Introduction to
Riemannian and Sub-Riemannian geometry

(from Hamiltonian viewpoint)

Andrei Agrachev

Davide Barilari

Ugo Boscain

May 29, 2014

last version available at `http://people.sissa.it/agrachev/agrachev_files/notes.html`

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).

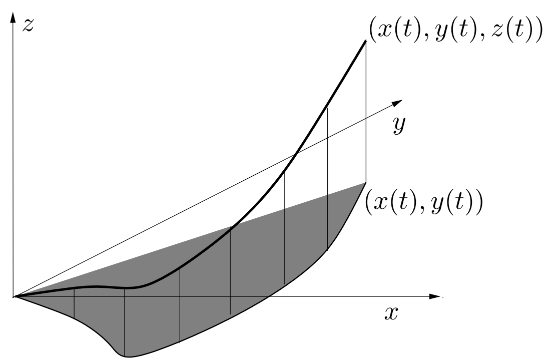


Figure 1: The Dido problem

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

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

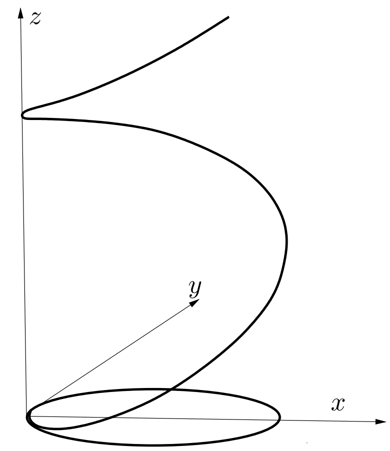


Figure 2: Solutions to the Dido problem

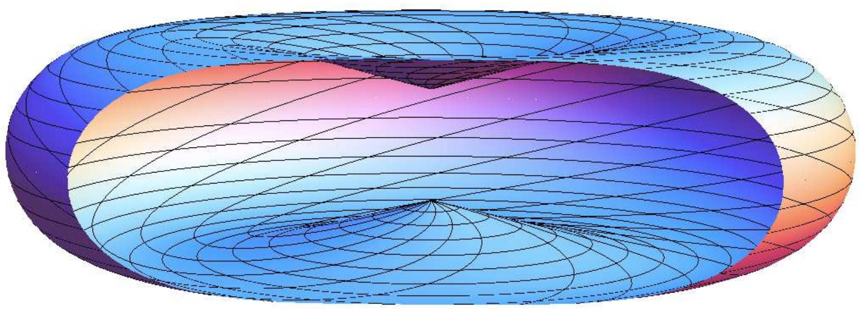


Figure 3: The Heisenberg sub-Riemannian sphere

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 8.5 and Chapter **??**.

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 it is hard to find an expert who could claim that sub-Riemannian geometry is his main field of expertise. Important motivations come from CR geometry, hyperbolic geometry, analysis of hypoelliptic operators, and some other domains. Our first motivation was control theory: length minimizing is a nice class of optimal control problems.

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

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

What is a natural parametrization of the space of geodesics? To understand this question, we adapt a classical "trajectory – wave front" duality. Given a length-parameterized geodesic $t \mapsto \gamma(t)$, we expect that the values at a fixed time t of geodesics starting at $\gamma(0)$ and close to 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".

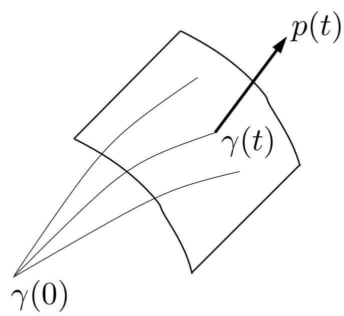


Figure 4: The "wave front" and the "impulse"

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 3.4. 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 **??**, we use this calculus for a fast introduction to the Lie group theory.

In Chapter 7, 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 conjugate points.

In Chapter 8, 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.

Chapter **??** is devoted to the explicit calculation of the sub-Riemannian distance for model spaces. In Chapter 10, 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 11, we turn to abnormal geodesics, which provide the deepest singularities of the distance. Abnormal geodesics are critical points of the endpoint map defined on the space of admissible paths, and the main tool for their study is the Hessian of the endpoint map.

This 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 $\Phi^t(T^*_qM)$, $q \in M$, 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^*_qM 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 12 is devoted to the elementary differential geometry of curves in the Lagrangian Grassmannian; in Chapter 13 we apply this geometry to Jacobi curves.

The language of Jacobi curves is translated to the traditional language in the Riemannian case in Chapter 14. We recover the Levi Civita connection and the Riemannian curvature and demonstrate their symplectic meaning. In Chapter 15, we explicitly compute the sub-Riemannian curvature for contact three-dimensional spaces. In the next Chapter 16 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.

In Chapter **??**, we consider two-dimensional sub-Riemannian metrics; such a metric differs from a Riemannian one only along a one-dimensional submanifold. In the last Chapter **??** 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.

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 this chapter we always assume $\mathbb{R}^3$ to be the ambient space, endowed with the standard Euclidean product, which we denote by $\langle \cdot, \cdot \rangle$.

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

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

Now we focus on minimizers, i.e. curves that minimize the distance between two points: $\ell(\gamma) = d(\gamma(0), \gamma(T))$.

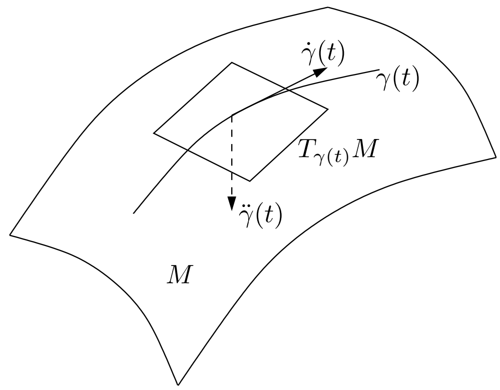


Figure 1.1: A smooth minimizer

Remark 1.3. Notice that, if $\gamma : [0,T] \to M$ is a minimizer, then the curve $\gamma|_{[0,t]}$ is also a minimizer, for all $0 < t < T$.

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) with control $u_{\tau,s}(t)$ and initial datum $x_{\tau,s}(0) = \gamma(0)$.

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

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

Proof. By Remark 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

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

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 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*

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

Proof. If $t \leq \tau$, then the conclusion follows from (1.4). 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

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

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

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

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 $a(\gamma(t)) = 0$. Computing the derivative 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 $\nu(t)$ such that

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

Hence we get

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

from which (1.12) follows. $\square$

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

Recall that every surface can be locally characterized as a regular level set of a smooth function (cf. Definition 1.1), hence equations (1.12) always characterize geodesics locally. 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. $\square$

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. $d(\gamma(\tau), \gamma(\tau+\varepsilon)) = \ell(\gamma|_{[\tau,\tau+\varepsilon]})$,

(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 vectors

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

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 ξ defined on W

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

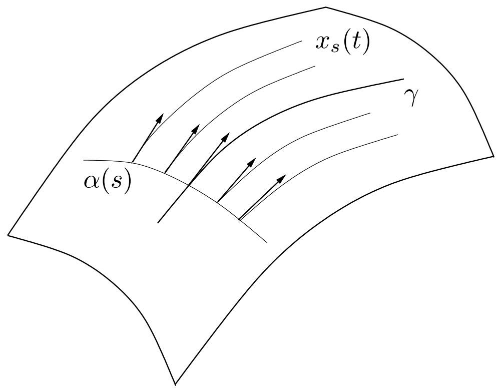


Figure 1.2: Proof of Theorem 1.12

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

Proof of Lemma 1.13. We first show that they 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 ξ. Actually they coincide since

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

$\square$

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

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

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

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

$$= \int_0^{\varepsilon'} \left\langle \nabla\phi(\beta(t)) \middle| \dot\beta(t) \right\rangle dt$$

$$= \int_0^{\varepsilon'} \left\langle \xi(\beta(t)) \middle| \dot\beta(t) \right\rangle dt \leq \varepsilon' = \ell(\beta), \tag{1.17}$$

The last inequality follows from the Cauchy-Schwartz inequality

$$\left\langle \xi(\beta(t)) \middle| \dot\beta(t) \right\rangle \leq \|\xi(\beta(t))\| \|\dot\beta(t)\| = 1 \tag{1.18}$$

which holds at every smooth point of $\beta(t)$. In addition, equality in (1.18) holds if and only if $\dot\beta(t) = \xi(\beta(t))$ (at the smooth points of β). Hence we get that $\ell(\beta) = \ell(\gamma|_{[0,\varepsilon]})$ if and only if β 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 interpret the integral in the Lebesgue sense.

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

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 $M = \mathbb{R}^2 \subset \mathbb{R}^3$ 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 $\langle \nabla_{\gamma(t)}a, \xi(t) \rangle = 0$. 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. $\square$

Remark 1.19. Notice that, since (1.19) 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 $\xi(0) = v$. Moreover the operator $\xi(0) \mapsto \xi(t)$ 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 scalar 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

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

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 $\varphi \in \mathcal{C}^\infty(M)$.

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) By Remark 1.23 the coordinate θ is defined up to a constant. However the vector field on SM defined by $\partial/\partial\theta$ is well-defined. Let us show that

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

Indeed consider a curve $t \mapsto \xi(t)$ on SM which corresponds to a rotation of on a single fibre. Then we have that velocity of this curve is exactly its orthogonal vector, i.e. $\dot{\xi}(t) = \eta(t)$ and equality above is proved. It remains to show that coefficients a_1, a_2 do not depend on θ, but this is due to the rotational invariance of ω.

(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 1.26, is the pull-back of an intrinsic 2-form on M, that in general is not exact. Since any 2-form on M is proportional to the area form dV, it makes sense to give the following definition:

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

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

Note that κ does not depend on the orientation of M, since both Ω and dV change sign if we reverse the orientation.

1.3 Gauss-Bonnet Theorems

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

Definition 1.29. Let $\gamma : [0, T] \to M$ be a smooth curve parametrized by 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 1.30. The geodesic curvature changes sign if we move along the curve in the opposite direction. Moreover, if $M = \mathbb{R}^2$, it coincides with the usual notion of curvature of a planar curve.

1.3.1 Gauss-Bonnet theorem: local version

Definition 1.31. 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 $\partial\Gamma = \cup_i \gamma_i(I_i)$ 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).

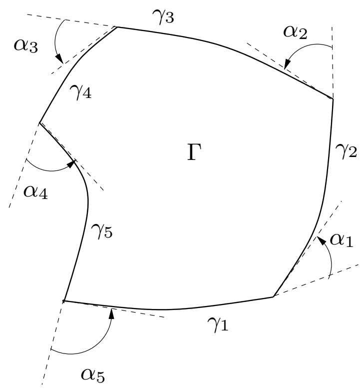


Figure 1.3: A curvilinear polygon

Notice that a curvilinear polygon is homeomorphic to a disk.

Theorem 1.32. *(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 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 $q_0 = F(0)$.

We refer to Figure 1.4. For ε small enough, consider

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

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

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

and compute the integral of the curvature κ on Γ_ε. First notice that

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

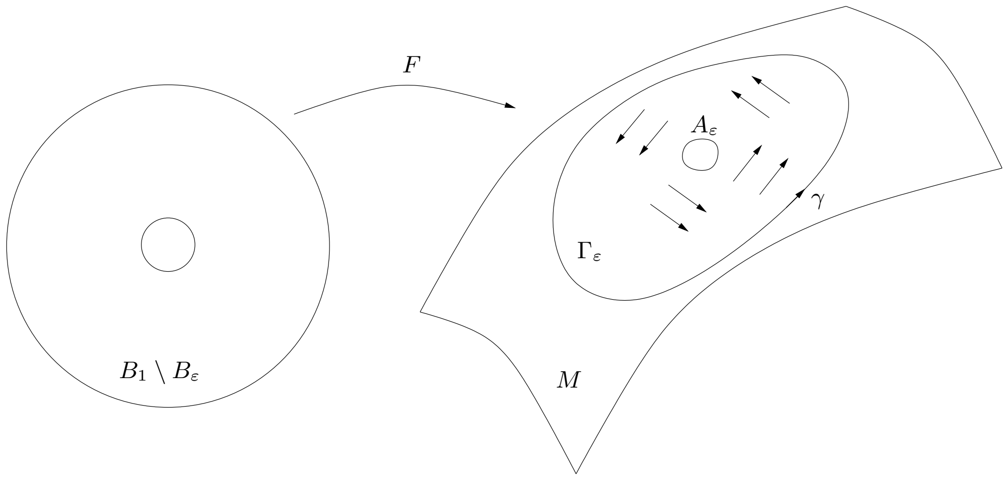


Figure 1.4: The map F

where we used the fact that $\pi(\phi(\Gamma_\varepsilon)) = \Gamma_\varepsilon$. Thus

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

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 1.26, 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. $\square$

Remark 1.33. Let us consider a geodesic triangle Γ, i.e. a curvilinear polygon with $m = 3$ and such that every γ_i is a geodesic. Denote with $A_i := \pi - \alpha_i$ the internal angles of the geodesic triangle. Using that the geodesic curvature of γ_i vanishes, the local version of Gauss-Bonnet Theorem (1.27) can be rewritten as

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

This formula shows that the Gaussian curvature measures how much the manifold M is far from being an Euclidean plane (that corresponds to the case $\kappa = 0$).

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 1.31). Notice that every compact surface admits a triangulation.[3]

Definition 1.34. Let $M \subset \mathbb{R}^3$ be a compact oriented surface with 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 1.35. 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, with show that the quantity (1.34) is independent on the triangulation.

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

Following the notation introduced in the previous section, for a given region Γ, we assume that $\partial\Gamma$ is oriented consistently with the orientation of M and $\partial\Gamma = \cup_i \gamma_i(I_i)$ 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).

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

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

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

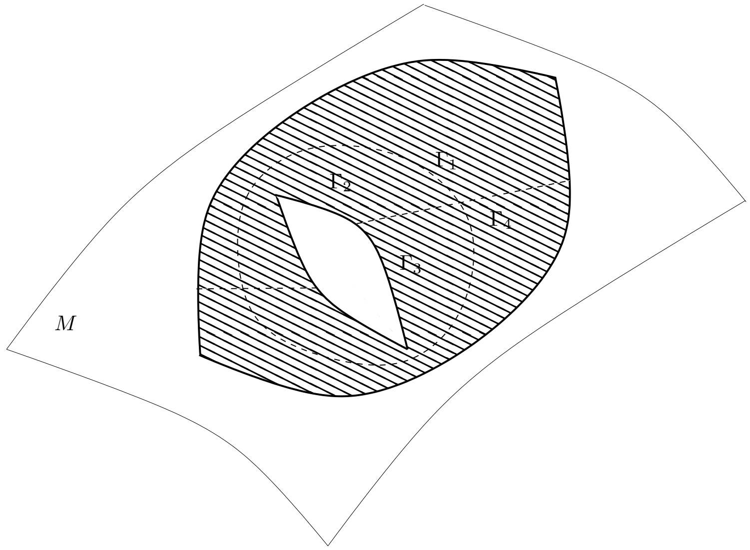


Figure 1.5: Gauss B

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 $\{\Gamma_j, j = 1, \ldots, n_2\}$ 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 1.32 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 elements 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

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

Note that 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:

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

$\square$

Corollary 1.38. *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 1.39. Let M, M' be two surfaces in $\mathbb{R}^3$. A map $\phi : M \to M'$ is called an *isometry* if for every $q \in M$ it satisfies

$$\langle v|w\rangle_M = \langle D_q\phi(v)|D_q\phi(w)\rangle_{M'}, \qquad \forall\, v, w \in T_qM \tag{1.44}$$

If the property (1.44) is satisfied by a map $\phi : U \subset M \to M'$ defined in a neighborhood U of q, 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'.

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.

From the local version of the Gauss-Bonnet Theorem (in particular from formula (1.33) and the fact that the angles are preserved by isometries) we obtain that the Gaussian curvature is intrinsic, in the following sense

Corollary 1.40. *(Gauss's Theorema Egregium)*
The Gaussian curvature is invariant by local isometries.

More precisely if $\phi : U \subset M \to M'$ is a local isometry, then for every $q \in U$ one has $\kappa(q) = \kappa'(\phi(q))$, where κ (resp. κ') is the Gaussian curvature of M (resp. M').

Corollary 1.41. *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 1.42. 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 1.41 we have

Theorem 1.43. *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 1.42. $\square$

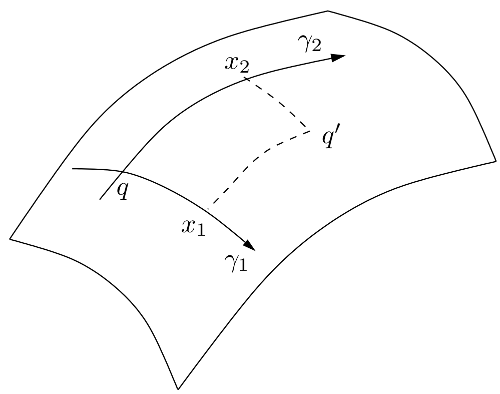


Figure 1.6: Proof of Theorem 1.43

1.3.4 The Gauss map

We end this section with a geometric characterization of the Gaussian curvature of a manifold M, using the Gauss map.

Definition 1.44. Let M be an oriented surface. We define the *Gauss map*

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

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

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

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

Lemma 1.46. *The differential of the Gauss map $D_q\mathcal{N}$ is a symmetric map, 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

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

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

Proof of Theorem 1.45. 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 relative to ξ 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 $\Omega = (\omega_{ij})$ (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 1.24). Moreover, since $e_3 = \mathcal{N} \circ \pi$, where $\pi : SM \to M$ is the canonical projection, one has

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

The proof is completed by the following

Exercise 1.47. Let V be a 2-dimensional Euclidean vector space and e_1, e_2 an orthonormal basis. Let $F : V \mapsto V$ 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.

□

Remark 1.48. Lemma 1.46 allows us to define the *principal curvatures* of M at the point q as the two real eigenvalues $k_1(q), k_2(q)$ of the map $D_q\mathcal{N}$. In particular

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

The principal curvatures 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

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

Chapter 2

Vector fields and vector bundles

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

2.1.1 Tangent vectors and vector fields

Definition 2.1. Let M be a smooth n-dimensional manifold. Two smooth curves $\gamma_1, \gamma_2 : (-\varepsilon, \varepsilon) \to M$ such that $\gamma_1(0) = \gamma_2(0) = q \in M$ are said to be *equivalent* if, in some coordinate chart, they have the same 1-st order Taylor polynomial. If $\gamma : (-\varepsilon, \varepsilon) \to M$ is a smooth curve its *tangent vector* at the point $q = \gamma(0)$, denoted by

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

is the equivalence class in the space of all smooth curves in M such that $\gamma(0) = q$.

It is easy to see, by the chain rule, that this is a well-defined object (i.e. it does not depend on the representative).

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

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

It is a standard fact that T_qM has a natural structure of n-dimensional vector space.

Definition 2.3. A *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 and only if its components $X^i(x)$ are smooth functions.

The value of a vector field X at a point q is denoted 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

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

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

$$\dot{\gamma}(t) = X(\gamma(t)), \qquad \forall\, t \in I \subset \mathbb{R}. \tag{2.3}$$

We also say that γ is an *integral curve* of the vector field X.

A standard theorem on ODE ensures that, for every point, there exists a unique integral curve of a vector field, defined on some interval and passing through this point.

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 a unique solution of (2.4), *denoted by*

$$\gamma(t; q_0), \qquad t \in (t_0, t_1), \quad t_0 < 0 < t_1 \tag{2.5}$$

defined on a sufficiently small interval (t_0, t_1). Moreover the solution $\gamma(t; q_0)$ smoothly depends on $(t, q_0) \in \mathbb{R} \times M$.

A vector field $X \in \mathrm{Vec}(M)$ is called *complete* if, for every $q_0 \in M$, the solution $\gamma(t; q_0)$ of the equation (2.2) can be extended for all $t \in \mathbb{R}$.

Remark 2.6. Standard results from ODE ensure completeness of the vector field $X \in \mathrm{Vec}(M)$ in the following cases

(i) M is a compact manifold,

(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 $\mathbb{R}^n$.

In what follows we will always assume that the the vector field X is complete. Indeed if the vector field X is not complete, but we are interested in the properties of its integral curves on a compact set $K \subset M$, it is sufficient to multiply X by a cut-off function $a : M \to \mathbb{R}$ that is identically 1 inside K, and that vanishes out of a suitably bigger compact K' such that $K \subset K'$. In this way we get a vector field that is complete and has the same integral curve of f inside K.

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 $\{\phi_t, t \in \mathbb{R}\}$ defined by

$$P_t : M \to M, \qquad q \mapsto \gamma(t; q). \tag{2.6}$$

In other words $P_t(q)$ is the shift for time t along the integral curve of X that starts from q. By Theorem 2.5 it follows that the map

$$(t, q) \mapsto \phi_t(q),$$

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

$$\begin{aligned} &\phi_0 = \mathrm{Id}, \\ &\phi_t \circ \phi_s = \phi_s \circ \phi_t = \phi_{t+s}, \qquad \forall\, t, s \in \mathbb{R}, \\ &(\phi_t)^{-1} = \phi_{-t}, \qquad\qquad\quad \forall\, t \in \mathbb{R}, \end{aligned} \tag{2.7}$$

Moreover it satisfies

$$\frac{\partial \phi_t(q)}{\partial t} = X(\phi_t(q)), \qquad \phi_0(q) = q, \quad \forall\, q \in M. \tag{2.8}$$

The family of maps ϕ_t defined by (2.6) 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 $t \in \mathbb{R}$.

Remark 2.7. When X is a linear vector field on $\mathbb{R}^n$, then $X(x) = Ax$ and it can be identified with the $n \times n$ matrix A. It is easy to show that the corresponding flow ϕ_t is precisely the matrix exponential, namely $\phi_t(x) = e^{tA}(x)$.

Following the exponential notation, the group properties (2.7) takes the form:

$$e^{0X} = \mathrm{Id}, \qquad e^{tX} \circ e^{sX} = e^{sX} \circ e^{tX} = e^{(t+s)X}, \qquad (e^{tX})^{-1} = e^{-tX}, \tag{2.9}$$

$$\frac{d}{dt} e^{tX} = X e^{tX}. \tag{2.10}$$

2.1.3 Nonautonomous vector fields

Definition 2.8. A *nonautonomous vector field* is a family of smooth vector fields $\{X_t\}_{t \in \mathbb{R}}$, where $X_t \in \mathrm{Vec}(M)$ for every $t \in \mathbb{R}$ and the map $t \mapsto X_t$ is measurable and locally bounded.[1]

Now we consider a *nonautonomous ODE*, i.e. an equation of the form

$$\dot{q} = X_t(q), \qquad q \in M, \tag{2.11}$$

where X_t is a nonautonomous vector field. If we consider local coordinates $x = (x_1, \ldots, x_n)$ in an open set O on the manifold M, the equation (2.11) is written in coordinates as

$$\dot{x} = f(t, x), \qquad x \in \mathbb{R}^n,$$

where the map $(t, x) \mapsto f(t, x)$ is defined on a subset of $\mathbb{R} \times \mathbb{R}^n$ and satisfies

[1] i.e. for every smooth function $a \in \mathcal{C}^\infty(M)$ the function $t \mapsto X_t a$ is L^∞.

(*i*) f is measurable and locally bounded with respect to t, for any fixed $x \in O$,

(*ii*) f is smooth in x for every fixed $t \in \mathbb{R}$,

(*iii*) f has locally bounded derivatives, i.e.,

$$\left|\frac{\partial f_i}{\partial x}(t,x)\right| \leq C_{I,K}, \qquad I \subset \mathbb{R},\ K \subset O \text{ compact}, \quad i = 1, \dots, n.$$

where we denote with $f = (f_1, \dots, f_n)$ the components of the vector function f.

The existence and uniqueness of the solution in the nonautonomous case is guaranteed by the following theorem (see [9]).

Theorem 2.9 (Carathéodory theorem)**.** *Assume that $f : \mathbb{R} \times \mathbb{R}^n \to \mathbb{R}^n$ satisfies (i)-(iii). Then the Cauchy problem*

$$\dot{x}(t) = f(t, x(t)), \qquad x(t_0) = x_0, \tag{2.12}$$

has locally a unique solution $x(t; t_0, x_0)$ such that (2.12) *is satisfied for almost every t and $x(t_0; t_0, x_0) = x_0$. Moreover the map $(t, x_0) \mapsto x(t; t_0, x_0)$ is Lipschitz with respect to t and smooth with respect to x_0.*

Let us assume now that the equation (2.9) is *complete*, i.e. for all $t_0 \in \mathbb{R}$ and $x_0 \in \mathbb{R}^n$ the solution $x(t; t_0, x_0)$ is defined for all $t \in \mathbb{R}$. Let us denote by $P_{t_0,t}(x_0) = x(t; t_0, x_0)$. The family of maps $P_{t_0,t}$ is the *nonautonomous flow* generated by X_t. It satisfies

$$\frac{\partial}{\partial t}\frac{\partial P_{t_0,t}}{\partial x}(x) = \frac{\partial f}{\partial x}(t, P_{t_0,t}(x_0))P_{t_0,t}(x)$$

Moreover the following algebraic identities are satisfied

$$\begin{aligned} &P_{t,t} = \mathrm{Id}, \\ &P_{t_2,t_3} \circ P_{t_1,t_2} = P_{t_1,t_3}, \qquad \forall\, t_1, t_2, t_3 \in \mathbb{R}, \\ &(P_{t_1,t_2})^{-1} = P_{t_2,t_1}, \qquad\qquad \forall\, t_1, t_2 \in \mathbb{R}, \end{aligned} \tag{2.13}$$

Conversely, to every family of smooth diffeomorphism $P_{t,s} : M \to M$ satisfying the relations (2.13) one can define its *infinitesimal generator* X_t as follows:

$$X_t(q) = \left.\frac{d}{ds}\right|_{s=0} P_{t,t+s}(q), \qquad \forall\, q \in M. \tag{2.14}$$

The following lemma characterizes the flows whose generator is autonomous.

Lemma 2.10. *Let $\{P_{t,s}\}_{t,s\in\mathbb{R}}$ be a family of smooth diffeomorphisms satisfying* (2.13)*. Its infinitesimal generator is an autonomous vector field if and only if*

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

2.1.4 Vector fields as operators on functions

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

$$X : \mathcal{C}^\infty(M) \to \mathcal{C}^\infty(M), \qquad a \mapsto Xa, \qquad a \in \mathcal{C}^\infty(M), \tag{2.15}$$

where

$$(Xa)(q) = \frac{d}{dt}\bigg|_{t=0} a(e^{tX}(q)), \qquad q \in M. \tag{2.16}$$

In other words it computes the derivative of the function a restricted on integral curves of the vector field X.

Remark 2.11. Let us denote $a_t := a \circ e^{tX}$. Clearly the map $t \mapsto a_t$ is smooth and from (2.16) it immediately follows that Xa represents the first order term in the expansion of a_t:

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

Exercise 2.12. Let $a \in \mathcal{C}^\infty(M)$ and $X \in \mathrm{Vec}(M)$, and denote $a_t = a \circ e^{tX}$. Prove the following formulas

$$\frac{d}{dt} a_t = X a_t, \tag{2.17}$$

$$a_t = a + t\,Xa + \frac{t^2}{2!}X^2 a + \frac{t^3}{3!}X^3 a + \ldots + \frac{t^k}{k!}X^k a + O(t^{k+1}). \tag{2.18}$$

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

$$X(ab) = (Xa)b + a(Xb), \qquad \forall\, a, b \in \mathcal{C}^\infty(M), \tag{2.19}$$

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

Remark 2.13. Notice that, in coordinates, if $a \in \mathcal{C}^\infty(M)$ 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 characherized by its action on functions.

Exercise 2.14. Let $f_1, \ldots, f_k \in \mathcal{C}^\infty(M)$ and assume that $N = \{f_1 = \ldots = f_k = 0\} \subset M$ where $df_1 \wedge \ldots \wedge df_k \neq 0$ on N. Show that $X \in \mathrm{Vec}(M)$ is tangent to the smooth submanifold N if and only if $Xf_i = 0$ for every $i = 1, \ldots, k$.

2.2 Differential of a map

A smooth map between manifolds induces a map between their tangent spaces, simply by transforming the smooth curves.

Definition 2.15. 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:

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

It is easily checked that this definition depends only on the equivalence class of γ.

Remark 2.16. Applying the definition, one immediately verifies that, if $\varphi : M \to N$, $\psi : N \to Q$ are two smooth maps between manifolds, then the differential of the composition $\psi \circ \varphi : M \to Q$ satisfies $(\psi \circ \varphi)_* = \psi_* \circ \varphi_*$.

Notation. The differential $\varphi_{*,q}$ of a smooth map $\varphi : M \to N$, sometimes called its *pushforward*, is also denoted by the following symbols

$$D_q\varphi, \qquad d_q\varphi, \tag{2.21}$$

and we will prefer the first one for smooth maps between manifolds while we use the second for smooth functions, i.e. when $N = \mathbb{R}$. In order to simplify the notation, we will sometimes omit the point q in writing $\varphi_{*,q}(v)$ (writing simply $\varphi_* v$) when there is no confusion at which point the tangent vector v is attached.

Sometimes it will also be useful to distinguish between intrinsic notation explained above and its coordinate representation, when we identify the linear map with the Jacobian matrix. In this case we will replace the notation (2.21) with

$$\frac{d\varphi}{dq}.$$

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 2.17. 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

$$(\varphi_* X)(\varphi(q)) := \varphi_*(X(q)), \qquad \forall\, q \in M. \tag{2.22}$$

If $P \in \mathrm{Diff}(M)$ is a diffeomorphism of M, we can rewrite the previous identity as

$$(P_* X)(q) = P_*(X(P^{-1}(q))), \qquad \forall\, q \in M. \tag{2.23}$$

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

Remark 2.18. From this definition it follows the useful formula for $X, Y \in \mathrm{Vec}(M)$

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

The following lemma shows that $P_* X$ is the vector field whose integral curves are the image under P of integral curves of X. Moreover it shows how the pushforward of a vector field acts on functions:

Lemma 2.19. *Let $P \in \mathrm{Diff}(M)$, $X \in \mathrm{Vec}(M)$ and $a \in \mathcal{C}^\infty(M)$ then*

$$e^{tP_* X} = P \circ e^{tX} \circ P^{-1}, \tag{2.24}$$

$$(P_* X)a = (X(a \circ P)) \circ P^{-1}. \tag{2.25}$$

Proof. From the formula

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

it follows that $t \mapsto P \circ e^{tX} \circ P^{-1}(q)$ is an integral curve of P_*X, from which (2.24) follows.

To prove (2.25) let us compute

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

Using (2.24) this is equal to

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

□

Remark 2.20. From this lemma it follows the following formula: for every $X, Y \in \mathrm{Vec}(M)$

$$(e_*^{tX} Y)a = Y(a \circ e^{tX}) \circ e^{-tX}. \tag{2.26}$$

2.3 Lie brackets

Now we introduce a fundamental notion of all our theory, 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[2]

$$[X, Y] := \left.\frac{\partial}{\partial t}\right|_{t=0} e_*^{-tX} Y. \tag{2.27}$$

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

$$[X, Y]\big|_q = \left.\frac{\partial}{\partial t}\right|_{t=0} e_*^{-tX} Y\big|_q = \left.\frac{\partial}{\partial t}\right|_{t=0} e_*^{-tX}(Y\big|_{e^{tX}(q)}) = \left.\frac{\partial}{\partial s \partial t}\right|_{t=s=0} e^{-tX} \circ e^{sY} \circ e^{tX}(q). \tag{2.28}$$

We recover its algebraic properties in the following

Proposition 2.23. *As a derivation on functions we have*

$$[X, Y] = XY - YX. \tag{2.29}$$

[2]Notice the conventional minus sign.

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 (2.26) we have

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

By Remark 2.11 we have $a \circ e^{-tX} = a_{-t} = a - t\,Xa + O(t^2)$, hence

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

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

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

Hence the first order term is $(XY - YX)a$. □

From this proposition it easily follows also the coordinate expression of the Lie bracket. Indeed if

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

we have

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

Proposition 2.23 shows that $\mathrm{Vec}(M)$, being an associative algebra with commutator as multiplication, is a Lie algebra with the Lie bracket.

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

Proposition 2.24. *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 $a \in \mathcal{C}^\infty(M)$, preliminarly we see, using (2.25), 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 (2.29)

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

□

To end this section, we want to show that the Lie bracket of two vector fields is zero, that means that they commute as operators, if and only if the same holds for their flows.

Proposition 2.25. *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. $[X, Y] = 0 \quad \Longrightarrow \quad e_*^{-tX} Y = Y$.

Proof of the Claim. 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

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

and the Claim is proved.

$(i) \Rightarrow (ii)$. Let us show that $P_s := e^{-tX} \circ e^{sY} \circ e^{tX}$ is the flow generated by Y. Indeed we have

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

where in the last equality we used the Claim. 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).

$(ii) \Rightarrow (i)$. For every function $a \in \mathcal{C}^\infty$ we have

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

Then (i) follows from (2.29). $\square$

Exercise 2.26. 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 tangent vector to the curve $\gamma(\sqrt{t})$ is exactly $[X, Y](q)$.

Exercise 2.27. Let $X, Y \in \mathrm{Vec}(M)$. Using the semigroup property of the flow, prove the following expansion

$$e_*^{-tX} Y = Y + t[X, Y] + \frac{t^2}{2}[X, [X, Y]] + \frac{t^3}{6}[X, [X, [X, Y]]] + \dots \tag{2.30}$$

Exercise 2.28. Let $X, Y \in \mathrm{Vec}(M)$ and $a \in \mathcal{C}^\infty(M)$. Prove the following Leibnitz rule for the Lie bracket:

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

Exercise 2.29. Let $X, Y, Z \in \mathrm{Vec}(M)$. Prove that the Lie bracket satisfies the *Jacobi identity*:

$$[X, [Y, Z]] + [Y, [Z, X]] + [Z, [X, Y]] = 0. \tag{2.31}$$

Hint: Differentiate the identity $e^{tX}_*[Y, Z] = [e^{tX}_* Y, e^{tX}_* Z]$.

2.4 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 2.30. 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, a smooth map yields a linear map between tangent spaces. Dualizing this map, we get a linear map on cotangent spaces going in the opposite direction.

Definition 2.31. 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 2.32. Let $a : M \to \mathbb{R}$ be a smooth function and $q \in M$. The differential d_qa of the function a at the point $q \in M$ is an element of T_q^*M since we have a well defined linear action

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

where $\gamma(t)$ is any smooth curve such that $\gamma(0) = q$ and $\dot\gamma(0) = v$.

Definition 2.33. A *differential 1-form* on a smooth manifold M is a smooth map

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

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

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

$$\langle \omega, X\rangle : q \mapsto \langle \omega(q), X(q)\rangle. \tag{2.32}$$

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

Definition 2.34. 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 $a \in \mathcal{C}^\infty(M)$, then

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

which is constant on fibers.

2.5 Vector bundles

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

Definition 2.35. 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-fiber diffeomorphism (also called *local trivialization*) $\psi : \pi^{-1}(O_q) \to O_q \times \mathbb{R}^k$ such that the following diagram commutes

$$\begin{tikzcd} \pi^{-1}(O_q) \arrow[r, "\psi"] \arrow[dr, "\pi"'] & O_q \times \mathbb{R}^k \arrow[d, "\pi_1"] \\ & O_q \end{tikzcd} \tag{2.33}$$

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 2.36. The existence of local trivialization maps ψ says that E, as 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. a local trivialization with $O_q = M$, then $E \simeq M \times \mathbb{R}^k$ and we say that E is *trivializable*.

Example 2.37. 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 where

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

define a local coordinate system. The differentials of the coordinate functions

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

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

$$\left.\frac{\partial}{\partial x_i}\right|_q \in T_qM, \qquad i = 1, \dots, n, \qquad q \in O, \tag{2.34}$$

$$\left\langle dx_i, \frac{\partial}{\partial x_j} \right\rangle = \delta_{ij}, \qquad i, j = 1, \dots, n. \tag{2.35}$$

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

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

and the maps

$$\psi_v : v \mapsto (x_1, \dots, x_n, v_1, \dots, v_n), \qquad \psi_\lambda : \lambda \mapsto (x_1, \dots, x_n, p_1, \dots, p_n), \tag{2.36}$$

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

Definition 2.38. 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

$$\begin{tikzcd} E \arrow[r, "f"] \arrow[dr, "\pi"'] & E' \arrow[d, "\pi'"] \\ & M \end{tikzcd} \tag{2.37}$$

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

Definition 2.39. Let $\pi : E \to M$ be a smooth vector bundle over M. A *section* of E is a smooth map[3] $\sigma : A \subset M \to E$ satisfying $\pi \circ \sigma = \mathrm{Id}_A$. 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*.

[3] as a map between manifolds.

Example 2.40. 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 2.41. Notice that vector fields and differential forms are, by definition, sections of the vector bundles TM and T^*M respectively.

Definition 2.42. 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* 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 2.43. (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

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

Remark 2.44. Let M be a smooth manifold and $a : M \to \mathbb{R}$ a smooth function. Assume that $c \in \mathbb{R}$ is a regular value of a[4], then

(i) $N_c = a^{-1}(c) = \{q \in M \,|\, a(q) = c\} \subset M$ is a smooth submanifold,

(ii) if $q \in N_c$ then $T_qN_c = \ker d_qa = \{v \in T_qM \mid \langle d_qa, v\rangle = 0\}$.

[4]that means $d_qa \neq 0$ for every $q \in a^{-1}(c)$.

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*[1] if

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

Definition 3.2. (sub-Riemannian manifold) 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 g_q, smooth with respect to q. For $u \in U_q$ we denote the norm of u as $|u| = \sqrt{(u|u)_q}$.

(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)

$$\begin{tikzcd} \mathbf{U} \arrow[r, "f"] \arrow[dr, "\pi_{\mathbf{U}}"'] & TM \arrow[d, "\pi"] \\ & M \end{tikzcd} \tag{3.2}$$

and f is *linear* on fibers.

(iii) The *set of horizontal vector fields* $\mathcal{D} := \{f(\sigma), \sigma \text{ smooth section of } \mathbf{U}\}$, is a bracket-generating family of vector fields.

[1] or that satisfies the *Hörmander condition*

When the vector bundle $\mathbf{U}$ admits a global trivialization we say that $(M, \mathbf{U}, f)$ 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

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

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.

The set of horizontal vector fields $\mathcal{D} \subset \mathrm{Vec}(M)$ has the structure of a finitely generated $\mathcal{C}^\infty(M)$-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

$$k(q) \leq m, \qquad \text{where } m = \operatorname{rank} \mathbf{U}, \tag{3.3}$$

$$k(q) \leq n, \qquad \text{where } n = \dim M. \tag{3.4}$$

In what follows we denote points in $\mathbf{U}$ as pairs (q, u), where $q \in M$ is an element of the base and $u \in U_q$ is an element of the fiber. Following this notation we can write the value of f at this point as

$$f(q, u) \qquad \text{or} \qquad f_u(q).$$

We use the second one when we want to emphasize that, for each $q \in M$, $f_u(q)$ is a vector in T_qM.

Definition 3.4. (Admissible Curves) 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 essentially bounded function

$$u : t \in [0, T] \mapsto u(t) \in U_{\gamma(t)}, \tag{3.5}$$

called the *control function*, such that

$$\dot\gamma(t) = f(\gamma(t), u(t)), \qquad \text{for a.e. } t \in [0, T]. \tag{3.6}$$

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

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

$$\dot\gamma(t) = \sum_{i=1}^m u_i(t) f_i(\gamma(t)), \qquad \text{for a.e. } t \in [0, T]. \tag{3.7}$$

Thanks to this local characterization and Theorem 2.9, for each initial condition $q \in M$ and $u \in L^\infty([0, T], \mathbb{R}^m)$ there exists an admissible curve γ, defined on a sufficiently small interval, such that u is the control associated with γ and $\gamma(0) = q$.

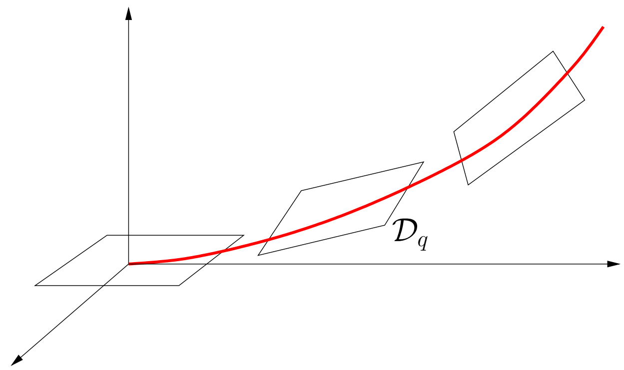


Figure 3.1: An horizontal curve

Remark 3.6. Notice that, for a curve to be admissible, it is not sufficient that it satisfies $\dot\gamma(t) \in \mathcal{D}_{\gamma(t)}$ 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

$$f(x,y,u_1,u_2) = (x,y,u_1,u_2x), \qquad f'(x,y,u_1,u_2) = (x,y,u_1,u_2x^2). \tag{3.8}$$

and let $\mathcal{D}$ and $\mathcal{D}'$ the corresponding moduli of horizontal vector fields. On one hand, 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 $t \in [-1,1]$.

On the other hand, γ is admissible for f, since its corresponding control is $(u_1,u_2) = (1,2)$ for a.e. $t \in [-1,1]$, but 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, in the case of constant rank distribution, we have that $\mathcal{D}_q = \mathcal{D}'_q$ 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 a 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

$$\|v\| := \min\{|u|,\ u \in U_q \ \text{ s.t. } \ v = f(q,u)\}. \tag{3.9}$$

Notice that since f is linear with respect to u, the minimum in (3.9) is always attained at a unique point. Indeed the condition $f(q,\cdot) = v$ defines an affine subspace of U_q (which is nonempty since $v \in \mathcal{D}_q$) and the minimum (3.9) is uniquely attained at the orthogonal projection of the origin onto this subspace (see Figure 3.2).

Exercise 3.8. Show that $\|\cdot\|$ is a norm in $\mathcal{D}_q$. Moreover prove that it satisfies the parallelogram law, i.e. it induce a scalar product on $\mathcal{D}_q$ by the polarization identity. We will denote this scalar product by $\langle\cdot|\cdot\rangle$.

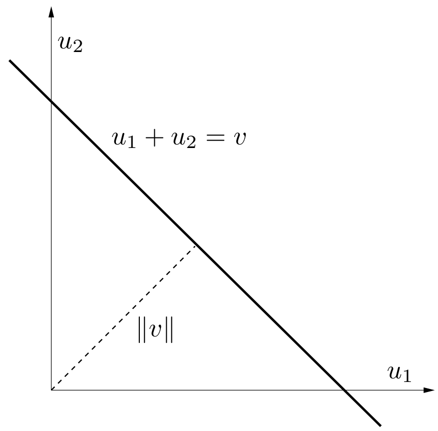


Figure 3.2: The norm of a vector v for $f(x, u_1, u_2) = u_1 + u_2$

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 where $\dot{\gamma}(t)$ exists. Hence it is well defined the unique control $t \mapsto u^*(t)$ associated with γ and realizing the minimum.

Definition 3.10. Given an admissible curve $\gamma : [0, T] \to M$, we say that the control $t \mapsto u^*(t)$ is the *minimal control* associated with γ.

We stress that $u^*(t)$ is pointwise defined for a.e. $t \in [0, T]$. In particular, if the admissible curve $\gamma : [0, T] \to M$ is C^1, the minimal control is defined everywhere on $[0, T]$.

Remark 3.11. Notice that, even if an admissible curve is smooth, its minimal control could be not continuous. Consider, as in Remark 3.6 the free sub-Riemannian structure on $\mathbb{R}^2$

$$f(x, y, u_1, u_2) = (x, y, u_1, u_2 x), \tag{3.10}$$

and let $\gamma : [-1, 1] \to \mathbb{R}^2$, $\gamma(t) = (t, t^2)$. 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.

The proof of the following crucial Lemma is postponed to the Section 3.A.

Lemma 3.12. *Let $\gamma : [0, T] \to M$ be an admissible curve. Then its minimal control $u^*(\cdot)$ is measurable and essentially bounded.*

Hence we are allowed to introduce the following definition.

Definition 3.13. Let $\gamma : [0, T] \to M$ be an admissible curve, we define its *sub-Riemannian length*

$$\ell(\gamma) := \int_0^T \|\dot{\gamma}(t)\| dt. \tag{3.11}$$

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

Notice that (3.11) says that the length of an admissible curve is the integral of the norm of its minimal control.

$$\ell(\gamma) = \int_0^T |u^*(t)| dt. \tag{3.12}$$

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 remark that γ_φ is a composition of Lipschitz functions, hence Lipschitz. Moreover it is admissible since, using 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

$$\ell(\gamma_\varphi) = \int_0^{T'} \|\dot{\gamma}_\varphi(s)\| ds = \int_0^{T'} |u^*(\varphi(s))||\dot{\varphi}(s)| ds = \int_0^T |u^*(t)| dt = \int_0^T \|\dot{\gamma}(t)\| dt = \ell(\gamma). \tag{3.13}$$

□

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

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

where K is a compact set such that $\psi([0,T]) \subset K$ and $C = \max_{x\in K}\left(\sum_{i=1}^m |f_i(x)|^2\right)^{1/2}$. Then if $s_1 = \varphi(t_1)$ and $s_0 = \varphi(t_0)$ one has

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

hence γ is Lipschitz. It follows that $\dot\gamma(s)$ 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 $\{s \in \mathbb{R} \mid s = \varphi(t), \dot\varphi(t) \text{ exists, and } \dot\varphi(t) = 0\}$ is zero.

By the previuos discussion, in what follows, it will be often convenient to assume that admissible curves are length-parametrized (or parametrized such that $\|\dot\gamma(t)\| = \text{const}$).

3.1.2 Equivalence of sub-Riemannian structures

In this section we discuss 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

$$\begin{tikzcd} & \mathbf{U} \arrow[dr, "f"] & \\ \mathbf{V} \arrow[ur, "p"] \arrow[dr, "p'"'] & & TM \\ & \mathbf{U}' \arrow[ur, "f'"'] & \end{tikzcd} \tag{3.14}$$

(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. Notice that if $(\mathbf{U}, f), (\mathbf{U}', f')$ 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 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 $\mathbf{U}'$ (cf. Definition 3.2) are equal.

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[2] of Euclidean bundles $\psi : \mathbf{U} \to \mathbf{U}'$ such that the following diagram is commutative

$$\begin{tikzcd} \mathbf{U} \arrow[r, "f"] \arrow[d, "\psi"'] & TM \arrow[d, "\phi_*"] \\ \mathbf{U}' \arrow[r, "f'"'] & TM' \end{tikzcd} \tag{3.15}$$

3.2 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] isomorphism of bundles in the broad sense, it is fiberwise but is not obliged to send fiber in the same fiber.

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

   Then, if we denote with $\{e_1, \ldots, e_m\}$ an orthonormal basis of $\mathbb{R}^m$, then we can define globally m smooth vector fields on M by $f_i(q) := f(q, e_i)$ for $i = 1, \ldots, m$. Then we have

   $$f(q,u) = f\left(q, \sum_{i=1}^{m} u_i e_i\right) = \sum_{i=1}^{m} u_i f_i(q), \qquad q \in M. \tag{3.16}$$

   In this case, the problem of finding an admissible curve joining two fixed points $q_0, q_1 \in M$ and with minimal length is rewritten as the optimal control problem

   $$\begin{cases} \dot\gamma(t) = \displaystyle\sum_{i=1}^{m} u_i(t) f_i(\gamma(t)) \\ \displaystyle\int_0^T |u(t)|dt \to \min \\ \gamma(0) = q_0, \quad \gamma(T) = q_1 \end{cases} \tag{3.17}$$

   For a free sub-Riemannian structure, the set of vector fields $f_1, \ldots, f_m$ build as above is called a *generating frame*. Notice that, in general, a generating frame 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 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

   $$f : \mathbf{U} \to TM, \qquad f(q,u) = \pi_q^\perp(u) \in T_qM. \tag{3.18}$$

   where $\pi_q^\perp : \mathbb{R}^3 \to T_qM \subset \mathbb{R}^3$ 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 frame for this structure.

   Exercise 3.24. Show that $(\mathbf{U}, f)$ defined in (3.18) is equivalent to the Riemannian structure on M induced by the embedding in $\mathbb{R}^3$.

3.2.1 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. If we denote with $T_M E$ the vector bundle $i^*(TE)$, i.e. the restriction of TE to the section M_0, we have the isomorphism (as vector bundles on M)

$$T_M E \simeq E \oplus TM. \tag{3.19}$$

Eq. (3.19) is a consequence of the fact that the tangent to every fibre E_q, being a vector space, is canonically isomorphic to its tangent space $T_q E_q$ so that

$$T_q E = T_q E_q \oplus T_q M \simeq E_q \oplus T_q M, \qquad \forall\, q \in M.$$

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

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

for $N = 2(n+m)$, and Ψ_* is injective as bundle map, i.e. $T_M E$ is a sub-bundle of $T\mathbb{R}^N \simeq \mathbb{R}^N \times \mathbb{R}^N$. Thus we can choose as a complement E', the orthogonal bundle (on the base M) with respect to the Euclidean metric in $\mathbb{R}^N$, i.e.[3]

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

and considering $E_0 := T_M E \oplus E'$ we have that E_0 is trivial since its fibers are sum of orthogonal complements and by (3.19) we are done.

□

Corollary 3.26. *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.*

Proof. By Lemma 3.25 there exists a vector bundle $\mathbf{U}'$ such that the direct sum $\widetilde{\mathbf{U}} := \mathbf{U} \oplus \mathbf{U}'$ is trivial. Endow $\mathbf{U}'$ with any metric structure. Define a metric on $\widetilde{\mathbf{U}}$ in such a way that $\widetilde{g}(u+u', v+v') = g(u, v) + g'(u', v')$ on each fiber $\widetilde{U}_q = U_q \oplus U'_q$. Notice that U_q and U'_q are orthogonal.

Let us define the sub-Riemannian structure $(\widetilde{\mathbf{U}}, \widetilde{f})$ on M by

$$\widetilde{f} : \widetilde{\mathbf{U}} \to TM, \qquad \widetilde{f} := f \circ p_1,$$

[3] we simplify notation, writing $T_q E_q^\perp$ for $\Psi_*(T_q E_q)^\perp$

where $p_1 : \mathbf{U} \oplus \mathbf{U}' \to \mathbf{U}$ denotes the projection on the first factor. By construction, the diagram

$$\begin{tikzcd} & \widetilde{\mathbf{U}} \arrow[dr, "\widetilde{f}"] & \\ \mathbf{U} \oplus \mathbf{U}' \arrow[ur, "\mathrm{Id}"] \arrow[dr, "p_1"'] & & TM \\ & \mathbf{U} \arrow[ur, "f"'] & \end{tikzcd} \tag{3.20}$$

is commutative. Moreover condition (ii) of Definition 3.17 is satisfied since for every $\widetilde{u} = u + u'$, with $u \in U_q$ and $u' \in U'_q$, we have $|\widetilde{u}|^2 = |u|^2 + |u'|^2$, hence $|u| = \min\{|\widetilde{u}|, p_1(\widetilde{u}) = u\}$. $\square$

This fact allow us to write every admissible curve of a sub-Riemannian structure in the form

$$\dot{\gamma}(t) = \sum_{i=1}^{m} u_i(t) f_i(\gamma(t)), \tag{3.21}$$

where its length is given by

$$\ell(\gamma) = \int_0^T |u^*(t)| dt = \int_0^T \sqrt{\sum_{i=1}^{m} u_i^*(t)^2} dt.$$

Here $f_1, \ldots, f_m$ are vector fields on M that are globally defined, with m big enough. As in Example 4 of Section 3.2 the set of vector fields $f_1, \ldots, f_m$ is called a *generating frame*.

We stress that this is equivalent to say that the modulus of horizontal vector fields $\mathcal{D}$ is globally generated by $f_1, \ldots, f_m$.

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

3.2.2 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 one of a *Riemannian foliation*, that is obtained when 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.

3.3 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 3.2.1, in what follows we can assume that the sub-Riemannian structure is free, with generating frame $\mathcal{F} = \{f_1, \ldots, f_m\}$. Notice that, by definition, $\mathcal{F}$ is assumed to be bracket generating.

Definition 3.28. 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

$$d(q_0, q_1) = \inf\{\ell(\gamma),\ \gamma \text{ admissible},\ \gamma(0) = q_0,\ \gamma(T) = q_1\}, \tag{3.22}$$

One of the purpose of this section is to show that, thanks to the bracket generating condition, (3.22) is well-defined since, for every $q_0, q_1 \in M$, there exists an admissible curve that joins q_0 to q_1 and $d(q_0, q_1) < +\infty$.

Theorem 3.29 (Chow-Raschevskii)**.** *Let M be a n-dimensional 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 $B(q, r)$ denotes the 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 3.29. To prove Theorem 3.29 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 $B(q_0, \delta) \subset O_{q_0}$.

3.3.1 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 $\widetilde{\gamma} : [0, T] \to M$ defined by $\widetilde{\gamma}(t) = \gamma(T - t)$ is admissible and $\ell(\widetilde{\gamma}) = \ell(\gamma)$.

The triangular inequality follows from the fact that the concatenation of two admissible curves is still admissible. This proves (c).

We divide the rest of the proof of the Theorem in the following steps.

Step 1. 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).

Step 2. We prove that d is finite on $M \times M$. This proves (a).

Step 3. We prove (b) and (e).

To prove Step 1 we first need the following lemmas:

Lemma 3.30. *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. $X(q) \in T_qN, \forall\, q \in N, X \in \mathcal{F}$. 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 (2.28)) 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 3.31. *Let M be an n-dimensional sub-Riemannian manifold with generating frame $\mathcal{F} = \{f_1, \ldots, f_m\}$. Then, 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 3.32. 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 in this case, for each choice of the vector fields $f_{i_1}, \ldots, f_{i_n} \in \mathcal{F}$, the image of the differential of ψ at $s = 0$ is $\mathrm{span}_{q_0}\{f_{i_j}, j = 1, \ldots, n\} \subset \mathcal{D}_{q_0}$ and the differential of ψ is not 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 3.31.

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 3.30, 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 3.30, $\dim \operatorname{Lie}_{q_1}\mathcal{D} = 2$, 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 3.31 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 $q_0 = \psi(0)$ is not contained $\psi(\widehat{V})$, cf. Remark 3.32.

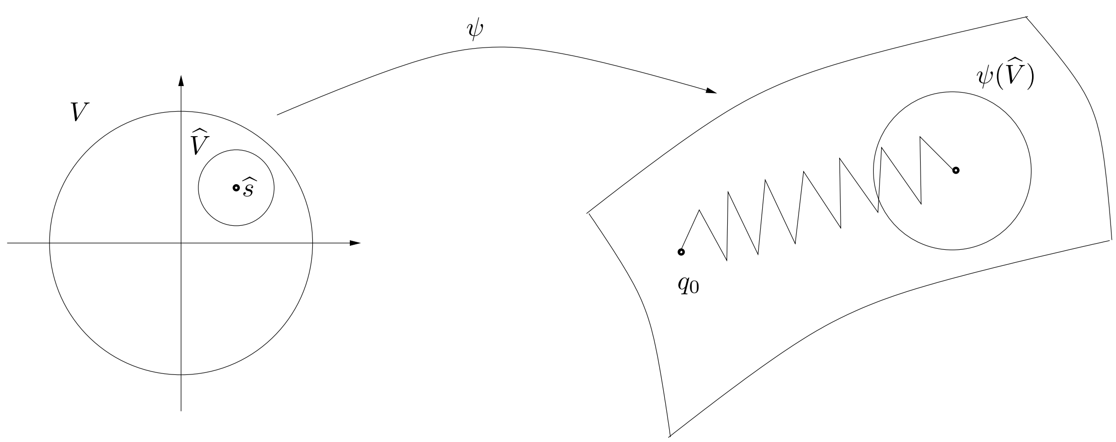


Figure 3.3: Proof of Lemma 3.31

To build a local diffeomorphism whose image contains q_0, we consider the map

$$\widehat{\psi} : \mathbb{R}^n \to M, \qquad \widehat{\psi}(s_1, \ldots, s_n) = e^{-\widehat{s}_1 f_{i_1}} \circ \cdots \circ e^{-\widehat{s}_n f_{i_n}} \circ \psi(s_1, \ldots, s_n),$$

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 $V = \{s \in \mathbb{R}^n, \sum_{i=1}^n |s_i| < \varepsilon\}$. 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 $s = (s_1, \dots, s_n)$ 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.

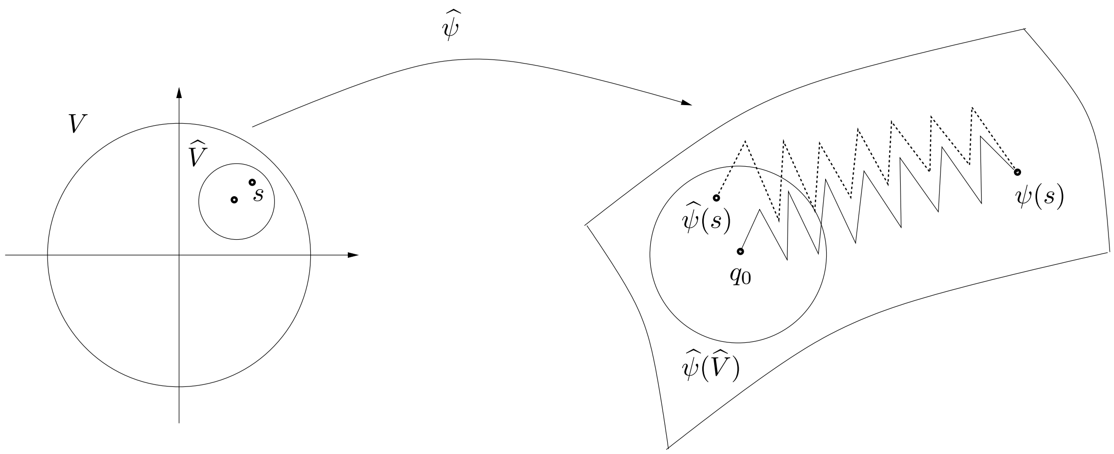


Figure 3.4: The map $\widehat{\psi}$

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:

$$d(q_0, q) \le \ell(\gamma) \le |s_1| + \ldots + |s_n| + |\widehat{s}_1| + \ldots + |\widehat{s}_n| < 2\varepsilon,$$

which ends the proof of Step 1.

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

$$q_1 \sim q_2 \qquad \text{if} \qquad d(q_1, q_2) < +\infty. \tag{3.23}$$

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. Since M is connected, it cannot be the union of open disjoint and nonempty subsets. It follows that there exists only one equivalence class.

Proof of Step 3. We start by proving the following lemma.

Lemma 3.33. *Fix $q_0 \in M$. For every compact set $K \subset M$ with $q_0 \in \operatorname{int} K$, there exists $\delta_K > 0$ such that every admissible curve $\gamma : [0,T] \to M$ satisfying $\gamma(0) = q_0$ and $\ell(\gamma) \le \delta_K$ 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

$$C_K := \max_{x \in K} \left(\sum_{i=1}^m |f_i(x)|^2 \right)^{1/2} \tag{3.24}$$

and fix $\delta_K > 0$ such that $\mathrm{dist}(q_0, \partial K) > C_K \delta_K$ (here dist is the Euclidean distance in coordinates).

Then 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

$$|\gamma(t^*) - \gamma(0)| \leq \int_0^{t^*} |\dot{\gamma}(t)| dt = \int_0^{t^*} \sum_{i=1}^{m} |u_i^*(t) f_i(\gamma(t))| \, dt \tag{3.25}$$

$$\leq \int_0^{t^*} \sqrt{\sum_{i=0}^{m} |f_i(\gamma(t))|^2} \sqrt{\sum_{i=0}^{m} u_i^*(t)^2} \, dt \tag{3.26}$$

$$\leq C_K \int_0^{t^*} \sqrt{\sum_{i=0}^{m} u_i^*(t)^2} \, dt \leq C_K \ell(\gamma) \tag{3.27}$$

$$\leq C_K \delta_K < \mathrm{dist}(q_0, \partial K). \tag{3.28}$$

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

Lemma 3.33 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 3.33, 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 a compact neighborhood K of q_0. Define C_K and δ_K as in Lemma 3.33, and set $\delta := \min\{\delta_K, \varepsilon / C_K\}$. Let us show that $|q - q_0| < \varepsilon$ whenever $d(q_0, q) < \delta$.

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 3.33, $\gamma_n([0, T]) \subset K$ for all n.

In particular we can repeat estimates (3.25)-(3.27) 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

$$|q - q_0| \leq C_K d(q_0, q) \leq C_K \delta < \varepsilon. \tag{3.29}$$

Corollary 3.34. *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. $\square$

3.4 Existence of minimizers

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

Definition 3.35. 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 3.36. The example $M = \mathbb{R}^2 \setminus \{0\}$ endowed with the Euclidean distance shows that in general there may be no minimizers between two points. However there may be several minimizers between two fixed points, as it happens for two antipodal points on the sphere S^2.

Before proving the existence of length minimizers we show a general property of the length functional.

Theorem 3.37. *Let γ_n be a sequence of admissible curves on M such that $\gamma_n \to \gamma$ uniformly. Then*

$$\ell(\gamma) \leq \liminf_{n\to\infty} \ell(\gamma_n). \tag{3.30}$$

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

Proof. Without loss of generality we assume that γ_n and γ are parametrized with constant speed on the interval $[0,1]$. Moreover, denote $L := \liminf \ell(\gamma_n)$ and choose a subsequence, which we still denote by the same symbol, such that $\ell(\gamma_n) \to L$. If $L = +\infty$ the inequality (3.30) is clearly true, thus assume $L < +\infty$.

Fix $\delta > 0$. By uniform convergence, it is not restrictive to assume that, for n large enough, $\ell(\gamma_n) \leq L + \delta$ and that the image of γ_n are all contained in a common compact set K. Since γ_n is parametrized by constant speed on $[0,1]$ we have that $\dot\gamma_n(t) \in V_{\gamma_n(t)}$ 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 $\ell(\gamma) \leq L + \delta$. Since δ is arbitrary, this implies $\ell(\gamma) \leq L$, that is (3.30).

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

$$\frac{1}{\varepsilon}(\gamma_n(t+\varepsilon) - \gamma_n(t)) = \frac{1}{\varepsilon}\int_t^{t+\varepsilon} f_{u_n(\tau)}(\gamma_n(\tau))d\tau \in \operatorname{conv}\{V_{\gamma_n(\tau)}, \tau \in [t, t+\varepsilon]\}. \tag{3.31}$$

Moreover, for n sufficiently large, we have for $\tau \in [t, t+\varepsilon]$

$$|\gamma_n(\tau) - \gamma(t)| \leq |\gamma_n(t) - \gamma_n(\tau)| + |\gamma_n(t) - \gamma(t)| \leq C'\varepsilon, \tag{3.32}$$

where C' is independent on n, ε. Indeed $|\gamma_n(t) - \gamma(t)| < \varepsilon$ (by uniform convergence) and an estimate similar to (3.27) gives for $\tau \in [t, t+\varepsilon]$

$$|\gamma_n(t) - \gamma_n(\tau)| \leq \int_t^\tau |\dot\gamma_n(s)|ds \leq C_K(L+\delta)\varepsilon. \tag{3.33}$$

where C_K is the constant (3.24) defined by the compact K. From the estimate (3.32) and the equivalence of the manifold and metric topology we have that, for all $\tau \in [t, t+\varepsilon]$ and n big enough, $\gamma_n(\tau) \in B_{\gamma(t)}(r(\varepsilon))$, where $r(\varepsilon) \to 0$ for $\varepsilon \to 0$. In particular

$$\operatorname{conv}\{V_{\gamma_n(\tau)}, \tau \in [t, t+\varepsilon]\} \subset \operatorname{conv}\{V_q, q \in B_{\gamma(t)}(r(\varepsilon))\}, \tag{3.34}$$

Plugging (3.34) in (3.31) and passing to the limit for $n \to \infty$ we get:

$$\frac{1}{\varepsilon}(\gamma(t+\varepsilon) - \gamma(t)) \in \operatorname{conv}\{V_q, q \in B_{\gamma(t)}(r(\varepsilon))\}. \tag{3.35}$$

Assume now that $t \in [0,1]$ is a differentiability point of γ. Then the limit for $\varepsilon \to 0$ in (3.35) gives $\dot\gamma(t) \in \operatorname{conv} V_{\gamma(t)} = V_{\gamma(t)}$. For every such t we can define the unique solution $u^*(t)$ to the problem $\dot\gamma(t) = f(\gamma(t), u^*(t))$ and $|u^*(t)| = \|\dot\gamma(t)\|$. Using the argument contained in Appendix 3.A it follows that $u^*(t)$ is measurable in t. Moreover it is bounded since, by construction, $|u^*(t)| \leq L + \delta$. Hence γ is admissible. Moreover $\ell(\gamma) \leq L + \delta$ since γ is parametrized on $[0,1]$.

□

Corollary 3.38 (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*

$$d(q_0, q_1) = \min\{\ell(\gamma), \gamma \text{ admissible}, \gamma(0) = q_0, \gamma(T) = q_1\}.$$

Proof. Fix $q_1 \in B_{q_0}(r)$ and consider a sequence of admissible trajectories $\gamma_n : [0,1] \to M$, parametrized with constant speed, joining q_0 and q and minimizing the lenght, i.e., $\ell(\gamma_n) \to d(q_0, q)$.

Since $d(q_0, q) < r$, we have $\ell(\gamma_n) \leq r$ for all n 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 (3.33) shows that for all n

$$|\gamma_n(t) - \gamma_n(\tau)| \leq \int_\tau^t |\dot{\gamma}_n(s)| ds \leq C_K r |t - \tau|, \qquad \forall\, t, \tau \in [0,1]. \tag{3.36}$$

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 we have uniform convergence $\gamma_n \to \gamma$. By Theorem 3.37 the curve γ is admissible and has length $\ell(\gamma) \leq \liminf \ell(\gamma_n) = d(q_0, q_1)$. ☐

Corollary 3.39. *Let $q_0 \in M$. Under the hypothesis of Corollary 3.38 there exists $\varepsilon > 0$ such that for all $r \leq \varepsilon$ and $q_1 \in B_{q_0}(r)$ there exists a minimizing curve joining q_0 and q_1.*

Proof. It is a direct consequence of Corollary 3.38 and Corollary 3.34. ☐

Remark 3.40. It is well known that a metric space is complete if and only if all closed balls are compact, see for instance [12]. In particular, if (M, d) is complete with respect to the sub-Riemannian distance, then for every $q_0, q_1 \in M$ there exists a length minimizer joining q_0 and q_1.

