Scuola Internazionale Superiore di Studi Avanzati - Trieste

The curvature of optimal control problems with applications to sub-Riemannian geometry

Supervisor:
Prof. Andrei A. Agrachev

Candidate:
Luca Rizzi

Thesis submitted for the degree of
"Doctor Philosophiæ"

Academic Year 2013/2014

SISSA - Via Bonomea 265 - 34136 TRIESTE - ITALY

Contents

Abstract

Optimal control theory is an extension of the calculus of variations, and deals with the optimal behaviour of a system under a very general class of constraints. This field has been pioneered by the group of mathematicians led by Lev Pontryagin in the second half of the 50s (see [61]) and nowadays has countless applications to the real worlds (robotics, trains, aerospace, models for human behaviour, human vision, image reconstruction, quantum control, motion of self-propulsed micro-organism, ...).

In this thesis we are mainly concerned with *affine optimal control problems*, that we now briefly introduce. Let M be a smooth manifold, representing the space of all possible states of a system subject to *non-holonomic constraints.* Its time evolution is described by an *admissible curve* $\gamma : [0,T] \to M$, namely a Lipschitz curve that satisfies

$$\frac{d\gamma}{dt} = f_0(\gamma(t)) + \sum_{i=1}^{k} u_i(t) f_i(\gamma(t)) \tag{1}$$

for some *control functions* $u_1, \dots, u_k : [0,T] \to \mathbb{R}$, where f_i, for $i = 0, \dots, k$ are smooth vector fields on M. The field f_0 is usually referred to as the *drift* field, the field $f_1, \dots, f_k$ are the *controllable fields.* Under mild regularity assumptions, and for any choice of the initial state $\gamma(0) = x_0$ and controls $u_1, \dots, u_k$, Eq. (1) has a unique solution that describes the evolution of the system for the given controls.

Eq. (1) restricts the class of admissible curves to those with velocity contained in the affine distribution $f_0 \oplus \operatorname{span}\{f_1, \dots, f_k\}$. Among all the admissible trajectories $\gamma : [t_0, t_1] \to M$ connecting two points x_0 and x_1, we are interested in those that minimize a *cost functional*, given in the form of an integral

$$J = \int_{t_0}^{t_1} L(x_1, \dots, x_n, u_1, \dots, u_k) dt,$$

where L is the *Lagrangian function*, satisfying suitable assumptions, while t_0 and t_1 are fixed. The admissible trajectories that minimize the cost (when they exist) are called *optimal trajectories* or simply *geodesics.*

Within this general framework, many heterogeneous geometrical structures (e.g. linear quadratic optimal control systems, (sub)-Riemannian or (sub)-Finsler structures) are included in the same category.

The main contribution of this thesis is the investigation of the concept of curvature for optimal control problems and its applications to sub-Riemannian geometry. The exposition is

organised in 4 chapters. Each chapter contains a more detailed mathematical and bibliographical description of the single problems, and begins with an ad-hoc introductions, meant to be self-contained, collecting all the main results. In order to give a flavour of the main topics, we now provide a brief summary, followed by a more detailed introduction.

Chapter 1. Curvature of optimal control problems. The first topic is a new definition of curvature of an optimal control problem, introduced as a fundamental invariant associated with the cost function along a given extremal. In this chapter we introduce the *geodesic growth vector*, a set of dimension-like invariants associated with a geodesic, that encodes the geometrical structure in the direction of the given extremal. Moreover, we define a generalized curvature operator, associated with a fixed geodesic, that extends the Riemannian concept of sectional curvature of the 2-planes containing the direction of the geodesic (see Theorems 1.A–1.B). Then we investigate the relation of the concepts introduced above with the symplectic invariants of the so-called *Jacobi curves*, namely curves of Lagrangian subspaces that extend the classical Jacobi fields to this more general setting.

At this point we specialise the discussion to the sub-Riemannian setting. In particular, we investigate the relation between the sub-Laplacian of the cost function and the curvature of the sub-Riemannian manifold (see Theorem 1.C). Then we define the class of *slow growth* distributions (see Sec. 1.5.5), and we specify Theorem 1.C to these (see Theorem 1.85).

Finally, we obtain a new dimensional-type invariant, that we call *geodesic dimension*, different from the Hausdorff and the topological ones, and we study its relation with the contraction of measures along geodesics (see Theorem 1.D).

Chapter 2. Comparison theorems in sub-Riemannian geometry. In the second chapter we investigate the interplay between curvature and optimality of sub-Riemannian geodesics. In particular, we obtain comparison theorems for the occurrence of conjugate points along the geodesics of a general sub-Riemannian manifold (see Theorems 2.A–2.B). In this unified setting, linear quadratic optimal control problems play the role of *constant curvature models*. As an application of these results, we prove a sub-Riemannian version of the celebrated Bonnet-Myers theorem (see Theorem 2.C) and obtain some new results on the existence of conjugate points for 3D-left invariant sub-Riemannian structures (see Theorem 2.D).

Chapter 3. On conjugate times of LQ optimal control problems. Motivated by the comparison theorems of Chapter 2, we investigate the well known class of linear quadratic optimal control systems (LQ in the following). We address the classical problem of finding necessary and sufficient condition for the existence of conjugate times, and we solve it in terms of the spectrum the Hamiltonian vector field of the LQ problem (see Theorem 3.A). Surprisingly, to our best knowledge, this classical question had remained unanswered until now.

Chapter 4. A formula for Popp's volume in sub-Riemannian geometry. In the last part of this thesis, we investigate a possible definition of intrinsic volume for a sub-Riemannian manifold. This is related with Chapter 1, since a canonical volume on sub-Riemannian manifold defines a canonical Laplace operator. In particular, we discuss Popp's volume, an intrinsic measure introduced by Montgomery (see [56]). In this chapter we obtain an explicit formula for Popp's volume (see Theorem 4.A) and we prove that, under some hypotheses, it is the unique smooth volume preserved by sub-Riemannian isometries (see Propositions 4.B–4.C).

The research presented in this thesis appears in the following preprints and publications, originating from various collaborations started during my PhD studies:

- A. Agrachev, D. Barilari, L. Rizzi, *The curvature: a variational approach*, arXiv:1306.5318
- D. Barilari, L. Rizzi, *Comparison theorems for conjugate points in sub-Riemannian geometry*, arXiv:1401.3193
- A. Agrachev, L. Rizzi, P. Silveira, *On conjugate times of LQ optimal control problems*, arXiv:1311.2009
- D. Barilari, L. Rizzi, *A formula for Popp's volume in sub-Riemannian geometry*, published in Analysis and Geometry in Metric Spaces, Volume 1 (2012), arXiv:1211.2325

Other material that is related with these topics and that has been part of the research developed during the PhD studies, but is not presented here, is contained in the following preprints in preparation:

- *Sub-Riemannian curvature for contact manifolds*, in preparation (with A. Agrachev, D. Barilari, P. Lee)
- *On the Measure Contraction Property of Sub-Riemannian manifolds and the MCP conjecture for Carnot groups* in preparation (with D. Barilari)
- *Counting geodesics on Carnot groups* in preparation (with A. Lerario)

Introduction

1 The curvature of optimal control problems

The curvature discussed here is a rather far going generalization of the Riemannian sectional curvature. We define it for a wide class of optimal control problems: a unified framework including geometric structures such as Riemannian, sub-Riemannian, Finsler and sub-Finsler structures; a special attention is paid to the sub-Riemannian (or Carnot–Carathéodory) metric spaces. Our construction of the curvature is direct and naive, and it is similar to the original approach of Riemann. Surprisingly, it works in a very general setting and, in particular, for *all* sub-Riemannian spaces.

The main idea is that the curvature is related with the distortion of the cost along the extremals. To fix the ideas, let's see how we can develop this concept starting from Riemannian geometry, the classical realm of curvature. In his seminal paper [62], Riemann thought at the curvature as a quantity that measures "how much the geometrical structure is different from the flat one". In a more modern language, this concept can be expressed very elegantly as follows. Let M be a Riemannian manifold. Consider two unit speed geodesics γ_v, γ_w, with initial tangent vector v and w, and let $\mathsf{d} : M \times M \to \mathbb{R}$ be the Riemannian distance. Then

$$\mathsf{d}^2(\gamma_v(t), \gamma_w(t)) = 2(1-\cos\theta)t^2 \left(1 - \tfrac{1}{3}\operatorname{Sec}(v,w)\cos^2(\theta/2)t^2\right) + O(t^6), \tag{2}$$

where $\operatorname{Sec}(v,w)$ is the sectional curvature of the plane spanned by v and w, and θ is the Riemannian angle between the two vectors (see [69]). The same naive idea can be extended to the more general realm of optimal control problems where, at least in principle, the Riemannian geodesics are replaced by optimal trajectories, and the squared distance is replaced by the value (or cost) function.

Still, some serious difficulties arise. Riemannian extremals are parametrized by their initial vector, and this is no longer true in the more general setting (or also in the closer sub-Riemannian one). More importantly, the cost function may be non-smooth on the diagonal. For example, in the sub-Riemannian case, the cost function is essentially the squared distance $\mathsf{d}^2 : M \times M \to \mathbb{R}$, that is *never* smooth on the diagonal. Thus, a naive Taylor expansion in the spirit of Eq. (2) is not possible.

Now that we have introduced the flavour of the main idea, we explain in more detail the nature of our curvature by describing the case of a contact sub-Riemannian structure. Then we move to the general construction.

Let M be an odd-dimensional Riemannian manifold endowed with a contact vector distribution $\mathscr{D} \subset TM$. Given $x_0, x_1 \in M$, the contact sub-Riemannian distance $\mathsf{d}(x_0, x_1)$ is the infimum of the lengths of Legendrian curves connecting x_0 and x_1. Recall that Legendrian curves are just integral curves of the distribution $\mathscr{D}$. The metric d is easily realized as the limit of a family of Riemannian metrics d^ε as $\varepsilon \to 0$. To define d^ε we start from the original Riemannian structure on M, keep fixed the length of vectors from $\mathscr{D}$ and multiply by $\frac{1}{\varepsilon}$ the length of the orthogonal to $\mathscr{D}$ tangent vectors to M. It is easy to see that $\mathsf{d}^\varepsilon \to \mathsf{d}$ uniformly on compacts in $M \times M$ as $\varepsilon \to 0$.

The distance converges, what about the curvature? Let ω be a contact differential form that annihilates $\mathscr{D}$, i.e. $\mathscr{D} = \omega^\perp$. Given $v_1, v_2 \in T_xM,\ v_1 \wedge v_2 \neq 0$, we denote by $\mathrm{Sec}^\varepsilon(v_1 \wedge v_2)$ the sectional curvature for the metric d^ε and section $\mathrm{span}\{v_1, v_2\}$. It is not hard to show that $\mathrm{Sec}^\varepsilon(v_1 \wedge v_2) \to -\infty$ if $v_1, v_2 \in \mathscr{D}$ and $d\omega(v_1, v_2) \neq 0$. Moreover, $\mathrm{Ric}^\varepsilon(v) \to -\infty$ as $\varepsilon \to 0$ for any nonzero vector $v \in \mathscr{D}$, where Ric^ε is the Ricci curvature for the metric d^ε. On the other hand, the distance between x and the conjugate locus of x tends to 0 as $\varepsilon \to 0$ and $\mathrm{Sec}^\varepsilon(v_1 \wedge v_2)$ tends to $+\infty$ for some $v_1, v_2 \in T_xM$, as well as $\mathrm{Ric}^\varepsilon(v)$ for some $v \in T_xM$.

What about the geodesics? For any $\varepsilon > 0$ and any $v \in T_xM$ there is a unique geodesic of the Riemannian metric d^ε that starts from x with velocity v. On the other hand, the velocities of all geodesics of the limit metric d belong to $\mathscr{D}$ and for any nonzero vector $v \in \mathscr{D}$ there exists a one-parametric family of geodesics whose initial velocity is equal to v. Too bad up to now, and here is the first encouraging fact: the family of geodesic flows converges if we re-write it as a family of flows on the cotangent bundle.

Indeed, any Riemannian structure on M induces a self-adjoint isomorphism $G : TM \to T^*M$, where $\langle Gv, v\rangle$ is the square of the length of the vector $v \in TM$, and $\langle \cdot, \cdot \rangle$ denotes the standard pairing between tangent and cotangent vectors. The geodesic flow, treated as flow on T^*M is a Hamiltonian flow associated with the Hamiltonian function $H : T^*M \to \mathbb{R}$, where $H(\lambda) = \frac{1}{2}\langle \lambda, G^{-1}\lambda\rangle,\ \lambda \in T^*M$. Let $(\lambda(t), \gamma(t))$ be a trajectory of the Hamiltonian flow, with $\lambda(t) \in T^*_{\gamma(t)}M$. The square of the Riemannian distance from x_0 is a smooth function on a neighbourhood of x_0 in M and the differential of this function at $\gamma(t)$ is equal to $2t\lambda(t)$ for any small $t \geq 0$. Let H^ε be the Hamiltonian corresponding to the metric d^ε. It is easy to see that H^ε converges with all derivatives to a Hamiltonian H^0. Moreover, geodesics of the limit sub-Riemannian metric are just projections to M of the trajectories of the Hamiltonian flow on T^*M associated to H^0.

We can recover the Riemannian curvature from the asymptotic expansion of the square of the distance from x_0 along a geodesic: this is essentially what Riemann did. Then we can write a similar expansion for the square of the limit sub-Riemannian distance to get an idea of the curvature in this case. Note that the metrics d^ε converge to d with all derivatives in any point of $M \times M$, where d is smooth. The metrics d^ε are not smooth at the diagonal but their squares are smooth. The point is that no power of d is smooth at the diagonal! Nevertheless, the desired asymptotic expansion can be controlled.

Fix a point $x_0 \in M$ and $\lambda_0 \in T^*_{x_0}M$ such that $\langle \lambda_0, \mathscr{D}\rangle \neq 0$. Let $(\lambda^\varepsilon(t), \gamma^\varepsilon(t))$, for $\varepsilon \geq 0$, be the trajectory of the Hamiltonian flow associated to the Hamiltonian H^ε and initial condition

(λ_0, x_0). We set:

$$c_t^\varepsilon(x) := -\frac{1}{2t}(\mathsf{d}^\varepsilon)^2(x, \gamma^\varepsilon(t)) \text{ if } \varepsilon > 0, \qquad c_t^0(x) := -\frac{1}{2t}\mathsf{d}^2(x, \gamma^0(t)).$$

There exists an interval $(0, \delta)$ such that the functions c_t^ε are smooth at x_0 for all $t \in (0, \delta)$ and all $\varepsilon \geq 0$. Moreover, $d_{x_0} c_t^\varepsilon = \lambda_0$. Let $\dot{c}_t^\varepsilon = \frac{\partial}{\partial t} c_t^\varepsilon$, then $d_{x_0} \dot{c}_t^\varepsilon = 0$. In other words, x_0 is a critical point of the function $\dot{c}_t^\varepsilon$ and its Hessian $d_{x_0}^2 \dot{c}_t^\varepsilon$ is a well-defined quadratic form on $T_{x_0}M$. Recall that $\varepsilon = 0$ is available, but t must be positive. We are going to study the asymptotics of the family of quadratic forms $d_{x_0}^2 \dot{c}_t^\varepsilon$ as $t \to 0$ for fixed ε. This asymptotic is a little bit different for $\varepsilon > 0$ and $\varepsilon = 0$. The difference reflects the structural difference of the Riemannian and sub-Riemannian metrics and emphasises the role of the curvature.

Given $v, w \in T_xM,\ \varepsilon > 0$, we denote $\langle v|w\rangle_\varepsilon = \langle G^\varepsilon v, w\rangle$ the inner product generating d^ε. Recall that $\langle v|v\rangle_\varepsilon$ does not depend on ε if $v \in \mathscr{D}$ and $\langle v|v\rangle_\varepsilon \to \infty$ $(\varepsilon \to 0)$ if $v \notin \mathscr{D}$; we will write $|v|^2 := \langle v|v\rangle_\varepsilon$ in the first case. For fixed $\varepsilon > 0$, we have:

$$d_{x_0}^2 \dot{c}_t^\varepsilon(v) = \frac{1}{t^2}\langle v|v\rangle_\varepsilon + \frac{1}{3}\langle R^\varepsilon(\dot{\gamma}^\varepsilon, v)\dot{\gamma}^\varepsilon)|v\rangle_\varepsilon + O(t), \qquad v \in T_{x_0}M,$$

where $\dot{\gamma}^\varepsilon = \dot{\gamma}^\varepsilon(0)$ and R^ε is the Riemannian curvature tensor of the metric d^ε. For $\varepsilon = 0$, only vectors $v \in \mathscr{D}$ have a finite length and the above expansion is modified as follows:

$$d_{x_0}^2 \dot{c}_t^0(v) = \frac{1}{t^2}\mathcal{I}_\gamma(v) + \frac{1}{3}\mathcal{R}_\gamma(v) + O(t), \qquad v \in \mathscr{D} \cap T_{x_0}M,$$

where $\mathcal{I}_\gamma(v) \geq |v|^2$ and $\mathcal{R}_\gamma$ is the *sub-Riemannian curvature* at x_0 along the geodesic $\gamma = \gamma^0$. Both $\mathcal{I}_\gamma$ and $\mathcal{R}_\gamma$ are quadratic forms on $\mathscr{D}_{x_0} := \mathscr{D} \cap T_{x_0}M$. The principal "structural" term $\mathcal{I}_\gamma$ has the following properties:

$$\max\{\mathcal{I}_\gamma(v) |\, v \in \mathscr{D}_{x_0},\ |v|^2 = 1\} = 4,$$
$$\mathcal{I}_\gamma(v) = |v|^2 \text{ if and only if } d\omega(v, \dot{\gamma}(0)) = 0.$$

In other words, the symmetric operator on $\mathscr{D}_{x_0}$ associated with the quadratic form $\mathcal{I}_\gamma$ has eigenvalue 1 of multiplicity $\dim \mathscr{D}_{x_0} - 1$ and eigenvalue 4 of multiplicity 1. The trace of this operator, which, in this case, does not depend on γ, equals $\dim \mathscr{D}_{x_0} + 3$. This trace has a simple geometric interpretation, it is equal to the *geodesic dimension* of the sub-Riemannian space.

The geodesic dimension is defined as follows. Let $\Omega \subset M$ be a bounded and measurable subset of positive volume and let $\Omega_{x_0,t}$, for $0 \leq t \leq 1$, be a family of subsets obtained from Ω by the homothety of Ω with respect to a fixed point x_0 along the shortest geodesics connecting x_0 with the points of Ω, so that $\Omega_{x_0,0} = \{x_0\}$, $\Omega_{x_0,1} = \Omega$. The volume of $\Omega_{x_0,t}$ has order $t^{\mathcal{N}_{x_0}}$, where $\mathcal{N}_{x_0}$ is the geodesic dimension at x_0 (see Section 1.5.6 for details).

Note that the topological dimension of our contact sub-Riemannian space is $\dim \mathscr{D}_{x_0} + 1$ and the Hausdorff dimension is $\dim \mathscr{D}_{x_0} + 2$. All three dimensions are obviously equal for Riemannian or Finsler manifolds. The structure of the term $\mathcal{I}_\gamma$ and comparison of the asymptotic expansions of $d_{x_0}^2 \dot{c}_t^\varepsilon$ for $\varepsilon > 0$ and $\varepsilon = 0$ explains why sectional curvature goes to $-\infty$ for certain sections.

The curvature operator which we define can be computed in terms of the symplectic invariants of the so-called Jacobi curve, namely a curve in the Lagrange Grassmannian related with the linearisation of the Hamiltonian flow. These symplectic invariants can be computed, in principle, via an algorithm which is, however, quite hard to implement. Explicit computations of the contact sub-Riemannian curvature will appear in a forthcoming paper. In the current chapter we deal with the general setting. A precise construction in full generality is presented in the forthcoming sections but, since this chapter is long, we find it worth to briefly describe the main ideas in the introduction.

Let M be a smooth manifold, $\mathscr{D} \subset TM$ be a vector distribution (not necessarily contact), f_0 be a vector field on M and $L : TM \to M$ be a Tonelli Lagrangian (i.e. $L\big|_{T_xM}$ has a superlinear growth and its Hessian is positive definite for any $x \in M$). *Admissible paths* on M are curves whose velocities belong to the "affine distribution" $f_0 + \mathscr{D}$. Let $\mathcal{A}_t$ be the space of admissible paths defined on the segment $[0,t]$ and $N_t = \{(\gamma(0), \gamma(t)) : \gamma \in \mathcal{A}_t\} \subset M \times M$. The optimal cost (or action) function $S_t : N_t \to \mathbb{R}$ is defined as follows:

$$S_t(x,y) = \inf \left\{ \int_0^t L(\dot\gamma(\tau))\, d\tau : \gamma \in \mathcal{A}_t,\ \gamma(0) = x,\ \gamma(t) = y \right\}.$$

The space $\mathcal{A}_t$ equipped with the $W^{1,\infty}$-topology is a smooth Banach manifold; the functional $J_t : \gamma \mapsto \int_0^t L(\dot\gamma(\tau))\, d\tau$ and the evaluation maps $F_\tau : \gamma \mapsto \gamma(\tau)$ are smooth on $\mathcal{A}_t$.

The optimal cost $S_t(x,y)$ is the solution of the conditional minimum problem for the functional J_t under conditions $F_0(\gamma) = x,\ F_t(\gamma) = y$. The Lagrange multipliers rule for this problem reads:

$$d_\gamma J_t = \lambda_t D_\gamma F_t - \lambda_0 D_\gamma F_0. \tag{3}$$

Here λ_t and λ_0 are "Lagrange multipliers", $\lambda_t \in T^*_{\gamma(t)}M,\ \lambda_0 \in T^*_{\gamma(0)}M$. We have:

$$D_\gamma F_t : T_\gamma \mathcal{A}_t \to T_{\gamma(t)}M, \qquad \lambda_t : T_{\gamma(t)}M \to \mathbb{R},$$

and the composition $\lambda_t D_\gamma F_t$ is a linear functional on $T_\gamma \mathcal{A}_t$. Moreover, Eq. (3) implies that

$$d_\gamma J_\tau = \lambda_\tau D_\gamma F_\tau - \lambda_0 D_\gamma F_0, \tag{4}$$

for some $\lambda_\tau \in T^*_{\gamma(\tau)}M$ and any $\tau \in [0,t]$. The curve $\tau \mapsto \lambda_\tau$ is a trajectory of the Hamiltonian system associated to the Hamiltonian $H : T^*M \to \mathbb{R}$ defined by

$$H(\lambda) = \max_{v \in f_0(x) + \mathscr{D}_x} \left(\langle \lambda, v \rangle - L(v) \right), \qquad \lambda \in T^*_xM,\, x \in M.$$

Moreover, any trajectory of this Hamiltonian system satisfies relation (4), where γ is the projection of the trajectory to M. Trajectories of the Hamiltonian system are called *normal extremals* and their projections to M are called *normal extremal trajectories*.

We recover the sub-Riemannian setting when $f_0 = 0$, $L(v) = \frac{1}{2}\langle Gv, v\rangle$. In this case, the optimal cost is related with the sub-Riemannian distance $S_t(x,y) = \frac{1}{2t}\mathsf{d}^2(x,y)$, and normal extremal trajectories are normal sub-Riemannian geodesics.

Let γ be an admissible path; the germ of γ at the point $x_0 = \gamma(0)$ defines a flag in $T_{x_0}M$ $\{0\} = \mathscr{F}^0_\gamma \subset \mathscr{F}^1_\gamma \subset \mathscr{F}^2_\gamma \subset \ldots \subset T_{x_0}M$ in the following way. Let V be a section of the vector distribution $\mathscr{D}$ such that $\dot\gamma(t) = f_0(\gamma(t)) + V(\gamma(t))$, $t \geq 0$, and $P_{0,t}$ be the local flow on M generated by the vector field $f_0 + V$; then $\gamma(t) = P_{0,t}(\gamma(0))$. We set:

$$\mathscr{F}^i_\gamma = \operatorname{span}\left\{ \left.\frac{d^j}{dt^j}\right|_{t=0} (P_{0,t})^{-1}_* \mathscr{D}_{\gamma(t)} : j = 0, \dots, i-1 \right\}.$$

The flag $\mathscr{F}^i_\gamma$ depends only on the germs of $f_0 + \mathscr{D}$ and γ at the initial point x_0.

A normal extremal trajectory γ is called *ample* if $\mathscr{F}^m_\gamma = T_{x_0}M$ for some $m > 0$. If γ is ample, then $J_t(\gamma) = S_t(x_0, \gamma(t))$ for all sufficiently small $t > 0$ and S_t is a smooth function in a neighbourhood of $(\gamma(0), \gamma(t))$. Moreover, $\frac{\partial S_t}{\partial y}\big|_{y=\gamma(t)} = \lambda_t$, $\frac{\partial S_t}{\partial x}\big|_{x=\gamma(0)} = -\lambda_0$, where λ_t is the normal extremal whose projection is γ.

We set $c_t(x) := -S_t(x, \gamma(t))$; then $d_{x_0} c_t = \lambda_0$ for any $t > 0$ and x_0 is a critical point of the function $\dot c_t$. The Hessian of this function $d^2_{x_0}\dot c_t$ is a well-defined quadratic form on $T_{x_0}M$. We are going to write an asymptotic expansion of $d^2_{x_0}\dot c_t|_{\mathscr{D}_{x_0}}$ as $t \to 0$ (see Theorem 1.A):

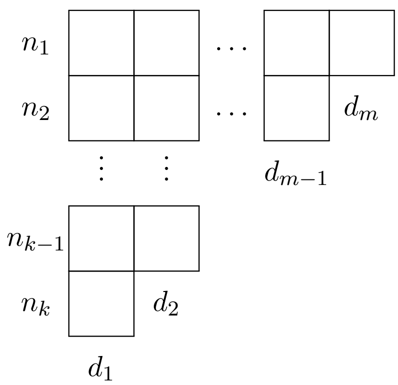
$$d^2_{x_0}\dot c_t(v) = \frac{1}{t^2}\mathcal{I}_\gamma(v) + \frac{1}{3}\mathcal{R}_\gamma(v) + O(t), \qquad \forall v \in \mathscr{D}_{x_0}.$$

Now we introduce a natural Euclidean structure on $T_{x_0}M$. Recall that $L|_{T_{x_0}M}$ is a strictly convex function, and $d^2_w(L|_{T_{x_0}M})$ is a positive definite quadratic form on $T_{x_0}M$, $\forall\, w \in T_{x_0}M$. If we set $|v|^2_\gamma = d^2_{\dot\gamma(0)}(L|_{T_{x_0}M})(v)$, $v \in T_{x_0}M$ we have the inequality

$$\mathcal{I}_\gamma(v) \geq |v|^2_\gamma, \qquad \forall\, v \in \mathscr{D}_{x_0}.$$

The inequality $\mathcal{I}_\gamma(v) \geq |v|^2_\gamma$ means that the eigenvalues of the symmetric operator on $\mathscr{D}_{x_0}$ associated with the quadratic form $\mathcal{I}_\gamma$ are greater or equal than 1. The quadratic form $\mathcal{R}_\gamma$ is the *curvature* of our constrained variational problem in the direction of the extremal trajectory γ.

A mild regularity assumption allows to explicitly compute the eigenvalues of $\mathcal{I}_\gamma$. We set $\gamma_\varepsilon(t) = \gamma(\varepsilon + t)$ and assume that $\dim \mathscr{F}^i_{\gamma_\varepsilon} = \dim \mathscr{F}^i_\gamma$ for all sufficiently small $\varepsilon \geq 0$ and all i. It turns out that $d_i = \dim \mathscr{F}^i_\gamma - \dim \mathscr{F}^{i-1}_\gamma$, for $i \geq 1$ is a non-increasing sequence of natural numbers with $d_1 = \dim \mathscr{D}_{x_0} = k$. We draw a Young tableau with d_i blocks in the i-th column and we define $n_1, \dots, n_k$ as the lengths of its rows (that may depend on γ).



The eigenvalues of the symmetric operator $\mathcal{I}_\gamma$ are $n_1^2, \dots, n_k^2$ (see Theorem 1.B). Some of these numbers may be equal (in the case of multiple eigenvalues) and are all equal to 1 in the Riemannian case. In the sub-Riemannian setting, the trace of $\mathcal{I}_\gamma$ is

$$\operatorname{tr} \mathcal{I}_\gamma = n_1^2 + \cdots + n_k^2 = \sum_{i=1}^{m} (2i-1) d_i,$$

When computed for the generic sub-Riemannian geodesic, this number is actually constant and depends only on x_0. This is what we called the *geodesic dimension* $\mathcal{N}_{x_0}$ of the manifolds. For Riemannian manifolds, this invariant is always equal to the topological dimension. For the $2n+1$-dimensional Heisenberg group, $\mathcal{N} = 2n+3$ (constantly on the manifold). Thus, the geodesic dimension is a new invariant, different from both the topological and the Hausdorff dimension of the sub-Riemannian space.

Let's see how this new dimension is related with the geometry of the sub-Riemannian manifold. Fix any smooth measure μ on the manifold, and let Ω be a measurable set with $0 < \mu(\Omega) < +\infty$. Fix $x_0 \in M$. For simplicity, assume that Ω does not intersect the cut locus of x_0. We define the *homothety* with center x_0 at time $t \in [0,1]$ of the set Ω as follows. Let $x \in \Omega$, and consider the unique geodesic γ such that $\gamma(0) = x_0$ and $\gamma(1) = x$. Then the homothethy of x at time t with center x_0 is the point $\gamma(t)$. Doing this for any $x \in \Omega$ defines a new set Ω_t, such that $\Omega_0 = \{x_0\}$ and $\Omega_1 = \Omega$. The main result of Sec. 1.5.6 is the following (see Theorem 1.D):

$$\mu(\Omega_t) \sim t^{\mathcal{N}_{x_0}}, \qquad t \to 0.$$

Namely geodesic dimension represents the critical exponent such that the volume of a measurable set shrinks to zero as $t \to 0$ along a sub-Riemannian homothety.

The last result of this chapter is related with the intrinsic Laplacian of a sub-Riemannian manifold. We sketch the general construction. For any fixed smooth volume form μ, we define the μ-divergence of a vector field $X \in \operatorname{Vec}(M)$ by the following formula

$$\mathcal{L}_X \mu = \operatorname{div}_\mu(X)\mu,$$

where $\mathcal{L}$ represents the Lie derivative. Moreover, for any $f \in C^\infty(M)$, we define the sub-Riemannian gradient ∇f as the unique horizontal vector field such that $g(\nabla f, \cdot) = df(\cdot)$. Thus, we define the μ-Laplacian as $\Delta_\mu f := \operatorname{div}_\mu \nabla f$. In the Riemannian setting, when μ is the Riemannian volume form, this construction leads to the familiar Laplace-Beltrami operator. In the sub-Riemannian setting, one can choose μ to be the canonical Popp's volume (see Chapter 4), and obtain an intrinsic sub-Laplacian operator. Still, we prefer to leave μ general here. The main result of Sec. 1.5 is the relation between the curvature and the geodesic dimension with the asymptotic behaviour of the sub-Laplacian of the sub-Riemannian squared distance from a geodesic, i.e. the function

$$\mathfrak{f}_t(\cdot) := -t c_t(\cdot) = \frac{1}{2}\mathsf{d}^2(\cdot, \gamma(t)), \qquad t \in (0,1],$$

where $\gamma(t)$ is an ample sub-Riemannian geodesic such that $\gamma(0) = x_0$. In particular we prove that (see Theorem 1.C)

$$\Delta_\mu \mathfrak{f}_t = \operatorname{tr} \mathcal{I}_\gamma - \dot{g}(0)t - \frac{1}{3} \operatorname{tr} \mathcal{R}_\gamma t^2 + O(t^3),$$

where $g : [0,T] \to \mathbb{R}$ is a smooth function that depends on the choice of the volume μ, whose precise definition is not needed here. This, in particular, implies that

$$\lim_{t\to 0} \Delta_\mu \mathfrak{f}_t \big|_{x_0} = \operatorname{tr} \mathcal{I}_\gamma,$$
$$\frac{d^2}{dt^2}\bigg|_{t=0} \Delta_\mu \mathfrak{f}_t \big|_{x_0} = -\frac{2}{3} \operatorname{tr} \mathcal{R}_\gamma.$$

The construction of the curvature presented here was preceded by a rather long research line (see [3,16–18,49,71]). For what concerns the alternative approaches to this topic, in recent years, several efforts have been made to introduce a notion of curvature to non-Riemannian situations, such as sub-Riemannian manifolds and, more in general, metric measure spaces. Motivated by the lack of classical Riemannian tools (such as the Levi-Civita connection and the theory of Jacobi fields) different approaches have been explored in order to extend some classical results in geometric analysis to such structures. In particular, to this extent, many synthetic notions of generalized Ricci curvature bound have been introduced. For instance, one can see [29,30] and references therein for a heat equation approach to the generalization of the curvature-dimension inequality and [19,51,67,68] and references therein for an optimal transport approach to the generalization of Ricci curvature.

2 Comparison theorems for conjugate points in sub-Riemannian geometry

Among the most celebrated results in Riemannian geometry, comparison theorems play a prominent role. These theorems allow to estimate properties of a manifold under investigation with the same property on the *model spaces* which, in the classical setting, are the simply connected manifolds with constant sectional curvature (the sphere, the Euclidean plane and the hyperbolic plane). The properties that may be investigated with these techniques are countless and include, among the others, the number of conjugate points along a given geodesic, the topology of loop spaces, the behaviour of volume of sets under homotheties, Laplacian comparison theorems, estimates for solutions of PDEs on the manifold, etc.

In this chapter we are concerned, in particular, with results of the following type. Until further notice, M is a Riemannian manifold, endowed with the Levi-Civita connection, and $\operatorname{Sec}(v,w)$ is the sectional curvature of the section $\operatorname{span}\{v,w\} \subset T_xM$.

Theorem 1. *Let $\gamma(t)$ be a unit-speed geodesic on M. If for all $t \geq 0$ and for all $v \in T_{\gamma(t)}M$ orthogonal to $\dot\gamma(t)$ with unit norm $\operatorname{Sec}(\dot\gamma(t), v) \geq k > 0$, then there exists $0 < t_c \leq \pi/\sqrt{k}$ such that $\gamma(t_c)$ is conjugate with $\gamma(0)$.*

Notice that the quadratic form $\mathrm{Sec}(\dot\gamma(t),\cdot) : T_{\gamma(t)}M \to \mathbb{R}$, which we call *directional curvature* (in the direction of $\dot\gamma$), computes the sectional curvature of the sections containing $\dot\gamma$. Theorem 1 compares the distance of the first conjugate point along γ with the same property computed on the sphere with sectional curvature $k > 0$, provided that the directional curvature along the geodesic on the reference manifold is bounded from below by k. Theorem 1 also contains all the basic ingredients of a comparison-type result:

- A micro-local condition, i.e. "along the geodesic", given in terms of a bound on curvature-type quantities, such as the sectional or Ricci curvature.
- Models for comparison, that is spaces in which the property under investigation can be computed explicitly.

As it is well known, Theorem 1 can be improved by replacing the bound on the directional curvature with a bound on the average, or Ricci curvature. Moreover, Theorem 1 leads immediately to the celebrated Bonnet-Myers theorem (see [57]).

In Riemannian geometry, the importance of conjugate points rests on the fact that geodesics cease to be minimizing after the first one. This remains true for strongly normal sub-Riemannian geodesics. Moreover, conjugate points, both in Riemannian and sub-Riemannian geometry, are also intertwined with the analytic properties of the underlying structure, for example they affect the behaviour of the heat kernel (see [22, 24] and references therein).

The main results of this chapter are comparison theorems on the existence of conjugate points, valid for any sub-Riemannian structure.

We briefly recall the concept of sub-Riemannian structure. A sub-Riemannian structure on a manifold M is defined by a distribution $\mathscr{D} \subseteq TM$ of constant rank, with a scalar product that, unlike the Riemannian case, is defined only for vectors in $\mathscr{D}$. Under mild assumptions on $\mathscr{D}$ (the Hörmander condition) any connected sub-Riemannian manifold is *horizontally* path-connected, namely any two points are joined by a path whose tangent vector belongs to $\mathscr{D}$.

Thus, a rich theory paralleling the classical Riemannian one can be developed, giving a meaning to the concept of geodesic, as an horizontal curve that locally minimises the length. Still, since in general there is no canonical completion of the sub-Riemannian metric to a Riemannian one, there is no way to define a connection à la Levi-Civita and thus the familiar Riemannian curvature tensor. The classical theory of Jacobi fields and its connection with the curvature plays a central role in the proof of many Riemannian comparison results, and the generalisation to the sub-Riemannian setting is not straightforward. The Jacobi equation itself, being defined in terms of the covariant derivative, cannot be formalised in the classical sense when a connection is not available.

In this chapter we focus on results in the spirit of Theorem 1 even tough there are no evident obstructions to the application of the same techniques, relying on the Riccati equations for sub-Riemannian geodesics, to other types of comparison results. We anticipate that the comparisons models will be linear quadratic optimal control problems (LQ problems in the following), i.e. minimization problems quite similar to the Riemannian one, where the length is replaced by a functional defined by a quadratic Lagrangian. More precisely one is interested in finding *admissible trajectories* of a *linear* control system in $\mathbb{R}^n$, namely curves $x : [0,t] \to \mathbb{R}^n$ for which there exists a control $u \in L^2([0,t],\mathbb{R}^k)$ such that

$$\dot{x} = Ax + Bu, \qquad x(0) = x_0, \qquad x(t) = x_1, \qquad x_0, x_1, t \text{ fixed},$$

that minimize a *quadratic* functional $\phi_t : L^2([0,t],\mathbb{R}^k) \to \mathbb{R}$ of the form

$$\phi_t(u) = \frac{1}{2}\int_0^t (u^* u - x^* Q x)\, dt.$$

Here A, B, Q are constant matrices of the appropriate dimension. The symmetric matrix Q is usually referred to as the *potential.* Notice that it makes sense to speak about *conjugate time* of a LQ problem: it is the time $t_c > 0$ at which extremal trajectories lose local optimality. It turns out that t_c does not depend on the data x_0, x_1, but it is an intrinsic feature of the problem. These kind of structures are well known in the field of optimal control theory, but to our best knowledge this is the first time they are employed as model spaces for comparison results.

With any ample, equiregular sub-Riemannian geodesic $\gamma(t)$ (see Definition 2.12), we associate: its *Young diagram* D, a scalar product $\langle\cdot|\cdot\rangle_{\gamma(t)} : T_{\gamma(t)}M \times T_{\gamma(t)}M \mapsto \mathbb{R}$ extending the sub-Riemannian one and a quadratic form $\mathfrak{R}_{\gamma(t)} : T_{\gamma(t)}M \mapsto \mathbb{R}$ (the sub-Riemannian directional curvature), all depending on the geodesic $\gamma(t)$. We stress that, for a Riemannian manifold, any non-trivial geodesic has the same Young diagram, composed by a single column with $n = \dim M$ boxes, the scalar product $\langle\cdot|\cdot\rangle_{\gamma(t)}$ coincides with the Riemannian one, and $\mathfrak{R}_{\gamma(t)}(v) = \mathrm{Sec}(v, \dot{\gamma}(t))$ for all $v \in T_{\gamma(t)}M$.

In this introduction, when we associate with a geodesic $\gamma(t)$ its Young diagram D, we implicitly assume that $\gamma(t)$ is ample and equiregular. Notice that these assumptions are true for the generic geodesic, as we discuss more precisely in Sec. 2.2.2.

In the spirit of Theorem 1, assume that the sub-Riemannian directional curvature is bounded from below by a quadratic form Q. Then, we associate a model LQ problem (i.e. matrices A and B, depending on γ) which, roughly speaking, represents the linearisation of the sub-Riemannian structure along γ itself, with potential Q. We dub this *model space* $\mathrm{LQ}(D;Q)$, where D is the Young diagram of γ, and Q represents the bound on the sub-Riemannian directional curvature. The first of our results can be stated as follows (see Theorem 2.A).

Theorem 2. *Let $\gamma(t)$ be a sub-Riemannian geodesic, with Young diagram D, such that $Q_- \geq \mathfrak{R}_{\gamma(t)} \geq Q_+$ for all $t \geq 0$. Then the first conjugate point along $\gamma(t)$ occurs at a time t_c not greater than the first conjugate time of the model* $\mathrm{LQ}(D;Q_+)$ *and not smaller than the first conjugate time of* $\mathrm{LQ}(D;Q_-)$.

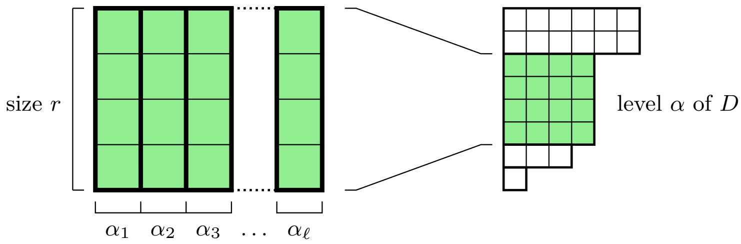
In the Riemannian case, any non-trivial geodesic γ has the same (trivial) Young diagram, and this leads to a simple LQ model with $A = 0$, $B = \mathbb{I}$. Moreover, $\langle\cdot|\cdot\rangle_\gamma$ is the Riemannian scalar product and $\mathfrak{R}_\gamma = \mathrm{Sec}(\dot{\gamma}, \cdot)$. Then, if Theorem 2 holds with $Q_+ = k\mathbb{I}$, the first conjugate point along the Riemannian geodesic, with directional curvature bounded by k occurs at a time t not greater than the first conjugate time of the LQ model

$$\dot{x} = u, \qquad \phi_t(u) = \frac{1}{2}\int_0^t \left(|u|^2 - k|x|^2\right) dt.$$

It is well known that, when $k > 0$, this problem represents a simple n-dimensional harmonic oscillator, whose extremal trajectories lose optimality at time $t_c = \pi/\sqrt{k}$. Thus we recover Theorem 1. In the sub-Riemannian setting, due to the intrinsic anisotropy of the structure different geodesics have different Young diagrams, resulting in a rich class of LQ models, with non-trivial drift terms. The directional sub-Riemannian curvature $\mathfrak{R}_{\gamma(t)}$ represents the potential "experienced" in a neighbourhood of the geodesic.

We stress that the generic $\mathrm{LQ}(D;Q)$ model may have infinite conjugate time. However, as we discuss in full detail in Chapter 3, there are necessary and sufficient conditions for its finiteness. Thus Theorem 2 can be employed to prove both existence or non-existence of conjugate points along a given geodesic.

As Theorem 1 can be improved by considering a bound on the Ricci curvature in the direction of the geodesic, instead of the whole sectional curvature, also Theorem 2 can be improved in the same spirit. In the sub-Riemannian case, however, the process of "taking the trace" is more delicate. Due to the anisotropy of the structure, it only makes sense to take *partial* traces, leading to a number of Ricci curvatures (each one obtained as a partial trace on an invariant subspace of $T_{\gamma(t)}M$, determined by the Young diagram D). In particular, for each *level* α of the Young diagram (namely the collection of all the rows with the same length equal to, say, ℓ) we have ℓ Ricci curvatures $\mathfrak{Ric}^{\alpha_i}_{\gamma(t)}$, for $i = 1, \dots, \ell$. The *size* of a level is the number r of boxes in each of its columns $\alpha_1, \dots, \alpha_\ell$.



The partial tracing process leads to our main result (see Theorem 2.B).

Theorem 3. *Let $\gamma(t)$ be a sub-Riemannian geodesic with Young diagram D. Consider a fixed level α of D, with length ℓ and size r. Then, if*

$$\frac{1}{r}\,\mathfrak{Ric}^{\alpha_i}_{\gamma(t)} \geq k_i, \qquad \forall i = 1, \dots, \ell, \quad \forall t \geq 0,$$

the first conjugate time $t_c(\gamma)$ along the geodesic satisfies $t_c(\gamma) \leq t_c(k_1, \dots, k_\ell)$.

In Theorem 3, $t_c(k_1, \dots, k_\ell)$ denotes the first conjugate time of the LQ model associated with a Young diagram with a single row, of length ℓ, and a diagonal potential $Q = \mathrm{diag}\{k_1, \dots, k_\ell\}$.

The hypotheses in Theorem 3 are no longer bounds on a quadratic form as in Theorem 2, but a finite number of *scalar* bounds. Observe that we have one comparison theorem for each level of the Young diagram of the given geodesic. In the Riemannian case, as we discussed earlier, D has only one level, of length $\ell = 1$, of size $r = \dim M$. In this case there is single Ricci curvature, namely $\mathfrak{Ric}^{\alpha_1}_{\gamma(t)} = \mathrm{Ric}(\dot\gamma(t))$ and, if $k_1 > 0$ in Theorem 3, $t_c(k_1) = \pi/\sqrt{k_1} < +\infty$. Back to the general case, we stress that in order to have $t_c(k_1, \dots, k_\ell) < +\infty$, the Riemannian condition

$\mathrm{Ric}(\dot{\gamma}) \geq k_1 > 0$ must be replaced by more complicated inequalities on the bounds $k_1, \dots, k_\ell$ on the sub-Riemannian Ricci curvatures. In particular, we allow also for some *negative values* of such constants.

As an application of Theorem 3, we prove a sub-Riemannian version of the classical Bonnet-Myers theorem (see Theorem 2.C).

Theorem 4. *Let M be a connected, complete sub-Riemannian manifold, such that the generic geodesic has the same Young diagram D. Assume that there exists a level α of length ℓ and size r and constants $k_1, \dots, k_\ell$ such that, for any length parametrized geodesic $\gamma(t)$*

$$\frac{1}{r}\,\mathfrak{Ric}^{\alpha_i}_{\gamma(t)} \geq k_i, \qquad \forall i = 1, \dots, \ell, \quad \forall t \geq 0.$$

Then, if the polynomial

$$P_{k_1,\dots,k_\ell}(x) := x^{2\ell} - \sum_{i=0}^{\ell-1} (-1)^{\ell-i} k_{\ell-i} x^{2i}$$

has at least one simple purely imaginary root, the manifold is compact, has diameter not greater than $t_c(k_1, \dots, k_\ell) < +\infty$. Moreover, its fundamental group is finite.

In the Riemannian setting we recover the classical Bonnet-Myers theorem, since $\ell = 1$, $r = \dim M$ and the condition on the roots of $P_{k_1}(x) = x^2 + k_1$ is equivalent to $k_1 > 0$.

Finally we apply our techniques to obtain information about the conjugate time of geodesics on 3D *unimodular* Lie groups. Left-invariant structures on 3D Lie groups are the basic examples of sub-Riemannian manifolds and the study of such structures is the starting point to understand the general properties of sub-Riemannian geometry.

A complete classification of such structures, up to local sub-Riemannian isometries, is given in [5, Thm. 1], in terms of the two basic geometric invariants $\chi \geq 0$ and κ, that are constant for left-invariant structures. In particular, for each choice of the pair (χ, κ), there exists a unique unimodular group in this classification. Even if left-invariant structures possess the symmetries inherited by the group structure, the sub-Riemannian geodesics and their conjugate loci have been studied only in some particular cases where explicit computations are possible.

The conjugate locus of left-invariant structures has been completely determined for the cases corresponding to $\chi = 0$, that are the Heisenberg group [40] and the semisimple Lie groups $SU(2), SL(2)$ where the metric is defined by the Killing form [34]. On the other hand, when $\chi > 0$, only few cases have been considered up to now. In particular, to our best knowledge, only the the sub-Riemannian structure on the group of motions of the Euclidean plane $SE(2)$, where $\chi = \kappa > 0$, has been considered [55, 65].

As an application of our results, we give a sufficient condition for the existence of finite conjugate times for geodesics in an unimodular Lie group, together with an estimate for it (non-sharp, in general). This condition is expressed in terms of a bound, depending on χ, k, on a constant of the motion $E(\gamma)$ associated with the given geodesic γ (see Theorem 2.D).

Theorem 5. *Let M be a 3D Lie group endowed with a contact left-invariant sub-Riemannian structure with invariants $\chi > 0$ and $\kappa \in \mathbb{R}$. Then there exists $\bar{E} = \bar{E}(\chi, \kappa)$ such that every length parametrized geodesic γ with $E(\gamma) \geq \bar{E}$ has a finite conjugate time.*

The cases corresponding to $\chi = 0$ are $\mathbb{H}$, $SU(2)$ and $SL(2)$, where $\kappa = 0, 1, -1$, respectively. For these structures we recover the exact estimates for the first conjugate time of a length parametrized geodesic (see Section 2.7.2).

The curvature employed in this chapter has been introduced for the first time by Agrachev and Gamkrelidze in [17], Agrachev and Zelenko in [16] and successively extended by Zelenko and Li in [71], where also the Young diagram is introduced for the first time in relation with the extremals of a variational problem. This research has been inspired by many recent works in this direction that we briefly review.

In [18] Agrachev and Lee investigate a generalisation of the measure contraction property (MCP) to 3D sub-Riemannian manifolds. The generalised MCP of Agrachev and Lee is expressed in terms of solutions of a particular 2D matrix Riccati equation for sub-Riemannian extremals, and this is one of the technical points that mostly inspired the present research.

In [48] Lee, Li and Zelenko pursue further progresses for sub-Riemannian contact structures with transversal symmetries. In this case, it is possible to exploit the Riemannian structure induced on the quotient space to write the curvature operator, and the authors recover sufficient condition for the contact manifold to satisfy the generalised MCP defined in [18]. Moreover, the authors perform the first step in the decoupling of the matrix Riccati equation for different levels of the Young diagram (see the splitting part of the proof of Theorem 2.B for more details).

The MCP for higher dimensional sub-Riemannian structures has also been investigated in [64] for Carnot groups.

We also mention that, in [49], Li and Zelenko prove comparison results for the number of conjugate points of curves in a Lagrange Grassmanian associated with sub-Riemannian structures with symmetries. In particular, [49, Cor. 4] is equivalent to Theorem 2.2, but obtained with differential topology techniques and with a different language. However, to our best knowledge, it is not clear how to obtain an averaged version of such comparison results with these techniques, and this is yet another motivation that led to Theorem 2.3.

In [29], Baudoin and Garofalo prove, with heat-semigroup techniques, a sub-Riemannian version of the Bonnet-Myers theorem for sub-Riemannian manifolds with transverse symmetries that satisfy an appropriate generalisation of the Curvature Dimension (CD) inequalities introduced in the same paper. In [30], Baudoin and Wang generalise the previous results to contact sub-Riemannian manifolds, removing the symmetries assumption. See also [27, 28] for other comparison results following from the generalised CD condition.

Even though in this chapter we discuss only sub-Riemannian structures, these techniques can be applied to the extremals of any affine optimal control problem, a general framework including (sub)-Riemannian, (sub)-Finsler manifolds, as discussed in [9]. For example, in [17], the authors prove a comparison theorem for conjugate points along extremals associated with *regular* Hamiltonian systems, such as those corresponding to Riemannian and Finsler geodesics. Finally, concerning comparison theorems for Finsler structures one can see, for example, [58, 60, 70].

Remark: $\mathcal{R}$ vs $\mathfrak{R}$

We made the effort to keep a uniform and coherent notation throughout the thesis, but to avoid confusion an important remark is in order. In Chapters 1 and 2, respectively, two seemingly different, but strongly related "curvatures" do appear. On one hand, in Chapter 1, the *curvature* $\mathcal{R} : \mathscr{D} \to \mathscr{D}$ is a symmetric operator defined on a subspace of the tangent space (notice that, in the sub-Riemannian case, $\mathscr{D}$ is the horizontal distribution). In this setting, $\mathcal{R}$ measures the distortion of the cost function in a neighbourhood of a fixed geodesic. On the other hand, in Chapter 2, where the setting is the sub-Riemannian one, the *directional curvature* $\mathfrak{R}$ is a symmetric operator defined on the *whole* tangent space, obtained by the symplectic invariants of the Jacobi curve associated with a fixed geodesic.

The two operators are strictly related: up to some constant coefficients, $\mathcal{R}$ is the restriction of $\mathfrak{R}$ to $\mathscr{D}$. This fact express the connection between the symplectic invariants of the Jacobi curve (contained in $\mathfrak{R}$) and the asymptotics of the cost function (trough $\mathcal{R}$), and is actually one of the main results of Chapter 1. The proof of this fact is a consequence of the asymptotics of the Jacobi curve of Theorem 1.122 (in particular, see Eq. 1.81, and compare it with the definition of $\mathfrak{R}$ in Chapter 2).

In the Riemannian setting, $\mathscr{D}$ is the whole tangent space and both curvatures coincide with the Riemannian sectional curvature in the direction of the fixed geodesic γ, namely for any unit vector $v \in T_\gamma M$,

$$\mathfrak{R}(v) = \mathcal{R}(v) = \mathrm{Sec}(v, \dot{\gamma})$$

Without entering into details, we believe that one can recover the *whole* $\mathfrak{R}$ (that is all the symplectic invariants of the Jacobi curve) in the asymptotic expansion of the geodesic cost function, but this is left to future investigations. .

3 On conjugate times of LQ optimal control problems

Linear Quadratic optimal control problems (LQ in the following) are a standard topic in control theory and dynamical systems, and are very popular in applications. They consist in a linear control system with quadratic Lagrangian. We briefly recall the general features of a LQ problem, and we refer to [15, Chapter 16] and [46, Chapter 7] for further details. We are interested in *admissible trajectories*, namely curves $x : [0, t_1] \to \mathbb{R}^n$ such that there exists a control $u \in L^2([0, t_1], \mathbb{R}^k)$ such that

$$\dot{x} = Ax + Bu, \qquad x(0) = x_0, \qquad x(t_1) = x_1, \qquad x_0, x_1, t_1 \text{ fixed},$$

that minimize a quadratic functional $\phi_{t_1} : L^2([0, t_1], \mathbb{R}^k) \to \mathbb{R}$ of the form

$$\phi_{t_1}(u) = \frac{1}{2} \int_0^{t_1} (u^* R u + x^* P u + x^* Q x)\, dt.$$

The condition $R \geq 0$ is necessary for existence of optimal control. We also assume $R > 0$ (for the singular case we refer to [46, Chapter 9]). Without loss of generality we may reduce to the case

$$\phi_{t_1}(u) = \frac{1}{2} \int_0^{t_1} (u^* u - x^* Q x)\, dt.$$

Here A, B, Q are constant matrices of the appropriate dimension. The vector Ax represents the *drift* field, while the columns of B represent the controllable directions. The meaning of the *potential* term Q will be clear later, when we will introduce the Hamiltonian associated with the LQ problem.

We assume that the system is *controllable*, namely there exists $m > 0$ such that

$$\operatorname{rank}(B, AB, \dots, A^{m-1}B) = n.$$

This hypothesis implies that, for any choice of t_1, x_0, x_1, the set of controls u such that the associated trajectory $x_u : [0, t_1] \to \mathbb{R}^n$ connects x_0 with x_1 in time t_1 is non-empty.

It is well known that the optimal trajectories of the LQ system are projections $(p, x) \mapsto x$ of the solutions of the Hamiltonian system

$$\dot{p} = -\partial_x H(p, x), \qquad \dot{x} = \partial_p H(p, x), \qquad (p, x) \in T^*\mathbb{R}^n = \mathbb{R}^{2n},$$

where the Hamiltonian function $H : \mathbb{R}^{2n} \to \mathbb{R}$ is defined by

$$H(p, x) = \frac{1}{2}(p, x)^* \mathbf{H} \begin{pmatrix} p \\ x \end{pmatrix}, \qquad \mathbf{H} = \begin{pmatrix} BB^* & A \\ A^* & Q \end{pmatrix}.$$

We denote by $P_t : \mathbb{R}^{2n} \to \mathbb{R}^{2n}$ the flow of the Hamiltonian system, which is defined for all $t \in \mathbb{R}$. To exploit the natural symplectic setting on $T^*\mathbb{R}^n = \mathbb{R}^{2n}$, we employ canonical coordinates (p, x) such that the symplectic form $\omega = \sum_{i=1}^n dp_i \wedge dx_i$ is represented by the matrix $\Omega = \left(\begin{smallmatrix} 0 & \mathbb{I}_n \\ -\mathbb{I}_n & 0 \end{smallmatrix}\right)$. The flow lines of P_t are precisely the integral lines of the *Hamiltonian vector field* $\vec{H} \in \operatorname{Vec}(\mathbb{R}^{2n})$, defined by $dH(\cdot) = \omega(\,\cdot\,, \vec{H})$. More explicitly

$$\vec{H}_{(p,x)} = \begin{pmatrix} -A^* & -Q \\ BB^* & A \end{pmatrix} \begin{pmatrix} p \\ x \end{pmatrix} = -\Omega \mathbf{H} \begin{pmatrix} p \\ x \end{pmatrix}.$$

By the term Hamiltonian vector field, we denote both the linear field $\vec{H}$ and the associated matrix $-\Omega\mathbf{H}$. The Hamiltonian flow can be explicitly written in terms of the latter as

$$P_t = e^{-t\Omega\mathbf{H}},$$

where the r.h.s. is the standard matrix exponential.

Conjugate times

We stress that not all the integral lines of the Hamiltonian flow lead to minimizing solutions of the LQ problem, since they only satisfy first order conditions for optimality. For this reason, they are usually called *extremals*. Sufficiently short segments, however, are optimal, but they lose optimality at some time $t_c > 0$, called the *first conjugate time*. In the following, we give a geometrical definition of conjugate time, in terms of curves in the Grassmannian of Lagrangian subspaces of $\mathbb{R}^{2n}$.

We say that a subspace $\Lambda \subset \mathbb{R}^{2n}$ is *Lagrangian* if $\omega|_\Lambda \equiv 0$, and $\dim \Lambda = n$. A notable example of Lagrangian subspace is the *vertical* subspace, that is $\mathcal{V} := \{(p, 0) |\, p \in \mathbb{R}^n\}$.

Definition 6. The Jacobi curve of the LQ problem $J(\cdot)$ is the following family of Lagrangian subspaces of $\mathbb{R}^{2n}$

$$J(t) := e^{t\Omega\mathbf{H}}\mathcal{V}, \qquad \mathcal{V} := \{(p,0)|\, p \in \mathbb{R}^n\}.$$

From the geometrical viewpoint, $J(\cdot)$ is a smooth curve in the submanifold of the Grassmannian of the n-dimensional subspaces of $\mathbb{R}^{2n}$ defined by the Lagrangian subspaces.

Definition 7. We say that t is a conjugate time if $J(t) \cap \mathcal{V} \neq 0$. The *multiplicity* of the conjugate time t is the dimension of the intersection.

In the language introduced by V. Arnold, these are times of *verticality* of the Jacobi curve. It is not hard to show that t is a conjugate time if and only if there exist solutions of the Hamilton equations such that $x(0) = x(t) = 0$.

The first conjugate time determines existence and uniqueness of minimizing solutions of the LQ problem, as specified by the following proposition (see [15, Chapter 16]).

Proposition 8. *Let t_c be the first conjugate time, namely $t_c := \inf\{t > 0|\, J(t) \cap \mathcal{V} \neq 0\}$.*

- *For $t_1 < t_c$, for any x_0, x_1 there exists a unique minimizer connecting x_0 with x_1 in time t_1.*
- *For $t_1 > t_c$, for any x_0, x_1 there exists no minimizer connecting x_0 with x_1 in time t_1.*
- *For $t_1 = t_c$, existence of minimizers depends on the initial data.*

In this chapter we completely characterise the occurrence of conjugate times for a controllable LQ problem. In particular, we prove the following result (see Theorem 3.A).

Theorem 9. *The conjugate times of a controllable linear quadratic optimal control problem obey the following dichotomy:*

- *If the Hamiltonian field $\vec{H}$ has at least one odd-dimensional Jordan block corresponding to a pure imaginary eigenvalue, the number of conjugate times in the interval $[0,T]$ grows to infinity for $T \to \pm\infty$.*
- *If the Hamiltonian field $\vec{H}$ has no odd-dimensional Jordan blocks corresponding to a pure imaginary eigenvalue, there are no conjugate times.*

As a corollary, the first conjugate time of a LQ optimal problem is finite if and only if the Hamiltonian field $\vec{H}$ has at least one odd-dimensional Jordan block corresponding to a pure imaginary eigenvalue. These are precisely the conditions we mentioned in the introduction of Chapter 2 for the existence of finite conjugate times.

Finally, in Sec. 3.3, we also provide estimates for the first conjugate time, in terms of the (signed) eigenvalues of $\vec{H}$ (see Corollaries 3.31 and 3.33).

4 A formula for Popp's volume in sub-Riemannian geometry

The problem to define a canonical volume on a sub-Riemannian manifold was first pointed out by Brockett in his seminal paper [35], motivated by the construction of a Laplace operator on a 3D sub-Riemannian manifold canonically associated with the metric structure, analogous to the Laplace-Beltrami operator on a Riemannian manifold. Recently, Montgomery addressed this problem in the general case (see [56, Chapter 10]).

Even on a Riemannian manifold, the Laplacian (defined as the divergence of the gradient) is a second order differential operator whose first order term depends on the choice of the volume on the manifold, which is required to define the divergence. Naively, in the Riemannian case, the choice of a canonical volume is determined by the metric, by requiring that the volume of a orthonormal parallelotope (i.e. whose edges are an orthonormal frame in the tangent space) is 1.

From a geometrical viewpoint, sub-Riemannian geometry is a natural generalization of Riemannian geometry under non-holonomic constraints. Formally speaking, a sub-Riemannian manifold is a smooth manifold M endowed with a bracket-generating distribution $\mathscr{D} \subset TM$, with $k = \operatorname{rank} \mathscr{D} < n = \dim M$, and a smooth fibre-wise scalar product on $\mathscr{D}$. From this structure, one derives a distance on M - the so-called *Carnot-Caratheodory metric* - as the infimum of the length of *horizontal* curves on M, i.e. the curves that are almost everywhere tangent to the distribution.

Nevertheless, sub-Riemannian geometry enjoys major differences with respect to the Riemannian case. For instance, a construction analogue to the one described above for the Riemannian volume is not possible. Indeed the inner product is defined only on a subspace of the tangent space, and there is no canonical way to extend it on the whole tangent space.

Popp's volume is a generalization of the Riemannian volume in sub-Riemannian setting. It was first defined by Octavian Popp but introduced only in [56] (see also [10]). Such a volume is smooth only for an equiregular sub-Riemannian manifold, i.e. when the dimensions of the higher order distributions $\mathscr{D}^1 := \mathscr{D}$, $\mathscr{D}^{i+1} := \mathscr{D}^i + [\mathscr{D}^i, \mathscr{D}]$, for every $i \geq 1$, do not depend on the point (for precise definitions, see Sec. 2).

Under the equiregularity hypothesis, the bracket-generating condition guarantees that there exists a minimal $m \in \mathbb{N}$, called *step* of the structure, such that $\mathscr{D}^m = TM$. Then, for each $q \in M$, it is well defined the graded vector space:

$$\operatorname{gr}_q(\mathscr{D}) := \bigoplus_{i=1}^{m} \mathscr{D}_q^i / \mathscr{D}_q^{i-1}, \qquad \text{where} \quad \mathscr{D}_q^0 = 0.$$

The vector space $\operatorname{gr}_q(\mathscr{D})$, which can be endowed with a natural sub-Riemannian structure, is called the *nilpotentization* of the structure at the point q, and plays a role analogous to the Euclidean tangent space in Riemannian geometry. Popp's volume is defined by inducing a canonical inner product on $\operatorname{gr}_q(\mathscr{D})$ via the Lie brackets, and then using a *non-canonical* isomorphism between $\operatorname{gr}_q(\mathscr{D})$ and T_qM to define an inner product on the whole T_qM. Interestingly, even though this construction depends on the choice of some complement to the distribution, the associated volume form (i.e. Popp's volume) is independent on this choice.

It is worth to recall that on a sub-Riemannian manifold, which is a metric space, the Haussdorff volume and the spherical Hausdorff volume, respectively $\mathcal{H}^Q$ and $\mathcal{S}^Q$, are canonically defined.[1] The relation between Popp's volume and $\mathcal{S}^Q$ has been studied in [7], where the authors show how the Radon-Nikodym derivative is related with the nilpotentization of the structure. In particular they prove that the Radon-Nikodym derivative could also be non smooth (see also [23, 32]). Remember that the Hausdorff and spherical Hausdorff volumes are both proportional to the Riemannian one on a Riemannian manifold. The relation between Hausdorff measures for curves and different notions of length in sub-Riemannian geometry is also investigated in [41].

On a contact sub-Riemannian manifold, Popp's volume coincides with the Riemannian volume obtained by "promoting" the Reeb vector field to an orthonormal complement to the distribution. In the general case, unfortunately, the definition is more involved. To the authors' best knowledge, explicit formulæ for Popp's volume appeared, for some specific cases, only in [7, 23, 32].

The goal of this chapter is to prove a general formula for Popp's volume, in terms of any *adapted* frame of the tangent bundle. In order to present the main results here, we briefly introduce some concepts which we will elaborate in details in the subsequent sections. Thus, 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 $\mathscr{D}^i$, where $k_i := \dim \mathscr{D}^i$, 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

$$[X_i, X_j] = \sum_{l=1}^{n} c_{ij}^l X_l. \tag{5}$$

With a standard abuse of notation we call them *structure constants*. For $j = 2, \dots, m$ we define the *adapted structure constants* $b_{i_1 \dots i_j}^l \in C^\infty(M)$ as follows:

$$[X_{i_1}, [X_{i_2}, \dots, [X_{i_{j-1}}, X_{i_j}]]] = \sum_{l=k_{j-1}+1}^{k_j} b_{i_1 i_2 \dots i_j}^l X_l \mod \mathscr{D}^{j-1}, \tag{6}$$

where $1 \leq i_1, \dots, i_j \leq k$. These are a generalization of the c_{ij}^l, with an important difference: the structure constants of Eq. (5) 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. (6) 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 $\mathscr{D}$), and considering the appropriate equivalence class. For $j = 2$, the adapted structure constants can be directly compared to the standard ones. Namely $b_{ij}^l = c_{ij}^l$ when both are defined, that is for $1 \leq i, j \leq k$, $l \geq k+1$.

Then, we define the $k_j - k_{j-1}$ dimensional square matrix B_j as follows:

$$[B_j]^{hl} = \sum_{i_1, i_2, \dots, i_j = 1}^{k} b_{i_1 i_2 \dots i_j}^h b_{i_1 i_2 \dots i_j}^l, \qquad j = 1, \dots, m, \tag{7}$$

[1] Recall that the Hausdorff dimension of a sub-Riemannian manifold M is given by the formula $Q = \sum_{i=1}^m i n_i$, where $n_i := \dim \mathscr{D}_q^i / \mathscr{D}_q^{i-1}$. In particular the Hausdorff dimension is always bigger than the topological dimension.

with the understanding that B_1 is the $k\times k$ identity matrix. It turns out that each B_j is positive definite.

Theorem 10. *Let $X_1,\dots,X_n$ be a local adapted frame, and let $\nu^1,\dots,\nu^n$ be the dual frame. Then Popp's volume $\mathcal{P}$ satisfies*

$$\mathcal{P} = \frac{1}{\sqrt{\prod_j \det B_j}}\, \nu^1 \wedge \ldots \wedge \nu^n, \tag{8}$$

where B_j is defined by (7) *in terms of the adapted structure constants* (6).

To clarify the geometric meaning of Eq. (8), let us consider more closely the case $m = 2$. If $\mathscr{D}$ is a step 2 distribution, we can build a local adapted frame $\{X_1,\dots,X_k,X_{k+1},\dots,X_n\}$ by completing any local orthonormal frame $\{X_1,\dots,X_k\}$ of the distribution to a local frame of the whole tangent bundle. Even though it may not be evident, it turns out that $B_2^{-1}(q)$ is the Gram matrix of the vectors $X_{k+1},\dots,X_n$, seen as elements of $T_qM/\mathscr{D}_q$. The latter has a natural structure of inner product space, induced by the surjective linear map $[,] : \mathscr{D}_q \otimes \mathscr{D}_q \to T_qM/\mathscr{D}_q$ (see Lemma 4.12). Therefore, the function appearing at the beginning of Eq. (8) is the volume of the parallelotope whose edges are $X_1,\dots,X_n$, seen as elements of the orthogonal direct sum $\mathrm{gr}_q(\mathscr{D}) = \mathscr{D}_q \oplus T_qM/\mathscr{D}_q$.

With a volume form at disposal, one can naturally define the associated divergence operator, which acts on vector fields. Moreover, the sub-Riemannian structure allows to define the horizontal gradient of a smooth function. Then, we define a canonical sub-Laplace operator as $\Delta := \mathrm{div} \circ \nabla$, which generalizes the Laplace-Beltrami operator. This is a second order differential operator, which has been studied in [10, 21]. As a corollary to Theorem 10, we obtain a formula for the sub-Laplacian Δ in terms of any local adapted frame.

Corollary 11. *Let $X_1,\dots,X_n$ be a local adapted frame. Let Δ be the canonical sub-Laplacian. Then*

$$\Delta = \sum_{i=1}^k X_i^2 - \left(\frac{1}{2}\sum_{j=1}^m \mathrm{Tr}(B_j^{-1}X_i(B_j)) + \sum_{l=1}^n c_{il}^l \right) X_i,$$

where c_{ij}^l are the structure constants (5)*, and B_j is defined by* (7) *in terms of the adapted structure constants* (6).

If M is a Carnot group (i.e. a connected, simply connected nilpotent group, whose Lie algebra is graded, and whose sub-Riemannian structure is left invariant) the B_j are constant. Moreover, $\forall i$ $\sum_{l=1}^n c_{il}^l = 0$, as a consequence of the graded structure. Then, in this case, the sub-Laplacian is a simple "sum of squares" $\Delta = \sum_{i=1}^k X_i^2$. This is a manifestation of the fact that Carnot groups are to sub-Riemannian geometry as Euclidean spaces are to Riemannian geometry. Indeed, on $\mathbb{R}^n$, the Laplace-Beltrami operator is a simple sum of squares.

More in general, in [10], the authors prove that for left-invariant structures on unimodular Lie groups the sub-Laplacian is a sum of squares.

In the last part of the chapter 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. A (local) diffeomorphism $\phi : M \to M$ is a *(local) isometry* if its differential $\phi_* : TM \to TM$ preserves the sub-Riemannian structure $(\mathscr{D}, \langle\cdot|\cdot\rangle)$, namely

i) $\phi_*(\mathscr{D}_q) = \mathscr{D}_{\phi(q)}$ for all $q \in M$,

ii) $\langle \phi_* X | \phi_* Y \rangle_{\phi(q)} = \langle X|Y\rangle_q$ for all $q \in M$, $X, Y \in \mathscr{D}_q$.

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. $\phi_*(\mathscr{D}^i_q) = \mathscr{D}^i_{\phi(q)}$, for $1 \leq i \leq m$.

Definition. 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 12. *Sub-Riemannian (local) isometries are volume preserving transformations for Popp's volume.*

Proposition 12 may be false for volumes different than Popp's one. We have the following.

Proposition 13. *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 12 holds true.*

Let M be a Lie group. We say that a sub-Riemannian structure $(M, \mathscr{D}, \langle\cdot|\cdot\rangle)$ 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 12 we recover a well-known result (see again [56]).

Corollary 14. *Let* $(M, \mathscr{D}, \langle\cdot|\cdot\rangle)$ *be a left-invariant sub-Riemannian structure. Then Popp's volume is left invariant, i.e.* $L_g^*\mathcal{P} = \mathcal{P}$ *for every* $g \in M$*.*

Propositions 12, 13 and Corollary 14 should shed some light about which is the "most natural" volume for sub-Riemannian manifold.

Chapter 1

The curvature of optimal control problems

1.1 Introduction

The curvature discussed here is a rather far going generalization of the Riemannian sectional curvature. We define it for a wide class of optimal control problems: a unified framework including geometric structures such as Riemannian, sub-Riemannian, Finsler and sub-Finsler structures; a special attention is paid to the sub-Riemannian (or Carnot–Carathéodory) metric spaces. Our construction of the curvature is direct and naive, and it is similar to the original approach of Riemann. Surprisingly, it works in a very general setting and, in particular, for *all* sub-Riemannian spaces.

The main idea is that the curvature is related with the distortion of the cost along the extremals. To fix the ideas, let's see how we can develop this concept starting from Riemannian geometry, the classical realm of curvature. In his seminal paper [62], Riemann thought at the curvature as a quantity that measures "how much the geometrical structure is different from the flat one". In a more modern language, this concept can be expressed very elegantly as follows. Let M be a Riemannian manifold. Consider two unit speed geodesics γ_v, γ_w, with initial tangent vector v and w, and let $\mathsf{d} : M \times M \to \mathbb{R}$ be the Riemannian distance. Then

$$\mathsf{d}^2(\gamma_v(t), \gamma_w(t)) = 2(1-\cos\theta)t^2 \left(1 - \tfrac{1}{3}\operatorname{Sec}(v,w)\cos^2(\theta/2)t^2\right) + O(t^6), \tag{1.1}$$

where $\operatorname{Sec}(v,w)$ is the sectional curvature of the plane spanned by v and w, and θ is the Riemannian angle between the two vectors (see [69]). The same naive idea can be extended to the more general realm of optimal control problems where, at least in principle, the Riemannian geodesics are replaced by optimal trajectories, and the squared distance is replaced by the value (or cost) function.

Still, some serious difficulties arise. Riemannian extremals are parametrized by their initial vector, and this is no longer true in the more general setting (or also in the closer sub-Riemannian one). More importantly, the cost function may be non-smooth on the diagonal. For example, in the sub-Riemannian case, the cost function is essentially the squared distance $\mathsf{d}^2 : M \times M \to \mathbb{R}$, that is *never* smooth on the diagonal. Thus, a naive Taylor expansion in the spirit of Eq. (1.1) is not possible.

Now that we have introduced the flavour of the main idea, we explain in more detail the nature of our curvature by describing the case of a contact sub-Riemannian structure. Then we move to the general construction.

Let M be an odd-dimensional Riemannian manifold endowed with a contact vector distribution $\mathscr{D} \subset TM$. Given $x_0, x_1 \in M$, the contact sub-Riemannian distance $\mathsf{d}(x_0, x_1)$ is the infimum of the lengths of Legendrian curves connecting x_0 and x_1. Recall that Legendrian curves are just integral curves of the distribution $\mathscr{D}$. The metric d is easily realized as the limit of a family of Riemannian metrics d^ε as $\varepsilon \to 0$. To define d^ε we start from the original Riemannian structure on M, keep fixed the length of vectors from $\mathscr{D}$ and multiply by $\frac{1}{\varepsilon}$ the length of the orthogonal to $\mathscr{D}$ tangent vectors to M. It is easy to see that $\mathsf{d}^\varepsilon \to \mathsf{d}$ uniformly on compacts in $M \times M$ as $\varepsilon \to 0$.

The distance converges, what about the curvature? Let ω be a contact differential form that annihilates $\mathscr{D}$, i.e. $\mathscr{D} = \omega^\perp$. Given $v_1, v_2 \in T_xM$, $v_1 \wedge v_2 \neq 0$, we denote by $\mathrm{Sec}^\varepsilon(v_1 \wedge v_2)$ the sectional curvature for the metric d^ε and section $\mathrm{span}\{v_1, v_2\}$. It is not hard to show that $\mathrm{Sec}^\varepsilon(v_1 \wedge v_2) \to -\infty$ if $v_1, v_2 \in \mathscr{D}$ and $d\omega(v_1, v_2) \neq 0$. Moreover, $\mathrm{Ric}^\varepsilon(v) \to -\infty$ as $\varepsilon \to 0$ for any nonzero vector $v \in \mathscr{D}$, where Ric^ε is the Ricci curvature for the metric d^ε. On the other hand, the distance between x and the conjugate locus of x tends to 0 as $\varepsilon \to 0$ and $\mathrm{Sec}^\varepsilon(v_1 \wedge v_2)$ tends to $+\infty$ for some $v_1, v_2 \in T_xM$, as well as $\mathrm{Ric}^\varepsilon(v)$ for some $v \in T_xM$.

What about the geodesics? For any $\varepsilon > 0$ and any $v \in T_xM$ there is a unique geodesic of the Riemannian metric d^ε that starts from x with velocity v. On the other hand, the velocities of all geodesics of the limit metric d belong to $\mathscr{D}$ and for any nonzero vector $v \in \mathscr{D}$ there exists a one-parametric family of geodesics whose initial velocity is equal to v. Too bad up to now, and here is the first encouraging fact: the family of geodesic flows converges if we re-write it as a family of flows on the cotangent bundle.

Indeed, any Riemannian structure on M induces a self-adjoint isomorphism $G : TM \to T^*M$, where $\langle Gv, v\rangle$ is the square of the length of the vector $v \in TM$, and $\langle \cdot, \cdot \rangle$ denotes the standard pairing between tangent and cotangent vectors. The geodesic flow, treated as flow on T^*M is a Hamiltonian flow associated with the Hamiltonian function $H : T^*M \to \mathbb{R}$, where $H(\lambda) = \frac{1}{2}\langle \lambda, G^{-1}\lambda\rangle$, $\lambda \in T^*M$. Let $(\lambda(t), \gamma(t))$ be a trajectory of the Hamiltonian flow, with $\lambda(t) \in T^*_{\gamma(t)}M$. The square of the Riemannian distance from x_0 is a smooth function on a neighbourhood of x_0 in M and the differential of this function at $\gamma(t)$ is equal to $2t\lambda(t)$ for any small $t \geq 0$. Let H^ε be the Hamiltonian corresponding to the metric d^ε. It is easy to see that H^ε converges with all derivatives to a Hamiltonian H^0. Moreover, geodesics of the limit sub-Riemannian metric are just projections to M of the trajectories of the Hamiltonian flow on T^*M associated to H^0.

We can recover the Riemannian curvature from the asymptotic expansion of the square of the distance from x_0 along a geodesic: this is essentially what Riemann did. Then we can write a similar expansion for the square of the limit sub-Riemannian distance to get an idea of the curvature in this case. Note that the metrics d^ε converge to d with all derivatives in any point of $M \times M$, where d is smooth. The metrics d^ε are not smooth at the diagonal but their squares are smooth. The point is that no power of d is smooth at the diagonal! Nevertheless, the desired asymptotic expansion can be controlled.

Fix a point $x_0 \in M$ and $\lambda_0 \in T^*_{x_0}M$ such that $\langle \lambda_0, \mathscr{D} \rangle \neq 0$. Let $(\lambda^\varepsilon(t), \gamma^\varepsilon(t))$, for $\varepsilon \geq 0$, be the trajectory of the Hamiltonian flow associated to the Hamiltonian H^ε and initial condition (λ_0, x_0). We set:

$$c_t^\varepsilon(x) := -\frac{1}{2t}(\mathsf{d}^\varepsilon)^2(x, \gamma^\varepsilon(t)) \text{ if } \varepsilon > 0, \qquad c_t^0(x) := -\frac{1}{2t}\mathsf{d}^2(x, \gamma^0(t)).$$

There exists an interval $(0, \delta)$ such that the functions c_t^ε are smooth at x_0 for all $t \in (0, \delta)$ and all $\varepsilon \geq 0$. Moreover, $d_{x_0} c_t^\varepsilon = \lambda_0$. Let $\dot{c}_t^\varepsilon = \frac{\partial}{\partial t} c_t^\varepsilon$, then $d_{x_0} \dot{c}_t^\varepsilon = 0$. In other words, x_0 is a critical point of the function $\dot{c}_t^\varepsilon$ and its Hessian $d^2_{x_0} \dot{c}_t^\varepsilon$ is a well-defined quadratic form on $T_{x_0}M$. Recall that $\varepsilon = 0$ is available, but t must be positive. We are going to study the asymptotics of the family of quadratic forms $d^2_{x_0} \dot{c}_t^\varepsilon$ as $t \to 0$ for fixed ε. This asymptotic is a little bit different for $\varepsilon > 0$ and $\varepsilon = 0$. The difference reflects the structural difference of the Riemannian and sub-Riemannian metrics and emphasises the role of the curvature.

Given $v, w \in T_xM$, $\varepsilon > 0$, we denote $\langle v|w \rangle_\varepsilon = \langle G^\varepsilon v, w \rangle$ the inner product generating d^ε. Recall that $\langle v|v \rangle_\varepsilon$ does not depend on ε if $v \in \mathscr{D}$ and $\langle v|v \rangle_\varepsilon \to \infty$ $(\varepsilon \to 0)$ if $v \notin \mathscr{D}$; we will write $|v|^2 := \langle v|v \rangle_\varepsilon$ in the first case. For fixed $\varepsilon > 0$, we have:

$$d^2_{x_0} \dot{c}_t^\varepsilon(v) = \frac{1}{t^2}\langle v|v \rangle_\varepsilon + \frac{1}{3}\langle R^\varepsilon(\dot{\gamma}^\varepsilon, v)\dot{\gamma}^\varepsilon)|v \rangle_\varepsilon + O(t), \qquad v \in T_{x_0}M,$$

where $\dot{\gamma}^\varepsilon = \dot{\gamma}^\varepsilon(0)$ and R^ε is the Riemannian curvature tensor of the metric d^ε. For $\varepsilon = 0$, only vectors $v \in \mathscr{D}$ have a finite length and the above expansion is modified as follows:

$$d^2_{x_0} \dot{c}_t^0(v) = \frac{1}{t^2}\mathcal{I}_\gamma(v) + \frac{1}{3}\mathcal{R}_\gamma(v) + O(t), \qquad v \in \mathscr{D} \cap T_{x_0}M,$$

where $\mathcal{I}_\gamma(v) \geq |v|^2$ and $\mathcal{R}_\gamma$ is the *sub-Riemannian curvature* at x_0 along the geodesic $\gamma = \gamma^0$. Both $\mathcal{I}_\gamma$ and $\mathcal{R}_\gamma$ are quadratic forms on $\mathscr{D}_{x_0} := \mathscr{D} \cap T_{x_0}M$. The principal "structural" term $\mathcal{I}_\gamma$ has the following properties:

$$\max\{\mathcal{I}_\gamma(v) |\, v \in \mathscr{D}_{x_0},\ |v|^2 = 1\} = 4,$$
$$\mathcal{I}_\gamma(v) = |v|^2 \text{ if and only if } d\omega(v, \dot{\gamma}(0)) = 0.$$

In other words, the symmetric operator on $\mathscr{D}_{x_0}$ associated with the quadratic form $\mathcal{I}_\gamma$ has eigenvalue 1 of multiplicity $\dim \mathscr{D}_{x_0} - 1$ and eigenvalue 4 of multiplicity 1. The trace of this operator, which, in this case, does not depend on γ, equals $\dim \mathscr{D}_{x_0} + 3$. This trace has a simple geometric interpretation, it is equal to the *geodesic dimension* of the sub-Riemannian space.

