Introduction to Riemannian and Sub-Riemannian geometry

From Hamiltonian viewpoint

ANDREI AGRACHEV
DAVIDE BARILARI
UGO BOSCAIN

This version: June 12, 2016
Preprint SISSA 09/2012/M
# Contents

## Introduction

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

   1.1 Geodesics and optimality ........................................... 17
     1.1.1 Existence and minimizing properties of geodesics .............. 17
     1.1.2 Absolutely continuous curves .................................. 21
   1.2 Parallel transport .................................................. 23
   1.3 Gauss-Bonnet Theorems .............................................. 26
     1.3.1 Gauss-Bonnet theorem: local version .......................... 26
     1.3.2 Gauss-Bonnet theorem: global version ......................... 30
     1.3.3 Consequences of the Gauss-Bonnet Theorems .................. 33
     1.3.4 The Gauss map ................................................ 34
   1.4 Surfaces in $\mathbb{R}^3$ with the Minkowski inner product ....... 37
   1.5 Model spaces of constant curvature .................................. 40
     1.5.1 Zero curvature: the Euclidean plane .......................... 40
     1.5.2 Positive curvature: spheres .................................. 40
     1.5.3 Negative curvature: the hyperbolic plane .................... 42

2 **Vector fields and vector bundles** ........................................... 45

   2.1 Differential equations on smooth manifolds .......................... 45
     2.1.1 Tangent vectors and vector fields ........................... 45
     2.1.2 Flow of a vector field ....................................... 47
     2.1.3 Vector fields as operators on functions ...................... 47
     2.1.4 Nonautonomous vector fields .................................. 48
   2.2 Differential of a smooth map ........................................ 50
   2.3 Lie brackets .......................................................... 51
   2.4 Cotangent space ..................................................... 53
   2.5 Vector bundles ...................................................... 55
   2.6 Submersions and level sets of smooth maps ......................... 58

3 **Sub-Riemannian structures** .................................................. 61

   3.1 Basic definitions .................................................... 61
     3.1.1 The minimal control and the length of an admissible curve .... 61
     3.1.2 Equivalence of sub-Riemannian structures .................... 63
     3.1.3 Examples ...................................................... 68

---

3
3.1.4 Every sub-Riemannian structure is equivalent to a free one 69
3.1.5 Proto sub-Riemannian structures 71
3.2 Sub-Riemannian distance and Chow-Rashevskii Theorem 71
3.2.1 Proof of Chow-Rashevskii Theorem 72
3.3 Existence of length-minimizers 76
3.4 Pontryagin extremals 79
3.4.1 The energy functional 80
3.4.2 Proof of Theorem 3.44 81
3.5 Measurability of the minimal control 84
3.5.1 Main lemma 84
3.5.2 Proof of Lemma 3.11 86
3.6 Lipschitz vs Absolutely continuous admissible curves 86
4 Characterization and local minimality of Pontryagin extremals 89
4.1 Geometric characterization of Pontryagin extremals 89
4.1.1 Lifting a vector field from $M$ to $T^*M$ 90
4.1.2 The Poisson bracket 91
4.1.3 Hamiltonian vector fields 93
4.2 The symplectic structure 94
4.2.1 The symplectic form vs the Poisson bracket 95
4.3 Characterization of normal and abnormal extremals 97
4.3.1 Normal extremals 97
4.3.2 Abnormal extremals 100
4.3.3 Example: codimension one distribution and contact distributions 102
4.4 Examples 103
4.4.1 2D Riemannian Geometry 103
4.4.2 Isoperimetric problem 105
4.4.3 Heisenberg group 108
4.5 Lie derivative 109
4.6 Symplectic geometry 110
4.7 Local minimality of normal trajectories 112
4.7.1 The Poincaré-Cartan one form 113
4.7.2 Normal trajectories are geodesics 115
5 Integrable Systems 119
5.1 Completely integrable systems 119
5.2 Arnold-Liouville theorem 121
5.3 Integrable geodesic flows 124
5.3.1 Geodesic flow 125
5.4 Geodesic flow on ellipsoids 128
6 Chronological calculus 131
6.1 Duality 131
6.2 Operator ODE and Volterra expansion 132
6.2.1 Volterra expansion 133
6.2.2 Adjoint representation 135
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.3</td>
<td>Variations Formulae</td>
<td>137</td>
</tr>
<tr>
<td>6.4</td>
<td>Whitney topology on smooth functions</td>
<td>138</td>
</tr>
<tr>
<td>6.5</td>
<td>Estimates of the Volterra series</td>
<td>139</td>
</tr>
<tr>
<td>7</td>
<td>Lie groups and left-invariant sub-Riemannian structures</td>
<td>143</td>
</tr>
<tr>
<td>7.1</td>
<td>Lie groups and Lie algebras</td>
<td>143</td>
</tr>
<tr>
<td>7.2</td>
<td>Left-invariant structures</td>
<td>143</td>
</tr>
<tr>
<td>7.3</td>
<td>Pontryagin extremals for left invariant structures</td>
<td>143</td>
</tr>
<tr>
<td>7.4</td>
<td>Bi-invariant metrics</td>
<td>143</td>
</tr>
<tr>
<td>7.5</td>
<td>Geodesics</td>
<td>143</td>
</tr>
<tr>
<td>8</td>
<td>End-point and Exponential map</td>
<td>145</td>
</tr>
<tr>
<td>8.1</td>
<td>The end-point map and its differential</td>
<td>145</td>
</tr>
<tr>
<td>8.2</td>
<td>Lagrange multipliers rule</td>
<td>148</td>
</tr>
<tr>
<td>8.3</td>
<td>Pontryagin extremals via Lagrange multipliers</td>
<td>148</td>
</tr>
<tr>
<td>8.4</td>
<td>Critical points and second order conditions</td>
<td>149</td>
</tr>
<tr>
<td>8.4.1</td>
<td>The manifold of Lagrange multipliers</td>
<td>152</td>
</tr>
<tr>
<td>8.5</td>
<td>Sub-Riemannian case</td>
<td>156</td>
</tr>
<tr>
<td>8.5.1</td>
<td>Free initial point problem</td>
<td>156</td>
</tr>
<tr>
<td>8.6</td>
<td>Exponential map</td>
<td>160</td>
</tr>
<tr>
<td>8.7</td>
<td>Conjugate points and minimality of extremal trajectories</td>
<td>162</td>
</tr>
<tr>
<td>8.7.1</td>
<td>Local minimality of normal extremal trajectories in the uniform topology</td>
<td>165</td>
</tr>
<tr>
<td>8.8</td>
<td>Global minimizers</td>
<td>166</td>
</tr>
<tr>
<td>8.9</td>
<td>An example: the first conjugate locus on perturbed sphere</td>
<td>166</td>
</tr>
<tr>
<td>9</td>
<td>2D-Almost-Riemannian Structures</td>
<td>173</td>
</tr>
<tr>
<td>9.1</td>
<td>Basic Definitions and properties</td>
<td>173</td>
</tr>
<tr>
<td>9.1.1</td>
<td>How big is the singular set?</td>
<td>178</td>
</tr>
<tr>
<td>9.1.2</td>
<td>Genuinely 2D-almost-Riemannian structures have always infinite area</td>
<td>179</td>
</tr>
<tr>
<td>9.1.3</td>
<td>Geodesics</td>
<td>180</td>
</tr>
<tr>
<td>9.2</td>
<td>The Grushin plane</td>
<td>181</td>
</tr>
<tr>
<td>9.2.1</td>
<td>Geodesics of the Grushin plane</td>
<td>181</td>
</tr>
<tr>
<td>9.3</td>
<td>Riemannian, Grushin and Martinet points</td>
<td>183</td>
</tr>
<tr>
<td>9.4</td>
<td>Generic 2D-almost-Riemannian structures</td>
<td>187</td>
</tr>
<tr>
<td>9.4.1</td>
<td>Proof of the genericity result</td>
<td>187</td>
</tr>
<tr>
<td>9.5</td>
<td>A Gauss-Bonnet theorem</td>
<td>189</td>
</tr>
<tr>
<td>9.5.1</td>
<td>Proof of Theorem [9.43]</td>
<td>192</td>
</tr>
<tr>
<td>9.5.2</td>
<td>Construction of trivializable 2-ARSs with no tangency points</td>
<td>195</td>
</tr>
<tr>
<td>10</td>
<td>Nonholonomic tangent space</td>
<td>197</td>
</tr>
<tr>
<td>10.1</td>
<td>Jet spaces</td>
<td>197</td>
</tr>
<tr>
<td>10.2</td>
<td>Admissible variations</td>
<td>200</td>
</tr>
<tr>
<td>10.3</td>
<td>Nilpotent approximation and privileged coordinates</td>
<td>203</td>
</tr>
<tr>
<td>10.3.1</td>
<td>Existence of privileged coordinates</td>
<td>212</td>
</tr>
<tr>
<td>10.4</td>
<td>Geometric meaning</td>
<td>216</td>
</tr>
<tr>
<td>10.5</td>
<td>Algebraic meaning</td>
<td>221</td>
</tr>
</tbody>
</table>
11 Regulation of the sub-Riemannian distance

11.1 General properties of the distance function ............................................. 223
11.2 Regularity of the squared distance .......................................................... 225
11.3 Locally Lipschitz functions and maps ......................................................... 232
  11.3.1 Locally Lipschitz map and Lipschitz submanifolds ............................... 235
  11.3.2 A non-smooth version of Sard Lemma .............................................. 238
11.4 Geodesic completeness and Hopf-Rinow theorem ....................................... 243

12 Abnormal extremals and second variation

12.1 Second variation ......................................................................................... 245
12.2 Abnormal extremals and regularity of the distance ..................................... 246
12.3 Goh and generalized Legendre conditions ............................................... 251
  12.3.1 Proof of Goh condition - (i) of Theorem 12.13 ................................. 253
  12.3.2 Proof of generalized Legendre condition - (ii) of Theorem 12.13 . . . 255
  12.3.3 More on Goh and generalized Legendre conditions ......................... 260
12.4 Rank 2 distributions and nice abnormal extremals ...................................... 261
12.5 Optimality of nice abnormal in rank 2 structures ..................................... 264
12.6 Conjugate points along abnormals ............................................................ 270
  12.6.1 Abnormals in dimension 3 ................................................................... 273
  12.6.2 Higher dimension ............................................................................. 274
12.7 Equivalence of local minimality ................................................................. 276

13 Some model spaces

13.1 Carnot groups of step 2 ............................................................................. 279
  13.1.1 Heisenberg ....................................................................................... 279
  13.1.2 (3, 6) ............................................................................................... 279
  13.1.3 (k, k(k + 1)/2) ................................................................................ 279
13.2 Other nilpotent structures ......................................................................... 279
  13.2.1 Grushin ............................................................................................ 279
  13.2.2 Martinet ........................................................................................... 279
13.3 Left invariant structures ............................................................................ 279
  13.3.1 SU(2), SO(3), SL(2) ....................................................................... 279
  13.3.2 SE(2) ............................................................................................... 279
  13.3.3 (3, 5) - Rolling sphere with twist ..................................................... 279

14 Curves in the Lagrange Grassmannian

14.1 The geometry of the Lagrange Grassmannian ........................................... 281
  14.1.1 The Lagrange Grassmannian ............................................................. 281
14.2 Regular curves in Lagrange Grassmannian .............................................. 284
14.3 Curvature of a regular curve .................................................................... 285
14.4 Reduction of non-regular curves in Lagrange Grassmannian .................... 289
14.5 Ample curves ........................................................................................... 293
14.6 From ample to regular ............................................................................ 295
14.7 Conjugate points in \( L(\Sigma) \) ................................................................ 299
14.8 Comparison theorems for regular curves .................................................. 300
15 Jacobi curves  
15.1 From Jacobi fields to Jacobi curves  
15.1.1 Jacobi curves  
15.2 Conjugate points and optimality  
15.3 Reduction of the Jacobi curves by homogeneity  
16 Riemannian curvature  
16.1 Ehresmann connection  
16.1.1 Curvature of an Ehresmann connection  
16.1.2 Linear Ehresmann connections  
16.1.3 Covariant derivative and torsion for linear connections  
16.2 Riemannian connection  
16.3 Relation with Hamiltonian curvature  
16.4 Locally flat spaces  
16.5 Example: curvature of the 2D Riemannian case  
17 Curvature in 3D contact sub-Riemannian geometry  
17.1 3D contact sub-Riemannian manifolds  
17.1.1 Curvature of a 3D contact structure  
18 Asymptotic expansion of the 3D contact exponential map  
18.1 Nilpotent case  
18.2 General case: second order asymptotic expansion  
18.3 General case: higher order asymptotic expansion  
18.3.1 Proof of Theorem 18.6 asymptotics of the exponential map  
18.3.2 Asymptotics of the conjugate locus  
18.3.3 Asymptotics of the conjugate length  
18.3.4 Stability of the conjugate locus  
19 The volume in sub-Riemannian geometry  
19.1 The Popp volume  
19.2 Popp volume for equiregular sub-Riemannian manifolds  
19.3 A formula for Popp volume  
19.4 Popp volume and isometries  
20 The sub-Riemannian heat equation  
20.1 The heat equation  
20.1.1 The heat equation in the Riemannian context  
20.1.2 The heat equation in the sub-Riemannian context  
20.1.3 Few properties of the sub-Riemannian Laplacian: the Hörmander theorem and the existence of the heat kernel  
20.1.4 The heat equation in the non-Lie-bracket generating case  
20.2 The heat-kernel on the Heisenberg group  
20.2.1 The Heisenberg group as a group of matrices  
20.2.2 The heat equation on the Heisenberg group  
A Hermite polynomials
B Elliptic functions

C Structural equations for curves in Lagrange Grassmannian
Introduction

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

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

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

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

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

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

The length of an admissible tangent vector $(\dot{x}, \dot{y}, \dot{z})$ is defined to be $(\dot{x}^2 + \dot{y}^2)^{1/2}$, that is the length of the projection of the vector to the $xy$-plane. We see that any smooth planar curve $(x(t), y(t))$ has a unique admissible lift $(x(t), y(t), z(t))$ in $\mathbb{R}^3$, where:

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

If $x(0) = y(0) = 0$, then $z(t)$ is the signed area of the domain bounded by the curve and the segment connecting $(0, 0)$ with $(x(t), y(t))$. By construction, the sub-Riemannian length of the admissible curve in $\mathbb{R}^3$ is equal to the Euclidean length of its projection to the plane.

We see that sub-Riemannian shortest paths are lifts to $\mathbb{R}^3$ of the solutions to the classical Dido isoperimetric problem: find a shortest planar curve among those connecting $(0, 0)$ with $(x_1, y_1)$ and such that the signed area of the domain bounded by the curve and the segment joining $(0, 0)$ and $(x_1, y_1)$ is equal to $z_1$ (see Figure 1).

Figure 1: The Dido problem

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

A piece of such a spiral is a shortest admissible path between its endpoints while the planar projection of this piece is an arc of the circle. The spiral ceases to be a shortest path when its planar projection starts to run the circle for the second time, i.e. when the spiral starts its second turn. Sub-Riemannian balls centered at the origin for this model look like apples with singularities at the poles (see Figure 3). Singularities are points on the sphere connected with the center by more than one shortest path. The dilation $(x, y, z) \mapsto (rx, ry, r^2z)$ transforms the ball of radius 1 into the ball of radius $r$. In particular, arbitrary small balls have singularities. This is always the case when admissible subspaces are proper subspaces.

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

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

transforms the ball of radius 1 into the ball of radius $r$. In particular, arbitrary small balls have singularities. This is always the case when admissible subspaces are proper subspaces.
turns $\mathbb{R}^3$ into a group, the Heisenberg group. The origin in $\mathbb{R}^3$ is the unit element of this group. It is easy to see that left translations of the group transform admissible curves into admissible ones and preserve the sub-Riemannian length. Hence left translations transform balls in balls of the same radius. A detailed description of this example and other models of sub-Riemannian spaces is done in Section 10.5 and Chapter 13.

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

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

Indeed, one can find a control theory spirit in our treatment of the subject. First of all, we include admissible paths in admissible flows that are flows generated by vector fields whose values in all points belong to admissible subspaces. The passage from admissible subspaces attached at different points of the manifold to a globally defined space of admissible vector fields makes the structure more flexible and well-adapted to algebraic manipulations. We pick generators \( f_1, \ldots, f_k \) of the space of admissible fields, and this allows us to describe all admissible paths as solutions to time-varying ordinary differential equations of the form:

\[
\dot{q}(t) = \sum_{i=1}^{k} u_i(t) f_i(q(t)).
\]

Different admissible paths correspond to the choice of different control functions \( u_i(\cdot) \) and initial points \( q(0) \) while the vector fields \( f_i \) are fixed at the very beginning.

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

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

![Figure 4: The “wave front” and the “impulse”](image)

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

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

The sub-Riemannian Hamiltonian \( H : T^*M \to \mathbb{R} \) is defined by the formula

\[
H(p, q) = \frac{1}{2} (p, v)^2,
\]

where \( p \in T^*_q M \), and \( v \in T_q M \) is an admissible velocity of length 1 that maximizes the inner product of \( p \) with admissible velocities of length 1 at \( q \in M \).

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

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

Recall that all velocities are admissible in the Riemannian case, and the Euclidean structure on the tangent bundle induces the identification of tangent vectors and covectors, i.e. of the velocities and impulses. We should however remember that this identification depends on the metric. One can think to a sub-Riemannian metric as the limit of a family of Riemannian metrics when the length of forbidden velocities tends to infinity, while the length of admissible velocities remains untouched.

It is easy to see that the Riemannian Hamiltonians defined by such a family converge with all derivatives to the sub-Riemannian Hamiltonian. Hence the Riemannian geodesics with a prescribed initial impulse converge to the sub-Riemannian geodesic with the same initial impulse. On the other hand, we cannot expect any reasonable convergence for the family of Riemannian geodesics with a prescribed initial velocity: those with forbidden initial velocities disappear at the limit while geodesics with admissible initial velocities multiply.

Outline of the book

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

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

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

Chapters 1–5 form a first part of the book where we do not use any tool from functional analysis. In fact, even the knowledge of the Lebesgue integration and elementary real analysis are not essential with a unique exception of the existence theorem in Section 3.3. In all other places the reader can substitute terms “Lipschitz” and “absolutely continuous” by “piecewise $C^1$” and
“measurable” by “piecewise continuous” without a loss for the understanding.

We start to use some basic functional analysis in Chapter 6. In this chapter, we give elements of an operator calculus that simplifies and clarifies calculations with non-stationary flows, their variations and compositions. In Chapter 7, we use this calculus for a fast introduction to the Lie group theory.

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

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

Chapter 13 is devoted to the explicit calculation of the sub-Riemannian distance for model spaces. In Chapter 11, we study general analytic properties of the sub-Riemannian distance as a function of points of the manifold. It is shown that the distance is smooth on an open dense subset and is semi-concave out of the points connected by abnormal length-minimizers. Moreover, generic sphere is a Lipschitz submanifold if we remove these bad points.

In Chapter 12, we turn to abnormal geodesics, which provide the deepest singularities of the distance. Abnormal geodesics are critical points of the endpoint map defined on the space of admissible paths, and the main tool for their study is the Hessian of the endpoint map.

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

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

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

Chapter 14 is devoted to the elementary differential geometry of curves in the Lagrangian Grassmannian; in Chapter 15 we apply this geometry to Jacobi curves.

The language of Jacobi curves is translated to the traditional language in the Riemannian case in Chapter 16. We recover the Levi Civita connection and the Riemannian curvature and demonstrate their symplectic meaning. In Chapter 17 we explicitly compute the sub-Riemannian curvature for contact three-dimensional spaces. In the next Chapter 18 we study the small distance asymptotics of the expowhree-dimensional contact case and see how the structure of the conjugate locus is encoded in the curvature.

In Chapter ??, we consider two-dimensional sub-Riemannian metrics; such a metric differs from a Riemannian one only along a one-dimensional submanifold. In the last Chapter 20 we define the
sub-Riemannian Laplace operator, the canonical volume form, and compute the density of the sub-Riemannian Hausdorff measure. We conclude with a discussion of the sub-Riemannian heat equation and an explicit formula for the heat kernel in the three-dimensional Heisenberg case.

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

---

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

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

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

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

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

### 1.1 Geodesics and optimality

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

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

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

**Remark 1.2.** Notice that the definition of length in (1.1) is invariant by reparametrizations of the curve. Indeed let $\varphi : [0, T'] \to [0, T]$ be a monotone smooth function. Define $\gamma_{\varphi} := \gamma \circ \varphi$. Using the change of variables $t = \varphi(s)$, one gets

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

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

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

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

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

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

17
Exercise 1.3. Prove that, if \( \gamma : [0, T] \to M \) is a length-minimizer, then the curve \( \gamma|_{[t_1, t_2]} \) is also a length-minimizer, for all \( 0 < t_1 < t_2 < T \).

The following proposition characterizes smooth minimizers. We prove later that all minimizers are smooth (cf. Corollary 1.15).

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

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

\[
\dot{f}_t(\gamma(t)) = \dot{\gamma}(t) \quad \text{and} \quad \|f_t(q)\| \equiv 1, \quad \forall (t,q) \in W.
\]

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

\[
\langle f_t(q) | g_t(q) \rangle = 0, \quad \|g_t(q)\| \equiv 1, \quad \forall (t,q) \in W.
\]

Piecewise smooth curves parametrized by length on \( M \) are solutions of the following ordinary differential equation

\[
\dot{x}(t) = \cos u(t)f_t(x(t)) + \sin u(t)g_t(x(t)), \quad (1.3)
\]

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

Let us consider the family of controls

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

and denote by \( x_{\tau,s}(t) \) the solution of (1.3) that corresponds to the control \( u_{\tau,s}(t) \) and with initial condition \( x_{\tau,s}(0) = \gamma(0) \).
Lemma 1.5. For every $\tau_1,\tau_2, t \in [0,T]$ the following vectors are linearly dependent

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

(1.5)

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

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

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

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

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

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

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

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

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

(1.6)

Proof. If $t \leq \tau$, then by construction (cf. (1.4)) the first vector is zero since there is no variation w.r.t. $s$ and the conclusion follows. Let us now assume that $t > \tau$. Again, by Remark 1.3 it is sufficient to prove the statement at $t = T$. Let us write the Taylor expansion of $\psi(t) = \frac{\partial}{\partial s}\bigg|_{s=0} x_{\tau,s}(t)$ in a right neighborhood of $t = \tau$. Observe that, for $t \geq \tau$

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

Hence

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

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

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

(1.7)

For $\tau$ sufficiently close to $T$, one can take $t = T$ in (1.7). Passing to the limit for $\tau \to T$ one gets

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

Now, by Lemma 1.5 all vectors in left hand side are parallel among them, hence they are parallel to $g_T(\gamma(T))$. The lemma is proved since $\dot{\gamma}(T) = f_T(\gamma(T))$ and $f_T$ and $g_T$ are orthogonal. \qed
Now we end the proposition by showing that $\ddot{\gamma}(t) \perp T_{\gamma(t)}M$. Notice that this is equivalent to show
\[
\langle \ddot{\gamma}(t) \mid f_t(\gamma(t)) \rangle = \langle \ddot{\gamma}(t) \mid g_t(\gamma(t)) \rangle = 0.
\] (1.8)
Recall that $\langle \dot{\gamma}(t) \mid \dot{\gamma}(t) \rangle = 1$. Differentiating this identity one gets
\[
0 = \frac{d}{dt} \langle \dot{\gamma}(t) \mid \dot{\gamma}(t) \rangle = 2 \langle \ddot{\gamma}(t) \mid \dot{\gamma}(t) \rangle,
\]
which shows that $\ddot{\gamma}(t)$ is orthogonal to $f_t(\gamma(t))$. Next, differentiating (1.6) with respect to $t$, we have
t for $t \neq \tau$
\[
\left\langle \frac{\partial}{\partial s} x_{\tau,s}(t) \mid \dot{x}_{\tau,s}(t) \right\rangle + \left\langle \frac{\partial}{\partial s} x_{\tau,s}(t) \mid \ddot{x}_{\tau,s}(t) \right\rangle = 0.
\] (1.9)
Now, from $\langle \dot{x}_{\tau,s}(t) \mid \dot{x}_{\tau,s}(t) \rangle = 1$ one gets
\[
\left\langle \frac{\partial}{\partial s} x_{\tau,s}(t) \mid \dot{x}_{\tau,s}(t) \right\rangle = 0, \quad \text{for } t \neq \tau.
\]
Evaluating at $s = 0$, using that $x_{\tau,0}(t) = \gamma(t)$, one has
\[
\left\langle \frac{\partial}{\partial s} x_{\tau,s}(t) \mid \dot{\gamma}(t) \right\rangle = 0, \quad \text{for } t \neq \tau.
\]
Hence, by (1.9), it follows that
\[
\left\langle \frac{\partial}{\partial s} x_{\tau,s}(t) \mid \ddot{\gamma}(t) \right\rangle = 0,
\]
which, by continuity, holds for every $t \in [0,T]$. Using that $\frac{\partial}{\partial s} x_{\tau,s}(t)$ is parallel to $g_t(\gamma(t))$ (see proof of Lemma 1.6), it follows that $\langle g_t(\gamma(t)) \mid \dot{\gamma}(t) \rangle = 0$. \hfill $\square$

**Definition 1.7.** A smooth curve $\gamma : [0,T] \to M$ parametrized with constant speed is called *geodesic* if it satisfies
\[
\ddot{\gamma}(t) \perp T_{\gamma(t)}M, \quad \forall t \in [0,T].
\] (1.10)
Proposition 1.4 says that a smooth curve that minimizes the length is a geodesic.

Now we get an explicit characterization of geodesics when the manifold $M$ is globally defined as the zero level of a smooth function. In other words there exists a smooth function $a : \mathbb{R}^3 \to \mathbb{R}$ such that
\[
M = a^{-1}(0), \quad \text{and} \quad \nabla a \neq 0 \text{ on } M.
\] (1.11)

**Remark 1.8.** Recall that for all $q \in M$ it holds $\nabla_q a \perp T_q M$. Indeed, for every $q \in M$ and $v \in T_q M$, let $\gamma : [0,T] \to M$ be a smooth curve on $M$ such that $\gamma(0) = q$ and $\dot{\gamma}(0) = v$. By definition of $M$ one has $a(\gamma(t)) = 0$. Differentiating this identity with respect to $t$ at $t = 0$ one gets $\langle \nabla_q a \mid v \rangle = 0$.

**Proposition 1.9.** A smooth curve $\gamma : [0,T] \to M$ is a geodesic if and only if it satisfies, in matrix notation:
\[
\ddot{\gamma}(t) = -\frac{\dot{\gamma}(t)^T (\nabla_{\gamma(t)}^2 a) \dot{\gamma}(t)}{\|\nabla_{\gamma(t)} a\|^2} \nabla_{\gamma(t)} a, \quad \forall t \in [0,T],
\] (1.12)
where $\nabla_{\gamma(t)}^2 a$ is the Hessian matrix of $a$.

\begin{footnote}{notice that $x_{\tau,s}$ is smooth on the set $[0,T] \setminus \{\tau\}$.} \end{footnote}
Proof. Differentiating the equality $\left< \nabla_{\gamma(t)}a \mid \dot{\gamma}(t) \right> = 0$ we get, in matrix notation:
\[
\dot{\gamma}(t)^T (\nabla^2_{\gamma(t)}a) \dot{\gamma}(t) + \ddot{\gamma}(t)^T \nabla_{\gamma(t)}a = 0.
\]
By definition of geodesic there exists a function $b(t)$ such that
\[
\ddot{\gamma}(t) = b(t) \nabla_{\gamma(t)}a.
\]
Hence we get
\[
\dot{\gamma}(t)^T (\nabla^2_{\gamma(t)}a) \dot{\gamma}(t) + b(t) \| \nabla_{\gamma(t)}a \|^2 = 0,
\]
from which (1.12) follows. \(\square\)

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

1.1.1 Existence and minimizing properties of geodesics

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

Corollary 1.11. Let $q \in M$ and $v \in T_qM$. There exists a unique geodesic $\gamma : [0, \varepsilon] \to M$, for $\varepsilon > 0$ small enough, such that $\gamma(0) = q$ and $\dot{\gamma}(0) = v$.

Proof. By Proposition 1.9 geodesics satisfy a second order ODE, hence they are smooth curves, characterized by their initial position and velocity. \(\square\)

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

Theorem 1.12. Let $\gamma : [0, T] \to M$ be a geodesic. For every $\tau \in [0, T]$ there exists $\varepsilon > 0$ such that

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

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

Proof. Without loss of generality let us assume that $\tau = 0$ and that $\gamma$ is length parametrized. Consider a length-parametrized curve $\alpha$ on $M$ such that $\alpha(0) = \gamma(0)$ and $\dot{\alpha}(0) \perp \dot{\gamma}(0)$ and denote by $(t, s) \mapsto x_s(t)$ the smooth variation of geodesics such that $x_0(t) = \gamma(t)$ and (see also Figure 1.2)
\[
x_s(0) = \alpha(s), \quad \dot{x}_s(0) \perp \dot{\alpha}(s).
\]
(1.13)

The map $\psi : (t, s) \mapsto x_s(t)$ is a local diffeomorphism near $(0, 0)$. Indeed the partial derivatives
\[
\left. \frac{\partial \psi}{\partial t} \right|_{t=s=0} = \left. \frac{\partial}{\partial t} \right|_{t=0} x_0(t) = \dot{\gamma}(0), \quad \left. \frac{\partial \psi}{\partial s} \right|_{t=s=0} = \left. \frac{\partial}{\partial s} \right|_{s=0} x_s(0) = \dot{\alpha}(0),
\]
are linearly independent. Thus $\psi$ maps a neighborhood $U$ of $(0, 0)$ on a neighborhood $W$ of $\gamma(0)$. We now consider the function $\phi$ and the vector field $X$ defined on $W$
\[
\phi : x_s(t) \mapsto t, \quad X : x_s(t) \mapsto \dot{x}_s(t).
\]
Lemma 1.13. $\nabla_q \phi = X(q)$ for every $q \in W$.

Proof of Lemma 1.13. We first show that the two vectors are parallel, and then that they actually coincide. To show that they are parallel, first notice that $\nabla \phi$ is orthogonal to its level set $\{t = \text{const}\}$, hence
\[ \left\langle \nabla x_s(t) \phi, \frac{\partial}{\partial s} x_s(t) \right\rangle = 0, \quad \forall (t, s) \in U. \] (1.14)

Now, let us show that
\[ \left\langle \frac{\partial}{\partial s} x_s(t), \dot{x}_s(t) \right\rangle = 0, \quad \forall (t, s) \in U. \] (1.15)

Computing the derivative with respect to $t$ of the left hand side of (1.15) one gets
\[ \left\langle \frac{\partial}{\partial s} \dot{x}_s(t), \dot{x}_s(t) \right\rangle + \left\langle \frac{\partial}{\partial s} x_s(t), \ddot{x}_s(t) \right\rangle, \]
which is identically zero. Indeed the first term is zero because $\dot{x}_s(t)$ has unit speed and the second one vanishes because of (1.10). Hence, the left hand side of (1.15) is constant and coincides with its value at $t = 0$, which is zero by the orthogonality assumption (1.13).

By (1.14) and (1.15) one gets that $\nabla \phi$ is parallel to $X$. Actually they coincide since
\[ \left\langle \nabla \phi \mid X \right\rangle = \frac{d}{dt} \phi(x_s(t)) = 1. \]

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

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

The last inequality follows from the Cauchy-Schwartz inequality

\[
\langle X(c(t)) \mid \dot{c}(t) \rangle \leq \|X(c(t))\| \|\dot{c}(t)\| = 1
\]

which holds at every smooth point of \( c(t) \). In addition, equality in (1.18) holds if and only if \( \dot{c}(t) = X(c(t)) \) (at the smooth points of \( c \)). Hence we get that \( \ell(c) = \ell(\gamma|_{[0, \varepsilon]}) \) if and only if \( c \) coincides with \( \gamma|_{[0, \varepsilon]} \).

Now let us show that there exists \( \bar{\varepsilon} \leq \varepsilon \) such that \( \gamma|_{[0, \bar{\varepsilon}]} \) is a global minimizer among all piecewise smooth curves joining \( \gamma(0) \) to \( \gamma(\bar{\varepsilon}) \). It is enough to take \( \bar{\varepsilon} < \text{dist}(\gamma(0), \partial W) \). Every curve that escape from \( W \) has length greater than \( \bar{\varepsilon} \).

From Theorem 1.12 it follows

**Corollary 1.14.** Any minimizer of the distance (in the class of piecewise smooth curves) is a geodesic, and hence smooth.

### 1.1.2 Absolutely continuous curves

Notice that formula (1.1) defines the length of a curve even in the class of absolutely continuous ones, if one understands the integral in the Lebesgue sense.

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

**Corollary 1.15.** Any minimizer of the distance (in the class of absolutely continuous curves) is a geodesic, and hence smooth.

### 1.2 Parallel transport

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

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

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

\[\text{The canonical isomorphism } \mathbb{R}^2 \simeq T_x \mathbb{R}^2 \text{ is written explicitly as follows: } y \mapsto \frac{d}{dt} \big|_{t=0} x + ty.\]
Definition 1.17. Let $\gamma : [0, T] \to M$ be a smooth curve. A smooth curve of tangent vectors $\xi(t) \in T_{\gamma(t)}M$ is said to be parallel if $\dot{\xi}(t) \perp T_{\gamma(t)}M$.

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

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

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

(1.19)

Proof. As in Remark 1.8, $\xi(t) \in T_{\gamma(t)}M$ implies $\langle \nabla_{\gamma(t)} a, \xi(t) \rangle = 0$. Moreover, by assumption $\dot{\xi}(t) = \alpha(t)\nabla_{\gamma(t)} a$ for some smooth function $\alpha$. With analogous computations as in the proof of Proposition 1.9 we get that

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

from which the statement follows.

Remark 1.19. Notice that, since (1.53) is a first order linear ODE with respect to $\xi$, for a given curve $\gamma : [0, T] \to M$ and initial datum $v \in T_{\gamma(0)}M$, there is a unique parallel curve of tangent vectors $\xi(t) \in T_{\gamma(t)}M$ along $\gamma$ such that $\xi(0) = v$. Since (1.53) is a linear ODE, the operator that associates with every initial condition $\xi(0)$ the final vector $\xi(t)$ is a linear operator, which is called parallel transport.

Next we state a key property of the parallel transport.

Proposition 1.20. The parallel transport preserves the inner product. In other words, if $\xi(t), \eta(t)$ are two parallel curves of tangent vectors along $\gamma$, then we have

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

(1.20)

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

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

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

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

In the following we assume that $M$ is oriented.

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

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

(1.21)
$SM$ is a smooth manifold of dimension 3. Moreover it has the structure of fiber bundle with base manifold $M$, typical fiber $S^1$, and canonical projection

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

Remark 1.23. Since every vector in the fiber $S_qM$ has norm one, we can parametrize every $v \in S_qM$ by an angular coordinate $\theta \in S^1$ through an orthonormal frame \{e_1(q), e_2(q)\} for $S_qM$, i.e. $v = \cos(\theta)e_1(q) + \sin(\theta)e_2(q)$.

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

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

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

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

Let $t \mapsto \xi(t) \in S_{\gamma(t)}M$ be a smooth curve of unit tangent vectors along $\gamma : [0, T] \to M$. Define $\eta(t), \nu(t) \in T_{\gamma(t)}M$ as above. Since $t \mapsto \xi(t)$ has constant speed, one has $\xi(t) \perp \dot{\xi}(t)$ and we can write

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

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

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

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

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

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

Lemma 1.25. A curve of unit tangent vectors $\xi(t)$ is parallel if and only if $\omega_\xi(t)(\dot{\xi}(t)) = 0$.

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

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

$$\omega_{\gamma(t)}(\dot{\gamma}(t)) = 0, \quad \forall t \in [0, T].$$  

(1.23)

Proposition 1.26. The Levi-Civita connection $\omega \in \Lambda^1(SM)$ satisfies:

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

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

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

25
(ii) \( d\omega = \pi^*\Omega \), where \( \Omega \) is a 2-form defined on \( M \) and \( \pi : SM \to M \) is the canonical projection.

**Proof.** (i) Fix a system of coordinates \((x_1, x_2, \theta)\) on \( SM \) and consider the vector field \( \partial/\partial \theta \) on \( SM \). Let us show that

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

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

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

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

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

**Definition 1.28.** The area form \( dV \) on a surface \( M \) is the differential two form that on every tangent space to the manifold agrees with the volume induced by the inner product. In other words, for every positively oriented orthonormal frame \( e_1, e_2 \) of \( T_qM \), one has \( dV(e_1, e_2) = 1 \).

Given a set \( \Gamma \subset M \) its **area** is the quantity \( |\Gamma| = \int_{\Gamma} dV \).

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

**Definition 1.29.** The **Gaussian curvature** of \( M \) is the function \( \kappa : M \to \mathbb{R} \) defined by the equality

\[
\Omega = -\kappa dV.
\]  \hspace{1cm} (1.25)

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

### 1.3 Gauss-Bonnet Theorems

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

**Definition 1.30.** Let \( \gamma : [0, T] \to M \) be a smooth curve parametrized by length. The **geodesic curvature** of \( \gamma \) is defined as

\[
\rho_\gamma(t) = \omega_{\dot{\gamma}(t)}(\ddot{\gamma}(t)).
\]  \hspace{1cm} (1.26)

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

**Remark 1.31.** The geodesic curvature changes sign if we move along the curve in the opposite direction. Moreover, if \( M = \mathbb{R}^2 \), it coincides with the usual notion of curvature of a planar curve.
1.3.1 Gauss-Bonnet theorem: local version

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

![Figure 1.3: A curvilinear polygon](image.png)

Notice that a curvilinear polygon is homeomorphic to a disk.

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

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

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

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

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

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

For $\varepsilon$ small enough, we define (cf. Figure 1.4)

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

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

$$\phi : \Gamma_\varepsilon \to SM, \quad \phi(q) = \frac{X(q)}{|X(q)|}.$$
First notice that
\[ \int_{\phi(\Gamma_\varepsilon)} d\omega = \int_{\phi(\Gamma_\varepsilon)} \pi^*\Omega = \int_{\pi(\phi(\Gamma_\varepsilon))} \Omega = \int_{\Gamma_\varepsilon} \Omega, \]  
where we used the fact that \( \pi(\phi(\Gamma_\varepsilon)) = \Gamma_\varepsilon \). Then let us compute the integral of the curvature \( \kappa \) on \( \Gamma_\varepsilon \)
\[ \int_{\Gamma_\varepsilon} \kappa dV = - \int_{\Gamma_\varepsilon} \Omega = - \int_{\phi(\Gamma_\varepsilon)} d\omega, \] (by (1.28))
\[ = - \int_{\partial\phi(\Gamma_\varepsilon)} \omega, \] (by Stokes Theorem)
\[ = \int_{\phi(A_\varepsilon)} \omega - \int_{\phi(\partial\Gamma)} \omega, \] (since \( \partial\phi(\Gamma_\varepsilon) = \phi(A_\varepsilon) \cup \phi(\partial\Gamma) \)) (1.29)

Notice that in the third equality we used the fact that the induced orientation on \( \partial\phi(\Gamma_\varepsilon) \) gives opposite orientation on the two terms. Let us treat separately these two terms. The first one, by Proposition 1.55 can be written as
\[ \int_{\phi(A_\varepsilon)} \omega = \int_{\phi(A_\varepsilon)} d\theta + \int_{\phi(A_\varepsilon)} a_1(x_1, x_2) dx_1 + a_2(x_1, x_2) dx_2 \] (1.30)

The first element of (1.30) is equal to \( 2\pi \) since we integrate the 1-form \( d\theta \) on a closed curve. The second element of (1.30), for \( \varepsilon \to 0 \), satisfies
\[ \left| \int_{\phi(A_\varepsilon)} a_1(x_1, x_2) dx_1 + a_2(x_1, x_2) dx_2 \right| \leq C\ell(\phi(A_\varepsilon)) \to 0, \] (1.31)

Indeed the functions \( a_i \) are smooth (hence bounded on compact sets) and the length of \( \phi(A_\varepsilon) \) goes to zero for \( \varepsilon \to 0 \).
Let us now consider the second term of (1.29). Since $\phi(\partial \Gamma)$ is parametrized by the curve $t \mapsto \dot{\gamma}(t)$ (as a curve on $SM$), we have

$$\int_{\phi(\partial \Gamma)} \omega = \int_I \omega \gamma'(t)dt = \int_I \rho_{\gamma}(t)dt.$$ 

Concluding we have from (1.29)

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

that is (1.27) in the smooth case (i.e. when $\alpha_i = 0$ for all $i$).

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

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

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

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

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

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

We are left to prove that

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

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

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

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

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

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

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

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

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

Then (1.32) follows. \qed
An important corollary is obtained by applying the Gauss-Bonnet Theorem to geodesic triangles. A geodesic triangle $T$ is a curvilinear polygon with $m = 3$ edges and such that every smooth piece of boundary $\gamma_i$ is a geodesic. For a geodesic triangle $T$ we denote by $A_i := \pi - \alpha_i$ its internal angles.

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

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

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

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

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

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

\[\square\]

### 1.3.2 Gauss-Bonnet theorem: global version

Now we state the global version of the Gauss-Bonnet theorem. In other words we want to generalize (1.27) to the case when $\Gamma$ is a region of $M$ not necessarily homeomorphic to the disk, see for instance Figure 1.5. As we will see that the result depends on the Euler characteristic $\chi(\Gamma)$ of this region.

In what follows, by a triangulation of $M$ we mean a decomposition of $M$ into curvilinear polygons (see Definition 1.32). Notice that every compact surface admits a triangulation. \[^3\]

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

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

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

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

**Remark 1.36.** The Euler characteristic is well-defined. Indeed one can show that the quantity (1.34) is invariant for refinement of a triangulation, since every at every step of the refinement the alternating sum does not change. Moreover, given two different triangulations of the same region, there always exists a triangulation that is a refinement of both of them. This shows that the quantity (1.34) is independent on the triangulation.

**Example 1.37.** For a compact connected orientable surface $M_g$ of genus $g$ (i.e., a surface that topologically is a sphere with $g$ handles) one has $\chi(M_g) = 2 - 2g$. For instance one has $\chi(S^2) = 2$, $\chi(\mathbb{T}^2) = 0$, where $\mathbb{T}^2$ is the torus. Notice also that $\chi(B_1) = 1$, where $B_1$ is the closed unit disk in $\mathbb{R}^2$.

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

30
Following the notation introduced in the previous section, for a given region $\Gamma$, we assume that $\partial \Gamma$ is oriented consistently with the orientation of $M$ and $\partial \Gamma = \bigcup_{i=1}^{m} \gamma_i(I_i)$ where $\gamma_i : I_i \to M$, for $i = 1, \ldots, m$, are smooth curves parametrized by length (with orientation consistent with $\partial \Gamma$). We denote by $\alpha_i$ the external angles at the points where $\partial \Gamma$ is not $C^1$ (see Figure 1.5).

Figure 1.5: Gauss-Bonnet Theorem

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

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

**Proof.** As in the proof of the local version of the Gauss-Bonnet theorem we consider two cases:

(i) Case $\partial \Gamma$ smooth (in particular $\alpha_i = 0$ for all $i$).

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

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

We have that

$$\sum_{j=1}^{n_2} \int_{\Gamma_j} \kappa dV = \int_{\Gamma} \kappa dV, \quad \sum_{j,k} \int_{I_k} \rho_{\gamma_k^{(j)}}(t)dt = \sum_{i=1}^{m} \int_{I_i} \rho_{\gamma_i}(t)dt.$$ \hspace{1cm} (1.37)

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

$$\sum_{j,k} \theta^{(j)}_k = 2\pi (n_1 - n_0), \quad (1.38)$$

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

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

Denote by $n^\partial_0$ the number of vertexes that belong to $\partial \Gamma$ and with $n^I_0 := n_0 - n^\partial_0$. Similarly we define $n^\partial_1$ and $n^I_1$. We have the following relations:

(i) $N = 2n^I_1 + n^\partial_1$,

(ii) $n^\partial_0 = n^\partial_1$.

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

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

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

$$\sum_{\nu=1}^N A_\nu = 2\pi n^I_0 + \pi n^\partial_0, \quad (1.41)$$

Combining (1.40), (1.41) and (i) one has

$$\sum_{i=1}^\nu \theta_\nu = (2n^I_1 + n^\partial_1)\pi - (2n^I_0 + n^\partial_0)\pi$$

Using (ii) one finally gets (1.39).

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

We consider a decomposition of $\Gamma$ into curvilinear polygons whose edges intersect the boundary in the smooth part (this is always possible). The proof is identical to the smooth case up to formula (1.37). Now, instead of (1.39), we have to check that

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

Now (1.42) can be rewritten as

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

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

32
Consider now a new region \( \tilde{\Gamma} \), obtained by smoothing the edges of \( \Gamma \), together with the decomposition induced by \( \Gamma \) (see Figure 1.5). Denote by \( \tilde{n}_1 \) and \( \tilde{n}_0 \) the number of edges and vertexes of the decomposition of \( \tilde{\Gamma} \). Notice that \( \{\theta_\nu, \nu \notin A\} \) is exactly the set of all angles of the decomposition of \( \tilde{\Gamma} \). Moreover \( \tilde{n}_1 - \tilde{n}_0 = n_1 - n_0 \), since \( n_0 = \tilde{n}_0 + m \) and \( n_1 = \tilde{n}_1 + m \), where \( m \) is the number of non-smooth points. Hence, by part (i) of the proof:

\[
\sum_{\nu \notin A} \theta_\nu = 2\pi(\tilde{n}_1 - \tilde{n}_0) = 2\pi(n_1 - n_0).
\]

\[\square\]

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

\[
\int_M \kappa dV = 2\pi \chi(M).
\] (1.43)

### 1.3.3 Consequences of the Gauss-Bonnet Theorems

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

\[
\langle v | w \rangle = \langle D_q \phi(v) | D_q \phi(w) \rangle, \quad \forall v, w \in T_q M.
\] (1.44)

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

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

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

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

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

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

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

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

As a converse of Corollary 1.32 we have the following.
Theorem 1.44. Assume that \( \kappa \equiv 0 \) in a neighborhood of a point \( q \in M \). Then \( M \) is locally Euclidean (i.e., locally isometric to \( \mathbb{R}^2 \)) around \( q \).

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

\[
\Omega = \kappa dV = 0.
\]

Hence \( d\omega = \pi^*\Omega = 0 \). From its explicit expression

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

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

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

Thus we can define a new angular coordinate on \( SM \), which we still denote by \( \theta \), in such a way that (see also Remark 1.27)

\[
\omega = d\theta.
\] (1.45)

Now, let \( \gamma \) be a length parametrized geodesic, i.e. \( \omega_{\dot{\gamma}(t)}(\dot{\gamma}(t)) = 0 \). Using the angular coordinate \( \theta \) just defined on the fibers of \( SM \), the curve \( t \mapsto \dot{\gamma}(t) \in S_{\gamma(t)}M \) is written as \( t \mapsto \theta(t) \). Using (1.45), we have then

\[
0 = \omega_{\dot{\gamma}(t)}(\dot{\gamma}(t)) = d\theta(\dot{\gamma}(t)) = \dot{\theta}(t).
\]

In other words the angular coordinate of a geodesic \( \gamma \) is constant.

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

Then, for each point \( q' \in U \) consider the two geodesics starting from \( q' \) and satisfying \( \theta_1(0) = 0 \) and \( \theta_2(0) = \pi/2 \). We assign coordinates \( (x_1, x_2) \) to each point \( q' \in U \) by considering the length parameter of the geodesic projection of \( q' \) on \( \gamma_1 \) and \( \gamma_2 \) (See Figure 1.6). Notice that the family of geodesics constructed in this way, and parametrized by \( q' \in U \), are mutually orthogonal at every point.

By construction, in this coordinate system the vectors \( \partial_{x_1} \) and \( \partial_{x_2} \) have length one (being the tangent vectors to length parametrized geodesics) and are everywhere mutually orthogonal. Hence the theorem follows from Exercise 1.43. \( \square \)

1.3.4 The Gauss map

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

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

\[
N : M \to S^2, \quad q \mapsto \nu_q,
\] (1.46)

where \( \nu_q \in S^2 \subset \mathbb{R}^3 \) denotes the external unit normal vector to \( M \) at \( q \).
Let us consider the differential of the Gauss map at the point $q$

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

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

**Theorem 1.46.** We have that $\kappa(q) = \det(D_qN)$.

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

**Lemma 1.47.** For every $q \in M$, the differential $D_qN$ of the Gauss map is a symmetric operator, i.e.,

$$\langle D_qN(\xi) \mid \eta \rangle = \langle \xi \mid D_qN(\eta) \rangle, \quad \forall \xi, \eta \in T_qM. \tag{1.47}$$

**Proof.** We prove the statement locally, i.e., for a manifold $M$ parametrized by a function $\phi : \mathbb{R}^2 \to M$. In this case $T_qM = \text{Im} \ D_u\phi$, where $\phi(u) = q$. Let $v, w \in \mathbb{R}^2$ such that $\xi = D_u\phi(v)$ and $\eta = D_u\phi(w)$. Since $N(q) \in T_qM^\perp$ we have $\langle N(q) \mid \eta \rangle = \langle N(q) \mid D_u\phi(w) \rangle = 0$. Taking the derivative in the direction of $\xi$ one gets

$$\langle D_qN(\xi) \mid \eta \rangle + \langle N(q) \mid D_u^2\phi(v, w) \rangle = 0,$$

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

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

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

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

$$d_\xi e_i(\eta) = \sum_{j=1}^{3} (\omega(\xi))_{ij}(\eta) e_j(\xi), \quad \omega_{ij} \in \Lambda^1 SM, \quad \eta \in T_\xi SM.$$
Dropping $\xi$ and $\eta$ from the notation one gets the relation

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

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

$$de_1 = \omega_{12} e_2 + \omega_{13} e_3,$$
$$de_2 = -\omega_{12} e_1 + \omega_{23} e_3,$$
$$de_3 = -\omega_{13} e_1 - \omega_{23} e_2.$$  

Let us now prove the following identity

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

Indeed, differentiating the first equation in (1.48) one gets, using that $d^2 = 0$,

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

which implies in particular (1.49).

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

$$\det(D_q N)dV = \kappa dV.$$  

Applying $\pi^*$ to both sides one gets

$$\pi^*(\det(D_q N)dV) = \pi^*\kappa dV = d\omega$$  

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

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

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

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

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

The proof is completed by the following

**Exercise 1.48.** Let $V$ be a 2-dimensional Euclidean vector space and $e_1, e_2$ an orthonormal basis. Let $F : V \to V$ a linear map and write $F = F_1 e_1 + F_2 e_2$, where $F_i : V \to \mathbb{R}$ are linear functionals. Prove that $F_1 \wedge F_2 = (\det F)dV$, where $dV$ is the area form induced by the inner product.
Remark 1.49. Lemma 1.47 allows us to define the principal curvatures of $M$ at the point $q$ as the two real eigenvalues $k_1(q), k_2(q)$ of the map $D_qN$. In particular

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

The principal curvatures can be geometrically interpreted as the maximum and the minimum of curvature of sections of $M$ with orthogonal planes.

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

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

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

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

$$\langle q_1, q_2 \rangle_h = x_1x_2 + y_1y_2 - z_1z_2. \quad (1.51)$$

Here $q_i = (x_i, y_i, z_i)$ for $i = 1, 2$, are two points in $\mathbb{R}^3$. When $\langle q, q \rangle_h \geq 0$, we denote by $\|q\|_h = \langle q, q \rangle_h^{1/2}$ the norm induced by the inner product (1.51).

For the metric structure to be defined on $M$, we require that the restriction of the inner product (1.51) to the tangent space to $M$ is positive definite at every point. Indeed, under this assumption, the inner product (1.51) can be used to define the length of a tangent vector to the surface (which is non-negative). Thus one can introduce the length of (piecewise) smooth curves on $M$ and its distance by the same formulas as in Section 1.1. These surfaces are also called space-like surfaces in the Minkowski space.

The structure of the inner product impose some condition on the structure of space-like surfaces, as the following exercise shows.

**Exercise 1.50.** Let $M$ be a space-like surface in $\mathbb{R}^3$ endowed with the inner product (1.51).

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

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

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

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

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

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

$$\nabla^h a = \left( \frac{\partial a}{\partial x}, \frac{\partial a}{\partial y}, -\frac{\partial a}{\partial z} \right).$$
If we assume that $M = a^{-1}(0)$ is a regular level set of a smooth function $a : \mathbb{R}^3 \to \mathbb{R}$. If $\gamma(t)$ is a curve contained in $M$, i.e. $a(\gamma(t)) = 0$, one has the identity
\[
0 = \left\langle \nabla_{\gamma(t)}^h a \mid \dot{\gamma}(t) \right\rangle_h.
\]
The same computation shows that $\nabla_{\gamma(t)}^h a$ is orthogonal to the level sets of $a$, where orthogonal always means with respect to $\langle \cdot \mid \cdot \rangle_h$. In particular, if $M = a^{-1}(0)$ is space-like, one has $\langle \nabla_q a, \nabla_q a \rangle_h < 0$.

**Exercise 1.51.** Let $\gamma$ be a geodesic on $M = a^{-1}(0)$. Show that $\gamma$ satisfies the equation (in matrix notation)
\[
\dot{\gamma}(t) = -\frac{\dot{\gamma}(t)^T (\nabla_{\gamma(t)}^2 a) \dot{\gamma}(t)}{\|\nabla_{\gamma(t)}^h a\|_h^2} \nabla_{\gamma(t)}^h a, \quad \forall t \in [0, T].
\] (1.52)
where $\nabla_{\gamma(t)}^2 a$ is the (classical) matrix of second derivatives of $a$.

Given a smooth curve $\gamma : [0, T] \to M$ on a surface $M$, a smooth curve of tangent vectors $\xi(t) \in T_{\gamma(t)}M$ is said to be parallel if $\dot{\xi}(t) \perp T_{\gamma(t)}M$, with respect to the hyperbolic inner product. It is then straightforward to check that, if $M$ is the zero level of a smooth function $a : \mathbb{R}^3 \to \mathbb{R}$, then $\dot{\xi}(t)$ is parallel along $\gamma$ if and only if it satisfies
\[
\dot{\dot{\xi}}(t) = -\frac{\dot{\xi}(t)^T (\nabla_{\xi(t)}^2 a) \dot{\xi}(t)}{\|\nabla_{\xi(t)}^h a\|_h^2} \nabla_{\xi(t)}^h a, \quad \forall t \in [0, T].
\] (1.53)
By definition a smooth curve $\gamma : [0, T] \to M$ is a geodesic if and only if $\dot{\gamma}$ is parallel along $\gamma$.

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

**Exercise 1.53.** Show that if $\xi(t), \eta(t)$ are two parallel curves of tangent vectors along $\gamma$, then we have
\[
\frac{d}{dt} \langle \xi(t) \mid \eta(t) \rangle_h = 0, \quad \forall t \in [0, T].
\] (1.54)
Assume that $M$ is oriented. Given an element $\xi \in S_\gamma M$ we can complete it to an orthonormal frame $(\xi, \eta, \nu)$ of $\mathbb{R}^3$ in the following unique way:

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

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

For a smooth curve of unit tangent vectors $\xi(t) \in S_{\gamma(t)}M$ along a curve $\gamma : [0, T] \to M$ we define $\eta(t), \nu(t) \in T_{\gamma(t)}M$ and we can write
\[
\dot{\xi}(t) = u_\xi(t) \eta(t) + v_\xi(t) \nu(t).
\]
Otherwise one can write the numerator of (1.52) as $\left\langle \nabla_{\gamma(t)}^2 h \dot{\gamma}(t) \mid \dot{\gamma}(t) \right\rangle_h$, where $\nabla_{\gamma(t)}^2 h$ is the hyperbolic Hessian.

38
Definition 1.54. The hyperbolic Levi-Civita connection on $M$ is the 1-form $\omega \in \Lambda^1(SM)$ defined by
\[
\omega_\xi : T_\xi SM \to \mathbb{R}, \quad \omega_\xi(z) = u_z,
\] (1.55)
where $z = u_z \eta + v_z \nu$ and $(\xi, \eta, \nu)$ is the orthonormal frame defined above.

It is again easy to check that a curve of unit tangent vectors $\xi(t)$ is parallel if and only if
\[
\omega_{\dot{\xi}(t)}(\ddot{\gamma}(t)) = 0, \quad \forall t \in [0, T].
\] (1.56)

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

(i) there exist two smooth functions $a_1, a_2 : M \to \mathbb{R}$ such that
\[
\omega = d\theta + a_1(x_1, x_2)dx_1 + a_2(x_1, x_2)dx_2,
\] (1.57)
where $(x_1, x_2, \theta)$ is a system of coordinates on $SM$.

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

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

Definition 1.56. The Gaussian curvature of a surface $M$ in the Minkowski 3-space is the function $\kappa : M \to \mathbb{R}$ defined by the equality
\[
\Omega = -\kappa dV.
\] (1.58)

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

Finally one can define the hyperbolic Gauss map

Definition 1.57. Let $M$ be an oriented surface. We define the Gauss map
\[
\mathcal{N} : M \to H^2, \quad q \mapsto \nu_q,
\] (1.59)
where $\nu_q \in H^2 \subset \mathbb{R}^3$ denotes the external unit normal vector to $M$ at $q$, with respect to the Minkovskky inner product.

Let us now consider the differential of the Gauss map at the point $q$:
\[
D_q\mathcal{N} : T_qM \to T_{\mathcal{N}(q)}H^2 \simeq T_qM
\]
where an element tangent to the hyperbolic plane $H^2$ at $\mathcal{N}(q)$, being orthogonal to $\mathcal{N}(q)$, is identified with a tangent vector to $M$ at $q$.

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

39
1.5 Model spaces of constant curvature

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

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

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

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

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

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

1.5.1 Zero curvature: the Euclidean plane

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

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

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

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

1.5.2 Positive curvature: spheres

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

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

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

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

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

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

It follows immediately by (1.61) that the Gaussian curvature of the sphere is $\kappa = 1/r^2$ at every point $q \in S^2_r$. Let us now recover the structure of geodesics and constant geodesic curvature curves on the sphere.
Proposition 1.60. Let $\gamma : [0, T] \to S^2_r$ be a curve with constant geodesic curvature equal to $c \in \mathbb{R}$. For every vector $w \in \mathbb{R}^3$ the function $\alpha(t) = \langle \dot{\gamma}(t) \mid w \rangle$ is a solution of the differential equation

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

Proof. Without loss of generality, we can assume that $\gamma$ is parametrized by unit speed. Differentiating twice the equality $a(\gamma(t)) = 0$, where $a$ is the function defined in (1.63), we get (in matrix notation):

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

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

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

(1.62)

where $c$ is the geodesic curvature of the curve and $\eta(t) = \dot{\gamma}(t)^\perp$ is the vector orthogonal to $\dot{\gamma}(t)$ in $T_{\gamma(t)}S^2_r$ (defined in such a way that $\dot{\gamma}(t)$ and $\eta(t)$ is a positively oriented frame). Reasoning as in the proof of Proposition 1.59 and noticing that $\nabla_{\gamma(t)}a$ is proportional to the vector $\gamma(t)$, one can compute $b(t)$ and obtains that $\gamma$ satisfies the differential equation

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

(1.63)

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

Proof of Lemma 1.61. The curve $\eta(t)$ has constant norm, hence $\dot{\eta}(t)$ is orthogonal to $\eta(t)$. Recall that the triple $(\dot{\gamma}(t), \ddot{\gamma}(t), \eta(t))$ defines an orthogonal frame at every point. Differentiating the identity $\langle \eta(t) \mid \gamma(t) \rangle = 0$ with respect to $t$ one has

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

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

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

where we used (1.63). Hence $\dot{\eta}(t) = \langle \dot{\eta}(t) \mid \dot{\gamma}(t) \rangle \dot{\gamma}(t) = -c\dot{\gamma}(t)$. \hfill \Box

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

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

(1.64)

Using Lemma 1.61 we have

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

(1.65)

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

(1.66)

which ends the proof of the Proposition 1.60. \hfill \Box
Corollary 1.62. Constant geodesic curvature curves are contained in the intersection of $S^2_r$ with an affine plane of $\mathbb{R}^3$. In particular, geodesics are contained in the intersection of $S^2_r$ with planes passing through the origin, i.e., great circles.

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

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

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

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

Remark 1.63. Curves with constant geodesic curvatures on the spheres are circles obtained as the intersection of the sphere with an affine plane. Moreover all these curves can be also characterized in the following two ways:

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

(ii) boundary of metric balls (spheres).

1.5.3 Negative curvature: the hyperbolic plane

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

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

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

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

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

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

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

Exercise 1.64. Prove that the Gaussian curvature of $H^2_r$ is $\kappa = -1/r^2$ at every point $q \in H^2_r$.

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

42
Proposition 1.65. Let \( \gamma : [0, T] \to H^2_r \) be a curve with constant geodesic curvature equal to \( c \in \mathbb{R} \). For every vector \( w \in \mathbb{R}^3 \) the function \( \alpha(t) = \langle \dot{\gamma}(t) \mid w \rangle_h \) is a solution of the differential equation

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

(1.70)

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

Corollary 1.66. Constant geodesic curvature curves on \( H^2_r \) are contained in the intersection of \( H^2_r \) with affine planes of \( \mathbb{R}^3 \). In particular, geodesics are contained in the intersection of \( H^2_r \) with planes passing through the origin.

Exercise 1.67. Prove Proposition 1.65 and Corollary 1.66.

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

These distinctions reflects in the value of the geodesic curvature. Indeed, as the form of (1.70) also suggest, the value \( c = \frac{1}{r} \) is a threshold and we have the following situation:

(i) if \( 0 \leq c < \frac{1}{r} \), then the curve is an hyperbola,

(ii) if \( c = \frac{1}{r} \), then the curve is a parabola,

(iii) if \( c > \frac{1}{r} \), then the curve is a circle.

This is not the only interesting feature of this classification. Indeed curves of type (i) are equidistant curves while curves of type (iii) are boundary of balls, i.e., spheres, in the hyperbolic plane. Finally, curves of type (ii) are also called horocycles (cf. Remark 1.63 for the difference with respect to the case of the positive constant curvature model).
Chapter 2

Vector fields and vector bundles

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

2.1 Differential equations on smooth manifolds

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

2.1.1 Tangent vectors and vector fields

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

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

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

(2.1)

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

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

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

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

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

45
Definition 2.3. A smooth vector field on a smooth manifold $M$ is a smooth map

$$X : q \mapsto X(q) \in T_q M,$$

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

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

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

$$\dot{q} = X(q), \quad q \in M,$$

(2.2)

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

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

(2.3)

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

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

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

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

(2.4)

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

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

If the maximal solution of (2.4) is defined on a bounded interval $I = (a, b)$, then the solution leaves every compact $K$ of $M$ in a finite time $t_K < b$.

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

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

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

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

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

where $| \cdot |$ denotes the Euclidean norm in $\mathbb{R}^n$. 46
When we are interested in the behavior of the trajectories of a vector field $X \in \text{Vec}(M)$ in a compact subset $K$ of $M$, the assumption of completeness is not restrictive.

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

### 2.1.2 Flow of a vector field

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

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

(2.5)

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

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

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

$$
\phi_0 = \text{Id},
\phi_t \circ \phi_s = \phi_{t+s}, \quad \forall t, s \in \mathbb{R},
(\phi_t)^{-1} = \phi_{-t}, \quad \forall t \in \mathbb{R},
$$

(2.6)

Moreover, by construction, we have

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

(2.7)

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

$$
e^{0X} = \text{Id}, \quad e^{tX} \circ e^{sX} = e^{sX} \circ e^{tX} = e^{(t+s)X}, \quad (e^{tX})^{-1} = e^{-tX},
$$

(2.8)

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

(2.9)

**Remark 2.7.** When $X(x) = Ax$ is a linear vector field on $\mathbb{R}^n$, where $A$ is a $n \times n$ matrix, the corresponding flow $\phi_t(x) = e^{tA}x$.

### 2.1.3 Vector fields as operators on functions

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

$$
X : C^\infty(M) \to C^\infty(M), \quad a \mapsto Xa, \quad a \in C^\infty(M),
$$

(2.10)

where

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

(2.11)

In other words $X$ differentiates the function $a$ along its integral curves.
Remark 2.8. Let us denote $a_t := a \circ e^{tX}$. The map $t \mapsto a_t$ is smooth and from (2.11) it immediately follows that $Xa$ represents the first order term in the expansion of $a_t$ with respect to $t$:

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

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

$$\frac{d}{dt}a_t = Xa, \quad (2.12)$$

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

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

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

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

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

**Exercise 2.11.** Let $f_1, \ldots, f_k \in C^\infty(M)$ and assume that $N = \{f_1 = \ldots = f_k = 0\} \subset M$ is a smooth submanifold. Show that $X \in \text{Vec}(M)$ is tangent to $N$, i.e., $X(q) \in T_q N$ for all $q \in N$, if and only if $X f_i = 0$ for every $i = 1, \ldots, k$.

### 2.1.4 Nonautonomous vector fields

**Definition 2.12.** A nonautonomous vector field is family of vector fields $\{X_t\}_{t \in \mathbb{R}}$ such that the map $X(t,q) = X(t,q)$ satisfies the following properties

(C1) $X(\cdot,q)$ is measurable for every fixed $q \in M$,

(C2) $X(t,\cdot)$ is smooth for every fixed $t \in \mathbb{R}$,

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

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

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

**Remark 2.13.** In these lecture notes we are mainly interested in nonautonomous vector fields of the following form

$$X_i(q) = \sum_{i=1}^m u_i(t) f_i(q) \quad (2.15)$$

48
where $u_i$ are $L^\infty$ functions and $f_i$ are smooth vector fields on $M$. For this class of nonautonomous vector fields assumptions (C1)-(C2) are trivially satisfied. For what concerns (C3), by the smoothness of $f_i$ for every compact set $K \subset \Omega$ we can find two positive constants $C_K, L_K$ such that for all $i = 1, \ldots, m$ and $j = 1, \ldots, n$ we have

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

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

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

The existence and uniqueness of integral curves of a nonautonomous vector field is guaranteed by the following theorem (see [9]).

**Theorem 2.14 (Carathéodory theorem).** Assume that the nonautonomous vector field $\{X_t\}_{t \in \mathbb{R}}$ satisfies (C1)-(C3). Then the Cauchy problem

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

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

Let us assume now that the equation $(2.14)$ is complete, i.e., for all $t_0 \in \mathbb{R}$ and $q_0 \in M$ the solution $\gamma(t;t_0,q_0)$ is defined on $I = \mathbb{R}$. Let us denote $P_{t_0,t}(q) = \gamma(t;t_0,q)$. The family of maps $P_{t_0,t} : M \to M$ is the (nonautonomous) flow generated by $X_t$. It satisfies

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

Moreover the following algebraic identities are satisfied

$$P_{t,t} = \text{Id},$$

$$P_{t_2,t_3} \circ P_{t_1,t_2} = P_{t_1,t_3}, \quad \forall t_1, t_2, t_3 \in \mathbb{R}, \quad (2.18)$$

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

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

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

The following lemma characterizes flows whose infinitesimal generator is autonomous.

**Lemma 2.15.** Let $\{P_{t,s}\}_{t,s \in \mathbb{R}}$ be a family of smooth diffeomorphisms satisfying $(2.18)$. Its infinitesimal generator is an autonomous vector field if and only if

$$P_{0,t} \circ P_{0,s} = P_{0,t+s}, \quad \forall t, s \in \mathbb{R}.$$
2.2 Differential of a smooth map

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

**Definition 2.16.** Let $\varphi : M \to N$ be a smooth map between smooth manifolds and $q \in M$. The **differential** of $\varphi$ at the point $q$ is the linear map
\[
\varphi_*,q : T_q M \to T_{\varphi(q)} N, \tag{2.20}
\]
defined as follows:
\[
\varphi_*,q(v) = \frac{d}{dt} \bigg|_{t=0} \varphi(\gamma(t)), \quad \text{if} \quad v = \frac{d}{dt} \bigg|_{t=0} \gamma(t), \quad q = \gamma(0).
\]
It is easily checked that this definition depends only on the equivalence class of $\gamma$.

![Figure 2.1: Differential of a map $\varphi : M \to N$](image)

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

**Exercise 2.17.** Let $\varphi : M \to N$, $\psi : N \to Q$ be smooth maps between manifolds. Prove that the differential of the composition $\psi \circ \varphi : M \to Q$ satisfies $(\psi \circ \varphi)_* = \psi_* \circ \varphi_*$.

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

**Definition 2.18.** Let $X \in \text{Vec}(M)$ and $\varphi : M \to N$ be a diffeomorphism. The **pushforward** $\varphi_* X \in \text{Vec}(N)$ is the vector field on $N$ defined by
\[
(\varphi_* X)(\varphi(q)) := \varphi_*(X(q)), \quad \forall q \in M. \tag{2.21}
\]
When $P \in \text{Diff}(M)$ is a diffeomorphism on $M$, we can rewrite the identity (2.21) as
\[
(P_* X)(q) = P_*(X(P^{-1}(q))), \quad \forall q \in M. \tag{2.22}
\]
Notice that, in general, if $\varphi$ is a smooth map, the pushforward of a vector field is not defined.

**Remark 2.19.** From this definition it follows the useful formula for $X, Y \in \text{Vec}(M)$
\[
(e^{tX} Y)|_q = e^{tX} (Y|_{e^{-tX}(q)}) = \frac{d}{ds} \bigg|_{s=0} e^{tX} \circ e^{sY} \circ e^{-tX}(q).
\]
If \( P \in \text{Diff}(M) \) and \( X \in \text{Vec}(M) \), then \( P \ast X \) is, by construction, the vector field whose integral curves are the image under \( P \) of integral curves of \( X \). The following lemma shows how it acts as operator on functions.

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

\[
\begin{align*}
e^{tP \ast X} &= P \circ e^{tX} \circ P^{-1}, \\
(P \ast X)a &= (X(a \circ P)) \circ P^{-1}.
\end{align*}
\]

**Proof.** From the formula

\[
\frac{d}{dt}\Bigg|_{t=0} P \circ e^{tX} \circ P^{-1}(q) = P(X(P^{-1}(q))) = (P \ast X)(q),
\]

it follows that \( t \mapsto P \circ e^{tX} \circ P^{-1}(q) \) is an integral curve of \( P \ast X \), from which (2.23) follows. To prove (2.24) let us compute

\[
(P \ast X)a := \frac{d}{dt}\bigg|_{t=0} a(e^{tP \ast X}(q)).
\]

Using (2.23) this is equal to

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

As a consequence of Lemma 2.20 one gets the following formula: for every \( X,Y \in \text{Vec}(M) \)

\[
(e^{tX}Y)a = Y(a \circ e^{tX}) \circ e^{-tX}.
\]

2.3 Lie brackets

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

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

\[
[X,Y] := \frac{\partial}{\partial t}\bigg|_{t=0} e^{-tX}Y.
\]

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

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

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

\[
[X,Y] = XY - YX.
\]
Proof. By definition of Lie bracket we have \([X, Y]a = \frac{\partial}{\partial t} |_{t=0} (e^{-tX}Y)a\). Hence we have to compute the first order term in the expansion, with respect to \(t\), of the map
\[
t \mapsto (e^{-tX}Y)a.
\]
Using formula (2.25) we have
\[
(e^{-tX}Y)a = Y(a \circ e^{-tX}) \circ e^{tX}.
\]
By Remark 2.8 we have \(a \circ e^{-tX} = a - tXa + O(t^2)\), hence
\[
(e^{-tX}Y)a = (Ya - tYXa + O(t^2)) \circ e^{tX} = (Ya - tYXa + O(t^2)) \circ e^{tX}
\]
Denoting \(b = Ya - tYXa + O(t^2)\), \(b_t = b \circ e^{tX}\), and using again the expansion above we get
\[
(e^{-tX}Y)a = (Ya - tYXa + O(t^2)) + tX(Ya - tYXa + O(t^2)) + O(t^2)
\]
that proves that the first order term with respect to \(t\) in the expansion is \((XY - YX)a\). \(\square\)

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

Exercise 2.24. Prove the coordinate expression of the Lie bracket: let
\[
X = \sum_{i=1}^{n} X_i \frac{\partial}{\partial x_i}, \quad Y = \sum_{j=1}^{n} Y_j \frac{\partial}{\partial x_j},
\]
be two vector fields in \(\mathbb{R}^n\). Show that
\[
[X, Y] = \sum_{i,j=1}^{n} \left( X_i \frac{\partial Y_j}{\partial x_i} - Y_j \frac{\partial X_i}{\partial x_i} \right) \frac{\partial}{\partial x_j}.
\]

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

Proposition 2.25. Let \(P \in \text{Diff}(M)\). Then \(P_*\) is a Lie algebra homomorphism of \(\text{Vec}(M)\), i.e.,
\[
P_*[X, Y] = [P_*X, P_*Y], \quad \forall X, Y \in \text{Vec}(M).
\]

Proof. We show that the two terms are equal as derivations on functions. Let \(a \in C^\infty(M)\), preliminarily we see, using (2.24), that
\[
P_*X(P_*Ya) = P_*X(Y(a \circ P) \circ P^{-1})
\]
\[
= X(Y(a \circ P) \circ P^{-1} \circ P) \circ P^{-1}
\]
\[
= X(Y(a \circ P)) \circ P^{-1},
\]
and using twice this property and (2.28)
\[
[P_*X, P_*Y]a = P_*X(P_*Ya) - P_*Y(P_*Xa)
\]
\[
= XY(a \circ P) \circ P^{-1} - YX(a \circ P) \circ P^{-1}
\]
\[
= (XY - YX)(a \circ P) \circ P^{-1}
\]
\[
= P_*[X, Y]a.
\]
\(\square\)
To end this section, we show that the Lie bracket of two vector fields is zero (i.e., they commute as operator on functions) if and only if their flows commute.