3.5 Pontryagin extremals

In this section we want to give necessary conditions to characterize the length minimizers. To begin with, we would like to motivate our Hamiltonian approach that we develop in the sequel.

In classical Riemannian geometry geodesics are local (in time) length-minimizers, appropriately parametrized. They satisfy a second order differential equation in M, which can be reduced to a first-order differential equation in TM. Hence the set of all geodesics can 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 geodesics. 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, geodesics starting from a point q_0 cover a full neighborhood of q_0.

The right approach is to parametrize the geodesics 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 $\{\gamma_{q_0}(\varepsilon),$ where γ_{q_0} is a geodesic starting from $q_0\}$ on the corresponding geodesic for $\varepsilon \to 0$.

Next theorem is the first version of Pontryagin maximum principle, whose proof is given in the next section.

Theorem 3.41 (Characterization of Pontryagin extremals)**.** *Let $\gamma : [0, T] \to M$ be an admissible curve which is a length-minimizer, parametrized by constant speed. Let $\widetilde{u}(\cdot)$ be the corresponding minimal control, i.e.,*

$$\dot{\gamma}(t) = \sum_{i=1}^{m} \widetilde{u}_i(t) f_i(\gamma(t)), \qquad \ell(\gamma) = \int_0^T |\widetilde{u}(t)| dt = d(\gamma(0), \gamma(T)), \qquad |\widetilde{u}(t)| = const. \ a.e.$$

Denote with $P_{0,t}$ the flow[4] *of the nonautonomous vector field $f_{\widetilde{u}(t)} = \sum_{i=1}^{k} \widetilde{u}_i(t) f_i$. Then there exists $\lambda_0 \in T^*_{\gamma(0)}M$ such that defining*

$$\lambda(t) := (P_{0,t}^{-1})^* \lambda_0, \qquad \lambda(t) \in T^*_{\gamma(t)}M, \tag{3.37}$$

we have that one of the following conditions is satisfied:

(N) $\widetilde{u}_i(t) \equiv \langle \lambda(t), f_i(\gamma(t)) \rangle, \quad \forall\, i = 1, \ldots, m,$

(A) $0 \equiv \langle \lambda(t), f_i(\gamma(t)) \rangle, \quad \forall\, i = 1, \ldots, m.$

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 $\widetilde{u}(t) \equiv 0$ a.e., i.e., γ is the trivial trajectory.

Definition 3.42. Let $\gamma : [0, T] \to M$ be an admissible curve with minimal control $\widetilde{u} \in L^\infty([0, T], \mathbb{R}^m)$. Fix $\lambda_0 \in T^*_{\gamma(0)}M \setminus \{0\}$, and define $\lambda(t)$ by (3.37).

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

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

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

Remark 3.44. If the sub-Riemannian structure is not surjective at q_0, i.e., $\text{span}_{q_0}\{f_1, \ldots, f_m\} \neq T_{q_0}M$, then the trivial trajectory, corresponding to $\widetilde{u}(t) \equiv 0$, 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).

Exercise 3.45. Prove that condition (N) of Theorem 3.37 implies that the minimal control $\widetilde{u}(t)$ is smooth. In particular normal extremals are smooth.

At this level it seems not obvious how to use Theorem 3.41 to find the explicit expression of extremals for a given problem. In the next chapter we provide another formulation of Theorem 3.41 which gives Pontryagin extremals as solutions of a Hamiltonian system.

The rest of this section is devoted to the proof of Theorem 3.41.

[4] defined for $t \in [0, T]$ and in a neighborhood of $\gamma(0)$

3.5.1 The action functional

Let $\gamma : [0,T] \to M$ be an admissible curve. We define the *action* functional J as follows

$$J(\gamma) = \frac{1}{2}\int_0^T \|\dot\gamma(t)\|^2 dt.$$

Remark 3.46. Notice that, 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 3.47. *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

$$\left(\int_0^T f(t)g(t)dt\right)^2 \le \int_0^T f(t)^2 dt \int_0^T g(t)^2 dt, \tag{3.38}$$

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

$$\ell(\gamma)^2 \le 2J(\gamma)T. \tag{3.39}$$

Moreover in (3.38) equality holds if and only if f is proportional to g, i.e. $\|\dot\gamma(t)\| = \text{const.}$ in (3.39). 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. □

3.5.2 Proof of Theorem 3.41

By Lemma 3.47 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

$$\bar J(u(\cdot)) := \frac{1}{2}\int_0^T |u(t)|^2 dt, \tag{3.40}$$

on the space of controls $u(\cdot) \in L^\infty([0,T], \mathbb{R}^m)$, the minimal control $\widetilde u(\cdot)$ of γ is a minimizer for the action functional $\bar J$

$$\bar J(\widetilde u(\cdot)) \le \bar J(u(\cdot)), \quad \forall\, u \in L^\infty([0,T], \mathbb{R}^m),$$

where trajectories corresponding to $u(\cdot)$ join $q_0, q_1 \in M$. In the following we denote the functional $\bar{J}$ by J.

Consider now a variation $u(\cdot) = \widetilde{u}(\cdot) + v(\cdot)$ of the control $\widetilde{u}(\cdot)$, and its associated trajectory $q(t)$, solution of the equation

$$\dot{q}(t) = f_{u(t)}(q(t)), \qquad q(0) = q_0, \tag{3.41}$$

Recall that $P_{0,t}$ denotes the local flow associated with the optimal control $\widetilde{u}(\cdot)$ 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

$$q(t) := P_{0,t}(x(t)). \tag{3.42}$$

In other words $x(t) = P_{0,t}^{-1}(q(t))$ is obtained by applying the inverse of the flow of $\widetilde{u}(\cdot)$ to the solution associated with the new control $u(\cdot)$ (see Figure 3.5). Notice that if $v(\cdot) = 0$, then $x(t) \equiv q_0$.

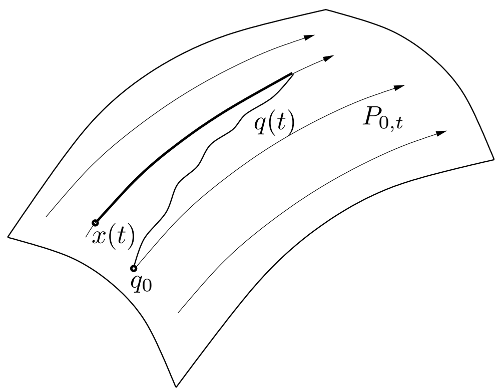


Figure 3.5: The trajectories $q(t)$, associated with $u(\cdot) = \widetilde{u}(\cdot) + v(\cdot)$, and the corresponding $x(t)$.

The next step is to write an ODE satisfied by $x(t)$. Differentiating (3.42) we get

$$\dot{q}(t) = f_{\widetilde{u}(t)}(q(t)) + (P_{0,t})_*(\dot{x}(t)) \tag{3.43}$$

$$= f_{\widetilde{u}(t)}(P_{0,t}(x(t)) + (P_{0,t})_*(\dot{x}(t)) \tag{3.44}$$

and using that $\dot{q}(t) = f_{u(t)}(q(t)) = f_{u(t)}(P_{0,t}(x(t))$ we can invert (3.44) with respect to $\dot{x}(t)$ and rewrite it as follows

$$\begin{aligned}
\dot{x}(t) &= (P_{0,t}^{-1})_* \left[(f_{u(t)} - f_{\widetilde{u}(t)})(P_{0,t}(x(t))) \right] \\
&= \left[(P_{0,t}^{-1})_*(f_{u(t)} - f_{\widetilde{u}(t)}) \right] (x(t)) \\
&= \left[(P_{0,t}^{-1})_*(f_{u(t)-\widetilde{u}(t)}) \right] (x(t)) \\
&= \left[(P_{0,t}^{-1})_* f_{v(t)} \right] (x(t))
\end{aligned} \tag{3.45}$$

If we define the nonautonomous vector field $g_{v(t)}^t = (P_{0,t}^{-1})_* f_{v(t)}$ we finally obtain by (3.45) the following Cauchy problem for $x(t)$

$$\dot{x}(t) = g_{v(t)}^t(x(t)), \qquad x(0) = q_0. \tag{3.46}$$

Notice that the vector field g_v^t is linear with respect to v, since f_u is linear with respect to u.

Now we fix the control $v(t)$ and consider the map

$$s \in \mathbb{R} \mapsto \begin{pmatrix} J(\widetilde{u}+sv) \\ x(T;\widetilde{u}+sv) \end{pmatrix} \in \mathbb{R} \times M$$

where $x(T;\widetilde{u}+sv)$ denote the solution at time T of (3.46), starting from q_0, corresponding to control $\widetilde{u}(\cdot)+sv(\cdot)$, and $J(\widetilde{u}+sv)$ is the associated cost.

Lemma 3.48. *There exists $\bar{\lambda} \in (\mathbb{R} \oplus T_{q_0}M)^*$, with $\bar{\lambda} \neq 0$, such that for all $v \in L^\infty([0,T],\mathbb{R}^m)$*

$$\bar{\lambda} \perp \left(\frac{\partial J(\widetilde{u}+sv)}{\partial s}\Big|_{s=0}, \frac{\partial x(T;\widetilde{u}+sv)}{\partial s}\Big|_{s=0} \right). \tag{3.47}$$

Proof of Lemma 3.48 . We argue by contradiction: if (3.47) is not true then there exist $v_0, \ldots, v_n \in L^\infty([0,T],\mathbb{R}^m)$ such that the vectors

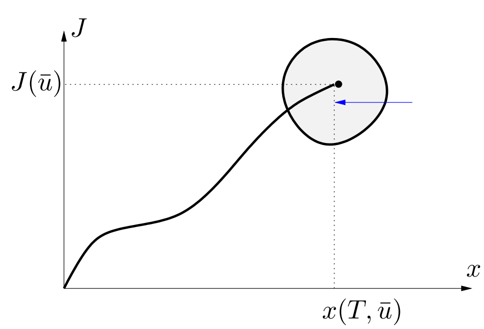
$$\begin{pmatrix} \dfrac{\partial J(\widetilde{u}+sv_0)}{\partial s}\Big|_{s=0} \\ \dfrac{\partial x(T;\widetilde{u}+sv_0)}{\partial s}\Big|_{s=0} \end{pmatrix}, \ \ldots \ , \begin{pmatrix} \dfrac{\partial J(\widetilde{u}+sv_n)}{\partial s}\Big|_{s=0} \\ \dfrac{\partial x(T;\widetilde{u}+sv_n)}{\partial s}\Big|_{s=0} \end{pmatrix} \tag{3.48}$$

are linearly independent. Let us now consider the map

$$\Phi : (s_0, \ldots, s_n) \in \mathbb{R}^{n+1} \ \mapsto \ \begin{pmatrix} J(\widetilde{u}+\sum_{i=0}^n s_i v_i) \\ x(T;\widetilde{u}+\sum_{i=0}^n s_i v_i) \end{pmatrix} \in \mathbb{R} \times M. \tag{3.49}$$

By differentiability properties of solution of smooth ODEs with respect to parameters, the map (3.49) is smooth. Moreover, since the vectors (3.48) 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 0 in $\mathbb{R}^{n+1}$ in a neighborhood of $(J(\widetilde{u}), q_0)$ 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 3.5.2)

$$x(T;\widetilde{u}+v) = q_0, \qquad J(\widetilde{u}+v) < J(\widetilde{u}).$$



In other words the curve $t \mapsto q(t;\widetilde{u}+v)$ join $q(0,\widetilde{u}+v) = q_0$ to

$$q(T;\widetilde{u}+v) = P_{0,T}(x(T;\widetilde{u}+v)) = P_{0,T}(q_0) = q_1,$$

with a cost smaller that the cost of $\gamma(t) = q(t,\widetilde{u})$, which is a contradiction $\square$

Notice that if $\bar{\lambda}$ satisfies (3.47), then for every $\alpha \in \mathbb{R}$, with $\alpha \neq 0$, $\alpha\bar{\lambda}$ satisfies (3.47) 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 non zero).

Hence condition (3.47) implies that there exists $\lambda_0 \in T^*_{q_0}M$ such that one of the following identities is satisfied for all $v \in L^\infty([0,T], \mathbb{R}^m)$:

$$\frac{\partial J(\widetilde{u}+sv)}{\partial s}\Big|_{s=0} = \left\langle \lambda_0, \frac{\partial x(T;\widetilde{u}+sv)}{\partial s}\Big|_{s=0} \right\rangle, \tag{3.50}$$

$$0 = \left\langle \lambda_0, \frac{\partial x(T;\widetilde{u}+sv)}{\partial s}\Big|_{s=0} \right\rangle. \tag{3.51}$$

with $\lambda_0 \neq 0$ in the second case. To end the proof we have to show that identities (3.50) and (3.51) are equivalent to conditions (N) and (A) of Theorem 3.41. Let us show that

$$\frac{\partial J(\widetilde{u}+sv)}{\partial s}\Big|_{s=0} = \int_0^T \sum_{i=1}^m \widetilde{u}_i(t) v_i(t) dt, \tag{3.52}$$

$$\frac{\partial x(T;\widetilde{u}+sv)}{\partial s}\Big|_{s=0} = \int_0^T g^t_{v(t)}(q_0) dt = \int_0^T \sum_{i=1}^m ((P^{-1}_{0,t})_* f_i)(q_0) v_i(t) dt. \tag{3.53}$$

Identity (3.52) follows from the definition of J

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

while (3.53) can be proved in coordinates. Indeed by (3.46) and the linearity of g_v with respect to v we have

$$x(T;\widetilde{u}+sv) = q_0 + s\int_0^T g^t_{v(t)}(x(t;\widetilde{u}+sv)) dt,$$

and differentiating with respect to s at $s=0$ one gets (3.53).

Let us show that (3.50) is equivalent to (N) of Theorem 3.41. Similarly, one gets that (3.51) is equivalent to (A). Using (3.52) and (3.53), equation (3.50) is rewritten as

$$\begin{aligned}\int_0^T \sum_{i=1}^m \widetilde{u}_i(t) v_i(t) dt &= \int_0^T \sum_{i=1}^m \left\langle \lambda_0, ((P^{-1}_{0,t})_* f_i)(q_0) \right\rangle v_i(t) dt \\ &= \int_0^T \sum_{i=1}^m \langle \lambda(t), f_i(\gamma(t)) \rangle v_i(t) dt,\end{aligned} \tag{3.55}$$

where we used, for every $i = 1, \ldots, m$, the identities

$$\left\langle \lambda_0, ((P^{-1}_{0,t})_* f_i)(q_0) \right\rangle = \left\langle \lambda_0, (P^{-1}_{0,t})_* f_i(\gamma(t)) \right\rangle = \left\langle (P^{-1}_{0,t})^* \lambda_0, f_i(\gamma(t)) \right\rangle = \langle \lambda(t), f_i(\gamma(t)) \rangle.$$

Since $v_i(\cdot) \in L^\infty([0,T], \mathbb{R}^m)$ are arbitrary, we get $\widetilde{u}_i(t) = \langle \lambda(t), f_i(\gamma(t)) \rangle$ for a.e. $t \in [0,T]$.

3.A Measurability of the minimal control

In this section we prove a general lemma that implies, for the specific case of sub-Riemannian geometry, that the minimal control associated with an admissible curve is measurable.

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) $g(t, u)$ is measurable with respect to t and continuous with respect to u.

(M2) $v(t)$ is measurable in t.

Moreover we assume that

(M3) 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 (M3) and consider the function $t \mapsto u^*(t)$.

Lemma 3.49. *The function $t \mapsto |u^*(t)|$ is measurable.*

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

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

is measurable. By our assumptions

$$A = \{t \in I : \exists\, u \in U \text{ s.t. } |u| \le r, g(t, u) = v(t)\}$$

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

$$A = \bigcap_{n\in\mathbb{N}} A_n = \bigcap_{n\in\mathbb{N}} \underbrace{\bigcup_{i\in\mathbb{N}} A_{i,n}}_{:=A_n} \tag{3.56}$$

where

$$A_{i,n} := \{t \in I : \ |g(t, u_i) - v(t)| < 1/n\}$$

Notice that the set $A_{i,n}$ is measurable by construction and if (3.56) is true, A is also measurable.

$\subset$ inclusion. Let $t \in A$. This means that there exists $\bar{u} \in U$ such that $|\bar{u}| \le r$ and $f(t, \bar{u}) = v(t)$. Since f 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 $i(n) = i_n$ 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 f in u one easily gets that $f(t, \bar{u}) = v(t)$. That is $t \in A$. ∎

Next we exploit the fact that the function $\varphi(t) := |u^*(t)|$ is measurable to show that the vector function $u^*(t)$ is measurable.

Lemma 3.50. *The vector function $t \mapsto u^*(t)$ is measurable.*

Proof. We have to prove that, for every closed ball O in $\mathbb{R}^n$ (for instance a closed ball) the set

$$B := \{t \in I : u^*(t) \in O\}$$

is measurable. Since the minimum is uniquely determined this is equivalent 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

$$B = \bigcap_{n\in\mathbb{N}} B_n = \bigcap_{n\in\mathbb{N}} \underbrace{\bigcup_{i\in\mathbb{N}} B_{i,n}}_{:=B_n} \tag{3.57}$$

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 (3.57) 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 $f(t,\bar{u}) = v(t)$. Since f 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 $f(t,\bar{u}) = v(t)$. Moreover $|\bar{u}| \leq \varphi(t)$. Hence $|\bar{u}| = \varphi(t)$. That is $t \in B$. ☐

3.A.1 Proof of Lemma 3.12

Consider an admissible curve $\gamma : [0,T] \to M$ and set $g(t,u) = f(\gamma(t),u)$, $v(t) = \dot{\gamma}(t)$.

Notice that assumptions (M1)-(M3) are satisfied. Indeed (M1) and (M2) follows from the fact that $g(t,u)$ is linear with respect to u and measurable in t. Moreover (M3) is also satisfied. .

3.B Lipschitz vs Absolutely continuous admissible curves

In this book 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 3.51. 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 3.52. *The length of an AC-admissible curve is invariant by AC reparametrization.*

Lemma 3.53. *Any AC-admissible curve of positive length is a AC reparametrization of a length-parametrized admissible one.*

The proof of Lemma 3.52 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 (3.13) still holds.

The proof of Lemma 3.53 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

$$d_{AC}(q_0, q_1) = \inf\{\ell(\gamma),\ \gamma\ AC\text{-admissible},\ \gamma(0) = q_0,\ \gamma(T) = q_1\}, \tag{3.58}$$

we have the following proposition.

Proposition 3.54. $d_{AC}(q_0, q_1) = d(q_0, q_1)$

Since $L^2([0,T]) \subset L^1([0,T])$, Lemmas 3.52, 3.53 and Proposition 3.54 are valid also in the framework of admissible curves associated with L^2 controls.

Bibliographical notes

Chapter 4

Hamiltonian setting

This chapter is devoted to the study of geometric properties of Pontryagin extremals. To this purpose we first rewrite Theorem 3.41 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 frame. Finally we prove that small pieces of normal extremal trajectories minimize the length.

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 previuos 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

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

we have that one of the following conditions is satisfied:

(N) $u_i(t) \equiv \langle \lambda(t), f_i(\gamma(t)) \rangle, \quad \forall\, i = 1, \ldots, m,$

(A) $0 \equiv \langle \lambda(t), f_i(\gamma(t)) \rangle, \quad \forall\, i = 1, \ldots, m, \qquad \lambda_0 \neq 0.$

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 is 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 first show that a vector field on T^*M is completely characterized by its action on function 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 $\mathcal{C}^\infty(M)$ are in a one-to-one correspondence with functions in $\mathcal{C}^\infty(T^*M)$ that are *constant on fibers* via the map $\alpha \mapsto \pi^*\alpha = \alpha \circ \pi$. In other words we have the isomorphism of algebras

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

In what follows, with abuse of notation, we often identify the function $\pi^*\alpha \in \mathcal{C}^\infty(T^*M)$ with the function $\alpha \in \mathcal{C}^\infty(M)$.

In a similar way smooth vector fields on M are in a one-to-one correspondence with functions in $\mathcal{C}^\infty(T^*M)$ that are *linear on fibers* via the map $Y \mapsto a_Y$, where $a_Y(\lambda) := \langle \lambda, Y(q)\rangle$ and $q = \pi(\lambda)$.

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

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

Definition 4.1. We say that a function $a \in \mathcal{C}^\infty(T^*M)$ is *affine on fibers* if there exists two functions $\alpha \in \mathcal{C}^\infty_{\mathsf{cst}}(T^*M)$ and $a_X \in \mathcal{C}^\infty_{\mathsf{lin}}(T^*M)$ 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 $\mathcal{C}^\infty_{\mathsf{aff}}(T^*M)$ 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 $\mathcal{C}^\infty(T^*M)$, is completely characterized by its action on affine functions,

Indeed for a vector field $V \in \mathrm{Vec}(T^*M)$ and $f \in \mathcal{C}^\infty(T^*M)$, 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 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 2.10 its 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 $\beta \in \mathcal{C}^\infty(M)$, coincides with the action of X. Indeed we have for all $\lambda \in T^*M$

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

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 $\lambda \in T^*M$

$$\begin{aligned}
\left.\frac{d}{dt}\right|_{t=0} a_Y((e^{-tX})^*\lambda) &= \left.\frac{d}{dt}\right|_{t=0} \left\langle (e^{-tX})^*\lambda, Y(e^{tX}(q)) \right\rangle \\
&= \left.\frac{d}{dt}\right|_{t=0} \left\langle \lambda, (e_*^{-tX} Y)(q) \right\rangle = \langle \lambda, [X,Y](q) \rangle \\
&= a_{[X,Y]}(\lambda).
\end{aligned} \tag{4.7}$$

Hence, by linearity, one gets that the action of V_X on functions of $\mathcal{C}^\infty_{\mathfrak{aff}}(T^*M)$ is

$$V_X(\beta + a_Y) = X\beta + a_{[X,Y]}. \tag{4.8}$$

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

More precisely we start by the following definition.

Definition 4.3. Let $a_X, a_Y \in \mathcal{C}^\infty_{\mathfrak{lin}}(T^*M)$ 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 $\mathcal{C}^\infty_{\mathfrak{lin}}(T^*M)$ 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 $X, Y \in \mathrm{Vec}(M)$:

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

As a consequence, the Poisson bracket is bilinear, skew-symmetric and satisfies the following relation

$$\{a_X, \alpha\, a_Y\} = \{a_X, a_{\alpha Y}\} = a_{[X,\alpha Y]} = \alpha\, a_{[X,Y]} + (X\alpha)\, a_Y, \qquad \forall\, \alpha \in \mathcal{C}^\infty(M). \tag{4.11}$$

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

Now we extend this definition on the whole $\mathcal{C}^\infty(T^*M)$.

Proposition 4.5. *There exists a unique bilinear and skew-simmetric map*

$$\{\cdot,\cdot\} : \mathcal{C}^\infty(T^*M) \times \mathcal{C}^\infty(T^*M) \to \mathcal{C}^\infty(T^*M)$$

that extends (4.9) *on* $\mathcal{C}^\infty(T^*M)$*, and that is a derivation in each argument, i.e. it satisfies*

$$\{a, bc\} = \{a,b\}c + \{a,c\}b, \qquad \forall\, a,b,c \in \mathcal{C}^\infty(T^*M). \tag{4.12}$$

We call this operation the Poisson bracket on $\mathcal{C}^\infty(T^*M)$*.*

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

By linearity and skew-symmetry we are reduced to compute Poisson brackets of kind $\{a_X,\alpha\}$ and $\{\alpha,\beta\}$, where $a_X\in\mathcal{C}^\infty_{\mathfrak{lin}}(T^*M)$ and $\alpha,\beta\in\mathcal{C}^\infty_{\mathfrak{cst}}(T^*M)$. 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 $\mathcal{C}^\infty(T^*M)$ in such a way that it is still a derivation. For $f,g\in\mathcal{C}^\infty(T^*M)$ 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 $\mathcal{C}^\infty_{\mathfrak{aff}}(T^*M)$ such that $d_\lambda f=d_\lambda(a_{f,\lambda})$ and $d_\lambda g=d_\lambda(a_{g,\lambda})$.

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

$$a_X(p,x)=\sum_{i=1}^n p_iX_i(x),\quad a_Y(p,x)=\sum_{i=1}^n p_iY_i(x).$$

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 $a, b \in \mathcal{C}^\infty(T^*M)$

$$\{a, b\} = \sum_{i=1}^{n} \frac{\partial a}{\partial p_i} \frac{\partial b}{\partial x_i} - \frac{\partial a}{\partial x_i} \frac{\partial b}{\partial p_i}, \qquad a, b \in \mathcal{C}^\infty(T^*M). \tag{4.19}$$

The explicit formula (4.19) shows that the extension of the Poisson bracket to $\mathcal{C}^\infty(T^*M)$ is still a derivation. $\square$

Remark 4.6. 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

$$\{\cdot, \cdot\}_\lambda : T^*_\lambda(T^*M) \times T^*_\lambda(T^*M) \to \mathbb{R}.$$

4.1.3 Hamiltonian vector fields

By construction, the linear operator defined by

$$\vec{a} : \mathcal{C}^\infty(T^*M) \to \mathcal{C}^\infty(T^*M) \qquad \vec{a}(b) := \{a, b\} \tag{4.20}$$

is a derivation of the algebra $\mathcal{C}^\infty(T^*M)$, therefore can be identified with an element of $\mathrm{Vec}(T^*M)$.

Definition 4.7. The vector field $\vec{a}$ on T^*M defined by (4.20) is called the *Hamiltonian vector field* associated with the smooth function $a \in \mathcal{C}^\infty(T^*M)$.

From (4.19) we can easily write the coordinate expression of $\vec{a}$ for any arbitrary function $a \in \mathcal{C}^\infty(T^*M)$

$$\vec{a} = \sum_{i=1}^{n} \frac{\partial a}{\partial p_i} \frac{\partial}{\partial x_i} - \frac{\partial a}{\partial x_i} \frac{\partial}{\partial p_i}. \tag{4.21}$$

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

Proposition 4.8. *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

$$\vec{a}_X(\alpha) = \{a_X, \alpha\} = X\alpha$$
$$\vec{a}_X(a_Y) = \{a_X, a_Y\} = a_{[X,Y]}.$$

Hence this action coincides with the action of V as in (4.6) and (4.7). □

Remark 4.9. 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

$$\vec{a}_X = \sum_{i=1}^n X_i \frac{\partial}{\partial x_i} - \sum_{i,j=1}^n p_i \frac{\partial X_i}{\partial x_j} \frac{\partial}{\partial p_j}. \tag{4.22}$$

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 4.10. *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 4.11. 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

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

hence we get

$$\langle s_\lambda, w\rangle = \langle \lambda, \pi_* w\rangle = \sum_{i=1}^n p_i\beta_i = \sum_{i=1}^n p_i \langle dx_i, w\rangle = \left\langle \sum_{i=1}^n p_i dx_i, w \right\rangle.$$

In other words the coordinate expression of the Liouville form s at the point λ coincides with the one of λ itself, namely

$$s_\lambda = \sum_{i=1}^n p_i dx_i. \tag{4.23}$$

Exercise 4.12. 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 $\pi \circ \omega = id$).

Definition 4.13. 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 σ is non degenerate and its expression in canonical coordinates (p, x) is written as follows

$$\sigma = \sum_{i=1}^n dp_i \wedge dx_i. \tag{4.24}$$

Remark 4.14 (The symplectic form in non-canonical coordinates). Given a basis of 1-forms $\omega_1, \ldots, \omega_n$ 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

$$s = \sum_{i=1}^n h_i\omega_i, \qquad \sigma = ds = \sum_{i=1}^n dh_i \wedge \omega_i + h_i d\omega_i. \tag{4.25}$$

Notice that, with respect to (4.24) 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

$$\left\{\begin{matrix}\text{non degenerate}\\ \text{bilinear forms on } V\end{matrix}\right\} \simeq \left\{\begin{matrix}\text{linear invertible maps}\\ V \to V^*\end{matrix}\right\} \simeq \left\{\begin{matrix}\text{non degenerate}\\ \text{bilinear forms on } V^*\end{matrix}\right\}. \tag{4.26}$$

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

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

2. 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 4.6) are dual under the identification (4.26). In particular show that

$$\{a, b\} = \vec{a}(b) = \langle db, \vec{a}\rangle = \sigma(\vec{a}, \vec{b}), \qquad \forall\, a, b \in \mathcal{C}^\infty(T^*M). \tag{4.27}$$

Remark 4.16. 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 (4.24) 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 4.17. *A function $a \in \mathcal{C}^\infty(T^*M)$ is a constant of the motion of the Hamiltonian system associated with $h \in \mathcal{C}^\infty(T^*M$ 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 $\lambda_0 \in T^*M$. Let us prove the following formula for the derivative of the function a along the solution

$$\frac{d}{dt}a(\lambda(t)) = \{h, a\}(\lambda(t)). \tag{4.28}$$

By (4.28) it is easy to see that, if $\{h, a\} = 0$, then the derivative of the function a along the flow vanishes for all t and then a is constant. Conversely, if a is constant along the flow then its derivative vanishes and the Poisson bracket is zero. $\square$

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

Corollary 4.18. *A function $h \in \mathcal{C}^\infty(T^*M)$ 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 the Pontryagin Maximum Principle (see Theorem 3.41) using the symplectic language developed in the last section.

Given a sub-Riemannian structure on M with generating frame $\{f_1, \ldots, f_m\}$, 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 4.19 (PMP)**.** *Let $\gamma : [0,T] \to M$ be an admissible curve which is a length-minimizer, parametrized by constant speed. Let $\widetilde{u}(\cdot)$ be the corresponding minimal control. Then there exists a Lipschitz curve $\lambda(t) \in T^*_{\gamma(t)}M$ such that*

$$\dot{\lambda}(t) = \sum_{i=1}^{m} \widetilde{u}_i(t) \vec{h}_i(\lambda(t)), \qquad a.e.\ t \in [0,T], \tag{4.29}$$

and one of the following conditions is satisfied:

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

(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 3.41, combining Proposition 4.8 and Exercise 4.10. □

Notice that Theorem 4.19 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 controls $\widetilde{u}_i(t)$ associated with the extremal trajectory.

Moreover, the actual formulation of Theorem 4.29 for the necessary condition for optimality still does not clarify if the extremals depend on the generating frame $\{f_1, \ldots, f_m\}$ 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 an *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 4.20. Let M be a sub-Riemannian manifold. The *sub-Riemannian Hamiltonian* is the smooth function on T^*M defined as follows

$$H : T^*M \to \mathbb{R}, \qquad H(\lambda) = \max_{u \in U_q} \left(\langle \lambda, f_u(q) \rangle - \frac{1}{2}|u|^2 \right), \quad q = \pi(\lambda). \tag{4.30}$$

Proposition 4.21. *The sub-Riemannian Hamiltonian H is quadratic on fibers. Moreover, for every generating frame $\{f_1, \ldots, f_m\}$ of the sub-Riemannian structure, the sub-Riemannian Hamiltonian H is written as follows*

$$H(\lambda) = \frac{1}{2} \sum_{i=1}^{k} \langle \lambda, f_i(q) \rangle^2, \qquad \lambda \in T_q^* M, \quad q = \pi(\lambda). \tag{4.31}$$

Proof. In terms of a generating frame $\{f_1, \ldots, f_m\}$, the sub-Riemannian Hamiltonian (3.55) is written as follows

$$H(\lambda) = \max_{u \in \mathbb{R}^m} \left(\sum_{i=1}^{m} u_i \langle \lambda, f_i(q) \rangle - \frac{1}{2} \sum_{i=1}^{m} u_i^2 \right). \tag{4.32}$$

Differentiating (4.32) with respect to u_i, one gets that the maximum is attained at $u_i = \langle \lambda, f_i(q) \rangle$, from which formula (4.31) follows. The fact that H is quadratic on fibers then easily follows from (4.31). $\square$

Exercise 4.22. Prove that two equivalent sub-Riemannian structures $(\mathbf{U}, f)$ and $(\mathbf{U}', f')$ on a manifold M define the same Hamiltonian.

Theorem 4.23. *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 frame 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 $h_i(\lambda(t)) = \widetilde{u}_i(t)$ by condition (N) of Theorem 4.19, one gets

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

Remark 4.24. In canonical coordinates $\lambda = (p, x)$, 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

$$\begin{cases} \dot{x} = \dfrac{\partial H}{\partial p} = \sum_{i=1}^{m} \langle p, f_i(x) \rangle f_i(x) \\ \dot{p} = -\dfrac{\partial H}{\partial x} = -\sum_{i=1}^{m} \langle p, f_i(x) \rangle \langle p, D_x f_i(x) \rangle \end{cases} \tag{4.33}$$

From here it is easy to see that if $\lambda(t) = (p(t), x(t))$ is a solution of (4.33) 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 $\alpha > 0$.

Proposition 4.25. *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. Thanks to Proposition 4.17 the sub-Riemannian Hamiltonian H is constant along the flow of $\vec{H}$, i.e. on normal extremals. Hence for every normal extremal $\lambda(t)$ associated with a control $u(\cdot)$ we have

$$\begin{aligned} \frac{1}{2}\int_0^T \|\gamma(t)\|^2 dt = \frac{1}{2}\int_0^T |u(t)|^2 dt = \frac{1}{2}\int_0^T \sum_{i=1}^k u_i(t)^2 dt \\ = \int_0^T H(\lambda(t))dt = H(\lambda_0)T, \end{aligned} \tag{4.34}$$

where we used the fact that, along a normal extremal, we have the relations for all $t \in [0,T]$

$$H(\lambda(t)) = H(\lambda_0) \qquad u_i(t) = \langle \lambda(t), f_i(\gamma(t)) \rangle . \tag{4.35}$$

The fact that H is constant along $\lambda(t)$, easily implies by (4.34) that $\|\dot\gamma(t)\|^2$ is constant. Moreover one easily gets that $\|\dot\gamma(t)\| = 1$ if and only if $H(\lambda(t)) \equiv 1/2$.

Moreover, by Remark 4.24, all normal extremal trajectories are reparametrization of length parametrized ones. □

Remark 4.26. Let $\lambda(t)$ be a normal extremal such that $\lambda(0) = \lambda_0 \in T^*_{q_0}M$. The corresponding normal extremal path $\gamma(t) = \pi(\lambda(t))$ can be written in the exponential notation

$$\gamma(t) = \pi \circ e^{t\vec{H}}(\lambda_0).$$

By the previous discussion 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 4.27. Let M be a smooth manifold and $\Omega \in \Lambda^k M$ a 2-form. A Lipschitz curve $\gamma : [0,T] \to M$ is said *characteristic* for Ω if for almost every $t \in [0,T]$ it holds

$$\dot\gamma(t) \in \operatorname{Ker}\Omega_{\gamma(t)}, \quad (\text{i.e. } \Omega_{\gamma(t)}(\dot\gamma(t),\cdot) = 0) \tag{4.36}$$

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

Proposition 4.28. *Let H be the sub-Riemannian Hamiltonian and assume that $c > 0$ is a regular value of H. Then a curve γ is a characteristic curve of $\sigma|_{H^{-1}(c)}$ 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.[1] 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 $\lambda \in H^{-1}(c)$.

[1] by Sard Theorem almost every $c > 0$ is regular value.

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 subspace that is skew-orthogonal to ξ, namely

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

Since, by antisymmetricity, $\sigma_\lambda(\xi, \xi) = 0$, 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

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

Exercise 4.29. The sub-Riemannian Hamiltonian encodes all the informations about the distribution and the metric defined on it.

1. Prove that a vector $v \in T_qM$ satisfies $v \in \mathcal{D}_q$ and $\|v\| \leq 1$ if and only if

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

2. Show that this implies the following characterization for the sub-Riemannian Hamiltonian

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

In particular if 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 their a priori 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 frame 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

$$H^{-1}(0) = \{\lambda \in T^*M, \langle \lambda, f_i(q)\rangle = 0, \, i = 1, \dots, m, q = \pi(\lambda)\}. \tag{4.37}$$

where H denotes the sub-Riemannian Hamiltonian (4.31).

Proposition 4.30. *Let H be the sub-Riemannian Hamiltonian and assume that 0 is a regular value of H. Then a curve γ is a characteristic curve of $\sigma|_{H^{-1}(0)}$ if and only if it is the reparametrization of a normal 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

$$\operatorname{Ker} \sigma_\lambda|_N = T_\lambda N^\angle = \operatorname{span}\{\vec{h}_i(\lambda),\, i = 1, \dots, m\}. \tag{4.38}$$

Indeed, from the definition of N, it follows that

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

and (4.38) follows by taking the skew-orthogonal. 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

$$\dot{\lambda}(t) = \sum_{i=1}^{m} u_i(t)\vec{h}_i(\lambda(t)). \qquad \square \tag{4.39}$$

The following exercise shows that the assumption of Proposition 4.30 is always satisfied in the case of a regular sub-Riemannian structure.

Exercise 4.31. Assume that the sub-Riemannian structure is *regular*, namely the following assumption holds

$$\dim \mathcal{D}_q = \dim \operatorname{span}_q\{f_1, \dots, f_m\} = \text{const}. \tag{4.40}$$

Shows that, under this assumption, 0 is a regular value for H. In particular, the set $H^{-1}(0)$ defined by (4.37) is a smooth submanifold of T^*M.

Remark 4.32. From Proposition 4.30 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 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 we can select (at least locally) a basis of 1-forms $\omega_1, \dots, \omega_m$ for the dual of the distribution

$$\mathcal{D}_q^\perp = \operatorname{span}\{\omega_i(q),\, i = 1, \dots, m\}, \tag{4.41}$$

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 4.14. In these coordinates the restriction of the symplectic structure $\mathcal{D}^\perp$ to is expressed as follows

$$\sigma|_{\mathcal{D}^\perp} = d(s|_{\mathcal{D}^\perp}) = \sum_{i=1}^{m} dh_i \wedge \omega_i + h_i d\omega_i, \tag{4.42}$$

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.

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

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

Since the symplectic form σ is skew-symmetric, by a dimensional argument we easily get 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 4.30.

Let us consider in more detail the simplest case $n = 3$. Assume that there exists a one form $\omega_0 \in \Lambda^1(M)$ such that $\mathcal{D} = \ker \omega$ (this is not restrictive, at least for a local description). Consider a basis of one forms $\omega_0, \omega_1, \omega_2$ such that $\omega_0 := \omega$ and the associated coordinates h_0, h_1, h_2 the coordinate associated to these forms (see Remark 4.14). By (4.42)

$$\sigma|_{\mathcal{D}^\perp} = dh_0 \wedge \omega + h_0\, d\omega, \tag{4.43}$$

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

$$\sigma \wedge \sigma|_{\mathcal{D}^\perp} = 2h_0\, dh_0 \wedge \omega \wedge d\omega. \tag{4.44}$$

Lemma 4.33. *Let N be a smooth $2k$-dimensional manifold and $\Omega \in \Lambda^2 M$. Then Ω is nondegenerate on N if and only if $\wedge^k \Omega \neq 0$.*[2]

Definition 4.34. Let M be a three dimensional manifold. We say that a constant rank distribution $\mathcal{D}$ on M of corank one is a *contact distribution* if $\omega \wedge d\omega \neq 0$.

Since M is three dimensional, the differential form $\omega \wedge d\omega$ is a top dimensional form, hence it is meaningful to consider the set, called *Martinet set*

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

Corollary 4.35. *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. Assume that the structure is contact. Then $\omega \wedge d\omega \neq 0$ and, thanks to (4.44), it follows that $\sigma \wedge \sigma|_{\mathcal{D}^\perp} \neq 0$. By Lemma 4.33 $\sigma|_{\mathcal{D}^\perp}$ is non degenerate (notice that dh_0 is always independent on $\omega \wedge d\omega$ since they depend on coordinates on the fibers and on the manifold, respectively). This shows that, under the contact assumption, the set $\mathfrak{M}$ is empty and there exists no nontrivial characteristic curve of $\sigma|_{\mathcal{D}^\perp}$. The first part of the statement follows by analogue arguments. $\square$

Remark 4.36. Since M is three dimensional, we can write

$$\omega \wedge d\omega = a dV$$

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

where $a \in C^\infty(M)$ and dV is some smooth volume form on M, that is 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 a is never vanishing. If 0 is a regular value of a, the set $a^{-1}(0)$ defines a two dimensional surface on M, called the *Martinet surface.* Recall that this condition is true for a generic choice of the distribution.

In this case abnormal extremal trajectories can be precisely characterized as the horizontal curves that are contained in the Martinet surface $\mathfrak{M}$. 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 are exactly (reparametrized) integral curves of this line field.

Exercise 4.37. 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 a reparametrization.

4.4 Lie derivatives

In this section we extend the notion of Lie derivative, already introduced for vector fields in Section 3.3), to differential forms. Recall that if $X, Y \in \mathrm{Vec}(M)$ are two vector fields we define

$$\mathcal{L}_X Y = \frac{d}{dt}\bigg|_{t=0} e_*^{-tX} Y = [X, Y].$$

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:

$$(P^*\omega)_q(\xi_1, \ldots, \xi_k) := \omega_{P(q)}(P_*\xi_1, \ldots, P_*\xi_k), \quad q \in M, \quad \xi_i \in T_q M. \tag{4.45}$$

It is an easy check that this operation is linear and satisfies the two following properties

$$P^*(\omega_1 \wedge \omega_2) = P^*\omega_1 \wedge P^*\omega_2, \tag{4.46}$$

$$P^* \circ d = d \circ P^*. \tag{4.47}$$

Definition 4.38. Let $X \in \mathrm{Vec}(M)$ and $\omega \in \Lambda^k M$. We define the *Lie derivative* of ω with respect to X as

$$\mathcal{L}_X : \Lambda^k M \to \Lambda^k M, \qquad \mathcal{L}_X \omega = \frac{d}{dt}\bigg|_{t=0} (e^{tX})^*\omega. \tag{4.48}$$

From (4.46) and (4.47), 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 4.39. Let $X \in \text{Vec}(M)$ and $\omega \in \Lambda^k M$. 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

$$(i_X\omega)(Y_1, \ldots, Y_{k-1}) := \omega(X, Y_1, \ldots, Y_{k-1}), \quad Y_i \in \text{Vec}(M). \tag{4.49}$$

One can show that the operator i_X is an *anti-derivation*, in the following sense:

$$i_X(\omega_1 \wedge \omega_2) = (i_X\omega_1) \wedge \omega_2 + (-1)^{k_1}\omega_1 \wedge (i_X\omega_2), \quad \omega_i \in \Lambda^{k_i} M, \quad i = 1, 2. \tag{4.50}$$

We end this section proving two classical formulas linking together these notions, and usually referred as Cartan's formulas.

Proposition 4.40 (Cartan's formula)**.** *The following identity holds true*

$$\mathcal{L}_X = i_X \circ d + d \circ i_X. \tag{4.51}$$

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 can write

$$d \circ D_X = d \circ i_X \circ d = D_X \circ d.$$

Moreover, 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 4.41. *Let* $X, Y \in \text{Vec}(M)$ *and* $\omega \in \Lambda^1 M$*, then*

$$d\omega(X, Y) = X\langle \omega, Y\rangle - Y\langle \omega, X\rangle - \langle \omega, [X, Y]\rangle. \tag{4.52}$$

Proof. On one hand Definition 4.38 implies, by Leibnitz rule

$$\begin{aligned} \langle \mathcal{L}_X\omega, Y\rangle_q &= \left.\frac{d}{dt}\right|_{t=0} \left\langle (e^{tX})^*\omega, Y\right\rangle_q \\ &= \left.\frac{d}{dt}\right|_{t=0} \left\langle \omega, e^{tX}_* Y\right\rangle_{e^{tX}(q)} \\ &= X\langle \omega, Y\rangle - \langle \omega, [X, Y]\rangle. \end{aligned}$$

On the other hand, Cartan's formula (4.51) 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 (4.52). $\square$

4.5 Examples

4.5.1 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 geodesics in M could be described as optimal control problem

$$\dot{q} = u_1 f_1(q) + u_2 f_2(q),$$

where length and action are expressed as

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

Equations of geodesics are projections of integral curves of the sub-Riemannian Hamiltonian in T^*M

$$H = \frac{1}{2}(h_1^2 + h_2^2), \qquad h_i(\lambda) = \langle \lambda, f_i(q) \rangle .$$

Now we consider coordinates (q, h_1, h_2) on T^*M and using that on solutions we have $u_i(t) = h_i(\lambda_t)$, we find equation on the base:

$$\dot{q} = h_1 f_1(q) + h_2 f_2(q), \tag{4.53}$$

and equation on the fiber (remember that along solutions $\dot{a} = \{H, a\}$)

$$\begin{cases} \dot{h}_1 = \{H, h_1\} = -\{h_1, h_2\} h_2 \\ \dot{h}_2 = \{H, h_2\} = \{h_1, h_2\} h_1 \end{cases} \tag{4.54}$$

from here we can also show directly that H is constant along solution, because

$$\dot{H} = h_1 \dot{h}_1 + h_2 \dot{h}_2 = 0.$$

We fix the level

$$H = \frac{1}{2} \Longleftrightarrow h_1^2 + h_2^2 = 1,$$

and we restrict to the spherical bundle SM (see Example 2.43)

$$h_1 = \cos\theta, \qquad h_2 = \sin\theta,$$

and equations (4.53) and (4.54) become:

$$\begin{cases} \dot{\theta} = \{h_1, h_2\} \\ \dot{q} = \cos\theta f_1(q) + \sin\theta f_2(q) \end{cases} \tag{4.55}$$

Moreover we have $\{h_1, h_2\}(\lambda) = \langle \lambda, [f_1, f_2] \rangle$ so that, if we let

$$[f_1, f_2] = a_1 f_1 + a_2 f_2, \qquad a_1, a_2 \in \mathcal{C}^\infty(M),$$

we have

$$\{h_1, h_2\} = a_1 h_1 + a_2 h_2.$$

$$
\begin{cases} \dot{\theta} = a_1(q)\cos\theta + a_2(q)\sin\theta \\ \dot{q} = \cos\theta f_1(q) + \sin\theta f_2(q) \end{cases} \tag{4.56}
$$

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

$$
\vec{H} = \cos\theta f_1 + \sin\theta f_2 + (a_1\cos\theta + a_2\sin\theta)\partial_\theta. \tag{4.57}
$$

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 (1.25) 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

$$
\begin{aligned}
d\omega &= -da_1 \wedge \mu_1 - da_2 \wedge \mu_2 - a_1 d\mu_1 - a_2 d\mu_2 \\
&= -(f_1(a_2) - f_2(a_1) - a_1^2 - a_2^2)\mu_1 \wedge \mu_2.
\end{aligned}
$$

□

4.5.2 Isoperimetric problem

Let M be a 2-dimensional Riemannian manifold, ν its volume form. $A \in \Lambda^1 M$ and $c \in \mathbb{R}$ fixed.
Problem. Fixed $q_0, q_1 \in M$, find (if exists) the minimum:

$$\min\{\ell(\gamma), \gamma(0) = q_0, \gamma(T) = q_1, \int_\gamma A = c\}. \tag{4.58}$$

Remark 4.43. Local minimizers depend only on dA, i.e. if we add an exact term to A we will find same minima for the problem (obviously value of c will change!).

Problem 1 can be reformulated as a sub-Riemannian problem on the extended manifold

$$\widehat{M} = \mathbb{R} \times M,$$

that means that solutions of the problem (4.58) turns to be geodesics for the sub-Riemannian structure on $\widehat{M}$.

So we define on the extended manifold the 1-form:

$$\omega = dy - A, \qquad \widehat{M} = \{(y, q), y \in \mathbb{R}, q \in M\}.$$

Admissible curves are pairs $z(t) = (y(t), \gamma(t))$ such that $\dot{z}(t) \in \Delta_{z(t)}$, i.e. $\omega(\dot{z}(t)) = 0$. This implies

$$\omega(\dot{z}(t)) = \dot{y}(t) - \langle A, \dot{\gamma}(t)\rangle = 0.$$

In other words $\gamma(t)$ is a curve on M and $y(t)$ satisfies the identity

$$y(t) = y_0 + \int_{\gamma_t} A, \qquad \text{where } \gamma_t = \gamma|_{[0,t]}.$$

In particular we can recover a basis for the distribution

$$\begin{cases} \dot{\gamma} = u_1 f_1 + u_2 f_2 \\ \dot{y} = u_1 \langle A, f_1\rangle \partial_y + u_2 \langle A, f_2\rangle \partial_y \end{cases} \Rightarrow \begin{pmatrix} \dot{\gamma} \\ \dot{y} \end{pmatrix} = u_1 \begin{pmatrix} f_1 \\ \langle A, f_1\rangle \partial_y \end{pmatrix} + u_2 \begin{pmatrix} f_2 \\ \langle A, f_2\rangle \partial_y \end{pmatrix}, \tag{4.59}$$

and $\mathcal{D} = \operatorname{span}(F_1, F_2)$ where

$$F_1 = f_1 + \langle A, f_1\rangle \partial_y, \qquad F_2 = f_2 + \langle A, f_2\rangle \partial_y.$$

Remark 4.44. Notice that the projection of the control system

$$\dot{z} = u_1 F_1(z) + u_2 F_2(z),$$

on the manifold M is

$$\dot{\gamma} = u_1 f_1(\gamma) + u_2 f_2(\gamma),$$

from which follows that the sub-Riemannian length on $\widehat{M}$ coincide exactly with the Riemannian one on M.

We denote with $h_i = \langle \lambda, F_i(q) \rangle$ the Hamiltonians linear on fibers of $T^*\widehat{M}$ and we want to compute normal and abnormal geodesics of this problem. With analogous computations of 2D case we get

$$\begin{cases} \dot{q} = \cos\theta F_1(q) + \sin\theta F_2(q) \\ \dot{\theta} = \{h_1, h_2\} \end{cases} \tag{4.60}$$

where we have to compute $\{h_1, h_2\} = \langle \lambda, [F_1, F_2] \rangle$. We set, as in the previous paragraph:

$$[f_1, f_2] = a_1 f_1 + a_2 f_2, \qquad a_1, a_2 \in \mathcal{C}^\infty(M). \tag{4.61}$$

so that

$$\begin{aligned} [F_1, F_2] &= [f_1 + \langle A, f_1 \rangle \partial_y, f_2 + \langle A, f_2 \rangle \partial_y] \\ &= [f_1, f_2] + (f_1 \langle A, f_2 \rangle - f_2 \langle A, f_1 \rangle)\partial_y \\ \text{(by (4.61))} \qquad &= a_1(F_1 - a_1 \langle A, f_1 \rangle) + a_2(F_2 - a_2 \langle A, f_2 \rangle) + f_1 \langle A, f_2 \rangle - f_2 \langle A, f_1 \rangle)\partial_y \\ &= a_1 F_1 + a_2 F_2 + dA(f_1, f_2)\partial_y. \end{aligned}$$

where in the last equality we use (4.52).

Being $dA \in \Lambda^2 M$ a volume form, $dA = b\mu_1 \wedge \mu_2$, for some $b \in \mathcal{C}^\infty(M)$. Then

$$[F_1, F_2] = a_1 F_1 + a_2 F_2 + b\partial_y.$$

Let we consider now h_0 linear on fibers Hamiltonian associated with ∂_y. From the previous we get

$$\{h_1, h_2\} = a_1 h_1 + a_2 h_2 + b h_0.$$

It follows that

$$\begin{cases} \dot{\theta} = a_1 \cos\theta + a_2 \sin\theta + b h_0 \\ \dot{h}_0 = 0 \Rightarrow h_0 = \text{ const.} \end{cases} \tag{4.62}$$

In other words

$$\kappa_g(\gamma) = \dot{\theta} - a_1(q)\cos\theta - a_2(q)\sin\theta = h_0 b. \tag{4.63}$$

Normal geodesics are curves with geodesic curvature proportional to the function b at every point. In the case $M = \mathbb{R}^2$ and $b = b_0$ costant we have that normal geodesics of this problem are circles on M (and helix on $\widehat{M}$).

Abnormal geodesics are contained in the set of points where $\omega \wedge d\omega = 0$.

$$\begin{aligned} \omega \wedge d\omega &= (dy - A) \wedge (b\mu_1 \wedge \mu_2) \\ &= b dy \wedge \mu_1 \wedge \mu_2. \end{aligned}$$

In other words abnormal geodesics are connected components of $b^{-1}(0)$. They are *independent* on the metric and, in general, they are not normal geodesics.

4.5.3 Heisenberg case

To be done

4.6 Symplectic geometry

In this section we generalize some of the construction we considered on the cotangent bundle T^*M to the case of a general symplectic manifold.

Definition 4.45. 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 necessarily 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 $h \in \mathcal{C}^\infty(N)$ by the formula $i_{\vec{h}}\sigma = -dh$, or equivalently $\sigma(\cdot, \vec{h}) = dh$.

Proposition 4.46. *A diffeomorphism $\phi : N \to N$ is a symplectomorphism if and only if for every $h \in \mathcal{C}^\infty(N)$:*

$$(\phi_*^{-1})\vec{h} = \overrightarrow{h \circ \phi}. \tag{4.64}$$

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 exactly (4.64). The converse implication follows analogously. □

Next we want to characterize those vector fields whose flow generates a one-parametric family of symplectomorphisms.

Lemma 4.47. *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 $t \in \mathbb{R}$:

$$\frac{d}{dt}(e^{tX})^*\sigma = \left.\frac{d}{ds}\right|_{s=0} (e^{tX})^*(e^{sX})^*\sigma = (e^{tX})^*\mathcal{L}_X\sigma.$$

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 (4.51) reads as follows

$$\mathcal{L}_X \sigma = d(i_X \sigma) + i_X(d\sigma) = d(i_X \sigma).$$

This proves the the equivalence between (ii) and (iii). $\square$

Corollary 4.48. *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 4.47 holds. $\square$

Notice that the converse of Corollary 4.48 is true when N is simply connected, since in this case every closed form is exact.

Definition 4.49. Let (N, σ) be a symplectic manifold and $a, b \in \mathcal{C}^\infty(N)$. 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 4.50. *The Poisson bracket satisfies the identities*

(i) $\{a, b\} \circ \phi = \{a \circ \phi, b \circ \phi\}, \qquad \forall\, a, b \in \mathcal{C}^\infty(N), \forall\, \phi \in \mathit{Sympl}(N),$

(ii) $\{a, \{b, c\}\} + \{c, \{a, b\}\} + \{b, \{c, a\}\} = 0, \qquad \forall\, a, b, c \in \mathcal{C}^\infty(N).$

Proof. Property (i) follows from (4.64). Property (ii) follows by considering $\phi = e^{t\vec{c}}$ in (i), for some $c \in \mathcal{C}^\infty(N)$,. and computing the derivative with respect to t at $t = 0$. $\square$

Finally we are able to prove the following generalization of (**??**).

Corollary 4.51. *For every $a, b \in \mathcal{C}^\infty(N)$ we have*

$$\overrightarrow{\{a, b\}} = [\vec{a}, \vec{b}]. \tag{4.65}$$

Proof. Property (ii) of Proposition 4.50 can be rewritten, by skew-symmetry of the Poisson bracket, as follows

$$\{\{a, b\}, c\} = \{a, \{b, c\}\} - \{b, \{a, c\}\}. \tag{4.66}$$

Using that $\{a, b\} = \sigma(\vec{a}, \vec{b}) = \vec{a}b$ one can rewrite again (4.66) as

$$\overrightarrow{\{a, b\}}c = \vec{a}(\vec{b}c) - \vec{b}(\vec{a}c) = [\vec{a}, \vec{b}]c.$$ $\square$

Remark 4.52. Property (ii) of Proposition 4.50 says that $\{a, \cdot\}$ is a derivation of the algebra $\mathcal{C}^\infty(N)$. Moreover, the space $\mathcal{C}^\infty(N)$ endowed with $\{\cdot, \cdot\}$ as a product is a Lie algebra isomorphic to a subalgebra of $\mathrm{Vec}(N)$. Indeed, by (4.65), the correspondence $a \mapsto \vec{a}$ is a Lie algebra homomorphism between $\mathcal{C}^\infty(N)$ 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 $a \in \mathcal{C}^\infty(M)$ and consider the smooth submanifold of T^*M defined by the graph of its differential

$$\mathcal{L}_0 = \{d_q a \,|\, q \in M\} \subset T^*M. \tag{4.67}$$

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 $\dim \mathcal{L}_0 = n$. Let us then consider the image $\mathcal{L}_t$ of $\mathcal{L}_0$ under the Hamiltonian flow

$$\mathcal{L}_t := e^{t\vec{H}}(\mathcal{L}_0), \qquad t > 0, \tag{4.68}$$

and define the $(n+1)$-dimensional manifold with boundary in $T^*M \times \mathbb{R}$ as follows

$$\begin{aligned} \mathcal{L} &= \{(t, \lambda) \in \mathbb{R} \times T^*M \,|\, \lambda \in \mathcal{L}_t,\, 0 \leq t \leq T\} && (4.69)\\ &= \{(t, e^{t\vec{H}}\lambda_0) \in \mathbb{R} \times T^*M \,|\, \lambda_0 \in \mathcal{L}_0,\, 0 \leq t \leq T\}. && (4.70)\end{aligned}$$

Here we assume that the Hamiltonian flow is defined on the interval $[0, T]$.

Finally, let us introduce the *Poincaré-Cartan* 1-form on $T^*M \times \mathbb{R} \simeq T^*(M \times \mathbb{R})$ defined by

$$s - H dt \in \Lambda^1(T^*M \times \mathbb{R})$$

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 4.53. $s|_{\mathcal{L}_0} = d(a \circ \pi)|_{\mathcal{L}_0}$

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 4.54. *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

$$d(s - Hdt) = \sigma - dH \wedge dt, \tag{4.71}$$

vanishes when applied to a 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 4.53 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:

$$\frac{\partial \Psi}{\partial t}(t,\lambda) = \vec{H}(\lambda) + \frac{\partial}{\partial t} \in T_{(t,\lambda)}\mathcal{L}. \tag{4.72}$$

It is then sufficient to show that the vector (4.72) is in the kernel of the two form $\sigma - dH \wedge dt$. In other words we have to prove

$$i_{\vec{H}+\partial_t}(\sigma - dH \wedge dt) = 0. \tag{4.73}$$

The last equality follows from 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

$$\int_\Gamma s - Hdt = 0. \tag{4.74}$$

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 $\mathcal{L}_0$):

$$\int_{\Gamma_0} s - H dt = \int_{\Gamma_0} s = \int_{\Gamma_0} d(a \circ \pi) = 0,$$

where we used Lemma 4.53 and the fact that the integral of an exact form over a closed curve is zero. Then (4.74) follows. $\square$

4.7.2 Normal trajectory 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

$$\dot{q} = f_u(q) = \sum_{i=1}^{m} u_i f_i(q), \tag{4.75}$$

where $f_1, \ldots, f_m$ is a generating frame for the sub-Riemannian structure are projections of integral curves of the Hamiltonian vector fields associated with the sub-Riemannian Hamiltonian

$$\dot{\lambda}(t) = \vec{H}(\lambda(t)), \qquad \text{(i.e. } \lambda(t) = e^{t\vec{H}}(\lambda_0)), \tag{4.76}$$

$$\gamma(t) = \pi(\lambda(t)), \qquad t \in [0, T]. \tag{4.77}$$

where

$$H(\lambda) = \max_{u \in U_q} \left\{ \langle \lambda, f_u(q) \rangle - \frac{1}{2}|u|^2 \right\} = \frac{1}{2} \sum_{i=1}^{m} \langle \lambda, f_i(q) \rangle^2 . \tag{4.78}$$

Theorem 4.55. *Assume that there exists $a \in \mathcal{C}^\infty(M)$ 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*

$$\widetilde{\gamma}(t) = \pi \circ e^{t\vec{H}}(\lambda_0), \qquad t \in [0, T], \tag{4.79}$$

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 $\widetilde{\gamma}(t)$, associated with the control $u(t)$ and such that $\gamma(0) = \widetilde{\gamma}(0)$ and $\gamma(T) = \widetilde{\gamma}(T)$. We denote by $\widetilde{u}(t)$ the control associated with the curve $\widetilde{\gamma}(t)$.

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 $\widetilde{\Gamma}$ in $\mathcal{L}$ defined by

$$\Gamma(t) = (t, \lambda(t)), \qquad \widetilde{\Gamma}(t) = (t, \widetilde{\lambda}(t)) \tag{4.80}$$

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

$$H(\lambda(t)) \leq \left\langle \lambda(t), f_{u(t)}(\gamma(t)) \right\rangle - \frac{1}{2}|u(t)|^2, \qquad \gamma(t) = \pi(\lambda(t)), \tag{4.81}$$

where $\lambda(t)$ is a lift of the trajectory $\gamma(t)$ associated with a control $u(t)$. Moreover, the equality holds in (4.81) if and only if $\lambda(t)$ is a solution of the Hamiltonian system $\dot{\lambda}(t) = H(\lambda(t))$. For this reason we have the relations

$$H(\lambda(t)) < \left\langle \lambda(t), f_{u(t)}(\gamma(t)) \right\rangle - \frac{1}{2}|u(t)|^2, \tag{4.82}$$

$$H(\widetilde{\lambda}(t)) = \left\langle \widetilde{\lambda}(t), f_{\widetilde{u}(t)}(\widetilde{\gamma}(t)) \right\rangle - \frac{1}{2}|\widetilde{u}(t)|^2. \tag{4.83}$$

since $\widetilde{\lambda}(t)$ is a solution of the Hamiltonian equation by assumptions, while $\lambda(t)$ is not. Indeed $\lambda(t)$ and $\widetilde{\lambda}(t)$ have the same initial condition, hence, by uniqueness of the solution of the Cauchy problem, it follows that $\dot{\lambda}(t) = H(\lambda(t))$ if and only if $\lambda(t) = \widetilde{\lambda}(t)$, that implies that $\widetilde{\gamma}(t) = \gamma(t)$.

Let us then show that the energy associated with the curve γ is bigger than the one of the curve $\widetilde{\gamma}$. Actually we prove the following chain of (in)equalities

$$\frac{1}{2}\int_0^T |\widetilde{u}(t)|^2 dt = \int_{\widetilde{\Gamma}} s - H dt = \int_{\Gamma} s - H dt < \frac{1}{2}\int_0^T |u(t)|^2 dt, \tag{4.84}$$

where Γ and $\widetilde{\Gamma}$ are the curves in $\mathcal{L}$ defined in (4.80).

By Lemma 4.54, the 1-form $s - Hdt$ is exact. Then the integral over the closed curve $\Gamma \cup \widetilde{\Gamma}$ vanishes, and one gets

$$\int_{\widetilde{\Gamma}} s - H dt = \int_{\Gamma} s - H dt.$$

The last inequality in (4.84) can be proved as follows

$$\begin{aligned}
\int_{\Gamma} s - H dt &= \int_0^T \langle \lambda(t), \dot{\gamma}(t) \rangle - H(\lambda(t)) dt \\
&= \int_0^T \left\langle \lambda(t), f_{u(t)}(\gamma(t)) \right\rangle - H(\lambda(t)) dt \\
&< \int_0^T \left\langle \lambda(t), f_{u(t)}(\gamma(t)) \right\rangle - \left(\left\langle \lambda(t), f_{u(t)}(\gamma(t)) \right\rangle - \frac{1}{2}|u(t)|^2 \right) dt \qquad (4.85) \\
&= \frac{1}{2}\int_0^T |u(t)|^2 dt.
\end{aligned}$$

where we used (4.82). A similar computation gives computation, using (4.83), gives

$$\int_{\widetilde{\Gamma}} s - H dt = \frac{1}{2}\int_0^T |\widetilde{u}(t)|^2 dt, \tag{4.86}$$

that ends the proof of (4.84). □

As a corollary we state a local version of the same theorem, that can be proved by adapting the above technique.

Corollary 4.56. *Assume that there exists $a \in \mathcal{C}^\infty(M)$ and neighborhoods Ω_t of $\widetilde{\gamma}(t)$, 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 (4.79) 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 4.57. *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 $a \in \mathcal{C}^\infty(M)$ 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 (4.67) and (4.68). 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.

Hence, for every fixed compact $K \subset M$ containing the curve γ, by continuity there exists $t_0 = t_0(K)$ such that the restriction on K of the map $\pi\big|_{\mathcal{L}_t}$ is also a diffeomorphism, for all $0 \le t < t_0$. Let us now denote δ_K the positive constant defined in Lemma 3.33 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 4.56 (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 4.55). ☐

Remark 4.58. 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 $[0,\varepsilon]$.

As we prove later, this is false as soon as $\mathcal{D}_{q_0} = T_{q_0}M$, see Theorem 10.21.

Bibliographical notes

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.

5.1 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

$$\begin{aligned} &h_i : T^*M \to \mathbb{R}, \qquad i = 1, \dots, n, \\ &\{h_i, h_j\} = 0, \qquad \forall\, i, j = 1, \dots, n. \end{aligned} \tag{5.1}$$

such that the differentials $d_\lambda h_1, \dots, d_\lambda h_n$ of the functions are independent at every point $\lambda \in T^*M$.

Definition 5.1. Under the assumptions (5.1), the Hamiltonian system defined by one of the Hamiltonian $h_i, i = 1, \dots, n$, is said to be *completely integrable*.

Let us consider the vector valued map, called *moment map*, defined by

$$h : T^*M \to \mathbb{R}^n, \qquad h = (h_1, \dots, h_n),$$

and let $c = (c_1, \dots, c_n) \in \mathbb{R}^n$ be a regular value of the map h.

Lemma 5.2. *The set $h^{-1}(c)$ is a n-dimensional submanifold in T^*M and we have*

$$T_\lambda h^{-1}(c) = \operatorname{span}\{\vec{h}_1(\lambda), \dots, \vec{h}_n(\lambda)\}, \qquad \forall\, \lambda \in h^{-1}(c). \tag{5.2}$$

Proof. Since c is a regular value of h, by Remark 2.44 the set $h^{-1}(c)$ is a submanifold of dimension n in T^*M. In particular $\dim T_\lambda h^{-1}(c) = n$. Moreover, by Exercise 2.14, 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 (5.2) it is then enough to show that these vector fields are linearly independent.