The geodesic dimension is defined as follows. Let $\Omega \subset M$ be a bounded and measurable subset of positive volume and let $\Omega_{x_0,t}$, for $0 \leq t \leq 1$, be a family of subsets obtained from Ω by the homothety of Ω with respect to a fixed point x_0 along the shortest geodesics connecting x_0 with the points of Ω, so that $\Omega_{x_0,0} = \{x_0\}$, $\Omega_{x_0,1} = \Omega$. The volume of $\Omega_{x_0,t}$ has order $t^{\mathcal{N}_{x_0}}$, where $\mathcal{N}_{x_0}$ is the geodesic dimension at x_0 (see Section 1.5.6 for details).

Note that the topological dimension of our contact sub-Riemannian space is $\dim \mathscr{D}_{x_0} + 1$ and the Hausdorff dimension is $\dim \mathscr{D}_{x_0} + 2$. All three dimensions are obviously equal for Riemannian or Finsler manifolds. The structure of the term $\mathcal{I}_\gamma$ and comparison of the asymptotic expansions of $d^2_{x_0}\dot{c}^\varepsilon_t$ for $\varepsilon > 0$ and $\varepsilon = 0$ explains why sectional curvature goes to $-\infty$ for certain sections.

The curvature operator which we define can be computed in terms of the symplectic invariants of the so-called Jacobi curve, namely a curve in the Lagrange Grassmannian related with the linearisation of the Hamiltonian flow. These symplectic invariants can be computed, in principle, via an algorithm which is, however, quite hard to implement. Explicit computations of the contact sub-Riemannian curvature will appear in a forthcoming paper. In the current chapter we deal with the general setting. A precise construction in full generality is presented in the forthcoming sections but, since this chapter is long, we find it worth to briefly describe the main ideas in the introduction.

Let M be a smooth manifold, $\mathscr{D} \subset TM$ be a vector distribution (not necessarily contact), f_0 be a vector field on M and $L : TM \to M$ be a Tonelli Lagrangian (i.e. $L|_{T_xM}$ has a superlinear growth and its Hessian is positive definite for any $x \in M$). *Admissible paths* on M are curves whose velocities belong to the "affine distribution" $f_0 + \mathscr{D}$. Let $\mathcal{A}_t$ be the space of admissible paths defined on the segment $[0, t]$ and $N_t = \{(\gamma(0), \gamma(t)) : \gamma \in \mathcal{A}_t\} \subset M \times M$. The optimal cost (or action) function $S_t : N_t \to \mathbb{R}$ is defined as follows:

$$S_t(x, y) = \inf \left\{ \int_0^t L(\dot\gamma(\tau))\, d\tau : \gamma \in \mathcal{A}_t,\ \gamma(0) = x,\ \gamma(t) = y \right\}.$$

The space $\mathcal{A}_t$ equipped with the $W^{1,\infty}$-topology is a smooth Banach manifold; the functional $J_t : \gamma \mapsto \int_0^t L(\dot\gamma(\tau))\, d\tau$ and the evaluation maps $F_\tau : \gamma \mapsto \gamma(\tau)$ are smooth on $\mathcal{A}_t$.

The optimal cost $S_t(x, y)$ is the solution of the conditional minimum problem for the functional J_t under conditions $F_0(\gamma) = x,\ F_t(\gamma) = y$. The Lagrange multipliers rule for this problem reads:

$$d_\gamma J_t = \lambda_t D_\gamma F_t - \lambda_0 D_\gamma F_0. \tag{1.2}$$

Here λ_t and λ_0 are "Lagrange multipliers", $\lambda_t \in T^*_{\gamma(t)}M,\ \lambda_0 \in T^*_{\gamma(0)}M$. We have:

$$D_\gamma F_t : T_\gamma \mathcal{A}_t \to T_{\gamma(t)}M, \qquad \lambda_t : T_{\gamma(t)}M \to \mathbb{R},$$

and the composition $\lambda_t D_\gamma F_t$ is a linear functional on $T_\gamma\mathcal{A}_t$. Moreover, Eq. (1.2) implies that

$$d_\gamma J_\tau = \lambda_\tau D_\gamma F_\tau - \lambda_0 D_\gamma F_0, \tag{1.3}$$

for some $\lambda_\tau \in T^*_{\gamma(\tau)}M$ and any $\tau \in [0, t]$. The curve $\tau \mapsto \lambda_\tau$ is a trajectory of the Hamiltonian system associated to the Hamiltonian $H : T^*M \to \mathbb{R}$ defined by

$$H(\lambda) = \max_{v \in f_0(x) + \mathscr{D}_x} \left(\langle \lambda, v \rangle - L(v) \right), \qquad \lambda \in T^*_xM,\, x \in M.$$

Moreover, any trajectory of this Hamiltonian system satisfies relation (1.3), where γ is the projection of the trajectory to M. Trajectories of the Hamiltonian system are called *normal extremals* and their projections to M are called *normal extremal trajectories*.

We recover the sub-Riemannian setting when $f_0 = 0$, $L(v) = \frac{1}{2}\langle Gv, v\rangle$. In this case, the optimal cost is related with the sub-Riemannian distance $S_t(x, y) = \frac{1}{2t}\mathsf{d}^2(x, y)$, and normal extremal trajectories are normal sub-Riemannian geodesics.

Let γ be an admissible path; the germ of γ at the point $x_0 = \gamma(0)$ defines a flag in $T_{x_0}M$ $\{0\} = \mathscr{F}^0_\gamma \subset \mathscr{F}^1_\gamma \subset \mathscr{F}^2_\gamma \subset \ldots \subset T_{x_0}M$ in the following way. Let V be a section of the vector distribution $\mathscr{D}$ such that $\dot\gamma(t) = f_0(\gamma(t)) + V(\gamma(t)),\ t \geq 0$, and $P_{0,t}$ be the local flow on M generated by the vector field $f_0 + V$; then $\gamma(t) = P_{0,t}(\gamma(0))$. We set:

$$\mathscr{F}^i_\gamma = \operatorname{span}\left\{ \left.\frac{d^j}{dt^j}\right|_{t=0} (P_{0,t})_*^{-1}\mathscr{D}_{\gamma(t)} : j = 0, \ldots, i-1 \right\}.$$

The flag $\mathscr{F}^i_\gamma$ depends only on the germs of $f_0 + \mathscr{D}$ and γ at the initial point x_0.

A normal extremal trajectory γ is called *ample* if $\mathscr{F}^m_\gamma = T_{x_0}M$ for some $m > 0$. If γ is ample, then $J_t(\gamma) = S_t(x_0, \gamma(t))$ for all sufficiently small $t > 0$ and S_t is a smooth function in a neighbourhood of $(\gamma(0), \gamma(t))$. Moreover, $\frac{\partial S_t}{\partial y}\big|_{y=\gamma(t)} = \lambda_t$, $\frac{\partial S_t}{\partial x}\big|_{x=\gamma(0)} = -\lambda_0$, where λ_t is the normal extremal whose projection is γ.

We set $c_t(x) := -S_t(x, \gamma(t))$; then $d_{x_0}c_t = \lambda_0$ for any $t > 0$ and x_0 is a critical point of the function $\dot c_t$. The Hessian of this function $d^2_{x_0}\dot c_t$ is a well-defined quadratic form on $T_{x_0}M$. We are going to write an asymptotic expansion of $d^2_{x_0}\dot c_t\big|_{\mathscr{D}_{x_0}}$ as $t \to 0$ (see Theorem 1.A):

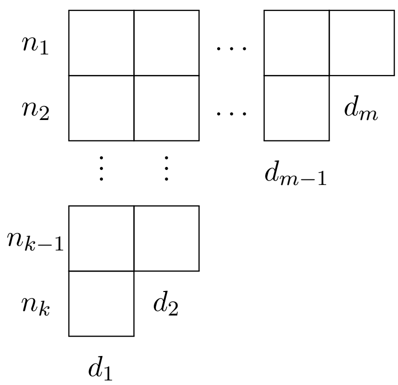
$$d^2_{x_0}\dot c_t(v) = \frac{1}{t^2}\mathcal{I}_\gamma(v) + \frac{1}{3}\mathcal{R}_\gamma(v) + O(t), \qquad \forall v \in \mathscr{D}_{x_0}.$$

Now we introduce a natural Euclidean structure on $T_{x_0}M$. Recall that $L|_{T_{x_0}M}$ is a strictly convex function, and $d^2_w(L|_{T_{x_0}M})$ is a positive definite quadratic form on $T_{x_0}M$, $\forall\, w \in T_{x_0}M$. If we set $|v|^2_\gamma = d^2_{\dot\gamma(0)}(L|_{T_{x_0}M})(v),\ v \in T_{x_0}M$ we have the inequality

$$\mathcal{I}_\gamma(v) \geq |v|^2_\gamma, \qquad \forall\, v \in \mathscr{D}_{x_0}.$$

The inequality $\mathcal{I}_\gamma(v) \geq |v|^2_\gamma$ means that the eigenvalues of the symmetric operator on $\mathscr{D}_{x_0}$ associated with the quadratic form $\mathcal{I}_\gamma$ are greater or equal than 1. The quadratic form $\mathcal{R}_\gamma$ is the *curvature* of our constrained variational problem in the direction of the extremal trajectory γ.

A mild regularity assumption allows to explicitly compute the eigenvalues of $\mathcal{I}_\gamma$. We set $\gamma_\varepsilon(t) = \gamma(\varepsilon + t)$ and assume that $\dim \mathscr{F}^i_{\gamma_\varepsilon} = \dim \mathscr{F}^i_\gamma$ for all sufficiently small $\varepsilon \geq 0$ and all i. It turns out that $d_i = \dim \mathscr{F}^i_\gamma - \dim \mathscr{F}^{i-1}_\gamma$, for $i \geq 1$ is a non-increasing sequence of natural numbers with $d_1 = \dim \mathscr{D}_{x_0} = k$. We draw a Young tableau with d_i blocks in the i-th column and we define $n_1, \ldots, n_k$ as the lengths of its rows (that may depend on γ).



The eigenvalues of the symmetric operator $\mathcal{I}_\gamma$ are $n_1^2, \dots, n_k^2$ (see Theorem 1.B). Some of these numbers may be equal (in the case of multiple eigenvalues) and are all equal to 1 in the Riemannian case. In the sub-Riemannian setting, the trace of $\mathcal{I}_\gamma$ is

$$\operatorname{tr} \mathcal{I}_\gamma = n_1^2 + \cdots + n_k^2 = \sum_{i=1}^{m} (2i-1) d_i,$$

When computed for the generic sub-Riemannian geodesic, this number is actually constant and depends only on x_0. This is what we called the *geodesic dimension* $\mathcal{N}_{x_0}$ of the manifolds. For Riemannian manifolds, this invariant is always equal to the topological dimension. For the $2n+1$-dimensional Heisenberg group, $\mathcal{N} = 2n+3$ (constantly on the manifold). Thus, the geodesic dimension is a new invariant, different from both the topological and the Hausdorff dimension of the sub-Riemannian space.

Let's see how this new dimension is related with the geometry of the sub-Riemannian manifold. Fix any smooth measure μ on the manifold, and let Ω be a measurable set with $0 < \mu(\Omega) < +\infty$. Fix $x_0 \in M$. For simplicity, assume that Ω does not intersect the cut locus of x_0. We define the *homothety* with center x_0 at time $t \in [0,1]$ of the set Ω as follows. Let $x \in \Omega$, and consider the unique geodesic γ such that $\gamma(0) = x_0$ and $\gamma(1) = x$. Then the homothethy of x at time t with center x_0 is the point $\gamma(t)$. Doing this for any $x \in \Omega$ defines a new set Ω_t, such that $\Omega_0 = \{x_0\}$ and $\Omega_1 = \Omega$. The main result of Sec. 1.5.6 is the following (see Theorem 1.D):

$$\mu(\Omega_t) \sim t^{\mathcal{N}_{x_0}}, \qquad t \to 0.$$

Namely geodesic dimension represents the critical exponent such that the volume of a measurable set shrinks to zero as $t \to 0$ along a sub-Riemannian homothety.

The last result of this chapter is related with the intrinsic Laplacian of a sub-Riemannian manifold. We sketch the general construction. For any fixed smooth volume form μ, we define the μ-divergence of a vector field $X \in \operatorname{Vec}(M)$ by the following formula

$$\mathcal{L}_X \mu = \operatorname{div}_\mu(X)\mu,$$

where $\mathcal{L}$ represents the Lie derivative. Moreover, for any $f \in C^\infty(M)$, we define the sub-Riemannian gradient ∇f as the unique horizontal vector field such that $g(\nabla f, \cdot) = df(\cdot)$. Thus, we define the μ-Laplacian as $\Delta_\mu f := \operatorname{div}_\mu \nabla f$. In the Riemannian setting, when μ is the Riemannian volume form, this construction leads to the familiar Laplace-Beltrami operator. In the sub-Riemannian setting, one can choose μ to be the canonical Popp's volume (see Chapter 4), and obtain an intrinsic sub-Laplacian operator. Still, we prefer to leave μ general here. The main result of Sec. 1.5 is the relation between the curvature and the geodesic dimension with the asymptotic behaviour of the sub-Laplacian of the sub-Riemannian squared distance from a geodesic, i.e. the function

$$\mathfrak{f}_t(\cdot) := -t c_t(\cdot) = \frac{1}{2}\mathsf{d}^2(\cdot, \gamma(t)), \qquad t \in (0,1],$$

where $\gamma(t)$ is an ample sub-Riemannian geodesic such that $\gamma(0) = x_0$. In particular we prove that (see Theorem 1.C)

$$\Delta_\mu \mathfrak{f}_t = \operatorname{tr} \mathcal{I}_\gamma - \dot{g}(0)t - \frac{1}{3} \operatorname{tr} \mathcal{R}_\gamma t^2 + O(t^3),$$

where $g : [0, T] \to \mathbb{R}$ is a smooth function that depends on the choice of the volume μ, whose precise definition is not needed here. This, in particular, implies that

$$\lim_{t \to 0} \Delta_\mu \mathfrak{f}_t \big|_{x_0} = \operatorname{tr} \mathcal{I}_\gamma,$$
$$\left.\frac{d^2}{dt^2}\right|_{t=0} \Delta_\mu \mathfrak{f}_t \big|_{x_0} = -\frac{2}{3} \operatorname{tr} \mathcal{R}_\gamma.$$

The construction of the curvature presented here was preceded by a rather long research line (see [3, 16–18, 49, 71]). For what concerns the alternative approaches to this topic, in recent years, several efforts have been made to introduce a notion of curvature to non-Riemannian situations, such as sub-Riemannian manifolds and, more in general, metric measure spaces. Motivated by the lack of classical Riemannian tools (such as the Levi-Civita connection and the theory of Jacobi fields) different approaches have been explored in order to extend some classical results in geometric analysis to such structures. In particular, to this extent, many synthetic notions of generalized Ricci curvature bound have been introduced. For instance, one can see [29, 30] and references therein for a heat equation approach to the generalization of the curvature-dimension inequality and [19, 51, 67, 68] and references therein for an optimal transport approach to the generalization of Ricci curvature.

1.1.1 Structure of the chapter

In Sections 1.2–1.4 we give a detailed exposition of the main constructions in a more general and flexible setting than in this introduction. Section 1.5 is devoted to the specification to the case of sub-Riemannian spaces and to some further results: an estimate of the Young tableau in terms of the nilpotent approximation (Lemma 1.67), an asymptotic expansion of the sub-Laplacian applied to the square of the distance (Theorem 1.C), the computation of the geodesic dimension (Theorem 1.D).

Before entering into details of the proofs, we end Section 1.5 by repeating our construction for one of the simplest sub-Riemannian structures: the Heisenberg group. In particular, we recover by a direct computation the results of Theorems 1.A, 1.B and 1.C.

The proofs of the main results are concentrated in Sections 1.6–1.8 where we introduce the main technical tools: Jacobi curves, their symplectic invariants and Li–Zelenko structural equations.

1.2 General setting

In this section we introduce a general framework that allows to treat smooth control system on a manifold in a coordinate free way, i.e. invariant under state and feedback transformations.

For the sake of simplicity, we will restrict our definition to the case of nonlinear affine control systems, although the construction of this section can be extended to any smooth control system (see [3]).

1.2.1 Affine control systems

Definition 1.1. Let M be a connected smooth n-dimensional manifold. An *affine control system* on M is a pair $(\mathbb{U}, f)$ where:

(i) $\mathbb{U}$ is a smooth rank k vector bundle with base M and fiber $\mathbb{U}_x$ i.e., for every $x \in M$, $\mathbb{U}_x$ is a k-dimensional vector space,

(ii) $f : \mathbb{U} \to TM$ is a smooth affine morphism of vector bundles, i.e. the diagram (1.4) is commutative and f is *affine* on fibers.

$$\begin{array}{ccc} \mathbb{U} & \xrightarrow{f} & TM \\ & \searrow^{\pi_{\mathbb{U}}} & \downarrow \pi \\ & & M \end{array} \tag{1.4}$$

The maps $\pi_{\mathbb{U}}$ and π are the canonical projections of the vector bundles $\mathbb{U}$ and TM, respectively.

We denote points in $\mathbb{U}$ as pairs (x, u), where $x \in M$ and $u \in \mathbb{U}_x$ is an element of the fiber. According to this notation, the image of the point (x, u) through f is $f(x, u)$ or $f_u(x)$ and we prefer the second one when we want to emphasize f_u as a vector on T_xM. Finally, let $L^\infty([0, T], \mathbb{U})$ be the set of measurable, essentially bounded functions $u : [0, T] \to \mathbb{U}$.

Definition 1.2. A Lipschitz curve $\gamma : [0, T] \to M$ is said to be *admissible* for the control system if there exists a *control* $u \in L^\infty([0, T], \mathbb{U})$ such that $\pi_{\mathbb{U}} \circ u = \gamma$ and

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

The pair (γ, u) of an admissible curve γ and its control u is called *admissible pair*.

We denote by $\overline{f} : \mathbb{U} \to TM$ the linear bundle morphism induced by f. In other words we write $f(x, u) = f_0(x) + \overline{f}(x, u)$, where $f_0(x) := f(x, 0)$ is the image of the zero section. In terms of a local frame for $\mathbb{U}$, $\overline{f}(x, u) = \sum_{i=1}^k u_i f_i(x)$.

Definition 1.3. The *distribution* $\mathscr{D} \subset TM$ is the family of subspaces

$$\mathscr{D} = \{\mathscr{D}_x\}_{x \in M}, \qquad \text{where} \qquad \mathscr{D}_x := \overline{f}(\mathbb{U}_x) \subset T_xM.$$

The family of *horizontal vector fields* $\overline{\mathscr{D}} \subset \mathrm{Vec}(M)$ is

$$\overline{\mathscr{D}} = \mathrm{span}\left\{\overline{f} \circ \sigma,\ \sigma : M \to \mathbb{U} \text{ is a smooth section of } \mathbb{U}\right\}.$$

Observe that, if the rank of $\overline{f}$ is not constant, $\mathscr{D}$ is not a sub-bundle of TM. Therefore the dimension of $\mathscr{D}_x$, in general, depends on $x \in M$.

Given a smooth function $L : \mathbb{U} \to \mathbb{R}$, called a *Lagrangian*, the *cost functional at time* T, called $J_T : L^\infty([0,T], \mathbb{U}) \to \mathbb{R}$, is defined by

$$J_T(u) := \int_0^T L(\gamma(t), u(t))dt,$$

where $\gamma(t) = \pi(u(t))$. We are interested in the problem of minimizing the cost among all admissible pairs (γ, u) that join two fixed points $x_0, x_1 \in M$ in time T. This corresponds to the optimal control problem

$$\begin{aligned} &\dot{x} = f(x,u) = f_0(x) + \sum_{i=1}^k u_i f_i(x), \qquad x \in M, \\ &x(0) = x_0,\ x(T) = x_1, \qquad J_T(u) \to \min, \end{aligned} \tag{1.5}$$

where we have chosen some local trivialization of $\mathbb{U}$.

Definition 1.4. Let $M' \subset M$ be an open subset with compact closure. For $x_0, x_1 \in M'$ and $T > 0$, we define the *value function*

$$S_T(x_0, x_1) := \inf\{J_T(u) \mid (\gamma, u) \text{ admissible pair},\ \gamma(0) = x_0,\ \gamma(T) = x_1,\ \gamma \subset M'\}.$$

The value function depends on the choice of a relatively compact subset $M' \subset M$. This choice, which is purely technical, is related with Theorem 1.19, concerning the regularity properties of S. We stress that all the objects defined in this chapter by using the value function do not depend on the choice of M'.

Assumptions. In what follows we make the following general assumptions:

(A1) The affine control system is *bracket generating*, namely

$$\text{Lie}_x \left\{(\text{ad}\, f_0)^i\, \overline{\mathscr{D}} \mid i \in \mathbb{N}\right\} = T_x M, \qquad \forall\, x \in M, \tag{1.6}$$

where $(\text{ad}\, X)Y = [X, Y]$ is the Lie bracket of two vector fields and $\text{Lie}_x \mathcal{F}$ denotes the Lie algebra generated by a family of vector fields $\mathcal{F}$, computed at the point x. Observe that the vector field f_0 is not included in the generators of the Lie algebra (1.6).

(A2) The function $L : \mathbb{U} \to \mathbb{R}$ is a *Tonelli Lagrangian*, i.e. it satisfies

- (A2.a) The Hessian of $L|_{\mathbb{U}_x}$ is positive definite for all $x \in M$. In particular, $L|_{\mathbb{U}_x}$ is strictly convex.
- (A2.b) L has superlinear growth, i.e. $L(x,u)/|u| \to +\infty$ when $|u| \to +\infty$.

Assumptions (A1) and (A2) are necessary conditions in order to have a nontrivial set of strictly normal minimizer and allow us to introduce a well defined smooth Hamiltonian (see Section 1.3).

State-feedback equivalence

All our considerations will be local. Hence, up to restricting our attention to a trivializable neighbourhood of M, we can assume that $\mathbb{U} \simeq M \times \mathbb{R}^k$. By choosing a basis of $\mathbb{R}^k$, we can write $f(x,u) = f_0(x) + \sum_{i=1}^k u_i f_i(x)$. Then, a Lipschitz curve $\gamma : [0,T] \to M$ is admissible if there exists a measurable, essentially bounded control $u : [0,T] \to \mathbb{R}^k$ such that

$$\dot\gamma(t) = f_0(\gamma(t)) + \sum_{i=1}^k u_i(t) f_i(\gamma(t)), \qquad \text{for a.e. } t \in [0,T].$$

We use the notation $u \in L^\infty([0,T],\mathbb{R}^k)$ to denote a measurable, essentially bounded control with values in $\mathbb{R}^k$. By choosing another (local) trivialization of $\mathbb{U}$, or another basis of $\mathbb{R}^k$, we obtain a different *presentation* of the same affine control system. Besides, by acting on the underlying manifold M via diffeomorphisms, we obtain equivalent affine control system starting from a given one. The following definition formalizes the concept of equivalent control systems.

Definition 1.5. Let $(\mathbb{U}, f)$ and $(\mathbb{U}', f')$ be two affine control systems on the same manifold M. A *state-feedback transformation* is a pair (ϕ, ψ), where $\phi : M \to M$ is a diffeomorphism and $\psi : \mathbb{U} \to \mathbb{U}'$ an invertible affine bundle map, such that the following diagram is commutative.

$$\begin{array}{ccc} \mathbb{U} & \xrightarrow{\ f\ } & TM \\ {\scriptstyle\psi}\downarrow & & \downarrow{\scriptstyle\phi_*} \\ \mathbb{U}' & \xrightarrow[\ f'\ ]{} & TM \end{array} \tag{1.7}$$

In other words, $\phi_* f(x,u) = f'(\phi(x), \psi(x,u))$ for every $(x,u) \in \mathbb{U}$. In this case $(\mathbb{U}, f)$ and $(\mathbb{U}', f')$ are said *state-feedback equivalent*.

Notice that, if $(\mathbb{U}, f)$ and $(\mathbb{U}', f')$ are state-feedback equivalent, then $\operatorname{rank}\mathbb{U} = \operatorname{rank}\mathbb{U}'$. Moreover, different presentations of the same control systems are indeed feedback equivalent (i.e. related by a state-feedback transformation with $\phi = \mathbb{I}$). Definition 1.5 corresponds to the classical notion of point-dependent reparametrization of the controls. The next lemma states that a state-feedback transformation preserves admissible curves.

Lemma 1.6. *Let $\gamma_{x_0,u}$ be the admissible curve starting from x_0 and associated with u. Then $\phi(\gamma_{x_0,u}(t)) = \gamma_{\phi(x_0),v}(t)$ where $v(t) = \psi(x(t), u(t))$.*

Proof. Denote $x(t) = \gamma_{x_0,u}(t)$ and set $y(t) := \phi(x(t))$. Then, by definition, $\dot x(t) = f(x(t), u(t))$ and $x(0) = x_0$. Hence $y(0) = \phi(x_0)$ and

$$\dot y(t) = \phi_* f(x(t), u(t)) = f'(\phi(x(t)), \psi(x(t), u(t))) = f'(y(t), v(t)).$$

□

Remark 1.7. Notice that every state-feedback transformation (ϕ, ψ) can be written as a composition of a pure state one, i.e. with $\psi = \mathbb{I}$, and a pure feedback one, i.e. with $\phi = \mathbb{I}$. For later convenience, let us discuss how two feedback equivalent systems are related. Consider a presentation of an affine control system

$$\dot{x} = f(x,u) = f_0(x) + \sum_{i=1}^{k} u_i f_i(x).$$

By the commutativity of diagram (1.7), a feedback transformation writes

$$\begin{cases} u' = \psi(x,u) \\ x' = \phi(x) \end{cases} \qquad u_i' = \psi_i(x,u) = \psi_{i,0}(x) + \sum_{j=1}^{k} \psi_{i,j}(x) u_j, \qquad i = 1,\dots,k,$$

where $\psi_{i,0}$ and $\psi_{i,j}$ denote, respectively, the affine and the linear part of the i-th component of ψ. In particular, for a pure feedback transformation, the original system is equivalent to

$$\dot{x} = f'(x,u') = f_0'(x) + \sum_{i=1}^{k} u_i' f_i'(x),$$

where $f_0(x) := f_0'(x) + \sum_{i=1}^{k} \psi_{i,0}(x) f_i'(x)$ and $f_i(x) := \sum_{j=1}^{k} \psi_{j,i}(x) f_j'(x)$.

We conclude this section recalling some well known facts about non-autonomous flows. By Caratheodory Theorem, for every control $u \in L^\infty([0,T],\mathbb{R}^k)$ and every initial condition $x_0 \in M$, there exists a unique Lipschitz solution to the Cauchy problem

$$\begin{cases} \dot{\gamma}(t) = f_0(\gamma(t)) + \sum_{i=1}^{k} u_i(t) f_i(\gamma(t)), \\ \gamma(0) = x_0, \end{cases} \tag{1.8}$$

defined for small time (see, e.g. [15,61]). We denote such a solution by $\gamma_{x_0,u}$ (or simply γ_u when the base point x_0 is fixed). Moreover, for a fixed control $u \in L^\infty([0,T],\mathbb{R}^k)$, it is well defined the family of diffeomorphisms $P_{0,t} : M \to M$, given by $P_{0,t}(x) := \gamma_{x,u}(t)$, which is Lipschitz with respect to t. Analogously one can define the flow $P_{s,t} : M \to M$, by solving the Cauchy problem with initial condition given at time s. Notice that $P_{t,t} = \mathbb{I}$ for all $t \in \mathbb{R}$ and $P_{t_1,t_2} \circ P_{t_0,t_1} = P_{t_0,t_2}$, whenever they are defined. In particular $(P_{t_1,t_2})^{-1} = P_{t_2,t_1}$.

1.2.2 End-point map

In this section, for convenience, we assume to fix some (local) presentation of the affine control system, hence $L^\infty([0,T],\mathbb{U}) \simeq L^\infty([0,T],\mathbb{R}^k)$. For a more intrinsic approach see [3, Sec. 1].

Definition 1.8. Fix a point $x_0 \in M$ and $T > 0$. The *end-point map at time T* of the system (1.8) is the map

$$E_{x_0,T} : \mathcal{U} \to M, \qquad u \mapsto \gamma_{x_0,u}(T),$$

where $\mathcal{U} \subset L^\infty([0,T],\mathbb{R}^k)$ is the open subset of controls such that the solution $t \mapsto \gamma_{x_0,u}(t)$ of the Cauchy problem (1.8) is defined on the whole interval $[0,T]$.

The end-point map is smooth. Moreover, its Fréchet differential is computed by the following well-known formula (see, e.g. [15]).

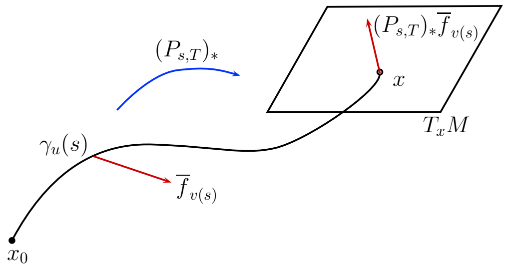


Figure 1.1: Differential of the end-point map.

Proposition 1.9. *The differential of $E_{x_0,T}$ at $u \in \mathcal{U}$, i.e. $D_uE_{x_0,T} : L^\infty([0,T],\mathbb{R}^k) \to T_xM$, where $x = \gamma_u(T)$, is*

$$D_uE_{x_0,T}(v) = \int_0^T (P_{s,T})_*\overline{f}_{v(s)}(\gamma_u(s))ds, \qquad \forall\, v \in L^\infty([0,T],\mathbb{R}^k). \tag{1.9}$$

In other words the differential $D_uE_{x_0,T}$ applied to the control v computes the integral mean of the linear part $\overline{f}_{v(t)}$ of the vector field $f_{v(t)}$ along the trajectory defined by u, by pushing it forward to the final point of the trajectory through the flow $P_{s,T}$ (see Fig. 1.1).

More explicitly, $f(x,u) = f_0(x) + \sum_{i=1}^k u_if_i(x)$, and Eq. (1.9) is rewritten as follows

$$D_uE_{x_0,T}(v) = \int_0^T \sum_{i=1}^k v_i(s)(P_{s,T})_*f_i(\gamma_u(s))ds, \qquad \forall\, v \in L^\infty([0,T],\mathbb{R}^k).$$

1.2.3 Lagrange multipliers rule

Fix $x_0, x \in M$. The problem of finding the infimum of the cost J_T for all admissible curves connecting the endpoints x_0 and x, respectively, in time T, can be naturally reformulated via the end-point map as a constrained extremal problem

$$S_T(x_0,x) = \inf\{J_T(u) \mid E_{x_0,T}(u) = x\} = \inf_{E^{-1}_{x_0,T}(x)} J_T. \tag{1.10}$$

Definition 1.10. We say that $u \in \mathcal{U}$ is an *optimal control* if it is a solution of Eq. (1.10).

Remark 1.11. When f is not injective, a curve γ may be associated with multiple controls. Nevertheless, among all the possible controls u associated with the same admissible curve, there exists a unique control u^* which, for a.e. $t \in [0,T]$, minimizes the Lagrangian function. Then, since we are interested in optimal controls, we assume that any admissible curve γ is always associated with the control u^* which minimizes the Lagrangian, and in this way we have a one-to-one correspondence between admissible curves and controls. With this observation, we say that the admissible curve γ is an *optimal trajectory* (or *minimizer*) if the associated control u^* is optimal according to Definition 1.10.

Notice that, in general, $D_uE_{x_0,T}$ is not surjective and the set $E^{-1}_{x_0,T}(x) \subset M$ is not a smooth submanifold. The Lagrange multipliers rule provides a necessary condition to be satisfied by a control u which is a constrained critical point for (1.10).

Proposition 1.12. *Let $u \in \mathcal{U}$ be an optimal control, with $x = E_{x_0,T}(u)$. Then (at least) one of the two following statements holds true*

(i) $\exists\, \lambda_T \in T_x^*M$ *s.t.* $\lambda_T\, D_u E_{x_0,T} = d_u J_T$,

(ii) $\exists\, \lambda_T \in T_x^*M$, $\lambda_T \neq 0$, *s.t.* $\lambda_T\, D_u E_{x_0,T} = 0$,

where $\lambda_T\, D_u E_{x_0,T}$ *denotes the composition of linear maps*

$$\begin{array}{ccc} L^\infty([0,T],\mathbb{R}^k) & \xrightarrow{\;D_u E_{x_0,T}\;} & T_x M \\ & \underset{d_u J_T}{\searrow} & \downarrow{\scriptstyle \lambda_T} \\ & & \mathbb{R} \end{array}$$

Definition 1.13. A control u, satisfying the necessary conditions for optimality of Proposition 1.12, is called *normal extremal* in case (i), while it is called *abnormal extremal* in case (ii). We use the same terminology to classify the associated extremal trajectory γ_u.

Notice that a single control $u \in \mathcal{U}$ can be associated with two different covectors (or *Lagrange multipliers*) such that both (i) and (ii) are satisfied. In other words, an optimal trajectory may be simultaneously normal and abnormal. We now introduce a key definition for what follows.

Definition 1.14. A normal extremal trajectory $\gamma : [0,T] \to M$ is called *strictly normal* if it is not abnormal. Moreover, if for all $s \in [0,T]$ the restriction $\gamma|_{[0,s]}$ is also strictly normal, then γ is called *strongly normal*.

Remark 1.15. A trajectory is abnormal if and only if the differential $D_u E_{x_0,T}$ is not surjective. By linearity of the integral, it is easy to show from Eq. (1.9) that this is equivalent to the relation

$$\operatorname{span}\{(P_{s,T})_* \mathscr{D}_{\gamma(s)}, s \in [0,T]\} \neq T_{\gamma(T)}M.$$

In particular γ is strongly normal if and only if a short segment $\gamma|_{[0,\varepsilon]}$ is strongly normal, for some $\varepsilon \leq T$.

1.2.4 Pontryagin Maximum Principle

In this section we recall a weak version of the Pontryagin Maximum Principle (PMP) for the optimal control problem, which rewrites the necessary conditions satisfied by normal optimal solutions in the Hamiltonian formalism. In particular it states that every normal optimal trajectory of problem (1.5) is the projection of a solution of a fixed Hamiltonian system defined on T^*M.

Let us denote by $\pi : T^*M \to M$ the canonical projection of the cotangent bundle, and by $\langle \lambda, v \rangle$ the pairing between a cotangent vector $\lambda \in T_x^*M$ and a vector $v \in T_xM$. The Liouville 1-form $\varsigma \in \Lambda^1(T^*M)$ is defined as follows: $\varsigma_\lambda = \lambda \circ \pi_*$, for every $\lambda \in T^*M$. The canonical symplectic structure on T^*M is defined by the non degenerate closed 2-form $\sigma = d\varsigma$. In canonical coordinates $(p, x) \in T^*M$ one has

$$\varsigma = \sum_{i=1}^n p_i dx_i, \qquad \sigma = \sum_{i=1}^n dp_i \wedge dx_i.$$

We denote by $\vec{h}$ the Hamiltonian vector field associated with a function $h \in C^\infty(T^*M)$. Namely, $d_\lambda h = \sigma(\cdot, \vec{h}(\lambda))$ for every $\lambda \in T^*M$ and the coordinates expression of $\vec{h}$ is

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

Let us introduce the smooth control-dependent Hamiltonian on T^*M:

$$\mathcal{H}(\lambda, u) = \langle \lambda, f(x, u)\rangle - L(x, u), \qquad \lambda \in T^*M,\ x = \pi(\lambda).$$

Assumption (A2) guarantees that, for each $\lambda \in T^*M$, the restriction $u \mapsto \mathcal{H}(\lambda, u)$ to the fibers of $\mathbb{U}$ has a unique maximum $\bar{u}(\lambda)$. Moreover, the fiber-wise strong convexity of the Lagrangian and an easy application of the implicit function theorem prove that the map $\lambda \mapsto \bar{u}(\lambda)$ is smooth. Therefore, it is well defined the *maximized Hamiltonian* (or simply, *Hamiltonian*) $H : T^*M \to \mathbb{R}$

$$H(\lambda) := \max_{v \in U_x} \mathcal{H}(\lambda, v) = \mathcal{H}(\lambda, \bar{u}(\lambda)), \qquad \lambda \in T^*M, x = \pi(\lambda).$$

Remark 1.16. When $f(x, u) = f_0(x) + \sum_{i=1}^k u_i f_i(x)$ is written in a local frame, then $\bar{u} = \bar{u}(\lambda)$ is characterized as the solution of the system

$$\frac{\partial \mathcal{H}}{\partial u_i}(\lambda, u) = \langle \lambda, f_i(x)\rangle - \frac{\partial L}{\partial u_i}(x, u) = 0, \qquad i = 1, \dots, k. \tag{1.11}$$

Theorem 1.17 (PMP, [15, 61])**.** *The admissible curve $\gamma : [0, T] \to M$ is a normal extremal trajectory if and only if there exists a Lipschitz lift $\lambda : [0, T] \to T^*M$, such that $\gamma(t) = \pi(\lambda(t))$ and*

$$\dot{\lambda}(t) = \vec{H}(\lambda(t)), \qquad t \in [0, T].$$

In particular, γ and λ are smooth. Moreover, the associated control can be recovered from the lift as $u(t) = \bar{u}(\lambda(t))$, and the final covector $\lambda_T = \lambda(T)$ is the normal Lagrange multiplier associated with u, namely $\lambda_T\, D_u E_{x_0,T} = d_u J_T$.

Thus, every normal extremal trajectory $\gamma : [0, T] \to M$ can be written as $\gamma(t) = \pi \circ e^{t\vec{H}}(\lambda_0)$, for some initial covector $\lambda_0 \in T^*M$ (although it may be non unique). This observation motivates the next definition. For simplicity, and without loss of generality, we assume that $\vec{H}$ is complete.

Definition 1.18. Fix $x_0 \in M$. The *exponential map* with base point x_0 is the map $\mathcal{E}_{x_0} : \mathbb{R}^+ \times T^*_{x_0}M \to M$, defined by $\mathcal{E}_{x_0}(t, \lambda_0) = \pi \circ e^{t\vec{H}}(\lambda_0)$.

When the first argument is fixed, we employ the notation $\mathcal{E}_{x_0,t} : T^*_{x_0}M \to M$ to denote the exponential map with base point x_0 and time t, namely $\mathcal{E}_{x_0,t}(\lambda) = \mathcal{E}_{x_0}(t, \lambda)$. Indeed, the exponential map is smooth.

From now on, we call *geodesic* any trajectory that satisfies the normal necessary conditions for optimality. In other words, geodesics are admissible curves associated with a normal Lagrange multiplier or, equivalently, projections of integral curves of the Hamiltonian flow.

1.2.5 Regularity of the value function

The next well known regularity property of the value function is crucial for the forthcoming sections (see Definition 1.4).

Theorem 1.19. *Let $\gamma : [0,T] \to M'$ be a strongly normal trajectory. Then there exist $\varepsilon > 0$ and an open neighbourhood $U \subset (0,\varepsilon) \times M' \times M'$ such that:*

(i) $(t,\gamma(0),\gamma(t)) \in U$ for all $t \in (0,\varepsilon)$,

(ii) For any $(t,x,y) \in U$ there exists a unique (normal) minimizer of the cost functional J_t, among all the admissible curves that connect x with y in time t, contained in M',

(iii) The value function $(t,x,y) \mapsto S_t(x,y)$ is smooth on U.

According to Definition 1.4, the function S, and henceforth U, depend on the choice of a relatively compact $M' \subset M$. For different relatively compacts, the correspondent value functions S agree on the intersection of the associated domains U: they define the same germ.

The proof of this result can be found in Appendix C. We end this section with a useful lemma about the differential of the value function at a smooth point.

Lemma 1.20. *Let $x_0, x \in M$ and $T > 0$. Assume that the function $x \mapsto S_T(x_0,x)$ is smooth at x and there exists an optimal trajectory $\gamma : [0,T] \to M$ joining x_0 to x. Then*

(i) γ is the unique minimizer of the cost functional J_T, among all the admissible curves that connect x_0 with x in time T, and it is strictly normal,

(ii) $d_x S_T(x_0,\cdot) = \lambda_T$, where λ_T is the final covector of the normal lift of γ.

Proof. Under the above assumptions the function

$$v \mapsto J_T(v) - S_T(x_0, E_{x_0,T}(v)), \qquad v \in L^\infty([0,T],\mathbb{R}^k),$$

is smooth and non negative. For every optimal trajectory γ, associated with the control u, that connects x_0 with x in time T, one has

$$0 = d_u(J_T(\cdot) - S_T(x_0, E_{x_0,T}(\cdot)) = d_u J_T - d_x S_T(x_0,\cdot) \circ D_u E_{x_0,T}.$$

Thus, γ is a normal extremal trajectory, with Lagrange multiplier $\lambda_T = d_x S_T(x_0,\cdot)$. By Theorem 1.17, we can recover γ by the formula $\gamma(t) = \pi \circ e^{(t-T)\vec{H}}(\lambda_T)$. Then, γ is the unique minimizer of J_T connecting its endpoints.

Next we show that γ is not abnormal. For y in a neighbourhood of x, consider the map

$$\Theta : y \mapsto e^{-T\vec{H}}(d_y S_T(x_0,\cdot)).$$

The map Θ, by construction, is a smooth right inverse for the exponential map at time T. This implies that x is a regular value for the exponential map and, a fortiori, u is a regular point for the end-point map at time T. □

1.3 Flag and growth vector of an admissible curve

For each smooth admissible curve, we introduce a family of subspaces, which is related with a micro-local characterization of the control system along the trajectory itself.

1.3.1 Growth vector of an admissible curve

Let $\gamma : [0,T] \to M$ be an admissible, smooth curve such that $\gamma(0) = x_0$, associated with a smooth control u. Let $P_{0,t}$ denote the flow defined by u. We define the family of subspaces of $T_{x_0}M$

$$\mathscr{F}_\gamma(t) := (P_{0,t})_*^{-1}\mathscr{D}_{\gamma(t)}. \tag{1.12}$$

In other words, the family $\mathscr{F}_\gamma(t)$ is obtained by collecting the distributions along the trajectory at the initial point, by using the flow $P_{0,t}$ (see Fig. 1.2).

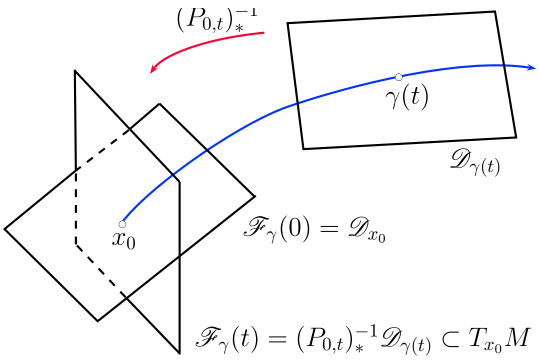


Figure 1.2: The family of subspaces $\mathscr{F}_\gamma(t)$.

Given a family of subspaces in a linear space it is natural to consider the associated flag.

Definition 1.21. The *flag of the admissible curve* γ is the sequence of subspaces

$$\mathscr{F}^i_\gamma(t) := \operatorname{span}\left\{ \frac{d^j}{dt^j} v(t) \,\middle|\, v(t) \in \mathscr{F}_\gamma(t) \text{ smooth}, \, j \leq i-1 \right\} \subset T_{x_0}M, \qquad i \geq 1.$$

Notice that, by definition, this is a filtration of $T_{x_0}M$, i.e. $\mathscr{F}^i_\gamma(t) \subset \mathscr{F}^{i+1}_\gamma(t)$, for all $i \geq 1$.

Definition 1.22. Let $k_i(t) := \dim \mathscr{F}^i_\gamma(t)$. The *growth vector of the admissible curve* γ is the sequence of integers $\mathcal{G}_\gamma(t) = \{k_1(t), k_2(t), \dots\}$.

An admissible curve is *ample at* t if there exists an integer $m = m(t)$ such that $\mathscr{F}^{m(t)}_\gamma(t) = T_{x_0}M$. We call the minimal $m(t)$ such that the curve is ample the *step at* t *of the admissible curve*. An admissible curve is called *equiregular at* t if its growth vector is locally constant at t. Finally, an admissible curve is *ample* (resp. *equiregular*) if it is ample (resp. equiregular) at each $t \in [0,T]$.

Remark 1.23. One can analogously introduce the family of subspaces (and the relevant filtration) at any base point $\gamma(s)$, for every $s \in [0,T]$, by defining the shifted curve $\gamma_s(t) := \gamma(s+t)$. Then $\mathscr{F}_{\gamma_s}(t) := (P_{s,s+t})_*^{-1}\mathscr{D}_{\gamma(s+t)}$. Notice that the relation $\mathscr{F}_{\gamma_s}(t) = (P_{0,s})_*\mathscr{F}_\gamma(s+t)$ implies that the growth vector of the original curve at t can be equivalently computed via the growth vector at time 0 of the curve γ_t, i.e. $k_i(t) = \dim \mathscr{F}^i_{\gamma_t}(0)$, and $\mathcal{G}_\gamma(t) = \mathcal{G}_{\gamma_t}(0)$.

Let us stress that the the family of subspaces (1.12) depends on the choice of the local frame (via the map $P_{0,t}$). However, we will prove that the flag of an admissible curve at $t = 0$ and its growth vector (for all t) are invariant by state-feedback transformation and, in particular, independent on the particular presentation of the system (see Section 1.3.3).

Remark 1.24. The following properties of the growth vector of an *ample* admissible curve highlight the analogy with the "classical" growth vector of the distribution.

(i) The functions $t \mapsto k_i(t)$, for $i = 1, \dots, m(t)$, are lower semicontinuous. In particular, being integer valued functions, this implies that the set of points t such that the growth vector is locally constant is open and dense on $[0,T]$.

(ii) The function $t \mapsto m(t)$ is upper semicontinuous. As a consequence, the step of an admissible curve is bounded on $[0,T]$.

(iii) If the admissible curve is equiregular at t, then $k_1(t) < \ldots < k_m(t)$ is a strictly increasing sequence. Let $i < m$. If $k_i(t) = k_{i+1}(t)$ for all t in a open neighbourhood then, using a local frame, it is easy to see that this implies $k_i(t) = k_{i+1}(t) = \ldots = k_m(t)$ contradicting the fact that the admissible curve is ample at t.

(iv) Assume that an admissible curve is equiregular with step m. The derivation of sections of $\mathscr{F}_\gamma(t)$ induces a well defined map on the quotients

$$\mathscr{F}^i_\gamma(t)/\mathscr{F}^{i-1}_\gamma(t) \longrightarrow \mathscr{F}^{i+1}_\gamma(t)/\mathscr{F}^i_\gamma(t), \qquad \forall\, t \in [0,T].$$

In this case, the maps defined above are surjective and the quotients $\mathscr{F}^i_\gamma/\mathscr{F}^{i-1}_\gamma$ have constant dimensions $d_i := k_{i+1} - k_i = \dim \mathscr{F}^i_\gamma - \dim \mathscr{F}^{i-1}_\gamma$ for $i = 1, \dots, m$. Therefore the sequence $d_1 \geq \ldots \geq d_m$ is decreasing, namely $\dim \mathscr{F}^i_\gamma - \dim \mathscr{F}^{i-1}_\gamma \geq \dim \mathscr{F}^{i+1}_\gamma - \dim \mathscr{F}^i_\gamma$.

Next, we show how the family $\mathscr{F}_\gamma(t)$ can be conveniently employed to characterize strictly and strongly normal geodesics.

Proposition 1.25. *Let $\gamma : [0,T] \to M$ be a geodesic. Then*

(i) γ is strictly normal if and only if $\operatorname{span}\{\mathscr{F}_\gamma(s), s \in [0,T]\} = T_{x_0}M$,

(ii) γ is strongly normal if and only if $\operatorname{span}\{\mathscr{F}_\gamma(s), s \in [0,t]\} = T_{x_0}M$ *for all* $0 < t \leq T$,

(iii) If γ is ample at $t = 0$, then it is strongly normal.

Proof. Recall that a geodesic $\gamma : [0,T] \to M$ is abnormal on $[0,T]$ if and only if the differential $D_uE_{x_0,T}$ is not surjective, which implies (see Remark 1.15)

$$\operatorname{span}\{(P_{s,T})_*\mathscr{D}_{\gamma(s)}, s \in [0,T]\} \neq T_{\gamma(T)}M.$$

By applying the inverse flow $(P_{0,T})_*^{-1} : T_{\gamma(T)}M \to T_{\gamma(0)}M$, we obtain

$$\operatorname{span}\{\mathscr{F}_\gamma(s), s \in [0,T]\} \neq T_{x_0}M.$$

This proves (i). In particular, this implies that a geodesic is strongly normal if and only if

$$\operatorname{span}\{\mathscr{F}_\gamma(s), s \in [0,t]\} = T_{x_0}M, \qquad \forall\, 0 < t \leq T,$$

which proves (ii). We now prove (iii). We argue by contradiction. If the geodesic is not strongly normal, there exists some $\lambda \in T_{x_0}^*M$ such that $\langle \lambda, \mathscr{F}_\gamma(t)\rangle = 0$, for all $0 < t \leq T$. Then, by taking derivatives at $t = 0$, we obtain that $\langle \lambda, \mathscr{F}_\gamma^i(0)\rangle = 0$, for all $i \geq 0$, which is impossible since the curve is ample at $t = 0$ by hypotesis. □

Remark 1.26. Ample geodesics play a crucial role in our approach to curvature, as we explain in Section 1.4. By Proposition 1.25, these geodesics are strongly normal. One may wonder whether the generic covector $\lambda_0 \in T_{x_0}^*M$ corresponds to a strongly normal (or even ample) geodesic. The answer to this question is trivial when there are no abnormal trajectories (e.g. in Riemannian geometry), but the matter is quite delicate in general. For this reason, in order to define the curvature of an affine control system, we assume in the following that the set of ample geodesics is non empty. Eventually, we address the problem of existence of ample geodesics for linear quadratic control systems and sub-Riemannian geometry. In these cases, we will prove that a generic normal geodesic is ample.

1.3.2 Linearised control system and growth vector

It is well known that the differential of the end-point map at a point $u \in \mathcal{U}$ is related with the linearisation of the control system along the associated trajectory. The goal of this section is to discuss the relation between the controllability of the linearised system and the ampleness of the geodesic.

Linearisation of a control system in $\mathbb{R}^n$

We start with some general considerations. Consider the nonlinear control system in $\mathbb{R}^n$

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

where $f : \mathbb{R}^n \times \mathbb{R}^k \to \mathbb{R}^n$ is smooth. Fix $x_0 \in \mathbb{R}^n$, and consider the end-point map $E_{x_0,t} : \mathcal{U} \to \mathbb{R}^n$ for $t \geq 0$. Consider a smooth solution $x_u(t)$, associated with the control $u(t)$, such that $x_u(0) = x_0$. The differential of the end-point map $D_uE_{x_0,t} : L^\infty([0,T],\mathbb{R}^k) \to \mathbb{R}^n$ at u is related with the end-point map of the *linearised system* at the pair $(x_u(t), u(t))$. More precisely, for every $v \in L^\infty([0,T],\mathbb{R}^k)$ the trajectory $y(t) := D_uE_{x_0,t}(v) \in \mathbb{R}^n$ is the solution of the non-autonomous linear system

$$\begin{cases} \dot{y}(t) = A(t)y(t) + B(t)v(t), \\ y(0) = 0, \end{cases} \tag{1.13}$$

where $A(t) := \dfrac{\partial f}{\partial x}(x_u(t), u(t))$ and $B(t) := \dfrac{\partial f}{\partial u}(x_u(t), u(t))$ are smooth families of $n \times n$ and $n \times k$ matrices, respectively. We have the formula

$$y(t) = D_u E_{x_0,t}(v) = M(t) \int_0^t M(s)^{-1} B(s) v(s) ds,$$

where $M(t)$ is the solution of the matrix Cauchy problem $\dot{M}(t) = A(t)M(t)$, with $M(0) = \mathbb{I}$. Indeed the solution $M(t)$ is defined on the whole interval $[0, T]$, and it is invertible therein.

Definition 1.27. The linear control system (1.13) is *controllable* in time $T > 0$ if, for any $y \in \mathbb{R}^n$, there exists $v \in L^\infty([0,T], \mathbb{R}^k)$ such that the associated solution $y_v(t)$ satisfies $y_v(T) = y$.

Let us recall the following classical controllability condition for a linear non-autonomous system, which is the non-autonomous generalization of the Kalman condition (see e.g. [37]). For a set $\{M_i\}$ of $n \times k$ matrices, we denote with $\operatorname{span}\{M_i\}$ the vector space generated by the columns of the matrices in $\{M_i\}$.

Proposition 1.28. *Consider the control system* (1.13)*, with $A(t), B(t)$ smooth, and define*

$$B_1(t) := B(t), \qquad B_{i+1}(t) := A(t)B_i(t) - \dot{B}_i(t). \tag{1.14}$$

Assume that there exist $t \in [0, T]$ and $m > 0$ such that $\operatorname{span}\{B_1(t), B_2(t), \dots, B_m(t)\} = \mathbb{R}^n$. *Then the system* (1.13) *is controllable in time T.*

Remark 1.29. Notice that, using $M(t)$ as a time-dependent change of variable, the new curve $\zeta(t) := M(t)^{-1} y(t) \in \mathbb{R}^n$ satisfies

$$\begin{cases} \dot{\zeta}(t) = M(t)^{-1} B(t) v(t), \\ \zeta(0) = 0. \end{cases} \tag{1.15}$$

If the controllability condition of Proposition (1.28) is satisfied for the pair $(A(t), B(t))$, then it is satisfied also for the pair $(0, C(t))$, with $C(t) = M(t)^{-1}B(t)$, as a consequence of the identity $C^{(i)}(t) = (-1)^i M(t)^{-1} B_{i+1}(t)$. Therefore, the controllability conditions for the control systems (1.13) and (1.15) are equivalent. Moreover, both systems are controllable if and only if one of them is controllable.

Linearisation of a control system in the general setting

Let us go back to the general setting. Let γ be a smooth admissible trajectory associated with the control u such that $\gamma(0) = x_0$. We are interested in the linearisation of the affine control system at γ. Consider the image of a fixed control $v \in L^\infty([0,T], \mathbb{R}^k)$ through the differential of the end-point map $E_{x_0,t}$, for every $t \geq 0$:

$$D_u E_{x_0,t} : L^\infty([0,T], \mathbb{R}^k) \to T_{\gamma(t)} M, \qquad \gamma(t) = E_{x_0,t}(u).$$

In this case, for each $t \geq 0$, the image of v belongs to a different tangent space. In order to obtain a well defined differential equation, we collect the family of vectors in a single vector space through the composition with the push forward $(P_{0,t})_*^{-1} : T_{\gamma(t)}M \to T_{x_0}M$:

$$(P_{0,t})_*^{-1} \circ D_u E_{x_0,t} : L^\infty([0,T], \mathbb{R}^k) \to T_{x_0}M.$$

Using formula (1.9) one easily finds

$$(P_{0,t})_*^{-1} \circ D_u E_{x_0,t}(v) = \int_0^t (P_{0,s})_*^{-1} \overline{f}_{v(s)}(\gamma(s)) ds.$$

Denoting $\zeta(t) := (P_{0,t})_*^{-1} \circ D_u E_{x_0,t}(v) \in T_{x_0}M$ one has that, in a local frame, this curve satisfies

$$\dot{\zeta}(t) = (P_{0,t})_*^{-1} \overline{f}_{v(t)}(\gamma(t)) = \sum_{i=1}^k v_i(t)(P_{0,t})_*^{-1} f_i(\gamma(t)).$$

Therefore, $\zeta(t)$ is a solution of the control system

$$\begin{cases} \dot{\zeta}(t) = C(t)v(t), \\ \zeta(0) = 0, \end{cases} \tag{1.16}$$

where the $n \times k$ matrix $C(t)$ has columns $C_i(t) := (P_{0,t})_*^{-1} f_i(\gamma(t))$ for $i = 1, \ldots, k$. Eq. (1.16) is the *linearised system along the admissible curve* γ. By hypotesis, γ is smooth. Then the linearised system is also smooth.

Remark 1.30. Notice that the composition of the end-point map with $(P_{0,t})_*^{-1}$ corresponds to the time dependent transformation $M(t)^{-1}$ of Remark 1.29.

Growth vector and controllability

From the definition of growth vector of an admissible curve, it follows that

$$\mathscr{F}_\gamma^i(t) = \operatorname{span}\{C(t), \dot{C}(t) \ldots, C^{(i-1)}(t)\}, \qquad i \geq 1.$$

This gives an efficient criterion to compute the geodesic growth vector of the admissible curve γ_u associated with the control u. Define in any local frame $f_1, \ldots, f_k$ and any coordinate system in a neighbourhood of γ, the $n \times n$ and $n \times k$ matrices, respectively:

$$A(t) := \frac{\partial f}{\partial x}(\gamma_u(t), u(t)) = \frac{\partial f_0}{\partial x}(\gamma_u(t)) + \sum_{i=1}^k u_i(t) \frac{\partial f_i}{\partial x}(\gamma_u(t)), \tag{1.17}$$

$$B(t) := \frac{\partial f}{\partial u}(\gamma_u(t), u(t)) = [f_i(\gamma_u(t))]_{i=1,\ldots,k}\,. \tag{1.18}$$

Denoting by $B_j(t)$ the matrices defined as in (1.14), and recalling Remark 1.29, we have

$$k_i(t) = \dim \mathscr{F}_\gamma^i(t) = \operatorname{rank}\{B_1(t), \ldots, B_i(t)\}.$$

Assume now that the admissible curve γ is actually a normal geodesic of the optimal control system. As a consequence of this discussion and Proposition 1.28, we obtain the following characterisation in terms of the controllability of the linearised system.

Proposition 1.31. *Let $\gamma : [0, T] \to M$ be a geodesic. Then*

(i) γ is strictly normal $\Leftrightarrow$ the linearised system is controllable in time T,

(ii) γ is strongly normal $\Leftrightarrow$ the linearised system is controllable in time t, $\forall\, t \in (0, T]$,

(iii) γ is ample at $t = 0$ $\Leftrightarrow$ the controllability condition of Proposition 1.28 is satisfied at $t = 0$.

In particular (iii)$\Rightarrow$(ii)$\Rightarrow$(i). Moreover, the three properties are equivalent in the analytic case.

The equivalence in the analytic case is a classical fact about the controllability of nonautonomous analytic linear systems. See, for example, [37, Sec. 1.3].

1.3.3 State-feedback invariance of the flag of an admissible curve

In this section we prove that, albeit the family $\mathscr{F}_\gamma(t)$ depends on the choice of the local trivialization, the flag of an admissible curve at $t = 0$ is invariant by state-feedback transformation, hence it does not depend on the presentation. This also implies that the growth vector of the admissible curve is well-defined (for all t). In this section we use the shorthand $\mathscr{F}^i_\gamma = \mathscr{F}^i_\gamma(0)$, when the flag is evaluated at $t = 0$.

Proposition 1.32. *The flag $\mathscr{F}^1_\gamma \subset \mathscr{F}^2_\gamma \subset \ldots \subset T_{x_0}M$ is state-feedback invariant. In particular it does not depend on the presentation of the control system.*

The next corollary is a direct consequence of Proposition 1.32 and Remark 1.23.

Corollary 1.33. *The growth vector of an admissible curve $\mathcal{G}_\gamma(t)$ is state-feedback invariant.*

Proof of Proposition 1.32. Recall that every state-feedback transformation is the composition of pure state and a pure feedback one. For pure state transformations the statement is trivial, since it is tantamount to a change of variables on the manifold. Thus, it is enough to prove the proposition for pure feedback ones. Recall that the subspaces $\mathscr{F}^i_\gamma$ are defined, in terms of a given presentation, as

$$\mathscr{F}^i_\gamma = \operatorname{span}\{C(0), \ldots, C^{(i-1)}(0)\}, \qquad i \geq 1,$$

where the columns of the matrices $C(t)$ are given by the vectors $C_i(t) = (P_{0,t})_*^{-1} f_i(\gamma(t))$. A pure feedback transformation corresponds to a change of presentation. Thus, let

$$\dot{x} = f(x, u) = f_0(x) + \sum_{i=1}^{k} u_i f_i(x), \qquad \dot{x} = f'(x, u') = f'_0(x) + \sum_{i=1}^{k} u'_i f'_i(x),$$

related by the pure feedback transformation $u'_i = \psi_i(x, u) = \psi_{i,0}(x) + \sum_{j=1}^k \psi_{i,j}(x) u_j$. In particular (see also Remark 1.7)

$$f_0(x) = f'_0(x) + \sum_{i=1}^{k} \psi_{i,0}(x) f'_i(x), \qquad f_i(x) = \sum_{j=1}^{k} \psi_{j,i}(x) f'_j(x). \tag{1.19}$$

Denote by $A(t), A'(t)$ and $B(t), B'(t)$ the matrices (1.17) and (1.18) associated with the two presentations, in some set of coordinates. According to Remark 1.29, $C(t) = M(t)^{-1}B(t)$, where $M(t)$ is the solution of $\dot{M}(t) = A(t)M(t)$, with $M(0) = \mathbb{I}$, and analogous formulae for the "primed" counterparts. In particular, since $C^{(i)}(t) = (-1)^i M(t)^{-1} B_{i+1}(t)$ and $M(0) = M(0)' = \mathbb{I}$, we get

$$\mathscr{F}^i_\gamma = \operatorname{span}\{B_1(0), \dots, B_i(0)\}, \qquad (\mathscr{F}^i_\gamma)' = \operatorname{span}\{B'_1(0), \dots, B'_i(0)\}, \tag{1.20}$$

where $B_i(t)$ and $B'_i(t)$ are the matrices defined in Proposition 1.28 for the two systems. Notice that Eq. (1.20) is true only at $t = 0$. We prove the following property, which implies our claim: there exists an invertible matrix $\Psi(t)$ such that

$$B_{i+1}(t) = B'_{i+1}(t)\Psi(t) \bmod \operatorname{span}\{B'_1(t), \dots, B'_i(t)\}, \tag{1.21}$$

where Eq. (1.21) is meant column-wise. Indeed, from Eq. (1.19) we obtain the relations

$$A(t) = A'(t) + B'(t)\Phi(t), \qquad B(t) = B'(t)\Psi(t), \tag{1.22}$$

where $\Psi(t)$ and $\Phi(t)$ are $k \times k$ and $k \times n$ matrices, respectively, with components

$$\Psi(t)_{i\ell} := \psi_{i,\ell}(x(t)), \qquad \Phi(t)_{i\ell} := \frac{\partial \psi_{i,0}}{\partial x_\ell}(x(t)) + \sum_{j=1}^k u_j(t) \frac{\partial \psi_{i,j}}{\partial x_\ell}(x(t)).$$

Notice that, by definition of feedback transformation, $\Psi(t)$ is invertible. We prove Eq. (1.21) by induction. For $i = 0$, it follows from (1.22). The induction assumption is (we omit t)

$$B_i = B'_i\Psi + \sum_{j=0}^{i-1} B'_j\Theta_j, \qquad \text{for some time dependent } k \times k \text{ matrices } \Theta_j.$$

Let $X \simeq Y$ denote $X = Y \bmod \operatorname{span}\{B'_1, \dots, B'_i\}$, column-wise. Then

$$\begin{aligned} B_{i+1} &= AB_i - \dot{B}_i \simeq \\ &\simeq (A'B'_i - \dot{B}'_i)\Psi + \sum_{j=0}^{i-1}(A'B'_j - \dot{B}'_j)\Theta_j \simeq B'_{i+1}\Psi. \end{aligned}$$

We used that $A = A' \bmod \operatorname{span}\{B'\}$, hence we can replace A by A'. Moreover all the terms with the derivatives of Θ_j belong to $\operatorname{span}\{B'_1, \dots, B'_i\}$. □

1.3.4 An alternative definition

In this section we present an alternative definition for the flag of an admissible curve, at $t = 0$. The idea is that the flag $\mathscr{F}_\gamma = \mathscr{F}_\gamma(0)$ of a smooth, admissible trajectory γ can be obtained by computing the Lie derivatives along the direction of γ of sections of the distribution, namely elements of $\overline{\mathscr{D}}$. In this sense, the flag of an admissible curve carries informations about the germ of the distribution along the given trajectory.

Let $\gamma : [0, T] \to M$ be a smooth admissible trajectory, such that $x_0 = \gamma(0)$. By definition, this means that there exists a smooth map $u : [0, T] \to \mathbb{U}$ such that $\dot{\gamma}(t) = f(\gamma(t), u(t))$.

Definition 1.34. We say that $\mathsf{T} \in f_0 + \overline{\mathscr{D}}$ is a smooth admissible extension of $\dot\gamma$ if there exists a smooth section $\sigma : M \to \mathbb{U}$ such that $\sigma(\gamma(t)) = u(t)$ and $\mathsf{T} = f \circ \sigma$.

In other words T is a vector field extending $\dot\gamma$ obtained through the bundle map $f : \mathbb{U} \to TM$ from an extension of the control u (seen as a section of $\mathbb{U}$ over the curve γ). Notice that, if $\dot\gamma(t) = f_0(\gamma(t)) + \sum_{i=1}^k u_i(t) f_i(\gamma(t))$, an admissible extension of $\dot\gamma$ is a smooth field of the form $\mathsf{T} = f_0 + \sum_{i=1}^k \alpha_i f_i$, where $\alpha_i \in C^\infty(M)$ are such that $\alpha_i(\gamma(t)) = u_i(t)$ for all $i = 1, \dots, k$.

With abuse of notation, we employ the same symbol $\mathscr{F}^i_\gamma$ for the following alternative definition.

Definition 1.35. The flag of the admissible curve γ is the sequence of subspaces

$$\mathscr{F}^i_\gamma := \operatorname{span}\{\mathcal{L}^j_{\mathsf{T}}(X)|_{x_0} \,|\, X \in \overline{\mathscr{D}},\ j \leq i-1\} \subset T_{x_0}M, \qquad i \geq 1,$$

where $\mathcal{L}_{\mathsf{T}}$ denotes the Lie derivative in the direction of T.

Notice that, by definition, this is a filtration of $T_{x_0}M$, i.e. $\mathscr{F}^i_\gamma \subset \mathscr{F}^{i+1}_\gamma$, for all $i \geq 1$. Moreover, $\mathscr{F}^1_\gamma = \mathscr{D}_{x_0}$. In the rest of this section, we show that Definition 1.35 is well posed, and is equivalent to the original Definition 1.21 at $t = 0$.

Proposition 1.36. *Definition 1.35 does not depend on the admissible extension of $\dot\gamma$.*

Proof. Let $\mathscr{F}^i_\gamma$ and $\widetilde{\mathscr{F}}^i_\gamma$ the subspaces obtained via Definition 1.35 with two different extensions T and $\widetilde{\mathsf{T}}$ of $\dot\gamma$, respectively. In particular, the field $V := \widetilde{\mathsf{T}} - \mathsf{T} \in \overline{\mathscr{D}}$ vanishes on the support of γ. We prove that $\widetilde{\mathscr{F}}^i_\gamma = \mathscr{F}^i_\gamma$ by induction. For $i = 1$ the statement is trivial. Then, assume $\widetilde{\mathscr{F}}^i_\gamma = \mathscr{F}^i_\gamma$. Since $\widetilde{\mathscr{F}}^{i+1}_\gamma = \widetilde{\mathscr{F}}^i_\gamma + \operatorname{span}\{\mathcal{L}^i_{\widetilde{\mathsf{T}}}(X)|_{x_0} \,|\, X \in \overline{\mathscr{D}}\}$, it sufficient to prove that

$$\mathcal{L}^i_{\widetilde{\mathsf{T}}}(X) = \mathcal{L}^i_{\mathsf{T}}(X) \bmod \mathscr{F}^i_\gamma, \qquad X \in \overline{\mathscr{D}}. \tag{1.23}$$

Notice that $\mathcal{L}^i_{\widetilde{\mathsf{T}}}(X) = \mathcal{L}^i_{\mathsf{T}}(X) + W$, where $W \in \operatorname{Vec}(M)$ is the sum of terms of the form

$$W = \mathcal{L}^\ell_{\mathsf{T}}([V, Y]), \qquad \text{for some} \quad Y \in \operatorname{Vec}(M), \quad 0 \leq \ell \leq i-1.$$

In terms of a local set of generators $f_1, \dots, f_k$ of $\overline{\mathscr{D}}$, $V = \sum_{i=1}^k v_j f_j$, where the functions v_i vanish identically on the support of γ, namely $v_j(\gamma(t)) = 0$ for $t \in [0, T]$. Then, an application of the binomial formula for derivations leads to

$$W = \sum_{j=1}^k \mathcal{L}^\ell_{\mathsf{T}}(v_j [f_j, Y]) - \mathcal{L}^\ell_{\mathsf{T}}(Y(v_j) f_j) =$$
$$= \sum_{j=1}^k \sum_{h=0}^\ell \binom{\ell}{h} \left(\mathcal{L}^h_{\mathsf{T}}(v_j) \mathcal{L}^{\ell-h}_{\mathsf{T}}([f_i, Y]) - \mathcal{L}^h_{\mathsf{T}}(Y(v_j)) \mathcal{L}^{\ell-h}_{\mathsf{T}}(f_i) \right).$$

Observe that $\mathcal{L}^h_{\mathsf{T}}(v_j)|_{x_0} = \frac{d^h v_j}{dt^h}\Big|_{t=0} = 0$, for all $h \geq 0$. Then, if we evaluate W at x_0, we obtain

$$W|_{x_0} = -\sum_{j=1}^k \sum_{h=0}^\ell \binom{\ell}{h} \mathcal{L}^h_{\mathsf{T}}(Y(v_j))|_{x_0} \mathcal{L}^{\ell-h}_{\mathsf{T}}(f_i).$$

Then, since $0 \leq \ell \leq i-1$, and by the induction hypotesis, $W|_{x_0} \in \mathscr{F}^i_\gamma$ and Eq. (1.23) follows. $\square$

Proposition 1.37. *Definition 1.35 is equivalent to Definition 1.21 at $t = 0$.*

Proof. Recall that, according to Definition 1.21, at $t = 0$

$$\mathscr{F}^i_\gamma = \mathscr{F}^i_\gamma(0) = \operatorname{span}\left\{ \left.\frac{d^j}{dt^j}\right|_{t=0} v(t) \,\middle|\, v(t) \in \mathscr{F}_\gamma(t) \text{ smooth}, \, j \leq i-1 \right\} \subset T_{x_0}M, \qquad i \geq 1.$$

where $\mathscr{F}_\gamma(t) = (P_{0,t})_*^{-1}\mathscr{D}_{\gamma(t)}$. By Proposition 1.32, the flag at $t = 0$ is state-feedback invariant. Then, up to a (local) pure feedback transformation, we assume that the fixed smooth admissible trajectory $\gamma : [0, T] \to M$ is associated with a constant control, namely $\dot{\gamma}(t) = f_0(\gamma(t)) + \sum_{i=1}^k u_i f_i(\gamma(t))$, where $u \in L^\infty([0, T], \mathbb{R}^k)$ is constant. In this case, the flow $P_{0,t} : M \to M$ is actually the flow of the autonomous vector field $\mathsf{T} := f_0 + \sum_{i=1}^k u_i f_i$, that is $P_{0,t} = e^{t\mathsf{T}}$.

Indeed $\mathsf{T} \in f_0 + \overline{\mathscr{D}}$ is an admissible extension of $\dot{\gamma}$. Moreover, any smooth $v(t) \in \mathscr{F}_\gamma(t)$ is of the form $v(t) = (P_{0,t})_*^{-1} X|_{\gamma(t)}$, where $X \in \overline{\mathscr{D}}$. Then

$$\left.\frac{d^j}{dt^j}\right|_{t=0} v(t) = \left.\frac{d^j}{dt^j}\right|_{t=0} (P_{0,t})_*^{-1} X|_{\gamma(t)} = \left.\frac{d^j}{dt^j}\right|_{t=0} e_*^{-t\mathsf{T}} X|_{\gamma(t)} = \mathcal{L}^j_{\mathsf{T}}(X)|_{x_0},$$

where in the last equality we have employed the definition of Lie derivative. □

Remark 1.38. To end this section, observe that, for any equiregular smooth admissible curve $\gamma : [0, T] \to M$, the Lie derivative in the direction of the curve defines surjective linear maps

$$\mathcal{L}_{\mathsf{T}} : \mathscr{F}^i_{\gamma(t)} / \mathscr{F}^{i-1}_{\gamma(t)} \to \mathscr{F}^{i+1}_{\gamma(t)} / \mathscr{F}^i_{\gamma(t)}, \qquad i \geq 1,$$

for any fixed $t \in [0, T]$ as follows. Let $\mathsf{T} \in \operatorname{Vec}(M)$ be any admissible extension of $\dot{\gamma}$. Similarly, for $X \in \mathscr{F}^i_{\gamma(t)}$, consider a smooth extension of X along the curve γ such that $X|_{\gamma(s)} \in \mathscr{F}^i_{\gamma(s)}$ for all $s \in [0, T]$. Then we define

$$\mathcal{L}_{\mathsf{T}}(X) := [T, X]|_{\gamma(t)} \bmod \mathscr{F}^i_{\gamma(t)}, \qquad t \in [0, T].$$

The proof that $\mathcal{L}_{\mathsf{T}}$ does not depend on the choice of the admissible extension T is the same of Proposition 1.36 and for this reason we omit it. The fact that it depends only on the value of $X \bmod \mathscr{F}^{i-1}_{\gamma(t)}$ at the point $\gamma(t)$ is similar, under the equiregularity assumption.

In particular, notice that the maps $\mathcal{L}^i_{\mathsf{T}} : \mathscr{D}_{\gamma(t)} \to \mathscr{F}^{i+1}_{\gamma(t)} / \mathscr{F}^i_{\gamma(t)}$, for $i \geq 1$, are well defined, surjective linear maps from the distribution (see also point (iv) of Remark 1.24).

1.4 Geodesic cost and its asymptotics

In this section we define the *geodesic cost function* and we state the main result about the existence of its asymptotics (see Theorem 1.A). This paves the way for the definition of curvature of an affine optimal control system (see Theorem 1.B).

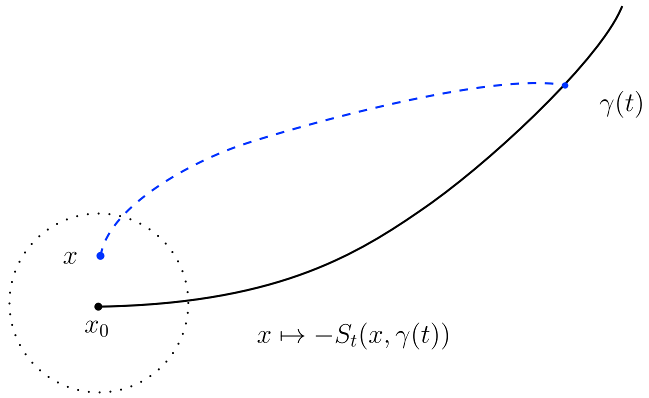


Figure 1.3: The geodesic cost function.

1.4.1 Geodesic cost

Definition 1.39. Let $x_0 \in M$ and consider a strongly normal geodesic $\gamma : [0, T] \to M$ such that $\gamma(0) = x_0$. The *geodesic cost* associated with γ is the family of functions

$$c_t(x) := -S_t(x, \gamma(t)), \qquad x \in M, \, t > 0,$$

The geodesic cost function is smooth in a neighbourhood of x_0, and for $t > 0$ sufficiently small. More precisely, Theorem 1.19, applied to the geodesic cost, can be rephrased as follows.

Theorem 1.40. *Let $x_0 \in M$ and $\gamma : [0, T] \to M$ be a strongly normal geodesic such that $\gamma(0) = x_0$. Then there exist $\varepsilon > 0$ and an open set $U \subset (0, \varepsilon) \times M$ such that*

(i) $(t, x_0) \in U$ for all $t \in (0, \varepsilon)$,

(ii) The geodesic cost function $(t, x) \mapsto c_t(x)$ is smooth on U.

Moreover, for any $(t, x) \in U$, there exists a unique (normal) minimizer of the cost functional J_t, among all the admissible curves that connect x with $\gamma(t)$.

In the following, $\dot{c}_t$ denotes the derivative of the geodesic cost with respect to t.

Proposition 1.41. *Under the assumptions above, $d_{x_0} c_t = \lambda_0$, for all $t \in (0, \varepsilon)$. In particular x_0 is a critical point for the function $\dot{c}_t$ for all $t \in (0, \varepsilon)$.*

Proof. First observe that, in general, if $\gamma(t)$ is an admissible curve for an affine control system, the "reversed" curve $\widetilde{\gamma}(t) := \gamma(T - t)$ is no longer admissible. As a consequence, the value function $(x_0, x_1) \mapsto S_T(x_0, x_1)$ is not symmetric and we cannot directly apply Lemma 1.20 To compute the differential of the value function $x \mapsto -S_t(x, \gamma(t))$ at x_0. Nevertheless, we can still exploit Lemma 1.20, by passing to an associated control problem with reversed dynamic.

Lemma 1.42. *Consider the control system with reversed dynamic*

$$\begin{aligned} &\dot{x} = \widetilde{f}(x, u), \qquad x \in M, \qquad \widetilde{f}(x, u) := -f(x, u), \\ &J_T(u) \to \min . \end{aligned}$$

Let $\widetilde{S}_T$ be the value function of this problem. Then $\widetilde{S}_T(x_0, x_1) = S_T(x_1, x_0)$, for all $x_0, x_1 \in M$.

Proof of Lemma 1.42. It is easy to see that the map $\gamma(t) \mapsto \widetilde{\gamma}(t) := \gamma(T-t)$ defines a one-to-one correspondence between admissible curves for the two problems. Moreover, if γ is associated with the control u, then $\widetilde{\gamma}$ is associated with control $\widetilde{u}(t) := u(T-t)$. Since the cost is invariant by this transformation, one has $\widetilde{S}_T(x_1, x_0) = S_T(x_0, x_1)$. Notice that this transformation preserves normal and abnormal trajectories and minimizers. □

The Hamiltonian of the reversed system is $\widetilde{H}(\lambda) = H(-\lambda)$. Let $i : T^*M \to T^*M$ be the fiberwise linear map $\lambda \mapsto -\lambda$. Then, $i_*\vec{\widetilde{H}}(\lambda) = -\vec{H}(-\lambda)$ (i.e. $\vec{\widetilde{H}}$ is i_*-related with $-\vec{H}$). This implies that, if $\lambda(t)$ is the lift of the geodesic $\gamma(t)$ for the original system, then $\widetilde{\lambda}(t) := -\lambda(T-t)$ is the lift of the geodesic $\widetilde{\gamma}(t) = \gamma(T-t)$ for the reversed system. In particular, the final covector of the reversed geodesic $\widetilde{\lambda}_T = \widetilde{\lambda}(T) = -\lambda(0) = -\lambda_0$ is equal to minus the initial covector of the original geodesic. Thus, we can apply Lemma 1.20 and obtain

$$d_{x_0} c_T = -d_{x_0} S_T(\cdot, \gamma(T)) = -d_{x_0}(\widetilde{S}_T(\gamma(T), \cdot)) = -\tilde{\lambda}(T) = \lambda_0.$$

where $\tilde{\gamma} : [0,T] \to M$ is the unique strictly normal minimizer of the cost functional $\widetilde{J}_T = J_T$ of the reversed system such that $\tilde{\gamma}(0) = \gamma(T)$ and $\tilde{\gamma}(T) = x_0$. □

1.4.2 Hamiltonian inner product

In this section we introduce an inner product on the distribution, which depends on a given geodesic. Namely, it is induced by the second derivative of Hamiltonian of the control system at a point $\lambda \in T^*M$, associated with a geodesic.

A non-negative definite quadratic form, defined on the dual of a vector space V^*, induces an inner product on a subspace of V as follows. Recall first that a quadratic form can be defined as a self-adjoint linear map $B : V^* \to V$. B is non-negative definite if, for all $\lambda \in V^*$, $\langle \lambda, B(\lambda) \rangle \geq 0$. Let us define a bilinear map on $\mathrm{Im}\,(B) \subset V$ by the formula

$$\langle w_1 | w_2 \rangle_B := \langle \lambda_1, B(\lambda_2) \rangle, \qquad \text{where } w_i = B(\lambda_i).$$

It is easy to prove that $\langle \cdot | \cdot \rangle_B$ is symmetric and does not depend on the representatives λ_i. Moreover, since B is non-negative definite, $\langle \cdot | \cdot \rangle_B$ is an inner product on $\mathrm{Im}\,(B)$.

Now we go back to the general setting. Fix a point $x \in M$, consider the restriction of the Hamiltonian H to the fiber $H_x := H|_{T_x^*M}$ and denote by $d^2_\lambda H_x$ its second derivative at the point $\lambda \in T_x^*M$. We show that $d^2_\lambda H_x$ is a non-negative quadratic form and, as a self-adjoint linear map $d^2_\lambda H_x : T_x^*M \to T_xM$, its image is exactly the distribution at the base point.

