Introduction to Riemannian and Sub-Riemannian geometry

from Hamiltonian viewpoint

ANDREI AGRACHEV
DAVIDE BARILARI
UGO BOSCAIN

This version: November 17, 2017
Preprint SISSA 09/2012/M

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. The point is 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 restictions (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 ω. In other words, a curve $t \mapsto (x(t), y(t), z(t))$ is an admissible path if and only if $\dot{z}(t) = \frac{1}{2}(y(t)\dot{x}(t) - x(t)\dot{y}(t))$.

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. We see that any smooth planar curve $(x(t), y(t))$ has a unique admissible lift $(x(t), y(t), z(t))$ in $\mathbb{R}^3$, where:

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

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

We see that sub-Riemannian shortest paths are lifts to $\mathbb{R}^3$ of the solutions to 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).

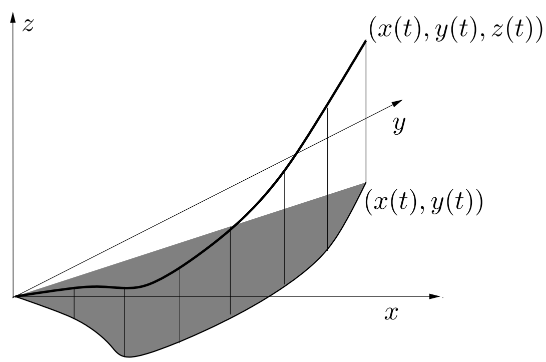


Figure 1: The Dido problem

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 $(x(t), y(t))$.

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 $(x, y, z) \mapsto (rx, ry, r^2 z)$ 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

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

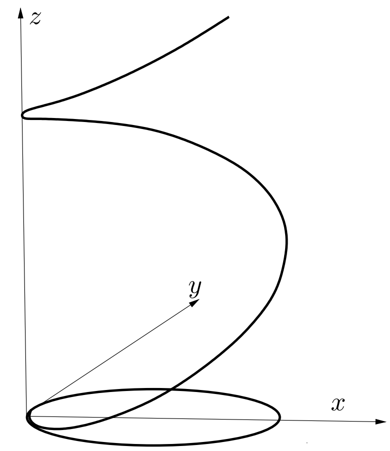


Figure 2: Solutions to the Dido problem

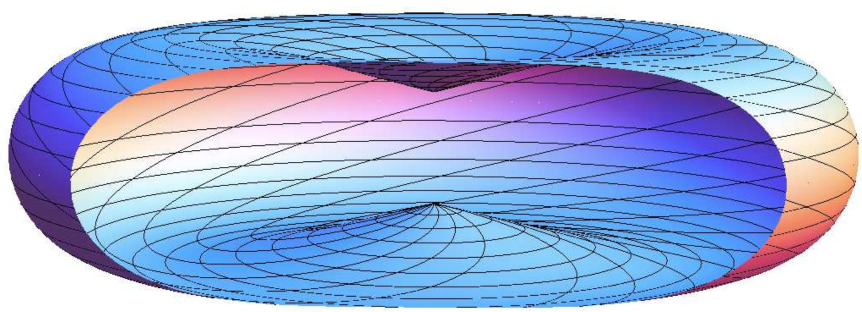


Figure 3: The Heisenberg sub-Riemannian sphere

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 Section **??** and Chapter 13.

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 eventual 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 few 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 $f_1, \dots, f_k$ 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 γ 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".

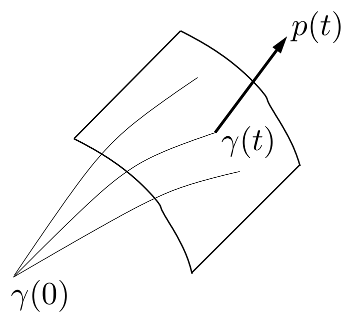


Figure 4: The "wave front" and the "impulse"

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 $p \in T^*_q M$, and $v \in T_q M$ is an admissible velocity of length 1 that maximizes the inner product of p with admissible velocities of length 1 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 geodesics with admissible initial velocities multiply.

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 prove three fundamental facts: the finiteness and the continuity of the sub-Riemannian distance; the existence of length-minimizers; the infinitesimal characterization of geodesics. The first is the classical Chow-Rashevski theorem, the second and the third one are simplified versions of the Filippov existence theorem and the Pontryagin maximum principle.

In Chapter 4, we introduce the symplectic language. We define the geodesic Hamiltonian flow, we consider an interesting class of three-dimensional problems and we prove a general sufficient condition for length-minimality of normal trajectories. Chapter 5 is devoted to applications 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 places the reader can substitute terms "Lipschitz" and "absolutely continuous" by "piecewise C^1" and "measurable" by "piecewise continuous" without a loss for the understanding.

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 differs from a Riemannian one only along a one-dimensional submanifold. 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 semi-concave out of the points connected by abnormal length-minimizers. Moreover, generic sphere is a Lipschitz submanifold if we remove these bad points.

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. Chapter 13 is devoted to the explicit calculation of the sub-Riemannian distance for model spaces.

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 Φ^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 λ 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 the next Chapter 18 we study the small distance asymptotics of the expowhree-dimensional contact case and see how the structure of the conjugate locus is encoded in the curvature.

Chapter 19 we address the problem of defining a canonical volume in sub-Riemannian geometry. We introduce the Popp volume, that is a canonical volume that is smooth for equiregular sub-Riemannian manifold, and study its basic properties.

In the last Chapter 20 we define the sub-Riemannian Laplace operator, the canonical volume form, and compute the density of the sub-Riemannian Hausdorff measure. 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.

We finish here this introduction into the Introduction... We hope that the reader won't be bored; comments to the chapters contain suggestions for further reading.[1]

[1] This research has been supported by the European Research Council, ERC StG 2009 "GeCoMethods", contract number 239748 and by the ANR project SRGI "Sub-Riemannian Geometry and Interactions", contract number ANR-15-CE40-0018.

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 "heating 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 Minskowski 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 γ 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 $\mathbb{R}^3$.

Remark 1.2. 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 γ.

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

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

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

Now we focus on *length-minimizers*, i.e., piece-wise smooth curves that realize the distance between their endpoints: $\ell(\gamma) = \mathsf{d}(\gamma(0), \gamma(T))$.

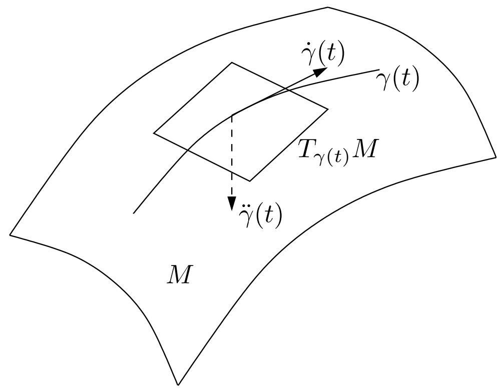


Figure 1.1: A smooth minimizer

Exercise 1.3. 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 minimizers are smooth (cf. Corollary 1.15).

Proposition 1.4. *Let $\gamma : [0,T] \to M$ be a smooth minimizer parametrized by length. Then $\ddot\gamma(t) \perp T_{\gamma(t)}M$ for all $t \in [0,T]$.*

Proof. Consider a smooth non-autonomous vector field $(t,q) \mapsto f_t(q) \in T_qM$ that extends the tangent vector to γ in a neighborhood W of the graph of the curve $\{(t,\gamma(t)) \in \mathbb{R} \times M\}$, i.e.

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

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

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

Piecewise smooth curves parametrized by 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 γ is the solution to (1.3) associated with the control $u(t) \equiv 0$ and initial condition $\gamma(0)$.

Let us consider 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 $x_{\tau,s}(0) = \gamma(0)$.

Lemma 1.5. *For every $\tau_1, \tau_2, t \in [0,T]$ the following vectors are linearly dependent*

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

Proof. By Exercice 1.3 is not restrictive to assume $t = T$. Fix $0 \leq \tau_1 \leq \tau_2 \leq T$ and consider the family of curves $\phi(t; h_1, h_2)$ solutions of (1.3) associated with controls

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

where h_1, h_2 belong to a neighborhood of 0 and ε is small enough (to guarantee the existence of the trajectory). Notice that ϕ is smooth in a neighborhood of $(t, h_1, h_2) = (T, 0, 0)$ and

$$\frac{\partial \phi}{\partial h_i}\bigg|_{(h_1,h_2)=0} = \frac{\partial}{\partial s}\bigg|_{s=0} x_{\tau_i,s}(T), \qquad i = 1, 2.$$

By contradiction assume that the vectors in (1.5) are linearly independent. Then $\frac{\partial \phi}{\partial h}$ is invertible and the classical 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 γ is a minimizer. □

Lemma 1.6. *For every $\tau, t \in [0,T]$ the following identity holds*

$$\left\langle \frac{\partial}{\partial s}\bigg|_{s=0} x_{\tau,s}(t) \,\bigg|\, \dot{\gamma}(t) \right\rangle = 0. \tag{1.6}$$

Proof. If $t \leq \tau$, then by construction (cf. (1.4)) the first vector is zero since there is no variation w.r.t. s and the conclusion follows. Let us now assume that $t > \tau$. Again, by Remark 1.3, it is sufficient to prove the statement at $t = T$. 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 $t = \tau$. Observe that, for $t \geq \tau$

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

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 $t \geq \tau$, we have

$$\psi(t) = (t - \tau) g_\tau(x_{\tau,s}(\tau)) + O((t - \tau)^2). \tag{1.7}$$

For τ sufficiently close to T, one can take $t = T$ in (1.7). 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 1.5 all vectors in 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. □

Now we end the proposition by showing that $\ddot{\gamma}(t) \perp T_{\gamma(t)}M$. Notice that this is equivalent to show

$$\langle \ddot{\gamma}(t) \,|\, f_t(\gamma(t)) \rangle = \langle \ddot{\gamma}(t) \,|\, g_t(\gamma(t)) \rangle = 0. \tag{1.8}$$

Recall that $\langle \dot{\gamma}(t) \,|\, \dot{\gamma}(t) \rangle = 1$. Differentiating this identity one gets

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

which shows that $\ddot{\gamma}(t)$ is orthogonal to $f_t(\gamma(t))$. Next, differentiating (1.6) with respect to t, we have[1] for $t \neq \tau$

$$\left\langle \frac{\partial}{\partial s}\bigg|_{s=0} \dot{x}_{\tau,s}(t) \,\bigg|\, \dot{\gamma}(t) \right\rangle + \left\langle \frac{\partial}{\partial s}\bigg|_{s=0} x_{\tau,s}(t) \,\bigg|\, \ddot{\gamma}(t) \right\rangle = 0. \tag{1.9}$$

Now, from $\langle \dot{x}_{\tau,s}(t) \,|\, \dot{x}_{\tau,s}(t) \rangle = 1$ one gets

$$\left\langle \frac{\partial}{\partial s} \dot{x}_{\tau,s}(t) \,\bigg|\, \dot{x}_{\tau,s}(t) \right\rangle = 0, \qquad \text{for } t \neq \tau.$$

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

$$\left\langle \frac{\partial}{\partial s}\bigg|_{s=0} \dot{x}_{\tau,s}(t) \,\bigg|\, \dot{\gamma}(t) \right\rangle = 0, \qquad \text{for } t \neq \tau.$$

Hence, by (1.9), it follows that

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

which, by continuity, holds for every $t \in [0, T]$. Using that $\frac{\partial}{\partial s}\big|_{s=0} x_{\tau,s}(t)$ is parallel to $g_t(\gamma(t))$ (see proof of Lemma 1.6), it follows that $\langle g_t(\gamma(t)) \,|\, \ddot{\gamma}(t) \rangle = 0$. $\square$

Definition 1.7. A smooth curve $\gamma : [0, T] \to M$ parametrized with constant speed is called *geodesic* if it satisfies

$$\ddot{\gamma}(t) \perp T_{\gamma(t)}M, \qquad \forall\, t \in [0, T]. \tag{1.10}$$

Proposition 1.4 says that a smooth curve that minimizes the length is a geodesic.

Now 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

$$M = a^{-1}(0), \qquad \text{and} \qquad \nabla a \neq 0 \text{ on } M. \tag{1.11}$$

Remark 1.8. Recall that for all $q \in M$ 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 1.9. *A smooth curve $\gamma : [0, T] \to M$ is a geodesic if and only if it satisfies, in matrix notation:*

$$\ddot{\gamma}(t) = -\frac{\dot{\gamma}(t)^T (\nabla^2_{\gamma(t)} a) \dot{\gamma}(t)}{\|\nabla_{\gamma(t)} a\|^2} \nabla_{\gamma(t)} a, \qquad \forall\, t \in [0, T], \tag{1.12}$$

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 $\langle \nabla_{\gamma(t)} a \,|\, \dot\gamma(t) \rangle = 0$ we get, in matrix notation:

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

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

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

Hence we get

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

from which (1.12) follows. □

Remark 1.10. Notice that formula (1.12) is always true locally since, by definition of surface, the assumptions (1.11) are always satisfied locally.

1.1.1 Existence and minimizing properties of geodesics

As a direct consequence of Proposition 1.9 one gets the following existence and uniqueness theorem for geodesics.

Corollary 1.11. *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 1.9, 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 1.12. *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.* $\mathsf{d}(\gamma(\tau), \gamma(\tau+\varepsilon)) = \ell(\gamma|_{[\tau,\tau+\varepsilon]})$,

(ii) $\gamma|_{[\tau,\tau+\varepsilon]}$ *is the unique minimizers 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 γ is length parametrized. Consider a length-parametrized curve α on M such that $\alpha(0) = \gamma(0)$ and $\dot\alpha(0) \perp \dot\gamma(0)$ and denote by $(t, s) \mapsto x_s(t)$ the smooth variation of geodesics such that $x_0(t) = \gamma(t)$ and (see also Figure 1.2)

$$x_s(0) = \alpha(s), \qquad \dot x_s(0) \perp \dot\alpha(s). \tag{1.13}$$

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

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

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

$$\phi : x_s(t) \mapsto t,$$
$$X : x_s(t) \mapsto \dot x_s(t).$$

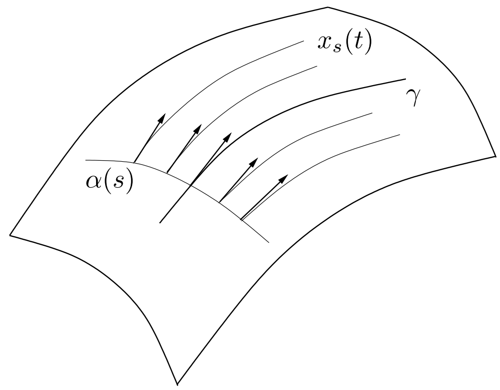


Figure 1.2: Proof of Theorem 1.12

Lemma 1.13. $\nabla_q \phi = X(q)$ *for every* $q \in W$.

Proof of Lemma 1.13. 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

$$\left\langle \nabla_{x_s(t)}\phi \,\middle|\, \frac{\partial}{\partial s} x_s(t) \right\rangle = 0, \qquad \forall\,(t,s) \in U. \tag{1.14}$$

Now, let us show that

$$\left\langle \frac{\partial}{\partial s} x_s(t) \,\middle|\, \dot{x}_s(t) \right\rangle = 0, \qquad \forall\,(t,s) \in U. \tag{1.15}$$

Computing the derivative with respect to t of the left hand side of (1.15) one gets

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

which is identically zero. Indeed the first term is zero because $\dot{x}_s(t)$ has unit speed and the second one vanishes because of (1.10). Hence, the left hand side of (1.15) is constant and coincides with its value at $t = 0$, which is zero by the orthogonality assumption (1.13).

By (1.14) and (1.15) one gets that $\nabla\phi$ is parallel to X. Actually they coincide since

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

□

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 γ 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

$$\ell(\gamma|_{[0,\varepsilon]}) = \phi(c(\varepsilon')) - \phi(c(0)) = \int_0^{\varepsilon'} \frac{d}{dt}\phi(c(t))dt \tag{1.16}$$

$$= \int_0^{\varepsilon'} \langle \nabla\phi(c(t)) \mid \dot{c}(t)\rangle \, dt$$

$$= \int_0^{\varepsilon'} \langle X(c(t)) \mid \dot{c}(t)\rangle \, dt \leq \varepsilon' = \ell(c), \tag{1.17}$$

The last inequality follows from the Cauchy-Schwartz inequality

$$\langle X(c(t)) \mid \dot{c}(t)\rangle \leq \|X(c(t))\|\|\dot{c}(t)\| = 1 \tag{1.18}$$

which holds at every smooth point of $c(t)$. In addition, equality in (1.18) 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 $\gamma|_{[0,\varepsilon]}$.

Now let us 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 $\bar{\varepsilon} < \mathrm{dist}(\gamma(0), \partial W)$. Every curve that escape from W has length greater than $\bar{\varepsilon}$. □

From Theorem 1.12 it follows

Corollary 1.14. *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 in the class of absolutely continuous ones, if one understands the integral in the Lebesgue sense.

In this setting, in the proof of Theorem 1.12, one can assume that the curve c is actually absolutely continuous. This proves that small pieces of geodesics are minimizers also in the class of absolutely continuous curves on M. Morever, this proves the following.

Corollary 1.15. *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, which let us to define the main geometric invariant of a surface: the Gaussian curvature.

Let us consider a curve $\gamma : [0, T] \to M$ and a vector $\xi \in T_{\gamma(0)}M$. We want to define the parallel transport of ξ along γ. Heuristically, it is a curve $\xi(t) \in T_{\gamma(t)}M$ such that the vectors $\{\xi(t), t \in [0, T]\}$ are all "parallel".

Remark 1.16. If $M = \mathbb{R}^2 \subset \mathbb{R}^3$ is the set $\{z = 0\}$ we can canonically identify every tangent space $T_{\gamma(t)}M$ with $\mathbb{R}^2$ so that every tangent vector $\xi(t)$ belong to the same vector space.[2] In this case, parallel simply means $\dot{\xi}(t) = 0$ as an element of $\mathbb{R}^3$. This is not the case if M is a manifold because tangent spaces at different points are different.

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

Definition 1.17. 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 $\dot{\xi}(t) \perp T_{\gamma(t)}M$.

Assume now that M is the zero level of a smooth function $a : \mathbb{R}^3 \to \mathbb{R}$ as in (1.11). We have the following description:

Proposition 1.18. *A smooth curve of tangent vectors $\xi(t)$ defined along $\gamma : [0, T] \to M$ is parallel if and only if it satisfies*

$$\dot{\xi}(t) = -\frac{\dot{\gamma}(t)^T(\nabla^2_{\gamma(t)}a)\xi(t)}{\|\nabla_{\gamma(t)}a\|^2}\nabla_{\gamma(t)}a, \qquad \forall\, t \in [0, T]. \tag{1.19}$$

Proof. As in Remark 1.8, $\xi(t) \in T_{\gamma(t)}M$ 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 α. With analogous computations as in the proof of Proposition 1.9 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 1.19. Notice that, since (1.53) is a first order linear ODE with respect to ξ, 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 γ such that $\xi(0) = v$. Since (1.53) 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 1.20. *The parallel transport preserves the inner product. In other words, if $\xi(t), \eta(t)$ are two parallel curves of tangent vectors along γ, then we have*

$$\frac{d}{dt}\langle \xi(t) \,|\, \eta(t) \rangle = 0, \qquad \forall\, t \in [0, T]. \tag{1.20}$$

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

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

The notion of parallel transport permits to give a new characterization of geodesics. Indeed, by definition

Corollary 1.21. *A smooth curve $\gamma : [0, T] \to M$ is a geodesic if and only if $\dot{\gamma}$ is parallel along γ.*

In the following we assume that M is oriented.

Definition 1.22. The *spherical bundle* SM on M is the disjoint union of all unit tangent vectors to M:

$$SM = \bigsqcup_{q \in M} S_qM, \qquad S_qM = \{v \in T_qM, \|v\| = 1\}. \tag{1.21}$$

SM is a smooth manifold of dimension 3. Moreover it has the structure 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 1.23. Since every vector in the fiber S_qM has norm one, we can parametrize every $v \in S_qM$ by an angular coordinate $\theta \in S^1$ through an orthonormal frame $\{e_1(q), e_2(q)\}$ for S_qM, i.e. $v = \cos(\theta)e_1(q) + \sin(\theta)e_2(q)$.

The choice of a positively oriented orthonormal frame $\{e_1(q), e_2(q)\}$ corresponds to fix the element in the fiber corresponding to $\theta = 0$. Hence, the choice of such an orthonormal frame at every point q induces coordinates on SM of the form $(q, \theta + \varphi(q))$, where $\varphi \in C^\infty(M)$.

Given an element $\xi \in S_qM$ we can complete it to an orthonormal frame (ξ, η, ν) of $\mathbb{R}^3$ in the following unique way:

(i) $\eta \in T_qM$ is orthogonal to ξ and (ξ, η) is positively oriented (w.r.t. the orientation of M),

(ii) $\nu \perp T_qM$ and (ξ, η, ν) is positively oriented (w.r.t. the orientation of $\mathbb{R}^3$).

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 $\eta(t), \nu(t) \in T_{\gamma(t)}M$ as above. Since $t \mapsto \xi(t)$ has constant speed, one has $\xi(t) \perp \dot{\xi}(t)$ and we can write

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

In particular this shows that every element of $T_\xi SM$, written in the basis (ξ, η, ν), has zero component along ξ.

Definition 1.24. The *Levi-Civita connection* on M is the 1-form $\omega \in \Lambda^1(SM)$ defined by

$$\omega_\xi : T_\xi SM \to \mathbb{R}, \qquad \omega_\xi(z) = u_z, \tag{1.22}$$

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

Notice that ω change sign if we change the orientation of M.

Lemma 1.25. *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 $\nu(t)$. ☐

In particular, a curve parametrized by length $\gamma : [0, T] \to M$ is a geodesic if and only if

$$\omega_{\dot{\gamma}(t)}(\ddot{\gamma}(t)) = 0, \qquad \forall\, t \in [0, T]. \tag{1.23}$$

Proposition 1.26. *The 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

$$\omega = d\theta + a_1(x_1, x_2)dx_1 + a_2(x_1, x_2)dx_2, \tag{1.24}$$

where (x_1, x_2, θ) is a system of coordinates on SM.

(ii) $d\omega = \pi^*\Omega$, where Ω 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, θ) on SM and consider the vector field $\partial/\partial\theta$ on SM. Let us show that

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

Indeed consider a curve $t \mapsto \xi(t)$ of unit tangent vector 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 ω. By construction, ω is invariant by rotations, hence the coefficients $a_i = \omega(\partial/\partial x_i)$ do not depend on the variable θ.

(ii) Follows directly from expression (1.24) noticing that $d\omega$ depends only on x_1, x_2. □

Remark 1.27. Notice that the functions a_1, a_2 in (1.24) are not invariant by change of coordinates on the fiber. Indeed the transformation $\theta \to \theta + \varphi(x_1, x_2)$ induces $d\theta \to d\theta + (\partial_{x_1}\varphi)dx_1 + (\partial_{x_2}\varphi)dx_2$ which gives $a_i \to a_i + \partial_{x_i}\varphi$ for $i = 1, 2$.

By definition ω is an intrinsic 1-form on SM. Its differential, by property (ii) of Proposition 1.55, is the pull-back of an intrinsic 2-form on M, that in general is not exact.

Definition 1.28. The area form dV on a surface M is the differential two 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 1.29. The *Gaussian curvature* of M is the function $\kappa : M \to \mathbb{R}$ defined by the equality

$$\Omega = -\kappa dV. \tag{1.25}$$

Note that κ does not depend on the orientation of M, since both Ω 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 1.30. Let $\gamma : [0, T] \to M$ be a smooth curve parametrized by length. The *geodesic curvature* of γ is defined as

$$\rho_\gamma(t) = \omega_{\dot{\gamma}(t)}(\ddot{\gamma}(t)). \tag{1.26}$$

Notice that if γ 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 1.31. 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

Definition 1.32. A *curvilinear polygon* Γ on an oriented surface M is the image of a closed polygon in $\mathbb{R}^2$ under a diffeomorphism. We assume that $\partial\Gamma$ is oriented consistently with the orientation of M. In the following we represent $\partial\Gamma = \cup_{i=1}^m \gamma_i(I_i)$ where $\gamma_i : I_i \to M$, for $i = 1, \ldots, m$, are smooth curves parametrized by length, with orientation consistent with $\partial\Gamma$. We denote by α_i the external angles at the points where $\partial\Gamma$ is not C^1 (see Figure 1.3).

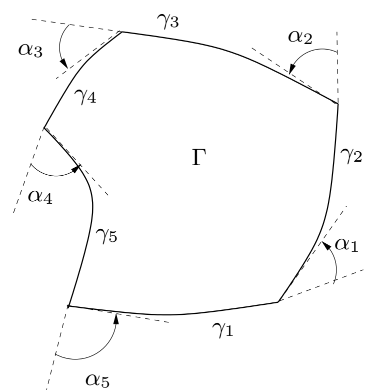


Figure 1.3: A curvilinear polygon

Notice that a curvilinear polygon is homeomorphic to a disk.

Theorem 1.33 (Gauss-Bonnet, local version)**.** *Let Γ be a curvilinear polygon on an oriented surface M. Then we have*

$$\int_\Gamma \kappa dV + \sum_{i=1}^m \int_{I_i} \rho_{\gamma_i}(t)dt + \sum_{i=1}^m \alpha_i = 2\pi. \tag{1.27}$$

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

In this case Γ 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 ε small enough, we define (cf. Figure 1.4)

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

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

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

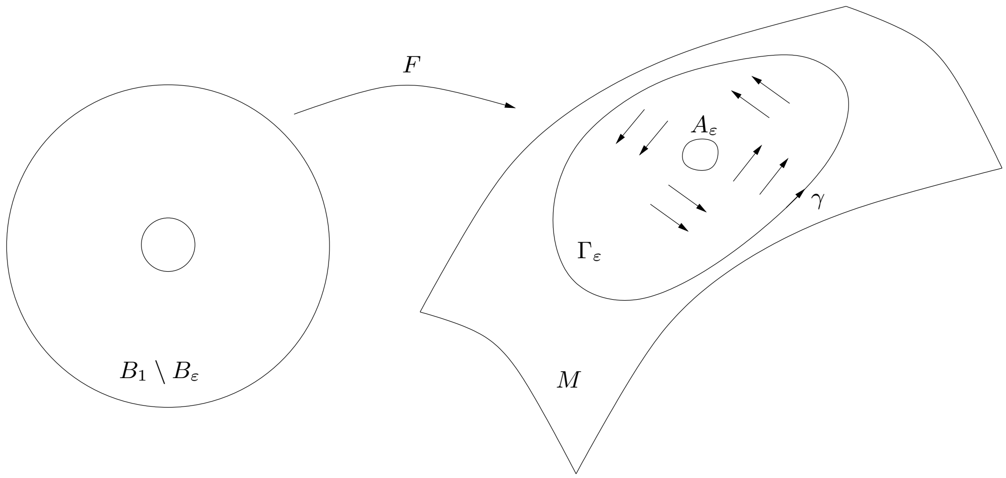


Figure 1.4: The map F

First notice that

$$\int_{\phi(\Gamma_\varepsilon)} d\omega = \int_{\phi(\Gamma_\varepsilon)} \pi^*\Omega = \int_{\pi(\phi(\Gamma_\varepsilon))} \Omega = \int_{\Gamma_\varepsilon} \Omega, \tag{1.28}$$

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

$$\begin{aligned}
\int_{\Gamma_\varepsilon} \kappa dV &= -\int_{\Gamma_\varepsilon} \Omega = -\int_{\phi(\Gamma_\varepsilon)} d\omega, && \text{(by (1.28))} \\
&= -\int_{\partial\phi(\Gamma_\varepsilon)} \omega, && \text{(by Stokes Theorem)} \\
&= \int_{\phi(A_\varepsilon)} \omega - \int_{\phi(\partial\Gamma)} \omega, && \text{(since } \partial\phi(\Gamma_\varepsilon) = \phi(A_\varepsilon) \cup \phi(\partial\Gamma)\text{)}
\end{aligned} \tag{1.29}$$

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 1.55, can be written as

$$\int_{\phi(A_\varepsilon)} \omega = \int_{\phi(A_\varepsilon)} d\theta + \int_{\phi(A_\varepsilon)} a_1(x_1, x_2)dx_1 + a_2(x_1, x_2)dx_2 \tag{1.30}$$

The first element of (1.30) is equal to 2π since we integrate the 1-form $d\theta$ on a closed curve. The second element of (1.30), for $\varepsilon \to 0$, satisfies

$$\left| \int_{\phi(A_\varepsilon)} a_1(x_1, x_2)dx_1 + a_2(x_1, x_2)dx_2 \right| \leq C\ell(\phi(A_\varepsilon)) \to 0, \tag{1.31}$$

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 (1.29). 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 (1.29)

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

that is (1.27) 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 with a sequence of polygons Γ_n such that $\partial\Gamma_n$ is smooth and Γ_n approximates Γ in a "smooth" way. In particular, we assume that $\partial\Gamma_n$ coincides with $\partial\Gamma$ excepts in neighborhoods U_i, for $i = 1, \dots, m$, of each point q_i where $\partial\Gamma$ is not smooth, in such a way that the curve $\sigma_i^{(n)}$ that parametrize $(\partial\Gamma_n \setminus \partial\Gamma) \cap U_i$ satisfies $\ell(\sigma_i^n) \leq 1/n$.

If we apply the statement of the Theorem for the smooth case to Γ_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 Γ_n tends to Γ as $n\to\infty$, then

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

We are left to prove that

$$\lim_{n\to\infty}\int \rho_{\gamma^{(n)}}(t)dt = \sum_{i=1}^m \int_{I_i} \rho_{\gamma_i}(t)dt + \sum_{i=1}^m \alpha_i. \tag{1.32}$$

For every n, let us split the curve $\gamma^{(n)}$ as the union of the smooth curves $\sigma_i^{(n)}$ and $\gamma_i^{(n)}$ as in Figure **??**. 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 γ_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 (1.32) 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 γ_i is a geodesic. For a geodesic triangle T we denote by $A_i := \pi - \alpha_i$ its internal angles.

Corollary 1.34. *Let T be a geodesic triangle and $A_i(T)$ its internal angles. Then*

$$\kappa(q) = \lim_{|T|\to 0} \frac{\sum_i A_i(T) - \pi}{|T|}$$

Proof. Fix a geodesic triangle T. Using that the geodesic curvature of γ_i vanishes, the local version of Gauss-Bonnet Theorem (1.27) can be rewritten as

$$\sum_{i=1}^{3} A_i = \pi + \int_\Gamma \kappa dV. \tag{1.33}$$

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

$$\kappa(q) = \lim_{|T|\to 0} \frac{1}{|T|} \int_T \kappa dV = \lim_{|T|\to 0} \frac{\sum_i A_i(T) - \pi}{|T|}$$

□

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 (1.27) to the case when Γ is a region of M not necessarily homeomorphic to the disk, see for instance Figure 1.5. 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 1.32). Notice that every compact surface admits a triangulation.[3]

Definition 1.35. Let $M \subset \mathbb{R}^3$ be a compact oriented surface with boundary ∂M (possibly with angles). Consider a triangulation of M. We define the *Euler characteristic* of M as

$$\chi(M) := n_2 - n_1 + n_0, \tag{1.34}$$

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

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

Remark 1.36. The Euler characteristic is well-defined. Indeed one can show that the quantity (1.34) is invariant for refinement of a triangulation, since every 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 (1.34) is independent on the triangulation.

Example 1.37. 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 $\mathbb{R}^2$.

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

Following the notation introduced in the previous section, for a given region Γ, 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 $i = 1, \dots, m$, are smooth curves parametrized by length (with orientation consistent with $\partial\Gamma$). We denote by α_i the external angles at the points where $\partial\Gamma$ is not C^1 (see Figure 1.5).

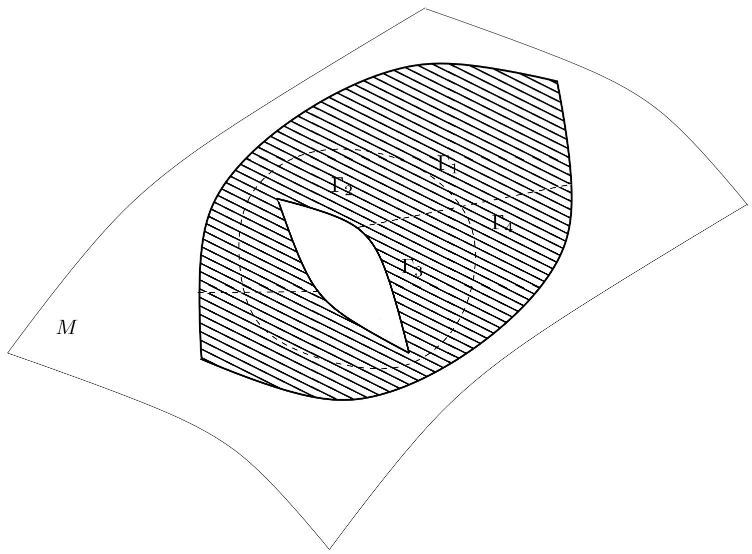


Figure 1.5: Gauss-Bonnet Theorem

Theorem 1.38 (Gauss-Bonnet, global version)**.** *Let Γ be a region of a surface on a compact oriented surface M. Then*

$$\int_\Gamma \kappa dV + \sum_{i=1}^m \int_{I_i} \rho_{\gamma_i}(t)dt + \sum_{i=1}^m \alpha_i = 2\pi\chi(\Gamma). \tag{1.35}$$

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 Γ and let $\{\Gamma_j, j = 1, \dots, n_2\}$ be the corresponding subdivision of Γ in curvilinear polygons. We denote by $\{\gamma_k^{(j)}\}$ the smooth curves parametrized by length whose image are the edges of Γ_j and by and $\theta_k^{(j)}$ the external angles of Γ_j. We assume that all orientations are chosen accordingly to the orientation of M. Applying Theorem 1.33 to every Γ_j and summing w.r.t. j we get

$$\sum_{j=1}^{n_2} \left(\int_{\Gamma_j} \kappa dV + \sum_k \int \rho_{\gamma_k^{(j)}}(t)dt + \sum_k \theta_k^{(j)} \right) = 2\pi n_2. \tag{1.36}$$

We have that

$$\sum_{j=1}^{n_2} \int_{\Gamma_j} \kappa dV = \int_\Gamma \kappa dV, \qquad \sum_{j,k} \int \rho_{\gamma_k^{(j)}}(t)dt = \sum_{i=1}^m \int \rho_{\gamma_i}(t)dt. \tag{1.37}$$

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

$$\sum_{j,k} \theta_k^{(j)} = 2\pi(n_1 - n_0), \tag{1.38}$$

Let us denote by N the total number of angles in the sum of the left hand side of (1.38). After reindexing we have to check that

$$\sum_{\nu=1}^{N} \theta_\nu = 2\pi(n_1 - n_0). \tag{1.39}$$

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

$$\sum_{\nu=1}^{N} \theta_\nu = N\pi - \sum_{\nu=1}^{N} A_\nu. \tag{1.40}$$

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

$$\sum_{\nu=1}^{N} A_\nu = 2\pi n_0^I + \pi n_0^\partial, \tag{1.41}$$

Combining (1.40), (1.41) 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 (1.39).

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

We consider a decomposition of Γ 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 (1.37). Now, instead of (1.39), we have to check that

$$\sum_{\nu=1}^{N} \theta_\nu = \sum_{i=1}^{m} \alpha_i + 2\pi(n_1 - n_0), \tag{1.42}$$

Now (1.42) 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 Γ, together with the decomposition induced by Γ (see Figure 1.5). 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:

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

□

Corollary 1.39. *Let M be a compact oriented surface without boundary. Then*

$$\int_M \kappa dV = 2\pi\chi(M). \tag{1.43}$$

1.3.3 Consequences of the Gauss-Bonnet Theorems

Definition 1.40. Let M, M' be two surfaces in $\mathbb{R}^3$. A smooth map $\phi : \mathbb{R}^3 \to \mathbb{R}^3$ is called an *isometry* between M and M' if $\phi(M) = M'$ and for every $q \in M$ it satisfies

$$\langle v \,|\, w\rangle = \langle D_q\phi(v) \,|\, D_q\phi(w)\rangle\,, \qquad \forall\, v, w \in T_qM. \tag{1.44}$$

If the property (1.44) is satisfied by a map defined locally in a neighborhood of every point q of M, then it is called a *local isometry*.

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

From (1.44) it follows that an isometry preserves the angles between vectors and, a fortiori, the length of a curve and the distance between two points.

Corollary 1.34, and 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.

Corollary 1.41 (Gauss's Theorema Egregium)**.** *Assume ϕ is a local isometry between M and M', then for every $q \in M$ one has $\kappa(q) = \kappa'(\phi(q))$, where κ (resp. κ') is the Gaussian curvature of M (resp. M').*

This Theorem says that the Gaussian curvature κ 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.

Corollary 1.42. *Let M be 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.*

Exercise 1.43. Prove that a surface M is locally isometric to the Euclidean plane $\mathbb{R}^2$ around a point $q \in M$ if and only if there exists a coordinate system (x_1, x_2) in a neighborhood U of $q \in M$ such that the vectors ∂_{x_1} and ∂_{x_2} have unit length and are everywhere orthonormal.

As a converse of Corollary 1.42 we have the following.

Theorem 1.44. *Assume that $\kappa \equiv 0$ in a neighborhood of a point $q \in M$. Then M is locally Euclidean (i.e., locally isometric to $\mathbb{R}^2$) around q.*

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

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

Hence $d\omega = \pi^*\Omega = 0$. From its 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, which we still denote by θ, in such a way that (see also Remark 1.27)

$$\omega = d\theta. \tag{1.45}$$

Now, let γ be a length parametrized geodesic, i.e. $\omega_{\dot\gamma(t)}(\ddot\gamma(t)) = 0$. Using the the angular coordinate θ 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 (1.45), 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 of a geodesic γ is constant.

We want to construct Cartesian coordinates in a neighborhood U of q. Consider the two length parametrized geodesics γ_1 and γ_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 γ_1 and γ_2 (See Figure 1.6). 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 ∂_{x_1} and ∂_{x_2} have length one (being the tangent vectors to length parametrized geodesics) and are everywhere mutually orthogonal. Hence the theorem follows from Exercise 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.

Definition 1.45. Let M be an oriented surface. We define the *Gauss map* associated to M as

$$\mathcal{N} : M \to S^2, \qquad q \mapsto \nu_q, \tag{1.46}$$

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

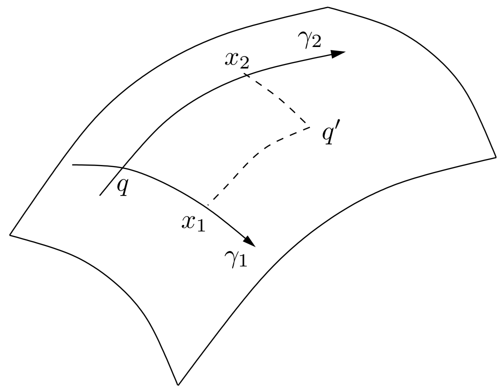


Figure 1.6: Proof of Theorem 1.44.

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

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

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

Theorem 1.46. *We have that* $\kappa(q) = \det(D_q\mathcal{N})$.

Before proving this theorem we prove an important property of the Gauss map.

Lemma 1.47. *For every $q \in M$, the differential $D_q\mathcal{N}$ of the Gauss map is a symmetric operator, i.e.,*

$$\langle D_q\mathcal{N}(\xi) \,|\, \eta\rangle = \langle \xi \,|\, D_q\mathcal{N}(\eta)\rangle, \qquad \forall\, \xi, \eta \in T_qM. \tag{1.47}$$

Proof. We prove the statement locally, i.e., for a manifold 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 $\langle \mathcal{N}(q) \,|\, \eta\rangle = \langle \mathcal{N}(q) \,|\, D_u\phi(w)\rangle = 0$. Taking the derivative in the direction of ξ one gets

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

where $D^2_u\phi$ is a bilinear symmetric map. Now (1.47) follows exchanging the role of v and w. $\square$

Proof of Theorem 1.46. We will use Cartan's moving frame method. Let $\xi \in SM$ and denote with

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

the orthonormal basis attached at ξ and constructed in Section 1.2. Let us compute the differentials of these vectors in the ambient space $\mathbb{R}^3$ and write them as a linear combination (with 1-form as coefficients) of the vectors e_i

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

Dropping ξ and η from the notation one gets the relation

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

Since for each ξ the basis $(e_1(\xi), e_2(\xi), e_3(\xi))$ is orthonormal (hence can be seen as an element of $SO(3)$) its derivative is expressed through a skew-symmentric matrix (i.e., $\omega_{ij} = -\omega_{ji}$) and one gets the equations

$$\begin{aligned} de_1 &= \omega_{12}e_2 + \omega_{13}e_3, \\ de_2 &= -\omega_{12}e_1 + \omega_{23}e_3, \\ de_3 &= -\omega_{13}e_1 - \omega_{23}e_2. \end{aligned} \tag{1.48}$$

Let us now prove the following identity

$$\omega_{13} \wedge \omega_{23} = d\omega_{12}. \tag{1.49}$$

Indeed, differentiating the first equation in (1.48) 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 (1.49).

The statement of the theorem can be rewritten as an identity between 2-forms as follows

$$\det(D_q\mathcal{N})dV = \kappa dV.$$

Applying π^* to both sides one gets

$$\pi^*(\det(D_q\mathcal{N})dV) = \pi^*\kappa dV = d\omega \tag{1.50}$$

where ω is the Levi-Civita connection. Let us show that (1.50) is equivalent to (1.49).

Indeed by construction ω_{12} computes the coefficient of the derivative of the first vector of the orthonormal basis along the second one, hence $\omega_{12} = \omega$ (see also Definition 1.54). It remains to show that

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

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 proof is completed by the following

Exercise 1.48. Let V be a 2-dimensional Euclidean vector space and e_1, e_2 an orthonormal basis. Let $F : V \to V$ a linear map and write $F = F_1e_1 + F_2e_2$, 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.

$\square$

Remark 1.49. Lemma 1.47 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 can be geometrically interpreted as the maximum and the minimum of curvature of sections of M with orthogonal planes.

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

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

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 Minkowski-type) inner product

$$\langle q_1, q_2\rangle_h = x_1x_2 + y_1y_2 - z_1z_2. \tag{1.51}$$

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 norm induced by the inner product (1.51).

For the metric structure to be defined on M, we require that the restriction of the inner product (1.51) to the tangent space to M is positive definite at every point. Indeed, under this assumption, the inner product (1.51) 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 Minkovski space.

The structure of the inner product impose some condition 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 (1.51).

(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 restriction to M of the projection $\pi(x, y, z) = (x, y)$ onto the xy-plane is a local diffeomorphism.

(iv) Show that M is locally a graph on the plane $\{z = 0\}$.

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* $\nabla_q^h a$ is defined as

$$\nabla_q^h a = \left(\frac{\partial a}{\partial x}, \frac{\partial a}{\partial y}, -\frac{\partial a}{\partial z}\right)$$

If we assume that $M = a^{-1}(0)$ is a regular level set of a smooth function $a : \mathbb{R}^3 \to \mathbb{R}$. If $\gamma(t)$ 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 γ be a geodesic on $M = a^{-1}(0)$. Show that γ satisfies the equation (in matrix notation)

$$\ddot\gamma(t) = -\frac{\dot\gamma(t)^T (\nabla^2_{\gamma(t)} a) \dot\gamma(t)}{\|\nabla^h_{\gamma(t)} a\|_h^2} \nabla^h_{\gamma(t)} a, \qquad \forall\, t \in [0, T]. \tag{1.52}$$

where $\nabla^2_{\gamma(t)} a$ is the (classical) matrix of second derivatives of a.[4]

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 γ if and only if it satisfies

$$\dot\xi(t) = -\frac{\dot\gamma(t)^T (\nabla^2_{\gamma(t)} a) \xi(t)}{\|\nabla^h_{\gamma(t)} a\|_h^2} \nabla^h_{\gamma(t)} a, \qquad \forall\, t \in [0, T]. \tag{1.53}$$

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

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 γ such that $\xi(0) = v$. Moreover the operator $\xi(0) \mapsto \xi(t)$ is a linear operator, which the *parallel transport* of v along γ.

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

$$\frac{d}{dt} \langle \xi(t) \,|\, \eta(t) \rangle_h = 0, \qquad \forall\, t \in [0, T]. \tag{1.54}$$

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

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

(ii) $\nu \perp T_q M$ with respect to $\langle \cdot \,|\, \cdot \rangle_h$ and (ξ, η, ν) 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

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

[4] otherwise one can write the numerator of (1.52) 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.*

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

$$\omega_\xi : T_\xi SM \to \mathbb{R}, \qquad \omega_\xi(z) = u_z, \tag{1.55}$$

where $z = u_z\eta + v_z\nu$ and (ξ, η, ν) 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 length $\gamma : [0, T] \to M$ is a geodesic if and only if

$$\omega_{\dot{\gamma}(t)}(\ddot{\gamma}(t)) = 0, \qquad \forall\, t \in [0, T]. \tag{1.56}$$

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

$$\omega = d\theta + a_1(x_1, x_2)dx_1 + a_2(x_1, x_2)dx_2, \tag{1.57}$$

where (x_1, x_2, θ) is a system of coordinates on SM.

(ii) $d\omega = \pi^*\Omega$, where Ω 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

$$\Omega = -\kappa dV. \tag{1.58}$$

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*

$$\mathcal{N} : M \to H^2, \qquad q \mapsto \nu_q, \tag{1.59}$$

where $\nu_q \in H^2 \subset \mathbb{R}^3$ denotes the external unit normal vector to M at q, with respect to the Minkovsky 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 Lorentzian 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 κ 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 of the classical 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 cases (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 curvature of this surface is zero (the Gauss map is constant) and to characterize geodesics and curves with constant curvature.

Exercise 1.59. Prove that geodesics on the Euclidean plane are lines. Moreover, show that curves with constant 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

$$S^2_r = a^{-1}(0), \qquad a(x, y, z) = x^2 + y^2 + z^2 - r^2. \tag{1.60}$$

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 since its is explicitly given by

$$\mathcal{N} : S^2_r \to S^2, \qquad \mathcal{N}(q) = \frac{1}{r} q, \tag{1.61}$$

It follows immediately by (1.69) 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 constant geodesic curvature curves on the sphere.

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

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

Proof. Without loss of generality, we can assume that γ is parametrized by unit speed. Differentiating twice the equality $a(\gamma(t)) = 0$, where a is the function defined in (1.68), 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 γ has constant geodesic curvature equal to c, there exists a function $b(t)$ such that

$$\ddot\gamma(t) = b(t)\nabla_{\gamma(t)}a + c\eta(t) \tag{1.62}$$

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 $T_{\gamma(t)}S_r^2$ (defined in such a way that $\dot\gamma(t)$ and $\eta(t)$ is a positively oriented frame). Reasoning as in the proof of Proposition 1.9 and noticing that $\nabla_{\gamma(t)}a$ is proportional to the vector $\gamma(t)$, one can compute $b(t)$ and obtains that γ satisfies the differential equation

$$\ddot\gamma(t) = -\frac{1}{r^2}\gamma(t) + c\eta(t). \tag{1.63}$$

Lemma 1.61. $\dot\eta(t) = -c\dot\gamma(t)$

Proof of Lemma 1.61. 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 nonvanishing 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 (1.63). Hence $\dot\eta(t) = \langle \dot\eta(t) \,|\, \dot\gamma(t)\rangle\,\dot\gamma(t) = -c\dot\gamma(t)$. ∎

Next we compute the derivatives of the function α as follows

$$\dot\alpha(t) = \langle \ddot\gamma(t) \,|\, w\rangle = -\frac{1}{r^2}\langle \gamma(t) \,|\, w\rangle + c\,\langle \eta(t) \,|\, w\rangle. \tag{1.64}$$

Using Lemma 1.61, we have

$$\ddot\alpha(t) = -\frac{1}{r^2}\langle \dot\gamma(t) \,|\, w\rangle + c\,\langle \dot\eta(t) \,|\, w\rangle \tag{1.65}$$

$$= -\frac{1}{r^2}\langle \dot\gamma(t) \,|\, w\rangle - c^2\,\langle \dot\gamma(t) \,|\, w\rangle = -\left(\frac{1}{r^2} + c^2\right)\alpha(t). \tag{1.66}$$

which ends the proof of the Proposition 1.60. ∎

Corollary 1.62. *Constant geodesic curvature curves are contained in the intersection of S^2_r with an affine plane of $\mathbb{R}^3$. In particular, geodesics are contained in the intersection of S^2_r with planes passing through the origin, i.e., great circles.*

Proof. Let us fix a vector $w \in \mathbb{R}^3$ that is orthogonal to $\dot{\gamma}(0)$ and $\ddot{\gamma}(0)$. Let us then prove that $\alpha(t) := \langle \dot{\gamma}(t) \,|\, w\rangle = 0$ for all $t \in [0, T]$. By Proposition 1.60, the function $\alpha(t)$ is a solution of the Cauchy problem

$$\begin{cases} \ddot{\alpha}(t) + (\frac{1}{r^2} + c^2)\alpha(t) = 0 \\ \alpha(0) = \dot{\alpha}(0) = 0 \end{cases} \tag{1.67}$$

Since (1.67) admits the unique solution $\alpha(t) = 0$ for all t.

If the curve is a geodesic, then $c = 0$ and the geodesic equation is written as $\ddot{\gamma}(t) = -\gamma(t)$. 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 $\Gamma(0) = \langle \gamma(0) \,|\, w\rangle = -\langle \ddot{\gamma}(0) \,|\, w\rangle = 0$, by the assumption on w. This proves that the curve γ is contained in a plane passing through the origin. ☐

Remark 1.63. 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^2_r obtained as the surface of $\mathbb{R}^3$, endowed with the hyperbolic metric, defined as the zero level set of the function

$$a(x, y, z) = x^2 + y^2 - z^2 + r^2. \tag{1.68}$$

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

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

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

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

$$\mathcal{N} : H^2_r \to H^2, \qquad \mathcal{N}(q) = \frac{1}{r}\nabla_q a, \tag{1.69}$$

Exercise 1.64. 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 constant geodesic curvature curves on the hyperbolic space. With start with a result than can be proved in an analogous way to Proposition 1.60.

Proposition 1.65. *Let $\gamma : [0, T] \to H^2_r$ be a curve with 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*

$$\ddot\alpha(t) + \left(c^2 - \frac{1}{r^2}\right)\alpha(t) = 0. \tag{1.70}$$

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

Corollary 1.66. *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 1.67. Prove Proposition 1.65 and Corollary 1.66.

Geodesics on H^2_r are hyperbolas, obtained as intersections of the hyperboloid with plane passing through the origin. The classification of constant geodesic curvature curves 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 (1.70) also suggest, the value $c = \frac{1}{r}$ is a threshold and we have the following situation:

(i) if $0 \le c < 1/r$, 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 curves of type (i) are equidistant curves while curves of type (iii) are boundary of balls, i.e., spheres, in the hyperbolic plane. Finally, curves of type (ii) are also called *horocycles* (cf. Remark 1.63 for the difference with respect to the case of the positive constant curvature model).

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 γ_1 and γ_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 $\gamma(0) = q$.

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 γ 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 $t_K < b$.

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 ensure 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 is sub-linear, 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 $\mathbb{R}^n$.

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 too.

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 ϕ_t defined by (2.5) is called the *flow* generated by X. For the flow ϕ_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 2.7. When $X(x) = Ax$ is a linear vector field on $\mathbb{R}^n$, where A is a $n \times n$ matrix, the corresponding flow ϕ_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 2.8. 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 $t \to 0$:

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

Exercise 2.9. Let $a \in C^\infty(M)$ and $X \in \mathrm{Vec}(M)$, and denote $a_t = a \circ e^{tX}$. Prove the following formulas

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

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

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

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

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

Remark 2.10. Notice that, in coordinates, if $a \in C^\infty(M)$ and $X = \sum_i X_i(x)\frac{\partial}{\partial x_i}$ then $Xa = \sum_i X_i(x)\frac{\partial a}{\partial x_i}$. In particular, when X is applied to the coordinate functions $a_i(x) = x_i$ then $Xa_i = X_i$, which shows that a vector field is completely characterized by its action on functions.

Exercise 2.11. Let $f_1, \ldots, f_k \in C^\infty(M)$ and assume that $N = \{f_1 = \ldots = f_k = 0\} \subset M$ 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 $Xf_i(q) = 0$ for every $q \in N$ and $i = 1, \ldots, k$.

2.1.4 Nonautonomous vector fields

Definition 2.12. 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) $X(\cdot, q)$ is measurable for every fixed $q \in M$,

(C2) $X(t, \cdot)$ 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 L^∞ functions $c(t), k(t)$ such that

$$\|X(t, x)\| \leq c(t), \qquad \|X(t, x) - X(t, y)\| \leq k(t)\|x - y\|, \qquad \forall\,(t, x), (t, y) \in I \times K$$

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

Remark 2.13. In these lecture notes 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^∞ 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), by 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 [34]).

Theorem 2.14 (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 Lipschitz with respect to t and smooth with respect to q_0.*

Let us assume now that the equation (2.14) is *complete*, i.e., for all $t_0 \in \mathbb{R}$ and $q_0 \in M$ the solution $\gamma(t; t_0, q_0)$ 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. It satisfies

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

Moreover the following algebraic identities are satisfied

$$\begin{aligned} &P_{t,t} = \mathrm{Id}, \\ &P_{t_2,t_3} \circ P_{t_1,t_2} = P_{t_1,t_3}, \qquad \forall\, t_1, t_2, t_3 \in \mathbb{R}, \\ &(P_{t_1,t_2})^{-1} = P_{t_2,t_1}, \qquad \forall\, t_1, t_2 \in \mathbb{R}, \end{aligned} \tag{2.18}$$

Conversely, with every family of smooth diffeomorphism $P_{t,s} : M \to M$ satisfying the relations (2.18), that is called a *flow* on M, one can associate its *infinitesimal generator* X_t as follows:

$$X_t(q) = \left.\frac{d}{ds}\right|_{s=0} P_{t,t+s}(q), \qquad \forall\, q \in M. \tag{2.19}$$

The following lemma characterizes flows whose infinitesimal generator is autonomous.

Lemma 2.15. *Let $\{P_{t,s}\}_{t,s\in\mathbb{R}}$ be a family of smooth diffeomorphisms satisfying* (2.18)*. 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 2.16. Let $\varphi : M \to N$ a smooth map between smooth manifolds and $q \in M$. The *differential* of φ at the point q is the linear map

$$\varphi_{*,q} : T_qM \to T_{\varphi(q)}N, \tag{2.20}$$

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 γ.

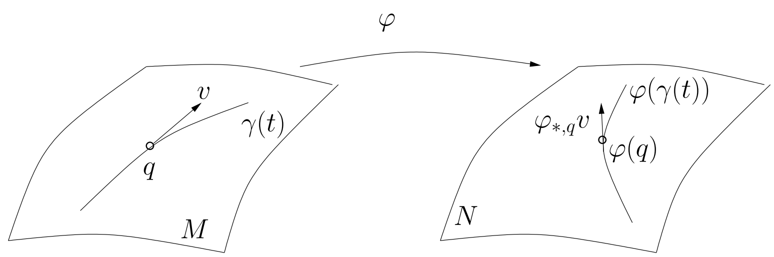


Figure 2.1: Differential of a map $\varphi : M \to N$

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 2.2).

Exercise 2.17. 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 2.18. 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

$$(\varphi_*X)(\varphi(q)) := \varphi_*(X(q)), \qquad \forall\, q \in M. \tag{2.21}$$

When $P \in \mathrm{Diff}(M)$ is a diffeomorphism on M, we can rewrite the identity (2.21) as

$$(P_*X)(q) = P_*(X(P^{-1}(q))), \qquad \forall\, q \in M. \tag{2.22}$$

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

Remark 2.19. From this definition it follows the useful formula for $X, Y \in \mathrm{Vec}(M)$

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

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

$$e^{tP_*X} = P \circ e^{tX} \circ P^{-1}, \tag{2.23}$$

$$(P_*X)a = (X(a \circ P)) \circ P^{-1}. \tag{2.24}$$

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 (2.23) follows. To prove (2.24) let us compute

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

Using (2.23) 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 2.20 one gets the following formula: for every $X, Y \in \mathrm{Vec}(M)$

$$(e_*^{tX}Y)a = Y(a \circ e^{tX}) \circ e^{-tX}. \tag{2.25}$$

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 expalined below, it measures how much Y is modified by the flow of X.

Definition 2.21. Let $X, Y \in \mathrm{Vec}(M)$. We define their *Lie bracket* as the vector field

$$[X, Y] := \frac{\partial}{\partial t}\bigg|_{t=0} e_*^{-tX}Y. \tag{2.26}$$

Remark 2.22. The geometric meaning of the Lie bracket can be understood by writing explicitly

$$[X, Y]\big|_q = \frac{\partial}{\partial t}\bigg|_{t=0} e_*^{-tX}Y\big|_q = \frac{\partial}{\partial t}\bigg|_{t=0} e_*^{-tX}(Y\big|_{e^{tX}(q)}) = \frac{\partial}{\partial s \partial t}\bigg|_{t=s=0} e^{-tX} \circ e^{sY} \circ e^{tX}(q). \tag{2.27}$$

Proposition 2.23. *As derivations on functions, one has the identity*

$$[X, Y] = XY - YX. \tag{2.28}$$

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 (2.25) we have

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

By Remark 2.8 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 $(XY - YX)a$. □

Proposition 2.23 shows that $(\mathrm{Vec}(M), [\cdot,\cdot])$ is a Lie algebra.

Exercise 2.24. 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 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 2.25. *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 $a \in C^\infty(M)$, preliminarly we see, using (2.24), 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 (2.28)

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

□

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 2.26. *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

$$[X, Y] = 0 \quad \Longrightarrow \quad e_*^{-tX} Y = Y, \quad \forall\, t \in \mathbb{R}. \tag{2.29}$$

To prove (2.29) 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 (2.29).

(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 (2.29). 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

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

Then (i) follows from (2.28). $\square$

Exercise 2.27. 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 tangent vector to the curve $t \mapsto \gamma(\sqrt{t})$ at $t = 0$ is $[X, Y](q)$.

Exercise 2.28. Let $X, Y \in \mathrm{Vec}(M)$. Using the semigroup property of the flow, prove that

$$\frac{d}{dt} e_*^{-tX} Y = e_*^{-tX}[X, Y] \tag{2.30}$$

Deduce the following expansion

$$\begin{aligned} e_*^{-tX} Y &= \sum_{n=0}^{\infty} \frac{t^n}{n!} (\mathrm{ad}\, X)^n Y \qquad (2.31) \\ &= Y + t[X, Y] + \frac{t^2}{2}[X, [X, Y]] + \frac{t^3}{6}[X, [X, [X, Y]]] + \dots \end{aligned}$$

Exercise 2.29. Let $X, Y \in \mathrm{Vec}(M)$ and $a \in C^\infty(M)$. Prove the following Leibnitz rule for the Lie bracket:

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

Exercise 2.30. Let $X, Y, Z \in \mathrm{Vec}(M)$. Prove that the Lie bracket satisfies the *Jacobi identity*:

$$[X, [Y, Z]] + [Y, [Z, X]] + [Z, [X, Y]] = 0. \tag{2.32}$$

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

Exercise 2.31. 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 $t = (t_1, \dots, t_n)$,

$$\frac{\partial \psi}{\partial t_i}(t) = e_*^{t_1 X_1} \circ \dots \circ e_*^{t_{i+1} X_{i+1}} X_i(\psi(t))$$

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 2.32. Let M be a smooth manifold. A *vector distribution* D of rank m on M is a family of vector subspaces $D_q \subset T_q M$ 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

$$D_q = \mathrm{span}\{X_1(q), \dots, X_m(q)\}, \qquad \forall\, q \in O_{q_0}. \tag{2.33}$$

Definition 2.33. 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 (2.33) and smooth functions a_{ij}^k on M such that

$$[X_i, X_k] = \sum_{j=1}^{m} a_{ij}^k X_j, \qquad \forall\, i, k = 1, \dots, m. \tag{2.34}$$

Exercise 2.34. 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 (2.33) there exist smooth functions a_{ij}^k such that (2.34) holds.

Definition 2.35. A smooth vector distribution D on M is said to be *flat* if for every point $q_0 \in M$ there exists a 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 2.36 (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

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

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

$$D_q = \text{span}\{X_1(q), \dots, X_m(q)\}, \qquad \forall\, q \in O_{q_0}. \tag{2.35}$$

and (2.34) are satisfied. We first need a lemma.

Lemma 2.37. *For every $k = 1, \dots, m$ we have $e_*^{tX_k} D = D$.*

Proof of Lemma 2.37. Let us define the time dependent vector fields

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

Using (2.34) and (2.30) 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

$$\dot{\Gamma}^k(t) = A^k(t)\Gamma^k(t), \qquad \Gamma^k(0) = I. \tag{2.36}$$

Then we have

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

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

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

which proves the claim. □

We can now end the proof of Theorem 2.36. Complete the family $X_1, \ldots, X_m$ to a basis of the tangent space

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

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 ψ is a local diffeomorphism at $(t, s) = (0, 0)$ and for (t, s) close to $(0, 0)$ we have that (cf. Exercice 2.31)

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

for every $i = 1, \ldots, m$. These vectors are linearly independent and, thanks to Lemma 2.37, 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. $\square$

2.5 Cotangent space

In this section we introduce tangent 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 2.38. 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 action of the covector λ 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 2.39. Let $\varphi : M \to N$ be a smooth map and $q \in M$. The *pullback* of φ 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

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

Example 2.40. Let $a : M \to \mathbb{R}$ be a smooth function and $q \in M$. The differential $d_q a$ of the function a at the point $q \in M$, defined through the formula

$$\langle d_q a, v \rangle := \left.\frac{d}{dt}\right|_{t=0} a(\gamma(t)), \qquad v \in T_qM, \tag{2.37}$$

where γ is any smooth curve such that $\gamma(0) = q$ and $\dot\gamma(0) = v$, is an element of T_q^*M, since (2.37) is linear with respect to v.

Definition 2.41. 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 $\omega \in \Lambda^1 M$ on $X \in \mathrm{Vec}(M)$ pointwise, defining a function on M.

$$\langle \omega, X \rangle : q \mapsto \langle \omega(q), X(q) \rangle . \tag{2.38}$$

The differential form ω 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 2.42. 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 manifold M, is a smooth family of vector spaces parametrized by points in M.

Definition 2.43. 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

$$\begin{tikzcd} \pi^{-1}(O_q) \arrow[r, "\psi"] \arrow[dr, "\pi"'] & O_q \times \mathbb{R}^k \arrow[d, "\pi_1"] \\ & O_q \end{tikzcd} \tag{2.39}$$

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

Remark 2.44. 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. 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 2.45. 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

$$\left.\frac{\partial}{\partial x_i}\right|_q \in T_q M, \qquad i = 1, \ldots, n, \qquad q \in O, \tag{2.40}$$

$$\left\langle dx_i, \frac{\partial}{\partial x_j} \right\rangle = \delta_{ij}, \qquad i, j = 1, \ldots, n. \tag{2.41}$$

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

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

and the maps

$$\psi : v \mapsto (x_1, \ldots, x_n, v_1, \ldots, v_n), \qquad \bar{\psi} : \lambda \mapsto (x_1, \ldots, x_n, p_1, \ldots, p_n), \tag{2.42}$$

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

Definition 2.46. 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

$$\begin{tikzcd} E \arrow[r, "f"] \arrow[dr, "\pi"'] & E' \arrow[d, "\pi'"] \\ & M \end{tikzcd} \tag{2.43}$$

where f is linear on fibers. Here π and π' denote the canonical projections.

Definition 2.47. 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 σ is defined on all M it is said to be a *global section*.

Example 2.48. 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 $M_0 := \zeta(M) \subset E$.

Remark 2.49. 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 construction on vector bundles.

Definition 2.50. 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)* φ^*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 2.51. (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

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

(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

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

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

2.7 Submersions and level sets of smooth maps

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

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

Theorem 2.52 (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 $(x_1, \dots, x_m)$ centered at q and $(y_1, \dots, y_n)$ centered at $\varphi(q)$ in which φ has the following coordinate representation:*

$$\varphi(x_1, \dots, x_m) = (x_1, \dots, x_k, 0, \dots, 0). \tag{2.44}$$

Remark 2.53. 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) φ 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 φ is linear.

In the case of a submersion, from Theorem 2.52 one can deduce the following result.

Corollary 2.54. *Assume $\varphi : M \to N$ is a smooth submersion at q. Then φ admits a local right inverse at $\varphi(q)$. Moreover φ is open at q. More precisely it exist $\varepsilon > 0$ and $C > 0$ such that*

$$B_{\varphi(q)}(C^{-1}r) \subset \varphi(B_q(r)), \qquad \forall\, r \in [0, \varepsilon), \tag{2.45}$$

where the balls in (2.45) *are considered with respect to some Euclidean norm in a coordinate chart.*

Remark 2.55. The constant C appearing in (2.45) 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 φ is a diffeomorphism, C is a bound on the norm of the differential of the inverse of φ. 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 2.56 (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 2.57. It is worth to specify the following two important sub cases of Theorem 2.56:

(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 2.58. 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 $N_c = a^{-1}(c) = \{q \in M \mid a(q) = c\} \subset M$ is a smooth submanifold. Prove that for every $q \in N_c$

$$T_q N_c = \ker d_q a = \{v \in T_q M \mid \langle d_q a, v\rangle = 0\}.$$

Bibliographical notes

The material presented in this chapter is classical and covered by many textbook in differential geometry, as for instance in [28, 73, 46, 92].

Theorem 2.14 is a well-known theorem in ODE. The statement presented here can be deduced from [35, Theorem 2.1.1, Exercice 2.4]. The functions $c(t), k(t)$ appearing in (C3) are assumed to be L^∞, that is stronger than L^1 (on compact intervals). This stronger assumptions imply that the solution is not only absolutely continuous with respect to t, but also locally Lipschitz.

Chapter 3

Sub-Riemannian structures

3.1 Basic definitions

In this section we introduce a definition of sub-Riemannian structure which is quite general. Indeed, this definition includes all the classical notions of Riemannian structure, constant-rank sub-Riemannian structure, rank-varying sub-Riemannian structure, almost-Riemannian structure etc.

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

$$\mathrm{Lie}_q \mathcal{F} := \{X(q), X \in \mathrm{Lie}\,\mathcal{F}\} = T_q M, \quad \forall\, q \in M.$$

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

$$\mathrm{Lie}^s_q \mathcal{F} := \{X(q), X \in \mathrm{Lie}^s \mathcal{F}\} = T_q M,$$

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 structures.

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 $(\cdot \mid \cdot)_q$, smooth with respect to q. For $u \in U_q$ we denote the norm of u as $|u|^2 = (u \mid u)_q$.

(ii) $f : \mathbf{U} \to TM$ is a smooth map that is a morphism of vector bundles, i.e. the following diagram is commutative (here $\pi_{\mathbf{U}} : \mathbf{U} \to M$ and $\pi : TM \to M$ are the canonical projections)

$$\begin{tikzcd} \mathbf{U} \arrow[r, "f"] \arrow[dr, "\pi_{\mathbf{U}}"'] & TM \arrow[d, "\pi"] \\ & M \end{tikzcd} \tag{3.3}$$

and f is *linear* on fibers.

(iii) The *set of horizontal vector fields* $\mathcal{D} := \{f(\sigma) \,|\, \sigma : M \to \mathbf{U} \text{ smooth section}\}$, 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 3.23).

Definition 3.3. Let $(M, \mathbf{U}, f)$ be a sub-Riemannian manifold. The *distribution* is the family of subspaces

$$\{\mathcal{D}_q\}_{q \in M}, \qquad \text{where} \qquad \mathcal{D}_q := f(U_q) \subset T_q M.$$

We call $k(q) := \dim \mathcal{D}_q$ 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 $k(q)$ is constant. Otherwise we say that the sub-Riemannian structure is *rank-varying*.

The set of horizontal vector fields $\mathcal{D} \subset \operatorname{Vec}(M)$ has the structure of a finitely generated $C^\infty(M)$-module, whose elements are vector fields tangent to the distribution at each point, i.e.

$$\mathcal{D}_q = \{X(q) |\, X \in \mathcal{D}\}.$$

The rank of a sub-Riemannian structure $(M, \mathbf{U}, f)$ satisfies

$$k(q) \leq m, \qquad \text{where } m = \operatorname{rank} \mathbf{U}, \tag{3.4}$$

$$k(q) \leq n, \qquad \text{where } n = \dim M. \tag{3.5}$$

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 $T_q M$.

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

$$u : t \in [0, T] \mapsto u(t) \in U_{\gamma(t)}, \tag{3.6}$$

called the *control function*, such that

$$\dot{\gamma}(t) = f(\gamma(t), u(t)), \qquad \text{for a.e. } t \in [0, T]. \tag{3.7}$$

In this case we say that $u(\cdot)$ is a *control corresponding* to γ. Notice that different controls could correspond to the same trajectory (see Figure 3.1).

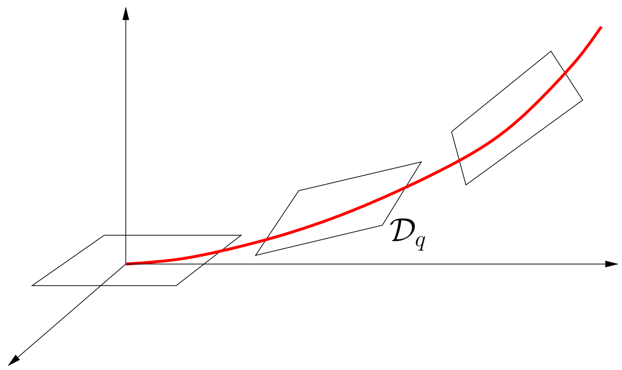


Figure 3.1: A horizontal curve

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 $u = (u_1, \dots, u_m) \in L^\infty([0,T], \mathbb{R}^m)$ such that

$$\dot{\gamma}(t) = \sum_{i=1}^m u_i(t) f_i(\gamma(t)), \qquad \text{for a.e. } t \in [0,T]. \tag{3.8}$$

Thanks to this local characterization and Theorem 2.14, 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 γ, defined on a sufficiently small interval, such that u is the control associated with γ 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 rank two and defined by

$$f(x,y,u_1,u_2) = (x,y,u_1,u_2 x), \qquad f'(x,y,u_1,u_2) = (x,y,u_1,u_2 x^2). \tag{3.9}$$

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, γ 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.

This example 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 $\mathcal{D} \neq \mathcal{D}'$. Notice, however, 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 3.7. Let $v \in \mathcal{D}_q$. We define the *sub-Riemannian norm* of v as follows

$$\|v\| := \min\{|u|,\, u \in U_q \ \text{ s.t. } \ v = f(q,u)\}. \tag{3.10}$$

Notice that since f is linear with respect to u, the minimum in (3.10) 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 (3.10) is uniquely attained at the orthogonal projection of the origin onto this subspace (see Figure 3.2).

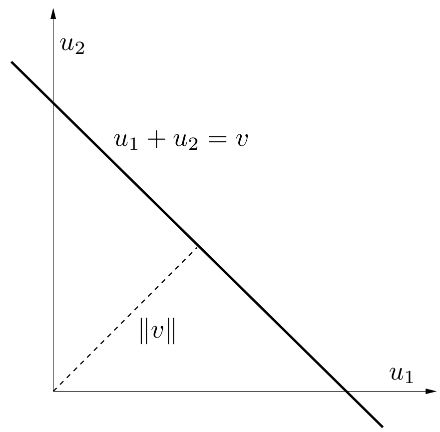


Figure 3.2: The norm of a vector v for $f(x, u_1, u_2) = u_1 + u_2$

Exercise 3.8. 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

$$\langle v \,|\, w \rangle_q = \frac{1}{4}\|v+w\|^2 - \frac{1}{4}\|v-w\|^2, \qquad v, w \in \mathcal{D}_q. \tag{3.11}$$

Exercise 3.9. 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 the unique control $t \mapsto u^*(t)$ associated with γ and realizing the minimum in (3.10).

Definition 3.10. Given an admissible curve $\gamma : [0,T] \to M$, we define

$$u^*(t) := \operatorname{argmin} \{|u|,\, u \in U_q \ \text{ s.t. } \ \dot\gamma(t) = f(\gamma(t), u)\}. \tag{3.12}$$

for all differentiability point of γ. We say that the control u^* is the *minimal control* associated with γ.

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 3.11. *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 3.12. 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.

Consider, as in Remark 3.6, the free sub-Riemannian structure on $\mathbb{R}^2$

$$f(x, y, u_1, u_2) = (x, y, u_1, u_2 x), \tag{3.13}$$

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 3.11 we are allowed to introduce the following definition.

Definition 3.13. Let $\gamma : [0, T] \to M$ be an admissible curve. We define the *sub-Riemannian length* of γ as

$$\ell(\gamma) := \int_0^T \|\dot{\gamma}(t)\| dt. \tag{3.14}$$

We say that γ is *length-parametrized* (or *arclength parametrized*) if $\|\dot{\gamma}(t)\| = 1$ for a.e. $t \in [0, T]$. Notice that for a length-parametrized curve we have that $\ell(\gamma) = T$.

Formula (3.14) says that the length of an admissible curve is the integral of the norm of its minimal control.

$$\ell(\gamma) = \int_0^T |u^*(t)| dt. \tag{3.15}$$

In particular any admissible curve has finite length.

Lemma 3.14. *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 γ_φ is a composition of Lipschitz functions, hence Lipschitz. Moreover γ_φ 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 γ. Using the change of variables $t = \varphi(s)$, one gets

$$\ell(\gamma_\varphi) = \int_0^{T'} \|\dot{\gamma}_\varphi(s)\| ds = \int_0^{T'} |u^*(\varphi(s))||\dot{\varphi}(s)| ds = \int_0^T |u^*(t)| dt = \int_0^T \|\dot{\gamma}(t)\| dt = \ell(\gamma). \tag{3.16}$$

□

Lemma 3.15. *Every admissible curve of positive length is a Lipschitz reparametrization of a length-parametrized admissible one.*

Proof. Let $\psi : [0, T] \to M$ be an admissible curve with minimal control u^*. Consider the Lipschitz monotone function $\varphi : [0, T] \to [0, \ell(\psi)]$ defined by

$$\varphi(t) := \int_0^t |u^*(\tau)| d\tau.$$

Notice that if $\varphi(t_1) = \varphi(t_2)$, the monotonicity of φ ensures $\psi(t_1) = \psi(t_2)$. Hence we are allowed to define $\gamma : [0, \ell(\psi)] \to M$ by

$$\gamma(s) := \psi(t), \qquad \text{if } s = \varphi(t) \text{ for some } t \in [0, T].$$

In other words, it holds $\psi = \gamma \circ \varphi$. To show that γ 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)

$$|\psi(t_1) - \psi(t_0)| \leq C \int_{t_0}^{t_1} |u^*(\tau)| d\tau.$$

Indeed fix $K \subset M$ a compact set such that $\psi([0, T]) \subset K$ and set $C := \max_{x \in K} \left(\sum_{i=1}^{m} |f_i(x)|^2 \right)^{1/2}$. Then

$$\begin{aligned}
|\psi(t_1) - \psi(t_0)| &\leq \int_{t_0}^{t_1} \sum_{i=1}^{m} |u_i^*(t) f_i(\psi(t))| \, dt \\
&\leq \int_{t_0}^{t_1} \sqrt{\sum_{i=1}^{m} |u_i^*(t)|^2} \sqrt{\sum_{i=1}^{m} |f_i(\psi(t))|^2 dt} \\
&\leq C \int_{t_0}^{t_1} |u^*(t)| dt,
\end{aligned}$$

Hence if $s_1 = \varphi(t_1)$ and $s_0 = \varphi(t_0)$ one has

$$|\gamma(s_1) - \gamma(s_0)| = |\psi(t_1) - \psi(t_0)| \leq C \int_{t_0}^{t_1} |u^*(\tau)| d\tau = C|s_1 - s_0|,$$

which proves that γ is Lipschitz. It particular $\dot{\gamma}(s)$ exists for a.e. $s \in [0, \ell(\psi)]$.

We are going to prove that γ 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

$$v(s) := \frac{u^*(t)}{\dot{\varphi}(t)} = \frac{u^*(t)}{|u^*(t)|}.$$

By Exercise 3.16 the control v is defined for a.e. s. Moreover, by construction, $|v(s)| = 1$ for a.e. s and v is the minimal control associated with γ. $\square$

Exercise 3.16. Show that for a Lipschitz and monotone function $\varphi : [0, T] \to \mathbb{R}$, the Lebesgue measure of the set $\{s \in \mathbb{R} \,|\, s = \varphi(t), \dot{\varphi}(t) \text{ exists}, \dot{\varphi}(t) = 0\}$ is zero.

By the previous discussion, in what follows, it will be often convenient to assume that admissible curves are 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 3.17. Let $(\mathbf{U}, f), (\mathbf{U}', f')$ be two sub-Riemannian structures on a smooth manifold M. They are said to be *equivalent* if the following conditions are satisfied

(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

$$\begin{tikzcd} & \mathbf{U} \arrow[dr, "f"] & \\ \mathbf{V} \arrow[ur, "p"] \arrow[dr, "p'"'] & & TM \\ & \mathbf{U}' \arrow[ur, "f'"'] & \end{tikzcd} \tag{3.17}$$

(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 3.18. If $(\mathbf{U}, f)$ and $(\mathbf{U}', f')$ are equivalent 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.

Remark 3.19. Notice that (i) is satisfied (with the vector bundle $\mathbf{V}$ possibly non Euclidean) if and only if the two moduli of horizontal vector fields $\mathcal{D}$ and $\mathcal{D}'$ defined by $\mathbf{U}$ and $\mathbf{U}'$ are equal (cf. Definition 3.2).

Definition 3.20. 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 3.21. 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 3.22. Two equivalent classes of sub-Riemannian manifolds are said to be *isometric* if there exist two representatives $(M, \mathbf{U}, f), (M', \mathbf{U}', f')$, a diffeomorphism $\phi : M \to M'$ and an isomorphism[1] of Euclidean bundles $\psi : \mathbf{U} \to \mathbf{U}'$ such that the following diagram is commutative

$$\begin{array}{ccc} \mathbf{U} & \xrightarrow{f} & TM \\ {\scriptstyle\psi}\downarrow & & \downarrow{\scriptstyle\phi_*} \\ \mathbf{U}' & \xrightarrow[f']{} & TM' \end{array} \tag{3.18}$$

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 3.23. Show that every Riemannian manifold in the sense of Definition 3.2 is indeed equivalent to a Riemannian structure in the classical sense above (cf. Exercise 3.8).

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 is equal to the dimension of the manifold, at every point.

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 $u = (u_1, \dots, u_m) \in \mathbb{R}^m$.

   If we denote by $\{e_1, \dots, e_m\}$ 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 $i = 1, \dots, m$. Then we have

$$f(q, u) = f\left(q, \sum_{i=1}^m u_i e_i\right) = \sum_{i=1}^m u_i f_i(q), \qquad q \in M. \tag{3.19}$$

   In this case, the problem of finding an admissible curve joining two fixed points $q_0, q_1 \in M$

[1] isomorphism of bundles in the broad sense, it is fiberwise but is not obliged to map a fiber in the same fiber.

and with minimal length is rewritten as the optimal control problem

$$\begin{cases} \dot{\gamma}(t) = \displaystyle\sum_{i=1}^{m} u_i(t) f_i(\gamma(t)) \\ \displaystyle\int_0^T |u(t)| dt \to \min \\ \gamma(0) = q_0, \quad \gamma(T) = q_1 \end{cases} \tag{3.20}$$

For a free sub-Riemannian structure, the set of vector fields $f_1, \dots, f_m$ 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 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 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

$$f : \mathbf{U} \to TM, \qquad f(q, u) = \pi_q^{\perp}(u) \in T_q M. \tag{3.21}$$

   where $\pi_q^{\perp} : \mathbb{R}^3 \to T_q M \subset \mathbb{R}^3$ is the orthogonal projection.

   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 3.24. Show that $(\mathbf{U}, f)$ defined in (3.21) 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 big rank.

Lemma 3.25. *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)

$$T_M E \simeq E \oplus TM. \tag{3.22}$$

Eq. (3.22) 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_qE_q so that

$$T_qE = T_qE_q \oplus T_qM \simeq E_q \oplus T_qM, \qquad \forall\, q \in M.$$

By Whitney theorem we have a (nonlinear on fibers, in general) immersion

$$\Psi : E \to \mathbb{R}^N, \qquad \Psi_* : T_ME \subset TE \hookrightarrow T\mathbb{R}^N,$$

for $N = 2(n+m)$, and Ψ_* is injective as bundle map, i.e., T_ME 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.

$$E' = \bigcup_{q\in M} E'_q, \qquad E'_q = (T_qE_q \oplus T_qM)^\perp,$$

and considering $E_0 := T_ME \oplus E'$ we have that E_0 is trivial since its fibers are sum of orthogonal complements and by (3.22) we are done.

□

Corollary 3.26. *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 3.25 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

$$\begin{tikzcd} & \overline{\mathbf{U}} \arrow[dr, "\bar{f}"] & \\ \mathbf{U} \oplus \mathbf{U}' \arrow[ur, "\mathrm{Id}"] \arrow[dr, "p_1"'] & & TM \\ & \mathbf{U} \arrow[ur, "f"'] & \end{tikzcd} \tag{3.23}$$

is commutative. Moreover condition (ii) of Definition 3.17 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

$$\dot{\gamma}(t) = \sum_{i=1}^{m} u_i(t) f_i(\gamma(t)), \tag{3.24}$$

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 γ 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 γ.

Notice that Corollary 3.26 implies that the modulus of horizontal vector fields $\mathcal{D}$ is globally generated by $f_1, \dots, f_m$.

Remark 3.27. The integral curve $\gamma(t) = e^{tf_i}$, defined on $[0, T]$, of an element f_i of a generating family $\mathcal{F} = \{f_1, \dots, f_m\}$ is admissible and $\ell(\gamma) \leq T$. If $\mathcal{F} = \{f_1, \dots, f_m\}$ are linearly independent then they are an orthonormal frame and $\ell(\gamma) = T$.

Exercise 3.28. Consider a sub-Riemannian structure $(\mathbf{U}, f)$ over M. Let $m = \text{rank}(\mathbf{U})$ and $h_{\max} = \max\{h(q) : q \in M\} \leq m$ 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 $\text{rank}(\overline{\mathbf{U}}) = h_{\max}$.

3.1.5 Proto sub-Riemannian 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. In what follows we call these structures *proto sub-Riemannian structures*.

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, and each of them is endowed with a Riemannian structure.

3.2 Sub-Riemannian distance and Chow-Rashevskii theorem

In this section we introduce the sub-Riemannian distance between two points as the infimum of the length of admissible curves joining them.

Recall that, in the definition of sub-Riemannian manifold, M is assumed to be connected. Moreover, thanks to the construction of Section 3.1.4, in what follows we can assume that the sub-Riemannian structure is free, with generating family $\mathcal{F} = \{f_1, \dots, f_m\}$. Notice that, by definition, $\mathcal{F}$ is assumed to be bracket generating.

Definition 3.29. 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

$$d(q_0, q_1) = \inf\{\ell(\gamma) \,|\, \gamma : [0, T] \to M \text{ admissible}, \ \gamma(0) = q_0, \ \gamma(T) = q_1\}, \tag{3.25}$$

One of the purpose of this section is to show that, thanks to the bracket generating condition, (9.1) is well-defined, namely for every $q_0, q_1 \in M$, there exists an admissible curve that joins q_0 to q_1, hence $d(q_0, q_1) < +\infty$.

Theorem 3.30 (Chow-Raschevskii). *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.

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 3.30. 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 Chow-Raschevskii 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

$$\gamma : [0, T_1 + T_2] \to M, \qquad \gamma(t) = \begin{cases} \gamma_1(t), & t \in [0, T_1], \\ \gamma_2(t - T_1), & t \in [T_1, T_1 + T_2]. \end{cases} \tag{3.26}$$

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 3.31. *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 (2.27)) 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. $\square$

Lemma 3.32. *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 3.33. Notice that, if $\mathcal{D}_{q_0} \neq T_{q_0}M$, then $\widehat{s} = 0$ cannot be a regular point of the map ψ. Indeed, for $s = 0$, the image of the differential of ψ at 0 is $\mathrm{span}_{q_0}\{f_{i_j}, j = 1, \ldots, n\} \subset \mathcal{D}_{q_0}$ and the differential of ψ 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 3.32. 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 Σ_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 Σ_1. Otherwise, by Lemma 3.31, dim $\mathrm{Lie}_q\mathcal{F} = 1$, 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 Σ_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 Σ_2. Otherwise, by Lemma 3.31, $\dim \mathrm{Lie}_{q_1}\mathcal{D} = 2$, 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 ϕ_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 3.32 there exists a neighborhood $\widehat{V} \subset V$ of $\widehat{s}$ such that ψ 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 3.33.

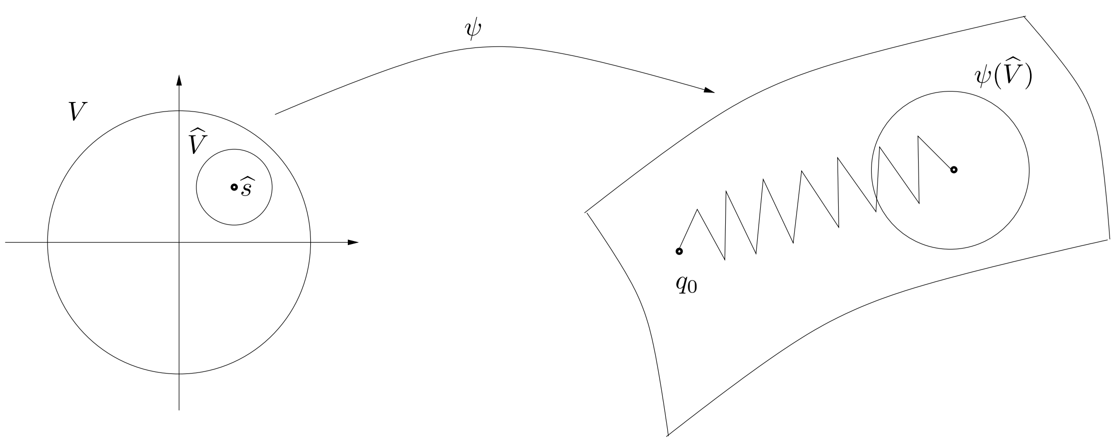


Figure 3.3: Proof of Lemma 3.32

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 γ the admissible curve joining q_0 to q, built by $2n$-pieces, as in Figure 3.4.

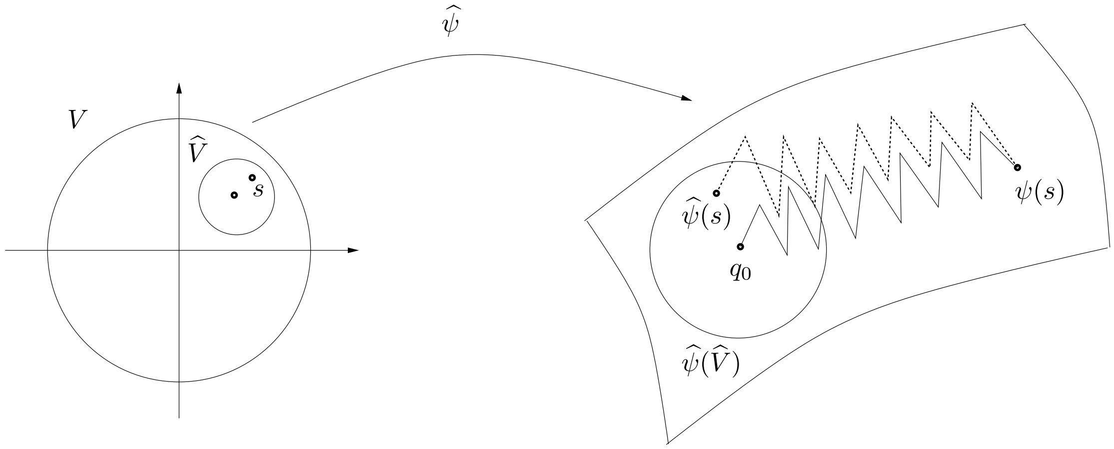


Figure 3.4: The map $\widehat{\psi}$

In other words γ 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 3.27). 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. $\square$

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

$$q_1 \sim q_2 \qquad \text{if} \qquad d(q_1, q_2) < +\infty. \tag{3.27}$$

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. $\square$

Lemma 3.34. *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 γ 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

$$C_K := \max_{x \in K} \left(\sum_{i=1}^{m} |f_i(x)|^2 \right)^{1/2} \tag{3.28}$$

and fix $\delta_K > 0$ such that $\operatorname{dist}(q_0, \partial K) > C_K \delta_K$ (here dist is the Euclidean distance, 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 $t^* := \sup\{t \in [0,T] : \gamma([0,t]) \subset K\}$, with $t^* < T$. Then

$$|\gamma(t^*) - \gamma(0)| \leq \int_0^{t^*} |\dot\gamma(t)| dt \leq \int_0^{t^*} \sum_{i=1}^m |u_i^*(t) f_i(\gamma(t))| \, dt \tag{3.29}$$

$$\leq \int_0^{t^*} \sqrt{\sum_{i=0}^m |f_i(\gamma(t))|^2} \sqrt{\sum_{i=0}^m u_i^*(t)^2} \, dt \tag{3.30}$$

$$\leq C_K \int_0^{t^*} \sqrt{\sum_{i=0}^m u_i^*(t)^2} \, dt \leq C_K \ell(\gamma) \tag{3.31}$$

$$\leq C_K \delta_K < \mathrm{dist}(q_0, \partial K). \tag{3.32}$$

which contradicts the fact that, at t^*, the curve γ leaves the compact K. Thus $t^* = T$. □

Proof of Step 3. Let us prove that Lemma 3.34 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 3.34, each admissible curve joining q_0 and q_1 has length greater than δ_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 δ_K as in Lemma 3.34, 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 3.34, $\gamma_n([0,T]) \subset K$ for all n.

We can repeat estimates (3.29)-(3.31) 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

$$|q - q_0| \leq C_K d(q_0, q) \leq C_K \delta < \varepsilon. \tag{3.33}$$

□

Corollary 3.35. *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. □

3.3 Existence of length-minimizers

In this section we want to discuss the existence of length-minimizers.

Definition 3.36. Let $\gamma : [0,T] \to M$ be an admissible curve. We say that γ 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 3.37. 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 3.38. *Let $\gamma_n : [0,T] \to M$ be a sequence of admissible curves on M such that $\gamma_n \to \gamma$ uniformly on $[0,T]$. Then*

$$\ell(\gamma) \leq \liminf_{n\to\infty} \ell(\gamma_n). \tag{3.34}$$

If moreover $\liminf_{n\to\infty} \ell(\gamma_n) < +\infty$, then γ is also admissible.

Proof. Let $L := \liminf_{n\to\infty} \ell(\gamma_n)$. If $L = +\infty$ the inequality (3.34) is true, thus we can assume $L < +\infty$ and choose a subsequence, still denoted by the same symbol, 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 γ_n are all contained in a common compact set K. Now we divide the proof into two steps

(i). We first prove that statement assuming that all γ_n 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

$$V_q = \{f_u(q), |u| \leq L+\delta\} \subset T_qM, \qquad f_u(q) = \sum_{i=1}^m u_i f_i(q).$$

Notice that V_q is convex for every $q \in M$, thanks to the linearity of f in u. Let us prove that γ is admissible and satisfies $\ell(\gamma) \leq L+\delta$. Once this is done, since δ is arbitrary, this implies $\ell(\gamma) \leq L$, that is (3.34).

Writing in local coordinates, we have for every $\varepsilon > 0$

$$\frac{1}{\varepsilon}(\gamma_n(t+\varepsilon) - \gamma_n(t)) = \frac{1}{\varepsilon}\int_t^{t+\varepsilon} f_{u_n(\tau)}(\gamma_n(\tau))d\tau \in \mathrm{conv}\{V_{\gamma_n(\tau)}, \tau \in [t,t+\varepsilon]\}. \tag{3.35}$$

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 (3.31) gives for $\tau \in [t, t+\varepsilon]$

$$|\gamma_n(t) - \gamma_n(\tau)| \leq \int_t^\tau |\dot\gamma_n(s)|ds \leq C_K(L+\delta)\varepsilon. \tag{3.36}$$

where C_K is the constant (3.28) 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 ε. 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

$$\mathrm{conv}\{V_{\gamma_n(\tau)},\, \tau \in [t, t+\varepsilon]\} \subset \mathrm{conv}\{V_q,\, q \in B_{\gamma(t)}(r_\varepsilon)\}. \tag{3.38}$$

Plugging (3.38) in (3.35) and passing to the limit for $n \to \infty$ we get finally to

$$\frac{1}{\varepsilon}(\gamma(t+\varepsilon)-\gamma(t)) \in \mathrm{conv}\{V_q,\, q \in B_{\gamma(t)}(r_\varepsilon)\}. \tag{3.39}$$

Assume now that $t \in [0,1]$ is a differentiability point of γ. Then the limit of the left hand side in (3.39) for $\varepsilon \to 0$ exists and gives $\dot\gamma(t) \in \mathrm{conv}\, V_{\gamma(t)} = V_{\gamma(t)}$. For every differentiability point t we can thus define the unique $u^*(t)$ satisfying $\dot\gamma(t) = f(\gamma(t), u^*(t))$ and $|u^*(t)| = \|\dot\gamma(t)\|$. Using the argument contained in Appendix 3.5 it follows that $u^*(t)$ is measurable in t. Moreover $|u^*(t)|$ is essentially bounded since, by construction, $|u^*(t)| \leq L+\delta$ for a.e. $t \in [0,T]$. Hence γ is admissible. Moreover $\ell(\gamma) \leq L+\delta$ since γ is defined on the interval $[0,1]$.

(ii) When $\gamma_n : [0,T] \to M$ is an arbitrary sequence converging uniformly to γ, let us consider the family $\widetilde\gamma_n : [0,1] \to M$ such that $\widetilde\gamma_n$ is parametrized by constant speed on $[0,1]$ (cf. Lemma 3.15). In particular

$$\gamma_n = \widetilde\gamma_n \circ \varphi_n, \qquad \varphi_n(t) = \frac{1}{\ell(\gamma_n)}\int_0^t |u_n^*(s)|ds$$

To prove the statement it is enough to prove that $\widetilde\gamma_n \to \widetilde\gamma$ where γ is some reparametrization of $\widetilde\gamma$, since length is invariant by reparametrization. Reasoning as in the proof of part (i) one gets

$$|\widetilde\gamma_n(s_1) - \widetilde\gamma_n(s_0)| \leq C_K(L+\delta)|s_1 - s_0|$$

then we can apply the Ascoli-Arzela theorem on the reparametrized sequence and we get that a subsequence is uniformly convergent to a curve, that is necessarily a curve $\widetilde\gamma$ whose γ is a reparametrization. □

Corollary 3.39. *Let γ_n be a sequence of length-minimizers on M such that $\gamma_n \to \gamma$ uniformly. Then γ is a length-minimizer.*

Proof. Since the length is invariant under reparametrization, it is not restrictive to assume that all curves γ_n and γ are parametrized on $[0,1]$. Since γ_n is a length-minimizer one has $\ell(\gamma_n) = d(\gamma_n(0), \gamma_n(1))$. By uniform convergence $\gamma_n(t) \to \gamma(t)$ for every $t \in [0,1]$ and, by continuity of the distance and semicontinuity of the length

$$\ell(\gamma) \leq \liminf_{n\to\infty} \ell(\gamma_n) = \liminf_{n\to\infty} d(\gamma_n(0), \gamma_n(1)) = d(\gamma(0), \gamma(1)),$$

that implies that $\ell(\gamma) = d(\gamma(0), \gamma(1))$, i.e., γ is a length-minimizer. □

The semicontinuity of the length implies the existence of minimizers, under a natural compactness assumption on the space.

Theorem 3.40 (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*

$$d(q_0, q_1) = \min\{\ell(\gamma) \,|\, \gamma : [0,T] \to M \text{ admissible}, \gamma(0) = q_0, \gamma(T) = q_1\}.$$

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 γ_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 γ_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 3.38, the curve γ satisfies $\ell(\gamma) \leq \liminf \ell(\gamma_n) = d(q_0, q_1)$, that implies $\ell(\gamma) = d(q_0, q_1)$. □

Remark 3.41. 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 $\bar{B}(q, r)$ is compact also, being a closed subset of a compact set $\overline{B}(q, r_0)$.

Combining Theorem 3.40 and Corollary 3.35 one gets the following corollary.

Corollary 3.42. *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 space. We start by proving a preliminary lemma.

Lemma 3.43. *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 $y \in B(x, r + \varepsilon) \setminus B(x, \varepsilon)$. Then there exists a length-parameterized curve γ connecting x with y such that $\ell(\gamma) = t + \varepsilon$ where $0 \leq t < r$. Let $t' \in (t, r)$; then $\gamma(t') \in B(x, r)$ and $y \in B(\gamma(t'), \varepsilon)$. □

Proposition 3.44. *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 $\{x_n\}$ in M is convergent. Fix $\varepsilon > 0$ satisfying the assumption. Since $\{x_n\}$ 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 Cauchy and admits a convergent subsequence, that implies that the whole sequence $\{x_n\}$ in M is convergent.

(ii) implies (iii). This is trivial.

(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 3.41, $r \in A$ 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 $r(y) < \varepsilon$ small enough such that $\bar{B}(y, r(y))$ is compact. We have

$$\bar{B}(x, r) \subset \bigcup_{y \in \bar{B}(x,r)} \bar{B}(y, r(y)).$$

By compactness of $\bar{B}(x, r)$ there exists a finite number of points $\{y_i\}_{i=1}^N$ in $\bar{B}(x, r)$ such that (denote $r_i := r(y_i)$)

$$\bar{B}(x, r) \subset \bigcup_{i=1}^N \bar{B}(y_i, r_i).$$

Moreover, there exists $\delta > 0$ such that the set of points $\bar{B}(x, r+\delta) = \{y \in M \,|\, \mathrm{dist}(y, B(x, r)) \leq \delta\}$, where the equality is given by Lemma 3.43, satisfies

$$\bar{B}(x, r + \delta) \subset \bigcup_{i=1}^N \bar{B}(y_i, r_i).$$

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 $B := \bar{B}(x, R)$ is compact. Since B is a closed set, it is enough to show that it is totally bounded, i.e. it admits an ε-net[2] 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 3.43 one has for every $y \in B$ that $\mathrm{dist}(y, B') < \varepsilon/3$. Then it is easy to show that

$$\mathrm{dist}(y, S) < \mathrm{dist}(y, B') + \varepsilon/3 < \varepsilon,$$

that is S is an ε-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 $R = +\infty$. □

Remark 3.45. Notice that only in the "(i) implies (ii)" part of the statement we used that the distance is sub-Riemannian. Actually the same statement, together with Lemma 3.43, remains true in the more general context of length metric space, see [38, Ch. 2].

For the relation with geodesic completeness of the sub-Riemannian manifold, see Section 11.5.

Corollary 3.46. *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.*

[2] an ε-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 $\mathsf{d}(y, z_i) < \varepsilon$).

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 3.47. *Let $\gamma : [0,T] \to M$ be a continuous curve. Then γ is Lipschitz with respect to the sub-Riemannian distance if and only if γ is admissible.*

Proof. (i). Assume γ is admissible and leu u be a control associated with γ. By definition u is essentially bounded. Then

$$d(\gamma(t), \gamma(s)) \leq \ell(\gamma|_{[t,s]}) \leq \int_s^t |u(\tau)| d\tau \leq C|t-s|,$$

for some constant $C > 0$. Then γ is Lipschitz with respect to the sub-Riemannian distance.

(ii). Conversely assume that γ 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 3.34 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 γ is Lipschitz in charts and γ is differentiable almost everywhere by Rademacher theorem.

Let us prove that γ 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 $t_{i,n} := i/2^n$ for $i = 1, \ldots, 2^n$. By compactness of small balls and compactness of $[0,T]$ for n large enough there exists a minimizer joining $\gamma(t_{i,n})$ and $\gamma(t_{i+1,n})$ for $i = 1, \ldots, 2^n - 1$.

Denote by γ_n the curve defined by the concatenation of 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

$$\ell(\gamma_n) = \sum_{i=1}^{2^n} d(\gamma(t_{i,n}), \gamma(t_{i+1,n})) \leq \sum_{i=1}^{2^n} L|t_{i,n} - t_{i+1,n}| \leq \sum_{i=1}^{2^n} \frac{L}{2^n} \leq L \tag{3.45}$$

Moreover, by construction, γ_n converge uniformly to γ when $n \to \infty$. By Theorem 3.38 γ is admissible and $\ell(\gamma) \leq L$. ☐

Exercise 3.48. 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 φ_n is Lipschitz for every n and $v_\gamma(t) = \sup_n |\dot\varphi_n(t)|$ for a.e $t \in [0,T]$.

Exercise 3.49. 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 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 $f : \mathbb{R}^m \times M \to TM$ of rank m.

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

$$\|f' - f\|_{0,K} = \sup\{|f'(q,v) - f(q,v)| : q \in K,\ |v| \le 1\}.$$

The family of seminorms $\|\cdot\|_{0,K}$ induces a topology on $\mathcal{S}_m$ with countable local bases 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 base of neighborhood of f is given by

$$U_{f,n} := \left\{ f' \in \mathcal{S}_m : \|f' - f\|_{0,K_n} \le \frac{1}{n} \right\}, \qquad n \in \mathbb{N}. \tag{3.48}$$

Exercise 3.50. (i) Prove that (3.48) defines a basis for a topology. (ii) Prove that this topology does not depend on the immersion of M into $\mathbb{R}^N$.

For $f \in \mathcal{S}_m$, we denote by d_f the sub-Riemannian distance on M associated with f.

Theorem 3.51. *Let $q_0, q_1 \in M$. The function $\mathrm{dist}_{q_0,q_1} : \mathcal{S}_m \to \mathbb{R}$ defined by $f \mapsto d_f(q_0, q_1)$ is continuous in the C^0 topology.*

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 ε 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.

Given $u \in L^\infty([0,1];\mathbb{R}^m)$, let us denote by $\gamma_f(t;u)$ the solution of the equation $\dot q = f(q,u)$ with initial condition $q(0) = q_0$. 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 $0 \le a \le 1$ and $a|_O \equiv 1$. By compactness, there exists $C > 0$ such that

$$|a(q')f(q',v) - a(q)f(q,v)| \le C|q' - q|, \quad \forall q, q' \in M,\ |v| \le 1. \tag{3.49}$$

Given $f' : \mathbb{R}^m \times M \to TM$ a complete sub-Riemannian structure, we set:

$$\delta_u(t) := |\gamma_{af'}(t;u) - \gamma_{af}(t;u)|.$$

Combining the definition of $\delta_u(t)$ and (3.49) one gets

$$\delta_u(t) \leq C \int_0^t \delta_u(s)\,ds + \|af' - af\|_{0,K} \int_0^t |u(s)|\,ds, \quad 0 \leq t \leq 1. \tag{3.50}$$

Using that $\|af' - af\|_{0,K} \leq \|f' - f\|_{0,K}$ and the Gronwall lemma, the inequality (3.50) 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 ε small enough we have that $\gamma_{af'}(t;u)$ belongs to O for every control u such that $\|u\|_{L^\infty} \leq r$. In particular, since $a = 1$ on O, we have $\gamma_{af'}(t;u) = \gamma_{f'}(t;u)$ for every $t \in [0,1]$ and the ball $B'_{q_0}(r) \subset O$, 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, $n = \dim M$) and

$$\widehat{\psi} : (s_1, \ldots, s_n) \mapsto e^{-\hat{s}_1 f_{i_1}} \circ \cdots \circ e^{-\hat{s}_n f_{i_n}} \circ e^{s_n f_{i_n}} \circ \cdots \circ e^{s_1 f_{i_1}}(q_1),$$

constructed as in the proof of the Chow–Rashevskii theorem, associated to the base point q_1 and defined for $|s| < \varepsilon$. 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 $\hat{\psi}$. The map $\hat{\psi}'$ 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 3.52. 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 3.52. *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 Φ 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. □

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-minimizer 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 Chow and Filippov theorems, 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 $\{\gamma_{q_0}(\varepsilon) \,|\, \gamma_{q_0}$ 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 3.53 (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[3] *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*

$$\lambda(t) := (P_{0,t}^{-1})^* \lambda_0, \qquad \lambda(t) \in T^*_{\gamma(t)}M, \tag{3.51}$$

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 $\lambda_0 \neq 0$.

Notice that, by definition, the curve $\lambda(t)$ is Lipschitz continuous. Moreover the conditions (N) and (A) are mutually exclusive, unless $\overline{u}(t) = 0$ for a.e. $t \in [0,T]$, i.e., γ is the trivial trajectory.

[3] $P_{0,t}(x)$ is defined for $t \in [0,T]$ and x in a neighborhood of $\gamma(0)$

Definition 3.54. 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 (3.51).

- 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*).

Remark 3.55. If the sub-Riemannian structure is not Riemannian at q_0, namely if

$$\mathcal{D}_{q_0} = \text{span}_{q_0}\{f_1, \dots, f_m\} \neq T_{q_0}M,$$

then the trivial trajectory, corresponding to $\overline{u}(t) \equiv 0$, is always normal and abnormal.

Notice that even a nontrivial admissible trajectory γ 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).

Remark 3.56. In the Riemannian case there are no abnormal extremals. Indeed, since the map f is fiberwise surjective, we can always find m vector fields $f_1, \dots, f_m$ on M such that

$$\text{span}_{q_0}\{f_1, \dots, f_m\} = T_{q_0}M,$$

and (A) would imply that $\langle \lambda_0, v\rangle = 0$, for all $v \in T_{q_0}M$, that gives the contradiction $\lambda_0 = 0$.

Exercise 3.57. Prove that condition (N) of Theorem 3.51 implies that the minimal control $\overline{u}(t)$ is smooth. In particular normal extremals are smooth.

At this level it seems not obvious how to use Theorem 3.53 to find the explicit expression of extremals for a given problem. In the next chapter we provide another formulation of Theorem 3.53 which gives Pontryagin extremals as solutions of a Hamiltonian system.

The rest of this section is devoted to the proof of Theorem 3.53.

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 γ.

Remark 3.58. While ℓ is invariant by reparametrization (see Remark 3.14), 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 ℓ.

Lemma 3.59. *Fix $T > 0$ and let Ω_{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 Ω_{q_0,q_1} if and only if it is a minimizer of ℓ on Ω_{q_0,q_1} and has constant speed.*

Proof. Applying the Cauchy-Schwarz inequality

$$\left(\int_0^T f(t)g(t)dt\right)^2 \leq \int_0^T f(t)^2 dt \int_0^T g(t)^2 dt, \tag{3.52}$$

with $f(t) = \|\dot\gamma(t)\|$ and $g(t) = 1$ we get

$$\ell(\gamma)^2 \leq 2J(\gamma)T. \tag{3.53}$$

Moreover in (3.52) equality holds if and only if f is proportional to g, i.e., $\|\dot\gamma(t)\| = \text{const.}$ in (3.53). Since, by Lemma 3.15, every curve is a Lipschitz reparametrization of a 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 3.53

By Lemma 3.59 we can assume that γ 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

$$\widetilde{J}(u(\cdot)) := \frac{1}{2}\int_0^T |u(t)|^2 dt, \tag{3.54}$$

on the space of controls $u(\cdot) \in L^\infty([0, T], \mathbb{R}^m)$, the minimal control $\overline{u}(\cdot)$ of γ 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

$$\dot q(t) = f_{u(t)}(q(t)), \qquad q(0) = q_0, \tag{3.55}$$

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

$$q(t) = P_{0,t}(x(t)). \tag{3.56}$$

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 $x(t) \equiv q_0$.

The next step is to write the ODE satisfied by $x(t)$. Differentiating (3.56) we get

$$\begin{aligned}\dot q(t) &= f_{\overline{u}(t)}(q(t)) + (P_{0,t})_*(\dot x(t)) &(3.57)\\ &= f_{\overline{u}(t)}(P_{0,t}(x(t))) + (P_{0,t})_*(\dot x(t)) &(3.58)\end{aligned}$$

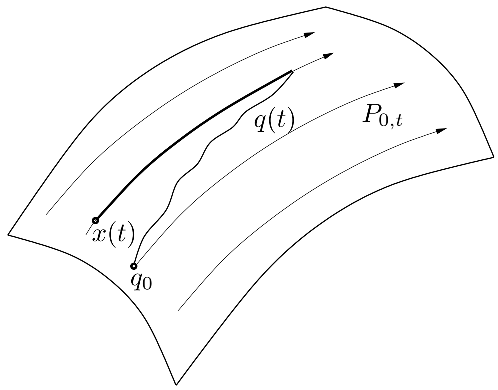


Figure 3.5: The trajectories $q(t)$, associated with $u(\cdot) = \overline{u}(\cdot) + v(\cdot)$, and the corresponding $x(t)$.

and using that $\dot{q}(t) = f_{u(t)}(q(t)) = f_{u(t)}(P_{0,t}(x(t)))$ we can invert (3.58) with respect to $\dot{x}(t)$ and rewrite it as follows

$$\begin{aligned}\dot{x}(t) &= (P_{0,t}^{-1})_* \left[(f_{u(t)} - f_{\overline{u}(t)})(P_{0,t}(x(t)))\right] \\ &= \left[(P_{0,t}^{-1})_*(f_{u(t)} - f_{\overline{u}(t)})\right](x(t)) \\ &= \left[(P_{0,t}^{-1})_*(f_{u(t)-\overline{u}(t)})\right](x(t)) \\ &= \left[(P_{0,t}^{-1})_* f_{v(t)}\right](x(t)) \end{aligned} \tag{3.59}$$

If we define the nonautonomous vector field $g^t_{v(t)} = (P_{0,t}^{-1})_* f_{v(t)}$ we finally obtain by (3.59) the following Cauchy problem for $x(t)$

$$\dot{x}(t) = g^t_{v(t)}(x(t)), \qquad x(0) = q_0. \tag{3.60}$$

Notice that the vector field g^t_v 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 (3.60), starting from q_0, corresponding to control $\overline{u}(\cdot) + sv(\cdot)$, and $J(\overline{u} + sv)$ is the associated cost.

Lemma 3.60. *There exists* $\bar{\lambda} \in (\mathbb{R} \oplus T_{q_0}M)^*$, *with* $\bar{\lambda} \neq 0$, *such that for all* $v \in L^\infty([0,T], \mathbb{R}^m)$

$$\left\langle \bar{\lambda}, \left(\frac{\partial J(\overline{u} + sv)}{\partial s}\Big|_{s=0}, \frac{\partial x(T; \overline{u} + sv)}{\partial s}\Big|_{s=0} \right) \right\rangle = 0. \tag{3.61}$$

Proof of Lemma 3.60. We argue by contradiction: assume that (3.61) is not true, then there exist $v_0, \ldots, v_n \in L^\infty([0,T], \mathbb{R}^m)$ such that the vectors in $\mathbb{R} \oplus T_{q_0}M$

$$\begin{pmatrix} \dfrac{\partial J(\overline{u} + sv_0)}{\partial s}\Big|_{s=0} \\ \dfrac{\partial x(T; \overline{u} + sv_0)}{\partial s}\Big|_{s=0} \end{pmatrix}, \ \ldots\ , \begin{pmatrix} \dfrac{\partial J(\overline{u} + sv_n)}{\partial s}\Big|_{s=0} \\ \dfrac{\partial x(T; \overline{u} + sv_n)}{\partial s}\Big|_{s=0} \end{pmatrix} \tag{3.62}$$

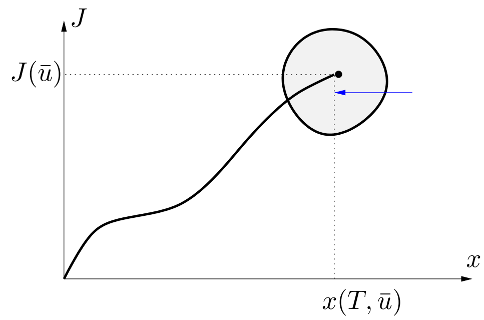
are linearly independent. Let us then consider the map

$$\Phi : \mathbb{R}^{n+1} \to \mathbb{R} \times M, \qquad \Phi(s_0, \ldots, s_n) = \begin{pmatrix} J(\overline{u} + \sum_{i=0}^n s_i v_i) \\ x(T; \overline{u} + \sum_{i=0}^n s_i v_i) \end{pmatrix}. \tag{3.63}$$

By differentiability properties of solution of smooth ODEs with respect to parameters, the map (3.63) is smooth in a neighborhood of $s = 0$. Moreover, since the vectors (3.62) are the components of the differential of Φ and they are independent, then the inverse function theorem implies that Φ 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 3.61. Notice that if $\bar{\lambda}$ satisfies (3.61), then for every $\alpha \in \mathbb{R}$, with $\alpha \neq 0$, $\alpha\bar{\lambda}$ satisfies (3.61) 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 (3.61) implies that there exists $\lambda_0 \in T^*_{q_0}M$ such that one of the following identities is satisfied for all $v \in L^\infty([0, T], \mathbb{R}^m)$:

$$\frac{\partial J(\overline{u} + sv)}{\partial s}\Big|_{s=0} = \left\langle \lambda_0, \frac{\partial x(T; \overline{u} + sv)}{\partial s}\Big|_{s=0} \right\rangle, \tag{3.64}$$

$$0 = \left\langle \lambda_0, \frac{\partial x(T; \overline{u} + sv)}{\partial s}\Big|_{s=0} \right\rangle. \tag{3.65}$$

with $\lambda_0 \neq 0$ in the second case (cf. Remark 3.61). To end the proof we have to show that identities (3.64) and (3.65) are equivalent to conditions (N) and (A) of Theorem 3.53. Let us show that

$$\frac{\partial J(\overline{u} + sv)}{\partial s}\Big|_{s=0} = \int_0^T \sum_{i=1}^m \overline{u}_i(t) v_i(t) dt, \tag{3.66}$$

$$\frac{\partial x(T; \overline{u} + sv)}{\partial s}\Big|_{s=0} = \int_0^T g^t_{v(t)}(q_0) dt = \int_0^T \sum_{i=1}^m ((P_{0,t}^{-1})_* f_i)(q_0) v_i(t) dt. \tag{3.67}$$

The identity (3.66) follows from the definition of J

$$J(\overline{u}+sv) = \frac{1}{2}\int_0^T |\overline{u}+sv|^2 dt. \tag{3.68}$$

Eq. (3.67) can be proved in coordinates. Indeed by (3.60) 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 (3.67).

Let us show that (3.64) is equivalent to (N) of Theorem 3.53. Similarly, one gets that (3.65) is equivalent to (A). Using (3.66) and (3.67), equation (3.64) is rewritten as

$$\begin{aligned}\int_0^T \sum_{i=1}^m \overline{u}_i(t)v_i(t)dt &= \int_0^T \sum_{i=1}^m \left\langle \lambda_0, ((P_{0,t}^{-1})_* f_i)(q_0)\right\rangle v_i(t)dt \\ &= \int_0^T \sum_{i=1}^m \langle \lambda(t), f_i(\gamma(t))\rangle\, v_i(t)dt, \end{aligned}\tag{3.69}$$

where we used, for every $i=1,\ldots,m$, 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 Main 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 3.62. *Under assumptions (M1)-(M4), the function $t \mapsto |u^*(t)|$ is measurable on I.*

Proof. Denote $\varphi(t) := |u^*(t)|$. To prove the lemma we show that for every fixed $r > 0$ the set

$$A = \{t \in I : \varphi(t) \leq r\}$$

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

$$A = \bigcap_{n\in\mathbb{N}} A_n = \bigcap_{n\in\mathbb{N}} \underbrace{\bigcup_{i\in\mathbb{N}} A_{i,n}}_{:=A_n} \tag{3.70}$$

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 (17.12) 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 3.63. *Under assumptions (M1)-(M4), the vector function $t \mapsto u^*(t)$ is measurable on I.*

Proof. 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

$$B = \{t \in I : \exists\, u \in O \ s.t.\ |u| = \varphi(t), g(t,u) = v(t)\}.$$

Let us fix the ball O and a countable dense set $\{u_i\}_{i\in\mathbb{N}}$ in O. Let show that

$$B = \bigcap_{n\in\mathbb{N}} B_n = \bigcap_{n\in\mathbb{N}} \underbrace{\bigcup_{i\in\mathbb{N}} B_{i,n}}_{:=B_n} \tag{3.71}$$

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 (3.71) 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 $t \in B$. $\square$

3.5.2 Proof of Lemma 3.11

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 3.15, we can assume that γ is 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 3.63 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^∞ controls). However, the theory can be equivalently developed in the framework of H^1 admissible curves (corresponding to L^2 controls) or in the framework of absolutely continuous admissible curves (corresponding to L^1 controls).

Definition 3.64. 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 H^1-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 3.14 and 3.15 can be rewritten in the absolutely continuous framework in the following form.

Lemma 3.65. *The length of an AC-admissible curve is invariant by AC reparametrization.*

Lemma 3.66. *Any AC-admissible curve of positive length is a AC reparametrization of a length-parametrized admissible one.*

The proof of Lemma 3.65 differs from the one of Lemma 3.14 only by the fact that, if $u^* \in L^1$ is the minimal control of γ then $(u^* \circ \varphi)\dot{\varphi}$ is the minimal control associated with $\gamma \circ \varphi$. Moreover $(u^* \circ \varphi)\dot{\varphi} \in L^1$, using the monotonicity of φ. Under these assumptions the change of variables formula (3.16) still holds.

The proof of Lemma 3.66 is unchanged. Notice that the statement of Exercise 3.16 remains true if we replace Lipschitz with absolutely continuous. We stress that the curve γ built in the proof is Lipschitz (since it is length-parametrized).

As a consequence of these results, if we define

$$d_{AC}(q_0, q_1) = \inf\{\ell(\gamma) \mid \gamma : [0, T] \to M \ AC\text{-admissible},\ \gamma(0) = q_0,\ \gamma(T) = q_1\}, \tag{3.72}$$

we have the following proposition.

Proposition 3.67. $d_{AC}(q_0, q_1) = d(q_0, q_1)$

Since $L^2([0, T]) \subset L^1([0, T])$, Lemmas 3.65, 3.66 and Proposition 3.67 are valid also in the framework of admissible curves associated with L^2 controls.

Bibliographical notes

Sub-Riemannian manifolds have been introduced, even if with different terminology, in several contexts starting from the end of 60s, see for instance [68, 63, 50, 64, 54] and [69, 70, 83, 55, 36, 19, 37, 100]. However, some pioneering ideas were already present in the work of Carathéodory and Cartan. The name sub-Riemannian geometry first appeared in [93].

Classical general references for sub-Riemannian geometry are [78, 18, 77, 57, 97]. Recent monographs [67, 88].

The definition of sub-Riemannian manifold using the language of bundles dates back to [7, 18]. For the original proof of the Raschevski-Chow theorem see [85, 44]. The problem of the measurability of the minimal control can be seen as a problem of differential inclusion [35]. The proof of existence of sub-Riemannian length minimizer presented here is an adaptation of the proof of Filippov theorem in optimal control. The fact that in sub-Riemannian geometry there exist abnormal length minimizers is due to Montgomery [76, 78]. The fact that the theory can be equivalently developed for Lipschitz or absolutely continuous curves is well known, a discussion can be found in [18]. A sub-Riemannian manifold, from the metric viewpoint, is a length space. A link with this theory is provided by Exercices 3.48-3.49, see also [38, Ch. 2].

The characterization of Pontryagin extremals given in Theorem 3.53 is a simplified version of the Pontryagin Maximum Priciple (PMP) [84]. The proof presented here is original and adapted to this setting. For more general versions of PMP see [8, 26]. 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 [88, 98].

Chapter 4

Characterization and local minimality of Pontryagin extremals

This chapter is devoted to the study of geometric properties of Pontryagin extremals. To this purpose we first rewrite Theorem 3.53 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 by the key concept of Poisson bracket.

4.1 Geometric characterization of Pontryagin extremals

In the previous chapter we proved that if $\gamma : [0,T] \to M$ is a length minimizer on a sub-Riemannian manifold, associated with a control $u(\cdot)$, then there exists $\lambda_0 \in T^*_{\gamma(0)}M$ such that defining

$$\lambda(t) = (P_{0,t}^{-1})^* \lambda_0, \qquad \lambda(t) \in T^*_{\gamma(t)}M, \tag{4.1}$$

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 (non-autonomous) 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.