Recall that the differentials of the functions h_i are linearly independent on $h^{-1}(c)$, namely

$$d_\lambda h_1 \wedge \ldots \wedge d_\lambda h_n \neq 0, \quad \forall\, \lambda \in h^{-1}(c). \tag{5.3}$$

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 the vectors $\vec{h}_1(\lambda), \dots, \vec{h}_n(\lambda)$ are linearly independent, hence they form a basis for $T_\lambda h^{-1}(c)$. ☐

Remark 5.3. 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 5.4. *Assume that the vector fields $\vec{h}_i$ are complete and define the map*

$$\Psi : \mathbb{R}^n \to \mathrm{Diff}(N_c), \qquad \Psi(s_1, \dots, s_n) := e^{s_1 \vec{h}_1} \circ \dots \circ e^{s_n \vec{h}_n}\Big|_{N_c}. \tag{5.4}$$

The map Ψ defines a transitive action of $\mathbb{R}^n$ onto N_c. In particular N_c is diffeomorphic to $T^k \times \mathbb{R}^{n-k}$ for some $0 \le k \le n$, where T^k denotes the k-dimensional torus.

Proof. The complete integrability assumption together with Corollary 4.51 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 2.25, this is equivalent to

$$e^{t\vec{h}_i} \circ e^{\tau \vec{h}_j} = e^{\tau \vec{h}_j} \circ e^{t\vec{h}_i}, \qquad \forall\, t, \tau \in \mathbb{R}. \tag{5.5}$$

Since the vector fields are complete by assumption, we can compute for every $s, s' \in \mathbb{R}^n$

$$\begin{aligned}
\Psi(s + s') &= e^{(s_1 + s_1')\vec{h}_1} \circ \dots \circ e^{(s_n + s_n')\vec{h}_n} \\
&= e^{s_1 \vec{h}_1} \circ e^{s_1' \vec{h}_1} \circ \dots \circ e^{s_n \vec{h}_n} \circ e^{s_n' \vec{h}_n} \\
&= e^{s_1 \vec{h}_1} \circ \dots \circ e^{s_n \vec{h}_n} \circ e^{s_1' \vec{h}_1} \circ \dots \circ e^{s_n' \vec{h}_n} \qquad\qquad \text{(by (5.5))} \\
&= \Psi(s) \circ \Psi(s'),
\end{aligned}$$

which proves that Ψ is a group action. Moreover, for every point $\lambda \in N_c$, we can consider its orbit under the action of Ψ, namely

$$\Omega_\lambda = \{\Psi(s)\lambda \,|\, s \in \mathbb{R}^n\}.$$

Notice that, for every λ, this defines a smooth local diffeomorphism between $\mathbb{R}^n$ and Ω_λ. Indeed the partial derivatives

$$\frac{\partial \Psi}{\partial s_i}(\Psi(s)\lambda) = \vec{h}_i(\Psi(s)\lambda), \quad i = 1, \dots, n,$$

are linearly independent on the level set N_c. As a consequence the stabilizer S_λ of the point λ, i.e. the set

$$S_\lambda = \{s \in \mathbb{R}^n \,|\, \Psi(s)\lambda = \lambda\},$$

is a discrete subgroup of $\mathbb{R}^n$. Then the proof of Proposition 5.4 is completed by the next lemma.

Lemma 5.5. *Let G be a non trivial discrete subgroup of $\mathbb{R}^n$. Then there exist $k \in \mathbb{N}$ with $1 \leq k \leq n$ and $e_1, \ldots, e_k \in \mathbb{R}^n$ such that*

$$G = \left\{ \sum_{i=1}^{k} m_i e_i, \ m_i \in \mathbb{Z} \right\}.$$

Proof. We prove the claim by induction on the dimension n of the ambient space $\mathbb{R}^n$.

(i). Let $n = 1$. Since G is a discrete subgroup of $\mathbb{R}$, then there exists an element $e_1 \neq 0$ closest to the origin $0 \in \mathbb{R}$. We claim that $G = \mathbb{Z}e_1 = \{me_1, m \in \mathbb{Z}\}$. By contradiction assume that there exists an element $f \in G$ such that $me_1 < f < (m+1)e_1$ for some $m \in \mathbb{Z}$. Then $\bar{f} := f - me_1$ belong to G and is closer to the origin with respect to e_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 $e_1 \in G$, closest to the origin. Moreover one can prove that $G_1 := G \cap \mathbb{R}e_1$ is a subgroup and, as in part (i) of the proof, that

$$G_1 := G \cap \mathbb{R}e_1 = \mathbb{Z}e_1.$$

If $G = G_1$ then the theorem is proved with $k = 1$. Otherwise one can consider the quotient G/G_1.

Exercise 5.6. (i). Prove that there exists a nonzero element $e_2 \in G/G_1$ that minimize the distance to the line $\ell = \mathbb{R}e_1$ 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 5.6 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 $e_2, \ldots, e_k$ such that

$$G/G_1 = \left\{ \sum_{i=2}^{k} m_i e_i, \ m_i \in \mathbb{Z} \right\}. \qquad \square$$

$\square$

From Proposition 5.4 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 5.7. *If N_c is compact, then $N_c \simeq T^n$.*

Remark 5.8. On any level set $\lambda \in N_c$ the map $\Psi_\lambda : \mathbb{R}^n \to N_c$ defined by $\Psi_\lambda(s) = \Psi(s)\lambda$ defines coordinates $(s_1, \ldots, s_n)$ in a neighborhood of the point λ. In these coordinate set (defined on N_c) the Hamiltonian vector fields $\vec{h}_i$ are constant.

5.2 Arnold-Liouville theorem

In this section we consider the moment map of a completely integrable system

$$h : T^*M \to \mathbb{R}^n, \qquad h = (h_1, \ldots, h_n),$$

and we assume that for all values of $c \in \mathbb{R}$ the level set $h^{-1}(c)$ is a smooth compact and connected manifold. In particular $N_c \simeq T^n$ for all $c \in \mathbb{R}$ by Corollary 5.7

Fix $c \in \mathbb{R}$ and a point $\lambda_c \in N_c$. Let us consider the basis $e_1, \ldots, e_n$ in $\mathbb{R}^n$ given by Lemma 5.5 and denote by $(\theta_1, \ldots, \theta_n)$ the coordinates defined in $\mathbb{R}^n$ by the choice of this basis.

Since $\theta_1, \ldots, \theta_n$ are obtained by $(s_1, \ldots, s_n)$ by a linear change of coordinates on each level set, the vector fields $\vec{h}_i$ are constant in these coordinates (see Remark 5.8) and the basis $\partial_{\theta_1}, \ldots, \partial_{\theta_n}$ can be expressed as follows

$$\partial_{\theta_i} = \sum_{j=1}^{n} b_{ij}(c)\vec{h}_j, \tag{5.6}$$

where the coefficients b_{ij} depend only on c, i.e., are constant on each level N_c.

Remark 5.9. Notice that the coordinate set $(\theta_1, \ldots, \theta_n)$ are not uniquely defined. Indeed every transformation of the kind $\theta_i \mapsto \theta_i + \psi_i(c)$ still defines a set of angular 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).

Notice that the vector fields ∂_{θ_i} are well defined and independent on this choice.

Let us now introduce the diffeomorphism

$$F_c : T^n \to N_c, \qquad F_c(\theta_1, \ldots, \theta_n) = \Psi(\theta_1 + 2\pi\mathbb{Z}, \ldots, \theta_n + 2\pi\mathbb{Z})(\lambda_c).$$

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 we have a well defined coordinate set $(c_1, \ldots, c_n, \theta_1, \ldots, \theta_n)$.

Theorem 5.10 (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*

$$(c_1, \ldots, c_n, \theta_1, \ldots, \theta_n) \mapsto (I_1, \ldots, I_n, \varphi_1, \ldots, \varphi_n) \tag{5.7}$$

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 5.11. The coordinates (I, φ) defined in Theorem 5.10 are called *action-angle* coordinates.

Remark 5.12. 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 in the angle-action coordinate (I,φ) is written as follows

$$\dot{I}_i = -\frac{\partial h_j}{\partial \varphi_i} = 0, \qquad \dot{\varphi}_i = \frac{\partial h_j}{\partial I_i}(I). \tag{5.8}$$

This explains also why this property is called *complete integrability.*

Proof of Theorem 5.10. In this proof we will use the following notation:

- if $c = (c_1, \dots, c_n) \in \mathbb{R}^n$ we set $c^{j,\varepsilon} = (c_1, \dots, c_j + \varepsilon, \dots, c_n)$,
- $\gamma_i(c)$ is the closed curve in the torus N_c parametrized by the i-th angular coordinate θ_i, namely
$$\gamma_i(c) = \{F_c(\theta_1, \dots, \theta_i + \tau, \dots, \theta_n) \in N_c \,|\, \tau \in [0, 2\pi]\}.$$
- $C_i^{j,\varepsilon}$ denotes the cylinder defined by the union of curves $\gamma_i(c^{j,\tau})$, for $0 \le \tau \le \varepsilon$.

Let us first define the coordinates $I_i = I_i(c_1, \dots, c_n)$ by the formula

$$I_i(c) = \frac{1}{2\pi} \int_{\gamma_i(c)} s,$$

where s is the tautological 1-form on T^*M. Being $\sigma|_{N_c} \equiv 0$, by Stokes Theorem the variable I_i depends only on the homotopy class of γ_i.[1]

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 (5.6)) one gets

$$\sigma(\cdot, \partial_{\theta_i}) = \sum_{j=1}^n b_{ij}(c) dh_j. \tag{5.9}$$

[1] Hence, in principle, we are free to choose any basis $\gamma_1, \dots, \gamma_n$ for the first homotopy group of T^n.

Moreover $dh_i = dc_i$ since they define the same coordinate set. Hence

$$\begin{aligned}
\frac{\partial I_i}{\partial c_j}(c) &= \frac{1}{2\pi}\int_{\gamma_i(c)} \left\langle \sum_{k=1}^{n} b_{ik} dc_k, \partial_{c_i} \right\rangle d\theta_i \\
&= \frac{1}{2\pi}\int_{\gamma_i(c)} b_{ij}(c) d\theta_i \\
&= b_{ij}(c)
\end{aligned}$$

Combining the last identity with (5.9) 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)

$$\sigma = \sum_{ij} a_{ij}(I) dI_i \wedge dI_j + \sum_i dI_i \wedge d\theta_i. \tag{5.10}$$

where the smooth functions a_{ij} depends only on the action variables, since the symplectic form σ and the term $\sum_i dI_i \wedge d\theta_i$ are closed form. Moreover it is easy to see that the first term of (5.10) 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 defining $\varphi_i := \theta_i - \beta_i(I)$. ☐

Remark 5.13. The notion of complete integrability introduced here is the classical one given by Liouville and Arnold. Sometimes, complete integrability of a dynamical system is also referred to systems whose solution can be reduced to a sequence of quadratures. Notice that by Theorem 5.10 complete integrability implies integrability by quadratures (see also Remark 5.12).

5.3 Integrable geodesic flows

In this section we want to discuss whether it is possible to apply the Arnold-Lioville's Theorem to the case of a geodesic flow on a Riemannian (or sub-Riemannian) manifold.

Recall that on a sub-Riemannian manifold, we denote by H the sub-Riemannian Hamiltonian.

Definition 5.14. We say that a complete smooth vector field $X \in \text{Vec}(M)$ 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 $t \in \mathbb{R}$.

Recall that, for every $X \in \text{Vec}(M)$, we can define the function $h_X \in \mathcal{C}^\infty(T^*M)$ linear on fibers associated with X by $h_X(\lambda) = \langle \lambda, X(q) \rangle$, where $q = \pi(\lambda)$.

The following lemma shows that, if X is a Killing vector field, i.e. a vector field on M whose flow generates isometries, then the Hamiltonian associated with it is in involution with the sub-Riemannian Hamiltonian.

Lemma 5.15. *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 $e^{tX}_* : T_q M \to T_{e^{tX}(q)} M$ 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

$$H(e^{tX*}\lambda) = H(\lambda), \qquad \forall\, \lambda \in T^*M.$$

On the other hand Proposition 4.8 implies that $(e^{tX})^* = e^{t\vec{h}_X}$, where h_X is the hamiltonian linear on fibers related to X. Hence differentiating 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 to every 1-parametric group of isometries of M we can associate an Hamiltonian in involution with H. Let us show the complete integrability of the geodesic flow in some very symmetric cases.

Example 5.16 (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. As a consequence the geodesic flow is completely integrable.

Example 5.17 (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 4.5.2). The sub-Riemannian structure on $M \times \mathbb{R}$ is determined by the function $b \in \mathcal{C}^\infty(M)$ 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) If both M and b are rotational invariant we find a first integral of the geodesic flow as in the previous example

(ii) By construction the problem is invariant by translation along the z-axis

Hence there exists three Hamitonian in involution and the geodesic flow is completely integrable.

5.3.1 Geodesic flow

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 $\mathbb{R}^n$. By 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 $\mathbb{R}^n$).

Adapting the arguments of Proposition 1.4 in Chapter 1, one can prove the following characterization of geodesics on a hypersurface M.

Proposition 5.18. *Let $\gamma : [0,1] \to M$ a lenght-parametrized curve on M. Then γ is a geodesic if and only if $\ddot{\gamma}(t) \perp T_{\gamma(t)}M$.*

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 $T^*\mathbb{R}^n$

$$T^*\mathbb{R}^n = \mathbb{R}^n \times \mathbb{R}^n = \{(x,p),\ x,p \in \mathbb{R}^n\}, \qquad \sigma = \sum_{i=1}^n dp_i \wedge dx_i,$$

For $x,p \in \mathbb{R}^n$ we will denote by $x + \mathbb{R}p$ the line of $\mathbb{R}^n$ $\{x + tp, t \in \mathbb{R}\}$.

Assumptions. In what follows we assume that the function $a : \mathbb{R}^n \to \mathbb{R}$ satisfies the following assumptions:

(i) the restriction of $a : \mathbb{R}^n \to \mathbb{R}$ to *every line* is strictly convex,

(ii) $a(x) \to +\infty$ when $|x| \to +\infty$.

Under these assumptions the restriction of the function a to each affine line in $\mathbb{R}^n$ always attains a minimum and we can define the function

$$h(x,p) = \min_{t \in \mathbb{R}} a(x + tp). \tag{5.11}$$

Remark 5.19. Given $x,p \in \mathbb{R}^n$ the line $x + \mathbb{R}p$ is tangent to the level set $a^{-1}(c)$ (with $c = a(x + \bar{t}p)$) at the point $\xi = x + \bar{t}p \in \mathbb{R}^n$ at which the minimum in (5.11) is attained. Indeed

$$0 = \left.\frac{d}{dt}\right|_{t=\bar{t}} a(x + tp) = \langle d_\xi a, p \rangle .$$

It is clear from the definition of h that actually it is a well-defined function on the space of affine lines in $\mathbb{R}^n$. This is formally proved in the following lemma.

Lemma 5.20. *The Hamiltonian $b(x,p) = \frac{1}{2}|p|^2$ satisfies $\{h,b\} = 0$, i.e. h it is constant along the flow of $\vec{b}$.*

Proof. The Hamiltonian system for $\vec{b}$ is easily solved for every initial condition $(x(0),p(0)) = (x_0,p_0)$

$$\begin{cases} \dot{x} = \frac{\partial b}{\partial p} = p \\ \dot{p} = -\frac{\partial b}{\partial x} = 0 \end{cases} \qquad \Rightarrow \qquad \begin{cases} x = x_0 + tp_0 \\ p = p_0 \end{cases} \tag{5.12}$$

and it is easy to see that, by its very definition, h is constant under this flow. □

Remark 5.21. Notice that to restrict to a level set of b is equivalent to restrict the function h to the space of affine lines in $\mathbb{R}^n$ since

$$\{(x,p) \in T^*\mathbb{R}^n,\ b(x,p) = 1/2\} = \{(x,p) \in T^*\mathbb{R}^n,\ |p| = 1\}.$$

Now we introduce the following function

$$\xi : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n, \qquad \xi(x,p) = x + s(x,p)p, \tag{5.13}$$

where $s(x,p) = \bar{t}$ is the point at which the function $f(t) = a(x + tp)$ attains its minimum.

The following proposition says that if we follow the flow of $\vec{h}$, as a flow on the space of lines, then the line is always tangent to the same quadric and actually describes a geodesic on it.

Proposition 5.22. *Let $(x(t), p(t))$ be a trajectory of the Hamiltonian vector field $\vec{h}$ associated to (5.11). Then the function*

$$t \mapsto \xi(t) := \xi(x(t), p(t)) \in \mathbb{R}^n, \tag{5.14}$$

(i) is contained in a 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 4.18, since every function is constant along the flow of its Hamiltonian vector field. Indeed, writing $h(x,p) = a(\xi(x,p))$ and denoting by $(x(t), p(t))$ the Hamiltonian flow, we get

$$a(\xi(t)) = a(\xi(x(t), p(t))) = h(x(t), p(t)) = \text{const},$$

i.e. the curve $\xi(t)$ is contained on a level set of a. Moreover by definition $s(x,p)$ denotes on the line $x + \mathbb{R}p$ where a attains its minimum, hence

$$\langle \nabla_{\xi(t)} a, p(t) \rangle = 0, \qquad \forall\, t. \tag{5.15}$$

The Hamiltonian system associated with h reads

$$\begin{cases} \dot{x} = s\nabla_\xi a \\ \dot{p} = -\nabla_\xi a \end{cases} \tag{5.16}$$

that immediately implies $\dot{x} + s\dot{p} = 0$. Computing the derivative

$$\dot{\xi} = \dot{x} + \dot{s}p + s\dot{p} = \dot{s}p,$$

it follows that $\dot{\xi}$ is parallel to p, and actually $p(t)$ is the velocity of the curve $\xi(t)$, when reparametrized with the parameter s, since $|p| = 1$ implies $|\dot{\xi}| = \dot{s}$.

Finally, the second derivative of the reparametrized of ξ is $\dot{p}$ and, since $\dot{p} \wedge \nabla_\xi a = 0$ from the Hamiltonian system, the second derivative of $\xi(t)$ (when reparametrized by the length) is orthogonal to the level set, i.e. $\xi(t)$ is a geodesic.

Notice also that s is a well defined parameter. Computing the derivative with respect to t in (5.15) we have that

$$\dot{s}\langle \nabla_\xi^2 a\, p, p\rangle - |\nabla_\xi a|^2 = 0.$$

and the strict convexity of a implies $\langle \nabla_\xi^2 a\, p, p\rangle \neq 0$. □

Remark 5.23. 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.

Consider now two functions $a, b : \mathbb{R}^n \to \mathbb{R}$ that satisfies our assumptions $(i), (ii)$. Following our notation, we set

$$
\begin{aligned}
h(x,p) &= a(\xi(x,p)), \qquad & \xi(x,p) &= x + s(x,p)p \\
g(x,p) &= b(\eta(x,p)), \qquad & \eta(x,p) &= x + t(x,p)p
\end{aligned}
$$

where $s(x,p)$ and $t(x,p)$ 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 5.24. *Under the previous assumptions*

$$
\{h, g\} = (s - t) \langle \nabla_\xi a, \nabla_\eta b \rangle . \tag{5.17}
$$

Proof. From the very definition of Poisson bracket

$$
\begin{aligned}
\{h, g\} &= \langle \nabla_p h, \nabla_x g \rangle - \langle \nabla_x h, \nabla_p g \rangle \\
&= (s - t) \langle \nabla_\xi a, \nabla_\eta b \rangle .
\end{aligned}
$$

where we used equations (5.16) for both h and g. $\square$

5.4 Geodesic flow on ellipsoids

It was Jacobi who first established that the geodesic flow on an ellipsoid is completely integrable, using the separation of variables method. Here we give a different derivation, essentially due to Moser, as an application of the theory developed in the previous section. 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 $\mathbb{R}^n$.

Definition 5.25. Let A be an $n \times n$ non degenerate symmetrix matrix. The *quadric* $\mathcal{Q}$ associated to A is the set

$$
\mathcal{Q} = \{x \in \mathbb{R}^n,\ \langle A^{-1}x, x \rangle = 1\}. \tag{5.18}
$$

For simplicity we consider 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 (5.18) 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 $\mathcal{Q}$

$$
\mathcal{Q}_\lambda = \{x \in \mathbb{R}^n,\ \sum_{i=1}^n \frac{x_i^2}{\alpha_i - \lambda} = 1\}, \qquad \forall\, \lambda \in \mathbb{R} \setminus \Lambda,
$$

where $\Lambda = \{\alpha_1, \ldots, \alpha_n\}$ denotes the set of eigenvalues of A. Note that $\mathcal{Q}_\lambda = \emptyset$ 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.

Exercise 5.26. 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 5.27. *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 $\lambda_1, \dots, \lambda_n$ in $\mathbb{R}\setminus\Lambda$ such that $x \in \mathcal{Q}_{\lambda_i}$ for every $i = 1, \dots, n$, and 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 5.26 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 $\lambda_1, \dots, \lambda_n$ such that $a_{\lambda_i}(x) = 1$ (that means $x \in \mathcal{Q}_{\lambda_i}$). Next, using part (ii) of Exercise 5.26 (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. $\square$

Now we define the family of Hamiltonians associated with the family of confocal quadrics

$$h_\lambda(x,p) = \min_t a_\lambda(x+tp) = a_\lambda(\xi_\lambda(x,p)), \tag{5.19}$$

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 5.28. *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

$$\langle A_\lambda \xi_\lambda, y\rangle = 0, \qquad \langle A_\lambda \xi_\lambda, \xi_\lambda\rangle = 1 \tag{5.20}$$

and an analogue relations is valid for $\mathcal{Q}_\mu$. Notice than from (5.20) 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 (5.26)

$$\begin{aligned}
\left\langle \nabla_{\xi_\lambda} a_\lambda, \nabla_{\xi_\mu} a_\mu\right\rangle &= 4\langle A_\lambda \xi_\lambda, A_\mu \xi_\mu\rangle \\
&= 4\langle A_\lambda A_\mu \xi_\lambda, \xi_\mu\rangle \\
&= \frac{4}{\mu - \lambda}(\langle A_\lambda \xi_\lambda, \xi_\mu\rangle - \langle A_\mu \xi_\mu, \xi_\lambda\rangle) = 0,
\end{aligned}$$

The last claim follows from Proposition (5.24). $\square$

Proposition 5.29. *A generic line in $\mathbb{R}^n$ is tangent to $n-1$ quadrics of a confocal family.*

Proof. Consider the projection along the fixed line $x + \mathbb{R}p$ of the quadrics of the confocal family onto an orthogonal hyperplane. The following exercise shows that this projection defines a confocal family of quadrics on the reduced space.

Exercise 5.30. (i). Show that the map $x \mapsto a_\lambda^p(x) := \langle A_\lambda(x + t_\lambda p), x + t_\lambda p\rangle$ is a quadratic form and that $p \in \operatorname{Ker} a_\lambda^p$. In particular this implies that a_λ^p is well defined on the quotient $\mathbb{R}^n/\mathbb{R}p$.
(ii). Prove that $\{a_\lambda^p\}_\lambda$ is a family of confocal quadric on the factor space (in $n-1$ variables).

Applying then Lemma 5.27 to the family $\{a_\lambda^p\}_\lambda$ we get that, for a generic choice of x, there exists $n-1$ quadrics passing through the point on the plane where the line is projected, i.e. the line $x + \mathbb{R}p$ is tangent to $n-1$ confocal quadrics of the family $\{a_\lambda\}_\lambda$. □

Remark 5.31. 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 5.29 mutually orthogonal and orthogonal to the line itself.

Theorem 5.32. *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 4.37, 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 5.28.

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. □

Chapter 6

Chronological calculus

In this chapter we develop some tools from chronological caluculs that will allow us to manage in a very efficient way with flows of nonautonomous vector fields.

The main idea is to replace a nonlinear object defined on the manifold M with its linear counterpart, when interpreted as an operator on the space $\mathcal{C}^\infty(M)$ of smooth functions on M.

6.1 Duality

We recall that the set $\mathcal{C}^\infty(M)$ of smooth functions on M is an $\mathbb{R}$-algebra with the usual operation of pointwise addition and multiplication

$$\begin{aligned} (a+b)(q) &= a(q)+b(q), \\ (\lambda a)(q) &= \lambda\, a(q), \qquad\qquad a,b \in \mathcal{C}^\infty(M),\ \lambda \in \mathbb{R}, \\ (a\cdot b)(q) &= a(q)b(q). \end{aligned}$$

Any point $q \in M$ can be interpreted as the linear functional

$$\widehat{q} : \mathcal{C}^\infty(M) \to \mathbb{R}, \qquad \widehat{q}(a) := a(q).$$

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 linear "change of variables" operator

$$\widehat{P} : \mathcal{C}^\infty(M) \to \mathcal{C}^\infty(M), \qquad \widehat{P}(a) := a(P(q)).$$

which is an automorphism of the algebra $\mathcal{C}^\infty(M)$.

Remark 6.1. Notice that every nontrivial homomorphism of algebras $\varphi : \mathcal{C}^\infty(M) \to \mathbb{R}$ is represented by some point, i.e. $\varphi = \widehat{q}$ for some $q \in M$. Moreover for every automorphism of algebras $\Phi : \mathcal{C}^\infty(M) \to \mathcal{C}^\infty(M)$ there exists a diffeomorphism $P \in \mathrm{Diff}(M)$ such that $\widehat{P} = \Phi$.

Now we want to characterize tangent vectors as functionals on $\mathcal{C}^\infty(M)$. As remarked 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

$$\widehat{v} : \mathcal{C}^\infty(M) \to \mathbb{R}, \qquad \widehat{v}(a) = \langle d_q a, v\rangle\,,$$

that satisfies the Leibnitz rule

$$\widehat{v}(a \cdot b) = \widehat{v}(a)b(q) + a(q)\widehat{v}(b), \qquad \forall\, a, b \in \mathcal{C}^\infty(M).$$

On the other hand, considering $v \in T_qM$ as the tangent vector of a curve $q(t)$ such that $q(0) = q$, it is also natural to consider the family of functionals $\widehat{q}(t) := \widehat{q(t)}$, and define

$$\widehat{v} := \frac{d}{dt}\bigg|_{t=0} \widehat{q}(t) : \mathcal{C}^\infty(M) \to \mathbb{R}. \tag{6.1}$$

It is easy to check that (6.1) agrees with our definition of $\widehat{v}$, it is sufficient to differentiate at $t = 0$ the identity

$$\widehat{q}(t)(a \cdot b) = \widehat{q}(t)a \cdot \widehat{q}(t)b$$

In the same spirit, a vector field $X \in \mathrm{Vec}(M)$ will be characterized, as a derivation of $\mathcal{C}^\infty(M)$ (for vector fields we already discussed this property in Chapter 2), as the infinitesimal version of a flow (i.e. family of diffeomorphisms) $P_t \in \mathrm{Diff}(M)$. Indeed if we set

$$\widehat{X} = \frac{d}{dt}\bigg|_{t=0} \widehat{P}_t : \mathcal{C}^\infty(M) \to \mathcal{C}^\infty(M),$$

we find that $\widehat{X}$ satisfies (see (2.19))

$$\widehat{X}(ab) = \widehat{X}(a)b + a\widehat{X}(b), \qquad \forall\, a, b \in \mathcal{C}^\infty(M).$$

Remark 6.2. It is possible to define on $\mathcal{C}^\infty(M)$ the Whitney topology and define regularity properties of family of functionals in a weak sense, i.e. we say that A_t is continuos (differentiable, etc.) if the map $t \mapsto A_t a$ has the same property for every $a \in \mathcal{C}^\infty(M)$. For instance, if X_t denotes some locally integrable family of vector fields we denote

$$\int_0^t X_s\, ds : a \mapsto \int_0^t X_s a\, ds$$

For a more detailed presentation see [4]. [1]

6.2 Operator ODE and Taylor expansion

Consider a nonautonomous vector field X_t and the correspondent nonautonomous ODE

$$\frac{d}{dt}q(t) = X_t(q(t)), \qquad q \in M. \tag{6.2}$$

Using the notation exploited in the previous section we can rewrite (6.2) in the following way

$$\frac{d}{dt}\widehat{q}(t) = \widehat{q}(t) \circ \widehat{X}_t. \tag{6.3}$$

[1] With this interpretation it makes sense to consider, for instance, the sum of a point q and a vector v

$$q + v : a \mapsto a(q) + \langle d_q a, v\rangle$$

Indeed assume that $q(t)$ satisfies (6.2) and let $a \in \mathcal{C}^\infty(M)$. We compute

$$\begin{aligned}
\left(\frac{d}{dt}\widehat{q}(t)\right) a &= \frac{d}{dt}\widehat{q}(t)a = \frac{d}{dt}a(q(t)) \\
&= \left\langle d_{q(t)}a, X_t(q(t))\right\rangle = (\widehat{X}_t a)(q(t)) \\
&= (\widehat{q}(t) \circ \widehat{X}_t)a
\end{aligned} \tag{6.4}$$

We discussed in Chapter 2 that, considering the solution to the nonautonomous ODE (6.2) , we have a well defined flow, i.e. family of diffeomorphisms, $P_t : q_0 \mapsto q(t)$. We call P_t the *right chronological exponential* and use the notation

$$P_t := \overrightarrow{\exp} \int_0^t X_s ds. \tag{6.5}$$

Lemma 6.3. *The flow P_t defined by* (6.5) *satisfies the differential equation*

$$\frac{d}{dt}\widehat{P}_t = \widehat{P}_t \circ \widehat{X}_t, \qquad \widehat{P}_0 = \mathrm{Id}. \tag{6.6}$$

Proof. Fix a point $q_0 \in M$ and denote by $q(t)$ the solution of the Cauchy problem (6.2) with initial condition $q(0) = q_0$. By the very definition of P_t we have that $q(t) = P_t(q_0)$, which easily implies $\widehat{q}(t) = \widehat{q}_0 \circ \widehat{P}_t$. □

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. The meaning of the notation will be clear from the context. Notice that there is no risk of confusion since, when using operatorial notation, composition works in the opposite side.

Our differential equation (6.6), namely

$$\begin{cases} \dot{P}_t = P_t \circ X_t \\ P_0 = \mathrm{Id} \end{cases} \tag{6.7}$$

can be rewritten as an integral equation as follows

$$P_t = \mathrm{Id} + \int_0^t P_s \circ X_s ds \tag{6.8}$$

Substituting into (6.8), and iterating we have

$$\begin{aligned}
P_t &= \mathrm{Id} + \int_0^t \left(\mathrm{Id} + \int_0^{s_1} P_{s_2} \circ X_{s_2} ds_2\right) \circ X_{s_1} ds_1 \\
&= \mathrm{Id} + \int_0^t X_s ds + \iint\limits_{0 \le s_2 \le s_1 \le t} P_{s_2} \circ X_{s_2} \circ X_{s_1} ds_1 ds_2 \\
&= \dots \\
&= \mathrm{Id} + \sum_{k=1}^N \int \cdots \int\limits_{0 \le s_k \le \ldots \le s_1 \le t} X_{s_k} \circ \cdots \circ X_{s_1} d^k s + R_N
\end{aligned}$$

where

$$R_N = \int \cdots \int_{0 \le s_N \le \ldots \le s_1 \le t} P_{s_N} \circ X_{s_N} \circ \cdots \circ X_{s_1} d^N s$$

Formally, letting $N \to \infty$, we can write the chronological series

$$\overrightarrow{\exp} \int_0^t X_s ds \approx \mathrm{Id} + \sum_{k=1}^{\infty} \int \cdots \int_{S_k(t)} X_{s_k} \circ \cdots \circ X_{s_1} d^k s \tag{6.9}$$

where $S_k(t) = \{(s_1, \ldots, s_k) \in \mathbb{R}^k | \, 0 \le s_k \le \ldots \le s_1 \le t\}$ denotes the k-dimensional symplex.

Remark 6.4. If we write expansion (6.9) when $X_t = X$ is an autonomous vector field, we find

$$e^{tX} = \overrightarrow{\exp} \int_0^t X ds \approx \mathrm{Id} + \sum_{k=1}^{\infty} \int \cdots \int_{S_k(t)} \underbrace{X \circ \cdots \circ X}_{k} d^k s = \sum_{k=0}^{\infty} \frac{t^k}{k!} X^k.$$

since $\mathrm{meas}(S_k(t)) = t^k/k!$. This also shows that in the nonautonomous case the order in which $s_1, \ldots, s_k$ are presented in the composition is very important. The key point is that for different time X_s and X_τ might not commute.

Remark 6.5. Notice that the chronological exponential cannot be written as the flow of an autonomous vector field

$$\overrightarrow{\exp} \int_0^t X_s ds \neq e^{\int_0^t X_s ds}.$$

One can show that a necessary condition for the equality holds is $[X_t, X_\tau] = 0$ for all t, τ.

Consider now the inverse flow $Q_t := P_t^{-1}$, where P_t satisfies (6.8), and try to characterize the differential equation satisfied by Q_t. First we differentiate the identity

$$P_t \circ Q_t = \mathrm{Id} \tag{6.10}$$

and Leibnitz rule give

$$\dot{P}_t \circ Q_t + P_t \circ \dot{Q}_t = 0$$

Using (6.7) then we get

$$P_t \circ X_t \circ Q_t + P_t \circ \dot{Q}_t = 0$$

hence we get, multiplying Q_t both sides, that Q_t satisfies

$$\begin{cases} \dot{Q}_t = -X_t \circ Q_t, \\ Q_0 = \mathrm{Id}. \end{cases} \tag{6.11}$$

which is dual to the Cauchy problem (6.7).

The solution to the problem (6.11) will be denoted by the *left chronological exponential*

$$Q_t := \overleftarrow{\exp} \int_0^t (-X_s) ds. \tag{6.12}$$

Repeating analogous reasoning, we find the formal expansion

$$\overleftarrow{\exp}\int_0^t (-X_s)ds \approx \mathrm{Id} + \sum_{k=1}^{\infty} \int\cdots\int_{0\le s_k\le\ldots\le s_1\le t} (-X_{s_1})\circ\cdots\circ(-X_{s_k})d^k s.$$

The difference with respect to the right chronological exponential is in the order of composition. In particular the arrow over the exp says in which direction the time increases.

We can summarize properties of the chronological exponential into the following

$$\frac{d}{dt}\overrightarrow{\exp}\int_0^t X_s ds = \overrightarrow{\exp}\int_0^t X_s ds \circ X_t, \tag{6.13}$$

$$\frac{d}{dt}\overleftarrow{\exp}\int_0^t X_s ds = X_t \circ \overleftarrow{\exp}\int_0^t X_s ds, \tag{6.14}$$

$$\left(\overrightarrow{\exp}\int_0^t X_s ds\right)^{-1} = \overleftarrow{\exp}\int_0^t (-X_s)ds. \tag{6.15}$$

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 $\mathcal{C}^\infty(M)$, we have

$$P_* v = v \circ P.$$

Indeed consider a curve $q(t)$ such that $\dot{q}(0) = v$ and compute

$$(P_*v)a = \left.\frac{d}{dt}\right|_{t=0} a(P(q(t))) = \left(\left.\frac{d}{dt}\right|_{t=0} q(t)\right)\circ Pa = v\circ Pa$$

Recall that, if $X \in \mathrm{Vec}(M)$ is a vector field we have $P_*X\big|_q = P_*(X\big|_{P^{-1}(q)})$. In a similar way we will find an expression for P_*X as derivation of $\mathcal{C}^\infty(M)$

$$P_*X = P^{-1}\circ X\circ P. \tag{6.16}$$

Remark 6.6. We can reinterpret the pushforward of a vector field in a totally algebraic way in the space of linear operator on $\mathcal{C}^\infty(M)$. Indeed

$$P_*X = (\mathrm{Ad}\, P^{-1})X,$$

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[2].

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

$$\begin{aligned}\left(\frac{d}{dt}\mathrm{Ad}\, P_t\right)Y &= \frac{d}{dt}(\mathrm{Ad}\, P_t)Y = \frac{d}{dt}(P_t\circ Y\circ P_t^{-1})\\ &= P_t\circ X_t\circ Y\circ P_t^{-1} + P_t\circ Y\circ(-X_t)\circ P_t^{-1}\\ &= P_t\circ(X_t\circ Y - Y\circ X_t)\circ P_t^{-1}\\ &= (\mathrm{Ad}\, P_t)[X_t, Y]\\ &= (\mathrm{Ad}\, P_t)(\mathrm{ad}\, X_t)Y\end{aligned}$$

[2]it is the differential of the conjugation $Q \mapsto P\circ Q\circ P^{-1}$, $Q \in \mathrm{Diff}(M)$

where

$$\operatorname{ad} X : Y \mapsto [X, Y],$$

is the adjoint action on the Lie algebra of vector fields.

In other words we proved that $\operatorname{Ad} P_t$ is a solution to the differential equation

$$\dot{A}_t = A_t \circ \operatorname{ad} X_t, \qquad A_0 = \mathrm{Id}.$$

Thus it can be expressed as chronological exponential and we have the identity

$$\operatorname{Ad}\left(\overrightarrow{\exp}\int_0^t X_s ds\right) = \overrightarrow{\exp}\int_0^t \operatorname{ad} X_s ds. \tag{6.17}$$

Exercise 6.7. Prove that, if $[X_t, Y] = 0$ for all t, then $(\operatorname{Ad} P_t)Y = Y$.

Remark 6.8. More explicitly we can write the following formula

$$(\operatorname{Ad} P_t)Y \simeq Y + \sum_{k=1}^{\infty} \int_{0 \leq s_k \leq \ldots \leq s_1 \leq t} \cdots \int [X_{s_n}, \ldots, [X_{s_2}, [X_{s_1}, Y]] d^k s, \tag{6.18}$$

which generalizes the formula (2.30). Indeed if $P_t = e^{tX}$ is the flow associated to an autonomous vector field we get

$$\begin{aligned}(\operatorname{Ad} e^{tX})Y \simeq e_*^{-tX} Y &= Y + \sum_{k=1}^{\infty} \frac{t^k}{k!}[X, \ldots, [X, Y]] \\ &= Y + t[X, Y] + \frac{t^2}{2}[X, [X, Y]] + \ldots\end{aligned}$$

Exercise 6.9. 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 $X \in \operatorname{Vec}(M)$ then
$$\operatorname{ad} X = \left.\frac{d}{dt}\right|_{t=0} \operatorname{Ad} P_t.$$
2. Show that, if $P \in \operatorname{Diff}(M)$, then P_* preserves Lie brackets, i.e. $P_*[X, Y] = [P_* X, P_* Y]$.
3. Show that the Jacobi identity in $\operatorname{Vec}(M)$ is the infinitesimal version of the identity proved in 2. (Hint. use $P_t = e^{tZ}$)

Exercise 6.10. Prove the following formula on the change of variables on a nonautonomous flow

$$P \circ \overrightarrow{\exp}\int_0^t X_s ds \circ P^{-1} = \overrightarrow{\exp}\int_0^t (\operatorname{Ad} P) X_s ds. \tag{6.19}$$

Notice that for an autonomous vector field it proves (2.24).

6.3 Variations Formulae

Consider the following ODE

$$\dot{q} = X_t(q) + Y_t(q) \tag{6.20}$$

where Y_t is some perturbation of our original equation (6.2). We want to describe the solution to the perturbed equation (6.20) as the perturbation of the solution of the unperturbed one.

Proposition 6.11. *Let X_t, Y_t be two nonautonomous vector fields. Then*

$$\overrightarrow{\exp}\int_0^t (X_s + Y_s)ds = \overrightarrow{\exp}\int_0^t \left(\overrightarrow{\exp}\int_0^s \operatorname{ad} X_\tau d\tau\right) Y_s ds \circ \overrightarrow{\exp}\int_0^t X_s ds \tag{6.21}$$

$$= \overrightarrow{\exp}\int_0^t (\operatorname{Ad} P_s) Y_s ds \circ P_t \tag{6.22}$$

where $P_t = \overrightarrow{\exp}\int_0^t X_s ds$ denote the flow of the original vector field.

Proof. Our goal is to find a flow R_t such that

$$Q_t := \overrightarrow{\exp}\int_0^t (X_s + Y_s)ds = R_t \circ P_t \tag{6.23}$$

By definition of right chronological exponential we have

$$\dot{Q}_t = Q_t \circ (X_t + Y_t) \tag{6.24}$$

On the other hand, from (6.23), we also find

$$\begin{aligned} \dot{Q}_t &= \dot{R}_t \circ P_t + R_t \circ \dot{P}_t \\ &= \dot{R}_t \circ P_t + R_t \circ P_t \circ X_t \\ &= \dot{R}_t \circ P_t + Q_t \circ X_t \end{aligned} \tag{6.25}$$

Hence, comparing (6.24) and (6.25), we get

$$Q_t \circ Y_t = \dot{R}_t \circ P_t$$

and we can write the ODE satisfied by R_t

$$\begin{aligned} \dot{R}_t &= Q_t \circ Y_t \circ P_t^{-1} \\ &= R_t \circ (\operatorname{Ad} P_t) Y_t \end{aligned}$$

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 (6.22). Using (6.17) we get (6.21) $\square$

Exercise 6.12. Prove the following

(i) Prove the second form of the variational formula, where the original flow appear to the left

$$\overrightarrow{\exp} \int_0^t (X_s + Y_s) ds = \overrightarrow{\exp} \int_0^t X_s ds \circ \overrightarrow{\exp} \int_0^t \left(\overrightarrow{\exp} \int_t^s \operatorname{ad} X_\tau d\tau \right) Y_s ds \tag{6.26}$$

(ii) For autonomous vector fields $X, Y \in \operatorname{Vec}(M)$ prove that

$$e^{t(X+Y)} = \overrightarrow{\exp} \int_0^t e^{s \operatorname{ad} X} Y ds \circ e^{tX} = \overrightarrow{\exp} \int_0^t e_*^{-sX} Y ds \circ e^{tX} \tag{6.27}$$

$$= e^{tX} \circ \overrightarrow{\exp} \int_0^t e^{(s-t) \operatorname{ad} X} Y ds \tag{6.28}$$

Chapter 7

End-point and Exponential map

7.1 First order conditions

In this section we introduce the end-point map (i.e. the map which associates to every control the final point of the associate trajectory) and we interpret the optimality condition obtained above as a Lagrange multipliers rule.

We start by defining the end-point map. Consider a smooth n-dimensional manifold M and a trivializable sub-Riemannian structure on it

$$\mathbf{U} = M \times \mathbb{R}^k, \qquad f : \mathbf{U} \to TM, \qquad (q, u) \mapsto f_u(q) = \sum_{i=1}^{k} u_i f_i(q)$$

For every measurable square integrable control function $t \mapsto u(t) \in L^2$ there exists an admissible trajectory $\gamma(t, u(\cdot))$ solution to the Cauchy problem

$$\dot{\gamma}(t) = f_{u(t)}(\gamma(t)), \qquad \gamma(0, u(\cdot)) = q_0.$$

In the sequel we fix the initial point $q_0 \in M$ and we consider the set of *admissible controls*

$$\mathcal{U} = \{u \in L^2([0,1], \mathbb{R}^k),\ \gamma(t, u(\cdot)) \text{ is defined for } t = 1\} \subset L^2$$

which is an open subset by ODE's continuous dependence theorem. Notice that the choice of L^2 will change topology in the space of controls but nothing change in our geometric space because we know that we can always consider length parametrized curves.

Definition 7.1. In the previous hypothesis we define the *end-point map*

$$F : \mathcal{U} \to M, \qquad F(u(\cdot)) := \gamma(1, u(\cdot)) = q_0 \circ \overrightarrow{\exp} \int_0^1 f_{u(t)} dt. \tag{7.1}$$

The end-point map is a map from an open set of an Hilbert space in a smooth finite-dimensional manifold. Using chronological calculus developed in Chapter **??**, it is easy to compute its (Frechét) differential.

Proposition 7.2. *Let $u \in \mathcal{U}$ and $q_1 = F(u(\cdot))$. The end-point map F is smooth and its differential at u is the map*

$$D_u F : L^2 \to T_{q_1} M, \qquad D_u F(v) = \int_0^1 P^1_{t*} f_{v(t)}(q_1) dt. \tag{7.2}$$

where $P^t_\tau : \gamma_u(\tau) \mapsto \gamma_u(t)$ is the flow generated by u.

Proof. The end-point map from q_0 can be rewritten as the chronological exponential

$$F(u(\cdot)) = q_0 \circ \overrightarrow{\exp} \int_0^1 f_{u(t)} dt.$$

Using the Volterra expansion (6.9) we can immediately compute the differential of the end-point map near $u \equiv 0$. Indeed we have that, for any control $v(\cdot)$ sufficiently close to 0:

$$F(v(\cdot)) = q_0 \circ \left(\mathrm{Id} + \int_0^1 f_{v(t)} dt + \iint\limits_{0 \le t_2 \le t_1 \le t} f_{v(t_2)} \circ f_{v(t_1)} ds_1 ds_2 + \dots \right) \tag{7.3}$$

From here we find that the linear term with respect to v is exactly where f appear only once,

$$D_0 F : L^2 \to T_{q_0} M, \qquad D_0 F(v) = q_0 \circ \int_0^1 f_{v(t)} dt = \int_0^1 f_{v(t)}(q_0) dt.$$

To compute the differential at a generic point $u \in \mathcal{U}$ we have to consider the expansion near 0 of the map

$$v(\cdot) \mapsto F(u(\cdot) + v(\cdot)) = q_0 \circ \overrightarrow{\exp} \int_0^1 f_{(u+v)(t)} dt.$$

The variation formula (6.21) let us to write (compare also with the proof of Proposition 3.41)

$$\begin{aligned} \overrightarrow{\exp} \int_0^1 f_{(u+v)(t)} dt &= \overrightarrow{\exp} \int_0^1 f_{u(t)} + f_{v(t)} dt \\ &= \overrightarrow{\exp} \int_0^1 \left(\overrightarrow{\exp} \int_0^t \operatorname{ad} f_{u(s)} ds \right) f_{v(t)} dt \circ \overrightarrow{\exp} \int_0^1 f_{u(t)} dt \\ &= \overrightarrow{\exp} \int_0^1 P_{t*}^0 f_{v(t)} dt \circ P_0^1 \end{aligned}$$

where $P_\tau^t : \gamma(\tau, u) \mapsto \gamma(t, u)$. In other words we rewrite

$$F(u) = P_0^1(G(u)), \qquad D_u F = P_{0*}^1 D_u G,$$

and reduced the problem to the expansion of G, which is easier. Indeed we have the Volterra expansion

$$\overrightarrow{\exp} \int_0^1 P_{t*}^0 f_{v(t)} dt \approx \mathrm{Id} + \int_0^1 P_{t*}^0 f_{v(t)} dt + \iint\limits_{0 \le t_2 \le t_1 \le t} P_{t*}^0 f_{v(t_2)} \circ P_{t*}^0 f_{v(t_1)} dt_1 dt_2 + \dots \tag{7.4}$$

from which we get, denoting $q_1 = F(u)$

$$D_u F : L^2 \to T_{q_1} M, \qquad D_u F(v) = P_{0*}^1 \int_0^1 P_{t*}^0 f_{v(t)}(q_0) dt = \int_0^1 P_{t*}^1 f_{v(t)}(q_1) dt.$$

□

Now we want to characterize sub-Riemannian extremals as critical points of the end point map.

Proposition 7.3. *The following properties hold:*

(i) $(u(t), \lambda(t))$ *is an abnormal extremal if and only if* $u(t)$ *is a critical point for* F*. Moreover there exists* $\lambda_1 \in T^*_{F(u)}M$ *such that*

$$\lambda_1 D_u F = 0, \qquad \lambda(t) = P_t^{1*}\lambda_1, \tag{7.5}$$

(ii) $(u(t), \lambda(t))$ *is a normal extremal if and only if there exists* $\lambda_1 \in T^*_{F(u)}M$ *such that*

$$\lambda_1 D_u F = u, \qquad \lambda(t) = P_t^{1*}\lambda_1. \tag{7.6}$$

where in the last equality we identify $u \in L^2$ *with the element* $(u, \cdot)_{L^2} \in (L^2)'$

Proof. (i). u is a critical point of F if and only if D_uF is not surjective. In other words there exists a covector

$$\lambda_1 : T_{F(u)}M \to \mathbb{R} \quad \text{such that} \quad \lambda_1 D_u F = 0, \tag{7.7}$$

where $\lambda_1 D_u F$ denotes the composition of maps

$$L^2 \xrightarrow{D_uF} T_{q_1}M \xrightarrow{\lambda_1} \mathbb{R} \tag{7.8}$$

Now, if we let $\lambda_t = P_t^{1*}\lambda_1$, it remains to prove that the curve λ_t satisfies the relation

$$h_i(\lambda(t)) = 0, \qquad h_i(\lambda(t)) = \langle \lambda(t), f_i(q(t)) \rangle$$

We have

$$\begin{aligned}
\langle \lambda_1, D_u F\rangle = 0 &\Leftrightarrow \int_0^1 \left\langle \lambda_1, P^1_{t*} f_{v(t)}(q_1)\right\rangle dt = 0, \forall v \\
&\Leftrightarrow \int_0^1 \left\langle \lambda(t), f_{v(t)}(q(t))\right\rangle dt = 0, \forall v \\
&\Leftrightarrow \sum_i \int_0^1 v_i \left\langle \lambda(t), f_i(q(t))\right\rangle dt = 0, \forall v \\
&\Leftrightarrow \langle \lambda(t), f_i(q(t))\rangle = 0, \forall i
\end{aligned}$$

(ii). With analogous proof. □

At the end we can rewrite this result in the following way

Corollary 7.4. *A control* $u \in \mathcal{U}$ *is an extremal if and only if there exist some "Lagrange multiplier"*

$$\lambda \in T^*_{F(u)}M, \qquad \lambda \neq 0$$

such that the following equality holds

$$\lambda D_u F = \nu u$$

where

(i) $\nu = 0$ *in the abnormal case,*

(ii) $\nu \neq 0$ *in the normal case.*

7.2 Lagrange points and Lagrange submanifolds

In this section we will work in the following setting:

Let $\mathcal{U}$ be an open set in a Hilbert space and let M be a smooth n-dimensional manifold. Assume we have a pair of smooth maps

$$F : \mathcal{U} \to M, \qquad \varphi : \mathcal{U} \to \mathbb{R}.$$

We want to characterize critical points of the functional φ when restricted to level set of F.

$$\min_{F^{-1}(q)} \varphi, \qquad q \in M. \tag{7.9}$$

Definition 7.5. Let $a : M \to \mathbb{R}$ be a smooth function and $N \subset M$ be a smooth submanifold. Then $q \in N$ is said a *critical point* of $a\big|_N$ if $d_q a\big|_{T_qN} = 0$.

We start with a geometric version of the Lagrange multipliers rule, which characterize constrained critical points.

Proposition 7.6 (Lagrange multipliers rule)**.** *Assume $u \in \mathcal{U}$ is a regular point of $F : \mathcal{U} \to M$ such that $F(u) = q$. Then u is a critical point of $\varphi\big|_{F^{-1}(q)}$ if and only if*

$$\exists\, \lambda \in T_q^*M,\ \lambda \neq 0, \qquad s.t. \qquad d_u\varphi = \lambda\, D_uF. \tag{7.10}$$

Proof. Recall that the differential of F is a well defined map

$$D_uF : T_u\mathcal{U} \to T_qM, \qquad q = F(u),$$

and, since u is a regular point, D_uF is surjective and the level set

$$A_q := F^{-1}(q) = \{u \in \mathcal{U},\ F(u) = q\} \subset \mathcal{U},$$

is a smooth submanifold, with $u \in A_q$.

Since u is a critical point of $\varphi\big|_{A_q}$, by definition $d_u\varphi\big|_{T_uA_q} = 0$. Moreover $T_uA_q = \operatorname{Ker} D_uF$. Thus we have that

$$\operatorname{Ker} D_uF \subset \operatorname{Ker} d_u\varphi. \tag{7.11}$$

Now consider the following diagram

$$\begin{tikzcd} T_u\mathcal{U} \arrow[r, "D_uF"] \arrow[dr, "d_u\varphi"'] & T_qM \arrow[d, "?"] \\ & \mathbb{R} \end{tikzcd} \tag{7.12}$$

From (7.11), using a standard lemma of linear algebra and the fact that D_uF is surjective, it follows that there exists a nontrivial linear map $\lambda : T_qM \to \mathbb{R}$ (that means $\lambda \in T_q^*M \setminus \{0\}$) that makes the diagram (7.12) commutative. □

Remark 7.7. In the case of sub-Riemannian geometry $\mathcal{U}$ represents the set of controls, F is the end-point map and φ is the length of the curve associated to controls. In this framework, the problem of finding constrained critical points means exactly to find critical points of the length when the initial point of the curve is fixed. Hence the solutions of the problem (7.9) represent exactly sub-Riemannian geodesics. In particular *abnormal extremals* corresponds to critical points of F, while *normal extremals* satisfy the Lagrange multipliers rule.

Now we want to consider second order information about our critical points. Recall that, if $\mathcal{V}$ is a submanifold in a Hilbert space $\mathcal{U}$, the first differential of a smooth function $\psi : \mathcal{V} \to \mathbb{R}$ at a point $u \in \mathcal{V}$ is well defined independently on coordinates

$$d_u\psi : T_u\mathcal{V} \to \mathbb{R}, \qquad d_u\psi(v) = \frac{d}{dt}\bigg|_{t=0} \psi(\gamma(t)),$$

where $\gamma : (-\varepsilon, \varepsilon) \to \mathcal{V}$ is a curve that satisfies $\gamma(0) = u,\ \dot\gamma(0) = v$.

This is not the case for the second differential. Indeed the second order derivative of a function ψ is meaningful only at its critical points (at a regular point, by implicit function theorem, one can always find coordinates such that ψ is locally linear). Hence, if u is a critical point for ψ it is intrinsically defined the quadratic map

$$\operatorname{Hess}_u \psi : T_u\mathcal{V} \to \mathbb{R}, \qquad v \mapsto \frac{d^2}{dt^2}\bigg|_{t=0} \psi(\gamma(t))$$

In our case $\mathcal{V} = F^{-1}(q)$, $\psi = \varphi\big|_{F^{-1}(q)}$, and since $T_uF^{-1}(q) = \operatorname{Ker} D_uF$, we have a well defined quadratic form

$$\operatorname{Hess}_u \varphi\big|_{F^{-1}(q)} : \operatorname{Ker} D_uF \to \mathbb{R}$$

which is computed as follows

Proposition 7.8. *For all $v \in \operatorname{Ker} D_uF$ we have*

$$\operatorname{Hess}_u \varphi\big|_{F^{-1}(q)}(v) = d_u^2\varphi(v) - \lambda\, D_u^2F(v). \tag{7.13}$$

where λ is defined by (7.10)*.*

Proof. Notice that $F^{-1}(q) \subset \mathcal{U}$ is a submanifold in a Hilbert space. Fix a point $q \in M$ and $u \in F^{-1}(q)$. Consider a path $u(s)$ in $\mathcal{U}$ such that $u(0) = u$ and $u(s) \in F^{-1}(q)$ for all s. Then in coordinates we have, differentiating twice with respect to u

$$\begin{aligned} F(u(s)) = q \quad &\Longrightarrow \quad \frac{dF}{du}\dot u = 0 \\ &\Longrightarrow \quad \langle \frac{d^2F}{du^2}\dot u, \dot u\rangle + \frac{dF}{du}\ddot u = 0. \end{aligned} \tag{7.14}$$

where we denoted[1] by $\dot u = \dot u(0)$ and $\ddot u = \ddot u(0)$. The same computation for φ gives

$$
\begin{aligned}
D_u^2\varphi(\dot u) &= \frac{d^2}{ds^2}\Big|_{s=0}\varphi(u(s)) \\
&= \langle \frac{d^2\varphi}{du^2}\dot u, \dot u\rangle + \frac{d\varphi}{du}\ddot u \\
&= \langle \frac{d^2\varphi}{du^2}\dot u, \dot u\rangle + \lambda\frac{dF}{du}\ddot u \qquad\qquad \text{(by (7.10))} \\
&= \langle \frac{d^2\varphi}{du^2}\dot u, \dot u\rangle - \lambda\langle\frac{d^2F}{du^2}\dot u, \dot u\rangle \qquad \text{(by (7.14))}
\end{aligned}
$$

□

Definition 7.9. In the previous setting let $u \in \mathcal{U}$ and $\lambda \in T^*_{F(u)}M$ be a non zero covector. We say that λ is a *Lagrange multiplier* for the problem (7.9) associated to u (equivalently that (u, λ) is a *Lagrange point*) if

$$
\exists\, \lambda \in T^*_{F(u)}M \qquad \text{s.t.} \qquad d_u\varphi = \lambda\, D_uF \tag{7.15}
$$

We denote the set of all Lagrange points by $C_{F,\varphi}$. More precisely

$$
C_{F,\varphi} = \{(u, \lambda) \in \mathcal{U} \times T^*M \mid F(u) = \pi(\lambda),\ d_u\varphi = \lambda\, D_uF\} \tag{7.16}
$$

The set $C_{F,\varphi}$ is a well-defined subset of the vector bundle $F^*(T^*M)$ (see Definition 2.42).

Now we give some transversality conditions that ensure F_c is an immersion.

Definition 7.10. The pair (F, φ) is said to be a *Morse pair* (or a *Morse problem*) if 0 is not a critical value for the smooth map

$$
\theta : F^*(T^*M) \to \mathcal{U}^* \simeq \mathcal{U}, \qquad (u, \lambda) \mapsto d_u\varphi - \lambda\, D_uF. \tag{7.17}
$$

Remark 7.11. Notice that, if $M = \{0\}$, 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.

In canonical coordinates $\lambda = (\xi, x)$ in T^*M we can describe the set $C_{F,\varphi} \subset F^*(T^*M)$ as the set of (u, ξ, x) that satisfy

$$
\begin{cases}
\dfrac{d\varphi}{du} - \xi\dfrac{dF}{du} = 0 \\
F(u) = x
\end{cases} \tag{7.18}
$$

The linearization of the system (7.18) at a point (u, ξ, x) is given by the set of points (u', ξ', x') that satisfy

$$
\begin{cases}
\dfrac{d^2\varphi}{du^2}u' - \xi\dfrac{d^2F}{du^2}u' - \xi'\dfrac{dF}{du} = 0 \\
\dfrac{dF}{du}u' = x'
\end{cases} \tag{7.19}
$$

Let us denote the linear map $Q : \mathcal{U} \to \mathcal{U}^* \simeq \mathcal{U}$ defined by

$$
Qu' = \xi\frac{d^2F}{du^2}u' - \frac{d^2\varphi}{du^2}u'.
$$

[1]Recall that the notation $\frac{dF}{du}$ stands for the differential of F in coordinates, while the notation $\langle d_uF, \cdot\rangle$ is intrinsic.

Recall that $\mathcal{U}$ is an Hilbert space and we can identify the space with its dual using the scalar product. Since Q is defined by second derivatives of the maps F and φ, it is a symmetric operator.

Note. We also need the extra assumption that $\operatorname{im} Q$ is closed. This is not restrictive for our purposes since, as explained later, for a Morse problem $\operatorname{Im} Q$ has finite codimension, hence it is a closed subspace.

The definition of Morse problem is immediately rewritten as follows: the pair (F, φ) define a Morse problem if and only if the following map is surjective.

$$\Theta : \mathcal{U} \times \mathbb{R}^{n*} \to \mathcal{U}, \qquad \Theta(u', \xi') = Qu' - \xi' \frac{dF}{du}. \tag{7.20}$$

Indeed the map Θ is exactly the coordinate expression of the differential of the first equation in (7.18) (that is the coordinate version of (7.17)).

Lemma 7.12. *If (F, φ) define a Morse problem, then $C_{F,\varphi}$ is a smooth n-dimensional manifold in $F^*(T^*M)$.*

Proof. First notice that, from (7.16) and (7.17), it immediately follows that

$$C_{F,\varphi} = \theta^{-1}(0) \tag{7.21}$$

Since 0 is not a critical value for θ, from the implicit function theorem it follows that $C_{F,\varphi}$ is a submanifold. A simple dimension argument let us to conclude under the additional assumption $\dim \mathcal{U} < +\infty$. Indeed in this case, since the differential of the map (7.17) is surjective we have that

$$\dim\, \phi^*(T^*M) - \dim\, C_{F,\varphi} = \dim \mathcal{U}$$

so we can compute the dimension of $C_{F,\varphi}$

$$\begin{aligned} \dim\, C_{F,\varphi} &= \dim\, \phi^*(T^*M) - \dim\, \mathcal{U} \\ &= (\dim \mathcal{U} + \operatorname{rank} T^*M) - \dim\, \mathcal{U} \\ &= \operatorname{rank} T^*M = n \end{aligned}$$

In the general case (when $\dim \mathcal{U} = +\infty$) the above argument is no more valid and we have to use explicitly that Q is self-adjoint. Let us denote with $B : \mathbb{R}^{n*} \to \mathcal{U}$ the map

$$B\xi' = \xi' \frac{dF}{du}, \qquad \text{so that} \qquad \Theta : (u', \xi') \mapsto Qu' - B\xi'$$

Since Θ is surjective and $\dim(\operatorname{Im} B) \leq n$ we get

$$\operatorname{codim} \operatorname{Im} Q \leq \dim \operatorname{Im} B \leq n$$

Moreover since Q is self-adjoint we have

$$\mathcal{U} = \operatorname{Ker} Q \oplus \operatorname{Im} Q, \qquad \dim \operatorname{Ker} Q = \dim(\operatorname{Im} Q)^{\perp} \leq n$$

Now, being Θ the coordinate expression of the differential of θ, the dimension of $C_{F,\varphi}$ coincide with the dimension of the kernel of Θ. In addition, if we denote with $\pi_{\mathrm{Ker}} : \mathcal{U} \to \operatorname{Ker} Q$ and $\pi_{\mathrm{Im}} : \mathcal{U} \to \operatorname{Im} Q$ the orthogonal projection onto the two subspaces, it is easy to see that

$$\Theta(u', \xi') = 0 \quad \iff \begin{cases} \pi_{\mathrm{Ker}} B\xi' = 0, \\ \pi_{\mathrm{Im}} B\xi' = Qu' \end{cases}$$

from which it immediately follows the identity

$$\dim \operatorname{Ker} \Theta = \dim \operatorname{Ker} Q + \dim \operatorname{Ker} (\pi_{\operatorname{Ker}} B) = n$$

since $\pi_{\operatorname{Ker}} B : \mathbb{R}^n \to \operatorname{Ker} Q$ is a surjective map between finite-dimensional spaces. $\square$

The last characterization of Morse problem leads to a convenient criterion to check, in coordinates, whether a pair (F, φ) defines a Morse problem or not.

Lemma 7.13. *Assume that* $\operatorname{Im} Q$ *is closed. Then the pair* (F, φ) *defines a Morse problem if and only if*

$$\operatorname{Ker} Q \cap \operatorname{Ker} D_u F = 0 \tag{7.22}$$

Proof. The problem is not Morse if and only if the image of the differential of the map (7.17) is not surjective, i.e. there exists $w \in \mathcal{U}$ that is orthogonal to $\operatorname{im} \Theta$,

$$\langle Qu', w \rangle - \langle \xi' \frac{dF}{du}, w \rangle = 0$$

Using that Q is self-adjoint we get

$$\langle u', Qw \rangle - \langle \xi' \frac{dF}{du}, w \rangle = 0, \qquad \forall\, \xi', u'$$

that is equivalent, since we have disjoint variables, to

$$Qw = 0 \qquad \text{and} \qquad \frac{dF}{du} w = 0$$

$\square$

Let us consider now the projection map $\phi_c : C_{F,\varphi} \longrightarrow T^*M$ defined by :

$$F_c(u, \lambda) = \lambda.$$

Definition 7.14. 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.

Proposition 7.15. *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 it is Lagragian.

(i). Recall that $F_c : C_{F,\varphi} \to T^*M$ where

$$C_{F,\varphi} = \{(u, \xi, x) \,|\, \text{equations (7.18) holds}\}$$

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 (7.18)

$$T_{(u,\lambda)} C_{F,\varphi} = \{(u', \xi', x') \,|\, \text{equations (7.19) holds}\}$$

where

$$D_{(u,\lambda)} F_c(u', \xi', x') = (\xi', x')$$

Now looking at (7.19) it easily seen that

$$D_{(u,\lambda)}F_c(u',\xi',x') = 0 \qquad \text{iff} \qquad Qu' = \frac{dF}{du}u' = 0.$$

Since (F,φ) defines a Morse problem we have by (7.22) 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 $\sigma = ds$ and pullback commutes with the differential it is sufficient to show that F_c^*s is closed. In particular we will show that

$$F_c^*s = d\varphi\big|_{C_{F,\varphi}}.$$

By definition the map F_c we have that the following diagram is commutative:

$$\begin{tikzcd} C_{F,\varphi} \arrow[r, "F_c"] \arrow[d, "\pi_U"'] & T^*M \arrow[d, "\pi_M"] \\ \mathcal{U} \arrow[r, "F"'] & M \end{tikzcd} \tag{7.23}$$

Moreover, notice that if $F : M \to N$ is smooth and $\omega \in \Lambda^1(N)$, by definition of pull-back we have $(F^*\omega)_q = \omega_{F(q)} \circ D_qF$. Hence

$$\begin{aligned} (F_c^*s)_{(u,\lambda)} &= s_\lambda \circ D_{(u,\lambda)}F_c && \text{(by definition of } s\text{)} \\ &= \lambda \circ \pi_{M*} \circ D_{(u,\lambda)}F_c && \text{(by (7.23))} \\ &= \lambda \circ D_uF \circ \pi_{U*} && \text{(by (7.10))} \\ &= d_u(\varphi \circ \pi_U) \end{aligned}$$

□

Remark 7.16. Recall that the set $\mathcal{L}_{F,\varphi}$ of Lagrange multipliers (see Definition 7.9) is the image of $C_{F,\varphi}$ under the map

$$F_c : C_{F,\varphi} \to T^*M, \qquad (u,\lambda) \mapsto \lambda, \qquad \mathcal{L}_{F,\varphi} := \operatorname{im} F_c$$

From the last proposition it follows that, if $\mathcal{L}_{F,\varphi}$ is a submanifold, then it is a Lagrangian submanifold.