Lemma 1.43. *For every $\lambda \in T_x^*M$, $d^2_\lambda H_x$ is non-negative definite and $\mathrm{Im}\,(d^2_\lambda H_x) = \mathscr{D}_x$.*

Proof. We prove the result by computing an explicit expression for $d^2_\lambda H_x$ in coordinates $\lambda = (p, x)$ on T^*M. Recall that the maximized Hamiltonian H is defined by the identity

$$H(p,x) = \mathcal{H}(p,x,\bar{u}) = \langle p, f_0(x) \rangle + \sum_{i=1}^k \bar{u}_i \langle p, f_i(x) \rangle - L(x, \bar{u}),$$

where $\bar{u} = \bar{u}(p, x)$ is the solution of the maximality condition

$$\langle p, f_i(x) \rangle = \frac{\partial L}{\partial u_i}(x, \bar{u}(p, x)), \qquad i = 1, \dots, k. \tag{1.24}$$

By the chain rule, we obtain

$$\frac{\partial H}{\partial p}(p, x) = f_0(x) + \sum_{i=1}^{k} \bar{u}_i f_i(x) + \underbrace{\frac{\partial \bar{u}_i}{\partial p} \langle p, f_i(x) \rangle - \frac{\partial L}{\partial u_i} \frac{\partial \bar{u}_i}{\partial p}}_{=0}.$$

By differentiating Eq. (1.24) with respect to p, we get

$$f_i(x) = \sum_{j=1}^{k} \frac{\partial^2 L}{\partial u_i \partial u_j} \frac{\partial \bar{u}_j}{\partial p}, \qquad i = 1, \dots, k.$$

Finally, we compute the second derivatives matrix

$$\frac{\partial^2 H}{\partial p^2}(p, x) = \sum_{i=1}^{k} \frac{\partial \bar{u}_i}{\partial p} f_i^*(x) = \sum_{i,j=1}^{k} f_i(x) \left(\frac{\partial^2 L}{\partial u_i \partial u_j} \right)^{-1} f_j^*(x). \tag{1.25}$$

Since the Hessian of L (with respect to u) is positive definite, Eq. (1.25) implies that $d^2_\lambda H_x$ is non-negative definite and $\operatorname{Im} d^2_\lambda H_x \subset \mathscr{D}_x$. Moreover, it is easy to see that $\operatorname{rank} \frac{\partial^2 H}{\partial p^2} = \dim \mathscr{D}_x$, therefore $\operatorname{Im}(d^2_\lambda H_x) = \mathscr{D}_x$. □

Definition 1.44. For any $\lambda \in T^*_x M$, the *Hamiltonian inner product* (associated with λ) is the inner product $\langle \cdot | \cdot \rangle_\lambda$ induced by $d^2_\lambda H_x$ on $\mathscr{D}_x$.

Remark 1.45. We stress that, for any fixed $x \in M$, the subspace $\mathscr{D}_x \subset T_x M$, where the inner product $\langle \cdot | \cdot \rangle_\lambda$ is defined, does not depend on the choice of the element λ in the fiber $T^*_x M$. When H_x itself is a quadratic form, $d^2_\lambda H_x = 2H_x$ for every $\lambda \in T^*_x M$. Therefore, the inner product $\langle \cdot | \cdot \rangle_\lambda$ does not depend on the choice of $\lambda \in T_x M$. This is the case, for example, of an optimal control system defined by a sub-Riemannian structure, in which the inner product just defined is precisely the sub-Riemannian one (see Section 1.5).

1.4.3 Asymptotics of the geodesic cost function and curvature

Let $f : M \to \mathbb{R}$ be a smooth function defined on a smooth manifold M. Its first differential at a point $x \in M$ is the linear map $d_x f : T_x M \to \mathbb{R}$.

The *second differential* of f, as a symmetric bilinear form, is well defined only at a critical point, i.e. at those points $x \in M$ such that $d_x f = 0$. Indeed, in this case, the map

$$d^2_x f : T_x M \times T_x M \to \mathbb{R}, \qquad d^2_x f(v, w) = V(W(f))(x),$$

where V, W are vector fields such that $V(x) = v$ and $W(x) = w$, respectively, is a well defined symmetric bilinear form which does not depend on the choice of the extensions.

The quadratic form associated with the second differential of f at x which, for simplicity, we denote by the same symbol $d^2_x f : T_x M \to \mathbb{R}$, is

$$d^2_x f(v) = \frac{d^2}{dt^2}\bigg|_{t=0} f(\gamma(t)), \qquad \gamma(0) = x, \quad \dot\gamma(0) = v.$$

Now, for $\lambda \in T^*_{x_0}M$, consider the geodesic cost function associated with the strongly normal geodesic $\gamma(t) = \mathcal{E}_{x_0}(t, \lambda)$, starting from x_0. By Proposition 1.41, for every $t \in (0, \varepsilon)$, the function $x \mapsto \dot c_t(x)$ has a critical point at x_0. Hence we can consider the family of quadratic forms defined on the distribution

$$d^2_{x_0} \dot c_t\big|_{\mathscr{D}_{x_0}} : \mathscr{D}_{x_0} \to \mathbb{R}, \qquad t \in (0, \varepsilon),$$

obtained by the restriction of the second differential of $\dot c_t$ to the distribution $\mathscr{D}_{x_0}$. Then, using the inner product $\langle\cdot|\cdot\rangle_\lambda$ induced by $d^2_\lambda H_x$ on $\mathscr{D}_x$ introduced in Section 1.4.2, we associate with this family of quadratic forms the family of symmetric operators on the distribution $\mathcal{Q}_\lambda(t) : \mathscr{D}_{x_0} \to \mathscr{D}_{x_0}$ defined by the identity

$$d^2_{x_0} \dot c_t(v) := \langle \mathcal{Q}_\lambda(t) v | v \rangle_\lambda, \qquad t \in (0, \varepsilon),\ v \in \mathscr{D}_{x_0}. \tag{1.26}$$

The assumption that the geodesic is strongly normal ensures the smoothness of $\mathcal{Q}_\lambda(t)$ for small $t > 0$. If the geodesic is also ample, we have a much stronger statement about the asymptotic behaviour of $\mathcal{Q}_\lambda(t)$ for $t \to 0$.

Theorem 1.A. *Let $\gamma : [0, T] \to M$ be an ample geodesic with initial covector $\lambda \in T^*_{x_0}M$, and let $\mathcal{Q}_\lambda(t) : \mathscr{D}_{x_0} \to \mathscr{D}_{x_0}$ be defined by (1.26). Then $t \mapsto t^2 \mathcal{Q}_\lambda(t)$ can be extended to a smooth family of operators on $\mathscr{D}_{x_0}$ for small $t \geq 0$, symmetric with respect to $\langle\cdot|\cdot\rangle_\lambda$. Moreover,*

$$\mathcal{I}_\lambda := \lim_{t \to 0^+} t^2 \mathcal{Q}_\lambda(t) \geq \mathbb{I} > 0,$$

as operators on $(\mathscr{D}_{x_0}, \langle\cdot|\cdot\rangle_\lambda)$. Finally

$$\frac{d}{dt}\bigg|_{t=0} t^2 \mathcal{Q}_\lambda(t) = 0.$$

As a consequence of Theorem 1.A we are allowed to introduce the following definitions.

Definition 1.46. Let $\lambda \in T^*_{x_0}M$ be the initial covector associated with an ample geodesic. The *curvature* is the symmetric operator $\mathcal{R}_\lambda : \mathscr{D}_{x_0} \to \mathscr{D}_{x_0}$ defined by

$$\mathcal{R}_\lambda := \frac{3}{2} \frac{d^2}{dt^2}\bigg|_{t=0} t^2 \mathcal{Q}_\lambda(t).$$

In particular, we have the following Laurent expansion for the family of symmetric operators $\mathcal{Q}_\lambda(t) : \mathscr{D}_{x_0} \to \mathscr{D}_{x_0}$:

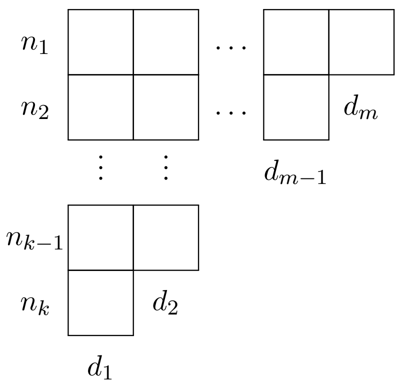
$$\mathcal{Q}_\lambda(t) = \frac{1}{t^2}\mathcal{I}_\lambda + \frac{1}{3}\mathcal{R}_\lambda + O(t), \qquad t > 0. \tag{1.27}$$

The normalization factor $1/3$ appearing in (1.27) in front of the operator $\mathcal{R}_\lambda$ is necessary for recovering the sectional curvature in the case of a control system defined by a Riemannian structure (see Section 1.4.4). We stress that, by construction, $\mathcal{I}_\lambda$ and $\mathcal{R}_\lambda$ are operators on the distributions, symmetric with respect to the inner product $\langle\cdot|\cdot\rangle_\lambda$.

Spectrum of $\mathcal{I}_\lambda$ for equiregular geodesics

Under the assumption that the geodesic is also equiregular, we can completely characterize the operator $\mathcal{I}_\lambda$, namely compute its spectrum.

Let us consider the growth vector $\mathcal{G}_\gamma = \{k_1, k_2, \dots, k_m\}$ of the geodesic γ which, by the equiregularity assumption, does not depend on t. Let $d_i := \dim \mathscr{F}_\gamma^i - \dim \mathscr{F}_\gamma^{i-1} = k_i - k_{i-1}$, for $i = 1, \dots, m$ (where $k_0 := 0$). Recall that d_i is a decreasing sequence (see (iv) of Remark 1.24). Then we can build a tableau with m columns of length d_i, for $i = 1, \dots, m$, as follows:



$$\sum_{i=1}^{m} d_i = n = \dim M,$$
$$d_1 = k_1 = k := \dim \mathscr{D}_{x_0}.$$

Finally, for $j = 1, \dots, k$, let n_j be the length of the j-th row of the tableau.

Theorem 1.B. *Let $\gamma : [0, T] \to M$ be an ample and equiregular geodesic with initial covector $\lambda \in T^*_{x_0}M$. Then the symmetric operator $\mathcal{I}_\lambda : \mathscr{D}_{x_0} \to \mathscr{D}_{x_0}$ satisfies*

(i) $\operatorname{spec} \mathcal{I}_\lambda = \{n_1^2, \dots, n_k^2\}$,

(ii) $\operatorname{tr} \mathcal{I}_\lambda = n_1^2 + \ldots + n_k^2$.

Remark 1.47. Notice that, although the family $\mathcal{Q}_\lambda(t)$ depends on the cost function, the operator $\mathcal{I}_\lambda$ depends only on the growth vector $\mathcal{G}_\gamma$, which is a state-feedback invariant not related with the cost. This is a consequence of the results of Section 1.3.3 and Theorem 1.B.

Remark 1.48. By the classical identity $\sum_{i=1}^n (2i-1) = n^2$, we rewrite the trace of $\mathcal{I}_\lambda$ as follows

$$\operatorname{tr} \mathcal{I}_\lambda = \sum_{i=1}^{m} (2i-1)(\dim \mathscr{F}_\gamma^i - \dim \mathscr{F}_\gamma^{i-1}).$$

Notice that the right hand side of the above equation makes sense also for a non-equiregular (tough still ample) geodesic, where the dimensions are computed at $t = 0$. This number also appears in Section 1.5, under the name of *geodesic dimension*, in connection with the asymptotics of the volume growth in sub-Riemannian geometry.

The proofs of Theorems 1.A and 1.B are postponed to Section 1.7, upon the introduction of the required technicals tools.

1.4.4 Examples

In this section, we discuss two relevant examples, namely an autonomous linear control system on $\mathbb{R}^n$ with quadratic cost and a Riemannian manifold. In the first example it is possible to compute $\mathcal{Q}_\lambda$ and its expansion, by a direct manipulation of the cost geodesic function. In the second example, we recover the sectional curvature of the Riemannian manifold.

Linear-quadratic control problem

Let us consider a classical linear-quadratic control system. Namely $M = \mathbb{R}^n$, $\mathbb{U} = \mathbb{R}^n \times \mathbb{R}^k$ and $f(x, u) = x + Bu$ is linear both in the state and in the control variables. Admissible curves are solutions of

$$\dot{x}(t) = Ax(t) + Bu(t), \qquad x \in \mathbb{R}^n,\ u \in \mathbb{R}^k,$$

where A and B are two $n \times n$ and $n \times k$ matrices, respectively. The cost of an admissible trajectory associated with u is proportional to the square of the L^2-norm of the control

$$J_T(u) = \frac{1}{2}\int_0^T u(t)^* u(t) dt.$$

Since $u : [0, T] \to \mathbb{R}^k$ is measurable and essentially bounded, the trajectory $x(t; x_0)$ associated with u such that $x(0; x_0) = x_0$ is explicitly computed by the Cauchy formula

$$x(t; x_0) = e^{tA}x_0 + \int_0^t e^{(t-s)A} Bu(s) ds.$$

In this case, the bracket-generating condition (A1) is the classical Kalman controllability condition, and reads

$$\operatorname{span}\{B, AB, \dots, A^m B\} = \mathbb{R}^n. \tag{1.28}$$

Since the system is linear, the linearisation along any admissible trajectory coincides with the system itself. Hence it follows that any geodesic is ample and equiregular. In fact, the geodesic growth vector is the same for any geodesic, and is equal to

$$\dim \mathscr{F}^i = \operatorname{rank}\{B, AB, \dots, A^{i-1}B\}, \qquad i \geq 1.$$

A standard computation shows that, under the assumption (1.28), there are no abnormal trajectories. Let us introduce canonical coordinates $(p, x) \in T^*\mathbb{R}^n \simeq \mathbb{R}^{n*} \times \mathbb{R}^n$. Here, it is convenient to treat $p \in \mathbb{R}^{n*}$ as a row vector, and $x \in \mathbb{R}^n$, $u \in \mathbb{R}^k$ as column vectors. The Hamiltonian of the system for normal extremals is

$$\mathcal{H}(p, x, u) = pAx + pBu - \frac{1}{2}u^* u.$$

The maximality condition gives $\bar{u}(p, x) = B^* p^*$. Then, the maximized Hamiltonian is

$$H(p, x) = pAx + \frac{1}{2} pBB^* p^*.$$

For a normal trajectory with initial covector $\lambda = (p_0, x_0)$, we have $p(t; x_0, p_0) = p_0 e^{-tA}$ and

$$x(t; x_0, p_0) = e^{tA} x_0 + e^{tA} \int_0^t e^{-sA} BB^* e^{-sA^*} ds\, p_0^*. \tag{1.29}$$

Let us denote by $C(t)$ the controllability matrix

$$C(t) := \int_0^t e^{-sA} BB^* e^{-sA^*} ds.$$

By Eq. (1.29), we can compute the optimal cost to reach the point $\widetilde{x}(t) = x(t; x_0, p_0)$, starting at point x (close to x_0), in time t, as follows

$$c_t(x) = -S_t(x, \widetilde{x}(t)) = -\frac{1}{2} p_0 C(t) p_0^* + p_0(x - x_0) - \frac{1}{2}(x - x_0)^* C(t)^{-1}(x - x_0).$$

Thus, $d_x^2 \dot{c}_t = -\frac{d}{dt} C(t)^{-1}$, and the family of quadratic forms $\mathcal{Q}_\lambda$, written in terms of the basis defined by the columns of B, is represented by the matrix

$$\mathcal{Q}_\lambda(t) = -B^* \frac{d}{dt} C(t)^{-1} B.$$

The operator $\mathcal{I}_\lambda$ is completely determined by Theorem 1.B. Its eigenvalues coincide with the squares of the Kronecker indices of the control system (see [15]). Moreover, the curvature $\mathcal{R}_\lambda$ is

$$\mathcal{R}_\lambda = -\frac{3}{2} \left.\frac{d^2}{dt^2}\right|_{t=0} \left(t^2 B^* \frac{d}{dt} C(t)^{-1} B \right) = -\frac{3}{2} \left.\frac{d^2}{dt^2}\right|_{t=0} \left(t B^* C(t)^{-1} B \right).$$

We stress that, for this specific case, the operators $\mathcal{I}_\lambda$ and $\mathcal{R}_\lambda$ do not depend neither on the geodesic nor on the initial point since the system is linear (hence it coincides with its linearisation along any geodesic starting at any point).

Remark 1.49. With straightforward but long computations one can generalize these formulae to the case of a quadratic cost with a potential of the form

$$J_T(u) = \frac{1}{2} \int_0^T u(t)^* u(t) + x_u(t)^* Q x_u(t) dt,$$

where Q is a symmetric $n \times n$ matrix, and $x_u(t)$ is the trajectory associated with the control u.

Riemannian geometry

In this example we characterize the operators $\mathcal{Q}_\lambda$ and $\mathcal{I}_\lambda$ for an optimal control system associated with a Riemannian structure. In particular, we show how the operator $\mathcal{R}_\lambda$ is related with the classical sectional curvature.

Let M be an n-dimensional Riemannian manifold. In this case, $\mathbb{U} = TM$, and $f : TM \to TM$ is the identity bundle map. Let $f_1, \ldots, f_n$ be a local orthonormal frame for the Riemannian structure. Any Lipschitz curve on M is admissible, and is a solution of

$$\dot{x} = \sum_{i=1}^n u_i f_i(x), \qquad x \in M,\ u \in \mathbb{R}^n.$$

The cost functional, whose extremals are the classical Riemannian geodesics, is

$$J_T(u) = \frac{1}{2}\int_0^T \sum_{i=1}^n u_i(t)^2 dt.$$

Every geodesic is ample and equiregular, and has trivial growth vector $\mathcal{G}_\gamma = \{n\}$ since, for all $x \in M$, $\mathscr{D}_x = T_x M$. Moreover, the Hamiltonian inner product is equal to the Riemannian inner product. As a standard consequence of the Cauchy-Schwartz inequality, and the fact that Riemannian geodesics have constant speed, the value function S_T can be written in terms of the Riemannian distance $\mathsf{d} : M \times M \to \mathbb{R}$ as follows

$$S_T(x,y) = \frac{1}{2T}\mathsf{d}^2(x,y), \qquad x, y \in M.$$

To any initial covector $\lambda \in T^*_{x_0}M$ corresponds, via the Riemannian structure, an initial vector $v \in T^*_{x_0}M$. We call $\gamma_v : [0,T] \to M$ the associated geodesic, such that $\gamma_v(0) = x_0$ and $\dot\gamma_v(0) = v$. Thus, the geodesic cost function associated with γ_v is

$$c_t(x) = -\frac{1}{2t}\mathsf{d}^2(x, \gamma_v(t)).$$

Then, in order to compute the operators $\mathcal{I}_\lambda$ and $\mathcal{R}_\lambda$ we essentially need an asymptotic expansion of the "squared distance from a geodesic".

Let $\gamma_v(t)$, $\gamma_w(s)$ be two arclength parametrized geodesics, with initial vectors $v, w \in T_{x_0}M$, respectively, starting from x_0. Let us define the function $C(t,s) := \frac{1}{2}d^2(\gamma_v(t), \gamma_w(s))$. It is well known that C is smooth at $(0,0)$ (this is not true in more general settings, such as sub-Riemannian geometry).

Lemma 1.50. *The following formula holds true for the Taylor expansion of $C(t,s)$ at $(0,0)$*

$$C(t,s) = \frac{1}{2}\left(t^2 + s^2 - 2\langle v|w\rangle ts\right) - \frac{1}{6}\langle R(v,w)v|w\rangle t^2 s^2 + t^2 s^2 o(|t| + |s|), \tag{1.30}$$

where $\langle\cdot|\cdot\rangle$ denotes the Riemannian inner product and R is the Riemann tensor.

Proof. Since the geodesics γ_v and γ_w are parametrised by arclength, we have

$$C(t,0) = t^2/2, \qquad C(0,s) = s^2/2, \qquad \forall\, t, s \geq 0. \tag{1.31}$$

Moreover, by standard computations, we obtain

$$\frac{\partial C}{\partial s}(t,0) = -t\langle v|w\rangle, \qquad \frac{\partial C}{\partial t}(0,s) = -s\langle v|w\rangle, \qquad \forall\, t, s \geq 0. \tag{1.32}$$

Eqs. (1.31) and (1.32) imply that the monomials t^n, st^n, s^n, ts^n with $n \geq 2$ do not appear in the Taylor polynomial. The statement is then reduced to the following identity:

$$-\frac{3}{2}\frac{\partial^4 C}{\partial t^2 \partial s^2}(0,0) = \langle R(v,w)v|w\rangle.$$

This identity appeared for the first time in [50, Th. 8.3], in the context of the Ma-Trudinger-Wang curvature tensor, and also in [69, Eq. 14.1]. For a detailed proof one can see also [39, Prop. 1.5.1]. Essentially, this is the very original definition of curvature introduced by Riemann in his famous *Habilitationsvortrag* (see [62]). □

Finally, fix $w \in T_{x_0}M$, and we compute the quadratic form $\langle \mathcal{Q}_\lambda(t)w|w\rangle = d^2_{x_0}\dot{c}_t(w)$

$$d^2_{x_0}\dot{c}_t(w) = \frac{\partial^2}{\partial s^2}\bigg|_{s=0} \frac{\partial}{\partial t}\left(-\frac{1}{2t}d^2(\gamma_v(t),\gamma_w(s))\right) = \frac{\partial}{\partial t}\left(-\frac{1}{t}\frac{\partial^2 C}{\partial s^2}(t,0)\right) = $$
$$= \frac{1}{t^2}\frac{\partial^2 C}{\partial s^2}(0,0) + \frac{1}{3}\left(-\frac{3}{2}\frac{\partial^4 C}{\partial t^2 \partial s^2}(0,0)\right) + O(t),$$

where, in the first equality, we can exchange the order of derivations by the smoothness of $C(t,s)$. By Lemma 1.50 we get $\mathcal{I}_\lambda = \mathbb{I}$ and $\mathcal{R}_\lambda = R(v,\cdot)v$. In particular, $\langle \mathcal{R}_\lambda w|w\rangle = \langle R(v,w)v|w\rangle$ is the Riemannian sectional curvature in the plane generated by $v, w \in T_{x_0}M$.

Remark 1.51. The relation between the operator $\mathcal{R}_\lambda$ and the Riemannian curvature tensor in the Riemannian setting was originally recovered in [17] by using the formalism of Jacobi curves (which we introduce in Section 1.6).

Remark 1.52. In Section 1.5, we apply our theory to the sub-Riemannian setting, where an analogue approach, leading to the Taylor expansion of Eq. (1.30) is not possible, for two major differences between the Riemannian and sub-Riemannian setting. First, geodesics cannot be parametrized by their initial tangent vector. Second, and crucial, for every $x_0 \in M$, the sub-Riemannian squared distance $x \mapsto \mathsf{d}^2(x_0, x)$ is *never* smooth at x_0.

Finsler geometry

The notion of curvature introduced in this chapter recovers not only the classical sectional curvature of Riemannian manifolds, but also the notion of *flag curvature* of Finsler manifolds. These structures can be realized as optimal control problems (in the sense of Section 1.2) by the choice $\mathbb{U} = TM$ and $f : TM \to TM$ equal to the identity bundle map. Moreover the Lagrangian is of the form $L = F^2/2$, where $F \in C^\infty(TM \setminus 0_{TM})$ (0_{TM} is the zero section), is non-negative and positive-homogeneous, i.e. $F(cv) = cF(v)$ for all $v \in TM$ and $c > 0$. Finally L satisfies the Tonelli assumption (A2).

In this setting, it is common to introduce the isomorphism $\tau^* : T^*M \to TM$ (the inverse *Legendre transform*) defined by

$$\tau^*(\lambda) := d_\lambda H_x, \qquad \lambda \in T^*_x M,$$

where H_x is the restriction to the fiber $T^*_x M$ of the Hamiltonian H of the system.

In this case, for all $x \in M$, $\mathscr{D}_x = T_x M$, and the operator $\mathcal{R}_\lambda : T_xM \to T_xM$ can be identified with the Finsler flag curvature operator $R^F_v : T_xM \to T_xM$, where $v = \tau^*(\lambda)$ is the flagpole. A more detailed discussion of Finsler structure and the aforementioned correspondence one can see, for instance, the recent work [59, Example 5.1].

1.5 Sub-Riemannian geometry

In this section we focus on the sub-Riemannian setting. After a brief introduction, we discuss the existence of ample geodesics, the regularity of the geodesic cost and the homogeneity properties of the family $\mathcal{Q}_\lambda$. Then we state the main result of this section about the sub-Laplacian of the sub-Riemannian distance. Finally, we define the concept of geodesic dimension and we investigate the asymptotic rate of growth of the volume of measurable set under sub-Riemannian geodesic homotheties.

1.5.1 Basic definitions

Sub-Riemannian structures are particular affine optimal control system, in the sense of Definition 1.1, where the "drift" vector field is zero and the Lagrangian L is induced by an Euclidean structure on the control bundle $\mathbb{U}$. For a general introduction to sub-Riemannian geometry from the control theory viewpoint we refer to [6]. Other classical references are [31, 56].

Definition 1.53. Let M be a connected, smooth n-dimensional manifold. A *sub-Riemannian structure* on M is a pair $(\mathbb{U}, f)$ where:

- (i) $\mathbb{U}$ is a smooth rank k *Euclidean* vector bundle with base M and fiber $\mathbb{U}_x$, i.e. for every $x \in M$, $\mathbb{U}_x$ is a k-dimensional vector space endowed with an inner product.
- (ii) $f : \mathbb{U} \to TM$ is a smooth *linear* morphism of vector bundles, i.e. f is *linear* on fibers and the following diagram is commutative:

$$\begin{array}{ccc} \mathbb{U} & \xrightarrow{\ f\ } & TM \\ & \underset{\pi_{\mathbb{U}}}{\searrow} & \downarrow{\scriptstyle \pi} \\ & & M \end{array}$$

The maps $\pi_{\mathbb{U}}$ and π are the canonical projections of the vector bundles $\mathbb{U}$ and TM, respectively. Notice that once we have chosen a local trivialization for the vector bundle $\mathbb{U}$, i.e. $\mathbb{U} \simeq M \times \mathbb{R}^k$, we can choose a basis in the fibers and the map f reads $f(x,u) = \sum_{i=1}^k u_i f_i(x)$.

It is always possible to reduce to the case when the control bundle $\mathbb{U}$ is trivial without changing the sub-Riemannian structure defined on it (see [6, 63]). In particular it is not restrictive to assume that the vector fields $f_1, \ldots, f_k$ are globally defined.

Remark 1.54. There is no assumption on the rank of the function f. In other words if we consider, in some choice of the trivialization of $\mathbb{U}$, the vector fields $f_1, \ldots, f_k$, they could be linearly dependent at some (or even at every) point. The structure is Riemannian if and only if $\dim \mathscr{D}_x = n$ for all $x \in M$.

Remark 1.55 (On the notation). Throughout this section, to adhere to the standard notation of the sub-Riemannian literature, we use the notation $X_i = f_i$ for the set of (local) vector fields which define the sub-Riemannian structure.

The Euclidean structure on the fibers induces a metric structure on the *distribution* $\mathscr{D}_x = f(\mathbb{U}_x)$ for all $x \in M$ as follows:

$$\|v\|_x^2 := \min \left\{ \|u\|^2 \;\middle|\; v = f(x,u) \right\}, \qquad \forall\, v \in \mathscr{D}_x. \tag{1.33}$$

It is possible to show that $\|\cdot\|_x$ is a norm on $\mathscr{D}_x$ that satisfies the parallelogram law, i.e. it is actually induced by an inner product $\langle\cdot|\cdot\rangle_x$ on $\mathscr{D}_x$. Notice that the minimum in (1.33) is always attained since we are minimizing an Euclidean norm in $\mathbb{R}^k$ on an affine subspace.

An admissible trajectory for the sub-Riemannian structure is also called *horizontal*, i.e. a Lipschitz curve $\gamma : [0,T] \to M$ such that

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

for some measurable and essentially bounded map $u : [0,T] \to \mathbb{R}^k$.

Remark 1.56. Given an admissible trajectory it is pointwise defined its *minimal control* $u : [0,T] \to \mathbb{R}^k$ such that $\|\dot\gamma(t)\|^2 = \|u(t)\|^2 = \sum_{i=1}^k u_i^2(t)$ for a.e. $t \in [0,T]$. In what follows, whenever we speak about the control associated with a horizontal trajectory, we implicitly assume to consider its minimal control. This is the sub-Riemannian implementation of Remark 1.11

For every admissible curve γ, it is natural to define its *length* by the formula

$$\ell(\gamma) = \int_0^T \|\dot\gamma(t)\| dt = \int_0^T \left(\sum_{i=1}^k u_i^2(t)\right)^{1/2} dt.$$

Since the length is invariant by reparametrization, we can always assume that $\|\dot\gamma(t)\|$ is constant. The *sub-Riemannian (or Carnot-Carathéodory) distance* between two points $x, y \in M$ is

$$\mathsf{d}(x,y) := \inf\{\ell(\gamma) \mid \gamma \text{ horizontal}, \ \gamma(0) = x, \gamma(T) = y\}.$$

It follows from the Cauchy-Schwartz inequality that, if the final time T is fixed, the minima of the length (parametrized with constant speed) coincide with the minima of the energy functional:

$$J_T(\gamma) = \frac{1}{2}\int_0^T \|\dot\gamma(t)\|^2 dt = \frac{1}{2}\int_0^T \sum_{i=1}^k u_i^2(t) dt.$$

Moreover, if γ is a minimizer with constant speed, one has the identity $\ell^2(\gamma) = 2T J_T(\gamma)$.

In particular, the problem of finding the sub-Riemannian geodesics, i.e. curves on M that minimize the distance between two points, coincides with the optimal control problem

$$\begin{aligned} &\dot x = \sum_{i=1}^k u_i X_i(x), \qquad x \in M, \\ &x(0) = x_0,\ x(T) = x_1, \qquad J_T(u) \to \min. \end{aligned}$$

Thus, a sub-Riemannian structure corresponds to an affine optimal control problem (1.5) where $f_0 = 0$ and the Lagrangian $L(x,u) = \frac{1}{2}\|u\|^2$ is induced by the euclidean structure on $\mathbb{U}$. Moreover the value function at time $T > 0$ is closely related with the sub-Riemannian distance as follows:

$$S_T(x,y) = \frac{1}{2T}\mathsf{d}^2(x,y), \qquad x, y \in M,$$

where we have chosen, in the definition of the value function, $M' = M$, even when the latter is not compact (see Definition 1.4). Indeed, the proof of the regularity of the value function in

Appendix C can be adapted by using the fact that small sub-Riemannian balls are compact. The smoothness properties of the sub-Riemannian square distance are discussed in Section 1.5.2 (see also [4]).

Remark 1.57. The assumption (A1) on the control system in the sub-Riemannian case reads $\mathrm{Lie}_x\overline{\mathscr{D}} = T_xM$, for every $x \in M$. This is the classical *bracket-generating* (or *Hörmander*) condition on the distribution $\mathscr{D}$, which implies the controllability of the system, i.e. $\mathsf{d}(x,y) < \infty$ for all $x, y \in M$. Moreover one can show that d induces on M the original manifold's topology. When (M, d) is complete as a metric space, Filippov Theorem guarantees the existence of minimizers joining x to y, for all $x, y \in M$ (see [6, 15]).

The maximality condition (1.11) of PMP reads $u_i(\lambda) = \langle \lambda, X_i(x)\rangle$, where $x = \pi(\lambda)$. Thus the maximized Hamiltonian is

$$H(\lambda) = \frac{1}{2}\sum_{i=1}^{k}\langle \lambda, X_i(x)\rangle^2, \qquad \lambda \in T^*M.$$

It is easily seen that $H : T^*M \to \mathbb{R}$ is also characterized as the dual of the norm on the distribution

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

Since, in this case, H is quadratic on fibers, we obtain immediately the following properties for the exponential map

$$\mathcal{E}_{x_0}(t, s\lambda_0) = \mathcal{E}_{x_0}(ts, \lambda_0), \qquad \lambda_0 \in T^*_{x_0}M, \quad t, s \geq 0,$$

which is tantamount to the fact that the normal geodesic associated with the covector λ_0 is the image of the ray $\{t\lambda_0, t \geq 0\} \subset T^*_{x_0}M$ through the exponential map: $\mathcal{E}_{x_0}(1, t\lambda_0) = \gamma(t)$.

Definition 1.58. Let $\gamma(t) = \pi \circ e^{t\vec{H}}(\lambda_0)$ be a strictly normal geodesic. We say that $\gamma(s)$ is *conjugate* to $\gamma(0)$ along γ if λ_0 is a critical point for $\mathcal{E}_{x_0,s}$, i.e. $D_{\lambda_0}\mathcal{E}_{x_0,s}$ is not surjective.

Remark 1.59. The sub-Riemannian maximized Hamiltonian is a quadratic function on fibers, which implies $d^2_\lambda H_x = 2H_x$, where $H_x = H|_{T^*_xM}$ and $\lambda \in T^*_xM$. In particular $d^2_\lambda H_x$ does not depend on λ and the inner product $\langle\cdot|\cdot\rangle_\lambda$ induced on the distribution $\mathscr{D}_x$ coincides with the sub-Riemannian inner product (see Section 1.4.2).

Nilpotent approximation and privileged coordinates

In this section we briefly recall the concept of nilpotent approximation. For more details we refer to [12, 13, 31, 44]. See also [54] for equiregular structures. The classical presentation that follows relies on the introduction of a set of privileged coordinates; an intrinsic construction can be found in [6].

Let M be a bracket-generating sub-Riemannian manifold. The *flag* of the distribution at a point $x \in M$ is the sequence of subspaces $\mathscr{D}^0_x \subset \mathscr{D}^1_x \subset \mathscr{D}^2_x \subset \ldots \subset T_xM$ defined by

$$\mathscr{D}^0_x := \{0\}, \qquad \mathscr{D}^1_x := \mathscr{D}_x, \qquad \mathscr{D}^{i+1}_x := \mathscr{D}^i_x + [\mathscr{D}^i, \mathscr{D}]_x,$$

where, with a standard abuse of notation, we understand that $[\mathscr{D}^i, \mathscr{D}]_x$ is the vector space generated by the iterated Lie brackets, up to length $i+1$, of local sections of the distribution, evaluated at x. We denote by $\mathfrak{m} = \mathfrak{m}_x$ the *step of the distribution* at x, i.e. the smallest integer such that $\mathscr{D}_x^{\mathfrak{m}_x} = T_xM$. The sub-Riemannian structure is called *equiregular* if $\dim \mathscr{D}_x^i$ does not depend on $x \in M$, for every $i \geq 1$.

Let O_x be an open neighbourhood of the point $x \in M$. We say that a system of coordinates $\psi : O_x \to \mathbb{R}^n$ is *linearly adapted* to the flag if, in these coordinates, $\psi(x) = 0$ and

$$\psi_*(\mathscr{D}_x^i) = \mathbb{R}^{h_1} \oplus \ldots \oplus \mathbb{R}^{h_i}, \qquad \forall\, i = 1, \ldots, \mathfrak{m}_x,$$

where $h_i = \dim \mathscr{D}_x^i - \dim \mathscr{D}_x^{i-1}$ for $i = 1, \ldots, \mathfrak{m}_x$. Indeed $h_1 + \ldots + h_{\mathfrak{m}_x} = n$.

In these coordinates, $x = (x_1, \ldots, x_{\mathfrak{m}_x})$, where $x_i = (x_i^1, \ldots, x_i^{h_i}) \in \mathbb{R}^{h_i}$, and $T_xM = \mathbb{R}^{h_1} \oplus \ldots \oplus \mathbb{R}^{h_{\mathfrak{m}_x}}$. The space of all differential operators in $\mathbb{R}^n$ with smooth coefficients forms an associative algebra with composition of operators as multiplication. The differential operators with polynomial coefficients form a subalgebra of this algebra with generators $1, x_i^j, \partial_{x_i^j}$, where $i = 1, \ldots, m$; $j = 1, \ldots, k_i$. We define weights of generators as follows: $\nu(1) = 0$, $\nu(x_i^j) = i$, $\nu(\partial_{x_i^j}) = -i$, and the weight of monomials accordingly. 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 $\delta_\alpha : \mathbb{R}^n \to \mathbb{R}^n$ defined by $\delta_\alpha(x_1, \ldots, x_m) = (\alpha x_1, \alpha^2 x_2, \ldots, \alpha^{\mathfrak{m}_x} x_{\mathfrak{m}_x}), \alpha > 0$. In particular for a homogeneous vector field X of weight h it holds $\delta_{\alpha*} X = \alpha^{-h} X$.

Let $X \in \mathrm{Vec}(\mathbb{R}^n)$, and consider its Taylor expansion at the origin as a first order differential operator. Namely, we can write the formal expansion

$$X \approx \sum_{h=-\mathfrak{m}_x}^{\infty} X^{(h)},$$

where $X^{(h)}$ is the homogeneous part of degree h of X (notice that every monomial of a first order differential operator has weight not smaller than $-\mathfrak{m}_x$). Define the filtration of $\mathrm{Vec}(\mathbb{R}^n)$

$$\mathrm{Vec}^{(h)}(\mathbb{R}^n) = \{X \in \mathrm{Vec}(\mathbb{R}^n) : X^{(i)} = 0, \forall\, i < h\}, \qquad h \in \mathbb{Z}.$$

Definition 1.60. A system of coordinates $\psi : O_x \to \mathbb{R}^n$ is called *privileged* for the sub-Riemannian structure if they are linearly adapted and $\psi_* X_i \in \mathrm{Vec}^{(-1)}(\mathbb{R}^n)$ for every $i = 1, \ldots, k$.

The existence of privileged coordinates is proved, e.g. in [12, 31]. Notice, however, that privileged coordinates are not unique. Now we are ready to define the sub-Riemannian tangent space of M at x.

Definition 1.61. Given a set of privileged coordinates, the *nilpotent approximation at x* is the sub-Riemannian structure on $T_xM = \mathbb{R}^n$ defined by the set of vector fields $\widehat{X}_1, \ldots, \widehat{X}_k$, where $\widehat{X}_i := (\psi_* X_i)^{(-1)} \in \mathrm{Vec}(\mathbb{R}^n)$.

The definition is well posed, in the sense that the structures obtained by different sets of privileged coordinates are isometric (see [31, Proposition 5.20]). Then, in what follows we omit the coordinate map in the notation above, identifying $T_xM = \mathbb{R}^n$ and a vector field with its coordinate expression in $\mathbb{R}^n$. The next proposition also justifies the name of the sub-Riemannian tangent space (see [31, Proposition 5.17]).

Proposition 1.62. *The vector fields* $\widehat{X}_1, \dots \widehat{X}_k$ *generate a nilpotent Lie algebra* $\mathrm{Lie}(\widehat{X}_1, \dots, \widehat{X}_k)$ *of step* $\mathfrak{m}_x$*. At any point* $z \in \mathbb{R}^n$ *they satisfy the bracket-generating assumption, namely* $\mathrm{Lie}_z(\widehat{X}_1, \dots, \widehat{X}_k) = \mathbb{R}^n$*.*

Remark 1.63*.* The sub-Riemannian distance $\widehat{\mathsf{d}}$ on the nilpotent approximation is homogeneous with respect to dilations δ_α, i.e. $\widehat{\mathsf{d}}(\delta_\alpha(x), \delta_\alpha(y)) = \alpha\, \widehat{\mathsf{d}}(x, y)$.

Definition 1.64. Let $X_1, \dots, X_k$ be a set of vector fields which defines the sub-Riemannian structure on M and fix a system of privileged coordinates at $x \in M$. The ε-approximated system at x is the sub-Riemannian structure induced by the vector fields $X_1^\varepsilon, \dots, X_k^\varepsilon$ defined by

$$X_i^\varepsilon := \varepsilon \delta_{1/\varepsilon *} X_i, \qquad i = 1, \dots, k.$$

The following lemma is a consequence of the definition of ε-approximated system and privileged coordinates.

Lemma 1.65. $X_i^\varepsilon \to \widehat{X}_i$ *in the* C^∞ *topology of uniform convergence of all derivatives on compact sets in* $\mathbb{R}^n$ *when* $\varepsilon \to 0$*, for* $i = 1, \dots, k$*.*

Therefore, the nilpotent approximation $\widehat{X}$ of a vector field X at a point x is the "principal part" in the expansion when one considers the blown up coordinates near the point x, with rescaled distances.

Approximating trajectories

In this subsection we show, in a system of privileged coordinates $\psi : O_x \to \mathbb{R}^n$, how the normal trajectories of the ε-approximated system converge to corresponding normal trajectories of the nilpotent approximation.

Let $H^\varepsilon : T^*\mathbb{R}^n \to \mathbb{R}$ be the maximized Hamiltonian for the ε-approximated system, and $\mathcal{E}^\varepsilon : T_0^*\mathbb{R}^n \to \mathbb{R}^n$ the corresponding exponential map (starting at 0). We denote by the symbols $\widehat{H}$ and $\widehat{\mathcal{E}}$ the analogous objects for the nilpotent approximation. The ε-approximated normal trajectory $\gamma^\varepsilon(t)$ converges to the corresponding nilpotent trajectory $\widehat{\gamma}(t)$.

Proposition 1.66. *Let* $\lambda_0 \in T_0^*\mathbb{R}^n$*. Let* $\gamma^\varepsilon : [0, T] \to \mathbb{R}^n$ *and* $\widehat{\gamma} : [0, T] \to \mathbb{R}^n$ *be the normal geodesics associated with* λ_0 *for the* ε*-approximating system and for the nilpotent system, respectively. Let* $u^\varepsilon : [0, T] \to \mathbb{R}^k$ *and* $\widehat{u} : [0, T] \to \mathbb{R}^k$ *be the associated controls. Then there exists a neighbourhood* $O_{\lambda_0} \subset T_0^*\mathbb{R}^n$ *of* λ_0 *such that for* $\varepsilon \to 0$

(i) $\mathcal{E}^\varepsilon \to \widehat{\mathcal{E}}$ *in the* C^∞ *topology of uniform convergence of all derivatives on* O_{λ_0}*,*

(ii) $\gamma^\varepsilon \to \widehat{\gamma}$ *in the* C^∞ *topology of uniform convergence of all derivatives on* $[0, T]$*,*

(iii) $u^\varepsilon \to \widehat{u}$ *in the* C^∞ *topology of uniform convergence of all derivatives on* $[0, T]$*.*

The proof of Proposition 1.66 is a consequence of a more general statement for the Hamiltonian flow of the approximating systems, which can be found in Appendix D.

1.5.2 Ample geodesics and regularity of the squared distance

In this section we discuss the existence of ample geodesics in sub-Riemannian geometry and its implications on the regularity of the sub-Riemannian distance.

We start by proving a bound on the growth vector of a normal geodesic in terms of the growth vector on the nilpotent system and the growth vector of the distribution. Let us write the control system and its nilpotent approximation at a point $x_0 \in M$ in privileged coordinates:

$$\dot{x} = f(x,u) = \sum_{i=1}^{k} u_i X_i(x), \qquad \dot{x} = \widehat{f}(x,u) = \sum_{i=1}^{k} u_i \widehat{X}_i(x), \qquad x \in \mathbb{R}^n.$$

Fix an initial covector λ_0, and let $\gamma : [0,T] \to M$ be the corresponding normal geodesic. Let $\mathscr{F}_\gamma(t)$ be the family associated with γ. Moreover, let $\widehat{\gamma} : [0,T] \to \mathbb{R}^n$ be the normal geodesic associated with the same initial covector in some set of privileged coordinates. Besides, let $\widehat{\mathscr{F}_\gamma}(t)$ be the family associated with $\widehat{\gamma}$. Analogously we define the steps $m(t)$ and $\widehat{m}(t)$ of γ and $\widehat{\gamma}$, respectively. Recall that $\mathfrak{m}_x$ is the step of the distribution $\mathscr{D}$ at x.

Lemma 1.67. *The following inequalities hold true, for $t = 0$:*

(i) $\dim \widehat{\mathscr{F}}^i_\gamma(t) \leq \dim \mathscr{F}^i_\gamma(t) \leq \dim \mathscr{D}^i_{\gamma(t)}$,

(ii) $\mathfrak{m}_{\gamma(t)} \leq m(t) \leq \widehat{m}(t)$.

Proof. Claim (ii) follows from (i). The right inequality in (i) is a direct consequence of the alternative definition of the flag of the geodesic given in Section 1.3.4.

The left inequality in (i) can be proved considering an ε-approximating structure for the nilpotent approximation in privileged coordinates, and applying the criterion of Section 1.3.2. Such a system writes, in privileged coordinates

$$\dot{x} = f^\varepsilon(x,u) = \sum_{i=1}^{k} u_i X_i^\varepsilon(x), \qquad x \in \mathbb{R}^n.$$

Let, as usual, $\gamma^\varepsilon : [0,T] \to \mathbb{R}^n$ and $u^\varepsilon : [0,T] \to \mathbb{R}^k$ be the normal geodesic and the normal control associated with the fixed covector λ_0 on the ε-approximating structures. Moreover, let $\widehat{\gamma}$, $\widehat{u}$ be the analogous objects on the nilpotent structure. The C^∞ uniform convergence guaranteed by Lemma 1.65 and Proposition 1.66, respectively, imply that, for $\varepsilon \to 0$

$$A^\varepsilon(t) := \frac{\partial f^\varepsilon}{\partial x}(\gamma^\varepsilon(t), u^\varepsilon(t)) \longrightarrow \widehat{A}(t) := \frac{\partial \widehat{f}}{\partial x}(\widehat{\gamma}(t), \widehat{u}(t)),$$
$$B^\varepsilon(t) := \frac{\partial f^\varepsilon}{\partial u}(\gamma^\varepsilon(t), u^\varepsilon(t)) \longrightarrow \widehat{B}(t) := \frac{\partial \widehat{f}}{\partial u}(\widehat{\gamma}(t), \widehat{u}(t)),$$

together with their derivatives, uniformly on $[0,T]$. For $t = 0$, the matrices $A^\varepsilon(t)$ and $B^\varepsilon(t)$ represent the linearisation of the ε-approximated system or, equivalently, of the original system in a set of ε-dependant coordinates. Besides they converge, together with their derivatives, to the same objects for the nilpotent system. The inequality is then a consequence of the criterion in Section 1.3.2 and the semicontinuity of the rank. □

Next, we provide a characterization for smooth points of the squared distance (which is the value function for the sub-Riemannian optimal control problem). Let $x_0 \in M$, and let $\Sigma_{x_0} \subset M$ be the set of points x such that there exists a unique non-conjugate and non-abnormal minimizer $\gamma : [0,1] \to M$ joining x_0 with x.

Theorem 1.68 (see [4])**.** *Let $x_0 \in M$ and set $\mathfrak{f} := \frac{1}{2}\mathsf{d}^2(x_0,\cdot)$. Then Σ_{x_0} is open, dense and $\mathfrak{f}$ is smooth on Σ_{x_0}. Moreover if $x \in \Sigma_{x_0}$ then $d_x\mathfrak{f} = \lambda(1)$, where $\lambda(t)$ is the normal lift of $\gamma(t)$.*

This result, together with Lemma 1.67, imply that the set of covectors such that the associated geodesic is ample is open and dense in the fiber.

Lemma 1.69. *The set $A_{x_0} \subseteq T^*_{x_0}M$ of covectors such that the corresponding normal geodesic is ample is a nonempty, Zariski open set, therefore dense.*

Proof. The fact that A_{x_0} is an open Zariski subset is a consequence of Remark 1.110, for any polynomial Hamiltonian, since in this case the non ampleness is an algebraic condition on the fibre. We only need to prove that A_{x_0} is nonempty. To this end, by Lemma 1.67, it is sufficient to prove that there is at least one ample geodesic on the nilpotent approximation at x_0. The nilpotent approximation structure is analytic, hence a geodesic is ample if and only if it is strictly normal (more precisely strictly $\Leftrightarrow$ strongly $\Leftrightarrow$ ample, see Proposition 1.25). The existence of at least one strictly normal geodesic then follows by Theorem 1.68. □

1.5.3 Reparametrization and homogeneity of the curvature operator

We already explained that a geodesic is not ample on a proper Zariski closed subset of the fibre. This set includes covectors associated to abnormal geodesics, since $\mathscr{D}_x^\perp \subset T^*_xM \setminus A_x$. On the other hand, for $\lambda \in A_x$, the curvature $\mathcal{R}_\lambda$ is well defined. Observe that A_x is invariant by rescaling, i.e. if $\lambda \in A_x$, then for $\alpha \neq 0$, also $\alpha\lambda \in A_x$. Therefore, we have the following:

Proposition 1.70. *The operators $\mathcal{I}_\lambda$ and $\mathcal{R}_\lambda$ are homogeneous of degree 0 and 2 with respect to λ, respectively. Namely, for $\lambda \in A_x$ and $\alpha > 0$*

$$\mathcal{I}_{\alpha\lambda} = \mathcal{I}_\lambda, \qquad \mathcal{R}_{\alpha\lambda} = \alpha^2 \mathcal{R}_\lambda. \tag{1.34}$$

Proof. Let c_t^λ be the geodesic cost associated with the covector $\lambda \in T^*_xM$. By homogeneity of the sub-Riemannian Hamiltonian, for $\alpha > 0$ we have

$$c_t^{\alpha\lambda} = \alpha\, c_{\alpha t}^\lambda.$$

In particular, this implies $d_x^2 \dot{c}_t^{\alpha\lambda} = \alpha^2 d_x^2 \dot{c}_{\alpha t}^\lambda$. The same relation is true for the restrictions to the distribution $\mathscr{D}_x$, therefore $\mathcal{Q}_{\alpha\lambda}(t) = \alpha^2 \mathcal{Q}_\lambda(\alpha t)$ as symmetric operators on $\mathscr{D}_x$. Applying Theorem 1.A to both families one obtains

$$\frac{1}{t^2}\mathcal{I}_{\alpha\lambda} + \frac{1}{3}\mathcal{R}_{\alpha\lambda} + O(t) = \alpha^2 \left(\frac{1}{\alpha^2 t^2}\mathcal{I}_\lambda + \frac{1}{3}\mathcal{R}_\lambda + O(\alpha t) \right),$$

which, in particular, implies Eq. (1.34). □

Notice that the same proof applies also to a general affine optimal control system, such that the Hamiltonian (or, equivalently, the Lagrangian) is homogeneous of degree two.

1.5.4 Asymptotics of the sub-Laplacian of the geodesic cost

In this section we discuss the asymptotic behaviour of the sub-Laplacian of the sub-Riemannian geodesic cost. On a Riemannian manifold, the Laplace-Beltrami operator is defined as the divergence of the gradient. This definition can be easily generalized to the sub-Riemannian setting. We will denote by $\langle\cdot|\cdot\rangle$ the inner product defined on the distribution.

Definition 1.71. Let $f \in C^\infty(M)$. The *horizontal gradient* of f is the unique horizontal vector field ∇f such that

$$\langle \nabla f | X\rangle = X(f), \qquad \forall X \in \overline{\mathscr{D}}.$$

For $x \in M$, the restriction of the sub-Riemannian Hamiltonian to the fiber $H_x : T_x^*M \to \mathbb{R}$ is a quadratic form. Then, as a consequence of the formula $\langle d_\lambda H_x | X\rangle = \langle \lambda, X\rangle$, we obtain

$$\nabla f = \sum_{i=1}^k X_i(f) X_i. \tag{1.35}$$

We want to stress that Eq. (1.35) is true in full generality, also when $\dim \mathscr{D}_x$ is not constant or the vectors $X_1, \dots, X_k$ are not independent.

Definition 1.72. Let $\mu \in \Omega^n(M)$ be a volume form, and $X \in \mathrm{Vec}(M)$. The *μ-divergence* of X is the smooth function $\mathrm{div}_\mu(X)$ defined by

$$\mathcal{L}_X \mu := \mathrm{div}_\mu(X)\mu,$$

where, we recall, $\mathcal{L}_X$ is the Lie derivative in the direction of X.

Notice that the definition of divergence does not depend on the orientation of M, namely the sign of μ. The divergence measures the rate at which the volume of a region changes under the integral flow of a field. Indeed, for any compact $\Omega \subset M$ and t sufficiently small, let $e^{tX} : \Omega \to M$ be the flow of $X \in \mathrm{Vec}(M)$, then

$$\left.\frac{d}{dt}\right|_{t=0} \int_{e^{tX}(\Omega)} \mu = -\int_\Omega \mathrm{div}_\mu(X)\mu.$$

The next proposition is an easy consequence of the definition of μ-divergence and is sometimes employed as an alternative definition of the latter.

Proposition 1.73. *Let $C_0^\infty(M)$ be the space of smooth functions with compact support. For any $f \in C_0^\infty(M)$ and $X \in \mathrm{Vec}(M)$*

$$\int_M f \mathrm{div}_\mu(X)\mu = -\int_M X(f)\mu.$$

With a divergence and a gradient at our disposal, we are ready to define the sub-Laplacian associated with the volume form μ.

Definition 1.74. Let $\mu \in \Omega^n(M)$, $f \in C^\infty(M)$. The *sub-Laplacian* associated with μ is the second order differential operator

$$\Delta_\mu f := \mathrm{div}_\mu\left(\nabla f\right),$$

On a Riemannian manifold, when μ is the Riemannian volume, this definition reduces to the Laplace-Beltrami operator. As a consequence of Eq. (1.35) and the Leibniz rule for the divergence $\operatorname{div}_\mu(fX) = X(f) + f\operatorname{div}_\mu(X)$, we can write the sub-Laplacian in terms of the fields $X_1, \dots, X_k$:

$$\operatorname{div}_\mu(\nabla f) = \sum_{i=1}^k \operatorname{div}_\mu(X_i(f)X_i) = \sum_{i=1}^k X_i(X_i(f)) + \operatorname{div}_\mu(X_i)X_i(f).$$

Then

$$\Delta_\mu = \sum_{i=1}^k X_i^2 + \operatorname{div}_\mu(X_i)X_i. \tag{1.36}$$

Remark 1.75. If we apply Proposition 1.73 to the horizontal gradient ∇g, we obtain

$$\int_M f\Delta g\mu = -\int_M \langle \nabla f|\nabla g\rangle \mu, \qquad \forall f, g \in C_0^\infty(M).$$

Then Δ_μ is symmetric and negative on $C_0^\infty(M)$. It can be proved that it is also essentially self-adjoint (see [66]). Hence it admits a unique self-adjoint extension to $L^2(M,\mu)$.

Observe that the principal symbol of Δ_μ, which is a function on T^*M, does not depend on the choice of μ, and is proportional to the sub-Riemannian Hamiltonian, namely $2H : T^*M \to \mathbb{R}$. The sub-Laplacian depends on the choice of the volume μ according to the following lemma.

Lemma 1.76. *Let $\mu, \mu' \in \Omega^n(M)$ be two volume forms such that $\mu' = e^a\mu$ for some $a \in C^\infty(M)$. Then*

$$\Delta_{\mu'} f = \Delta_\mu f + \langle \nabla a|\nabla f\rangle.$$

Proof. It follows from the Leibniz rule $\mathcal{L}_X(a\mu) = X(a)\mu + a\mathcal{L}_X\mu = (X(\log a) + \operatorname{div}_\mu(X))a\mu$ for every $a \in C^\infty(M)$. □

The sub-Laplacian, computed at critical points, does not depend on the choice of the volume.

Lemma 1.77. *Let $f \in C^\infty(M)$, and let $x \in M$ be a critical point of f. Then, for any choice of the volume μ,*

$$\Delta_\mu f|_x = \sum_{i=1}^k X_i^2(f)|_x.$$

Proof. The proof follows from Eq. (1.36), and the fact that $X_i(f)|_x = 0$. □

From now on, when computing the sub-Laplacian of a function at a critical point, we employ the notation $\Delta_\mu f|_x = \Delta f|_x$, since it does not depend on the volume.

Lemma 1.78. *Let $f \in C^\infty(M)$, and let $x \in M$ be a critical point of f. Then $\Delta f|_x = \operatorname{tr} d_x^2 f|_{\mathscr{D}_x}$.*

Proof. Recall that if x is a critical point of f, then the second differential $d_x^2 f$ is the quadratic form associated with the symmetric bilinear form

$$d_x^2 f : T_xM \times T_xM \to \mathbb{R}, \qquad (X, Y) \mapsto X(Y(f))|_x.$$

The restriction of $d_x^2 f$ to the distribution can be associated, via the inner product, with a symmetric operator defined on $\mathscr{D}_x$, whose trace is computed in terms of $X_1, \dots, X_k$ as follows

$$\operatorname{tr} d_x^2 f|_{\mathscr{D}_x} = \sum_{i=1}^{k} X_i^2(f)|_x, \tag{1.37}$$

We stress that Eq. (1.37) holds true for any set of generators, not necessarily linearly independent, of the sub-Riemannian structure $X_1, \dots, X_k$ such that $H(\lambda) = \frac{1}{2}\sum_{i=1}^{k}\langle \lambda, X_i\rangle^2$. The statement now is a direct consequence of Lemma 1.77. □

Remember that the derivative of the geodesic cost function $\dot{c}_t$ has a critical point at $x_0 = \gamma(0)$. As a direct consequence of Theorem 1.A, 1.B, Lemma 1.78 and the fact that, in the sub-Riemannian case, the Hamiltonian inner product is the sub-Riemannian one (see Remark 1.45), we get the following asymptotic expansion:

Theorem 1.79. *Let c_t be the geodesic cost associated with a geodesic γ such that $\gamma(0) = x_0$. Then*

$$\Delta \dot{c}_t|_{x_0} = \frac{\operatorname{tr} \mathcal{I}_\lambda}{t^2} + \frac{1}{3}\operatorname{tr} \mathcal{R}_\lambda + O(t).$$

The next result is an explicit expression for the asymptotic of the sub-Laplacian of the geodesic cost computed at the initial point x_0 of the geodesic γ. In the sub-Riemannian case, the geodesic cost is essentially the squared distance from the geodesic, i.e. the function

$$\mathfrak{f}_t(\cdot) := -tc_t(\cdot) = \frac{1}{2}\mathsf{d}^2(\,\cdot\,, \gamma(t)), \qquad t \in (0, 1].$$

For this reason, we may state the theorem equivalently in terms of $\mathfrak{f}_t$ or the geodesic cost c_t. Remember also that, since x_0 is not a critical point of $\mathfrak{f}_t$, its sub-Laplacian depends on the choice of the volume form μ.

Theorem 1.C. *Let γ be an equiregular geodesic with initial covector $\lambda \in T_{x_0}^*M$. Assume also that $\dim \mathscr{D}$ is constant in a neighbourhood of x_0. Then there exists a smooth n-form ω defined along γ, such that for any volume form μ on M, $\mu_{\gamma(t)} = e^{g(t)}\omega_{\gamma(t)}$, we have*

$$\Delta_\mu \mathfrak{f}_t|_{x_0} = \operatorname{tr} \mathcal{I}_\lambda - \dot{g}(0)t - \frac{1}{3}\operatorname{tr} \mathcal{R}_\lambda t^2 + O(t^3), \tag{1.38}$$

As a consequence, for any choice of the volume form μ

$$\lim_{t\to 0} \Delta_\mu \mathfrak{f}_t\big|_{x_0} = \operatorname{tr} \mathcal{I}_\lambda,$$
$$\left.\frac{d^2}{dt^2}\right|_{t=0} \Delta_\mu \mathfrak{f}_t\big|_{x_0} = -\frac{2}{3}\operatorname{tr} \mathcal{R}_\lambda.$$

The proof Theorem 1.C is postponed to Section 1.8.

Observe that only the first order term in t of Eq. (1.38) depends on the choice of the volume. The explicit expression of ω is not relevant here, and requires the premature introduction of some technical tools which we deemed not necessary at this point. We only anticipate that ω, which indeed depends on γ, is related with a generalization of the parallel transport of the volume form along the geodesic. On a Riemannian manifold, ω does not depend on γ and, up to a sign, is equal to the Riemannian volume form. Therefore the first order term in Eq. (1.38) vanishes. This is not true, in general, for sub-Riemannian manifolds.

1.5.5 Equiregular distributions

In this section we focus on equiregular sub-Riemannian structures, endowed with a smooth, intrinsic volume form, called Popp's volume. Then we introduce a special class of equiregular distributions, that we call *slow growth*. In this case, we define a family of smooth operators in terms of which the asymptotic expansion of Theorem 1.C (and in particular its linear term) can be expressed explicitly.

Recall that a bracket generating sub-Riemannian manifold M is *equiregular* if $\dim \mathscr{D}^i_x$ does not depend on $x \in M$, for every $i \geq 0$, where $\mathscr{D}^0_x \subset \mathscr{D}^1_x \subset \mathscr{D}^2_x \subset \ldots \subset T_xM$ is the flag of the distribution at a point $x \in M$ (see Section 1.5).

Popp's volume

In this section we provide the definition of Popp's volume for an equiregular sub-Riemannian structure. Our presentation follows closely the one of [26, 56]. The definition rests on the following lemmas, whose proof is not repeated here.

Lemma 1.80. *Let E be an inner product space, and let $\pi : E \to V$ be a surjective linear map. Then π induces an inner product on V such that the norm of $v \in V$ is*

$$\|v\|_V = \min\{\|e\|_E \text{ s.t. } \pi(e) = v\}.$$

Lemma 1.81. *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 $\theta : \wedge^n E \to \wedge^n \mathrm{gr}(F)$.*

The idea behind Popp's volume is to define an inner product on each $\mathscr{D}^i_x/\mathscr{D}^{i-1}_x$ which, in turn, induces an inner product on the orthogonal direct sum

$$\mathrm{gr}_x(\mathscr{D}) = \mathscr{D}_x \oplus \mathscr{D}^2_x/\mathscr{D}_x \oplus \ldots \oplus \mathscr{D}^{\mathfrak{m}}_x/\mathscr{D}^{\mathfrak{m}-1}_x.$$

The latter has a natural volume form, which is the canonical volume of an inner product space obtained by wedging the elements of an orthonormal dual basis. Then, we employ Lemma 1.81 to define an element of $(\wedge^n T_xM)^* \simeq \wedge^n T^*_xM$, which is Popp's volume form computed at x.

Fix $x \in M$. Then, let $v, w \in \mathscr{D}_x$, and let V, W be any horizontal extensions of v, w. Namely, $V, W \in \overline{\mathscr{D}}$ and $V(x) = v$, $W(x) = w$. The linear map $\pi : \mathscr{D}_x \otimes \mathscr{D}_x \to \mathscr{D}^2_x/\mathscr{D}_x$

$$\pi(v \otimes w) := [V, W]_x \mod \mathscr{D}_x, \tag{1.39}$$

is well defined, and does not depend on the choice the horizontal extensions. Similarly, let $1 \leq i \leq \mathfrak{m}$. The linear maps $\pi_i : \otimes^i \mathscr{D}_x \to \mathscr{D}_x^i/\mathscr{D}_x^{i-1}$

$$\pi_i(v_1 \otimes \cdots \otimes v_i) = [V_1, [V_2, \ldots, [V_{i-1}, V_i]]]_x \mod \mathscr{D}_x^{i-1}, \tag{1.40}$$

are well defined and do not depend on the choice of the horizontal extensions $V_1, \ldots, V_i$ of $v_1, \ldots, v_i$.

By the bracket-generating condition, the maps π_i are surjective and, by Lemma 1.80, they induce an inner product space structure on $\mathscr{D}_x^i/\mathscr{D}_x^{i-1}$. Therefore, the nilpotentization of the distribution at x, namely $\mathrm{gr}_x(\mathscr{D})$, 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) $\eta_x \in \wedge^n \mathrm{gr}_x(\mathscr{D})^*$, which is the volume form obtained by wedging the elements of an orthonormal dual basis.

Finally, Popp's volume (computed at the point x) is obtained by transporting the volume of $\mathrm{gr}_x(\mathscr{D})$ to T_xM through the map $\theta_x : \wedge^n T_xM \to \wedge^n \mathrm{gr}_x(\mathscr{D})$ defined in Lemma 1.81. Namely

$$\mathcal{P}_x = \eta_x \circ \theta_x, \tag{1.41}$$

where we employ the canonical identification $(\wedge^n T_xM)^* \simeq \wedge^n T_x^*M$. Eq. (1.41) is defined only in the domain of the chosen local frame. If M is orientable, with a standard argument, these n-forms can be glued together to obtain Popp's volume $\mathcal{P} \in \Omega^n(M)$. Notice that Popp's volume is smooth by construction.

Remark 1.82. From Eq. (1.39) and (1.40) it follows that, for any $i \geq 0$ and $V \in \mathscr{D}_x$ the linear maps $\mathrm{ad}_x^i V : \mathscr{D}_x \to \mathscr{D}_x^{i+1}/\mathscr{D}_x^i$ given by

$$\mathrm{ad}_x^i V(W) := \underbrace{[V, [V, \ldots, [V}_{i \text{ times}}, W]]]_x \mod \mathscr{D}_x^i, \qquad W \in \mathscr{D}_x,$$

are well-defined.

Slow growth distributions

Now we are ready to introduce the following class of equiregular distributions.

Definition 1.83. An equiregular distribution is *slow growth* at $x \in M$ if there exists a vector $\mathsf{T} \in \mathscr{D}_x$ such that the linear map $\mathrm{ad}_x^i \mathsf{T}$ is surjective for all $i \geq 0$.

This condition is actually generic in T, as stated by the following proposition.

Proposition 1.84. *Let $\mathscr{D}$ be a slow growth distribution at x. Then, for T in a nonempty open Zariski subset of $\mathscr{D}_x$, all the linear maps $\mathrm{ad}_x^i \mathsf{T}$ are surjective.*

Proof. Let X_i be an orthonormal basis for $\mathscr{D}_x$ and write $\mathsf{T} = \sum_{j=1}^k \alpha_j X_j$, where $k = \dim \mathscr{D}_x$ and the α_j are constant. The definition of slow growth is a maximal rank condition on the operators $\mathrm{ad}_x^i \mathsf{T} = (\sum_{j=1}^k \alpha_j \mathrm{ad}_x X_j)^i$, which is satisfied by at least one element of $\mathscr{D}_x$. Then, the result follows from the fact that $\mathrm{ad}_x^i \mathsf{T}$ depends polynomially on the α_j. □

We say that a distribution $\mathscr{D}$ is *slow growth* if it is slow growth at every point $x \in M$. Familiar sub-Riemannian structures such as contact, quasi-contact, fat, Engel, Goursat-Darboux distributions (see [36]) are examples of slow growth distributions.

Now, for any fixed equiregular, ample (of step m) geodesic $\gamma : [0,T] \to M$, with flag $0 = \mathscr{F}^0_{\gamma(t)} \subset \mathscr{F}^1_{\gamma(t)} \subset \ldots \subset \mathscr{F}^m_{\gamma(t)} = T_{\gamma(t)}M$ recall the smooth families of operators

$$\mathcal{L}^i_{\mathsf{T}} : \mathscr{F}_{\gamma(t)} \to \mathscr{F}^{i+1}_{\gamma(t)}/\mathscr{F}^i_{\gamma(t)}, \qquad i = 0, \ldots, m-1,$$

defined for all $t \in [0,T]$ in terms of an admissible extension T of $\dot\gamma$ (see Remark 1.38). If the distribution is slow growth, we have the identities $\mathcal{L}^i_{\mathsf{T}} = \mathrm{ad}^i_{\gamma(t)}\mathsf{T}$ which, in particular, say that $\mathcal{L}^i_{\mathsf{T}}$ depend only on the value of T at $\gamma(t)$. Moreover, the following growth condition is satisfied

$$\dim \mathscr{F}^i_\gamma = \dim \mathscr{D}^i, \qquad \forall\, i \geq 0. \tag{1.42}$$

As a consequence of Proposition 1.84 it follows that, for a nonempty Zariski open set of initial covectors, the corresponding geodesic is ample (of step $m = \mathfrak{m}$, the step of the distribution), equiregular and satisfies the growth condition of Eq. (1.42).

Next, recall that given V, W inner product spaces, any surjective linear map $L : V \to W$ descends to an isomorphism $L : V/\ker L \to W$. Then, thanks to the inner product structure, we can consider the map $L^* \circ L : V/\ker L \to V/\ker L$ obtained by composing L with its adjoint L^*, which is a symmetric invertible operator. Applying this construction to our setting, we define the smooth families of symmetric operators

$$M_i(t) := (\mathcal{L}^{i-1}_{\mathsf{T}})^* \circ \mathcal{L}^{i-1}_{\mathsf{T}} : \mathscr{D}_{\gamma(t)}/\ker \mathcal{L}^{i-1}_{\mathsf{T}} \to \mathscr{D}_{\gamma(t)}/\ker \mathcal{L}^{i-1}_{\mathsf{T}}, \qquad i = 1, \ldots, m. \tag{1.43}$$

We are now ready to specify Theorem 1.C for any ample, equiregular geodesic satisfying the growth condition of Eq. (1.42). First, let us discuss the zeroth order term of the expansion. Recall that the Hausdorff dimension of an equiregular sub-Riemannian manifold is computed by Mitchell's formula (see [31, 54]), namely

$$Q = \sum_{i=1}^m i(\dim \mathscr{D}^i - \dim \mathscr{D}^{i-1}).$$

Thus, for a slow growth distribution and a geodesic γ with initial covector $\lambda \in T^*_{x_0}M$ satisfying the growth condition of Eq. (1.42), we have the following identity (see also Remark 1.48)

$$\begin{aligned}\operatorname{tr} \mathcal{I}_\lambda &= \sum_{i=1}^m (2i-1)(\dim \mathscr{F}^i_\gamma - \dim \mathscr{F}^{i-1}_\gamma) = \\ &= \sum_{i=1}^m (2i-1)(\dim \mathscr{D}^i - \dim \mathscr{D}^{i-1}) = 2Q - n.\end{aligned}$$

This formula gives the zeroth order term of the following theorem.

Theorem 1.85. *Let M be a sub-Riemannian manifold with a slow growth distribution $\mathscr{D}$. Let γ be an ample, equiregular geodesic with initial covector $\lambda \in T^*_{x_0}M$ satisfying the growth condition* (1.42)*. Then*

$$\Delta_\mu \mathfrak{f}_t|_{x_0} = (2Q - n) - \frac{1}{2}\sum_{i=1}^{m} \operatorname{tr}\left(M_i(0)^{-1}\dot{M}_i(0)\right) t - \frac{1}{3}\operatorname{tr}\mathcal{R}_\lambda t^2 + O(t^3). \tag{1.44}$$

where the smooth families of operators $M_i(t)$ are defined by Eq. (1.43)*.*

Remark 1.86. Equivalently we can write Eq. (1.44) in the following form

$$\Delta_\mu \mathfrak{f}_t|_{x_0} = (2Q - n) - \frac{1}{2}\left(\frac{d}{ds}\bigg|_{s=0}\sum_{i=1}^{m} \log\det M_i(s)\right) t - \frac{1}{3}\operatorname{tr}\mathcal{R}_\lambda t^2 + O(t^3).$$

The proof of Theorem 1.85 is postponed to the end of Section 1.8. We end this section with an example.

Example 1.87 (Riemannian structures)**.** In a Riemannian structure (see Section 1.4.4), any nontrivial geodesic has the same flag $\mathscr{F}_{\gamma(t)} = \mathscr{D}_{\gamma(t)} = T_{\gamma(t)}M$. In particular, it is a trivial example of slow growth distribution. Notice that Popp's volume reduces to the usual Riemannian volume form. Since every geodesic is ample with step $m = 1$, there is only one family of operators associated with $\gamma(t)$, namely the constant operator $M_1(t) = \mathbb{I}|_{T_{\gamma(t)}M}$. Thus, in this case, the linear term of Theorem 1.85 vanishes, and we obtain

$$\Delta \mathfrak{f}_t|_{x_0} = n - \frac{1}{3}\operatorname{tr}\mathcal{R}_\lambda t^2 + O(t^3),$$

where $\operatorname{tr}\mathcal{R}_\lambda = \operatorname{Ric}(\dot{\gamma}(0))$ is the classical Ricci curvature in the direction of the geodesic.

In Section 1.5.7 we compute explicitly the asymptotic expansion of Theorem 1.85 in the case of the Heisenberg group, endowed with its canonical volume. A more general class of slow growth sub-Riemannian distributions, in which the operators $M_i(t)$ are not trivial and can be computed explicitly, namely contact structures, will appear in a forthcoming paper [8].

1.5.6 Geodesic dimension and sub-Riemannian homotheties

In this section, M is a complete, connected, orientable sub-Riemannian manifold, endowed with a smooth volume form μ. With a slight abuse of notation, we denote by the same symbol the induced measure on M. We are interested in sub-Riemannian homotheties, namely contractions along geodesics. To this end, let us fix $x_0 \in M$, which will be the center of the homothety. Recall that Σ_{x_0} is the set of points $x \in M$ such that there exists a unique non-conjugate and non-abnormal minimizer $\gamma : [0,1] \to M$ that joins x_0 with x. Recall also that, by Theorem 1.68, $\Sigma_{x_0} \subset M$ is the open and dense set where the function $\mathfrak{f} = \frac{1}{2}\mathsf{d}^2(x_0, \cdot)$ is smooth.

Definition 1.88. For any $x \in \Sigma_{x_0}$ and $t \in [0,1]$, the *sub-Riemannian geodesic homothety of center x_0 at time t* is the map $\phi_t : \Sigma_{x_0} \to M$ that associates x with the point at time t of the unique geodesic connecting x_0 with x.

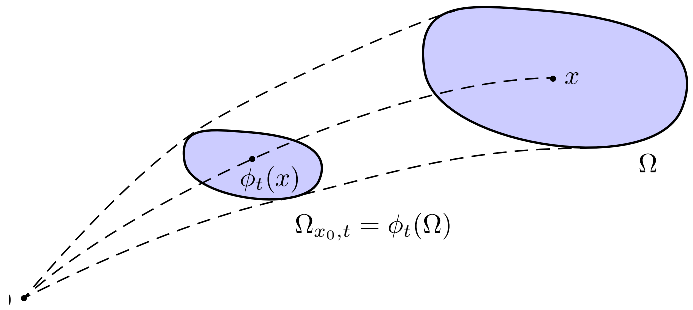


Figure 1.4: Sub-Riemannian homothety of the set Ω with center x_0.

As a consequence of Theorem 1.68 and the smooth dependence on initial data, it is easy to prove that $(t,x) \mapsto \phi_t(x)$ is smooth on $[0,1] \times \Sigma_{x_0}$, and is given by the explicit formula

$$\phi_t(x) = \pi \circ e^{(t-1)\vec{H}}(d_x\mathfrak{f}). \tag{1.45}$$

Let now $\Omega \subset \Sigma_{x_0}$ be a bounded, measurable set, with $0 < \mu(\Omega) < +\infty$, and let $\Omega_{x_0,t} := \phi_t(\Omega)$. The map $t \mapsto \mu(\Omega_{x_0,t})$ is smooth on $[0,1]$. As shown in Fig. 1.4, the homothety shrinks Ω to the center x_0. Indeed $\Omega_{x_0,0} = \{x_0\}$, and $\mu(\Omega_{x_0,t}) \to 0$ for $t \to 0$. For a Riemannian structure, a standard computation in terms of Jacobi fields shows that

$$\mu(\Omega_{x_0,t}) \sim t^{\dim M}, \qquad \text{for } t \to 0, \tag{1.46}$$

where we write $f(t) \sim g(t)$ if there exists $C \neq 0$ such that $f(t) = g(t)(C + o(1))$.

In the sub-Riemannian case, we have a similar power-law behaviour, but the exponent is a different dimensional invariant, which we call *geodesic dimension*. The main result of this section is a formula for the geodesic dimension, in terms of the growth vector of the geodesic.

Definition 1.89. Let $\lambda \in T^*_{x_0}M$. Assume that the corresponding geodesic $\gamma : [0,1] \to M$ is ample (at $t = 0$) of step m, with growth vector $\mathcal{G}_\gamma = \{k_1, k_2, \dots, k_m\}$ (at $t = 0$). Then we define

$$\mathcal{N}_\lambda := \sum_{i=1}^{m} (2i-1)(k_i - k_{i-1}), \tag{1.47}$$

and $\mathcal{N}_\lambda := +\infty$ if the geodesic is not ample.

Observe that Eq. (1.47) closely resembles the formula for Hausdorff dimension of an equiregular sub-Riemannian manifold. In the latter, each direction has a weight according to the flag of the distribution, while in Eq. (1.47), the weights depend on the flag of the geodesic.

Remark 1.90. Assume that λ is associated with an equiregular geodesic γ. Then, by Remark 1.48 and Eq. (1.47) it follows that

$$\mathcal{N}_\lambda = \operatorname{tr} \mathcal{I}_\lambda.$$

Moreover, as a consequence of Theorem 1.C, under these assumption $\mathcal{N}_\lambda$ can be recovered from the sub-Laplacian of $\mathfrak{f}_t$ by the following formula

$$\mathcal{N}_\lambda = \lim_{t \to 0} \Delta_\mu \mathfrak{f}_t\big|_{x_0}.$$

Proposition 1.91. *The function* $\lambda \mapsto \mathcal{N}_\lambda$ *is constant a.e. on* $T^*_{x_0}M$*, assuming its minimum value. Therefore, we define the* geodesic dimension *at* x_0 *as*

$$\mathcal{N}_{x_0} := \min\{\mathcal{N}_\lambda | \, \lambda \in T^*_{x_0}M\} < +\infty.$$

Remark 1.92. For every $x_0 \in M$ we have the inequality $\mathcal{N}_{x_0} \geq \dim M$ and the equality holds if and only if the structure is Riemannian at x_0. Notice that, if the distribution is equiregular at x_0, it follows from Lemma 1.67 and Mitchell's formula for Hausdorff dimension (see [54]) that $\mathcal{N}_{x_0} > \dim_{\mathcal{H}} M$. For genuine sub-Riemannian structures then, the geodesic dimension is a new invariant, related with the structure of the distribution along geodesics.

The geodesic dimension is the exponent of the sub-Riemannian analogue of Eq. (1.46).

Theorem 1.D. *Let* μ *be a smooth volume. For any bounded, measurable set* $\Omega \subset \Sigma_{x_0}$*, with* $0 < \mu(\Omega) < +\infty$ *we have*

$$\mu(\Omega_{x_0,t}) \sim t^{\mathcal{N}_{x_0}}, \qquad \text{for } t \to 0.$$

Observe also that homotheties with different center may have different asymptotic exponents. This can happen, for example, in non-equiregular sub-Riemannian structures.

The proof of Proposition 1.91 and Theorem 1.D is postponed to the end of Section 1.6.

Example 1.93 (Geodesic dimension in contact structures)**.** Let $(M, \mathscr{D}, \langle\cdot|\cdot\rangle)$ be a contact sub-Riemannian structure. In this case, for any $x_0 \in M$, $\dim M = 2\ell + 1$ and $\dim \mathscr{D}_{x_0} = 2\ell$. Any non-trivial geodesic γ is ample with the same growth vector $\mathcal{G}_\gamma = \{2\ell, 2\ell + 1\}$. Therefore, by Eq. (1.47), $\mathcal{N}_{x_0} = 2\ell + 3$ (notice that it does not depend on x_0). Theorem 1.D is an asymptotic generalization of the results obtained in [45], where the exponent $2\ell+3$ appears in the context of measure contraction property in the Heisenberg group. For a more recent overview on measure contraction property in Carnot groups, see [64].

1.5.7 Heisenberg group

Before entering into details of the proofs, we repeat the construction introduced in the previous sections for one of the simplest sub-Riemannian structures: the Heisenberg group. We provide an explicit expression for the geodesic cost function and, applying Definition 1.46, we obtain a formula for the operators $\mathcal{I}_\lambda$ and $\mathcal{R}_\lambda$. In particular, we recover by a direct computation the results of Theorems 1.A, 1.B and 1.C.

The Heisenberg group $\mathbb{H}$ is the equiregular sub-Riemannian structure on $\mathbb{R}^3$ defined by the global (orthonormal) frame

$$X = \partial_x - \frac{y}{2}\partial_z, \qquad Y = \partial_y + \frac{x}{2}\partial_z. \tag{1.48}$$

Notice that the distribution is bracket-generating, for $Z := [X, Y] = \partial_z$. Let us introduce the linear on fibers functions $h_x, h_y, h_z : T^*\mathbb{R}^3 \to \mathbb{R}$

$$h_x := p_x - \frac{y}{2}p_z, \qquad h_y := p_y + \frac{x}{2}p_z, \qquad h_z := p_z,$$

where (x, y, z, p_x, p_y, p_z) are canonical coordinates on $T^*\mathbb{R}^3$ induced by coordinates (x, y, z) on $\mathbb{R}^3$. Notice that h_x, h_y, h_z are the linear on fibers functions associated with the fields X, Y, Z, respectively (i.e. $h_x(\lambda) = \langle \lambda, X\rangle$, and analogously for h_y, h_z).

The sub-Riemannian Hamiltonian is $H = \frac{1}{2}(h_x^2 + h_y^2)$ and the coordinates (x, y, z, h_x, h_y, h_z) define a global chart for T^*M. It is useful to introduce the identification $\mathbb{R}^3 = \mathbb{C}\times\mathbb{R}$, by defining the complex variable $w := x + iy$ and the complex "momentum" $h_w := h_x + ih_y$. Let $q = (w, z)$ and $q' = (w', z')$ be two points in $\mathbb{H}$. The Heisenberg group law, in complex coordinates, is given by

$$q \cdot q' = \left(w + w', z + z' - \frac{1}{2}\Im\left(w\overline{w'}\right)\right). \tag{1.49}$$

Observe that the frame (1.48) is left-invariant for the group action defined by Eq. (1.49). Notice also that h_z is constant along any geodesic due to the identity $[X, Z] = [Y, Z] = 0$.

The geodesic $\gamma(t) = (w(t), z(t))$ starting from $(w_0, z_0) \in \mathbb{H}$ and corresponding to the initial covector $(h_{w,0}, h_z)$, with $h_z \neq 0$ is given by

$$w(t) = w_0 + \frac{h_{w,0}}{ih_z}\left(e^{ih_z t} - 1\right),$$
$$z(t) = z_0 + \frac{1}{2}\int_0^t \Im(\overline{w}dw).$$

In the following, we assume that the geodesic is parametrized by arc length, i.e. $|h_{w,0}|^2 = 1$. We fix $h_{w,0} = ie^{i\phi}$, i.e. ϕ parametrizes the (unit) velocity of the geodesic $\dot\gamma(0) = -\sin\phi X + \cos\phi Y$. Finally, the geodesics corresponding to covectors with $h_z = 0$ are straight lines

$$w(t) = w_0 + h_{w,0}t,$$
$$z(t) = z_0 + \frac{1}{2}\Im(h_{w,0}\overline{w_0})t.$$

In the following, we employ both real (x, y, z, h_x, h_y, h_z) and complex (w, z, h_w, h_z) coordinates when convenient.

