A.21}$$

The curve $\Lambda(t)$ is called *equiregular* if for every $j > 0$ the dimension of $\Lambda^{(j)}(t)$ is constant. As the integer-valued function $\dim \Lambda^{(j)}(t)$ is lower semi-continuous, it is locally constant on an open dense set of $I$, which will imply that for a generic $t$ the curve $\Lambda$ is equiregular in a neighborhood of $t$. So, from now on we will assume that the curve $\Lambda(t)$ is equiregular. For an equiregular curve passing to the osculating flag, we get a curve in a flag variety, i.e., the equivalence of curves in Grassmannians (Lagrangian Grassmannians) is reduced to the equivalence of the osculating curves in the corresponding flag varieties.

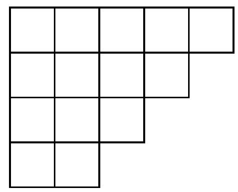
*Remark* A.5. For an equiregular curve $\Lambda(t)$ the subspaces $\Lambda^{(j)}(t)$ can be described by means of the identification of tangent vectors to Grassmannians with certain linear maps as in subsection A.2. Namely, $\Lambda^{(j-1)}(t)$ is the preimage under the canonical projection from $V$ to $V/\Lambda^{(j)}(t)$ of $\mathrm{Im}\frac{d}{dt}\Lambda^{(j)}(t)$. For an equiregular curve $\Lambda(t)$ we also have

$$\Lambda^{(j)}(t) = \mathrm{Ker}\frac{d}{dt}\Lambda^{(j-1)}(t), \quad j > 0$$

**Exercise A.6.** Prove that if the curve $\Lambda$ is equiregular, then

$$\dim \Lambda^{(j-1)}(t) - \dim \Lambda^{(j)}(t) \leq \dim \Lambda^{(j)}(t) - \dim \Lambda^{(j+1)}(t), j < 0.$$

Recall that a Young diagram is a finite collection of boxes, arranged in columns (equivalently, rows) with the column (rows) lengths in non-increasing order , aligned from the top and the left, as shown in the example below:



**Definition A.7.** The Young diagram $D$ such that the number of boxes in the $-j$th column, $j < 0$, of $D$ is equal to $\dim \Lambda^{(j)} - \dim \Lambda^{(j+1)}$ is called the Young diagram $D$ of the curve $\Lambda(t)$ in Grassmannian

This notion is especially important for monotonic curves in Lagrangian Grassmannians as shown in Proposition A.22.

**Example A.8.** The curve in Lagrangian Grassmannian $\Lambda$ is regular in the sense of Definition 14.18 if and only if $\Lambda^{(-1)} = V$. In this case the Young diagram consists of one row with the number of boxes equal to $\frac{1}{2}\dim V$. $\Box$.

Let

$$V_j(t) := \Lambda^{(j)}(t)/\Lambda^{(j+1)}(t) \tag{A.22}$$

and

$$\mathrm{gr}V(t) := \bigoplus_{j\in\mathbb{Z}} V_j(t) \tag{A.23}$$

be the graded space, associated with the filtration (A.20). By constructions, for any $j \in \mathbb{Z}$ we have the following inclusion

$$(\Lambda^{(j)})^{(-1)}(t) \subseteq \Lambda^{(j-1)}(t). \tag{A.24}$$

Hence, the velocity $\frac{d}{dt}\Lambda^{(j)}(t)$ of the curve $\Lambda^{(j)}$ at $t$, which is the map from $\Lambda^{(j)}(t)$ to $V/\Lambda^{(j)}(t)$ factors through the map from $V_j(t)$ to $V_{j-1}(t)$. Thus, the velocity of the curve of osculating flags (A.20) at $t$ factors through the endomorphism $\delta_t$ of the graded space $\mathrm{gr}V(t)$, sending $V_j(t)$ to $V_{j-1}(t)$ for any $j \in \mathbb{Z}$, i.e., the degree $-1$ endomorphism of the graded space $\mathrm{gr}V(t)$. This endomorphism is called the *symbol* of the curve $\Lambda$ at $t$ in the Grassmannian $Gr_k(V)$.

*Remark* A.9. By constructions $\delta_t : V_j(t) \to V_{j-1}(t)$ is injective for $j \geq 0$ and surjective for $j \leq 0$.

The natural equivalence relation on the space of endomorphisms of graded vector spaces is defined via conjugation: two endomorphisms $\delta$ and $\tilde{\delta}$ acting on graded vector spaces $X = \bigoplus_{j\in\mathbb{Z}} X_j$ and $\widetilde{X} = \bigoplus_{j\in\mathbb{Z}} \tilde{X}_j$, respectively, are called equivalent, if there exists an isomorphism $Q : X \to \tilde{X}$, preserving the grading, i.e., such that $Q(X_j) = \widetilde{X}_j$, and conjugating $\delta$ with $\tilde{\delta}$, i.e., such that

$$Q\delta = \tilde{\delta}Q. \tag{A.25}$$

So, it is in fact more correct to call the symbol of the curve $\Lambda$ at $t$ the equivalence class of $\delta_t$ in the set of degree $-1$ endomorphisms of graded spaces instead of a single degree $-1$ endomorphisms $\delta_t$. Also note that if $\tilde{X} = X$, then $\delta$ and $\tilde{\delta}$ are equivalent if and only if they lie in the same orbit under adjoint action of the isomorphisms of $X$ preserving the grading on $\mathfrak{gl}(X)$.

In the case of a curve in Lagrangian Grassmannians $L(V)$ the symplectic form $\sigma$ on $V$ induces the symplectic form $\sigma_t$ on each graded space $\mathrm{gr}V(t)$ as follows: if $\bar{x} \in V_j(t)$ and $\bar{y} \in V_{\tilde{j}}(t)$ with $j + \tilde{j} = 1$, then $\sigma_t(\bar{x}, \bar{y}) := w(x, y)$, where $x$ and $y$ are representatives of $\bar{x}$ and $\bar{y}$ in $\Lambda^{(j)}(t)$ and $\Lambda^{(\tilde{j})}(t)$ respectively; if $j + \tilde{j} \neq 1$, then $\sigma_t(\bar{x}, \bar{y}) = 0$. From (A.29) below it will follow that the symbol $\delta_t$ is not only an endomorphism of $\mathrm{gr}V(t)$ but also an element of the symplectic algebra $\mathfrak{sp}\big(\mathrm{gr}\, V(t)\big)$.

We say that a graded space $X = \bigoplus_{j \in \mathbb{Z}} X_j$ with a symplectic form $\sigma$ is a *symplectic graded space*, if the flag $\{X^j\}_{j \in \mathbb{Z}}$, where $X^j := \bigoplus_{i \geq j} X_i$, is a symplectic flag and after the identification of $X^j/X^{j+1}$ with $X_j$ the symplectic form induced by $\sigma$ on the graded space $X = \bigoplus_{j \in \mathbb{Z}} X^j/X^{j-1}$ coincides with $\sigma$. Equivalently, it means that the spaces $X_j \oplus X_{\bar{j}}$ with $j + \bar{j} \neq 1$ are isotropic with respect to $\sigma$. To define the equivalence relation on the space of endomorphisms of symplectic graded spaces one has to require that the conjugating isomorphism $Q$, as in (A.25), preserves the symplectic form.

Note that the notion of symbol as above can be defined not only for a curve in a Grassmannian (a Lagrangian Grassmannian) via its curve of osculating flags but for any curve of flags (symplectic flags) $\{\Lambda^j(t)\}_{j \in \mathbb{Z}}$ such that

$$\Lambda^j(t) \subseteq \Lambda^{j-1}(t), \quad (\Lambda^j)^{(-1)}(t) \subseteq \Lambda^{j-1}(t). \tag{A.26}$$

For this, spaces $\Lambda^{(j)}(t)$ in all previous formulas should be replaced by $\Lambda^j(t)$. So, the subsequent theory will be developed to the more general case of such curves of flags, which will be called *curves of flags compatible with osculation*.

Fix a flag (a symplectic flag) $\{V^j\}_{j \in Z}$ with $V^{j+1} \subset V^j$ in $V$ and the grading (symplectic grading) of $V$,

$$V = \bigoplus_{j \in \mathbb{Z}} V_j \tag{A.27}$$

such that $V^j = V^{j+1} \oplus V_j$. Recall that an endomorphism $A$ of the graded space $V$ is of degree $k$ if

$$A(V_j) \subset V_{j+k}.$$

Further, let $G$ denote either $\mathrm{GL}(V)$ or $\mathrm{Sp}(V)$ and by $G^0$ the subgroup of $G$ preserving the chosen flag $\{V^j\}_{j \in Z}$. The grading on $V$ induces the grading on the Lie algebra $\mathfrak{g}$:

$$\mathfrak{g} = \bigoplus_{k \in \mathbb{Z}} \mathfrak{g}_k, \tag{A.28}$$

where $\mathfrak{g}_k$ is the space of degree $k$ endomorphism of $V$, belonging to $\mathfrak{g}$. Given $a \in \mathfrak{g}$ we will denote by $a_k$ its degree $k$ component, i.e., the $\mathfrak{g}_k$-component of $A$ with respect to the splitting (A.28).

If $\Gamma(t)$ is a smooth lift of curve of flags $\{\Lambda^j(t)\}_{j \in \mathbb{Z}}$ compatible with osculation from $G/G^0$ to $G$, then by constructions

$$\delta_t := \big(\Gamma'(t)\Gamma(t)^{-1}\big)_{-1}. \tag{A.29}$$

**Definition A.10.** We say that a curve of flags (symplectic flags) $\{\Lambda^j(t)\}_{j\in\mathbb{Z}}$ compatible with osculation is of *constant type* if for all $t$ symbols $\delta_t$ belong to the same equivalence class. If $\delta$ is a degree $-1$ element of a graded space (a symplectic graded space), then we say that a curve of flags $\{\Lambda^j(t)\}_{j\in\mathbb{Z}}$ compatible with osculation is of *constant type* $\delta$ if for all $t$ the symbol $\delta_t$ is equivalent to $\delta$.

**Lemma A.11.** *The curve of flags $\{\Lambda^j(t)\}_{j\in\mathbb{Z}}$ is of constant type $\delta$ if and only if for any smooth lift $\Gamma(t)$ of the curve the degree $-1$ component of the structure function $C_\Gamma$ of $\Gamma$ lies in the orbit of $\delta$ under the adjoint action of $G^0$ on $\mathfrak{g}$.*

*Proof.* Indeed, from (A.29) it follows that

$$\Big(C_\Gamma(t)\Big)_{-1} = \Big(\Omega_{\Gamma(t)}\big(\Gamma'(t)\big)\Big)_{-1} = \Big(\Gamma(t)^{-1}\Gamma'(t)\Big)_{-1} = \Big(\Gamma(t)^{-1}\delta_t\Gamma(t)\Big)_{-1}. \tag{A.30}$$

Therefore, $\delta_t$ is equivalent to $\delta$ if and only if $\Big(C_\Gamma(t)\Big)_{-1}$ lies in the orbit of $\delta$ under the adjoint action of $G^0$ on $\mathfrak{g}$. $\square$

The set of all equivalence classes of degree $-1$ endomorphisms of a graded space (a symplectic graded space) is finite and all equivalence classes in this case are explicitly described in [DZ12]. The finiteness of these equivalence classes implies that a curve of flags (symplectic flags) compatible with osculation is of constant type in a neighborhood of its generic point. Moreover, as it will be shown for completeness later, any equiregular monotonically nondecreasing curve in a Lagrangian Grassmannian is of constant type; the space of equivalence classes of symbols of such curves is in fact in one-to-one correspondence with the tuples $\{\dim \Lambda^{(j)}\}_{j\leq 0}$ or , equivalently, with the set of all Young diagrams (see Proposition A.22 below). Note also that the finiteness of the set of equivalence classes follows in fact from more general result of E.B. Vinberg [Vin76] on finiteness of orbits of degree $-1$ elements of a graded semisimple Lie algebra under the adjoint action of the group of automorphisms of this graded Lie algebra.

### Flat curves of constant type and their symmetries

Fix $\delta \in \mathfrak{g}_{-1}$. The ultimate goal of the entire section is to describe the unified construction of canonical bundle of moving frames for all curves of flags (symplectic flags) of constant type $\delta$.

According to the general discussions at the end of the previous section, first we need to distinguish the most symmetric curves within this class. The natural candidate is an orbit under the action of the one-parametric group $\exp(t\delta)$ on the corresponding flag variety, for example, the curve of flags $t \to \{\exp(t\delta)V^j\}_{j\in\mathbb{Z}}$. Such curve is called the *flat curve* of constant type $\delta$ and will be denoted by $F(\delta)$. As we will see later, this curve is indeed the right candidate for the most symmetric curve in the considered class. The Lie algebra of the symmetry group of $F_\delta$ has an explicit algebraic description. Since for a parametrized curves the group of non-effective symmetries are important, in this subsection we will focus on the corresponding Lie algebra, referred as the algebra of infinitesimal non-effective symmetries.

Let

$$\mathfrak{u}_0(\delta) := \{x \in \mathfrak{g}_0 : [x,\delta] = 0\} \tag{A.31}$$

and define recursively

$$\mathfrak{u}_k(\delta) := \{x \in \mathfrak{g}_k : [x,\delta] \in \mathfrak{u}_{k-1}(\delta)\}, \quad k > 0. \tag{A.32}$$

**Exercise A.12.** Show that

$$\mathfrak{u}(\delta) := \bigoplus_{k\geq 0} \mathfrak{u}_k(\delta) \tag{A.33}$$

is a subalgebra of $\mathfrak{g}$.

The algebra $\mathfrak{u}(\delta)$ is called the *universal algebraic prolongation of the symbol* $\delta$, and its degree $k$ homogeneous component $\mathfrak{u}_k(\delta)$ is called the *kth algebraic prolongation of the symbol* $\delta$.

*Remark* A.13. Let

$$U_0(\delta) = \{A \in G : \operatorname{Ad} A\,\delta = \delta \text{ and } A(V_j) = V_j\ \forall j\} \tag{A.34}$$

$$U^0(\delta) = \{A \in G : \big(\operatorname{Ad} A\,\delta\big)_{-1} = \delta \text{ and } A(V^j) = V^j\ \forall j\} \tag{A.35}$$

Then the Lie algebra of $U_0(\delta)$ is equal to $\mathfrak{u}_0(\delta)$ and the Lie algebra of $U^0(\delta)$ is equal to $\mathfrak{u}_0 \oplus \bigoplus_{k>0} \mathfrak{g}_k$.

**Theorem A.14.** *The algebra of infinitesimal non-effective symmetries* $\mathrm{sym}^{\mathrm{ne}}_{F(\delta)}$ *of the flat curve* $F(\delta)$ *of constant type* $\delta$ *is equal to the algebra* $\mathfrak{u}(\delta)$.