We resume the results obtained above in the following

Proposition 7.17. *Let (F,φ) be a Morse problem and assume (u,λ) is a Lagrange point such that u is a regular point for F, where $F(u) = q$. The following properties are equivalent:*

(i) $\operatorname{Hess}_u \varphi\big|_{F^{-1}(q)}$ *is degenerate,*

(ii) *(u,λ) is a critical point for the map $\pi \circ F_c = F\big|_{C_{F,\varphi}} : C_{F,\varphi} \to M$,*

(iii) *if $\mathcal{L}_{F,\varphi}$ is a submanifold, 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

$$\mathrm{Hess}_u\varphi\big|_{F^{-1}(q)}(v) = \langle Qv, v\rangle, \qquad \forall\, v \in \operatorname{Ker} D_uF.$$

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 $u' \in \operatorname{Ker} D_uF$ such that

$$\langle Qu', v\rangle = 0, \qquad \forall\, v \in \operatorname{Ker} D_uF.$$

In other words $Qu' \perp \operatorname{Ker} D_uF$ that is equivalent to say that Qu' is a linear combination of the row of the Jacobian matrix

$$\exists\, \xi' \quad \text{such that} \quad Qu' = \xi' \frac{dF}{du}$$

From equations (7.19) it follows immediately that (i) is equivalent to (ii). The fact that (ii) is equivalent to (iii) is obvious. □

7.3 Sub-Riemannian case

In this section we want to specify all the theory we developed in the previous one to the sub-Riemannian case. As we mentioned, we will consider the functional J defined by $J(u) = \frac{1}{2}\int |u|^2$ and we consider its critical points constrained to level set of the end point map F, that means that we fix the final point of our trajectory (as usual we assume that the starting point q_0 is fixed by the very beginning).

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.

Lemma 7.18. *Let $q_1 \in M$ and (u, λ) be a critical point of $J\big|_{F^{-1}(q_1)}$. Then for every $v \in \operatorname{Ker} D_uF$*

$$\mathrm{Hess}_u J\big|_{F^{-1}(q_1)}(v) = \|v\|^2_{L^2} - \langle \lambda\,, \iint\limits_{0\le\tau\le t\le 1} [P^1_{\tau *} f_{v(\tau)}, P^1_{t*} f_{v(t)}] d\tau dt\rangle \tag{7.24}$$

where $P^t_s : \gamma(s) \mapsto \gamma(t)$ is the flow defined by the control u.

Proof. By Proposition 7.8 we have

$$\mathrm{Hess}_u J\big|_{F^{-1}(q_1)}(v) = d^2_u J - \lambda D^2_u F.$$

It is easy to compute derivatives of J. Indeed we can rewrite it as $J(u) = \frac{1}{2}(u,u)_{L^2}$, hence

$$d_uJ(v) = (u, v)_{L^2}, \qquad d^2_u J(v) = (v, v)_{L^2} = \|v\|^2_{L^2}, \qquad \forall\, v \in \operatorname{Ker} D_uF$$

It remains to compute the second derivative of the end-point map. From the Volterra expansion (7.4) we get

$$D^2_u F(v, v) = 2\, q_1 \circ \iint\limits_{0\le\tau\le t\le 1} P^1_{\tau *} f_{v(\tau)} \circ P^1_{t*} f_{v(t)} d\tau dt$$

To end the proof we use the following lemma on chronological calculus, which we will use to symmetrize the second derivative

Lemma 7.19. *Let X_t be a nonautonomous vector field on M. Then*

$$\iint\limits_{0\leq s\leq t\leq 1} X_s \circ X_t ds dt = \frac{1}{2}\int_0^1 X_s ds \circ \int_0^1 X_t dt + \frac{1}{2}\iint\limits_{0\leq s\leq t\leq 1} [X_s, X_t] ds dt \tag{7.25}$$

Proof of the Lemma. It is a simple computation

$$\begin{aligned} 2\iint\limits_{0\leq s\leq t\leq 1} X_s \circ X_t ds dt &= 2\iint\limits_{0\leq s\leq t\leq 1} X_s \circ X_t ds dt - \iint\limits_{0\leq s\leq t\leq 1} X_t \circ X_s ds dt + \iint\limits_{0\leq s\leq t\leq 1} X_t \circ X_s ds dt \\ &= \int_0^1\int_0^1 X_s \circ X_t ds dt + \iint\limits_{0\leq s\leq t\leq 1} [X_t, X_s] ds dt \\ &= \int_0^1 X_s ds \circ \int_0^1 X_t dt + \iint\limits_{0\leq s\leq t\leq 1} [X_s, X_t] ds dt \end{aligned}$$

where in the second line we exchange the role of s and t in the integral. □

□

Proposition 7.20. *The sub-Riemannian problem (F, J) is Morse.*

Proof. We use the characterization of Lemma 7.13. We have to show that, in canonical coordinates $\lambda = (\xi, x)$,

$$\operatorname{Im}\left(\xi\frac{d^2F}{du^2} - \mathrm{Id}\right) \text{ is closed}, \qquad \operatorname{Ker}\left(\xi\frac{d^2F}{du^2} - \mathrm{Id}\right) \cap \operatorname{Ker}\left(\frac{dF}{du}\right) = 0. \tag{7.26}$$

Using the previous notation and defining $g_v^t := P_{t*}^1 f_v$, we can write

$$\frac{dF}{du} v(\cdot) = q_1 \circ \int_0^1 g_{v(t)}^t dt$$

Moreover we have

$$\left\langle \xi\frac{d^2F}{du^2} v(\cdot), v(\cdot)\right\rangle = \xi\iint\limits_{0\leq\tau\leq t\leq 1} g_{v(\tau)}^\tau \circ g_{v(t)}^t d\tau dt$$

Since we want to find the kernel of the bilinear form we need to recover the linear operator associated to it, i.e. to symmetrize the form

$$(Av)(t) := \left(\xi\frac{d^2F}{du^2} v(\cdot)\right)(t) = \xi\int_0^t g_{v(\tau)}^\tau d\tau \circ g_{v(t)}^t + \xi\, g_{v(t)}^t \circ \int_t^1 g_{v(\tau)}^\tau d\tau \tag{7.27}$$

Since (7.27) is an compact integral operator, then $A - \mathrm{Id}$ is Fredholm, and the closedness of $\operatorname{Im}(A - \mathrm{Id})$ follows from the fact that it is of finite codimension. On the other hand, for every control $v \in \operatorname{Ker} D_u F$ we can compute (see (7.2))

$$q_1 \circ \int_0^t g_{v(\tau)}^\tau d\tau = -q_1 \circ \int_t^1 g_{v(\tau)}^\tau d\tau$$

Hence we have that v belong to the intersection (7.26) if and only if it satisfies

$$\left(I - \xi \frac{d^2F}{du^2} v(\cdot)\right)(t) = v(t) + \xi \left[g^t_{v(t)}, \int_t^1 g^\tau_{v(\tau)} d\tau \right](q_1)$$

which has trivial kernel since is a Volterra operator, of the form $v(t) + \int_0^t K(t,\tau)v(\tau)d\tau$.

$\square$

Corollary 7.21. *The manifold of Lagrange multilpliers of the sub-Riemannian problem (F, J) is a smooth n-dimensional submanifold of T^*M, namely*

$$\mathcal{L}_{(F,J)} := \{\lambda_1 \in T^*M |\, \lambda_1 = e^{\vec{H}}(\lambda_0),\, \lambda_0 \in T^*_{q_0}M\}$$

where H is the sub-Riemannian Hamiltonian.

To end this chapter we consider the *free initial point* problem, i.e. we consider the free end point map

$$\mathbb{F} : M \times \mathbf{U} \to M, \qquad (q, u) \mapsto \gamma(1, q, u),$$

where

$$\gamma(t, q, u) = q \circ \overrightarrow{\exp} \int_0^t f_{u(s)} ds,$$

is the solution to the Cauchy problem

$$\dot{\gamma}(t) = f_{u(t)}(\gamma(t)), \qquad \gamma(0) = q.$$

We look for solution of the problem

$$\min_{\mathbb{F}^{-1}(q_1)} J(u) + a(q), \qquad a \in \mathcal{C}^\infty(M) \tag{7.28}$$

Critical points of this problem can be found with the Lagrange multiplier rule, where, following the notation exploited in the previuos Chapter

$$F = \mathbb{F}, \qquad \varphi = J + a.$$

Fix a point $(q_0, \widetilde{u}) \in M \times \mathbf{U}$. It is easy to see that

$$\mathbb{F}\big|_{\{q_0\}\times\mathbf{U}} = F, \qquad \mathbb{F}\big|_{M\times\{\widetilde{u}\}} = P_0^1$$

and it is easy to see that the equation

$$\lambda_1 D_{(q_0,\widetilde{u})} F = d_{(q_0,\widetilde{u})}\varphi$$

splits into

$$\begin{cases} \lambda_1 D_u F = d_u J = u, \\ \lambda_1 P^1_{0*} = d_{q_0} a \end{cases}$$

In other words, by PMP, we have that to every critical point of the problem (7.28) we can associate the normal extremal

$$\lambda_t = P_t^{0*} \lambda_0, \qquad \lambda_0 = d_{q_0} a,$$

where the initial condition is defined by the function a.

Exercise 7.22. Consider the *free endpoint* problem, i.e. find solution of the problem

$$\min_u J(u) - a(F(u)), \qquad a \in \mathcal{C}^\infty(M) \tag{7.29}$$

In other words now we do not restrict to the sublevel $F^{-1}(q_1)$ (we do not fix the final point of the trajectory) but we consider a penalty in the functional we want to minimize.

Prove that u is a critical point of this functional if and only if

$$\lambda_1 D_u F = u, \qquad \lambda_1 = d_{F(u)}a.$$

7.4 Exponential map

Now we can define the sub-Riemannian exponential map.

Definition 7.23. Let $q_0 \in M$. We define the *exponential map* (from q_0) of the sub-Riemannian problem the mapping

$$\mathcal{E}_{q_0} : T^*_{q_0}M \to M, \qquad \mathcal{E}_{q_0}(\lambda_0) = \pi \circ e^{\vec{H}}(\lambda_0). \tag{7.30}$$

When the initial point q_0 is fixed we omit it in the notation, writing simply $\mathcal{E}$.

The homogeneity of the sub-Riemannian Hamiltonian H yields to the following homogeneity property of the flow of $\vec{H}$.

Lemma 7.24. *Let H be the sub-Riemannian Hamiltonian. Then, for every $\lambda \in T^*M$*

$$e^{t\vec{H}}(\alpha\lambda) = \alpha e^{\alpha t\vec{H}}(\lambda), \qquad \forall\, \alpha, t > 0. \tag{7.31}$$

Proof. By Remark **??** we know that if $\lambda(t) = e^{t\vec{H}}(\lambda_0)$ is a solution of the Hamiltonian system, then also $\lambda_\alpha(t) := \alpha\lambda(\alpha t)$ is a solution. The result follows from the uniqueness of the solution and the identity that $\lambda_\alpha(0) = \alpha\lambda(0)$. □

The exponential map sends a covector λ_0 to the point at time 1 of the normal extremal path with initial condition λ_0. The homogeneity property let us to recover the whole geodesic as the image of the ray that join 0 to λ_0 in the fiber $T^*_{q_0}M$.

Corollary 7.25. *Let $\lambda(t)$, $t \in [0,T]$, 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

$$\gamma(t) = \mathcal{E}(t\lambda_0), \qquad t \in [0,T]$$

Proof. Using (7.31) we get

$$\mathcal{E}(t\lambda_0) = \pi(e^{\vec{H}}(t\lambda_0)) = \pi(e^{t\vec{H}}(\lambda_0)) = \pi(\lambda(t)) = \gamma(t).$$

□

Remark 7.26. Due to the homogeneity property we can consider the following map

$$\mathbb{E}_{q_0} : \mathbb{R}^+ \times C_{q_0} \to M, \qquad \mathbb{E}_{q_0}(t, \lambda_0) = \mathcal{E}_{q_0}(t\lambda_0)$$

where C_{q_0} is the hypercilynder of normalized covectors

$$C_{q_0} = \{\lambda \in T^*_{q_0}M |\ H(\lambda) = 1/2\}$$

In other words we restrict to length parametrized extremal paths, considering the time as an extra variable.

We end this section by the Hamiltonian version of the Gauss' Lemma

Proposition 7.27 (Gauss' Lemma). *Let (u, λ_1) be associated with a normal minimizer starting from q_0. The covector λ_1 annihilates the tangent space to the sub-Riemannian front $\mathcal{E}_{q_0}(T^*_{q_0}M)$.*

Proof. It is enough to show that for every smooth variation η^s of initial covectors such that $\eta^0 = \lambda$ we have

$$\left\langle \lambda(1), \frac{d}{dt}\bigg|_{s=0} \mathcal{E}_{q_0}(\eta^s) \right\rangle = 0$$

Let us consider a family of initial covectors $\lambda^s \in H^{-1}(1/2)$ and their associated controls $u^s(\cdot)$ defined by the identities

$$u^s_i(t) = \langle \eta^s(t), f_i(\gamma^s(t)) \rangle, \qquad \|u^s\|_{L^2} = 1$$

where $\eta^s(t)$ is the solution of the Hamiltonian equation with initial value η^s and $\gamma^s(t)$ is the corresponding trajectory. For these controls one has $\mathcal{E}_{q_0}(\eta^s) = F(u^s)$ hence

$$\frac{d}{ds}\bigg|_{s=0} \mathcal{E}_{q_0}(\eta^s) = \frac{d}{ds}\bigg|_{s=0} F(u^s) = D_u F(v), \qquad v := \frac{d}{ds}\bigg|_{s=0} u^s \tag{7.32}$$

Notice that v is orthogonal to u since $\|u^s\| = const$. Thus by the normal equation (7.5) and (7.32)

$$\left\langle \lambda(1), \frac{d}{ds}\bigg|_{s=0} \mathcal{E}_{q_0}(\eta^s) \right\rangle = \langle \lambda(1), D_u F(v) \rangle = (u, v)_{L^2} = 0. \tag{7.33}$$

□

7.5 Conjugate points and minimality properties of geodesics

Consider now an extremal pair $(u(t), \lambda(t))$, $t \in [0, 1]$, such that the corresponding extremal path $\gamma(t)$ is strictly normal. Recall that by Corollary 4.57, the curve γ is a geodesic. Moreover, $\gamma|_{[0,s]}$ also is a geodesic, for every $s > 0$, and if we reparametrize it as $\gamma_s(t) := \gamma(st), t \in [0, 1]$ it corresponds to the control $u_s(t) = su(st)$.

Definition 7.28. A geodesic $\gamma(t)$ is said to be *strongly normal*, if $\gamma|_{[0,s]}$ is stricly normal $\forall\, s > 0$.

Proposition 7.29. *Let γ be a strongly normal geodesic. The following are equivalent:*

(*i*) $\mathrm{Hess}_u J\big|_{F^{-1}(\gamma(1))}$ *is positive definite,*

(*ii*) $\mathrm{Hess}_{u_s} J\big|_{F^{-1}(\gamma_s(1))}$ *is non degenerate for all $s > 0$.*

Proof. Recall that

$$\mathrm{Hess}_{u_s} J\big|_{F^{-1}(\gamma_s(1))}(v) = \|v\|^2_{L^2} - \langle \lambda_s\,, D^2_{u_s}F(v,v)\rangle \tag{7.34}$$

which is a well defined quadratic form of the kind $\mathrm{Id} - Q_s$, with Q_s compact, since it is a Volterra operator (see also the proof of Proposition 7.20). Then define the function

$$\begin{aligned}\alpha(s) :&= \inf_{\|v\|=1} \left\{ \|v\|^2_{L^2} - \langle \lambda_s\,, D^2_{u_s}F(v,v)\rangle \right\} \\ &= 1 - \sup_{\|v\|=1} \left\langle \lambda_s, D^2_{u_s}F(v,v)\right\rangle \end{aligned}\tag{7.35}$$

Notice that, being a compact operator, if the quadratic form is not negative definite then the maximum is attained in (7.35) and it is the maximum eigenvalue[2]. On the other hand, if the quadratic form is negative definite the supremum is always zero (it is sufficient to evaluate it on any orthonormal sequence).

Now we prove the following claim, which immediately implies the proposition, using that the Hessian is degenerate at some point $\bar{s}$ then $\alpha(\bar{s}) = 0$ (indeed $\alpha(\bar{s}) = 0$ means that the quadratic form is nonnegative and has infimum zero, hence has zero as eigenvalue, by compactness).

Claim. $\alpha(s)$ is a continuous and monotonic decreasing function, with $\alpha(0) = 1$.

Proof of the Claim. It is easy to show that the following formulas hold for the first and second differentials computed at points u_s

$$D_{u_s}F(v) = \int_0^s P^1_{t*} f_{v(t)} dt, \qquad D^2_{u_s}F(v,v) = \iint\limits_{0\le\tau\le t\le s} [P^1_{\tau *} f_{v(\tau)}, P^1_{t*} f_{v(t)}] d\tau dt \tag{7.36}$$

Now consider $0 \le s \le \widehat{s} \le 1$ and $v \in \mathrm{Ker}\, D_{u_s}F$ and define the control

$$\widehat{v}(t) = \begin{cases} v(\frac{\widehat{s}}{s}t), & 0 \le t \le \frac{s}{\widehat{s}}, \\ 0, & \frac{s}{\widehat{s}} < t \le 1. \end{cases}$$

Then $\|\widehat{v}\| = \|v\|$, $\widehat{v} \in \mathrm{Ker}\, D_{u_{\widehat{s}}}F$ and $D^2_{u_s}F(v,v) = D^2_{u_{\widehat{s}}}F(\widehat{v},\widehat{v})$, hence $\alpha(s) \ge \alpha(\widehat{s})$.

On the other hand, if we consider $\gamma_s(t) = \gamma(st)$ as defined on the whole segment $[0,1]$, we can rewrite (7.36) as follows

$$D_{u_s}F(v) = s\int_0^1 P^1_{st*} f_{v(t)} dt, \qquad D^2_{u_s}F(v,v) = s^2 \iint\limits_{0\le\tau\le t\le 1} [P^1_{s\tau *} f_{v(\tau)}, P^1_{st*} f_{v(t)}] d\tau dt \tag{7.37}$$

To prove that α is continuous we need that both the integrand in the expression of $D_{u_s}F$ and the kernel $\mathrm{Ker}\, D_{u_s}F$ of these quadratic form is continuous with respect to s. This follows from our main assumption on γ. Indeed, since every restriction $\gamma|_{[0,s]}$ is strictly abnormal we have that rank of the quadratic form is always equal[3] to n, and the kernel continuously depend on s. ☐

[2] a compact operator in a Hilbert space is diagonalizable and the set of eigenvalues is countable, bounded, and can be ordered in such a way that $\mu_n \to 0$.

[3] a piece of curve γ_s is abnormal if and only if it is a critical point of F, that means that the rank of the derivative is not maximum at this point

Remark 7.30. Notice that (i) implies only that u is local minimizer in the L^2-topology. We will discuss more stronger minimality conditions in next sections.

As we said the definition of exponential map is nothing but the map of the Proposition 7.17. Is natural then to give the following definition:

Definition 7.31. Fix $q_0 \in M$ and consider the exponential map $\mathcal{E} = \mathcal{E}_{q_0}$ starting from q_0. A point $q \neq q_0$ is said *conjugate* to q_0 if q is a critical value for $\mathcal{E}$.

We say that q is *conjugate* to q_0 *along the geodesic* $\gamma(t) = \mathcal{E}(t\lambda)$ if $q = \gamma(s)$ and $s\lambda$ is a critical point of the exponential map $\mathcal{E}$.

Remark 7.32. Recall that $\mathcal{E}(\lambda)$

Proposition 7.33. *Let $\gamma(t)$ be a strongly normal geodesic and $0 \leq s \leq t$. Then $\gamma(s)$ is conjugate to $\gamma(0)$ if and only if* $\mathrm{Hess}_{u_s}\, c\big|_{F^{-1}(\gamma_s)}$ *is degenerate.*

Proof. We apply Proposition 7.17. Indeed $\gamma(s)$ is a conjugate point if and only if u_s is a critical point of the exponential map, that is equivalent to the fact that $\mathrm{Hess}_{u_s}\, c\big|_{F^{-1}(\gamma_s)}$ is degenerate. □

Corollary 7.34. *Let $\gamma(t)$ be a strongly normal geodesic and assume that there are no conjugate points. Then* $\mathrm{Hess}_u c\big|_{F^{-1}(q_1)} > 0$. *In particular $\gamma(t)$ is a local minimizer in the L^2-topology for controls.*

Proof. Indeed, since there are no conjugate points, by Proposition 7.33 it follows that $\mathrm{Hess}_{u_s}\, c\big|_{F^{-1}(\gamma_s)}$ is non degenerate for every $s \in [0,1]$, hence $\mathrm{Hess}_u c\big|_{F^{-1}(q_1)} > 0$ by Proposition 7.29. □

Corollary 7.35. *Let $\gamma(t)$ be a strongly normal geodesic. Then the set $\{s > 0, \gamma(s) \text{ is conjugate}\}$ is isolated from 0.*

Proof. It follows from the fact that small pieces of a normal geodesic are minimizers and Proposition 7.33. □

Hence we have a good characterization of minimizers for the sub-Riemannian distance in terms of conjugate points, but only in the L^2-topology for controls, that is equivalent to the H^1-topology for the trajectories. Now we want to prove that, if there are no conjugate points, the trajectory is also a minimizer in the $\mathcal{C}^0$-topology, that is more strong.

Proposition 7.36. *Let γ be a strongly normal geodesic. If $\gamma(s)$ is not conjugate to $\gamma(0)$ for every $0 < s \leq 1$, then γ is a strong miminum in the $\mathcal{C}^0$-topology for trajectories.*

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 4.55 are satisfied. We will use the following lemma, which we prove at the end of the proposition.

Lemma 7.37. *There exists $a \in \mathcal{C}^\infty(M)$ such that*

$$\lambda_0 = d_{q_0}a, \qquad \mathrm{Hess}_{(q_0,u)} J + a\Big|_{\mathbb{F}^{-1}(\gamma_s)} > 0,$$

In this case $(\mathbb{F}, J + a)$ is a Morse problem and

$$\mathcal{L}_{(\mathbb{F}, J+a)} = \{e^{\vec{H}}(d_q a),\ q \in M\}$$

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 (7.31) we can rewrite this saying that

$$\pi \circ e^{s\vec{H}}\big|_{\mathcal{L}_0} \text{ is an immersion at } \lambda_0, \quad \forall\, s \in [0,1],$$

In particular it is a local diffeomorphism. Hence we can apply the local version of Theorem 4.55. $\square$

Proof of Lemma 7.37. First we notice that

$$\operatorname{Ker} D_{(q_0,u)}\mathbb{F} \subset T_{q_0}M \oplus \mathbf{U}, \qquad \mathbf{U} \quad \text{Hilbert}$$

In particular

$$\operatorname{Ker} D_{(q_0,u)}\mathbb{F} \cap (0 \oplus \mathbf{U}) = \operatorname{Ker} D_u F$$

Since there are no conjugate points, it follows that

$$\operatorname{Hess}_{(q_0,u)} J + a\Big|_{0 \oplus \operatorname{Ker} D_u F} = \operatorname{Hess}_u J > 0 \tag{7.38}$$

Then it is sufficient to show that there exists a choice of the function $a \in \mathcal{C}^\infty(M)$ such that the Hessian is positive definite also in the complement. We define

$$W_s := \{\xi \oplus v \in \operatorname{Ker} D_{(q_0,u_s)}\mathbb{F} |\, \operatorname{Hess}(J+a)(\xi \oplus v, 0 \oplus \operatorname{Ker} D_{u_s} F) = 0\}$$

Notice from (7.38) that, if there is some $\xi \oplus v \in W_s$, then $\xi \neq 0$. Now we prove that there exists a map

$$B_s : T_q M \to \mathbf{U}, \qquad W_s = \{\xi \oplus B_s \xi,\ \xi \in T_q M\}$$

Then we will have

$$\operatorname{Ker} D_{(q_0,u_s)}\mathbb{F} = (0 \oplus \operatorname{Ker} D_{u_s} F) + W_s$$

and we get

$$\begin{aligned}
\operatorname{Hess}(J+a)(\xi \oplus B_s\xi + 0 \oplus v, \xi \oplus B_s \xi + 0 \oplus v) = \\
&= \operatorname{Hess} J(v,v) + \operatorname{Hess}(J+a)(\xi \oplus B_s \xi, \xi \oplus B_s \xi) \\
&= \operatorname{Hess} J(v,v) + d^2 a(\xi,\xi) + Q(\xi)
\end{aligned}$$

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, we can choose a in such a way that it remains positive. $\square$

Remark 7.38. The assumption that the curve γ is strictly normal is essential in what we proved. Indeed if a curve γ is both normal and abnormal we have that there exists two covectors $\lambda_1, \nu_1 \neq 0$ that satisfy

$$\lambda_1 D_u F = u, \qquad \nu_1 D_u F = 0,$$

that implies

$$(\lambda_1 + s\nu_1) D_u F = u, \qquad \forall\, s \in \mathbb{R}$$

and the whole one parameter family of covectors projects on the same geodesic, and γ would be a critical point of the projection. In this case the definition of conjugate point should be changed.

Up to now we proved a sufficient condition for a strictly normal geodesic to be a strong minimum of the sub-Riemannian distance. Indeed Proposition 7.36 says that, if γ contains no conjugate points, then it is optimal with respect to sufficiently $\mathcal{C}^0$-closed curves.

On the other hand, if we consider a control u such that the corresponding trajectory

$$\gamma(t) = q_0 \circ \overrightarrow{\exp} \int_0^t f_{u(s)} ds$$

is strictly normal, that means u is not a critical point of the end-point map F, then it is well defined the Hessian of $J\big|_{F^{-1}(q_1)}$, where $q_1 = F(u)$ at the point u. Moreover, if γ is locally optimal, also in a very weak sense, then necessarily we have

$$\operatorname{Hess}_u J\big|_{F^{-1}(q_1)} \geq 0$$

Indeed if the Hessian is sign-indefinite, then the map is locally open around the point u and we have that small perturbations give rise to a smaller cost.

As in the proof of Proposition 7.29 we consider the family of rescaled controls (and corresponding trajectories)

$$u_s(t) = su(st), \qquad \gamma_s(t) = \gamma(st), \qquad s, t \in [0, 1],$$

and we define the function

$$\alpha(s) = \min_{\|v\|=1} \operatorname{Hess}_{u_s} J\big|_{F^{-1}(\gamma_s(1))}$$

that is well defined, continuous and non-increasing, under the assumption that γ_s is strictly normal for every $s \in [0, 1]$. Notice that $\alpha(\bar{s}) = 0$ if and only if $\gamma(\bar{s})$ is a conjugate point. Since $\alpha(0) = 1$ we have only three cases

(*a*) $\alpha(1) > 0$. By monotonicity this implies $\alpha(s) > 0$ for all s and we have no conjugate points. Hence, by Proposition 7.36, γ is a minimum in the strong topology.

(*b*) $\alpha(1) < 0$. Then the Hessian at u is sign indefinite and γ is not a minimum, also in the weak topology.

(*c*) $\alpha(1) = 0$. In this case the Hessian is semi-definite and we cannot conclude anything on the minimality of γ.

Notice that in cases (*b*) and (*c*) also a segment of conjugate point can appear. To analyze in details case (*c*) and to understand better the properties of a segment of conjugate point we introduce the notion of Jacobi curves, which is some sense generalize the notion of Jacobi fields in Riemannanian geometry. (see Chapter 13)

7.6 Application: Conjugate locus on perturbed $\mathbb{S}^2$

In this section we prove that the conjugate locus of a generic C^2 perturbation of the standard metric on S^2, generates a conjugate locus which has at least 4 cusps. Recall that the conjugate locus from a point q on the standard sphere S^2 coincide with the point that is antipodal to q, where all geodesics starting from q meets and lose their optimality.

Let us then consider a point q_0 on S^2 with a Riemannian Hamiltonian H sufficiently close to H_0 (with respect to the C^2 topology). Normal geodesic starting from q_0 can be parametrized by an angle $\theta \in S^1$, that describes the set of normal extremal paths parametrized by length, or equivalently covectors $\lambda \in T^*_{q_0}M$ such that $H(\lambda) = 1/2$.

For a fixed initial condition $\lambda = (q_0, \theta)$ we have that

$$\lambda(t) = e^{t\vec{H}}(\lambda) = (p(t,\theta), \gamma(t,\theta)),$$

and we denote by $\mathcal{E}_{q_0}$ the exponential map based at q_0

$$\mathcal{E}_{q_0}(t,\lambda) = \pi \circ e^{t\vec{H}}(\lambda) = \gamma(t,\theta)$$

For every initial condition $\theta \in S^1$ let us denote by $\gamma_\theta(t)$ (also $\gamma(t,\theta)$) the normal extremal path associated with θ and starting from q_0, and by $t_c(\theta)$ the first conjugate time along γ_θ. The conjugate locus is the set $\mathrm{Con}(q_0) = \{\gamma(t_c(\theta),\theta), \theta \in S^1\}$.

Proposition 7.39. *The conjugate time along γ_θ is characterized as follows*

$$t_c(\theta) = \min\left\{ t > 0 \;\middle|\; \frac{\partial \mathcal{E}}{\partial \theta}(t,\theta) = 0 \right\}. \tag{7.39}$$

Proof. The conjugate point corresponds to points (t,θ) such that the differential of the exponential map is not surjective, i.e. when

$$\mathrm{rank}\left\{ \frac{\partial \mathcal{E}}{\partial t}(t,\theta), \frac{\partial \mathcal{E}}{\partial \theta}(t,\theta) \right\} = 1. \tag{7.40}$$

Let us show that the two vector cannot be proportional unless $\frac{\partial \mathcal{E}}{\partial \theta}(t,\theta) = 0$. Indeed it follows from Proposition 7.27 that

$$\left\langle p, \frac{\partial \mathcal{E}}{\partial t}(t,\theta) \right\rangle = 1, \qquad \left\langle p, \frac{\partial \mathcal{E}}{\partial \theta}(t,\theta) \right\rangle = 0,$$

thus, whenever $\frac{\partial \mathcal{E}}{\partial \theta}(t,\theta) \neq 0$, the two vectors appearing in (7.40) are always linearly independent. □

Let us now consider solutions of the equation

$$\frac{\partial \mathcal{E}}{\partial \theta}(t,\theta) = 0. \tag{7.41}$$

In other words introduce the function $\beta : \theta \mapsto \mathcal{E}(t_c(\theta),\theta)$. By the chain rule and (7.41), it is easy to see that

$$\beta'(\theta) = t_c'(\theta)\frac{\partial \mathcal{E}}{\partial \theta}(t_c(\theta),\theta) + \underbrace{\frac{\partial \mathcal{E}}{\partial \theta}(t_c(\theta),\theta)}_{=0} \tag{7.42}$$

Let us denote by $g : S^1 \to \mathbb{R}^2$ the function $g(\theta) := \frac{\partial \mathcal{E}}{\partial \theta}(t_c(\theta),\theta)$. When H corresponds to the Hamiltonian H_0 of the standard Riemannian structure on the sphere then the function g describes a circle:

$$g_0(\theta) = \begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix}$$

By assumption the perturbation of the metric is small in the C^2-topology, hence the perturbation does not change the convexity property of g. Then the cuspidal point of the conjugate locus corresponds exactly to those points where the function $\theta \mapsto t_c'(\theta)$ change sign.

Theorem 7.40. *The conjugate locus of the perturbed sphere has at least 4 cuspidal points.*

Proof. Notice that the function $\theta \mapsto t_c'(\theta)$, seen as a periodic function defined on $\mathbb{R}$, can change sign only an even number of times on an interval $[0, 2\pi]$. Moreover it is has zero mean since

$$\int_0^{2\pi} t_c'(\theta)d\theta = t_c(2\pi) - t_c(0) = 0 \tag{7.43}$$

that implies that, if it is not identically zero, it has to change sign at least twice on $[0, 2\pi]$. Notice also that

$$\int_0^{2\pi} t_c'(\theta)g(\theta)d\theta = \int_0^{2\pi} \beta'(\theta)d\theta = \beta(2\pi) - \beta(0) = 0. \tag{7.44}$$

Let us now assume by contradiction that the function $\theta \mapsto t_c'(\theta)$ changes sign exactly twice at points $\theta_1, \theta_2 \in S^1$. Then, by convexity, there exists a covector $\lambda \in (\mathbb{R}^2)^*$ such that $\langle \lambda, g(\theta_i) \rangle = 0$ for $i = 1, 2$ and such that $t_c'(\theta) \langle \lambda, g(\theta) \rangle \geq 0$, that implies in particular

$$\int_0^{2\pi} t_c'(\theta) \langle \lambda, g(\theta) \rangle \, d\theta \neq 0$$

which contradicts (7.44). $\square$

Remark 7.41. The same argument can be applied for every small C^2 perturbation H of the Riemannian Hamiltonian H_0 associated with the standard Riemannian structure on S^2, and not necessarily a quadratic Hamiltonian coming from a Riemannian metric.

7.7 Global minimizers

Before going to the analysis of global minimality of geodesics, let us resume in the following Theorem our results about local minimality.

Theorem 7.42. *Let M be complete and $\gamma(s)$ with $\gamma|_{[0,s]}$ and $\gamma|_{[s,1]}$ strictly normal $0 \leq s \leq 1$.*

(i) if γ has no conjugate point then its a minimizer in the C_0-topology for the trajectories,

(ii) if γ has at least a conjugate point then its not minimizer in the L^2-topology for controls.

Remark 7.43. Notice that the hypotheses of the above theorem imply that in the case (ii) it not possible to have ha segment of full conjugate point up to $t = 1$.

Definition 7.44. We say that a point q is in the cut locus of q_0 if there exists two length minimizers joining q_0 and q.

Our previous analysis of conjugate points let us to state the following result.

Theorem 7.45. *Let M be a complete sub-Riemannian manifold and $\gamma : [0, 1] \to M$ be a normal extremal path. Then*

(i) assume that $\gamma|_{[0,s]}$ is strictly normal for all $s > 0$ and that γ is not a minimizer. Then there exists $\tau \in]0, 1]$ such that $\gamma(\tau)$ is either cut or conjugate to $\gamma(0)$,

(ii) assume that $\gamma|_{[s,1]}$ is strictly normal for all $s > 0$ and that there exists $\tau \in]0,1]$ such that $\gamma(s)$ is either cut or conjugate to $\gamma(0)$. Then γ not a minimizer.

In particular if γ is strongly normal then we have that γ is not a minimizer if and only if there exists a cut or a conjugate point along γ.

Proof. (i). Let us assume that γ is not a minimizer and that there are no conjugate points along γ. We prove that this implies the presence of a cut point. Define

$$t_* := \sup\{t \mid \gamma|_{[0,t]} \text{ is minimizing}\}$$

Let us show that $0 < t_* < 1$. Indeed $t_* > 0$ since small pieces of a normal extremal path are minimizers. Moreover, since $\gamma|_{[0,1]}$ is not a minimizer, by continuity of the distance also $t_* < 1$. Denote by $\alpha_s(\cdot)$ a minimizer joining $\gamma(0)$ to $\gamma(s)$, for each $s > t_*$. The existence of such a minimizers follows from the completeness assumption.

Let us consider the trajectories α_{t_n} for a sequence $t_n \to t_*$. By compactness of minimizers (up to considering a subsequence t_{n_k}, which we still demote by t_n) there exists a limit minimizer $\alpha_{t_n} \to \alpha$ joining $\gamma(0)$ to $\gamma(t_*)$.

Moreover, in the segment $\gamma|_{[0,t_*]}$ there are no conjugate points (by definition of t_*), hence the curve $\gamma|_{[0,t_*]}$ is a minimizer in the strict C_0-topology. Thus α, that by continuity is not shorter than $\gamma|_{[0,t_*]}$, is not contained in a neighborhood γ. From this it follows that $\gamma(t_*)$ is a cut point.

(ii). Assume that there exists a conjugate point $\gamma(\tau)$ in the segment $[0,1]$. Then γ is not a local (hence global) minimizer, as proved in Theorem 7.42. It remains to show that the same remains true if $\gamma(\tau)$ is a cut point. Indeed in this case we have a minimizer $\widehat{\gamma}$ such that $\widehat{\gamma}(\tau) = \gamma(\tau)$. From this it follows that the curve built with $\widehat{\gamma}|_{[0,\tau]}$ and $\gamma|_{[\tau,1]}$ is also a minimizer and the piece $\gamma_{[\tau,1]}$, by uniqueness of the covector, would be associated with two different normal covectors, hence abnormal, that contradicts our assumptions.

□

Theorem 7.46. *Let $\gamma : [0,1] \to M$ be a normal extremal path. Assume that for some $s > 0$*

(i) $\gamma|_{[0,s]}$ is a global minimizer,

(ii) at each point in a neighborhood of $\gamma(s)$ there exists a unique normal minimizer joining $\gamma(0)$ to $\gamma(s)$.

Then $\gamma(s)$ is not conjugate to $\gamma(0)$.

Proof. Let us consider a neighborhood O of $\gamma(s)$ and, for each $q \in O$, let us denote by u^q (resp. γ^q) the minimizing control (resp. trajectory) joining $\gamma(0)$ to q.

The map $q \mapsto u^q$ is continuous in the L^2 topology. Hence we can consider the family λ_1^q of covectors such that

$$\lambda_1^q D_{u^q} F = u^q, \qquad \forall\, q \in O.$$

By the smoothness of F and the contiuity of the map $q \mapsto D_{u^q}F$ we have that the map $q \mapsto \lambda_1^q$ is continuous. Thus the map $q \mapsto \lambda_0^q$ is continuous too, being the composition of the previous one with $(P_{0,1}^*)^{-1}$.

Moreover, the map $q \mapsto \lambda_0^q$ is also injective. Indeed it is an inverse of the exponential map. By the invariance of domain theorem we have that $O' = \{\lambda_0^q, q \in O\}$ is open in T_q^*M.

Thus $(1+\varepsilon)\lambda_0^{\gamma(s)} \in O'$ for $|\varepsilon|$ small enough. This proves that for points in $\gamma([0,1]) \cap \mathcal{O}$ that are close to $\gamma(s)$, the restriction of γ is the unique minimizer. Hence $\gamma(s)$ is not conjugate. □

Chapter 8

Nonholonomic tangent space

8.1 Jet spaces

Consider a smooth curve γ on a smooth manifold M and assume $\gamma(0) = q \in M$ is fixed. In coordinates we can write

$$\gamma(t) = q + t\dot{\gamma}(0) + O(t^2) \tag{8.1}$$

and we defined a tangent vector $v \in T_qM$ as equivalence classes of curves such that, in some coordinate chart, they have the same 1-st order Taylor polynomial.

In the same spirit we can consider, given a smooth curve such that $\gamma(0) = q$, its m-th order Taylor polynomial at q

$$\gamma(t) = q + t\dot{\gamma}(0) + \frac{t^2}{2}\ddot{\gamma}(0) + \ldots + \frac{t^m}{m!}\gamma^{(m)}(0) \tag{8.2}$$

Exercise 8.1. Let γ, γ' be two curves starting from q. We say that γ is (m-)equivalent to γ', and we write $\gamma \sim_m \gamma'$, if their Taylor polynomial of order m in some coordinate chart coincide. Prove that $\sim_m$ is a well-defined equivalence relation on the set of curves starting from q.

Definition 8.2. Let $m > 0$ be an integer and $q \in M$. We define the set of *m-th jets of curves* at point $q \in M$ as the equivalence classes of curves starting from q with respect to $\sim_m$. We denote with $J_q^m\gamma$ the equivalence class of a curve γ and with

$$J_q^m := \{J_q^m\gamma, \gamma \text{ smooth curve on } M\}$$

Remark 8.3. It is easy to show from coordinates representation (8.2) that J_q^m is a smooth manifold and $\dim J_q^m = nm$. Indeed in (8.2) every term $\gamma^{(i)}(0)$ is an n-dimensional vector.

Notice that J_q^m is not a vector space since a change of coordinates does not act linearly on the m-th Taylor polynomial. For instance, given a smooth curve γ such that $\dot{\gamma}(0) \neq 0$ there always exists a coordinate chart in which γ is a straight line, hence higher order derivatives have no intrinsic meaning.

In the following we always assume that $q \in M$ is fixed and we simplify the notation assuming that it coincides with the origin in our coordinate chart. The Taylor expansion then is written as follows

$$J_q^m\gamma = \sum_{i=1}^{m} \frac{t^i}{i!}\gamma^{(i)}(0)$$

To better understand the structure of J_q^m as a smooth manifold we consider the map which "forget about" the m-th derivative

$$\Pi_{m-1}^m : J_q^m \longrightarrow J_q^{m-1}$$
$$\sum_{i=1}^m \frac{t^i}{i!}\gamma^{(i)}(0) \mapsto \sum_{i=1}^{m-1} \frac{t^i}{i!}\gamma^{(i)}(0)$$

Proposition 8.4. *J_q^m is an affine bundle over J_q^{m-1} with projection Π_{m-1}^m, whose fibers are affine spaces over T_qM.*

Fix an element $j \in J_q^{m-1}$, then the fiber at point j is the set of all m^{th}-jets with fixed $m-1$ terms. To show that it is an affine space we should define properly an action of tangent vectors on m^{th}-jets with $(m-1)^{th}$-jet fixed.

The geometric meaning of the fact that J_q^m is an affine bundle (and not an vector bundle) is that we cannot complete in a canonic way a $(m-1)^{th}$-jet to a m^{th}-jet, i.e. we cannot fix an origin in the fiber. On the other hand we can choose as a "global" origin on J_q^m the jet of the constant curve $\gamma(t) \equiv q$.

Proof. Let $j = J_q^m\gamma$ be the m^{th}-jet of a smooth curve in M and let $v \in T_qM$. Extend the vector v to a vector filed $V \in \mathrm{Vec}(M)$ such that $V(q) = v$ and define the action of v on j as

$$J_q^m\gamma + v := J_q^m(e^{t^mV}(\gamma(t)))$$

It is easily seen that
$$J_q^m(e^{t^mV}(\gamma(t))) = J_q^m\gamma + t^mV(q)$$
hence $J_q^{m-1}(\gamma(t))) = J_q^{m-1}(e^{t^mV}(\gamma(t)))$ and one can check that all is well defined. $\square$

Now we want to define dilations on jet spaces, analogously to homothety in Euclidean spaces. Since we have no vector space structure we have to find an appropriate notion

Definition 8.5. Let $\alpha \in \mathbb{R}$ and define $\gamma_\alpha(t) := \gamma(\alpha t)$ for every $t \in \mathbb{R}$. Define the *dilation* of factor α as
$$\delta_\alpha : J_q^m \to J_q^m, \qquad \delta_\alpha(J_q^m\gamma) = J_q^m(\gamma_\alpha)$$

This definition does not depend on the representative and in coordinates it is a *quasi-homogeneous* multiplication
$$\delta_\alpha(\sum_{i=1}^m t^i\xi_i) = \sum_{i=1}^m t^i\alpha^i\xi_i$$

Now we extend the notion of jets also for vector fields. To start with we consider flows on the manifold

Definition 8.6. A *flow* on M is a smooth family of diffeomorphisms

$$P_\cdot = \{P_t \in \mathrm{Diff}(M),\ t \in \mathbb{R}\}, \qquad P_0 = \mathrm{Id}$$

Notice that we do not ask P to be a one parametric group (i.e. $P_t \circ P_s \neq P_{t+s}$) and this in general is charachterized as the flow of the nonautonomous vector field

$$X_t := \frac{d}{d\varepsilon}\Big|_{\varepsilon=0} P_t^{t+\varepsilon}.$$

The set of all flows on M is a group with the point-wise product, i.e. the product of the flows P_t and Q_t is given by

$$(P \circ Q)_t := P_t \circ Q_t$$

Clearly we can act with a flow on a smooth curve on M as $(P.\gamma)(t) = P_t(\gamma(t))$. Moreover, since $P_0 = id$, it makes sense to consider if this action is well-behaved with respect to J_q^m. Indeed the new curve start from q and it is easy to see from the chain rule that $J_q^m(P.\gamma)$ depends only on first m derivatives of γ. Then we can define

Definition 8.7. Let P_t be a smooth flow on M and $j = J_q^m\gamma \in J_q^m$. The *action* of P on J_q^m is defined by

$$P.(J_q^m\gamma) := J_q^m(P.\gamma)$$

It can be easily checked that $(P \circ Q).j = P.(Q.j)$ for every $j \in J_q^m$.

Given a vector field $V \in \mathrm{Vec}(M)$ we want to define its m^{th}-jet $J_q^m V$ which should be an element of $\mathrm{Vec}(J_q^m)$.

Let us denote with e^{tV} the 1-parametric group defined by the flow of V. As we explained we can act on jets

$$e^{\cdot V} : j \mapsto e^{\cdot V}(j)$$

By the way we need a family of flows to acts on a family of curves so we consider the 1-parametric group of flows $s \mapsto e^{stV}$

Definition 8.8. The map $J_q^m V : J_q^m \to TJ_q^m$ is defined as

$$(J_q^m V)(J_q^m\gamma) := \frac{\partial}{\partial s}\Big|_{s=0} e^{\cdot sV}(J_q^m\gamma) = \frac{\partial}{\partial s}\Big|_{s=0} J_q^m(e^{tsV}(\gamma(t))) \tag{8.3}$$

Exercise 8.9. A vector field, when written in coordinates, can be identified with a vector function. Prove that

$$(J_q^m V)(J_q^m\gamma) = \sum_{i=1}^{m} \frac{t^i}{i!}\frac{d^i}{dt^i}\Big|_{t=0}(tV(\gamma(t)))$$

To end this section we study the interplay between dilations and jets of vector fields. Since δ_α is a map on J_q^m its differential $(\delta_\alpha)_*$ acts on elements of $\mathrm{Vec}(J_q^m)$, in particular on jets of vector fields on M. Surprisingly, its action on these fields is linear with respect to α:

Proposition 8.10. $(\delta_\alpha)_*(J_q^m V) = J_q^m(\alpha V) = \alpha\, J_q^m V$

Proof. From the very definition of the differential of a map (see also Chapter 2) we have

$$\begin{aligned}
((\delta_\alpha)_* J_q^m V)(J_q^m \gamma) &= \frac{\partial}{\partial s}\bigg|_{s=0} J_q^m(\delta_\alpha\, e^{tsV}\, \delta_{1/\alpha}(\gamma(t))) \\
&= \frac{\partial}{\partial s}\bigg|_{s=0} J_q^m(\delta_\alpha\, e^{tsV}(\gamma(t/\alpha))) \\
&= \frac{\partial}{\partial s}\bigg|_{s=0} J_q^m(e^{\alpha tsV}(\gamma(t))) \\
&= J_q^m(\alpha V) = \alpha J_q^m V
\end{aligned}$$

□

8.2 Admissible variations

In this section we define the appropriate notion of tangent vector to a sub-Riemannian manifold. Our goal is to define the "tangent structure" to a sub-Riemannian one.

As we know, we can assume that the sub-Riemannian structure is defined by the orthonormal frame $\{f_1, \dots, f_k\}$. Admissible curves on M are maps $\gamma : [0, T] \to M$ such that there exists a control function $u(t) \in L^\infty$ such that

$$\dot\gamma(t) = f_{u(t)}(\gamma(t)) = \sum_{i=1}^{k} u_i(t) f_i(\gamma(t))$$

To have a good definition of tangent vector we could not restrict to family of admissible curves, because in this way we loose all the information about directions that are not in the distribution. Indeed we want the tangent space to be a first order approximation of the structure, containing informations about all directions.

We need a proper definition of tangent vector, that means a proper definition of variation of a point, in order to give a precise meaning to its "principal term", that is going to be the "tangent vector".

The idea is to introduce the notion of admissible variation

Definition 8.11. A curve $\gamma : [0, T] \to M$ such that $\gamma(0) = q$ is said a *smooth admissible variation* of q if there exists a family of controls $\{u(t, s)\}_{t\in[0,T]}$ such that

(*i*) $u(t, \cdot)$ is measurable and bounded for all $t \in [0, T]$,

(*ii*) $u(\cdot, s)$ is smooth with bounded derivatives, for all $s \in [0, \tau]$,

(*iii*) $u(0, s) = 0$ for all $s \in [0, \tau]$,

(*iv*) $\gamma(t) = q \circ \overrightarrow{\exp} \int_0^\tau f_{u(t,s)} ds$

In other words γ is an admissible variation if it can be parametrized as the final point of a smooth family of admissible curves. We stress that γ is not admissible, in general.

Remark 8.12. We recall that two distributions are said to be equivalent (see also Definition 3.3 and 3.17) if and only if the corresponding modulus of horizontal vector fields are isomorphic $\mathcal{D} \simeq \mathcal{D}'$, where we recall that

$$\mathcal{D} = \text{span}\{f(\sigma),\ \sigma \text{ smooth section of } \mathbf{U}\}.$$

which is finitely generated by a basis $f_1, \ldots, f_k$.

Now we show that the definition of admissible variation does not depend on the frame $f_1, \ldots, f_k$. Notice that $\gamma(t)$ is an admissible variation if $\gamma(t) = q(t, \tau)$ where $q(t, s)$ is a solution of

$$\frac{\partial}{\partial s} q(t,s) = \sum_{i=1}^{k} u_i(t,s) f_i(q(t,s)), \qquad s \in [0, \tau]$$

Let now $\widetilde{f}_1, \ldots, \widetilde{f}_k$ be another set of local generators of the modulus. There exist functions $a_{ij} \in \mathcal{C}^\infty(M)$ such that

$$\widetilde{f}_i(q) = \sum_{j=1}^{k} a_{ij}(q) f_j(q), \qquad \forall\, q \in M, \qquad \forall\, i = 1, \ldots, k \tag{8.4}$$

and assume that γ is an admissible variation with respect to $u(t,s)$ in this new frame, i.e.

$$\frac{\partial}{\partial s} q(t,s) = \sum_{i=1}^{k} u_i(t,s) \widetilde{f}_i(q(t,s)), \qquad s \in [0, \tau] \tag{8.5}$$

Now we prove that there exist a control $\widetilde{u}(t,s)$ such that γ is an admissible variation of the old frame with respect to this control. From (11.23) we get

$$\begin{aligned} \widetilde{f}(u,q) &= \sum_i u_i(t,s) \widetilde{f}_i(q) \\ &= \sum_{i,j} u_i(t,s) a_{ij}(q) f_j(q) \\ &= \sum_j v_j(t,s,q) f_j(q) \\ &= f(v(u,q), q) \end{aligned}$$

Then we could define, using the solution $q(t,s)$ of (8.5), the new control

$$\widetilde{u}_j(t,s) = \sum_i u_i(t,s) a_{ij}(q(t,s))$$

and we see from identities above that

$$\frac{\partial}{\partial s} q(t,s) = \sum_{i=1}^{k} \widetilde{u}_j(t,s) f_j(q(t,s)), \qquad s \in [0, \tau] \tag{8.6}$$

Note. We assume that the sub-Riemannian structure is bracket generating at q and let m the degree of nonholonomy of the distribution, i.e. such that $\mathcal{D}^m_q = T_q M$.

Definition 8.13. The set of *admissible jets* with respect to the sub-Riemannian structure is

$$J_q^f := \{J_q^m \gamma,\ \gamma \text{ is an admissible variation}\}$$

Example 8.14. Consider two vector fields $X, Y \in \mathrm{Vec}(M)$ and the curve

$$\gamma : t \mapsto e^{-tY} \circ e^{-tX} \circ e^{tY} \circ e^{tX}(q)$$

It is easily seen that γ is an admissible variation if we set

$$\gamma(t) = \overrightarrow{\exp} \int_0^4 f_{tv(s)}(q) ds$$

where

$$v(s) = \begin{cases} (1,0), & \text{if } s \in [0,1], \\ (0,1), & \text{if } s \in [1,2], \\ (-1,0), & \text{if } s \in [2,3], \\ (0,-1), & \text{if } s \in [3,4]. \end{cases}$$

In coordinates we have expansion $\gamma(t) = q + t^2[X,Y] + o(t^2)$.

Now we want to describe the nonholonomic tangent space in an intrinsic coordinate free way. Then we will see how it can be described in special coordinates.

Definition 8.15. The *group of flows of admissible variations* is

$$\mathcal{P}^f := \left\{ \overrightarrow{\exp} \int_0^\tau f_{u(t,s)} ds,\ u(t,s) \text{ smooth variation} \right\}$$

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 $\gamma(t) \equiv q$ for all t. Then we have

$$J_q^f = \{J_q^m(P_.(q)), P_. \in \mathcal{P}^f\}$$

and the set of admissible jets is exactly the orbit of q under the action of the group $\mathcal{P}^f$.

Remark 8.16. It is easy to see that $\mathcal{P}^f$ is a group since the following equality holds

$$\overrightarrow{\exp} \int_0^{\tau_1} f_{u(t,s)} ds \circ \overrightarrow{\exp} \int_0^{\tau_2} f_{v(t,s)} ds = \overrightarrow{\exp} \int_0^{\tau_1+\tau_2} f_{w(t,s)} ds$$

where

$$w(t,s) = \begin{cases} u(t,s), & 0 \le s \le \tau_1, \\ v(t, s-\tau_1), & \tau_1 \le s \le \tau_1 + \tau_2. \end{cases}$$

is the concatenation of controls.[1]

Now we want to describe the tangent space as the quotient of this set with respect to some subgroup of "slow" flows. The heuristic idea is that a flow is slow if the first nonzero jet of its associated trajectory $J_q^i \gamma$ belong to a subspace Δ^j, with $j < i$.

[1] Here we see that is useful not to fix τ in the definition, otherwise we need to rescale controls.

Definition 8.17. Let $P^{\cdot} \in \mathcal{P}^f$. $P^{\cdot}$ is said to be *purely slow* if it is associated to a smooth variation $u(t,s)$ such that satisfies $u(0,s) = \dfrac{\partial u}{\partial t}(0,s) = 0$.

The *subgroup of slow flows* is the normal subgruop of $\mathcal{P}^f$ generated by purely slow flows, i.e.

$$\mathcal{P}_0^f := \{(Q_t)^{-1} \circ P_t \circ Q_t,\ Q_t \in \mathcal{P}^f,\ P_t \text{ purely slow}\}$$

Remark 8.18. Notice that, from the definition and the linearity of f, a purely slow flow can be written as follows: $u(t,s) = tv(t,s)$, where $v(0,s) = 0$. Moreover we have

$$P_t = \overrightarrow{\exp}\int_0^\tau f_{u(t,s)}ds = \overrightarrow{\exp}\int_0^\tau f_{tv(t,s)}ds = \overrightarrow{\exp}\int_0^\tau tf_{v(t,s)}ds = t\,\overrightarrow{\exp}\int_0^\tau f_{v(t,s)}ds,$$

Definition 8.19. Let $\gamma, \widetilde{\gamma}$ be two curves on M. We say that $J_q^m\gamma$ and $J_q^m\widetilde{\gamma}$ are equivalent if $\widetilde{\gamma}(t) = P_t(\gamma(t))$ for some $P_t \in \mathcal{P}_0^f$. The *nonholonomic tangent space* T_q^f is defined as

$$T_q^f := J_q^f / \sim$$

We end this section with the coordinate presentation of jets of horizontal vector fields of the sub-Riemannian structure

Proposition 8.20. *Let $X \in \mathcal{D}$ be an horizontal vector field for the sub-Riemannian structure on M. Then the one parametric group e^{tX} acts on the set J_q^f. Moreover the action is well defined on the equivalence classes with respect to $\sim$.*

Proof. From the very definition of J_q^f it is easy to see that if $J_q^m\gamma$ is the jet of an admissible variation then the right hand side of (8.3) is an admissible variation for every s. We are left to show that if

$$\gamma(t) \sim \gamma'(t) \implies e^{tX}\gamma(t) \sim e^{tX}\gamma'(t).$$

From our assumption we get $\gamma'(t) = \gamma(t) \circ Q_t$ for a slow flow $Q_{\cdot} \in \mathcal{P}_0^f$. It follows that

$$\begin{aligned}
\gamma'(t) \circ e^{tX} &= \gamma(t) \circ Q_t \circ e^{tX} \\
&= \gamma(t) \circ e^{tX} \circ e^{-tX} \circ Q_t \circ e^{tX} \\
&= (\gamma(t) \circ e^{tX}) \circ \widetilde{Q}_t
\end{aligned}$$

where $\widetilde{Q}_t := e^{-tX} \circ Q_t \circ e^{tX}$ is also a slow flow. This shows that e^{tX} is independent on the representative and is well defined on the quotient. $\square$

8.3 Nilpotent approximation and privileged coordinates

In this section we want to introduce some coordinates in which we have a good description of the nonholonomic tangent space.

Consider some non negative integers $k_1, \ldots, k_m$ such that $n = k_1 + \ldots + k_m$ and the splitting

$$\mathbb{R}^n = \mathbb{R}^{k_1} \oplus \ldots \oplus \mathbb{R}^{k_m}, \qquad x = (x_1, \ldots, x_m)$$

where every $x_i = (x_i^1, \ldots, x_i^{k_i}) \in \mathbb{R}^{k_i}$.

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 $i = 1, \ldots, m;\ j = 1, \ldots, k_i$. We define weights of generators as

$$\nu(1) = 0, \qquad \nu(x_i^j) = i, \qquad \nu(\frac{\partial}{\partial x_i^j}) = -i.$$

Then for any monomial

$$\nu(y_1 \cdots y_\alpha \frac{\partial^\beta}{\partial z_1 \cdots \partial z_\beta}) = \sum_{i=1}^{\alpha} \nu(y_i) - \sum_{j=1}^{\beta} \nu(z_j).$$

We say that a polynomial differential operator D is *homogeneous* if it is a sum of monomial terms all of same weight.

Lemma 8.21. *Let D_1, D_2 be two homogeneous differential operators. Then $D_1 \circ D_2$ is homogeneous and*

$$\nu(D_1 \circ D_2) = \nu(D_1) + \nu(D_2) \tag{8.7}$$

Proof. It is sufficent to check for monomials of kind $D_1 = \frac{\partial}{\partial x_{i_1}^{j_1}}$ and $D_2 = x_{i_2}^{j_2}$ and formula (8.7) follows from identity

$$\frac{\partial}{\partial x_{i_1}^{j_1}} \circ x_{i_2}^{j_2} = x_{i_2}^{j_2} \frac{\partial}{\partial x_{i_1}^{j_1}} + \frac{\partial x_{i_2}^{j_2}}{\partial x_{i_1}^{j_1}}$$

□

A special case is when we consider vector fields. 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 multiindex 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

$$\mathcal{D}^{(h)} = \{D \in \mathrm{Der}(\mathbb{R}^n) : D^{(i)} = 0, \forall\, i < h\}, \qquad h \in \mathbb{Z}$$

It is easy to see that it is a decreasing filtration, i.e. $\mathcal{D}^{(h)} \subset \mathcal{D}^{(h-1)}$ for every h, and 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*) $\mathrm{Vec}(\mathbb{R}^n) \subset \mathcal{D}^{(-m)}$,

(*ii*) $V \in \mathrm{Vec}(\mathbb{R}^n) \cap \mathcal{D}^{(0)}$ implies $V(0) = 0$.

and every vector field that is not zero at the origin is necessarily in $D^{(-1)}$. This motivates the folowing definition

Definition 8.22. A system of coordinates near the point q is said *privileged* for a sub-Riemannian structure M if the following conditions are satisfied

(*i*) $\mathcal{D}^i_q = \mathbb{R}^{k_1} \oplus \ldots \oplus \mathbb{R}^{k_i}, \qquad \forall\, i = 1, \ldots, m,$

(*ii*) $f \in \mathcal{D}^{(-1)}$ for every $f \in \mathcal{D}$.

Condition (*i*) says that our coordinates are *linearly adapted* to the flag $\mathcal{D}^1_q \subset \mathcal{D}^2_q \subset \ldots \subset \mathcal{D}^m_q$. Notice that this condition can be always satisfied with a linear change of coordinates.

Example 8.23. We analyze the meaning of privileged coordinates in the easiest cases $m = 1, 2$ and we show that in general not all system of linearly adapted coordinates are privileged.

(1) If $m = 1$ all sets of coordinates are privileged because $\mathrm{Vec}(M) \subset \mathcal{D}^{(-1)}$ since $\nu(\partial_{x_i}) = -1$ for all i.

(2) If $m = 2$ then all systems of coordinates that are linearly adapted to the flag are privileged. Indeed we have $\nu(\partial_{x_1^j}) = -1$ and $\nu(\partial_{x_2^j}) = -2$ and a vector field that belong to $\mathcal{D}^{(-2)} \setminus \mathcal{D}^{(-1)}$ must contain a monomial of the second kind, with constant coefficient. On the other hand vector fields $f_1, \ldots, f_k$ cannot contain such a monomial since, by our assumption

$$\mathrm{span}\{f_1(0), \ldots, f_k(0)\} = \mathcal{D}^1_0 = \mathbb{R}^{k_1}.$$

(3) Let we consider the following set of vector fields in $\mathbb{R}^3 = \mathbb{R} \oplus \mathbb{R} \oplus \mathbb{R}$

$$f_1 = \partial_{x_1} + x_1 \partial_{x_3}, \qquad f_2 = x_1 \partial_{x_2}, \qquad f_3 = x_2 \partial_{x_3}$$

where we put $\nu(x_i) = i$ for $i = 1, 2, 3$. All nontrivial commutators are computed as follows

$$[f_1, f_2] = \partial_{x_2}, \qquad [f_2, f_3] = x_1 \partial_{x_3}, \qquad [[f_1, f_2], f_3] = \partial_{x_3},$$

and it is easy to see that the flag (computed at $x = 0$) is

$$\mathcal{D}^1_0 = \mathrm{span}\{\partial_{x_1}\}, \qquad \mathcal{D}^2_0 = \mathrm{span}\{\partial_{x_1}, \partial_{x_2}\}, \qquad \mathcal{D}^3_0 = \mathrm{span}\{\partial_{x_1}, \partial_{x_2}, \partial_{x_3}\}$$

Then this set of coordinates are linearly adapted to the flag but are not privileged since $\nu(x_1 \partial_{x_3}) = -2$

Theorem 8.24. *Let M be a sub-Riemannian manifold and $q \in M$. There always exists a system of privileged coordinates near q.*

We postpone the proof of this theorem to the end of this section, after having analyzed the meaning of privileged coordinates.

Theorem 8.25. *Let M be a sub-Riemannian manifold and $q \in M$. In privileged coordinates we have the following*

(i) $J_q^f = \{\sum_{i=1}^m t^i \xi_i, \xi_i \in \mathcal{D}_q^i\}$ *and* $\dim J_q^f = mk_1 + (m-1)k_2 + \ldots + k_m$.

(ii) *Let* $j_1, j_2 \in J_q^f$. *Then* $j_1 \sim j_2$ *if and only if* $j_1 - j_2 = \sum_{i=1}^m t^i \eta_i$, *where* $\eta_i \in \mathcal{D}_q^{i-1}$.

First part of proof of Theorem 8.25. We start by proving the inclusion $J_q^f \subset \{\sum_{i=1}^m t^i \xi_i, \xi_i \in \mathcal{D}_q^i\}$. For any smooth variation $\gamma(t)$ we can write

$$\gamma(t) = q \circ \overrightarrow{\exp} \int_0^\tau f_{u(t,s)} ds$$

Taylor expansion leads to

$$\gamma(t) = q + \sum_{j=1}^{i} \int \cdots \int_{0 \le s_j \le \ldots \le s_1 \le s} q \circ f_{u(t,s_1)} \circ \ldots \circ f_{u(t,s_j)} \, ds_1 \ldots ds_j + O(t^{i+1})$$

Indeed using the fact that f is linear in u, we can factor out t from every term since $u(0,s) = 0$. If we want compute our curve in privileged coordinates (to compute weights) it is sufficient to apply all to the coordinate function. In particular, since $f_u \in \mathcal{D}^{(-1)}$ we have that

$$f_{u(t,s_1)} \circ \ldots \circ f_{u(t,s_j)} \in \mathcal{D}^{(-i)}$$

and applying to a coordinate function x_α^β, where $\alpha = 1, \ldots, m$ and $\beta = 1, \ldots, k_\alpha$ we have

$$f_{u(t,s_1)} \circ \ldots \circ f_{u(t,s_j)} x_\alpha^\beta \in \mathcal{D}^{(-i+\alpha)}$$

because $\nu(x_\alpha^\beta) = \alpha$. Then, if $\alpha > i$ we have that this function has positive weight. Thus, when evaluated at $x = 0$ it is zero.

In other words we proved that, for every $i = 1, \ldots, m$, up to the i^{th}-term we can find only element in $\mathcal{D}_q^i$. □

To prove the converse inclusion we have to show that, given some elements $\xi^i \in \mathcal{D}_q^i$ we can find a smooth variation that has these vectors as elements of its jet. We start with some preliminary lemmas.

Lemma 8.26. *Let m, n be two integers. Assume that we have two flows*

$$P_t = \mathrm{Id} + t^n V + O(t^{n+1})$$
$$Q_t = \mathrm{Id} + t^m W + O(t^{m+1})$$

Then $P_t Q_t P_t^{-1} Q_t^{-1} = \mathrm{Id} + t^{n+m}[V, W] + O(t^{n+m+1})$.

Proof. Denoting V_t the nonautonomous vector filed associated to P_t it is easily check that

$$V_t = nt^{n-1} V + O(t^n)$$

Moreover for the inverse flow we have

$$P_t^{-1} = \mathrm{Id} - t^n V + O(t^{n+1})$$
$$Q_t^{-1} = \mathrm{Id} - t^m W + O(t^{m+1})$$

Define $R(t,s) := P_t Q_s P_t^{-1} Q_s^{-1}$. Since $P_0 = Q_0 = \mathrm{Id}$ we have that R is constant on the axes,i.e. $R(0,s) = R(t,0) = \mathrm{Id}$. Hence the only derivative that enter in our expansion, that coincide with $F(t) = R(t,t)$, are mixed derivatives. This remark let us to expand the product $P_t Q_t P_t^{-1} Q_t^{-1}$ and consider only terms with mixed power of t and s to get

$$(\mathrm{Id} + t^n V + O(t^{n+1}))(\mathrm{Id} + t^m W + O(t^{m+1}))(\mathrm{Id} - t^n V + O(t^{n+1}))(\mathrm{Id} - t^m W + O(t^{m+1})) =$$

$$= \mathrm{Id} + t^n s^m (VW - WV) + \dots$$
$$= \mathrm{Id} + t^n s^m [V, W]$$

and the lemma is proved. □

Lemma 8.27. *For all $l \geq h$ and $\forall\, i_1, \dots, i_h \in \{1, \dots, k\}$, there exists an admissible variation $u(t,s)$ such that*

$$q \circ \overrightarrow{\exp} \int_0^\tau f_{u(t,s)} ds = q + t^l [f_{i_1}, \dots, [f_{i_{h-1}}, f_{i_h}]](q) + O(t^{l+1})$$

Proof. By induction

- $\forall\, l \geq 1$ and $\forall\, i = 1, \dots, k$ there exists an admissible variation $u(t,s)$ such that

$$q \circ \overrightarrow{\exp} \int_0^\tau f_{u(t,s)} ds = q + t^l f_i(q) + O(t^{l+1})$$

  It is sufficient to consider $u = (u_1, \dots, u_k)$ where $u_i = t^l$ and $u_h = 0$ for all $h \neq i$.