**Proposition 2.26.** Let $X, Y \in \text{Vec}(M)$. The following properties are equivalent:

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

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

**Proof.** We start the proof with the following claim

\[
[X, Y] = 0 \implies e^{-tX} \circ Y = Y, \quad \forall t \in \mathbb{R}.
\]  

(2.29)

To prove (2.29) let us show that $[X, Y] = 0$ implies that $\frac{d}{dt} e^{-tX} Y = 0$ for all $t \in \mathbb{R}$.

Indeed we have

\[
\frac{d}{dt} e^{-tX} Y = \frac{d}{ds} \bigg|_{s=0} e^{-tX} \circ e^{(s+\varepsilon)Y} = \frac{d}{ds} \bigg|_{s=0} e^{-tX} e^{-\varepsilon Y} Y
\]

which proves (2.29).

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

\[
\frac{\partial}{\partial s} \phi_s = \frac{\partial}{\partial s} \bigg|_{s=0} e^{-tX} \circ e^{(s+\varepsilon)Y} \circ e^{tX}
\]

\[
= \frac{\partial}{\partial \varepsilon} \bigg|_{\varepsilon=0} e^{-tX} \circ e^{sY} \circ e^{tX} \circ e^{-tX} \circ e^{sY} \circ e^{tX}
\]

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

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

which is equivalent to (ii).

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

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

Then (i) follows from (2.28). \qed

**Exercise 2.27.** Let $X, Y \in \text{Vec}(M)$ and $q \in M$. Consider the curve on $M$

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

Prove that the tangent vector to the curve $t \mapsto \gamma(\sqrt{t})$ at $t = 0$ is $[X, Y](q)$. 53
Exercise 2.28. Let $X, Y \in \text{Vec}(M)$. Using the semigroup property of the flow, prove the following expansion
\[
e^{-tX}Y = \sum_{n=0}^{\infty} \frac{t^n}{n!} (\text{ad } X)^n Y
= Y + t[X, Y] + \frac{t^2}{2} [X, [X, Y]] + \frac{t^3}{6} [X, [X, [X, Y]]] + \ldots
\]

Exercise 2.29. Let $X, Y \in \text{Vec}(M)$ and $a \in C^\infty(M)$. Prove the following Leibnitz rule for the Lie bracket:
\[
[X, aY] = a[X, Y] + (Xa)Y.
\]

Exercise 2.30. Let $X, Y, Z \in \text{Vec}(M)$. Prove that the Lie bracket satisfies the Jacobi identity:
\[
[X, [Y, Z]] + [Y, [Z, X]] + [Z, [X, Y]] = 0.
\] (2.30)

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

2.4 Cotangent space

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

Definition 2.31. Let $M$ be a $n$-dimensional smooth manifold. The cotangent space at a point $q \in M$ is the set
\[
T^*_q M := (T_q M)^* = \{ \lambda : T_q M \to \mathbb{R}, \lambda \text{ linear} \}.
\]

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

As we have seen, a smooth map yields a linear map between tangent spaces. Dualizing this map, we get a linear map on cotangent spaces.

Definition 2.32. Let $\varphi : M \to N$ be a smooth map and $q \in M$. The pullback of $\varphi$ at point $\varphi(q)$, where $q \in M$, is the map
\[
\varphi^* : T^*_{\varphi(q)} N \to T^*_q M, \quad \lambda \mapsto \varphi^* \lambda,
\]
defined by duality in the following way
\[
\langle \varphi^* \lambda, v \rangle := \langle \lambda, \varphi_\ast v \rangle, \quad \forall v \in T_q M, \forall \lambda \in T^*_{\varphi(q)} M.
\]

Example 2.33. Let $a : M \to \mathbb{R}$ be a smooth function and $q \in M$. The differential $d_q a$ of the function $a$ at the point $q \in M$, defined through the formula
\[
\langle d_q a, v \rangle := \left. \frac{d}{dt} \right|_{t=0} a(\gamma(t)), \quad v \in T_q M,
\] (2.31)

where $\gamma$ is any smooth curve such that $\gamma(0) = q$ and $\dot{\gamma}(0) = v$, is an element of $T^*_q M$, since (2.31) is linear with respect to $v$. 

54
**Definition 2.34.** A **differential 1-form** on a smooth manifold \( M \) is a smooth map

\[
\omega : q \mapsto \omega(q) \in T^*_q M,
\]
that associates to every point \( q \) in \( M \) a cotangent vector at \( q \). We denote by \( \Lambda^1(M) \) the set of differential forms on \( M \).

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

\[
\langle \omega, X \rangle : q \mapsto \langle \omega(q), X(q) \rangle.
\] (2.32)

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

**Definition 2.35.** Let \( \varphi : M \to N \) be a smooth map and \( a : N \to \mathbb{R} \) be a smooth function. The **pullback** \( \varphi^* a \) is the smooth function on \( M \) defined by

\[
(\varphi^* a)(q) = a(\varphi(q)), \quad q \in M.
\]

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

\[
(\pi^* a)(\lambda) = a(\pi(\lambda)), \quad \lambda \in T^* M,
\]

which is constant on fibers.

### 2.5 Vector bundles

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

**Definition 2.36.** Let \( M \) be a \( n \)-dimensional manifold. A **smooth vector bundle** of rank \( k \) over \( M \) is a smooth manifold \( E \) with a surjective smooth map \( \pi : E \to M \) such that

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

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

\[
\begin{array}{ccc}
\pi^{-1}(O_q) & \xrightarrow{\psi} & O_q \times \mathbb{R}^k \\
\downarrow \pi & & \downarrow \pi_1 \\
O_q & & 
\end{array}
\]

(2.33)

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

**Remark 2.37.** A vector bundle \( E \), as a smooth manifold, has dimension

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

In the case when there exists a global trivialization map, i.e. one can choose a local trivialization with \( O_q = M \) for all \( q \in M \), then \( E \) is diffeomorphic to \( M \times \mathbb{R}^k \) and we say that \( E \) is **trivializable**.
Example 2.38. For any smooth $n$-dimensional manifold $M$, the tangent bundle $TM$, defined as the disjoint union of the tangent spaces at all points of $M$,

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

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

$$\pi : TM \rightarrow M, \quad \pi(v) = q \quad\text{if} \quad v \in T_qM.$$

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

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

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

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

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

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

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

a local coordinate system. The differentials of the coordinate functions

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

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

$$\frac{\partial}{\partial x_i}_q \in T_qM, \quad i = 1, \ldots, n, \quad q \in O, \quad (2.34)$$

$$\left\langle dx_i, \frac{\partial}{\partial x_j}_q \right\rangle = \delta_{ij}, \quad i, j = 1, \ldots, n. \quad (2.35)$$

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

$$v = \sum_{i=1}^n v_i \frac{\partial}{\partial x_i}_q, \quad \lambda = \sum_{i=1}^n p_i dx_i|_q,$$

and the maps

$$\psi : v \mapsto (x_1, \ldots, x_n, v_1, \ldots, v_n), \quad \tilde{\psi} : \lambda \mapsto (x_1, \ldots, x_n, p_1, \ldots, p_n), \quad (2.36)$$

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

56
**Definition 2.39.** A morphism \( f : E \to E' \) between two vector bundles \( E, E' \) on the base \( M \) (also called a bundle map) is a smooth map such that the following diagram is commutative

\[
\begin{array}{ccc}
E & \xrightarrow{f} & E' \\
\downarrow{\pi} & & \downarrow{\pi'} \\
M & & M
\end{array}
\]

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

**Definition 2.40.** Let \( \pi : E \to M \) be a smooth vector bundle over \( M \). A local section of \( E \) is a smooth map \( \sigma : A \subset M \to E \) satisfying \( \pi \circ \sigma = \text{Id}_A \), where \( A \) is an open set of \( M \). In other words \( \sigma(q) \) belongs to \( E_q \) for each \( q \in A \), smoothly with respect to \( q \). If \( \sigma \) is defined on all \( M \) it is said to be a global section.

**Example 2.41.** Let \( \pi : E \to M \) be a smooth vector bundle over \( M \). The zero section of \( E \) is the global section \( \zeta : M \to E, \quad \zeta(q) = 0 \in E_q, \quad \forall q \in M. \)

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

**Remark 2.42.** Notice that smooth vector fields and smooth differential forms are, by definition, sections of the vector bundles \( TM \) and \( T^*M \) respectively.

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

**Definition 2.43.** Let \( \varphi : M \to N \) be a smooth map between smooth manifolds and \( E \) be a vector bundle on \( N \), with fibers \( \{ E_{q'}, q' \in N \} \). The induced bundle (or pullback bundle) \( \varphi^*E \) is a vector bundle on the base \( M \) defined by

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

Notice that rank \( \varphi^*E = \text{rank } E \), hence dim \( \varphi^*E = \text{dim } M + \text{rank } E \).

**Example 2.44.** (i). Let \( M \) be a smooth manifold and \( TM \) its tangent bundle, endowed with an Euclidean structure. The spherical bundle \( SM \) is the vector subbundle of \( TM \) defined as follows

\[
SM = \bigcup_{q \in M} S_q M, \quad S_q M = \{v \in T_q M \mid |v| = 1\}.
\]

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

\[
(E \oplus E')_q := E_q \oplus E'_q.
\]

\(^1\)Hetre smooth means as a map between manifolds.
2.6 Submersions and level sets of smooth maps

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

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

**Theorem 2.45 (Rank Theorem).** Suppose \( M \) and \( N \) are smooth manifolds of dimensions \( m \) and \( n \), respectively, and \( \varphi : M \to N \) is a smooth map with constant rank \( k \) in a neighborhood of \( q \in M \). Then there exist coordinates \( (x_1, \ldots, x_m) \) centered at \( q \) and \( (y_1, \ldots, y_n) \) centered at \( \varphi(q) \) in which \( \varphi \) has the following coordinate representation:

\[
\varphi(x_1, \ldots, x_m) = (x_1, \ldots, x_k, 0, \ldots, 0).
\]

**Remark 2.46.** The previous theorem can be rephrased in the following way. Let \( \varphi : M \to N \) be a smooth map between two smooth manifolds. Then the following are equivalent:

(i) \( \varphi \) has constant rank in a neighborhood of \( q \in M \).

(ii) There exist coordinates near \( q \in M \) and \( \varphi(q) \in N \) in which the coordinate representation of \( \varphi \) is linear.

In the case of a submersion, from Theorem 2.45 one can deduce the following result.

**Corollary 2.47.** Assume \( \varphi : M \to N \) is a smooth submersion at \( q \). Then \( \varphi \) admits a local right inverse at \( \varphi(q) \). Moreover \( \varphi \) is open at \( q \). More precisely it exist \( \varepsilon > 0 \) and \( C > 0 \) such that

\[
B_{\varphi(q)}(C^{-1}r) \subset \varphi(B_q(r)), \quad \forall r \in [0, \varepsilon].
\]

**Remark 2.48.** The constant \( C \) appearing in (2.39) is the norm of the differential of the local right inverse. When \( \varphi \) is a diffeomorphism, \( C \) is a bound on the norm of the differential of the inverse of \( \varphi \). This recover the classical quantitative statement of the inverse function theorem.

Using these results, one can give some very general criteria for level sets of smooth maps (or smooth functions) to be submanifolds.

**Theorem 2.49 (Constant Rank Level Set Theorem).** Let \( M \) and \( N \) be smooth manifolds, and let \( \varphi : M \to N \) be a smooth map with constant rank \( k \). Each level set \( \varphi^{-1}(y) \), for \( y \in N \) is a closed embedded submanifold of codimension \( k \) in \( M \).

**Remark 2.50.** It is worth to specify the following two important sub cases of Theorem 2.49

(a) If \( \varphi : M \to N \) is a submersion at every \( q \in \varphi^{-1}(y) \) for some \( y \in N \), then \( \varphi^{-1}(y) \) is a closed embedded submanifold whose codimension is equal to the dimension of \( N \).

(b) If \( a : M \to \mathbb{R} \) is a smooth function such that \( d_q a \neq 0 \) for every \( q \in a^{-1}(c) \), where \( c \in \mathbb{R} \), then the level set \( a^{-1}(c) \) is a smooth hypersurface of \( M \)

**Exercise 2.51.** Let \( a : M \to \mathbb{R} \) be a smooth function. Assume that \( c \in \mathbb{R} \) is a regular value of \( a \), i.e., \( d_q a \neq 0 \) for every \( q \in a^{-1}(c) \). Then \( N_c = a^{-1}(c) = \{ q \in M | a(q) = c \} \subset M \) is a smooth submanifold. Prove that for every \( q \in N_c \)

\[
T_q N_c = \ker d_q a = \{ v \in T_q M \mid \langle d_q a, v \rangle = 0 \}.
\]
Bibliographical notes

The material presented in this chapter is classical and covered by many textbook in differential geometry, as for instance in [6, 24, 14, 32].

Theorem 2.14 is a well-known theorem in ODE. The statement presented here can be deduced from [10, Theorem 2.1.1, Exercice 2.4]. The functions \( c(t), k(t) \) appearing in (C3) are assumed to be \( L^\infty \), that is stronger than \( L^1 \) (on compact intervals). This stronger assumptions implies that the solution is not only absolutely continuous with respect to \( t \), but also locally Lipschitz.
Chapter 3

Sub-Riemannian structures

3.1 Basic definitions

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

**Definition 3.1.** Let $M$ be a smooth manifold and let $\mathcal{F} \subset \text{Vec}(M)$ be a family of smooth vector fields. The **Lie algebra generated** by $\mathcal{F}$ is the smallest sub-algebra of $\text{Vec}(M)$ containing $\mathcal{F}$, namely

\[
\text{Lie}\mathcal{F} := \text{span}\{[X_1, \ldots, [X_{j-1}, X_j]], X_i \in \mathcal{F}, j \in \mathbb{N}\}.
\]  

(3.1)

We will say that $\mathcal{F}$ is **bracket-generating** (or that satisfies the Hörmander condition) if

\[
\text{Lie}_q\mathcal{F} := \{X(q), X \in \text{Lie}\mathcal{F}\} = T_qM, \quad \forall q \in M.
\]

Moreover, for $s \in \mathbb{N}$, we define

\[
\text{Lie}^s\mathcal{F} := \text{span}\{[X_1, \ldots, [X_{j-1}, X_j]], X_i \in \mathcal{F}, j \leq s\}.
\]  

(3.2)

We say that the family $\mathcal{F}$ is of step $s$ at $q$ if $m$ is the minimal integer satisfying

\[
\text{Lie}^m_q\mathcal{F} := \{X(q), X \in \text{Lie}^m\mathcal{F}\} = T_qM,
\]

Notice that, in general, the step may depend on the point on $M$ and $s = s(q)$ can be unbounded on $M$ even for bracket-generating structure.

**Definition 3.2.** Let $M$ be a connected smooth manifold. A **sub-Riemannian structure** on $M$ is a pair $(U, f)$ where:

(i) $U$ is an Euclidean bundle with base $M$ and Euclidean fiber $U_q$, i.e., for every $q \in M$, $U_q$ is a vector space equipped with a scalar product $(\cdot | \cdot)_q$, smooth with respect to $q$. For $u \in U_q$ we denote the norm of $u$ as $|u|^2 = (u | u)_q$.

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

\[
\begin{array}{ccc}
U & \xrightarrow{f} & TM \\
\downarrow{\pi_U} & & \downarrow{\pi} \\
M & & 
\end{array}
\]  

(3.3)
and $f$ is linear on fibers.

(iii) The set of horizontal vector fields $\mathcal{D} := \{f(\sigma) \mid \sigma : M \to U \text{ smooth section} \}$, is a bracket-generating family of vector fields. We call step of the sub-Riemannian structure at $q$ the step of $\mathcal{D}$.

When the vector bundle $U$ admits a global trivialization we say that $(U, f)$ is a free sub-Riemannian structure.

A smooth manifold endowed with a sub-Riemannian structure (i.e., the triple $(M, U, f)$) is called a sub-Riemannian manifold. When the map $f : U \to TM$ is fiberwise surjective, $(M, U, f)$ is called a Riemannian manifold (cf. Exercise 3.23).

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

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

We call $k(q) := \dim \mathcal{D}_q$ the rank of the sub-Riemannian structure at $q \in M$. We say that the sub-Riemannian structure $(U, f)$ on $M$ has constant rank if $k(q)$ is constant. Otherwise we say that the sub-Riemannian structure is rank-varying.

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

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

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

$$k(q) \leq m, \quad \text{where} \quad m = \text{rank } U, \quad (3.4)$$  

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

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

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

We prefer the second notation to stress that, for each $q \in M$, $f_u(q)$ is a vector in $T_q M$.

**Definition 3.4.** A Lipschitz curve $\gamma : [0, T] \to M$ is said to be admissible (or horizontal) for a sub-Riemannian structure if there exists a measurable and essentially bounded function

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

called the control function, such that

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

In this case we say that $u(\cdot)$ is a control corresponding to $\gamma$. Notice that different controls could correspond to the same trajectory.
Remark 3.5. Once we have chosen a local trivialization \( O_q \times \mathbb{R}^m \) for the vector bundle \( U \), where \( O_q \) is a neighborhood of a point \( q \in M \), we can choose a basis in the fibers and the map \( f \) is written \( f(q,u) = \sum_{i=1}^{m} u_i f_i(q) \), where \( m \) is the rank of \( U \). In this trivialization, a Lipschitz curve \( \gamma : [0,T] \to M \) is admissible if there exists \( u = (u_1, \ldots, u_m) \in L^\infty([0,T], \mathbb{R}^m) \) such that

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

(3.8)

Thanks to this local characterization and Theorem 2.14 for each initial condition \( q \in M \) and \( u \in L^\infty([0,T], \mathbb{R}^m) \) there exists an admissible curve \( \gamma \), defined on a sufficiently small interval, such that \( u \) is the control associated with \( \gamma \) and \( \gamma(0) = q \).

Remark 3.6. Notice that, for a curve to be admissible, it is not sufficient to satisfy \( \dot{\gamma}(t) \in \mathcal{D}_{\gamma(t)} \) for almost every \( t \in [0,T] \). Take for instance the two free sub-Riemannian structures on \( \mathbb{R}^2 \) having rank two and defined by

\[
f(x,y,u_1,u_2) = (x,y,u_1,u_2x), \quad f'(x,y,u_1,u_2) = (x,y,u_1,u_2x^2).\]

(3.9)

and let \( \mathcal{D} \) and \( \mathcal{D}' \) the corresponding moduli of horizontal vector fields. It is easily seen that the curve \( \gamma : [-1,1] \to \mathbb{R}^2 \), \( \gamma(t) = (t,t^2) \) satisfies \( \dot{\gamma}(t) \in \mathcal{D}_{\gamma(t)} \) and \( \dot{\gamma}(t) \in \mathcal{D}'_{\gamma(t)} \) for every \( t \in [-1,1] \).

Moreover, \( \gamma \) is admissible for \( f \), since its corresponding control is \((u_1,u_2) = (1,2)\) for a.e. \( t \in [-1,1] \), but it is not admissible for \( f' \), since its corresponding control is uniquely determined as \((u_1(t),u_2(t)) = (1,2/t)\) for a.e. \( t \in [-1,1] \), which is not essentially bounded.

This example shows that, for two different sub-Riemannian structures \((U,f)\) and \((U',f')\) on the same manifold \( M \), one can have \( \mathcal{D}_q = \mathcal{D}'_q \) for every \( q \in M \), but \( \mathcal{D} \neq \mathcal{D}' \). Notice, however, that if the distribution has constant rank one has \( \mathcal{D}_q = \mathcal{D}'_q \) for every \( q \in M \) if and only if \( \mathcal{D} = \mathcal{D}' \).

### 3.1.1 The minimal control and the length of an admissible curve

We start by defining the sub-Riemannian norm for vectors that belong to the distribution.

**Definition 3.7.** Let \( v \in \mathcal{D}_q \). We define the sub-Riemannian norm of \( v \) as follows

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

(3.10)
Notice that since $f$ is linear with respect to $u$, the minimum in (3.10) is always attained at a unique point. Indeed the condition $f(q, \cdot) = v$ defines an affine subspace of $U_q$ (which is nonempty since $v \in \mathcal{D}_q$) and the minimum in (3.10) is uniquely attained at the orthogonal projection of the origin onto this subspace (see Figure 3.2).

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

**Exercise 3.8.** Show that $\| \cdot \|$ is a norm in $\mathcal{D}_q$. Moreover prove that it satisfies the parallelogram law, i.e., it is induced by a scalar product $\langle \cdot | \cdot \rangle_q$ on $\mathcal{D}_q$, that can be recovered by the polarization identity

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

**Exercise 3.9.** Let $u_1, \ldots, u_m \in U_q$ be an orthonormal basis for $U_q$. Define $v_i = f(q, u_i)$. Show that if $f(q, \cdot)$ is injective then $v_1, \ldots, v_m$ is an orthonormal basis for $\mathcal{D}_q$.

An admissible curve $\gamma : [0, T] \to M$ is Lipschitz, hence differentiable at almost every point. Hence it is well defined the unique control $t \mapsto u^*(t)$ associated with $\gamma$ and realizing the minimum in (3.10).

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

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

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

We stress that $u^*(t)$ is pointwise defined for a.e. $t \in [0, T]$. The proof of the following crucial Lemma is postponed to the Section 3.A.

**Lemma 3.11.** Let $\gamma : [0, T] \to M$ be an admissible curve. Then its minimal control $u^*(\cdot)$ is measurable and essentially bounded on $[0, T]$. 
Remark 3.12. If the admissible curve \( \gamma : [0, T] \to M \) is differentiable, its minimal control is defined everywhere on \([0, T]\). Nevertheless, it could be not continuous, in general.

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

\[
f(x, y, u_1, u_2) = (x, y, u_1, u_2x),
\]

and let \( \gamma : [-1, 1] \to \mathbb{R}^2 \) defined by \( \gamma(t) = (t, t^2) \). Its minimal control \( u^*(t) \) satisfies \((u_1^*(t), u_2^*(t)) = (1, 2) \) when \( t \neq 0 \), while \((u_1^*(0), u_2^*(0)) = (1, 0) \), hence is not continuous.

Thanks to Lemma 3.11 we are allowed to introduce the following definition.

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

\[
\ell(\gamma) := \int_0^T \| \dot{\gamma}(t) \| dt.
\]

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

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

\[
\ell(\gamma) = \int_0^T |u^*(t)| dt.
\]

In particular any admissible curve has finite length.

**Lemma 3.14.** The length of an admissible curve is invariant by Lipschitz reparametrization.

**Proof.** Let \( \gamma : [0, T] \to M \) be an admissible curve and \( \varphi : [0, T'] \to [0, T] \) a Lipschitz reparametrization, i.e., a Lipschitz and monotone surjective map. Consider the reparametrized curve

\[
\gamma_\varphi : [0, T'] \to M, \quad \gamma_\varphi := \gamma \circ \varphi.
\]

First observe that \( \gamma_\varphi \) is a composition of Lipschitz functions, hence Lipschitz. Moreover \( \gamma_\varphi \) is admissible since, by the linearity of \( f \), it has minimal control \((u^* \circ \varphi) \dot{\varphi} \in L^\infty\), where \( u^* \) is the minimal control of \( \gamma \). Using the change of variables \( t = \varphi(s) \), one gets

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

\[\square\]

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

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

\[
\varphi(t) := \int_0^t |u^*(\tau)| d\tau.
\]
Notice that if $\varphi(t_1) = \varphi(t_2)$, the monotonicity of $\varphi$ ensures $\psi(t_1) = \psi(t_2)$. Hence we are allowed to define $\gamma: [0, \ell(\psi)] \to M$ by

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

In other words, it holds $\psi = \gamma \circ \varphi$. To show that $\gamma$ is Lipschitz let us first show that there exists a constant $C > 0$ such that, for every $t_0, t_1 \in [0, T]$ one has, in some local coordinates (where $| \cdot |$ denotes the Euclidean norm in coordinates)

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

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

$$|\psi(t_1) - \psi(t_0)| \leq \int_{t_0}^{t_1} \sum_{i=1}^{m} |u^*_i(t) f_i(\psi(t))| dt$$

$$\leq \int_{t_0}^{t_1} \sqrt{\sum_{i=1}^{m} |u^*_i(t)|^2} \sqrt{\sum_{i=1}^{m} |f_i(\psi(t))|^2} dt$$

$$\leq C \int_{t_0}^{t_1} |u^*(t)| dt,$$

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

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

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

We are going to prove that $\gamma$ is admissible and its minimal control has norm one. Define for every $s$ such that $s = \varphi(t)$, $\dot{\varphi}(t)$ exists and $\dot{\varphi}(t) \neq 0$, the control

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

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

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

By the previous discussion, in what follows, it will be often convenient to assume that admissible curves are length-parametrized (or parametrized such that $\|\dot{\gamma}(t)\|$ is constant).
3.1.2 Equivalence of sub-Riemannian structures

In this section we introduce the notion of equivalence for sub-Riemannian structures on the same base manifold $M$ and the notion of isometry between sub-Riemannian manifolds.

**Definition 3.17.** Let $(U, f), (U', f')$ be two sub-Riemannian structures on a smooth manifold $M$. They are said to be equivalent if the following conditions are satisfied

(i) there exist an Euclidean bundle $V$ and two surjective vector bundle morphisms $p : V \to U$ and $p' : V \to U'$ such that the following diagram is commutative

\[
\begin{array}{ccc}
U & \xrightarrow{f} & TM \\
\downarrow{p} & & \downarrow{f'} \\
V & \xrightarrow{p'} & U'
\end{array}
\]

(ii) the projections $p, p'$ are compatible with the scalar product, i.e., it holds

\[
|u| = \min\{|v|, p(v) = u\}, \quad \forall u \in U,
\]

\[
|u'| = \min\{|v|, p'(v) = u'\}, \quad \forall u' \in U',
\]

**Remark 3.18.** If $(U, f)$ and $(U', f')$ are equivalent sub-Riemannian structures on $M$, then:

(a) the distributions $\mathcal{D}_q$ and $\mathcal{D}'_q$ defined by $f$ and $f'$ coincide, since $f(U_q) = f'(U'_q)$ for all $q \in M$.

(b) for each $w \in \mathcal{D}_q$ we have $\|w\| = \|w\|'$, where $\| \cdot \|$ and $\| \cdot \|'$ are the norms induced by $(U, f)$ and $(U', f')$ respectively.

In particular the length of an admissible curve for two equivalent sub-Riemannian structures is the same.

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

**Definition 3.20.** Let $M$ be a sub-Riemannian manifold. We define the minimal bundle rank of $M$ as the infimum of rank of bundles that induce equivalent structures on $M$. Given $q \in M$ the local minimal bundle rank of $M$ at $q$ is the minimal bundle rank of the structure restricted on a sufficiently small neighborhood $O_q$ of $q$.

**Exercise 3.21.** Prove that the free sub-Riemannian structure on $\mathbb{R}^2$ defined by $f : \mathbb{R}^2 \times \mathbb{R}^3 \to T\mathbb{R}^2$ defined by

\[
f(x, y, u_1, u_2, u_3) = (x, y, u_1, u_2x + u_3y)
\]

has non constant local minimal bundle rank.

For equivalence classes of sub-Riemannian structures we introduce the following definition.
Definition 3.22. Two equivalent classes of sub-Riemannian manifolds are said to be isometric if there exist two representatives \((M, U, f), (M', U', f')\), a diffeomorphism \(\phi : M \to M'\) and an isomorphism of Euclidean bundles \(\psi : U \to U'\) such that the following diagram is commutative

\[
\begin{array}{ccc}
U & \xrightarrow{f} & TM \\
\downarrow{\psi} & & \downarrow{\phi_*} \\
U' & \xrightarrow{f'} & TM'
\end{array}
\] (3.18)

3.1.3 Examples

Our definition of sub-Riemannian manifold is quite general. In the following we list some classical geometric structures which are included in our setting.

1. Riemannian structures.
   Classically a Riemannian manifold is defined as a pair \((M, \langle \cdot | \cdot \rangle)\), where \(M\) is a smooth manifold and \(\langle \cdot | \cdot \rangle_q\) is a family of scalar product on \(T_q M\), smoothly depending on \(q \in M\).
   This definition is included in Definition 3.2 by taking \(U = TM\) endowed with the Euclidean structure induced by \(\langle \cdot | \cdot \rangle\) and \(f : TM \to TM\) the identity map.

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

2. Constant rank sub-Riemannian structures.
   Classically a constant rank sub-Riemannian manifold is a triple \((M, D, \langle \cdot | \cdot \rangle)\), where \(D\) is a vector subbundle of \(TM\) and \(\langle \cdot | \cdot \rangle_q\) is a family of scalar product on \(D_q\), smoothly depending on \(q \in M\).
   This definition is included in Definition 3.2 by taking \(U = D\), endowed with its Euclidean structure, and \(f : D \to TM\) the canonical inclusion.

3. Almost-Riemannian structures.
   An almost-Riemannian structure on \(M\) is a sub-Riemannian structure \((U, f)\) on \(M\) such that its local minimal bundle rank is equal to the dimension of the manifold, at every point.

4. Free sub-Riemannian structures.
   Let \(U = M \times \mathbb{R}^m\) be the trivial Euclidean bundle of rank \(m\) on \(M\). A point in \(U\) can be written as \((q, u)\), where \(q \in M\) and \(u = (u_1, \ldots, u_m) \in \mathbb{R}^m\).
   If we denote by \(\{e_1, \ldots, e_m\}\) an orthonormal basis of \(\mathbb{R}^m\), then we can define globally \(m\) smooth vector fields on \(M\) by \(f_i(q) := f(q, e_i)\) for \(i = 1, \ldots, m\). Then we have

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

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

---

1. Isomorphism of bundles in the broad sense, it is fiberwise but is not obliged to map a fiber in the same fiber.
and with minimal length is rewritten as the optimal control problem

\[
\begin{aligned}
\dot{\gamma}(t) &= \sum_{i=1}^{m} u_i(t) f_i(\gamma(t)) \\
\int_{0}^{T} |u(t)| dt &\rightarrow \min \\
\gamma(0) &= q_0, \quad \gamma(T) = q_1
\end{aligned}
\]  

(3.20)

For a free sub-Riemannian structure, the set of vector fields \( f_1, \ldots, f_m \) build as above is called a \textit{generating family}. Notice that, in general, a generating family is not orthonormal when \( f \) is not injective.

5. \textbf{Surfaces in} \( \mathbb{R}^3 \) \textbf{as free sub-Riemannian structures}

Due to topological constraints, in general it is not possible to regard a surface as a free sub-Riemannian structure of rank 2, i.e., defined by a pair of globally defined orthonormal vector fields. However, it is always possible to regard it as a free sub-Riemannian structure of rank 3.

Indeed, for an embedded surface \( M \) in \( \mathbb{R}^3 \), consider the trivial Euclidean bundle \( U = M \times \mathbb{R}^3 \), where points are denoted as usual \((q, u)\), with \( u \in \mathbb{R}^3, q \in M \), and the map

\[
f : U \rightarrow TM, \quad f(q, u) = \pi_q^+(u) \in T_q M.
\]  

(3.21)

where \( \pi_q^+ : \mathbb{R}^3 \rightarrow T_q M \subset \mathbb{R}^3 \) is the orthogonal projection.

Notice that \( f \) is a surjective bundle map and the set of vector fields \( \{\pi_q^+(\partial_x), \pi_q^+(\partial_y), \pi_q^+(\partial_z)\} \) is a generating family for this structure.

\textbf{Exercise 3.24.} Show that \((U, f)\) defined in (3.21) is equivalent to the Riemannian structure on \( M \) induced by the embedding in \( \mathbb{R}^3 \).

3.1.4 \textbf{Every sub-Riemannian structure is equivalent to a free one}

The purpose of this section is to show that every sub-Riemannian structure \((U, f)\) on \( M \) is equivalent to a sub-Riemannian structure \((U', f')\) where \( U' \) is a trivial bundle with sufficiently big rank.

\textbf{Lemma 3.25.} Let \( M \) be a \( n \)-dimensional smooth manifold and \( \pi : E \rightarrow M \) a smooth vector bundle of rank \( m \). Then, there exists a vector bundle \( \pi_0 : E_0 \rightarrow M \) with rank \( E_0 \leq 2n + m \) such that \( E \oplus E_0 \) is a trivial vector bundle.

\textit{Proof.} Remember that \( E \), as a smooth manifold, has dimension

\[
\dim E = \dim M + \text{rank } E = n + m
\]

Consider the map \( i : M \hookrightarrow E \) which embeds \( M \) into the vector bundle \( E \) as the zero section \( M_0 = i(M) \). If we denote with \( T_M E := i^*(TE) \) the pullback vector bundle, i.e., the restriction of \( TE \) to the section \( M_0 \), we have the isomorphism (as vector bundles on \( M \))

\[
T_M E \simeq E \oplus TM.
\]  

(3.22)
Eq. (3.22) is a consequence of the fact that the tangent to every fibre \( E_q \), being a vector space, is canonically isomorphic to its tangent space \( T_q E_q \) so that
\[
T_q E = T_q E_q \oplus T_q M \cong E_q \oplus T_q M, \quad \forall q \in M.
\]
By Whitney theorem we have a (nonlinear on fibers, in general) immersion
\[
\Psi : E \rightarrow \mathbb{R}^N, \quad \Psi_* : T_M E \subset TE \rightarrow T\mathbb{R}^N,
\]
for \( N = 2(n+m) \), and \( \Psi_* \) is injective as bundle map, i.e., \( T_M E \) is a sub-bundle of \( T\mathbb{R}^N \cong \mathbb{R}^N \times \mathbb{R}^N \). Thus we can choose as a complement \( E' \), the orthogonal bundle (on the base \( M \)) with respect to the Euclidean metric in \( \mathbb{R}^N \), i.e.
\[
E' = \bigcup_{q \in M} E'_q, \quad E'_q = (T_q E_q \oplus T_q M)^{\perp},
\]
and considering \( E_0 := T_M E \oplus E' \) we have that \( E_0 \) is trivial since its fibers are sum of orthogonal complements and by (3.22) we are done.

\( \square \)

**Corollary 3.26.** Every sub-Riemannian structure \((U, f)\) on \( M \) is equivalent to a sub-Riemannian structure \((\overline{U}, \overline{f})\) where \( \overline{U} \) is a trivial bundle.

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

Let us define the sub-Riemannian structure \((\overline{U}, \overline{f})\) on \( M \) by
\[
\overline{f} : \overline{U} \rightarrow TM, \quad \overline{f} := f \circ p_1,
\]
where \( p_1 : U \oplus U' \rightarrow U \) denotes the projection on the first factor. By construction, the diagram

\[
\begin{array}{ccc}
\overline{U} & \xrightarrow{\overline{f}} & TM \\
\downarrow \text{Id} & & \downarrow f \\
U \oplus U' & \xrightarrow{p_1} & U
\end{array}
\]

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

\( \square \)

Since every sub-Riemannian structure is equivalent to a free one, in what follows we can assume that there exists a global generating family, i.e., a family of \( f_1, \ldots, f_m \) of vector fields globally defined on \( M \) such that every admissible curve of the sub-Riemannian structure satisfies
\[
\dot{\gamma}(t) = \sum_{i=1}^{m} u_i(t) f_i(\gamma(t)), \quad (3.24)
\]
Moreover, by the classical Gram-Schmidt procedure, we can assume that $f_i$ are the image of an orthonormal frame defined on the fiber. (cf. Example [1] of Section [3.1.3])

Under these assumptions the length of an admissible curve $\gamma$ is given by

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

where $u^*(t)$ is the minimal control associated with $\gamma$.

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

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

Exercise 3.28. Consider a sub-Riemannian structure $(U, f)$ over $M$. Let $m = \text{rank}(U)$ and $h_{\text{max}} = \max\{h(q) : q \in M\} \leq m$ where $h(q)$ is the local minimal bundle rank at $q$. Prove that there exists a sub-Riemannian structure $(\tilde{U}, \tilde{f})$ equivalent to $(U, f)$ such that $\text{rank}(\tilde{U}) = h_{\text{max}}$.

3.1.5 Proto sub-Riemannian structures

Sometimes can be useful to consider structures that satisfy only property (i) and (ii) of Definition 3.2, but that are not bracket generating. In what follows we call these structures proto sub-Riemannian structures.

The typical example is the one of a Riemannian foliation, that is obtained when the family of horizontal vector fields $D$ satisfies

(i) $[D, D] \subset D$,

(ii) $\dim D_q$ does not depend on $q \in M$.

In this case the manifold $M$ is foliated by integral manifolds of the distribution, and each of them is endowed with a Riemannian structure.

3.2 Sub-Riemannian distance and Chow-Rashevskii Theorem

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

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

Definition 3.29. Let $M$ be a sub-Riemannian manifold and $q_0, q_1 \in M$. The sub-Riemannian distance (or Carnot-Caratheodory distance) between $q_0$ and $q_1$ is

$$d(q_0, q_1) = \inf\{\ell(\gamma) : [0, T] \to M \text{ admissible, } \gamma(0) = q_0, \gamma(T) = q_1\}, \quad (3.25)$$
One of the purpose of this section is to show that, thanks to the bracket generating condition, (9.1) is well-defined, namely for every \( q_0, q_1 \in M \), there exists an admissible curve that joins \( q_0 \) to \( q_1 \), hence \( d(q_0, q_1) < +\infty \).

**Theorem 3.30** (Chow-Raschevskii). Let \( M \) be a sub-Riemannian manifold. Then

(i) \((M, d)\) is a metric space,

(ii) the topology induced by \((M, d)\) is equivalent to the manifold topology.

In particular, \( d : M \times M \to \mathbb{R} \) is continuous.

In what follows \( B(q, r) \) (sometimes denoted also \( B_r(q) \)) is the (open) sub-Riemannian ball of radius \( r \) and center \( q \):

\[
B(q, r) := \{ q' \in M \mid d(q, q') < r \}.
\]

The rest of this section is devoted to the proof of Theorem 3.30. To prove it, we have to show that \( d \) is actually a distance, i.e.,

(a) \( 0 \leq d(q_0, q_1) < +\infty \) for all \( q_0, q_1 \in M \),

(b) \( d(q_0, q_1) = 0 \) if and only if \( q_0 = q_1 \),

(c) \( d(q_0, q_1) = d(q_1, q_0) \) and \( d(q_0, q_2) \leq d(q_0, q_1) + d(q_1, q_2) \) for all \( q_0, q_1, q_2 \in M \),

and the equivalence between the metric and the manifold topology: for every \( q_0 \in M \) we have

(d) for every \( \varepsilon > 0 \) there exists a neighborhood \( O_{q_0} \) of \( q_0 \) such that \( O_{q_0} \subset B(q_0, \varepsilon) \),

(e) for every neighborhood \( O_{q_0} \) of \( q_0 \) there exists \( \delta > 0 \) such that \( B(q_0, \delta) \subset O_{q_0} \).

### 3.2.1 Proof of Chow-Raschevskii Theorem

The symmetry of \( d \) is a direct consequence of the fact that if \( \gamma : [0, T] \to M \) is admissible, then the curve \( \tilde{\gamma} : [0, T] \to M \) defined by \( \tilde{\gamma}(t) = \gamma(T - t) \) is admissible and \( \ell(\tilde{\gamma}) = \ell(\gamma) \). The triangular inequality follows from the fact that, given two admissible curves \( \gamma_1 : [0, T_1] \to M \) and \( \gamma_2 : [0, T_2] \to M \) such that \( \gamma_1(T_1) = \gamma_2(0) \), their concatenation

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

is still admissible. These two arguments prove item (c).

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

S1. We prove that, for every \( q_0 \in M \), there exists a neighborhood \( O_{q_0} \) of \( q_0 \) such that \( d(q_0, \cdot) \) is finite and continuous in \( O_{q_0} \). This proves (d).

S2. We prove that \( d \) is finite on \( M \times M \). This proves (a).

S3. We prove (b) and (e).

To prove Step 1 we first need the following lemmas:
Lemma 3.31. Let $N \subset M$ be a submanifold and $\mathcal{F} \subset \text{Vec}(M)$ be a family of vector fields tangent to $N$, i.e., $X(q) \in T_qN$, for every $q \in N$ and $X \in \mathcal{F}$. Then for all $q \in N$ we have $\text{Lie}_q \mathcal{F} \subset T_q N$. In particular $\dim \text{Lie}_q \mathcal{F} \leq \dim N$.

Proof. Let $X \in \mathcal{F}$. As a consequence of the local existence and uniqueness of the two Cauchy problems

\[
\begin{align*}
\dot{q} &= X(q), & q &\in M, \\
q(0) &= q_0, & q_0 &\in N.
\end{align*}
\]

and

\[
\begin{align*}
\dot{q} &= X|_N(q), & q &\in N, \\
q(0) &= q_0, & q_0 &\in N.
\end{align*}
\]

it follows that $e^{tX}(q) \in N$ for every $q \in N$ and $t$ small enough. This property, together with the definition of Lie bracket (see formula (2.27)) implies that, if $X, Y$ are tangent to $N$, the vector field $[X,Y]$ is tangent to $N$ as well. Iterating this argument we get that $\text{Lie}_q \mathcal{F} \subset T_q N$ for every $q \in N$, from which the conclusion follows.

Lemma 3.32. Let $M$ be an $n$-dimensional sub-Riemannian manifold with generating family $\mathcal{F} = \{f_1, \ldots, f_n\}$. For every $q_0 \in M$ and every neighborhood $V$ of the origin in $\mathbb{R}^n$ there exist $\hat{s} = (\hat{s}_1, \ldots, \hat{s}_n) \in V$, and a choice of $n$ vector fields $f_1, \ldots, f_n \in \mathcal{F}$, such that $\hat{s}$ is a regular point of the map

\[
\psi : \mathbb{R}^n \rightarrow M, \quad \psi(s_1, \ldots, s_n) = e^{s_n f_n} \circ \cdots \circ e^{s_1 f_1}(q_0).
\]

Remark 3.33. Notice that, if $\mathcal{D}_{q_0} \neq T_{q_0} M$, then $\hat{s} = 0$ cannot be a regular point of the map $\psi$. Indeed, for $s = 0$, the image of the differential of $\psi$ at 0 is $\text{span}_{q_0} \{f_j, j = 1, \ldots, n\} \subset \mathcal{D}_{q_0}$ and the differential of $\psi$ cannot be surjective.

We stress that, in the choice of $f_1, \ldots, f_n \in \mathcal{F}$, a vector field can appear more than once, as for instance in the case $m < n$.

Proof of Lemma 3.32. We prove the lemma by steps.

1. There exists a vector field $f_{i_1} \in \mathcal{F}$ such that $f_{i_1}(q_0) \neq 0$, otherwise all vector fields in $\mathcal{F}$ vanish at $q_0$ and $\dim \text{Lie}_{q_0} \mathcal{F} = 0$, which contradicts the bracket generating condition. Then, for $|s|$ small enough, the map

\[
\phi_1 : s_1 \mapsto e^{s_1 f_{i_1}}(q_0),
\]

is a local diffeomorphism onto its image $\Sigma_1$. If $\dim M = 1$ the Lemma is proved.

2. Assume $\dim M \geq 2$. Then there exist $t_i^1 \in \mathbb{R}$, with $|t_i^1|$ small enough, and $f_{i_2} \in \mathcal{F}$ such that, if we denote by $q_1 = e^{t_1^1 f_{i_1}}(q_0)$, the vector $f_{i_2}(q_1)$ is not tangent to $\Sigma_1$. Otherwise, by Lemma 3.31 $\dim \text{Lie}_q \mathcal{F} = 1$, which contradicts the bracket generating condition. Then the map

\[
\phi_2 : (s_1, s_2) \mapsto e^{s_2 f_{i_2}} \circ e^{s_1 f_{i_1}}(q_0),
\]

is a local diffeomorphism near $(t_1^1, 0)$ onto its image $\Sigma_2$. Indeed the vectors

\[
\frac{\partial \phi_2}{\partial s_1} \bigg|_{(t_1^1, 0)} \in T_{q_1} \Sigma_1, \quad \frac{\partial \phi_2}{\partial s_2} \bigg|_{(t_1^1, 0)} = f_{i_2}(q_1),
\]

are linearly independent by construction. If $\dim M = 2$ the Lemma is proved.
3. Assume \( \dim M \geq 3 \). Then there exist \( t_1^1, t_2^2 \), with \( |t_2^2 - t_1^1| \) and \( |t_2^2| \) small enough, and \( f_{i_3} \in \mathcal{F} \) such that, if \( q_2 = e^{t_2^1 f_{i_2}} \circ e^{t_1^1 f_{i_1}}(q_0) \) we have that \( f_{i_3}(q_2) \) is not tangent to \( \Sigma_2 \). Otherwise, by Lemma 3.31 \( \dim \text{Lie}_{q_1} \mathcal{D} = 2 \), which contradicts the bracket generating condition. Then the map

\[
\phi_3 : (s_1, s_2, s_3) \mapsto e^{s_3 f_{i_3}} \circ e^{s_2 f_{i_2}} \circ e^{s_1 f_{i_1}}(q_0),
\]

is a local diffeomorphism near \((t_1^1, t_2^2, 0)\). Indeed the vectors

\[
\frac{\partial \phi_3}{\partial s_1}(t_1^1, t_2^2, 0), \quad \frac{\partial \phi_3}{\partial s_2}(t_1^1, t_2^2, 0) \in T_{q_2} \Sigma_2,
\]

\[
\frac{\partial \phi_3}{\partial s_3}(t_1^1, t_2^2, 0) = f_{i_3}(q_2),
\]

are linearly independent since the last one is transversal to \( T_{q_2} \Sigma_2 \) by construction, while the first two are linearly independent since \( \phi_3(s_1, s_2, 0) = \phi_2(s_1, s_2) \) and \( \phi_2 \) is a local diffeomorphism at \((t_1^1, t_2^2)\) which is close to \((t_1^1, 0)\).

Repeating the same argument \( n \) times (with \( n = \dim M \)), the lemma is proved.

**Proof of Step 1.** Thanks to Lemma 3.32 there exists a neighborhood \( \hat{V} \subset V \) of \( \hat{s} \) such that \( \psi \) is a diffeomorphism from \( \hat{V} \) to \( \psi(\hat{V}) \), see Figure 3.3. We stress that in general \( q_0 = \psi(0) \) does not belong to \( \psi(\hat{V}) \), cf. Remark 3.33.

![Figure 3.3: Proof of Lemma 3.32](image)

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

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

which has the following property: \( \hat{\psi} \) is a diffeomorphism from a neighborhood of \( \hat{s} \in V \), that we still denote \( \hat{V} \), to a neighborhood of \( \hat{\psi}(\hat{s}) = q_0 \).

Fix now \( \epsilon > 0 \) and apply the construction above where \( V \) is the neighborhood of the origin in \( \mathbb{R}^n \) defined by \( V = \{ s \in \mathbb{R}^n, \sum_{i=1}^n |s_i| < \epsilon \} \). Let us show that the claim of Step 1 holds with \( O_{q_0} = \hat{\psi}(\hat{V}) \). Indeed, for every \( q \in \hat{\psi}(\hat{V}) \), let \( s = (s_1, \ldots, s_n) \) such that \( q = \hat{\psi}(s) \), and denote by \( \gamma \) the admissible curve joining \( q_0 \) to \( q \), built by \( 2n \)-pieces, as in Figure 3.4.
In other words $\gamma$ is the concatenation of integral curves of the vector fields $f_{ij}$, i.e., admissible curves of the form $t \mapsto e^{t f_{ij}}(q)$ defined on some interval $[0, T]$, whose length is less or equal than $T$ (cf. Remark 3.27). Since $s, s \in \hat{V} \subset V$, it follows that:

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

which ends the proof of Step 1.

Proof of Step 2. To prove that $d$ is finite on $M \times M$ let us consider the equivalence classes of points in $M$ with respect to the relation

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

From the triangular inequality and the proof of Step 1, it follows that each equivalence class is open. Moreover, by definition, the equivalence classes are disjoint and nonempty. Since $M$ is connected, it cannot be the union of open disjoint and nonempty subsets. It follows that there exists only one equivalence class.

**Lemma 3.34.** Let $q_0 \in M$ and $K \subset M$ a compact set with $q_0 \in \text{int } K$. Then there exists $\delta_K > 0$ such that every admissible curve $\gamma$ starting from $q_0$ and with $\ell(\gamma) \leq \delta_K$ is contained in $K$.

**Proof.** Without loss of generality we can assume that $K$ is contained in a coordinate chart of $M$, where we denote by $|\cdot|$ the Euclidean norm in the coordinate chart. Let us define

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

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

Let us show that for any admissible curve $\gamma : [0, T] \to M$ such that $\gamma(0) = q_0$ and $\ell(\gamma) \leq \delta_K$ we have $\gamma([0, T]) \subset K$. Indeed, if this is not true, there exists an admissible curve $\gamma : [0, T] \to M$
Moreover, the sub-Riemannian metric \(d\) and \(n\) are enough, the sub-Riemannian ball is bounded. Thus small sub-Riemannian balls are compact.

Then
\[
|\gamma(t^*) - \gamma(0)| \leq \int_0^{t^*} |\dot{\gamma}(t)| \, dt = \int_0^{t^*} \sum_{i=1}^{m} |u_i^*(t)f_i(\gamma(t))| \, dt 
\]
\[
\leq \int_0^{t^*} \sqrt{\sum_{i=0}^{m} |f_i(\gamma(t))|^2} \sqrt{\sum_{i=0}^{m} u_i^*(t)^2} \, dt 
\]
\[
\leq C_K \int_0^{t^*} \sum_{i=0}^{m} u_i^*(t)^2 \, dt \leq C_K \ell(\gamma) 
\]
\[
\leq C_K \delta_K < \text{dist}(q_0, \partial K). 
\]
which contradicts the fact that, at \(t^*\), the curve \(\gamma\) leaves the compact \(K\). Thus \(t^* = T\).

\[\square\]

Proof of Step 3. Let us prove that Lemma 3.34 implies property (b). Indeed the only nontrivial implication is that \(d(q_0, q_1) > 0\) whenever \(q_0 \neq q_1\). To prove this, fix a compact neighborhood \(K\) of \(q_0\) such that \(q_1 \notin K\). By Lemma 3.34, each admissible curve joining \(q_0\) and \(q_1\) has length greater than \(\delta_K\), hence \(d(q_0, q_1) \geq \delta_K > 0\).

Let us now prove property (e). Fix \(\varepsilon > 0\) and a compact neighborhood \(K\) of \(q_0\). Define \(C_K\) and \(\delta_K\) as in Lemma 3.34 and set \(\delta := \min\{\delta_K, \varepsilon/C_K\}\). Let us show that \(|q - q_0| < \varepsilon\) whenever \(d(q_0, q) < \delta\), where again \(|\cdot|\) is the Euclidean norm in a coordinate chart.

Consider a minimizing sequence \(\gamma_n : [0, T] \to M\) of admissible trajectories joining \(q_0\) and \(q\) such that \(\ell(\gamma_n) \to d(q_0, q)\) for \(n \to \infty\). Without loss of generality, we can assume that \(\ell(\gamma_n) \leq \delta\) for all \(n\). By Lemma 3.34, \(\gamma_n([0, T]) \subset K\) for all \(n\).

We can repeat estimates (3.29)-(3.31) proving that \(|q - q_0| = |\gamma_n(T) - \gamma_n(0)| \leq C_K \ell(\gamma_n)\) for all \(n\). Passing to the limit for \(n \to \infty\), one gets
\[
|q - q_0| \leq C_K d(q_0, q) \leq C_K \delta < \varepsilon. 
\]

\[\square\]

Corollary 3.35. The metric space \((M, d)\) is locally compact, i.e., for any \(q \in M\) there exists \(\varepsilon > 0\) such that the closed sub-Riemannian ball \(\overline{B}(q, r)\) is compact for all \(0 \leq r \leq \varepsilon\).

Proof. By the continuity of \(d\), the set \(\overline{B}(q, r) = \{d(q, \cdot) \leq r\}\) is closed for all \(q \in M\) and \(r \geq 0\). Moreover the sub-Riemannian metric \(d\) induces the manifold topology on \(M\). Hence, for radius small enough, the sub-Riemannian ball is bounded. Thus small sub-Riemannian balls are compact.

\[\square\]

### 3.3 Existence of length-minimizers

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

**Definition 3.36.** Let \(\gamma : [0, T] \to M\) be an admissible curve. We say that \(\gamma\) is a *length-minimizer* if it minimizes the length among admissible curves with same endpoints, i.e., \(\ell(\gamma) = d(\gamma(0), \gamma(T))\).
Remark 3.37. Notice that the existence length-minimizers between two points is not guaranteed in general, as it happens for two points in \( M = \mathbb{R}^2 \setminus \{0\} \) (endowed with the Euclidean distance) that are symmetric with respect to the origin. On the other hand, when length-minimizers exist between two fixed points, they may not be unique, as it happens for two antipodal points on the sphere \( S^2 \).

We now show a general semicontinuity property of the length functional.

**Theorem 3.38.** Let \( \gamma_n : [0, T] \to M \) be a sequence of admissible curves on \( M \) such that \( \gamma_n \to \gamma \) uniformly on \( [0, T] \). Then

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

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

**Proof.** Without loss of generality we assume that \( \gamma_n \) and \( \gamma \) are parametrized with constant speed on the interval \([0, 1]\). Moreover, denote \( L := \liminf \ell(\gamma_n) \) and choose a subsequence, which we still denote by the same symbol, such that \( \ell(\gamma_n) \to L \). If \( L = +\infty \) the inequality (3.34) is clearly true, thus assume \( L < +\infty \).

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

\[
V_q = \{ f_u(q), |u| \leq L + \delta \} \subset T_q M, \quad f_u(q) = \sum_{i=1}^m u_i f_i(q).
\]

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

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

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

Next we want to estimate the right hand side of (3.35) uniformly. For \( n \geq n_0 \) sufficiently large, we have \(|\gamma_n(t) - \gamma(t)| < \varepsilon \) (by uniform convergence) and an estimate similar to (3.31) gives for \( \tau \in [t, t+\varepsilon] \)

\[
|\gamma_n(t) - \gamma(\tau)| \leq \int_t^T |\gamma_n(s)| ds \leq C_K (L + \delta) \varepsilon. \tag{3.36}
\]

where \( C_K \) is the constant (3.28) defined by the compact \( K \). Hence we deduce for every \( \tau \in [t, t+\varepsilon] \) and every \( n \geq n_0 \)

\[
|\gamma_n(\tau) - \gamma(\tau)| \leq |\gamma_n(t) - \gamma(\tau)| + |\gamma_n(t) - \gamma(t)| \leq C' \varepsilon, \tag{3.37}
\]

where \( C' \) is independent on \( n \) and \( \varepsilon \). From the estimate (3.37) and the equivalence of the manifold and metric topology we have that, for all \( \tau \in [t, t+\varepsilon] \) and \( n \geq n_0 \), \( \gamma_n(\tau) \in B_{\gamma(t)}(r_\varepsilon) \), with \( r_\varepsilon \to 0 \) when \( \varepsilon \to 0 \). In particular

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

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

\[
\frac{1}{\varepsilon}(\gamma(t + \varepsilon) - \gamma(t)) \in \text{conv}\{V_q, q \in B_{\gamma(t)}(r_\varepsilon)\}. \tag{3.39}
\]
Assume now that \( t \in [0, 1] \) is a differentiability point of \( \gamma \). Then the limit of the left hand side in (3.39) for \( \varepsilon \to 0 \) exists and gives \( \dot{\gamma}(t) \in \text{conv} \, V_{\gamma(t)} = V_{\gamma(t)} \). For every differentiability point \( t \) we can thus define the unique \( u^*(t) \) satisfying \( \dot{\gamma}(t) = f(\gamma(t), u^*(t)) \) and \( |u^*(t)| = |\dot{\gamma}(t)| \). Using the argument contained in Appendix 3.A it follows that \( u^*(t) \) is measurable in \( t \). Moreover \( |u^*(t)| \) is essentially bounded since, by construction, \( |u^*(t)| \leq L + \delta \) for a.e. \( t \in [0, T] \). Hence \( \gamma \) is admissible. Moreover \( \ell(\gamma) \leq L + \delta \) since \( \gamma \) is length-parametrized on the interval \([0, 1]\).

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

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

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

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

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

**Theorem 3.40 (Existence of minimizers).** Let \( M \) be a sub-Riemannian manifold and \( q_0 \in M \). Assume that the ball \( B_{q_0}(r) \) is compact, for some \( r > 0 \). Then for all \( q_1 \in B_{q_0}(r) \) there exists a length minimizer joining \( q_0 \) and \( q_1 \), i.e., we have

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

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

Since \( d(q_0, q_1) < r \), we have \( \ell(\gamma_n) \leq r \) for all \( n \geq n_0 \) large enough, hence we can assume without loss of generality that the image of \( \gamma_n \) is contained in the common compact \( K = \overline{B}_{q_0}(r) \) for all \( n \). In particular, the same argument leading to (3.38) shows that for all \( n \geq n_0 \)

\[
|\gamma_n(t) - \gamma_n(\tau)| \leq \int_{\tau}^{t} |\ddot{\gamma}_n(s)| ds \leq C_K r |t - \tau|, \quad \forall t, \tau \in [0, 1].
\]

where \( C_K \) depends only on \( K \). In other words, all trajectories in the sequence \( \{\gamma_n\}_{n \in \mathbb{N}} \) are Lipschitz with the same Lipschitz constant. Thus the sequence is equicontinuous and uniformly bounded.

By the classical Ascoli-Arzelà Theorem there exist a subsequence of \( \gamma_n \), which we still denote by the same symbol, and a Lipschitz curve \( \gamma : [0, T] \to M \) such that \( \gamma_n \to \gamma \) uniformly. By Theorem 3.38 the curve \( \gamma \) satisfies \( \ell(\gamma) \leq \liminf \ell(\gamma_n) = d(q_0, q_1) \), that implies \( \ell(\gamma) = d(q_0, q_1) \).

**Remark 3.41.** Assume that \( \bar{B}(q, r_0) \) is compact for some \( r_0 > 0 \). Then for every \( 0 < r \leq r_0 \) we have that \( \bar{B}(q, r) \) is compact also, being a closed subset of a compact set \( \bar{B}(q, r_0) \).

**Corollary 3.42.** Let \( q_0 \in M \). Under the hypothesis of Corollary 3.40 there exists \( \varepsilon > 0 \) such that for all \( r \leq \varepsilon \) and \( q_1 \in B_{q_0}(r) \) there exists a minimizing curve joining \( q_0 \) and \( q_1 \).

**Proof.** It is a direct consequence of Theorem 3.40 and Corollary 3.35.

**Remark 3.43.** It is well known that a length space is complete if and only if all closed balls are compact, see [11, Ch. 2]. In particular, if \((M, d)\) is complete with respect to the sub-Riemannian distance, then for every \( q_0, q_1 \in M \) there exists a length minimizer joining \( q_0 \) and \( q_1 \).
3.4 Pontryagin extremals

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

In classical Riemannian geometry length-minimizer trajectories satisfy a necessary condition given by a second order differential equation in \( M \), which can be reduced to a first-order differential equation in \( TM \). Hence the set of all length-minimizers is contained in the set of extremals, i.e., trajectories that satisfy the necessary condition, that are be parametrized by initial position and velocity.

In our setting (which includes Riemannian and sub-Riemannian geometry) we cannot use the initial velocity to parametrize length-minimizer trajectories. This can be easily understood by a dimensional argument. If the rank of the sub-Riemannian structure is smaller than the dimension of the manifold, the initial velocity \( \dot{\gamma}(0) \) of an admissible curve \( \gamma(t) \) starting from \( q_0 \), belongs to the proper subspace \( D_{q_0} \) of the tangent space \( T_{q_0} M \). Hence the set of admissible velocities form a set whose dimension is smaller than the dimension of \( M \), even if, by the Chow and Filippov theorems, length-minimizer trajectories starting from a point \( q_0 \) cover a full neighborhood of \( q_0 \).

The right approach is to parametrize length-minimizers by their initial point and an initial covector \( \lambda_0 \in T^*_{q_0} M \), which can be thought as the linear form annihilating the “front”, i.e., the set \( \{ \gamma_{q_0}(\varepsilon) | \gamma_{q_0} \text{ is a length-minimizer starting from } q_0 \} \) on the corresponding length-minimizer trajectory for \( \varepsilon \to 0 \).

The next theorem gives the necessary condition satisfied by length-minimizers in sub-Riemannian geometry. Curves satisfying this condition are called Pontryagin extremals. The proof the following theorem is given in the next section.

**Theorem 3.44 (Characterization of Pontryagin extremals).** Let \( \gamma : [0, T] \to M \) be an admissible curve which is a length-minimizer, parametrized by constant speed. Let \( \overline{u}(\cdot) \) be the corresponding minimal control, i.e., for a.e. \( t \in [0, T] \)

\[
\dot{\gamma}(t) = \sum_{i=1}^{m} \overline{u}_i(t) f_i(\gamma(t)), \quad \ell(\gamma) = \int_{0}^{T} |\overline{u}(t)| dt = d(\gamma(0), \gamma(T)),
\]

with \( |\overline{u}(t)| \) constant a.e. on \( [0, T] \). Denote with \( P_{0,t} \) the flow\(^2\) of the nonautonomous vector field \( f_{\overline{u}(t)} = \sum_{i=1}^{k} \overline{u}_i(t) f_i \). Then there exists \( \lambda_0 \in T^*_{\gamma(0)} M \) such that defining

\[
\lambda(t) := (P_{0,t}^{-1})^* \lambda_0, \quad \lambda(t) \in T^*_{\gamma(t)} M,
\]

we have that one of the following conditions is satisfied:

- **(N)** \( \overline{u}_i(t) \equiv \langle \lambda(t), f_i(\gamma(t)) \rangle \), \( \forall i = 1, \ldots, m \),
- **(A)** \( 0 \equiv \langle \lambda(t), f_i(\gamma(t)) \rangle \), \( \forall i = 1, \ldots, m \).

Moreover in case **(A)** one has \( \lambda_0 \neq 0 \).

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

\(^2\)\( P_{0,t}(x) \) is defined for \( t \in [0, T] \) and \( x \) in a neighborhood of \( \gamma(0) \)
Definition 3.45. Let $\gamma : [0, T] \to M$ be an admissible curve with minimal control $\overline{\pi} \in L^\infty([0, T], \mathbb{R}^m)$. Fix $\lambda_0 \in T_{\gamma(0)}^* M \setminus \{0\}$, and define $\lambda(t)$ by (3.41).

- If $\lambda(t)$ satisfies (N) then it is called normal extremal (and $\gamma(t)$ a normal extremal trajectory).
- If $\lambda(t)$ satisfies (A) then it is called abnormal extremal (and $\gamma(t)$ a abnormal extremal trajectory).

Remark 3.46. In the Riemannian case there are no abnormal extremals. Indeed, since the map $f$ is fiberwise surjective, we can always find $m$ vector fields $f_1, \ldots, f_m$ on $M$ such that $\text{span}_{q_0}\{f_1, \ldots, f_m\} = T_{q_0} M$, and (A) would imply that $\langle \lambda_0, v \rangle = 0$, for all $v \in T_{q_0} M$, that gives the contradiction $\lambda_0 = 0$.

Remark 3.47. If the sub-Riemannian structure is not Riemannian at $q_0$, namely if $D_{q_0} = \text{span}_{q_0}\{f_1, \ldots, f_m\} \neq T_{q_0} M$, then the trivial trajectory, corresponding to $\overline{\pi}(t) \equiv 0$, is always normal and abnormal.

Notice that even a nontrivial admissible trajectory $\gamma$ can be both normal and abnormal, since there may exist two different lifts $\lambda(t), \lambda'(t) \in T_{\gamma(t)}^* M$, such that $\lambda(t)$ satisfies (N) and $\lambda'(t)$ satisfies (A).

Exercise 3.48. Prove that condition (N) of Theorem 3.44 implies that the minimal control $\overline{\pi}(t)$ is smooth. In particular normal extremals are smooth.

At this level it seems not obvious how to use Theorem 3.44 to find the explicit expression of extremals for a given problem. In the next chapter we provide another formulation of Theorem 3.44 which gives Pontryagin extremals as solutions of a Hamiltonian system.

The rest of this section is devoted to the proof of Theorem 3.44.

3.4.1 The energy functional

Let $\gamma : [0, T] \to M$ be an admissible curve. We define the energy functional $J$ on the space of Lipschitz curves on $M$ as follows

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

Notice that $J(\gamma) < +\infty$ for every admissible curve $\gamma$.

Remark 3.49. While $\ell$ is invariant by reparametrization (see Remark 3.14), $J$ is not. Indeed consider, for every $\alpha > 0$, the reparametrized curve $\gamma_\alpha : [0, T/\alpha] \to M, \quad \gamma_\alpha(t) = \gamma(\alpha t).$

Using that $\dot{\gamma}_\alpha(t) = \alpha \dot{\gamma}(\alpha t)$, we have

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

Thus, if the final time is not fixed, the infimum of $J$, among admissible curves joining two fixed points, is always zero.
The following lemma relates minimizers of $J$ with fixed final time with minimizers of $\ell$.

**Lemma 3.50.** Fix $T > 0$ and let $\Omega_{q_0,q_1}$ be the set of admissible curves joining $q_0, q_1 \in M$. An admissible curve $\gamma : [0,T] \to M$ is a minimizer of $J$ on $\Omega_{q_0,q_1}$ if and only if it is a minimizer of $\ell$ on $\Omega_{q_0,q_1}$ and has constant speed.

**Proof.** Applying the Cauchy-Schwarz inequality
\[
\left( \int_0^T f(t)g(t)dt \right)^2 \leq \int_0^T f(t)^2 dt \int_0^T g(t)^2 dt,
\]
with $f(t) = \|\dot{\gamma}(t)\|$ and $g(t) = 1$ we get
\[
\ell(\gamma)^2 \leq 2J(\gamma)T.
\]
Moreover in (3.42) equality holds if and only if $f$ is proportional to $g$, i.e., $\|\dot{\gamma}(t)\| = \text{const.}$ in (3.43). Since, by Lemma 3.15, every curve is a Lipschitz reparametrization of a length-parametrized one, the minima of $J$ are attained at admissible curves with constant speed, and the statement follows.

### 3.4.2 Proof of Theorem 3.44

By Lemma 3.50 we can assume that $\gamma$ is a minimizer of the functional $J$ among admissible curves joining $q_0 = \gamma(0)$ and $q_1 = \gamma(T)$ in fixed time $T > 0$. In particular, if we define the functional
\[
\tilde{J}(u(\cdot)) := \frac{1}{2} \int_0^T |u(t)|^2 dt,
\]
on the space of controls $u(\cdot) \in L^\infty([0,T],[\mathbb{R}^m])$, the minimal control $\overline{u}(\cdot)$ of $\gamma$ is a minimizer for the energy functional $\tilde{J}$
\[
\tilde{J}(\overline{u}(\cdot)) \leq \tilde{J}(u(\cdot)), \quad \forall u \in L^\infty([0,T],[\mathbb{R}^m]),
\]
where trajectories corresponding to $u(\cdot)$ join $q_0, q_1 \in M$. In the following we denote the functional $\tilde{J}$ by $\tilde{J}$.

Consider now a variation $u(\cdot) = \overline{u}(\cdot) + v(\cdot)$ of the control $\overline{u}(\cdot)$, and its associated trajectory $q(t)$, solution of the equation
\[
\dot{q}(t) = f_{u(t)}(q(t)), \quad q(0) = q_0,
\]
Recall that $P_{0,t}$ denotes the local flow associated with the optimal control $\overline{u}(\cdot)$ and that $\gamma(t) = P_{0,t}(q_0)$ is the optimal admissible curve. We stress that in general, for $q$ different from $q_0$, the curve $t \mapsto P_{0,t}(q)$ is not optimal. Let us introduce the curve $x(t)$ defined by the identity
\[
q(t) = P_{0,t}(x(t)).
\]
In other words $x(t) = P_{0,t}^{-1}(q(t))$ is obtained by applying the inverse of the flow of $\overline{u}(\cdot)$ to the solution associated with the new control $u(\cdot)$ (see Figure 3.5). Notice that if $v(\cdot) = 0$, then $x(t) \equiv q_0$.

The next step is to write the ODE satisfied by $x(t)$. Differentiating (3.46) we get
\[
\dot{q}(t) = f_{\overline{u}(t)}(q(t)) + (P_{0,t})_*(\dot{x}(t))
\]
(3.47)
\[
= f_{\overline{u}(t)}(P_{0,t}(x(t))) + (P_{0,t})_*(\dot{x}(t))
\]
(3.48)
Figure 3.5: The trajectories $q(t)$, associated with $u(\cdot) = \pi(\cdot) + v(\cdot)$, and the corresponding $x(t)$.

and using that $\dot{q}(t) = f_u(t)(q(t)) = f_u(t)(P_{0,t}(x(t)))$ we can invert (3.48) with respect to $\dot{x}(t)$ and rewrite it as follows

$$\dot{x}(t) = \left(P_{0,t}^{-1}\right)_* \left[\left(f_u(t) - f_{\pi(t)}\right)(P_{0,t}(x(t)))\right]$$

$$= \left[P_{0,t}^{-1}\right]_* (f_u(t) - f_{\pi(t)}) (x(t))$$

$$= \left[P_{0,t}^{-1}\right]_* f_v(t) (x(t))$$

(3.49)

If we define the nonautonomous vector field $g_{v(t)}^t = \left(P_{0,t}^{-1}\right)_* f_v(t)$ we finally obtain by (3.49) the following Cauchy problem for $x(t)$

$$\dot{x}(t) = g_{v(t)}^t(x(t)), \quad x(0) = q_0. \quad (3.50)$$

Notice that the vector field $g_{v(t)}^t$ is linear with respect to $v$, since $f_u$ is linear with respect to $u$. Now we fix the control $v(t)$ and consider the map

$$s \in \mathbb{R} \mapsto \left(\frac{\partial J(\pi + sv)}{\partial s}, \frac{\partial x(T; \pi + sv)}{\partial s}\right) \in \mathbb{R} \times M$$

where $x(T; \pi + sv)$ denote the solution at time $T$ of (3.50), starting from $q_0$, corresponding to control $\pi(\cdot) + sv(\cdot)$, and $J(\pi + sv)$ is the associated cost.