*Proof.* By definition, $y \in \mathrm{sym}^{\mathrm{ne}}_{F(\delta)}$ if and only if $\exp(sy) \in \mathrm{Sym}^{\mathrm{ne}}_{F(\delta)}$ for every $s$ sufficiently close to 0. The latter means that

$$\exp(sy).F_\delta(t) = F_\delta(t) \tag{A.36}$$

for every $s$ sufficiently close to 0 and every $t \in \mathbb{R}$ or, equivalently

$$\exp(sy) \circ \exp(t\delta)V^j = \exp(t\delta)V^j \Leftrightarrow \exp(-t\delta) \circ \exp(sy) \circ \exp(t\delta)V^j = V^j$$

for every $s$ sufficiently close to 0, every $t \in \mathbb{R}$ and $j \in \mathbb{Z}$, which in turn equivalently can be written as

$$\exp(-t\delta) \circ \exp(sy) \circ \exp(t\delta) \in G^0, \tag{A.37}$$

where, as before, $G^0$ denotes the subgroup of $G$ consisting of all element preserving the flag $\{V^j\}_{j\in\mathbb{Z}}$.

Differentiating (A.37) with respect to $s$ at $s = 0$, we get

$$\operatorname{Ad}\Big(\exp(-t\delta)\Big)y \in \mathfrak{g}^0 \tag{A.38}$$

where $\mathfrak{g}^0$ is the Lie algebra of $G^0$. Pass to the Taylor series of the left-hand side of (A.38),

$$\operatorname{Ad}\Big(\exp(-t\delta)\Big)y = \sum_{k=0}^{\infty} \frac{t^k}{k!}\big(\operatorname{ad}(-\delta)\big)^k y, \tag{A.39}$$

where the sum is actually finite for every element $y$. Then we get that (A.38) is equivalent to

$$\big(\operatorname{ad}\delta\big)^k y \in \mathfrak{g}^0, \quad k \geq 0. \tag{A.40}$$

Obviously, $\mathfrak{g}^0 = \bigoplus_{k\geq 0} \mathfrak{g}_k$. So, $y$ can be represented as $y = \sum_{k\geq 0} y_k$ with $y_k \in \mathfrak{g}_k$. Since $\delta$ has degree $-1$ and all elements of $\mathfrak{g}^0$ are of nonnegative degree, from (A.40) with $k = 1$ it follows that $[\delta, y_0] = 0$, i.e. by (A.31) we have $y_0 \in \mathfrak{u}_0(\delta)$. Further, from (A.40) with $k = 2$ it follows that $\operatorname{ad}\delta^2 y_1 = 0$, which yields that $[\delta, y_1] \in u_0(\delta)$ and so by (A.32) we have $y_1 \in \mathfrak{u}_1(\delta)$. In this way by induction in $k$ one proves that (A.40) implies that $y_k \in \mathfrak{u}_k(\delta)$ for all $k \geq 0$, i.e., that $\mathrm{sym}^{\mathrm{ne}}_{F(\delta)} \subset \mathfrak{u}(\delta)$. Finally, the opposite inclusion is valid because (A.40) implies (A.36). ☐

*Remark* A.15. By analogy, one can describe the whole algebra of infinitesimal symmetries $\mathrm{sym}_{F(\delta)}$ of a curve $F_\delta$ considered as an unparametrized curve (a one-dimensional submanifold of the corresponding flag variety). For this set

$$\widetilde{\mathfrak{u}}_{-1}(\delta) := \mathrm{span}\{\delta\}$$

and define recursively

$$\widetilde{\mathfrak{u}}_k(\delta) := \{x \in \mathfrak{g}_k : [x, \delta] \in \widetilde{\mathfrak{u}}_{k-1}(\delta)\}, \quad k \geq 0.$$

Then $\mathrm{sym}_{F(\delta)} = \bigoplus_{k \geq -1} \widetilde{\mathfrak{u}}_k(\delta)$. The algebra $\mathrm{sym}_{F(\delta)}$ is in fact the maximal graded subalgebra among all graded subalgebras of algebras with negative part equal to $\mathrm{span}\{\delta\}$. Moreover the algebra $\mathrm{sym}^{\mathrm{ne}}_{F(\delta)}$ is the largest ideal of $\mathrm{sym}_{F(\delta)}$ concentrated in nonnegative degree and $\mathrm{sym}_{F(\delta)}/\mathrm{sym}^{\mathrm{ne}}_{F(\delta)}$ is isomorphic to the algebra $\mathrm{sl}_2$.

### Construction of canonical frames for curves of constant type $\delta$

Fix $\delta \in \mathfrak{g}_{-1}$ again. Now we will describe the unified construction of canonical bundle of moving frames for all curves of constant type $\delta$. Fix a curve $\gamma$. Let $\pi : G \to G/G^0$ be the canonical projection. By a *moving frame bundle* $B$ over $\gamma$ we mean any subbundle (not necessarily principal) of the $G^0$-bundle $\pi^{-1}(\gamma) \to \gamma$.

Let $B(t) = \pi^{-1}\big(\gamma(t)\big) \cap B$ be the fiber of $B$ over the point $\gamma(t)$. Given any $\Gamma \in B(t)$ consider the tangent space $T_\Gamma B(t)$ to the fiber $B(t)$ at $\Gamma$. This space can be identified with the following subspace $W_\Gamma$ of the Lie algebra $\mathfrak{g}^0$ via the left Maurer-Cartan form $\Omega$:

$$W_\Gamma := \Omega_\Gamma\left(T_\Gamma B(t)\right). \tag{A.41}$$

If $B$ is a principal bundle over our curve, which is a reduction of the bundle $\pi^{-1}(\gamma) \to \gamma$, then the space $L_\Gamma$ is independent of $\Gamma$ and equal to the Lie algebra of the structure group of the bundle $B$. For our purposes here we need to consider more general class of fiber subbundles of $\pi^{-1}(\gamma) \to \gamma$. To define this class first consider the decreasing filtration $\{\mathfrak{g}^k\}_{k\geq 0}$of the graded space $\mathfrak{g}^0$ where

$$\mathfrak{g}^k = \bigoplus_{i \geq k} \mathfrak{g}_i.$$

Given a subspace $U$ of $\mathfrak{g}^0$ let $U^k = U \cap \mathfrak{g}^k$. Note that the quotient space $\mathfrak{g}^k/\mathfrak{g}^{k+1}$ is naturally identified with $\mathfrak{g}_k$. Therefore, since the quotient space $U^k/U^{k+1}$ can be considered as a subspace of $\mathfrak{g}^k/\mathfrak{g}^{k+1}$, $U^k/U^{k+1}$ can be naturally identified with a subspace $U_k$ of $\mathfrak{g}_k$. With this notation, we assign to each subspace $U$ of $\mathfrak{g}^0$ a graded subspace $\mathrm{gr}\, U := \bigoplus_{k \geq 0} U_k$ of $\mathfrak{g}^0$. Note that this space is in general different from the original space $U$.

The space $\mathrm{gr} W_\Gamma$, where $W_\Gamma$ is as in (A.41), is called the *symbol* of the bundle $B$ at the point $\Gamma$.

**Definition A.16.** We say that the fiber subbundle $B$ of $\pi^{-1}(\gamma) \to \gamma$ has a *constant symbol* $\mathfrak{s}$ if its symbols at different points coincide with $\mathfrak{s}$. In this case we call $B$ the *quasi-principal subbundle of the bundle* $\pi^{-1}(\gamma) \to \gamma$ *with symbol* $\mathfrak{s}$.

Let $[\delta, \mathfrak{g}_k] := \{[\delta, y] : y \in \mathfrak{g}_k\}$.

**Definition A.17.** Let $N = \bigoplus_{k \geq 0} N_k$ be a graded subspace of $\mathfrak{g}^0$, i.e., such that $N_k \subset \mathfrak{g}_k$. We say that $N$ defines a normalization condition if for any $k \geq 0$ the subspace $N_k$ is complementary to $\mathfrak{u}_k + [\delta, \mathfrak{g}_{k+1}]$ in $\mathfrak{g}_k$.

$$\mathfrak{g}_k = \big(\mathfrak{u}_k(\delta) + [\delta, \mathfrak{g}_{k+1}]\big) \oplus N_k, \ \ k \geq 0. \tag{A.42}$$

In this case we also say that the graded subspace $N_k$ is complementary to $\big(\mathfrak{u}(\delta) + [\delta, \mathfrak{g}^0]\big) \cap \mathfrak{g}^0$ in $\mathfrak{g}^0$.

**Theorem A.18.** *Given a normalization condition $N$, for any curve $\gamma$ of constant type $\delta$ the set of moving frames $\Gamma(t)$ such that its structure function $C_\Gamma$ satisfies*

$$C_\Gamma(t) - \delta \in N, \quad \forall t, \tag{A.43}$$

*foliates the fiber subbundle of the bundle $\pi^{-1}(\gamma) \to \gamma$ of constant symbol $\mathfrak{u}(\delta)$. Moreover, if $N$ is invariant with respect to the adjoint action of the group $\mathrm{Sym}^{\mathrm{ne}}_{F(\delta)}$ of noneffective symmetries of the flat curve $F(\delta)$ of constant type $\delta$, then the resulting bundle is a principle $\mathrm{Sym}^{\mathrm{ne}}_{F(\delta)}$-subbundle of $\pi^{-1}(\gamma) \to \gamma$ and the foliation of moving frames, satisfying* (A.43)*, is invariant with respect to the principal $\mathrm{Sym}^{\mathrm{ne}}_{F(\delta)}$-action.*

*Proof.* We will say that a moving frame $\Gamma(t)$ is *normal up to order $k \geq 0$*, if its structure function $C_\Gamma$ satisfies

$$\big(C_\Gamma(t)\big)_{-1} = \delta \tag{A.44}$$

for all $t$ and

$$\big(C_\Gamma(t)\big)_i \in N_i \tag{A.45}$$

for all $t$ and $0 \leq i < k$.

We will construct by induction the decreasing sequence of subbundles

$$B^{-1} = \pi^{-1}(\gamma) \supset B^0 \subset B^1 \supset \dots \tag{A.46}$$

such that $B^k$ is the union of all normal up to order $k$ moving frames. Moreover, the bundle $B^k$ has constant symbol $\bigoplus_{i=0}^{k} \mathfrak{u}_i(\delta) \oplus \mathfrak{g}^{k+1}$.

Let us describe this inductive procedure. For $k = 0$, the condition (A.45) is void and by (A.30) the condition (A.44) is equivalent to

$$\Big(\Gamma(t)^{-1}\delta_t\Gamma(t)\Big)_{-1} = \delta.$$

By Definition A.10 of curves of constant type $\delta$ such $\Gamma(t)$ exists for any $t$ , i.e., $B^0$ is not empty. Moreover, by Remark A.13 $B_0$ is the principal reduction of the principal bundle $B_{-1}$ with a structure group $U^0(\delta)$ as in (A.35) and with the Lie algebra $\mathfrak{u}_0 \oplus \mathfrak{g}^1$. In particular, the latter is the symbol of this algebra.

Now assume by induction that the bundle $B_{k-1}$ with the properties above is constructed for some $k \geq 1$ and construct the next bundle $B^k$. Since, by assumptions, (A.45) holds for $i < k-1$, the symbol of $B^{k-1}$ is $\bigoplus_{i=0}^{k-1} \mathfrak{u}_i(\delta) \oplus \mathfrak{g}^k$, and by (A.42) the spaces $N_i$ and $\mathfrak{u}_i(\delta)$ intersect trivially, we have that if two normal up to order $k-1$ moving frames $\Gamma$ and $\widetilde{\Gamma}$ pass through the same point at $t=t_0$, i.e., $\Gamma(t_0)=\widetilde{\Gamma}(t_0)$, then

$$\Big(C_\Gamma(t_0)\Big)_i = \Big(C_{\widetilde{\Gamma}}(t_0)\Big)_i, \quad 0 \le i \le k-2; \tag{A.47}$$

$$\Big(C_\Gamma(t_0)\Big)_{k-1} = \Big(C_{\widetilde{\Gamma}}(t_0)\Big)_{k-1} \quad \mod \mathfrak{u}_{k-1}(\delta). \tag{A.48}$$

In other words, for $0 \le i \le k-2$ the degree $i$ component of the structure function of a normal up to order $k-1$ frame, passing through a point $b \in B^{k-1}$, depends not on the frame but on $b$ only, while the degree $k-1$ component depends on $b$ modulo $\mathfrak{u}_{k-1}(\delta)$. For the latter case, let us denote by $\xi_{k-1}(b)$ the corresponding element of $\mathfrak{g}_{k-1} \mathrm{mod}\, \mathfrak{u}_{k-1}$ or, equivalently, of $\mathfrak{g}_{k-1}/\mathfrak{u}_{k-1}(\delta)$.

Now let us explore how the function $R_{k-1}$ changes along the fiber of $B^{k-1}$. Denote by $R_a$ the right translation by $a$ in the group $G$.

**Lemma A.19.** *The following identity holds*

$$\xi_{k-1}\big(R_{\exp x} b\big) = \xi_{k-1}(b) + [\delta, x]_{k-1} \quad \mod \mathfrak{u}_{k-1}(\delta), \quad \forall x \in g^k, \tag{A.49}$$

*or, equivalently,*

$$\xi_{k-1}\big(R_{\exp x} b\big) = \begin{cases} \xi_{k-1}(b) + [\delta, x] \quad \mod \mathfrak{u}_{k-1}(\delta), & x \in g_k; \\ \xi_{k-1}(b) \quad \mod \mathfrak{u}_{k-1}(\delta), & x \in g^{k+1}. \end{cases} \tag{A.50}$$

*Proof.* We use the following equivariance property of the left Maurer -Cartan form:

$$R_a^* \Omega = \Big(\operatorname{Ad} a^{-1}\Big)\Omega, \quad \forall a \in G. \tag{A.51}$$

Let us prove (A.51) for completeness. Indeed, using the definition of the left Maurer-Cartan form given by (A.3) and the fact that any left translation commutes with the right translation, we have

$$R_a^* \Omega_b(y) = \Omega_{ba}\big((R_a)_* y\big) = \Big((L_{(ba)^{-1}})_* \circ (R_a)_*\Big)(y) = \Big((L_{a^{-1}})_* \circ (R_a)_* \circ (L_{b^{-1}}\Big)_*(y) = \Big(\operatorname{Ad} a^{-1}\Big)\Omega_b(y),$$

for any $y \in \mathfrak{g}$, which proves (A.51).