- $\forall\, l \geq 2$ and $\forall\, i, j = 1, \dots, k$, there exists an admissible variation $u(t,s)$ such that

$$q \circ \overrightarrow{\exp} \int_0^\tau f_{u(t,s)} ds = q + t^l [f_i, f_j](q) + O(t^{l+1})$$

  It is sufficient to use the previous lemma where P_t and Q_t are flows respectively of nonautonomous vector fields $V_t = t^{l-1} f_{i_1}$ and $W_t = t f_{i_2}$.

  With analogous arguments we can prove by induction the lemma

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 $\mathcal{D}^i$ id the linear span of all i-th order brackets). □

Remark 8.28. The previuous construction of $u(t,s)$ does *not* depend on the sub-Riemannian structure but only on the structure of the Lie bracket.

Lemma 8.29. *Let $\pi = \pi(f_1, \dots, f_k)$ a bracket polynomial of degree $\deg \pi \leq l$. There exists an admissible variation $u(t,s)$ such that*

$$q \circ \overrightarrow{\exp} \int_0^\tau f_{u(t,s)} ds = q + t^l \pi(f_1, \dots, f_k)(q) + O(t^{l+1})$$

Proof. Let $\pi(f_1, \dots, f_k) = \sum_j V_j(f_1, \dots, f_k)$ where V_j are monomials. By our previous argument we can find $u^j(t,s), s \in [0, \tau_j]$ such that

$$q \circ \overrightarrow{\exp} \int_0^\tau f_{u^j(t,s)} ds = q + t^l V_j(f_1, \dots, f_k)(q) + O(t^{l+1})$$

Now consider the concatenation of controls $u(t,s)$, where $s \in [0, \tau]$ and $\tau = \sum \tau_j$ defined as follows

$$u(t,s) = u^j(t, s - \sum_{i=1}^{j} \tau_i), \qquad \text{if } \tau_j \leq s \leq \tau_{j+1}$$

Exercise 8.30. End the previous proof showing that the flow relative to u has as l-th term $\sum_j V_j$. Then prove, by rescaling that also any monomial of type αV can be presented.

□

Now we can complete the proof of the first statemet of Theorem 8.25 proving the following inclusion $\{\sum_{i=1}^m t^i \xi_i, \xi_i \in \mathcal{D}_q^i\} \subset J_q^f$.

Second part of Theorem 8.25. Let we consider a m-th jet $\sum_{i=1}^m t^i \xi_i, \xi_i \in \mathcal{D}_q^i$. We prove by induction

- From previous lemmas there exists an admissible variation $\gamma(t)$ such that

  $$\gamma(t) = q \circ \overrightarrow{\exp} \int_0^\tau f_{u(t,s)} ds, \qquad \dot{\gamma}(t) = \xi_1$$

  Then we will have $\gamma(t) = t\xi_1 + t^2\eta_2 + \dots$ where $\eta_2 \in \mathcal{D}^2$ from first part of the proof. We want to correct the second order term

- From previous lemma there exists an admissible variation $\gamma_1(t)$ such that

  $$\gamma_1(t) = q \circ \overrightarrow{\exp} \int_0^\tau f_{v(t,s)} ds, \qquad \gamma(t) = t^2(\xi_2 - \eta_2) + o(t^2)$$

  Defining $\gamma_2(t) = \gamma_1(t) \circ \gamma(t)$ we have

  $$\begin{aligned} \gamma_2(t) &\simeq t\xi_1 + t^2\eta_2 + t^2(\xi_2 - \eta_2) + t^3\eta_3 \\ &\simeq t\xi_1 + t^2\xi_2 + t^3\eta_3 \end{aligned}$$

  where $\eta_3 \in \mathcal{D}^3$.

At every step we can correct the right term of the jet and prove the inclusion.

(ii) We have to prove that

$$j \sim j' \Longleftrightarrow j - j' = \sum_{i=1}^{m} t^i \eta_i, \qquad \eta_i \in \mathcal{D}_q^{i-1}.$$

$(\Rightarrow)$. Assume that $j \sim j'$, where $j = J_q^m \gamma = \sum t^i \xi_i$ and $j' = J_q^m \gamma' = \sum t^i \xi_i'$. 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 $P^i \in \mathcal{P}^f, i = 1, \ldots, h$. For simplicity we prove only the case $h = 1$. By formula (6.19) we have that

$$Q_t = P_t \circ \overrightarrow{\exp} \int_0^\tau f_{tv(t,s)} ds \circ P_t^{-1} = \overrightarrow{\exp} \int_0^\tau P_t \circ f_{tv(t,s)} \circ P_t^{-1} ds$$

then by linearity of f we have

$$Q_t = \overrightarrow{\exp} \int_0^\tau t \,\mathrm{Ad}\, P_t \, f_{v(t,s)} ds$$

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 (6.17) we get

$$Q_t = \overrightarrow{\exp} \int_0^\tau t \ \overrightarrow{\exp} \int_0^s \mathrm{ad} f_{w(t,\theta)} d\theta \ f_{v(t,s)} ds$$

Finally, if $\gamma(t) = q \circ \overrightarrow{\exp} \int_0^\tau f_{u(t,s)} ds$ we can write

$$\gamma'(t) = q \circ \overrightarrow{\exp} \int_0^\tau f_{u(t,s)} ds \circ \overrightarrow{\exp} \int_0^\tau t \ \overrightarrow{\exp} \int_0^s \mathrm{ad} f_{w(t,\theta)} d\theta \ f_{v(t,s)} ds$$

Expanding with respect to $tQ_t \simeq (Id + t \sum t^i V_i) = Id + \sum t^{i+1} V_i$ 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 $j = J_q^m \gamma = \sum t^i \xi_i$ and $j' = J_q^m \gamma' = \sum t^i \xi_i'$, with

$$j - j' = \sum_{i=1}^{m} t^i \eta_i, \qquad \eta_i \in \mathcal{D}_q^{i-1}.$$

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 $\sum_{i=1}^m t^i \eta_i,\ \eta_i \in \mathcal{D}_q^{i-1}$. To this purpose we can repeat arguments of proof of part (i), using the following

Lemma 8.31. *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 $\mathcal{P}_0^f$. ☐

☐

Corollary 8.32. *In privileged coordinates $(x_1,\dots,x_m)$ defined by the splitting $\mathbb{R}^n=\mathbb{R}^{k_1}\oplus\mathbb{R}^{k_m}$ we have*

$$J_q^f=\left\{\begin{pmatrix} tx_1+O(t^2)\\ t^2x_2+O(t^3)\\ \vdots\\ t^m x_m\end{pmatrix}, x_i\in\mathbb{R}^{k_i}, i=1,\dots,m\right\}$$

Proof. Indeed we know that $\mathcal{D}^i=\mathbb{R}^{k_1}\oplus\mathbb{R}^{k_i}$ and writing

$$\xi_i=x_{i,1}+\ldots+x_{i,i},\qquad x_{i,j}\in\mathbb{R}^{k_j}$$

we have, expanding and collecting terms

$$\begin{aligned}\sum t^i\xi_i&=t\xi_1+t^2\xi_2+\ldots+t^m\xi_m\\ &=tx_{1,1}+t^2(x_{2,1}+x_{2,2})+\ldots+t^m(x_{m,1}+\ldots+x_{m,m})\\ &=(tx_{1,1}+t^2x_{2,1}+\ldots+t^mx_{m,1},t^2x_{2,2}+\ldots+t^mx_{m,2},t^mx_{m,m})\end{aligned}$$

□

Corollary 8.33. *The nonholonomic tangent space T_q^f is a smooth manifold of dimension $\dim T_q^f=\sum_{i=1}^{m(q)}k_i(q)$. In privileged coordinates we can write*

$$T_q^f=\left\{\begin{pmatrix} tx_1\\ t^2x_2\\ \vdots\\ t^m x_m\end{pmatrix}, x_i\in\mathbb{R}^{k_i}, i=1,\dots,m\right\}$$

and dilations δ_α acts on T_q^f in a quasi-homogeneous way

$$\delta_\alpha(tx_1,\dots,t^mx_m)=(\alpha tx_1,\dots,\alpha^mt^mx_m),\qquad \alpha>0.$$

Proof. It follows directly from the representation of the equivalence relation. Indeed two elements j and j' can be written in coordinates as

$$\begin{aligned}j&=(tx_1+O(t^2),t^2x_2+O(t^3),\dots,t^mx_m)\\ j'&=(ty_1+O(t^2),t^2y_2+O(t^3),\dots,t^my_m)\end{aligned}$$

and $j\sim j'$ if and only if $x_j=y_j$ for all j. □

Remark 8.34. 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

$$\delta_t(x_1,\dots,x_m)=(tx_1,t^2x_2,\dots,t^mx_m),\qquad t>0. \tag{8.8}$$

In particular for a homogeneous vector field X of weight h it holds $\delta_*X=t^{-h}X$.

Now we can improve Proposition 8.20 and see that actually the jet of a horizontal vector field is a vector field on the tangent space and belongs to $\mathcal{D}^{(-1)}$ (in privileged coordinates).

Lemma 8.35. *Fix a set of privileged coordinate. Let $V \in \mathcal{D}^{(-1)}$, then the jet $J_q^m V$ is tangent to the submanifold J_q^f. Moreover it is well defined as vector field $\widehat{V}$ on the nonhonolomic tangent space. In other words $\widehat{V} \in \mathrm{Vec}(T_q^f)$ and we have*

$$V = \begin{pmatrix} v_1(x) \\ v_2(x) \\ \vdots \\ v_m(x) \end{pmatrix} \quad \Longrightarrow \quad \widehat{V} = \begin{pmatrix} \widehat{v}_1(x) \\ \widehat{v}_2(x) \\ \vdots \\ \widehat{v}_m(x) \end{pmatrix} \tag{8.9}$$

where $\widehat{v}_i$ is the $i-1$ order term of v_i.

Proof. Let $V \in \mathcal{D}^{(-1)}$ and $\gamma(t)$ be an admissible variation. When expressed in coordinates we have (see ...)

$$V = \begin{pmatrix} v_1(x) \\ v_2(x) \\ \vdots \\ v_m(x) \end{pmatrix}, \qquad \gamma(t) = \begin{pmatrix} tx_1 + O(t^2) \\ t^2 x_2 + O(t^3) \\ \vdots \\ t^m x_m \end{pmatrix}$$

We know that $(J_q^m V)(J_q^m \gamma)$ is expressed as the m-th jet of $tV(\gamma(t))$ by Exercise ...Hence we compute

$$(J_q^m V)(J_q^m \gamma) = \begin{pmatrix} tv_1(tx_1 + O(t^2), \dots, t^m x_m) \\ tv_2(tx_1 + O(t^2), \dots, t^m x_m) \\ \vdots \\ tv_m(tx_1 + O(t^2), \dots, t^m x_m) \end{pmatrix} \tag{8.10}$$

Notice that $V \in \mathcal{D}^{(-1)}$ means exactly that

$$V = \sum v_i(x) \frac{\partial}{\partial x_i} = \sum v_i^j(x) \frac{\partial}{\partial x_i^j}, \qquad \nu(\frac{\partial}{\partial x_i^j}) = -i$$

and v_i is a function of order at least $i-1$. Let we 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, then evaluate and neglect the higher order part (that corresponds to the projection on the factor space) in order to have

$$v_i(tx_1, \dots, t^m x_m) = t^{i-1} \widehat{v}_i(x_1, \dots, x_m) + O(t^i)$$

and using equality we have

$$(J_q^m V)\Big|_{T_q^f} = \begin{pmatrix} tv_1(tx_1, \dots, t^m x_m) \\ tv_2(tx_1, \dots, t^m x_m) \\ \vdots \\ tv_m(tx_1, \dots, t^m x_m) \end{pmatrix} = \begin{pmatrix} t\widehat{v}_1 + O(t^2) \\ t^2 \widehat{v}_2 + O(t^3) \\ \vdots \\ t^m \widehat{v}_m + O(t^{m+1}) \end{pmatrix} \tag{8.11}$$

From this easily follows (8.9). □

Remark 8.36. 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

$$\widehat{V}(x) = \begin{pmatrix} \widehat{v}_1 \\ \widehat{v}_2(x_1) \\ \vdots \\ \widehat{v}_m(x_1, \ldots, x_{m-1}) \end{pmatrix} \tag{8.12}$$

Moreover the flow of a vector field of this kind can be easily computed by a step by step substitution.

Now we prove existence of privileged coordinates

Proof of Theorem 8.24. Consider our sub-Riemannian structure on M defined by the orthonormal frame $\{f_1, \ldots, f_k\}$ and its flag $\mathcal{D}_q^1 \subset \mathcal{D}_q^2 \subset \ldots \subset \mathcal{D}_q^m = T_qM$, with

$$n_j := \dim \mathcal{D}_q^j \qquad (n_j = k_1 + \ldots + k_j)$$

Let we consider a basis $\{V_1, \ldots, V_n\}$ of the tangent space adapted to the flag, i.e.

$$V_i = \pi_i(f_1, \ldots, f_k)$$
$$\pi_i \text{ bracket polynomial,} \quad \deg \pi_i \leq j \quad \text{if } i \leq n_j$$
$$\mathcal{D}_q^j = \operatorname{span}\{V_1(q), \ldots, V_{n_j}(q)\}, \qquad j = 1, \ldots, m$$

In particular $V_1, \ldots, V_{n_1}$ are selected in $\{f_i, i = 1, \ldots, k\}$, $V_{n_1+1}, \ldots, V_{n_2}$ are selected from $\{[f_i, f_j], i, j = 1, \ldots, k\}$ and so on.

Define the map

$$\Psi : (s_1, \ldots, s_n) \mapsto q \circ e^{s_1 V_1} \circ \ldots \circ e^{s_n V_n} \tag{8.13}$$

We want to show that Ψ^{-1} defines privileged coordinates around q. It is easy to show that (8.13) is a local diffeomorphism since

$$\frac{\partial \Psi}{\partial s_i}\Big|_{s=0} = \Psi_* \frac{\partial}{\partial s_i}\Big|_{s=0} = V_i(q), \qquad i = 1, \ldots, n \tag{8.14}$$

Hence it remains to show that

(*i*) $\Psi_*^{-1}(\mathcal{D}_q^i) = \operatorname{span}\{\frac{\partial}{\partial s_1}, \ldots, \frac{\partial}{\partial s_{n_i}}\}$,

(*ii*) $\Psi_*^{-1} f_i \in \mathcal{D}^{(-1)}$ for every $i = 1, \ldots, k$

Part (*i*) easily follows from our choice of adapted frame to the flag and (8.14). On the other hand the second part is not trivial since we need to compute differential of Ψ at every point and not only at $s = 0$.

Remark 8.37. In what follows we consider on T_qM the weight defined by coordinates $(y_1, \ldots, y_n)$ induced by the flag. In other words we consider the basis $V_1(q), \ldots, V_n(q)$ in T_qM and write

$$v = (y_1, \ldots, y_n) = \sum y_i V_i(q), \qquad \text{where} \quad \nu(y_i) := w_i = j \quad \text{if} \quad n_{j-1} < i \leq n_j$$

Moreover we can think at $v \in T_qM$ as the constant vector field on T_qM identically equal to v. In this way it makes sense to consider the value of a polynomial bracket at $\pi(f_1, \ldots, f_k)$ at the point q and consider its weight $\nu(\pi)$.

We prove the following auxiliary

Lemma 8.38. *Let $X = \pi(f_1, \ldots, f_k)(q) \in \mathrm{Vec}(T_qM)$, $\nu(X) \le d$. Consider now the polynomial vector field on T_qM*

$$\begin{aligned} Y(y) &= \sum y_{i_l} \cdots y_{i_1} (\operatorname{ad} V_{i_l} \circ \cdots \circ \operatorname{ad} V_{i_1} X)(q) \\ &= \sum p_i(y) V_i(q) \end{aligned} \tag{8.15}$$

for some polynomial p_i. Then $p_i \in \mathcal{D}^{(w_i - d)}$.

Proof of Lemma. It easily follows from definition of weights that

$$\operatorname{ad} V_{i_l} \circ \cdots \circ \operatorname{ad} V_{i_1}(X) \in \mathcal{D}^{(-\sum w_{i_j} - d)}$$

hence every summand of (8.15) belong to $\mathcal{D}^{(-d)}$. Then if we rewrite the sum in terms of the basis $V_i(q), i = 1, \ldots, k$ we have that every coefficient $p_i(y)$ must belong to $\mathcal{D}^{(w_i - d)}$, since $\nu(V_i(q)) = w_i$. $\square$

Now we prove the following claim: for every bracket polynomial $X = \pi(f_1, \ldots, f_k)$ we have $\Psi_*^{-1} X \in \mathcal{D}^{(-d)}$. In particular part (*ii*) will follow when $d = 1$. Clearly we can write in coordinates

$$\Psi_*^{-1} X = \sum_{i=1}^{n} a_i(s) \frac{\partial}{\partial s_i} \tag{8.16}$$

and our claim is equivalent to show that $a_i \in \mathcal{D}^{(w_i - d)}$. First we notice that

$$\begin{aligned} \Psi_* \frac{\partial}{\partial s_i} &= \frac{\partial}{\partial \varepsilon}\Big|_{\varepsilon=0} q \circ e^{s_1 V_1} \circ \cdots \circ e^{(s_i + \varepsilon) V_i} \circ \cdots \circ e^{s_n V_n} \\ &= q \circ e^{s_1 V_1} \circ \cdots \circ e^{s_i V_i} \circ V_i \circ e^{s_{i+1} V_{i+1}} \circ \cdots \circ e^{s_n V_n} \\ &= \underbrace{q \circ e^{s_1 V_1} \circ \cdots \circ e^{s_n V_n}}_{\Psi(s)} \circ e^{-s_n V_n} \circ \cdots \circ e^{-s_{i+1} V_{i+1}} \circ V_i \circ e^{s_{i+1} V_{i+1}} \circ \cdots \circ e^{s_n V_n} \end{aligned}$$

In geometric notation we can write

$$\Psi_* \frac{\partial}{\partial s_i} = e_*^{s_n V_n} \cdots e_*^{s_{i+1} V_{i+1}} V_i \Big|_{\Psi(s)} \tag{8.17}$$

Remember that, as operator on functions, $e_*^{tY} = e^{-t \operatorname{ad} Y}$. This implies that in (8.17) we have a series of bracket polynomials. Apply Ψ_* to (8.16) we get

$$X\Big|_{\Psi(s)} = \sum_{i=1}^{n} a_i(s) e_*^{s_n V_n} \cdots e_*^{s_{i+1} V_{i+1}} V_i \Big|_{\Psi(s)}$$

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

$$e_*^{-s_1 V_1} \cdots e_*^{-s_n V_n} X\Big|_q = \sum_{i=1}^{n} a_i(s) e_*^{-s_1 V_1} \cdots e_*^{-s_{i-1} V_{i-1}} V_i \Big|_q \tag{8.18}$$

Rewriting this identity in coordinates

$$\sum_i b_i(s)V_i(q) = \sum_{i,j} a_i(s)(\varphi_{ij}(s)V_j(q) + V_i(q)) \tag{8.19}$$

where $\varphi_{ij}(0) = 0$. Indeed we split the zero order term since we know that for $s = 0$ the pushforward of the vector fields is exactly V_i. Using Lemma above with X and $V_i, i = 1, \ldots, n$ we have

$$b_i \in \mathcal{D}^{w_i - d}, \qquad \varphi_{ij} \in \mathcal{D}^{w_j - w_i}$$

On the other hand we can rewrite relation between coefficients as follows

$$B(s) = A(s)(\Phi(s) + I)$$

where we denote $B(s) = (b_1(s), \ldots, b_n(s))$, $A(s) = (a_1(s), \ldots, a_n(s))$ and $\Phi(s) = (\varphi_{ij})_{ij}$ Thus we get

$$\begin{aligned} A(s) &= B(s)(I + \Phi(s))^{-1} \\ &= B(s)(I - \Phi(s) + \Phi(s)^2 - \ldots) \\ &= B(s) - (B\Phi)(s) + (B\Phi^2)(s) - \ldots \end{aligned}$$

and we can finish the proof noticing that

$$\begin{aligned} (B)_i &= b_i \in \mathcal{D}^{w_i - d} \\ (B\Phi)_i &= \sum b_j \varphi_{ji} \in \mathcal{D}^{w_j - d + (w_i - w_j)} = \mathcal{D}^{w_i - d} \\ &\vdots \end{aligned}$$

and so on. Hence we get $a_i \in \mathcal{D}^{w_i - d}$. $\square$

Remark 8.39. One can repeat all calculation in chronological notation and recover the proof in a purely algebraic way. In the above computations nothing change if we consider any permutation $\sigma = (i_1, \ldots, i_n)$ of $(1, \ldots, n)$ and the coordinate map

$$\Psi_\sigma : (s_1, \ldots, s_n) \mapsto q \circ e^{s_{i_n} V_{i_n}} \circ \ldots \circ e^{s_{i_1} V_{i_1}}$$

In particular we can consider the coordinate map

$$\Phi : (x_1, \ldots, x_n) \mapsto q \circ e^{x_n V_n} \circ \ldots \circ e^{x_1 V_1}$$

and it is easy to see that it satisfies

$$\begin{aligned} &\Phi_*^{-1} V_1 = \partial_{x_1} \\ &\Phi_*^{-1} V_2 \Big|_{x_1 = 0} = \partial_{x_2} \\ &\quad \vdots \\ &\Phi_*^{-1} V_i \Big|_{x_1 = \ldots = x_{i-1} = 0} = \partial_{x_i} \end{aligned} \tag{8.20}$$

for $i = 1, \ldots, n_1$, the set of vector fields among $f_1, \ldots, f_k$ that generates $\mathcal{D}_q$.

In Riemannian geometry the tangent space depends only on the dimension of the manifold (i.e. all tangent spaces to a n-dimensional manifold are isometric). Now we can prove that in sub-Riemannian geometry this is not true. Indeed we see that, even in dimension 3, we can have non isometric tangent space, depending on the growth vector $(n_1, \ldots, n_m)$.

In bigger dimension it is also possible to prove that, for a fixed growth vector, we have non isometric tangent space depending on the point on the manifold.

Example 8.40. (Heisenberg)
Assume $n = 3$ and that growth vector is $(2, 3)$. Then we consider coordinates (x_1, x_2, x_3) and weights $(w_1, w_2, w_3) = (1, 1, 2)$. We can assume that

$$V_1 = f_1, \qquad V_2 = f_2, \qquad V_3 = [f_1, f_2]$$

From last Remark we have that, in privileged coordinates we can assume

$$f_1 = \partial_{x_1}, \qquad f_2 = \partial_{x_2} + \alpha x_1 \partial_{x_3}, \qquad \alpha \in \mathbb{R} \tag{8.21}$$

because $f_i = \partial_{x_i} +$ something that has weight -1 and depend only on $\partial_{x_j}, j > n_1$. On the other hand from (8.20) we have

$$[f_1, f_2]\Big|_0 = \partial_{x_3} \qquad \Longrightarrow \qquad \alpha = 1$$

and we get the Heisenberg algebra

$$f_1 = \partial_{x_1}, \qquad f_2 = \partial_{x_2} + x_1 \partial_{x_3}, \qquad f_3 = \partial_{x_3} \tag{8.22}$$

Example 8.41. (Martinet)
Assume $n = 3$ and that growth vector is $(2, 2, 3)$. Then we consider coordinates (x_1, x_2, x_3) and weights $(w_1, w_2, w_3) = (1, 1, 3)$. We can assume, up to change indices, that

$$V_1 = f_1, \qquad V_2 = f_2, \qquad V_3 = [f_1, [f_1, f_2]]$$

From last Remark we have that, in privileged coordinates we can write

$$f_1 = \partial_{x_1}, \qquad f_2 = \partial_{x_2} + (\alpha x_1^2 + \beta x_1 x_2)\partial_{x_3}, \qquad \alpha, \beta \in \mathbb{R} \tag{8.23}$$

since we assume $f_2|_{x_1=0} = \partial_{x_2}$ that implies $f_2 = \partial_{x_2} + x_1 a(x)\partial_{x_3}$, but $\nu(f_2) = -1$ and so (8.23) follows.

From $V_3|_{x=0} = \partial_{x_3}$ we have

$$[f_1, [f_1, f_2]] = 2\alpha \partial_{x_3} \Longrightarrow \alpha = 1/2.$$

Moreover, since we are interested to normalize sub-Riemannian structure and not only the pair of vector fields, we consider rotations of the orthonormal frame.

Remark 8.42. Notice that

$$\begin{array}{l} \widetilde{f_1} = \cos\theta f_1 - \sin\theta f_2 \\ \widetilde{f_2} = \sin\theta f_1 + \cos\theta f_2 \end{array} \qquad \Longrightarrow \qquad [\widetilde{f_1}, \widetilde{f_2}] = [f_1, f_2].$$

Thus, denoting as usual

$$f_u = u_1 f_1 + u_2 f_2$$

we can consider the linear map

$$\varphi : u \mapsto [f_u, [f_1, f_2]]/\mathcal{D}$$

which vanish on some line on the plane $\mathcal{D} = \operatorname{span}\{f_1, f_2\}$. Up to a rotation of the frame we can assume that $f_2 \in \ker \varphi$ so that $[f_2, [f_1, f_2]] = 0$, hence $\beta = 0$.

$$f_1 = \partial_{x_1}, \qquad f_2 = \partial_{x_2} + \frac{1}{2} x_1^2 \partial_{x_3}, \qquad f_3 = \partial_{x_3} \tag{8.24}$$

8.4 Geometric meaning

In the previous section we very clearly found how $\widehat{V}$ is analitically recovered from V. It is nothing else but the principal part of V in privileged coordinates. But now we want to discuss in which sense $\widehat{V}$ is an approximation of V. It turns out that in this nonholonomic setting it plays the same role that linearization of a vector filed does in the Euclidean case.

Lemma 8.43. *Let V a vector field. In privileged coordinates we have equality*

$$\varepsilon \delta_{\frac{1}{\varepsilon}*} V = \widehat{V} + \varepsilon W_\varepsilon, \qquad \text{where } W_\varepsilon \text{ is smooth}$$

Proof. Write $V = \widehat{V} + W$ and applying the dilation we find

$$\delta_{\frac{1}{\varepsilon}*} V = \delta_{\frac{1}{\varepsilon}*} \widehat{V} + \delta_{\frac{1}{\varepsilon}*} W$$

Since $\widehat{V}$ is homogeneous of degree -1 we have $\delta_{\frac{1}{\varepsilon}*} \widehat{V} = \frac{1}{\varepsilon} \widehat{V}$ and setting $W_\varepsilon = \varepsilon \delta_{\frac{1}{\varepsilon}*} W$ we are done. $\square$

Remark 8.44. Geometrically this procedure means that we consider a small neighborhood of the point q and we make a dilation. Then we properly rescale in order to catch the principal term. This is a blow-up procedure. Notice that we are blowing-up in a nonisotropic way and it contains information about local structure of the bracket .

Now we can give a very precise meaning of the fact that nilpotent approximation is the principal part of the sub-Riemannian structure, which knows local geometry near the point q. Let us consider the *end point map*

$$F : \mathcal{U} \to M, \qquad u(\cdot) \mapsto q \circ \overrightarrow{\exp} \int_0^1 f_{u(t)} dt$$

where $\mathcal{U} = L_2^k(0,1) = L^2([0,1], \mathbb{R}^k)$ is the set of admissible controls. Let we denote by ρ the sub-Riemannian distance from the fixed point

$$\rho(x) := d(x, q) = \inf\{\|u\|, F(u) = x\} \tag{8.25}$$

From Lemma 8.43 we can write for $\varepsilon > 0$

$$f_u^\varepsilon := \varepsilon \delta_{\frac{1}{\varepsilon}*} f_u = \widehat{f}_u + \varepsilon W_u^\varepsilon$$

Denote now with f^ε and $\widehat{f}$ respectively the sub-Riemannian structures on $\mathbb{R}^n$ and by d^ε and $\widehat{d}$ the associated sub-Riemannian distance. Notice that, from the very definition of d^ε we have

$$d^\varepsilon(x,y) = \frac{1}{\varepsilon} d(\delta_\varepsilon(x), \delta_\varepsilon(y))$$

that says d^ε is d when we look infinitesimally near the point q and rescale.

Let $\rho^\varepsilon, \widehat{\rho}$ and $F^\varepsilon, \widehat{F}$ have analogous meaning. We start from an auxiliary proposition.

Proposition 8.45. *$F^\varepsilon \to \widehat{F}$ uniformly on balls in $L_2^k(0,1)$ (actually in $\mathcal{C}^\infty$ sense).*

Proof. Consider the solution $x^\varepsilon(t)$ and $\widehat{x}(t)$ of the two systems starting from $q = 0$

$$\dot{\widehat{x}}(t) = \widehat{f}_{u(t)}(\widehat{x}(t)), \qquad \dot{x}^\varepsilon(t) = f^\varepsilon_{u(t)}(x^\varepsilon(t))$$

Using Lemma 8.43 we rewrite the second equation as

$$\dot{x}^\varepsilon(t) = \widehat{f}_{u(t)}(x^\varepsilon(t)) + \varepsilon W_t^\varepsilon(x^\varepsilon(t))$$

and standard estimates from ODE theory prove that $x^\varepsilon \to \widehat{x}$.

Notice that, since nilpotent vector fields are complete, the solution $\widehat{x}(t)$ is defined for all $t \in \mathbb{R}$. ☐

Lemma 8.46. *$\{\rho^\varepsilon\}_{\varepsilon>0}$ is an equicontinuous family.*

Proof. We will prove the following: for every compact $K \subset \mathbb{R}^n$ there exists $\varepsilon_0, C > 0$, depending on K, such that

$$d^\varepsilon(x,y) \leq C|x-y|^{1/m}, \qquad \forall\, \varepsilon < \varepsilon_0, \forall\, x, y \in K. \tag{8.26}$$

where m is the degree of nonholonomy. Notice that from (8.26) we get, using triangle inequality

$$|\rho^\varepsilon(x) - \rho^\varepsilon(y)| = |d^\varepsilon(0,x) - d^\varepsilon(0,y)| \leq d^\varepsilon(x,y) \leq C|x-y|^{1/m}$$

which proves the lemma. We are then reduced to prove (8.26). Idea is to cover a fixed neighborhood of the origin using controls with bounded norms, uniformly in ε.

Let $\widehat{V}_1, \ldots, \widehat{V}_n$ an adapted basis of the nilpotent system $\widehat{f}$, such that $\widehat{V}_i = \pi_i(\widehat{f}_1, \ldots, \widehat{f}_k)$ for some bracket polynomials $\pi_i, i = 1, \ldots, n$. From the very definition we have

$$\widehat{V}_1(0) \wedge \ldots \wedge \widehat{V}_n(0) \neq 0$$

On the other hand, by continuity, this implies that they are linearly independent also in a small neighborhood of the origin and by quasi-homogeneity we get

$$\widehat{V}_1(x) \wedge \ldots \wedge \widehat{V}_n(x) \neq 0, \qquad \forall\, x \in \mathbb{R}^n.$$

Let $V_i^\varepsilon = \pi_i(f_1^\varepsilon, \ldots, f_k^\varepsilon)$ denote vector fields defined by the same bracket polynomials but in terms of the vector fields of the approximating system. For every $K \subset \mathbb{R}^n$ there exists $\varepsilon_0 = \varepsilon_0(K)$ such that

$$V_1^\varepsilon(x) \wedge \ldots \wedge V_n^\varepsilon(x) \neq 0, \qquad \forall\, x \in K, \forall\, \varepsilon \leq \varepsilon_0.$$

Recall that by Lemma 8.29, given a bracket polynomial $\pi_i(g_1,\dots,g_k), \deg \pi_i = w_i$ 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 $w_i = \deg \widehat{V}_i = \deg V_i^\varepsilon$. Now consider the map

$$\Phi^\varepsilon(t_1,\dots,t_n,x) = x\circ\overrightarrow{\exp}\int_0^1 f^\varepsilon_{u_1(t_1^{1/w_1},s)}ds\circ\ldots\circ\overrightarrow{\exp}\int_0^1 f^\varepsilon_{u_n(t_n^{1/w_n},s)}ds \tag{8.27}$$

Remark 8.47. We have the expansion

$$x\circ\overrightarrow{\exp}\int_0^1 f^\varepsilon_{u_i(t_i^{1/w_i},s)}ds = x + t_iV_i^\varepsilon(x) + O(t_i^{\frac{w_i+1}{w_i}})$$

In particular this is a $\mathcal{C}^1$ map with respect to t. Notice that it is not $\mathcal{C}^2$ if $w_i > 1$ for some i (i.e. a "real" subriemannian problem).

From this remark it follows that $\Phi^\varepsilon \in \mathcal{C}^1$ as a function of t, being a composition of $\mathcal{C}^1$ maps. Moreover we get the expansion

$$\Phi^\varepsilon(t_1,\dots,t_n,x) = x + \sum_{i=1}^n t_iV_i^\varepsilon(x) + O(|t|) \quad\Longrightarrow\quad \frac{\partial\Phi^\varepsilon}{\partial t_i}\Big|_{t=0} = V_i^\varepsilon(x)$$

Hence the map Φ^ε is a local diffeomorphism near the origin $t = (t_1,\dots,t_n) = 0$ and by Implicit Function Theorem there exists a constant $c > 0$ such that

$$x + c\nu B \subset \Phi^\varepsilon(\nu B, x), \qquad B = B(0,1)\subset\mathbb{R}^n, \quad x\in K, \tag{8.28}$$

where c is independent of ε and ν is small enough.

Let us denote now with F_x the end-point map starting from the point $x\in\mathbb{R}^n$ (with analogous meaning for $F_x^\varepsilon, \widehat{F}_x$), and with B_{L^2} the unit ball in $L_2^k[0,1]$.

We claim that (8.28) implies that there exists a constant c' such that

$$x + c'\nu B \subset F_x^\varepsilon(\nu^{\frac{1}{m}}B_{L^2}), \qquad \forall\,\nu,\varepsilon > 0 \tag{8.29}$$

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

$$t\in\nu B \Rightarrow u_i(t,\cdot)\in c_i\nu B_{L^2}, \tag{8.30}$$

$$\Rightarrow u_i(t^{1/w_i},\cdot)\in c_i\nu^{1/w_i}B_{L^2}, \tag{8.31}$$

for all $\nu > 0$ small enough.

For such ν we have by inclusion (8.29) that

$$|x-y| \le c\nu \Longrightarrow d^\varepsilon(x,y) \le \nu^{1/m}$$

where we used the fact that d^ε is the infimum of norm of u such that $F_x^\varepsilon(u) = y$. From this easily follows

$$d^\varepsilon(x,y) \le c^{-\frac{1}{m}} |x-y|^{\frac{1}{m}} \tag{8.32}$$

□

Remark 8.48. All estimates are valid also for $\varepsilon \to 0$, i.e. for the nilpotent approximation. In particular, using homogeneity

$$\widehat{d}(x,y) \le C|x-y|^{\frac{1}{m}}, \qquad \forall\, x,y \in \mathbb{R}^n \tag{8.33}$$

Indeed from the proof of Lemma 8.46 it follows that the estimate (8.33) holds in a compact K containing the origin. Consider two arbitrary points $x, y \in \mathbb{R}^n$ and $\varepsilon > 0$ such that $\delta_\varepsilon x, \delta_\varepsilon y \in K$. By the homogeneity of the distance

$$\widehat{d}(\delta_\varepsilon x, \delta_\varepsilon y) = \varepsilon\, \widehat{d}(x,y).$$

Moreover since the estimate (8.33) holds in K

$$\begin{aligned} \widehat{d}(\delta_\varepsilon x, \delta_\varepsilon y) &\le C|\delta_\varepsilon x - \delta_\varepsilon y|^{1/m} \\ &\le C\varepsilon |x-y|^{1/m} \end{aligned}$$

We can state now the main result

Theorem 8.49. *$\rho^\varepsilon \to \widehat{\rho}$ uniformly on compacts in $\mathbb{R}^n$.*

Proof. By Lemma 8.46 it is sufficient to prove pointwise convergence. We prove the following inequalities

$$\limsup_{\varepsilon\to 0^+} \rho^\varepsilon(x) \le \widehat{\rho}(x) \le \liminf_{\varepsilon\to 0^+} \rho^\varepsilon(x) \tag{8.34}$$

(i) Fix a point x and a control $\widehat{u}$ such that

$$\widehat{F}(\widehat{u}) = x, \qquad \|\widehat{u}\| = \widehat{\rho}(x),$$

i.e. such that the corresponding trajectory is a minimizer for the system $\widehat{f}$. Now consider $x^\varepsilon := F^\varepsilon(\widehat{u})$. From Proposition 8.45 we get $x^\varepsilon \to x$ for $\varepsilon \to 0$. Moreover, from the definition of ρ^ε we have $\rho^\varepsilon(x^\varepsilon) \le \widehat{\rho}(x)$. Hence

$$\begin{aligned} \rho^\varepsilon(x) &= \rho^\varepsilon(x^\varepsilon) + \rho^\varepsilon(x) - \rho^\varepsilon(x^\varepsilon) \\ &\le \widehat{\rho}(x) + |\rho^\varepsilon(x) - \rho^\varepsilon(x^\varepsilon)| \end{aligned}$$

Using that ρ^ε is an equicontinuous family and that $x^\varepsilon \to x$ we have the left inequality in (8.34).

(ii) Let now u^ε be a control such that

$$F^\varepsilon(u^\varepsilon) = x, \qquad \|u^\varepsilon\| = \rho^\varepsilon(x)$$

and define $x^\varepsilon := \widehat{F}(u^\varepsilon)$. As before we have $\widehat{\rho}(x^\varepsilon) \leq \rho^\varepsilon(x)$. Then

$$\begin{aligned}\widehat{\rho}(x) &= \widehat{\rho}(x^\varepsilon) + \widehat{\rho}(x) - \widehat{\rho}(x^\varepsilon) \\ &\leq \rho^\varepsilon(x) + |\widehat{\rho}(x) - \widehat{\rho}(x^\varepsilon)|\end{aligned}$$

and now it is sufficient to notice that $x^\varepsilon = F^\varepsilon(u^\varepsilon) \to \widehat{F}(u^\varepsilon) = x$ since $F^\varepsilon \to \widehat{F}$ uniformly on balls of L^2 and u^ε bounded since ρ^ε are equicontinuous.. □

In privileged coordinates $x = (x_1, \ldots, x_m) \in \mathbb{R}^{k_1} \oplus \ldots \oplus \mathbb{R}^{k_m} = \mathbb{R}^n$ we set

$$\Pi_\varepsilon = \{x \in \mathbb{R}^n, |x_i| \leq \varepsilon^i, i = 1, \ldots, m\}$$

Corollary 8.50 (Ball-Box Theorem)**.** *There exists constants $c_1, c_2 > 0$ such that*

$$c_1\Pi_\varepsilon \subset B(x, \varepsilon) \subset c_2\Pi_\varepsilon$$

where $B(x, \varepsilon)$ is the subriemannian ball in privileged coordinates.

Notice that this is a weaker statement with respect to Theorem 8.49.

Exercise 8.51. Prove Corollary 8.50.

Definition 8.52. Let f and $\widetilde{f}$ be two sub-Riemannian structures on the same manifold M. We say that the structures are *locally Lipschitz equivalent* if, for any compact $K \subset M$ there exist $c_1, c_2 > 0$ such that

$$c_1 d(x, y) \leq \widetilde{d}(x, y) \leq c_2 d(x, y)$$

where μ and $\widetilde{\mu}$ are respectively the sub-Riemannian distances induced by f and $\widetilde{f}$.

From the Ball-Box Theorem we easily get a characterization of locally Lipschitz equivalent structures in term of the distribution.

Corollary 8.53. *Two sub-Riemannian structures are locally Lipschitz equivalent if and only if the two flags are equal at al points, i.e.*

$$\mathcal{D}^i_q = \widetilde{\mathcal{D}}^i_q, \qquad \forall\, q \in M, \quad \forall\, i \geq 1.$$

Corollary 8.54. *Two regular sub-Riemannian structures are locally Lipschitz equivalent if and only if their distributions are equal at al points, i.e.*

$$\mathcal{D}_q = \widetilde{\mathcal{D}}_q, \qquad \forall\, q \in M.$$

In other words, in the regular case, the distribution define the metric up to locally Lipschitz equivalence.

Remark 8.55. In the proof of Theorem 8.49 we showed that, in some coordinates, the sub-Riemannian metric has an holder estimate with respect to the Euclidean one. The fact that the metric is Lipschitz equivalent to the Euclidean one characterize exactly Riemannian structures on M.

Moreover we notice that this is only local property since we do not study the behaviour of the constants c_1, c_2 when K become big.

8.5 Algebraic meaning

In the last section we proved in which sense the sub-Riemannian tangent space approximate the sub-Riemannian structure on the manifold. Now we also show that, at least in the regular case, the nilpotent approximation has a structure of Lie group, endowed with a left-invariant sub-Riemannian structure.

Recall that given an orthonormal frame $\{f_1, \ldots, f_k\}$ for the sub-Riemannian structure, by Proposition 8.20 the vector field $J_q^m f_i$, jet of a vector field on M, is a well defined vector field on the quotient $T_q^f := J_q^f / \sim$, which we denote $\widehat{f}_i$.

Proposition 8.56. *The Lie algebra* $\mathrm{Lie}\{\widehat{f}_1, \ldots, \widehat{f}_k\}$ *is a nilpotent Lie algebra of step m, where m is the nonholonomic degree of f at q.*

Proof. Consider privileged coordinates around the point q. Then $\widehat{f}_i$ has weight -1 and is homogeneous with respect to the dilation δ_λ. Moreover for any bracket monomial we have

$$\nu([\widehat{f}_{i_1}, \ldots, [\widehat{f}_{i_{j-1}}, \widehat{f}_{i_j}]]) = -j$$

Since every vector field V, when written in privileged coordinates, satisfies $\nu(V) \geq -m$, then every bracket of m vector fileds is necessarily zero. $\square$

Consider now the group generated by the flows of these vector fields

$$G = Gr\{e^{t\widehat{f}_1}, \ldots, e^{t\widehat{f}_k}\}$$

which acts on T_q^f on the right, and is by definition a nilpotent Lie group.[2] Moreover in the proof of Theorem 8.25 we showed that this action is also transitive (i.e. we can realize every element of T_q^f with this action)

Collecting together all these results we have

Corollary 8.57. *The nilpotent approximation T_q^f is a homogeneous space, diffeomorphic to the quotient G/G_0, where G_0 is the isotropy group of the trivial element of T_q^f.*

Before interpreting this contruction at the level of Lie algebras, we recall some definitions.

The *free associative algebra* on k generators $x_1, \ldots, x_k$ is the associative algebra A_k of linear combinations of words of its generators, where the product of two element is defined by juxtaposition. The *free Lie algebra* on k generators, denoted $\mathcal{L}_k$, is the algebra of Lie elements of A_k where the product of two elements x, y is defined by the commutator $[x, y] = xy - yx$.

The *nilpotent step m free Lie algebra* on k generators $x_1, \ldots, x_k$, is the quotient of the free Lie algebra by the ideal $\mathcal{I}^{m+1}$ generated as follows: $\mathcal{I}^1 = \mathcal{L}$, and $\mathcal{I}^j = [\mathcal{I}^{j-1}, \mathcal{L}]$.

Let $\mathrm{Lie}_m\{X_1, \ldots, X_k\}$ be the nilpotent step m free Lie algebra generated by the vector fields $X_1, \ldots, X_k$ and consider the subalgebra

$$C := \{\pi \in \mathrm{Lie}_m\{X_1, \ldots, X_k\} \,|\, \pi(\widehat{f}_1, \ldots, \widehat{f}_k)(0) = 0\}$$

of all polynomial bracket such that if we replace X_i with $\widehat{f}_i$ are zero when evaluated at zero. Then

$$\mathrm{Lie}\, T_q^f \simeq \mathrm{Lie}_m\{X_1, \ldots, X_k\}/C$$

[2]A Lie group G is *nilpotent* if its Lie algebra $\mathfrak{g}$ is nilpotent. The fact that G acts on the right is because right action satisfies $R_{hg} = R_g R_h$ (i.e. $x \cdot (hg) = (x \cdot h) \cdot g$).

Remark 8.58. To discuss regularity properties of T_q^f with respect to q, we can restate this characterization in such a way that does not depend on the nilpotent approximation:

$$\mathrm{Lie}\, T_q^f \simeq \mathrm{Lie}_m\{X_1, \ldots, X_k\}/C_q$$

where C_q is the core subalgebra

$$C_q := \{\pi \in \mathrm{Lie}_m\{X_1, \ldots, X_k\} \,|\, \pi(f_1, \ldots, f_k)(q) \in \mathcal{D}_q^{\deg \pi - 1}\} \tag{8.35}$$

Lemma 8.59. *Assume that the sub-Riemannian structure has constant growth vector, i.e. that $n_i(q) = \dim \mathcal{D}_q^i$ does not depend on q. Then C_q is an ideal.*

In particular T_q^f is a Lie group.

Proof. It is sufficent to prove that

$$X \in C_q \quad \Longrightarrow \quad [f_i, X] \in C_q, \quad \forall\, i = 1, \ldots, k$$

Since the structure has constant growth vector, we can consider an adapted basis $V_1, \ldots, V_n$, well defined in a neighborhood O_q of q. In particular if $X = \pi(f_1, \ldots, f_k)$ is a bracket polynomial of degree $\deg \pi = d$ we can write

$$X(q') = \sum_{i: w_i \leq d} a_i(q') V_i(q'), \qquad \forall\, q' \in O_q$$

where a_i are suitable smooth functions. From (8.35) we have that $X \in C_q$ if and only if it belongs to $\mathcal{D}_q^{d-1}$, i.e. $a_i(q) = 0$, $\forall\, i$ s.t. $w_i = d$. On the other hand

$$\begin{aligned} [f_i, X] &= [f_i, \sum_{w_j \leq d} a_j V_j] \\ &= \sum_{w_j \leq d} a_j [f_i, V_j] + f_i(a_j) V_j \end{aligned} \tag{8.36}$$

From this equality it is easy to check that every coefficient of degree $d+1$ in this sum is null at q, since they can appear only in the first summand of (8.36). □

Corollary 8.60. *Under previuos assumptions $\widehat{f}_1, \ldots, \widehat{f}_k$ are a basis of left-invariant vector fields on T_q^f.*

Proof. All relies on the fact that if we consider a left invariant vector field X on a Lie group G, and we consider the right action of a normal subgroup H on it, then X is a well defined left-invariant vector field on the quotient G/H, which is still a Lie group. □

Examples

Heisenberg

Martinet

Grushin

Chapter 9

The volume in sub-Riemannian geometry

9.1 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.

9.2 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 $i = 1, 2, \ldots, m$, $k_i(q) = \dim(\mathcal{D}_q^i)$ does not depend on $q \in M$. In this case the subspaces $\mathcal{D}_q^i$ 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 9.1. 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

$$\mathrm{gr}_q(\mathcal{D}) = \mathcal{D}_q \oplus \mathcal{D}_q^2/\mathcal{D}_q \oplus \ldots \oplus \mathcal{D}_q^m/\mathcal{D}_q^{m-1}.$$

The vector space $\mathrm{gr}_q(\mathcal{D})$ can be endowed with a Lie algebra structure, which respects the grading. Then, there is a unique connected, simply connected group, $\mathrm{Gr}_q(\mathcal{D})$, such that its Lie algebra is $\mathrm{gr}_q(\mathcal{D})$. The global, left-invariant vector fields obtained by the group action on any orthonormal basis of $\mathcal{D}_q \subset \mathrm{gr}_q(\mathcal{D})$ defines 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 [**?**]. (See also [16]). The definition rests on the following lemmas.

Lemma 9.2. *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 length of $v \in V$ is*

$$\|v\|_V = \min\{\|e\|_E \ s.t.\ \pi(e) = v\}\,. \tag{9.1}$$

Proof. It is easy to check that Eq. (9.1) 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 length of $v \in V$ is the length of the shortest element $e \in \pi^{-1}(v)$. □

Lemma 9.3. *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 [16, Lemma 10.4]. □

The idea behind Popp's volume is to define an inner product on each $\mathcal{D}_q^i/\mathcal{D}_q^{i-1}$ which, in turn, induces an inner product on the orthogonal direct sum $\mathrm{gr}_q(\mathcal{D})$. The latter has a natural volume form, which is the canonical volume of an inner product space obtained by wedging the elements an orthonormal dual basis. Then, we employ Lemma 9.3 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 $\pi : \mathcal{D}_q \otimes \mathcal{D}_q \to \mathcal{D}_q^2/\mathcal{D}_q$

$$\pi(v \otimes w) := [V, W]_q \mod \mathcal{D}_q\,, \tag{9.2}$$

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 $\mathcal{D}$

$$\widetilde{V} = V + \sum_{i=1}^k f_i X_i\,, \qquad \widetilde{W} = W + \sum_{i=1}^k g_i X_i\,, \tag{9.3}$$

where, for $1 \leq i \leq k$, $f_i, g_i \in C^\infty(M)$ and $f_i(q) = g_i(q) = 0$. Therefore

$$[\widetilde{V}, \widetilde{W}] = [V, W] + \sum_{i=1}^k \left(V(g_i) - W(f_i)\right) X_i + \sum_{i,j=1}^k f_i g_j [X_i, X_j]\,. \tag{9.4}$$

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 $\pi_i : \otimes^i \mathcal{D}_q \to \mathcal{D}_q^i/\mathcal{D}_q^{i-1}$

$$\pi_i(v_1 \otimes \cdots \otimes v_i) = [V_1, [V_2, \ldots, [V_{i-1}, V_i]]]_q \mod \mathcal{D}_q^{i-1}\,, \tag{9.5}$$

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 9.2, 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

$$\mathrm{gr}_q(\mathcal{D}) = \mathcal{D}_q \oplus \mathcal{D}_q^2/\mathcal{D}_q \oplus \ldots \oplus \mathcal{D}_q^m/\mathcal{D}_q^{m-1}\,, \tag{9.6}$$

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 9.3. Namely

$$\mathcal{P}_q = \theta_q^*(\mu_q) = \mu_q \circ \theta_q\,, \tag{9.7}$$

where θ_q^* denotes the dual map and we employ the canonical identification $(\wedge^n T_qM)^* \simeq \wedge^n T_q^*M$. Eq. (9.7) 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 9.5.

Remark 9.4. 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 9.3.

9.3 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. Even though it is not needed right now, it is useful to define the functions $c_{ij}^l \in C^\infty(M)$ by

$$[X_i, X_j] = \sum_{l=1}^{n} c_{ij}^l X_l\,. \tag{9.8}$$

With a standard abuse of notation we call them *structure constants*. For $j = 2, \dots, m$ we define the *adapted structure constants* $b^l_{i_1 \dots i_j} \in C^\infty(M)$ as follows:

$$[X_{i_1}, [X_{i_2}, \dots, [X_{i_{j-1}}, X_{i_j}]]] = \sum_{l=k_{j-1}+1}^{k_j} b^l_{i_1 i_2 \dots i_j} X_l \mod \mathcal{D}^{j-1}\,, \tag{9.9}$$

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. (9.8) 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. (9.9) 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:

$$[B_j]^{hl} = \sum_{i_1, i_2, \ldots, i_j = 1}^{k} b^h_{i_1 i_2 \ldots i_j} b^l_{i_1 i_2 \ldots i_j} \,, \qquad j = 1, \ldots, m \,, \tag{9.10}$$

with the understanding that B_1 is the $k \times k$ identity matrix. It turns out that each B_j is positive definite.

Theorem 9.5. *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*

$$\mathcal{P} = \frac{1}{\sqrt{\prod_j \det B_j}} \, \nu^1 \wedge \ldots \wedge \nu^n \,, \tag{9.11}$$

where B_j is defined by (9.10) *in terms of the adapted structure constants* (9.9).

To clarify the geometric meaning of Eq. (9.11), 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 9.2). Therefore, the function appearing at the beginning of Eq. (9.11) 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 9.5

We are now ready to prove Theorem 9.5. 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 $\mathcal{D}^i_q$ of the appropriate higher order distribution.

Step 2 distribution

If $\mathcal{D}$ is a step 2 distribution, then $\mathcal{D}^2 = TM$. The growth vector is $\mathcal{G} = (k, n)$. We choose $n - k$ independent vector fields $\{Y_l\}_{l=k+1}^n$ such that $X_1, \ldots, X_k, Y_{k+1}, \ldots, Y_n$ is a local adapted frame for TM. Then

$$[X_i, X_j] = \sum_{l=k+1}^{n} b_{ij}^l Y_l \mod \mathcal{D} \,. \tag{9.12}$$

For each $l = k+1, \ldots, n$, we can think to b^l_{ij} 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 9.2, the norm on $\mathcal{D}^2/\mathcal{D}$ is induced by the linear map $\pi : \otimes^2\mathcal{D} \to \mathcal{D}^2/\mathcal{D}$

$$\pi(X_i \otimes X_j) = [X_i, X_j] \mod \mathcal{D}\,. \tag{9.13}$$

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

$$\|\mathcal{D}^2/\mathcal{D}\|_Y = \min\{\| \otimes^2 \mathcal{D}\|_Z \text{ s.t. } \pi(Z) = Y\}\,. \tag{9.14}$$

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. (9.14) writes

$$\|\mathcal{D}^2/\mathcal{D}\|_Y = \min\{|a| \text{ s.t. } a \cdot b^l = c^l,\ l = k+1, \ldots, n\}\,, \tag{9.15}$$

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 $l = k+1, \ldots, n$. In order to find an explicit expression for $\|\mathcal{D}^2/\mathcal{D}\|^2_Y$ in terms of the b^l, we employ the Lagrange multipliers technique. Then, we look for extremals of

$$L(a, b^{k+1}, \ldots, b^n, \lambda_{k+1}, \ldots, \lambda_n) = |a|^2 - 2\sum_{l=k+1}^n \lambda_l (a \cdot b^l - c^l)\,. \tag{9.16}$$

We obtain the following system

$$\begin{cases} \displaystyle\sum_{l=k+1}^n \lambda_l \cdot b^l - a = 0, \\ \displaystyle\sum_{l=k+1}^n \lambda_l b^l \cdot b^r = c^r\,, \qquad r = k+1, \ldots, n. \end{cases} \tag{9.17}$$

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 (9.17) by employing the matrix B^{-1}, which has components B^{-1}_{hl}. Indeed a straightforward computation leads to

$$\|\mathcal{D}^2/\mathcal{D}\|^2_{c^s Y_s} = c^h B^{-1}_{hl} c^l\,. \tag{9.18}$$

By polarization, the inner product on $\mathcal{D}^2/\mathcal{D}$ is defined, in the basis Y_l, by

$$\langle Y_l, Y_h\rangle_{\mathcal{D}^2/\mathcal{D}} = B^{-1}_{lh}\,. \tag{9.19}$$

Observe that B^{-1} is the Gram matrix of the vectors $Y_{k+1}, \ldots, Y_n$ seen as elements of $\mathcal{D}^2/\mathcal{D}$. Then, by the definition of Popp's volume, if $\nu^1, \ldots, \nu^k, \mu^{k+1}, \ldots, \mu^n$ is the dual basis associated with $X_1, \ldots, X_k, Y_{k+1}, \ldots, Y_n$, the following formula holds true

$$\mathcal{P} = \frac{1}{\sqrt{\det B}}\, \nu^1 \wedge \cdots \wedge \nu^k \wedge \mu^{k+1} \wedge \cdots \wedge \mu^n\,. \tag{9.20}$$

General case

In the general case, the procedure above can be carried out with no difficulty. Let $X_1, \ldots, X_n$ be a local adapted frame for the flag $\mathcal{D}^0 \subset \mathcal{D} \subset \mathcal{D}^2 \subset \cdots \subset \mathcal{D}^m$. As usual $k_i = \dim(\mathcal{D}^i)$. For $j = 2, \ldots, m$ we define the adapted structure constants $b^l_{i_1 \ldots i_j} \in C^\infty(M)$ by

$$[X_{i_1}, [X_{i_2}, \ldots, [X_{i_{j-1}}, X_{i_j}]]] = \sum_{l=k_{j-1}+1}^{k_j} b^l_{i_1 i_2 \ldots i_j} X_l \mod \mathcal{D}^{j-1}\,, \tag{9.21}$$

where $1 \leq i_1, \ldots, i_j \leq k$. Again, $b^l_{i_1 \ldots i_j}$ can be seen as the components of a vector $b^l \in \mathbb{R}^{k^j}$.

Recall that for each j we defined the surjective linear map $\pi_j : \otimes^j \mathcal{D} \to \mathcal{D}^j/\mathcal{D}^{j-1}$

$$\pi_j(X_{i_1} \otimes X_{i_2} \otimes \cdots \otimes X_{i_j}) = [X_{i_1}, [X_{i_2}, \ldots, [X_{i_{j-1}}, X_{i_j}]]] \mod \mathcal{D}^{j-1}\,. \tag{9.22}$$

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 $1 \leq j \leq m$, the $k_j - k_{j-1}$ dimensional square matrix B_j, of components

$$[B_j]^{hl} = \sum_{i_1, i_2, \ldots, i_j = 1}^{k} b^h_{i_1 i_2 \ldots i_j} b^l_{i_1 i_2 \ldots i_j}\,. \tag{9.23}$$

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 $X_{k_{j-1}+1}, \ldots, X_{k_j} \in \mathcal{D}^j/\mathcal{D}^{j-1}$, in other words

$$\langle X_{k_{j-1}+l}, X_{k_{j-1}+h} \rangle_{\mathcal{D}^j/\mathcal{D}^{j-1}} = B^{-1}_{lh}\,. \tag{9.24}$$

Therefore, if $\nu^1, \ldots, \nu^n$ is the dual frame associated with $X_1, \ldots, X_n$, Popp's volume is

$$\mathcal{P} = \frac{1}{\sqrt{\prod_{j=1}^m \det B_j}}\, \nu^1 \wedge \ldots \wedge \nu^n\,. \tag{9.25}$$

9.4 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 9.6. 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 9.7. 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 $1 \leq i \leq m$.

Definition 9.8. 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 9.9. *Sub-Riemannian (local) isometries are volume preserving transformations for Popp's volume.*

Proposition 9.9 may be false for volumes different than Popp's one. We have the following.

Proposition 9.10. *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 9.9 holds true.*

Definition 9.11. 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 9.9 we recover a well-known result (see again [16]).

Corollary 9.12. *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 9.9 and 9.10.

Proof of Proposition 9.9

Let $\phi \in \mathrm{Iso}(M)$ be a (local) isometry, and $1 \leq i \leq m$. The differential ϕ_* induces a linear map

$$\widetilde{\phi}_* : \otimes^i \mathcal{D}_q \to \otimes^i \mathcal{D}_{\phi(q)} . \tag{9.26}$$

Moreover ϕ_* preserves the flag $\mathcal{D} \subset \ldots \subset \mathcal{D}^m$. Therefore, it induces a linear map

$$\widehat{\phi}_* : \mathcal{D}^i_q/\mathcal{D}^{i-1}_q \to \mathcal{D}^i_{\phi(q)}/\mathcal{D}^{i-1}_{\phi(q)} . \tag{9.27}$$

The key to the proof of Proposition 9.9 is the following lemma.

Lemma 9.13. $\widetilde{\phi}_*$ *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. (9.5). Namely, let $Y \in \mathcal{D}^i_q/\mathcal{D}^{i-1}_q$. Then

$$\|Y\|_{\mathcal{D}^i_q/\mathcal{D}^{i-1}_q} = \min\{\|Z\|_{\otimes\mathcal{D}_q} \text{ s.t. } \pi_i(Z) = Y\} . \tag{9.28}$$

As a consequence of the properties of the Lie brackets, $\pi_i \circ \widetilde{\phi}_* = \widehat{\phi}_* \circ \pi_i$. Therefore

$$\|Y\|_{\mathcal{D}^i_q/\mathcal{D}^{i-1}_q} = \min\{\|\widetilde{\phi}_* Z\|_{\otimes\mathcal{D}_{\phi(q)}} \text{ s.t. } \pi_i(\widetilde{\phi}_* Z) = \widehat{\phi}_* Y\} = \|\widehat{\phi}_* Y\|_{\mathcal{D}^i_{\phi(q)}/\mathcal{D}^{i-1}_{\phi(q)}} . \tag{9.29}$$

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. $\square$

Proof of Proposition 9.10

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 $\phi \in \mathrm{Iso}(M)$

$$f\mu = \mathcal{P} = \phi^*\mathcal{P} = (f \circ \phi)\, \phi^*\mu = (f \circ \phi)\, \mu\,, \tag{9.30}$$

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. $\square$

Hausdorff dimension and Hausdorff volume

Density of the Hausdorff volume with respect to a smooth volume

Bibliographical notes

Chapter 10

Regularity of the sub-Riemannian distance

In this chapter we focus our attention on the analytical properties of the sub-Riemannian distance d. In particular we want to answer to the following questions:

(i) Which is the minimal regularity of d that we can always expect?

(ii) Is the sub-Riemannian distance smooth? If not, can we characterize smooth points?

10.1 General properties of the distance function

In the following we work in the usual setting

$$f : \mathbf{U} \to TM, \qquad \dot{q} = \sum_{i=1}^{k} u_i f_i(q)$$

where $\mathbf{U}$ is a rank k trivial Euclidean bundle on M and $u_i(t) \in L^2$, and $f_1, \ldots, f_k$ are smooth vector fields.

Definition 10.1. Fix a point $q \in M$. The *flag* of the sub-Riemannian structure at the point q is the sequence of subspaces $\mathcal{D}_q^1 \subset \mathcal{D}_q^2 \subset \ldots$ defined by

$$\mathcal{D}_q^i := \operatorname{span}\{[f_{i_1}, \ldots, [f_{i_{l-1}}, f_{i_l}]](q),\ l \leq i\}$$

Notice that $\mathcal{D}_q^1 = \mathcal{D}_q$ is the set of admissible directions, $\mathcal{D}_q^2$ is the set of directions that are admissible with one bracket etc.

The bracket generating assumptions implies that

$$\forall\, q \in M,\ \exists\, m(q) > 0 \quad \text{s.t.} \quad \mathcal{D}_q^{m(q)} = T_qM$$

and $m(q)$ is called the *step* of the sub-Riemannian structure at q.

Exercise 10.2. Prove that the filtration defined by the subspaces $\mathcal{D}_q^i, i \geq 1$ is intrinsic for the sub-Riemannian structure. In particular this implies that $m(q)$ does not depend on the basis of vector fileds (i.e. on the trivialization of $\mathbf{U}$).

In the Chapter **??** we already proved that the sub-Riemannian distance is Hölder continuous. For the reader's convenience, we recall here the statement.

Proposition 10.3. *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.

Now fix a point $q_0 \in M$ and define the following functions

$$\rho(q) = d(q_0, q), \qquad \mu(q) = \frac{1}{2}d^2(q_0, q).$$

Note. In what follows we always consider the point q_0 to be fixed and $r_0 > 0$ such that $B = B_{q_0}(r_0)$ is a closed compact ball centered in q_0.

Our main goal is to prove the following

Theorem 10.4. *The function $\mu\big|_B : B \to \mathbb{R}$ is smooth on a open dense subset of B.*

Let us denote by $F = F_{q_0} : \mathcal{U} \to M$ the end-point map with fixed point $q_0 \in M$, i.e. the map that associates to every control $u(\cdot) \in \mathcal{U} \subset L^2$ the end-point $q_u(1)$ of the solution associated to the control u and denote with $\mathcal{B}$ the ball of radius r_0 (defined above) in L^2. Notice that since B is compact then $\mathcal{B} \subset \mathcal{U}$.

The proof of Theorem 10.4 is the object of all this section. Let us start recalling the following result and its corollary, some of which were already obtained in the previous chapters.

Proposition 10.5. *$F\big|_{\mathcal{B}}$ is weakly continuous in L^2. In other words if $u_n \rightharpoonup u$ in the weak topology then $F(u_n) \to F(u)$.*

Remark 10.6. Actually we prove that all trajectories converge uniformly and not only their end-points.

Proof. Consider the solution of the problem

$$\dot\gamma(t) = f_{u(t)}(\gamma(t)), \qquad \gamma(0) = q_0, \qquad u \in \mathcal{B}.$$

Since the ball B is compact, all trajectories are Lipschitzian with the same Lipchitz constant. In particular they form a precompact set in the C^0 topology.

Assume now that $u_n \rightharpoonup u$ and consider the family of curves $\gamma_n(t)$ associated to u_n, that satisfy

$$\gamma_n(t) = q_0 + \int_0^t f_{u_n(\tau)}(\gamma_n(\tau))d\tau.$$

By compactness there exists a subsequence, which we still denote γ_n, such that $\gamma_n \to \gamma$ uniformly, for some curve γ, in particular their endpoints converge. It remains to show that γ is the trajectory associated to u.

Since $u_n \rightharpoonup u$ we have that $f_{u_n(t)}(\gamma_n(t)) \to f_{u(t)}(\gamma(t))$ being the product between strong and weak convergent sequences.[1] taking the limit we find

$$\gamma(t) = q_0 + \int_0^t f_{u(\tau)}(\gamma(\tau))d\tau,$$

i.e. γ is the trajectory associated to u. □

[1] one can write the coordinate expression $\sum u_k^i f_i(q_k(t))$

The previous proposition reproves the existence of minimizers

Corollary 10.7 (Existence of minimizers)**.** *For any $q_1 \in B_{q_0}(r)$ there exists $\widetilde{u}$ (with $\|\widetilde{u}\| \le r$) that join q_0 and q_1 and is a minimizer. i.e. $\|\widetilde{u}\| = d(q_0, q_1) = \rho(q_1)$.*

Proof. Consider a point q_1 in the compact ball B. Then take a minimizing sequence u_n such that $F(u_n) = q_1$ and $\|u_n\| \to \rho(q_1)$. Since $u_n \in \mathcal{B}$ and bounded sets are compact in the weak L^2 topology, we can assume $u_n \rightharpoonup \widetilde{u}$. On the other hand the standard semicontinuity result of the norm gives

$$\|\widetilde{u}\| \le \liminf_n \|u_n\| = \rho(q_1),$$

and the equality holds.

□

Definition 10.8. A control u is called *minimizer* if it satisfies $\mu(F(u)) = J(u)$. Notice that in this case we have $|u(t)| = const$ and moreover $|u(t)| = \|u\| = \rho(F(u))$.