**Lemma 3.51.** There exists $\bar{\lambda} \in (\mathbb{R} \oplus T_{q_0} M)^*$, with $\bar{\lambda} \neq 0$, such that for all $v \in L^\infty([0,T], \mathbb{R}^m)$

$$\left\langle \bar{\lambda}, \left(\frac{\partial J(\pi + sv)}{\partial s}, \frac{\partial x(T; \pi + sv)}{\partial s}\right)_{s=0}\right\rangle = 0. \quad (3.51)$$

**Proof of Lemma 3.51.** We argue by contradiction: assume that (3.51) is not true, then there exist $v_0, \ldots, v_n \in L^\infty([0,T], \mathbb{R}^m)$ such that the vectors in $\mathbb{R} \oplus T_{q_0} M$

$$\left(\frac{\partial J(\pi + sv_0)}{\partial s}, \frac{\partial x(T; \pi + sv_0)}{\partial s}\right)_{s=0}, \ldots, \left(\frac{\partial J(\pi + sv_n)}{\partial s}, \frac{\partial x(T; \pi + sv_n)}{\partial s}\right)_{s=0} \quad (3.52)$$
are linearly independent. Let us then consider the map
\[
\Phi : \mathbb{R}^{n+1} \rightarrow \mathbb{R} \times M, \quad \Phi(s_0, \ldots, s_n) = \left( \begin{array}{c} J(\overline{u} + \sum_{i=0}^{n} s_i v_i) \\ x(T; \overline{\omega} + \sum_{i=0}^{n} s_i v_i) \end{array} \right).
\]
(3.53)

By differentiability properties of solution of smooth ODEs with respect to parameters, the map \((3.53)\) is smooth in a neighborhood of \(s = 0\). Moreover, since the vectors \((3.52)\) are the components of the differential of \(\Phi\) and they are independent, then the inverse function theorem implies that \(\Phi\) is a local diffeomorphism sending a neighborhood of \(s = 0\) in \(\mathbb{R}^{n+1}\) in a neighborhood of \((J(\overline{u}), q_0)\) in \(\mathbb{R} \times M\). As a result we can find \(v(\cdot) = \sum_i s_i v_i(\cdot)\) such that (see also Figure 3.4.2)
\[
x(T; \overline{u} + v) = q_0, \quad J(\overline{u} + v) < J(\overline{u}).
\]
In other words the curve \(t \mapsto q(t; \overline{u} + v)\) joins \(q(0; \overline{u} + v) = q_0\) to
\[q(T; \overline{u} + v) = P_{0,T}(x(T; \overline{u} + v)) = P_{0,T}(q_0) = q_1,\]
with a cost smaller that the cost of \(\gamma(t) = q(t; \overline{u})\), which is a contradiction \(\square\)

**Remark 3.52.** Notice that if \(\lambda\) satisfies \((3.51)\), then for every \(\alpha \in \mathbb{R}\), with \(\alpha \neq 0\), \(\alpha \lambda\) satisfies \((3.51)\) too. Thus we can normalize \(\lambda\) to be \((-1, \lambda_0)\) or \((0, \lambda_0)\), with \(\lambda_0 \in T_{q_0}^* M\), and \(\lambda_0 \neq 0\) in the second case (since \(\lambda\) is not zero).

Condition \((3.51)\) implies that there exists \(\lambda_0 \in T_{q_0}^* M\) such that one of the following identities is satisfied for all \(v \in L^\infty([0, T], \mathbb{R}^m)\):
\[
\frac{\partial J(\overline{u} + sv)}{\partial s} \bigg|_{s=0} = \left< \lambda_0, \frac{\partial x(T; \overline{u} + sv)}{\partial s} \right|_{s=0},
\]
(3.54)
\[
0 = \left< \lambda_0, \frac{\partial x(T; \overline{u} + sv)}{\partial s} \right|_{s=0}.
\]
(3.55)
with \(\lambda_0 \neq 0\) in the second case (cf. Remark 3.52). To end the proof we have to show that identities \((3.54)\) and \((3.55)\) are equivalent to conditions (N) and (A) of Theorem 3.44. Let us show that
\[
\frac{\partial J(\overline{u} + sv)}{\partial s} \bigg|_{s=0} = \int_0^T \sum_{i=1}^m \overline{\omega}_i(t) v_i(t) dt,
\]
(3.56)
\[
\frac{\partial x(T; \overline{u} + sv)}{\partial s} \bigg|_{s=0} = \int_0^T \sum_{i=1}^m (P_{0,t}^{-1})_* f_i(q_0) v_i(t) dt.
\]
(3.57)
The identity (3.56) follows from the definition of $J$

$$J(\pi + sv) = \frac{1}{2} \int_0^T |\pi + sv|^2 dt.$$  \hfill (3.58)

Eq. (3.57) can be proved in coordinates. Indeed by (3.50) and the linearity of $g_v$ with respect to $v$ we have

$$x(T; \pi + sv) = q_0 + s \int_0^T g_v(t)(x(t; \pi + sv))dt,$$

and differentiating with respect to $s$ at $s = 0$ one gets (3.57).

Let us show that (3.54) is equivalent to (N) of Theorem 3.44. Similarly, one gets that (3.55) is equivalent to (A). Using (3.56) and (3.57), equation (3.54) is rewritten as

$$\int_0^T \sum_{i=1}^m v_i(t)dt = \int_0^T \sum_{i=1}^m \langle \lambda_0, ((P_0^{-1})_*, f_i)(q_0) \rangle v_i(t)dt$$

$$= \int_0^T \sum_{i=1}^m \langle \lambda(t), f_i(\gamma(t)) \rangle v_i(t)dt,$$  \hfill (3.59)

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

$$\langle \lambda_0, ((P_0^{-1})_*, f_i)(q_0) \rangle = \langle \lambda_0, (P_0^{-1})_* f_i(\gamma(t)) \rangle = \langle (P_0^{-1})_* \lambda_0, f_i(\gamma(t)) \rangle = \langle \lambda(t), f_i(\gamma(t)) \rangle.$$

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

### 3.A Measurability of the minimal control

In this appendix we prove a technical lemma about measurability of solutions to a class of minimization problems. This lemma when specified to the sub-Riemannian context, implies that the minimal control associated with an admissible curve is measurable.

#### 3.A.1 Main lemma

Let us fix an interval $I = [a, b] \subset \mathbb{R}$ and a compact set $U \subset \mathbb{R}^m$. Consider two functions $g : I \times U \rightarrow \mathbb{R}^n$, $v : I \rightarrow \mathbb{R}^n$ such that

(M1) $g(\cdot, u)$ is measurable in $t$ for every fixed $u \in U$,

(M2) $g(t, \cdot)$ is continuous in $u$ for every fixed $t \in I$,

(M3) $v(t)$ is measurable with respect to $t$.

Moreover we assume that

(M4) for every fixed $t \in I$, the problem $\min\{ |u| : g(t, u) = v(t), u \in U \}$ has a unique solution.

Let us denote by $u^*(t)$ the solution of (M4) for a fixed $t \in I$. 84
Lemma 3.53. Under assumptions (M1)-(M4), the function $t \mapsto |u^*(t)|$ is measurable on $I$.

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

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

is measurable in $\mathbb{R}$. By our assumptions

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

Let us fix $r > 0$ and a countable dense set $\{u_i\}_{i \in \mathbb{N}}$ in the ball of radius $r$ in $U$. Let show that

$$A = \bigcap_{n \in \mathbb{N}} A_n = \bigcap_{n \in \mathbb{N}} \bigcup_{i \in \mathbb{N}} A_{i,n}$$ (3.60)

where

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

Notice that the set $A_{i,n}$ is measurable by construction and if (3.60) is true, $A$ is also measurable.

$\subset$ inclusion. Let $t \in A$. This means that there exists $\bar{u} \in U$ such that $|\bar{u}| \leq r$ and $g(t, \bar{u}) = v(t)$. Since $g$ is continuous with respect to $u$ and $\{u_i\}_{i \in \mathbb{N}}$ is a dense, for each $n$ we can find $u_{i,n}$ such that $|g(t, u_{i,n}) - v(t)| < 1/n$, that is $t \in A_n$ for all $n$. □

$\supset$ inclusion. Assume $t \in \bigcap_{n \in \mathbb{N}} A_n$. Then for every $n$ there exists $i_n$ such that the corresponding $u_{i,n}$ satisfies $|g(t, u_{i,n}) - v(t)| < 1/n$. From the sequence $u_{i,n}$, by compactness, it is possible to extract a convergent subsequence $u_{i,n} \to \bar{u}$. By continuity of $g$ with respect to $u$ one easily gets that $g(t, \bar{u}) = v(t)$. That is $t \in A$.

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

Lemma 3.54. Under assumptions (M1)-(M4), the vector function $t \mapsto u^*(t)$ is measurable on $I$.

Proof. It is sufficient to prove that, for every closed ball $O$ in $\mathbb{R}^n$ the set

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

is measurable. Since the minimum in (M4) is uniquely determined, this is equivalent to

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

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

$$B = \bigcap_{n \in \mathbb{N}} B_n = \bigcap_{n \in \mathbb{N}} \bigcup_{i \in \mathbb{N}} B_{i,n}$$ (3.61)

where

$$B_{i,n} := \{ t \in I : |u_i| < \varphi(t) + 1/n, |g(t, u_i) - v(t)| < 1/n \}$$
Notice that the set $B_{i,n}$ is measurable by construction and if \((3.61)\) is true, $B$ is also measurable.

Let $t \in B$. This means that there exists $\bar{u} \in O$ such that $|\bar{u}| = \varphi(t)$ and $g(t, \bar{u}) = v(t)$. Since $g$ is continuous with respect to $u$ and $\{u_i\}_{i \in \mathbb{N}}$ is a dense in $O$, for each $n$ we can find $u_{i_n}$ such that $|g(t, u_{i_n}) - v(t)| < 1/n$ and $|u_{i_n}| < \varphi(t) + 1/n$, that is $t \in B_n$ for all $n$.

Assume $t \in \bigcap_{n \in \mathbb{N}} B_n$. Then for every $n$ it is possible to find $i_n$ such that the corresponding $u_{i_n}$ satisfies $|g(t, u_{i_n}) - v(t)| < 1/n$ and $|u_{i_n}| < \varphi(t) + 1/n$. From the sequence $u_{i_n}$, by compactness of the closed ball $O$, it is possible to extract a convergent subsquence $u_{i_n} \to \bar{u}$. By continuity of $f$ in $u$ one easily gets that $g(t, \bar{u}) = v(t)$. Moreover $|\bar{u}| \leq \varphi(t)$. Hence $|\bar{u}| = \varphi(t)$.

That is $t \in B$.

\[ \square \]

### 3.1.2 Proof of Lemma 3.11

Consider an admissible curve $\gamma : [0, T] \to M$. Since measurability is a local property it is not restrictive to assume $M = \mathbb{R}^n$. Moreover, by Lemma 3.55 we can assume that $\gamma$ is length-parametrized so that its minimal control belong to the compact set $U = \{|u| \leq 1\}$. Define $g : [0, T] \times U \to \mathbb{R}^n$ and $v : [0, T] \to \mathbb{R}^n$ by

$$g(t, u) = f(\gamma(t), u), \quad v(t) = \dot{\gamma}(t).$$

Assumptions (M1)-(M4) are satisfied. Indeed (M1)-(M3) follow from the fact that $g(t, u)$ is linear with respect to $u$ and measurable in $t$. Moreover (M4) is also satisfied by linearity with respect to $u$ of $f$. Applying Lemma 3.54 one gets that the minimal control $u^*(t)$ is measurable in $t$.

### 3.2 Lipschitz vs Absolutely continuous admissible curves

In these lecture notes sub-Riemannian geometry is developed in the framework of Lipschitz admissible curves (that correspond to the choice of $L^\infty$ controls). However, the theory can be equivalently developed in the framework of $H^1$ admissible curves (corresponding to $L^2$ controls) or in the framework of absolutely continuous admissible curves (corresponding to $L^1$ controls).

**Definition 3.55.** An absolutely continuous curve $\gamma : [0, T] \to M$ is said to be $AC$-admissible if there exists an $L^1$ function $u : t \in [0, T] \mapsto u(t) \in U_{\gamma(t)}$ such that $\dot{\gamma}(t) = f(\gamma(t), u(t))$, for a.e. $t \in [0, T]$. We define $H^1$-admissible curves similarly.

Being the set of absolutely continuous curve bigger than the set of Lipschitz ones, one could expect that the sub-Riemannian distance between two points is smaller when computed among all absolutely continuous admissible curves. However this is not the case thanks to the invariance by reparametrization. Indeed Lemmas 3.14 and 3.15 can be rewritten in the absolutely continuous framework in the following form.

**Lemma 3.56.** *The length of an AC-admissible curve is invariant by AC reparametrization.*

**Lemma 3.57.** *Any AC-admissible curve of positive length is a AC reparametrization of a length-parametrized admissible one.*
The proof of Lemma 3.56 differs from the one of Lemma 3.14 only by the fact that, if \( u^* \in L^1 \) is the minimal control of \( \gamma \) then \( (u^* \circ \varphi)\dot{\varphi} \) is the minimal control associated with \( \gamma \circ \varphi \). Moreover \( (u^* \circ \varphi)\dot{\varphi} \in L^1 \), using the monotonicity of \( \varphi \). Under these assumptions the change of variables formula (3.16) still holds.

The proof of Lemma 3.57 is unchanged. Notice that the statement of Exercise 3.16 remains true if we replace Lipschitz with absolutely continuous. We stress that the curve \( \gamma \) built in the proof is Lipschitz (since it is length-parametrized).

As a consequence of these results, if we define

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

we have the following proposition.

**Proposition 3.58.** \( d_{AC}(q_0, q_1) = d(q_0, q_1) \)

Since \( L^2([0, T]) \subset L^1([0, T]) \), Lemmas 3.56, 3.57 and Proposition 3.58 are valid also in the framework of admissible curves associated with \( L^2 \) controls.

**Bibliographical notes**

Sub-Riemannian manifolds have been introduced, even if with different terminology, in several contexts starting from the end of 60s, see for instance [23, 20, 15, 21, 16]. However, some pioneering ideas were already present in the work of Carathéodory and Cartan. The name sub-Riemannian geometry first appeared in [33].

Classical general references for sub-Riemannian geometry are [27, 4, 26, 17, 35]. Recent monographs [22, 31].

The definition of sub-Riemannian manifold using the language of bundles dates back to [2, 4]. For the original proof of the Rashevski-Chow theorem see [29, 12]. The proof of existence of sub-Riemannian length minimizer presented here is an adaptation of the proof of Filippov theorem in optimal control. The fact that in sub-Riemannian geometry there exist abnormal length minimizers is due to Montgomery [25, 27]. The fact that the theory can be equivalently developed for Lipschitz or absolutely continuous curves is well known, a discussion can be found in [4].

The definition of the length by using the minimal control is, up to our best knowledge, original. The problem of the measurability of the minimal control can be seen as a problem of differential inclusion [10]. The characterization of Pontryagin extremals given in Theorem 3.44 is a simplified version of the Pontryagin Maximum Principle (PMP) [28]. The proof presented here is original and adapted to this setting. For more general versions of PMP see [3, 5]. The fact that every sub-Riemannian structure is equivalent to a free one (cf. Section 3.1.4) is a consequence of classical results on fiber bundles. A different proof in the case of classical (constant rank) distribution was also considered in [31, 36].
Chapter 4

Characterization and local minimality of Pontryagin extremals

This chapter is devoted to the study of geometric properties of Pontryagin extremals. To this purpose we first rewrite Theorem 3.44 in a more geometric setting, which permits to write a differential equation in $T^*M$ satisfied by Pontryagin extremals and to show that they do not depend on the choice of a generating family. Finally we prove that small pieces of normal extremal trajectories are length-minimizers.

To this aim, all along this chapter we develop the language of symplectic geometry, starting by the key concept of Poisson bracket.

4.1 Geometric characterization of Pontryagin extremals

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

$$\lambda(t) = (P_{0,t}^{-1})^* \lambda_0, \quad \lambda(t) \in T_{\gamma(t)}^* M,$$

one of the following conditions is satisfied:

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

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

Here $P_{0,t}$ denotes the flow associated with the nonautonomous vector field $f_{u(t)} = \sum_{i=1}^m u_i(t) f_i$ and

$$(P_{0,t}^{-1})^* : T_q^* M \to T_{P_{0,t}(q)}^* M.$$  \hspace{1cm} (4.2)

is the induced flow on the cotangent space.

The goal of this section is to characterize the curve (4.1) as the integral curve of a suitable (non-autonomous) vector field on $T^*M$. To this purpose, we start by showing that a vector field on $T^*M$ is completely characterized by its action on functions that are affine on fibers. To fix the ideas, we first focus on the case in which $P_{0,t} : M \to M$ is the flow associated with an autonomous vector field $X \in \text{Vec}(M)$, namely $P_{0,t} = e^{tX}$.
4.1.1 Lifting a vector field from $M$ to $T^*M$

We start by some preliminary considerations on the algebraic structure of smooth functions on $T^*M$. As usual $\pi : T^*M \to M$ denotes the canonical projection.

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

$$C^\infty(M) \simeq C^\infty_{\text{pt}}(T^*M) := \{\pi^*\alpha \mid \alpha \in C^\infty(M)\} \subset C^\infty(T^*M).$$

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

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

$$\text{Vec}(M) \simeq C^\infty_{\text{lin}}(T^*M) := \{ a_Y \mid Y \in \text{Vec}(M) \} \subset C^\infty(T^*M).$$

Notice that this is an isomorphism as modules over $C^\infty(M)$. Indeed, as $\text{Vec}(M)$ is a module over $C^\infty(M)$, we have that $C^\infty_{\text{lin}}(T^*M)$ is a module over $C^\infty(M)$ as well. For any $\alpha \in C^\infty(M)$ and $a_X \in C^\infty_{\text{lin}}(T^*M)$ their product is defined as $\alpha a_X := (\pi^*\alpha)a_X = a_{\alpha X} \in C^\infty_{\text{lin}}(T^*M)$.

**Definition 4.1.** We say that a function $a \in C^\infty(T^*M)$ is affine on fibers if there exist two functions $\alpha \in C^\infty_{\text{lin}}(T^*M)$ and $a_X \in C^\infty_{\text{lin}}(T^*M)$ such that $a = \alpha + a_X$. In other words

$$a(\lambda) = \alpha(q) + \langle \lambda, X(q) \rangle, \quad q = \pi(\lambda).$$

We denote by $C^\infty_{\text{aff}}(T^*M)$ the set of affine function on fibers.

**Remark 4.2.** Linear and affine functions on $T^*M$ are particularly important since they reflects the linear structure of the cotangent bundle. In particular every vector field on $T^*M$, as a derivation of $C^\infty(T^*M)$, is completely characterized by its action on affine functions.

Indeed for a vector field $V \in \text{Vec}(T^*M)$ and $f \in C^\infty(T^*M)$, one has that

$$\left(\frac{d}{dt}|_{t=0} f(e^{tV}(\lambda))\right) = \langle d\lambda f, V(\lambda) \rangle, \quad \lambda \in T^*M.$$ (4.5)

which depends only on the differential of $f$ at the point $\lambda$. Hence, for each fixed $\lambda \in T^*M$, to compute (4.5) one can replace the function $f$ with any affine function whose differential at $\lambda$ coincide with $d\lambda f$. Notice that such a function is not unique.

Let us now consider the infinitesimal generator of the flow $(P_{0,t})^* = (e^{-tX})^*$. Since it satisfies the group law

$$(e^{-tX})^* \circ (e^{-sX})^* = (e^{-(t+s)X})^* \quad \forall t, s \in \mathbb{R},$$

by Lemma 2.15 its infinitesimal generator is an autonomous vector field $V_X$ on $T^*M$. In other words we have $(e^{-tX})^* = e^{tV_X}$ for all $t$.

Let us then compute the right hand side of (4.5) when $V = V_X$ and $f$ is either a function constant on fibers or a function linear on fibers.

The action of $V_X$ on functions that are constant on fibers, of the form $\beta \circ \pi$ with $\beta \in C^\infty(M)$, coincides with the action of $X$. Indeed we have for all $\lambda \in T^*M$

$$\left.\frac{d}{dt}\right|_{t=0} \beta \circ \pi((e^{-tX})^*\lambda)) = \left.\frac{d}{dt}\right|_{t=0} \beta(e^{tX}(q)) = (X\beta)(q), \quad q = \pi(\lambda).$$ (4.6)
For what concerns the action of $V_X$ on functions that are linear on fibers, of the form $a_Y(\lambda) = \langle \lambda, Y(q) \rangle$, we have for all $\lambda \in T^*M$

$$\frac{d}{dt} a_Y((e^{-tX})^*\lambda) = \frac{d}{dt} \langle (e^{-tX})^*\lambda, Y(e^{tX}(q)) \rangle$$

$$= \frac{d}{dt} \langle \lambda, (e^{-tX}Y)(q) \rangle = \langle \lambda, [X,Y](q) \rangle$$

$$= a_{[X,Y]}(\lambda).$$

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

$$V_X(\beta + a_Y) = X\beta + a_{[X,Y]}.$$  \hspace{1cm} (4.8)

As explained in Remark 4.2, formula (4.8) characterizes completely the generator $V_X$ of $(P_{0,t})^*$. To find its explicit form we introduce the notion of Poisson bracket.

### 4.1.2 The Poisson bracket

The purpose of this section is to introduce an operation $\{\cdot, \cdot\}$ on $C^\infty(T^*M)$, called Poisson bracket. First we introduce it in $C^\infty_{lin}(T^*M)$, where it reflects the Lie bracket of vector fields in $\text{Vec}(M)$, seen as elements of $C^\infty_{lin}(T^*M)$. Then it is uniquely extended to $C^\infty_{aff}(T^*M)$ and $C^\infty(T^*M)$ by requiring that it is a derivation of the algebra $C^\infty(T^*M)$ in each argument.

More precisely we start by the following definition.

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

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

(4.9)

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

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

$$[X,\alpha Y] = \alpha [X,Y] + (X\alpha) Y, \quad \forall \alpha \in C^\infty(M).$$

(4.10)

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

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

(4.11)

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

Next, we extend this definition on the whole $C^\infty(T^*M)$.

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

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

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

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

(4.12)

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

91
Proof. We start by proving that, as a consequence of the requirement that \{·, ·\} is a derivation in each argument, it is uniquely extended to $C^\infty_{aff}(T^*M)$.

By linearity and skew-symmetry we are reduced to compute Poisson brackets of kind \{\(a_X, \alpha\}\} and \{\(\alpha, \beta\}\}, where \(a_X \in C^\infty_{lin}(T^*M)\) and \(\alpha, \beta \in C^\infty_{aff}(T^*M)\). Using that \(a_{\alpha Y} = a a_Y\) and (4.12) one gets

\[
\{a_X, a_{\alpha Y}\} = \{a_X, \alpha a_Y\} = \alpha\{a_X, a_Y\} + \{a_X, \alpha\}a_Y. \tag{4.13}
\]

Comparing (4.11) and (4.13) one gets

\[
\{a_X, \alpha\} = X\alpha. \tag{4.14}
\]

Next, using (4.12) and (4.14), one has

\[
\{a_{\alpha Y}, \beta\} = \{\alpha a_Y, \beta\} = \alpha\{a_Y, \beta\} + \{\alpha, \beta\}a_Y \tag{4.15}
\]

Using again (4.14) one also has \(\{a_{\alpha Y}, \beta\} = \alpha Y\beta\), hence \(\{\alpha, \beta\} = 0\).

Combining the previous formulas one obtains the following expression for the Poisson bracket between two affine functions on \(T^*M\)

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

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

Next we extend this definition to \(C^\infty(T^*M)\) in such a way that it is still a derivation. For \(f, g \in C^\infty(T^*M)\) we define

\[
\{f, g\}|_\lambda := \{a_{f,\lambda}, a_{g,\lambda}\}|_\lambda \tag{4.18}
\]

where \(a_{f,\lambda}\) and \(a_{g,\lambda}\) are two functions in \(C^\infty_{aff}(T^*M)\) such that \(d_\lambda f = d_\lambda (a_{f,\lambda})\) and \(d_\lambda g = d_\lambda (a_{g,\lambda})\).

Remark 4.6. The definition (4.18) is well posed, since if we take two different affine functions \(a_{f,\lambda}\) and \(a'_{f,\lambda}\), their difference satisfy \(d_\lambda (a_{f,\lambda} - a'_{f,\lambda}) = d_\lambda (a_{f,\lambda}) - d_\lambda (a'_{f,\lambda}) = 0\), hence by bilinearity of the Poisson bracket

\[
\{a_{f,\lambda}, a_{g,\lambda}\}|_\lambda = \{a'_{f,\lambda}, a_{g,\lambda}\}|_\lambda.
\]

Let us now compute the coordinate expression of the Poisson bracket. In canonical coordinates \((p, x)\) in \(T^*M\), if

\[
X = \sum_{i=1}^{n} X_i(x) \frac{\partial}{\partial x_i}, \quad Y = \sum_{i=1}^{n} Y_i(x) \frac{\partial}{\partial x_i},
\]

we have

\[
a_X(p, x) = \sum_{i=1}^{n} p_i X_i(x), \quad a_Y(p, x) = \sum_{i=1}^{n} p_i Y_i(x).
\]

92
and, denoting \( f = a_X + \alpha, \ g = a_Y + \beta \) we have

\[
\{ f, g \} = a_{[X,Y]} + X\beta - Y\alpha
\]

\[
= \sum_{i,j=1}^n p_j \left( X_i \frac{\partial Y_j}{\partial x_i} - Y_i \frac{\partial X_j}{\partial x_i} \right) + X_i \frac{\partial \beta}{\partial p_i} - Y_i \frac{\partial \alpha}{\partial p_i}
\]

\[
= \sum_{i,j=1}^n X_i \left( p_j \frac{\partial Y_j}{\partial x_i} + \frac{\partial \beta}{\partial p_i} \right) - Y_i \left( p_j \frac{\partial X_j}{\partial x_i} + \frac{\partial \alpha}{\partial p_i} \right)
\]

\[
= \sum_{i=1}^n \frac{\partial f}{\partial p_i} \frac{\partial g}{\partial x_i} - \frac{\partial f}{\partial x_i} \frac{\partial g}{\partial p_i}.
\]

From these computations we get the formula for Poisson brackets of two functions \( a, b \in C^\infty(T^*M) \)

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

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

**Remark 4.7.** We stress that the value \( \{ a, b \}|_\lambda \) at a point \( \lambda \in T^*M \) depends only on \( d_\lambda a \) and \( d_\lambda b \). Hence the Poisson bracket computed at the point \( \lambda \in T^*M \) can be seen as a skew-symmetric and nondegenerate bilinear form

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

### 4.1.3 Hamiltonian vector fields

By construction, the linear operator defined by

\[
\bar{a} : C^\infty(T^*M) \to C^\infty(T^*M) \quad \bar{a}(b) := \{ a, b \}
\]

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

**Definition 4.8.** The vector field \( \bar{a} \) on \( T^*M \) defined by (4.20) is called the **Hamiltonian vector field** associated with the smooth function \( a \in C^\infty(T^*M) \).

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

\[
\bar{a} = \sum_{i=1}^n \frac{\partial a}{\partial p_i} \frac{\partial}{\partial x_i} - \frac{\partial a}{\partial x_i} \frac{\partial}{\partial p_i}. \tag{4.21}
\]

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

**Proposition 4.9.** Let \( X \in \text{Vec}(M) \) be complete and let \( P_{0,t} = e^{tX} \). The flow on \( T^*M \) defined by \( (P_{0,t})^* = (e^{-tX})^* \) is generated by the Hamiltonian vector field \( \bar{a}_X \), where \( a_X(\lambda) = \langle \lambda, X(q) \rangle \) and \( q = \pi(\lambda) \).
Proof. To prove that the generator $V$ of $(P_{0,t}^{-1})^*$ coincides with the vector field $\vec{a}_X$ it is sufficient to show that their action is the same. Indeed, by definition of Hamiltonian vector field, we have

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

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

Remark 4.10. In coordinates $(p,x)$ if the vector field $X$ is written $X = \sum_{i=1}^n X_i \frac{\partial}{\partial x_i}$ then $a_X(p,x) = \sum_{i=1}^n p_i X_i$ and the Hamiltonian vector field $\vec{a}_X$ is written as follows

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

Notice that the projection of $\vec{a}_X$ onto $M$ coincides with $X$ itself, i.e., $\pi_*(\vec{a}_X) = X$.

This construction can be extended to the case of nonautonomous vector fields.

**Proposition 4.11.** Let $X_t$ be a nonautonomous vector field and denote by $P_{0,t}$ the flow of $X_t$ on $M$. Then the nonautonomous vector field on $T^*M$ $V_t := \overrightarrow{a_X}$, $a_X(\lambda) = \langle \lambda, X_t(q) \rangle$, is the generator of the flow $(P_{0,t}^{-1})^*$.

### 4.2 The symplectic structure

In this section we introduce the symplectic structure of $T^*M$ following the classical construction. In subsection 4.2.1 we show that the symplectic form can be interpreted as the “dual” of the Poisson bracket, in a suitable sense.

**Definition 4.12.** The tautological (or Liouville) 1-form $s \in \Lambda^1(T^*M)$ is defined as follows:

$$s : \lambda \mapsto s_\lambda \in T^*_\lambda(T^*M), \quad \langle s_\lambda, w \rangle := \langle \lambda, \pi_*w \rangle, \quad \forall \lambda \in T^*M, w \in T_\lambda(T^*M),$$

where $\pi : T^*M \to M$ denotes the canonical projection.

The name “tautological” comes from its expression in coordinates. Recall that, given a system of coordinates $x = (x_1, \ldots, x_n)$ on $M$, canonical coordinates $(p,x)$ on $T^*M$ are coordinates for which every element $\lambda \in T^*M$ is written as follows

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

For every $w \in T_\lambda(T^*M)$ we have the following

$$w = \sum_{i=1}^n \alpha_i \frac{\partial}{\partial p_i} + \beta_i \frac{\partial}{\partial x_i} \implies \pi_*w = \sum_{i=1}^n \beta_i \frac{\partial}{\partial x_i}.$$
hence we get
\[ \langle s_\lambda, w \rangle = \langle \lambda, \pi_\ast w \rangle = \sum_{i=1}^{n} p_i \beta_i = \sum_{i=1}^{n} p_i \langle dx_i, w \rangle = \left\langle \sum_{i=1}^{n} p_i dx_i, w \right\rangle. \]

In other words the coordinate expression of the Liouville form \( s \) at the point \( \lambda \) coincides with the one of \( \lambda \) itself, namely
\[ s_\lambda = \sum_{i=1}^{n} p_i dx_i. \quad (4.23) \]

**Exercise 4.13.** Let \( s \in \Lambda^1(T^*M) \) be the tautological form. Prove that
\[ \omega^\ast s = \omega, \quad \forall \omega \in \Lambda^1(M). \]
(Recall that a 1-form \( \omega \) is a section of \( T^*M \), i.e. a map \( \omega : M \to T^*M \) such that \( \pi \circ \omega = id_M \)).

**Definition 4.14.** The differential of the tautological 1-form \( \sigma := ds \in \Lambda^2(T^*M) \) is called the **canonical symplectic structure** on \( T^*M \).

By construction \( \sigma \) is a closed 2-form on \( T^*M \). Moreover its expression in canonical coordinates \((p,x)\) shows immediately that is a nondegenerate two form
\[ \sigma = \sum_{i=1}^{n} dp_i \wedge dx_i. \quad (4.24) \]

**Remark 4.15 (The symplectic form in non-canonical coordinates).** Given a basis of 1-forms \( \omega_1, \ldots, \omega_n \) in \( \Lambda^1(M) \), one can build coordinates on the fibers of \( T^*M \) as follows.

Every \( \lambda \in T^*M \) can be written uniquely as \( \lambda = \sum_{i=1}^{n} h_i \omega_i \). Thus \( h_i \) become coordinates on the fibers. Notice that these coordinates are not related to any choice of coordinates on the manifold, as the \( p \) were. By definition, in these coordinates, we have
\[ s = \sum_{i=1}^{n} h_i \omega_i, \quad \sigma = ds = \sum_{i=1}^{n} dh_i \wedge \omega_i + h_i d\omega_i. \quad (4.25) \]

Notice that, with respect to (4.24) in the expression of \( \sigma \) an extra term appears since, in general, the 1-forms \( \omega_i \) are not closed.

### 4.2.1 The symplectic form vs the Poisson bracket

Let \( V \) be a finite dimensional vector space and \( V^* \) denotes its dual (i.e. the space of linear forms on \( V \)). By classical linear algebra arguments one has the following identifications
\[ \left\{ \text{non degenerate bilinear forms on } V \right\} \cong \left\{ \text{linear invertible maps } V \to V^* \right\} \cong \left\{ \text{non degenerate bilinear forms on } V^* \right\}. \quad (4.26) \]

Indeed to every bilinear form \( B : V \times V \to \mathbb{R} \) we can associate a linear map \( L : V \to V^* \) defined by \( L(v) = B(v, \cdot) \). On the other hand, given a linear map \( L : V \to V^* \), we can associate with it a bilinear map \( B : V \times V \to \mathbb{R} \) defined by \( B(v, w) = \langle L(v), w \rangle \), where \( \langle \cdot, \cdot \rangle \) denotes as usual the
pairing between a vector space and its dual. Moreover $B$ is non-degenerate if and only if the map $B(v, \cdot)$ is an isomorphism for every $v \in V$, that is if and only if $L$ is invertible.

The previous argument shows how to identify a bilinear form on $B$ on $V$ with an invertible linear map $L$ from $V$ to $V^*$. Applying the same reasoning to the linear map $L^{-1}$ one obtain a bilinear map on $V^*$.

**Exercise 4.16.** (a). Let $h \in C^\infty(T^*M)$. Prove that the Hamiltonian vector field $\vec{h} \in \text{Vec}(T^*M)$ satisfies the following identity

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

(b). Prove that, for every $\lambda \in T^*M$ the bilinear forms $\sigma_{\lambda}$ on $T_{\lambda}(T^*M)$ and $\{\cdot, \cdot\}_{\lambda}$ on $T^*_{\lambda}(T^*M)$ (cf. Remark 4.17) are dual under the identification (4.26). In particular show that

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

**Remark 4.17.** Notice that $\sigma$ is nondegenerate, which means that the map $w \mapsto \sigma_{\lambda}(\cdot, w)$ defines a linear isomorphism between the vector spaces $T_{\lambda}(T^*M)$ and $T^*_{\lambda}(T^*M)$. Hence $\vec{h}$ is the vector field canonically associated by the symplectic structure with the differential $dh$. For this reason $\vec{h}$ is also called symplectic gradient of $h$.

From formula (4.24) we have that in canonical coordinates $(p, x)$ the Hamiltonian vector filed associated with $h$ is expressed as follows

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

and the Hamiltonian system $\dot{\lambda} = \vec{h}(\lambda)$ is rewritten as

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

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

**Proposition 4.18.** A function $a \in C^\infty(T^*M)$ is a constant of the motion of the Hamiltonian system associated with $h \in C^\infty(T^*M)$ if and only if $\{h, a\} = 0$.

**Proof.** Let us consider a solution $\lambda(t) = e^{t\vec{h}}(\lambda_0)$ of the Hamiltonian system associated with $\vec{h}$, with $\lambda_0 \in T^*M$. Let us prove the following formula for the derivative of the function $a$ along the solution

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

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

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

**Corollary 4.19.** A function $h \in C^\infty(T^*M)$ is a constant of the motion of the Hamiltonian system defined by $\vec{h}$.
4.3 Characterization of normal and abnormal extremals

Now we can rewrite Theorem 3.44 using the symplectic language developed in the last section.

Given a sub-Riemannian structure on \( M \) with generating family \( \{f_1, \ldots, f_m\} \), and define the fiberwise linear functions on \( T^* M \) associated with these vector fields \( h_i : T^* M \to \mathbb{R} \), \( h_i(\lambda) := \langle \lambda, f_i(q) \rangle \), \( i = 1, \ldots, m \).

**Theorem 4.20** (PMP). Let \( \gamma : [0, T] \to M \) be an admissible curve which is a length-minimizer, parametrized by constant speed. Let \( \overline{u}(\cdot) \) be the corresponding minimal control. Then there exists a Lipschitz curve \( \lambda(t) \in T^* \gamma(t) M \) such that

\[
\dot{\lambda}(t) = \sum_{i=1}^{m} \overline{u}_i(t) \overline{h}_i(\lambda(t)), \quad \text{a.e.} \; t \in [0, T],
\]

and one of the following conditions is satisfied:

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

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

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

**Proof.** The statement is a rephrasing of Theorem 3.44 obtained by combining Proposition 4.9 and Exercise 4.11.

Notice that Theorem 4.20 says that normal and abnormal extremals appear as solution of an Hamiltonian system. Nevertheless, this Hamiltonian system is non autonomous and depends on the trajectory itself by the presence of the control \( \overline{u}(t) \) associated with the extremal trajectory.

Moreover, the actual formulation of Theorem 4.20 for the necessary condition for optimality still does not clarify if the extremals depend on the generating family \( \{f_1, \ldots, f_m\} \) for the sub-Riemannian structure. The rest of the section is devoted to the geometric intrinsic description of normal and abnormal extremals.

### 4.3.1 Normal extremals

In this section we show that normal extremals are characterized as solutions of a smooth autonomous Hamiltonian system on \( T^* M \), where the Hamiltonian \( H \) is a function that encodes all the informations on the sub-Riemannian structure.

**Definition 4.21.** Let \( M \) be a sub-Riemannian manifold. The sub-Riemannian Hamiltonian is the function on \( T^* M \) defined as follows

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

**Proposition 4.22.** The sub-Riemannian Hamiltonian \( H \) is smooth and quadratic on fibers. Moreover, for every generating family \( \{f_1, \ldots, f_m\} \) of the sub-Riemannian structure, the sub-Riemannian Hamiltonian \( H \) is written as follows

\[
H(\lambda) = \frac{1}{2} \sum_{i=1}^{m} \langle \lambda, f_i(q) \rangle^2, \quad \lambda \in T^*_q M, \quad q = \pi(\lambda). \tag{4.31}
\]
Proof. In terms of a generating family \( \{ f_1, \ldots, f_m \} \), the sub-Riemannian Hamiltonian (4.30) is written as follows

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

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

Exercise 4.23. Prove that two equivalent sub-Riemannian structures \((U, f)\) and \((U', f')\) on a manifold \(M\) define the same Hamiltonian.

Theorem 4.24. Every normal extremal is a solution of the Hamiltonian system \( \dot{\lambda}(t) = \vec{H}(\lambda(t)) \).

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

\[
\vec{H}(\lambda(t)) = \sum_{i=1}^{m} h_i(\lambda(t)) \vec{h}_i(\lambda(t)) = \sum_{i=1}^{m} \vec{u}_i(t) \vec{h}_i(\lambda(t)). \tag{4.33}
\]

In particular, since along a normal extremal \( h_i(\lambda(t)) = \vec{u}_i(t) \) by condition (N) of Theorem 4.20, one gets

\[
\vec{H}(\lambda(t)) = \sum_{i=1}^{m} h_i(\lambda(t)) \vec{h}_i(\lambda(t)) = \sum_{i=1}^{m} \vec{u}_i(t) \vec{h}_i(\lambda(t)). \tag{4.33}
\]

Remark 4.25. In canonical coordinates \( \lambda = (p, x) \) in \( T^*M \), \( H \) is quadratic with respect to \( p \) and

\[
H(p, x) = \frac{1}{2} \sum_{i=1}^{m} \langle p, f_i(x) \rangle^2.
\]

The Hamiltonian system associated with \( H \), in these coordinates, is written as follows

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

From here it is easy to see that if \( \lambda(t) = (p(t), x(t)) \) is a solution of (4.33) then also the rescaled extremal \( \alpha \lambda(\alpha t) = (\alpha p(\alpha t), x(\alpha t)) \) is a solution of the same Hamiltonian system, for every \( \alpha > 0 \).

Lemma 4.26. Let \( \lambda(t) \) be a normal extremal and \( \gamma(t) = \pi(\lambda(t)) \) be the corresponding normal extremal trajectory. Then for all \( t \in [0, T] \) one has

\[
\frac{1}{2} ||\dot{\gamma}(t)||^2 = H(\lambda(t)).
\]
Proof. For every normal extremal $\lambda(t)$ associated with the (minimal) control $\overline{u}(\cdot)$ we have

$$\frac{1}{2}\|\dot{\gamma}(t)\|^2 = \frac{1}{2}\|\overline{u}(t)\|^2 = \frac{1}{2} \sum_{i=1}^{k} \overline{u}_i(t)^2 = H(\lambda(t))$$

(4.34)

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

$$\overline{u}_i(t) = \langle \lambda(t), f_i(\gamma(t)) \rangle .$$

Corollary 4.27. A normal extremal trajectory is parametrized by constant speed. In particular it is length parametrized if and only if its extremal lift is contained in the level set $H^{-1}(1/2)$.

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

Finally, by Remark 4.25 all normal extremal trajectories are reparametrization of length parametrized ones.

Let $\lambda(t)$ be a normal extremal such that $\lambda(0) = \lambda_0 \in T^*_0 M$. The corresponding normal extremal trajectory $\gamma(t) = \pi(\lambda(t))$ can be written in the exponential notation

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

By Corollary 4.27 length-parametrized normal extremal trajectories corresponds to the choice of $\lambda_0 \in H^{-1}(1/2)$.

We end this section by characterizing normal extremal trajectory as characteristic curves of the canonical symplectic form contained in the level sets of $H$.

Definition 4.28. Let $M$ be a smooth manifold and $\Omega \in \Lambda^2 M$ a 2-form. A Lipschitz curve $\gamma : [0,T] \to M$ is a characteristic curve for $\Omega$ if for almost every $t \in [0,T]$ it holds

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

(4.35)

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

Proposition 4.29. Let $H$ be the sub-Riemannian Hamiltonian and assume that $c > 0$ is a regular value of $H$. Then a Lipschitz curve $\gamma$ is a characteristic curve for $\sigma|_{H^{-1}(c)}$ if and only if it is the reparametrization of a normal extremal on $H^{-1}(c)$.

Proof. Recall that if $c$ is a regular value of $H$, then the set $H^{-1}(c)$ is a smooth $(2n-1)$-dimensional manifold in $T^* M$ (notice that by Sard Theorem almost every $c > 0$ is regular value for $H$).

For every $\lambda \in H^{-1}(c)$ let us denote by $E_\lambda = T_\lambda H^{-1}(c)$ its tangent space at this point. Notice that, by construction, $E_\lambda$ is an hyperplane (i.e., dim $E_\lambda = 2n-1$) and $d_\lambda H|_{E_\lambda} = 0$. The restriction $\sigma|_{H^{-1}(c)}$ is computed by $\sigma_\lambda|_{E_\lambda}$, for each $\lambda \in H^{-1}(c)$.

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

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

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

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

**Exercise 4.30.** Prove that if two smooth Hamiltonians $h_1, h_2 : T^*M \to \mathbb{R}$ define the same level set, i.e. $E = \{h_1 = c_1\} = \{h_2 = c_2\}$ for some $c_1, c_2 \in \mathbb{R}$, then their Hamiltonian flow $\vec{h}_1, \vec{h}_2$ coincide on $E$, up to reparametrization.

**Exercise 4.31.** The sub-Riemannian Hamiltonian $H$ encodes all the information about the sub-Riemannian structure.

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

$$\frac{1}{2} \langle \lambda, v \rangle^2 \leq H(\lambda), \quad \forall \lambda \in T^*_qM.$$ 

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

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

When the structure is Riemannian, $H$ is the “inverse” norm defined on the cotangent space.

### 4.3.2 Abnormal extremals

In this section we provide a geometric characterization of abnormal extremals. Even if for abnormal extremals it is not possible to determine a priori their regularity, we show that they can be characterized as characteristic curves of the symplectic form. This gives an unified point of view of both class of extremals.

We recall that an abnormal extremal is a non zero solution of the following equations

$$\dot{\lambda}(t) = \sum_{i=1}^m u_i(t) \vec{h}_i(\lambda(t)), \quad h_i(\lambda(t)) = 0, \; i = 1, \ldots, m.$$ 

where $\{f_1, \ldots, f_m\}$ is a generating family for the sub-Riemannian structure and $h_1, \ldots, h_m$ are the corresponding functions on $T^*M$ linear on fibers. In particular every abnormal extremal is contained in the set

$$H^{-1}(0) = \{\lambda \in T^*M \mid \langle \lambda, f_i(q) \rangle = 0, \; i = 1, \ldots, m, \; q = \pi(\lambda)\}. \quad (4.36)$$

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

**Proposition 4.32.** Let $H$ be the sub-Riemannian Hamiltonian and assume that $H^{-1}(0)$ is a smooth manifold. Then a Lipschitz curve $\gamma$ is a characteristic curve for $\sigma|_{H^{-1}(0)}$ if and only if it is the reparametrization of a abnormal extremal on $H^{-1}(0)$.
Proof. In this proof we denote for simplicity $N := H^{-1}(0) \subset T^* M$. For every $\lambda \in N$ we have the identity
\[
\text{Ker} \sigma|_N = T_\lambda N^\perp = \text{span}\{\vec{h}_i(\lambda) \mid i = 1, \ldots, m\}.
\] (4.37)
Indeed, from the definition of $N$, it follows that
\[
T_\lambda N = \{w \in T_\lambda (T^* M) \mid \langle d_\lambda h_i, w \rangle = 0, i = 1, \ldots, m\}
\]
\[
= \{w \in T_\lambda (T^* M) \mid \sigma(w, \vec{h}_i(\lambda)) = 0, i = 1, \ldots, m\}
\]
\[
= \text{span}\{\vec{h}_i(\lambda) \mid i = 1, \ldots, m\}^\perp.
\]
and (4.37) follows by taking the skew-orthogonal on both sides. Thus $w \in T_\lambda H^{-1}(0)$ if and only if $w$ is a linear combination of the vectors $\vec{h}_i(\lambda)$. This implies that $\lambda(t)$ is a characteristic curve for $\sigma|_{H^{-1}(0)}$ if and only if there exists controls $u_i(\cdot)$ for $i = 1, \ldots, m$ such that
\[
\dot{\lambda}(t) = \sum_{i=1}^m u_i(t) \vec{h}_i(\lambda(t)).
\] (4.38)

Notice that 0 is never a regular value of $H$. Nevertheless, the following exercise shows that the assumption of Proposition 4.32 is always satisfied in the case of a regular sub-Riemannian structure.

Exercise 4.33. Assume that the sub-Riemannian structure is regular, namely the following assumption holds
\[
\text{dim} \mathcal{D}_q = \text{dim} \text{span}_q \{f_1, \ldots, f_m\} = \text{const.}
\] (4.39)
Then prove that the set $H^{-1}(0)$ defined by (4.36) is a smooth submanifold of $T^* M$.

Remark 4.34. From Proposition 4.32 it follows that abnormal extremals do not depend on the sub-Riemannian metric, but only on the distribution. Indeed the set $H^{-1}(0)$ is characterized as the annihilator $\mathcal{D}^\perp$ of the distribution
\[
H^{-1}(0) = \{\lambda \in T^* M \mid \langle \lambda, v \rangle = 0, \ \forall v \in \mathcal{D}_{\pi(\lambda)}\} = \mathcal{D}^\perp \subset T^* M.
\]
Here the orthogonal is meant in the duality sense.

Under the regularity assumption (4.39) we can select (at least locally) a basis of 1-forms $\omega_1, \ldots, \omega_m$ for the dual of the distribution
\[
\mathcal{D}^\perp = \text{span}\{\omega_i(q) \mid i = 1, \ldots, m\},
\] (4.40)
Let us complete this set of 1-forms to a basis $\omega_1, \ldots, \omega_n$ of $T^* M$ and consider the induced coordinates $h_1, \ldots, h_n$ as defined in Remark 4.15. In these coordinates the restriction of the symplectic structure $\mathcal{D}^\perp$ to is expressed as follows
\[
\sigma|_{\mathcal{D}^\perp} = d(s|_{\mathcal{D}^\perp}) = \sum_{i=1}^m dh_i \wedge \omega_i + h_i d\omega_i,
\] (4.41)
We stress that the restriction $\sigma|_{\mathcal{D}^\perp}$ can be written only in terms of the elements $\omega_1, \ldots, \omega_m$ (and not of a full basis of 1-forms) since the differential $d$ commutes with the restriction.
4.3.3 Example: codimension one distribution and contact distributions

Let $M$ be a $n$-dimensional manifold endowed with a constant rank distribution $\mathcal{D}$ of codimension one, i.e., $\dim \mathcal{D}_q = n - 1$ for every $q \in M$. In this case $\mathcal{D}$ and $\mathcal{D}^\perp$ are sub-bundles of $TM$ and $T^*M$ respectively and their dimension, as smooth manifolds, are

$$\dim \mathcal{D} = \dim M + \text{rank} \mathcal{D} = 2n - 1,$$

$$\dim \mathcal{D}^\perp = \dim M + \text{rank} \mathcal{D}^\perp = n + 1.$$

Since the symplectic form $\sigma$ is skew-symmetric, a dimensional argument implies that for $n$ even, the restriction $\sigma|_{\mathcal{D}^\perp}$ has always a nontrivial kernel. Hence there always exist characteristic curves of $\sigma|_{\mathcal{D}^\perp}$, that correspond to reparametrized abnormal extremals by Proposition 4.32.

Let us consider in more detail the case $n = 3$. Assume that there exists a one form $\omega \in \Lambda^1(M)$ such that $\mathcal{D} = \ker \omega$ (this is not restrictive for a local description). Consider a basis of one forms $\omega_0, \omega_1, \omega_2$ such that $\omega_0 := \omega$ and the coordinates $h_0, h_1, h_2$ associated to these forms (see Remark 4.15). By (4.41)

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

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

$$\sigma \wedge \sigma|_{\mathcal{D}^\perp} = 2h_0 dh_0 \wedge \omega \wedge d\omega. \quad (4.43)$$

Lemma 4.35. Let $N$ be a smooth $2k$-dimensional manifold and $\Omega \in \Lambda^2 M$. Then $\Omega$ is nondegenerate on $N$ if and only if $\wedge^k \Omega \neq 0$. \[1\]

Definition 4.36. Let $M$ be a three dimensional manifold. We say that a constant rank distribution $\mathcal{D} = \ker \omega$ on $M$ of corank one is a contact distribution if $\omega \wedge d\omega \neq 0$.

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

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

Corollary 4.37. Under the previous assumptions all nontrivial abnormal extremal trajectories are contained in the Martinet set $\mathfrak{M}$. In particular, if the structure is contact, there are no nontrivial abnormal extremal trajectories.

Proof. By Proposition 4.32 any abnormal extremal $\lambda(t)$ is a characteristic curve of $\sigma|_{\mathcal{D}^\perp}$. By Lemma 4.35 $\sigma|_{\mathcal{D}^\perp}$ is degenerate if and only if $\sigma \wedge \sigma|_{\mathcal{D}^\perp} = 0$, which is in turn equivalent to $\omega \wedge d\omega = 0$ thanks to (4.43) (notice that $dh_0$ and $\omega \wedge d\omega$ are independent since they depend on coordinates on the fibers and on the manifold, respectively).

This shows that, if $\gamma(t)$ is an abnormal trajectory and $\lambda(t)$ is the associated abnormal extremal, then $\lambda(t)$ is a characteristic curve of $\sigma|_{\mathcal{D}^\perp}$ if and only if $(\omega \wedge d\omega)|_{\gamma(t)} = 0$, that is $\gamma(t) \in \mathfrak{M}$. By definition of $\mathfrak{M}$ it follows that, if $\mathcal{D}$ is contact, then $\mathfrak{M}$ is empty. \[\square\]

Remark 4.38. Since $M$ is three dimensional, we can write $\omega \wedge d\omega = adV$ where $a \in C^\infty(M)$ and $dV$ is some smooth volume form on $M$, i.e., a never vanishing 3-form on $M$. \[1\]

\[1\] Here $\wedge^k \Omega = \Omega \wedge \ldots \wedge \Omega$. 102
In particular the Martinet set is \( \mathcal{M} = a^{-1}(0) \) and the distribution is contact if and only if the function \( a \) is never vanishing. When 0 is a regular value of \( a \), the set \( a^{-1}(0) \) defines a two dimensional surface on \( M \), called the Martinet surface. Notice that this condition is satisfied for a generic choice of the (one form defining the) distribution.

Abnormal extremal trajectories are the horizontal curves that are contained in the Martinet surface. When \( \mathcal{M} \) is smooth, the intersection of the tangent bundle to the surface \( \mathcal{M} \) and the 2-dimensional distribution of admissible velocities defines, generically, a line field on \( \mathcal{M} \). Abnormal extremal trajectories coincide with the integral curves of this line field, up to a reparametrization.

4.4 Examples

4.4.1 2D Riemannian Geometry

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

\[
\dot{q}(t) = u_1(t)f_1(q(t)) + u_2(t)f_2(q(t)),
\]

where length and energy are expressed as

\[
\ell(q(\cdot)) = \int_0^T \sqrt{u_1(t)^2 + u_2(t)^2} \, dt, \quad J(q(\cdot)) = \frac{1}{2} \int_0^T (u_1(t)^2 + u_2(t)^2) \, dt.
\]

Geodesics are projections of integral curves of the sub-Riemannian Hamiltonian in \( T^*M \)

\[
H(\lambda) = \frac{1}{2}(h_1(\lambda)^2 + h_2(\lambda)^2), \quad h_i(\lambda) = \langle \lambda, f_i(q) \rangle, \quad i = 1, 2.
\]

Since the vector fields \( f_1 \) and \( f_2 \) are linearly independent, the functions \( (h_1, h_2) \) defines a system of coordinates on fibers of \( T^*M \). In what follows it is convenient to use \( (q, h_1, h_2) \) as coordinates on \( T^*M \) (even if coordinates on the manifold are not necessarily fixed).

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

\[
\langle \lambda(t), f_1(\gamma(t)) \rangle = \langle \lambda(t), f_2(\gamma(t)) \rangle = 0, \quad \forall t \in [0, T], \tag{4.44}
\]

that implies that \( \lambda(t) = 0 \) for all \( t \in [0, T] \) since \( \left\{f_1, f_2\right\} \) is a basis of the tangent space at every point. This is a contradiction since \( \lambda(t) \neq 0 \) by Theorem 3.44.

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

\[
\dot{q} = h_1 f_1(q) + h_2 f_2(q). \tag{4.45}
\]

For the equation on the fiber we have (remember that along solutions \( \dot{a} = \{H, a\} \))

\[
\begin{cases}
\dot{h}_1 = \{H, h_1\} = -\{h_1, h_2\}h_2 \\
\dot{h}_2 = \{H, h_2\} = \{h_1, h_2\}h_1.
\end{cases} \tag{4.46}
\]

From here one can see directly that \( H \) is constant along solutions. Indeed

\[
\dot{H} = h_1 \dot{h}_1 + h_2 \dot{h}_2 = 0.
\]
If we require that extremals are parametrized by arclength \( u_1(t)^2 + u_2(t)^2 = 1 \) for a.e. \( t \in [0, T] \), we have

\[
H(\lambda(t)) = \frac{1}{2} \iff h_1^2(\lambda(t)) + h_2^2(\lambda(t)) = 1.
\]

It is then convenient to restrict to the spherical cotangent bundle \( S^*M \) (see Example 2.44) of coordinates \((q, \theta)\), by setting

\[
h_1 = \cos \theta, \quad h_2 = \sin \theta.
\]

Let \( a_1, a_2 \in C^\infty(M) \) be such that

\[
[f_1, f_2] = a_1 f_1 + a_2 f_2. \tag{4.47}
\]

Since \( \{h_1, h_2\}(\lambda) = \langle \lambda, [f_1, f_2] \rangle \), we have \( \{h_1, h_2\} = a_1 h_1 + a_2 h_2 \) and equations (4.53) and (4.54) are rewritten in \((\theta, q)\) coordinates

\[
\begin{cases}
\dot{\theta} = a_1(q) \cos \theta + a_2(q) \sin \theta \\
\dot{q} = \cos \theta f_1(q) + \sin \theta f_2(q)
\end{cases} \tag{4.48}
\]

In other words we are saying that an arc-length parametrized curve on \( M \) (i.e. a curve which satisfies the second equation) is a geodesic if and only if it satisfies the first. Heuristically this suggests that the quantity

\[
\dot{\theta} - a_1(q) \cos \theta - a_2(q) \sin \theta,
\]

has some relation with the geodesic curvature on \( M \).

Let \( \mu_1, \mu_2 \) the dual frame of \( f_1, f_2 \) (so that \( dV = \mu_1 \wedge \mu_2 \)) and consider the Hamiltonian field in these coordinates

\[
\vec{H} = \cos \theta f_1 + \sin \theta f_2 + (a_1 \cos \theta + a_2 \sin \theta) \partial \theta. \tag{4.49}
\]

The Levi-Civita connection on \( M \) is expressed by some coefficients (see Chapter ??)

\[
\omega = d\theta + b_1 \mu_1 + b_2 \mu_2,
\]

where \( b_i = b_i(q) \). On the other hand geodesics are projections of integral curves of \( \vec{H} \) so that

\[
\langle \omega, \vec{H} \rangle = 0 \implies b_1 = -a_1, \quad b_2 = -a_2.
\]

In particular if we apply

\[
\lambda = \cos \theta f_1 + \sin \theta f_2 + \dot{\theta} \partial \theta
\]

which projects on \( \gamma \) we find geodesic curvature

\[
\kappa_g(\gamma) = \dot{\theta} - a_1(q) \cos \theta - a_2(q) \sin \theta,
\]

as we infer above. To end this section we prove a useful formula for the Gaussian curvature of \( M \)

**Corollary 4.39.** If \( \kappa \) denotes the Gaussian curvature of \( M \) we have

\[
\kappa = f_1(a_2) - f_2(a_1) - a_1^2 - a_2^2.
\]
Proof. From (1.58) we have $d\omega = -\kappa dV$ where $dV = \mu_1 \wedge \mu_2$ is the Riemannian volume form. On the other hand, using the following identities

$$d\mu_i = -a_i \mu_1 \wedge \mu_2, \quad da_i = f_1(a_i) \mu_1 + f_2(a_i) \mu_2, \quad i = 1, 2.$$ 

we can compute

$$d\omega = -da_1 \wedge \mu_1 - da_2 \wedge \mu_2 - a_1 d\mu_1 - a_2 d\mu_2$$
$$= -(f_1(a_2) - f_2(a_1) - a_1^2 - a_2^2) \mu_1 \wedge \mu_2.$$

\[ \square \]

4.4.2 Isoperimetric problem

Let $M$ be a 2-dimensional orientable Riemannian manifold and $\nu$ its Riemannian volume form. Fix a smooth one-form $A \in \Lambda^1 M$ and $c \in \mathbb{R}$.

**Problem 1.** Fix $c \in \mathbb{R}$ and $q_0, q_1 \in M$. Find, whenever it exists, the solution to

$$\min \left\{ \ell(\gamma) : \gamma(0) = q_0, \gamma(T) = q_1, \int_\gamma A = c \right\}. \quad (4.50)$$

**Remark 4.40.** Minimizers depend only on $dA$, i.e., if we add an exact term to $A$ we will find same minima for the problem (with a different value of $c$).

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

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

in the sense that solutions of the problem \([12,76]\) turns to be length minimizers for a suitable sub-Riemannian structure on $\overline{M}$, that we are going to construct.

To every curve $\gamma$ on $M$ satisfying $\gamma(0) = q_0$ and $\gamma(T) = q_1$ we can associate the function

$$z(t) = \int_{\gamma|\{0, t\}} A = \int_0^t A(\dot{\gamma}(s)) ds.$$

The curve $\xi(t) = (\gamma(t), z(t))$ defined on $\overline{M}$ satisfies $\omega(\dot{\xi}(t)) = 0$ where $\omega = dz - A$ is a one form on $\overline{M}$, since

$$\omega(\dot{\xi}(t)) = \dot{z}(t) - A(\dot{\gamma}(t)) = 0.$$

Equivalently, $\dot{\xi}(t) \in \mathcal{D}_{\xi(t)}$ where $\mathcal{D} = \ker \omega$. We define a metric on $\mathcal{D}$ by defining the norm of a vector $v \in \mathcal{D}$ as the Riemannian norm of its projection $\pi_* v$ on $M$, where $\pi : \overline{M} \to M$ is the canonical projection on the first factor. This endows $\overline{M}$ with a sub-Riemannian structure.

If we fix a local orthonormal frame $f_1, f_2$ for $M$, the pair $(\gamma(t), z(t))$ satisfies

$$\left( \begin{array}{c} \dot{\gamma} \\ \dot{z} \end{array} \right) = u_1 \left( \begin{array}{c} f_1 \\ \langle A, f_1 \rangle \end{array} \right) + u_2 \left( \begin{array}{c} f_2 \\ \langle A, f_2 \rangle \end{array} \right). \quad (4.51)$$

Hence the two vector fields on $\overline{M}$

$$F_1 = f_1 + \langle A, f_1 \rangle \partial_z, \quad F_2 = f_2 + \langle A, f_2 \rangle \partial_z,$$
defines an orthonormal frame for the metric defined above on $\mathcal{D} = \text{span}(F_1, F_2)$.

Problem 1 is then equivalent to the following:

**Problem 2.** Fix $c \in \mathbb{R}$ and $q_0, q_1 \in M$. Find, whenever it exists, the solution to

$$\min \left\{ \ell(\xi) : \xi(0) = (q_0, 0), \xi(T) = (q_1, c), \xi(t) \in \mathcal{D}_{\xi(t)} \right\}. \quad (4.52)$$

Notice that, by construction, $\mathcal{D}$ is a distribution of constant rank (equal to 2) but is not necessarily bracket-generating. Let us now compute normal and abnormal extremals associated to the sub-Riemannian structure just introduced on $\overline{M}$. In what follows we denote with $h_i(\lambda) = \langle \lambda, F_i(q) \rangle$ the Hamiltonians linear on fibers of $T^*M$.

**Normal extremals**

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

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

Let us introduce $F_0 = \partial_z$ and $h_0(\lambda) = \langle \lambda, F_0(q) \rangle$. Since $F_1, F_2$ and $F_0$ are linearly independent, then $(h_1, h_2, h_0)$ defines a system of coordinates on fibers of $T^*M$. In what follows it is convenient to use $(q, h_1, h_2, h_0)$ as coordinates on $T^*M$.

For a normal extremal we have $u_i(t) = h_i(\lambda(t))$ for $i = 1, 2$ and the equation on the base is

$$\dot{\lambda} = h_1 F_1(\lambda) + h_2 F_2(\lambda). \quad (4.53)$$

For the equation on the fibers we have (remember that along solutions $\dot{a} = \{H, a\}$)

$$\begin{aligned}
\dot{h}_1 &= \{H, h_1\} = -\{h_1, h_2\} h_2 \\
\dot{h}_2 &= \{H, h_2\} = \{h_1, h_2\} h_1. \\
\dot{h}_0 &= \{H, h_0\} = 0.
\end{aligned} \quad (4.54)$$

If we require that extremals are parametrized by arclength we can restrict to the cylinder of the cotangent bundle $T^*M$ defined by

$$h_1 = \cos \theta, \quad h_2 = \sin \theta.$$ 

Let $a_1, a_2 \in C^\infty(M)$ be such that

$$[f_1, f_2] = a_1 f_1 + a_2 f_2. \quad (4.55)$$

Then

$$[F_1, F_2] = [f_1 + \langle A, f_1 \rangle \partial_z, f_2 + \langle A, f_2 \rangle \partial_z]
= [f_1, f_2] + (f_1 \langle A, f_2 \rangle - f_2 \langle A, f_1 \rangle) \partial_z$$

(by (4.55))

$$= a_1(F_1 - \langle A, f_1 \rangle) + a_2(F_2 - \langle A, f_2 \rangle) + f_1 \langle A, f_2 \rangle - f_2 \langle A, f_1 \rangle) \partial_z$$

$$= a_1 F_1 + a_2 F_2 + dA(f_1, f_2) \partial_z.$$

where in the last equality we use Cartan formula (cf. (4.74) for a proof). Let $\mu_1, \mu_2$ be the dual forms to $f_1$ and $f_2$. Then $\nu = \mu_1 \wedge \mu_2$ and we can write $dA = b \mu_1 \wedge \mu_2$, for a suitable function $b \in C^\infty(M)$. In this case

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

106
and
\[ \{h_1, h_2\} = \langle \lambda, [F_1, F_2]\rangle = a_1 h_1 + a_2 h_2 + b h_0. \]  
(4.56)

With computations analogous to the 2D case we obtain the Hamiltonian system associated to \( H \) in the \((q, \theta, h_0)\) coordinates
\[
\begin{aligned}
\dot{\xi} &= \cos \theta F_1(\xi) + \sin \theta F_2(\xi) \\
\dot{\theta} &= a_1 \cos \theta + a_2 \sin \theta + bh_0 \\
\dot{h}_0 &= 0
\end{aligned}
\]  
(4.57)

In other words if \( q(t) = \bar{\pi}(\xi(t)) \) is the projection of a normal extremal path on \( M \) (here \( \bar{\pi} : \overline{M} \to M \)), its geodesic curvature
\[ \kappa_g(q(t)) = \dot{\theta}(t) - a_1(q(t)) \cos \theta(t) - a_2(q(t)) \sin \theta(t) \]  
(4.58)

satisfies
\[ \kappa_g(q(t)) = b(q(t)) h_0. \]  
(4.59)

Namely, projections on \( M \) of normal extremal paths are curves with geodesic curvature proportional to the function \( b \) at every point. The case \( b \) equal to constant is treated in the example of Section 4.4.3.

**Abnormal extremals**

We prove the following characterization of abnormal extremal

**Lemma 4.41.** Abnormal extremal trajectories are contained in the Martinet set \( \mathcal{M} = \{ b = 0 \} \).

**Proof.** Assume that \( \lambda(t) \) is an abnormal extremal whose projection is a curve \( \xi(t) = \pi(\lambda(t)) \) that is not reduced to a point. Then we have
\[ h_1(\lambda(t)) = \langle \lambda(t), F_1(\xi(t)) \rangle = 0, \quad h_2(\lambda(t)) = \langle \lambda(t), F_2(\xi(t)) \rangle = 0, \quad \forall t \in [0, T], \]  
(4.60)

We can differentiate the two equalities with respect to \( t \in [0, T] \) and we get
\[
\begin{aligned}
\frac{d}{dt} h_1(\lambda(t)) &= u_2(t)\{h_1, h_2\}\vert_{\lambda(t)} = 0 \\
\frac{d}{dt} h_2(\lambda(t)) &= -u_1(t)\{h_1, h_2\}\vert_{\lambda(t)} = 0
\end{aligned}
\]

Since the pair \((u_1(t), u_2(t)) \neq (0, 0)\) we have that \( \{h_1, h_2\}\vert_{\lambda(t)} = 0 \) that implies
\[ 0 = \langle \lambda(t), [F_1, F_2](\xi(t))\rangle = b(\xi(t)) h_0, \]  
(4.61)

where in the last equality we used (4.56) and the fact that \( h_1(\lambda(t)) = h_2(\lambda(t)) = 0 \). Recall that \( h_0 \neq 0 \) otherwise the covector is identically zero (that is not possible for abnormalities), then \( b(\xi(t)) = 0 \) for all \( t \in [0, T] \).

The last result shows that abnormal extremal trajectories are forced to live in connected components of \( b^{-1}(0) \).

**Exercise 4.42.** Prove that the set \( b^{-1}(0) \) is independent on the Riemannian metric chosen on \( M \) (and the corresponding sub-Riemannian metric defined on \( D \)).
4.4.3 Heisenberg group

The Heisenberg group is a basic example in sub-Riemannian geometry. It is the sub-Riemannian structure defined by the isoperimetric problem in \( M = \mathbb{R}^2 = \{(x,y)\} \) endowed with its Euclidean scalar product and the 1-form (cf. previous section)

\[
A = \frac{1}{2}(xdy - ydx).
\]

Notice that \( dA = dx \wedge dy \) defines the area form on \( \mathbb{R}^2 \), hence \( b \equiv 1 \) in this case. On the extended manifold \( \overline{M} = \mathbb{R}^3 = \{(x,y,z)\} \) the one-form \( \omega \) is written as

\[
\omega = dz - \frac{1}{2}(xdy - ydx).
\]

Following the notation of the previous paragraph we can choose as an orthonormal frame for \( \mathbb{R}^2 \) the frame \( f_1 = \partial_x \) and \( f_2 = \partial_y \). This induced the choice

\[
F_1 = \partial_x - \frac{y}{2} \partial_z, \quad F_2 = \partial_y + \frac{x}{2} \partial_z,
\]

for the orthonormal frame on \( \mathcal{D} = \ker \omega \). Notice that \([F_1, F_2] = \partial_z\), that implies that \( \mathcal{D} \) is bracket-generating at every point. Defining \( F_0 = \partial_z \) and \( h_i = \langle \lambda, F_i(q) \rangle \) for \( i = 0, 1, 2 \), the Hamiltonians linear on fibers of \( T^*\overline{M} \), we have

\[
\{h_1, h_2\} = h_0,
\]

hence the equation (4.57) for normal extremals become

\[
\begin{cases}
\dot{q} = \cos \theta F_1(q) + \sin \theta F_2(q) \\
\dot{\theta} = h_0 \\
\dot{h}_0 = 0
\end{cases}
\quad (4.62)
\]

It follows that the two last equation can be immediately solved

\[
\begin{cases}
\theta(t) = \theta_0 + h_0 t \\
h_0(t) = h_0
\end{cases}
\quad (4.63)
\]

Moreover

\[
\begin{cases}
h_1(t) = \cos(\theta_0 + h_0 t) \\
h_2(t) = \sin(\theta_0 + h_0 t)
\end{cases}
\quad (4.64)
\]

From these formulas and the explicit expression of \( F_1 \) and \( F_2 \) it is immediate to recover the normal extremal trajectories starting from the origin \((x_0 = y_0 = z_0 = 0)\) in the case \( h_0 \neq 0 \)

\[
x(t) = \frac{1}{h_0}(\sin(\theta_0 + h_0 t) - \sin(\theta_0)) \quad y(t) = \frac{1}{h_0}(\cos(\theta_0 + h_0 t) - \cos(\theta_0))
\quad (4.65)
\]

and the vertical coordinate \( z \) is computed as the integral

\[
z(t) = \frac{1}{2} \int_0^t x(t)y'(t) - y(t)x'(t) dt = \frac{1}{2h_0^2}(h_0 t - \sin(h_0 t))
\]

108
When \( h_0 = 0 \) the curve is simply a straight line
\[
x(t) = \sin(\theta_0)t \quad y(t) = \cos(\theta_0)t \quad z(t) = 0
\] (4.66)

Notice that, as we know from the results of the previous paragraph, normal extremal trajectories are curves whose projection on \( \mathbb{R}^2 = \{(x, y)\} \) has constant geodesic curvature, i.e., straight lines or circles on \( \mathbb{R}^2 \) (that correspond to horizontal lines and helix on \( M \)). There are no non trivial abnormal geodesics since \( b = 1 \).

### 4.5 Lie derivative

In this section we extend the notion of Lie derivative, already introduced for vector fields in Section 3.2, to differential forms. Recall that if \( X, Y \in \text{Vec}(M) \) are two vector fields we define
\[
\mathcal{L}_X Y = [X, Y] = \frac{d}{dt} \bigg|_{t=0} e^{-tX}Y.
\]

If \( P : M \to M \) is a diffeomorphism we can consider the pullback \( P^* : T^*_qM \to T^*_qM \) and extend its action to \( k \)-forms. Let \( \omega \in \Lambda^k M \), we define \( P^* \omega \in \Lambda^k M \) in the following way:
\[
(P^*\omega)_q(\xi_1, \ldots, \xi_k) := \omega_{P(q)}(P_*\xi_1, \ldots, P_*\xi_k), \quad q \in M, \quad \xi_i \in T_q M.
\] (4.67)

It is an easy check that this operation is linear and satisfies the two following properties
\[
P^*(\omega_1 \wedge \omega_2) = P^*\omega_1 \wedge P^*\omega_2,
\] (4.68)
\[
P^* \circ d = d \circ P^*.
\] (4.69)

**Definition 4.43.** Let \( X \in \text{Vec}(M) \) and \( \omega \in \Lambda^k M \), with \( k \geq 1 \). We define the **Lie derivative** of \( \omega \) with respect to \( X \) as
\[
\mathcal{L}_X : \Lambda^k M \to \Lambda^k M, \quad \mathcal{L}_X \omega = \frac{d}{dt} \bigg|_{t=0} (e^{tX})^* \omega.
\] (4.70)

When \( k = 0 \) this definition recovers the Lie derivative of smooth functions \( \mathcal{L}_X f = Xf \), for \( f \in C^\infty(M) \). From (1.68) and (4.69), we easily deduce the following properties of the Lie derivative:

(i) \( \mathcal{L}_X (\omega_1 \wedge \omega_2) = (\mathcal{L}_X \omega_1) \wedge \omega_2 + \omega_1 \wedge (\mathcal{L}_X \omega_2) \),

(ii) \( \mathcal{L}_X \circ d = d \circ \mathcal{L}_X \).