Take a moving frame $\Gamma$ over $\gamma$ and consider the moving frame $R_{\exp x}(\Gamma)$. Using formula (A.51) and the Taylor expansion as in (A.39) with $\delta = x$, one can relate the structure functions of the frames $\Gamma$ and $R_{\exp x}(\Gamma)$ in the following way:

$$C_{R_{\exp x}(\Gamma)}(t) = C_\Gamma(t) + \sum_{k=1}^{\infty} \frac{1}{k!} \big(\operatorname{ad}(-x)\big)^k C_\Gamma(t), \tag{A.52}$$

where the sum is actually finite. Comparing the homogeneous components of degree less than $k-1$ in both sides of (A.52), we get that if $x \in \mathfrak{g}^k$ and the moving frame $\Gamma$ is normal up to the order $k-1$, then the moving frame $R_{\exp x(\Gamma)}$ is normal up to the order $k-1$. Therefore, comparing the degree $k-1$ components in both sides of (A.52) and using (A.44), we can replace the structure function by the function $\xi_{k-1}$ evaluated at the appropriate points to obtain (A.49). $\square$

Now, from (A.50) and the fact that $N_{k-1}$ is complementary to $[x, \mathfrak{g}_k]$ mod $\mathfrak{u}_{k-1}(\delta)$ it follows that for $b \in B^{k-1}$, one can find $x \in \mathfrak{g}_k$ such that

$$\xi_{k-1}(R_{\exp x} b) \in N_{k-1} \text{ mod } \mathfrak{u}_{k-1}(\delta). \tag{A.53}$$

Moreover, since the degree $k-1$ component of the symbol of the bundle $B^{k-1}$ is equal to $\mathfrak{u}_{k-1}(\delta)$ one can find a normal up to order $k$ moving frame $\Gamma$ passing through $R_{\exp x} b$ which implies that $R_{\exp x} b \in B^k$, i.e., $B^k(t_0)$ is not empty, where $t_0$ is such that $\pi(b) = \gamma(t_0)$.

Now, if $b \in B^k$ , then (A.50) implies that $R_{\exp x} b \in B^k$ for $x \in \mathfrak{g}_k$ if and only if

$$[x, \delta] \in N_{k-1} \oplus \mathfrak{u}_{k-1}(\delta). \tag{A.54}$$

Since by (A.42) $N_{k-1}$ is transversal to $\mathfrak{u}_{k-1}(\delta) + [\delta, \mathfrak{g}_k]$ in $\mathfrak{g}_{k-1}$, the relation (A.54) implies that $[\delta, x] \in \mathfrak{u}_{k-1}(\delta)$, hence $x \in \mathfrak{u}_k(\delta)$. This implies that $B_k$ is the fiber subbundle of $B^{-1}$ with constant symbol $\bigoplus_{i=0}^{k} \mathfrak{u}_i(\delta) \oplus \mathfrak{g}^{k+1}$, which concludes the proof of the induction step.

Since there exists an integer $m$ such that $\mathfrak{g}_i = 0$ for all $i \geq m$, the sequence of bundles (A.46) will be stabilized, i.e., $B^i = B^m$ for all $i \geq m$. Moreover, the normal up to order $m$ moving frames will foliate $B^m$, because for any point $b \in B^m$ the structure function of any normal up to order $m$ moving frame and therefore the tangent line to such a moving frame will depend on the point $b$ only. So, there is a unique normal up to order $m$ moving frame which passes through $b$. So, $B^m$ is the desired bundle of moving frames, which completes the proof of the first part of the theorem.

The moving frames, which are normal up to order $m$, will be called simply *normal*. If $\mathcal{N}$ is invariant with respect to the adjoint action of $\mathrm{Sym}^{\mathrm{ne}}_{F(\delta)}$, then by (A.51), if the moving frame $\Gamma$ is normal, then for any $u \in \mathrm{sym}^{\mathrm{ne}}_{F(\delta)}$ the moving frame $R_{exp\, u}(\Gamma)$ is normal as well. Hence the bundle $B^m$ is a principal $U^0(\delta)$-bundle and the foliation of normal moving frames is invariant with respect to $R_{exp\, u}$, which completes the proof of the second part of our theorem. $\square$

## A.3 Application to differential geometry of monotonic parametrized curves in Lagrangian Grassmannians

Now we apply the general algebraic theory, developed in the previous section, to construct the canonical bundle of moving frames for monotonic parametrized curves in Lagrangian Grassmannians. For this we will first classify all possible symbols of their osculating flags, compute their algebraic prolongation, and find the natural invariant normalization conditions.

### Classification of symbols of monotonic curves in Lagrangian Grassmannians

Let $\Lambda(t)$ be a parametrized equiregular curve in Lagrangian Grassmannians $L(V)$. As in formula (A.20) of subsection A.2, let $\{\Lambda^j\}_{j \in Z}$ be the osculating flag. We do not lose much by assuming that there exists a negative integer $p$ such that

$$\Lambda^{(p)}(t) = V \tag{A.55}$$

Otherwise, if $\Lambda^{(p-1)}(t) = \Lambda^{(p)}(t) \subset V$, then the subspace $\tilde{V} = \Lambda^{(p)}(t)$ does not depend on $t$ and one can work with the curve $\Lambda(t)/\widetilde{V}^{\angle}$ in the symplectic space $\widetilde{V}/\widetilde{V}^{\angle}$ instead of $\Lambda(t)$.

**Definition A.20.** The curve $\Lambda$ is called ample, if for any $t$ there exists $p$ for which (A.55) holds.

Let $D$ be the Young diagram of the curve $\Lambda$ (see Definition A.7). Let the length of the rows of $D$ be $p_1$ repeated $r_1$ times, $p_2$ repeated $r_2$ times, ..., $p_s$ repeated $r_s$ times with $p_1 > p_2 > \ldots > p_s$. *The reduction or the reduced Young diagram of the Young diagram* $D$ is the Young diagram $\Delta$, consisting of $k$ rows such that the $i$th row has $p_i$ boxes.

Make the mirror reflection of the Young diagram $\Delta$ with respect to its left vertical edge . Denote the skew-diagram obtained by union of this mirror reflection and $\Delta$ by $\widetilde{\Delta}$. Denote by $r$ and $l$ the right and left shifts on $\widetilde{\Delta}$, respectively. In other words given a box $a$ of $\widetilde{\Delta}$ denote by $r(a)$ and $l(a)$ the boxes next to $a$ to the right and to the left, respectively, in the same row of $\Delta$. Also let $m : \widetilde{\Delta} \to \widetilde{\Delta}$ be the mirror reflection with respect to the left vertical edge of the diagram $\Delta$, i.e., the map sending a box $a$ of $\widetilde{\Delta}$ to the box which is mirror-symmetric to $a$ with respect to this left edge.

We say that the basis $\{E_a\}_{a\in\widetilde{\Delta}}$ of $V$, where for a box $a$ from the $i$th row of $\widetilde{\Delta}$ $E_a$ is the tuple of $r_i$ vectors in $V$, $E_a = (e_a^1, \ldots e_a^{r_i})$, forms a *Darboux basis indexed by the diagram* $\widetilde{D}$, if any vector from the tuple $E_a$ is skew-orthogonal to any vector from the tuple $E_b$ for $b \neq m(a)$ and

$$\sigma(e^i_{m(a)}, e^j_a) = \delta_{ij}, \quad a \in \Delta. \tag{A.56}$$

Given a tuple of vectors $E$ in $V$ and an endomorphism $X$ of $V$, by $XE$ we denote the tuple of vectors obtained by applying $X$ to vectors of $E$. If $Y$ is a matrix with the same number of rows as the number of vectors in E, then by $EY$ we mean the new tuple with the $j$th vector equal to the linear combination of vectors $E$ with coefficients in $j$th column of $Y$.

**Exercise A.21.** Show that the map $X \in \mathfrak{sp}(X)$ if and only if it has the representation in a Darboux basis $\{E_a\}_{a\in\widetilde{\Delta}}$

$$X(E_a) = \sum_{b\in\widetilde{\Delta}} E_b X_{ba} \tag{A.57}$$

such that

$$X_{ab} = X^T_{m(b)m(a)}, \quad a \in \Delta, b \in m(\Delta) \tag{A.58}$$

$$X_{ab} = -X^T_{m(b)m(a)} \; a, b \in \Delta \tag{A.59}$$

If we denote

$$\varepsilon(a) = \begin{cases} -1, & a \in \Delta; \\ 1, & a \in m(\Delta) \end{cases} \tag{A.60}$$

then (A.58) and (A.59) can be written as

$$X_{ab} = -\varepsilon(a)\varepsilon(b) X^T_{m(b)m(a)} \tag{A.61}$$

**Proposition A.22.** *Any monotonic equiregular ample curve with Young Diagram $D$ in a Lagrangian Grassmannian has the unique symbol represented by the endomorphism $\delta$ acting on a Darboux basis $\{E_a\}_{a\in\widetilde{\Delta}}$ as follows:*

$$\delta(E_a) = \varepsilon(a) E_{r(a)}, \tag{A.62}$$

*where $\varepsilon(a)$ is defined in* (A.60). *In particular, there is one-to-one correspondence between the set of Young diagrams and the set of symbols of monotonic curve in Lagrangian Grassmannians.*

*Proof.* Let $V_j$ and $\mathrm{gr}V(t)$ be as (A.22) and (A.23). Let $\delta_t$ be the symbol of the curve of osculating flags $\{\Lambda^j\}_{j\in Z}$ at $t$.

**Lemma A.23.** *If $\Lambda$ satisfies the assumption of Proposition A.22, then for any $j \geq 0$ the map $\delta_t^{2j+1} : V_j(t) \to V_{-j-1}(t)$ is an isomorphism.*

*Proof.* Let $\sigma_t$ be the natural symplectic form on $\mathrm{gr}V_t$ induced by the symplectic form on $V$ as described in subsection A.2. Note that the bilinear form $(x, y) \mapsto \sigma_t(\delta_t x, y)$ on $V_0(t)$ is symmetric and nondegenerate, as follows from (A.17) and the construction of the spaces $\Lambda^{(1)}(t)$. From the fact that the osculating flags are symplectic it follows that

$$\sigma_t(\delta_t^{2j+1}x, y) = (-1)^s \sigma_t(\delta_t^{2j+1-s}x, \delta_t^s y) \tag{A.63}$$

for all $x, y \in V_j(t)$. In particular, for $s = j$

$$\sigma_t(\delta_t^{2j+1}x, y) = (-1)^j \sigma_t(\delta_t^{j+1}x, \delta_t^j y)$$

for all $x, y \in V_j(t)$. This implies that the bilinear form $\sigma_t(\delta_t^{2j+1}x, y)$ on $V_j(t)$ is symmetric and the desired condition that $\delta_t^{2j+1} : V_j(t) \to V_{-j-1}(t)$ is an isomorphism is equivalent to the fact that this form is nondegenerate. Besides, since $\delta_t^j : V_j(t) \mapsto V_0(t)$ is injective (see Remark A.9) the latter statement is equivalent to the fact that the symmetric bilinear form $\sigma_t(\delta_t x, y)$ (and hence the corresponding quadratic form) is nondegenerate on the subspace $\delta_T^j(V_j)$ of $V_0(t)$. Finally, since by the assumption the curve $\Lambda$ is monotonic, the quadratic form $\sigma_t(\delta_t x, y)$ is positive definite on $V_0(t)$, and hence its restriction to any subspace of $V_0(t)$ is positive definite and hence nondegenerate. This completes the proof of the lemma. □

Let $\rho_i$ be the last (i.e., the most right) box of the $i$th row of the reduced Young diagram $\Delta$ of $D$ and let again $p_i$ is the number of boxes in the $i$th row of $\Delta$. Let $E_{m(\rho_1)}(t)$ be a basis of $V_{p_1}(t)$ orthonormal with respect to the inner product $(-1)^{p_1}\sigma_t(\delta_t^{2p_1+1}x, y)$. Then for $0 \leq s < p_1 - p_2 - 1$ set

$$E_{r^s\left(m(\rho_1)(t)\right)} := \delta_t^s E_{m(\rho_1)}(t).$$

Note that by (A.63) the tuples $\delta_t^s E_{m(\rho_1)}(t)$ are orthonormal with respect to the inner product

$$(-1)^{p_1-s}\sigma_t(\delta_t^{2(p_1-s)+1}x, x).$$

Further, let $E_{m(\rho_2)}(t)$ be a completion of the tuple $\delta_t^{p_2-p_1-1}E_{m(\rho_1)}(t)$ to an orthonormal basis of $V_{p_2}(t)$ with respect to the inner product $(-1)^{p_2}\sigma_t(\delta^{2p_2+1}x, x)$. So, we defined $E_a(t)$ for all $a$ not located to the right of $\rho_2$ in $m(\Delta)$.

In the same way, by applying $\delta$ and completing the constructed tuples to orthonormal bases of the corresponding $V_j(t)$ with respect to the corresponding inner product, we can define $E_a$ for all $a \in m(\Delta)$ such that $\delta_t E_a(t)$ is either equal to $E_{r(a)}(t)$ or is subtuple of $E_{r(a)}(t)$. Further, for the box $a$ in the $j$th column of the diagram $\Delta$ set $E_a(t) := (-1)^{j-1}\delta_t^{2j-1}E_{m(a)}(t)$. Then again from (A.63) it follows that $\{E_a(t)\}_{a\in\widetilde{\Delta}}$ is a Darboux frame of $\mathrm{gr}V(t)$. Moreover, by constructions $\delta_t$ acts on the basis $\{E_a(t)\}_{\widetilde{D}}$ as in (A.62), which proves the statement. □

### Calculation of the universal algebraic prolongation of $\delta$

We assume that $\delta$ has the form (A.62) in a Darboux basis $\{E_a\}_{a\in\widetilde{\Delta}}$. First, let us calculate the commutator of $\delta$ with $X \in \mathfrak{sp}(V)$ in the Darboux basis $\{E_a\}_{a\in\widetilde{\Delta}}$. We need it both for calculation of the algebraic prolongation and for the choice of the normalization conditions.

**Lemma A.24.** *If*

$$[\delta, X](E_a) = \sum_{b\in\widetilde{\Delta}} E_b Y_{ba}, \tag{A.64}$$

*then*

$$Y_{ba} = \varepsilon\big(l(b)\big) X_{l(b)a} - \varepsilon(a) X_{b\,r(a)}, \tag{A.65}$$

*where the terms involving non-existing boxes of the diagram $\widetilde{\Delta}$ are considered to be equal to zero.*

*Proof.* Using (A.62) and (A.57), we have

$$[\delta, X](E_a) = \delta \circ X(E_a) - X\delta(A) = \delta\Big(\sum_{b\in\widetilde{\Delta}} E_b X_{ba}\Big) - \varepsilon(a) X(E_{r(a)}) =$$

$$\sum_{b\in\widetilde{\Delta}} \varepsilon(b) E_{r(b)} X_{ba} - \varepsilon(a) \sum_{b\in\widetilde{\Delta}} E_b X_{b\,r(a)} = \sum_{b\in\widetilde{\Delta}} E_b Y_{ba}$$

with $Y_{ba}$ as in (A.65), which completes the proof of the lemma. □

**Proposition A.25.** *The following holds:*

$$\mathfrak{u}(\delta) = \mathfrak{u}_0(\delta) \cong \bigoplus_{i=1}^{s} \mathfrak{so}_{r_i} \tag{A.66}$$

*In more detail, $\mathfrak{u}(\delta)$ consists of all $X \in \mathfrak{g}^0$ such that if $X$ is represented in the Darboux frame $\{E_a\}_{a\in\widetilde{\Delta}}$ by* (A.57)*, then the only possibly nonzero matrices $X_{ba}$ are when $a = b$ and in this case each $X_{aa}$ is skew-symmetric and $X_{l(a)l(a)} = X_{(a,a)}$ for any box $a \in \widetilde{\Delta}$, which is not the first box of a row.*

*Proof.* First let us describe all $X$ in $\mathfrak{g}^0$ (and not necessarily in $\mathfrak{g}_0$) that commute with $\delta$. This will allow us to calculate $\mathfrak{u}_0$ and will be used to prove that $\mathfrak{u}_i = 0$ for $i > 0$.