Theorem 10.9 (Compactness)**.** *The set of all minimizer controls with value in a compact ball*

$$\mathcal{M} = \{u \text{ minimizer}, \ F(u) \in B\},$$

is compact in the strong L^2 topology.

Proof. Consider a sequence $u_n \in \mathcal{M}$, that we can assume weakly convergent $u_n \rightharpoonup u$ (since bounded sets are weakly compact). If we prove that $\|u_n\| \to \|u\|$ we are done by a standard argument.

From Proposition it follows that $F(u_n) \to F(u)$ strongly and the continuity of the distance implies $\rho(F(u_n)) \to \rho(F(u))$. Moreover since $u_n \in \mathcal{M}$ we have that $\|u_n\| = \rho(F(u_n))$ and by weak semicontinuity of the norm we get

$$\|u\| \le \liminf_n \|u_n\| = \liminf_n \rho(F(u_n)) = \rho(F(u)),$$

that implies $u_n \to u$ and $u \in \mathcal{M}$. □

Now we focus on normal extremal paths starting from the fixed point q_0 and we want to understand how they cover a neighborhood of the initial point.

Recall that normal extremal paths are projections of the Hamiltonian flow on T^*M

$$\dot{\lambda}(t) = \vec{H}(\lambda(t)), \qquad H(\lambda) = \frac{1}{2}\sum_{i=1}^{k} \langle \lambda, f_i(q) \rangle^2 ,$$

where H is the sub-Riemannian Hamiltonian.

In particular the exponential map can be interpreted as the restriction of the end-point map to a special class of controls parametrized by a covector $\lambda_0 \in T^*_{q_0}M$

$$\mathcal{E}_{q_0}(\lambda_0) = \pi \circ e^{\vec{H}}(\lambda_0) = F(u^{\lambda_0}),$$

where

$$u^{\lambda_0}(t) = (u_i^{\lambda_0}(t))_{i=1,\dots,k}, \qquad u_i^{\lambda_0}(t) = \langle \lambda(t), f_i(\gamma(t)) \rangle .$$

Recall that, if we denote by $\gamma_{\lambda_0}(t)$ the normal extremal path with initial covector λ_0, from the homogeneity property of H it follows that

$$\mathcal{E}(t\lambda(0)) = \gamma_{\lambda_0}(t),$$

from which one get

$$D_0\mathcal{E}(\lambda_0) = \dot{\gamma}_{\lambda_0}(0) = \frac{\partial H}{\partial p}(q_0, p_0) = D_0(H|_{T^*_{q_0}M}).$$

It follows that 0 is a regular point of $\mathcal{E}$ if and only if $\mathcal{D}_{q_0} = T_{q_0}M$.

Remark 10.10. In the Riemannian case $\mathcal{E}$ 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 d^2 is a smooth function around q_0.

As we show next, as soon as $\mathcal{D}_{q_0} \neq T_{q_0}M$ singularities appear. Recall the notation $\mu(q) = \frac{1}{2}d^2(q_0, q)$ and let us consider, as before, a compact ball $B = B_{q_0}(r_0)$, where we have existence of minimizers by Proposition 10.7.

Theorem 10.11. *The function μ is smooth in a neighborhood O_q of $q \in B$ if and only if for all $q' \in O_q$*

(i) q' is connected with q_0 by a unique *minimizer,*

(ii) $q' = \mathcal{E}_{q_0}(\lambda'_0)$, where λ'_0 is a regular point of $\mathcal{E}_{q_0}$,

In this case the final covector on the geodesic is $\lambda'_1 := e^{\vec{H}}(\lambda'_0) = d_{q'}\mu$.

We divide the proof of this Theorem into two part, proving separately the two implications.

First part of the Proof of Theorem 10.11. Assume that μ is smooth in a neighborhood O_q of $q \in B$, we want to show that (i) and (ii) holds. Moreover, we show that $d_{q_1}\mu$ is the covector associated to the minimizer.

Denote by F the endpoint map from q_0 and consider the function

$$\Psi : u \mapsto \frac{1}{2}\int_0^1 |u(t)|^2 dt - \mu(F(u)),$$

Notice that $\Psi \geq 0$ and $\Psi(u) = 0$ if and only if u is lenght-minimizer. Since μ is smooth we have that u is lenght-minimizer implies $D_u\Psi = 0$, but

$$D_u\Psi = u - (d_{F(u)}\mu)\, D_uF = 0,$$

that means exactly that u is normal and $\lambda_1 = d_{F(u)}\mu$. Moreover u is unique, since from the covector $\lambda_1 \in T_{q_1}M$ we can uniquely recover the extremal $\lambda(t)$ and its projection $\gamma(t) = \pi(\lambda(t))$, that uniquely determines $u_i(t) = \langle \lambda(t), f_i(\gamma(t))\rangle$.

It remains to prove (ii). For any $q_1 \in O_q$ define the following map

$$\Phi : q_1 \mapsto e^{-\vec{H}}(d_{q_1}\mu) \in T^*_{q_0}M.$$

Clearly it is smooth and is a right inverse for the exponential map, since

$$\mathcal{E}(\Phi(q_1)) = \pi \circ e^{\vec{H}}(e^{-\vec{H}}(d_{q_1}\mu)) = \pi(d_{q_1}\mu) = q_1. \tag{10.1}$$

The existence of a smooth right inverse, using the chain rule, implies that q is a regular point. $\square$

Remark 10.12. Notice that (ii) proves that u is strictly normal, being a regular point of the exponential map.

Before proving the converse we show an important corollary of this result. Denote by $S_r := \mu^{-1}(r)$ the sub-Riemannian sphere centered at q_0

Corollary 10.13. *Assume that $\mathcal{D}_{q_0} \neq T_{q_0}M$. Any non empty level set S_r contains a non smooth point of the function μ.*

Proof. Assume, by contradiction, that μ is smooth at every point of the sphere S_r. Then, since $d_q\mu \neq 0$ for every $q \in S_r$ (since $d_q\mu$ is the nonzero covector attached at the final point of a geodesic, see the Theorem 10.11) it follows that S_r is a submanifold of dimension $n-1$. Moreover, being the level set of a continuous function, it is closed, hence compact.

Consider the map

$$\Phi : S_r \to T^*_{q_0}M, \qquad q \mapsto e^{-\vec{H}}(d_q\mu) \in T^*_{q_0}M,$$

which defines an inverse of the exponential map (see also (10.1)). Moreover

$$H(\Phi(q)) = r, \qquad \text{since} \quad \mu(q) = H(\lambda) = r,$$

from which it follows that actually Φ defines a map

$$\Phi : S_r \to H^{-1}(r) \cap T^*_{q_0}M, \tag{10.2}$$

that is a diffeomorphism of S_r, onto some connected and compact $n-1$ dimensional region in the image. On the other hand the set

$$H^{-1}(r) \cap T^*_{q_0}M = \{\lambda \in T^*_{q_0}M : \frac{1}{2}\sum_{i=1}^{k} \langle \lambda, f_i(q)\rangle^2 = r\}$$

is a connected $n-1$ dimensional submanifold, being diffeomorphic to the cylinder $S^\ell \times \mathbb{R}^{n-\ell}$, where ℓ is the rank of the structure at the point. Since the cylinder $S^\ell \times \mathbb{R}^{n-\ell}$ is connected, but not compact, we get a contradiction. □

Second part of the Proof of Theorem 10.11. We start with the definition of smooth point for μ.

Definition 10.14. A point $q \in M$ is called a *smooth point* if there exists a unique minimizer joining q_0 to q such that it is strictly normal, associated to a control u s.t. $\lambda_1 D_u F = u$ for some $\lambda_1 \in T^*_q M$, and $\lambda_0 = e^{-\vec{H}}(\lambda_1)$ is a regular point for $\mathcal{E}_{q_0}$. The set of smooth points is denoted by Σ.

Notice that this definition does not involve directly the regularity of μ. Actually, the proof of Theorem 10.11 is completed by the following

Proposition 10.15. *Σ is an open set and μ is smooth at every point of Σ.*

Proof. Let us start proving that Σ is open. It is sufficient to show that

$$\forall\, q_n \to q, \quad \exists\, n_0 \in \mathbb{N} \quad \text{such that} \quad q_n \in \Sigma, \quad \forall\, n > n_0.$$

Due to the existence of minimizers there exists a sequence of controls u_n, such that u_n is a minimizer control that join q_0 to q_n. Moreover, by Proposition 10.9, the set of minimizers is strongly compact in L_2. Then there exist v such that $u_n \to v$. On the other hand the control v is forced to be a minimizer joining q and q_0, and by our assumption on uniqueness, we have $v = u$.

By smoothness of the end point map and $u_n \to u$ we get $D_{u_n}F \to D_u F$. Since $D_u F$ has full rank (u is strictly normal, hence is not a critical point for F) we have that, for n big enough, u_n is not a critical point of F. There exists a sequence $\lambda_1^n \in T^*_{q_n}M$ such that

$$\lambda_1^n D_{u_n}F = u_n, \qquad D_{u_n}F : L^2_k \to T_{q_n}M \quad \text{has full rank}$$

Considering the dual map

$$(D_{u_n}F)^*\lambda_1^n = u_n, \qquad D_{u_n}F^* : T^*_{q_n}M \to L^2_k \quad \text{is injective}$$

From here we get that also $\lambda_1^n \to \lambda_1$ and λ_1 must be the unique solution for q.

Remark 10.16. In other words we proved that $\forall V_\lambda$ nieghborhood of λ in $T^*_{q_0}M$ there exists a neighborhood O_q of q such that every point of $q' \in O_q$ is joined to q_0 by a minimal control u', whose corresponding covector is $\lambda' \in V_\lambda$.

Now it remains to prove that the covector λ_1 is unique and that μ is smooth at these points. Since at λ the exponential map is regular we have that $\mathcal{E}\big|_{V_\lambda} : \lambda' \mapsto \mathcal{E}(\lambda')$ is locally invertible. Hence

$$\forall\, q \in O_q \quad \exists!\, \lambda' \in V_\lambda \quad \text{s.t.} \quad q' = \mathcal{E}(\lambda') \quad \text{and} \quad q' \mapsto \lambda' \quad \text{is smooth.}$$

The conclusion follows from the equality $\mu(q') = H(\lambda')$ we are done.

□

□

Now we continue proving Theorem 10.4.

Proof of Theorem 10.4. Our goal is to show that Σ is a dense set in B. We start by characterizing some larger set than the set of smooth points[2].

Definition 10.17. A point $q \in B$ is said to be a

- *fair point* if there exists a unique normal minimizer joining q_0 to q.
- *good point* if it is a fair point and the control is strictly normal.

We will denote by Σ_f and Σ_g the set of fair and good points respectively. Clearly $\Sigma \subset \Sigma_g \subset \Sigma_f$.

Notice that a smooth point is a good point which is also a regular point for the exponential map [3]. We proceed into the following steps.

(i) Σ_f is a dense set,

(ii) Σ_g is a dense set,

(iii) μ is Lipschitz in a neighborhood of points of Σ_g,

(iv) Σ is a dense set.

[2]recall that we always work in the compact ball B

[3]i.e. not only for the endpoint map, but also for its restriction to the manifold of critical control

Proof of (i). Let $O \subset B$ an open set. We want to 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. Then consider the function

$$\Psi : q \in O \mapsto \mu(q) - a(q)$$

Clearly Ψ attains its minimum at some point $q_1 \in O$.

Now consider also the map

$$\Psi : u \in \mathcal{U} \mapsto J(u) - a(F(u))$$

Since $J \geq \mu$ we get that Ψ attain its minimum at the control u_1 that satisfy $F(u_1) = q_1$ and is a minimizer. Then

$$D_{u_1}\Phi = u_1 - (d_{q_1}a)D_{u_1F} = 0$$

form which it follows that u_1 is normal with $\lambda_1 = d_{q_1}a$ its covector.

Remark 10.18. In the Riemannian case $\Sigma_f = \Sigma_g$ since there are no abnormal extremal.

Proof of (ii). We want to show that $\Sigma_g \cap O \neq \emptyset$ for any open subset $O \subset B$.
For any $q \in \Sigma_f \cap O$ (which is nonempty by (i)) define $\operatorname{rank} q = \operatorname{rank} D_u F$, where u is the control associated to the unique minimizer γ that join q_0 to q. To prove (ii) 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, for the unique control u' associated to q'. Assume by contradiction that

$$k_O := \max_{q \in \Sigma_f \cap O} \operatorname{rank} q < n,$$

and consider a point $\widehat{q}$ such that $\operatorname{rank} \widehat{q} = k_O$.

Claim: all points sufficiently close to $\widehat{q}$ have the same rank.

Indeed, if it is not the case, there exists a sequence of points $q_n \to \widehat{q}$ such that $q_n \in \Sigma_f \cap O$, $\operatorname{rank} q_n < k_O$. But this implies that the sequence of controls u_n associated to q_n satisfies $u_n \to \widehat{u}$ strongly in L_2, by uniqueness and compactness (see also proof of (a) of Proposition 10.15). By smoothness of F it follows that $D_{u_n}F \to D_{\widehat{u}}F$ which implies the contradiction $\operatorname{rank} D_{\widehat{u}}F < k_O$.

Hence we can assume that $\operatorname{rank} q = k_O < n$ for every $q \in \Sigma_f \cap O$ (maybe restricting our neighborhood).

We introduce the following set

$$\Pi_q = e^{-\vec{H}}\{\xi \in T_q^*M \mid \xi D_u F = \lambda_1 D_u F\} \subset T_{q_0}^*M$$

which is an affine subset of $T_{q_0}^*M$ such that $\dim \Pi_q = k_O$. Indeed, let P_t be the (local) nonautonomous flow associated to the control u (which is the same for every ξ), then the the map $e^{-\vec{H}}$ acts linearly on the subspace of covectors associated to the same control u since we know that $\lambda(0) = e^{-\vec{H}}(\lambda_1) = P_t^*\lambda_1$, (see also Chapter 4).

Moreover, the map $q \mapsto \Pi_q$ is continuous on $\Sigma_f \cap O$. Indeed if we consider a sequence $q_n \to q$ we have that $u_n \to u$ strongly and $D_{u_n}F \to D_uF$. Since rank $D_{u_n}F$ is constant the kernel also is continuous.

Consider now $B \subset T_{q_0}^*M$ a k_O-dimensional ball that contains $\lambda_0 = e^{-\vec{H}}(\lambda_1)$ and is transversal to Π_q. By continuity B is transversal also to $\Pi_{q'}$, for $q' \in \Sigma_f \cap O$ close to q. In particular $\Pi_{q'} \cap B \neq \emptyset$.

This implies, since $\mathcal{E}(\Pi_q) = q$, that $\Sigma_f \cap O \subset \mathcal{E}(B)$. By (i) $\Sigma_f \cap O$ is a dense set, hence $\mathcal{E}(B)$ is also dense. On the other hand, since $\mathcal{E}$ is a smooth map and B is a compact ball of positive codimension ($k_O < n$), by Sard Lemma it follows that $\mathcal{E}(B)$ has measure zero, that is a contradiction.

Proof of (iii). We start with the following

Theorem 10.19. *Let $K \subset B$ a compact in our ball such that any minimizer connecting q_0 to $q \in K$ is strictly normal. Then μ is Lipschitz on K.*

Corollary 10.20. *If $q \in \Sigma_g$, then μ is Lipschitz in a neighborhood of q.*

Proof. It is sufficient to prove that in a neighborhood of good points there can be only points reached by strictly normal minimizers (no uniqueness is required). Assume the contrary, then there exists a sequence of points q_n such that u_n is also 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. $\square$

Proof of Theorem 10.19. Consider some point $q \in K$ and a minimizing control u (maybe is not unique). By our assumptions D_uF is surjective (u is strictly normal). Then, using inverse function theorem, there is a smooth right inverse. In other words there exists $\varepsilon > 0$ and $C > 0$ such that

$$B_q(Cs) \subset F(\mathcal{B}_u(s)), \qquad \forall\, 0 < s < \varepsilon$$

where $\mathcal{B}_u(r)$ is the ball of radius r in L^2. From this it follows that for every q and any minimizer u which connect q_0 to q:

$$\mu(q') \leq \mu(q) + C|q - q'| \tag{10.3}$$

By compactness of minimizers and compactness of K it follows that we can find ε and C such that 10.3 is valid for all points and minimizers. Then we can exchange the role of q and q' getting

$$|\mu(q') - \mu(q)| \leq C|q - q'| \tag{10.4}$$

$\square$

Proof of (iv). Since we know that μ is Lipschitz on O, μ is differentiable almost everywhere in O. Moreover, every point of differentiability of μ is a fair point.

Indeed consider the functional

$$\Psi : u \mapsto J(u) - \mu(F(u))$$

As noticed in (i), Ψ attains minimum at minimizers of μ. Moreover, if μ is differentiable at the point $F(u)$ (where u is the minimizer) then Ψ is differentiable at u and

$$D_u\Psi = u - (d_{F(u)}\mu)D_uF = 0$$

that implies that the point is fair. Hence the set of fair point has full measure in O. Moreover, by Sard Lemma, the set of regular values of the exponential map is also a dense set. Since all fair points are in the image of exponential map, for which thanks to Sard lemma almost all points are regular. The intersection is still dense. $\square$

We end this section by proving the following

Theorem 10.21. *For every $\varepsilon > 0$ there exists $q_0 \in M$ and a normal extremal path γ starting from q_0 such that $\ell(\gamma) = \varepsilon$ and γ is not a minimizer.*

10.2 Lipschitz functions and maps

Recall that if $\varphi : M \to \mathbb{R}$ is a Lipschitz function, the differential $d_q\varphi$ is well defined for a.e. $q \in M$, by Rademacher Theorem. Now we introduce a weaker definition of differentiability.

Definition 10.22. Let $\varphi : M \to \mathbb{R}$ be locally Lipschitz function. The (Clarke) *sub-differential* of ϕ at the point $q \in M$ is

$$\partial_q\varphi := \mathrm{conv}\{\xi \in T_q^*M \,|\, \xi = \lim_{q_n \to q} d_{q_n}\varphi,\ q_n \text{ diff. point}\} \tag{10.5}$$

Notice that by definition the set $\partial_q\varphi$ is bounded and closed, hence compact.

Note. In what follows if $\varphi : M \to \mathbb{R}$ is a locally Lipschitz function the notation $d_q\varphi$ means that $q \in M$ is a differentiability point of φ.

Example 10.23. 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 $x \geq 0$, then $\partial_0\varphi = [1, 2]$.

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.

This notion permits to extend some classical properties of critical points of smooth functions.

Proposition 10.24. *Let $\varphi : M \to \mathbb{R}$ be locally Lipschitz and q be a minimum for φ. Then $0 \in \partial_q\varphi$.*

Proof. Assume by contradiction that $0 \notin \partial_q\varphi$. Then by compactness the set $\partial_q\varphi$ is separated from 0. In particular from this follows that

$$\exists\, \varepsilon > 0,\ \exists\, v \in T_qM \qquad \text{s.t.} \qquad \langle \xi, v\rangle \leq -\varepsilon < 0, \quad \forall\, \xi \in \partial_q\varphi,$$

By continuity there exists a neighborhood O_q of q and V_v of v such that

$$\left\langle d_{q'}\varphi, v'\right\rangle \leq -\varepsilon/2 < 0, \qquad \forall\, q' \in O_q, \quad \forall\, v' \in V_v,$$

where q' is a differentiability point. Since for a.e. direction v' the intersection of the set of differentiable points with the line $\{q + tv'\}$ has full measure, the function $a(t) = \varphi(q + tv')$ cannot have a minimum in q. Indeed

$$a(t) - a(0) = \varphi(q + tv') - \varphi(q) = \int_0^t \left\langle d_q\varphi, v'\right\rangle \leq -\varepsilon t/2.$$

$\square$

The following lemma gives an estimate for the sub-differential of some special class of function.

Lemma 10.25. *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:*

$$(\omega, q) \mapsto \varphi_\omega(q), \qquad (\omega, q) \mapsto d_q\varphi_\omega(q)$$

Then the function $a(q) := \min_{\omega\in\Omega} \varphi_\omega(q)$ is locally Lipschitz and

$$\partial_q a \subset \mathrm{conv}\{d_q\varphi_\omega \,|\, \forall\, \omega \in \Omega \ s.t.\ \varphi_\omega(q) = a(q)\}. \tag{10.6}$$

Proof. We divide the proof into two steps. First we prove that a is locally Lipschitz and then we prove the estimate (10.6).

(i). It is enough to prove the statement for Lipschitz functions on a compact $K \subset M$. Since every φ_ω is Lipschitz on K and Ω is compact, there exists a constant $C > 0$ such that

$$\varphi_\omega(q) - \varphi_\omega(q') \leq C|q - q'|, \qquad \forall\, q, q' \in K, \quad \omega \in \Omega,$$

Clearly

$$\min_{\omega \in \Omega} \varphi_\omega(q) - \varphi_\omega(q') \leq C|q - q'|$$

and passing to the min with respect to q' we get

$$a(q) - a(q') \leq C|q - q'|, \qquad \forall\, q, q' \in K,$$

Since the constant C depends only on K we can exchange the role of q and q', proving

$$|a(q) - a(q')| \leq C|q - q'|, \qquad \forall\, q, q' \in K,$$

(ii). To prove (10.6) it is sufficient to show that, at every differentiable point $q \in M$

$$d_q a \in D, \qquad D := \operatorname{conv}\{d_q \varphi_\omega |\, \forall\, \omega \in \Omega \ s.t.\ \varphi_\omega(q) = a(q)\}.$$

Take $\xi \notin D$. By separation theorem

$$\exists\, \varepsilon > 0,\ \exists\, v \in T_q M \quad \text{s.t.} \quad \langle d_q \varphi_\omega, v\rangle > \langle \xi, v\rangle + \varepsilon, \quad \forall\, \omega \text{ s.t. } \varphi_\omega(q) = a(q),$$

By continuity there exists a neighborhood O_q of q and V_ω of ω such that

$$\left\langle d_{q'} \varphi_{\omega'}, v\right\rangle > \langle \xi, v\rangle + \varepsilon/2, \qquad \forall\, q' \in O_q, \quad \forall\, \omega' \in V_\omega \text{ s.t. } \varphi_{\omega'}(q') = a(q'),$$

A similar argument let us to prove that

$$\frac{1}{t}(\varphi_\omega(q + tv) - \varphi_\omega(q)) > \langle \xi, v\rangle + \varepsilon/4$$

which implies, repeating the argument used in (i),

$$\frac{1}{t}(a(q + tv) - a(q)) \geq \langle \xi, v\rangle + \varepsilon/4$$

and passing to the limit we get

$$\langle d_q a, v\rangle \geq \langle \xi, v\rangle + \varepsilon/4 \tag{10.7}$$

If $d_q a \notin D$ we can replace ξ with $d_q a$ in (10.7) getting $\langle d_q a, v\rangle \geq \langle d_q a, v\rangle + \varepsilon/4$. $\square$

For a Lipschitz map between manifolds $f : M \to N$ the (Clarke) *sub-differential* is defined in analogous way to the scalar case

$$\partial_q f := \operatorname{conv}\{L \in \operatorname{Hom}(T_q M, T_{f(q)} N) |\, L = \lim_{q_n \to q} D_{q_n} \varphi,\ q_n \text{ diff. point}\},$$

The following lemma shows how the standard chain rule extends to the Lipschitz case.

Lemma 10.26. *Let M be a C^1 manifold and $f : M \to N$ be a Lipschitz map.*

(a) *If $\phi : M \to M$ is a diffeomorphism and $q \in M$ we have*

$$\partial_q(f \circ \phi) = \partial_{\varphi(q)} f \cdot D_q \phi. \tag{10.8}$$

(b) *If $\varphi : N \to W$ is a C^1 map, and $q \in M$ we have*

$$\partial_q(\varphi \circ f) = D_{f(q)} \varphi \cdot \partial_q f. \tag{10.9}$$

Moreover the sub-differential, as a set, is upper semicontinuous*, i.e. for every neighborhood $\Omega \in \mathrm{Hom}(T_qM, T_{f(q)}N)$ of $\partial_q f$ there exisxt a neighborhood O_q of q such that $\partial_{q'} f \in \Omega$, for every $q' \in O_q$.*

Proof. (i). Since ϕ is a diffeomorphism, it sends every differentiability point q of $f \circ \phi$ to a differentiability point $\phi(q)$ for f. Then (10.8) is true at differentiability point and passing to the limit it is also valid for sub-differential (you prove both inclusion using ϕ and ϕ^{-1}).

With analogous reasoning one can prove (ii). The semicontinuity can be easily proved by separation theorem or the Caratheodory Lemma. □

Definition 10.27. Let $f : M \to N$ be a 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*.

Theorem 10.28. *Let $f : \mathbb{R}^n \to \mathbb{R}^n$ be a Lipschitz map and $q \in M$ be a regular point. Then there exists neighborhood $O_{f(q)}$ and a 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 10.29. The C^1 version of inverse function theorem can be proved from Theorem 10.28 and the chain rule. Indeed Theorem 10.28 implies that there exists a Lipschitz inverse g and using the chain rule it is easy to show that the sub-differential of g contains only one element (hence is differentiable at that point) and the differential is the inverse of the differential of f.

We start the proof of the Theorem 10.28 with two lemmas

Lemma 10.30. *There exists a neighborhood O_q and $\varepsilon > 0$ such that*

$$\forall\, v, |v| = 1,\ \exists\, \xi_v, |\xi_v| = 1 \qquad \text{s.t.} \qquad \langle \xi_v, \partial_x f(v) \rangle > \varepsilon, \quad \forall\, x \in O_q.$$

Proof. Since $\partial_q f(v)$ is a compact convex set that does not contain 0 (all matrix in $\partial_q f$ are invertibles) by separation theorem we can find ξ_v such that $\langle \xi_v, \partial_x f(v) \rangle > \varepsilon(v), \quad \forall\, x \in O_q$. By compactness of the set of v, there exists $\varepsilon > 0$ that works for all $|v| = 1$. □

Lemma 10.31. $|f(x) - f(y)| \geq \varepsilon |x - y|$, *for all $x, y \in O_q$.*

Proof. Write $y = x + sv$, where $s = |x-y|$ and v is a vector of norm 1. Consider a direction v' close to v such that almost every point in that direction is a point of differentiability, and let $y' = x + sv'$. Then we can write

$$f(y') - f(x) = \int_0^s (D_{x+tv'} f) v' dt$$

Hence

$$\begin{aligned} |f(y') - f(x)| &\geq \left\langle \xi_{v'}, f(y') - f(x) \right\rangle \\ &= \int_0^s \left\langle \xi_{v'}, (D_{x+tv'} f) v' \right\rangle dt \\ &\geq \varepsilon |y' - x| \end{aligned}$$

Then we can pass to the limit for $v' \to v$ (i.e. $y' \to y$) since ε does not depend on v. □

Proof of Theorem 10.28. Lemma 10.31 shows that f is injective in a neighborhood O_q of the point q and that the inverse function (which is well defined) is Lipschitz. It remains to show that $f(O_q)$ covers a neighborhood of the point $f(q)$.

Without lack of generality we can assume that the estimate of the Lemma 10.31 holds also on ∂O_q (maybe considering some smaller neighborhood). Lemma 10.31 also says that

$$\operatorname{dist}(f(q), \partial f(O_q)) \geq \varepsilon \operatorname{dist}(q, O_q) > 0$$

Then consider $W \subset f(O_q)$ such that $|y - q| < \operatorname{dist}(y, \partial f(O_q))$, for every $y \in W$.

Fix $y \in W$ and define the function

$$\varphi : \overline{O_q} \to \mathbb{R}, \qquad \varphi(x) = |f(x) - y|^2$$

By construction $\varphi(q) < \varphi(z)$, for all $z \in \partial f(O_q)$, hence by continuity φ attains the minimum on some point $x \in O_q$. By Proposition 10.24 $0 \in \partial_x \varphi$. Moreover, using the chain rule

$$\partial_x \varphi = (f(x) - y)^T \cdot \partial_x f$$

and since x is a regular point of f, the set $\partial_x \varphi$ contains zero if and only if $f(x) = y$. □

Corollary 10.32. *Let $\varphi : M \to \mathbb{R}$ be Lipschitz and assume that $y \in \mathbb{R}$ is a regular value of a, i.e. $\varphi^{-1}(y) \neq \emptyset$ and every $x \in \varphi^{-1}(y)$ is regular). Then $\varphi^{-1}(y)$ is a Lipschitz submanifold of M of codimension 1.*

Proof. We show that in any neighborhood O_x of $x \in \varphi^{-1}(y)$ the set $O_x \cap \varphi^{-1}(y)$ can be described as the zero locus of a Lipschitz function. Since $\partial_x a$ does not contain 0 there exists v, of norm 1, such that $\langle \partial_x \varphi, v \rangle > 0$ for every x in the compact neighborhood $O_x \cap \varphi^{-1}(y)$.

Then complete v to a orthonormal basis $(v, \xi_2, \dots, \xi_n)$ of $\mathbb{R}^n$ and consider the map

$$f : O_x \to \mathbb{R}^n, \qquad f(x') = \begin{pmatrix} \varphi(x') \\ \langle \xi_2, x' \rangle \\ \vdots \\ \langle \xi_n, x' \rangle \end{pmatrix}$$

By construction f is Lipschitz and x is a regular value of f.. Hence there exists, by Theorem 10.28 a Lipschitz inverse g of f. In particular the inverse map transforms the hyperplane $y_1 = \mathit{const}$ into the level of φ. Hence the level of φ is a Lipschitz submanifold. □

10.2.1 A non-smooth version of Sard Lemma

In this section we prove a Sard Lemma-type result for some special class of Lipschitz functions.

Recall that the classical Sard Lemma says that, if

$$\varphi : \mathbb{R}^n \to \mathbb{R}^m, \qquad \varphi \in C^k, \qquad k \geq \max\{n - m + 1, 1\},$$

and X denotes the critical set of φ, i.e. the set of points x in $\mathbb{R}^n$ at which the Jacobian matrix of φ has rank smaller than m, then $\varphi(X)$ has Lebesgue measure 0 in $\mathbb{R}^m$.

In particular, it does not apply even for C^1 functions $\varphi : \mathbb{R}^n \to \mathbb{R}$, when $n \geq 1$.

Theorem 10.33. *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) $\Omega = \bigcup_{\alpha\in\mathbb{N}} N_\alpha$ *is the union of smooth submanifold, and is compact,*

(ii) *the map* $(\omega, q) \mapsto \varphi_\omega(q)$ *and* $(\omega, q) \mapsto d_q\varphi_\omega$ *are continuous on* $\Omega \times M$*,*

(iii) *the maps* $\psi_\alpha : N_\alpha \times M \to \mathbb{R}$, $(\omega, q) \mapsto \varphi_\omega(q)$ *are smooth.*

Then the set of regular values of the function $a(q) = \min\limits_{\omega\in\Omega} \varphi_\omega(q)$ *has full measure in* $\mathbb{R}$*.*

Proof. It is enough to find a countable set of smooth functions such that any critical point of a is a critical point of one of these fucntions with the same value. Then this result reduce to standard Sard Lemma.

By Lemma 10.25 we already know that a is Lipschitz. Assume that q is a critical point of a, then

$$0 \in \partial_q a \subset \operatorname{conv}\{d_q\varphi_\omega | \,\forall\, \omega \in \Omega \ s.t. \ \varphi_\omega(q) = a(q)\}.$$

In other words there exists finite number $\lambda_1, \dots, \lambda_\ell$ such that $\lambda_i > 0, \sum_{i=1}^{\ell} \lambda_i = 1$ and

$$0 = \sum_{i=1}^{\ell} \lambda_i d_q\varphi_{\omega_i}, \qquad \varphi_{\omega_i}(q) = a(q), \quad \forall\, i = 1, \dots, \ell.$$

Moreover, since $\varphi_{\omega_i}(q) = a(q) = \min_\Omega \varphi_\omega(q)$, if $\omega_i \in N_{\alpha_i}$ then ω_i is critical for the restriction function $\omega \mapsto \varphi_\omega(q)\big|_{N_{\alpha_i}}$.

Then denote by $\Lambda_\ell = \{(\lambda_1, \dots, \lambda_\ell) | \,\lambda_i > 0, \sum \lambda_i = 1\}$ and consider the map

$$B_\ell : \bigcup_{i=1}^{\ell} N_{\alpha_i} \times \Lambda_\ell \times M \to \mathbb{R}$$

$$(\omega_0, \dots, \omega_\ell, \lambda_0, \dots, \lambda_\ell, q) \mapsto \sum_{i=0}^{\ell} \lambda_i \varphi_{\omega_i}(q) \tag{10.10}$$

It is easy to see that

$$d_z B_\ell = 0 \iff \begin{cases} \dfrac{\partial \varphi_{\omega_i}}{\partial \omega_i} = 0, \qquad i = 1, \dots, \ell \\ \sum_{i=1}^{\ell} \lambda_i d_q\varphi_{\omega_i} = 0 \\ \varphi_{\omega_0}(q) = \ldots = \varphi_{\omega_\ell}(q) \end{cases} \tag{10.11}$$

□

Corollary 10.34. *Let $\phi : N \to M$ be a smooth map between finite dimensional manifolds and $\varphi : N \to \mathbb{R}$ be a smooth function. Assume that*

(i) ϕ *is a submersion*

(ii) $\forall\, q \in M$ *the set* $C_q = \{x \in N, \, \varphi(x) = \min\limits_{\phi^{-1}(q)} \varphi(x)\} \neq \emptyset$ *is compact.*

Then the function $x \mapsto \min\limits_{\phi^{-1}(q)} \varphi(x)$ *satisfies the Theorem 10.33.*

Proof. It is sufficient to prove the statement locally, i.e. for the restriction of our function to every element of a countable covering of M. Consider a point $q \in M$. From (i) we know that $\phi^{-1}(q)$ is a smooth submanifold in N. Let us consider now a function $a : N \to \mathbb{R}$ and $c > 0$ such that the following assumptions are satisfied

(a) $A_\alpha := a^{-1}([0, \alpha])$ is compact for every $\alpha > 0$.

(b) $C_q \subset \operatorname{int} A_c$,

(c) c is a regular level of $a\big|_{\phi^{-1}(q)}$.

By continuity it follows that (a)-(c) are satisfied also for every q' in some neighborhood O_q of the point q. Noting that (c) is equivalent to the fact that level set of ϕ are transversal to level of a we can conclude that $\phi^{-1}(O_q) \cap A_c$ has the structure of locally trivial bundle. Maybe restricting the neighborhood of q then we can assume

$$\phi^{-1}(q) \cap A_c = \Omega, \qquad \phi^{-1}(O_q) \cap A_c \simeq O_q \times \Omega,$$

where Ω is a smooth manifold with boundary. In other words in this neighborhood we can split variables and our function is rewritten as $\min_{\omega \in \Omega} \varphi(\omega, q)$. We can apply the previous theorem to the function restricted to this domain (note that Ω is compact and is the union of its interior and its boundary, which are smooth by (a)-(c)). Since we can cover M by a countable family of this neighborhoods we are done. □

Remark 10.35. Notice that we do not assume that N is compact. In that case the proof is easier since every submersion $\phi : N \to M$ with N compact automatically endows N with a locally trivial bundle structure.

Now we apply the previous result to the sub-Riemannian distance.

Theorem 10.36. *Assume that* $B_{q_0}(r_0)$ *does not contain abnormal minimizers. Then* μ *is Lipschitz and the sphere* $S_{q_0}(r)$ *is a Lipschitz submanifold, for a.e.* $r \leq r_0$, .

Proof. Since there are no abnormal extremals every normal extremal is strongly normal.

Remark 10.37. Recall that for every strongly normal extremal $\lambda(t)$ there exists t_0 such that $t\lambda(0)$ is a regular point of the exponential map, for $0 < t \leq t_0$. Moreover, given a compact

$$K \subset T^*_{q_0}M \setminus (H^{-1}(0) \cap T^*_{q_0}M)$$

this t_0 can be chosen in a uniform way with respect to the initial condition $\lambda(0) \in K$. (see also Corollary 7.35)

We want to construct a map from a finite dimensional manifold into M, to apply the previous result and prove that almost every level set of the distance is a Lipschitz submanifold.

The idea is to find an appropriate modification of the exponential map. Remove a small ball B_δ, with $\delta > 0$ around the point q_0 and consider the set $D_\delta = B_r \setminus B_\delta$

Claim: the set $C = \{\lambda_0 \in T^*_{q_0}M,\ \lambda_0 \text{ minimizer},\ \mathcal{E}(\lambda_0) \in \overline{D}_\delta\}$ is precompact.

Indeed assume that there exists a sequence λ_n of covectors (and the associate sequence u_n of controls) such that $|\lambda_n| \to +\infty$. Since they are all normal extremals they satisfies

$$\lambda_n D_{u_n} F = u_n \Longrightarrow \frac{\lambda_n}{|\lambda_n|} D_{u_n} F = \frac{u_n}{|\lambda_n|} \tag{10.12}$$

and using compactness of minimizers we can assume that $\lambda_n \to \lambda$ and $u_n \to u$. Passing to the limit in (10.12) we find $\lambda D_u F = 0$, that is not possible since only normal minimizer reach points of $\overline{D}_\delta$. (the only abnormal is the zero control $u \equiv 0$, which is removed in our construction)

Hence the set $C_1 := \overline{C} \subset T^*_{q_0} M$ is compact. Moreover define

$$C_2 = \{\lambda_1 \in C_1 \cap H^{-1}(]0, \varepsilon[)\}$$

where ε is chosen in such a way that $A_{\lambda_0}\lambda_1$ is a regular point of $\mathcal{E}_{\mathcal{E}_{q_0}(\lambda_0)}$ for every λ_0, λ_1. [4] where

$$A_{\lambda_0} : T^*_{q_0} M \to T^*_{\mathcal{E}_{q_0}(\lambda_0)} M$$

is the pullback of the flow defined by the control u_0 associated to λ_0.

Define the map

$$\Psi : C_1 \times C_2 \to D_\delta \subset M, \qquad \Psi(\lambda_0, \lambda_1) = \mathcal{E}_{\mathcal{E}_{q_0}(\lambda_0)}(A_{\lambda_0}\lambda_1)$$

By construction Ψ is a submersion. Moreover, since $\Psi(\lambda_0, s\lambda_0) = \mathcal{E}((1+s)\lambda_0)$ for $0 < s < \varepsilon$, it follows that Ψ attains miminum exactly at the same points as the sub-Riemannian distance.

Since $\delta > 0$ is arbitrary we are done. $\square$

[4] it is possible to find $\varepsilon = \varepsilon(\lambda_0)$ that works for every $\lambda_1 \in C_2$, once λ_0 is fixed. Since C_1 is compact it is possible to find a unique ε that works for all.

Chapter 11

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.

11.1 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 $x \in V$

$$D_x F : V \to T_{F(x)}M, \qquad D_x F(v) = \frac{d}{dt}\Big|_{t=0} F(x+tv), \quad v \in V,$$

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

$$D_x^2 F(v) = \frac{d^2}{dt^2}\Big|_{t=0} F(x+tv) \tag{11.1}$$

has not geometric meaning if $D_x F(v) \neq 0$. 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 $D_x^2 F(v)$ vanish in these coordinates.

In general, the linear structure on V let us to define the second differential of F as a quadratic map

$$D_x^2 F : \operatorname{Ker} D_x F \to T_{F(x)}M \tag{11.2}$$

On the other hand the map (11.2) 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

$$D_x F : T_x N \to T_{F(x)}M, \qquad x \in N.$$

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):

$$D_x^2F(v) = \frac{d^2}{dt^2}\bigg|_{t=0} F(\gamma(t)), \qquad \gamma(0) = x, \quad \dot\gamma(0) = v \in \operatorname{Ker} D_xF, \tag{11.3}$$

Computing in coordinates we find that

$$\frac{d^2}{dt^2}\bigg|_{t=0} F(\gamma(t)) = \frac{d^2F}{dx^2}(\dot\gamma(0), \dot\gamma(0)) + \frac{dF}{dx}\ddot\gamma(0) \tag{11.4}$$

that shows that term (11.4) 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

$$\operatorname{Hess}_x F : \operatorname{Ker} D_xF \to T_{F(x)}M / \operatorname{Im} D_xF$$

11.2 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 11.1. *Assume that there are no normal minimizers that join q_0 to $\widehat{q}$. Then μ is not Lipschitz in a neighborhood of $\widehat{q}$. Moreover*

$$\lim_{\substack{q\to\widehat{q}\\ q\in\Sigma}} |d_q\mu| = +\infty. \tag{11.5}$$

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

$$\lambda_n D_{u_n}F = u_n, \qquad \lambda_n = d_{q_n}\mu.$$

Assume by contradiction that $|d_{q_n}\mu| \leq M$ 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

$$\operatorname{Hess}_u F : \operatorname{Ker} D_uF \to \operatorname{Coker} D_uF = T_{F(u)}M/\operatorname{Im} D_uF.$$

Remark 11.2. 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

$$\lambda \operatorname{Hess}_u F : \operatorname{Ker} D_uF \to \mathbb{R}$$

Notice that the dimension of the space $\operatorname{Im} D_uF^\perp$ of such covectors coincide with $\dim \operatorname{Coker} D_uF$.

Definition 11.3. 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:

$$\operatorname{ind} Q = \sup\{\dim W \mid Q\big|_{W\setminus\{0\}} < 0\}. \tag{11.6}$$

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 11.4. Let $F : \mathcal{U} \to M$ and $u \in \mathcal{U}$ be a critical point for F. The *index* of F at u is

$$\operatorname{Ind}_u F = \min_{\lambda \perp \operatorname{Im} D_u F} \operatorname{ind}(\lambda \operatorname{Hess}_u F) - \operatorname{codim} \operatorname{Im} D_u F$$

Theorem 11.5. *If* $\operatorname{Ind}_u F \geq 1$*, then* u *is not a strictly abnormal minimizer.*

We state without proof the following result (see Lemma 20.8 of [4])

Lemma 11.6. *Let* $Q : \mathbb{R}^N \to \mathbb{R}^n$ *be a vector valued quadratic form. Assume that* $\operatorname{Ind}_0 Q \geq 0$*. Then there exists a regular point* $x \in \mathbb{R}^n$ *of* Q *such that* $Q(x) = 0$*.*

Definition 11.7. 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 $C > 0$, $\bar{\varepsilon} > 0$ such that

$$\forall\, \varepsilon < \bar{\varepsilon}, \quad \exists\, \delta(\varepsilon) > 0, \qquad B_{\widehat{\Phi}(x)}(C\varepsilon^r) \subset \widehat{\Phi}(B_x(\varepsilon)), \qquad \forall\, \widehat{\Phi} \in \mathcal{C}^0,\ \|\widehat{\Phi} - \Phi\|_{\mathcal{C}^0} < \delta. \tag{11.7}$$

Exercise 11.8. Prove that if x is a regular point of Φ, then Φ is 1-solid at x.
(Use implicit function theorem to prove that Φ satisfies (11.7) and Brower theorem to show that the same holds for some small perturbation)

We can assume that $x = 0$ and that $\Phi(0) = 0$.

Proposition 11.9. *Assume that* $\operatorname{Ind}_0 \Phi \geq 0$*. Then* Φ *is 2-solid at* $x = 0$*.*

Proof. 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 $k := \dim \operatorname{Coker} D_0\Phi$ and consider the Hessian

$$\operatorname{Hess}_0 \Phi : \operatorname{Ker} D_0\Phi \to \operatorname{Coker} D_0\Phi$$

We can rewrite the assumption on the index of Φ as follows

$$\operatorname{ind} \lambda \operatorname{Hess}_0 \Phi \geq k, \qquad \forall\, \lambda \in \operatorname{Im} D_0\Phi^{\perp} \setminus \{0\}. \tag{11.8}$$

Since property (11.8) 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

$$E = E_1 \oplus E_2 \qquad E_2 := \operatorname{Ker} D_0\Phi$$

The Hessian is a map

$$\mathrm{Hess}_0\Phi : E_2 \to \mathbb{R}^n / D_0\Phi(E_1)$$

According to Lemma 11.6 there exists $e_2 \in E_2$, regular point of $\mathrm{Hess}_0\Phi$, such that

$$\mathrm{Hess}_0\Phi(e_2) = 0 \qquad \Longrightarrow \qquad D_0^2\Phi(e_2) = D_0\Phi(e_1), \quad \text{for some } e_1 \in E_1.$$

Define the map $Q : E \to \mathbb{R}^n$ by the formula

$$Q(v_1 + v_2) := D_0\Phi(v_1) + \frac{1}{2} D_0^2\Phi(v_2), \qquad v = v_1 + v_2 \in E = E_1 \oplus E_2.$$

and 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. Consider then the map

$$\Phi_\varepsilon : v_1 + v_2 \mapsto \frac{1}{\varepsilon^2}\Phi(\varepsilon^2 v_1 + \varepsilon v_2) \tag{11.9}$$

Using that $v_2 \in \operatorname{Ker} D_0\Phi$ we compute the Taylor expansion with respect to ε

$$\Phi_\varepsilon(v_1 + v_2) = Q(v_1 + v_2) + O(\varepsilon) \tag{11.10}$$

hence for small ε the image of Φ_ε contain a ball around 0 from which it follows that

$$B_{\phi(0)}(c\varepsilon^2) \subset \Phi(B_0(\varepsilon)) \tag{11.11}$$

Moreover as soon as ε is fixed we can perturb the map Φ and still the estimate (11.11) holds. $\square$

Actually we proved the following statement, that is stronger than 2-solideness of Φ:

Lemma 11.10. *Under the assumptions of the Theorem 11.9, there exists $C > 0$ such that for every ε small enough*

$$B_{\phi(0)}(C\varepsilon^2) \subset \Phi(B_0'(\varepsilon^2) \times B_0''(\varepsilon)) \tag{11.12}$$

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 (11.9) and (11.10) show. [1]

Proof of Theorem 11.5. 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

$$\Phi : \mathcal{U} \to \mathbb{R} \times M, \qquad u \mapsto \begin{pmatrix} J(u) \\ F(u) \end{pmatrix}$$

is locally open at u. Since u is strictly abnormal it means that

$$d_u J\big|_{\mathrm{Ker}\, D_u F} \neq 0$$

Indeed recall that $d_u J = \lambda\, D_u F$, for some $\lambda \in T_{F(u)} M$ is equivalent to $d_u J\big|_{\mathrm{Ker}\, D_u F} = 0$ (see Proposition 7.6).

Moreover

$$\mathrm{Ker}\, D_u \Phi = \mathrm{Ker}\, d_u J \cap \mathrm{Ker}\, D_u F, \qquad \dim \mathrm{Im}\, d_u J = 1$$

and from this it follows that

$$\mathrm{Hess}_u \Phi = \mathrm{Hess}_u F\big|_{\mathrm{Ker}\, d_u J \cap \mathrm{Ker}\, D_u F}$$

since differential of the first coordinate is independent from the other.

From here it follows that

$$\mathrm{Ind}_u \Phi \geq \mathrm{Ind}_u F - 1 \geq 0$$

Applying Proposition 11.9 we find that Φ is locally open at u. Hence u cannot be a minimizer. $\square$

Now we prove that, under the same assumptions on the index of the endpoint map given in Theorem 11.5, the sub-Riemannian is Lipschitz even if some abnormal minimizers are present.

Theorem 11.11. *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 μ is Lipschitz on K.*

Proof. Recall that if there are no abnormal minimizers, Theorem 10.36 ensures that μ is Lipschitz. Then, using compactness of the set of all minimizers, it is sufficient to prove the estimate near a point $q = F(u)$, where u is abnormal.

Since $\mathrm{Ind}_u F \geq 1$ by assumption, Theorem 11.5 implies that every abnormal minimizer is not strictly abnormal. Then we can assume that every abnormal minimizer u is both normal and abnormal. We have

$$\mathrm{Hess}_u F : \mathrm{Ker}\, D_u F \to \mathrm{Coker}\, D_u F, \qquad \text{with} \qquad \mathrm{Ind}_u F \geq 1.$$

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 $\mathrm{Ker}\, D_u F \subset \mathrm{Ker}\, d_u J$.

The assumption of Lemma 11.10 are satisfied, hence splitting the the space of controls

$$L^2_k([0,1]) = E_1 \oplus E_2, \qquad E_2 := \mathrm{Ker}\, D_u F$$

[1] $B_0(c) \subset \Phi_\varepsilon(B(1)) \Leftrightarrow B_0(c\varepsilon^2) \subset \Phi(\varepsilon^2 v_1 + \varepsilon v_2), v_i \in B^i(1) \Leftrightarrow B_0(c\varepsilon^2) \subset \Phi(B'_{\varepsilon^2} \times B''_\varepsilon)$

we have that there exists $C > 0$ such that for ε small enough

$$B_q(C\varepsilon^2) \subset F(\mathcal{B}_\varepsilon), \qquad \mathcal{B}_\varepsilon := \mathcal{B}'_u(\varepsilon^2) \times \mathcal{B}''_u(\varepsilon), \qquad q = F(u), \tag{11.13}$$

where $\mathcal{B}'_u(r)$ and $\mathcal{B}''_u(r)$ are the ball of radius r in E_1 and E_2 respectively.

Consider now coordinates on M and an element $x \in K$ such that $|x - q| = C\varepsilon^2$. Then (11.13) implies that there exists $v = (v_1, v_2) \in \mathcal{B}_\varepsilon$ such that $F(v) = x$. It follows that

$$\begin{aligned}
\mu(x) - \mu(q) &\leq J(v) - \mu(q) && \text{(by definition of } \mu) \\
&= J(v) - J(u) && \text{(since } u \text{ is a minimizer)} \\
&\leq C'|v_2| + C''|v_1 - u|^2 && \text{(using } d_uJ = 0 \text{ on } E_2) \\
&\leq \widetilde{C}\varepsilon^2 && \text{(by definition of } \mathcal{B}_\varepsilon) \\
&= \frac{\widetilde{C}}{C}|x - q|
\end{aligned} \tag{11.14}$$

Notice that $\widetilde{C}$ and C' does not depend on ε, hence by compactness of K we can find constants $C, C' > 0$ that does not depend on q and exchange the role of x and q in the formulas above, getting

$$|\mu(x) - \mu(q)| \leq c|x - q|$$

□

Now we present some necessary conditions for the index of the quadratic form along an abnormal extremal to be finite.

Theorem 11.12. *Let u be an abnormal minimizer and let $\lambda_1 \in T^*_{F(u)}M$ satisfies $\lambda_1 D_uF = 0$. Assume that* $\operatorname{ind} \lambda_1 \mathrm{Hess}_u F < +\infty$. *Then the following condition are satisfied :*

(i) $\langle \lambda(t), [f_i, f_j](\gamma(t)) \rangle \equiv 0,$ *for a.e. t, $\forall\, i, j = 1, \ldots, k$,* *(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.

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

$$\dot{\lambda}(t) = \sum_{i=1}^{k} u_i(t)\vec{h}_i(\lambda(t)), \qquad \langle \lambda(t), f_i(\gamma(t)) \rangle = 0,\ \forall\, i = 1, \ldots, k,$$

i.e. $\lambda(t) \in \mathcal{D}^\perp_{\gamma(t)}$. Goh condition are equivalent to require $\lambda(t) \in (\mathcal{D}^2_{\gamma(t)})^\perp$.

Corollary 11.13. *Assume that the sub-Riemannian structure is 2-generating, i.e. $\mathcal{D}^2_q = T_qM$ for all $q \in M$. Then there are no strictly abnormal minimizers. In particular μ is globally Lipschitz.*

Proof. Since $\mathcal{D}^2_q = T_qM$ implies $(\mathcal{D}^2_{\gamma(t)})^\perp = 0$ for every $q \in M$, no abnormal extremal can satisfy the Goh condition. Hence by Theorem 11.12 it follows that $\mathrm{Ind}_u F = +\infty$, for any abnormal minimizer u.

In particular, from Theorem 11.5 it follows that the minimizer cannot be strictly abnormal Hence μ is globally Lipschitz by Theorem 11.11. □

Remark 11.14. Notice that μ is globally Lipschitz 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 (Corollary 8.50) it follows that μ is not Lipschitz at q.

If Goh condition is satisfied, generalized Legendre condition can also be characterized as an intrinsic property of the module. Indeed one can see that the quadratic map

$$U_{\gamma(t)} \to \mathbb{R}, \qquad v \mapsto \big\langle \lambda(t), [[f_{u(t)}, f_v], f_v](\gamma(t)) \big\rangle$$

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

$$\{h_v, h_w\}(\lambda(t)) \equiv 0, \qquad \forall\, v, w \in \mathbb{R}^k \tag{11.15}$$

while strong Legendre conditions reads

$$\{\{h_{u(t)}, h_v\}, h_v\} \geq 0, \qquad \forall\, v \in \mathbb{R}^k \tag{11.16}$$

Taking derivative of (11.15) 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

$$(v, w) \mapsto \{\{h_{u(t)}, h_v\}, h_w\}(\lambda) \tag{11.17}$$

is symmetric. Hence the generalized Legendre condition says that the quadratic form associated to (11.17) is nonnegative.

Proof of Theorem 11.12. 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 7.2 we can write the end-point map as the composition

$$F(u) = P_1(G(u)), \qquad D_u F = P_{1*} D_0 G,$$

and reduced the problem to the expansion of G, which is easier. Indeed denoting $g_i^t := P_{t*}^{-1} f_i$, the map G can be interpreted as the end-point map for the system

$$\dot{q}(t) = g_{v(t)}^t(q(t)) = \sum_{i=1}^k v_i(t) g_i^t(q(t))$$

and the Hessian of F can be computed easily starting from the Hessian of G at $v = 0$

$$\mathrm{Hess}_u F = P_{1*} \mathrm{Hess}_0 G$$

from which we get, using that $\lambda_0 = P_1^*\lambda_1$,

$$\lambda_1 \mathrm{Hess}_u F = \lambda_1 P_{1*} \mathrm{Hess}_0 G = \lambda_0 \mathrm{Hess}_0 G$$

Moreover computing

$$\begin{aligned} \langle \lambda(t), [f_i, f_j](\gamma(t)) \rangle &= \left\langle \lambda_0, P_{t*}^{-1}[f_i, f_j](\gamma(t)) \right\rangle \\ &= \left\langle \lambda_0, [g_i^t, g_j^t](\gamma(0)) \right\rangle \end{aligned}$$

the Goh and generalized Legendre conditions can also be rewritten as

$$\left\langle \lambda_0, [g_i^t, g_j^t]\gamma(0) \right\rangle \equiv 0, \qquad \text{for a.e. } t \in [0,1], \quad \forall\, i,j = 1, \dots, k, \tag{$G.1$}$$
$$\langle \lambda_0, [[g_{u(t)}^t, g_i^t], g_i^t]](\gamma(0)) \rangle \geq 0, \qquad \text{for a.e. } t \in [0,1], \quad \forall\, i = 1, \dots, k. \tag{$L.1$}$$

Now we want to compute the Hessian of the map G. Using the Volterra expansion computed in Chapter **??** we have

$$G(v(\cdot)) \simeq q_0 \circ \left(\mathrm{Id} + \int_0^1 g_{v(t)}^t dt + \iint\limits_{0 \leq \tau \leq t \leq 1} g_{v(\tau)}^\tau \circ g_{v(t)}^t d\tau dt \right) + O(\|v\|^3)$$

where we used that g_v^t 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

$$D_0 G(v) = \int_0^1 g_{v(t)}^t (q_0) dt,$$

On the other hand the expression for the quadratic part, i.e. the second differential

$$D_0^2 G(v) = 2\, q_0 \circ \iint\limits_{0 \leq \tau \leq t \leq 1} g_{v(\tau)}^\tau \circ g_{v(t)}^t d\tau dt.$$

has not an immediate geometrical interpretation.

Recall that the second differential $D_0^2 G$ is defined on the set

$$\mathrm{Ker}\, D_0 G = \{ v \in L_k^2[0,1], \int_0^1 g_{v(t)}^t (q_0) dt = 0 \} \tag{11.18}$$

and, for such a v, $D_0^2 G(v)$ belong to the tangent space $T_{q_0}M$. Indeed, using Lemma 7.19, and that v belong to the set (11.18), we can symmetrize the second derivative, getting the formula

$$D_0^2 G(v) = \iint\limits_{0 \leq \tau \leq t \leq 1} [g_{v(\tau)}^\tau, g_{v(t)}^t](q_0) d\tau dt,$$

which shows that the second differential is computed by the integral mean of the commutator of the vector field $g_{v(t)}^t$ for different times.

Now consider an element $\lambda_0 \in \operatorname{Im} D_0 G^{\perp}$, i.e. that satisfies

$$\left\langle \lambda_0, g_v^t(q_0) \right\rangle = 0, \qquad \text{for a.e. } t \in [0,1], \forall\, v \in \mathbb{R}^k.$$

Then we can compute the Hessian

$$\lambda_0 \mathrm{Hess}_0 G(v) = \iint\limits_{0 \le \tau \le t \le 1} \langle \lambda_0, [g^{\tau}_{v(\tau)}, g^t_{v(t)}](q_0) \rangle d\tau dt \tag{11.19}$$

Remark 11.15. Denoting by K the bilinear form

$$K(\tau,t)(v,w) = \left\langle \lambda_0, [g_v^{\tau}, g_w^t](q_0) \right\rangle,$$

the Goh and generalized Legendre conditions are rewritten as follows

$$K(t,t)(v,w) = 0, \qquad \forall\, v, w \in \mathbb{R}^k, \quad \text{for a.e. } t \in [0,1], \tag{G.2}$$

$$\frac{\partial K}{\partial \tau}(\tau,t)\bigg|_{\tau=t}(v,v) \ge 0, \qquad \forall\, v \in \mathbb{R}^k, \quad \text{for a.e. } t \in [0,1]. \tag{L.2}$$

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

$$\frac{\partial}{\partial t} g_v^t = [g^t_{u(t)}, g^t_v]$$

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 $\lambda_0 \in \operatorname{Im} D_0 G^{\perp}$:

$$\eta_v^t := \vec{h}_v^t(\lambda_0) \in T_{\lambda_0} T^*M \tag{11.20}$$

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

$$\lambda_0 \mathrm{Hess}_0 G(v(\cdot)) = \iint\limits_{0 \le \tau \le t \le 1} \sigma_{\lambda_0}(\eta^{\tau}_{v(\tau)}, \eta^t_{v(t)}) dt d\tau. \tag{11.21}$$

where the control $v \in \operatorname{Ker} D_0G$ 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 $\operatorname{ind} \lambda_0 \mathrm{Hess}_0 G < +\infty$, 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 D_0G that are concentrated on small segments $[t, t+s]$.

To do this we consider an arbitrary vector control function $v : [0,1] \to \mathbb{R}^k$ such that its support is concentrated on $[0,1]$ and we build, for every $t \in [0,1]$ and s small enough the control

$$v_s(\tau) = v\left(\frac{\tau - t}{s}\right), \qquad \operatorname{supp} v_s \subset [t, t+s] \tag{11.22}$$

Then we apply the Hessian to this particular control functions and we compute the asymptotics for $s \to 0$.

Moreover, 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

$$E_s := \{v_s \in \operatorname{Ker} D_0G,\ v_s \text{ defined by (11.22)}, \int_0^1 v(\tau) d\tau = 0\}.$$

Notice in particular that $\operatorname{codim} E_s$ does not depend on s.

Remark 11.16. We will use the following identities (writing σ for σ_{λ_0}) , which holds for every control function $v : [0,1] \to \mathbb{R}^k$

$$\iint\limits_{\alpha \leq \tau \leq t \leq \beta} \sigma(\eta_{v(\tau)}^\tau, \eta_{v(t)}^t) dt d\tau = \int_\alpha^\beta \sigma(\int_\alpha^t \eta_{v(\tau)}^\tau d\tau, \eta_{v(t)}^t) dt = \int_\alpha^\beta \sigma(\eta_{v(\tau)}^\tau, \int_\tau^\beta \eta_{v(t)}^t dt) d\tau. \tag{11.23}$$

Moreover we have the integration by parts formula, where $w(t) = \int_0^t v(\tau) d\tau$:

$$\int_\alpha^\beta \eta_{v(t)}^t dt = \eta_{w(\beta)}^\beta - \eta_{w(\alpha)}^\alpha - \int_\alpha^\beta \dot{\eta}_{w(t)}^t dt. \tag{11.24}$$

Then we use equality (11.23) and we apply the Hessian to the function v_s (since the control is concentrated on the segment $[t, t+s]$ we can restrict the extrema of the integral)

$$\lambda_0 \mathrm{Hess}_0 G(v_s(\cdot)) = \int_t^{t+s} \sigma(\int_t^\tau \eta_{v_s(\theta)}^\theta d\theta, \eta_{v_s(\tau)}^\tau) d\tau. \tag{11.25}$$

The integration by parts, by our boundary conditions, gives

$$\int_t^\tau \eta^\theta_{v_s(\theta)} d\theta = \eta^\tau_{w_s(\tau)} - \int_t^\tau \dot\eta^\theta_{w_s(\theta)} d\theta. \tag{11.26}$$

where

$$w_s(\theta) = \int_t^\theta v_s(\tau) d\tau, \qquad \theta \in [t, t+s].$$

Using (11.26)

$$\begin{aligned} \lambda_0 \mathrm{Hess}_0 G(v_s(\cdot)) &= \int_t^{t+s} \sigma(\eta^\tau_{w_s(\tau)}, \eta^\tau_{v_s(\tau)}) d\tau - \int_t^{t+s} \sigma(\int_t^\tau \dot\eta^\theta_{w_s(\theta)} d\theta, \eta^\tau_{v_s(\tau)}) d\tau \\ &= \int_t^{t+s} \sigma(\eta^\tau_{w_s(\tau)}, \eta^\tau_{v_s(\tau)}) d\tau - \int_t^{t+s} \sigma(\dot\eta^\tau_{w_s(\tau)}, \int_\tau^{t+s} \eta^\theta_{v_s(\theta)} d\theta) d\tau \end{aligned} \tag{11.27}$$

where the second equality follows from (11.23).

Now consider the second term in (11.27) and apply again the integration by part formula (now we use the assumption $w(t+s) = 0$)

$$\begin{aligned} \int_t^{t+s} \sigma(\dot\eta^\tau_{w_s(\tau)}, \int_\tau^{t+s} \eta^\theta_{v_s(\theta)} d\theta) d\tau = &- \int_t^{t+s} \sigma(\dot\eta^\tau_{w_s(\tau)}, \eta^\tau_{w_s(\tau)}) d\tau \\ &- \int_t^{t+s} \sigma(\dot\eta^\tau_{w_s(\tau)}, \int_\tau^{t+s} \dot\eta^\theta_{w_s(\theta)} d\theta) d\tau \end{aligned}$$

Collecting together all the computations

$$\begin{aligned} \lambda_0 \mathrm{Hess}_0 G(v_s(\cdot)) = &\int_t^{t+s} \sigma(\eta^\tau_{w_s(\tau)}, \eta^\tau_{v_s(\tau)}) d\tau \\ &+ \int_t^{t+s} \sigma(\dot\eta^\tau_{w_s(\tau)}, \eta^\tau_{w_s(\tau)}) d\tau \\ &+ \int_t^{t+s} \sigma(\dot\eta^\tau_{w_s(\tau)}, \int_\tau^{t+s} \dot\eta^\theta_{w_s(\theta)} d\theta) d\tau \end{aligned}$$

Then using the identity

$$w_s(\theta) = s\, w\left(\frac{\theta - t}{s}\right),$$

and performing the change of variables

$$\zeta = \frac{\tau - t}{s}, \qquad \tau \in [t, t+s],$$

we come to the following expression for the Hessian:

$$\begin{aligned} \lambda_0 \mathrm{Hess}_0 G(v_s(\cdot)) = \; & s^2 \int_0^1 \sigma(\eta^{t+s\theta}_{w(\theta)}, \eta^{t+s\theta}_{v(\theta)}) d\theta \\ &+ s^3 \int_0^1 \sigma(\dot\eta^{t+s\theta}_{w(\theta)}, \eta^{t+s\theta}_{w(\theta)}) d\theta \\ &+ s^4 \int_0^1 \sigma(\dot\eta^{t+s\theta}_{w(\theta)}, \int_\theta^1 \dot\eta^{t+s\zeta}_{w(\zeta)} d\zeta) d\theta \end{aligned} \tag{11.28}$$

from which we get

$$\lambda_0 \mathrm{Hess}_0 G(v(\cdot)) = \; s^2 \int_0^1 \sigma(\eta^t_{w(\theta)}, \eta^t_{v(\theta)}) d\theta + O(s^3). \tag{11.29}$$

Since we assume $\operatorname{ind} \lambda_0 \mathrm{Hess}_0 G < +\infty$, this implies that the quadratic form

$$w(\cdot) \mapsto \int_0^1 \sigma(\eta^t_{w(\theta)}, \eta^t_{\dot w(\theta)}) d\theta \tag{11.30}$$

has finite index, otherwise by continuity every sufficiently small perturbation of (11.30) would have infinite index, contradicting our assumption on (11.29).

To prove that Goh conditions hold then it is sufficient to prove that if (11.30) has finite index then the integrand is zero, which is guaranteed by the following

Lemma 11.17. *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 $\operatorname{ind} Q < +\infty$ if and only if $A \equiv 0$.