Distance in the Heisenberg group

Let $\mathsf{d}_0 = \mathsf{d}(0, \cdot) : \mathbb{H} \to \mathbb{R}$ be the sub-Riemannian distance from the origin and introduce cylindrical coordinates (r, φ, z) on $\mathbb{H}$ defined by $x = r\cos\varphi$, $y = r\sin\varphi$. In order to write an explicit formula for d recall that

(i) $\mathsf{d}_0^2(r, \varphi, z)$ does not depend on φ.

(ii) $\mathsf{d}_0^2(\alpha r, \varphi, \alpha^2 z) = \alpha^2\mathsf{d}_0^2(r, \varphi, z)$, where $\alpha > 0$.

Then, for $r \neq 0$, one has

$$\mathsf{d}_0^2(r, \varphi, z) = r^2\mathsf{d}_0^2\left(1, 0, \frac{z}{r^2}\right). \tag{1.50}$$

It is then sufficient to compute the squared distance of the point $q = (1, 0, \xi)$ from the origin.

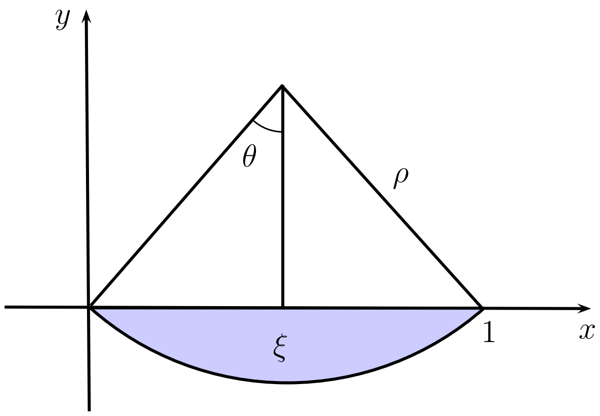


Figure 1.5: Projection of the geodesic joining the origin with $(1, 0, \xi)$ in $\mathbb{H}$.

Consider the minimizing geodesic joining the origin with the point $(1, 0, \xi)$. Its projection on the xy-plane is an arc of circle with radius ρ, connecting the origin with the point $(1, 0)$. In what follows we refer to notation of Fig. 1.5.

The highlighted circle segment has area equal to ξ. Observe that $\theta \in (-\pi, \pi)$, with $\theta = 0$ corresponding to $\xi = 0$ and $\theta \to \pm\pi$ corresponding to $\xi \to \pm\infty$. Then

$$\xi = \theta\rho^2 - \frac{\rho\cos\theta}{2}.$$

Since $2\rho\sin\theta = 1$, we obtain the following equation

$$4\xi = \frac{\theta}{\sin^2\theta} - \cot\theta. \tag{1.51}$$

The right hand side of Eq. (1.51) is a smooth and strictly monotone function of θ, for $\theta \in (-\pi, \pi)$. Therefore the function $\theta : \xi \mapsto \theta(\xi)$ is well defined and smooth. Moreover θ is an odd function and, by Eq. (1.51), it satisfies the following differential equation

$$\frac{d}{d\xi}\left(\frac{\theta^2}{\sin^2\theta}\right) = 4\theta.$$

Finally, the squared distance from the origin of the point $(1, 0, \xi)$ is the Euclidean squared length of the arc, i.e.

$$\mathsf{d}_0^2(1, 0, \xi) = \frac{\theta^2(\xi)}{\sin^2\theta(\xi)}. \tag{1.52}$$

Plugging Eq. (1.52) in Eq. (1.50), we obtain the formula for the squared distance:

$$\mathsf{d}_0^2(r, \phi, z) = r^2\frac{\theta^2(z/r^2)}{\sin^2\theta(z/r^2)}. \tag{1.53}$$

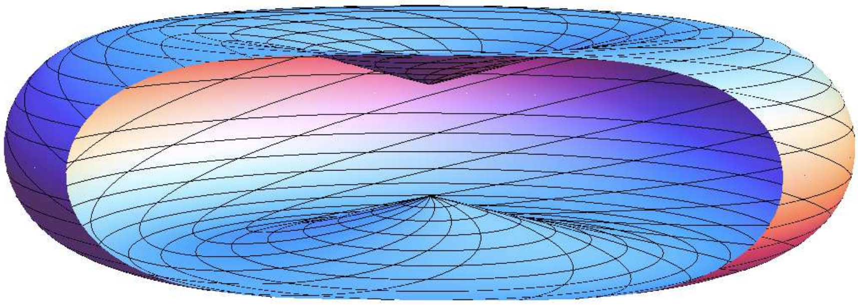


Figure 1.6: A picture of the sub-Riemannian sphere defined by $\mathsf{d}_0 = 1$.

Asymptotic expansion of the distance

Next we investigate, for two given geodesics γ_1, γ_2 in $\mathbb{H}$ starting from the origin and associated with covectors $\lambda_1, \lambda_2 \in T_0^*M$, the regularity of the function

$$C(t,s) := \frac{1}{2}\mathsf{d}^2(\gamma_1(t), \gamma_2(s))$$

in a neighbourhood of $(t,s) = (0,0)$. By left-invariance, one has

$$C(t,s) = \frac{1}{2}\mathsf{d}_0^2(\gamma_1(t)^{-1} \cdot \gamma_2(s)).$$

Let $(W_{t,s}, Z_{t,s})$ be the complex coordinates for the point $\gamma_1(t)^{-1} \cdot \gamma_2(s) \in \mathbb{H}$. Moreover, let $R_{t,s}^2 := |W_{t,s}|^2$, and $\xi_{t,s} := Z_{t,s}/R_{t,s}^2$. Then, by Eq. (1.53),

$$C(t,s) = \frac{1}{2}R_{t,s}^2 \frac{\theta^2(\xi_{t,s})}{\sin^2\theta(\xi_{t,s})}.$$

A long computation, that is sketched in Appendix E, leads to the following result.

Proposition 1.94. *The function $C(t,s)$ is C^1 in a neighbourhood of the origin, but not C^2. In particular, the function $\partial_{ss}C(t,0)$ is not continuous at the origin. However, the singularity at $t=0$ is removable, and the following expansion holds, for $t > 0$*

$$\frac{\partial^2 C}{\partial s^2}(t,0) = 1 + 3\sin^2(\phi_2 - \phi_1) + \frac{1}{2}[2h_{z,2}\sin(\phi_2 - \phi_1) - h_{z,1}\sin(2\phi_2 - 2\phi_1)]t - \\ - \frac{2}{15}h_{z,1}^2\sin^2(\phi_2 - \phi_1)t^2 + O(t^3).$$

If the geodesic γ_2 is chosen to be a straight line (i.e. $h_{z,2} = 0$), then

$$\frac{\partial^2 C}{\partial s^2}(t,0) = 1 + 3\sin^2(\phi_2 - \phi_1) - \frac{h_{z,1}}{2}\sin(2\phi_2 - 2\phi_1)t - \frac{2}{15}h_{z,1}^2\sin^2(\phi_2 - \phi_1)t^2 + O(t^3), \quad (1.54)$$

where $\lambda_j = (-\sin\phi_j, \cos\phi_j, h_{z,j}) \in T_0^*M$ is the initial covector of the geodesic γ_j.

We stress once again that, for a Riemannian structure, the function $C(t,s)$ (which can be defined in a completely analogous way as the squared distance between two Riemannian geodesics) is smooth at the origin.

Second differential of the geodesic cost

We are now ready to compute explicitly the asymptotic expansion of $\mathcal{Q}_\lambda$. Fix $w \in T_{x_0}M$ and let $\alpha(s)$ be any geodesic in $\mathbb{H}$ such that $\dot\alpha(0) = w$. Then we compute the quadratic form $d^2_{x_0}\dot c_t(w)$ for $t > 0$

$$\begin{aligned}\langle \mathcal{Q}_\lambda(t)w|w\rangle = d^2_{x_0}\dot c_t(w) &= \left.\frac{\partial^2}{\partial s^2}\right|_{s=0} \frac{\partial}{\partial t} c_t(\alpha(s)) = \\ &= \left.\frac{\partial^2}{\partial s^2}\right|_{s=0} \frac{\partial}{\partial t}\left(-\frac{1}{2t}\mathsf{d}^2(\gamma(t),\alpha(s))\right) = \frac{\partial}{\partial t}\left(-\frac{1}{t}\frac{\partial^2 C}{\partial s^2}(t,0)\right) = \\ &= \frac{1}{t^2}\left(\lim_{t\to 0^+}\frac{\partial^2 C}{\partial s^2}(t,0)\right) + \frac{1}{3}\left(-\frac{3}{2}\lim_{t\to 0^+}\frac{\partial^4 C}{\partial t^2 \partial s^2}(t,0)\right) + O(t),\end{aligned}$$

where, in the second line, we exchanged the order of derivations by smoothness of $C(t,s)$ for $t > 0$. It is enough to compute the value of $\mathcal{Q}_\lambda(t)$ on an orthonormal basis $v := \dot\gamma(0)$ and $v^\perp := \dot\gamma(0)^\perp$. By using the results of Proposition 1.94, we obtain

$$\langle \mathcal{Q}_\lambda(t)v|v\rangle = \frac{1}{t^2} + O(t), \qquad \langle \mathcal{Q}_\lambda(t)v^\perp|v^\perp\rangle = \frac{4}{t^2} + \frac{2}{15}h_z^2 + O(t).$$

By polarization we obtain $\langle \mathcal{Q}_\lambda(t)v|v^\perp\rangle = O(t)$. Thus the matrices representing the symmetric operators $\mathcal{I}_\lambda$ and $\mathcal{R}_\lambda$ in the basis $\{v, v^\perp\}$ of $\mathscr{D}_{x_0}$ are

$$\mathcal{I}_\lambda = \begin{pmatrix} 1 & 0 \\ 0 & 4 \end{pmatrix}, \qquad \mathcal{R}_\lambda = \frac{2}{5}h_z^2\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}, \tag{1.55}$$

where, we recall, λ has coordinates (h_x, h_y, h_z).

Another way to obtain Eq. (1.55) is to exploit the connection between the curvature operator and the invariants of the Jacobi curves obtained in the proof of Theorem 1.B (see Eqs. (1.80)–(1.81)), in terms of a canonical frame. The latter is not easy to compute, even though, in principle, an algorithmic construction is possible. The interested reader can see [71] for the general case, [49] for contact structures with symmetries, and [18] for a more explicit expression for the 3D contact case and, in particular, the Heisenberg group.

Explicit computations for the curvature of a contact sub-Riemannian structure will appear in a forthcoming paper [8].

Sub-Laplacian of the geodesic cost

By using the results of Proposition 1.94, we explicitly compute the asymptotics of the sub-Laplacian Δ_μ of the function $\mathfrak{f}_t = \frac{1}{2}\mathsf{d}^2(\cdot,\gamma(t))$ at x_0, at the second order in t. In the Heisenberg group, we fix $\mu = dx \wedge dy \wedge dz$ (i.e. the Popp's volume of $\mathbb{H}$), and we suppress the explicit dependence of Δ_μ from the volume form.

Since the sub-Riemannian structure of the Heisenberg group is left-invariant, we can reduce the computation of the asymptotic of $\Delta\mathfrak{f}_t$ to the case of a geodesic γ starting from the origin. Indeed, let us denote by $L_g : \mathbb{H} \to \mathbb{H}$ the left multiplication by $g \in \mathbb{H}$. It is easy to show that if $\gamma(t) = \mathcal{E}_{x_0}(t,\lambda)$ is a geodesic, then $\widetilde{\gamma}(t) := L_g(\gamma(t))$ is a geodesic too. If $\mathfrak{f}_t$ and $\widetilde{\mathfrak{f}}_t$ denote the squared distance along the geodesics γ and $\widetilde{\gamma}$, respectively, we have

$$\widetilde{\mathfrak{f}}_t(L_g(x)) = \frac{1}{2}\mathsf{d}^2(L_g(x), \widetilde{\gamma}(t)) = \frac{1}{2}\mathsf{d}^2(L_g(x), L_g(\gamma(t))) = \frac{1}{2}\mathsf{d}^2(x,\gamma(t)) = \mathfrak{f}_t(x).$$

Moreover, by using Proposition 1.41, and recalling the relation $c_t = -t\mathfrak{f}_t$, it is easy to show that

$$\widetilde{\gamma}(t) = \mathcal{E}_{y_0}(t,\eta), \qquad \text{where} \qquad y_0 = L_g(x_0), \quad \eta = (L_g^*)^{-1}\lambda \in T_{y_0}^*M.$$

Moreover Δ is left-invariant hence $\Delta(f \circ L_g) = \Delta f \circ L_g$ for every $f \in C^\infty(M)$, and we have

$$\Delta\widetilde{\mathfrak{f}}_t|_{y_0} = \Delta\mathfrak{f}_t|_{x_0}.$$

In terms of an orthonormal frame, the sub-Laplacian is $\Delta = X^2 + Y^2$ hence

$$\Delta\mathfrak{f}_t|_{x_0} = \left.\frac{d^2}{ds^2}\right|_{s=0} \mathfrak{f}_t(e^{sX}(x_0)) + \left.\frac{d^2}{ds^2}\right|_{s=0} \mathfrak{f}_t(e^{sY}(x_0)), \tag{1.56}$$

where $e^{sX}(x_0)$ denote the integral curve of the vector field X starting from x_0 (and similarly for Y). Observe that the integral curves of the vector fields X and Y, starting from the origin, are two orthogonal straight lines contained in the xy-plane. Thus we can compute Eq. (1.56) (where $x_0 = 0$) by summing two copies of Eq. (1.54) for $\phi_2 = -\pi/2$ and $\phi_2 = 0$ respectively. By left-invariance we immediately find, for any $x_0 \in \mathbb{H}$

$$\Delta\mathfrak{f}_t|_{x_0} = 5 - \frac{2}{15}h_z^2 t^2 + O(t^3),$$

where, we recall, the initial covector associated with the geodesic γ is $\lambda = (h_x, h_y, h_z) \in T_{x_0}^*M$.

1.6 Jacobi curves

In this section we introduce the notion of Jacobi curve associated with a normal geodesic, that is a curve of Lagrangian subspaces in a symplectic vector space. This curve arises naturally from the geometric interpretation of the second derivative of the geodesic cost, and is closely related with the asymptotic expansion of Theorem 1.A.

We start with a brief description of the properties of curves in the Lagrange Grassmannian. For more details, see [16, 17, 71].

1.6.1 Curves in the Lagrange Grassmannian

Let (Σ, σ) be a $2n$-dimensional symplectic vector space. A subspace $\Lambda \subset \Sigma$ is called *Lagrangian* if it has dimension n and $\sigma|_\Lambda \equiv 0$. The *Lagrange Grassmannian* $L(\Sigma)$ is the set of all n-dimensional Lagrangian subspaces of Σ.

Proposition 1.95. *$L(\Sigma)$ is a compact $n(n+1)/2$-dimensional submanifold of the Grassmannian of n-planes in Σ.*

Proof. Let $\Delta \in L(\Sigma)$, and consider the set $\Delta^{\pitchfork} := \{\Lambda \in L(\Sigma) \,|\, \Lambda \cap \Delta = 0\}$ of all Lagrangian subspaces transversal to Δ. Clearly, the collection of these sets for all $\Delta \in L(\Sigma)$ is an open cover of $L(\Sigma)$. Then it is sufficient to find submanifold coordinates on each $\Delta^{\pitchfork}$.

Let us fix any Lagrangian complement Π of Δ (which always exists, tough it is not unique). Every n-dimensional subspace $\Lambda \subset \Sigma$ that is transversal to Δ is the graph of a linear map from Π to Δ. Choose an adapted Darboux basis on Σ, namely a basis $\{e_i, f_i\}_{i=1}^n$ such that

$$\Delta = \operatorname{span}\{f_1, \dots, f_n\}, \qquad \Pi = \operatorname{span}\{e_1, \dots, e_n\},$$
$$\sigma(e_i, f_j) - \delta_{ij} = \sigma(f_i, f_j) = \sigma(e_i, e_j) = 0, \qquad i, j = 1, \dots, n.$$

In these coordinates, the linear map is represented by a matrix S_Λ such that

$$\Lambda \cap \Delta = 0 \Leftrightarrow \Lambda = \{z = (p, S_\Lambda p),\, p \in \Pi \simeq \mathbb{R}^n\}.$$

Moreover it is easily seen that $\Lambda \in L(\Sigma)$ if and only if $S_\Lambda = S_\Lambda^*$. Hence, the open set $\Delta^{\pitchfork}$ of all Lagrangian subspaces transversal to Δ is parametrized by the set of symmetric matrices, and this gives smooth submanifold coordinates on $\Delta^{\pitchfork}$. This also proves that the dimension of $L(\Sigma)$ is $n(n+1)/2$. Finally, as a closed subset of a compact manifold, $L(\Sigma)$ is compact. □

Fix now $\Lambda \in L(\Sigma)$. The tangent space $T_\Lambda L(\Sigma)$ to the Lagrange Grassmannian at the point Λ can be canonically identified with the set of quadratic forms on the space Λ itself, namely

$$T_\Lambda L(\Sigma) \simeq Q(\Lambda).$$

Indeed, consider a smooth curve $\Lambda(\cdot)$ in $L(\Sigma)$ such that $\Lambda(0) = \Lambda$, and denote by $\dot{\Lambda} \in T_\Lambda L(\Sigma)$ its tangent vector. For any point $z \in \Lambda$ and any smooth extension $z(t) \in \Lambda(t)$, we define the quadratic form

$$\dot{\Lambda} := z \mapsto \sigma(z, \dot{z}),$$

where $\dot{z} := \dot{z}(0)$. A simple check shows that the definition does not depend on the extension $z(t)$. Finally, if in local coordinates $\Lambda(t) = \{(p,\, S(t)p),\, p \in \mathbb{R}^n\}$, the quadratic form $\dot{\Lambda}$ is represented by the matrix $\dot{S}(0)$. In other words, if $z \in \Lambda$ has coordinates $p \in \mathbb{R}^n$, then $\dot{\Lambda} : p \mapsto p^* \dot{S}(0) p$.

Ample, equiregular, monotone curves

Let $J(\cdot) \in L(\Sigma)$ be a smooth curve in the Lagrange Grassmannian. For $i \in \mathbb{N}$, consider

$$J^{(i)}(t) = \operatorname{span}\left\{ \frac{d^j}{dt^j}\ell(t) \,\middle|\, \ell(t) \in J(t),\, \ell(t) \text{ smooth},\, 0 \leq j \leq i \right\} \subset \Sigma, \qquad i \geq 0.$$

Definition 1.96. The subspace $J^{(i)}(t)$ is the *i-th extension* of the curve $J(\cdot)$ at t. The flag

$$J(t) = J^{(0)}(t) \subset J^{(1)}(t) \subset J^{(2)}(t) \subset \ldots \subset \Sigma,$$

is the *associated flag of the curve* at the point t. The curve $J(\cdot)$ is called:

(i) *equiregular* at t if $\dim J^{(i)}(\cdot)$ is locally constant at t, for all $i \in \mathbb{N}$,

(ii) *ample* at t if there exists $N \in \mathbb{N}$ such that $J^{(N)}(t) = \Sigma$,

(iii) *monotone increasing* (resp. *decreasing*) at t if $\dot{J}(t)$ is non-negative definite (resp. non-positive definite) as a quadratic form.

The *step* of the curve at t is the minimal $N \in \mathbb{N}$ such that $J^{(N)}(t) = \Sigma$.

In coordinates, $J(t) = \{(p, S(t)p) | \ p \in \mathbb{R}^n\}$ for some smooth family of symmetric matrices $S(t)$. The curve is ample at t if and only if there exists $N \in \mathbb{N}$ such that

$$\operatorname{rank}\{\dot{S}(t), \ddot{S}(t), \dots, S^{(N)}(t)\} = n.$$

The *rank* of the curve at t is the rank of $\dot{J}(t)$ as a quadratic form (or, equivalently, the rank of $\dot{S}(t)$). We say that the curve is equiregular, ample or monotone (increasing or decreasing) if it is equiregular, ample or monotone for all t in the domain of the curve.

In the subsequent sections we show that with any ample (resp. equiregular) geodesic, we can associate in a natural way an ample (resp. equiregular) curve in an appropriate Lagrange Grassmannian. This justifies the terminology introduced in Definition 1.96.

An important property of ample, monotone curves is described in the following lemma.

Lemma 1.97. *Let $J(\cdot) \in L(\Sigma)$ be a monotone, ample curve at t_0. Then, there exists $\varepsilon > 0$ such that $J(t) \cap J(t_0) = \{0\}$ 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, $J(t) = \{(p, S(t)p) | \ p \in \mathbb{R}^n\}$, 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 τ, $J(\tau) \cap J(0) \neq \{0\}$ (w.l.o.g. assume $\tau > 0$). This means that $\exists p \in \mathbb{R}^n$ such that $S(\tau)p = 0$. Indeed also $p^*S(\tau)p = 0$. The function $t \mapsto p^*S(t)p = 0$ is monotone, vanishing at $t = 0$ and $t = \tau$. Therefore $p^*S(t)p = 0$ for all $0 \leq t \leq \tau$. Being a semidefinite, symmetric matrix, $p^*S(t)p = 0$ if and only if $S(t)p = 0$. Therefore, we conclude that $p \in \ker S(t)$ for $0 \leq t \leq \tau$. This implies that, for any $i \in \mathbb{N}$, $p \in \ker S^{(i)}(0)$, which is a contradiction, since the curve is ample at 0. □

Remark 1.98. Ample curves with $N = 1$ are also called *regular*. See in particular [16,17], where the authors discuss geometric invariants of these curves. Notice that a curve $J(\cdot)$ is regular at t if and only if its tangent vector at t is a non degenerate quadratic form, i.e. the matrix $\dot{S}(t)$ is invertible.

The Young diagram of an equiregular Jacobi curve

Let $J(\cdot) \in L(\Sigma)$ be smooth, ample and equiregular. We can associate in a standard way a Young diagram with the curve $J(\cdot)$ as follows. Consider the restriction of the curve to a neighbourhood of t such that, for all $i \in N$, $\dim J^{(i)}(\cdot)$ is constant. Let $h_i := \dim J^{(i)}(\cdot)$. By hypotesis, there exists a minimal $N \in \mathbb{N}$ such that $h_i = \dim \Sigma$ for all $i \geq N$. It follows from the definition of extension that, for $i \in \mathbb{N}$, we have the inequalities $h_{i+1} - h_i \leq h_i - h_{i-1}$. Then, we build a Young diagram with N columns, with $h_i - h_{i-1}$ boxes in the i-th column. This is the *Young diagram of the curve* $J(\cdot)$. In particular, notice that the number of boxes in the first column is equal to the rank of $J(\cdot)$.

1.6.2 The Jacobi curve and the second differential of the geodesic cost

Recall that T^*M has a natural structure of symplectic manifold, with the canonical symplectic form defined as the differential of the Liouville form, namely $\sigma = d\varsigma$. In particular, for any $\lambda \in T^*M$, $T_\lambda(T^*M)$ is a symplectic vector space with the canonical symplectic form σ. Therefore, we can specify the construction above to $\Sigma := T_\lambda(T^*M)$. In this section we show that the second derivative of the geodesic cost (associated with an ample geodesic γ with initial covector $\lambda \in T^*M$) can be naturally interpreted as a curve in the Lagrange Grassmannian of $T_\lambda(T^*M)$, which is ample in the sense of Definition 1.96.

Second differential at a non critical point

Let $f \in C^\infty(M)$. As we explained in Section 1.4.3, the second differential of f, which is a symmetric bilinear form on the tangent space, is well defined only at critical points of f. If $x \in M$ is not a critical point, it is still possible to define the second differential of f, as the differential of df, thought as a section of T^*M.

Definition 1.99. Let $f \in C^\infty(M)$, and

$$df : M \to T^*M, \qquad df : x \mapsto d_x f.$$

Fix $x \in M$, and let $\lambda := d_x f \in T^*M$. The *second differential* of f at $x \in M$ is the linear map

$$d_x^2 f := d_x(df) : T_xM \to T_\lambda(T^*M), \qquad d_x^2 f : v \mapsto \left.\frac{d}{ds}\right|_{s=0} d_{\gamma(s)} f,$$

where $\gamma(\cdot)$ is a curve on M such that $\gamma(0) = x$ and $\dot\gamma(0) = v$.

Definition 1.99 generalizes the concept of "second derivatives" of f, as the linearisation of the differential.

Remark 1.100. The image of the differential $df : M \to T^*M$ is a Lagrangian submanifold of T^*M. Thus, by definition, the image of the second differential $d_x^2 f(T_xM)$ at a point x is the tangent space of $df(M)$ at $\lambda = d_x f$, which is an n-dimensional Lagrangian subspace of $T_\lambda(T^*M)$ transversal to the vertical subspace $T_\lambda(T_x^*M)$.

By a dimensional argument and the fact that $\pi \circ df = \mathbb{I}_M$ (hence $\pi_* \circ d_x^2 f = \mathbb{I}_{T_xM}$), we obtain the following formula for the image of a subspace through the second differential.

Lemma 1.101. *Let $f : M \to \mathbb{R}$ and $W \subset T_xM$. Then $d_x^2 f(W) = d_x^2 f(T_xM) \cap \pi_*^{-1}(W)$.*

The next lemma describes the affine structure on the space of second differentials.

Lemma 1.102. *Let $\lambda \in T_x^*M$. The set $\mathcal{L}_\lambda := \{d_x^2 f \,|\, f \in C^\infty(M), d_x f = \lambda\}$ is an affine space over the vector space $Q(T_xM)$ of the quadratic forms over T_xM.*

Proof. Consider two functions f_1, f_2 such that $d_x f_1 = d_x f_2 = \lambda$. Then $f_1 - f_2$ has a critical point at x. We define the difference between $d_x^2 f_1$ and $d_x^2 f_2$ as the quadratic form $d_x^2(f_1 - f_2)$. □

Remark 1.103. When $\lambda = 0 \in T_x^*M$, $\mathcal{L}_\lambda$ is the space of the second derivatives of the functions with a critical point at x. In this case we can fix a canonical origin in $\mathcal{L}_\lambda$, namely the second differential of any constant function. This gives the identification of $\mathcal{L}_\lambda$ with the space of quadratic forms on T_xM, recovering the standard notion of Hessian discussed in Section 1.4.3.

Second differential of the geodesic cost

Let $\gamma : [0, T] \to M$ be a strongly normal geodesic. Let $x = \gamma(0)$. Without loss of generality, we can choose T sufficiently small so that the map $(t, x) \to c_t(x)$ is smooth in a neighbourhood of $(0, T) \times \{x\} \subset \mathbb{R} \times M$, and $d_x c_t = \lambda$ is the initial covector associated with γ (see Theorem 1.40 and Proposition 1.41).

The second differential of c_t defines a curve in the Lagrange Grassmannian $L(T_\lambda(T^*M))$. For any $\lambda \in T^*M$, $\pi(\lambda) = x$, we denote with the symbol $\mathcal{V}_\lambda = T_\lambda(T_x^*M) \subset T_\lambda(T^*M)$ the vertical subspace, namely the tangent space to the fiber T_x^*M. Observe that, if $\pi : T^*M \to M$ is the bundle projection, $\mathcal{V}_\lambda = \ker \pi_*$.

Definition 1.104. The *Jacobi curve* associated with γ is the smooth curve $J_\lambda : [0, T] \to L(T_\lambda(T^*M))$ defined by

$$J_\lambda(t) := d_x^2 c_t(T_x M),$$

for $t \in (0, T]$, and $J_\lambda(0) := \mathcal{V}_\lambda$.

The Jacobi curve is smooth as a consequence of the next proposition, which provides an equivalent characterization of the Jacobi curve in terms of the Hamiltonian flow on T^*M.

Proposition 1.105. *Let $\lambda : [0, T] \to T^*M$ be the unique lift of γ such that $\lambda(t) = e^{t\vec{H}}(\lambda)$. Then the associated Jacobi curve satisfies the following properties for all t, s such that both sides of the statements are defined:*

(i) $J_\lambda(t) = e_*^{-t\vec{H}} \mathcal{V}_{\lambda(t)}$,

(ii) $J_\lambda(t + s) = e_*^{-t\vec{H}} J_{\lambda(t)}(s)$,

(iii) $\dot{J}_\lambda(0) = -d_\lambda^2 H_x$ *as quadratic forms on* $\mathcal{V}_\lambda \simeq T_x^*M$.

Proof. In order to prove (i) it is sufficient to show that $\pi_* \circ e_*^{t\vec{H}} \circ d_x^2 c_t = 0$. Then, let $v \in T_x M$, and $\alpha(\cdot)$ a smooth arc such that $\alpha(0) = x$, $\dot{\alpha}(0) = v$. Recall that, for s sufficiently small, $d_{\alpha(s)} c_t$ is the initial covector of the unique normal geodesic which connects $\alpha(s)$ with $\gamma(t)$ in time t, i.e. $\pi \circ e^{t\vec{H}} \circ d_{\alpha(s)} c_t = \gamma(t)$. Then

$$\pi_* \circ e_*^{t\vec{H}} \circ d_x^2 c_t(v) = \left.\frac{d}{ds}\right|_{s=0} \pi \circ e^{t\vec{H}} \circ d_{\alpha(s)} c_t = 0.$$

Statement (ii) follows from (i) and the group property of the Hamiltonian flow. To prove (iii), introduce canonical coordinates (p, x) in the cotangent bundle. Let $\xi \in \mathcal{V}_\lambda$, such that $\xi = \sum_{i=1}^n \xi_i \partial_{p_i}|_\lambda$. By (i), the smooth family of vectors in $\mathcal{V}_\lambda$ defined by

$$\xi(t) := e_*^{-t\vec{H}} \left(\sum_{i=1}^n \xi^i \partial_{p_i}|_{\lambda(t)} \right),$$

satisfies $\xi(0) = \xi$ and $\xi(t) \in J_\lambda(t)$. Therefore

$$\dot{J}_\lambda(0)\xi = \sigma(\xi, \dot{\xi}) = -\sum_{i,j=1}^n \frac{\partial^2 H}{\partial p_i \partial p_j} \xi^i \xi^j = -\langle \xi, (d_\lambda^2 H_x)\xi \rangle,$$

where the last equality follows from the definition of $d^2_\lambda H_x$ after the identification $\mathcal{V}_\lambda \simeq T^*_x M$ (see Section 1.4.2). □

Remark 1.106. Point (i) of Proposition 1.105 can be used to associate a Jacobi curve with any integral curve of the Hamiltonian flow, without any further assumptions on the underlying trajectory on the manifold. In particular we associate with any initial covector $\lambda \in T_x M$ the Jacobi curve $J_\lambda(t) := e^{-t\vec{H}} \mathcal{V}_{\lambda(t)}$. Observe that, in general, $\gamma(\cdot) := \pi \circ \lambda(\cdot)$ may be also abnormal.

Proposition 1.105 and the fact that the quadratic form $d^2_\lambda H_x$ is non-negative definite imply the next corollary.

Corollary 1.107. *The Jacobi curve J_λ is monotone nonincreasing for every $\lambda \in T^*M$.*

The following proposition provides the connection between the flag of a normal geodesic and the flag of the associated Jacobi curve.

Proposition 1.108. *Let $\gamma(t) = \pi \circ e^{t\vec{H}}(\lambda)$ be a normal geodesic associated with the initial covector λ. The flag of the Jacobi curve J_λ projects to the flag of the geodesic γ at $t = 0$, namely*

$$\pi_* J^{(i)}_\lambda(0) = \mathscr{F}^i_\gamma(0), \qquad \forall\, i \in \mathbb{N}. \tag{1.57}$$

Moreover, $\dim J^{(i)}_\lambda(t) = n + \dim \mathscr{F}^i_\gamma(t)$*. Therefore γ is ample of step m (resp. equiregular) if and only if J_λ is ample of step m (resp. equiregular).*

Proof. The last statement follows directly from Eq. (1.57), Proposition 1.105 (point (ii)) and the definition of $\mathscr{F}_{\gamma(s)}(t) = (P_{s,s+t})^{-1}_* \mathscr{D}_{\gamma(s+t)}$. In order to prove Eq. (1.57), let $\bar{u} : T^*M \to L^\infty([0,T], \mathbb{R}^k)$ be the map that associates to any covector the corresponding normal control:

$$\bar{u}_i(\lambda)(\cdot) = \langle e^{\cdot \vec{H}}(\lambda), f_i \rangle, \qquad i = 1, \dots, k,$$

where we assume, without loss of generality, that the Hamiltonian field $\vec{H}$ is complete. For any control $v \in L^\infty([0,T], \mathbb{R}^k)$ and initial point $x \in M$, consider the non-autonomous flow $P^v_{0,t}(x)$. We have the following identity, for any $\lambda \in T^*M$ and $t \in [0,T]$

$$\pi \circ e^{t\vec{H}}(\lambda) = P^{\bar{u}(\lambda)}_{0,t}(\pi(\lambda)).$$

Remember that, as a function of the control, $P^v_{0,t}(x) = E_{x,t}(v)$ (i.e. the endpoint map with basepoint x and endtime t). Therefore, by taking the differential at λ (such that $\pi(\lambda) = x$), we obtain

$$\pi_* \circ e^{t\vec{H}}_* |_\lambda = \left(P^{\bar{u}(\lambda)}_{0,t}\right)_* \circ \pi_*|_\lambda + D_{\bar{u}(\lambda)} E_{x,t} \circ \bar{u}_*|_\lambda,$$

Then, by the explicit formula for the differential of the endpoint map, we obtain, for any vertical field $\xi(t) \in \mathcal{V}_{e^{t\vec{H}}(\lambda)}$

$$\pi_* \circ e^{-t\vec{H}}_* \xi(t) = -\int_0^t (P_{0,\tau})^{-1}_* \overline{f}(v(t,\tau), \gamma(t)) d\tau,$$

where $\gamma(t) = \pi \circ e^{t\vec{H}}(\lambda)$ is the normal geodesic with initial covector λ and, for any $t \in [0,T]$,

$$v_i(t,\cdot) := \bar{u}_* \circ e^{-t\vec{H}}_* \xi(t) = \left(\bar{u} \circ e^{-t\vec{H}}\right)_* \xi(t), \qquad v(t,\cdot) \in L^\infty([0,T], \mathbb{R}^k).$$

More precisely, $v(t,\cdot)$ has components

$$v_i(t,\tau) = \left.\frac{d}{d\varepsilon}\right|_{\varepsilon=0} \langle e^{(\tau-t)\vec{H}}(\lambda(t)+\varepsilon\xi(t)), f_i\rangle, \qquad i=1,\dots,k,$$

where $\lambda(t) = e^{t\vec{H}}(\lambda)$, and we identified $\mathcal{V}_{e^{t\vec{H}}(\lambda)} \simeq T^*_{\gamma(t)}M$. Observe that, on the diagonal, $v_i(t,t) = \langle \xi(t), f_i\rangle = \xi_i(t)$. It is now easy to show that, for any positive $i \in \mathbb{N}$

$$\left.\frac{d^i}{dt^i}\right|_{t=0} \pi_* \circ e_*^{-t\vec{H}}\xi(t) = -\left.\frac{d^{i-1}}{dt^{i-1}}\right|_{t=0} \left[(P_{0,t})_*^{-1}\sum_{j=1}^k \xi_j(t)\overline{f}_j(\gamma(t))\right] \mod \mathscr{F}^{i-1}_\gamma(0). \tag{1.58}$$

By point (i) of Proposition 1.105, any smooth family $\ell(t) \in J_\lambda(t)$ is of the form $e_*^{-t\vec{H}}\xi(t)$ for some smooth $\xi(t) \in \mathcal{V}_{e^{t\vec{H}}(\lambda)}$. Therefore, Eq. (1.58) for $i=1$ implies that $J^{(1)}_\lambda = \mathscr{F}^1_\gamma(0)$. The same equation and an easy induction argument, together with the definitions of the flags show that $J^{(i)}_\lambda(0) = \mathscr{F}^i_\gamma(0)$ for any positive $i \in \mathbb{N}$. □

Remark 1.109. Observe that, if γ is equiregular, ample of step m with growth vector $\mathcal{G}_\lambda = (k_1, k_2, \dots, k_m)$, the Young diagram of J_λ has m columns, with $d_i := k_i - k_{i-1}$ boxes in the i-th column (recall that $k_0 = \dim \mathscr{F}^0_\gamma(t) = 0$).

Remark 1.110. We already observed that the ampleness condition is a rank condition on the derivatives of the symmetric matrix $S_\lambda(t)$ that represents the Jacobi curve $J_\lambda(t)$. In particular the curve is ample at $t=0$ if and only if there exists $N \in \mathbb{N}$ such that

$$\operatorname{rank}\{\dot{S}_\lambda(0), \ddot{S}_\lambda(0), \dots, S^{(N)}_\lambda(0)\} = n.$$

By point (i) of Proposition 1.105 it follows that, for any fibre-wise polynomial Hamiltonian, $S^{(i)}_\lambda(0)$ is a polynomial function of the initial covector $\lambda \in T^*_xM$, for any $i \in \mathbb{N}$. Therefore, under this assumption (which is true, for example, in the sub-Riemannian case), $J_\lambda(\cdot)$ is ample on an open Zariski subset of the fibre T^*_xM.

1.6.3 The Jacobi curve and the Hamiltonian inner product

The following is an elementary, albeit very useful property of the symplectic form σ.

Lemma 1.111. *Let $\xi \in \mathcal{V}_\lambda$ a vertical vector. Then, for any $\eta \in T_\lambda(T^*M)$*

$$\sigma(\xi,\eta) = \langle \xi, \pi_*\eta\rangle,$$

where we employed the canonical identification $\mathcal{V}_\lambda = T^*_xM$.

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. □

In Section 1.4.2 we introduced the Hamiltonian inner product on $\mathscr{D}_x$, which, in general, depends on λ. Such an inner product is defined by the quadratic form $d^2_\lambda H_x : T^*_xM \to T_xM$ on $\mathscr{D}_x = \operatorname{Im}(d^2_\lambda H_x)$. The following lemma allows the practical computation of the Hamiltonian inner product through the Jacobi curve.

Lemma 1.112. *Let* $\xi \in T_x^*M$*. Then*

$$d_\lambda^2 H_x(\xi) = -\pi_* \dot{\xi},$$

where $\dot{\xi}$ *is the derivative, at* $t = 0$*, of any extension* $\xi(t)$ *of* ξ *such that* $\xi(0) = \xi$ *and* $\xi(t) \in J_\lambda(t)$*.*

Proof. By point (iii) of Proposition 1.105, $d_\lambda^2 H_x = -\dot{J}_\lambda(0)$. By definition of $\dot{J}_\lambda(0) : \mathcal{V}_\lambda \to \mathbb{R}$ as a quadratic form, $\dot{J}_\lambda(0)(\xi) = \sigma(\xi, \dot{\xi})$. Then, by Lemma 1.111, $\dot{J}_\lambda(0)(\xi) = \langle \xi, \pi_* \dot{\xi} \rangle$. This implies the statement after identifying again the quadratic form with the associated symmetric map. □

By Lemma 1.112, for any $v \in \mathscr{D}_x$ there exists a $\xi \in \mathcal{V}_\lambda$ such that, for any extension $\xi(t) \in J_\lambda(t)$, with $\xi(0) = \xi$, we have $v = \pi_* \dot{\xi}$. Indeed ξ may not be unique. Besides, if $v = \pi_* \dot{\xi}$ and $w = \pi_* \dot{\eta}$, the Hamiltonian inner product rewrites

$$\langle v | w \rangle_\lambda = \sigma(\xi, \dot{\eta}) = -\sigma(\eta, \dot{\xi}). \tag{1.59}$$

We now have all the tools required for the proof of Theorem 1.A.

1.6.4 Proof of Theorem 1.A

The statement of Theorem 1.A is related with the analytic properties of the functions $t \mapsto \langle \mathcal{Q}_\lambda(t) v | v \rangle_\lambda$ for $v \in \mathscr{D}_x$. By definition, $\langle \mathcal{Q}_\lambda(t) v | v \rangle_\lambda = d_x^2 \dot{c}_t(v)$.

As a first step, we compute a coordinate formula for such a function in terms of a splitting $\Sigma = \mathcal{V}_\lambda \oplus \mathcal{H}_\lambda$, where $\mathcal{V}_\lambda$ is the vertical space and $\mathcal{H}_\lambda$ is any Lagrangian complement. Observe that $\mathcal{V}_\lambda = J_\lambda(0) = \ker \pi_*$ and π_* induces an isomorphism between $\mathcal{H}_\lambda$ and T_xM. $J_\lambda(t)$ is the graph of a linear map $S(t) : \mathcal{V}_\lambda \to \mathcal{H}_\lambda$. Equivalently, by Lemma 1.97, for $0 < t < \varepsilon$, $J_\lambda(t)$ is the graph of $S(t)^{-1} : \mathcal{H}_\lambda \to \mathcal{V}_\lambda$. Once a Darboux basis (adapted to the splitting) is fixed, as usual one can identify these maps with the representative matrices.

Fix $v \in \mathscr{D}_x \subset T_xM$ and let $\widetilde{v} \in \mathcal{H}_\lambda$ be the unique horizontal lift such that $\pi_* \widetilde{v} = v$. Then, by definition of Jacobi curve, and the standard identification $\mathcal{V}_\lambda \simeq T_x^*M$

$$\langle \mathcal{Q}_\lambda(t) v | v \rangle_\lambda = \frac{d}{dt} \sigma(S(t)^{-1} \widetilde{v}, \widetilde{v}). \tag{1.60}$$

Since $J_\lambda(0) = \mathcal{V}_\lambda$, it follows that $S(t)^{-1}$ is singular at $t = 0$. In what follows we prove Theorem 1.A, by computing the asymptotic expansion of the matrix $S(t)^{-1}$. More precisely, from (1.60) it is clear that we need only a "block" of $S(t)^{-1}$ since it acts only on vectors $\widetilde{v} \in \pi_*^{-1}(\mathscr{D}_x) \cap \mathcal{H}_\lambda$. In what follows we build natural coordinates on the space Σ in such a way that Eq. (1.60) is given by the derivative of the first $k \times k$ block of $S(t)^{-1}$ where, we recall, $k = \dim \mathscr{D}_x$. Notice that this restriction is crucial in the proof since only the aforementioned block has a simple pole. This is not true, in general, for the whole matrix $S(t)^{-1}$.

Coordinate presentation of the Jacobi curve

In order to obtain a convenient expression for the matrix $S(t)$ we introduce a set of coordinates (p,x) induced by a particular Darboux frame adapted to the splitting $\Sigma = \mathcal{V}_\lambda \oplus \mathcal{H}_\lambda$. Namely

$$\Sigma = \{(p,x) | \, p,x \in \mathbb{R}^n\}, \qquad \mathcal{V}_\lambda = \{(p,0) | \, p \in \mathbb{R}^n\}, \qquad \mathcal{H}_\lambda = \{(0,x) | \, x \in \mathbb{R}^n\}.$$

Besides, if $\xi = (p,x)$, $\bar{\xi} = (\bar{p},\bar{x}) \in \Sigma$ the symplectic product is $\sigma(\xi,\bar{\xi}) = p^*\bar{x} - \bar{p}^*x$. In these coordinates, $J_\lambda(t) = \{(p, S(t)p) | \, p \in \mathbb{R}^n\}$, and $S(0) = 0$. The symmetric matrix $S(t)$ represents a monotone Jacobi curve, hence $\dot{S}(t) \leq 0$. Moreover, since the curve is ample, by Lemma 1.97, $S(t) < 0$ for $0 < t < \varepsilon$. Moreover we introduce the coordinate splitting $\mathbb{R}^n = \mathbb{R}^k \oplus \mathbb{R}^{n-k}$ (accordingly we write $p = (p_1, p_2)$ and $x = (x_1, x_2)$), such that $\pi_*(\mathbb{R}^k) = \mathscr{D}_x$. In blocks notation

$$S(t) = \begin{pmatrix} S_{11}(t) & S_{12}(t) \\ S_{12}^*(t) & S_{22}(t) \end{pmatrix}, \qquad \text{with} \quad S_{11}(t), S_{22}(t) < 0 \quad \text{for} \quad 0 < t < \varepsilon.$$

By point (iii) of Proposition 1.105, in these coordinates we also have

$$\dot{S}(0) = \begin{pmatrix} \dot{S}_{11}(0) & 0 \\ 0 & 0 \end{pmatrix}, \qquad \text{with} \quad \operatorname{rank} \dot{S}_{11}(0) = \dim \mathscr{D}_x.$$

Therefore, we obtain the following coordinate formula for the Hamiltonian inner product. Let $v, w \in \mathscr{D}_x$, with coordinates $v = (v_1, 0)$, $w = (w_1, 0)$ then

$$\langle v | w \rangle_\lambda = -v_1^* \dot{S}_{11}(0)^{-1} w_1, \qquad v_1, w_1 \in \mathbb{R}^k,$$

Remark 1.113. In other words, the quadratic form associated with the operator $\mathbb{I} : \mathscr{D}_x \to \mathscr{D}_x$ via the Hamiltonian inner product is represented by the matrix $-\dot{S}_{11}(0)^{-1}$.

Moreover the horizontal lift of v is $\widetilde{v} = ((0,0),(v_1,0))$ and analogously for w. Thus, by (1.60)

$$\langle \mathcal{Q}_\lambda(t) v | w \rangle_\lambda = \frac{d}{dt} v_1^* [S(t)^{-1}]_{11} w_1, \qquad v_1, w_1 \in \mathbb{R}^k, \quad t > 0. \tag{1.61}$$

For convenience, for $t > 0$, we introduce the smooth family of $k \times k$ matrices $S^\flat(t)$ defined by

$$S^\flat(t)^{-1} := [S(t)^{-1}]_{11}, \qquad t > 0.$$

Then, the quadratic form associated with the operator $\mathcal{Q}_\lambda(t) : \mathscr{D}_x \to \mathscr{D}_x$ via the Hamiltonian inner is represented by the matrix $\frac{d}{dt} S^\flat(t)^{-1}$.

The proof of Theorem 1.A is based upon the following result.

Theorem 1.114. *The map $t \mapsto S^\flat(t)^{-1}$ has a simple pole at $t = 0$.*

Proof. The expression of $S^\flat(t)$ in terms of the blocks of $S(t)$ is given by the following lemma.

Lemma 1.115. *Let $A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$ be a sign definite matrix, and denote by $[A^{-1}]_{11}$ the first block of the inverse of A. Then $[A^{-1}]_{11} = (A_{11} - A_{12} A_{22}^{-1} A_{21})^{-1}$.*

Then, by definition of $S^\flat$, we have the following formula (where we suppress t):

$$S^\flat = S_{11} - S_{12}S_{22}^{-1}S_{12}^*. \tag{1.62}$$

Lemma 1.116. *As quadratic forms on $\mathbb{R}^k$, $S_{11}(t) \leq S^\flat(t) < 0$ for $t > 0$.*

Proof of Lemma 1.116. Let $t > 0$. $S(t)$ is symmetric and negative, then also its inverse $S(t)^{-1}$ is symmetric and negative. This implies that $S^\flat(t)^{-1} = [S(t)^{-1}]_{11} < 0$ and so is $S^\flat(t)$. This proves the right inequality. By Eq. (1.62) and the fact that $S_{22}(t)$ is negative definite (and so is $S_{22}^{-1}(t)$) one also gets (we suppress $t > 0$)

$$p_1^*(S_{11} - S^\flat)p_1 = p_1^*S_{12}S_{22}^{-1}S_{12}^*p_1 = (S_{12}^*p_1)^*S_{22}^{-1}(S_{12}^*p_1) \leq 0, \qquad p_1 \in \mathbb{R}^k.$$

□

Lemma 1.117. *The map $t \mapsto S^\flat(t)$ can be extended by smoothness at $t = 0$.*

Proof. Indeed, by the coordinate expression of Eq. (1.62), it follows that the only term that can give rise to singularities is the inverse matrix $S_{22}^{-1}(t)$. Since, by assumption, the curve is ample, $t \mapsto \det S_{22}(t)$ has a finite order zero at $t = 0$, thus the singularity can be only a finite order pole. On the other hand $S(t) \to 0$ for $t \to 0$, thus $S_{11}(t) \to 0$ as well. Then, by Lemma 1.116, $S^\flat(t) \to 0$ for $t \to 0$, hence can be extended by smoothness at $t = 0$. □

We are now ready to prove that $t \mapsto S^\flat(t)^{-1}$ has a simple pole at $t = 0$. As a byproduct, we obtain an explicit form for its residue. As usual, for $i > 0$, we set $k_i := \dim J_\lambda^{(i)}(0) - n$, and $d_i := k_i - k_{i-1}$. In coordinates, this means that

$$\operatorname{rank}\{\dot{S}(0), \dots, S^{(i)}(0)\} = k_i, \qquad i = 1, \dots, m.$$

By hypotesis, the curve is ample at $t = 0$, then there exists m such that $k_m = n$. Since we are only interested in Taylor expansions, we may assume $S(t)$ to be real-analytic in $[0, \varepsilon]$ by replacing, if necessary, $S(t)$ with its Taylor polynomial of sufficient high order. Then, let us consider the analytic family of symmetric matrices $\dot{S}(t)$. For $i = 1, \dots, n$, the family $w_i(t)$ of eigenvectors of $\dot{S}(t)$ (and the relative eigenvalues) are an analytic family (see [47, Theorem 6.1, Chapter II]). Therefore, $\dot{S}(t) = W(t)D(t)W(t)^*$, where $W(t)$ is the $n \times n$ matrix whose columns are the vectors $w_i(t)$, and $D(t)$ is a diagonal matrix. Recall that $\dot{S}(t)$ is non-positive definite. Then $\dot{S}(t) = -V(t)V(t)^*$, for some analytic family of $n \times n$ matrices $V(t)$. Let $v_i(t)$ denote the columns of $V(t)$.

Now, let us consider the flag $E_1 \subset E_2 \subset \ldots \subset E_m = \mathbb{R}^n$ defined as follows

$$E_i = \operatorname{span}\{v_j^{(\ell)}(0),\ 1 \leq j \leq n,\ 0 \leq \ell \leq i-1\}.$$

Let $\operatorname{span}\{A\}$ denote the column space of a matrix A. Indeed $\operatorname{span}\{\dot{S}(t)\} \subseteq \operatorname{span}\{V(t)\}$. Besides, $\operatorname{rank}\{\dot{S}(t)\} = \operatorname{rank}\{V(t)V(t)^*\} = \operatorname{rank}\{V(t)\} = \dim \operatorname{span}\{V(t)\}$. Therefore, $\operatorname{span}\{\dot{S}(t)\} = \operatorname{span}\{V(t)\}$, for all $|t| < \varepsilon$. Thus, for $i = 1, \dots, m$

$$E_i = \operatorname{span}\{V(0), V^{(1)}(0), \dots, V^{(i-1)}(0)\} = \operatorname{span}\{\dot{S}(0), \dots, S^{(i)}(0)\}.$$

Therefore $\dim E_i = k_i$. Choose coordinates in $\mathbb{R}^n$ adapted to this flag, i.e. $\text{span}\{e_1, \dots, e_{k_i}\} = E_i$. In these coordinates, $V(t)$ has a peculiar structure, namely

$$V(t) = \begin{pmatrix} \widehat{v}_1 \\ t\widehat{v}_2 \\ \vdots \\ t^{m-1}\widehat{v}_m \end{pmatrix} + \begin{pmatrix} O(t) \\ O(t^2) \\ \vdots \\ O(t^m) \end{pmatrix},$$

where $\widehat{v}_i$ is a $d_i \times n$ matrix of maximal rank (notice that the $\widehat{v}_i$ are not directly related with the columns $v_i(t)$ of $V(t)$). Let $\widehat{V}(t)$ denote the "principal part" of $V(t)$. In other words, $\widehat{V}(t) = (\widehat{v}_1, t\widehat{v}_2, \dots, t^{m-1}\widehat{v}_m)^*$. Then, remember that $S(0) = 0$ and

$$S(t) = \int_0^t \dot{S}(\tau)d\tau = -\int_0^t V(\tau)V(\tau)^* d\tau = -\int_0^t \widehat{V}(\tau)\widehat{V}(\tau)^* d\tau + r(t),$$

where $r(t)$ is a remainder term. Observe that the matrix

$$\widehat{S}(t) = -\int_0^t \widehat{V}(\tau)\widehat{V}(\tau)^* d\tau$$

is negative definite for $t > 0$. In fact, a non trivial kernel for some $t > 0$ would contradict the hypotesis $\text{span}\{V(0), V^{(1)}(0), \dots, V^{(m-1)}(0)\} = \mathbb{R}^n$. In components, we write $S(t)$ as a $m \times m$ block matrix, $S_{ij}(t)$ being a $d_i \times d_j$ block, as follows:

$$S_{ij}(t) = \int_0^t \dot{S}_{ij}(\tau)d\tau = -\left(\frac{\widehat{v}_i\widehat{v}_j^*}{i+j-1}\right) t^{i+j-1} + O(t^{i+j}) = \chi_{ij}t^{i+j-1} + O(t^{i+j}),$$

where we introduced the negative definite constant matrix $\chi := \widehat{S}(1) < 0$. By computing the determinant of $\widehat{S}(t)$, we obtain

$$\det \widehat{S}(t) = \det \begin{pmatrix} t\chi_{11} & t^2\chi_{12} & \cdots & t^m\chi_{1m} \\ t^2\chi_{21} & t^3\chi_{22} & \cdots & t^{m+1}\chi_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ t^m\chi_{m1} & t^{m+1}\chi_{m2} & \cdots & t^{2m-1}\chi_{mm} \end{pmatrix} = t^{d_1+3d_2+\ldots+(2m-1)d_m} \det \chi. \tag{1.63}$$

We now compute the inverse of $S(t)$. First, the inverse of the principal part $\widehat{S}(t)$ is

$$\widehat{S}(t)^{-1}_{ij} = \frac{(\chi^{-1})_{ij}}{t^{i+j-1}},$$

as we readily check:

$$\sum_{\ell=1}^m \widehat{S}(t)^{-1}_{i\ell}\widehat{S}(t)_{\ell j} = \sum_{\ell=1}^m (\chi^{-1})_{i\ell}\chi_{\ell j}\frac{t^{\ell+j-1}}{t^{i+\ell-1}} = \sum_{\ell=1}^m (\chi^{-1})_{i\ell}\chi_{\ell j}t^{j-i} = \delta_{ij}.$$

The (block-wise) principal part of the inverse $S(t)^{-1}$ is equal to the inverse of the (block-wise) principal part of $S(t)$. Then we obtain, in blocks notation, for $i = 1, \dots, m$

$$[S(t)^{-1}]_{ij} = \frac{(\chi^{-1})_{ij}}{t^{i+j-1}} + O\left(\frac{1}{t^{i+j-2}}\right).$$

Finally, by definition, $(S^\flat)^{-1} = [S^{-1}]_{11}$. Thus

$$S^\flat(t)^{-1} = \frac{(\chi^{-1})_{11}}{t} + O(1).$$

Therefore the reduced curve has a simple pole at $t = 0$, with a negative definite residue, and the proof is complete. □

Remark 1.118. Notice that, as a consequence of Eq. (1.63), the order of $\det S(t)$ at $t = 0$ is equal to the order of its principal part $\widehat{S}(t)$. Namely

$$\det S(t) \sim \det \widehat{S}(t) \sim t^{\mathcal{N}}, \qquad \mathcal{N} = \sum_{i=1}^{m}(2i-1)d_i. \tag{1.64}$$

Proof of the Theorem 1.A. It is now clear that, in coordinates

$$\mathcal{Q}_\lambda(t) = \frac{d}{dt} S^\flat(t)^{-1},$$

as quadratic forms on $(\mathscr{D}_x, \langle \cdot, \cdot \rangle_\lambda)$ (see Eq. (1.61)). By Theorem 1.114, the map $t \mapsto S^\flat(t)^{-1}$ has a simple pole at $t = 0$, and its residue is a negative definite matrix. Then, $\mathcal{Q}_\lambda(t)$ has a second order pole at $t = 0$, and $t^2\mathcal{Q}_\lambda(t)$ can be extended smoothly also at $t = 0$. In particular, $\mathcal{I}_\lambda := \lim_{t\to 0^+} t^2 \mathcal{Q}_\lambda(t) > 0$.

Besides, by Lemma 1.116, $S_{11}(t) \leq S^\flat(t) < 0$, which implies $S^\flat(t)^{-1} \leq S_{11}(t)^{-1} < 0$. Then,

$$\mathcal{I}_\lambda = \lim_{t\to 0^+} t^2 \frac{d}{dt} S^\flat(t)^{-1} = -\lim_{t\to 0^+} tS^\flat(t)^{-1} \geq -\lim_{t\to 0^+} tS_{11}(t)^{-1} = -\dot{S}_{11}(0)^{-1} > 0,$$

which, according to Remark 1.113, implies $\mathcal{I}_\lambda \geq \mathbb{I} > 0$ as operators on $\mathscr{D}_x$.

Finally, $\mathcal{Q}_\lambda(t)$ cannot have a term of order -1 in the Laurent expansion, which is tantamount to $\frac{d}{dt}\Big|_{t=0} t^2\mathcal{Q}_\lambda(t) = 0$. □

1.6.5 Proof of Theorem 1.D

The purpose of this section is the proof of the main result of Section 1.5.6, namely a formula for the exponent of the asymptotic volume growth of geodesic homotheties.

Fix $x_0 \in M$ and let $\gamma : [0,1] \to M$ be the geodesic associated with the covector $\lambda \in T^*_{x_0}M$. Moreover, let J_λ be the associated Jacobi curve. As usual, we fix a Lagrangian splitting $T_\lambda(T^*M) = \mathcal{V}_\lambda \oplus \mathcal{H}_\lambda$, in terms of which $J_\lambda(t)$ is the graph of the map $S(t) : \mathcal{V}_\lambda \to \mathcal{H}_\lambda$. The reader can easily check that the statements that follow do not depend on the choice of the Lagrangian subspaces $\mathcal{H}_\lambda$. The following lemma relates $\mathcal{N}_\lambda$ with the Jacobi curve.

Lemma 1.119. *Assume that γ is ample, of step m, with growth vector $\mathcal{G}_\lambda = \{k_1, \dots, k_m\}$ (at $t = 0$). Then the order of $\det S(t)$ at $t = 0$ is*

$$\det S(t) \sim t^{\mathcal{N}_\lambda}, \qquad \mathcal{N}_\lambda = \sum_{i=1}^{m}(2i-1)(k_i - k_{i-1}).$$

If γ is not ample, the order of $\det S(t)$ at $t = 0$ is $+\infty$.

Proof. Indeed the order of $\det S(t)$ does not depend on the choice of the horizontal complement $\mathcal{H}_\lambda$ and Darboux coordinates. Then, for an ample curve, the statement is precisely Eq. (1.64), obtained in the proof of Theorem 1.114. Finally, if γ is not ample, the Taylor polynomial of arbitrary order of $S(t)$ is singular, thus the order of $\det S(t)$ at $t = 0$ is $+\infty$. □

Proof of Proposition 1.91. By Lemma 1.69, on a sub-Riemannian manifold there always exists at least an ample geodesic γ, with covector λ. Then it is well defined

$$\mathcal{N}_{x_0} = \min\{\mathcal{N}_\lambda \,|\, \lambda \in T^*_{x_0}M\} < +\infty.$$

Let $\lambda \in T^*_{x_0}M$ be a covector at which the minimum is attained. Then, by definition of ample Jacobi curve

$$\operatorname{rank}\{\dot{S}_\lambda(0), \ddot{S}_\lambda(0), \ldots, S_\lambda^{(\mathcal{N}_{x_0})}(0)\} = n,$$

where $S_\lambda(t)$ is the matrix associated with the Jacobi curve $J_\lambda(t)$. We already observed (see Remark 1.110) that, for any Hamiltonian that is fibre-wise polynomial, $S_\lambda^{(i)}(0)$ is a polynomial function of the initial covector $\lambda \in T^*_{x_0}M$. Then $\operatorname{rank}\{\dot{S}_\lambda(0), \ddot{S}_\lambda(0), \ldots, S_\lambda^{(\mathcal{N}_{x_0})}(0)\} < n$ on a closed Zariski subset $\mathcal{Z} \subset T^*_{x_0}M$, which has indeed zero measure. □

We are now ready to prove the main result of Section 1.5.6.

Proof of Theorem 1.D. Without loss of generality, we can assume that Ω is contained in a single coordinate patch $\{x_i\}_{i=1}^n$. In terms of such coordinates, $\mu = e^a dx^1 \wedge \ldots \wedge dx^n$ and

$$\mu(\Omega_{x_0,t}) = \int_\Omega |\det(d_x\phi_t)| e^{a\circ\phi_t(x)} dx. \tag{1.66}$$

By smoothness, it is clear that the order of $\mu(\Omega_{x_0,t})$ at $t = 0$ is equal to the order of the map $t \mapsto \det(d_x\phi_t)$. In the following, $\mathcal{E}_{x_0} : T^*_{x_0}M \to M$ denotes the sub-Riemannian exponential map at time 1. Let us define $\Sigma^*_{x_0} := \mathcal{E}^{-1}_{x_0}(\Sigma_{x_0}) \subset T^*_{x_0}M$. Indeed, if $\lambda \in \Sigma^*_{x_0}$, the associated geodesic $\gamma(t) = \mathcal{E}_{x_0}(t\lambda)$ is the unique one connecting x_0 with $x = \mathcal{E}_{x_0}(\lambda)$. We now compute the order of the map $t \mapsto \det(d_x\phi_t)$.

Lemma 1.120. *For every $x \in \Sigma_{x_0}$ the order of $t \mapsto \det(d_x\phi_t)$ is equal to $\mathcal{N}_\lambda$, where $\lambda = \mathcal{E}^{-1}_{x_0}(x)$.*

Proof. Recall that the order of a family of linear maps does not depend on the choice of the representative matrices. By Eq. (1.45),

$$d_x\phi_t = \pi_* \circ e_*^{(t-1)\vec{H}} \circ d_x^2\mathfrak{f}.$$

Let us focus on the linear map $e_*^{(t-1)\vec{H}} \circ d_x^2\mathfrak{f} : T_xM \to T_{\lambda(t)}(T^*M)$, where $\lambda(t) = e^{t\vec{H}}(\lambda)$ is the normal lift of γ. Let us choose a smooth family of Darboux bases $\{E_i|_{\lambda(t)}, F_i|_{\lambda(t)}\}_{i=1}^n$ of $T_{\lambda(t)}(T^*M)$, such that $\mathcal{V}_{\lambda(t)} = \operatorname{span}\{E_i|_{\lambda(t)}\}_{i=1}^n$ and $\mathcal{H}_{\lambda(t)} = \operatorname{span}\{F_i|_{\lambda(t)}\}_{i=1}^n$. Let us define the column vectors $E|_{\lambda(t)} := (E_1|_{\lambda(t)}, \ldots, E_n|_{\lambda(t)})^*$ and $F|_{\lambda(t)} := (F_1|_{\lambda(t)}, \ldots, F_n|_{\lambda(t)})^*$. Observe that the elements of $\pi_*F|_{\lambda(t)}$ are a smooth family of bases for $T_{\gamma(t)}M$. Then

$$e_*^{(t-1)\vec{H}} \circ d_x^2\mathfrak{f}(\pi_*F|_{\lambda(1)}) = A(t)E|_{\lambda(t)} + B(t)F|_{\lambda(t)}, \tag{1.67}$$

for some smooth families of $n \times n$ matrices $A(t)$ and $B(t)$. Then, by definition, the order of the map $t \mapsto \det(d_x\phi_t)$ is the order of $\det B(t)$ at $t = 0$. By acting with $e_*^{-t\vec{H}}$ in Eq. (1.67), we obtain

$$A(t)e_*^{-t\vec{H}} E|_{\lambda(t)} = e_*^{-\vec{H}} \circ d_x^2\mathfrak{f}(\pi_* F|_{\lambda(1)}) - B(t)e_*^{-t\vec{H}} F|_{\lambda(t)}. \tag{1.68}$$

Notice that $A(0)$ is nonsingular. Then, for t sufficiently close to 0, the l.h.s. of Eq. (1.68) is a smooth basis for the Jacobi curve J_λ. We rewrite the r.h.s. of Eq. (1.68) in terms of the fixed basis $\{E|_{\lambda(0)}, F|_{\lambda(0)}\}$. To this end, observe that

$$e_*^{-t\vec{H}} F|_{\lambda(t)} = C(t)E|_{\lambda(0)} + D(t)F|_{\lambda(0)},$$
$$e_*^{-\vec{H}} \circ d_x^2\mathfrak{f}(\pi_* F|_{\lambda(1)}) = GE|_{\lambda(0)}.$$

For some $n \times n$ smooth matrices $C(t), D(t), G$. Observe that $C(0) = 0$ and $D(t)$ is nonsingular for t sufficiently close to 0. Moreover, since $x \in \Sigma_{x_0}$ is a regular value for the sub-Riemannian exponential map $\mathcal{E}_{x_0} = \pi \circ e^{\vec{H}}$, G is nonsingular. Then

$$A(t)e_*^{-t\vec{H}} E|_{\lambda(t)} = [G - B(t)C(t)]E|_{\lambda(0)} - B(t)D(t)F|_{\lambda(0)}.$$

Therefore, the representative matrix of $J_\lambda(t)$ in terms of the basis $\{E|_{\lambda(0)}, F|_{\lambda(0)}\}$ is

$$S(t) = -[G - B(t)C(t)]^{-1}B(t)D(t), \qquad |t| < \varepsilon.$$

By the properties of the matrices G, $C(t)$ and $D(t)$ for sufficiently small t, $\det S(t) \sim \det B(t)$, and the two determinants have the same order. Then the statement follows from Lemma 1.119. □

By Proposition 1.91, $\mathcal{N}_\lambda = \mathcal{N}_{x_0}$ a.e. on $T^*_{x_0}M$. Then the order of $t \mapsto \det(d_x\phi_t)$ is equal to $\mathcal{N}_{x_0}$ up to a zero measure set on Σ_{x_0} and the statement of Theorem 1.D follows from (1.66), since $\mu(\Omega) > 0$. □

1.7 Asymptotics of the Jacobi curve: equiregular case

In this section, we introduce a key technical tool, the so-called *canonical frame*, associated with a monotone, ample, equiregular curve in the Lagrange Grassmannian $L(\Sigma)$. This is a special moving frame in the symplectic space Σ which satisfies a set of differential equations encoding the dynamics of the underlying curve, which has been introduced for the first time in [71].

The main result of this section is an asymptotic formula for the curve, written in coordinates induced by the canonical frame. Finally, we exploit this result to prove Theorem 1.B.

1.7.1 The canonical frame

Let $J(\cdot) \subset L(\Sigma)$ be an ample, monotone nonincreasing, equiregular curve of rank k. Suppose that its Young diagram D has k rows, of length n_a, for $a = 1, \ldots, k$. Let us fix some terminology about the frames, indexed by the boxes of the Young diagram D. Each box of the diagram is labelled "ai", where $a = 1, \dots, k$ is the row index, and $i = 1, \dots, n_a$ is the progressive box number, starting from the left, in the specified row. Indeed n_a is the length of the a-th row, and $n_1 + \dots + n_k = n = \dim \Sigma$. Briefly, the notation $ai \in D$ denotes a generic box of the diagram.

From now on, we employ letters from the beginning of the alphabet $a, b, c, d, \dots$ for rows, and letters from the middle of the alphabet $i, j, h, k, \dots$ for the position of the box in the row. According to this notation, a frame $\{E_{ai}, F_{ai}\}_{ai \in D}$ for Σ is Darboux if, for any $ai, bj \in D$,

$$\sigma(E_{ai}, E_{bj}) = \sigma(F_{ai}, F_{bj}) = \sigma(E_{ai}, F_{bj}) - \delta_{ab}\delta_{ij} = 0,$$

where $\delta_{ab}\delta_{ij}$ is the Kronecker delta defined on $D \times D$.

A remark on the notation

Any Darboux frame indexed by the boxes of the Young diagram defines a Lagrangian splitting $\Sigma = \mathcal{V} \oplus \mathcal{H}$, where

$$\mathcal{V} = \operatorname{span}\{E_{ai}\}_{ai \in D}, \qquad \mathcal{H} = \operatorname{span}\{F_{ai}\}_{ai \in D}.$$

In the following, we deal with linear maps $S : \mathcal{V} \to \mathcal{H}$ (and their inverses), written in coordinates induced by the frame. The corresponding matrices have a peculiar block structure, associated with the Young diagram. The F_{bj} component of $S(E_{ai})$ is denoted by $S_{ab,ij}$. As a matrix, S can be naturally thought as a $k \times k$ block matrix. The block ab is a $n_a \times n_b$ matrix. This structure is the key of the calculations that follow, and we provide an example. Consider the Young diagram D, together with the "reflected" diagram $\overline{D}$ in Fig. 1.7. We labelled the boxes

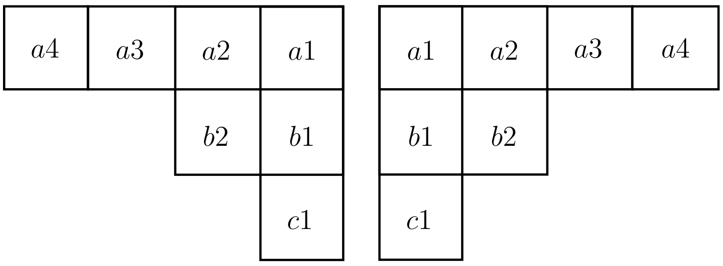


Figure 1.7: The Young diagrams $\overline{D}$ (left) and D (right).

of the diagrams according to the convention introduced above. It is useful to think at each box of the diagram D as a one dimensional subspace of $\mathcal{V}$, and at each box of the diagram $\overline{D}$ as a one dimensional subspace of $\mathcal{H}$. Namely, the box $ai \in D$ corresponds to the subspace $\mathbb{R}E_{ai}$ (respectively, the box $bj \in \overline{D}$ corresponds to the subspace $\mathbb{R}F_{bj}$). Then the matrix S has the following block structure.

$$S = \begin{pmatrix} S_{aa} & S_{ab} & S_{ac} \\ S_{ba} & S_{bb} & S_{bc} \\ S_{ca} & S_{cb} & S_{cc} \end{pmatrix},$$

where each block is a matrix of the appropriate dimension, e.g. S_{ab} is a 4×2 matrix as explained pictorially in Fig. 1.8.

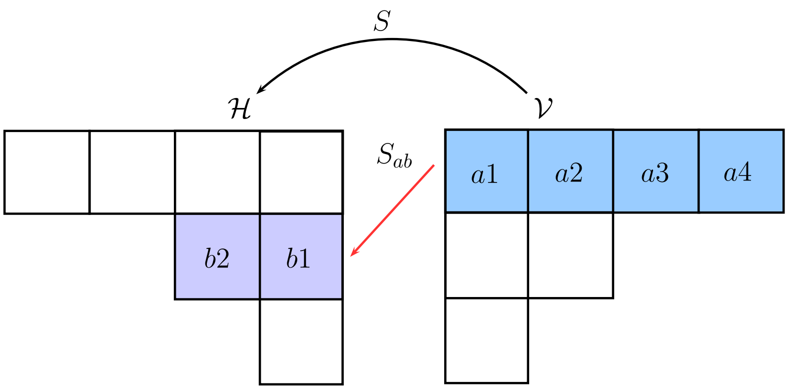


Figure 1.8: The 4×2 block S_{ab} of the map S.

Definition 1.121. A moving Darboux frame $\{E_{ai}(t), F_{ai}(t)\}_{ai \in D}$ is called a *canonical frame* of a monotonically nonincreasing curve $J(\cdot)$ with Young diagram D if $J(t) = \text{span}\{E_{ai}(t)\}_{ai \in D}$ for any t, and there exists a one-parametric family of $n \times n$ symmetric matrices $R(t)$ such that the moving frame satisfies the *structural equations*

$$\begin{aligned}
\dot{E}_{ai}(t) &= E_{a(i-1)}(t), && a = 1, \dots, k, i = 2, \dots, n_a, \\
\dot{E}_{a1}(t) &= -F_{a1}(t), && a = 1, \dots, k, \\
\dot{F}_{ai}(t) &= \sum_{b=1}^{k} \sum_{j=1}^{n_b} R_{ab,ij}(t) E_{bj}(t) - F_{a(i+1)}(t), && a = 1, \dots, k, i = 1, \dots, n_a - 1, \\
\dot{F}_{an_a}(t) &= \sum_{b=1}^{k} \sum_{j=1}^{n_b} R_{ab,n_a j}(t) E_{bj}(t), && a = 1, \dots, k.
\end{aligned}$$

Notice that the matrix $R(t)$ is labelled according to the convention introduced above. The canonical frame for curves in a Lagrange Grassmannian has been introduced for the first time in [71]. In the aforementioned reference, the authors prove that such a frame always exists. Moreover, by requiring some algebraic condition on the family $R(t)$, the authors also proved that the canonical frame is unique up to orthogonal transformations which, in a sense, preserve the structure of the Young diagram. In this case, the family $R(t)$ (which is said to be *normal*) can be associated with a well defined operator which, together with the Young diagram D, completely classify the curve up to symplectic transformations (see also Section 1.7.2). At the end of this section, we also find a formula which connects the curvature operator $\mathcal{R}_\lambda$ of Definition 1.46 with some of the symplectic invariants $R(t)$ of the Jacobi curve (see Eq. (1.81)).

1.7.2 Main result

Fix a canonical frame, associated with $J(\cdot)$. Let $\mathcal{V} = \text{span}\{E_{ai}(0)\}_{ai \in D}$ be the *vertical* subspace, and $\mathcal{H} = \text{span}\{F_{bj}(0)\}_{bj \in D}$ be the *horizontal* subspace of Σ. Observe that $\mathcal{V} = J(0)$. The splitting $\Sigma = \mathcal{V} \oplus \mathcal{H}$ induces a coordinate chart in $L(\Sigma)$, such that $J(t) = \{(p, S(t)p) | \, p \in \mathbb{R}^n\}$. Recall that $S(0) = 0$ and, being the curve ample, $S(t)$ is invertible for $|t| < \varepsilon$ (see Lemma 1.97).

We introduce the constant $n \times n$ symmetric matrices, $\widehat{S}$, its inverse $\widehat{S}^{-1}$ and C, defined by

$$\widehat{S}_{ab,ij} = \frac{\delta_{ab}(-1)^{i+j-1}}{(i-1)!(j-1)!(i+j-1)},$$
$$\widehat{S}^{-1}_{ab,ij} = \frac{-\delta_{ab}}{i+j-1}\binom{n_a+i-1}{i-1}\binom{n_b+j-1}{j-1}\frac{(n_a)!(n_b)!}{(n_a-i)!(n_b-j)!},$$
$$C_{ab,ij} = \frac{(-1)^{i+j}(i+j+2)}{(i-1)!(j-1)!(i+j+1)(i+1)(j+1)}.$$

where, as usual, $a, b = 1, \dots, k$, $i = 1, \dots, n_a$, $j = 1, \dots, n_b$.

Theorem 1.122. *Let $J(\cdot)$ be a monotone, ample, equiregular curve of rank k, with a given Young diagram D with k rows, of length n_a, for $a = 1, \dots, k$. Then, for $|t| < \varepsilon$*

$$S_{ab,ij}(t) = \widehat{S}_{ab,ij}t^{i+j-1} - R_{ab,11}(0)C_{ab,ij}t^{i+j+1} + O(t^{i+j+2}). \tag{1.69}$$

Moreover, for $0 < |t| < \varepsilon$, the following asymptotic expansion holds for the inverse matrix:

$$S^{-1}_{ab,ij}(t) = \frac{\widehat{S}^{-1}_{ab,ij}}{t^{i+j-1}} + R_{ab,11}(0)\frac{(\widehat{S}^{-1}C\widehat{S}^{-1})_{ab,ij}}{t^{i+j-3}} + O\left(\frac{1}{t^{i+j-4}}\right). \tag{1.70}$$

Eqs. (1.69) and (1.70) highlight the block structure of the S matrix and its inverse at the leading orders. In particular, they give the leading order of the principal part of S^{-1} on the diagonal blocks (i.e. when $a = b$). The leading order terms of the diagonal blocks of S (and its inverse S^{-1}) only depend on the structure of the given Young diagram. Indeed the dependence on $R(t)$ appears in the higher order terms of Eqs. (1.69) and (1.70).

Restriction

At the end of this section, we apply Theorem 1.122 to compute the expansion of the family of operators $\mathcal{Q}_\lambda(t)$. According to the discussion that follows Eq. (1.60), we only need a block of the matrix $S(t)^{-1}$, namely $S^\flat(t)^{-1}$. As we explain below, it turns out that this corresponds to consider only the restriction of S^{-1} to the first columns of the Young diagram D and $\overline{D}$ (see Fig. 1.9). In terms of the frame $\{F_{a1}(0), E_{a1}(0)\}_{a=1}^k$, the map $S^\flat(t)^{-1}$ is a $k \times k$ matrix, with entries $S^\flat(t)^{-1}_{ab} = (S^{-1})_{ab,11}$. The following corollary is a consequence of Theorem 1.122, and gives the principal part of the aforementioned block.

Corollary 1.123. *Let $J(\cdot)$ be a monotone, ample, equiregular curve of rank k, with a given Young diagram D with k rows, of length n_a, for $a = 1, \dots, k$. Then, for $0 < |t| < \varepsilon$*

$$S^\flat(t)^{-1}_{ab} = -\delta_{ab}\frac{n_a^2}{t} + R_{ab,11}(0)\Omega(n_a, n_b)t + O(t^2), \tag{1.71}$$

where

$$\Omega(n_a, n_b) = \begin{cases} 0 & |n_a - n_b| \geq 2, \\ \frac{1}{4(n_a+n_b)} & |n_a - n_b| = 1, \\ \frac{n_a}{4n_a^2-1} & n_a = n_b. \end{cases} \tag{1.72}$$

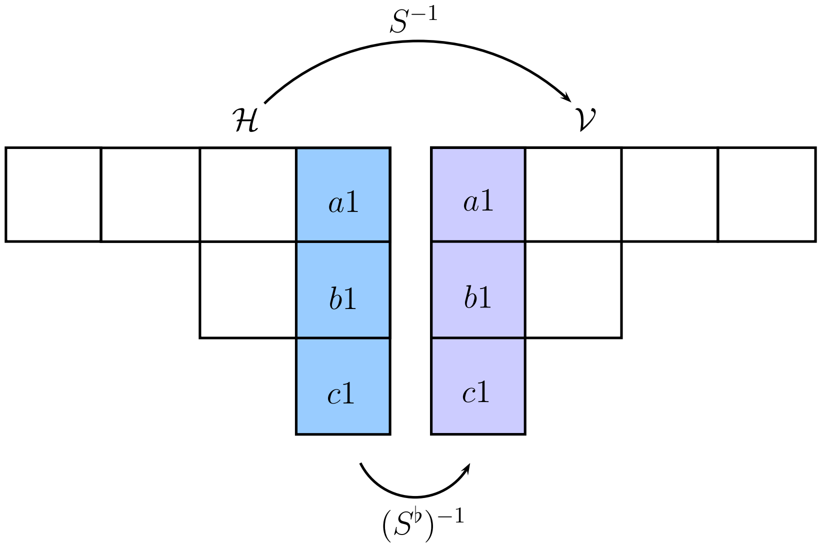


Figure 1.9: The block $S^\flat(t)^{-1}$ of the map $S(t)^{-1}$. Namely $(S^\flat)^{-1}_{ab} = S^{-1}_{ab,11}$.

Remark 1.124. If the Young diagram consists in a single column, with n boxes, $n_a = 1$ for all $a = 1, \dots, n$ and

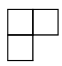
$$S^\flat(t)^{-1}_{ab} = -\frac{\delta_{ab}}{t} + \frac{1}{3} R_{ab}(0) t + O(t^2).$$

A remark on the coefficients

Let us discuss the consequences of the peculiar form of the coefficients of Eq. (1.72). If $|n_a - n_b| \geq 2$, $\Omega(n_a, n_b) = 0$ and the corresponding $R_{ab,11}$ does not appear in the first order asymptotic. Nevertheless, if we assume that $R(t)$ is a *normal family* in the sense of [71], the "missing" entries are precisely the ones that vanish due to the assumptions on $R(t)$. It is natural to expect that some of the $R_{ab,ij}$ do not appear also in the higher orders of the asymptotic expansion. This may suggest the algebraic conditions to enforce on a generic family $R_{ab,ij}$ in order to obtain a truly canonical moving frame for the Jacobi curve.

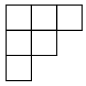
Examples

In this section we provide two practical examples of the asymptotic form of $S^\flat(t)^{-1}$. We suppress the subscript "$_{11}$" and the evaluation at $t = 0$ from each entry $R_{ab,11}(0)$.

A) Consider the 3-dimensional Jacobi curve with Young diagram: 

$$S^\flat(t)^{-1} = -\frac{1}{t}\begin{pmatrix} 4 & 0 \\ 0 & 1 \end{pmatrix} + \frac{1}{3}\begin{pmatrix} \frac{2}{5}R_{11} & \frac{1}{4}R_{12} \\ \frac{1}{4}R_{21} & R_{22} \end{pmatrix} t + O(t^2).$$

This corresponds to the case of the Jacobi curve associated with the geodesics of a 3D contact sub-Riemannian structure.