Functions in $C^\infty(M)$ are in a one-to-one correspondence with functions in $C^\infty(T^*M)$ that are *constant on fibers* via the map $\alpha \mapsto \pi^*\alpha = \alpha \circ \pi$. In other words we have the isomorphism of algebras

$$C^\infty(M) \simeq C^\infty_{\mathsf{cst}}(T^*M) := \{\pi^*\alpha \,|\, \alpha \in C^\infty(M)\} \subset C^\infty(T^*M). \tag{4.3}$$

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 $q = \pi(\lambda)$.

$$\mathrm{Vec}(M) \simeq C^\infty_{\mathsf{lin}}(T^*M) := \{a_Y \,|\, Y \in \mathrm{Vec}(M)\} \subset C^\infty(T^*M). \tag{4.4}$$

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 $C^\infty_{\mathsf{lin}}(T^*M)$ is a module over $C^\infty(M)$ as well. For any $\alpha \in C^\infty(M)$ and $a_X \in C^\infty_{\mathsf{lin}}(T^*M)$ their product is defined as $\alpha a_X := (\pi^*\alpha)a_X = a_{\alpha X} \in C^\infty_{\mathsf{lin}}(T^*M)$.

Definition 4.1. We say that a function $a \in C^\infty(T^*M)$ is *affine on fibers* if there exist two functions $\alpha \in C^\infty_{\mathsf{cst}}(T^*M)$ and $a_X \in C^\infty_{\mathsf{lin}}(T^*M)$ 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 $C^\infty_{\mathsf{aff}}(T^*M)$ 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 λ. 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 λ 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 2.15 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, of the form $\beta \circ \pi$ with $\beta \in C^\infty(M)$, coincides with the action of X. Indeed we have for all $\lambda \in T^*M$

$$\frac{d}{dt}\bigg|_{t=0} \beta \circ \pi((e^{-tX})^*\lambda)) = \frac{d}{dt}\bigg|_{t=0} \beta(e^{tX}(q)) = (X\beta)(q), \qquad q = \pi(\lambda). \tag{4.6}$$

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 $C^\infty_{\mathfrak{aff}}(T^*M)$ 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 in $C^\infty_{\mathfrak{lin}}(T^*M)$, where it reflects the Lie bracket of vector fields in $\mathrm{Vec}(M)$, seen as elements of $C^\infty_{\mathfrak{lin}}(T^*M)$. Then it is uniquely extended to $C^\infty_{\mathfrak{aff}}(T^*M)$ and $C^\infty(T^*M)$ by requiring that it is a derivation of the algebra $C^\infty(T^*M)$ in each argument.

More precisely we start by the following definition.

Definition 4.3. Let $a_X, a_Y \in C^\infty_{\mathfrak{lin}}(T^*M)$ be associated with vector fields $X, Y \in \mathrm{Vec}(M)$. Their *Poisson bracket* is defined by

$$\{a_X, a_Y\} := a_{[X,Y]}, \tag{4.9}$$

where $a_{[X,Y]}$ is the function in $C^\infty_{\mathfrak{lin}}(T^*M)$ associated with the vector field $[X,Y]$.

Remark 4.4. Recall that the Lie bracket is a bilinear, skew-symmetric map defined on $\mathrm{Vec}(M)$, that satisfies the Leibnitz rule for $X, Y \in \mathrm{Vec}(M)$:

$$[X, \alpha Y] = \alpha[X,Y] + (X\alpha)Y, \qquad \forall\, \alpha \in C^\infty(M). \tag{4.10}$$

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 relation makes sense since the product between $\alpha \in C^\infty_{\mathfrak{cst}}(T^*M)$ and $a_X \in C^\infty_{\mathfrak{lin}}(T^*M)$ belong to $C^\infty_{\mathfrak{lin}}(T^*M)$, namely $\alpha a_X = a_{\alpha X}$.

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 $C^\infty_{\mathfrak{aff}}(T^*M)$.

By linearity and skew-symmetry we are reduced to compute Poisson brackets of kind $\{a_X, \alpha\}$ and $\{\alpha, \beta\}$, where $a_X \in C^\infty_{\mathfrak{lin}}(T^*M)$ and $\alpha, \beta \in C^\infty_{\mathfrak{cst}}(T^*M)$. 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

$$\{a_X + \alpha, a_Y + \beta\} := a_{[X,Y]} + X\beta - Y\alpha. \tag{4.17}$$

From the explicit formula (4.17) it is easy to see that the Poisson bracket computed at a fixed $\lambda \in T^*M$ depends only on the differential of the two functions $a_X + \alpha$ and $a_Y + \beta$ at λ.

Next we extend this definition to $C^\infty(T^*M)$ in such a way that it is still 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 $C^\infty_{\mathfrak{aff}}(T^*M)$ 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

$$\begin{aligned}
\{f,g\} &= a_{[X,Y]} + X\beta - Y\alpha \\
&= \sum_{i,j=1}^n p_j \left(X_i \frac{\partial Y_j}{\partial x_i} - Y_i \frac{\partial X_j}{\partial x_i} \right) + X_i \frac{\partial \beta}{\partial p_i} - Y_i \frac{\partial \alpha}{\partial p_i} \\
&= \sum_{i,j=1}^n X_i \left(p_j \frac{\partial Y_j}{\partial x_i} + \frac{\partial \beta}{\partial p_i} \right) - Y_i \left(p_j \frac{\partial X_j}{\partial x_i} + \frac{\partial \alpha}{\partial p_i} \right) \\
&= \sum_{i=1}^n \frac{\partial f}{\partial p_i}\frac{\partial g}{\partial x_i} - \frac{\partial f}{\partial x_i}\frac{\partial g}{\partial p_i}.
\end{aligned}$$

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. $\square$

Remark 4.7. We stress that the value $\{a,b\}|_\lambda$ at a point $\lambda \in T^*M$ depends only on $d_\lambda a$ and $d_\lambda b$. Hence 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 $f = (f_1, \dots, f_k) : T^*M \to \mathbb{R}^k$, $g : T^*M \to \mathbb{R}$ and $\varphi : \mathbb{R}^k \to \mathbb{R}$ be smooth functions. Denote by $\varphi_f := \varphi \circ f$. Prove that

$$\{\varphi_f, g\} = \sum_{i=1}^k \frac{\partial \varphi}{\partial f_i}\{f_i, g\}. \tag{4.20}$$

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 $a \in C^\infty(T^*M)$

$$\vec{a} = \sum_{i=1}^{n} \frac{\partial a}{\partial p_i}\frac{\partial}{\partial x_i} - \frac{\partial a}{\partial x_i}\frac{\partial}{\partial p_i}. \tag{4.22}$$

The following proposition gives the explicit form of the vector field V on T^*M generating the flow $(P_{0,t}^{-1})^*$.

Proposition 4.10. *Let $X \in \mathrm{Vec}(M)$ be complete 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 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.

Proposition 4.12. *Let X_t be a nonautonomous vector field and 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 $(P_{0,t}^{-1})^*$.

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 $x = (x_1, \dots, x_n)$ on M, canonical coordinates (p, x) on T^*M are coordinates for which 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

$$w = \sum_{i=1}^{n} \alpha_i \frac{\partial}{\partial p_i} + \beta_i \frac{\partial}{\partial x_i} \qquad \Longrightarrow \qquad \pi_* w = \sum_{i=1}^{n} \beta_i \frac{\partial}{\partial x_i},$$

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 λ coincides with the one of λ 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 ω 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 σ is a closed 2-form on T^*M. Moreover its expression in canonical coordinates (p, x) shows immediately that is a nondegenerate two form

$$\sigma = \sum_{i=1}^{n} dp_i \wedge dx_i. \tag{4.25}$$

Remark 4.16 (The symplectic form in non-canonical coordinates)*.* Given a basis of 1-forms $\omega_1, \dots, \omega_n$ in $\Lambda^1(M)$, 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 h_i become coordinates on the fibers. Notice that these coordinates are not related to any 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 σ an extra term appears since, in general, the 1-forms ω_i are not closed.

4.2.1 The symplectic form vs the Poisson bracket

Let V be a finite dimensional vector space and V^* denotes 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 $v \in V$, that is 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^*. Applying the same reasoning to the linear map L^{-1} one obtain a bilinear map on V^*.

Exercise 4.17. (a). Let $h \in C^\infty(T^*M)$. Prove that the Hamiltonian vector field $\vec{h} \in \mathrm{Vec}(T^*M)$ satisfies the following identity

$$\sigma(\cdot, \vec{h}(\lambda)) = d_\lambda h, \qquad \forall\, \lambda \in T^*M.$$

(b). Prove that, for every $\lambda \in T^*M$ the bilinear forms σ_λ 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 show that

$$\{a, b\} = \vec{a}(b) = \langle db, \vec{a}\rangle = \sigma(\vec{a}, \vec{b}), \qquad \forall\, a, b \in C^\infty(T^*M). \tag{4.28}$$

Remark 4.18. Notice that σ is nondegenerate, which means that the map $w \mapsto \sigma_\lambda(\cdot, w)$ defines a linear isomorphism between the vector spaces $T_\lambda(T^*M)$ and $T^*_\lambda(T^*M)$. Hence $\vec{h}$ is the vector field canonically associated by the symplectic structure with the differential dh. For this reason $\vec{h}$ is also called *symplectic gradient* of h.

From formula (4.25) we have that in canonical coordinates (p, x) the Hamiltonian vector filed 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

$$\begin{cases} \dot{x}_i = \dfrac{\partial h}{\partial p_i} \\ \dot{p}_i = -\dfrac{\partial h}{\partial x_i} \end{cases}, \qquad i = 1, \ldots, n.$$

We conclude this section with two classical but rather important results:

Proposition 4.19. *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 (4.28), we have the following formula for the derivative of the function a along the solution

$$\frac{d}{dt}a(\lambda(t)) = \{h, a\}(\lambda(t)). \tag{4.29}$$

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 is a is constant. □

The skew-simmetry of the Poisson brackets immediately implies the following corollary.

Corollary 4.20. *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 extremals

Now we can rewrite Theorem 3.53 using the symplectic language developed in the last section.

Given a sub-Riemannian structure on M with generating family $\{f_1, \ldots, f_m\}$, and define the fiberwise linear functions on T^*M associated with these vector fields

$$h_i : T^*M \to \mathbb{R}, \qquad h_i(\lambda) := \langle \lambda, f_i(q) \rangle, \quad i = 1, \ldots, m.$$

Theorem 4.21 (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*

$$\dot{\lambda}(t) = \sum_{i=1}^{m} \overline{u}_i(t)\vec{h}_i(\lambda(t)), \qquad a.e.\ t \in [0, T], \tag{4.30}$$

and one of the following conditions is satisfied:

(N) $h_i(\lambda(t)) \equiv \overline{u}_i(t), \quad i = 1, \ldots, m,\ \forall\, t,$

(A) $h_i(\lambda(t)) \equiv 0, \qquad i = 1, \ldots, m,\ \forall\, t.$

Moreover in case (A) one has $\lambda(t) \neq 0$ for all $t \in [0, T]$.

Proof. The statement is a rephrasing of Theorem 3.53, obtained by combining Proposition 4.10 and Exercise 4.12. □

Notice that Theorem 4.21 says that normal and abnormal extremals appear as solution of an Hamiltonian system. Nevertheless, this Hamiltonian system is non autonomous 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 4.21 for the necessary condition for optimality still does not clarify if the extremals depend on the generating family $\{f_1, \ldots, f_m\}$ for 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 4.22. Let M be a sub-Riemannian manifold. The *sub-Riemannian Hamiltonian* is the function on T^*M defined as follows

$$H : T^*M \to \mathbb{R}, \qquad H(\lambda) = \max_{u \in U_q} \left(\langle \lambda, f_u(q) \rangle - \frac{1}{2}|u|^2 \right), \quad q = \pi(\lambda). \tag{4.31}$$

Proposition 4.23. *The sub-Riemannian Hamiltonian H is smooth and quadratic on fibers. Moreover, for every generating family $\{f_1, \ldots, f_m\}$ of the sub-Riemannian structure, the sub-Riemannian Hamiltonian H is written as follows*

$$H(\lambda) = \frac{1}{2} \sum_{i=1}^{m} \langle \lambda, f_i(q) \rangle^2, \qquad \lambda \in T_q^*M, \quad q = \pi(\lambda). \tag{4.32}$$

Proof. In terms of a generating family $\{f_1, \ldots, f_m\}$, the sub-Riemannian Hamiltonian (4.31) is written as follows

$$H(\lambda) = \max_{u \in \mathbb{R}^m} \left(\sum_{i=1}^{m} u_i \langle \lambda, f_i(q) \rangle - \frac{1}{2} \sum_{i=1}^{m} u_i^2 \right). \tag{4.33}$$

Differentiating (4.33) with respect to u_i, one gets that the maximum in the r.h.s. is attained at $u_i = \langle \lambda, f_i(q) \rangle$, from which formula (4.32) follows. The fact that H is smooth and quadratic on fibers then easily follows from (4.32). $\square$

Exercise 4.24. Prove that two equivalent sub-Riemannian structures $(\mathbf{U}, f)$ and $(\mathbf{U}', f')$ on a manifold M define the same Hamiltonian.

Exercise 4.25. Consider the sub-Riemannian Hamiltonian $H : T^*M \to \mathbb{R}$. Denote by $H_q : T_q^*M \to \mathbb{R}$ its restriction on fiber 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: use that, if $f_1, \ldots, f_m$ is a generating frame, then

$$d_\lambda H_q = \sum_{i=1}^{m} \langle \lambda, f_i(q) \rangle f_i(q).$$

Theorem 4.26. *Every normal extremal is a solution of the Hamiltonian system $\dot{\lambda}(t) = \vec{H}(\lambda(t))$. In particular, every normal extremal trajectory is smooth.*

Proof. Denoting, as usual, $h_i(\lambda) = \langle \lambda, f_i(q) \rangle$ for $i = 1, \ldots, m$, the functions linear on fibers associated with a generating family and using the identity $\overrightarrow{h_i^2} = 2h_i\vec{h}_i$ (see (4.12)), it follows that

$$\vec{H} = \overrightarrow{\frac{1}{2}\sum_{i=1}^{m} h_i^2} = \sum_{i=1}^{m} h_i\vec{h}_i.$$

In particular, since along a normal extremal $h_i(\lambda(t)) = \overline{u}_i(t)$ by condition (N) of Theorem 4.21, one gets

$$\vec{H}(\lambda(t)) = \sum_{i=1}^{m} h_i(\lambda(t))\vec{h}_i(\lambda(t)) = \sum_{i=1}^{m} \overline{u}_i(t)\vec{h}_i(\lambda(t)).$$

□

Remark 4.27. 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

$$\begin{cases} \dot{x} = \dfrac{\partial H}{\partial p} = \sum_{i=1}^{m} \langle p, f_i(x) \rangle f_i(x) \\ \dot{p} = -\dfrac{\partial H}{\partial x} = -\sum_{i=1}^{m} \langle p, f_i(x) \rangle \langle p, D_x f_i(x) \rangle \end{cases} \tag{4.34}$$

From here it is easy to see that if $\lambda(t) = (p(t), x(t))$ is a solution of (4.34) 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 $\alpha > 0$.

Lemma 4.28. *Let $\lambda(t)$ be an integral curve of the Hamiltonian vector field $\vec{H}$ and $\gamma(t) = \pi(\lambda(t))$ be the corresponding normal extremal trajectory. Then for all $t \in [0, T]$ one has*

$$\frac{1}{2}\|\dot{\gamma}(t)\|^2 = H(\lambda(t)).$$

Proof. Fix a generating frame $f_1, \ldots, f_m$. Since $\lambda(t)$ is a solution of the Hamiltonian system we have

$$\dot{\gamma}(t) = \sum_{i=1}^{m} \langle \lambda(t), f_i(\gamma(t)) \rangle f_i(\gamma(t)) \tag{4.35}$$

hence $\overline{u}_i(t) = \langle \lambda(t), f_i(\gamma(t) \rangle$ defines a control for the curve γ. This control is indeed the minimal one as it follows from Exercice 4.25 and

$$\frac{1}{2}\|\dot{\gamma}(t)\|^2 = \frac{1}{2}\sum_{i=1}^{m} \overline{u}_i(t)^2 = \frac{1}{2}\sum_{i=1}^{m} \langle \lambda(t), f_i(\gamma(t)) \rangle^2 = H(\lambda(t)) \tag{4.36}$$

□

Remark 4.29. Notice that from (4.35) 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

$$\dot{\gamma}(0) = \sum_{i=1}^{m} \langle \lambda_0, f_i(q_0) \rangle f_i(q_0). \tag{4.37}$$

Corollary 4.30. *A normal extremal trajectory is parametrized by constant speed. In particular it is 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 (4.36) 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 4.27, all normal extremal trajectories are reparametrization of length parametrized ones. □

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 4.30, 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 4.31. Let M be a smooth manifold and $\Omega \in \Lambda^k M$ a 2-form. A Lipschitz curve $\gamma : [0, T] \to M$ is a *characteristic curve* for Ω if for almost every $t \in [0, T]$ it holds

$$\dot{\gamma}(t) \in \ker \Omega_{\gamma(t)}, \quad (\text{i.e. } \Omega_{\gamma(t)}(\dot{\gamma}(t), \cdot) = 0) \tag{4.38}$$

Notice that this notion is independent on the parametrization of the curve.

Proposition 4.32. *Let H be the sub-Riemannian Hamiltonian and assume that $c > 0$ is a regular value of H. Then a Lipschitz curve γ is a characteristic curve for $\sigma|_{H^{-1}(c)}$ if and only if it is the reparametrization of a normal extremal on $H^{-1}(c)$.*

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 by Sard Theorem almost every $c > 0$ is regular value for H).

For every $\lambda \in H^{-1}(c)$ let us denote by $E_\lambda = T_\lambda H^{-1}(c)$ its tangent space at this point. Notice that, by construction, E_λ is an hyperplane (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 $\lambda \in H^{-1}(c)$.

One one hand $\ker \sigma_\lambda|_{E_\lambda}$ is non trivial since the dimension of E_λ is odd. On the other hand the symplectic 2-form σ is nondegenerate on T^*M, hence the dimension of $\ker \sigma_\lambda|_{E_\lambda}$ cannot be greater than one. It follows that $\dim \ker \sigma_\lambda|_{E_\lambda} = 1$.

We are left to show that $\ker \sigma_\lambda|_{E_\lambda} = \vec{H}(\lambda)$. Assume that $\ker \sigma_\lambda|_{E_\lambda} = \mathbb{R}\xi$, for some $\xi \in T_\lambda(T^*M)$. By construction, E_λ coincides with the skew-orthogonal to ξ, namely

$$E_\lambda = \xi^\angle = \{w \in T_\lambda(T^*M)) \,|\, \sigma_\lambda(\xi, w) = 0\}.$$

Since, by skew-symmetry, $\sigma_\lambda(\xi, \xi) = 0$, it follows that $\xi \in E_\lambda$. Moreover, by definition of Hamiltonian vector field $\sigma(\cdot, \vec{H}) = dH$, hence for the restriction to E_λ one has

$$\sigma_\lambda(\cdot, \vec{H}(\lambda))\big|_{E_\lambda} = d_\lambda H\big|_{E_\lambda} = 0.$$

□

Exercise 4.33. Prove that if 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 $c_1, c_2 \in \mathbb{R}$, then their Hamiltonian flow $\vec{h}_1, \vec{h}_2$ coincide on E, up to reparametrization.

Exercise 4.34. The sub-Riemannian Hamiltonian H encodes all the information about the sub-Riemannian structure.

(a) Prove that a vector $v \in T_qM$ is *sub-unit*, i.e., it satisfies $v \in \mathcal{D}_q$ and $\|v\| \leq 1$ if and only if

$$\frac{1}{2}|\langle \lambda, v\rangle|^2 \leq H(\lambda), \qquad \forall \lambda \in T_q^*M.$$

(b) Show that this implies the following characterization for the sub-Riemannian Hamiltonian

$$H(\lambda) = \frac{1}{2}\|\lambda\|^2, \qquad \|\lambda\| = \sup_{v \in \mathcal{D}_q, |v|=1} |\langle \lambda, v\rangle|.$$

When the structure is Riemannian, H is the "inverse" norm defined on the cotangent space.

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, we show that they can be characterized as characteristic curves of the symplectic form. This gives an unified point of view of both class of extremals.

We recall that an abnormal extremal is a non zero 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 linear on fibers. In particular every abnormal extremal is contained in the set

$$H^{-1}(0) = \{\lambda \in T^*M \mid \langle \lambda, f_i(q)\rangle = 0, \, i = 1, \dots, m, \, q = \pi(\lambda)\}. \tag{4.39}$$

where H denotes the sub-Riemannian Hamiltonian (4.32).

Proposition 4.35. *Let H be the sub-Riemannian Hamiltonian and assume that $H^{-1}(0)$ is a smooth manifold. Then a Lipschitz curve γ is a characteristic curve for $\sigma|_{H^{-1}(0)}$ if and only if it is the reparametrization of a abnormal extremal on $H^{-1}(0)$.*

Proof. In this proof we denote for simplicity $N := H^{-1}(0) \subset T^*M$. For every $\lambda \in N$ we have the identity

$$\ker \sigma_\lambda|_N = T_\lambda N^{\angle} = \operatorname{span}\{\vec{h}_i(\lambda) \,|\, i = 1, \dots, m\}. \tag{4.40}$$

Indeed, from the definition of N, it follows that

$$\begin{aligned} T_\lambda N &= \{w \in T_\lambda(T^*M) \mid \langle d_\lambda h_i, \, w\rangle = 0, \, i = 1, \dots, m\} \\ &= \{w \in T_\lambda(T^*M) \,|\, \sigma(w, \vec{h}_i(\lambda)) = 0, \, i = 1, \dots, m\} \\ &= \operatorname{span}\{\vec{h}_i(\lambda) \,|\, i = 1, \dots, m\}^{\angle}. \end{aligned}$$

and (4.40) follows by taking the skew-orthogonal on both sides. Thus $w \in T_\lambda H^{-1}(0)$ 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 there exists controls $u_i(\cdot)$ for $i = 1, \dots, m$ such that

$$\dot{\lambda}(t) = \sum_{i=1}^{m} u_i(t)\vec{h}_i(\lambda(t)). \qquad \square \tag{4.41}$$

Notice that 0 is never a regular value of H. Nevertheless, the following exercise shows that the assumption of Proposition 4.35 is always satisfied in the case of a regular sub-Riemannian structure.

Exercise 4.36. Assume that the sub-Riemannian structure is *regular*, namely the following assumption holds

$$\dim \mathcal{D}_q = \dim \operatorname{span}_q\{f_1, \dots, f_m\} = \text{const}. \tag{4.42}$$

Then prove that the set $H^{-1}(0)$ defined by (4.39) is a smooth submanifold of T^*M.

Remark 4.37. From Proposition 4.35 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 (4.42) we can select (at least locally) a basis of 1-forms $\omega_1, \dots, \omega_m$ for the dual of the distribution

$$\mathcal{D}_q^\perp = \operatorname{span}\{\omega_i(q) \,|\, i = 1, \dots, m\}, \tag{4.43}$$

Let us complete this set of 1-forms to a basis $\omega_1, \dots, \omega_n$ of T^*M and consider the induced coordinates $h_1, \dots, h_n$ as defined in Remark 4.16. In these coordinates the restriction of the symplectic structure $\mathcal{D}^\perp$ to is expressed as follows

$$\sigma|_{\mathcal{D}^\perp} = d(s|_{\mathcal{D}^\perp}) = \sum_{i=1}^{m} dh_i \wedge \omega_i + h_i d\omega_i, \tag{4.44}$$

We stress that the restriction $\sigma|_{\mathcal{D}^\perp}$ can be written only in terms of the elements $\omega_1, \dots, \omega_m$ (and not of a full basis of 1-forms) since the differential d commutes with the restriction.

4.3.3 Example: codimension one distribution 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 dimension, as smooth manifolds, are

$$\begin{aligned} \dim \mathcal{D} &= \dim M + \operatorname{rank} \mathcal{D} = 2n - 1, \\ \dim\, \mathcal{D}^\perp &= \dim M + \operatorname{rank} \mathcal{D}^\perp = n + 1. \end{aligned}$$

Since the symplectic form σ 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 4.35.

Let us consider in more detail the case $n = 3$. 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). Consider a basis of one forms $\omega_0, \omega_1, \omega_2$ such that $\omega_0 := \omega$ and the coordinates h_0, h_1, h_2 associated to these forms (see Remark 4.16). By (4.44)

$$\sigma|_{\mathcal{D}^\perp} = dh_0 \wedge \omega + h_0\, d\omega, \tag{4.45}$$

and we can easily compute (recall that $\mathcal{D}^\perp$ is 4-dimensional)

$$\sigma \wedge \sigma|_{\mathcal{D}^\perp} = 2h_0\, dh_0 \wedge \omega \wedge d\omega. \tag{4.46}$$

Lemma 4.38. *Let N be a smooth $2k$-dimensional manifold and $\Omega \in \Lambda^2 M$. Then Ω is nondegenerate on N if and only if $\wedge^k \Omega \neq 0$.*[1]

Definition 4.39. Let M be a three dimensional manifold. We say that a constant rank distribution $\mathcal{D} = \ker \omega$ on M of corank one is a *contact distribution* if $\omega \wedge d\omega \neq 0$.

For a three dimensional manifold M endowed with a distribution $\mathcal{D} = \ker \omega$ we define the *Martinet set* as

$$\mathfrak{M} = \{q \in M |\, (\omega \wedge d\omega)|_q = 0\} \subset M.$$

Corollary 4.40. *Under the previous assumptions all nontrivial abnormal extremal trajectories are contained in the Martinet set $\mathfrak{M}$. In particular, if the structure is contact, there are no nontrivial abnormal extremal trajectories.*

Proof. By Proposition 4.35 any abnormal extremal $\lambda(t)$ is a characteristic curve of $\sigma|_{\mathcal{D}^\perp}$. By Lemma 4.38 $\sigma|_{\mathcal{D}^\perp}$ is degenerate if and only if $\sigma \wedge \sigma|_{\mathcal{D}^\perp} = 0$, which is in turn equivalent to $\omega \wedge d\omega = 0$ thanks to (4.46) (notice that dh_0 and $\omega \wedge d\omega$ are independent since they depend on coordinates on the fibers and on the manifold, respectively).

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 is $\gamma(t) \in \mathfrak{M}$. By definition of $\mathfrak{M}$ it follows that, if $\mathcal{D}$ is contact, then $\mathfrak{M}$ is empty. $\square$

Remark 4.41. Since M is three dimensional, we can write $\omega \wedge d\omega = a dV$ where $a \in C^\infty(M)$ and dV is some smooth volume form on M, i.e., a never vanishing 3-form on M.

In particular 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 set $a^{-1}(0)$ defines a two dimensional surface on M, called the *Martinet surface*. Notice that this condition is satisfied for a generic choice of the (one form defining the) distribution.

Abnormal extremal trajectories are the horizontal curves that are contained in the Martinet surface. When $\mathfrak{M}$ is smooth, 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 with the integral curves of this line field, up to a reparametrization.

[1] Here $\wedge^k \Omega = \underbrace{\Omega \wedge \ldots \wedge \Omega}_{k}$.

4.4 Examples

4.4.1 2D Riemannian Geometry

Let M be a 2-dimensional manifold and $f_1, f_2 \in \mathrm{Vec}(M)$ a local orthonormal frame for the Riemannian structure. The problem of finding length-minimizers on M could be described 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

$$\ell(q(\cdot)) = \int_0^T \sqrt{u_1(t)^2 + u_2(t)^2}\, dt, \qquad J(q(\cdot)) = \frac{1}{2}\int_0^T \left(u_1(t)^2 + u_2(t)^2\right) dt.$$

Geodesics 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 (even if coordinates on the manifold are not necessarily fixed).

Let us start by showing that there are no abnormal extremals. Indeed if $\lambda(t)$ is an abnormal extremal and $\gamma(t)$ is the associated abnormal trajectory we have

$$\langle \lambda(t), f_1(\gamma(t)) \rangle = \langle \lambda(t), f_2(\gamma(t)) \rangle = 0, \qquad \forall\, t \in [0, T], \tag{4.47}$$

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 3.53.

Suppose now that $\lambda(t)$ is a normal extremal. Then $u_i(t) = h_i(\lambda(t))$ and the equation on the base is

$$\dot{q} = h_1 f_1(q) + h_2 f_2(q). \tag{4.48}$$

For the equation on the fiber we have (remember that along solutions $\dot{a} = \{H, a\}$)

$$\begin{cases} \dot{h}_1 = \{H, h_1\} = -\{h_1, h_2\} h_2 \\ \dot{h}_2 = \{H, h_2\} = \{h_1, h_2\} h_1. \end{cases} \tag{4.49}$$

From here one can see directly that H is constant along solutions. Indeed

$$\dot{H} = h_1 \dot{h}_1 + h_2 \dot{h}_2 = 0.$$

If we require that extremals are parametrized by arclength $u_1(t)^2 + u_2(t)^2 = 1$ for a.e. $t \in [0, T]$, we have

$$H(\lambda(t)) = \frac{1}{2} \quad \iff \quad h_1^2(\lambda(t)) + h_2^2(\lambda(t)) = 1.$$

It is then convenient to restrict to the spherical cotangent bundle S^*M (see Example 2.51) of coordinates (q, θ), by setting

$$h_1 = \cos\theta, \qquad h_2 = \sin\theta.$$

Let $a_1, a_2 \in C^\infty(M)$ be such that

$$[f_1, f_2] = a_1 f_1 + a_2 f_2. \tag{4.50}$$

Since $\{h_1, h_2\}(\lambda) = \langle \lambda, [f_1, f_2] \rangle$, we have $\{h_1, h_2\} = a_1 h_1 + a_2 h_2$ and equations (7.28) and (4.57) are rewritten in (θ, q) coordinates

$$\begin{cases} \dot\theta = a_1(q)\cos\theta + a_2(q)\sin\theta \\ \dot q = \cos\theta f_1(q) + \sin\theta f_2(q) \end{cases} \tag{4.51}$$

In other words we are saying that an arc-length parametrized curve on M (i.e. a curve which satisfies the second equation) is a geodesic if and only if it satisfies the first. Heuristically this suggests that the quantity

$$\dot\theta - a_1(q)\cos\theta - a_2(q)\sin\theta,$$

has some relation with the geodesic curvature on M.

Let μ_1, μ_2 the dual frame of f_1, f_2 (so that $dV = \mu_1 \wedge \mu_2$) and consider the Hamiltonian field in these coordinates

$$\vec H = \cos\theta f_1 + \sin\theta f_2 + (a_1\cos\theta + a_2\sin\theta)\partial_\theta. \tag{4.52}$$

The Levi-Civita connection on M is expressed by some coefficients (see Chapter **??**)

$$\omega = d\theta + b_1\mu_1 + b_2\mu_2,$$

where $b_i = b_i(q)$. On the other hand geodesics are projections of integral curves of $\vec H$ so that

$$\langle \omega, \vec H \rangle = 0 \implies b_1 = -a_1, \quad b_2 = -a_2.$$

In particular if we apply $\omega = d\theta - a_1\mu_1 - a_2\mu_2$ to a generic curve (not necessarily a geodesic)

$$\lambda = \cos\theta f_1 + \sin\theta f_2 + \dot\theta\,\partial_\theta,$$

which projects on γ we find geodesic curvature

$$\kappa_g(\gamma) = \dot\theta - a_1(q)\cos\theta - a_2(q)\sin\theta,$$

as we infer above. To end this section we prove a useful formula for the Gaussian curvature of M

Corollary 4.42. *If κ denotes the Gaussian curvature of M we have*

$$\kappa = f_1(a_2) - f_2(a_1) - a_1^2 - a_2^2.$$

Proof. From (1.58) we have $d\omega = -\kappa dV$ where $dV = \mu_1 \wedge \mu_2$ is the Riemannian volume form. On the other hand, using the following identities

$$d\mu_i = -a_i\mu_1 \wedge \mu_2, \qquad da_i = f_1(a_i)\mu_1 + f_2(a_i)\mu_2, \quad i = 1, 2.$$

we can compute

$$\begin{aligned} d\omega &= -da_1 \wedge \mu_1 - da_2 \wedge \mu_2 - a_1 d\mu_1 - a_2 d\mu_2 \\ &= -(f_1(a_2) - f_2(a_1) - a_1^2 - a_2^2)\mu_1 \wedge \mu_2. \end{aligned}$$

□

4.4.2 Isoperimetric problem

Let M be a 2-dimensional orientable Riemannian manifold and ν 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

$$\min \left\{ \ell(\gamma) : \gamma(0) = q_0, \gamma(T) = q_1, \int_\gamma A = c \right\}. \tag{4.53}$$

Remark 4.43. 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 (4.53) turns to be length minimizers for a suitable sub-Riemannian structure on $\overline{M}$, that we are going to construct.

To every curve γ 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 $\xi(t) = (\gamma(t), z(t))$ defined on $\overline{M}$ satisfies $\omega(\dot\xi(t)) = 0$ where $\omega = dz - A$ is a one form on $\overline{M}$, since

$$\omega(\dot\xi(t)) = \dot z(t) - A(\dot\gamma(t)) = 0.$$

Equivalently, $\dot\xi(t) \in \mathcal{D}_{\xi(t)}$ where $\mathcal{D} = \ker \omega$. We define a metric on $\mathcal{D}$ by defining the norm of a vector $v \in \mathcal{D}$ as the Riemannian norm of its projection $\bar\pi_* v$ on M, where $\bar\pi : \overline{M} \to M$ is the canonical 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

$$\begin{pmatrix} \dot\gamma \\ \dot z \end{pmatrix} = u_1 \begin{pmatrix} f_1 \\ \langle A, f_1 \rangle \end{pmatrix} + u_2 \begin{pmatrix} f_2 \\ \langle A, f_2 \rangle \end{pmatrix}. \tag{4.54}$$

Hence the two vector fields on $\overline{M}$

$$F_1 = f_1 + \langle A, f_1 \rangle \partial_z, \qquad F_2 = f_2 + \langle A, f_2 \rangle \partial_z,$$

defines an orthonormal frame for the metric defined above on $\mathcal{D} = \operatorname{span}(F_1, F_2)$. 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

$$\min \left\{ \ell(\xi) : \xi(0) = (q_0, 0), \xi(T) = (q_1, c), \dot\xi(t) \in \mathcal{D}_{\xi(t)} \right\}. \tag{4.55}$$

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 extremals associated to the sub-Riemannian structure just introduced on $\overline{M}$. In what follows we denote with $h_i(\lambda) = \langle \lambda, F_i(q) \rangle$ the Hamiltonians linear on fibers of $T^*\overline{M}$.

Normal extremals

Equations of normal extremals are projections of integral curves of the sub-Riemannian Hamiltonian in $T^*\overline{M}$

$$H(\lambda) = \frac{1}{2}(h_1^2(\lambda) + h_2^2(\lambda)), \qquad h_i(\lambda) = \langle \lambda, f_i(q) \rangle, \quad i = 1, 2.$$

Let us introduce $F_0 = \partial_z$ and $h_0(\lambda) = \langle \lambda, F_0(q) \rangle$. 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 $T^*\overline{M}$. In what follows it is convenient to use (q, h_1, h_2, h_0) as coordinates on T^*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

$$\dot{\xi} = h_1 F_1(\xi) + h_2 F_2(\xi). \tag{4.56}$$

For the equation on the fibers we have (remember that along solutions $\dot{a} = \{H, a\}$)

$$\begin{cases} \dot{h}_1 = \{H, h_1\} = -\{h_1, h_2\} h_2 \\ \dot{h}_2 = \{H, h_2\} = \{h_1, h_2\} h_1. \\ \dot{h}_0 = \{H, h_0\} = 0 \end{cases} \tag{4.57}$$

If we require that extremals are parametrized by arclength we can restrict to the cylinder of the cotangent bundle T^*M defined by

$$h_1 = \cos\theta, \qquad h_2 = \sin\theta.$$

Let $a_1, a_2 \in C^\infty(M)$ be such that

$$[f_1, f_2] = a_1 f_1 + a_2 f_2. \tag{4.58}$$

Then

$$\begin{aligned} [F_1, F_2] &= [f_1 + \langle A, f_1 \rangle \partial_z, f_2 + \langle A, f_2 \rangle \partial_z] \\ &= [f_1, f_2] + (f_1 \langle A, f_2 \rangle - f_2 \langle A, f_1 \rangle)\partial_z \\ \text{(by (4.58))} \qquad &= a_1(F_1 - \langle A, f_1 \rangle) + a_2(F_2 - \langle A, f_2 \rangle) + f_1 \langle A, f_2 \rangle - f_2 \langle A, f_1 \rangle)\partial_z \\ &= a_1 F_1 + a_2 F_2 + dA(f_1, f_2)\partial_z. \end{aligned}$$

where in the last equality we use Cartan formula (cf. (4.77) for a proof). Let μ_1, μ_2 be the dual forms to f_1 and f_2. Then $\nu = \mu_1 \wedge \mu_2$ and we can write $dA = b\mu_1 \wedge \mu_2$, for a suitable function $b \in C^\infty(M)$. In this case

$$[F_1, F_2] = a_1 F_1 + a_2 F_2 + b\partial_z.$$

and

$$\{h_1, h_2\} = \langle \lambda, [F_1, F_2] \rangle = a_1 h_1 + a_2 h_2 + b h_0. \tag{4.59}$$

With computations analogous to the 2D case we obtain the Hamiltonian system associated to H in the (q, θ, h_0) coordinates

$$\begin{cases} \dot{\xi} = \cos\theta F_1(\xi) + \sin\theta F_2(\xi) \\ \dot{\theta} = a_1 \cos\theta + a_2 \sin\theta + b h_0 \\ \dot{h}_0 = 0 \end{cases} \tag{4.60}$$

In other words if $q(t) = \bar{\pi}(\xi(t))$ is the projection of a normal extremal path on M (here $\bar{\pi} : \overline{M} \to M$), its geodesic curvature

$$\kappa_g(q(t)) = \dot{\theta}(t) - a_1(q(t)) \cos \theta(t) - a_2(q(t)) \sin \theta(t) \tag{4.61}$$

satisfies

$$\kappa_g(q(t)) = b(q(t))h_0. \tag{4.62}$$

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 extremals

We prove the following characterization of abnormal extremal

Lemma 4.44. *Abnormal extremal trajectories are contained in the Martinet set $\mathcal{M} = \{b = 0\}$.*

Proof. Assume that $\lambda(t)$ is an abnormal extremal whose projection is a curve $\xi(t) = \pi(\lambda(t))$ that is not reduced to a point. Then we have

$$h_1(\lambda(t)) = \langle \lambda(t), F_1(\xi(t)) \rangle = 0, \qquad h_2(\lambda(t)) = \langle \lambda(t), F_2(\xi(t)) \rangle = 0, \qquad \forall\, t \in [0, T], \tag{4.63}$$

We can differentiate the two equalities with respect to $t \in [0, T]$ and we get

$$\frac{d}{dt} h_1(\lambda(t)) = u_2(t) \{h_1, h_2\}|_{\lambda(t)} = 0$$
$$\frac{d}{dt} h_2(\lambda(t)) = -u_1(t) \{h_1, h_2\}|_{\lambda(t)} = 0$$

Since the pair $(u_1(t), u_2(t)) \neq (0, 0)$ we have that $\{h_1, h_2\}|_{\lambda(t)} = 0$ that implies

$$0 = \langle \lambda(t), [F_1, F_2](\xi(t)) \rangle = b(\xi(t))h_0, \tag{4.64}$$

where in the last equality we used (4.59) 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 $b(\xi(t)) = 0$ for all $t \in [0, T]$. $\square$

The last result shows that abnormal extremal trajectories are forced to live in connected components of $b^{-1}(0)$.

Exercise 4.45. 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* 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 ω 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 induced 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 (4.60) for normal extremals become

$$\begin{cases} \dot{q} = \cos\theta F_1(q) + \sin\theta F_2(q) \\ \dot{\theta} = h_0 \\ \dot{h}_0 = 0 \end{cases} \tag{4.65}$$

It follows that the two last equation can be immediately solved

$$\begin{cases} \theta(t) = \theta_0 + h_0 t \\ h_0(t) = h_0 \end{cases} \tag{4.66}$$

Moreover

$$\begin{cases} h_1(t) = \cos(\theta_0 + h_0 t) \\ h_2(t) = \sin(\theta_0 + h_0 t) \end{cases} \tag{4.67}$$

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 $h_0 \neq 0$

$$x(t) = \frac{1}{h_0}(\sin(\theta_0 + h_0 t) - \sin(\theta_0)) \qquad y(t) = \frac{1}{h_0}(\cos(\theta_0 + h_0 t) - \cos(\theta_0)) \tag{4.68}$$

and the vertical coordinate z is computed as the integral

$$z(t) = \frac{1}{2}\int_0^t x(t)y'(t) - y(t)x'(t)dt = \frac{1}{2h_0^2}(h_0 t - \sin(h_0 t))$$

When $h_0 = 0$ the curve is simply a straight line

$$x(t) = \sin(\theta_0)t \qquad y(t) = \cos(\theta_0)t \qquad z(t) = 0 \tag{4.69}$$

Notice that, as we know from the results of the previous paragraph, 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 $\overline{M}$). There are no non trivial abnormal geodesics since $b = 1$.

Remark 4.46. 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, as explained in 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 we define

$$\mathcal{L}_X Y = [X, Y] = \left.\frac{d}{dt}\right|_{t=0} e_*^{-tX} Y.$$

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. Let $\omega \in \Lambda^k M$, we define $P^*\omega \in \Lambda^k M$ in the following way:

$$(P^*\omega)_q(\xi_1, \ldots, \xi_k) := \omega_{P(q)}(P_*\xi_1, \ldots, P_*\xi_k), \quad q \in M, \quad \xi_i \in T_q M. \tag{4.70}$$

It is an easy check that this operation is linear and satisfies the two following properties

$$P^*(\omega_1 \wedge \omega_2) = P^*\omega_1 \wedge P^*\omega_2, \tag{4.71}$$

$$P^* \circ d = d \circ P^*. \tag{4.72}$$

Definition 4.47. Let $X \in \mathrm{Vec}(M)$ and $\omega \in \Lambda^k M$, where $k \geq 0$. We define the *Lie derivative* of ω with respect to X as

$$\mathcal{L}_X : \Lambda^k M \to \Lambda^k M, \qquad \mathcal{L}_X \omega = \left.\frac{d}{dt}\right|_{t=0} (e^{tX})^* \omega. \tag{4.73}$$

When $k = 0$ this definition recovers the Lie derivative of smooth functions $\mathcal{L}_X f = Xf$, for $f \in C^\infty(M)$. From (4.71) and (4.72), we easily deduce 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) $\mathcal{L}_X \circ d = d \circ \mathcal{L}_X$.

The first of these properties can be also expressed by saying that $\mathcal{L}_X$ is a *derivation* of the exterior algebra of k-forms.

The Lie derivative combines together a k-form and a vector field defining a new k-form. A second way of combining these two object is to define their inner product, by defining a $(k-1)$-form.

Definition 4.48. Let $X \in \mathrm{Vec}(M)$ and $\omega \in \Lambda^k M$, with $k \geq 1$. We define the *inner product of ω and X* as the operator $i_X : \Lambda^k M \to \Lambda^{k-1} M$, where we set

$$(i_X \omega)(Y_1, \ldots, Y_{k-1}) := \omega(X, Y_1, \ldots, Y_{k-1}), \quad Y_i \in \mathrm{Vec}(M). \tag{4.74}$$

One can show that the operator i_X is an *anti-derivation*, in the following sense:

$$i_X(\omega_1 \wedge \omega_2) = (i_X \omega_1) \wedge \omega_2 + (-1)^{k_1} \omega_1 \wedge (i_X \omega_2), \quad \omega_i \in \Lambda^{k_i} M, \quad i = 1, 2. \tag{4.75}$$

We end this section proving two classical formulas linking together these notions, and usually referred as Cartan's formulas.

Proposition 4.49 (Cartan's formula)**.** *The following identity holds true*

$$\mathcal{L}_X = i_X \circ d + d \circ i_X. \tag{4.76}$$

Proof. Define $D_X := i_X \circ d + d \circ i_X$. 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 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 checked by

$$D_X f = i_X(df) + \underbrace{d(i_X f)}_{=0} = \langle df, X \rangle = Xf = \mathcal{L}_X f. \qquad \square$$

Corollary 4.50. *Let $X, Y \in \mathrm{Vec}(M)$ and $\omega \in \Lambda^1 M$, then*

$$d\omega(X, Y) = X \langle \omega, Y \rangle - Y \langle \omega, X \rangle - \langle \omega, [X, Y] \rangle . \tag{4.77}$$

Proof. On one hand Definition 4.47 implies, by Leibnitz rule

$$\begin{aligned} \langle \mathcal{L}_X \omega, Y \rangle_q &= \frac{d}{dt}\bigg|_{t=0} \left\langle (e^{tX})^* \omega, Y \right\rangle_q \\ &= \frac{d}{dt}\bigg|_{t=0} \left\langle \omega, e_*^{tX} Y \right\rangle_{e^{tX}(q)} \\ &= X \langle \omega, Y \rangle - \langle \omega, [X, Y] \rangle . \end{aligned}$$

On the other hand, Cartan's formula (4.76) 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 (4.77). $\square$

4.6 Symplectic geometry

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, σ) 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 a symplectic manifold N is necessarily even-dimensional. We stress that, in general, the symplectic form σ is not exact, as in the case of $N = T^*M$.

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 $h \in C^\infty(N)$:*

$$(\phi_*^{-1}) \vec{h} = \overrightarrow{h \circ \phi}. \tag{4.78}$$

Proof. Assume that ϕ 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 (4.78). The converse implication follows analogously. $\square$

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, σ). 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 σ is, by definition, a closed form. Then $d\sigma = 0$ and Cartan's formula (4.76) 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 defines a flow of symplectomorphisms.*

Proof. This is a direct consequence of the fact that, for an Hamitonian vector field $\vec{h}$, one has $i_{\vec{h}}\sigma = -dh$. Hence $i_{\vec{h}}\sigma$ is a cloded 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, σ) 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 (4.78). 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 $t = 0$. ☐

Corollary 4.57. *For every $a, b \in C^\infty(N)$ we have*

$$\overrightarrow{\{a, b\}} = [\vec{a}, \vec{b}]. \tag{4.79}$$

Proof. Property (ii) of Proposition 4.56 can be rewritten, by skew-symmetry of the Poisson bracket, as follows

$$\{\{a, b\}, c\} = \{a, \{b, c\}\} - \{b, \{a, c\}\}. \tag{4.80}$$

Using that $\{a, b\} = \sigma(\vec{a}, \vec{b}) = \vec{a}b$ one rewrite (4.80) as

$$\overrightarrow{\{a, b\}}c = \vec{a}(\vec{b}c) - \vec{b}(\vec{a}c) = [\vec{a}, \vec{b}]c.$$

☐

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 (4.79), 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 trajectories

In this section we prove a fundamental result about local optimality of normal trajectories. More precisely we show 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

$$\mathcal{L}_0 = \{d_q a \,|\, q \in M\} \subset T^*M. \tag{4.81}$$

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

$$\mathcal{L}_t := e^{t\vec{H}}(\mathcal{L}_0), \qquad t \in [0, T]. \tag{4.82}$$

Define the $(n+1)$-dimensional manifold with boundary in $\mathbb{R} \times T^*M$ as follows

$$\begin{aligned} \mathcal{L} &= \{(t, \lambda) \in \mathbb{R} \times T^*M \,|\, \lambda \in \mathcal{L}_t,\, 0 \le t \le T\} & (4.83)\\ &= \{(t, e^{t\vec{H}}\lambda_0) \in \mathbb{R} \times T^*M \,|\, \lambda_0 \in \mathcal{L}_0,\, 0 \le t \le T\}. & (4.84) \end{aligned}$$

Finally, let us introduce the *Poincaré-Cartan* 1-form on $T^*M \times \mathbb{R} \simeq T^*(M \times \mathbb{R})$ 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

$$d(s - Hdt) = \sigma - dH \wedge dt, \tag{4.85}$$

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 Ψ with respect to t:

$$\frac{\partial \Psi}{\partial t}(t, \lambda) = \frac{\partial}{\partial t} + \vec{H}(\lambda) \in T_{(t,\lambda)}\mathcal{L}. \tag{4.86}$$

It is then sufficient to show that the vector (4.86) is in the kernel of the two form $\sigma - dH \wedge dt$. In other words we have to prove

$$i_{\partial_t + \vec{H}}(\sigma - dH \wedge dt) = 0. \tag{4.87}$$

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 Γ in $\mathcal{L}$ one has

$$\int_{\Gamma} s - Hdt = 0. \tag{4.88}$$

Every curve Γ 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

$$K : [0, T] \times \mathcal{L} \to \mathcal{L}, \qquad K(\tau, (t, e^{t\vec{H}}\lambda_0)) = (t - \tau, e^{(t-\tau)\vec{H}}\lambda_0)$$

defines an homotopy of $\mathcal{L}$ such that $K(0, (t, e^{t\vec{H}}\lambda_0)) = (t, e^{t\vec{H}}\lambda_0)$ and $K(t, (t, e^{t\vec{H}}\lambda_0)) = (0, \lambda_0)$. Then the curve Γ 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 Γ_0 is computed as follows (recall that $\Gamma_0 \subset \mathcal{L}_0$ and $dt = 0$ on $\mathcal{L}_0$):

$$\int_{\Gamma_0} s - Hdt = \int_{\Gamma_0} s = \int_{\Gamma_0} d(a \circ \pi) = 0,$$

where we used Lemma 4.59 and the fact that the integral of an exact form over a closed curve is zero. Then (4.88) follows. $\square$

4.7.2 Normal trajectories are geodesics

Now we are ready to prove a sufficient condition that ensures the optimality of small pieces of normal trajectories. As a corollary we will get that small pieces of normal trajectories are geodesics.

Recall that normal trajectories for the problem

$$\dot{q} = f_u(q) = \sum_{i=1}^{m} u_i f_i(q), \tag{4.89}$$

where $f_1, \ldots, f_m$ is a generating family for the sub-Riemannian structure are projections of integral curves of the Hamiltonian vector fields associated with the sub-Riemannian Hamiltonian

$$\dot{\lambda}(t) = \vec{H}(\lambda(t)), \qquad (\text{i.e. } \lambda(t) = e^{t\vec{H}}(\lambda_0)), \tag{4.90}$$

$$\gamma(t) = \pi(\lambda(t)), \qquad t \in [0, T]. \tag{4.91}$$

where

$$H(\lambda) = \max_{u \in U_q} \left\{ \langle \lambda, f_u(q) \rangle - \frac{1}{2}|u|^2 \right\} = \frac{1}{2} \sum_{i=1}^{m} \langle \lambda, f_i(q) \rangle^2 . \tag{4.92}$$

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 to H as is (4.81) and (4.82).

Theorem 4.61. *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 $t \in [0,T]$. Then for any $\lambda_0 \in \mathcal{L}_0$ the normal geodesic*

$$\overline{\gamma}(t) = \pi \circ e^{t\vec{H}}(\lambda_0), \qquad t \in [0,T], \tag{4.93}$$

is a strict length-minimizer among all admissible curves γ with the same boundary conditions.

Proof. Let $\gamma(t)$ be an admissible trajectory, different from $\overline{\gamma}(t)$, associated with the control $u(t)$ and such that $\gamma(0) = \overline{\gamma}(0)$ and $\gamma(T) = \overline{\gamma}(T)$. We denote by $\overline{u}(t)$ the control associated with the curve $\overline{\gamma}(t)$.

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 $\lambda(t) \in \mathcal{L}_t$. Notice that the corresponding curves Γ and $\overline{\Gamma}$ in $\mathcal{L}$ defined by

$$\Gamma(t) = (t, \lambda(t)), \qquad \overline{\Gamma}(t) = (t, \overline{\lambda}(t)) \tag{4.94}$$

have the same boundary conditions, since for $t = 0$ and $t = T$ they project to the same base point on M and their lift is uniquely determined by the diffeomorphisms $\pi|_{\mathcal{L}_0}$ and $\pi|_{\mathcal{L}_T}$, respectively.

Recall now that, by definition of the sub-Riemannian Hamiltonian, we have

$$H(\lambda(t)) \geq \left\langle \lambda(t), f_{u(t)}(\gamma(t)) \right\rangle - \frac{1}{2}|u(t)|^2, \qquad \gamma(t) = \pi(\lambda(t)), \tag{4.95}$$

where $\lambda(t)$ is a lift of the trajectory $\gamma(t)$ associated with a control $u(t)$. Moreover, the equality holds in (4.95) if and only if $\lambda(t)$ is a solution of the Hamiltonian system $\dot{\lambda}(t) = \vec{H}(\lambda(t))$. For this reason we have the relations

$$H(\lambda(t)) > \left\langle \lambda(t), f_{u(t)}(\gamma(t)) \right\rangle - \frac{1}{2}|u(t)|^2, \tag{4.96}$$

$$H(\overline{\lambda}(t)) = \left\langle \overline{\lambda}(t), f_{\overline{u}(t)}(\overline{\gamma}(t)) \right\rangle - \frac{1}{2}|\overline{u}(t)|^2. \tag{4.97}$$

since $\overline{\lambda}(t)$ is a solution of the Hamiltonian equation by assumptions, while $\lambda(t)$ is not. Indeed $\lambda(t)$ and $\overline{\lambda}(t)$ have the same initial condition, hence, by uniqueness of the solution of the Cauchy problem, it follows that $\dot{\lambda}(t) = \vec{H}(\lambda(t))$ if and only if $\lambda(t) = \overline{\lambda}(t)$, that implies that $\overline{\gamma}(t) = \gamma(t)$.

Let us then show that the energy associated with the curve γ is bigger than the one of the curve $\overline{\gamma}$. Actually we prove the following chain of (in)equalities

$$\frac{1}{2}\int_0^T |\overline{u}(t)|^2 dt = \int_{\overline{\Gamma}} s - H dt = \int_{\Gamma} s - H dt < \frac{1}{2}\int_0^T |u(t)|^2 dt, \tag{4.98}$$

where Γ and $\overline{\Gamma}$ are the curves in $\mathcal{L}$ defined in (4.94).

By Lemma 4.60, the 1-form $s - Hdt$ is exact. Then the integral over the closed curve $\Gamma \cup \overline{\Gamma}$ vanishes, and one gets

$$\int_{\overline{\Gamma}} s - Hdt = \int_{\Gamma} s - Hdt.$$

The last inequality in (4.98) can be proved as follows

$$\begin{aligned}
\int_{\Gamma} s - Hdt &= \int_0^T \langle \lambda(t), \dot{\gamma}(t) \rangle - H(\lambda(t))dt \\
&= \int_0^T \left\langle \lambda(t), f_{u(t)}(\gamma(t)) \right\rangle - H(\lambda(t))dt \\
&< \int_0^T \left\langle \lambda(t), f_{u(t)}(\gamma(t)) \right\rangle - \left(\left\langle \lambda(t), f_{u(t)}(\gamma(t)) \right\rangle - \frac{1}{2}|u(t)|^2 \right) dt \qquad (4.99) \\
&= \frac{1}{2} \int_0^T |u(t)|^2 dt.
\end{aligned}$$

where we used (4.96). A similar computation, using (4.97), gives

$$\int_{\overline{\Gamma}} s - Hdt = \frac{1}{2} \int_0^T |\overline{u}(t)|^2 dt, \qquad (4.100)$$

that ends the proof of (4.98). □

As a corollary we state a local version of the same theorem, that can be proved by adapting the above technique.

Corollary 4.62. *Assume that there exists $a \in C^\infty(M)$ and neighborhoods Ω_t of $\overline{\gamma}(t)$, such that $\pi \circ e^{t\vec{H}} \circ da|_{\Omega_0} : \Omega_0 \to \Omega_t$ is a diffeomorphism for every $t \in [0,T]$. Then (4.93) is a strict length-minimizer among all admissible trajectories γ with 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 4.63. *Let $\gamma : [0,T] \to M$ be a sub-Riemannian normal trajectory. Then for every $\tau \in [0,T[$ there exists $\varepsilon > 0$ such that*

(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 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 (4.81) and (4.82). By construction, for the extremal lift associated with γ one has $\lambda(t) = e^{t\vec{H}}(\lambda_0) \in \mathcal{L}_t$ for all t. Moreover the projection $\pi\big|_{\mathcal{L}_0}$ is a diffeomorphism, since $\mathcal{L}_0$ is a section of T^*M.

Fix a compact $K \subset M$ containing the curve γ and consider the restriction $\pi_{t,K} : \mathcal{L}_t \cap \pi^{-1}(K) \to K$ of the map $\pi\big|_{\mathcal{L}_t}$. By continuity there exists $t_0 = t_0(K)$ such that $\pi_{t,K}$ is a diffeomorphism, for all $0 \leq t < t_0$. Let us now denote $\delta_K > 0$ the constant defined in Lemma 3.34 such that every curve starting from $\gamma(0)$ and leaving K is necessary longer than δ_K.

Then, defining $\varepsilon = \varepsilon(K) := \min\{\delta_K, t_0(K)\}$, we have that the curve $\gamma|_{[0,\varepsilon]}$ is contained in K and is shorter than any other curve contained in K with the same boundary condition by Corollary 4.62 (applied to $\Omega_t = K$ for all $t \in [0, T]$). Moreover $\ell(\gamma|_{[0,\varepsilon]}) = \varepsilon$ since γ is length parametrized, hence it is shorter than any admissible curve that is not contained in K. Thus $\gamma|_{[0,\varepsilon]}$ is a global minimizer. Moreover it is unique up to reparametrization by uniqueness of the solution of the Hamiltonian equation (see proof of Theorem 4.61). $\square$

Remark 4.64. 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 $\varepsilon = \varepsilon(\lambda_0)$, with λ_0 belonging to a compact set. 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 ε, see Theorem 12.17.

Bibliographical notes

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 and is inspired by [**?**].

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 [45] and references therein. The link between the two definition is clarified in Exercice 4.34

The proof that normal extremal are geodesics is an adaptation of a more general condition for optimality given in [8] for a more general class of problems. This is inspired by the classical idea of "fields of extremals" in classical Calculus of Variation.

Chapter 5

Integrable systems

In this chapter we present some applications of the Hamiltonian formalism developed in the previous chapter. 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 sub-Riemannian completely integrable systems, together with a proof that all left-invariant sub-Riemannian geodesic flows on 3D Lie groups are completely integrable, are presented in Chapter 13.

5.1 Reduction of Hamiltonian systems with symmetries

Recall that a symplectic manifold (N, σ) is a smooth manifold wendowed with a closed non-degenerate two-form σ (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 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

$$\frac{d}{dt} h \circ e^{t\vec{g}} = 0. \tag{5.1}$$

namely, h is preserved along the flow of $\vec{g}$.

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 $d_{x_0} g = \sigma(\cdot, \vec{g}(x_0))$ with $d_{x_0} g \neq 0$ and σ 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 to its 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 4.32, 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 is saying that the flow of $\vec{g}$ is well defined on the level set.

Consider then the quotient

$$N/_{\sim} := \{x \in U \cap N_{g,c} \mid x_1 \sim x_2 \text{ if } x_2 = e^{s\vec{g}}(x_1), s \in \mathbb{R}, \cup_{t\in[0,s]} e^{t\vec{g}}(x_1) \subset U\}.$$

In other words $N/_{\sim}$ is 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 (and not leaving U). Under our assumptions, the quotient has the structure of smooth manifold of dimension $2n-2$. To build a chart close to a point $[x_0] \in N/_{\sim}$ (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 $N/_{\sim}$.

The construction of the above quotient is classical (see for instance [9]). The restriction of the symplectic structure σ to the quotient $N/_{\sim}$ is necessarily non-degenerate (since σ is non-degenerate on the whole space N), hence gives to $N/_{\sim}$ the structure of symplectic space.

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 terms are 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 $N/_{\sim}$. 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 $N/_{\sim}$.

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 $N/_{\sim}$ with $C^\infty(N/_{\sim}) \simeq \{f \in C^\infty(N) \mid \{f, g\} = 0\}$.

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] $T^2 \simeq [0,1]^2/_{\equiv}$, 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

$$\vec{g}(x, y) = \frac{\partial g}{\partial p}\frac{\partial}{\partial x} - \frac{\partial g}{\partial x}\frac{\partial}{\partial p} = \frac{\partial}{\partial x} + \alpha\frac{\partial}{\partial p},$$

whose trajectories are given by

$$x(t) = x_0 + t, \qquad p(t) = p_0 + \alpha t.$$

It is well known that, 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.

[1] with the equivalence relation $(x, 0) \equiv (x, 1)$ and $(0, p) \equiv (1, p)$.

5.1.1 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 $N/_\sim$ is a smooth (symplectic) manifold of dimension $2n-2$ globally diffeomorphic to the space of 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 $N/_\sim$ of $N_{g,c}$ with respect to orbits of the Hamiltonian vector field $\vec{g}$ is the space of affine lines of $\mathbb{R}^n$ and is globally defined. The proof is completed by Proposition 5.5. $\square$

Proposition 5.5. *The set $A(n)$ of 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 $U_i^+ = S^{n-1} \cap \{x_i > 0\}$ an open subset 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 line $L \in W_i$ there exists a unique $\bar{x} \in H_i$ and $\bar{v} \in U_i^+$ such that $L = \{\bar{x} + t\bar{v} \mid t \in \mathbb{R}\}$. Then, for $i = 1, \ldots, n$, we define the coordinate chart

$$\phi_i : W_i \to H_i \times U_i^+, \qquad \phi_i(L) = (\bar{x}, \bar{v}).$$

Using the standard identification $H_i \simeq \mathbb{R}^{n-1}$ and the stereographic projection $W_i \simeq \mathbb{R}^{n-1}$, we build coordinate maps $\phi_i : W_i \to \mathbb{R}^{2n-2}$ for $i = 1, \ldots, n$. $\square$

Exercise 5.6. Check that $\{W_i\}_{i=1,\ldots,n}$ is an open cover of $A(n)$, and that the change of coordinates $\phi_i \circ \phi_j^{-1} : \mathbb{R}^{2n-2} \to \mathbb{R}^{2n-2}$ is smooth for every $i, j = 1, \ldots, n$.

5.2 Riemannian geodesic flow on hypersurfaces

In this section we want to show that the Riemannian geodesic flow on an hypersurface of $\mathbb{R}^n$, that is an Hamiltonian flow on a $2n-2$ dimension, can be seen as the restriction of the Hamiltonian flow of $\mathbb{R}^{2n}$ to the (reduced) symplectic space of affine lines in $\mathbb{R}^n$ (cf. Section 5.1.1).

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 1.4 in Chapter 1, one can prove the following characterization of geodesics on a hypersurface M_c.

Proposition 5.7. *Let $\gamma : [0,T] \to M$ be a smooth minimizer parametrized by length. Then $\ddot{\gamma}(t) \perp T_{\gamma(t)}M$.*

Exercise 5.8. Prove Proposition 5.7.

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 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

$$\langle \nabla_\xi a \,|\, p \rangle = \left.\frac{d}{dt}\right|_{t=s(x,p)} a(x+tp) = 0, \tag{5.5}$$

where $\langle \cdot \,|\, \cdot \rangle$ denotes the scalar product in $\mathbb{R}^n$.

The following proposition says that 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 quadric and the point of tangency describes a geodesic on it.

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 function*

$$t \mapsto \xi(t) := \xi(x(t), p(t)) \in \mathbb{R}^n, \tag{5.6}$$

(i) is contained in a fixed level set $M_c = a^{-1}(c)$, for some $c \in \mathbb{R}$,

(ii) is a geodesic on M_c.

Proof. Property (i) is a simple consequence of Corollary 4.20, 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

$$\dot{s}(t)\left\langle \nabla^2_{\xi(t)}a\,p(t)\,\middle|\,p(t)\right\rangle - |\nabla_{\xi(t)}a|^2=0.$$

and the strict convexity of a implies $\left\langle \nabla^2_{\xi(t)}a\,p(t)\,\middle|\,p(t)\right\rangle \neq 0$ and

$$\dot{s}(t)=\frac{|\nabla_{\xi(t)}a|^2}{\left\langle \nabla^2_{\xi(t)}a\,p(t)\,\middle|\,p(t)\right\rangle}\neq 0.$$

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 $|\dot{\xi}(t)|=\dot{s}(t)$.

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 ξ reparametrized by the 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 perpendicular 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 ξ, η 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*

$$\{h,g\}=(s-\tau)\left\langle \nabla_\xi a\,|\,\nabla_\eta b\right\rangle. \tag{5.9}$$

Proof. The coordinate expression of the Poisson bracket (4.19) can be rewritten as

$$\{h,g\}=\left\langle \nabla_p h\,|\,\nabla_x g\right\rangle-\left\langle \nabla_x h\,|\,\nabla_p g\right\rangle, \tag{5.10}$$

and using equation (5.8) for both h and g one gets

$$\{h,g\}=(s-\tau)\left\langle \nabla_\xi a\,|\,\nabla_\eta b\right\rangle. \tag{5.11}$$

□

5.3 Sub-Riemannian structures with symmetries

Recall that, for a sub-Riemannian manifold, we denote by H the sub-Riemannian Hamiltonian.

Definition 5.13. 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 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 5.14. *Let M be a sub- Riemannian manifold and H the sub-Riemannian Hamiltonian. For a vector field $X \in \mathrm{Vec}(M)$ is a Killing vector field if and only if $\{H, h_X\} = 0$.*

Proof. A vector field X generates isometries if and only if, by definition, 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

$$H((e^{tX})^*\lambda) = H(\lambda), \qquad \forall\, \lambda \in T^*M. \tag{5.12}$$

On the other hand Proposition 4.10 implies that $(e^{tX})^* = e^{t\vec{h}_X}$, where h_X is the Hamiltonian linear on fibers related to X. Differentiating (5.12) with respect to t we find the equivalence

$$H \circ e^{tX*} = H \quad \Leftrightarrow \quad \vec{h}_X H = 0 \quad \Leftrightarrow \quad \{H, h_X\} = 0.$$

□

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 5.15 (Revolution surfaces in $\mathbb{R}^3$)**.** Let M be a 2-dimensional revolution surface in $\mathbb{R}^3$. Since the rotation around the revolution axis preserves the Riemannian structure, by definition, we have that the Hamiltonian generated by this flow and the Riemannian Hamiltonian H are in involution.

Example 5.16 (Isoperimetric sub-Riemannian problem)**.** Let us consider a sub-Riemannian structure associated with an isoperimetric problem defined on a 2-dimensional revolution 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.

(i) By construction the problem is invariant by translation along the z-axis

(ii) If, moreover, both M and b are rotational invariant we find a first integral of the geodesic flow as in the previous example

5.4 Completely integrable systems

Let M be an n-dimensional smooth manifold and assume that there exist n independent Hamiltonians in involution in T^*M, i.e. a set of n smooth functions

$$\begin{aligned} &h_i : T^*M \to \mathbb{R}, \qquad i = 1, \dots, n, \\ &\{h_i, h_j\} = 0, \qquad i, j = 1, \dots, n. \end{aligned} \tag{5.13}$$

such that the differentials $d_\lambda h_1, \dots, d_\lambda h_n$ of the functions are independent on an open dense set of point $\lambda \in T^*M$.

Let us consider the vector valued map, called *moment map*, defined by

$$h : T^*M \to \mathbb{R}^n, \qquad h = (h_1, \dots, h_n).$$

Definition 5.17. Under the assumptions (5.13), then we say that the map h is *completely integrable*. The same terminology applies to any of the Hamiltonian system defined by one of the Hamiltonian h_i, for $i = 1, \dots, n$.

Lemma 5.18. *Assume that h is completely integrable and $c \in \mathbb{R}^n$ be a regular value of h. Then the set $h^{-1}(c)$ is a n-dimensional submanifold in T^*M and we have*

$$T_\lambda h^{-1}(c) = \operatorname{span}\{\vec{h}_1(\lambda), \dots, \vec{h}_n(\lambda)\}, \qquad \forall\, \lambda \in h^{-1}(c). \tag{5.14}$$

Proof. Since c is a regular value of h, by Remark 2.58 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 2.11, 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 (5.14) 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

$$\dim \operatorname{span}\{d_\lambda h_1, \dots, d_\lambda h_n\} = n, \quad \forall\, \lambda \in h^{-1}(c). \tag{5.15}$$

Moreover the symplectic form σ on T^*M induces for all λ 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

$$\dim \operatorname{span}\{\vec{h}_1(\lambda), \dots, \vec{h}_n(\lambda)\} = n, \quad \forall\, \lambda \in h^{-1}(c). \tag{5.16}$$

hence they form a basis for $T_\lambda h^{-1}(c)$. $\square$

Remark 5.19. 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 $i, j = 1, \dots, n$.

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 5.20. *Assume that the vector fields $\vec{h}_i$ are complete and define the map*

$$\Psi : \mathbb{R}^n \to \operatorname{Diff}(N_c), \qquad \Psi(s_1, \dots, s_n) := e^{s_1 \vec{h}_1} \circ \ldots \circ e^{s_n \vec{h}_n}\Big|_{N_c}. \tag{5.17}$$

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 $i, j = 1, \dots, n$ since

$$[\vec{h}_i, \vec{h}_j] = \overrightarrow{\{h_i, h_j\}} = 0.$$

By Proposition 2.26, this is equivalent to

$$e^{t\vec{h}_i} \circ e^{\tau\vec{h}_j} = e^{\tau\vec{h}_j} \circ e^{t\vec{h}_i}, \qquad \forall\, t, \tau \in \mathbb{R}. \tag{5.18}$$

Thus, for every λ, the map Ψ_λ is a smooth local diffeomorphism between at each point. Indeed, using (5.18), one has (cf. also Exercice 2.31)

$$\frac{\partial \Psi_\lambda}{\partial s_i}(\Psi_\lambda(s)) = \vec{h}_i(\Psi_\lambda(s)), \quad i = 1, \dots, n,$$

and the partial derivatives are linearly independent at each point of N_c.

Since the vector fields are complete by assumption, we can compute for every $s, s' \in \mathbb{R}^n$

$$\begin{aligned}
\Psi(s + s') &= e^{(s_1+s_1')\vec{h}_1} \circ \ldots \circ e^{(s_n+s_n')\vec{h}_n} \\
&= e^{s_1\vec{h}_1} \circ e^{s_1'\vec{h}_1} \circ \ldots \circ e^{s_n\vec{h}_n} \circ e^{s_n'\vec{h}_n} \\
&= e^{s_1\vec{h}_1} \circ \ldots \circ e^{s_n\vec{h}_n} \circ e^{s_1'\vec{h}_1} \circ \ldots \circ e^{s_n'\vec{h}_n} \qquad \text{(by (5.18))} \\
&= \Psi(s) \circ \Psi(s'),
\end{aligned}$$

which proves that Ψ is a group action. Denote, for every point $\lambda \in N_c$, its orbit under the group action, namely

$$\Omega_\lambda = \operatorname{im} \Psi_\lambda = \{\Psi_\lambda(s) \mid s \in \mathbb{R}^n\}.$$

Exercise 5.21. Using the fact that N_c is connected, prove that $\Omega_\lambda = N_c$ for every $\lambda \in N_c$.

Hence the map Ψ_λ is surjective, but in general it is not injective (as for instance in the case when M is compact). As a consequence we consider the stabiliser S_λ of the point λ, i.e. the set

$$S_\lambda = \{s \in \mathbb{R}^n \mid \Psi_\lambda(s) = \lambda\},$$

Exercise 5.22. Prove that S_λ is a discrete[2] subgroup of $\mathbb{R}^n$, independent on $\lambda \in N_c$.

Then the proof of Proposition 5.20 is completed by the next lemma.

Lemma 5.23. *Let G be a non trivial discrete subgroup of $\mathbb{R}^n$. Then there exist $k \in \mathbb{N}$ with $1 \le k \le n$ and $v_1, \dots, v_k \in \mathbb{R}^n$ such that*

$$G = \left\{ \sum_{i=1}^{k} m_i v_i,\ m_i \in \mathbb{Z} \right\}.$$

[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 $U \cap G = \{g\}$.

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 $e_1 \neq 0$ 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 5.24. (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 ℓ that does not contain elements of G/G_1.

By Exercise 5.24 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 $v_2, \dots, v_k$ such that

$$G/G_1 = \left\{ \sum_{i=2}^{k} m_i v_i,\ m_i \in \mathbb{Z} \right\}. \qquad \square$$

$\square$

Corollary 5.25. *The connected manifold 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 $(\theta_1, \dots, \theta_k) \in T^k$ and $(\theta_{k+1}, \dots, \theta_n) \in \mathbb{R}^{n-k}$, then we have*

$$\vec{h}_i = \sum_{j=1}^{n} b_{ij}(c) \partial_{\theta_j}, \tag{5.19}$$

for some constants $b_{ij}(c)$ independent on $\lambda \in N_c$.

Proof. Fix $c \in \mathbb{R}^n$ and a point $\lambda \in N_c$. Let us consider the elements $v_1, \dots, v_k \in \mathbb{R}^n$ generators of the stabiliser S_λ (independent on λ) given by Lemma 5.23 and complete it to a global basis $v_1, \dots, v_n$. Denote by $e_1, \dots, e_n$ the canonical basis of $\mathbb{R}^n$ and by $B : \mathbb{R}^n \to \mathbb{R}^n$ any isomorphism such that $Be_i = v_i$ for $i = 1, \dots, n$. We stress again that B does not depend on $\lambda \in N_c$ and is thus a function of c only.

Then clearly the map $B \circ \Psi_\lambda : \mathbb{R}^n \to N_c$ is a local diffeomorphism and, due to the fact that S_λ is the stabiliser of Ψ_λ, 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 $(\theta_1, \dots, \theta_n)$ in $\mathbb{R}^n$ induced by the choice of the basis $v_1, \dots, v_n$.

Since $(\theta_1, \dots, \theta_n)$ are obtained by $(s_1, \dots, s_n)$ 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 $\partial_{\theta_1}, \dots, \partial_{\theta_n}$ can be expressed as follows

$$\vec{h}_i = \partial_{s_i} = \sum_{j=1}^{n} b_{ij}(c)\partial_{\theta_j}, \tag{5.20}$$

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, due to the fact that the level set N_c is not compact, the set (c, θ) do not define local coordinates on T^*M. *If we assume* that (c, θ) define a set of local coordinates, then the Hamiltonian system defined by h_i takes the form (on the whole space T^*M)

$$\begin{cases} \dot{c} = 0 \\ \dot{\theta}_j = b_{ij}(c) \end{cases}, \qquad i = 1, \dots, n. \tag{5.21}$$

Notice that, as soon as (c, θ) define local coordinates, the coordinate set $(\theta_1, \dots, \theta_n)$ are not uniquely defined. In particular, every transformation of the kind $\theta_i \mapsto \theta_i + \psi_i(c)$ still defines a set of cylindirical coordinates on each level set. The choice of the functions $\psi_i(c)$ corresponds to the choice of the initial value of θ_i at a point (for every choice of c). However, the vector fields ∂_{θ_i} are independent on this choice.

5.5 Arnold-Liouville theorem

In this section we consider in detail the case when the level set of a completely integrable system defined by

$$h : T^*M \to \mathbb{R}^n, \qquad h = (h_1, \dots, h_n),$$

are *compact*. More precisely we assume that for all values of $c \in \mathbb{R}$ the level set $h^{-1}(c)$ is a smooth compact and connected manifold. From Proposition 5.20 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

$$F_c : T^n \to N_c, \qquad F_c(\theta_1, \dots, \theta_n) = \Psi_\lambda(\theta_1 + 2\pi\mathbb{Z}, \dots, \theta_n + 2\pi\mathbb{Z}).$$

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 $(c_1, \dots, c_n, \theta_1, \dots, \theta_n)$ is a well defined coordinate set on $\mathcal{O}$.

Theorem 5.28 (Arnold-Liouville)**.** *Let us consider a moment map $h : T^*M \to \mathbb{R}^n$ associated with a completely integrable system such that every level set N_c is compact and connected. Then for every $\bar{c} \in \mathbb{R}$ there exists a neighborhood $\mathcal{O}$ of $N_{\bar{c}}$ and a change of coordinates*

$$(c_1, \dots, c_n, \theta_1, \dots, \theta_n) \mapsto (I_1, \dots, I_n, \varphi_1, \dots, \varphi_n) \tag{5.22}$$

such that

(i) $I = \Phi \circ h$*, where* $\Phi : h(\mathcal{O}) \to \mathbb{R}^n$ *is a diffeomorphism,*

(ii) $\sigma = \sum_{j=1}^n dI_j \wedge d\varphi_j$.

Definition 5.29. The coordinates (I, φ) 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 θ_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 $0 \leq \tau \leq \varepsilon$.

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 γ_i.[3]

Let us compute the Jacobian of the change of variables.
$$\begin{aligned}
\frac{\partial I_i}{\partial c_j}(c) &= \frac{1}{2\pi} \frac{\partial}{\partial \varepsilon}\bigg|_{\varepsilon=0} \left(\int_{\gamma_i(c^{j,\varepsilon})} s - \int_{\gamma_i(c)} s \right) \\
&= \frac{1}{2\pi} \frac{\partial}{\partial \varepsilon}\bigg|_{\varepsilon=0} \int_{\partial C_i^{j,\varepsilon}} s \\
&= \frac{1}{2\pi} \frac{\partial}{\partial \varepsilon}\bigg|_{\varepsilon=0} \int_{C_i^{j,\varepsilon}} \sigma \qquad\qquad \text{(where } \sigma = ds\text{)} \\
&= \frac{1}{2\pi} \frac{\partial}{\partial \varepsilon}\bigg|_{\varepsilon=0} \int_{c_j}^{c_j+\varepsilon} \int_{\gamma_i(c^{j,\tau})} \sigma(\partial_{c_j}, \partial_{\theta_i}) d\theta_i d\tau \\
&= \frac{1}{2\pi} \int_{\gamma_i(c)} \sigma(\partial_{c_j}, \partial_{\theta_i}) d\theta_i.
\end{aligned}$$

Using that $\partial_{\theta_i} = \sum_{j=1}^n b^{ij}(c) \vec{h}_j$ (see (5.20)) (where b^{ij} are the entries of the inverse matrix of b_{ij}) one gets
$$\sigma(\cdot, \partial_{\theta_i}) = \sum_{j=1}^n b^{ij}(c) dh_j. \tag{5.23}$$

[3]Hence, in principle, we are free to choose any basis $\gamma_1, \dots, \gamma_n$ for the fundamental group of T^n.

Moreover $dh_i = dc_i$ since they define the same coordinate set. Hence

$$\begin{aligned}\frac{\partial I_i}{\partial c_j}(c) &= \frac{1}{2\pi}\int_{\gamma_i(c)} \left\langle \sum_{k=1}^n b^{ik}(c)dc_k, \partial_{c_i} \right\rangle d\theta_i \\ &= \frac{1}{2\pi}\int_{\gamma_i(c)} b^{ij}(c)d\theta_i \\ &= b^{ij}(c)\end{aligned}$$

Combining the last identity with (5.23) 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, θ)

$$\sigma = \sum_{i,j=1}^n a_{ij}(I)dI_i \wedge dI_j + \sum_{i=1}^n dI_i \wedge d\theta_i. \tag{5.24}$$

where the smooth functions a_{ij} depends only on the action variables, since the symplectic form σ 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 (5.24) 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 σ 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.

There then exist, as mentioned above, special sets of canonical coordinates on the phase space such that the invariant tori are the level sets of the action variables, and the angle variables are the natural periodic coordinates on the torus. 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, φ) is written as follows

$$\dot{I}_i = -\frac{\partial h_j}{\partial \varphi_i} = 0, \qquad \dot{\varphi}_i = \frac{\partial h_j}{\partial I_i}(I). \tag{5.25}$$

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 (and, as a consequence, on quadrics) is completely integrable. More precisely we consider the particular case when the function a is a quadratic polynomial, i.e. every level set of our function is a quadric in $\mathbb{R}^n$. The presentation follows the arguments of Moser [80].

Definition 5.31. Let A be an $n \times n$ non degenerate symmetrix matrix. The *quadric* $\mathcal{Q}$ associated to A is the set

$$\mathcal{Q} = \{x \in \mathbb{R}^n,\, \langle A^{-1}x, x\rangle = 1\}. \tag{5.26}$$

For simplicity we deal with the case when A has simple distinct eigenvalues $\alpha_1 < \ldots < \alpha_n$. Define, for every λ that is not an eigenvalue of A,

$$a_\lambda(x) = \langle (A - \lambda I)^{-1}x, x\rangle, \qquad \mathcal{Q}_\lambda = \{x \in \mathbb{R}^n,\, a_\lambda(x) = 1\}.$$

If $A = \mathrm{diag}(\alpha_1, \ldots, \alpha_n)$ is a diagonal matrix then (5.26) 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 quadrics that are confocal to $\mathcal{Q}$

$$\mathcal{Q}_\lambda = \left\{ x \in \mathbb{R}^n,\, \sum_{i=1}^n \frac{x_i^2}{\alpha_i - \lambda} = 1 \right\}, \qquad \forall\, \lambda \in \mathbb{R} \setminus \Lambda,$$

where $\Lambda = \{\alpha_1, \ldots, \alpha_n\}$ denotes the set of eigenvalues of A. Note 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. 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 $\{\mathcal{Q}_\lambda\}_{\lambda \in \Lambda}$ be the 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 $\mathbb{R} \setminus \Lambda$

$$\frac{\partial a_\lambda}{\partial \lambda}(x) = \left\langle A_\lambda^2 x, x\right\rangle = |A_\lambda x|^2 \geq 0, \qquad \text{where} \qquad A_\lambda := (A - \lambda I)^{-1},$$

as follows from part (i) of Exercise 5.32 and the fact that A (hence A_λ) is self-adjoint. Thus $a_\lambda(x)$ is monotone increasing as a function of λ, and takes values from $-\infty$ to $+\infty$ in each interval $]\alpha_i, \alpha_{i+1}[$ contained between two eigenvalues of A. This implies that, for a fixed x, there exist exactly n values

$\lambda_1, \ldots, \lambda_n$ such that $a_{\lambda_i}(x) = 1$ (that means $x \in \mathcal{Q}_{\lambda_i}$). 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 $x \in \mathcal{Q}_{\lambda_i} \cap \mathcal{Q}_{\lambda_j}$:

$$\begin{aligned}
\left\langle \nabla_x a_{\lambda_i}, \nabla_x a_{\lambda_j} \right\rangle &= 4 \left\langle A_{\lambda_i} x, A_{\lambda_j} x \right\rangle \\
&= 4 \left\langle A_{\lambda_i} A_{\lambda_j} x, x \right\rangle \\
&= \frac{4}{\lambda_j - \lambda_i} (\langle A_{\lambda_i} x, x\rangle - \left\langle A_{\lambda_j} x, x \right\rangle) = 0,
\end{aligned}$$

where again we used the fact that A_λ is selfadjoint and $\langle A_\lambda x, x \rangle = 1$ for all λ. □

Now we define the family of Hamiltonians associated with the family of confocal quadrics

$$h_\lambda(x,p) = \min_t a_\lambda(x+tp) = a_\lambda(\xi_\lambda(x,p)), \tag{5.27}$$

Remark 5.34. Notice that the minimum in (5.27) is attained at a unique point, and the function a_λ satisfies the assumptions (A1)-(A2) introduced in Section **??**, only if the corresponding quadric is an ellipsoid.

In what follows we generalize the considerations to all quadrics associated to $\lambda \in \mathbb{R} \setminus \Lambda$. Indeed we can still define the hamiltonian h_λ as the value of the function a_λ at its critical point along an affine line (hence defining h_λ as an Hamiltonian on the set of affine lines as well).

Now we prove another 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 5.35. *Assume that two confocal quadrics are tangent to a given line, i.e. there exist $x, y \in \mathbb{R}^n$ 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 $x + \mathbb{R}y$ at ξ_λ 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)

$$\begin{aligned}
\left\langle \nabla_{\xi_\lambda} a_\lambda, \nabla_{\xi_\mu} a_\mu \right\rangle &= 4 \langle A_\lambda \xi_\lambda, A_\mu \xi_\mu \rangle \\
&= 4 \langle A_\lambda A_\mu \xi_\lambda, \xi_\mu \rangle \\
&= \frac{4}{\mu - \lambda} (\langle A_\lambda \xi_\lambda, \xi_\mu \rangle - \langle A_\mu \xi_\mu, \xi_\lambda \rangle) = 0,
\end{aligned}$$

This implies also $\{h_\lambda, h_\mu\} = 0$, thanks to Proposition 5.12. □

Proposition 5.36. *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 5.37. (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_λ is well defined on the quotient $\mathbb{R}^n/\mathbb{R}p$.
(ii). Prove that $\{a^p_\lambda\}_\lambda$ is a family of confocal quadric 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 $\{a_\lambda\}_\lambda$. $\square$

Remark 5.38. 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 5.36 mutually orthogonal and orthogonal to the line itself.

Theorem 5.39. *The geodesic flow on an ellipsoid is completely integrable. In particular, the tangents of any geodesics on an ellipsoid are tangent to the same set of its confocal quadrics, i.e. independently on the point on the geodesic.*

Proof. We want to show that the functions $\lambda_1(x,p), \ldots, \lambda_{n-1}(x,p)$ (as functions defined on the set of lines in $\mathbb{R}^n$) that assign to each line $x + \mathbb{R}p$ in $\mathbb{R}^n$ the $n-1$ values of λ such that the line is tangent to $\mathcal{Q}_\lambda$ are independent and in involution.

First notice that each level set $\lambda_i(x,p) = c$ coincide with the level set $h_c = 1$. Hence, by Exercise 4.33, the two functions defines the same Hamiltonian flow on this level set (up to reparametrization). We are then reduced to prove that the functions $h_{c_1}, \ldots, h_{c_{n-1}}$ are independent and in involution, which is a consequence of Proposition 5.35.

Since the lines that are tangent to a geodesic on the ellipsoid $\mathcal{Q}_\lambda$ form an integral curve of the Hamiltoian flow of the associated function h_λ, 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. $\square$

Bibliographical notes

The notion of complete integrability introduced here is the classical one given by Liouville and Arnold [9]. 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 [65]. 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 [80], 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 [79, 10, 72].

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.

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 γ 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 explain 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 following 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 we are 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 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 ϕ_i for $i = 1, \ldots, n$ 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 the geometric meaning of the object we are dealing with.

6.2 Duality

The basic idea behind this formal construction is to replace nonlinear objects defined on the manifold M with their linear counterpart, when interpreted as maps on the space $C^\infty(M)$ of smooth functions on M.

We recall that the set $C^\infty(M)$ of smooth functions on M is an $\mathbb{R}$-algebra with the usual operation of pointwise addition and multiplication

$$\begin{aligned} (a+b)(q) &= a(q) + b(q), \\ (\lambda a)(q) &= \lambda\, a(q), \qquad\qquad a, b \in C^\infty(M),\ \lambda \in \mathbb{R}, \\ (a \cdot b)(q) &= a(q) b(q). \end{aligned}$$

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

$$\widehat{P} : C^\infty(M) \to C^\infty(M), \qquad \widehat{P}(a) := a(P(q)).$$

which is an automorphism of the algebra $C^\infty(M)$.

Remark 6.1. One can prove that for every nontrivial 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 [8, Appendix A].

Next we want to characterize tangent vectors as functionals on $C^\infty(M)$. As explained in Chapter 2, a tangent vector $v \in T_qM$ defines in a natural way the derivation in the direction of v, i.e. the functional

$$\widehat{v} : C^\infty(M) \to \mathbb{R}, \qquad \widehat{v}(a) = \langle d_q a, v \rangle\,,$$

that satisfies the Leibniz rule

$$\widehat{v}(a \cdot b) = \widehat{v}(a) b(q) + a(q) \widehat{v}(b), \qquad \forall\, a, b \in C^\infty(M).$$

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

$$\widehat{v} = \left.\frac{d}{dt}\right|_{t=0} \widehat{q}(t) : C^\infty(M) \to \mathbb{R}. \tag{6.3}$$

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 $f \in C^\infty(M)$

$$\|f\|_{\alpha,K} = \sup_{q \in K,} \{|(X_{i_\ell} \odot \cdots \odot X_{i_1} f)(q)| : 1 \le i_j \le r, 0 \le \ell \le \alpha\}$$

The family of seminorms $\|\cdot\|_{\alpha,K}$ induces a topology on $C^\infty(M)$ with countable local bases 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 $f \in C^\infty(M)$, a countable local base of neighborhood of f is given by

$$U_{f,n} := \left\{ g \in C^\infty(M) : \|f - g\|_{n,K_n} \le \frac{1}{n} \right\}, \qquad n \in \mathbb{N}. \tag{6.4}$$

Exercise 6.2. (i). Prove that (6.4) 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 [62, 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 one can see, for instance, [62, Chapter 2].

Example 6.4. Prove that, given a diffeomorphism $P \in \mathrm{Diff}(M)$ and $\alpha \in \mathbb{N}$, there exists a constant $C_{\alpha,P} > 0$ such that for all $f \in C^\infty(M)$ one has

$$\|Pf\|_{\alpha,K} \leq C_{\alpha,P}\|f\|_{\alpha,P(K)}, \qquad \forall\, K \subset M.$$

In other words the diffeomorphism P, when interpreted as a linear operator on $C^\infty(M)$, is continuous in the Whitnhey topology. One can then define its seminorm

$$\|P\|_{\alpha,K} := \sup\{\|Pf\|_{\alpha,K} : \|f\|_{\alpha,P(K)} \leq 1\}.$$

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 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
$$\int_{t_0}^{t_1} \|a_t\|_{\alpha,K}dt < \infty,$$
for every $\alpha \in \mathbb{N}$ and $K \subset M$ compact.
- *absolutely continuous*, if there exists a locally integrable family of functions b_t such that
$$a_t = a_{t_0} + \int_{t_0}^{t} b_s ds.$$
- *Lipschitz*, if
$$\|a_t - a_s\|_{\alpha,K} \leq C_{s,K}|t - s|,$$
for every $\alpha \in \mathbb{N}$ and $K \subset M$ compact.

Analogous regularity property 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 continuos (differentiable, etc.) if the map $t \mapsto A_t a$ has the same property for every $a \in C^\infty(M)$.

We define a *non-autonomous vector field* as a family of vector fields X_t that is locally bounded. A *non-autonomous flow* is a family of diffeomorphisms P_t that is absolutely continuous. Hence, for any non-autonomous 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 non-autonomous 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 family of vector fields we denote

$$\int_0^t X_s \, ds : a \mapsto \int_0^t X_s a \, ds$$

$$\frac{d}{dt} X_t : a \mapsto \frac{d}{dt}(X_t a)$$

One can show that if A_t and B_t are continuous families of operators on $C^\infty(M)$ wich are differentiable at some t_0, then the family $A_t \circ B_t$ is differentiable at t_0 and satisfies the Leibnitz rule

$$\left.\frac{d}{dt}\right|_{t=t_0} (A_t \circ B_t) = \left(\left.\frac{d}{dt}\right|_{t=t_0} A_t \right) \circ B_{t_0} + A_{t_0} \circ \left(\left.\frac{d}{dt}\right|_{t=t_0} B_t \right). \tag{6.5}$$

The same result holds true for the composition of functionals with operators. For a proof of the last fact one can see [8, Chapter 2 and Appendix A].

6.4 Operator ODE and Volterra expansion

Consider a nonautonomous vector field X_t and the corresponding nonautonomous ODE

$$\frac{d}{dt} q(t) = X_t(q(t)), \qquad q \in M. \tag{6.6}$$

Using the notation introduced in the previous section we can rewrite (6.6) in the following way

$$\frac{d}{dt} q(t) = q(t) \circ X_t. \tag{6.7}$$

Indeed assume that $q(t)$ satisfies (6.6) and let $a \in C^\infty(M)$. Using "hat" notation of Section 6.2

$$\left(\frac{d}{dt} \widehat{q}(t) \right) a = \frac{d}{dt} \widehat{q}(t) a = \frac{d}{dt} a(q(t)) = \langle d_{q(t)} a, X_t(q(t)) \rangle = (\widehat{X}_t a)(q(t)) = (\widehat{q}(t) \circ \widehat{X}_t) a. \tag{6.8}$$

As discussed in Chapter 2, the solution to the nonautonomous ODE (6.6) defines a flow, i.e., family of diffeomorphisms, $P_{s,t}$.

Lemma 6.5. *The flow $P_{s,t}$ defined by* (6.10) *satisfies the operator differential equation*

$$\frac{d}{dt} P_{s,t} = P_{s,t} \circ X_t, \qquad P_{s,s} = \mathrm{Id}. \tag{6.9}$$

Proof. Fix a point $q_0 \in M$ and denote by $q(t)$ the solution of the Cauchy problem (6.6) with initial condition $q(s) = q_0$. By the very definition of $P_{s,t}$ we have that $q(t) = P_{s,t}(q_0)$, which rewrites as $q(t) = q_0 \circ P_{s,t}$. □

Definition 6.6. We call $P_{s,t}$ the *right chronological exponential* and use the notation

$$P_{s,t} := \overrightarrow{\exp} \int_s^t X_\tau d\tau. \tag{6.10}$$

Notice that the arrow in the notation recalls in which "position" the vector field appears when differentiating the flow (cf. (6.9)).

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 (6.9) rewrites as

$$\begin{cases} \dot{P}_t = P_t \circ X_t \\ P_0 = \mathrm{Id} \end{cases} \tag{6.11}$$

and can be rewritten as an integral operator equation as follows

$$P_t = \mathrm{Id} + \int_0^t P_s \circ X_s ds \tag{6.12}$$

Replacing iteratively P_s in the right hand side of (6.12) with the equation (6.12) itself, we have

$$\begin{aligned} P_t &= \mathrm{Id} + \int_0^t \left(\mathrm{Id} + \int_0^{s_1} P_{s_2} \circ X_{s_2} ds_2 \right) \circ X_{s_1} ds_1 \\ &= \mathrm{Id} + \int_0^t X_s ds + \iint\limits_{0 \le s_2 \le s_1 \le t} P_{s_2} \circ X_{s_2} \circ X_{s_1} ds_1 ds_2 \\ &= \dots \\ &= \mathrm{Id} + \sum_{k=1}^{N-1} \int\cdots\int\limits_{0 \le s_k \le \dots \le s_1 \le t} X_{s_k} \circ \cdots \circ X_{s_1} d^k s + R_N \end{aligned}$$

where the remainder term is defined as follows

$$R_N := \int\cdots\int\limits_{0 \le s_N \le \dots \le s_1 \le t} P_{s_N} \circ X_{s_N} \circ \cdots \circ X_{s_1} d^N s$$

Formally, letting $N \to \infty$ and assuming that $R_N \to 0$, we can write the flow P_t as the chronological series

$$\mathrm{Id} + \sum_{k=1}^{\infty} \int\cdots\int\limits_{\Delta_k(t)} X_{s_k} \circ \cdots \circ X_{s_1} d^k s \tag{6.13}$$

where $\Delta_k(t) = \{(s_1, \dots, s_k) \in \mathbb{R}^k | \, 0 \le s_k \le \dots \le s_1 \le t\}$ denotes the k-dimensional symplex.

A discussion about the convergence of the series is contained in Section 6.A.

Remark 6.7. If we write expansion (6.13) when $X_t = X$ is an autonomous vector field, we find that the chronological exponential coincides with the exponential of the vector field

$$\overrightarrow{\exp} \int_0^t X ds \simeq \mathrm{Id} + \sum_{k=1}^{\infty} \int \cdots \int_{\Delta_k(t)} \underbrace{X \circ \cdots \circ X}_{k} d^k s$$
$$\simeq \sum_{k=0}^{\infty} \mathrm{vol}(\Delta_k(t)) X^k = \sum_{k=0}^{\infty} \frac{t^k}{k!} X^k = e^{tX},$$

since $\mathrm{vol}(\Delta_k(t)) = t^k/k!$. In the nonautonomous case for different time X_{s_1} and X_{s_2} might not commute, hence the order in which the vector fields appears in the composition is crucial. The arrow in the notation recalls in which "direction" the parameters are increasing.

Exercise 6.8. Prove that in general, for a nonautonomous vector field X_t, one has

$$\overrightarrow{\exp} \int_0^t X_s ds \neq e^{\int_0^t X_s ds}. \tag{6.14}$$

Prove that if $[X_t, X_\tau] = 0$ for all $t, \tau \in \mathbb{R}$ then the equality holds in (6.14)

Proposition 6.9. *Assume that* P_t *satisfies* (6.11) *and consider the inverse flow* $Q_t := (P_t)^{-1}$*. Then* Q_t *satisfies the Cauchy problem*

$$\begin{cases} \dot{Q}_t = -X_t \circ Q_t, \\ Q_0 = \mathrm{Id}. \end{cases} \tag{6.15}$$

Proof. From the definition of inverse flow we have the identity $P_t \circ Q_t = \mathrm{Id}$, for every $t \in \mathbb{R}$. Differentiating and using the Leibnitz rule one obtains

$$\dot{P}_t \circ Q_t + P_t \circ \dot{Q}_t = 0. \tag{6.16}$$

Using (6.11) then we get

$$P_t \circ X_t \circ Q_t + P_t \circ \dot{Q}_t = 0 \tag{6.17}$$

Multiplying both sides by Q_t on the left, one gets (6.15). □

The solution to the problem (6.15) will be denoted by the *left chronological exponential*

$$Q_t := \overleftarrow{\exp} \int_0^t (-X_s) ds. \tag{6.18}$$

Repeating analogous reasoning, we find the formal expansion

$$\overleftarrow{\exp} \int_0^t (-X_s) ds \simeq \mathrm{Id} + \sum_{k=1}^{\infty} \int \cdots \int_{0 \leq s_k \leq \ldots \leq s_1 \leq t} (-X_{s_1}) \circ \cdots \circ (-X_{s_k}) d^k s.$$

The difference with respect to the right chronological exponential is in the order of composition. Again, the arrow over the exponential says in which direction the time increases and in which position the vector field appears when differentiating the flow.

We can summarize all the properties of the chronological exponential as follows

$$\frac{d}{dt}\overrightarrow{\exp}\int_0^t X_s ds = \overrightarrow{\exp}\int_0^t X_s ds \circ X_t, \tag{6.19}$$

$$\frac{d}{dt}\overleftarrow{\exp}\int_0^t X_s ds = X_t \circ \overleftarrow{\exp}\int_0^t X_s ds, \tag{6.20}$$

$$\left(\overrightarrow{\exp}\int_0^t X_s ds\right)^{-1} = \overleftarrow{\exp}\int_0^t (-X_s) ds. \tag{6.21}$$

6.4.2 Adjoint representation

Now we can 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

$$(P_* v)a = \frac{d}{dt}\bigg|_{t=0} a(P(q(t))) = \left(\frac{d}{dt}\bigg|_{t=0} q(t)\right) \circ Pa = v \circ Pa$$

Recall that, if $X \in \mathrm{Vec}(M)$ is a vector field we have $P_* X\big|_q = P_*(X\big|_{P^{-1}(q)})$. In a similar way we will find an expression for $P_* X$ as derivation of $C^\infty(M)$

$$P_* X = P^{-1} \circ X \circ P. \tag{6.22}$$

Remark 6.10. 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

$$P_* X = (\mathrm{Ad}\, P^{-1}) X, \tag{6.23}$$

where

$$\mathrm{Ad}\, P : X \mapsto P \circ X \circ P^{-1}, \qquad \forall\, X \in \mathrm{Vec}(M)$$

is the adjoint action of P on the space of vector fields[1].

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 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

$$\mathrm{ad}\, X : Y \mapsto [X, Y],$$

[1] this is the differential of the conjugation $Q \mapsto P \circ Q \circ P^{-1}$, $Q \in \mathrm{Diff}(M)$

is the adjoint action on the Lie algebra of vector fields.

In other words we proved that $\mathrm{Ad}\, P_t$ is a solution to the differential equation

$$\dot{A}_t = A_t \circ \mathrm{ad}\, X_t, \qquad A_0 = \mathrm{Id}.$$

Thus it can be expressed as chronological exponential and we have the identity

$$\mathrm{Ad}\left(\overrightarrow{\exp}\int_0^t X_s ds\right) = \overrightarrow{\exp}\int_0^t \mathrm{ad}\, X_s ds. \tag{6.24}$$

Notice that combining (6.24) with (6.23) in the case of an autonomous vector field one gets

$$e_*^{-tX} = e^{t\,\mathrm{ad}\, X} \tag{6.25}$$

Exercise 6.11. Prove that, if $[X_t, Y] = 0$ for all t, then $(\mathrm{Ad}\, P_t)Y = Y$.

Remark 6.12. More explicitly we can write the following formal expansion

$$(\mathrm{Ad}\, P_t)Y \simeq Y + \sum_{k=1}^{\infty} \int\cdots\int_{0\leq s_k\leq\ldots\leq s_1\leq t} [X_{s_n}, \ldots, [X_{s_2}, [X_{s_1}, Y]]d^k s, \tag{6.26}$$

which generalizes the formula (2.31). Indeed if $P_t = e^{tX}$ is the flow associated with an autonomous vector field we get

$$\begin{aligned}(\mathrm{Ad}\, e^{tX})Y = e_*^{-tX}Y &= Y + \sum_{k=1}^{\infty}\frac{t^k}{k!}[X, \ldots, [X, Y]] \\ &\simeq Y + t[X, Y] + \frac{t^2}{2}[X, [X, Y]] + o(t^2)\end{aligned}$$

Exercise 6.13. Prove the following using operator notation:

1. Show that ad is the infinitesimal version of the operator Ad, i.e. if P_t is a flow generated by the vector field $X \in \mathrm{Vec}(M)$ then

$$\mathrm{ad}\, X = \frac{d}{dt}\bigg|_{t=0} \mathrm{Ad}\, P_t.$$

2. Show that, if $P \in \mathrm{Diff}(M)$, then P_* preserves Lie brackets, i.e. $P_*[X, Y] = [P_*X, P_*Y]$.

3. Show that the Jacobi identity in $\mathrm{Vec}(M)$ is the infinitesimal version of the identity proved in 2. (Hint. use $P_t = e^{tZ}$)

Exercise 6.14. Prove the following change of variables formula for a nonautonomous flow:

$$P \circ \overrightarrow{\exp}\int_0^t X_s ds \circ P^{-1} = \overrightarrow{\exp}\int_0^t (\mathrm{Ad}\, P)X_s ds. \tag{6.27}$$

Notice that for an autonomous vector field this identity reduces to (2.23).

6.5 Variations Formulae

Consider the following ODE

$$\dot{q} = X_t(q) + Y_t(q) \tag{6.28}$$

where Y_t is thought as a perturbation term of the equation (6.6). We want to describe the solution to the perturbed equation (6.28) as the perturbation of the solution of the original one.

Proposition 6.15. *Let X_t, Y_t be two nonautonomous vector fields. Then*

$$\overrightarrow{\exp}\int_0^t (X_s + Y_s)ds = \overrightarrow{\exp}\int_0^t \left(\overrightarrow{\exp}\int_0^s \operatorname{ad} X_\tau d\tau\right) Y_s ds \circ \overrightarrow{\exp}\int_0^t X_s ds \tag{6.29}$$

$$= \overrightarrow{\exp}\int_0^t (\operatorname{Ad} P_s) Y_s ds \circ P_t \tag{6.30}$$

where $P_t = \overrightarrow{\exp}\int_0^t X_s ds$ denotes the flow of the original vector field.

Proof. Our goal is to find a flow R_t such that

$$Q_t := \overrightarrow{\exp}\int_0^t (X_s + Y_s)ds = R_t \circ P_t \tag{6.31}$$

By definition of right chronological exponential we have

$$\dot{Q}_t = Q_t \circ (X_t + Y_t) \tag{6.32}$$

On the other hand, from (6.31), we also have

$$\begin{aligned}\dot{Q}_t &= \dot{R}_t \circ P_t + R_t \circ \dot{P}_t \\ &= \dot{R}_t \circ P_t + R_t \circ P_t \circ X_t \\ &= \dot{R}_t \circ P_t + Q_t \circ X_t \end{aligned} \tag{6.33}$$

Comparing (6.32) and (6.33), 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 $R_0 = \mathrm{Id}$ we find 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 (6.30). Plugging (6.24) in (6.30) one gets (6.29). □

Exercise 6.16. Prove the following versions of the variation formula:

(i) For every non autonomous vector fields X_t, Y_t on M

$$\overrightarrow{\exp}\int_0^t (X_s+Y_s)ds = \overrightarrow{\exp}\int_0^t X_s ds \circ \overrightarrow{\exp}\int_0^t \left(\overrightarrow{\exp}\int_t^s \operatorname{ad} X_\tau d\tau\right) Y_s ds \tag{6.34}$$

(ii) For every autonomous vector fields $X, Y \in \mathrm{Vec}(M)$ prove that

$$e^{t(X+Y)} = \overrightarrow{\exp}\int_0^t e^{s \operatorname{ad} X} Y ds \circ e^{tX} = \overrightarrow{\exp}\int_0^t e_*^{-sX} Y ds \circ e^{tX} \tag{6.35}$$

$$= e^{tX} \circ \overrightarrow{\exp}\int_0^t e^{(s-t)\operatorname{ad} X} Y ds \tag{6.36}$$

6.A Estimates and Volterra expansion

In this section we discuss the convergence of the Volterra expansion

$$\mathrm{Id} + \sum_{k=1}^{\infty} \int \cdots \int_{\Delta_k(t)} X_{s_k} \circ \cdots \circ X_{s_1} d^k s \tag{6.37}$$

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. Recall that if $X_s = X$ is autonomous then the series (6.37) simplifies in

$$\sum_{k=0}^{\infty} \frac{t^k}{k!} X^k \tag{6.38}$$

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 6.17. *Let X be a nonzero smooth vector field. Then there exists $a \in C^\infty(M)$ such that the Volterra expansion*

$$\sum_{k=0}^{\infty} \frac{t^k}{k!} X^k a \tag{6.39}$$

is not convergent at some point $q \in M$.

Proof. Fix a point $q \in M$ such that $X(q) \neq 0$ and consider 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 $X = \partial_{x_1}$ in $\mathbb{R}^n$. Fix an arbitrary sequence $(c_n)_{n\in\mathbb{N}}$ and let $f : I \to \mathbb{R}$ defined in a neighborhood I of 0 such that $f^{(n)}(0) = c_n$, for every $n \in \mathbb{N}$. The existence of such a function is guaranteed by Lemma 6.18. 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

$$\sum_{k=0}^{\infty} \frac{t^k}{k!} X^k a|_q = \sum_{k=0}^{\infty} \frac{t^k}{k!} c_k \tag{6.40}$$

which is not convergent for a suitable choice of the sequence $(c_n)_{n\in\mathbb{N}}$. $\square$

Lemma 6.18 (Borel lemma)**.** *Let $(c_n)_{n\in\mathbb{N}}$ be a real sequence. Then there exist a C^∞ 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^∞ bump function $\phi : \mathbb{R} \to \mathbb{R}$ with compact support and such that $\phi(0) = 1$ and $\phi^{(j)}(0) = 0$ for every $j \geq 1$. Then set

$$g_k(x) := \frac{c_k}{k!} x^k \phi\left(\frac{x}{\varepsilon_k}\right) \tag{6.41}$$

Notice that $g_k^{(j)}(0) = \delta_{jk} c_k$, where δ_{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

$$|g_k^{(j)}(x)| \leq 2^{-j}, \qquad \forall\, j \leq k-1, \forall\, x \in \mathbb{R}, \tag{6.42}$$

and define the function

$$f(x) := \sum_{k=0}^{\infty} g_k(x). \tag{6.43}$$

The series (6.43) converges uniformly with all the derivatives by (6.42) and, by differentiating under the sum, one obtains

$$f^{(j)}(x) := \sum_{k=0}^{\infty} g_k^{(j)}(x), \qquad f^{(j)}(0) := \sum_{k=0}^{\infty} g_k^{(j)}(0) = a_j.$$

□

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

$$S_N(t) := \mathrm{Id} + \sum_{k=1}^{N-1} \int \cdots \int_{\Delta_k(t)} X_{s_k} \circ \cdots \circ X_{s_1} d^k s$$

the N-th partial sum, we have the following estimate.

Theorem 6.19. *For every $t > 0$, $\alpha, N \in \mathbb{N}$, $K \subset M$ compact, we have*

$$\left\| \left(\overrightarrow{\exp} \int_0^t X_s ds - S_N(t) \right) a \right\|_{\alpha,K} \leq \frac{C}{N!} e^{C \int_0^t \|X_s\|_{\alpha,K'} ds} \left(\int_0^t \|X_s\|_{\alpha+N-1,K'} ds \right)^N \|a\|_{\alpha+N,K'}, \tag{6.44}$$

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.B. Let us specify this estimate for a non autonomous 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 $u \in L^2([0,T], \mathbb{R}^m)$.

Theorem 6.20. *For every $t > 0$, $\alpha, N \in \mathbb{N}$, $K \subset M$ compact, we have (denoting $\|u\|_{1,t} = \|u\|_{L^1([0,t],\mathbb{R}^m)}$)*

$$\left\| \left(\overrightarrow{\exp} \int_0^t \sum_{i=1}^{m} u_i(t) X_i - S_N(t) \right) a \right\|_{\alpha,K} \leq \frac{C}{N!} e^{C\|u\|_{1,t}} \|u\|_{1,t}^N \|a\|_{\alpha+N,K'} \tag{6.45}$$

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

$$\int_0^t \|X_s\|_{\alpha,K'} ds \leq \|u\|_{L^1([0,t],\mathbb{R}^m)} \tag{6.46}$$

Indeed we have for every f such that $\|f\|_{\alpha+1,K'} \leq 1$ that

$$\left\| \sum_{i=1}^m u_i(s) X_i f \right\|_{\alpha,K'} \leq \sup_{x\in K'} \left| X_{i_\ell} \circ \cdots \circ X_{i_1} \left(\sum_{i=1}^m u_i(s) X_i f \right) \right| \tag{6.47}$$

$$\leq \sup_{x\in K'} \sum_{i=1}^m |u_i(s)| |X_{i_\ell} \circ \cdots \circ X_{i_1} \circ X_i f| \leq \sum_{i=1}^m |u_i(s)| \tag{6.48}$$

□

To complete the discussion, let us describe a special case when the Volterra expansion is actually convergent. One can prove the following convergence result.

Proposition 6.21. *Let X_t be a nonautonomous vector field, locally bounded w.r.t. $t \in I$. 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) $\sup\{\|X_t a\| : a \in L, \|a\| \leq 1, t \in I\} < \infty$

Then the Volterra expansion (6.37) *converges on L for every $t \in I$.*

Proof. We can bound the general term of the sum with respect to the norm $\|\cdot\|$ of L

$$\left\| \int\cdots\int_{\Delta_k(t)} X_{s_k} \circ \cdots \circ X_{s_1} a \, d^k s \right\| \leq \int\cdots\int_{\Delta_k(t)} \|X_{s_k}\| \cdots \|X_{s_1}\| d^k s \, \|a\| \tag{6.49}$$

$$= \frac{1}{k!} \left(\int_0^t \|X_s\| ds \right)^k \|a\| \tag{6.50}$$

then the norm of the k-th term of the Volterra expansion is bounded above by the exponential series, and the Volterra expansion converges on L uniformly. □

Remark 6.22. The assumption in the theorem is satisfied in particular for a linear vector field X on $M = \mathbb{R}^n$ and $L \subset C^\infty(\mathbb{R}^n)$ the set of linear functions.

If 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 seet [5].

6.B Remainder term of the Volterra expansion

In this Appendix we prove Theorem 6.19. We start with the following key result.

Proposition 6.23. *Let X_t be a complete non autonomous 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*

$$\|P_{0,t}a\|_{\alpha,K} \leq Ce^{\int_0^t \|X_s\|_{\alpha,K'}ds}\|a\|_{\alpha,K'} \tag{6.51}$$

Proof. Define the compact set

$$K_t := \bigcup_{s\in[0,t]} P_{0,s}(K),$$

and the real function

$$\beta(t) := \sup\left\{\frac{\|P_{0,t}f\|_{\alpha,K}}{\|f\|_{\alpha+1,K_t}} \,\Big|\, f \in C^\infty(M), \|f\|_{\alpha+1,K_t} \neq 0\right\} \tag{6.52}$$

Notice that the function β is measurable in 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 6.24. *For every $t > 0$, $\alpha \in \mathbb{N}$ and $K \subset M$ compact, there exists $C > 0$ such that*

$$\|P_{0,t}f\|_{\alpha,K} \leq C\beta(t)\|f\|_{\alpha,K_t}, \qquad \forall\, f \in C^\infty(M). \tag{6.53}$$

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 6.24 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

$$\frac{\|P_{0,t}a\|_{\alpha,K}}{\|a\|_{\alpha+1,K_t}} \leq 1 + C\int_0^t \beta(s)\|X_s\|_{\alpha,K_t} ds$$

By definition (6.52) of the function β we have the inequality

$$\beta(t) \leq 1 + C\int_0^t \beta(s)\|X_s\|_{\alpha,K_t} ds \tag{6.54}$$

that by Gronwall inequality implies

$$\beta(t) \le e^{C\int_0^t \|X_s\|_{\alpha,K_t} ds} \tag{6.55}$$

and (6.51) follows combining the last inequality and (6.53) choosing f equal to a and for every compact set K' containing K_t. $\square$

Now we complete the proof of the main result, namely Theorem 6.19. Recall that we can write

$$\overrightarrow{\exp}\int_0^t X_s ds - S_N(t) = \int\!\!\cdots\!\!\int_{0\le s_N\le\ldots\le s_1\le t} P_{0,s_N}\circ X_{s_N}\circ\cdots\circ X_{s_1} ds$$

hence

$$\left\|\left(\overrightarrow{\exp}\int_0^t X_s ds - S_N(t)\right) a\right\|_{\alpha,K} \le \int\!\!\cdots\!\!\int_{0\le s_N\le\ldots\le s_1\le t} \|P_{0,s_N}\circ X_{s_N}\circ\cdots\circ X_{s_1}a\|_{\alpha,K}\, ds$$

Applying Proposition 6.23 to the function $X_{s_N}\circ\cdots\circ X_{s_1}a$ one obtains

$$\left\|\left(\overrightarrow{\exp}\int_0^t X_s ds - S_N(t)\right) a\right\|_{\alpha,K} \le Ce^{\int_0^t \|X_s\|_{\alpha,K} ds} \int\!\!\cdots\!\!\int_{0\le s_N\le\ldots\le s_1\le t} \|X_{s_N}\circ\cdots\circ X_{s_1}a\|_{\alpha,K'}\, ds \tag{6.56}$$

for some compact K' containing K. Now let us estimate the integral

$$\int\!\!\cdots\!\!\int_{0\le s_N\le\ldots\le s_1\le t} \|X_{s_N}\circ\cdots\circ X_{s_1}a\|_{\alpha,K'}\, ds \tag{6.57}$$

$$\le \int\!\!\cdots\!\!\int_{0\le s_N\le\ldots\le s_1\le t} \|X_{s_N}\|_{\alpha,K'} \left\|X_{s_{N-1}}\right\|_{\alpha+1,K'}\cdots\circ \|X_{s_1}\|_{\alpha+N-1,K'} \|a\|_{\alpha+N,K'}\, ds \tag{6.58}$$

$$\le \|a\|_{\alpha+N,K'} \int\!\!\cdots\!\!\int_{0\le s_N\le\ldots\le s_1\le t} \|X_{s_N}\|_{\alpha+N-1,K'} \left\|X_{s_{N-1}}\right\|_{\alpha+N-1,K'}\cdots\circ \|X_{s_1}\|_{\alpha+N-1,K'}\, ds \tag{6.59}$$

$$\le \|a\|_{\alpha+N,K'} \frac{1}{N!}\left(\int_0^t \|X_s\|_{\alpha+N-1,K'}\, ds\right)^N \tag{6.60}$$

and combining this inequality with (6.56) we are done.

Proof of Lemma 6.24. 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 $\{X_i\}_{i=1,\ldots,r}$ 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

$$\|P_{0,t}f\|_{\alpha,K} = \sup_{q\in K,}\{|(X_{i_\ell}\circ\cdots\circ X_{i_1}\circ P_{0,t}f)(q)| : 1\le i_j\le r, 0\le \ell\le\alpha\}$$

is attained (the existence guaranteed by compactness of K) and let p_f be the polynomial in $\mathbb{R}^n$ and of degree $\leq \alpha$ that coincides with the Taylor polynomial of degree α of f at $q_t = P_{0,t}(q_0)$. Then by construction we have

$$\|P_{0,t}f\|_{\alpha,K} \leq \|P_{0,t}p_f\|_{\alpha,K}, \qquad \|p_f\|_{\alpha,q_t} \leq \|f\|_{\alpha,K_t} \tag{6.61}$$

Moreover in the finite-dimensional space of polynomials in $\mathbb{R}^n$ of degree $\leq \alpha$ all norms are equivalent then there exist $C > 0$ such that

$$\|p_f\|_{\alpha,K_t} \leq C\|p_f\|_{\alpha,q_t} \tag{6.62}$$

Combining (6.61) and (6.62) with $\|p_f\|_{\alpha,K_t} = \|p_f\|_{\alpha+1,K_t}$ (since p_f is a polynomial of degree α) and the definition of β, we have

$$\frac{\|P_{0,t}f\|_{\alpha,K}}{\|f\|_{\alpha,K_t}} \leq \frac{\|P_{0,t}p_f\|_{\alpha,K}}{\|p_f\|_{\alpha,q_t}} \leq C\frac{\|P_{0,t}p_f\|_{\alpha,K}}{\|p_f\|_{\alpha,K_t}} \leq C\frac{\|P_{0,t}p_f\|_{\alpha,K}}{\|p_f\|_{\alpha+1,K_t}} \leq C\beta(t).$$

$\square$

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 groups G. Such a structures provide most of the examples in which normal Pontryagin extremal can be computed explicitly in terms of elementary functions.

We introduce a Lie groups as a sub-group 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 are always constant rank and, if they are of rank k, they can be generated by exactly k linearly independent vector fields defined globally. On such a structure we have always global existence of minimizers.

We then discuss Hamiltonian systems on Lie groups with left-invariant Hamiltonians. Such Hamiltonian systems are particularly simple since their tangent and cotangent bundles are always trivial. They have always a certain number of constant of the motion that for systems on a Lie group of dimension 3 are sufficient for the complete integrability.

We study in details some classes of systems in which one can obtain the explicit expression of normal Pontryagin extremals.

7.1 Sub-groups of Diff(M) generated by a finite dimensional Lie algebra of vector fields

Let M be a 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

$$\mathcal{G} := \{e^{X_1} \circ \ldots \circ e^{X_k} \mid k \in \mathbb{N}, X_1, \ldots, X_k \in L\} \subset \mathrm{Diff}(M), \tag{7.1}$$

that 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 holds true:*

(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 all $P' \in \Phi_P(U)$ there exists $V \subset U$ neighborhood of 0 such that $\Phi_{P'}(V) \subset \Phi_P(U)$.

Thanks to the previous result, one can introduce the following basis of neighborhoods[1] on $\mathcal{G}$:

$$\mathcal{B} = \{\Phi_P(W) \mid P \in \mathcal{G}, W \subset U, 0 \in W\}. \tag{7.2}$$

where U is determined as in (i) of Proposition 7.2. Part (ii) of Proposition 7.2 ensures that (7.2) satisfies the axioms of a basis for generates a unique topology on $\mathcal{G}$. Indeed it is sufficient to apply it twice to show that if $\Phi_P(W) \cap \Phi_{P'}(W') \neq \emptyset$ then there exists $Q \in \Phi_P(W) \cap \Phi_{P'}(W')$ and $V \subset U$ with $0 \in V$ such that $\Phi_Q(V) \subset \Phi_P(W) \cap \Phi_{P'}(W')$.

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 $\Phi_P^{-1} \cap \Phi_{P'} : W \cap W' \to W \cap W'$ is smooth with respect to the smooth structure defined on the vector space L (cf. Exercice 7.10). Hence $\mathcal{G}$ has the structure of smooth manifold.

7.1.1 Proof of Proposition 7.2

To prove this theorem we use a reduction to a finite dimensional setting, by evaluating elements of G, that are diffeomorphisms of M, on a special set of ℓ points, where ℓ is the dimension of L.

To identify this set of points, we first need a general lemma.

Lemma 7.3. *For every $k \in \mathbb{N}$ and $F_1, \ldots, F_k : \mathbb{R}^m \to \mathbb{R}^n$ 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 $(\mathbb{R}^n)^k = \mathbb{R}^n \times \ldots \times \mathbb{R}^n$.

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 $x_1, \ldots, x_k$ to be the set of

[1] Recall that a collection $\mathcal{B}$ of subset 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 all $B_1, B_2 \in \mathcal{B}$ with $B_1 \cap B_2 \neq \emptyset$ there exists nonempty $B_3 \in \mathcal{B}$ such that $B_3 \subset B_1 \cap B_2$.

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 equation (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. $\square$

As an immediate consequence of the previous lemma one obtains the following property.