**Lemma A.26.** *If $X \in \mathfrak{g}^0$ and $[\delta, X] = 0$, then the only possibly nonzero matrices $X_{ba}$ in the representation* (A.57) *are when $a = b$ and in this case each $X_{aa}$ is skew-symmetric and $X_{l(a)l(a)} = X_{(a,a)}$ for any box $a \in \widetilde{\Delta}$, which is not the first box of a row.*

*Proof.* Let, as before, $\rho_i$ be the last (i.e., the most right) box in the $i$th row of $\widetilde{\Delta}$. Since $r(\sigma)$ does not exist, by (A.65) we get that the condition $Y_{b\rho_i} = 0$ implies that

$$X_{l(b)\rho_i} = 0. \tag{A.67}$$

This means that $X_{b\rho_i} = 0$ for all $b \neq \rho_j$. Since $X \in \mathfrak{g}^0$, $X_{\rho_j\rho_i} = 0$ if $i < j$. Also, if $j > i$, then (A.59) implies that $X_{\rho_j\rho_i} = -X^T_{m(\rho_i)m(\rho_j)}$ and the latter is 0 from the condition $X \in \mathfrak{g}^0$ again. So, we got that

$$X_{b\rho_i} = 0, \quad b \neq \rho_i. \tag{A.68}$$

Using relations (A.67) and (A.65) we get that the condition $Y_{l(b)l(\rho_i)} = 0$ implies that

$$X_{l^2(b)l(\rho_i)} = 0.$$

In the same way, by induction we will get that if $[\delta, X] = 0$, then

$$X_{l^{i+1}(b)l^i(\rho_i)} = 0,$$

which together with (A.68) yields

$$X_{ba} = 0, \quad b \neq a.$$

Now treat the case $b = a$. From (A.65) the condition $Y_{r(a)a} = 0$, where $a$ is not the last box of a row $\widetilde{\Delta}$, is equivalent to $X_{aa} = X_{r(a)r(a)}$, which implies that $X_{aa} = X_{bb}$ if boxes $a$ and $b$ lie in the same row of $\widetilde{\Delta}$. Since boxes $a$ and $m(a)$ lie in the same row, we get from this and (A.59) that

$$X_{aa} = X_{m(a)m(a)} = -X_{aa}^T,$$

i.e., $X_{aa}$ is skew-symmetric, which completes the proof of the lemma. □

From the previous lemma we get that $\mathfrak{u}_0$ consists of all $X \in \mathfrak{g}_0$ such that the only possibly nonzero matrices $X_{ba}$ in the representation (A.57) are when $a = b$ and in this case each $X_{aa}$ is skew-symmetric and $X_{l(a)l(a)} = X_{(a,a)}$ for any box $a \in \widetilde{\Delta}$, which is not the first box of a row.

Also, from the previous lemma it follows that in order to complete the proof of the proposition it is enough to prove that $\mathfrak{u}_1 = 0$, because there is no nontrivial elements of degree $\geq 2$ in $\mathfrak{g}$ , which commute with $\delta$.

To calculate $\mathfrak{u}_1$, let $Y_{ba}$ be as in (A.64). If $\rho_i$ is the last box of the $i$ row of $\widetilde{\Delta}$ and $[\delta, X] \in \mathfrak{u}_0$, then $Y_{b\rho_i} = 0$ form $b \neq \rho_i$. Hence by (A.65) we have (A.67) for $b \neq \rho_i$ and by exactly the same arguments as in Lemma A.24 and the assumption that $X \in \mathfrak{g}_1$ one gets that

$$X_{ba} = 0, \quad b \neq l(a). \tag{A.69}$$

Now, applying (A.65) for $b = a$, we get

$$Y_{aa} = \varepsilon\big(l(a)\big)X_{l(a)a} - \varepsilon(a)X_{a\,r(a)}, \tag{A.70}$$

where $Y_{aa}$ is the same skew-symmetric matrix for all $a$ on the same row of $\widetilde{\Delta}$. For $a = \rho_i$ formula (A.70) implies

$$X_{l(\rho_i)\rho_i} = \varepsilon\big(l(\rho_i)\big)Y_{\rho_i\rho_i} \tag{A.71}$$

Now use (A.70) for $a = l(\rho_i)$

$$Y_{l(\rho_i)l(\rho_i)} = \varepsilon\big(l^2(\rho_i)\big)X_{l^2(\rho_i)l(\rho_i)} - \varepsilon\big(l(\rho_i)\big)X_{l(\rho_i)\rho_i} \tag{A.72}$$

Substituting (A.71) into (A.72) and using that $Y_{l(\rho_i)l(\rho_i)} = Y_{\rho_i\rho_i}$, we get

$$X_{l^2(\rho_i)l(\rho_i)} = 2\varepsilon\big(l^2(\rho_i)\big)Y_{\rho_i\rho_i}.$$

Continuing by induction we get

$$X_{l^j(\rho_i)l^{j-1]}(\rho_i)} = j\varepsilon\big(l^j(\rho_i)\big)Y_{\rho_i\rho_i}, \tag{A.73}$$

which implies that for every $a$ in the $i$th row of $\widetilde{\Delta}$ the matrix $X_{l(a)a}$ is a nonzero multiple of the same skew-symmetric matrix $Y_{\rho_i\rho_i}$.

Now assume that $a_i$ is the first box in the $i$th row of $\Delta$. Then $l(a_i) = m(a_i)$ hence by (A.58) we have

$$X_{l(a_i)a_i} = X_{m(a_i)a_i} = X^T_{m(a_i)a_i} = X^T_{l(a_i)a_i},$$

i.e., $X_{l(a_i)a_i}$ is simultaneously symmetric and skew-symmetric and hence it is equal to zero. This implies that $X_{l(a)a} = 0$, which together with (A.69) yields that $X = 0$. So, we proved that $\mathfrak{u}_1 = 0$ and hence $u_i = 0$ for all $i \geq 2$. The proof of the proposition is completed. □

*Remark* A.27. As a matter of fact, $\mathfrak{u}_0(\delta)$ can be found without calculations from the fact that each space $V_j(t)$, $j \geq 0$ is endowed with the Euclidean structure given by the quadratic form $\sigma_t(\delta_t^{2j+1}x, x)$. Also, from this and the fact that $\mathfrak{u}(\delta) = \mathfrak{u}_0(\delta)$ it is obvious that $\mathrm{Sym}^{\mathrm{ne}}_{F(\delta)} \cong O_{r_1} \times \ldots \times O_{r_s}$. The adjoint action of this group on $\mathfrak{g}$ can be described as follows: If $U = (U_1, \ldots, U_s)$, where $U_i \in O_{r_i}$ and $X \in \mathfrak{g}$, then

$$(\mathrm{Ad}\, U X)_{ba} = U_j X_{ba} U_i^{-1}, \tag{A.74}$$

where $a$ and $b$ are in the $i$th and $j$th rows of $\widetilde{\Delta}$, respectively.

## Calculation of $[\delta, \mathfrak{g}^0]$

Before choosing the normalization condition, i.e., a graded subspace complementary to $\mathfrak{u}(\delta) + [\delta, \mathfrak{g}^0] \cap \mathfrak{g}^0$ in $\mathfrak{g}^0$, we have to describe the space $[\delta, \mathfrak{g}^0] \cap \mathfrak{g}^0$. The following notation will be useful for this purpose. Given $Y \in \mathfrak{g}^0$, which has the form $Y = \sum_{b \in \widetilde{\Delta}} E_b Y_{ba}$, let

$$D(Y)_{ba} := Y_{ba} + \frac{\varepsilon\big(l(b)\big)}{\varepsilon\big(l(a)\big)} Y_{l(b)l(a)} + \frac{\varepsilon\big(l(b)\big)}{\varepsilon\big(l(a)\big)} \frac{\varepsilon\big(l^2(b)\big)}{\varepsilon\big(l^2(a)\big)} Y_{l^2(b)l^2(a)} + \ldots = \sum_{j \geq 0} \left( \prod_{s=1}^{j} \frac{\varepsilon\big(l^s(b)\big)}{\varepsilon\big(l^s(a)\big)} \right) Y_{l^j(b)l^j(a)}, \tag{A.75}$$

where the sum is finite as we reach the first box of a row after finite number of applications of $l$.

**Proposition A.28.** *$Y \in [\delta, \mathfrak{g}^0] \cap \mathfrak{g}^0$ if and only if for every last box $\rho$ of the diagram $\widetilde{\Delta}$ and every box $b \in \widetilde{\Delta}$ that is not higher than $\rho$ in $\widetilde{\Delta}$ the following identity holds*

$$D(Y)_{b\rho} = 0. \tag{A.76}$$

*Proof.* Let $(a, b)$ be a pair of boxes of $\widetilde{\Delta}$ such that $b$ is not to the right and not higher than $a$. Note that from (A.61) an element $X \in \mathfrak{g}^0$ is determined uniquely from the knowledge of $X_{ba}$ for all such pairs.

If $Y = [\delta, X]$ for some $X \in \mathfrak{g}^0$, then applying (A.65) to pairs of boxes $(b, a)$, $\big(l(a), l(b)\big)$

$,\big(l^2(a), l^2(b)\big), \dots$ we get the following chain of identities:

$$
\begin{aligned}
&Y_{ba} = \varepsilon\big(l(b)\big)X_{l(b)a} - \varepsilon(a)X_{b\,r(a)},\\
&Y_{l^2(b)l^2(a)} = \varepsilon\big(l^2(b)\big)X_{l^2(b)l(a)} - \varepsilon\big(l(a)\big)X_{l(b)\,a)},\\
&Y_{l^3(b)l^3(a)} = \varepsilon\big(l^3(b)\big)X_{l^2(b)l(a)} - \varepsilon\big(l^2(a)\big)X_{l^2(b)\,l(a)},\\
&\vdots\\
&Y_{l^{j-1}(b)l^{j-1}(a)} = \varepsilon\big(l^j(b)\big)X_{l^j(b)l^{j-1}(a)} - \varepsilon\big(l^{j-1}(a)\big)X_{l^{j-1}(b)\,l^{j-2}(a)},\\
&Y_{l^j(b)l^j(a)} = -\varepsilon\big(l^j(a)\big)X_{l^j(b)\,l^{j-1}(a)},
\end{aligned}
\tag{A.77}
$$

where $j$ is such that $l^j(b)$ is the first box in the corresponding row of $\widetilde{\Delta}$. Note that by the assumption on the location of $b$ with respect to $a$ all indices appearing in (A.77), except maybe $r(a)$, are well defined. Eliminating $X_{l(b)a}$ by taking an appropriate linear combination of the first identities in (A.77), then eliminating $X_{X_{l^2(b)l(a)}}$ from the resulting combination by adding the third identity of (A.77), and continuing this successive eliminating procedure we get that

$$
\varepsilon(a)X_{br(a)} = -D(Y)_{ba} \tag{A.78}
$$

This implies (A.76) in the case when $a = \rho$, i.e., the last box of a row in $\widetilde{\Delta}$, which proves necessity of (A.76).

To prove sufficiency, given $Y \in \mathfrak{g}^0$, satisfying (A.76), define $X$ such that it satisfies (A.78) for all pair $(a, b)$ , where $a$ is not the last box in a row and $b$ is not to the right and not higher than $r(a)$, and also such that (A.61) holds. It can be shown that conditions (A.78) and (A.61) are consistent in the case when $a$ and $b$ lie in the same row, so such $X$ indeed can be constructed and $X \in \mathfrak{g}^0$. Moreover, by reversion of the procedure of going from (A.77) to (A.78) we can show that $[\delta, X] = Y$, which completes the proof of sufficiency. □

Now in order to choose a normalization condition the following lemma is useful:

**Lemma A.29.** *Assume that $\rho$ is the last box of a row of $\widetilde{\Delta}$ and $b$ be the $k$th box in the same row (from the left). Then if $Y \in \mathfrak{g}^0$ the matrix $D(Y)_{b\rho}$ is symmetric if $k$ is odd and skew-symmetric if $k$ is even.*

*Proof.* In the considered case in the sum (A.75) defining $D(Y)_{b\rho}$ the terms are subdivided into pairs satisfying relation (A.61). Indeed, for any $j$, $0 \le j \le k$ it is easy to show that

$$
m\big(l^j(b)\big) = l^{k-1-j}(\rho), \quad m\big(l^j(\rho)\big) = l^{k-1-j}(b).
$$

Hence, by (A.61),

$$
Y_{l^j(b)l^j(\rho)} = -\varepsilon\big(l^j(b)\big)\varepsilon\big(l^j(\rho)\big)\big(Y_{l^{k-1-j}(b)l^{k-1-j}(\rho)}\big)^T \tag{A.79}
$$

Assume that the number of boxes in the considered row of $\widetilde{D}$ is equal to $2p$ Consider the following two cases separately

*Case 1.* Assume that $k \le p$. Then for all $j$, $0 \le j \le k-1$,

$$
l^j(b) \in m(\Delta), \quad l^j(\rho) \in \Delta \tag{A.80}
$$

Hence, $\varepsilon\big(l^j(b)\big) = -\varepsilon\big(l^j(\rho)\big)$. Therefore, by (A.75)

$$D(Y)_{b\rho} = \sum_{j=0}^{k-1} (-1)^j Y_{l^j(b)l^j(\rho)} \tag{A.81}$$

and by (A.79)

$$Y_{l^j(b)l^j(\rho)} = \big(Y_{l^{k-1-j}(b)l^{k-1-j}(\rho)}\big)^T \tag{A.82}$$

Consequently,

$$(-1)^j Y_{l^j(b)l^j(\rho)} + (-1)^{k-1-j} Y_{l^{k-1-j}(b)l^{k-1-j}(\rho)} = (-1)^j Y_{l^j(b)l^j(\rho)} + (-1)^{k-1-j}\big(Y_{l^j(b)l^j(\rho)}\big)^T$$

and the latter matrix is symmetric if $k$ is odd and skew-symmetric if $k$ is even. This together with (A.81) implies the statement of the lemma.

*Case 2.* Assume that $k > p$. We have 3 subcases:

1. $k - p \le j \le p - 1$;
2. $0 \le j < k - p$;
3. $p \le j < k$.