B) Consider the diagram: 

$$S^\flat(t)^{-1} = -\frac{1}{t}\begin{pmatrix} 9 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 1 \end{pmatrix} + \frac{1}{3}\begin{pmatrix} \frac{9}{35}R_{11} & \frac{3}{20}R_{12} & 0 \\ \frac{3}{20}R_{21} & \frac{2}{5}R_{22} & \frac{1}{4}R_{23} \\ 0 & \frac{1}{4}R_{23} & R_{33} \end{pmatrix} t + O(t^2).$$

This corresponds to the case of the Jacobi curve associated with a generic ample geodesics of a $(3,6)$ Carnot group. In this example we can appreciate that some of the $R_{ab,11}$ do not appear in the linear term of the reduced matrix.

1.7.3 Proof of Theorem 1.122

The proof boils down to a careful manipulation of the structural equations, and matrices inversions. We prove Theorem 1.122 in three steps.

1. First, we consider the case of a rank 1 curve, and we assume $R(t) = 0$. In this case, the Young diagram is a single row and the structural equations are very simple. The canonical frame at time t is a polynomial in terms of the canonical frame at $t = 0$, and we compute explicitly the matrix $S(t)$ and its inverse.

2. Then, we consider a general rank 1 curve. The canonical frame at time t is no longer a polynomial in terms of the canonical frame at $t = 0$, but we can control the higher order terms. The non-vanishing $R(t)$ gives a contribution of higher order in t in each entry of the matrix $S(t)$ and its inverse.

3. Finally, we consider a general rank k curve. We show that, at the leading orders, we can "split" the curve in k rank 1 curves, and employ the results of the previous steps.

Rank 1 curve with vanishing $R(t)$

With these assumptions, the canonical frame is $\{E_i(t), F_i(t)\}_{i=1}^n$ (we suppress the row index, as D has a single row). The structural equations are

$$\begin{aligned} \dot{E}_1(t) &= -F_1(t), & \dot{F}_1(t) &= -F_2(t), \\ \dot{E}_2(t) &= E_1(t), & \dot{F}_2(t) &= -F_3(t), \\ &\vdots & &\vdots \\ \dot{E}_n(t) &= E_{n-1}(t), & \dot{F}_n(t) &= 0. \end{aligned}$$

Pictorially, in the double Young diagram the derivative shifts each element of the frame to the left by one box (see Fig. 1.10).

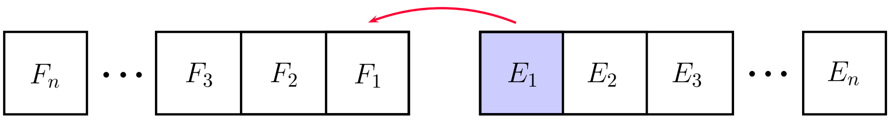


Figure 1.10: The action of the derivative on E_1.

Let $E(t) = (E_1, \dots, E_n)^*$ and $F(t) = (F_1, \dots, F_n)^*$, where each element is computed at t. Then there exist one parameter families of $n \times n$ matrices $A(t), B(t)$ such that

$$E(t) = A(t)E(0) + B(t)F(0).$$

$A(t)$ and $B(t)$ have monomial entries w.r.t. t. For $i, j = 1, \dots, n$

$$A_{ij}(t) = \frac{t^{i-j}}{(i-j)!} = \widehat{A}_{ij} t^{i-j}, \qquad (i \geq j), \tag{1.73}$$

$$B_{ij}(t) = \frac{(-1)^j t^{i+j-1}}{(i+j-1)!} = \widehat{B}_{ij} t^{i+j-1}. \tag{1.74}$$

Observe that A is a lower triangular matrix. A straightforward computation shows that

$$A^{-1}_{ij}(t) = \frac{(-1)^{i-j} t^{i-j}}{(i-j)!} = \widehat{A}^{-1}_{ij} t^{i-j}, \qquad (i \geq j). \tag{1.75}$$

Eqs. (1.73), (1.74) and (1.75) implicitly define the constant matrices $\widehat{A}$, $\widehat{B}$ and $\widehat{A}^{-1}$. The matrix $S(t)$ can be computed directly in terms of $A(t)$ and $B(t)$. Indeed $S(t) = A(t)^{-1}B(t)$.

Proposition 1.125 (Special case of Theorem 1.122)**.** *Let $J(\cdot)$ a curve of rank 1, with vanishing $R(t)$. The matrix $S(t)$, in terms of a canonical frame, is*

$$S(t)_{ij} = \frac{(-1)^{i+j-1}}{(i-1)!(j-1)!} \frac{t^{i+j-1}}{(i+j-1)} = \widehat{S}_{ij} t^{i+j-1}. \tag{1.76}$$

Its inverse is

$$S^{-1}(t)_{ij} = \frac{-1}{i+j-1} \binom{n+i-1}{i-1} \binom{n+j-1}{j-1} \frac{(n!)^2}{(n-i)!(n-j)!} = \frac{\widehat{S}^{-1}_{ij}}{t^{i+j-1}}. \tag{1.77}$$

As expected, $S(t)$ is symmetric, since the canonical frame is Darboux. The proof of Proposition 1.125 is a straightforward but long computation, which can be found in Appendix A. Eqs. (1.76) (1.77) implicitly define the constant matrix $\widehat{S}$ and its inverse $\widehat{S}^{-1}$. Observe that the entries of the latter depend explicitly on the dimension n.

General rank 1 curve

Now consider a general rank 1 curve. Its Young diagram is still a single row but, in general, $R(t) \neq 0$. As a consequence, the elements of the moving frame are no longer polynomial in t. However, we can still expand each $E_i(t)$ and obtain a Taylor approximation of its components w.r.t. the frame at $t = 0$. Each derivative at $t = 0$, up to order $i - 1$, is still a vertical vector

$$\frac{d^k E_i}{dt^k}(0) = E_{i-k}(0), \qquad k = 0, \ldots, i-1.$$

The i-th derivative at $t = 0$ gives the lowest order horizontal term, i.e.

$$\frac{d^i E_i}{dt^i}(0) = -F_1(0).$$

Henceforth, each additional derivative, computed at $t = 0$, gives higher order horizontal terms, but also new vertical terms, depending on $R(t)$. Let us see a particular example, for $E_1(t)$. $\dot{E}_1(0) = -F_1(0)$, and $\ddot{E}_1(0) = F_2(0) - \sum_{j=1}^n R_{1j}(0)E_j(0)$ (see Fig. 1.11).

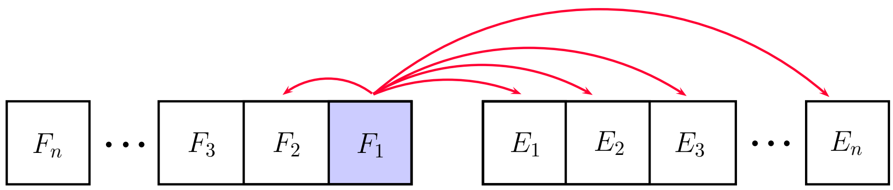


Figure 1.11: The action of the derivative of an horizontal element of the frame when $R \neq 0$.

Indeed $E_1(t)$ has a zeroth order term (w.r.t. the variable t) in the direction $E_1(0)$. The next term in the direction $E_1(0)$ is of order 2 or more. Besides, $E_1(t)$ has vanishing zeroth order term in each other vertical direction (i.e. $E_j(0)$, $j \neq 1$), but non vanishing components in each other vertical direction can appear, at orders greater or equal than 2. Let us turn to the horizontal components. $E_1(t)$ has a first order term in the direction $F_1(0)$. The next term in the same direction can appear only after two additional derivatives, or more. Therefore, the next term in the direction $F_1(0)$ is of order 3 or more in t. The “gaps” in the orders appearing in a given directions are precisely the key to the proof.

Let $E(t) = (E_1, \ldots, E_n)^*$ and $F(t) = (F_1, \ldots, F_n)^*$, where each element is computed at t. Then, as in the previous step, there exist one parameter families of $n \times n$ matrices $A(t)$, $B(t)$ such that

$$E(t) = A(t)E(0) + B(t)F(0).$$

The discussion above, and a careful application of the structural equations give us asymptotic formulae for the matrices $A(t)$ and $B(t)$. Let $\widehat{A}$ and $\widehat{B}$ defined as in Eqs. (1.73)-(1.74), corresponding to the case of a rank 1 curve with vanishing $R(t)$. Then, for $i, j = 1, \ldots, n$

$$A(t)_{ij} = \widehat{A}_{ij} t^{i-j} - R_{1j}(0)\frac{t^{i+1}}{(i+1)!} + O(t^{i+2}),$$

$$B(t)_{ij} = \widehat{B}_{ij} t^{i+j-1} + R_{11}(0)\frac{(-1)^{j+1} t^{i+j+1}}{(i+j+1)!} + O(t^{i+j+2}).$$

The matrix A is no longer triangular, due to the presence of higher order terms in each entry. Besides, the order of the remainder grows only with the row index for $A(t)$ and it grows with both the column and row indices for $B(t)$. This reflects the different role played by the horizontal and vertical terms in the structural equations. We are now ready to consider the general case.

General rank k curve

The last step, which concludes the proof of the theorem, is built upon the previous cases. It is convenient to split a frame in subframes, relative to the rows of the Young diagram. For $a = 1, \ldots, k$, the symbol E_a denotes the n_a-dimensional column vector

$$E_a = (E_{a1}, E_{a2}, \ldots, E_{an_a})^* \in \Sigma^{n_a},$$

and analogously for F_a. Similarly, the symbol E denotes the n-dimensional column vector

$$E = (E_1, \ldots, E_k)^* \in \Sigma^n,$$

and similarly for F. Once again, we express the elements of the Jacobi curves $E(t)$ in terms of the canonical frame at $t = 0$. With the notation introduced above

$$E(t) = A(t)E(0) + B(t)F(0).$$

This time, $A(t)$ and $B(t)$ are $k \times k$ block matrices, the ab block being a $n_a \times n_b$ matrix. For $a, b = 1, \ldots, k$, $i = 1, \ldots, n_a$, $j = 1, \ldots, n_b$

$$A(t)_{ab,ij} = \delta_{ab}\widehat{A}_{ij} t^{i-j} - R_{ab,1j}(0)\frac{t^{i+1}}{(i+1)!} + O(t^{i+2}), \tag{1.78}$$

$$B(t)_{ab,ij} = \delta_{ab}\widehat{B}_{ij} t^{i+j-1} + R_{ab,11}(0)\frac{(-1)^{j+1} t^{i+j+1}}{(i+j+1)!} + O(t^{i+j+2}),$$

where, once again, the constant matrices $\widehat{A}$, $\widehat{B}$ correspond to the matrices defined for the rank 1 and $R(t) = 0$ case, of the appropriate dimension. Notice that we do not need explicitly the leading terms on the off-diagonal blocks. The knowledge of the leading terms on the diagonal blocks is sufficient for our purposes.

Remember that $S(t) = A(t)^{-1}B(t)$. In order to compute the inverse of $A(t)$ at the relevant order, we rewrite the matrix $A(t)$ as

$$A(t) = \widehat{A}(t) - M(t),$$

where $\widehat{A}(t)$ is the matrix corresponding to a rank k curve with vanishing $R(t)$, namely

$$\widehat{A}(t)_{ab,ij} = \delta_{ab}\widehat{A}_{ij} t^{i-j}, \qquad i = 1, \ldots, n_a, \quad j = 1, \ldots, n_b,$$

and, from Eq. (1.78), we get

$$M(t)_{ab,ij} = R_{ab,1j}(0)\frac{t^{i+1}}{(i+1)!} + O(t^{i+2}).$$

A standard inversion of the Neumann series leads to

$$A(t)^{-1} = \widehat{A}(t)^{-1} + \widehat{A}(t)^{-1}M(t)\widehat{A}(t)^{-1} + \sum_{n=2}^{\infty}\left(\widehat{A}(t)^{-1}M(t)\right)^n \widehat{A}(t)^{-1},$$

where the reminder term in the r.h.s. converges uniformly in the operator norm small t. Then, a long computation gives

$$A(t)^{-1}_{ab,ij} = \delta_{ab}\widehat{A}^{-1}_{ij}t^{i-j} - R_{ab,11}(0)\frac{(-1)^i t^{i+1}}{(i+1)(i-1)!} + O(t^{i+2}).$$

The matrix $S(t)$ can be computed explicitly, at the leading order, by the usual formula $S(t) = A(t)^{-1}B(t)$, and we obtain, for $a, b = 1, \dots, k$, $i = 1, \dots, n_a$, $j = 1, \dots, n_b$,

$$S(t)_{ab,ij} = \widehat{S}_{ab,ij}t^{i+j-1} - R_{ab,11}(0)C_{ab,ij}t^{i+j+1} + O(t^{i+j+2}),$$

where $\widehat{S}_{ab,ij} = \delta_{ab}\widehat{S}_{ij}$ of the appropriate dimension, and

$$C_{ab,ij} = \frac{(-1)^{i+j}(i+j+2)}{(i-1)!(j-1)!(i+j+1)(i+1)(j+1)}, \qquad i = 1, \dots, n_a, \quad j = 1, \dots, n_b.$$

The computation of $S(t)^{-1}$ follows from another inversion of the Neumann series, and a careful estimate of the remainder. We obtain

$$S^{-1}_{ab,ij}(t) = \frac{\widehat{S}^{-1}_{ab,ij}}{t^{i+j-1}} + R_{ab,11}(0)\frac{(\widehat{S}^{-1}C\widehat{S}^{-1})_{ab,ij}}{t^{i+j-3}} + O\left(\frac{1}{t^{i+j-4}}\right),$$

where

$$\widehat{S}^{-1}_{ab,ij} = \frac{-\delta_{ab}}{i+j-1}\binom{n_a+i-1}{i-1}\binom{n_b+j-1}{j-1}\frac{n_a!n_b!}{(n_a-i)!(n_b-j)!}.$$

This concludes the proof of Theorem 1.122. □

Proof of Corollary 1.123

Corollary 1.123 follows easily from Theorem 1.122. The only non trivial part is the explicit form of the coefficient $\Omega(n_a, n_b)$ in Eq. (1.71). By the results of Theorem 1.122,

$$\Omega(n_a, n_b) = (\widehat{S}^{-1}C\widehat{S}^{-1})_{ab,11}.$$

By replacing the explicit expression of $\widehat{S}^{-1}$ and C, the proof of Corollary 1.123 is reduced to the following lemma, which we prove in Appendix B.

Lemma 1.126. *Let $\Omega(n,m)$ be defined by the formula*

$$\Omega(n,m) = \frac{nm}{(n+1)(m+1)} \sum_{j=1}^{n}\sum_{i=1}^{m}(-1)^{i+j}\binom{n+i-1}{i-1}\binom{n+1}{i+1}\binom{m+j-1}{j-1}\binom{m+1}{j+1}\frac{i+j+2}{i+j+1}.$$

Then

$$\Omega(n,m) = \begin{cases} 0 & |n-m| \geq 2, \\ \frac{1}{4(n+m)} & |n-m| = 1, \\ \frac{n}{4n^2-1} & n = m. \end{cases}$$

The proof of Corollary 1.123 is now complete. □

1.7.4 Proof of Theorem 1.B

In this section $J_\lambda : [0,T] \to L(T_\lambda(T^*M))$ is the Jacobi curve associated with an ample, equiregular geodesic γ, with initial covector $\lambda \in T_x^*M$. The next lemma shows that the projection of the horizontal part of the canonical frame corresponding to the first column of the Young diagram is an orthonormal basis for the Hamiltonian product on the distribution.

Lemma 1.127. *Let $X_a := \pi_* F_{a1}(0) \in T_xM$. Then, the set $\{X_a\}_{a=1}^k$ is an orthonormal basis for $(\mathscr{D}_x, \langle\cdot|\cdot\rangle_\lambda)$.*

Proof. First, recall that $F_{a1}(0) = -\dot{E}_{a1}(0)$. Therefore $X_a = -\pi_*\dot{E}_{a1}(0)$. Then, by Eq. (1.59)

$$\langle X_a|X_b\rangle_\lambda = -\sigma(E_{a1}(0), \dot{E}_{b1}(0)) = \sigma(E_{a1}(0), F_{b1}(0)) = \delta_{ab}.$$

where we used the structural equations and the fact that the canonical frame is Darboux. □

We are now ready to prove one of the main results of Section 1.4.3, namely the one concerning the spectrum of the operator $\mathcal{I}_\lambda : \mathscr{D}_x \to \mathscr{D}_x$.

Proof of Theorem 1.B. Actually, we prove something more: we use the basis $\{X_a\}_{a=1}^k$ obtained above to compute an asymptotic formula for the family $\mathcal{Q}_\lambda(t)$ introduced in Section 1.4.3.

Let $\Sigma = \mathcal{V}_\lambda \oplus \mathcal{H}_\lambda$ be the splitting induced by the canonical frame in $\Sigma = T_\lambda(T^*M)$. Let $S(t) : \mathcal{V}_\lambda \to \mathcal{H}_\lambda$ be the map which represents the Jacobi curve in terms of the canonical splitting. Then, by definition of Jacobi curve, it follows that, for any $v \in T_xM$ (see also Eq. (1.60)),

$$\langle \mathcal{Q}_\lambda(t)v|v\rangle_\lambda = \frac{d}{dt}\sigma(S(t)^{-1}\widetilde{v}, \widetilde{v}).$$

where $\tilde{v} \in \mathcal{H}_\lambda$ is the unique horizontal lift such that $\pi_*\widetilde{v} = v$. In particular, if $v = \sum_{a=1}^k v_a X_a \in \mathscr{D}_x$, we have $\widetilde{v} = \sum_{a=1}^k v_a F_{a1}(0)$. Thus,

$$\langle \mathcal{Q}_\lambda(t)v|v\rangle_\lambda = \frac{d}{dt}\sum_{a,b=1}^k S(t)^{-1}_{ab,11} v_a v_b = \frac{d}{dt}\sum_{a,b=1}^k S^\flat(t)^{-1}_{ab} v_a v_b.$$

By Corollary 1.123, we obtain finally the following asymptotic formula for $\mathcal{Q}_\lambda(t)$.

$$\langle \mathcal{Q}_\lambda(t)v|v\rangle_\lambda = \sum_{a,b=1}^{k} \left(\delta_{ab}\frac{n_a^2}{t^2} + R_{ab,11}(0)\Omega(n_a,n_b)\right) v_a v_b + O(t). \tag{1.79}$$

Equation (1.79), together with Lemma 1.127 imply that, for $a, b = 1, \dots, k$,

$$\mathcal{I}_\lambda X_a = n_a^2 X_a, \tag{1.80}$$

$$\mathcal{R}_\lambda X_a = 3R_{ab,11}(0)\Omega(n_a,n_b)X_b. \tag{1.81}$$

Equation (1.80) completely characterizes the spectrum and the eigenvectors of $\mathcal{I}_\lambda$. □

Equation (1.81) is the anticipated formula which connects the curvature operator of Definition 1.46 with some of the symplectic invariants of the Jacobi curve, namely the elements of the matrix $R_{ab,ij}$ corresponding to the first column of the Young diagram.

1.8 Sub-Laplacian and Jacobi curves

Throughout this section, we assume M to be an equiregular sub-Riemannian manifold (that is, the rank of the distribution $\mathscr{D}$ is constant, equal to k). Nevertheless, most of the statements of this section hold true in the general case, by replacing the sub-Riemannian inner product on $\mathscr{D}$ with the Hamiltonian inner product. The final goal of this section is the proof of Theorem 1.C, that is an asymptotic formula for the sub-Laplacian of the cost function. We start with a general discussion about the computation of the sub-Laplacian at a fixed point.

Let $f \in C^\infty(M)$, $x \in M$ and $\lambda = d_x f \in T_x^*M$. Moreover, let $X_1, \dots, X_k$ be a local orthonormal frame for the sub-Riemannian structure. All our considerations are local, then we assume without loss of generality that the frame $X_1, \dots, X_n$ is globally defined. Then, by Eq. (1.36), the sub-Laplacian associated with the volume form μ writes

$$\Delta_\mu f = \sum_{i=1}^{k} X_i^2(f) + \mathrm{div}_\mu(X_i)X_i(f).$$

As one can see, the sub-Laplacian is the sum of two terms. The first term, $\sum_{i=1}^k X_i^2(f)$, is a "sum of squares" which does not depend on the choice of the volume form. On the other hand, the second term, namely $\sum_{i=1}^k \mathrm{div}_\mu(X_i)X_i(f)$ depends on μ through the divergence operator. When x is a critical point for f, the second term vanishes, and the sub-Laplacian can be computed by taking the trace of the ordinary second differential of f (see Lemma 1.78). On the other hand, if x is non-critical, we need to compute both terms explicitly.

We start with the second term. Let $\theta_1, \dots, \theta_n$ be the coframe dual to $X_1, \dots, X_n$. Namely $\theta_i(X_j) = \delta_{ij}$. Then, there exists a smooth function $g \in C^\infty(M)$ such that $\mu = e^g \theta_1 \wedge \dots \wedge \theta_n$. Finally, let $c_{ij}^k \in C^\infty(M)$ be the *structure functions* defined by $[X_i, X_j] = \sum_{k=1}^n c_{ij}^k X_k$. A standard computation using the definition of divergence gives

$$\mathrm{div}_\mu(X_i) = X_i(g) - \sum_{j=1}^{n} c_{ij}^j.$$

Thus, the second term of the sub-Laplacian is

$$\sum_{i=1}^{k} \mathrm{div}_\mu(X_i) X_i(f) = \langle \nabla f | \nabla g \rangle - \sum_{i=1}^{k} \sum_{j=1}^{n} c_{ij}^{j} X_i(f). \tag{1.82}$$

The first term of the sub-Laplacian can be computed through the generalized second differential introduced with Definition 1.99. Recall that the second differential at a non critical point x is a linear map $d_x^2 f : T_x M \to T_\lambda(T^*M)$.

1.8.1 Coordinate lift of a local frame

We introduce a special basis of $T_\lambda(T^*M)$, associated with a choice of the local frame $X_1, \dots, X_n$, which is a powerful tool for explicit calculations. We define an associated frame on T^*M as follows. For $i = 1, \dots, n$ let $h_i : T^*M \to \mathbb{R}$ be the linear-on-fibres function defined by $\lambda \mapsto h_i(\lambda) := \langle \lambda, X_i \rangle$. The action of the derivations on T^*M is completely determined by the action on affine functions, namely functions $a \in C^\infty(T^*M)$ such that $a(\lambda) = \langle \lambda, Y \rangle + \pi^* g$ for some $Y \in \mathrm{Vec}(M)$, $g \in C^\infty(M)$. Then, we define the *coordinate lift of a field* $X \in \mathrm{Vec}(M)$ as the field $\widetilde{X} \in \mathrm{Vec}(T^*M)$ such that $\widetilde{X}(h_i) = 0$ for $i = 1, \dots, n$ and $\widetilde{X}(\pi^* g) = X(g)$. This, together with Leibniz rule, characterize the action of $\widetilde{X}$ on affine functions, and then completely define $\widetilde{X}$. Indeed, by definition, $\pi_* \widetilde{X} = X$. On the other hand, we define the (vertical) fields ∂_{h_i} such that $\partial_{h_i}(\pi^* g) = 0$, and $\partial_{h_i}(h_j) = \delta_{ij}$. It is easy to check that $\{\partial_{h_i}, \widetilde{X}_i\}_{i=0}^{n}$ is a frame on T^*M. We call such a frame the *coordinate lifted frame*, and we employ the shorthand $\partial_i := \partial_{h_i}$. Observe that, by the same procedure, we can define the coordinate lift of a vector $X \in T_x M$ (i.e. not necessarily a field) at any point $\lambda \in T_x^* M$.

Remark 1.128. Remember that we require $X_1, \dots, X_n$ to be *fields* (and not simple vectors in $T_x M$) in order to define the coordinate lift. In particular, the lift $\widetilde{X}|_\lambda \in T_\lambda(T^*M)$ depends on the germ at x of the chosen frame $X_1, \dots, X_n$. On the other hand, $\partial_i|_\lambda$ depends only on the value of $X_1, \dots, X_n$ at x.

Lemma 1.129. *Let $X \in T_x M$. In terms of a coordinate lifted frame,*

$$d_x^2 f(X) = \widetilde{X} + \sum_{i=1}^{n} X(X_i(f)) \partial_i,$$

where $X(X_i(f))$ is understood to be computed at x and $\widetilde{X}, \partial_i \in T_\lambda(T^*M)$.

Proof. We explicitly compute the action of the vector $d_x^2 f(X) \in T_\lambda(T^*M)$ on affine functions. First, for any $g \in C^\infty(M)$, $d_x^2 f(X)(\pi^* g) = \pi_* \circ d_x^2 f(X)(g) = X(g)$. Moreover, $d_x^2 f(X)(h_i) = X(h_i \circ df) = X(\langle df, X_i \rangle) = X(X_i(f))$. □

Lemma 1.129, when applied to the vectors $X_1, \dots, X_k$, completely characterize the second order component of the sub-Laplacian, in terms of the second differential $d_x^2 f$.

1.8.2 Sub-Laplacian of the geodesic cost

Assume $f = c_t$, that is the geodesic cost associated with an ample, equiregular geodesic $\gamma : [0,T] \to M$. As usual, let $x = \gamma(0)$ be the initial point, $\lambda = d_x c_t$ the initial covector, and $J_\lambda(\cdot)$ the associated Jacobi curve, with Young diagram D. As discussed in Section 1.7, there is a class of preferred frames in $T_\lambda(T^*M)$, namely the canonical moving frame $\{E_{ai}(t), F_{ai}(t)\}_{ai\in D}$. In order to employ the results of Theorem 1.122 for the computation of Δc_t, we first relate the canonical frame with a coordinate lifted frame. As a first step, we need the following lemma, which is an extension of Lemma 1.127 along the geodesic.

Lemma 1.130. *Let $\{E_{ai}(t), F_{ai}(t)\}_{ai\in D}$ be a canonical moving frame for $J_\lambda(\cdot)$ and consider the following vector fields along γ:*

$$X_{ai}(t) := \pi_* \circ e^{t\vec{H}}_* F_{ai}(t) \in T_{\gamma(t)}M, \qquad ai \in D.$$

The set $\{X_{ai}(t)\}_{ai\in D}$ is a basis for $T_{\gamma(t)}M$. Moreover $\{X_{a1}(t)\}_{a=1}^k$ is an orthonormal basis for $\mathscr{D}_{\gamma(t)}$ along the geodesic. Finally, consider any smooth extension of $\{X_{ai}(t)\}_{ai\in D}$ in a neighbourhood of γ, and the associated coordinate lifted frame. Then

$$E_{ai}(t) = e^{-t\vec{H}}_* \partial_{ai}|_{\lambda(t)},$$

Lemma 1.130 states that the projection of the horizontal elements of the canonical frame (the "F"s) corresponding to the first column of the Young diagram are an orthonormal frame for the sub-Riemannian distribution along the geodesic. Moreover, if we complete the frame with the projections of the other horizontal elements, and we introduce the associated coordinate lifted frame along the extremal $e^{t\vec{H}}(\lambda)$, the vertical elements of the canonical frame (the "E"s) have a simple expression. Observe that, according to Remark 1.128, the last statement of the lemma does not depend on the choice of the extension of the vectors $X_{ai}(t)$ in a neighbourhood of γ.

Proof. Assume first that the statement is true at $t = 0$. Then, let $0 < t < T$. Point (ii) of Proposition 1.105 gives the relation between the Jacobi curves "attached" at different points $\lambda(t) = e^{t\vec{H}}(\lambda)$ along the lift of γ. Namely

$$J_{\lambda(t)}(\cdot) = e^{t\vec{H}}_* J_\lambda(t + \cdot).$$

As a consequence of this, and the definition of canonical frame, if $\{E_{ai}(\cdot), F_{ai}(\cdot)\}_{ai\in D}$ is a canonical frame for the Jacobi curve $J_\lambda(\cdot)$, it follows that, for any fixed t,

$$\widetilde{E}_{ai}(\cdot) := e^{t\vec{H}}_* E_{ai}(t + \cdot),$$
$$\widetilde{F}_{ai}(\cdot) := e^{t\vec{H}}_* F_{ai}(t + \cdot),$$

is a canonical frame for the Jacobi curve $J_{\lambda(t)}(\cdot)$. In particular, $X_{ai}(t) = \pi_* \widetilde{F}_{ai}(0)$, and the statements now follow from the assumption that the lemma is true at the initial time of the Jacobi curve $J_{\lambda(t)}(\cdot)$.

Then, we only need to prove the statement at $t = 0$. For clarity, we suppress the explicit evaluation at $t = 0$. As usual, let $\mathcal{H}_\lambda = \mathrm{span}\{F_{ai}\}_{ai\in D}$ be the horizontal subspace and $\mathcal{V}_\lambda = \mathrm{span}\{E_{ai}\}_{ai\in D}$ be the vertical subspace. By definition of canonical frame, $T_\lambda(T^*M) = \mathcal{H}_\lambda \oplus \mathcal{V}_\lambda$. Since $\mathcal{V}_\lambda = \ker \pi_*$, and π_* is a submersion, $\pi_*\mathcal{H}_\lambda = T_xM$. Thus $\{X_{ai}\}_{ai\in D}$ is a basis for T_xM. By Lemma 1.127, the set $\{X_{a1}\}_{a=1}^k$ is an orthonormal frame for the Hamiltonian inner product $\langle \cdot | \cdot \rangle_\lambda$ which, in the sub-Riemannian case, does not depend on λ and coincides with the sub-Riemannian inner product (see Remark 1.45). Now, we show that $E_{ai} = \partial_{ai}|_\lambda$. Since the canonical frame is Darboux, this is equivalent to $\sigma(\partial_{ai}, F_{bj}) = \delta_{ab}\delta_{ij}$. Indeed, in terms of the coframe $\{\theta_{ai}\}_{ai\in D}$, dual to $\{X_{ai}\}_{ai\in D}$

$$\sigma = \sum_{ai\in D} dh_{ai} \wedge \pi^*\theta_{ai} + h_{ai}\pi^* d\theta_{ai}.$$

Therefore

$$\sigma(\partial_{ai}, F_{bj}) = \theta_{ai}(\pi_* F_{bj}) = \theta_{ai}(X_{bj}) = \delta_{ab}\delta_{ij}.$$

□

We now have all the tools we need in order to prove Theorem 1.C, concerning the asymptotic behaviour of Δc_t.

Proof of Theorem 1.C. The idea is the compute the "hard" term of Δc_t, namely the sum of squares term, through the coordinate representation of the Jacobi curve. By Lemma 1.129, written in terms of the frame $X_{ai} := X_{ai}(0) = \pi_* F_{ai}(0)$ of T_xM, and its coordinate lift, we have

$$d_x^2 c_t(X_\rho) = \widetilde{X}_\rho + \sum_{\nu\in D} X_\rho(X_\nu(c_t))\partial_\nu, \tag{1.83}$$

where we used greek letters as a shorthand for boxes of the Young diagram D. When ρ belongs to the first column of the Young diagram D, namely $\rho = a1$ (in this case, we simply write a), we have, as a consequence of Lemma 1.130 and the structural equations

$$F_a(0) = -\dot{E}_a(0) = -[\vec{H}, \partial_a] = \widetilde{X}_a + \sum_{\nu\in D}\left(\sum_{\kappa\in D} c_{a\nu}^\kappa h_\kappa + \sum_{b=1}^k h_b c_{b\nu}^a\right)\partial_\nu,$$

where everything is evaluated at λ. Therefore, from Eq. (1.83), we obtain

$$d_x^2 c_t(X_a) = F_a(0) + \sum_{\nu\in D}\left(X_a(X_\nu(c_t)) - \sum_{\kappa\in D} c_{a\nu}^\kappa h_\kappa - \sum_{b=1}^k h_b c_{b\nu}^a\right) E_\nu(0).$$

Recall that $S(t)^{-1} : \mathcal{H}_\lambda \to \mathcal{V}_\lambda$ is the matrix that represents the Jacobi curve in the coordinates induced by the canonical frame (at $t = 0$). More explicitly

$$d_x^2 c_t(X_\rho) = F_\rho(0) + \sum_{\nu\in D} S(t)^{-1}_{\rho\nu} E_\nu(0).$$

Moreover, since we restricted $d_x^2 c_t$ to elements of $\mathscr{D}_x$, we obtain

$$\sum_{a=1}^{k} X_a^2(c_t) = \sum_{a=1}^{k} S^\flat(t)^{-1}_{aa} + \sum_{a=1}^{k}\sum_{b=1}^{k} h_a c_{ab}^b. \tag{1.84}$$

Now observe that, if ρ *does not* belong to the first column of the Young diagram, we have

$$\dot{E}_\rho(0) = [\vec{H}, \partial_\rho] = \sum_{a=1}^{k}\sum_{\nu\in D} h_a c_{a\nu}^\rho E_\nu(0).$$

On the other hand, by the structural equations, $\dot{E}_\rho(0)$ is a vertical vector that does not have $E_\rho(0)$ components. Then, when ρ is not in the first column of D, $\sum_{a=1}^k h_a c_{a\rho}^\rho = 0$. Thus we rewrite Eq. (1.84) as

$$\sum_{a=1}^{k} X_a^2(c_t) = \sum_{a=1}^{k} S^\flat(t)^{-1}_{aa} + \sum_{a=1}^{k}\sum_{\rho\in D} h_a c_{a\rho}^\rho. \tag{1.85}$$

By taking the sum of Eq. (1.82) and Eq. (1.85), we obtain

$$\Delta_\mu c_t|_x = \sum_{a=1}^{k} S^\flat(t)^{-1}_{aa} + \langle \nabla_x c_t | \nabla_x g\rangle,$$

where we recall that the function g is implicitly defined (in a neighbourhood of γ) by $\mu = e^g \theta_1 \wedge \ldots \wedge \theta_n$. Remember that, at $x = \gamma(0)$, $\nabla_x c_t = \dot\gamma(0)$. Then

$$\Delta_\mu c_t|_x = \sum_{a=1}^{k} S^\flat(t)^{-1}_{aa} + \left.\frac{d}{dt}\right|_{t=0} g(\gamma(t)).$$

Remark 1.131. Observe that if $P_t := X_1(t) \wedge \ldots \wedge X_n(t) \in \bigwedge^n T_{\gamma(t)}M$ is the parallelotope whose edges are the elements of the frame $\{X_i(t)\}_{i=1}^n$, then $g(\gamma(t)) = \log|\mu(P_t)|$, that is the logarithm of the volume of the parallelotope P_t.

Thus, by replacing the results of Corollary 1.123 about the asymptotics of the reduced Jacobi curve, we obtain

$$\Delta_\mu c_t|_x = -\frac{\operatorname{tr}\mathcal{I}_\lambda}{t} + \dot g(0) + \frac{1}{3}\operatorname{tr}\mathcal{R}_\lambda t + O(t^2),$$

where $\dot g(0) := \left.\frac{d}{dt}\right|_{t=0} g(\gamma(t))$. Since $\mathfrak{f}_t = -tc_t$, we obtain

$$\Delta_\mu \mathfrak{f}_t|_x = \operatorname{tr}\mathcal{I}_\lambda - \dot g(0)t - \frac{1}{3}\operatorname{tr}\mathcal{R}_\lambda t^2 + O(t^3),$$

which is the sought expansion, valid for small t. □

Computation of the linear term

Recall that, for any equiregular smooth admissible curve $\gamma : [0,T] \to M$, the Lie derivative in the direction of the curve defines surjective linear maps

$$\mathcal{L}_{\mathsf{T}} : \mathscr{F}^i_{\gamma(t)}/\mathscr{F}^{i-1}_{\gamma(t)} \to \mathscr{F}^{i+1}_{\gamma(t)}/\mathscr{F}^i_{\gamma(t)}, \qquad i \geq 1,$$

as defined in Section 1.5.5. In particular, notice that $\mathcal{L}^i_{\mathsf{T}} : \mathscr{D}_{\gamma(t)} \to \mathscr{F}^{i+1}_{\gamma(t)}/\mathscr{F}^i_{\gamma(t)}$, for $i \geq 1$ is a well defined, surjective linear map from the distribution (see also point (iv) of Remark 1.24).

Lemma 1.132. *For $t \in [0,T]$, we recover the projections $X_{ai}(t) = e^{t\vec{H}}_* F_{ai}(t) \in T_{\gamma(t)}M$ as*

$$X_{ai}(t) = (-1)^{i-1}\mathcal{L}^{i-1}_{\mathsf{T}}(X_{a1}(t)) \bmod \mathscr{F}^{i-1}_{\gamma(t)}, \qquad a = 1,\dots,k, \quad i = 1,\dots,n_a.$$

Proof. Fix $a = 1,\dots,k$. For $i = 1$ the statement is trivial. Assume the statement to be true for $j \leq i$. Recall that we can see $F_{ai}|_{\lambda(t)} = e^{t\vec{H}}_* F_{ai}(t)$ as a field along the extremal $\lambda(t)$. Then, by the structural equations for the canonical frame, $X_{a(i+1)} = -\pi_*[\vec{H}, F_{ai}]$. A quick computation in terms of a coordinate lifted frame proves that

$$X_{a(i+1)}(t) = -[\mathsf{T}, X_{ai}]|_{\gamma(t)} \bmod \mathscr{F}^i_{\gamma(t)},$$

for an admissible extension T of $\dot\gamma$. Thus, by induction, we obtain the statement. □

Proof of Theorem 1.85. We consider equiregular distributions and ample geodesics γ that obey the growth condition

$$\dim \mathscr{F}^i_{\gamma(t)} = \dim \mathscr{D}^i, \qquad \forall\, i \geq 0. \tag{1.86}$$

We only need to compute explicitly the term $\dot g(0)$ of the asymptotic expansion in Theorem 1.C. Recall that, according to the proof of Theorem 1.C, the coefficient of the linear term is given by the following formula (see Remark 1.131)

$$\dot g(0) = \left.\frac{d}{dt}\right|_{t=0} \log|\mu(P_t)|,$$

where P_t is the parallelotope whose edges are the projections $\{X_{ai}(t)\}_{ai\in D}$ of the horizontal part of the canonical frame $X_{ai} = \pi_* \circ e^{t\vec{H}}_* F_{ai}(t) \in T_{\gamma(t)}M$, namely

$$P_t = \bigwedge_{ai\in D} X_{ai}(t). \tag{1.87}$$

By definition of canonical frame, Proposition 1.108, and the growth condition (1.86) we have that the elements $\{X_{ai}(t)\}_{ai\in D}$ are a frame along the curve $\gamma(t)$ adapted to the flag of the distribution. More precisely

$$\mathscr{D}^i_{\gamma(t)} = \operatorname{span}\{X_{aj}(t) \mid aj \in D,\ 1 \leq j \leq i\}.$$

By Lemma 1.132 we can write the adapted frame $\{X_{ai}\}_{ai \in D}$ in terms of the smooth linear maps $\mathcal{L}_{\mathsf{T}}$, and we obtain the following formula for the parallelotope

$$P_t = \bigwedge_{i=1}^{m} \bigwedge_{a_i=1}^{d_i} X_{a_i i}(t) = \bigwedge_{i=1}^{m} \bigwedge_{a_i=1}^{d_i} \mathcal{L}_{\mathsf{T}}^{i-1}(X_{a_i 1}(t)).$$

Then, a standard linear algebra argument and the very definition of Popp's volume leads to

$$|\mu(P_t)| = \sqrt{\prod_{i=1}^{m} \det M_i(t)},$$

where the smooth families of operators $M_i(t)$, for $i = 1, \ldots, m$ are the one defined in Eq. (1.43). This, together with Eq. (1.87) completes the computation of the linear term of Theorem 1.C for any ample geodesic satisfying the growth condition (1.86). □

Chapter 2

Comparison theorems for conjugate points in sub-Riemannian geometry

2.1 Introduction

Among the most celebrated results in Riemannian geometry, comparison theorems play a prominent role. These theorems allow to estimate properties of a manifold under investigation with the same property on the *model spaces* which, in the classical setting, are the simply connected manifolds with constant sectional curvature (the sphere, the Euclidean plane and the hyperbolic plane). The properties that may be investigated with these techniques are countless and include, among the others, the number of conjugate points along a given geodesic, the topology of loop spaces, the behaviour of volume of sets under homotheties, Laplacian comparison theorems, estimates for solutions of PDEs on the manifold, etc.

In this chapter we are concerned, in particular, with results of the following type. Until further notice, M is a Riemannian manifold, endowed with the Levi-Civita connection, and $\mathrm{Sec}(v, w)$ is the sectional curvature of the section $\mathrm{span}\{v, w\} \subset T_xM$.

Theorem 2.1. *Let $\gamma(t)$ be a unit-speed geodesic on M. If for all $t \geq 0$ and for all $v \in T_{\gamma(t)}M$ orthogonal to $\dot{\gamma}(t)$ with unit norm $\mathrm{Sec}(\dot{\gamma}(t), v) \geq k > 0$, then there exists $0 < t_c \leq \pi/\sqrt{k}$ such that $\gamma(t_c)$ is conjugate with $\gamma(0)$.*

Notice that the quadratic form $\mathrm{Sec}(\dot{\gamma}(t), \cdot) : T_{\gamma(t)}M \to \mathbb{R}$, which we call *directional curvature* (in the direction of $\dot{\gamma}$), computes the sectional curvature of the sections containing $\dot{\gamma}$. Theorem 2.1 compares the distance of the first conjugate point along γ with the same property computed on the sphere with sectional curvature $k > 0$, provided that the directional curvature along the geodesic on the reference manifold is bounded from below by k. Theorem 2.1 also contains all the basic ingredients of a comparison-type result:

- A micro-local condition, i.e. "along the geodesic", given in terms of a bound on curvature-type quantities, such as the sectional or Ricci curvature.
- Models for comparison, that is spaces in which the property under investigation can be computed explicitly.

As it is well known, Theorem 2.1 can be improved by replacing the bound on the directional curvature with a bound on the average, or Ricci curvature. Moreover, Theorem 2.1 leads immediately to the celebrated Bonnet-Myers theorem (see [57]).

In Riemannian geometry, the importance of conjugate points rests on the fact that geodesics cease to be minimizing after the first one. This remains true for strongly normal sub-Riemannian geodesics. Moreover, conjugate points, both in Riemannian and sub-Riemannian geometry, are also intertwined with the analytic properties of the underlying structure, for example they affect the behaviour of the heat kernel (see [22, 24] and references therein).

The main results of this chapter are comparison theorems on the existence of conjugate points, valid for any sub-Riemannian structure.

We briefly recall the concept of sub-Riemannian structure. A sub-Riemannian structure on a manifold M is defined by a distribution $\mathscr{D} \subseteq TM$ of constant rank, with a scalar product that, unlike the Riemannian case, is defined only for vectors in $\mathscr{D}$. Under mild assumptions on $\mathscr{D}$ (the Hörmander condition) any connected sub-Riemannian manifold is *horizontally* path-connected, namely any two points are joined by a path whose tangent vector belongs to $\mathscr{D}$.

Thus, a rich theory paralleling the classical Riemannian one can be developed, giving a meaning to the concept of geodesic, as an horizontal curve that locally minimises the length. Still, since in general there is no canonical completion of the sub-Riemannian metric to a Riemannian one, there is no way to define a connection à la Levi-Civita and thus the familiar Riemannian curvature tensor. The classical theory of Jacobi fields and its connection with the curvature plays a central role in the proof of many Riemannian comparison results, and the generalisation to the sub-Riemannian setting is not straightforward. The Jacobi equation itself, being defined in terms of the covariant derivative, cannot be formalised in the classical sense when a connection is not available.

In this chapter we focus on results in the spirit of Theorem 2.1 even tough there are no evident obstructions to the application of the same techniques, relying on the Riccati equations for sub-Riemannian geodesics, to other types of comparison results. We anticipate that the comparisons models will be linear quadratic optimal control problems (LQ problems in the following), i.e. minimization problems quite similar to the Riemannian one, where the length is replaced by a functional defined by a quadratic Lagrangian. More precisely one is interested in finding *admissible trajectories* of a *linear* control system in $\mathbb{R}^n$, namely curves $x : [0, t] \to \mathbb{R}^n$ for which there exists a control $u \in L^2([0, t], \mathbb{R}^k)$ such that

$$\dot{x} = Ax + Bu, \qquad x(0) = x_0, \qquad x(t) = x_1, \qquad x_0, x_1, t \text{ fixed},$$

that minimize a *quadratic* functional $\phi_t : L^2([0, t], \mathbb{R}^k) \to \mathbb{R}$ of the form

$$\phi_t(u) = \frac{1}{2} \int_0^t (u^* u - x^* Q x)\, dt.$$

Here A, B, Q are constant matrices of the appropriate dimension. The symmetric matrix Q is usually referred to as the *potential*. Notice that it makes sense to speak about *conjugate time* of a LQ problem: it is the time $t_c > 0$ at which extremal trajectories lose local optimality. It turns out that t_c does not depend on the data x_0, x_1, but it is an intrinsic feature of the problem.

These kind of structures are well known in the field of optimal control theory, but to our best knowledge this is the first time they are employed as model spaces for comparison results.

With any ample, equiregular sub-Riemannian geodesic $\gamma(t)$ (see Definition 2.12), we associate: its *Young diagram* D, a scalar product $\langle\cdot|\cdot\rangle_{\gamma(t)} : T_{\gamma(t)}M \times T_{\gamma(t)}M \mapsto \mathbb{R}$ extending the sub-Riemannian one and a quadratic form $\mathfrak{R}_{\gamma(t)} : T_{\gamma(t)}M \mapsto \mathbb{R}$ (the sub-Riemannian directional curvature), all depending on the geodesic $\gamma(t)$. We stress that, for a Riemannian manifold, any non-trivial geodesic has the same Young diagram, composed by a single column with $n = \dim M$ boxes, the scalar product $\langle\cdot|\cdot\rangle_{\gamma(t)}$ coincides with the Riemannian one, and $\mathfrak{R}_{\gamma(t)}(v) = \mathrm{Sec}(v, \dot\gamma(t))$ for all $v \in T_{\gamma(t)}M$.

In this introduction, when we associate with a geodesic $\gamma(t)$ its Young diagram D, we implicitly assume that $\gamma(t)$ is ample and equiregular. Notice that these assumptions are true for the generic geodesic, as we discuss more precisely in Sec. 2.2.2.

In the spirit of Theorem 2.1, assume that the sub-Riemannian directional curvature is bounded from below by a quadratic form Q. Then, we associate a model LQ problem (i.e. matrices A and B, depending on γ) which, roughly speaking, represents the linearisation of the sub-Riemannian structure along γ itself, with potential Q. We dub this *model space* $\mathrm{LQ}(D;Q)$, where D is the Young diagram of γ, and Q represents the bound on the sub-Riemannian directional curvature. The first of our results can be stated as follows (see Theorem 2.A).

Theorem 2.2. *Let $\gamma(t)$ be a sub-Riemannian geodesic, with Young diagram D, such that $Q_- \geq \mathfrak{R}_{\gamma(t)} \geq Q_+$ for all $t \geq 0$. Then the first conjugate point along $\gamma(t)$ occurs at a time t_c not greater than the first conjugate time of the model $\mathrm{LQ}(D;Q_+)$ and not smaller than the first conjugate time of $\mathrm{LQ}(D;Q_-)$.*

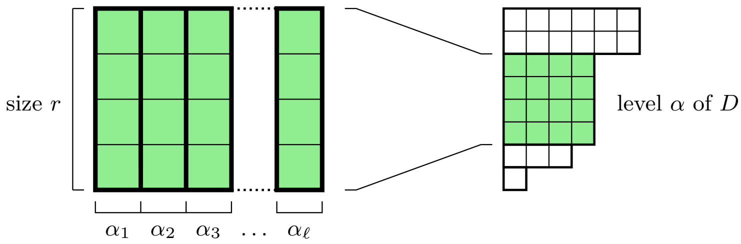
In the Riemannian case, any non-trivial geodesic γ has the same (trivial) Young diagram, and this leads to a simple LQ model with $A = 0$, $B = \mathbb{I}$. Moreover, $\langle\cdot|\cdot\rangle_\gamma$ is the Riemannian scalar product and $\mathfrak{R}_\gamma = \mathrm{Sec}(\dot\gamma, \cdot)$. Then, if Theorem 2.2 holds with $Q_+ = k\mathbb{I}$, the first conjugate point along the Riemannian geodesic, with directional curvature bounded by k occurs at a time t not greater than the first conjugate time of the LQ model

$$\dot x = u, \qquad \phi_t(u) = \frac{1}{2}\int_0^t \left(|u|^2 - k|x|^2\right) dt.$$

It is well known that, when $k > 0$, this problem represents a simple n-dimensional harmonic oscillator, whose extremal trajectories lose optimality at time $t_c = \pi/\sqrt{k}$. Thus we recover Theorem 2.1. In the sub-Riemannian setting, due to the intrinsic anisotropy of the structure different geodesics have different Young diagrams, resulting in a rich class of LQ models, with non-trivial drift terms. The directional sub-Riemannian curvature $\mathfrak{R}_{\gamma(t)}$ represents the potential "experienced" in a neighbourhood of the geodesic.

We stress that the generic $\mathrm{LQ}(D;Q)$ model may have infinite conjugate time. However, as we discuss in full detail in Chapter 3, there are necessary and sufficient conditions for its finiteness. Thus Theorem 2.2 can be employed to prove both existence or non-existence of conjugate points along a given geodesic.

As Theorem 2.1 can be improved by considering a bound on the Ricci curvature in the direction of the geodesic, instead of the whole sectional curvature, also Theorem 2.2 can be improved in the same spirit. In the sub-Riemannian case, however, the process of "taking the trace" is more delicate. Due to the anisotropy of the structure, it only makes sense to take *partial* traces, leading to a number of Ricci curvatures (each one obtained as a partial trace on an invariant subspace of $T_{\gamma(t)}M$, determined by the Young diagram D). In particular, for each *level* α of the Young diagram (namely the collection of all the rows with the same length equal to, say, ℓ) we have ℓ Ricci curvatures $\mathfrak{Ric}^{\alpha_i}_{\gamma(t)}$, for $i = 1, \dots, \ell$. The *size* of a level is the number r of boxes in each of its columns $\alpha_1, \dots, \alpha_\ell$.



The partial tracing process leads to our main result (see Theorem 2.B).

Theorem 2.3. *Let $\gamma(t)$ be a sub-Riemannian geodesic with Young diagram D. Consider a fixed level α of D, with length ℓ and size r. Then, if*

$$\frac{1}{r}\mathfrak{Ric}^{\alpha_i}_{\gamma(t)} \geq k_i, \qquad \forall i = 1, \dots, \ell, \quad \forall t \geq 0,$$

the first conjugate time $t_c(\gamma)$ along the geodesic satisfies $t_c(\gamma) \leq t_c(k_1, \dots, k_\ell)$.

In Theorem 2.3, $t_c(k_1, \dots, k_\ell)$ denotes the first conjugate time of the LQ model associated with a Young diagram with a single row, of length ℓ, and a diagonal potential $Q = \operatorname{diag}\{k_1, \dots, k_\ell\}$.

The hypotheses in Theorem 2.3 are no longer bounds on a quadratic form as in Theorem 2.2, but a finite number of *scalar* bounds. Observe that we have one comparison theorem for each level of the Young diagram of the given geodesic. In the Riemannian case, as we discussed earlier, D has only one level, of length $\ell = 1$, of size $r = \dim M$. In this case there is single Ricci curvature, namely $\mathfrak{Ric}^{\alpha_1}_{\gamma(t)} = \operatorname{Ric}(\dot\gamma(t))$ and, if $k_1 > 0$ in Theorem 2.3, $t_c(k_1) = \pi/\sqrt{k_1} < +\infty$. Back to the general case, we stress that in order to have $t_c(k_1, \dots, k_\ell) < +\infty$, the Riemannian condition $\operatorname{Ric}(\dot\gamma) \geq k_1 > 0$ must be replaced by more complicated inequalities on the bounds $k_1, \dots, k_\ell$ on the sub-Riemannian Ricci curvatures. In particular, we allow also for some *negative values* of such constants.

As an application of Theorem 2.3, we prove a sub-Riemannian version of the classical Bonnet-Myers theorem (see Theorem 2.C).

Theorem 2.4. *Let M be a connected, complete sub-Riemannian manifold, such that the generic geodesic has the same Young diagram D. Assume that there exists a level α of length ℓ and size r and constants $k_1, \dots, k_\ell$ such that, for any length parametrized geodesic $\gamma(t)$*

$$\frac{1}{r}\mathfrak{Ric}^{\alpha_i}_{\gamma(t)} \geq k_i, \qquad \forall i = 1, \dots, \ell, \quad \forall t \geq 0.$$

Then, if the polynomial

$$P_{k_1,\dots,k_\ell}(x) := x^{2\ell} - \sum_{i=0}^{\ell-1}(-1)^{\ell-i}k_{\ell-i}x^{2i}$$

has at least one simple purely imaginary root, the manifold is compact, has diameter not greater than $t_c(k_1,\dots,k_\ell) < +\infty$*. Moreover, its fundamental group is finite.*

In the Riemannian setting we recover the classical Bonnet-Myers theorem, since $\ell = 1$, $r = \dim M$ and the condition on the roots of $P_{k_1}(x) = x^2 + k_1$ is equivalent to $k_1 > 0$.

Finally we apply our techniques to obtain information about the conjugate time of geodesics on 3D *unimodular* Lie groups. Left-invariant structures on 3D Lie groups are the basic examples of sub-Riemannian manifolds and the study of such structures is the starting point to understand the general properties of sub-Riemannian geometry.

A complete classification of such structures, up to local sub-Riemannian isometries, is given in [5, Thm. 1], in terms of the two basic geometric invariants $\chi \geq 0$ and κ, that are constant for left-invariant structures. In particular, for each choice of the pair (χ, κ), there exists a unique unimodular group in this classification. Even if left-invariant structures possess the symmetries inherited by the group structure, the sub-Riemannian geodesics and their conjugate loci have been studied only in some particular cases where explicit computations are possible.

The conjugate locus of left-invariant structures has been completely determined for the cases corresponding to $\chi = 0$, that are the Heisenberg group [40] and the semisimple Lie groups $SU(2), SL(2)$ where the metric is defined by the Killing form [34]. On the other hand, when $\chi > 0$, only few cases have been considered up to now. In particular, to our best knowledge, only the sub-Riemannian structure on the group of motions of the Euclidean plane $SE(2)$, where $\chi = \kappa > 0$, has been considered [55, 65].

As an application of our results, we give a sufficient condition for the existence of finite conjugate times for geodesics in an unimodular Lie group, together with an estimate for it (non-sharp, in general). This condition is expressed in terms of a bound, depending on χ, k, on a constant of the motion $E(\gamma)$ associated with the given geodesic γ (see Theorem 2.D).

Theorem 2.5. *Let M be a 3D Lie group endowed with a contact left-invariant sub-Riemannian structure with invariants $\chi > 0$ and $\kappa \in \mathbb{R}$. Then there exists $\bar{E} = \bar{E}(\chi, \kappa)$ such that every length parametrized geodesic γ with $E(\gamma) \geq \bar{E}$ has a finite conjugate time.*

The cases corresponding to $\chi = 0$ are $\mathbb{H}$, $SU(2)$ and $SL(2)$, where $\kappa = 0, 1, -1$, respectively. For these structures we recover the exact estimates for the first conjugate time of a length parametrized geodesic (see Section 2.7.2).

The curvature employed in this chapter has been introduced for the first time by Agrachev and Gamkrelidze in [17], Agrachev and Zelenko in [16] and successively extended by Zelenko and Li in [71], where also the Young diagram is introduced for the first time in relation with the extremals of a variational problem. This research has been inspired by many recent works in this direction that we briefly review.

In [18] Agrachev and Lee investigate a generalisation of the measure contraction property (MCP) to 3D sub-Riemannian manifolds. The generalised MCP of Agrachev and Lee is expressed in terms of solutions of a particular 2D matrix Riccati equation for sub-Riemannian extremals, and this is one of the technical points that mostly inspired the present research.

In [48] Lee, Li and Zelenko pursue further progresses for sub-Riemannian contact structures with transversal symmetries. In this case, it is possible to exploit the Riemannian structure induced on the quotient space to write the curvature operator, and the authors recover sufficient condition for the contact manifold to satisfy the generalised MCP defined in [18]. Moreover, the authors perform the first step in the decoupling of the matrix Riccati equation for different levels of the Young diagram (see the splitting part of the proof of Theorem 2.B for more details).

The MCP for higher dimensional sub-Riemannian structures has also been investigated in [64] for Carnot groups.

We also mention that, in [49], Li and Zelenko prove comparison results for the number of conjugate points of curves in a Lagrange Grassmanian associated with sub-Riemannian structures with symmetries. In particular, [49, Cor. 4] is equivalent to Theorem 2.2, but obtained with differential topology techniques and with a different language. However, to our best knowledge, it is not clear how to obtain an averaged version of such comparison results with these techniques, and this is yet another motivation that led to Theorem 2.3.

In [29], Baudoin and Garofalo prove, with heat-semigroup techniques, a sub-Riemannian version of the Bonnet-Myers theorem for sub-Riemannian manifolds with transverse symmetries that satisfy an appropriate generalisation of the Curvature Dimension (CD) inequalities introduced in the same paper. In [30], Baudoin and Wang generalise the previous results to contact sub-Riemannian manifolds, removing the symmetries assumption. See also [27, 28] for other comparison results following from the generalised CD condition.

Even though in this chapter we discuss only sub-Riemannian structures, these techniques can be applied to the extremals of any affine optimal control problem, a general framework including (sub)-Riemannian, (sub)-Finsler manifolds, as discussed in [9]. For example, in [17], the authors prove a comparison theorem for conjugate points along extremals associated with *regular* Hamiltonian systems, such as those corresponding to Riemannian and Finsler geodesics. Finally, concerning comparison theorems for Finsler structures one can see, for example, [58, 60, 70].

2.1.1 Structure of the chapter

The plan of the chapter is as follows. In Sec. 2.2 we provide the basic definitions of sub-Riemannian geometry, and in particular the growth vector and the Young diagram of a sub-Riemannian geodesic. In Sec. 2.3 we revisit the theory of Jacobi fields. In Sec. 2.4 we introduce the main technical tool, that is the generalised matrix Riccati equation, and the appropriate comparison models. Then, in Sec. 2.5 we provide the "average" version of our comparison theorems, transitioning from sectional-curvature type results to Ricci-curvature type ones. In Sec. 2.6, as an application, we prove a sub-Riemannian Bonnet-Myers theorem. Finally, in Sec. 2.7, we apply our theorems to obtain some new results on conjugate points for 3D left-invariant sub-Riemannian structures.

2.2 Preliminaries

Let us recall some basic facts in sub-Riemannian geometry. We refer to [6] for further details.

Let M be a smooth, connected manifold of dimension $n \geq 3$. A sub-Riemannian structure on M is a pair $(\mathscr{D}, \langle\cdot|\cdot\rangle)$ where $\mathscr{D}$ is a smooth vector distribution of constant rank $k \leq n$ satisfying the *Hörmander condition* (i.e. $\mathrm{Lie}_x\mathscr{D} = T_xM$, $\forall x \in M$) and $\langle\cdot|\cdot\rangle$ is a smooth Riemannian metric on $\mathscr{D}$. A Lipschitz continuous curve $\gamma : [0,T] \to M$ is *horizontal* (or *admissible*) if $\dot\gamma(t) \in \mathscr{D}_{\gamma(t)}$ for a.e. $t \in [0,T]$. Given a horizontal curve $\gamma : [0,T] \to M$, the *length of* γ is

$$\ell(\gamma) = \int_0^T \|\dot\gamma(t)\| dt,$$

where $\|\cdot\|$ denotes the norm induced by $\langle\cdot|\cdot\rangle$. The *sub-Riemannian distance* is the function

$$\mathsf{d}(x,y) := \inf\{\ell(\gamma) \mid \gamma(0) = x, \gamma(T) = y, \gamma \text{ horizontal}\}.$$

The connectedness of M and the Hörmander condition guarantee the finiteness and the continuity of $d : M \times M \to \mathbb{R}$ with respect to the topology of M (Rashevsky-Chow theorem).

Example 2.6. A sub-Riemannian manifold of odd dimension is *contact* if $\mathscr{D} = \ker\omega$, where ω is a one-form and $d\omega|_{\mathscr{D}}$ is non degenerate. The *Reeb vector field* $X_0 \in \mathrm{Vec}(M)$ is the unique vector field such that $d\omega(X_0, \cdot) = 0$ and $\omega(X_0) = 1$.

Example 2.7. Let M be a Lie group, and $L_x : M \to M$ be the left translation by $x \in M$. A sub-Riemannian structure $(\mathscr{D}, \langle\cdot|\cdot\rangle)$ is *left-invariant* if $d_yL_x : \mathscr{D}_y \to \mathscr{D}_{L_xy}$ and is an isometry w.r.t. $\langle\cdot|\cdot\rangle$ for all $x, y \in M$. Any Lie group admits left invariant structures obtained by choosing a scalar product on its Lie algebra and transporting it on the whole M by left translation.

Locally, the pair $(\mathscr{D}, \langle\cdot|\cdot\rangle)$ can be given by assigning a set of k smooth vector fields that span $\mathscr{D}$, orthonormal for $\langle\cdot|\cdot\rangle$. In this case, the set $\{X_1, \dots, X_k\}$ is called a *local orthonormal frame* for the sub-Riemannian structure. Finally, we can write the system in "control form", namely for any horizontal curve $\gamma : [0,T] \to M$ there is a *control* $u \in L^\infty([0,T], \mathbb{R}^k)$ such that

$$\dot\gamma(t) = \sum_{i=1}^k u_i(t) X_i|_{\gamma(t)}, \qquad \text{a.e. } t \in [0,T].$$

2.2.1 Minimizers and geodesics

A sub-Riemannian *geodesic* is an admissible curve $\gamma : [0,T] \to M$ such that $\|\dot\gamma(t)\|$ is constant and for every sufficiently small interval $[t_1, t_2] \subseteq [0,T]$, the restriction $\gamma|_{[t_1,t_2]}$ minimizes the length between its endpoints. The length of a geodesic is invariant by reparametrization of the latter. Geodesics for which $\|\dot\gamma(t)\| = 1$ are called *length parametrized* (or of *unit speed*). A sub-Riemannian manifold is said to be *complete* if (M, d) is complete as a metric space.

With any sub-Riemannian structure we associate the Hamiltonian function $H \in C^\infty(T^*M)$

$$H(\lambda) = \frac{1}{2}\sum_{i=1}^k \langle\lambda, X_i\rangle^2, \qquad \forall \lambda \in T^*M,$$

in terms of any local frame $X_1, \dots, X_k$, where $\langle\lambda, \cdot\rangle$ denotes the action of the covector λ on vectors. Let σ be the canonical symplectic form on T^*M. With the symbol $\vec{a}$ we denote the

Hamiltonian vector field on T^*M associated with a function $a \in C^\infty(T^*M)$. Indeed $\vec{a}$ is defined by the formula $da = \sigma(\cdot, \vec{a})$. For $i = 1, \dots, k$ let $h_i \in C^\infty(T^*M)$ be the linear-on-fibers functions defined by $h_i(\lambda) := \langle \lambda, X_i \rangle$. Notice that

$$H = \frac{1}{2}\sum_{i=1}^k h_i^2, \qquad \vec{H} = \sum_{i=1}^k h_i \vec{h}_i.$$

Trajectories minimizing the distance between two points are solutions of first-order necessary conditions for optimality, which in the case of sub-Riemannian geometry are given by a weak version of the Pontryagin Maximum Principle ([61]), see also [6] for an elementary proof). We denote by $\pi : T^*M \to M$ the standard bundle projection.

Theorem 2.8. *Let $\gamma : [0,T] \to M$ be a sub-Riemannian geodesic associated with a non-zero control $u \in L^\infty([0,T], \mathbb{R}^k)$. Then there exists a Lipschitz curve $\lambda : [0,T] \to T^*M$, such that $\pi \circ \lambda = \gamma$ and only one of the following conditions holds for a.e. $t \in [0,T]$:*

(i) $\dot\lambda(t) = \vec{H}|_{\lambda(t)}$ *and* $h_i(\lambda(t)) = u_i(t)$,

(ii) $\dot\lambda(t) = \sum_{i=1}^k u_i(t)\vec{h}_i|_{\lambda(t)}$, $\lambda(t) \neq 0$ *and* $h_i(\lambda(t)) = 0$.

If $\lambda : [0,T] \to M$ is a solution of (i) (resp. (ii)) it is called a *normal* (resp. *abnormal*) *extremal*). It is well known that if $\lambda(t)$ is a normal extremal, then its projection $\gamma(t) := \pi(\lambda(t))$ is a smooth geodesic. This does not hold in general for abnormal extremals. On the other hand, a geodesic can be at the same time normal and abnormal, namely it admits distinct extremals, satisfying (i) and (ii). In the Riemannian setting there are no abnormal extremals.

Definition 2.9. A geodesic $\gamma : [0,T] \to M$ is *strictly normal* if it is not abnormal. It is called *strongly normal* if for every $t \in (0,T]$, the segment $\gamma|_{[0,t]}$ is not abnormal.

Notice that extremals satisfying (i) are simply integral lines of the Hamiltonian field $\vec{H}$. Thus, let $\lambda(t) = e^{t\vec{H}}(\lambda_0)$ denote the integral line of $\vec{H}$, with initial condition $\lambda(0) = \lambda_0$. The sub-Riemannian *exponential map* starting from x_0 is

$$\mathcal{E}_{x_0} : T^*_{x_0}M \to M, \qquad \mathcal{E}_{x_0}(\lambda_0) := \pi(e^{t\vec{H}}(\lambda_0)).$$

Unit speed normal geodesics correspond to initial covectors $\lambda_0 \in T^*M$ such that $H(\lambda) = 1/2$.

2.2.2 Geodesic flag and Young diagram

In this section we introduce a set of invariants of a sub-Riemannian geodesic, namely the geodesic flag, and a useful graphical representation of the latter: the Young diagram. The concept of Young diagram in this setting appeared for the first time in [71], as a fundamental invariant for curves in the Lagrange Grassmanian. The proof that the original definition in [71] is equivalent to forthcoming one can be found in [9, Sec. 6], in the general setting of affine control systems.

Let $\gamma(t)$ be a normal sub-Riemannian geodesic. By definition $\dot\gamma(t) \in \mathscr{D}_{\gamma(t)}$ for all times. Consider a smooth horizontal extension of the tangent vector, namely an horizontal vector field $\mathsf{T} \in \mathrm{Vec}_{\mathcal{H}}(M)$ such that $\mathsf{T}|_{\gamma(t)} = \dot\gamma(t)$.

Definition 2.10. The *flag of the geodesic* $\gamma(t)$ is the sequence of subspaces

$$\mathscr{F}^i_\gamma(t) := \operatorname{span}\{\mathcal{L}^j_{\mathsf{T}}(X)|_{\gamma(t)} \mid X \in \operatorname{Vec}_{\mathcal{H}}(M),\, j \le i-1\} \subseteq T_{\gamma(t)}M, \qquad i \ge 1,$$

where $\mathcal{L}_{\mathsf{T}}$ denotes the Lie derivative in the direction of T.

By definition, this is a filtration of $T_{\gamma(t)}M$, i.e. $\mathscr{F}^i_\gamma(t) \subseteq \mathscr{F}^{i+1}_\gamma(t)$, for all $i \ge 1$. Moreover, $\mathscr{F}^1_\gamma(t) = \mathscr{D}_{\gamma(t)}$. Definition 2.10 is well posed, namely does not depend on the choice of the horizontal extension T (see [9, Sec. 3.4]).

For each time t, the flag of the geodesic contains informations about how new directions can be obtained by taking the Lie derivative in the direction of the geodesic itself. In this sense it carries informations about the germ of the distribution along the given trajectory, and is the microlocal analogue of the flag of the distribution.

Definition 2.11. The *growth vector* of the geodesic $\gamma(t)$ is the sequence of integer numbers

$$\mathcal{G}_\gamma(t) := \{\dim \mathscr{F}^1_\gamma(t), \dim \mathscr{F}^2_\gamma(t), \ldots\}.$$

Notice that, by definition, $\dim \mathscr{F}^1_\gamma(t) = \dim \mathscr{D}_{\gamma(t)} = k$.

Definition 2.12. Let $\gamma(t)$ be a normal sub-Riemannian geodesic, with growth vector $\mathcal{G}_\gamma(t)$. We say that the geodesic is:

- *equiregular* if $\dim \mathscr{F}^i_\gamma(t)$ does not depend on t for all $i \ge 1$,
- *ample* if for all t there exists $m \ge 1$ such that $\dim \mathscr{F}^m_\gamma(t) = \dim T_{\gamma(t)}M$.

We stress that equiregular (resp. ample) geodesics are the microlocal counterpart of equiregular (resp. bracket-generating) distributions. Let $d_i := \dim \mathscr{F}^i_\gamma - \dim \mathscr{F}^{i-1}_\gamma$, for $i \ge 1$ be the increment of dimension of the flag of the geodesic at each step (with the convention $k_0 := 0$).

Lemma 2.13. *For an equiregular, ample geodesic,* $d_1 \ge d_2 \ge \ldots \ge d_m$.

Proof. By the equiregularity assumption, the Lie derivative $\mathcal{L}_{\mathsf{T}}$ defines surjective linear maps

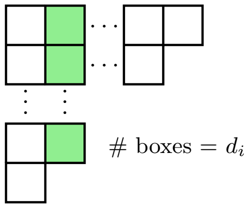
$$\mathcal{L}_{\mathsf{T}} : \mathscr{F}^i_\gamma(t)/\mathscr{F}^{i-1}_\gamma(t) \to \mathscr{F}^{i+1}_\gamma(t)/\mathscr{F}^i_\gamma(t), \qquad \forall t, \quad i \ge 1,$$

where we set $\mathscr{F}^0_\gamma(t) = \{0\}$. The quotients $\mathscr{F}^i_\gamma/\mathscr{F}^{i-1}_\gamma$ have constant dimension $d_i := \dim \mathscr{F}^i_\gamma - \dim \mathscr{F}^{i-1}_\gamma$. Therefore the sequence $d_1 \ge d_2 \ge \ldots \ge d_m$ is non-increasing. □

Notice that any ample geodesic is strongly normal, and for real-analytic sub-Riemannian structures also the converse is true (see [9, Prop. 3.11]). The generic geodesic is ample and equiregular. More precisely, the set of points $x \in M$ such there a exists non-empty Zariski open set $A_x \subseteq T^*_xM$ of initial covectors for which the associated geodesic is ample and equiregular with the same (maximal) growth vector, is open and dense in M. For more details, see [71, Sec. 5] and also [9, Lemma 5.17].

Young diagram

For an ample, equiregular geodesic, the sequence of dimension stabilises, namely $\dim \mathscr{F}_\gamma^m = \dim \mathscr{F}_\gamma^{m+j} = n$ for $j \geq 0$, and we write $\mathcal{G}_\gamma = \{\dim \mathscr{F}_\gamma^1, \dots, \dim \mathscr{F}_\gamma^m\}$. Thus, we associate with any ample, equiregular geodesic its Young diagram as follows. Recall that $d_i = \dim \mathscr{F}_\gamma^i - \dim \mathscr{F}_\gamma^{i-1}$ defines a decreasing sequence by Lemma 2.13. Then we can build a tableau D with m columns of length d_i, for $i = 1, \dots, m$, as follows:



Indeed $\sum_{i=1}^m d_i = n = \dim M$ is the total number of boxes in D. Let us discuss some examples.

Example 2.14. For a Riemannian structure, the flag of any non-trivial geodesic consists in a single space $\mathscr{F}_\gamma^1(t) = T_{\gamma(t)}M$. Therefore $\mathcal{G}_\gamma(t) = \{n\}$ and all the geodesics are ample and equiregular. Roughly speaking, all the directions have the same (trivial) behaviour w.r.t. the Lie derivative.

Example 2.15. Consider a contact, sub-Riemannian manifold with $\dim M = 2n+1$, and a non-trivial geodesic γ with tangent field $\mathsf{T} \in \mathrm{Vec}_{\mathcal{H}}(M)$. Let $X_1, \dots, X_{2n}$ be a local frame in a neighbourhood of the geodesic and X_0 the Reeb vector field. Let ω be the contact form. We define the invertible bundle map $J : \mathscr{D} \to \mathscr{D}$ by $\langle X | JY \rangle = d\omega(X, Y)$, for $X, Y \in \mathrm{Vec}_{\mathcal{H}}(M)$. Finally, we split $\mathscr{D} = J\mathsf{T} \oplus J\mathsf{T}^\perp$ along the geodesic $\gamma(t)$. We obtain

$$\mathcal{L}_{\mathsf{T}}(Y) = \langle J\mathsf{T} | Y \rangle X_0 \mod \mathrm{Vec}_{\mathcal{H}}(M), \qquad Y \in \mathrm{Vec}_{\mathcal{H}}(M).$$

Therefore, the Lie derivative of fields in $J\mathsf{T}^\perp$ does not generate "new directions". On the other hand, $\mathcal{L}_{\mathsf{T}}(J\mathsf{T}) = X_0$ up to elements in $\mathrm{Vec}_{\mathcal{H}}(M)$. In this sense, the subspaces $J\mathsf{T}$ and $J\mathsf{T}^\perp$ are different w.r.t. Lie derivative: the former generates new directions, the latter does not. In the Young diagram, the subspace $J\mathsf{T}^\perp$ corresponds to the rectangular sub-diagram D_2, while the subspace $J\mathsf{T} \oplus X_0$ corresponds to the rectangular sub-diagram D_1 in Fig. 2.1.b.

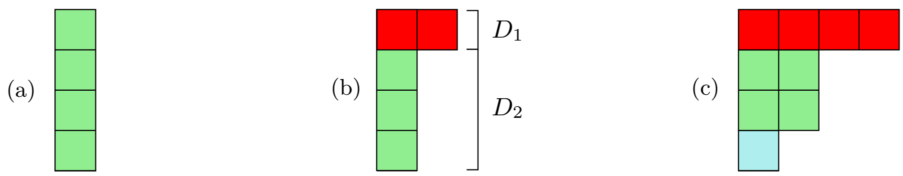


Figure 2.1: Young diagrams for (a) Riemannian, (b) contact, (c) a more general structure.

See Fig. 2.1 for some examples of Young diagrams. The number of boxes in the i-th row (i.e. d_i) is the number of new independent directions in $T_{\gamma(t)}M$ obtained by taking $(i-1)$-th Lie derivatives in the direction of T.

2.3 Jacobi fields revisited: conjugate points and Riccati equation

Let $\lambda \in T^*M$ be the covector associated with a strongly normal geodesic, projection of the extremal $\lambda(t) = e^{t\vec{H}}(\lambda)$. For any $\xi \in T_\lambda(T^*M)$ we define the field along the extremal $\lambda(t)$ as

$$X(t) := e_*^{t\vec{H}}\xi \in T_{\lambda(t)}(T^*M).$$

The set of vector fields obtained in this way is a $2n$-dimensional vector space, that we call *the space of Jacobi fields along the extremal*. In the Riemannian case, the projection π_* is an isomorphisms between the space of Jacobi fields along the extremal and the classical space of Jacobi fields along the geodesic γ. Thus, this definition is equivalent to the standard one in Riemannian geometry, does not need curvature or connection, and works for any strongly normal sub-Riemannian geodesic.

In Riemannian geometry, the study of one half of such a 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 $\lambda \in T^*M$, let $\mathcal{V}_\lambda := \ker \pi_*|_\lambda \subset T_\lambda(T^*M)$ be the *vertical subspace*. We define the family of Lagrangian subspaces along the extremal

$$\mathcal{L}(t) := e_*^{t\vec{H}}\mathcal{V}_\lambda \subset T_{\lambda(t)}(T^*M).$$

Definition 2.16. A time $t > 0$ is a *conjugate time* for γ if $\mathcal{L}(t) \cap \mathcal{V}_{\lambda(t)} \neq \{0\}$. Equivalently, we say that $\gamma(t) = \pi(\lambda(t))$ is a conjugate point w.r.t. $\gamma(0)$ along $\gamma(t)$. The *first conjugate time* is the smallest conjugate time, namely $t_c(\gamma) = \inf\{t > 0 \mid \mathcal{L}(t) \cap \mathcal{V}_{\lambda(t)} \neq \{0\}\}$.

Since the geodesic is strongly normal, the first conjugate time is separated from zero, namely there exists $\varepsilon > 0$ such that $\mathcal{L}(t) \cap \mathcal{V}_{\lambda(t)} = \{0\}$ for all $t \in (0, \varepsilon)$. Notice that conjugate points correspond to the critical values of the sub-Riemannian exponential map with base in $\gamma(0)$. In other words, if $\gamma(t)$ is conjugate with $\gamma(0)$ along γ, there exists a one-parameter family of geodesics starting at $\gamma(0)$ and ending at $\gamma(t)$ at first order. Indeed, let $\xi \in \mathcal{V}_\lambda$ such that $\pi_* \circ e_*^{t\vec{H}}\xi = 0$, then the vector field $\tau \mapsto \pi_* \circ e_*^{\tau\vec{H}}\xi$ is a classical Jacobi field along γ which vanishes at the endpoints, and this is precisely the vector field of the aforementioned variation.

In Riemannian geometry geodesics cease to be minimizing after the first conjugate time. This remains true for strongly normal sub-Riemannian geodesics (see, for instance, [6]).

2.3.1 Riemannian interlude

In this section, we recall the concept of parallely transported frame along a geodesic in Riemannian geometry, and we give an equivalent characterisation in terms of a Darboux moving frame along the corresponding extremal lift. Let $(M, \langle\cdot|\cdot\rangle)$ be a Riemannian manifold, endowed with the Levi-Civita connection $\nabla : \mathrm{Vec}(M) \to \mathrm{Vec}(M)$. In terms of a local orthonormal frame

$$\nabla_{X_j} X_i = \sum_{k=1}^n \Gamma_{ij}^k X_k, \qquad \Gamma_{ij}^k = \frac{1}{2}\left(c_{ij}^k + c_{ki}^j + c_{kj}^i\right),$$

where $\Gamma_{ij}^k \in C^\infty(M)$ are the Christoffel symbols written in terms of the orthonormal frame. Notice that $\Gamma_{ij}^k = -\Gamma_{ik}^j$.

Let $\gamma(t)$ be a geodesic and $\lambda(t)$ be the associated (normal) extremal, such that $\dot{\lambda}(t) = \vec{H}|_{\lambda(t)}$ and $\gamma(t) = \pi \circ \lambda(t)$. Let $\{X_1, \dots, X_n\}$ a parallely transported frame along the geodesic $\gamma(t)$, i.e. $\nabla_{\dot{\gamma}} X_i = 0$. Let $h_i : T^*M \to \mathbb{R}$ be the linear-on-fibers functions associated with X_i, defined by $h_i(\lambda) := \langle \lambda, X_i \rangle$. We define the (vertical) fields $\partial_{h_i} \in \mathrm{Vec}(T^*M)$ such that $\partial_{h_i}(\pi^* g) = 0$, and $\partial_{h_i}(h_j) = \delta_{ij}$ for any $g \in C^\infty(M)$ and $i, j = 1, \dots, n$. We define a moving frame along the extremal $\lambda(t)$ as follows

$$E_i := \partial_{h_i}, \qquad F_i := -[\vec{H}, E_i],$$

where the frame is understood to be evaluated at $\lambda(t)$. Notice that we can recover the parallely transported frame by projection, namely $\pi_* F_i|_{\lambda(t)} = X_i|_{\gamma(t)}$ for all i. In the following, for any vector field Z along an extremal $\lambda(t)$ we employ the shorthand

$$\dot{Z}|_{\lambda(t)} := \left.\frac{d}{d\varepsilon}\right|_{\varepsilon=0} e_*^{-\varepsilon \vec{H}} Z|_{\lambda(t+\varepsilon)} = [\vec{H}, Z]|_{\lambda(t)}$$

to denote the vector field along $\lambda(t)$ obtained by taking the Lie derivative in the direction of $\vec{H}$ of any smooth extension of Z. Notice that this is well defined, namely its value at $\lambda(t)$ does not depend on the choice of the extension. We state the properties of the moving frame in the following proposition.

Proposition 2.17. *The smooth moving frame $\{E_i, F_i\}_{i=1}^n$ has the following properties:*

(i) $\mathrm{span}\{E_i|_{\lambda(t)}\} = \mathcal{V}_{\lambda(t)}$.

(ii) It is a Darboux basis, namely

$$\sigma(E_i, E_j) = \sigma(F_i, F_j) = \sigma(E_i, F_j) - \delta_{ij} = 0, \qquad i, j = 1, \dots, n.$$

(iii) The frame satisfies structural equations

$$\dot{E}_i = -F_i, \qquad \dot{F}_i = \sum_{j=1}^n R_{ij}(t) E_j,$$

for some smooth family of $n \times n$ symmetric matrices $R(t)$.

Properties (i)-(iii) uniquely define the moving frame up to orthogonal transformations. More precisely if $\{\widetilde{E}_i, \widetilde{F}_j\}_{i=1}^n$ is another smooth moving frame along $\lambda(t)$ satisfying (i)-(iii), with some matrix $\widetilde{R}(t)$ then there exist a constant, orthogonal matrix O such that

$$\widetilde{E}_i|_{\lambda(t)} = \sum_{j=1}^n O_{ij} E_j|_{\lambda(t)}, \qquad \widetilde{F}_i|_{\lambda(t)} = \sum_{j=1}^n O_{ij} F_j|_{\lambda(t)}, \qquad \widetilde{R}(t) = O R(t) O^*. \tag{2.2}$$

A few remarks are in order. Property (ii) implies that $\mathrm{span}\{E_1,\dots,E_n\}$, $\mathrm{span}\{F_1,\dots,F_n\}$, evaluated at $\lambda(t)$, are Lagrangian subspaces of $T_{\lambda(t)}(T^*M)$. Eq. (2.2) reflects the fact that a parallely transported frame is defined up to constant orthogonal transformations. In particular, one could use properties (i)-(iii) to *define* the parallel transport along $\gamma(t)$ by $X_i|_{\gamma(t)} := \pi_* F_i|_{\lambda(t)}$. Finally, the symmetric matrix $R(t)$ induces a well defined quadratic form $\mathfrak{R}_{\gamma(t)} : T_{\gamma(t)}M \to \mathbb{R}$

$$\mathfrak{R}_{\gamma(t)}(v) := \sum_{i,j=1}^{n} R_{ij}(t) v_i v_j, \qquad v = \sum_{i=1}^{n} v_i X_i|_{\gamma(t)} \in T_{\gamma(t)}M.$$

Indeed Proposition 2.17 implies that the definition of $\mathfrak{R}_{\gamma(t)}$ does not depend on the choice of the parallely transported frame.