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*

$$(X_i(q_1), \ldots, X_i(q_\ell)), \qquad i = 1, \ldots, \ell,$$

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 ϕ_* is equal to ℓ. Computing the partial derivatives at $0 \in L$ of ϕ in the directions $X_1, \ldots, X_\ell$ we have

$$\frac{\partial \phi}{\partial X_i}(0) = \left.\frac{d}{dt}\right|_{t=0} (e^{tX_i}(q_1), \ldots, e^{tX_i}(q_\ell)) = (X_i(q_1), \ldots, X_i(q_\ell)), \qquad i = 1, \ldots, \ell,$$

and these are linearly independent as elements of $T_{q_1}M \times \ldots \times T_{q_\ell}M$ by Lemma 7.4. $\square$

[2]here $M^\ell = \underbrace{M \times \ldots \times M}_{\ell \text{ times}}$.

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:

$$[(X_1, \ldots, X_k), (Y_1, \ldots, Y_k)] = ([X_1, Y_1], \ldots, [X_k, Y_k]).$$

Lemma 7.6. *For every $k \in \mathbb{N}$ the map $i : L \to \mathrm{Vec}(M)^k$ defined by $i(X) = (X, \ldots, X)$ defines an involutive distribution on M^k.*

Proof. It follows from the identity $[i(X), i(Y)] = i([X, Y])$, since

$$[(X, \ldots, X), (Y, \ldots, Y)] = ([X, Y], \ldots, [X, Y]). \qquad \square$$

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 (6.25) 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. $\square$

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 ϕ_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^ℓ 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 ℓ) has the expression

$$N = \{(\hat{P}(q_1), \ldots, \hat{P}(q_\ell)) \mid \hat{P} = e^{X_1} \circ \ldots \circ^{X_k}, k \in \mathbb{N}, X_1, \ldots, X_k \in L\},$$

while for each $P \in \mathcal{G}$,

$$\phi_P(U) = \{(P \circ e^X(q_1), \dots, P \circ e^X(q_\ell)) \mid P \in \mathcal{G}, X \in U \subset L\},$$

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 ℓ that is tangent to L thanks to Lemma 7.7, and passes through the point $\phi_P(0) = (P(q_1), \dots, P(q_\ell)) \in M^\ell$.

□

Remark 7.9. The previous result implies that for every $(q_1', \dots, q_\ell') \in \phi_P(U) \cap \phi_{P'}(U)$ there exists *uniques* $X, X' \in U$ such that

$$(P \circ e^X(q_1), \dots, P \circ e^X(q_\ell)) = (P' \circ e^{X'}(q_1), \dots, P' \circ e^{X'}(q_\ell)) = (q_1', \dots, q_\ell'). \tag{7.6}$$

In other words we are saying that the two diffeomorphisms $P \circ e^X$ and $P' \circ e^{X'}$ coincides when evaluated on the set of points $\{q_1, \dots, q_\ell\}$.

Exercise 7.10. Prove that the maps that associates $X \mapsto X'$ defined in (7.6) is smooth.

The argument that is developed in the next section shows that actually, not only one has the identity (7.6), but also $P \circ e^X = P' \circ e^{X'}$ 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".

Fix $q \in M$. Let us introduce

$$\overline{\phi} : U \to M^{\ell+1}, \qquad \overline{\phi}(X) = (e^X(q), e^X(q_1), \dots, e^X(q_\ell)) \in M^{\ell+1}.$$

Moreover, we define for every $P \in \mathcal{G}$

$$\overline{\phi}_P : U \to M^{\ell+1}, \qquad \overline{\phi}_P(X) = (P \circ e^X(q), P \circ e^X(q_1), \dots, P \circ e^X(q_\ell)) \in M^{\ell+1}.$$

The following Proposition can be proved following the same arguments as the one of Proposition 7.8.

Proposition 7.11. *Fix $q \in M$. For all $P \in \mathcal{G}$ we have that $\overline{\phi}_P(U)$ is an integral manifold of dimension ℓ in $M^{\ell+1}$ of a foliation defined by L (seen as distribution in* $\mathrm{Vec}(M)^{\ell+1}$*) and passing through the point $(P(q), P(q_1), \dots, P(q_\ell)) \in M^{\ell+1}$. 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 $\pi(q_0, q_1, \dots, q_\ell) = (q_1, \dots, q_\ell)$ 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).

We can now prove the main result.

Proof of Proposition 7.2. (i). It is enough to show that Φ_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

$$\phi_P(X) = (P \circ e^X(q_1), \dots, P \circ e^X(q_\ell)) = (P \circ e^Y(q_1), \dots, P \circ e^Y(q_\ell)) = \phi_P(Y)$$

hence by invertibility of ϕ_P on U we have that $X = Y$.

(ii). Recall that, by construction, one has the following relation between Φ_P and its finite-dimensional representation ϕ_P

$$\phi_P(W) = \{(Q(q_1), \dots, Q(q_\ell)) : Q \in \Phi_P(W)\}, \qquad W \subset U.$$

For every $V \subset U$, with $0 \in V$, one has that $\phi_{P'}(V)$ and $\phi_P(U)$ are integral submanifold of M^ℓ belonging to the same leaf of the foliation, thanks to Proposition 7.8.

Since by assumption $P' \in \Phi_P(U)$, it follows that the intersection $\phi_{P'}(V) \cap \phi_P(U)$ is open and non empty in M^ℓ and contains the point $(P'(q_1), \dots, P'(q_\ell))$. We can then choose V small enough such that $\phi_{P'}(V) \subset \phi_P(U)$.

This inclusion of the finite-dimensional images implies the following: for every $X' \in V$ there exists a unique element $X \in U$ such that $P' \circ e^{X'} = P \circ e^X$ when evaluated on the special set of points, namely

$$(P \circ e^X(q_1), \dots, P \circ e^X(q_\ell)) = (P' \circ e^{X'}(q_1), \dots, P' \circ e^{X'}(q_\ell)). \tag{7.8}$$

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 firs prove that, for V defined as before, one has $\overline{\phi}_{P'}(V) \subset \overline{\phi}_P(U)$ (indepedently on q). Assume that $\overline{\phi}_P(U) \setminus \overline{\phi}_{P'}(V) \neq \emptyset$, then we have

$$\begin{aligned} \pi(\overline{\phi}_{P'}(V)) &= \pi(\overline{\phi}_{P'}(V) \cap \overline{\phi}_P(U)) \cup \pi(\overline{\phi}_P(U) \setminus \overline{\phi}_{P'}(V)) & (7.9)\\ &= \phi_{P'}(V) \cup \pi(\overline{\phi}_P(U) \setminus \overline{\phi}_{P'}(V)) & (7.10) \end{aligned}$$

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 $X' \in V \cup W$ there exists a unique element $\widehat{X} \in U$ (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 ϕ_P on 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 multiplication 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.

Definition 7.15. The *Lie algebra associated with a Lie group* G is the Lie algebra $\mathfrak{g}$ of its left-invariant vector fields.

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 (Third 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 product $G_1 \times G_2$. Build the group $G \subset G_1 \times G_2$ associated with this 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 group 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 $\mathrm{Diff}(M)$ 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 Lie group.*

We now want to prove a converse statement for connected group, i.e., every connetected 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 a connected Lie group and L the Lie algebra of left invariant vector fields. Then G is isomorphic to $\mathcal{G}^{G,L}$.*

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 low given by the standard composition.

Lemma 7.22. *The flow of a left invariant vector fields on a Lie group G commutes with left translations.*

Proof. If ϕ is a diffeomorphism and X a vector field we have that (see Lemma 2.20)

$$e^{t\phi_* X} = \phi \circ e^{tX} \circ \phi^{-1}.$$

Composing on the right with ϕ, we have

$$e^{t\phi_* X} \circ \phi = \phi \circ e^{tX}.$$

Now taking $\phi = L_g$ for some g, 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. $\square$

A similar statement holds for right invariant vector fields.

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 $h \in H$, we have

$$P(h) = P(L_h \mathbb{1}) = L_h P(\mathbb{1}) = L_h g = hg$$

hence $P = R_g$. $\square$

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 $\mathcal{G}^{G,L}$ is isomorphic to G_R. Indeed we are going to prove that $\mathcal{G}^{G,L} = G_R$.

To prove that $\mathcal{G}^{G,L} \subseteq G_R$ observe that every element of $\mathcal{G}^{G,L}$ is a composition of the flow of left invariant vector fields and hence it is a right translation.

To prove that $\mathcal{G}^{G,L} = G_R$, observe that by the argument above $\mathcal{G}^{G,L}$ is a subgroup of G_R. Moreover since $\dim(\mathcal{G}^{G,L}) = \dim(G_R)$. It follows that $\mathcal{G}^{G,L}$ 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. $\square$

$\square$

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 $gl(n) = \{M \in \mathbb{R}^{n \times n}\}$ (resp. $gl(n, \mathbb{C}) = \{M \in \mathbb{C}^{n \times n}\}$).

Definition 7.27. A group of matrices is a sub group of $GL(n)$ or of $GL(n, \mathbb{C})$.

Remark 7.28. The Lie algebra of a sub-group of $GL(n)$ (resp. $GL(n, \mathbb{C})$) is a sub-algebra of $gl(n)$ (resp. $gl(n, \mathbb{C})$).

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 $sl(n) = \{M \in \mathbb{R}^{n \times n} \mid \mathrm{trace}(M) = 0\}$.

- 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 $so(n) = \{M \in \mathbb{R}^{n \times n} \mid M = -M^T\}$. $SO(n)$ is the connected component of $O(n)$ to the identity.

- The special unitary group

  $$SU(n) = \{M \in \mathbb{C}^{n \times n} \mid MM^\dagger = \mathbb{1}\},$$

  where $M^\dagger$ is the transpose of the complex conjugate of M. The Lie algebra of $SU(n)$ is $su(n) = \{M \in \mathbb{C}^{n \times n} \mid M = -M^\dagger\}$.

- The group of (positively oriented) Euclidean transformations of $\mathbb{R}^n$

  $$SE(n) = \left\{ \left(\begin{array}{c|c} & a_1 \\ R & \vdots \\ & a_n \\ \hline 0 & 1 \end{array} \right) \mid R \in SO(n), \quad a_1, \ldots, a_n \in \mathbb{R} \right\}.$$

  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

$$se(n) = \left\{ \left(\begin{array}{c|c} M & \begin{matrix} b_1 \\ \vdots \\ b_n \end{matrix} \\ \hline 0 & 0 \end{array} \right) \mid M \in so(n), \quad b_1, \dots, b_n \in \mathbb{R} \right\}.$$

Exercise 7.29. Prove that $o(3)$ and $su(2)$ 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 $g(t) = g_0 e^{tA}$ 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 $X_{ij} = \sum_k g_{ik}A_{kj}$ and $Y_{ij} = \sum_k g_{ik}B_{kj}$, $[X, Y]_{ij} = \sum_{kl} \left(\frac{\partial Y_{ij}}{\partial g_{kl}} X_{kl} - \frac{\partial X_{ij}}{\partial g_{kl}} Y_{kl} \right)$.)

(ii). Prove that for 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 fields on a group of matrices it is often convenient to use the abuse of notation $X(g) = gX$. This formula clarify well 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

- 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 to 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

Recall that a *pseudo-Riemannian metric* is a family of non-degenerate, symmetric metric bilinear form 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 metric that respects 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

$$\langle [X, Y] \,|\, Z \rangle = \langle X \,|\, [Y, Z] \rangle\,, \qquad \forall\, X, Y, Z \in \mathfrak{g}. \tag{7.15}$$

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 [16].

Next we define the natural 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

$$\operatorname{Ad}_g(X) = gXg^{-1}, \quad X \in T_{\mathbb{1}}G.$$

Recall that, given $x \in \mathfrak{g}$, its adjoint representation $\operatorname{ad} x : \mathfrak{g} \to \mathfrak{g}$ is given by $\operatorname{ad} x(y) = [x, y]$.

Definition 7.37. The *Killing form* on a Lie algebra $\mathfrak{g}$ is the symmetric bilinear form

$$K : \mathfrak{g} \times \mathfrak{g} \to \mathbb{R}, \qquad K(x, y) = \operatorname{trace}(\operatorname{ad} x \circ \operatorname{ad} y) \tag{7.16}$$

Exercise 7.38. Prove that the Killing form has the associativity property

$$K([x, y], z) = K(x, [y, z]). \tag{7.17}$$

Exercise 7.39. Prove that the Killing form of a nilpotent Lie algebra is identically zero.

Definition 7.40. A Lie algebra is said to be *semisimple* if the Killing form is non-degenerate.

Exercise 7.41. Prove that for semisimple Lie algebras, the Killing form is a bi-invariant pseudo-metric. Prove that for compact semisimple Lie algebras the Killing form is negative definite.

From the algebraic viewpoint a semisimple Lie algebra can be equivalently defined as a a Lie algebra $\mathfrak{g}$ satisfying $[\mathfrak{g}, \mathfrak{g}] = \mathfrak{g}$. See for instance [16].

7.2.4 The Levi-Malcev decomposition

A very important result in the theory of Lie algebras (see for instance [43, Ch. 4, Sect. 4, Thm. 4]) states that every Lie algebra can be decomposed as

$$\mathfrak{g} = \mathfrak{r} \ni \mathfrak{s}, \tag{7.18}$$

where

- $\mathfrak{r}$ is the so called *radical*, i.e., the maximal solvable ideal of $\mathfrak{g}$. A solvable Lie algebra is defined in the following way. 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 $\mathfrak{l}^0 = \mathfrak{l}$, $\mathfrak{l}^{(1)} = [\mathfrak{l}^{(0)}, \mathfrak{l}^{(0)}]$, ..., $\mathfrak{l}^{(n+1)} = [\mathfrak{l}^{(n)}, \mathfrak{l}^{(n)}]$. The Lie algebra $\mathfrak{l}$ is said to be solvable if there exists n such that $\mathfrak{l}^{(n)} = 0$.

- $\mathfrak{s}$ 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 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 7.42. 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 groups that one obtains taking as manifold $G_1 \times G_2$ 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 7.43. *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 consider 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. □

We have then an isomorphism between TG and $G \times T_{\mathbb{1}}G$. This isomorphism is 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 $G \times T_{\mathbb{1}}G$ extend to the dual. Hence T^*G is isomorphic to $G \times T^*_{\mathbb{1}}G$, the isomorphism being given by L^*_g, i.e.

$$T^*G \ni (g, p) \mapsto (g, \xi) \in G \times T^*_{\mathbb{1}}G,$$

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 $(G, \mathcal{D}, \langle \cdot \mid \cdot \rangle)$ (cf. Section 3.1.3, Example 2) where

- G is a 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 Lie bracket generating or equivalently that the smallest Lie sub-algebra of $\mathfrak{g}$ containing $\mathcal{D}$ is $\mathfrak{g}$ itself;
- $\langle \cdot \mid \cdot \rangle$ 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 $\langle v \mid w \rangle_g = \langle \nu \mid \mu \rangle_{\mathbb{1}}$.

Remark 7.44. 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 3.20).

Given a left-invariant sub-Riemannian structure we can always find m linearly independent vectors $e_1, \dots, e_m$ in $T_{\mathbb{1}}G$ such that

(i) $\mathcal{D}(g) = \{\sum_{i=1}^m u_i L_{g*}e_i \mid u_1, \dots u_m \in \mathbb{R}\}$

(ii) $\langle e_i \mid e_j \rangle_{\mathbb{1}} = \delta_{ij}$.

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

$$\begin{cases} \dot\gamma(t) = \sum_{i=1}^m u_i(t) L_{g*}e_i \\ \int_0^T \sqrt{\sum_{i=1}^m u_i(t)^2}\, dt \to \min \\ \gamma(0) = g_1, \quad \gamma(T) = g_2, \end{cases} \tag{7.19}$$

Exercise 7.45. (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

$$L_g(B(h, r)) = B(gh, r).$$

Existence of minimizers

Proposition 3.44 immediately implies the following.

Corollary 7.46. *Any left-invariant sub-Riemannian structure on a Lie group G is complete.*

Proof. By Proposition 3.35 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 7.45) $\overline{B}(g, \varepsilon) = L_g(\overline{B}(\mathbb{1}, \varepsilon))$ is compact for every $g \in G$, independently on ε. By Proposition 3.44, the sub-Riemannian structure is complete. $\square$

7.5 Carnot groups of step 2

The Heisenberg sub-Riemannian structure 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 7.47. Prove that the Lie algebra of the Heisenberg group can be written as $\mathfrak{g} = \mathfrak{g}_1 \oplus \mathfrak{g}_2$ where

$$\mathfrak{g}_1 = \operatorname{span}\{\partial_x - \frac{y}{2}\partial_z, \partial_y + \frac{x}{2}\partial_z\}, \quad \text{and} \quad \mathfrak{g}_2 = \operatorname{span}\{\partial_z\}.$$

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 generalization of the Heisenberg group, namely Lie groups G on $\mathbb{R}^n$ such that its Lie algebra $\mathfrak{g}$ satisfies

$$\mathfrak{g} = \mathfrak{g}_1 \oplus \mathfrak{g}_2, \qquad [\mathfrak{g}_1, \mathfrak{g}_1] = \mathfrak{g}_2, \qquad [\mathfrak{g}_1, \mathfrak{g}_2] = [\mathfrak{g}_2, \mathfrak{g}_2] = 0. \tag{7.20}$$

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 (7.20). Notice that $\mathfrak{g}$ is a nilpotent Lie algebra and that we have the inequality

$$n \leq \frac{k(k+1)}{2}, \qquad k = \dim \mathfrak{g}_1,\ n = \dim \mathfrak{g}.$$

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

$$\mathfrak{g}_1 = \operatorname{span}\{X_1, \ldots, X_k\}, \qquad \mathfrak{g}_2 = \operatorname{span}\{Y_1, \ldots, Y_{n-k}\},$$

where $\{X_1, \ldots, X_k\}$ 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:

$$\begin{cases} [X_i, X_j] = \sum_{h=1}^{n-k} c_{ij}^h Y_h, & i, j = 1, \ldots, k, \quad \text{where} \quad c_{ij}^h = -c_{ji}^h, \\ [X_i, Y_j] = [Y_j, Y_h] = 0, & i = 1, \ldots, k, \qquad j, h = 1, \ldots, n-k. \end{cases} \tag{7.21}$$

Define the the $n - k$ skew-symmetric matrices $C_h = (c_{ij}^h)$, for $h = 1, \ldots, n-k$. We stress that since the vector fields are left-invariant, then the structure functions c_{ij}^h are constant.

Given an adapted basis, we can associate with the family of matrices $\{C_1, \ldots, C_{n-k}\}$ the subspace

$$\mathcal{C} = \operatorname{span}\{C_1, \ldots, C_{n-k}\} \subset so(\mathfrak{g}_1) \tag{7.22}$$

of skew-symmetric operators on $\mathfrak{g}_1$ that are represented by linear combination of this family of matrices.

Proposition 7.48 (2-step Carnot algebras and subspaces of $so(\mathfrak{g}_1)$)**.** *For a given a 2-step Carnot algebra $\mathfrak{g}$, the subspace $\mathcal{C} \subset so(\mathfrak{g}_1)$ is independent on the choice of the adapted basis on $\mathfrak{g}$*

Proof. Assume that we fix another adapted basis

$$\mathfrak{g}_1 = \mathrm{span}\{X'_1, \dots, X'_k\}, \qquad \mathfrak{g}_2 = \mathrm{span}\{Y'_1, \dots, Y'_{n-k}\}.$$

where $\{X'_1, \dots, X'_k\}$ is orthonormal for the inner prodict. Then there exists $A = (a_{ij})$ an orthogonal matrix and $B = (b_{hl})$ an invertible matrix such that

$$X'_i = \sum_{j=1}^{k} a_{ij} X_j, \qquad Y'_h = \sum_{l=1}^{n-k} b_{hl} Y_l.$$

A direct computation shows that, denoting $B^{-1} = (b^{hl})$, we have

$$[X'_i, X'_j] = \sum_{h,l=1}^{k} a_{ih} a_{jl} [X_h, X_l] = \sum_{h,l=1}^{k} a_{ih} a_{jl} \sum_{r=1}^{n-k} c^r_{hl} Y_r \tag{7.23}$$

$$= \sum_{s=1}^{n-k} \left(\sum_{r=1}^{n-k} \sum_{h,l=1}^{k} a_{ih} a_{jl} c^r_{hl} b^{rs} \right) Y'_s \tag{7.24}$$

it follows that

$$C'_s = \sum_{h=1}^{n-k} b^{hs} (A C_h A^*) \tag{7.25}$$

Recall that two matrices C and C' represents the same element of $so(\mathfrak{g}_1)$ with respect to the two basis if and only if $C' = ACA^*$. Then formula (7.25) implies that elements of $\mathcal{C}'$ are written as linear combination of elements of $\mathcal{C}$ that represents the same linear operator, as claimed. ☐

Remark 7.49. We have the following basis-independent interpretation of Proposition 7.48. 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 $so(\mathfrak{g}_1)$, 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

$$\Psi : \mathfrak{g}_2^* \to so(\mathfrak{g}_1)$$

The subspace $\mathcal{C}$ introduced in (7.22) coincides with $\mathrm{im}\,\Psi \subset so(\mathfrak{g}_1)$.

Definition 7.50. 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 7.51. *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 $so(\mathfrak{g}_1)$.*

7.5.1 Pontryagin extremals for 2-step Carnot groups

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 $\mathbb{R}^n = \mathbb{R}^k \oplus \mathbb{R}^{n-k}$ (using coordinates $g = (x, z) \in \mathbb{R}^k \oplus \mathbb{R}^{n-k}$) and satisfying (13.11) is given by

$$X_i = \frac{\partial}{\partial x_i} - \frac{1}{2}\sum_{j=1}^{k}\sum_{\ell=1}^{n-k} c_{ij}^{\ell} x_j \frac{\partial}{\partial z_\ell}, \qquad i = 1, \ldots, k, \tag{7.26}$$

$$Z_\ell = \frac{\partial}{\partial z_\ell}, \qquad \ell = 1, \ldots, n-k. \tag{7.27}$$

The group G is $\mathbb{R}^n = \mathbb{R}^k \oplus \mathbb{R}^{n-k}$ 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 $(n-k)$-tuple $C = (C_1, \ldots, C_{n-k})$ of $k \times k$ matrices, the product

$$Cx \cdot x' = (C_1 x \cdot x', \ldots, C_{n-k} x \cdot x') \in \mathbb{R}^{n-k}.$$

and $x \cdot x'$ denotes the Euclidean inner product in $\mathbb{R}^k$.

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 $\{X_1, \ldots, X_k, Z_1, \ldots, Z_{n-k}\}$ are linearly independent, the functions (h_i, w_ℓ) defines a system of coordinates on fibers of T^*G. In what follows it is convenient to use (x, y, h, w) as coordinates on T^*G.

Geodesics are projections of integral curves of the sub-Riemannian Hamiltonian in T^*G

$$H = \frac{1}{2}\sum_{i=1}^{k} h_i^2$$

Suppose now that $\lambda(t) = (x(t), y(t), h(t), \omega(t))$ is a normal Pontryagin extremal. Then $u_i(t) = h_i(\lambda(t))$ and the equation on the base is

$$\dot{g} = \sum_{i=1}^{k} h_i X_i(g). \tag{7.28}$$

that rewrites as

$$\begin{cases} \dot{x}_i = h_i \\ \dot{z}_h = -\frac{1}{2}\sum_{i,j=1}^{k} c_{ij}^{\ell} h_i x_j \end{cases} \tag{7.29}$$

For the equations on the fiber we have (remember that along solutions $\dot{a} = \{H, a\}$)

$$\begin{cases} \dot{h}_i = \{H, h_i\} = -\sum_{j=1}^{k}\{h_i, h_j\}h_j = -\sum_{\ell=1}^{n-k}\sum_{j=1}^{k} c_{ij}^{\ell} h_j w_\ell \\ \dot{w}_\ell = \{H, w_\ell\} = 0. \end{cases} \tag{7.30}$$

H is constant along solutions and if we require that extremals are parametrized by arclength. From (7.30) we easily get that ω_h is constant and the vector $h = (h_1, \dots, h_k) \in \mathbb{R}^k$ satisfies the linear equation

$$\dot{h} = -\Omega_w h, \qquad \Omega_w = \sum_{\ell=1}^{n-k} w_\ell C_\ell$$

where we recall that the vector $w = (w_1, \dots, w_{n-k})$ 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 7.52. *The projection $x(t)$ on the layer $\mathfrak{g}_1 \simeq \mathbb{R}^k$ of a Pontryagin extremal such that $x(0) = 0$ is the image of the origin through a one-parametric group of isometries of $\mathbb{R}^k$.*

Proof. The action of a 1-parametric group of isometries can be recovered by exponentiating an element of its Lie algebra (cf. Exercice 7.53). This reduces to compute the solution of the differential equation

$$\dot{x} = Ax + b$$

where A is skew-symmetric and $b \in \mathbb{R}^k$. 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 $\mathfrak{g}_1 \simeq \mathbb{R}^k$ of a Pontryagin extremal satisfies this equation with $\bar{x} = x(0) = 0$, $A = -\Omega_w$ and $b = h(0)$. $\square$

Exercise 7.53. (i). Show that the group of (positively oriented) affine isometries on $\mathbb{R}^n$ can be identified with the matrix group

$$SE(n) = \left\{ \begin{pmatrix} M & c \\ 0 & 1 \end{pmatrix}, M \in SO(n), c \in \mathbb{R}^n \right\},$$

through the identification of an element $x \in \mathbb{R}^n$ with the vector $\begin{pmatrix} x \\ 1 \end{pmatrix}$ in $\mathbb{R}^{n+1}$.

(ii). Prove that the Lie algebra of $SE(n)$ is given by

$$\mathfrak{se}(n) = \left\{ \begin{pmatrix} A & b \\ 0 & 0 \end{pmatrix}, A \in so(n), b \in \mathbb{R}^n \right\}.$$

(iii). Prove the following formula for the exponential of an element of the Lie algebra

$$\exp\left(t \begin{pmatrix} A & b \\ 0 & 0 \end{pmatrix} \right) = \begin{pmatrix} e^{tA} & \int_0^t e^{sA} b ds \\ 0 & 1 \end{pmatrix}.$$

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 get

$$h(t) = \begin{pmatrix} \cos(wt) & -\sin(wt) \\ \sin(wt) & \cos(wt) \end{pmatrix}\begin{pmatrix} -\sin\theta \\ \cos\theta \end{pmatrix} = \begin{pmatrix} -\sin(wt+\theta) \\ \cos(wt+\theta) \end{pmatrix}$$

$$x(t) = \frac{1}{w}\begin{pmatrix} \sin(wt) & \cos(wt) - 1 \\ -\cos(wt) + 1 & \sin(wt) \end{pmatrix}\begin{pmatrix} -\sin\theta \\ \cos\theta \end{pmatrix} = \frac{1}{w}\begin{pmatrix} \cos(wt+\theta) - \cos\theta \\ \sin(wt+\theta) - \sin\theta \end{pmatrix}$$

This recovers the formulas already computed in Section 4.4.3. Notice that the z component is recovered simply by integrating the last equation, that in this case gives

$$\dot{z} = \frac{1}{2}(-h_1x_2 + h_2x_1)$$

$$\begin{aligned} z(t) &= \frac{1}{2w}\int_0^t \sin(ws+\theta)(\sin(ws+\theta) - \sin\theta) + \cos(ws+\theta)(\cos(ws+\theta) - \cos\theta)ds \\ &= \frac{1}{2w}\int_0^t 1 - \sin(ws+\theta)\sin\theta - \cos(ws+\theta)\cos\theta ds = \frac{1}{2w}\int_0^t 1 - \cos(ws)ds \\ &= \frac{1}{2w^2}(wt - \sin(wt)). \end{aligned}$$

Analogous computation are performed for higher dimensional Heisenberg groups in Section 13.1.

7.6 Left-invariant Hamiltonian systems on Lie groups

In this section we study Hamiltonian systems non necessarily coming from a sub-Riemnnian problem.

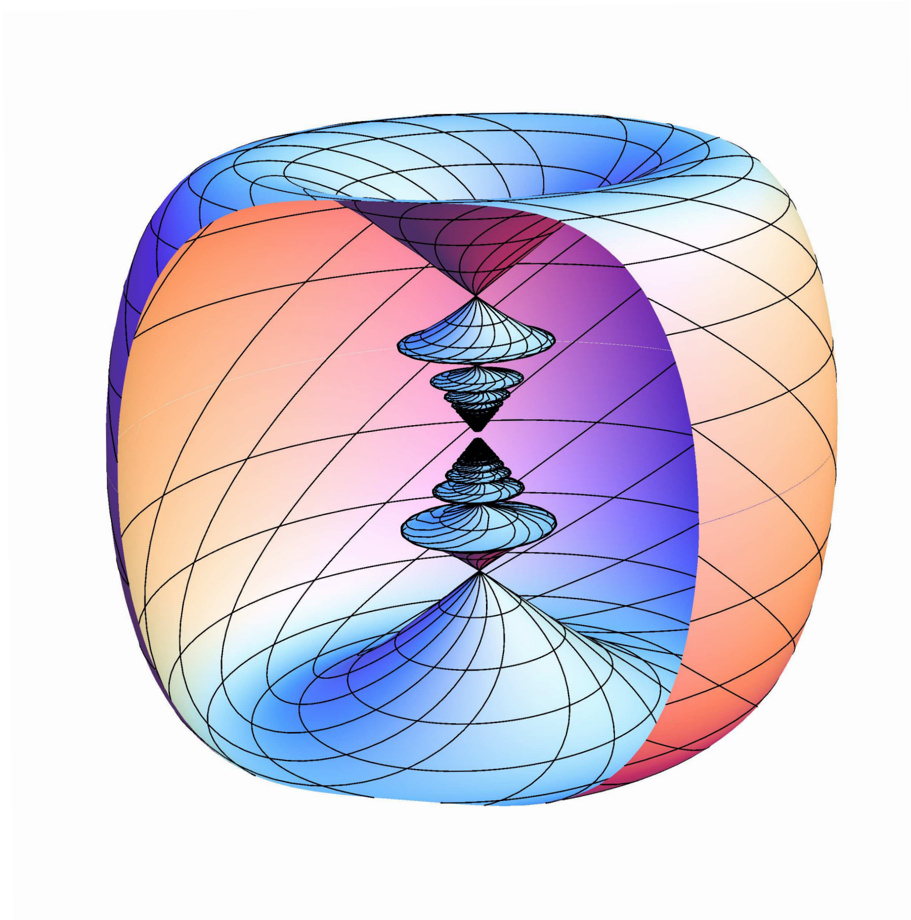


Figure 7.1: The set of end points of length 1 Pontryagin extremals for the 3D Heisenberg group. Notice the singularities accumulating at the origin.

7.6.1 Vertical coordinates in TG and T^*G

Thanks to the isomorphism between TG and $G \times T_{\mathbb{1}}G$, a bases $\{e_1, \ldots, e_n\}$ of $T_{\mathbb{1}}G$ induces global coordinates on TG. Indeed a base of T_gG is $L_{g*}e_1, \ldots, L_{g*}e_n$ and every element (g, v) of TG can be written as

$$(g, v) = (g, \sum_{i=1}^{n} v_i L_{g*} e_i).$$

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 $(g, v) = (g, \sum_{i=1}^{n} v^i L_{g*} e_i) \in TG$, then the corresponding point in $G \times T_{\mathbb{1}}G$ is $(g, \xi) = (g, \sum_{i=1}^{n} v^i e_i)$ hence, in coordinates, both are representedby $(g, v_1, \ldots, v_n)$.

If $\{e_1^*, \ldots, e_n^*\}$ is the dual base 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 (g, p) of T^*G can be written as

$$(g, p) = (g, \sum_{i=1}^{n} h_i L_{g^{-1}}^* e_i^*).$$

The coordinates $h_1, \dots h_n$ are called *vertical coordinates* in T^*G. For the same reason as above, in vertical coordinates $(g, h_1, \dots, h_n)$ 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 7.54. 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)

$$\{h_i, h_j\} = \langle p, [X_i, X_j]\rangle_g = \langle \xi, [e_i, e_j]\rangle_{\mathbb{1}}. \tag{7.31}$$

Remark 7.55. 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 (7.31) 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

$$\mathcal{H}(g, \xi) = H(g, L^*_{g^{-1}}\xi), \quad \mathcal{H} : G \times T^*_{\mathbb{1}}G \to \mathbb{R}.$$

We say that H is left-invariant if $\mathcal{H}(g, \xi)$ 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 7.56. 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

$$H(g, p) = \mathcal{H}(L^*_g p).$$

Hence a left invariant-Hamiltonian can be interpreted as a function on $T^*_{\mathbb{1}}G$.

Example 7.57. 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 $H(g, p) = \frac{1}{2}\sum_{i=1}^m \langle p, f_i(g)\rangle^2$ is a left-invariant Hamiltonian. Indeed

$$\mathcal{H}(g, \xi) = \frac{1}{2}\sum_{i=1}^m \langle L^*_{g^{-1}}\xi, L_{g*}w_i\rangle^2 = \frac{1}{2}\sum_{i=1}^m \langle \xi, w_i\rangle^2,$$

which is independent from g.

Remark 7.58. If we write $p = \sum_{j=1}^{n} h_j L_{g^{-1}}^* e_j^*$ then

$$H(g, \sum L_{g^{-1}}^* h_j e_j^*) = \mathcal{H}(L_g^* \sum h_j L_{g^{-1}}^* e_j^*) = \mathcal{H}(\sum h_j e_j^*).$$

In other words in vertical coordinates $h_1, \dots h_n$, we have for a left-invariant Hamiltonian

$$H(g, h_1, \dots, h_n) = \mathcal{H}(h_1, \dots, h_n).$$

and we can identify H and $\mathcal{H}$.

Remark 7.59. In the context of Lie groups, to write Hamiltonian equations is convenient avoiding 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 $H(g, h_1, \dots, h_n)$. Let us write the vertical part of the Hamiltonian equations. We are going to see that this equation is particularly simple. We have

$$\dot{h}_i = \{H, h_i\}, \qquad i = 1, \dots, n. \tag{7.32}$$

Using Exercice 4.8 we have for $i = 1, \dots, n$,

$$\dot{h}_i = \sum_{j=1}^{n} \frac{\partial H}{\partial h_j} \{h_j, h_i\} = \sum_{j=1}^{n} \frac{\partial H}{\partial h_j} \langle \xi, [e_j, e_i] \rangle = \left\langle \xi, \left[\sum_{j=1}^{n} \frac{\partial H}{\partial h_j} e_j, e_i \right] \right\rangle. \tag{7.33}$$

Notice that since $\mathcal{H}$ is a function on the linear space $T_{\mathbb{1}}^*G$, then $d\mathcal{H}(h_1, \dots, h_n)$ is an element of $T_{\mathbb{1}}^{**}G = T_{\mathbb{1}}G$. If we write an element of $T_{\mathbb{1}}^*G$ as $h_1 e_1^* + \dots + h_n e_n^*$, then an element of its tangent at $(h_1, \dots, h_n)$ is written as $v_1 \partial_{h_1}, \dots, v_n \partial_{h_n}$ with the identification $\partial_{h_i} = e_i^*$ due to the linear structure. An element of its cotangent space $T_{\mathbb{1}}^{**}G$ at $(h_1, \dots, h_n)$ is then written as $\omega_1 dh_1 + \dots + \omega_n dh_n$ with the identification $dh_i = (e_i^*)^* = e_i$ again due to the linear structure. Then

$$d\mathcal{H}(h_1, \dots, h_n) = \sum_{j=1}^{n} \frac{\partial \mathcal{H}}{\partial h_j} dh_j = \sum_{j=1}^{n} \frac{\partial \mathcal{H}}{\partial h_j} e_j = \sum_{j=1}^{n} \frac{\partial H}{\partial h_j} e_j. \tag{7.34}$$

Hence the vertical part of the Hamiltonian equations can be written as

$$\begin{aligned} \dot{h}_i &= \langle \xi, [d\mathcal{H}, e_i] \rangle \\ &= \langle \xi, (\operatorname{ad} d\mathcal{H}) e_i \rangle \\ &= \langle (\operatorname{ad} d\mathcal{H})^* \xi, e_i \rangle \end{aligned} \tag{7.35}$$

or more compactly recalling that $\xi = \sum_{i=1}^{k} h_i e_i^*$,

$$\dot{\xi} = (\operatorname{ad} d\mathcal{H})^* \xi. \tag{7.36}$$

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

$$\frac{d}{dt}\beta(g(t)) = \{H, \beta\}_{(g(t),p(t))} = \sum_{j=1}^{n} \frac{\partial H}{\partial h_j}\{h_j, \beta\}_{(g(t),p(t))}.$$

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

$$\frac{d}{dt}\beta(g(t)) = \sum_{j=1}^{n} \frac{\partial H}{\partial h_j}(L_{g*}e_j)\beta\bigg|_{g(t)} = \left(L_{g*}\sum_{j=1}^{n}\frac{\partial H}{\partial h_j}e_j\right)\beta\bigg|_{g(t)} = L_{g*}d\mathcal{H}|_{g(t)}.$$

Since the function β is arbitrary we have

$$\dot{g} = L_{g*}d\mathcal{H}.$$

We have then proved the following

Proposition 7.60. *Let H be a left invariant Hamiltonian on a Lie group G, i.e. $H(g,p) = \mathcal{H}(L_g^* p)$ where $(g,p) \in T^*G$ 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 $\frac{d}{dt}(g,p) = \vec{H}(g,p)$ are,*

$$\begin{cases} \dot{g} = L_{g*}d\mathcal{H} \\ \dot{\xi} = (\operatorname{ad} d\mathcal{H})^*\xi. \end{cases} \tag{7.37}$$

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 (it does not involve g).

When we have available a bi-invariant metric equation (7.36) can be written in a simpler form. Indeed in this case we can identify $T_{\mathbb{1}}G$ with $T_{\mathbb{1}}^*G$ via

$$\xi \in T_{\mathbb{1}}^*G \longleftrightarrow M \in T_{\mathbb{1}}G \iff \langle M \,|\, v\rangle = \langle \xi, v\rangle, \quad \forall v \in T_{\mathbb{1}}G.$$

Using (7.36) and (7.15), for every $v \in T_{\mathbb{1}}G$ let us compute

$$\left\langle \frac{dM}{dt} \,\middle|\, v \right\rangle = \left\langle \frac{d\xi}{dt}, v \right\rangle = \langle (\operatorname{ad} d\mathcal{H})^*\xi, v\rangle = \langle \xi, (\operatorname{ad} d\mathcal{H})v\rangle = \langle \xi, [d\mathcal{H}, v]\rangle = \langle M \,|\, [d\mathcal{H}, v]\rangle = \langle [M, d\mathcal{H}] \,|\, v\rangle.$$

Hence the Hamiltonian equations for a left-invariant Hamiltonian, when we have a bi-invariant peseudometric are:

$$\begin{cases} \dot{g} = L_{g*}d\mathcal{H} \\ \frac{dM}{dt} = [M, d\mathcal{H}]. \end{cases} \tag{7.38}$$

7.7 First integrals for Hamiltonian systems on Lie groups*

7.7.1 Integrability of left invariant sub-Riemannian structures on 3D Lie groups*

7.8 Normal Extremals for left-invariant sub-Riemannian structures

Consider a left-invariant sub-Riemannian structure of rank m (cf. (7.19)) 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

$$H(g,p) = \frac{1}{2}\sum_{i=1}^{m} \langle p, X_i(g)\rangle^2 = \frac{1}{2}\sum_{i=1}^{m} \langle p, L_{g*}e_i\rangle^2 ,$$

hence it is left invariant (cf. Example 7.57). The corresponding trivialized Hamiltonian is

$$\mathcal{H}(\xi) = \frac{1}{2}\sum_{i=1}^{m} \langle \xi, e_i\rangle^2 .$$

Now $\langle \xi, e_i\rangle = h_i(g,p)$ hence in vertical coordinates we have

$$H(h_1, \ldots, h_m) = \frac{1}{2}\sum_{i=1}^{m} h_i^2.$$

7.8.1 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 7.61. 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 square of its matrix elements.

Exercise 7.62. Prove that the distribution defined in Remark 7.61 is Lie 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 $y \in \mathbf{s}$. Let $e_1, \ldots e_m$ 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 $\langle \cdot \,|\, \cdot \rangle$ we have

$$h_i = \langle \xi, e_i \rangle = \langle M \,|\, e_i \rangle = x_i.$$

Hence

$$\mathcal{H} = \frac{1}{2}\sum_{i=1}^{n} h_i^2 = \frac{1}{2}\sum_{i=1}^{n} x_i^2 = \frac{1}{2}\langle x \,|\, x \rangle = \frac{1}{2}\|x\|^2. \tag{7.39}$$

Notice that (cf. (7.34)) $d\mathcal{H} = \sum_{i=1}^{n} \frac{\partial \mathcal{H}}{\partial h_i} e_i = \sum_{i=1}^{n} \frac{\partial \mathcal{H}}{\partial x_i} e_i = \sum_{i=1}^{n} x_i e_i = x$. Hence the vertical part of the Hamiltonian equation $dM/dt = [M, d\mathcal{H}]$ become

$$\dot{x} + \dot{y} = [x + y, x] = [y, x]. \tag{7.40}$$

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 (7.15) and for the last equality that facts 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 (7.40) become

$$\dot{x} = [y, x]$$
$$\dot{y} = 0$$

Hence all y component are constant of the motion and we have

$$y(t) = y_0$$
$$\dot{x} = [y_0, x] = (\operatorname{ad} y_0)x$$

The solution of the last equation is

$$x(t) = e^{t \operatorname{ad} y_0} x_0. \tag{7.41}$$

Then for the horizontal part we have

$$\dot{g} = L_{g*} d\mathcal{H} = L_{g*} x(t) = L_{g*} e^{t \operatorname{ad} y(0)} x(0). \tag{7.42}$$

Using the variation formula for smooth vector fields (cf. (6.35)),

$$e^{t(Y+X)} = \overrightarrow{\exp} \int_0^t e^{s \operatorname{ad} Y} X ds \circ e^{tY}, \tag{7.43}$$

we have that the solution of (7.42) starting from g_0 and corresponding to x_0, y_0 is [3]

$$g(x_0, y_0; t) = g_0 e^{t(x_0 + y_0)} e^{-t y_0} \tag{7.44}$$

[3]For a group of matrices: formula (7.41) reads as $e^{t y_0} x_0 e^{-t y_0}$, while (7.42) is $g e^{t y_0} x_0 e^{-t y_0}$.

The parameterization by arclength is obtained requiring $H = 1/2$. From (7.39) at $t = 0$ we obtain that the normal Pontryagin extremals (7.44) 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 (7.44) are

$$u_i(t) = \langle p(t), X_i(g(t))\rangle = h_i(g(t), p(t)) = x_i(t) = \left\langle e^{t\,\mathrm{ad}\, y_0} x_0 \,\middle|\, e_i \right\rangle, \quad i = 1, \dots, m.$$

Exercise 7.63. Study abnormal extremals for this problem.

7.8.2 Example: The $\mathbf{d} \oplus \mathbf{s}$ problem on $SO(3)$

The Lie group $SO(3)$ is the group of special orthogonal 3×3 real matrices

$$SO(3) = \left\{ g \in \mathrm{Mat}(3, \mathbb{R}) \mid gg^T = \mathrm{Id}, \det(g) = 1 \right\}.$$

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 $g(0) = e$. 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 $g(0) = -g^T(0)$. Hence the Lie algebra of $SO(3)$ is the space of skew symmetric 3×3 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

$$x_0 + y_0 = \underbrace{\cos(\theta)e_1 + \sin(\theta)e_2}_{x_0} + \underbrace{ce_3}_{y_0}, \quad \theta \in S^1, \quad c \in \mathbb{R}.$$

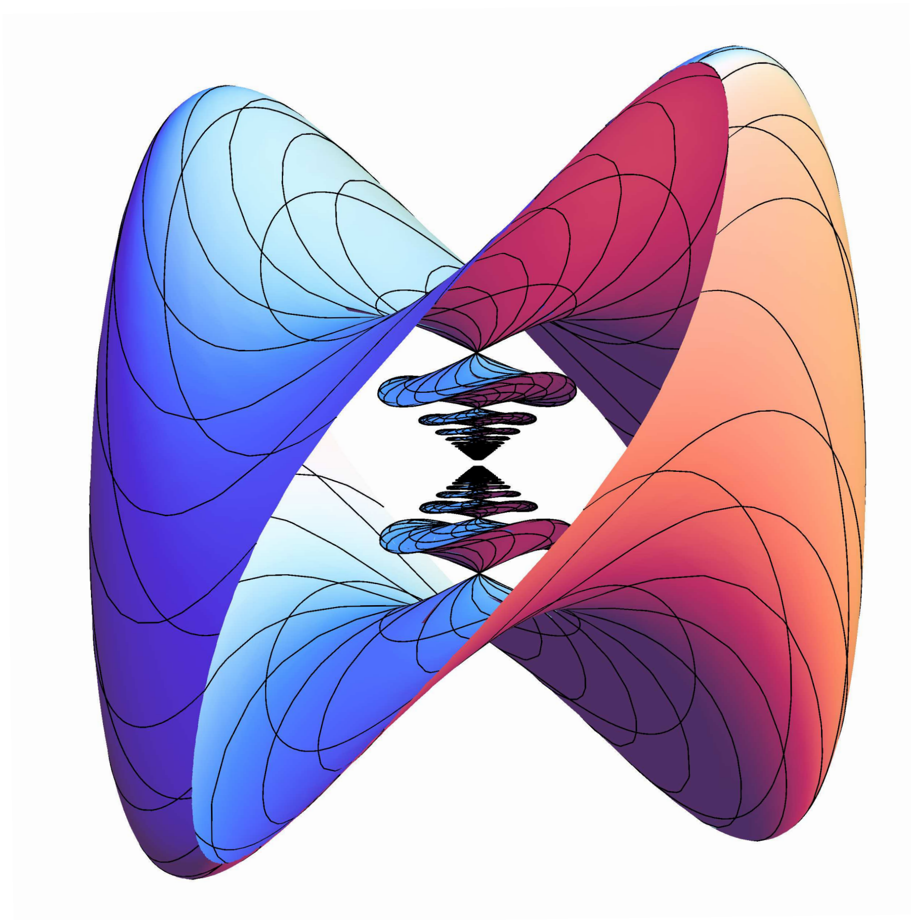


Figure 7.2: The set of end points of Pontryagin extremals of length 1 for the $\mathbf{d}\oplus\mathbf{s}$ sub-Riemannian problem on $SO(3)$. In the picture the x-axis is the element $(g)_{23}$, the z-axis is the element $(g)_{13}$, the z-axis is the element $(g)_{12}$. Notice the singularities accumulating 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 consider end points of geodesics of length r and makes r smaller and smaller. For r big the two pictures become very different due to the different topology of $\mathbb{R}^3$ and $SO(3)$.

Using formula (7.44), we have that the normal Pontryagin extremals starting from the identity are

$$g(\theta, c; t) := e^{(\cos(\theta)e_1+\sin(\theta)e_2+ce_3)t}e^{-ce_3t} = \tag{7.45}$$

$$= \begin{pmatrix} K_1\cos(ct)+K_2\cos(2\theta+ct)+K_3c\sin(ct) & K_1\sin(ct)+K_2\sin(2\theta+ct)-K_3c\cos(ct) & K_4\cos(\theta)+K_3\sin(\theta) \\ -K_1\sin(ct)+K_2\sin(2\theta+ct)+K_3c\cos(ct) & K_1\cos(ct)-K_2\cos(2\theta+ct)+K_3c\sin(ct) & -K_3\cos(\theta)+K_4\sin(\theta) \\ K_4\cos(\theta+ct)-K_3\sin(\theta+ct) & K_3\cos(\theta+ct)+K_4\sin(\theta+ct) & \frac{\cos\left(\sqrt{1+c^2}t\right)+c^2}{1+c^2} \end{pmatrix}$$

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.

7.8.3 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 1.22) 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 $gx_1 = x_2,\ gv_1 = v_2$. 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 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 7.64. Show that the differential of the map

$$SO(3) \to S^2, \qquad g \mapsto \left(g\left(\begin{smallmatrix}0\\0\\1\end{smallmatrix}\right), g\left(\begin{smallmatrix}1\\0\\0\end{smallmatrix}\right)\right) \tag{7.46}$$

transforms the left-invariant distribution $\mathbf{d}$ into the kernel of the Levi-Civita connection (cf. Definition 1.54) on $S\,S^2$.

Let ω 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 ω on $\mathbb{R} \times M$, where ω 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 ω 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 ω 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 $so(2,1)$ and it is isomorphic to $sl(2)$. We set $\langle X|Y\rangle = -\frac{1}{2}\mathrm{trace}(XY)$, a bi-invariant pseudo-metric on $so(2,1)$. 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 7.65. Show that the differential of the map

$$SO_+(2,1) \to H^2, \qquad g \mapsto \left(g \left(\begin{smallmatrix} 0 \\ 0 \\ 1 \end{smallmatrix} \right), g \left(\begin{smallmatrix} 1 \\ 0 \\ 0 \end{smallmatrix} \right) \right) \tag{7.47}$$

transforms the left-invariant distribution $\mathbf{d}$ into the kernel of the Levi-Civita connection on SH^2.

The transformation (7.47) 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.

7.8.4 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);

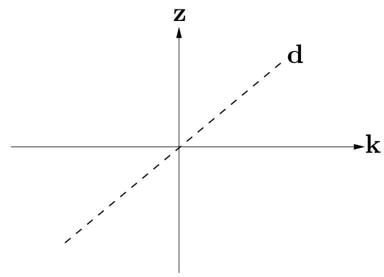


Figure 7.3: The $\mathbf{k} \oplus \mathbf{z}$ problem

- we assume that the distribution (that we assume to be Lie 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 $\langle x \,|\, y \rangle = 0$.

Exercise 7.66. 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 (7.38) 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 7.67. 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.

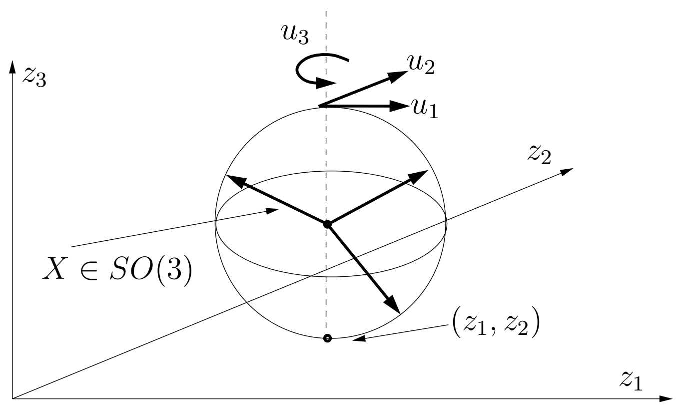


Figure 7.4: Rolling sphere with twisting.

Let us write the maximized Hamiltonian. Let $e_1, \dots, e_m$ be an orthonormal frame for $\mathbf{k}$. Then an orthonormal frame for $\mathbf{d}$ is $e_1 + Ae_1, \dots, e_m + Ae_m$. We have

$$H(g,p) = \frac{1}{2}\sum_{i=1}^{m} \langle p, L_{g*}(e_i + Ae_i)\rangle^2 .$$

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

$$\mathcal{H}(x,y) = \frac{1}{2}\sum_{i=1}^{m} \langle x + y \,|\, (e_i + Ae_i)\rangle_{T_{\mathbb{1}}G}^2 = \frac{1}{2}\sum_{i=1}^{m} (\langle x \,|\, e_i\rangle + \langle y \,|\, Ae_i\rangle)^2. \tag{7.48}$$

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

$$\mathcal{H}(x,y) = \frac{1}{2}\sum_{i=1}^{m} (\langle x \,|\, e_i\rangle + \langle A^*y \,|\, e_i\rangle_{\mathbf{k}})^2 = \frac{1}{2}\|x + A^*y\|_{\mathbf{k}}^2. \tag{7.49}$$

The vertical part of the Hamiltonian equations are (cf. the second equation of (7.38) with M replaced by x+y)

$$\dot{x} + \dot{y} = [x + y, d\mathcal{H}]. \tag{7.50}$$

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 (7.50) and we get

$$\dot{x} + \dot{y} = [x + y, x + A^*y] = [x, A^*y],$$

from which we deduce

$$\dot{x} = [x, A^* y],$$
$$\dot{y} = 0.$$

Hence all y components are constant of the motion and we have

$$y(t) = y_0$$
$$\dot{x} = [x, A^* y_0] = -[A^* y_0, x] = -(\mathrm{ad}\,(A^* y_0))x$$

The solution of the last equation is

$$x(t) = e^{-t\mathrm{ad}\,(A^* y_0)} x_0. \tag{7.51}$$

For the horizontal part of the Hamiltonian equations we have

$$\dot{g}(t) = L_{g(t)*} d\mathcal{H}(x(t), y(t)) = L_{g(t)*}(\underbrace{x(t) + A^* y_0}_{\in \mathbf{k}} + \underbrace{Ax(t) + AA^* y_0}_{\in \mathbf{z}}). \tag{7.52}$$

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 (7.52) splits in the following way

$$\dot{g}_1 = L_{g_1 *}(x(t) + A^* y_0) \tag{7.53}$$
$$\dot{g}_2 = Ax(t) + AA^* y_0 \tag{7.54}$$

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 (7.53) and (7.53) we have the initial conditions $g_1(0) = g_{01}$ and $g_2(0) = g_{02}$.

Let us solve (7.53). Using (7.51) this equation is reduced to

$$\dot{g}_1 = L_{g_1 *}(e^{-t\,\mathrm{ad}\,(A^* y_0)} x_0 + A^* y_0) = L_{g_1 *} e^{-t\,\mathrm{ad}\,(A^* y_0)}(x_0 + A^* y_0), \tag{7.55}$$

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. (6.35)),

$$e^{t(Y+X)} = \overrightarrow{\exp} \int_0^t e^{s\,\mathrm{ad}\,Y} X ds \circ e^{tY}, \tag{7.56}$$

with $Y \to -A^* y_0$ and $X \to x_0 + A^* y_0$, we get

$$g_1(t) = g_{01} e^{t\,x_0} e^{t\,A^* y_0}. \tag{7.57}$$

For (7.54), using (7.51) and using the fact that $G_{\mathbf{z}}$ is Abelian, we have

$$g_2(t) = g_{02} + \int_0^t (Ax(s) + AA^* y_0)\, ds = g_{02} + \int_0^t \left(Ae^{-s\mathrm{ad}\,(A^* y_0)} x_0 + AA^* y_0 \right) ds. \tag{7.58}$$

The parameterization by arclength is obtained requiring $H = \frac{1}{2}$. From (7.49) 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 7.48):

$$u_i(t) = \langle x(t) + y_0 \,|\, e_i + Ae_i \rangle = \langle x(t) \,|\, e_i \rangle + \langle y_0 \,|\, Ae_i \rangle = \langle x(t) + A^* y_0 \,|\, e_i \rangle = \left\langle e^{-t\mathrm{ad}\,(A^* y_0)} x_0 + A^* y_0 \,\middle|\, e_i \right\rangle.$$

Exercise 7.68. Study abnormal extremals for this problem.

7.9 Rolling spheres

7.9.1 $(3,5)$ - Rolling sphere 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 $SO(3)$,

$$e_1 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}, \quad e_2 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix}, \quad e_3 = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \tag{7.59}$$

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 $i = 1, 2, 3$).

$$\left\{ \begin{array}{l} \dot{z}_1 = u_1(t) \\ \dot{z}_2 = u_2(t) \\ \dot{X} = X(u_2(t)e_1 - u_1(t)e_2 + u_3(t)e_3). \end{array} \right. \tag{7.60}$$

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 7.69. *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 7.69 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 $v = u_1(f_1-e_2)+u_2(f_2+e_1)+u_3e_3 \in \mathbf{d}$ we define its norm as $\|v\| = \sqrt{u_1^2+u_2^2+u_3^2}$. If $\pi : \mathfrak{so}(3) \times \mathbb{R}^2 \to \mathbb{R}^2$ is the canonical projection, this norm coincide 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 7.8.2.

The corresponding sub-Riemannian problem is then

$$\dot{g} = g\big(u_1(t)(f_1 - e_2) + u_2(t)(f_2 + e_1) + u_3e_3\big), \tag{7.61}$$

$$g(0) = g_0, \quad g(T) = g_1, \tag{7.62}$$

$$\int_0^T \sqrt{\sum_{i=1}^3 u_i(t)^2\, dt} \;\to \min, \tag{7.63}$$

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 (7.60).

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 7.70. 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

$$A^*f_1 = -e_2, \quad A^*f_2 = e_1. \tag{7.64}$$

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.

Geodesics

Geodesics are parametrized by arclength if we take $x_0 \in \mathfrak{so}(3)$ and $y_0 \in \mathbb{R}^2$ satisfying

$$\|x_0 + A^*y_0\| = 1. \tag{7.65}$$

Now writing $y_0 = y_{01}f_1 + y_{02}f_2$ and using (7.64) 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 (7.65) 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

$$y_{01} \in \mathbb{R}, \quad y_{02} \in \mathbb{R}, \quad \theta \in [0, \pi], \quad \varphi \in [0, 2\pi], \tag{7.66}$$

taking

$$\begin{aligned} x_{01} &= -y_{02} + \cos(\theta)\cos(\varphi) && (7.67)\\ x_{02} &= y_{01} + \cos(\theta)\sin(\varphi) && (7.68)\\ x_{03} &= \sin(\theta) && (7.69)\\ & && (7.70) \end{aligned}$$

The $\mathbf{z}$ part of the geodesics is given by the formula (7.58), with $g_2 \to (z_1, z_2)^T$, i.e.,

$$\begin{aligned} \begin{pmatrix} z_1(t) \\ z_2(t) \end{pmatrix} &= \begin{pmatrix} z_{01} \\ z_{02} \end{pmatrix} + \int_0^t \left(A e^{-s\mathrm{ad}\,(A^* y_0)} x_0 + A A^* y_0 \right) ds. \\ &= \begin{pmatrix} z_{01} \\ z_{02} \end{pmatrix} + \int_0^t \left(A e^{-s(A^* y_0)} x_0 e^{s(A^* y_0)} + \begin{pmatrix} y_{01} \\ y_{02} \end{pmatrix} \right) ds. \end{aligned} \tag{7.71}$$

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 7.71. Prove that each trajectory $(z_1(t), z_2(t))$ is a rototranslation of a sinusoid and that φ determines its initial direction, r its frequence, θ its amplitude and a its rotation on the plane.

The $\mathbf{k}$ part of the geodesics can be obtained with the formula

$$X(t) = e^{t\,x_0} e^{t\,A^* y_0}.$$

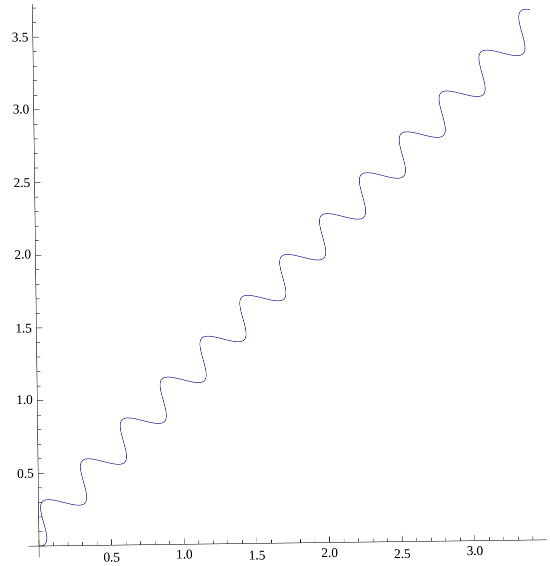


Figure 7.5: A Pontryagin extremals for the rolling sphere with twist

7.9.2 $(2,3,5)$ - Rolling without twisting

We now consider a sphere rolling on a plane without slipping and without twisting. Similarly to what done in Section 7.9, 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 (7.61) 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 (7.59). 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.

The commutators among f_1, f_2, e_1, e_2, e_3 are

$$\begin{gathered}
[f_1, f_2] = 0 \\
[f_i, e_j] = 0, \quad i = 1,2, \quad j = 1,2,3, \\
[e_1, e_2] = e_3, \quad [e_2, e_3] = e_1, \quad [e_3, e_1] = e_2.
\end{gathered} \tag{7.73}$$

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 7.72. 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).

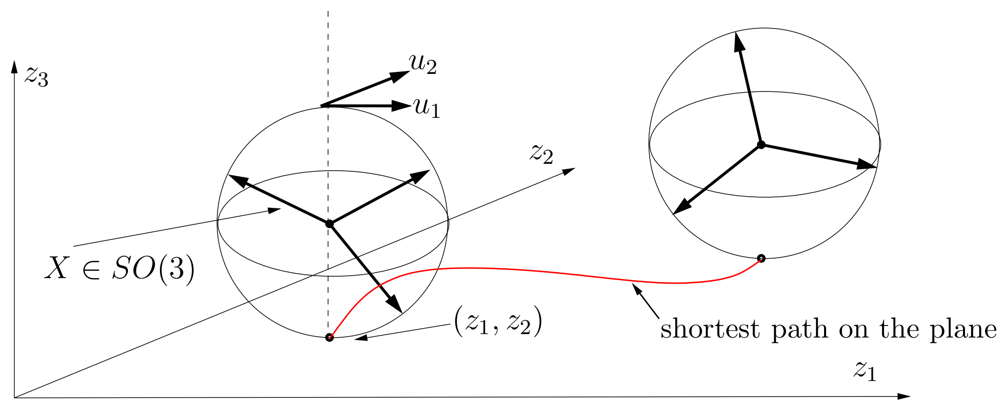


Figure 7.6: The sub-Riemannian problem of rolling a sphere without slipping and twisting.

Contrarily to what happens to the problem of rolling a sphere with twisting (Section 7.9.1), this time the problem is not of the form $\mathbf{k}+\mathbf{z}$. Indeed the distribution is two dimensional while 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

$$H(g,p) = \frac{1}{2}\left(\langle p, L_{g*}(f_1 - e_2)\rangle^2 + \langle p, L_{g*}(f_2 + e_1)\rangle^2\right).$$

The trivialized Hamiltonian is

$$\mathcal{H}(\xi) = \frac{1}{2}\left(\langle \xi, (f_1 - e_2)\rangle^2 + \langle \xi, (f_2 + e_1)\rangle^2\right), \quad \xi \in T_{\mathbb{1}}^* G.$$

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

$$\begin{gathered}
\{h_{f_1}, h_{f_2}\} = \langle \xi, [f_1, f_2]\rangle = 0, \\
\{h_{f_i}, h_{e_j}\} = \langle \xi, [f_i, e_j]\rangle = 0, \quad i = 1,2, \quad j = 1,2,3, \\
\{h_{e_1}, h_{e_2}\} = \langle \xi, [e_1, e_2]\rangle = \langle \xi, e_3\rangle = h_{e_3}, \quad \{h_{e_2}, h_{e_3}\} = h_{e_1}, \quad \{h_{e_3}, h_{e_1}\} = h_{e_2}.
\end{gathered}$$

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

$$\begin{aligned}
\dot{h}_{f_2} &= 0, \\
\dot{h}_{e_1} &= (h_{f_1} - h_{e_2})h_{e_3}, \\
\dot{h}_{e_2} &= (h_{f_2} + h_{e_1})h_{e_3}, \\
\dot{h}_{e_3} &= -h_{f_1}h_{e_1} - h_{f_2}h_{e_2}.
\end{aligned}$$

Now if we consider normal Pontryagin extremals parametrized by length, i.e., if we work on the level $\{\mathcal{H} = 1/2\} \simeq S^1 \times \mathbb{R}^3$, it is convenient to use the coordinates r, α, θ, 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 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, α_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, θ_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 constant of the motion that up to a constant is the energy of the pendulum:

$$H_{\mathrm{p}} = \frac{1}{2}c^2 - r_0\cos(\theta).$$

Fixed (r_0, c_0), one compute H_{p} and the corresponding trajectory in the (θ, 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_{p} are shown in Figure (7.8). 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.

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 7.73. Find all abnormal extremals for this problem.

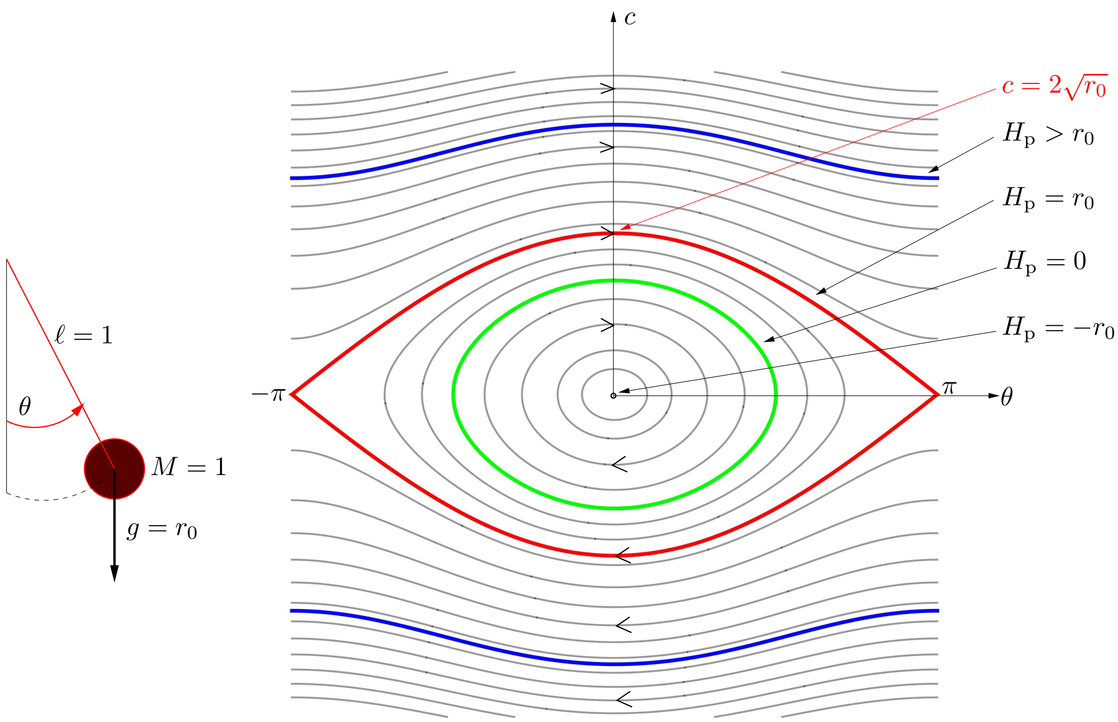


Figure 7.7: Level set of the pendulum for $r_0 \neq 0$. The vertical line $\theta = \pi$ is identified with the veritical line $\theta = -\pi$. We have also indicated the direction of parameterization that one gets from the equation $\dot{\theta} = c$. Notice that the only critical points are $(\theta, c) = (0, 0)$ (stable equilibrium) and $(\theta, c) = (\pi, 0)$ (unstable equilibrium).

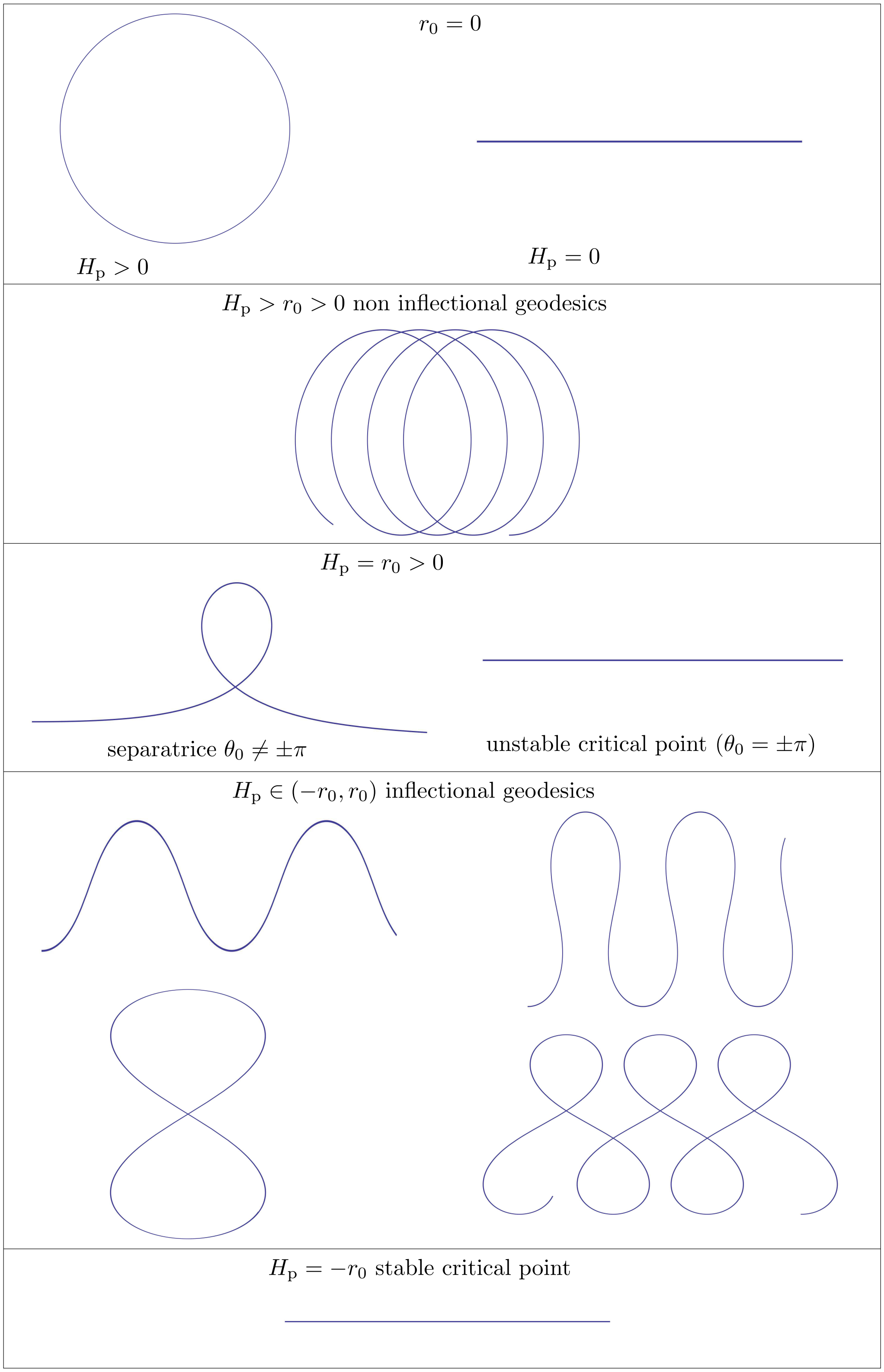


Figure 7.8: z-geodesics. Notice the presence of a periodic trajectories.

7.9.3 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 [47].

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: $r_0 > 0$, α_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;
- θ_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_{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 7.74. *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.*

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 7.75. 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 geodesics. 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

$$X_1 = \partial_{x_1}, \quad X_2 = \partial_{x_2} - x_1\partial_{x_3} + \frac{x_1^2}{2}\partial_{x_4} + x_1 x_2\partial_{x_5}.$$

7.9.4 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 geodesic 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 7.8.4 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 $D_q \subset T_qN,\ q \in N$ be a connection. We have $T_qN = \mathfrak{g} \oplus D_q$; a linear projection $\omega_q : T_qN \to \mathfrak{g}$ such that $\ker \omega_q = D_q$ defines a non-degenerate G-invariant $\mathfrak{g}$-valued vector differential form ω 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 γ. 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 7.8.4, 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.

Bibliographical notes

Chapter 8

End-point map and Exponential map

In Chapter 4 we started to study necessary conditions for an horizontal trajectory to be a minimizer of the sub-Riemannian length between two fixed points. By applying first order variations we found two different class of candidates, namely normal and abnormal extremals. We also proved that normal extremal trajectories are 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 to a control u the final point $E_{q_0}(u)$ of the admissible trajectory associated to u and starting from q_0. Then we treat the problem of minimizing the energy J of curves joining two fixed points $q_0, q_1 \in M$ as the problem of minimization with constraint

$$\min J|_{E_{q_0}^{-1}(q_1)}, \qquad q_1 \in M. \tag{8.1}$$

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 abnormal extremals as candidates for optimality that are critical points of the map E_{q_0} defining the constraint.

In this chapter we take advantage of the invariance by reparametrization to assume all the trajectories to be defined on the same interval $I = [0, 1]$. Also, since the energy of a curve coincides 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). This is useful since $L^2([0, 1], \mathbb{R}^m)$ has a natural structure of Hilbert space.

8.1 The end-point map and its differential

Recall that every sub-Riemannian manifold $(M, \mathbf{U}, f)$ is equivalent to a free one, as explained in Section 3.1.4. In this chapter we always assume that the sub-Riemannian structure is free of rank m, i.e., $\mathbf{U} = M \times \mathbb{R}^m$. In the following $\{f_1, \dots, f_m\}$ denotes a generating frame.

Fix $q_0 \in M$. Recall that, for every control $u \in L^2([0, 1], \mathbb{R}^m)$, the corresponding trajectory γ_u is the unique 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 γ_u starting at q_0 is defined on $[0, 1]$.

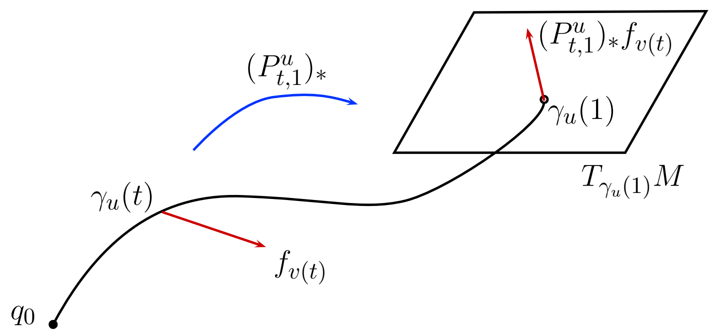


Figure 8.1: Differential of the end-point map

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 closure of the sub-Riemannian ball $\overline{B}_{q_0}(r_0)$ is compact (cf. Corollary 3.35), 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 of rank m 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 γ_u is the unique solution to 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 that is denoted by $E^t_{q_0} : \mathcal{U}^t_{q_0} \to M$ and defined by the identity $E^t_{q_0}(u) := \gamma_u(t)$ defined on the set $\mathcal{U}^t_{q_0}$ of controls u for which the corresponding trajectory γ_u is defined on $[0, t]$.

Now we prove that the end-point map is differentiable (and actually smooth) and we compute its (Fréchet) differential.

Proposition 8.4. *The end-point map E_{q_0} is smooth on $\mathcal{U}_{q_0}$ and for every $u \in \mathcal{U}_{q_0}$ we have*

$$D_uE_{q_0} : L^2([0,1],\mathbb{R}^m) \to T_{\gamma_u(1)}M, \qquad D_uE_{q_0}(v) = \int_0^1 (P^u_{t,1})_* f_{v(t)}\big|_{\gamma_u(1)} dt. \tag{8.4}$$

for every $v \in L^2([0,1],\mathbb{R}^m)$. Here $P^u_{t,s}$ is the flow generated by u.

From the geometric viewpoint, the differential $D_uE_{q_0}(v)$ computes the integral mean of the vector field $f_{v(t)}$ defined by v along the trajectory γ_u defined by u, where all the vectors are pushed forward in the same tangent space $T_{\gamma_u(1)}$ with $P^u_{t,1}$ (see Figure 8.1). We stress that, since $\mathcal{U}_{q_0}$ is an open set of $L^2([0,1],\mathbb{R}^m)$, the differential is defined on the tangent space to $\mathcal{U}_{q_0}$ that is $L^2([0,1],\mathbb{R}^m)$.

Proof of Proposition 8.4. The end-point map from q_0 is a map $E_{q_0} : \mathcal{U}_{q_0} \to M$. Instead of proving the smoothness of the end-point map in coordinates (on M), we will evaluate the end point on a function $a : M \to \mathbb{R}$ and obtain $a \circ E_{q_0} : \mathcal{U}_{q_0} \to \mathbb{R}$, adopting the viewpoint of chronological calculus.

Employing the notation $f_u(q) := \sum_{i=1}^m u_i f_i(q)$. the end-point map from q_0 can be rewritten as the chronological exponential (cf. Chapter 6)

$$E_{q_0}(u) = q_0 \circ \overrightarrow{\exp} \int_0^1 f_{u(t)}\, dt. \tag{8.5}$$

We will show that for every control $\bar{u}$ in the set $\mathcal{U}_{q_0}$ we can write a Taylor expansion around $\bar{u}$ and control the rest at the corresponding order.

Step 1. Let us first show the Taylor expansion of E_{q_0} near the control $\bar{u} = 0$. We remove the subscript q_0 and write

$$E(v) = \overrightarrow{\exp} \int_0^1 f_{v(t)}\, dt. \tag{8.6}$$

splitting it into the sum of the two parts of the Volterra series

$$E(v(\cdot)) = S_N(v) + R_N(v) \tag{8.7}$$

where

$$S_N(v) = \mathrm{Id} + \sum_{k=1}^{N-1} \int \cdots \int_{\Delta_k(1)} f_{v(s_k)} \circ \cdots \circ f_{v(s_1)} ds$$

$$R_N(v) = \int \cdots \int_{\Delta_N(1)} P^v_{0,s_N} \circ f_{v(s_N)} \circ \cdots \circ f_{v(s_1)} ds$$

By linearity of f_v with respect to v, the k-th term in the sum S_N is k-linear. Moreover, applying Theorem 6.19 with $t = 1$, there exists $C > 0$ such that

$$\|R_N(v)a\|_{\alpha,K} \leq \frac{C}{N!} e^{C\|v\|_2} \|v\|_2^N \|a\|_{\alpha+N,K'} \tag{8.8}$$

We stress that the previous inequality holds (for suitable values of the constants) for every $N \in \mathbb{N}$, and in the particular case when $N = 2$ 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. This says in particular that the end-point map is differentiable at $\bar{u} = 0$ and, since the map $v \mapsto \int_0^1 f_{v(t)}(q_0) dt$ is linear and the right hand side is $o(\|v\|_2)$, computes its differential.

Step 2. To compute the Taylor expansion at an arbitrary point $\bar{u} \in \mathcal{U}_{q_0}$, let us consider the expansion in a neighborhood of $v = 0$ of the map

$$v \mapsto E_{q_0}(\bar{u} + v) = q_0 \circ \overrightarrow{\exp} \int_0^1 f_{(\bar{u}+v)(t)} dt.$$

Using the variation formula (6.29) one can write

$$\begin{aligned}
\overrightarrow{\exp} \int_0^1 f_{(\bar{u}+v)(t)} dt &= \overrightarrow{\exp} \int_0^1 f_{\bar{u}(t)} + f_{v(t)} dt \\
&= \overrightarrow{\exp} \int_0^1 \left(\overrightarrow{\exp} \int_0^t \mathrm{ad}\, f_{\bar{u}(s)} ds \right) f_{v(t)} dt \circ \overrightarrow{\exp} \int_0^1 f_{\bar{u}(t)} dt \qquad (8.11) \\
&= \overrightarrow{\exp} \int_0^1 (P^{\bar{u}}_{0,t})^{-1}_* f_{v(t)} dt \circ P^{\bar{u}}_{0,1}
\end{aligned}$$

Indeed we have

$$E_{q_0}(\bar{u}+v) = P^{\bar{u}}_{0,1}(G^{\bar{u}}_{q_0}(v)) = G^{\bar{u}}_{q_0}(v) \circ P^{\bar{u}}_{0,1} \tag{8.12}$$

where $G^{\bar{u}}_{q_0}$ is the map defined as follows

$$G^{\bar{u}}_{q_0}(v) := q_0 \circ \overrightarrow{\exp} \int_0^1 (P^{\bar{u}}_{0,t})^{-1}_* f_{v(t)} dt$$

Then, the expansion of (8.12) near $v = 0$ is obtained by the Volterra expansion of the map $G^{\bar{u}}_{q_0}$ with respect to v. Using the same computations and estimate as above one obtains

$$D_0 G^{\bar{u}}_{q_0}(v) = q_0 \circ \int_0^1 (P^{\bar{u}}_{0,t})^{-1}_* f_{v(t)} dt = \int_0^1 (P^{\bar{u}}_{0,t})^{-1}_* f_{v(t)}(q_0) dt \tag{8.13}$$

and, by composition,

$$\begin{aligned} D_{\bar{u}} E_{q_0}(v) &= (P^{\bar{u}}_{0,1})_* \circ D_0 G^{\bar{u}}_{q_0}(v) = (P^{\bar{u}}_{0,1})_* \int_0^1 (P^{\bar{u}}_{0,t})^{-1}_* f_{v(t)}(q_0) dt \\ &= \int_0^1 (P^{\bar{u}}_{t,1})_* f_{v(t)}(q_1) dt. \end{aligned}$$

where we denote $q_1 := E_{q_0}(\bar{u})$. □

Remark 8.5. Notice that the decomposition of the non autonomous flow associated with $\bar{u} + v$ into the one associated with $\bar{u}$ 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 3.53 (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 φ under the constraint defined by F. More precisely, we want to write necessary conditions satisfied by the solutions of the problem

$$\min \varphi\big|_{F^{-1}(q)}, \qquad q \in M. \tag{8.15}$$

Theorem 8.6. *Assume $u \in \mathcal{U}$ is solution of the minimization problem* (8.15)*. Then there exists a covector $(\lambda, \nu) \in T^*_q M \times \mathbb{R}$ such that $(\lambda, \nu) \neq (0, 0)$ and*

$$\lambda\, D_u F + \nu D_u \varphi = 0. \tag{8.16}$$

Remark 8.7. Formula (8.16) means that for every $v \in \mathcal{H}$ one has

$$\langle \lambda, D_u F(v) \rangle + \nu D_u \varphi(v) = 0.$$

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 Ψ, then $D_u\Psi$ is surjective. By implicit function theorem, this implies that Ψ is locally surjective at u. In particular, for every neighborhood V of u it 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 $D_u\Psi = (D_uF, D_u\varphi)$ is not surjective and there exists a non zero covector (λ, ν) such that $\lambda\, D_uF + \nu D_u\varphi = 0$. □

8.3 Pontryagin extremals via Lagrange multipliers

Applying the previous result to the case when $F = E_{q_0}$ is the end-point map and $\varphi = J$ is the sub-Riemannian energy, one 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*

$$\lambda\, D_uE_{q_0} + \nu D_uJ = 0. \tag{8.17}$$

Let us now prove that these necessary conditions are equivalent to those obtained in Chapter 4. Recall that, since $J(u) = \frac{1}{2}\|u\|^2_{L^2}$, then $D_uJ(v) = (u, v)_{L^2}$ and, identifying $L^2([0, 1], \mathbb{R}^m)$ with its dual, we have $D_uJ = u$.

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 $\lambda(t) = (P^u_{t,1})^*\lambda_1$ and u satisfies* (8.17) *with $(\lambda, \nu) = (\lambda_1, -1)$, namely*

$$\lambda_1 D_uE_{q_0} = u. \tag{8.18}$$

(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 $\lambda(t) = (P^u_{t,1})^*\lambda_1$ and u satisfies* (8.17) *with $(\lambda, \nu) = (\lambda_1, 0)$, namely*

$$\lambda_1 D_uE_{q_0} = 0. \tag{8.19}$$

where in (8.18) *we identify $u \in L^2$ with the element $(u, \cdot)_{L^2} \in (L^2)'$*

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 $i = 1, \dots, m$, where $\gamma(t) = \pi(\lambda(t))$.

Assume that u satisfies (8.18) for some λ_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

$$\langle\lambda_1, D_uE_{q_0}(v)\rangle = (u, v)_{L^2} \tag{8.20}$$

Using (8.4), the left hand side is rewritten as follows

$$\begin{aligned}\langle \lambda_1, D_u E_{q_0}(v)\rangle &= \int_0^1 \left\langle \lambda_1, (P^u_{t,1})_* f_{v(t)}(q_1)\right\rangle dt = \int_0^1 \left\langle (P^u_{t,1})^* \lambda_1, f_{v(t)}(\gamma(t))\right\rangle dt \\ &= \int_0^1 \left\langle \lambda(t), f_{v(t)}(\gamma(t))\right\rangle dt = \int_0^1 \sum_{i=1}^m \langle \lambda(t), f_i(\gamma(t))\rangle\, v_i(t) dt,\end{aligned}$$

where we used that $\gamma(t) = (P^u_{t,1})^{-1}(q_1)$. Then (8.20) becomes

$$\int_0^1 \sum_{i=1}^m \langle \lambda(t), f_i(\gamma(t))\rangle\, v_i(t) dt = \int_0^1 \sum_{i=1}^m u_i(t) v_i(t) dt. \tag{8.21}$$

and since $v(t)$ 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. □

8.4 Critical points and second order conditions

In this chapter, we develop second order conditions for constrained critical points in the case in which the constraint is regular. When applied to the sub-Riemannian case, this gives second order conditions for normal extremals (that are not abnormal). Cf. also Section 8.5.

In the following $\mathcal{H}$ always denote an 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 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 $\phi(\mathcal{V} \cap Y) = \mathcal{W} \cap U$ for U a closed linear subspace of $\mathcal{H}$.

We now recall the implicit function theorem in this setting.

Proposition 8.10 (Implicit function theorem). *Let $F : \mathcal{H} \to M$ be a smooth map and fix $q \in M$. If F is a submersion at every $u \in F^{-1}(q)$, i.e., the Fréchet differential $D_u F : \mathcal{H} \to T_q M$ is surjective for every $u \in F^{-1}(q)$, then $F^{-1}(q)$ is a smooth submanifold whose codimension is equal to the dimension of M. Moreover $T_u F^{-1}(q) = \ker D_u F$.*

We now define critical points.

Definition 8.11. 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 $D_u\varphi\big|_{T_u N} = 0$.

We start with a geometric version of the Lagrange multipliers rule, which caracterizes constrained critical points (not just minima). This construction is then used to develop a second order analysis.

Proposition 8.12 (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 $F : \mathcal{U} \to M$. 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*

$$\lambda\, D_u F = D_u \varphi. \tag{8.22}$$

Proof. Recall that the differential of F is a well-defined map

$$D_uF : T_u\mathcal{U} \to T_qM, \qquad q = F(u).$$

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 $T_u\mathcal{V}_q = \ker D_uF$. Since u is a critical point of $\varphi\big|_{\mathcal{V}_q}$, by definition $D_u\varphi\big|_{T_u\mathcal{V}_q} = D_u\varphi\big|_{\ker D_uF} = 0$, i.e.,

$$\ker D_uF \subset \ker D_u\varphi. \tag{8.23}$$

Now consider the following diagram

$$\begin{tikzcd} T_u\mathcal{U} \arrow[r, "D_uF"] \arrow[dr, "d_u\varphi"'] & T_qM \arrow[d, "?"] \\ & \mathbb{R} \end{tikzcd} \tag{8.24}$$

From (8.23), using Exercice 8.13, it follows that there exists a linear map $\lambda : T_qM \to \mathbb{R}$ (that means $\lambda \in T_q^*M$) that makes the diagram (8.24) commutative. □

Exercise 8.13. 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 $\lambda \circ G = \phi$.

Now we want to consider second order information at critical points. Recall that, for a function $\varphi : \mathcal{U} \to \mathbb{R}$ defined on an open set $\mathcal{U}$ of an Hilbert space $\mathcal{H}$, the *first and second differential* are defined in the following way,

$$D_u\varphi(v) = \left.\frac{d}{ds}\right|_{s=0} \varphi(u + sv), \qquad D_u^2\varphi(v) = \left.\frac{d^2}{ds^2}\right|_{s=0} \varphi(u + sv)$$

For a function $F : \mathcal{U} \to M$ whose target space is a manifold its first differential $D_uF : \mathcal{H} \to T_{F(u)}M$ is still well defined while the second differential D_u^2F is meaningful only if we fix a set of coordinates in the target space.

If $\mathcal{V}$ is a submanifold in $\mathcal{H}$, the first differential of a smooth function $\psi : \mathcal{V} \to \mathbb{R}$ at a point $u \in \mathcal{V}$ is defined as

$$D_u\psi : T_u\mathcal{V} \to \mathbb{R}, \qquad D_u\psi(v) = \left.\frac{d}{ds}\right|_{s=0} \psi(w(s)),$$

where $w : (-\varepsilon, \varepsilon) \to \mathcal{V}$ is a curve that satisfies $w(0) = u, \ \dot{w}(0) = v$. If $\psi = \varphi|_{\mathcal{V}}$ is the restriction of a function $\varphi : \mathcal{H} \to \mathbb{R}$ defined globally on $\mathcal{H}$, then $D_u\psi = D_u\phi|_{T_u\mathcal{V}}$ coincides with the restriction of the differential defined on the ambient space $\mathcal{H}$. For the second differential things are more delicate. Indeed the formula

$$v \in T_u\mathcal{V} \mapsto \left.\frac{d^2}{ds^2}\right|_{s=0} \psi(w(s)) \tag{8.25}$$

where $w : (-\varepsilon, \varepsilon) \to \mathcal{V}$ is a curve that satisfies $w(0) = u, \ \dot{w}(0) = v$, is a well-defined object (i.e., the right hand side depends only on v) only if u is a critical point of ψ. Indeed, if this is not the case, the quantity (8.25) depends also on the second derivative of w, as it is easily checked.

If u is a critical point of $\psi : \mathcal{V} \to \mathbb{R}$ (i.e., $D_u\psi = 0$) the second order differential (8.25) is a well-defined quadratic form $T_u\mathcal{V}$, that is called the *Hessian* of ψ at u:

$$\operatorname{Hess}_u \psi : T_u\mathcal{V} \to \mathbb{R}, \qquad v \mapsto \left.\frac{d^2}{ds^2}\right|_{s=0} \psi(w(s)) \tag{8.26}$$

We stress that if $\psi = \varphi|_{\mathcal{V}}$ is the restriction of a function $\varphi : \mathcal{H} \to \mathbb{R}$ defined globally on $\mathcal{H}$, then the Hessian of ψ at a critical point u *does not* coincide, in general, with the restriction of the second differential of φ to the tangent space $T_u\mathcal{V}$.

Let us compute the Hessian of the restriction in the case when $\mathcal{V} = F^{-1}(q)$ is a smooth submanifold of $\mathcal{H}$, and $\psi = \varphi\big|_{F^{-1}(q)}$. Using that $T_uF^{-1}(q) = \ker D_uF$, the Hessian is a well-defined quadratic form

$$\operatorname{Hess}_u \varphi\big|_{F^{-1}(q)} : \ker D_uF \to \mathbb{R}$$

that is computed in terms of the second differentials of φ and F as follows.

Proposition 8.14. *For all $v \in \ker D_uF$ we have*

$$\operatorname{Hess}_u \varphi\big|_{F^{-1}(q)}(v) = D_u^2\varphi(v) - \lambda\, D_u^2F(v). \tag{8.27}$$

where λ is satisfies the identity $\lambda\, D_uF = D_u\varphi$.

Remark 8.15. We stress again that in (8.27), 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^2F needs also a choice of coordinates in the manifold M.

Proof of Proposition 8.14. 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. Differentiating twice with respect to u, with respect to some local coordinates on M, we have

$$D_uF(\dot u) = 0, \qquad \langle D_u^2F(\dot u), \dot u\rangle + D_uF(\ddot u) = 0. \tag{8.28}$$

where we denoted by $\dot u = \dot u(0)$ and $\ddot u = \ddot u(0)$. Analogous computations for φ gives

$$\begin{aligned}
\operatorname{Hess}_u \varphi\big|_{F^{-1}(q)}(\dot u) &= \left.\frac{d^2}{ds^2}\right|_{s=0} \varphi(w(s)) \\
&= \langle D_u^2\varphi(\dot u), \dot u\rangle + D_u\varphi(\ddot u) \\
&= \langle D_u^2\varphi(\dot u), \dot u\rangle + \lambda D_uF(\ddot u) \qquad &&(\text{by } \lambda\, D_uF = D_u\varphi) \\
&= \langle D_u^2\varphi(\dot u), \dot u\rangle - \lambda \langle D_u^2F(\dot u), \dot u\rangle \qquad &&(\text{by } (8.28))
\end{aligned}$$

□

8.4.1 The manifold of Lagrange multipliers

As above, let us consider the two smooth maps $\varphi : \mathcal{U} \to \mathbb{R}$ and $F : \mathcal{U} \to M$ defined on an open set $\mathcal{U}$ of an Hilbert space $\mathcal{H}$.

Definition 8.16. We say that a pair (u,λ), with $u \in \mathcal{U}$ and $\lambda \in T^*M$, is a *Lagrange point* for the pair (F,φ) if $\lambda \in T^*_{F(u)}M$ and $D_u\varphi = \lambda\, D_uF$. We denote the set of all Lagrange points by $C_{F,\varphi}$. More precisely

$$C_{F,\varphi} = \{(u,\lambda) \in \mathcal{U} \times T^*M \mid F(u) = \pi(\lambda),\ D_u\varphi = \lambda\, D_uF\}. \tag{8.29}$$

The set $C_{F,\varphi}$ is a well-defined subset of the vector bundle $F^*(T^*M)$, that we recall is defined as follows (cf. also Definition 2.50)

$$F^*(T^*M) = \{(u,\lambda) \in \mathcal{U} \times T^*M \mid F(u) = \pi(\lambda)\}. \tag{8.30}$$

We now study the structure of the set $C_{F,\varphi}$. It turns to be a smooth manifold under some regularity conditions on the maps (F,φ).

Definition 8.17. The pair (F,φ) is said to be a *Morse pair* (or a *Morse problem*) if 0 is a regular value for the smooth map

$$\theta : F^*(T^*M) \to \mathcal{U}^* \simeq \mathcal{U}, \qquad (u,\lambda) \mapsto D_u\varphi - \lambda\, D_uF. \tag{8.31}$$

Remark 8.18. Notice that, if M is a single point, then F is the trivial map and with this definition we have that (F,φ) is a Morse pair if and only if φ is a Morse function. Indeed in this case $D_uF = 0$, and 0 is a critical value for θ if, by definition, the second differential $D^2_u\varphi$ is non-degenerate.

Proposition 8.19. *If (F,φ) define a Morse problem, then $C_{F,\varphi}$ is a smooth manifold in $F^*(T^*M)$.*

Proof. To prove that $C_{F,\varphi}$ is a smooth manifold it is sufficient to notice that $C_{F,\varphi} = \theta^{-1}(0)$ and, by definition of Morse pair, 0 is a regular value of θ. The result follows from the version of the implicit function theorem stated in Lemma 8.20 □

Lemma 8.20. *Let N be a smooth 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 θ, were finite, a simple dimensional argument would permit to compute the dimension of $C_{F,\varphi} = \theta^{-1}(0)$ (as in Proposition 8.10). In this case, since the differential of θ 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 no more valid, and we need the explicit expression of the differential of θ.

Proposition 8.21. *Under the assumption of Proposition 8.19, then $\dim C_{F,\varphi} = \dim M = n$.*

Proof. To prove the statement, let us choose a set of coordinates $\lambda = (\xi, x)$ in T^*M and describe the set $C_{F,\varphi} \subset F^*(T^*M)$ as follows

$$\begin{cases} D_u\varphi - \xi D_u F = 0 \\ F(u) = x \end{cases} \tag{8.32}$$

where here ξ is thought as a row vector. 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 (u, ξ, x). The tangent space $T_{(u,\xi,x)}C_{F,\varphi}$ is described in coordinates by the set of points (u', ξ', x') satisfying the equations[1]

$$\begin{cases} D_u^2\varphi(u', \cdot) - \xi D_u^2 F(u', \cdot) - \xi' D_u F(\cdot) = 0 \\ D_u F(u') = x' \end{cases} \tag{8.33}$$

Let us denote the linear map $Q : \mathcal{U} \to \mathcal{U}^* \simeq \mathcal{U}$ defined by

$$Q(u') = D_u^2\varphi(u', \cdot) - \xi D_u^2 F(u', \cdot).$$

Since Q is defined by second derivatives of the maps F and φ, it is a symmetric operator. on the Hilbert space $\mathcal{U}$.

The definition of Morse problem is immediately rewritten as follows: the pair (F, φ) defines a Morse problem if and only if the following map is surjective.

$$\Theta : \mathcal{U} \times \mathbb{R}^{n*} \to \mathcal{U}^* \simeq \mathcal{U}, \qquad \Theta(u', \xi') = Q(u') - B(\xi'). \tag{8.34}$$

where we denoted with $B : \mathbb{R}^{n*} \to \mathcal{U}^* \simeq \mathcal{U}$ the map

$$B(\xi') = \xi' D_u F(\cdot).$$

Indeed the map Θ is exactly the first equation in (8.33). The dimension of $C_{F,\varphi}$ coincides with the dimension of $\ker \Theta$. Indeed for each element $(u', \xi') \in \ker \Theta$ by setting $x' = D_u F(u')$ we find a unique $(u', \xi', x') \in T_{(u,\xi,x)}C_{F,\varphi}$. Since Q is self-adjoint, we have

$$\mathcal{U} = \ker Q \oplus \overline{\operatorname{im} Q}, \qquad \dim \ker Q = \operatorname{codim} \operatorname{im} Q.$$

Using that Θ is surjective and $\dim(\operatorname{im} B) \leq n$ we get that

$$\dim \ker Q = \operatorname{codim} \operatorname{im} Q \leq \dim \operatorname{im} B \leq n,$$

is finite dimensional (in particular $\operatorname{im} Q$ is closed and $\mathcal{U} = \ker Q \oplus \operatorname{im} Q$).

If we denote with $\pi_{\ker} : \mathcal{U} \to \ker Q$ and $\pi_{\operatorname{im}} : \mathcal{U} \to \operatorname{im} Q$ the orthogonal projection onto the two subspaces, it is easy to see that

$$\Theta(u', \xi') = 0 \quad \Longleftrightarrow \quad \begin{cases} \pi_{\ker} B\xi' = 0 \\ \pi_{\operatorname{im}} B\xi' = Qu' \end{cases}$$

Moreover $\pi_{\ker} B : \mathbb{R}^n \to \ker Q$ is a surjective map between finite-dimensional spaces (the surjectivity is a consequence of the fact that Θ is surjective). 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 $\square$

[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 $T_z C$ at a point $z \in C$ is described by the linear equation $\{z' \in Z \mid D_z\Psi(z') = 0\}$.

The last characterization of Morse problem leads to a convenient criterion to check whether a pair (F, φ) defines a Morse problem.

Lemma 8.22. *The pair (F, φ) defines a Morse problem if and only if*

(i) $\operatorname{im} Q$ *is closed,*

(ii) $\ker Q \cap \ker D_u F = \{0\}$.

Proof. Assume that (F, φ) is a Morse problem. Then, following the lines of the proof of Proposition 8.21, $\operatorname{im} Q$ has finite codimension, hence is closed, and (i) is proved. Moreover, since the problem is Morse, then the image of the differential of the map (8.31) is surjective, i.e. if there exists $w \in \mathcal{U}$ that is orthogonal to $\operatorname{im} \Theta$, namely

$$\langle Q(u'), w\rangle - \langle \xi' D_u F(\cdot), w\rangle = 0, \qquad \forall\, (\xi', u'),$$

then $w = 0$. Using that Q is self-adjoint we can rewrite the previous identity as

$$\langle u', Q(w)\rangle - \langle \xi' D_u F(\cdot), w\rangle = 0, \qquad \forall\, (\xi', u'),$$

that is equivalent, since ξ', u' are arbitrary, to

$$Q(w) = 0 \qquad \text{and} \qquad D_u F(w) = 0.$$

This proves (ii). The converse implications are proved in a similar way. □

Definition 8.23. Let N be a n-dimensional submanifold. An immersion $F : N \to T^*M$ is said to be a *Lagrange immersion* if $F^*\sigma = 0$, where σ 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 8.24. *If the pair (F, φ) defines a Morse problem, then F_c is a Lagrange immersion.*

Proof. First we prove that F_c is an immersion and then that $F_c^*\sigma = 0$.

(i). Recall that $F_c : C_{F,\varphi} \to T^*M$ where

$$C_{F,\varphi} = \{(u, \xi, x) \,|\, \text{equations (8.32) holds}\}$$

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 (8.32)

$$T_{(u,\lambda)}C_{F,\varphi} = \{(u', \xi', x') \,|\, \text{equations (8.33) holds}\}$$

where

$$D_{(u,\lambda)}F_c(u', \xi', x') = (\xi', x')$$

Now looking at (8.33) it easily seen that

$$D_{(u,\lambda)}F_c(u', \xi', x') = 0 \qquad \text{iff} \qquad Q(u') = D_u F(u') = 0.$$

Since (F, φ) defines a Morse problem we have by Lemma 8.22 that such a u' does not exists. Hence the differential is never zero and F_c is an immersion.

(ii). We now show that $F_c^*\sigma = 0$. Since $\sigma = ds$ is the differential of the tautological form s, and $F_c^*\sigma = dF_c^*s$ since the pullback commutes with the differential, it is sufficient to show that F_c^*s is closed. Let us show the identity

$$F_c^*s = D(\varphi \circ \pi_{\mathcal{U}})\big|_{C_{F,\varphi}}.$$

By definition of the map F_c, the following diagram is commutative:

$$\begin{tikzcd} C_{F,\varphi} \arrow[r, "F_c"] \arrow[d, "\pi_{\mathcal{U}}"'] & T^*M \arrow[d, "\pi_M"] \\ \mathcal{U} \arrow[r, "F"'] & M \end{tikzcd} \tag{8.35}$$

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

$$\begin{aligned}
(F_c^*s)_{(u,\lambda)} &= s_\lambda \circ D_{(u,\lambda)}F_c \\
&= \lambda \circ \pi_{M*} \circ D_{(u,\lambda)}F_c \qquad &&\text{(by definition } s_\lambda = \lambda \circ \pi_{M*}\text{)} \\
&= \lambda \circ D_uF \circ \pi_{\mathcal{U}*} &&\text{(by (8.35))} \\
&= D_u(\varphi \circ \pi_{\mathcal{U}}) &&\text{(by (8.22))}
\end{aligned}$$

□

Definition 8.25. The set $\mathcal{L}_{F,\varphi} \subset T^*M$ of *Lagrange multipliers* associated with the pair (F,φ) is the image of $C_{F,\varphi}$ under the map F_c.

From Proposition 8.24 it follows that, if $\mathcal{L}_{F,\varphi}$ is a smooth manifold, 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 8.26. *Let (F,φ) be a Morse pair and assume (u,λ) is a Lagrange point such that u is a regular point for F, where $F(u) = q = \pi(\lambda)$. The following properties are equivalent:*

(i) $\mathrm{Hess}_u\,\varphi\big|_{F^{-1}(q)}$ *is degenerate,*

(ii) (u,λ) *is a critical point for the map* $\pi \circ F_c = F\big|_{C_{F,\varphi}} : C_{F,\varphi} \to M$,

Moreover, if $\mathcal{L}_{F,\varphi}$ is a submanifold, then (i) and (ii) are equivalent to

(iii) λ *is a critical point for the map* $\pi\big|_{\mathcal{L}_{F,\varphi}} : \mathcal{L}_{F,\varphi} \to M$.

Proof. In coordinates we have the following expression for the Hessian

$$\mathrm{Hess}_u\varphi\big|_{F^{-1}(q)}(v) = \langle Q(v), v\rangle, \qquad \forall\, v \in \ker D_uF.$$

and Q is the linear operator associated to the bilinear form. Assume that $\mathrm{Hess}_u\,\varphi\big|_{F^{-1}(q)}$ is degenerate, i.e. there exists $u' \in \ker D_uF$ such that

$$\langle Qu', v\rangle = 0, \qquad \forall\, v \in \ker D_uF.$$

In other words $Q(u') \perp \ker D_u F$ that is equivalent to say that $Q(u')$ is a linear combination of the row of the Jacobian matrix of F, namely

$$Q(u') = \xi' D_u F(\cdot),$$

for some row vector ξ'. From equations (8.33) it follows immediately that (i) is equivalent to (ii). The fact that, if $\mathcal{L}_{F,\varphi}$ is a submanifold, (ii) is equivalent to (iii) is obvious. □

8.5 Sub-Riemannian case

In this section we want to specify the theory that we developed in the previous ones to the case of sub-Riemannian normal extremal. Hence, we will consider the action functional J defined by $J(u) = \frac{1}{2}\int_0^1 |u(t)|^2 dt$ and we consider its critical points constrained to a regular level set of the end-point map E, that means that we fix the final point of our trajectory (as usual we assume that the starting point q_0 is fixed).

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 $J\big|_{E^{-1}(q_1)}$.

Lemma 8.27. *Let $q_1 \in M$ and (u, λ) be a critical point of $J\big|_{E^{-1}(q_1)}$. Then for every $v \in \ker D_u F$*

$$\mathrm{Hess}_u J\big|_{E^{-1}(q_1)}(v) = \|v\|^2_{L^2} - \left\langle \lambda, D^2_u E(v) \right\rangle, \tag{8.36}$$

where

$$D^2_u E(v,v) = 2 \iint\limits_{0 \le s \le t \le 1} [(P_{s,1})_* f_{v(s)}, (P_{t,1})_* f_{v(t)}](q_1)\, ds dt. \tag{8.37}$$

and $P_{t,s}$ denotes the nonautonomous flow defined by the control u.

Proof. By Proposition 8.14 we have

$$\mathrm{Hess}_u J\big|_{E^{-1}(q_1)}(v) = D^2_u J - \lambda D^2_u E.$$

It is easy to compute derivatives of J. Indeed we can rewrite it as $J(u) = \frac{1}{2}(u,u)_{L^2}$, hence

$$D_u J(v) = (u,v)_{L^2}, \qquad D^2_u J(v) = (v,v)_{L^2} = \|v\|^2_{L^2}, \qquad \forall\, v \in \ker D_u E$$

It remains to compute the second derivative of the end-point map. From the Volterra expansion (8.13) we get

$$D^2_u E(v,v) = 2\, q_1 \circ \iint\limits_{0 \le s \le t \le 1} (P_{s,1})_* f_{v(s)} \circ (P_{t,1})_* f_{v(t)} ds dt \tag{8.38}$$

To end the proof we use the following lemma on chronological calculus, which we will use to symmetrize the second derivative.

Lemma 8.28. *Let X_t be a nonautonomous vector field on M. Then*

$$\iint\limits_{0 \le s \le t \le 1} X_s \circ X_t ds dt = \frac{1}{2} \int_0^1 X_s ds \circ \int_0^1 X_t dt + \frac{1}{2} \iint\limits_{0 \le s \le t \le 1} [X_s, X_t] ds dt. \tag{8.39}$$

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 8.28 we obtain from (8.38)

$$D_u^2E(v,v) = q_1\odot 2\iint\limits_{0\le s\le t\le 1} [(P_{s,1})_* f_{v(s)}, (P_{t,1})_* f_{v(t)}] dsdt \tag{8.40}$$

where we used that $\int_0^1 (P_{t,1})_* f_{v(t)} dt = 0$ since $v\in\ker D_uE$. □

Proposition 8.29. *The sub-Riemannian problem (E, J) is a Morse pair.*

Proof. We use the characterization of Lemma 8.22. We have to show that

$$\operatorname{im}\left(\mathrm{Id}-\lambda D_u^2E\right) \text{ is closed}, \qquad \ker\left(\mathrm{Id}-\lambda D_u^2E\right)\cap\ker\,(D_uE) = 0. \tag{8.41}$$

Using the previous notation and defining $g_v^t := (P_{t,1})_* f_v$, we can write

$$D_uE(v) = q_1\odot\int_0^1 g_{v(t)}^t dt$$

Moreover we have

$$\begin{aligned}
\left\langle\lambda D_u^2E(v), v\right\rangle &= 2\iint\limits_{0\le s\le t\le 1} g_{v(s)}^s\odot g_{v(t)}^t dsdt\odot a & (8.42)\\
&= \iint\limits_{0\le s\le t\le 1} g_{v(s)}^s\odot g_{v(t)}^t dsdt\odot a + \iint\limits_{0\le t\le s\le 1} g_{v(t)}^t\odot g_{v(s)}^s dsdt\odot a & (8.43)\\
&= \int_0^1\int_0^t g_{v(s)}^s\odot g_{v(t)}^t dsdt\odot a + \int_0^1\int_t^1 g_{v(t)}^t\odot g_{v(s)}^s dsdt\odot a & (8.44)
\end{aligned}$$

where a is any smooth function such that $d_{q_1}a = \lambda$.

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

$$\left\langle \lambda D_u^2 E(v), v \right\rangle = (Qv, v)_{L^2} = \int_0^1 (Qv)(t)v(t)dt.$$

Then it follows that

$$(Qv)(t) = \left(\int_0^t g_{v(s)}^s ds \odot g^t + g^t \odot \int_t^1 g_{v(s)}^s ds \right) \odot a \tag{8.45}$$

where g^t denotes the vector $(g_1^t, \ldots, g_m^t)$ and we recall that $g_i^t = (P_{t,1})_* f_i$ for $i = 1, \ldots, m$. Let us now prove the following technical lemma.

Lemma 8.30. *Let us consider the linear operator $A : L^2([0,T], \mathbb{R}^m) \to L^2([0,T], \mathbb{R}^m)$ defined by*

$$(Av)(t) = v(t) - \int_0^t K(t,s)v(s)ds \tag{8.46}$$

where $K(t,s)$ is a function in $L^2([0,T]^2, \mathbb{R}^m)$. Then

(i) $A = I - Q$, *where Q is a compact operator,*

(ii) $\ker A = \{0\}$.

Moreover, if $K(t,s) = K(s,t)$ for all t, s, then A is a symmetric operator.

Proof. The fact that the integral operator $Q : L^2([0,T], \mathbb{R}^m) \to L^2([0,T], \mathbb{R}^m)$ defined by

$$(Qv)(t) = \int_0^t K(t,s)v(s)ds \tag{8.47}$$

is compact is classical (see for instance [61, Chapter 6]). We then prove statement (ii) in two steps. (a) we prove it for small T. (b) we prove it for arbitrary T.

(a). Fix $T > 0$ and consider a solution in $L^2([0,T], \mathbb{R}^m)$ to the equation

$$v(t) = \int_0^t K(t,s)v(s)ds, \qquad t \in [0,T]. \tag{8.48}$$

We multiply (8.48) by $v(t)$ and integrate over $t \in [0,T]$, obtaining

$$\int_0^T v(t)^2 dt = \int_0^T \int_0^t K(t,s)v(s)v(t)dsdt$$

By applying twice the Cauchy-Schwartz identity, one obtains

$$\int_0^T v(t)^2 dt \leq \left(\int_0^T \int_0^T |K(t,s)|^2 dtds \right)^{1/2} \int_0^T v(t)^2 dt.$$

or, equivalently

$$\|v\|_{L^2}^2 \leq \|K\|_{L^2} \|v\|_{L^2}^2.$$

Since for $T \to 0$ we have $\|K\|_{L^2([0,T]^2, \mathbb{R}^m)} \to 0$, this implies that $v = 0$ on $[0,T]$.

(b). Consider a solution of the identity (8.48) and define $T^* = \sup\{\tau > 0 \,|\, v(t) = 0, t \in [0,\tau]\}$. By part (a) one has $T^* > 0$. Since the set $X := \{v \in L^2([0,T], \mathbb{R}^m) \,|\, v(t) = 0 \text{ a.e. on } [0,T^*]\}$ is preserved by A (namely $A(X) \subset X$) using again part (a) one obtains that v indeed vanishes on $[0, T^* + \varepsilon]$, for some $\varepsilon > 0$, contradicting the fact that that T^* is the supremum. □

Let us go back to the proof of Proposition 8.29. Since (8.45) 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, for every control $v \in \ker D_u E$ we have the identity (cf. (8.4))

$$q_1 \odot \int_0^t g^s_{v(s)} ds = -q_1 \odot \int_t^1 g^s_{v(s)} ds$$

Hence we have that v belong to the intersection in (8.41) if and only if it satisfies

$$\left(I - \lambda D^2_u E\right) v(\cdot)(t) = v(t) + \lambda \int_0^t \left[g^s_{v(s)}, g^t_{v(t)}\right](q_1) ds$$

which has trivial kernel thanks to Lemma 8.30. $\square$

Combining the last result with Proposition 8.24 we obtain the following corollary.

Corollary 8.31. *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 parametrizes normal extremals through their initial covectors.

Definition 8.32. Let $q_0 \in M$. The *sub-Riemannian exponential map* (based at q_0) is the map

$$\exp_{q_0} : \mathscr{A}_{q_0} \subset T^*_{q_0}M \to M, \qquad \exp_{q_0}(\lambda_0) = \pi \circ e^{\vec{H}}(\lambda_0). \tag{8.49}$$

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 8.33. *Let H be the sub-Riemannian Hamiltonian. Then, for every $\lambda \in T^*M$*

$$e^{t\vec{H}}(\alpha\lambda) = \alpha e^{\alpha t\vec{H}}(\lambda), \tag{8.50}$$

for any $\alpha > 0$ and $t > 0$ such that both sides of the identity are defined.

Proof. By Remark 4.27 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 (8.50) follows from the uniqueness of the solution and the fact that $\lambda_\alpha(0) = \alpha\lambda(0)$. $\square$

The homogeneity property (8.50) permits to recover the whole extremal trajectory as the image of the ray joining 0 to λ_0 in the fiber $T^*_{q_0}M$.

Corollary 8.34. *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 (8.50) 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 8.35 (Unit speed normal extremals)*.* Due to the homogeneity property one can introduce the cylinder Λ_{q_0} of normalized covectors

$$\Lambda_{q_0} = \{\lambda \in T^*_{q_0}M |\ H(\lambda) = 1/2\},$$

and consider the following exponential map with 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 8.36. *If the metric space (M, d) is complete, then $\mathscr{A}_{q_0} = T^*_{q_0}M$. Moreover, if there are no strictly abnormal 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 gives a contradiction to the fact that $\lambda(t)$ is not extendable. By Hamilton equations (4.34)

$$\dot{p}(t) = -\frac{\partial H}{\partial x}(p(t), x(t)) = -\sum_{i=1}^{m} \langle p(t), f_i(\gamma(t))\rangle \langle p(t), D_x f_i(\gamma(t))\rangle .$$

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 $i = 1, \dots, m$. Moreover by smoothness of f_i, the derivatives $|D_x f_i| \leq C$ are locally bounded in the neighborhood 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 minimizers, cf. Proposition 3.44 and Corollary 3.46. □

Corollary 8.37. *If the metric space (M, d) is complete, then every normal extremal trajectory is extendable on* $[0, +\infty[$.

We end this section by an Hamiltonianian version of the Gauss' Lemma.

Proposition 8.38 (Cotangent Gauss' Lemma)**.** *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 $q := \exp_{q_0}(\lambda_0)$.*

Proof. It is enough to show that for every smooth variation $\eta^s \in \Lambda_{q_0}, s \in [0, 1]$, of initial covectors such that $\eta^0 = \lambda_0$ we have

$$\left\langle \lambda(1), \left.\frac{d}{ds}\right|_{s=0} \exp_{q_0}(\eta^s) \right\rangle = 0.$$

Let $\eta^s(\tau) := e^{\tau\vec{H}}(\eta^s)$ and $\gamma^s(t) = \pi(\eta^s(t))$ be the corresponding trajectory. Define the family of controls $u^s(\cdot)$ satisfying for a.e. $\tau \in [0, 1]$

$$u_i^s(\tau) := \langle \eta^s(\tau), f_i(\gamma^s(\tau)) \rangle\,, \qquad i = 1, \dots, m, \tag{8.51}$$

where $f_1, \dots, f_m$ denotes as usual a generating frame. By definition (8.51) of u^s we have $\exp_{q_0}(\eta^s) = E_{q_0}(u^s)$ hence we can compute

$$\left.\frac{d}{ds}\right|_{s=0} \exp_{q_0}^t(\eta^s) = \left.\frac{d}{ds}\right|_{s=0} E_{q_0}^t(u^s) = D_u E_{q_0}(v), \tag{8.52}$$

where we denoted $v := \left.\frac{d}{ds}\right|_{s=0} u^s$. Notice that v is orthogonal to u in L^2 since, by Lemma 4.28 the map $s \mapsto \|u^s\|_{L^2}^2$ is constant. Thus we have

$$\left\langle \lambda(1), \left.\frac{d}{ds}\right|_{s=0} \exp_{q_0}(\eta^s) \right\rangle = \langle \lambda(1), D_u E_{q_0}(v) \rangle = (u, v)_{L^2} = 0, \tag{8.53}$$

where the second identity follows from the normal condition (8.18) and (8.52). □

Exercise 8.39. Deduce from Proposition (8.38) 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 $q_t := \exp_{q_0}^t(\lambda_0)$.

We end this section with 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 diffemorphism 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

$$D_0\exp_{q_0}(\xi) = \left.\frac{d}{dt}\right|_{t=0} \exp_{q_0}(0 + t\xi) = \left.\frac{d}{dt}\right|_{t=0} \gamma_\xi(t) = \dot{\gamma}_\xi(0). \tag{8.54}$$

where γ_ξ 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 (4.37) one has

$$\dot{\gamma}_\xi(0) = \sum_{i=1}^m \langle \xi, f_i(q_0) \rangle\, f_i(q_0). \tag{8.55}$$

Since $\xi \in T_{q_0}^*M$ is arbitrary, the proof is completed. □

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 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 11.8 and Theorem 12.17, singularities appear naturally.

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.

Definition 8.41. 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 $s\lambda$ 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 $s = \inf\{\tau > 0 \,|\, \tau\lambda \text{ is a critical point of } \exp_{q_0}\}$.

We denote by Con_{q_0} the set of all first conjugate points to q_0 along some normal extremal trajectory starting from q_0.

Remark 8.42. Notice that, given a normal extremal trajectory $\gamma : [0,1] \to M$ defined by $\gamma(t) = \exp_{q_0}(t\lambda_0)$, if γ 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 γ is a critical point for E_{q_0}, i.e., the differential $D_uE_{q_0}$ is not surjectuve. Since, by definition of the exponential map, one has $\operatorname{im} D_{\lambda_0}\exp_{q_0} \subset \operatorname{im} D_uE_{q_0}$, 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 8.42 is saying that an abnormal extremal is made of conjugate points. The following theorem discuss somehow a converse statement.

Theorem 8.43. *Let $\gamma : [0,T] \to M$ be a normal extremal path. Assume that $t_0 > 0$ is a limit of a decreasing (resp. increasing) sequence of conjugate times. Then there exists $\varepsilon > 0$ such that*

(a) all points of the segment $[t_0, t + \varepsilon]$ (resp. $[t_0 - \varepsilon, t_0]$) are conjugate,

(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 of conjugate times and leave to the reader to make necessary modifications in the case of an increasing sequence.

Let $(u(t), \lambda(t)),\ 0 \le t \le T$, be a normal extremal, where

$$\gamma(t) = \pi(\lambda(t)), \quad \dot\gamma = f_u(\gamma).$$

We set $P_{0,t} = \overrightarrow{\exp} \int_0^t f_{u(\tau)}\, d\tau$. We consider the maps

$$\mathcal{F}_t : \lambda \mapsto \pi \circ P^*_{0,t} \circ e^{\vec{H}}(t\lambda)$$

defined on a neighborhood of λ_0 in $T^*_{q_0}M$, where $q_0 = \gamma(0)$. According to the construction, $\mathcal{F}_t(\lambda_t) = \lambda_0$ for all t. I claim that $t \in (0,T]$ is a conjugate time for γ if and only if λ_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 $\exp_{q_0} = \pi \circ e^{\vec{H}}\big|_{T^*_{q_0}M}$, i.e., if $T_{\lambda(t)}e^{\vec{H}}(T^*_{q_0}M) \cap T_{\lambda(t)}(T^*_{\gamma(t)}M) \neq 0$, and the diffeomorphism $P^*_{0,t}$ transforms $T^*_{\gamma(t)}M$ into $T^*_{q_0}M$.