In subcase (1) (A.80) holds, In subcase (2) $l^j(b) \in m(\Delta)$ and $l^j(\rho) \in m(\Delta)$ and in subcase (3) $l^j(b) \in \Delta$ and $l^j(\rho) \in \Delta$. In both of these subcases $\varepsilon\big(l^j(b)\big) = \varepsilon\big(l^j(\rho)\big)$. Hence, by (A.75)

$$D(Y)_{b\rho} = \sum_{j=0}^{k-p-1} Y_{l^j(b)l^j(\rho)} + \sum_{j=k-p}^{p-1} (-1)^{j-k+p+1} Y_{l^j(b)l^j(\rho)} + (-1)^{2p-k} \sum_{j=p}^{k-1} Y_{l^j(b)l^j(\rho)} \tag{A.83}$$

and by (A.79)

$$Y_{l^j(b)l^j(\rho)} = -\big(Y_{l^{k-1-j}(b)l^{k-1-j}(\rho)}\big)^T \tag{A.84}$$

The middle sum in (A.83) corresponds to subcase (1) and can be treated as in the previous case. To treat the other two sums note that by (A.84)

$$Y_{l^j(b)l^j(\rho)} + (-1)^{k-2p} Y_{l^{k-1-j}(b)l^{k-1-j}(\rho)} = Y_{l^j(b)l^j(\rho)} + (-1)^{k-1}\big(Y_{l^j(b)l^j(\rho)}\big)^T.$$

and the latter is symmetric if $k$ is odd and skew-symmetric if $k$ is even. This implies that the sum of the first and the third sums in (A.83) symmetric if $k$ is odd and skew-symmetric if $k$ is even, which completes the proof of the lemma. $\square$

### A class of $\mathrm{Ad}$-invariant normalization conditions

Now we are ready to choose a normalization condition. Of course, this choice is not unique, but Proposition A.28 and Lemma A.29 immediately suggest an entire class of normalization conditions invariant with respect to the adjoint action of the group $\mathrm{Sym}^{\mathrm{ne}}_{F(\delta)}$ (recall that it has the Lie algebra $\mathfrak{u}(\delta) = \mathfrak{u}_0(\delta)$).

Let us describe this class of normalization conditions. For any last box $\rho$ of a row of $\widetilde{\Delta}$ and a box $b \neq \rho$ , which is not higher than $\rho$, in the following set of pairs of boxes

$$\{(b,\rho), \big(l(b), l(\rho)\big), \big(l^2(b), l^2(\rho)\big), \ldots\} \tag{A.85}$$

we choose exactly one pair of boxes, denoted $\varphi(b,\rho)$.

Let $N_\varphi$ be the subspace of $\mathfrak{g}^0$ consisting of all $Y$ such that the only possibly nonzero matrices $Y_{ba}$ are one of the following:

1. $Y_{\varphi(b,\rho)}$ , if $b$ and $\rho$ are not in the same row of $\widetilde{\Delta}$;
2. $Y_{\varphi(b,\rho)}$ and $Y_{m\left(\varphi(b,\rho)\right)}$, if $b$ and $\rho$ are in the same row of $\widetilde{\Delta}$ and $b \neq \rho$. Moreover, if $b$ is the $k$th box of the row, then $Y_{g\varphi(b,\rho)}$ is symmetric and $Y_{m\left(\varphi(b,\rho)\right)} = Y_{\varphi(b,\rho)}$, if $k$ is odd, and $Y_{\varphi(b,\rho)}$ is skew-symmetric and $Y_{m\left(\varphi(b,\rho)\right)} = -Y_{\varphi(b,\rho)}$, if $k$ is even.

As a direct consequence of Propositions A.25, A.28, and Lemma A.29 we have the following

**Theorem A.30.** *The subspace $N_\varphi$ corresponding to an assignment $\varphi$ as above is an invariant normalization condition with respect to the adjoint action of the group* $\mathrm{Sym}^{\mathrm{ne}}_{F(\delta)}$ *on* $\mathfrak{g}_0$.

The invariancy follows from the form of the adjoint action given by (A.74). The condition that $b \neq \rho$ comes from the fact that we need to find a complement to $\mathfrak{u}(\delta) + [\delta, \mathfrak{g}^0] \cap \mathfrak{g}^0$ and not to $[\delta, \mathfrak{g}^0] \cap \mathfrak{g}^0$.

Note that the normalization condition chosen in the original works [ZL07, ZL09] belongs to the class of normalization of the previous theorem and it corresponds to the following assignment $\varphi_0$:

1. Assume that $b$ and $\rho$ are not in the same row of $\widetilde{\Delta}$ and assume that $c$ is the first (i.e., the most left) box of the row of $b$ in $\widetilde{\Delta}$ and $d$ is a box in the row of $\rho$ such that $(c,d)$ belongs to the set (A.85), then
   - if $m(c)$ is located to the left of $d$, then $\varphi_0(b,\rho) = (c,d)$;
   - if $m(c)$ is not located to the left of $d$, then $\varphi_0(b,\rho)$ is the only pair in the set (A.85) of the form $(m(b_1), a_1)$ or $(m \circ r(b_1), a_1)$, where $a_1$ and $b_1$ lie in the same column;
2. Assume that $b$ and $\rho$ are in the same row of $\widetilde{\Delta}$ and $b$ is the $k$th box of this row.
   - If $k$ is odd then $\varphi_0(b,\rho)$ is the unique pair in the set (A.85) of the form $(m(e), e)$;
   - If $k$ is odd then $\varphi_0(b,\rho)$ is the unique pair in the set (A.85) of the form $(m \circ r(e), e)$.

In the light of the more general theory developed here and based on [DZ12, DZ13] this particular normalization $N_{\varphi_0}$ for general Young diagram does not have any advantage compared to any other normalization of Theorem A.30.

Finally, assume that the assignment $\varphi$ is chosen and we used Theorem A.43 to construct the bundle of moving frames. Then for any $a \in \widetilde{\Delta}$ the space

$$V_a(t) := \Gamma(t)(E_a),$$

is independent of the choice of the normal frame $\Gamma$ and it is endowed with the canonical Euclidean structure (see Remark A.27). Besides from the invariancy with respect to the adjoint action and (A.74) it follows that for any $a$ and $b$ the corresponding matrix block $\left(C_\Gamma(t)\right)_{ba}$ of the structure function of $\Gamma$ defines the linear map from $V_a$ to $V_b$ , which is independent of the choice of the normal basis. We call it the $(a,b)$-*curvature map of the curve* $\gamma$ at $t$. [4]

We conclude with several examples.

[4]The $(a,b)$-curvature defined in the original work [ZL07, ZL09] for the particular normalization condition chosen there corresponds to $\big(a, m(b)\big)$-curvature here.

**Example A.31.** This is the continuation of Example A.8. The reduced diagram of a regular curve in Lagrangian Grassmannian consists of one box, say $\rho$. In this case there is only one box $b = m(\rho)$ in $\widetilde{\Delta}$, which differs from $\rho$ and hence, there is only one choice of the assignment $\varphi$, acting as the identity on the pair $\big(m(\rho), \rho\big)$. There is the unique nontrivial curvature map in this case, the $\big(m(\rho), \rho\big)$-curvature map, and it coincides with the curvature map of the regular curve defined in Chapter 14.

**Example A.32.** *(The case of rectangular Young diagram)* Assume that the Young diagram $D$ of $\Lambda$ is rectangular. Then the reduced diagram $\Delta$ consists of only one row. Hence, for an assignment $\varphi$ the condition (1) for the corresponding normalization condition $N_\varphi$ is void. Now, if we use the assignment $\varphi_0$ as above, then for a normal frame $\Gamma$ the only possibly nonzero blocks $\Big(C_\Gamma(t)\Big)_{ba}$ for its structure function (where $b$ is not located to the right of $a$) are when $(b, a) = (m(e), e)$ or $(b, a) = (m \circ r(e), e)$, where $e \in \Delta$. Moreover, the matrices $\Big(C_\Gamma(t)\Big)_{m(e)e}$ are symmetric and the matrices $\Big(C_\Gamma(t)\Big)_{m\circ r(e)e}$ are skew-symmetric.

**Example A.33.** (*The case of rank* 1 *curves in Lagrangian Grassmannians*) Let $\Lambda$ be an equiregular and ample and that

$$\dim \Lambda^{(-1)} - \dim \Lambda = 1. \tag{A.86}$$

The last condition is equivalent to the fact that the rank of the linear map $\frac{d}{dt}\Lambda(t)$ is equal to 1. Such curves are called *rank* 1 *curves* in Lagrangian Grassmannians and they appear as Jacobi curves of sub-Riemannian structures on rank 2 distributions. From (A.86) and the assumptions that the curve is ample and equiregular it follows that

$$\dim \Lambda^{(j-1)} - \dim \Lambda^{(j)} = 1, \quad 0 \leq -j < \frac{1}{2}\dim V.$$

Hence, the Young diagram $D$ of $\Lambda$ consists of one row of length $\frac{1}{2}\dim V$, i.e., this is a particular case of the previous example. In this case the corresponding matrices $\Big(C_\Gamma(t)\Big)_{ba}$ are $1 \times 1$ matrix valued functions, i.e., they are usual (scalar-valued) functions. and if we use the normalization condition $N_{\varphi_0}$, then the only possibly nonzero entries $\Big(C_\Gamma(t)\Big)_{ba}$ of the structure functions with $b$ located not to the right of $a$ are $\Big(C_\Gamma(t)\Big)_{m(e)e}$ with $e \in \Delta$, as a skew-symmmetric $1 \times 1$ matrices are zero. Besides, by Remark A.27 the group $\mathrm{Sym}^{\mathrm{ne}}_{F(\delta)}$ is isomorphic to $\{\pm I\}$ and there are exactly two normal frames which differ by a sign. The tuple of functions

$$\left\{\Big(C_\Gamma(t)\Big)_{m(e)e}\right\}_{e\in\Delta} \tag{A.87}$$

for the complete system of invariants of the curve $\Lambda$, i.e., two curves are $Sp(V)$-equivalent if and only if the corresponding tuples as in (A.87) are equal.

# Bibliography


[AB12] Andrei Agrachev and Davide Barilari. Sub-Riemannian structures on 3D Lie groups. *J. Dyn. and Contr. Syst.*, 18(1):21–44, 2012.

[ABB12] Andrei Agrachev, Davide Barilari, and Ugo Boscain. On the Hausdorff volume in sub-Riemannian geometry. *Calc. Var. and PDE's*, 43(3-4):355–388, 2012.

[ABC$^+$10] A. A. Agrachev, U. Boscain, G. Charlot, R. Ghezzi, and M. Sigalotti. Two-dimensional almost-Riemannian structures with tangency points. *Ann. Inst. H. Poincaré Anal. Non Linéaire*, 27(3):793–807, 2010.

[ABCK97] A. Agrachev, B. Bonnard, M. Chyba, and I. Kupka. Sub-Riemannian sphere in Martinet flat case. *ESAIM Control Optim. Calc. Var.*, 2:377–448 (electronic), 1997.

[ABGR09] Andrei Agrachev, Ugo Boscain, Jean-Paul Gauthier, and Francesco Rossi. The intrinsic hypoelliptic Laplacian and its heat kernel on unimodular Lie groups. *J. Funct. Anal.*, 256(8):2621–2655, 2009.

[ABR17] Andrei Agrachev, Davide Barilari, and Luca Rizzi. Sub-Riemannian curvature in contact geometry. *J. Geom. Anal.*, 27(1):366–408, 2017.

[ABR18] A. Agrachev, D. Barilari, and L. Rizzi. Curvature: a variational approach. *Memoirs of the AMS*, 256(1225):142, 2018.

[ABS08] Andrei Agrachev, Ugo Boscain, and Mario Sigalotti. A Gauss-Bonnet-like formula on two-dimensional almost-Riemannian manifolds. *Discrete Contin. Dyn. Syst.*, 20(4):801–822, 2008.

[ACEAG98] A. A. Agrachev, El-H. Chakir El-A., and J. P. Gauthier. Sub-Riemannian metrics on $\mathbf{R}^3$. In *Geometric control and non-holonomic mechanics (Mexico City, 1996)*, volume 25 of *CMS Conf. Proc.*, pages 29–78. Amer. Math. Soc., Providence, RI, 1998.

[ACGZ00] A. A. Agrachev, G. Charlot, J. P. A. Gauthier, and V. M. Zakalyukin. On sub-Riemannian caustics and wave fronts for contact distributions in the three-space. *J. Dynam. Control Systems*, 6(3):365–395, 2000.

[AG76] A. A. Agračev and R. V. Gamkrelidze. The principle of second order optimality for time-optimal problems. *Mat. Sb. (N.S.)*, 100(142)(4):610–643, 648, 1976.

[AG78] A. A. Agrachev and R. V. Gamkrelidze. Exponential representation of flows and a chronological enumeration. *Mat. Sb. (N.S.)*, 107(149)(4):467–532, 639, 1978.

[AG80] A. A. Agrachev and R. V. Gamkrelidze. Chronological algebras and nonstationary vector fields. In *Problems in geometry, Vol. 11 (Russian)*, pages 135–176, 243. Akad. Nauk SSSR, Vsesoyuz. Inst. Nauchn. i Tekhn. Informatsii, Moscow, 1980.

[AG97] A. A. Agrachev and R. V. Gamkrelidze. Feedback-invariant optimal control theory and differential geometry. I. Regular extremals. *J. Dynam. Control Systems*, 3(3):343–389, 1997.

[AG98] A. Agrachev and R. Gamkrelidze. Symplectic methods for optimization and control. In *Geometry of feedback and optimal control*, volume 207 of *Monogr. Textbooks Pure Appl. Math.*, pages 19–77. Dekker, New York, 1998.

[AG99] Andrei A. Agrachev and Jean-Paul A. Gauthier. On the Dido problem and plane isoperimetric problems. *Acta Appl. Math.*, 57(3):287–338, 1999.

[AG01a] Andrei Agrachev and Jean-Paul Gauthier. On the subanalyticity of Carnot-Caratheodory distances. *Ann. Inst. H. Poincaré Anal. Non Linéaire*, 18(3):359–382, 2001.

[AG01b] Andrei Agrachev and Jean-Paul Gauthier. Subanalyticity of distance and spheres in sub-Riemannian geometry. In *Nonlinear control in the year 2000, Vol. 1 (Paris)*, volume 258 of *Lect. Notes Control Inf. Sci.*, pages 1–8. Springer, London, 2001.

[Agr77] A. A. Agračev. A necessary condition for second order optimality in the general nonlinear case. *Mat. Sb. (N.S.)*, 102(144)(4):551–568, 632, 1977.

[Agr95] Andrei A. Agrachev. Methods of control theory in nonholonomic geometry. In *Proceedings of the International Congress of Mathematicians, Vol. 1, 2 (Zürich, 1994)*, pages 1473–1483, Basel, 1995. Birkhäuser.

[Agr96] A. A. Agrachev. Exponential mappings for contact sub-Riemannian structures. *J. Dynam. Control Systems*, 2(3):321–358, 1996.

[Agr98a] A. Agrachev. Compactness for sub-Riemannian length-minimizers and subanalyticity. *Rend. Sem. Mat. Univ. Politec. Torino*, 56(4):1–12 (2001), 1998. Control theory and its applications (Grado, 1998).

[Agr98b] A. A. Agrachev. Feedback-invariant optimal control theory and differential geometry. II. Jacobi curves for singular extremals. *J. Dynam. Control Systems*, 4(4):583–604, 1998.