The first of these properties can be also expressed by saying that \( \mathcal{L}_X \) is a *derivation* of the exterior algebra of \( k \)-forms.

The Lie derivative combines together a \( k \)-form and a vector field defining a new \( k \)-form. A second way of combining these two object is to define their inner product, by defining a \( (k-1) \)-form.

**Definition 4.44.** Let \( X \in \text{Vec}(M) \) and \( \omega \in \Lambda^k M \), with \( k \geq 1 \). We define the **inner product** of \( \omega \) and \( X \) as the operator \( i_X : \Lambda^k M \to \Lambda^{k-1} M \), where we set
\[
(i_X \omega)(Y_1, \ldots, Y_{k-1}) := \omega(X, Y_1, \ldots, Y_{k-1}), \quad Y_i \in \text{Vec}(M).
\] (4.71)
One can show that the operator $i_X$ is an *anti-derivation*, in the following sense:

$$i_X(\omega_1 \wedge \omega_2) = (i_X \omega_1) \wedge \omega_2 + (-1)^{k_1} \omega_1 \wedge (i_X \omega_2), \quad \omega_i \in \Lambda^{k_i} M, \quad i = 1, 2. \quad (4.72)$$

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

**Proposition 4.45** (Cartan’s formula). *The following identity holds true*

$$\mathcal{L}_X = i_X \circ d + d \circ i_X. \quad (4.73)$$

**Proof.** Define $D_X := i_X \circ d + d \circ i_X$. It is easy to check that $D_X$ is a derivation on the algebra of $k$-forms, since $i_X$ and $d$ are anti-derivations. Let us show that $D_X$ commutes with $d$. Indeed, using that $d^2 = 0$, one gets

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

Since any $k$-form can be expressed in coordinates as $\omega = \sum \omega_{i_1...i_k} dx_{i_1} ... dx_{i_k}$, it is sufficient to prove that $\mathcal{L}_X$ coincide with $D_X$ on functions. This last property is easily checked by

$$D_X f = i_X(df) + d(i_X f) = \langle df, X \rangle = X f = \mathcal{L}_X f. \quad \square$$

**Corollary 4.46.** *Let $X, Y \in \text{Vec}(M)$ and $\omega \in \Lambda^1 M$, then*

$$d\omega(X, Y) = X \langle \omega, Y \rangle - Y \langle \omega, X \rangle - \langle \omega, [X, Y] \rangle. \quad (4.74)$$

**Proof.** On one hand Definition 4.43 implies, by Leibnitz rule

$$\langle \mathcal{L}_X \omega, Y \rangle_q = \left. \frac{d}{dt} \right|_{t=0} \langle (e^{tX})^* \omega, Y \rangle_q$$

$$= \left. \frac{d}{dt} \right|_{t=0} \langle \omega, e^{tX} Y \rangle_{e^{tX}(q)}$$

$$= X \langle \omega, Y \rangle - \langle \omega, [X, Y] \rangle.$$

On the other hand, Cartan’s formula (4.73) gives

$$\langle \mathcal{L}_X \omega, Y \rangle = \langle i_X (d\omega), Y \rangle + \langle d(i_X \omega), Y \rangle$$

$$= d\omega(X, Y) + Y \langle \omega, X \rangle.$$

Comparing the two identities one gets (4.74). \square

### 4.6 Symplectic geometry

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

**Definition 4.47.** A *symplectic manifold* $(N, \sigma)$ is a smooth manifold $N$ endowed with a closed, non degenerate 2-form $\sigma \in \Lambda^2(N)$. A *symplectomorphism* of $N$ is a diffeomorphism $\phi : N \to N$ such that $\phi^* \sigma = \sigma$. 

110
Notice that a symplectic manifold $N$ is necessarily even-dimensional. We stress that, in general, the symplectic form $\sigma$ is not exact, as in the case of $N = T^*M$.

The symplectic structure on a symplectic manifold $N$ permits us to define the Hamiltonian vector field $\vec h \in \text{Vec}(N)$ associated with a function $h \in C^\infty(N)$ by the formula $i_{\vec h}\sigma = -dh$, or equivalently $\sigma(\cdot, \vec h) = dh$.

**Proposition 4.48.** A diffeomorphism $\phi : N \to N$ is a symplectomorphism if and only if for every $h \in C^\infty(N)$:

$$ (\phi^{-1}_* \vec h) = \vec h \circ \phi. $$

**(4.75)**

**Proof.** Assume that $\phi$ is a symplectomorphism, namely $\phi^*\sigma = \sigma$. More precisely, this means that for every $\lambda \in N$ and every $v, w \in T\lambda N$ one has

$$ \sigma_\lambda(v, w) = (\phi^*\sigma)_\lambda(v, w) = \sigma_{\phi(\lambda)}(\phi_* v, \phi_* w), $$

where the second equality is the definition of $\phi^*\sigma$. If we apply the above equality at $w = \phi^{-1}_* \vec h$ one gets, for every $\lambda \in N$ and $v \in T\lambda N$

$$ \sigma_\lambda(v, \phi^{-1}_* \vec h) = (\phi^*\sigma)_\lambda(v, \phi^{-1}_* \vec h) = \sigma_{\phi(\lambda)}(\phi_* v, \vec h) $$

$$ = \langle d\phi(\lambda) h, \phi_* v \rangle = \langle \phi^* d\phi(\lambda) h, v \rangle. $$

This shows that $\sigma(\cdot, \phi^{-1}_* \vec h) = d(h \circ \phi)$, that is (4.75). The converse implication follows analogously. □

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

**Lemma 4.49.** Let $X \in \text{Vec}(N)$ be a complete vector field on a symplectic manifold $(N, \sigma)$. The following properties are equivalent

(i) $(e^{tX})^* \sigma = \sigma$ for every $t \in \mathbb{R}$,

(ii) $L_X \sigma = 0$,

(iii) $i_X \sigma$ is a closed 1-form on $N$.

**Proof.** By the group property $e^{(t+s)X} = e^{tX} \circ e^{sX}$ one has the following identity for every $t \in \mathbb{R}$:

$$ \frac{d}{dt} (e^{tX})^* \sigma \bigg|_{t=0} = (e^{tX})^* (e^{sX})^* \sigma = (e^{tX})^* L_X \sigma. $$

This proves the equivalence between (i) and (ii), since the map $(e^{tX})^*$ is invertible for every $t \in \mathbb{R}$.

Recall now that the symplectic form $\sigma$ is, by definition, a closed form. Then $d\sigma = 0$ and Cartan’s formula (4.73) reads as follows

$$ L_X \sigma = d(i_X \sigma) + i_X (d\sigma) = d(i_X \sigma), $$

which proves the the equivalence between (ii) and (iii). □
Corollary 4.50. The flow of a Hamiltonian vector field defines a flow of symplectomorphisms.

Proof. This is a direct consequence of the fact that, for an Hamiltonian vector field $\vec{h}$, one has $i_{\vec{h}}\sigma = -dh$. Hence $i_{\vec{h}}\sigma$ is a closed form (actually exact) and property (iii) of Lemma 4.49 holds.

Notice that the converse of Corollary 4.50 is true when $N$ is simply connected, since in this case every closed form is exact.

Definition 4.51. Let $(N,\sigma)$ be a symplectic manifold and $a, b \in C^\infty(N)$. The Poisson bracket between $a$ and $b$ is defined as $\{a, b\} = \sigma(\vec{a}, \vec{b})$.

We end this section by collecting some properties of the Poisson bracket that follow from the previous results.

Proposition 4.52. The Poisson bracket satisfies the identities

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

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

Proof. Property (i) follows from (4.75). Property (ii) follows by considering $\phi = e^{t\vec{c}}$ in (i), for some $c \in C^\infty(N)$, and computing the derivative with respect to $t$ at $t = 0$.

Corollary 4.53. For every $a, b \in C^\infty(N)$ we have

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

Proof. Property (ii) of Proposition 4.52 can be rewritten, by skew-symmetry of the Poisson bracket, as follows

$$\{\{a, b\}, c\} = \{a, \{b, c\}\} - \{b, \{a, c\}\}. \quad (4.77)$$

Using that $\{a, b\} = \sigma(\vec{a}, \vec{b}) = \vec{a}\vec{b}$ one rewrite (4.77) as

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

Remark 4.54. Property (ii) of Proposition 4.52 says that $\{a, \cdot\}$ is a derivation of the algebra $C^\infty(N)$. Moreover, the space $C^\infty(N)$ endowed with $\{\cdot, \cdot\}$ as a product is a Lie algebra isomorphic to a sub-algebra of $\text{Vec}(N)$. Indeed, by (4.76), the correspondence $a \mapsto \vec{a}$ is a Lie algebra homomorphism between $C^\infty(N)$ and $\text{Vec}(N)$.

4.7 Local minimality of normal trajectories

In this section we prove a fundamental result about local optimality of normal trajectories. More precisely we show small pieces of a normal trajectory are length minimizers.
4.7.1 The Poincaré-Cartan one form

Fix a smooth function $a \in C^\infty(M)$ and consider the smooth submanifold of $T^*M$ defined by the graph of its differential
\[
\mathcal{L}_0 = \{d_q a \mid q \in M\} \subset T^*M.
\] (4.78)

Notice that the restriction of the canonical projection $\pi : T^*M \to M$ to $\mathcal{L}_0$ defines a diffeomorphism between $\mathcal{L}_0$ and $M$, hence $\dim \mathcal{L}_0 = n$. Assume that the Hamiltonian flow is complete and consider the image of $\mathcal{L}_0$ under the Hamiltonian flow
\[
\mathcal{L}_t := e^{t\hat{H}}(\mathcal{L}_0), \quad t \in [0, T].
\] (4.79)

Define the $(n + 1)$-dimensional manifold with boundary in $\mathbb{R} \times T^*M$ as follows
\[
\mathcal{L} = \{(t, \lambda) \in \mathbb{R} \times T^*M \mid \lambda \in \mathcal{L}_t, 0 \leq t \leq T\}
= \{(t, e^{t\hat{H}}a_0) \in \mathbb{R} \times T^*M \mid a_0 \in \mathcal{L}_0, 0 \leq t \leq T\}.\] (4.80)

Finally, let us introduce the Poincaré-Cartan 1-form on $T^*M \times \mathbb{R} \simeq T^*(M \times \mathbb{R})$ defined by
\[
s - H dt \in \Lambda^1(T^*M \times \mathbb{R})
\]
where $s \in \Lambda^1(T^*M)$ denotes, as usual, the tautological 1-form of $T^*M$. We start by proving a preliminary lemma.

**Lemma 4.55.** $s|_{\mathcal{L}_0} = d(a \circ \pi)|_{\mathcal{L}_0}$

**Proof.** By definition of tautological 1-form $s_\lambda(w) = (\lambda, \pi_*w)$, for every $w \in T\lambda(T^*M)$. If $\lambda \in \mathcal{L}_0$ then $\lambda = d_q a$, where $q = \pi(\lambda)$. Hence for every $w \in T\lambda(T^*M)$
\[
s_\lambda(w) = (\lambda, \pi_*w) = (d_q a, \pi_*w) = (\pi^*d_q a, w) = (d_q (a \circ \pi), w). \]

**Proposition 4.56.** The 1-form $(s - H dt)|_{\mathcal{L}}$ is exact.

**Proof.** We divide the proof in two steps: (i) we show that the restriction of the Poincare-Cartan 1-form $(s - H dt)|_{\mathcal{L}}$ is closed and (ii) that it is exact.

(i). To prove that the 1-form is closed we need to show that the differential
\[
d(s - H dt) = \sigma - dH \wedge dt,
\] (4.82)
vanishes when applied to every pair of tangent vectors to $\mathcal{L}$. Since, for each $t \in [0, T]$, the set $\mathcal{L}_t$ has codimension 1 in $\mathcal{L}$, there are only two possibilities for the choice of the two tangent vectors:

(a) both vectors are tangent to $\mathcal{L}_t$, for some $t \in [0, T]$.

(b) one vector is tangent to $\mathcal{L}_t$ while the second one is transversal.

Case (a). Since both tangent vectors are tangent to $\mathcal{L}_t$, it is enough to show that the restriction of the one form $\sigma - dH \wedge dt$ to $\mathcal{L}_t$ is zero. First let us notice that $dt$ vanishes when applied to tangent vectors to $\mathcal{L}_t$, thus $\sigma - dH \wedge dt|_{\mathcal{L}_t} = \sigma|_{\mathcal{L}_t}$. Moreover, since by definition $\mathcal{L}_t = e^{t\hat{H}}(\mathcal{L}_0)$ one has
\[
\sigma|_{\mathcal{L}_t} = \sigma|_{e^{t\hat{H}}(\mathcal{L}_0)} = (e^{t\hat{H}})^*\sigma|_{\mathcal{L}_0} = \sigma|_{\mathcal{L}_0} = ds|_{\mathcal{L}_0} = d^2(a \circ \pi)|_{\mathcal{L}_0} = 0.
\]
where in the last line we used Lemma 4.55 and the fact that \((e^{t\bar{H}})^*\sigma = \sigma\), since \(e^{t\bar{H}}\) is an Hamiltonian flow and thus preserves the symplectic form.

Case (b). The manifold \(\mathcal{L}\) is, by construction, the image of the smooth mapping

\[
\Psi : [0, T] \times \mathcal{L}_0 \to [0, T] \times T^*M, \quad \Psi(t, \lambda) \mapsto (t, e^{t\bar{H}}\lambda),
\]

Thus a tangent vector to \(\mathcal{L}\) that is transversal to \(\mathcal{L}\) can be obtained by differentiating the map \(\Psi\) with respect to \(t\):

\[
\frac{\partial \Psi}{\partial t}(t, \lambda) = \frac{\partial}{\partial t} + \bar{H}(\lambda) \in T_{(t, \lambda)}\mathcal{L}.
\]

(4.83)

It is then sufficient to show that the vector (4.83) is in the kernel of the two form \(\sigma - dH \wedge dt\). In other words we have to prove

\[
i_{\partial_t + \bar{H}}(\sigma - dH \wedge dt) = 0.
\]

(4.84)

The last equality is a consequence of the following identities

\[
i_{\bar{H}}\sigma = \sigma(\bar{H}, \cdot) = -dH, \quad i_{\partial_t}\sigma = 0,
\]

\[
i_{\bar{H}}(dH \wedge dt) = (i_{\bar{H}}dH) \wedge dt - dH \wedge (i_{\bar{H}}dt) = 0,
\]

\[
i_{\partial_t}(dH \wedge dt) = (i_{\partial_t}dH) \wedge dt - dH \wedge (i_{\partial_t}dt) = -dH.
\]

where we used that \(i_{\bar{H}}dH = dH(\bar{H}) = \{H, H\} = 0\).

(ii). Next we show that the form \(s - H dt\mid_{\mathcal{L}}\) is exact. To this aim we have to prove that for every closed curve \(\Gamma\) in \(\mathcal{L}\) one has

\[
\int_{\Gamma} s - H dt = 0.
\]

(4.85)

Every curve \(\Gamma\) in \(\mathcal{L}\) can be written as follows

\[
\Gamma : [0, T] \to \mathcal{L}, \quad \Gamma(s) = (t(s), e^{t(s)\bar{H}}\lambda(s)), \quad \text{where } \lambda(s) \in \mathcal{L}_0.
\]

Moreover, it is easy to see that the continuous map defined by

\[
K : [0, T] \times \mathcal{L} \to \mathcal{L}, \quad K(\tau, (t, e^{t\bar{H}}\lambda_0)) = (t - \tau, e^{(t-\tau)\bar{H}}\lambda_0)
\]

defines an homotopy of \(\mathcal{L}\) such that \(K(0, (t, e^{t\bar{H}}\lambda_0)) = (t, e^{t\bar{H}}\lambda_0)\) and \(K(t, (t, e^{t\bar{H}}\lambda_0)) = (0, \lambda_0)\). Then the curve \(\Gamma\) is homotopic to the curve \(\Gamma_0(s) = (0, \lambda(s))\). Since the 1-form \(s - H dt\) is closed, the integral is invariant under homotopy, namely

\[
\int_{\Gamma} s - H dt = \int_{\Gamma_0} s - H dt.
\]

Moreover, the integral over \(\Gamma_0\) is computed as follows (recall that \(\Gamma_0 \subset \mathcal{L}_0\) and \(dt = 0\) on \(\mathcal{L}_0\)):

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

where we used Lemma 4.55 and the fact that the integral of an exact form over a closed curve is zero. Then (4.85) follows.
4.7.2 Normal trajectories are geodesics

Now we are ready to prove a sufficient condition that ensures the optimality of small pieces of normal trajectories. As a corollary we will get that small pieces of normal trajectories are geodesics.

Recall that normal trajectories for the problem

\[ \dot{q} = f_u(q) = \sum_{i=1}^{m} u_i f_i(q), \]  

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

\[ \dot{\lambda}(t) = \tilde{H}(\lambda(t)), \quad (\text{i.e.} \, \lambda(t) = e^{t\tilde{H}}(\lambda_0)), \]  

\[ \gamma(t) = \pi(\lambda(t)), \quad t \in [0, T]. \]  

where

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

Recall that, given a smooth function \( a \in C^\infty(M) \), we can consider the image of its differential \( L_0 \) and its evolution \( L_t \) under the Hamiltonian flow associated to \( H \) as in (4.78) and (4.79).

**Theorem 4.57.** Assume that there exists \( a \in C^\infty(M) \) such that the restriction of the projection \( \pi|_{L_t} \) is a diffeomorphism for every \( t \in [0, T] \). Then for any \( \lambda_0 \in L_0 \) the normal geodesic

\[ \gamma(t) = \pi \circ e^{t\tilde{H}}(\lambda_0), \quad t \in [0, T], \]  

is a strict length-minimizer among all admissible curves \( \gamma \) with the same boundary conditions.

**Proof.** Let \( \gamma(t) \) be an admissible trajectory, different from \( \gamma(t) \), associated with the control \( u(t) \) and such that \( \gamma(0) = \gamma(0) \) and \( \gamma(T) = \gamma(T) \). We denote by \( \gamma(t) \) the control associated with the curve \( \gamma(t) \).

By assumption, for every \( t \in [0, T] \) the map \( \pi|_{L_t} : L_t \to M \) is a local diffeomorphism, thus the trajectory \( \gamma(t) \) can be uniquely lifted to a smooth curve \( \lambda(t) \in L_t \). Notice that the corresponding curves \( \Gamma \) and \( \Gamma \) in \( L \) defined by

\[ \Gamma(t) = (t, \lambda(t)), \quad \Gamma(t) = (t, \lambda(t)) \]  

have the same boundary conditions, since for \( t = 0 \) and \( t = T \) they project to the same base point on \( M \) and their lift is uniquely determined by the diffeomorphisms \( \pi|_{L_0} \) and \( \pi|_{L_T} \), respectively.

Recall now that, by definition of the sub-Riemannian Hamiltonian, we have

\[ H(\lambda(t)) = \langle \lambda(t), f_u(t)(\gamma(t)) \rangle - \frac{1}{2} |u(t)|^2, \quad \gamma(t) = \pi(\lambda(t)), \]  

where \( \lambda(t) \) is a lift of the trajectory \( \gamma(t) \) associated with a control \( u(t) \). Moreover, the equality holds in (4.92) if and only if \( \lambda(t) \) is a solution of the Hamiltonian system \( \dot{\lambda}(t) = H(\lambda(t)) \). For this reason we have the relations

\[ H(\lambda(t)) < \langle \lambda(t), f_u(t)(\gamma(t)) \rangle - \frac{1}{2} |u(t)|^2, \]  

\[ H(\gamma(t)) = \langle \gamma(t), f_u(t)(\gamma(t)) \rangle - \frac{1}{2} |\gamma(t)|^2. \]
since \( \bar{\lambda}(t) \) is a solution of the Hamiltonian equation by assumptions, while \( \lambda(t) \) is not. Indeed \( \lambda(t) \) and \( \bar{\lambda}(t) \) have the same initial condition, hence, by uniqueness of the solution of the Cauchy problem, it follows that \( \dot{\lambda}(t) = H(\lambda(t)) \) if and only if \( \lambda(t) = \bar{\lambda}(t) \), that implies that \( \bar{\gamma}(t) = \gamma(t) \).

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

\[
\frac{1}{2} \int_0^T |u(t)|^2 dt = \int_T s - Hdt = \int_T s - Hdt < \frac{1}{2} \int_0^T |u(t)|^2 dt, \tag{4.95}
\]

where \( \Gamma \) and \( \bar{\Gamma} \) are the curves in \( \mathcal{L} \) defined in (4.91).

By Lemma 4.56, the 1-form \( s - Hdt \) is exact. Then the integral over the closed curve \( \Gamma \cup \bar{\Gamma} \) vanishes, and one gets

\[
\int_T s - Hdt = \int_T s - Hdt.
\]

The last inequality in (4.95) can be proved as follows

\[
\int_T s - Hdt = \int_0^T \langle \lambda(t), \dot{\gamma}(t) \rangle - H(\lambda(t)) dt \\
= \int_0^T \langle \lambda(t), f_u(t)(\gamma(t)) \rangle - H(\lambda(t)) dt \\
< \int_0^T \langle \lambda(t), f_u(t)(\gamma(t)) \rangle - \left( \langle \lambda(t), f_u(t)(\gamma(t)) \rangle - \frac{1}{2} |u(t)|^2 \right) dt \\
= \frac{1}{2} \int_0^T |u(t)|^2 dt, \tag{4.96}
\]

where we used (4.93). A similar computation gives computation, using (4.94), gives

\[
\int_T s - Hdt = \frac{1}{2} \int_0^T |\pi(t)|^2 dt, \tag{4.97}
\]

that ends the proof of (4.95).

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

**Corollary 4.58.** Assume that there exists \( a \in C^\infty(M) \) and neighborhoods \( \Omega_t \) of \( \bar{\gamma}(t) \), such that \( \pi \circ e^{tH} \circ da|_{\Omega_0} : \Omega_0 \to \Omega_t \) is a diffeomorphism for every \( t \in [0,T] \). Then (4.90) is a strict length-minimizer among all admissible trajectories \( \gamma \) with same boundary conditions and such that \( \gamma(t) \in \Omega_t \) for all \( t \in [0,T] \).

We are in position to prove that small pieces of normal trajectories are global length minimizers.

**Theorem 4.59.** Let \( \gamma : [0,T] \to M \) be a sub-Riemannian normal trajectory. Then for every \( \tau \in [0,T] \) there exists \( \varepsilon > 0 \) such that

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

(ii) \( \gamma|_{[\tau,\tau+\varepsilon]} \) is the unique length minimizer joining \( \gamma(\tau) \) and \( \gamma(\tau + \varepsilon) \), up to reparametrization.
Proof. Without loss of generality we can assume that the curve is parametrized by length and prove the theorem for $\tau = 0$. Let $\gamma(t)$ be a normal extremal trajectory, such that $\gamma(t) = \pi(e^{tH}(\lambda_0))$, for $t \in [0, T]$. Consider a smooth function $a \in C^\infty(M)$ such that $d_q a = \lambda_0$ and let $L_t$ be the family of submanifold of $T^*M$ associated with this function by (4.78) and (4.79). By construction, for the extremal lift associated with $\gamma$ one has $\lambda(t) = e^{tH}(\lambda_0) \in L_t$ for all $t$. Moreover the projection $\pi|_{L_0}$ is a diffeomorphism, since $L_0$ is a section of $T^*M$.

Hence, for every fixed compact $K \subset M$ containing the curve $\gamma$, by continuity there exists $t_0 = t_0(K)$ such that the restriction on $K$ of the map $\pi|_{L_t}$ is also a diffeomorphism, for all $0 \leq t < t_0$. Let us now denote $\delta_K$ the positive constant defined in Lemma 3.34 such that every curve starting from $\gamma(0)$ and leaving $K$ is necessary longer than $\delta_K$.

Then, defining $\varepsilon(K) := \min\{\delta_K, t_0(K)\}$ we have that the curve $\gamma|_{[0, \varepsilon]}$ is contained in $K$ and is shorter than any other curve contained in $K$ with the same boundary condition by Corollary 4.58 (applied to $\Omega_t = K$ for all $t \in [0, T]$). Moreover $\ell(\gamma|_{[0, \varepsilon]}) = \varepsilon$ since $\gamma$ is length parametrized, hence it is shorter than any admissible curve that is not contained in $K$. Thus $\gamma|_{[0, \varepsilon]}$ is a global minimizer. Moreover it is unique up to reparametrization by uniqueness of the solution of the Hamiltonian equation (see proof of Theorem 4.57) \hfill \Box

Remark 4.60. When $D_{q_0} = T_{q_0} M$, as it is the case for a Riemannian structure, the level set of the Hamiltonian

$$\{H = 1/2\} = \{\lambda \in T^*_q M | H(\lambda) = 1/2\},$$

is diffeomorphic to an ellipsoid, hence compact. Under this assumption, for each $\lambda_0 \in \{H = 1/2\}$, the corresponding geodesic $\gamma(t) = \pi(e^{tH}(\lambda_0))$ is optimal up to a time $\varepsilon = \varepsilon(\lambda_0)$, with $\lambda_0$ belonging to a compact set. It follows that it is possible to find a common $\varepsilon > 0$ (depending only on $q_0$) such that each normal trajectory with base point $q_0$ is optimal on the interval $[0, \varepsilon]$. It can be proved that this is false as soon as $D_{q_0} \neq T_{q_0} M$. Indeed in this case, for every $\varepsilon > 0$ there exists a normal extremal path that lose optimality in time $\varepsilon$, see Theorem 12.17.

Bibliographical notes

The Hamiltonian approach to sub-Riemannian geometry is nowadays classical. However the construction of the symplectic structure, obtained by extending the Poisson bracket from the space of affine functions, is not standard and is inspired by [?].

Historically, in the setting of PDE, the sub-Riemannian distance (also called Carnot-Carathéodory distance) is introduced by means of sub-unit curves, see for instance [13] and references therein. The link between the two definition is clarified in Exercise 13.11.

The proof that normal extremal are geodesics is an adaptation of a more general condition for optimality given in [3] for a more general class of problems. This is inspired by the classical idea of “fields of extremals” in classical Calculus of Variation.
Chapter 5

Integrable Systems

In this chapter we present some applications of the Hamiltonian formalism developed in the previous chapter. In particular we give a proof the well-known Arnold-Liouville’s Theorem and, as an application, we study the complete integrability of the geodesic flow on a special class of Riemannian manifolds.

5.1 Completely integrable systems

Let $M$ be an $n$-dimensional smooth manifold and assume that there exist $n$ independent Hamiltonians in involution in $T^*M$, i.e. a set of $n$ smooth functions

$$h_i : T^*M \to \mathbb{R}, \quad i = 1, \ldots, n,$$

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

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

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

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

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

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

$$T_\lambda h^{-1}(c) = \text{span}\{\vec{h}_1(\lambda), \ldots, \vec{h}_n(\lambda)\}, \quad \forall \lambda \in h^{-1}(c).$$

**Proof.** Since $c$ is a regular value of $h$, by Remark [2.51] the set $h^{-1}(c)$ is a submanifold of dimension $n$ in $T^*M$. In particular $\dim T_\lambda h^{-1}(c) = n$. Moreover, by Exercise [2.11] each vector field $\vec{h}_i$ is tangent to $h^{-1}(c)$, since $\vec{h}_i h_j = \{h_i, h_j\} = 0$ by assumption. To prove (5.2) it is then enough to show that these vector fields are linearly independent.

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

$$d_\lambda h_1 \wedge \ldots \wedge d_\lambda h_n \neq 0, \quad \forall \lambda \in h^{-1}(c).$$
Moreover the symplectic form $\sigma$ on $T^*M$ induces for all $\lambda$ an isomorphism $T_\lambda(T^*M) \to T^*_\lambda(T^*M)$ defined by $w \mapsto \sigma_\lambda(\cdot, w)$. By nondegeneracy of the symplectic form, this implies that the vectors $\vec{h}_1(\lambda), \ldots, \vec{h}_n(\lambda)$ are linearly independent, hence they form a basis for $T_\lambda h^{-1}(c)$. \qed

Remark 5.3. Notice that the symplectic form vanishes on $T_\lambda h^{-1}(c)$. Indeed this is a consequence of the fact that $\sigma(\vec{h}_i, \vec{h}_j) = h_i, h_j = 0$ for all $i, j = 1, \ldots, n$.

In what follows we denote by $N_c = h^{-1}(c)$ the level set of $h$. If $h^{-1}(c)$ is not connected, $N_c$ will denote a connected component of $h^{-1}(c)$.

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

$$
\Psi : \mathbb{R}^n \to \text{Diff}(N_c), \quad \Psi(s_1, \ldots, s_n) := \left. e^{s_1 \vec{h}_1} \circ \cdots \circ e^{s_n \vec{h}_n} \right|_{N_c}.
$$

(5.4)

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

Proof. The complete integrability assumption together with Corollary 4.53 implies that the flows of $\vec{h}_i$ and $\vec{h}_j$ commute for every $i, j = 1, \ldots, n$ since $[\vec{h}_i, \vec{h}_j] = 0$.

By Proposition 2.26 this is equivalent to

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

(5.5)

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

$$
\Psi(s + s') = e^{(s_1 + s'_1) \vec{h}_1} \circ \cdots \circ e^{(s_n + s'_n) \vec{h}_n}
= e^{s_1 \vec{h}_1} \circ \cdots \circ e^{s_n \vec{h}_n} \circ e^{s'_1 \vec{h}_1} \circ \cdots \circ e^{s'_n \vec{h}_n}
= \Psi(s) \circ \Psi(s')
$$

(by (5.5))

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

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

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

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

are linearly independent on the level set $N_c$. As a consequence the stabilizer $S_\lambda$ of the point $\lambda$, i.e. the set

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

is a discrete subgroup of $\mathbb{R}^n$. Then the proof of Proposition 5.4 is completed by the next lemma.
Lemma 5.5. Let $G$ be a non trivial discrete subgroup of $\mathbb{R}^n$. Then there exist $k \in \mathbb{N}$ with $1 \leq k \leq n$ and $e_1, \ldots, e_k \in \mathbb{R}^n$ such that

$$G = \left\{ \sum_{i=1}^{k} m_i e_i, \ m_i \in \mathbb{Z} \right\}.$$

Proof. We prove the claim by induction on the dimension $n$ of the ambient space $\mathbb{R}^n$.

(i). Let $n = 1$. Since $G$ is a discrete subgroup of $\mathbb{R}$, then there exists an element $e_1 \neq 0$ closest to the origin $0 \in \mathbb{R}$. We claim that $G = Ze_1 = \{me_1, m \in \mathbb{Z}\}$. By contradiction assume that there exists an element $f \in G$ such that $me_1 < f < (m+1)e_1$ for some $m \in \mathbb{Z}$. Then $\bar{f} := f - me_1$ belong to $G$ and is closer to the origin with respect to $e_1$, that is a contradiction.

(ii). Assume the statement is true for $n-1$ and let us prove it for $n$. The discreteness of $G$ guarantees the existence of an element $e_1 \in G$, closest to the origin. Moreover one can prove that $G_1 := G \cap \mathbb{R}e_1$ is a subgroup and, as in part (i) of the proof, that

$$G_1 := G \cap \mathbb{R}e_1 = Ze_1.$$ 

If $G = G_1$ then the theorem is proved with $k = 1$. Otherwise one can consider the quotient $G/G_1$.

Exercise 5.6. (i). Prove that there exists a nonzero element $e_2 \in G/G_1$ that minimize the distance to the line $\ell = \mathbb{R}e_1$ in $\mathbb{R}^n$.

(ii). Show that there exists a neighborhood of the line $\ell$ that does not contain elements of $G/G_1$.

By Exercise 5.6 the quotient group $G/G_1$ is a discrete subgroup in $\mathbb{R}^n/\ell \simeq \mathbb{R}^{n-1}$. Hence, by the induction step there exists $e_2, \ldots, e_k$ such that

$$G/G_1 = \left\{ \sum_{i=2}^{k} m_i e_i, \ m_i \in \mathbb{Z} \right\}. \quad \Box$$

From Proposition 5.4 and the fact that $T^k \times \mathbb{R}^{n-k}$ is compact if and only if $k = n$ we have the following corollary.

Corollary 5.7. If $N_c$ is compact, then $N_c \simeq T^n$.

Remark 5.8. On any level set $\lambda \in N_c$ the map $\Psi_{\lambda} : \mathbb{R}^n \to N_c$ defined by $\Psi_{\lambda}(s) = \Psi(s)\lambda$ defines coordinates $(s_1, \ldots, s_n)$ in a neighborhood of the point $\lambda$. In these coordinate set (defined on $N_c$) the Hamiltonian vector fields $\vec{h}_i$ are constant.

5.2 Arnold-Liouville theorem

In this section we consider the moment map of a completely integrable system

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

and we assume that for all values of $c \in \mathbb{R}$ the level set $h^{-1}(c)$ is a smooth compact and connected manifold. In particular $N_c \simeq T^n$ for all $c \in \mathbb{R}$ by Corollary 5.7.

Fix $c \in \mathbb{R}$ and a point $\lambda_c \in N_c$. Let us consider the basis $e_1, \ldots, e_n$ in $\mathbb{R}^n$ given by Lemma 5.5 and denote by $(\theta_1, \ldots, \theta_n)$ the coordinates defined in $\mathbb{R}^n$ by the choice of this basis.
Since $\theta_1, \ldots, \theta_n$ are obtained by $(s_1, \ldots, s_n)$ by a linear change of coordinates on each level set, the vector fields $\vec{h}_i$ are constant in these coordinates (see Remark 5.8) and the basis $\partial_{\theta_1}, \ldots, \partial_{\theta_n}$ can be expressed as follows

$$\partial_{\theta_i} = \sum_{j=1}^{n} b_{ij}(c) \vec{h}_j,$$

where the coefficients $b_{ij}$ depend only on $c$, i.e., are constant on each level $N_c$.

**Remark 5.9.** Notice that the coordinate set $(\theta_1, \ldots, \theta_n)$ are not uniquely defined. Indeed every transformation of the kind $\theta_i \mapsto \theta_i + \psi_i(c)$ still defines a set of angular coordinates on each level set. The choice of the functions $\psi_i(c)$ corresponds to the choice of the initial value of $\theta_i$ at a point (for every choice of $c$).

Notice that the vector fields $\partial_{\theta_i}$ are well defined and independent on this choice.

Let us now introduce the diffeomorphism

$$F_c : T^n \to N_c, \quad F_c(\theta_1, \ldots, \theta_n) = \Psi(\theta_1 + 2\pi \mathbb{Z}, \ldots, \theta_n + 2\pi \mathbb{Z})(\lambda_c).$$

Next we want to analyze the dependence of this construction with respect to $c$. Fix $\bar{c} \in \mathbb{R}^n$ and consider a neighborhood $\mathcal{O}$ of the submanifold $N_{\bar{c}}$ in the cotangent space $T^*M$. Being $N_{\bar{c}}$ compact, in $\mathcal{O}$ we have a foliation of invariant tori $N_c$, for $c$ close to $\bar{c}$. In other words we have a well defined coordinate set $(c_1, \ldots, c_n, \theta_1, \ldots, \theta_n)$.

**Theorem 5.10** (Arnold-Liouville). Let us consider a moment map $h : T^*M \to \mathbb{R}^n$ associated with a completely integrable system such that every level set $N_c$ is compact and connected. Then for every $\bar{c} \in \mathbb{R}$ there exists a neighborhood $\mathcal{O}$ of $N_{\bar{c}}$ and a change of coordinates

$$(c_1, \ldots, c_n, \theta_1, \ldots, \theta_n) \mapsto (I_1, \ldots, I_n, \varphi_1, \ldots, \varphi_n) \quad (5.7)$$

such that

(i) $I = \Phi \circ h$, where $\Phi : h(\mathcal{O}) \to \mathbb{R}^n$ is a diffeomorphism,

(ii) $\sigma = \sum_{j=1}^{n} dI_j \wedge d\varphi_j$.

**Definition 5.11.** The coordinates $(I, \varphi)$ defined in Theorem 5.10 are called action-angle coordinates.

**Remark 5.12.** This proves that there exists a regular foliation of the phase space by invariant manifolds, that are actually tori, such that the Hamiltonian vector fields associated to the invariants of the foliation span the tangent distribution. There then exist, as mentioned above, special sets of canonical coordinates on the phase space such that the invariant tori are the level sets of the action variables, and the angle variables are the natural periodic coordinates on the torus. The motion on the invariant tori, expressed in terms of these canonical coordinates, is linear in the angle variables.

Indeed, since the $h_j$ are functions on $I$ variables only, we have

$$\vec{h}_j = \sum_{i=1}^{n} \frac{\partial h_j}{\partial I_i} \partial_{\varphi_i}. $$
In other words, the Hamiltonian system in the angle-action coordinate \((I, \varphi)\) is written as follows

\[
\dot{I}_i = -\frac{\partial h}{\partial \varphi_i} = 0, \quad \dot{\varphi}_i = \frac{\partial h}{\partial I_i}(I) . \tag{5.8}
\]

This explains also why this property is called complete integrability.

**Proof of Theorem 5.10.** In this proof we will use the following notation:

- if \(c = (c_1, \ldots, c_n) \in \mathbb{R}^n\) we set \(c^j, \varepsilon = (c_1, \ldots, c_j + \varepsilon, \ldots, c_n)\),
- \(\gamma_i(c)\) is the closed curve in the torus \(N_c\) parametrized by the \(i\)-th angular coordinate \(\theta_i\), namely
  \[
  \gamma_i(c) = \{ F_c(\theta_1, \ldots, \theta_i + \tau, \ldots, \theta_n) \in N_c \mid \tau \in [0, 2\pi] \} .
  \]
- \(C_j^i, \varepsilon\) denotes the cylinder defined by the union of curves \(\gamma_i(c^j, \tau)\), for \(0 \leq \tau \leq \varepsilon\).

Let us first define the coordinates \(I_i = I_i(c_1, \ldots, c_n)\) by the formula

\[
I_i(c) = \frac{1}{2\pi} \int_{\gamma_i(c)} s ,
\]

where \(s\) is the tautological 1-form on \(T^*M\). Being \(\sigma|_{N_c} \equiv 0\), by Stokes Theorem the variable \(I_i\) depends only on the homotopy class of \(\gamma_i\)\(^1\).

Let us compute the Jacobian of the change of variables.

\[
\frac{\partial I_i}{\partial c_j}(c) = \frac{1}{2\pi} \frac{\partial}{\partial \varepsilon} \bigg|_{\varepsilon=0} \left( \int_{\gamma_i(c^j, \varepsilon)} s - \int_{\gamma_i(c)} s \right) = \frac{1}{2\pi} \frac{\partial}{\partial \varepsilon} \bigg|_{\varepsilon=0} \int_{C_j^i, \varepsilon} s = \frac{1}{2\pi} \frac{\partial}{\partial \varepsilon} \bigg|_{\varepsilon=0} \int_{C_j^i, \varepsilon} \sigma \quad \text{(where } \sigma = ds)\]

\[
= \frac{1}{2\pi} \frac{\partial}{\partial \varepsilon} \bigg|_{\varepsilon=0} \int_{c_j}^{c_j + \varepsilon} \int_{\gamma_i(c^j, \tau)} \sigma(\partial_{c_j}, \partial_{\theta_i}) d\theta_i d\tau = \frac{1}{2\pi} \int_{\gamma_i(c)} \sigma(\partial_{c_j}, \partial_{\theta_i}) d\theta_i .
\]

Using that \(\partial_{\theta_i} = \sum_{j=1}^n b_{ij}(c)\hat{h}_j\) (see (5.6)) one gets

\[
\sigma(\cdot, \partial_{\theta_i}) = \sum_{j=1}^n b_{ij}(c)\hat{h}_j . \tag{5.9}
\]

\(^1\)Hence, in principle, we are free to choose any basis \(\gamma_1, \ldots, \gamma_n\) for the first homotopy group of \(T^n\).
Moreover, $dh_i = dc_i$ since they define the same coordinate set. Hence

$$\frac{\partial I_i}{\partial c_j}(c) = \frac{1}{2\pi} \int_{\gamma_i(c)} \left( \sum_{k=1}^{n} b_{ik} dc_k, \partial_{c_i} \right) d\theta_i$$

$$= \frac{1}{2\pi} \int_{\gamma_i(c)} b_{ij}(c) d\theta_i$$

$$= b_{ij}(c)$$

Combining the last identity with (5.9) one gets

$$\sigma(\cdot, \partial \theta_i) = dI_i$$

In particular, this implies that the symplectic form has the following expression in the coordinates $(I, \theta)$

$$\sigma = \sum_{ij} a_{ij}(I) dI_i \wedge dI_j + \sum_i dI_i \wedge d\theta_i. \tag{5.10}$$

where the smooth functions $a_{ij}$ depends only on the action variables, since the symplectic form $\sigma$ and the term $\sum_i dI_i \wedge d\theta_i$ are closed form. Moreover it is easy to see that the first term of (5.10) can be rewritten as

$$\sum_{i,j=1}^{n} a_{ij}(I) dI_i \wedge dI_j = d \left( \sum_{i=1}^{n} \beta_i(I) \right) \wedge dI_i$$

and $\sigma$ can be rewritten as

$$\sigma = \sum_{i=1}^{n} dI_i \wedge d(\theta_i - \beta_i(I))$$

The proof is completed by defining $\varphi_i := \theta_i - \beta_i(I)$.

Remark 5.13. The notion of complete integrability introduced here is the classical one given by Liouville and Arnold. Sometimes, complete integrability of a dynamical system is also referred to systems whose solution can be reduced to a sequence of quadratures. Notice that by Theorem 5.10, complete integrability implies integrability by quadratures (see also Remark 5.12).

5.3 Integrable geodesic flows

In this section we want to discuss whether it is possible to apply the Arnold-Liouville’s Theorem to the case of a geodesic flow on a Riemannian (or sub-Riemannian) manifold.

Recall that on a sub-Riemannian manifold, we denote by $H$ the sub-Riemannian Hamiltonian.

Definition 5.14. We say that a complete smooth vector field $X \in \text{Vec}(M)$ is a Killing vector field if it generates a one parametric flow of isometries, i.e. $e^{tX} : M \rightarrow M$ is an isometry for all $t \in \mathbb{R}$.

Recall that, for every $X \in \text{Vec}(M)$, we can define the function $h_X \in C^\infty(T^*M)$ linear on fibers associated with $X$ by $h_X(\lambda) = \langle \lambda, X(q) \rangle$, where $q = \pi(\lambda)$.

The following lemma shows that, if $X$ is a Killing vector field, i.e. a vector field on $M$ whose flow generates isometries, then the Hamiltonian associated with it is in involution with the sub-Riemannian Hamiltonian.
Lemma 5.15. Let $M$ be a sub-Riemannian manifold and $H$ the sub-Riemannian Hamiltonian. For a vector field $X \in \text{Vec}(M)$ is a Killing vector field if and only if $\{H, h_X\} = 0$.

Proof. A vector field $X$ generates isometries if and only if, by definition, the differential of its flow $e^{sX} : T_qM \to T_{e^{sX}(q)}M$ preserves the sub-Riemannian distribution and the norm on it, i.e.

$$e^{sX} v \in D_{e^{sX}(q)}$$

for every $v \in D_q$ and $\|e^{sX} v\| = \|v\|$. By definition of $H$, this is equivalent to the identity

$$H(e^{sX} \lambda) = H(\lambda), \quad \forall \lambda \in T^*M.$$

On the other hand Proposition 4.9 implies that $(e^{tX})^* = e^{th_X}$, where $h_X$ is the hamiltonian linear on fibers related to $X$. Hence differentiating with respect to $t$ we find the equivalence

$$H \circ e^{tX} = H \iff \vec{h}_X H = 0 \iff \{H, h_X\} = 0.$$

In other words to every 1-parametric group of isometries of $M$ we can associate an Hamiltonian in involution with $H$. Let us show the complete integrability of the geodesic flow in some very symmetric cases.

Example 5.16 (Revolution surfaces in $\mathbb{R}^3$). Let $M$ be a 2-dimensional revolution surface in $\mathbb{R}^3$. Since the rotation around the revolution axis preserves the Riemannian structure, by definition, we have that the Hamiltonian generated by this flow and the Riemannian Hamiltonian $H$ are in involution. As a consequence the geodesic flow is completely integrable.

Example 5.17 (Isoperimetric sub-Riemannian problem). Let us consider a sub-Riemannian structure associated with an isoperimetric problem defined on a 2-dimensional revolution surface $M$ (see Section 4.4.2). The sub-Riemannian structure on $M \times \mathbb{R}$ is determined by the function $b \in C^\infty(M)$ satisfying $dA = bdV$, where $A \in \Lambda^1(M)$ is the 1-form defining the isoperimetric problem and $dV$ is the volume form on $M$.

(i) If both $M$ and $b$ are rotational invariant we find a first integral of the geodesic flow as in the previous example

(ii) By construction the problem is invariant by translation along the $z$-axis

Hence there exists three Hamiltonian in involution and the geodesic flow is completely integrable.

5.3.1 Geodesic flow

Let us consider now a smooth function $a : \mathbb{R}^n \to \mathbb{R}$ and consider the family of hypersurfaces defined by the level sets of $a$

$$M_c := a^{-1}(c) \subset \mathbb{R}^n, \quad c \text{ is a regular value of } a,$$

endowed with the Riemannian structure induced by the ambient space $\mathbb{R}^n$. By Sard’s Lemma for almost every $c \in \mathbb{R}$, $c$ is a regular value for $a$ (in particular, $M_c$ is a smooth submanifold of codimension one in $\mathbb{R}^n$).

Adapting the arguments of Proposition 1.4 in Chapter 1 one can prove the following characterization of geodesics on a hypersurface $M$.
Proposition 5.18. Let \( \gamma : [0, 1] \rightarrow M \) a length-parametrized curve on \( M \). Then \( \gamma \) is a geodesic if and only if \( \ddot{\gamma}(t) \perp T_{\gamma(t)}M \).

For a large class of functions \( a \), we will find an Hamiltonian, defined on the ambient space \( T^*\mathbb{R}^n \), whose (reparametrized) flow generates the geodesic flow when restricted to each level set \( M_c \).

Consider the standard symplectic structure on \( T^*\mathbb{R}^n \)

\[
T^*\mathbb{R}^n = \mathbb{R}^n \times \mathbb{R}^n = \{(x, p), x, p \in \mathbb{R}^n\}, \quad \sigma = \sum_{i=1}^{n} dp_i \wedge dx_i,
\]

For \( x, p \in \mathbb{R}^n \) we will denote by \( x + \mathbb{R}p \) the line of \( \mathbb{R}^n \) \{\( x + tp, t \in \mathbb{R} \)\}.

**Assumptions.** In what follows we assume that the function \( a : \mathbb{R}^n \rightarrow \mathbb{R} \) satisfies the following assumptions:

(i) the restriction of \( a : \mathbb{R}^n \rightarrow \mathbb{R} \) to every line is strictly convex,

(ii) \( a(x) \rightarrow +\infty \) when \( |x| \rightarrow +\infty \).

Under these assumptions the restriction of the function \( a \) to each affine line in \( \mathbb{R}^n \) always attains a minimum and we can define the function

\[
h(x, p) = \min_{t \in \mathbb{R}} a(x + tp). \tag{5.11}
\]

**Remark 5.19.** Given \( x, p \in \mathbb{R}^n \) the line \( x + \mathbb{R}p \) is tangent to the level set \( a^{-1}(c) \) (with \( c = a(x + \bar{t}p) \)) at the point \( \xi = x + tp \in \mathbb{R}^n \) at which the minimum in (5.11) is attained. Indeed

\[
0 = \left. \frac{d}{dt} \right|_{t=\bar{t}} a(x + tp) = \langle d\xi a, p \rangle.
\]

It is clear from the definition of \( h \) that actually it is a well-defined function on the space of affine lines in \( \mathbb{R}^n \). This is formally proved in the following lemma.

**Lemma 5.20.** The Hamiltonian \( b(x, p) = \frac{1}{2}|p|^2 \) satisfies \( \{h, b\} = 0 \), i.e. \( h \) it is constant along the flow of \( \vec{b} \).

**Proof.** The Hamiltonian system for \( \vec{b} \) is easily solved for every initial condition \( (x(0), p(0)) = (x_0, p_0) \)

\[
\begin{align*}
\dot{x} &= \frac{\partial b}{\partial p} = p \\
\dot{p} &= -\frac{\partial b}{\partial x} = 0
\end{align*}
\Rightarrow \begin{align*}
x &= x_0 + tp_0 \\
p &= p_0
\end{align*} \tag{5.12}
\]

and it is easy to see that, by its very definition, \( h \) is constant under this flow. \( \square \)

**Remark 5.21.** Notice that to restrict to a level set of \( b \) is equivalent to restrict the function \( h \) to the space of affine lines in \( \mathbb{R}^n \) since

\[
\{(x, p) \in T^*\mathbb{R}^n, b(x, p) = 1/2\} = \{(x, p) \in T^*\mathbb{R}^n, |p| = 1\}.
\]
Now we introduce the following function

\[ \xi : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n, \quad \xi(x, p) = x + s(x, p)p, \quad (5.13) \]

where \( s(x, p) = \ell \) is the point at which the function \( f(t) = a(x + tp) \) attains its minimum.

The following proposition says that if we follow the flow of \( \vec{h} \), as a flow on the space of lines, then the line is always tangent to the same quadric and actually describes a geodesic on it.

**Proposition 5.22.** Let \( (x(t), p(t)) \) be a trajectory of the Hamiltonian vector field \( \vec{h} \) associated to \( (5.11) \). Then the function

\[ t \mapsto \xi(t) := \xi(x(t), p(t)) \in \mathbb{R}^n, \quad (5.14) \]

(i) is contained in a level set \( M_c = a^{-1}(c) \), for some \( c \in \mathbb{R} \),

(ii) is a geodesic on \( M_c \).

**Proof.** Property (i) is a simple consequence of Corollary 4.19, since every function is constant along the flow of its Hamiltonian vector field. Indeed, writing \( h(x, p) = a(\xi(x, p)) \) and denoting by \( (x(t), p(t)) \) the Hamiltonian flow, we get

\[ a(\xi(t)) = a(\xi(x(t), p(t))) = h(x(t), p(t)) = \text{const}, \]

i.e. the curve \( \xi(t) \) is contained on a level set of \( a \). Moreover by definition \( s(x, p) \) denotes on the line \( x + \mathbb{R}p \) where \( a \) attains its minimum, hence

\[ \langle \nabla_{\xi(t)}a, p(t) \rangle = 0, \quad \forall t. \quad (5.15) \]

The Hamiltonian system associated with \( h \) reads

\[
\begin{cases}
\dot{x} = s\nabla_{\xi}a \\
\dot{p} = -\nabla_{\xi}a
\end{cases}
\quad (5.16)
\]

that immediately implies \( \dot{x} + s\dot{p} = 0 \). Computing the derivative

\[ \dot{\xi} = \dot{x} + s\dot{p} + sp = \dot{s}p, \]

it follows that \( \dot{\xi} \) is parallel to \( p \), and actually \( p(t) \) is the velocity of the curve \( \xi(t) \), when reparametrized with the parameter \( s \), since \( |p| = 1 \) implies \( |\dot{\xi}| = \dot{s} \).

Finally, the second derivative of the reparametrized of \( \xi \) is \( \dot{p} \) and, since \( \dot{p} \wedge \nabla_{\xi}a = 0 \) from the Hamiltonian system, the second derivative of \( \xi(t) \) (when reparametrized by the length) is orthogonal to the level set, i.e. \( \xi(t) \) is a geodesic.

Notice also that \( s \) is a well defined parameter. Computing the derivative with respect to \( t \) in \( (5.15) \) we have that

\[ \dot{s}\langle \nabla_{\xi}^2a p, p \rangle - |\nabla_{\xi}a|^2 = 0. \]

and the strict convexity of \( a \) implies \( \langle \nabla_{\xi}^2a p, p \rangle \neq 0 \). □

**Remark 5.23.** Thus we can visualize the solutions of \( \vec{h} \) as a motion of lines: the lines move in such a way to be tangent to one and the same geodesic. The tangency point \( x \) on the line moves perpendicular to this line in this process. We will also refer to this flow as the “line flow” associated with \( a \).
Consider now two functions $a, b : \mathbb{R}^n \to \mathbb{R}$ that satisfies our assumptions (i), (ii). Following our notation, we set
\[
 h(x, p) = a(\xi(x, p)), \quad \xi(x, p) = x + s(x, p)p
\]
\[
 g(x, p) = b(\eta(x, p)), \quad \eta(x, p) = x + t(x, p)p
\]
where $s(x, p)$ and $t(x, p)$ are defined as above, and $\xi, \eta$ denote the tangency point of the line $x + \mathbb{R}p$ with the level set of $a$ and $b$ respectively. The following proposition computes the Poisson bracket of these Hamiltonian functions

**Proposition 5.24.** **Under the previous assumptions**
\[
 \{h, g\} = (s - t) \langle \nabla_\xi a, \nabla_\eta b \rangle. \tag{5.17}
\]

**Proof.** From the very definition of Poisson bracket
\[
 \{h, g\} = \langle \nabla_p h, \nabla_x g \rangle - \langle \nabla_x h, \nabla_p g \rangle
\]
\[
 = (s - t) \langle \nabla_\xi a, \nabla_\eta b \rangle.
\]
where we used equations (5.16) for both $h$ and $g$. \hfill \Box

## 5.4 Geodesic flow on ellipsoids

It was Jacobi who first established that the geodesic flow on an ellipsoid is completely integrable, using the separation of variables method. Here we give a different derivation, essentially due to Moser, as an application of the theory developed in the previous section. More precisely we consider the particular case when the function $a$ is a quadratic polynomial, i.e. every level set of our function is a quadric in $\mathbb{R}^n$.

**Definition 5.25.** Let $A$ be an $n \times n$ non degenerate symmetrix matrix. The *quadric* $Q$ associated to $A$ is the set
\[
 Q = \{x \in \mathbb{R}^n, \langle A^{-1}x, x \rangle = 1\}. \tag{5.18}
\]

For simplicity we consider the case when $A$ has simple distinct eigenvalues $\alpha_1 < \ldots < \alpha_n$. Define, for every $\lambda$ that is not an eigenvalue of $A$,
\[
 a_{\lambda}(x) = \langle (A - \lambda I)^{-1}x, x \rangle, \quad Q_\lambda = \{x \in \mathbb{R}^n, a_{\lambda}(x) = 1\}.
\]
If $A = \text{diag}(\alpha_1, \ldots, \alpha_n)$ is a diagonal matrix then (5.18) reads
\[
 Q = \{x \in \mathbb{R}^n, \sum_{i=1}^{n} \frac{x_i^2}{\alpha_i} = 1\},
\]
and $Q_\lambda$ represents the family quadrics that are confocal to $Q$
\[
 Q_\lambda = \{x \in \mathbb{R}^n, \sum_{i=1}^{n} \frac{x_i^2}{\alpha_i - \lambda} = 1\}, \quad \forall \lambda \in \mathbb{R} \setminus \Lambda,
\]
where $\Lambda = \{\alpha_1, \ldots, \alpha_n\}$ denotes the set of eigenvalues of $A$. Note that $Q_\lambda = \emptyset$ when $\lambda > \alpha_n$.

**Note.** In what follows by a “generic” point $x$ for $A$ we mean a point $x$ that does not belong to any proper invariant subspace of $A$. In the diagonal case it is equivalent to say that $x = (x_1, \ldots, x_n)$, with $x_i \neq 0$ for every $i$. 

128
Exercise 5.26. Denote by $A_\lambda := (A - \lambda I)^{-1}$. Prove the two following formulas:

(i) $\frac{d}{d\lambda} A_\lambda = A_\lambda^2$,

(ii) $A_\lambda - A_\mu = (\mu - \lambda)A_\lambda A_\mu$.

Lemma 5.27. Let $x \in \mathbb{R}^n$ be a generic point for $A$ and let $\{Q_\lambda\}_{\lambda \in \Lambda}$ be the family of confocal quadrics. Then there exists exactly $n$ distinct real numbers $\lambda_1, \ldots, \lambda_n$ in $\mathbb{R} \setminus \Lambda$ such that $x \in Q_{\lambda_i}$ for every $i = 1, \ldots, n$, and the quadrics $Q_{\lambda_i}$ are pairwise orthogonal at the point $x$.

Proof. For a fixed $x$, the function $\lambda \mapsto a_\lambda(x) = \langle A_\lambda x, x \rangle$ satisfies in $\mathbb{R} \setminus \Lambda$

$$\frac{\partial a_\lambda}{\partial \lambda}(x) = \langle A_\lambda^2 x, x \rangle = |A_\lambda x|^2 \geq 0,$$

as follows from part (i) of Exercise 5.26 and the fact that $A$ (hence $A_\lambda$) is self-adjoint. Thus $a_\lambda(x)$ is monotone increasing as a function of $\lambda$, and takes values from $-\infty$ to $+\infty$ in each interval $]a_i, a_{i+1}[$ contained between two eigenvalues of $A$. This implies that, for a fixed $x$, there exist exactly $n$ values $\lambda_1, \ldots, \lambda_n$ such that $a_{\lambda_i}(x) = 1$ (that means $x \in Q_{\lambda_i}$). Next, using part (ii) of Exercise 5.26 (also known as resolvent formula) we can compute, for two distinct values $\lambda_i \neq \lambda_j$ and $x \in Q_{\lambda_i} \cap Q_{\lambda_j}$:

$$\langle \nabla_x a_{\lambda_i}, \nabla_x a_{\lambda_j} \rangle = 4 \langle A_{\lambda_i} x, A_{\lambda_i} x \rangle$$

$$= 4 \langle A_{\lambda_i} A_{\lambda_j} x, x \rangle$$

$$= \frac{4}{\lambda_j - \lambda_i} \langle \langle A_{\lambda_i} x, x \rangle - \langle A_{\lambda_j} x, x \rangle \rangle = 0,$$

where again we used the fact that $A_\lambda$ is self-adjoint and $\langle A_\lambda x, x \rangle = 1$ for all $\lambda$. \hfill \square

Now we define the family of Hamiltonians associated with the family of confocal quadrics

$$h_\lambda(x, p) = \min_t a_\lambda(x + tp) = a_\lambda(\xi_\lambda(x, p)),$$

(5.19)

Now we prove another interesting “orthogonality” property of the family. We show that if two confocal quadrics are tangent to the same line, then their gradient are orthogonal at the tangency points.

Proposition 5.28. Assume that two confocal quadrics are tangent to a given line, i.e. there exist $x, y \in \mathbb{R}^n$ such that

$$a_\lambda(\xi_\lambda) = a_\mu(\xi_\mu), \quad \text{where} \quad \xi_\lambda = x + t_\lambda y, \quad \xi_\mu = x + t_\mu y.$$

Then $\langle \nabla_\xi a_\lambda, \nabla_\xi a_\mu \rangle = 0$. In particular $\{h_\lambda, h_\mu\} = 0$.

Proof. The condition that the quadric $Q_\lambda$ is tangent to the line $x + \mathbb{R}y$ at $\xi_\lambda$ is expressed by the following two equality

$$\langle A_\lambda \xi_\lambda, y \rangle = 0, \quad \langle A_\lambda \xi_\lambda, \xi_\lambda \rangle = 1$$

(5.20)

and an analogue relations is valid for $Q_\mu$. Notice than from (5.20) one also gets $\langle A_\lambda \xi_\lambda, \xi_\mu \rangle = \langle A_\mu \xi_\mu, \xi_\lambda \rangle = 1$. Then, with the same computation as before using (5.26)

$$\langle \nabla_\xi a_\lambda, \nabla_\xi a_\mu \rangle = 4 \langle A_\lambda \xi_\lambda, A_\mu \xi_\mu \rangle$$

$$= 4 \langle A_\lambda A_\mu \xi_\lambda, \xi_\mu \rangle$$

$$= \frac{4}{\mu - \lambda} (\langle A_\lambda \xi_\lambda, \xi_\mu \rangle - \langle A_\mu \xi_\mu, \xi_\lambda \rangle) = 0,$$

The last claim follows from Proposition 5.24. \hfill \square
**Proposition 5.29.** A generic line in \( \mathbb{R}^n \) is tangent to \( n - 1 \) quadrics of a confocal family.

**Proof.** Consider the projection along the fixed line \( x + \mathbb{R}p \) of the quadrics of the confocal family onto an orthogonal hyperplane. The following exercise shows that this projection defines a confocal family of quadrics on the reduced space.

**Exercise 5.30.** (i). Show that the map \( x \mapsto a^p_\lambda(x) := \langle A_\lambda(x + t_\lambda p), x + t_\lambda p \rangle \) is a quadratic form and that \( p \in \text{Ker} \ a^p_\lambda \). In particular this implies that \( a^p_\lambda \) is well defined on the quotient \( \mathbb{R}^n / \mathbb{R}p \).

(ii). Prove that \( \{a^p_\lambda\}_\lambda \) is a family of confocal quadric on the factor space (in \( n - 1 \) variables).

Applying then Lemma 5.27 to the family \( \{a^p_\lambda\}_\lambda \) we get that, for a generic choice of \( x \), there exists \( n - 1 \) quadrics passing through the point on the plane where the line is projected, i.e. the line \( x + \mathbb{R}p \) is tangent to \( n - 1 \) confocal quadrics of the family \( \{a_\lambda\}_\lambda \).

**Remark 5.31.** Notice that this proves that every generic line in \( \mathbb{R}^n \) is associated with an orthonormal frame of \( \mathbb{R}^n \), being all the normal vectors to the \( n - 1 \) quadrics given by Proposition 5.29 mutually orthogonal and orthogonal to the line itself.

**Theorem 5.32.** The geodesic flow on an ellipsoid is completely integrable. In particular, the tangents of any geodesics on an ellipsoid are tangent to the same set of its confocal quadrics, i.e. independently on the point on the geodesic.

**Proof.** We want to show that the functions \( \lambda_1(x, p), \ldots, \lambda_{n-1}(x, p) \) (as functions defined on the set of lines in \( \mathbb{R}^n \)) that assign to each line \( x + \mathbb{R}p \) in \( \mathbb{R}^n \) the \( n - 1 \) values of \( \lambda \) such that the line is tangent to \( Q_\lambda \) are independent and in involution.

First notice that each level set \( \lambda_i(x, p) = c \) coincide with the level set \( h_c = 1 \). Hence, by Exercise 4.30 the two functions defines the same Hamiltonian flow on this level set (up to reparametrization). We are then reduced to prove that the functions \( h_{c_1}, \ldots, h_{c_{n-1}} \) are independent and in involution, which is a consequence of Proposition 5.28.

Since the lines that are tangent to a geodesic on the ellipsoid \( Q_\lambda \) form an integral curve of the Hamiltonian flow of the associated function \( h_\lambda \), and all the Poisson brackets with the other Hamiltonians are zero, it follows that the line remains tangent to the same set of \( n - 1 \) quadrics. \( \square \)
Chapter 6

Chronological calculus

In this chapter we develop some tools from chronological calculus that will allow us to manage in a very efficient way with flows of nonautonomous vector fields.

The main idea is to replace a nonlinear object defined on the manifold $M$ with its linear counterpart, when interpreted as an operator on the space $C^\infty(M)$ of smooth functions on $M$.

6.1 Duality

We recall that the set $C^\infty(M)$ of smooth functions on $M$ is an $\mathbb{R}$-algebra with the usual operation of pointwise addition and multiplication

$$
(a + b)(q) = a(q) + b(q),
$$

$$
(\lambda a)(q) = \lambda a(q), \quad a, b \in C^\infty(M), \ \lambda \in \mathbb{R},
$$

$$
(a \cdot b)(q) = a(q)b(q).
$$

Any point $q \in M$ can be interpreted as the linear functional

$$
\hat{q} : C^\infty(M) \to \mathbb{R}, \quad \hat{q}(a) := a(q).
$$

For every $q \in M$, the functional $\hat{q}$ is a homomorphism of algebras, i.e. it satisfies

$$
\hat{q}(a \cdot b) = \hat{q}(a)\hat{q}(b).
$$

A diffeomorphism $P \in \text{Diff}(M)$ can be thought as the linear “change of variables” operator

$$
\hat{P} : C^\infty(M) \to C^\infty(M), \quad \hat{P}(a) := a(P(q)).
$$

which is an automorphism of the algebra $C^\infty(M)$.

Remark 6.1. Notice that every nontrivial homomorphism of algebras $\varphi : C^\infty(M) \to \mathbb{R}$ is represented by some point, i.e., $\varphi = \hat{q}$ for some $q \in M$. Moreover for every automorphism of algebras $\Phi : C^\infty(M) \to C^\infty(M)$ there exists a diffeomorphism $P \in \text{Diff}(M)$ such that $\hat{P} = \Phi$. For a proof of these facts one can see [3, Appendix A].
Next we want to characterize tangent vectors as functionals on $C^\infty(M)$. As explained in Chapter 2, a tangent vector $v \in T_qM$ defines in a natural way the derivation in the direction of $v$, i.e. the functional
\[
\widehat{v} : C^\infty(M) \to \mathbb{R}, \quad \widehat{v}(a) = \langle d_qa, v \rangle,
\]
that satisfies the Leibnitz rule
\[
\widehat{v}(a \cdot b) = \widehat{v}(a)b(q) + a(q)\widehat{v}(b), \quad \forall a, b \in C^\infty(M).
\]
If $v \in T_qM$ is the tangent vector of a curve $q(t)$ such that $q(0) = q$, it is also natural to check the identity as operators
\[
\widehat{v} = \left. \frac{d}{dt} \right|_{t=0} \widehat{q}(t) : C^\infty(M) \to \mathbb{R}.
\]
Indeed, it is sufficient to differentiate at $t = 0$ the following identity
\[
\widehat{q}(t)(a \cdot b) = \widehat{q}(t) a \cdot \widehat{q}(t)b.
\]
In the same spirit, a vector field $X \in \text{Vec}(M)$ is characterized, as a derivation of $C^\infty(M)$ (cf. also the discussion in Chapter 2), as the infinitesimal version of a flow (i.e., family of diffeomorphisms) $P_t \in \text{Diff}(M)$. Indeed if we set
\[
\widehat{X} = \left. \frac{d}{dt} \right|_{t=0} \widehat{P}_t : C^\infty(M) \to C^\infty(M),
\]
we find that $\widehat{X}$ satisfies (see (2.14))
\[
\widehat{X}(ab) = \widehat{X}(a)b + a\widehat{X}(b), \quad \forall a, b \in C^\infty(M).
\]

Remark 6.2. It is possible to define on $C^\infty(M)$ the Whitney topology and define regularity properties of family of functionals in a weak sense: we say that a family of operators $A_t$ is continuous (differentiable, etc.) if the map $t \mapsto A_t a$ has the same property for every $a \in C^\infty(M)$. For instance, if $X_t$ denotes some locally integrable family of vector fields we denote
\[
\int_0^t X_s ds : a \mapsto \int_0^t X_s a ds
\]
For a more detailed presentation see [3].

6.2 Operator ODE and Volterra expansion

Consider a nonautonomous vector field $X_t$ and the corresponding nonautonomous ODE
\[
\frac{d}{dt} q(t) = X_t(q(t)), \quad q \in M.
\]
Using the notation introduced in the previous section we can rewrite (6.2) in the following way

\[
\frac{d}{dt} \hat{q}(t) = \hat{q}(t) \circ \hat{X}_t.
\] (6.3)

Indeed assume that \( q(t) \) satisfies (6.2) and let \( a \in C^\infty(M) \). We compute

\[
\left( \frac{d}{dt} \hat{q}(t) \right) a = \frac{d}{dt} \hat{q}(t) a = \frac{d}{dt} a(q(t)) = \langle d_{q(t)}a, X_t(q(t)) \rangle = (\hat{X}_t a)(q(t)) = (\hat{q}(t) \circ \hat{X}_t) a
\] (6.4)

As discussed in Chapter 2, the solution to the nonautonomous ODE (6.2) defines a flow, i.e., family of diffeomorphisms, \( P_{s,t} \). We call \( P_{s,t} \) the right chronological exponential and use the notation

\[
P_{s,t} := \exp \int_s^t X_\tau d\tau.
\] (6.5)

Sometimes it is useful to set the initial time \( s = 0 \). In this case we use the short notation \( P_t := P_{0,t} \).

**Lemma 6.3.** The flow \( P_t \) defined by (6.5) satisfies the differential equation

\[
\frac{d}{dt} \hat{P}_t = \hat{P}_t \circ \hat{X}_t, \quad \hat{P}_0 = \text{Id}.
\] (6.6)

**Proof.** Fix a point \( q_0 \in M \) and denote by \( q(t) \) the solution of the Cauchy problem (6.2) with initial condition \( q(0) = q_0 \). By the very definition of \( P_t \) we have that \( q(t) = P_t(q_0) \), which easily implies \( \hat{q}(t) = \hat{q}_0 \circ \hat{P}_t \).

\( \square \)

**Remark 6.4.** In the following we will identify any object with its dual interpretation as operator on functions and stop to use a different notation for the same object when acting on the space of smooth functions. The meaning of the notation will be clear from the context. Notice that there is no risk of confusion since, when using operatorial notation, composition works in the opposite side.

### 6.2.1 Volterra expansion

The operator differential equation

\[
\begin{cases}
\hat{P}_t = P_t \circ X_t \\
P_0 = \text{Id}
\end{cases}
\] (6.7)

can be rewritten as an integral equation as follows

\[
P_t = \text{Id} + \int_0^t P_s \circ X_s ds
\] (6.8)
Substituting into (6.8), and iterating we have

\[ P_t = \text{Id} + \int_0^t \left( \text{Id} + \int_0^{s_1} P_{s_2} \circ X_{s_2} ds_2 \right) \circ X_{s_1} ds_1 \]

\[ = \text{Id} + \int_0^t X_s ds + \int \int_{0 \leq s_2 \leq s_1 \leq t} P_{s_2} \circ X_{s_2} \circ X_{s_1} ds_1 ds_2 \]

\[ = \ldots \]

\[ = \text{Id} + \sum_{k=1}^N \int \cdots \int X_{s_k} \circ \cdots \circ X_{s_1} d^k s + R_N \]

where

\[ R_N = \int \cdots \int_{0 \leq s_N \leq \ldots \leq s_1 \leq t} P_{s_N} \circ X_{s_N} \circ \cdots \circ X_{s_1} d^N s \]

Formally, letting \( N \to \infty \) and assuming that \( R_N \to 0 \), we can write the chronological series

\[ \exp \int_0^t X_s ds = \text{Id} + \sum_{k=1}^\infty \int \cdots \int_{S_k(t)} X_{s_k} \circ \cdots \circ X_{s_1} d^k s \]

(6.9)

where \( S_k(t) = \{(s_1, \ldots, s_k) \in \mathbb{R}^k | 0 \leq s_k \leq \ldots \leq s_1 \leq t\} \) denotes the \( k \)-dimensional simplex.

A detailed discussion about the convergence of the series is contained in Section 6.5.

**Remark 6.5.** If we write expansion (6.9) when \( X_t = X \) is an autonomous vector field, we find that the chronological exponential coincides with the exponential of the vector field

\[ \overrightarrow{\exp} \int_0^t X_s ds = \text{Id} + \sum_{k=1}^\infty \int_{S_k(t)} X_{s_k} \circ \cdots \circ X_{s_1} d^k s \]

since \( \text{vol}(S_k(t)) = t^k/k! \). In the nonautonomous case for different time \( X_{s_1} \) and \( X_{s_2} \) might not commute, hence the order in which the vector fields appears in the composition is very important. The arrow in the notation recalls in which “direction” the parameters are increasing.