As we know, $(P^*_{0,t})^{-1} = \overrightarrow{\exp}\int_0^t \vec{h}_{u(t)}\,dt$, where $h_u(\lambda) = \langle \lambda, f_u \rangle$. The variations formula and formula (4.64???) imply that the depending on $t \in [0,T]$ family of diffeomorphisms

$$\lambda \mapsto P^*_{0,t} \circ e^{\vec{H}}(t\lambda) = P^*_{0,t} \circ e^{t\vec{H}}(\lambda), \quad \lambda \in T^*M,$$

is a time-varying Hamiltonian flow generated by the Hamiltonian $g_t : T^*M \to \mathbb{R}$ defined by

$$g_t := (H - h_{u(t)}) \circ (P^*_{0,t})^{-1}.$$

We have: $g_t \geq 0$ and $g_t(\lambda_0) = 0$. It follows that $d_{\lambda_0} g_t = 0$ and $d^2_{\lambda_0} g_t$ is a nonnegative quadratic form on the symplectic space $T_{\lambda_0}(T^*M)$. We introduce the following notations:

$$\Sigma := T_{\lambda_0}(T^*M), \qquad \Pi := T_{\lambda_0}(T^*_{q_0}M), \qquad Q_t := \frac{1}{2} d^2_{\lambda_0} g_t. \tag{8.56}$$

The linear Hamiltonian flow $\overrightarrow{\exp}\int_0^t \vec{Q}_\tau\,d\tau$ on Σ is the linearization of the flow $\overrightarrow{\exp}\int_0^t \vec{g}_\tau\,d\tau$ at the equilibrium λ_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 Σ are n-dimensional subspaces on which the symplectic form σ vanishes identically. In particular, Π is a Lagrange subspace. J_t is also a Lagrange subspace because symplectic flows preserve the symplectic form. A Darboux basis for Σ is a basis $e_1, \ldots, e_n, f_1, \ldots, f_n$ satisfying

$$\sigma(e_i, f_j) = \delta_{ij}, \quad \sigma(f_i, f_j) = \sigma(e_i, e_j) = 0, \qquad i, j = 1, \ldots, n. \tag{8.57}$$

We'll need the following simple lemma:

Lemma 8.44. *Let Λ_0, Λ_1 be Lagrange subspaces of Σ, with $\dim(\Lambda_0 \cap \Lambda_1) = k$. Then there exist Darboux basis $e_1, \ldots, e_n, f_1, \ldots, f_n$ in Σ such that*

$$\Lambda_0 = \operatorname{span}\{e_1, \ldots, e_n\}, \quad \Lambda_1 = \operatorname{span}\{e_1, \ldots, e_k, e_{k+1} + f_{k+1}, \ldots, e_n + f_n\}.$$

Proof. Consider any arbitrary basis $e_1, \ldots, e_n$ of Λ_0 satisfying

$$\Lambda_0 \cap \Lambda_1 = \operatorname{span}\{e_1, \ldots, e_k\}.$$

The nondegeneracy of σ implies the existence of $f_1 \in \Sigma$ such that

$$\sigma(e_1, f_1) = 1, \qquad \sigma(e_2, f_1) = \cdots = \sigma(e_n, f_1) = 0.$$

Chosen f_1, the nondegeneracy of σ 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 $f_1, \dots, f_k$ 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

$$\Gamma = \{v \in \Lambda_1 : \sigma(f_1, v) = \cdots = \sigma(f_n, v) = 0\}.$$

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 $e_1, \dots, e_n, f_1, \dots, f_n$. □

We apply the previous lemma to the pair of Lagrange subspaces Π 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 nonnegative 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 close to J_{t_0} n-dimensional subspace Λ 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, Λ is a Lagrange subspace if and only if A is a symmetric matrix. Indeed,

$$\sigma((p_1, Ap_1), (p_2, Ap_2)) = p_1^T A p_2 - p_2^T A p_1 = p_1^T (A - A^*) p_2.$$

where v^T denotes the transpose of a vector v. Let $J_t = \{(p, S_t p) : p \in \mathbb{R}^n\}$ for t close to t_0; then S_t is a symmetric matrix smoothly depending on t. Moreover,

$$\Pi \cap J_t = \{(p, 0) \in \mathbb{R}^n \times \mathbb{R}^n : S_t p = 0\}.$$

Lemma 8.45. *For every $p \in \mathbb{R}^n$ one has $p^T \dot{S}_t p \geq 0$.*

Proof. We keep symbol Q_t for the matrix of the quadratic form Q_t on Σ. 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:

$$\sigma((p, S_t p), (\dot{p}, S_t \dot{p}) + (0, \dot{S}_t p)) = \langle p, \dot{S}_t p \rangle \geq 0.$$

□

Lemma 8.46. *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 8.45. Indeed,

$$0 \leq \langle S_{t_0} \bar{p}, \bar{p} \rangle \leq \langle S_t \bar{p}, \bar{p} \rangle \leq \langle S_{t_1} \bar{p}, \bar{p} \rangle = 0.$$

Hence $\langle S_t \bar{p}, \bar{p} \rangle = 0$. Since $p \mapsto \langle S_t p, p \rangle$ is a nonnegative quadratic form, we obtain that $S_t \bar{p} = 0$. □

Lemma 8.46 implies claim (a) of the theorem (for a decreasing sequence). Let us prove claim (b), whose proof is also based on Lemma 8.46.

The fiber $T_{q_0}^*M$ is a vector space, it is naturally identified with its tangent space Π, and the coordinates $p \in \mathbb{R}^n$ on Π introduced above serve as coordinates on $T_{q_0}^*M$. The restriction of the Hamiltonian g_t to $T_{q_0}^*M$ has a form:

$$g_t(p) = \frac{1}{2}\sum_{i=1}^{k}\langle p, (P_{0,t*}^{-1} f_i)(q_0)\rangle^2 - \langle p, (P_{0,t*}^{-1} f_{u(t)})(q_0)\rangle.$$

Hence

$$\langle Q_t(p,0),(p,0)\rangle = \frac{1}{2}\sum_{i=1}^{k}\langle p, (P_{0,t*}^{-1} f_i)(q_0)\rangle^2. \tag{8.58}$$

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 8.46, 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 (8.58),

$$\langle \bar{p}, (P_{0,t*}^{-1} f_i)(q_0)\rangle = 0, \quad i = 1,\ldots,k, \quad t \in [t_0,t_1].$$

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 8.43 the following important corollary.

Corollary 8.47. *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*

$$\mathcal{T}_c := \{t > 0 \mid \gamma(t) \text{ is conjugate to } \gamma(0)\}.$$

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, that is the topology induced on the space of horizontal trajectories by the L^2 norm on the space of controls,
- the C^0 topology, also called strong topology, that is the usual uniform topology on the space of continuous curves on M.

The main result of this section is the following one.

Theorem 8.48. *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.*

Remark 8.49. Notice that claim (i) of Theorem 8.48 is a direct consequence of Corollary 8.47. Nevertheless we will obtain in this section an independent proof. The proof of part (ii) and (iii) need some preliminary results.

Some of these preliminary results holds true under weaker assumptions. For the sake of simplicity in this section we state them for normal extremal trajectory that does not contain abnormal segments. A discussion on the validity of these statements under different assumptions is contained in Exercice 8.54.

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 λ 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 8.50. *Assume that* $\gamma_u : [0,1] \to M$ *contains no abnormal segments. Then* $\gamma_u(s)$ *is conjugate to* $\gamma_u(0)$ *if and only if* $\mathrm{Hess}_{u^s} J|_{E_{q_0}^{-1}(\gamma^s(1))}$ *is a degenerate quadratic form.*

Proof. Since the curve contains no abnormal segments, the control u^s is a regular point for the end-point map. Hence, thanks to Proposition 8.26 combined with Proposition 8.29 and Corollary 8.31, one has that $\gamma_u(s)$ is conjugate to $\gamma_u(0)$ if and only if λ^s is a critical point of the exponential map, that is equivalent to the fact that $\mathrm{Hess}_{u^s} J\big|_{E_{q_0}^{-1}(\gamma^s(1))}$ is degenerate. ☐

The following lemma, studying the family of quadratic form $s \mapsto \mathrm{Hess}_{u_s} J\big|_{E_{q_0}^{-1}(\gamma^s(1))}$, is crucial in what follows.

Lemma 8.51. *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*

$$\alpha(s) := \inf\left\{\|v\|^2_{L^2} - \left\langle \lambda^s, D^2_{u^s}E_{q_0}(v)\right\rangle \mid \|v\|^2_{L^2} = 1,\ v \in \ker D_{u^s}E_x\right\}. \tag{8.59}$$

Then α *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) α *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 8.51. Notice that one can write

$$\|v\|_{L^2}^2 - \lambda^s \circ D_{u^s}^2 E_{q_0}(v) = \langle (I - Q_s)(v)|v\rangle_{L^2}, \tag{8.60}$$

where $Q_s : L^2([0,1],\mathbb{R}^m) \to L^2([0,1],\mathbb{R}^m)$ is a compact and symmetric operator thanks to Lemma 8.30. A compact symmetric operator on a Hilbert space is diagonalizable and the set of eigenvalues is countable $\{\mu_n\}_{n\in\mathbb{N}}$, bounded, and can be ordered in such a way that $\mu_n \to 0$ (see [71, III Thm. 6.26]). As a consequence, one can prove that the infimum in (8.59) is attained.

Observe 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 norm topology), we reduce (8.59) to

$$\alpha(s) = 1 - \sup\{\langle \phi_s \circ Q_s \circ \phi_s^{-1}(w)|w\rangle_{\mathcal{H}} \mid w \in \mathcal{H},\ \|w\|_{\mathcal{H}} = 1\}, \tag{8.61}$$

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 [71, V Thm. 4.10]). This proves that α is continuous.

Let us recall that

$$D_{u^s}E_{q_0}(v) = \int_0^s (P_{t,1})_* f_{v(t)}|_{\gamma_u(s)} dt, \tag{8.62}$$

$$D_{u^s}^2 E_{q_0}(v,v) = \iint\limits_{0\le\tau\le t\le s} [(P_{\tau,1})_* f_{v(\tau)}, (P_{t,1})_* f_{v(t)}]|_{\gamma_u(s)} d\tau dt. \tag{8.63}$$

By a rescaling one can see that

$$D_{u^s}E_{q_0}(v) = s\int_0^1 (P_{st,1})_* f_{v(st)}|_{\gamma_u(s)} dt, \tag{8.64}$$

$$D_{u^s}^2 E_{q_0}(v,v) = s^2 \iint\limits_{0\le\tau\le t\le 1} [(P_{s\tau,1})_* f_{v(s\tau)}, (P_{st,1})_* f_{v(st)}]|_{\gamma_u(s)} d\tau dt. \tag{8.65}$$

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 that there exists a sequence $v_n \in \ker D_{u^{\bar{s}}}E_{q_0}$ of controls with $\|v_n\| = 1$ and such that $\|v_n\|_{L^2}^2 - \langle Q_{\bar{s}}(v_n)|v_n\rangle_{L^2} \to 0$ 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 $\bar{v}$. By compactness of $Q_{\bar{s}}$, we deduce that $\langle Q_{\bar{s}}(\bar{v})|\bar{v}\rangle_{L^2} = 1$. Since $\|\bar{v}\|_{L^2}^2 \leq 1$, we have $\langle (I - Q_{\bar{s}})(\bar{v})|\bar{v}\rangle_{L^2} = 0$. 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.

Exercise 8.52. 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 $\mathcal{Q}(\bar{v},\cdot) = 0$. Prove that a non negative quadratic form is degenerate if and only if there exists $\bar{v}$ such that $\mathcal{Q}(\bar{v},\bar{v}) = 0$.

To prove (c) let us fix $0 \leq s \leq s' \leq 1$ and $v \in \ker D_{u^s} E_x$. Define

$$\widehat{v}(t) := \begin{cases} \sqrt{\dfrac{s'}{s}}\, v\left(\dfrac{s'}{s} t\right), & 0 \leq t \leq \dfrac{s}{s'}, \\ 0, & \dfrac{s}{s'} < t \leq 1. \end{cases}$$

It follows that $\|\widehat{v}\|_{L^2}^2 = \|v\|_{L^2}^2$, $\widehat{v} \in \ker D_{u^{s'}} E_x$, 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 $\bar{s} \leq s \leq s_1$. 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 8.43, the segment $\gamma|_{[\bar{s}, s_1]}$ is also abnormal, contradicting the assumption on γ. $\square$

Proof of Theorem 8.48. Thanks to Lemma 8.51 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 8.50. This proves claim (i).

To prove claim (ii) notice that if $\gamma|_{[0,s]}$ does not contain conjugate points, by Proposition 8.50 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 8.51.

Let $\tau > t_c$ and assume by contradiction that the trajectory is a length-minimizer. Then, using the terminology of Lemma 8.51, 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. $\square$

Remark 8.53. Notice that claim (i) of Theorem 8.48 is also an immediate consequence of Corollary 8.47. However the previous argument gives another proof which is independent on the argument contained in the proof of Theorem 8.43 in the previous section.

Exercise 8.54. Introduce the following definitions: a normal extremal trajectory $\gamma : [0, T] \to M$ is said to be

- *left strongly normal*, if for every $s \in (0, T]$ the curve $\gamma|_{[0,s]}$ does not admit abnormal lifts.
- *right strongly normal*, if for every $s \in [0, T)$ the curve $\gamma|_{[s,T]}$ does not admit abnormal lifts.
- *strongly normal*, if γ 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 8.48 claim (i)-(ii), Proposition 8.50, Lemma 8.51 claim (a)-(b)-(c), hold under the weaker assumption that the normal extremal trajetory is left strongly normal.

8.8.1 Local length-minimality in the strong topology

A direct consequence of Theorem 8.48 proved in the previous section is the following.

Corollary 8.55. *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 γ is a local miminum for the length in the $W^{1,2}$ topology in the space of admissible trajectories with the same endpoints.*

The main goal of this section is to prove that indeed the same conclusion holds true in the uniform topology. The proof of this result, which is based upon the arguments of Theorem 4.61, 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, given a final point $q_1 \in M$, if the initial point q_0 is not fixed the minimization problem

$$\min_{q \in M, u \in E_q^{-1}(q_1)} J \tag{8.66}$$

has only the trivial solution $(q,u) = (q_1, 0)$.

In this case it is meaningful to introduce a penalty function $a \in C^\infty(M)$ and consider the minimization problem

$$\min_{q \in M, u \in E_q^{-1}(q_1)} J(u) + a(q) \tag{8.67}$$

Let us introduce the *extendend end-point map*

$$\mathbb{E} : M \times \mathcal{U} \to M, \qquad (q,u) \mapsto E_q(u),$$

where $E_q(u)$ is the end-point map based at q. Notice that $\mathbb{E}$ is trivially a submersion since for every $q \in M$ one has $\mathbb{E}(q,0) = q$. Moreover denoting $P^u_{t,s}$ the nonautonomous flow associated with u one has

$$\mathbb{E}\big|_{\{q_0\}\times\mathcal{U}} = E_{q_0}, \qquad \mathbb{E}\big|_{M\times\{u\}} = P^u_{0,1}. \tag{8.68}$$

The minimization problem (8.67) is then rewritten as

$$\min_{\mathbb{E}^{-1}(q_1)} \varphi \tag{8.69}$$

where $\varphi : M \times \mathcal{U} \to \mathbb{R}$ is defined by $\varphi(q,u) := J(u) + a(q)$ and choosing $F = \mathbb{E}$ this constrained minimization problem is of the type studied in Section 8.4.[2]

Notice that every level set $\mathbb{E}^{-1}(q_1)$ is regular since the map $\mathbb{E}$ is a submersion. The Lagrange multiplier equation (8.22) is rewritten as follows: the point $(q_0, u) \in M \times \mathcal{U}$ is a critical point of the problem (8.69) if and only if there exists a $\lambda_1 \in T^*M$ such that

$$\lambda_1 D_{(q_0,u)}\mathbb{E} = D_{(q_0,u)}(J + a) \tag{8.70}$$

Since the differentials $D_{(q_0,u)}\mathbb{E}$ and $D_{(q_0,u)}(J+a)$ are defined on the product space $T_{(q_0,u)}M \times \mathcal{U} \simeq T_{q_0}M \times \mathcal{U}$, and thanks to the identity

$$D_{(q_0,u)}\mathbb{E} = (D_u E_{q_0}, (P^u_{0,1})_*), \qquad D_{(q_0,u)}(J+a) = (D_u J, d_{q_0}a)$$

[2]to be precise, here the problem is defined on a Hilber manifold and not on a subspace an Hilber space, but since M is finite dimensional the theory applies with essentially no modifications.

the equation (8.70) 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, to every critical point of the problem (8.69) we can associate a normal extremal

$$\lambda(t) = (P^{-1}_{0,t})^* \lambda_1,$$

where the initial condition is defined by the function a by $\lambda_0 = d_{q_0} a$.

Proposition 8.56. *A point $(q_0, u) \in M \times \mathcal{U}$ is a critical point of the problem (8.69) if and only if the corresponding horizontal trajectory $\gamma_u(t)$ 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 $F^{-1}(q_1)$ but considers a penalty in the functional at the end-point.

Exercise 8.57. Fix $q_0 \in M$ and $a \in C^\infty(M)$. Prove that every critical point $\bar{u} \in \mathcal{U}$ of the *free endpoint* problem

$$\min_{u \in \mathcal{U}} J(u) - a(E_{q_0}(u)), \tag{8.71}$$

we can associate a normal extremal trajectory satisfying

$$\lambda D_{\bar{u}} F = \bar{u}, \qquad \lambda = d_{F(\bar{u})} a.$$

Proof of local length-minimality in the strong topology

We can now prove the following result.

Proposition 8.58. *Let $\gamma : [0, 1] \to M$ be a normal extremal trajectory that does not contain abnormal segments. If γ does not contain conjugate points, then it is a local miminum for the length in the C^0 topology in the space of admissible trajectories with the same endpoints.*

Proof. Assume that

$$\gamma(t) = \pi \circ e^{t\vec{H}}(\lambda_0), \qquad \lambda_0 \in T^*_q M$$

We want to show that hypothesis of Theorem 4.61 are satisfied. We will use the following lemma, which we prove at the end of the proposition.

Lemma 8.59. *There exists $a \in C^\infty(M)$ such that*

$$\lambda_0 = d_{q_0} a, \qquad \mathrm{Hess}_{(q_0,u)} J + a\Big|_{\mathbb{E}^{-1}(\gamma_s)} > 0,$$

Moroever $(\mathbb{E}, J + a)$ is a Morse problem and

$$\mathcal{L}_{(\mathbb{E}, J+a)} = \{e^{\vec{H}}(d_q a),\ q \in M\}$$

From this Lemma it follows that $s\lambda_0$ is a regular point of the map $\pi \circ e^{\vec{H}}\big|_{\mathcal{L}_0}$, where as usual $\mathcal{L}_0 = \{d_q a, q \in M\}$ denotes the graph of the differential. Using the homogeneity property (8.50) we can rewrite this saying that

$$\pi \circ e^{s\vec{H}}\big|_{\mathcal{L}_0} \quad \text{is an immersion at } \lambda_0, \quad \forall\, s \in [0,1],$$

In particular it is a local diffeomorphism. Hence we can apply the local version of Theorem 4.61. $\square$

We end the section with the proof of the technical lemma.

Proof of Lemma 8.59. First we notice that

$$\ker D_{(q_0,u)}\mathbb{E} \subset T_{q_0}M \oplus L^2([0,1],\mathbb{R}^m)$$

In particular

$$\ker D_{(q_0,u)}\mathbb{E} \cap (0 \oplus L^2([0,1],\mathbb{R}^m)) = \ker D_u E_{q_0}$$

Since there are no conjugate points, it follows that

$$\mathrm{Hess}_{(q_0,u)} J + a\Big|_{0\oplus \ker D_u E} = \mathrm{Hess}_u J > 0 \tag{8.72}$$

Then it is sufficient to show that there exists a choice of the function $a \in C^\infty(M)$ such that the Hessian is positive definite also in the complement. We define

$$W_s := \{\xi \oplus v \in \ker D_{(q_0,u_s)}\mathbb{E} \mid \mathrm{Hess}(J+a)(\xi\oplus v, 0 \oplus \ker D_{u_s}E) = 0\}$$

Notice from (8.72) that, if there is some $\xi \oplus v \in W_s$, then $\xi \neq 0$. Now we prove the existence of a map $B_s : T_qM \to L^2([0,1],\mathbb{R}^m)$ such that

$$W_s = \{\xi \oplus B_s\xi \mid \xi \in T_qM\}$$

Then we will have

$$\ker D_{(q_0,u_s)}\mathbb{E} = (0 \oplus \ker D_{u_s}F) + W_s.$$

Let us compute

$$\begin{aligned} \mathrm{Hess}(J+a)(\xi \oplus B_s\xi + 0 \oplus v, \xi \oplus B_s\xi + 0 \oplus v) &= \\ &= \mathrm{Hess}J(v,v) + \mathrm{Hess}(J+a)(\xi \oplus B_s\xi, \xi \oplus B_s\xi) \\ &= \mathrm{Hess}J(v,v) + d^2a(\xi,\xi) + Q(\xi) \end{aligned}$$

where we used that mixed terms give no contribution and denote with $Q(\xi)$ a quadratic form that does not depend on second derivatives of a. In particular, since the first term is positive and does not depend on ξ, we can choose a in such a way that it remains positive. $\square$

Combining the results obtained in the previous sections we have the following result.

Theorem 8.60. *Let $\gamma : [0,1] \to M$ be a normal extremal trajectory that does not contain abnormal segments.*

(i) if γ has no conjugate point then its a local length-minimizer in the C^0 topology in the space of admissible trajectories with the same endpoints,

(ii) if γ has at least a conjugate point then its not a local length-minimizer in the $W^{1,2}$ topology in 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.

Moreover, thanks to reparametrization, we assume that trajectories are parametrized by constant speed on the interval $[0,1]$. Notice that in this case if γ_u is the horizontal curve corresponding to a control u one has $\ell(\gamma_u) = \|u\|_{L^1} = \|u\|_{L^2}$. Recall that

$$\|u\|_{L^1} = \int_0^1 |u(t)|dt, \qquad \|u\|_{L^2} = \left(\int_0^1 |u(t)|^2 dt\right)^{\frac{1}{2}}.$$

where $|\cdot|$ denotes the standard norm on $\mathbb{R}^m$.

Proposition 8.61. *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-L^2 topology then $E_{q_0}(u_n) \to E_{q_0}(u)$.*

Proof. First notice that since $u_n \rightharpoonup u$ in the weak-L^2 topology 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 γ_n of the Cauchy problem

$$\dot{\gamma}(t) = f_{u_n(t)}(\gamma(t)), \qquad \gamma(0) = q_0$$

satisfies the integral identity

$$\gamma_n(t) = q_0 + \int_0^t f_{u_n(\tau)}(\gamma_n(\tau))d\tau, \tag{8.73}$$

Since $\|u_n\| \leq r_0$ for every n, all trajectories γ_n are contained in the compact ball B, they are Lipschitzian with the same Lipchitz constant. In particular 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 γ_n) and a limit continuous curve γ such that $\gamma_n \to \gamma$ uniformly. Let us show that γ is the horizontal trajectory associated to u.

Since u_n weakly converges to u we have that $f_{u_n(t)}(\gamma_n(t)) \to f_{u(t)}(\gamma(t))$, since this can be seen as a product between strongly and weakly convergent sequences.[3] Passing to the limit for $n \to \infty$ in (8.73), one finds that

$$\gamma(t) = q_0 + \int_0^t f_{u(\tau)}(\gamma(\tau))d\tau,$$

namely that γ is the trajectory associated to u. This completes the proof. □

Remark 8.62. 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 minimizers, cf. Theorem 3.40.

Corollary 8.63 (Existence of 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 γ_u joins q_0 and q and is a minimizer, i.e., $\ell(\gamma_u) = d(q_0, q)$.*

[3] writing the coordinate expression $\sum_{i=1}^m u_{n,i} f_i(\gamma_n(t))$.

Proof. Consider a point q in the compact ball B. Then take a minimizing sequence u_n such that $E_{q_0}(u_n) = q$ and $\|u_n\|_{L^2} \to d(q_0, q)$. The sequence $(\|u_n\|_{L^2})_n$ is bounded, hence by weak compactness of balls in L^2 there exists a subsequence, still denoted by the same symbol, such that $u_n \rightharpoonup u$ for some u. By weak continuity $E_{q_0}(u) = q$. Moreover the semicontinuity of the L^2 norm proves that u corresponds to a minimizer joining q_0 to q since

$$\|u\|_{L^2} \leq \liminf_{n\to\infty} \|u_n\|_{L^2} = d(q_0, q).$$

□

Definition 8.64. 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 8.65 (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 L^2 topology.

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 8.61 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

$$\|u\|_{L^2} \leq \liminf_{n\to\infty} \|u_n\|_{L^2} = \liminf_{n\to\infty} d(q_0, E_{q_0}(u_n)) = d(q_0, E_{q_0}(u)). \tag{8.74}$$

Since by definition of distance $d(q_0, E_{q_0}(u)) \leq \ell(\gamma_u) \leq \|u\|_{L^2}$ we have that all inequalities are equalities in (8.74), 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. □

This implies the following continuity property.

Proposition 8.66. *Let M be complete and assume that $q \in M$ is reached by a unique minimizer starting from q_0 associated with u. If u_n is any sequence of minimizer 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 u_{k_n} admit a convergent subsequence $u_{k_n} \to \widehat{u}$, for some control $\widehat{u} \in \mathcal{M}_K$. The trajectory corresponding to $\widehat{u}$ is a minimizer joining q_0 to q. Hence by uniqueness $\widehat{u} = u$.

This proves that every subsequence of u_n admits a subsequence converging to the same element u. A general topological argument implies that the whole sequence u_n converges to u. □

Remark 8.67. If M is not complete, all the results of this section holds 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 8.68. A horizontal trajectory $\gamma : [0, T] \to M$ is called a *geodesic* if it is parametrized by unit speed and for every $t \in [0, T]$ there exists $\varepsilon > 0$ such that $\gamma|_{[t-\varepsilon, t+\varepsilon]}$ realizes 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 what follows when we speak about a geodesic we always assume that it is maximal.

Recall that a normal extremal trajectory parametrized by unit speed is a geodesic by Theorem 4.63. When M is complete, it is extendable to $[0, +\infty[$ thanks to Corollary 8.37.

Exercise 8.69. Let γ be a geodesic. Introduce the set $A = \{t > 0 : \gamma|_{[0,t]}$ is length-minimizing$\}$. Prove that A is an interval either of the form $(0, t_*]$ or $(0, +\infty)$.

Definition 8.70. Let γ be a geodesic and define

$$t_* := \sup\{t > 0 : \gamma|_{[0,t]} \text{ is length-minimizing}\}.$$

If $t_* < +\infty$ we say that $\gamma(t_*)$ is the *cut point* of $\gamma(0)$ along γ. If $t_* = +\infty$ we say that γ has no cut point. We denote by Cut_{q_0} the set of all cut points of geodesics starting from a point $q_0 \in M$.

Cut points along geodesics detect the segments on which they are global length-minimizer. The following is the fundamental property of cut locus along normal extremal trajectories.

Theorem 8.71. *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. Suppose that there exists $t_0 \in (0, T)$ such that*

(a) either $\gamma(t_0)$ is the first conjugate point along γ,

(b) or there exists a length-minimizer $\widehat{\gamma} \neq \gamma$ joining $\gamma(0)$ and $\gamma(t_0)$ with $\ell(\widehat{\gamma}) = \ell(\gamma|_{[0,t_0]})$.

then there exist $t_* \in (0, t_0]$ such that $\gamma(t_*)$ is the cut point along γ.

Conversely, if $\gamma(t_0)$ is the cut point from $\gamma(0)$ along γ, then either (a) or (b) are satisfied.

Proof. Let us first assume that there exists $t_0 > 0$ such that (a) is satisfied and that the cut time t_* is strictly bigger than t_0. This implies that $\gamma|_{[0,t_*]}$ is a minimizer contradicting Theorem 8.60, claim (ii).

Assume now that assumption (b) is satisfied and there exists a 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 would built two different normal lifts of the normal extremal trajectory $\gamma|_{[t_0,T]}$, hence $\gamma|_{[t_0,T]}$ would be an abnormal segment, contradicting our assumption on γ.

Assume now that $\gamma(t_0)$ is the cut point from $\gamma(0)$ along γ 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 γ_n joining $\gamma(0)$ to $\gamma(t_n)$, namely $\ell(\gamma_n) = d(\gamma(0), \gamma(t_n))$.

By compactness of minimizers there exists (up to extraction of a convergent subsequence) a limit 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 uniform C_0 topology. Thus $\widehat{\gamma}$ cannot be contained in a neighborhood γ and necessarily $\widehat{\gamma} \neq \gamma$, ending the proof.

□

Theorem 8.72. *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 there every points 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. Indeed this fact, together with assumptions (i) and (ii), imply that the cut time t_* along γ satisfies $t_* \geq t_0 + \varepsilon$.

Fix a neighborhood U of $\gamma(t_0)$ and, for each $q \in U$, let us denote by u^q (resp. γ^q) the minimizing control (resp. trajectory) joining $\gamma(0)$ to q. Thanks to Proposition 8.66 the map $q \mapsto u^q$ is continuous in the L^2 topology.

Hence we can consider the family λ_1^q of normal final covectors associated with u^q, i.e., satisfying the identity

$$\lambda_1^q D_{u^q} F = u^q, \qquad \forall\, q \in U.$$

By the smoothness of the end-point map E_{q_0}, the map $q \mapsto D_{u^q} E_{q_0}$ is continuous and; 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 λ_1^q is the unique solution to the linear equation $(D_{u^q} F)^* \xi = u^q$ (unicity of covector is guaranteed since the trajectory is strict abnormal by assumption (ii)). Since the coefficient 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 λ_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, Φ^0 is a right inverse of the exponential map $\exp_{q_0}$. Thus the map Φ^0 is injective on U and, by the invariance of domain theorem, Φ^0 is an open map and $A := \{\Phi^0(q) \mid q \in U\}$ is an open set in $T_q^* M$ containing $\lambda_0^{\gamma(t_0)}$.

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 minimizer joining q_0 with $\gamma((1+\delta_0)t_0)$ is γ itself. Thus γ deos not contain conjugate points in the segment $[0, t_0 + \varepsilon]$ for every $\varepsilon < \delta_0 t_0$. □

We end this section by explicitly stating the converse of Theorem 8.72, in the case when the structure admits no abnormal minimizers.

Corollary 8.73. *Assume that the sub-Riemannian structure admits no abnormal 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 length-minimizers.*

Recall that, thanks to Theorem 8.71, if the sub-Riemannian structure admits no abnormals, points where geodesics lose global optimality can be of two types: (a) (first) conjugate points, or (b) points reached by two minimizers.

Corollary 8.73 says that, if there are no abnormal 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.

8.11 An example: the first conjugate locus on perturbed sphere

In this section we prove that a C^∞ 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 and lose their local and global 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^∞ 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 θ. 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. $t_c(\theta) = \inf\{\tau > 0 \,|\, \gamma(\tau,\theta) \text{ is conjugate to } q_0 \text{ along } \gamma(\cdot,\theta)\}$.

Proposition 8.74. *The first conjugate time* $t_c(\theta)$ *is characterized as follows*

$$t_c(\theta) = \inf\left\{ t > 0 \;\middle|\; \frac{\partial \exp}{\partial \theta}(t,\theta) = 0 \right\}. \tag{8.75}$$

Proof. Conjugate points correspond to critical points of the exponential map, i.e., points $\exp(t,\theta)$ such that

$$\operatorname{rank}\left\{ \frac{\partial \exp}{\partial t}(t,\theta), \frac{\partial \exp}{\partial \theta}(t,\theta) \right\} = 1. \tag{8.76}$$

Notice that $\frac{\partial \exp}{\partial t}(t,\theta) = \dot\gamma(t,\theta) \neq 0$. Let us show that condition (8.76) occurs only if $\frac{\partial \exp}{\partial \theta}(t,\theta) = 0$. Indeed, by Proposition 8.38, one has that

$$\left\langle p, \frac{\partial \exp}{\partial t}(t,\theta)\right\rangle = 1, \qquad \left\langle p, \frac{\partial \exp}{\partial \theta}(t,\theta)\right\rangle = 0,$$

thus, whenever $\frac{\partial \exp}{\partial \theta}(t,\theta) \neq 0$, the two vectors appearing in (8.76) are always linearly independent. □

Lemma 8.75. *The function $\theta \mapsto t_c(\theta)$ is C^1.*

Proof. By Proposition 8.74, $t_c(\theta)$ is a solution to the equation (with respect to t)

$$\frac{\partial \exp}{\partial \theta}(t,\theta) = 0. \tag{8.77}$$

Let us first remark that, for the exponential map $\exp_0$ associated with the Hamitonian H_0 we have

$$\frac{\partial \exp_0}{\partial \theta}(t_c^0(\theta),\theta) = 0, \qquad \frac{\partial^2 \exp_0}{\partial t \partial \theta}(t_c^0(\theta),\theta) \neq 0 \tag{8.78}$$

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^∞ topology, by continuity with respect to the data of solution of ODEs, we have that exp is close to $\exp_0$ in the C^∞ topology too. Moreover the condition (8.78) ensures the existence of a solution $t_c(\theta)$ of (8.77) that is close to $t_c^0(\theta)$. Hence we have that

$$\frac{\partial^2 \exp}{\partial t \partial \theta}(t_c(\theta),\theta) \neq 0 \tag{8.79}$$

By the implicit function the function $\theta \mapsto t_c(\theta)$ is 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 β. 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 8.75) the map β takes value in a neighborhood of the point $\widehat{q}_0$ antipodal to q_0. Let us take stereographic coordinates around this point and consider β as a function from S^1 to $\mathbb{R}^2$. By the chain rule and (8.77), we have

$$\beta'(\theta) = t_c'(\theta)\frac{\partial \exp}{\partial t}(t_c(\theta),\theta) + \underbrace{\frac{\partial \exp}{\partial \theta}(t_c(\theta),\theta)}_{=0} \tag{8.80}$$

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

$$C_0 = \{\rho g_0(\theta) \mid \theta \in S^1, \rho \in [0,1]\}$$

is 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^∞-topology, hence

$$C = \{\rho g(\theta) \mid \theta \in S^1, \rho \in [0,1]\}, \tag{8.81}$$

remains convex.

Theorem 8.76. *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 $S^1 = [0, 2\pi]/\sim$. Moreover

$$\int_0^{2\pi} t_c'(\theta) d\theta = t_c(2\pi) - t_c(0) = 0. \tag{8.82}$$

A function with zero integral mean on $[0, 2\pi]$ which is not identically zero has to change sign at least twice on the interval. Notice also that

$$\int_0^{2\pi} t_c'(\theta) g(\theta) d\theta = \int_0^{2\pi} \beta'(\theta) d\theta = \beta(2\pi) - \beta(0) = 0. \tag{8.83}$$

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 (8.83). $\square$

Remark 8.77. 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 (8.81).

The same argument can be applied for every arbitrary small C^∞ (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.

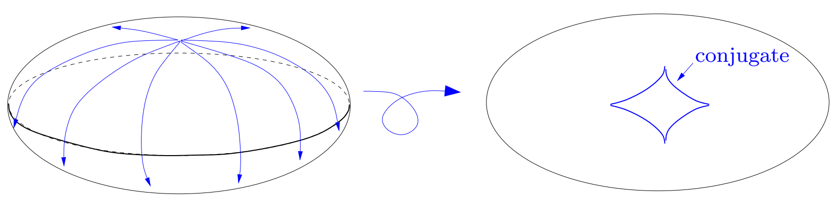


Figure 8.2: Perturbed sphere or ellipsoid

Chapter 9

2D-Almost-Riemannian Structures

Almost-Riemannian structures are examples of sub-Riemannian strucures such that the local minimum bundle rank (cf. Definition 3.20) 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 9.19), 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 3.28, 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 $(\mathbf{U}, f)$ where

- $\mathbf{U}$ is an Euclidean bundle over M of rank 2. We denote each fiber by U_q, the scalar product on U_q by $(\cdot \mid \cdot)_q$ and the norm of $u \in U_q$ as $|u| = \sqrt{(u \mid u)_q}$.
- $f : \mathbf{U} \to TM$ 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.

As for a general sub-Riemannian structure, we define:

- the *distribution* as $\mathcal{D}(q) = \{X(q) \mid X \in \mathcal{D}\} = f(U_q) \subseteq T_qM$,
- the *norm* of a vector $v \in \mathcal{D}_q$ as $\|v\| := \min\{|u|, u \in U_q \text{ s.t. } v = f(q,u)\}$.

- *admissible curve* as a Lipschitz curve $\gamma : [0,T] \to M$ such that there exists a measurable and essentially bounded function $u : t \in [0,T] \mapsto u(t) \in U_{\gamma(t)}$, 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.

- *minimal control* of an admissible curve γ as $u^*(t) := \text{argmin}\{|u|,\, u \in U_{\gamma(t)}$ s.t. $\dot\gamma(t) = f(\gamma(t), u)\}$ (for all differentiability point of γ). Recall that the minimal control is measurable (cf. Section 3.5)

- *(almost-Riemannian) length* of an admissible curve $\gamma : [0,T] \to M$ as $\ell(\gamma) := \int_0^T \|\dot\gamma(t)\| dt = \int_0^T |u^*(t)| dt$.

- *distance* between two points $q_0, q_1 \in M$ as

$$d(q_0, q_1) = \inf\{\ell(\gamma) \,|\, \gamma : [0,T] \to M \text{ admissible, } \gamma(0) = q_0,\ \gamma(T) = q_1\}. \tag{9.1}$$

Recall that thanks to the Lie-bracket generating condition, the Chow-Rashevskii Theorem 3.30 guarantees that (M,d) is a metric space and that the topology induced by (M,d) is equivalent to the manifold topology.

Definition 9.2. If (σ_1, σ_2) is an orthonormal frame for $(\cdot\,|\,\cdot)_q$ on a local trivialization $\Omega \times \mathbb{R}^2$ of $\mathbf{U}$, an *orthonormal frame* for the 2D-almost-Riemannian structure on Ω is the pair of vector fields $(F_1, F_2) := (f \circ \sigma_1, f \circ \sigma_2)$. In $\Omega \times \mathbb{R}^2$ the map f can be written as $f(q,u) = u_1 F_1(q) + u_2 F_2(q)$. When this can be done globally, we say that the 2D-almost-Riemannian structure is *free*.

In this chapter we do not work with an equivalent structure of higher bundle rank that is free. Technically such a structure fits Definition 3.20 (i.e., that local minimum bundle rank is equal to the dimension of the manifold at each point) but not Definition 9.1. We rather work with local orthonormal frames that, as explained below, are orthonormal in the standard sense out of the singular set.

This point of view permits to understand how global properties of $\mathbf{U}$ (as its orientability, its topology) are transferred in properties of the almost-Riemannian structure.

Definition 9.3. A 2D-almost-Riemannian structure $(\mathbf{U}, f)$ over a 2D manifold 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.

Remark 9.4. 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 positive oriented orthonormal frames defined respectively on two open subsets Ω and Ξ, 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 $(x, -\pi) \sim (x, \pi)$, $(-\pi, y) \sim (\pi, -y)$.

Let $\mathbf{U} = M \times \mathbb{R}^2$ with the standard Euclidean metric and consider the morphism of vector bundles given by

$$f : \mathbf{U} \to TM, \quad f(x_1, x_2, u_1, u_2) = (x_1, x_2, u_1, u_2 \sin(2x_1)).$$

This structure is Lie bracket generating and the two vector fields

$$F_1(x_1, x_2) = f(x_1, x_2, 1, 0) = (x_1, x_2, 1, 0), \quad F_2(x_1, x_2) = (x_1, x_2, 0, \sin(2x_1)),$$

which are well defined on M, provide a global orthonormal frame. This structure is orientable since $\mathbf{U}$ is trivial.

Exercise 9.5. Construct a non orientable almost-Riemannian structure on the 2D torus.

We now define Euler number of $\mathbf{U}$ that measures how far the vector bundle $\mathbf{U}$ is from the trivial one.

Definition 9.6. Consider a 2D almost-Riemannian structure $(\mathbf{U}, f)$ on a 2D manifold M. The Euler number of $\mathbf{U}$, denoted by $e(\mathbf{U})$ 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 E is defined up to a sign, depending on the orientations of both $\mathbf{U}$ and M. Since reversing the orientation on M also reverses the orientation of TM, the Euler number of TM is defined unambiguously and is equal to $\chi(M)$, the Euler characteristic of M.

Remark 9.7. Assume that $\sigma \in \Gamma(E)$ has only isolated zeros, i.e.,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 associate 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 σ.

Notice that if $i(p, \sigma) \neq 0$, the limit $\lim_{q \to p} \tilde{\sigma}(q)$ does not exist.

Remark 9.8. Notice that $\mathbf{U}$ is trivial if and only if $e(\mathbf{U}) = 0$.

Remark 9.9. Consider a 2D-almost-Riemannian structure $(\mathbf{U}, f)$ on a 2D manifold M. Let σ be a section of $\mathbf{U}$ and $\mathbf{z}_\sigma$ the set of its zeros. As in Remark 9.7, define on $M \setminus \mathbf{z}_\sigma$ the normalization $\tilde{\sigma}$ of σ 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 $(f \circ \tilde{\sigma}, f \circ \tilde{\sigma}^\perp)$.

Exercise 9.10. Consider a 2D-almost-Riemannian structure $(\mathbf{U}, f)$ on a 2D manifold M. Prove that $(\mathbf{U}, f)$ is free when restricted to $M \setminus \{q_0\}$ where q_0 is any point on M.

Definition 9.11. The *singular set $\mathcal{Z}$ of a 2D-almost-Riemannian structure* $(\mathbf{U}, f)$ over a 2D manifold 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 $k(q) := \dim(\mathcal{D}_q)$ is less than 2.

Notice if $q \in \mathcal{Z}$ then $k(q) = 1$. Indeed at q we have $k(q) = 0$ then the structure could not be bracket generated at q.

Since outside the singular set $\mathcal{Z}$, f is fiberwise surjective, we have the following

Proposition 9.12. *A 2D-almost-Riemannian structure is Riemannian structure on $M \setminus \mathcal{Z}$.*

On Riemannian points, the Riemannian metric g is reconstructed with the polarization identity (see Exercice 3.8). We have that if $v = v_1 F_1(q) + v_2 F_2(q) \in T_q M$ and $w = w_1 F_1(q) + w_2 F_2(q) \in T_q M$ then

$$g_q(v, w) = v_1 w_1 + v_2 w_2.$$

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 9.13. 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 9.14. *For every point q_0 of M there exists a neighborhood Ω of this point and a system of coordinates (x_1, x_2) in Ω such that an orthonormal frame for the 2D-almost-Riemannian structure can be written in Ω as:*

$$F_1(q) = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad F_2 = \begin{pmatrix} 0 \\ \mathfrak{f}(x_1, x_2) \end{pmatrix}, \tag{9.2}$$

where $\mathfrak{f} : \Omega \to \mathbb{R}$ is a smooth function. Moreover

(i) the integral curves of F_1 are normal Pontryagin extremals;

(ii) if the step of the structure at q is equal to s, we have $\partial_{x_1}^r \mathfrak{f} = 0$ for $r = 1, 2, \ldots, s-2$ and $\partial_{x_1}^{s-1} \mathfrak{f} \neq 0$;

Remark 9.15. Notice that using the system of coordinates and the orthonormal frame given by Proposition 9.14, we have that $\mathcal{Z} \cap \Omega = \{(x_1, x_2) \in \Omega \mid \mathfrak{f}(x_1, x_2) = 0\}$.

Before proving Proposition 9.14, let us prove the following Lemma

Lemma 9.16. *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., such that $\mathcal{D}(q) + T_q W = T_q M$ for every $q \in W$. Then, for every $q \in W$ there exists an open neighborhood U of q such that for every $\varepsilon > 0$ small enough, the set*

$$\{q' \in U \mid d(q', W) = \varepsilon\} \tag{9.3}$$

is a smooth embedded one-dimensional submanifold of U.

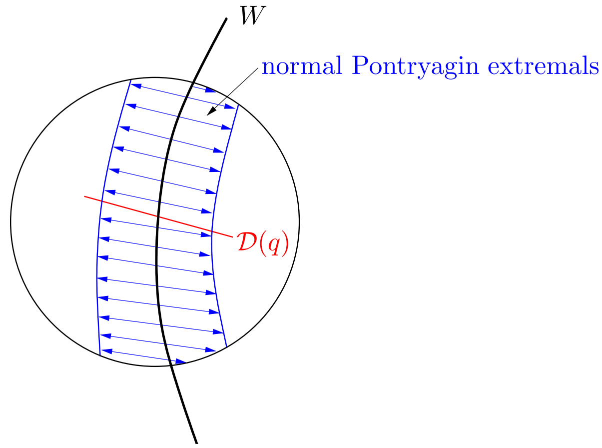


Figure 9.1: Normal Pontryagin extremals starting from the singular set

Proof. Let $H(\lambda)$ be sub-Riemannian Hamiltonian 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 do so 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 not parallel. 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 λ,

$$\langle \lambda_0(\bar{\alpha}), v_2 \rangle = \langle \lambda_0(\bar{\alpha}), \frac{\partial H}{\partial \lambda}(\lambda_0(\bar{\alpha})) \rangle = 2H(\lambda_0(\bar{\alpha})) = 1. \tag{9.5}$$

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 U of q, using the fact that small arcs of normal extremal paths are minimizers, we have that for $\varepsilon > 0$ small enough, the set $A = \{q' \in U \mid d(q', W) = \varepsilon\}$ contains the intersection of U with the images of $E(\varepsilon, \cdot)$ and $E(-\varepsilon, \cdot)$. By possibly restricting U, we are in the situation of Figure 9.1 and the set A coincides with the intersection of U with the images of $E(\varepsilon, \cdot)$ and $E(-\varepsilon, \cdot)$. $\square$

Remark 9.17. 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 non trivial abnormal extremals.

Proof of Proposition 9.14. Following the notation of the proof of Lemma 9.16 let us take (t,α) as a system of coordinates on U 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 $q' \in U$ the vector $X(q')$ belongs to $\mathcal{D}(q')$ and $\|F_1(q')\| = 1$. In the coordinates (t, α) we have $F_1 = (1, 0)$ and by construction its integral curves are normal Pontryagin extremals. Let F_2 be a vector field on U such that (F_1, F_2) is an orthonormal frame for the 2D almost-Riemannian structure in U.

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 B. We have

$$F_3 := [F_1, F_2] = \begin{pmatrix} 0 \\ \partial_{x_1}\mathfrak{f}(x_1, x_2) \end{pmatrix} \tag{9.7}$$

and beside (9.7), the only brackets among F_1, F_2 and F_3 that could be different from zero are of the form

$$\underbrace{[F_3, \dots, [F_3, F_1], F_1]}_{r \text{ times}} = \begin{pmatrix} 0 \\ \partial^r_{x_1}\mathfrak{f}(x_1, x_2) \end{pmatrix}.$$

Hence if the structure has step s at q we have $\partial^r_{x_1}\mathfrak{f} = 0$ for $r = 1, 2, \dots, s-2$ and $\partial^{s-1}_{x_1}\mathfrak{f} \neq 0$. $\square$

The form (9.2) is very useful to express the Riemannian quantities on $M \setminus \mathcal{Z}$. Indeed one has

Lemma 9.18. *Assume that on an open set $\Omega \subset M$ a system of coordinates (x_1, x_2) is fixed and 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*

$$g_{(x_1,x_2)} = \begin{pmatrix} 1 & 0 \\ 0 & \frac{1}{\mathfrak{f}(x_1,x_2)^2} \end{pmatrix}, \tag{9.8}$$

$$dA_{(x_1,x_2)} = \frac{1}{|\mathfrak{f}(x_1, x_2)|} dx_1\, dx_2, \tag{9.9}$$

$$K(x_1, x_2) = \frac{\mathfrak{f}(x_1, x_2)\partial^2_{x_1}\mathfrak{f}(x_1, x_2) - 2\left(\partial_{x_1}\mathfrak{f}(x_1, x_2)\right)^2}{\mathfrak{f}(x_1, x_2)^2}. \tag{9.10}$$

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

$$K(q) = -\alpha_1^2 - \alpha_2^2 + F_1\alpha_2 - F_2\alpha_1$$

where α_1 and α_2 are the two functions defined by $[F_1, F_2] = \alpha_1 F_1 + \alpha_2 F_2$ (see Corollary 4.42). $\square$

Hence in a 2D-almost-Riemannian structure all Riemannian quantities explodes while approaching to $\mathcal{Z}$.

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 9.19. *Consider a system of coordinates (x_1, x_2) defined on an open set Ω and let $dx_1\, dx_2$ be the corresponding Lebesgue measure. Then $\mathcal{Z} \cap \Omega$ has zero $dx_1\, dx_2$-measure.*

Proof. Without loss of generality we can assume that Ω has the following properties:

- it is the product of two non-empty intervals:

$$\Omega = (x_1^A, x_1^B) \times (x_2^A, x_2^B),$$

- on Ω we have an orthonormal frame of the form

$$F_1(q) = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad F_2 = \begin{pmatrix} 0 \\ \mathfrak{f}(x_1, x_2) \end{pmatrix}, \tag{9.11}$$

- on Ω the step of the structure is $s \in \mathbf{N}$.

If some of the properties above are not satisfied, one can prove the theorem on a countable union of sets where the properties above hold.

Let $\mathbf{1}_{\mathcal{Z}} : \Omega \to \{0, 1\}$ be the characteristic function of $\mathcal{Z}$. Using Fubini theorem,

$$\int_{\mathcal{Z}\cap\Omega} dx_1 dx_2 = \int_{\Omega} \mathbf{1}_{\mathcal{Z}}(x_1, x_2)\, dx_1 dx_2 = \int_{x_2^A}^{x_2^B} \left(\int_{x_1^A}^{x_1^B} \mathbf{1}_{\mathcal{Z}}(x_1, x_2) dx_1 \right) dx_2.$$

We now prove that for every fixed $\bar{x}_2 \in (x_2^A, x_2^B)$, we have $\int_{x_1^A}^{x_1^B} \mathbf{1}_{\mathcal{Z}}(x_1, \bar{x}_2) dx_1 = 0$ from which the conclusion of the theorem follows.

Indeed B. of Proposition 9.14 guarantees that there exists $r \leq s - 1$ such that $\partial_{x_1}^r \mathfrak{f}(x_1, \bar{x}_2) \neq 0$ for every $x_1 \in (x_1^A, x_1^B)$. Hence $\mathfrak{f}(\cdot, \bar{x}_2)$ has only isolated zeros and $\int_{x_1^A}^{x_1^B} \mathbf{1}_{\mathcal{Z}}(x_1, \bar{x}_2) dx_1 = 0$. ☐

Exercise 9.20. Show that from the proof of Theorem 9.19 it follows that the singular set is locally the countable union of zero- and one-dimensional manifolds and hence that it is rectifiable.

9.1.2 Genuinely 2D-almost-Riemannian structures have always infinite area

Theorem 9.21. *Let Ω be a bounded open set such that $\Omega \cap \mathcal{Z} \neq \emptyset$. Then*

$$\mathrm{diam}(\Omega) \leq +\infty \qquad \textit{and} \qquad \int_{\Omega\setminus\mathcal{Z}} dA = +\infty$$

where $\mathrm{diam}(\Omega)$ *is the diameter of* Ω *computed with respect to the almost-Riemannian distance and* dA *is the Riemannian area associated with the almost-Riemannian structure on* $\Omega \setminus \mathcal{Z}$*.*

Proof. Take a point $q_0 \in \Omega \setminus \mathcal{Z}$ and a system of coordinates (x_1, x_2) on a neighborhood $\Omega_0 \subset \Omega$ of q_0. Expanding $\mathfrak{f}$ in Taylor series, we have

$$\mathfrak{f}(x_1, x_2) = a_1 x_1 + a_2 x_2 + O(x_1^2 + x_2^2). \tag{9.12}$$

According to (9.9), the (almost-Riemannian) area of Ω_0 is

$$\int_{\Omega_0} \frac{1}{|\mathfrak{f}(x_1, x_2)|} dx_1\, dx_2.$$

But the inverse of a function of the form (9.12) is never integrable around the origin in the plane. ☐

9.1.3 Normal 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 extremals. Normal extremals are geodesics while abnormal extremals could or could not be geodesics. An important fact is the following.

Theorem 9.22. *For a 2D-almost-Riemannian structure, all abnormal extremal are trivial. Moreover a trivial trajectory $\gamma : [a,b] \to M$, $\gamma(t) = q_0$ is the projection of an abnormal extremal if and only if $q_0 \in \mathcal{Z}$.*

Proof. It is immediate to verify that if $\gamma(t) = q_0 \in \mathcal{Z}$ for every $t \in [a,b]$ then γ admits an abnormal lift.

Let $\gamma : [a,b] \to M$, $(a < b)$ be the projection of an abnormal extremal and let us prove that $\gamma([a,b]) = q_0$ for some $q_0 \in \mathcal{Z}$.

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 non trivial interval $[c,d] \subset]a,b[$ such that $\gamma([c,d]) \cap \mathcal{Z} = \emptyset$. Then $\gamma_{[c,d]}$ is a Riemannian geodesic and hence cannot be abnormal. Recall that if an arc of a geodesic is not abnormal, then the geodesic if not abnormal too, hence it follows that γ is not abnormal. This contradicts the hypothesis that γ is the projection of an abnormal extremal.

Let us fix a local system of coordinates such that an orthonormal frame is given in the form (9.2). If this is not possible globally on a neighborhood of $\gamma([a,b])$, one can repeat the proof on different coordinate charts.

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 γ admits an abnormal lift.
- If γ_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

  $$\bar{\mathcal{Z}} = \{(x_1, c) \mid x_1 \in [x_1^A, x_1^B]\} \text{ with } x_1^A < x_1^B.$$

  Hence $\mathfrak{f} = 0$ on $\bar{\mathcal{Z}}$. It follows that $\partial_{x_1}^r \mathfrak{f} = 0$ on $\bar{\mathcal{Z}}$ for every $r = 1, 2, \ldots$. As in the proof of Theorem 9.19 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 $\gamma([a,b]) \subset \mathcal{Z}$ for an abnormal extremal. Hence also this case is not possible.

∎

Hence all non-trivial geodesics are normal and are projection on M of the solution of the Hamiltonian system whose Hamiltonian is (cf. (4.31))

$$H : T^*M \to \mathbb{R}, \qquad H(\lambda) = \max_{u \in U_q} \left(\langle \lambda, f(q,u) \rangle - \frac{1}{2}|u|^2 \right), \quad q = \pi(\lambda). \tag{9.13}$$

Locally, if an orthonormal frame $\{F_1, F_2\}$ is assigned, we have

$$H(\lambda) = \frac{1}{2}\left(\langle\lambda, F_1(q)\rangle^2 + \langle\lambda, F_2(q)\rangle^2\right).$$

For a system of coordinates and a choice of an orthonormal frame as those of Proposition 9.14, we have

$$H(x_1, x_2, p_1, p_2) = \frac{1}{2}\left(p_1^2 + p_2^2\,\mathfrak{f}(x_1, x_2)^2\right). \tag{9.14}$$

As a consequence of the fact that all geodesics are projections of solutions of a smooth Hamiltonian system and that our structure is Riemannian on $M \setminus \mathcal{Z}$, we have

Proposition 9.23. *In 2D almost-Riemannian geometry all geodesics are smooth and they coincide with Riemannian geodesics on $M \setminus \mathcal{Z}$.*

The only particular property of geodesics in almost-Riemannian geometry is that on the singular set their velocity is constrained to belong to the distribution (otherwise their length could not be finite). 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-Riemannain structure on $\mathbb{R}^2$ for which a global orthonormal frame is given by

$$F_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad F_2 = \begin{pmatrix} 0 \\ x_1 \end{pmatrix}$$

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 $f((x_1, x_2), (u_1, u_2)) = ((x_1, x_2), (u_1, u_2 x_1))$.

Here the singular set $\mathcal{Z}$ is the x_2-axis and on $\mathbb{R}^2 \setminus \mathcal{Z}$ the Riemannian metric, the Riemannian area and the Gaussian curvature are given respectively by:

$$g = \begin{pmatrix} 1 & 0 \\ 0 & \frac{1}{x_1^2} \end{pmatrix}, \quad dA = \frac{1}{|x_1|} dx_1\, dx_2, \quad K = -\frac{2}{x_1^2}. \tag{9.15}$$

Notice that the (almost-Riemannian) area of an open set intersecting the x_2-axis is always infinite.

9.2.1 Normal Pontryagin extremals of the Grushin plane

In this section we recall how to compute the normal Pontryagin extremals for the Grushin plane, with the purpose of stressing that they can cross the singular set with no singularities.

In this case the Hamiltonian (9.14) is given by

$$H(x_1, x_2, p_1, p_2) = \frac{1}{2}(p_1^2 + x_1^2 p_2^2) \tag{9.16}$$

and the corresponding Hamiltonian equations are:

$$\begin{aligned} \dot{x}_1 &= p_1, & \dot{p}_1 &= -x_1 p_2^2 \\ \dot{x}_2 &= x_1^2 p_2, & \dot{p}_2 &= 0 \end{aligned} \tag{9.17}$$

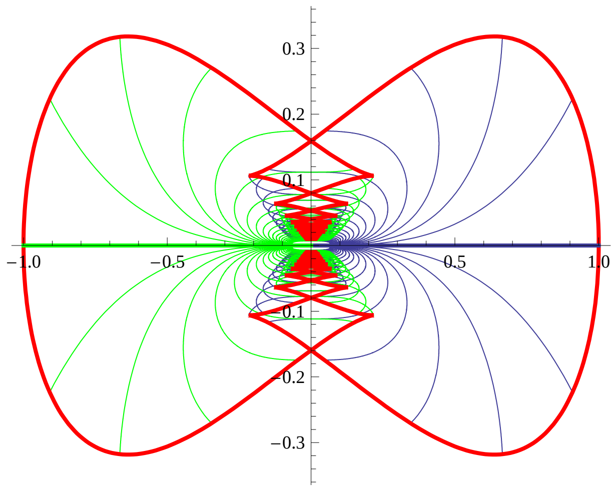


Figure 9.2: Normal Pontryagin extremals and the front for the Grushin plane, starting from the singular set.

Normal Pontryagin extremals parameterized by arclength are projections on the (x_1, x_2) plane of solutions of these equations, lying on the level set $H = 1/2$. We study the normal Pontryagin extremals starting from: **i)** a point on $\mathcal{Z}$, e.g. $(0,0)$; **ii)** an ordinary 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 normal Pontryagin extremals 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:

$$\begin{cases} x_1(t) = \pm t, & x_2(t) = 0, \qquad \text{if } a = 0 \\ x_1(t) = \pm \dfrac{\sin(at)}{a}, & x_2(t) = \dfrac{2at - \sin(2at)}{4a^2}, \qquad \text{if } a \neq 0 \end{cases} \tag{9.18}$$

Some normal Pontryagin extremals are plotted in Figure 9.2 together with the "front", i.e., the end point of all normal Pontryagin extremals at time $t = 1$. Notice that normal Pontryagin extremals start horizontally. The particular form of the front shows the presence of a conjugate locus accumulating to the origin.

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 $p_1 = \cos(\theta), p_2 = \sin(\theta), \quad \theta \in S^1$. The expression of the normal Pontryagin extremals is given by:

$$\begin{cases} x_1(t) = t - 1, \quad x_2(t) = 0, & \text{if } \theta = 0 \\ x_1(t) = -t - 1, \quad x_2(t) = 0, & \text{if } \theta = \pi \\ \left. \begin{array}{l} x_1(t) = -\dfrac{\sin(\theta - t\sin(\theta))}{\sin(\theta)}, \\ x_2(t) = \dfrac{2t - 2\cos(\theta) + \frac{\sin(2\theta - 2t\sin(\theta))}{\sin(\theta)}}{4\sin(\theta)} \end{array} \right\} & \text{if } \theta \notin \{0, \pi\} \end{cases}$$

Some normal Pontryagin extremals are plotted in Figure 9.3 together with the "front" at time $t = 4.8$. Notice that normal Pontryagin extremals pass horizontally through $\mathcal{Z}$, with no singularities. The particular form of the front shows the presence of a conjugate locus. Normal Pontryagin extremals can have conjugate times only after intersecting $\mathcal{Z}$. Before it is impossible since they are Riemannian and the curvature is negative.

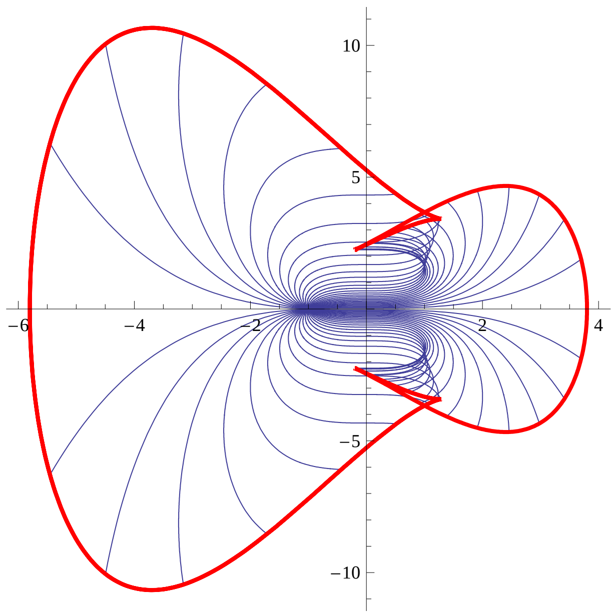


Figure 9.3: Normal Pontryagin extremals and the front for the Grushin plane, starting from a Riemannian point.

9.3 Riemannian, Grushin and Martinet points

In 2D almost-Riemannian structures there are 3 kind of important points, namely Riemannian, Grushin and Martinet points. As we are going to see in Section 9.4, these points are important in the following sense: if a system has only this type of points, then this remains true also after a small perturbation of the system. Moreover arbitrarily close to any system there is a system where only these points are present.

First we study under which conditions $\mathcal{Z}$ has the structure of a 1D manifold. To this purpose we are going to study $\mathcal{Z}$ as the set of zeros of a function.

Definition 9.24. Let $\{F_1, F_2\}$ be a local orthonormal frame on an open set Ω and let ω be a volume form on Ω. On Ω define the function $\Phi = \omega(F_1, F_2)$.

Exercise 9.25. Prove that Φ is invariant by a positive oriented change of orthonormal frame defined on the same open set Ω.

Since a volume form can be globally defined when M is orientable we have that Φ can be globally defined on fully orientable 2D almost-Riemannian structures (cf. Definition 9.3), just defining it as above on positive oriented orthonormal frames.

For structure that are not fully orientable, Φ 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 Φ 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

$$\Phi(x_1, x_2) = h(x_1, x_2) \det \begin{pmatrix} F_1^1 & F_2^1 \\ F_1^2 & F_2^2 \end{pmatrix} \Big|_{(x_1, x_2)}.$$

Remark 9.26. For a system of coordinates and a choice of an orthonormal frame as those of Proposition 9.14, and taking $\omega = dx_1 \wedge dx_2$, we have $\Phi(x_1, x_2) = \mathfrak{f}(x_1, x_2)$.

The function Φ permits to write,

$$\mathcal{Z} = \{q \in M \mid \Phi(q) = 0\}.$$

We are now going to consider the following assumptions

H0$_{q_0}$ If $\Phi(q_0) = 0$ then $d\Phi(q_0) \neq 0$.

H0 The condition **H0**$_{q_0}$ holds for every $q_0 \in M$.

Exercise 9.27. Prove that the conditions above do not depend on the choice of the volume form ω.

By definition of submanifold we have

Proposition 9.28. *Assume that* **H0** *holds. Then $\mathcal{Z}$ is a one dimensional embedded submanifold of M.*

As usual define $\mathcal{D}_1 = \mathcal{D}$, $\mathcal{D}_{i+1} = \mathcal{D}_i + [\mathcal{D}_i, \mathcal{D}_i]$, for $i \geq 1$. We are now ready to define Riemannian, Grushin and Martinet points.

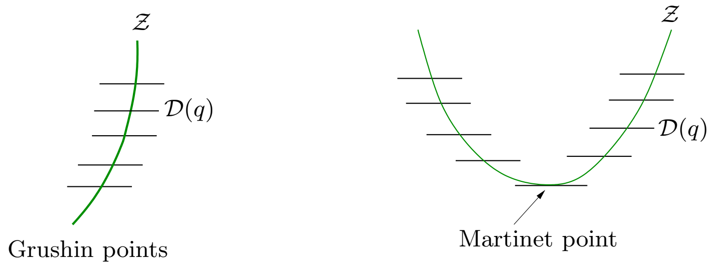


Figure 9.4: Grushin and Martinet points

Definition 9.29. 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 $q_0 \in \mathcal{Z}$), $\mathbf{H0}_{q_0}$ holds then
  - if $\mathcal{D}_2(q_0) = T_qM$ we say that q_0 is a *Grushin* point.
  - if $\mathcal{D}_2(q_0) \neq T_qM$ we say that q_0 is a *Martinet* point.

Remark 9.30. Hence under **H0** every point is either a Riemannian or a Grushin or a Martinet point.

Exercise 9.31. By using the system of coordinate given by Proposition 9.14 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 9.4).

Proposition 9.32. *We have the following:*

(i) $\mathcal{Z}$ is an embedded 1D manifold around Grushin or Martinet points;

(ii) if q_0 is a Grushin point then $\mathcal{D}(q_0)$ is transversal to $T_{q_0}\mathcal{Z}$;

(iii) if q_0 is a Martinet point then $\mathcal{D}(q_0)$ is parallel to $T_{q_0}\mathcal{Z}$;

(iv) Martinet points are isolated.

Proof. We use the system of coordinates and an orthonormal frame as the one given by Proposition 9.14, with $q_0 = (0,0)$,

$$F_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad F_2 = \begin{pmatrix} 0 \\ \mathfrak{f} \end{pmatrix}.$$

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 by definition $d\Phi \neq 0$ at Grushin and Martinet points.

To prove (ii), notice that $\mathcal{D}(q_0) = \text{span}(F_1(q_0)) = (1,0)$ while $T_{q_0}\mathcal{Z} = \text{span}\{(-\partial_{x_2}\mathfrak{f}(q_0), \partial_{x_1}\mathfrak{f}(q_0))\}$ that are not parallel 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 $T_{q_0}\mathcal{Z} = \mathrm{span}\{(-\partial_{x_2}\mathfrak{f}, 0)\}$ since the condition $\mathcal{D}_2(q_0) \neq T_{q_0}M$ implies $\partial_{x_1}\mathfrak{f}(q_0) = 0$.

To prove (iv), simply observe that if Martinet points were accumulating at q_0 then at that point we cold not have $\partial_{x_1}^{s-1}\mathfrak{f} \neq 0$, where s is the step of the structure at q_0. □

Examples

- All points on the x_2-axis for the Grushin plane are Grushin points.
- The origin 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 for 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 (althugh 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*

Proposition 9.33. *Let q_0 be a Riemannian, Grushin or a Martinet point. There exists a neighborhood Ω of q_0 and a system of coordinates (x_1, x_2) in Ω such that an orthonormal frame for the 2D-almost-Riemannian structure can be written in Ω 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
$$F_1(x_1,x_2) = (1,0), \qquad F_2(x_1,x_2) = (0, xe^{\phi(x_1,x_2)})$$

(NF3) if q_0 is a Martinet point, then
$$F_1(x_1,x_2) = (1,0), \qquad F_2(x_1,x_2) = (0, (x_2 - x_1^{s-1}\psi(x))e^{\xi(x_1,x_2)}),$$

where ϕ, ξ and ψ are smooth real-valued functions such that $\phi(0,x_2) = 0$ and $\psi(0) \neq 0$. Moreover $s \geq 2$ is an integer, that is the step of the structure at the Martinet point.

Proof. To be written. □

9.4 Generic 2D-almost-Riemannian structures

Recall hypothesis $\mathbf{H0}_{q_0}$ and $\mathbf{H0}$:

$\mathbf{H0}_{q_0}$ If $\Phi(q_0) = 0$ then $d\Phi(q_0) \neq 0$.

H0 The condition $\mathbf{H0}_{q_0}$ holds for every $q_0 \in M$.

Recall the **H0** is independent from the volume form used to define the function Φ. We have seen (cf. Remark 9.30) 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 9.34. Fix a rank 2 Euclidean bundle $\mathbf{U}$ over a 2D *compact* manifold M. Let $\mathbf{F}$ be the set of all morphism of bundle from $\mathbf{U}$ to TM such that $(\mathbf{U}, f)$, $f \in \mathbf{F}$ is a 2D almost-Riemannian structure. Endow $\mathbf{F}$ with the C^1 norm. We say that a subset of $\mathbf{F}$ is generic if it is open and dense in $\mathbf{F}$.

Theorem 9.35. *Under the same hypothesis of Definition 9.34, let $\bar{\mathbf{F}} \subset \mathbf{F}$ the subset of morphisms satisfying* **H0**. *Then $\bar{\mathbf{F}}$ is generic.*

Remark 9.36. In Theorem 9.35 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 gets that $\bar{\mathbf{F}}$ is a countable union of open and dense subsets of $\mathbf{F}$ and one should use a suitable topology (the Whitney one). 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.

9.4.1 Proof of the genericity result

Cover M with a finite number of *compact* coordinate neighborhood $\mathcal{U}_i$, $i = 1 \ldots N$, in such a way that an orthonormal frame for the 2-ARS in $\mathcal{U}_i$ is given by

$$F_i(x_1^i, x_2^i) = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad G_i(x_1^i, x_2^i) = \begin{pmatrix} 0 \\ \mathfrak{f}_i(x_1, x_2) \end{pmatrix}. \tag{9.19}$$

Let us consider the following hypothesis

$\mathbf{H}_i$ The condition $\mathbf{H0}_{q_0}$ holds for every $q_0 \in \mathcal{U}_i$.

Proposition 9.37. *Let $\mathbf{F}_i$ be the subset of $\mathbf{F}$ satisfying $\mathbf{H0}_i$. Then $\mathbf{F}_i$ is generic.*

Once Proposition 9.37 is proved, the conclusion of Theorem 9.35 follows immediately. Indeed $\mathbf{F}_i$ is open and dense in $\mathbf{F}$ and the open and dense set $\bar{\mathbf{F}} := \cap_{i=1}^{N} \mathbf{F}_i$ is made by systems satisfying **H0** in all M.

Proof of Proposition 9.37. Since the map that to (F_i, G_i) associates Φ is continuous in the C^1 topology, a small perturbation of F_i and G_i will induce a small perturbation of Φ. Fixed q_0, condition $\mathbf{H0}_{q_0}$ is clearly open in the set of maps from $\mathcal{U}_i$ to $\mathbb{R}$ for the C^1 topology. As a consequence of the compactness of $\mathcal{U}_i$, condition $\mathbf{H0}_i$ is open as well.

We are now going to prove that $\mathbf{H0}_i$ 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 $\mathbf{H0}_i$ is satisfied.

Lemma 9.38. *For every $\varepsilon \in \mathbb{R}$ with $|\varepsilon|$ small enough there exists a perturbation $(F_i^\varepsilon, G_i^\varepsilon)$ of (F_i, G_i) such that $\|F_i^\varepsilon - F_i\|_{C^1} \leq C\varepsilon$, $\|G_i^\varepsilon - G_i\|_{C^1} \leq C\varepsilon$ (for some $C > 0$ independent from ε) and on $\mathcal{U}_i$ we have $\Phi_\varepsilon := \omega(F_i^\varepsilon, G_i^\varepsilon) = \Phi + \varepsilon$;*

Once Lemma 9.38 is proved, the density of $\mathbf{F}_i$ follows easily. Indeed let now apply the Sard Lemma to the C^∞ function Φ 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 ε, condition $\mathbf{H0}_i$ is realized for $(F_i^\varepsilon, G_i^\varepsilon)$. $\square$

Proof of Lemma 9.38. If in $\mathcal{U}_i$ we write in coordinates

$$\omega = h_i(x_1^i, x_2^i) dx_1^i \wedge dx_2^i,$$

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^ε of G_i of the form

$$G_i^\varepsilon(x_1^i, x_2^i) = \begin{pmatrix} 0 \\ \mathfrak{f}_i(x_1^i, x_2^i) + \frac{\varepsilon}{h_i(x_1^i, x_2^i)} \end{pmatrix}. \tag{9.20}$$

and let us define $F_i^\varepsilon = F_i$. It follows that in $\mathcal{U}_i$,

$$\Phi_\varepsilon = \omega(F_i^\varepsilon, G_i^\varepsilon) = h_i(x_1^i, x_2^i)\left(\mathfrak{f}_i(x_1^i, x_2^i) + \frac{\varepsilon}{h_i(x_1^i, x_2^i)}\right) = h_i(x_1^i, x_2^i)\mathfrak{f}_i(x_1^i, x_2^i) + \varepsilon = \Phi + \varepsilon.$$

Notice that by construction G_i^ε is close to G_i in the C^1 norm. $\square$.

9.5 A Gauss-Bonnet theorem

For an compact orientable 2D-Riemannian manifold, the 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 admit an interesting generalization in the context of 2D almost-Riemannian structures that are fully orientable. This generalization is not trivial 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, namely we will assume

HG : The base manifold M is compact. The 2D almost-Riemannian structure is fully orientable, **H0** holds and every point of $\mathcal{Z}$ is a Grushin point.

The hypotheses that the structure is fully orientable is crucial and it is the almost-Riemannian version of the classical orientability hypothesis that one need in Riemannian geometry. The hypothesis **H0** is the basic hypothesis to have a reasonable description of the asymptotics of K in a neighborhood of $\mathcal{Z}$. The hypotesis that every point is a Grushin point is a technical hypothesis. A version of a Gauss Bonnet Theorem in presence of Martinet points can also be written, but is more technical and outside the purpose of this book.

With an argument similar to the one of the beginning of Section 9.4.1, one get

Theorem 9.39. *Hypothesis* **HG** *is open in the set of smooth map $f : \mathbf{U} \to TM$ endowed with C^1 topology:*

Clearly hypothesis **HG** is not dense since Martinet points do not disappear for small C^1 perturbations of the system.

It is important to notice that **HG** is not empty. Indeed we have

Lemma 9.40. *Every oriented compact surface can be endowed with an oriented almost-Riemannian structure satisfying the requirement that there are no Martinet points.*

We are going to prove Lemma 9.40 in Section 9.5.2.

Definition 9.41. Consider a 2D almost-Riemannian structure $(\mathbf{U}, f)$ over a 2D manifold M and assume that **HG** holds.

Let ν 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 positive oriented local orthonormal frame for $(\cdot \mid \cdot)_q$. Let Ξ be an orientation on M. We define:

- The *signed area form* dA^s on M as the two-form on $M \setminus \mathcal{Z}$ given by the pushforward of ν 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 Ξ_q are the same $\}$.[1]
- $M^- = \{q \in M \setminus \mathcal{Z},$ s.t. the orientation given by dA^s_q and Ξ_q are opposite $\}$.

Notice that given a measurable function $h : \Omega \subset M^\pm \setminus \mathcal{Z} \to \mathbb{R}$, we have

$$\int_\Omega h \; dA_s = \pm \int_\Omega h \; dA \quad \text{(if it exists).} \tag{9.21}$$

Definition 9.42. Under the same hypotheses of Definition 9.41, 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
$$\lim_{\varepsilon \to 0} \int_{M_\varepsilon} h \; dA_s \tag{9.22}$$
exists and is finite. In this case we denote such a limit by $\int h \, dA_s$.

Remark 9.43. Notice that (9.22) is equivalent to

$$\lim_{\varepsilon \to 0} \left(\int_{M_\varepsilon^+} h \; dA - \int_{M_\varepsilon^-} h \; dA \right)$$

[1] i.e., $dA^s_q(F_1, F_2) = \alpha \Xi(F_1, F_2)$ with $\alpha > 0$

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

$$Y_1 = \begin{pmatrix} 0 \\ -y_3 \\ y_2 \end{pmatrix} \text{ (rotation along the } y_1\text{-axis)} \tag{9.23}$$

$$Y_2 = \begin{pmatrix} -y_3 \\ 0 \\ y_1 \end{pmatrix} \text{ (rotation along the } y_2\text{-axis)} \tag{9.24}$$

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

$$\begin{aligned} X_1 &= \cos(\phi - \pi/2)Y_1 + \sin(\phi - \pi/2)Y_2 \\ X_2 &= -\sin(\varphi - \pi/2)Y_1 + \cos(\phi - \pi/2)Y_2 \end{aligned}$$

we get that an orthonormal frame is given by

$$X_1 = \begin{pmatrix} 0 \\ \tan(x) \end{pmatrix}, \quad X_2 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

Notice that the singularity at $x = \pi/2$ is due to the spherical coordinates. Instead $\mathcal{Z} = \{x = 0\}$. 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 9.5.

The main result of this section is the following.

Theorem 9.44. *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*

$$\int K\, dA^s = e(\mathbf{U})$$

where $e(\mathbf{U})$ denotes the Euler number of E. Moreover we have

$$e(\mathbf{U}) = \chi(M^+) - \chi(M^-)$$

where $\chi(M^{\pm})$ denotes the Euler characteristic of $M^{\pm}$.

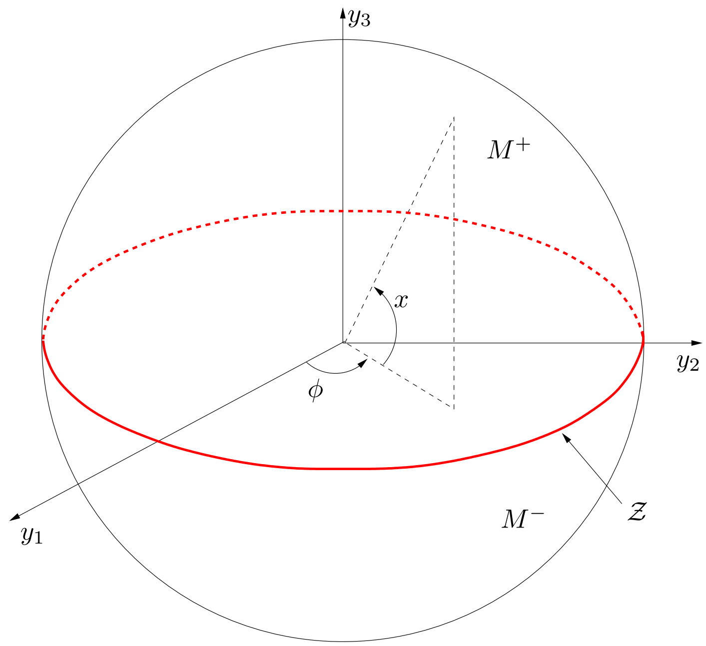


Figure 9.5: The Grushin sphere

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 9.44 contains the classical Gauss-Bonnet theorem.

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 constraints 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 what happens 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 we are constrained to be parallelizable (i.e.,we are constrained to be on the torus). In the AR context, M could be any compact orientable manifolds, but the metric is constrained to be singular somehwere. In any case, the integral of the curvature will be zero.

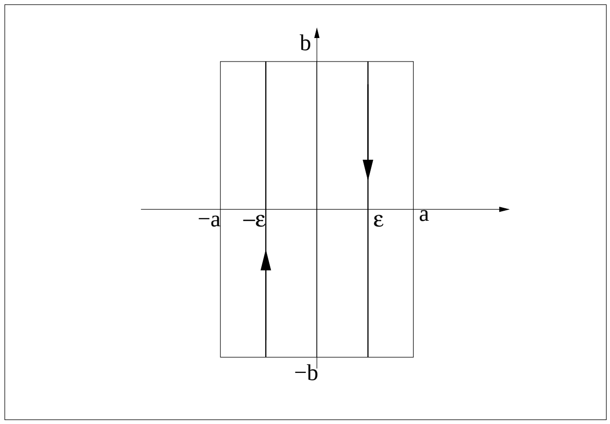
9.5.1 Proof of Theorem 9.44*

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 9.16 one has:

Lemma 9.45. *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*



∂M_ε *is smooth and the set* $M \setminus M_\varepsilon$ *is diffeomorphic to* $\mathcal{Z} \times [0,1]$.

Under **HG** the almost-Riemannian structure can be described, around each point of $\mathcal{Z}$, by a normal form of type (NF2).

Take ε_0 as in the statement of Lemma 9.45. 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 Gauss-Bonnet formula asserts that for every compact oriented Riemannian manifold (N, g) with smooth boundary ∂N, we have

$$\int_N K dA + \int_{\partial N} k_g ds = 2\pi\chi(N),$$

where K is the curvature of (N, g), dA is the Riemannian density, k_g is the geodesic curvature of ∂N (whose orientation is induced by the one of N), and ds is the length element.

Applying the Gauss-Bonnet formula to the Riemannian manifolds (M_ε^+, g) and (M_ε^-, g) (whose boundary smoothness is guaranteed by Lemma 9.45), 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 9.45, $\chi(M_\varepsilon^\pm) = \chi(M^\pm)$. 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 $\varepsilon < a$. 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

$$\int_{\partial M_\varepsilon \cap U_q} k_g \, ds = O(\varepsilon). \tag{9.27}$$

Then (9.26) follows from the compactness of $\mathcal{Z}$. (Indeed, $\{-\varepsilon\} \times [-b, b]$ and $\{\varepsilon\} \times [-b, b]$, the horizontal edges of ∂U_q, are normal Pontryagin extremals 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_ε^+ induces on $\partial M_\varepsilon^+ = \{\varepsilon\} \times [-b, b]$ a downwards orientation (see Figure 9.5.1). 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 arclength of ∂M_ε^+, making a constant angle with F_1. Let (θ_1, θ_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 [**?**, Corollary 3, p. 389, Vol. III], the geodesics curvature of ∂M_ε^+ at $c(s)$ is equal to $\lambda(\dot{c}(s))$, where $\lambda \in \Lambda^1(U_q)$ is the unique one-form satisfying

$$d\theta_1 = \lambda \wedge \theta_2, \quad d\theta_2 = -\lambda \wedge \theta_1.$$

A trivial computation shows that

$$\lambda = \partial_{x_1}(x_1^{-1} e^{-\phi(x_1, x_2)}) dx_2.$$

Thus,

$$k_g(c(s)) = -\partial_{x_1}(x_1^{-1} e^{-\phi(c(s))}) \, (dx_2(F_2))(c(s)) = \frac{1}{\varepsilon} + \partial_{x_1}\phi(\varepsilon, x_2(s)).$$

Denote by L_1 and L_2 the lengths of, respectively, $\{\varepsilon\} \times [0, b]$ and $\{\varepsilon\} \times [-b, 0]$. Then,

$$\begin{aligned} \int_{\partial M_\varepsilon^+ \cap U_q} k_g ds &= \int_{-L_1}^{L_2} k_g(c(s)) ds \\ &= \int_{-L_1}^{L_2} \left(\frac{1}{\varepsilon} + \partial_{x_1}\phi(\varepsilon, (s)) \right) ds \\ &= \int_{-b}^{b} \left(\frac{1}{\varepsilon} + \partial_{x_1}\phi(\varepsilon, x_2) \right) \frac{1}{\varepsilon e^{\phi(\varepsilon, x_2)}} dx_2, \end{aligned}$$

where the last equality is obtained taking $x_2 = x_2(-s)$ as the new variable of integration.

We reason similarly on $\partial M_\varepsilon^- \cap U_q$, on which M_ε^- induces the upwards 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

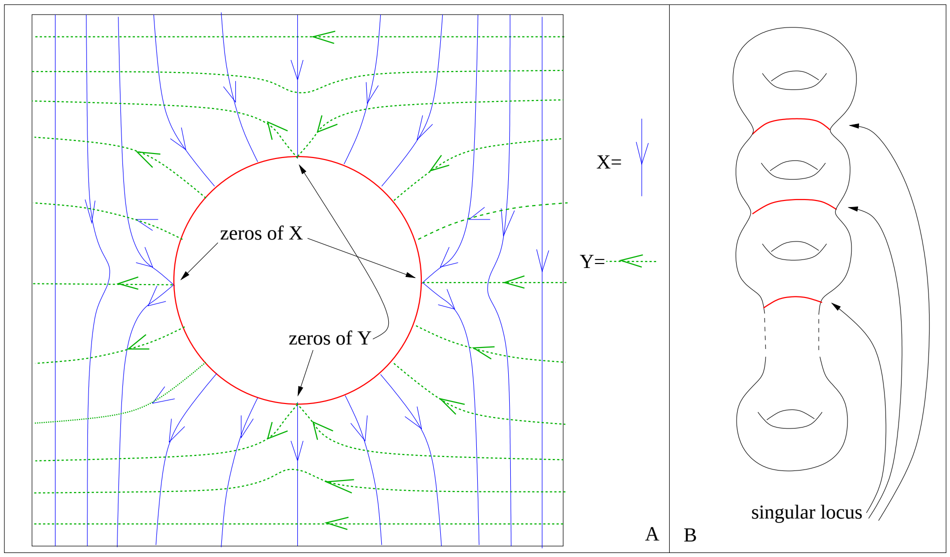
$$F(\varepsilon, x_2) = (1 + \varepsilon \partial_{x_1}\phi(\varepsilon, x_2)) e^{-\phi(\varepsilon, x_2)}. \tag{9.28}$$

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 ε 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

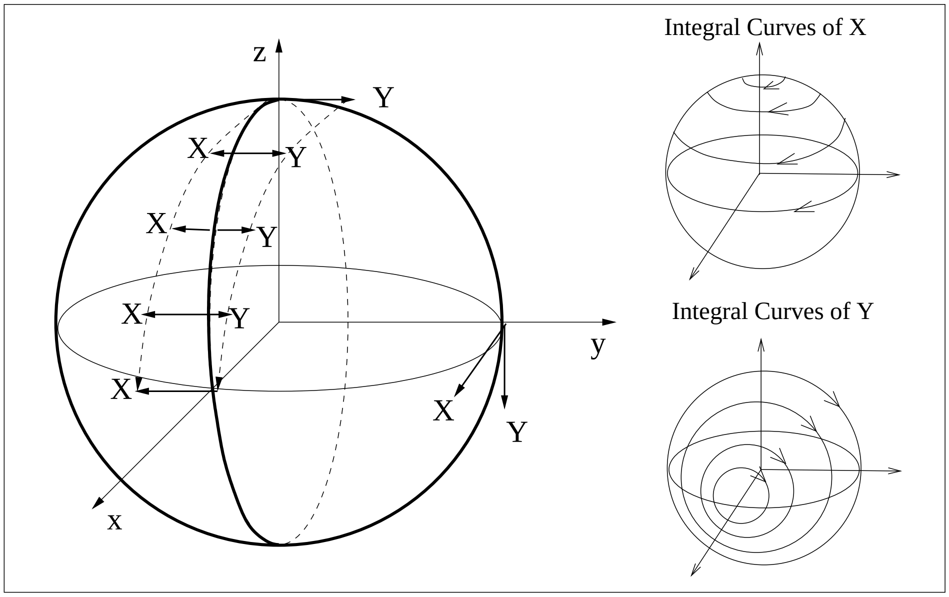
The idea of the proof is to find a section σ of SE with isolated singularities $p_1,\dots,p_m$ such that $\sum_{j=1}^m i(p_j,\sigma)=\chi(M^+)-\chi(M^-)+\tau(\mathcal{S})$. In the sequel, we consider $\mathcal{Z}$ to be oriented with the orientation induced by M^+. To be finished.

9.5.2 Construction of trivializable 2-ARSs with no tangency points

In this section we prove Lemma 9.40, by showing how to construct a trivializable 2-ARS with no tangency points on every compact orientable two-dimensional manifold.

Without loss of generality we can assume M connected. 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 9.5.2A, 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 9.5.2B.

We are left to check the existence of a trivializable 2-ARS with no tangency points on a sphere. A simple example can be found in the literature and arises from a model of control of quantum systems (see [29, 30]). Let M be a sphere in $\mathbb{R}^3$ centered at the origin and take $F_1(x,y,z)=(y,-x,0)$,



$F_2(x, y, z) = (0, z, -y)$ as orthonormal frame. Then F_1 (respectively, F_2) is an infinitesimal rotation around the third (respectively, first) axis. The singular locus is therefore given by the intersection of the sphere with the plane $\{y = 0\}$ and none of its points exhibit tangency (see Figure 9.5.2). Notice that hypothesis **HG** is satisfied.

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 homogeneous space of Carnot group, defined as follows.

Definition 10.1 (Carnot Groups)**.** A Carnot group G is a connected and simply connected Lie group whose Lie algebra $\mathfrak{g}$ admits a decomposition

$$\mathfrak{g} = \mathfrak{g}_1 \oplus \mathfrak{g}_2 \oplus \ldots \oplus \mathfrak{g}_r \tag{10.1}$$

satisfying the following properties

$$[\mathfrak{g}_1, \mathfrak{g}_i] = \mathfrak{g}_{i+1}, \qquad [\mathfrak{g}_1, \mathfrak{g}_r] = 0, \qquad i = 1, \ldots, r-1. \tag{10.2}$$

The smallest integer r such that (10.1)-(10.2) are satisfied is called *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), that is bracket generating thanks to (10.2).

Notice that Carnot groups of step 2 as defined in Section 7.5 are included in Definition 10.1.

Remark 10.2. 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 (10.1) and thanks to the property (10.2) they posses a family $\{\delta_\alpha\}_{\alpha\in\mathbb{R}}$ of authomorphisms on $\mathfrak{g}$ (called *dilations*) defined by

$$\delta_\alpha(v) = \sum_{i=1}^{r} \alpha^i v_i, \qquad \text{if} \quad v = \sum_{i=1}^{r} v_i,\ v_i \in \mathfrak{g}_i.$$

Carnot groups play a crucial role in sub-Riemannian geometry : these are left-invariant sub-Riemannian structure arising as metric tangent space of equiregular sub-Riemannian manifolds. In this sense they 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 based on the notion of admissible variation, providing both a geometric and an algebraic interpretation of this construction. We prove the existence of privileged coordinates, i.e., special sets of coordinates where the nonholonomic tangent space writes conveniently to perform computations.

Moreover this chapter contains also 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 Jet spaces

In this chapter, given a point $q \in M$, the symbol Ω_q denotes the set of smooth curves γ 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 γ 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

$$\gamma(t) = q + \dot{\gamma}(0)t + O(t^2). \tag{10.3}$$

The tangent vector $v \in T_qM$ to γ at $t = 0$ is by definition the equivalence class of curves in Ω_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

$$\gamma(t) = q + \dot{\gamma}(0)t + \ddot{\gamma}(0)\frac{t^2}{2} + \ldots + \gamma^{(k)}(0)\frac{t^k}{k!} + O(t^{k+1}), \tag{10.4}$$

and define analogously an equivalence class on higher order Taylor polynomial.

Exercise 10.3. Let $\gamma, \gamma' \in \Omega_q$. We say that γ is equivalent up to order k at q to γ', 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 10.4. 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 Ω_q with respect to $\sim_{q,k}$. We denote with $J_q^k\gamma$ the equivalence class of a curve γ and with

$$J_q^kM := \{J_q^k\gamma \mid \gamma \in \Omega_q\}.$$

Exercise 10.5. Prove that J_q^kM has the structure of smooth manifold and $\dim J_q^kM = kn$. *Hint*: use the coordinates representation (10.4) 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 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 10.6. *Let $k \geq 2$. Then $J_q^k M$ is an affine bundle over $J_q^{k-1} M$ with projection Π_{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

$$J_q^k \gamma + v := J_q^k(\gamma^v), \qquad \gamma^v(t) = e^{t^k V}(\gamma(t)) \tag{10.5}$$

It is easy to see that, due to the presence of the factor t^k, the $(k-1)^{th}$ Taylor polynomial of γ and γ^v coincide. Indeed

$$J_q^k(e^{t^k V}(\gamma(t))) = J_q^k \gamma + t^k V(q)$$

Hence the sum (10.5) gives to $(\Pi_{k-1}^k)^{-1}(j)$ the structure of affine space over $T_q M$. Notice that this definition does not depend on the representative curve γ defining j. □

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 10.7. 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 α 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.

Definition 10.8. 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) and its infinitesimal generator is the nonautonomous vector field

$$X_t := \frac{d}{d\varepsilon}\bigg|_{\varepsilon=0} P_{t+\varepsilon} \circ P_t^{-1}. \tag{10.6}$$

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 10.8) on a smooth curve γ is defined as

$$(P\gamma)(t) := P_t(\gamma(t)). \tag{10.7}$$

Proposition 10.9. *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*

$$Pj := J_q^k(P\gamma), \qquad \textit{if} \quad j = J_q^k\gamma. \tag{10.8}$$

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 γ 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. □

10.1.1 Jets of vector fields

As explained in Proposition 10.9, a flow on M induces a diffeomeorphism in Ω_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 pass 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 10.10. 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

$$(J_q^k V)(J_q^k\gamma) := \frac{\partial}{\partial s}\bigg|_{s=0} P_V^s(J_q^k\gamma) = \frac{\partial}{\partial s}\bigg|_{s=0} J_q^k(e^{tsV}(\gamma(t))). \tag{10.9}$$

Exercise 10.11. Prove the following formula for every $V \in \mathrm{Vec}(M)$

$$(J_q^k V)(J_q^k\gamma) = \sum_{i=1}^{k} \frac{t^i}{i!}\frac{d^i}{dt^i}\bigg|_{t=0}(tV(\gamma(t))),$$

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 δ_α 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 α.

Proposition 10.12. *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

$$\begin{aligned}
((\delta_\alpha)_* J_q^k V))(J_q^k \gamma) &= \frac{\partial}{\partial s}\bigg|_{s=0} J_q^k(\delta_\alpha\, e^{tsV}\, \delta_{1/\alpha}(\gamma(t))) \\
&= \frac{\partial}{\partial s}\bigg|_{s=0} J_q^k(\delta_\alpha\, e^{tsV}(\gamma(t/\alpha))) \\
&= \frac{\partial}{\partial s}\bigg|_{s=0} J_q^k(e^{\alpha tsV}(\gamma(t))) \\
&= J_q^k(\alpha V) = \alpha J_q^k V
\end{aligned}$$

$\square$

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.

10.2 Admissible variations

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 $\{f_1, \ldots, f_m\}$ 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.

Definition 10.14. A curve $\gamma : [0, T] \to M$ in Ω_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 γ 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 10.15. 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 10.16. 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 γ 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 $\mathcal{D}^i_q \setminus \mathcal{D}^{i-1}_q$ (cf. Lemmas 10.34-10.35 for a precise statement).

Proposition 10.17. *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 3.17) 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

$$\frac{\partial}{\partial s} q(t,s) = \sum_{i=1}^m u_i(t,s) f_i(q(t,s)), \qquad s \in [0,\tau], \tag{10.10}$$

Assume that $f_1', \dots, f_{m'}'$ is another set of local generators of the modulus. Then there exist functions $a_{ij} \in C^\infty(M)$ for $i = 1, \dots, m$ and $j = 1, \dots, m'$, such that

$$f_i(q) = \sum_{j=1}^{m} a_{ij}(q) f_j'(q), \qquad \forall\, q \in M, \qquad \forall\, i = 1, \dots, m. \tag{10.11}$$

Next we prove that there exist a family $\widetilde{u}(t,s)$ of controls such that γ is an admissible variation for the frame $f_1', \dots, f_{m'}'$. From (10.11) we get

$$\sum_{i=1}^{m} u_i(t,s) f_i(q) = \sum_{i=1}^{m} \sum_{j=1}^{m'} u_i(t,s) a_{ij}(q) f_j'(q). \tag{10.12}$$

Then we could define, through the solution $q(t,s)$ of (10.10), the new family of controls

$$u_j'(t,s) := \sum_{i=1}^{m} u_i(t,s) a_{ij}(q(t,s)), \qquad j = 1, \dots, m',$$

and we see from identities above that

$$\frac{\partial}{\partial s} q(t,s) = \sum_{j=1}^{m'} u_j'(t,s) f_j'(q(t,s)), \qquad s \in [0, \tau]. \tag{10.13}$$

Since the role of $f_1, \dots, f_m$ and $f_1', \dots, f_{m'}'$ 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 10.18. 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., $\mathcal{D}_q^k = T_q M$.

Next we want to introduce the nonholonomic tangent space in a coordinate-free way. In the next section we will see how it can be described in some special set of coordinates.

Definition 10.19. 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

$$v(t,s) := \begin{cases} u_1(t,s), & 0 \le s \le \tau_1, \\ u_2(t, s - \tau_1), & \tau_1 \le s \le \tau_1 + \tau_2. \end{cases}$$

Remark 10.20. 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 10.21. A smooth admissible variation $u(t,s)$ for $\mathcal{D}$ is said to be a *slow variation* if

$$u(0,s) = \frac{\partial u}{\partial t}(0,s) = 0, \qquad \forall\, s \in [0,\tau]. \tag{10.14}$$

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

$$\mathcal{P}_0^f := \left\{(P_t)^{-1} \circ Q_t \circ P_t \mid P \in \mathcal{P}^f,\, Q \text{ purely slow}\right\}. \tag{10.15}$$

Remark 10.22. Notice that, 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 10.15). 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 defines an equivalence relation on the space of jets.

Exercise 10.23. Let $j = J_q^k \gamma$ and $j' = J_q^k \gamma'$ for some $\gamma, \gamma' \in \Omega_q$. Prove that

$$J_q^k \gamma \sim J_q^k \gamma', \qquad \text{if} \qquad \gamma'(t) = P_t(\gamma(t)) \tag{10.16}$$

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 10.24. 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 (10.16).

Finally, every horizontal vector field induces a vector field on the noholonomic tangent space at every point.

Proposition 10.25. *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 nonhonolomic 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

$$(J_q^k X)(J_q^k\gamma) := \frac{\partial}{\partial s}\bigg|_{s=0} P_X^s(J_q^k\gamma) = \frac{\partial}{\partial s}\bigg|_{s=0} J_q^k(e^{tsX}(\gamma(t))). \tag{10.17}$$

It is easily seen that if $\gamma(t)$ is an admissible variation, then for every s the curve $e^{tsV}(\gamma(t))$ 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 $T_q^f M$. □

10.3 Nilpotent approximation and privileged coordinates

In this section we want to introduce some special set of coordinates in which we have a good description of the nonholonomic tangent space $T_q^f M$.

Consider some 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 $i = 1, \ldots, k$; $j = 1, \ldots, n_i$. 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 10.26. *Let D_1, D_2 be two homogeneous differential operators. Then $D_1 \circ D_2$ is homogeneous and*

$$\nu(D_1 \circ D_2) = \nu(D_1) + \nu(D_2). \tag{10.18}$$

Proof. By linearity, it is sufficent to check formula (10.18) 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 (10.18) is easily checked in this case. □

A special case is when we consider first order differential operators, namely vector fields.

Corollary 10.27. *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

$$V \in \mathrm{Vec}(\mathbb{R}^n) \quad \Rightarrow \quad V^{(i)} = 0, \quad \forall\, i < -m.$$

Indeed every monomial of a N^{th}-order differential operator has weight not smaller than $-mN$. In other words we have

(i) $\mathrm{Vec}(\mathbb{R}^n) \subset \mathcal{F}^{(-m)}$,

(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 10.28. (i). A system of coordinates near the point q is said *linearly adapted* to the flag $\mathcal{D}_q^1 \subset \mathcal{D}_q^2 \subset \ldots \subset \mathcal{D}_q^k$ if

$$\mathcal{D}_q^i = \mathbb{R}^{n_1} \oplus \ldots \oplus \mathbb{R}^{n_i}, \qquad \forall\, i = 1, \ldots, k. \tag{10.19}$$

(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 10.29 (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 $X(0) \in \mathcal{D}_0^1 = \mathbb{R}^{n_1}$.

(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

$$\mathcal{D}_0^1 = \mathrm{span}\{\partial_{x_1}\}, \qquad \mathcal{D}_0^2 = \mathrm{span}\{\partial_{x_1}, \partial_{x_2}\}, \qquad \mathcal{D}_0^3 = \mathrm{span}\{\partial_{x_1}, \partial_{x_2}, \partial_{x_3}\}.$$

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 10.30. *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 10.3.2.

10.3.1 Properties of privileged coordinates

We showed in Proposition 10.25 that given a horizontal vector field X it induces a well defined vector field $\widehat{X}$ on the nonhonolomic 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.

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 10.31. *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 10.31, Claim (i), part 1. We start by proving the following inclusion

$$J_q^f M \subset \left\{ \sum_{i=1}^k t^i \xi_i \mid \xi_i \in \mathcal{D}_q^i \right\}. \tag{10.20}$$

For any smooth variation $\gamma(t) = q \circ \overrightarrow{\exp} \int_0^\tau f_{u(t,s)} ds$, we can write the Volterra expansion

$$\gamma(t) = q + \sum_{i=1}^k \int_{0 \le s_i \le \ldots \le s_1 \le \tau} \cdots \int q \circ f_{u(t,s_1)} \circ \ldots \circ f_{u(t,s_i)} \, ds_1 \ldots ds_i + O(t^{k+1}). \tag{10.21}$$

Let us write (cf. Remark 10.15) the controls $u(t, s_i) = t\bar{u}(t, s_i)$ for some suitable families $\bar{u}(t, s_i)$. Then (10.21) becomes, using the fact that f is linear in u, as follows

$$\gamma(t) = q + \sum_{i=1}^k t^i \int_{0 \le s_i \le \ldots \le s_1 \le \tau} \cdots \int q \circ f_{\bar{u}(t,s_1)} \circ \ldots \circ f_{\bar{u}(t,s_i)} \, ds_1 \ldots ds_i + O(t^{k+1}). \tag{10.22}$$

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

$$f_{\bar{u}(t,s_1)} \circ \ldots \circ f_{\bar{u}(t,s_i)} \in \mathcal{F}^{(-j)} \tag{10.23}$$

Let us apply the differential operator (10.23) to a coordinate function x_α^β, with $\alpha = 1, \ldots, k$ and $\beta = 1, \ldots, n_\alpha$. Since $\nu(x_\alpha^\beta) = \alpha$ we have

$$f_{\bar{u}(t,s_1)} \circ \ldots \circ f_{\bar{u}(t,s_i)} x_\alpha^\beta \in \mathcal{F}^{(-i+\alpha)} \tag{10.24}$$

Therefore, for every $\alpha > i$, this function has positive weight and vanishes when evaluated at $x = 0$.

In privileged coordinates satisfying (10.19), this says that, for every $i = 1, \ldots, k$, the sum in (10.21) 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 $\xi_i \in \mathcal{D}_q^i$ for $i = 1, \ldots, k$, 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 10.32. *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 gives contribution. Using that

$$P_t^{-1} = \mathrm{Id} - t^n V + O(t^{n+1}), \qquad Q_t^{-1} = \mathrm{Id} - t^m W + O(t^{m+1}).$$

one gets

$$\begin{aligned}(\mathrm{Id} + t^n V + O(t^{n+1}))(\mathrm{Id} + s^m W + O(s^{m+1}))(\mathrm{Id} - t^n V + O(t^{n+1}))(\mathrm{Id} - s^m W + O(s^{m+1})) = \\ = \mathrm{Id} + t^n s^m (VW - WV) + O(t^{n+m+1}) \\ = \mathrm{Id} + t^n s^m [V, W] + O(t^{n+m+1})\end{aligned}$$

and the lemma is proved. □

Exercise 10.33. 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 to P_t satisfies $V_t = nt^{n-1}V + O(t^n)$.

Lemma 10.34. *For all $i_1, \ldots, i_h \in \{1, \ldots, k\}$ and $l \geq h$, there exists an admissible variation $u(t,s)$, depending only on the Lie bracket structure, such that*

$$q \circ \overrightarrow{\exp} \int_0^\tau f_{u(t,s)} ds = q + t^l [f_{i_1}, \ldots, [f_{i_{h-1}}, f_{i_h}]](q) + O(t^{l+1}). \tag{10.25}$$

Proof. The lemma is proved by induction on h.

(i) For all $i = 1, \ldots, k$ 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 $u = (u_1, \ldots, u_k)$ such that $u_i = t^l$ and $u_j = 0$ for all $j \neq i$.

(ii) For all $i, j \in \{1, \ldots, k\}$ 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 10.32 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.

Iterating this argument the lemma is proved.

□

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 10.35. *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*

$$q \circ \overrightarrow{\exp} \int_0^\tau f_{u(t,s)} ds = q + t^l \pi(f_1, \ldots, f_m)(q) + O(t^{l+1}). \tag{10.26}$$

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 (10.26) 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 $j = 1$. □

Exercise 10.36. 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 αV 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 10.31

Proof of Theorem 10.31, Claim (i), part 2. We have to prove the remaining inclusion

$$\left\{ \sum_{i=1}^k t^i \xi_i \mid \xi_i \in \mathcal{D}_q^i \right\} \subset J_q^f M. \tag{10.27}$$

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 10.35, 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 10.35, 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. □

[1] we define the product of two curves $\gamma(t) = q \circ P_t$ and $\gamma'(t) = q \circ P'_t$ as follows: $(\gamma' * \gamma)(t) := q \circ P_t \circ P'_t$.

Proof of Theorem 10.31, 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

$$Q_t = Q_t^1 \circ \cdots \circ Q_t^h,$$
$$Q_t^i = P_t^i \circ \overrightarrow{\exp} \int_0^\tau f_{tv^i(t,s)} ds \circ (P_t^i)^{-1},$$

for some $P^i \in \mathcal{P}^f$ and some admissible variations $v_i(t,s)$, for $i = 1, \ldots, h$. It is sufficient to prove it for the case $h = 1$. By formula (6.27) 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 (6.24) 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 γ' we will find the same terms of γ plus something that belong to $\mathcal{D}^{i-1}$.

($\Leftarrow$). 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 10.32.

Lemma 10.37. *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 $P_t Q_t P_t^{-1} Q_t^{-1} \in \mathcal{P}_0^f$.*

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 $P_t Q_t P_t^{-1} \in \mathcal{P}_0^f$. Hence also their composition is in $\mathcal{P}_0^f$. □

□

We have the following corollary of Theorem 10.31, part (i).

Corollary 10.38. *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*

$$J_q^f M = \left\{ \begin{pmatrix} tx_1 + O(t^2) \\ t^2 x_2 + O(t^3) \\ \vdots \\ t^k x_k \end{pmatrix} : x_i \in \mathbb{R}^{n_i}, i = 1, \ldots, k \right\}. \tag{10.28}$$

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

$$\begin{aligned} \sum_{i=1}^k t^i \xi_i &= t\xi_1 + t^2 \xi_2 + \ldots + t^k \xi_k \\ &= tx_{1,1} + t^2(x_{2,1} + x_{2,2}) + \ldots + t^k(x_{k,1} + \ldots + x_{k,k}) \\ &= (tx_{1,1} + t^2 x_{2,1} + \ldots + t^k x_{k,1}, t^2 x_{2,2} + \ldots + t^k x_{k,2}, t^k x_{k,k}) \end{aligned}$$

□

Corollary 10.39. *The nonholonomic tangent space* $T_q^f M$ *is a smooth manifold of dimension* $\dim T_q^f M = \sum_{i=1}^{k(q)} n_i(q)$. *In privileged coordinates we have*

$$T_q^f M = \left\{ \begin{pmatrix} tx_1 \\ t^2 x_2 \\ \vdots \\ t^k x_k \end{pmatrix} : x_i \in \mathbb{R}^{n_i}, i = 1, \ldots, k \right\}, \tag{10.29}$$

and dilations $\{\delta_\alpha\}_{\alpha>0}$ *acts on* $T_q^f M$ *in the following quasi-homogeneous way*

$$\delta_\alpha(tx_1, \ldots, t^k x_k) = (\alpha t x_1, \ldots, \alpha^k t^k x_k).$$

Proof. It follows directly from Corollary 10.38 that two elements j and j' can be written in coordinates as

$$\begin{aligned} j &= (tx_1 + O(t^2), t^2 x_2 + O(t^3), \ldots, t^k x_k), \\ j' &= (ty_1 + O(t^2), t^2 y_2 + O(t^3), \ldots, t^k y_k). \end{aligned}$$

Moreover, thanks to Theorem 10.31, claim (ii), we have that $j \sim j'$ if and only if $x_i = y_i$ for all $i = 1, \ldots, k$. □

Remark 10.40. Notice that a polynomial differential operator homogeneous with respect to ν (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

$$\delta_t(x_1, \dots, x_k) = (tx_1, t^2 x_2, \dots, t^k x_k), \qquad t > 0. \tag{10.30}$$

In particular for a homogeneous vector field X of weight h it holds $\delta_{t*}X = t^{-h}X$.

Now we can improve Proposition 10.25 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 10.41. *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*

$$V = \begin{pmatrix} v_1(x) \\ v_2(x) \\ \vdots \\ v_k(x) \end{pmatrix} \quad \Longrightarrow \quad \widehat{V} = \begin{pmatrix} \widehat{v}_1(x) \\ \widehat{v}_2(x) \\ \vdots \\ \widehat{v}_k(x) \end{pmatrix} \tag{10.31}$$

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 10.11, 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

$$(J_q^k V)(J_q^k \gamma) = \begin{pmatrix} tv_1(tx_1 + O(t^2), \dots, t^k x_k) \\ tv_2(tx_1 + O(t^2), \dots, t^k x_k) \\ \vdots \\ tv_k(tx_1 + O(t^2), \dots, t^k x_k) \end{pmatrix} \tag{10.32}$$

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 10.32 we have

$$(J_q^k V)\Big|_{T_q^f M} = \begin{pmatrix} tv_1(tx_1 + O(t^2), \ldots, t^k x_k) \\ tv_2(tx_1 + O(t^2), \ldots, t^k x_k) \\ \vdots \\ tv_k(tx_1 + O(t^2), \ldots, t^k x_k) \end{pmatrix} = \begin{pmatrix} t\widehat{v}_1 + O(t^2) \\ t^2\widehat{v}_2 + O(t^3) \\ \vdots \\ t^k\widehat{v}_k + O(t^{m+1}) \end{pmatrix} \tag{10.33}$$

from which (10.31) follows. $\square$

Remark 10.42. Notice that, since $\widehat{v}_i$ is a homogeneous function of weight $i-1$, it depends only on variables $x_1, \ldots, x_{i-1}$ of weight equal of smaller than its weight. Hence $\widehat{V}$ has the following triangular form

$$\widehat{V}(x) = \begin{pmatrix} \widehat{v}_1 \\ \widehat{v}_2(x_1) \\ \vdots \\ \widehat{v}_k(x_1, \ldots, x_{k-1}) \end{pmatrix} \tag{10.34}$$

A triangular vector field of the kind (10.34) is complete and its flow can be easily computed by a step by step substitution.

10.3.2 Existence of privileged coordinates: proof of Theorem 10.30.

Fix a generating frame $\{f_1, \ldots, f_m\}$ of the distribution $\mathcal{D}$. Assume that $\mathcal{D}$ is bracket generating of step k at the point q

$$\mathcal{D}_q^1 \subset \mathcal{D}_q^2 \subset \ldots \subset \mathcal{D}_q^k = T_q M. \tag{10.35}$$

Denote by $d_j := \dim \mathcal{D}_q^j$ the dimension of the elements of the flag, for $j = 1, \ldots, k$.

Definition 10.43. A set $V_1, \ldots, V_n$ 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) $V_i = \pi_i(f_1, \ldots, f_m)$, where π_i is some bracket polynomial, for $i = 1, \ldots, n$,

(b) $\deg \pi_i \leq j$ for every $i \leq d_j$,

(c) $\mathcal{D}_q^j = \operatorname{span}\{V_1(q), \ldots, V_{d_j}(q)\}$, for $j = 1, \ldots, k$.

A privileged frame can be constructed as follows: choose $V_1, \ldots, V_{d_1}$ among the vector fields $\{f_1, \ldots, f_m\}$ in such a way that $\mathcal{D}_q = \operatorname{span}\{V_1(q), \ldots, V_{d_1}(q)\}$, then fix $V_{d_1+1}, \ldots, V_{d_2}$ among the set $\{[f_i, f_j] : i, j = 1, \ldots, m\}$ in such a way that $\mathcal{D}_q^2 = \operatorname{span}\{V_1(q), \ldots, V_{d_2}(q)\}$, and so on.

Remark 10.44. Given a privileged frame $V_1, \ldots, V_n$, one can introduce on $T_q M$ the weight on the coordinates $(y_1, \ldots, y_n)$ induced by the flag. In other words we write every element v in $T_q M$ along the basis $V_1(q), \ldots, V_n(q)$ and set

$$v = (y_1, \ldots, y_n) = \sum_{i=1}^n y_i V_i(q), \qquad \text{where} \quad \nu(y_i) = w_i := j \quad \text{if} \quad d_{j-1} < i \leq d_j$$

Identifying $v \in T_q M$ with a constant vector field, it makes sense to consider the value of a polynomial bracket $X = \pi(f_1, \ldots, f_m)$ at the point q and consider its weight $\nu(X)$.

Privileged coordinates are then easily build in terms of a privileged frame.

Theorem 10.45. *Let $V_1, \dots, V_n$ be a privileged frame at q. Then the map*

$$\Psi : \mathbb{R}^n \to M, \qquad \Psi(s_1, \dots, s_n) = q \circ e^{s_1 V_1} \circ \dots \circ e^{s_n V_n}, \tag{10.36}$$

is a local diffeomorphism at $s = 0$ and its inverse Ψ^{-1} defines privileged coordinates around q.

Proof. The map (10.46) is a local diffeomorphism at $s = 0$ since

$$\left.\frac{\partial \Psi}{\partial s_i}\right|_{s=0} = V_i(q), \qquad i = 1, \dots, n \tag{10.37}$$

and these vectors are linearly independent by property (c) of privileged frame. To complete the proof we have to show that:

(i) $\Psi_*^{-1}(\mathcal{D}_q^j) = \operatorname{span}\left\{\frac{\partial}{\partial s_1}, \dots, \frac{\partial}{\partial s_{d_j}}\right\}$, 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 Ψ defines linearly adapted coordinates, easily follows from property (c) of privileged frame and (10.37). On the other hand, claim (ii) is not trivial since requires the computation of the differential of Ψ at every point, and not only at $s = 0$.

We prove the following preliminary result.

Lemma 10.46. *Let $X = \pi(f_1, \dots, f_m)(q) \in \mathrm{Vec}(T_q M)$ be a bracket polynomial with $\nu(X) \leq h$. Given a polynomial vector field on $T_q M$*

$$Y(y) := \sum y_{i_l} \cdots y_{i_1} (\operatorname{ad} V_{i_l} \circ \cdots \circ \operatorname{ad} V_{i_1} X)(q) \tag{10.38}$$

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 10.46. 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 (10.38) belongs to $\mathcal{F}^{(-h)}$. Then if we rewrite the sum (10.38) in terms of the basis $V_i(q)$, for $i = 1, \dots, n$ we have that every coefficient $p_i(y)$ must belong to $\mathcal{F}^{(w_i - h)}$, since $\nu(V_i(q)) = w_i$. $\square$

The proof of existence of privileged coordinates is completed by the following proposition, applied in the particula case $h = 1$.

Proposition 10.47. *Let $X = \pi(f_1, \ldots, f_m)$ be a bracket polynomial with $\nu(X) \leq h$ and Ψ be the map defined in* (10.46)*. Then $\Psi_*^{-1} X \in \mathcal{F}^{(-h)}$.*

Proof. Writing the vector field $\Psi_*^{-1} X$ in coordinates

$$\Psi_*^{-1} X = \sum_{i=1}^{n} a_i(s) \frac{\partial}{\partial s_i}, \tag{10.39}$$

the statement is proved if we show that $a_i \in \mathcal{F}^{(w_i - h)}$. We compute the differential of Ψ (cf. also Exercice 2.31)

$$\begin{aligned}
\Psi_* \frac{\partial}{\partial s_i} &= \frac{\partial}{\partial \varepsilon}\bigg|_{\varepsilon=0} q \circ e^{s_1 V_1} \circ \cdots \circ e^{(s_i+\varepsilon)V_i} \circ \cdots \circ e^{s_n V_n} \\
&= q \circ e^{s_1 V_1} \circ \cdots \circ e^{s_i V_i} \circ V_i \circ e^{s_{i+1} V_{i+1}} \circ \cdots \circ e^{s_n V_n} \\
&= \underbrace{q \circ e^{s_1 V_1} \circ \cdots \circ e^{s_n V_n}}_{\Psi(s)} \circ e^{-s_n V_n} \circ \cdots \circ e^{-s_{i+1} V_{i+1}} \circ V_i \circ e^{s_{i+1} V_{i+1}} \circ \cdots \circ e^{s_n V_n}.
\end{aligned}$$

In geometric notation we can write

$$\Psi_* \frac{\partial}{\partial s_i} = e_*^{s_n V_n} \cdots e_*^{s_{i+1} V_{i+1}} V_i \Big|_{\Psi(s)}. \tag{10.40}$$

Remember that, as operator on functions, $e_*^{tY} = e^{-t \operatorname{ad} Y}$. This implies that in (10.40) we have a series of bracket polynomials. Applying Ψ_* to (10.39) 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

$$e_*^{-s_1 V_1} \cdots e_*^{-s_n V_n} X \Big|_q = \sum_{i=1}^{n} a_i(s) e_*^{-s_1 V_1} \cdots e_*^{-s_{i-1} V_{i-1}} V_i \Big|_q. \tag{10.41}$$

Rewriting the last identity in the basis $V_1(q), \ldots, V_n(q)$ we have

$$\sum_{i=1}^{n} b_i(s) V_i(q) = \sum_{i,j=1}^{n} a_i(s)(V_i(q) + \varphi_{ij}(s) V_j(q)), \tag{10.42}$$

for some smooth functions b_i, φ_{ij} such that $\varphi_{ij}(0) = 0$. Applying Lemma 10.46 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 $B(s) = (b_1(s), \dots, b_n(s))$, $A(s) = (a_1(s), \dots, a_n(s))$ 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 10.48. The previous proof can be rewritten in purely algebraic way through chronological notation. In the above proof nothing changes if we consider some permutation $\sigma = (i_1, \dots, i_n)$ of $(1, \dots, n)$ and work with the map

$$\Psi_\sigma : (s_1, \dots, s_n) \mapsto q \circ e^{s_{i_n} V_{i_n}} \circ \dots \circ e^{s_{i_1} V_{i_1}}.$$

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 $i = 1, \dots, n$.

We summarize the previous considerations in the next corollary.

Corollary 10.49. *Let $V_1, \dots, V_n$ be a privileged frame at q and $\sigma = (i_1, \dots, i_n)$ a permutation of $\{1, \dots, n\}$. Then the map*

$$\Psi_\sigma : \mathbb{R}^n \to M, \qquad \Psi_\sigma(s_1, \dots, s_n) = q \circ e^{s_{i_n} V_{i_n}} \circ \dots \circ e^{s_{i_1} V_{i_1}}, \tag{10.43}$$

is a local diffeomorphism at $s = 0$ and its inverse Ψ_σ^{-1} defines privileged coordinates around q.

Remark 10.50. As a particular case of Corollary 10.49 we can consider the coordinate map

$$\Phi : (x_1, \dots, x_n) \mapsto q \circ e^{x_n V_n} \circ \dots \circ e^{x_1 V_1}.$$

Computing the differential Φ_* (cf. also Exercice 2.31) it is easy to see that for every $i = 1, \dots, n$

$$\Phi_*^{-1} V_i \Big|_{x_1 = \cdots = x_{i-1} = 0} = \partial_{x_i}. \tag{10.44}$$

This implies in particular that for $i = 1, \dots, d_1$ we have in coordinates

$$V_i = \partial_{x_i} + \sum_{j \geq d_1} a_{ij}(x_1, \dots, x_{d_1}) \partial_{x_j}, \tag{10.45}$$

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 10.51. Let $V_1, \dots, V_n$ be a privileged frame at q. Prove that the map

$$\Psi_+ : \mathbb{R}^n \to M, \qquad \Psi_+(s_1, \dots, s_n) = q \circ e^{\sum_{i=1}^n s_i V_i} \tag{10.46}$$

is a local diffeomorphism at $s = 0$ and its inverse Ψ_+^{-1} defines privileged coordinates around q.

10.3.3 Nonholonomic tangent spaces in low dimension

In Riemannian geometry the above procedure becomes very easy since when $k = 1$ we have that $J_q^k M = T_q M$ 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 $(d_1, \ldots, d_k)$ 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 10.52 (Heisenberg group)**.** Assume $n = 3$ and that 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

$$V_1 = f_1, \qquad V_2 = f_2, \qquad V_3 = [f_1, f_2]. \tag{10.47}$$

Using privileged coordinates defined in Remark 10.50, we have that

$$V_1 = f_1 = \partial_{x_1}, \qquad V_2 = f_2 = \partial_{x_2} + \alpha x_1 \partial_{x_3}, \tag{10.48}$$

for some $\alpha \in \mathbb{R}$. On the other hand since

$$V_3 = [f_1, f_2] = \alpha \partial_{x_3} \tag{10.49}$$

and $V_3(0) = \partial_{x_3}$ from (10.44) we get $\alpha = 1$. This gives the following normal form for the generating frame of the nonholonomic tangent space

$$f_1 = \partial_{x_1}, \qquad f_2 = \partial_{x_2} + x_1 \partial_{x_3}. \tag{10.50}$$

□

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 space.