[Agr98c] A. A. Agrachev. On the equivalence of different types of local minima in sub-riemannian problems. In *Proceedings of the 37th IEEE Conference on Decision and Control (Cat. No.98CH36171)*, volume 2, pages 2240–2243 vol.2, 1998.

[Agr08] A. A. Agrachev. Geometry of optimal control problems and Hamiltonian systems. In *Nonlinear and optimal control theory*, volume 1932 of *Lecture Notes in Math.*, pages 1–59. Springer, Berlin, 2008.

[Agr09] A. A. Agrachëv. Any sub-Riemannian metric has points of smoothness. *Dokl. Akad. Nauk*, 424(3):295–298, 2009.

[Agr14] Andrei A. Agrachev. *Some open problems*, pages 1–13. Springer International Publishing, Cham, 2014.

[AGS89] A. A. Agrachev, R. V. Gamkrelidze, and A. V. Sarychev. Local invariants of smooth control systems. *Acta Appl. Math.*, 14(3):191–237, 1989.

[Ahd89] Shahla Ahdout. Fanning curves of Lagrangian manifolds and geodesic flows. *Duke Math. J.*, 59(2):537–552, 1989.

[AL09] Andrei Agrachev and Paul Lee. Optimal transportation under nonholonomic constraints. *Trans. Amer. Math. Soc.*, 361(11):6019–6047, 2009.

[AL14] Andrei Agrachev and Paul W. Y. Lee. Generalized ricci curvature bounds for three dimensional contact subriemannian manifolds. *Mathematische Annalen*, 360(1):209–253, Oct 2014.

[AM03] A. Agrachev and A. Marigo. Nonholonomic tangent spaces: intrinsic construction and rigid dimensions. *Electron. Res. Announc. Amer. Math. Soc.*, 9:111–120, 2003.

[APD09] J. C. Álvarez Paiva and Carlos E. Durán. Geometric invariants of fanning curves. *Adv. in Appl. Math.*, 42(3):290–312, 2009.

[Arn89] V. I. Arnol′d. *Mathematical methods of classical mechanics*, volume 60 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, second edition, 1989. Translated from the Russian by K. Vogtmann and A. Weinstein.

[AS87] A. A. Agrachev and A. V. Sarychev. Filtrations of a Lie algebra of vector fields and the nilpotent approximation of controllable systems. *Dokl. Akad. Nauk SSSR*, 295(4):777–781, 1987.

[AS95] A. A. Agrachev and A. V. Sarychev. Strong minimality of abnormal geodesics for 2-distributions. *Journal of Dynamical and Control Systems*, 1(2):139–176, Apr 1995.

[AS04] Andrei A. Agrachev and Yuri L. Sachkov. *Control theory from the geometric viewpoint*, volume 87 of *Encyclopaedia of Mathematical Sciences*. Springer-Verlag, Berlin, 2004. Control Theory and Optimization, II.

[Aud94] Michèle Audin. Courbes algébriques et systèmes intégrables: géodésiques des quadriques. *Exposition. Math.*, 12(3):193–226, 1994.

[AW86] A. M. Arthurs and G. R. Walsh. On Hammersley's minimum problem for a rolling sphere. *Math. Proc. Cambridge Philos. Soc.*, 99(3):529–534, 1986.

[AZ02a] A. Agrachev and I. Zelenko. Geometry of Jacobi curves. II. *J. Dynam. Control Systems*, 8(2):167–215, 2002.

[AZ02b] A. A. Agrachev and I. Zelenko. Geometry of Jacobi curves. I. *J. Dynam. Control Systems*, 8(1):93–140, 2002.

[BA88] G. Ben Arous. Développement asymptotique du noyau de la chaleur hypoelliptique hors du cut-locus. *Ann. Sci. École Norm. Sup. (4)*, 21(3):307–331, 1988.

[BA89] Gérard Ben Arous. Développement asymptotique du noyau de la chaleur hypoelliptique sur la diagonale. *Ann. Inst. Fourier (Grenoble)*, 39(1):73–99, 1989.

[BAL91a] G. Ben Arous and R. Léandre. Décroissance exponentielle du noyau de la chaleur sur la diagonale. II. *Probab. Theory Related Fields*, 90(3):377–402, 1991.

[BAL91b] G. Ben Arous and R. Léandre. Décroissance exponentielle du noyau de la chaleur sur la diagonale. II. *Probab. Theory Related Fields*, 90(3):377–402, 1991.

[Bao67] Mohamed Salah Baouendi. Sur une classe d'opérateurs elliptiques dégénérés. *Bull. Soc. Math. France*, 95:45–87, 1967.

[Bar13] D. Barilari. Trace heat kernel asymptotics in 3D contact sub-Riemannian geometry. *J. Math. Sci. (N.Y.)*, 195(3):391–411, 2013. Translation of Sovrem. Mat. Prilozh. No. 82 (2012).

[BB09] Fabrice Baudoin and Michel Bonnefont. The subelliptic heat kernel on SU(2): representations, asymptotics and gradient bounds. *Math. Z.*, 263(3):647–672, 2009.

[BB18] Davide Barilari and Francesco Boarotto. On the set of points of smoothness for the value function of affine optimal control problems. *SIAM J. Control Optim.*, 56(2):649–671, 2018.

[BBCN17] Davide Barilari, Ugo Boscain, Grégoire Charlot, and Robert W. Neel. On the heat diffusion for generic Riemannian and sub-Riemannian structures. *Int. Math. Res. Not. IMRN*, (15):4639–4672, 2017.

[BBG12] Davide Barilari, Ugo Boscain, and Jean-Paul Gauthier. On 2-step, corank 2 sub-Riemannian metrics. *SIAM Journal of Control and Optimization*, 50(1):559–582, 2012.

[BBI01] Dmitri Burago, Yuri Burago, and Sergei Ivanov. *A course in metric geometry*, volume 33 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2001.

[BBLss] D. Barilari, I. Beschastnyi, and A. Lerario. Volume of small balls and sub-Riemannian curvature in 3D contact manifolds. *Journal of Symplectic Geometry*, 2019, in press.

[BBN12] Davide Barilari, Ugo Boscain, and Robert Neel. Small time asymptotics of the heat kernel at the sub-Riemannian cut locus. *J. Differential Geometry*, 92(3):373–416, 2012.

[BBNss] D. Barilari, U. Boscain, and R. W. Neel. Heat kernel asymptotics on sub-Riemannian manifolds with symmetries and applications to the bi-Heisenberg group. *Annales Faculté Sciences Toulouse*, 2019, in press.

[BBS16a] Davide Barilari, Ugo Boscain, and Mario Sigalotti, editors. *Geometry, analysis and dynamics on sub-Riemannian manifolds. Vol. 1.* EMS Series of Lectures in Mathematics. European Mathematical Society (EMS), Zürich, 2016. Lecture notes from the IHP Trimester held at the Institut Henri Poincaré, Paris and from the CIRM Summer School "Sub-Riemannian Manifolds: From Geodesics to Hypoelliptic Diffusion" held in Luminy, Fall 2014.

[BBS16b] Davide Barilari, Ugo Boscain, and Mario Sigalotti, editors. *Geometry, analysis and dynamics on sub-Riemannian manifolds. Vol. II.* EMS Series of Lectures in Mathematics. European Mathematical Society (EMS), Zürich, 2016. Lecture notes from the IHP Trimester held at the Institut Henri Poincaré, Paris and from the CIRM Summer School "Sub-Riemannian Manifolds: From Geodesics to Hypoelliptic Diffusion" held in Luminy, Fall 2014.

[BC03] Bernard Bonnard and Monique Chyba. *Singular trajectories and their role in control theory*, volume 40 of *Mathématiques & Applications (Berlin) [Mathematics & Applications]*. Springer-Verlag, Berlin, 2003.

[BC14] B. Bonnard and J.-B. Caillau. Metrics with equatorial singularities on the sphere. *Ann. Mat. Pura Appl. (4)*, 193(5):1353–1382, 2014.

[BCC05] Ugo Boscain, Thomas Chambrion, and Grégoire Charlot. Nonisotropic 3-level quantum systems: complete solutions for minimum time and minimum energy. *Discrete Contin. Dyn. Syst. Ser. B*, 5(4):957–990 (electronic), 2005.

[BCG02a] U. Boscain, T. Chambrion, and J.-P. Gauthier. On the $K + P$ problem for a three-level quantum system: optimality implies resonance. *J. Dynam. Control Systems*, 8(4):547–572, 2002.

[BCG$^+$02b] Ugo Boscain, Grégoire Charlot, Jean-Paul Gauthier, Stéphane Guérin, and Hans-Rudolf Jauslin. Optimal control in laser-induced population transfer for two- and three-level quantum systems. *J. Math. Phys.*, 43(5):2107–2132, 2002.

[BCG13] U. Boscain, G. Charlot, and R. Ghezzi. Normal forms and invariants for 2-dimensional almost-Riemannian structures. *Differential Geom. Appl.*, 31(1):41–62, 2013.

[BCGJ11] B. Bonnard, G. Charlot, R. Ghezzi, and G. Janin. The sphere and the cut locus at a tangency point in two-dimensional almost-Riemannian geometry. *J. Dyn. Control Syst.*, 17(1):141–161, 2011.

[BCGM15] Ugo Boscain, Gregoire Charlot, Moussa Gaye, and Paolo Mason. Local properties of almost-Riemannian structures in dimension 3. *Discrete Contin. Dyn. Syst.*, 35(9):4115–4147, 2015.

[BCGS13] U. Boscain, G. Charlot, R. Ghezzi, and M. Sigalotti. Lipschitz classification of almost-Riemannian distances on compact oriented surfaces. *J. Geom. Anal.*, 23(1):438–455, 2013.

[BCJ$^+$19] D. Barilari, Y. Chitour, F. Jean, D. Prandi, and M. Sigalotti. On the regularity of abnormal minimizers for rank 2 sub-Riemannian structures. *J. Math. Pures Appl.*, 2019.

[BdSR18] André Belotto da Silva and Ludovic Rifford. The Sard conjecture on Martinet surfaces. *Duke Math. J.*, 167(8):1433–1471, 2018.

[Bel96] André Bellaïche. The tangent space in sub-Riemannian geometry. In *Sub-Riemannian geometry*, volume 144 of *Progr. Math.*, pages 1–78. Birkhäuser, Basel, 1996.

[Ber88] V. N. Berestovskiĭ. Homogeneous spaces with an intrinsic metric. *Dokl. Akad. Nauk SSSR*, 301(2):268–271, 1988.

[Ber03] Marcel Berger. *A panoramic view of Riemannian geometry.* Springer-Verlag, Berlin, 2003.

[Bes14] I. Yu. Beschastnyĭ. On the optimal rolling of a sphere with twisting but without slipping. *Mat. Sb.*, 205(2):3–38, 2014.

[BFPR18] A Belotto da Silva, A Figalli, A Parusiński, and L Rifford. Strong Sard Conjecture and regularity of singular minimizing geodesics for analytic sub-Riemannian structures in dimension 3. *ArXiv e-prints*, October 2018.

[BG92] Marcel Berger and Bernard Gostiaux. *Géométrie différentielle: variétés, courbes et surfaces.* Mathématiques. [Mathematics]. Presses Universitaires de France, Paris, second edition, 1992.

[BG13] Ugo Boscain and Jean-Paul Gauthier. On the spherical Hausdorff measure in step 2 corank 2 sub-Riemannian geometry. *SIAM J. Control Optim.*, 51(6):4450–4462, 2013.

[BL13] Ugo Boscain and Camille Laurent. The Laplace-Beltrami operator in almost-Riemannian geometry. *Ann. Inst. Fourier (Grenoble)*, 63(5):1739–1770, 2013.

[Blo15] A. M. Bloch. *Nonholonomic mechanics and control*, volume 24 of *Interdisciplinary Applied Mathematics*. Springer, New York, second edition, 2015. With the collaboration of J. Bailieul, P. E. Crouch, J. E. Marsden and D. Zenkov, With scientific input from P. S. Krishnaprasad and R. M. Murray.

[BLU07] A. Bonfiglioli, E. Lanconelli, and F. Uguzzoni. *Stratified Lie groups and potential theory for their sub-Laplacians.* Springer Monographs in Mathematics. Springer, Berlin, 2007.

[BNR17] Ugo Boscain, Robert Neel, and Luca Rizzi. Intrinsic random walks and sub-Laplacians in sub-Riemannian geometry. *Adv. Math.*, 314:124–184, 2017.

[Bon12] Michel Bonnefont. The subelliptic heat kernels on $\mathrm{SL}(2,\mathbb{R})$ and on its universal covering $\widetilde{\mathrm{SL}(2,\mathbb{R})}$: integral representations and some functional inequalities. *Potential Anal.*, 36(2):275–300, 2012.

[Boo86] William M. Boothby. *An introduction to differentiable manifolds and Riemannian geometry*, volume 120 of *Pure and Applied Mathematics*. Academic Press, Inc., Orlando, FL, second edition, 1986.

[Bor06] Max Born. Stabilitat der elastischen linie in ebene und raum. *Preisschrift Und Dissertation*, 1:5–101, 1906.

[BP05] Ugo Boscain and Benedetto Piccoli. A short introduction to optimal control. In *Contrôle Non Linéaire et Applications,* T. Sari, editor, pages 19–66. Hermann, Paris, 2005.

[BP07] Alberto Bressan and Benedetto Piccoli. *Introduction to the mathematical theory of control*, volume 2 of *AIMS Series on Applied Mathematics*. American Institute of Mathematical Sciences (AIMS), Springfield, MO, 2007.

[BR86] Asim O. Barut and Ryszard Rączka. *Theory of group representations and applications*. World Scientific Publishing Co., Singapore, second edition, 1986.

[BR96] André Bellaïche and Jean-Jacques Risler, editors. *Sub-Riemannian geometry*, volume 144 of *Progress in Mathematics*. Birkhäuser Verlag, Basel, 1996.

[BR08] Ugo Boscain and Francesco Rossi. Invariant Carnot-Caratheodory metrics on $S^3$, SO(3), SL(2), and lens spaces. *SIAM J. Control Optim.*, 47(4):1851–1878, 2008.

[BR13] Davide Barilari and Luca Rizzi. A formula for Popp's volume in sub-Riemannian geometry. *Anal. Geom. Metr. Spaces*, 1:42–57, 2013.

[BR17] Davide Barilari and Luca Rizzi. On Jacobi fields and a canonical connection in sub-Riemannian geometry. *Arch. Math. (Brno)*, 53(2):77–92, 2017.

[BR18] Davide Barilari and Luca Rizzi. Sharp measure contraction property for generalized H-type carnot groups. *Communications in Contemporary Mathematics*, 20(6):1750081, 2018.

[Bra14] Marco Bramanti. *An invitation to hypoelliptic operators and Hörmander's vector fields*. SpringerBriefs in Mathematics. Springer, Cham, 2014.

[Bro73] R. W. Brockett. Lie theory and control systems defined on spheres. *SIAM J. Appl. Math.*, 25:213–225, 1973. Lie algebras: applications and computational methods (Conf., Drexel Univ., Philadelphia, Pa., 1972).