Lemma 2.18. *Let* $R^\nabla : \mathrm{Vec}(M) \times \mathrm{Vec}(M) \times \mathrm{Vec}(M) \to \mathrm{Vec}(M)$ *the Riemannian curvature tensor w.r.t. the Levi-Civita connection. Then*

$$\mathfrak{R}_\gamma(v) = \langle R^\nabla(v,\dot\gamma)\dot\gamma | v\rangle, \qquad v \in T_\gamma M,$$

where we suppressed the explicit dependence on time.

In other words, for any unit vector $v \in T_\gamma M$, $\mathfrak{R}_\gamma(v) = \mathrm{Sec}(v,\dot\gamma)$ is the sectional curvature of the plane generated by v and $\dot\gamma$, i.e. the *directional curvature* in the direction of the geodesic. The proof of Proposition 2.17 and Lemma 2.18 can be found in Appendix I.

2.3.2 Canonical frame

The concept of Levi-Civita connection and covariant derivative is not available for general sub-Riemannian structures, and it is not clear how to parallely transport a frame along a sub-Riemannian geodesic. Nevertheless, in [71], the authors introduce a parallely transported frame along the corresponding extremal $\lambda(t)$ which, in the spirit of Proposition 2.17, generalises the concept of parallel transport also to (sufficiently regular) sub-Riemannian extremals.

Consider an ample, equiregular geodesic, with Young diagram D, with k rows, of length $n_1,\dots,n_k$. Indeed $n_1+\dots+n_k = n$. The moving frame we are going to introduce is indexed by the boxes of the Young diagram, so we fix some terminology first. Each box is labelled "ai", where $a = 1,\dots,k$ is the row index, and $i = 1,\dots,n_a$ is the progressive box number, starting from the left, in the specified row. Briefly, the notation $ai \in D$ denotes the generic box of the diagram. We employ letters from the beginning of the alphabet $a,b,c,\dots$ for rows, and letters from the middle of the alphabet $i,j,h,\dots$ for the position of the box in the row.

We collect the rows with the same length in D, and we call them *levels* of the Young diagram. In particular, a level is the union of r rows $D_1,\dots,D_r$, and r is called the *size* of the level. The set of all the boxes $ai \in D$ that belong to the same column and the same level of D is called *superbox*. We use greek letters $\alpha,\beta,\dots$ to denote superboxes. Notice that that two boxes ai, bj are in the same superbox if and only if ai and bj are in the same column of D and in possibly distinct row but with same length, i.e. if and only if $i = j$ and $n_a = n_b$. See Fig. 2.2 for examples of levels and superboxes for Riemannian, contact and more general structures.

Theorem 2.19 (See [71])**.** *There exists a smooth moving frame* $\{E_{ai},F_{ai}\}_{ai\in D}$ *along the extremal* $\lambda(t)$ *such that*

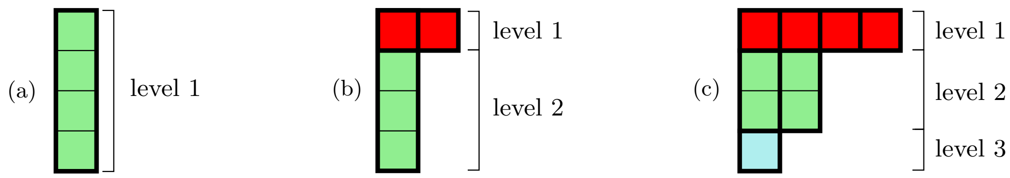


Figure 2.2: Levels (shaded regions) and superboxes (delimited by bold lines) for the Young diagram of (a) Riemannian, (b) contact, (c) a more general structure. The Young diagram for any Riemannian geodesic has a single level and a single superbox. The Young diagram of any contact sub-Riemannian geodesic has levels two levels containing 2 and 1 superboxes, respectively. The Young diagram (c) has three levels with 4, 2, 1 superboxes, respectively.

(i) $\mathrm{span}\{E_{ai}|_{\lambda(t)}\} = \mathcal{V}_{\lambda(t)}$.

(ii) It is a Darboux basis, namely

$$\sigma(E_{ai}, E_{bj}) = \sigma(F_{ai}, F_{bj}) = \sigma(E_{ai}, F_{bj}) = \delta_{ab}\delta_{ij}, \qquad ai, bj \in D.$$

(iii) The frame satisfies structural equations

$$\begin{cases} \dot{E}_{ai} = E_{a(i-1)} & a = 1, \dots, k, \quad i = 2, \dots, n_a, \\ \dot{E}_{a1} = -F_{a1} & a = 1, \dots, k, \\ \dot{F}_{ai} = \sum_{bj \in D} R_{ai,bj}(t) E_{bj} - F_{a(i+1)} & a = 1, \dots, k, \quad i = 1, \dots, n_a - 1, \\ \dot{F}_{an_a} = \sum_{bj \in D} R_{bj,an_a}(t) E_{bj} & a = 1, \dots, k, \end{cases} \tag{2.3}$$

for some smooth family of $n \times n$ symmetric matrices $R(t)$, with components $R_{ai,bj}(t) = R_{bj,ai}(t)$, indexed by the boxes of the Young diagram D. The matrix $R(t)$ is normal *in the sense of [71].*

Properties (i)-(iii) uniquely define the frame up to orthogonal transformation that preserve the Young diagram. More precisely, if $\{\widetilde{E}_{ai}, \widetilde{F}_{ai}\}_{ai \in D}$ is another smooth moving frame along $\lambda(t)$ satisfying i)-iii), with some normal matrix $\widetilde{R}(t)$, then for any superbox α of size r there exists an orthogonal (constant) $r \times r$ matrix O^α such that

$$\widetilde{E}_{ai} = \sum_{bj \in \alpha} O^\alpha_{ai,bj} E_{bj}, \qquad \widetilde{F}_{ai} = \sum_{bj \in \alpha} O^\alpha_{ai,bj} F_{bj}, \qquad ai \in \alpha.$$

Theorem 2.19 implies that the following objects are well defined:

- The scalar product $\langle \cdot | \cdot \rangle_{\gamma(t)}$, depending on $\gamma(t)$, such that the fields $X_{ai}|_{\gamma(t)} := \pi_* F_{ai}|_{\lambda(t)}$ along $\gamma(t)$ are an orthonormal frame.
- A splitting of $T_{\gamma(t)}M$, orthogonal w.r.t. $\langle \cdot | \cdot \rangle_{\gamma(t)}$

$$T_{\gamma(t)}M = \bigoplus_\alpha S^\alpha_{\gamma(t)}, \qquad S^\alpha_{\gamma(t)} := \mathrm{span}\{X_{ai}|_{\gamma(t)} \mid ai \in \alpha\},$$

where the sum is over the superboxes α of D. Notice that the dimension of $S^\alpha_{\gamma(t)}$ is equal to the size r of the level in which the superbox α is contained.

- The sub-Riemannian directional curvature, defined as the quadratic form $\mathfrak{R}_{\gamma(t)} : T_{\gamma(t)}M \to \mathbb{R}$ whose representative matrix, in terms of an orthonormal frame $\{X_{ai}\}_{ai\in D}$ is $R_{ai,bj}(t)$.
- For each superbox α, the *sub-Riemannian Ricci curvatures*

$$\mathfrak{Ric}^{\alpha}_{\gamma(t)} := \sum_{ai\in\alpha} \mathfrak{R}_{\gamma(t)}(X_{ai}),$$

  which is precisely the partial trace of $\mathfrak{R}_{\gamma(t)}$, identified through the scalar product with an operator on $T_{\gamma(t)}M$, on the subspace $S^{\alpha}_{\gamma(t)} \subseteq T_{\gamma(t)}M$.

In this sense, each superbox α in the Young diagram corresponds to a well defined subspace $S^{\alpha}_{\gamma(t)}$ of $T_{\gamma(t)}M$. Notice that, for Riemannian structures, the Young diagram is trivial with n rows of length 1, there is a single superbox, Theorem 2.19 reduces to Proposition 2.17, the scalar product $\langle\cdot|\cdot\rangle_{\gamma(t)}$ reduces to the Riemannian product computed along the geodesic $\gamma(t)$, the orthogonal splitting is trivial, the directional curvature is the sectional curvature of the planes containing $\dot\gamma(t)$ and there is only one Ricci curvature.

A compact form for the structural equations

We rewrite system (2.3) in a compact form. In the sequel it will be convenient to split a frame $\{E_{ai}, F_{ai}\}_{ai\in D}$ in subframes, relative to the rows of the Young diagram. For $a = 1,\dots,k$, the symbol E_a denotes the n_a-dimensional row vector

$$E_a = (E_{a1}, E_{a2}, \dots, E_{an_a}),$$

with analogous notation for F_a. Similarly, E denotes the n-dimensional row vector

$$E = (E_1, \dots, E_k),$$

and similarly for F. Let $\Gamma_1 = \Gamma_1(D), \Gamma_2 = \Gamma_2(D)$ be $n \times n$ matrices, depending on the Young diagram D, defined as follows: for $a, b = 1,\dots,k$, $i = 1,\dots,n_a$, $j = 1,\dots,n_b$, we set

$$(\Gamma_1)_{ai,bj} := \delta_{ab}\delta_{i,j-1}, \tag{2.4}$$

$$(\Gamma_2)_{ai,bj} := \delta_{ab}\delta_{i1}\delta_{j1}. \tag{2.5}$$

It is convenient to see Γ_1 and Γ_2 as block diagonal matrices, the a-th block on the diagonal being a $n_a \times n_a$ matrix with components $\delta_{i,j-1}$ and $\delta_{i1}\delta_{j1}$, respectively (see also Eq. (2.13)). Notice that Γ_1 is nilpotent and Γ_2 is idempotent. Then, we rewrite the system (2.3) as follows

$$\begin{pmatrix} \dot E & \dot F \end{pmatrix} = \begin{pmatrix} E & F \end{pmatrix} \begin{pmatrix} \Gamma_1 & R(t) \\ -\Gamma_2 & -\Gamma_1^* \end{pmatrix}.$$

By exploiting the structural equations, we write a linear differential equation in $\mathbb{R}^{2n}$ that rules the evolution of the Jacobi fields along the extremal.

2.3.3 Linearized Hamiltonian

Let $\xi \in T_\lambda(T^*M)$ and $X(t) := e_*^{t\vec{H}}\xi$ be the associated Jacobi field along the extremal. In terms of any moving frame $\{E_{ai}, F_{ai}\}_{ai\in D}$ along $\lambda(t)$, it has components $(p(t), x(t)) \in \mathbb{R}^{2n}$, namely

$$X(t) = \sum_{ai\in D} p_{ai}(t) E_{ai}|_{\lambda(t)} + x_{ai}(t) F_{ai}|_{\lambda(t)}.$$

If we choose the canonical frame, using the structural equations, we obtain that the coordinates of the Jacobi field satisfy the following system of linear ODEs:

$$\begin{pmatrix} \dot{p} \\ \dot{x} \end{pmatrix} = \begin{pmatrix} -\Gamma_1 & -R(t) \\ \Gamma_2 & \Gamma_1^* \end{pmatrix} \begin{pmatrix} p \\ x \end{pmatrix}. \tag{2.6}$$

In this sense, the canonical frame is a tool to write the linearisation of the Hamiltonian flow along the geodesic in a canonical form. The r.h.s. of Eq. (2.6) is the “linearised Hamiltonian vector field”, written in its normal form (see also Eq. (3.1)). The linearised Hamiltonian field is, in general, non-autonomous. Notice also that the canonical form of the linearisation depends on the Young diagram D (through the matrices Γ_1 and Γ_2) and the curvature matrix $R(t)$.

In the Riemannian case, $D = \square$ for any geodesic, $\Gamma_1 = 0$, $\Gamma_2 = \mathbb{I}$ and we recover the classical Jacobi equation, written in terms of an orthonormal frame along the geodesic

$$\ddot{x} + R(t)x = 0.$$

2.3.4 Riccati equation: blowup time and conjugate time

Now we study, with a single matrix equation, the space of Jacobi fields along the extremal associated with an ample, equiregular geodesic. We write the generic element of $\mathcal{L}(t)$ in terms of the frame along the extremal. Let $E_{\lambda(t)}, F_{\lambda(t)}$ be row vectors, whose entries are the elements of the frame. The action of $e_*^{t\vec{H}}$ is meant entry-wise. Then

$$\mathcal{L}(t) \supset e_*^{t\vec{H}} E_{\lambda(0)} = E_{\lambda(t)} M(t) + F_{\lambda(t)} N(t),$$

for some smooth families $M(t), N(t)$ of $n \times n$ matrices. Notice that

$$M(0) = \mathbb{I}, \qquad N(0) = 0, \qquad \det N(t) \neq 0 \text{ for } t \in (0, \varepsilon).$$

The first $t > 0$ such that $\det N(t) = 0$ is indeed the first conjugate time. By using once again the structural equations, we obtain the following system of linear ODEs:

$$\frac{d}{dt}\begin{pmatrix} M \\ N \end{pmatrix} = \begin{pmatrix} -\Gamma_1 & -R(t) \\ \Gamma_2 & \Gamma_1^* \end{pmatrix} \begin{pmatrix} M \\ N \end{pmatrix}.$$

The solution of the Cauchy problem with the initial datum $M(0) = \mathbb{I}$, $N(0) = 0$ is defined on the whole interval on which $R(t)$ is defined. The columns of the $2n \times n$ matrix $\left(\begin{smallmatrix} M \\ N \end{smallmatrix}\right)$ are the components of Jacobi fields along the extremal w.r.t. the given frame, and they generate the n-dimensional subspace of Jacobi fields $X(t)$ along the extremal $\lambda(t)$ such that $\pi_* X(0) = 0$.

Since, for small $t > 0$, $\mathcal{L}(t) \cap \mathcal{L}(0) = \{0\}$, we have that

$$\mathcal{L}(t) = \operatorname{span}\{F_{\lambda(t)} + E_{\lambda(t)} V(t)\}, \qquad t > 0,$$

where $V(t) := M(t)N(t)^{-1}$ is well defined and smooth for $t > 0$ until the first conjugate time. Since $\mathcal{L}(t)$ is a Lagrangian subspace and the canonical frame is Darboux, $V(t)$ is a symmetric matrix. Moreover it satisfies the following Riccati equation:

$$\dot{V} = -\Gamma_1 V - V\Gamma_1^* - R(t) - V\Gamma_2 V. \tag{2.7}$$

We characterize $V(t)$ as the solution of a Cauchy problem with limit initial condition.

Lemma 2.20. *The matrix $V(t)$ is the unique solution of the Cauchy problem*

$$\dot{V} = -\Gamma_1 V - V\Gamma_1^* - R(t) - V\Gamma_2 V, \qquad \lim_{t\to 0^+} V^{-1} = 0, \tag{2.8}$$

in the sense that $V(t)$ is the unique solution such that $V(t)$ is invertible for small $t > 0$ and $\lim_{t\to 0^+} V(t)^{-1} = 0$.

Proof. As we already observed, $V(t)$ satisfies Eq. (2.7). Moreover $V(t)$ is invertible for $t > 0$ small enough, $V(t)^{-1} = N(t)M(t)^{-1}$ and $\lim_{t\to 0^+} V^{-1} = 0$. The uniqueness follows from the well-posedness of the limit Cauchy problem. See Lemma G.1 in Appendix G. □

It is well known that the solutions of Riccati equations are not, in general, defined for all t, but they may blow up at finite time. The next proposition relates the occurrence of such blow up time with the first conjugate point along the geodesic.

Proposition 2.21. *Let $V(t)$ the unique solution of* (2.8)*, defined on its maximal interval $I \subseteq (0, +\infty)$. Let $t_c := \inf\{t > 0 | \, \mathcal{L}(t) \cap \mathcal{V}_{\lambda(t)} = \{0\}\}$ be the first conjugate point along the geodesic. Then $I = (0, t_c)$.*

Proof. First, we prove that $I \supseteq (0, t_c)$. For any $t \in (0, t_c)$, $\mathcal{L}(t)$ is transversal to $\mathcal{V}_{\lambda(t)}$. Then the matrix $N(t)$ is non-degenerate for all $t \in (0, t_c)$. Then $V(t) := M(t)N(t)^{-1}$ is the solution of (2.8), and $I \supseteq (0, t_c)$.

On the other hand, let $V(t)$ be the maximal solution of (2.8), and let $t \in I$. Then the family of symplectic subspaces $\widetilde{\mathcal{L}}(t) := \operatorname{span}\{F_{\lambda(t)} + E_{\lambda(t)} V(t)\}$ is a family transversal to the vertical bundle, namely $\widetilde{\mathcal{L}}(t) \cap V_{\lambda(t)} = \{0\}$ for all $t \in I$. It is possible to show, following the argument of [46, Ch. 8] that the evolution of $\widetilde{\mathcal{L}}(t)$ is ruled by the Hamiltonian flow, namely $\widetilde{\mathcal{L}}(t+s) = e_*^{s\vec{H}} \widetilde{\mathcal{L}}(t)$. Then, since $\widetilde{\mathcal{L}}(\varepsilon) = \mathcal{L}(\varepsilon)$, we have that $\widetilde{\mathcal{L}}(t) = \mathcal{L}(t)$ and $I \subseteq (0, t_c)$. □

Proposition 2.21 states that the problem of finding the first conjugate time is equivalent to the study of the blow up time of the Cauchy problem (2.8) for the Riccati equation.

2.4 Microlocal comparison theorem

In Sec. 2.3, we reduced the problem of finding the conjugate points along an ample, equiregular sub-Riemannian geodesic to the study of the blow-up time of the solution of the Cauchy problem

$$\dot{V} + \Gamma_1 V + V\Gamma_1^* + R(t) + V\Gamma_2 V = 0, \qquad \lim_{t\to 0^+} V^{-1} = 0.$$

It is well known that the same equation controls the conjugate times of a LQ optimal control problems, defined by appropriate matrices A, B, Q, where $A = \Gamma_1^*$, $BB^* = \Gamma_2$, and the potential Q replaces $R(t)$. In this sense, for what concerns the study of conjugate points, LQ problems represent the natural *constant curvature models.*

2.4.1 LQ optimal control problems

Linear quadratic optimal control problems (LQ in the following) are a classical topic in control theory. They consist in a linear control system with a cost given by a quadratic Lagrangian. We briefly recall the general features of a LQ problem, and we refer to [15, Ch. 16] and [46, Ch. 7] for further details. We are interested in *admissible trajectories*, namely curves $x : [0, t] \to \mathbb{R}^n$ for which there exists a control $u \in L^2([0, t], \mathbb{R}^k)$ such that

$$\dot{x} = Ax + Bu, \qquad x(0) = x_0, \qquad x(t) = x_1, \qquad x_0, x_1, t \text{ fixed}, \tag{2.9}$$

that minimize a quadratic functional $\phi_t : L^2([0, t], \mathbb{R}^k) \to \mathbb{R}$ of the form

$$\phi_t(u) = \frac{1}{2}\int_0^t (u^* u - x^* Q x)\, dt. \tag{2.10}$$

Here A, B, Q are constant matrices of the appropriate dimension. The vector Ax represents the *drift*, while the columns of B are the controllable directions. The meaning of the *potential* term Q will be clear later, when we will introduce the Hamiltonian of the LQ problem.

We only deal with *controllable* systems, i.e. we assume that there exists $m > 0$ such that

$$\operatorname{rank}(B, AB, \dots, A^{m-1}B) = n.$$

This hypothesis implies that, for any choice of t, x_0, x_1, the set of controls u such that the associated trajectory $x_u : [0, t] \to \mathbb{R}^n$ connects x_0 with x_1 in time t is not empty.

It is well known that the optimal trajectories of the LQ system are projections $(p, x) \mapsto x$ of the solutions of the Hamiltonian system

$$\dot{p} = -\partial_x H(p, x), \qquad \dot{x} = \partial_p H(p, x), \qquad (p, x) \in T^*\mathbb{R}^n = \mathbb{R}^{2n},$$

where the Hamiltonian function $H : \mathbb{R}^{2n} \to \mathbb{R}$ is defined by

$$H(p, x) = \frac{1}{2}\begin{pmatrix} p^* & x^* \end{pmatrix}\begin{pmatrix} BB^* & A \\ A^* & Q \end{pmatrix}\begin{pmatrix} p \\ x \end{pmatrix}.$$

We denote by $P_t : \mathbb{R}^{2n} \to \mathbb{R}^{2n}$ the flow of the Hamiltonian system, which is defined for all $t \in \mathbb{R}$. We employ canonical coordinates (p, x) on $T^*\mathbb{R}^n = \mathbb{R}^{2n}$ such that the symplectic form is written

$\sigma = \sum_{i=1}^n dp_i \wedge dx_i$. The flow lines of P_t are the integral lines of the *Hamiltonian vector field* $\vec{H} \in \mathrm{Vec}(\mathbb{R}^{2n})$, defined by $dH(\cdot) = \sigma(\,\cdot\,, \vec{H})$. More explicitly

$$\vec{H}_{(p,x)} = \begin{pmatrix} -A^* & -Q \\ BB^* & A \end{pmatrix} \begin{pmatrix} p \\ x \end{pmatrix}. \tag{2.11}$$

We stress that not all the integral lines of the Hamiltonian flow lead to minimizing solutions of the LQ problem, since they only satisfy first order conditions for optimality. Sufficiently short segments, however, are optimal, but they lose optimality at some time $t > 0$, called the *first conjugate time*.

Definition 2.22. We say that t is a conjugate time if there exists a solution of the Hamiltonian equations such that $x(0) = x(t) = 0$.

The first conjugate time determines existence and uniqueness of minimizing solutions of the LQ problem, as specified by the following proposition (see [15, Sec. 16.4]).

Proposition 2.23. *Let t_c be the first conjugate time of the LQ problem* (2.9)-(2.10)

- *For $t < t_c$, for any x_0, x_1 there exists a unique minimizer connecting x_0 with x_1 in time t.*
- *For $t > t_c$, for any x_0, x_1 there exists no minimizer connecting x_0 with x_1 in time t.*

The first conjugate time can be also characterised in terms of blow-up time of a matrix Riccati equation. Consider the vector subspace of solutions of Hamilton equations such that $x(0) = 0$. A basis of such a space is given by the solutions $(p_i(t), x_i(t))$ with initial condition $p_i(0) := e_i$, $x_i(0) = 0$, where e_i, for $i = 1, \dots, n$ is the standard basis of $\mathbb{R}^n$. Consider the matrices M, N, whose columns are the vectors $p_i(t)$ and $x_i(t)$, respectively. They solve the following equation:

$$\frac{d}{dt}\begin{pmatrix} M \\ N \end{pmatrix} = \begin{pmatrix} -A^* & -Q \\ BB^* & A \end{pmatrix} \begin{pmatrix} M \\ N \end{pmatrix},$$

where $M(0) = \mathbb{I}$ and $N(0) = 0$. Under the controllability condition, $N(t)$ is non-singular for $t > 0$ sufficiently small. By definition, the first conjugate time of the LQ problem is the first $t > 0$ such that $N(t)$ is singular. Thus, consider $V(t) := M(t)N(t)^{-1}$. The matrix $V(t)$ is symmetric and is the unique solution of the following Cauchy problem with limit initial condition:

$$\dot{V} + A^*V + VA + VBB^*V = 0, \qquad \lim_{t \to 0^+} V^{-1} = 0.$$

Thus we have the following characterization of the first conjugate time of the LQ problem.

Lemma 2.24. *The maximal interval of definition of the unique solution of the Cauchy problem*

$$\dot{V} + A^*V + VA + VBB^*V = 0, \qquad \lim_{t \to 0^+} V^{-1} = 0,$$

is $I = (0, t_c)$, where t_c is the first conjugate time of the associated LQ optimal control problem.

The same characterisation holds also for conjugate points along sub-Riemannian geodesics (see Proposition 2.21), and in this sense LQ problems provide models for computing conjugate times along sub-Riemannian geodesics.

2.4.2 Constant curvature models

Let D be a Young diagram associated with some ample, equiregular geodesic, and let $\Gamma_1 = \Gamma_1(D)$, $\Gamma_2 = \Gamma_2(D)$ the matrices defined in Sec. 2.2. Let Q be a symmetric $n \times n$ matrix.

Definition 2.25. We denote by $\mathrm{LQ}(D;Q)$ the *constant curvature model*, associated with a Young diagram D and constant curvature equal to Q, defined by the LQ problem with Hamiltonian

$$H(p,x) = \frac{1}{2}\left(p^*BB^*p + 2p^*Ax + x^*Qx\right), \qquad A = \Gamma_1^*, \quad BB^* = \Gamma_2.$$

We denote by $t_c(D;Q) \leq +\infty$ the first conjugate time of $\mathrm{LQ}(D;Q)$.

Remark 2.26. Indeed there are many matrices B such that $BB^* = \Gamma_2$, namely LQ problems with the same Hamiltonian, but their first conjugate time is the same. In particular, without loss of generality, one may choose $B = BB^* = \Gamma_2$.

In general, it is not trivial to deduce whether $t_c(D;Q) < +\infty$ or not, and this will be crucial in our comparison theorems. Nevertheless we have the following result in terms of the representative matrix of the Hamiltonian vector field $\vec{H}$ given by Eq. (2.11) (see Theorem 3.A or [14]).

Theorem 2.27. *The following dichotomy holds true for a controllable LQ optimal control system:*

- *If $\vec{H}$ has at least one odd-dimensional Jordan block corresponding to a pure imaginary eigenvalue, the number of conjugate times in $[0,T]$ grows to infinity for $T \to \pm\infty$.*
- *If $\vec{H}$ has no odd-dimensional Jordan blocks corresponding to a pure imaginary eigenvalue, there are no conjugate times.*

Thus, it is sufficient to put the Hamiltonian vector field $\vec{H}$ of $\mathrm{LQ}(D;Q)$, given by

$$\vec{H} \simeq \begin{pmatrix} -\Gamma_1 & -Q \\ \Gamma_2 & \Gamma_1^* \end{pmatrix},$$

in its Jordan normal form, to obtain necessary and sufficient condition for the finiteness of the first conjugate time.

Example 2.28. If D is the Young diagram associated with a Riemannian geodesic, with a single column with $n = \dim M$ boxes (or, equivalently, one single level with 1 superbox), $\Gamma_1 = 0$, $\Gamma_2 = \mathbb{I}$, and $\mathrm{LQ}(D;k\mathbb{I})$ is given by

$$H(p,x) = \frac{1}{2}\left(|p|^2 + k|x|^2\right),$$

which is the Hamiltonian of an harmonic oscillator (for $k > 0$), a free particle (for $k = 0$) or an harmonic repulsor (for $k < 0$). Extremal trajectories satisfy $\ddot{x} + kx = 0$. Moreover

$$t_c(D;k\mathbb{I}) = \begin{cases} \frac{\pi}{\sqrt{k}}, & k > 0 \\ +\infty & k \leq 0. \end{cases}$$

Indeed, for $k > 0$, all extremal trajectories starting from the origin are periodic, and they return to the origin at $t = \pi/\sqrt{k}$. On the other hand, for $k \leq 0$, all trajectories escape at least linearly from the origin, and we cannot have conjugate times (small variations of any extremal spread at least linearly for growing time). In this case, the Hamiltonian vector field $\vec{H}$ of $\mathrm{LQ}(D; k\mathbb{I})$ has characteristic polynomial $P(\lambda) = (\lambda^2 + k)^n$. Therefore Theorem 2.27 correctly gives that the first conjugate time is finite if and only if $k > 0$.

Example 2.29. For any Young diagram D, consider the model $\mathrm{LQ}(D; 0)$. Indeed in this case all the eigenvalues of $\vec{H}$ vanish. Thus, by Theorem 2.27, one has $t_c(D; Q) = +\infty$.

In the following, when considering average comparison theorems, we will consider a particular class of models, that we discuss in the following example.

Example 2.30. Let $D = \square \dots \square$ be a Young diagram with a single row of length ℓ, and $Q = \mathrm{diag}\{k_1, \dots, k_\ell\}$. We denote these special LQ models simply $\mathrm{LQ}(k_1, \dots, k_\ell)$.

In the case $\ell = 2$, Theorem 2.27 says that $t_c(k_1, k_2) < +\infty$ if and only if

$$\begin{cases} k_1 > 0, \\ k_2 > -k_1^2/4, \end{cases} \qquad \text{or} \qquad \begin{cases} k_1 \leq 0, \\ k_2 > 0. \end{cases}$$

In particular, by explicit integration of the Hamiltonian flow, one can compute that, if $k_1 > 0$ and $k_2 = 0$, the first conjugate time of $\mathrm{LQ}(k_1, 0)$ is $t_c(k_1, 0) = 2\pi/\sqrt{k_1}$.

2.4.3 General microlocal comparison theorem

We are now ready to prove the main result on estimates for conjugate times in terms of the constant curvature models $\mathrm{LQ}(D; Q)$.

Theorem 2.A. *Let $\gamma(t)$ be an ample, equiregular geodesic, with Young diagram D. Let $\mathfrak{R}_{\gamma(t)} : T_{\gamma(t)}M \to \mathbb{R}$ be directional curvature in the direction of the geodesic. Then, if*

$$Q_- \geq \mathfrak{R}_{\gamma(t)} \geq Q_+, \tag{2.12}$$

for some constant quadratic forms $Q_\pm : \mathbb{R}^n \to \mathbb{R}$, the first conjugate time $t_c(\gamma)$ along the geodesic satisfies

$$t_c(D; Q_-) \leq t_c(\gamma) \leq t_c(D; Q_+),$$

where, in Eq. (2.12), we understand the identification of $T_{\gamma(t)}M \simeq \mathbb{R}^n$ through any orthonormal basis for the scalar product $\langle \cdot | \cdot \rangle_{\gamma(t)}$.

In particular, since $t_c(D; 0) = +\infty$ (see Example 2.29), we have the following corollary.

Corollary 2.31. *Let $\gamma(t)$ be an ample, equiregular geodesic, with Young diagram D. Let $\mathfrak{R}_{\gamma(t)} : T_{\gamma(t)}M \to \mathbb{R}$ be directional curvature in the direction of the geodesic. Then, if $\mathfrak{R}_{\gamma(t)} \leq 0$, there are no conjugate points along the geodesic.*

In other words, the first conjugate times of $\mathrm{LQ}(D; Q)$ gives an estimate for the first conjugate time along geodesics with directional curvature $\mathfrak{R}_{\gamma(t)}$ controlled by Q.

Remark 2.32. There is no curvature along the direction of motion, that is $\mathfrak{R}_{\gamma(t)}(\dot\gamma(t)) = 0$. As it is well known in Riemannian geometry, it is possible to "take out the direction of the motion", considering the restriction of $\mathfrak{R}_{\gamma(t)}$ to the orthogonal complement of $\dot\gamma(t)$, with respect to $\langle\cdot|\cdot\rangle_{\gamma(t)}$, effectively reducing the dimension by one. To simplify the discussion, we do not go into such details since there is no variation with respect to the classical Riemannian case.

Remark 2.33. These microlocal theorems apply very nicely to geodesics in the Heisenberg group. In this example we have both geodesics with $\mathfrak{R}_{\gamma(t)} = 0$ (the straight lines) and geodesics with $\mathfrak{R}_{\gamma(t)} > 0$ (all the others). The former do not have conjugate times (by Theorem 2.31), while the latter do all have a finite conjugate time (by Theorem 2.A). For more details see Section 2.7.

Proof of Theorem 2.A. By Proposition 2.21, the study of the first conjugate time is reduced to the study of the blowup time of the solutions of the Riccati equation. We compare the solution of the Cauchy problem (2.8) for the matrix $V(t)$ for our extremal:

$$\dot V = -\begin{pmatrix}\mathbb{I} & V\end{pmatrix}\begin{pmatrix}R(t) & \Gamma_1\\ \Gamma_1^* & \Gamma_2\end{pmatrix}\begin{pmatrix}\mathbb{I}\\ V\end{pmatrix}, \qquad \lim_{t\to 0^+} V^{-1} = 0,$$

and the analogous solution $V_{D;Q}$ for any normal extremal of the model $\mathrm{LQ}(D;Q_\pm)$:

$$\dot V_{D;Q_\pm} = -\begin{pmatrix}\mathbb{I} & V_{D;Q_\pm}\end{pmatrix}\begin{pmatrix}Q_\pm & \Gamma_1\\ \Gamma_1^* & \Gamma_2\end{pmatrix}\begin{pmatrix}\mathbb{I}\\ V_{D;Q_\pm}\end{pmatrix}, \qquad \lim_{t\to 0^+} V_{D;Q_\pm}^{-1} = 0.$$

By Lemma G.1 in Appendix G, both solutions are well defined and positive definite for $t > 0$ sufficiently small. By hypothesis, $Q_- \geq R(t) \geq Q_+$. Therefore

$$-\begin{pmatrix}Q_+ & \Gamma_1\\ \Gamma_1^* & \Gamma_2\end{pmatrix} \geq -\begin{pmatrix}R(t) & \Gamma_1\\ \Gamma_1^* & \Gamma_2\end{pmatrix} \geq -\begin{pmatrix}Q_- & \Gamma_1\\ \Gamma_1^* & \Gamma_2\end{pmatrix}.$$

Moreover, by definition, $\lim_{t\to 0^+} V_{D;Q_\pm}^{-1}(t) = \lim_{t\to 0^+} V^{-1}(t) = 0$. Therefore, by Riccati comparison techniques (Theorem F.3 in Appendix F), we obtain

$$V_{D;Q_+}(t) \geq V(t) \geq V_{D;Q_-}(t),$$

for all $t > 0$ such that both solutions are defined. We need the following lemma.

Lemma 2.34. *For any D and Q, the solution $V_{D;Q}$ is monotone non-increasing.*

Proof of Lemma 2.34. It is a general fact that any solution of the symmetric Riccati differential equation with constant coefficients is monotone (see [1, Thm. 4.1.8]). In other words, for any solution $X(t)$ of a Cauchy problem with a Riccati equation with constant coefficients

$$\dot X + A^*X + XA + XQX = 0, \qquad \dot X(t_0) = X_0,$$

we have that $\dot X \geq 0$ (for $t \geq t_0$, where defined) if and only if $\dot X(t_0) \geq 0$ (true also with reversed and/or strict inequalities). Thus, in order to complete the proof of the lemma, it only suffices to compute the sign of $\dot V_{D;Q}(\varepsilon)$. This is easily done by exploiting the relationship with the inverse matrix. Observe that $\dot W_{D;Q}(0) = \Gamma_2 \geq 0$. Then $W_{D;Q}(t)$ is monotone non-decreasing. In particular $\dot W_{D;Q}(\varepsilon) \geq 0$. This, together with the fact that $W_{D;Q}(\varepsilon) > 0$ for ε sufficiently small (see Appendix G), implies that $\dot V_{D;Q}(\varepsilon) \leq 0$, and the lemma is proved. □

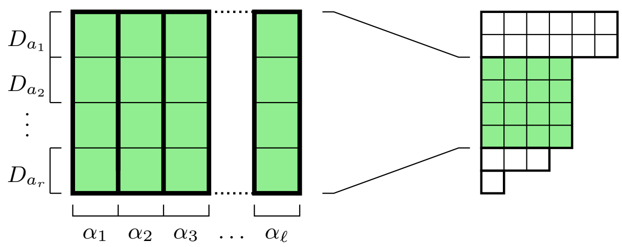


Figure 2.3: Detail of a single level of D of length ℓ and size r. It consists of the rows $D_{a_1}, \dots, D_{a_r}$, each one of length ℓ. The sets of boxes in each column are the superboxes $\alpha_1, \dots, \alpha_\ell$.

Now we conclude the proof. If $V_{D;Q_\pm}(t)$ blows up at $\bar{t}_\pm := t_c(D; Q_\pm) > 0$, it must tend to $-\infty$, namely $\lim_{t\to\bar{t}_\pm} V_{D;Q_\pm}(t) = -\infty$. Therefore $V(t) \leq V_{D;Q_+}(t)$ must blow up at a time not greater than $\bar{t}_+$. Analogously, $V(t) \geq V_{D;Q_-}(t)$ can blow up only at time greater than $\bar{t}_-$. □

2.5 Average microlocal comparison theorem

In this section we prove the average version of Theorem 2.A. Recall that, with any ample, equiregular geodesic $\gamma(t)$ we associate its Young diagram D. The latter is partitioned in levels, namely the sets of rows with the same length. Let $\alpha_1, \dots, \alpha_\ell$ be the superboxes in some given level, of length ℓ. The size r of the level is the number of rows contained in the level (see Fig. 2.3). To the superboxes α_i we associated the Ricci curvatures $\mathfrak{Ric}^{\alpha_i}_{\gamma(t)}$ for $i = 1, \dots, \ell$. Finally, we recall the definition anticipated in Example 2.30.

Definition 2.35. With the symbol $\mathrm{LQ}(k_1, \dots, k_\ell)$ we denote the LQ model associated with the Young diagram D with a single row of length ℓ, and with diagonal potential $Q = \mathrm{diag}(k_1, \dots, k_\ell)$. With the symbol $t_c(k_1, \dots, k_\ell)$ we denote the first conjugate time of $\mathrm{LQ}(k_1, \dots, k_\ell)$.

Theorem 2.B. *Let $\gamma(t)$ be an ample, equiregular geodesic, with Young diagram D. Let $\alpha_1, \dots, \alpha_\ell$ be the superboxes in some fixed level, of length ℓ and size r. Then, if*

$$\frac{1}{r}\,\mathfrak{Ric}^{\alpha_i}_{\gamma(t)} \geq k_i, \qquad \forall i = 1, \dots, \ell, \quad \forall t \geq 0,$$

the first conjugate time $t_c(\gamma)$ along the geodesic satisfies $t_c(\gamma) \leq t_c(k_1, \dots, k_\ell)$.

The hypotheses in Theorem 2.B are no longer bounds on a quadratic form, but a finite number of *scalar* bounds. Observe that we have one comparison theorem for each level of the Young diagram of the given geodesic.

Consider the Young diagram of any geodesic of a Riemannian structure. It consists of a single level of length $\ell = 1$, with one superbox α, of size $r = n = \dim M$ and $\mathfrak{Ric}^{\alpha}_{\gamma(t)} = \mathrm{Ric}(\dot\gamma(t))$. This, together with the computation of $t_c(k)$ of Example 2.28, recovers the following well known result.

Corollary 2.36. *Let $\gamma(t)$ be a Riemannian geodesic, such that $\mathrm{Ric}(\dot\gamma) \geq nk > 0$. Then the first conjugate time $t_c(\gamma)$ along the geodesic satisfies $t_c(\gamma) \leq \pi/\sqrt{k}$.*

Corollary 2.36 can be refined by taking out the direction of the motion, effectively reducing the dimension by 1. A similar reduction can be performed in Theorem 2.B, in the case of a "Riemannian" level of length 1 and size r, effectively reducing the size of by one. We do not go into details, since such a reduction can be obtained exactly as in the Riemannian case (see [69, Chapter 14] and also Remark 2.32).

We recall how the averaging procedure is carried out in Riemannian geometry. In this setting, one considers the average of the diagonal elements of $V(t)$, namely the trace, and employs the Cauchy-Schwarz inequality to obtain a scalar Riccati equation for $\operatorname{tr} V(t)$, where the curvature matrix is replaced by its trace, namely the Ricci curvature in the direction of the geodesic. On the other hand, in the sub-Riemannian setting, non-trivial terms containing matrices $\Gamma_1(D)$ and $\Gamma_2(D)$ appear in the Riccati equation. These terms, upon tracing, cannot be controlled in terms of $\operatorname{tr} V(t)$ alone. The failure of such a procedure in genuine sub-Riemannian manifolds is somehow expected: different directions have a different "behaviour", according to the structure of the Young diagram, and it makes no sense to average over all of them. The best we can do is to average among the directions corresponding to the rows of D that have the same length, namely rows in the same level. The proof of Theorem 2.B is based on the following two steps.

Splitting: The idea is to split the Cauchy problem

$$\dot{V} + \Gamma_1 V + V\Gamma_1^* + R(t) + V\Gamma_2 V = 0, \qquad \lim_{t\to 0^+} V^{-1} = 0,$$

in several, lower-dimensional Cauchy problems for particular blocks of $V(t)$. In these equations, only some blocks of $R(t)$ appear. In particular, we obtain one Riccati equation for each row of the Young diagram D, of dimension equal to the length of the row. The blow up of a block of $V(t)$ imples a blow up time for $V(t)$. Therefore, the presence of finite blow up time in any one of these lower dimensional blocks implies a conjugate time for the original problem.

Tracing: After the splitting step, we sum the Riccati equations corresponding to the rows with the same length, since all these equations are, in some sense, compatible (they have the same Γ_1, Γ_2 matrices). In the Riemannian case, this procedure leads to a single, scalar Riccati equation. In the sub-Riemannian case, we obtain one Riccati equation for each level of the Young diagram, of dimension equal to the length ℓ of the level. In this case the curvature matrix is replaced by a diagonal matrix, whose diagonal elements are the Ricci curvatures of the superboxes $\alpha_1, \dots, \alpha_\ell$ in the given level. This leads to a finite number of scalar conditions.

Proof of Theorem 2.B. We split the blocks of the Riccati equation corresponding to the rows of the Young diagram D, with k rows $D_1, \dots, D_k$, of length $n_1, \dots, n_k$. Recall that the matrices $\Gamma_1(D)$, $\Gamma_2(D)$, defined in Eqs. (2.4)-(2.5), are $n \times n$ block diagonal matrices

$$\Gamma_i(D) := \begin{pmatrix} \Gamma_i(D_1) & & \\ & \ddots & \\ & & \Gamma_i(D_k) \end{pmatrix}, \qquad i = 1, 2,$$

the a-th block being the $n_a \times n_a$ matrices

$$\Gamma_1(D_a) := \begin{pmatrix} 0 & \mathbb{I}_{n_a-1} \\ 0 & 0 \end{pmatrix}, \qquad \Gamma_2(D_a) := \begin{pmatrix} 1 & 0 \\ 0 & 0_{n_a-1} \end{pmatrix}, \tag{2.13}$$

where $\mathbb{I}_m$ is the $m \times m$ identity matrix and 0_m is the $m \times m$ zero matrix. Consider the maximal solution of the Cauchy problem

$$\dot{V} + \Gamma_1 V + V\Gamma_1^* + R(t) + V\Gamma_2 V = 0, \qquad \lim_{t\to 0^+} V^{-1} = 0.$$

The blow up of a block of $V(t)$ implies a finite blow up time for the whole matrix, hence a conjugate time. Thus, consider $V(t)$ as a block matrix. In particular, in the notation of Sec. 2.3, the block ab, denoted $V_{ab}(t)$ for $a, b = 1, \dots, k$, is a $n_a \times n_b$ matrix with components $V_{ai,bj}(t)$, $i = 1, \dots, n_a$, $j = 1, \dots, n_b$. Let us focus on the diagonal blocks

$$V(t) = \begin{pmatrix} V_{11}(t) & & * \\ & \ddots & \\ * & & V_{kk}(t) \end{pmatrix}.$$

Consider the equation for the a-th block on the diagonal, which we call $V_{aa}(t)$, and is a $n_a \times n_a$ matrices with components $V_{ai,aj}(t)$, $i, j = 1, \dots, n_a$. We obtain

$$\dot{V}_{aa} + \Gamma_1 V_{aa} + V_{aa}\Gamma_1^* + \widetilde{R}_{aa}(t) + V_{aa}\Gamma_2 V_{aa} = 0,$$

where $\Gamma_i = \Gamma_i(D_a)$, for $i = 1, 2$ are the matrix in Eq. (2.13), i.e. the a-th diagonal blocks of the matrices $\Gamma_i(D)$. Moreover

$$\widetilde{R}_{aa}(t) = R_{aa}(t) + \sum_{b\neq a} V_{ab}(t)\Gamma_2(D_b)V_{ba}(t).$$

The ampleness assumption implies the following limit condition for the block V_{aa}.

Lemma 2.37. $\lim_{t\to 0^+} (V_{aa})^{-1} = 0.$

Proof. Without loss of generality, consider the first block V_{11}. We partition the matrix V and $W = V^{-1}$ in blocks as follows

$$V = \begin{pmatrix} V_{11} & V_{10} \\ V_{10}^* & V_{00} \end{pmatrix},$$

where the index "0" collects all indices different from 1. By block-wise inversion, $W_{11} = (V^{-1})_{11} = (V_{11} - V_{10}V_{00}^{-1}V_{10}^*)^{-1}$. By Lemma G.1 in Appendix G, for small $t > 0$, $V(t) > 0$, hence $V_{00} > 0$ as well. Therefore $V_{11} - (W_{11})^{-1} = V_{10}V_{00}^{-1}V_{10}^* \geq 0$. Thus $V_{11} \geq (W_{11})^{-1} > 0$ and, by positivity, $0 < (V_{11})^{-1} \leq W_{11}$ for small $t > 0$. By taking the limit for $t \to 0^+$, since $W_{11} \to 0$, we obtain the statement. □

We proved that the block $V_{aa}(t)$ is solution of the Cauchy problem

$$\dot{V}_{aa} + \Gamma_1 V_{aa} + V_{aa}\Gamma_1^* + \widetilde{R}_{aa}(t) + V_{aa}\Gamma_2 V_{aa} = 0, \qquad \lim_{t\to 0^+}(V_{aa})^{-1} = 0. \tag{2.14}$$

The crucial observation is the following (see [48] for the original argument in the contact case with symmetries). Since $\Gamma_2(D_b) \geq 0$ and $V_{ba} = V_{ab}^*$ for all $a, b = 1, \ldots, k$, we obtain

$$\widetilde{R}_{aa}(t) = R_{aa}(t) + \sum_{b\neq a} V_{ab}(t)\Gamma_2(D_b)V_{ab}^*(t) \geq R_{aa}(t). \tag{2.15}$$

We now proceed with the second step of the proof, namely tracing over the level. Consider Eq. (2.14) for the diagonal blocks of $V(t)$, with $\widetilde{R}_{aa}(t) \geq R_{aa}(t)$. Now, we average over all the rows in the same level α. Let ℓ be the length of the level, namely $\ell = n_a$, for any row $D_{a_1}, \ldots, D_{a_r}$ in the given level (see Fig. 2.3). Then define the $\ell \times \ell$ symmetric matrix:

$$V_\alpha := \frac{1}{r}\sum_{a\in\alpha} V_{aa},$$

where the sum is taken on the indices $a \in \{a_1, \ldots, a_r\}$ of the rows D_a in the given level α. Once again, the blow up of $V_\alpha(t)$ implies also a blow up for $V(t)$. A computation shows that V_α is the solution of the following Cauchy problem

$$\dot{V}_\alpha + \Gamma_1 V_\alpha + V_\alpha\Gamma_1^* + R_\alpha(t) + V_\alpha\Gamma_2 V_\alpha = 0, \qquad \lim_{t\to 0^+} V_\alpha = 0,$$

where $\Gamma_2 = \Gamma_2(D_a)$ for any $a \in \alpha$, and the $\ell \times \ell$ matrix $R_\alpha(t)$ is defined by

$$\begin{aligned} R_\alpha(t) &:= \frac{1}{r}\sum_{a\in\alpha}\widetilde{R}_{aa}(t) + \frac{1}{r}\sum_{a\in\alpha} V_{aa}\Gamma_2 V_{aa} - V_\alpha\Gamma_2 V_\alpha \\ &= \frac{1}{r}\sum_{a\in\alpha}\widetilde{R}_{aa}(t) + \frac{1}{r}\left[\sum_{a\in\alpha}(V_{aa}\Gamma_2)(V_{aa}\Gamma_2)^* - \frac{1}{r}\left(\sum_{a\in\alpha}V_{aa}\Gamma_2\right)\left(\sum_{a\in\alpha}V_{aa}\Gamma_2\right)^*\right]. \end{aligned}$$

The key observation is that the term in square brackets is non-negative, as a consequence of the following lemma, whose proof is in Appendix H.

Lemma 2.38. *Let $\{X_a\}_{a=1}^r$, $\{Y_a\}_{a=1}^r$ be two sets of $\ell \times \ell$ matrices. Then*

$$\left(\sum_{a=1}^r X_a^* Y_a\right)\left(\sum_{b=1}^r X_b^* Y_b\right)^* \leq \left\|\sum_{a=1}^r Y_a^* Y_a\right\| \sum_{b=1}^r X_b^* X_b. \tag{2.16}$$

Here $\|\cdot\|$ denotes the operator norm.

Remark 2.39. Lemma 2.38 is a generalisation of the Cauchy-Schwarz inequality, in which the scalar product in $\mathbb{R}^r$ is replaced by a non-commutative product $\odot : \mathrm{Mat}(\ell)^r \times \mathrm{Mat}(\ell)^r \to \mathrm{Mat}(\ell)$, such that, if $X = \{X_a\}_{a=1}^r$, $Y = \{Y_a\}_{a=1}^r$, the product $X \odot Y := \sum_{a=1}^r X_a^* Y_a$. Eq. (2.16) becomes

$$(X \odot Y)(X \odot Y)^* \leq \|Y \odot Y\| X \odot X.$$

Then the l.h.s. of Eq. 2.16 is just the "square of the scalar product". For $\ell = 1$, we recover the classical Cauchy-Schwarz inequality.

We apply Lemma 2.38 to $X_a = \Gamma_2 V_{aa}$ and $Y_a = \Gamma_2$, for $a \in \alpha = \{a_1, \dots, a_r\}$. We obtain

$$\sum_{a\in\alpha}(V_{aa}\Gamma_2)(V_{aa}\Gamma_2)^* - \frac{1}{r}\left(\sum_{a\in\alpha} V_{aa}\Gamma_2\right)\left(\sum_{a\in\alpha} V_{aa}\Gamma_2\right)^* \geq 0,$$

which implies, together with Eq. (2.15)

$$R_\alpha(t) \geq \frac{1}{r}\sum_{a\in\alpha}\widetilde{R}_{aa}(t) \geq \frac{1}{r}\sum_{a\in\alpha} R_{aa}(t). \tag{2.17}$$

Notice that the ij-th component of the sum in the r.h.s. of Eq. (2.17) is $\frac{1}{r}\sum_{a\in\alpha} R_{ai,aj}(t)$, where $i,j = 1,\dots,\ell$. Thus, for any two fixed indices i,j we are considering, in coordinates, the trace of the restriction $\mathfrak{R}_{\gamma(t)} : S^{\alpha_i}_{\gamma(t)} \to S^{\alpha_j}_{\gamma(t)}$, written in terms of any orthonormal basis for $(T_{\gamma(t)}M, \langle\cdot|\cdot\rangle_{\gamma(t)})$. The matrix $R(t)$ is normal (see Theorem 2.19). Thus, according to [71], such a trace is always zero, unless $i = j$. Thus only the diagonal elements are non-vanishing and

$$\frac{1}{r}\sum_{a\in\alpha} R_{aa}(t) = \frac{1}{r}\begin{pmatrix} \mathfrak{Ric}^{\alpha_1}_{\gamma(t)} & & 0 \\ & \ddots & \\ 0 & & \mathfrak{Ric}^{\alpha_\ell}_{\gamma(t)} \end{pmatrix}.$$

Thus, for any level α, the average over the level V_α satisfies the $\ell\times\ell$ matrix Riccati equation

$$\dot{V}_\alpha + \Gamma_1 V_\alpha + V_\alpha\Gamma_1^* + R_\alpha(t) + V_\alpha\Gamma_2 V_\alpha = 0, \qquad \lim_{t\to 0^+} V_\alpha^{-1} = 0,$$

and, under our hypotheses, $R_\alpha(t) \geq \operatorname{diag}\{k_1,\dots,k_\ell\}$. Then we proceed as in the proof of Theorem 2.A, with $\operatorname{diag}\{k_1,\dots,k_\ell\}$ in place of Q_+, and we obtain the statement. □

2.6 A sub-Riemannian Bonnet-Myers theorem

As an application of Theorem 2.B, we prove a sub-Riemannian analogue of the classical Bonnet-Myers theorem. In the following, we say that a property (P) holds for the *generic normal geodesic* if, for any $x \in M$, there exists an open non-empty Zariski subset $A_x \subseteq T^*_x M$ such that (P) is true for any normal geodesic whose initial covector is in A_x. In particular A_x is dense in $T^*_x M$. We make the following general assumptions on the sub-Riemannian manifold M:

(i) The generic normal geodesic is equiregular.

By [9, Lemma 5.17] (that is Lemma 1.69 in Chapter 1), the generic normal geodesic is ample, with the same (maximal) growth vector, and thus the same Young diagram D_x. Thus, we also assume that

(ii) The Young diagram D_x is the same for any point x.

Assumptions (i)-(ii) are satisfied, for instance, by any slow-growth distribution, a large class of sub-Riemannian structures including contact, quasi-contact, fat, Engel, Goursat-Darboux distributions (see [9, Sec. 5.5], that is Sec. 1.5.5). Under these assumptions, there exists a unique Young diagram D such that the generic ample geodesic has Young diagram D. Thus, with a generic geodesic $\gamma(t)$ we can associate the directional curvature $\mathfrak{R}_{\gamma(t)} : T_{\gamma(t)}M \to \mathbb{R}$ and the corresponding Ricci curvatures $\mathfrak{Ric}^\alpha_{\gamma(t)}$, one for each superbox α in D.

Theorem 2.C. *Let M be a complete, connected sub-Riemannian manifold satisfying (i)-(ii). Assume that there exists a level α of length ℓ and size r of the Young diagram D and constants $k_1, \dots, k_\ell$ such that, for any length parametrized geodesic $\gamma(t)$*

$$\frac{1}{r}\,\mathfrak{Ric}^{\alpha_i}_{\gamma(t)} \geq k_i, \qquad \forall i = 1, \dots, \ell, \quad \forall t \geq 0.$$

Then, if the polynomial

$$P_{k_1,\dots,k_\ell}(x) := x^{2\ell} - \sum_{i=0}^{\ell-1} (-1)^{\ell-i} k_{\ell-i} x^{2i}$$

has at least one simple purely imaginary root, the manifold is compact, has diameter not greater than $t_c(k_1, \dots, k_\ell) < +\infty$. Moreover, its fundamental group is finite.

Proof. First, we show that $\operatorname{diam}(M) := \sup\{\mathsf{d}(x, y) | \, x, y \in M\} \leq t_c(k_1, \dots, k_\ell)$. Let $x_0 \in M$, and let $\Sigma_{x_0} \subseteq M$ be the set of points x such that there exists a unique minimizing geodesic connecting x_0 with x, strictly normal and with no conjugate points. The following fundamental result is due to Agrachev and Rifford (see [4] or also Theorem 1.68).

Theorem 2.40. *Let $x_0 \in M$. The set Σ_{x_0} is open, dense and the sub-Riemannian squared distance $x \mapsto \mathsf{d}^2(x_0, x)$ is smooth on Σ_{x_0}.*

Indeed, the sub-Riemannian exponential map $\mathcal{E}_{x_0} : T^*_{x_0}M \to M$ is a smooth diffeomorphism between $\overline{\Sigma}_{x_0} := \mathcal{E}^{-1}_{x_0}(\Sigma_{x_0}) \subseteq T^*_{x_0}M$ and Σ_{x_0}. Now consider all the normal geodesics connecting x_0 with points in Σ_{x_0}, associated with initial covectors in $\overline{\Sigma}_{x_0}$. The generic normal geodesic, with covector in $A_{x_0} \subseteq T^*_{x_0}M$ is ample and equiregular, with the same growth vector, and thus the same Young diagram $D_{x_0} = D$. Thus, for an open dense set $\Sigma'_{x_0} := \mathcal{E}_{x_0}(A_{x_0}) \cap \Sigma_{x_0} \subseteq M$, there exists a unique geodesic connecting x_0 with $x \in \Sigma'_{x_0}$, and it has Young diagram D.

Now we apply Theorem 2.B to all the geodesics connecting x_0 with points $x \in \Sigma'_{x_0}$, and we obtain that the first conjugate time t_c along these geodesics satisfies $t_c \leq t_c(k_1, \dots, k_\ell)$. These geodesics lose optimality after the first conjugate point and, since the geodesics are parametrised by length, we have that, for any $x_0 \in M$, $\sup\{\mathsf{d}(x_0, x) | \, x \in \Sigma'_{x_0}\} \leq t_c(k_1, \dots, k_\ell)$. By density of Σ'_{x_0} in M, we obtain that $\operatorname{diam}(M) \leq t_c(k_1, \dots, k_\ell)$. The condition on the roots of $P_{k_1,\dots,k_\ell}$ implies that $t_c(k_1, \dots, k_\ell) < +\infty$, by Theorem 2.27.

By completeness of M, closed sub-Riemannian balls are compact, hence M is compact. For the result about the fundamental group, the argument is the classical one. □

Remark 2.41. In the Riemannian case, $P_{k_1}(x) = x^2 + k_1$. Then we recover the classical Bonnet-Myers theorem since, by Example 2.28, $t_c(k_1) = \pi/\sqrt{k_1}$.

2.7 Applications to left-invariant structures on 3D unimodular Lie groups

Consider a contact left-invariant sub-Riemannian structure on a 3D manifold. Any non-trivial geodesic is ample, equiregular and has the same Young diagram, with two boxes on the first row, and one in the second row (see Example 2.15). The subspace associated with the box in the second row corresponds to the direction of the motion, i.e., the tangent vector to the geodesic. Since the curvature always vanishes in this direction (see Remark 2.32), we can restrict to a single-level Young diagram α of length $\ell = 2$ and size $r = 1$. We denote by α_1, α_2 the two boxes of this level. Then, Theorem 2.B rewrites as follows.

Theorem 2.42. *Let $\gamma(t)$ be a length parametrized geodesic of a contact left-invariant sub-Riemannian structure on a 3D manifold. Assume that*

$$\mathfrak{Ric}^{\alpha_i}_{\gamma(t)} \geq k_i, \qquad i = 1, 2,$$

for some k_1, k_2 such that

$$\begin{cases} k_1 > 0, \\ 4k_2 > -k_1^2, \end{cases} \qquad or \qquad \begin{cases} k_1 \leq 0, \\ k_2 > 0. \end{cases} \tag{2.18}$$

Then $t_c(\gamma) \leq t_c(k_1, k_2) < +\infty$. If the hypotheses are satisfied for every length parametrised geodesic, then the manifold is compact, with diameter not greater than $t_c(k_1, k_2)$. Moreover, its fundamental group is finite.

In the statement of Theorem 2.42 we allow also for negative Ricci curvatures. Indeed, in general, $\mathfrak{Ric}^{\alpha_2}_{\gamma(t)}$ is not sign-definite along the geodesic.

2.7.1 Invariants of a 3D contact structure

In this section we introduce the invariants χ, κ of 3D contact sub-Riemannian structures, not necessarily left-invariant. For left-invariant structures, χ and κ are constant and we write the expression for $\mathfrak{Ric}^{\alpha_1}_{\gamma(t)}$ and $\mathfrak{Ric}^{\alpha_2}_{\gamma(t)}$ in terms of these quantities. The presentation follows closely the one contained in [5], where the interested reader can find more details.

Recall that a three-dimensional sub-Riemannian structure is contact if $\mathscr{D} = \ker \omega$, where $d\omega|_{\mathscr{D}_x}$ is non degenerate, for every $x \in M$. In what follows we normalize the contact structure by requiring that $d\omega|_{\mathscr{D}_x}$ agrees with the volume induced by the inner product on $\mathscr{D}$. The *Reeb vector field* associated with the contact structure is the unique vector field X_0 such that $\omega(X_0) = 1$ and $d\omega(X_0, \cdot) = 0$. Notice that X_0 depends only on the sub-Riemannian structure. For every orthonormal frame X_1, X_2 on the distribution, we have

$$\begin{aligned} [X_1, X_0] &= c_{01}^1 X_1 + c_{01}^2 X_2, \\ [X_2, X_0] &= c_{02}^1 X_1 + c_{02}^2 X_2, \\ [X_2, X_1] &= c_{12}^1 X_1 + c_{12}^2 X_2 + X_0, \end{aligned} \tag{2.19}$$

where $c_{ij}^k \in C^\infty(M)$. The sub-Riemannian Hamiltonian is

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

where $h_i(\lambda) = \langle \lambda, X_i(q) \rangle$ are the linear-on-fibers functions on T^*M associated with the vector fields X_i, for $i = 0, 1, 2$. Length parametrized geodesics are projections of solutions of the Hamiltonian system associated with H on T^*M that are contained in the level set $H = 1/2$.

The Poisson bracket $\{H, h_0\}$ is an invariant of the sub-Riemannian structure and, by definition, it vanishes everywhere if and only if the flow of the Reeb vector field e^{tX_0} is a one-parameter family of sub-Riemannian isometries. A standard computation gives

$$\{H, h_0\} = c_{01}^1 h_1^2 + (c_{01}^2 + c_{02}^1) h_1 h_2 + c_{02}^2 h_2^2.$$

For every $x \in M$, the restriction of $\{H, h_0\}$ to $T_x^* M$, that we denote by $\{H, h_0\}_x$, is a quadratic form on the dual of the distribution $\mathscr{D}_x^*$, hence it can be interpreted as a symmetric operator on the distribution $\mathscr{D}_x$ itself. In particular its determinant and its trace are well defined. Moreover one can show that $\mathrm{tr}\{H, h_0\}_x = c_{01}^1 + c_{02}^2 = 0$, for every $x \in M$. The first invariant χ is defined as the positive eigenvalue of this operator, namely

$$\chi(x) := \sqrt{-\det\{H, h_0\}_x} \geq 0. \tag{2.20}$$

The second invariant κ can be defined via the structure constants (2.19) as follows:

$$\kappa(x) := X_2(c_{12}^1) - X_1(c_{12}^2) - (c_{12}^1)^2 - (c_{12}^2)^2 + \frac{c_{01}^2 - c_{02}^1}{2}. \tag{2.21}$$

One can prove that the expression (2.21) is invariant by rotation of the orthonormal frame.

Remark 2.43. The quantities χ, κ were first introduced in [2] as differential invariants appearing in the asymptotic expansion of the cut and conjugate locus of the sub-Riemannian exponential map near to the base point.

2.7.2 Left-invariant structures

For left-invariant structures, the functions c_{ij}^k and the invariants χ, κ, are constant on M can be used to classify the left-invariant structures on three-dimensional Lie groups. In particular, when $\chi = 0$, the unique left-invariant structures (up to local isometries) are the Heisenberg group $\mathbb{H}$ and the Lie groups $SU(2)$ and $SL(2)$, with metric given by the Killing form, corresponding to the choice of $\kappa = 0, 1, -1$, respectively. When $\chi > 0$, for each choice of (χ, κ) there exists exactly one unimodular Lie group with these values (see [5, Thm. 1]).

Case $\chi = 0$

When $\chi = 0$, the flow of the Reeb vector field is a one-parameter family of sub-Riemannian isometries. In particular from the computations contained in [18, Thm. 6.2] one gets that the directional curvature $\mathfrak{R}_{\gamma(t)}$ is diagonal with entries $\mathfrak{Ric}_{\gamma(t)}^{\alpha_1}$ and $\mathfrak{Ric}_{\gamma(t)}^{\alpha_2}$, where

$$\mathfrak{Ric}_{\gamma(t)}^{\alpha_1} = h_0^2(t) + \kappa(h_1^2(t) + h_2^2(t)), \qquad \mathfrak{Ric}_{\gamma(t)}^{\alpha_2} = 0.$$

We stress that $\lambda(t) = (h_0(t), h_1(t), h_2(t))$ is the solution of the Hamiltonian system associated with H and $\lambda = (h_1(0), h_2(0), h_0(0))$ is the initial covector associated with the geodesic $\gamma(t)$.

For any length-parametrized geodesic $H(\lambda(t)) = 1/2$, namely $h_1^2(t) + h_2^2(t) = 1$. Moreover $h_0(t) = h_0$ is a constant of the motion. Thus

$$\mathfrak{Ric}_{\gamma(t)}^{\alpha_1} = h_0^2 + \kappa, \qquad \mathfrak{Ric}_{\gamma(t)}^{\alpha_2} = 0.$$

Notice that $\mathfrak{Ric}^{\alpha_1}_{\gamma(t)}$ and $\mathfrak{Ric}^{\alpha_2}_{\gamma(t)}$ are constant in t and $\mathfrak{R}_{\gamma(t)}$ is diagonal, so for all these cases we can apply Theorem 2.A, computing the exact value of the first conjugate time. In particular, this recovers the following well known results obtained in [34, 40].

- $\mathbb{H}$. In this case $\kappa = 0$. If $h_0 = 0$ we have $\mathfrak{Ric}^{\alpha_1}_{\gamma(t)} = \mathfrak{Ric}^{\alpha_2}_{\gamma(t)} = 0$ and the geodesic has no conjugate point. If $h_0 \neq 0$ then $t_c = 2\pi/|h_0|$.
- $SU(2)$. In this case $\kappa = 1$. We have $\mathfrak{Ric}^{\alpha_1}_{\gamma(t)} = h_0^2 + 1$, $\mathfrak{Ric}^{\alpha_2}_{\gamma(t)} = 0$ and every geodesic has conjugate time $t_c = 2\pi/\sqrt{h_0^2 + 1}$.
- $SL(2)$. In this case $\kappa = -1$. We have $\mathfrak{Ric}^{\alpha_1}_{\gamma(t)} = h_0^2 - 1$, $\mathfrak{Ric}^{\alpha_2}_{\gamma(t)} = 0$ and we have two cases. If $h_0 \leq 1$ then $t_c = +\infty$. If $h_0 > 1$ every geodesic has conjugate time $t_c = 2\pi/\sqrt{h_0^2 - 1}$.

Let us mention that, for $SU(2)$, the first condition of (2.18) holds for any geodesic. Hence, thanks to Theorem 2.42, we recover its compactness and the exact estimate on its diameter, equal to 2π.

Case $\chi > 0$

In this section we prove our result on 3D unimodular Lie groups with $\chi > 0$. Let us recall that under these assumptions, there exists a special orthonormal frame for the sub-Riemannian structure. In terms of the latter we provide the explicit expression of a constant of the motion.

Proposition 2.44. *Let M be a 3D unimodular Lie group, endowed with a contact left-invariant structure, with $\chi > 0$. Then there exists a left-invariant orthonormal frame X_1, X_2 on the distribution such that*

$$\{H, h_0\} = 2\chi h_1 h_2.$$

Moreover the Lie algebra defined by the frame X_0, X_1, X_2 satisfies

$$\begin{aligned} [X_1, X_0] &= (\chi + \kappa) X_2, \\ [X_2, X_0] &= (\chi - \kappa) X_1, \\ [X_2, X_1] &= X_0. \end{aligned} \tag{2.22}$$

The function $E : T^*M \to \mathbb{R}$, defined by

$$E = \frac{h_0^2}{2\chi} + h_2^2,$$

is a constant of the motion, i.e., $\{H, E\} = 0$. Finally the curvatures $\mathfrak{Ric}^{\alpha_1}_{\gamma}$ and $\mathfrak{Ric}^{\alpha_2}_{\gamma}$ satisfy:

$$\mathfrak{Ric}^{\alpha_1}_{\gamma} = h_0^2 + 3\chi(h_1^2 - h_2^2) + \kappa(h_1^2 + h_2^2), \tag{2.23}$$

$$\mathfrak{Ric}^{\alpha_2}_{\gamma} = 6\chi(h_1^2 - h_2^2)h_0^2 - 2\chi(\chi + \kappa)h_1^4 - 12\chi^2 h_1^2 h_2^2 - 2\chi(\chi - \kappa)h_2^4. \tag{2.24}$$

In Eqs. (2.23) and (2.24) we suppressed the explicit dependence on t.

Proof. From [5, Prop. 13] it follows that there exists a unique (up to a sign) canonical frame X_0, X_1, X_2 such that

$$[X_1, X_0] = c_{01}^2 X_2,$$
$$[X_2, X_0] = c_{02}^1 X_1,$$
$$[X_2, X_1] = c_{12}^1 X_1 + c_{12}^2 X_2 + X_0.$$

In particular, if the Lie group is unimodular, then the left and the right Haar measures coincide. This implies $c_{12}^1 = c_{12}^2 = 0$ (cf. proof of [5, Thm. 1]). Then, from (2.20) and (2.21), it follows that $\chi = (c_{01}^2 + c_{02}^1)/2$, and $\kappa = (c_{01}^2 - c_{02}^1)/2$, which imply (2.22).

Let us show that, if (2.22) holds, then $\{H, E\} = 0$. Using that $\{H, h_0\} = 2\chi h_1 h_2$ and $\{H, h_2\} = \{h_1, h_2\}h_1 = -h_0 h_1$ one gets

$$\{H, E\} = \frac{1}{\chi}\{H, h_0\}h_0 + 2\{H, h_2\}h_2 = 2h_1h_2h_0 - 2h_1h_2h_0 = 0.$$

Finally, Eqs. (2.23) and (2.24) are simply formulae form [18, Thm. 6.2] specified for left-invariant structures and rewritten in terms of χ, κ in the frame introduced above (notice that the constants c_{ij}^k appearing here are the opposite of those used in [18]). □

Since E is a constant of the motion, for any length parametrized geodesic $\gamma(t)$ we denote by $E(\gamma)$ the (constant) value of $E(\lambda(t))$, where $\lambda(t)$ is the solution of the Hamiltonian system associated with H such that $\gamma(t) = \pi(\lambda(t))$.

Theorem 2.D. *Let M be a 3D unimodular Lie group, endowed with a contact left-invariant structure, with $\chi > 0$ and $\kappa \in \mathbb{R}$. Then there exists $\bar{E} = \bar{E}(\chi, \kappa)$ such that every length parametrized geodesic γ with $E(\gamma) \geq \bar{E}$ has a finite conjugate time.*

Proof. We prove that the assumptions of Theorem 2.42 are satisfied for every geodesic when E is large enough. Since E is a constant of the motion and $H = 1/2$ we have

$$h_2^2 = E - \frac{h_0^2}{2\chi}, \qquad h_1^2 = 1 - E + \frac{h_0^2}{2\chi}. \tag{2.25}$$

Plugging Eq. (2.25) into Eqs. (2.23) and (2.24), $\mathfrak{Ric}_{\gamma(t)}^{\alpha_1}$ and $\mathfrak{Ric}_{\gamma(t)}^{\alpha_2}$ are rewritten as follows

$$\mathfrak{Ric}_{\gamma(t)}^{\alpha_1} = 4h_0^2 - 3\chi(2E-1) + \kappa,$$
$$\mathfrak{Ric}_{\gamma(t)}^{\alpha_2} = 8h_0^4 - [2\kappa + 10\chi(2E-1)]h_0^2 + [2\chi\kappa(2E-1) + \chi^2(8E^2 - 8E - 2)]. \tag{2.26}$$

Since $h_1^2 + h_2^2 = 1$ one has $|h_2| \leq 1$, from (2.25) one has the following bound for h_0 along the curve

$$2\chi(E-1) \leq h_0^2(t) \leq 2\chi E. \tag{2.27}$$

Then we have easily a lower bound for $\mathfrak{Ric}_{\gamma(t)}^{\alpha_1}$

$$\mathfrak{Ric}_{\gamma(t)}^{\alpha_1} \geq 8\chi(E-1) - 3\chi(2E-1) + \kappa$$
$$\geq 2\chi E - 5\chi + \kappa =: k_1$$

Since we want to prove the result for E large enough, we assume that

$$E \geq \frac{1}{2}\left(5 - \frac{\kappa}{\chi}\right),$$

so that $k_1 > 0$, and the coefficient of h_0^2 in (2.26) is negative. Then using (2.27) one estimates

$$\mathfrak{Ric}^{\alpha_2}_{\gamma(t)} \geq 2\chi^2(15 - 26E) - 2\chi\kappa =: k_2$$

In order to show that the first condition of Eq. (2.18) of Theorem 2.42 is satisfied we also compute

$$4k_2 + k_1^2 = 4\chi^2 E^2 + a(\chi,\kappa)E + b(\chi,\kappa), \tag{2.28}$$

where a and b are the following quadratic functions

$$a(\chi,\kappa) = 4\chi\kappa - 228\chi^2, \qquad b(\chi,\kappa) = 145\chi^2 - 18\chi\kappa + \kappa^2.$$

Since the coefficient of E^2 in Eq. (2.28) is positive, there exists $\bar{E} = \bar{E}(\chi,\kappa)$, the largest positive root of Eq. (2.28), such that $4k_2 + k_1^2 > 0$ for all $E > \bar{E}$, which ends the proof. □

Remark 2.45. The roots of Eq. (2.28), and in particular $\bar{E}(\chi,\kappa)$, depend only on the ratio κ/χ. This means that this number is invariant by rescaling of the sub-Riemannian structure. This could seem strange at a first glance but is a consequence of the fact that we consider only length parametrized geodesics. We also stress that, in general, the value $\bar{E}(\chi,\kappa)$ given by this computation is not sharp.

Chapter 3

On conjugate times of Linear Quadratic optimal control problems

3.1 Introduction

Linear Quadratic optimal control problems (LQ in the following) are a standard topic in control theory and dynamical systems, and are very popular in applications. They consist in a linear control system with quadratic Lagrangian. We briefly recall the general features of a LQ problem, and we refer to [15, Chapter 16] and [46, Chapter 7] for further details. We are interested in *admissible trajectories*, namely curves $x : [0, t_1] \to \mathbb{R}^n$ such that there exists a control $u \in L^2([0, t_1], \mathbb{R}^k)$ such that

$$\dot{x} = Ax + Bu, \qquad x(0) = x_0, \qquad x(t_1) = x_1, \qquad x_0, x_1, t_1 \text{ fixed},$$

that minimize a quadratic functional $\phi_{t_1} : L^2([0, t_1], \mathbb{R}^k) \to \mathbb{R}$ of the form

$$\phi_{t_1}(u) = \frac{1}{2} \int_0^{t_1} (u^* R u + x^* P u + x^* Q x)\, dt.$$

The condition $R \geq 0$ is necessary for existence of optimal control. We also assume $R > 0$ (for the singular case we refer to [46, Chapter 9]). Without loss of generality we may reduce to the case

$$\phi_{t_1}(u) = \frac{1}{2} \int_0^{t_1} (u^* u - x^* Q x)\, dt.$$

Here A, B, Q are constant matrices of the appropriate dimension. The vector Ax represents the *drift* field, while the columns of B represent the controllable directions. The meaning of the *potential* term Q will be clear later, when we will introduce the Hamiltonian associated with the LQ problem.

We assume that the system is *controllable*, namely there exists $m > 0$ such that

$$\operatorname{rank}(B, AB, \ldots, A^{m-1}B) = n.$$

This hypothesis implies that, for any choice of t_1, x_0, x_1, the set of controls u such that the associated trajectory $x_u : [0, t_1] \to \mathbb{R}^n$ connects x_0 with x_1 in time t_1 is non-empty.

It is well known that the optimal trajectories of the LQ system are projections $(p, x) \mapsto x$ of the solutions of the Hamiltonian system

$$\dot{p} = -\partial_x H(p, x), \qquad \dot{x} = \partial_p H(p, x), \qquad (p, x) \in T^*\mathbb{R}^n = \mathbb{R}^{2n},$$

where the Hamiltonian function $H : \mathbb{R}^{2n} \to \mathbb{R}$ is defined by

$$H(p, x) = \frac{1}{2}(p, x)^* \mathbf{H} \begin{pmatrix} p \\ x \end{pmatrix}, \qquad \mathbf{H} = \begin{pmatrix} BB^* & A \\ A^* & Q \end{pmatrix}. \tag{3.1}$$

We denote by $P_t : \mathbb{R}^{2n} \to \mathbb{R}^{2n}$ the flow of the Hamiltonian system, which is defined for all $t \in \mathbb{R}$. To exploit the natural symplectic setting on $T^*\mathbb{R}^n = \mathbb{R}^{2n}$, we employ canonical coordinates (p, x) such that the symplectic form $\omega = \sum_{i=1}^n dp_i \wedge dx_i$ is represented by the matrix $\Omega = \left(\begin{smallmatrix} 0 & \mathbb{I}_n \\ -\mathbb{I}_n & 0 \end{smallmatrix}\right)$. The flow lines of P_t are precisely the integral lines of the *Hamiltonian vector field* $\vec{H} \in \mathrm{Vec}(\mathbb{R}^{2n})$, defined by $dH(\cdot) = \omega(\,\cdot\,, \vec{H})$. More explicitly

$$\vec{H}_{(p,x)} = \begin{pmatrix} -A^* & -Q \\ BB^* & A \end{pmatrix} \begin{pmatrix} p \\ x \end{pmatrix} = -\Omega \mathbf{H} \begin{pmatrix} p \\ x \end{pmatrix}.$$

By the term Hamiltonian vector field, we denote both the linear field $\vec{H}$ and the associated matrix $-\Omega\mathbf{H}$. The Hamiltonian flow can be explicitly written in terms of the latter as

$$P_t = e^{-t\Omega\mathbf{H}},$$

where the r.h.s. is the standard matrix exponential.

Conjugate times

We stress that not all the integral lines of the Hamiltonian flow lead to minimizing solutions of the LQ problem, since they only satisfy first order conditions for optimality. For this reason, they are usually called *extremals*. Sufficiently short segments, however, are optimal, but they lose optimality at some time $t > 0$, called the *first conjugate time*. In the following, we give a geometrical definition of conjugate time, in terms of curves in the Grassmannian of Lagrangian subspaces of $\mathbb{R}^{2n}$.

We say that a subspace $\Lambda \subset \mathbb{R}^{2n}$ is *Lagrangian* if $\omega|_\Lambda \equiv 0$, and $\dim \Lambda = n$. A notable example of Lagrangian subspace is the *vertical* subspace, that is $\mathcal{V} := \{(p, 0) \mid p \in \mathbb{R}^n\}$.

Definition 3.1. The Jacobi curve $J(\cdot)$ is the following family of Lagrangian subspaces of $\mathbb{R}^{2n}$

$$J(t) := e^{t\Omega\mathbf{H}}\mathcal{V}, \qquad \mathcal{V} := \{(p, 0) \mid p \in \mathbb{R}^n\}.$$

From the geometrical viewpoint, $J(\cdot)$ is a smooth curve in the submanifold of the Grassmannian of the n-dimensional subspaces of $\mathbb{R}^{2n}$ defined by the Lagrangian subspaces.

Definition 3.2. We say that t is a conjugate time if $J(t) \cap \mathcal{V} \neq 0$. The *multiplicity* of the conjugate time t is the dimension of the intersection.

In the language introduced by V. Arnold, these are times of *verticality* of the Jacobi curve. It is not hard to show that t is a conjugate time if and only if there exist solutions of the Hamilton equations such that $x(0) = x(t) = 0$.

We briefly recall the connection between conjugate times and second order conditions for optimality. The solutions of the LQ problems can be seen as constrained minima of the quadratic functional J_{t_1} on $\mathcal{U}(x_0, x_1) \subset L^2([0, t_1], \mathbb{R}^k)$ given by all the controls u such that $x_u(0) = x_0$ and $x_u(t_1) = x_1$. It is easy to check that $\mathcal{U}(x_0, x_1) = u^* + \mathcal{U}(0, 0)$ for any $u^* \in \mathcal{U}(x_0, x_1)$, that is $\mathcal{U}(x_0, x_1)$ is an affine space over the vector space $\mathcal{U}(0, 0)$. For this reason, the behaviour of J_{t_1}, restricted to $\mathcal{U}(0, 0)$ provides all the informations about optimality. It is a well known fact that the number of conjugate times in the interval $(0, t_1)$, counted with their multiplicity, is equal to the negative inertia index of the quadratic form $J_{t_1} : \mathcal{U}(0, 0) \to \mathbb{R}$ (this can be proved directly with the techniques in [15, Propositions 16.2, 16.3], see also [3, Theorem I.2] for a more general setting). The occurrence of conjugate times implies that an extremal cannot be a minimizer, since one can find a small variation of u^* that decreases the value of J_{t_1}. The first conjugate time determines existence and uniqueness of minimizing solutions of the LQ problem, as specified by the following proposition.

Proposition 3.3. *Let t_c be the first conjugate time, namely $t_c := \inf\{t > 0 |\, J(t) \cap \mathcal{V} \neq 0\}$.*

- *For $t_1 < t_c$, for any x_0, x_1 there exists a unique minimizer connecting x_0 with x_1 in time t_1.*
- *For $t_1 > t_c$, for any x_0, x_1 there exists no minimizer connecting x_0 with x_1 in time t_1.*
- *For $t_1 = t_c$, existence of minimizers depends on the initial data.*

We completely characterise the occurrence of conjugate times for a controllable LQ problem. In particular, we prove the following result.

Theorem 3.A. *The conjugate times of a controllable linear quadratic optimal control problem obey the following dichotomy:*

- *If the Hamiltonian field $\vec{H}$ has at least one odd-dimensional Jordan block corresponding to a pure imaginary eigenvalue, the number of conjugate times in the interval $[0, T]$ grows to infinity for $T \to \pm\infty$.*
- *If the Hamiltonian field $\vec{H}$ has no odd-dimensional Jordan blocks corresponding to a pure imaginary eigenvalue, there are no conjugate times.*

In Sec. 3.3, we also provide estimates for the first conjugate time, in terms of the (signed) eigenvalues of $\vec{H}$ (see Corollaries 3.31 and 3.33).

Before passing to a more detailed description of curves of Lagrangian subspaces, we stress that the concept of Jacobi curves is not limited to LQ optimal control problems and can be defined for way more general geometrical structures, such as control systems with Tonelli Lagrangian including, among the others Riemannian, sub-Riemannian, Finsler and sub-Finsler manifolds. In these more general settings, however, we cannot exploit the natural linear structure of $\mathbb{R}^n$, and the Jacobi curve is a curve of subspaces of the tangent space to the cotangent bundle, associated with a fixed "geodesic" (i.e. locally minimizing curve) of the underlying structure. We refer the interested reader to [9, 16, 17].