Proof. Clearly if $A = 0$, then $Q = 0$ and $\operatorname{ind} Q = 0$. Assume then that $A \neq 0$ and we prove that $\operatorname{ind} Q = +\infty$. 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 $x = (x_1, \dots, x_N) \in \mathbb{R}^N$ one can built the function

$$w_x(t) = x_i\, w(Nt - i), \qquad t \in [\frac{i}{N}, \frac{i+1}{N}], \qquad i = 1, \dots, N.$$

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. □

Applying Lemma 11.17 for any t we prove that the s^2 order term in (11.28) vanish and we get to

$$\begin{aligned} \lambda_0 \mathrm{Hess}_0 G(v(\cdot)) &= s^3 \int_0^1 \sigma(\dot{\eta}^{t+s\theta}_{w(\theta)}, \eta^{t+s\theta}_{w(\theta)})d\theta + O(s^4) \\ &= s^3 \int_0^1 \sigma(\dot{\eta}^{t+s\theta}_{w(\theta)}, \eta^{t}_{w(\theta)})d\theta + O(s^4) \end{aligned}$$

where the last equalily follows from the fact that η^t_v is Lipschitz with respect to t (see also (11.20)), i.e.

$$\eta_v^{t+s\theta} = \eta_v^t + O(s)$$

On the other hand $\dot{\eta}^t_v$ 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 $\dot{\eta}^t_{w(\cdot)}$ is well defined and we can write

$$\begin{aligned} \lambda_0 \mathrm{Hess}_0 G(v(\cdot)) = s^3 &\int_0^1 \sigma(\dot{\eta}^{t}_{w(\theta)}, \eta^{t}_{w(\theta)})d\theta \\ &- s^3 \left(\int_0^1 \sigma(\dot{\eta}^{t+s\theta}_{w(\theta)}, \eta^{t}_{w(\theta)}) - \sigma(\dot{\eta}^{t}_{w(\theta)}, \eta^{t}_{w(\theta)})d\theta \right) + O(s^4) \end{aligned}$$

Using the linearity of σ and the boundedness of the vector fields we can estimate

$$\begin{aligned} \left| \int_0^1 \sigma(\dot{\eta}^{t+s\theta}_{w(\theta)}, \eta^{t}_{w(\theta)}) - \sigma(\dot{\eta}^{t}_{w(\theta)}, \eta^{t}_{w(\theta)})d\theta \right| &\leq C \int_0^1 |\dot{\eta}^{t+s\theta}_{w(\theta)} - \dot{\eta}^{t}_{w(\theta)}| d\theta \\ &\leq C \sup_{|v|\leq 1} \frac{1}{s}\int_0^1 |\dot{\eta}^{t+\tau}_v - \dot{\eta}^t_v| d\tau \underset{s\to 0}{\longrightarrow} 0 \end{aligned}$$

where the last term tends to zero by definition of Lebesgue point. Hence we come to

$$\lambda_0 \mathrm{Hess}_0 G(v(\cdot)) = s^3 \int_0^1 \sigma(\dot{\eta}^t_{w(\theta)}, \eta^t_{w(\theta)}) d\theta + o(s^3) \tag{11.31}$$

Then 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 11.17.

Lemma 11.18. *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.

□

Now we want to characterize the trajectories that satisfy these conditions. Recall that, if $\lambda(t)$ is an abnormal geodesic, we have

$$\dot{\lambda}(t) = \vec{h}_{u(t)}(\lambda(t)), \qquad h_i(\lambda(t)) \equiv 0, \quad 0 \le t \le 1. \tag{11.32}$$

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

$$h_{i_1 \dots i_k}(\lambda) = \{h_{i_1}, \dots, \{h_{i_{k-1}}, h_{i_k}\}\}(\lambda) \tag{11.33}$$

$$= \left\langle \lambda, [f_{i_1}, \dots, [f_{i_{k-1}}, f_{i_k}]](q) \right\rangle, \qquad q = \pi(\lambda) \tag{11.34}$$

Differentiating the identities in (11.32), using (11.33), we get

$$h_i(\lambda(t)) = 0 \qquad \Rightarrow \qquad \sum_{j=1}^k u_j(t) h_{ji}(\lambda(t)) = 0, \quad \forall\, t. \tag{11.35}$$

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

$$h_{ij}(\lambda(t)) \equiv 0. \tag{11.36}$$

Hence you cannot recover the control u from the linear system (11.35). We differentiate again equations (11.36) and we find

$$\sum_{l=1}^k u_l(t) h_{lij}(\lambda(t)) \equiv 0. \tag{11.37}$$

For every fixed t, these are $k(k-1)/2$ equations in k variables $u_1, \dots, u_k$. 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 11.19. For generic system 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.

11.3 Rank 2 distributions

Consider a rank 2 distribution, whose Hamiltonian equation for abnormal extremals is written as follows

$$\dot{\lambda}(t) = u_1(t)\vec{h}_1(\lambda(t)) + u_2(t)\vec{h}_2(\lambda(t)), \qquad h_1(\lambda(t)) = h_2(\lambda(t)) = 0. \tag{11.38}$$

Lemma 11.20. *Every abnormal extremal satisfy the Goh condition.*

Proof. Indeed differentiating the identities above we get, (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 u_1 and u_2 are arbitrary and at least one of them is nonzero, we have that $h_{12}(\lambda(t)) \equiv 0$, that is Goh condition. □

The fact that, in the rank 2 case, every abnormal extremal satisfies Goh conditions can be rewritten as

$$\lambda(t) \in (\mathcal{D}^2)^\perp,$$

and the system (11.37) reads

$$u_1 h_{112}(\lambda) = u_2 h_{221}(\lambda). \tag{11.39}$$

Assume now that $\lambda \in (\mathcal{D}^2)^\perp \setminus (\mathcal{D}^3)^\perp$, then at least one coefficient in (11.39) is nonzero and we can uniquely recover u up to a scalar. Under this assumption we find the control

$$u_1(t) = h_{221}(\lambda(t)), \qquad u_2(t) = h_{112}(\lambda(t)). \tag{11.40}$$

If we plug this control into the original equation we find that $\lambda(t)$ solve

$$\dot{\lambda} = h_{221}(\lambda)\vec{h}_1(\lambda) + h_{112}(\lambda)\vec{h}_2(\lambda).$$

In particular if we define the quadratic Hamitonian

$$H = h_{221}h_1 + h_{112}h_2 \tag{11.41}$$

using that any abnormal extremal belong to the subset $\{h_1(\lambda(t)) = h_2(\lambda(t)) = 0\}$, we have that $\lambda(t)$ satisfies

$$\dot{\lambda}(t) = \vec{H}(\lambda(t)) \tag{11.42}$$

Remark 11.21. Notice that, as soon as $n > 3$, the set $(\mathcal{D}_q^2)^\perp \setminus (\mathcal{D}_q^3)^\perp$ is nonempty for an open dense set of $q \in M$. Indeed assume that we have $\mathcal{D}_q^2 = \mathcal{D}_q^3$ for any q in a open neighborhood O_{q_0} of a point q_0 in M. Then it follows that

$$\mathcal{D}_{q_0}^2 = \mathcal{D}_{q_0}^3 = \mathcal{D}_{q_0}^4 = \dots$$

and the structure cannot be bracket generating, since $\dim \mathcal{D}_{q_0}^i < \dim M$ for every $i > 1$.

From now on we consider extremals associated with covectors $\lambda_0 \in (\mathcal{D}_q^2)^\perp \setminus (\mathcal{D}_q^3)^\perp$. This represents the less degenerate case. The case $n = 3$ will be treated separately.

Now we prove that the flow of the Hamiltonian H defined by (11.41) characterize exactly these extremals

Theorem 11.22. *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* (11.42) *with initial condition* $\lambda_0 \in (\mathcal{D}^2)^\perp \setminus (\mathcal{D}^3)^\perp$.

Proof. It remains to prove that a solution of the system

$$\dot\lambda(t) = \vec{H}(\lambda(t)), \qquad \lambda_0 \in (\mathcal{D}^2)^\perp \setminus (\mathcal{D}^3)^\perp, \tag{11.43}$$

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 (11.43) (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

$$\begin{cases}
\dot h_1 = a_1h_1 + a_2h_2 + a_{12}h_{12} \\
\dot h_2 = b_1h_1 + b_2h_2 + b_{12}h_{12} \\
\dot h_{12} = c_1h_1 + c_2h_2
\end{cases} \tag{11.44}$$

If we plug the solution $\lambda(t)$ into the equation of (11.43), 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.

$$h_1(\lambda_0) = h_2(\lambda_0) = h_{12}(\lambda_0) = 0. \tag{11.45}$$

Hence

$$h_1(\lambda(t)) = h_2(\lambda(t)) = h_{12}(\lambda(t)) = 0, \qquad \forall\, t.$$

$\square$

Definition 11.23. An abnormal extremal $\lambda(t)$ is called *nice abnormal* if, for every $t \in [0,1]$, it satisfies

$$\lambda(t) \in (\mathcal{D}^2)^\perp \setminus (\mathcal{D}^3)^\perp.$$

Lemma 11.24. *Let $\lambda(t)$ be a nice abnormal. Then $\lambda(t)$ or $-\lambda(t)$ satisfy*[2] *the generalized Legendre condition.*

Proof. It is sufficient to prove that the quadratic form

$$Q_t : v \mapsto \left\langle \lambda(t), [[f_{u(t)}, f_v], f_v] \right\rangle, \qquad v \in \mathbb{R}^2 \tag{11.46}$$

is semi-definite. We know that the bilinear form

$$B_t : (v,w) \mapsto \left\langle \lambda(t), [[f_{u(t)}, f_v], f_w] \right\rangle, \qquad v, w \in \mathbb{R}^2 \tag{11.47}$$

is symmetric. From (11.47) it is easy to see that $u(t) \in \operatorname{Ker} B_t$ 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 (11.46) 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. $\square$

Up to now we proved that every nice abnormal extremal automatically satisfies the necessary condition for optimality. Now we prove that actually they are strict local minimizers

Theorem 11.25. *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.*[3]

Remark 11.26. 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 sub-Riemannian structure.

It follows that, as soon as the metric is fixed, small pieces of nice abnormal are also global minimizers.

Before proving the Theorem we prove the following

Lemma 11.27. *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 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 0 is an isolated point on the level set.

[2] Recall that if $\lambda(t)$ is an abnormal extremal, $-\lambda(t)$ is also an abnormal extremal.

[3] which is equivalent to H^1-topology for trajectories.

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

$$E = E_1 \oplus E_2, \qquad E_2 = \operatorname{Ker} D_0\Phi, \tag{11.48}$$

$$\mathbb{R}^n = \mathbb{R}^{k_1} \oplus \mathbb{R}^{k_2}, \qquad \mathbb{R}^{k_1} = \operatorname{Im} D_0\Phi, \tag{11.49}$$

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, 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, by Implicit function theorem 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 $x_2 \in E_2 = \operatorname{Ker} D_0\Phi$. Notice that, by construction of the coordinate set, the function $x_2 \mapsto \Phi_2(0, x_2)$ coincide 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

$$\left.\frac{d}{ds}\right|_{s=0} a(\Phi_2(0, x_2)) = \left.\frac{d}{ds}\right|_{s=0} a(sv) = \langle \lambda, v \rangle \leq 0 \quad \Rightarrow \quad a(sv) \leq 0 \quad \text{for} \quad s \geq 0$$

which is a contradiction. $\square$

11.3.1 Optimality of nice abnormal

Let $\lambda(t)$ be an abnormal extremal and let $\gamma(t)$ be corresponding abnormal trajectory.

$$\dot{\gamma} = u_1 f_1(\gamma) + u_2 f_2(\gamma). \tag{11.50}$$

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 (11.50) 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

$$\dot{q} = (u_1 - 1)e_*^{-tf_1} f_1 + u_2 e_*^{-tf_1} f_2. \tag{11.51}$$

Since $e_*^{-tf_1} f_1 = f_1$, denoting $g_t := e_*^{-tf_1} f_2$ and defining the primitives

$$w(t) = \int_0^t (1 - u_1(\tau))d\tau, \qquad v(t) = \int_0^t u_2(\tau)d\tau, \tag{11.52}$$

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

$$\lambda_0 \mathrm{Hess}_0 G(u_1, \dot{v}) = \int_0^1 \langle \lambda_0, [\int_0^t -\dot{w}(\tau) f_1 + \dot{v}(\tau) g_\tau d\tau, -\dot{w}(t) f_1 + \dot{v}(t) g_t](q_0) \rangle dt. \tag{11.53}$$

Recall that

$$\begin{aligned} D_0 G(u_1, \dot{v}) &= \int_0^1 -\dot{w}(t) f_1(q_0) + \dot{v}(t) g_t(q_0) dt \\ &= -w(1) f_1(q_0) + \int_0^1 \dot{v}(t) g_t(q_0) dt \end{aligned}$$

and the condition $\lambda_0 \in \operatorname{Im} D_0 G^\perp$ is rewritten as

$$\langle \lambda_0, f_1(q_0) \rangle = \langle \lambda_0, g_t(q_0) \rangle = 0, \qquad \forall\, t. \tag{11.54}$$

Notice that since equality (11.54) is valid for all t then we have that

$$\langle \lambda_0, \dot{g}_t(q_0) \rangle = \langle \lambda_0, [f_1, g_t](q_0) \rangle = 0, \tag{11.55}$$

Then we can rewrite our quadratic form only as a function of $\dot{v}$, since all terms containing $\dot{w}$ disappear

$$\lambda_0 \mathrm{Hess}_0 G(\dot{v}) = \int_0^1 \langle \lambda_0, [\int_0^t \dot{v}(\tau) g_\tau d\tau, \dot{v}(t) g_t](q_0) \rangle dt \tag{11.56}$$

with the extra condition

$$\int_0^1 \dot{v}(t) g_t(q_0) dt = w(1) f_1(q_0). \tag{11.57}$$

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

$$\int_0^t \dot{v}(\tau) g_\tau d\tau = v(t) g_t - \int_0^t v(\tau) \dot{g}_\tau d\tau \tag{11.58}$$

we have

$$\lambda_0 \mathrm{Hess}_0 G(\dot{v}) = \int_0^1 \langle \lambda_0, [v(t)g_t, \dot{v}(t)g_t](q_0)\rangle dt \\ - \int_0^1 \langle \lambda_0, [\int_0^t v(\tau)\dot{g}_\tau d\tau, \dot{v}(t)g_t](q_0)\rangle dt$$

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

$$\lambda \mathrm{Hess}_0 G(\dot{v}) = \int_0^1 \langle \lambda_0, [\dot{g}_t, g_t](q_0)\rangle v(t)^2 dt \\ + \int_0^1 \langle \lambda_0, [\int_0^t v(\tau)\dot{g}_\tau, v(t)\dot{g}_t - v(1)g_1](q_0)\rangle dt \tag{11.59}$$

which can also be rewritten as follows

$$\lambda \mathrm{Hess}_0 G(\dot{v}) = \int_0^1 \langle \lambda_0, [\dot{g}_t, g_t](q_0)\rangle v(t)^2 \, dt \\ + \int_0^1 \langle \lambda_0, [\int_1^t v(\tau)\dot{g}_\tau \, d\tau + v(1)g_1, v(t)\dot{g}_t](q_0) \, dt. \tag{11.60}$$

Moreover, again integrating by parts the extra condition (11.57), we find

$$\int_0^1 v(t)\dot{g}_t(q_0)dt = -w(1)f_1(q_0) + v(1)g_1(q_0) \tag{11.61}$$

Remark 11.28. Notice that we cannot plug in the expression (11.61) directly into the formula since this equality is valid only at the point q_0, while in (11.59) 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 (11.61) it follows that for every pair (w, v) in the kernel the following estimates are valid

$$|w(1)| \leq C\|v\|_{L^2}, \qquad |v(1)| \leq C\|v\|_{L^2}. \tag{11.62}$$

Theorem 11.29. *Let $\gamma : [0,1] \to M$ be an abnormal trajectory and assume that the quadratic form* (11.59) *satisfies*

$$\lambda_0 \mathrm{Hess}_0 G(\dot{v}) \geq \alpha \|v\|_{L^2}^2. \tag{11.63}$$

Then the curve is locally minimizer in the L^2 topology of controls.

Remark 11.30. Notice that the estimate (11.63) depends only on v, while the map G is a smooth map of $\dot{v}$ and $\dot{w}$. Hence Lemma 11.27 does not apply.

Moreover, the statement of Lemma 11.27 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 11.29. 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 (11.52), 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 $a \in \mathcal{C}^\infty(M)$ such that $d_q a = \lambda \in \operatorname{Im} D_0 F^\perp$, we have

$$a \circ F\big|_D(\dot{w}, \dot{v}) = \lambda \mathrm{Hess}_0 F(\dot{w}, \dot{v}) + R(w, v), \qquad \text{where} \quad \frac{R(w,v)}{\|v\|^2} \underset{\|(\dot{w},\dot{v})\| \to 0}{\longrightarrow} 0 \tag{11.64}$$

in the domain

$$D = \{(\dot{w}, \dot{v}) \in \operatorname{Ker} D_0 F, \dot{w} \geq 0\}$$

Indeed if we prove (11.64) 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 (6.21) 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 (11.64), 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

$$(1 - \dot{w}(t)) e_*^{-v(t)f_2} f_1 = f_1 + X^0(v(t)) + \dot{w}(t) X^1(v(t)) \tag{11.65}$$

and using the variation formula (6.22), 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 $Y_t^0(0) = Y_t^1(0) = 0$,

(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 $o(\|v\|^2)$.[4]

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 $(\dot{w}, \dot{v}) \in \operatorname{Ker} D_0 F$, which gives the estimates (11.62), we compute

$$\begin{aligned} \left| \int_0^1 \dot{w}(t) v^2(t) dt \right| &\leq \int_0^1 |\dot{w}(t)| v^2(t) dt \\ &= \int_0^1 \dot{w}(t) v^2(t) dt \\ &= w(1) v^2(1) - \int_0^1 w(t) v(t) \dot{v}(t) dt \\ &\leq \|v\|^3 + \varepsilon \|v\|^2, \end{aligned}$$

[4] where $o(\|v\|^2)$ have the same meaning as in (11.64).

where we estimate for the second term follows from

$$\begin{aligned}\left|\int_0^1 w(t)v(t)\dot v(t)dt\right| &\le \max w(t)\left|\int_0^1 v(t)\dot v(t)dt\right|\\ &\le w(1)\|v\|\|\dot v\|\\ &\le C\|\dot v\|\|v\|^2\end{aligned}$$

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

$$\begin{aligned}\left|\int_0^1 \dot w(t)\int_0^t \dot w(\tau)v(\tau)d\tau dt\right| &\le 2|w(1)|\int_0^1 v(t)\dot w(t)dt\\ &\le C\|\dot w\|\|v\|^2\end{aligned}$$

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 11.25 . 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[5] that (see also Lemma 11.24)

$$\langle\lambda_0, [g_t, \dot g_t](q_0)\rangle \ge C > 0$$

then the norm

$$\|v\|_g = \left(\int_0^s \langle\lambda_0, [g_t, \dot g_t](q_0)\rangle v(t)^2 dt\right)^{1/2} \tag{11.66}$$

[5]it is semidefinite and we already know that f_1 is in the kernel

is equivalent to the standard L^2-norm. Hence the Hessian can be rewritten as

$$\lambda \mathrm{Hess}_0 G_s(v) = \|v\|_g + \langle Tv, v\rangle \tag{11.67}$$

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 $s > 0$. $\square$

11.4 Conjugate points

In this section, we give an effective way to check the inequality (11.63) that implies local minimality of the nice abnormal geodesic according to Theorem 11.29.

We set: $Q_1(v) \doteq \lambda \mathrm{H}ess_0 G(\dot{v})$. 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

$$\left\{ v \in L_2[0,1] : \int_0^1 v(t)\dot{g}_t(q_0)\, dt \in span\{f_1(q_0), g_1(q_0)\} \right\}.$$

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 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 a family of quadratic forms for $0 < s \leq 1$

$$Q_s(v) := \int_0^s l(t)v(t)^2\, dt + \int_0^s \langle \lambda_0, [X_t, \dot{X}_t](q_0)\rangle\, dt,$$

$$\dot{X}_t = v(t)\dot{g}_t, \quad X_s \wedge g_s = 0,\ X_0(q_0) \wedge f_1(q_0) = 0. \tag{1}$$

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 7.29. We leave details to the reader.

Proposition 11.31. *The form Q_1 is positive definite if and only if* $\ker Q_s = 0,\ \forall s \in (0,1]$.

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 $f \in \mathrm{V}ecM$ 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,

$$l(t) = \sigma(\dot{\zeta}_t, \zeta_t), \quad 0 \le t \le 1. \tag{11.68}$$

Let $E = span\{\xi_1, \zeta_t, 0 \le t \le 1\}$. 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 11.32. $\dim \Sigma \le 2\,(\dim span\{f_1(q_0), g_t(q_0), 0 \le t \le 1\} - 1)$.

Proof. Dimension of Σ equals the double 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, $\pi_*(E) = span\{f_1(q_0), g_t(q_0), 0 \le t \le 1\}$ and $E \cap \ker \pi_*$ is an isotropic subspace of $\sigma|_E$. □

We denote by $\underline{\zeta}_t \in \Sigma$ the projection of 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 (11.68) implies that $\underline{\zeta}_t \ne 0$. $\forall t \in [0,1]$. Final expression of Q_s is as follows:

$$Q_s(v) = \int_0^s l(t)v(t)^2\,dt + \int_0^s \sigma(x(t),\dot{x}(t))\,dt,$$

$$\dot{x} = v(t)\dot{\underline{\zeta}}_t, \quad x(s) \wedge \underline{\zeta}_s = 0,\ x(0) \in \Pi. \tag{4}$$

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

$$l(t)v(t) + \sigma(x(t), \dot{\underline{\zeta}}_t) = \sigma(\nu, \dot{\underline{\zeta}}_t), \quad 0 \le t \le s.$$

We set $y(t) = x(t) - \nu$ and obtain the following:

Theorem 11.33. *A time moment $s \in (0,1]$ is conjugate to 0 if and only if there exists a nontrivial solution of the equation*

$$l(t)\dot{y} = \sigma(\dot{\underline{\zeta}}_t, y)\dot{\underline{\zeta}}_t \tag{11.69}$$

that satisfy the following boundary conditions:

$$\exists\, \nu \in \Pi^\angle \cap \underline{\zeta}_s^\angle \quad \textit{such that} \quad (y(s) + \nu) \wedge \underline{\zeta}_s = 0,\ (y(0) + \nu) \in \Pi. \tag{11.70}$$

Remark 11.34. Identity (11.68) implies that $y(t) = \underline{\zeta}_t$ $0 \le t \le 1$, is a solution to the equation (11.69). However this solution may violate the boundary conditions.

Let us consider a special case: $\dim span\{f_1(q_0), g_t(q_0), 0\} = 2$; 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 (11.70) must be 0 and $y(s) = c\underline{\zeta}_s$, where c is a nonzero constant. Hence $y(t) = c\underline{\zeta}_t$ for $0 \le t \le 1$ and $y(0) = c\underline{\zeta}_0 \notin \Pi$. We obtain:

Corollary 11.35. *If* $\dim span\{f_1(q_0), g_t(q_0), 0 \le t \le 1\} = 2$*, then the segment* $[0,1]$ *does not contain conjugate time moments and assumption of Theorem 11.29 is satisfied.*

We can apply this corollary to the isoperimetric problem studied in Section 4.5.2. Abnormal geodesics correspond to connected components of the zero locus of the function b (see notations in Sec. 4.5.2). 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 11.29 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!

Now consider another important special case that is typical if dimension of the ambient manifold is greater than 3. Namely, assume that, for some $k \ge 2$, the vector fields

$$f_1,\ f_2,\ (\mathrm{ad} f_1)f_2,\ \ldots,\ (\mathrm{ad} f_1)^{k-1} f_2 \tag{11.71}$$

are linearly independent in any point of a neigborhood of our nice abnormal geodesic γ, while $(\mathrm{ad} f_1)^k f_2$ is a linear combination of the vector fields (11.71) 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:

$$\zeta_t^{(k)} = \sum_{i=0}^{k-1} a_i(\gamma(t))\zeta^{(i)} + \alpha(\gamma(t))\xi_1. \quad 0 \le t \le 1. \tag{11.72}$$

Identity (11.72) implies that $\dim E = k$ and $\Pi = 0$. The boundary conditions (11.70) take the form:

$$y(0) \in \underline{\zeta}_s^\angle, \quad (y(s) - y(0)) \wedge \underline{\zeta}_s = 0. \tag{11.73}$$

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 (11.71) 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 (11.69). Hence boundary conditions (11.73) are equivalent to the condition

$$\underline{\zeta}_s \wedge \underline{\zeta}_0 = 0. \tag{11.74}$$

It is easy to re-write relation (11.74) in the intrinsic way without special notations we used to simplify calculations. We have:
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 11.36. *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, 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 $\mathcal{D}$. 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.* □

11.5 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.

Notice that the Theorem holds for general structure and not only for rank 2 distributions.

Theorem 11.37. *Assume that the sub-Riemannian structure is extendable to a Riemannian structure G on M. Let $\gamma(t)$ be a (strict) local minimizer in the L^2 topology for the controls, that has no self-intersection. If $\gamma \in \mathcal{C}^1$, then it is a (strict) local minimizer in the $\mathcal{C}^0$ topology for the trajectories.*

Proof. Since γ has no self intersections, as before we can assume to choose coordinates $x = (x_1, y)$ in the cylinder

$$M = I_\varepsilon \times B_{n-1} = \{(x, y) \in \mathbb{R}^n, x \in]-\varepsilon, 1+\varepsilon[, y \in \mathbb{R}^{n-1}, |y| < \varepsilon\},$$

where our curve γ is rectified, $\gamma(t) = e^{tf_1}(0)$. Moreover we can set $\ell(\gamma) = 1$.

In these coordinates $\gamma(t) = (\xi(t), \eta(t)) = (t, 0)$. Then we need the following

Lemma 11.38. *There exists ε and a set of coordinates such that $G(x, 0) = \mathrm{Id}$.*

Proof of the Lemma. Our normalization of the curve γ implies that

$$g = \begin{pmatrix} g_{11} & G_{12} \\ G_{21} & G_{22} \end{pmatrix}, \qquad \text{with} \quad g_{11}(x, 0) = 1$$

Now consider the orthogonal complement with respect to G, to our line at every point $(x, 0)$, that can be written

$$\{(\langle a_x, y\rangle, y), y \in \mathbb{R}^{n-1}\}$$

for some vector $a_x \in \mathbb{R}^{n-1}$. The change of coordinates

$$(x, y) \mapsto (x - \langle a_x, y\rangle, y)$$

is regular (choosing ε small enough) and normalize our metric in such a way that

$$g(x,0) = \begin{pmatrix} 1 & 0 \\ 0 & G \end{pmatrix}, \qquad \text{with} \quad g_{11}(x,0) = 1$$

With a linear change of cooordinates in the y space

$$(x,y) \mapsto (x, G(x,0)^{1/2} y)$$

we can then normalize all the matrix in such a way that $G(x,0) = \mathrm{Id}$ $\square$

The action of a curve γ is defined as follows

$$J(\gamma) = \int_0^1 \left\langle G_{\gamma(t)} \dot\gamma(t), \dot\gamma(t) \right\rangle dt$$

Now notice that, in the Euclidean case $G = \mathrm{Id}$, we have

$$J_{eu}(\gamma) = \int_0^1 \dot\xi^2(t) + \dot\eta^2(t) dt \tag{11.75}$$

and by definition of H^1 norm we get the equality

$$\begin{aligned} \|\gamma - \gamma_0\|_{H^1}^2 &= \int_0^1 |\dot\xi(t) - 1|^2 + |\dot\eta(t)|^2 dt \\ &= J_{eu}(\gamma) - 1 \\ &= J_{eu}(\gamma) - J_{eu}(\gamma_0) \end{aligned}$$

Thus, our assumption can be rewritten as follows: there exists $\varepsilon > 0$ such that γ is admissible and

$$J_{eu}(\gamma) \le 1 + \varepsilon \qquad \Longrightarrow \qquad J(\gamma) \ge 1$$

Now take $\delta > 0$ and a curve γ contained in our domain such that the curve $\gamma_\delta := F_\delta(\gamma)$, image of γ under the dilation

$$F_\delta : (x,y) \mapsto (x, \delta y)$$

is admissible. There are two possibilities:

(*i*) $J_{eu}(\gamma_\delta) \le 1 + \varepsilon$. Then by our assumption $J(\gamma_\delta) > 1$

(*ii*) $J_{eu}(\gamma_\delta) > 1 + \varepsilon$. Then since $G(x,0) = \mathrm{Id}$ and $|y| < \delta$, by smoothness of G we have that in this neighborhood

$$G(v) = |v| + O(\delta) \quad \Rightarrow \quad J(\gamma_\delta) = J_{eu}(\gamma) + O(\delta)$$

from which it follows that $J(\gamma_\delta) \ge 1 + \varepsilon + O(\delta) > 1$, choosing δ in an appropriate way.

Hence every curve γ that is contained in the δ-strip is longer that γ_0.

$\square$

Remark 11.39. Notice that this Theorem implies in particular the statement of Theorem 4.55, since normal extremals are always smooth. On the other hand, the argument of Theorem 4.55 can be adapted for any coercive functional (see [4]) while this proof use explicitly estimates that holds only for our specific cost (distance).

Remark 11.40. Notice that *nice* abnormals are smooth. Hence we can apply this result and every nice abnormal is also a $\mathcal{C}^0$ local minimizer.

Chapter 12

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.

12.1 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 12.1. 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 12.2. *Let $W \in G_k(V)$. We have a canonical isomorphism*

$$T_W G_k(V) \simeq \mathrm{Hom}(W, V/W).$$

Proof. Consider a smooth curve on $G_k(V)$ which starts from W, i.e. a smooth family of k-dimensional subspaces defined by a moving frame

$$W(t) = \mathrm{span}\{e_1(t), \dots, e_k(t)\}, \qquad W(0) = W.$$

We want to associate in a canonical way with the tangent vector $\dot{W}(0)$ a linear operator from W to the quotient V/W. Fix $w \in W$ and consider any smooth extension $w(t) \in W(t)$, with $w(0) = w$. Then define the map

$$W \to V/W, \qquad w \mapsto \dot{w}(0) \pmod{W}. \tag{12.1}$$

We are left to prove that the map (12.1) 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

$$\dot{w}_1(t) = \dot{w}(t) + \sum_{i=1}^{k} \dot{\alpha}_i(t) e_i(t) + \sum_{i=1}^{k} \alpha_i(t) \dot{e}_i(t), \tag{12.2}$$

and evaluating (12.2) 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 $\dot{w}_1(0) = \dot{w}(0) \pmod{W}$, hence the map (12.1) 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. $\square$

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 12.3. 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

$$\dim \Pi + \dim \Pi^{\angle} = \dim \Sigma. \tag{12.3}$$

where as usual we denote the symplectic orthogonal by $\Pi^{\angle} = \{x \in \Sigma \,|\, \sigma(x, y) = 0, \ \forall\, y \in \Pi\}$.

Exercise 12.4. 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 12.5. 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 12.6. Any symplectic vector space admits Lagrangian subspaces. Indeed fix any non-zero element $e_1 := e \neq 0$ in Σ. Choose iteratively

$$e_i \in \mathrm{span}\{e_1, \dots, e_{i-1}\}^{\angle} \setminus \mathrm{span}\{e_1, \dots, e_{i-1}\}, \quad i = 2, \dots, n. \tag{12.4}$$

Then $\Pi := \mathrm{span}\{e_1, \dots, e_n\}$ is a Lagrangian subspace by construction. Notice that the choice (12.4) is possible by (12.3)

Lemma 12.7. *Let $\Pi = \mathrm{span}\{e_1, \dots, e_n\}$ be a Lagrangian subspace of Σ. Then there exists vectors $f_1, \dots, f_n \in \Sigma$ such that*

(*i*) $\Sigma = \Pi \oplus \Delta, \qquad \Delta := \mathrm{span}\{f_1, \dots, f_n\},$

(*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 $\mathrm{span}\{e_n, f_n\}$ is nondegerate, hence by (a) of Exercise 12.4

$$\mathrm{span}\{e_n, f_n\} \cap \mathrm{span}\{e_n, f_n\}^{\angle} = 0, \tag{12.5}$$

And we can apply induction on the $2(n-1)$ subspace $\Sigma' := \mathrm{span}\{e_n, f_n\}^{\angle}$. Notice that (12.5) implies that σ is non degenerate also on Σ'. $\square$

Remark 12.8. In particular the complementary subspace $\Delta = \mathrm{span}\{f_1, \dots, f_n\}$ defined in Lemma 12.7 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 12.7) we find that in coordinates, if $x_1 = (\zeta_1, z_1), x_2 = (\zeta_2, z_2)$ we get

$$\sigma(x_1, x_2) = \zeta_1 z_2 - \zeta_2 z_1, \tag{12.6}$$

where we denote with ζz the standard rows by columns product.

Lemma 12.7 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 12.9. 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

$$\Lambda_1 = \{(p, 0)\}, \qquad \Lambda_2 = \{((p_1, \dots, p_k, 0, \dots, 0), (0, \dots, 0, q_{k+1}, \dots, q_n))\}.$$

12.1.1 The Lagrange Grassmannian

Definition 12.10. The *Lagrange Grassmannian* $L(\Sigma)$ of a symplectic vector space Σ is the set of its n-dimensional Lagrangian subspaces.

Proposition 12.11. *$L(\Sigma)$ is a compact submanifold of the Grassmannian $G_n(\Sigma)$ of n-dimensional subspaces. Moreover*

$$\dim L(\Sigma) = \frac{n(n+1)}{2}. \tag{12.7}$$

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 12.7 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 (12.6) 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 (12.7). 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 12.12. *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

$$\underline{\dot{\Lambda}} : x \mapsto \sigma(x, \dot{x}), \tag{12.8}$$

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

$$\begin{aligned} \sigma(x, \dot{x}) &= z^T(\dot{S}_\Lambda z + S_\Lambda \dot{z}) - \dot{z}^T S_\Lambda z \\ &= z^T \dot{S}_\Lambda z + z^T S_\Lambda \dot{z} - \dot{z}^T S_\Lambda z \\ &= z^T \dot{S}_\Lambda z. \end{aligned} \tag{12.9}$$

□

Exercise 12.13. 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 12.14. 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 12.2. 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

$$\Sigma/\Lambda \simeq \Lambda^*, \qquad \langle [z]_\Lambda, x\rangle = \sigma(z, x). \tag{12.10}$$

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 12.15. 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 $h \in \mathcal{C}^\infty(\Sigma)$. (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 12.16. *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. □

12.2 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 12.17. 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 12.18. 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 12.12 that the quadratic form $\underline{\dot{\Lambda}}(t)$ is represented by the matrix $\dot{S}_\Lambda(t)$ (see also (12.9)). 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 12.7) 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 12.19. *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*

$$S_B(t) = (S(t)^{-1} - B)^{-1}. \tag{12.11}$$

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

$$\begin{cases} p' = p - Bq \\ q' = q \end{cases} \tag{12.12}$$

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 (12.12) 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 12.20. *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 (12.11) (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 (12.13). 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 12.21. 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 12.22. 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

$$S^\circ(t) = B(t)^{-1} + S(t), \quad \text{where} \quad B(t) := -\frac{1}{2}\dot{S}(t)^{-1}\ddot{S}(t)\dot{S}(t)^{-1}, \tag{12.13}$$

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 12.23. 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

$$\Delta_2 - \Delta_1 \mapsto A \in \mathrm{Hom}(\Sigma/\Lambda, \Lambda), \qquad A([z]_\Lambda) = z_2 - z_1 \in \Lambda, \tag{12.14}$$

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 (12.10)) 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 12.24. Prove that the operator A defined by (12.14), in the case when Λ is Lagrangian, is a self-adjoint operator.

Remark 12.25. 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 12.23) 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 (12.11) for the change of coordinates can be rewritten as follows

$$S_B(t)^{-1} = S^{-1}(t) - B, \tag{12.15}$$

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.

12.3 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 12.1, 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:

$$\underline{\dot{\Lambda}}(t) : \Lambda(t) \to \Lambda^\circ(t), \qquad \underline{\dot{\Lambda}}^\circ(t) : \Lambda^\circ(t) \to \Lambda(t). \tag{12.16}$$

Definition 12.26. 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 12.27. 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 12.28. Let Λ_1, Λ_2 be two transversal Lagrangian subspaces of Σ. We denote

$$\pi_{\Lambda_1\Lambda_2} : \Sigma \to \Lambda_2, \tag{12.17}$$

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 12.29. 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 $\pi_{\Lambda_1\Lambda_2}$:

$$\pi_{\Lambda_1\Lambda_2} = \begin{pmatrix} S_{12}^{-1} S_1 & -S_{12}^{-1} \\ S_2 S_{12}^{-1} S_1 & -S_2 S_{12}^{-1} \end{pmatrix}, \qquad S_{12} := S_1 - S_2. \tag{12.18}$$

From the very definition of the derivative of our curve we can get the following geometric characterization of the curvature of a curve.

Proposition 12.30. *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 (12.30) 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 12.31. *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)*

$$R_\Lambda = ((2\dot{S})^{-1}\ddot{S})^{\cdot} - ((2\dot{S})^{-1}\ddot{S})^2 \tag{12.19}$$

$$= \frac{1}{2}\dot{S}^{-1}\dddot{S} - \frac{3}{4}(\dot{S}^{-1}\ddot{S})^2. \tag{12.20}$$

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 (12.18)) 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

$$R = (S^\circ - S)^{-1}\dot{S}^\circ(S^\circ - S)^{-1}\dot{S}. \tag{12.21}$$

We prove formula (12.20) 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 (12.13) 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 (12.20). $\square$

Remark 12.32. 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 (12.20) 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 12.33. Let $f : \mathbb{R} \to \mathbb{R}$ be a smooth function. The *Schwartzian derivative* of f is defined as

$$\mathcal{S}f := \left(\frac{f''}{2f'}\right)' - \left(\frac{f''}{2f'}\right)^2 \tag{12.22}$$

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 12.34. The previous proposition says that the curvature R is the matrix version of the Schwartzian derivative of the matrix S (cfr. (12.19) and (12.22)).

Example 12.35. 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 12.36. *Let $\varphi : \mathbb{R} \to \mathbb{R}$ a diffeomorphism and define the curve $\Lambda_\varphi(t) := \Lambda(\varphi(t))$. Then*

$$R_{\Lambda_\varphi}(t) = \dot{\varphi}^2(t) R_\Lambda(\varphi(t)) + R_\varphi(t)\mathrm{Id}. \tag{12.23}$$

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 $L(\mathbb{R}^2)$. $\square$

Exercise 12.37. (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 (12.16) we can treat each tangent vector as a linear operator

$$\xi_0 : \Lambda_0 \to \Lambda_1, \qquad \xi_1 : \Lambda_1 \to \Lambda_0, \tag{12.24}$$

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 $t \to 0$

$$[\underline{\dot{\Lambda}}(t), \underline{\dot{\Lambda}}(0)] \simeq \frac{1}{t^2} Id + \frac{1}{3} R_\Lambda(0) + O(t). \tag{12.25}$$

12.4 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 12.38. *There is a natural identification of $L(\Gamma^\angle/\Gamma)$ as a subspace of $L(\Sigma)$:*

$$L(\Gamma^\angle/\Gamma) \simeq \{\Lambda \in L(\Sigma), \Gamma \subset \Lambda\} \subset L(\Sigma). \tag{12.26}$$

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 (12.26).

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 12.5. (The identity $(\Lambda \cap \Gamma^\angle) + \Gamma = (\Lambda + \Gamma) \cap \Gamma^\angle$ is also a consequence of the same exercise.) ☐

Remark 12.39. 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 12.40. 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 $\Gamma = \mathrm{span}\{e_1\}$ be a one dimensional isotropic subspace and define

$$\Lambda_\varepsilon = \mathrm{span}\{e_1 + \varepsilon f_2, e_2 + \varepsilon f_1\}, \qquad \forall\, \varepsilon > 0.$$

It is easy to see that Λ_ε is Lagrangian for every ε and that

$$\begin{aligned}\Lambda_\varepsilon^\Gamma &= \mathrm{span}\{e_1, f_2\}, \qquad \forall\, \varepsilon > 0, \\ \Lambda_0^\Gamma &= \mathrm{span}\{e_1, e_2\}.\end{aligned} \tag{12.27}$$

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 (12.27) holds. The case $\varepsilon = 0$ is trivial.

12.5 Ample curves

In this section we introduce ample curves.

Definition 12.41. 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

$$\Sigma = \mathrm{span}\{\lambda^{(i)}(t_0) |\ \lambda(t) \in \Lambda(t), \lambda(t) \text{ smooth}, 0 \leq i \leq N\}. \tag{12.28}$$

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

$$\Sigma = \{(p, q) |\ p, q \in \mathbb{R}^n\}, \qquad \Lambda(t) = \{(p, S(t)p) |\ p \in \mathbb{R}^n\}.$$

Exercise 12.42. 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:

$$\mathrm{rank}\{\dot S(0), \ddot S(0), \dots, S^{(N)}(0)\} = n. \tag{12.29}$$

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 12.43. *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 12.44. 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 12.45. *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

$$\widetilde{\Lambda}(0) \cap \widetilde{\Lambda}(t) = \{0\}, \qquad \text{in} \quad L(\Upsilon^{\angle}/\Upsilon) \tag{12.30}$$

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. $\square$

12.6 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 12.46. *Let $\Lambda(t)$ be a smooth ample monotone curve and set $\Gamma := \operatorname{Ker} \dot{\underline{\Lambda}}(0)$. Then the reduced curve $t \mapsto \Lambda^{\Gamma}(t)$ is a smooth regular curve. In particular $\dot{\underline{\Lambda}}^{\Gamma}(0) > 0$.*

Before proving Theorem 12.46, let us discuss two useful lemmas.

Lemma 12.47. *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. $\square$

Lemma 12.48. *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. $\square$

Proof of Theorem 12.46. 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

$$S(t) = \begin{pmatrix} S_{11}(t) & S_{12}(t) \\ S_{12}^*(t) & S_{22}(t) \end{pmatrix}, \qquad \text{with} \qquad \dot{S}(0) = \begin{pmatrix} \mathbb{I}_k & 0 \\ 0 & 0 \end{pmatrix}. \tag{12.31}$$

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

$$S_{11}(t) > 0, \qquad S_{22}(t) > 0, \qquad \forall\, t > 0. \tag{12.32}$$

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

$$\Lambda(t) \cap \Gamma^\angle = \{(p, S(t)p), S(t)p \in \mathbb{R}^k\} = \left\{ \left(S^{-1}(t) \begin{pmatrix} q_1 \\ 0 \end{pmatrix}, \begin{pmatrix} q_1 \\ 0 \end{pmatrix} \right), q_1 \in \mathbb{R}^k \right\} \tag{12.33}$$

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

$$S^\Gamma = S_{11} - S_{12}S_{22}^{-1}S_{12}^*. \tag{12.34}$$

(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 $0 \le S^\Gamma(t) \le S_{11}(t)$.

Indeed $S(t)$ symmetric and positive one has that its inverse $S(t)^{-1}$ is symmetric and positive also. This implies that $S^\Gamma(t)^{-1} = (S(t)^{-1})_{11} > 0$ and so is $S^\Gamma(t)$. This proves the left inequality of the Claim I.

Moreover using (12.34) and the fact that S_{22} is positive definite (and so S_{22}^{-1}) one gets

$$\langle (S_{11} - S^\Gamma)p_1, p_1 \rangle = \langle S_{12}S_{22}^{-1}S_{12}^* p_1, p_1 \rangle = \langle S_{22}^{-1}(S_{12}^* p_1), (S_{12}^* p_1) \rangle \ge 0.$$

Since $S(t) \to 0$ for $t \to 0$, clearly $S_{11}(t) \to 0$ when $t \to 0$, that proves that $S^\Gamma(t) \to 0$ also.

(iii). We are reduced to show that the derivative of $t \mapsto S^\Gamma(t)$ at 0 is non degenerate matrix, which is equivalent to show that $t \mapsto S^\Gamma(t)^{-1}$ has a simple pole at $t = 0$.

We need the following lemma, whose proof is postponed at the end of the proof of Theorem 12.46.

Lemma 12.49. *Let $A(t)$ be a smooth family of symmetric nonnegative $n \times n$ matrices. If the condition* $\operatorname{rank}(A, \dot{A}, \dots, A^{(N)})|_{t=0} = n$ *is satisfied for some N, then there exists $\varepsilon_0 > 0$ such that $\varepsilon t A(0) < \int_0^t A(\tau)d\tau$ for all $\varepsilon < \varepsilon_0$ and $t > 0$ small enough.*

Applying the Lemma to the family $A(t) = \dot{S}(t)$ one obtains (see also (12.31))

$$\langle S(t)p, p \rangle > \varepsilon t |p_1|^2$$

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

$$\langle S^\Gamma(t)p_1, p_1 \rangle = \langle S(t)p, p \rangle \ge \varepsilon t |p_1|^2,$$

This identity can be rewritten as $S^\Gamma(t) > \varepsilon t\, \mathbb{I}_k > 0$ and implies by Lemma 12.48

$$0 < S^\Gamma(t)^{-1} < \frac{1}{\varepsilon t}\mathbb{I}_k$$

which completes the proof. $\square$

Proof of Lemma 12.49. 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 $A(t) \ge 0$ 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

$$E_i = \operatorname{span}\{v_j^{(l)}, 1 \le j \le n, 0 \le l \le i\}.$$

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 12.47, 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

$$c := \det\left(\frac{c_{ij}}{m_i+m_j+1}\right), \qquad \widehat{N} := n + 2\sum_{i=1}^{N} m_i$$

$\square$

12.7 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 12.50. 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 12.43, we have the following immediate corollary.

Corollary 12.51. *Conjugate points on a monotone and ample curve in $L(\Sigma)$ are isolated.*

The following two results describe general properties of conjugate points

Theorem 12.52. *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 $S_{\Lambda}(\infty) = \lim_{t\to+\infty} S_{\Lambda}(t)$. $\square$

Theorem 12.53. *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 12.7).

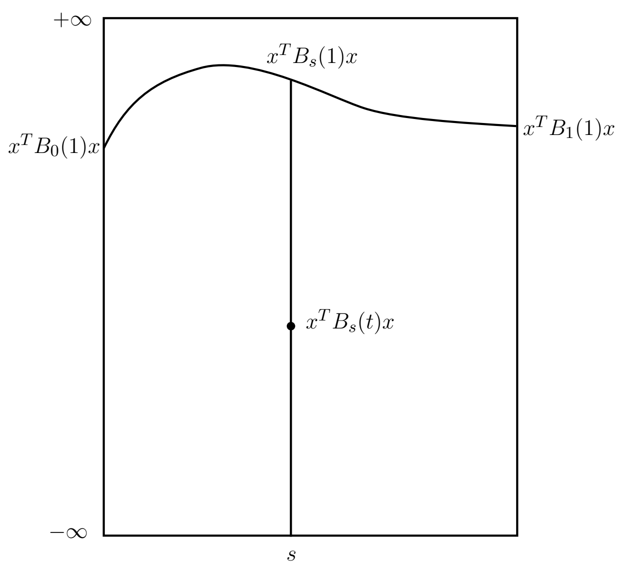


Figure 12.1: Proof of Theorem 12.53

∎

12.8 Comparison theorems for regular curves

In this last section we prove two comparison theorems for regular monotone curves in the Lagrange Grassmannian.

Corollary 12.54. *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 12.52 ∎

Theorem 12.55. *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 12.56. Notice that the estimates of Theorem 12.55 are sharp, as it is immediately seen by considering the example of a 1-dimensional curve of constant velocity (see Example 12.35).

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 12.36

$$\begin{aligned} R_{\Lambda_\varphi}(t) &= \dot{\varphi}^2 R_\Lambda(\varphi(t)) + R_\varphi(t)\mathrm{Id} \\ &= \frac{1}{(1 + kt^2)^2}(R_\Lambda(\varphi(t)) - k\mathrm{Id}) \leq 0. \end{aligned}$$

By Corollary 12.54 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 $b(t) := \mathrm{trace}\, B(t)$. Computing the trace in both sides of equality

$$\dot{B}(t) = B^2(t) + R_\Lambda(t),$$

we get

$$\dot{b}(t) = \mathrm{trace}(B^2(t)) + \mathrm{trace}\, R_\Lambda(t). \tag{12.35}$$

Lemma 12.57. *For every $n \times n$ symmetric matrix S the following inequality holds true*

$$\mathrm{trace}(S^2) \geq \frac{1}{n}(\mathrm{trace}\, S)^2. \tag{12.36}$$

Proof. For every symmetric matrix S there exists a matrix M such that $MSM = D$ is diagonal. Since $\mathrm{trace}(MAM^{-1}) = \mathrm{trace}(A)$ for every matrix A, it is enough to prove the inequality (12.36) for a diagonal matrix $D = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$. In this case (12.36) reduces to the Cauchy-Schwartz inequality

$$\sum_{i=1}^{n} \lambda_i^2 \geq \frac{1}{n} \left(\sum_{i=1}^{n} \lambda_i \right)^2 .$$

$\square$

Applying Lemma 12.57 to (12.35) and using the assumption (ii) one gets

$$\dot{b}(t) \geq \frac{1}{n} b^2(t) + nk, \tag{12.37}$$

By standard results in ODE theory we have $b(t) \geq \varphi(t)$, where $\varphi(t)$ is the solution of the differential equation

$$\dot{\varphi}(t) = \frac{1}{n} \varphi^2(t) + nk \tag{12.38}$$

The solution for (12.38), 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 13

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.

13.1 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 $T_{\lambda(t)}(T^*M)$.

13.1.1 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 13.1. Let $\lambda \in T^*M$. The *Jacobi curve* at the point λ is defined as follows

$$J_\lambda(t) := e_*^{-t\vec{H}}\mathcal{V}_{\lambda(t)}, \tag{13.1}$$

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 12, 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 13.2. *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 (12.8)

$$\dot{J}_\lambda(0)\xi = \sigma(\xi, \dot{\xi}) = \sigma\left(\xi, \frac{d}{dt}\bigg|_{t=0} e_*^{-t\vec{H}}\xi\right) = \sigma(\xi, [\vec{H}, \xi]). \tag{13.2}$$

To compute the last quantity we use the following elementary, although very useful, property of the symplectic form σ.

Lemma 13.3. *Let $\xi \in \mathcal{V}_\lambda$ a vertical vector. Then, for any $\eta \in T_\lambda(T^*M)$*

$$\sigma(\xi, \eta) = \langle \xi, \pi_*\eta \rangle, \tag{13.3}$$

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. $\square$

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 13.3 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

$$\underline{\dot{J}}_\lambda(t) \circ e_*^{t\vec{H}} = \underline{\dot{J}}_{\lambda(t)}(0). \tag{13.4}$$

Since $e_*^{-t\vec{H}}$ is a symplectic transformation, it preserves the sign and the rank of the quadratic form.[1] $\square$

Remark 13.4. Notice that claim (iii) of Proposition 13.2 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 13.5. *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 $J_\lambda(t) = e_*^{-t\vec{H}}J_{\lambda(t)}(0)$.

13.2 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 13.6. *Let $\gamma(t) = \mathcal{E}_q(t\lambda)$ 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 7.31, $\gamma(s)$ is conjugate to $\gamma(0)$ if $s\lambda$ is a critical point of the exponential map $\mathcal{E}_q$. 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_*

$$e_*^{s\vec{H}} J_\lambda(0) \cap T_{\lambda(s)} T^*_{\gamma(s)} M \neq \{0\}. \tag{13.5}$$

Applying the linear invertible transformation $e_*^{-s\vec{H}}$ to both subspaces one gets that (13.5) 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 13.7. *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 13.8. 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 13.7 necessarily $t_0 > 0$.

Proof. Let us denote by $J_\lambda(t)$ the Jacobi curve associated with $\gamma(t)$. From Proposition 13.6 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

$$J_\lambda(0) \cap \bigcap_{t \in [t_0,t_1]} J_\lambda(t) \neq \{0\}. \tag{13.6}$$

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

$$J_\lambda(0) \cap J_\lambda(t_1) \subset J_\lambda(0) \cap \bigcap_{t \in [t_0,t_1]} J_\lambda(t) \tag{13.7}$$

that implies that actually we have the equality in (13.7).

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 13.9. 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

$$\sum_{i=1}^k h_i(\eta(t))^2 = 0, \qquad \dot{\eta}(t) = \sum_{i=1}^k h_i(\lambda(t))\vec{h}_i(\eta(t)). \tag{13.8}$$

Exercise 13.9 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.

$\square$

Corollary 13.10. *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 $0 \leq \tau \leq t$.*

13.3 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 13.11. This procedure was already exploited in Section 7.6, obtained by a direct argument via Proposition 7.27. 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 13.12. *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 13.13. 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 13.14. 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 13.15. *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 (7.31) 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

$$\delta_{-s} \circ e^{-t\vec{H}} \circ \delta_s = e^{-st\vec{H}}. \tag{13.9}$$

Computing the 2^{nd} order mixed partial derivative at $(t, s) = (0, 1)$ in (13.9) one gets, by (2.28), that $[\vec{H}, \vec{E}] = -\vec{H}$. Thus, by (2.30) we have $e_*^{-t\vec{H}}\vec{E} = \vec{E} - t\vec{H}$, since every higher order commutator vanishes.

□

Proposition 13.16. *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:

$$\sigma(\vec{E},\vec{H}) = \vec{E}(H) = \frac{d}{dt}\bigg|_{t=0} H(e^{t\vec{E}}(p,x)) = p\frac{\partial H}{\partial p} = 2H = 1. \tag{13.10}$$

It follows that $\{\vec{E},\vec{H}\}$ is a Darboux basis for $\widetilde{\Sigma}$. □

In particular we can consider the the symplectic splitting $\Sigma = \widetilde{\Sigma} \oplus \widetilde{\Sigma}^{\angle}$.

Exercise 13.17. 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 13.12 are satisfied and we could split our Jacobi curve.

Definition 13.18. The *reduced Jacobi curve* is defined as follows

$$\widehat{J}_\lambda(t) := J_\lambda(t) \cap \widetilde{\Sigma}^{\angle}. \tag{13.11}$$

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 13.19. 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 14

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.

14.1 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 14.1. 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 14.2. Let $X \in \mathrm{Vec}(M)$. The *horizontal lift* of X is the unique vector field $\nabla_X \in \mathrm{Vec}(E)$ such that

$$\nabla_X(z) \in \mathcal{H}_z, \qquad \pi_* \nabla_X = X, \qquad \forall\, z \in E. \tag{14.1}$$

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 14.3. 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}

$$\Phi_{t_0,t_1} := \overrightarrow{\exp}\int_{t_0}^{t_1} \nabla_{X_s} ds : E_{\gamma(t_0)} \to E_{\gamma(t_1)}, \qquad \text{for } 0 < t_0 < t_1 < T. \tag{14.2}$$

In the general case we need some extra assumptions on the vector field to ensure that (14.2) 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 14.4. 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.)

14.1.1 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 14.5. Let E be a smooth fiber bundle and ∇ a connection on E. Let $X, Y \in \mathrm{Vec}(M)$ and define

$$R(X,Y) := [\nabla_X, \nabla_Y]_{ver} \tag{14.3}$$

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 14.6. *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

$$R(aX,Y) = aR(X,Y), \qquad \text{where} \quad a \in \mathcal{C}^\infty(M).$$

Notice that the symbol a in the right hand side stands for the function $\pi^* a = a \circ \pi$ in $\mathcal{C}^\infty(E)$, 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 $a \in \mathcal{C}^\infty(M)$. Applying the definition of ∇ and the Leibnitz rule for the Lie bracket one gets

$$\begin{aligned} R(aX,Y) &= [\nabla_{aX}, \nabla_Y] - \nabla_{[aX,Y]} \\ &= a[\nabla_X, \nabla_Y] - (\nabla_Y a)\nabla_X - \nabla_{a[X,Y]-(Ya)X} \\ &= a[\nabla_X, \nabla_Y] - (Ya)\nabla_X - a\nabla_{[X,Y]} + (Ya)\nabla_X \\ &= aR(X,Y). \end{aligned}$$

□

14.1.2 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 $\mathcal{C}^\infty_L(E)$ the set of smooth functions on E that are linear on fibers.

Remark 14.7. 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 2.40). The "dual" version of this identification is the inclusion $i : \mathcal{C}^\infty(M) \to \mathcal{C}^\infty(E)$. Indeed any function in $\mathcal{C}^\infty(M)$ can be considered as a functions in $\mathcal{C}^\infty(E)$ which is constant on fibers, i.e. more precisely $a \in \mathcal{C}^\infty(M) \mapsto \pi^* a \in \mathcal{C}^\infty(E)$.

Exercise 14.8. 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 $\mathcal{C}^\infty(E)$, it maps the subspace $\mathcal{C}^\infty(M)$ into itself.

After this discussion it is natural to give the following definition.

Definition 14.9. 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 $a \in \mathcal{C}^\infty_L(E)$ it holds $\nabla_X a \in \mathcal{C}^\infty_L(E)$.

Remark 14.10. 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 $\pi_* \nabla_{X_i} = X_i$)

$$\nabla_{X_i} = X_i + \sum_{j=1}^{n} a_{ij}(u,q)\partial_{u_j}, \qquad i = 1, \ldots, n, \tag{14.4}$$

where $a_{ij} \in \mathcal{C}^\infty(E)$ 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 14.11. 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.

14.1.3 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

$$\Phi_{t_0,t_1} := \overrightarrow{\exp} \int_{t_0}^{t_1} \nabla_{X_s} ds : E_{\gamma(t_0)} \to E_{\gamma(t_1)}, \qquad \text{for } 0 < t_0 < t_1 < T. \tag{14.5}$$

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)

$$\Phi_{t_0,t_1}^* := \left(\overrightarrow{\exp} \int_{t_1}^{t_0} \nabla_{X_s} ds\right)^* : E_{\gamma(t_0)}^* \to E_{\gamma(t_1)}^*, \qquad \text{for } 0 < t_0 < t_1 < T. \tag{14.6}$$

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 $Y^* \in \mathcal{C}^\infty(T^*M)$ 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 14.12. 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 14.13. *Let $\{X_1, \ldots, X_n\}$ 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 14.14. The *torsion* of a linear connection ∇ is the map $T : \mathrm{Vec}(M)^2 \to \mathrm{Vec}(M)$ defined by the identity

$$T(X,Y)^* := \sigma(\nabla_X, \nabla_Y), \qquad \forall\, X, Y \in \mathrm{Vec}(M). \tag{14.7}$$

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 14.15. *The torsion tensor satisfies the identity*

$$T(X,Y) = \nabla_X Y - \nabla_Y X - [X,Y]. \tag{14.8}$$

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 (4.52)

$$\begin{aligned} T(X,Y)^* &= ds(\nabla_X, \nabla_Y) \\ &= \nabla_X \langle s, \nabla_Y \rangle - \nabla_Y \langle s, \nabla_X \rangle - \langle s, [\nabla_X, \nabla_Y] \rangle \\ &= \nabla_X \langle s, \nabla_Y \rangle - \nabla_Y \langle s, \nabla_X \rangle - \left\langle s, \nabla_{[X,Y]} \right\rangle \\ &= \nabla_X Y^* - \nabla_Y X^* - [X,Y]^*, \end{aligned}$$

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 14.16. Show that a linear connection ∇ on a vector bundle E satisfies the following Leibnitz rule

$$\nabla_X(aY) = a\nabla_X Y + (Xa)Y, \qquad \text{for each } a \in \mathcal{C}^\infty(M).$$

[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.

14.2 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 12, 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 14.17. 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 14.18. *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.

$$\mathcal{H}_{c\lambda} = \delta_{c*}\mathcal{H}_\lambda, \qquad \forall\, c > 0. \tag{14.9}$$

Indeed in this case the functions $a_{ij} \in \mathcal{C}^\infty(T^*M)$ defining the connection (see (14.4)) are 1-homogeneous, hence linear as a consequence of Exercise 14.19.

Let us prove (14.9). 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

$$e^{t\vec{H}} \circ \delta_c = \delta_c \circ e^{ct\vec{H}}, \qquad \forall\, c > 0, \tag{14.10}$$

one easily gets that

$$J_{c\lambda}(t) = \delta_{c*}J_\lambda(ct), \qquad \forall\, t \geq 0, \lambda \in T^*M. \tag{14.11}$$

Indeed one has the following chain of identities

$$\begin{aligned}
J_{c\lambda}(t) &= e^{-t\vec{H}}_*(T_{c\lambda}(T^*_qM)) \\
&= e^{-t\vec{H}}_* \circ \delta_{c*}(T_\lambda(T^*_qM)) \qquad \text{(by (14.10))} \\
&= \delta_{c*} \circ e^{-ct\vec{H}}_*(T_\lambda(T^*_qM)) \\
&= \delta_{c*}J_\lambda(ct).
\end{aligned}$$

Now we show that the same relation holds true also for the derivative curve, i.e.

$$J^\circ_{c\lambda}(t) = \delta_{c*} J^\circ_\lambda(ct), \qquad \forall\, t \geq 0,\, \lambda \in T^*M. \tag{14.12}$$

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 (14.11) 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 (14.12) follows from the identity $S^\circ(t) = B^{-1}(t) + S(t)$. (See also Exercise 12.22). In particular at $t = 0$ the identity (14.12) 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 14.19. 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 14.20. *There is a unique Ehresmann connection ∇ satisfying the properties (i), (ii), and (iii) of Proposition 14.18, that is the Levi-Civita connection. Its Christoffel symbols are computed by*

$$\Gamma^k_{ij} = \frac{1}{2}(c^k_{ij} - c^i_{jk} + c^j_{ki}), \tag{14.13}$$

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 $X_1, \dots, X_n$ be a local orthonormal frame for the Riemannian structure and let us consider coordinates (q, u) in T^*M, where the fiberwise coordinates $u = (u_1, \dots, u_n)$ 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

$$\nabla_{X_i} = X_i + \sum_{j=1}^n \Gamma^k_{ij} u_k \partial_{u_j}, \qquad i = 1, \dots, n.$$

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 ($\nu_1, \dots, \nu_n$ denotes the dual basis to $X_1, \dots, X_n$)

$$\sigma = \sum_{i,j,k=1}^n du_k \wedge \nu_k - c^k_{ij} u_k \nu_i \wedge \nu_k.$$

[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

$$0 = \sigma(\nabla_{X_i}, \nabla_{X_j}) = \sum_{k=1}^{n} (\Gamma_{ij}^k - \Gamma_{ji}^k - c_{ij}^k) u_k, \qquad \forall\, i, j = 1, \dots, n,$$

hence $c_{ij}^k = \Gamma_{ij}^k - \Gamma_{ji}^k$ for all i, j, k. Moreover the connection is metric, i.e. it satisfies

$$0 = \nabla_{X_i} H = \sum_{j,k=1}^{n} \Gamma_{ij}^k u_k u_j, \qquad \forall\, i = 1, \dots, n.$$

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 14.21. Notice that in the classical approach one can recover formula (14.13) 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 $X_1, \dots, X_n$. 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 14.22 (Bianchi identity)**.** *For every $X, Y, Z \in \mathrm{Vec}(M)$ the following identity holds*

$$R(X,Y)Z + R(Y,Z)X + R(Z,X)Y = 0. \tag{14.14}$$

Proof. We will show that (14.14) is a consequence of the Jacobi identity (2.31). 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

$$\begin{aligned} 0 &= [X,[Y,Z]] + [Y,[Z,X]] + [Z,[X,Y]] \\ &= \nabla_X \nabla_Y Z - \nabla_X \nabla_Z Y - \nabla_{[Y,Z]} X \\ &\quad + \nabla_Z \nabla_X Y - \nabla_Z \nabla_Y X - \nabla_{[X,Y]} Z \\ &\qquad + \nabla_Y \nabla_Z X - \nabla_Y \nabla_X Z - \nabla_{[Z,X]} Y \\ &= R(X,Y)Z + R(Y,Z)X + R(Z,X)Y. \end{aligned}$$

□

Exercise 14.23. 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:

$$(X,Y,Z,W) + (Z,X,Y,W) + (Y,Z,X,W) = 0, \qquad \forall\, X,Y,Z,W \in \mathrm{Vec}(M). \tag{14.15}$$

Remark 14.24. The property of the Riemann tensor can be reformulated as follows

$$(X,Y,Z,W) = -(Y,X,Z,W), \qquad (X,Y,Z,W) = -(X,Y,W,Z). \tag{14.16}$$

Proposition 14.25. *For every $X,Y,Z,W \in \mathrm{Vec}(M)$ we have $(X,Y,Z,W) = (Z,W,X,Y)$.*

Proof. Using (14.15) 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 (14.16), one gets $(X,Z,W,Y) = (W,Y,X,Z)$. □

Proposition 14.26. *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 (14.16) of the Riemann tensor we have that $(X,Y,Z,W) = 0$ whenever any two of the vector fields coincide. In particular

$$0 = (X, Y+W, Z, Y+W) = (X,Y,Z,W) + (X,W,Z,Y). \tag{14.17}$$

since the two extra terms that should appear in the expansion vanish by assumptions. Then (14.17) 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 (14.15), hence $(X,Y,Z,W) = 0$. □

We end this section by summarizing the symmetry property of the Riemann curvature as follows

Corollary 14.27. *There is a well defined map*

$$\overline{R} : \wedge^2 T_q M \to \wedge^2 T_q M, \qquad \overline{R}(X \wedge Y) := R(X,Y).$$

Moreover $\bar{R}$ is skew-adjoint with respect to the induced scalar product on $\wedge^2 T_q M$, that means

$$\left\langle \overline{R}(X \wedge Y), Z \wedge W \right\rangle = \left\langle X \wedge Y, \overline{R}(Z \wedge W) \right\rangle.$$

14.3 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 14.28. 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 14.29. *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 12.30. 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. $\eta \in \mathcal{H}_\lambda$). $\square$

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 12.36 one gets

Corollary 14.30. *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 $\lambda \in T^*_q M$ and $v \in T_q M$ if $\langle \lambda, w \rangle = \langle v | w \rangle$ for all $w \in T_q M$.

Let denote $H_q = H|_{T^*_q M}$. Recall that the differential of H_q can be interpreted as a linear map $DH_q : T^*_q M \to T_q M$ that sends $\lambda \in T^*_q M$ into $D_\lambda H_q$ seen as a linear functional on $T^*_q M$, i.e. a tangent vector. This map is actually the inverse of the isomorphism I.

Lemma 14.31. $D_\lambda H_q = I^{-1}(\lambda)$.

Proof. It is a simple consequence of the formula $H(\lambda) = \frac{1}{2} \langle \lambda, I^{-1}(\lambda) \rangle$. $\square$

Corollary 14.32. *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 14.31. 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 $v = \pi_*\vec{H}(\lambda)$. $\square$

Theorem 14.33. *We have the following identity*

$$\mathcal{R}_{I(X)}(I(Y)) = R(X,Y)X, \qquad \forall\, X, Y \in T_qM. \tag{14.18}$$

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

$$-[\vec{H}, [\vec{H}, I(Y)]_{hor}]_{ver}(I(X)) = [\nabla_X, \nabla_Y]_{ver}(I(X)) = R(X,Y)(X).$$

$\square$

Definition 14.34. 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 14.35. 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.

14.4 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 14.36. *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 14.37. 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

$$\sigma = ds = \sum_{i=1}^n d\psi_i \wedge \nu_i + \psi_i d\nu_i. \tag{14.19}$$

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 $H(\psi, q) = \frac{1}{2}|\psi|^2$. $\square$

The next result shows that the Hamiltonian curvature can detect if a manifold is flat or not.

Corollary 14.38. *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 (14.18).