Example 10.53 (Martinet space)**.** 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

$$V_1 = f_1, \qquad V_2 = f_2, \qquad V_3 = [f_1, [f_1, f_2]]. \tag{10.51}$$

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

$$\varphi : \mathbb{R}^2 \to T_qM/\mathcal{D}_q, \qquad \varphi(u_1, u_2) := [f_u, [f_1, f_2]](q) \mod \mathcal{D}_q.$$

Since φ is surjective (by bracket-generating assumption) and $\dim T_qM/\mathcal{D}_q = 1$, then $\ker \varphi$ is one dimensional. Thus, up to a rotation of constant angle of the generating frame 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

$$[f_2, [f_1, f_2]] = 0. \tag{10.52}$$

Using privileged coordinates defined in Remark 10.50, we have that

$$V_1 = f_1 = \partial_{x_1}, \qquad V_2 = f_2 = \partial_{x_2} + x_1 a(x_1, x_2)\partial_{x_3}, \tag{10.53}$$

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

$$f_1 = \partial_{x_1}, \qquad f_2 = \partial_{x_2} + (\alpha x_1^2 + \beta x_1 x_2)\partial_{x_3}. \tag{10.54}$$

Since $[f_1, [f_1, f_2]] = 2\alpha\partial_{x_3}$, the requirement $V_3|_{x=0} = \partial_{x_3}$ in (10.51) gives $\alpha = 1/2$. Moreover for this value o α we have $[f_2, [f_1, f_2]] = \beta\partial_{x_3}$ and the condition (10.52) gives $\beta = 0$. We have then the normal form for the generating frame of the nonholonomic tangent space

$$f_1 = \partial_{x_1}, \qquad f_2 = \partial_{x_2} + \frac{1}{2}x_1^2\partial_{x_3}, \qquad f_3 = \partial_{x_3} \tag{10.55}$$

$\square$

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 nonholonomic tangent space is the Grushin plane.

Example 10.54 (Grushin plane)**.** Assume $n = 2$ and that 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 rame 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 10.50, 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 frame of the nonholonomic tangent space

$$f_1 = \partial_{x_1}, \qquad f_2 = x_1\partial_{x_2}. \tag{10.56}$$

10.4 Metric meaning

In this section we study the interplay between the distance and the nonholonomic tangent space. In other words we consider a sub-Riemannian manifold $(M, \mathbf{U}, f)$ and we want to understand what is the metric structure which is naturally defined on the nonholonomic tangent space and in which sense the latter gives a good approximation of the original structure in a neighborhood of a point.

To this aim, we start by exploring in more details, 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 10.55. *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*

$$\varepsilon\delta_{\frac{1}{\varepsilon}*}V = \widehat{V} + \varepsilon W^\varepsilon, \tag{10.57}$$

where $\{\delta_\alpha\}_{\alpha>0}$ denotes the family of dilations defined in (10.30) *and W^ε depends smoothly on the parameter ε. In particular $\widehat{V}$ is characterized as follows*

$$\widehat{V} = \lim_{\varepsilon\to 0} \varepsilon\delta_{\frac{1}{\varepsilon}*}V. \tag{10.58}$$

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

$$\delta_{\frac{1}{\varepsilon}*}V = \delta_{\frac{1}{\varepsilon}*}\widehat{V} + \delta_{\frac{1}{\varepsilon}*}W = \frac{1}{\varepsilon}\widehat{V} + \delta_{\frac{1}{\varepsilon}*}W, \tag{10.59}$$

where we used the homogeneity of $\widehat{V}$ (cf. Remark 10.40). Noting that $W \in \mathcal{F}^{(0)}$, hence setting $W^\varepsilon := \varepsilon\delta_{\frac{1}{\varepsilon}*}W$ we have that W^ε is smooth with respect to ε and $\varepsilon W^\varepsilon \to 0$ for $\varepsilon \to 0$. □

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 filed in the Euclidean case.

10.4.1 Convergence of the sub-Riemannian distance and the Ball-Box theorem

Given a sub-Riemannian structure on M, with $\dim M = n$, let us denote by $\{f_1, \ldots, f_m\}$ a generating frame and fix a point q where the structure has step k.

Once we have fixed a privileged coordinate chart, we can treat the vector fields $\{f_1, \ldots, f_m\}$ as vector fields in $\mathbb{R}^n$, introduce the family of dilations $\{\delta_\alpha\}_{\alpha>0}$ defined in (10.30) and introduce the vector fields

$$f_i^\varepsilon := \varepsilon\delta_{\frac{1}{\varepsilon}*}f_i, \qquad i = 1, \ldots, m. \tag{10.60}$$

Thanks to Lemma 10.55 we have that $f_i^\varepsilon \to \widehat{f}_i$ for $i = 1, \ldots, m$ and we can define the sub-Riemannian structure f^ε and $\widehat{f}$ on $\mathbb{R}^n$ defined by the generating frames $\{f_1^\varepsilon, \ldots, f_m^\varepsilon\}$ and $\{\widehat{f}_1, \ldots, \widehat{f}_m\}$ respectively.

From the definition (10.60) of the vector fields f_i^ε, it follows directly that the sub-Riemannian distance defined by these vector fields is, up to a rescaling, the original sub-Riemannian distance in the dilated coordinates. More precisely we have the following relation.

Proposition 10.56. *Let d^ε and d be the sub-Riemannian distances on $\mathbb{R}^n$ associated with the sub-Riemannian structures f^ε and f, respectively. Then for every $x, y \in \mathbb{R}^n$ we have*

$$d^\varepsilon(x,y) = \frac{1}{\varepsilon} d(\delta_\varepsilon(x), \delta_\varepsilon(y)). \tag{10.61}$$

Proposition 10.56 is saying that d^ε 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 10.57. *Let $B(x,r)$ (resp. $B^\varepsilon(x,r)$) be the sub-Riemannian ball with respect to the distance d (resp. d^ε). Then for every $r > 0$ and $\varepsilon > 0$ one has*

$$\delta_\varepsilon(B^\varepsilon(x,r)) = B(\delta_\varepsilon x, \varepsilon r). \tag{10.62}$$

In particular $\delta_\varepsilon(B^\varepsilon(0,1)) = B(0,\varepsilon)$ for every $\varepsilon > 0$.

The previous results relates the original distance d with the approximating one d^ε. Next we move to the convergence of d^ε for $\varepsilon \to 0$.

We start from an auxiliary proposition, studying the convergence of the end-point maps. Denote E^ε_x and $\widehat{E}_x$ the end-point map of the approximating frame and the nilpotent one based at a point $x \in \mathbb{R}^n$.

Proposition 10.58. *Let $x \in \mathbb{R}^n$. Then $E^\varepsilon_x \to \widehat{E}_x$ 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

$$\dot{x} = \sum_{i=1}^m u_i(t) f^\varepsilon_i(x), \qquad \dot{x} = \sum_{i=1}^m u_i(t) \widehat{f}_i(x), \tag{10.63}$$

with some fixed initial condition $x(0) = x_0 \in \mathbb{R}^n$. Using Lemma 10.55, we write $f^\varepsilon_i = \widehat{f}_i + \varepsilon W^\varepsilon_i$ and the first equation in (10.63) becomes

$$\dot{x} = \sum_{i=1}^m u_i(t) \widehat{f}_i(x) + \varepsilon \sum_{i=1}^m u_i(t) W^\varepsilon_i(x). \tag{10.64}$$

In the right hand side the term

$$W^\varepsilon_t(x) := \varepsilon \sum_{i=1}^m u_i(t) W^\varepsilon_i(x), \tag{10.65}$$

is a non-autonomous vector field smoothly depending on the parameter ε. Moreover $W^\varepsilon_t(x) \to 0$ when $\varepsilon \to 0$. 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 converges and the convergence can be taken uniform Notice that, since nilpotent vector fields are complete (cf. Remark 10.42), 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 10.58 holds in the C^∞ sense.

We now prove a key uniform Hölder estimate (with respect to ε) for the approximating sub-Riemannian distance.

Proposition 10.59. *For every compact $K \subset \mathbb{R}^n$ there exists $\varepsilon_0, C > 0$, depending on K, such that*

$$d^\varepsilon(x,y) \leq C|x-y|^{1/k}, \qquad \forall\, \varepsilon \in (0,\varepsilon_0),\ \forall\, x,y \in K. \tag{10.66}$$

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 10.43), such that $\widehat{V}_i = \pi_i(\widehat{f}_1, \dots, \widehat{f}_k)$ for some bracket polynomials π_i, where $i = 1, \dots, n$. By construction we have

$$\widehat{V}_1(0) \wedge \ldots \wedge \widehat{V}_n(0) \neq 0. \tag{10.67}$$

By continuity, this implies that they are linearly independent also in a small neighborhood of the origin and, thanks to quasi-homogeneity, this implies

$$\widehat{V}_1(x) \wedge \ldots \wedge \widehat{V}_n(x) \neq 0, \qquad \forall\, x \in \mathbb{R}^n. \tag{10.68}$$

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

$$V_1^\varepsilon(x) \wedge \ldots \wedge V_n^\varepsilon(x) \neq 0, \qquad \forall\, x \in K, \forall\, \varepsilon \leq \varepsilon_0. \tag{10.69}$$

Recall that by Lemma 10.35, 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 π_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 $\varepsilon > 0$

$$\Phi^\varepsilon(t_1, \dots, t_n, x) := x \circ \overrightarrow{\exp} \int_0^1 f^\varepsilon_{u_1(t_1^{1/w_1}, s)} ds \circ \ \ldots\ \circ \overrightarrow{\exp} \int_0^1 f^\varepsilon_{u_n(t_n^{1/w_n}, s)} ds. \tag{10.70}$$

Notice that we have the expansion

$$x \circ \overrightarrow{\exp} \int_0^1 f^\varepsilon_{u_i(t_i^{1/w_i}, s)} ds = x + t_i V_i^\varepsilon(x) + O(t_i^{\frac{w_i+1}{w_i}}). \tag{10.71}$$

In particular (10.71) 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 Φ^ε is C^1 as a function of t, being a composition of C^1 maps. Clearly Φ^ε is smooth as a function of x. Combining the contributions of (10.71) we obtain the expansion

$$\Phi^\varepsilon(x; t_1, \dots, t_n) = x + \sum_{i=1}^n t_i V_i^\varepsilon(x) + o(|t|), \tag{10.72}$$

This implies that the partial derivatives

$$\frac{\partial \Phi^\varepsilon}{\partial t_i}\Big|_{t=0} = V_i^\varepsilon(x), \tag{10.73}$$

are linearly independent at the origin thanks to (10.69) and Φ^ε is a local diffeomorphism at $t = (t_1, \ldots, t_n) = 0$. Applying classical Implicit Function Theorem (see Corollary 2.54) we have that there exists a constant $c > 0$ satifying

$$B(x, cr) \subset \Phi^\varepsilon(x; B(0, r)), \quad x \in K, \tag{10.74}$$

where here $B(x, r)$ denotes the ball in $\mathbb{R}^n$ and c is independent of x, ε 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 $E_x^\varepsilon, \widehat{E}_x$), and with $\mathcal{B}$ the unit ball in $L_2^k[0,1]$.

We claim that (10.74) implies that there exists a constant c' such that for all $r > 0$ and $\varepsilon > 0$ small enough

$$B(x, c'r) \subset E_x^\varepsilon(r^{\frac{1}{m}}\mathcal{B}), \tag{10.75}$$

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

$$t \in B(0, r) \Rightarrow u_i(t, \cdot) \in c_i r\mathcal{B}, \tag{10.76}$$

$$\Rightarrow u_i(t^{1/w_i}, \cdot) \in c_i r^{1/w_i}\mathcal{B}, \tag{10.77}$$

for all $r > 0$ small enough. For such values of $r > 0$ we have thanks to the inclusion (10.75) 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^ε is the infimum of norm of u such that $E_x^\varepsilon(u) = y$. From this it follows the inequality for every $x, y \in K$

$$d^\varepsilon(x, y) \leq c^{-\frac{1}{k}}|x - y|^{\frac{1}{k}} \tag{10.78}$$

□

We are now ready to prove the main result of this section.

Theorem 10.60. *$d^\varepsilon \to \widehat{d}$ uniformly on compacts sets in $\mathbb{R}^n \times \mathbb{R}^n$.*

Proof. By Proposition 10.59 it is sufficient to prove the pointwise convergence. We prove the following inequalities

$$\lim_{\varepsilon \to 0^+} d^\varepsilon(x, y) = \widehat{d}(x, y) \tag{10.79}$$

but (10.79) is a consequence of Theorem 3.51 and the fact that the vector fields f_i^ε converge to $\widehat{f}_i$ thanks to Lemma 10.55. □

Combining Proposition 10.59 and Theorem 10.60 we obtain the following corollary.

Corollary 10.61. *For every compact $K \subset \mathbb{R}^n$ there exists $C > 0$, depending on K, such that*

$$\widehat{d}(x, y) \leq C|x - y|^{1/k}, \qquad \forall\, x, y \in K, \tag{10.80}$$

where k is the degree of nonholonomy of the sub-Riemannian structure.

The uniform convergence given in Theorem 10.60 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

$$\mathrm{Box}(\varepsilon) = \{x \in \mathbb{R}^n : |x_i| \leq \varepsilon^i, i = 1, \ldots, k\}. \tag{10.81}$$

Theorem 10.62 (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 10.60.

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 10.60 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^ε. Applying the dilation δ_ε 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^ε we have that $\delta_\varepsilon(B^\varepsilon(0, 1)) = B(0, \varepsilon)$ (cf. also Corollary 10.57). $\square$

10.5 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 frame $\{f_1, \ldots, f_m\}$ 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 $T_q^f M$.

We start with a basic observation on the structure of the Lie algebra generated by $\{\widehat{f}_1, \ldots, \widehat{f}_m\}$.

Proposition 10.63. *The Lie algebra* $\mathrm{Lie}\{\widehat{f}_1, \ldots, \widehat{f}_m\}$ *is a finite-dimensional nilpotent Lie algebra 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. $\square$

Consider now the Lie algebra of vector fields $L := \mathrm{Lie}\{\widehat{f}_1, \ldots, \widehat{f}_m\}$. This Lie algebra is finite-dimensional and nilpotent thanks to Proposition 10.63. Denote by $\mathcal{G}$ the Lie group of associated flows (cf. Section 7.1)

$$\mathcal{G} = \{e^{t_1 \widehat{f}_{i_1}} \circ \ldots \circ e^{t_j \widehat{f}_{i_j}} : t_i \in \mathbb{R}, j \in \mathbb{N}\}. \tag{10.82}$$

endowed with the product $\circ$. By construction this is a nilpotent Lie 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

$$e^{t\widehat{f}_i} \cdot [j] := [\gamma \circ e^{t\widehat{f}_i}], \qquad j = J_q^k \gamma \in J_q^k M. \tag{10.83}$$

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 10.3, and in particular Theorem 10.31, we have the following result

Theorem 10.64. *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 $\mathcal{G}/\mathcal{G}_0$.*

Remark 10.65. The diffeomorphism given by Theorem 10.64 was built explicitly thanks to privileged coordinates in in Section 10.3.

Notice that indeed this could also be seen as a consequence of the theory of Lie groups. Indeed it is not difficult to see that actually in the proof of Theorem 10.31 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 [73].

Next we give a purely algebraic interpretation of this construction at the level of Lie algebras. Let us first recall some definitions.

Definition 10.66. The *free associative algebra* $\mathcal{A}_m$ (or $\mathcal{A}(x_1, \ldots, x_m)$) generated by $x_1, \ldots, x_m$ 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* Lie_m or $\mathrm{Lie}\{x_1, \ldots, x_m\}$ 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 Lie_m^k or $\mathrm{Lie}^k\{x_1, \ldots, x_m\}$, is the quotient $\mathrm{Lie}_m^k = \mathrm{Lie}_m/\mathcal{I}^{k+1}$ of the free Lie algebra 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 $\mathrm{Lie}_k\{x_1, \ldots, x_m\}$ be the free Lie algebra nilpotent of step k generated by the elements $x_1, \ldots, x_m$. Notice when taking an element $\pi \in \mathrm{Lie}_k\{x_1, \ldots, x_m\}$ we can define a vector field $\pi(X_1, \ldots, X_m)$ replacing generators with vector fields $X_1, \ldots, X_m$ (on $\mathbb{R}^n$).

Definition 10.67. Given a sub-Riemannian structure defined by the generating frame $\{f_1, \ldots, f_m\}$ that is bracket generating of step k at a point q, we define the *core algebra*

$$C_q := \{\pi \in \mathrm{Lie}_k\{X_1, \ldots, X_m\} \,|\, \pi(f_1, \ldots, f_m)(q) \in \mathcal{D}_q^{\deg \pi - 1}\}. \tag{10.84}$$

Exercise 10.68. (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 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 10.69. *There exists a canonical isomorphism*

$$\phi : \mathcal{G}_m^k / \mathcal{C}_q \to T_q^f M.$$

Its differential ϕ_* sends generators $X_1, \dots, X_m$ to $\widehat{f}_1, \dots, \widehat{f}_m$.

Remark 10.70. The core algebra can be rewritten in privileged coordinates in terms of the nilpotentb approximation of the generators as follows $\{\widehat{f}_1, \dots, \widehat{f}_m\}$ 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 10.71 (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 10.54.

We have shown that the nonholonomic tangent space has the following normal form

$$\widehat{f}_1 = \partial_{x_1}, \qquad \widehat{f}_2 = x_1 \partial_{x_2}. \tag{10.85}$$

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

$$\mathrm{Lie}\{\widehat{f}_1, \widehat{f}_2\} = \mathrm{span}\{\widehat{f}_1, \widehat{f}_2, \widehat{f}_3\} \tag{10.86}$$

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 (does not belong to $\mathcal{D}_0^1$), hence $C_0 = \mathrm{span}\{\widehat{f}_2\}$.

10.5.1 The equiregular case

The last two statements concerns the case of a equiregular distribution. In this case one can show that the subgroup $\mathcal{G}_0$ of $\mathcal{G}$ is trivial.

Proposition 10.72. *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 $T_q^f M$ is a Lie 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 (10.84), this is equivalent to prove the following claim: for every $X = \pi(f_1, \dots, f_m)$ bracket polynomial of degree $\deg \pi \leq h$ 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

$$\mathcal{D}^i_x = \operatorname{span}\{V_1(x), \dots, V_{d_i}(x)\}. \tag{10.87}$$

Let $X = \pi(f_1, \dots, f_m)$ be a bracket polynomial of degree $\deg \pi \leq h$. Then there exist smooth functions a_j such that

$$X(x) = \sum_{j : w_j \leq h} a_j(x) V_j(x), \qquad \forall\, x \in O_q. \tag{10.88}$$

Thanks to (10.87), $X(q) \in \mathcal{D}^{h-1}_q$ is equivalent to require that $a_j(q) = 0$ for every j such that $w_j = h$. Let us compute

$$[f_i, X] = \left[f_i, \sum_{w_j \leq h} a_j V_j \right] = \sum_{w_j \leq h} a_j [f_i, V_j] + f_i(a_j) V_j. \tag{10.89}$$

Evaluating (10.89) at the point q and using that $a_j(q) = 0$ for every j such that $w_j = h$, it follows that $[f_i, X](q) \in \mathcal{D}^h_q$ for every $i = 1, \dots, m$, that is our claim. $\square$

Corollary 10.73. *Assume that the sub-Riemannian structure is equiregular and* $\{f_1, \dots, f_m\}$ *is a generating frame. Then* $\widehat{f}_1, \dots, \widehat{f}_m$ *are a basis of left-invariant vector fields on* $T^f_q M$.

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 fields 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 10.74. Prove the two statement contained in the proof of Corollary 10.73.

10.6 Carnot groups: normal forms in low dimension

In this section we provide normal forms for Carnot groups in dimension less 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 $d_i = \dim \mathcal{D}^i_q$ is independent on q. Denote by k the step of the sub-Riemannian structure, namely k is the smallest integer such that $d_k = \dim M$. The sequence of integers $(d_1, \dots, d_k)$ is called *growth vector* of the sub-Riemannian structure.

Exercise 10.75. Prove that if the structure is equiregular of step k, then 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 10.75 it easily follows that the possibilities for the growth vector in dimension less or equal than 5 are the following:

- $(2, 3)$, 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 prevuois list. In every case but the last one, the normal form is unique.

Theorem 10.76. *Let $(M, \mathbf{U}, f)$ be an equiregular sub-Riemannian manifold. 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*

$$f_1 = \partial_{x_1},$$
$$f_2 = \partial_{x_2} + x_1 \partial_{x_3}.$$

- (Engel). *If the growth vector is $(2,3,4)$, then the orthonormal frame can be chosen as*

$$f_1 = \partial_{x_1},$$
$$f_2 = \partial_{x_2} + x_1 \partial_{x_3} + x_1 x_2 \partial_{x_4}.$$

- (Quasi-Heisenberg). *If the growth vector is $(3,4)$, then the orthonormal frame can be chosen as*

$$f_1 = \partial_{x_1},$$
$$f_2 = \partial_{x_2} + x_1 \partial_{x_4},$$
$$f_3 = \partial_{x_3}.$$

- (Cartan rank 2). *If the growth vector is $(2,3,5)$, then the orthonormal frame can be chosen as*

$$f_1 = \partial_{x_1},$$
$$f_2 = \partial_{x_2} + x_1 \partial_{x_3} + \frac{1}{2} x_1^2 \partial_{x_4} + x_1 x_2 \partial_{x_5}.$$

- (Goursat rank 2). *If the growth vector is $(2,3,4,5)$, then the orthonormal frame can be chosen as*

$$f_1 = \partial_{x_1},$$
$$f_2 = \partial_{x_2} + x_1 \partial_{x_3} + \frac{1}{2} x_1^2 \partial_{x_4} + \frac{1}{6} x_1^3 \partial_{x_5}.$$

- (Cartan rank 3). *If the growth vector is $(3,5)$, then the orthonormal frame can be chosen as*

$$f_1 = \partial_{x_1} - \frac{1}{2} x_2 \partial_{x_4},$$
$$f_2 = \partial_{x_2} + \frac{1}{2} x_1 \partial_{x_4} - \frac{1}{2} x_3 \partial_{x_5},$$
$$f_3 = \partial_{x_3} + \frac{1}{2} x_2 \partial_{x_5}.$$

- (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 $X_1,\dots,X_m$ a basis of a Lie algebra $\mathfrak{g}$. The coefficients c_{ij}^{ℓ} 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

$$\dim \operatorname{span}\{[X_1,X_2],[X_1,X_3],[X_2,X_3]\}/\mathcal{D}_0 = 1. \tag{10.90}$$

In other words, we can define the skew-simmetric bilinear map

$$\Phi(\cdot,\cdot) : \mathcal{D}_0 \times \mathcal{D}_0 \to T_0G/\mathcal{D}_0, \qquad \Phi(v,w) = [V,W](0) \mod \mathcal{D}_0 \tag{10.91}$$

where V,W are smooth vector fields such that $V(0)=v$ and $W(0)=w$. The condition (10.90) 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 10.53, 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

$$[f_1, f_2] = f_4, \qquad [f_1, f_4] = f_5.$$

(b). Growth vector equal to $(2, 3, 4, 5)$. Let $(M, \mathbf{U}, f)$ be a nilpotent $(3, 4, 5)$ 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

$$\begin{aligned} \dim \operatorname{span}\{X_1, X_2, [X_1, X_2]\} &= 3 \\ \dim \operatorname{span}\{X_1, X_2, [X_1, X_2], [X_1, [X_1, X_2]], [X_2, [X_1, X_2]]\} &= 4. \end{aligned} \tag{10.92}$$

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 (10.92) 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 the case $(4, 5)$ let us consider again the map

$$\Phi(\cdot, \cdot) : \mathcal{D}_0 \times \mathcal{D}_0 \to T_0G/\mathcal{D}_0, \qquad \Phi(v, w) = [V, W](0) \mod \mathcal{D}_0 \tag{10.93}$$

since $\dim T_0G/\mathcal{D}_0 = 1$, the map (10.93) is represented by a single 4×4 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. Up to relabelling 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 $f_5 := [f_1, f_2]$ we have that and setting $\alpha := \alpha_2/\alpha_1$ we get $[f_3, f_4] = \alpha f_5$. $\square$

Remark 10.77. In the proof of Theorem 10.76 we showed that the structure of Lie brackets can is uniquely determined (in the last example modulo a real parameter α) by the choice of a suitable orthonormal frame.

Of course the coordinate representation of vector fields satisfying these structural equation is not unique (compare for instance the vector fields in the case of the Heisenberg group with respect to those used in the prevuois chapters). Nevertheless all of them are obtained from the one described here with a change of variable, thanks to the Nagano principle [82].

Exercise 10.78. Prove that in the three examples described in Section 10.3.3 there is a unique normal form for the generating frame, even if the distribution is endowed with an inner product.

Chapter 11

Regularity of the sub-Riemannian distance

In this chapter we focus our attention on the analytical properties of the sub-Riemannian squared distance from a fixed point. In particular we want to answer to the following questions:

(i) Which is the (minimal) regularity of d^2 that one can expect?

(ii) Is the sub-Riemannian distance d^2 smooth? If not, can we characterize smooth points?

11.1 General properties of the distance function

In this section we recall and collect some general properties of the sub-Riemannian distance and results related to it, some of which we already proved in the previous chapters.

Let us consider a free sub-Riemannian structure $(M, \mathbf{U}, f)$ where the vector fields $f_1, \ldots, f_m$ define a generating family, i.e.

$$f : \mathbf{U} \to TM, \qquad f(u,q) = \sum_{i=1}^{m} u_i f_i(q)$$

Here $\mathbf{U}$ is a trivial Euclidean bundle on M of rank m.

Definition 11.1. Fix a point $q \in M$. The *flag* of the sub-Riemannian structure at the point q is the sequence of subspaces $\{\mathcal{D}_q^i\}_{i\in\mathbb{N}}$ defined by

$$\mathcal{D}_q^i := \operatorname{span}\{[f_{j_1}, \ldots, [f_{j_{l-1}}, f_{j_l}]](q),\ \forall\, l \leq i\}$$

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 all $i \geq 1$.

The bracket generating assumptions implies that

$$\forall\, q \in M,\ \exists\, m(q) > 0 \quad \text{s.t.} \quad \mathcal{D}_q^{m(q)} = T_q M$$

and $m(q)$ is called the *step* of the sub-Riemannian structure at q.

Exercise 11.2. 1. Prove that the filtration defined by the subspaces $\mathcal{D}^i_q$, for $i \geq 1$, is independent on the choice of a generating family (i.e., on the trivialization of $\mathbf{U}$).

2. Show that $m(q)$ does not depend on the generating frame. Prove that the map $q \mapsto m(q)$ is upper semicontinuous.

In Chapter 10 we already proved that the sub-Riemannian distance is Hölder continuous. For the reader's convenience, we recall here the statement.

Proposition 11.3. *For every $q \in M$ there exists a neighborhood O_q such that $\forall\, q_0, q_1 \in O_q$ and for every coordinate map $\phi : O_q \to \mathbb{R}^n$*

$$d(q_0, q_1) \leq C|\phi(q_0) - \phi(q_1)|^{1/m}$$

where $m = m(q)$ is the step of the sub-Riemannian structure at q.

11.2 Regularity of the sub-Riemannian distance

In this section we fix once for all a point $q_0 \in M$ and a closed ball $B = \overline{B}_{q_0}(r_0)$ such that B is compact. In particular for each $q \in B$ there exists a minimizer joining q_0 and q (see Corollary 8.63). In what follows we denote by $\mathfrak{f}$ the squared distance from q_0

$$\mathfrak{f}(\cdot) = \frac{1}{2}d^2(q_0, \cdot). \tag{11.1}$$

The main result of this chapter is the following.

Theorem 11.4. *The function $\mathfrak{f}\big|_B : B \to \mathbb{R}$ is smooth on a open dense subset of B.*

In the case of complete sub-Riemannian structures, since balls are compact for all radii, we have immediately the following corollary

Corollary 11.5. *Assume that M is a complete sub-Riemannian manifold and $q_0 \in M$. Then $\mathfrak{f}$ is smooth on an open and dense subset of M.*

We start by looking for necessary conditions for $\mathfrak{f}$ to be C^∞ around a point.

Proposition 11.6. *Let $q \in B$ and assume that $\mathfrak{f}$ is C^∞ at q. Then*

(i) there exists a unique length minimizer γ joining q_0 with q. Moreover γ is not abnormal and not conjugate.

(ii) $d_q\mathfrak{f} = \lambda_1$, where λ_1 is the final covector of the normal lift of γ.

Proof. Under the above assumptions the functional

$$\Psi : v \mapsto J(v) - \mathfrak{f}(F(v)), \qquad v \in L^\infty([0, T], \mathbb{R}^k), \tag{11.2}$$

is smooth and non negative. For every optimal trajectory γ, associated with the control u, that connects q_0 with q in time 1, one has

$$0 = d_u\Psi = d_uJ - d_q\mathfrak{f} \circ D_uF. \tag{11.3}$$

Thus, γ is a normal extremal trajectory, with Lagrange multiplier $\lambda_1 = d_q\mathfrak{f}$. By Theorem 4.26, we can recover γ by the formula $\gamma(t) = \pi \circ e^{(t-1)\vec{H}}(\lambda_1)$. Then, γ is the unique minimizer of J connecting its endpoints, and is normal.

Next we show that γ is not abnormal and not conjugate. For y in a neighbourhood O_q of q, let us consider the map

$$\Phi : O_q \mapsto T^*_{q_0}M, \qquad \Phi(y) = e^{-\vec{H}}(d_y\mathfrak{f}). \tag{11.4}$$

The map Φ, by construction, is a smooth right inverse for the exponential map, since

$$\exp(\Phi(y)) = \pi \circ e^{\vec{H}}(e^{-\vec{H}}(d_y\mathfrak{f})) = \pi(d_y\mathfrak{f}) = y. \tag{11.5}$$

This implies that q is a regular value for the exponential map. Since 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 11.7. Notice that from the proof it follows that if we only assume that $\mathfrak{f}$ is differentiable at q, we can still conclude that there exists a unique minimizer γ joining q_0 to q, and it is normal.

Moreover leu us notice that to conclude it is enough to assume that $\mathfrak{f}$ is twice differentiable at q. In particular a posteriori we can prove that whenever $\mathfrak{f}$ is is twice differentiable at q then it is C^∞.

Before going further in the study of the smoothness property of the distance function, we are already able to prove an important corollary of this result.

Denote, for $r > 0$, $S_r := \mathfrak{f}^{-1}(\frac{r^2}{2})$ the sub-Riemannian sphere of radius r centered at q_0

Corollary 11.8. *Assume that $\mathcal{D}_{q_0} \neq T_{q_0}M$. For every $r \leq r_0$, the sphere S_r contains a non smooth point of the function $\mathfrak{f}$.*

Proof. Since $r \leq r_0$, the sphere S_r is non empty and contained in a compact ball. Assume, by contradiction, that $\mathfrak{f}$ is smooth at every point of S_r. Then S_r is a level set defined by $\mathfrak{f}$ and $d_q\mathfrak{f} \neq 0$ for every $q \in S_r$ (since $d_q\mathfrak{f}$ is the nonzero covector attached at the final point of a geodesic, see Proposition 11.6). It follows 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

$$\Phi : S_r \to T^*_{q_0}M, \qquad \Phi(q) = e^{-\vec{H}}(d_q\mathfrak{f}),$$

By assumption $\mathfrak{f}$ is smooth, hence Φ is a smooth right inverse of the exponential map (see also (11.5)). In particular the differential of Φ is injective at every point. Moreover $H(\Phi(q)) = r$ since $\mathfrak{f}(q) = H(\lambda) = r$ for every $q \in S_r$. It follows that actually Φ defines a smooth immersion

$$\Phi : S_r \to H^{-1}(r) \cap T^*_{q_0}M \tag{11.6}$$

of the sphere S_r into the set

$$C_r := H^{-1}(r) \cap T^*_{q_0}M = \left\{ \lambda \in T^*_{q_0}M : \frac{1}{2}\sum_{i=1}^{k} \langle \lambda, f_i(q_0) \rangle^2 = r \right\}.$$

Notice that C_r is a smooth connected and non compact $n-1$ dimensional submanifold of the fiber $T^*_{q_0}M$, indeed diffeomorphic to the cylinder $S^{k-1} \times \mathbb{R}^{n-k}$ (here $k = \dim \mathcal{D}_{q_0} < n$ is the rank of the structure at the point q_0). By continuity of Φ, 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, namely $\Phi(S_r) = C_r$. This is a contradiction since, by continuity, $\Phi(S_r)$ is compact, while C_r is not. $\square$

Next we go back to the proof of the main result. Recall that $q_0 \in M$ is fixed and $\mathfrak{f}$ is the one half of the distance squared from q_0. After Proposition 11.6, it is natural to introduce the following definition.

Definition 11.9. Fix a point $q_0 \in M$. The set of *smooth point* from q_0 is the set $\Sigma \subset M$ of $q \in M$ such that there exists a unique lenght-minimizer γ joining q_0 to q, that it is strictly normal, and not conjugate.

From the proof of Proposition 11.6 (see also Remark 11.7) it follows that if the squared distance $\mathfrak{f}$ from q_0, is smooth at q then $q \in \Sigma$. The name smooth point of $\mathfrak{f}$ is justified by the following theorem.

Theorem 11.10. *The set Σ is open and dense in B. Moreover $\mathfrak{f}$ is smooth at every point of Σ.*

Proof. We divide the proof into three parts: (a) the set Σ is open, (b) the function $\mathfrak{f}$ is smooth in a neighborhood of every point of Σ, (c) the set Σ is dense in B.

(a). To prove that Σ is open we have to show that for every $q \in \Sigma$ there exists a neighborhood O_q of q such that every $q' \in O_q$ is also in Σ.

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 minimizer.

By contradiction, if this property is not true, there exists a sequence q_n of points in B converging to q such that (at least) two minimizers γ_n and γ'_n joining q_0 and q_n. Let us denote by u_n and v_n the corresponding minimizing controls.

By Proposition 8.65, the set of controls associated with minimizers whose endpoint is in the compact ball B is compact in L^2 (w.r.t. the strong topology). Then there exist, up to considering a subsequence, two controls u, v such that $u_n \to u$ and $v_n \to v$. Moreovers the limits u and v are both minimizers and join q_0 with q. Since by assumption there is a unique minimizer γ joining q_0 with q, it follows that $u = v$ is the corresponding control.

By smoothness of the end point map both $D_{u_n}F$ and $D_{v_n}F$ tends to D_uF, which has has full rank (u is strictly normal, hence is not a critical point for F). Hence, for n big enough, both $D_{u_n}F$ and $D_{v_n}F$ are surjective, i.e., u_n and v_n are strictly normal, and we can build the sequence λ_1^n and ξ_1^n of corresponding final covectors in $T^*_{q_n}M$ satisfying

$$\lambda_1^n D_{u_n}F = u_n, \qquad \xi_1^n D_{v_n}F = v_n.$$

These relations can be rewritten in terms of the adjoint linear maps

$$(D_{u_n}F)^* \lambda_1^n = u_n, \qquad (D_{v_n}F)^* \xi_1^n = v_n.$$

Since both $(D_{u_n}F)^*$ and $(D_{v_n}F)^*$ are a family of injective linear maps converging to $(D_uF)^*$ and $u_n, v_n \to u$, it follows that the corresponding (unique) solutions λ_1^n and ξ_1^n also converge to the solution of the limit problem $(D_uF)^*\lambda_1 = u$, i.e, both converge to the final covector λ_1 corresponding to γ. By using the flow defined by the corresponding controls we can deduce the convergence of the sequences λ_0^n and ξ_0^n of the initial covectors associated to u_n and v_n to the unique initial covector λ_0 corresponding to γ.

Finally, since λ_0 by assumption is a regular point of the exponential map, i.e., the unique minimizer γ joining q_0 to q is not conjugate, it follows that the exponential map is invertible in a neighborhood V_{λ_0} of λ_0 onto its image $O_q := \exp(V_{\lambda_0})$, that is a neighborhood of q. In particular this proves our initial claim.

More precisely 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 Σ is open.

(b). Now we prove that $\mathfrak{f}$ 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_{λ_0} of λ_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,

$$\mathfrak{f}(q') = \frac{1}{2}d^2(q_0, q') = H(\Phi(q')),$$

that is a composition of smooth functions, hence smooth.

(c). Our next goal is to show that Σ is a dense set in B. We start by a preliminary definition.

Definition 11.11. 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 Σ_f and Σ_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:

$(c1)$ Σ_f is a dense set in B,

$(c2)$ Σ_g is a dense set in B,

$(c3)$ $\mathfrak{f}$ is Lipschitz in a neighborhood of every point of Σ_g,

$(c4)$ Σ 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

$$\psi : O \to \mathbb{R}, \qquad \psi(q) = \mathfrak{f}(q) - a(q)$$

The function ψ is continuous on O and, since $\mathfrak{f}$ 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 ψ attains its minimum at some point $q_1 \in O$. Define a control u_1 associated with a minimizer γ joining q_0 and $F(u_1) = q_1$.

Since $J(u) \geq \mathfrak{f}(F(u))$ for every u, it is easy to see that the map

$$\Phi : \mathcal{U} \to \mathbb{R}, \qquad \Phi(u) = J(u) - a(F(u))$$

attains its minimum at u_1. In particular it holds

$$0 = D_{u_1}\Phi = u_1 - (d_{q_1}a)D_{u_1}F.$$

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 4.26, the corresponding trajectory γ is uniquely recovered by the formula $\gamma(t) = \pi \circ e^{(t-1)\vec{H}}(d_{q_1}a)$. In particular γ is the unique minimizer joining q_0 to $q_1 \in O$, and is normal, i.e. $q_1 \in \Sigma_f \cap O$.

Remark 11.12. In the Riemannian case $\Sigma_f = \Sigma_g$ since there are no abnormal extremal.

(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 $\operatorname{rank} q := \operatorname{rank} D_u F$, where u is the control associated to the unique minimizer γ 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., $D_{u'}F$ 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 exist 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 F it follows that $D_{u_n}F \to D_{\widehat{u}}F$ which, by semicontinuity of the rank, implies the contradiction

$$\operatorname{rank} \widehat{q} = \operatorname{rank} D_{\widehat{u}}F \leq \liminf_{n\to\infty} \operatorname{rank} D_{u_n}F \leq k_O - 1.$$

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

$$\Pi_q = e^{-\vec{H}}\{\xi \in T_q^*M \mid \xi D_u F = \lambda_1 D_u F\} \subset T_{q_0}^*M.$$

The set Π_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 11.13. *Π_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 $\widehat{\Pi}_q = \{\xi \in T_q^*M \mid \xi D_u F = \lambda_1 D_u F\}$ is an affine subspace of $T_{q_0}^*M$. Indeed $\xi \in \Pi_q$ if and only if $(D_uF)^*(\xi - \lambda_1) = 0$, that is

$$\widehat{\Pi}_q = \{\xi \in T_q^*M \mid \xi D_u F = \lambda_1 D_u F\} = \lambda_1 + \ker (D_uF)^*,$$

Moreover $\dim \ker (D_uF)^* = n - \dim \operatorname{im} D_uF = n - k_O$. 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 Π_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 $D_{u_n}F \to D_uF$. Since rank $D_{u_n}F$ is constant, it follows that $\ker (D_{u_n}F)^* \to \ker (D_uF)^*$, as subspaces.

$\square$

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 Π_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.

(c3) The proof of this claim relies on the following result, which is of independent interest.

Theorem 11.14. *Let $K \subset B$ a compact in our ball such that any minimizer connecting q_0 to $q \in K$ is strictly normal. Then $\mathfrak{f}$ is Lipschitz on K.*

Proof of Theorem 11.14. Let us first notice that, since K is compact, it is sufficient to show that $\mathfrak{f}$ 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 $D_u F$ 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 $F(\Phi(q')) = q'$ for all $q' \in O_q$ (see also Theorem 2.54).

Fix then local coordinates around q. Since Φ is smooth, there exists $R > 0$ and $C_0 > 0$ such that

$$B_q(C_0 r) \subset F(\mathcal{B}_u(r)), \qquad \forall\, 0 \leq r < R, \tag{11.7}$$

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

$$J(u') - J(u) \leq C_1 \|u' - u\|_2$$

Pick then any point $q' \in K$ such that $|q' - q| = C_0 r$, with $0 \leq r \leq R$. By (11.7), there exists $u' \in \mathcal{B}_u(R)$ with $\|u' - u\|_2 \leq r$ such that $F(u') = q'$. Using that $\mathfrak{f}(q') \leq J(u')$ and $\mathfrak{f}(q) = J(u)$, since u is a minimizer, we have

$$\mathfrak{f}(q') - \mathfrak{f}(q) \leq J(u') - J(u) \leq C_1 \|u' - u\|_2 \leq C' |q' - q|,$$

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

$$|\mathfrak{f}(q') - \mathfrak{f}(q)| \leq C' |q' - q|,$$

for every pair of points q, q' such that $|q' - q| \leq C_0 R$.

□

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 $\mathfrak{f}$ is Lipschitz on O'. It is then sufficient to show that $\Sigma \cap O' \neq \emptyset$.

By Proposition 11.6 (see also Remark 11.7) every differentiability point of $\mathfrak{f}$ is reached by a unique minimizer that is normal, hence is a fair point. Since we know that $\mathfrak{f}$ 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 Σ_f that is a regular value for the exponential map is in Σ, 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. □

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 11.15. *Assume that M is a complete sub-Riemannian structure and that there are no abnormal minimizers. Then* $\text{meas}(M \setminus \Sigma) = 0$.

This result is not known in general, and it is indeed a main open problem of sub-Riemannian geometry to establish whether Corollary 11.15 remains true in presence of abnormal minimizers.

We stress that the assumptions of the theorem are satisfied in the case of Riemannian structure. Indeed in this case, following the same arguments of the proof, we have the following result.

Proposition 11.16. *Let M be a sub-Riemannian structure that is Riemannian at q_0,i.e., such that* $\dim \mathcal{D}_{q_0} = \dim M$. *Then there exists a neighborhood O_{q_0} of q_0 such that $\mathfrak{f}$ is smooth on O_{q_0}.*

11.3 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 containing S. It is characterized as the set of $v \in V$ such that there exists a finite number of elements $v_0, \ldots, v_\ell \in S$ such that

$$v = \sum_{i=0}^{\ell} \lambda_i v_i, \qquad \lambda_i \geq 0, \quad \sum_{i=0}^{n} \lambda_i = 1.$$

If $\varphi : M \to \mathbb{R}$ is a function defined on a smooth manifold M, we say that φ is *locally Lipschitz* is φ 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 we 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 φ.

Definition 11.17. Let $\varphi : M \to \mathbb{R}$ be a locally Lipschitz function. The *(Clarke) generalized differential* of φ at the point $q \in M$ is the set

$$\partial_q\varphi := \text{conv}\{\xi \in T_q^*M \,|\, \xi = \lim_{q_n \to q} d_{q_n}\varphi\} \tag{11.8}$$

Notice that, by definition, $\partial_q\varphi$ is a subset of T_q^*M. It is closed by definition and bounded since the function is locally Lipschitz, hence compact.

Exercise 11.18. (i). Show that the mapping $q \mapsto \partial_q\varphi$ is *upper semicontinuous* in the following sense: if $q_n \to q$ in M and $\xi_n \to \xi$ in T^*M where $\xi_n \in \partial_{q_n}\varphi$, then $\xi \in \partial_q\varphi$.

(ii). We say that q is *regular* for φ if $0 \notin \partial_q\varphi$. Prove that the set of regular point for φ is open in M.

From the very definition of generalized differential we have the following result.

Lemma 11.19. *Let $\varphi : M \to \mathbb{R}$ be a locally Lipschitz function and $q \in M$. The following 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, i.e., for every sequence of differentiability point $q_n \to q$ we have $d_{q_n}\varphi \to d_q\varphi$.*

Remark 11.20. Let A be a subset of $\mathbb{R}^n$ of measure zero and consider the set of half-lines $L_v = \{q + tv, t \geq 0\}$ 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 for almost all $v \in S^{n-1}$, the function $t \mapsto \varphi(q + tv)$ is differentiable for a.e. $t \geq 0$.

Example 11.21. 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 φ 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 11.25 and Theorem 11.29).

The following is a classical hyperplane separation theorem for closed convex sets in $\mathbb{R}^n$.

Lemma 11.22. *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*

$$\langle x, v\rangle > \langle y, v\rangle + \varepsilon, \qquad \forall\, x \in K, \forall\, y \in C. \tag{11.9}$$

We also recall here another useful result from convex analysis.

Lemma 11.23 (Carathéodory)**.** *Let $S \subset \mathbb{R}^n$ and $x \in \operatorname{conv}(S)$. Then there exists $x_0, \dots, x_n \in S$ 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.

Proposition 11.24. *Let $\varphi : M \to \mathbb{R}$ be locally Lipschitz and q be a local minimum for φ. 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 φ. 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 11.22, 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 φ one has

$$\langle d_{q'}\varphi, v'\rangle \leq -\varepsilon/2, \qquad \forall\, v' \in V_v.$$

Fix $q' \in O_q$ where φ is differentiable and a vector $w \in V_v$ such that the set of differentiable points with the line $\{q + tw\}$ has full measure (cf. Remark 11.20). 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 φ cannot have a minimum at q. $\square$

The following proposition gives an estimate for the generalized differential of some special class of function.

Proposition 11.25. *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:*

$$(\omega, q) \mapsto \varphi_\omega(q), \qquad (\omega, q) \mapsto d_q\varphi_\omega$$

Then the function $a(q) := \min_{\omega\in\Omega} \varphi_\omega(q)$ is locally Lipschitz on M and

$$\partial_q a \subset \mathrm{conv}\{d_q\varphi_\omega \,|\, \forall\, \omega \in \Omega \text{ s.t. } \varphi_\omega(q) = a(q)\}. \tag{11.10}$$

Proof. As in the proof of Proposition 11.24 we can assume that $M = \mathbb{R}^n$. Notice that, if we denote by $\Omega_q = \{\omega \in \Omega, \varphi_\omega(q) = a(q)\}$ we have by compactness of Ω that Ω_q is non empy for every $q \in M$ and we can rewrite the claim as follows

$$\partial_q a \subset \mathrm{conv}\{d_q\varphi_\omega \,|\, \omega \in \Omega_q\}. \tag{11.11}$$

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 (11.11).

(i). Fix a compact $K \subset M$. Since every φ_ω is Lipschitz on K and Ω is compact, 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 ω 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 $D_q := \operatorname{conv}\{d_q\varphi_\omega | \forall\, \omega \in \Omega_q\}$. 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 11.22 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 Ω_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 ω 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

$$\langle d_q a, v\rangle \geq \langle \xi, v\rangle + \varepsilon/4 \tag{11.12}$$

If $d_q a \notin D_q$ we can choose $\xi = d_q a$ in the above reasoning and (11.12) 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 11.18) and the fact that Ω is compact. □

Exercise 11.26. Complete the second part of the proof of Proposition 11.25. Hint: use Carathéodory lemma.

11.3.1 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 $\mathbb{R}^n$ is locally Lipschitz.

For a locally Lipschitz map between manifolds $f : M \to N$ the *(Clarke) generalized differential* is defined as follows

$$\partial_q f := \operatorname{conv}\{L \in \operatorname{Hom}(T_qM, T_{f(q)}N) |\, L = \lim_{q_n\to q} D_{q_n} f,\ q_n \text{ diff. point of } f\},$$

The following lemma shows how the standard chain rule extends to the Lipschitz case.

Lemma 11.27. *Let M be a smooth manifold 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*

$$\partial_q(f \circ \phi) = \partial_{\varphi(q)} f \cdot D_q\phi. \tag{11.13}$$

(b) *If $\varphi : N \to W$ is a C^1 map, and $q \in M$ we have*

$$\partial_q(\varphi \circ f) = D_{f(q)}\varphi \cdot \partial_q f. \tag{11.14}$$

Moreover the generalized differential, as a set, is upper semicontinuous. *More precisely 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 **??**. Here we only give the main ideas.

(a). Since ϕ is a diffeomorphism, it sends every differentiability point q of $f \circ \phi$ to a differentiability point $\phi(q)$ for f. Then (11.13) is true at differentiability point and passing to the limit it is also valid for sub-differential (one proves both inclusions using ϕ and ϕ^{-1}). Part (b) can be proved along the same lines. The semicontinuity can be proved by using the hyperplane separation theorem and the Carathéodory Lemma. □

Definition 11.28. Let $f : M \to N$ be a locally Lipschitz map. A point $q \in M$ is said *critical* for f if $\partial_q f$ contains a non-surjective map. If $q \in M$ is not critical it is said *regular*.

Notice that by the semicountinuity property of Lemma 11.27, it follows that the set of regular point of a locally Lipschitz map f is open.

Theorem 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 exists neighborhood $O_{f(q)}$ and a locally Lipschitz map $g : O_{f(q)} \subset \mathbb{R}^n \to \mathbb{R}^n$ such that $f \circ g = g \circ f = \mathrm{Id}$.*

Remark 11.30. The classical C^1 version of the inverse function theorem (cf. Theorem **??**) can be proved from Theorem 11.29 and the chain rule (Lemma 11.27). Indeed Theorem 11.29 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 11.29 we need the following technical lemma.

Lemma 11.31. *Let $f : \mathbb{R}^n \to \mathbb{R}^n$ be a locally Lipschitz map and $q \in M$ be a regular point. Then there exists a neighborhood O_q of q and $\varepsilon > 0$ such that*

$$\forall\, v \in S^{n-1},\ \exists\, \xi_v \in S^{n-1} \qquad \text{s.t.} \qquad \langle \xi_v, \partial_x f(v)\rangle > \varepsilon, \quad \forall\, x \in O_q. \tag{11.15}$$

Moreover $|f(x) - f(y)| \geq \varepsilon|x - y|$, for all $x, y \in O_q$.

We stress that (11.15) 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 ξ_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 (11.15).

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 11.20), and set $y' = x + sv'$ and $\xi_{v'}$ the vector associated to v' defined by (11.15). 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 ε 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 11.29. The inequality proved in Lemma 11.31 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 11.31 holds also on the topological boundary ∂O_q. Lemma 11.31 also implies that

$$\operatorname{dist}(f(q), \partial f(O_q)) \geq \varepsilon \operatorname{dist}(q, \partial O_q) > 0,$$

where $\operatorname{dist}(x, A) = \inf_{y\in A} |x-y|$ denotes the Euclidean distance from x to the set A. Then consider 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

$$\psi : \overline{O_q} \to \mathbb{R}, \qquad \psi(x) = |f(x) - \bar{y}|^2$$

By construction $\psi(q) < \psi(z)$, for all $z \in \partial O_q$, hence by continuity ψ attains the minimum on some point $\bar{x} \in O_q$. By Proposition 11.24, we have $0 \in \partial_{\bar{x}}\psi$. Moreover, using the chain rule

$$\partial_{\bar{x}}\psi = (f(\bar{x}) - \bar{y})^T \cdot \partial_{\bar{x}} f$$

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 11.32. *Let $\varphi : M \to \mathbb{R}$ be locally Lipschitz and assume that $c \in \mathbb{R}$ is a regular value for φ. 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. Hence there exists, by Theorem 11.29 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. $\square$

11.3.2 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 11.33 (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 11.34 (Rifford [86])**.** *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 $a(q) = \min_{\omega \in \Omega} \varphi_\omega(q)$ *has measure zero in* $\mathbb{R}$.

Proof. We are going to define a countable set of smooth functions Φ_α indexed by $\alpha = (\alpha_0, \dots, \alpha_n) \in \mathbb{N}^{n+1}$, where $n = \dim M$, such that to every critical point q of a there corresponds a critical point z_q of some Φ_α. Moreover we have $\Phi_\alpha(z_q) = a(q)$.

Denote by $\Lambda_n = \{(\lambda_0, \ldots, \lambda_n) | \, \lambda_i \geq 0, \sum \lambda_i = 1\}$. For every $\alpha = (\alpha_0, \ldots, \alpha_n) \in \mathbb{N}^{n+1}$ let us consider the map

$$\Phi_\alpha : N_{\alpha_0} \times \ldots \times N_{\alpha_n} \times \Lambda_n \times M \to \mathbb{R}$$
$$\Phi_\alpha(\omega_0, \ldots, \omega_n, \lambda_0, \ldots, \lambda_n, q) = \sum_{i=0}^{n} \lambda_i \varphi_{\omega_i}(q). \tag{11.16}$$

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 Φ_α id and only if it satisfies the following relations:

$$\begin{cases} \sum_{i=0}^{n} \lambda_i \dfrac{\partial \psi_{\alpha_i}}{\partial \omega}(\omega_i, q) = 0, \qquad i = 0, \ldots, n, \\ \sum_{i=0}^{n} \lambda_i d_q \varphi_{\omega_i} = 0 \qquad i = 0, \ldots, n, \\ \varphi_{\omega_0}(q) = \ldots = \varphi_{\omega_n}(q) \end{cases} \tag{11.17}$$

Recall that ψ_i is simply the restriction of the map $(\omega, q) \mapsto \varphi_\omega(q)$ for $\omega \in N_i$.

Let us now show that every critical point q of a can be associated to a critical point z_q of some Φ_α. By Proposition 11.25, the function a is locally Lipschitz. Assume that q is a critical point of a, then we have

$$0 \in \partial_q a \subset \operatorname{conv}\{d_q \varphi_\omega | \, \forall \, \omega \in \Omega \; s.t. \; \varphi_\omega(q) = a(q)\}.$$

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 ψ_{α_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 (11.17) 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_α the set of critical point of Φ_α 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 all α by classical Sard lemma. $\square$

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 11.35. *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 $\text{meas}(a(C_a) \cap O_{q_n}) = 0$.

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) $N_q \subset \text{int}\, A_c$,

(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 11.33), 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 ψ fixed. We observe that (A2) is equivalent to require that level set of F are transversal to level of ψ. 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 Ω 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 Ω 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 11.34 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

$$\mathrm{meas}(a(C_a)) \leq \sum_{n\in\mathbb{N}} \mathrm{meas}(a(C_a) \cap O_{q_n}) = 0$$

since $\mathrm{meas}(a(C_a \cap O_q) = 0$ for every q.

□

Remark 11.36. 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.

11.4 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 11.37. *Fix $q_0 \in M$ and let $\mathcal{K} \subset T^*_{q_0}M \setminus (H^{-1}(0) \cap T^*_{q_0}M)$ be a compact set 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 the $\exp_{q_0}$ for all $0 < t \leq \varepsilon$.*

Proof. By Corollary **??** for every strongly normal extremal $\gamma(t) = \exp(t\lambda_0)$, with $\lambda_0 \in T^*_{q_0}M$, there exists $\varepsilon = \varepsilon(\lambda_0) > 0$ such that $\gamma|_{]0,\varepsilon]}$ does not contain points conjugate to q_0, or equivalently, $t\lambda_0$ is a regular point for the $\exp_{q_0}$ for all $0 < t \leq \varepsilon$. Since $\mathcal{K}$ is compact, it follows that there exists $\varepsilon = \varepsilon(\mathcal{K})$ such that the above property holds uniformly on $\mathcal{K}$. □

Lemma 11.38. *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*

$$C = \{\lambda_0 \in T^*_{q_0}M |\ \lambda_0 \text{ minimizer},\ \exp(\lambda_0) \in K\}.$$

Then $\overline{C}$ is compact.

Proof. It is 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 γ_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

$$\lambda_n D_{u_n}F = u_n, \qquad \forall\, n \in \mathbb{N}. \tag{11.18}$$

and dividing by $|\lambda_n|$ one obtain the identity

$$\frac{\lambda_n}{|\lambda_n|} D_{u_n}F = \frac{u_n}{|\lambda_n|}, \qquad \forall\, n \in \mathbb{N}. \tag{11.19}$$

Using compactness of minimizers whose endpoints stay in a compact region, 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 λ. Using that $D_{u_n}F \to D_uF$ and the fact that $|\lambda_n| \to +\infty$, passing to the limit for $n \to \infty$ in (11.19) we obtain $\lambda D_uF = 0$. This implies in particular that the minimizers γ_n converge to a minimizer γ (associated to λ) that is abnormal and reaches a point of K that is a contradiction. □

Theorem 11.39 (Rifford [87]). *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 $\bar{B}_{q_0}(r_0)$ 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

$$C = \{\lambda_0 \in T^*_{q_0}M \mid \lambda_0 \text{ minimizer}, \ \exp(\lambda_0) \in \overline{A}_\delta\}$$

By Lemma 11.38 the set $C_0 := \overline{C}$ is compact. Moreover define

$$C_1 := \{\lambda_0 \in C_0 \cap H^{-1}([0, \varepsilon_0])\},$$

for some $\varepsilon_0 > 0$ that is chosen later. Notice that C_1 is compact. For every $\lambda_0 \in T^*M$ let us consider the control u associated with $\gamma(t) = \exp(t\lambda_0)$ and denote by

$$\Phi_{\lambda_0} := (P^{-1}_{0,t})_* : T^*_{q_0}M \to T^*_{\exp_{q_0}(\lambda_0)}M,$$

the pullback of the flow defined by the control u, computed at q_0.

For a fixed $\lambda_0 \in C_0$, using that C_1 is compact, let us choose $\varepsilon = \varepsilon(\lambda_0)$ satisfying the following property: for every $\lambda_1 \in C_1$, the covector $\Phi_{\lambda_0}(\lambda_1) \in T^*_{\exp_{q_0}(\lambda_0)}M$, is a regular point of $\exp_{\exp_{q_0}(\lambda_0)}$. Being C_0 also compact, we can define $\varepsilon_0 = \min\{\varepsilon(\lambda_0), \lambda_0 \in C_0\}$. Define the map

$$\Psi : C_0 \times C_1 \to D_\delta \subset M, \qquad \Psi(\lambda_0, \lambda_1) = \exp_{\exp_{q_0}(\lambda_0)}(\Phi_{\lambda_0}(\lambda_1)).$$

By construction Ψ is a submersion. We want to apply Theorem 11.35 to the submersion Ψ and the scalar function

$$\mathcal{H} : C_0 \times C_1 \to \mathbb{R}, \qquad \mathcal{H}(\lambda_0, \lambda_1) = H(\lambda_0) + H(\lambda_1).$$

Let us show that the assumption of Theorem 11.35 are satisfied. Indeed we have to show that the set

$$N_q = \{(\lambda_0, \lambda_1) \in C_0 \times C_1 \,|\, \mathcal{H}(\lambda_0, \lambda_1) = \min_{\Psi(\lambda_0,\lambda_1)=q} \mathcal{H}(\lambda_0, \lambda_1)\}, \qquad \forall\, q \in \overline{A}_\delta,$$

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 C_0, for each $q \in \overline{A}_\delta$ there exists $\bar{\lambda}_0 \in C_0$ 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

$$\min_{\Psi(\lambda_0,\lambda_1)=q} \mathcal{H}(\lambda_0, \lambda_1) = \min_{\exp_{q_0}(\lambda_0)=q} H(\lambda_0) = \mathfrak{f}(q), \qquad \forall\, q \in \overline{A}_\delta.$$

This implies that N_q is non empty for every q. Moreover one can show that N_q is compact. By applying Theorem 11.35 one gets that the function

$$a(q) = \min_{\Psi(\lambda_0,\lambda_1)=q} \mathcal{H}(\lambda_0, \lambda_1) = \mathfrak{f}(q),$$

is locally Lipschitz in $\overline{A}_\delta$ and the set of its critical values has measure zero in $\overline{A}_\delta$. Since $\delta > 0$ is arbitrary we let $\delta \to 0$ and we have that $\mathfrak{f}$ 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 $\mathfrak{f}$. Then, applying Corollary 11.32, the sphere $\mathfrak{f}^{-1}(r^2/2)$ is a Lipschitz submanifold for almost every $r \leq r_0$. $\square$

11.5 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 11.40 (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. For the fixed $x \in M$, let us consider

$$A = \{r > 0 \,|\, B(x,r) \text{ is compact }\}, \qquad R := \sup A.$$

As in the proof of Theorem 3.44, 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 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 have to show that $\bar{B}(x,R)$ is compact, i.e., for every sequence y_i in $\bar{B}(x,R)$ we can extract a convergent subsequence. Define $r_i := \mathsf{d}(y_i, x)$. It is not restrictive to assume that $r_i \to R$ (if it is not the case, the sequence stays in a compact ball and the existence of a convergent subsequence is clear). Since the ball $\bar{B}(x, r_i)$ is compact, by Theorem 3.40 there exists a length minimizing trajectory $\gamma_i : [0, r_i] \to M$ joining x and y_i, parametrized by unit speed.

Due to the completeness of the vector field $\vec{H}$, we can extend each curve γ_i, parametrized by length, to the common interval $[0, R]$. By construction this sequence of trajectory is normal

$$\gamma_i(t) = \exp(t\lambda_i) = \pi \circ e^{t\vec{H}}(\lambda_i),$$

for some $\lambda_i \in T_xM$, and is contained in the compact set $\bar{B}(x,R)$. Since there is no abnormal minimizer, by Lemma 11.38 the sequence $\{\lambda_i\}$ is bounded in T_x^*M, thus there exists a subsequence λ_{i_n} converging to λ. Then $r_{i_n}\lambda_{i_n} \to R\lambda$ and by continuity of exp we have that $\{y_i\}$ has a convergent subsequence

$$y_{i_n} = \gamma_{i_n}(r_{i_n}) = \exp(r_{i_n}\lambda_{i_n}) \to \exp(R\lambda) =: y$$

To end the proof, one should just notice that an arbitrary Cauchy sequence in M is bounded, hence contained in a suitable ball centered at x, which is compact since $R = +\infty$. Thus it admits a convergent subsequence. $\square$

As an immediate corollary we have the following version of geodesic completeness theorem.

Corollary 11.41. *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.*

11.6 Equivalence of sub-Riemannian distances*

Chapter 12

Abnormal extremals and second variation

In this chapter we are going to discuss in more details abnormal extremals and how the regularity of the sub-Riemannian distance is affected by the presence of these extremals.

12.1 Second variation

We want to introduce the notion of Hessian (and second derivative) for smooth maps between manifolds. We first discuss the case of the second differential of a map between linear spaces.

Let $F : V \to M$ be a smooth map from a linear space V on a smooth manifold M. As we know, the first differential of F at a point $x \in V$

$$D_xF : V \to T_{F(x)}M, \qquad D_xF(v) = \frac{d}{dt}\bigg|_{t=0} F(x+tv), \quad v \in V,$$

and is a well defined linear map independent on the linear structure on V. This is not the case for the second differential. Indeed it is easy to see that the second order derivative

$$D_x^2F(v) = \frac{d^2}{dt^2}\bigg|_{t=0} F(x+tv) \tag{12.1}$$

has not invariant meaning if $D_xF(v) \neq 0$. Indeed in this case the curve $\gamma : t \mapsto F(x+tv)$ is a smooth curve in M with nonzero tangent vector. Then there exists some local coordinates on M such that the curve γ is a straight line. Hence the second derivative $D_x^2F(v)$ vanish in these coordinates.

In general, the linear structure on V let us to define the second differential of F as a quadratic map

$$D_x^2F : \ker D_xF \to T_{F(x)}M \tag{12.2}$$

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 a linear map between the tangent spaces

$$D_xF : T_xN \to T_{F(x)}M, \qquad x \in N.$$

and the definition of second order derivative should be modified using smooth curves with fixed tangent vector (that belong to the kernel of D_xF):

$$D_x^2F(v) = \left.\frac{d^2}{dt^2}\right|_{t=0} F(\gamma(t)), \qquad \gamma(0) = x, \quad \dot\gamma(0) = v \in \ker D_xF, \tag{12.3}$$

Computing in coordinates we find that

$$\left.\frac{d^2}{dt^2}\right|_{t=0} F(\gamma(t)) = \frac{d^2F}{dx^2}(\dot\gamma(0), \dot\gamma(0)) + \frac{dF}{dx}\ddot\gamma(0) \tag{12.4}$$

that shows that term (12.4) is defined only up to $\operatorname{im} D_xF$.

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

In the previuos chapter we proved that if we have abnormal minimizer that reach some point q, then the sub-Riemannian distance is not smooth at q. If we also have that no normal minimizers reach q we can say that it is not even Lipschitz.

Proposition 12.1. *Assume that there are no normal minimizers that join q_0 to $\widehat{q}$. Then $\mathfrak{f}$ is not Lipschitz in a neighborhood of $\widehat{q}$. Moreover*

$$\lim_{\substack{q\to\widehat{q}\\ q\in\Sigma}} |d_q\mathfrak{f}| = +\infty. \tag{12.5}$$

In the previous theorem $|\cdot|$ is an arbitrary norm of the fibers of T^*M.

Proof. Consider a sequence of smooth points $q_n \in \Sigma$ such that $q_n \to \widehat{q}$. Since q_n are smooth we know that there exists unique controls u_n and covectors λ_n such that

$$\lambda_n D_{u_n}F = u_n, \qquad \lambda_n = d_{q_n}\mathfrak{f}.$$

Assume by contradiction that $|d_{q_n}\mathfrak{f}| \leq M$ then, using compactness we find that $u_n \to u$, $\lambda_n \to \lambda$ with $\lambda D_uF = u$, that means that the associate geodesic reach $\widehat{q}$. In other words, there exists a normal minimizer that goes at $\widehat{q}$, that is a contradiction. □

Let us now consider the end-point map $F : \mathcal{U} \to M$. As we explained in the previous section, its Hessian at a point $u \in \mathcal{U}$ is the quadratic vector function

$$\operatorname{Hess}_u F : \ker D_uF \to \operatorname{Coker} D_uF = T_{F(u)}M/\operatorname{im} D_uF.$$

Remark 12.2. Recall that $\lambda\, D_uF = 0$ if and only if $\lambda \in (\operatorname{im} D_uF)^\perp$. In other words, for every abnormal extremal there is a well defined scalar quadratic form

$$\lambda \operatorname{Hess}_u F : \ker D_uF \to \mathbb{R}$$

Notice that the dimension of the space $\operatorname{im} D_uF^\perp$ of such covectors coincide with $\dim \operatorname{Coker} D_uF$.

Definition 12.3. Let $Q : V \to \mathbb{R}$ be a quadratic form defined on a vector space V. The *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 this number coincide with the number of negative eigenvalues in the diagonal form of Q.

The following notion of *index* of the map F will be also useful:

Definition 12.4. Let $F : \mathcal{U} \to M$ and $u \in \mathcal{U}$ be a critical point for F. The *index* of F at u is

$$\mathrm{Ind}_u F = \min_{\substack{\lambda \in \mathrm{im}\, D_u F^{\perp} \\ \lambda \neq 0}} \mathrm{ind}^-(\lambda\, \mathrm{Hess}_u F) - \mathrm{codim}\, \mathrm{im}\, D_u F$$

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 $\mathrm{Ind}_u F = \mathrm{ind}^-(\lambda\, \mathrm{Hess}_u F) - 1$.

Theorem 12.6. *If* $\mathrm{Ind}_u F \geq 1$, *then u is not a strictly abnormal minimizer.*

We state without proof the following result (see Lemma 20.8 of [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 $Q(x) = 0$.*

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 Φ is *r-solid* at a point $x \in E$ if there exists 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 Φ, then Φ is 1-solid at x.
(Hint: Use implicit function theorem to prove that Φ satisfies (12.7) and Brower theorem to show that the same holds for some small perturbation)

Proposition 12.10. *Assume that* $\mathrm{Ind}_x \Phi \geq 0$. *Then Φ 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 Φ 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 property (12.8) is invariant by multiplication of the covector by a positive scalar we are reduced to the sphere

$$\lambda \in S^{k-1} = \{\lambda \in \operatorname{im} D_0\Phi^{\perp}, |\lambda| = 1\}.$$

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

$$\lambda \operatorname{Hess}_u \Phi\big|_{E_\lambda \setminus \{0\}} < 0$$

By the continuity of the form with respect to λ, there exists a neighborhood O_λ of λ 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

$$S^{k-1} = O_{\lambda_1} \cup \ldots \cup O_{\lambda_N}$$

Then it is sufficient to consider the finitedimensional 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

$$\operatorname{Hess}_0\Phi : E_2 \to \mathbb{R}^n / D_0\Phi(E_1)$$

According to Lemma 12.7 there exists $e_2 \in E_2$, regular point of $\operatorname{Hess}_0\Phi$, such that

$$\operatorname{Hess}_0\Phi(e_2) = 0 \qquad \Longrightarrow \qquad D_0^2\Phi(e_2) = D_0\Phi(e_1), \quad \text{for some } e_1 \in E_1.$$

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 the vector $e := -e_1/2 + e_2$. 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 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 ε

$$\Phi_\varepsilon(v_1 + v_2) = Q(v_1 + v_2) + O(\varepsilon) \tag{12.10}$$

hence for small ε the image of Φ_ε contain a ball around 0 from which it follows that

$$B_{\phi(0)}(c\varepsilon^2) \subset \Phi(B_0(\varepsilon)) \tag{12.11}$$

Moreover as soon as ε is fixed we can perturb the map Φ and still the estimate (12.11) holds. $\square$

Actually we proved the following statement, that is stronger than 2-solideness of Φ:

Lemma 12.11. *Under the assumptions of the Theorem 12.10, there exists $C > 0$ such that for every ε 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 Φ vanish, the ball of radius ε is mapped into a ball of radius ε^2, while the restriction on the other subspace "preserves" the order, as the estimates (12.9) and (12.10) show. [1]

Proof of Theorem 12.6. We prove that if $\mathrm{Ind}_u F \geq 1$, where u is a strictly abnormal geodesic, then u cannot be a minimizer. It is sufficient to show that the "extended" endpoint map

$$\Phi : \mathcal{U} \to \mathbb{R} \times M, \qquad \Phi(u) = \begin{pmatrix} J(u) \\ F(u) \end{pmatrix},$$

is locally open at u. Recall that $d_u J = \lambda\, D_u F$, for some $\lambda \in T_{F(u)} M$, if and only if $d_u J\big|_{\ker D_u F} = 0$ (see also Proposition 8.12). Since u is strictly abnormal, it follows that

$$d_u J\big|_{\ker D_u F} \neq 0. \tag{12.13}$$

Moreover from the definition of Φ and (12.13) one has

$$\ker D_u \Phi = \ker d_u J \cap \ker D_u F, \qquad \dim \mathrm{im}\, d_u J = 1.$$

Moreover, a covector $\bar\lambda = (\alpha, \lambda)$ in $\mathbb{R} \times T^*_{F(u)} M$ annihilates the image of $D_u \Phi$ if and only if $\alpha = 0$ and $\lambda \in \mathrm{im}\, D_u F^{\perp}$, indeed if

$$0 = \bar\lambda D_u \Phi = \alpha d_u J + \lambda D_u F$$

with $\alpha \neq 0$, this would imply that u is also normal. In other words we proved the equality

$$\mathrm{im}\, D_u \Phi^{\perp} = \{(0, \lambda) \in \mathbb{R} \times T^*_{F(u)} M \mid \lambda \in \mathrm{im}\, D_u F^{\perp}\} \tag{12.14}$$

Combining (12.13) and (12.14) one obtains for every $\bar\lambda = (0, \lambda) \in \mathrm{im}\, D_u \Phi^{\perp}$

$$\bar\lambda\, \mathrm{Hess}_u \Phi = \lambda\, \mathrm{Hess}_u F\big|_{\ker d_u J \cap \ker D_u F} \tag{12.15}$$

Moreover $\mathrm{codim}\, \mathrm{im}\, D_u \Phi = \mathrm{codim}\, \mathrm{im}\, D_u F$ since $\dim \mathrm{im}\, D_u \Phi = \dim \mathrm{im}\, D_u F + 1$ by (12.13) and $D_u \Phi$ takes values in $\mathbb{R} \times T_{F(u)} M$. Then for every $\bar\lambda = (0, \lambda) \in \mathrm{im}\, D_u \Phi^{\perp}$

$$\begin{aligned} \mathrm{ind}^-(\bar\lambda\, \mathrm{Hess}_u \Phi) - \mathrm{codim}\, \mathrm{im}\, D_u \Phi &= \mathrm{ind}^-(\lambda\, \mathrm{Hess}_u F\big|_{\ker d_u J \cap \ker D_u F}) - \mathrm{codim}\, \mathrm{im}\, D_u F \\ &\geq \mathrm{ind}^-(\lambda\, \mathrm{Hess}_u F) - 1 - \mathrm{codim}\, \mathrm{im}\, D_u F \end{aligned}$$

and passing to the infimum with respect to $\bar\lambda$ we get

$$\mathrm{Ind}_u \Phi \geq \mathrm{Ind}_u F - 1 \geq 0.$$

By Proposition 12.10 this implies that Φ is locally open at u. Hence u cannot be a minimizer. $\square$

[1] $B_0(c) \subset \Phi_\varepsilon(B(1)) \Leftrightarrow B_0(c\varepsilon^2) \subset \Phi(\varepsilon^2 v_1 + \varepsilon v_2), v_i \in B^i(1) \Leftrightarrow B_0(c\varepsilon^2) \subset \Phi(B'_{\varepsilon^2} \times B''_{\varepsilon})$

Now we prove that, under the same assumptions on the index of the endpoint map given in Theorem 12.6, the sub-Riemannian is Lipschitz even if some abnormal minimizers are present.

Theorem 12.12. *Let $K \subset B_{q_0}(r_0)$ be a compact and assume that* $\mathrm{Ind}_u F \geq 1$ *for every abnormal minimizer u such that $F(u) \in K$. Then $\mathfrak{f}$ is Lipschitz on K.*

Proof. Recall that if there are no abnormal minimizers reaching K, Theorem 11.39 ensures that $\mathfrak{f}$ is Lipschitz on K. Then, using compactness of the set of all minimizers, it is sufficient to prove the estimate in neighborhood of a point $q = F(u)$, where u is abnormal.

Since $\mathrm{Ind}_u F \geq 1$ by assumption, Theorem 12.6 implies that every abnormal minimizer u is not strictly abnormal, i.e., has also a normal lift. We have

$$\mathrm{Hess}_u F : \ker D_u F \to \mathrm{Coker}\, D_u F, \qquad \text{with} \qquad \mathrm{Ind}_u F \geq 1.$$

and, since u is also normal, it follows that $d_u J = \lambda D_u F$ for some $\lambda \in T^*_{F(u)} M$, hence $\ker D_u F \subset \ker d_u J$. The assumption of Lemma 12.11 are satisfied, hence splitting the the space of controls

$$L^2_k([0,1]) = E_1 \oplus E_2, \qquad E_2 := \ker D_u F$$

we have that there exists $C_0 > 0$ and $R > 0$ such that for $0 \leq \varepsilon < R$ we have

$$B_q(C_0 \varepsilon^2) \subset F(\mathcal{B}_\varepsilon), \qquad \mathcal{B}_\varepsilon := \mathcal{B}'_u(\varepsilon^2) \times \mathcal{B}''_u(\varepsilon), \qquad q = F(u), \tag{12.16}$$

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 $F(u') = q'$. Using that $\mathfrak{f}(q') \leq J(u')$ and $\mathfrak{f}(q) = J(u)$, since u is a minimizer, we have

$$\begin{aligned} \mathfrak{f}(q') - \mathfrak{f}(q) &\leq J(u') - J(u) \leq C_1 \|u'_1 - u_1\| + C_2 \|u'_2 - u_2\|^2 && (12.17) \\ &\leq C \varepsilon^2 = C' |q' - q| && (12.18) \end{aligned}$$

where we can choose $C = \max\{C_1, C_2\}$ 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

$$|\mathfrak{f}(q') - \mathfrak{f}(q)| \leq C' |q' - q|,$$

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 u be an abnormal minimizer and let $\lambda_1 \in T^*_{F(u)}M$ satisfy $\lambda_1 D_u F = 0$. Assume that* $\mathrm{ind}^- \lambda_1 \mathrm{Hess}_u F < +\infty$*. Then the following condition are satisfied :*

(i) $\langle \lambda(t), [f_i, f_j](\gamma(t)) \rangle \equiv 0,$ *for a.e. t, $\forall\, i, j = 1, \dots, k$,* *(Goh condition)*

(ii) $\big\langle \lambda(t), [[f_{u(t)}, f_v], f_v](\gamma(t)) \big\rangle \geq 0,$ *for a.e. t, $\forall\, v \in \mathbb{R}^k$,* *(Generalized Legendre condition)*

where $\lambda(t)$ and $\gamma(t) = \pi(\lambda(t))$ are respectively the extremal and the trajectory associated to λ_1.

Remark 12.14. Notice that, in the statement of the previous theorem, if λ_1 satisfies the assumption $\lambda_1 D_u F = 0$, then also $-\lambda_1$ satisfies the same assumptions. Since $\mathrm{ind}^-(-\lambda_1 \mathrm{Hess}_u F) = \mathrm{ind}^+ \lambda_1 \mathrm{Hess}_u F$ this implies that the statement holds under the assumption $\mathrm{ind}^+ \lambda_1 \mathrm{Hess}_u F < +\infty$. Indeed the proof shows that as soon as the Goh condition is not satisfied, both the positive and the negative index of this form are infinity.

Notice that these condition 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 abnormal if and only if it satisfies

$$\dot{\lambda}(t) = \sum_{i=1}^{k} u_i(t) \vec{h}_i(\lambda(t)), \qquad \langle \lambda(t), f_i(\gamma(t)) \rangle = 0, \ \forall\, i = 1, \dots, k,$$

i.e. $\lambda(t)$ satifies the Hamiltonian equation and belongs to $\mathcal{D}^{\perp}_{\gamma(t)}$. Goh condition are equivalent to require that $\lambda(t) \in (\mathcal{D}^2_{\gamma(t)})^{\perp}$.

Corollary 12.15. *Assume that the sub-Riemannian structure is 2-generating, i.e. $\mathcal{D}^2_q = T_q M$ for all $q \in M$. Then there are no strictly abnormal minimizers. In particular $\mathfrak{f}$ 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 $\mathrm{Ind}_u F = +\infty$, for any abnormal minimizer u. In particular, from Theorem 12.6 it follows that the minimizer cannot be strictly abnormal Hence $\mathfrak{f}$ is globally Lipschitz by Theorem 12.12. □

Remark 12.16. Notice that $\mathfrak{f}$ 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, then from Ball-Box Theorem (Theorem 10.62) it follows that the squared distance $\mathfrak{f}$ is not Lipschitz at the base point q_0.

On the other hand, on the set where $\mathfrak{f}$ is positive, we have that $\mathfrak{f}$ is Lipschitz if and only if the sub-Riemannian distance $d(q_0, \cdot)$ is.

Before going into the proof of the Goh conditions (Theorem 12.13) we discuss an important corollary.

Theorem 12.17. *Assume that $\mathcal{D}_{q_0} \neq \mathcal{D}^2_{q_0}$. Then for every $\varepsilon > 0$ there exists a normal extremal path γ starting from q_0 such that $\ell(\gamma) = \varepsilon$ and γ is not a length-minimizer.*

Before the proof, this is the idea: fix an element $\xi \in \mathcal{D}_{q_0}^{\perp} \setminus (\mathcal{D}_{q_0}^2)^{\perp}$ which is non empty by assumptions. We want to build an abnormal minimizing trajectory that has ξ as initial covector and that is the limit of a sequence of stricly normal lenth-minimizers. In this way this abnormal will have finite index (the abnormal quadratic form will be the limit of positive ones) and then by Goh condition $\xi \cdot \mathcal{D}_{q_0}^2 = 0$, which is a contradiction.

Proof. Assume by contradiction that there exists $T > 0$ such that all normal extremal paths γ_λ associated with initial covector $\lambda \in H^{-1}(1/2) \cap T_{q_0}^*M$ minimize on the segment $[0, T]$. Since restriction of length-minimizers are still length-minimizers, by suitably reducing $T > 0$, we can assume, thanks to Lemma 3.34, that there exists[2] a compact set K such that $\{\gamma_\lambda(T) \mid \lambda \in H^{-1}(1/2)\} \subset K$.

Fix an element $\xi \in \mathcal{D}_{q_0}^{\perp} \setminus (\mathcal{D}_{q_0}^2)^{\perp}$, which is non empty by assumptions. Then consider, given any $\lambda_0 \in H^{-1}(1/2) \cap T_{q_0}^*M$, 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 γ_s, and defined on $[0, T]$. Due to Theorem 11.4, 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 γ_{s_n} is a strictly normal length-minimizer for every $n \in \mathbb{N}$.

Denote $\Phi_t^n = P_{0,t}^{u_{s_n}}$ the non autonomous flow generated by the control u_{s_n}. The family $\lambda_{s_n}(t)$ satisfies

$$\lambda_{s_n}(t) = e^{t\vec{H}}(\lambda_0 + s_n\xi) = (\Phi_t^n)^*(\lambda_0 + s_n\xi).$$

Moreover, by continuity of the flow with respect to convergence of controls, we have that $\Phi_t^n \to \Phi_t$ for $n \to \infty$, where Φ_t denotes the flow associated with the control $\bar{u}$. Hence we have that the rescaled family

$$\frac{1}{s_n}\lambda_{s_n}(t) = (\Phi_t^n)^* \left(\frac{1}{s_n}\lambda_0 + \xi\right)$$

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 to the minimizing control $\bar{u}$, and with initial covector ξ.

The fact that u_{s_n} is a strictly normal minimizer says that the Hessian of the energy J restricted to the level set $F^{-1}(q_n)$ is non negative. Recall that

$$\mathrm{Hess}_u J|_{F^{-1}(q)} = I - \lambda_1 D_u^2 F,$$

where $\lambda_1 \in T_{F(u)}M$ 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

$$\|v\|^2 - \lambda_{s_n}(T) D_{u_{s_n}}^2 F(v, v) \geq 0.$$

This immediately implies

$$\frac{1}{s_n}\|v\|^2 - \frac{1}{s_n}\lambda_{s_n}(T) D_{u_{s_n}}^2 F(v, v) \geq 0,$$

[2]indeed it is enough to fix an arbitrary compact K with $q_0 \in \mathrm{int}(K)$ such that the corresponding δ_K defined by Lemma 3.34 is smaller than T.

and passing to the limit for $n \to \infty$ one gets

$$-\bar{\lambda}(T)D^2_{\bar{u}}F(v,v) \geq 0.$$

In particular one has that

$$\mathrm{ind}^+\bar{\lambda}(T)\mathrm{Hess}_{\bar{u}}F = \mathrm{ind}^-(-\bar{\lambda}(T)D^2_{\bar{u}}F) = 0.$$

Hence the abnormal extremal has finite (positive) index and we can apply Goh conditions (see Theorem 12.13 and Remark 12.14). Thus ξ is orthogonal to $\mathcal{D}^2_{q_0}$, which is a contradiction since $\xi \in \mathcal{D}^{\perp}_{q_0} \setminus (\mathcal{D}^2_{q_0})^{\perp}$. $\square$

Remark 12.18 (About the assumptions 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 $\mathcal{D}_{q_0} \neq \mathcal{D}^2_{q_0}$ for every $q \in D$.

Indeed assume that $\mathcal{D}_q \neq \mathcal{D}^2_q$ for all q in an open set A, then it is easy to see that $\mathcal{D}^i_q = \mathcal{D}_q \neq T_qM$ for all $q \in A$, since the structure is not Riemannian. Hence the structure is not bracket-generating in 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 $P_t = \overrightarrow{\exp}\int_0^t f_{u(s)}ds$ the nonautonomous flow generated by u. Following the argument used in the proof of Proposition 8.4 we can write the end-point map as the composition

$$E(u+v) = P_1(G(v)), \qquad D_uE = P_{1*}D_0G,$$

and reduced the problem to the expansion of G, which is easier. Indeed denoting $g^t_i := P^{-1}_{t*}f_i$, the map G can be interpreted as the end-point map for the system

$$\dot{q}(t) = g^t_{v(t)}(q(t)) = \sum_{i=1}^k v_i(t)g^t_i(q(t))$$

and the Hessian of F can be computed easily starting from the Hessian of G at $v = 0$

$$\mathrm{Hess}_uF = P_{1*}\mathrm{Hess}_0G$$

from which we get, using that $\lambda_0 = P_1^*\lambda_1$,

$$\lambda_1\mathrm{Hess}_uF = \lambda_1P_{1*}\mathrm{Hess}_0G = \lambda_0\mathrm{Hess}_0G$$

Moreover computing

$$\begin{aligned}\langle\lambda(t), [f_i, f_j](\gamma(t))\rangle &= \left\langle\lambda_0, P^{-1}_{t*}[f_i, f_j](\gamma(t))\right\rangle \\ &= \left\langle\lambda_0, [g^t_i, g^t_j](\gamma(0))\right\rangle\end{aligned}$$

the Goh and generalized Legendre conditions can also be rewritten as

$$\left\langle\lambda_0, [g^t_i, g^t_j]\gamma(0)\right\rangle \equiv 0, \qquad \text{for a.e. } t \in [0,1], \quad \forall\, i,j = 1,\dots,k, \tag{G.1}$$

$$\langle\lambda_0, [[g^t_{u(t)}, g^t_i], g^t_i]](\gamma(0))\rangle \geq 0, \qquad \text{for a.e. } t \in [0,1], \quad \forall\, i = 1,\dots,k. \tag{L.1}$$

Now we want to compute the Hessian of the map G. Using the Volterra expansion computed in Chapter 6 we have

$$G(v(\cdot)) \simeq q_0 \circ \left(\mathrm{Id} + \int_0^1 g^t_{v(t)} dt + \iint\limits_{0\le\tau\le t\le 1} g^\tau_{v(\tau)} \circ g^t_{v(t)} d\tau dt \right) + O(\|v\|^3)$$

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

$$D_0^2 G(v) = 2\, q_0 \circ \iint\limits_{0\le\tau\le t\le 1} g^\tau_{v(\tau)} \circ g^t_{v(t)} d\tau dt.$$

has not an immediate geometrical interpretation. Recall that the second differential $D_0^2 G$ is defined on the set

$$\ker D_0 G = \{v \in L^2_k[0,1],\ \int_0^1 g^t_{v(t)}(q_0) dt = 0\} \tag{12.19}$$

and, for such a v, $D_0^2 G(v)$ belong to the tangent space $T_{q_0}M$. Indeed, using Lemma 8.28, and that v belong to the set (12.19), we can symmetrize the second derivative, getting the formula

$$D_0^2 G(v) = \iint\limits_{0\le\tau\le t\le 1} [g^\tau_{v(\tau)}, g^t_{v(t)}](q_0) d\tau dt,$$

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

$$\left\langle \lambda_0, g^t_v(q_0) \right\rangle = 0, \qquad \text{for a.e. } t \in [0,1], \forall\, v \in \mathbb{R}^k.$$

Then we can compute the Hessian

$$\lambda_0 \mathrm{Hess}_0 G(v) = \iint\limits_{0\le\tau\le t\le 1} \langle \lambda_0, [g^\tau_{v(\tau)}, g^t_{v(t)}](q_0) \rangle d\tau dt \tag{12.20}$$

Remark 12.19. Denoting by K the bilinear form

$$K(\tau, t)(v, w) = \left\langle \lambda_0, [g^\tau_v, g^t_w](q_0) \right\rangle,$$

the Goh and generalized Legendre conditions are rewritten as follows

$$K(t,t)(v,w) = 0, \qquad \forall\, v, w \in \mathbb{R}^k, \quad \text{for a.e. } t \in [0,1], \tag{$G.2$}$$

$$\left.\frac{\partial K}{\partial \tau}(\tau, t)\right|_{\tau=t} (v,v) \ge 0, \qquad \forall\, v \in \mathbb{R}^k, \quad \text{for a.e. } t \in [0,1]. \tag{$L.2$}$$

Indeed, the first one easily follows from $(G.1)$. Moreover recall that $g_v^t = P_{t*}^{-1} f_v$, hence the map $t \mapsto g_v^t$ is Lipschitz for every fixed v. By definition of $P_t = \overrightarrow{\exp} \int_0^t f_{u(t)} dt$ it follows that

$$\frac{\partial}{\partial t} g_v^t = [g_{u(t)}^t, g_v^t]$$

which shows that $(L.2)$ is equivalent to $(L.1)$.

Finally we want to express the Hessian of G in Hamiltonian terms. To this end, consider the family of functions on T^*M which are linear on fibers, associated to the vector fields g_v^t:

$$h_v^t(\lambda) := \left\langle \lambda, g_v^t(q) \right\rangle, \qquad \lambda \in T^*M, \quad q = \pi(\lambda).$$

and 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

$$\sigma_{\lambda_0}(\eta_v^t, \eta_w^t) = \left\langle \lambda_0, [g_v^t, g_w^t](q_0) \right\rangle$$

and we get the final expression for the Hessian

$$\lambda_0 \mathrm{Hess}_0 G(v(\cdot)) = \iint\limits_{0 \le \tau \le t \le 1} \sigma_{\lambda_0}(\eta_{v(\tau)}^\tau, \eta_{v(t)}^t) dt d\tau. \tag{12.22}$$

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 "Hamiltonian" version of Goh and Legendre conditions is expressed as follows:

$$\sigma_{\lambda_0}(\eta_v^t, \eta_w^t) = 0, \qquad \forall\, v, w \in \mathbb{R}^k, \text{ for a.e. } t \in [0,1], \tag{G.3}$$

$$\sigma_{\lambda_0}(\dot{\eta}_v^t, \eta_v^t) \geq 0, \qquad \forall\, v \in \mathbb{R}^k, \text{ for a.e. } t \in [0,1]. \tag{L.3}$$

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 $\ker D_0 G$ 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 $v : [0,1] \to \mathbb{R}^k$ with compact support in $[0,1]$ and build, for $s > 0$ small enough, the control

$$v_s(\tau) = v\left(\frac{\tau - t}{s}\right), \qquad \operatorname{supp} v_s \subset [t, t+s]. \tag{12.23}$$

The idea is to apply the Hessian to this particular control functions and then compute the asymptotics for $s \to 0$.

indice finito allora e finito anche qui sopra.

Actually, since 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, it is not restrictive to restrict also to the subspace of zero average controls

$$E_s := \{v_s \in \ker D_0 G,\ v_s \text{ defined by (12.23)}, \int_0^1 v(\tau)d\tau = 0\}.$$

Notice that this space depend on the choice of t, while $\operatorname{codim} E_s$ does not depend on s.

Remark 12.20. We will use the following identity (writing σ for σ_{λ_0}), which holds for arbitrary control functions $v, w : [0,1] \to \mathbb{R}^k$

$$\iint_{\alpha \le \tau \le t \le \beta} \sigma(\eta^\tau_{v(\tau)}, \eta^t_{w(t)}) dt d\tau = \int_\alpha^\beta \sigma\Big(\int_\alpha^t \eta^\tau_{v(\tau)} d\tau, \eta^t_{w(t)}\Big) dt = \int_\alpha^\beta \sigma\Big(\eta^\tau_{v(\tau)}, \int_\tau^\beta \eta^t_{w(t)} dt\Big) d\tau. \tag{12.24}$$

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 (13.21), we rewrite the Hessian applied to v_s as follows

$$\lambda_0 \mathrm{Hess}_0 G(v_s(\cdot)) = \int_t^{t+s} \sigma\Big(\int_t^\tau \eta^\theta_{v_s(\theta)} d\theta, \eta^\tau_{v_s(\tau)}\Big) d\tau. \tag{12.26}$$

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

$$\begin{aligned}
\lambda_0 \mathrm{Hess}_0 G(v_s(\cdot)) &= \int_t^{t+s} \sigma(\eta^\tau_{w_s(\tau)}, \eta^\tau_{v_s(\tau)}) d\tau - \int_t^{t+s} \sigma\Big(\int_t^\tau \dot\eta^\theta_{w_s(\theta)} d\theta, \eta^\tau_{v_s(\tau)}\Big) d\tau \\
&= \int_t^{t+s} \sigma(\eta^\tau_{w_s(\tau)}, \eta^\tau_{v_s(\tau)}) d\tau - \int_t^{t+s} \sigma\Big(\dot\eta^\tau_{w_s(\tau)}, \int_\tau^{t+s} \eta^\theta_{v_s(\theta)} d\theta\Big) d\tau \qquad (12.28)
\end{aligned}$$

where the second equality uses (13.21).

Next consider the second term in (12.28) and apply again the integration by part formula (recall that $w_s(t+s) = 0$)

$$\begin{aligned}
\int_t^{t+s} \sigma\Big(\dot\eta^\tau_{w_s(\tau)}, \int_\tau^{t+s} \eta^\theta_{v_s(\theta)} d\theta\Big) d\tau = &- \int_t^{t+s} \sigma(\dot\eta^\tau_{w_s(\tau)}, \eta^\tau_{w_s(\tau)}) d\tau \\
&- \int_t^{t+s} \sigma\Big(\dot\eta^\tau_{w_s(\tau)}, \int_\tau^{t+s} \dot\eta^\theta_{w_s(\theta)} d\theta\Big) d\tau.
\end{aligned}$$

Collecting together all these results one obtains

$$\begin{aligned}\lambda_0 \mathrm{Hess}_0 G(v_s(\cdot)) = &\int_t^{t+s} \sigma(\eta^\tau_{w_s(\tau)},\eta^\tau_{v_s(\tau)})d\tau \\ &+\int_t^{t+s} \sigma(\dot\eta^\tau_{w_s(\tau)},\eta^\tau_{w_s(\tau)})d\tau \\ &+\int_t^{t+s} \sigma(\dot\eta^\tau_{w_s(\tau)},\int_\tau^{t+s} \dot\eta^\theta_{w_s(\theta)}d\theta)d\tau\end{aligned}$$

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:

$$\begin{aligned}\lambda_0 \mathrm{Hess}_0 G(v_s(\cdot)) = \ & s^2 \int_0^1 \sigma(\eta^{t+s\theta}_{w(\theta)},\eta^{t+s\theta}_{v(\theta)})d\theta \\ &+s^3 \int_0^1 \sigma(\dot\eta^{t+s\theta}_{w(\theta)},\eta^{t+s\theta}_{w(\theta)})d\theta \\ &+ s^4 \int_0^1 \sigma(\dot\eta^{t+s\theta}_{w(\theta)},\int_\theta^1 \dot\eta^{t+s\zeta}_{w(\zeta)}d\zeta)d\theta\end{aligned} \tag{12.29}$$

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 $\mathrm{ind}^- \lambda_0 \mathrm{Hess}_0 G < +\infty$. 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 too. Hence, for s small enough, the quadratic form $\lambda_0 \mathrm{Hess}_0 G$ 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 $A : \mathbb{R}^k \times \mathbb{R}^k \to \mathbb{R}$ 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 $\mathcal{U} := \{w(\cdot) \in \mathrm{Lip}[0,1], w(0) = w(1) = 0\}$. Then $\mathrm{ind}^- Q < +\infty$ if and only if $A \equiv 0$.

Proof. Clearly if $A = 0$, then $Q = 0$ and $\mathrm{ind}^- Q = 0$. Assume then that $A \neq 0$ and we prove that $\mathrm{ind}^- Q = +\infty$. 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.

Assume that there exists $x, y \in \mathbb{R}^k$ such that $A(x, y) \neq 0$, and consider a smooth nonconstant 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 $z \in \mathbb{R}^k$ 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 zero 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 built the function

$$w_x(t) = x_i\, w(Nt - i), \qquad t \in [\frac{i}{N}, \frac{i+1}{N}], \qquad i = 1, \ldots, N.$$

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 is nondegenerate. $\square$

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

$$\begin{aligned}\lambda_0 \mathrm{Hess}_0 G(v(\cdot)) &= s^3 \int_0^1 \sigma(\dot{\eta}_{w(\theta)}^{t+s\theta}, \eta_{w(\theta)}^{t+s\theta}) d\theta + O(s^4) \\ &= s^3 \int_0^1 \sigma(\dot{\eta}_{w(\theta)}^{t+s\theta}, \eta_{w(\theta)}^{t}) d\theta + O(s^4)\end{aligned}$$

where the last equalily follows from the fact that η_v^t is Lipschitz with respect to t (see also (12.21)), i.e.

$$\eta_v^{t+s\theta} = \eta_v^t + O(s)$$

On the other hand $\dot{\eta}_v^t$ is only measurable bounded, but the Lebesgue points of u are the same of $\dot{\eta}$. In particular if t is a Lebesgue point of $\dot{\eta}$, the quantity $\dot{\eta}_{w(\cdot)}^t$ is well defined and we can write

$$\begin{aligned}\lambda_0 \mathrm{Hess}_0 G(v(\cdot)) = s^3 & \int_0^1 \sigma(\dot{\eta}_{w(\theta)}^{t}, \eta_{w(\theta)}^{t}) d\theta \\ & - s^3 \left(\int_0^1 \sigma(\dot{\eta}_{w(\theta)}^{t+s\theta}, \eta_{w(\theta)}^{t}) - \sigma(\dot{\eta}_{w(\theta)}^{t}, \eta_{w(\theta)}^{t}) d\theta \right) + O(s^4)\end{aligned}$$

Using the linearity of σ and the boundedness of the vector fields we can estimate

$$\begin{aligned}\left| \int_0^1 \sigma(\dot{\eta}_{w(\theta)}^{t+s\theta}, \eta_{w(\theta)}^{t}) - \sigma(\dot{\eta}_{w(\theta)}^{t}, \eta_{w(\theta)}^{t}) d\theta \right| &\leq C \int_0^1 |\dot{\eta}_{w(\theta)}^{t+s\theta} - \dot{\eta}_{w(\theta)}^{t}| d\theta \\ &\leq C \sup_{|v|\leq 1} \frac{1}{s} \int_0^s |\dot{\eta}_v^{t+\tau} - \dot{\eta}_v^t| d\tau \xrightarrow[s\to 0]{} 0\end{aligned}$$

where the last term tends to zero by definition of Lebesgue point. Hence we come to

$$\lambda_0 \mathrm{Hess}_0 G(v(\cdot)) = s^3 \int_0^1 \sigma(\dot{\eta}_{w(\theta)}^{t}, \eta_{w(\theta)}^{t}) d\theta + o(s^3) \tag{12.32}$$

□

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 12.22. *Let $Q : \mathbb{R}^k \to \mathbb{R}$ be a quadratic form on $\mathbb{R}^k$ and*

$$\mathcal{U} := \{w(\cdot) \in \mathrm{Lip}[0,1], w(0) = w(1) = 0\}.$$

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

$$U_{\gamma(t)} \to \mathbb{R}, \qquad v \mapsto \left\langle \lambda(t), [[f_{u(t)}, f_v], f_v](\gamma(t)) \right\rangle$$

is well defined and does not depend on the extension of f_v to a vector field $f_{v(t)}$ on $\mathbf{U}$.

Notice that, using the notation $h_v(\lambda) = \langle \lambda, f_v(q) \rangle$ an abnormal extremal satisfies

$$h_v(\lambda_t) \equiv 0, \qquad \forall\, v \in \mathbb{R}^k$$

Recalling that the Poisson bracket between linear functions on T^*M is computed by the Lie bracket

$$\{h_v, h_w\}(\lambda) = \langle \lambda, [f_v, f_w](q) \rangle$$

we can rewrite the Goh condition as follows

$$\{h_v, h_w\}(\lambda(t)) \equiv 0, \qquad \forall\, v, w \in \mathbb{R}^k \tag{12.33}$$

while strong Legendre conditions reads

$$\{\{h_{u(t)}, h_v\}, h_v\} \geq 0, \qquad \forall\, v \in \mathbb{R}^k \tag{12.34}$$

Taking derivative of (12.33) with respect to t we find

$$\{h_{u(t)}, \{h_v, h_w\}\}(\lambda(t)) \equiv 0, \qquad \forall\, v, w \in \mathbb{R}^k$$

and using Jacobi identity of the Poisson bracket we get that the bilinear form

$$(v, w) \mapsto \{\{h_{u(t)}, h_v\}, h_w\}(\lambda) \tag{12.35}$$

is symmetric. Hence the generalized Legendre condition says that the quadratic form associated to (12.35) is nonnegative.

Now we want to characterize the trajectories that satisfy these conditions. Recall that, if $\lambda(t)$ is an abnormal geodesic, we have

$$\dot{\lambda}(t) = \vec{h}_{u(t)}(\lambda(t)), \qquad h_i(\lambda(t)) \equiv 0, \quad 0 \leq t \leq 1. \tag{12.36}$$

where $\vec{h}_{u(t)} = \sum_{i=1}^k u_i(t) \vec{h}_i(t)$. Moreover for any smooth function $a : T^*M \to \mathbb{R}$

$$\frac{d}{dt} a(\lambda(t)) = \{h_{u(t)}, a\}(\lambda(t)) = \sum_{i=1}^k u_i(t) \{h_i, a\}(\lambda(t))$$

Notation. We will denote the iterated Poisson brackets

$$\begin{aligned} h_{i_1 \dots i_k}(\lambda) &= \{h_{i_1}, \dots, \{h_{i_{k-1}}, h_{i_k}\}\}(\lambda) && (12.37)\\ &= \left\langle \lambda, [f_{i_1}, \dots, [f_{i_{k-1}}, f_{i_k}]](q) \right\rangle, \qquad q = \pi(\lambda) && (12.38) \end{aligned}$$

Differentiating the identities in (12.36), using (12.37), we get

$$h_i(\lambda(t)) = 0 \qquad \Rightarrow \qquad \sum_{j=1}^{k} u_j(t) h_{ji}(\lambda(t)) = 0, \quad \forall\, t. \tag{12.39}$$

If k is odd we always have a nontrivial solution of the system, if k is even is possible only for those λ that satisfy $\det\{h_{ij}(\lambda)\} = 0$. But we want to characterize only those controls that satisfy Goh conditions, i.e. such that

$$h_{ij}(\lambda(t)) \equiv 0. \tag{12.40}$$

Hence you cannot recover the control u from the linear system (12.39). We differentiate again equations (12.40) and we find

$$\sum_{l=1}^{k} u_l(t) h_{lij}(\lambda(t)) \equiv 0. \tag{12.41}$$

For every fixed t, these are $k(k-1)/2$ equations in k variables $u_1, \ldots, u_k$. Hence

- (i) If $k = 2$, we have 1 equation in 2 variables and we can recover the control u_1, u_2 up to a scalar mutilplier, if at least one of the coefficients does not vanish. Since we can always deal with lengh-parametrized curve this uniquely determine the control u.
- (ii) If $k \geq 3$, we have that the system is overdetermined.

Remark 12.23. For generic systems it is proved that, when $k \geq 3$, Goh conditions are not satisfied. On the other hand, in the case of Carnot groups, for big codimension of the distribution, abnormal minimizers always appear.

12.4 Rank 2 distributions and nice abnormal extremals

Consider a rank 2 distribution generated by a local frame f_1, f_2 and let h_1, h_2 be the associated linear Hamiltonian. An abnormal extremal $\lambda(t)$ associated with a control $u(t)$ satisfies the system of equations

$$\begin{aligned} \dot{\lambda}(t) &= u_1(t)\vec{h}_1(\lambda(t)) + u_2(t)\vec{h}_2(\lambda(t)), \\ h_1(\lambda(t)) &= h_2(\lambda(t)) = 0. \end{aligned} \tag{12.42}$$

Define the linear Hamiltonian associated with the $h_{12}(\lambda(t)) = \langle \lambda, [f_1, f_2](q) \rangle$. Notice that in this special framework the Goh condition is rewritten as $h_{12}(\lambda(t)) = 0$ for a.e. t.

Equivalently, every abnormal extremal satisfies Goh conditions if and only if

$$\lambda(t) \in (\mathcal{D}^2)^{\perp}.$$

Lemma 12.24. *Every nontrivial abnormal extremal on a rank 2 sub-Riemannian structure satisfies the Goh condition.*

Proof. Indeed differentiating the identity (12.42) one gets (we omit t in the notation for simplicity)

$$\begin{aligned} u_2\{h_2, h_1\} &= u_2 h_{21}(\lambda) = 0, \\ u_1\{h_1, h_2\} &= -u_1 h_{21}(\lambda) = 0, \end{aligned}$$

Since at least one among u_1 and u_2 is not identically zero, we have that $h_{12}(\lambda(t)) \equiv 0$, that is Goh condition. $\square$

From now on we focus on a special class of abnormal extremals.

Definition 12.25. An abnormal extremal $\lambda(t)$ is called *nice abnormal* if, for every $t \in [0,1]$, it satisfies

$$\lambda(t) \in (\mathcal{D}^2)^\perp \setminus (\mathcal{D}^3)^\perp.$$

Remark 12.26. Assume that $\lambda(t)$ is a nice abnormal extremal. The system (12.41) obtained by differentiating twice the equations (12.42) reads

$$u_1 h_{112}(\lambda) = u_2 h_{221}(\lambda). \tag{12.43}$$

Under our assumption, at least one coefficient in (12.43) is nonzero and we can uniquely recover the control $u = (u_1, u_2)$ up to a scalar as follows

$$u_1(t) = h_{221}(\lambda(t)), \qquad u_2(t) = h_{112}(\lambda(t)). \tag{12.44}$$

If we plug this control into the original equation we find that $\lambda(t)$ is a solution of

$$\dot{\lambda} = h_{221}(\lambda)\vec{h}_1(\lambda) + h_{112}(\lambda)\vec{h}_2(\lambda). \tag{12.45}$$

Let us now introduce the quadratic Hamitonian

$$H_0 = h_{221}h_1 + h_{112}h_2. \tag{12.46}$$

Theorem 12.27. *Any abnormal extremal belong to $(\mathcal{D}^2)^\perp$. Moreover we have that $\lambda(t) \in (\mathcal{D}^2)^\perp \setminus (\mathcal{D}^3)^\perp$ for all $t \in [0,1]$ if and only if $\lambda(t)$ satisfies*

$$\dot{\lambda}(t) = \vec{H}_0(\lambda(t)) \tag{12.47}$$

with initial condition $\lambda_0 \in (\mathcal{D}^2_q)^\perp \setminus (\mathcal{D}^3_q)^\perp$.

Remark 12.28. Notice that, as soon as $n > 3$, the set $(\mathcal{D}^2_q)^\perp \setminus (\mathcal{D}^3_q)^\perp$ is nonempty for an open dense set of $q \in M$. Indeed assume that we have $\mathcal{D}^2_q = \mathcal{D}^3_q$ for any q in a open neighborhood O_{q_0} of a point q_0 in M. Then it follows that

$$\mathcal{D}^2_{q_0} = \mathcal{D}^3_{q_0} = \mathcal{D}^4_{q_0} = \dots$$

and the structure cannot be bracket generating, since $\dim \mathcal{D}^i_{q_0} < \dim M$ for every $i > 1$. The case $n = 3$ will be treated separately.

Proof. Using that any abnormal extremal belong to the subset $\{h_1(\lambda(t)) = h_2(\lambda(t)) = 0\}$, it is easy to show that an abnormal extremal $\lambda(t)$ satisfies (12.45) if and only if it is an integral curve of the Hamiltonian vector field $\vec{H}_0$.

It remains to prove that a solution of the system

$$\dot{\lambda}(t) = \vec{H}(\lambda(t)), \qquad \lambda_0 \in (\mathcal{D}^2)^\perp \setminus (\mathcal{D}^3)^\perp, \tag{12.48}$$

satisfies $\lambda(t) \in (\mathcal{D}^2)^\perp \setminus (\mathcal{D}^3)^\perp$ for every t. First notice that the solution cannot intersect the set $(\mathcal{D}^3)^\perp$ since these are equilibrium points of the system (12.48) (since at these points the Hamiltonian has a root of order two).

We are reduced to prove that $(\mathcal{D}^2)^\perp$ is an invariant subset for $\vec{H}$. Hence we prove that the functions h_1, h_2, h_{12} are constantly zero when computed on the extremal.

To do this we find the differential equation satisfied by these Hamiltonians. 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. We see that there exists smooth functions a_1, a_2, a_{12} and b_1, b_2, b_{12} such that

$$\begin{cases}\dot{h}_1 = a_1h_1 + a_2h_2 + a_{12}h_{12} \\ \dot{h}_2 = b_1h_1 + b_2h_2 + b_{12}h_{12} \\ \dot{h}_{12} = c_1h_1 + c_2h_2\end{cases} \tag{12.49}$$

If we plug the solution $\lambda(t)$ into the equation of (12.48), 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.

$$h_1(\lambda_0) = h_2(\lambda_0) = h_{12}(\lambda_0) = 0. \tag{12.50}$$

Hence

$$h_1(\lambda(t)) = h_2(\lambda(t)) = h_{12}(\lambda(t)) = 0, \qquad \forall\, t.$$

□

We also can prove easily 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 12.29. *Let $\lambda(t)$ be a nice abnormal. 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

$$Q_t : v \mapsto \left\langle \lambda(t), [[f_{u(t)}, f_v], f_v] \right\rangle, \qquad v \in \mathbb{R}^2 \tag{12.51}$$

is non negative definite. We already proved (cf. **??**) that the bilinear form

$$B_t : (v, w) \mapsto \left\langle \lambda(t), [[f_{u(t)}, f_v], f_w] \right\rangle, \qquad v, w \in \mathbb{R}^2 \tag{12.52}$$

is symmetric. From (12.52) it is easy to see that $u(t) \in \ker B_t$ for every t. Hence Q_t is degenerate for every t. On the other hand if the quadratic form is identically zero we have $\lambda(t) \in (\mathcal{D}^3)^\perp$, which is a contradiction.

Hence the quadratic form has rank 1 and is semi-definite and we can choose $\pm\lambda_0$ in such a way that (12.51) is positive at $t = 0$. Since the sign of the quadratic form does not change along the curve (it is continuous and it cannot vanish) we have that it is positive for all t. □

12.5 Optimality 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 12.30. *Let $\lambda(t)$ be a nice abnormal extremal and let $\gamma(t)$ be corresponding abnormal trajectory. Then there exists $s > 0$ such that $\gamma|_{[0,s]}$ is a strict local length minimizer in the L^2-topology for the controls (equivalently the H^1-topology for trajectories).*

Remark 12.31. Notice that this property of γ does not depend on the metric but only on the distribution. In particular the value of s will be independent on the metric structure defined on the distribution.

It follows that, as soon as the metric is fixed, small pieces of nice abnormal are also global minimizers.

Before proving Theorem 12.30 we prove the following technical result.

Lemma 12.32. *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 Φ*

$$\lambda\, D_0\Phi = 0, \qquad \lambda \in \mathbb{R}^{n*},\ \lambda \neq 0.$$

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

$$\Phi(x) \notin \mathbb{R}^+ v, \qquad \forall\, x \in O, x \neq 0, \qquad \mathbb{R}^+ = \{\alpha \in \mathbb{R}, \alpha > 0\}.$$

In particular the map Φ 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.

Split the domain and the range of the map as follows

$$E = E_1 \oplus E_2, \qquad E_2 = \ker D_0\Phi, \tag{12.53}$$

$$\mathbb{R}^n = \mathbb{R}^{k_1} \oplus \mathbb{R}^{k_2}, \qquad \mathbb{R}^{k_1} = \mathrm{im}\, D_0\Phi, \tag{12.54}$$

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 Φ_1 is a submersion by construction, the Implicit function theorem implies that by a smooth change of coordinates we can linearize Φ_1 and assume that Φ 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 Φ 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 non negative in a neighborhood of 0.

Assume now that $\Phi(x_1, x_2) = sv$ for some $s \geq 0$. Since $v \in \mathbb{R}^{k_2}$ 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

$$\frac{d}{ds}\bigg|_{s=0} a(\Phi_2(0, x_2)) = \frac{d}{ds}\bigg|_{s=0} a(sv) = \langle \lambda, v \rangle \leq 0 \quad \Rightarrow \quad a(sv) \leq 0 \quad \text{for} \quad s \geq 0$$

which is a contradiction. □

Let $\lambda(t)$ be an abnormal extremal and let $\gamma(t)$ be corresponding abnormal trajectory.

$$\dot{\gamma} = u_1 f_1(\gamma) + u_2 f_2(\gamma). \tag{12.55}$$

In what follows we always assume that $\bar{\gamma} \doteq \{\gamma(t) : t \in [0, 1]\}$ is a smooth one-dimensional submanifold of M, with or without border. Then either the curve γ has no self-intersection or $\bar{\gamma}$ is diffeomorfic to S^1. In both cases we can chose a basis f_1, f_2 in a neighborhood of $\bar{\gamma}$ in such a way that γ is the integral curve of f_1

$$\dot{\gamma} = f_1(\gamma)$$

Then γ is the solution of (12.55) with associated control $\widetilde{u} = (1, 0)$. 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

$$F : (u_1, u_2) \mapsto \gamma(1)$$

as the composition

$$F = e^{f_1} \circ G$$

where G is the end point map for the system

$$\dot{q} = (u_1 - 1) e_*^{-tf_1} f_1 + u_2 e_*^{-tf_1} f_2. \tag{12.56}$$

Since $e_*^{-tf_1} f_1 = f_1$, denoting $g_t := e_*^{-tf_1} f_2$ and defining the primitives

$$w(t) = \int_0^t (1 - u_1(\tau)) d\tau, \qquad v(t) = \int_0^t u_2(\tau) d\tau, \tag{12.57}$$

we can rewrite the system, whose endpoint map is G, as follows

$$\dot{q} = -\dot{w} f_1(q) + \dot{v} g_t(q).$$

The Hessian of G is computed

$$\lambda_0 \mathrm{Hess}_0 G(u_1, \dot{v}) = \int_0^1 \langle \lambda_0, [\int_0^t -\dot{w}(\tau) f_1 + \dot{v}(\tau) g_\tau d\tau, -\dot{w}(t) f_1 + \dot{v}(t) g_t](q_0) \rangle dt. \tag{12.58}$$

Recall that

$$\begin{aligned} D_0G(u_1, \dot v) &= \int_0^1 -\dot w(t) f_1(q_0) + \dot v(t) g_t(q_0) dt \\ &= -w(1) f_1(q_0) + \int_0^1 \dot v(t) g_t(q_0) dt \end{aligned}$$

and the condition $\lambda_0 \in \operatorname{im} D_0G^{\perp}$ is rewritten as

$$\langle \lambda_0, f_1(q_0) \rangle = \langle \lambda_0, g_t(q_0) \rangle = 0, \qquad \forall\, t. \tag{12.59}$$

Notice that since equality (12.59) is valid for all t then we have that

$$\langle \lambda_0, \dot g_t(q_0) \rangle = \langle \lambda_0, [f_1, g_t](q_0) \rangle = 0, \tag{12.60}$$

Then we can rewrite our quadratic form only as a function of $\dot v$, since all terms containing $\dot w$ disappear

$$\lambda_0 \mathrm{Hess}_0 G(\dot v) = \int_0^1 \langle \lambda_0, [\int_0^t \dot v(\tau) g_\tau d\tau, \dot v(t) g_t](q_0) \rangle dt \tag{12.61}$$

with the extra condition

$$\int_0^1 \dot v(t) g_t(q_0) dt = w(1) f_1(q_0). \tag{12.62}$$

Now we rearrange these formulas, using integration by parts, rewriting the Hessian as a quadratic form on the space of primitives

$$v(t) = \int_0^t \dot v(\tau) d\tau$$

Using the equality

$$\int_0^t \dot v(\tau) g_\tau d\tau = v(t) g_t - \int_0^t v(\tau) \dot g_\tau d\tau \tag{12.63}$$

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 addend 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

$$\lambda \mathrm{Hess}_0 G(\dot v) = \int_0^1 \langle \lambda_0, [\dot g_t, g_t](q_0)\rangle v(t)^2 dt$$
$$+ \int_0^1 \langle \lambda_0, [\int_0^t v(\tau)\dot g_\tau, v(t)\dot g_t - v(1)g_1](q_0)\rangle dt \tag{12.64}$$

which can also be rewritten as follows

$$\lambda \mathrm{Hess}_0 G(\dot v) = \int_0^1 \langle \lambda_0, [\dot g_t, g_t](q_0)\rangle v(t)^2 \, dt$$
$$+ \int_0^1 \langle \lambda_0, [\int_1^t v(\tau)\dot g_\tau \, d\tau + v(1)g_1, v(t)\dot g_t](q_0) \, dt. \tag{12.65}$$

Moreover, again integrating by parts the extra condition (12.62), we find

$$\int_0^1 v(t)\dot g_t(q_0) dt = -w(1)f_1(q_0) + v(1)g_1(q_0) \tag{12.66}$$

Remark 12.33. Notice that we cannot plug in the expression (12.66) directly into the formula since this equality is valid only at the point q_0, while in (12.64) we have to compute the bracket.

Notice that the vectors $f_1(q_1)$ and $f_2(q_1)$ are linearly independent, then also

$$f_1(q_0) = e_*^{-f_1}(f_1(q_1)), \qquad \text{and} \qquad g_1(q_0) = e_*^{-f_1}(f_2(q_1)),$$

are linearly independent. From (12.66) it follows that for every pair (w, v) in the kernel the following estimates are valid

$$|w(1)| \leq C\|v\|_{L^2}, \qquad |v(1)| \leq C\|v\|_{L^2}. \tag{12.67}$$

Theorem 12.34. *Let $\gamma : [0,1] \to M$ be an abnormal trajectory and assume that the quadratic form* (12.64) *satisfies*

$$\lambda_0 \mathrm{Hess}_0 G(\dot v) \geq \alpha \|v\|_{L^2}^2. \tag{12.68}$$

Then the curve is locally minimizer in the L^2 topology of controls.

Remark 12.35. Notice that the estimate (12.68) depends only on v, while the map G is a smooth map of $\dot v$ and $\dot w$. Hence Lemma 12.32 does not apply.

Moreover, the statement of Lemma 12.32 violates for the endpoint map, since it is locally open as soon as the bracket generating condition is satisfied (this is equivalent to the Chow-Rashevsky Theorem). Moreover the final point of the trajectory is never isolated in the level set.

What we are going to use is part of the proof of this Lemma, to show that the statements holds for the restriction of the endpoint map to some subset of controls

Proof of Theorem 12.34. Our goal is to prove that there are no curves shorter than γ that join q_0 to $q_1 = \gamma(1)$.

To this extent we consider the *restriction* of the endpoint map to the set of curves that are shorter or have the same lenght than the original curve. Hence we need to fix some sub-Riemannian structure on M.

We can then assume the orthonormal frame f_1, f_2 to be fixed and that the length of our curve is exactly 1 (we can always dilate all the distances on our manifold and the local optimality of the curve is not affected).

The set of curves of length less or equal than 1 can be parametrized, using Lemma 3.15, by the set

$$\{(u_1, u_2)|\, u_1^2 + u_2^2 \leq 1\}$$

Following the notation (12.57), notice that

$$\{(u_1, u_2)|\, u_1^2 + u_2^2 \leq 1\} \subset \{(w, v)|\, \dot{w} \geq 0\}.$$

We want to show that, for some function $a \in C^\infty(M)$ such that $d_q a = \lambda \in \operatorname{im} D_0 F^\perp$, we have

$$a \circ F\big|_D(\dot{w}, \dot{v}) = \lambda \mathrm{Hess}_0 F(\dot{w}, \dot{v}) + R(w, v), \qquad \text{where} \quad \frac{R(w, v)}{\|v\|^2} \underset{\|(\dot{w},\dot{v})\| \to 0}{\longrightarrow} 0 \tag{12.69}$$

in the domain

$$D = \{(\dot{w}, \dot{v}) \in \ker D_0 F, \dot{w} \geq 0\}$$

Indeed if we prove (12.69) we have that the point $(\dot{w}, \dot{v}) = (0, 0)$ is locally optimal for F. This means that the curve γ, i.e. the curve associated to controls $u_1 = 1, u_2 = 0$, is also 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 (6.29) to the endpoint map F we get

$$\begin{aligned} F(\dot{w}, \dot{v}) &= q_0 \circ \overrightarrow{\exp} \int_0^1 (1 - \dot{w}(t)) f_1 + \dot{v}(t) f_2 \, dt \\ &= q_0 \circ \overrightarrow{\exp} \int_0^1 (1 - \dot{w}(t)) e_*^{-v(t) f_2} f_1 \, dt \circ e^{v(1) f_2} \end{aligned}$$

Hence we can express the endpoint map as a smooth function of the pair $(\dot{w}, v)$.

Now, to compute (12.69), 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:

$$a \circ F(\dot{w}, \dot{v}) = q_0 \circ \overrightarrow{\exp} \int_0^1 (1 - \dot{w}(t)) e_*^{-v(t) f_2} f_1 \, dt \, a$$

Writing

$$(1 - \dot{w}(t)) e_*^{-v(t) f_2} f_1 = f_1 + X^0(v(t)) + \dot{w}(t) X^1(v(t)) \tag{12.70}$$

and using the variation formula (6.30), setting $Y_t^i = e_*^{(t-1) f_1} X^i$ for $i = 0, 1$, we get (recall that $q_1 = q_0 \circ e^{f_1}(q_0)$)

$$a \circ F(\dot{w}, \dot{v}) = q_1 \circ \overrightarrow{\exp} \int_0^1 Y_t^0(v(t)) + \dot{w}(t) Y_t^1(v(t)) dt \, a, \qquad Y_t^0(0) = Y_t^1(0) = 0,$$

Expanding the chronological exponential we find that

(a) the zero order term vanish since $Y^0_t(0) = Y^1_t(0) = 0$,

(b) all first order terms vanish since the vector fields f_1 and $[f_1, f_2]$ spans the image of the differential (hence are orthogonal to $\lambda = d_q a$)

(c) the second order terms are in the Hessian, since our domain D is contained in the kernel of the differential

In other words it remains to show that every term in v, w of order greater or equal than 3 in the expansion can be estimated with $o(\|v\|^2)$.[3]

Let us prove first the claim for monomial of order three:

$$\int_0^1 \dot{w}(t)v^2(t)dt = o(\|v\|^2), \qquad \int_0^1 \dot{w}(t)\int_0^t \dot{w}(\tau)v(\tau)d\tau dt = o(\|v\|^2)$$

$$\int_0^1 \dot{w}(t)\int_0^t \dot{w}(\tau)\int_0^\tau \dot{w}(s)ds d\tau dt = o(\|v\|^2)$$

Using that $\dot{w} \geq 0$, which is the key assumption, and the fact that $(\dot{w}, \dot{v}) \in \ker D_0F$, which gives the estimates (12.67), 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 we 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

$$\begin{aligned}\left|\int_0^1 \dot{w}(t)\int_0^t \dot{w}(\tau)v(\tau)d\tau dt\right| &\leq 2|w(1)|\int_0^1 v(t)\dot{w}(t)dt \\ &\leq C\|\dot{w}\|\|v\|^2\end{aligned}$$

[3]where $o(\|v\|^2)$ have the same meaning as in (12.69).