[Bro82] R. W. Brockett. Control theory and singular Riemannian geometry. In *New directions in applied mathematics (Cleveland, Ohio, 1980)*, pages 11–27. Springer, New York, 1982.

[Bro84] R. W. Brockett. Nonlinear control theory and differential geometry. In *Proceedings of the International Congress of Mathematicians, Vol. 1, 2 (Warsaw, 1983)*, pages 1357–1368. PWN, Warsaw, 1984.

[Bro99] Roger W Brockett. Explicitly solvable control problems with nonholonomic constraints. In *Decision and Control, 1999. Proceedings of the 38th IEEE Conference on*, volume 1, pages 13–16. IEEE, 1999.

[BS90] Rosa Maria Bianchini and Gianna Stefani. Graded approximations and controllability along a trajectory. *SIAM J. Control Optim.*, 28(4):903–924, 1990.

[BZ15a] V. N. Berestovskiĭ and I. A. Zubareva. Geodesics and shortest arcs of a special sub-Riemannian metric on the Lie group SO(3). *Sibirsk. Mat. Zh.*, 56(4):762–774, 2015.

[BZ15b] V. N. Berestovskiĭ and I. A. Zubareva. Sub-Riemannian distance in the Lie groups SU(2) and SO(3). *Mat. Tr.*, 18(2):3–21, 2015.

[BZ16] V. N. Berestovskiĭ and I. A. Zubareva. Geodesics and shortest arcs of a special sub-Riemannian metric on the Lie group $SL(2)$. *Sibirsk. Mat. Zh.*, 57(3):527–542, 2016.

[Car09] C. Carathéodory. Untersuchungen über die Grundlagen der Thermodynamik. *Math. Ann.*, 67(3):355–386, 1909.

[Car33] Élie Cartan. Sur la géométrie pseudo-conforme des hypersurfaces de l'espace de deux variables complexes. *Ann. Mat. Pura Appl.*, 11(1):17–90, 1933.

[Car67] Henri Cartan. *Calcul différentiel*. Hermann, Paris, 1967.

[CDPT07] Luca Capogna, Donatella Danielli, Scott D. Pauls, and Jeremy T. Tyson. *An introduction to the Heisenberg group and the sub-Riemannian isoperimetric problem*, volume 259 of *Progress in Mathematics*. Birkhäuser Verlag, Basel, 2007.

[CF10] Thomas Cass and Peter Friz. Densities for rough differential equations under Hörmander's condition. *Ann. of Math. (2)*, 171(3):2115–2141, 2010.

[Cha06] Isaac Chavel. *Riemannian geometry*, volume 98 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, second edition, 2006. A modern introduction.

[Che55] Claude Chevalley. *Théorie des groupes de Lie. Tome III. Théorèmes généraux sur les algèbres de Lie*. Actualités Sci. Ind. no. 1226. Hermann & Cie, Paris, 1955.

[CHLT15] Thomas Cass, Martin Hairer, Christian Litterer, and Samy Tindel. Smoothness of the density for solutions to Gaussian rough differential equations. *Ann. Probab.*, 43(1):188–239, 2015.

[Cho39] Wei-Liang Chow. Über Systeme von linearen partiellen Differentialgleichungen erster Ordnung. *Math. Ann.*, 117:98–105, 1939.

[CJT06] Y. Chitour, F. Jean, and E. Trélat. Genericity results for singular curves. *J. Differential Geom.*, 73(1):45–73, 2006.

[Cla90] F. H. Clarke. *Optimization and nonsmooth analysis*, volume 5 of *Classics in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, second edition, 1990.

[CR08] P. Cannarsa and L. Rifford. Semiconcavity results for optimal control problems admitting no singular minimizing controls. *Ann. Inst. H. Poincaré Anal. Non Linéaire*, 25(4):773–802, 2008.

[CTDL92] C. Cohen-Tannoudji, B. Diu, and F. Laloe. *Quantum Mechanics, 2 Volume Set*. Wiley, 1992.

[dC76] Manfredo P. do Carmo. *Differential geometry of curves and surfaces*. Prentice-Hall, Inc., Englewood Cliffs, N.J., 1976. Translated from the Portuguese.

[dC92] Manfredo Perdigão do Carmo. *Riemannian geometry*. Mathematics: Theory & Applications. Birkhäuser Boston, Inc., Boston, MA, 1992. Translated from the second Portuguese edition by Francis Flaherty.

[Die60] J. Dieudonné. *Foundations of modern analysis.* Pure and Applied Mathematics, Vol. X. Academic Press, New York-London, 1960.

[DZ12] Boris Doubrov and Igor Zelenko. On geometry of curves of flags of constant type. *Cent. Eur. J. Math.*, 10(5):1836–1871, 2012.

[DZ13] Boris Doubrov and Igor Zelenko. Geometry of curves in generalized flag varieties. *Transform. Groups*, 18(2):361–383, 2013.

[EAGK96] El-H. Ch. El-Alaoui, J.-P. Gauthier, and I. Kupka. Small sub-Riemannian balls on $\mathbf{R}^3$. *J. Dynam. Control Systems*, 2(3):359–421, 1996.

[Eul] Leonhard Euler. De Miris Proprietatibvs Cvrvae Elasticae.

[Eva98] Lawrence C. Evans. *Partial differential equations*, volume 19 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 1998.

[Fed69] Herbert Federer. *Geometric measure theory.* Die Grundlehren der mathematischen Wissenschaften, Band 153. Springer-Verlag New York Inc., New York, 1969.

[FG96] Elisha Falbel and Claudio Gorodski. Sub-Riemannian homogeneous spaces in dimensions 3 and 4. *Geom. Dedicata*, 62(3):227–252, 1996.

[FL83] Bruno Franchi and Ermanno Lanconelli. Une métrique associée à une classe d'opérateurs elliptiques dégénérés. *Rend. Sem. Mat. Univ. Politec. Torino*, (Special Issue):105–114 (1984), 1983. Conference on linear partial and pseudodifferential operators (Torino, 1982).

[Fla70] Harley Flanders. The Schwarzian as a curvature. *J. Differential Geometry*, 4:515–519, 1970.

[Fol73] G. B. Folland. A fundamental solution for a subelliptic operator. *Bull. Amer. Math. Soc.*, 79:373–376, 1973.

[Fol75] G. B. Folland. Subelliptic estimates and function spaces on nilpotent Lie groups. *Ark. Mat.*, 13(2):161–207, 1975.

[FOT94] Masatoshi Fukushima, Yōichi Ōshima, and Masayoshi Takeda. *Dirichlet forms and symmetric Markov processes*, volume 19 of *De Gruyter Studies in Mathematics*. Walter de Gruyter & Co., Berlin, 1994.

[FS74] G. B. Folland and E. M. Stein. Estimates for the $\bar{\partial}_b$ complex and analysis on the Heisenberg group. *Comm. Pure Appl. Math.*, 27:429–522, 1974.

[FSSC15] Bruno Franchi, Raul Paolo Serapioni, and Francesco Serra Cassano. Area formula for centered Hausdorff measures in metric spaces. *Nonlinear Anal.*, 126:218–233, 2015.

[Gaf54] Matthew P. Gaffney. The heat equation method of Milgram and Rosenbloom for open Riemannian manifolds. *Ann. of Math. (2)*, 60:458–466, 1954.

[Gaf55] Matthew P. Gaffney. Hilbert space methods in the theory of harmonic integrals. *Trans. Amer. Math. Soc.*, 78:426–444, 1955.

[Gam14] Revaz V. Gamkrelidze. Differential-geometric and invariance properties of the equations of maximum principle (MP). In *Geometric control theory and sub-Riemannian geometry*, volume 5 of *Springer INdAM Ser.*, pages 167–175. Springer, Cham, 2014.

[Gar16] Nicola Garofalo. Hypoelliptic operators and some aspects of analysis and geometry of sub-Riemannian spaces. In *Geometry, analysis and dynamics on sub-Riemannian manifolds. Vol. 1*, EMS Ser. Lect. Math., pages 123–257. Eur. Math. Soc., Zürich, 2016.

[Gav77] Bernard Gaveau. Principe de moindre action, propagation de la chaleur et estimées sous elliptiques sur certains groupes nilpotents. *Acta Math.*, 139(1-2):95–153, 1977.

[GG73] M. Golubitsky and V. Guillemin. *Stable mappings and their singularities*. Springer-Verlag, New York-Heidelberg, 1973. Graduate Texts in Mathematics, Vol. 14.

[GHL90] Sylvestre Gallot, Dominique Hulin, and Jacques Lafontaine. *Riemannian geometry*. Universitext. Springer-Verlag, Berlin, second edition, 1990.

[GJ14] Roberta Ghezzi and Frédéric Jean. Hausdorff measure and dimensions in non equiregular sub-Riemannian manifolds. In *Geometric control theory and sub-Riemannian geometry*, volume 5 of *Springer INdAM Ser.*, pages 201–218. Springer, Cham, 2014.

[GJ15] R. Ghezzi and F. Jean. Hausdorff volume in non equiregular sub-Riemannian manifolds. *Nonlinear Anal.*, 126:345–377, 2015.

[Goh66] B. S. Goh. Necessary conditions for singular extremals involving multiple control variables. *SIAM J. Control*, 4:716–731, 1966.

[Goo76] Roe W. Goodman. *Nilpotent Lie groups: structure and applications to analysis*. Lecture Notes in Mathematics, Vol. 562. Springer-Verlag, Berlin-New York, 1976.

[Gro81] Mikhael Gromov. *Structures métriques pour les variétés riemanniennes*, volume 1 of *Textes Mathématiques [Mathematical Texts]*. CEDIC, Paris, 1981. Edited by J. Lafontaine and P. Pansu.

[Gro96] Mikhael Gromov. Carnot-Carathéodory spaces seen from within. In *Sub-Riemannian geometry*, volume 144 of *Progr. Math.*, pages 79–323. Birkhäuser, Basel, 1996.

[Gru70] V. V. Grušin. A certain class of hypoelliptic operators. *Mat. Sb. (N.S.)*, 83 (125):456–473, 1970.

[GV88] V. Gershkovich and A. Vershik. Nonholonomic manifolds and nilpotent analysis. *J. Geom. Phys.*, 5(3):407–452, 1988.

[Had98] J. Hadamard. Les surfaces a courbures opposees et leurs lignes geodesique. *J. Math. Pures Appl.*, 4:27–73, 1898.

[Hai11] Martin Hairer. On Malliavin's proof of Hörmander's theorem. *Bull. Sci. Math.*, 135(6-7):650–666, 2011.

[Ham83] J. M. Hammersley. Oxford commemoration ball. In *Probability, statistics and analysis*, volume 79 of *London Math. Soc. Lecture Note Ser.*, pages 112–142. Cambridge Univ. Press, Cambridge-New York, 1983.

[Hat02] Allen Hatcher. *Algebraic topology*. Cambridge University Press, Cambridge, 2002.

[Hir76] Morris W. Hirsch. *Differential topology*. Springer-Verlag, New York-Heidelberg, 1976. Graduate Texts in Mathematics, No. 33.

[HLD16] Eero Hakavuori and Enrico Le Donne. Non-minimality of corners in subriemannian geometry. *Invent. Math.*, 206(3):693–704, 2016.

[Hör67] Lars Hörmander. Hypoelliptic second order differential equations. *Acta Math.*, 119:147–171, 1967.

[HS74] Morris W. Hirsch and Stephen Smale. *Differential equations, dynamical systems, and linear algebra*. Academic Press [A subsidiary of Harcourt Brace Jovanovich, Publishers], New York-London, 1974. Pure and Applied Mathematics, Vol. 60.

[Hug95] Walker Keener Hughen. *The sub-Riemannian geometry of three-manifolds*. ProQuest LLC, Ann Arbor, MI, 1995. Thesis (Ph.D.)–Duke University.

[Hul76] A. Hulanicki. The distribution of energy in the Brownian motion in the Gaussian field and analytic-hypoellipticity of certain subelliptic operators on the Heisenberg group. *Studia Math.*, 56(2):165–173, 1976.

[IK04] Jin-ichi Itoh and Kazuyoshi Kiyohara. The cut loci and the conjugate loci on ellipsoids. *Manuscripta Math.*, 114(2):247–264, 2004.

[IV15] Jin-ichi ITOH and Costin VILCU. Every graph is a cut locus. *J. Math. Soc. Japan*, 67(3):1227–1238, 07 2015.

[Jac39] C. G. J. Jacobi. Note von der geodätischen Linie auf einem Ellipsoid und den verschiedenen Anwendungen einer merkwürdigen analytischen Substitution. *J. Reine Angew. Math.*, 19:309–313, 1839.

[Jac62] Nathan Jacobson. *Lie algebras*. Interscience Tracts in Pure and Applied Mathematics, No. 10. Interscience Publishers (a division of John Wiley & Sons), New York-London, 1962.

[Jac99] S. Jacquet. Subanalyticity of the sub-Riemannian distance. *J. Dynam. Control Systems*, 5(3):303–328, 1999.

[Jea14] Frédéric Jean. *Control of nonholonomic systems: from sub-Riemannian geometry to motion planning*. SpringerBriefs in Mathematics. Springer, Cham, 2014.

[JSC87] David Jerison and Antonio Sánchez-Calle. Subelliptic, second order differential operators. In *Complex analysis, III (College Park, Md., 1985–86)*, volume 1277 of *Lecture Notes in Math.*, pages 46–77. Springer, Berlin, 1987.

[Jur93] V. Jurdjevic. The geometry of the plate-ball problem. *Arch. Rational Mech. Anal.*, 124(4):305–328, 1993.

[Jur97] Velimir Jurdjevic. *Geometric control theory*, volume 52 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 1997.

[Jur99] V. Jurdjevic. Optimal control, geometry, and mechanics. In *Mathematical control theory*, pages 227–267. Springer, New York, 1999.

[Jur01] V. Jurdjevic. Hamiltonian point of view of non-Euclidean geometry and elliptic functions. *Systems Control Lett.*, 43(1):25–41, 2001. Lie theory and its applications in control (Würzburg, 1999).

[Jur16] Velimir Jurdjevic. *Optimal Control and Geometry: Integrable Systems*. Cambridge University Press, Cambridge, 2016.

[K1̈5] Wolfgang Kühnel. *Differential geometry*, volume 77 of *Student Mathematical Library*. American Mathematical Society, Providence, RI, 2015. Curves—surfaces—manifolds.

[Kat95] Tosio Kato. *Perturbation theory for linear operators*. Classics in Mathematics. Springer-Verlag, Berlin, 1995. Reprint of the 1980 edition.

[Kaw99] M. Kawski. Chronological algebras: combinatorics and control. In *Geometric control theory (Russian) (Moscow, 1998)*, volume 64 of *Itogi Nauki Tekh. Ser. Sovrem. Mat. Prilozh. Temat. Obz.*, pages 144–178. Vseross. Inst. Nauchn. i Tekhn. Inform. (VINITI), Moscow, 1999.