**Exercise 6.6.** Prove that in general, for a nonautonomous vector field \( X_t \), one has

\[ \overrightarrow{\exp} \int_0^t X_s ds \neq e^{\int_0^t X_s ds} \]

(6.10)

Prove that if \([X_t, X_\tau] = 0\) for all \( t, \tau \in \mathbb{R} \) then the equality holds in (6.10).

Assume now that \( P_t \) satisfies (6.8) and consider the inverse flow \( Q_t := P_t^{-1} \). Let us characterize the differential equation satisfied by \( Q_t \). First, by differentiating the identity

\[ P_t \circ Q_t = \text{Id}, \]

(6.11)
and using the Leibnitz rule one gets to
\[ \dot{P}_t \circ Q_t + P_t \circ \dot{Q}_t = 0. \]
Using (6.7) then we get
\[ P_t \circ X_t \circ Q_t + P_t \circ \dot{Q}_t = 0 \]
hence we get, multiplying \( Q_t \) on the right both sides, that \( Q_t \) satisfies the equation
\[
\begin{cases}
\dot{Q}_t = -X_t \circ Q_t, \\
Q_0 = \text{Id}.
\end{cases}
\tag{6.12}
\]
The solution to the problem (6.12) will be denoted by the left chronological exponential
\[ Q_t := \leftarrow \exp \int_0^t (-X_s) ds. \tag{6.13} \]
Repeating analogous reasoning, we find the formal expansion
\[ \leftarrow \exp \int_0^t (-X_s) ds = \text{Id} + \sum_{k=1}^{\infty} \int \cdots \int_{0 \leq s_k \leq \cdots \leq s_1 \leq t} (-X_{s_1}) \circ \cdots \circ (-X_{s_k}) d^k s. \]
The difference with respect to the right chronological exponential is in the order of composition. In particular the arrow over the exp says in which direction the time increases.
We can summarize properties of the chronological exponential into the following
\[
\begin{align*}
\frac{d}{dt} \exp \int_0^t X_s ds &= \exp \int_0^t X_s ds \circ X_t, \\
\frac{d}{dt} \leftarrow \exp \int_0^t X_s ds &= X_t \circ \leftarrow \exp \int_0^t X_s ds, \\
\left( \exp \int_0^t X_s ds \right)^{-1} &= \exp \int_0^t (-X_s) ds. \tag{6.16}
\end{align*}
\]
6.2.2 Adjoint representation
Now we can study the action of diffeomorphisms on vectors and vector fields. Let \( v \in T_q M \) and \( P \in \text{Diff}(M) \). We claim that, as functionals on \( C^\infty(M) \), we have
\[ P_* v = v \circ P. \]
Indeed consider a curve \( q(t) \) such that \( \dot{q}(0) = v \) and compute
\[ (P_* v)a = \frac{d}{dt} \bigg|_{t=0} a(P(q(t))) = \left( \frac{d}{dt} \bigg|_{t=0} q(t) \right) \circ Pa = v \circ Pa \]
Recall that, if \( X \in \text{Vec}(M) \) is a vector field we have \( P_* X\big|_q = P_*(X|_{P^{-1}(q)}) \). In a similar way we will find an expression for \( P_* X \) as derivation of \( C^\infty(M) \)
\[ P_* X = P^{-1} \circ X \circ P. \tag{6.17} \]
Remark 6.7. We can reinterpret the pushforward of a vector field in a totally algebraic way in the space of linear operator on $C^\infty(M)$. Indeed

$$P_*X = (\text{Ad } P^{-1})X,$$

where

$$\text{Ad } P : X \mapsto P \circ X \circ P^{-1}, \quad \forall X \in \text{Vec}(M)$$

is the adjoint action of $P$ on the space of vector fields $2$.

Assume now that $P_t = \exp \int_0^t X_s ds$. We try to characterize the flow $\text{Ad } P_t$ by looking for the ODE it satisfies. Applying to a vector field $Y$ we have

$$\left( \frac{d}{dt} \text{Ad } P_t \right) Y = \frac{d}{dt} (\text{Ad } P_t)Y = \frac{d}{dt} (P_t \circ Y \circ P^{-1}_t)$$

$$= P_t \circ X_t \circ Y \circ P^{-1}_t + P_t \circ Y \circ (-X_t) \circ P^{-1}_t$$

$$= P_t \circ (X_t \circ Y - Y \circ X_t) \circ P^{-1}_t$$

$$= (\text{Ad } P_t)[X_t, Y]$$

$$= (\text{Ad } P_t)(\text{ad } X_t)Y$$

where

$$\text{ad } X : Y \mapsto [X, Y],$$

is the adjoint action on the Lie algebra of vector fields.

In other words we proved that $\text{Ad } P_t$ is a solution to the differential equation

$$\dot{A}_t = A_t \circ \text{ad } X_t, \quad A_0 = \text{Id}.$$ 

Thus it can be expressed as chronological exponential and we have the identity

$$\text{Ad } \left( \exp \int_0^t X_s ds \right) = \exp \int_0^t \text{ad } X_s ds.$$  \hspace{1cm} (6.18)

Exercise 6.8. Prove that, if $[X_t, Y] = 0$ for all $t$, then $(\text{Ad } P_t)Y = Y$.

Remark 6.9. More explicitly we can write the following formula

$$(\text{Ad } P_t)Y = Y + \sum_{k=1}^{\infty} \int_{0 \leq s_k \leq \ldots \leq s_1 \leq t} [X_{s_n}, \ldots, [X_{s_2}, [X_{s_1}, Y]]] d^k s,$$  \hspace{1cm} (6.19)

which generalizes the formula (??). Indeed if $P_t = e^{tX}$ is the flow associated to an autonomous vector field we get

$$(\text{Ad } e^{tX})Y = e^{-tX}Y = Y + \sum_{k=1}^{\infty} \frac{t^k}{k!} [X, \ldots, [X, Y]]$$

$$= Y + t[X, Y] + \frac{t^2}{2} [X, [X, Y]] + o(t^2)$$

$^2$this is the differential of the conjugation $Q \mapsto P \circ Q \circ P^{-1}$, $Q \in \text{Diff}(M)$
Exercise 6.10. Prove the following using operator notation:

1. Show that $\text{ad}$ is the infinitesimal version of the operator $\text{Ad}$, i.e. if $P_t$ is a flow generated by the vector field $X \in \text{Vec}(M)$ then
   \[ \text{ad} X = \frac{d}{dt} \big|_{t=0} \text{Ad} P_t. \]

2. Show that, if $P \in \text{Diff}(M)$, then $P_*$ preserves Lie brackets, i.e. $P_* [X, Y] = [P_* X, P_* Y]$.

3. Show that the Jacobi identity in $\text{Vec}(M)$ is the infinitesimal version of the identity proved in 2.
   (Hint. use $P_t = e^{tZ}$)

Exercise 6.11. Prove the following change of variables formula for a nonautonomous flow

\[ P \circ \exp \int_0^t X_s ds \circ P^{-1} = \exp \int_0^t (\text{Ad} P) X_s ds. \]  \hspace{1cm} (6.20)

Notice that for an autonomous vector field this identity reduces to (2.23).

6.3 Variations Formulae

Consider the following ODE

\[ \dot{q} = X_t(q) + Y_t(q) \]  \hspace{1cm} (6.21)

where $Y_t$ is thought as a perturbation of our original equation (6.2). We want to describe the solution to the perturbed equation (6.21) as the perturbation of the solution of the original one.

Proposition 6.12. Let $X_t, Y_t$ be two nonautonomous vector fields. Then

\[ \exp \int_0^t (X_s + Y_s) ds = \exp \int_0^t \left( \exp \int_0^s \text{ad} X_{\tau} d\tau \right) Y_s ds \circ \exp \int_0^t X_s ds \]  \hspace{1cm} (6.22)

\[ = \exp \int_0^t (\text{Ad} P_t) Y_s ds \circ P_t \]  \hspace{1cm} (6.23)

where $P_t = \exp \int_0^t X_s ds$ denotes the flow of the original vector field.

Proof. Our goal is to find a flow $R_t$ such that

\[ Q_t := \exp \int_0^t (X_s + Y_s) ds = R_t \circ P_t \]  \hspace{1cm} (6.24)

By definition of right chronological exponential we have

\[ \dot{Q}_t = Q_t \circ (X_t + Y_t) \]  \hspace{1cm} (6.25)

On the other hand, from (6.24), we also have

\[ \dot{Q}_t = \dot{R}_t \circ P_t + R_t \circ \dot{P}_t \]
\[ = \dot{R}_t \circ P_t + R_t \circ P_t \circ X_t \]
\[ = \dot{R}_t \circ P_t + Q_t \circ X_t \]  \hspace{1cm} (6.26)
Comparing (6.25) and (6.26), one gets

\[
Q_t \circ Y_t = \dot{R}_t \circ P_t
\]

and the ODE satisfied by \( R_t \) is

\[
\dot{R}_t = Q_t \circ Y_t \circ P_t^{-1} = R_t \circ (\text{Ad } P_t) Y_t
\]

Since \( R_0 = \text{Id} \) we find that \( R_t \) is a chronological exponential and

\[
\exp \int_0^t (X_s + Y_s) ds = \exp \int_0^t (\text{Ad } P_s) Y_s ds \circ P_t
\]

which is (6.23). Plugging (6.18) in (6.23) one gets (6.22). \( \square \)

**Exercise 6.13.** Prove the following versions of the variation formula:

(i) For every non autonomous vector fields \( X_t, Y_t \) on \( M \)

\[
\exp \int_0^t (X_s + Y_s) ds = \exp \int_0^t X_s ds \circ \exp \int_0^t \left( \exp \int_0^t \text{ad } X_\tau d\tau \right) Y_s ds
\]

(ii) For every autonomous vector fields \( X, Y \in \text{Vec}(M) \) prove that

\[
e^{t(X+Y)} = \exp \int_0^t e^{s \text{ad } X} Y ds \circ e^{tX} = \exp \int_0^t e^{-sX} Y ds \circ e^{tX}
\]

\[
e^{tX} \circ \exp \int_0^t e^{(s-t) \text{ad } X} Y ds
\]

### 6.4 Whitney topology on smooth functions

We introduce the Whitney topology on the space \( C^\infty(M) \). Denote by \( X_1, \ldots, X_N \) a family of vector fields such that

\[
\text{span}\{X_1, \ldots, X_N\}|_q = T_qM, \quad \forall q \in M.
\]

For \( s \in \mathbb{N} \) and \( K \subset M \) compact, define the following seminorm of a function \( f \in C^\infty(M) \)

\[
\|f\|_{s,K} = \sup_{q \in K} \{|(X_{i_1} \circ \cdots \circ X_{i_\ell} f)(q)| : 1 \leq i_j \leq N, 0 \leq \ell \leq s\}
\]

The family of seminorms \( \| \cdot \|_{s,K} \) induces a topology on \( C^\infty(M) \) as follows: take a family of compact sets \( \{K_n\}_{n \in \mathbb{N}} \) such that \( K_n \subset K_{n+1} \subset M \) for every \( n \in \mathbb{N} \) and \( M = \bigcup_{n \in \mathbb{N}} K_n \). For every \( f \in C^\infty(M) \), a local base of neighborhood of \( f \) in this topology is given by

\[
U_{f,n} := \left\{ g \in C^\infty(M) : \|f - g\|_{n,K_n} \leq \frac{1}{n} \right\}, \quad n \in \mathbb{N}.
\]

**Example 6.14.** Prove that the topology does not depend on the family of vector fields \( X_1, \ldots, X_N \) generating the tangent space to \( M \) and on the family of compact sets \( \{K_n\}_{n \in \mathbb{N}} \) invading \( M \).
This topology turns $C^\infty(M)$ into a Fréchet space, i.e., a complete, metrizable, locally convex topological vector space. More details about this topology, and the topology of the space $C^k(M,N)$ of $C^k$ maps among two smooth manifolds $M$ and $N$, can be found, for instance, in [19].

**Example 6.15.** Prove that, given a diffeomorphism $P \in \text{Diff}(M)$ and $s \in \mathbb{N}$, there exists a constant $C_{s,P} > 0$ such that for all $f \in C^\infty(M)$ one has

$$\|Pf\|_{s,K} \leq C_{s,P} \|f\|_{s,P(K)}, \quad \forall K \subset M.$$  

In other words the diffeomorphism $P$, when interpreted as a linear operator on $C^\infty(M)$, is continuous in the Whitney topology.

Given a vector field $X$ on $M$, we define its seminorms as follows

$$\|X\|_{s,K} = \sup\{\|Xf\|_{s,K} : \|f\|_{s+1,K} \leq 1\}, \quad \forall K \subset M.$$  

Convergence of functions, norm of vector fields and diffeo.

### 6.5 Estimates of the Volterra series

In this section we discuss the convergence of the Volterra series

$$\text{Id} + \sum_{k=1}^\infty \int \cdots \int_{S_k(t)} X_{s_k} \circ \cdots \circ X_{s_1} d^k s$$  \hspace{1cm} (6.30)

where $S_k(t) = \{(s_1, \ldots, s_k) \in \mathbb{R}^k | 0 \leq s_k \leq \ldots \leq s_1 \leq t\}$ denotes the $k$-dimensional symplex. Recall that if $X = X$ is autonomous then the series (6.30) simplifies in

$$\sum_{k=0}^\infty \frac{t^k}{k!} X^k$$  \hspace{1cm} (6.31)

We prove the following result, saying that in general, if the vector field is not zero, the chronological exponential is never convergent on the whole space $C^\infty(M)$.

**Proposition 6.16.** Let $X$ be a nonzero smooth vector field. Then there exists $a \in C^\infty(M)$ such that the Volterra series

$$\sum_{k=0}^\infty \frac{t^k}{k!} X^k a$$  \hspace{1cm} (6.32)

is not convergent at some point $q \in M$.

**Proof.** Fix a point $q \in M$ such that $X(q) \neq 0$ and consider a smooth coordinate chart around $q$ such that $X$ is rectified in this chart. We are then reduced to study the case when $X = \partial_{x_1}$ in $\mathbb{R}^n$. Fix a sequence $(c_n)_{n \in \mathbb{N}}$ and let $f : I \to \mathbb{R}$ defined in a neighborhood $I$ of 0 such that $f^{(n)}(0) = c_n$, for every $n \in \mathbb{N}$. The existence of such a function is guaranteed by Lemma. Then define $a(x) = f(x_1)$, where $x = (x_1, x') \in \mathbb{R}^n$. In this case $X^k a(q) = \partial_{x_1}^k f(0) = c_k$ and

$$\sum_{k=0}^\infty \frac{t^k}{k!} X^k a \big|_{q} = \sum_{k=0}^\infty \frac{t^k}{k!} c_k$$  \hspace{1cm} (6.33)

which is not convergent for a suitable choice of the sequence $(a_n)$.
Lemma 6.17 (Borel lemma). Let \((c_n)_{n \in \mathbb{N}}\) be a sequence of real numbers. Then there exist a \(C^\infty\) function \(f : I \to \mathbb{R}\) defined in a neighborhood \(I\) of \(0\) such that \(f^{(n)}(0) = c_n\), for every \(n \in \mathbb{N}\).

**Proof.** Fix a \(C^\infty\) function \(\phi : \mathbb{R} \to \mathbb{R}\) with compact support and such that \(\phi(0) = 1\) and \(\phi^j(0) = 0\) for every \(j \geq 1\). Then set

\[
g_k(x) := \frac{c_k}{k!} x^k \phi \left( \frac{x}{\varepsilon_k} \right) \tag{6.34}
\]

Notice that \(g_k^{(j)}(0) = \delta_{jk} c_k\), where \(\delta_{jk}\) is the Kronecker symbol, and \(|g_k^{(j)}(x)| \leq C_{j,k} \varepsilon_k^{-j}\) for every \(x \in \mathbb{R}\), for some constant \(C_{j,k} > 0\). Then choose \(\varepsilon_k > 0\) in such a way that

\[
|g_k^{(j)}(x)| \leq 2^{-j}, \quad \forall j \leq k - 1, \forall x \in \mathbb{R}, \tag{6.35}
\]

and define the function

\[
f(x) := \sum_{k=0}^{\infty} g_k(x).
\]

The series converges uniformly with all the derivatives by (6.35) and, by differentiating under the sum one obtains

\[
f^{(j)}(x) := \sum_{k=0}^{\infty} g_k^{(j)}(x), \quad f^{(j)}(0) := \sum_{k=0}^{\infty} g_k^{(j)}(0) = a_j
\]

Theorem 6.18. For every \(a \in C^\infty(M), s \in \mathbb{N}, K \subset M\) compact, we have

\[
\left\| \left( \exp \int_0^t X_s ds - S_m(t) \right) a \right\|_{s,K} \leq C \frac{m+1}{m+1} \left( \int_0^t \|X_s\|_{s+m,K'} ds \right)^{m+1} \|a\|_{s+m+1,K'}
\]

for some \(K'\) compact set containing \(K\) and some positive constant \(C > 0\).

**Proof.**

Let us specify this estimate for a non autonomous vector field of the form

\[
X_t = \sum_{i=1}^{m} u_i(t) X_i
\]

where \(X_1, \ldots, X_m\) are smooth vector fields on \(M\) and \(u \in L^2([0,T], \mathbb{R}^m)\).
Theorem 6.19. For every $a \in C^\infty(M)$, $s \in \mathbb{N}$, $K \subset M$ compact, we have

$$\left\| \left( \exp \int_0^t \sum_{i=1}^m u_i(t) X_i - S_m(t) \right) a \right\|_{s,K} \leq \frac{C}{(m+1)!} e^{C \|u\|_2} \|u\|_2^{m+1} \|a\|_{s+m+1,K'}$$

(6.37)

for some $K'$ compact set containing $K$ and some positive constant $C = C_{s,m,K'} > 0$.

Proof.

To complete the discussion, let us describe one special case when the whole Volterra series is convergent. One can prove, for instance, the following convergence result.

Proposition 6.20. Let $X_t$ be a nonautonomous vector field, locally bounded w.r.t. $t$. Assume that there exists a normed subspace $(L, \| \cdot \|) \subset C^\infty(M)$ such that

(a) $X_t a \in L$ for all $a \in L$ and all $t \in I$

(b) $\sup \{\|X_t a\| : a \in L, \|a\| \leq 1, t \in I\} < \infty$

Then the Volterra series (6.30) converges on $L$ for every $t \in I$.

Proof. We can bound the general term of the sum with respect to the norm $\| \cdot \|$ of $L$

$$\left\| \int \cdots \int_{S_k(t)} X_{s_k} \circ \cdots \circ X_{s_1} a \, ds \right\| \leq \int \cdots \int_{S_k(t)} \|X_{s_k}\| \cdots \|X_{s_1}\| \, ds \|a\|$$

(6.38)

then the norm of the $n$-th term of the Volterra series is bounded above by the exponential series, and the Volterra series converges on $L$ uniformly. □

Remark 6.21. The assumption in the theorem is satisfied in particular for a linear vector field $X$ on $M = \mathbb{R}^n$ and $L \subset C^\infty(\mathbb{R}^n)$ the set of linear functions.

A statement about analytic vector fields and [?].
Chapter 7

Lie groups and left-invariant sub-Riemannian structures

7.1 Lie groups and Lie algebras
7.2 Left-invariant structures
7.3 Pontryagin extremals for left invariant structures
7.4 Bi-invariant metrics
7.5 Geodesics
Chapter 8

End-point and Exponential map

In Chapter 4 we started to study necessary conditions for an horizontal trajectory to be a minimizer of the sub-Riemannian length between two fixed points. By applying first order variations we found two different class of candidates, namely normal and abnormal extremals. We also proved that normal extremal trajectories are geodesics, i.e., short arcs realize the sub-Riemannian distance.

In this chapter we go further and we study second order conditions. To this purpose, we introduce the end-point map $E_{q_0}$ that associates to a control $u$ the final point $E_{q_0}(u)$ of the admissible trajectory associated to $u$ and starting from $q_0$. Then we treat the problem of minimizing the energy $J$ of curves joining two fixed points $q_0, q_1 \in M$ as the problem of minimization with constraint

$$\min J|_{E_{q_0}^{-1}(q_1)}, \quad q_1 \in M.$$  \hfill (8.1)

It is then natural to introduce Lagrange multipliers. First order conditions recover Pontryagin extremals, while second order conditions give new information. This viewpoint permits to interpret abnormal extremals as candidates for optimality that are critical points of the map $E_{q_0}$ defining the constraint.

In this chapter we take advantage of the invariance by reparametrization to assume all the trajectories to be defined on the same interval $[0, 1]$. Also, since the energy of a curve coincides with the $L^2$-norm of the corresponding control, it is natural to take $L^2([0, 1], \mathbb{R}^m)$ as class of admissible controls (cf. Section 3.1). This is useful since $L^2([0, 1], \mathbb{R}^m)$ has a natural structure of Hilbert space.

8.1 The end-point map and its differential

Recall that every sub-Riemannian manifold $(M, U, f)$ is equivalent to a free one (cf. Chapter 3). In this chapter we always assume that the sub-Riemannian structure is free of rank $m$, i.e., $U = M \times \mathbb{R}^m$. In the following $\{f_1, \ldots, f_m\}$ denotes a generating frame.

Fix $q_0 \in M$. Recall that, for every control $u \in L^2([0, 1], \mathbb{R}^m)$, the corresponding trajectory $\gamma_u$ is the unique solution of the Cauchy problem

$$\dot{\gamma}(t) = \sum_{i=1}^{m} u_i(t)f_i(\gamma(t)), \quad \gamma(0) = q_0.$$  \hfill (8.2)
**Definition 8.1.** Let \((M, U, f)\) be a free sub-Riemannian manifold of rank \(m\) and fix \(q_0 \in M\). Define \(U_{q_0} \subset L^2([0,1],\mathbb{R}^m)\) the set of controls \(u\) such that the corresponding trajectory \(\gamma_u\) starting at \(q_0\) is defined on \([0,1]\). The end-point map based at \(q_0\) is the map

\[ E_{q_0}: U_{q_0} \to M, \quad E_{q_0}(u) = \gamma_u(1). \quad (8.3) \]

**Exercise 8.2.** Prove that \(U_{q_0}\) is an open subset of \(L^2([0,1],\mathbb{R}^m)\).

In what follows we employ the usual notation \(f_u(q) = \sum_{i=1}^{m} u_i f_i(q)\).

**Remark 8.3.** With the notation of Chapter 6 the end-point map is rewritten as the chronological exponential

\[ E_{q_0}(u) = q_0 \circ \exp \int_0^1 f_u(t) \, dt. \quad (8.4) \]

Now we prove that the end-point map is differentiable and we compute its (Fréchet) differential.

**Proposition 8.4.** The end-point map \(E_{q_0}\) is smooth on \(U_{q_0}\) and for every \(u \in U_{q_0}\) we have

\[ D_u E_{q_0} : L^2([0,1],\mathbb{R}^m) \to T_{\gamma_u(1)} M, \quad D_u E_{q_0}(v) = \int_0^1 (P_{t,1}^u)_{*} f(v(t)) \big|_{\gamma_u(1)} dt. \quad (8.5) \]

for every \(v \in L^2([0,1],\mathbb{R}^m)\). Here \(P_{t,s}^u\) is the flow generated by \(u\).

From the geometric viewpoint, the differential \(D_u E_{q_0}(v)\) computes the integral mean of the vector field \(f_u(t)\) defined by \(v\) along the trajectory \(\gamma_u\) defined by \(u\), where all the vectors are pushed forward in the same tangent space \(T_{\gamma_u(1)}\) with \(P_{t,1}^u\) (see Figure 8.1). We stress that, since \(U_{q_0}\) is an open set of \(L^2([0,1],\mathbb{R}^m)\), the differential is defined on the whole tangent space (identified with) \(L^2([0,1],\mathbb{R}^m)\).

**Proof of Proposition 8.4.** The end-point map from \(q_0\) can be rewritten as the chronological exponential (8.4). Let us first consider the smoothness near the control \(u \equiv 0\).

\[ E(v(\cdot)) = S_m(v) + R_m(v) \quad (8.6) \]

where

\[ S_m(v) = \text{Id} + \sum_{k=1}^{m} \int_{s_k(1)} \cdots \int_{s_1(1)} f_{v(t_k)} \circ \cdots \circ f_{v(t_1)} \, ds \]

Finally, we investigate the differential of \(E(v(\cdot))\) with respect to \(v\).
\[ R_m(v) = \int S_m(1) P_{0,t_m}^v \circ f_{v(t_m)} \circ \cdots \circ f_{v(t_1)} \, ds \]

By estimate (??)

\[ \| R_m(v) a \|_{s,K} \leq \frac{C}{m} e^{C \| v \|_2^2} \| a \|_{s+m,K'} \]  

(8.7)

that proves that the end-point map is differentiable at \( u = 0 \) (indeed \( m \)-times differentiable, for every \( m \in \mathbb{N} \)) and the previous inequality for \( m = 1 \) gives that

\[ \left\| \left( E(v(\cdot)) - \int_0^1 f_{v(t)} dt \right) a \right\|_{s,K} \leq C e^{C \| v \|_2^2} \| a \|_{s+1,K'} \]  

(8.8)

To compute the differential at a point \( \bar{u} \in \mathcal{U} \), have to consider the expansion near 0 of the map

\[ v(\cdot) \mapsto F(\bar{u}(\cdot) + v(\cdot)) = q_0 \circ \exp \int_0^1 f_{(\bar{u}+v)(t)} \, dt. \]

We reproduce the argument used in the proof of Proposition ??, i.e. we write

\[ E(\bar{u} + v) = P_{0,1}^{\bar{u}} \circ G_{\bar{u}}(v) \]

where \( G_{\bar{u}} \) is the map defined as follows

\[ G_{\bar{u}}(v) = \exp \int_0^1 (P_{0,t}^u)^{-1} f_{v(t)} \, dt \]

Indeed this is easily seen by using the variation formula (6.22) (compare also with the proof of Proposition 3.41)

\[ \exp \int_0^1 f_{(u+v)(t)} \, dt = \exp \int_0^1 f_u(t) + f_v(t) \, dt = \exp \int_0^1 \left( \exp \int_0^t ad f_u(s) \right) f_v(t) \, dt \circ \exp \int_0^1 f_u(t) \, dt = \exp \int_0^1 (P_{0,t}^u)^{-1} f_v(t) \, dt \circ P_{0,1}^u \]

Then, the expansion of \( v \mapsto G_{\bar{u}}(v) \) near \( v = 0 \) is obtained again by estimate (??) and one obtains

\[ D_0G_{\bar{u}} = \int_0^1 (P_{0,t}^u)^{-1} f_v(t) \, dt \]  

(8.9)

from which we get, denoting \( q_1 = E(u) \)

\[ D_{\bar{u}}E(v) = (P_{0,1})_* \int_0^1 (P_{0,t}^u)^{-1} f_v(t)(q_0) \, dt = \int_0^1 (P_{t,1})_* f_v(t)(q_1) \, dt. \]  

\[ \square \]
8.2 Lagrange multipliers rule

Let $U$ be an open set of an Hilbert space $H$, and let $M$ be a smooth $n$-dimensional manifold. Consider two smooth maps

$$\varphi : U \rightarrow \mathbb{R}, \quad F : U \rightarrow M.$$ (8.10)

In this section we discuss the Lagrange multipliers rule for the minimization of the function $\varphi$ under the constraint defined by $F$. More precisely, we want to write necessary conditions for the solutions of the problem

$$\min \varphi|_{F^{-1}(q)}, \quad q \in M.$$ (8.11)

**Theorem 8.5.** Assume $u \in U$ is solution of the minimization problem (8.11). Then there exists a covector $(\lambda, \nu) \in T^*qM \times \mathbb{R}$ such that $(\lambda, \nu) \neq (0, 0)$ and

$$\lambda D_u F + \nu D_u \varphi = 0.$$ (8.12)

**Remark 8.6.** Formula (8.18) means that for every $v \in T_qM$ one has

$$\langle \lambda, D_u F(v) \rangle + \nu D_u \varphi(v) = 0.$$

**Proof.** Let us prove that if $u \in U$ is solution of the minimization problem (8.11), then $u$ is a critical point for the extended map $\Psi : U \rightarrow M \times \mathbb{R}$ defined by $\Psi(v) = (F(v), \varphi(v))$.

Indeed, if $u$ is not a critical point for $\Psi$, then $D_u \Psi$ is surjective. By implicit function theorem, this implies that $\Psi$ is locally surjective at $u$. In particular, for every neighborhood $V$ of $u$ it exists $v \in V$ such that $F(v) = F(u) = q$ and $\varphi(v) < \varphi(u)$, that contradicts that $u$ is a constrained minimum.

Hence $D_u \Psi = (D_u F, D_u \varphi)$ is not surjective and there exists a non zero covector $(\lambda, \nu)$ such that $\lambda D_u F + \nu D_u \varphi = 0$. □

8.3 Pontryagin extremals via Lagrange multipliers

Applying the previous result to the case when $F = E_{q_0}$ is the end-point map and $\varphi = J$ is the sub-Riemannian energy, one obtains the following result.

**Corollary 8.7.** Assume that a control $u \in U$ is a solution of the minimization problem (8.11), then there exists $(\lambda, \nu) \in T^*qM \times \mathbb{R}$ such that $(\lambda, \nu) \neq (0, 0)$ and

$$\lambda D_u E_{q_0} + \nu D_u J = 0.$$ (8.13)

Let us now prove that these necessary conditions are equivalent to those obtained in Chapter 4. Recall that, since $J(u) = \frac{1}{2}\|u\|_{L^2}^2$, then $D_u J(v) = (u, v)_{L^2}$ and, identifying $L^2([0, 1], \mathbb{R}^m)$ with its dual, we have $D_u J = u$.

**Proposition 8.8.** We have the following:

($N$) $(u(t), \lambda(t))$ is a normal extremal if and only if there exists $\lambda_1 \in T^*q_1 M$, where $q_1 = E_{q_0}(u)$, such that $\lambda(t) = (P_{t_1}^u)^*\lambda_1$ and $u$ satisfies (8.13) with $(\lambda, \nu) = (\lambda_1, -1)$, namely

$$\lambda_1 D_u E_{q_0} = u.$$ (8.14)

148
(A) \((u(t), \lambda(t))\) is an abnormal extremal if and only if there exists \(\lambda_1 \in T_{q_1}^* M\), where \(q_1 = E_{q_0}(u)\), such that \(\lambda(t) = (P_{t,1}^n)^* \lambda_1\) and \(u\) satisfies (8.13) with \((\lambda, \nu) = (\lambda_1, 0)\), namely

\[
\lambda_1 D_u E_{q_0} = 0. \tag{8.15}
\]

where in (8.14) we identify \(u \in L^2\) with the element \((u, \cdot)_{L^2} \in (L^2)'\).

**Proof.** Let us prove (N). The proof of (A) is similar.

Recall that the pair \((u(t), \lambda(t))\) is a normal extremal if the curve \(\lambda(t)\) satisfies \(\lambda(t) = (P_{t,1}^n)^* \lambda(1)\) (that is equivalent to say that \(\lambda(t)\) is a solution of the Hamiltonian system, cf. Chapter 1) and \(\langle \lambda(t), f_i(\gamma(t)) \rangle = u_i(t)\) for every \(i = 1, \ldots, m\), where \(\gamma(t) = \pi(\lambda(t))\).

Assume that \(u\) satisfies (8.14) for some \(\lambda_1\), let us prove that the curve defined by \(\lambda(t) := (P_{t,1}^n)^* \lambda_1\) is a normal extremal. Condition (8.14) means that for every \(v \in L^2([0, T], \mathbb{R}^m)\) we have

\[
\langle \lambda_1, D_u E_{q_0}(v) \rangle = (u, v)_{L^2} \tag{8.16}
\]

Using (8.3), the left hand side is rewritten as follows

\[
\langle \lambda_1, D_u E_{q_0}(v) \rangle = \int_0^1 \langle \lambda_1, (P_{t,1}^n)^* f_{v(t)}(q_1) \rangle \, dt = \int_0^1 \langle (P_{t,1}^n)^* \lambda_1, f_{v(t)}(\gamma(t)) \rangle \, dt
\]

\[
= \int_0^1 \langle \lambda(t), f_{v(t)}(\gamma(t)) \rangle \, dt = \int_0^1 \sum_{i=1}^m \langle \lambda(t), f_i(\gamma(t)) \rangle \, v_i(t) \, dt,
\]

where we used that \(\gamma(t) = P_{t,1}^{-1}(q_1)\). Then (8.16) becomes

\[
\int_0^1 \sum_{i=1}^m \langle \lambda(t), f_i(\gamma(t)) \rangle \, v_i(t) \, dt = \int_0^1 \sum_{i=1}^m u_i(t) \, v_i(t) \, dt. \tag{8.17}
\]

and since \(v(t)\) is arbitrary, this implies \(\langle \lambda(t), f_i(\gamma(t)) \rangle = u_i(t)\) for every \(i = 1, \ldots, m\). Following the same computations in the opposite direction we have that if \((u(t), \lambda(t))\) is a normal extremal then the identity (8.14) is satisfied. \(\square\)

### 8.4 Critical points and second order conditions

In this chapter, we develop second order conditions for constrained critical points in the case in which the constrained is regular. When applied to the sub-Riemannian case (cf. Section 8.5), this gives second order conditions for normal extremals.

In the following \(\mathcal{H}\) always denote an Hilbert space. Recall that a smooth submanifold of \(\mathcal{H}\) is a subset \(\mathcal{V} \subset \mathcal{H}\) such that for every point \(v \in \mathcal{V}\) there is an open neighborhood \(Y\) of \(v\) in \(\mathcal{H}\) and a smooth diffeomorphism \(\phi : \mathcal{V} \to \mathcal{W}\) to an open subset \(\mathcal{W} \subset \mathcal{H}\) such that \(\phi(\mathcal{V} \cap Y) = \mathcal{W} \cap U\) for \(U\) a closed linear subspace of \(\mathcal{H}\).

We now recall the implicit function theorem in our setting.

**Proposition 8.9** (Implicit function theorem). Let \(F : \mathcal{H} \to M\) be a smooth map and fix \(q \in M\). If \(F\) is a submersion at every \(u \in F^{-1}(q)\), i.e., the Fréchet differential \(D_u F : \mathcal{H} \to T_q M\) is surjective for every \(u \in F^{-1}(q)\), then \(F^{-1}(q)\) is a smooth submanifold whose codimension is equal to the dimension of \(M\). Moreover \(T_u F^{-1}(q) = \ker D_u F\).
We now define critical points.

**Definition 8.10.** Let \( \varphi : \mathcal{H} \to \mathbb{R} \) be a smooth function and \( N \subset \mathcal{H} \) be a smooth submanifold. Then \( u \in N \) is called a *critical point* of \( \varphi|_N \) if \( D_u \varphi|_{T_u N} = 0 \).

We start with a geometric version of the Lagrange multipliers rule, which characterize constrained critical points (not just minima). This construction is then used to develop a second order analysis.

**Proposition 8.11 (Lagrange multipliers rule).** Let \( U \) be an open subset of \( \mathcal{H} \) and assume that \( u \in U \) is a regular point of \( F : U \to M \). Let \( q = F(u) \), then \( u \) is a critical point of \( \varphi|_{F^{-1}(q)} \) if and only if it exists \( \lambda \in T^*_q M \) such that
\[
\lambda D_u F = D_u \varphi. \tag{8.18}
\]

**Proof.** Recall that the differential of \( F \) is a well-defined map
\[
D_u F : T_u U \to T_q M, \quad q = F(u).
\]
Since \( u \) is a regular point, \( D_u F \) is surjective and, by implicit function theorem, the level set \( V_q := F^{-1}(q) \) is a smooth submanifold (of codimension \( n = \dim M \)), with \( u \in V_q \) and \( T_u V_q = \text{Ker} D_u F \).

Since \( u \) is a critical point of \( \varphi|_{V_q} \), by definition \( D_u \varphi|_{T_u V_q} = D_u \varphi|_{\text{Ker} D_u F} = 0 \), i.e.,
\[
\text{Ker} D_u F \subset \text{Ker} D_u \varphi. \tag{8.19}
\]

Now consider the following diagram
\[
\begin{array}{ccc}
T_u U & \xrightarrow{D_u F} & T_q M \\
\downarrow & & \downarrow \\
\mathbb{R} & \xleftarrow{d_u \varphi} & V_q
\end{array}
\]

From (8.19), using Exercise 8.12 it follows that there exists a linear map \( \lambda : T_q M \to \mathbb{R} \) (that means \( \lambda \in T^*_q M \)) that makes the diagram (8.20) commutative.

**Exercise 8.12.** Let \( V \) be a separable Hilbert spaces and \( W \) be a finite-dimensional vector space. Let \( G : V \to W \) and \( \phi : V \to \mathbb{R} \) two linear maps such that \( \text{ker} G \subset \text{ker} \phi \). Then show that there exists a linear map \( \lambda : W \to \mathbb{R} \) such that \( \lambda \circ G = \phi \).

Now we want to consider second order information at critical points. Recall that, for a function \( \varphi : U \to \mathbb{R} \) defined on an open set \( U \) of an Hilbert space \( \mathcal{H} \), the *first and second differential* are defined in the following way,
\[
D_u \varphi(v) = \frac{d}{ds} \bigg|_{s=0} \varphi(u + sv), \quad D_u^2 \varphi(v) = \frac{d^2}{ds^2} \bigg|_{s=0} \varphi(u + sv)
\]
For a function \( F : U \to M \) whose target space is a manifold its first differential \( D_u F : \mathcal{H} \to T_{F(u)} M \) is still well defined while the second differential \( D_u^2 F \) is meaningful only if we fix a set of coordinates in the target space.

If \( \mathcal{V} \) is a submanifold in \( \mathcal{H} \), the first differential of a smooth function \( \psi : \mathcal{V} \to \mathbb{R} \) at a point \( u \in \mathcal{V} \) is defined as
\[
D_u \psi : T_u \mathcal{V} \to \mathbb{R}, \quad D_u \psi(v) = \frac{d}{ds} \bigg|_{s=0} \psi(w(s)),
\]
where \( w : (-\varepsilon, \varepsilon) \to \mathcal{V} \) is a curve that satisfies \( w(0) = u, \ w'(0) = v \). If \( \psi = \varphi|_{\mathcal{V}} \) is the restriction of a function \( \varphi : \mathcal{H} \to \mathbb{R} \) defined globally on \( \mathcal{H} \), then \( D_u \psi = D_u \phi|_{T_u \mathcal{V}} \) coincides with the restriction of the differential defined on the ambient space \( \mathcal{H} \). For the second differential things are more delicate. Indeed the formula

\[
v \in T_u \mathcal{V} \mapsto \frac{d^2}{ds^2} \bigg|_{s=0} \psi(w(s)) \tag{8.21}
\]

where \( w : (-\varepsilon, \varepsilon) \to \mathcal{V} \) is a curve that satisfies \( w(0) = u, \ w'(0) = v \), is a well-defined object (i.e., the right hand side depends only on \( v \)) only if \( u \) is a critical point of \( \psi \). Indeed, if this is not the case, the quantity \((8.21)\) depends also on the second derivative of \( w \), as it is easily checked.

If \( u \) is a critical point of \( \psi : \mathcal{V} \to \mathbb{R} \) (i.e., \( D_u \psi = 0 \)) the second order differential \((8.21)\) is a well-defined quadratic form \( T_u \mathcal{V} \), that is called the Hessian of \( \psi \) at \( u \):

\[
\text{Hess}_u \psi : T_u \mathcal{V} \to \mathbb{R}, \quad v \mapsto \frac{d^2}{ds^2} \bigg|_{s=0} \psi(w(s)) \tag{8.22}
\]

We stress that if \( \psi = \varphi|_{\mathcal{V}} \) is the restriction of a function \( \varphi : \mathcal{H} \to \mathbb{R} \) defined globally on \( \mathcal{H} \), then the Hessian of \( \psi \) at a critical point \( u \) does not coincide, in general, with the restriction of the second differential of \( \varphi \) to the tangent space \( T_u \mathcal{V} \).

Let us compute the Hessian of the restriction in the case when \( \mathcal{V} = F^{-1}(q) \) is a smooth submanifold of \( \mathcal{H} \), and \( \psi = \varphi|_{F^{-1}(q)} \). Using that \( T_u F^{-1}(q) = \text{Ker} \ D_u F \), the Hessian is a well-defined quadratic form

\[
\text{Hess}_u \varphi|_{F^{-1}(q)} : \text{Ker} \ D_u F \to \mathbb{R}
\]

that is computed in terms of the second differentials of \( \varphi \) and \( F \) as follows.

**Proposition 8.13.** For all \( v \in \text{Ker} \ D_u F \) we have

\[
\text{Hess}_u \varphi|_{F^{-1}(q)}(v) = D_u^2 \varphi(v) - \lambda \ D_u^2 F(v). \tag{8.23}
\]

where \( \lambda \) is satisfies the identity \( \lambda \ D_u F = D_u \varphi \).

**Remark 8.14.** We stress again that in \((8.23)\), while the left hand side is a well defined object, in the right hand side \( D_u^2 \varphi \) is well-defined thanks to the linear structure of \( \mathcal{H} \), while \( D_u^2 F \) needs also a choice of coordinates in the manifold \( M \).

**Proof of Proposition 8.13** By assumption \( F^{-1}(q) \subset \mathcal{U} \) is a smooth submanifold in a Hilbert space. Fix \( u \in F^{-1}(q) \) and consider a smooth path \( w(s) \) in \( \mathcal{U} \) such that \( w(0) = u \) and \( w(s) \in F^{-1}(q) \) for all \( s \). Differentiating twice with respect to \( u \), with respect to some local coordinates on \( M \), we have

\[
D_u F(\dot{u}) = 0, \quad (D_u^2 F(\dot{u}), \dot{u}) + D_u F(\ddot{u}) = 0. \tag{8.24}
\]

where we denoted by \( \dot{u} = \dot{u}(0) \) and \( \ddot{u} = \ddot{u}(0) \). Analogous computations for \( \varphi \) gives

\[
\text{Hess}_u \varphi|_{F^{-1}(q)}(\dot{u}) = \frac{d^2}{ds^2} \bigg|_{s=0} \varphi(w(s)) \\
= \langle D_u^2 \varphi(\dot{u}), \dot{u} \rangle + D_u \varphi(\ddot{u}) \\
= \langle D_u^2 \varphi(\dot{u}), \dot{u} \rangle + \lambda D_u F(\ddot{u}) \quad \text{(by } \lambda D_u F = D_u \varphi) \\
= \langle D_u^2 \varphi(\dot{u}), \dot{u} \rangle - \lambda \langle D_u^2 F(\dot{u}), \ddot{u} \rangle \quad \text{(by } (8.24))
\]

\[\square\]
8.4.1 The manifold of Lagrange multipliers

As above, let us consider the two smooth maps \( \varphi : U \to \mathbb{R} \) and \( F : U \to M \) defined on an open set \( U \) of an Hilbert space \( \mathcal{H} \).

**Definition 8.15.** We say that a pair \( (u, \lambda) \), with \( u \in U \) and \( \lambda \in T^*M \), is a Lagrange point for the pair \( (F, \varphi) \) if \( \lambda \in T^*_F(u) \) and \( D_u \varphi = \lambda D_u F \). We denote the set of all Lagrange points by \( C_{F, \varphi} \). More precisely

\[
C_{F, \varphi} = \{(u, \lambda) \in U \times T^*M \mid F(u) = \pi(\lambda), \ D_u \varphi = \lambda D_u F\}.
\] (8.25)

The set \( C_{F, \varphi} \) is a well-defined subset of the vector bundle \( F^*(T^*M) \), that we recall is defined as

\[
F^*(T^*M) = \{(u, \lambda) \in U \times T^*M \mid F(u) = \pi(\lambda)\}.
\] (8.26)

We now study the structure of the set \( C_{F, \varphi} \). It turns to be a smooth manifold under some regularity conditions on the maps \((F, \varphi)\).

**Definition 8.16.** The pair \((F, \varphi)\) is said to be a Morse pair (or a Morse problem) if 0 is not a critical value for the smooth map

\[
\theta : F^*(T^*M) \to U^* \simeq U, \quad (u, \lambda) \mapsto D_u \varphi - \lambda D_u F.
\] (8.27)

**Remark 8.17.** Notice that, if \( M \) is a single point, then \( F \) is the trivial map and with this definition we have that \((F, \varphi)\) is a Morse pair if and only if \( \varphi \) is a Morse function. Indeed in this case \( D_u F = 0 \), and 0 is a critical value for \( \theta \) if, by definition, the second differential \( D^2_u \varphi \) is non-degenerate.

**Proposition 8.18.** If \((F, \varphi)\) define a Morse problem, then \( C_{F, \varphi} \) is a smooth manifold in \( F^*(T^*M) \).

**Proof.** To prove that \( C_{F, \varphi} \) is a smooth manifold it is sufficient to notice that \( C_{F, \varphi} = \theta^{-1}(0) \) and, by definition of Morse pair, 0 is a regular value of \( \theta \). The result follows from the version of the implicit function theorem stated in Lemma 8.19. \( \square \)

**Lemma 8.19.** Let \( N \) be a smooth manifold and \( \mathcal{H} \) a Hilbert space. Consider a smooth map \( f : M \to \mathcal{H} \) and assume that 0 is a regular value of \( f \). Then \( f^{-1}(0) \) is a smooth submanifold of \( N \).

If the dimension of \( U \), the target space of \( \theta \), were finite, a simple dimensional argument would permit to compute the dimension of \( C_{F, \varphi} = \theta^{-1}(0) \) (as in Proposition 8.9). In this case, since the differential of \( \theta \) is surjective we would have that

\[
\dim F^*(T^*M) - \dim C_{F, \varphi} = \dim U
\]

so we could compute the dimension of \( C_{F, \varphi} \)

\[
\dim C_{F, \varphi} = \dim F^*(T^*M) - \dim U
= (\dim U + \text{rank} T^*M) - \dim U
= \text{rank} T^*M = n
\]

However, in the case \( \dim U = +\infty \) the above argument is no more valid, and we need the explicit expression of the differential of \( \theta \).
Proposition 8.20. Under the assumption of Proposition 8.18, then \( \dim C_{F,\varphi} = \dim M = n \).

Proof. To prove the statement, let us choose a set of coordinates \( \lambda = (\xi, x) \) in \( T^*M \) and describe the set \( C_{F,\varphi} \subset F^*(T^*M) \) as follows

\[
\begin{align*}
D_u\varphi - \xi D_u F &= 0 \\
F(u) &= x
\end{align*}
\]  
(8.28)

where here \( \xi \) is thought as a row vector. To compute \( \dim C_{F,\varphi} \), it will be enough to compute the dimension of its tangent space \( T_{(u,\xi,x)}C_{F,\varphi} \) at every \( (u,\xi,x) \). The tangent space \( T_{(u,\xi,x)}C_{F,\varphi} \) is described in coordinates by the set of points \( (u',\xi',x') \) satisfying the equations

\[
\begin{align*}
D^2_{u'}\varphi(u',\cdot) - \xi D^2_{u} F(u',\cdot) - \xi'D_u F(\cdot) &= 0 \\
D_u F(u') &= x'
\end{align*}
\]  
(8.29)

Let us denote the linear map \( Q : \mathcal{U} \to \mathcal{U}^* \simeq \mathcal{U} \) defined by

\[
Q(u') = D^2_{u'}\varphi(u',\cdot) - \xi D^2_{u} F(u',\cdot).
\]

Since \( Q \) is defined by second derivatives of the maps \( F \) and \( \varphi \), it is a symmetric operator on the Hilbert space \( \mathcal{U} \).

The definition of Morse problem is immediately rewritten as follows: the pair \( (F,\varphi) \) defines a Morse problem if and only if the following map is surjective.

\[
\Theta : \mathcal{U} \times \mathbb{R}^n \to \mathcal{U}^* \simeq \mathcal{U}, \quad \Theta(u',\xi') = Q(u') - B(\xi').
\]  
(8.30)

where we denoted with \( B : \mathbb{R}^n \to \mathcal{U}^* \simeq \mathcal{U} \) the map

\[
B(\xi') = \xi' D_u F(\cdot).
\]

Indeed the map \( \Theta \) is exactly the first equation in (8.29). The dimension of \( C_{F,\varphi} \) coincides with the dimension of \( \ker \Theta \). Indeed for each element \( (u',\xi') \in \ker \Theta \) by setting \( x' = D_u F(u') \) we find a unique \( (x',u',\xi') \in C_{F,\varphi} \). Since \( Q \) is self-adjoint, we have

\[
\mathcal{U} = \ker Q \oplus \overline{\im Q}, \quad \dim \ker Q = \text{codim } \im Q.
\]

Using that \( \Theta \) is surjective and \( \dim(\im B) \leq n \) we get that

\[
\dim \ker Q = \text{codim } \im Q \leq \dim \im B \leq n,
\]

is finite dimensional (in particular \( \im Q \) is closed and \( \mathcal{U} = \ker Q \oplus \im Q \)).

If we denote with \( \pi_{\ker} : \mathcal{U} \to \ker Q \) and \( \pi_{\im} : \mathcal{U} \to \im Q \) the orthogonal projection onto the two subspaces, it is easy to see that

\[
\Theta(u',\xi') = 0 \iff \begin{cases} 
\pi_{\ker} B\xi' = 0 \\
\pi_{\im} B\xi' = Qu'
\end{cases}
\]

\[1\]If a manifold \( C \) is described as the set \( \{ z : \Psi(z) = 0 \} \), then its tangent space \( T_z C \) at a point \( z \in C \) is described by the linear equation \( \{ z' : D_z \Psi(z') = 0 \} \).
Moreover \( \pi_{\ker B} : \mathbb{R}^n \to \ker Q \) is a surjective map between finite-dimensional spaces (the surjectivity is a consequence of the fact that \( \Theta \) is surjective). In particular we have \( \dim \ker (\pi_{\ker B}) = n - \dim \ker Q \). Then we get the identity

\[
\dim \ker \Theta = \dim \ker Q + \dim \ker (\pi_{\ker B}) = \dim \ker Q + (n - \dim \ker Q) = n
\]
since \( \pi_{\ker B} : \mathbb{R}^n \to \ker Q \) is a surjective map.

The last characterization of Morse problem leads to a convenient criterion to check whether a pair \((F, \varphi)\) defines a Morse problem.

**Lemma 8.21.** The pair \((F, \varphi)\) defines a Morse problem if and only if

(i) \( \text{Im } Q \) is closed,

(ii) \( \ker Q \cap \ker D_u F = \{0\} \).

**Proof.** Assume that \((F, \varphi)\) is a Morse problem. Then, following the lines of the proof of Proposition 8.20, \( \text{Im } Q \) has finite codimension, hence is closed, and (i) is proved. Moreover, since the problem is Morse, then the image of the differential of the map (8.27) is surjective, i.e. if there exists \( w \in U \) that is orthogonal to \( \text{Im } \Theta \), namely

\[
\langle Q(u'), w \rangle - \langle \xi' D_u F(\cdot), w \rangle = 0, \quad \forall (\xi', u'),
\]
then \( w = 0 \). Using that \( Q \) is self-adjoint we can rewrite the previous identity as

\[
\langle u', Q(w) \rangle - \langle \xi' D_u F(\cdot), w \rangle = 0, \quad \forall (\xi', u'),
\]
that is equivalent, since \( \xi', u' \) are arbitrary, to

\[
Q(w) = 0 \quad \text{and} \quad D_u F(w) = 0.
\]

This proves (ii). The converse implications are proved in a similar way.

**Definition 8.22.** Let \( N \) be a \( n \)-dimensional submanifold. An immersion \( F : N \to T^*M \) is said to be a Lagrange immersion if \( F^* \sigma = 0 \), where \( \sigma \) denotes the standard symplectic form on \( T^*M \).

Let us consider now the projection map \( F_c : C_{F, \varphi} \to T^*M \) defined by:

\[
F_c(u, \lambda) = \lambda.
\]

**Proposition 8.23.** If the pair \((F, \varphi)\) defines a Morse problem, then \( F_c \) is a Lagrange immersion.

**Proof.** First we prove that \( F_c \) is an immersion and then that \( F_c^* \sigma = 0 \).

(i). Recall that \( F_c : C_{F, \varphi} \to T^*M \) where

\[
C_{F, \varphi} = \{(u, \xi, x) \mid \text{equations (8.28) holds}\}
\]

The differential \( D_{(u, \lambda)}F_c : T_{(u, \lambda)}C_{F, \varphi} \to T_{(u, \lambda)}T^*M \) is defined by the linearization of equations (8.28)

\[
T_{(u, \lambda)}C_{F, \varphi} = \{(u', \xi', x') \mid \text{equations (8.29) holds}\}
\]
where

\[ D_{(u,\lambda)} F_c(u',\xi', x') = (\xi', x') \]

Now looking at (8.29) it easily seen that

\[ D_{(u,\lambda)} F_c(u',\xi', x') = 0 \quad \text{iff} \quad Q(u') = D_u F(u') = 0. \]

Since \((F, \varphi)\) defines a Morse problem we have by Lemma 8.21 that such a \(u'\) does not exists. Hence the differential is never zero and \(F_c\) is an immersion.

(ii). We now show that \(F_c^* \sigma = 0\). Since \(\sigma = ds\) is the differential of the tautological form \(s\), and \(F_c^* \sigma = dF_c^* s\) since the pullback commutes with the differential, it is sufficient to show that \(F_c^* s\) is closed. Let us show the identity

\[ F_c^* s = D(\varphi \circ \pi_U)|_{C_{F,\varphi}}. \]

By definition of the map \(F_c\), the following diagram is commutative:

\[ \begin{array}{ccc}
C_{F,\varphi} & \xrightarrow{F_c} & T^* M \\
\pi_U \downarrow & & \downarrow \pi_M \\
U & \xrightarrow{F} & M 
\end{array} \tag{8.31} \]

Moreover, notice that if \(\phi : M \to N\) is smooth and \(\omega \in \Lambda^1(N)\), by definition of pull-back we have \((\phi^* \omega)_q = \omega_{\phi(q)} \circ D_q \phi\). Hence

\[
(F_c^* s)_{(u,\lambda)} = s_\lambda \circ D_{(u,\lambda)} F_c \\
= \lambda \circ \pi_M^* \circ D_{(u,\lambda)} F_c \quad \text{(by definition} \ s_\lambda = \lambda \circ \pi_M^* ) \\
= \lambda \circ D_u F \circ \pi_U^* \quad \text{(by } \tag{8.31} \text{)} \\
= D_u (\varphi \circ \pi_U) \quad \text{(by } \tag{8.18} \text{)}
\]

\[ \square \]

**Definition 8.24.** The set \(\mathcal{L}_{F,\varphi} \subset T^* M\) of Lagrange multipliers associated with the pair \((F, \varphi)\) is the image of \(C_{F,\varphi}\) under the map \(F_c\).

From Proposition 8.23 it follows that, if \(\mathcal{L}_{F,\varphi}\) is a smooth manifold, then it is a Lagrangian submanifold of \(T^* M\), i.e., \(\sigma|_{\mathcal{L}_{F,\varphi}} = 0\).

Collecting the results obtained above, we have the following proposition.

**Proposition 8.25.** Let \((F, \varphi)\) be a Morse pair and assume \((u, \lambda)\) is a Lagrange point such that \(u\) is a regular point for \(F\), where \(F(u) = q = \pi(\lambda)\). The following properties are equivalent:

(i) \(\text{Hess}_u \varphi|_{F^{-1}(q)}\) is degenerate,

(ii) \((u, \lambda)\) is a critical point for the map \(\pi \circ F_c = F|_{C_{F,\varphi}} : C_{F,\varphi} \to M\),

Moreover, if \(\mathcal{L}_{F,\varphi}\) is a submanifold, then (i) and (ii) are equivalent to

(iii) \(\lambda\) is a critical point for the map \(\pi|_{\mathcal{L}_{F,\varphi}} : \mathcal{L}_{F,\varphi} \to M\).
Proof. In coordinates we have the following expression for the Hessian
\[
\text{Hess}_u \varphi|_{E^{-1}(q)}(v) = \langle Q(v), v \rangle, \quad \forall v \in \text{Ker} \, D_u F.
\]
and \( Q \) is the linear operator associated to the bilinear form. Assume that \( \text{Hess}_u \varphi|_{E^{-1}(q)} \) is degenerate, i.e. there exists \( u' \in \text{Ker} \, D_u F \) such that
\[
\langle Qu', v \rangle = 0, \quad \forall v \in \text{Ker} \, D_u F.
\]
In other words \( Q(u') \perp \text{Ker} \, D_u F \) that is equivalent to say that \( Q(u') \) is a linear combination of the row of the Jacobian matrix of \( F \)
\[
\exists \xi' \text{ such that } Q(u') = \xi'D_u F(\cdot).
\]
From equations (8.29) it follows immediately that (i) is equivalent to (ii). The fact that (ii) is equivalent to (iii) is obvious. \( \square \)

8.5 Sub-Riemannian case

In this section we want to specify all the theory that we developed in the previous ones to the case of sub-Riemannian normal extremal. Hence, we will consider the functional \( J \) defined by
\[
J(u) = \frac{1}{2} \int_0^1 |u(t)|^2 dt
\]
and we consider its critical points constrained to a regular level set of the end-point map \( E \), that means that we fix the final point of our trajectory (as usual we assume that the starting point \( q_0 \) is fixed by the very beginning).

We already characterized critical points by means of Lagrange multipliers, now we want to consider second order informations. We start by computing the Hessian of \( J|_{E^{-1}(q_1)} \).

Lemma 8.26. Let \( q_1 \in M \) and \( (u, \lambda) \) be a critical point of \( J|_{E^{-1}(q_1)} \). Then for every \( v \in \text{Ker} \, D_u F \)
\[
\text{Hess}_u J|_{E^{-1}(q_1)}(v) = \|v\|^2_{L^2} - \langle \lambda, D_u^2 E(v) \rangle, \tag{8.32}
\]
where
\[
D_u^2 E(v, v) = 2 q_1 \circ \int_0^1 \int_{s \leq t \leq 1} [(P_{s,1})_s f_v(s), (P_{t,1})_s f_v(t)] ds dt. \tag{8.33}
\]
and \( P_{t,s} \) is the nonautonomous flow defined by the control \( u \).

Proof. By Proposition 8.13 we have
\[
\text{Hess}_u J|_{E^{-1}(q_1)}(v) = D_u^2 J - \lambda D_u^2 E.
\]
It is easy to compute derivatives of \( J \). Indeed we can rewrite it as \( J(u) = \frac{1}{2} \langle u, u \rangle_{L^2} \), hence
\[
D_u J(v) = (u, v)_{L^2}, \quad D_u^2 J(v) = (v, v)_{L^2} = \|v\|^2_{L^2}, \quad \forall v \in \text{Ker} \, D_u E
\]
It remains to compute the second derivative of the end-point map. From the Volterra expansion (8.9) we get
\[
D_u^2 E(v, v) = 2 q_1 \circ \int_0^1 \int_{s \leq t \leq 1} (P_{s,1})_s f_v(s) \circ (P_{t,1})_s f_v(t) ds dt \tag{8.34}
\]
To end the proof we use the following lemma on chronological calculus, which we will use to symmetrize the second derivative.
Lemma 8.27. Let $X_t$ be a nonautonomous vector field on $M$. Then
\[
\iint_{0 \leq s \leq t \leq 1} X_s \circ X_t dsdt = \frac{1}{2} \int_0^1 X_s ds \circ \int_0^1 X_t dt + \frac{1}{2} \iint_{0 \leq s \leq t \leq 1} [X_s, X_t] dsdt. \tag{8.35}
\]

Proof of the Lemma. We have
\[
2 \iint_{0 \leq s \leq t \leq 1} X_s \circ X_t dsdt = \iint_{0 \leq s \leq t \leq 1} X_s \circ X_t dsdt + \iint_{0 \leq s \leq t \leq 1} X_s \circ X_t dsdt - \iint_{0 \leq s \leq t \leq 1} X_t \circ X_s dsdt + \iint_{0 \leq s \leq t \leq 1} X_t \circ X_s dsdt
\]
\[
= \iint_{0 \leq s \leq t \leq 1} X_s \circ X_t dsdt + \iint_{0 \leq s \leq t \leq 1} [X_s, X_t] dsdt + \iint_{0 \leq s \leq t \leq 1} X_t \circ X_s dsdt
\]
\[
= \int_0^1 \int_0^1 X_s \circ X_t dsdt + \iint_{0 \leq s \leq t \leq 1} [X_s, X_t] dsdt
\]
\[
= \int_0^1 X_s ds \circ \int_0^1 X_t dt + \iint_{0 \leq s \leq t \leq 1} [X_s, X_t] dsdt.
\]

Using Lemma 8.27 we obtain from (8.34)
\[
D_u^2 E(v, v) = 2q_1 \circ \iint_{0 \leq s \leq t \leq 1} [(P_{s,1})_s f_v(s), (P_{t,1})_t f_v(t)] dsdt \tag{8.36}
\]
where we used that $\int_0^1 (P_{t,1})_s f_v(t) dt = 0$ since $v \in \text{ker } D_u E$. □

Proposition 8.28. The sub-Riemannian problem $(E, J)$ is a Morse pair.

Proof. We use the characterization of Lemma 8.21. We have to show that
\[
\text{Im } (\text{Id} - \lambda D_u^2 E) \text{ is closed, } \text{Ker } (\text{Id} - \lambda D_u^2 E) \cap \text{Ker } (D_u E) = 0. \tag{8.37}
\]
Using the previous notation and defining $g^t_v := (P_{t,1})_s f_v$, we can write
\[
D_u E(v) = q_1 \circ \int_0^1 g^t_v dt
\]
Moreover we have
\[
\langle \lambda D_u^2 E(v), v \rangle = 2 \iint_{0 \leq s \leq t \leq 1} g^s_v(s) \circ g^t_v(t) dsdt \circ a \tag{8.38}
\]
\[
= \iint_{0 \leq s \leq t \leq 1} g^s_v(s) \circ g^t_v(t) dsdt \circ a + \iint_{0 \leq s \leq t \leq 1} g^t_v(t) \circ g^s_v(s) dsdt \circ a \tag{8.39}
\]
\[
= \int_0^1 \int_0^t g^s_v(s) \circ g^t_v(t) dsdt \circ a + \int_0^1 \int_t^1 g^t_v(t) \circ g^s_v(s) dsdt \circ a \tag{8.40}
\]
where $a$ is a smooth function such that $d_{aq}a = \lambda$.

The kernel of the bilinear form is the kernel of the symmetric linear operator associated to it, the unique symmetric operator $Q$ satisfying
\[
\langle \lambda D^2_u E(v), v \rangle = (Qv, v)_{L^2} = \int_0^1 (Av)(t)v(t)dt.
\]

Then it follows
\[
(Qv)(t) = \left( \int_0^t g^s_v(s)ds \circ g^t + g^t \circ \int_t^1 g^s_v(s)ds \right) \circ a
\] (8.41)

Since (8.41) is a compact integral operator, then $I - Q$ is Fredholm, and the closedness of $\text{Im}(I - Q)$ follows from the fact that it is of finite codimension. On the other hand, for every control $v \in \text{Ker } D_uE$ we can compute (see (8.5))
\[
q_1 \circ \int_0^t g^s_v(s)ds = -q_1 \circ \int_t^1 g^s_v(s)ds
\]

Hence we have that $v$ belong to the intersection (8.37) if and only if it satisfies
\[
(I - \lambda D^2_u E) v(\cdot)(t) = v(t) + \lambda \int_0^t \left[ g^s_v(s), g^t_v(t) \right] (q_1)ds
\]
which has trivial kernel as it follows from the next lemma.

**Lemma 8.29.** Let us consider the linear operator $A : L^2([0, T], \mathbb{R}^m) \rightarrow L^2([0, T], \mathbb{R}^m)$ defined by
\[
(Av)(t) = v(t) - \int_0^t K(t, s)v(s)ds
\] (8.42)

where $K(t, s)$ is a function in $L^2([0, T]^2, \mathbb{R}^m)$. Then
(i) $A = I - Q$, where $Q$ is a compact operator,
(ii) $\ker A = \{0\}$.

Moreover, if $K(t, s) = K(s, t)$ for all $t, s$, then $A$ is a symmetric operator.

**Proof.** The fact that the integral operator $Q : L^2([0, T], \mathbb{R}^m) \rightarrow L^2([0, T], \mathbb{R}^m)$ defined by
\[
(Qv)(t) = \int_0^t K(t, s)v(s)ds
\] (8.43)

is compact is classical (see for instance [18, Chapter 6]). We then prove statement (ii) in two steps. a) we prove it for small $T$. b) we prove it for arbitrary $T$.

(a). Fix $T > 0$ and consider a solution in $L^2([0, T], \mathbb{R}^m)$ to the equation
\[
v(t) = \int_0^t K(t, s)v(s)ds, \quad t \in [0, T].
\] (8.44)

We multiply (8.41) by $v(t)$ and integrate on $[0, T]$, obtaining
\[
\int_0^T v(t)^2dt = \int_0^T \int_0^t K(t, s)v(s)v(t)dsdt
\]

158
By applying twice the Cauchy-Schwartz identity, one obtains
\[ \int_0^T v(t)^2 dt \leq \left( \int_0^T \int_0^T |K(t,s)|^2 ds \right)^{1/2} \int_0^T v(t)^2 dt. \]
or, equivalently
\[ \|v\|_{L^2}^2 \leq \|K\|_{L^2} \|v\|_{L^2}^2. \]
Since for \( T \to 0 \) we have \( \|K\|_{L^2([0,T],\mathbb{R}^m)} \to 0 \), this implies that \( v = 0 \) on \([0,T]\).

(b). Consider a solution of the identity (8.44) and define \( T^* = \sup \{ \tau > 0 \mid v(t) = 0, t \in [0,\tau] \} \).

By part (a) one has \( T^* > 0 \). Since the set \( \{v \in L^2([0,T],\mathbb{R}^m) \mid v(t) = 0 \text{ a.e. on } [0,T^*] \} \) is preserved by \( A \) then again by part (a) one obtains that \( v \) indeed vanishes on \([0,T^* + \varepsilon]\), for some \( \varepsilon > 0 \), contradicting the fact that \( T^* \) is the supremum.

Combining the last result with Proposition 8.23 we obtain the following corollary.

**Corollary 8.30.** The manifold of Lagrange multipliers of the sub-Riemannian problem \((E,J)\)
\[ \mathcal{L}_{(E,J)} := \{ \lambda_1 \in T^* M \mid \lambda_1 = e^{\bar{H}}(\lambda_0), \lambda_0 \in T^* q_0 M \} \]
is a smooth \( n \)-dimensional submanifold of \( T^* M \).

### 8.5.1 Free initial point problem

Let us consider the free initial point problem, i.e., consider the map
\[ E : M \times \mathcal{U} \to M, \quad (q,u) \mapsto E_q(u), \]
where \( E_q(u) \) is the end-point map based at \( q \). Notice that \( E \) is a submersion and
\[ E|_{\{q_0\} \times \mathcal{U}} = E_{q_0}, \quad E|_{M \times \{u\}} = P^u_{0,1} \]
where \( P^u_{t,s} \) is the nonautonomous flow associated with \( u \). Since the initial point is not fixed, the minimization problem
\[ \min_{E^{-1}(q_1)} J \] (8.45)
is the only trivial solution. We can try to look for solutions of the problem
\[ \min_{E^{-1}(q_1)} J(u) + a(q) \] (8.46)
where \( a \in C^\infty(M) \) is a suitable smooth function.

Critical points of this constrained minimization problem can be found with the Lagrange multiplier rule studied in the previous sections with
\[ F = E, \quad \varphi = J + a. \]

Fix a point \((q_0,u) \in M \times \mathcal{U}\). Notice that every level set \( E^{-1}(q_1) \) is regular since the map \( E \) is a submersion. Then the equation (8.18) is written as
\[ \lambda D_{(q_0,u)}E = D_{(q_0,u)}(J + a) \] (8.47)
Since $D(q_0,u)E = (D_uE_q^0, (P_0u^-1)_\ast)$, $D(q_0,u)(J + a) = (D_Jd_q^0a)$, the equation (8.47) splits into
\[ \{ \lambda D F = D(J - aE_q^0), \lambda (P_0u^-1)_\ast = d_q^0a \} \]

Exercise 8.31. Fix $q_0 \in M$ and $a \in C_\infty(M)$. Prove that to every critical point of the free endpoint problem
\[ \min J(u) - a(E_q^0(u)), \lambda(0) = (P_0u^-1)_\ast a \]
we can associate a normal extremal satisfying
\[ \lambda D F = u, \lambda = d_q^0a. \]

In other words, to every critical point of the problem (8.46) we can associate a normal extremal satisfying
\[ \lambda D F = u, \lambda(0) = (P_0u^-1)_\ast a, \lambda(1) = d_q^0a. \]

8.6 Exponential map

A key object in sub-Riemannian geometry is the exponential map, that parametrize normal extremals by their initial covectors.

Proof. By Remark 4.25 we know that if $(t) = e^{t\vec{H}}(0)$ is a solution. The result follows from the uniqueness of the solution and the identity $\lambda(0) = \alpha\lambda(0)$.

Lemma 8.33. Let $H$ be the sub-Riemannian Hamiltonian. Then, for every $\lambda \in TM$ where the domain $S_0$ is the set of covectors such that the corresponding solution of the Hamiltonian system is defined on $[0,1]$.

Definition 8.32. Let $q_0 \in M$. The sub-Riemannian exponential map (based at $q_0$) is the map $E_{q_0}: S_0 \subset T^{*}q_0 M \to M$, $E_{q_0}(\lambda) = \pi \circ e_{\vec{H}}(\lambda)$, where $\lambda = e^{t\vec{H}}(\lambda_0)$ is a solution of the Hamiltonian system.

When there is no confusion on the point where the exponential map is based at, we omit it in the notation, writing $E_{\lambda} = E_{q_0}(\lambda)$. The homogeneity of the sub-Riemannian Hamiltonian $H$ yields to the following homogeneity property of the flow of $\vec{H}$.

Lemma 8.33. Let $H$ be the sub-Riemannian Hamiltonian. Then, for every $\lambda \in TM$, $e^{t\vec{H}}(\alpha\lambda) = \alpha e^{\alpha t\vec{H}}(\lambda)$, for every $\alpha > 0$ and $t > 0$ such that both sides are defined.

Proof. By Remark 4.25 we know that if $(t) = e^{t\vec{H}}(0)$ is a solution. The result follows from the uniqueness of the solution and the identity $\lambda(0) = \alpha\lambda(0)$.

In other words now we do not restrict to the sublevel $F^{-1}(q_0)$ (we do not fix the final point of the trajectory) but we consider a penalty in the functional we want to minimize.

Exercise 8.31. Fix $q_0 \in M$ and $a \in C_\infty(M)$. Prove that to every critical point of the problem (8.46) we can associate a normal extremal satisfying
\[ \lambda D F = u, \lambda(0) = (P_0u^-1)_\ast a, \lambda(1) = d_q^0a. \]
The homogeneity property \([\text{S.}30]\) permits to recover the whole extremal trajectory as the image of the ray that join 0 to \(\lambda_0\) in the fiber \(T_{q_0}^* M\).

**Corollary 8.34.** Let \(\lambda(t), t \in [0, T]\), be the normal extremal that satisfies the initial condition
\[
\lambda(0) = \lambda_0 \in T_{q_0}^* M.
\]
Then the normal extremal path \(\gamma(t) = \pi(\lambda(t))\) satisfies
\[
\gamma(t) = \mathcal{E}_{q_0}(t\lambda_0), \quad t \in [0, T]
\]

**Proof.** Using \([\text{S.}30]\) we get
\[
\mathcal{E}_{q_0}(t\lambda_0) = \pi(e^{tH}(t\lambda_0)) = \pi(e^{tH}(\lambda_0)) = \pi(\lambda(t)) = \gamma(t).
\]

**Remark 8.35.** Due to the homogeneity property we can consider the following map
\[
\mathbb{R}^+ \times C_{q_0} \to M, \quad (t, \lambda_0) \mapsto \mathcal{E}_{q_0}(t\lambda_0)
\]
where \(C_{q_0}\) is the hypercylinder of normalized covectors
\[
C_{q_0} = \{ \lambda \in T_{q_0}^* M \mid H(\lambda) = 1/2 \}
\]
With an abuse of notation in what follows we define
\[
\mathcal{E}_{q_0}(t, \lambda_0) := \mathcal{E}_{q_0}(t\lambda_0),
\]
whenever the right hand side is defined. In other words we restrict to length parametrized extremal paths, considering the time as an extra variable.