Finally, the last integral is very easy to estimate using the equality

$$\begin{aligned}\int_0^1 \dot{w}(t)\int_0^t \dot{w}(\tau)\int_0^\tau \dot{w}(s)dsd\tau dt &= \frac{1}{6}\int_0^1 \dot{w}(t)^3 dt\\ &\le C\|\dot{w}\|\|v\|^2\end{aligned}$$

Starting from these estimate it is easy to show that any mixed monomial of order greater that three satisfies these estimates 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 12.30 . If we apply the arguments above to a small piece $\gamma_s = \gamma|_{[0,s]}$ of the curve γ it is easy to see that the Hessian rescale as follows,

$$\begin{aligned}\lambda_0 \mathrm{Hess}_0 G_s(v) = &\int_0^s \langle \lambda_0, [g_t, \dot{g}_t](q_0)\rangle v(t)^2 dt\\ &+ \int_0^s \langle \lambda_0, [\int_0^t v(\tau)\dot{g}_\tau d\tau, v(t)\dot{g}_t - v(s)g_s](q_0)\rangle dt\end{aligned}$$

Since the generalized Legendre condition ensures[4] that (see also Lemma 12.29)

$$\langle \lambda_0, [g_t, \dot{g}_t](q_0)\rangle \ge C > 0$$

then the norm

$$\|v\|_g = \left(\int_0^s \langle \lambda_0, [g_t, \dot{g}_t](q_0)\rangle v(t)^2 dt\right)^{1/2} \tag{12.71}$$

is equivalent to the standard L^2-norm. Hence the Hessian can be rewritten as

$$\lambda \mathrm{Hess}_0 G_s(v) = \|v\|_g + \langle Tv, v\rangle \tag{12.72}$$

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 $\|T\|^2 = \|K\|^2_{L^2} \to 0$ for $s \to 0$, it follows that the Hessian is positive definite for small $s > 0$. ☐

12.6 Conjugate points along abnormals

In this section, we give an effective way to check the inequality (12.68) that implies local minimality of the nice abnormal geodesic according to Theorem 12.34.

[4] it is semidefinite and we already know that f_1 is in the kernel

We define $Q_1(v) := \lambda \mathrm{Hess}_0 G(\dot v)$. 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

$$D(Q_1) = \left\{ v \in L_2[0,1] : \int_0^1 v(t)\dot g_t(q_0)\,dt \in \mathrm{span}\{f_1(q_0), g_1(q_0)\} \right\}.$$

The extension of Q_1 to this closure is denoted by the same symbol Q_1. We set:

$$l(t) = \langle \lambda_0, [\dot g_t, g_t](q_0)\rangle, \quad X_t = v_1 g_1 + \int_1^t v(\tau)\dot g_\tau\,d\tau$$

and we rewrite the form Q_1 in these more compact notations:

$$Q_1(v) = \int_0^1 l(t)v(t)^2\,dt + \int_0^1 \langle \lambda_0, [X_t, \dot X_t](q_0)\rangle\,dt,$$

$$\dot X_t = v(t)\dot g_t, \quad X_1 \wedge g_1 = 0,\ X_0(q_0) \wedge f_1(q_0) = 0. \tag{1}$$

Moreover, we introduce the family of quadratic forms Q_s, for $0 < s \le 1$, as follows

$$Q_s(v) := \int_0^s l(t)v(t)^2\,dt + \int_0^s \langle \lambda_0, [X_t, \dot X_t](q_0)\rangle\,dt,$$

$$\dot X_t = v(t)\dot g_t, \quad X_s \wedge g_s = 0,\ X_0(q_0) \wedge f_1(q_0) = 0. \tag{1}$$

Recall that $l(t)$ is a strictly positive continuous function. In particular, $\int_0^1 l(t)v(t)^2\,dt$ is the square of a norm of v that is equivalent to the standard L_2-norm. Next statement is proved by the same arguments as Proposition **??**. We leave details to the reader.

Proposition 12.36. *The form Q_1 is positive definite if and only if* $\ker Q_s = 0,\ \forall s \in (0,1]$.

Definition 12.37. A time moment $s \in (0,1]$ is called *conjugate* to 0 for the abnormal geodesic γ 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 λ_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:

$$Q_s(v) = \int_0^s l(t)v(t)^2\,dt + \int_0^s \sigma(x(t), \dot x(t))\,dt,$$

$$\dot x(t) = v(t)\dot\zeta_t, \quad x(s) \wedge \zeta_s = 0,\ \pi_* x(0) \wedge \pi_*\xi_1 = 0, \tag{2}$$

where σ is the standard symplectic product on $T_{\lambda_0}(T^*M)$ and $\pi : T^*M \to M$ is the standard projection. Moreover

$$l(t) = \sigma(\dot\zeta_t, \zeta_t), \quad 0 \le t \le 1. \tag{12.73}$$

Let $E = \mathrm{span}\{\xi_1, \zeta_t, 0 \le t \le 1\}$. We use only the restriction of σ to E in the expression of Q_s and we are going to get rid of unnecessary variables. Namely, we set: $\Sigma \doteq E/(\ker \sigma|_E)$.

Lemma 12.38. $\dim \Sigma \leq 2\left(\dim \operatorname{span}\{f_1(q_0), g_t(q_0), 0 \leq t \leq 1\} - 1\right)$.

Proof. Dimension of Σ 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, $\pi_*(E) = \operatorname{span}\{f_1(q_0), g_t(q_0), 0 \leq t \leq 1\}$ and $E \cap \ker \pi_*$ is an isotropic subspace of $\sigma|_E$. $\square$

We denote by $\underline{\zeta}_t \in \Sigma$ the projection of ζ_t to Σ and by $\Pi \subset \Sigma$ the projection of $E \cap \ker \pi_*$. Note that the projection of ξ_1 to Σ is 0; moreover, equality (12.73) implies that $\underline{\zeta}_t \neq 0,\ \forall t \in [0,1]$. The final expression of Q_s is as follows:

$$Q_s(v) = \int_0^s l(t) v(t)^2 \, dt + \int_0^s \sigma(x(t), \dot{x}(t)) \, dt,$$

$$\dot{x}(t) = v(t) \dot{\underline{\zeta}}_t, \quad x(s) \wedge \underline{\zeta}_s = 0,\ x(0) \in \Pi. \tag{4}$$

We have: $v \in \ker Q_s$ if and only if

$$\int_0^s \left(l(t) v(t) + \sigma(x(t), \dot{\underline{\zeta}}_t) \right) w(t) \, dt = 0,$$

for any $w(\cdot)$ such that

$$\int_0^s \dot{\underline{\zeta}}_t w(t) \, dt \in \Pi + \mathbb{R}\underline{\zeta}_s. \tag{5}$$

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

$$l(t) v(t) + \sigma(x(t), \dot{\underline{\zeta}}_t) = \sigma(\nu, \dot{\underline{\zeta}}_t), \quad 0 \leq t \leq s.$$

We set $y(t) = x(t) - \nu$ and obtain the following:

Theorem 12.39. *A time moment $s \in (0,1]$ is conjugate to 0 if and only if there exists a nontrivial solution of the equation*

$$l(t) \dot{y} = \sigma(\dot{\underline{\zeta}}_t, y) \dot{\underline{\zeta}}_t \tag{12.74}$$

that satisfy the following boundary conditions:

$$\exists\, \nu \in \Pi^\angle \cap \underline{\zeta}_s^\angle \quad \textit{such that} \quad (y(s) + \nu) \wedge \underline{\zeta}_s = 0,\ (y(0) + \nu) \in \Pi. \tag{12.75}$$

Remark 12.40. Notice that identity (12.73) implies that $y(t) = \underline{\zeta}_t$ for $t \in [0,1]$ is a solution to the equation (12.74). However this solution may violate the boundary conditions.

Let us consider the special case: $\dim \operatorname{span}\{f_1(q_0), g_t(q_0), 0 \leq t \leq 1\} = 2$; this is what we automatically have for abnormal geodesics in a 3-dimensional sub-Riemannian manifold. In this case, $\dim E = 2,\ \dim \Pi = 1$; hence $\Pi^\angle = \Pi, \underline{\zeta}_s^\angle = \mathbb{R}\underline{\zeta}_s$ and $\Pi^\angle \cap \underline{\zeta}_s^\angle = 0$. Then ν in the boundary conditions (12.75) must be 0 and $y(s) = c\underline{\zeta}_s$, where c is a nonzero constant. Hence $y(t) = c\underline{\zeta}_t$ for $0 \leq t \leq 1$ and $y(0) = c\underline{\zeta}_0 \notin \Pi$. We obtain:

Corollary 12.41. *If $\dim \operatorname{span}\{f_1(q_0), g_t(q_0), 0 \leq t \leq 1\} = 2$, then the segment $[0,1]$ does not contain conjugate time moments and assumption of Theorem 12.34 is satisfied.*

We can apply this corollary to the isoperimetric problem studied in Section 4.4.2. Abnormal geodesics correspond to connected components of the zero locus of the function b (see notations in Sec. 4.4.2). All these abnormal geodesics are nice if and only if zero is a regular value of b. Take a compact connected component of $b^{-1}(0)$; this is a smooth closed curve. Our corollary together with Theorem 12.34 implies that this closed curve passed once, twice, three times or arbitrary number of times is a locally optimal solution of the isoperimetric problem. Moreover, this is true for any Riemannian metric on the surface M!

12.6.1 Abnormals in dimension 3

Nice abnormals for the isoperimetric problem on surfaces

Recall the isoperimetric problem: given two points x_0, x_1 on a 2-dimensional Riemannian manifold N, a 1-form $\nu \in \Lambda^1 N$ and $c \in \mathbb{R}$, we have to find (if it exists) the minimum:

$$\min\{\ell(\gamma), \gamma(0) = x_0, \gamma(T) = x_1, \int_\gamma \nu = c\} \tag{12.76}$$

As shown in Section 4.4.2, this problem can be reformulated as a sub-Riemannian problem on the extended manifold

$$M = N \times \mathbb{R} = \{(x, y), x \in N, y \in \mathbb{R}\},$$

where the sub-Riemannian structure is defined by the contact form

$$\mathcal{D} = \ker (dy - \nu)$$

and the sub-Riemannian length of a curve coincides with the Riemannian length of its projection on N. If we write $d\nu = b\, dV$, where b is a smooth function and dV denote the Riemannian volume on N, we have that the Martinet surface is defined by the cilynder

$$\mathcal{M} = \mathbb{R} \times b^{-1}(0),$$

where, generically, the set $b^{-1}(0)$ is a regular level of b.

Since the distribution is well behaved with the projection on N by construction, it follows that the distribution is always transversal to the Martinet surface and all abnormal are nice, since $\mathcal{D}^3_q = T_q M$ for all q.

Thus the projection of abnormal geodesics on N are the connected components of the set $b^{-1}(0)$ and we can recover the whole abnormal extremal integrating the 1-form ν to find the missing component. In other words the abnormal extremals are spirals on $\mathcal{M}$ with step equal to $\int_A d\nu$, (if $d\nu$ is the volume form on N, it coincide with the area of the region A inside the curve defined on N by the connected component of $b^{-1}(0)$).

Corollary 12.42. *Let M be a sub-Riemannian manifold,* $\dim M = 3$*, and let $\gamma : [0, 1] \to M$ be a nice abnormal geodesic. Then γ is a strict local minimizer for the L^2 control topology, for any metric.*

Remark 12.43. Notice that we do not require that the curve does not self-intersect since in the 3D case this is automatically guaranteed by the fact that nice abnormal are integral curves of a smooth vector fields on 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 $\mathbb{R}^2 = \{(x_1, x_2),\, x_i \in \mathbb{R}\}$ defined by the 1-form ν such that

$$d\nu = x_1 x_2 dx_1 dx_2.$$

Here $b(x_1, x_2) = x_1x_2$ and the set $b^{-1}(0)$ consists of the union of the two axes, with moreover $db|_0 = 0$.

Let us fix $\bar{x}_1, \bar{x}_2 > 0$ and consider the curve joining $(0, \bar{x}_2)$ and $(\bar{x}_1, 0)$ that is the union of two segment contained in the coordinate axes

$$\gamma : [-\bar{x}_2, \bar{x}_1] \to \mathbb{R}^2, \qquad \gamma(t) = \begin{cases} (0, -t), & t \in [-\bar{x}_2, 0], \\ (t, 0), & t \in [0, \bar{x}_1]. \end{cases}$$

Proposition 12.44. *The curve γ is a projection of an abnormal extremal that is not a length minimizer.*

Proof of Proposition 12.44. Let us built a family of "variations" $\gamma_{\varepsilon,\delta}$ of the curve γ defined as in Figure 12.1. Namely in $\gamma_{\varepsilon,\delta}$ we cut a corner of size ε at the origin and we turn around a small circle of radius δ before reaching the endpoint. Denoting by D_ε and D_δ the two region 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 $d\nu = x_1x_2dx_1dx_2$ 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

$$\ell(\gamma_{\varepsilon,\delta}) - \ell(\gamma) = 2\pi\delta - (2 - \sqrt{2})\varepsilon \tag{12.77}$$

Choosing ε in such a way that $c_1\varepsilon^4 = c_2\delta^3$ it is an easy exercise to show that the quantity (12.77) is negative when $\delta > 0$ is very small. $\square$

Remark 12.45. If you consider some plane curve $\widetilde{\gamma}$ that is a projection of a normal extremal having the same endpoint γ 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.

It is still an open problem to find which is the length minimizer joining these two points. We know that it should be a projection of a normal extremal (hence smooth) but for instance we do not know how many self-intersection it has.

12.6.2 Higher dimension

Now consider another important special case that is typical if dimension of the ambient manifold is greater than 3. Namely, assume that, for some $k \geq 2$, the vector fields

$$f_1,\ f_2,\ (\mathrm{ad} f_1)f_2,\ \ldots,\ (\mathrm{ad} f_1)^{k-1}f_2 \tag{12.78}$$

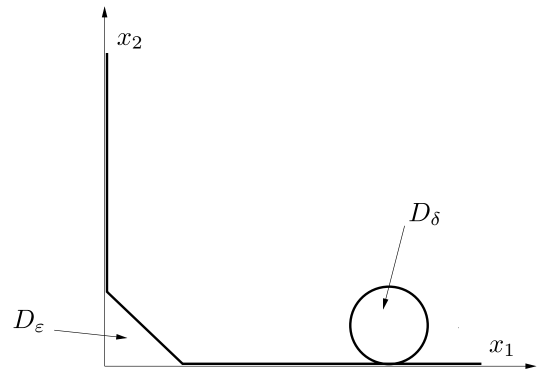


Figure 12.1: An abnormal extremal that is not length minimizer

are linearly independent in any point of a neighborhood of our nice abnormal geodesic γ, while $(\mathrm{a}df_1)^k f_2$ is a linear combination of the vector fields (12.78) 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, α are smooth functions. In this case, all closed to γ solutions of the equation $\dot{q} = f_1(q)$ are abnormal geodesics.

A direct calculation based on the fact that $\langle \lambda_t, (\mathrm{ad}f_1^i)f_2)(\gamma(t)\rangle = 0,\ 0 \le t \le 1$, gives the identity:

$$\zeta_t^{(k)} = \sum_{i=0}^{k-1} a_i(\gamma(t))\zeta^{(i)} + \alpha(\gamma(t))\xi_1. \quad 0 \le t \le 1. \tag{12.79}$$

Identity (12.79) implies that $\dim E = k$ and $\Pi = 0$. The boundary conditions (12.75) take the form:

$$y(0) \in \underline{\zeta}_s^{\angle}, \quad (y(s) - y(0)) \wedge \underline{\zeta}_s = 0. \tag{12.80}$$

The caracterization of conjugate points is especially simple and geometrically clear if the ambient manifold has dimension 4. Let Δ be a rank 2 equiregular distribution in a 4-dimensional manifold (the Engel distribution). Then abnormal geodesics form a 1-foliation of the manifold and condition (12.78) is satisfied with $k = 2$. Moreover, $\dim E = 3$, $\dim \Sigma = 2$ and $\underline{\zeta}_s^{\angle} = \mathbb{R}\underline{\zeta}_s$. Recall that $y(t) = \underline{\zeta}_t,\ 0 \le t \le s$, is a solution to (12.74). Hence boundary conditions (12.80) are equivalent to the condition

$$\underline{\zeta}_s \wedge \underline{\zeta}_0 = 0. \tag{12.81}$$

It is easy to re-write relation (12.81) in the intrinsic way without special notations we used to simplify calculations. We have the following characterization of conjugate times.

Lemma 12.46. *A time moment t is conjugate to 0 for the abnormal geodesic γ 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 it 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. Conjugate moment is a moment when the plane makes a complete revolution. Collecting all the information we obtain:

Theorem 12.47. *Let $\mathcal{D}$ be the Engel 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 γ is an abnormal geodesic. Moreover*

(i) if $e_*^{tf_1}\mathcal{D}_{\gamma(0)} \neq \mathcal{D}_{\gamma(t)}$, $\forall t \in (0,1]$, then γ 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 γ is not a normal geodesic, then γ is not a local length minimizer.

12.7 Equivalence of local minimality

Now we prove that, under the assumption that our trajectory is smooth, it is equivalent to be locally optimal in the H^1-topology or in the 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 γ that is C^0-close to γ satisfying the same boundary conditions, while it is called a H^1-*local length-minimizer* if $\ell(\bar{\gamma}) \leq \ell(\gamma)$ for every curve γ such that the control u corresponding to γ is close in the L^2 topology to the control $\bar{u}$ associated with $\bar{\gamma}$ and γ satisfies the same boundary conditions.

Any C^0-local minimizer is automatically a H^1-local 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

$$|\gamma_v(t) - \gamma_w(t)| \leq C\|u - v\|_{L^2}, \qquad \forall\, t \in [0, T],$$

where γ_v and γ_w are the trajectories associated to controls v,w respectively.

Theorem 12.48. *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^1 and has no self intersections. If $\bar{\gamma}$ is a (strict) local minimizer in the L^2 topology for the controls then $\bar{\gamma}$ is also a (strict) local minimizer in the C^0 topology for the trajectories.*

Proof. Since $\bar{\gamma}$ has no self intersections, we can look for a preferred system of coordinates on an open neighborhood Ω 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

$$I_\varepsilon \times B_\varepsilon^{n-1} = \{(x, y) \in \mathbb{R}^n : x \in]-\varepsilon, 1+\varepsilon[, y \in \mathbb{R}^{n-1}, |y| < \varepsilon\}, \tag{12.82}$$

We need the following technical lemma.

Lemma 12.49. *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 Φ^*g is the identity matrix at $(t, 0)$,i.e., along $\bar{\gamma}$.

Proof of the Lemma. As in the proof of Theorem **??**, 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 Φ^*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 Φ^*g corresponding to the splitting $\mathbb{R}^n = \mathbb{R} \times \mathbb{R}^{n-1}$ defined in (12.82). 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 (in this proof $\langle \cdot,\cdot \rangle$ is the Euclidean product in $\mathbb{R}^n$)

$$T(x,0) = \left\{ \left(\langle v_x, y \rangle, y \right), y \in \mathbb{R}^{n-1} \right\}$$

for some family[5] of vectors $v_x \in \mathbb{R}^{n-1}$, depending smoothly with respect to x. 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 Ψ to $I_\varepsilon \times B_\varepsilon^{n-1}$ is invertible. Notice that this is possible since

$$\det D\Psi(x,y) = 1 - \left\langle \frac{\partial v_x}{\partial x}, y \right\rangle .$$

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

$$\begin{aligned} \|\gamma - \bar{\gamma}\|_{H^1}^2 &= \int_0^1 |\dot{x}(t) - 1|^2 + |\dot{y}(t)|^2 dt \\ &= \int_0^1 |\dot{x}(t)|^2 + |\dot{y}(t)|^2 dt - 2 \int_0^1 \dot{x}(t) dt + 1 \\ &= \int_0^1 |\dot{x}(t)|^2 + |\dot{y}(t)|^2 dt - 1 \end{aligned}$$

where we used that $x(0) = 0$ and $x(1) = 1$ since γ satisfies the boundary conditions. If we denote by

$$J(\gamma) = \int_0^1 \langle G(\gamma(t))\dot{\gamma}(t), \dot{\gamma}(t) \rangle \, dt, \qquad J_e(\gamma) = \int_0^1 |\dot{x}(t)|^2 + |\dot{y}(t)|^2 dt \tag{12.83}$$

respectively the energy of γ and the "Euclidean" energy, we have $\|\gamma - \bar{\gamma}\|_{H^1}^2 = J_e(\gamma) - 1$ and the H^1-local minimality can be rewritten as follows:

[5] 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}.

($*$) there exists $\varepsilon > 0$ such that for every γ 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 $I_\varepsilon \times B_\varepsilon^{n-1}$ 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 γ is contained in $I_\varepsilon \times B_{\delta\varepsilon}^{n-1}$. 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 $(x, y) \in I_\varepsilon \times B_{\delta\varepsilon}^{n-1}$ 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 γ is contained in $I_\varepsilon \times B_{\delta\varepsilon}^{n-1}$ 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 12.50. Notice that this result implies in particular Theorem 4.61, since normal extremals are always smooth. Nevertheless, the argument of Theorem 4.61 can be adapted for more general coercive functional (see [8]), while this proof use specific estimates that hold only for our explicit cost (i.e., the distance).

12.8 Non optimality of corners

Is any sub-Riemannian shortest path smooth? We still do not know if this is always true. We know that normal geodesics are smooth as well as nice abnormal. It is easy to construct abnormal extremal paths but all known examples are not shortest. See, for instance, an example of the nonsmooth abnormal in Sec. 12.6.1: it is a local length minimizer in the L^∞-topology for controls but it is not a shortest path (and not a local length minimizer in the L^p-topology $\forall\, p < \infty$). The following important regularity result shows that "corners" are not shortest paths.

Theorem 12.51 (Hakavuori, Le Donne [60])**.** *Any piecewise smooth parameterized by the length shortest path is of class C^1.*

Proof. Let $q \in M$, $\gamma_i : [0, t_i] \to M$, $i = 1, 2$, are smooth horizontal curves, $\gamma_1(0) = \gamma_2(0) = q$, $|\dot{\gamma}_1(t)| = |\dot{\gamma}_2(t)| = 1$, $\dot{\gamma}_1(0) + \dot{\gamma}_2(0) \neq 0$. We have to show that the concatenation of the curves $t \mapsto \gamma_1(\varepsilon - t)$ and $t \mapsto \gamma_2(t)$, $0 \leq t \leq \varepsilon$, is not a shortest path between $\gamma_1(\varepsilon)$ and $\gamma_2(\varepsilon)$ for an arbitrary small $\varepsilon > 0$.

First we consider the main case of linearly independent $\dot{\gamma}_1(0)$ and $\dot{\gamma}_2(0)$ and then explain what to do in 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 is divided in several steps.

1. Let f_i be horizontal vector fields such that

$$\dot{\gamma}_i(t) = f_i(\gamma_i(t)), \quad 0 \le t \le 1, \quad i = 1, 2.$$

Assume that $d(\gamma_1(t), \gamma_2(t)) = 2t$ for all sufficiently small $t > 0$, where $d(\cdot, \cdot)$ is the sub-Riemannian distance. We are going to show that this assumption leads to a contradiction.

Let $\delta_\varepsilon : O_q \to O_q,\ \varepsilon > 0$, be the dilation associated to some privileged coordinates in a neighborhood O_q of the point q in M (see Chapter 10). We set $d_\varepsilon(q_1, q_2) = \frac{1}{\varepsilon} d(\delta_\varepsilon(q_1), \delta_\varepsilon(q_2)),\ q_1, q_2 \in O_q$, and denote:

$$f_i^\varepsilon = \varepsilon \delta_{\frac{1}{\varepsilon}*} f_i, \quad \gamma_i^\varepsilon(t) = e^{t f_i^\varepsilon}, \quad i = 1, 2;$$

then $d_\varepsilon(\gamma_1^\varepsilon(t), \gamma_2^\varepsilon(t)) = 2t$. Moreover, f_i^ε converges to $\hat{f}_i$ in the C^∞-topology and d_ε uniformly converges to $\hat{d}$ as $\varepsilon \to 0$, where the vector fields $\hat{f}_i,\ i = 1, 2$, are two of generators of the Carnot algebra acting on the nonholonomic tangent space at q and $\hat{d}(\cdot, \cdot)$ is the metric on the nonholonomic tangent space at q (see Section 10.4). We obtain that $\hat{d}\left(e^{t\hat{f}_1}(q), e^{t\hat{f}_2}(q)\right) = 2t$.

2. Nonholonomic tangent space is a homogeneous space of the Carnot group and the distance $\hat{d}(\hat{q}_1, \hat{q}_2)$ is, by definition, minimum of the Carnot group distances between elements of the stable subgroups of the points $\hat{q}_1, \hat{q}_2$ for this action. We keep symbol $\hat{d}$ for the Carnot group distance; then $\hat{d}\left(\hat{(}e^{t\hat{f}_1}, e^{t\hat{f}_2}\right) = 2t$ (it cannot be greater than $2t$ because the length of the concatenation of the curves $\tau \to e^{(t-\tau)\hat{f}_1}$ and $\tau \to e^{\tau \hat{f}_2},\ 0 \le \tau \le t$, equals $2t$).

3. The Carnot algebra may have more than two generators. Let us consider the subalgebra generated by $\hat{f}_1, \hat{f}_2$ 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.

4. We arrived to the key step of the proof and would like to simplify notations. Let G be a Carnot group with a Carnot algebra $\mathfrak{g}$. We assume that $\mathfrak{g}$ is a step k Carnot algebra with two generators, i. e.

$$\mathfrak{g} = g_1 \oplus \cdots \oplus g_k, \quad \mathfrak{g} = Lie\{g_1\}, \quad g_1 = span\{x_1, x_2\}.$$

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"). The statement of Theorem 1 in the no cusps case is reduced to the following:

Proposition 12.52. *$d(e^{x_1}, e^{x_2}) < 2$.*

Proof. We prove this statement by induction in k. For $k = 2$, G is the Heisenberg group where we already know all shortest paths and they are smooth.

Induction step. Assume that the statement is valid for the $(k-1)$-step Carnot groups. Note that g_k is contained in the center of G and e^{g_k} takes part of the center of G. Then G/e^{g_k} is a Carnot group with a step $(k-1)$ Carnot algebra $g_1 \oplus \cdots \oplus g_{k-1}$. Moreover, the sub-Riemannian distance between two points in G/e^{g_k} is simply minimum of the distances between the points of the correspondent residue classes. Taking into account the left-invariance of the distance, we can write:

$$d(e^{g_1} q_1, e^{g_2} q_2) = \min_{z \in g_k} d(e^z q_1, q_2).$$

Our induction assumption implies that there exists $z \in g_k$ such that

$$d(e^z e^{x_1}, e^{x_2}) = 2 - \nu,$$

where $\nu > 0$. Moreover, left-invariance of the distance implies that $d(e^z e^{x_1}, e^{x_2}) = d(1, e^{-x_1} e^{-z} e^{x_2})$.

We have to show that the distance between e^{x_1} and e^{x_2} is smaller than the length of the concatenation of the curves $t \mapsto e^{(1-t)x_1}$ and $t \mapsto e^{tx_2}$, $0 \le t \le 1$. The trick is to demonstrate it playing with non-horizontal curves. First we insert a short piece of the form $t \mapsto e^{-t\varepsilon^k z}$, $0 \le t \le 1$.

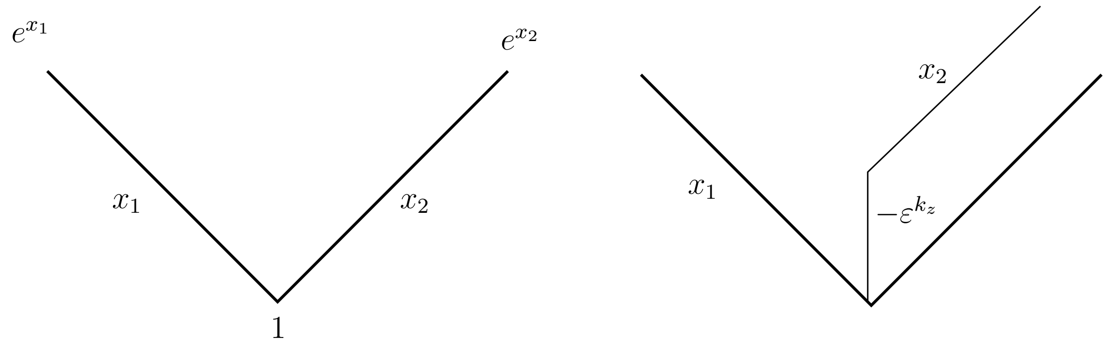


Figure 12.2: Adding one piece

New curve contains a horizontal part of the length 2 but the distance between its endpoints is smaller than 2. I claim that $d(e^{x_1}, e^{-\varepsilon^k z} e^{x_2}) \le 2 - \varepsilon\nu$. Indeed, $d(e^{x_1}, e^{-\varepsilon^k z} e^{x_2}) = d(1, e^{-x_1} e^{-\varepsilon^k z} e^{x_2})$ and

$$e^{-x_1} e^{-\varepsilon^k z} e^{x_2} = e^{(\varepsilon-1)x_1} \left(e^{-\varepsilon x_1} e^{-\varepsilon^k z} e^{\varepsilon x_2} \right) e^{(1-\varepsilon)x_2}.$$

We have: $e^{-\varepsilon x_1} e^{-\varepsilon^k z} e^{\varepsilon x_2} = \delta_\varepsilon \left(e^{x_1} e^{-z} e^{x_2} \right)$, where $\delta.$ is the dilation of the Carnot group. Moreover, $d(1, \delta_\varepsilon(q)) = \varepsilon d(1, q)$, $\forall q \in G$. The triangle inequality for left-invariant metrics reads: $d(1, ab) \le d(1, a) + d(1, b)$, therefore

$$\begin{aligned} d(1, e^{-x_1} e^{-z} e^{x_2}) &\le d(1, e^{(\varepsilon-1)x_1}) + \varepsilon(2 - \nu) + d(1, e^{(1-\varepsilon)x_2}) \\ &= (1 - \varepsilon) + \varepsilon(2 - \nu) + (1 - \varepsilon) = 2 - \varepsilon\nu. \end{aligned}$$

Now we would like to compensate the deviation of the endpoint 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 $y_i \in g_{k-1}$. Each piece costs $O(\varepsilon^{\frac{k}{k-1}})$ of the distance since $e^{\varepsilon^k y_i} = \delta_{\varepsilon^{\frac{k}{k-1}}}(e^{y_i})$. Hence the distance between the endpoints of the resulting curve remains smaller than 2 if ε is small enough.

It is actually sufficient to insert three pieces as follows:

We are looking for y_1, y_2, y_3 such that

$$e^{x_1} e^{\varepsilon^k y_1} e^{-x_1} e^{-\varepsilon^k z} e^{\frac{1}{2}x_2} e^{\varepsilon^k y_2} e^{\frac{1}{2}x_2} e^{\varepsilon^k y_3} = e^{x_2}$$

for all $\varepsilon > 0$. To find them we use the fact that $e^{-\varepsilon^k z}$ commutes with all elements of the group and re-write the last equation in the form:

$$\left(e^{x_1} e^{\varepsilon^k y_1} e^{-x_1} \right) \left(e^{\frac{1}{2}x_2} e^{\varepsilon^k y_2} e^{-\frac{1}{2}x_2} \right) \left(e^{x_2} e^{\varepsilon^k y_3} e^{-x_2} \right) = e^{\varepsilon^k z}$$

Now we use a universal identity: $e^x e^y e^{-x} = e^{(e^{\operatorname{ad} x} y)}$. Moreover, since $\mathfrak{g}$ is a step k nilpotent Lie algebra and $y_i \in g_{k-1}$, we obtain:

$$e^{\operatorname{ad} x_j} y_i = y_i + \frac{1}{2}[x_j, y_i], \quad i = 1, 2, 3, \ j = 1, 2.$$

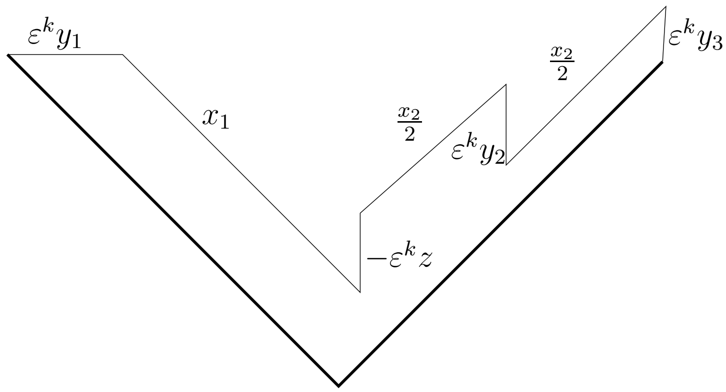


Figure 12.3: Adding more pieces

All elements y_i, $[x_j, y_i]$ are mutually commuting because $k \geq 3$ and $[y_i, y_j] \in g_{2k-2} = 0$. Hence product of the exponents equals the exponent of the sum and we arrive to the equation:

$$e^{\varepsilon^k(\sum_{i=1}^{3} y_i + \frac{1}{2}[x_1,y_1] + \frac{1}{4}[x_2,y_2] + \frac{1}{2}[x_2,y_3])} = e^{\varepsilon^k z}$$

that is equivalent to the system

$$\sum_{i=1}^{3} y_i = 0, \qquad [x_1, y_1] + \frac{1}{2}[x_2, y_2] + [x_2, y_3] = 2z.$$

We insert $y_3 = -y_1 - y_2$ in the second equation and obtain:

$$[x_1 - x_2, y_1] - \frac{1}{2}[x_2, y_2] = 2z.$$

Existence of the desired y_1, y_2 now follows from the relations:

$$g_1 = span\{x_1, x_2\} = span\{x_1 - x_2, x_2\}, \qquad [g_1, g_{k-1}] = g_k \ni z.$$

□

Now we return to the beginning of the proof of Theorem 1 and consider the case of a cusp: $\dot\gamma_1(0) = \dot\gamma_2(0)$. In this case, there exists a horizontal field f_1 and smooth control $t \mapsto u(t)$ such that

$$\dot\gamma_t(t) = f_1(\gamma_1(t)), \quad \dot\gamma_2(t) = f_1(\gamma_2(t)) + t f_{u(t)}(\gamma_2(t)).$$

If the concatenation of the curves $t \mapsto \gamma_1(\varepsilon - t)$ and $t \mapsto \gamma_2(t)$, $0 \leq t \leq \varepsilon$, is a shortest path then $d(\gamma_1(t), \gamma_2(t)) = 2t$. We apply the blow-up procedure and lift to the Carnot group as in steps **1, 2** of the proof in the no cusp case and obtain that $\hat d\left(e^{t\hat f_1}, \overrightarrow{exp} \int_0^t \hat f_1 + \tau \hat f_{u(\tau)}\, d\tau\right) = 2t$. We have:

$$\hat d\left(e^{t\hat f_1}, \overrightarrow{exp} \int_0^t \hat f_1 + \tau \hat f_{u(\tau)}\, d\tau\right) = \hat d\left(1, e^{-t\hat f_1} \overrightarrow{exp} \int_0^t \hat f_1 + \tau \hat f_{u(\tau)}\, d\tau\right)$$

since $\hat{d}$ is a left-invariant metric. Moreover,

$$e^{-t\hat{f}_1}\overrightarrow{exp}\int_0^t \hat{f}_1 + \tau\hat{f}_{u(\tau)}\,d\tau = \overrightarrow{exp}\int_0^t g_\tau^t\,d\tau,$$

where $g_\tau^t = \tau e^{(t-\tau)\mathrm{ad}\hat{f}_1}\hat{f}_{u(\tau)}$, according to the variations formula (see Chapter 6). If the Carnot group is of step k, then:

$$g_\tau^t = \sum_{i=0}^{k-1}\frac{\tau(t-\tau)^i}{i!}(\mathrm{ad}\hat{f}_1)^i\hat{f}_{u(\tau)}.$$

The i-th term of the sum belongs to the $(i+1)$-th level of the Carnot algebra and has order t^{i+1} as $t \to 0$.

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 started at the origin solution of the ordinary differential equation whose right-hand side has order t^i as $t \to 0$.

The ball-box estimates imply that $\hat{d}\left(1, \overrightarrow{exp}\int_0^t g_\tau^t\,d\tau\right) \leq Ct^{\frac{k}{k+1}}$ for some constant C. The obtained contradition completes the proof of the theorem. $\square$

Chapter 13

Some model spaces

In this chapter we are going to construct explicitly the full set of optimal arclength geodesics starting from a point for certain relevant sub-Riemannian structures. This is what is called the problem of constructing the optimal synthesis.

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 explicitly computes 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 techniques to identify the cut locus, that generalize 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 exercise. 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 distribution. 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 complete as metric space. Let γ be an archlength geodesic. The cut time along γ is

$$t_{cut} := \sup\{t > 0 : \gamma|_{[0,t]} \text{ is length-minimizing}\}.$$

If $t_{cut} < +\infty$ we say that $\gamma(t_{cut})$ is the cut point of $\gamma(0)$ along γ. If $t_{cut} = +\infty$ we say that γ has no cut point. We denote by 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 groups 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

$$\mathfrak{g}_1 = \operatorname{span}\{X_1, \ldots, X_k\}, \qquad \mathfrak{g}_2 = \operatorname{span}\{Z_1, \ldots, Z_{n-k}\},$$

where $\{X_1, \ldots, X_k\}$ 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

$$\begin{cases} [X_i, X_j] = \sum_{\ell=1}^{n-k} c_{ij}^{\ell} Z_\ell, & i, j = 1, \ldots, k, \quad \text{with} \quad c_{ij}^{\ell} = -c_{ji}^{\ell}, \\ [X_i, Z_j] = [Z_j, Z_\ell] = 0, & i = 1, \ldots, k, \qquad j, \ell = 1, \ldots, n-k. \end{cases} \tag{13.2}$$

Given an adapted basis, we can introduce the family of skew-symmetric matrices $\{C_1, \ldots, C_{n-k}\}$ encoding the structure constants of the Lie algebra, defined by $C_\ell = (c_{ij}^{\ell})$, for $\ell = 1, \ldots, n-k$, and the corresponding the subspace of skew-symmetric operators on $\mathfrak{g}_1$ that are represented by linear combination of this family of matrices

$$\mathcal{C} := \operatorname{span}\{C_1, \ldots, C_{n-k}\} \subset \mathfrak{so}(\mathfrak{g}_1) \tag{13.3}$$

We stress that since the vector fields of the basis are left-invariant, then c_{ij}^{ℓ} 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 $\ell = 1, \ldots, n-k$, defines a basis of $\mathcal{C}$.

A representation of the Lie algebra defined above is given by the family of vector fields on $\mathbb{R}^n = \mathbb{R}^k \oplus \mathbb{R}^{n-k}$ (using coordinates $g = (x, y) \in \mathbb{R}^k \oplus \mathbb{R}^{n-k}$)

$$X_i = \frac{\partial}{\partial x_i} - \frac{1}{2} \sum_{j=1}^{k} \sum_{\ell=1}^{n-k} c_{ij}^\ell x_j \frac{\partial}{\partial z_\ell}, \qquad i = 1, \ldots, k, \tag{13.4}$$

$$Z_\ell = \frac{\partial}{\partial z_\ell}, \qquad \ell = 1, \ldots, n-k. \tag{13.5}$$

The group law on G, when identified with $\mathbb{R}^n = \mathbb{R}^k \oplus \mathbb{R}^{n-k}$, reads as follows

$$(x, y) * (x', y') = \left(x + x', z + z' + \frac{1}{2} Cx \cdot x' \right),$$

where we denoted for the $(n-k)$-tuple $C = (C_1, \ldots, C_{n-k})$ of $k \times k$ matrices, the product

$$Cx \cdot x' = (C_1 x \cdot x', \ldots, C_{n-k} x \cdot x') \in \mathbb{R}^{n-k}.$$

and $a \cdot b$ denotes here the Euclidean inner product between two vectors $a, b \in \mathbb{R}^k$. The choice of the linearly independent vector fields $\{X_1, \ldots, X_k, Z_1, \ldots, Z_{n-k}\}$ 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:

$$H = \frac{1}{2} \sum_{i=1}^{k} h_i^2. \tag{13.6}$$

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

$$\begin{cases} \dot x_i = h_i \\ \dot z_\ell = -\frac{1}{2} \sum_{i,j=1}^{k} c_{ij}^\ell h_i x_j \end{cases} \qquad \begin{cases} \dot h_i = -\sum_{\ell=1}^{n-k} \sum_{j=1}^{k} c_{ij}^\ell h_j w_\ell \\ \dot w_\ell = 0 \end{cases} \tag{13.7}$$

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

$$\begin{cases} \dot h_i = \{H, h_i\} = -\sum_{j=1}^{k} \{h_i, h_j\} h_j = -\sum_{\ell=1}^{n-k} \sum_{j=1}^{k} c_{ij}^\ell h_j w_\ell \\ \dot w_\ell = \{H, w_\ell\} = 0. \end{cases} \tag{13.8}$$

Recall moreover that H is constant along solutions, in particulat $H = 1/2$ along extremals parametrized by arclength. From (13.8) we easily get that w_ℓ is constant for every $\ell = 1, \ldots, n-k$, hence the first equation rewrites as an autonomous linear equation for $h = (h_1, \ldots, h_k) \in \mathbb{R}^k$

$$\dot h = -\left(\sum_{\ell=1}^{n-k} w_\ell C_\ell \right) h,$$

It follows that

$$h(t) = e^{-t\Omega_w} h(0), \qquad \Omega_w := \sum_{\ell=1}^{n-k} w_\ell C_\ell. \tag{13.9}$$

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 α_i are strictly positive. This is equivalent to require that the structure is of contact type.

According to this decomposition we denote by $\{X_1, \dots, X_l, Y_1, \dots, Y_l, Z\}$ the orthonormal basis of $\mathfrak{g}_1$, where the vector fields satisfy the relations

$$\begin{cases} \mathfrak{g}_1 = \mathrm{span}\{X_1, \dots, X_l, Y_1, \dots, Y_l\}, & \\ [X_i, Y_i] = \alpha_i Z, & i = 1, \dots, l, \\ [X_i, Y_j] = 0, & i \neq j, \\ [X_i, Z] = [Y_i, Z] = 0, & i = 1, \dots, l, \end{cases} \tag{13.13}$$

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

$$\begin{aligned} X_i &= \partial_{x_i} - \frac{1}{2} \alpha_i y_i \partial_z, \qquad i = 1, \dots, l, \\ Y_i &= \partial_{y_i} + \frac{1}{2} \alpha_i x_i \partial_z, \qquad i = 1, \dots, l, \\ Z &= \partial_z. \end{aligned} \tag{13.15}$$

where $x = (x_1, \dots, x_l), y = (y_1, \dots, y_l) \in \mathbb{R}^l$ 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 geodesics 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 $(x_1, \dots, x_n, y_1, \dots, y_n)$.

$$\ell(\gamma) = \int_0^T \left(\sum_{i=1}^{l} (u_i^2(t) + v_i^2(t)) \right)^{1/2} dt.$$

Now we solve (13.16) with initial conditions (corresponding to arclength parametrized trajectories starting from the origin)

$$(x^0, y^0, z^0) = (0, 0, 0), \tag{13.17}$$
$$(u^0, v^0, w^0) = (u_1^0, \dots, u_l^0, v_1^0, \dots, v_l^0, w^0) \in S^{2l-1} \times \mathbb{R}. \tag{13.18}$$

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 the following version of polar coordinates, using the following change of variables

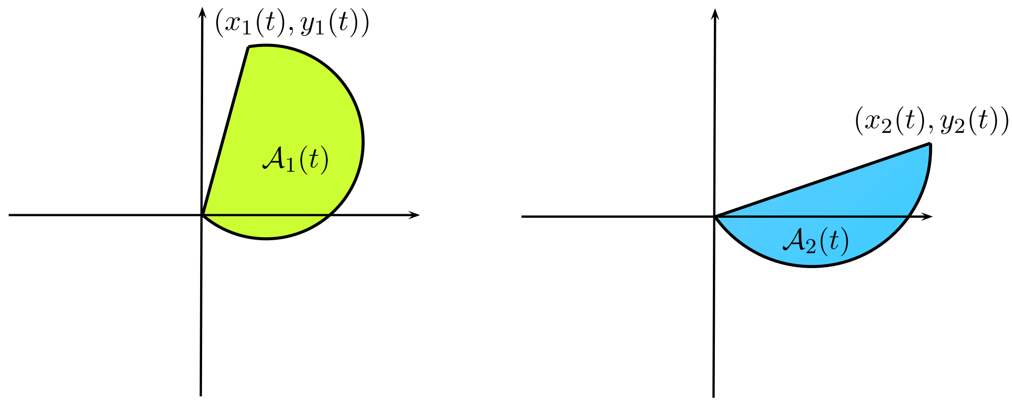
$$u_i^0 = -r_i \sin\theta_i, \qquad v_i^0 = r_i \cos\theta_i, \qquad i = 1, \dots, l. \tag{13.21}$$

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 $r = (r_1, \dots, r_l) \in S^l$. 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 ρ_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 α_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

We start the analysis of the optimal synthesis with the following general lemma. Recall that here we assume $\alpha_i > 0$ for every $i = 1, \dots, l$.

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 γ 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 8.71, 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 assumption, $\alpha_1 = \max_i \alpha_i$. From (13.22) it is easily seen that projection on the (x_1, y_1)-plane of the trajectory γ 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 critical time for exponential map, hence 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 2π) by relations (13.24). Finally, from the assumption that $\widetilde{\gamma}$ also reach optimally the point $\widetilde{\gamma}(t_0)$, 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 $\widetilde{w} > w$ (the case $\widetilde{w} < w$ being analogous). 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 frequences are positive there are no abnormal extremals, Theorem 8.71 and Corollary 8.73 permits to conclude that $\gamma(t_0)$ is not a cut point. $\square$

The next proposition computes 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). $\square$

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 Cut_0.

Generalize the previous formulas to all other cases for $0 = \alpha_l \leq \ldots \leq \alpha_l$, and compute the dimension of 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 the Carnot group of step 2 is *free* if the matrices $C_1, \ldots, C_{n-k}$ define a basis of the space of skew-symmetric matrices. In particular $n = k + \frac{k(k-1)}{2}$ and it is convenient to treat $\mathbb{R}^n = \mathbb{R}^k \oplus \mathbb{R}^{n-k}$ as the sum

$$\mathbb{R}^n = \mathbb{R}^k \oplus (\mathbb{R}^k \wedge \mathbb{R}^k).$$

In what follows we denote by $\mathbb{G}_k := \mathbb{R}^k \oplus \wedge^2 \mathbb{R}^k$ the free Carnot groups of step 2 and we identify $\wedge^2 \mathbb{R}^k$ with the vector space of skew-symmetric real matrices, that is $v \wedge w = vw^* - wv^*$ for $v, w \in \mathbb{R}^k$.

It is convenient to employ the following notation: we denote points $(x, Z) \in \mathbb{G}_k$, where $x \in \mathbb{R}^k$ and Z is a skew-symmetric matrix. We fix the canonical basis $\{E_{\ell m j}\}_{1\leq \ell<m\leq k}$ of $\mathfrak{so}(\mathbb{R}^k)$ and we write $Z = \sum_{\ell<m} Z_{\ell m}E_{\ell m}$.

As discussed in Section 13.1 we can can choose a suitable basis in such a way that the sub-Riemannian structure is generated by the set of global orthonormal vector fields:

$$X_i := \partial_{x_i} - \frac{1}{2}\sum_{1\leq \ell<m\leq k}(e_i \wedge x)_{\ell m}\partial_{Z_{\ell m}}, \qquad i = 1, \dots, k, \tag{13.33}$$

where $\{e_1, \dots, e_k\}$ is the standard basis of $\mathbb{R}^k$. More precisely, the horizontal distribution is defined by $\mathcal{D} := \mathrm{span}\{X_1, \dots, X_k\}$ 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 k generators:

$$\mathfrak{g} = \mathfrak{g}_1 \oplus \mathfrak{g}_2, \qquad \text{where} \qquad \mathfrak{g}_1 = \mathrm{span}\{X_1, \dots, X_k\}, \quad \mathfrak{g}_2 = \mathrm{span}\{\partial_{Z_{ij}}\}_{i<j}. \tag{13.34}$$

There Lie group structure on $\mathbb{G}_k$ 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 $C_1, \dots, C_{n-k}$ coincide in this case with the standard basis of $\mathfrak{so}(k)$ hence the matrix Ω_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 $\mathfrak{so}(k)$

$$\Omega_w = \sum_{1\leq \ell<m\leq k} w_{\ell m}C_{\ell m} = \sum_{1\leq \ell<m\leq k} w_{\ell m}E_{\ell m}.$$

For this reason in what follows we drop the w from the notation and simply write Ω for Ω_w.

Example 13.12. The case $k = 2$ 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 $k = 3$ 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 3×3 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 $V = \{(0, Z) \mid Z \in \wedge^2\mathbb{R}^k\}$. In particular we give the explicit formula of the distance from the origin to every point of V.

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

$$h(t) = e^{-t\Omega}h(0), \qquad \Omega \in \mathfrak{so}(k).$$

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

$$\dot{Z}(t) = \frac{1}{2}x(t) \wedge h(t). \tag{13.39}$$

that gives the following formula (recall $Z(0) = 0$)

$$Z(t) = \frac{1}{2}\int_0^1 \int_0^t e^{-s\Omega}h(0) \wedge e^{-t\Omega}h(0)dsdt, \tag{13.40}$$

$$= \frac{1}{2}\int_0^1 \int_0^t (e^{-s\Omega}Pe^{t\Omega} - e^{-t\Omega}Pe^{-s\Omega})dsdt. \tag{13.41}$$

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 Ω is written in normal form. The main linear algebra ingredient is given by the following lemma.

Lemma 13.14. *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 Ω is skew-symmetric there exists $\alpha_1, \ldots, \alpha_r$ 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 Ω 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 $e^{t\Omega}x_0 = (e^{t\Omega_1}x_{0,1}, \ldots, e^{t\Omega_r}x_{0,r}, 0)$.

For every $j > 0$ such that $x_{0,j} \neq 0$ we the corresponding block Ω_j can be put to zero without changing the value of $e^{t\Omega}x_0$.

If there exists a block with multiple eigenvalues (i.e., there exists $j > 0$ such that $\dim E_j > 2$) then, thanks to Exercice 13.15 we have $\dim\mathrm{span}\{e^{t\Omega_j}x_{0,j} \mid t \in \mathbb{R}\} = \dim\mathrm{span}\{x_0, \Omega x_0\} = 2$, thus we can write

$$E_j = \mathrm{span}\{x_{0,j}, \Omega_j x_{0,j}\} \oplus \mathrm{span}\{x_{0,j}, \Omega_j x_{0,j}\}^{\perp}. \tag{13.42}$$

Choosing a basis in E_j corresponding to the splitting (13.42), we can put to zero the block of Ω_j corresponding to $\mathrm{span}\{x_{0,j}, \Omega_j x_{0,j}\}^\perp$ and the new matrix has $\pm i\alpha_j$ as simple eigenvalues, and kernel of dimension $\dim(E_j) - 2$. This proves the existence of the matrix $\overline{\Omega}$. □

Exercise 13.15. Let $\Omega \in \mathfrak{so}(n)$ and assume $\mathrm{spec}(\Omega) = \{\pm i\alpha\}$. Then for $x_0 \in \mathbb{R}^n$

$$\mathrm{span}\{e^{t\Omega}x_0 \mid t \in \mathbb{R}\} = \mathrm{span}\{x_0, \Omega x_0\}.$$

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 Ω is presented as a block-diagonal matrix

$$\Omega = (\Omega_1, \ldots, \Omega_\ell, \mathbb{O}),$$

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 2×2 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 $\mathbb{G}_k$ is contained in the cut locus.

Lemma 13.16. *The set of points* $\{(0, Z) \mid Z \in \wedge^2\mathbb{R}^k \setminus \{0\}\}$ *is contained in* Cut_0.

Proof. Fix a point $(0, Z) \in \mathbb{G}_k$ with $Z \neq 0$. Thanks to Exercice13.17 there exists a non zero orthogonal matrix $M \in SO(k)$ 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 exists (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 γ and $\overline{\gamma}$ are two horizontal length-minimizers joining the same end-points. This proves the claim. □

Exercise 13.17. Let $Z \in \mathfrak{so}(k)$ be a non zero skew-symmetric matrix.

(a). Prove that there exists an orthogonal matrix $M \in SO(k), M \neq \mathbb{I}$, 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.

We then compute the distance from the origin of vertical points in $\mathbb{G}_k$. A very close formula appears as the second statement of [36, Thm. 2], and differs from ours by a factor 4π.

Proposition 13.18. *Let $(0, Z) \in \mathbb{G}_k$, 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,*

$$d((0,0),(0,Z))^2 = 4\pi \sum_{i=1}^{r} i\alpha_i. \tag{13.43}$$

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 Ω are of the form $\pm i2\pi\phi$, with $\phi \in \mathbb{N}$. By Lemma 13.14, and up to an orthogonal transformation, we may assume that $\Omega = (2\pi\phi_1\mathbb{J}, \dots, 2\pi\phi_\ell\mathbb{J}, 0_{k-2\ell})$, with all simple eigenvalues, $2\ell = \mathrm{rank}\,(\Omega)$, and with distinct $\phi_i \in \mathbb{N}$. We split accordingly $h_0 = (h_{0,1}, \dots, h_{0,\ell}, h_{0,0})$, with $h_{0,i} \in \mathbb{R}^2$ for $i = 1, \dots, \ell$ and $h_{0,0} \in \mathbb{R}^{k-2\ell}$. 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

$$Z(1) = \left(\frac{|h_{0,1}|^2}{4\pi\phi_1}\mathbb{J}, \dots, \frac{|h_{0,\ell}|^2}{4\pi\phi_\ell}\mathbb{J}, 0_{k-2\ell} \right). \tag{13.45}$$

Then $|h_{0,j}|^2 = 4\pi\phi_j\alpha_j$ for all $j = 1, \dots, r$. The squared length of γ 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 $j = 1, \dots, r$. ☐

For more details we refer to [?] (see also [36]).

13.4 An extended Hadamard technique to compute the cut locus

Let us consider a sub-Riemannian structure, complete as metric space and fix $q_0 \in M$. Assume that we are able to solve the problems (A) and (B) above. 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 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 has nowhere vanishing Jacobian then it is a covering.

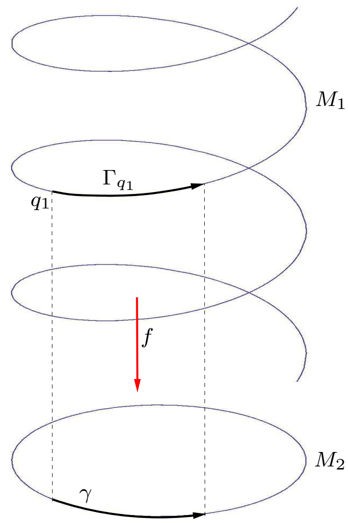


Figure 13.2: Uniqueness of the lift for a covering map.

Definition 13.19. 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 ∂M_1 is mapped in ∂M_2 through f.

Definition 13.20. 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 number of preimages 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 Γ_{q_1} is called the lift of γ (through q_1).

P3: Consider two homotopic loop $\gamma, \gamma' : [0,1] \to M_2$ and a point q_1 in M_1 such that $f(q_1) = \gamma(0) = \gamma'(0)$. Let Γ_{q_1} and Γ'_{q_1} the corresponding lift. Then the final point of Γ_{q_1} and Γ'_{q_1} are the same, namely $\Gamma_{q_1}(1) = \Gamma'_{q_1}(1)$.

Theorem 13.21. *Let M_1 and M_2 two smooth connected differentiable manifolds and $f : M_1 \to M_2$ be smooth. If*

- *f is proper,*
- *the Jacobian of f vanishes nowhere,*

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 13.22 (of Theorem 13.21)**.** *Under the assumptions of Theorem 13.21, if M_2 is simply connected, then f 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 is given by the following result

Corollary 13.23 (of Theorem 13.21)**.** *Under the assumptions of Theorem 13.21, if M_2 is not simply connected, but it is homeomorphic to $S^1 \times N$, where N is simply connected, and we find a loop in M_1 that project via f in a loop in M_2 that is homotopic to S^1, then f is a global diffeomorphism.*

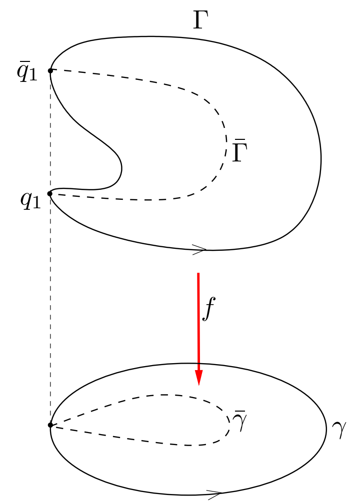


Figure 13.3: Proof of Corollary 13.23

Proof. Assume by contradiction that the number of pre-images of a point is not one. We refer to Figure 13.3. Let $\Gamma : [0,1] \to M_1$ be loop in M_1, $q_1 = \Gamma(0)$ and let γ be the corresponding projection in M_2 as in the statement of the Corollary. Let $\bar{q}_1$ be another preimages of $\gamma(0)$. We are going to prove that $q_1 = \bar{q}_1$.

Consider a continuous curve $\bar{\Gamma} : [0,1] \to M_1$ connecting q_1 and $\bar{q}_1$ (this is possible since M_1 is connected a manifold and hence path connected). Consider its projection on M_2 that is $\bar{\gamma} := f(\bar{\Gamma})$. Because of the topology of M_2, $\bar{\gamma}$ is a loop winded n times around S_1 ($n = 0, 1, 2\ldots$).

If $\bar{\gamma}$ is homotopic to S_1 then it is homotopic to γ. Hence since $\Gamma(0) = \bar{\Gamma}(0) = q_1$ and because of property P3 we have that $\bar{\Gamma}(1) = \Gamma(1)$. As a consequence $q_1 = \bar{q}_1$.

If $\bar{\gamma}$ is winded n times around S_1 with ($n > 1$) then we consider the loop $\Gamma^n : [0,n] \to M_1$ obtained concatenating n times Γ. Let us call γ^n its projection on M_2. We have that $\bar{\gamma}$ is homotopic to γ^n. The same reasoning as before gives again $q_1 = \bar{q}_1$.

If $\bar{\gamma}$ is winded 0 times around S_1 (i.e., if it 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 γ^0 be its projection. Since a covering is a continuous map, the projection of a contractible loop is a contractible loop. Hence γ^0 is contractible and we have that $\bar{\gamma}$ and γ^0 are homotopic. The same reasoning as before gives again $q_1 = \bar{q}_1$.

□

Finding the cut locus via Theorem 13.21 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 n) or not. Recall that if the structure is Riemannian at q_0, then Λ_{q_0} has the topology of S^{n-1} while if the structure has rank $k < n$ at q_0 then Λ_{q_0} has the topology of $S^{k-1} \times \mathbb{R}^{n-k}$.

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 $\overline{\mathrm{Cut}_{q_0}}$ and the conjectured cut times $\overline{t_{cut}(\theta)}$, $\theta \in \Lambda_{q_0}$ (notice that it may happen that $\overline{t_{cut}(\theta)}$ is $+\infty$).

Notice that if $\overline{\mathrm{Cut}_{q_0}}$ has a boundary then the points on the boundary are expected to be conjugate points (since the set $\overline{\mathrm{Cut}_{q_0}}$ 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 $\overline{\mathrm{Cut}_{q_0}}$ must be included in $\overline{\mathrm{Cut}_{q_0}}$.

We have two cases:

- If the structure is Riemannian at q_0 define $N_1 = \{t\,\theta \mid \theta \in \Lambda_{q_0},\ \ t \in [0, \overline{t_{cut}(\theta)})\} \subset T^*_{q_0}M$. 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 $N_1 = \{t\,\theta \mid \theta \in \Lambda_{q_0},\ \ t \in (0, \overline{t_{cut}(\theta)})\}$; 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 $N_2 = M \setminus \overline{\mathrm{Cut}_{q_0}}$. If this is not the case then the conjectured cut locus and cut times were wrong. Indeed if there exists $q \in N_2 \setminus (M \setminus \overline{\mathrm{Cut}_{q_0}})$ then in q is reached by a geodesic at its conjectured cut time and by another geodesic before its conjectured cut time and hence the conjectured cut times was wrong. On the other side if there exists $q \in (M \setminus \overline{\mathrm{Cut}_{q_0}}) \setminus N_2$ then $\exp_{q_0}|_{N_1}$ is not covering M up to the conjectured cut locus.

Remark 13.24. 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.

Remark 13.25. 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 π. We have only two geodesics parametrized by arclength: $q^+(t) = t$ and $q^-(t) = -t$. By symmetry the two geodesics meet at $t = 0, \pi, 2\pi, 3\pi, \ldots$ etc. Assume that we make the (false) conjecture that the cut time is $\overline{t_{cut}} = 3\pi$ (instead than $\overline{t_{cut}} = \pi$). We have $\overline{\mathrm{Cut}_0} = S^1 \setminus \{\pi\}$. In this case **Step 1** fails because $N_2 = S^1 \neq S^1 \setminus \overline{\mathrm{Cut}_0}$.

Step 2 Prove that the Jacobian of $\exp_{q_0}$ vanishes nowhere 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 (and non trivial too) then there are always conjugate points (cf. Remark 8.42). 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 13.24). After having verified that N_2 is simply connected, Corollary 13.22 (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 13.21 permits to prove that $\exp_{q_0}|_{N_1}$ is a covering but one cannot conclude that f is a diffeomorphism using Corollary 13.22 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 13.23. Notice that if $n = 3$ and the structure is not Riemannian at q_0 then N_2 is never simply connected.

Writing $\gamma_\theta(\cdot) = \exp_{q_0}(\cdot, \theta)_{[0,\overline{t_{cut}(\theta)}]}$ the optimal synthesis is then the collection of trajectories

$$\left\{\gamma_\theta(\cdot) \mid \theta \in H^{-1}(1/2)\right\}.$$

Remark 13.26. 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 remove it from N_1. This should be done to satisfy the hypothesis of Theorem 13.21 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, N_2 is simply connected and one obtain directly that the exponential map 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 $\mathbb{R}^2$ for which a global orthonormal frame is given by

$$F_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \qquad F_2 = \begin{pmatrix} 0 \\ x \end{pmatrix}.$$

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

$$H(x, y, p_1, p_2) = \frac{1}{2}(\langle p, F_1\rangle^2 + \langle p, F_2\rangle^2) = \frac{1}{2}(p_1^2 + x^2 p_2^2), \tag{13.47}$$

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 $p_1 = \cos(\theta)$, $p_2 = \sin(\theta)/x_0$, $\theta \in S^1$. 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

$$\begin{cases} x(t,0) = t + x_0, \quad y(t,0) = 0, \\ y(t,\pi) = -t + x_0, \quad y(t,\pi) = 0, \\ \left.\begin{aligned} x(t) &= x_0 \dfrac{\sin(\theta + \frac{t\sin(\theta)}{x_0})}{\sin(\theta)}, \\ y(t) &= x_0 \dfrac{2t + 2x_0\cos(\theta) - x_0 \frac{\sin(2\theta + 2\frac{t\sin(\theta)}{x_0})}{\sin(\theta)}}{4\sin(\theta)} \end{aligned}\right\} \quad \text{if } \theta \notin \{0,\pi\} \end{cases} \tag{13.48}$$

Theorem 13.27. *The cut time for the geodesic $q(\cdot,\theta)$ is*

$$t_{cut}(\theta) = \left| x_0 \frac{\pi}{\sin(\theta)} \right|.$$

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 $q(t_{cut}(\theta),\theta)$ 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 $q(t_{cut}(\theta),\theta)$ is reached exactly by one optimal geodesic for which $t_{cut}(\theta)$ is also a conjugate point.

By direct computation one gets

Corollary 13.28. *The cut locus starting from $(x_0, 0)$ is*

$$\mathrm{Cut}_{x_0} = \{(-x_0, y) \in \mathbb{R}^2 \mid y \in (-\infty, -\frac{\pi}{2}x_0^2] \cup [\frac{\pi}{2}x_0^2, \infty)\}.$$

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 13.27

We are going to apply the extended Hadamard technique to the case in which 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:

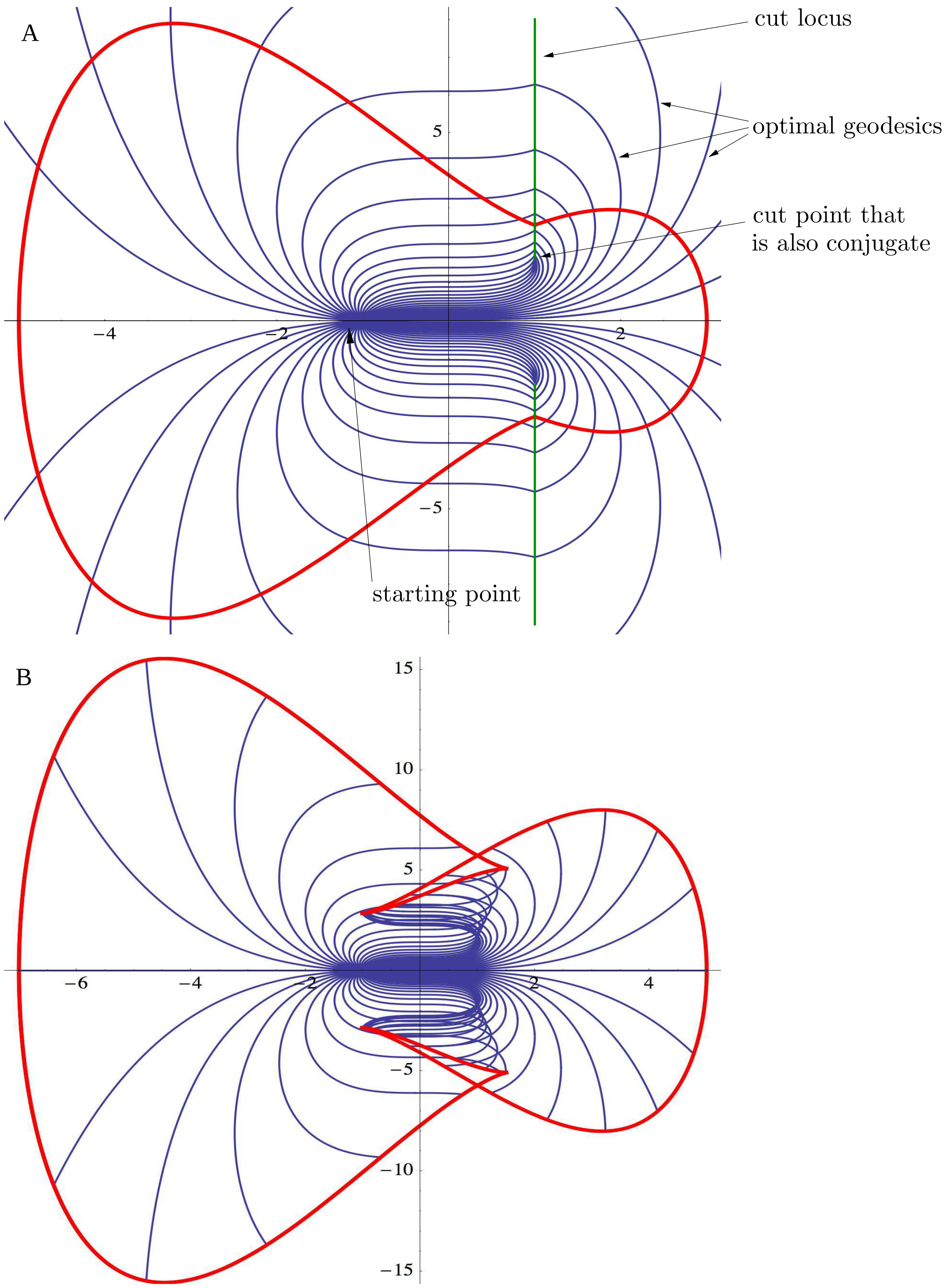


Figure 13.4: A: the optimal synthesis for the Grushin plane starting from the point $(-1, 0)$, together with the sub-Riemannian sphere of radius 4. B: all geodesics up to length 6 with the corresponding wave front.

Lemma 13.29. *For $\theta \neq \{0,\pi\}$, we have*

$$q\left(\left|x_0\frac{\pi}{\sin(\theta)}\right|,\theta,x_0\right) = q\left(\left|x_0\frac{\pi}{\sin(\theta)}\right|,\pi-\theta,x_0\right) = (-x_0,\frac{\pi}{2}x_0^2\frac{1}{\sin(\theta)^2}).$$

Moreover the determinant of the differential of the exponential map is:

$$D(t,\theta,x_0) = \begin{pmatrix} \partial_t x(t,\theta) & \partial_\theta x(t,\theta) \\ \partial_t y(t,\theta) & \partial_\theta y(t,\theta) \end{pmatrix} = \begin{cases} t^2 + \frac{t^3}{3x_0} + tx_0 & \text{if } \theta = 0, \\ -t^2 + \frac{t^3}{3x_0} + tx_0 & \text{if } \theta = \pi, \\ \dfrac{x_0\left(x_0\frac{\sin\left(\frac{t\sin(\theta)}{x_0}\right)}{\sin(\theta)} - t\cos(\theta)\cos\left(\theta+\frac{t\sin(\theta)}{x_0}\right)\right)}{\sin^2(\theta)}, & \text{if } \theta \notin \{0,\pi\}. \end{cases}$$

In particular $D(|x_0\pi|,\pi/2,x_0) = 0$.

We then conjecture that the cut time of the geodesic $q(t,\theta)$ is $\overline{t_{cut}(\theta)} = \left|x_0\frac{\pi}{\sin(\theta)}\right|$ and that the cut locus is

$$\overline{\mathrm{Cut}_{x_0}} = \{(-x_0,y) \in \mathbb{R}^2 \mid y \in (-\infty,-\frac{\pi}{2}x_0^2] \cup [\frac{\pi}{2}x_0^2,\infty)\}.$$

We have then in polar coordinates

$$N_1 = \{(\rho,\theta) \mid \rho < \left|x_0\frac{\pi}{\sin(\theta)}\right|\}.$$

In cartesian coordinates

$$N_1 = \{(p_1,p_2) \in T^*\mathbb{R}^2 : |p_2| < \pi\}.$$

And

$$N_2 = \exp(N_1) = \{(x,y) \in \mathbb{R}^2 \mid (x,y) \notin \overline{\mathrm{Cut}_{x_0}}\}$$

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 13.30. *The geodesic $q(\cdot,\theta)$ has no conjugate points in $[0,\overline{t_{cut}(\theta)})$.*

Proof. Since the zeros of $D(\cdot,\theta,x_0)$ 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

- $D(0,\theta,x_0) = 0$. Notice however that this does not mean that $t = 0$ is a conjugate time. Indeed in x_0 the structure is Riemannian and $D(0,\theta,x_0)$ vanishes only as a consequence of the choice of polar coordinates.
- $D(\overline{t_{cut}(\theta)},\theta,x_0) = \pi x_0^2\frac{\cos^2\theta}{\sin^3\theta}$. This quantity is always larger than zero except for $\theta = \pi/2$ where it is zero.
- $\partial_t D(t,\theta,x_0) = \dfrac{(x_0 + t\cos\theta)\left(\sin(\theta + \frac{t\sin\theta}{x_0})\right)}{\sin\theta}$. Notice that this function is positive in $t = 0$. Let us study when this function is zero in the interval $(0,\overline{t_{cut}(\theta)})$. 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 $(0, \overline{t_{cut}(\theta)})$ when $\theta \in (\bar{\theta}, \pi]$ where $\bar{\theta} = -\arctan(\pi) \simeq 1.88$. One immediately verify that this zero correspond to a minimum of $D(\cdot, \theta, x_0)$ 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 $(0, \overline{t_{cut}(\theta)})$ if and only if $k = 1$. One immediately verify that this zero correspond to a maximum of $D(\cdot, \theta, x_0)$ and that the value of this maximum is positive.

By this analysis it follows that $D(\cdot, \theta, x_0)$ is a function that is zero in zero; it has positive derivative in zero; it is positive at $\overline{t_{cut}(\theta)}$ (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 $D(\cdot, \theta, x_0)$ is never zero in $(0, \overline{t_{cut}(\theta)})$. Since $t = 0$ is not a conjugate point, it follows that there are no conjugate points in $[0, \overline{t_{cut}(\theta)})$.

□

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 $\overline{\mathrm{Cut}_{x_0}}$;
- 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 13.22 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

$$\begin{cases} x^{\pm}(t, 0) = \pm t, \quad y(t, 0) = 0, \\[2ex] \left. x^{\pm}(t) = \pm\dfrac{\sin(at)}{a}, \quad y(t) = \dfrac{2at - \sin(2at)}{4a^2} \right\} \quad \text{if } a \neq 0 \end{cases} \tag{13.49}$$

Theorem 13.31. *The cut time for the geodesic $q^{\pm}(\cdot, a)$ is*

$$t_{cut}(a) = \frac{\pi}{|a|}$$

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

$$\mathrm{Cut}_{(0,0)} = \{(0, y) \in \mathbb{R}^2 \mid y \neq 0\}.$$

and each point of the cut locus is reached exactly by two optimal geodesic.

The optimal synthesis starting from the origin for Grushin plane is depicted in Figure 13.5.

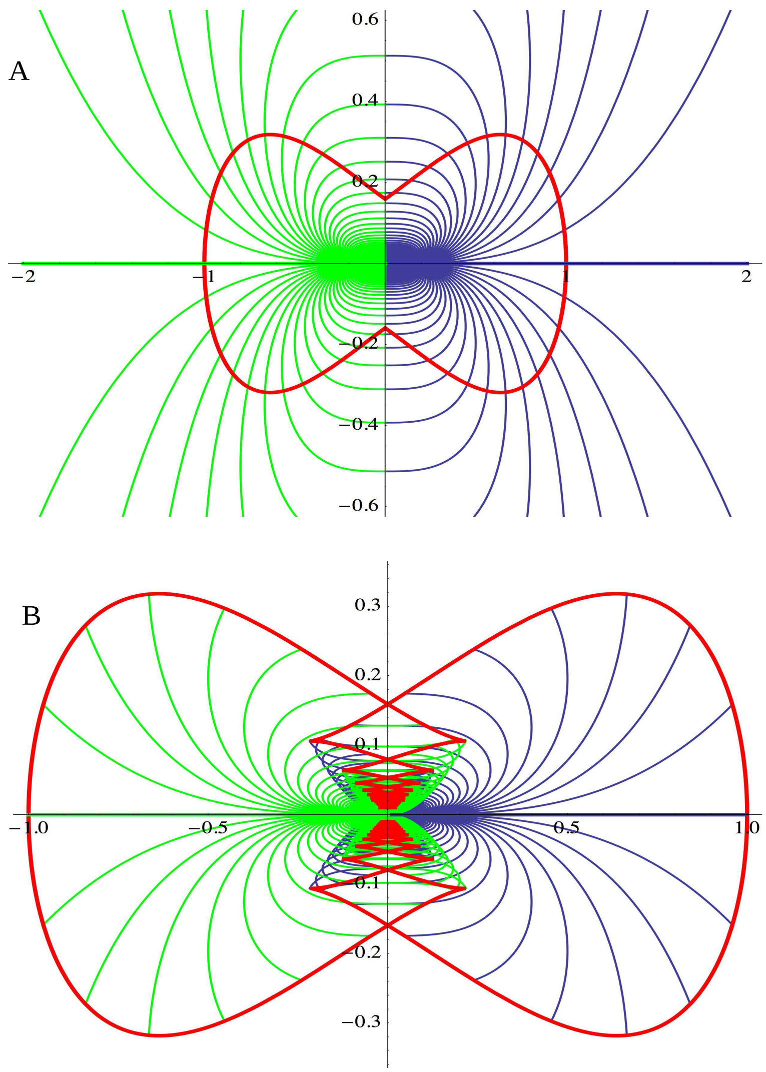


Figure 13.5: A: the optimal synthesis for the Grushin plane starting from the origin, together with the sub-Riemannian sphere for $t = 1$. B: all geodesics up to time 1 with the corresponding wave front.

Proof of Theorem 13.31

We give a proof of Theorem 13.31 by making a direct computation, without using the extended Hadamard technique. See also Exercise 13.32.

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

$$x^+(t,a)=\bar{x} \tag{13.50}$$

$$y(t,a)=\bar{y}. \tag{13.51}$$

Proof of the Claim. Fix a. Let us try to find $t(a)$ from equation (13.50). 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 take 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 take all values from ∞ to $\pi\bar{x}^2/4$.

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 $t_1(a)$.

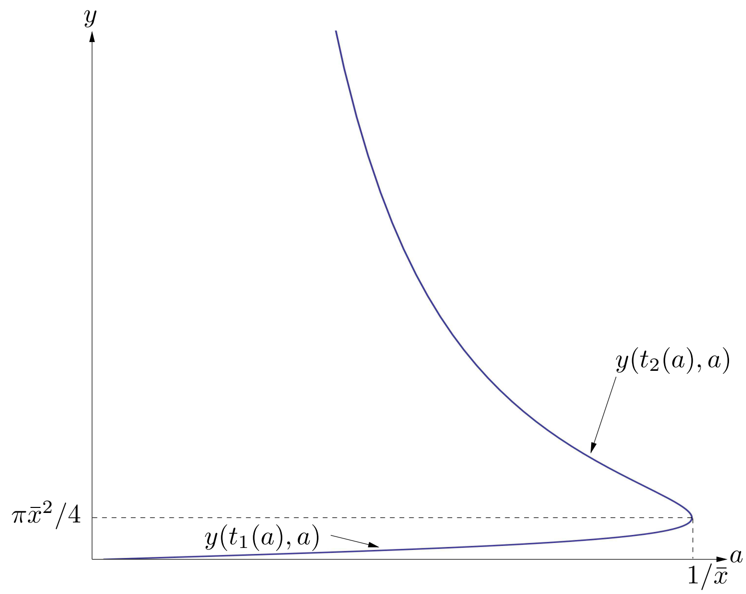


Figure 13.6: Proof of Theorem 13.31.

- 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 13.32. Prove Theorem 13.31 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 2×2 complex matrices

$$SU(2) = \left\{ \begin{pmatrix} \alpha & \beta \\ -\overline{\beta} & \overline{\alpha} \end{pmatrix} \in \mathrm{Mat}(2, \mathbb{C}) \mid |\alpha|^2 + |\beta|^2 = 1 \right\}.$$

The Lie algebra of $SU(2)$ is the algebra of antihermitian traceless 2×2 complex matrices

$$su(2) = \left\{ \begin{pmatrix} i\alpha & \beta \\ -\overline{\beta} & -i\alpha \end{pmatrix} \in \mathrm{Mat}(2, \mathbb{C}) \mid \alpha \in \mathbb{R}, \beta \in \mathbb{C} \right\}.$$

A basis of $su(2)$ is $\{p_1, p_2, k\}$ where

$$p_1 = \frac{1}{2}\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \quad p_2 = \frac{1}{2}\begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix} \quad k = \frac{1}{2}\begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}, \tag{13.52}$$

whose commutation relations are $[p_1, p_2] = k, \quad [p_2, k] = p_1, \quad [k, p_1] = p_2.$

For $su(2)$ 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

$$\langle \cdot \,|\, \cdot \rangle = -\frac{1}{2}\mathrm{Kil}(\cdot,\cdot)$$

is a positive definite bi-invariant metric on $su(2)$ (cf. Section 7.2.3 and Exercice 7.41).

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 $\langle \cdot \,|\, \cdot \rangle\,|_{\mathbf{d}}$ we get a sub-Riemannian structre of the type $\mathbf{d} \oplus \mathbf{s}$ (cf. 7.8.1).

Remark 13.33. 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

$$\left\{ \begin{pmatrix} \alpha \\ 0 \end{pmatrix} \mid |\alpha|^2 = 1 \right\}.$$

Expression of geodesics

Let us write an initial covector in $su(2)$ 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 (7.44), 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 13.34. 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 13.35. *The cut time for the geodesic $g(\theta, c, t)$ coincides with its first conjugate time. It is independent from θ and it is given by the formula*

$$t_{cut}(c) = \frac{2\pi}{\sqrt{1+c^2}}.$$

Moreover $g(\theta, c; t_{cut}(c))$ is independent from θ. Hence each cut point is reached by an infinite number of geodesics (a one parameter family parameterized by θ).

Since the largest cut time is obtained for $c = 0$ we have

Corollary 13.36. *The diameter of $SU(2)$ with the standard sub-Riemannian structure is 2π.*

By a direct computation one gets

Corollary 13.37. *The cut locus starting from the identity is*

$$\mathrm{Cut}_{\mathrm{id}} = e^{\mathbb{R}k} \setminus \{id\} = \left\{ \begin{pmatrix} \alpha \\ 0 \end{pmatrix} \mid |\alpha|^2 = 1, \alpha \neq 0 \right\}.$$

Moreover each cut point is also a conjugate point.

Remark 13.38. Notice that with our definition of cut locus, the starting point is never a cut point.

Proof of Theorem 13.35. 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 obtain:

Lemma 13.39. *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

$$\sin\left(\sqrt{1+c^2}\frac{t}{2}\right)\left(2\sin\left(\sqrt{1+c^2}\frac{t}{2}\right) - \sqrt{1+c^2}\,t\cos\left(\sqrt{1+c^2}\frac{t}{2}\right)\right) = 0. \tag{13.53}$$

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 $\overline{t_{cut}(c)} = \frac{2\pi}{\sqrt{1+c^2}}$ and that the cut locus is

$$\overline{\mathrm{Cut}_{\mathrm{id}}} = e^{\mathbb{R}k} = \left\{ \begin{pmatrix} \alpha \\ 0 \end{pmatrix} \mid |\alpha|^2 = 1, \alpha \neq 0 \right\}.$$

We define

$$N_1 = \{ap_1 + bp_2 + ck \in su(2) \mid (a, b) \neq (0, 0), |c| \leq \sqrt{2\pi - 1}\}$$

and

$$N_2 = \exp(N_1) = \{g \in SU(2) \mid g \notin \overline{\text{Cut}_{\text{Id}}}\}$$

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 $\exp_{\text{Id}}|_{N_1}$. Conjugate times are given by formula (13.53). 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 $\overline{t_{cut}(c)}$, 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 ∂N_1 are mapped in points of ∂N_2.

Step 4 (SR) By Theorem 13.21 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 13.22 since N_2 is not simply connected. Let us show that the hypotheses of Corollary 13.23 are verified. The topology of N_2 is those of $S^1 \times \mathbb{R}^2$. 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 ℓ_2 winding once the cut locus is the image through the exponential map of a loop ℓ_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 $\overline{\text{Cut}_{\text{id}}}$ is the image through the exponential map of a loop ℓ_1. Then Corollary 13.23 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 13.40. The argument above apply to any 3 dimensional structure that is genuinely sub-Riemannian at the starting point.

Exercise 13.41. Corollary 13.36 says that the diameter of $SU(2)$ for the standard sub-Riemannian structure is 2π. 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 2π as well.

A representation of the cut locus for $SU(2)$ is given in Figure 13.7.

Exercise 13.42. Consider the $\mathbf{d} \oplus \mathbf{s}$ sub-Riemannian structure on $SO(3)$ introduced in Section 7.8.2. 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 13.43. 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.

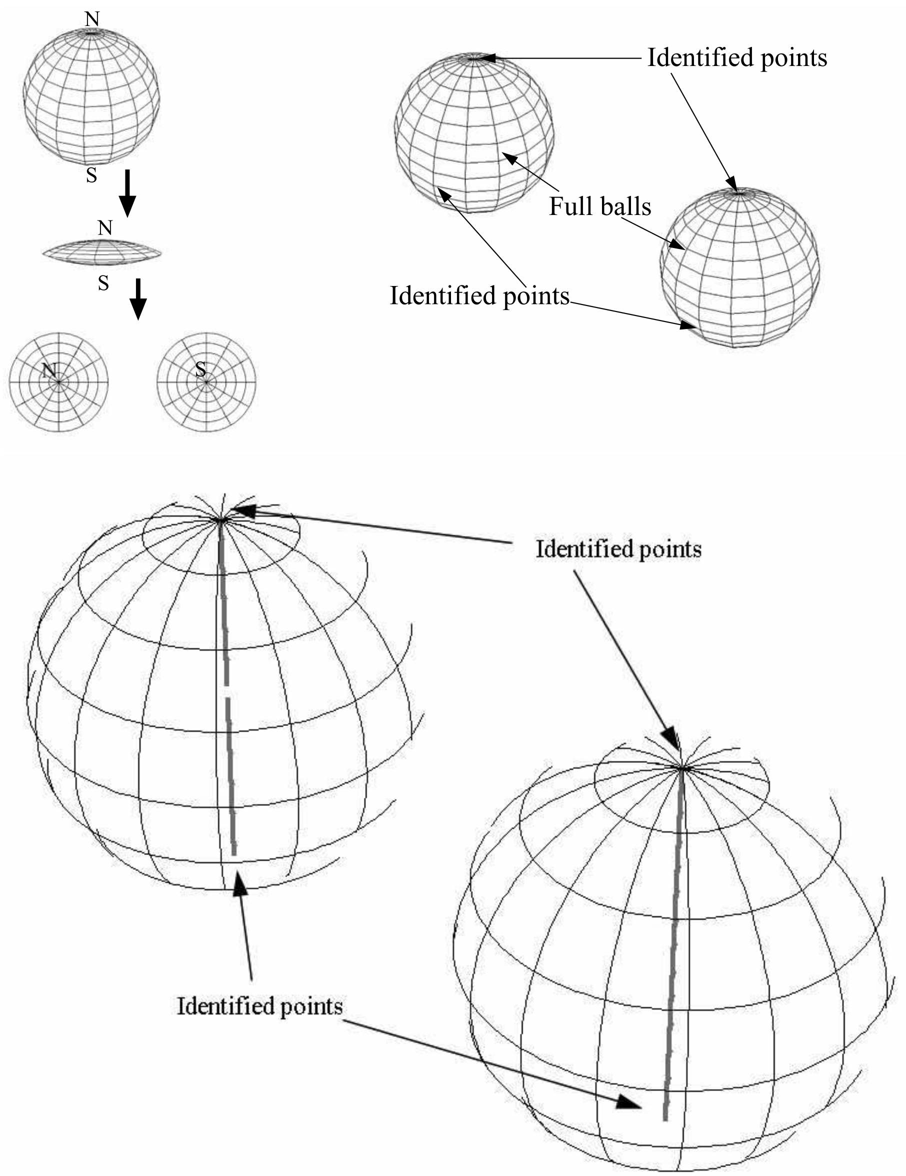


Figure 13.7: We recall a standard construction for representing S^2 in a two dimensional space and S^3 in a three dimensional one. Consider $S^2 \subset \mathbb{R}^3$ and flatten it on the equator plane, pushing the northern hemisphere down and the southern hemisphere up, getting two disks D^2 joined 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 the the great circle passing through the identity, the north and the south pole (the identity should then be removed, cf. Remark 13.2).

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 7.8.3. 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 7.8.3).

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 [24, 23, 25] but we advise the reader to recover them by him(her)self.

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 a 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 θ_0 be the angle between $\dot\gamma(0)$ and v_0 Then the parallel transport of v_0 along γ 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 a constant geodesic curvature (recall that angles are defined modulo 2π). 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 ξ be the shortest path connecting $\gamma(1)$ with $\gamma(0)$ and ϕ 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 curvature and the sign of ϕ.

To compute the parallel transport along the curve γ 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

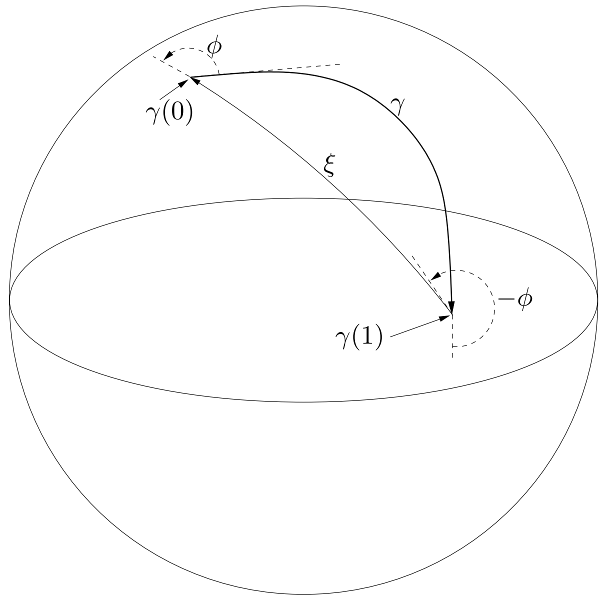


Figure 13.8: Construction of the optimal synthesis on $SO(3)$ and $SO_+(2,1)$. Definition of the angle ϕ. (The picture refers to $SO(3)$)

of $\dot\gamma(0)$ is $-\phi$ and the direction of $\dot\gamma(1)$ is $+\phi$. Hence the parallel transport of $\dot\xi(1)$ along γ 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 both curves coincide if and only if

$$2(\rho\ell + 2\phi) \equiv 0 \bmod 2\pi.$$

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 (1.27) applied to Γ gives a relation:

$$\rho\ell + 2\phi \pm \text{Area}(\Gamma) = 2\pi.$$

If M is the unit sphere, then $\rho\ell + 2\phi = 2\pi - \text{Area}(\Gamma)$. The case $\rho\ell + 2\phi = \pi$ is a natural candidate to cut. If M is the Lobachevsky plane, then $\rho\ell + 2\phi = 2\pi + \text{Area}(\Gamma)$ and a natural candidate to cut is the case $\rho\ell + 2\phi = 3\pi$. Both cases are characterized by the identity:

$$\text{Area}(\Gamma) = \pi.$$

We are now ready to describe the optimal synthesis. Let M be either unit sphere in the three-dimensional Euclidean space or hyperbolic plane in the Minkowsky space.

1. Geodesics are parallel transports along curves of a constant geodesic curvature in M, and curves of a constant geodesic curvature are just the intersections of $M \subset \mathbb{R}^3$ with affine planes.

2. Let $t \mapsto \gamma(t)$ is a parameterized curve of a constant geodesic curvature in M and $\Gamma_t \subset M$ be the smaller domain among two domains whose boundary is the concatenation of $\gamma|_{[0,t]}$ and the shortest path connecting $\gamma(t)$ with $\gamma(0)$. We assume that γ is oriented in such a way that Γ_t stays to the right from γ (as in the figure). The cut time t_γ for the parallel transport along γ is as follows:
$$t_\gamma = \min\{t > 0 : \gamma(t) = \gamma(0) \text{ or } \mathrm{Area}(\Gamma_t) = \pi\}.$$

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 $\mathrm{Area}(\Gamma_t) = \pi$ 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 vector $(y_1, y_2, 1)^t$ then the action of a matrix of $SE(2)$ produces a rotation of angle θ and a translation of (x_1, x_2) (cf. Section 7.2.2). The Lie algebra of $SE(2)$ is

$$\mathfrak{se}(2) = \mathrm{span}\{e_1, e_2, e_r\},$$

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:

$$[e_1, e_2] = 0, \quad [e_1, e_r] = -e_2, \quad [e_2, e_r] = e_1. \tag{13.54}$$

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 $SE(2)$),

$$\dot{g} = g(ue_1 + ve_r), \tag{13.55}$$

$$\int_0^T \sqrt{u(t)^2 + v(t)^2}\, dt, \to \min, \tag{13.56}$$

$$g(0) = g_0, \quad g(T) = g_1. \tag{13.57}$$

Notice that since we are in dimension 3 and with one bracket one get 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 become

$$\dot{q} = uX_1(q) + vX_r(q), \tag{13.58}$$

$$\int_0^T \sqrt{u(t)^2 + v(t)^2}\, dt \to \min, \tag{13.59}$$

$$q(0) = q_0, \quad q(T) = q_1. \tag{13.60}$$

where

$$X_1 = \begin{pmatrix} \cos(\theta) \\ \sin(\theta) \\ 0 \end{pmatrix}, \quad X_r = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}. \tag{13.61}$$

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 (13.54) i.e., $[X_1, X_2] = 0$, $[X_1, X_r] = -X_2$ and $[X_2, X_r] = X_1$.

Exercise 13.44. 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 θ with respect to the x_1 axis (see Figure 13.9 (A)). Then the optimal control problem (13.58)-(13.61) 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, θ 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 (13.58), (13.59), (13.60), (13.61) is

$$H(q,p) = \frac{1}{2}\left(\langle p, X_1\rangle^2 + \langle p, X_2\rangle^2\right).$$

Setting $p = (p_1, p_2, p_\theta)$, $p_1 = P\cos(p_a)$, $p_2 = P\sin(p_a)$ we have

$$H = \frac{1}{2}\left((p_1\cos\theta + p_2\sin\theta)^2 + p_\theta^2\right) = \frac{1}{2}\left(P^2\cos^2(\theta - p_a) + p_\theta^2\right).$$

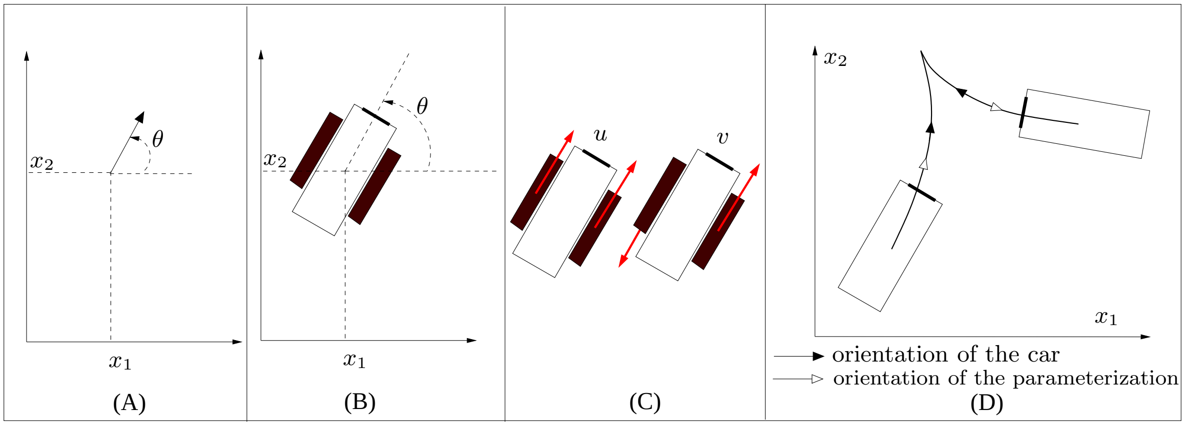


Figure 13.9: Mechanical interpretation of the problem on $SE(2)$.

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 integrable in the sense of Liouville, since we have enough constants of the motion in involution (i.e. H, p_1, p_2 or equivalently H, P, θ). The last two equations gives 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

$$\ddot{\bar{\theta}} = P^2\sin\bar{\theta}. \tag{13.62}$$

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 (13.62) with gravity P^2 and initial condition

$$\bar{\theta}(0) = -2p_a, \qquad \dot{\bar{\theta}}(0) = 2p_\theta(0).$$

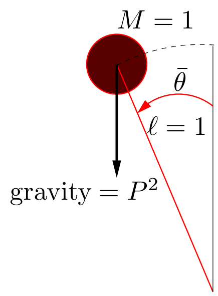


Figure 13.10: The inverted pendulum

Once that the pendulum equation has been solved one obtains

$$\theta(t) = \frac{\bar{\theta}(t)}{2} + p_a \tag{13.63}$$

$$x_1(t) = \int_0^t \dot{x}_1(s)\,ds = P\int_0^t \cos(\theta(s) - p_a)\cos\theta(s)\,ds = P\int_0^t \cos\left(\frac{\bar{\theta}(s)}{2}\right)\cos\left(\frac{\bar{\theta}(t)}{2} + p_a\right)\,ds \tag{13.64}$$

$$x_2(t) = \int_0^t \dot{x}_2(s)\,ds = P\int_0^t \cos(\theta(s) - p_a)\sin\theta(s)\,ds = P\int_0^t \cos\left(\frac{\bar{\theta}(s)}{2}\right)\sin\left(\frac{\bar{\theta}(t)}{2} + p_a\right)\,ds \tag{13.65}$$

Qualitative behaviour of the geodesics.
Equation (13.62) 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 a constant of the motion (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 π, then the projection on the (x_1, x_2) plane of the geodesic has a cusp.

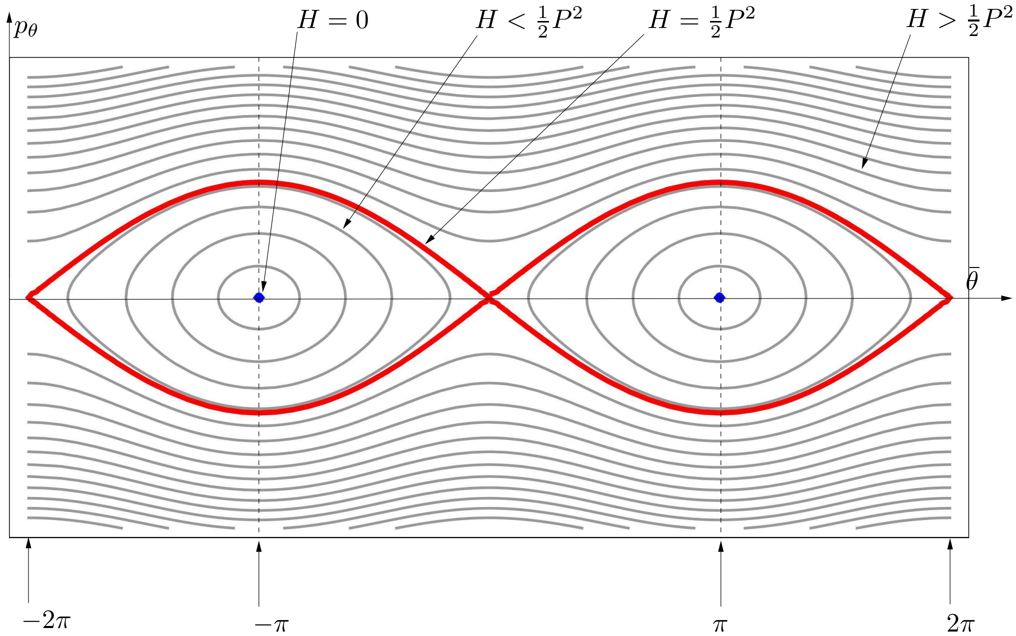


Figure 13.11: Trajectories of the inverted pendulum

- 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 $k \in \mathbb{N}$.

- Trajectories with $P > 0$ and corresponding to the oscillating pendulum. In this case $\bar\theta(t)$ is oscillating either around π 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 π 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 (13.62) 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 $\theta(t) = \pm t$, $x_1(t) = 0$, $x_2(t) = 0$.

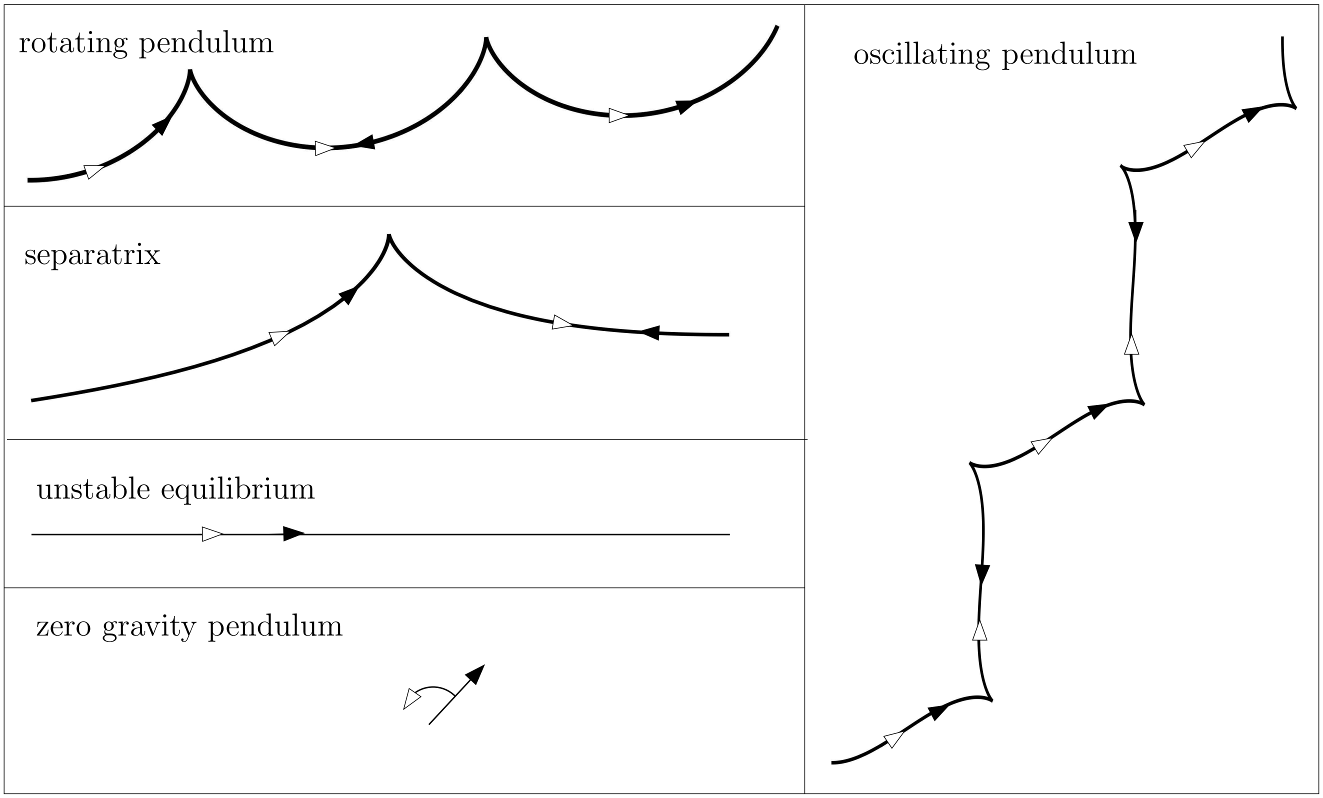


Figure 13.12: Geodesics for $SE(2)$

Remark 13.45. 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 ℓ 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 ℓ and orthogonal to ℓ.

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 ℓ.

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.

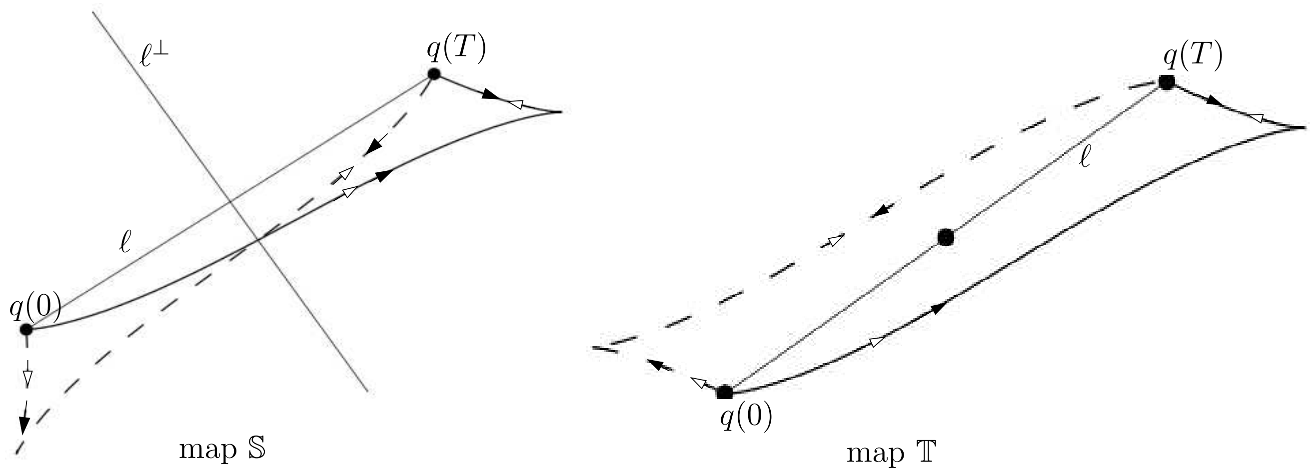


Figure 13.13: Maps $\mathbb{S}$ and $\mathbb{T}$. Courtesy of Y. Sachkov.

Remark 13.46. 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 13.47. 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. We have the following

Theorem 13.48 (Yuri Sachkov)**.** *A geodesic $q(.)$ 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.

13.9 The Martinet sub-Riemannian structure

Let us write a point of $\mathbb{R}^3$ as (x,y,z). The Martinet sub-Riemannian structure is the structure in $\mathbb{R}^3$ for which an orthonormal frame is given by

$$X_1 = \begin{pmatrix} 1 \\ 0 \\ \frac{y^2}{2} \end{pmatrix}, \quad X_2 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}. \tag{13.66}$$

Remark 13.49. 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 form $A = \frac{y^2}{2}dx$.

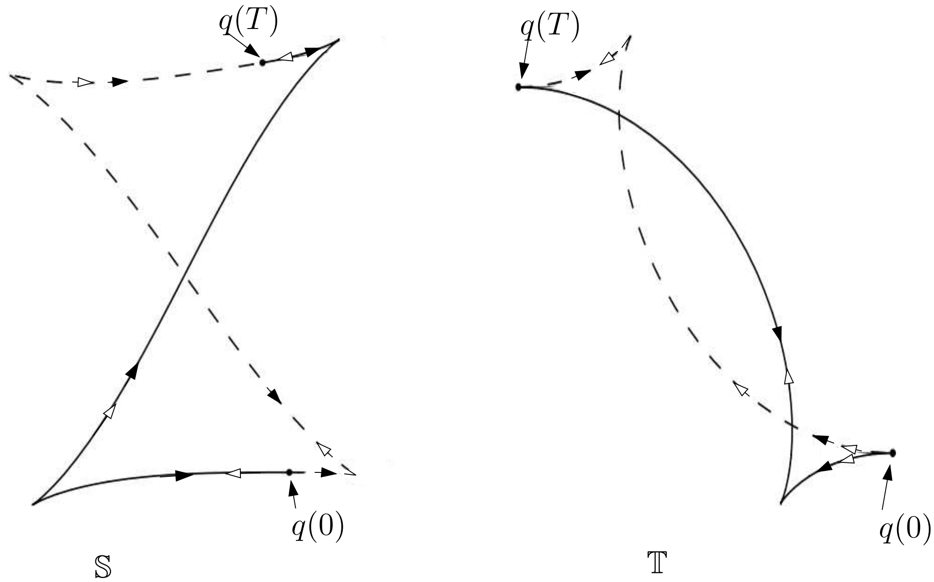


Figure 13.14: Cut loci corresponding to $\mathbb{S}$ and $\mathbb{T}$. Courtesy of Y. Sachkov.

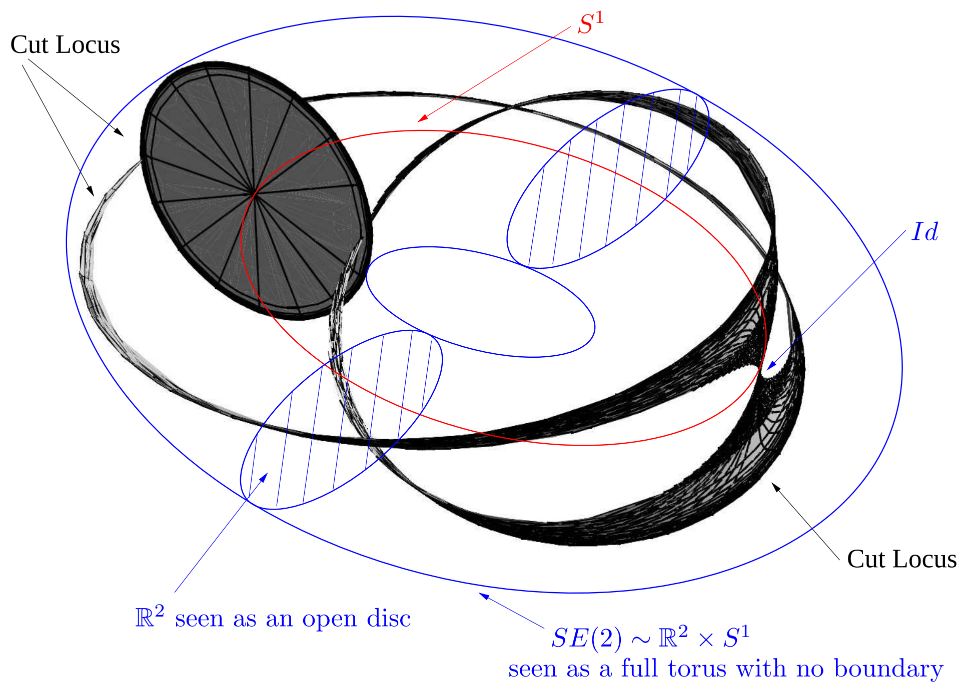


Figure 13.15: Cut locus (dark region) from the identity for the sub-Riemannian problem on $SE(2)$. Courtesy of Y. Sachkov. In this picture $SE(2)$ (that has the topology of $\mathbb{R}^2 \times S^1$) is represented as a solid torus without boundary given by $B_2 \times S^1$, where B_2 is the 2D disc without boundary.

In other words the trajectory realizing the sub-Riemannian distance for the Martinet problem 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

$$\int_\gamma A = \int_0^T A(\dot\gamma(t))dt = \int_0^T \frac{y(t)}{2}\dot x(t)dt = z_1,$$

and whose projection in the (x,y)-plane is the shortest for the Euclidean distance.

This structure is bracket generating, but it is not equiregular. Indeed we have

$$X_3 := [X_1,X_2] = \begin{pmatrix} 0 \\ 0 \\ -y \end{pmatrix}, \qquad [X_3,X_2] = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.$$

Hence the structure is 3D-contact out of $\{y=0\}$ and to get the full tangent space in every point one need one more bracket.

In the following two sections we are going to construct the Pontryagin extremals. We already know Section 4.4.2 that the support of abnormal extremals should be contained in the set $\{A=0\}$ that is the plane $\{y=0\}$. Such set is called the Martinet 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., $\gamma(t)\equiv\gamma(0)$) or we have

$$\langle p(t), X_3(q(t))\rangle = p_z(t)y(t) \equiv 0. \tag{13.67}$$

Since (p_x,p_y,p_x) cannot vanish, we have that γ is contained in the Martinet surface i.e., $\gamma([a,b])\subset\{y=0\}$.

To obtain the controls corresponding to γ let us differentiate once more (13.67). 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 $[X_1,X_3]=0$ and $[X_2,X_3]=(0,0,-1)^t$. Since again (p_x,p_y,p_x) cannot vanish we obtain

$$u_2(t) = 0 \quad \text{ for almost every } t.$$

Indeed we already knew this fact since the only way to stay on the Martinet surface is to have $u_2=0$ almost everywhere. The value of u_1 is then obtained by requiring that γ 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 $|u_1(t)| = 1$. Such control can switch arbitrarily between 1 and -1. Because of Remark 13.49 only trajectories corresponding to a control that is almost everywhere constant are optimal. We then obtain the following.

Proposition 13.50. *Arclength parametrized trajectories admitting an abnormal lift are Lipschitz trajectories $\gamma : [a,b] \to \mathbb{R}^3$ lying on the Martinet surface 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

$$H(q,p) = \frac{1}{2}(h_1(q,p)^2 + h_2(q,p)^2),$$

where

$$h_1(q,p) = p_x + \frac{y^2}{2}p_z, \qquad h_2(q,p) = p_y.$$

The Hamiltonian equations are then

$$\dot{x} = \frac{\partial H}{\partial p_x} = h_1, \qquad \dot{p}_x = -\frac{\partial H}{\partial x} = 0, \tag{13.68}$$

$$\dot{y} = \frac{\partial H}{\partial p_y} = p_y, \qquad \dot{p}_y = -\frac{\partial H}{\partial y} = -h_1 y\, p_z, \tag{13.69}$$

$$\dot{z} = \frac{\partial H}{\partial p_z} = h_1\frac{y^2}{2}, \qquad \dot{p}_z = -\frac{\partial H}{\partial z} = 0. \tag{13.70}$$

Notice that this Hamiltonian system is integrable in the sense of Liouville, since we have enough constants of the motion in involution (i.e. H, p_x, p_z).

From (13.70) we have that p_z is constant. Let us set $p_z = a$. We can solve (13.68) and (13.69) since these equations are independent from z. Let us use as coordinates (x, y, h_1, h_2). We have

$$\dot{x} = h_1, \qquad \dot{h}_1 = \dot{p}_x + y\underbrace{\dot{y}}_{p_y}a = a\,y\,h_2, \tag{13.71}$$

$$\dot{y} = p_y = h_2, \qquad \dot{h}_2 = \dot{p}_y = -a\,y\,h_1. \tag{13.72}$$

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 convinient to set

$$h_1(t) = \cos\theta(t), \qquad h_2(t) = \sin\theta(t).$$

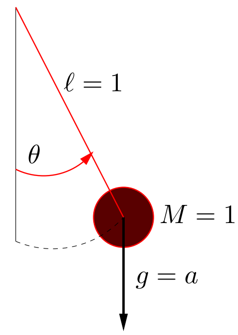


Figure 13.16: The pendulum for the Martinet distribution

The equations for h_1 and h_2 in (13.71) and (13.72) give then

$$\begin{aligned} -\sin(\theta)\dot{\theta} &= ay\sin(\theta), \\ \cos(\theta)\dot{\theta} &= -ay\cos(\theta), \end{aligned}$$

from which we have

$$\dot{\theta} = -ay. \tag{13.73}$$

This equation together with $\dot{y} = h_2 = \sin\theta$ (see the equation for $\dot{y}$ in (13.72)) gives

$$\ddot{\theta} = -a\sin\theta \tag{13.74}$$

We obtain again a pendulum 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 (13.73) we have that $\dot{\theta}(0) = 0$.

Once equation the pendulum equation has been solved, one gets

$$x(t) = \int_0^t \dot{x}(s)\,ds = \int_0^t \ h_1(q(s), p(s))\,ds = \int_0^t \cos\theta(s)\,ds, \tag{13.75}$$

$$y(t) = \int_0^t \dot{y}(s)\,ds = \int_0^t \ h_2(q(s), p(s))\,ds = \int_0^t \sin\theta(s)\,ds, \tag{13.76}$$

$$z(t) = \int_0^t \dot{z}(s)\,ds = \int_0^t \ h_1(q(s), p(s))\frac{y^2(s)}{2}\,ds = \int_0^t \cos(\theta(s))\frac{y^2(s)}{2}\,ds. \tag{13.77}$$

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 constant of the motion

$$H_p(\theta, \dot{\theta}) = \frac{1}{2}\dot{\theta}^2 - a\cos(\theta).$$

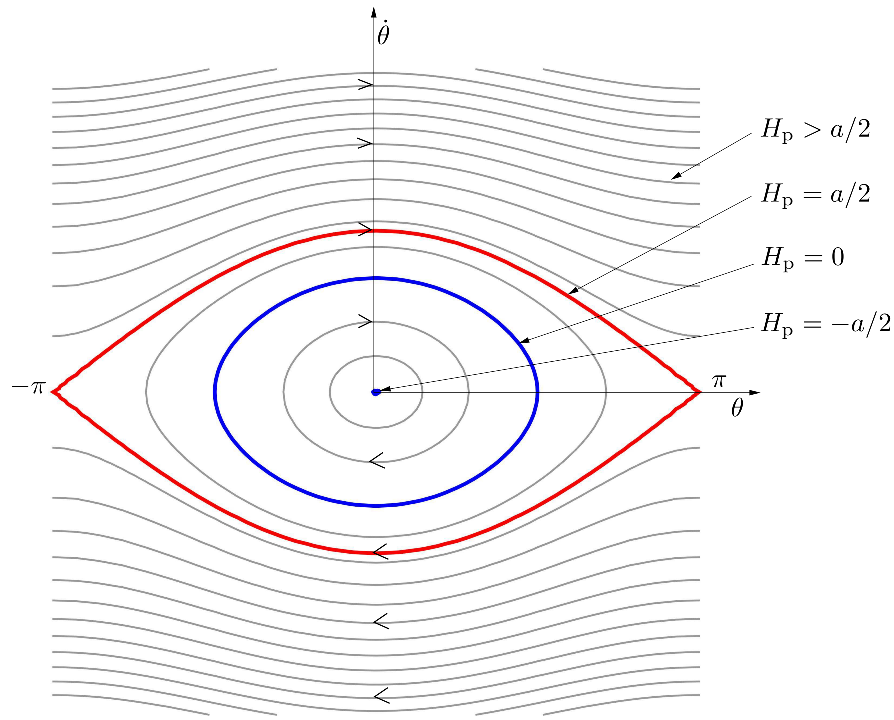


Figure 13.17: The phase portrait of the pendulum for the Martinet problem

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 (13.75)-(13.77) 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 surface 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 surface 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 13.49 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* [**?**, p. 62]

$$\mathrm{Eps}(\phi, m) := \int_0^{\phi} \mathrm{dn}^2(w, m)\, dw.$$

Let us define the following functions of t, θ_0, a (here we assume $a > 0$, $\theta_0 \in (0, \pi)$).

$$k = \sqrt{\frac{1 - \cos(\theta_0)}{2}}, \tag{13.78}$$

$$k' = \sqrt{\frac{1 + \cos(\theta_0)}{2}}, \tag{13.79}$$

$$u(t, k, a) = K(k^2) + t\sqrt{a}, \tag{13.80}$$

$$\Upsilon(t, k, a) = \mathrm{Eps}(u(t, k, a), k^2) - \mathrm{Eps}(K(k^2), k^2), \tag{13.81}$$

Theorem 13.51 (Agrachev, Bonnard, Chyba, Kupka)**.** *The normal geodesics starting from the origin for $\theta_0 \in (0, \pi)$ and $a > 0$ are given by:*

$$x(t) = -t + \frac{2}{\sqrt{a}}\Upsilon(t, k, a) \tag{13.82}$$

$$y(t) = -2\frac{k}{\sqrt{a}}\mathrm{cn}(u(t, k, a), k^2) \tag{13.83}$$

$$z(t) = \frac{2}{3a^{3/2}}\left[(2k^2 - 1)\Upsilon(t, k, a) + k'^2 t\sqrt{a} + 2k^2 \mathrm{sn}(u(t, k, a), k^2)\mathrm{cn}(u(t, k, a), k^2)\mathrm{dn}(u(t, k, a), k^2)\right] \tag{13.84}$$

For negative values of θ_0 and/or a, the formulas are obtained from the previous ones considering that a change in sign of θ_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 13.52. 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 [**?**, 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 13.53 (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 surface $\{y = 0\}$. The cut time is given by the formula*

$$t_{\text{cut}} = \begin{cases} 2\frac{K(k^2)}{\sqrt{a}}, & \text{for} a > 0, \\ 2\frac{K(k'^2)}{\sqrt{-a}}, & \text{for} a < 0. \end{cases}$$

The Martinet sphere for $t = 1$ is drawn in Figure 13.18. Its intersection with the Martinet surface (that is also the cut locus) is drawn 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, the cut locus is the image via the exponential map of an unbounded curve on the cylinder $H = 1/2$. Points on this curve that having high values of a correspond to the part of the sphere that become tangent to the abnormal as pictured.

13.10 Bibliographical Note

Explicit computations of Pontryagin extremals and the cut locus for the Heisenberg group and its higher dimensional generalizations are well known. [1, **?**, 1, **?**, **?**, **?**, **?**]

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 [58]). Its sub-Riemannian variant was used to construct the optimal syntheses in several cases. See for instance [2, 81, 90, 91]. 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 sub-Riemannian structure and in particular to prove Theorem 13.53. See [2].

The shape of the synthesis for the Grushin plane starting from a Riemannian point was drawn in [4, 31]. 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 [33] but using a different technique. These optimal syntheses, together with the one for $SO_+(2, 1)$, were also constructed in [23, 24, 25] 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 [74, 90, 91] that also proved Theorem 13.48 in full details.