[Kaw02] Matthias Kawski. The combinatorics of nonlinear controllability and noncommuting flows. In *Mathematical control theory, Part 1, 2 (Trieste, 2001)*, ICTP Lect. Notes, VIII, pages 223–311. Abdus Salam Int. Cent. Theoret. Phys., Trieste, 2002.

[Kaw12] Matthias Kawski. Chronological calculus in systems and control theory. In *Mathematics of complexity and dynamical systems. Vols. 1–3*, pages 88–101. Springer, New York, 2012.

[KKM67] Henry J. Kelley, Richard E. Kopp, and H. Gardner Moyer. Singular extremals. In *Topics in Optimization*, pages 63–101. Academic Press, New York, 1967.

[KL15] Robert J. Kipka and Yuri S. Ledyaev. Extension of chronological calculus for dynamical systems on manifolds. *J. Differential Equations*, 258(5):1765–1790, 2015.

[Knö80] Horst Knörrer. Geodesics on the ellipsoid. *Invent. Math.*, 59(2):119–143, 1980.

[Kre73] Arthur J. Krener. The high order maximal principle. In D. Q. Mayne and R. W. Brockett, editors, *Geometric Methods in System Theory*, pages 174–184, Dordrecht, 1973. Springer Netherlands.

[Kre77] Arthur J. Krener. The high order maximal principle and its application to singular extremals. *SIAM J. Control Optimization*, 15(2):256–293, 1977.

[KS97] Matthias Kawski and Héctor J. Sussmann. Noncommutative power series and formal Lie-algebraic techniques in nonlinear control theory. In *Operators, systems, and linear algebra (Kaiserslautern, 1997)*, European Consort. Math. Indust., pages 111–128. Teubner, Stuttgart, 1997.

[L9́2] Rémi Léandre. Développement asymptotique de la densité d'une diffusion dégénérée. *Forum Math.*, 4(1):45–75, 1992.

[Law89] Derek F. Lawden. *Elliptic functions and applications*, volume 80 of *Applied Mathematical Sciences*. Springer-Verlag, New York, 1989.

[LDMO+16] Enrico Le Donne, Richard Montgomery, Alessandro Ottazzi, Pierre Pansu, and Davide Vittone. Sard property for the endpoint map on some Carnot groups. *Ann. Inst. H. Poincaré Anal. Non Linéaire*, 33(6):1639–1666, 2016.

[Lee97] John M. Lee. *Riemannian manifolds*, volume 176 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1997. An introduction to curvature.

[Lee13] John M. Lee. *Introduction to smooth manifolds*, volume 218 of *Graduate Texts in Mathematics*. Springer, New York, second edition, 2013.

[LM08] Gian Paolo Leonardi and Roberto Monti. End-point equations and regularity of sub-Riemannian geodesics. *Geom. Funct. Anal.*, 18(2):552–582, 2008.

[LS95] Wensheng Liu and Héctor J. Sussman. Shortest paths for sub-Riemannian metrics on rank-two distributions. *Mem. Amer. Math. Soc.*, 118(564):x+104, 1995.

[M7́6] Guy Métivier. Fonction spectrale et valeurs propres d'une classe d'opérateurs non elliptiques. *Comm. Partial Differential Equations*, 1(5):467–519, 1976.

[Mey12] Robert A. Meyers, editor. *Mathematics of complexity and dynamical systems. Vols. 1–3*. Springer, New York, 2012. Selections from ıt The encyclopedia of complexity and systems science [Springer, New York, 2009].

[Mit85] John Mitchell. On Carnot-Carathéodory metrics. *J. Differential Geom.*, 21(1):35–45, 1985.

[MM95] G. A. Margulis and G. D. Mostow. The differential of a quasi-conformal mapping of a Carnot-Carathéodory space. *Geom. Funct. Anal.*, 5(2):402–433, 1995.

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

[MM17] Annamaria Montanari and Daniele Morbidelli. On the subRiemannian cut locus in a model of free two-step Carnot group. *Calc. Var. Partial Differential Equations*, 56(2):Art. 36, 26, 2017.

[Mol75] S. A. Molčanov. Diffusion processes, and Riemannian geometry. *Uspehi Mat. Nauk*, 30(1(181)):3–59, 1975.

[Mon94] Richard Montgomery. Abnormal minimizers. *SIAM J. Control Optim.*, 32(6):1605–1620, 1994.

[Mon96] Richard Montgomery. Survey of singular geodesics. In *Sub-Riemannian geometry*, volume 144 of *Progr. Math.*, pages 325–339. Birkhäuser, Basel, 1996.

[Mon02] Richard Montgomery. *A tour of subriemannian geometries, their geodesics and applications*, volume 91 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2002.

[Mon14] Roberto Monti. Regularity results for sub-riemannian geodesics. *Calculus of Variations and Partial Differential Equations*, 49(1):549–582, Jan 2014.

[Mos80a] J. Moser. Geometry of quadrics and spectral theory. In *The Chern Symposium 1979 (Proc. Internat. Sympos., Berkeley, Calif., 1979)*, pages 147–188. Springer, New York-Berlin, 1980.

[Mos80b] J. Moser. Various aspects of integrable Hamiltonian systems. In *Dynamical systems (C.I.M.E. Summer School, Bressanone, 1978)*, volume 8 of *Progr. Math.*, pages 233–289. Birkhäuser, Boston, Mass., 1980.

[MPAM06] F. Monroy-Pérez and A. Anzaldo-Meneses. The step-2 nilpotent $(n, n(n+1)/2)$ sub-Riemannian geometry. *J. Dyn. Control Syst.*, 12(2):185–216, 2006.

[MS10] Igor Moiseev and Yuri L. Sachkov. Maxwell strata in sub-Riemannian problem on the group of motions of a plane. *ESAIM Control Optim. Calc. Var.*, 16:380–399, 2010.

[MS15] A. P. Mashtakov and Yu. L. Sachkov. Superintegrability of left-invariant sub-Riemannian structures on unimodular three-dimensional Lie groups. *Differ. Equ.*, 51(11):1476–1483, 2015. Translation of Differ. Uravn. **51** (2015), no. 11, 1482–1488.

[Mya02] O. Myasnichenko. Nilpotent $(3, 6)$ sub-Riemannian problem. *J. Dynam. Control Systems*, 8(4):573–597, 2002.

[Mya06] Oleg Myasnichenko. Nilpotent $(n, n(n+1)/2)$ sub-Riemannian problem. *J. Dyn. Control Syst.*, 12(1):87–95, 2006.

[Mye36] Sumner Byron Myers. Connections between differential geometry and topology II. Closed surfaces. *Duke Math. J.*, 2(1):95–102, 1936.

[Nac82] Adrian I. Nachman. The wave equation on the Heisenberg group. *Comm. Partial Differential Equations*, 7(6):675–714, 1982.

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

[NSW85] Alexander Nagel, Elias M. Stein, and Stephen Wainger. Balls and metrics defined by vector fields. I. Basic properties. *Acta Math.*, 155(1-2):103–147, 1985.

[OV17] A. Ottazzi and D. Vittone. On the codimension of the abnormal set in step two Carnot groups. *ArXiv e-prints*, September 2017.

[Ovs89] V. Yu. Ovsienko. A Lagrangian Schwarzian derivative. *Vestnik Moskov. Univ. Ser. I Mat. Mekh.*, (6):42–45, 98, 1989.

[Ovs93] Valentin Ovsienko. Lagrange Schwarzian derivative and symplectic Sturm theory. *Ann. Fac. Sci. Toulouse Math. (6)*, 2(1):73–96, 1993.

[Pan89] Pierre Pansu. Métriques de Carnot-Carathéodory et quasiisométries des espaces symétriques de rang un. *Ann. of Math. (2)*, 129(1):1–60, 1989.

[PBGM62] L. S. Pontryagin, V. G. Boltyanskii, R. V. Gamkrelidze, and E. F. Mishchenko. *The mathematical theory of optimal processes.* Translated from the Russian by K. N. Trirogoff; edited by L. W. Neustadt. Interscience Publishers John Wiley & Sons, Inc. New York-London, 1962.

[Pet16] Peter Petersen. *Riemannian geometry*, volume 171 of *Graduate Texts in Mathematics*. Springer, Cham, third edition, 2016.

[Poi05] Henri Poincaré. Sur les lignes géodésiques des surfaces convexes. *Trans. Amer. Math. Soc.*, 6(3):237–274, 1905.

[Ras38] P.K. Rashevsky. Any two points of a totally nonholonomic space may be connected by an admissible line. *Uch. Zap. Ped Inst. im. Liebknechta*, 2:83–84, 1938.

[Rif04] Ludovic Rifford. A Morse-Sard theorem for the distance function on Riemannian manifolds. *Manuscripta Math.*, 113(2):251–265, 2004.

[Rif06] L. Rifford. À propos des sphères sous-riemanniennes. *Bull. Belg. Math. Soc. Simon Stevin*, 13(3):521–526, 2006.

[Rif14] Ludovic Rifford. *Sub-Riemannian geometry and Optimal Transport.* Springerbriefs in Mathematics, 2014.

[Ros97] Steven Rosenberg. *The Laplacian on a Riemannian manifold*, volume 31 of *London Mathematical Society Student Texts*. Cambridge University Press, Cambridge, 1997. An introduction to analysis on manifolds.

[RS76] Linda Preiss Rothschild and E. M. Stein. Hypoelliptic differential operators and nilpotent groups. *Acta Math.*, 137(3-4):247–320, 1976.

[RS07] Franco Rampazzo and Héctor J. Sussmann. Commutators of flow maps of nonsmooth vector fields. *J. Differential Equations*, 232(1):134–175, 2007.

[RS17] Luca Rizzi and Ulysse Serres. On the cut locus of free, step two Carnot groups. *Proc. Amer. Math. Soc.*, 145(12):5341–5357, 2017.

[RT05] L. Rifford and E. Trélat. Morse-Sard type results in sub-Riemannian geometry. *Math. Ann.*, 332(1):145–159, 2005.

[Rud87] Walter Rudin. *Real and complex analysis.* McGraw-Hill Book Co., New York, third edition, 1987.

[Sac08a] Yu. L. Sachkov. Conjugate points in the Euler elastic problem. *J. Dyn. Control Syst.*, 14(3):409–439, 2008.

[Sac08b] Yu. L. Sachkov. Maxwell strata in the Euler elastic problem. *J. Dyn. Control Syst.*, 14(2):169–234, 2008.

[Sac10] Yuri L. Sachkov. Conjugate and cut time in the sub-Riemannian problem on the group of motions of a plane. *ESAIM Control Optim. Calc. Var.*, 16:1018–1039, 2010.

[Sac11] Yuri L. Sachkov. Cut locus and optimal synthesis in the sub-Riemannian problem on the group of motions of a plane. *ESAIM Control Optim. Calc. Var.*, 17(2):293–321, 2011.

[Sar04] A. V. Sarychev. Lie extensions of nonlinear control systems. *Sovrem. Mat. Prilozh.*, (21, Geom. Zadachi Teor. Upr.):106–133, 2004.

[Sim83] Leon Simon. *Lectures on geometric measure theory*, volume 3 of *Proceedings of the Centre for Mathematical Analysis, Australian National University*. Australian National University, Centre for Mathematical Analysis, Canberra, 1983.

[Spi79] Michael Spivak. *A comprehensive introduction to differential geometry. Vol. I.* Publish or Perish, Inc., Wilmington, Del., second edition, 1979.

[SS14] Yu. L. Sachkov and E. F. Sachkova. Exponential mapping in Euler's elastic problem. *J. Dyn. Control Syst.*, 20(4):443–464, 2014.

[Str86] Robert S. Strichartz. Sub-Riemannian geometry. *J. Differential Geom.*, 24(2):221–263, 1986.

[Str89] Robert S. Strichartz. Corrections to: "Sub-Riemannian geometry" [J. Differential Geom. **24** (1986), no. 2, 221–263; MR0862049 (88b:53055)]. *J. Differential Geom.*, 30(2):595–596, 1989.

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

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

[Sus83] H. J. Sussmann. Lie brackets, real analyticity and geometric control. In *Differential geometric control theory (Houghton, Mich., 1982)*, volume 27 of *Progr. Math.*, pages 1–116. Birkhäuser Boston, Boston, MA, 1983.

[Sus96] Héctor J. Sussmann. A cornucopia of four-dimensional abnormal sub-Riemannian minimizers. In *Sub-Riemannian geometry*, volume 144 of *Progr. Math.*, pages 341–364. Birkhäuser, Basel, 1996.

[Sus08] Hector J. Sussmann. Smooth distributions are globally finitely spanned. In *Analysis and design of nonlinear control systems*, pages 3–8. Springer, Berlin, 2008.

[Tay96] Michael E. Taylor. *Partial differential equations. I*, volume 115 of *Applied Mathematical Sciences*. Springer-Verlag, New York, 1996. Basic theory.

[Tré00] E. Trélat. Some properties of the value function and its level sets for affine control systems with quadratic cost. *Journal of Dynamical and Control Systems*, 6(4):511–541, Oct 2000.

[VG87] A. M. Vershik and V. Ya. Gershkovich. Nonholonomic dynamical systems. Geometry of distributions and variational problems. In *Current problems in mathematics. Fundamental directions, Vol. 16 (Russian)*, Itogi Nauki i Tekhniki, pages 5–85, 307. Akad. Nauk SSSR, Vsesoyuz. Inst. Nauchn. i Tekhn. Inform., Moscow, 1987.

[Vin76] È. B. Vinberg. The Weyl group of a graded Lie algebra. *Izv. Akad. Nauk SSSR Ser. Mat.*, 40(3):488–526, 709, 1976.

[Whi35] J. H. C. Whitehead. On the covering of a complete space by the geodesics through a point. *Annals of Mathematics*, 36(3):679–704, 1935.

[Whi55] Hassler Whitney. On singularities of mappings of euclidean spaces. I. Mappings of the plane into the plane. *Ann. of Math. (2)*, 62:374–410, 1955.

[Zel00] M. I. Zelikin. *Control theory and optimization. I*, volume 86 of *Encyclopaedia of Mathematical Sciences*. Springer-Verlag, Berlin, 2000. Homogeneous spaces and the Riccati equation in the calculus of variations, A translation of ıt Homogeneous spaces and the Riccati equation in the calculus of variations (Russian), "Faktorial", Moscow, 1998, Translation by S. A. Vakhrameev.

[Zel06] Igor Zelenko. On variational approach to differential invariants of rank two distributions. *Differential Geom. Appl.*, 24(3):235–259, 2006.

[ZL07] Igor Zelenko and Chengbo Li. Parametrized curves in Lagrange Grassmannians. *C. R. Math. Acad. Sci. Paris*, 345(11):647–652, 2007.

[ZL09] Igor Zelenko and Chengbo Li. Differential geometry of curves in Lagrange Grassmannians with given Young diagram. *Differential Geom. Appl.*, 27(6):723–742, 2009.

[ZZ95] I. Zelenko and M. Zhitomirskiĭ. Rigid paths of generic 2-distributions on 3-manifolds. *Duke Math. J.*, 79(2):281–307, 1995.

# Index