The plan of the chapter is as follows. In Sec. 3.2 we recall some basic facts about geometry of curves in Lagrange Grassmannian, and the main technical tool: the Maslov index. Then, in Sec. 3.3 we prove the main result.

3.2 Curves in the Lagrange Grassmannian

Let (Σ, ω) be a $2n$-dimensional symplectic vector space. Recall that subspace $\Lambda \subset \Sigma$ is called *Lagrangian* if it has dimension n and $\omega|_\Lambda \equiv 0$. The *Lagrange Grassmannian* $\mathcal{L}(\Sigma)$ is the set of all n-dimensional Lagrangian subspaces of Σ.

Proposition 3.4. *$\mathcal{L}(\Sigma)$ is a compact $n(n+1)/2$-dimensional submanifold of the Grassmannian of n-planes in Σ.*

Proof. Let $\Delta \in \mathcal{L}(\Sigma)$, and consider the set $\Delta^{\pitchfork} := \{\Lambda \in \mathcal{L}(\Sigma) \,|\, \Lambda \cap \Delta = 0\}$ of all Lagrangian subspaces transversal to Δ. Clearly, the collection of these sets for all $\Delta \in \mathcal{L}(\Sigma)$ is an open cover of $\mathcal{L}(\Sigma)$. Then it is sufficient to find submanifold coordinates on each $\Delta^{\pitchfork}$.

Let us fix any Lagrangian complement Π of Δ (which always exists, though it is not unique). Every n-dimensional subspace $\Lambda \subset \Sigma$ that is transversal to Δ is the graph of a linear map from Π to Δ. Choose an adapted Darboux basis on Σ, namely a basis $\{e_i, f_i\}_{i=1}^n$ such that

$$\Delta = \operatorname{span}\{f_1, \dots, f_n\}, \qquad \Pi = \operatorname{span}\{e_1, \dots, e_n\},$$
$$\omega(e_i, f_j) - \delta_{ij} = \omega(f_i, f_j) = \omega(e_i, e_j) = 0, \qquad i, j = 1, \dots, n.$$

In these coordinates, the linear map is represented by a matrix S_Λ such that

$$\Lambda \cap \Delta = 0 \Leftrightarrow \Lambda = \{(p, S_\Lambda p) \,|\, p \in \Pi \simeq \mathbb{R}^n\}.$$

Moreover it is easy to see that $\Lambda \in \mathcal{L}(\Sigma)$ if and only if $S_\Lambda = S_\Lambda^*$. Hence, the open set $\Delta^{\pitchfork}$ of all Lagrangian subspaces transversal to Δ is parametrized by the set of symmetric matrices, and this gives smooth submanifold coordinates on $\Delta^{\pitchfork}$. This also proves that the dimension of $\mathcal{L}(\Sigma)$ is $n(n+1)/2$. Finally, as a closed subset of a compact manifold, $\mathcal{L}(\Sigma)$ is compact. □

Fix now $\Lambda \in \mathcal{L}(\Sigma)$. The tangent space $T_\Lambda \mathcal{L}(\Sigma)$ to the Lagrange Grassmannian at the point Λ can be canonically identified with the set of quadratic forms on the space Λ itself, namely

$$T_\Lambda \mathcal{L}(\Sigma) \simeq Q(\Lambda).$$

Indeed, consider a smooth curve $\Lambda(\cdot)$ in $\mathcal{L}(\Sigma)$ such that $\Lambda(0) = \Lambda$, and denote by $\dot{\Lambda} \in T_\Lambda \mathcal{L}(\Sigma)$ its tangent vector. For any point $z \in \Lambda$ and any smooth extension $z(t) \in \Lambda(t)$, we define the quadratic form

$$\dot{\Lambda} := z \mapsto \omega(z, \dot{z}),$$

where $\dot{z} = \dot{z}(0)$. A simple check shows that the definition does not depend on the extension $z(t)$. Finally, if in local coordinates $\Lambda(t) = \{(p, S(t)p) \,|\, p \in \mathbb{R}^n\}$, the quadratic form $\dot{\Lambda}$ is represented by the matrix $\dot{S}(0)$. In other words, if $z \in \Lambda$ has coordinates $p \in \mathbb{R}^n$, then $\dot{\Lambda}[z] = p^* \dot{S}(0) p$.

3.2.1 Transversality properties

In this section we introduce some important properties of curves in the Lagrange Grassmannian. Then we discuss the specific case of a Jacobi curve. Let $J(\cdot) \in \mathcal{L}(\Sigma)$ be a smooth curve in the Lagrange Grassmannian. For $i \in \mathbb{N}$, consider

$$J^{(i)}(t) = \operatorname{span}\left\{ \frac{d^j}{dt^j}\ell(t) \middle|\, \ell(t) \in J(t),\, \ell(t) \text{ smooth},\, 0 \leq j \leq i \right\} \subset \Sigma, \qquad i \geq 0.$$

The subspace $J^{(i)}(t)$ is the *i-th extension* of the curve $J(\cdot)$ at t. The flag

$$J(t) = J^{(0)}(t) \subset J^{(1)}(t) \subset J^{(2)}(t) \subset \ldots \subset \Sigma,$$

is the *associated flag of the curve* at the point t. The curve $J(\cdot)$ is called:

(i) *equiregular* at t if $\dim J^{(i)}(\cdot)$ is locally constant at t, for all $i \in \mathbb{N}$,

(ii) *ample* at t if there exists $N \in \mathbb{N}$ such that $J^{(N)}(t) = \Sigma$,

(iii) *monotone increasing* (resp. *decreasing*) at t if $\dot{J}(t)$ is non-negative definite (resp. non-positive definite) as a quadratic form.

In coordinates, $J(t) = \{(p, S(t)p)|\ p \in \mathbb{R}^n\}$ for some smooth family of symmetric matrices $S(t)$. The curve is ample at t if and only if there exists $N \in \mathbb{N}$ such that

$$\operatorname{rank}\{\dot{S}(t), \ddot{S}(t), \ldots, S^{(N)}(t)\} = n.$$

We say that the curve is equiregular, ample or monotone (increasing or decreasing) if it is equiregular, ample or monotone for all t in the domain of the curve.

A crucial property of ample, monotone curves is described in the following lemma.

Lemma 3.5. *Let $J(\cdot) \in \mathcal{L}(\Sigma)$ a monotone, ample curve at t_0. Then, for any fixed Lagrangian subspace Λ, there exists $\varepsilon > 0$ such that $J(t) \cap \Lambda = 0$ for $0 < |t - t_0| < \varepsilon$.*

In other words, ample, monotone curves can intersect any fixed Lagrange subspace Λ only at a discrete set of times.

Proof. Without loss of generality, assume $t_0 = 0$. Choose a Lagrangian splitting $\Sigma = \Lambda \oplus \Pi$, such that, for $|t| < \varepsilon$, the curve is contained in the chart defined by such a splitting. In coordinates, $J(t) = \{(p, S(t)p)|\ p \in \mathbb{R}^n\}$, with $S(t)$ symmetric. The curve is monotone, then $\dot{S}(t)$ is a semidefinite symmetric matrix. Without loss of generality, we assume $\dot{S}(t) \geq 0$. Assume that $J(0) \cap \Lambda \neq 0$. In coordinate, this means that $S(0)$ has some vanishing eigenvalues. We now show that the whole spectrum of $S(t)$ is strictly increasing in t, hence it moves away from zero for t sufficiently small.

Notice that $S(t) - S(0) = \int_0^t \dot{S}(\tau)d\tau \geq 0$, by the monotonicity assumption. Then, for any $z \in \mathbb{R}^n$, consider the smooth function $t \mapsto f_z(t) := z^*[S(t) - S(0)]z$, which is non-decreasing and vanishes at $t = 0$. Moreover, $f_z(t)$ cannot be constantly zero on any interval of the form $[0, \delta)$, otherwise z would be in the kernel of $S(t) - S(0)$ for all $t \in [0, \delta)$ and, a fortiori, in the kernel of all the derivatives $S^{(N)}(0)$, which is absurd by the ampleness hypothesis. Therefore, $f_z(t) > 0$ for $0 < t < \varepsilon$. Since z is arbitrary

$$S(t) > S(0), \qquad 0 < t < \varepsilon. \tag{3.2}$$

Now, denote by $\lambda_1(t) \geq \ldots \geq \lambda_n(t)$ the eigenvalues of $S(t)$ at each fixed t. Then, by the Courant min-max principle, we have the following variational characterisation

$$\lambda_k(t) = \max\{\min\{x^* S(t)x |\, x \in U \subset \mathbb{R}^n,\ |x| = 1\} |\, \dim U = k\}, \qquad k = 1, \ldots, n.$$

Thus, by Eq. (3.2), each eigenvalue is strictly increasing for $0 < t < \varepsilon$. So, even if $S(0)$ has a non-trivial kernel, it becomes non-degenerate for sufficiently small small $t > 0$. The same argument shows that this is true also for $t < 0$. □

Observe that, if $\Lambda = J(0)$, then $S(0) = 0$ in any chart given by the splitting $\Sigma = J(0) \oplus \Pi$. Therefore, the proof of Lemma 3.5 implies that all the eigenvalues of $S(t)$ are strictly non-zero for all $|t| < \varepsilon$, $t \neq 0$. If the curve is also monotone and ample, the only restriction on ε comes from the fact that $J(t)$ must belong to the given coordinate chart. In particular, the eigenvalues of $S(t)$ are strictly positive for all $t > 0$ (and strictly negative for $t < 0$) at least until the first intersection of $J(\cdot)$ with Π occurs. This means that $J(\cdot)$ cannot have further intersections with $J(0)$ until it crosses Π. Thus, we obtain the following.

Corollary 3.6. *Let $J(\cdot) \in \mathcal{L}(\Sigma)$ a monotone, ample curve, such that $J(\cdot) \cap \Pi = 0$, for some Lagrangian subspace Π. Then $J(\cdot)$ has no self-intersections, namely $J(t_1) \cap J(t_2) = 0$ for all $t_1 \neq t_2$.*

3.2.2 Reduction

Let (Σ, ω) be a symplectic vector space, and let $\Gamma \subset \Sigma$ be an isotropic subspace, namely $\omega|_\Gamma \equiv 0$. For any subspace $V \subset \Sigma$, we denote by the symbol $V^\angle$ the corresponding ω-orthogonal subspace.

Definition 3.7. The reduction of (Σ, ω) with respect to an isotropic subspace Γ is the symplectic space (Σ^Γ, ω), where

$$\Sigma^\Gamma := \Gamma^\angle / \Gamma.$$

The definition is well posed, since ω descends to a well-defined symplectic form on the quotient. Moreover, if $\dim \Sigma = 2n$ and $\dim \Gamma = k$, then Σ^Γ is a $2(n-k)$-dimensional symplectic space.

The projection $\pi^\Gamma : \mathcal{L}(\Sigma) \to \mathcal{L}(\Sigma^\Gamma)$, defined by $\Lambda \mapsto \Lambda \cap \Gamma^\angle / \Gamma$, is not even continuous in general. Nevertheless, the following lemma holds true.

Lemma 3.8. *The restriction of π^Γ to $\Gamma^\pitchfork := \{\Lambda \in \mathcal{L}(\Sigma) |\, \Lambda \cap \Gamma = 0\}$ is smooth.*

Proof. Let $\Lambda \in \Gamma^\pitchfork$. We can always find a Lagrangian space Π which contains Γ and such that $\Pi \cap \Lambda = 0$. The proof is now trivial in charts given by a Darboux basis on the splitting $\Pi \oplus \Lambda$. Indeed, in these charts, the projection corresponds to take a $n - k \times n - k$ block of the representative matrix. □

The next lemma provides condition under which a monotone, ample Jacobi curve remains monotone and ample upon projection.

Lemma 3.9. *Let $J(\cdot) \in \mathcal{L}(\Sigma)$ a monotone, ample (at t_0) curve such that $J(\cdot) \in \Gamma^{\pitchfork}$. Then the projection $J^\Gamma(\cdot) := \pi^\Gamma(J(\cdot))$ is a monotone, ample (at t_0) curve in $\mathcal{L}(\Sigma^\Gamma)$.*

Proof. By Lemma 3.8, the projection $J^\Gamma(\cdot)$ is still smooth. We prove the Lemma by analysing the coordinate presentation of the curve. Without loss of generality, we choose $t_0 = 0$. We find $\Pi \in \mathcal{L}(\Sigma)$ such that $\Gamma \subset \Pi$, and $\Sigma = J(0) \oplus \Pi$. Therefore, we introduce Darboux coordinates $(p, x) \in \mathbb{R}^{2n}$ such that, for small t

$$\Pi = \{(0, x) |\, x \in \mathbb{R}^n\}, \qquad J(t) = \{(p, S(t)p) |\, p \in \mathbb{R}^n\}, \qquad S(0) = 0.$$

Moreover, if $\dim \Gamma = k$, we split $\mathbb{R}^n = \mathbb{R}^k \oplus \mathbb{R}^{n-k}$, and we write $x = (x_1, x_2)$ and $p = (p_1, p_2)$. Thus

$$\Gamma = \{((0, 0), (x_1, 0)) |\, x_1 \in \mathbb{R}^k\}, \qquad \Gamma^\angle = \{((0, p_2), (x_1, x_2)) |\, p_2, x_2 \in \mathbb{R}^{n-k},\ x_1 \in \mathbb{R}^k\}.$$

Accordingly, the matrix $S(t)$ splits as $\begin{pmatrix} S_{11}(t) & S_{12}(t) \\ S_{12}^*(t) & S_{22}(t) \end{pmatrix}$. In terms of these coordinates, and analogous coordinates on $\Sigma^\Gamma = \Gamma^\angle/\Gamma$, we obtain that the matrix representing the reduced curve is $S^\Gamma(t) := S_{22}(t)$ (which is a $n - k \times n - k$ symmetric matrix). More precisely

$$J^\Gamma(t) = \{(p_2, S_{22}(t)p_2) |\, p_2 \in \mathbb{R}^{n-k}\}.$$

The original curve is monotone (say non-decreasing), then $\dot{S}(t) \geq 0$. Therefore, also $\dot{S}_{22}(t) \geq 0$, and $J^\Gamma(\cdot)$ is monotone too.

We now prove that the reduced curve is still ample at 0 if the original curve was. We assume $S(t)$ to be real-analytic, otherwise, it is sufficient to replace $S(t)$ with its Taylor polynomial of sufficiently high order. From the proof of Lemma 3.5, $S(t) > S(0) = 0$ for $t > 0$ sufficiently small. Thus, for all $y \in \mathbb{R}^{n-k}$, the function $t \mapsto y^* S_{22}(t) y$ is zero at $t = 0$, and strictly positive for $t > 0$. But an analytic function with these properties has at least a non-vanishing (strictly positive) derivative. Hence, for some $i > 0$, $y^* S_{22}^{(i)}(0) y > 0$. Since this construction holds for any $y \in \mathbb{R}^{n-k}$, this implies

$$\operatorname{rank}\{\dot{S}_{22}(0), \ldots, S_{22}^{(N)}\} = n - k,$$

for some sufficiently large $N > 0$. □

3.2.3 Maslov index and conjugate times

In this section we review a very useful homotopy invariant of curves in the Lagrange Grassmannian: the Maslov index, that is the intersection number of a curve with a certain pseudo-manifold in $\mathcal{L}(\Sigma)$. There are many things called Maslov index in different contexts, for a modern review we suggest [25]. Here we follow mainly the approach in [3] and [11].

Let $\Pi \in \mathcal{L}(\Sigma)$, consider the following subset of $\mathcal{L}(\Sigma)$,

$$\mathcal{M}_\Pi = \mathcal{L}(\Sigma) \setminus \Pi^{\pitchfork} = \{\Lambda \in \mathcal{L}(\Sigma) |\, \Lambda \cap \Pi \neq 0\},$$

which is called the *train* of the Lagrangian subspace Π due to V. Arnold. To see how $\mathcal{M}_\Pi$ looks locally, let $\Delta \in \mathcal{L}(\Sigma)$, $\Delta \cap \Pi = 0$. In coordinates induced by the splitting $\Sigma = \Pi \oplus \Delta$,

$$\Delta^{\pitchfork} = \{(p, Sp) | \, p \in \Pi \simeq \mathbb{R}^n, \, S \in Q(\mathbb{R}^n)\}.$$

Therefore, in coordinates,

$$\Delta^{\pitchfork} \setminus \Pi^{\pitchfork} \simeq \{S \in Q(\mathbb{R}^n) | \, \ker S \neq 0\}.$$

Hence the intersection of $\mathcal{M}_\Pi$ with the coordinate neighbourhood $\Delta^{\pitchfork}$ coincides with the set of all degenerate quadratic forms on $\mathbb{R}^n$. Notice that to a subspace Λ, which has k-dimensional intersection with Π there corresponds a form with k-dimensional kernel. The set of degenerate forms constitute an algebraic hypersurface in the space of all quadratic forms $Q(\mathbb{R}^n)$.

We want to define the intersection number of a curve in $\mathcal{L}(\Sigma)$ with the train $\mathcal{M}_\Pi$. To do this, we need a "co-orientation" on $\mathcal{M}_\Pi$, so first we start describing its singular locus. We see that a point $\Lambda \in \mathcal{M}_\Pi$ is singular if its associated quadratic form has at least two-dimensional kernel, so $\mathcal{M}_\Pi$ is an algebraic hypersurface in $\mathcal{L}(\Sigma)$ and its singular locus is an algebraic subset of codimension three in $\mathcal{M}_\Pi$. Thus $\mathcal{M}_\Pi$ is a pseudo-manifold.

Now, let us define a *canonical co-orientation* of the hypersurface $\mathcal{M}_\Pi$ at a non-singular point Λ, i.e. we indicate the "positive and negative sides" of $\mathcal{M}_\Pi$ in $\mathcal{L}(\Sigma)$. It is not difficult to see that vectors from $T_\Lambda \mathcal{L}(\Sigma)$ corresponding to positive definite and negative definite quadratic forms on Λ are not tangent to $\mathcal{M}_\Pi$, then we have the following.

Definition 3.10. Let Λ be a non-singular point of $\mathcal{M}_\Pi$. We consider as positive (negative) that side of $\mathcal{M}_\Pi$ towards which the positive (negative) definite elements of $T_\Lambda \mathcal{L}(\Sigma)$ are directed.

We say that a curve $J(\cdot)$ in $\mathcal{L}(\Sigma)$ is in *general position* (with respect to $\mathcal{M}_\Pi$) if $J(\cdot)$ intersects the non-singular locus of $\mathcal{M}_\Pi$ smoothly and transversally. The above co-orientation permits to define correctly the intersection number (or Maslov index) of a continuous curve in general position, with endpoints outside $\mathcal{M}_\Pi$, with the hypersurface $\mathcal{M}_\Pi$.

Definition 3.11. Let $J(t)$, $t_0 \leq t \leq t_1$ be a continuous curve in general position in $\mathcal{L}(\Sigma)$ with respect to the train $\mathcal{M}_\Pi$ such that $J(t_0), J(t_1) \notin \mathcal{M}_\Pi$. The Maslov index $J_{[t_0,t_1]} \cdot \mathcal{M}_\Pi$ is the number of points where $J(\cdot)$ intersects $\mathcal{M}_\Pi$ in the positive direction minus the number of points where this curves intersects $\mathcal{M}_\Pi$ in the negative direction.

A crucial property of the Maslov index is that it is a homotopy invariant of the curve, indeed a homotopy between curves in general position that leaves fixed the endpoints does not change the Maslov index. The proof of this fact is the same as for usual intersection number of a curve with a closed oriented hypersurface (see e.g. [53]). Notice that, since the singular locus of $\mathcal{M}_\Pi$ has codimension three, the generic homotopy moves the curve in general position. Thus, the Maslov index of any curve with endpoints not in $\mathcal{M}_\Pi$ is defined by putting the curve in general position.

Definition 3.12. Let $J(t)$, $t_0 \leq t \leq t_1$ be a continuous curve (not necessarily in general position) in $\mathcal{L}(\Sigma)$ such that $J(t_0), J(t_1) \notin \mathcal{M}_\Pi$. The Maslov index $J_{[t_0,t_1]} \cdot \mathcal{M}_\Pi$ is defined as $J'_{[t_0,t_1]} \cdot \mathcal{M}_\Pi$, where $J'(t)$, $t_0 \leq t \leq t_1$ is any curve in general position homotopic to $J_{[t_0,t_1]}$, with the same endpoints.

A weak point of the definition of Maslov index is the necessity of putting the curve in general position. This does not look like a very efficient way to compute the intersection number since putting the curve in general position could imply the modification of maybe a nice object, but the fact that the Maslov index is homotopy invariant leads to a very simple and effective way to compute it.

Lemma 3.13. *Assume that the piece of curve $J_{[t_0,t_1]}$ belongs to the chart $\Delta^{\pitchfork}$, $\Delta \cap \Pi = J(t_0) \cap \Pi = J(t_1) \cap \Pi = 0$. Let $S(t_i)$ be the symmetric matrix representing the subspace $J(t_i)$ in coordinates given by the splitting $\Sigma = \Delta \oplus \Pi$, that is $J(t_i) = \{(p, S(t_i)p) | p \in \Pi \simeq \mathbb{R}^n\}$. Thus*

$$J_{[t_0,t_1]} \cdot \mathcal{M}_\Pi = \operatorname{ind} S(t_0) - \operatorname{ind} S(t_1),$$

where $\operatorname{ind} S$ *is the index of the quadratic form $z \mapsto z^* S z$, $z \in \mathbb{R}^n$.*

In general the whole curve is not contained in a chart, but we can split it into segments $J_{[\tau_i,\tau_{i+1}]}$, $i = 0, \dots, \ell$, in such a way that $J(\tau) \in \Delta_i^{\pitchfork}$ $\forall \tau \in [\tau_i, \tau_{i+1}]$, where $\Delta_i \cap \Pi = 0$, $i = 0, \dots, \ell$. Hence

$$J(\cdot) \cdot \mathcal{M}_\Pi = \sum_{i=0}^{\ell} J_{[\tau_i,\tau_{i+1}]} \cdot \mathcal{M}_\Pi.$$

Remark 3.14. In particular, if $J(\cdot)$ is a Jacobi curve (which is monotone and ample) then the absolute value of the Maslov index $J_{[t_0,t_1]} \cdot \mathcal{M}_\mathcal{V}$ is the number of conjugate times of $J(\cdot)$ counted with multiplicity in the interval $[t_0, t_1]$.

We finish this section with the two main propositions about the Maslov index we need in the following. The first one provides an estimate of the difference of the Maslov index of a curve with respect to two different trains.

Proposition 3.15. *Let $J(t)$, $t_0 \le t \le t_1$, be a continuous curve in $\mathcal{L}(\Sigma)$ and suppose that $\Pi, \Pi' \in \mathcal{L}(\Sigma)$ satisfy $\Pi \cap J(t_i) = \Pi' \cap J(t_i) = 0$, $i = 0, 1$. Then*

$$|J_{[t_0,t_1]} \cdot \mathcal{M}_\Pi - J_{[t_0,t_1]} \cdot \mathcal{M}_{\Pi'}| \le n.$$

In the second one we consider a continuous curve P_t in $Sp(\Sigma)$, i.e. a one-parameter subgroup of the group $Sp(\Sigma)$ of *symplectic transformations* of Σ and we estimate the difference of the indices between two curves generated by P_t with respect to the same train.

Proposition 3.16. *Let $P_t \in Sp(\Sigma)$, $t_0 \le t \le t_1$ be a continuous curve in $Sp(\Sigma)$, $P_{t_0} = \mathbb{I}$, and suppose $\Lambda, \Lambda' \in \mathcal{L}(\Sigma)$. Set $J(t) = P_t \Lambda$ and $J'(t) = P_t \Lambda'$. Then, for all $\Pi \in L(\Sigma)$ such that $\Pi \cap J(t_i) = \Pi \cap J'(t_i) = 0$, $i = 0, 1$, the following inequality holds*

$$|J_{[t_0,t_1]} \cdot \mathcal{M}_\Pi - J'_{[t_0,t_1]} \cdot \mathcal{M}_\Pi| \le n.$$

The proofs of Propositions 3.15 and 3.16 can be found in [11, Propositions 5, 6].

3.3 Main results

We start this section by defining, more precisely, the class of dynamical systems under investigation. Let (Σ, σ) be a symplectic vector space.

Definition 3.17. A LQ optimal control problem is a pair $(H, \mathcal{V})$, where $H : \Sigma \to \mathbb{R}$ is a quadratic form (the Hamiltonian) and $\mathcal{V} \subset \Sigma$ is a Lagrangian subspace, such that $H|_{\mathcal{V}} \geq 0$.

By choosing appropriate Darboux coordinates, $\Sigma = \mathbb{R}^{2n}$, $\omega = \Omega$, $\mathcal{V} = \{(p, 0)|\, p \in \mathbb{R}^n\}$ and the Hamiltonian is

$$H(p, x) = \frac{1}{2}(p, x)^* \mathbf{H} \begin{pmatrix} p \\ x \end{pmatrix}, \qquad \mathbf{H} = \begin{pmatrix} BB^* & A \\ A^* & Q \end{pmatrix}. \tag{3.3}$$

Thus, Definition 3.17 is a coordinate-free characterization of the systems introduced in Sec. 3.1. With the pair $(H, \mathcal{V})$ we associate the Jacobi curve $J(t) = e^{t\Omega\mathbf{H}}\mathcal{V}$, which is a smooth curve in the Lagrange Grassmannian $\mathcal{L}(\Sigma)$. The assumption $H|_{\mathcal{V}} \geq 0$ is equivalent to the monotonicity of $J(\cdot)$.

Lemma 3.18. *The Jacobi curve of the system $(H, \mathcal{V})$ is monotone and equiregular.*

Proof. Let $z \in J(t)$, then there exists $z_0 \in \mathcal{V}$ such that $z = e^{t\Omega\mathbf{H}}z_0$. The last formula also provides a smooth extension of z belonging to the Jacobi curve for times close to t. Then, by definition of the quadratic form $\dot{J}(t)$, we obtain

$$\dot{J}(t)[z] = \omega(z, \dot{z}) = \omega\left(e^{t\Omega\mathbf{H}}z_0, e^{t\Omega\mathbf{H}}\Omega\mathbf{H}z_0\right) = \omega(z_0, \Omega\mathbf{H}z_0) = -z_0^* BB^* z_0 \leq 0,$$

where we have used the fact that the Hamiltonian flow is a one-parameter group of symplectomorphisms. This proves that $\dot{J}(t) \leq 0$ as a quadratic form and the curve is monotone.

Now observe that $J(t + \varepsilon) = e^{t\Omega\mathbf{H}}J(\varepsilon)$. This imples, by definition of i-th extension, that

$$J^{(i)}(t) = e^{t\Omega\mathbf{H}}J^{(i)}(0), \qquad i \geq 0,$$

hence the i-th extensions have the same dimension for all t, and the curve is equiregular. □

Lemma 3.19. *The system $(H, \mathcal{V})$ is controllable if and only if the Jacobi curve $J(\cdot)$ is ample.*

Proof. By definition, the system $(H, \mathcal{V})$, which can be written as in Eq. (3.3), is controllable if

$$\operatorname{rank}(B, AB, \ldots, A^{m-1}B) = n.$$

It is sufficient to prove that this is equivalent to ampleness at $t = 0$, since ampleness at all t follows from the equiregularity of the curve. Indeed, for small t, $J(t) = \{(p, S(t)p)|\, p \in \mathbb{R}^n\}$. We explicitly compute $S(t)$ as follows. Observe that

$$J(t) = e^{t\Omega\mathbf{H}} \begin{pmatrix} p \\ 0 \end{pmatrix} = \begin{pmatrix} \phi_{11}(t) & \phi_{12}(t) \\ \phi_{21}(t) & \phi_{22}(t) \end{pmatrix} \begin{pmatrix} p \\ 0 \end{pmatrix}, \qquad p \in \mathbb{R}^n.$$

It is clear that $S(t) = \phi_{21}(t)\phi_{11}(t)^{-1}$. Then we can compute iteratively the derivatives of $S(t)$ at $t = 0$, and we obtain, for any $m > 0$

$$\operatorname{rank}\{\dot{S}(0), \ddot{S}(0), \ldots, S^{m-1}(0)\} = \operatorname{rank}\{B, AB, \ldots, A^{m-1}B\}.$$

Therefore controllability is equivalent to ampleness of the curve at $t = 0$. □

We employ the symbol $\mathcal{H}$ to denote the set of controllable dynamical systems $(H, \mathcal{V})$ or, with no risk of confusion, the associated Hamiltonian vector fields $\vec{H}$. Since the associated Jacobi curve is monotone, ample and equiregular, Lemma 3.5 and Corollary 3.6 apply. This has important consequences on conjugate times.

Definition 3.20. We say that $\Gamma \subset \Sigma$ is an $\vec{H}$-invariant subspace if $P_t(\Gamma) = \Gamma$ for all $t \in \mathbb{R}$.

Proposition 3.21. *Let $\vec{H} \in \mathcal{H}$. Suppose there exists an $\vec{H}$-invariant Lagrangian subspace $\Gamma \subset \Sigma$, then the Jacobi curve $J(\cdot)$ has no conjugate times.*

Proof. Indeed, by Lemma 3.5, the Jacobi curve remains transversal to Γ for all times. Then, by Corollary 3.6, the only intersection with $\mathcal{V} = J(0)$ can occur at $t = 0$. □

Notice that the Lagrangian hypothesis is crucial. Indeed, Proposition 3.21 is false if the $\vec{H}$-invariant subspace is simply isotropic.

3.3.1 Proof of Theorem 3.A

Now we are ready to prove Theorem 3.A. By "eigenvalues of the Hamiltonian" we will mean the eigenvalues of $\Omega\mathbf{H}$, that is the matrix representing the Hamiltonian vector field $\vec{H}$. We first discuss the case in which the Hamiltonian field is diagonalizable, and the two "extreme" cases:

(i) $\vec{H}$ has no purely imaginary eigenvalues.

(ii) $\vec{H}$ has only purely imaginary eigenvalues.

In case (i) we build an $\vec{H}$-invariant Lagrangian subspace of Σ. Then by Proposition 3.21 there are no conjugate times. In case (ii) we directly prove that there are infinitely many conjugate times. Then, by the reduction process introduced in Sec. 3.2.2, we reduce the "intermediate" cases to case (ii). Finally, we extend our construction to the general (non diagonalizable) case.

We start by recalling a very important property of the spectrum of Hamiltonian matrices as $\vec{H}$. Namely, if λ is an eigenvalue, then also $\pm\lambda, \pm\bar{\lambda}$ are eigenvalues with the same multiplicity, where the bar denotes complex conjugation. Then, eigenvalues always appear in pairs (if $\lambda = \beta$ or $\lambda = i\beta$ for $\beta \in \mathbb{R}$) or in quadruples otherwise.

We denote by $E_\lambda \subseteq \mathbb{R}^{2n}$ the real invariant subspace corresponding to the eigenvalues $\lambda, \bar{\lambda}$ of $\vec{H}$. This is the real vector space generated by the generalized eigenvectors ξ, $\bar{\xi}$ corresponding to the eigenvalues λ and $\bar{\lambda}$, respectively. More precisely

$$E_\lambda := \operatorname{span}\{u, v \in \mathbb{R}^{2n} \,|\, u + iv \in \ker(\vec{H} - \lambda\mathbb{I})^k,\ k \geq 0\}.$$

It is clear that $E_\lambda = E_{\bar{\lambda}}$.

Diagonalizable case

Throughout this section, we assume $\vec{H}$ to be diagonalizable. We first prove an important orthogonality property of the invariant subspaces.

Lemma 3.22. *Let λ and λ' be eigenvalues of $\vec{H}$ (not necessarily distinct). If $\lambda + \lambda' \neq 0$ and $\bar{\lambda} + \lambda' \neq 0$ then $E_\lambda \Omega E_{\lambda'} = 0$.*

Proof. Recall that $\vec{H} = -\Omega \mathbf{H}$ and $\Omega^2 = -\mathbb{I}$. Let ξ and ξ' be eigenvectors corresponding to λ and λ' respectively. Since $\Omega^2 = -\mathbb{I}$, we have $\xi' \mathbf{H} \xi = \lambda \xi' \Omega \xi$ and $\xi \mathbf{H} \xi' = \lambda' \xi \Omega \xi'$ so $(\lambda + \lambda')\xi\Omega\xi' = 0$. Analogously, we obtain $\xi' H \bar{\xi} = \bar{\lambda} \xi' \Omega \bar{\xi}$ and $\bar{\xi} H \xi' = \lambda' \bar{\xi} \Omega \xi'$. Then $(\bar{\lambda} + \lambda')\bar{\xi}\Omega\xi' = 0$. Since $\lambda + \lambda' \neq 0$ and $\bar{\lambda} + \lambda' \neq 0$ it follows that $E_\lambda \Omega E_{\lambda'} = 0$. □

Remark 3.23. The above result still holds if $\vec{H}$ is not diagonalizable (see [52, Lemma D.1, Chapter II]).

Remark 3.24. In particular if $\lambda = \alpha + i\beta$, with $\alpha \neq 0$ then $\Omega|_{E_\lambda} \equiv 0$, i.e. $E_{\alpha+i\beta}$ is isotropic if $\alpha \neq 0$.

It follows that the invariant subspaces associated with purely imaginary eigenvalues, non-purely imaginary eigenvalues, and E_0 are pairwise Ω-orthogonal. This, together with the non-degeneracy of Ω, implies the following decomposition in Ω-orthogonal symplectic subspaces

$$\mathbb{R}^{2n} = \underbrace{E_0 \oplus \left(\bigoplus_{\alpha \neq 0} E_{\alpha+i\beta} \right)}_{\text{non pure imaginary}} \oplus \underbrace{\left(\bigoplus_{\beta \neq 0} E_{i\beta} \right)}_{\text{pure imaginary}} .$$

In the following, with the term “pure imaginary eigenvalue” we understand all the eigenvalues $\lambda = i\beta$, with $\beta \neq 0$.

Lemma 3.25. *There exists an $\vec{H}$-invariant, Lagrangian subspace Γ_+ of the symplectic space $\bigoplus_{\substack{\lambda \text{ non pure} \\ \text{imaginary}}} E_\lambda$.*

Proof. If zero is not an eigenvalue of $\vec{H}$, we take $\Gamma_+ = \bigoplus_{\alpha>0} E_{\alpha+i\beta}$, which is $\vec{H}$-invariant by definition. If zero is an eigenvalue of $\vec{H}$, let us consider the corresponding invariant subspace E_0, with $\dim E_0 = 2m$. Choose an isotropic m-dimensional subspace $\Gamma_0 \subset E_0$ (which is indeed $\vec{H}$-invariant). Hence $\Gamma_+ = \Gamma_0 \oplus \bigoplus_{\alpha>0} E_{\alpha+i\beta}$ satisfies the required properties. □

Proposition 3.26. *Let $\vec{H} \in \mathcal{H}$. Assume that $\vec{H}$ is diagonalizable and has no pure imaginary eigenvalues. Then the Jacobi curve $J(\cdot)$ has no conjugate times.*

Proof. By Proposition 3.21 it is sufficient to find a Lagrangian $\vec{H}$-invariant subspace of Σ. Since, in this case, $\Sigma = \bigoplus_{\substack{\lambda \text{ non pure} \\ \text{imaginary}}} E_\lambda$, the result follows from Lemma 3.25. □

Proposition 3.27. *Let $\vec{H} \in \mathcal{H}$. Suppose that $\vec{H}$ is diagonalizable and has a pure imaginary spectrum. Then the Jacobi curve $J(\cdot)$ has infinitely many conjugate times.*

Proof. If $\vec{H}$ has only pure imaginary eigenvalues, it is well known (see e.g. [20, Appendix 6]) that there exists a symplectic change of coordinates such that the Hamiltonian can be written as

$$H(p,x) = \frac{1}{2}\sum_{j=1}^{n} \omega_j (p_j^2 + x_j^2), \qquad \omega_1 \geq \omega_2 \geq \cdots \geq \omega_n. \tag{3.4}$$

Notice that the eigenvalues of $\vec{H}$ are $\pm i\omega_j$, $j = 1, \ldots, n$. The signs of the ω_j are precisely the signs of H on the real eigenspaces $E_{i\omega_j}$. The following two lemmas are crucial.

Lemma 3.28 (Givental' [42])**.** *There exists a Lagrangian subspace $\Lambda \subset \mathbb{R}^{2n}$ such that $H|_\Lambda > 0$ if and only if $\omega_j + \omega_{n-j+1} > 0$, $j = 1, \ldots, n$.*

Lemma 3.29 (Faĭbusovich [38])**.** *Under the controllability assumption (or, equivalently, the ampleness of the Jacobi curve), there exists a Lagrangian subspace $\Lambda \subset \mathbb{R}^{2n}$ such that $H|_\Lambda > 0$.*

Lemmas 3.29 and 3.28 imply the following inequality:

$$\sum_{j=1}^{n} \omega_j > 0. \tag{3.5}$$

Now, let us define a new curve $L(t) := P_t(L_0)$ in $\mathcal{L}(\mathbb{R}^{2n})$, where $L_0 := \{(p,0) : p \in \mathbb{R}^n\} \subset \mathbb{R}^{2n}$, $L_0 \in \mathcal{L}(\mathbb{R}^{2n})$.

Remark 3.30. Notice that, in order to bring the Hamiltonian to the normal form of Eq. (3.4), we have done a symplectic change of basis. Thus, in general, $L_0 \neq \mathcal{V}$.

If we reorder coordinates in such a way that $(p,x) \mapsto (p_1, x_1, \ldots, p_n, x_n)$, we can write

$$L(t) = \begin{pmatrix} r(t\omega_1) & & \\ & \ddots & \\ & & r(t\omega_n) \end{pmatrix} L_0,$$

where $r(t\omega_j)$ is a rotation of angle $t\omega_j$ in the 2-dimensional subspace (p_j, x_j). Observe that, given $t > 0$ we can choose $\varepsilon > 0$ sufficiently small such that $L(\varepsilon) \cap L_0 = L(t+\varepsilon) \cap L_0 = 0$. Therefore the Maslov index $L_{[\varepsilon,t+\varepsilon]} \cdot \mathcal{M}_{L_0}$ is well defined, since the endpoints of the curve are transversal to the train. We employ the shorthand $L_{(0,t)} \cdot \mathcal{M}_{L_0} = L_{[\varepsilon,t+\varepsilon]} \cdot \mathcal{M}_{L_0}$, for any ε sufficiently small, and similar notation is understood every time a small variation of the end-times is required.

We now prove that the index $L_{(0,+\infty)} \cdot \mathcal{M}_{L_0}$ is infinite. Intersections with the train occur at each half-rotation in each 2-dimensional subspace (p_j, x_j), with a sign given by the sign of ω_j. Therefore, by a direct computation, we have

$$L_{(0,T)} \cdot \mathcal{M}_{L_0} = \sum_{j=1}^{n} \lfloor \frac{T\omega_j}{\pi} \rfloor > \sum_{j=1}^{n} \frac{T\omega_j}{\pi} - n.$$

Inequality (3.5) implies that there are no compensations in the sum of the signs in the computation of the Maslov index. Indeed, let $N > 0$ fixed. Since $\sum_{j=1}^n \omega_j > 0$ we can take $T \geq \frac{(N+n)\pi}{\sum_{j=1}^n \omega_j}$ so that

$$L_{(0,T)} \cdot \mathcal{M}_{L_0} > N.$$

This implies that the Maslov index of the curve $L(t) = P_t(L_0)$ with the train $\mathcal{M}_{L_0}$ grows to infinity for $T \to \infty$. On the other hand, the number of conjugate times (counted with multiplicity) is the Maslov index of the Jacobi curve $J(t) = P_t(\mathcal{V})$ with the train $\mathcal{M}_{\mathcal{V}}$. Thus, by combining Proposition 3.15 and 3.16, we obtain

$$|J_{(0,T)} \cdot \mathcal{M}_{\mathcal{V}} - L_{(0,T)} \cdot \mathcal{M}_{L_0}| \leq 2n.$$

Therefore

$$J_{(0,T)} \cdot \mathcal{M}_{\mathcal{V}} > \frac{\sum_{j=1}^n \omega_j}{\pi} T - 3n.$$

Thus $J(\cdot)$ has infinitely many conjugate times. □

As a corollary of the proof of Proposition 3.27, we can give an estimate for the first conjugate time of a LQ optimal control problem.

Corollary 3.31. *Suppose the Hamiltonian can be written as* $H(p,x) = \frac{1}{2}\sum_{j=1}^n \omega_j(p_j^2 + x_j^2)$*. Then, if* $T \geq \frac{(N+3n-1)\pi}{\sum_{j=1}^n \omega_j}$ *there are at least* N *conjugate times (counted with multiplicity) in the interval* $(0,T]$*. In particular, the first conjugate time* t_c *satisfies* $t_c \leq \frac{3n\pi}{\sum_{j=1}^n \omega_j}$*.*

Now we are ready to discuss the case in which both pure and non pure imaginary eigenvalues occur in the spectrum of $\vec{H}$.

Proposition 3.32. *Let* $\vec{H} \in \mathcal{H}$*. Assume that* $\vec{H}$ *is diagonalizable and has at least one pure imaginary eigenvalue. Then the associated Jacobi curve* $J(\cdot)$ *has infinitely many conjugate times.*

Proof. We reduce the problem to the extremal case of Proposition 3.27. Consider Γ_+ as in Lemma 3.25, and let $\dim \Gamma = k$ (we drop the index $+$ from now on). Recall that Γ is an $\vec{H}$-invariant isotropic subspace of $\Sigma = \mathbb{R}^{2n}$. We will consider the Lagrange Grassmannian of the reduced space $\Sigma^\Gamma = \Gamma^\angle/\Gamma$. Notice that, by Lemma 3.5, the Jacobi curve remains transversal to Γ for all times. Thus, by Lemma 3.9, the reduced Jacobi curve $J^\Gamma(\cdot)$ is a smooth, ample, monotone curve in $\mathcal{L}(\Sigma^\Gamma)$. By construction, we have

$$\Gamma^\angle = \Gamma \oplus \bigoplus_{\substack{\lambda \text{ pure} \\ \text{imaginary}}} E_\lambda, \qquad \Sigma^\Gamma = \bigoplus_{\substack{\lambda \text{ pure} \\ \text{imaginary}}} E_\lambda.$$

Therefore we reduced the problem to the case of purely imaginary spectrum, and we can apply Proposition 3.27 to conclude that $J^\Gamma(\cdot)$ has infinitely many conjugate times. Notice that conjugate times for $J^\Gamma(\cdot)$ are intersections with $\mathcal{V}^\Gamma := \pi(\mathcal{V}) = (\Gamma^\angle \cap \mathcal{V})/\Gamma$. This means that the original curve has infinitely many intersections with $\mathcal{V}^\Gamma \oplus \Gamma$. More precisely, as we obtained in the proof of Proposition 3.27, and recalling that $\dim \Sigma^\Gamma = 2(n-k)$ we have

$$J_{(0,T)} \cdot \mathcal{M}_{\mathcal{V}^\Gamma \oplus \Gamma} > \frac{\sum_{j=1}^{n-k} \omega_j}{\pi} T - 3(n-k).$$

By applying again Proposition 3.15, we obtain

$$|J_{(0,T)} \cdot \mathcal{M}_{\mathcal{V}} - J_{(0,T)} \cdot \mathcal{M}_{\mathcal{V}^\Gamma \oplus \Gamma}| \leq n.$$

Therefore

$$J_{(0,T)} \cdot \mathcal{M}_{\mathcal{V}} > \frac{\sum_{j=1}^{n-k} \omega_j}{\pi} T - 4n + 3k.$$

Then $J(\cdot)$ has infinitely many conjugate times as well. □

Again, we give an estimate for the number of conjugate times as a separate corollary.

Corollary 3.33. *Suppose the Hamiltonian, restricted to* $\bigoplus_{\substack{\lambda \text{ pure} \\ \text{imaginary}}} E_\lambda$*, can be written as* $H(p,x) = \frac{1}{2}\sum_{j=1}^{n-k} \omega_j (p_j^2 + x_j^2)$*. Then if* $T \geq \frac{(N+4n-3k-1)\pi}{\sum_{j=1}^{n-k} \omega_j}$*, there are at least* N *conjugate times (counted with multiplicity) in the interval* $(0,T]$*. In particular, the first conjugate time* t_c *satisfies* $t_c \leq \frac{(4n-3k)\pi}{\sum_{j=1}^{n-k} \omega_j}$.

General case

Now, let us consider an arbitrary $\vec{H}$. We approach the problem with the same basic techniques devised for the diagonalizable case. Let $\lambda = i\beta$, $\beta \neq 0$ a pure imaginary eigenvalue of $\vec{H}$. Recall that, by Lemma 3.22, $E_{i\beta}$ is Ω-orthogonal to all the others $E_{\lambda'}$, with $\lambda' \neq \pm i\beta$. Therefore $E_{i\beta}$ is symplectic. It is well known that there exists a symplectic change of coordinates on $E_{i\beta}$ such that the Hamiltonian $H|_{E_{i\beta}}$ has one of the following normal forms (see [20, Appendix 6]).

(a) If $\pm i\beta$ correspond to a pair of Jordan blocks of even order $2k$:

$$H(p,x) = \pm\frac{1}{2}\left[\sum_{j=1}^{k}\left(\frac{1}{\beta^2}x_{2j-1}x_{2k-2j+1} + x_{2j}x_{2k-2j+2}\right) - \beta^2\sum_{j=1}^{k} p_{2j-1}x_{2j} + \right.$$
$$\left. + \sum_{j=1}^{k} p_{2j}x_{2j-1} - \sum_{j=1}^{k-1}\left(\beta^2 p_{2j+1}p_{2k-2j+1} + p_{2j+2}p_{2k-2j+2}\right)\right]. \quad (3.6)$$

(b) If $\pm\lambda$ correspond to a pair of Jordan blocks of odd order $2k+1$:

$$H(p,x) = \pm\frac{1}{2}\left[\sum_{j=1}^{k}\left(\beta^2 p_{2j}p_{2k-2j+2} + x_{2j}x_{2k-2j+2}\right) - \sum_{j=1}^{2k} p_j x_{j+1} - \right.$$
$$\left. - \sum_{j=1}^{k+1}\left(\beta^2 p_{2j-1}p_{2k-2j+3} + x_{2j-1}x_{2k-2j+3}\right)\right]. \quad (3.7)$$

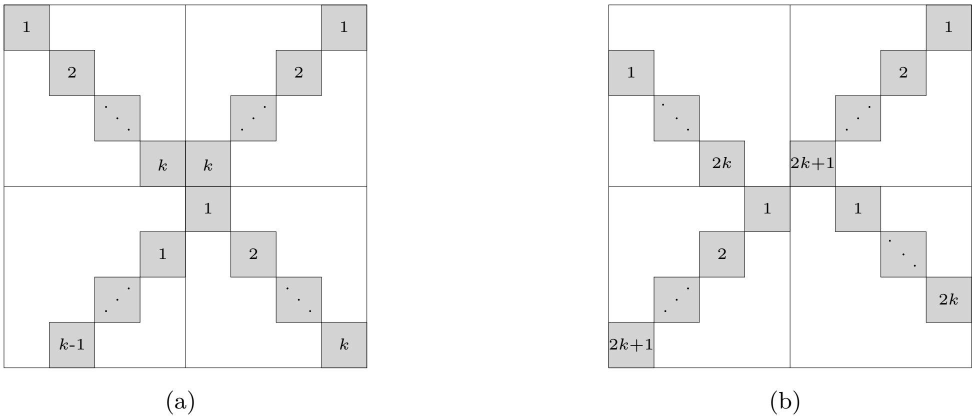


Figure 3.1: Block structure of the normal form of $\vec{H}|_{E_\lambda}$ for a pair of Jordan blocks of even order (case a) and odd order (case b). In case (a), $\dim E_\lambda = 4k$, and each box denotes the presence of a non-vanishing 2×2 block. In case (b), $\dim E_\lambda = 4k+2$, and each box denotes the presence of a non-vanishing 1×1 block. All other entries are zero.

Notice that the dimension of E_λ is $4k$ or $4k+2$, respectively.

Lemma 3.34. *Let $\lambda = i\beta$ a pure imaginary eigenvalue of $\vec{H}$. Thus*

(a) If the Jordan block corresponding to λ has even order $2k$ then there exists a Lagrangian $\vec{H}$-invariant subspace $\Gamma \subset E_\lambda$ (of dimension $2k$).

(b) If the Jordan block corresponding to λ has odd order $2k+1$ then there exists an isotropic $\vec{H}$-invariant subspace $\Gamma \subset E_\lambda$ of dimension $2k$.

Proof. Let us consider the first case. As we recall above, $H|_{E_\lambda}$ can be written as in Eq. (3.6). Then, a careful inspection shows that $\vec{H}|_{E_\lambda} = -\Omega\mathbf{H}|_{E_\lambda}$ has the structure, in coordinates $(p,x) \in \mathbb{R}^{4k}$, displayed in Fig. 3.1(a). Notice that, for what follows, we do not need to know the explicit form of each box. If k is even, we choose $\Gamma = \{(p,x) \in \mathbb{R}^{4k} | \, p_{k+1} = \ldots = p_{2k} = x_1 = \ldots = x_k = 0\}$ and if k is odd we set $\Gamma = \{(p,x) \in \mathbb{R}^{4k} | \, p_{k+2} = \ldots = p_{2k} = x_1 = \ldots = x_{k+1} = 0\}$. It is a simple check that, in both cases, Γ is a $2k$-dimensional $\vec{H}$-invariant space, which is also isotropic by construction, and thus Lagrangian (since $\dim E_\lambda = 4k$).

Now, suppose that the Jordan block corresponding to λ has odd order $2k+1$. Thus $H|_{E_\lambda}$ can be written as in Eq. (3.7) and $\vec{H}|_{E_\lambda} = -\Omega\mathbf{H}|_{E_\lambda}$ has the structure, in coordinates $(p,x) \in \mathbb{R}^{4k+2}$, displayed in Fig. 3.1(b). Once again, we stress that we do not need the explicit form of each box. By choosing $\Gamma = \{(p,x) \in \mathbb{R}^{2n} | \, p_1 = \ldots = p_{k+1} = x_{k+1} = \ldots = x_{2k+1} = 0\}$, we get the required subspace. □

Proposition 3.35. *Let $\vec{H} \in \mathcal{H}$. If all Jordan blocks of $\vec{H}$ corresponding to pure imaginary eigenvalues are of even order, the Jacobi curve has no conjugate times.*

Proof. By Lemma 3.21 it is enough to find an $\vec{H}$-invariant Lagrangian subspace $\Gamma \subset \Sigma$. Let $\pm\lambda_1, \ldots, \pm\lambda_m$ be the pure imaginary eigenvalues of $\vec{H}$. By Lemma 3.34 there exists a Lagrangian

$\vec{H}$-invariant subspace $\Gamma_i \subset E_{\lambda_i}$ for each i. Set $\Gamma = \Gamma_1 \oplus \cdots \oplus \Gamma_m \oplus \Gamma_+$, where $\Gamma_+ \subset \bigoplus_{\substack{\lambda \text{ non pure} \\ \text{imaginary}}} E_\lambda$ is as in Lemma 3.25. □

Proposition 3.36. *Let $\vec{H} \in \mathcal{H}$. Suppose there exists at least one Jordan block of odd order corresponding to a pure imaginary eigenvalue of $\vec{H}$. Thus the Jacobi curve has infinitely many conjugate times.*

Proof. We will reduce the problem to the diagonalizable case by studying the curve in a reduced space $\Sigma^\Gamma = \Gamma^\angle/\Gamma$. Let $\pm\lambda_1, \ldots, \pm\lambda_m$ be the pure imaginary eigenvalues of $\vec{H}$ and let us consider, for each i, the quotient spaces $E_{\lambda_i}^{\Gamma_i} := E_{\lambda_i} \cap \Gamma_i^\angle/\Gamma_i$, where the subspaces $\Gamma_i \subset E_{\lambda_i}$ are as in Lemma 3.34. Notice that $\dim E_{\lambda_i}^{\Gamma_i} = 0$ or 2 depending on whether the Jordan block corresponding to λ_i is even or odd, respectively.

Now set $\Gamma = \Gamma_1 \oplus \cdots \oplus \Gamma_m \oplus \Gamma_+$ as in the previous proposition, hence $\Sigma^\Gamma = E_{\lambda_1}^{\Gamma_1} \oplus \cdots \oplus E_{\lambda_m}^{\Gamma_m}$, so if there is at least one λ_i for which the corresponding Jordan block has odd order then $\vec{H}|_{\Sigma^\Gamma}$ has nonempty pure imaginary spectrum and it is diagonalizable. Moreover, since the original Jacobi curve is ample and monotone, the reduced Jacobi curve $J^\Gamma(\cdot)$ is ample and monotone too by Lemma 3.9. Thus the result follows from Proposition 3.27. □

Chapter 4

A formula for Popp's volume in sub-Riemannian geometry

4.1 Introduction

The problem to define a canonical volume on a sub-Riemannian manifold was first pointed out by Brockett in his seminal paper [35], motivated by the construction of a Laplace operator on a 3D sub-Riemannian manifold canonically associated with the metric structure, analogous to the Laplace-Beltrami operator on a Riemannian manifold. Recently, Montgomery addressed this problem in the general case (see [56, Chapter 10]).

Even on a Riemannian manifold, the Laplacian (defined as the divergence of the gradient) is a second order differential operator whose first order term depends on the choice of the volume on the manifold, which is required to define the divergence. Naively, in the Riemannian case, the choice of a canonical volume is determined by the metric, by requiring that the volume of a orthonormal parallelotope (i.e. whose edges are an orthonormal frame in the tangent space) is 1.

From a geometrical viewpoint, sub-Riemannian geometry is a natural generalization of Riemannian geometry under non-holonomic constraints. Formally speaking, a sub-Riemannian manifold is a smooth manifold M endowed with a bracket-generating distribution $\mathscr{D} \subset TM$, with $k = \operatorname{rank} \mathscr{D} < n = \dim M$, and a smooth fibre-wise scalar product on $\mathscr{D}$. From this structure, one derives a distance on M - the so-called *Carnot-Caratheodory metric* - as the infimum of the length of *horizontal* curves on M, i.e. the curves that are almost everywhere tangent to the distribution.

Nevertheless, sub-Riemannian geometry enjoys major differences with respect to the Riemannian case. For instance, a construction analogue to the one described above for the Riemannian volume is not possible. Indeed the inner product is defined only on a subspace of the tangent space, and there is no canonical way to extend it on the whole tangent space.

Popp's volume is a generalization of the Riemannian volume in sub-Riemannian setting. It was first defined by Octavian Popp but introduced only in [56] (see also [10]). Such a volume is smooth only for an equiregular sub-Riemannian manifold, i.e. when the dimensions of the higher order distributions $\mathscr{D}^1 := \mathscr{D}$, $\mathscr{D}^{i+1} := \mathscr{D}^i + [\mathscr{D}^i, \mathscr{D}]$, for every $i \geq 1$, do not depend on the point (for precise definitions, see Sec. 2).

Under the equiregularity hypothesis, the bracket-generating condition guarantees that there exists a minimal $m \in \mathbb{N}$, called *step* of the structure, such that $\mathscr{D}^m = TM$.

Then, for each $q \in M$, it is well defined the graded vector space:

$$\mathrm{gr}_q(\mathscr{D}) := \bigoplus_{i=1}^{m} \mathscr{D}_q^i / \mathscr{D}_q^{i-1}, \qquad \text{where} \quad \mathscr{D}_q^0 = 0.$$

The vector space $\mathrm{gr}_q(\mathscr{D})$, which can be endowed with a natural sub-Riemannian structure, is called the *nilpotentization* of the structure at the point q, and plays a role analogous to the Euclidean tangent space in Riemannian geometry. Popp's volume is defined by inducing a canonical inner product on $\mathrm{gr}_q(\mathscr{D})$ via the Lie brackets, and then using a *non-canonical* isomorphism between $\mathrm{gr}_q(\mathscr{D})$ and T_qM to define an inner product on the whole T_qM. Interestingly, even though this construction depends on the choice of some complement to the distribution, the associated volume form (i.e. Popp's volume) is independent on this choice.

It is worth to recall that on a sub-Riemannian manifold, which is a metric space, the Hausdorff volume and the spherical Hausdorff volume, respectively $\mathcal{H}^Q$ and $\mathcal{S}^Q$, are canonically defined.[1] The relation between Popp's volume and $\mathcal{S}^Q$ has been studied in [7], where the authors show how the Radon-Nikodym derivative is related with the nilpotentization of the structure. In particular they prove that the Radon-Nikodym derivative could also be non smooth (see also [23, 32]). Remember that the Hausdorff and spherical Hausdorff volumes are both proportional to the Riemannian one on a Riemannian manifold. The relation between Hausdorff measures for curves and different notions of length in sub-Riemannian geometry is also investigated in [41].

On a contact sub-Riemannian manifold, Popp's volume coincides with the Riemannian volume obtained by "promoting" the Reeb vector field to an orthonormal complement to the distribution. In the general case, unfortunately, the definition is more involved. To the authors' best knowledge, explicit formulæ for Popp's volume appeared, for some specific cases, only in [7, 23, 32].

The goal of this chapter is to prove a general formula for Popp's volume, in terms of any *adapted* frame of the tangent bundle. In order to present the main results here, we briefly introduce some concepts which we will elaborate in details in the subsequent sections. Thus, 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 $\mathscr{D}^i$, where $k_i := \dim \mathscr{D}^i$, 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

$$[X_i, X_j] = \sum_{l=1}^{n} c_{ij}^l X_l. \tag{4.1}$$

With a standard abuse of notation we call them *structure constants*. For $j = 2, \dots, m$ we define

[1] Recall that the Hausdorff dimension of a sub-Riemannian manifold M is given by the formula $Q = \sum_{i=1}^m i n_i$, where $n_i := \dim \mathscr{D}_q^i / \mathscr{D}_q^{i-1}$. In particular the Hausdorff dimension is always bigger than the topological dimension.

the *adapted structure constants* $b^l_{i_1\dots i_j} \in C^\infty(M)$ as follows:

$$[X_{i_1},[X_{i_2},\dots,[X_{i_{j-1}},X_{i_j}]]] = \sum_{l=k_{j-1}+1}^{k_j} b^l_{i_1 i_2\dots i_j} X_l \mod \mathscr{D}^{j-1}, \tag{4.2}$$

where $1 \leq i_1,\dots,i_j \leq k$. These are a generalization of the c^l_{ij}, with an important difference: the structure constants of Eq. (4.1) 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. (4.2) 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 $\mathscr{D}$), and considering the appropriate equivalence class. For $j = 2$, the adapted structure constants can be directly compared to the standard ones. Namely $b^l_{ij} = c^l_{ij}$ when both are defined, that is for $1 \leq i,j \leq k$, $l \geq k+1$.

Then, we define the $k_j - k_{j-1}$ dimensional square matrix B_j as follows:

$$[B_j]^{hl} = \sum_{i_1,i_2,\dots,i_j=1}^{k} b^h_{i_1 i_2\dots i_j} b^l_{i_1 i_2\dots i_j}, \qquad j = 1,\dots,m, \tag{4.3}$$

with the understanding that B_1 is the $k \times k$ identity matrix. It turns out that each B_j is positive definite.

Theorem 4.A. *Let $X_1,\dots,X_n$ be a local adapted frame, and let $\nu^1,\dots,\nu^n$ be the dual frame. Then Popp's volume $\mathcal{P}$ satisfies*

$$\mathcal{P} = \frac{1}{\sqrt{\prod_j \det B_j}} \nu^1 \wedge \dots \wedge \nu^n, \tag{4.4}$$

where B_j is defined by (4.3) *in terms of the adapted structure constants* (4.2).

To clarify the geometric meaning of Eq. (4.4), let us consider more closely the case $m = 2$. If $\mathscr{D}$ is a step 2 distribution, we can build a local adapted frame $\{X_1,\dots,X_k,X_{k+1},\dots,X_n\}$ by completing any local orthonormal frame $\{X_1,\dots,X_k\}$ of the distribution to a local frame of the whole tangent bundle. Even though it may not be evident, it turns out that $B_2^{-1}(q)$ is the Gram matrix of the vectors $X_{k+1},\dots,X_n$, seen as elements of $T_qM/\mathscr{D}_q$. The latter has a natural structure of inner product space, induced by the surjective linear map $[,] : \mathscr{D}_q \otimes \mathscr{D}_q \to T_qM/\mathscr{D}_q$ (see Lemma 4.12). Therefore, the function appearing at the beginning of Eq. (4.4) is the volume of the parallelotope whose edges are $X_1,\dots,X_n$, seen as elements of the orthogonal direct sum $\mathrm{gr}_q(\mathscr{D}) = \mathscr{D}_q \oplus T_qM/\mathscr{D}_q$.

With a volume form at disposal, one can naturally define the associated divergence operator, which acts on vector fields. Moreover, the sub-Riemannian structure allows to define the horizontal gradient of a smooth function. Then, we define a canonical sub-Laplace operator as $\Delta := \mathrm{div} \circ \nabla$, which generalizes the Laplace-Beltrami operator. This is a second order differential operator, which has been studied in [10, 21]. As a corollary to Theorem 4.A, we obtain a formula for the sub-Laplacian Δ in terms of any local adapted frame.

Corollary 4.1. *Let $X_1, \dots, X_n$ be a local adapted frame. Let Δ be the canonical sub-Laplacian. Then*

$$\Delta = \sum_{i=1}^{k} X_i^2 - \left(\frac{1}{2} \sum_{j=1}^{m} \mathrm{Tr}(B_j^{-1} X_i(B_j)) + \sum_{l=1}^{n} c_{il}^l \right) X_i,$$

where c_{ij}^l are the structure constants (4.1)*, and B_j is defined by* (4.3) *in terms of the adapted structure constants* (4.2)*.*

Remark 4.2. If M is a Carnot group (i.e. a connected, simply connected nilpotent group, whose Lie algebra is graded, and whose sub-Riemannian structure is left invariant) the B_j are constant. Moreover, $\forall i \ \sum_{l=1}^n c_{il}^l = 0$, as a consequence of the graded structure. Then, in this case, the sub-Laplacian is a simple "sum of squares" $\Delta = \sum_{i=1}^k X_i^2$. This is a manifestation of the fact that Carnot groups are to sub-Riemannian geometry as Euclidean spaces are to Riemannian geometry. Indeed, on $\mathbb{R}^n$, the Laplace-Beltrami operator is a simple sum of squares.

More in general, in [10], the authors prove that for left-invariant structures on unimodular Lie groups the sub-Laplacian is a sum of squares.

In the last part of the chapter 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 4.3. A (local) diffeomorphism $\phi : M \to M$ is a *(local) isometry* if its differential $\phi_* : TM \to TM$ preserves the sub-Riemannian structure $(\mathscr{D}, \langle \cdot | \cdot \rangle)$, namely

i) $\phi_*(\mathscr{D}_q) = \mathscr{D}_{\phi(q)}$ for all $q \in M$,

ii) $\langle \phi_* X | \phi_* Y \rangle_{\phi(q)} = \langle X | Y \rangle_q$ for all $q \in M$, $X, Y \in \mathscr{D}_q$.

Remark 4.4. 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. $\phi_*(\mathscr{D}_q^i) = \mathscr{D}_{\phi(q)}^i$, for $1 \leq i \leq m$.

Definition 4.5. 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 4.B. *Sub-Riemannian (local) isometries are volume preserving transformations for Popp's volume.*

Proposition 4.B may be false for volumes different than Popp's one. We have the following.

Proposition 4.C. *Let* Iso(M) *be the group of isometries of the sub-Riemannian manifold M. If* Iso(M) *acts transitively on M, then Popp's volume is the unique volume (up to multiplication by scalar constant) such that Proposition 4.B holds true.*

Definition 4.6. Let M be a Lie group. A sub-Riemannian structure $(M, \mathscr{D}, \langle\cdot|\cdot\rangle)$ 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 4.B we recover a well-known result (see again [56]).

Corollary 4.7. *Let $(M, \mathscr{D}, \langle\cdot|\cdot\rangle)$ be a left-invariant sub-Riemannian structure. Then Popp's volume is left invariant, i.e. $L_g^*\mathcal{P} = \mathcal{P}$ for every $g \in M$.*

Propositions 4.B, 4.C and Corollary 4.7 should shed some light about which is the "most natural" volume for sub-Riemannian manifold.

4.2 Sub-Riemannian geometry

We recall some basic definitions in sub-Riemannian geometry. For a more detailed introduction, see [6, 31, 56].

Definition 4.8. A *sub-Riemannian manifold* is a triple $(M, \mathscr{D}, \langle\cdot|\cdot\rangle)$, where

(i) M is a connected orientable smooth manifold of dimension $n \geq 3$;

(ii) $\mathscr{D} \subset TM$ is a smooth distribution of constant rank $k < n$;

(iii) $\langle\cdot|\cdot\rangle_q$ is an inner product on the fibres $\mathscr{D}_q$, smooth as a function of q.

Let $\Gamma(\mathscr{D}) \subset \mathrm{Vec}(M)$ be the $C^\infty(M)$-module of the smooth sections of $\mathscr{D}$. Throughout this chapter we assume that the sub-Riemannian manifold M satisfies the *bracket-generating condition*, i.e.

$$\mathrm{span}\{[X_1, [X_2, \dots, [X_{j-1}, X_j]]](q) \mid X_i \in \Gamma(\mathscr{D}),\, j \in \mathbb{N}\} = T_qM, \quad \forall\, q \in M. \tag{4.5}$$

In other words, the iterated Lie brackets of smooth sections of $\mathscr{D}$ span the whole tangent bundle TM. Condition (4.5) is also called *Hörmander condition*, and bracket-generating distribution are also referred to as *completely nonholonomic* distributions.

An absolutely continuous curve $\gamma : [0, T] \to M$ is said to be *horizontal* (or *admissible*) if

$$\dot\gamma(t) \in \mathscr{D}_{\gamma(t)} \qquad \text{for a.e. } t \in [0, T].$$

Given an horizontal curve $\gamma : [0, T] \to M$, the *length of* γ is

$$\ell(\gamma) = \int_0^T |\dot\gamma(t)|\, dt.$$

The *distance* induced by the sub-Riemannian structure on M is the function

$$d(q_0, q_1) = \inf\{\ell(\gamma) \mid \gamma(0) = q_0, \gamma(T) = q_1, \gamma \text{ horizontal}\}.$$

The connectedness of M and the bracket-generating condition guarantee the finiteness and the continuity of the sub-Riemannian distance with respect to the topology of M (Chow-Rashevsky

Theorem, see, for instance, [15]). The function $d(\cdot,\cdot)$ is called the *Carnot-Caratheodory distance* and gives to M the structure of metric space (see [31, 43]).

Locally (i.e. on an open set $U \subset M$), there always exists a set of k smooth vector fields $X_1, \dots, X_k$ such that, $\forall q \in U$, it is an orthonormal basis of $\mathscr{D}_q$. The set $\{X_1, \dots, X_k\}$ is called a *local orthonormal frame* for the sub-Riemannian structure.

Definition 4.9. Let $\mathscr{D}$ be a distribution. Its *flag at* $q \in M$ is the sequence of vector spaces $\mathscr{D}_q^0 \subset \mathscr{D}_q^1 \subset \mathscr{D}_q^2 \subset \ldots \subset T_qM$ defined by

$$\mathscr{D}_q^0 := \{0\}, \qquad \mathscr{D}_q^1 := \mathscr{D}_q, \qquad \mathscr{D}_q^{i+1} := \mathscr{D}_q^i + [\mathscr{D}^i, \mathscr{D}]_q,$$

where, with a standard abuse of notation, we understand that $[\mathscr{D}^i, \mathscr{D}]_q$ is the vector space generated by the iterated Lie brackets, up to length i, of local sections of the distribution, evaluated at q.

Even though the rank of $\mathscr{D}$ is constant, the dimensions of the subspaces of the flag, i.e. the numbers $k_i(q) := \dim(\mathscr{D}_q^i)$ may depend on the point. Observe that the bracket-generating condition can be rewritten as follows

$$\forall q \in M \;\; \exists \;\text{ minimal }\; m(q) \in \mathbb{N} \quad \text{ such that } \quad k_m(q) = \dim T_qM.$$

The number $m(q)$ is called the *step* of the distribution at the point q. The vector $\mathcal{G}(q) := (k_1(q), k_2(q), \dots, k_m(q))$ is called the *growth vector* of the distribution at q.

Definition 4.10. A distribution $\mathscr{D}$ is *equiregular* if the growth vector is constant, i.e. for each $i = 1, 2, \dots, m$, $k_i(q) = \dim(\mathscr{D}_q^i)$ does not depend on $q \in M$. In this case the subspaces $\mathscr{D}_q^i$ are fibres of the *higher order distributions* $\mathscr{D}^i \subset TM$.

For equiregular distributions we will simply talk about growth vector and step of the distribution, without any reference to the point q.

Finally, we introduce the nilpotentization of the distribution at the point q, which is fundamental for the definition of Popp's volume.

Definition 4.11. Let $\mathscr{D}$ be an equiregular distribution of step m. The *nilpotentization of* $\mathscr{D}$ at the point $q \in M$ is the graded vector space

$$\mathrm{gr}_q(\mathscr{D}) = \mathscr{D}_q \oplus \mathscr{D}_q^2/\mathscr{D}_q \oplus \ldots \oplus \mathscr{D}_q^m/\mathscr{D}_q^{m-1}.$$

The vector space $\mathrm{gr}_q(\mathscr{D})$ can be endowed with a Lie algebra structure, which respects the grading. Then, there is a unique connected, simply connected group, $\mathrm{Gr}_q(\mathscr{D})$, such that its Lie algebra is $\mathrm{gr}_q(\mathscr{D})$. The global, left-invariant vector fields obtained by the group action on any orthonormal basis of $\mathscr{D}_q \subset \mathrm{gr}_q(\mathscr{D})$ defines a sub-Riemannian structure on $\mathrm{Gr}_q(\mathscr{D})$, which is called the *nilpotent approximation* of the sub-Riemannian structure at the point q.

4.3 Popp's volume

In this section we provide the definition of Popp's volume, and we prove Theorem 4.A. Our presentation follows closely the one that can be found in [56]. The definition rests on the following lemmas.

Lemma 4.12. *Let E be an inner product space, and 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*

$$\|v\|_V = \min\{\|e\|_E \ s.t. \ \pi(e) = v\}. \tag{4.6}$$

Proof. It is easy to check that Eq. (4.6) 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 4.13. *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 [56, Lemma 10.4]. □

The idea behind Popp's volume is to define an inner product on each $\mathscr{D}^i_q/\mathscr{D}^{i-1}_q$ which, in turn, induces an inner product on the orthogonal direct sum $\mathrm{gr}_q(\mathscr{D})$. 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 4.13 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 $v, w \in \mathscr{D}_q$, and let V, W be any horizontal extensions of v, w. Namely, $V, W \in \Gamma(\mathscr{D})$ and $V(q) = v$, $W(q) = w$. The linear map $\pi : \mathscr{D}_q \otimes \mathscr{D}_q \to \mathscr{D}^2_q/\mathscr{D}_q$

$$\pi(v \otimes w) := [V, W]_q \mod \mathscr{D}_q,$$

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 $\mathscr{D}$

$$\widetilde{V} = V + \sum_{i=1}^k f_i X_i, \qquad \widetilde{W} = W + \sum_{i=1}^k g_i X_i,$$

where, for $1 \leq i \leq k$, $f_i, g_i \in C^\infty(M)$ and $f_i(q) = g_i(q) = 0$. Therefore

$$[\widetilde{V}, \widetilde{W}] = [V, W] + \sum_{i=1}^{k} (V(g_i) - W(f_i))\, X_i + \sum_{i,j=1}^{k} f_i g_j [X_i, X_j].$$

Thus, evaluating at q, $[\widetilde{V}, \widetilde{W}]_q = [V, W]_q \mod \mathscr{D}_q$, as claimed. Similarly, let $1 \leq i \leq m$. The linear maps $\pi_i : \otimes^i \mathscr{D}_q \to \mathscr{D}_q^i / \mathscr{D}_q^{i-1}$

$$\pi_i(v_1 \otimes \cdots \otimes v_i) = [V_1, [V_2, \ldots, [V_{i-1}, V_i]]]_q \mod \mathscr{D}_q^{i-1}, \tag{4.7}$$

are well defined and do not depend on the choice of the horizontal extensions $V_1, \ldots, V_i$ of $v_1, \ldots, v_i$.

By the bracket-generating condition, π_i are surjective and, by Lemma 4.12, they induce an inner product space structure on $\mathscr{D}_q^i / \mathscr{D}_q^{i-1}$. Therefore, the nilpotentization of the distribution at q, namely

$$\mathrm{gr}_q(\mathscr{D}) = \mathscr{D}_q \oplus \mathscr{D}_q^2 / \mathscr{D}_q \oplus \ldots \oplus \mathscr{D}_q^m / \mathscr{D}_q^{m-1},$$

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) $\mu_q \in \wedge^n \mathrm{gr}_q(\mathscr{D})^*$, 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 $\mathrm{gr}_q(\mathscr{D})$ to T_qM through the map $\theta_q : \wedge^n T_qM \to \wedge^n \mathrm{gr}_q(\mathscr{D})$ defined in Lemma 4.13. Namely

$$\mathcal{P}_q = \theta_q^*(\mu_q) = \mu_q \circ \theta_q, \tag{4.8}$$

where θ_q^* denotes the dual map and we employ the canonical identification $(\wedge^n T_qM)^* \simeq \wedge^n T_q^*M$. Eq. (4.8) 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 4.A.

Remark 4.14. The definition of Popp's volume can be restated as follows. Let $(M, \mathscr{D})$ be an oriented sub-Riemannian manifold. Popp's volume is the unique volume $\mathcal{P}$ such that, for all $q \in M$, the following diagram is commutative:

$$\begin{array}{ccc} (M, \mathscr{D}) & \xrightarrow{\mathcal{P}} & (\wedge^n T_qM)^* \\ {\scriptstyle \mathrm{gr}_q}\big\downarrow & & \big\downarrow{\scriptstyle \theta_q^*} \\ \mathrm{gr}_q(\mathscr{D}) & \xrightarrow[\mu]{} & (\wedge^n \mathrm{gr}_q(\mathscr{D}))^* \end{array}$$

where μ associates the inner product space $\mathrm{gr}_q(\mathscr{D})$ with its canonical volume μ_q, and θ_q^* is the dual of the map defined in Lemma 4.13.

4.3.1 Proof of Theorem 4.A

We are now ready to prove Theorem 4.A. 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 $\mathrm{gr}(\mathscr{D})$ we mean the nilpotentization of $\mathscr{D}$ at the point q, and by $\mathscr{D}^i$ we mean the fibre $\mathscr{D}_q^i$ of the appropriate higher order distribution.

Step 2 distribution

If $\mathscr{D}$ is a step 2 distribution, then $\mathscr{D}^2 = TM$. 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

$$[X_i, X_j] = \sum_{l=k+1}^{n} b_{ij}^l Y_l \mod \mathscr{D}.$$

For each $l = k+1, \ldots, n$, we can think to b_{ij}^l as the components of an Euclidean vector in $\mathbb{R}^{k^2}$, which we denote by the symbol b^l. According to the general construction of Popp's volume, we need first to compute the inner product on the orthogonal direct sum $\mathrm{gr}(\mathscr{D}) = \mathscr{D} \oplus \mathscr{D}^2/\mathscr{D}$. By Lemma 4.12, the norm on $\mathscr{D}^2/\mathscr{D}$ is induced by the linear map $\pi : \otimes^2\mathscr{D} \to \mathscr{D}^2/\mathscr{D}$

$$\pi(X_i \otimes X_j) = [X_i, X_j] \mod \mathscr{D}.$$

The vector space $\otimes^2\mathscr{D}$ inherits an inner product from the one on $\mathscr{D}$, namely $\forall X, Y, Z, W \in \mathscr{D}$, $\langle X \otimes Y, Z \otimes W\rangle = \langle X, Z\rangle\langle Y, W\rangle$. π is surjective, then we identify the range $\mathscr{D}^2/\mathscr{D}$ with $\ker \pi^\perp \subset \otimes^2\mathscr{D}$, and define an inner product on $\mathscr{D}^2/\mathscr{D}$ by this identification. In order to compute explicitly the norm on $\mathscr{D}^2/\mathscr{D}$ (and then, by polarization, the inner product), let $Y \in \mathscr{D}^2/\mathscr{D}$. Then

$$\|Y\|_{\mathscr{D}^2/\mathscr{D}} = \min\{\|Z\|_{\otimes^2\mathscr{D}} \text{ s.t. } \pi(Z) = Y\}. \tag{4.9}$$

Let $Y = \sum_{l=k+1}^n c^l Y_l$ and $Z = \sum_{i,j=1}^k a_{ij} X_i \otimes X_j \in \otimes^2\mathscr{D}$. We can think to a_{ij} as the components of a vector $a \in \mathbb{R}^{k^2}$. Then, Eq. (4.9) writes

$$\|Y\|_{\mathscr{D}^2/\mathscr{D}} = \min\{|a| \text{ s.t. } a \cdot b^l = c^l,\ l = k+1, \ldots, n\},$$

where $|a|$ is the Euclidean norm of a, and the dot denotes the Euclidean inner product. Indeed, $\|Y\|_{\mathscr{D}^2/\mathscr{D}}$ is the Euclidean distance of the origin from the affine subspace of $\mathbb{R}^{k^2}$ defined by the equations $a \cdot b^l = c^l$ for $l = k+1, \ldots, n$. In order to find an explicit expression for $\|Y\|^2_{\mathscr{D}^2/\mathscr{D}}$ in terms of the b^l, we employ the Lagrange multipliers technique. Then, we look for extremals of

$$L(a, b^{k+1}, \ldots, b^n, \lambda_{k+1}, \ldots, \lambda_n) = |a|^2 - 2\sum_{l=k+1}^{n} \lambda_l (a \cdot b^l - c^l).$$

We obtain the following system

$$\begin{cases} \displaystyle\sum_{l=k+1}^{n} \lambda_l \cdot b^l - a = 0, \\ \displaystyle\sum_{l=k+1}^{n} \lambda_l b^l \cdot b^r = c^r, \qquad r = k+1, \ldots, n. \end{cases} \tag{4.10}$$

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 $\pi : \otimes^2\mathscr{D} \to \mathscr{D}^2/\mathscr{D}$, which has rank $n-k$. Therefore B is symmetric and positive definite, hence invertible. It is now easy to write the solution of system (4.10) by employing the matrix B^{-1}, which has components B^{-1}_{hl}. Indeed a straightforward computation leads to

$$\|c^s Y_s\|^2_{\mathscr{D}^2/\mathscr{D}} = c^h B^{-1}_{hl} c^l \,.$$

By polarization, the inner product on $\mathscr{D}^2/\mathscr{D}$ is defined, in the basis Y_l, by

$$\langle Y_l, Y_h\rangle_{\mathscr{D}^2/\mathscr{D}} = B^{-1}_{lh}.$$

Observe that B^{-1} is the Gram matrix of the vectors $Y_{k+1},\dots,Y_n$ seen as elements of $\mathscr{D}^2/\mathscr{D}$. Then, by the definition of Popp's volume, if $\nu^1,\dots,\nu^k,\mu^{k+1},\dots,\mu^n$ is the dual basis associated with $X_1,\dots,X_k,Y_{k+1},\dots,Y_n$, the following formula holds true

$$\mathcal{P} = \frac{1}{\sqrt{\det B}}\,\nu^1\wedge\dots\wedge\nu^k\wedge\mu^{k+1}\wedge\dots\wedge\mu^n.$$

General case

In the general case, the procedure above can be carried out with no difficulty. Let $X_1,\dots,X_n$ be a local adapted frame for the flag $\mathscr{D}^0\subset\mathscr{D}\subset\mathscr{D}^2\subset\dots\subset\mathscr{D}^m$. As usual $k_i=\dim(\mathscr{D}^i)$. For $j=2,\dots,m$ we define the adapted structure constants $b^l_{i_1\dots i_j}\in C^\infty(M)$ by

$$[X_{i_1},[X_{i_2},\dots,[X_{i_{j-1}},X_{i_j}]]] = \sum_{l=k_{j-1}+1}^{k_j} b^l_{i_1 i_2\dots i_j} X_l \mod \mathscr{D}^{j-1},$$

where $1\leq i_1,\dots,i_j\leq k$. Again, $b^l_{i_1\dots i_j}$ can be seen as the components of a vector $b^l\in\mathbb{R}^{k^j}$.

Recall that for each j we defined the surjective linear map $\pi_j:\otimes^j\mathscr{D}\to\mathscr{D}^j/\mathscr{D}^{j-1}$

$$\pi_j(X_{i_1}\otimes X_{i_2}\otimes\dots\otimes X_{i_j}) = [X_{i_1},[X_{i_2},\dots,[X_{i_{j-1}},X_{i_j}]]] \mod \mathscr{D}^{j-1}.$$

Then, we compute the norm of an element of $\mathscr{D}^j/\mathscr{D}^{j-1}$ 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

$$[B_j]^{hl} = \sum_{i_1,i_2,\dots,i_j=1}^{k} b^h_{i_1 i_2\dots i_j} b^l_{i_1 i_2\dots i_j}.$$

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 $\mathscr{D}^j/\mathscr{D}^{j-1}$ shows that the matrices B_j^{-1} are precisely the Gram matrices of the vectors $X_{k_{j-1}+1},\dots,X_{k_j}\in\mathscr{D}^j/\mathscr{D}^{j-1}$, in other words

$$\langle X_{k_{j-1}+l}, X_{k_{j-1}+h}\rangle_{\mathscr{D}^j/\mathscr{D}^{j-1}} = [B_j^{-1}]_{lh}.$$

Therefore, if $\nu^1,\dots,\nu^n$ is the dual frame associated with $X_1,\dots,X_n$, Popp's volume is

$$\mathcal{P} = \frac{1}{\sqrt{\prod_{j=1}^m \det B_j}}\,\nu^1\wedge\ldots\wedge\nu^n.$$

4.3.2 Examples

In this section we compute Popp's volume for some specific equiregular sub-Riemannian structures. We also discuss, through an example, the non-equiregular case.