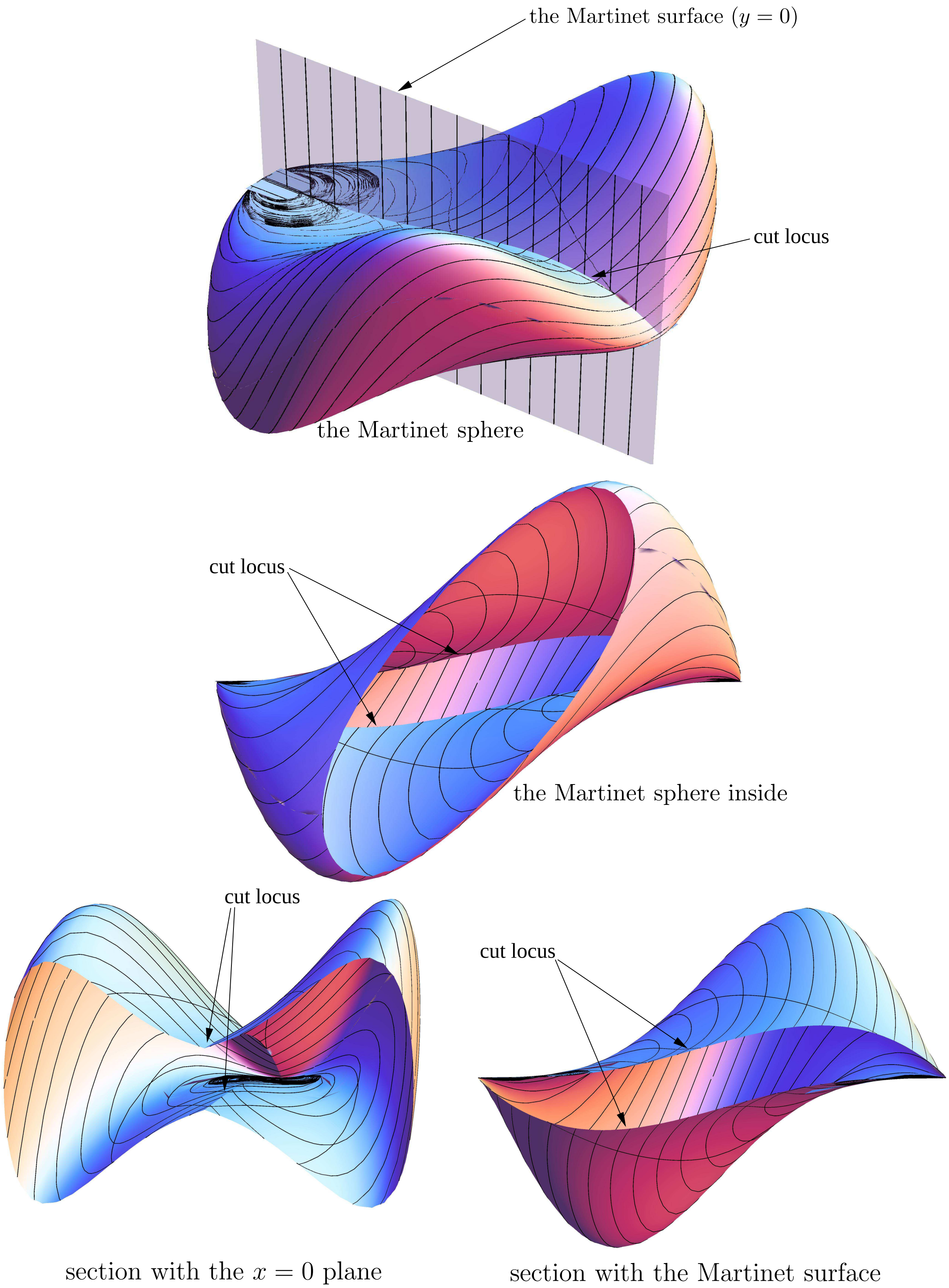


Figure 13.18: The Martinet sphere for $t = 1$.

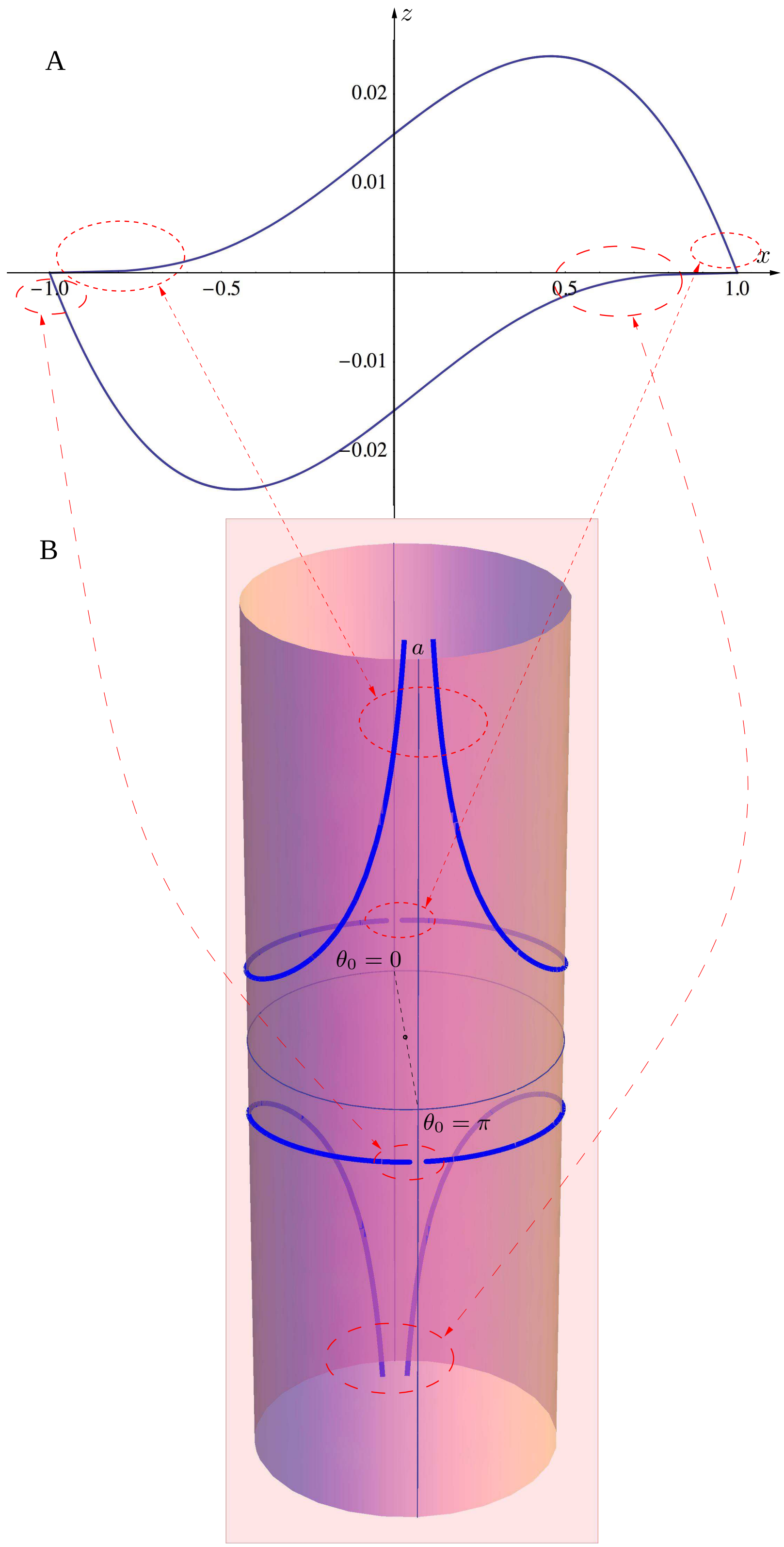


Figure 13.19: A: the intersection of the Martinet sphere for $t = 1$ with the Martinet surface, that is also the cut locus. B: the cut locus seen on the cotangent bundle on $H = \frac{1}{2}$.

The optimal synthesis for the Martinet sub-Riemannian structure was constructed in [2]. In the same paper one can also find the proof of Theorem 13.53. See also [26].

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.

14.1 The geometry of the Lagrange Grassmannian

In this section we recall some basic facts about Grassmanians of k-dimensional subspaces of an n-dimensional vector space and then we 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 *Grassmanian 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 $k(n-k)$.

Now we describe the tangent space to this manifold.

Proposition 14.2. *Let $W \in G_k(V)$. We have a canonical isomorphism*

$$T_W G_k(V) \simeq \operatorname{Hom}(W, V/W).$$

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

$$W(t) = \operatorname{span}\{e_1(t), \ldots, e_k(t)\}, \qquad W(0) = W.$$

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. In the same way 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 space have the same dimension. □

Let us now consider a symplectic vector space (Σ, σ), i.e. a $2n$-dimensional vector space Σ 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 σ, one has

$$\dim \Pi + \dim \Pi^\angle = \dim \Sigma. \tag{14.3}$$

where as usual we denote the symplectic orthogonal by $\Pi^\angle = \{x \in \Sigma \,|\, \sigma(x, y) = 0,\ \forall\, y \in \Pi\}$.

Exercise 14.4. Prove the following properties for a vector subspace $\Pi \subset \Sigma$:

(i) Π is symplectic iff $\Pi \cap \Pi^\angle = \{0\}$,

(ii) Π is isotropic iff $\Pi \subset \Pi^\angle$,

(iii) Π is Lagrangian iff $\Pi = \Pi^\angle$.

Exercise 14.5. Prove that, given two subspaces $A, B \subset \Sigma$, one has the 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 admits Lagrangian subspaces. Indeed fix any non-zero element $e_1 := e \neq 0$ in Σ. Choose iteratively

$$e_i \in \operatorname{span}\{e_1, \dots, e_{i-1}\}^{\angle} \setminus \operatorname{span}\{e_1, \dots, e_{i-1}\}, \quad i = 2, \dots, n. \tag{14.4}$$

Then $\Pi := \operatorname{span}\{e_1, \dots, e_n\}$ is a Lagrangian subspace by construction. Notice that the choice (14.4) is possible by (14.3)

Lemma 14.7. *Let $\Pi = \operatorname{span}\{e_1, \dots, e_n\}$ be a Lagrangian subspace of Σ. Then there exists vectors $f_1, \dots, f_n \in \Sigma$ such that*

(*i*) $\Sigma = \Pi \oplus \Delta, \qquad \Delta := \operatorname{span}\{f_1, \dots, f_n\}$,

(*ii*) $\sigma(e_i, f_j) = \delta_{ij}, \quad \sigma(e_i, e_j) = \sigma(f_i, f_j) = 0, \qquad \forall\, i, j = 1, \dots, n.$

Proof. We prove the lemma by induction. By nondegeneracy of σ there exists a non-zero $x \in \Sigma$ such that $\sigma(e_n, x) \neq 0$. Then we define the vector

$$f_n := \frac{x}{\sigma(e_n, x)}, \qquad \Longrightarrow \qquad \sigma(e_n, f_n) = 1.$$

The last equality implies that σ restricted to $\operatorname{span}\{e_n, f_n\}$ is nondegerate, hence by (a) of Exercise 14.4

$$\operatorname{span}\{e_n, f_n\} \cap \operatorname{span}\{e_n, f_n\}^{\angle} = 0, \tag{14.5}$$

And we can apply induction on the $2(n-1)$ subspace $\Sigma' := \operatorname{span}\{e_n, f_n\}^{\angle}$. Notice that (14.5) implies that σ is non degenerate also on Σ'. $\square$

Remark 14.8. In particular the complementary subspace $\Delta = \operatorname{span}\{f_1, \dots, f_n\}$ defined in Lemma 14.7 is Lagrangian and transversal to Π

$$\Sigma = \Pi \oplus \Delta.$$

Considering coordinates induced from the basis chosen for this splitting we can write $\Sigma = \mathbb{R}^{n*} \oplus \mathbb{R}^n$, (denoting $\mathbb{R}^{n*}$ denotes the set of row vectors). More precisely $x = (\zeta, z)$ if

$$x = \sum_{i=1}^{n} \zeta^i e_i + z^i f_i, \qquad \zeta = \begin{pmatrix} \zeta^1 & \cdots & \zeta^n \end{pmatrix}, \quad z = \begin{pmatrix} z^1 \\ \vdots \\ z^n \end{pmatrix},$$

and using canonical form of σ on our basis (see Lemma 14.7) we find that in coordinates, if $x_1 = (\zeta_1, z_1), x_2 = (\zeta_2, z_2)$ we get

$$\sigma(x_1, x_2) = \zeta_1 z_2 - \zeta_2 z_1, \tag{14.6}$$

where we denote with ζz the standard rows by columns product.

Lemma 14.7 shows that the group of symplectomorphisms acts transitively on pairs of transversal Lagrangian subspaces. The next exercise, whose proof is an adaptation of the previous one, describes all the orbits of the action of the group of symplectomorphisms on pairs of subspaces of a symplectic vector spaces.

Exercise 14.9. Let Λ_1, Λ_2 be two subspaces in a symplectic vector space Σ, and assume that $\dim \Lambda_1 \cap \Lambda_2 = k$. Show that there exists Darboux coordinates (p, q) in Σ such that

$$\Lambda_1 = \{(p, 0)\}, \qquad \Lambda_2 = \{((p_1, \dots, p_k, 0, \dots, 0), (0, \dots, 0, q_{k+1}, \dots, q_n)\}.$$

14.1.1 The Lagrange Grassmannian

Definition 14.10. The *Lagrange Grassmannian* $L(\Sigma)$ of a symplectic vector space Σ is the set of its n-dimensional Lagrangian subspaces.

Proposition 14.11. *$L(\Sigma)$ is a compact submanifold of the Grassmannian $G_n(\Sigma)$ of n-dimensional subspaces. 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

$$\Delta^{\pitchfork} = \{\Lambda \in L(\Sigma) : \Lambda \cap \Delta = 0\}.$$

Clearly $\Delta^{\pitchfork} \subset L(\Sigma)$ is an open subset and 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. Every n-dimensional subspace $\Lambda \subset \Sigma$ which is transversal to Δ is the graph of a linear map from Π to Δ. More precisely there exists a matrix S_Λ such that

$$\Lambda \cap \Delta = 0 \Leftrightarrow \Lambda = \{(z^T, S_\Lambda z), z \in \mathbb{R}^n\}.$$

(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

$$\sigma((z_1^T, S_\Lambda z_1), (z_2^T, S_\Lambda z_2)) = z_1^T S_\Lambda z_2 - z_2^T S_\Lambda z_1 = 0,$$

which means exactly S_Λ symmetric. Hence the open set of all subspaces that are transversal to Λ is parametrized by the set of symmetric matrices, that gives coordinates in this open set. This also proves that the dimension of $L(\Sigma)$ coincide with the dimension of the space of symmetric matrices, hence (14.7). 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 14.12. *Let $\Lambda \in L(\Sigma)$. Then we have a canonical isomorphism*

$$T_\Lambda L(\Sigma) \simeq Q(\Lambda),$$

where $Q(\Lambda)$ denote the set of quadratic forms on Λ.

Proof. Consider a smooth curve $\Lambda(t)$ in $L(\Sigma)$ such that $\Lambda(0) = \Lambda$ and $\dot{\Lambda}(0) \in T_\Lambda L(\Sigma)$ its tangent vector. As before consider a point $x \in \Lambda$ and a smooth extension $x(t) \in \Lambda(t)$ and denote with $\dot{x} := \dot{x}(0)$. We define the map

$$\underline{\dot{\Lambda}} : x \mapsto \sigma(x, \dot{x}), \tag{14.8}$$

that is nothing else but the quadratic map associated to the self adjoint map $x \mapsto \dot{x}$ by the symplectic structure. We show that in coordinates $\underline{\dot{\Lambda}}$ is a well defined quadratic map, independent on all choices. Indeed

$$\Lambda(t) = \{(z^T, S_{\Lambda(t)} z), z \in \mathbb{R}^n\},$$

and the curve $x(t)$ can be written

$$x(t) = (z(t)^T, S_{\Lambda(t)} z(t)), \qquad x = x(0) = (z^T, S_\Lambda z),$$

for some curve $z(t)$ where $z = z(0)$. Taking derivative we get

$$\dot{x}(t) = (\dot{z}(t)^T, \dot{S}_{\Lambda(t)} z(t) + S_{\Lambda(t)} \dot{z}(t)),$$

and evaluating at $t = 0$ (we simply omit t when we evaluate at $t = 0$) we have

$$x = (z^T, S_\Lambda z), \qquad \dot{x} = (\dot{z}^T, \dot{S}_\Lambda z + S_\Lambda \dot{z}),$$

and finally get, using the simmetry of S_Λ, that

$$\begin{aligned} \sigma(x, \dot{x}) &= z^T(\dot{S}_\Lambda z + S_\Lambda \dot{z}) - \dot{z}^T S_\Lambda z \\ &= z^T \dot{S}_\Lambda z + z^T S_\Lambda \dot{z} - \dot{z}^T S_\Lambda z \\ &= z^T \dot{S}_\Lambda z. \end{aligned} \tag{14.9}$$

□

Exercise 14.13. Let $\Lambda(t) \in L(\Sigma)$ such that $\Lambda = \Lambda(0)$ and σ be the symplectic form. Prove that the map $S : \Lambda \times \Lambda \to \mathbb{R}$ defined by $S(x, y) = \sigma(x, \dot{y})$, where $\dot{y} = \dot{y}(0)$ is the tangent vector to a smooth extension $y(t) \in \Lambda(t)$ of y, is a symmetric bilinear map.

Remark 14.14. We have the following natural interpretation of this result: since $L(\Sigma)$ is a submanifold of the Grassmanian $G_n(\Sigma)$, 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 Proposition 14.2. Being Λ a Lagrangian subspace of Σ, the symplectic structure identifies in a canonical way the factor space Σ/Λ with the dual space Λ^* defining

$$\Sigma/\Lambda \simeq \Lambda^*, \qquad \langle [z]_\Lambda, x\rangle = \sigma(z, x). \tag{14.10}$$

Hence the tangent space to the Lagrange Grassmanian consist of those linear maps in the space $\mathrm{Hom}(\Lambda, \Lambda^*)$ that are self-adjoint, which are naturally identified with quadratic forms on Λ itself. [1]

Remark 14.15. 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, i.e. the curve on $L(\Sigma)$ induced by the action on Λ by the flow of the linear Hamiltonian vector field $\vec{h}$ associated with a quadratic Hamiltonian $h \in C^\infty(\Sigma)$. (Recall that a Hamiltonian vector field transform Lagrangian subspaces into Lagrangian subspaces.)

[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 $q(v) = B(v, v)$.

Proposition 14.16. *Let $\Lambda \in L(\Sigma)$ and define $\Lambda(t) = e^{t\vec{h}}(\Lambda)$. Then $\underline{\dot{\Lambda}} = 2h|_{\Lambda}$.*

Proof. Consider $x \in \Lambda$ and the smooth extension $x(t) = e^{t\vec{h}}(x)$. Then $\dot{x} = \vec{h}(x)$ and by definition of Hamiltonian vector field we find

$$\begin{aligned}\sigma(x, \dot{x}) &= \sigma(x, \vec{h}(x)) \\ &= \langle d_x h, x\rangle \\ &= 2h(x),\end{aligned}$$

where in the last equality we used that h is quadratic on fibers. □

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 Λ as a quadratic map)

Definition 14.17. Let $\Lambda(t) \in L(\Sigma)$ be a smooth curve in the Lagrange Grassmannian. We say that the curve is

(i) *monotone* increasing (descreasing) if $\underline{\dot{\Lambda}}(t) \geq 0$ ($\underline{\dot{\Lambda}}(t) \leq 0$).

(ii) *strictly monotone* increasing (decreasing) if the inequality in (i) is strict.

(iii) *regular* if its derivative $\underline{\dot{\Lambda}}(t)$ is a non degenerate quadratic form.

Remark 14.18. 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 14.12 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. (i.e. another curve $\Lambda^{\circ}(t)$ in the Lagrange Grassmannian defined by $\Lambda(t)$ and such that $\Sigma = \Lambda(t) \oplus \Lambda^{\circ}(t)$.)

Consider an arbitrary Lagrangian splitting $\Sigma = \Lambda(0) \oplus \Delta$ defined by a complement Δ to $\Lambda(0)$ (see Lemma 14.7) and fix coordinates in such a way that that

$$\Sigma = \{(p, q),\, p, q \in \mathbb{R}^n\}, \qquad \Lambda(0) = \{(p, 0),\, p \in \mathbb{R}^n\}, \qquad \Delta = \{(0, q),\, q \in \mathbb{R}^n\}.$$

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 complement to $\Lambda(0)$ are parametrized by a symmetrix matrix B as follows

$$\Delta_B = \{(Bq, q), q \in \mathbb{R}^n\}, \qquad B = B^T.$$

The following lemma shows how the coordinate expression of our curve $\Lambda(t)$ change in the new coordinate set defined by the splitting $\Sigma = \Lambda(0) \oplus \Delta_B$.

Lemma 14.19. *Let $S_B(t)$ the one parametric family of symmetric matrices defining $\Lambda(t)$ in coordinates w.r.t. 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, q) and (p', q') denotes coordinates with respect to the splitting defined by the subspaces Δ and Δ_B we have

$$\begin{cases} p' = p - Bq \\ q' = q \end{cases} \tag{14.12}$$

The matrix $S_B(t)$ by definition is the matrix that satisfies the identity $q' = S_B(t)p'$. Using that $q = S(t)p$ by definition of $\Lambda(t)$, from (14.12) we find

$$q' = q = S(t)p = S(t)(p' + Bq'),$$

and with straightforward computations we finally get

$$S_B(t) = (I - S(t)B)^{-1}S(t) = (S(t)^{-1} - B)^{-1}.$$

□

Since $\dot{S}(t)$ represents the tangent vectors to the regular curve $\Lambda(t)$, its properties are invariant with respect to change of coordinates. Hence it is natural to look for a change of coordinates (i.e. a choice of the matrix B) that simplifies the second derivative our curve.

Corollary 14.20. *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 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 degerate, 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 (14.13). Notice that by construction $\Lambda^\circ(0)$ is a Lagrangian subspace and it is transversal to $\Lambda(0)$. The same argument can be applied to define $\Lambda^\circ(t)$ for every t.

Definition 14.21. 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 14.22. 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 $\Delta = \{(0,q), q \in \mathbb{R}^n\}$. (Actually this condition is equivalent to the invertibility of $B(t)$.) Notice that if $S(0) = 0$ then $S^\circ(0) = B(0)^{-1}$.

Remark 14.23. The set Λ^{tr} of all n-dimensional spaces transversal to a fixed subspace Λ 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 Λ 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 Λ^* and Λ, that it isomorphic to $Q(\Lambda^*)$.

Notice that if we fix a distinguished complement of Λ, 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 14.24. Prove that the operator A defined by (14.14), in the case when Λ is Lagrangian, is a self-adjoint operator.

Remark 14.25. 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 Δ (see Remark 14.23) that is given by $S^{-1}(t)$ and write the Laurent expansion of this curve in the affine space

$$\begin{aligned} S(t)^{-1} &= \left(t\dot{S} + \frac{t^2}{2}\ddot{S} + O(t^3)\right)^{-1} \\ &= \frac{1}{t}\dot{S}^{-1}\left(I + \frac{t}{2}\ddot{S}\dot{S}^{-1} + O(t^2)\right)^{-1} \\ &= \frac{1}{t}\dot{S}^{-1} \underbrace{-\frac{1}{2}\dot{S}^{-1}\ddot{S}\dot{S}^{-1}}_{B} + O(t). \end{aligned}$$

It is 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 (i.e. $S_B(t)^{-1}$ has no term of order zero). This is equivalent to say that a regular curve let 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 14.26. 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 14.27. In the monotonic case, when $|\underline{\dot{\Lambda}}(t)|$ 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 $\underline{\dot{\Lambda}}^\circ(t)\,\mathrm{sign}(\underline{\dot{\Lambda}}^\circ(t))$.

Definition 14.28. Let Λ_1, Λ_2 be two transversal Lagrangian subspaces of Σ. We denote

$$\pi_{\Lambda_1\Lambda_2} : \Sigma \to \Lambda_2, \tag{14.17}$$

the projection on Λ_2 parallel to Λ_1, i.e. the linear operator such that

$$\pi_{\Lambda_1\Lambda_2}|_{\Lambda_1} = 0 \qquad \pi_{\Lambda_1\Lambda_2}|_{\Lambda_2} = Id.$$

Exercise 14.29. Assume Λ_1 and Λ_2 be two Lagrangian subspaces in Σ and assume that, in some coordinate set, $\Lambda_i = \{(x, S_i x), \in \mathbb{R}^n\}$ 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 $\pi_{\Lambda_1\Lambda_2}$:

$$\pi_{\Lambda_1\Lambda_2} = \begin{pmatrix} S_{12}^{-1} S_1 & -S_{12}^{-1} \\ S_2 S_{12}^{-1} S_1 & -S_2 S_{12}^{-1} \end{pmatrix}, \qquad S_{12} := S_1 - S_2. \tag{14.18}$$

From the very definition of the derivative of our curve we can get the following geometric characterization of the curvature of a curve.

Proposition 14.30. *Let $\Lambda(t)$ a regular curve in $L(\Sigma)$ and $\Lambda^\circ(t)$ its derivative curve. Then*

$$\underline{\dot{\Lambda}}(t)(x_t) = \pi_{\Lambda(t)\Lambda^\circ(t)}(\dot{x}_t), \qquad \underline{\dot{\Lambda}}^\circ(t)(x_t) = -\pi_{\Lambda^\circ(t)\Lambda(t)}(\dot{x}_t).$$

In particular the curvature is the composition $R_\Lambda(t) = \underline{\dot{\Lambda}}^\circ(t) \circ \underline{\dot{\Lambda}}(t)$.

[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 $\Lambda^* \simeq \Delta$.

Proof. Recall that, by definition, the linear operator $\underline{\dot{\Lambda}} : \Lambda \to \Sigma/\Lambda$ associated with the quadratic form is the map $x \mapsto \dot{x} \pmod{\Lambda}$. Hence to build the map $\Lambda \to \Lambda^\circ$ it is enough to compute the projection of $\dot{x}$ onto the complement Λ°, that is exactly $\pi_{\Lambda\Lambda^\circ}(\dot{x})$. Notice that the minus sign in equation (14.30) 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. $\square$

The curvature $R_\Lambda(t)$ of the curve $\Lambda(t)$ is a kind of relative velocity between the two curves $\Lambda(t)$ and $\Lambda^\circ(t)$. In particular notice that if the two curves moves in the same direction we have $R_\Lambda(t) > 0$.

Now we compute the expression of the curvature $R_\Lambda(t)$ in coordinates.

Proposition 14.31. *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 Λ (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

$$\Lambda(t) = \{(p, S(t)p)\}, \qquad \Lambda^\circ(t) = \{(p, S^\circ(t)p)\}.$$

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 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

$$\pi_{\Lambda\Lambda^\circ}(p, Sp) = (p', S^\circ p'), \qquad p' = -(S - S^\circ)^{-1}\dot{S}p.$$

Exchanging the role of Λ and Λ°, and taking into account of the minus sign one finds that the coordinate representation of R is given by

$$R = (S^\circ - S)^{-1}\dot{S}^\circ(S^\circ - S)^{-1}\dot{S}. \tag{14.21}$$

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 coordinate independent by construction, this is not restrictive.

Under this extra assumption, it follows from (14.13) that

$$\Lambda(t) = \{(p, S(t)p)\}, \qquad \Lambda^\circ(t) = \{(p, S^\circ(t)p)\},$$

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 $t = 0$

$$\begin{aligned} R &= (S^\circ - S)^{-1}\dot{S}^\circ(S^\circ - S)^{-1}\dot{S} \\ &= B\left(\left.\frac{d}{dt}\right|_{t=0} B(t)^{-1} + S(t)\right)B\dot{S} \\ &= (B\dot{S})^2 - \dot{B}\dot{S}. \end{aligned}$$

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 14.32. The formula for the curvature $R_\Lambda(t)$ of a curve $\Lambda(t)$ in $L(\Sigma)$ 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

$$\Lambda(0) = \{(p, 0), p \in \mathbb{R}^n\}, \qquad \Lambda^\circ(0) = \{(0, q), q \in \mathbb{R}^n\}.$$

Indeed using a symplectic change of coordinates in Σ that preserves both Λ and Λ° (i.e. of the kind $p' = Ap$, $q' = (A^{-1})^* q$) we can choose the matrix A in such a way that $\dot{S}(0) = I$. Moreover we know from Proposition that the fact that $\Lambda^\circ = \{(0, q), q \in \mathbb{R}^n\}$ is equivalent to $\ddot{S}(0) = 0$. Hence one finds from (14.20) that

$$R = \frac{1}{2}\dddot{S}$$

When the curve $\Lambda(t)$ is strictly monotone, the curvature R 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 14.33. Let $f : \mathbb{R} \to \mathbb{R}$ be a smooth function. The *Schwartzian derivative* of f is defined as

$$\mathcal{S}f := \left(\frac{f''}{2f'}\right)' - \left(\frac{f''}{2f'}\right)^2 \tag{14.22}$$

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 14.34. The previous proposition says that the curvature R is the matrix version of the Schwartzian derivative of the matrix S (cfr. (14.19) and (14.22)).

Example 14.35. Let Σ 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 $R(t) = \alpha\dot{S}(t) - \alpha^2 S^2(t) = \alpha^2$, i.e. the curve has constant curvature.

We end this section with a useful formula on the curvature of a reparametrized curve.

Proposition 14.36. *Let $\varphi : \mathbb{R} \to \mathbb{R}$ a diffeomorphism and define the curve $\Lambda_\varphi(t) := \Lambda(\varphi(t))$. Then*

$$R_{\Lambda_\varphi}(t) = \dot{\varphi}^2(t) R_\Lambda(\varphi(t)) + R_\varphi(t)\mathrm{Id}. \tag{14.23}$$

Proof. It is a simple check that the Schwartzian 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 14.37. (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

$$\xi_0 : \Lambda_0 \to \Lambda_1, \qquad \xi_1 : \Lambda_1 \to \Lambda_0, \tag{14.24}$$

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[3]

$$[\xi_1, \xi_0] = (S_1 - S_0)^{-1} \dot{S}_1 (S_1 - S_0)^{-1} \dot{S}_0.$$

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 $t \to 0$

$$[\underline{\dot{\Lambda}}(t), \underline{\dot{\Lambda}}(0)] \simeq \frac{1}{t^2} Id + \frac{1}{3} R_\Lambda(0) + O(t). \tag{14.25}$$

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 (Σ, σ) be a symplectic vector space and $L(\Sigma)$ denote the Lagrange Grassmannian. We start by describing a natural subspace of $L(\Sigma)$ associated with an isotropic subspace Γ of Σ. This will allow us to define a reduction procedure for a non regular curve.

Let Γ be a k-dimensional isotropic subspace of Σ, 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 with the restriction of σ.

Lemma 14.38. *There is a natural identification of $L(\Gamma^\angle/\Gamma)$ as a subspace of $L(\Sigma)$:*

$$L(\Gamma^\angle/\Gamma) \simeq \{\Lambda \in L(\Sigma), \Gamma \subset \Lambda\} \subset L(\Sigma). \tag{14.26}$$

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 Λ is Lagrangian, $\Lambda = \Lambda^\angle \subset \Gamma^\angle$, hence the identification (14.26).

Assume now that $\Lambda \in L(\Gamma^\angle/\Gamma)$ and let us show that $\pi^\Gamma(\Lambda) = \Lambda$, i.e. π^Γ is a projection. Indeed from the inclusions $\Gamma \subset \Lambda \subset \Gamma^\angle$ one has $\pi^\Gamma(\Lambda) = \Lambda^\Gamma = (\Lambda \cap \Gamma^\angle) + \Gamma = \Lambda + \Gamma = \Lambda$.

[3]here $\dot{S}_i$ denotes the matrix associated with ξ_i.

We are left to check that Λ^Γ 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. (The identity $(\Lambda \cap \Gamma^\angle) + \Gamma = (\Lambda + \Gamma) \cap \Gamma^\angle$ is also a consequence of the same exercise.) □

Remark 14.39. Let $\Gamma^\pitchfork = \{\Lambda \in L(\Sigma), \Lambda \cap \Gamma = \{0\}\}$. The restriction $\pi^\Gamma\big|_{\Gamma^\pitchfork}$ is smooth. Indeed it can be shown that π^Γ is defined by a rational function, since it is expressed via the solution of a linear system.

The following example shows that the projection π^Γ is not globally continous on $L(\Sigma)$.

Example 14.40. Consider the symplectic structure σ on $\mathbb{R}^4$, with 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

$$\Lambda_\varepsilon = \operatorname{span}\{e_1 + \varepsilon f_2, e_2 + \varepsilon f_1\}, \qquad \forall\, \varepsilon > 0.$$

It is easy to see that Λ_ε is Lagrangian for every ε and that

$$\begin{aligned}\Lambda_\varepsilon^\Gamma &= \operatorname{span}\{e_1, f_2\}, \qquad \forall\, \varepsilon > 0, \\ \Lambda_0^\Gamma &= \operatorname{span}\{e_1, e_2\}.\end{aligned} \tag{14.27}$$

Indeed $f_2 \in e_1^\angle$, 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 (14.27) holds. The case $\varepsilon = 0$ is trivial.

14.5 Ample curves

In this section we introduce ample curves.

Definition 14.41. 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

$$\Sigma = \operatorname{span}\{\lambda^{(i)}(t_0) \mid \lambda(t) \in \Lambda(t), \lambda(t) \text{ smooth}, 0 \le i \le N\}. \tag{14.28}$$

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

$$\Sigma = \{(p,q) \mid p, q \in \mathbb{R}^n\}, \qquad \Lambda(t) = \{(p, S(t)p) \mid p \in \mathbb{R}^n\}.$$

Exercise 14.42. 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:

$$\operatorname{rank}\{\dot S(0), \ddot S(0), \dots, S^{(N)}(0)\} = n. \tag{14.29}$$

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 14.43. *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. $\square$

Exercise 14.44. Prove that a monotone curve $\Lambda(t)$ is ample at t_0 if and only if one of the 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 an *analytic* monotone curve on $L(\Sigma)$. Without loss of generality we can assume the curve to be non increasing, 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 Υ_t is a decreasing family of subspaces, i.e. $\Upsilon_t \subset \Upsilon_\tau$ if $\tau \leq t$. Hence the family Υ_t for $t \to 0$ stabilizes and the limit subspace Υ is well defined

$$\Upsilon := \lim_{t \to 0} \Upsilon_t$$

The symplectic reduction of the curve by the isotropic subspace Υ defines a new curve $\widetilde{\Lambda}(t) := \Lambda(t)^{\Upsilon} \in L(\Upsilon^{\angle}/\Upsilon)$.

Proposition 14.45. *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

$$\widetilde{\Lambda}(0) \cap \widetilde{\Lambda}(t) = \{0\}, \qquad \text{in} \quad L(\Upsilon^{\angle}/\Upsilon) \tag{14.30}$$

In particular, if $\widetilde{S}(t)$ denotes the symmetric matrix representing $\widetilde{\Lambda}(t)$ such that $\widetilde{S}(0) = \widetilde{\Lambda}(t_0)$, 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. $\square$

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 14.46. *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 $\underline{\dot{\Lambda}}^{\Gamma}(0) > 0$.*

Before proving Theorem 14.46, let us discuss two useful lemmas.

Lemma 14.47. *Let $v_1(t), \ldots, v_k(t) \in \mathbb{R}^n$ 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) $\operatorname{span}\{v_i(\tau) \,|\, i = 1, \ldots, k; \tau \in [0,t]\} = \mathbb{R}^n$.

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 $\operatorname{im}_{\tau\in[0,t]} V(\tau) = \operatorname{span}\{v_i(\tau) \,|\, i = 1, \ldots, k,\ \tau \in [0,t]\} = \mathbb{R}^n$, that is a contradiction. The converse is similar. □

Lemma 14.48. *Let A, B 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. □

Proof of Theorem 14.46. 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 Σ such that

$$\Sigma = \{(p, q) : p, q \in \mathbb{R}^n\}, \qquad \Lambda(t) = \{(p, S(t)p)|\, p \in \mathbb{R}^n\}, \qquad S(0) = 0.$$

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 $p = (p_1, p_2)$ and $q = (q_1, q_2)$. The subspaces Γ and $\Gamma^\angle$ are described by the equations

$$\Gamma = \{(p, q) : p_1 = 0, q = 0\}, \qquad \Gamma^\angle = \{(p, q) : q_2 = 0\}$$

and (p_1, q_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

$$S(t) = \begin{pmatrix} S_{11}(t) & S_{12}(t) \\ S_{12}^*(t) & S_{22}(t) \end{pmatrix}, \qquad \text{with} \qquad \dot{S}(0) = \begin{pmatrix} \mathbb{I}_k & 0 \\ 0 & 0 \end{pmatrix}. \tag{14.31}$$

where $\mathbb{I}_k$ is the $k \times k$ identity matrix. Finally, from the fact that S is monotone and ample, that implies $S(t) > 0$ for each $t > 0$, it follows

$$S_{11}(t) > 0, \qquad S_{22}(t) > 0, \qquad \forall\, t > 0. \tag{14.32}$$

Then we can compute the coordinate expression of the reduced curve, i.e. the matrix $S^\Gamma(t)$ such that

$$\Lambda^\Gamma(t) = \{(p_1, S^\Gamma(t)p_1), p_1 \in \mathbb{R}^k\}.$$

From the identity

$$\Lambda(t) \cap \Gamma^\angle = \{(p, S(t)p), S(t)p \in \mathbb{R}^k\} = \left\{ \left(S^{-1}(t) \begin{pmatrix} q_1 \\ 0 \end{pmatrix}, \begin{pmatrix} q_1 \\ 0 \end{pmatrix} \right), q_1 \in \mathbb{R}^k \right\} \tag{14.33}$$

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 recovered simply by considering it as a map of (p_1, q_1) only, i.e.

$$S(t)p = \begin{pmatrix} S_{11} & S_{12} \\ S_{12}^* & S_{22} \end{pmatrix} \begin{pmatrix} p_1 \\ p_2 \end{pmatrix} = \begin{pmatrix} S_{11}p_1 + S_{12}p_2 \\ S_{12}^*p_1 + S_{22}p_2 \end{pmatrix}$$

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

$$S^\Gamma = S_{11} - S_{12}S_{22}^{-1}S_{12}^*. \tag{14.34}$$

(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 the 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 inequality $0 \leq S^\Gamma(t) \leq S_{11}(t)$.

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 (14.34) and the fact that S_{22} is positive definite (and so S_{22}^{-1}) one gets

$$\langle (S_{11} - S^\Gamma)p_1, p_1 \rangle = \langle S_{12}S_{22}^{-1}S_{12}^* p_1, p_1 \rangle = \langle S_{22}^{-1}(S_{12}^* p_1), (S_{12}^* p_1) \rangle \geq 0.$$

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 14.46.

Lemma 14.49. *Let $A(t)$ be a smooth family of symmetric nonnegative $n \times n$ matrices. If the condition* $\operatorname{rank}(A, \dot{A}, \ldots, A^{(N)})|_{t=0} = n$ *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 (14.31))

$$\langle S(t)p, p \rangle > \varepsilon t |p_1|^2$$

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 $(p, S(t)p) \in \Lambda(t) \cap \Gamma^\angle$ (i.e. $S(t)p = (q_1\ 0)^T$ or equivalently $S(t)^{-1}(q_1, 0) = (p_1, p_2)$). This implies in particular that $S^\Gamma(t)p_1 = q_1$ and

$$\langle S^\Gamma(t)p_1, p_1 \rangle = \langle S(t)p, p \rangle \geq \varepsilon t |p_1|^2,$$

This identity can be rewritten as $S^\Gamma(t) > \varepsilon t\, \mathbb{I}_k > 0$ and implies by Lemma 14.48

$$0 < S^\Gamma(t)^{-1} < \frac{1}{\varepsilon t}\mathbb{I}_k$$

which completes the proof. □

Proof of Lemma 14.49. We reduce the proof of the Lemma to the following statement:

Claim II. *There exists $c, \widehat{N} > 0$ such that for any sufficiently small $\varepsilon, t > 0$*

$$\det\left(\int_0^t A(\tau) - \varepsilon A(0)\, d\tau\right) > c\, t^{\widehat{N}}.$$

Moreover $c, \widehat{N}$ depends only on the $2N$-th Taylor polynomial of $A(t)$.

Indeed fix $t_0 > 0$. Since $A(t) \geq 0$ and $A(t)$ is not the zero family, then $\int_0^{t_0} A(\tau)d\tau > 0$. Hence, for a fixed t_0, there exists ε small enough such that $\int_0^{t_0} A(\tau) - \varepsilon A(0)\, d\tau > 0$. Assume now that the matrix $S_t = \int_0^t A(\tau) - \varepsilon A(0)\, d\tau > 0$ is not strictly positive for some $0 < t < t_0$, then $\det S(\tau) = 0$ 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 continuity 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 **??**), 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 $\widehat{A}(t)$ satisfies the assumption of Lemma 14.47, 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

$$\widehat{a}_{ij}(t) = c_{ij}t^{m_i+m_j}, \qquad a_{ij}(t) = \widehat{a}_{ij}(t) + O(t^{m_i+m_j+1}),$$

for suitable constants c_{ij} (some of them may be zero).

Then we let $A^\varepsilon(t) := A(t) - \varepsilon A(0) = \{a^\varepsilon_{ij}(t)\}_{i,j=1}^n$. Of course $a^\varepsilon_{ij}(t) = c^\varepsilon_{ij}t^{m_i+m_j} + O(t^{m_i+m_j+1})$ where

$$c^\varepsilon_{ij} = \begin{cases} (1-\varepsilon)c_{ij}, & \text{if } m_i + m_j = 0, \\ c_{ij}, & \text{if } m_i + m_j > 0. \end{cases}$$

From the equality

$$\int_0^t a^\varepsilon_{ij}(\tau)d\tau = t^{m_i+m_j+1}\left(\frac{c^\varepsilon_{ij}}{m_i + m_j + 1} + O(t)\right)$$

one gets

$$\det\left(\int_0^t A^\varepsilon(\tau)d\tau\right) = t^{n+2\sum_{i=1}^N m_i}\left(\det\left(\frac{c^\varepsilon_{ij}}{m_i+m_j+1}\right) + O(t)\right)$$

On the other hand

$$\det\left(\int_0^t \widehat{A}(\tau)d\tau\right) = t^{n+2\sum_{i=1}^N m_i}\left(\det\left(\frac{c_{ij}}{m_i+m_j+1}\right) + O(t)\right) > 0$$

hence $\det\left(\frac{c^\varepsilon_{ij}}{m_i+m_j+1}\right) > 0$ for small ε. The proof is completed by setting

$$c := \det\left(\frac{c_{ij}}{m_i+m_j+1}\right), \qquad \widehat{N} := n + 2\sum_{i=1}^{N} m_i$$

$\square$

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 14.50. Let $\Lambda(t)$ be a monotone curve in $L(\Sigma)$. We say that $\Lambda(t)$ *is conjugate to* $\Lambda(0)$ if $\Lambda(t) \cap \Lambda(0) \neq \{0\}$.

As a consequence of Lemma 14.43, we have the following immediate corollary.

Corollary 14.51. *Conjugate points on a monotone and ample curve in $L(\Sigma)$ are isolated.*

The following two results describe general properties of conjugate points

Theorem 14.52. *Let $\Lambda(t), \Delta(t)$ two ample monotone curves in $L(\Sigma)$ defined on $\mathbb{R}$ such that*

(i) $\Sigma = \Lambda(t) \oplus \Delta(t)$ *for every* $t \geq 0$,

(ii) $\dot{\Lambda}(t) \leq 0$, $\dot{\Delta}(t) \geq 0$, *as quadratic forms.*

Then there exists no $\tau > 0$ such that $\Lambda(\tau)$ is conjugate to $\Lambda(0)$. Moreover $\exists \lim_{t\to+\infty} \Lambda(t) = \Lambda(\infty)$.

Proof. Fix coordinates induced by some Lagrangian splitting of Σ 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 14.53. *Let $\Lambda_s(t)$, for $t, s \in [0,1]$ be an homotopy of curves in $L(\Sigma)$ such that $\Lambda_s(0) = \Lambda$ for $s \in [0,1]$. Assume that*

(i) $\Lambda_s(\cdot)$ *is monotone and ample for every* $s \in [0,1]$,

(ii) $\Lambda_0(\cdot), \Lambda_1(\cdot)$ *and* $\Lambda_s(1)$, *for* $s \in [0,1]$, *contains no conjugate points to* Λ.

Then no curve $t \mapsto \Lambda_s(t)$ contains conjugate points to Λ.

Proof. Let us consider the open chart $\Lambda^{\pitchfork}$ defined by all the Lagrangian subspaces traversal to Λ. The statement is equivalent to prove that $\Lambda_s(t) \in \Lambda^{\pitchfork}$ for all $t > 0$ and $s \in [0,1]$. Let us fix coordinates induced by some Lagrangian splitting $\Sigma = \Lambda \oplus \Delta$ in such a way that $\Lambda = \{(p,0)\}$ and

$$\Lambda_s(t) = \{(B_s(t)q, q)\}$$

for all s and $t > 0$ (at least for t small enough, indeed by ampleness $\Lambda_s(t) \in \Lambda^{\pitchfork}$ for t small). Moreover we can assume that $B_s(t)$ is a monotone increasing family of symmetric matrices.

Notice that $x^T B_s(\tau)x \to -\infty$ for every $x \in \mathbb{R}^n$ when $\tau \to 0^+$, due to the fact that $\Lambda_s(0) = \Lambda$ is out of the coordinate chart. Moreover, a necessary condition for $\Lambda_s(t)$ to be conjugate to Λ 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 increasing and bounded up to $t = 1$. Hence the continuous family of values $M_s := x^T B_s(1)x$ is weel defined and bounded for all s. The monotonicity implies that actually $x^T B_s(t)x < +\infty$ for all values of $t, s \in [0,1]$. (See also Figure 14.7).

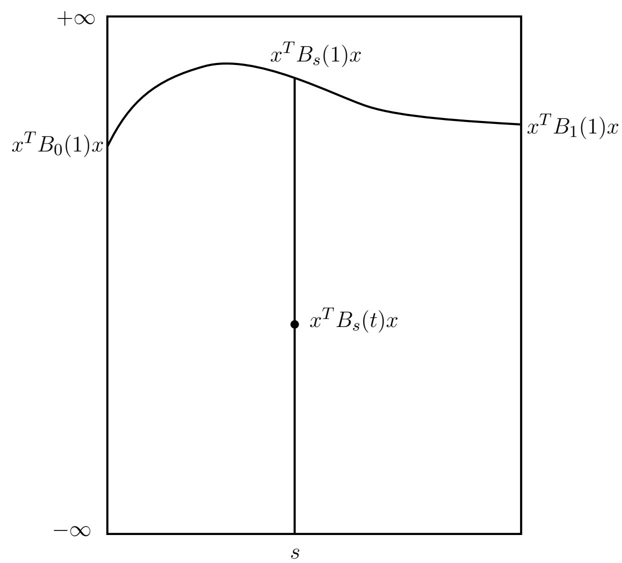


Figure 14.1: Proof of Theorem 14.53

∎

14.8 Comparison theorems for regular curves

In this last section we prove two comparison theorems for regular monotone curves in the Lagrange Grassmannian.

Corollary 14.54. *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 $\Lambda(0)$.*

Proof. This is a direct consequence of Theorem 14.52 ∎

Theorem 14.55. *Let $\Lambda(t)$ be a monotone and regular curve in the Lagrange Grassmannian. Assume that there exists $k \geq 0$ such that for all $t \geq 0$*

(i) $R_\Lambda(t) \leq k\,\mathrm{Id}$. *Then, if $\Lambda(t)$ is conjugate to $\Lambda(0)$, we have $t \geq \frac{\pi}{\sqrt{k}}$.*

(ii) $\frac{1}{n}\mathrm{trace}\, R_\Lambda(t) \geq k$. *Then for every $t \geq 0$ there exists $\tau \in [t, t + \frac{\pi}{\sqrt{k}}]$ such that $\Lambda(\tau)$ is conjugate to $\Lambda(0)$.*

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 14.56. Notice that the estimates of Theorem 14.55 are sharp, as it is immediately seen by considering the example of a 1-dimensional curve of constant velocity (see Example 14.35).

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 φ is

$$R_\varphi(t) = -\frac{k}{(1 + kt^2)^2}.$$

Thus using φ as a reparametrization we find, by Proposition 14.36

$$\begin{aligned} R_{\Lambda_\varphi}(t) &= \dot{\varphi}^2 R_\Lambda(\varphi(t)) + R_\varphi(t)\mathrm{Id} \\ &= \frac{1}{(1 + kt^2)^2}(R_\Lambda(\varphi(t)) - k\mathrm{Id}) \leq 0. \end{aligned}$$

By Corollary 14.54 the curve $\Lambda \circ \varphi$ has no conjugate points, i.e. Λ 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 one 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 equality

$$\dot{B}(t) = B^2(t) + R_\Lambda(t),$$

we get

$$\dot{b}(t) = \operatorname{trace}(B^2(t)) + \operatorname{trace} R_\Lambda(t). \tag{14.35}$$

Lemma 14.57. *For every $n \times n$ symmetric matrix S the following inequality holds true*

$$\operatorname{trace}(S^2) \geq \frac{1}{n}(\operatorname{trace} S)^2. \tag{14.36}$$

Proof. For every symmetric matrix S there exists a matrix M such that $MSM = D$ is diagonal. Since $\operatorname{trace}(MAM^{-1}) = \operatorname{trace}(A)$ for every matrix A, it is enough to prove the inequality (14.36) for a diagonal matrix $D = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$. In this case (14.36) reduces to the 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 14.57 to (14.35) and using the assumption (ii) one gets

$$\dot{b}(t) \geq \frac{1}{n} b^2(t) + nk, \tag{14.37}$$

By standard results in ODE theory we have $b(t) \geq \varphi(t)$, where $\varphi(t)$ is the solution of the differential equation

$$\dot{\varphi}(t) = \frac{1}{n} \varphi^2(t) + nk \tag{14.38}$$

The solution for (14.38), with initial datum $\varphi(t_0) = 0$, is explicit and 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. $\square$

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 λ, 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 π_* gives an isomorphisms 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 **??** 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 λ:

$$e_*^{-t\vec{H}} : T_{\lambda(t)}(T^*M) \to T_\lambda(T^*M).$$

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 $\mathcal{V}_\lambda := T_\lambda(T_q^*M)$ for the vertical subspace at the point $\lambda \in T^*M$, i.e. the tangent space at λ to the fiber T_q^*M, 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 $\mathcal{V}_\lambda \simeq T_q^*M$.

Definition 15.1. Let $\lambda \in T^*M$. The *Jacobi curve* at the point λ 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 $J_\lambda(0) = \mathcal{V}_\lambda = T_\lambda(T_q^*M)$ 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) $\underline{\dot{J}}_\lambda(0) = -2H|_{T_q^*M}$ *as quadratic forms on* $\mathcal{V}_\lambda \simeq T_q^*M$.

(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 ξ as a constant vertical vector field) is a smooth extension of ξ, i.e. it satisfies $\xi(0) = \xi$ and $\xi(t) \in J_\lambda(t)$. Therefore, by (14.8)

$$\dot{J}_\lambda(0)\xi = \sigma(\xi, \dot{\xi}) = \sigma\left(\xi, \left.\frac{d}{dt}\right|_{t=0} e_*^{-t\vec{H}}\xi\right) = \sigma(\xi, [\vec{H}, \xi]). \tag{15.2}$$

To compute the last quantity we use the following elementary, although very useful, property of the symplectic form σ.

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 definition 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 λ. Denote by $J_\lambda(t)$ its Jacobi curve. Then for $s > 0$*

$$\gamma(s) \text{ is conjugate to } \gamma(0) \qquad \Longleftrightarrow \qquad J_\lambda(s) \text{ is conjugate to } J_\lambda(0).$$

Proof. By Definition 8.41, $\gamma(s)$ 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 λ, i.e. the image of the differential $e_*^{s\vec{H}}$ has a nontrivial intersection with the kernel of the projection π_*

$$e_*^{s\vec{H}} J_\lambda(0) \cap T_{\lambda(s)} T^*_{\gamma(s)} M \neq \{0\}. \tag{15.5}$$

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 $J_\lambda(0)$. □

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.

Theorem 15.7. *Let $\gamma : [0,1] \to M$ be a normal extremal path such that $\gamma|_{[0,s]}$ is not abnormal for all $0 < s \leq 1$. 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 assume that $S(t) \leq 0$ for every $t_0 \leq t \leq t_1$, i.e. is non positive definite and monotone decreasing, [2] 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 $0 \leq \tau \leq \tau$, 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 vector field ξ is preserved by the Hamiltonian field, i.e. $e_*^{t\vec{H}}\xi = \xi$, that implies $[\vec{H}, \xi](\lambda(t)) = 0$. Then the statement is proved by the following

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 $[\vec{H}, \xi](\lambda(t)) = 0$ 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 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.

□

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 $\gamma|_{[0,\tau]}$ is not abnormal for all $0 \leq \tau \leq t$ if and only if $J_\lambda(\tau) \cap J_\lambda(0) = \{0\}$ for all $0 \leq \tau \leq t$.*

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 λ. 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.

[2] Indeed it is proved that the only invariant of a pair of two Lagrangian subspaces in a symplectic space is the dimension of the intersection, i.e. the rank of the difference $\mathrm{rank}(S(t) - S(0))$. Add exercise

Remark 15.11. This procedure was already exploited in Section 8.11, obtained by a direct argument via Proposition 8.38. 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* $\vec{E} \in \mathrm{Vec}(T^*M)$ is the vertical vector field defined by

$$\vec{E}(\lambda) = \left.\frac{d}{ds}\right|_{s=1} \delta_s(\lambda), \qquad \lambda \in T^*M.$$

It is easy to see that in canonical coordinates (x, ξ) it satisfies $\vec{E} = \sum_{i=1}^n \xi_i \frac{\partial}{\partial \xi_i}$ and the following identity holds

$$e^{t\vec{E}}\lambda = e^t\lambda, \qquad \text{i.e. } e^{t\vec{E}}(\xi, x) = (e^t\xi, x).$$

Exercise 15.14. Prove that the Euler vector field is characterized by the identity

$$i_{\vec{E}}\,\sigma = s, \qquad\qquad s = \text{ Liouville 1-form in } T^*M.$$

Lemma 15.15. *We have the identity $e_*^{-t\vec{H}}\vec{E} = \vec{E} - t\vec{H}$. In particular $[\vec{H}, \vec{E}] = -\vec{H}$.*

Proof. The homogeneity property (8.50) 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 δ_{-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 (2.27), that $[\vec{H}, \vec{E}] = -\vec{H}$. Thus, by (2.31) we have $e_*^{-t\vec{H}}\vec{E} = \vec{E} - t\vec{H}$, since every higher order commutator vanishes.

□

Proposition 15.16. *The subspace $\widetilde{\Sigma} = \mathrm{span}\{\vec{E}, \vec{H}\}$ is invariant under the action of the Hamiltonian flow. Moreover $\{\vec{E}, \vec{H}\}$ 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

$$e_*^{-t\vec{H}}\vec{E} = \vec{E} - t\vec{H}, \qquad e_*^{-t\vec{H}}\vec{H} = 0.$$

Moreover, on the level set $H^{-1}(1/2)$, we have by homogeneity of H w.r.t. p:

$$\sigma(\vec{E},\vec{H}) = \vec{E}(H) = \left.\frac{d}{dt}\right|_{t=0} H(e^{t\vec{E}}(p,x)) = p\frac{\partial H}{\partial p} = 2H = 1. \tag{15.10}$$

It follows that $\{\vec{E},\vec{H}\}$ 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 and we could split our Jacobi curve.

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

$$J_\lambda(t) = \widehat{J}_\lambda(t) \oplus \mathbb{R}(\vec{E} - t\vec{H}).$$

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 $\widetilde{\Sigma}$ is always zero.

(ii) Prove that $J_\lambda(0) \cap J_\lambda(s) \neq \{0\}$ if and only if $\widehat{J}_\lambda(0) \cap \widehat{J}_\lambda(s) \neq \{0\}$.

Chapter 16

Riemannian curvature

On a manifold, in general there is no canonical method for identifying tangent spaces at different points, (or more generally fibers of a vector bundle at different points). Thus, we have to expect that a notion of derivative for vector fields (or sections of a vector bundle), has to depend on certain choices.

In our presentation we introduce the general notion of *Ehresmann connection* and we then we discuss how this notion is related with the notion of parallel transport and covariant derivative usually introduced in classical Riemannian geometry.

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 π_*, 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 $\ker \pi_{*,z} = \mathcal{V}_z$.

Notation. In the following we will refer also at ∇ as the connection on E.

Given a smooth curve $\gamma : [0,T] \to M$ on the manifold M, the connection let us to define the *parallel transport* along γ, i.e. a way to identify tangent vectors belonging to tangent spaces at different points of the curve. Let X_t be a nonautonomous smooth vector field defined on a neighborhood of γ, that is an extension of the velocity vector field of the curve[1], i.e. such that

$$\dot{\gamma}(t) = X_t(\gamma(t)), \qquad \forall\, t \in [0,T].$$

Then consider the non autonomous vector field $\nabla_{X_t} \in \mathrm{Vec}(E)$ obtained by its lift.

Definition 16.3. Let $\gamma : [0,T] \to M$ be a smooth curve. The *parallel transport* along γ is the map Φ defined by the flow of ∇_{X_t}

$$\Phi_{t_0,t_1} := \overrightarrow{\exp} \int_{t_0}^{t_1} \nabla_{X_s} ds : E_{\gamma(t_0)} \to E_{\gamma(t_1)}, \qquad \text{for } 0 < t_0 < t_1 < T. \tag{16.2}$$

In the general case we need some extra assumptions on the vector field to ensure that (16.2) exists (even for small time $t > 0$) since the existence time of a solution also depend on the point on the fiber. For instance if we the fibers are compact, then it is possible to find such $t > 0$.

Exercise 16.4. Show that the parallel transport map 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 ∇ be a 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 computes if the same holds true for two horizontal vector fields.

Definition 16.5. Let E be a smooth fiber bundle and ∇ 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.

Notice that, given a vector field on E, its horizontal part coincide, 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 and $R \equiv 0$ if and only if the horizontal distribution $\mathcal{H}$ is involutive.

The curvature $R(X,Y)$ is also rewritten in the following more classical way

$$\begin{aligned} R(X,Y) &= [\nabla_X, \nabla_Y] - \nabla_{[X,Y]}. \\ &= \nabla_X \nabla_Y - \nabla_X \nabla_Y - \nabla_{[X,Y]}. \end{aligned}$$

Next we show that R is actually a tensor on T_qM, i.e. the value of $R(X,Y)$ at q depends only on the value of X and Y at the point q.

[1]this is always possible with a (maybe non autonomous) vector field.

Proposition 16.6. *R is a skew symmetric tensor on M.*

Proof. The skew-symmetry is immediate. To prove that the value of $R(X,Y)$ at q depends only on the value of X and Y at the point q, it is sufficient to prove that R is linear on functions. By skew-symmetry, we are reduced to prove that R is linear in the first argument, namely

$$R(aX,Y) = aR(X,Y), \qquad \text{where} \quad a \in C^\infty(M).$$

Notice that the symbol a in the right hand side stands for the function $\pi^*a = a \circ \pi$ in $C^\infty(E)$, that is constant on fibers.

By definition of lift of a vector field it is easy to prove the identities $\nabla_{aX} = a\nabla_X$ and $\nabla_X a = Xa$ for every $a \in C^\infty(M)$. Applying the definition of ∇ and the Leibnitz 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 E is a *vector* bundle on M (i.e. each fiber $E_q = \pi^{-1}(q)$ has a natural structure of vector space). In this case it is natural to introduce a notion of *linear* Ehresmann connection ∇ on E.

Given a vector bundle $\pi : E \to M$, we denote by $C^\infty_L(E)$ the set of smooth functions on E that are linear on fibers.

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 2.48). 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 $a \in C^\infty(M) \mapsto \pi^*a \in C^\infty(E)$.

Exercise 16.8. Show that a vector field on E is the lift of a vector field on M if and only if, as a differential operator on $C^\infty(E)$, it maps the subspace $C^\infty(M)$ into itself.

After this discussion it is natural to give the following definition.

Definition 16.9. A *linear connection* on a vector bundle E on the base M is an Ehresmann connection ∇ such that the lift ∇_X of a vector field $X \in \mathrm{Vec}(M)$ satisfies the following property: for every $a \in C^\infty_L(E)$ it holds $\nabla_X a \in C^\infty_L(E)$.

Remark 16.10. Given a local basis of vector fields $X_1, \dots, X_n$ on M we can build dual coordinates $(u_1, \dots, u_n)$ on the fibers of E defining the functions $u_i(z) = \langle z, X_i(q)\rangle$ where $q = \pi(z)$. In this way

$$E = \{(u,q), q \in M, u \in \mathbb{R}^n\},$$

and the tangent space to E is splitted in $T_zE \simeq T_qM \oplus T_zE_q$. A connection on E is determined by the lift of the vector fields $X_i, i = 1, \ldots, n$ on the base manifold (recall that $\pi_* \nabla_{X_i} = X_i$)

$$\nabla_{X_i} = X_i + \sum_{j=1}^{n} a_{ij}(u,q)\partial_{u_j}, \qquad i = 1, \ldots, n, \tag{16.4}$$

where $a_{ij} \in C^\infty(E)$ are suitable smooth functions. Then ∇ is linear if and only if for every i, j the function $a_{ij}(u,q) = \sum_{k=1}^{n} \Gamma_{ij}^k(q) u_k$ is linear with respect to u .

The smooth functions Γ_{ij}^k are also called the *Christoffel symbols* of the linear connection.

Exercise 16.11. Let γ 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 γ are written as $\dot u_j = \sum_{i,k} \Gamma_{ij}^k v_i u_k$ where $(u_1, \ldots, u_n)$ are the vertical coordinates of ξ.

Notice that, for a linear connection, the parallel transport is defined by a first order linear (nonautonomous) ODE. The existence of the flow is then guaranteed from stantard results form ODE theory. Moreover, when it exists, the map Φ_{t_0,t_1} is a linear transformation between fibers.

16.1.3 Covariant derivative and torsion for linear connections

Once a connection on a linear vector bundle E is given, we have a well defined linear parallel transport map

$$\Phi_{t_0,t_1} := \overrightarrow{\exp} \int_{t_0}^{t_1} \nabla_{X_s} ds : E_{\gamma(t_0)} \to E_{\gamma(t_1)}, \qquad \text{for } 0 < t_0 < t_1 < T. \tag{16.5}$$

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 exchange of t_0, t_1 in the integral)

$$\Phi^*_{t_0,t_1} := \left(\overrightarrow{\exp} \int_{t_1}^{t_0} \nabla_{X_s} ds\right)^* : E^*_{\gamma(t_0)} \to E^*_{\gamma(t_1)}, \qquad \text{for } 0 < t_0 < t_1 < T. \tag{16.6}$$

The infinitesimal generator of this "adjoint" flow defines a linear parallel transport, hence a linear connection, on the dual bundle E^*.

In what follows we will restrict our attention to the case of the vector bundle $E = T^*M$ and we assume that a linear connection ∇ 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 every vector field $Y \in \mathrm{Vec}(M)$ let us denote with $Y^* \in C^\infty(T^*M)$ the function

$$Y^*(\lambda) = \langle \lambda, Y(q) \rangle, \quad q = \pi(z),$$

namely the smooth function on E associated with Y 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 16.12. Let $X, Y \in \mathrm{Vec}(M)$. We define $\nabla_X Y = Z$ if and only if $\nabla_X Y^* = Z^*$ with $Z \in \mathrm{Vec}(M)$.

Notice that the definition is well-posed since ∇ is linear, hence $\nabla_X Y^*$ is a linear function and there exists $Z \in \mathrm{Vec}(M)$ such that $\nabla_X Y^* = Z^*$.[2]

Lemma 16.13. *Let $\{X_1, \dots, X_n\}$ be a local frame on M. Then $\nabla_{X_i} X_j = \Gamma_{ij}^k X_k$, where Γ_{ij}^k are the Christoffel symbols of the connection ∇.*

Proof. Let us prove this in the coordinates dual to our frame. In these coordinates, the linear connection is specified by the lifts

$$\nabla_{X_i} = X_i + \Gamma_{ij}^k u_k \partial_{u_j}, \qquad \text{where} \qquad u_j(\lambda) = \langle \lambda, X_j \rangle .$$

Moreover $X_j^* = u_j$. Hence it is immediate to show $\nabla_{X_i} X_j^* = \Gamma_{ij}^k X_k^*$, and the lemma is proved. $\square$

We now introduce the torsion tensor of a linear connection on T^*M. As usual, σ denotes the canonical symplectic structure on T^*M.

Definition 16.14. The *torsion* of a linear connection ∇ is the map $T : \mathrm{Vec}(M)^2 \to \mathrm{Vec}(M)$ defined by the identity

$$T(X,Y)^* := \sigma(\nabla_X, \nabla_Y), \qquad \forall\, X, Y \in \mathrm{Vec}(M). \tag{16.7}$$

It is easy to check that T is actually a tensor, i.e. the value of $T(X,Y)$ at a point q depends only on the values of X, Y at the point. 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 $T \equiv 0$.

The classical formula for the torsion tensor, in terms of the covariant derivative, is recovered in the following lemma.

Lemma 16.15. *The torsion tensor satisfies the identity*

$$T(X,Y) = \nabla_X Y - \nabla_Y X - [X,Y]. \tag{16.8}$$

Proof. We have to prove that $T(X,Y)^* = \nabla_X Y^* - \nabla_Y X^* - [X,Y]^*$. Notice that by definition of the Liouville 1-form $s \in \Lambda^1(T^*M)$, $s_\lambda = \lambda \circ \pi_*$ we have $X^*(\lambda) = \langle \lambda, X \rangle = \langle s_\lambda, \nabla_X \rangle$. Then we have, using that $\sigma = ds$ and the Cartan formula (4.77)

$$\begin{aligned} T(X,Y)^* &= ds(\nabla_X, \nabla_Y) \\ &= \nabla_X \langle s, \nabla_Y \rangle - \nabla_Y \langle s, \nabla_X \rangle - \langle s, [\nabla_X, \nabla_Y] \rangle \\ &= \nabla_X \langle s, \nabla_Y \rangle - \nabla_Y \langle s, \nabla_X \rangle - \left\langle s, \nabla_{[X,Y]} \right\rangle \\ &= \nabla_X Y^* - \nabla_Y X^* - [X,Y]^*, \end{aligned}$$

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. $\square$

Exercise 16.16. Show that a linear connection ∇ on a vector bundle E satisfies the following Leibnitz rule

$$\nabla_X(aY) = a\nabla_X Y + (Xa)Y, \qquad \text{for each } a \in C^\infty(M).$$

[2]There is no confusion in the notation above since, by definition, ∇_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.

16.2 Riemannian connection

In this section we want to introduce the Levi-Civita connection on a Riemannian manifold M by defining an Ehresmann connection on T^*M via the Jacobi curve approach.

Recall that every Jacobi curve associated with a trajectory on a Riemannian manifold is regular. Moreover, as showed in Chapter 14, every regular curve in the Lagrangian Grassmannian admits a derivative curve, which defines a canonical complement to the curve itself. Hence, following this approach, it is natural to introduce the Riemannian connection at $\lambda \in T^*M$ as the canonical complement to the Jacobi curve defined at λ.

Definition 16.17. The *Levi-Civita connection* on T^*M is the Ehresmann connection $\mathcal{H}$ is defined by

$$\mathcal{H}_\lambda = J^\circ_\lambda(0), \qquad \lambda \in T^*M,$$

where as usual $J_\lambda(t)$ denotes the Jacobi curve defined at the point $\lambda \in T^*M$ and J°_λ denotes its derivative curve.

The next proposition characterizes the Levi-Civita connection as the unique linear connection on T^*M that is linear, metric preserving and torsion free.

Proposition 16.18. *The Levi-Civita connection satisfies the following properties:*

(i) is a linear connection,

(ii) is torsion free,

(iii) 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, i.e.

$$\mathcal{H}_{c\lambda} = \delta_{c*}\mathcal{H}_\lambda, \qquad \forall\, c > 0. \tag{16.9}$$

Indeed in this case the functions $a_{ij} \in C^\infty(T^*M)$ defining the connection (see (16.4)) are 1-homogeneous, hence linear as a consequence of Exercise 16.19.

Let us prove (16.9). The differential of the dilation on the fibers $\delta_c : T^*M \to T^*M$ satisfies the property $\delta_{c*}(T_\lambda(T^*_qM)) = T_{c\lambda}(T^*_qM)$. From this identity and differentiating the identity

$$e^{t\vec{H}} \circ \delta_c = \delta_c \circ e^{ct\vec{H}}, \qquad \forall\, c > 0, \tag{16.10}$$

one easily gets that

$$J_{c\lambda}(t) = \delta_{c*} J_\lambda(ct), \qquad \forall\, t \geq 0, \lambda \in T^*M. \tag{16.11}$$

Indeed one has the following chain of identities

$$\begin{aligned} J_{c\lambda}(t) &= e^{-t\vec{H}}_*(T_{c\lambda}(T^*_qM)) \\ &= e^{-t\vec{H}}_* \circ \delta_{c*}(T_\lambda(T^*_qM)) \qquad \text{(by (16.10))} \\ &= \delta_{c*} \circ e^{-ct\vec{H}}_*(T_\lambda(T^*_qM)) \\ &= \delta_{c*} J_\lambda(ct). \end{aligned}$$

Now we show that the same relation holds true also for the derivative curve, i.e.

$$J^\circ_{c\lambda}(t) = \delta_{c*} J^\circ_\lambda(ct), \qquad \forall\, t \geq 0,\, \lambda \in T^*M. \tag{16.12}$$

Indeed one can check in coordinates (we denote as usual $J_\lambda(t) = \{(p, S_\lambda(t)p), p \in \mathbb{R}^n\}$) that the identity (16.11) is written as $S_{c\lambda}(t) = \frac{1}{c} S_\lambda(ct)$ thus $S_{c\lambda}(t)^{-1} = cS_\lambda(ct)^{-1}$. From here[3] one also gets $B_{c\lambda}(t) = cB_\lambda(ct)$ and (16.12) follows from the identity $S^\circ(t) = B^{-1}(t) + S(t)$. (See also Exercise 14.22). In particular at $t = 0$ the identity (16.12) says that $\mathcal{H}_{c\lambda} = \delta_{c*}\mathcal{H}_\lambda$.

(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 ∇_X and $\vec{H}$ are horizontal vector field. Since the horizontal space is Lagrangian

$$\nabla_X H = \sigma(\nabla_X, \vec{H}) = 0.$$

□

Exercise 16.19. 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 $\alpha \geq 0$. Then f is linear.

The following theorem says that a connection satisfying the three properties above is unique. Then it characterize the Levi-Civita connection in terms of the structure constants of the Lie algebra defined by an orthonormal frame.

Theorem 16.20. *There is a unique Ehresmann connection ∇ satisfying the properties (i), (ii), and (iii) of Proposition 16.18, that is the Levi-Civita connection. Its Christoffel symbols are computed by*

$$\Gamma^k_{ij} = \frac{1}{2}(c^k_{ij} - c^i_{jk} + c^j_{ki}), \tag{16.13}$$

where c^k_{ij} are the smooth functions defined by the identity $[X_i, X_j] = \sum_{k=1}^n c^k_{ij} X_k$.

Proof. Let $X_1, \ldots, X_n$ be a local orthonormal frame for the Riemannian structure and let us consider coordinates (q, u) in T^*M, where the fiberwise coordinates $u = (u_1, \ldots, u_n)$ are dual to the orthonormal frame. From the linearity of the connection it follows that there exist smooth functions $\Gamma^k_{ij} : M \to \mathbb{R}$ (depending on q only) such that

$$\nabla_{X_i} = X_i + \sum_{j=1}^n \Gamma^k_{ij} u_k \partial_{u_j}, \qquad i = 1, \ldots, n.$$

In particular $\nabla_{X_i} X_j = \Gamma^k_{ij} X_k$. In these coordinates the Hamiltonian vector field associated with the Riemannian Hamiltonian $H = \frac{1}{2}\sum_{i=1}^n u_i^2$ reads (see also Exercise **??**)

$$\vec{H} = \sum_{i,j,k=1}^n u_i X_i + c^k_{ij} u_i u_k \partial_{u_j},$$

while the symplectic form σ is written ($\nu_1, \ldots, \nu_n$ denotes the dual basis to $X_1, \ldots, X_n$)

$$\sigma = \sum_{i,j,k=1}^n du_k \wedge \nu_k - c^k_{ij} u_k \nu_i \wedge \nu_k.$$

[3] recall that B is the zero order term of the expansion of S^{-1}.

Since the horizontal space is Lagrangian, one has the relations

$$0 = \sigma(\nabla_{X_i}, \nabla_{X_j}) = \sum_{k=1}^{n} (\Gamma_{ij}^k - \Gamma_{ji}^k - c_{ij}^k) u_k, \qquad \forall\, i, j = 1, \ldots, n,$$

hence $c_{ij}^k = \Gamma_{ij}^k - \Gamma_{ji}^k$ for all i, j, k. Moreover the connection is metric, i.e. it satisfies

$$0 = \nabla_{X_i} H = \sum_{j,k=1}^{n} \Gamma_{ij}^k u_k u_j, \qquad \forall\, i = 1, \ldots, n.$$

The last identity implies that Γ_{ij}^k is skew-symmetric with respect to the pair (j,k), i.e. $\Gamma_{ij}^k = -\Gamma_{ik}^j$. Thus combining the two identities one gets

$$\begin{aligned} c_{ij}^k - c_{jk}^i + c_{ki}^j &= (\Gamma_{ij}^k - \Gamma_{ji}^k) - (\Gamma_{jk}^i + \Gamma_{kj}^i) + (\Gamma_{ki}^j - \Gamma_{ik}^j) \\ &= \Gamma_{ij}^k - \Gamma_{ik}^j = 2\Gamma_{ij}^k. \end{aligned}$$

□

Remark 16.21. Notice that in the classical approach one can recover formula (16.13) from the following particular case of the Koszul formula

$$\Gamma_{ij}^k = g(\nabla_{X_i} X_j, X_k) = \frac{1}{2} \left(g([X_i, X_j], X_k) - g([X_j, X_k], X_i) + g([X_k, X_i], X_j) \right),$$

that holds for every orthonormal basis $X_1, \ldots, X_n$. Notice also that the Hamiltonian vector field is written in coordinates $\vec{H} = \sum_{i=1}^n u_i \nabla_{X_i}$, which gives another proof of the fact that it is horizontal.

Let $X, Y, Z, W \in \mathrm{Vec}(M)$. We define $R(X,Y)Z = W$ if $R(X,Y)Z^* = W^*$.

Proposition 16.22 (Bianchi identity)**.** *For every $X, Y, Z \in \mathrm{Vec}(M)$ the following identity holds*

$$R(X,Y)Z + R(Y,Z)X + R(Z,X)Y = 0. \tag{16.14}$$

Proof. We will show that (16.14) is a consequence of the Jacobi identity (2.32). Using that ∇ 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

$$\begin{aligned} 0 &= [X,[Y,Z]] + [Y,[Z,X]] + [Z,[X,Y]] \\ &= \nabla_X \nabla_Y Z - \nabla_X \nabla_Z Y - \nabla_{[Y,Z]} X \\ &\quad + \nabla_Z \nabla_X Y - \nabla_Z \nabla_Y X - \nabla_{[X,Y]} Z \\ &\quad\quad + \nabla_Y \nabla_Z X - \nabla_Y \nabla_X Z - \nabla_{[Z,X]} Y \\ &= R(X,Y)Z + R(Y,Z)X + R(Z,X)Y. \end{aligned}$$

□

Exercise 16.23. Prove the second Bianchi identity

$$(\nabla_X R)(Y, Z, W) + (\nabla_Y R)(Z, X, W) + (\nabla_Z R)(X, Y, W) = 0, \qquad \forall\, X, Y, Z, W \in \mathrm{Vec}(M).$$

(Hint: Expand the identity $\nabla_{[X,[Y,Z]]+[Y,[Z,X]]+[Z,[X,Y]]}W = 0$.)

Let us denote $(X, Y, Z, W) := \langle R(X, Y)Z, W\rangle$. Following this notation, the first Bianchi identity can be rewritten as follows:

$$(X, Y, Z, W) + (Z, X, Y, W) + (Y, Z, X, W) = 0, \qquad \forall\, X, Y, Z, W \in \mathrm{Vec}(M). \tag{16.15}$$

Remark 16.24. The property of the Riemann tensor can be reformulated as follows

$$(X, Y, Z, W) = -(Y, X, Z, W), \qquad (X, Y, Z, W) = -(X, Y, W, Z). \tag{16.16}$$

Proposition 16.25. *For every $X, Y, Z, W \in \mathrm{Vec}(M)$ we have $(X, Y, Z, W) = (Z, W, X, Y)$.*

Proof. Using (16.15) four times we can write the identities

$$\begin{aligned}
(X, Y, Z, W) + (Z, X, Y, W) + (Y, Z, X, W) = 0,\\
(Y, Z, W, X) + (W, Y, Z, X) + (Z, W, Y, X) = 0,\\
(Z, W, X, Y) + (X, Z, W, Y) + (W, X, Z, Y) = 0,\\
(W, X, Y, Z) + (Y, W, X, Z) + (X, Y, W, Z) = 0.
\end{aligned}$$

Summing all together and using the skew symmetry (16.16), one gets $(X, Z, W, Y) = (W, Y, X, Z)$. ☐

Proposition 16.26. *Assume that $(X, Y, X, W) = 0$ for every $X, Y, W \in \mathrm{Vec}(M)$. Then*

$$(X, Y, Z, W) = 0 \qquad \forall\, X, Y, Z, W \in \mathrm{Vec}(M).$$

Proof. By assumptions and the skew-simmetry properties (16.16) of the Riemann tensor we have that $(X, Y, Z, W) = 0$ whenever any two of the vector fields coincide. In particular

$$0 = (X, Y + W, Z, Y + W) = (X, Y, Z, W) + (X, W, Z, Y). \tag{16.17}$$

since the two extra terms that should appear in the expansion vanish by assumptions. Then (16.17) can be rewritten as

$$(X, Y, Z, W) = (Z, X, Y, W),$$

i.e. the quantity (X, Y, Z, W) is invariant by ciclic permutations of X, Y, Z. But the cyclic sum of terms is zero by (16.15), hence $(X, Y, Z, W) = 0$. ☐

We end this section by summarizing the symmetry property of the Riemann curvature as follows

Corollary 16.27. *There is a well defined map*

$$\overline{R} : \wedge^2 T_qM \to \wedge^2 T_qM, \qquad \overline{R}(X \wedge Y) := R(X, Y).$$

Moreover $\bar{R}$ is skew-adjoint with respect to the induced scalar product on $\wedge^2 T_qM$, that means

$$\left\langle \overline{R}(X \wedge Y), Z \wedge W \right\rangle = \left\langle X \wedge Y, \overline{R}(Z \wedge W) \right\rangle.$$

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 to a geodesic is a kind of sectional curvature operator in the direction of the geodesic itself.

Definition 16.28. The *Hamiltonian curvature tensor* at $\lambda \in T^*M$ is the operator

$$\mathcal{R}_\lambda := \mathcal{R}_{J_\lambda(0)} : \mathcal{V}_\lambda \to \mathcal{V}_\lambda.$$

In other words $\mathcal{R}_\lambda$ is the curvature of the Jacobi curve associated with λ at $t = 0$.

Proposition 16.29. *Let $\xi \in \mathcal{V}_\lambda$ and V be a smooth vertical vector field extending ξ. Then*

$$\mathcal{R}_\lambda(\xi) = -[\vec{H}, [\vec{H}, V]_{hor}]_{ver}(\lambda)$$

Proof. This is a direct consequence of Proposition 14.30. Indeed 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

$$\pi_{J(0)J^\circ(0)}(\xi) = \xi_{hor}, \qquad \pi_{J^\circ(0)J(0)}(\eta) = \eta_{ver}.$$

FInally we can extend vectors in $J_\lambda(0)$ (resp. $J^\circ_\lambda(0)$) by applying the Hamiltonian vector field since $J_\lambda(t) = e^{t\vec{H}}_* J_\lambda(0)$ (resp. $J^\circ_\lambda(t) = e^{t\vec{H}}_* J^\circ_\lambda(0)$). From these remarks we obtain the following formulas

$$\underline{\dot{J}}_\lambda(0)\xi = [\vec{H}, V]_{hor}, \qquad \underline{\dot{J}}^\circ_\lambda(0)\eta = -[\vec{H}, W]_{ver}$$

for some V vertical (resp. W horizontal) extension of the vector $\xi \in \mathcal{V}_\lambda$ (reps. $\eta \in \mathcal{H}_\lambda$). □

Another immediate property of the curvature tensor is the homogeneity with respect to the rescaling of the covector (that corresponds to reparametrization of the trajectory). Indeed by choosing $\varphi(t) = ct$, with $c > 0$, in Proposition 14.36 one gets

Corollary 16.30. *For every $c > 0$ we have $\mathcal{R}_{c\lambda} = c^2\mathcal{R}_\lambda$.*

If we use the Riemannian product to identify the tangent and the cotangent space at a point, we recognize that $\mathcal{R}_\lambda$ is nothing but the sectional curvature operator where one entry is the tangent vector $\dot{\gamma}$ of the geodesic.

Let us denote by $I : TM \to T^*M$ the isomorphism defined by the Riemannian scalar product $\langle\cdot|\cdot\rangle$. In particular $I(v) = \lambda$ for $\lambda \in T^*_qM$ and $v \in T_qM$ if $\langle\lambda, w\rangle = \langle v|w\rangle$ for all $w \in T_qM$.

Let denote $H_q = H|_{T^*_qM}$. Recall that the differential of H_q can be interpreted as a linear map $DH_q : T^*_qM \to T_qM$ that sends $\lambda \in T^*_qM$ into $D_\lambda H_q$ seen as a linear functional on T^*_qM, i.e. a tangent vector. This map is actually the inverse of the isomorphism I.

Lemma 16.31. $D_\lambda H_q = I^{-1}(\lambda)$.

Proof. It is a simple consequence of the formula $H(\lambda) = \frac{1}{2}\left\langle\lambda, I^{-1}(\lambda)\right\rangle$. □

Corollary 16.32. *Assume $I(v) = \lambda$, then $\vec{H}(\lambda) = \nabla_v$.*

Proof. Indeed, since $\vec{H}$ is an horizontal vector field, it is sufficient to show that $\pi_* \vec{H}(\lambda) = v$, which is a consequence of Lemma 16.31. Indeed for every vertical vector $\xi \in T_\lambda(T_q^*M)$ one has

$$\langle \xi, v \rangle = \left\langle \xi, I^{-1}(\lambda) \right\rangle = D_\lambda H(\xi) = \sigma(\xi, \vec{H}(\lambda)) = \left\langle \xi, \pi_* \vec{H}(\lambda) \right\rangle.$$

By arbitrary of $\xi \in T_\lambda(T_q^*M)$ one has the equality $v = \pi_* \vec{H}(\lambda)$. □

Theorem 16.33. *We have the following identity*

$$\mathcal{R}_{I(X)}(I(Y)) = R(X,Y)X, \qquad \forall\, X, Y \in T_qM. \tag{16.18}$$

Proof. We have to compute the quantity

$$\mathcal{R}_{I(X)}(I(Y)) = -[\vec{H}, [\vec{H}, IY]_{hor}]_{ver}(I(X))$$

First notice that $\pi_*[\vec{H}, I(Y)] = -Y$ hence $[\vec{H}, I(Y)]_{hor} = -\nabla_Y$. Then

$$-[\vec{H}, [\vec{H}, I(Y)]_{hor}]_{ver}(I(X)) = [\nabla_X, \nabla_Y]_{ver}(I(X)) = R(X,Y)(X).$$

□

Definition 16.34. The *Ricci tensor* at λ is defined as the trace of the curvature operator at λ, $\mathrm{Ric}(\lambda) := \mathrm{trace}\, \mathcal{R}_\lambda$.

Exercise 16.35. Prove the following expression for the Ricci tensor, where $X_1, \ldots, X_n$ is a local orthonormal frame and $\dot{\gamma}(0) = v = I^{-1}(\lambda)$ is the tangent vector to the geodesic:

$$\begin{aligned}\mathrm{Ric}(\lambda) &= \sum_{i=1}^n \langle R(v, X_i)v | X_i \rangle \\ &= \sum_{i=1}^n \sigma_\lambda([\vec{H}, \nabla_{X_i}], \nabla_{X_i}).\end{aligned}$$

This shows that $\mathrm{Ric}(\lambda) = \mathrm{Ric}(v)$ coincide with the classical Riemannian Ricci tensor.

16.4 Locally flat spaces

In this section we want to show that the Riemannian curvature is the only obstruction for a Riemannian manifold to be locally Euclidean. Finally we show that the Riemannian curvature is also completely recovered by the Hamiltonian curvature $\mathcal{R}_\lambda$.

A Riemannian manifold M is called *flat* if $R(X,Y) = 0$ for every $X, Y \in \mathrm{Vec}(M)$.

Theorem 16.36. *M is flat if and only if M is locally isometric to $\mathbb{R}^n$.*

Proof. If M is locally isometric to $\mathbb{R}^n$, then its curvature tensor at every point in a neighborhood is identically 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 coordinate such that the Hamiltonian, in these set of coordinates, is written as the Hamiltonian of the Euclidean $\mathbb{R}^n$.

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 transversal to the fiber T_q^*M, where $q = \pi(\lambda)$.

Fix a point $q_0 \in M$ and a neighborhood O_{q_0} where R is identically zero. 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 λ the intersection of the leaf passing through this point and $T_{q_0}^*M$.

Exercise 16.37. Show that Ψ is a linear, orthogonal transformation, i.e. $H(\Psi(\lambda)) = H(\lambda)$ for all $\lambda \in \pi^{-1}(O_{q_0})$. (Hint: use the linearity of the connection and the fact that $\vec{H}$ is horizontal).

Fix now a basis $\{\nu_1, \ldots, \nu_n\}$ in $T_{q_0}^*M$ that is orthonormal (with respect to the dual metric). Being Ψ 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 $X_1, \ldots, X_n$ in the neighborhood O_{q_0}. Moreover $X_1, \ldots, X_n$ is an orthonormal basis since Ψ is an orthogonal map.

We want to show that $\{X_1, \ldots, X_n\}$ is an orthonormal basis of vector fields that commutes everywhere.

Let us show that the fact that the foliation is Lagrangian implies $[X_i, X_j] = 0$ for all $i, j = 1, \ldots, n$.

Indeed the tautological 1-form is written in these coordinates as $s = \sum_{i=1}^n \psi_i \, \nu_i$ and

$$\sigma = ds = \sum_{i=1}^n d\psi_i \wedge \nu_i + \psi_i d\nu_i. \tag{16.19}$$

Since on each leaf the function ψ_i is constant by definition (hence $d\psi_i|_{\mathfrak{L}} = 0$), 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

$$0 = d\nu_i(X_j, X_k) = -\nu_i([X_j, X_k]), \qquad \forall\, i, j, k.$$

This proves that $[X_i, X_j] = 0$ for each $i, j = 1, \ldots, n$. Hence, in the coordinate set (ψ, q), we have $H(\psi, q) = \frac{1}{2}|\psi|^2$. ☐

The next result shows that the Hamiltonian curvature can detect if a manifold is flat or not.

Corollary 16.38. *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 $\mathcal{R}_\lambda = 0$ from (16.18).

Let us prove the converse. Recall that $\mathcal{R}_\lambda = 0$ implies, again by (16.18), that

$$(X, Y, X, W) = 0, \qquad \forall\, X, Y, W \in \mathrm{Vec}(M).$$

Then the statement is a consequence of Proposition 16.26. $\square$

Exercise 16.39. Prove that actually the Riemann tensor R is completely determined by $\mathcal{R}$.

16.5 Example: curvature of the 2D Riemannian case

In this section we apply the definition of curvature discussed in this chapter to a two dimensional Riemannian surface. As we explain, we recover that the Riemannian curvature tensor is determined by the Gauss curvature of the manifold.

Let M be a 2-dimensional surface and $f_1, f_2 \in \mathrm{Vec}(M)$ be a local orthonormal frame for the Riemannian metric. The Riemannian Hamiltonian H is written as follows (we use canonical coordinates $\lambda = (p, x)$ on T^*M)

$$H(p, x) = \frac{1}{2}(\langle p, f_1(x)\rangle^2 + \langle p, f_2(x)\rangle^2) \tag{16.20}$$

Here, for a covector $\lambda = (p, x) \in T^*M$, the symplectic vector space $\Sigma_\lambda = T_\lambda(T^*M)$ is 4-dimensional.

Recall that, being M 2-dimensional, the level set $H^{-1}(1/2) \cap T_q^*M$ is a circle. Hence, there is a well defined vector field that produces rotation on the reduced fiber. Let us define the angle θ on the level $H^{-1}(1/2) \cap T_x^*M$ by setting

$$\langle p, f_1(x)\rangle = \cos\theta, \qquad \langle p, f_2(x)\rangle = \sin\theta,$$

in such a way that $\theta = 0$ corresponds to the direction of f_1. Denote by ∂_θ the rotation in the fiber of the unit tangent bundle and by $\vec{E}$, the Euler vector field. Denote finally by $\vec{H}' := [\partial_\theta, \vec{H}]$.

Notice that $\Sigma_\lambda = \mathcal{V}_\lambda \oplus \mathcal{H}_\lambda$ where $\mathcal{V}_\lambda = \mathrm{span}\{\vec{E}, \partial_\theta\}$ and $\mathcal{H}_\lambda = \mathrm{span}\{\vec{H}, \vec{H}'\}$.

Lemma 16.40. *The vector fields $\{\vec{E}, \partial_\theta, \vec{H}, \vec{H}'\}$ at λ form a Darboux basis for Σ_λ.*

Proof. We want to compute the following symplectic products of the vector fields:

$$\sigma(\partial_\theta, \vec{E}) = 0, \qquad \sigma(\partial_\theta, \vec{H}) = 0, \qquad \sigma(\vec{E}, \vec{H}) = 1. \tag{16.21}$$

$$\sigma(\partial_\theta, \vec{H}') = 1, \qquad \sigma(\vec{E}, \vec{H}') = 0, \qquad \sigma(\vec{H}, \vec{H}') = 0. \tag{16.22}$$

Indeed, let us prove first (16.21). 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, ∂_θ 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 (16.22) notice that, if $s = i_{\vec{E}}\sigma$ denotes the tautological Liouville form, one has

$$\langle s, \vec{H}\rangle = 1, \qquad \langle s, \vec{H}'\rangle = 0. \tag{16.23}$$

These two identities follows from

$$\langle s, \vec{H} \rangle = \sigma(\vec{E}, \vec{H}) = 1, \tag{16.24}$$

$$\langle s, \vec{H}' \rangle = \langle s, [\partial_\theta, \vec{H}] \rangle = ds(\partial_\theta, \vec{H}) = \sigma(\partial_\theta, \vec{H}) = 0, \tag{16.25}$$

where in the second line we used the Cartan formula (4.77) and the fact that ∂_θ is vertical.

Let us now prove (16.22). Being $[\partial_\theta, \vec{H}'] = [\partial_\theta, [\partial_\theta, \vec{H}]] = -\vec{H}$, we have again by Cartan formula and (16.23)

$$\sigma(\partial_\theta, \vec{H}') = ds(\partial_\theta, \vec{H}') = -\langle s, [\partial_\theta, \vec{H}'] \rangle = \langle s, \vec{H} \rangle = \sigma(\vec{E}, \vec{H}) = 1$$

Moreover by (16.23)

$$\sigma(\vec{E}, \vec{H}') = \langle s, \vec{H}' \rangle = 0.$$

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 (with dH as 1-form).

$$dH([\partial_\theta, \vec{H}]) = d^2H(\partial_\theta, \vec{H}) - \partial_\theta \langle dH, \vec{H} \rangle + \vec{H}\, \langle dH, \partial_\theta \rangle = 0$$

since the three terms are all equal to zero.

□

Now we compute the curvature via the Jacobi curve, reduced by homogeneity. Notice that by Lemma 16.40 we can remove the symplectic space spanned by $\{\vec{E}, \vec{H}\}$ and, being $\{\vec{E}, \vec{H}\}^{\angle} = \{\partial_\theta, \vec{H}'\}$, we have

$$\widehat{J}_\lambda(t) = \mathrm{span}\{e_*^{-t\vec{H}} \partial_\theta\}.$$

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

$$\sigma(V_t, \dot{V}_t) = -1, \qquad \text{for every } t \geq 0. \tag{16.26}$$

Indeed it is true for $t = 0$ and the equality is valid for all t since the transformation $e_*^{t\vec{H}}$ is symplectic. To compute the curvature of the Jacobi curve let us write

$$V_t = \alpha(t) V_0 - \beta(t) \dot{V}_0 \tag{16.27}$$

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

$$V_t = \alpha(t) V_0 - \beta(t) \dot{V}_0 = \alpha(t) \left(V_0 - \frac{\beta(t)}{\alpha(t)} \dot{V}_0 \right), \tag{16.28}$$

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 $\dot{V}_0$ is a Darboux basis, it is easy to compute

$$\sigma(V_0, V_t) = \alpha(t)\underbrace{\sigma(V_0, V_0)}_{=0} - \beta(t)\underbrace{\sigma(V_0, \dot{V}_0)}_{=-1} = \beta(t), \tag{16.29}$$

$$\sigma(\dot{V}_0, V_t) = \alpha(t)\underbrace{\sigma(\dot{V}_0, V_0)}_{=1} - \beta(t)\underbrace{\sigma(\dot{V}_0, \dot{V}_0)}_{=0} = \alpha(t). \tag{16.30}$$

Differentiating the identity (16.26) with respect to t one gets the relations

$$\sigma(V_t, \ddot{V}_t) = 0, \qquad \sigma(V_t, V_t^{(3)}) = -\sigma(\dot{V}_t, \ddot{V}_t)$$

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

$$S(t) = \frac{-t + \frac{t^3}{6}\sigma(V_0, \dddot{V}_0) + O(t^5)}{1 + \frac{t^2}{2}\sigma(\dot{V}_0, \ddot{V}_0) + O(t^4)} \tag{16.31}$$

$$= \left(-t + \frac{t^3}{6}\sigma(V_0, \dddot{V}_0) + O(t^5)\right)\left(1 - \frac{t^2}{2}\sigma(\dot{V}_0, \ddot{V}_0) + O(t^4)\right) \tag{16.32}$$

and one gets for the derivative of $S(t)$ at $t = 0$

$$\dot{S}(0) = -1, \qquad \ddot{S}(0) = 0, \qquad \dddot{S}(0) = 2\sigma(\dot{V}_0, \ddot{V}_0).$$

The formula for the curvature $\mathcal{R}$ is finally computed in terms of $S(t)$ as follows:

$$\mathcal{R} = -\frac{1}{2}\dddot{S}(0) = \sigma(\ddot{V}_0, \dot{V}_0) \tag{16.33}$$

Using that $V_t = e_*^{-t\vec{H}}\partial_\theta$ we can expand V_t 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 (16.33) is rewritten as

$$\mathcal{R} = \sigma([\vec{H}, [\vec{H}, \partial_\theta]], [\vec{H}, \partial_\theta]) \tag{16.34}$$

$$= \sigma([\vec{H}, \vec{H}'], \vec{H}') \tag{16.35}$$

To end this section, we compute the curvature $\mathcal{R}$ with respect to the orthonormal frame f_1, f_2. Denote the Hamiltonians

$$h_i(p, x) = \langle p, f_i(x) \rangle, \qquad i = 1, 2.$$

The PMP reads

$$\begin{cases} \dot{x} = h_1 f_1(x) + h_2 f_2(x) \\ \dot{h}_1 = \{H, h_1\} = \{h_2, h_1\}h_2 \\ \dot{h}_2 = \{H, h_2\} = -\{h_2, h_1\}h_1 \end{cases} \tag{16.36}$$

Moreover $\{h_2, h_1\}(p, x) = \langle p, [f_2, f_1](x)\rangle$. Assume that

$$[f_1, f_2] = a_1 f_1 + a_2 f_2, \qquad a_i \in C^\infty(M).$$

Then

$$\{h_2, h_1\} = -a_1 h_1 - a_2 h_2.$$

If we restrict to $h_1 = \cos\theta$ and $h_2 = \sin\theta$ equations (16.36) become

$$\begin{cases} \dot{x} = \cos\theta f_1 + \sin\theta f_2 \\ \dot{\theta} = a_1 \cos\theta + a_2 \sin\theta \end{cases}$$

and it is easy to compute the following expression for $\vec{H}$ and commutators[4]

$$\begin{aligned} \vec{H} &= h_1 f_1 + h_2 f_2 + (a_1 h_1 + a_2 h_2)\partial_\theta, \\ \vec{H}' &= -h_2 f_1 + h_1 f_2 + (-a_1 h_2 + a_2 h_1)\partial_\theta, \\ [\vec{H}, \vec{H}'] &= (f_1 a_2 - f_2 a_1 - a_1^2 - a_2^2)\partial_\theta. \end{aligned}$$

Recall that

$$\kappa = f_1 a_2 - f_2 a_1 - a_1^2 - a_2^2,$$

is the Gaussian curvature of the surface M (see also Chapter 4). Since $\sigma(\partial_\theta, \vec{H}') = 1$ one gets

$$\mathcal{R} = \sigma([\vec{H}, \vec{H}'], \vec{H}') = \sigma(\kappa\partial_\theta, \vec{H}') = \kappa.$$

Exercise 16.41. In this exercise we recover the previous computations introducing dual coordinates to our frame. Let ν_1, ν_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 $b := a_1 h_1 + a_2 h_2$ on T^*M. In these notation

$$\vec{H} = f_\theta + b\partial_\theta, \qquad \vec{H}' = f_{\theta'} + b'\partial_\theta,$$

where $'$ denotes the derivative with respect to θ. Then, using that in these coordinates the tautological form is $s = \nu_\theta$, show that the symplectic form is written as

$$\sigma = ds = d\theta \wedge \nu_{\theta'} - b\, \nu_1 \wedge \nu_2,$$

and compute the following expressions

$$\begin{aligned} i_{\vec{H}'}\sigma &= (b' - b)\nu_{\theta'} - d\theta, \\ [\vec{H}, \vec{H}'] &= (f_\theta b' - f_{\theta'} b - b^2 - b'^2)\partial_\theta, \end{aligned}$$

showing that this gives an alternative proof of the above computation of the curvature.

[4] here we still use the notation h_1, h_2 as functions of θ satisfying $\partial_\theta h_1 = -h_2, \partial_\theta h_2 = h_1$

Chapter 17

Curvature in 3D contact sub-Riemannian geometry

The main goal of this chapter is to compute the curvature of the three dimensional contact sub-Riemannian case. Then we will discuss how the invariant contained in the sub-Riemannian curvature classify 3D left-invariant structures on Lie groups.

17.1 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 contact 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

$$\mathcal{D}_q = \ker \omega_q = \operatorname{span}\{f_1(q), f_2(q)\}$$

Recall that the 1-form $\omega \in \Lambda^1(M)$ defines a contact distribution if and only if $\omega \wedge d\omega \neq 0$ is never vanishing.

Exercise 17.1. Let M be a 3D manifold, $\omega \in \Lambda^1 M$ and $\mathcal{D} = \ker \omega$. The following are equivalent:

(*i*) ω is a contact 1-form,

(*ii*) $d\omega\big|_{\mathcal{D}} \neq 0$,

(*iii*) $\forall\, f_1, f_2 \in \overline{\mathcal{D}}$ linearly independent, then $[f_1, f_2] \notin \overline{\mathcal{D}}$.

Remark 17.2. The contact form ω is defined up to a smooth function, i.e. if ω is a contact form, $a\omega$ is a contact form for every $a \in C^\infty(M)$. This let 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 ν_1, ν_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 17.3. The *Reeb vector field* of the contact structure is the unique vector field $f_0 \in \operatorname{Vec}(M)$ 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 ω, ν_1, ν_2 is the dual basis to this frame.

Remark 17.4. The flow generated by the Reeb vector field $e^{tf_0} : M \to M$ is a group of diffeomorphisms that satisfy $(e^{tf_0})^*\omega = \omega$. Indeed

$$\mathcal{L}_{f_0}\omega = d(i_{f_0}\omega) + i_{f_0}d\omega = 0$$

since $i_{f_0}\omega = \omega(f_0) = 1$ is constant and $i_{f_0}d\omega = d\omega(f_0, \cdot) = 0$.

In what follows, to simplify the notation, we will replace the contact form ω by ν_0, as the dual element to the vector field f_0. We can write the structure equations of this basis of 1-forms

$$\begin{cases} d\nu_0 = \nu_1 \wedge \nu_2 \\ d\nu_1 = c_{01}^1 \nu_0 \wedge \nu_1 + c_{02}^1 \nu_0 \wedge \nu_2 + c_{12}^1 \nu_1 \wedge \nu_2 \\ d\nu_2 = c_{01}^2 \nu_0 \wedge \nu_1 + c_{02}^2 \nu_0 \wedge \nu_2 + c_{12}^2 \nu_1 \wedge \nu_2 \end{cases} \tag{17.1}$$

The structure constants c_{ij}^k are smooth functions on the manifold. Recall that the equation

$$d\nu_k = \sum_{i,j=0}^{2} c_{ij}^k \nu_i \wedge \nu_j \qquad \text{if and only if} \qquad [f_j, f_i] = \sum_{k=0}^{2} c_{ij}^k f_k.$$

Introduce the coordinates (h_0, h_1, h_2) in each fiber of T^*M induced by the dual frame

$$\lambda = h_0\nu_0 + h_1\nu_1 + h_2\nu_2$$

where $h_i(\lambda) = \langle \lambda, f_i(q) \rangle$ are the Hamiltonians linear on fibers associated to f_i, for $i = 0, 1, 2$. The sub-Riemannian Hamiltonian is written as follows

$$H = \frac{1}{2}(h_1^2 + h_2^2).$$

We now compute the Poisson bracket $\{H, h_0\}$, denoting with $\{H, h_0\}_q$ its restriction to the fiber T_q^*M.

Proposition 17.5. *The Poisson bracket $\{H, h_0\}_q$ is a quadratic form. Moreover we have*

$$\{H, h_0\} = c_{01}^1 h_1^2 + (c_{01}^2 + c_{02}^1) h_1 h_2 + c_{02}^2 h_2^2, \tag{17.2}$$

$$c_{01}^1 + c_{02}^2 = 0. \tag{17.3}$$

Notice that $\Delta_q^\perp \subset \ker \{H, h_0\}_q$ and $\{H, h_0\}_q$ can be treated as a quadratic form on $T_q^*M/\Delta_q^\perp = \Delta_q^*$.

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 (17.1) one gets:

$$
\begin{aligned}
0 = d^2\nu_0 &= d\nu_1 \wedge \nu_2 - \nu_1 \wedge \nu_2 \\
&= (c_{01}^1 \nu_0 \wedge \nu_1) \wedge \nu_2 - \nu_1 \wedge (c_{02}^2 \nu_0 \wedge \nu_2) \\
&= (c_{01}^1 + c_{02}^2)\nu_0 \wedge \nu_1 \wedge \nu_2
\end{aligned}
$$

which proves (17.3). □

Remark 17.6. 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 (17.3) we get

$$
\text{trace}\,\{H, h_0\}_q = c_{01}^1 + c_{02}^2 = 0.
$$

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.

It is natural then to define our *first invariant* as the positive eigenvalue of this operator, namely:

$$
\chi(q) = \sqrt{-\det\{H, h_0\}_q}. \tag{17.4}
$$

Notice that the function χ 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 $\{H, h_0\}$ compute the derivative of H along the flow of $\vec{h}_0$, i.e. the obstruction to the fact that the flow of the Reeb field f_0 (which preserves the distribution and the volume form on it) to preserve the metric. Notice that, by definition $\chi \geq 0$.

Corollary 17.7. *Assume that the vector field f_0 is complete. Then $\{e^{tf_0}\}_{t\in\mathbb{R}}$ is a group of sub-Riemannian isometries if and only if $\chi \equiv 0$.*

In the case when $\chi \equiv 0$ one can consider (locally) the quotient of M with respect to the action of this group, i.e. the space of trajectories described by f_0. The two dimensional surface defined by the quotient strucure is endowed with a well defined Riemannian metric.

The sub-Riemannian structure on M coincide with the isoperimetric Dido problem constructed on this surface. The Heisenberg case corresponds with the case when the surface has zero Gaussian curvature.

17.2 Canonical frames

In this section we want to show that it is always possible to select a canonical orthonormal frame for the sub-Riemannian structure. In this way we are able to find missing discrete invariants and to classify sub-Riemannian structures simply knowing structure constants c_{ij}^k for the canonical frame. We study separately the two cases $\chi \neq 0$ and $\chi = 0$.

We start by rewriting and improving Proposition 17.5 when $\chi \neq 0$.

Proposition 17.8. *Let M be a 3D contact sub-Riemannian manifold and $q \in M$. If $\chi(q) \neq 0$, then there exists a local frame such that*

$$\{h, h_0\} = 2\chi h_1 h_2. \tag{17.5}$$

In particular, in the Lie group case with left-invariant stucture, there exists a unique (up to a sign) canonical frame (f_0, f_1, f_2) such that

$$\begin{aligned} [f_1, f_0] &= c_{01}^2 f_2, \\ [f_2, f_0] &= c_{02}^1 f_1, \\ [f_2, f_1] &= c_{12}^1 f_1 + c_{12}^2 f_2 + f_0. \end{aligned} \tag{17.6}$$

Moreover we have

$$\chi = \frac{c_{01}^2 + c_{02}^1}{2}, \qquad \kappa = -(c_{12}^1)^2 - (c_{12}^2)^2 + \frac{c_{01}^2 - c_{02}^1}{2}. \tag{17.7}$$

Proof. From Proposition 17.5 we know that the Poisson bracket $\{h, h_0\}_q$ 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 (remember that f_0 depends only on the structure and not on the orthonormal frame on the distribution). It is easily seen that in both of these cases we obtain the expression (17.5). □

Remark 17.9. Notice that, if we change sign to f_1 or f_2, then c_{12}^2 or c_{12}^1, respectively, change sign in (17.6), while c_{02}^1 and c_{01}^2 are unaffected. Hence equalities (17.7) do not depend on the orientation of the sub-Riemannian structure.

If $\chi = 0$ the above procedure cannot apply. Indeed both trace and determinant of the operator vanish, hence we have $\{h, h_0\}_q = 0$. From (17.2) we get the identities

$$c_{01}^1 = c_{02}^2 = 0, \quad c_{01}^2 + c_{02}^1 = 0. \tag{17.8}$$

so that commutators (**??**) simplify in (where $c = c_{01}^2$)

$$\begin{aligned} [f_1, f_0] &= c f_2, \\ [f_2, f_0] &= -c f_1, \\ [f_2, f_1] &= c_{12}^1 f_1 + c_{12}^2 f_2 + f_0. \end{aligned} \tag{17.9}$$

We want to show, with an explicit construction, that also in this case there always exists a rotation of our frame, by an angle that smoothly depends on the point, such that in the new frame κ is the only structure constant which appear in (17.9).

Lemma 17.10. *Let f_1, f_2 be an orthonormal frame on M. If we denote with $\widehat{f}_1, \widehat{f}_2$ the frame obtained from the previous one with a rotation by an angle $\theta(q)$ and with $\widehat{c}_{ij}^k$ structure constants of rotated frame, we have:*

$$\begin{aligned} \widehat{c}_{12}^1 &= \cos\theta(c_{12}^1 - f_1(\theta)) - \sin\theta(c_{12}^2 - f_2(\theta)), \\ \widehat{c}_{12}^2 &= \sin\theta(c_{12}^1 - f_1(\theta)) + \cos\theta(c_{12}^2 - f_2(\theta)). \end{aligned}$$

Now we can prove the main result of this section.

Proposition 17.11. *Let M be a 3D simply connected contact sub-Riemannian manifold such that $\chi = 0$. Then there exists a rotation of the original frame $\widehat{f}_1, \widehat{f}_2$ such that:*

$$
\begin{aligned}
[\widehat{f}_1, f_0] &= \kappa \widehat{f}_2, \\
[\widehat{f}_2, f_0] &= -\kappa \widehat{f}_1, \\
[\widehat{f}_2, \widehat{f}_1] &= f_0.
\end{aligned}
\tag{17.10}
$$

Proof. Using Lemma 17.10 we can rewrite the statement in the following way: there exists a function $\theta : M \to \mathbb{R}$ such that

$$
f_1(\theta) = c_{12}^1, \quad f_2(\theta) = c_{12}^2. \tag{17.11}
$$

Indeed, this would imply $\widehat{c}_{12}^1 = \widehat{c}_{12}^2 = 0$ and $\kappa = c$.

Let us introduce simplified notations $c_{12}^1 = \alpha_1,\ c_{12}^2 = \alpha_2$. Then

$$
\kappa = f_2(\alpha_1) - f_1(\alpha_2) - (\alpha_1)^2 - (\alpha_2)^2 + c. \tag{17.12}
$$

If (ν_0, ν_1, ν_2) denotes the dual basis to (f_0, f_1, f_2) we have

$$
d\theta = f_0(\theta)\nu_0 + f_1(\theta)\nu_1 + f_2(\theta)\nu_2.
$$

and from (17.9) 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 (17.11) are satisfied, we get

$$
d\theta = (\kappa - c)\nu_0 + \alpha_1\nu_1 + \alpha_2\nu_2 =: \eta. \tag{17.13}
$$

with the r.h.s. independent from θ.

To prove the theorem we have to show that η is an exact 1-form. Since the manifold is simply connected, it is sufficient to prove that η is closed. If we denote $\nu_{ij} := \nu_i \wedge \nu_j$ dual equations of (17.9) are:

$$
\begin{aligned}
d\nu_0 &= \nu_{12}, \\
d\nu_1 &= -c\nu_{02} + \alpha_1\nu_{12}, \\
d\nu_2 &= c\nu_{01} - \alpha_2\nu_{12}.
\end{aligned}
$$

and differentiating we get two nontrivial relations:

$$
f_1(c) + c\alpha_2 + f_0(\alpha_1) = 0, \tag{17.14}
$$

$$
f_2(c) - c\alpha_1 + f_0(\alpha_2) = 0. \tag{17.15}
$$

Recollecting all these computations we prove the closure of η

$$\begin{aligned}
d\eta &= d(\kappa - c) \wedge \nu_0 + (\kappa - c)d\nu_0 + d\alpha_1 \wedge \nu_1 + \alpha_1 d\nu_1 + d\alpha_2 \wedge \nu_2 + \alpha_2 d\nu_2 \\
&= -dc \wedge \nu_0 + (\kappa - c)\nu_{12} + \\
&\qquad + f_0(\alpha_1)\nu_{01} - f_2(\alpha_1)\nu_{12} + \alpha_1(\alpha_1\nu_{12} - c\nu_{02}) \\
&\qquad\qquad + f_0(\alpha_2)\nu_{02} + f_1(\alpha_2)\nu_{12} + \alpha_2(c\nu_{01} - \alpha_2\nu_{12}) \\
&= (f_0(\alpha_1) + \alpha_2 c + f_1(c))\nu_{01} \\
&\qquad + (f_0(\alpha_2) - \alpha_1 c + f_2(c))\nu_{02} \\
&\qquad\qquad + (\kappa - c - f_2(\alpha_1) + f_1(\alpha_2) + \alpha_1^2 + \alpha_2^2)\nu_{12} \\
&= 0.
\end{aligned}$$

where in the last equality we use (17.12) and (17.14)-(17.15). □

17.3 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 16.5 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 θ 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, θ, 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

$$\begin{cases} \dot{x} = h_1 f_1(x) + h_2 f_2(x) \\ \dot{h}_1 = \{H, h_1\} = \{h_2, h_1\}h_2 \\ \dot{h}_2 = \{H, h_2\} = -\{h_2, h_1\}h_1 \\ \dot{h}_0 = \{H, h_0\} \end{cases} \tag{17.16}$$

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

$$\begin{cases} \dot{x} = \cos\theta f_1 + \sin\theta f_2 \\ \dot{\theta} = -h_0 - b \\ \dot{h}_0 = a \end{cases} \tag{17.17}$$

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, θ, h_0) (where we use h_1, h_2 as a shorthand for $\cos\theta$ and $\sin\theta$):

$$\vec{H} = h_1 f_1 + h_2 f_2 - (h_0 + b)\partial_\theta + a\partial_{h_0} \tag{17.18}$$

$$[\partial_\theta, \vec{H}] = \vec{H}' = -h_2 f_1 + h_1 f_2 + a'\partial_{h_0} - b'\partial_\theta \tag{17.19}$$

where we denoted by $'$ the derivative with respect to θ, e.g. $h_1' = -h_2$ 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 $\partial_\theta, \partial_{h_0}, \vec{E}$. Hence the Jacobi curve is

$$J_\lambda(t) = \operatorname{span}\{e_*^{-t\vec{H}}\partial_\theta, e_*^{-t\vec{H}}\partial_{h_0}, e_*^{-t\vec{H}}\vec{E}\}$$

The first reduction, by homogeneity, let us to split the space $\Sigma_\lambda = \operatorname{span}\{\vec{E}, \vec{H}\} \oplus \operatorname{span}\{\vec{E}, \vec{H}\}^\angle$ and consider the reduced Jacobi curve $\Lambda(t) := \widehat{J}_\lambda(t)$ in the 4-dimensional symplectic space

$$\Lambda(t) := e_*^{-t\vec{H}}\widehat{\mathcal{V}}_\lambda/\mathbb{R}\vec{H} = \operatorname{span}\{e_*^{-t\vec{H}}\partial_\theta, e_*^{-t\vec{H}}\partial_{h_0}\}/\mathbb{R}\vec{H}$$

Next we describe 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 $\Gamma = \mathbb{R}\partial_{h_0}$ and $\Gamma^\angle$ is 3-dimensional in $\widehat{\mathcal{V}}_\lambda/\mathbb{R}\vec{H}$.

Proposition 17.12. *We have the following characterizations:*

(i) $\Gamma^\angle = \operatorname{span}\{\partial_{h_0}, \partial_\theta, \vec{H}'\}$ *in* $\widehat{\mathcal{V}}_\lambda/\mathbb{R}\vec{H}$*,*

(ii) $\{\partial_\theta, \vec{H}'\}$ *is a Darboux basis for* $\Gamma^\angle/\Gamma$*.*

Proof. Since ∂_{h_0} and ∂_θ are vertical to prove (i) it is enough to show that $\vec{H}'$ is skew-orthongonal to ∂_{h_0}. It is easy to compute, by Cartan 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 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 ∂_{h_0} and $[\partial_{h_0}, \vec{H}']$ are both vertical, as can be computed from (17.19).

To complete the proof of (ii) it is enough to show, using $[\partial_\theta, \vec{H}'] = -\vec{H}$, that

$$\sigma(\partial_\theta, \vec{H}') = \partial_\theta\langle s, \vec{H}'\rangle - \vec{H}'\langle s, \partial_\theta\rangle - \langle s, [\partial_\theta, \vec{H}']\rangle = \langle s, \vec{H}\rangle = 1.$$

$\square$

Next we compute the curvature in terms of the Hamiltonian vector field and its commutators. For a vector field W we use the notations

$$\dot{W} := [\vec{H}, W], \qquad W' := [\partial_\theta, W].$$

Let us consider the vector field $V_t = e_*^{-t\vec{H}} \partial_{h_0}$. Notice that

$$\dot{V}_0 = \partial_\theta, \qquad \ddot{V}_0 = -\vec{H}'.$$

The fact that ∂_θ and ∂_{h_0} are vertical implies that

$$\sigma(V_t, \dot{V}_t) = 0, \qquad \forall\, t \geq 0$$

Differentiating the above identity at $t = 0$ we get (from now on, we omit t when we evaluate at $t = 0$)

$$\sigma(\dot{V}, \dot{V}) + \sigma(V, \ddot{V}) = 0 \qquad \Longrightarrow \qquad \sigma(V, \ddot{V}) = 0.$$

Differentiating once more the last identity and using $\sigma(\dot{V}, \ddot{V}) = -\sigma(\partial_\theta, \vec{H}') = -1$ one gets

$$\sigma(\dot{V}, \ddot{V}) + \sigma(V, V^{(3)}) = 0 \qquad \Longrightarrow \qquad \sigma(V, V^{(3)}) = 1.$$

With similar computations one can show that $\sigma(\dot{V}, V^{(3)}) = \sigma(V, V^{(4)}) = 0$. Evaluating all derivatives of order 4 one can see that

$$r := \sigma(\ddot{V}, V^{(3)}) = -\sigma(\dot{V}, V^{(4)}) = \sigma(V, V^{(5)}).$$

Proposition 17.13. *The sub-Riemannian curvature is*

$$\mathcal{R} = \frac{1}{10}\sigma([\vec{H}, \vec{H}'], \vec{H}') = -\frac{r}{10}$$

Proof. The second equality follows from the definition of r and the fact that $\ddot{V} = -\vec{H}'$ and $V^{(3)} = [\vec{H}, \vec{H}']$.

To prove the first identity we have to compute the Schwartzian derivative of the bi-reduced curve, in the symplectic basis $(\dot{V}, -\ddot{V})$ of the space $\Gamma^\angle/\Gamma$ (notice the minus sign).

Recall that $\Lambda(t) = \operatorname{span}\{V_t, \dot{V}_t\}$. 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). In other words we look for $x(t)$ such that

$$\sigma(\dot{V}_t + x(t)V_t, V_0) = 0 \qquad \Longrightarrow \qquad x(t) = -\frac{\sigma(\dot{V}_t, V_0)}{\sigma(V_t, V_0)}. \tag{17.20}$$

Then we write this vector as a linear combination of the Darboux basis (cf. (16.28) for the 2D Riemannian case)

$$\dot{V}_t + x(t)V_t = \alpha(t)\dot{V}_0 - \beta(t)\ddot{V}_0 + \xi(t)V_0 \tag{17.21}$$

To see it as a curve in the space $\Gamma/\Gamma^\angle$ we simply ignore the coefficient along V_0. In these coordinates the matrix $S(t)$, which is a scalar, representing the curve is

$$S(t) = \frac{\beta(t)}{\alpha(t)} \tag{17.22}$$

Notice that this is a one-dimensional non-degenerate curve. These coefficients are computed by the symplectic products

$$\alpha(t) = -\sigma(\dot{V}_t + x(t)V_t, \ddot{V}_0) \tag{17.23}$$

$$\beta(t) = -\sigma(\dot{V}_t + x(t)V_t, \dot{V}_0) \tag{17.24}$$

Combining (17.23),(17.24) with (17.22) and (17.20) one gets

$$S(t) = \frac{\sigma(\dot{V}_t, \dot{V}_0)\sigma(V_t, V_0) - \sigma(V_t, \dot{V}_0)\sigma(\dot{V}_t, V_0)}{\sigma(\dot{V}_t, \ddot{V}_0)\sigma(\dot{V}_t, \dot{V}_0) - \sigma(\dot{V}_t, \ddot{V}_0)\sigma(\dot{V}_t, \dot{V}_0)} \tag{17.25}$$

After some computations, by Taylor expansion one gets

$$S(t) = \frac{t}{4} - \frac{t^3}{120}r + O(t^4) \tag{17.26}$$

Since $\ddot{S}_0 = 0$ the curvature is computer by

$$\mathcal{R} = \frac{\dddot{S}_0}{2\dot{S}_0} = -\frac{r}{10}$$

$\square$

We end this section by computing the expression of the curvature in terms of the orthonormal frame for the distribution and the Reeb vector filed. As usual we restrict to the level set $H^{-1}(1/2)$ where

$$h_1^2 + h_2^2 = 1, \qquad h_1 = \cos\theta, \quad h_2 = \sin\theta.$$

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 θ 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 commutator $[\vec{H}, \vec{H}']$. From (17.18) and (17.19) 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 ν_0, ν_1, ν_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 σ 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 that $d\nu_0 = \nu_1 \wedge \nu_2 = \nu_\theta \wedge \nu_{\theta'}$. Computing the symplectic product then one finds the value of

$$10\mathcal{R} = h_0^2 + \frac{3}{2}a' + \kappa$$

where

$$\kappa = f_2 c_{12}^1 - f_1 c_{12}^2 - (c_{12}^1)^2 - (c_{12}^2)^2 + \frac{c_{01}^2 - c_{02}^1}{2} \tag{17.27}$$

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 $\lambda = (h_0, h_1, h_2) \in T_x^*M$

$$10\mathcal{R} = h_0^2 + \frac{3}{2}a' + \kappa(h_1^2 + h_2^2)$$

Remark 17.14. The restriction of $\mathcal{R}$ to the 1-dimensional subspace $\lambda \in \mathcal{D}^\perp$ (that corresponds to $\lambda = (h_0, 0, 0)$), 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 ω and its normalization.