Let us prove the converse. Recall that $\mathcal{R}_\lambda = 0$ implies, again by (14.18), that

$$(X, Y, X, W) = 0, \qquad \forall\, X, Y, W \in \mathrm{Vec}(M).$$

Then the statement is a consequence of Proposition 14.26. $\square$

Exercise 14.39. Prove that actually the Riemann tensor R is completely determined by $\mathcal{R}$.

14.5 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)

$$H(p, x) = \frac{1}{2}(\langle p, f_1(x)\rangle^2 + \langle p, f_2(x)\rangle^2) \tag{14.20}$$

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 14.40. *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:

$$\sigma(\partial_\theta, \vec{E}) = 0, \qquad \sigma(\partial_\theta, \vec{H}) = 0, \qquad \sigma(\vec{E}, \vec{H}) = 1. \tag{14.21}$$

$$\sigma(\partial_\theta, \vec{H}') = 1, \qquad \sigma(\vec{E}, \vec{H}') = 0, \qquad \sigma(\vec{H}, \vec{H}') = 0. \tag{14.22}$$

Indeed, let us prove first (14.21). 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 (13.10).

As a preliminary step for the proof of (14.22) notice that, if $s = i_{\vec{E}}\sigma$ denotes the tautological Liouville form, one has

$$\langle s, \vec{H}\rangle = 1, \qquad \langle s, \vec{H}'\rangle = 0. \tag{14.23}$$

These two identities follows from

$$\langle s, \vec{H}\rangle = \sigma(\vec{E}, \vec{H}) = 1, \tag{14.24}$$

$$\langle s, \vec{H}'\rangle = \langle s, [\partial_\theta, \vec{H}]\rangle = ds(\partial_\theta, \vec{H}) = \sigma(\partial_\theta, \vec{H}) = 0, \tag{14.25}$$

where in the second line we used the Cartan formula (4.52) and the fact that ∂_θ is vertical.

Let us now prove (14.22). Being $[\partial_\theta, \vec{H}'] = [\partial_\theta, [\partial_\theta, \vec{H}]] = -\vec{H}$, we have again by Cartan formula and (14.23)

$$\sigma(\partial_\theta, \vec{H}') = ds(\partial_\theta, \vec{H}') = -\langle s, [\partial_\theta, \vec{H}']\rangle = \langle s, \vec{H}\rangle = \sigma(\vec{E}, \vec{H}) = 1$$

Moreover by (14.23)

$$\sigma(\vec{E}, \vec{H}') = \langle s, \vec{H}'\rangle = 0.$$

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 14.40 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

$$\widehat{J}_\lambda(t) = \operatorname{span}\{e_*^{-t\vec{H}}\partial_\theta\}.$$

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

$$\sigma(V_t, \dot{V}_t) = -1, \qquad \text{for every } t \geq 0. \tag{14.26}$$

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

$$V_t = \alpha(t)V_0 - \beta(t)\dot{V}_0 \tag{14.27}$$

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

$$V_t = \alpha(t)V_0 - \beta(t)\dot{V}_0 = \alpha(t)\left(V_0 - \frac{\beta(t)}{\alpha(t)}\dot{V}_0\right), \tag{14.28}$$

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

$$\sigma(V_0, V_t) = \alpha(t)\underbrace{\sigma(V_0, V_0)}_{=0} - \beta(t)\underbrace{\sigma(V_0, \dot{V}_0)}_{=-1} = \beta(t), \tag{14.29}$$

$$\sigma(\dot{V}_0, V_t) = \alpha(t)\underbrace{\sigma(\dot{V}_0, V_0)}_{=1} - \beta(t)\underbrace{\sigma(\dot{V}_0, \dot{V}_0)}_{=0} = \alpha(t). \tag{14.30}$$

Differentiating the identity (14.26) 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

$$S(t) = \frac{-t + \dfrac{t^3}{6}\sigma(V_0, \dddot{V}_0) + O(t^5)}{1 + \dfrac{t^2}{2}\sigma(\dot{V}_0, \ddot{V}_0) + O(t^4)} \tag{14.31}$$

$$= \left(-t + \frac{t^3}{6}\sigma(V_0, \dddot{V}_0) + O(t^5)\right)\left(1 - \frac{t^2}{2}\sigma(\dot{V}_0, \ddot{V}_0) + O(t^4)\right) \tag{14.32}$$

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:

$$\mathcal{R} = -\frac{1}{2}\dddot{S}(0) = \sigma(\ddot{V}_0, \dot{V}_0) \tag{14.33}$$

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 (14.33) is rewritten as

$$\mathcal{R} = \sigma([\vec{H}, [\vec{H}, \partial_\theta]], [\vec{H}, \partial_\theta]) \tag{14.34}$$

$$= \sigma([\vec{H}, \vec{H}'], \vec{H}') \tag{14.35}$$

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

$$\begin{cases} \dot{x} = h_1 f_1(x) + h_2 f_2(x) \\ \dot{h}_1 = \{H, h_1\} = \{h_2, h_1\}h_2 \\ \dot{h}_2 = \{H, h_2\} = -\{h_2, h_1\}h_1 \end{cases} \tag{14.36}$$

Moreover $\{h_2, h_1\}(p,x) = \langle p, [f_2, f_1](x)\rangle$. Assume that

$$[f_1, f_2] = a_1 f_1 + a_2 f_2, \qquad a_i \in \mathcal{C}^\infty(M).$$

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 (14.36) 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 14.41. 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 15

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 some general features of the curvature in sub-Riemannian geometry.

15.1 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 15.1. 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 15.2. 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 $a \in \mathcal{C}^\infty(M)$. 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 15.3. 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 15.4. 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

$$\begin{cases} d\nu_0 = \nu_1 \wedge \nu_2 \\ d\nu_1 = c_{01}^1\nu_0 \wedge \nu_1 + c_{02}^1\nu_0 \wedge \nu_2 + c_{12}^1\nu_1 \wedge \nu_2 \\ d\nu_2 = c_{01}^2\nu_0 \wedge \nu_1 + c_{02}^2\nu_0 \wedge \nu_2 + c_{12}^2\nu_1 \wedge \nu_2 \end{cases} \tag{15.1}$$

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 15.5. *The Poisson bracket $\{H, h_0\}_q$ is a quadratic form. Moreover we have*

$$\{H, h_0\} = c_{01}^1 h_1^2 + (c_{01}^2 + c_{02}^1)h_1h_2 + c_{02}^2 h_2^2, \tag{15.2}$$

$$c_{01}^1 + c_{02}^2 = 0. \tag{15.3}$$

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

$$\begin{aligned} \{H, h_0\} &= \frac{1}{2}\{h_1^2 + h_2^2, h_0\} = h_1\{h_1, h_0\} + h_2\{h_2, h_0\} \\ &= h_1(c_{01}^1 h_1 + c_{01}^2 h_2) + h_2(c_{02}^1 h_1 + c_{02}^2 h_2) \\ &= c_{01}^1 h_1^2 + (c_{01}^2 + c_{02}^1)h_1h_2 + c_{02}^2 h_2^2. \end{aligned}$$

Differentiating the first equation in (15.1) 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 (15.3). □

Remark 15.6. 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 (15.3) 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:

$$\chi(q) = \sqrt{-\det\{h, h_0\}_q}. \tag{15.4}$$

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 15.7. *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.

15.1.1 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 14.5 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

$$\begin{cases} \dot{x} = h_1 f_1(x) + h_2 f_2(x) \\ \dot{h}_1 = \{H, h_1\} = \{h_2, h_1\} h_2 \\ \dot{h}_2 = \{H, h_2\} = -\{h_2, h_1\} h_1 \\ \dot{h}_0 = \{H, h_0\} \end{cases} \tag{15.5}$$

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

$$\begin{cases} \dot{x} = \cos\theta f_1 + \sin\theta f_2 \\ \dot{\theta} = -h_0 - b \\ \dot{h}_0 = a \end{cases} \tag{15.6}$$

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 $\sin\theta$):

$$\vec{H} = h_1 f_1 + h_2 f_2 - (h_0 + b)\partial_\theta + a\partial_{h_0} \tag{15.7}$$

$$[\partial_\theta, \vec{H}] = \vec{H}' = -h_2 f_1 + h_1 f_2 + a'\partial_{h_0} - b'\partial_\theta \tag{15.8}$$

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 $t = 0$

$$\Gamma := \operatorname{Ker} \underline{\dot{\Lambda}}(0)$$

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 15.8. *We have the following characterizations:*

(i) $\Gamma^{\angle} = \mathrm{span}\{\partial_{h_0}, \partial_\theta, \vec{H}'\}$ *in* $\widehat{\mathcal{V}}_\lambda/\mathbb{R}\vec{H}$*,*

(ii) $\{\partial_\theta, \vec{H}'\}$ *is a Darboux basis for* $\Gamma^{\angle}/\Gamma$*.*

Proof. Since ∂_{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 (15.8).

To complete the proof of (ii) it is enough to show, using $[\partial_\theta, \vec{H}'] = -\vec{H}$, that

$$\sigma(\partial_\theta, \vec{H}') = \partial_\theta\langle s, \vec{H}'\rangle - \vec{H}'\,\langle s, \partial_\theta\rangle - \langle s, [\partial_\theta, \vec{H}']\rangle = \langle s, \vec{H}\rangle = 1.$$

□

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 15.9. *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 $\Lambda(t) = \mathrm{span}\{V_t, \dot{V}_t\}$. 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

$$\sigma(\dot{V}_t + x(t)V_t, V_0) = 0 \qquad \Longrightarrow \qquad x(t) = -\frac{\sigma(\dot{V}_t, V_0)}{\sigma(V_t, V_0)}. \tag{15.9}$$

Then we write this vector as a linear combination of the Darboux basis (cf. (14.28) for the 2D Riemannian case)

$$\dot{V}_t + x(t)V_t = \alpha(t)\dot{V}_0 - \beta(t)\ddot{V}_0 + \xi(t)V_0 \tag{15.10}$$

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

$$S(t) = \frac{\beta(t)}{\alpha(t)} \tag{15.11}$$

Notice that this is a one-dimensional non-degenerate curve. These coefficients are computed by the symplectic products

$$\alpha(t) = -\sigma(\dot{V}_t + x(t)V_t, \ddot{V}_0) \tag{15.12}$$

$$\beta(t) = -\sigma(\dot{V}_t + x(t)V_t, \dot{V}_0) \tag{15.13}$$

Combining (15.12),(15.13) with (15.11) and (15.9) one gets

$$S(t) = \frac{\sigma(\dot{V}_t, \dot{V}_0)\sigma(V_t, V_0) - \sigma(V_t, \dot{V}_0)\sigma(\dot{V}_t, V_0)}{\sigma(\dot{V}_t, \ddot{V}_0)\sigma(\dot{V}_t, \dot{V}_0) - \sigma(\dot{V}_t, \ddot{V}_0)\sigma(\dot{V}_t, \dot{V}_0)} \tag{15.14}$$

After some computations, by Taylor expansion one gets

$$S(t) = \frac{t}{4} - \frac{t^3}{120}r + O(t^4) \tag{15.15}$$

Since $\ddot{S}_0 = 0$ the curvature is computer by

$$\mathcal{R} = \frac{\dddot{S}_0}{2\dot{S}_0} = -\frac{r}{10}$$

□

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 (15.7) and (15.8) 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

$$\kappa = f_2 c_{12}^1 - f_1 c_{12}^2 - (c_{12}^1)^2 - (c_{12}^2)^2 + \frac{c_{01}^2 - c_{02}^1}{2} \tag{15.16}$$

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 15.10. 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 15.11. (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

$$\operatorname{trace}(10\mathcal{R}\big|_{h_0=0}) = 2\kappa, \qquad \operatorname{discr}(10\mathcal{R}\big|_{h_0=0}) = 36\chi.$$

(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 15.12. 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

$$\begin{aligned}
\widehat{c}_{12}^1 &= \cos\theta(c_{12}^1 - f_1(\theta)) - \sin\theta(c_{12}^2 - f_2(\theta)), \\
\widehat{c}_{12}^2 &= \sin\theta(c_{12}^1 - f_1(\theta)) + \cos\theta(c_{12}^2 - f_2(\theta)).
\end{aligned}$$

Exercise 15.13. Show that the expression (15.16) for κ does not depend on the choice of an orthonormal frame f_1, f_2 for the sub-Riemannian structure.

Chapter 16

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 15.1.1)

$$\vec{H} = h_1 f_1 + h_2 f_2 - (h_0 + b)\partial_\theta + a\partial_{h_0} \tag{16.1}$$

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 distribution. 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 (16.1) on the level set $H^{-1}(1/2)$ is rewritten in the following form:

$$\begin{cases} \dot{q} = \cos\theta f_1 + \sin\theta f_2 \\ \dot{\theta} = \rho - b \\ \dot{\rho} = -a \end{cases} \tag{16.2}$$

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 $(\theta_0, \rho_0) \in T^*_{q_0}M$ assigns the solution at time t of the system (16.2), denoted by $\mathcal{E}_{q_0}(t, \theta_0, \rho_0)$, or simply $\mathcal{E}(t, \theta_0, \rho_0)$.

Conjugate points are points where the differential of the exponential map is not surjective, i.e. solutions to the equation

$$\frac{\partial \mathcal{E}}{\partial \theta_0} \wedge \frac{\partial \mathcal{E}}{\partial \rho_0} \wedge \frac{\partial \mathcal{E}}{\partial t} = 0. \tag{16.3}$$

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 7.6 and Proposition 7.27). This implies that (16.3) is equivalent to

$$\frac{\partial \mathcal{E}}{\partial \theta_0} \wedge \frac{\partial \mathcal{E}}{\partial \rho_0} = 0. \tag{16.4}$$

16.1 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

$$\begin{cases} \dot{q} = \cos\theta f_1 + \sin\theta f_2 \\ \dot{\theta} = \rho \\ \dot{\rho} = 0 \end{cases} \tag{16.5}$$

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

$$q = (x, y), \qquad x = (x_1, x_2) \in \mathbb{R}^2, y \in \mathbb{R}. \tag{16.6}$$

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 (16.4) is equivalent to the following

$$\frac{\partial x}{\partial \theta_0} \wedge \frac{\partial x}{\partial \rho_0} = 0. \tag{16.7}$$

By the Gauss' Lemma (Proposition 7.27) the covector $p = (p_x, \rho)$ at the final point annihilates the differential of the exponential map restricted to the level set, i.e.

$$\left\langle p, \frac{\partial \mathcal{E}}{\partial \theta_0} \right\rangle = \left\langle p_x, \frac{\partial x}{\partial \theta_0} \right\rangle + \rho \frac{\partial y}{\partial \theta_0} = 0 \tag{16.8}$$

$$\left\langle p, \frac{\partial \mathcal{E}}{\partial \rho_0} \right\rangle = \left\langle p_x, \frac{\partial x}{\partial \rho_0} \right\rangle + \rho \frac{\partial y}{\partial \rho_0} = 0 \tag{16.9}$$

and since $\rho \neq 0$ it follows that among the three vectors

$$\begin{pmatrix} \dfrac{\partial x_1}{\partial \theta_0} \\ \dfrac{\partial x_1}{\partial \rho_0} \end{pmatrix} \quad \begin{pmatrix} \dfrac{\partial x_2}{\partial \theta_0} \\ \dfrac{\partial x_2}{\partial \rho_0} \end{pmatrix} \quad \begin{pmatrix} \dfrac{\partial y}{\partial \theta_0} \\ \dfrac{\partial y}{\partial \rho_0} \end{pmatrix} \tag{16.10}$$

the third one is always a linear combination of the first two.

Proposition 16.1. *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

$$\begin{cases} \dot{x}_1 = \cos\theta \\ \dot{x}_2 = \sin\theta \end{cases} \tag{16.11}$$

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 $\rho > 0$)

$$x(t; \theta_0, \rho_0) = \int_0^t \begin{pmatrix} \cos(\theta_0 + \rho s) \\ \sin(\theta_0 + \rho s) \end{pmatrix} ds = \int_{\theta_0}^{\theta_0 + t} \begin{pmatrix} \cos \rho s \\ \sin \rho s \end{pmatrix} ds \tag{16.12}$$

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 16.2. $\square$

Exercise 16.2. 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.

16.2 General case: second order asymptotic expansion

Let us consider the Hamiltonian system for the general 3D contact case

$$\begin{cases} \dot{q} = f_\theta := \cos \theta f_1 + \sin \theta f_2 \\ \dot{\theta} = \rho - b \\ \dot{\rho} = -a \end{cases} \tag{16.13}$$

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 16.3. *The conjugate time $t_c(\theta_0, \nu)$ is a smooth function of the parameter ν. Moreover*

$$t_c(\theta_0, \nu) = 2\pi |\nu| + O(|\nu|^3).$$

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 (16.13) is rewritten in the new coordinate system (q, θ, r) (where we recall $r = 1/\rho$), as follows

$$\begin{cases} \dot{q} = r f_\theta \\ \dot{\theta} = 1 - rb \\ \dot{r} = r^3 a \\ \dot{t} = r \end{cases} \tag{16.14}$$

To compute the asymptotic 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 $f_1^\varepsilon, \dots, f_k^\varepsilon$ (and analogously for $a^\varepsilon, b^\varepsilon$).

Let us now define, for fixed ε, the variable w such that $r = \varepsilon w$. The system (16.14) is finally rewritten in the following form

$$\begin{cases} \dot{q} = w f_\theta^\varepsilon \\ \dot{\theta} = 1 - w b^\varepsilon \\ \dot{w} = \varepsilon w^3 a^\varepsilon \\ \dot{t} = \varepsilon w \end{cases} \tag{16.15}$$

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 $w(0) = w_1$ 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 8.22). If we denote by $q^\varepsilon = (x^\varepsilon, y^\varepsilon)$ the solution of the horizontal part of the ε-system (16.15), 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 16.1, 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

$$\varphi(\tau, \varepsilon, \theta_0) := \left.\frac{\partial x^\varepsilon}{\partial \theta_0} \wedge \frac{\partial x^\varepsilon}{\partial w_0}\right|_{w_0 = 1} = 0 \tag{16.16}$$

For $\varepsilon = 0$, i.e. the nilpotent approximation, the the first conjugate time is $\tau_c = 2\pi$, and moreover it is a simple root. Thus one gets

$$\varphi(2\pi, 0, \theta_0) = 0, \qquad \frac{\partial \varphi}{\partial \tau}(2\pi, 0, \theta_0) \neq 0. \tag{16.17}$$

Hence the implicit function theorem guarantees that there exists a smooth function $\tau_c(\varepsilon, \theta_0)$ such that

$$\varphi(\tau_c(\varepsilon, \theta_0), \varepsilon, \theta_0) = 0. \tag{16.18}$$

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

$$\tau_c(\varepsilon, \theta_0) = 2\pi + O(\varepsilon^2). \tag{16.19}$$

Differentiating the identity (16.18) 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 (16.17), the expansion (16.19) holds if and only if $\dfrac{\partial \varphi}{\partial \varepsilon}(2\pi, 0, \theta_0) = 0$.

Moreover differentiating the expression (16.16) 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 vanish 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

$$\frac{\partial^2 x^\varepsilon}{\partial \varepsilon \partial \theta_0}\bigg|_{\varepsilon=0, \tau=2\pi} = 0. \tag{16.20}$$

which is a consequence of the following lemma.

Lemma 16.4. *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 (16.15).

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

$$\left.\frac{\partial \theta}{\partial \varepsilon}\right|_{\varepsilon=0} = -\int_0^t \langle \beta, u(\theta(\tau)) \rangle \, d\tau = \big\langle \beta, u'(\theta_0 + t) - u'(\theta_0) \big\rangle \tag{16.21}$$

where $u'(\theta) = (-\sin\theta, \cos\theta)$.

Next we are going to use (16.21) to compute the derivative of x^ε wrt ε. The equation for the horizontal part of (16.15) 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 16.5. *Let $\phi(\varepsilon, t)$ denotes 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 16.5 we have

$$\frac{d}{dt}\frac{\partial x}{\partial \varepsilon} = \frac{\partial F^x}{\partial x}\frac{\partial x}{\partial \varepsilon} + \frac{\partial F^x}{\partial \theta}\frac{\partial \theta}{\partial \varepsilon} + \frac{\partial F^x}{\partial \varepsilon}$$

Computing the derivatives at $\varepsilon = 0$ gives

$$\left.\frac{\partial F^x}{\partial x}\right|_{\varepsilon=0} = 0, \qquad \left.\frac{\partial F^x}{\partial \theta}\right|_{\varepsilon=0} = u'(\bar{\theta}(t)), \qquad \left.\frac{\partial F^x}{\partial \varepsilon}\right|_{\varepsilon=0} = f^{(0)}_{u(\bar{\theta}(t))}(\bar{x}(t))$$

and we obtain the equation for $\frac{\partial x}{\partial \varepsilon}$

$$\left.\frac{d}{dt}\frac{\partial x}{\partial \varepsilon}\right|_{\varepsilon=0} = \left.\frac{\partial \theta}{\partial \varepsilon}\right|_{\varepsilon=0} u'(\theta_0 + t) + f^{(0)}_{u(\theta_0+t)}(u'(\theta_0) - u'(\theta_0 + t))$$

[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.

If we set $s = \theta_0 + t$ we can rewrite this equation

$$\frac{d}{ds}\frac{\partial x}{\partial \varepsilon}\Big|_{\varepsilon=0} = \frac{\partial \theta}{\partial \varepsilon}u'(s) + f^{(0)}_{u(s)}(u'(\theta_0) - u'(s))$$

and integrating one has

$$\begin{aligned}\frac{\partial x}{\partial \varepsilon}\Big|_{(2\pi,0)} &= \int_{\theta_0}^{\theta_0+2\pi} \big\langle \beta, u'(s) - u'(\theta_0)\big\rangle\, u'(s)ds \\ &\qquad + \int_{\theta_0}^{\theta_0+2\pi} f^{(0)}_{u(s)}(u'(\theta_0) - u'(s))ds\end{aligned}$$

In the last expression it is easy to see that all terms where θ_0 appear 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 16.4, hence the proof of the Proposition 16.3. □

16.3 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 16.3.1.

Theorem 16.6. *In a system of local coordinates around $q_0 \in M$ one has the expansion*

$$Con_{q_0}(\theta_0, \nu) = q_0 \pm \pi f_0|\nu|^2 \pm \pi(a' f_{\theta_0} - a f_{\theta_0'})|\nu|^3 + O(|\nu|^4), \qquad \nu \to 0^{\pm}. \tag{16.22}$$

If we choose coordinates such that $a = 2\chi u_1 u_2$ one gets

$$Con_{q_0}(\theta_0, \nu) = q_0 \pm \pi f_0|\nu|^2 \pm 2\pi\chi(q_0)(\cos^3\theta f_2 - \sin^3\theta f_1)|\nu|^3 + O(|\nu|^4), \qquad \nu \to 0^{\pm}. \tag{16.23}$$

Moreover for the conjugate length we have the expansion

$$\ell_c(\theta_0, \nu) = 2\pi|\nu| - \pi\kappa|\nu|^3 + O(|\nu|^4), \qquad \nu \to 0^{\pm}. \tag{16.24}$$

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 16.7. *Assume $\chi(q_0) \neq 0$. In a system of local coordinates around $q_0 \in M$ such that $a = 2\chi u_1 u_2$ one gets*

$$Cut_{q_0}(\theta, \nu) = q_0 \pm \pi\nu^2 f_0(q_0) \pm 2\pi\chi(q_0)\cos\theta f_1(q_0)\nu^3 + O(\nu^4), \qquad \nu \to 0^{\pm}$$

Moreover the cut length satisfies

$$\ell_{cut}(\theta, \nu) = 2\pi|\nu| - \pi(\kappa + 2\chi\sin^2\theta)|\nu|^3 + O(\nu^4), \qquad \nu \to 0^{\pm} \tag{16.25}$$

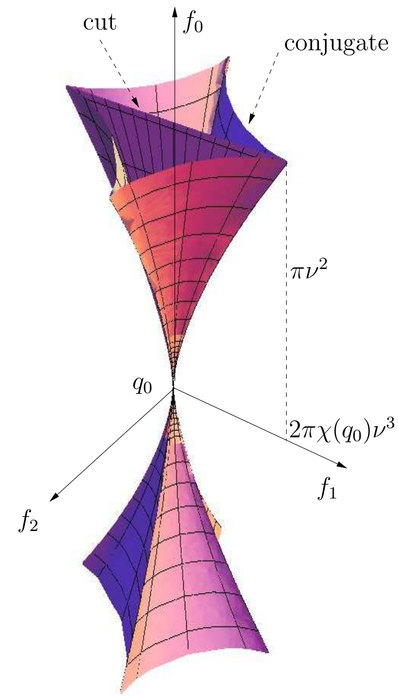


Figure 16.1: Asymptotic structure of cut and conjugate locus

We can collect the information given by the asymptotics of the conjugate and the cut locus in Figure 16.1.

All geometrical information about the structure of this 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 4.29).

Let us consider the kernel of the sub-Riemannian Hamiltonian

$$\ker H = \{\lambda \in T_q^*M : \langle \lambda, v\rangle = 0,\ \forall\, v \in \mathcal{D}_q\} = \mathcal{D}_q^{\perp}. \tag{16.26}$$

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 15.1.1)

$$10\mathcal{R} = (\kappa + 3\chi)h_1^2 + (\kappa - 3\chi)h_2^2.$$

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 16.8. The expression of the cut locus given in Theorem 16.7 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 we compute in the next section).

As we show in the next section, 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 omeomorphism of the the singular set of the truncated map. This proves that the shape of the conjugate locus (and the one of the cur locus) described in Figure 16.1 obtained via its third order approximation is indeed a picture of the *true* shape. The full statement of this fact can be found in [3].

16.3.1 Proof of Theorem 16.6: asymptotics of the exponential map

The proof of Theorem 16.6 requires a careful analysis of the asymptotic of the exponential map. Let us consider again our Hamiltonian system in the form (16.14)

$$\begin{cases} \dot{q} = r f_\theta \\ \dot{\theta} = 1 - rb \\ \dot{r} = r^3 a \\ \dot{t} = r \end{cases} \tag{16.27}$$

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:

$$\begin{cases} \dot{q} = \dfrac{r}{1-rb} f_\theta \\ \dot{\theta} = 1 \\ \dot{r} = \dfrac{r^3}{1-rb} a \\ \dot{\ell} = \dfrac{r}{1-rb} \end{cases} \tag{16.28}$$

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 (16.28) 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

$$Q_t = \overrightarrow{\exp} \int_{\theta_0}^{\theta_0+t} V_\tau d\tau, \qquad V_\tau := \frac{r(\tau)}{1-r(\tau)b(\tau)} f_\tau. \tag{16.29}$$

We start by analyzing the asymptotics of the end point map after time $t = 2\pi$.

Lemma 16.9. $F(2\pi;\theta_0,\nu) = -\pi f_0(q_0)\nu^2 + O(\nu^3)$

Proof. From (16.28), 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. (6.9))

$$\overrightarrow{\exp} \int_{\theta_0}^{\theta_0+2\pi} V_\tau d\tau = \left(\mathrm{Id} + \int_{\theta_0}^{\theta_0+2\pi} V_\tau d\tau + \iint\limits_{\theta_0 \le \tau_2 \le \tau_1 \le \theta_0+2\pi} V_{\tau_2} \circ V_{\tau_1} d\tau_1 d\tau_2 + \dots \right) \tag{16.30}$$

and substituting $r(\tau) \equiv \nu$ we have the following expansion for the first term in (16.30):

$$\begin{aligned} \int_{\theta_0}^{\theta_0+2\pi} V_\tau d\tau = \int_{\theta_0}^{\theta_0+2\pi} \frac{\nu}{1-\nu b(\tau)} f_\tau d\tau &= \int_{\theta_0}^{\theta_0+2\pi} \nu(1+\nu b(\tau) + O(\nu^2)) f_\tau \, d\tau, \\ &= \nu \int_{\theta_0}^{\theta_0+2\pi} f_\tau d\tau + \nu^2 \int_{\theta_0}^{\theta_0+2\pi} b(\tau) f_\tau d\tau + O(\nu^3) \\ &= \nu^2 \int_{\theta_0}^{\theta_0+2\pi} b(\tau) f_\tau d\tau + O(\nu^3) \end{aligned}$$

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 (16.28) can be rewritten using Lemma 7.19

$$\begin{aligned}
\iint\limits_{0\le\tau_2\le\tau_1\le t} V_{\tau_2}\circ V_{\tau_1} d\tau_1 d\tau_2 &= \frac{1}{2}\int_{\theta_0}^{\theta_0+2\pi} V_\tau dt \circ \int_{\theta_0}^{\theta_0+2\pi} V_\tau dt + \iint\limits_{\theta_0\le\tau_2\le\tau_1\le\theta_0+2\pi} [V_{\tau_2}, V_{\tau_1}] d\tau_1 d\tau_2 \\
&= \frac{\nu^2}{2}\left(\int_{\theta_0}^{\theta_0+2\pi} f_\tau d\tau \circ \int_{\theta_0}^{\theta_0+2\pi} f_\tau d\tau + \iint\limits_{\theta_0\le\tau_2\le\tau_1\le\theta_0+2\pi} [f_{\tau_2}, f_{\tau_1}] d\tau_1 d\tau_2\right) \\
&= \frac{\nu^2}{2}\iint\limits_{\theta_0\le\tau_2\le\tau_1\le\theta_0+2\pi} [f_{\tau_2}, f_{\tau_1}] d\tau_1 d\tau_2
\end{aligned}$$

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

$$\begin{aligned}
F(2\pi;\theta_0,\nu) &= \nu^2\left(\int_{\theta_0}^{\theta_0+2\pi} b(t) f_t\, dt + \frac{1}{2}\left[\int_{\theta_0}^{t} f_\tau d\tau, f_t\right] dt\right) + O(\nu^3) \\
&= -\pi\nu^2 f_0(q_0) + O(\nu^3) \qquad (16.31)
\end{aligned}$$

Notice that the factor π in (16.31) 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 16.10. $F(t;\theta_0,\nu) = F(t;\theta_0+\pi,-\nu)$

Proof. It is a direct consequence of our geodesic equation. Recall that $F(t;\theta_0,\nu) = q(t+\theta_0;\theta_0,\nu)$, is the solution of the system, with initial condition $q(\theta_0;\theta_0,\nu) = q_0$.

Applying the transformation $t \mapsto t+\pi$ and $\nu \to -\nu$ we see that the right hand side of $\dot q$ in (16.28) 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 16.11. *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) $q_{2n+1}(\theta) = -\dfrac{1}{2}\displaystyle\int_{\theta}^{\theta+\pi} \dfrac{dq_{2n+1}}{d\theta}(\tau)d\tau.$

Proof. This is an immediate consequence of Lemma 16.10 and the identity

$$2q_{2n+1}(\theta) = q_{2n+1}(\theta) - q_{2n+1}(\theta+\pi) = -\int_{\theta}^{\theta+\pi} \frac{dq_{2n+1}}{d\theta}(\tau)d\tau.$$

□

We already computed the terms $q_1(\theta)$ and $q_2(\theta)$. To find $q_3(\theta)$ we start by computing the derivative of the map F with respect to θ.

Lemma 16.12. $\dfrac{\partial F}{\partial \theta_0}(2\pi;\theta_0,\nu) = -\pi[f_0, f_{\theta_0}]_{q_0}\nu^3 + O(\nu^4)$

Proof. We stress that, since we are now interested to third order term in ν, we can no more assume that $r(\tau)$ is constant. Differentiating (3.55) with respect to θ gives two terms as follows:

$$\begin{aligned}\frac{\partial F}{\partial \theta_0} &= \frac{\partial}{\partial \theta_0}(q_0 \circ Q_t) = q_0 \circ \frac{\partial}{\partial \theta_0}\left(\overrightarrow{\exp}\int_{\theta}^{\theta+2\pi} V_\tau d\tau\right) \\ &= q_0 \circ \left(Q_{2\pi} \circ V_{\theta_0+2\pi} - V_{\theta_0} \circ Q_{2\pi}\right) \qquad (16.32)\end{aligned}$$

Next let us rewrite

$$\begin{aligned}Q_{2\pi} \circ V_{\theta_0+2\pi} &= Q_{2\pi} \circ V_{\theta_0+2\pi} \circ Q_{2\pi}^{-1} \circ Q_{2\pi} \\ &= \mathrm{Ad}\, Q_{2\pi} \circ V_{\theta_0+2\pi}\end{aligned}$$

so that (16.32) can be rewritten as

$$\frac{\partial F}{\partial \theta_0} = q_0 \circ (\mathrm{Ad}\, Q_{2\pi} \circ V_{\theta_0+2\pi} - V_{\theta_0}) \circ Q_{2\pi} \qquad (16.33)$$

Thanks to Lemma 16.10 we can write

$$Q_{2\pi} = \mathrm{Id} - \pi\nu^2 f_0 + O(\nu^3) \qquad (16.34)$$

that implies the following asymptotics for the action of its adjoint by (6.17)

$$\mathrm{Ad}\, Q_{2\pi} = \mathrm{Id} - \pi\nu^2 \mathrm{ad}\, f_0 + O(\nu^3)$$

We are left to compute the asymptotic expansion of (16.33). 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

$$r(t) = \nu + \nu^3\int_{\theta_0}^{t} a(\tau)d\tau + O(\nu^4) \qquad (16.35)$$

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 gets

$$\begin{aligned}\mathrm{Ad}\,Q_{2\pi}\circ V_{\theta_0+2\pi}-V_{\theta_0}&=(\mathrm{Id}-\pi\nu^2\mathrm{ad}\,f_0+O(\nu^3))\left(\frac{\nu}{1-\nu b}f_{\theta_0}\right)-\left(\frac{\nu}{1-\nu b}f_{\theta_0}\right)+O(\nu^4)\\&=-\pi\nu^2\mathrm{ad}\,f_0(\nu f_{\theta_0})+O(\nu^4)\end{aligned}\tag{16.36}$$

and finally plugging (16.34) and (16.36) into (16.33) one obtains

$$\begin{aligned}\frac{\partial F}{\partial\theta}&=q_0\circ\left(-\pi\nu^2\mathrm{ad}\,f_0(\nu f_{\theta_0})+O(\nu^4)\right)\circ(\mathrm{Id}+O(\nu))\\&=q_0\circ(-\pi\nu^3[f_0,f_{\theta_0}]+O(\nu^4))\end{aligned}$$

□

16.3.2 Asymptotics of the conjugate locus

In this section we finally prove Theorem 16.6, by computing the expansion of the conjugate time $t_c(\theta_0,\nu)$. We know that

$$t_c(\theta_0,\nu)=2\pi+\nu^2 s(\theta_0)+O(\nu^3)$$

By definition of conjugate point, the function $s=s(\theta_0)$ is characterized as the solution of the equation

$$\left.\frac{\partial F}{\partial s}\wedge\frac{\partial F}{\partial\theta}\wedge\frac{\partial F}{\partial\nu}\right|_{(2\pi+\nu^2 s,\theta,\nu)}=0,\tag{16.37}$$

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 16.12 gives

$$\frac{\partial F}{\partial\theta}(2\pi+\nu^2 s;\theta,\nu)=\nu^3(\pi[f_\theta,f_0]+s f_{\theta'})+O(\nu^4)$$

where as usual $f_{\theta'}$ denotes the derivative with respect to θ.

Then, collecting together all these computations, the equation for conjugate points (16.37) can be rewritten as

$$f_\theta\wedge(s f_{\theta'}+\pi[f_\theta,f_0])\wedge f_0=O(\nu)\tag{16.38}$$

Since $f_\theta, f_{\theta'}$ are an orthonormal frame on $\mathcal{D}$ and f_0 is transversal to the distribution, (16.38) 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 16.11 and Lemma one gets

$$\mathrm{Con}(q_0; \theta_0, \nu) = -\pi\nu^2 f_0(q_0) - \frac{\nu^3}{2} \int_{\theta_0}^{\theta_0+\pi} \frac{dq_3}{d\tau} d\tau + \nu^3 s(\theta_0) f_{\theta_0} + O(\nu^4)$$

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

$$\begin{aligned}
\mathrm{Con}(q_0, \theta_0, \nu) &= -\nu^2 f_0(q_0) - \frac{\nu^3}{2} \int_{\theta_0}^{\theta_0+\pi} \pi [f_{\theta_0}, f_0] d\tau + \nu^3 s(\theta_0) f_{\theta_0} \\
&= -\nu^2 f_0(q_0) - \frac{\nu^3}{2} \int_{\theta_0}^{\theta_0+\pi} \pi [f_{\theta_0}, f_0] + s'(t) f_{\theta_0} + s(t) f_{\theta_0'} dt \qquad (16.39)
\end{aligned}$$

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 (16.39) 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

$$\begin{aligned}
\mathrm{Con}(q_0; \theta_0, \nu) &= -\nu^2 f_0(q_0) - \frac{3\nu^3}{2} \pi \int_{\theta_0}^{\theta_0+\pi} a(\tau) f_\tau d\tau + O(\nu^4) \\
&= -\nu^2 f_0(q_0) + \nu^3 \pi (a' f_{\theta_0} - a f_{\theta_0'}) + O(\nu^4)
\end{aligned}$$

16.3.3 Asymptotics of the conjugate lenght

Similarly, we consider conjugate lenght. 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

$$\ell_c(\theta_0,\nu) = \int_{\theta_0}^{\theta_0+2\pi} \frac{r(t)}{1-r(t)Q_t^{\theta_0,\nu}b(t)}dt + \nu^3 s(\theta_0) + O(\nu^4) \tag{16.40}$$

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 (16.35). 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 (16.40).

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

$$r(t) = \nu + \nu^3\int_0^t a(t)dt + O(\nu^4) \tag{16.41}$$

$$r^2(t)Q_t^\theta(\nu)b(t) = (\nu^2+O(\nu^4))\left(\mathrm{Id}+\nu\int_0^t f_\tau d\tau + O(\nu)\right)b(t) \tag{16.42}$$

$$= \nu^2 b(t) + \nu^3\int_0^t f_\tau d\tau b(t) + O(\nu^4) \tag{16.43}$$

$$r^3(t)b(t)^2 = \nu^3 b(t)^2 + O(\nu^4) \tag{16.44}$$

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. □

Chapter 17

The sub-Riemannian heat equation

In this chapter we derive the sub-Riemannian heat equation and we discuss the strictly related question of how to define an intrinsic volume in sub-Riemannian geometry.

17.1 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.

17.1.1 The heat equation in the Riemannian context

Let (M, g) be an oriented Riemannian manifold of dimension n and let ω be a volume form on M, i.e., a never-vanishing n-form on M.[1] The most natural choice for ω is of course the Riemannian volume defined by

$$\omega(X_1, \ldots, X_n) = 1, \text{ where } \{X_1, \ldots, X_n\} \text{ is a local orthonormal frame.}$$

In coordinates if g is represented by a matrix (g_{ij}), we have

$$\omega = \sqrt{\det(g_{ij})}\, dx_1 \wedge \ldots \wedge dx_n.$$

However in the following we write the heat equation for a general volume form that not necessarily coincides with the Riemannian one. This point of view is useful in sub-Riemannian geometry, where a canonical volume exists only in certain cases.

Let ϕ be a quantity (depending on the position q and the time t) subjects to a diffusion process e.g. 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:

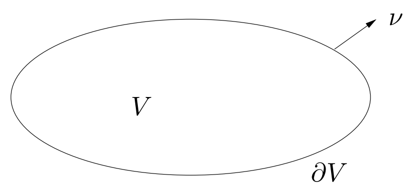
[1] For simplicity here we assume that M is orientable, but since this construction is essentially local, this hypothesis it is not restrictive

(R1) the flux is proportional to minus the gradient of ϕ i.e., normalizing the proportionality constant to one, we assume that

$$\mathbf{F} = -\mathrm{grad}(\phi); \tag{17.1}$$

(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

$$-\frac{d}{dt}\int_V \phi\,\omega = \int_{\partial V} \mathbf{F}\cdot\nu\,\mathrm{dS}. \tag{17.2}$$



Here ν is the external (Riemannian) normal to ∂V and dS is the element of area induced by ω on M, thanks to the Riemannian structure, i.e., $\mathrm{dS} = \omega(\nu,\cdot)$. The quantity $\mathbf{F}\cdot\nu$ is a notation for $g_q(\mathbf{F}(q,t),\nu(q))$.

Applying the Riemannian divergence theorem to (17.2) and using (17.1) we have then

$$-\frac{d}{dt}\int_V \phi\,\omega = \int_{\partial V} \mathbf{F}\cdot\nu\,\mathrm{dS} = \int_V \mathrm{div}_\omega(\mathbf{F})\,\omega = -\int_V \mathrm{div}_\omega(\mathrm{grad}(\phi))\,\omega.$$

By the arbitrarity of V and defining the Riemannian Laplace operator as

$$\triangle\phi = \mathrm{div}_\omega(\mathrm{grad}(\phi)) \tag{17.3}$$

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 div_ω in formula (17.13).

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 $q \in M$,

$$g_q(\mathrm{grad}(\varphi), v) = d\varphi(v), \quad \text{for every } v \in T_qM. \tag{17.4}$$

In coordinates, if g is represented by a matrix (g_{ij}), and calling (g^{ij}) its inverse, we have

$$\mathrm{grad}(\varphi)^i = \sum_{j=1}^n g^{ij}\partial_j\varphi. \tag{17.5}$$

If $\{X_1, \dots, X_n\}$ is a local orthonormal frame for g, we have the useful formula

$$\operatorname{grad}(\varphi) = \sum_{i=1}^{n} X_i(\varphi) X_i. \tag{17.6}$$

Exercise 17.1. Prove that if the Riemannian metric is defined globally via a generating frame $\{X_1, \dots, X_k\}$ with $k \geq n$, in the sense of Section .... then $\operatorname{grad}(\varphi) = \sum_{i=1}^{k} X_i(\varphi) X_i$.

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

$$L_X \omega = \operatorname{div}_\omega(X)\omega.$$

Now using $d\omega = 0$ and the Cartan formula we have that $L_X\omega = i_X d\omega + d(i_X\omega) = d(i_X\omega)$. Hence the divergence of a vector field X can be defined by

$$d(i_X\omega) = \operatorname{div}_\omega(X)\omega. \tag{17.7}$$

In coordinates, if $\omega = h(x)dx^1 \wedge \dots dx^n$ we have

$$\operatorname{div}_\omega(X) = \frac{1}{h(x)} \sum_{i=1}^{n} \partial_i (h(x) X^i). \tag{17.8}$$

Remark 17.2. Notice that to define the divergence of a vector field it is not necessary a Riemannian structure, but only a volume form.

If we put together formula 17.5 and formula 17.8, with $X = \operatorname{grad}(\varphi)$ we get the well known expression

$$\triangle(\varphi) = \operatorname{div}_\omega(\operatorname{grad}(\varphi)) = \frac{1}{h(x)} \sum_{i,j=1}^{n} \partial_i (h(x) g^{ij} \partial_j \varphi). \tag{17.9}$$

Combining formula 17.6 with the property $\operatorname{div}(aX) = a \operatorname{div}(X) + X(a)$ where X is a vector field and a a function, we get

$$\triangle(\varphi) = \sum_{i=1}^{n} \left(X_i^2 \varphi + \operatorname{div}_\omega(X_i) X_i(\varphi) \right) \quad \text{where } \{X_1, \dots X_n\} \text{ is a local orthonormal frame.} \tag{17.10}$$

Similarly, defining the Riemannian structure via a generating frame we get

$$\triangle(\varphi) = \sum_{i=1}^{k} \left(X_i^2 \varphi + \operatorname{div}_\omega(X_i) X_i(\varphi) \right) \quad \text{where } \{X_1, \dots X_k\},\ k \geq n, \text{ is a generating frame} \tag{17.11}$$

Remark 17.3. Notice that the choice of the volume form does not affect the second order terms, but only the first order ones.

When $\triangle$ is built with respect to the Riemannian volume form, it is called the Laplace-Beltrami operator.

17.1.2 The heat equation in the sub-Riemannian context

Let M be a sub-Riemannian manifold of dimension n. Let $\mathcal{D}$ be the associated set of horizontal vector fields and g_q the corresponding metric on the distribution $\mathcal{D}_q$.

As in the Riemannian case, we assume by simplicity that M is oriented and we assume that a volume form ω has been assigned on M. In Section **??** we show that, in the equiregular case, the sub-Riemannian structure induces, canonically, a volume form on M. For the moment we assume that the volume form is assigned independently of the sub-Riemannian structure.

As in the previous section, we denote by ϕ the quantity subject to the diffusion process, by $\mathbf{F}$ the corresponding flux, 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 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.

To derive the heat equation in the Riemannian case, we have used the following ingredients that are not directly available in the sub-Riemannian context:

- the Riemannian gradient;
- the Riemannian normal to ∂V, and the inner product to define the conservation 17.2;
- the Riemannian divergence theorem.

Hence the standard Riemannian construction fails in the sub-Riemannian context and we have to reason in a different way to derive the heat equation. Let us analyse one by one the ingredients above and let us see how to generalise them in sub-Riemannian geometry.

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 $q \in M$,

$$g_q(\mathrm{grad}_H(\varphi), v) = d\varphi(v), \quad \text{for every } v \in \mathcal{D}_q M. \tag{17.12}$$

If $\{X_1, \ldots, X_k\}$ is a generating frame then

$$\mathrm{grad}_H(\varphi) = \sum_{i=1}^{k} X_i(\varphi) X_i.$$

The postulate (SR1) is then written as

$$\mathbf{F} = -\mathrm{grad}_H(\phi).$$

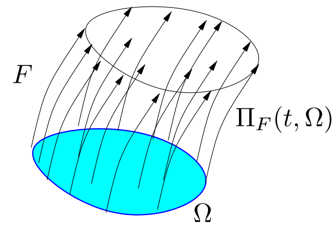


Figure 17.1:

The conservation of the heat

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 17.4. *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 be a (possible 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 17.1):*

$$\Pi_F(t,\Omega) = \{P_{0,t}(\Omega) \mid s \in [0,t]\}.$$

Then

$$\frac{d}{dt}\bigg|_{t=0} \int_{\Pi_F(t,\Omega)} \omega = \int_\Omega i_F\big|_{t=0}\,\omega.$$

With the notation of this Lemma, the postulate (SR2) is 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 17.7 and using that $\mathbf{F} = -\mathrm{grad}_H\phi$ we have

$$\int_{\partial V} i_{\mathbf{F}}\,\omega = \int_V d(i_{\mathbf{F}}\,\omega) = \int_V \mathrm{div}_\omega(\mathbf{F})\,\omega = -\int_V \mathrm{div}(\mathrm{grad}_H(\phi))\,\omega.$$

By the arbitrarity of V and defining

$$\triangle_H\phi = \mathrm{div}_\omega(\mathrm{grad}_H(\phi)), \tag{17.13}$$

we get the sub-Riemannian heat equation

$$\frac{\partial}{\partial t}\phi(q,t) = \triangle_H\phi(q,t).$$

Definition 17.5. 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.*

When it is possible to construct a volume from the sub-Riemannian structure, then the corresponding sub-Riemannian Laplacian is called the *intrinsic sub-Laplacian.* The construction of a canonical volume form in a sub-Riemannian manifold is the purpose of Section **??**. Here let us just remark that in the case of left-invariant structures on Lie groups, a canonical volume can be built naturally from the group structure. This will be done in Section 17.2 for the Heisenberg group.

17.1.3 Few properties of the sub-Riemannian Laplacian: the Hörmander theorem and the existence of the heat kernel

The same computation of the Riemannian case provides the following expression for the sub-Riemannian Laplacian,

$$\triangle_H(\phi) = \sum_{i=1}^{k} \left(X_i^2 \phi + \mathrm{div}_\omega(X_i) X_i(\phi) \right) \quad \text{where } \{X_1, \ldots X_k\}, \text{ is a generating frame.} \tag{17.14}$$

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 17.6 (Hörmander)**.** *Let $Y_0, Y_1 \ldots Y_k$ 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 $\mathcal{C}^\infty$, then φ is $\mathcal{C}^\infty$ in Ω.*

Remark 17.7. Notice that elliptic operators with $\mathcal{C}^\infty$ coefficients are hypoelliptic. The heat operator $\partial_t - \Delta$, where Δ is the standard Laplacian in $\mathbb{R}^n$ is not elliptic (since the matrix of coefficients of the second order derivatives in $\mathbb{R}^{n+1}$ has one zero eigenvalue), but it is hypoelliptic since $\partial_{x_1}, \ldots, \partial_{x_n}, \partial_t$ are Lie Bracket generating in $\mathbb{R}^{n+1}$.

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 $(\partial_t - \Delta_H)\varphi = 0$ exists in distributional sense in an open set Ω of $\mathbb{R} \times M$, then, being $0 \in \mathcal{C}^\infty$, it follows that φ is $\mathcal{C}^\infty$ in Ω.

A standard result for the existence of a solution in $L^2(M, \omega)$ is given by the following theorem. See for instance [**?**].

Theorem 17.8. *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*

$$\begin{cases} (\partial_t - \triangle)\phi = 0 \\ \phi(q, 0) = \phi_0(q) \in L^2(M, \omega), \end{cases} \tag{17.15}$$

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 (17.15) is indeed $\mathcal{C}^\infty$ in $]0, \infty[\times M$.

The discussion of the theory of self-adjoint operators is out of the purpose of this book. However the essential self-adjointness of Δ_H is guaranteed by the completeness of the sub-Riemannian manifold as metric space. This condition guarantees also the existence of the solution to the Cauchy problem in the form of a convolution kernel.

Theorem 17.9 (Strichartz)**.** *Consider a sub-Riemannian manifold that is complete as metric space. Let ω be a volume on M. Then Δ_H is essentially self-adjoint on $L^2(M,\omega)$. Moreover the unique solution to the Cauchy problem*

$$\begin{cases} (\partial_t - \triangle_H)\phi = 0 \\ \phi(q,0) = \phi_0(q) \in L^2(M,\omega), \end{cases} \tag{17.16}$$

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 $K_t(q,\cdot) \in L^2(M,\omega)$.

Typical cases in which the sub-Riemannian manifold is complete are let-invariant structure on Lie groups, sub-Riemannian structures obtained as restriction of complete Riemannian structures, sub-Riemannian structures defined in $\mathbb{R}^n$ having as generating frame a set of sublLinear vector fields.

Let us just remark that if the sub-Riemannian structure is not Lie-bracket generated,[2] then in general the operator is not hypoelliptic and the heat evolution does not smooth the initial condition.

Consider for example the operator $L = \partial_x^2 + \partial_y^2$ on $\mathbb{R}^3$. This operator is not obtained from Lie-bracket generating vector fields. Consider the corresponding heat operator $\partial_t - L$ 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 $\partial_t - L = 0$ on $[0,\infty] \times \mathbb{R}^2$, then $\psi(x,y,t)\theta(z)$ is a solution of the heat equation in $[0,\infty] \times \mathbb{R}^3$, where θ is the Heaviside function.

17.2 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 to build in a canonical way a volume form and hence the sub-Riemannian Laplacian. Moreover this point of view permits to look for a simplified version of the heat kernel using the group law.

17.2.1 The Heisenberg group as a group of matrices

The Heisenberg group H_2 can be seen as the 3-dimensional group of matrices

$$H_2 = \left\{ \begin{pmatrix} 1 & x & z + \frac{1}{2}xy \\ 0 & 1 & y \\ 0 & 0 & 1 \end{pmatrix} \mid x,y,z \in \mathbb{R} \right\}$$

endowed with the standard matrix product. H_2 is indeed $\mathbb{R}^3$, endowed with the group law

$$(x_1,y_1,z_1)\cdot(x_2,y_2,z_2) = \left(x_1 + x_2, y_1 + y_2, z_1 + z_2 + \frac{1}{2}(x_1y_2 - x_2y_1) \right).$$

[2] i.e. a proto-sub-Riemannian structure

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 H_2 is $\{p_1, p_2, k\}$ where

$$p_1 = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad p_2 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} \quad k = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \tag{17.17}$$

They satisfy the following commutation rules: $[p_1, p_2] = k$, $[p_1, k] = [p_2, k] = 0$, hence H_2 is a 2-step nilpotent group.

Remark 17.10. Notice that if one write an element of the algebra as $xp_1 + yp_2 + zk$, one has that

$$\exp(xp_1 + yp_2 + zk) = \begin{pmatrix} 1 & x & z+\frac{1}{2}xy \\ 0 & 1 & y \\ 0 & 0 & 1 \end{pmatrix}. \tag{17.18}$$

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 H_2 is the one having as generating frame:

$$X_1(g) = gp_1, \quad X_2(g) = gp_2.$$

With a straightforward computation one get the following coordinate expression for the generating frame:

$$X_1 = \partial_x - \frac{y}{2}\partial_z, \qquad X_2 = \partial_y + \frac{x}{2}\partial_z,$$

that we already met several times in the previous chapters.

Let L_g (reap. R_g) be the left (resp. right) multiplication on H_2:

$$L_g : H_2 \ni h \mapsto gh \ \ (\text{resp. } R_g : H_2 \ni h \mapsto hg).$$

Exercise Prove that, up to a multiplicative constant, there exist one and only one 3-form dh_L on H_2 which is left-invariant, i.e. such that $L_g^* dh = dh$ 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 called the left and right Haar measures. Since they coincide up to a 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

$$dh_L(X_1, X_2, [X_1, X_2]) = dh_R(X_1, X_2, [X_1, X_2]) = 1. \tag{17.19}$$

The 3-form obtained in this way coincide with the Lebesgue measure and in the following we call it simply the "Haar measure"

$$dh = dx \wedge dy \wedge dz.$$

Exercise Prove that the two conditions (17.19) are invariant by change of the orthonormal frame.

17.2.2 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,

$$\triangle_H(\phi) = \big(X_1^2 + X_2^2 + \operatorname{div}_\omega(X_1)X_1 + \operatorname{div}_\omega(X_2)X_2\big)\phi. \tag{17.20}$$

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

$$\triangle_H(\phi) = (X_1)^2 + (X_2)^2 = (\partial_x - \frac{y}{2}\partial_z)^2 + (\partial_y + \frac{x}{2}\partial_z)^2. \tag{17.21}$$

The heat equation on the Heisenberg group is then

$$\triangle_H(\phi) = \big((\partial_x - \frac{y}{2}\partial_z)^2 + (\partial_y + \frac{x}{2}\partial_z)^2\big)\phi(x, y, z, t) = \partial_t\phi(x, y, z, t).$$

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

$$\begin{cases} (\partial_t - \triangle_H)\phi = 0 \\ \phi(q, 0) = \phi_0(q) \in L^2(\mathbb{R}^3, dh) \end{cases} \tag{17.22}$$

can be expressed as

$$\phi(q, t) = \int_{\mathbb{R}^3} K_t(q, \bar{q})\phi_0(\bar{q})dh(\bar{q}). \tag{17.23}$$

The existence of a heat kernel that is smooth, positive and symmetric is guaranteed by Theorem 17.9 since the Heisenberg group (as sub-Riemannian structure) is complete.

The construction of the explicit expression of the heat kernel on the Heisenberg group was an important achievement of the end of the seventies. 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 3 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 have that for $K_t(q,\bar{q}) = K_t(p\cdot q, p\cdot\bar{q})$ for every $p \in H_2$. Taking $p = q^{-1}$ we have that $K_t(q,\bar{q}) = K_t(0, q^{-1}\bar{q})$ hence we can write

$$K_t(q,\bar{q}) = p_t(q^{-1}\cdot\bar{q}) = p_t(\bar{x}-x, \bar{y}-y, \bar{z}-z) = p_t(x-\bar{x}, y-\bar{y}, z-\bar{z}),$$

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 17.10). In the new variables we have that the Haar measure is $dh = dx \wedge dy \wedge dw$. The generating frame and the sub-Riemannian Laplacian become

$$X_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = \partial_x \tag{17.24}$$

$$X_2 = \begin{pmatrix} 0 \\ 1 \\ x \end{pmatrix} = \partial_y + x\partial_w \tag{17.25}$$

$$\triangle_H(\phi) = (X_1)^2 + (X_2)^2 = \partial_x^2 + (\partial_y + x\partial_w)^2. \tag{17.26}$$

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

$$\begin{cases} \partial_t\varphi(x,y,w,t) = \triangle_H(\varphi(x,y,w,t)) = \big(\partial_x^2 + (\partial_y + x\partial_w)^2\big)\varphi(x,y,w,t) \\ \varphi(x,y,w,0) = \varphi_0(x,y,w) \in L^2(\mathbb{R}^3, dh) \end{cases} \tag{17.27}$$

where $\varphi(x,y,w,t) = \phi(x,y,w-\frac{1}{2}xy)$.

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

$$\begin{cases} \partial_t\hat{\varphi}(x,\mu,\nu,t) = \big(\partial_x^2 - (\mu+\nu x)^2\big)\hat{\varphi}(x,\mu,\nu,t) \\ \hat{\varphi}(x,\mu,\nu,0) = \hat{\varphi}_0(x,\mu,\nu). \end{cases} \tag{17.28}$$

By making the change of variable $x \to \theta$, where $\mu + \nu x = \nu\theta$, i.e., $\theta = x + \frac{\mu}{\nu}$ we get:

$$\begin{cases} \partial_t\bar{\varphi}^{\mu,\nu}(\theta,t) = \big(\partial_\theta^2 - \nu^2\theta^2\big)\bar{\varphi}^{\mu,\nu}(\theta,t) \\ \bar{\varphi}^{\mu,\nu}(\theta,0) = \bar{\varphi}_0^{\mu,\nu}(\theta), \end{cases} \tag{17.29}$$

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 17.11. *The solution of the Cauchy problem for the evolution of the heat in an harmonic potential, i.e.*

$$\begin{cases} \partial_t \psi(\theta,t) = \big(\partial_\theta^2 - \nu^2\theta^2\big)\psi(\theta,t) \\ \psi(\theta,0) = \psi_0(\theta) \in L^2(\mathbb{R}, d\theta) \end{cases} \tag{17.30}$$

can be written in the form of a convolution kernel

$$\psi(\theta,t) = \int_{\mathbb{R}} Q_t^\nu(\theta,\bar\theta)\psi_0(\bar\theta)d\bar\theta.$$

where

$$Q_t^\nu(\theta,\bar\theta) := \sqrt{\frac{\nu}{2\pi\sinh(2\nu t)}}\exp\left(-\frac{1}{2}\frac{\nu\cosh(2\nu t)}{\sinh(2\nu t)}(\theta^2+\bar\theta^2) + \frac{\nu\theta\bar\theta}{\sinh(2\nu t)}\right). \tag{17.31}$$

Remark 17.12. In the case $\nu = 0$ we interpret $Q_t^0(\theta,\bar\theta)$ as

$$\lim_{\nu\to 0} Q_t^\nu(\theta,\bar\theta) = \frac{1}{\sqrt{4\pi t}}\exp[-\frac{(\theta-\bar\theta)^2}{4t}]. \tag{17.32}$$

Proof. For $\nu = 0$, equation (17.30) is the standard heat equation on $\mathbb{R}$ and the heat kernel is given by formula (17.32). See for instance [**?**]. In the following we assume $\nu \neq 0$. The eigenvalues and the eigenfunctions of the operator $\partial_\theta^2 - \nu^2\theta^2$ on $\mathbb{R}$ are (see Appendix **??**)

$$\begin{aligned} E_j &= -2\nu(j+1/2) \\ \varphi_j^\nu(\theta) &= \frac{1}{\sqrt{2^j j!}}\left(\frac{\nu}{\pi}\right)^{\frac14}\exp(-\frac{\nu\theta^2}{2})H_j(\sqrt{\nu}\,\theta) \end{aligned} \tag{17.33}$$

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 $\{\varphi_j^\nu\}_{j\in\mathbb{N}}$ an orthonormal frame of $L^2(\mathbb{R})$, we can write

$$\psi(\theta,t) = \sum_j C_j(t)\varphi_j^\nu(\theta).$$

Using equation (17.30), we obtain that

$$C_j(t) = C_j(0)\exp(tE_j)$$

where $C_j(0) = \int_{\mathbb{R}}\varphi_j^\nu(\bar\theta)\psi_0(\bar\theta)\,d\bar\theta$. Hence

$$\psi(\theta,t) = \int_{\mathbb{R}} Q_t^\nu(\theta,\bar\theta)\psi_0(\bar\theta)\,d\bar\theta$$

where

$$Q_t^\nu(\theta,\bar\theta) = \sum_j \varphi_j^\nu(\theta)\varphi_j^\nu(\bar\theta)\exp(tE_j).$$

After some algebraic manipulations and using the Mehler formula for Hermite polynomials

$$\sum_j \frac{H_j(\theta)H_j(\bar\theta)}{2^j j!}(w)^j = (1-w^2)^{-\frac{1}{2}}\exp\left(\frac{2\theta\bar\theta w-(\theta^2+\bar\theta^2)w^2}{1-w^2}\right), \quad \forall\, w\in\mathbb{R}$$

with $\theta\to\sqrt{\nu}\theta$, $\bar\theta\to\sqrt{\nu}\bar\theta$, $w\to\exp(-2\nu t)$, one get formula (17.31). □

Using Theorem 17.11 we can write the solution to 17.30 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

$$\phi(x,y,z,t)=\varphi(x,y,z+\frac{1}{2}xy)=\int_{\mathbb{R}^3}K_t(x,y,z,\bar x,\bar y,\bar z)\phi_0(\bar x,\bar y,\bar z)d\bar x\,d\bar y\,d\bar z.$$

where

$$K_t(x,y,z,\bar x,\bar y,\bar z)=\frac{1}{(2\pi)^2}\int_{\mathbb{R}}\int_{\mathbb{R}}Q_t^\nu(x+\frac{\mu}{\nu},\bar x+\frac{\mu}{\nu})e^{i\mu(y-\bar y)}e^{i\nu(z-\bar z+\frac{1}{2}(xy-\bar x\bar y))}d\mu\,d\nu.$$

Setting $\bar x,\bar y,\bar z$ to zero and after some algebraic manipulations we get for the fundamental solution

$$p_t(x,y,z)=\frac{1}{(2\pi t)^2}\int_{\mathbb{R}}\frac{2\tau}{\sinh(2\tau)}\exp\left(-\frac{\tau(x^2+y^2)}{2t\tanh(2\tau)}\right)\cos(2\frac{z\tau}{t})d\tau. \tag{17.34}$$

The integral representation (17.34) can be computed explicitly on the origin and on the z axis. Indeed we have

$$K_t(0,0,0;0,0,0)=p_t(0,0,0)=\frac{1}{16t^2} \tag{17.35}$$

$$K_t(0,0,0;0,0,z)=p_t(0,0,z)=\frac{1}{8t^2\left(1+\cosh\left(\frac{\pi z}{t}\right)\right)}=\frac{1}{4t^2}\exp\left(-\frac{d^2(0,0,0;0,0,z)}{4t}\right)f(t) \tag{17.36}$$

In the last equality we have used the fact that for the Heisenberg group $d(0,0,0;0,0,z) = \sqrt{4\pi z}$. Here $f(t)$ is a smooth function of t such that $f(0) = 1$ (here $z \neq 0$ is fixed). A more detailed analysis permits to get for every fixed (x,y,z) such that $x^2 + y^2 \neq 0$

$$K_t(0,0,0;x,y,z) = p_t(x,y,z) = \frac{C + O(t)}{t^{3/2}} \exp\left(-\frac{d^2(0,0,0;x,y,z)}{4t}\right). \tag{17.37}$$

Notice that the asymptotics (17.35), (17.36), (17.37) 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

$$K_t(0,0,0;x,y,z) = \frac{1}{(4\pi t)^{n/2}} \exp\left(-\frac{x^2 + y^2 + z^2}{4t}\right). \tag{17.38}$$

Comparing (17.38) with (17.35), (17.36), (17.37), 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. While on all the other points it is similar to to the one in an Euclidean space of dimension 3. 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 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 [**?**]. 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 $\bar{q}$ close enough to q we have $K_t(q,\bar{q}) = \frac{1+O(t)}{(4\pi t)^{n/2}} \exp\left(-\frac{d^2(q,\bar{q})}{4t}\right)$.

Index

Bibliography

[1] A. Agrachev, D. Barilari, and U. Boscain. On the Hausdorff volume in sub-Riemannian geometry. *Calc. Var. and PDE's*, 43(3-4):355–388, 2012.

[2] A. Agrachev, U. Boscain, J.-P. Gauthier, and F. Rossi. The intrinsic hypoelliptic Laplacian and its heat kernel on unimodular Lie groups. *J. Funct. Anal.*, 256(8):2621–2655, 2009.

[3] A. A. Agrachev. Exponential mappings for contact sub-Riemannian structures. *J. Dynam. Control Systems*, 2(3):321–358, 1996.

[4] A. A. Agrachev and Y. L. Sachkov. *Control theory from the geometric viewpoint*, volume 87 of *Encyclopaedia of Mathematical Sciences*. Springer-Verlag, Berlin, 2004. Control Theory and Optimization, II.

[5] D. Barilari, U. Boscain, and J.-P. Gauthier. On 2-step, corank 2 sub-Riemannian metrics. *SIAM Journal of Control and Optimization*, 50(1):559–582, 2012.

[6] A. Bellaïche. The tangent space in sub-Riemannian geometry. In *Sub-Riemannian geometry*, volume 144 of *Progr. Math.*, pages 1–78. Birkhäuser, Basel, 1996.

[7] U. Boscain and J.-P. Gauthier. On the spherical Hausdorff measure in step 2 corank 2 sub-Riemannian geometry. *arXiv:1210.2615 [math.DG], Preprint*, 2012.

[8] U. Boscain and C. Laurent. The Laplace-Beltrami operator in almost-Riemannian geometry. *arXiv:1105.4687v1 [math.SP], Preprint*, 2011.

[9] A. Bressan and B. 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.

[10] A. Bressan and B. 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.

[11] 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, Warsaw, 1984. PWN.

[12] D. Burago, Y. Burago, and S. Ivanov. *A course in metric geometry*, volume 33 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2001.

[13] M. Gromov. *Metric structures for Riemannian and non-Riemannian spaces*. Modern Birkhäuser Classics. Birkhäuser Boston Inc., Boston, MA, english edition, 2007. Based on the 1981 French original, With appendices by M. Katz, P. Pansu and S. Semmes, Translated from the French by Sean Michael Bates.

[14] V. Jurdjevic. *Geometric control theory*, volume 52 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 1997.

[15] R. Montgomery. *A tour of subriemannian geometries, their geodesics and applications*, volume 91 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2002.

[16] R. Montgomery. *A tour of subriemannian geometries, their geodesics and applications*, volume 91 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2002.

[17] R. S. Strichartz. Sub-Riemannian geometry. *J. Differential Geom.*, 24(2):221–263, 1986.