Contact manifolds

Contact manifolds are a well-known class of sub-Riemannian structures. We recall the basic definition first, then we compute Popp's volume in terms of a canonical operator associated with a contact structure.

Definition 4.15. Let $\omega \in \Omega^1(M)$ be a one-form on M. Let $\mathscr{D}$ be the $n-1$ dimensional distribution $\mathscr{D} := \ker \omega$. We say that ω is a *contact form* if $d\omega|_{\mathscr{D}}$ is non degenerate. In this case, $\mathscr{D}$ is a called *contact distribution.* A sub-Riemannian structure $(M, \mathscr{D}, \langle\cdot|\cdot\rangle)$, where $\mathscr{D}$ is a contact distribution, is is called *contact sub-Riemannian manifold.*

Notice that the non-degeneracy assumption implies that the dimension of M is odd. Observe that any contact manifold satisfies the bracket-generating condition, is equiregular, has step 2, and its growth vector is $\mathcal{G} = (n-1, n)$.

With any contact form ω we can associate its *Reeb vector field*, which is the unique vector field X_0 satisfying the conditions $\omega(X_0) = 1$ and $d\omega(X_0, \cdot) = 0$. Notice that, given a local orthonormal frame $X_1, \dots, X_k$ for the distribution, then $X_1, \dots, X_k, X_0$ is a local adapted frame, since X_0 is transversal to $\mathscr{D}$.

The contact form ω induces a linear bundle map (i.e. a fibre-wise linear map) $J : \mathscr{D} \to \mathscr{D}$, defined by $\langle JX|Y\rangle = d\omega(X, Y)$, $\forall X, Y \in \mathscr{D}$. Observe that the restriction J_q of J to the fibres of $\mathscr{D}$ is a linear skew-symmetric operator on the inner product space $(\mathscr{D}_q, \langle\cdot|\cdot\rangle_q)$. Hence its Hilbert-Schmidt norm $\|J_q\|$ is well defined by the formula $\|J_q\|^2 = \sum_{i,j=1}^k \langle X_i|JX_j\rangle^2$.

Proposition 4.16. *Let M be a contact sub-Riemannian manifold and $J : \mathscr{D} \to \mathscr{D}$ as above. Let $\nu^1, \dots, \nu^k, \nu^0$ be the dual frame associated with the local adapted frame $X_1, \dots, X_k, X_0$, where X_0 is the Reeb vector field. Then*

$$\mathcal{P} = \frac{1}{\|J_q\|} \nu^1 \wedge \ldots \wedge \nu^k \wedge \nu^0, \tag{4.11}$$

where $\|J_q\|$ is the Hilbert-Schmidt norm of J_q.

Proof. Let $X_1, \dots, X_k, X_0$ be a local adapted frame, where X_0 is the Reeb vector field associated with the contact form. Then, for $1 \leq i, j \leq k$, the structure constants satisfy

$$[X_i, X_j] = \sum_{l=1}^k c_{ij}^l X_l + c_{ij}^0 X_0,$$

$$[X_i, X_0] = \sum_{l=1}^k c_{i0}^l X_l.$$

By Eq. (4.4), $\mathcal{P} = \sqrt{g}\,\nu^1 \wedge \ldots \wedge \nu^k \wedge \nu^0$ where $g = 1/\sum_{i,j=1}^k (c_{ij}^0)^2$. Then the statement follows from the identity

$$\|J\|^2 = \sum_{i,j=1}^k \langle X_i | J X_j \rangle^2 = \sum_{i,j=1}^k d\omega(X_i, X_j)^2 = \sum_{i,j=1}^k \omega([X_i, X_j])^2 = \sum_{i,j=1}^k (c_{ij}^0)^2. \tag{4.12}$$

Observe that, in the last equality of Eq. (4.12), we employed Cartan formula for the differential of a one-form, and the fact that $\omega(X_i) = 0$. □

Eq. (4.11) can be expressed in terms of the eigenvalues of J. See also [7, Remark 30], where the authors exhibit this formula for the case $\mathcal{G} = (4, 5)$.

Remark 4.17. Let $f \in C^\infty(M)$ be a smooth, non-vanishing function. Then ω and $\omega' := f\omega$ define the same contact distribution $\mathscr{D}$. However $d\omega' \neq f d\omega$ and, in general, the associated Reeb vector field is different. On the other hand, as a consequence of the identity $d\omega'|_{\mathscr{D}} = f d\omega|_{\mathscr{D}}$, it follows that $J' = fJ$. Therefore, it is convenient to choose a "normalized" contact form, which is uniquely specified (up to a sign) by the condition $\|J_q\|^2 = 1$, $\forall q \in M$. Then, in terms of the Reeb vector field associated with the normalized contact form, $\mathcal{P} = \nu^1 \wedge \ldots \wedge \nu^k \wedge \nu^0$.

Carnot groups of step 2

A Carnot group $\mathbb{G}$ of step 2 is a left-invariant sub-Riemannian structure on a nilpotent, connected, simply connected Lie group whose Lie algebra $\mathfrak{g}$ admits a stratification $\mathfrak{g} = V_1 \oplus V_2$ with $[V_1, V_1] = V_2$ and $[V_1, V_2] = [V_2, V_2] = \{0\}$. The sub-Riemannian structure is defined by left translation of the subspace V_1, where we choose an orthonormal basis $X_1, \ldots, X_k$. It is possible to choose a basis $Y_{k+1}, \ldots, Y_n$ of V_2 such that

$$[X_i, X_j] = \sum_{h=k+1}^n b_{ij}^h Y_h, \qquad [X_i, Y_h] = [Y_h, Y_l] = 0.$$

Using the standard exponential coordinates (i.e. the identification of the Lie group and its Lie algebra via the exponential map) the explicit expression for the associated left-invariant vector fields in $\mathbb{R}^n = \{(x, y) \,|\, x \in \mathbb{R}^k, y \in \mathbb{R}^{n-k}\}$ is

$$X_i = \partial_{x_i} - \frac{1}{2}\sum_{j,h} b_{ij}^h x_j \partial_{y_h}, \qquad i = 1, \ldots, k,$$
$$Y_h = \partial_{y_h}, \qquad h = k+1, \ldots, n.$$

In [23], the authors employed the skew-symmetric matrices L^h, $k+1 \leq h \leq n$, of components $[L^h]_{ij} = b_{ij}^h$ in order to investigate the nilpotent approximation of a step 2 sub-Riemannian structure. In terms of these matrices,

$$B^{hl} = (L^h, L^l),$$

where $(M, N) := \mathrm{Tr}(M^T N)$ is the Hilbert-Schmidt inner product on $\mathrm{GL}(k, \mathbb{R})$. If the L matrices are orthonormal, Eq. (4.4) gives

$$\mathcal{P} = dx^1 \wedge \ldots \wedge dx^k \wedge dy^{k+1} \wedge \ldots \wedge dy^n.$$

The last formula is (up to a constant factor) the definition of Popp's volume employed in [23, Definition 4] and [32], given in terms of a global adapted frame.

Non-equiregular case

The basic example of a bracket-generating, non-equiregular sub-Riemannian structure is the so-called *Martinet distribution*. This is the distribution on $\mathbb{R}^3$ defined by the kernel of the one-form $\theta := dz - y^2 dx$. A global frame for $\mathscr{D}$, which we declare orthonormal, is

$$X = \partial_x + y^2\partial_z, \qquad Y = \partial_y.$$

Let $Z := \partial_z$. Then $[X, Y] = -2yZ$ and $[Y, [X, Y]] = 2Z$. Observe that X, Y, Z is a global (adapted) frame for TM, therefore Martinet distribution is bracket-generating. However, its growth vector is

$$\mathcal{G}(x, y, z) = \begin{cases} (2,3) & \text{if } y \neq 0, \\ (2,2,3) & \text{if } y = 0, \end{cases}$$

and the distribution is not equiregular on the hyperplane $y = 0$. Nevertheless, if we restrict to the connected components of $\{y \neq 0\}$, we obtain a step 2 equiregular sub-Riemannian manifold. Here, Theorem 4.A gives the following expression:

$$\mathcal{P} = \frac{1}{|y|} dx \wedge dy \wedge dz. \tag{4.13}$$

Eq. (4.13) shows that singularities arise precisely on the hypersurface where the equiregularity hypotesis fails. In [33], the authors investigate the properties of the sub-Laplacian associated with this volume in the Martinet structure. They show that the sub-Laplacian is essentially self-adjoint in each connected component of $\{y \neq 0\}$, hence the hyperplane $\{y = 0\}$ acts as a barrier for the heat propagation.

4.4 Sub-Laplacian

In this section we define the canonical sub-Laplacian associated with a generic volume form and we prove Corollary 4.1, namely an explicit formula for the sub-Laplacian associated with Popp's volume.

On a Riemannian manifold, the Laplace-Beltrami operator is defined as the divergence of the gradient. This definition can be easily generalized to the sub-Riemannian setting.

Definition 4.18. Let $f \in C^\infty(M)$. The *horizontal gradient* of f is the unique horizontal vector field ∇f such that

$$\langle \nabla f | X \rangle = X(f), \qquad \forall X \in \Gamma(\mathscr{D}).$$

It follows from the definition that, in terms of a local frame $X_1, \dots, X_k$ for $\mathscr{D}$

$$\nabla f = \sum_{i=1}^k X_i(f) X_i. \tag{4.14}$$

Definition 4.19. Let $\mu \in \Omega^n(M)$ be a positive volume form, and $X \in \mathrm{Vec}(M)$. The *μ-divergence* of X is the smooth function $\mathrm{div}_\mu X$ defined by

$$\mathcal{L}_X \mu = \mathrm{div}_\mu X \mu.$$

where $\mathcal{L}_X$ is the Lie derivative in the direction X.

Notice that the definition of divergence does not depend on the orientation of M, namely the sign of μ. The divergence measures the rate at which the volume of a region changes under the integral flow of a field. Indeed, for any compact $\Omega \subset M$ and t sufficiently small, let $e^{tX} : \Omega \to M$ be the flow of $X \in \mathrm{Vec}(M)$, then

$$\left.\frac{d}{dt}\right|_{t=0} \int_{e^{tX}(\Omega)} \mu = -\int_\Omega \mathrm{div}_\mu X \mu.$$

The next proposition is sometimes employed as an alternative definition of divergence. Let $C_0^\infty(M)$ be the space of smooth functions with compact support.

Proposition 4.20. *For any $f \in C_0^\infty(M)$ and $X \in \mathrm{Vec}(M)$*

$$\int_M f \mathrm{div}_\mu X \mu = -\int_M X(f) \mu.$$

Proof. The proof is an easy consequence of the definition of μ-divergence. □

The next lemma gives the relation between divergences associated with different volumes.

Lemma 4.21. *Let $\mu, \mu' \in \Omega^n(M)$ be volume forms. Let $f \in C^\infty(M)$, $f \neq 0$ such that $\mu' = f\mu$. Then, for any $X \in \mathrm{Vec}(M)$*

$$\mathrm{div}_{\mu'} X = \mathrm{div}_\mu X + X(\log f).$$

Proof. It follows from the Leibniz rule $\mathcal{L}_X(f\mu) = (Xf)\mu + f\mathcal{L}_X\mu = (X(\log f) + \mathrm{div}_\mu X) f\mu$. □

When no confusion may arise, we write "div", without any reference to the volume form μ. In the following, we fix the reference volume to be Popp's one. Lemma 4.21 can be used to generalize the results to the case of a generic μ-divergence.

With a divergence and a gradient at our disposal, we are ready to define the sub-Laplacian associated with the volume form μ.

Definition 4.22. Let $\mu \in \Omega^n(M)$, $f \in C^\infty(M)$. The *sub-Laplacian* associated with μ is the second order differential operator

$$\Delta f := \mathrm{div}\,(\nabla f)\,,$$

This definition reduces to the Laplace-Beltrami operator when μ is the Riemannian volume. As a consequence of Eq. (4.14) and the Leibniz rule for the divergence $\mathrm{div}(fX) = Xf + f\,\mathrm{div}(X)$, we can find the expression of the sub-Laplacian in terms of any local frame $X_1, \dots, X_k$:

$$\mathrm{div}\,(\nabla f) = \sum_{i=1}^k \mathrm{div}\,(X_i(f) X_i) = \sum_{i=1}^k X_i(X_i(f)) + \mathrm{div}(X_i) X_i(f)\,.$$

Then

$$\Delta = \sum_{i=1}^{k} X_i^2 + \mathrm{div}(X_i)X_i. \tag{4.15}$$

Remark 4.23. Observe that the second order term of Δ, namely the "sum of squares" in Eq. (4.15), does not depend on the choice of the volume. Indeed, only the first order terms depend on it through the divergence operator, which changes according to Lemma 4.25 upon a change of volume.

Remark 4.24. If we apply Proposition 4.20 to the horizontal gradient ∇g, we obtain

$$\int_M f\Delta g\mu = -\int_M \langle \nabla f | \nabla g \rangle \mu, \qquad \forall f, g \in C_0^\infty(M).$$

Then Δ is symmetric and negative on $C_0^\infty(M)$. It can be proved that it is also essentially self-adjoint (see [66]).

Now we prove a useful formula for the divergence associated with Popp's volume. Analogous formulae for μ-divergences are easily obtained by an application of Lemma 4.21.

Lemma 4.25. *Let $X_1, \dots, X_n$ be a local adapted frame. Let* div *be the divergence associated with Popp's volume. Then, for $i = 1, \dots, n$*

$$\mathrm{div} X_i = -\left(\frac{1}{2} \sum_{j=1}^{m} \mathrm{Tr}(B_j^{-1} X_i(B_j)) + \sum_{l=1}^{n} c_{il}^l \right) X_i. \tag{4.16}$$

Proof. Let $\nu \in \Omega^1(M)$, and $X, Y \in \mathrm{Vec}(M)$. The Lie derivative obeys Leibniz rule:

$$\mathcal{L}_X (\nu(Y)) = (\mathcal{L}_X \nu)(Y) + \nu(\mathcal{L}_X Y).$$

Then, if $\nu^1, \dots, \nu^n$ is the dual frame associated with $X_1, \dots, X_n$

$$\mathcal{L}_{X_i} \nu^j = -\sum_{l=1}^{n} c_{il}^j \nu^l, \tag{4.17}$$

which is the "dual formulation" of Eq. (4.1). By Theorem 4.A, Popp's volume is

$$\mathcal{P} = \frac{1}{\sqrt{\prod_j \det B_j}} \nu^1 \wedge \dots \wedge \nu^n.$$

Then, for $i = 1, \dots, n$,

$$\mathcal{L}_{X_i} \mathcal{P} = \sqrt{\prod_j \det B_j}\, X_i \left(\frac{1}{\sqrt{\prod_j \det B_j}} \right) \mathcal{P} + \\ + \frac{1}{\sqrt{\prod_j \det B_j}} \left(\mathcal{L}_{X_i} \nu^1 \wedge \dots \wedge \nu^n + \dots + \nu^1 \wedge \dots \wedge \mathcal{L}_{X_i} \nu^n \right). \tag{4.18}$$

Eq. (4.16) now follows from the definition of divergence, Eq. (4.17) and Eq. (4.18). □

Finally, Corollary 4.1 is a straightforward consequence of Lemma 4.25 and Eq. (4.15). □

4.5 Volume preserving transformations

This section is devoted to the proof of Propositions 4.B and 4.C.

4.5.1 Proof of Proposition 4.B

Let $\phi \in \mathrm{Iso}(M)$ be a (local) isometry, and $1 \leq i \leq m$. The differential ϕ_* induces a linear map

$$\widetilde{\phi}_* : \otimes^i \mathscr{D}_q \to \otimes^i \mathscr{D}_{\phi(q)}.$$

Moreover ϕ_* preserves the flag $\mathscr{D} \subset \ldots \subset \mathscr{D}^m$. Therefore, it induces a linear map

$$\widehat{\phi}_* : \mathscr{D}^i_q / \mathscr{D}^{i-1}_q \to \mathscr{D}^i_{\phi(q)} / \mathscr{D}^{i-1}_{\phi(q)}.$$

The key to the proof of Proposition 4.B is the following lemma.

Lemma 4.26. $\widetilde{\phi}_*$ *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 $\mathscr{D}^i/\mathscr{D}^{i-1}$ is induced by the surjective maps $\pi_i : \otimes^i \mathscr{D} \to \mathscr{D}^i/\mathscr{D}^{i-1}$ defined by Eq. (4.7). Namely, let $Y \in \mathscr{D}^i_q/\mathscr{D}^{i-1}_q$. Then

$$\|Y\|_{\mathscr{D}^i_q/\mathscr{D}^{i-1}_q} = \min\{\|Z\|_{\otimes \mathscr{D}_q} \text{ s.t. } \pi_i(Z) = Y\}.$$

As a consequence of the properties of the Lie brackets, $\pi_i \circ \widetilde{\phi}_* = \widehat{\phi}_* \circ \pi_i$. Therefore

$$\|Y\|_{\mathscr{D}^i_q/\mathscr{D}^{i-1}_q} = \min\{\|\widetilde{\phi}_* Z\|_{\otimes \mathscr{D}_{\phi(q)}} \text{ s.t. } \pi_i(\widetilde{\phi}_* Z) = \widehat{\phi}_* Y\} = \|\widehat{\phi}_* Y\|_{\mathscr{D}^i_{\phi(q)}/\mathscr{D}^{i-1}_{\phi(q)}}.$$

By polarization, $\widehat{\phi}_*$ is an isometry. □

Since $\mathrm{gr}_q(\mathscr{D}) = \oplus_{i=1}^m \mathscr{D}^i_q/\mathscr{D}^{i-1}_q$ is an orthogonal direct sum, $\widehat{\phi}_* : \mathrm{gr}_q(\mathscr{D}) \to \mathrm{gr}_{\phi(q)}(\mathscr{D})$ is also an isometry of inner product spaces.

Finally, Popp's volume is the canonical volume of $\mathrm{gr}_q(\mathscr{D})$ 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. □

4.5.2 Proof of Proposition 4.C

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 $\phi \in \mathrm{Iso}(M)$

$$f\mu = \mathcal{P} = \phi^*\mathcal{P} = (f \circ \phi)\, \phi^*\mu = (f \circ \phi)\, \mu,$$

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. □

Appendix A

Asymptotics of Jacobi curves

Proposition (Special case of Theorem 1.122)**.** *Let* $\Lambda(\cdot)$ *a Jacobi curve of rank* 1*, with vanishing* $R(t)$*. The matrix* S*, in terms of the canonical frame, is*

$$S_{ij}(t) = \frac{(-1)^{i+j-1}}{(i-1)!(j-1)!}\frac{t^{i+j-1}}{(i+j-1)} = \widehat{S}_{ij}t^{i+j-1}, \qquad i,j = 1,\dots,n.$$

Its inverse is

$$S^{-1}(t)_{ij} = \frac{-1}{i+j-1}\binom{n+i-1}{i-1}\binom{n+j-1}{j-1}\frac{(n!)^2}{(n-i)!(n-j)!} = \frac{\widehat{S}^{-1}_{ij}}{t^{i+j-1}}, \qquad i,j = 1,\dots,n.$$

Proof. From Eqs. (1.74) and (1.75), we obtain

$$\begin{aligned} S_{ij}(t) &= \sum_{k=1}^{n} A^{-1}_{ik}B_{kj} = \sum_{k=1}^{i}\frac{(-1)^{i-k}t^{i-k}}{(i-k)!}\frac{(-1)^{j}t^{k+j-1}}{(k+j-1)!} = (-1)^{j}t^{i+j-1}\sum_{k=1}^{i}\frac{(-1)^{i-k}}{(k+j-1)!(i-k)!} = \\ &= (-1)^{j}t^{i+j-1}\sum_{\ell=0}^{i-1}\frac{(-1)^{\ell}}{(i+j-1-\ell)!\ell!} = \frac{(-1)^{j}t^{i+j-1}}{(i+j-1)!}\sum_{\ell=0}^{i-1}\binom{i+j-1}{\ell}(-1)^{\ell} = \\ &= \frac{(-1)^{i+j-1}t^{i+j-1}}{(i+j-1)!}\binom{i+j-2}{j-1} = \frac{(-1)^{i+j-1}}{(i-1)!(j-1)!}\frac{t^{i+j-1}}{(i+j-1)}. \end{aligned}$$

By Cramer's rule, the inverse of $S(t)$ is

$$S^{-1}_{ij}(t) = \frac{(-1)^{i+j}\det\left[\frac{(-1)^{\ell+k-1}}{(\ell-1)!(k-1)!}\frac{t^{\ell+k-1}}{(\ell+k-1)}\right]_{\substack{\ell\neq j\\ k\neq i}}}{\det\left[\frac{(-1)^{\ell+k-1}}{(\ell-1)!(k-1)!}\frac{t^{\ell+k-1}}{(\ell+k-1)}\right]} = \frac{-(i-1)!(j-1)!\det\left[\frac{1}{\ell+k-1}\right]_{\substack{\ell\neq j\\ k\neq i}}}{t^{i+j-1}\det\left[\frac{1}{\ell+k-1}\right]}, \tag{A.1}$$

Now we compute the ratio of determinants in the last factor of Eq. (A.1). Consider a generic matrix of the form $H_{\ell k} = \frac{1}{x_\ell + x_k}$, for $\ell, k = 1,\dots,n$. For fixed $i,j \in \{1,\dots,n\}$, we can express the determinant of H in terms of the the i,j-th minor, by rows and columns operations as follows. First, subtract the i-th column from each other column. We obtain a new matrix, H', whose i-th column is the same of H, while, for $k \neq i$

$$H'_{\ell k} = \frac{1}{x_\ell + y_k} - \frac{1}{x_\ell + y_i} = \frac{y_i - y_k}{(x_\ell + y_i)(x_\ell + y_k)}, \qquad \ell, k = 1, \dots, n.$$

Indeed $\det H' = \det H$. Then, we collect the factor $\frac{1}{x_\ell + y_i}$ from each row, and the factor $(y_i - y_k)$ from each column but the i-th. We obtain

$$\det \left[\frac{1}{x_\ell + x_k} \right] = \prod_{\ell=1}^n \frac{1}{x_\ell + y_i} \prod_{\substack{k=1 \\ k \neq i}}^n (y_i - y_k) \det \begin{bmatrix} \frac{1}{x_1+y_1} & \frac{1}{x_1+y_2} & \dots & 1 & \dots & \frac{1}{x_1+y_n} \\ \frac{1}{x_2+y_1} & \frac{1}{x_2+y_2} & \dots & 1 & \dots & \frac{1}{x_2+y_n} \\ \vdots & \vdots & & \vdots & & \vdots \\ \frac{1}{x_n+y_1} & \frac{1}{x_n+y_2} & \dots & 1 & \dots & \frac{1}{x_n+y_n} \end{bmatrix},$$

where the entries of the i-th column are equal to 1. Now, subtract the j-th row from each other row, but the j-th itself. Collect again the common factors. We obtain

$$\det \left[\frac{1}{x_\ell + x_k} \right] = (-1)^{i+j} \prod_{\ell=1}^n \frac{1}{x_\ell + y_i} \prod_{\substack{k=1 \\ k \neq i}}^n (y_i - y_k) \prod_{\substack{k=1 \\ k \neq i}}^n \frac{1}{x_j + y_k} \prod_{\substack{\ell=1 \\ \ell \neq j}}^n (x_j - x_\ell) \det \left[\frac{1}{x_\ell + x_k} \right]_{\substack{\ell \neq j \\ k \neq i}}. \tag{A.2}$$

Now we apply the result of Eq. (A.2) to our case, i.e. $x_\ell = y_\ell = \ell - \frac{1}{2}$. Therefore we obtain

$$\begin{aligned} \frac{\det \left[\frac{1}{\ell + k - 1} \right]_{\substack{\ell \neq j \\ k \neq i}}}{\det \left[\frac{1}{\ell + k - 1} \right]} &= (-1)^{i+j} \prod_{\ell=1}^n (\ell + i - 1) \prod_{\substack{k=1 \\ k \neq i}}^n \frac{1}{i - k} \prod_{\substack{k=1 \\ k \neq i}}^n (j + k - 1) \prod_{\substack{\ell=1 \\ \ell \neq j}}^n \frac{1}{j - \ell} = \\ &= \frac{1}{i + j - 1} \frac{(n!)^2}{(i-1)!(j-1)!} \binom{i+n-1}{i-1} \binom{j+n-1}{j-1}. \end{aligned} \tag{A.3}$$

Eq. (A.1) and Eq. (A.3), together, give the desired formula. □

Appendix B

Curvature coefficients

Lemma. *Let*

$$\Omega(n,m) = \frac{nm}{(n+1)(m+1)} \sum_{j=1}^{n} \sum_{i=1}^{m} (-1)^{i+j} \binom{n+i-1}{i-1} \binom{n+1}{i+1} \binom{m+j-1}{j-1} \binom{m+1}{j+1} \frac{i+j+2}{i+j+1}.$$

Then

$$\Omega(n,m) = \begin{cases} 0 & |n-m| \geq 2, \\ \frac{1}{4(n+m)} & |n-m| = 1, \\ \frac{n}{4n^2-1} & n = m. \end{cases}$$

Proof. It is clear that $\Omega(n,m) = \Omega(m,n)$, then we can assume without loss of generality that $n \leq m$. The case $m = n = 1$ can be easily proved by a direct computation. Then, we also assume $m \geq 2$. Let us write $\Omega(n,m)$ in a more compact form. In order to do that, let $M(n,m)$ be the $n \times m$ matrix of components

$$M(n,m)_{ij} := (-1)^{i+j} \frac{i+j+2}{i+j+1}, \qquad i = 1, \dots, n, \quad j = 1, \dots, m.$$

and let $v(m)$ be the m-dimensional column vector of components

$$v(m)_j = \frac{m}{m+1} \binom{m+1}{j+1} \binom{m+j-1}{j-1}, \qquad j = 1, \dots, m.$$

Then

$$\Omega(n,m) = v(n)^* M(n,m) v(m).$$

Consider first the $i-th$ component of the n-dimensional vector $w(n,m) := M(n,m)v(m)$, namely

$$w(n,m)_i = \sum_{j=1}^{m} (-1)^{i+j} \frac{i+j+2}{i+j+1} \frac{m}{m+1} \binom{m+1}{j+1} \binom{m+j-1}{j-1} = \frac{(-1)^i}{(m-1)!} \sum_{j=0}^{m} (-1)^j \binom{m}{j} Q_i(j),$$

where, for each $i = 1, \dots, n$, $Q_i(j)$ is a rational function (in the variable j) defined by

$$Q_i(j) = \frac{(m+j-1)!}{(j-1)!(j+1)} \frac{i+j+2}{i+j+1} = j(j+2)(j+3)\dots(j+m-1)\frac{i+j+2}{i+j+1}.$$

Notice that the factor $(j+1)$ does not appear (remember also that $m \geq 2$). The idea is to exploit the following beautiful identity.

Lemma B.1. *Let $m \geq 2$. Let $P(x)$ be any polynomial of degree smaller than m, then*

$$\sum_{j=0}^{m} (-1)^j \binom{m}{j} P(j) = 0.$$

Proof. It is sufficient to prove the statement for $P(x) = x^i$, with $0 \leq i < m$, since any polynomial of degree smaller than m is a linear combination of such monomials. By Newton's binomial formula, we have

$$(x-1)^m = (-1)^m \sum_{j=0}^{m} (-1)^j \binom{m}{j} x^j.$$

The result easily follows observing that any derivative of order strictly smaller than m, evaluated at $x = 1$ vanishes. □

We will see that, for many values of i, the denominator of $Q_i(j)$ factors the numerator, and then $Q_i(j)$ is actually a polynomial of degree $m-1$ in the variable j. Then we apply Lemma B.1 to show that $w(n,m)_i \neq 0$ only if $i = m-1, m$. In particular, since $w(n,m)$ is a n-dimensional vector, if $n \leq m-2$ then $w(n,m) = 0$ and $\Omega(n,m)$ vanishes too. Then we will explicitly compute the coefficient for $n = m-1$ and $n = m$.

Observe that, for each $i = 1, \dots, n$, the numerator of $Q_i(j)$ is a polynomial of degree m in the variable j. Therefore there exists a polynomial $P_i(j)$ (of degree strictly smaller than m) and a number R_i such that

$$Q_i(j) = P_i(j) + \frac{R_i}{i+j+1}.$$

It is easy to compute the remainder. Observe that

$$R_i = -(i+j+1)P_i(j) + Q_i(j)(i+j+1).$$

Then, evaluating at $j = -i-1$, we obtain

$$R_i = \begin{cases} 0 & i = 1, 2, \dots, m-2, \\ (-1)^{m-1}\dfrac{m!}{m-1} & i = m-1, \\ (-1)^{m-1}\dfrac{(m+1)!}{m} & i = m. \end{cases} \tag{B.1}$$

By Lemma B.1 we have

$$w(n,m)_i = \frac{(-1)^i}{(m-1)!} \sum_{j=0}^{m} (-1)^j \binom{m}{j} \frac{R_i}{i+j+1},$$

which, by Eq. (B.1), is indeed zero if $i = 1, 2, \ldots, m-2$. Then, since $\Omega(n,m) = v(n)^* w(n,m)$, we obtain after some straightforward computations the following formula:

$$\Omega(n,m) = \begin{cases} 0 & m-n > 2, \\ \binom{2m-3}{m-2} \sum_{j=0}^{m} \binom{m}{j} \frac{(-1)^j}{j+m} & n = m-1, \\ \binom{2m-2}{m-2} (m+1) \sum_{j=0}^{m} \binom{m}{j} \frac{(-1)^j}{j+m} - \binom{2m-1}{m-1} m \sum_{j=0}^{m} \binom{m}{j} \frac{(-1)^j}{j+m+1} & n = m. \end{cases} \tag{B.2}$$

In order to obtain the result, it only remains to compute the sums appearing in Eq. (B.2). Indeed these are of the form

$$S_k := \sum_{j=0}^{m} (-1)^j \binom{m}{j} \frac{1}{j+k},$$

where k is a positive integer. We have the following, remarkable identity.

$$\sum_{j=0}^{m} (-1)^j \binom{m}{j} \frac{1}{j+k} = \frac{m!(k-1)!}{(m+k)!}. \tag{B.3}$$

By plugging Eq. (B.3) in Eq. (B.2) we obtain the result. Then we only need to prove Eq. (B.3). Indeed, for k a positive integer, let us define the following function

$$f_k(x) := \sum_{j=0}^{m} (-1)^j \binom{m}{j} \frac{(-x)^{j+k}}{j+k}.$$

Indeed $S_k = f_k(-1)$. Let us compute the derivative of f.

$$f'(x) = -\sum_{j=0}^{m} (-1)^j \binom{m}{j} (-x)^{j+k-1} = (-1)^k x^{k-1} (1+x)^m.$$

where we used Newton's binomial formula. Then

$$S_k = f(-1) = (-1)^k \int_0^{-1} x^{k-1} (1+x)^m.$$

By integrating by parts $k-1$ times, we obtain the result

$$S_k = f(-1) = \frac{m!(k-1)!}{(m+k)!}.$$

□

Appendix C

Smoothness of the value function

The goal of this section is to prove Theorem 1.19 on the smoothness of the value function. All the relevant definitions can be found in Section 1.2. As a first step, we generalize the classical definition of conjugate points to our setting.

Definition C.1. Let $\gamma : [0,T] \to M$ be a strictly normal trajectory, such that $x_0 = \gamma(0)$ and $\gamma(t) = \mathcal{E}_{x_0}(t, \lambda_0)$. We say that $\gamma(t)$ is *conjugate with* x_0 *along* γ if λ_0 is a critical point for $\mathcal{E}_{x_0,t}$.

Observe that the relation "being conjugate with" is not reflexive in general. Indeed, even if $\gamma(t)$ is conjugate with x_0, there might not even exist an admissible curve starting from $\gamma(t)$ and ending at x_0.

We stress that, if γ is also abnormal, any $\gamma(t)$ is a critical value of the sub-Riemannian exponential map. Indeed, this is a consequence of the inclusion

$$\operatorname{Im} D_{\lambda_0}\mathcal{E}_{x_0,t} \subset \operatorname{Im} D_u E_{x_0,t} \subsetneq T_{x_0}M$$

for abnormal trajectories; being strongly normal is a necessary condition for the absence of critical values along a normal trajectory. Actually, a converse of this statement is true.

Proposition C.2. *Let $\gamma : [0,T] \to M$ be a strongly normal trajectory. Then, there exists an $\varepsilon > 0$ such that $\gamma(t)$ is not conjugate with $\gamma(0)$ along γ for all $t \in (0,\varepsilon)$.*

The proof of Proposition C.2 in the sub-Riemannian setting can be found in [6] and can be adapted to a general affine optimal control system. See also [15] for a more general approach.

We are now ready to prove Theorem 1.19 about smoothness of the value function which, for the reader's convenience, we restate here. Recall that $M' \subset M$ is the relatively compact subset chosen for the definition of the value function.

Theorem. *Let $\gamma : [0,T] \to M'$ be a strongly normal trajectory. Then there exists an $\varepsilon > 0$ and an open neighbourhood $U \subset (0,\varepsilon) \times M' \times M'$ such that:*

(i) $(t, \gamma(0), \gamma(t)) \in U$ *for all* $t \in (0,\varepsilon)$,

(ii) For any $(t,x,y) \in U$ there exists a unique (normal) minimizer of the cost functional J_t, among all the admissible curves that connect x with y in time t, contained in M',

(iii) The value function $(t, x, y) \mapsto S_t(x, y)$ is smooth on U.

Proof. We first prove the theorem in the case $M' = M$ compact. We need the following sufficient condition for optimality of normal trajectory. Let $a \in C^\infty(M)$. The graph of its differential is a smooth submanifold $\mathcal{L}_0 := \{d_x a | \, x \in M\} \subset T^*M$, $\dim \mathcal{L}_0 = \dim M$. Translations of $\mathcal{L}_0$ by the flow of the Hamiltonian field $\mathcal{L}_\tau = e^{\tau \vec{H}}(\mathcal{L}_0)$ are also smooth submanifolds of the same dimension.

Lemma C.3 (see [15, Theorem 17.1])**.** *Assume that the restriction $\pi : \mathcal{L}_\tau \to M$ is a diffeomorphism for any $\tau \in [0, \varepsilon]$. Then, for any $\lambda_0 \in \mathcal{L}_0$, the normal trajectory*

$$\gamma(\tau) = \pi \circ e^{\tau \vec{H}}(\lambda_0), \qquad \tau \in [0, \varepsilon],$$

is a strict minimum of the cost functional J_ε among all admissible trajectories connecting $\gamma(0)$ with $\gamma(\varepsilon)$ in time ε.

Lemma C.3 is a sufficient condition for the optimality of a single normal trajectory. By building a suitable family of smooth functions $a \in C^\infty(M)$, one can prove that, for any sufficiently small compact set $K \subset T^*M$, we can find a $\varepsilon = \varepsilon(K) > 0$ sufficiently small such that, for any $\lambda_0 \in K$, and for any $t \leq \varepsilon$, the normal trajectory

$$\gamma(\tau) = \pi \circ e^{\tau \vec{H}}(\lambda_0), \qquad \tau \in [0, t], \qquad t \leq \varepsilon$$

is a strict minimum of the cost functional J_t among all admissible curves connecting $\gamma(0)$ with $\gamma(t)$ in time t.

We sketch the explicit construction of such a family. Let $K \subset T^*M$ sufficiently small such that it is contained in a trivial neighbourhood $\mathbb{R}^n \times U \subset T^*M$. Let (p, x) be coordinates on K induced by a choice of coordinates x on $O \subset M$. Then, consider the function $a : K \times O \to \mathbb{R}$, defined in coordinates by $a(p_0, x_0; y) = p_0^* y$. Extend such a function to $a : K \times M \to \mathbb{R}$. For any $\lambda_0 \in K$, denote by $a^{(\lambda_0)} = a(\lambda_0; \cdot) \in C^\infty(M)$. Indeed, for $x_0 = \pi(\lambda_0)$, we have $\lambda_0 = d_{x_0} a^{(\lambda_0)}$. In other words we can recover any initial covector in K by taking the differential at x_0 of an appropriate element of the family. Therefore, let $\mathcal{L}_0^{(\lambda_0)} := \{d_x a^{(\lambda_0)} | \, x \in M\}$, and $\mathcal{L}_\tau^{(\lambda_0)} := e^{\tau \vec{H}}(\mathcal{L}_0^{(\lambda_0)})$. M is compact, then there exists $\varepsilon(K) = \sup\{\tau \geq 0 | \, \pi : \mathcal{L}_s^{(\lambda_0)} \to M$ is a diffeomorphism for all $s \in [0, \tau]$, $\lambda_0 \in K\} > 0$.

Let us go back to the proof. Set $x_0 = \gamma(0)$, and let $\gamma(t) = \mathcal{E}_{x_0}(t, \lambda_0)$. By Proposition C.2, we can assume that $\gamma(t)$ is not conjugate with $\gamma(0)$ along γ for all $t \in (0, \varepsilon)$. In particular, $D_{\lambda_0} \mathcal{E}_{x_0, t}$ has maximal rank for all $t \in (0, \varepsilon)$. Without loss of generality, assume that $\vec{H}$ is complete. Then, consider the map $\phi : \mathbb{R}^+ \times T^*M \to \mathbb{R}^+ \times M \times M$, defined by

$$\phi(t, \lambda) = (t, \pi(\lambda), \mathcal{E}_{\pi(\lambda)}(t, \lambda)).$$

The differential of ϕ, computed at (t, λ_0), is

$$D_{(t,\lambda_0)} \phi = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \mathbb{I} & 0 \\ * & * & D_{\lambda_0} \mathcal{E}_{x_0, t} \end{pmatrix}, \qquad \forall t \in (0, \varepsilon),$$

which has maximal rank. Therefore, by the inverse function theorem, for each $t \in (0, \varepsilon)$, there exist an interval I_t and open sets W_t, U_t, V_t such that

$$t \in I_t \subset (0, \varepsilon), \qquad \lambda_0 \in W_t \subset T^*M, \qquad \gamma(0) \in U_t \subset M, \qquad \gamma(t) \in V_t \subset M,$$

and such that the restriction

$$\phi : I_t \times W_t \to I_t \times U_t \times V_t$$

is a smooth diffeomorphism. In particular, for any $(\tau, x, y) \in I_t \times U_t \times V_t$ there exists an unique initial covector $\lambda_0(\tau, x, y) := \phi^{-1}(\tau, x, y)$ such that the corresponding normal trajectory starts from x and arrives at y in time τ, i.e. $\mathcal{E}_x(\tau, \lambda_0(\tau, x, y)) = y$. Moreover, we can choose $W_t \subset K$. Then such a normal trajectory is also a strict minimizer of J_τ among all the admissible curves connecting x with y in time τ. In particular, it is unique.

As a consequence of the smoothness of the local inverse, the value function $(t, x, y) \mapsto S_t(x, y)$ is smooth on each open set $I_t \times U_t \times V_t$. Indeed, for any $(\tau, x, y) \in I_t \times U_t \times V_t$, $S_t(x, y)$ is equal to the cost J_τ of the unique (normal) minimizer connecting x with y in time τ, namely

$$S_\tau(x, y) = \int_0^\tau L(\mathcal{E}_{x_0}(s, \lambda_0(\tau, x, y)), \bar{u}(e^{s\vec{H}}(\lambda_0(\tau, x, y))))ds, \qquad (\tau, x, y) \in I_t \times U_t \times V_t,$$

where $\bar{u} : T^*M \to \mathbb{R}^k$ is the smooth map which recovers the control associated with the lift on T^*M of the trajectory (see Theorem 1.17). Therefore the value function is smooth on $I_t \times U_t \times V_t$, as a composition of smooth functions. We conclude the proof by defining the open set

$$U := \bigcup_{t \in (0, \varepsilon)} I_t \times U_t \times V_t \subset (0, \varepsilon) \times M \times M,$$

which is indeed open and contains $(t, \gamma(0), \gamma(t))$ for all $t \in (0, \varepsilon)$.

In the general case the proof follows the same lines, although the optimality of small segments of geodesics is only among all the trajectories not leaving M'. If we choose a different relatively compact $M'' \subset M$, we find a common ε such that the restriction to the interval $[0, \varepsilon]$ of all the normal geodesics with initial covector in K is a strict minimum of the cost function among all the admissible trajectories not leaving $M'' \cup M'$. Therefore, the value functions associated with the two different choices of the relatively compact subset agree on the intersection of the associated domains U. $\square$

Appendix D

Convergence of the approximating systems

The goal of this section is the proof of Proposition 1.66. Actually, we discuss a more general statement for the associated Hamiltonian system. All the relevant definitions can be found in Section 1.5.1.

Let $\lambda = (p, x) \in T^*\mathbb{R}^n = \mathbb{R}^{2n}$ any initial datum. Let ϕ^ε and $\widehat{\phi}$, respectively, the Hamiltonian flow of the ε-approximated system and of the nilpotent system, respectively. A priori, these local flows are defined in a neighbourhood of the initial condition and for small time which, in general, depend on ε. Notice that, by abuse of notation $\phi^0 = \widehat{\phi}$.

Lemma. *For $\varepsilon \geq 0$ sufficiently small, there exist common neighbourhood $I_0 \subset \mathbb{R}$ of 0 and $O_{\lambda_0} \subset \mathbb{R}^{2n}$ of λ_0, such that $\phi^\varepsilon : I_0 \times O_{\lambda_0} \to \mathbb{R}^{2n}$ is well defined. Moreover, $\phi^\varepsilon \to \widehat{\phi}$ in the C^∞ topology of uniform convergence of all derivatives on $I_0 \times O_{\lambda_0}$.*

Proof. Indeed, for any $\varepsilon \geq 0$, the Hamiltonian flow ϕ^ε is associated with the Cauchy problem

$$\dot{\lambda}(t) = H^\varepsilon(\lambda(t)), \qquad \lambda(0) = \lambda_0.$$

Moreover, ϕ^ε is well defined and smooth in a neighbourhood $I_0^\varepsilon \times O_{\lambda_0}^\varepsilon \subset \mathbb{R} \times \mathbb{R}^{2n}$ (that depends on ε). To find a common domain of definition, consider the associated Cauchy problem in $\mathbb{R}^{2n+1}$.

$$\begin{pmatrix} \dot{\lambda}(t) \\ \dot{\varepsilon}(t) \end{pmatrix} = \begin{pmatrix} H(\varepsilon(t), \lambda(t)) \\ 0 \end{pmatrix}, \qquad \begin{pmatrix} \lambda(0) \\ \varepsilon(0) \end{pmatrix} = \begin{pmatrix} \lambda_0 \\ \varepsilon_0 \end{pmatrix}, \tag{D.1}$$

where $H(\varepsilon, \lambda) := H^\varepsilon(\lambda)$ is smooth in both variables by construction. We denote by $\Phi(t; \lambda_0, \varepsilon_0)$ the flow associated with the Cauchy problem (D.1). By classical ODE theory, there exists a neighbourhood $I_0 \subset \mathbb{R}$ of 0 and $U_{\lambda_0,\varepsilon_0} \subset \mathbb{R}^{2n+1}$ of $(\lambda_0, \varepsilon_0)$ such that $\Phi : I_0 \times U_{\lambda_0,\varepsilon_0} \to \mathbb{R}^{2n+1}$ is well defined and smooth. Indeed $\Phi(t; \lambda_0, \varepsilon) = \phi^\varepsilon(t; \lambda_0)$ and $\Phi(t; \lambda_0, 0) = \widehat{\phi}(t; \lambda_0)$. Then, we can find an open neighbourhood $O_{\lambda_0} \subset \mathbb{R}^{2n}$ of λ_0 such that $O_{\lambda_0} \times [0, \delta] \subset U_{\lambda_0,0}$. Thus, the sought common domain of definition for all the ϕ^ε, with $0 \leq \varepsilon \leq \delta$, is $I_0 \times O_{\lambda_0}$.

Finally, Φ is smooth on $I_0 \times U_{\lambda_0,0}$. Then ϕ^ε (and all its derivatives) converge to $\widehat{\phi}$ (and all the corresponding derivatives) on $I_0 \times O_{\lambda_0}$. Up to restricting the domain of definition of Φ, we can always assume I_0 and O_λ to be compact, hence the convergence is also uniform. □

Without loss of generality, by homogeneity, we can always reduce to $I_0 = [0, T]$. Now Proposition 1.66 easily follows, since the exponential map is the projection of the Hamiltonian flow, restricted to the fiber $T_0^*\mathbb{R}^n$.

Appendix E

Distance expansion in the Heisenberg group

For the reader's convenience, we briefly recall the statement of Proposition 1.94. We refer to Section 1.5.7 for all the relevant definitions.

Proposition. *The function $C(t,s)$ is C^1 in a neighbourhood of the origin, but not C^2. In particular, the function $\partial_{ss}C(t,0)$ is not continuous at the origin. However, the singularity at $t=0$ is removable, and the following expansion holds, for $t>0$*

$$\frac{\partial^2 C}{\partial s^2}(t,0) = 1 + 3\sin^2(\phi_2-\phi_1) + \frac{1}{2}[2h_{z,2}\sin(\phi_2-\phi_1) - h_{z,1}\sin(2\phi_2-2\phi_1)]t - \\ - \frac{2}{15}h_{z,1}^2\sin^2(\phi_2-\phi_1)t^2 + O(t^3).$$

If the geodesic γ_2 is chosen to be a straight line (i.e. $h_{z,2}=0$), then

$$\frac{\partial^2 C}{\partial s^2}(t,0) = 1 + 3\sin^2(\phi_2-\phi_1) - \frac{h_{z,1}}{2}\sin(2\phi_2-2\phi_1)t - \frac{2}{15}h_{z,1}^2\sin^2(\phi_2-\phi_1)t^2 + O(t^3). \quad \text{(E.1)}$$

where $\lambda_j = (ie^{i\phi_j}, h_{z,j}) = (-\sin\phi_j, \cos\phi_j, h_{z,h}) \in T_0^*M$ is the initial covector of the geodesic γ_j.

Proof. The proof is essentially a brute force computation. In the following, we show the relevant calculation to obtain the zeroth order term in Eq. (E.1), which is sufficient to prove the non-continuity of the function $t \mapsto \partial_{ss}C(t,0)$ at $t=0$. Indeed, since $C(0,s)=s^2/2$, we obtain $\partial_{ss}C(0,0)=1$, while from Eq. (E.1), $\lim_{t\to 0^+}\partial_{ss}C(0,s) = 1+3\sin^2(\phi_2-\phi_1)$.

For $i=1,2$, let $\gamma_i(\tau) = (w_i(\tau), z_i(\tau))$. Then

$$w_i(\tau) = \frac{e^{i\phi_i}}{a_i}\left(e^{ia_i\tau}-1\right) = ie^{i\phi_i}\tau - \frac{1}{2}a_i e^{i\phi_i}\tau^2 + O(\tau^3),$$
$$z_i(\tau) = \frac{a_i\tau - \sin(a_i\tau)}{2a_i^2} = O(\tau^3).$$

For $(t,s) \neq (0,0)$, dropping the subscripts from $R_{t,s}$ and $\xi_{t,s}$, we have

$$\partial_{tt}C(t,s) = \frac{1}{2}\partial_{tt}R^2\frac{\theta^2(\xi)}{\sin^2\theta(\xi)} + 4\partial_t R^2\theta(\xi)\partial_t\xi + 2R^2\dot{\theta}(\xi)(\partial_t\xi)^2 + 2R^2\theta(\xi)\partial_{tt}\xi =$$
$$= A_1(t,s) + A_2(t,s) + A_3(t,s) + A_4(t,s), \quad \text{(E.2)}$$

where A_i are the four addends of the upper line of Eq. (E.2). In order to compute Eq. (E.2), we employ the following calculations

$$\begin{aligned}
R_{t,s}^2 &= |w_2(s) - w_1(t)|^2, \\
\partial_t R_{t,s}^2 &= \dot{w}_1(t)[\overline{w}_1(t) - \overline{w}_2(s)] + [w_1(t) - w_2(s)]\dot{\overline{w}}_1(t), \\
\partial_{tt} R_{t,s}^2 &= \ddot{w}_1(t)[\overline{w}_1(t) - \overline{w}_2(s)] + 2|\dot{w}_1(t)|^2 + \ddot{\overline{w}}_1(t)[w_1(t) - w_2(s)], \\
Z_{t,s} &= -z_1(t) + z_2(s) + \frac{1}{2}\Im(w_1(t)\overline{w}_2(s)), \\
\partial_t Z_{t,s} &= -\dot{z}_1(t) + \frac{1}{2}\Im(\dot{w}_1(t)\overline{w}_2(s)), \\
\partial_{tt} Z_{t,s} &= -\ddot{z}_1(t) + \frac{1}{2}\Im(\ddot{w}_1(t)\overline{w}_2(s)), \\
\xi_{t,s} &= Z_{t,s}/R_{t,s}^2, \\
\partial_t\xi_{t,s} &= \frac{\partial_t Z}{R^2} - \frac{Z}{R^4}\partial_t R^2, \\
\partial_{tt}\xi_{t,s} &= \frac{\partial_{tt} Z}{R^2} - 2\frac{\partial_t Z}{R^4}\partial_t R^2 - \frac{Z}{R^4}\partial_{tt}R^2 + 4\frac{Z}{R^6}(\partial_t R^2)^2,
\end{aligned}$$

where $\Im$ is the imaginary part, the overline is the complex conjugate, and the dot is the derivative w.r.t. the argument. Moreover, the Taylor series for θ is

$$\theta(x) = 6x + O(x^3).$$

By computing everything at $t = 0$, and then taking the limit $s \to 0$, we obtain

$$\begin{aligned}
\lim_{s\to 0} A_1(0,s) &= 1, \\
\lim_{s\to 0} A_2(0,s) &= 0, \\
\lim_{s\to 0} A_3(0,s) &= 3\sin^2(\phi_1 - \phi_2), \\
\lim_{s\to 0} A_4(0,s) &= 0,
\end{aligned}$$

therefore $\lim_{s\to 0}\partial_{tt}C(0,s) = 1 + 3\sin^2(\phi_1 - \phi_2)$, which is the zeroth order term of Eq. (E.1). The term arising from the addend $A_3(0,s)$ is responsible for the discontinuity of $\partial_{tt}C(0,s)$ at $s = 0$. The remaining terms can be obtained by taking expansions up to the fourth order of R^2, Z, θ, and replacing them in Eq. (E.2). □

Appendix F

Comparison theorems for the matrix Riccati equation

The general, non-autonomous, symmetric matrix Riccati equation can be written as follows:

$$\dot{X} = \mathrm{R}(X;t) := M(t)_{11} + XM(t)_{12} + M(t)_{12}^{*}X + XM(t)_{22}X = \begin{pmatrix} \mathbb{I} & X \end{pmatrix} M(t) \begin{pmatrix} \mathbb{I} \\ X \end{pmatrix},$$

where $M(t)$ is a smooth family of $2n \times 2n$ symmetric matrices. We always assume a symmetric initial datum, then the solution must be symmetric as well on the maximal interval of definition. All the comparison results are based upon the following theorems.

Theorem F.1 (Riccati comparison theorem 1)**.** *Let $M_1(t)$, $M_2(t)$ be two smooth families of $2n \times 2n$ symmetric matrices. Let $X_i(t)$ be smooth solution of the Riccati equation*

$$\dot{X}_i = \mathrm{R}_i(X_i;t), \qquad i = 1,2,$$

on a common interval $I \subseteq \mathbb{R}$. Let $t_0 \in I$ and (i) *$M_1(t) \geq M_2(t)$ for all $t \in I$,* (ii) *$X_1(t_0) \geq X_2(t_0)$. Then for any $t \in [t_0, +\infty) \cap I$, we have $X_1(t) \geq X_2(t)$.*

Proof. The proof is a simplified version of [1, Thm. 4.1.4]. Let $U := X_1 - X_2$. Notice that U is symmetric on the interval I where both solutions are defined. A computation shows that

$$\dot{U} = \theta(t)U + U\theta(t)^{*} + \begin{pmatrix} \mathbb{I} & X_1 \end{pmatrix} (M_1 - M_2) \begin{pmatrix} \mathbb{I} \\ X_1 \end{pmatrix},$$

where

$$\theta(t) = M_2(t)_{12}^{*} + \frac{1}{2}X_1(t)M_2(t)_{22} + \frac{1}{2}X_2(t)M_2(t)_{22}.$$

Taking in account that $M_1(t) - M_2(t) \geq 0$, the matrix U satisfies

$$\dot{U} \geq \theta(t)U + U\theta(t)^{*}.$$

Indeed $U(t_0) \geq 0$. Then, the statement follows from the next lemma. □

Lemma F.2. *(see [1, Thm. 4.1.2]) Let U be a symmetric solution of the Lyapunov differential inequality*

$$\dot{U} \geq \theta(t)U + U\theta(t)^*, \qquad t \in I \subseteq \mathbb{R},$$

where $\theta(t)$ is smooth. Then $U(t_0) \geq 0$ implies $U(t) \geq 0$ for all $t \in I \cap [t_0, +\infty)$.

Proof. Let $t, \tau \in I$. Let $\phi(t,\tau) : \mathbb{R}^n \to \mathbb{R}^n$ be the non-autonomous flow associated with $-\theta^*(t)$, namely

$$\partial_t \phi(t,\tau) = -\theta^*(t)\phi(t,\tau), \qquad \phi(\tau,\tau) = \mathbb{I}.$$

Notice that $\partial_t \phi^*(t,\tau) = -\phi^*(t,\tau)\theta(t)$. Then, let $P(t,\tau) := \phi^*(t,\tau)U(t)\phi(t,\tau)$. Since $U(t_0) \geq 0$, and $\phi(t,\tau)$ is not degenerate, we have

$$P(t_0,\tau) \geq 0 \qquad \tau \in I.$$

Moreover

$$\partial_t P(t,\tau) = \phi^*(t,\tau)[-\theta(t)U(t) + \dot{U}(t) - U(t)\theta(t)^*]\phi(t,\tau) \geq 0,$$

by hypotesis. This implies that

$$P(t,\tau) \geq P(t_0,\tau) \geq 0 \qquad \tau \in I,\, t \in I \cap [t_0, +\infty).$$

Then, setting $\tau = t$, we obtain

$$U(t) = P(t,t) \geq 0, \qquad t \in I \cap [t_0, +\infty).$$

and the statement follows. □

The assumptions of Theorem F.1 involve comparison on coefficients of Riccati equations and on initial data. It can be generalised also for limit initial data as follows.

Theorem F.3 (Riccati comparison theorem 2)**.** *Let $M_1(t)$, $M_2(t)$ be two smooth families of $2n \times 2n$ symmetric matrices. Let $X_i(t)$ be smooth solutions of the Riccati equation*

$$\dot{X}_i = \mathrm{R}_i(X_i; t), \qquad i = 1, 2,$$

on a common interval $I \subseteq \mathbb{R}$. Let $t_0 \in \bar{I}$. Assume that (i) *$M_1(t) \geq M_2(t)$ for all $t \in \bar{I}$,* (ii) *$X_i(t) > 0$ for $t > t_0$ sufficiently small,* (iii) *there exist $Y_i(t_0) := \lim_{t \to t_0+} X_i^{-1}(t)$ and* (iv) *$Y_1(t_0) \leq Y_2(t_0)$. Then, for any $t \in (t_0, +\infty) \cap I$, we have $X_1(t) \geq X_2(t)$.*

Proof. Let $Y_i(t) := X_i(t)^{-1}$, defined on some interval $(t_0, \varepsilon) \subseteq I$. They satisfy

$$\dot{Y}_i = \begin{pmatrix} \mathbb{I} & Y_i \end{pmatrix} N_i(t) \begin{pmatrix} \mathbb{I} \\ Y_i \end{pmatrix}, \qquad N_i(t) := -\begin{pmatrix} 0 & \mathbb{I} \\ \mathbb{I} & 0 \end{pmatrix} M_i(t) \begin{pmatrix} 0 & \mathbb{I} \\ \mathbb{I} & 0 \end{pmatrix}, \qquad i = 1, 2.$$

Indeed $Y_i(t)$ can be prolonged on $[t_0, \varepsilon]$ for ε sufficiently small by (iii). Moreover $N_2(t) \geq N_1(t)$ by (i) and $Y_2(t_0) \geq Y_1(t_0)$ by (iv). The point t_0 belongs to the interval of definition of Y_i then, by Theorem F.1, $Y_2(\varepsilon) \geq Y_1(\varepsilon)$. By (ii), this implies that $X_1(\varepsilon) \geq X_1(\varepsilon)$. Then we can apply again Theorem F.1 to X_1 and X_2, with $t_0 = \varepsilon$, and we obtain that $X_1(t) \geq X_2(t)$ for all $t \in [\varepsilon, +\infty) \cap I$. Since ε can be chosen arbitrarily close to t_0, we obtain the statement. □

Appendix G

Well posedness of limit Cauchy problem

The following lemma justifies the savage use of the Cauchy problem with limit initial condition. Let A, B, be $n \times n$ and $n \times k$ matrices, respectively, satisfying the controllability condition

$$\operatorname{span}\{B, AB, \ldots, A^m B\} = \mathbb{R}^n,$$

for some $m \geq 0$. Thus, since the column space of B is equal to the column space of BB^*, we have that, if we put $\Gamma_1 := A^*$ and $\Gamma_2 := BB^* \geq 0$,

$$\operatorname{span}\{\Gamma_2, \Gamma_2\Gamma_1, \ldots, \Gamma_2\Gamma_1^m\} = \mathbb{R}^n.$$

This condition is indeed satisfied for the matrices Γ_1, Γ_2 introduced in Sec. 2.3.

Lemma G.1. *For any smooth $R(t)$, the Cauchy problem with limit initial condition*

$$\dot{V} = -\Gamma_1 V - V\Gamma_1^* - R(t) - V\Gamma_2 V, \qquad \lim_{t\to 0^+} V^{-1} = 0,$$

is well posed, in the sense that there exists a solution of the Riccati equation, invertible for small $t > 0$ such that $\lim_{t\to 0^+} V^{-1} = 0$. The solution is unique on some maximal interval of definition $I \subseteq (0, +\infty)$. In addition, $V(t) > 0$ for small $t > 0$.

Proof. We first prove uniqueness. If two solutions V_1, V_2 exist, their inverses W_1 and W_2 (defined for $t > 0$ sufficiently small) can be extended to smooth matrices on $[0, \varepsilon)$, by setting $W_1(0) = W_2(0) = 0$. Moreover, they both satisfy the following Cauchy problem:

$$\dot{W} = \Gamma_1^* W + W\Gamma_1 + \Gamma_2 + WR(t)W, \qquad W(0) = 0.$$

By uniqueness of the standard Cauchy problem, $W_1(\varepsilon) = W_2(\varepsilon)$. Therefore also $V_1^{-1}(\varepsilon) = V_2^{-1}(\varepsilon)$, and uniqueness follows. The choice of $\varepsilon > 0$ for setting the Cauchy datum is irrelevant, since different choices bring to the same solution. Finally, any solution can be extended uniquely to a maximal solution, defined on some interval $I \subseteq (0, +\infty)$.

Now, we prove the existence. Consider the Cauchy problem

$$\dot{W} = \Gamma_1^* W + W\Gamma_1 + \Gamma_2 + WR(t)W, \qquad W(0) = 0.$$

Its solution is well defined for $t \in [0, \varepsilon)$. We will soon prove that, for $t \in (0, \varepsilon)$, such a solution is positive. Thus $V(t) := W(t)^{-1}$, defined for $t \in (0, \varepsilon)$, is a solution of the original Cauchy problem with limit initial datum, by construction.

We are left to prove that, for $t > 0$ small enough, $W(t) > 0$. Since $R(t)$ is smooth, for $t \in [0, \varepsilon)$ we can find k such that $R(t) \geq k\mathbb{I}$. By comparison Theorem 2.A, we have that our solution is bounded below by the solution with $R(t) = k\mathbb{I}$. We write $W(t) \geq W_k(t) \geq 0$ for $t \in [0, \varepsilon)$ (the last inequality follows again from Theorem 2.A, by considering the trivial solution of the Cauchy problem obtained by setting $\Gamma_2 = 0$).

Assume that, for some small $t > 0$ and $x \neq 0$, we have $W(t)x = 0$. This imples $W_k(t)x = 0$. Being a solution of a Riccati equation with constant coefficients, $W_k(t)$ is monotone non-decreasing (indeed $\dot{W}(0) = \Gamma_2 \geq 0$, and the same holds true for $t \in [0, \varepsilon)$ by Lemma 2.34). Therefore $W_k(t)x = 0$ identically. Therefore all the derivatives, computed at $t = 0$, vanish identically. This imples, after careful examination of the higher derivatives, that

$$\Gamma_2 x = \Gamma_2\Gamma_1 x = \ldots \Gamma_2\Gamma_1^m x = \ldots = 0,$$

that leads to $x = 0$. This contradicts the assumption, hence $W(t) > 0$ for t sufficiently small. □

Appendix H

A generalised Cauchy-Schwarz inequality

Lemma. *Let* $\{X_a\}_{a=1}^r$*,* $\{Y_a\}_{a=1}^r$ *be sets of* $\ell \times \ell$ *matrices. Then*

$$\left(\sum_{a=1}^r X_a^* Y_a\right)\left(\sum_{b=1}^r X_b^* Y_b\right)^* \leq \left\|\sum_{a=1}^r Y_a^* Y_a\right\| \sum_{b=1}^r X_b^* X_b.$$

Proof. Let $v \in \mathbb{R}^\ell$. Then

$$v^* \left(\sum_{a=1}^r X_a^* Y_a\right)\left(\sum_{b=1}^r X_b^* Y_b\right)^* v = \left\|\sum_{a=1}^r Y_a^* X_a v\right\|^2.$$

Notice the change in position of the transpose. Now, let $u \in \mathbb{R}^\ell$, such that $\|u\| = 1$, and

$$u^* \left(\sum_{a=1}^r Y_a^* X_a v\right) = \left\|\sum_{a=1}^r Y_a^* X_a v\right\|.$$

Then, by the Cauchy-Schwarz inequality, we obtain

$$\left\|\sum_{a=1}^r Y_a^* X_a v\right\|^2 = \left|\sum_{a=1}^r (Y_a u)^* (X_a v)\right|^2 \leq \sum_{a=1}^r \|Y_a u\|^2 \sum_{a=1}^r \|X_a v\|^2. \tag{H.1}$$

Now, observe that

$$\sum_{a=1}^r \|Y_a u\|^2 = u^* \sum_{a=1}^r Y_a^* Y_a u \leq \left\|\sum_{a=1}^r Y_a^* Y_a\right\|. \tag{H.2}$$

Then, plugging Eq. (H.2) in Eq. (H.1), we obtain

$$\left\|\sum_{a=1}^r Y_a^* X_a v\right\|^2 \leq \left\|\sum_{a=1}^r Y_a^* Y_a\right\| \sum_{b=1}^r \|X_b v\|^2,$$

which implies the statement. □

Appendix I

The canonical frame in the Riemannian setting

In order to prove Proposition 2.17 and Lemma 2.18 we define a local frame on T^*M, associated with the choice of a local frame $X_1, \dots, X_n$ on M. For $i = 1, \dots, n$ let $h_i : T^*M \to \mathbb{R}$ be the linear-on-fibres function defined by $h_i(\lambda) := \langle \lambda, X_i \rangle$. The action of derivations on T^*M is completely determined by the action on affine functions, namely functions $a \in C^\infty(T^*M)$ such that $a(\lambda) = \langle \lambda, Y \rangle + \pi^* g$ for some $Y \in \mathrm{Vec}(M)$, $g \in C^\infty(M)$. Then, we define the *coordinate lift of a field* $X \in \mathrm{Vec}(M)$ as the field $\widetilde{X} \in \mathrm{Vec}(T^*M)$ such that $\widetilde{X}(h_i) = 0$ for $i = 1, \dots, n$ and $\widetilde{X}(\pi^* g) = X(g)$. This, together with Leibniz rule, characterize the action of $\widetilde{X}$ on affine functions, and then completely define $\widetilde{X}$. Indeed, by definition, $\pi_* \widetilde{X} = X$. On the other hand, we define the (vertical) fields ∂_{h_i} such that $\partial_{h_i}(\pi^* g) = 0$, and $\partial_{h_i}(h_j) = \delta_{ij}$. It is easy to check that $\{\partial_{h_i}, \widetilde{X}_i\}_{i=0}^n$ is a local frame on T^*M. We call such a frame the *coordinate lifted frame*.

Proof of Proposition 2.17 and Lemma 2.18. Point (i) is trivial and follows from the definition of the coordinate lifted frame ∂_{h_i}. In order to prove point (ii), we compute explicitly

$$\vec{H} = \sum_{i=1}^n \left(h_i \widetilde{X}_i + \sum_{j,k=1}^n h_i c_{ij}^k h_k \partial_{h_j} \right) = \sum_{i=1}^n \left(h_i \widetilde{X}_i + \sum_{j,k=1}^n h_i \Gamma_{ij}^k h_k \partial_{h_j} \right),$$

where we used the identities $c_{ij}^k = \Gamma_{ij}^k - \Gamma_{ji}^k$ and $\Gamma_{ij}^k = -\Gamma_{ik}^j$. Then, a direct computation gives

$$F_i = -[\vec{H}, \partial_{h_i}] = \widetilde{X}_i + \sum_{j,k=1}^n h_k \left(\Gamma_{ij}^k + \Gamma_{kj}^i \right) \partial_{h_j} = \widetilde{X}_i + \sum_{j,k=1}^n h_k \Gamma_{ij}^k \partial_{h_j}, \tag{I.1}$$

where we used the fact that, for a parallely transported frame, $\sum_{k=1}^n h_k \Gamma_{kj}^i = 0$ and we suppressed the explicit evaluation at $\lambda(t)$. Now we are ready to prove point (ii). Indeed $\sigma_{\lambda(t)}(\partial_{h_i}, \partial_{h_j}) = 0$, since $\mathcal{V}_\lambda$ is Lagrangian for all λ. Then

$$\sigma_{\lambda(t)}(\partial_{h_i}, -[\vec{H}, \partial_{h_j}]) = -\langle X_i | \pi_* [\vec{H}, \partial_{h_j}] \rangle = \delta_{ij},$$

where we used that $\pi_*[H, \partial_{h_j}] = -X_j$, and that for any vertical vector $\xi \in \mathcal{V}_\lambda$ and $\eta \in T_\lambda(T^*M)$, $\sigma(\xi, \eta) = \langle \xi | \pi_* \eta \rangle$, where we identified ξ with an element of $T_{\pi(\lambda)}M$ through the scalar product. Finally, by using the r.h.s. of Eq. (I.1), we obtain

$$\sigma(F_i, F_j) = \sum_{k=1}^n \left(\Gamma_{ij}^k h_k - \Gamma_{ji}^k h_k - h_k c_{ij}^k \right) = \sum_{k=1}^n \langle h_k X_k | \nabla_{X_i} X_j - \nabla_{X_j} X_i - [X_i, X_j] \rangle = 0,$$

where we suppressed the explicit dependence on t and the last equality is implied by the vanishing of the torsion of Levi-Civita connection. For what concerns point (iii), the first structural equation is the definition of F_i. By taking the derivative of F_i, we obtain

$$\dot{F}_i = [\vec{H}, F_i] = \sum_{\ell,k,j=1}^n h_\ell h_k \langle \nabla_{X_i} \nabla_{X_\ell} X_k - \nabla_{X_\ell} \nabla_{X_i} X_k - \nabla_{[X_i, X_\ell]} X_k | X_j \rangle E_j.$$

In particular, this implies Lemma 2.18, since

$$R_{ij}(t) = \sum_{\ell,j=1}^n h_\ell h_k \langle \nabla_{X_i} \nabla_{X_\ell} X_k - \nabla_{X_\ell} \nabla_{X_i} X_k - \nabla_{[X_i, X_\ell]} X_k | X_j \rangle = \langle R^\nabla(X_i, \dot{\gamma}) \dot{\gamma}, X_j \rangle,$$

by definition of Riemann tensor, and the fact that $\dot{\gamma}(t) = \sum_{i=1}^n h_i(\lambda(t)) X_i|_{\gamma(t)}$. Finally, let $\widetilde{E}_i, \widetilde{F}_j$ be any smooth moving frame along $\lambda(t)$ satisfying (i)-(iii). We can write, in full generality

$$\widetilde{E}_i = \sum_{j=1}^n A_{ij}(t) E_j + B_{ij}(t) F_j, \qquad \widetilde{F}_i = \sum_{j=1}^n C_{ij}(t) E_j + D_{ij}(t) F_j,$$

for some smooth families of $n \times n$ matrices $A(t), B(t), C(t), D(t)$, where the frame is understood to be evaluated at $\lambda(t)$. By imposing conditions (i)-(iii), we obtain that the latter are actually constant, orthogonal matrices, and $B = C = 0$, thus proving the uniqueness property. $\square$

Acknowledgements

First and foremost, I owe my gratitude to my advisor Prof. A. Agrachev, for his support and guidance during my PhD studies. Not only he has been an invaluable teacher, but also a model for how I believe a true Mathematician should be.

A special thanks goes to my friend and invaluable collaborator D. Barilari. It has been really fun doing research with you, and I really hope that our collaboration may continue in the future.

I wish to thanks SISSA for the stimulating environment and the resources that puts at our disposal. Hardly I could have imagined a better place for my doctoral studies. I also thank the Institut Henri Poincaré for its hospitality in Paris.

I am indebted to Prof. Paul Lee for his kind invitation to the Chinese University of Hong Kong, where part of the research contained in this thesis has been carried out.

A special mention goes also to my office mate and collaborator P. Silveira, for accepting - and sharing, when needed - my unconventional work ethic, and to Antonio Lerario, for our newborn but intense collaboration.

Finally, a special thank to my wife, Francisca and to all the people who shared with me these years in Trieste, in particular to the best "colleagues" I could have asked for: Alessandro, Fabio, Mattia, Riccardo and Roby.

Bibliography

[1] H. Abou-Kandil. *Matrix Riccati Equations: In Control and Systems Theory.* Systems & Control: Foundations & Applications. Springer Verlag NY, 2003.

[2] A. Agrachev. Exponential mappings for contact sub-Riemannian structures. *J. Dynam. Control Systems*, 2(3):321–358, 1996.

[3] A. Agrachev. Geometry of optimal control problems and Hamiltonian systems. In *Nonlinear and optimal control theory*, volume 1932 of *Lecture Notes in Math.*, pages 1–59. Springer, Berlin, 2008.

[4] A. Agrachev. Any sub-Riemannian metric has points of smoothness. *Dokl. Akad. Nauk*, 424(3):295–298, 2009.

[5] A. Agrachev and D. Barilari. Sub-Riemannian structures on 3D Lie groups. *J. Dyn. and Contr. Syst.*, 18(1):21–44, 2012.

[6] A. Agrachev, D. Barilari, and U. Boscain. Introduction to Riemannian and sub-Riemannian geometry (Lecture Notes), http://people.sissa.it/agrachev/agrachev_files/notes.html. 2012.

[7] 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.

[8] A. Agrachev, D. Barilari, P. W. Y. Lee, and L. Rizzi. Curvature for contact sub-Riemannian manifolds. *In preparation*, 2014.

[9] A. Agrachev, D. Barilari, and L. Rizzi. The curvature: a variational approach. *arXiv:1306.5318v3*, 2013.

[10] 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.

[11] A. Agrachev and R. Gamkrelidze. Symplectic methods for optimization and control. In *Geometry of feedback and optimal control*, volume 207 of *Monogr. Textbooks Pure Appl. Math.*, pages 19–77. Dekker, New York, 1998.

[12] A. Agrachev, R. V. Gamkrelidze, and A. V. Sarychev. Local invariants of smooth control systems. *Acta Appl. Math.*, 14(3):191–237, 1989.

[13] A. Agrachev and J.-P. Gauthier. On the subanalyticity of Carnot-Caratheodory distances. *Ann. Inst. H. Poincaré Anal. Non Linéaire*, 18(3):359–382, 2001.

[14] A. Agrachev, L. Rizzi, and P. Silveira. On conjugate times of LQ optimal control problems. *ArXiv e-prints*, Nov. 2013.

[15] 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.

[16] A. Agrachev and I. Zelenko. Geometry of Jacobi curves. I. *J. Dynam. Control Systems*, 8(1):93–140, 2002.

[17] A. A. Agrachev and R. V. Gamkrelidze. Feedback-invariant optimal control theory and differential geometry. I. Regular extremals. *J. Dynam. Control Systems*, 3(3):343–389, 1997.

[18] A. A. Agrachev and P. Lee. Generalized Ricci Curvature Bounds for Three Dimensional Contact Subriemannian manifolds. *ArXiv preprint, arXiv:0903.2550 (to appear on Math. Ann)*, Mar. 2009.

[19] L. Ambrosio, N. Gigli, and G. Savaré. Metric measure spaces with Riemannian Ricci curvature bounded from below. *ArXiv preprint, arXiv:1109.0222*, Sept. 2011.

[20] V. I. Arnol′d. *Mathematical methods of classical mechanics*, volume 60 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1997. Translated from the 1974 Russian original by K. Vogtmann and A. Weinstein, Corrected reprint of the second (1989) edition.

[21] D. Barilari. Trace heat kernel asymptotics in 3d contact sub-Riemannian geometry. *To appear on Journal of Mathematical Sciences*, 2011.

[22] D. Barilari, U. Boscain, G. Charlot, and R. W. Neel. On the heat diffusion for generic Riemannian and sub-Riemannian structures. *ArXiv e-prints*, Oct. 2013.

[23] 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.

[24] D. Barilari, U. Boscain, and R. W. Neel. Small-time heat kernel asymptotics at the sub-Riemannian cut locus. *J. Differential Geom.*, 92(3):373–416, 2012.

[25] D. Barilari and A. Lerario. Geometry of Maslov cycles. In *Geometric Control Theory and sub-Riemannian Geometry*, INdAM Series 5. Springer, 2013. arXiv preprint arXiv:1301.0268.

[26] D. Barilari and L. Rizzi. A formula for Popp's volume in sub-Riemannian geometry. *Analysis and Geometry in Metric Spaces*, 2013.

[27] F. Baudoin, M. Bonnefont, and N. Garofalo. A sub-riemannian curvature-dimension inequality, volume doubling property and the poincaré inequality. *Mathematische Annalen*, pages 1–28, 2013.

[28] F. Baudoin, M. Bonnefont, N. Garofalo, and I. H. Munive. Volume and distance comparison theorems for sub-Riemannian manifolds. *ArXiv e-prints*, Nov. 2012.

[29] F. Baudoin and N. Garofalo. Curvature-dimension inequalities and Ricci lower bounds for sub-Riemannian manifolds with transverse symmetries. *arXiv:1101.3590*, Jan. 2011.

[30] F. Baudoin and J. Wang. Curvature-dimension inequalities and subelliptic heat kernel gradient bounds on contact manifolds. *To appear on Potential Analysis*, 2013.

[31] A. Bellaïche. The tangent space in sub-Riemannian geometry. In *Sub-Riemannian geometry*, volume 144 of *Progr. Math.*, pages 1–78. Birkhäuser, Basel, 1996.

[32] 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.

[33] U. Boscain and C. Laurent. The Laplace-Beltrami operator in almost-Riemannian geometry. *arXiv:1105.4687v1 [math.SP], Preprint*, 2011.

[34] U. Boscain and F. Rossi. Invariant Carnot-Caratheodory metrics on S^3, SO(3), SL(2), and lens spaces. *SIAM J. Control Optim.*, 47(4):1851–1878, 2008.

[35] R. W. Brockett. Control theory and singular Riemannian geometry. In *New directions in applied mathematics (Cleveland, Ohio, 1980)*, pages 11–27. Springer, New York, 1982.

[36] R. L. Bryant, S. S. Chern, R. B. Gardner, H. L. Goldschmidt, and P. A. Griffiths. *Exterior differential systems*, volume 18 of *Mathematical Sciences Research Institute Publications*. Springer-Verlag, New York, 1991.

[37] J.-M. Coron. *Control and nonlinearity*, volume 136 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2007.

[38] L. E. Faĭbusovich. Existence and uniqueness of extremal solutions of the Riccati equation and symplectic geometry. *Funktsional. Anal. i Prilozhen.*, 19(1):85–86, 1985.

[39] T. Gallouët. *Transport optimal : régularité et applications*. These, Ecole normale supérieure de lyon - ENS LYON, Dec. 2012.

[40] V. Gershkovich and A. Vershik. Nonholonomic manifolds and nilpotent analysis. *J. Geom. Phys.*, 5(3):407–452, 1988.

[41] R. Ghezzi and F. Jean. A new class of $(\mathcal{H}^k, 1)$-rectifiable subsets of metric spaces. *ArXiv preprint, arXiv:1109.3181 [math.MG]*, Sept. 2011.

[42] A. B. Givental′. The maximum number of singular points on a projective hypersurface. *Funktsional. Anal. i Prilozhen.*, 17(3):73–74, 1983.

[43] M. Gromov. Carnot-Carathéodory spaces seen from within. In *Sub-Riemannian geometry*, volume 144 of *Progr. Math.*, pages 79–323. Birkhäuser, Basel, 1996.

[44] F. Jean. *Control of Nonholonomic Systems: from Sub-Riemannian Geometry to Motion Planning.* to appear.

[45] N. Juillet. Geometric inequalities and generalized Ricci bounds in the Heisenberg group. *Int. Math. Res. Not. IMRN*, (13):2347–2373, 2009.

[46] V. Jurdjevic. *Geometric control theory*, volume 52 of *Cambridge Studies in Advanced Mathematics.* Cambridge University Press, Cambridge, 1997.

[47] T. Kato. *Perturbation theory for linear operators.* Classics in Mathematics. Springer-Verlag, Berlin, 1995. Reprint of the 1980 edition.

[48] P. W. Y. Lee, C. Li, and I. Zelenko. Measure contraction properties of contact sub-Riemannian manifolds with symmetry. *ArXiv e-prints*, Apr. 2013.

[49] C. Li and I. Zelenko. Jacobi equations and comparison theorems for corank 1 sub-Riemannian structures with symmetries. *J. Geom. Phys.*, 61(4):781–807, 2011.

[50] G. Loeper. On the regularity of solutions of optimal transportation problems. *Acta Math.*, 202(2):241–283, 2009.

[51] J. Lott and C. Villani. Ricci curvature for metric-measure spaces via optimal transport. *Ann. of Math. (2)*, 169(3):903–991, 2009.

[52] K. R. Meyer and G. R. Hall. *Introduction to Hamiltonian dynamical systems and the N-body problem*, volume 90 of *Applied Mathematical Sciences.* Springer-Verlag, New York, 1992.

[53] J. W. Milnor. *Topology from the differentiable viewpoint.* Princeton Landmarks in Mathematics. Princeton University Press, Princeton, NJ, 1997. Based on notes by David W. Weaver, Revised reprint of the 1965 original.

[54] J. Mitchell. On Carnot-Carathéodory metrics. *J. Differential Geom.*, 21(1):35–45, 1985.

[55] I. Moiseev and Y. L. Sachkov. Maxwell strata in sub-Riemannian problem on the group of motions of a plane. *ESAIM Control Optim. Calc. Var.*, 16:380–399, 2010.

[56] R. Montgomery. *A tour of subriemannian geometries, their geodesics and applications*, volume 91 of *Mathematical Surveys and Monographs.* American Mathematical Society, Providence, RI, 2002.

[57] S. B. Myers. Riemannian manifolds with positive mean curvature. *Duke Math. J.*, 8:401–404, 1941.

[58] S.-i. Ohta. On the curvature and heat flow on hamiltonian systems. *arXiv:1308.5746*, 2013.

[59] S.-i. Ohta. On the curvature and heat flow on Hamiltonian systems. *ArXiv e-prints*, Aug. 2013.

[60] S.-i. Ohta and K.-T. Sturm. Bochner–Weitzenböck formula and Li–Yau estimates on Finsler manifolds. *Adv. Math.*, 252:429–448, 2014.

[61] L. S. Pontryagin, V. G. Boltyanskiĭ, R. V. Gamkrelidze, and E. F. Mishchenko. *Selected works. Vol. 4*. Classics of Soviet Mathematics. Gordon & Breach Science Publishers, New York, 1986. The mathematical theory of optimal processes, Edited and with a preface by R. V. Gamkrelidze, Translated from the Russian by K. N. Trirogoff, Translation edited by L. W. Neustadt, With a preface by L. W. Neustadt and K. N. Trirogoff, Reprint of the 1962 English translation.

[62] B. Riemann. Über die hypothesen, welche der geometrie zu grunde liegen. Göttengen, 1854.

[63] L. Rifford. *Sub-Riemannian Geometry and Optimal Transport.* Lecture Notes, 146 pp., 2012.

[64] L. Rifford. Ricci curvatures in Carnot groups. *Math. Control Relat. Fields*, 3(4):467–487, 2013.

[65] Y. L. Sachkov. Conjugate and cut time in the sub-Riemannian problem on the group of motions of a plane. *ESAIM Control Optim. Calc. Var.*, 16:1018–1039, 2010.

[66] R. S. Strichartz. Sub-Riemannian geometry. *J. Differential Geom.*, 24(2):221–263, 1986.

[67] K.-T. Sturm. On the geometry of metric measure spaces. I. *Acta Math.*, 196(1):65–131, 2006.

[68] K.-T. Sturm. On the geometry of metric measure spaces. II. *Acta Math.*, 196(1):133–177, 2006.

[69] C. Villani. *Optimal transport*, volume 338 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 2009.

[70] B. Y. Wu and Y. L. Xin. Comparison theorems in Finsler geometry and their applications. *Math. Ann.*, 337(1):177–196, 2007.

[71] I. Zelenko and C. Li. Differential geometry of curves in Lagrange Grassmannians with given Young diagram. *Differential Geom. Appl.*, 27(6):723–742, 2009.