**Proposition 8.36.** If \((M, d)\) is complete, then \(\mathcal{D}_{q_0} = T_{q_0}^* M\). Moreover, if there are no strictly abnormal minimizers, the exponential map is surjective.

**Proof.** To prove that \(\mathcal{D}_{q_0} = T_{q_0}^* M\), it is enough to show that any normal extremal \(\lambda(t)\) starting from \(\lambda_0 \in T_{q_0}^* M\) with \(H(\lambda_0) = 1/2\) is defined for all \(t \in \mathbb{R}\). Assume that the extremal \(\lambda(t)\) is defined on \([0, T]\), and assume that it is extendable to any interval \([0, T + \varepsilon]\). The projection \(\gamma(t) = \pi(\lambda(t))\) defined on \([0, T]\) is a curve with unit speed hence for any sequence \(t_j \to T\) the sequence \(\gamma(t_j)\) is a Cauchy sequence on \(M\) since
\[
d(\gamma(t_i), \gamma(t_j)) \leq |t_i - t_j|
\]
hence convergent to a point \(q_1\) by completeness of \(M\). Let us now consider coordinates around the point \(q_1\) and show that, in coordinates \(\lambda(t) = (p(t), x(t))\), also \(p(t)\) is uniformly bounded. This will give a contradiction to the fact that \(\lambda(t)\) is not extendable. By Hamilton equations \([1.33]\)
\[
\dot{p}(t) = -\frac{\partial H}{\partial x}(p(t), x(t)) = -\sum_{i=1}^{m} \langle p(t), f_i(\gamma(t)) \rangle \langle p(t), D_x f_i(\gamma(t)) \rangle
\]
Since \(H(\lambda(t)) = \frac{1}{2} \sum_{i=1}^{m} \langle p(t), f_i(\gamma(t)) \rangle^2 = 1/2\) then \(|\langle p(t), f_i(\gamma(t)) \rangle| \leq 1\). Moreover by smoothness of \(f_i\), the derivatives \(|D_x f_i| \leq C\) are locally bounded in the neighborhood and one get the inequality
\[
|\dot{p}(t)| \leq C|p(t)|
\]
which by Gronwall’s lemma implies that \(|p(t)|\) is uniformly bounded. The second part of the statement follows from the existence of minimizers. \(\square\)
We end this section by the Hamiltonian version of the Gauss’ Lemma

**Proposition 8.37** (Gauss’ Lemma). Let \( \lambda_0 \in C_{q_0} \) and let \( \lambda(t) = e^{t\bar{H}}(\lambda_0) \) for \( t \in [0, 1] \) be a normal extremal. Assume that the sub-Riemannian front \( E_{q_0}(C_{q_0}) \) is smooth at \( E_{q_0}(\lambda(0)) \). Then the covector \( \lambda(1) \) annihilates the tangent space to \( E_{q_0}(C_{q_0}) \).

**Proof.** It is enough to show that for every smooth variation \( \eta^s \in T^*_{q_0} M \cap H^{-1}(1/2) \) of initial covectors such that \( \eta^0 = \lambda(0) \) we have

\[
\left\langle \lambda(1), \frac{d}{ds} \bigg|_{s=0} E_{q_0}(\eta^s) \right\rangle = 0.
\]

Let us consider the family of associated controls \( u^s(\cdot) \) defined by the identities

\[
u^s(t) = \langle \eta^s(t), f_i(\gamma^s(t)) \rangle, \quad \|u^s\|_{L^2} = 1
\]

where \( \eta^s(t) \) is the solution of the Hamiltonian equation with initial value \( \eta^s \) and \( \gamma^s(t) \) is the corresponding trajectory. For these controls one has \( E_{q_0}(\eta^s) = E_{q_0}(u^s) \) hence

\[
\frac{d}{ds} \bigg|_{s=0} E_{q_0}(\eta^s) = \frac{d}{ds} \bigg|_{s=0} E_{q_0}(u^s) = D_u E_{q_0} (v), \quad v := \frac{d}{ds} \bigg|_{s=0} u^s \quad (8.51)
\]

Notice that \( v \) is orthogonal to \( u \) since \( \|u^s\| = \text{const} \). Thus by the normal equation \( (8.14) \) and \( (8.51) \)

\[
\left\langle \lambda(1), \frac{d}{ds} \bigg|_{s=0} E_{q_0}(\eta^s) \right\rangle = \left\langle \lambda(1), D_u E_{q_0}(v) \right\rangle = (u, v)_{L^2} = 0. \quad (8.52)
\]

**Proposition 8.38.** The sub-Riemannian exponential map \( E_{q_0} : T^*_{q_0} M \to M \) is a local diffeomorphism at 0 if and only if \( D_{q_0} = T_{q_0} M \).

**Proof.** It follows from \( D_0 E(\lambda_0) = \dot{\gamma}_0(0) \) that \( \text{im} D_0 E = D_{q_0} \). \( \square \)

### 8.7 Conjugate points and minimality of extremal trajectories

Consider now an extremal pair \( (u(t), \lambda(t)) \), \( t \in [0, 1] \), such that the corresponding extremal path \( \gamma(t) \) is strictly normal. Recall that by Corollary 4.59 the curve \( \gamma \) is a geodesic. Moreover, \( \gamma_{[0, s]} \) is a geodesic too, for every \( s > 0 \). If we define \( \gamma_s(t) := \gamma(st) \), with \( t \in [0, 1] \), then \( \gamma_s \) corresponds to the control \( u_s(t) = su(st) \). Notice that \( \gamma_s \) is the curve \( \gamma_{[0, s]} \) reparametrized with constant speed on \( [0, 1] \).

**Definition 8.39.** An extremal trajectory \( \gamma : [0, T] \to M \) is said to be strongly normal, if \( \gamma_{[0, s]} \) is strictly normal \( \forall s > 0 \).

**Proposition 8.40.** Let \( \gamma \) be a strongly normal extremal trajectory. The following are equivalent:

(i) \( \text{Hess}_{u^s} J |_{E^{-1}(\gamma(1))} \) is positive definite,

(ii) \( \text{Hess}_{u^s} J |_{E^{-1}(\gamma_s(1))} \) is non degenerate for all \( s > 0 \).
Proof. Recall that every operator $A$ on $L^2([0,T],\mathbb{R}^m)$ can be associated with the quadratic form $Q(v) = (Av,v)_{L^2}$ and viceversa. The quadratic form
\[ Q(v) = (Av,v)_{L^2}, \]
defined for $v \in \ker D_{u_s}E$, is written as $(\cdot,\cdot)_{L^2} - Q_s$, with
\[ Q_s(v) = (\lambda(s), D^2_{u_s} E(v,v)), \]
that is associated with a compact operator, thanks to Lemma 8.29. Then define the function
\[ \alpha(s) := \inf_{\|v\|=1} \left\{ \|v\|^2_{L^2} - (\lambda(s), D^2_{u_s} E(v,v)) \right\} = 1 - \sup_{\|v\|=1} \langle \lambda(s), D^2_{u_s} E(v,v) \rangle \] (8.54)

Let us prove the following properties
(a) $\alpha(0) = 1$
(b) $\alpha(\bar{s}) = 0$ implies that $\text{Hess}_{u_s} J\big|_{E^{-1}(\gamma_s(1))}$ is degenerate
(c) $\alpha(s)$ is a continuous and monotone decreasing function

To prove (a), let us notice that
\[ D_{u_s} E(v) = \int_0^s P_{t*} f(v(t)) dt, \quad D^2_{u_s} E(v,v) = \int_0^1 \int_{0 \leq \tau \leq t} [P_{t*} f(v(\tau)), P_{t*} f(v(t))] d\tau dt. \] (8.55)

A change of variables in the integral gives
\[ D_{u_s} E(v) = s \int_0^1 P_{s*} f(v(st)) dt, \quad D^2_{u_s} E(v,v) = s^2 \int_0^1 \int_{0 \leq \tau \leq t \leq s} [P_{s*} f(v(\tau)), P_{s*} f(v(st))] d\tau dt. \] (8.56)

Hence for $s = 0$ we have $D^2_{u_s} E(v,v) = 0$ and $\alpha(0) = 1$.

To prove (b), notice that $\alpha(\bar{s}) = 0$ means that the quadratic form $(\cdot,\cdot)_{L^2} - Q_s$ is nonnegative and has infimum zero. Hence there exists a sequence of $v_j$ such that $\|v_j\|^2 = 1$ and
\[ \|v_j\|^2 - Q(v_j) \to 0. \] (8.57)

Since the unit ball in $L^2$ is weakly compact we can extract a convergent subsequence, that we still denote by the same symbol, $v_j \rightharpoonup \bar{v}$. By compactness of $Q_s$ we have
\[ \|v_j\|^2 - Q(v_j) = 1 - Q(v_j) \to 1 - Q(\bar{v}). \] (8.58)

Comparing (8.57) and (8.58) we get
\[ \text{Hess}_{u_s} J\big|_{E^{-1}(\gamma_s(1))} (\bar{v}) = 0 \]
Thanks to Exercise 8.41 the quadratic form $\text{Hess}_{u_s} J\big|_{E^{-1}(\gamma_s(1))}$ is degenerate.
Exercise 8.41. Let \( V \) be a vector space and \( Q : V \times V \to \mathbb{R} \) be a quadratic form on \( V \). Recall that \( Q \) is degenerate if there exists \( \tilde{v} \in V \) such that \( Q(\tilde{v}, \cdot) = 0 \). Prove that a non-negative quadratic form is degenerate if and only if there exists \( \tilde{v} \) such that \( Q(\tilde{v}, \tilde{v}) = 0 \).

Finally to prove (c), let us write the second differential applied to functions \( v \in L^2([0,1], \mathbb{R}^m) \)
\[
D^2_{uu} E(v,v) = s^2 \int_{0 \leq \tau \leq t \leq 1} [P^1_{s \tau} f_v(\tau), P^1_{s t} f_v(t)] d\tau dt.
\] (8.59)
consider \( 0 \leq s \leq s' \leq 1 \) and \( v \in \text{Ker} \ D_{uu} E \) and define the control
\[
\hat{v}(t) = \begin{cases} \sqrt{s} v \left( \frac{s'}{s} t \right), & 0 \leq t \leq \frac{s}{s'}, \\ 0, & \frac{s}{s'} < t \leq 1. \end{cases}
\]
Then \( \|\hat{v}\| = \|v\|, \hat{v} \in \text{Ker} \ D_{uu} E \) and \( D^2_{uu} E(v,v) = D^2_{uu} E(\hat{v}, \hat{v}) \), hence \( \alpha(s) \geq \alpha(s') \).

To prove that \( \alpha \) is continuous we need that both the integrand in the expression of \( D_{uu} E \) and the kernel \( \text{Ker} \ D_{uu} E \) of these quadratic form is continuous with respect to \( s \). This follows from our main assumption on \( \gamma \). Indeed, since every restriction \( \gamma|_{[0,s]} \) is strictly normal we have that rank of \( D_{uu} E \) is always equal to \( n \), and the kernel continuously depend on \( s \).

Remark 8.42. Notice that (i) implies only that \( u \) is local minimizer in the \( L^2 \)-topology. We will discuss more stronger minimality conditions in next sections.

Definition 8.43. Let \( q_0 \in M \) and \( \mathcal{E}_{q_0} \) be the exponential map based at \( q_0 \). We say that \( q \) is conjugate to \( q_0 \) along \( \gamma(t) = \mathcal{E}_{q_0}(t\lambda) \) if \( q = \gamma(s) \) and \( s \lambda \) is a critical point of the exponential map \( \mathcal{E}_{q_0} \). We say that \( q \) is the first conjugate point to \( q_0 \) along \( \gamma(t) = \mathcal{E}_{q_0}(t\lambda) \) if \( q = \gamma(s) \) and \( s = \inf\{\tau > 0 | \tau \lambda \text{ is a critical point of } \mathcal{E}_{q_0}\} \).

We denote by \( \text{Con}_{q_0} \) the set of all first conjugate points to \( q_0 \) along some normal extremal trajectory starting from \( q_0 \).

Proposition 8.44. Let \( \gamma : [0,T] \to M \) be a strongly normal extremal trajectory and \( s \in ]0, T] \). Then \( \gamma(s) \) is conjugate to \( \gamma(0) \) along \( \gamma \) if and only if \( \text{Hess}_{uu} J |_{E^{-1}(\gamma_s)} \) is degenerate.

Proof. We apply Proposition 8.25. Indeed \( \gamma(s) \) is a conjugate point if and only if \( u_s \) is a critical point of the exponential map, that is equivalent to the fact that \( \text{Hess}_{uu} J |_{E^{-1}(\gamma_s)} \) is degenerate.

Corollary 8.45. Let \( \gamma : [0,T] \to M \) be a strongly normal extremal trajectory and assume that \( \gamma(t) \) is not conjugate to \( \gamma(0) \) along \( \gamma \) for every \( t > 0 \). Then \( \text{Hess}_{uu} J |_{E^{-1}(\gamma_1)} > 0 \). In particular \( \gamma(t) \) is a local minimizer in the \( L^2 \)-topology for controls.

Proof. Since \( \gamma \) contains no conjugate points, by Proposition 8.44 it follows that \( \text{Hess}_{uu} J |_{E^{-1}(\gamma_1)} \) is non degenerate for every \( s \in [0,1] \), hence \( \text{Hess}_{uu} c |_{E^{-1}(\gamma_1)} > 0 \) by Proposition 8.40.

Corollary 8.46. Let \( \gamma : [0,T] \to M \) be a strongly normal extremal trajectory. Then the set
\[
\{ s > 0 | \gamma(s) \text{ is conjugate to } \gamma(0) \}
\]
is isolated from 0.

Proof. It follows from the fact that small pieces of a normal extremal trajectory are minimizers and Proposition 8.44.
8.7.1 Local minimality of normal extremal trajectories in the uniform topology

In the previous section we proved that a normal extremal trajectory that contains no conjugate points, is a local minimizer for the length in the space of admissible trajectories with fixed endpoints, endowed with the $H^1$-topology, i.e., the $L^2$-topology for controls.

In this section we prove that, in absence of conjugate points, a normal extremal trajectory is a local minimizer with respect to the stronger uniform (i.e., $C^0$) topology in the space of admissible trajectories with fixed endpoints.

**Proposition 8.47.** Let $\gamma : [0, T] \to M$ be a strongly normal extremal trajectory. If $\gamma(s)$ is not conjugate to $\gamma(0)$ for every $s > 0$, then $\gamma$ is a local miminum for the length in the $C^0$-topology in the space of admissible trajectories with the same endpoints.

**Proof.** Assume that $\gamma(t) = \pi \circ e^{t\vec{H}}(\lambda_0), \quad \lambda_0 \in T^*_q M$

We want to show that hypothesis of Theorem 4.57 are satisfied. We will use the following lemma, which we prove at the end of the proposition.

**Lemma 8.48.** There exists $a \in C^\infty(M)$ such that

$$\lambda_0 = d_{q_0} a, \quad \text{Hess}_{(q_0, u)} J + a \big|_{E^{-1}(\gamma_s)} > 0,$$

In this case $(\mathbb{E}, J + a)$ is a Morse problem and

$$\mathcal{L}_{(\mathbb{E}, J + a)} = \{e^{t\vec{H}}(d_q a), q \in M\}$$

From this Lemma it follows that $s\lambda_0$ is a regular point of the map $\pi \circ e^{s\vec{H}}|_{\mathcal{L}_0}$, where as usual $\mathcal{L}_0 = \{d_q a, q \in M\}$ denotes the graph of the differential. Using the homogeneity property (8.50) we can rewrite this saying that

$$\pi \circ e^{s\vec{H}}|_{\mathcal{L}_0} \text{ is an immersion at } \lambda_0, \quad \forall s \in [0, 1],$$

In particular it is a local diffeomorphism. Hence we can apply the local version of Theorem 4.57.

**Proof of Lemma 8.48.** First we notice that

$$\text{Ker} D_{(q_0, u)} \mathbb{E} \subset T_{q_0} M \oplus \mathbb{U}, \quad \mathbb{U} \text{ Hilbert}$$

In particular

$$\text{Ker} D_{(q_0, u)} \mathbb{E} \cap (0 \oplus \mathbb{U}) = \text{Ker} D_u \mathbb{E}$$

Since there are no conjugate points, it follows that

$$\text{Hess}_{(q_0, u)} J + a \big|_{0 \oplus \text{Ker} D_u \mathbb{E}} = \text{Hess}_u J > 0 \quad (8.60)$$

Then it is sufficient to show that there exists a choice of the function $a \in C^\infty(M)$ such that the Hessian is positive definite also in the complement. We define

$$W_s := \{\xi \oplus v \in \text{Ker} D_{(q_0, u_s)} \mathbb{E} | \text{Hess}(J + a)(\xi \oplus v, 0 \oplus \text{Ker} D_{u_s} \mathbb{E}) = 0\}$$

165
Notice from [8.60] that, if there is some $\xi \oplus v \in W_s$, then $\xi \neq 0$. Now we prove that there exists a map
$$B_s : T_qM \to U, \quad W_s = \{\xi \oplus B_s\xi, \xi \in T_qM\}$$
Then we will have
$$\text{Ker} \, D_{(q_0,u_s)}E = (0 \oplus \text{Ker} \, D_{u_s}F) + W_s$$
and we get
$$\text{Hess}(J + a)(\xi \oplus B_s\xi + 0 \oplus v, \xi \oplus B_s\xi + 0 \oplus v) =$$
$$= \text{Hess}J(v,v) + \text{Hess}(J + a)(\xi \oplus B_s\xi, \xi \oplus B_s\xi)$$
$$= \text{Hess}J(v,v) + d^2a(\xi,\xi) + Q(\xi)$$
where we used that mixed terms give no contribution and denote with $Q(\xi)$ a quadratic form that does not depend on second derivatives of $a$. In particular, since the first term is positive, we can choose $a$ in such a way that it remains positive.

Up to now we proved a sufficient condition for a strictly normal extremal trajectory to be a strong minimum of the sub-Riemannian distance. Indeed Proposition 8.47 says that, if $\gamma$ contains no conjugate points, then it is optimal with respect to sufficiently $C^0$-closed curves.

On the other hand, if we consider a control $u$ such that the corresponding trajectory
$$\gamma(t) = q_0 \circ \overrightarrow{\exp} \int_0^t f(u(s))ds$$
is strictly normal, that means $u$ is not a critical point of the end-point map $E$, then it is well defined the Hessian of $J|_{E^{-1}(q_1)}$, where $q_1 = E(u)$ at the point $u$. Moreover, if $\gamma$ is locally optimal, also in a very weak sense, then necessarily we have
$$\text{Hess}_u J|_{E^{-1}(q_1)} \geq 0$$
Indeed if the Hessian is sign-indefinite, then the map is locally open around the point $u$ and we have that small perturbations give rise to a smaller cost.

As in the proof of Proposition 8.30 we consider the family of rescaled controls (and corresponding trajectories)
$$u_s(t) = su(st), \quad \gamma_s(t) = \gamma(st), \quad s, t \in [0,1],$$
and we define the function
$$\alpha(s) = \min_{\|v\| = 1} \text{Hess}_{u_s} J|_{E^{-1}(\gamma_s(1))}$$
that is well defined, continuous and non-increasing, under the assumption that $\gamma_s$ is strictly normal for every $s \in [0,1]$. Notice that $\alpha(s) = 0$ if and only if $\gamma(s)$ is a conjugate point. Since $\alpha(0) = 1$ we have only three cases

(a) $\alpha(1) > 0$. By monotonicity this implies $\alpha(s) > 0$ for all $s$ and we have no conjugate points. Hence, by Proposition 8.47 $\gamma$ is a minimum in the strong topology.

(b) $\alpha(1) < 0$. Then the Hessian at $u$ is sign indefinite and $\gamma$ is not a minimum, also in the weak topology.
(c) \( \alpha(1) = 0 \). In this case the Hessian is semi-definite and we cannot conclude anything on the minimality of \( \gamma \).

Notice that in cases (b) and (c) also a segment of conjugate point can appear. To analyze in details case (c) and to understand better the properties of a segment of conjugate point we introduce the notion of Jacobi curves, which is some sense generalize the notion of Jacobi fields in Riemannanian geometry. (see Chapter 15)

8.8 Global minimizers

Before going to the analysis of global minimality of extremal trajectories, let us resume in the following Theorem our results about local minimality.

**Theorem 8.49.** Let \( M \) be complete and \( \gamma(s) \) with \( \gamma|_{[0,s]} \) and \( \gamma|_{[s,1]} \) strictly normal \( 0 \leq s \leq 1 \).

(i) if \( \gamma \) has no conjugate point then its a minimizer in the \( C_0 \)-topology for the trajectories,

(ii) if \( \gamma \) has at least a conjugate point then its not minimizer in the \( L^2 \)-topology for controls.

**Remark 8.50.** The assumption that the curve \( \gamma \) is strictly normal is essential in what we proved. Indeed if a curve \( \gamma \) is both normal and abnormal we have that there exists two covectors \( \lambda_1, \nu_1 \neq 0 \) that satisfy

\[
\lambda_1 D_u F = u, \quad \nu_1 D_u F = 0,
\]

that implies

\[
(\lambda_1 + s\nu_1)D_u F = u, \quad \forall s \in \mathbb{R}
\]

and the whole one parameter family of covectors projects on the same extremal trajectory, and \( \gamma \) would be a critical point of the projection. In this case the definition of conjugate point should be changed.

**Remark 8.51.** Notice that the hypotheses of the above theorem imply that in the case (ii) it not possible to have ha segment of full conjugate point up to \( t = 1 \).

**Definition 8.52.** We say that a point \( q \) is in the cut locus of \( q_0 \) if there exists two length minimizers joining \( q_0 \) and \( q \).

Our previous analysis of conjugate points let us to state the following result.

**Theorem 8.53.** Let \( M \) be a complete sub-Riemannian manifold and \( \gamma : [0,1] \rightarrow M \) be a normal extremal path. Then

(i) assume that \( \gamma|_{[0,s]} \) is strictly normal for all \( s > 0 \) and that \( \gamma \) is not a minimizer. Then there exists \( \tau \in [0,1] \) such that \( \gamma(\tau) \) is either cut or conjugate to \( \gamma(0) \),

(ii) assume that \( \gamma|_{[s,1]} \) is strictly normal for all \( s > 0 \) and that there exists \( \tau \in [0,1] \) such that \( \gamma(s) \) is either cut or conjugate to \( \gamma(0) \). Then \( \gamma \) not a minimizer.

In particular if \( \gamma \) is strongly normal then we have that \( \gamma \) is not a minimizer if and only if there exists a cut or a conjugate point along \( \gamma \).
Proof. (i). Let us assume that $\gamma$ is not a minimizer and that there are no conjugate points along $\gamma$. We prove that this implies the presence of a cut point. Define

$$t_* := \sup\{t \in [0,1] \mid \gamma|_{[0,t]} \text{ is minimizing}\}$$

Let us show that $0 < t_* < 1$. Indeed $t_* > 0$ since small pieces of a normal extremal path are minimizers. Moreover, since $\gamma|_{[0,1]}$ is not a minimizer, by continuity of the distance also $t_* < 1$ and $\ell(\gamma|_{[0,t_*]}) = d(\gamma(0), \gamma(t_*))$.

Fix now a sequence $t_n \to t_*$ such that $t_n > t_*$ for all $n$ and denote by $\gamma_n(\cdot)$ a minimizer joining $\gamma(0)$ to $\gamma(t_n)$ such that $\ell(\gamma_n) = d(\gamma(0), \gamma(t_n))$ (the existence of such a minimizers follows from the completeness assumption).

By compactness of minimizers (up to considering a subsequence) there exists a limit minimizer $\gamma_n \to \hat{\gamma}$ joining $\gamma(0)$ to $\gamma(t_*)$. In particular $\ell(\hat{\gamma}|_{[0,t_*]}) = d(\gamma(0), \gamma(t_*)) = \ell(\gamma|_{[0,t_*]})$.

On the other hand, since the segment $\gamma|_{[0,t_*]}$ contains no conjugate points (by definition of $t_*$), the curve $\gamma|_{[0,t_*]}$ is a minimizer in the strict $C_0$-topology. Thus $\hat{\gamma}$ cannot be contained in a neighborhood $\gamma$. From this it follows that $\gamma(t_*)$ is a cut point.

(ii). Assume that there exists a conjugate point $\gamma(\tau)$ in the segment $[0,1]$. Then $\gamma$ is not a local (hence global) minimizer, as proved in Theorem 8.49. It remains to show that the same remains true if $\gamma(\tau)$ is a cut point. Indeed in this case we have a minimizer $\hat{\gamma}$ such that $\hat{\gamma}(\tau) = \gamma(\tau)$. From this it follows that the curve built with $\hat{\gamma}|_{[0,\tau]}$ and $\gamma|_{[\tau,1]}$ is also a minimizer and the piece $\gamma|_{[\tau,1]}$, by uniqueness of the covector, would be associated with two different normal covectors, hence abnormal, that contradicts our assumptions.

\[\square\]

Theorem 8.54. Let $\gamma : [0,1] \to M$ be a strictly normal extremal path. Assume that for some $s > 0$

(i) $\gamma|_{[0,s]}$ is a global minimizer,

(ii) at each point in a neighborhood of $\gamma(s)$ there exists a unique minimizer joining $\gamma(0)$ to $\gamma(s)$, that is not abnormal.

Then there exists $\varepsilon > 0$ such that $\gamma|_{[0,s+\varepsilon]}$ is a global minimizer.

Proof. Let us consider a neighborhood $O$ of $\gamma(s)$ and, for each $q \in O$, let us denote by $u^q$ (resp. $\gamma^q$) the minimizing control (resp. trajectory) joining $\gamma(0)$ to $q$.

The map $q \mapsto u^q$ is continuous in the $L^2$ topology. Hence we can consider the family $\lambda_1^q$ of covectors such that

$$\lambda_1^q D_{u^q} F = u^q, \quad \forall q \in O.$$  

By the smoothness of $F$ and the continuity of the map $q \mapsto D_{u^q} F$ we have that the map $q \mapsto \lambda_1^q$ is continuous. Indeed since the trajectory associated with $u^q$ is not abnormal by assumptions, one has $D_{u^q} F$ is onto. Thus its adjoint $(D_{u^q} F)^*$ is injective and satisfies $\lambda_1^q = (D_{u^q} F)^* u^q$. Thus the map $q \mapsto \lambda_0^q$ is continuous too, being the composition of the previous one with $(P_{0,1})^{-1}$.

Moreover, the map $q \mapsto \lambda_0^q$ is also injective. Indeed it is an inverse of the exponential map. By the invariance of domain theorem we have that $O' = \{\lambda_0^q, q \in O\}$ is open in $T_q^* M$.

Thus $(1 + \varepsilon) \lambda_0^{\gamma(s)} \in O'$ for $|\varepsilon|$ small enough. Since $(1 + \varepsilon) \lambda_0^{\gamma(s)} = \lambda_0^{((1+\varepsilon)s)}$, this means that $\gamma$ is minimizer on the interval $[0, (1+\varepsilon)s]$. Hence $\gamma(s)$ is not a conjugate point. \[\square\]
Corollary 8.55. If we assume in Theorem 8.54 that $\gamma$ is strongly normal, then $\gamma(s)$ is not a conjugate point.

Corollary 8.56. Assume that the sub-Riemannian structures admits no abnormal minimizer. Let $\gamma : [0, 1] \rightarrow M$ be a length minimizer such that $\gamma(1)$ is conjugate to $\gamma(0)$. Then any neighborhood of $\gamma(1)$ contains a cut point.

8.9 An example: the first conjugate locus on perturbed sphere

In this section we prove that a $C^\infty$ small perturbation of the standard metric on $S^2$ has a first conjugate locus with at least 4 cusps. See Figure ???. Recall that geodesics for the standard metric on $S^2$ are great circles, and the first conjugate locus from a point $q_0$ coincides with its antipodal point $\hat{q}_0$. Indeed all geodesics starting from $q_0$ meet and lose their local and global optimality at $\hat{q}_0$.

Denote $H_0$ the Hamiltonian associated with the standard metric on the sphere and let $H$ be an Hamiltonian associated with a Riemannian metric on $S^2$ such that $H$ is sufficiently close to $H_0$, with respect to the $C^\infty$ topology for smooth functions in $T^*M$.

Fix a point $q_0 \in S^2$. Normal extremal trajectories starting from $q_0$ and parametrized by length (with respect to the Hamiltonian $H$) can be parametrized by covectors $\lambda \in T^*_qM$ such that $H(\lambda) = 1/2$. The set $H^{-1}(1/2)$ is diffeomorphic to a circle $S^1$ and can be parametrized by an angle $\theta$. For a fixed initial condition $\lambda_0 = (q_0, \theta)$, where $q_0 \in M$ and $\theta \in S^1$ we write

$$\lambda(t) = e^{tH}(\lambda_0) = (p(t, \theta), \gamma(t, \theta)),$$

and we denote by $E = E_{q_0}$ the exponential map based at $q_0$

$$E_{q_0}(t, \lambda_0) = \pi \circ e^{tH}(\lambda_0) = \gamma(t, \theta)$$

For every initial condition $\theta \in S^1$ denote by $t_c(\theta)$ the first conjugate time along $\gamma(\cdot, \theta)$, i.e. $t_c(\theta) = \inf\{\tau > 0 | \gamma(\tau, \theta) \text{ is conjugate to } q_0 \text{ along } \gamma(\cdot, \theta)\}$.

Proposition 8.57. The first conjugate time $t_c(\theta)$ is characterized as follows

$$t_c(\theta) = \inf \left\{ t > 0 \left| \frac{\partial E}{\partial \theta}(t, \theta) = 0 \right. \right\}. \quad (8.61)$$

Proof. Conjugate points correspond to critical points of the exponential map, i.e., points $E(t, \theta)$ such that

$$\text{rank} \left\{ \frac{\partial E}{\partial t}(t, \theta), \frac{\partial E}{\partial \theta}(t, \theta) \right\} = 1. \quad (8.62)$$

Notice that $\frac{\partial E}{\partial t}(t, \theta) = \dot{\gamma}(t, \theta) \neq 0$. Let us show that condition (8.62) occurs only if $\frac{\partial E}{\partial \theta}(t, \theta) = 0$. Indeed, by Proposition 8.37 one has that

$$\left\langle p, \frac{\partial E}{\partial t}(t, \theta) \right\rangle = 1, \quad \left\langle p, \frac{\partial E}{\partial \theta}(t, \theta) \right\rangle = 0,$$

thus, whenever $\frac{\partial E}{\partial \theta}(t, \theta) \neq 0$, the two vectors appearing in (8.62) are always linearly independent. \qed
Lemma 8.58. The function $\theta \mapsto t_c(\theta)$ is $C^1$.

Proof. By Proposition 8.57, $t_c(\theta)$ is a solution to the equation (with respect to $t$)

$$\frac{\partial \mathcal{E}}{\partial \theta}(t, \theta) = 0. \quad (8.63)$$

Let us first remark that, for the exponential map $\mathcal{E}_0$ associated with the Hamiltonian $H_0$ we have

$$\frac{\partial \mathcal{E}_0}{\partial \theta}(t_0^0(\theta), \theta) = 0, \quad \frac{\partial^2 \mathcal{E}_0}{\partial t \partial \theta}(t_0^0(\theta), \theta) \neq 0 \quad (8.64)$$

where $t_0^0(\theta)$ is the first conjugate time with respect to the metric induced by $H_0$, as it is easily checked.

Since $H$ is close to $H_0$ in the $C^\infty$ topology, by continuity with respect to the data of solution of ODEs, we have that $\mathcal{E}$ is close to $\mathcal{E}_0$ in the $C^\infty$ topology too. Moreover the condition $(8.64)$ ensures the existence of a solution $t_c(\theta)$ of $(8.63)$ that is close to $t_0^0(\theta)$. Hence we have that

$$\frac{\partial^2 \mathcal{E}}{\partial t \partial \theta}(t_c(\theta), \theta) \neq 0 \quad (8.65)$$

By the implicit function the function $\theta \mapsto t_c(\theta)$ is $C^1$. $\square$

Let us introduce the function $\beta : S^1 \to M$ defined by $\beta(\theta) = \mathcal{E}(t_c(\theta), \theta)$. The first conjugate locus, by definition, is the image of the map $\beta$. The cuspidal point of the conjugate locus are by definition those points where the function $\theta \mapsto t'_c(\theta)$ change sign. By continuity (cf. proof of Lemma 8.58) the map $\beta$ takes value in a neighborhood of the point $\hat{q}_0$ antipodal to $q_0$. Let us take stereographic coordinates around this point and consider $\beta$ as a function from $S^1$ to $\mathbb{R}^2$. By the chain rule and $(8.63)$, we have

$$\beta'(\theta) = t'_c(\theta) \frac{\partial \mathcal{E}}{\partial t}(t_c(\theta), \theta) + \frac{\partial \mathcal{E}}{\partial \theta}(t_c(\theta), \theta) = 0 \quad (8.66)$$

Let us define $g, g_0 : S^1 \to \mathbb{R}^2$ by $g(\theta) := \frac{\partial \mathcal{E}}{\partial t}(t_c(\theta), \theta)$ and $g_0(\theta) := \frac{\partial \mathcal{E}_0}{\partial t}(t_0^0(\theta), \theta)$. The set

$$C_0 = \{\rho g_0(\theta) \mid \theta \in S^1, \rho \in [0, 1]\}$$

is convex, since

$$g_0(\theta) = \begin{pmatrix} \cos \theta \\ \sin \theta \end{pmatrix}$$

By assumption the perturbation of the metric is small in the $C^\infty$-topology, hence

$$C = \{\rho g(\theta) \mid \theta \in S^1, \rho \in [0, 1]\}, \quad (8.67)$$

remains convex.

Theorem 8.59. The conjugate locus of the perturbed sphere has at least 4 cuspidal points.
Proof. Notice that the function $\theta \mapsto t_c'(\theta)$ can change sign only an even number of times on $S^1 = [0, 2\pi]/\sim$. Moreover
\[
\int_0^{2\pi} t_c'(\theta) d\theta = t_c(2\pi) - t_c(0) = 0. \tag{8.68}
\]
A function with zero integral mean on $[0, 2\pi]$ which is not identically zero has to change sign at least twice on the interval. Notice also that
\[
\int_0^{2\pi} t_c'(\theta) g(\theta) d\theta = \int_0^{2\pi} \beta'(\theta) d\theta = \beta(2\pi) - \beta(0) = 0. \tag{8.69}
\]
Let us now assume by contradiction that the function $\theta \mapsto t_c'(\theta)$ changes sign exactly twice at $\theta_1, \theta_2 \in S^1$. Then, by convexity of $C$, there exists a covector $\eta \in (\mathbb{R}^2)^*$ such that $\langle \eta, g(\theta_i) \rangle = 0$ for $i = 1, 2$ and such that $t_c'(\theta) \langle \eta, g(\theta) \rangle > 0$ if $\theta \neq \theta_i$ for $i = 1, 2$. This implies in particular
\[
\left\langle \eta, \int_0^{2\pi} t_c'(\theta) g(\theta) d\theta \right\rangle = \int_0^{2\pi} t_c'(\theta) \langle \eta, g(\theta) \rangle d\theta \neq 0
\]
which contradicts (8.69). \qed

Remark 8.60. A careful analysis of the proof shows that the statement remains true if one considers a small perturbation of the Hamiltonian (or equivalently, the metric) in the $C^4$ topology. Indeed the key point is that $g$ is close to $g_0$ in the $C^2$ topology, to preserve the convexity of the set $C$ defined by (8.67).

The same argument can be applied for every arbitrary small $C^\infty$ (and actually $C^4$) perturbation $H$ of the Riemannian Hamiltonian $H_0$ associated with the standard Riemannian structure on $S^2$, without requiring that $H$ comes from a Riemannian metric.
Chapter 9

2D-Almost-Riemannian Structures

Almost-Riemannian structures are examples of sub-Riemannian structures such that the local minimum bundle rank (cf. Definition 3.20) is equal to the dimension of the manifold at each point (cf. Section 3.1.3). They are the prototype of rank-varying sub-Riemannian structures. In this chapter we study the 2-dimensional case, that is very simple since it is Riemannian almost everywhere (see Theorem 9.19), but presents already some interesting phenomena as for instance the presence of sets of finite diameter but infinite area and the presence of conjugate points even when the curvature is always negative (where it is defined). Also the Gauss-Bonnet theorem has a surprising form in this context.

9.1 Basic Definitions and properties

Thanks to Exercise 3.28 given a structure having constant local minimum bundle rank $m$ one can find an equivalent one having bundle rank $m$. In dimension 2, due to the Lie bracket generating assumption, also the opposite holds true in the following sense: a structure having bundle rank 2 has local minimal bundle rank 2. Hence we can define a 2D-almost-Riemannian structure in the following simpler way.

**Definition 9.1.** Let $M$ be a 2-D connected smooth manifold. A 2D-almost-Riemannian structure on $M$ is a pair $(U, f)$ where

- $U$ is an Euclidean bundle over $M$ of rank 2. We denote each fiber by $U_q$, the scalar product on $U_q$ by $(\cdot | \cdot)_q$ and the norm of $u \in U_q$ as $|u| = \sqrt{u | u}_q$.

- $f : U \to TM$ is a smooth map that is a morphism of vector bundles i.e. $f(U_q) \subseteq T_qM$ and $f$ is linear on fibers.

- $\mathcal{D} = \{ f(\sigma) \mid \sigma : M \to U \text{ smooth section} \}$, is a bracket-generating family of vector fields.

As for a general sub-Riemannian structure, we define:

- the **distribution** as $\mathcal{D}(q) = \{ X(q) \mid X \in \mathcal{D} \} = f(U_q) \subseteq T_qM$,

- the **norm** of a vector $v \in \mathcal{D}_q$ as $\|v\| := \min\{|u|, u \in U_q \text{ s.t. } v = f(q, u)|$. 

173
• **admissible curve** as a Lipschitz curve $\gamma : [0, T] \to M$ such that there exists a measurable and essentially bounded function $u : t \in [0, T] \mapsto u(t) \in U_{\gamma(t)}$, called control function, such that $\dot{\gamma}(t) = f(\gamma(t), u(t))$, for a.e. $t \in [0, T]$. Recall that there may be more than one control corresponding to the same admissible curve.

• **minimal control** of an admissible curve $\gamma$ as $u^*(t) := \arg\min\{|u|, u \in U_{\gamma(t)} \text{ s.t. } \dot{\gamma}(t) = f(\gamma(t), u)\}$ (for all differentiability point of $\gamma$). Recall that the minimal control is measurable (cf. Section 3.1).

• (almost-Riemannian) length of an admissible curve $\gamma : [0, T] \to M$ as $\ell(\gamma) := \int_0^T \|\dot{\gamma}(t)\|dt = \int_0^T |u^*(t)|dt$.

• **distance** between two points $q_0, q_1 \in M$ as

$$d(q_0, q_1) = \inf\{\ell(\gamma) \mid \gamma : [0, T] \to M \text{ admissible}, \ \gamma(0) = q_0, \ \gamma(T) = q_1\}. \quad (9.1)$$

Recall that thanks to the Lie-bracket generating condition, the Chow-Rashevskii Theorem guarantees that $(M, d)$ is a metric space and that the topology induced by $(M, d)$ is equivalent to the manifold topology.

**Definition 9.2.** If $(\sigma_1, \sigma_2)$ is an orthonormal frame for $(\cdot \mid \cdot)_q$ on a local trivialization $\Omega \times \mathbb{R}^2$ of $U$, an orthonormal frame for the 2D-almost-Riemannian structure on $\Omega$ is the pair of vector fields $(F_1, F_2) := (f \circ \sigma_1, f \circ \sigma_2)$. In $\Omega \times \mathbb{R}^2$ the map $f$ can be written as $f(q, u) = u_1 F_1(q) + u_2 F_2(q)$. When this can be done globally, we say that the 2D-almost-Riemannian structure is free.

In this chapter we do not work with an equivalent structure of higher bundle rank that is free. Technically such a structure fits Definition 3.20 (i.e., that local minimum bundle rank is equal to the dimension of the manifold at each point) but not Definition 9.1. We rather work with local orthonormal frames that, as explained below, are orthonormal in the standard sense out of the singular set.

This point of view permits to understand how global properties of $U$ (as its orientability, its topology) are transferred in properties of the almost-Riemannian structure.

**Definition 9.3.** A 2D-almost-Riemannian structure $(U, f)$ over a 2D manifold $M$ is said to be **orientable** if $U$ is orientable. It is said to be **fully orientable** if both $U$ and $M$ are orientable.

**Remark 9.4.** Free 2D almost-Riemannian structures are always orientable.

On an orientable 2D almost-Riemannian structure if $\{F_1, F_2\}$ and $\{G_1, G_2\}$ are two positive oriented orthonormal frames defined respectively on two open subsets $\Omega$ and $\Sigma$ then on $\Omega \cap \Sigma$ there exists a smooth function $\theta : M \to S^1$ such that

$$(G_1(q) \quad G_2(q)) = \left(\begin{array}{cc}
\cos(\theta(q)) & \sin(\theta(q)) \\
-\sin(\theta(q)) & \cos(\theta(q))
\end{array}\right) \left(\begin{array}{c}
F_1(q) \\
F_2(q)
\end{array}\right).$$

As shown by the following examples, one can construct orientable 2D-almost-Riemannian structures on non-orientable manifolds and viceversa.

**An orientable 2D almost-Riemannian structure on the Klein bottle.** Let $M$ be the Klein bottle seen as the square $[-\pi, \pi] \times [-\pi, \pi]$ with the identifications $(x, -\pi) \sim (x, \pi)$, $(-\pi, y) \sim (\pi, y)$. 

174
Let \( U = M \times \mathbb{R}^2 \) with the standard Euclidean metric and consider the morphism of vector bundles given by

\[
f : U \to TM, \quad f(x_1, x_2, u_1, u_2) = (x_1, x_2, u_1, u_2 \sin(2x_1))
\]

This structure is Lie bracket generating and the two vector fields

\[
F_1(x_1, x_2) = f(x_1, x_2, 1, 0) = (x_1, x_2, 1, 0), \quad F_2(x_1, x_2) = (x_1, x_2, 0, \sin(2x_1)),
\]

which are well defined on \( M \), provide a global orthonormal frame. This structure is orientable since \( U \) is trivial.

**Exercise 9.5.** Construct a non orientable almost-Riemannian structure on the 2D torus.

We now define Euler number of \( U \) that measures how far the vector bundle \( U \) is from the trivial one.

**Definition 9.6.** Consider a 2D-almost-Riemannian structure \((U, f)\) on a 2D manifold \( M \). The Euler number of \( U \), denoted by \( e(U) \) is the self-intersection number of \( M \) in \( U \), where \( M \) is identified with the zero section. To compute \( e(U) \), consider a smooth section \( \sigma : M \to U \) transverse to the zero section. Then, by definition,

\[
e(U) = \sum_{p | \sigma(p) = 0} i(p, \sigma),
\]

where \( i(p, \sigma) = 1 \), respectively \(-1\), if \( d_p\sigma : T_pM \to T_{\sigma(p)}U \) preserves, respectively reverses, the orientation. Notice that if we reverse the orientation on \( M \) or on \( U \) then \( e(U) \) changes sign. Hence, the Euler number of an orientable vector bundle \( E \) is defined up to a sign, depending on the orientations of both \( U \) and \( M \). Since reversing the orientation on \( M \) also reverses the orientation of \( TM \), the Euler number of \( TM \) is defined unambiguously and is equal to \( \chi(M) \), the Euler characteristic of \( M \).

**Remark 9.7.** Assume that \( \sigma \in \Gamma(E) \) has only isolated zeros, i.e. the set \( \{p | \sigma(p) = 0\} \) is finite. Since \( U \) is endowed with a smooth scalar product \( \langle \cdot | \cdot \rangle_q \) we can define \( \tilde{\sigma} : M \setminus \{p | \sigma(p) = 0\} \to SU \) by \( \tilde{\sigma}(q) = \frac{\sigma(q)}{\sqrt{\langle \sigma | \sigma \rangle_q}} \) (here \( SU \) denotes the spherical bundle of \( U \)). Then if \( \sigma(p) = 0 \), \( i(p, \tilde{\sigma}) = i(p, \sigma) \) is equal to the degree of the map \( \partial B \to S^1 \) that associate with each \( q \in \partial B \) the value \( \tilde{\sigma}(q) \), where \( B \) is a neighborhood of \( p \) diffeomorphic to an open ball in \( \mathbb{R}^n \) that does not contain any other zero of \( \sigma \).

Notice that if \( i(p, \sigma) \neq 0 \), the limit \( \lim_{q \to p} \tilde{\sigma}(q) \) does not exist.

**Remark 9.8.** Notice that \( U \) is trivial if and only if \( e(U) = 0 \).

**Remark 9.9.** Consider a 2D-almost-Riemannian structure \((U, f)\) on a 2D manifold \( M \). Let \( \sigma \) be a section of \( U \) and \( z_\sigma \) the set of its zeros. As in Remark 9.6, define on \( M \setminus z_\sigma \) the normalization \( \tilde{\sigma} \) of \( \sigma \) and let \( \tilde{\sigma}^\perp \) (still defined on \( M \setminus z_\sigma \)) its orthogonal with respect to \( \langle \cdot | \cdot \rangle_q \). Then the original structure is free when restricted to \( M \setminus z_\sigma \) and \( \{\tilde{\sigma}, \tilde{\sigma}^\perp\} \) is a global orthonormal frame for \( \langle \cdot | \cdot \rangle_q \). The global orthonormal frame for the corresponding 2D-almost-Riemannian structure is then \((f \circ \tilde{\sigma}, f \circ \tilde{\sigma}^\perp)\).

**Exercise 9.10.** Consider a 2D-almost-Riemannian structure \((U, f)\) on a 2D manifold \( M \). Prove that \((U, f)\) is free when restricted to \( M \setminus \{q_0\} \) where \( q_0 \) is any point on \( M \).
Definition 9.11. The singular set \( Z \) of a 2D-almost-Riemannian structure \((U, f)\) over a 2D manifold \( M \) is the set of points \( q \) of \( M \) such that \( f \) is not fiberwise surjective, i.e., such that the rank of the distribution \( k(q) := \dim(D_q) \) is less than 2.

Notice if \( q \in Z \) then \( k(q) = 1 \). Indeed at \( q \) we have \( k(q) = 0 \) then the structure could not be bracket generated at \( q \).

Since outside the singular set \( Z \), \( f \) is fiberwise surjective, we have the following

Proposition 9.12. A 2D-almost-Riemannian structure is Riemannian structure on \( M \setminus Z \).

On Riemannian points, the Riemannian metric \( g \) is reconstructed with the polarization identity (see Exercise 3.8). We have that if \( v = v_1 F_1(q) + v_2 F_2(q) \in T_q M \) and \( w = w_1 F_1(q) + w_2 F_2(q) \in T_q M \) then
\[
g_q(v, w) = v_1 w_1 + v_2 w_2.
\]

By construction, at Riemannian points, \( \{F_1, F_2\} \) is an orthonormal frame in the usual sense
\[
g_q(F_i(q), F_j(q)) = \delta_{ij}, \quad i, j = 1, 2.
\]

Exercise 9.13. Assume that in a local system of coordinates an orthonormal frame is given by
\[
F_1 = \begin{pmatrix} F_1^1 \\ F_2^1 \end{pmatrix}, \quad F_2 = \begin{pmatrix} F_1^2 \\ F_2^2 \end{pmatrix}
\]
and let
\[
F = (F_i^j)_{i,j=1,2} = \begin{pmatrix} F_1^1 & F_2^1 \\ F_1^2 & F_2^2 \end{pmatrix}.
\]

Prove that at Riemannian points the Riemannian metric is represented by the matrix
\[
g = {}^t(F^{-1})F^{-1}.
\]

The following Proposition is very useful to study local properties of 2D-almost-Riemannian structures

Proposition 9.14. For every point \( q_0 \) of \( M \) there exists a neighborhood \( \Omega \) of this point and a system of coordinates \((x_1, x_2)\) in \( \Omega \) such that an orthonormal frame for the 2D-almost-Riemannian structure can be written in \( \Omega \) as:
\[
F_1(q) = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad F_2 = \begin{pmatrix} 0 \\ f(x_1, x_2) \end{pmatrix},
\]
where \( f : \Omega \to \mathbb{R} \) is a smooth function. Moreover

(i) the integral curves of \( F_1 \) are geodesics;

(ii) if the step of the structure at \( q \) is equal to \( s \), we have \( \partial_{x_1}^r f = 0 \) for \( r = 1, 2, \ldots, s - 2 \) and \( \partial_{x_1}^{s-1} f \neq 0 \);

Remark 9.15. Notice that using the system of coordinates and the orthonormal frame given by Proposition 9.14 we have that \( Z \cap \Omega = \{(x_1, x_2) \in \Omega \mid f(x_1, x_2) = 0\} \).

Before proving Proposition 9.14 let us prove the following Lemma

Lemma 9.16. Consider a 2D-almost-Riemannian structure and let \( W \) be a smooth embedded one-dimensional submanifold of \( M \). Assume that \( W \) is transversal to the distribution \( D \), i.e., such that \( D(q) + T_q W = T_q M \) for every \( q \in W \). Then, for every \( q \in W \) there exists an open neighborhood \( U \) of \( q \) such that for every \( \epsilon > 0 \) small enough, the set
\[
\{q' \in U \mid d(q', W) = \epsilon\}
\]
is a smooth embedded one-dimensional submanifold of \( U \).
Proof. Let $H(\lambda)$ be sub-Riemannian Hamiltonian and consider a smooth regular parametrization $\alpha \mapsto w(\alpha)$ of $W$. Let $\alpha \mapsto \lambda_0(\alpha) \in T^*_w(\alpha)M$ be a smooth map satisfying $H(\lambda_0(\alpha)) = 1/2$ and $\lambda_0(\alpha) \perp T^*_w(\alpha)W$.

Let $E(t, \alpha)$ be the solution at time $t$ of the Hamiltonian system with Hamiltonian $H$ and with initial condition $\lambda(t) = \lambda_0(\alpha)$. Fix $q \in W$ and define $\tilde{\alpha}$ by $q = w(\tilde{\alpha})$. Now let us prove that $E(t, \alpha)$ is a local diffeomorphism around the point $(0, \tilde{\alpha})$. To do so let us show that the two vectors

$$v_1 = \frac{\partial E}{\partial \alpha}(0, \tilde{\alpha}) \quad \text{and} \quad v_2 = \frac{\partial E}{\partial t}(0, \tilde{\alpha})$$

(9.4)

are not parallel. On one hand, since $v_1$ is equal to $dw/d\alpha(\tilde{\alpha})$, then it spans $T_qW$. On the other hand, being $H$ quadratic in $\lambda$,

$$\langle \lambda_0(\tilde{\alpha}), v_2 \rangle = \langle \lambda_0(\tilde{\alpha}), \frac{\partial H}{\partial \lambda}(\lambda_0(\tilde{\alpha})) \rangle = 2H(\lambda_0(\tilde{\alpha})) = 1. \quad (9.5)$$

Thus $v_2$ does not belong to the orthogonal to $\lambda_0(\tilde{\alpha})$, that is, to $T_qW$.

Therefore for a small enough neighborhood $U$ of $q$, using the fact that small arcs of normal extremal paths are minimizers, we have that for $\varepsilon > 0$ small enough, the set $A = \{q' \in U \mid d(q', W) = \varepsilon\}$ contains the intersection of $U$ with the images of $E(\varepsilon, \cdot)$ and $E(-\varepsilon, \cdot)$. By possibly restricting $U$, we are in the situation of Figure 9.1 and the set $A$ coincides with the intersection of $U$ with the images of $E(\varepsilon, \cdot)$ and $E(-\varepsilon, \cdot)$.

Remark 9.17. Notice that in this proof we did not make any hypothesis on abnormal extremals. In Section 9.1.3 we are going to see that for 2D almost-Riemannian structures there are no non trivial abnormal extremals.

Proof of Proposition 9.14. Following the notation of the proof of Lemma 9.16 let us take $(t, \alpha)$ as a system of coordinates on $U$ and define the vector field $F_1$ by

$$F_1(t, \alpha) = \frac{\partial E(t, \alpha)}{\partial t}. \quad (9.6)$$

177
Notice that, by construction, for every \( q' \in U \) the vector \( X(q') \) belongs to \( D(q') \) and \( \| F_1(q') \| = 1 \). In the coordinates \((t, \alpha)\) we have \( F_1 = (1, 0) \) and by construction its integral curves are geodesics. Let \( F_2 \) be a vector field on \( U \) such that \((F_1, F_2)\) is an orthonormal frame for the 2D almost-Riemannian structure in \( U \).

We claim that the first component of \( F_2 \) is identically equal to zero. Indeed, were this not the case, the norm of \( F_1 \) would not be equal to one.

We are left to prove \( B \). We have

\[
F_3 := [F_1, F_2] = \begin{pmatrix} 0 \\ \partial_{x_1} f(x_1, x_2) \end{pmatrix}
\]

(9.7)

and beside (9.7), the only brackets among \( F_1, F_2 \) and \( F_3 \) that could be different from zero are of the form

\[
[F_3, \ldots, [F_3, F_1], F_1] = \begin{pmatrix} 0 \\ \partial_{x_1} f(x_1, x_2) \end{pmatrix}.
\]

Hence if the structure has step \( s \) at \( q \) we have \( \partial_{x_1} f = 0 \) for \( r = 1, 2, \ldots, s - 2 \) and \( \partial_{x_1}^{-1} f \neq 0 \). □

The form (9.2) is very useful to express the Riemannian quantities on \( M \setminus Z \). Indeed one has

**Lemma 9.18.** Assume that on an open set \( \Omega \subset M \) a system of coordinates \((x_1, x_2)\) is fixed and an orthonormal frame for the 2D-almost-Riemannian is given in the form (9.2). Then on \( \Omega \cap (M \setminus Z) \) the Riemannian metric, the element of Riemannian area and the Gaussian curvatures are given by

\[
g(x_1, x_2) = \begin{pmatrix} 1 & 0 \\ 0 & \frac{1}{f(x_1, x_2)^2} \end{pmatrix},
\]

(9.8)

\[
dA(x_1, x_2) = \frac{1}{|f(x_1, x_2)|} dx_1 dx_2,
\]

(9.9)

\[
K(x_1, x_2) = \frac{f(x_1, x_2) \partial_{x_1}^2 f(x_1, x_2) - 2 (\partial_{x_1} f(x_1, x_2))^2}{f(x_1, x_2)^2}.
\]

(9.10)

**Proof.** Formula (9.8) is a direct consequence of (9.1). Formula (9.9) comes from the definition of the Riemannian area \( dA(F_1, F_2) = 1 \) where \( \{F_1, F_2\} \) is a local orthonormal frame. Formula (9.10) comes from the formula

\[
K(q) = -\alpha_1^2 - \alpha_2^2 + F_1 \alpha_2 - F_2 \alpha_1
\]

where \( \alpha_1 \) and \( \alpha_2 \) are the two functions defined by \([F_1, F_2] = \alpha_1 F_1 + \alpha_2 F_2\) (see Corollary 4.39). □

Hence in a 2D-almost-Riemannian structure all Riemannian quantities explodes while approaching to \( Z \).

### 9.1.1 How big is the singular set?

A natural question is how big could be the singular set. The answer is given by the following Lemma.

**Theorem 9.19.** Consider a system of coordinates \((x_1, x_2)\) defined on an open set \( \Omega \) and let \( dx_1 dx_2 \) be the corresponding Lebesgue measure. Then \( Z \cap \Omega \) has zero \( dx_1 dx_2 \)-measure.
Proof. Without loss of generality we can assume that $\Omega$ has the following properties:

- it is the product of two non-empty intervals:
  \[ \Omega = (x_1^A, x_1^B) \times (x_2^A, x_2^B), \]
- on $\Omega$ we have an orthonormal frame of the form
  \[ F_1(q) = \begin{pmatrix} 1 & 0 \\ 0 & f(x_1, x_2) \end{pmatrix}, \]
  \[ F_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \]
  \(9.11\)
- on $\Omega$ the step of the structure is $s \in \mathbb{N}$.

If some of the properties above are not satisfied, one can prove the theorem on a countable union of sets where the properties above hold.

Let $1_Z : \Omega \to \{0, 1\}$ be the characteristic function of $Z$. Using Fubini theorem,
\[
\int_{\Omega \cap Z} dx_1 dx_2 = \int_\Omega 1_Z(x_1, x_2) dx_1 dx_2 = \int_{x_2^A}^{x_2^B} \left( \int_{x_1^A}^{x_1^B} 1_Z(x_1, x_2) dx_1 \right) dx_2.
\]

We now prove that for every fixed $\bar{x}_2 \in (x_2^A, x_2^B)$, we have $\int_{x_1^A}^{x_1^B} 1_Z(x_1, \bar{x}_2) dx_1 = 0$ from which the conclusion of the theorem follows.

Indeed B. of Proposition 9.14 guarantees that there exists $r \leq s - 1$ such that $\partial_{x_1}^r f(x_1, \bar{x}_2) \neq 0$ for every $x_1 \in (x_1^A, x_1^B)$. Hence $f(\cdot, \bar{x}_2)$ has only isolated zeros and $\int_{x_1^A}^{x_1^B} 1_Z(x_1, \bar{x}_2) dx_1 = 0$. \(\square\)

**Exercise 9.20.** Show that from the proof of Theorem 9.19 it follows that the singular set is locally the countable union of zero- and one-dimensional manifolds and hence that it is rectifiable.

#### 9.1.2 Genuinely 2D-almost-Riemannian structures have always infinite area

**Theorem 9.21.** Let $\Omega$ be a bounded open set such that $\Omega \cap Z \neq \emptyset$. Then

\[ \text{diam}(\Omega) \leq \infty \text{ and } \int_{\Omega \setminus Z} dA = \infty \]

where $\text{diam}(\Omega)$ is the diameter of $\Omega$ computed with respect to the almost-Riemannian distance and $dA$ is the Riemannian area associated to the almost-Riemannian structure on $\Omega \setminus Z$.

**Proof.** Take a a point $q_0 \in \Omega \setminus Z$ and a system of coordinates $(x_1, x_2)$ on a neighborhood $\Omega_0 \subset \Omega$ of $q_0$. Expanding $f$ in Taylor series, we have
\[ f(x_1, x_2) = a_1 x_1 + a_2 x_2 + O(x_1^2 + x_2^2). \]
\(9.12\)

According to (9.9), the (almost-Riemannian) area of $\Omega_0$ is
\[ \int_{\Omega_0} \frac{1}{|f(x_1, x_2)|} dx_1 dx_2. \]

But the inverse of a function of the form (9.12) is never integrable around the origin in the plane. \(\square\)
9.1.3 Geodesics

Since 2D almost Riemannian structures are particular cases of sub-Riemannian structures, there are two kind of candidate optimal trajectories: normal and abnormal extremals. Normal extremals are geodesics while abnormal extremals could or could not be geodesics. An important fact is the following.

**Theorem 9.22.** For a 2D-almost-Riemannian structure, all abnormal extremal are trivial. Moreover a trivial trajectory \( \gamma : [a, b] \to M, \gamma(t) = q_0 \) is the projection of an abnormal extremal if and only if \( q_0 \in \mathcal{Z} \).

**Proof.** It is immediate to verify that if \( \gamma(t) = q_0 \in \mathcal{Z} \) for every \( t \in [a, b] \) then \( \gamma \) admits an abnormal lift.

Let \( \gamma : [a, b] \to M, (a < b) \) be the projection of an abnormal extremal and let us prove that \( \gamma([a, b]) = q_0 \) for some \( q_0 \in \mathcal{Z} \).

Let us first prove that \( \gamma([a, b]) \subset \mathcal{Z} \). By contradiction assume that there exists \( \bar{t} \in ]a, b[ \) such that \( \gamma(\bar{t}) \notin \mathcal{Z} \). By continuity there exists a non trivial interval \([c, d] \subset ]a, b[ \) such that \( \gamma([c, d]) \cap \mathcal{Z} = \emptyset \). Then \( \gamma|_{[c, d]} \) is a Riemannian geodesic and hence cannot be abnormal. Recall that if an arc of a geodesic is not abnormal, then the geodesic if not abnormal too, hence it follows that \( \gamma \) is not abnormal. This contradicts the hypothesis that \( \gamma \) is the projection of an abnormal extremal.

Let us fix a local system of coordinates such that an orthonormal frame is given in the form (9.2). If this is not possible globally on a neighborhood of \( \gamma([a, b]) \), one can repeat the proof on different coordinate charts.

Let us write in coordinates \( \gamma(t) = (\gamma_1(t), \gamma_2(t)) \). We have different cases.

- If \( (\gamma_1(t), \gamma_2(t)) = (c_1, c_2) \) for every \( t \in [a, b] \) we already know that \( \gamma \) admits an abnormal lift.
- If \( \gamma_1 \) is not constant and \( \gamma_2 = c \in [a, b] \), then \( \dot{\gamma}_2 = 0 \) in \([a, b]\) and \( \mathcal{Z} \) contains a set of the type \( \mathcal{Z} = \{(x_1, c) \mid x_1 \in [x_1^A, x_1^B]\} \) with \( x_1^A < x_1^B \).

Hence \( f = 0 \) on \( \mathcal{Z} \). It follows that \( \partial^r_{x_1} f = 0 \) on \( \mathcal{Z} \) for every \( r = 1, 2, \ldots \). As in the proof of Theorem [9.19] it follows that all brackets between \( F_1 \) and \( F_2 \) are zero on \( \mathcal{Z} \) and that the bracket generating condition is violated. Hence this case is not possible.

- There exists \( \bar{t} \in ]a, b[ \) such that \( \dot{\gamma}_2(\bar{t}) \) is defined and \( \dot{\gamma}_2(\bar{t}) \neq 0 \). Now since

\[
\dot{\gamma}(\bar{t}) = \begin{pmatrix} v_1 \\ v_2 f(\gamma(\bar{t})) \end{pmatrix},
\]

for some \( v_1, v_2 \in \mathbb{R} \), we have \( f(\gamma(\bar{t})) \neq 0 \) and hence \( \dot{\gamma}(\bar{t}) \notin \mathcal{Z} \) violating the condition \( \gamma([a, b]) \subset \mathcal{Z} \) for an abnormal extremal. Hence also this case is not possible.

\[\square\]

Hence all non-trivial geodesics are normal and are projection on \( M \) of the solution of the Hamiltonian system whose Hamiltonian is (cf. (4.30))

\[
H : T^* M \to \mathbb{R}, \quad H(\lambda) = \max_{u \in U_q} \left( \langle \lambda, f(q, u) \rangle - \frac{1}{2} |u|^2 \right), \quad q = \pi(\lambda). \quad (9.13)
\]
Locally, if an orthonormal frame \( \{ F_1, F_2 \} \) is assigned, we have
\[
H(\lambda) = \frac{1}{2} \left( \langle \lambda, F_1(q) \rangle^2 + \langle \lambda, F_2(q) \rangle^2 \right).
\]

For a system of coordinates and a choice of an orthonormal frame as those of Proposition 9.14, we have
\[
H(x_1, x_2, p_1, p_2) = \frac{1}{2} \left( p_1^2 + p_2^2 f(x_1, x_2)^2 \right). \tag{9.14}
\]

As a consequence of the fact that all geodesics are projections of solutions of a smooth Hamiltonian system and that our structure is Riemannian on \( M \setminus Z \), we have

**Proposition 9.23.** In 2D almost-Riemannian geometry all geodesics are smooth and they coincide with Riemannian geodesics on \( M \setminus Z \).

The only particular property of geodesics in almost-Riemannian geometry is that on the singular set their velocity is constrained to belong to the distribution (otherwise their length could not be finite). All this is illustrated in the next section for the Grushin plane.

### 9.2 The Grushin plane

The Grushin plane is the simplest example of genuinely almost-Riemannian structure. It is the free almost-Riemannian structure on \( \mathbb{R}^2 \) for which a global orthonormal frame is given by
\[
F_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad F_2 = \begin{pmatrix} 0 \\ x_1 \end{pmatrix}
\]

In the sense of Definition 9.1, it can be seen as the pair \((U, f)\) where \( U = \mathbb{R}^2 \times \mathbb{R}^2 \) and \( f((x_1, x_2), (u_1, u_2)) = ((x_1, x_2), (u_1, u_2x_1)) \).

Here the singular set \( Z \) is the \( x_2 \) axis and on \( \mathbb{R}^2 \setminus Z \) the Riemannian metric, the Riemannian area and the Gaussian curvature are given respectively by:
\[
g = \begin{pmatrix} 1 & 0 \\ 0 & \frac{1}{x_1^2} \end{pmatrix}, \quad dA = \frac{1}{|x_1|} dx_1 dx_2, \quad K = -\frac{2}{x_1^2}. \tag{9.15}
\]

Notice that the (almost-Riemannian) area of an open set intersecting the \( x_2 \) axis is always infinite.

#### 9.2.1 Geodesics of the Grushin plane

In this section we recall how to compute the geodesics for the Grushin plane, with the purpose of stressing that they can cross the singular set with no singularities.

In this case the Hamiltonian (9.14) is given by
\[
H(x_1, x_2, p_1, p_2) = \frac{1}{2} \left( p_1^2 + x_1^2 p_2^2 \right) \tag{9.16}
\]

and the corresponding Hamiltonian equations are:
\[
\begin{align*}
\dot{x}_1 &= p_1, & \dot{p}_1 &= -x_1 p_2^2 \\
\dot{x}_2 &= x_1^2 p_2, & \dot{p}_2 &= 0
\end{align*} \tag{9.17}
\]
Figure 9.2: Geodesics and the front for the Grushin plane, starting from the singular set.

Geodesics parameterized by arclength are projections on the \((x_1,x_2)\) plane of solutions of these equations, lying on the level set \(H = 1/2\). We study the geodesics starting from: i) a point on \(\mathcal{Z}\), e.g. \((0,0)\); ii) an ordinary point, e.g. \((-1,0)\).

**Case** \((x_1(0),x_2(0)) = (0,0)\)

In this case the condition \(H(x_1(0),x_2(0),p_1(0),p_2(0)) = 1/2\) implies that we have two families of geodesics corresponding respectively to \(p_1(0) = \pm 1\), \(p_2(0) \equiv a \in \mathbb{R}\). Their expression can be easily obtained and it is given by:

\[
\begin{align*}
\begin{cases}
  x_1(t) &= \pm t, & x_2(t) &= 0 \\
  x_1(t) &= \pm \frac{\sin(at)}{a}, & x_2(t) &= \frac{2at - \sin(2at)}{4a^2} 
\end{cases}
\end{align*}
\]  

if \(a = 0\) if \(a \neq 0\)  

(9.18)

Some geodesics are plotted in Figure 9.2 together with the “front”, i.e., the end point of all geodesics at time \(t = 1\). Notice that geodesics start horizontally. The particular form of the front shows the presence of a conjugate locus accumulating to the origin.

**Case** \((x_1(0),x_2(0)) = (-1,0)\)

In this case the condition \(H(x_1(0),x_2(0),p_1(0),p_2(0)) = 1/2\) becomes \(p_1^2 + p_2^2 = 1\) and it is convenient to set \(p_1 = \cos(\theta), p_2 = \sin(\theta)\), \(\theta \in \mathbb{S}^1\). The expression of the geodesics is given by:

\[
\begin{align*}
\begin{cases}
  x_1(t) &= t - 1, & x_2(t) &= 0, & \text{if } \theta = 0 \\
  x_1(t) &= \frac{\sin(t) - t \sin(\theta)}{\sin(\theta)}, & x_2(t) &= \frac{2t - 2 \cos(\theta) + \frac{\sin(2\theta - 2t \sin(\theta))}{\sin(\theta)}}{4 \sin(\theta)} & \text{if } \theta \notin \{0, \pi\} 
\end{cases}
\end{align*}
\]  

182
Some geodesics are plotted in Figure 9.3 together with the “front” at time $t = 4.8$. Notice that geodesics pass horizontally through $Z$, with no singularities. The particular form of the front shows the presence of a conjugate locus. Geodesics can have conjugate times only after intersecting $Z$. Before it is impossible since they are Riemannian and the curvature is negative.

![Image](image.png)

Figure 9.3: Geodesics and the front for the Grushin plane, starting from a Riemannian point.

### 9.3 Riemannian, Grushin and Martinet points

In 2D almost-Riemannian structures there are 3 kind of important points, namely Riemannian, Grushin and Martinet points. As we are going to see in Section 9.4 these points are important in the sense that if a system has only this type of points then this is true also after a small perturbation of the system. Moreover arbitrarily close to any system there is a system where only these points are present.

First we study under which conditions $Z$ has the structure of a 1D manifold. To this purpose we are going to study $Z$ as the set of zeros of a function.

**Definition 9.24.** Let $\{F_1, F_2\}$ be a local orthonormal frame on an open set $\Omega$ and let $\omega$ be a volume form on $\Omega$. On $\Omega$ define the function $\Phi = \omega(F_1, F_2)$.

**Exercise 9.25.** Prove that $\Phi$ is invariant by a positive oriented change of orthonormal frame defined on the same open set $\Omega$.

Since a volume form can be globally defined when $M$ is orientable we have that $\Phi$ can be globally defined on fully orientable 2D almost-Riemannian structures (cf. Definition 9.3), just defining it as above on positive oriented orthonormal frames.
For structure that are not fully orientable, Φ can be defined only locally and up to a sign. (notice however that |Φ| is always well defined). This is what should be taken in mind every time that the function Φ appears in the following.

If in a system of coordinates \((x_1, x_2)\), we write
\[
F_1 = \begin{pmatrix} F_{11}^1 \\ F_{11}^2 \end{pmatrix}, \quad F_2 = \begin{pmatrix} F_{21}^1 \\ F_{21}^2 \end{pmatrix}, \quad \omega(x_1, x_2) = h(x_1, x_2) dx_1 \wedge dx_2
\]
then
\[
\Phi(x_1, x_2) = h(x_1, x_2) \det \left( \begin{pmatrix} F_{11}^1 & F_{12}^1 \\ F_{21}^1 & F_{22}^1 \end{pmatrix} \right) \bigg|_{(x_1, x_2)}.
\]

**Remark 9.26.** For a system of coordinates and a choice of an orthonormal frame as those of Proposition 9.14 and taking \(\omega = dx_1 \wedge dx_2\), we have \(\Phi(x_1, x_2) = f(x_1, x_2)\).

The function \(\Phi\) permits to write,
\[
Z = \{q \in M \mid \Phi(q) = 0\}.
\]

We are now going to consider the following assumptions

**H0**

If \(\Phi(q_0) = 0\) then \(d\Phi(q_0) \neq 0\).

**H0**

The condition \(\text{H0}_{q_0}\) holds for every \(q_0 \in M\).

**Exercise 9.27.** Prove that the conditions above do not depend on the choice of the volume form \(\omega\).

By definition of submanifold we have

**Proposition 9.28.** Assume that \(\text{H0}\) holds. Then \(Z\) is a one dimensional embedded submanifold of \(M\).

As usual define \(D_1 = D, D_{i+1} = D_i + [D_i, D_i], i = 1, 2, \ldots\) We are now ready to define Riemannian, Grushin and Martinet points.
**Definition 9.29.** Consider a 2D-almost Riemannian structure. Fix \( q_0 \in M \).

- If \( \mathcal{D}_1(q_0) = T_{q_0} M \) (equivalently if \( q_0 \notin Z \)) we say that \( q_0 \) is a *Riemannian* point.
- If \( \mathcal{D}_1(q_0) \neq T_{q_0} M \) (equivalently if \( q_0 \in Z \)), \( H_0(q_0) \) holds then
  - if \( \mathcal{D}_2(q_0) = T_{q_0} M \) we say that \( q_0 \) is a *Grushin* point.
  - if \( \mathcal{D}_2(q_0) \neq T_{q_0} M \) we say that \( q_0 \) is a *Martinet* point.