On the orthogonal complement (with respect to $\mathcal{R}$) $\{h_0 = 0\}$ we have that $\mathcal{R}$ is treated as a quadratic form

$$\mathcal{R} = \frac{3}{2}a' + \kappa(h_1^2 + h_2^2).$$

Remark 17.15. (i). If $a \neq 0$ there always exists a frame such that

$$a = 2\chi h_1 h_2$$

and in this frame we can express $\mathcal{R}$ as a quadratic form on the whole T^*M

$$\mathcal{R} = h_0^2 + (\kappa + 3\chi)h_1^2 + (\kappa - 3\chi)h_2^2.$$

It is easily seen from this formulas that we can recover the two invariants χ, κ considering

$$\mathrm{trace}(10\mathcal{R}\big|_{h_0=0}) = 2\kappa, \qquad \mathrm{discr}(10\mathcal{R}\big|_{h_0=0}) = 36\chi.$$

(ii). When $a = 0$ the eigenvalues of $\mathcal{R}$ coincide and $\chi = 0$. In this case κ represents the Riemannian curvature of the surface defined by the quotient of M with respect to the flow of the Reeb vector field.

Indeed the flow $e_*^{tf_0}$ preserves the metric and it is easy to see that the identities

$$e_*^{tf_0} f_i = f_i, \quad i = 1, 2.$$

implies $[f_0, f_1] = [f_0, f_2] = 0$. Hence $c_{01}^2, c_{02}^1 = 0$ and the expression of κ reduces to the Riemannian curvature of a surface whose orthonormal frame is f_1, f_2.

Exercise 17.16. Let f_1, f_2 be an orthonormal frame for M and denote by $\widehat{f_1}, \widehat{f_2}$ the frame obtained rotating f_1, f_2 by an angle $\theta = \theta(q)$. Show that the structure constants $\widehat{c}_{ij}^k$of rotated frame satisfies

$$\widehat{c}_{12}^1 = \cos\theta(c_{12}^1 - f_1(\theta)) - \sin\theta(c_{12}^2 - f_2(\theta)),$$
$$\widehat{c}_{12}^2 = \sin\theta(c_{12}^1 - f_1(\theta)) + \cos\theta(c_{12}^2 - f_2(\theta)).$$

Exercise 17.17. Show that the expression (17.27) for κ does not depend on the choice of an orthonormal frame f_1, f_2 for the sub-Riemannian structure.

17.4 Application: classification of 3D left-invariant structures*

In this section we exploit the local invariants χ, κ introduced before to provide a complete classification of left-invariant structures on 3D Lie groups. 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 is uniquely determined by a two dimensional subspace of the Lie algebra of the group. The distribution is bracket generating (and contact) if and only if the subspace is not a Lie subalgebra.

A standard result on the classification of 3D Lie algebras (see, for instance, [66]) reduce the analysis on the Lie algebras of the following Lie groups:

H_3, the Heisenberg group,

$A^+(\mathbb{R}) \oplus \mathbb{R}$, where $A^+(\mathbb{R})$ is the group of orientation preserving affine maps on $\mathbb{R}$,

$SOLV^+, SOLV^-$ are Lie groups whose Lie algebra is solvable and has 2-dim square,

$SE(2)$ and $SH(2)$ are the groups of orientation preserving motions of Euclidean and Hyperbolic plane respectively,

$SL(2)$ and $SU(2)$ are the three dimensional simple Lie groups.

Moreover it is easy to show that in each of these cases but one all left-invariant bracket generating distributions are equivalent by automorphisms of the Lie algebra. The only case where there exists two non-equivalent distributions is the Lie algebra $\mathfrak{sl}(2)$. More precisely a 2-dimensional subspace of $\mathfrak{sl}(2)$ is called *elliptic* (*hyperbolic*) if the restriction of the Killing form on this subspace is sign-definite (sign-indefinite). Accordingly, we use notation $SL_e(2)$ and $SL_h(2)$ to specify on which subspace the sub-Riemannian structure on $SL(2)$ is defined.

For a left-invariant structure on a Lie group the invariants χ and κ are constant functions and allow us to distinguish non isometric structures. To complete the classification we can restrict ourselves to *normalized* sub-Riemannian structures, i.e. structures that satisfy

$$\chi = \kappa = 0, \qquad \text{or} \qquad \chi^2 + \kappa^2 = 1. \tag{17.28}$$

Indeed χ and κ are homogeneous with respect to dilations of the orthonormal frame, that means rescaling of distances on the manifold. Thus we can always rescale our structure in such a way that (17.28) is satisfied.

To find missing discrete invariants, i.e. to distinguish between normalized structures with same χ and κ, we then show that it is always possible to select a canonical orthonormal frame for the sub-Riemannian structure such that all structure constants of the Lie algebra of this frame are invariant with respect to local isometries. Then the commutator relations of the Lie algebra generated by the canonical frame determine in a unique way the sub-Riemannian structure.

Falbel and Gorodski in [49] present a complete classification of 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, by means of invariants associated with an adapted connection.

In what follows we recover these result in the case of 3D Lie groups, using our invariants χ and κ, which coincide, up to a normalization factor, with those used in [49] and denoted τ_0 and K.

Theorem 17.18. *All left-invariant sub-Riemannian structures on 3D Lie groups are classified up to local isometries and dilations as in Figure 17.1, where a structure is identified by the point* (κ, χ) *and two distinct points represent non locally isometric structures.*

Moreover

(i) If $\chi = \kappa = 0$ *then the structure is locally isometric to the Heisenberg group,*

(ii) If $\chi^2 + \kappa^2 = 1$ *then there exist no more than three non isometric normalized sub-Riemannian structures with these invariants; in particular there exists a unique normalized structure on a unimodular Lie group (for every choice of* χ, κ*).*

(iii) If $\chi \neq 0$ *or* $\chi = 0, \kappa \geq 0$*, then two structures are locally isometric if and only if their Lie algebras are isomorphic.*

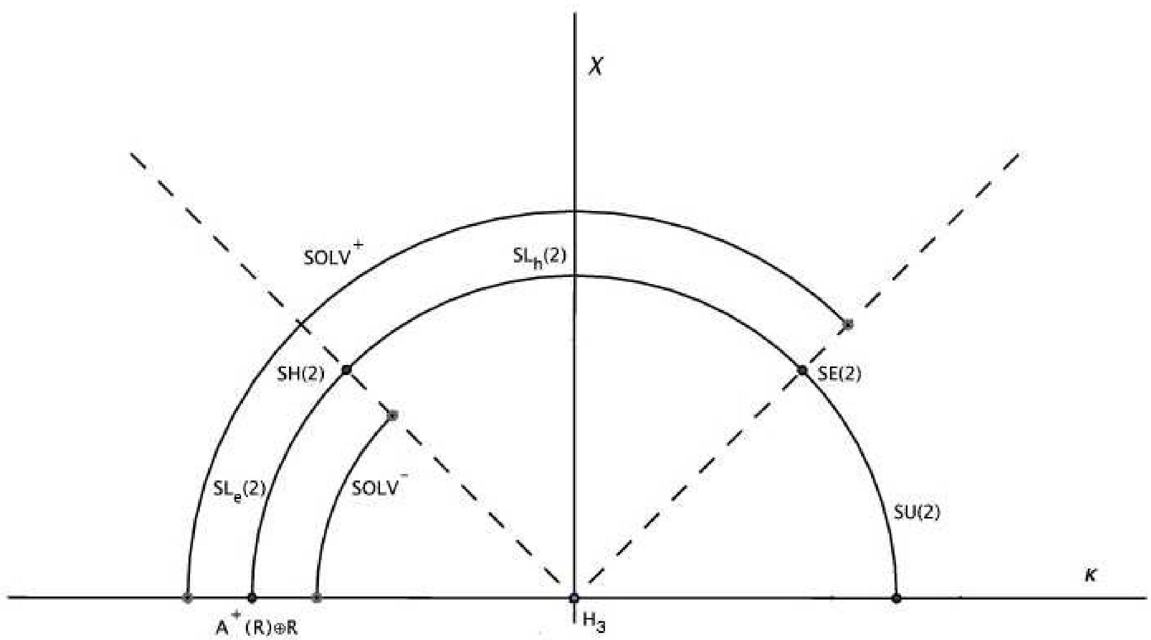


Figure 17.1: Classification

In other words every left-invariant sub-Riemannian structure is locally isometric to a normalized one that appear in Figure 17.1, where we draw points on different circles since we consider equivalence classes of structures up to dilations. In this way it is easier to understand how many normalized structures there exist for some fixed value of the local invariants. Notice that unimodular Lie groups are those that appear in the middle circle (except for $A^+(\mathbb{R}) \oplus \mathbb{R}$).

From the proof of Theorem 17.18 we get also a uniformization-like theorem for "constant curvature" manifolds in the sub-Riemannian setting:

Corollary 17.19. *Let* M *be a complete simply connected 3D contact sub-Riemannian manifold. Assume that* $\chi = 0$ *and* κ *is costant 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 H_3,

(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 covering 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 17.18, 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 17.19.

This fact was already noted in [49] as a consequence of the classification. In this paper we explicitly compute the global sub-Riemannian isometry between $A^+(\mathbb{R}) \oplus \mathbb{R}$ and the universal covering of $SL(2)$ by means of Nagano principle. 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[1]

$$A^+(\mathbb{R}) \oplus \mathbb{R} := \left\{ \begin{pmatrix} a & 0 & b \\ 0 & 1 & c \\ 0 & 0 & 1 \end{pmatrix},\ a > 0, b, c \in \mathbb{R} \right\}.$$

The standard left-invariant sub-Riemannian structure on $A^+(\mathbb{R}) \oplus \mathbb{R}$ is defined by the orthonormal frame $\mathcal{D} = \operatorname{span}\{e_2, e_1 + e_3\}$, where

$$e_1 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad e_2 = \begin{pmatrix} -1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad e_3 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix},$$

is a basis of the Lie algebra of the group, satisfying $[e_1, e_2] = e_1$.

The subgroup $A^+(\mathbb{R})$ is topologically homeomorphic to the half-plane $\{(a, b) \in \mathbb{R}^2, a > 0\}$ which can be descirbed in standard polar coordinates as $\{(\rho, \theta) | \, \rho > 0, -\pi/2 < \theta < \pi/2\}$.

Theorem 17.20. *The diffeomorphism $\Psi : A^+(\mathbb{R}) \times S^1 \longrightarrow SL(2)$ defined by*

$$\Psi(\rho, \theta, \varphi) = \frac{1}{\sqrt{\rho \cos \theta}} \begin{pmatrix} \cos \varphi & \sin \varphi \\ \rho \sin(\theta - \varphi) & \rho \cos(\theta - \varphi) \end{pmatrix}, \tag{17.29}$$

where $(\rho, \theta) \in A^+(\mathbb{R})$ and $\varphi \in S^1$, is a global sub-Riemannian isometry.

[1] We can recover the action as an affine map identifying $(x, y) \in \mathbb{R}^2$ with $(x, y, 1)^T$ and

$$\begin{pmatrix} a & 0 & b \\ 0 & 1 & c \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = \begin{pmatrix} ax + b \\ y + c \\ 1 \end{pmatrix}.$$

Using this global sub-Riemannian isometry as a change of coordinates one can recover the geometry of the sub-Riemannian structure on the group $A^+(\mathbb{R}) \times S^1$, starting from the analogous properties of $SL(2)$ (e.g. explicit expression of the sub-Riemannian distance, the cut locus).

Remark 17.21 (Comments). χ and κ are functions defined on the manifold; they reflect intrinsic geometric properties of the sub-Riemannian structure and are preserved by the sub-Riemannian isometries. In particular, χ and κ are constant functions for left-invariant structures on Lie groups (since left translations are isometries).

17.5 Proof of Theorem 17.18

Now we use the results of the previous sections to prove Theorem 17.18.

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.

From the fact that, for left-invariant structures, local invariants are constant functions (see Remark **??**) we obtain a necessary condition for two structures to be locally isometric.

Proposition 17.22. *Let G, H be 3D Lie groups with locally isometric sub-Riemannian structures. Then $\chi_G = \chi_H$ and $\kappa_G = \kappa_H$.*

Notice that this condition is not sufficient. It turns out that there can be up to three mutually non locally isometric normalized structures with the same invariants χ, κ.

Remark 17.23. It is easy to see that χ and κ are homogeneous of degree 2 with respect to dilations of the frame. Indeed assume that the sub-Riemannian structure $(M, \mathcal{D}, \mathbf{g})$ is locally defined by the orthonormal frame f_1, f_2, i.e.

$$\mathcal{D} = \operatorname{span}\{f_1, f_2\}, \qquad \mathbf{g}(f_i, f_j) = \delta_{ij}.$$

Consider now the dilated structure $(M, \mathcal{D}, \widetilde{\mathbf{g}})$ defined by the orthonormal frame $\lambda f_1, \lambda f_2$

$$\mathcal{D} = \operatorname{span}\{f_1, f_2\}, \qquad \widetilde{\mathbf{g}}(f_i, f_j) = \frac{1}{\lambda^2}\delta_{ij}, \qquad \lambda > 0.$$

If χ, κ and $\widetilde{\chi}, \widetilde{\kappa}$ denote the invariants of the two structures respectively, we find

$$\widetilde{\chi} = \lambda^2 \chi, \qquad \widetilde{\kappa} = \lambda^2 \kappa, \qquad \lambda > 0.$$

A dilation of the orthonormal frame corresponds to a multiplication by a factor $\lambda > 0$ of all distances in our manifold. Since we are interested in a classification by local isometries, we can always suppose (for a suitable dilation of the orthonormal frame) that the local invariants of our structure satisfy

$$\chi = \kappa = 0, \qquad \text{or} \qquad \chi^2 + \kappa^2 = 1,$$

and we study equivalence classes with respect to local isometries.

Since χ is non negative by definition (see Remark **??**), we study separately the two cases $\chi > 0$ and $\chi = 0$.

17.5.1 Case $\chi > 0$

Let G be a 3D Lie group with a left-invariant sub-Riemannian structure such that $\chi \neq 0$. From Proposition 17.8 we can assume that $\mathcal{D} = \mathrm{span}\{f_1, f_2\}$ where f_1, f_2 is the canonical frame of the structure. From (17.6) we obtain the dual equations

$$\begin{aligned} d\nu_0 &= \nu_1 \wedge \nu_2, \\ d\nu_1 &= c_{02}^1 \nu_0 \wedge \nu_2 + c_{12}^1 \nu_1 \wedge \nu_2, \\ d\nu_2 &= c_{01}^2 \nu_0 \wedge \nu_1 + c_{12}^1 \nu_1 \wedge \nu_2. \end{aligned} \tag{17.30}$$

Using $d^2 = 0$ we obtain structure equations

$$\begin{cases} c_{02}^1 c_{12}^2 = 0, \\ c_{01}^2 c_{12}^1 = 0. \end{cases} \tag{17.31}$$

We know that the structure constants of the canonical frame are invariant by local isometries (up to change signs of c_{12}^1, c_{12}^2, see Remark 17.9). Hence, every different choice of coefficients in (17.6) which satisfy also (17.31) will belong 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 (17.7)), 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 (17.7) 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 $\mathbb{R}^2$, and it holds $\chi = \kappa$.

(b) If $c_{01}^2 = 0$ we get the Lie algebra $\mathfrak{sh}(2)$ of the group $SH(2)$ of the Hyperbolic isometries of $\mathbb{R}^2$, and it holds $\chi = -\kappa$.

(c) If $c_{01}^2 > 0$ and $c_{02}^1 < 0$ we get the Lie algebra $\mathfrak{su}(2)$ and $\chi - \kappa < 0$.

(d) If $c_{01}^2 < 0$ and $c_{02}^1 > 0$ we get the Lie algebra $\mathfrak{sl}(2)$ with $\chi + \kappa < 0$.

(e) If $c_{01}^2 > 0$ and $c_{02}^1 > 0$ we get the Lie algebra $\mathfrak{sl}(2)$ with $\chi + \kappa > 0, \chi - \kappa > 0$.

(*ii*) $c^1_{02} = 0$ and $c^1_{12} = 0$. In this case we have

$$\begin{aligned} [f_1, f_0] &= c^2_{01} f_2, \\ [f_2, f_0] &= 0, \\ [f_2, f_1] &= c^2_{12} f_2 + f_0, \end{aligned} \tag{17.32}$$

and necessarily $c^2_{01} \neq 0$. Moreover we get

$$\chi = \frac{c^2_{01}}{2} > 0, \qquad \kappa = -(c^2_{12})^2 + \frac{c^2_{01}}{2},$$

from which it follows

$$\chi - \kappa \geq 0.$$

The Lie algebra $\mathfrak{g} = \operatorname{span}\{f_1, f_2, f_3\}$ defined by (17.32) satisfies $\dim [\mathfrak{g}, \mathfrak{g}] = 2$, hence it can be interpreted as the operator $A = \operatorname{ad} f_1$ which acts on the subspace $\operatorname{span}\{f_0, f_2\}$. Moreover, it can be easily computed that

$$\operatorname{trace} A = -c^2_{12}, \qquad \det A = c^2_{01} > 0,$$

and we can find the useful relation

$$2\frac{\operatorname{trace}^2 A}{\det A} = 1 - \frac{\kappa}{\chi}. \tag{17.33}$$

(*iii*) $c^2_{01} = 0$ and $c^2_{12} = 0$. In this last case we get

$$\begin{aligned} [f_1, f_0] &= 0, \\ [f_2, f_0] &= c^1_{02} f_1, \\ [f_2, f_1] &= c^1_{12} f_1 + f_0, \end{aligned} \tag{17.34}$$

and $c^1_{02} \neq 0$. Moreover we get

$$\chi = \frac{c^1_{02}}{2} > 0, \qquad \kappa = -(c^1_{12})^2 - \frac{c^1_{02}}{2},$$

from which it follows

$$\chi + \kappa \leq 0.$$

As before, the Lie algebra $\mathfrak{g} = \operatorname{span}\{f_1, f_2, f_3\}$ defined by (17.34) has two-dimensional square and it can be interpreted as the operator $A = \operatorname{ad} f_2$ which acts on the plane $\operatorname{span}\{f_0, f_1\}$. It can be easily seen that it holds

$$\operatorname{trace} A = c^1_{12}, \qquad \det A = -c^1_{02} < 0,$$

and we have an analogous relation

$$2\frac{\operatorname{trace}^2 A}{\det A} = 1 + \frac{\kappa}{\chi}. \tag{17.35}$$

Remark 17.24. Lie algebras of cases (*ii*) and (*iii*) are *solvable* algebras and we will denote respectively $\mathfrak{solv}^+$ and $\mathfrak{solv}^-$, where the sign depends on the determinant of the operator it represents. In particular, formulas (17.33) and (17.35) permits to recover the ratio between invariants (hence to determine a unique normalized structure) only from intrinsic properties of the operator. Notice that if $c_{12}^2 = 0$ we recover the normalized structure (*i*)-(*a*) while if $c_{12}^1 = 0$ we get the case (*i*)-(*b*).

Remark 17.25. The algebra $\mathfrak{sl}(2)$ is the only case where we can define two nonequivalent distributions which corresponds to the case that Killing form restricted on the distribution is positive definite (case (*d*)) or indefinite (case (*e*)). We will refer to the first one as the *elliptic* structure on $\mathfrak{sl}(2)$, denoted $\mathfrak{sl}_e(2)$, and with *hyperbolic* structure in the other case, denoting $\mathfrak{sl}_h(2)$.

17.5.2 Case $\chi = 0$

A direct consequence of Proposition 17.11 for left-invariant structures is the following

Corollary 17.26. *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 very 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 κ.

A straightforward calculation leads to the following list of the left-invariant structures on simply connected three dimensional Lie groups with $\chi = 0$:

- H_3 is the Heisenberg nilpotent group; then $\kappa = 0$.
- $SU(2)$ with the Killing inner product; then $\kappa > 0$.
- $\widetilde{SL}(2)$ with the elliptic distribution and Killing inner product; then $\kappa < 0$.
- $A^+(\mathbb{R}) \oplus \mathbb{R}$; then $\kappa < 0$.

Remark 17.27. In particular, we have the following:

(*i*) All left-invariant sub-Riemannian structures on H_3 are locally isometric,

(*ii*) There exists on $A^+(\mathbb{R}) \oplus \mathbb{R}$ a unique (modulo dilations) left-invariant sub-Riemannian structure, which is locally isometric to $SL_e(2)$ with the Killing metric.

Proof of Theorem 17.18 is now completed and we can recollect our result as in Figure 17.1, where we associate to every normalized structure a point in the (κ, χ) plane: either $\chi = \kappa = 0$, or (κ, χ) belong to the semicircle

$$\{(\kappa, \chi) \in \mathbb{R}^2, \chi^2 + \kappa^2 = 1, \chi > 0\}.$$

Notice that different points means that sub-Riemannian structures are not locally isometric.

17.6 Proof of Theorem 17.20

In this section we want to 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

$$\Delta_{\mathfrak{sl}} = \mathrm{span}\{g_1, g_2\}, \qquad \text{and} \qquad g_0 := -g_3, \tag{17.36}$$

is the Reeb vector field. Notice that this frame is already canonical since equations (17.10) are satisfied. Indeed

$$[g_1, g_0] = -g_2 = \kappa g_2.$$

Recall that the universal covering of $SL(2)$, which we denote $\widetilde{SL}(2)$, is a simply connected Lie group with Lie algebra $\mathfrak{sl}(2)$. Hence (17.36) define a left-invariant structure also on the universal covering.

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

$$A^+(\mathbb{R}) \oplus \mathbb{R} := \left\{ \begin{pmatrix}-y & 0 & x\\ 0 & 1 & z\\ 0 & 0 & 1\end{pmatrix}, \quad y < 0, x, z \in \mathbb{R} \right\}. \tag{17.37}$$

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

$$\begin{aligned} \mathcal{D}_{\mathfrak{a}} &= \mathrm{span}\{f_1, f_2\},\\ f_1 &:= e_2 = -y\partial_y,\\ f_2 &:= e_1 + e_3 = -y\partial_x + \partial_z. \end{aligned} \tag{17.38}$$

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 (17.10). Hence we can apply Proposition 17.11 to find the canonical frame, that will be no more left-invariant.

Following the notation of Proposition 17.11 we have

Lemma 17.28. *The canonical orthonormal frame on $A^+(\mathbb{R}) \oplus \mathbb{R}$ has the form:*

$$\begin{aligned} \widehat{f}_1 &= \quad y \sin z \, \partial_x - y \cos z \, \partial_y - \sin z \, \partial_z, \\ \widehat{f}_2 &= -y \cos z \, \partial_x - y \sin z \, \partial_y + \cos z \, \partial_z. \end{aligned} \tag{17.39}$$

Proof. It is equivalent to show that the rotation defined in the proof of Proposition 17.11 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 equation (17.13) we get $c = 0, \alpha_1 = 0, \alpha_2 = 1$. Hence

$$d\theta = -\nu_0 + \nu_2 = dz.$$

□

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 commutator relations:

$$\begin{aligned} [\widehat{f}_1, f_0] &= -\widehat{f}_2, & [g_1, g_0] &= -g_2, \\ [\widehat{f}_2, f_0] &= \widehat{f}_1, & [g_2, g_0] &= g_1, \\ [\widehat{f}_2, \widehat{f}_1] &= f_0, & [g_2, g_1] &= 0. \end{aligned} \tag{17.40}$$

Let us consider the two control systems

$$\begin{aligned} \dot{q} &= u_1 \widehat{f}_1(q) + u_2 \widehat{f}_2(q) + u_0 f_0(q), \quad q \in A^+(\mathbb{R}) \oplus \mathbb{R}, \\ \dot{x} &= u_1 g_1(x) + u_2 g_2(x) + u_0 g_0(x), \quad x \in \widetilde{SL}(2). \end{aligned}$$

and denote with $x_u(t), q_u(t),\ t \in [0, T]$ the solutions of the equations relative to the same control $u = (u_1, u_2, u_0)$. Nagano Principle (see [**?**] and also [82, 95, 96]) ensure that the map

$$\widetilde{\Psi} : A^+(\mathbb{R}) \oplus \mathbb{R} \to \widetilde{SL}(2), \qquad q_u(T) \mapsto x_u(T). \tag{17.41}$$

that sends the final point of the first system to the final point of the second one, is well-defined and does not depend on the control u.

Thus we can find the endpoint map of both systems relative to constant controls, i.e. considering maps

$$\widetilde{F} : \mathbb{R}^3 \to A^+(\mathbb{R}) \oplus \mathbb{R}, \qquad (t_1, t_2, t_0) \mapsto e^{t_0 f_0} \circ e^{t_2 \widehat{f}_2} \circ e^{t_1 \widehat{f}_1}(1_A), \tag{17.42}$$

$$\widetilde{G} : \mathbb{R}^3 \to SL(2), \qquad (t_1, t_2, t_0) \mapsto e^{t_0 g_0} \circ e^{t_2 g_2} \circ e^{t_1 g_1}(1_{SL}). \tag{17.43}$$

where we denote with 1_A and 1_{SL} identity element of $A^+(\mathbb{R}) \oplus \mathbb{R}$ and $\widetilde{SL}(2)$, respectively.

The composition of these two maps makes the following diagram commutative

$$\begin{tikzcd} A^+(\mathbb{R}) \oplus \mathbb{R} \arrow[r, "\widetilde{\Psi}"] \arrow[d, "\widetilde{F}^{-1}"'] \arrow[dr, "\Psi"] & \widetilde{SL}(2) \arrow[d, "\pi"] \\ \mathbb{R}^3 \arrow[r, "\widetilde{G}"] & SL(2) \end{tikzcd} \tag{17.44}$$

where $\pi : \widetilde{SL}(2) \to SL(2)$ is the canonical projection and we set $\Psi := \pi \circ \widetilde{\Psi}$.

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 differential equations we get from (17.42) the following expressions

$$F(t_1, t_2, t_0) = \left(2e^{-2t_1}\frac{\tanh t_2}{1+\tanh^2 t_2},\ -e^{-2t_1}\frac{1-\tanh^2 t_2}{1+\tanh^2 t_2},\ 2(\arctan(\tanh t_2) - t_0)\right). \tag{17.45}$$

The function F is globally invertible on its image and its inverse

$$F^{-1}(x, y, z) = \left(-\frac{1}{2}\log\sqrt{x^2+y^2},\ \operatorname{arctanh}(\frac{y+\sqrt{x^2+y^2}}{x}),\ \arctan(\frac{y+\sqrt{x^2+y^2}}{x}) - \frac{z}{2}\right).$$

is defined for every $y < 0$ and for every x (it is extended by continuity at $x = 0$).

On the other hand, the map (17.43) can be expressed by the product of exponential matrices as follows

$$G(t_1, t_2, t_0) = \begin{pmatrix} e^{t_1} & 0 \\ 0 & e^{-t_2} \end{pmatrix}\begin{pmatrix} \cosh t_2 & \sinh t_2 \\ \sinh t_2 & \cosh t_2 \end{pmatrix}\begin{pmatrix} \cos t_0 & -\sin t_0 \\ \sin t_0 & \cos t_0 \end{pmatrix}. \tag{17.46}$$

To simplify the computations, we consider standard polar coordinates (ρ, θ) on the half-plane $\{(x, y), y < 0\}$, where $-\pi/2 < \theta < \pi/2$ is the angle that the point (x, y) defines with y-axis. In particular, it is easy to see that the expression that appear in F^{-1} is naturally related to these coordinates:

$$\xi = \xi(\theta) := \tan\frac{\theta}{2} = \begin{cases} \dfrac{y+\sqrt{x^2+y^2}}{x}, & \text{if} \quad x \neq 0, \\ 0, & \text{if} \quad x = 0. \end{cases}$$

Hence we can rewrite

$$F^{-1}(\rho, \theta, z) = \left(-\frac{1}{2}\log\rho,\ \operatorname{arctanh}\xi,\ \arctan\xi - \frac{z}{2}\right).$$

and compute the composition $\Psi = G \circ F^{-1} : A^+(\mathbb{R}) \oplus \mathbb{R} \longrightarrow SL(2)$. Once we substitute these expressions in (17.46), the third factor is a rotation matrix by an angle $\arctan\xi - z/2$. Splitting this matrix in two consecutive rotations and using standard trigonometric identities $\cos(\arctan\xi) = \frac{1}{\sqrt{1+\xi^2}}$, $\sin(\arctan\xi) = \frac{\xi}{\sqrt{1+\xi^2}}$, $\cosh(\operatorname{arctanh}\xi) = \frac{1}{\sqrt{1-\xi^2}}$, $\sinh(\operatorname{arctanh}\xi) = \frac{\xi}{\sqrt{1-\xi^2}}$, for $\xi \in (-1, 1)$, 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

$$\Psi(\rho,\theta,z) = \begin{pmatrix} \rho^{-1/2} & 0 \\ 0 & \rho^{1/2} \end{pmatrix} \begin{pmatrix} \dfrac{1+\xi^2}{\sqrt{1-\xi^4}} & 0 \\ \dfrac{2\xi}{\sqrt{1-\xi^4}} & \dfrac{1-\xi^2}{\sqrt{1-\xi^4}} \end{pmatrix} \begin{pmatrix} \cos\dfrac{z}{2} & \sin\dfrac{z}{2} \\ -\sin\dfrac{z}{2} & \cos\dfrac{z}{2} \end{pmatrix}$$

$$= \sqrt{\frac{1+\xi^2}{1-\xi^2}} \begin{pmatrix} \rho^{-1/2} & 0 \\ 0 & \rho^{1/2} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ \dfrac{2\xi}{1+\xi^2} & \dfrac{1-\xi^2}{1+\xi^2} \end{pmatrix} \begin{pmatrix} \cos\dfrac{z}{2} & \sin\dfrac{z}{2} \\ -\sin\dfrac{z}{2} & \cos\dfrac{z}{2} \end{pmatrix}$$

$$= \frac{1}{\sqrt{\rho\cos\theta}} \begin{pmatrix} 1 & 0 \\ 0 & \rho \end{pmatrix} \begin{pmatrix} 1 & 0 \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} \cos\dfrac{z}{2} & \sin\dfrac{z}{2} \\ -\sin\dfrac{z}{2} & \cos\dfrac{z}{2} \end{pmatrix}$$

$$= \frac{1}{\sqrt{\rho\cos\theta}} \begin{pmatrix} \cos\dfrac{z}{2} & \sin\dfrac{z}{2} \\ \rho\sin(\theta-\dfrac{z}{2}) & \rho\cos(\theta-\dfrac{z}{2}) \end{pmatrix}.$$

Lemma 17.29. *The set $\Psi^{-1}(I)$ is a normal subgroup of $A^+(\mathbb{R})\oplus\mathbb{R}$.*

Proof. It is easy to show that $\Psi^{-1}(I) = \{F(0,0,2k\pi), k\in\mathbb{Z}\}$. From (17.45) we see that $F(0,0,2k\pi) = (0,-1,-4k\pi)$ and (17.37) implies that this is a normal subgroup. Indeed it is enoough to prove that $\Psi^{-1}(I)$ is a subgroup of the centre, that 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 17.30. With a standard topological argument it is possible to prove that actually $\Psi^{-1}(A)$ is a discrete countable set for every $A\in SL(2)$, and Ψ is a representation of $A^+(\mathbb{R})\oplus\mathbb{R}$ as universal covering of $SL(2)$.

By Lemma 17.29 the map Ψ is 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 Ψ on $z\in[-2\pi,2\pi]$.

If we consider the new variable $\varphi = z/2$, defined on $[-\pi,\pi]$, we can finally write the global isometry as

$$\Psi(\rho,\theta,\varphi) = \frac{1}{\sqrt{\rho\cos\theta}} \begin{pmatrix} \cos\varphi & \sin\varphi \\ \rho\sin(\theta-\varphi) & \rho\cos(\theta-\varphi) \end{pmatrix}, \tag{17.47}$$

where $(\rho,\theta)\in A^+(\mathbb{R})$ and $\varphi\in S^1$.

Remark 17.31. In the coordinate set defined above we have that $1_A = (1, 0, 0)$ and

$$\Psi(1_A) = \Psi(1, 0, 0) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = 1_{SL}.$$

On the other hand Ψ is not a homomorphism since in $A^+(\mathbb{R}) \oplus \mathbb{R}$ it holds

$$\big(\frac{\sqrt{2}}{2}, \frac{\pi}{4}, \pi\big)\big(\frac{\sqrt{2}}{2}, -\frac{\pi}{4}, -\pi\big) = 1_A,$$

while it can be easily checked from (17.47) that

$$\Psi\big(\frac{\sqrt{2}}{2}, \frac{\pi}{4}, \pi\big)\Psi\big(\frac{\sqrt{2}}{2}, -\frac{\pi}{4}, -\pi\big) = \begin{pmatrix} 2 & 0 \\ 1/2 & 1/2 \end{pmatrix} \neq 1_{SL}.$$

Bibliographical Notes

Chapter 18

Asymptotic expansion of the 3D contact exponential map

In this chapter we study the small time asymptotics of the exponential map in the three-dimensional contact case and see how the structure of the cut and the conjugate locus is encoded in the curvature.

Let us consider the sub-Riemannian Hamiltonian of a 3D contact structure (cf. Section 17.3)

$$\vec{H} = h_1 f_1 + h_2 f_2 - (h_0 + b)\partial_\theta + a\partial_{h_0} \tag{18.1}$$

written in the dual coordinates (h_0, h_1, h_2) of a local frame f_0, f_1, f_2, where ν_0 is the normalized contact form, f_0 is the Reeb vector field and f_1, f_2 is a local orthonormal frame for the sub-Riemannian structure. As usual the coordinate θ on the level set $H^{-1}(1/2)$ is defined such a way that $h_1 = \cos\theta$ and $h_2 = \sin\theta$.

In this chapter it will be convenient to introduce the notation $\rho := -h_0$ for the function linear on fibers of T^*M associated with the opposite of the Reeb vector field. The Hamiltonian system (18.1) on the level set $H^{-1}(1/2)$ is rewritten in the following form:

$$\begin{cases} \dot{q} = \cos\theta f_1 + \sin\theta f_2 \\ \dot{\theta} = \rho - b \\ \dot{\rho} = -a \end{cases} \tag{18.2}$$

The exponential map starting from the initial point $q_0 \in M$ is the map that to each time $t > 0$ and every initial covector $(\theta_0, \rho_0) \in T^*_{q_0}M \cap H^{-1}(1/2)$ assigns the first component of the solution at time t of the system (18.2), denoted by $\exp_{q_0}(t, \theta_0, \rho_0)$, or simply $\exp(t, \theta_0, \rho_0)$.

Conjugate points are points where the differential of the exponential map is not surjective, i.e. solutions to the equation

$$\frac{\partial \exp}{\partial \theta_0} \wedge \frac{\partial \exp}{\partial \rho_0} \wedge \frac{\partial \exp}{\partial t} = 0. \tag{18.3}$$

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 8.38). This implies that (18.3) is equivalent to

$$\frac{\partial \exp}{\partial \theta_0} \wedge \frac{\partial \exp}{\partial \rho_0} = 0. \tag{18.4}$$

18.1 Nilpotent case

The nilpotent case, i.e. the Heisenberg group, corresponds to the case when the functions a and b vanish identically, i.e. the system

$$\begin{cases} \dot{q} = \cos\theta f_1 + \sin\theta f_2 \\ \dot{\theta} = \rho \\ \dot{\rho} = 0 \end{cases} \tag{18.5}$$

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

$$q = (x, y), \qquad x = (x_1, x_2) \in \mathbb{R}^2, y \in \mathbb{R}. \tag{18.6}$$

Notice moreover that in this case the Reeb vector field is proportional to ∂_y and its dual coordinate ρ is constant along trajectories. There are two possible cases:

(i) $\rho = 0$. Then the solution is a straight line contained in the plane $y = 0$ and is optimal for all time.

(ii) $\rho \neq 0$. In this case we claim that the equation (18.4) is equivalent to the following

$$\frac{\partial x}{\partial \theta_0} \wedge \frac{\partial x}{\partial \rho_0} = 0. \tag{18.7}$$

By the Gauss' Lemma (Proposition 8.38) the covector $p = (p_x, \rho)$ at the final point annihilates the differential of the exponential map restricted to the level set, i.e.

$$\left\langle p, \frac{\partial \exp}{\partial \theta_0} \right\rangle = \left\langle p_x, \frac{\partial x}{\partial \theta_0} \right\rangle + \rho \frac{\partial y}{\partial \theta_0} = 0 \tag{18.8}$$

$$\left\langle p, \frac{\partial \exp}{\partial \rho_0} \right\rangle = \left\langle p_x, \frac{\partial x}{\partial \rho_0} \right\rangle + \rho \frac{\partial y}{\partial \rho_0} = 0 \tag{18.9}$$

and since $\rho \neq 0$ it follows that among the three vectors

$$\begin{pmatrix} \dfrac{\partial x_1}{\partial \theta_0} \\ \dfrac{\partial x_1}{\partial \rho_0} \end{pmatrix} \quad \begin{pmatrix} \dfrac{\partial x_2}{\partial \theta_0} \\ \dfrac{\partial x_2}{\partial \rho_0} \end{pmatrix} \quad \begin{pmatrix} \dfrac{\partial y}{\partial \theta_0} \\ \dfrac{\partial y}{\partial \rho_0} \end{pmatrix} \tag{18.10}$$

the third one is always a linear combination of the first two.

Proposition 18.1. *The first conjugate time is* $t_c(\theta_0, \rho_0) = 2\pi/|\rho_0|$.

Proof. In the standard coordinates (x_1, x_2, y) the two vector fields f_1 and f_2 defining the orthonormal frame are

$$f_1 = \partial_{x_1} - \frac{x_2}{2}\partial_y, \qquad f_2 = \partial_{x_2} + \frac{x_1}{2}\partial_y$$

Thus, the first two coordinates of the horizontal part of the Hamiltonian system satisfy

$$\begin{cases} \dot{x}_1 = \cos\theta \\ \dot{x}_2 = \sin\theta \end{cases} \tag{18.11}$$

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 $\rho > 0$)

$$x(t; \theta_0, \rho_0) = \int_0^t \begin{pmatrix} \cos(\theta_0 + \rho s) \\ \sin(\theta_0 + \rho s) \end{pmatrix} ds = \int_{\theta_0}^{\theta_0 + t} \begin{pmatrix} \cos \rho s \\ \sin \rho s \end{pmatrix} ds \tag{18.12}$$

Due to the symmetry of the Heisenberg group, the determinant of the Jacobian map will not depend on θ_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 Exercise 18.2. $\square$

Exercise 18.2. Prove that $\tau_c = 2\pi$ is the first positive root of the equation $\tau \sin \tau + 2 \cos \tau - 2 = 0$. Moreover show that τ_c is a simple root.

18.2 General case: second order asymptotic expansion

Let us consider the Hamiltonian system for the general 3D contact case

$$\begin{cases} \dot{q} = f_\theta := \cos \theta f_1 + \sin \theta f_2 \\ \dot{\theta} = \rho - b \\ \dot{\rho} = -a \end{cases} \tag{18.13}$$

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 ρ 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 18.3. *The conjugate time $t_c(\theta_0, \nu)$ is a smooth function of the parameter ν for $\nu > 0$. Moreover for $\nu \to 0$*

$$t_c(\theta_0, \nu) = 2\pi|\nu| + O(|\nu|^3).$$

Proof. Let us introduce a new time variable τ such that $\frac{dt}{d\tau} = r$. If we now denote by $\dot{F}$ the derivative of a function F with respect to the new time τ, the system (18.13) is rewritten in the new coordinate system (q, θ, r) (where we recall $r = 1/\rho$), as follows

$$\begin{cases} \dot{q} = r f_\theta \\ \dot{\theta} = 1 - rb \\ \dot{r} = r^3 a \\ \dot{t} = r \end{cases} \tag{18.14}$$

To compute the asymptotics of the conjugate time, it is also convenient to consider a system of coordinates, depending on a parameter ε, corresponding to the quasi-homogeneous blow up of the sub-Riemannian structure at q_0 and converging to the nilpotent approximation. In other words we consider the change of coordinates Φ_ε such that $f_\theta \mapsto \frac{1}{\varepsilon} f_\theta^\varepsilon$ where

$$f_\theta^\varepsilon = \widehat{f} + \varepsilon f^{(0)} + \varepsilon^2 f^{(1)} + \dots$$

Accordingly to this change of coordinates we have the equalities

$$f_i = \frac{1}{\varepsilon} f_i^\varepsilon, \qquad f_0 = \frac{1}{\varepsilon^2} f_0^\varepsilon, \qquad b = \frac{1}{\varepsilon} b^\varepsilon, \qquad a = \frac{1}{\varepsilon^2} a^\varepsilon$$

where f_0^ε is the Reeb vector field defined by the orthonormal frame $f_1^\varepsilon, f_2^\varepsilon$ (and analogously for $a^\varepsilon, b^\varepsilon$).

Let us now define, for fixed ε, the variable w such that $r = \varepsilon w$.

Proposition 18.4. *The system* (18.14) *is rewritten in these variables as follows*

$$\begin{cases} \dot{q} = w f_\theta^\varepsilon \\ \dot{\theta} = 1 - w b^\varepsilon \\ \dot{w} = \varepsilon w^3 a^\varepsilon \\ \dot{t} = \varepsilon w \end{cases} \tag{18.15}$$

Notice that the dynamical system is written in a coordinate system that depends on ε. Moreover the initial asymptotic for $\rho_0 \to \infty$, corresponding to $r \to 0$, is now reduced to fix an initial value $w(0) = 1$ and send $\varepsilon \to 0$.

Consider some linearly adapted coordinates (x, y), with $x \in \mathbb{R}^2$ and $y \in \mathbb{R}$ (cf. Definition 10.28). If we denote by $q^\varepsilon = (x^\varepsilon, y^\varepsilon)$ the solution of the horizontal part of the ε-system (18.15), 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 18.1, 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, θ_0, ρ_0) parametrizing the trajectories in the exponential map correspond to a conjugate point if the corresponding parameters $(\tau, \theta_0, \varepsilon)$ satisfy

$$\varphi(\tau, \varepsilon, \theta_0) := \frac{\partial x^\varepsilon}{\partial \theta_0} \wedge \frac{\partial x^\varepsilon}{\partial w_0}\bigg|_{w_0=1} = 0 \tag{18.16}$$

For $\varepsilon = 0$, i.e. the nilpotent approximation, the first conjugate time is $\tau_c = 2\pi$, and moreover it is a simple root. Thus one gets

$$\varphi(2\pi, 0, \theta_0) = 0, \qquad \frac{\partial \varphi}{\partial \tau}(2\pi, 0, \theta_0) \neq 0. \tag{18.17}$$

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

$$\varphi(\tau_c(\varepsilon, \theta_0), \varepsilon, \theta_0) = 0. \tag{18.18}$$

In other words $\tau_c(\varepsilon, \theta_0)$ computes the conjugate time τ associated with parameters ε, θ_0. By smoothness of τ_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 τ_c as follows

$$\tau_c(\varepsilon, \theta_0) = 2\pi + O(\varepsilon^2). \tag{18.19}$$

Differentiating the identity (18.18) with respect to ε one has

$$\frac{\partial \varphi}{\partial \tau}\frac{\partial \tau_c}{\partial \varepsilon} + \frac{\partial \varphi}{\partial \varepsilon} = 0,$$

hence, thanks to (18.17), the expansion (18.19) holds if and only if $\dfrac{\partial \varphi}{\partial \varepsilon}(2\pi, 0, \theta_0) = 0$.

Moreover differentiating the expression (18.16) with respect to ε one has

$$\frac{\partial \varphi}{\partial \varepsilon}(2\pi, 0, \theta_0) = \frac{\partial^2 x^\varepsilon}{\partial \varepsilon \partial \theta_0} \wedge \frac{\partial x^\varepsilon}{\partial w_0} - \frac{\partial^2 x^\varepsilon}{\partial \varepsilon \partial w_0} \wedge \frac{\partial x^\varepsilon}{\partial \theta_0}\bigg|_{w_0=1, \varepsilon=0, \tau=2\pi}$$

The second one vanishes since at $\varepsilon = 0$ is the Heisenberg case, whose horizontal part at $\tau = 2\pi$ does not depend on θ_0. Hence we are reduced to prove that

$$\frac{\partial^2 x^\varepsilon}{\partial \varepsilon \partial \theta_0}\bigg|_{\varepsilon=0, \tau=2\pi} = 0. \tag{18.20}$$

which is a consequence of the following lemma.

Lemma 18.5. *The quantity* $\dfrac{\partial x^\varepsilon}{\partial \varepsilon}\bigg|_{\varepsilon=0, \tau=2\pi}$ *does not depend on* θ_0.

Proof of Lemma. To prove the lemma it will be enough to find the first order expansion in ε of the solution of the system (18.15).

Recall that when $\varepsilon = 0$ the system corresponds to the Heisenberg case, i.e. we have $a^\varepsilon|_{\varepsilon=0} = 0, b^\varepsilon|_{\varepsilon=0} = 0$. This gives the expansion of w (recall that $w(0) = w_0 = 1$)

$$w(t) = w(0) + \int_0^t \varepsilon a^\varepsilon(\tau) w^3(\tau) d\tau \quad \Rightarrow \quad w = 1 + O(\varepsilon^2)$$

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 β denotes the (constant) coefficient of weight zero in the expansion of b with respect to ε.

Denoting $u(\theta) = (\cos\theta, \sin\theta)$, the equation for θ 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

$$\left.\frac{\partial \theta}{\partial \varepsilon}\right|_{\varepsilon=0} = -\int_0^t \langle \beta, u(\theta(\tau)) \rangle \, d\tau = \big\langle \beta, u'(\theta_0 + t) - u'(\theta_0) \big\rangle \tag{18.21}$$

where $u'(\theta) = (-\sin\theta, \cos\theta)$.

Next we are going to use (18.21) to compute the derivative of x^ε wrt ε. The equation for the horizontal part of (18.15) can be expanded in ε as follows

$$\dot{x}^\varepsilon = u(\theta) + \varepsilon f^{(0)}_{u(\theta)}(x) + O(\varepsilon^2)$$

where the first term is Heisenberg, and $f^{(0)}_{u(\theta)}$ is the term of weight zero of f_u, which is linear with respect to x_1 and x_2 because of the weight.[1] To compute the derivative of the solution with respect to parameter we use the following general fact

Lemma 18.6. *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 18.6 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

$$\left.\frac{\partial F^x}{\partial x}\right|_{\varepsilon=0} = 0, \qquad \left.\frac{\partial F^x}{\partial \theta}\right|_{\varepsilon=0} = u'(\bar{\theta}(t)), \qquad \left.\frac{\partial F^x}{\partial \varepsilon}\right|_{\varepsilon=0} = f^{(0)}_{u(\bar{\theta}(t))}(\bar{x}(t))$$

[1] Recall that this is the zero order part of the vector field f_u along ∂_x, hence only x variables appear and have order 1.

and we obtain the equation for $\frac{\partial x}{\partial \varepsilon}$

$$\frac{d}{dt}\frac{\partial x}{\partial \varepsilon}\bigg|_{\varepsilon=0} = \frac{\partial \theta}{\partial \varepsilon}\bigg|_{\varepsilon=0} u'(\theta_0+t) + f^{(0)}_{u(\theta_0+t)}(u'(\theta_0) - u'(\theta_0+t))$$

If we set $s = \theta_0 + t$ we can rewrite this equation

$$\frac{d}{ds}\frac{\partial x}{\partial \varepsilon}\bigg|_{\varepsilon=0} = \frac{\partial \theta}{\partial \varepsilon} u'(s) + f^{(0)}_{u(s)}(u'(\theta_0) - u'(s))$$

and integrating one has

$$\begin{aligned}\frac{\partial x}{\partial \varepsilon}\bigg|_{(2\pi,0)} = &\int_{\theta_0}^{\theta_0+2\pi} \big\langle \beta, u'(s) - u'(\theta_0)\big\rangle\, u'(s) ds \\ &+ \int_{\theta_0}^{\theta_0+2\pi} f^{(0)}_{u(s)}(u'(\theta_0) - u'(s)) ds\end{aligned}$$

In the last expression it is easy to see that all terms where θ_0 appears are zero, while the others vanish since we compute integrals of periodic functions over a period (which does not dep on θ_0). This finishes the proof of Lemma 18.5, hence the proof of the Proposition 18.3. □

18.3 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 the higher order asymptotic. 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 18.3.1.

Theorem 18.7. *In a system of local coordinates around $q_0 \in M$ one has the expansion*

$$\mathrm{Con}_{q_0}(\theta_0, \nu) = q_0 \pm \pi f_0 |\nu|^2 \pm \pi(a' f_{\theta_0} - a f_{\theta_0'})|\nu|^3 + O(|\nu|^4), \qquad \nu \to 0^{\pm}. \tag{18.22}$$

If we choose coordinates such that $a = 2\chi h_1 h_2$ one gets

$$\mathrm{Con}_{q_0}(\theta_0, \nu) = q_0 \pm \pi f_0 |\nu|^2 \pm 2\pi\chi(q_0)(\cos^3\theta f_2 - \sin^3\theta f_1)|\nu|^3 + O(|\nu|^4), \qquad \nu \to 0^{\pm}. \tag{18.23}$$

Moreover for the conjugate length we have the expansion

$$\ell_c(\theta_0, \nu) = 2\pi|\nu| - \pi\kappa|\nu|^3 + O(|\nu|^4), \qquad \nu \to 0^{\pm}. \tag{18.24}$$

Analogous formulas can be obtained for the asymptotics of the cut locus at a point q_0 where the invariant χ is non vanishing.

Theorem 18.8. *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

$$\ell_{cut}(\theta, \nu) = 2\pi|\nu| - \pi(\kappa + 2\chi\sin^2\theta)|\nu|^3 + O(\nu^4), \qquad \nu \to 0^{\pm} \tag{18.25}$$

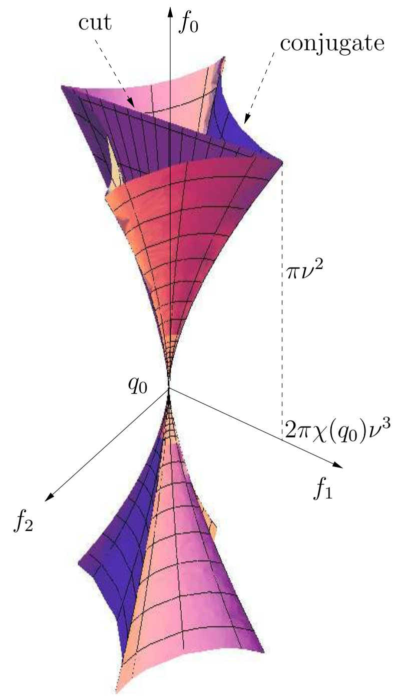


Figure 18.1: Asymptotic structure of cut and conjugate locus

We can collect the information given by the asymptotics of the conjugate and the cut loci in Figure 18.1.

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 4.34).

Let us consider the kernel of the sub-Riemannian Hamiltonian

$$\ker H = \{\lambda \in T_q^*M : \langle \lambda, v\rangle = 0,\ \forall\, v \in \mathcal{D}_q\} = \mathcal{D}_q^{\perp}. \tag{18.26}$$

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 ω and its normalization.

If we denote by $\mathcal{D}_q^*$ the orthogonal complement of $\mathcal{D}_q^{\perp}$ in the fiber with respect to $\mathcal{R}^2$, we have that $\mathcal{R}$ is a quadratic form on $\mathcal{D}_q^*$ and, by using the Euclidean metric defined by H on $\mathcal{D}_q$, as a symmetric operator.

As we explained in the previous chapter, at each q_0 where $\chi(q_0) \neq 0$ there always exists a frame such that

$$\{H, h_0\} = 2\chi h_1 h_2$$

[2] this is indeed isomorphic to the space of linear functionals defined on $\mathcal{D}_q$.

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 17.3)

$$10\mathcal{R} = (\kappa + 3\chi)h_1^2 + (\kappa - 3\chi)h_2^2.$$

From this formulae it is easy to recover the two invariants χ, κ considering

$$\mathrm{trace}(10\mathcal{R}\big|_{h_0=0}) = 2\kappa, \qquad \mathrm{discr}(10\mathcal{R}\big|_{h_0=0}) = 36\chi^2,$$

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 $\mathrm{discr}(Q) = \mathrm{trace}^2(Q) - 4\det(Q)$.

The cubic term of the conjugate locus (for a fixed value of ν) parametrizes an astroid. The cuspidal directions of the astroid are given by the eigenvectors of R, and the cut locus intersect the conjugate locus exactly at the cuspidal points in the direction of the eigenvector of R corresponding to the larger eigenvalue.

Finally the "size" of the cut locus increases for bigger values of χ, while κ is involved in the length of curves arriving at cut/conjugate locus

Remark 18.9. The expression of the cut locus given in Theorem 18.8 gives the truncation up to order 3 of the asymptotics of the cut locus of the exponential map. It is possible to show that this is actually the *exact* cut locus corresponding to the *truncated* exponential map at order 3, which is the object of the next sections (see Section 18.3.4).

18.3.1 Proof of Theorem 18.7: asymptotics of the exponential map

The proof of Theorem 18.7 requires a careful analysis of the asymptotic of the exponential map. Let us consider again our Hamiltonian system in the form (18.14)

$$\begin{cases} \dot{q} = r f_\theta \\ \dot{\theta} = 1 - rb \\ \dot{r} = r^3 a \\ \dot{t} = r \end{cases} \tag{18.27}$$

where we recall that equations are written with respect to the time τ. 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 ℓ.

Next, we consider a last change of the time variable. Namely we parametrize trajectories by the coordinate θ. 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 θ. The equations are rewritten in the following form:

$$\begin{cases} \dot{q} = \dfrac{r}{1 - rb} f_\theta \\ \dot{\theta} = 1 \\ \dot{r} = \dfrac{r^3}{1 - rb} a \\ \dot{\ell} = \dfrac{r}{1 - rb} \end{cases} \tag{18.28}$$

where we recall that $f_\theta = \cos\theta f_1 + \sin\theta f_2$. 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 θ_0 start from q_0 at time equal to θ_0.

Notice that in (18.28) 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

$$Q_t = \overrightarrow{\exp}\int_{\theta_0}^{\theta_0+t} V_\tau d\tau, \qquad V_\tau := \frac{r(\tau)}{1-r(\tau)b(\tau)} f_\tau. \tag{18.29}$$

We start by analyzing the asymptotics of the end point map after time $t = 2\pi$.

Lemma 18.10. $F(2\pi;\theta_0,\nu) = q_0 - \pi f_0(q_0)\nu^2 + O(\nu^3)$

Proof. From (18.28), 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 ν 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. (6.13))

$$\overrightarrow{\exp}\int_{\theta_0}^{\theta_0+2\pi} V_\tau d\tau = \left(\mathrm{Id} + \int_{\theta_0}^{\theta_0+2\pi} V_\tau d\tau + \iint\limits_{\theta_0\le\tau_2\le\tau_1\le\theta_0+2\pi} V_{\tau_2}\circ V_{\tau_1} d\tau_1 d\tau_2 + \dots\right) \tag{18.30}$$

and substituting $r(\tau)\equiv\nu$ we have the following expansion for the first term in (18.30):

$$\begin{aligned}\int_{\theta_0}^{\theta_0+2\pi} V_\tau d\tau = \int_{\theta_0}^{\theta_0+2\pi} \frac{\nu}{1-\nu b(\tau)} f_\tau d\tau &= \int_{\theta_0}^{\theta_0+2\pi} \nu(1+\nu b(\tau)+O(\nu^2)) f_\tau\, d\tau,\\ &= \nu\int_{\theta_0}^{\theta_0+2\pi} f_\tau d\tau + \nu^2\int_{\theta_0}^{\theta_0+2\pi} b(\tau) f_\tau d\tau + O(\nu^3)\\ &= \nu^2\int_{\theta_0}^{\theta_0+2\pi} b(\tau) f_\tau d\tau + O(\nu^3)\end{aligned}$$

Notice that the first order term in ν vanishes since we integrate over a period and $\int_{\theta_0}^{\theta_0+2\pi} f_\tau d\tau = 0$.

The second term in (18.30) can be rewritten using Lemma 8.28

$$\begin{aligned}
\iint\limits_{0\le\tau_2\le\tau_1\le t} V_{\tau_2}\circ V_{\tau_1}d\tau_1 d\tau_2 &= \frac{1}{2}\int_{\theta_0}^{\theta_0+2\pi} V_\tau d\tau \circ \int_{\theta_0}^{\theta_0+2\pi} V_\tau d\tau + \iint\limits_{\theta_0\le\tau_2\le\tau_1\le\theta_0+2\pi} [V_{\tau_2}, V_{\tau_1}]d\tau_1 d\tau_2 \\
&= \frac{\nu^2}{2}\left(\int_{\theta_0}^{\theta_0+2\pi} f_\tau d\tau \circ \int_{\theta_0}^{\theta_0+2\pi} f_\tau d\tau + \iint\limits_{\theta_0\le\tau_2\le\tau_1\le\theta_0+2\pi} [f_{\tau_2}, f_{\tau_1}]d\tau_1 d\tau_2\right) \\
&= \frac{\nu^2}{2}\iint\limits_{\theta_0\le\tau_2\le\tau_1\le\theta_0+2\pi} [f_{\tau_2}, f_{\tau_1}]d\tau_1 d\tau_2
\end{aligned}$$

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

$$\begin{aligned}
F(2\pi;\theta_0,\nu) &= q_0 + \nu^2\left(\int_{\theta_0}^{\theta_0+2\pi} b(t)f_t\,dt + \frac{1}{2}\left[\int_{\theta_0}^{t} f_\tau d\tau, f_t\right]dt\right) + O(\nu^3) \\
&= q_0 - \pi\nu^2 f_0(q_0) + O(\nu^3)
\end{aligned} \tag{18.31}$$

Notice that the factor π in (18.31) 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 18.11. $F(t;\theta_0,\nu) = F(t;\theta_0+\pi,-\nu)$

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 (18.28) is preserved while the right hand side of $\dot r$ change 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 then $(q(t+\pi), -r(t+\pi))$ is also a solution. The lemma follows. □

The symmetry property just proved permits to characterize all odd terms in the expansion in ν of the exponential map at $t = 2\pi$, as follows.

Corollary 18.12. *Consider the 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 18.11 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 θ.

Lemma 18.13. $\dfrac{\partial F}{\partial \theta_0}(2\pi;\theta_0,\nu) = -\pi[f_0, f_{\theta_0}]_{q_0}\nu^3 + O(\nu^4)$

Proof. We stress that, since we are now interested to third order term in ν, we can no more assume that $r(\tau)$ is constant. Differentiating (3.69) with respect to θ gives two terms as follows:

$$\begin{aligned}\frac{\partial F}{\partial \theta_0} &= \frac{\partial}{\partial \theta_0}(q_0 \circ Q_t) = q_0 \circ \frac{\partial}{\partial \theta_0}\left(\overrightarrow{\exp}\int_\theta^{\theta+2\pi} V_\tau d\tau\right) \\ &= q_0 \circ (Q_{2\pi} \circ V_{\theta_0+2\pi} - V_{\theta_0} \circ Q_{2\pi}) \end{aligned} \tag{18.32}$$

Next let us rewrite

$$\begin{aligned} Q_{2\pi} \circ V_{\theta_0+2\pi} &= Q_{2\pi} \circ V_{\theta_0+2\pi} \circ Q_{2\pi}^{-1} \circ Q_{2\pi} \\ &= \mathrm{Ad}\, Q_{2\pi} \circ V_{\theta_0+2\pi} \end{aligned}$$

so that (18.32) can be rewritten as

$$\frac{\partial F}{\partial \theta_0} = q_0 \circ (\mathrm{Ad}\, Q_{2\pi} \circ V_{\theta_0+2\pi} - V_{\theta_0}) \circ Q_{2\pi} \tag{18.33}$$

Thanks to Lemma 18.10 we can write

$$Q_{2\pi} = \mathrm{Id} - \pi\nu^2 f_0 + O(\nu^3) \tag{18.34}$$

that implies the following asymptotics for the action of its adjoint by (6.24)

$$\mathrm{Ad}\, Q_{2\pi} = \mathrm{Id} - \pi\nu^2 \mathrm{ad}\, f_0 + O(\nu^3)$$

We are left to compute the asymptotic expansion of (18.33). 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 ν

$$r(t) = \nu + \nu^3 \int_{\theta_0}^t a(\tau)d\tau + O(\nu^4) \tag{18.35}$$

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 allows us to replace $r(\cdot)$ with ν 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

$$\begin{aligned}\operatorname{Ad} Q_{2\pi}\circ V_{\theta_0+2\pi}-V_{\theta_0}&=(\mathrm{Id}-\pi\nu^2\operatorname{ad} f_0+O(\nu^3))\left(\frac{\nu}{1-\nu b}f_{\theta_0}\right)-\left(\frac{\nu}{1-\nu b}f_{\theta_0}\right)+O(\nu^4)\\ &=-\pi\nu^2\operatorname{ad} f_0(\nu f_{\theta_0})+O(\nu^4)\end{aligned} \tag{18.36}$$

and finally plugging (18.34) and (18.36) into (18.33) one obtains

$$\begin{aligned}\frac{\partial F}{\partial\theta}&=q_0\circ\left(-\pi\nu^2\operatorname{ad} f_0(\nu f_{\theta_0})+O(\nu^4)\right)\circ(\mathrm{Id}+O(\nu))\\ &=q_0\circ(-\pi\nu^3[f_0,f_{\theta_0}]+O(\nu^4))\end{aligned}$$

□

18.3.2 Asymptotics of the conjugate locus

In this section we finally prove Theorem 18.7, by computing the expansion of the conjugate time $t_c(\theta_0,\nu)$. We know from Proposition 18.3 that

$$\tau_c(\theta_0,\nu)=2\pi+\nu^2 s(\theta_0)+O(\nu^3)$$

By definition of conjugate point, the function $s=s(\theta_0)$ is characterized as the solution of the equation

$$\left.\frac{\partial F}{\partial s}\wedge\frac{\partial F}{\partial\theta}\wedge\frac{\partial F}{\partial\nu}\right|_{(2\pi+\nu^2 s,\theta,\nu)}=0, \tag{18.37}$$

where s is considered as a parameter. Notice that the derivative with respect to s is computed by

$$\frac{\partial F}{\partial s}=\frac{\partial F}{\partial t}\frac{\partial t}{\partial s}=(\nu f_\theta+O(\nu^2))\nu^2\simeq\nu^3 f_\theta+O(\nu^4)$$

Moreover, from the expansion of F with respect to ν 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 θ_0 together with Lemma 18.13 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 θ.

Then, collecting together all these computations, the equation for conjugate points (18.37) can be rewritten as

$$f_\theta\wedge(s f_{\theta'}+\pi[f_\theta,f_0])\wedge f_0=O(\nu) \tag{18.38}$$

Since $f_\theta, f_{\theta'}$ are an orthonormal frame on $\mathcal{D}$ and f_0 is transversal to the distribution, (18.38) is equivalent to

$$f_\theta\wedge(s f_{\theta'}+\pi[f_\theta,f_0])=O(\nu)$$

that implies

$$s(\theta) = \pi \langle [f_0, f_\theta], f_{\theta'} \rangle + O(\nu)$$

where $\langle \cdot, \cdot \rangle$ denotes the the scalar product on the distribution. Hence

$$t_c(\theta, \nu) = 2\pi + \pi\nu^2 \langle [f_0, f_\theta], f_{\theta'} \rangle_{q_0} + O(\nu^3)$$

To find the expression of conjugate locus, we evaluate the ecponential 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 ν of $\partial_t F$ is νf_θ 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 18.12 and Lemma 18.10 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

$$\begin{aligned}
\mathrm{Con}_{q_0}(\theta_0, \nu) &= q_0 - \nu^2 f_0(q_0) - \frac{\nu^3}{2} \int_{\theta_0}^{\theta_0+\pi} \pi [f_{\theta_0}, f_0] d\tau + \nu^3 s(\theta_0) f_{\theta_0} \\
&= q_0 - \nu^2 f_0(q_0) - \frac{\nu^3}{2} \int_{\theta_0}^{\theta_0+\pi} \pi [f_{\theta_0}, f_0] + s'(t) f_{\theta_0} + s(t) f_{\theta'_0} dt
\end{aligned} \tag{18.39}$$

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 (18.39) as follows

$$\begin{aligned}
\pi [f_{\theta_0}, f_0] + s'(t) f_{\theta_0} + s(t) f_{\theta'_0} &= \pi [f_{\theta_0}, f_0] + 2\pi a f_{\theta_0} + \pi \left\langle [f_0, f_{\theta_0}], f_{\theta'_0} \right\rangle f_{\theta'_0} \\
&= \pi \langle [f_{\theta_0}, f_0], f_{\theta_0} \rangle f_{\theta_0} + 2\pi a f_{\theta_0} \\
&= 3\pi a f_{\theta_0}
\end{aligned}$$

Finally

$$\begin{aligned}
\mathrm{Con}_{q_0}(\theta_0, \nu) &= q_0 - \nu^2 f_0(q_0) - \frac{3\nu^3}{2} \pi \int_{\theta_0}^{\theta_0+\pi} a(\tau) f_\tau d\tau + O(\nu^4) \\
&= q_0 - \nu^2 f_0(q_0) + \nu^3 \pi (a' f_{\theta_0} - a f_{\theta'_0}) + O(\nu^4)
\end{aligned}$$

18.3.3 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 ν. Namely

$$\ell_c(\theta_0,\nu)=\int_{\theta_0}^{\theta_0+2\pi}\frac{r(t)}{1-r(t)Q_t^{\theta_0,\nu}b(t)}dt+\nu^3 s(\theta_0)+O(\nu^4) \tag{18.40}$$

Indeed $t_c(\theta_0,\nu)=2\pi+\nu^2 s(\theta)+O(\nu^3)$ and the first order term w.r.t. ν in the integrand is exactly ν by (18.35). In what follows we use again the notation $Q_t:=Q_t^{\theta_0,\nu}$, and we compute the expansion in ν of the integral appearing in (18.40).

First notice that

$$\frac{r(t)}{1-r(t)Q_tb(t)}=r(t)\left(1+r(t)Q_tb(t)+r^2(t)[Q_tb(t)\circ Q_tb(t)]+O(r(t)^3)\right)$$

Using that $r(t)=\nu+O(\nu^3)$ and $Q_tb(t)=b(t)+O(\nu)$ we have that

$$\frac{r(t)}{1-r(t)Q_tb(t)}=r(t)+r^2(t)Q_tb(t)+r^3(t)b(t)^2+O(\nu^4)$$

Now each addend of the sum expands as follows

$$r(t)=\nu+\nu^3\int_0^t a(t)dt+O(\nu^4) \tag{18.41}$$

$$r^2(t)Q_t(\nu)b(t)=(\nu^2+O(\nu^4))\left(\mathrm{Id}+\nu\int_0^t f_\tau d\tau+O(\nu)\right)b(t) \tag{18.42}$$

$$=\nu^2 b(t)+\nu^3\int_0^t f_\tau d\tau b(t)+O(\nu^4) \tag{18.43}$$

$$r^3(t)b(t)^2=\nu^3 b(t)^2+O(\nu^4) \tag{18.44}$$

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 ν^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. $\square$

18.3.4 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 18.14. *In these coordinates we have*

$$\begin{aligned}\frac{1}{\pi}F(2\pi + \pi\eta^2\tau, \theta, \nu) &= (x_0(\tau,\theta,\nu), x_1(\tau,\theta,\nu), x_2(\tau,\theta,\nu)) \\ &= (-\nu^2, (\tau - c_{02}^1)\cos(\theta)\nu^3, (\tau + c_{01}^2)\sin(\theta)\nu^3) + O(\nu^4)\end{aligned} \tag{18.45}$$

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 (18.45) is rewritten as follows

$$\frac{1}{\pi}F(2\pi + \pi\eta^2\tau, \theta, \nu) = (-\zeta^2, (\tau - c_{02}^1)\cos(\theta)\zeta^3 + O(\zeta^4), (\tau + c_{01}^2)\sin(\theta)\zeta^3 + O(\zeta^4)) \tag{18.46}$$

Notice that the first coordinate function of this map is constant in the new variables, when ζ is constant. The map (18.46) can be interpreted as a family of maps, parametrized by ζ, depending on two variables

$$\frac{1}{\pi}F(2\pi + \pi\eta^2\tau, \theta, \nu) = (-\zeta^2, \zeta^3\Phi_\zeta(\tau,\theta)) \tag{18.47}$$

where we have defined

$$\Phi_\zeta(\tau,\theta) = ((\tau - c_{02}^1)\cos(\theta), (\tau + c_{01}^2)\sin(\theta)) + O(\zeta) \tag{18.48}$$

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

$$\tau = c_{02}^1 \sin^2(\theta) - c_{01}^2 \cos^2(\theta). \tag{18.49}$$

The critical values of this map, that is the image under the map Φ_0 of the set defined by (18.49), is the astroid

$$\mathcal{A}_0 = \{2\chi(-\sin^3(\theta), \cos^3(\theta)), \theta \in S^1\} \tag{18.50}$$

The restriction to Φ_0 to the set $\mathcal{A}_0$ is a one-to-one map. Moreover every critical point of Φ_0 is a fold or a cusp. This implies that Φ_0 is a Whitney map. Hence it is stable, in the sense of Thom-Mather theory, see [101, 56].

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 ζ.

Collecting these results, we have proved that the shape of the conjugate locus described in Figure 18.1 obtained via third order approximation of the end-point map is indeed a picture of the *true* shape.

Theorem 18.15. *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 [6].

Chapter 19

The volume in sub-Riemannian geometry

19.1 The Popp volume

For an equiregular sub-Riemannian manifold M, Popp's volume is a smooth volume which is canonically associated with the sub-Riemannian structure, and it is a natural generalization of the Riemannian one. In this chapter we define the Popp volume and we prove a general formula for its expression, written in terms of a frame adapted to the sub-Riemannian distribution.

As a first application of this result, we prove an explicit formula for the canonical sub-Laplacian, namely the one associated with Popp's volume. Finally, we discuss sub-Riemannian isometries, and we prove that they preserve Popp's volume.

19.2 Popp volume for equiregular sub-Riemannian manifolds

Recall that a distribution $\mathcal{D}$ is *equiregular* if the growth vector is constant, i.e. for each $i = 1, 2, \ldots, m$, $k_i(q) = \dim(\mathcal{D}^i_q)$ does not depend on $q \in M$. In this case the subspaces $\mathcal{D}^i_q$ are fibres of the *higher order distributions* $\mathcal{D}^i \subset TM$.

For equiregular distributions we will simply talk about growth vector and step of the distribution, without any reference to the point q.

Next, we introduce the nilpotentization of the distribution at the point q, which is fundamental for the definition of Popp's volume.

Definition 19.1. 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

$$\mathrm{gr}_q(\mathcal{D}) = \mathcal{D}_q \oplus \mathcal{D}^2_q/\mathcal{D}_q \oplus \ldots \oplus \mathcal{D}^m_q/\mathcal{D}^{m-1}_q.$$

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 [15]. (See also [78]). The definition rests on the following lemmas.

Lemma 19.2. *Let E be an inner product space and V a vector space. Let $\pi : E \to V$ be a surjective linear map. Then π induces an inner product on V such that the norm of $v \in V$ is*

$$\|v\|_V = \min\{\|e\|_E \text{ s.t. } \pi(e) = v\}\,. \tag{19.1}$$

Proof. It is easy to check that Eq. (19.1) 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 $e \in \pi^{-1}(v)$. □

Lemma 19.3. *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 $X_1, \ldots, X_n$ be a adapted basis for E, i.e. $X_1, \ldots, X_{k_i}$ 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 $X_{k_j+1}, \ldots, X_{k_{j+1}}$ 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 $\widehat{\theta}(X_1) \wedge \ldots \wedge \widehat{\theta}(X_n)$. The proof that θ does not depend on the choice of the adapted basis is "dual" to the proof of [78, Lemma 10.4]. □

The idea behind Popp's volume is to define an inner product on each $\mathcal{D}^i_q/\mathcal{D}^{i-1}_q$ 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 19.3 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 $\pi : \mathcal{D}_q \otimes \mathcal{D}_q \to \mathcal{D}^2_q/\mathcal{D}_q$

$$\pi(v \otimes w) := [V, W]_q \mod \mathcal{D}_q\,, \tag{19.2}$$

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 $\mathcal{D}$

$$\widetilde{V} = V + \sum_{i=1}^k f_i X_i\,, \qquad \widetilde{W} = W + \sum_{i=1}^k g_i X_i\,, \tag{19.3}$$

where, for $1 \leq i \leq k$, $f_i, g_i \in C^\infty(M)$ and $f_i(q) = g_i(q) = 0$. Therefore

$$[\widetilde{V}, \widetilde{W}] = [V, W] + \sum_{i=1}^k (V(g_i) - W(f_i))\, X_i + \sum_{i,j=1}^k f_i g_j [X_i, X_j]\,. \tag{19.4}$$

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 $\pi_i : \otimes^i \mathcal{D}_q \to \mathcal{D}^i_q/\mathcal{D}^{i-1}_q$

$$\pi_i(v_1 \otimes \cdots \otimes v_i) = [V_1, [V_2, \ldots, [V_{i-1}, V_i]]]_q \mod \mathcal{D}^{i-1}_q\,, \tag{19.5}$$

are well defined and do not depend on the choice of the horizontal extensions $V_1, \dots, V_i$ of $v_1, \dots, v_i$.

By the bracket-generating condition, π_i are surjective and, by Lemma 19.2, 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

$$\mathrm{gr}_q(\mathcal{D}) = \mathcal{D}_q \oplus \mathcal{D}_q^2/\mathcal{D}_q \oplus \ldots \oplus \mathcal{D}_q^m/\mathcal{D}_q^{m-1}\,, \tag{19.6}$$

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) $\mu_q \in \wedge^n \mathrm{gr}_q(\mathcal{D})^*$, 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 19.3. Namely

$$\mathcal{P}_q = \theta_q^*(\mu_q) = \mu_q \circ \theta_q\,, \tag{19.7}$$

where θ_q^* denotes the dual map and we employ the canonical identification $(\wedge^n T_qM)^* \simeq \wedge^n T_q^*M$. Eq. (19.7) 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 19.5.

Remark 19.4. 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:

$$\begin{tikzcd}
(M,\mathcal{D}) \arrow[r, "\mathcal{P}"] \arrow[d, "\mathrm{gr}_q"'] & (\wedge^n T_qM)^* \arrow[d, "\theta_q^*"] \\
\mathrm{gr}_q(\mathcal{D}) \arrow[r, "\mu"'] & (\wedge^n \mathrm{gr}_q(\mathcal{D}))^*
\end{tikzcd}$$

where μ associates the inner product space $\mathrm{gr}_q(\mathcal{D})$ with its canonical volume μ_q, and θ_q^* is the dual of the map defined in Lemma 19.3.

19.3 A formula for Popp volume

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

$$[X_i, X_j] = \sum_{l=1}^{n} c_{ij}^l X_l\,. \tag{19.8}$$

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:

$$[X_{i_1}, [X_{i_2}, \dots, [X_{i_{j-1}}, X_{i_j}]]] = \sum_{l=k_{j-1}+1}^{k_j} b_{i_1 i_2 \dots i_j}^l X_l \mod \mathcal{D}^{j-1}\,, \tag{19.9}$$

where $1 \leq i_1, \ldots, i_j \leq k$. These are a generalization of the c_{ij}^l, with an important difference: the structure constants of Eq. (19.8) 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. (19.9) 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:

$$[B_j]^{hl} = \sum_{i_1, i_2, \ldots, i_j = 1}^{k} b_{i_1 i_2 \ldots i_j}^h b_{i_1 i_2 \ldots i_j}^l \,, \qquad j = 1, \ldots, m \,, \tag{19.10}$$

with the understanding that B_1 is the $k \times k$ identity matrix. It turns out that each B_j is positive definite.

Theorem 19.5. *Let $X_1, \ldots, X_n$ be a local adapted frame, and let $\nu^1, \ldots, \nu^n$ be the dual frame. Then Popp's volume $\mathcal{P}$ satisfies*

$$\mathcal{P} = \frac{1}{\sqrt{\prod_j \det B_j}} \, \nu^1 \wedge \ldots \wedge \nu^n \,, \tag{19.11}$$

where B_j is defined by (19.10) *in terms of the adapted structure constants* (19.9).

To clarify the geometric meaning of Eq. (19.11), let us consider more closely the case $m = 2$. If $\mathcal{D}$ is a step 2 distribution, we can build a local adapted frame $\{X_1, \ldots, X_k, X_{k+1}, \ldots, X_n\}$ by completing any local orthonormal frame $\{X_1, \ldots, X_k\}$ 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 $X_{k+1}, \ldots, X_n$, 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 19.2). Therefore, the function appearing at the beginning of Eq. (19.11) is the volume of the parallelotope whose edges are $X_1, \ldots, X_n$, 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 19.5

We are now ready to prove Theorem 19.5. 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 $\mathcal{D}_q^i$ 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 $X_1, \ldots, X_k, Y_{k+1}, \ldots, Y_n$ is a local adapted frame for TM. Then

$$[X_i, X_j] = \sum_{l=k+1}^{n} b_{ij}^l Y_l \mod \mathcal{D} \,. \tag{19.12}$$

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 19.2, the norm on $\mathcal{D}^2/\mathcal{D}$ is induced by the linear map $\pi : \otimes^2\mathcal{D} \to \mathcal{D}^2/\mathcal{D}$

$$\pi(X_i \otimes X_j) = [X_i, X_j] \mod \mathcal{D}\,. \tag{19.13}$$

The vector space $\otimes^2\mathcal{D}$ inherits an inner product from the one on $\mathcal{D}$, namely $\forall X, Y, Z, W \in \mathcal{D}$, $\langle X \otimes Y, Z \otimes W\rangle = \langle X, Z\rangle\langle Y, W\rangle$. π 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

$$\|\mathcal{D}^2/\mathcal{D}\|_Y = \min\{\| \otimes^2 \mathcal{D}\|_Z \text{ s.t. } \pi(Z) = Y\}\,. \tag{19.14}$$

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. (19.14) writes

$$\|\mathcal{D}^2/\mathcal{D}\|_Y = \min\{|a| \text{ s.t. } a \cdot b^l = c^l,\ l = k+1, \dots, n\}\,, \tag{19.15}$$

where $|a|$ is the Euclidean norm of a, and the dot denotes the Euclidean inner product. Indeed, $\|\mathcal{D}^2/\mathcal{D}\|_Y$ 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 $\|\mathcal{D}^2/\mathcal{D}\|_Y^2$ in terms of the b^l, we employ the Lagrange multipliers technique. Then, we look for extremals of

$$L(a, b^{k+1}, \dots, b^n, \lambda_{k+1}, \dots, \lambda_n) = |a|^2 - 2\sum_{l=k+1}^n \lambda_l (a \cdot b^l - c^l)\,. \tag{19.16}$$

We obtain the following system

$$\begin{cases} \displaystyle\sum_{l=k+1}^n \lambda_l \cdot b^l - a = 0, \\ \displaystyle\sum_{l=k+1}^n \lambda_l b^l \cdot b^r = c^r\,, \qquad r = k+1, \dots, n. \end{cases} \tag{19.17}$$

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 (19.17) by employing the matrix B^{-1}, which has components B_{hl}^{-1}. Indeed a straightforward computation leads to

$$\|\mathcal{D}^2/\mathcal{D}\|_{c^s Y_s}^2 = c^h B_{hl}^{-1} c^l\,. \tag{19.18}$$

By polarization, the inner product on $\mathcal{D}^2/\mathcal{D}$ is defined, in the basis Y_l, by

$$\langle Y_l, Y_h\rangle_{\mathcal{D}^2/\mathcal{D}} = B_{lh}^{-1}\,. \tag{19.19}$$

Observe that B^{-1} is the Gram matrix of the vectors $Y_{k+1}, \dots, Y_n$ seen as elements of $\mathcal{D}^2/\mathcal{D}$. Then, by the definition of Popp's volume, if $\nu^1, \dots, \nu^k, \mu^{k+1}, \dots, \mu^n$ is the dual basis associated with $X_1, \dots, X_k, Y_{k+1}, \dots, Y_n$, the following formula holds true

$$\mathcal{P} = \frac{1}{\sqrt{\det B}}\, \nu^1 \wedge \cdots \wedge \nu^k \wedge \mu^{k+1} \wedge \cdots \wedge \mu^n\,. \tag{19.20}$$

General case

In the general case, the procedure above can be carried out with no difficulty. Let $X_1, \dots, X_n$ be a local adapted frame for the flag $\mathcal{D}^0 \subset \mathcal{D} \subset \mathcal{D}^2 \subset \dots \subset \mathcal{D}^m$. As usual $k_i = \dim(\mathcal{D}^i)$. For $j = 2, \dots, m$ we define the adapted structure constants $b^l_{i_1 \dots i_j} \in C^\infty(M)$ by

$$[X_{i_1}, [X_{i_2}, \dots, [X_{i_{j-1}}, X_{i_j}]]] = \sum_{l=k_{j-1}+1}^{k_j} b^l_{i_1 i_2 \dots i_j} X_l \mod \mathcal{D}^{j-1}, \tag{19.21}$$

where $1 \le i_1, \dots, i_j \le k$. Again, $b^l_{i_1 \dots i_j}$ 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 $\pi_j : \otimes^j \mathcal{D} \to \mathcal{D}^j / \mathcal{D}^{j-1}$

$$\pi_j(X_{i_1} \otimes X_{i_2} \otimes \dots \otimes X_{i_j}) = [X_{i_1}, [X_{i_2}, \dots, [X_{i_{j-1}}, X_{i_j}]]] \mod \mathcal{D}^{j-1}. \tag{19.22}$$

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 $1 \le j \le m$, the $k_j - k_{j-1}$ dimensional square matrix B_j, of components

$$[B_j]^{hl} = \sum_{i_1, i_2, \dots, i_j = 1}^{k} b^h_{i_1 i_2 \dots i_j} b^l_{i_1 i_2 \dots i_j}. \tag{19.23}$$

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 π_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 $X_{k_{j-1}+1}, \dots, X_{k_j} \in \mathcal{D}^j / \mathcal{D}^{j-1}$, in other words

$$\langle X_{k_{j-1}+l}, X_{k_{j-1}+h} \rangle_{\mathcal{D}^j / \mathcal{D}^{j-1}} = B^{-1}_{lh}. \tag{19.24}$$

Therefore, if $\nu^1, \dots, \nu^n$ is the dual frame associated with $X_1, \dots, X_n$, Popp's volume is

$$\mathcal{P} = \frac{1}{\sqrt{\prod_{j=1}^m \det B_j}} \, \nu^1 \wedge \dots \wedge \nu^n. \tag{19.25}$$

19.4 Popp volume and 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 19.6. A (local) diffeomorphism $\phi : M \to M$ is a *(local) isometry* if its differential $\phi_* : TM \to TM$ preserves the sub-Riemannian structure $(\mathcal{D}, \langle \cdot | \cdot \rangle)$, namely

i) $\phi_*(\mathcal{D}_q) = \mathcal{D}_{\phi(q)}$ for all $q \in M$,

ii) $\langle \phi_* X \,|\, \phi_* Y \rangle_{\phi(q)} = \langle X \,|\, Y \rangle_q$ for all $q \in M$, $X, Y \in \mathcal{D}_q$.

Remark 19.7. Condition *i*), which is trivial 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 ϕ_*, i.e. $\phi_*(\mathcal{D}^i_q) = \mathcal{D}^i_{\phi(q)}$, for $1 \le i \le m$.

Definition 19.8. Let M be a manifold equipped with a volume form $\mu \in \Omega^n(M)$. We say that a (local) diffeomorphism $\phi : M \to M$ is a *(local) volume preserving transformation* if $\phi^*\mu = \mu$.

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 19.9. *Sub-Riemannian (local) isometries are volume preserving transformations for Popp's volume.*

Proposition 19.9 may be false for volumes different than Popp's one. We have the following.

Proposition 19.10. *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 19.9 holds true.*

Definition 19.11. Let M be a Lie group. A sub-Riemannian structure $(M, \mathcal{D}, \langle \cdot | \cdot \rangle)$ is left invariant if $\forall g \in M$, the left action $L_g : M \to M$ is an isometry.

As a trivial consequence of Proposition 19.9 we recover a well-known result (see again [78]).

Corollary 19.12. *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* $g \in M$*.*

This section is devoted to the proof of Propositions 19.9 and 19.10.

Proof of Proposition 19.9

Let $\phi \in \mathrm{Iso}(M)$ be a (local) isometry, and $1 \leq i \leq m$. The differential ϕ_* induces a linear map

$$\widetilde{\phi}_* : \otimes^i \mathcal{D}_q \to \otimes^i \mathcal{D}_{\phi(q)} \,. \tag{19.26}$$

Moreover ϕ_* preserves the flag $\mathcal{D} \subset \ldots \subset \mathcal{D}^m$. Therefore, it induces a linear map

$$\widehat{\phi}_* : \mathcal{D}_q^i / \mathcal{D}_q^{i-1} \to \mathcal{D}_{\phi(q)}^i / \mathcal{D}_{\phi(q)}^{i-1} \,. \tag{19.27}$$

The key to the proof of Proposition 19.9 is the following lemma.

Lemma 19.13. $\widetilde{\phi}_*$ *and* $\widehat{\phi}_*$ *are isometries of inner product spaces.*

Proof. The proof for $\widetilde{\phi}_*$ is trivial. The proof for $\widehat{\phi}_*$ 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. (19.5). Namely, let $Y \in \mathcal{D}_q^i/\mathcal{D}_q^{i-1}$. Then

$$\|Y\|_{\mathcal{D}_q^i/\mathcal{D}_q^{i-1}} = \min\{\|Z\|_{\otimes \mathcal{D}_q} \text{ s.t. } \pi_i(Z) = Y\} \,. \tag{19.28}$$

As a consequence of the properties of the Lie brackets, $\pi_i \circ \widetilde{\phi}_* = \widehat{\phi}_* \circ \pi_i$. Therefore

$$\|Y\|_{\mathcal{D}_q^i/\mathcal{D}_q^{i-1}} = \min\{\|\widetilde{\phi}_* Z\|_{\otimes \mathcal{D}_{\phi(q)}} \text{ s.t. } \pi_i(\widetilde{\phi}_* Z) = \widehat{\phi}_* Y\} = \|\widehat{\phi}_* Y\|_{\mathcal{D}_{\phi(q)}^i/\mathcal{D}_{\phi(q)}^{i-1}} \,. \tag{19.29}$$

By polarization, $\widehat{\phi}_*$ is an isometry. $\square$

Since $\mathrm{gr}_q(\mathcal{D}) = \oplus_{i=1}^m \mathcal{D}_q^i/\mathcal{D}_q^{i-1}$ is an orthogonal direct sum, $\widehat{\phi}_* : \mathrm{gr}_q(\mathcal{D}) \to \mathrm{gr}_{\phi(q)}(\mathcal{D})$ 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 ϕ_* is equal to $\widehat{\phi}_*$ under such an identification, and the latter is an isometry of inner product spaces, the result follows. □

Proof of Proposition 19.10

Let μ be a volume form such that $\phi^*\mu = \mu$ for any isometry $\phi \in \mathrm{Iso}(M)$. There exists $f \in C^\infty(M)$, $f \neq 0$ such that $\mathcal{P} = f\mu$. It follows that, for any $\phi \in \mathrm{Iso}(M)$

$$f\mu = \mathcal{P} = \phi^*\mathcal{P} = (f \circ \phi)\, \phi^*\mu = (f \circ \phi)\, \mu\,, \tag{19.30}$$

where we used the $\mathrm{Iso}(M)$-invariance of Popp's volume. Then also f is $\mathrm{Iso}(M)$-invariant, namely $\phi^* f = f$ for any $\phi \in \mathrm{Iso}(M)$. By hypothesis, the action of $\mathrm{Iso}(M)$ is transitive, then f is constant. □

19.5 Hausdorff dimension and Hausdorff volume*

Bibliographical notes

The problem to define a canonical volume on a sub-Riemannian manifold was first pointed out by Brockett in his seminal paper [36], 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.

Recently, Montgomery addressed this problem in the general case (see [78, Chapter 10]). Popp's volume was first defined by Octavian Popp but introduced only in [78] (see also [3, 15]).

Chapter 20

The sub-Riemannian heat equation

In this chapter we derive the sub-Riemannian heat equation and we briefly discuss the strictly related question of how to define an 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 Lie 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 famous Gaveau-Hulanicki formula) and we briefly discuss the relation between the small-time heat kernel asymptotic and the sub-Riemannian distance.

20.1 The heat equation

To write the heat equation in a sub-Riemannian manifold, let us recall how to write it in the Riemannian context and let us see which mathematical structures are missing in the sub-Riemannian one.

20.1.1 The heat equation in the Riemannian context

Let (M, g) be an oriented[1] Riemannian manifold of dimension n and let $\mathcal{R}$ 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 ϕ 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 ϕ, i.e., how much of ϕ is flowing through the unity of surface in unitary time.

Our purpose is to get a partial differential equation describing the evolution of ϕ. The Riemannian heat equation is obtained by postulating the following two facts:

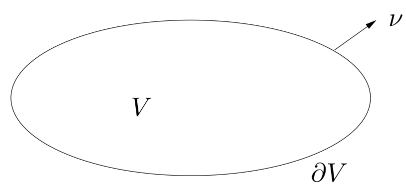
[1] we chose an oriented manifold for simplicity of presentation. In the non-orientable case, a never vanishing globally defined n form does not exist, but one can repeat the same arguments using densities. See for instance [99], Section 2.2.

(R1) the flux is proportional to minus the gradient of ϕ i.e., normalizing the proportionality constant to one, we assume that

$$\mathbf{F} = -\mathrm{grad}(\phi); \tag{20.1}$$

(R2) the quantity ϕ satisfies a conservation law, i.e. for every bounded open set V having a smooth boundary ∂V we have the following: the rate of decreasing of ϕ inside V is equal to the rate of flowing of ϕ via $\mathbf{F}$, out of V, through ∂V. In formulas this is written as

$$-\frac{d}{dt}\int_V \phi\, \mathcal{R} = \int_{\partial V} \mathbf{F}\cdot\nu\, \mathrm{dS}. \tag{20.2}$$



Here ν is the external (Riemannian) normal to ∂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 quantity $\mathbf{F}\cdot\nu$ is a notation for $g_q(\mathbf{F}(q,t),\nu(q))$.

Applying the Riemannian divergence theorem to (20.2) and using (20.1) 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

$$\triangle\phi = \mathrm{div}_{\mathcal{R}}(\mathrm{grad}(\phi)), \tag{20.3}$$

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 (20.3).

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 φ and its magnitude is the derivative of φ in that direction. In formulas it is the unique vector field $\mathrm{grad}(\varphi)$ satisfying for every $q \in M$,

$$g_q(\mathrm{grad}(\varphi), v) = d\varphi(v), \quad \text{for every } v \in T_qM. \tag{20.4}$$

In coordinates, if g is represented by a matrix (g_{ij}), and calling (g^{ij}) its inverse, we have

$$\mathrm{grad}(\varphi)^i = \sum_{j=1}^n g^{ij}\partial_j\varphi. \tag{20.5}$$

If $\{X_1, \ldots, X_n\}$ is a local orthonormal frame for g, we have the useful formula

$$\operatorname{grad}(\varphi) = \sum_{i=1}^{n} X_i(\varphi) X_i. \tag{20.6}$$

Exercise 20.1. Prove that if the Riemannian metric is defined globally via a generating family $\{X_1, \ldots, X_m\}$ with $m \geq n$, in the sense of Chapter 3, then $\operatorname{grad}(\varphi) = \sum_{i=1}^{m} X_i(\varphi) X_i$.

Recall that the divergence of a smooth vector field X says how much the flow of X is increasing or decreasing the volume. 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 "point-wise" constant of proportionality is a smooth function that by definition is the divergence of X. In formulas

$$L_X \mathcal{R} = \operatorname{div}_{\mathcal{R}}(X) \mathcal{R}.$$

Now using $d\mathcal{R} = 0$ and the Cartan formula we have that $L_X \mathcal{R} = i_X d\mathcal{R} + d(i_X \mathcal{R}) = d(i_X \mathcal{R})$. Hence the divergence of a vector field X can be defined by

$$d(i_X \mathcal{R}) = \operatorname{div}_{\mathcal{R}}(X) \mathcal{R}. \tag{20.7}$$

In coordinates, if $\mathcal{R} = h(x) dx^1 \wedge \ldots dx^n$ we have

$$\operatorname{div}_{\mathcal{R}}(X) = \frac{1}{h(x)} \sum_{i=1}^{n} \partial_i (h(x) X^i). \tag{20.8}$$

Remark 20.2. 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 20.5 and formula 20.8, with $X = \operatorname{grad}(\varphi)$ we get the well known expression for the Laplace Beltrami operator,

$$\triangle(\varphi) = \operatorname{div}_{\mathcal{R}}(\operatorname{grad}(\varphi)) = \frac{1}{h(x)} \sum_{i,j=1}^{n} \partial_i (h(x) g^{ij} \partial_j \varphi). \tag{20.9}$$

Combining formula 20.6 with the property $\operatorname{div}(aX) = a \operatorname{div}(X) + X(a)$ where X is a vector field and a is a function, we get

$$\triangle(\varphi) = \sum_{i=1}^{n} \left(X_i^2 \varphi + \operatorname{div}_{\mathcal{R}}(X_i) X_i(\varphi) \right) \quad \text{where } \{X_1, \ldots X_n\} \text{ is a local orthonormal frame.} \tag{20.10}$$

Similarly, defining the Riemannian structure via a generating family we get

$$\triangle(\varphi) = \sum_{i=1}^{m} \left(X_i^2 \varphi + \operatorname{div}_{\mathcal{R}}(X_i) X_i(\varphi) \right) \quad \text{where } \{X_1, \ldots X_m\},\ m \geq n, \text{ is a generating family} \tag{20.11}$$

Remark 20.3. 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 $\omega \neq \mathcal{R}$. In this case one would get a heat equation of the form

$$\frac{\partial}{\partial t} \phi(q,t) = \triangle \phi(q,t), \text{ where } \triangle \phi = \operatorname{div}_{\omega}(\operatorname{grad}(\phi)),$$

(to do this explicitly use Lemma 20.4 below). From Formula 20.10 one gets that the choice of the volume form does not affect the second order terms, but only the first order ones.

20.1.2 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 reason in a different way to derive the heat equation. We denote by ϕ the quantity subject to the diffusion process, and we postulate that:

(SR1) the heat flows in the direction where ϕ is varying more but only among horizontal directions;

(SR2) the quantity ϕ satisfies a conservation law, i.e. for every bounded open set V having a smooth and orientable boundary ∂V, the rate of decreasing of ϕ inside V is equal to the rate of flowing of ϕ, out of V, through ∂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 ν to ∂V, the scalar product between ν 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 φ and its magnitude is the derivative of φ in that direction. In formulas it is the unique vector field $\mathrm{grad}_H(\varphi)$ satisfying for every $q \in M$,

$$\langle \mathrm{grad}_H(\varphi) \,|\, v \rangle_q = d\varphi(v), \quad \text{for every } v \in \mathcal{D}_q M. \tag{20.12}$$

Here $\langle \cdot \,|\, \cdot \rangle_q$ is the scalar product induced by the sub-Riemannian structure on $\mathcal{D}_q$ (see Exercise 3.8).

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) is a subtle problem. In Chapter 19 we have seen that, in the equiregular case, a 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, different construction give rise to the same volume form (up to a multiplicative constant). In others cases give rise to a different volume form. We are not going to discuss here the details of this problem. Let us just recall that the three situations that one can meet are (see the bibligraphical note for some references):

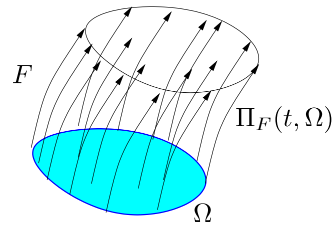


Figure 20.1:

- rank-varying or non-equiregular cases. In the first case a construction of a canonical smooth volume form is not known.
- equiregular cases for which the nilpotent approximation is the same in every point. 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 the nilpotent approximation changes with the point. 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 approximation is the same at each point and we are in the second case. For these structures Popp's volume is a left-invariant volume form and hence it coincide (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 ω (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 for exercise.

Lemma 20.4. *Let M be a smooth manifold provided with a smooth volume form ω. Let Ω be an embedded bounded sub-manifold (possible with boundary) of codimension 1. Let $F(q,t)$ be a time dependent complete smooth vector field and $P_{0,t}$ be the corresponding flow. Consider the cylinder formed by the images of Ω translated by the flow of F for times between 0 and t (see Figure 20.1):*

$$\Pi_F(t,\Omega) = \{P_{0,t}(\Omega) \mid s \in [0,t]\}.$$

Then

$$\frac{d}{dt}\bigg|_{t=0} \int_{\Pi_F(t,\Omega)} \omega = \int_\Omega i_{F(q,0)}\, \omega(q).$$

[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.

The heat equation

The postulate (SR1) consist then in declaring that the heat is flowing via a flux $\mathbf{F}$ given by

$$\mathbf{F} = -\mathrm{grad}_H(\phi).$$

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 result of the lemma.

Now, using the Stokes theorem, the definition of divergence 20.7 and using that $\mathbf{F} = -\mathrm{grad}_H\phi$ 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 20.5. Let M be a sub-Riemannian manifolds and let ω 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

$$\frac{\partial}{\partial t}\phi(q,t) = \triangle_H\phi(q,t).$$

20.1.3 Few properties of the sub-Riemannian Laplacian: the Hörmander theorem and the existence of the heat kernel

Remark 20.6. Notice that 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,

$$\triangle_H(\phi) = \sum_{i=1}^m \left(X_i^2\phi + \mathrm{div}_\omega(X_i)X_i(\phi)\right) \quad \text{where } \{X_1,\dots X_m\}, \text{ is a generating family.} \tag{20.13}$$

In the Riemannian case, the operator Δ_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, Δ_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 Lie bracket generating condition Δ_H is hypoelliptic. More precisely we have the following.

Theorem 20.7 (Hörmander)**.** *Let $\{Y_0, Y_1 \dots Y_k\}$ be a set of Lie bracket generating vector fields on a smooth manifold M. Then the operator $L = Y_0 + \sum_{i=1}^{k} Y_i^2$ is hypoellptic which means that if φ is a distribution defined on an open set $\Omega \subset M$, such that $L\varphi$ is C^∞, then φ is C^∞ in Ω.*

Notice that:

- Elliptic operators with C^∞ 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), but 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 Lie 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 Lie 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 Lie 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 $(\Delta_H - \partial_t)\varphi = 0$ exists in distributional sense in an open set Ω of $\mathbb{R} \times M$, then, being $0 \in C^\infty$, it follows that φ is C^∞ in Ω.

A standard result for the existence of a solution in $L^2(M, \omega)$ is given by the following theorem.[3]

Theorem 20.8. *Let M be a smooth manifold and ω a volume on M. If Δ is a non negative and essentially self-adjoint operator on $L^2(M, \omega)$, then, there exists a unique solution to the Cauchy problem*

$$\begin{cases} (\partial_t - \triangle)\phi = 0 \\ \phi(q, 0) = \phi_0(q) \in L^2(M, \omega), \end{cases} \tag{20.14}$$

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 Δ_H is non-negative and symmetric on $L^2(M, \omega)$. If in addition one can prove that Δ_H is essentially self-adjoint, then thanks to the Hörmander theorem, one has that the solution of (20.14) is indeed C^∞ 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 Δ_H is guaranteed by the completeness of the sub-Riemannian manifold as metric space.

Theorem 20.9 (Strichartz)**.** *Consider a sub-Riemannian manifold that is complete as metric space. Let ω be a volume on M. Then Δ_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 ω

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 (20.14) one need to specify more the problem (e.g., boundary conditions).

As a consequence of the hypoellipticity of $\triangle_H - \partial_t$, of Therem 20.8 and of Theorem 20.9, we have

Corollary 20.10. *Consider a sub-Riemannian manifold that is complete as metric space. Let ω be a volume on M. There exists a unique solution to the Cauchy problem (20.14), that is C^∞ in $]0,\infty[\times M$.*

Under the hypothesis of completeness of the manifold one can also guarantee the existence of a convolution kernel.

Theorem 20.11 (Strichartz)**.** *Consider a sub-Riemannian manifold that is complete as metric space. Let ω be a volume on M. Then the unique solution to the Cauchy problem* (20.14) *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, 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.

20.1.4 The heat equation in the non-Lie-bracket generating case

If the sub-Riemannian structure is not Lie-bracket generated, i.e., when we are dealing with a proto-sub-Riemannian structure in the sense of Section 3.1.5 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 the proto-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 Lie-bracket generating vector fields. Consider the corresponding heat operator $\Delta_H - \partial_t$ 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 $\Delta_H - \partial_t = 0$ 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 θ is the Heaviside function.

20.2 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...).

20.2.1 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

$$(x_1, y_1, z_1) \cdot (x_2, y_2, z_2) = \left(x_1 + x_2, y_1 + y_2, z_1 + z_2 + \frac{1}{2}(x_1 y_2 - x_2 y_1) \right). \tag{20.15}$$

This group law comes from the matrix product after making 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 its Lie algebra of $\mathbb{H}^1$ is $\{p_1, p_2, k\}$ where

$$p_1 = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad p_2 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}, \quad k = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \tag{20.16}$$

They satisfy the following commutation rules: $[p_1, p_2] = k$, $[p_1, k] = [p_2, k] = 0$, hence $\mathbb{H}^1$ is a 2-step nilpotent group.

Remark 20.12. Notice that if one write an element of the algebra as $xp_1 + yp_2 + zk$, one has that

$$\exp(xp_1 + yp_2 + zk) = \begin{pmatrix} 1 & x & z + \frac{1}{2}xy \\ 0 & 1 & y \\ 0 & 0 & 1 \end{pmatrix}. \tag{20.17}$$

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. They are called coordinates of the "first type". As we will see later, coordinate $x, y, w = z + \frac{1}{2}xy$, that are more adapted to the group, are also useful.

The standard sub-Riemannian structure on $\mathbb{H}^1$ is the one having as generating family:

$$X_1(g) = gp_1, \quad X_2(g) = gp_2.$$

With a straightforward computation one get 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) multiplication on $\mathbb{H}^1$:

$$L_g : \mathbb{H}^1 \ni h \mapsto gh \quad (\text{resp. } R_g : \mathbb{H}^1 \ni h \mapsto hg).$$

Exercise 20.13. Prove that, up to a multiplicative constant, there exist one and only one 3-form dh_L on $\mathbb{H}^1$ which is left-invariant, i.e. such that $L_g^* dh_L = dh_L$ and that in coordinates coincide (up to a 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 $\mathbb{H}^1$. 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

$$dh_L(X_1, X_2, [X_1, X_2]) = dh_R(X_1, X_2, [X_1, X_2]) = 1. \tag{20.18}$$

The 3-form obtained in this way on $\mathbb{H}^1$ coincide with the Lebesgue measure on $\mathbb{R}^3$ and in the following we call it simply the "Haar measure"

$$dh = dx \wedge dy \wedge dz. \tag{20.19}$$

As already remarked above, since we are on a Lie group this 3-form coincides (up to a multiplicative constant) with Popp's measure.

Exercise 20.14. Prove that (20.19) is indeed Popp's measure (i.e. that the multiplicative constant is indeed one).

Exercise 20.15. Prove that the two conditions (20.18) are invariant by change of the orthonormal frame.

20.2.2 The heat equation on the Heisenberg group

Given a volume form ω on $\mathbb{R}^3$, the sub-Riemannian Laplacian for the Heisenberg sub-Riemannian structure is given by the formula,

$$\triangle_H(\phi) = \big(X_1^2 + X_2^2 + \mathrm{div}_\omega(X_1)X_1 + \mathrm{div}_\omega(X_2)X_2\big)\phi. \tag{20.20}$$

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

$$\triangle_H(\phi) = (X_1)^2 + (X_2)^2 = (\partial_x - \frac{y}{2}\partial_z)^2 + (\partial_y + \frac{x}{2}\partial_z)^2. \tag{20.21}$$

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

$$\begin{cases} (\triangle_H - \partial_t)\phi = 0 \\ \phi(q, 0) = \phi_0(q) \in L^2(\mathbb{R}^3, dh) \end{cases} \tag{20.22}$$

can be expressed as

$$\phi(q,t)=\int_{\mathbb{R}^3} K_t(q,\bar q)\phi_0(\bar q)dh(\bar q),\quad t>0. \tag{20.23}$$

The existence of a heat kernel that is smooth, positive and symmetric is guaranteed by Theorem 20.9 since the Heisenberg group (as sub-Riemannian structure) is complete. Its explicit expression (indeed in a form of a Fourier transform) is given by the following Theorem.

Theorem 20.16 (Gaveau, Hulanicki)**.** *The heat kernel for the heat equation for the standard sub-Riemannian structure on the Heisenberg group namely for 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* (20.15)*)*

$$K_t(q,\bar q)=P_t(q^{-1}\cdot\bar q),$$

where

$$P_t(x,y,z)=\frac{1}{(2\pi t)^2}\int_{\mathbb{R}}\frac{2\tau}{\sinh(2\tau)}\exp\left(-\frac{\tau(x^2+y^2)}{2t\tanh(2\tau)}\right)\cos(2\frac{z\tau}{t})d\tau,\quad t>0. \tag{20.24}$$

Formula 20.24 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).

$$P_t(q)=K_t(q,0)=\int_{\mathbb{R}^3}K_t(q,\bar q)\delta_0(\bar q)dh(\bar q).$$

20.2.3 Construction of the Gaveau-Hunalicki 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 step:

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 than two.

STEP 3. By using the Fourier transform in two variables, we transform the heat equation (that was a PDE in three spatial variable 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 for $K_t(q,\bar{q}) = K_t(h\cdot q, h\cdot\bar{q})$ for every $h \in \mathbb{H}^1$. Taking $h = q^{-1}$ we then look for a heat kernel having the propery $K_t(q,\bar{q}) = K_t(0, q^{-1}\bar{q})$. Hence setting $q = (x,y,z)$ and $\bar{q} = (\bar{x},\bar{y},\bar{z})$ we can write

$$K_t(q,\bar{q}) = P_t(q^{-1}\cdot\bar{q}) = P_t(\bar{x}-x, \bar{y}-y, \bar{z}-z) = P_t(x-\bar{x}, y-\bar{y}, z-\bar{z}), \tag{20.25}$$

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 20.12). 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 become

$$X_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = \partial_x, \tag{20.26}$$

$$X_2 = \begin{pmatrix} 0 \\ 1 \\ x \end{pmatrix} = \partial_y + x\partial_w, \tag{20.27}$$

$$\triangle_H(\phi) = (X_1)^2 + (X_2)^2 = \partial_x^2 + (\partial_y + x\partial_w)^2. \tag{20.28}$$

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

$$\begin{cases} \partial_t\varphi(x,y,w,t) = \triangle_H(\varphi(x,y,w,t)) = \big(\partial_x^2 + (\partial_y + x\partial_w)^2\big)\varphi(x,y,w,t) \\ \varphi(x,y,w,0) = \varphi_0(x,y,w) \in L^2(\mathbb{R}^3, dh) \end{cases} \tag{20.29}$$

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

$$\begin{cases} \partial_t\hat{\varphi}(x,\mu,\nu,t) = \big(\partial_x^2 - (\mu+\nu x)^2\big)\hat{\varphi}(x,\mu,\nu,t) \\ \hat{\varphi}(x,\mu,\nu,0) = \hat{\varphi}_0(x,\mu,\nu). \end{cases} \tag{20.30}$$

By making the change of variable $x \to \theta$, where $\mu + \nu x = \nu\theta$, i.e., $\theta = x + \frac{\mu}{\nu}$ we get:

$$\begin{cases} \partial_t\bar{\varphi}^{\mu,\nu}(\theta,t) = \big(\partial_\theta^2 - \nu^2\theta^2\big)\bar{\varphi}^{\mu,\nu}(\theta,t) \\ \bar{\varphi}^{\mu,\nu}(\theta,0) = \bar{\varphi}_0^{\mu,\nu}(\theta), \end{cases} \tag{20.31}$$

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 20.17. *The solution of the Cauchy problem for the evolution of the heat in an harmonic potential, i.e.*

$$\begin{cases} \partial_t\psi(\theta,t) = \big(\partial_\theta^2 - \nu^2\theta^2\big)\psi(\theta,t) \\ \psi(\theta,0) = \psi_0(\theta) \in L^2(\mathbb{R}, d\theta) \end{cases} \tag{20.32}$$

can be written in the form of a convolution kernel

$$\psi(\theta,t)=\int_{\mathbb{R}} Q_t^\nu(\theta,\bar\theta)\psi_0(\bar\theta)d\bar\theta,\quad t>0.$$

where

$$Q_t^\nu(\theta,\bar\theta):=\sqrt{\frac{\nu}{2\pi\sinh(2\nu t)}}\exp\left(-\frac{1}{2}\frac{\nu\cosh(2\nu t)}{\sinh(2\nu t)}(\theta^2+\bar\theta^2)+\frac{\nu\theta\bar\theta}{\sinh(2\nu t)}\right). \tag{20.33}$$

Remark 20.18. In the case $\nu=0$ we interpret $Q_t^0(\theta,\bar\theta)$ in the limit

$$\lim_{\nu\to 0}Q_t^\nu(\theta,\bar\theta)=\frac{1}{\sqrt{4\pi t}}\exp\left(-\frac{|\theta-\bar\theta|^2}{4t}\right). \tag{20.34}$$

Proof. For $\nu=0$, equation (20.32) is the standard heat equation on $\mathbb{R}$ and the heat kernel is given by formula (20.34). See for instance [48]. In the following of this proof we assume $\nu>0$. The eigenvalues and the eigenfunctions of the operator $\partial_\theta^2-\nu^2\theta^2$ on $\mathbb{R}$ are[4]

$$E_j=-2\nu(j+1/2),$$
$$\Phi_j^\nu(\theta)=\frac{1}{\sqrt{2^j j!}}\left(\frac{\nu}{\pi}\right)^{\frac14}\exp\left(-\frac{\nu\theta^2}{2}\right)H_j(\sqrt{\nu}\,\theta),$$

where H_j are the Hermite polynomials $H_j(\theta)=(-1)^j\exp(\theta^2)\frac{d^j}{d\theta^j}\exp(-\theta^2)$. Being $\{\Phi_j^\nu\}_{j\in\mathbb{N}}$ an orthonormal frame of $L^2(\mathbb{R})$, we can write

$$\psi(\theta,t)=\sum_j C_j(t)\Phi_j^\nu(\theta).$$

Using equation (20.32), 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

$$Q_t^\nu(\theta,\bar\theta)=\sum_j\Phi_j^\nu(\theta)\Phi_j^\nu(\bar\theta)\exp(tE_j).$$

After some algebraic manipulations and using the Mehler formula[5] for Hermite polynomials

$$\sum_j\frac{H_j(\xi)H_j(\bar\xi)}{2^j j!}(w)^j=(1-w^2)^{-\frac12}\exp\left(\frac{2\xi\bar\xi w-(\xi^2+\bar\xi^2)w^2}{1-w^2}\right),\quad\forall\, w\in(-1,1),$$

[4] see for instance https://en.wikipedia.org/wiki/Quantum_harmonic_oscillator
[5] See for instance https://en.wikipedia.org/wiki/Hermite_polynomials

with $\xi = \sqrt{\nu}\theta$, $\bar{\xi} = \sqrt{\nu}\bar{\theta}$, $w = \exp(-2\nu t)$, one get formula (20.33). In the case $\nu < 0$ we get the same result. $\square$

Using Theorem 20.17 we can write the solution to 20.32 as

$$\bar{\varphi}^{\mu,\nu}(\theta, t) = \int_{\mathbb{R}} Q_t^{\nu}(\theta, \bar{\theta})\bar{\varphi}_0^{\mu,\nu}(\bar{\theta})d\bar{\theta}.$$

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 integration 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

$$\begin{aligned}
\varphi(x, y, w, t) &= \frac{1}{(2\pi)^2}\int_{\mathbb{R}}\int_{\mathbb{R}} \hat{\varphi}(x, \mu, \nu, t)e^{i\mu y}e^{i\nu w}d\mu\,d\nu \\
&= \int_{\mathbb{R}^3}\left(\frac{1}{(2\pi)^2}\int_{\mathbb{R}}\int_{\mathbb{R}} Q_t^{\nu}(x + \frac{\mu}{\nu}, \bar{x} + \frac{\mu}{\nu})e^{i\mu(y-\bar{y})}e^{i\nu(w-\bar{w})}d\mu\,d\nu\right)\varphi_0(\bar{x}, \bar{y}, \bar{w})d\bar{x}\,d\bar{y}\,d\bar{w}.
\end{aligned}$$

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. (20.25))

$$P_t(x, y, z) = K_t(x, y, z; 0, 0, 0) = \frac{1}{(2\pi)^2}\int_{\mathbb{R}}\int_{\mathbb{R}} Q_t^{\nu}(x + \frac{\mu}{\nu}, \frac{\mu}{\nu})e^{i\mu y}e^{i\nu(z+\frac{1}{2}xy)}d\mu\,d\nu.$$

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 (20.33)

$$\begin{aligned}
Q_t^{\nu}(x + \frac{\mu}{\nu}, \frac{\mu}{\nu}) &= A(\nu, t)\exp\left(-\frac{\nu}{2\tanh(2\nu t)}\left(\left(x + \frac{\mu}{\nu}\right)^2 + \frac{\mu^2}{\nu^2}\right) + \frac{\nu\left(x + \frac{\mu}{\nu}\right)\frac{\mu}{\nu}}{\sinh(2\nu t)}\right) \\
&= A(\nu, t)\exp\left(-\frac{\nu}{2\tanh(2\nu t)}x^2 + (\mu\nu x + \mu^2)\alpha(\nu, t)\right),
\end{aligned}$$

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 ω appear only in the term in $\mathtt{T}_0$. The integral in $d\omega$ can then be calculated. 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 τ we can substitute $e^{i\frac{2}{t}\tau z}$ with $\cos(\frac{2}{t}\tau z)$ and we get

$$P_t(x, y, z) = \frac{1}{(2\pi t)^2} \int_{\mathbb{R}} \frac{2\tau}{\sinh(2\tau)} \exp\left(-\frac{\tau(x^2 + y^2)}{2t\tanh(2\tau)} \right) \cos(2\frac{z\tau}{t}) d\tau. \tag{20.35}$$

Exercise 20.19. 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.

20.2.4 Small-time asymptotics for the Gaveau-Hulanicki fundamental solution

The integral representation (20.24) 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

$$K_t(q_0, q_0) = P_t(0, 0, 0) = \frac{1}{16t^2} \tag{20.36}$$

$$K_t(q_0, q_z) = P_t(0, 0, z) = \frac{1}{8t^2\left(1 + \cosh\left(\frac{\pi z}{t}\right)\right)} = \frac{1}{4t^2} \exp\left(-\frac{d^2(q_0, q_z)}{4t} \right) f_z(t) \tag{20.37}$$

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 $x^2 + y^2 \neq 0$

$$K_t(q_0, q) = P_t(x, y, z) = \frac{C + O(t)}{t^{3/2}} \exp\left(-\frac{d^2(q_0, q)}{4t} \right). \tag{20.38}$$

Notice that the asymptotics (20.36), (20.37), (20.38) 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

$$K_t(q_0, q) = \frac{1}{(4\pi t)^{n/2}} \exp\left(-\frac{d_E(q_0, q)^2}{4t} \right). \tag{20.39}$$

Here $q_0, q \in \mathbb{R}^n$ and d_E is the standard Euclidean norm. Comparing (20.39) with (20.36), (20.37), (20.38), 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 $\sim \frac{1}{t^2} \exp(-\frac{d^2(q_0,q)}{4t})$ 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 $\sim \frac{1}{t^{(3/2)}} \exp(-\frac{d^2(q_0,q)}{4t})$ 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 **??**). 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 value depend on the structure of optimal geodesics starting from q_0 and arriving in a neighborhood of q.

20.3 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 [36]. The problem was then studied by Montgomery in [78] who introduced the Popp measure, and in [3]. Concerning the uniqueness of an intrinsic volume see [1, 32].

For the heat equation in Riemannian geometry, we refer to [89] and references therein. For an elementary introduction in $\mathbb{R}^n$ we refer to the book of Evans [48].

Theorem 20.9 has been proved in [93, 94]. This result has been first proved in the Riemannian context in [52, 53]. In [93, 94] one finds also the proof of Theorem 20.11. For the proof of Theorem 20.8, see for instance [51]. Hörmander theorem was proved in [63]. Today there are althernative proofs based on stochastic analysis. See for instance [59, 41, 42].

The fundamental solution of the heat equation on the Heisenberg group was obtained by Gaveau using a kind of Hamilton-Jacobi theory [54] and by Hulanicki using non-commutative Fourier analysis [64]. For this second method applied on other 3-dimensional Lee groups see also [3, 17, 27]. 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 (20.36), (20.37), (20.38) have been obtained in [54]. 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 [21, 22], see also [11]. 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 [20] (out of the cut locus) and in [12, 13, 14] on the cut locus, adapting a technique due to Molchanov [75].

Bibliography

[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.

[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.

[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.

[AG78] Andrei Agrachev and Revaz Gamkrelidze. The exponential representation of flows and the chronological calculus. *Mat. Sb. (N.S.)*, 107(149)(4(12)):467–532, 1978.

[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.

[Agr96] A. A. Agrachev. Exponential mappings for contact sub-Riemannian structures. *J. Dynam. Control Systems*, 2(3):321–358, 1996.

[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.

[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.

[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.

[BAL91] 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.

[Bar11] Davide Barilari. Trace heat kernel asymptotics in 3d contact sub-Riemannian geometry. *To appear on Journal of Mathematical Sciences*, 2011.

[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.

[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.

[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.

[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.

[BBN16] Davide Barilari, Ugo Boscain, and Robert W Neel. Heat kernel asymptotics on sub-riemannian manifolds with symmetries and applications to the bi-Heisenberg group. *Ann. Fac. Sci. Toulouse, in press. ArXiv preprint arXiv:1606.01159*, 2016.

[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.

[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.

[BCG+02] 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.

[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.

[BL11] Ugo Boscain and Camille Laurent. The Laplace-Beltrami operator in almost-Riemannian geometry. *arXiv:1105.4687v1 [math.SP], Preprint*, 2011.

[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.

[BP07a] 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.

[BP07b] 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.

[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-Berlin, 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.

[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.

[CF10] Thomas Cass and Peter Friz. Densities for rough differential equations under Hörmander's condition. *Ann. of Math. (2)*, 171(3):2115–2141, 2010.

[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.

[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.

[DGN07] D. Danielli, N. Garofalo, and D. M. Nhieu. Sub-Riemannian calculus on hypersurfaces in Carnot groups. *Adv. Math.*, 215(1):292–378, 2007.

[Eul] Lehonard 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.

[FG96] Elisha Falbel and Claudio Gorodski. Sub-Riemannian homogeneous spaces in dimensions 3 and 4. *Geom. Dedicata*, 62(3):227–252, 1996.

[Fol73] G. B. Folland. A fundamental solution for a subelliptic operator. *Bull. Amer. Math. Soc.*, 79:373–376, 1973.

[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.

[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.

[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.

[Gro96] Mikhael Gromov. Carnot-Carathéodory spaces seen from within. In *Sub-Riemannian geometry*, volume 144 of *Progr. Math.*, pages 79–323. Birkhäuser, Basel, 1996.

[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.

[Hir76] Morris W. Hirsch. *Differential topology*. Springer-Verlag, New York-Heidelberg, 1976. Graduate Texts in Mathematics, No. 33.

[HL99] Francis Hirsch and Gilles Lacombe. *Elements of functional analysis*, volume 192 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1999. Translated from the 1997 French original by Silvio Levy.

[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.

[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.

[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.

[Jea14] Frédéric Jean. *Control of Nonholonomic Systems: from Sub-Riemannian Geometry to Motion Planning*. Springerbriefs in Mathematics, 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.

[Jur97] Velimir Jurdjevic. *Geometric control theory*, volume 52 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 1997.

[Jur16] Velimir Jurdjevic. *Optimal Control and Geometry: Integrable Systems*. Cambridge University Press, Cambridge, 2016.

[Kat95] Tosio Kato. *Perturbation theory for linear operators*. Classics in Mathematics. Springer-Verlag, Berlin, 1995. Reprint of the 1980 edition.

[Knö80] Horst Knörrer. Geodesics on the ellipsoid. *Invent. Math.*, 59(2):119–143, 1980.

[Lee13] John M. Lee. *Introduction to smooth manifolds*, volume 218 of *Graduate Texts in Mathematics*. Springer, New York, second edition, 2013.

[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.

[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.

[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.

[Mya02] O. Myasnichenko. Nilpotent $(3,6)$ sub-Riemannian problem. *J. Dynam. Control Systems*, 8(4):573–597, 2002.

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

[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.

[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.

[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.

[Spi79] Michael Spivak. *A comprehensive introduction to differential geometry. Vol. I.* Publish or Perish, Inc., Wilmington, Del., second edition, 1979.

[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.

[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.

[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.

[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.

Index