**Remark 9.30.** Hence under \( H_0 \) every point is either a Riemannian or a Grushin or a Martinet point.

**Exercise 9.31.** By using the system of coordinate given by Proposition 9.14 prove the following:

- \( q_0 \) is a Grushin point if and only if \( q_0 \in Z \) and \( L_v \Phi(q_0) \neq 0 \) for \( v \in \mathcal{D}(q_0), \|v\| = 1 \).
- \( q_0 \) is a Martinet point if and only if \( q_0 \in Z \), \( d\Phi(q_0) \neq 0 \), and for \( v \in \mathcal{D}(q_0), \|v\| = 1 \), we have \( L_v \Phi(q_0) = 0 \).

The following proposition describes properties of Grushin and Martinet points (see Figure 9.4).

**Proposition 9.32.** We have the following:

(i) \( Z \) is an embedded 1D manifold around Grushin or Martinet points;

(ii) if \( q_0 \) is a Grushin point then \( \mathcal{D}(q_0) \) is transversal to \( T_{q_0} Z \);

(iii) if \( q_0 \) is a Martinet point then \( \mathcal{D}(q_0) \) is parallel to \( T_{q_0} Z \);

(iv) Martinet points are isolated.

**Proof.** We use the system of coordinates and an orthonormal frame as the one given by Proposition 9.14 with \( q_0 = (0,0) \),

\[
F_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad F_2 = \begin{pmatrix} 0 \\ f \end{pmatrix}.
\]

If we take \( \omega = dx \wedge dy \), we have \( \Phi = f, \ d\Phi = (\partial_{x_1} f, \partial_{x_2} f) \).

To prove (i), it is sufficient to notice that by definition \( d\Phi \neq 0 \) at Grushin and Martinet points.

To prove (ii), notice that \( \mathcal{D}(q_0) = \text{span}(F_1(q_0)) = (1,0) \) while \( T_{q_0} Z = \text{span}\{-\partial_{x_2} f(q_0), \partial_{x_1} f(q_0)\} \) that are not parallel since \( \partial_{x_1} f(q_0) \neq 0 \).
To prove (iii), notice that \( D(q_0) = \text{span}(F_1(q_0)) = (1, 0) \) while \( T_{q_0}Z = \text{span}\{(-\partial_{x_2} f, 0)\} \) since the condition \( D_2(q_0) \neq T_{q_0}M \) implies \( \partial_{x_1} f(q_0) = 0 \).

To prove (iv), simply observe that if Martinet points were accumulating at \( q_0 \) then at that point we could not have \( \partial_{x_1}^{-1} f \neq 0 \), where \( s \) is the step of the structure at \( q_0 \).

\[ \square \]

**Examples**

- All points on the \( x_2 \) axis for the Grushin plane are Grushin points.
- The origin the following structure is the simplest example of Martinet point

\[
F_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad F_2 = \begin{pmatrix} 0 \\ x_2 - x_1^2 \end{pmatrix},
\]

- The origin for the following example

\[
F_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad \text{and} \quad F_2 = \begin{pmatrix} 0 \\ x_2 - x_1^2 \end{pmatrix},
\]

is not a Martinet point since the condition \( d\Phi(0,0) \neq 0 \) is not satisfied. Outside the origin all points are either Riemannian or Grushin points, but at the origin \( Z \) is not a manifold.
- The \( x_2 \) axis of the following example

\[
F_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad \text{and} \quad F_2 = \begin{pmatrix} 0 \\ x_1^2 \end{pmatrix},
\]

is not made by Grushin points since \( D^2((0, x_2)) \neq T(0, x_2)M \) and it is not made by Martinet points since \( d\Phi(0, x_2) \neq 0 \) is not satisfied (although in this case \( Z \) is a manifold). In this case \( D(0, x_2) \) is transversal to \( Z \).

**Proposition 9.33.** Let \( q_0 \) be a Riemannian, Grushin or a Martinet point. There exists a neighborhood \( \Omega \) of \( q_0 \) and a system of coordinates \((x_1, x_2)\) in \( \Omega \) such that an orthonormal frame for the 2D-almost-Riemannian structure can be written in \( \Omega \) as:

**(NF1)** if \( q_0 \) is a Riemannian point, then

\[
F_1(x_1, x_2) = (1, 0), \quad F_2(x_1, x_2) = (0, e^{\phi(x_1, x_2)}),
\]

**(NF2)** if \( q_0 \) is a Grushin point, then

\[
F_1(x_1, x_2) = (1, 0), \quad F_2(x_1, x_2) = (0, xe^{\phi(x_1, x_2)})
\]

**(NF3)** if \( q_0 \) is a Martinet point, then

\[
F_1(x_1, x_2) = (1, 0), \quad F_2(x_1, x_2) = (0, (x_2 - x_1^s) e^{\xi(x_1, x_2)}),
\]

where \( \phi, \xi \) and \( \psi \) are smooth real-valued functions such that \( \phi(0, x_2) = 0 \) and \( \psi(0) \neq 0 \). Moreover \( s \geq 2 \) is an integer, that is the step of the structure at the Martinet point.
9.4 Generic 2D-almost-Riemannian structures

Recall hypothesis $H_{0q_0}$ and $H_0$:

$H_{0q_0}$ If $\Phi(q_0) = 0$ then $d\Phi(q_0) \neq 0$.

$H_0$ The condition $H_{0q_0}$ holds for every $q_0 \in M$.

Recall the $H_0$ is independent from the volume form used to define the function $\Phi$. We have seen (cf. Remark 9.30) that under hypothesis $H_0$ every point is either a Riemannian or a Grushin or a Martinet point.

In this section we are going to prove that hypothesis $H_0$ holds for most of the systems. More precisely we are going to prove that hypothesis $H_0$ is generic in the following sense.

**Definition 9.34.** Fix a rank 2 Euclidean bundle $U$ over a 2D compact manifold $M$. Let $F$ be the set of all morphism of bundle from $U$ to $TM$ such that $(U,f), f \in F$ is a 2D almost-Riemannian structure. Endowed $F$ with the $C^1$ norm. We say that a subset of $F$ is generic if it is open and dense in $F$.

**Theorem 9.35.** Under the same hypothesis of Definition 9.34 let $\bar{F} \subset F$ the subset of morphisms satisfying $H_0$. Then $\bar{F}$ is generic.

**Remark 9.36.** In Theorem 9.35 we have assumed that $M$ is compact. A similar result holds also in the case in which $M$ is not compact. However, in the non compact case, one gets that $\bar{F}$ is a countable union of open and dense subsets of $F$ and one should use a suitable topology (the Whitney one). In this book we have decided not to enter inside transversality theory and we have provided a statement that can be proved easily via the Sard lemma.

9.4.1 Proof of the genericity result

Since the map that to $f : U \to TM$ associate $\Phi : M \to \mathbb{R}$ is continuous in the $C^1$ topology, a small perturbation of $f$ will provoke a small perturbation of $\Phi$. Fixed $q_0$, condition $H_{0q_0}$ is clearly open in the set of maps from $M$ to $\mathbb{R}$ for the $C^1$ topology. As a consequence, since the manifold is compact, condition $H_0$ is open as well.

We are now going to prove that $H_0$ is dense. To this purpose we consider an almost Riemannian structure $(U,f)$ over $M$ with the corresponding function $\Phi$ and we construct an arbitrarily small perturbation (in the $C^1$ norm) of $f$ for which $H_0$ is satisfied.

Fix a finite number of points points $q_1, \ldots, q_r$ in such a way that

- the structure is Riemannian in a neighborhood $\Omega$ of $\{q_1, \ldots, q_r\}$;
- if we consider another open set $\Omega_0$ compacy contained in $\Omega$, the structure $(U,f)$ when restricted to $M := \operatorname{int}(M \setminus \Omega_0)$ is free (cf. Remark 9.9 and Exercise 9.10); In the following we call $(U,f)|_M$ this structure.

For every $\varepsilon \in \mathbb{R}$ with $|\varepsilon|$ small enough, consider a perturbation $f_\varepsilon$ of $f$ such that:

- $\|f - f_\varepsilon\|_{C^1} \leq C_\varepsilon$ (for some $C > 0$ independent from $\varepsilon$);
- the corresponding function $\Phi_\varepsilon$ satisfies:
(A) on \( \tilde{M} \) we have \( \Phi_{\varepsilon} = \Phi + \varepsilon \);
(B) on \( \Omega_0 \), where the structure is Riemannian, we do not add any non-Riemannian point.

The difficult point is to realize (A). This can be done thanks to Lemma 9.37 below. Once that this is done, (B) can be easily realized since the perturbation \( f_{\varepsilon} \) is small in the \( C^1 \) norm and the property of having only Riemannian points is open.

Let now apply the Sard Lemma to the \( C^\infty \) function \( \Phi|_{\tilde{M}} \). We have that the set

\[
\{ c \in \mathbb{R} \text{ such that there exists } q \in \tilde{M} \text{ such that } \Phi(q) = c \text{ and } d\Phi(q) = 0 \}
\]

has measure zero. As a consequence, since \( \Phi_{\varepsilon} = \Phi + \varepsilon \), we have that the set

\[
\{ \varepsilon \in \mathbb{R} \text{ such that there exists } q \in \tilde{M} \text{ such that } \Phi_{\varepsilon}(q) = 0 \text{ and } d\Phi_{\varepsilon}(q) = 0 \}
\]

has measure zero. It follows that for almost every \( \varepsilon \) condition \( H_0 \) is realized for \( f_{\varepsilon} \).

To conclude the proof we have to show that (A) can be realized. This can be done thanks to the following Lemma.

**Lemma 9.37.** For every \( \varepsilon \in \mathbb{R} \) with \( |\varepsilon| \) small enough there exists a perturbation \( f_{\varepsilon} \) of \( f \) such that

\[ \| f - f_{\varepsilon} \|_{C^1} \leq C\varepsilon \] (for some \( C > 0 \) independent from \( \varepsilon \)) and on \( \tilde{M} \) we have \( \Phi_{\varepsilon} = \Phi + \varepsilon \);

**Proof.** Let \( \sigma \) be a never vanishing section of \((\mathcal{U}, f)\) restricted to \( \tilde{M} \) of unitary norm.

Let \( \sigma^\perp \) the orthogonal section to \( \sigma \) i.e. satisfying \( (\sigma|_{\sigma^\perp})_q = 0 \) and \( (\sigma^\perp|_{\sigma^\perp})_q = 1 \). A global orthonormal frame for the structure is \((F, F^\perp) := (f \circ \sigma, f \circ \sigma^\perp)\).

Let us first assume that \( F \) is never vanishing.

Cover \( \tilde{M} \) with a finite number of coordinate neighborhood \( \mathcal{U}_i, i = 1 \ldots N \), and in each \( \mathcal{U}_i \), construct coordinates \((x_1^i, x_2^i)\) in such a way that \( F \) is represented by:

\[
F_i(x_1^i, x_2^i) = \begin{pmatrix} 1 \\ 0 \end{pmatrix}.
\] (9.19)

In other words we use coordinates where \( F \) is rectified. In these coordinates

\[
F_i^\perp(x_1^i, x_2^i) = \begin{pmatrix} a_i(x_1^i, x_2^i) \\ b_i(x_1^i, x_2^i) \end{pmatrix}.
\]

Let's assume that \( F \) is never vanishing.

Notice that to have that \( (9.19) \) holds in every coordinate chart, the only admitted change of coordinates have the form

\[
x_1^i = x_1^i + \alpha_{ji}(x_2^i),
\] (9.20)
\[
x_2^i = \beta_{ji}(x_2^i).
\] (9.21)

The Jacobian of each change of coordinates has then the form

\[
\begin{pmatrix} 1 & \alpha_{ji}' \\ 0 & \beta_{ji}' \end{pmatrix}.
\]

188
In each coordinate chart we have that $\omega$ is represented by
$$h_i(x^i_1, x^i_2)dx^i_1 \wedge dx^i_2.$$ 
Hence $\Phi = \omega(F, F^{\perp})$ is represented by
$$h_i(x^i_1, x^i_2)b_i(x^i_1, x^i_2)$$
Consider now a perturbation $F^{\perp\varepsilon}$ of $F^{\perp}$ that in each coordinate chart is
$$F^{\perp\varepsilon}_i(x^i_1, x^i_2) = \left( \frac{a_i(x^i_1, x^i_2)}{b_i(x^i_1, x^i_2)} + \frac{\varepsilon}{h_i(x^i_1, x^i_2)} \right). \quad (9.22)$$
Notice that equation (9.22) defines $F^{\perp\varepsilon}$ globally. Indeed for a change of coordinates of the type (9.20), (9.21) we have that
$$b_j = \beta^j_{ji}b_i, \quad \text{and} \quad h_j = \beta^j_{ji}^{-1}h_i.$$ 
It follows that in each coordinate chart $\omega(F, F^{\perp\varepsilon})$ is represented by
$$h_i(x^i_1, x^i_2)\left( b_i(x^i_1, x^i_2) + \frac{\varepsilon}{h_i(x^i_1, x^i_2)} \right) = h_i(x^i_1, x^i_2)b_i(x^i_1, x^i_2) + \varepsilon$$
Since this is true in each coordinate chart, we have that
$$\Phi_{\varepsilon} = \omega(F, F^{\perp\varepsilon}) = \Phi + \varepsilon.$$ 
Notice that by construction $F^{\perp\varepsilon}$ is close to $F^{\perp}$ in the $C^1$ norm and hence this is true also for $f$ and $f_{\varepsilon}$.
In the case in which $F$ has some zeros ...........

### 9.5 A Gauss-Bonnet theorem

For an compact orientable 2D-Riemannian manifold, the Gauss-Bonnet theorem asserts that the integral of the curvature is a topological invariant that is the Euler characteristic of the manifold (see Section 1.3).

This theorem admit an interesting generalization in the context of 2D almost-Riemannian structures that are fully orientable. This generalization is not trivial since one needs to integrate the Gaussian curvature (that in general is diverging while approaching to the singular set) on the manifold (that has always infinite volume).

This generalization holds under certain natural assumptions on the 2D almost-Riemannian structure, namely we will assume

\textbf{HG} : The base manifold $M$ is compact. The 2D almost-Riemannian structure is fully orientable, \textbf{H0} holds and every point of $Z$ is a Grushin point.

The hypotheses that the structure is fully orientable is crucial and it is the almost-Riemannian version of the classical orientability hypothesis that one need in Riemannian geometry. The hypothesis \textbf{H0} is the basic hypothesis to have a reasonable description of the asymptotics of $K$ in a neighborhood of $Z$. The hypothesis that every point is a Grushin point is a technical hypothesis. A version of a Gauss Bonnet Theorem in presence of Martinet points can also be written, but is more technical and outside the purpose of this book.

With an argument similar to the one of the beginning of Section 9.4.1 one get

\[189\]
Theorem 9.38. Hypothesis HG is open in the set of smooth map \( f : U \to TM \) endowed with \( C^1 \) topology:

Clearly hypothesis HG is not dense since Martinet points do not disappear for small \( C^1 \) perturbations of the system.

It is important to notice that HG is not empty. Indeed we have

Lemma 9.39. Every oriented compact surface can be endowed with an oriented almost-Riemannian structure satisfying the requirement that there are no Martinet points.

We are going to prove Lemma 9.39 in Section 9.5.2.

Definition 9.40. Consider a 2D almost-Riemannian structure \((U, f)\) over a 2D manifold \( M \) and assume that HG holds.

Let \( \nu \) a volume form for the Euclidean structure on \( U \), i.e. a never vanishing 2-form s.t. \( \nu(\sigma_1, \sigma_2) = 1 \) on every positive oriented local orthonormal frame for \((\cdot | \cdot)_q\). Let \( \Xi \) be an orientation on \( M \). We define:

- The signed area form \( dA^s \) on \( M \) as the two-form on \( M \setminus Z \) given by the pushforward of \( \nu \) along \( f \). Notice that the Riemannian area \( dA \) on \( M \setminus Z \) is the density associated to the volume form \( dA^s \).

- \( M^+ = \{ q \in M \setminus Z, \text{ s.t. the orientation given by } dA_q^s \text{ and } \Xi_q \text{ are the same} \} \)

- \( M^- = \{ q \in M \setminus Z, \text{ s.t. the orientation given by } dA_q^s \text{ and } \Xi_q \text{ are opposite} \} \).

Notice that given a measurable function \( h : \Omega \subset M^\pm \setminus Z \to \mathbb{R} \), we have

\[
\int_{\Omega} h \ dA_s = \pm \int_{\Omega} h \ dA \quad \text{(if it exists).} \tag{9.23}
\]

Definition 9.41. Under the same hypotheses of Definition 9.40 define

- \( M_\varepsilon = \{ q \in M \mid d(q, Z) > \varepsilon \} \) where \( d(\cdot, \cdot) \) is the 2D-almost-Riemannian structure on \( M \).

- \( M_\varepsilon^\pm = M_\varepsilon \cap M^\pm \)

- Given a measurable function \( h : M \setminus Z \to \mathbb{R} \), we say that it is AR-integrable if

\[
\lim_{\varepsilon \to 0} \int_{M_\varepsilon} h \ dA_s \quad \text{(9.24)}
\]

exists and is finite. In this case we denote such a limit by \( \int h \ dA_s \).

Remark 9.42. Notice that (9.24) is equivalent to

\[
\lim_{\varepsilon \to 0} \left( \int_{M_\varepsilon^+} h \ dA - \int_{M_\varepsilon^-} h \ dA \right)
\]

\( ^1 \text{i.e. } dA_q^s(F_1, F_2) = \alpha \Xi(F_1, F_2) \text{ with } \alpha > 0 \)
Example: the Grushin sphere

The Grushin sphere is the free 2D-almost Riemannian structure on the sphere \( S^2 = \{ y_1^2 + y_2^2 + y_3^2 = 1 \} \) for which an orthonormal frame is given by two orthogonal rotations for instance

\[
Y_1 = \begin{pmatrix} 0 \\ -y_3 \\ y_2 \end{pmatrix} \text{ (rotation along the } y_1 \text{ axis)} \tag{9.25}
\]

\[
Y_2 = \begin{pmatrix} -y_3 \\ 0 \\ y_1 \end{pmatrix} \text{ (rotation along the } y_2 \text{ axis)} \tag{9.26}
\]

In this case \( Z = \{ y_3 = 0, \ y_1^2 + y_2^2 = 1 \} \). Passing in spherical coordinates

\[
y_1 = \cos(x) \cos(\phi) \\
y_2 = \cos(x) \sin(\phi) \\
y_3 = \sin(x)
\]

and letting

\[
X_1 = \cos(\phi - \pi/2)Y_1 + \sin(\phi - \pi/2)Y_2 \\
X_2 = -\sin(\phi - \pi/2)Y_1 + \cos(\phi - \pi/2)Y_2
\]

we get that an orthonormal frame is given by

\[
X_1 = \begin{pmatrix} 0 \\ \tan(x) \end{pmatrix}, \quad X_2 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}
\]

Notice that the singularity at \( x = \pi/2 \) is due to the spherical coordinates. Instead \( Z = \{ x = 0 \} \). In this case we have.

\[
dA = \frac{1}{|\tan(x)|} dx \, d\phi, \quad dA_s = \frac{1}{\tan(x)} dx \wedge d\phi, \quad K = \frac{-2}{\sin(x)^2}
\]

The loci \( Z, M^\pm \), are illustrated in Figure 9.5.

The main result of this section is the following.

**Theorem 9.43.** Consider a 2D-almost-Riemannian structure satisfying hypothesis HG. Let \( dA^s \) be the signed area form and \( K \) be the Riemannian curvature, both defined on \( M \setminus Z \). Then \( K \) is \( AR\)-integrable and we have

\[
\int K \, dA^s = e(U)
\]

where \( e(U) \) denotes the Euler number of \( E \). Moreover we have

\[
e(U) = \chi(M^+) - \chi(M^-)
\]

where \( \chi(M^\pm) \) denotes the Euler characteristic of \( M^\pm \).
Notice that in the Riemannian case \( \int K dA^s \) is the standard integral of the Riemannian curvature and \( e(U) = \chi(M) \) since \( U = TM \). Hence Theorem 9.43 contains the classical Gauss-Bonnet theorem.

In a sense, in Riemannian geometry the topology of the surface gives a constraint on the total curvature, while in 2D almost-Riemannian geometry such constraints is determined by the topology of the bundle \( U \).

For a free almost-Riemannian structure we have that \( U \) is a rank 2 trivial bundle over \( M \). As a consequence we get that \( \int K dA^s = 0 \), generalizing what happens on the torus.

We could interpret this result in the following way. Take a metric that is determined by a single pair of vector fields. In the Riemannian context we are constrained to be parallelizable (i.e. we are constrained to be on the torus). In the AR context, \( M \) could be any compact orientable manifolds, but the metric is constrained to be singular somewhere. In any case, the integral of the curvature will be zero.

### 9.5.1 Proof of Theorem 9.43

The proof is divided in two steps. First we prove that \( \int K dA^s = \chi(M^+) - \chi(M^-) \). Then we prove that \( e(U) = \chi(M^+) - \chi(M^-) \)

#### Step 1

As a consequence of the compactness of \( M \) and of Lemma 9.16 one has:

**Lemma 9.44.** Assume that \( HG \) holds. Then the set \( Z \) is the union of finitely many curves diffeomorphic to \( S^1 \). Moreover, there exists \( \varepsilon_0 > 0 \) such that, for every \( 0 < \varepsilon < \varepsilon_0 \), we have that
$\partial M_\varepsilon$ is smooth and the set $M \setminus M_\varepsilon$ is diffeomorphic to $Z \times [0, 1]$.

Under $\text{HG}$ the almost-Riemannian structure can be described, around each point of $Z$, by a normal form of type (NF2).

Take $\varepsilon_0$ as in the statement of Lemma 9.44. For every $\varepsilon \in (0, \varepsilon_0)$, let $M^\pm_\varepsilon = M^\pm \cap M_\varepsilon$. By definition of $dA_s$ and $M^\pm_\varepsilon$,

$$\int_{M_\varepsilon} KdA_s = \int_{M^+_\varepsilon} KdA - \int_{M^-_\varepsilon} KdA.$$  

The Gauss-Bonnet formula asserts that for every compact oriented Riemannian manifold $(N, g)$ with smooth boundary $\partial N$, we have

$$\int_N KdA + \int_{\partial N} k_g ds = 2\pi \chi(N),$$

where $K$ is the curvature of $(N, g)$, $dA$ is the Riemannian density, $k_g$ is the geodesic curvature of $\partial N$ (whose orientation is induced by the one of $N$), and $ds$ is the length element.

Applying the Gauss-Bonnet formula to the Riemannian manifolds $(M^+_\varepsilon, g)$ and $(M^-_\varepsilon, g)$ (whose boundary smoothness is guaranteed by Lemma 9.44), we have

$$\int_{M_\varepsilon} KdA_s = 2\pi(\chi(M^+_\varepsilon) - \chi(M^-_\varepsilon)) - \int_{\partial M^+_\varepsilon} k_g ds + \int_{\partial M^-_\varepsilon} k_g ds. \quad (9.27)$$

Thanks again to Lemma 9.44, $\chi(M^\pm_\varepsilon) = \chi(M^\pm)$. We are left to prove that

$$\lim_{\varepsilon \to 0} \left( \int_{\partial M^+_\varepsilon} k_g ds - \int_{\partial M^-_\varepsilon} k_g ds \right) = 0. \quad (9.28)$$

Fix $q \in Z$ and a (NF2)-type local system of coordinates $(x_1, x_2)$ in a neighborhood $U_q'$ of $q$. We can assume that $U_q'$ is given, in the coordinates $(x_1, x_2)$, by a rectangle $[-a, a] \times [-b, b]$, $a, b > 0$. Assume that $\varepsilon < a$. Notice that $Z \cap U_q = \{0\} \times [-b, b]$ and $\partial M_\varepsilon \cap U_q = \{-\varepsilon, \varepsilon\} \times [-b, b]$.

We are going to prove that

$$\int_{\partial M_\varepsilon \cap U_q} k_g \, ds = O(\varepsilon). \quad (9.29)$$
Thus, $\{0, a\} \times [-b, b]$. Therefore, $M^+_\varepsilon$ induces on $\partial M^+_\varepsilon = \{\varepsilon\} \times [-b, b]$ a downwards orientation (see Figure 9.5.1). The curve $s \mapsto c(s) = (\varepsilon, x_2(s))$ satisfying

$$
\dot{c}(s) = -F_2(c(s)), \quad c(0) = (\varepsilon, 0),
$$

is an oriented parametrization by arclength of $\partial M^+_\varepsilon$, making a constant angle with $F_1$. Let $(\theta_1, \theta_2)$ be the dual basis to $(F_1, F_2)$ on $U_q \cap M^+$, i.e., $\theta_1 = dx_1$ and $\theta_2 = x_1^{-1}e^{-\phi(x_1, x_2)}dx_2$. According to [?], Corollary 3, p. 389, Vol. III], the geodesic curvature of $\partial M^+_\varepsilon$ at $c(s)$ is equal to $\lambda(\dot{c}(s))$, where $\lambda \in \Lambda^1(U_q)$ is the unique one-form satisfying

$$
d\theta_1 = \lambda \wedge \theta_2, \quad d\theta_2 = -\lambda \wedge \theta_1.
$$

A trivial computation shows that

$$
\lambda = \partial x_1(x_1^{-1}e^{-\phi(x_1, x_2)})dx_2.
$$

Thus,

$$
k_\eta(c(s)) = -\partial x_1(x_1^{-1}e^{-\phi(c(s))})(dx_2(F_2))(c(s)) = \frac{1}{\varepsilon} + \partial x_1\phi(\varepsilon, x_2(s)).
$$

Denote by $L_1$ and $L_2$ the lengths of, respectively, $\{\varepsilon\} \times [0, b]$ and $\{\varepsilon\} \times [-b, 0]$. Then,

$$
\int_{\partial M^+_\varepsilon \cap U_q} k_\eta ds = \int_{-L_1}^{L_2} k_\eta(c(s))ds
$$

$$
= \int_{-L_1}^{L_2} \left( \frac{1}{\varepsilon} + \partial x_1\phi(\varepsilon, s) \right) \frac{1}{\varepsilon} \frac{1}{e^{\phi(\varepsilon, x_2)}}dx_2,
$$

where the last equality is obtained taking $x_2 = x_2(-s)$ as the new variable of integration.

We reason similarly on $\partial M^- \cap U_q$, on which $M^- \varepsilon$ induces the upwards orientation. An orthonormal frame on $M^- \cap U_q$, oriented consistently with $M$, is given by $(F_1, -F_2)$, whose dual basis is $(\theta_1, -\theta_2)$. The same computations as above lead to

$$
\int_{\partial M^- \cap U_q} k_\eta ds = \int_{-b}^{b} \left( \frac{1}{\varepsilon} - \partial x_1\phi(-\varepsilon, x_2) \right) \frac{1}{\varepsilon} e^{\phi(-\varepsilon, x_2)}dx_2.
$$

Define

$$
F(\varepsilon, x_2) = (1 + \varepsilon \partial x_1\phi(\varepsilon, x_2))e^{-\phi(\varepsilon, x_2)}, \quad (9.30)
$$

Then

$$
\int_{\partial M^+_\varepsilon \cap U_q} k_\eta ds - \int_{\partial M^- \cap U_q} k_\eta ds = \frac{1}{\varepsilon^2} \int_{-b}^{b} (F(\varepsilon, x_2) - F(-\varepsilon, x_2)) dx_2.
$$

By Taylor expansion with respect to $\varepsilon$ we get

$$
F(\varepsilon, x_2) - F(-\varepsilon, x_2) = 2\partial_\varepsilon F(0, x_2)\varepsilon + O(\varepsilon^3) = O(\varepsilon^3)
$$

194
where the last equality follows from the relation \( \partial_{\varepsilon} F(0, x_2) = 0 \) (see equation (9.30)). Therefore,

\[
\int_{\partial M^+ \cap U_q} k_g ds - \int_{\partial M^- \cap U_q} k_g ds = O(\varepsilon),
\]

and (9.29) is proved.

**Step 2**

The idea of the proof is to find a section \( \sigma \) of \( SE \) with isolated singularities \( p_1, \ldots, p_m \) such that \( \sum_{j=1}^m i(p_j, \sigma) = \chi(M^+) - \chi(M^-) + \tau(S) \). In the sequel, we consider \( Z \) to be oriented with the orientation induced by \( M^+ \).

### 9.5.2 Construction of trivializable 2-ARSs with no tangency points

In this section we prove Lemma 9.39 by showing how to construct a trivializable 2-ARS with no tangency points on every compact orientable two-dimensional manifold.

Without loss of generality we can assume \( M \) connected. For the torus, an example of such structure is provided by the standard Riemannian one. The case of a connected sum of two tori can be treated by gluing together two copies of the pair of vector fields \( F_1 \) and \( F_2 \) represented in Figure 9.5.2A, which are defined on a torus with a hole cut out. In the figure the torus is represented as a square with the standard identifications on the boundary. The vector fields \( F_1 \) and \( F_2 \) are parallel on the boundary of the disk which has been cut out. Each vector field has exactly two zeros and the distribution spanned by \( F_1 \) and \( F_2 \) is transversal to the singular locus. Examples on the connected sum of three or more tori can be constructed similarly by induction. The resulting singular locus is represented in Figure 9.5.2B.

We are left to check the existence of a trivializable 2-ARS with no tangency points on a sphere. A simple example can be found in the literature and arises from a model of control of quantum systems (see [7, 8]). Let \( M \) be a sphere in \( \mathbb{R}^3 \) centered at the origin and take \( F_1(x, y, z) = (y, -x, 0) \),
$F_2(x, y, z) = (0, z, -y)$ as orthonormal frame. Then $F_1$ (respectively, $F_2$) is an infinitesimal rotation around the third (respectively, first) axis. The singular locus is therefore given by the intersection of the sphere with the plane $\{y = 0\}$ and none of its points exhibit tangency (see Figure 9.5.2). Notice that hypothesis HG is satisfied.
In this chapter, for a point \( q \in M \), the symbol \( \Omega_q \) denotes the set of smooth curves \( \gamma \) on \( M \) that are based at \( q \), that is \( \gamma(0) = q \).

### 10.1 Jet spaces

Fix \( q \) in \( M \) and a curve \( \gamma \in \Omega_q \). In every coordinate chart it is meaningful to write the Taylor expansion

\[
\gamma(t) = q + \dot{\gamma}(0)t + O(t^2) \quad (10.1)
\]

The tangent vector \( v \in T_qM \) to \( \gamma \) at \( t = 0 \) is by definition the equivalence class of curves in \( \Omega_q \) such that, in some coordinate chart, they have the same 1-st order Taylor polynomial. (This requirement indeed implies that the same is true for every coordinate chart, by the chain rule.)

In the same spirit we can consider, given a smooth curve such that \( \gamma(0) = q \), its \( m \)-th order Taylor polynomial at \( q \)

\[
\gamma(t) = q + \dot{\gamma}(0)t + \ddot{\gamma}(0)\frac{t^2}{2} + \ldots + \gamma^{(m)}(0)\frac{t^m}{m!} + O(t^{m+1}) \quad (10.2)
\]

**Exercise 10.1.** Let \( \gamma, \gamma' \in \Omega_q \). We say that \( \gamma \) is \( (m-) \)-equivalent to \( \gamma' \) at \( q \), and we write \( \gamma \sim_{q,m} \gamma' \), if their Taylor polynomial at \( q \) of order \( m \) in some coordinate chart coincide. Prove that \( \sim_{q,m} \) is a well-defined equivalence relation on the set of curves based at \( q \).

**Definition 10.2.** Let \( m > 0 \) be an integer and \( q \in M \). We define the set of \( m \)-th jets of curves at point \( q \in M \) as the equivalence classes of curves based at \( q \) with respect to \( \sim_{q,m} \). We denote with \( J^m_q \gamma \) the equivalence class of a curve \( \gamma \) and with

\[
J^m_q := \{ J^m_q \gamma : \gamma \in \Omega_q \}
\]

**Remark 10.3.** From coordinates representation \( (10.2) \), one can prove that \( J^m_q \) is a smooth manifold and \( \dim J^m_q = mn \). Indeed the \( m \)-th order Taylor polynomial is characterized by the \( n \)-dimensional vectors \( \gamma^{(i)}(0) \) for \( i=1,\ldots,m \) (cf. \( (10.2) \)).

In the following we always assume that \( q \in M \) is fixed together with a coordinate chart around \( q \), where \( q = 0 \). The Taylor expansion of a curve \( \gamma \in \Omega_q \) is then written as follows

\[
J^m_q \gamma = \sum_{i=1}^{m} \gamma^{(i)}(0)\frac{t^i}{i!}.
\]

197
To better understand the structure of \( J_m^q \) as a smooth manifold we consider the map which “forget about” the \( m \)-th derivative

\[
\Pi_{m-1}^m : J_m^q \rightarrow J_{m-1}^q
\]

\[
\sum_{i=1}^m \gamma^{(i)}(0) \frac{t^i}{i!} \mapsto \sum_{i=1}^{m-1} \gamma^{(i)}(0) \frac{t^i}{i!}
\]

**Proposition 10.4.** \( J_m^q \) is an affine bundle over \( J_{m-1}^q \) with projection \( \Pi_{m-1}^m \), whose fibers are affine spaces over \( T_qM \).

**Proof.** Fix an element \( j \in J_{m-1}^q \), then the fiber \( (\Pi_{m-1}^m)^{-1}(j) \) is the set of all \( m \)-th-jets with fixed \((m-1)\)-th jet equal to \( j \). To show that it is an affine space over \( T_qM \) we should define the sum of a tangent vector and an \( m \)-th-jet, with \((m-1)\)-th-jet fixed, having as a result another \( m \)-th-jet with the same \((m-1)\)-th-jet.

Let \( j = J_m^q \gamma \) be the \( m \)-th-jet of a smooth curve in \( M \) and let \( v \in T_qM \). Consider a smooth vector field \( V \in \text{Vec}(M) \) such that \( V(q) = v \) and define the sum

\[
J_m^q \gamma + v := J_m^q (\gamma v), \quad \gamma v(t) = e^{tmV}(\gamma(t)) \quad (10.3)
\]

It is easy to see that, due to the presence of the power \( t^m \), the \((m-1)\)-th Taylor polynomial of \( \gamma \) and \( \gamma v \) coincide. Indeed

\[
J_m^q(e^{tmV}(\gamma(t))) = J_m^q \gamma + tmV(q)
\]

Hence the sum (10.3) gives to \( (\Pi_{m-1}^m)^{-1}(j) \) the structure of affine space over \( T_qM \). Indeed it is enough to check that the definition does not depend on the representative.

The geometric meaning of the fact that \( J_m^q \) is an affine bundle (and not an vector bundle) is that we cannot complete in a canonic way a \((m-1)\)-th-jet to a \( m \)-th-jet, i.e. we cannot fix an origin in the fiber. On the other hand there exists a sort of “global” origin on \( J_m^q \), that is the jet of the constant curve equal to \( q \).

Now we want to define dilations on jet spaces, analogously to homothety in Euclidean spaces. Since we have no vector space structure we have to find an appropriate notion

**Definition 10.5.** Let \( \alpha \in \mathbb{R} \) and define \( \gamma_\alpha(t) := \gamma(\alpha t) \) for every \( t \in \mathbb{R} \). Define the *dilation* of factor \( \alpha \) on \( J_m^q \) as

\[
\delta_\alpha : J_m^q \rightarrow J_m^q, \quad \delta_\alpha(J_m^q \gamma) = J_m^q (\gamma_\alpha)
\]

One can check that this definition does not depend on the representative and, in coordinates, it is written as a *quasi-homogeneous* multiplication

\[
\delta_\alpha \left( \sum_{i=1}^m t^i \xi_i \right) = \sum_{i=1}^m t^i \alpha^i \xi_i
\]

Next we extend the notion of jets also for vector fields. To start with we consider flows on the manifold.

**Definition 10.6.** A *flow* on \( M \) is a smooth family of diffeomorphisms

\[
P = \{ P_t \in \text{Diff}(M), \ t \in \mathbb{R} \}, \quad P_0 = \text{Id}
\]
Notice that we do not require the family to be a one parametric group (i.e., the group law $P_t \circ P_s = P_{t+s}$ is not satisfied) and this in general is characterized as the flow of the nonautonomous vector field

$$X_t := \frac{d}{d\varepsilon} \bigg|_{\varepsilon=0} P_{t+\varepsilon} \circ P_t^{-1}.$$  

The set of all flows on $M$ is a group with the point-wise product, i.e. the product of the flows $P = \{P_t\}$ and $Q = \{Q_t\}$ is given by

$$(P \circ Q)_t := P_t \circ Q_t$$

Clearly we can act with a flow on a smooth curve on $M$ as follows: $(P\gamma)(t) = P_t(\gamma(t))$. Moreover, since $P_0 = \text{Id}$, every flow defines a map on $\Omega_q$.

This action is well-behaved with respect to equivalence relations $\sim_{m,q}$, i.e., it defines a map on $J^m_q$. Indeed if $\gamma \in \Omega_q$, then $P\gamma \in \Omega_q$ and from the chain rule it follows that $J^m_q(P\gamma)$ depends only on first $m$ derivatives of $\gamma$ at $q$, i.e., on $J^m_q\gamma$.

**Definition 10.7.** Let $P$ be a smooth flow on $M$. The action of $P$ on $J^m_q\gamma$ is defined by

$$P_j := J^m_q(P\gamma), \quad \text{if } j = J^m_q\gamma.$$  

It can be easily checked that the definition is well-posed and $(P \circ Q)j = P(Qj)$ for every $j \in J^m_q$.

**Jets of vector fields**

Given a vector field $V \in \text{Vec}(M)$ we want to define its $m$th-jet $J^m_qV$ which should be naturally an element of $\text{Vec}(J^m_q)$.  

Let us denote with $P_V = \{e^{tV}\}$ the 1-parametric group defined by the flow of $V$. As we explained we can act on jets

$$P_V : j \mapsto e^{tV}(j)$$

To act on a family of curves we need a family of flows, then let us consider the 1-parametric group of flows $P^*_V = \{e^{sV}\}$

**Definition 10.8.** For every $V \in \text{Vec}(M)$ we define the vector field $J^m_qV \in \text{Vec}(J^m_q)$ is the section $J^m_qV : J^m_q \to T J^m_q$ defined as follows

$$(J^m_qV)(J^m_q\gamma) := \frac{\partial}{\partial s} \bigg|_{s=0} P^*_V(J^m_q\gamma) = \frac{\partial}{\partial s} \bigg|_{s=0} J^m_q(e^{tsV}(\gamma(t))) \quad (10.4)$$

**Exercise 10.9.** Prove the following formula for every $V \in \text{Vec}(M)$

$$(J^m_qV)(J^m_q\gamma) = \sum_{i=1}^{m} \frac{t^i}{i!} \frac{d}{dt}l(tV(\gamma(t)))$$

where $V$ is identified with a vector function $V : \mathbb{R}^n \to \mathbb{R}^n$ in coordinates.

To end this section we study the interplay between dilations and jets of vector fields. Since $\delta_\alpha$ is a map on $J^m_q$ its differential $(\delta_\alpha)_*\gamma$, acts on elements of $\text{Vec}(J^m_q)$, and in particular on jets of vector fields on $M$. Surprisingly, its action on these fields is linear with respect to $\alpha$.  

199
Proposition 10.10. For every $\alpha \in \mathbb{R}$ and $V \in \text{Vec}(M)$ one has

$$(\delta_\alpha)_*(J^m_qV) = J^m_q(\alpha V) = \alpha J^m_qV.$$ 

Proof. From the very definition of the differential of a map (see also Chapter 2) we have

$$((\delta_\alpha)_*J^m_qV)(J^m_q\gamma) = \frac{\partial}{\partial s} \bigg|_{s=0} J^m_q(\delta_\alpha e^{tsV} \delta_1/\alpha(\gamma(t)))$$

$$= \frac{\partial}{\partial s} \bigg|_{s=0} J^m_q(\delta_\alpha e^{tsV}(\gamma(t/\alpha)))$$

$$= \frac{\partial}{\partial s} \bigg|_{s=0} J^m_q(e^{tsV}(\gamma(t)))$$

$$= J^m_q(\alpha V) = \alpha J^m_qV$$

\square

10.2 Admissible variations

In this section we define the appropriate notion of tangent vector to a sub-Riemannian manifold. Our goal is to define the “tangent structure” to a sub-Riemannian one.

As usual, we assume that the sub-Riemannian structure is defined by the generating family $\{f_1, \ldots, f_m\}$. Admissible curves on $M$ are maps $\gamma : [0, T] \to M$ such that there exists a control function $u \in L^\infty$ such that

$$\dot{\gamma}(t) = f_u(t)(\gamma(t)) = \sum_{i=1}^m u(t_i)f_i(\gamma(t)).$$

To have a good definition of tangent vector we could not restrict to family of admissible curves, because in this way we lose all the information about directions that are not in the distribution. Indeed we want the tangent space to be a first order approximation of the structure, containing informations about all directions.

We need a proper definition of tangent vector, that means a proper definition of “variation of a point”, in order to give a precise meaning to its “principal term”, that is going to be the tangent vector.

We now introduce the notion of smooth admissible variation.

Definition 10.11. A curve $\gamma : [0, T] \to M$ in $\Omega_q$ is said a smooth admissible variation if there exists a family of controls $\{u(t, s)\}_{s \in [0, \tau]}$ such that

(i) $u(t, \cdot)$ is measurable and essentially bounded for all $t \in [0, T]$,

(ii) $u(\cdot, s)$ is smooth with bounded derivatives, for all $s \in [0, \tau]$,

(iii) $u(0, s) = 0$ for all $s \in [0, \tau]$,

(iv) $\gamma(t) = q \circ \exp \int_0^t f_u(t, s)ds$
In other words $\gamma$ is a smooth admissible variation (or shortly, admissible variation) it can be parametrized as the final point of a smooth family of admissible curves. We stress that an admissible variation is not an admissible curve, in general.

Remark 10.12. Recall that two distributions are said to be equivalent (see also Definition 3.3 and 3.17) if and only if the corresponding modulus of horizontal vector fields are isomorphic $\mathcal{D} \simeq \mathcal{D}'$, where

$$\mathcal{D} = \text{span}\{f(\sigma), \sigma \text{ smooth section of } U\}.$$ 

is finitely generated by a basis $f_1, \ldots, f_m$.

Let us show that the definition of admissible variation does not depend on the frame $f_1, \ldots, f_m$. By definition any admissible variation $\gamma(t)$ is associated with a family $q(t, s)$, for $s \in [0, \tau]$ solution of

$$\frac{\partial}{\partial s} q(t, s) = \sum_{i=1}^{m} u_i(t, s) f_i(q(t, s)), \quad (10.5)$$

such that $\gamma(t) = q(t, \tau)$. Assume that $\tilde{f}_1, \ldots, \tilde{f}_m$ is another set of local generators of the modulus. Then there exist functions $a_{ij} \in C^\infty(M)$ such that

$$f_i(q) = \sum_{j=1}^{m} a_{ij}(q) \tilde{f}_j(q), \quad \forall q \in M, \quad \forall i = 1, \ldots, m. \quad (10.6)$$

and assume that $\gamma$ is an admissible variation with respect to $u(t, s)$, i.e., it satisfies (10.6).

Now we prove that there exist a family $\tilde{u}(t, s)$ of controls such that $\gamma$ is an admissible variation in the new frame. From (10.6) we get

$$\sum_{i=1}^{m} u_i(t, s) f_i(q) = \sum_{i,j=1}^{m} u_i(t, s) a_{ij}(q) \tilde{f}_j(q)$$

Then we could define, using the solution $q(t, s)$ of (10.5), the new family of controls

$$\tilde{u}_j(t, s) = \sum_{i=1}^{m} u_i(t, s) a_{ij}(q(t, s))$$

and we see from identities above that

$$\frac{\partial}{\partial s} q(t, s) = \sum_{i=1}^{m} \tilde{u}_j(t, s) \tilde{f}_j(q(t, s)), \quad s \in [0, \tau] \quad (10.7)$$

Assumption. From now on, we assume that the sub-Riemannian structure is bracket generating at $q$ with step $m$, i.e. $\mathcal{D}_q^m = T_qM$.

Definition 10.13. Let $(M, U, f)$ be a sub-Riemannian structure. The set of admissible jets with respect to the sub-Riemannian structure is

$$J_q^f := \{J_q^m \gamma, \gamma \in \Omega_q \text{ is an admissible variation}\}$$

201
Example 10.14. Consider two vector fields \(X, Y \in \text{Vec}(M)\) and the curve 
\[
\gamma : [0, T] \to M, \quad \gamma(t) = e^{-tY} \circ e^{-tX} \circ e^{tY} \circ e^{tX}(q)
\]
It is easily seen that \(\gamma\) is an admissible variation if we set 
\[
\gamma(t) = \exp \int_0^4 f_{tv(s)}(q) ds
\]
where 
\[
v(s) = \begin{cases} 
(1, 0), & \text{if } s \in [0, 1], \\
(0, 1), & \text{if } s \in [1, 2], \\
(-1, 0), & \text{if } s \in [2, 3], \\
(0, -1), & \text{if } s \in [3, 4].
\end{cases}
\]
In coordinates we have expansion \(\gamma(t) = q + t^2[X,Y](q) + o(t^2)\).

Now we want to introduce the nonholonomic tangent space in a coordinate-free way. In the next section we will see how it can be described in some special set of coordinates.

Definition 10.15. The group of flows of admissible variations is 
\[
\mathcal{P}^f := \left\{ \exp \int_0^\tau f_{u(t,s)} ds, \ u(t,s) \text{ smooth variation} \right\}
\]
Any admissible variation is given by \(\gamma(t) = P_t(q)\) for some \(P \in \mathcal{P}^f\), where we identify \(q\) with the constant curve.

Remark 10.16. \(\mathcal{P}^f\) is a group, indeed the following equality holds 
\[
\exp \int_0^{\tau_1} f_{u(t,s)} ds \circ \exp \int_0^{\tau_2} f_{v(t,s)} ds = \exp \int_0^{\tau_1+\tau_2} f_{w(t,s)} ds
\]
where 
\[
w(t,s) = \begin{cases} 
u(t,s), & 0 \leq s \leq \tau_1, \\
v(t,s-\tau_1), & \tau_1 \leq s \leq \tau_1 + \tau_2.
\end{cases}
\]
is the concatenation of controls. Then we have that 
\[
J_q^f = \{J_q^m(P(q)), P \in \mathcal{P}^f\}
\]
is exactly the orbit of \(q\) under the action of the group \(\mathcal{P}^f\).

The nonholonomic tangent space is the quotient of \(\mathcal{P}^f\) with respect to the action of the subgroup of “slow” flows. Heuristically, a flow is slow if the first nonzero jet \(J_q^i \gamma\) of its associated trajectory \(\gamma\) belongs to a subspace \(\mathcal{D}^j\), with \(j < i\).

Definition 10.17. Let \(Q \in \mathcal{P}^f\). \(Q\) is said to be a slow flow if it is associated to a smooth variation \(u(t,s)\) such that satisfies \(u(0,s) = \frac{\partial u}{\partial t}(0,s) = 0\).

The subgroup of slow flows is the normal subgroup \(\mathcal{P}^f_0\) of \(\mathcal{P}^f\) generated by slow flows, i.e.
\[
\mathcal{P}^f_0 := \left\{ (P_t)^{-1} \circ Q_t \circ P_t : P \in \mathcal{P}^f, \ Q \text{ slow flow} \right\}
\]
Remark 10.18. Notice that, by definition of slow flow and the linearity of \( f \), a slow flow is associated
with a family of control that can be written in the form \( u(t,s) = tv(t,s) \), where \( v(0,s) = 0 \). Moreover we have
\[
P_t = \exp \int_0^t f_{u(t,s)} ds = \exp \int_0^t f_{tv(t,s)} ds = \exp \int_0^t t f_{v(t,s)} ds = t \exp \int_0^t f_{v(t,s)} ds,
\]
In other words a slow flow \( Q \in \mathcal{P}_f^0 \) is of the form \( Q_t = tP_t \) for some \( P \in \mathcal{P}_f^f \).

Exercise 10.19. Let \( j = J_q^m \gamma \) and \( j' = J_q^m \gamma' \) for some \( \gamma, \gamma' \in \Omega_q \). Prove that
\[
J_q^m \gamma \sim J_q^m \gamma', \quad \text{if} \quad \gamma'(t) = P_t(\gamma(t)) \tag{10.8}
\]
for some \( P \in \mathcal{P}_f^f \) is a well defined equivalence relation on \( J_q^f \).

Definition 10.20. The nonholonomic tangent space \( T_q^f \) is defined as
\[
T_q^f := J_q^f / \sim
\]
where \( \sim \) is the equivalence relation \( (10.8) \).

Proposition 10.21. Let \( X \in \mathcal{D} \) be an horizontal vector field for the sub-Riemannian structure on \( M \). The action of the one parametric group \( e^{tX} \) on \( J_q^f \) defined by \( (10.4) \) passes to the quotient with respect to the equivalence classes with respect to \( \sim \).

Proof. From the very definition of \( J_q^f \) it is easy to see that if \( J_q^m \gamma \) is the jet of an admissible variation then the right hand side of \( (10.4) \) is an admissible variation for every \( s \). We are left to show that if
\[
\gamma(t) \sim \gamma'(t) \quad \Rightarrow \quad e^{tX} \gamma(t) \sim e^{tX} \gamma'(t).
\]
From our assumption we get \( \gamma'(t) = \gamma(t) \circ Q_t \) for a slow flow \( Q \in \mathcal{P}_0^f \). It follows that
\[
\gamma'(t) \circ e^{tX} = \gamma(t) \circ Q_t \circ e^{tX} = \gamma(t) \circ e^{tX} \circ e^{-tX} \circ Q_t \circ e^{tX} = (\gamma(t) \circ e^{tX}) \circ \tilde{Q}_t
\]
where \( \tilde{Q}_t := e^{-tX} \circ Q_t \circ e^{tX} \) is also a slow flow. This shows that \( e^{tX} \) is independent on the representative and its action is well defined on the quotient. \( \square \)

10.3 Nilpotent approximation and privileged coordinates

In this section we want to introduce some coordinates in which we have a good description of the nonholonomic tangent space.

Consider some non negative integers \( k_1, \ldots, k_m \) such that \( n = k_1 + \ldots + k_m \) and the splitting
\[
\mathbb{R}^n = \mathbb{R}^{k_1} \oplus \ldots \oplus \mathbb{R}^{k_m}, \quad x = (x_1, \ldots, x_m)
\]
where every \( x_i = (x_i^1, \ldots, x_i^{k_i}) \in \mathbb{R}^{k_i} \).
The space $\text{Der}(\mathbb{R}^n)$ of all differential operators in $\mathbb{R}^n$ with smooth coefficients form an associative algebra with composition of operators as multiplication. The differential operators with polynomial coefficients form a subalgebra of this algebra with generators $1, x_i^j, \frac{\partial}{\partial x_i^j}$, where $i = 1, \ldots, m; j = 1, \ldots, k_i$. We define weights of generators as

$$\nu(1) = 0, \quad \nu(x_i^j) = i, \quad \nu\left(\frac{\partial}{\partial x_i^j}\right) = -i.$$ 

Then for any monomial

$$\nu\left(y_1 \cdots y_\alpha \frac{\partial^\beta}{\partial z_1 \cdots \partial z_\beta}\right) = \sum_{i=1}^\alpha \nu(y_i) - \sum_{j=1}^\beta \nu(z_j).$$

We say that a polynomial differential operator $D$ is \textit{homogeneous} if it is a sum of monomial terms all of same weight. We stress that this definition depends on the coordinate set and the choice of the weights.

\textbf{Lemma 10.22.} Let $D_1, D_2$ be two homogeneous differential operators. Then $D_1 \circ D_2$ is homogeneous and

$$\nu(D_1 \circ D_2) = \nu(D_1) + \nu(D_2) \quad (10.9)$$

\textbf{Proof.} It is sufficient to check formula (10.9) for monomials of kind $D_1 = \frac{\partial}{\partial x_i^j}$ and $D_2 = x_i^j$. This follows from the identity

$$\frac{\partial}{\partial x_i^j_1} \circ x_i^j_2 = x_i^j_2 \frac{\partial}{\partial x_i^j_1} + \frac{\partial x_i^j_2}{\partial x_i^j_1}.$$ 

\hfill $\Box$

A special case is when we consider, as differential operators, vector fields.

\textbf{Corollary 10.23.} If $V_1, V_2 \in \text{Vec}(\mathbb{R}^n)$ are homogeneous vector fields then $[V_1, V_2]$ is homogeneous and $\nu([V_1, V_2]) = \nu(V_1) + \nu(V_2)$.

With these properties we can define a filtration in the space of all smooth differential operators. Indeed we can write (in multiindex notation)

$$D = \sum_\alpha \varphi_\alpha(x) \frac{\partial^{|\alpha|}}{\partial x^\alpha}$$

Considering the Taylor expansion at 0 of every coefficient we can split $D$ as a sum of its homogeneous components

$$D \approx \sum_{i=-\infty}^{\infty} D^{(i)}$$

and define the filtration

$$\mathcal{D}^{(h)} = \{ D \in \text{Der}(\mathbb{R}^n) : D^{(i)} = 0, \forall i < h \}, \quad h \in \mathbb{Z}$$
It is easy to see that it is a decreasing filtration, i.e. $\mathcal{D}^{(h)} \subset \mathcal{D}^{(h-1)}$ for every $h$, and if we restrict our attention to vector fields we get

$$V \in \text{Vec}(\mathbb{R}^n) \Rightarrow V^{(i)} = 0, \quad \forall i < -m$$

Indeed every monomial of a $N^{th}$-order differential operator has weight not smaller than $-mN$. In other words we have

(i) $\text{Vec}(\mathbb{R}^n) \subset \mathcal{D}^{(-m)}$,

(ii) $V \in \text{Vec}(\mathbb{R}^n) \cap \mathcal{D}^{(0)}$ implies $V(0) = 0$.

and every vector field that is not zero at the origin is at least in $\mathcal{D}^{(-1)}$. This motivates the following definition

**Definition 10.24.** A system of coordinates near the point $q$ is said linearly adapted to the flag $\mathcal{D}^1_q \subset \mathcal{D}^2_q \subset \ldots \subset \mathcal{D}^m_q$ if

$$\mathcal{D}^i_q = \mathbb{R}^{k_1} \oplus \ldots \oplus \mathbb{R}^{k_i}, \quad \forall i = 1, \ldots, m. \tag{10.10}$$

A system of coordinates near the point $q$ is said privileged if it is linearly adapted to the flag and $f \in \mathcal{D}^{(-1)}$ for every $f \in \mathcal{D}$.

Notice that condition (i) can always be satisfied after a suitable linear change of coordinates.

Condition (ii) says that each horizontal vector field has no homogeneous component of degree less than $-1$.

**Example 10.25.** We analyze the meaning of privileged coordinates in the basic cases $m = 1, 2$ and then for $m = 3$ we show that in general not all system of linearly adapted coordinates are privileged.

1. If $m = 1$ all sets of coordinates are privileged because $\text{Vec}(M) \subset \mathcal{D}^{(-1)}$ since $\nu(x_i) = -1$ for all $i$.

2. If $m = 2$ then all systems of coordinates that are linearly adapted to the flag are privileged. Indeed, since $\nu(x_1) = -1$ and $\nu(x_2) = -2$, a vector field belonging to $\mathcal{D}^{(-2)} \setminus \mathcal{D}^{(-1)}$ must contain a monomial vector field of the kind $x_2 \partial x_2$, with constant coefficients. On the other hand a vector field $f \in \mathcal{D}$ cannot contain such a monomial since, by our assumption $f(0) \in \mathcal{D}^0 = \mathbb{R}^{k_1}$.

3. Let us consider the following set of vector fields in $\mathbb{R}^3 = \mathbb{R} \oplus \mathbb{R} \oplus \mathbb{R}$

$$f_1 = \partial_{x_1} + x_1 \partial_{x_3}, \quad f_2 = x_1 \partial_{x_2}, \quad f_3 = x_2 \partial_{x_3}$$

and set $\nu(x_i) = i$ for $i = 1, 2, 3$. The nontrivial commutators between these vector fields are

$$[f_1, f_2] = \partial_{x_2}, \quad [f_2, f_3] = x_1 \partial_{x_3}, \quad [[f_1, f_2], f_3] = \partial_{x_1}.$$

Then the flag (computed at $x = 0$) is given by

$$\mathcal{D}^1_0 = \text{span}\{\partial_{x_1}\}, \quad \mathcal{D}^2_0 = \text{span}\{\partial_{x_1}, \partial_{x_2}\}, \quad \mathcal{D}^3_0 = \text{span}\{\partial_{x_1}, \partial_{x_2}, \partial_{x_3}\}.$$  

These coordinates are then linearly adapted to the flag but they are not privileged since $\nu(x_1 \partial_{x_3}) = -2$ and $f_1 \in \mathcal{D}^{(-2)} \setminus \mathcal{D}^{(-1)}$. 

205
Theorem 10.26. Let $M$ be a sub-Riemannian manifold and $q \in M$. There always exists a system of privileged coordinates around $q$.

We postpone the proof of this theorem to the end of this section, after having analyzed in more detail the structure of privileged coordinates.

Theorem 10.27. Let $M$ be a sub-Riemannian manifold and $q \in M$. In privileged coordinates we have the following

(i) $J^f_q = \{ \sum_{i=1}^m t^i \xi_i, \xi_i \in \mathcal{D}^i_q \}$ and $\dim J^f_q = mk_1 + (m-1)k_2 + \ldots + k_m$.

(ii) Let $j_1, j_2 \in J^f_q$. Then $j_1 \sim j_2$ if and only if $j_1 - j_2 = \sum_{i=1}^m t^i \eta_i$, where $\eta_i \in \mathcal{D}^{i-1}_q$.

First part of proof of Theorem 10.27. We start by proving the inclusion $J^f_q \subset \{ \sum_{i=1}^m t^i \xi_i, \xi_i \in \mathcal{D}^i_q \}$.

For any smooth variation $\gamma(t)$ we can write

$$\gamma(t) = q \circ \exp \int_0^t f_u(t,s) ds$$

Taylor expansion leads to

$$\gamma(t) = q + \sum_{j=1}^i \int_{0 \leq s_j \leq \ldots \leq s_1 \leq s} q \circ f_u(t,s_1) \circ \ldots \circ f_u(t,s_j) ds_1 \ldots ds_j + O(t^{i+1})$$

Indeed using the fact that $f$ is linear in $u$, we can factor out $t$ from every term since $u(0,s) = 0$. If we want compute our curve in privileged coordinates (to compute weights) it is sufficient to apply all to the coordinate function. In particular, since by definition of privileged coordinates $f_u \in \mathcal{D}^{(-1)}$ for each $u$, we have that

$$f_u(t,s_1) \circ \ldots \circ f_u(t,s_j) \in \mathcal{D}^{(-j)}$$

and applying to a coordinate function $x_\alpha^\beta$, where $\alpha = 1, \ldots, m$ and $\beta = 1, \ldots, k_\alpha$ we have

$$f_u(t,s_1) \circ \ldots \circ f_u(t,s_j)x_\alpha^\beta \in \mathcal{D}^{(-j+\alpha)}$$

because $\nu(x_\alpha^\beta) = \alpha$. Then, if $\alpha > i$ we have that this function has positive weight. Thus, when evaluated at $x = 0$ it is zero.

In other words we proved that, for every $i = 1, \ldots, m$, up to the $i^{th}$-term we can find only element in $\mathcal{D}^i_q$.

To prove the converse inclusion we have to show that, given some elements $\xi_i^j \in \mathcal{D}^i_q$, we can find a smooth variation that has these vectors as elements of its jet. We start with some preliminary lemmas.

Lemma 10.28. Let $m, n$ be two integers. Assume that we have two flows such that

$$P_t = \text{Id} + Vt^n + O(t^{n+1})$$
$$Q_t = \text{Id} + Wt^m + O(t^{m+1})$$

in the operator sense. Then $P_tQ_tP_t^{-1}Q_t^{-1} = \text{Id} + [V,W]t^{n+m} + O(t^{n+m+1})$. 

206
Exercise 10.29. Assume that the flow $P_t$ satisfies $P_t = \text{Id} + V t^n + O(t^{n+1})$. Show that the nonautonomous vector field $V_t$ associated to $P_t$ satisfies $V_t = nt^{n-1}V + O(t^n)$.

Proof. Define $R(t,s) := P_tQ_sP_t^{-1}Q_s^{-1}$. Since $P_0 = Q_0 = \text{Id}$ we have that $R$ satisfies $R(0,s) = R(t,0) = \text{Id}$ for every $t,s \in \mathbb{R}$. Hence the only derivative that enter in our expansion, that coincide with $F(t) = R(t,t)$, are mixed derivatives. This remark let us to expand the product $P_tQ_sP_t^{-1}Q_s^{-1}$ and keep only terms with mixed power of $t$ and $s$ in the expansion. Using that for the inverse flow we have the expansions

$$P_t^{-1} = \text{Id} - t^nV + O(t^{n+1}), \quad Q_t^{-1} = \text{Id} - t^mW + O(t^{m+1}).$$

one gets

$$(\text{Id} + t^nV + O(t^{n+1}))(\text{Id} + s^mW + O(s^{m+1}))(\text{Id} - t^nV + O(t^{n+1}))(\text{Id} - s^mW + O(s^{m+1})) =$$

$$= \text{Id} + t^n s^m(VW - WV) + O(t^{n+m+1})$$

and the lemma is proved. \hfill \square

Lemma 10.30. For all $l \geq h$ and $\forall i_1, \ldots, i_h \in \{1, \ldots, k\}$, there exists an admissible variation $u(t,s)$ such that

$$q \circ \exp^t \int_0^s f_{u(t,s)} ds = q + t^l[f_{i_1}, \ldots, [f_{i_{h-1}}, f_{i_h}]](q) + O(t^{l+1}). \quad(10.11)$$

Proof. We prove the lemma by induction on $h$

- $\forall l \geq 1$ and $\forall i = 1, \ldots, k$ we have to show that there exists an admissible variation $u(t,s)$ such that

$$q \circ \exp^t \int_0^s f_{u(t,s)} ds = q + t^l f_i(q) + O(t^{l+1})$$

To this aim, it is sufficient to consider a control $u = (u_1, \ldots, u_k)$ where $u_i = t^l$ and $u_j = 0$ for all $j \neq i$.

- $\forall l \geq 2$ and $\forall i,j = 1, \ldots, k$, we have to show that there exists an admissible variation $u(t,s)$ such that

$$q \circ \exp^t \int_0^s f_{u(t,s)} ds = q + t^l[f_{i}, f_j](q) + O(t^{l+1})$$

To this aim, it is sufficient to use the previous lemma where $P_t$ and $Q_t$ are flows respectively of nonautonomous vector fields $V_t = t^{l-1}f_{i_1}$ and $W_t = tf_{i_2}$.

With analogous arguments we can prove by induction the lemma.

In other words we proved that every bracket monomial of degree $i$ can be presented as the $i$-th term of a jet of some admissible variation. Now we prove that we can do the same for any linear combination of such monomials (recall that $\mathcal{D}^i$ is the linear span of all $i$-th order brackets). \hfill \square

Remark 10.31. The previous construction of $u(t,s)$ does not depend on the sub-Riemannian structure but only on the structure of the Lie bracket.
Lemma 10.32. Let $\pi = \pi(f_1, \ldots, f_k)$ a bracket polynomial of degree $\deg \pi \leq l$. There exists an admissible variation $u(t, s)$ such that
\[
q \circ \text{exp} \int_0^t f_u(t, s)ds = q + t^l \pi(f_1, \ldots, f_k)(q) + O(t^{l+1})
\]

Proof. Let $\pi(f_1, \ldots, f_k) = \sum_{j=1}^N V_j(f_1, \ldots, f_k)$ where $V_j$ are monomials. By our previous argument we can find $u^j(t, s), s \in [0, \tau_j]$ such that
\[
q \circ \text{exp} \int_0^t f_{u^j(t, s)}ds = q + t^j V_j(f_1, \ldots, f_k)(q) + O(t^{l+1})
\]

Now consider the concatenation of controls $u(t, s)$, where $s \in [0, \tau]$ and $\tau = \sum_{j=1}^N \tau_j$ defined as follows
\[
u(t, s) = u^j \left( t, s - \sum_{i=1}^j \tau_i \right), \quad \text{if} \quad \sum_{i=1}^j \tau_i \leq s < \sum_{i=1}^{j+1} \tau_i, \quad 1 \leq j \leq N.
\]

Exercise 10.33. Complete the previous proof by showing that the flow associated with $u$ has as main term in the Taylor expansion $\sum_j V_j$ at order $l$. Then prove, by using a time rescaling argument, that also any monomial of type $\alpha V$ for $\alpha \in \mathbb{R}$ can be presented in this way.

Second part of Theorem [10.27] Now we can complete the proof of the first statement of Theorem [10.27] proving the following inclusion $\{ \sum_{i=1}^m \xi_i, \xi_i \in D^i_q \} \subset J_q^m$.

Let us consider a $m$-th jet $j = \sum_{i=1}^m \xi_i, \xi_i \in D^i_q$. We prove the statement by steps: at $i$-th step we built an admissible variation whose $i$-th Taylor polynomial coincide with the one of $j$.

- From Lemma ?? there exists a smooth admissible variation $\gamma(t)$ such that
\[
\gamma(t) = q \circ \text{exp} \int_0^t f_u(t, s)ds, \quad \dot{\gamma}(t) = \xi_1
\]
Then we will have $\gamma(t) = t\xi_1 + t^2 \eta_2 + O(t^3)$ where $\eta_2 \in D^2$ from first part of the proof. In the second step we correct the second order term.

- From Lemma ?? there exists a smooth admissible variation $\gamma_1(t)$ such that
\[
\gamma_1(t) = q \circ \text{exp} \int_0^t f_u(t, s)ds, \quad \gamma_1(t) = t^2(\xi_2 - \eta_2) + O(t^3)
\]
Defining $\gamma_2(t) := \gamma_1(t) \circ \gamma(t)$ we have
\[
\gamma_2(t) \simeq t\xi_1 + t^2 \eta_2 + t^2(\xi_2 - \eta_2) + t^3 \eta_3
\]
\[
\simeq t\xi_1 + t^2 \xi_2 + t^3 \eta_3
\]
where $\eta_3 \in D^3$.

At every step we can correct the right term of the jet and after $m$ steps we have the inclusion.
(ii) We have to prove that

\[ j \sim j' \iff j - j' = \sum_{i=1}^{m} t^i \eta_i, \quad \eta_i \in D_q^{i-1}. \]

(\Rightarrow). Assume that \( j \sim j' \), where \( j = J_q^m \gamma = \sum t^i \xi_i \) and \( j' = J_q^m \gamma' = \sum t^i \xi'_i \). Then \( \gamma' = \gamma \circ Q_t \) for some slow flow \( Q_t \in \mathcal{P}_0^f \) of the form

\[ Q_t = Q_t^1 \circ \cdots \circ Q_t^h, \quad Q_t^i = P_t^i \circ \exp \int_0^\tau f_{tv(t,s)} ds \circ (P_t^i)^{-1}. \]

for some \( P^i \in \mathcal{P}_i, i = 1, \ldots, h \). For simplicity we prove only the case \( h = 1 \). By formula (6.20) we have that

\[ Q_t = P_t \circ \exp \int_0^\tau f_{tv(t,s)} ds \circ P_t^{-1} = \exp \int_0^\tau P_t \circ f_{tv(t,s)} \circ P_t^{-1} ds \]

then by linearity of \( f \) we have

\[ Q_t = \exp \int_0^\tau t \text{Ad} P_t f_{tv(t,s)} ds \]

Now recall that \( P_t = \exp \int_0^\tau f_{w(t,\theta)} d\theta \) for some admissible variation \( w(t, \theta) \) and from (6.18) we get

\[ Q_t = \exp \int_0^\tau t \exp \int_0^s \text{Ad} f_{w(t,\theta)} d\theta f_{v(t,s)} ds \]

Finally, if \( \gamma(t) = q \circ \exp \int_0^\tau f_{u(t,s)} ds \) we can write

\[ \gamma'(t) = q \circ \exp \int_0^\tau f_{v(t,s)} ds \circ \exp \int_0^\tau t \exp \int_0^s \text{Ad} f_{w(t,\theta)} d\theta f_{v(t,s)} ds \]

Expanding with respect to \( t \) we have \( Q_t \simeq (I_d + t \sum t^i V_i) = I_d + \sum t^{i+1} V_i \) where \( V_i \) is a bracket polynomial of degree \( \leq i \). Due to the presence of \( t \) it is easy to see that in the expansion of \( \gamma' \) we will find the same terms of \( \gamma \) plus something that belong to \( D^{i-1} \).

(\Leftarrow). Assume now that \( j = J_q^m \gamma = \sum t^i \xi_i \) and \( j' = J_q^m \gamma' = \sum t^i \xi'_i \), with

\[ j - j' = \sum_{i=1}^{m} t^i \eta_i, \quad \eta_i \in D_q^{i-1}. \]

We need to find a slow flow \( Q_t \) such that \( \gamma' = \gamma \circ Q_t \). In other words it is sufficient to prove that we can realize with a slow flow every jet of type \( \sum_{i=1}^{m} t^i \eta_i, \quad \eta_i \in D_q^{i-1} \). To this purpose we can repeat arguments of proof of part (i), using the following

**Lemma 10.34.** Let \( P_t, Q_t \) be two flows with \( P_t \in \mathcal{P}_0^f \) and \( Q_t \in \mathcal{P}_0^f \) (or \( P_t \in \mathcal{P}_0^f \) and \( Q_t \in \mathcal{P}_0^f \)). Then \( P_t Q_t P_t^{-1} Q_t^{-1} \in \mathcal{P}_0^f \).

**Proof.** If \( Q_t \in \mathcal{P}_0^f \) then \( Q_t^{-1} \in \mathcal{P}_0^f \). Moreover from the definition of \( \mathcal{P}_0^f \) we have that \( P_t Q_t P_t^{-1} \in \mathcal{P}_0^f \). Hence also their composition is in \( \mathcal{P}_0^f \). \( \square \)

\[ \square \]
We have the following corollary of Theorem 10.27, part (i).

**Corollary 10.35.** In privileged coordinates \((x_1, \ldots, x_m)\) defined by the splitting \(\mathbb{R}^n = \mathbb{R}^{k_1} \oplus \cdots \oplus \mathbb{R}^{k_m}\) we have

\[
J_q^f = \left\{ \begin{pmatrix} tx_1 + O(t^2) \\ t^2x_2 + O(t^3) \\ \vdots \\ t^mx_m \end{pmatrix} : x_i \in \mathbb{R}^{k_i}, i = 1, \ldots, m \right\}
\]

**Proof.** Indeed we know that \(D^i = \mathbb{R}^{k_1} \oplus \cdots \oplus \mathbb{R}^{k_i}\) and writing

\[\xi_i = x_{i,1} + \cdots + x_{i,m}, \quad x_{i,j} \in \mathbb{R}^{k_j}\]

we have, expanding and collecting terms

\[
\sum t^i \xi_i = t\xi_1 + t^2\xi_2 + \cdots + t^m\xi_m
\]

\[
= tx_{1,1} + t^2(x_{2,1} + x_{2,2}) + \cdots + t^m(x_{m,1} + \cdots + x_{m,m})
\]

\[
= (tx_{1,1} + t^2x_{2,1} + \cdots + t^m x_{m,1}, t^2x_{2,2} + \cdots + t^m x_{m,2}, t^m x_{m,m})
\]

\[\square\]

**Corollary 10.36.** The nonholonomic tangent space \(T_q^f\) is a smooth manifold of dimension \(\dim T_q^f = \sum_{i=1}^{m(q)} k_i(q)\). In privileged coordinates we can write

\[
T_q^f = \left\{ \begin{pmatrix} tx_1 \\ t^2x_2 \\ \vdots \\ t^mx_m \end{pmatrix} : x_i \in \mathbb{R}^{k_i}, i = 1, \ldots, m \right\}
\]

and dilations \(\delta_\alpha\) acts on \(T_q^f\) in a quasi-homogeneous way as follows

\[
\delta_\alpha(tx_1, \ldots, t^mx_m) = (\alpha tx_1, \ldots, \alpha^m t^m x_m), \quad \alpha > 0.
\]

**Proof.** It follows directly from the representation of the equivalence relation. Indeed two elements \(j\) and \(j'\) can be written in coordinates as

\[
j = (tx_1 + O(t^2), t^2x_2 + O(t^3), \ldots, t^mx_m)
\]

\[
j' = (ty_1 + O(t^2), t^2y_2 + O(t^3), \ldots, t^my_m)
\]

and \(j \sim j'\) if and only if \(x_i = y_i\) for all \(i = 1, \ldots, m\). \[\square\]

**Remark 10.37.** Notice that a polynomial differential operator homogeneous with respect to \(\nu\) (i.e. whose monomials are all of same weight) is homogeneous with respect to dilations \(\delta_t : \mathbb{R}^n \rightarrow \mathbb{R}^n\) defined by

\[
\delta_t(x_1, \ldots, x_m) = (tx_1, t^2x_2, \ldots, t^mx_m), \quad t > 0.
\]

(10.12)

In particular for a homogeneous vector field \(X\) of weight \(h\) it holds \(\delta_*X = t^{-h}X\).
Now we can improve Proposition 10.21 and see that actually the jet of a horizontal vector field is a vector field on the tangent space and belongs to $D^{(-1)}$ (in privileged coordinates).

**Lemma 10.38.** Fix a set of privileged coordinate. Let $V \in D^{(-1)}$, then the jet $J^m_q V$ is tangent to the submanifold $J^I_q$. Moreover it is well defined as vector field $\hat{V}$ on the nonholonomic tangent space. In other words $\hat{V} \in \text{Vec}(T^I_q)$ and we have

$$V = \begin{pmatrix} v_1(x) \\ v_2(x) \\ \vdots \\ v_m(x) \end{pmatrix} \implies \hat{V} = \begin{pmatrix} \hat{v}_1(x) \\ \hat{v}_2(x) \\ \vdots \\ \hat{v}_m(x) \end{pmatrix}$$  \hspace{1cm} (10.13)

where $\hat{v}_i$ is the term of order $i - 1$ of $v_i$.

**Proof.** Let $V \in D^{(-1)}$ and $\gamma(t)$ be an admissible variation. When expressed in coordinates we have

$$V = \begin{pmatrix} v_1(x) \\ v_2(x) \\ \vdots \\ v_m(x) \end{pmatrix}, \quad \gamma(t) = \begin{pmatrix} tx_1 + O(t^2) \\ t^2 x_2 + O(t^3) \\ \vdots \\ t^m x_m \end{pmatrix}$$

We know that $(J^m_q V)(j^m_q \gamma)$ is expressed as the $m$-th jet of $tV(\gamma(t))$ by Exercise 10.9. Hence we compute

$$(J^m_q V)(j^m_q \gamma) = \begin{pmatrix} tv_1(tx_1 + O(t^2), \ldots, t^m x_m) \\ tv_2(tx_1 + O(t^2), \ldots, t^m x_m) \\ \vdots \\ tv_m(tx_1 + O(t^2), \ldots, t^m x_m) \end{pmatrix}$$  \hspace{1cm} (10.14)

Notice that $V \in D^{(-1)}$ means exactly that

$$V = \sum_{i=1}^m v_i(x) \frac{\partial}{\partial x_i} = \sum v_i^j(x) \frac{\partial}{\partial x_i^j}, \quad \nu \left( \frac{\partial}{\partial x_i} \right) = -i$$

and $v_i$ is a function of order at least $i - 1$. Let we denote with $\hat{v}_i$ the homogeneous part of $v_i$ of order $i - 1$. To compute the value of $\hat{V}$ then we have to restrict its action on admissible variations from $T^I_q$, then evaluate and neglect the higher order part (that corresponds to the projection on the factor space) in order to have

$$v_i(tx_1, \ldots, t^m x_m) = t^{i-1} \hat{v}_i(x_1, \ldots, x_m) + O(t^i)$$

and using equality we have

$$(J^m_q V)_{T^I_q} = \begin{pmatrix} tv_1(tx_1, \ldots, t^m x_m) \\ tv_2(tx_1, \ldots, t^m x_m) \\ \vdots \\ tv_m(tx_1, \ldots, t^m x_m) \end{pmatrix} = \begin{pmatrix} t\hat{v}_1 + O(t^2) \\ t^2 \hat{v}_2 + O(t^3) \\ \vdots \\ t^m \hat{v}_m + O(t^{m+1}) \end{pmatrix}$$  \hspace{1cm} (10.15)

Then (10.13) follows.
Remark 10.39. Notice that, since $\hat{v}_i$ is a homogeneous function of weight $i - 1$, it depends only on variables $x_1, \ldots, x_{i-1}$ of weight equal of smaller than its weight. Hence $\hat{V}$ has the following triangular form
\[
\hat{V}(x) = \begin{pmatrix}
\hat{v}_1 \\
\hat{v}_2(x_1) \\
\vdots \\
\hat{v}_m(x_1, \ldots, x_{m-1})
\end{pmatrix}
\] (10.16)

Moreover the flow of a vector field of this kind can be easily computed by a step by step substitution.

10.3.1 Existence of privileged coordinates

Now we prove existence of privileged coordinates

Proof of Theorem 10.26. Consider our sub-Riemannian structure on $M$ defined by the orthonormal frame $\{f_1, \ldots, f_k\}$ and its flag $D_q^1 \subset D_q^2 \subset \ldots \subset D_q^n = T_q M$, with
\[
n_j := \dim D_q^j \quad (n_j = k_1 + \ldots + k_j)
\]

Let we consider a basis $\{V_1, \ldots, V_n\}$ of the tangent space adapted to the flag, i.e.
\[
\pi_i = \pi_i(f_1, \ldots, f_k)
\]
$\pi_i$ bracket polynomial, $\deg \pi_i \leq j$ if $i \leq n_j$
\[
D_q^j = \text{span}\{V_1(q), \ldots, V_{n_j}(q)\}, \quad j = 1, \ldots, m
\]

In particular $V_1, \ldots, V_{n_1}$ are selected in $\{f_i, i = 1, \ldots, k\}$, $V_{n_1+1}, \ldots, V_{n_2}$ are selected from $\{[f_i, f_j], i, j = 1, \ldots, k\}$ and so on.

Define the map
\[
\Psi : (s_1, \ldots, s_n) \mapsto q \circ e^{s_1 V_1} \circ \ldots \circ e^{s_n V_n}
\] (10.17)

We want to show that $\Psi^{-1}$ defines privileged coordinates around $q$. It is easy to show that (10.17) is a local diffeomorphism since
\[
\frac{\partial \Psi}{\partial s_i} \bigg|_{s=0} = \Psi^{-1} \frac{\partial}{\partial s_i} \bigg|_{s=0} = V_i(q), \quad i = 1, \ldots, n
\] (10.18)

Hence it remains to show that
\[
(i) \quad \Psi^{-1}_{e^{-s}}(D_q^j) = \text{span}\{\frac{\partial}{\partial s_{1}}, \ldots, \frac{\partial}{\partial s_{n_j}}\},
\]
\[
(ii) \quad \Psi^{-1}_{e^{-s}} f_i \in D^{(-1)} \text{ for every } i = 1, \ldots, k
\]

Part (i) easily follows from our choice of adapted frame to the flag and (10.18). On the other hand the second part is not trivial since we need to compute differential of $\Psi$ at every point and not only at $s = 0$. 212
and our claim is equivalent to show that
\[ \Psi^{-1} \in \text{D}^{(w_1-d)} \].

Remark 10.40. In what follows we consider on \( T_q M \) the weight defined by coordinates \((y_1, \ldots, y_n)\) induced by the flag. In other words, we consider the basis \( V_1(q), \ldots, V_n(q) \) in \( T_q M \) and write
\[ v = (y_1, \ldots, y_n) = \sum y_i V_i(q), \quad \text{where} \quad \nu(y_i) := w_i = j \quad \text{if} \quad n_{j-1} < i \leq n_j \]
Moreover, we can think at \( v \in T_q M \) as the constant vector field on \( T_q M \) identically equal to \( v \). In this way, it makes sense to consider the value of a polynomial bracket at \( \pi(f_1, \ldots, f_k) \) at the point \( q \) and consider its weight \( \nu(\pi) \).

We prove the following auxiliary

Lemma 10.41. Let \( X = \pi(f_1, \ldots, f_k)(q) \in \text{Vec}(T_q M), \ \nu(X) \leq d. \) Consider now the polynomial vector field on \( T_q M \)
\[
Y(y) = \sum y_{i_1} \cdots y_{i_k} (\text{ad } V_{i_1} \circ \cdots \circ \text{ad } V_{i_k} X)(q)
= \sum p_i(y) V_i(q)
\]
for some polynomial \( p_i \). Then \( p_i \in \text{D}^{(w_i-d)} \).

Proof of Lemma. It easily follows from definition of weights that
\[ \text{ad } V_{i_1} \circ \cdots \circ \text{ad } V_{i_k}(X) \in \text{D}^{(-\sum w_{j_i} - d)} \]
hence every summand of (10.19) belong to \( \text{D}^{(-d)} \). Then if we rewrite the sum in terms of the basis \( V_i(q), i = 1, \ldots, k \) we have that every coefficient \( p_i(y) \) must belong to \( \text{D}^{(w_i-d)} \), since \( \nu(V_i(q)) = w_i \).

Now we prove the following claim: for every bracket polynomial \( X = \pi(f_1, \ldots, f_k) \) we have \( \Psi^{-1} X \in \text{D}^{(-d)} \). In particular part (ii) will follow when \( d = 1 \). Clearly, we can write in coordinates
\[
\Psi^{-1} X = \sum_{i=1}^{n} a_i(s) \frac{\partial}{\partial s_i}
\]
and our claim is equivalent to show that \( a_i \in \text{D}^{(w_i-d)} \). First we notice that
\[
\Psi \frac{\partial}{\partial s_i} = \frac{\partial}{\partial \epsilon} \bigg|_{\epsilon=0} q \circ e^{s_1 V_1} \circ \cdots \circ e^{(s_i+\epsilon)V_i} \circ \cdots \circ e^{s_n V_n}
= q \circ e^{s_1 V_1} \circ \cdots \circ e^{s_i V_i} \circ V_i \circ e^{s_{i+1} V_{i+1}} \circ \cdots \circ e^{s_n V_n}
= q \circ e^{s_1 V_1} \circ \cdots \circ e^{s_n V_n} \circ e^{-s_n V_n} \circ \cdots \circ e^{-s_{i+1} V_{i+1}} \circ V_i \circ e^{s_{i+1} V_{i+1}} \circ \cdots \circ e^{s_n V_n}
\]
In geometric notation we can write
\[
\Psi \frac{\partial}{\partial s_i} = e_{s_i}^{s_n V_n} \cdots e_{s_i+1}^{s_{i+1} V_{i+1}} V_i \bigg|_{\Psi(s)}
\]
Remember that, as operator on functions, \( e^Y_\ast = e^{-\ast \text{ad } Y} \). This implies that in (10.21) we have a series of bracket polynomials. Apply \( \Psi \) to (10.20) we get
\[
X \bigg|_{\Psi(s)} = \sum_{i=1}^{n} a_i(s) e_{s_i}^{s_n V_n} \cdots e_{s_i+1}^{s_{i+1} V_{i+1}} V_i \bigg|_{\Psi(s)}
\]
Now we apply $e_{s_1}^{-s_1}V_1 \cdot \cdot \cdot e_{s_n}^{-s_n}V_n$ to both sides to compute the vector field at the point $q$

$$e_{s_1}^{-s_1}V_1 \cdot \cdot \cdot e_{s_n}^{-s_n}V_n X \bigg|_q = \sum_{i=1}^{n} a_i(s)e_{s_1}^{-s_1}V_1 \cdot \cdot \cdot e_{s_{i-1}}^{-s_{i-1}}V_{i-1}V_i$$  \hspace{1cm} (10.22)

Rewriting this identity in coordinates

$$\sum_i b_i(s)V_i(q) = \sum_{i,j} a_i(s)(\varphi_{ij}(s)V_j(q) + V_i(q))$$  \hspace{1cm} (10.23)

where $\varphi_{ij}(0) = 0$. Indeed we split the zero order term since we know that for $s = 0$ the pushforward of the vector fields is exactly $V_i$. Using Lemma above with $X$ and $V_i, i = 1, \ldots, n$ we have

$$b_i \in \mathcal{D}^{w_i-d}, \quad \varphi_{ij} \in \mathcal{D}^{w_j-w_i}$$

On the other hand we can rewrite relation between coefficients as follows

$$B(s) = A(s)(\Phi(s) + I)$$

where we denote $B(s) = (b_1(s), \ldots, b_n(s)), A(s) = (a_1(s), \ldots, a_n(s))$ and $\Phi(s) = (\varphi_{ij})_{ij}$ Thus we get

$$A(s) = B(s)(I + \Phi(s))^{-1} = B(s)(I - \Phi(s) + \Phi(s)^2 - \ldots) = B(s) - (B\Phi)(s) + (B\Phi^2)(s) - \ldots$$

and we can finish the proof noticing that

$$(B)_i = b_i \in \mathcal{D}^{w_i-d}$$

$$(B\Phi)_i = \sum b_j \varphi_{ji} \in \mathcal{D}^{w_j-d+(w_i-w_j)} = \mathcal{D}^{w_i-d}$$

and so on. Hence we get $a_i \in \mathcal{D}^{w_i-d}$.

**Remark 10.42.** One can repeat all calculation in chronological notation and recover the proof in a purely algebraic way. In the above computations nothing change if we consider any permutation $\sigma = (i_1, \ldots, i_n)$ of $(1, \ldots, n)$ and the coordinate map

$$\Psi_\sigma : (s_1, \ldots, s_n) \mapsto q \circ e_{s_n}^{s_n}V_n \circ \ldots \circ e_{s_1}^{s_1}V_1$$

In particular we can consider the coordinate map

$$\Phi : (x_1, \ldots, x_n) \mapsto q \circ e_{x_n}^{x_n}V_n \circ \ldots \circ e_{x_1}^{x_1}V_1$$

and it is easy to see that it satisfies

$$\Phi^{-1}_*V_1 = \partial_{x_1}$$

$$\Phi^{-1}_*V_2 \bigg|_{x_1=0} = \partial_{x_2}$$

$$\vdots$$

$$(10.24)$$

for $i = 1, \ldots, n_1$, the set of vector fields among $f_1, \ldots, f_k$ that generates $\mathcal{D}_q$. 1214
In Riemannian geometry the tangent space depends only on the dimension of the manifold (i.e. all tangent spaces to a \( n \)-dimensional manifold are isometric). Now we can prove that in sub-Riemannian geometry this is not true. Indeed we see that, even in dimension 3, we can have non isometric tangent space, depending on the growth vector \((n_1, \ldots, n_m)\).

In bigger dimension it is also possible to prove that, for a fixed growth vector, we have non isometric tangent space depending on the point on the manifold.

**Example 10.43.** (Heisenberg)
Assume \( n = 3 \) and that growth vector is \((2, 3)\). Then we consider coordinates \((x_1, x_2, x_3)\) and weights \((w_1, w_2, w_3) = (1, 1, 2)\). We can assume that

\[
V_1 = f_1, \quad V_2 = f_2, \quad V_3 = [f_1, f_2]
\]

From last Remark we have that, in privileged coordinates we can assume

\[
f_1 = \partial_{x_1}, \quad f_2 = \partial_{x_2} + \alpha x_1 \partial_{x_3}, \quad \alpha \in \mathbb{R}
\]

(10.25)

because \( f_i = \partial_{x_i} + \) something that has weight \(-1\) and depend only on \( \partial_{x_j}, j > n_1 \). On the other hand from (10.24) we have

\[
[f_1, f_2] = \partial_{x_3} \implies \quad \alpha = 1
\]

and we get the Heisenberg algebra

\[
f_1 = \partial_{x_1}, \quad f_2 = \partial_{x_2} + x_1 \partial_{x_3}, \quad f_3 = \partial_{x_3}
\]

(10.26)

**Example 10.44.** (Martinet)
Assume \( n = 3 \) and that growth vector is \((2, 2, 3)\). Then we consider coordinates \((x_1, x_2, x_3)\) and weights \((w_1, w_2, w_3) = (1, 1, 3)\). We can assume, up to change indices, that

\[
V_1 = f_1, \quad V_2 = f_2, \quad V_3 = [f_1, [f_1, f_2]]
\]

From last Remark we have that, in privileged coordinates we can write

\[
f_1 = \partial_{x_1}, \quad f_2 = \partial_{x_2} + (\alpha x_1^2 + \beta x_1 x_2) \partial_{x_3}, \quad \alpha, \beta \in \mathbb{R}
\]

(10.27)

since we assume \( f_2 |_{x_1 = 0} = \partial_{x_2} \) that implies \( f_2 = \partial_{x_2} + x_1 a(x) \partial_{x_3}, \) but \( \nu(f_2) = -1 \) and so (10.27) follows.

From \( V_3 |_{x = 0} = \partial_{x_3} \) we have

\[
[f_1, [f_1, f_2]] = 2\alpha \partial_{x_3} \implies \alpha = 1/2.
\]

Moreover, since we are interested to normalize sub-Riemannian structure and not only the pair of vector fields, we consider rotations of the orthonormal frame.

**Remark 10.45.** Notice that

\[
\bar{f}_1 = \cos \theta f_1 - \sin \theta f_2 \quad \bar{f}_2 = \sin \theta f_1 + \cos \theta f_2
\]

\[
\implies \quad [\bar{f}_1, \bar{f}_2] = [f_1, f_2].
\]
Thus, denoting as usual
\[ f_u = u_1 f_1 + u_2 f_2 \]
we can consider the linear map
\[ \varphi : u \mapsto [f_u, [f_1, f_2]]/D \]
which vanish on some line on the plane \( D = \text{span}\{f_1, f_2\} \). Up to a rotation of the frame we can assume that \( f_2 \in \ker \varphi \) so that \( [f_2, [f_1, f_2]] = 0 \), hence \( \beta = 0 \).

\[ f_1 = \partial_{x_1}, \quad f_2 = \partial_{x_2} + \frac{1}{2} x_1^2 \partial_{x_3}, \quad f_3 = \partial_{x_3} \quad (10.28) \]

10.4 Geometric meaning

In the previous section we very clearly found how \( \hat{V} \) is analytically recovered from \( V \). It is nothing else but the principal part of \( V \) in privileged coordinates. But now we want to discuss in which sense \( \hat{V} \) is an approximation of \( V \). It turns out that in this nonholonomic setting it plays the same role that linearization of a vector field does in the Euclidean case.

**Lemma 10.46.** Let \( V \) a vector field. In privileged coordinates we have equality
\[ \varepsilon \delta_{\frac{1}{\varepsilon}*} V = \hat{V} + \varepsilon W_\varepsilon, \quad \text{where } W_\varepsilon \text{ is smooth} \]

*Proof.* Write \( V = \hat{V} + W \) and applying the dilation we find
\[ \delta_{\frac{1}{\varepsilon}*} V = \delta_{\frac{1}{\varepsilon}*} \hat{V} + \delta_{\frac{1}{\varepsilon}*} W \]
Since \( \hat{V} \) is homogeneous of degree \(-1\) we have \( \delta_{\frac{1}{\varepsilon}*} \hat{V} = \frac{1}{\varepsilon} \hat{V} \) and setting \( W_\varepsilon = \varepsilon \delta_{\frac{1}{\varepsilon}*} W \) we are done. \( \square \)

**Remark 10.47.** Geometrically this procedure means that we consider a small neighborhood of the point \( q \) and we make a dilation. Then we properly rescale in order to catch the principal term. This is a blow-up procedure. Notice that we are blowing-up in a nonisotropic way and it contains information about local structure of the bracket.

Now we can give a very precise meaning of the fact that nilpotent approximation is the principal part of the sub-Riemannian structure, which knows local geometry near the point \( q \). Let us consider the *end point map*
\[ E : \mathcal{U} \to M, \quad u(\cdot) \mapsto q \circ \exp_{\hat{p}} \int_0^1 f_{u(t)} dt \]
where \( \mathcal{U} = L^2_k(0,1) = L^2([0,1], \mathbb{R}^k) \) is the set of admissible controls. Let we denote by \( \rho \) the sub-Riemannian distance from the fixed point
\[ \rho(x) := d(x, q) = \inf\{\|u\|, E(u) = x\} \quad (10.29) \]
From Lemma 10.46 we can write for \( \varepsilon > 0 \)
\[ f_u^\varepsilon := \varepsilon \delta_{\frac{1}{\varepsilon}*} f_u = \tilde{f}_u + \varepsilon W_u^\varepsilon \]
Denote now with $f^\varepsilon$ and $\hat{f}$ respectively the sub-Riemannian structures on $\mathbb{R}^n$ and by $d^\varepsilon$ and $\hat{d}$ the associated sub-Riemannian distance. Notice that, from the very definition of $d^\varepsilon$ we have

$$d^\varepsilon(x, y) = \frac{1}{\varepsilon}d(\delta_\varepsilon(x), \delta_\varepsilon(y))$$

that says $d^\varepsilon$ is $d$ when we look infinitesimally near the point $q$ and rescale.

Let $\rho^\varepsilon, \hat{\rho}$ and $E^\varepsilon, \hat{E}$ have analogous meaning. We start from an auxiliary proposition.

**Proposition 10.48.** $E^\varepsilon \to \hat{E}$ uniformly on balls in $L^2(0, 1)$ (actually in $C^\infty$ sense).

**Proof.** Consider the solution $x^\varepsilon(t)$ and $\hat{x}(t)$ of the two systems based at $q = 0$

$$\dot{\hat{x}}(t) = \hat{f}u(t)(\hat{x}(t)), \quad \dot{x}^\varepsilon(t) = f^\varepsilon u(t)(x^\varepsilon(t))$$

Using Lemma 10.46 we rewrite the second equation as

$$\dot{x}^\varepsilon(t) = \hat{f}u(t)(x^\varepsilon(t)) + \varepsilon W^\varepsilon_t(x^\varepsilon(t))$$

and standard estimates from ODE theory prove that $x^\varepsilon \to \hat{x}$.

Notice that, since nilpotent vector fields are complete, the solution $\hat{x}(t)$ is defined for all $t \in \mathbb{R}$.

**Lemma 10.49.** \{${}\rho^\varepsilon\}_\varepsilon>0$ is an equicontinuous family.

**Proof.** We will prove the following: for every compact $K \subset \mathbb{R}^n$ there exists $\varepsilon_0, C > 0$, depending on $K$, such that

$$d^\varepsilon(x, y) \leq C|x - y|^{1/m}, \quad \forall \varepsilon < \varepsilon_0, \forall x, y \in K. \quad (10.30)$$

where $m$ is the degree of nonholonomy. Notice that from (10.30) we get, using triangle inequality

$$|\rho^\varepsilon(x) - \rho^\varepsilon(y)| = |d^\varepsilon(0, x) - d^\varepsilon(0, y)| \leq d^\varepsilon(x, y) \leq C|x - y|^{1/m}$$

which proves the lemma. We are then reduced to prove (10.30). Idea is to cover a fixed neighborhood of the origin using controls with bounded norms, uniformly in $\varepsilon$.

Let $\hat{V}_1, \ldots, \hat{V}_n$ an adapted basis of the nilpotent system $\hat{f}$, such that $\hat{V}_i = \pi_i(\hat{f}_1, \ldots, \hat{f}_k)$ for some bracket polynomials $\pi_i, i = 1, \ldots, n$. From the very definition we have

$$\hat{V}_1(0) \land \ldots \land \hat{V}_n(0) \neq 0$$

On the other hand, by continuity, this implies that they are linearly independent also in a small neighborhood of the origin and by quasi-homogeneity we get

$$\hat{V}_1(x) \land \ldots \land \hat{V}_n(x) \neq 0, \quad \forall x \in \mathbb{R}^n.$$

Let $V_i^\varepsilon = \pi_i(f_1^\varepsilon, \ldots, f_k^\varepsilon)$ denote vector fields defined by the same bracket polynomials but in terms of the vector fields of the approximating system. For every $K \subset \mathbb{R}^n$ there exists $\varepsilon_0 = \varepsilon_0(K)$ such that

$$V_1^\varepsilon(x) \land \ldots \land V_n^\varepsilon(x) \neq 0, \quad \forall x \in K, \forall \varepsilon \leq \varepsilon_0. \quad 217$$
Recall that by Lemma 10.32, given a bracket polynomial \( \pi_i(g_1, \ldots, g_k) \), \( \deg \pi = w_i \) there exists an admissible variation \( u_i(t, s) \), depending only on \( \pi_i \), such that
\[
\exp \int_0^1 g_{u_i(t, s)} ds = \text{Id} + t^{w_i} \pi_i(g_1, \ldots, g_k) + O(t^{w_i+1})
\]
If we apply this lemma for \( g_i = f^\varepsilon_i \) we find \( u_i(t, s) \) such that
\[
\exp \int_0^1 f^\varepsilon_{u_i(t, s)} ds = \text{Id} + t^{w_i} V^\varepsilon_i + O(t^{w_i+1}), \quad \forall \varepsilon > 0
\]
where \( w_i = \deg \hat{V}_i = \deg V^\varepsilon_i \). Now consider the map
\[
\Phi^\varepsilon(t_1, \ldots, t_n, x) = x \circ \exp \int_0^1 f^\varepsilon_{u_{t_1}(t_1, s)} ds \circ \cdots \circ \exp \int_0^1 f^\varepsilon_{u_n(t_n, s)} ds
\] (10.31)

Remark 10.50. We have the expansion
\[
x \circ \exp \int_0^1 f^\varepsilon_{u_{t_1}(t_1, s)} ds = x + t_i V^\varepsilon_i(x) + O(t^{w_i+1}_i)
\]
In particular this is a \( C^1 \) map with respect to \( t \). Notice that it is not \( C^2 \) if \( w_i > 1 \) for some \( i \) (i.e. a “real” subriemannian problem).

From this remark it follows that \( \Phi^\varepsilon \in C^1 \) as a function of \( t \), being a composition of \( C^1 \) maps. Moreover we get the expansion
\[
\Phi^\varepsilon(t_1, \ldots, t_n, x) = x + \sum_{i=1}^n t_i V^\varepsilon_i(x) + O(|t|) \quad \Rightarrow \quad \frac{\partial \Phi^\varepsilon}{\partial t_i} \bigg|_{t=0} = V^\varepsilon_i(x)
\]
Hence the map \( \Phi^\varepsilon \) is a local diffeomorphism near the origin \( t = (t_1, \ldots, t_n) = 0 \) and by Implicit Function Theorem there exists a constant \( c > 0 \) such that
\[
x + c\nu B \subset \Phi^\varepsilon(\nu B, x), \quad B = B(0, 1) \subset \mathbb{R}^n, \quad x \in K, \quad (10.32)
\]
where \( c \) is independent of \( \varepsilon \) and \( \nu \) is small enough.

Let us denote now with \( E_x \) the end-point map based at the point \( x \in \mathbb{R}^n \) (with analogous meaning for \( E^\varepsilon_x, \hat{E}_x \)) and with \( B_{L^2} \) the unit ball in \( L^2_x[0, 1] \).

We claim that (10.32) implies that there exists a constant \( c' \) such that
\[
x + c'\nu B \subset E^\varepsilon_x(\nu \frac{1}{w_i} B_{L^2}), \quad \forall \nu, \varepsilon > 0 \quad (10.33)
\]
Since \( t \mapsto u_i(t, \cdot) \) is a smooth map for every \( i \), and \( u_i(0, \cdot) = 0 \) we have that there exist a constant \( c_i \) such that
\[
t \in \nu B \Rightarrow u_i(t, \cdot) \in c_i \nu B_{L^2}, \quad (10.34)
\]
\[
t \in \nu B \Rightarrow u_i(t^{1/w_i}, \cdot) \in c_i \nu^{1/w_i} B_{L^2}, \quad (10.35)
\]
for all \( \nu > 0 \) small enough.
For such \( \nu \) we have by inclusion (10.33) that

\[ |x - y| \leq c \nu \implies d^\nu(x, y) \leq \nu^{1/m} \]

where we used the fact that \( d^\nu \) is the infimum of norm of \( u \) such that \( E^\nu_x(u) = y \). From this easily follows

\[ d^\nu(x, y) \leq c \frac{1}{m} |x - y|^\frac{1}{m} \quad (10.36) \]

\[ \square \]

Remark 10.51. All estimates are valid also for \( \varepsilon \to 0 \), i.e. for the nilpotent approximation. In particular, using homogeneity

\[ \hat{d}(x, y) \leq C|x - y|^\frac{1}{m}, \quad \forall x, y \in \mathbb{R}^n \quad (10.37) \]

Indeed from the proof of Lemma 10.49 it follows that the estimate (10.37) holds in a compact \( K \) containing the origin. Consider two arbitrary points \( x, y \in \mathbb{R}^n \) and \( \varepsilon > 0 \) such that \( \delta_x, \delta_y \in K \). By the homogeneity of the distance

\[ \hat{d}(\delta_x, \delta_y) = \varepsilon \hat{d}(x, y). \]

Moreover since the estimate (10.37) holds in \( K \)

\[ \hat{d}(\delta_x, \delta_y) \leq C|\delta_x - \delta_y|^\frac{1}{m} \leq C|\delta_x - \delta_y|^\frac{1}{m} \]

We can state now the main result

**Theorem 10.52.** \( \rho^\varepsilon \to \hat{\rho} \) uniformly on compacts in \( \mathbb{R}^n \).

**Proof.** By Lemma 10.49 it is sufficient to prove pointwise convergence. We prove the following inequalities

\[ \limsup_{\varepsilon \to 0^+} \rho^\varepsilon(x) \leq \hat{\rho}(x) \leq \liminf_{\varepsilon \to 0^+} \rho^\varepsilon(x) \quad (10.38) \]

(i) Fix a point \( x \) and a control \( \hat{u} \) such that

\[ \hat{E}(\hat{u}) = x, \quad ||\hat{u}|| = \hat{\rho}(x), \]

i.e. such that the corresponding trajectory is a minimizer for the system \( \hat{f} \). Now consider \( x^\varepsilon := E^\varepsilon(\hat{u}) \). From Proposition 10.48 we get \( x^\varepsilon \to x \) for \( \varepsilon \to 0 \). Moreover, from the definition of \( \rho^\varepsilon \) we have \( \rho^\varepsilon(x^\varepsilon) \leq \hat{\rho}(x) \). Hence

\[ \rho^\varepsilon(x) = \rho^\varepsilon(x^\varepsilon) + \rho^\varepsilon(x) - \rho^\varepsilon(x^\varepsilon) \leq \hat{\rho}(x) + |\rho^\varepsilon(x) - \rho^\varepsilon(x^\varepsilon)| \]

Using that \( \rho^\varepsilon \) is an equicontinuous family and that \( x^\varepsilon \to x \) we have the left inequality in (10.38).
Let now $u^\varepsilon$ be a control such that 

$$E^\varepsilon(u^\varepsilon) = x, \quad \|u^\varepsilon\| = \rho^\varepsilon(x)$$

and define $x^\varepsilon := \hat{E}(u^\varepsilon)$. As before we have $\hat{\rho}(x^\varepsilon) \leq \rho^\varepsilon(x)$. Then 

$$\hat{\rho}(x) = \hat{\rho}(x^\varepsilon) + \hat{\rho}(x) - \hat{\rho}(x^\varepsilon)$$

$$\leq \rho^\varepsilon(x) + |\hat{\rho}(x) - \hat{\rho}(x^\varepsilon)|$$

and now it is sufficient to notice that $x^\varepsilon = E^\varepsilon(u^\varepsilon) \to \hat{E}(u^\varepsilon) = x$ since $E^\varepsilon \to \hat{E}$ uniformly on balls of $L^2$ and $u^\varepsilon$ bounded since $\rho^\varepsilon$ are equicontinuous.

In privileged coordinates $x = (x_1, \ldots, x_m) \in \mathbb{R}^{k_1} \oplus \cdots \oplus \mathbb{R}^{k_m} = \mathbb{R}^n$ we set 

$$\Pi_\varepsilon = \{x \in \mathbb{R}^n, |x_i| \leq \varepsilon^i, i = 1, \ldots, m\}$$

**Corollary 10.53** (Ball-Box Theorem). There exists constants $c_1, c_2 > 0$ such that 

$$c_1 \Pi_\varepsilon \subset B(x, \varepsilon) \subset c_2 \Pi_\varepsilon$$

where $B(x, \varepsilon)$ is the subriemannian ball in privileged coordinates.

Notice that this is a weaker statement with respect to Theorem [10.52]

**Exercise 10.54.** Prove Corollary [10.53]

**Definition 10.55.** Let $f$ and $\tilde{f}$ be two sub-Riemannian structures on the same manifold $M$. We say that the structures are locally Lipschitz equivalent if, for any compact $K \subset M$ there exist $c_1, c_2 > 0$ such that 

$$c_1 d(x, y) \leq \tilde{d}(x, y) \leq c_2 d(x, y)$$

where $\mu$ and $\tilde{\mu}$ are respectively the sub-Riemannian distances induced by $f$ and $\tilde{f}$.

From the Ball-Box Theorem we easily get a characterization of locally Lipschitz equivalent structures in term of the distribution.

**Corollary 10.56.** Two sub-Riemannian structures are locally Lipschitz equivalent if and only if the two flags are equal at al points, i.e. 

$$D^i_q = \tilde{D}^i_q, \quad \forall q \in M, \quad \forall i \geq 1.$$ 

**Corollary 10.57.** Two regular sub-Riemannian structures are locally Lipschitz equivalent if and only if their distributions are equal at al points, i.e. 

$$D_q = \tilde{D}_q, \quad \forall q \in M.$$ 

In other words, in the regular case, the distribution define the metric up to locally Lipschitz equivalence.

**Remark 10.58.** In the proof of Theorem [10.52] we showed that, in some coordinates, the sub-Riemannian metric has a holder estimate with respect to the Euclidean one. The fact that the metric is Lipschitz equivalent to the Euclidean one characterize exactly Riemannian structures on $M$.

Moreover we notice that this is only local property since we do not study the behaviour of the constants $c_1, c_2$ when $K$ become big.
10.5 Algebraic meaning

In the last section we proved in which sense the sub-Riemannian tangent space approximate the sub-Riemannian structure on the manifold. Now we also show that, at least in the regular case, the nilpotent approximation has a structure of Lie group, endowed with a left-invariant sub-Riemannian structure.

Recall that given an orthonormal frame \( \{ f_1, \ldots, f_k \} \) for the sub-Riemannian structure, by Proposition 10.21 the vector field \( J_q^m f_i \), jet of a vector field on \( M \), is a well defined vector field on the quotient \( T^f_q := J_q^f / \sim \), which we denote \( \hat{f}_i \).

**Proposition 10.59.** The Lie algebra \( \text{Lie}\{\hat{f}_1, \ldots, \hat{f}_k\} \) is a nilpotent Lie algebra of step \( m \), where \( m \) is the nonholonomic degree of \( f \) at \( q \).

**Proof.** Consider privileged coordinates around the point \( q \). Then \( \hat{f}_i \) has weight \(-1\) and is homogeneous with respect to the dilation \( \delta_\lambda \). Moreover for any bracket monomial we have
\[
\nu([\hat{f}_{i_1}, \ldots, [\hat{f}_{i_{j-1}}, \hat{f}_{i_j}]]) = -j
\]
Since every vector field \( V \), when written in privileged coordinates, satisfies \( \nu(V) \geq -m \), then every bracket of \( m \) vector fields is necessarily zero. \( \square \)

Consider now the group generated by the flows of these vector fields
\[
G = \text{Gr}\{e^{t\hat{f}_1}, \ldots, e^{t\hat{f}_k}\}
\]
which acts on \( T^f_q \) on the right, and is by definition a nilpotent Lie group.\(^1\) Moreover in the proof of Theorem 10.27 we showed that this action is also transitive (i.e. we can realize every element of \( T^f_q \) with this action).

Collecting together all these results we have

**Corollary 10.60.** The nilpotent approximation \( T^f_q \) is a homogeneous space, diffeomorphic to the quotient \( G/G_0 \), where \( G_0 \) is the isotropy group of the trivial element of \( T^f_q \).

Before interpreting this contruction at the level of Lie algebras, we recall some definitions.

The free associative algebra on \( k \) generators \( x_1, \ldots, x_k \) is the associative algebra \( A_k \) of linear combinations of words of its generators, where the product of two element is defined by juxtaposition. The free Lie algebra on \( k \) generators, denoted \( \mathcal{L}_k \), is the algebra of Lie elements of \( A_k \) where the product of two elements \( x, y \) is defined by the commutator \( [x, y] = xy - yx \).

The nilpotent step \( m \) free Lie algebra on \( k \) generators \( x_1, \ldots, x_k \) is the quotient of the free Lie algebra by the ideal \( \mathcal{I}^{m+1} \) generated as follows: \( \mathcal{I}^1 = \mathcal{L} \), and \( \mathcal{I}^j = [\mathcal{I}^{j-1}, \mathcal{L}] \).

Let \( \text{Lie}_m\{X_1, \ldots, X_k\} \) be the nilpotent step \( m \) free Lie algebra generated by the vector fields \( X_1, \ldots, X_k \) and consider the subalgebra
\[
C := \{ \pi \in \text{Lie}_m\{X_1, \ldots, X_k\} \mid \pi(\hat{f}_1, \ldots, \hat{f}_k)(0) = 0 \}
\]
of all polynomial bracket such that if we replace \( X_i \) with \( \hat{f}_i \) are zero when evaluated at zero. Then
\[
\text{Lie} T^f_q \simeq \text{Lie}_m\{X_1, \ldots, X_k\}/C
\]
\(^1\)A Lie group \( G \) is nilpotent if its Lie algebra \( g \) is nilpotent. The fact that \( G \) acts on the right is because right action satisfies \( R_{hg} = R_h R_g \) (i.e. \( x \cdot (hg) = (x \cdot h) \cdot g \)).
Remark 10.61. To discuss regularity properties of $T_q^f$ with respect to $q$, we can restate this characterization in such a way that does not depend on the nilpotent approximation:

$$\text{Lie } T_q^f \simeq \text{Lie}_m\{X_1, \ldots, X_k\}/C_q$$

where $C_q$ is the core subalgebra

$$C_q := \{ \pi \in \text{Lie}_m\{X_1, \ldots, X_k\} \mid \pi(f_1, \ldots, f_k)(q) \in \mathcal{D}_q^{\deg \pi - 1} \}$$

Lemma 10.62. Assume that the sub-Riemannian structure has constant growth vector, i.e. that $n_i(q) = \dim \mathcal{D}_q^i$ does not depend on $q$. Then $C_q$ is an ideal.

In particular $T_q^f$ is a Lie group.

Proof. It is sufficient to prove that

$$X \in C_q \implies [f_i, X] \in C_q, \quad \forall i = 1, \ldots, k$$

Since the structure has constant growth vector, we can consider an adapted basis $V_1, \ldots, V_n$, well defined in a neighborhood $O_q$ of $q$. In particular if $X = \pi(f_1, \ldots, f_k)$ is a bracket polynomial of degree $\deg \pi = d$ we can write

$$X(q') = \sum_{i:w_i\leq d} a_i(q')V_i(q'), \quad \forall q' \in O_q$$

where $a_i$ are suitable smooth functions. From (10.39) we have that $X \in C_q$ if and only if it belongs to $\mathcal{D}_q^{d-1}$, i.e. $a_i(q) = 0, \forall i \text{ s.t. } w_i = d$. On the other hand

$$[f_i, X] = [f_i, \sum_{w_j \leq d} a_jV_j]$$

$$= \sum_{w_j \leq d} a_j[f_i, V_j] + f_i(a_j)V_j$$

(10.40)

From this equality it is easy to check that every coefficient of degree $d + 1$ in this sum is null at $q$, since they can appear only in the first summand of (10.40).

Corollary 10.63. Under previous assumptions $\hat{f}_1, \ldots, \hat{f}_k$ are a basis of left-invariant vector fields on $T_q^f$.

Proof. All relies on the fact that if we consider a left invariant vector field $X$ on a Lie group $G$, and we consider the right action of a normal subgroup $H$ on it, then $X$ is a well defined left-invariant vector field on the quotient $G/H$, which is still a Lie group.

Examples

Heisenberg
Martinet
Grushin
Chapter 11

Regularity of the sub-Riemannian distance

In this chapter we focus our attention on the analytical properties of the sub-Riemannian squared distance from a fixed point. In particular we want to answer to the following questions:

(i) Which is the (minimal) regularity of $d^2$ that one can expect?

(ii) Is the sub-Riemannian distance $d^2$ smooth? If not, can we characterize smooth points?

11.1 General properties of the distance function

In this section we recall and collect some general properties of the sub-Riemannian distance and results related to it, some of which we already proved in the previous chapters.

Let us consider a free sub-Riemannian structure $(M, U, f)$ where the vector fields $f_1, \ldots, f_m$ define a generating family, i.e.

$$f : U \to TM, \quad f(u, q) = \sum_{i=1}^{m} u_i f_i(q)$$

Here $U$ is a trivial Euclidean bundle on $M$ of rank $m$.

Definition 11.1. Fix a point $q \in M$. The flag of the sub-Riemannian structure at the point $q$ is the sequence of subspaces $\{D^i_q\}_{i \in \mathbb{N}}$ defined by

$$D^i_q := \text{span}\{[f_{j_1}, \ldots, [f_{j_{l-1}}, f_{j_l}]](q), \forall l \leq i\}$$

Notice that $D^1_q = D_q$ is the set of admissible directions. Moreover, by construction, $D^i_q \subset D^{i+1}_q$ for all $i \geq 1$.

The bracket generating assumptions implies that

$$\forall q \in M, \ \exists m(q) > 0 \ \text{s.t.} \ D^m_q = T_q M$$

and $m(q)$ is called the step of the sub-Riemannian structure at $q$. 

223
Exercise 11.2. 1. Prove that the filtration defined by the subspaces $D_q^i$, for $i \geq 1$, is independent on the choice of a generating family (i.e., on the trivialization of $U$).

2. Show that $m(q)$ does not depend on the generating frame. Prove that the map $q \mapsto m(q)$ is upper semicontinuous.

In Chapter [10] we already proved that the sub-Riemannian distance is Hölder continuous. For the reader’s convenience, we recall here the statement.

**Proposition 11.3.** For every $q \in M$ there exists a neighborhood $O_q$ such that $\forall q_0, q_1 \in O_q$ and for every coordinate map $\phi : O_q \to \mathbb{R}^n$

$$d(q_0, q_1) \leq C|\phi(q_0) - \phi(q_1)|^{1/m}$$

where $m = m(q)$ is the step of the sub-Riemannian structure at $q$.

In what follows we fix a point $q_0 \in M$ to be fixed and $r_0 > 0$ such that $B = B_{q_0}(r_0)$ is a closed compact ball centered in $q_0$.

Let us denote by $E = E_{q_0} : U \to M$ the end-point map based at $q_0 \in M$, i.e., the map that associates to every control $u(\cdot) \in U \subset L^2$ the end-point $q_u(1)$ of the solution associated to the control $u$ (we recall that $U$ is the open set of $L^2$ such that the corresponding solution $q_u(\cdot)$ is defined on $[0, 1]$). Denote with $B$ the ball of radius $r_0$ in $L^2$ (where $r_0$ is chosen in such a way that the closure of $B_{q_0}(r_0)$ is compact). Notice that since $B$ is compact then $B \subset U$.

**Proposition 11.4.** $F \mid_B : B \to M$ is continuous in the weak topology. In other words if $u_n \rightharpoonup u$ in the weak-$L^2$ topology then $F(u_n) \to F(u)$.

Proof. Consider the solution of the problem

$$\dot{\gamma}(t) = f_u(t)(\gamma(t)), \quad \gamma(0) = q_0, \quad u \in B.$$  

Since the ball $B$ is compact, all trajectories are Lipschitzian with the same Lipchitz constant. In particular this set has compact closure in the $C^0$ topology.

Assume now that $u_n \rightharpoonup u$ and consider the family of curves $\gamma_n(t)$ associated to $u_n$, that satisfy

$$\gamma_n(t) = q_0 + \int_0^t f_{u_n(\tau)}(\gamma_n(\tau))d\tau.$$  

By compactness there exists a subsequence, which we still denote $\gamma_n$, such that $\gamma_n \to \gamma$ uniformly, for some curve $\gamma$, in particular their endpoints converge. It remains to show that $\gamma$ is the trajectory associated to $u$.

Since $u_n \rightharpoonup u$ we have that $f_{u_n(t)}(\gamma_n(t)) \rightharpoonup f_u(t)(\gamma(t))$ being the product between strong and weak convergent sequences,\footnote{one can write the coordinate expression $\sum u_k^i f_i(q_k(t))$} taking the limit we find

$$\gamma(t) = q_0 + \int_0^t f_u(\tau)(\gamma(\tau))d\tau,$$

i.e. $\gamma$ is the trajectory associated to $u$. \qed
Remark 11.5. Actually we prove that all trajectories converge uniformly and not only their endpoints.

The previous proposition given another proof of the existence of minimizers

**Corollary 11.6 (Existence of minimizers).** For any \( q \in \overline{B}_{q_0}(r) \) there exists \( u \) (with \( \|u\| \leq r \)) that join \( q_0 \) and \( q \) and is a minimizer, i.e. \( \|u\| = d(q_0, q) \).

**Proof.** Consider a point \( q \) in the compact ball \( B \). Then take a minimizing sequence \( u_n \) such that \( F(u_n) = q \) and \( \|u_n\| \to d(q_0, q) \). The sequence \( \|u_n\| \) is bounded, hence by weak compactness of balls in \( L^2 \) there exists a subsequence, that we still call \( u_n \) such that \( u_n \rightharpoonup u \) for some \( u \). By continuity \( F(u) = q \). Moreover the semicontinuity of the \( L^2 \) norm proves that \( u \) corresponds to a minimizer joining \( q_0 \) to \( q \) since

\[
\|u\| \leq \liminf_{n \to \infty} \|u_n\| = d(q_0, q).
\]

**Definition 11.7.** A control \( u \) is called a minimizer if it satisfies \( J(u) = \frac{1}{2}d^2(q_0, F(u)) \). Notice that in this case we have \( \|u\| = d(q_0, F(u)) \).

We denote by \( \mathcal{M} \subset L^2 \) the set of all minimizing controls.

**Theorem 11.8 (Compactness).** Let \( K \subset M \) be compact. The set of all minimal controls associated with trajectories reaching \( K \)

\[
\mathcal{M}_K = \{ u \in \mathcal{M} \mid F(u) \in K \},
\]

is compact in the strong \( L^2 \) topology.

**Proof.** Consider a sequence \( u_n \in \mathcal{M}_K \). Since \( K \) is compact, the sequence \( \|u_n\| \) is bounded. Since bounded sets in \( L^2 \) are weakly compact, we can assume that \( u_n \rightharpoonup u \). Let us show that we also have \( \|u_n\| \to \|u\| \).

From Proposition 11.4 it follows that \( F(u_n) \to F(u) \) in \( M \) and the continuity of the distance implies \( d(q_0, F(u_n)) \to d(q_0, F(u)) \). Moreover since \( u_n \in \mathcal{M} \) we have that \( \|u_n\| = d(q_0, F(u_n)) \) and by weak semicontinuity of the \( L^2 \) norm we get

\[
\|u\| \leq \liminf_{n \to \infty} \|u_n\| = \liminf_{n \to \infty} d(q_0, F(u_n)) = d(q_0, F(u)).
\]

Hence \( u_n \to u \) strongly in \( L^2 \) and \( u \in \mathcal{M} \). 

### 11.2 Regularity of the squared distance

In this section we fix once for all a point \( q_0 \in M \) and a closed ball \( B = \overline{B}_{q_0}(r_0) \) such that \( B \) is compact. In particular for each \( q \in B \) there exists a minimizer joining \( q_0 \) and \( q \) (see Corollary 11.6). In what follows we denote by \( f \) the squared distance from \( q_0 \)

\[
f(\cdot) = \frac{1}{2}d^2(q_0, \cdot).
\]

The main result of this chapter is the following.
Theorem 11.9. The function $\mathcal{f}|_B : B \to \mathbb{R}$ is smooth on a open dense subset of $B$.

In the case of complete sub-Riemannian structures, since balls are compact for all radii, we have immediately the following corollary

Corollary 11.10. Assume that $M$ is a complete sub-Riemannian manifold and $q_0 \in M$. Then $\mathcal{f}$ is smooth on an open and dense subset of $M$.

We start by looking for necessary conditions for $\mathcal{f}$ to be $C^\infty$ around a point.

Proposition 11.11. Let $q \in B$ and assume that $\mathcal{f}$ is $C^\infty$ at $q$. Then

(i) there exists a unique length minimizer $\gamma$ joining $q_0$ with $q$. Moreover $\gamma$ is not abnormal and not conjugate.

(ii) $d_q \mathcal{f} = \lambda_1$, where $\lambda_1$ is the final covector of the normal lift of $\gamma$.

Proof. Under the above assumptions the functional

$$\Psi : v \mapsto J(v) - \mathcal{f}(F(v)), \quad v \in L^\infty([0,T], \mathbb{R}^k),$$

is smooth and non negative. For every optimal trajectory $\gamma$, associated with the control $u$, that connects $q_0$ with $q$ in time 1, one has

$$0 = d_u \Psi = d_u J - d_q \mathcal{f} \circ D_u F.$$  \hspace{1cm} (11.3)

Thus, $\gamma$ is a normal extremal trajectory, with Lagrange multiplier $\lambda_1 = d_q \mathcal{f}$. By Theorem 4.24, we can recover $\gamma$ by the formula $\gamma(t) = \pi \circ e^{(t-1)\tilde{H}}(\lambda_1)$. Then, $\gamma$ is the unique minimizer of $J$ connecting its endpoints, and is normal.

Next we show that $\gamma$ is not abnormal and not conjugate. For $y$ in a neighbourhood $O_q$ of $q$, let us consider the map

$$\Phi : O_q \mapsto T^*_{q_0}M, \quad \Phi(y) = e^{-\tilde{H}}(d_q \mathcal{f}).$$

(11.4)

The map $\Phi$, by construction, is a smooth right inverse for the exponential map, since

$$E(\Phi(y)) = \pi \circ e^{\tilde{H}}(e^{-\tilde{H}}(d_q \mathcal{f})) = \pi(d_q \mathcal{f}) = y.$$  \hspace{1cm} (11.5)

This implies that $q$ is a regular value for the exponential map. Since $q$ is a regular value for the exponential map and, a fortiori, $u$ is a regular point for the end-point map. This proves that $u$ corresponds to a trajectory that is at the same time strictly normal and not conjugate. \hfill \Box

Remark 11.12. Notice that from the proof it follows that if we only assume that $\mathcal{f}$ is differentiable at $q$, we can still conclude that there exists a unique minimizer $\gamma$ joining $q_0$ to $q$, and it is normal.

Before going further in the study of the smoothness property of the distance function, we are already able to prove an important corollary of this result.
Denote, for \( r > 0 \), \( S_r := f^{-1}(\frac{r^2}{2}) \) the sub-Riemannian sphere of radius \( r \) centered at \( q_0 \).

**Corollary 11.13.** Assume that \( D_{q_0} \neq T_{q_0} M \). For every \( r \leq r_0 \), the sphere \( S_r \) contains a non smooth point of the function \( f \).

**Proof.** Since \( r \leq r_0 \), the sphere \( S_r \) is non empty and contained in a compact ball. Assume, by contradiction, that \( f \) is smooth at every point of \( S_r \). Then \( S_r \) is a level set defined by \( f \) and \( d_{q} f \neq 0 \) for every \( q \in S_r \) (since \( d_{q} f \) is the nonzero covector attached at the final point of a geodesic, see Proposition 11.11). It follows that \( S_r \) is a smooth submanifold of dimension \( n - 1 \), without boundary. Moreover, being the level set of a continuous function, \( S_r \) is closed, hence compact.

Let us consider the map

\[
\Phi : S_r \to T_{q_0}^* M, \quad \Phi(q) = e^{-H}(d_q f),
\]

By assumption \( f \) is smooth, hence \( \Phi \) is a smooth right inverse of the exponential map (see also (11.5)). In particular the differential of \( \Phi \) is injective at every point. Moreover \( H(\Phi(q)) = r \) since \( f(q) = H(\lambda) = r \) for every \( q \in S_r \). It follows that actually \( \Phi \) defines a smooth immersion

\[
\Phi : S_r \to H^{-1}(r) \cap T_{q_0}^* M
\]

of the sphere \( S_r \) into the set

\[
C_r := H^{-1}(r) \cap T_{q_0}^* M = \left\{ \lambda \in T_{q_0}^* M : \frac{1}{2} \sum_{i=1}^{k} \langle \lambda, f_i(q_0) \rangle^2 = r \right\}
\]

Notice that \( C_r \) is a smooth connected and non compact \( n - 1 \) dimensional submanifold of the fiber \( T_{q_0}^* M \), indeed diffeomorphic to the cylinder \( S^{k-1} \times \mathbb{R}^{n-k} \) (here \( k = \dim D_{q_0} < n \) is the rank of the structure at the point \( q_0 \)). By continuity of \( \Phi \), the image \( \Phi(S_r) \) is closed in \( C_r \). Moreover, since every immersion is a local submersion and \( \dim S_r = \dim C_r \), the set \( \Phi(S_r) \) is also open in \( C_r \). Hence it is connected. Since \( \Phi(S_r) \) has no boundary, it is a connected component of \( C_r \), namely \( \Phi(S_r) = C_r \). This is a contradiction since, by continuity, \( \Phi(S_r) \) is compact, while \( C_r \) is not. \( \square \)

Next we go back to the proof of the main result. Recall that \( q_0 \in M \) is fixed and \( f \) is the one half of the distance squared from \( q_0 \). After Proposition 11.11 it is natural to introduce the following definition.

**Definition 11.14.** Fix a point \( q_0 \in M \). The set of smooth point from \( q_0 \) is the set \( \Sigma \subset M \) of \( q \in M \) such that there exists a unique length-minimizer \( \gamma \) joining \( q_0 \) to \( q \), that it is strictly normal, and not conjugate.

From the proof of Proposition 11.11 (see also Remark 11.12), it follows that if the squared distance \( f \) from \( q_0 \), is smooth at \( q \) then \( q \in \Sigma \). The name smooth point of \( f \) is justified by the following theorem.

**Theorem 11.15.** The set \( \Sigma \) is open and dense in \( B \). Moreover \( f \) is smooth at every point of \( \Sigma \).

**Proof.** We divide the proof into three parts: (a) the set \( \Sigma \) is open, (b) the function \( f \) is smooth in a neighborhood of every point of \( \Sigma \), (c) the set \( \Sigma \) is dense in \( B \).
(a). To prove that \( \Sigma \) is open we have to show that for every \( q \in \Sigma \) there exists a neighborhood \( O_q \) of \( q \) such that every \( q' \in O_q \) is also in \( \Sigma \).

Let us start by proving the following claim: there exists a neighborhood of \( q \) in \( B \) such that every point in this neighborhood is reached by exactly one minimizer.

By contradiction, if this property is not true, there exists a sequence \( q_n \) of points in \( B \) converging to \( q \) such that (at least) two minimizers \( \gamma_n \) and \( \gamma'_n \) joining \( q_0 \) and \( q_n \). Let us denote by \( u_n \) and \( v_n \) the corresponding minimizing controls.

By Proposition 1.1.8 the set of controls associated with minimizers whose endpoint is in the compact ball \( B \) is compact in \( L^2 \) (w.r.t. the strong topology). Then there exist, up to considering a subsequence, two controls \( u, v \) such that \( u_n \to u \) and \( v_n \to v \). Moreover the limits \( u \) and \( v \) are both minimizers and join \( q_0 \) with \( q \). Since by assumption there is a unique minimizer \( \gamma \) joining \( q_0 \) with \( q \), it follows that \( u = v \) is the corresponding control.

By smoothness of the end point map both \( D_{u_n} F \) and \( D_{v_n} F \) tends to \( D_u F \), which has has full rank (\( u \) is strictly normal, hence is not a critical point for \( F \)). Hence, for \( n \) big enough, both \( D_{u_n} F \) and \( D_{v_n} F \) are surjective, i.e., \( u_n \) and \( v_n \) are strictly normal, and we can build the sequence \( \lambda^n_1 \) and \( \xi^n_1 \) of corresponding final covectors in \( T_{q_n} M \) satisfying

\[
\lambda^n_1 D_{u_n} F = u_n, \quad \xi^n_1 D_{v_n} F = v_n.
\]

These relations can be rewritten in terms of the adjoint linear maps

\[
(D_{u_n} F)^* \lambda^n_1 = u_n, \quad (D_{v_n} F)^* \xi^n_1 = v_n.
\]

Since both \((D_{u_n} F)^*\) and \((D_{v_n} F)^*\) are a family of injective linear maps converging to \((D_u F)^*\) and \( u_n, v_n \to u \), it follows that the corresponding (unique) solutions \( \lambda^n_1 \) and \( \xi^n_1 \) also converge to the solution of the limit problem \((D_u F)^* \lambda_1 = u \), i.e., both converge to the final covector \( \lambda_1 \) corresponding to \( \gamma \). By using the flow defined by the corresponding controls we can deduce the convergence of the sequences \( \lambda^n_0 \) and \( \xi^n_0 \) of the initial covectors associated to \( u_n \) and \( v_n \) to the unique initial covector \( \lambda_0 \) corresponding to \( \gamma \).

Finally, since \( \lambda_0 \) by assumption is a regular point of the exponential map, i.e., the unique minimizer \( \gamma \) joining \( q_0 \) to \( q \) is not conjugate, it follows that the exponential map is invertible in a neighborhood \( V_{\lambda_0} \) of \( \lambda_0 \) onto its image \( O_q := E(V_{\lambda_0}) \), that is a neighborhood of \( q \). In particular this proves our initial claim.

More precisely we have proved that for every point \( q' \in O_q \) there exists a unique minimizer joining \( q_0 \) to \( q' \), whose initial covector \( \lambda' \in V_{\lambda} \) is a regular point of the exponential map. This implies that every \( q' \in O_q \) is a smooth point, and \( \Sigma \) is open.

(b). Now we prove that \( f \) is smooth in a neighborhood of each point \( q \in \Sigma \). From the part (a) of the proof it follows that if \( q \in \Sigma \) there exists a neighborhood \( V_{\lambda_0} \) of \( \lambda_0 \) and \( O_q \) of \( q \) such that \( E|_{V_{\lambda_0}} : V_{\lambda_0} \to O_q \) is a smooth invertible map. Denote by \( \Phi : O_q \to V_{\lambda_0} \) its smooth inverse. Since for every \( q' \in O_q \) there is only one minimizer joining \( q_0 \) to \( q' \) with initial covector \( \Phi(q') \) it follows that,

\[
f(q') = \frac{1}{2} d^2(q_0, q') = H(\Phi(q')),
\]

that is a composition of smooth functions, hence smooth.

(c). Our next goal is to show that \( \Sigma \) is a dense set in \( B \). We start by a preliminary definition.
Definition 11.16. A point \( q \in B \) is said to be

(i) a *fair point* if there exists a unique minimizer joining \( q_0 \) to \( q \), that is normal.

(ii) a *good point* if it is a fair point and the unique minimizer joining \( q_0 \) to \( q \) is strictly normal.

We denote by \( \Sigma_f \) and \( \Sigma_g \) the set of fair and good points, respectively.

We stress that a fair point can be reached by a unique minimize \( r \) that is both normal and abnormal. From the definition it is immediate that \( \Sigma \subset \Sigma_g \subset \Sigma_f \). The proof of (c) relies on the following four steps:

(c1) \( \Sigma_f \) is a dense set in \( B \),

(c2) \( \Sigma_g \) is a dense set in \( B \),

(c3) \( f \) is Lipschitz in a neighborhood of every point of \( \Sigma_g \),

(c4) \( \Sigma \) is a dense set in \( B \).

(c1). Fix an open set \( O \subset B \) and let us show that \( \Sigma_f \cap O \neq \emptyset \). Consider a smooth function \( a : O \to \mathbb{R} \) such that \( a^{-1}([s, +\infty[) \) is compact for every \( s \in \mathbb{R} \). Then consider the function

\[
\psi : O \to \mathbb{R}, \quad \psi(q) = f(q) - a(q)
\]

The function \( \psi \) is continuous on \( O \) and, since \( f \) is nonnegative, the set \( \psi^{-1}(-\infty, s[) \) are compact for every \( s \in \mathbb{R} \) due to the assumption on \( a \). It follows that \( \psi \) attains its minimum at some point \( q_1 \in O \). Define a control \( u_1 \) associated with a minimizer \( \gamma \) joining \( q_0 \) and \( F(u_1) = q_1 \).

Since \( J(u) \geq f(F(u)) \) for every \( u \), it is easy to see that the map

\[
\Phi : \mathcal{U} \to \mathbb{R}, \quad \Phi(u) = J(u) - a(F(u))
\]

attains its minimum at \( u_1 \). In particular it holds

\[
0 = D_{u_1} \Phi = u_1 - (d_{q_1} a) D_{u_1} F.
\]

The last identity implies that \( u_1 \) is normal and \( \lambda_1 = d_{q_1} a \) is the final covector associated with the trajectory. By Theorem 4.24 the corresponding trajectory \( \gamma \) is uniquely recovered by the formula \( \gamma(t) = \pi \circ e^{(t-1)\tilde{H}}(d_{q_1} a) \). In particular \( \gamma \) is the unique minimizer joining \( q_0 \) to \( q_1 \in O \), and is normal, i.e. \( q_1 \in \Sigma_f \cap O \).

Remark 11.17. In the Riemannian case \( \Sigma_f = \Sigma_g \) since there are no abnormal extremal.

(c2). As in the proof of (c1), we shall prove that \( \Sigma_g \cap O \neq \emptyset \) for any open \( O \subset B \). By (c1) the set \( \Sigma_f \cap O \) is nonempty. For any \( q \in \Sigma_f \cap O \) we can define \( \text{rank } q := \text{rank } D_u F \), where \( u \) is the control associated to the unique minimizer \( \gamma \) joining \( q_0 \) to \( q \). To prove (c2) it is sufficient to prove that there exists a point \( q' \in \Sigma_f \cap O \) such that \( \text{rank } q' = n \) (i.e., \( D_{u'} F \) is surjective, where \( u' \) is the control associated to the unique minimizer joining \( q_0 \) and \( q' \)). Assume by contradiction that

\[
k_O := \max_{q \in \Sigma_f \cap O} \text{rank } q < n,
\]

and consider a point \( \tilde{q} \) where the maximum is attained, i.e., such that \( \text{rank } \tilde{q} = k_O \).
We claim that all points of $\Sigma_f \cap O$ that are sufficiently close to $q$ have the same rank (we stress that the existence of points in $\Sigma_f \cap O$ arbitrary close to $q$ is also guaranteed by (c1)).

Assume that the claim is not true, i.e., there exist a sequence of points $q_n \in \Sigma_f \cap O$ such that $q_n \to q$ and $\text{rank } q_n \leq k_O - 1$. Reasoning as in the proof of (a), using uniqueness and compactness of the minimizers, one can prove that the sequence of controls $u_n$ associated to the unique minimizers joining $q_0$ to $q_n$ satisfies $u_n \to \hat{u}$ strongly in $L_2$, where $\hat{u}$ is the control associated to the unique minimizer joining $q_0$ with $q$. By smoothness of the end-point map $F$ it follows that $D_{u_n} F \to D_{\hat{u}} F$ which, by semicontinuity of the rank, implies the contradiction
\[
\text{rank } \hat{q} = \text{rank } D_{\hat{u}} F \leq \liminf_{n \to \infty} \text{rank } D_{u_n} F \leq k_O - 1.
\]

Thus, without loss of generality, we can assume that $\text{rank } q = k_O < n$ for every $q \in \Sigma_f \cap O$ (maybe by restricting our neighborhood $O$). We introduce the following set
\[
\Pi_q = e^{-H} \{ \xi \in T_q^* M \mid \xi D_u F = \lambda_1 D_u F \} \subset T_{q_0}^* M.
\]

The set $\Pi_q$ is the set of initial covector $\lambda_0 \in T_{q_0}^* M$ whose image via the exponential map is the point $q$.

**Lemma 11.18.** $\Pi_q$ is an affine subset of $T_{q_0}^* M$ such that $\dim \Pi_q = n - k_O$. Moreover the map $q \mapsto \Pi_q$ is continuous.

**Proof.** It is easy to check that the set $\hat{\Pi}_q = \{ \xi \in T_q^* M \mid \xi D_u F = \lambda_1 D_u F \}$ is an affine subspace of $T_{q_0}^* M$. Indeed $\xi \in \Pi_q$ if and only if $(D_u F)^*(\xi - \lambda_1) = 0$, that is
\[
\hat{\Pi}_q = \{ \xi \in T_q^* M \mid \xi D_u F = \lambda_1 D_u F \} = \lambda_1 + \text{Ker } (D_u F)^*.
\]

Moreover $\dim \text{Ker } (D_u F)^* = n - \dim \text{Im } D_u F = n - k_O$. Since all elements $\xi \in \hat{\Pi}_q$ are associated with the same control $u$, we have that $\Pi_q = e^{-H}(\hat{\Pi}_q) = P_{\hat{\Pi}_q}(\hat{\Pi}_q)$, hence $\Pi_q$ is an affine subspace of $T_{q_0}^* M$.

Let us now show that the map $q \mapsto \Pi_q$ is continuous on $\Sigma_f \cap O$. Consider a sequence of points $q_n$ in $\Sigma_f \cap O$ such that $q_n \to q \in \Sigma_f \cap O$. Let $u_q$ (resp. $u$) be the unique control associated with the minimizing trajectory joining $q_0$ and $q_n$ (resp. $q$). By the uniqueness-compactness argument already used in the previous part of the proof we have that $u_n \to u$ strongly and moreover $D_{u_n} F \to D_u F$. Since $\text{rank } D_{u_n} F$ is constant, it follows that $\text{Ker } (D_{u_n} F)^* \to \text{Ker } (D_u F)^*$, as subspaces.

\[\square\]

Consider now $A \subset T_{q_0}^* M$ a $k_O$-dimensional ball that contains $\lambda_0 = e^{-H}(\lambda_1)$ and is transversal to $\Pi_q$. By continuity $A$ is transversal also to $\Pi_{q'}$, for $q' \in \Sigma_f \cap O$ close to $q$. In particular $\Pi_{q'} \cap A \neq \emptyset$.

Since $E(\Pi_q) = q$, this implies that $\Sigma_f \cap O \subset E(A)$. By (c1), $\Sigma_f \cap O$ is a dense set, hence $E(A)$ is also dense in $O$. On the other hand, since $E$ is a smooth map and $A$ is a compact ball of positive codimension ($k_O < n$), by Sard Lemma it follows that $E(A)$ is a closed dense set of $O$ that has measure zero, that is a contradiction.

(c3) The proof of this claim relies on the following result, which is of independent interest.

**Theorem 11.19.** Let $K \subset B$ a compact in our ball such that any minimizer connecting $q_0$ to $q \in K$ is strictly normal. Then $\tilde{f}$ is Lipschitz on $K$.

230
Proof of Theorem 11.19. Let us first notice that, since \( K \) is compact, it is sufficient to show that \( f \) is locally Lipschitz on \( K \).

Fix a point \( q \in K \) and some control \( u \) associated with a minimizer joining \( q_0 \) and \( q \) (it may be not unique). By our assumptions \( D_uF \) is surjective, since \( u \) is strictly normal. Thus, by inverse function theorem, there exist neighborhoods \( V \) of \( u \) in \( U \) and \( O_q \) of \( q \) in \( K \), together with a smooth map \( \Phi : O_q \to V \) that is a local right inverse for the end-point map, namey \( F(\Phi(q')) = q' \) for all \( q' \in O_q \) (see also Theorem 2.37).

Fix then local coordinates around \( q \). Since \( \Phi \) is smooth, there exists \( R > 0 \) and \( C_0 > 0 \) such that

\[
B_q(C_0r) \subset F(B_u(r)), \quad \forall 0 \leq r < R,
\]

(11.7)

where \( B_u(r) \) is the ball of radius \( r \) in \( L^2 \) and \( B_q(r) \) is the ball of radius \( r \) in coordinates on \( M \). Let us also observe that, since \( J \) is smooth on, there exists \( C_1 > 0 \) such that for every \( u, u' \in B_u(R) \) one has

\[
J(u') - J(u) \leq C_1\|u' - u\|_2.
\]

Pick then any point \( q' \in K \) such that \( |q' - q| = C_0r \), with \( 0 \leq r < R \). By (11.7), there exists \( u' \in B_u(R) \) with \( \|u' - u\|_2 \leq r \) such that \( F(u') = q' \). Using that \( f(q') \leq J(u') \) and \( f(q) = J(u) \), since \( u \) is a minimizer, we have

\[
f(q') - f(q) \leq J(u') - J(u) \leq C_1\|u' - u\|_2 \leq C'|q' - q|,
\]

where \( C' = C_1/C_0 \). Notice that the above inequality is true for all \( q' \) such that \( |q' - q| \leq C_0R \).

Since \( K \) is compact, and the set of control \( u \) associated with minimizers that reach the compact set \( K \) is also compact, the constants \( R > 0 \) and \( C_0, C_1 \) can be chosen uniformly with respect to \( q \in K \). Hence we can exchange the role of \( q' \) and \( q \) in the above reasoning and get

\[
|f(q') - f(q)| \leq C'|q' - q|,
\]

for every pair of points \( q, q' \) such that \( |q' - q| \leq C_0R \).

To end the proof of (c3) it is sufficient to show that if \( q \in \Sigma_q \) there exists a (compact) neighborhood \( O_q \) of \( q \) such that every point in \( O_q \) is reached by only strictly normal minimizers (we stress that no uniqueness is required here). By contradiction, assume that the claim is not true. Then there exists a sequence of points \( q_n \) converging to \( q \) and a choice of controls \( u_n \), such that the corresponding minimizers are abnormal. By compactness of minimizers there exists \( u \) such that \( u_n \to u \) and by uniqueness of the limit \( u \) is abnormal for the point \( q \), that is a contradiction.

(c4). We have to prove that \( \Sigma \cap O \) is non empty for every open neighborhood \( O \) in \( B \). By (c3) we can choose \( q' \in \Sigma_q \cap O \) and fix \( O' \subset O \) neighborhood of \( q \) such that \( f \) is Lipschitz on \( O' \). It is then sufficient to show that \( \Sigma \cap O' \neq \emptyset \).

By Proposition 11.14 (see also Remark 11.12) every differentiability point of \( f \) is reached by a unique minimizer that is normal, hence is a fair point. Since we know that \( f \) is Lipschitz on \( O' \), it follows by Rademacher Theorem that almost every point of \( O' \) is fair, namely \( \text{meas}(\Sigma_f \cap O') = \text{meas}(O') \).

Let us also notice that the set \( \Sigma_f \cap O' \) of fair points of \( O' \) is also contained in the image of the exponential map. Thanks to the Sard Lemma, the set of regular values of the exponential map in
$O'$ is also a set of full measure in $O'$. Since by definition a point in $\Sigma_f$ that is a regular value for the exponential map is in $\Sigma$, this implies that $\text{meas}(\Sigma \cap O') = \text{meas}(\Sigma_f \cap O') = \text{meas}(O')$. This in particular proves that $\Sigma \cap O'$ is not empty. $\square$

As a corollary of this result we can prove that if there are no abnormal minimizers, then the set of smooth points has full measure

**Corollary 11.20.** Assume that $M$ is a complete sub-Riemannian structure and that there are no abnormal minimizers. Then $\text{meas}(M \setminus \Sigma) = 0$.

This result is not known in general, and it is indeed a main open problem of sub-Riemannian geometry to establish whether Corollary 11.20 remains true in presence of abnormal minimizers.

We stress that the assumptions of the theorem are satisfied in the case of Riemannian structure. Indeed in this case, following the same arguments of the proof, we have the following result.

**Proposition 11.21.** Let $M$ be a sub-Riemannian structure that is Riemannian at $q_0$, i.e., such that $\dim D_{q_0} = \dim M$. Then there exists a neighborhood $O_{q_0}$ of $q_0$ such that $f$ is smooth on $O_{q_0}$.

### 11.3 Locally Lipschitz functions and maps

If $S$ is a subset of a vector space $V$, we denote by $\text{conv}(S)$ the convex hull of $S$, that is the smallest convex set containing $S$. It is characterized as the set of $v \in V$ such that there exists a finite number of elements $v_0, \ldots, v_\ell \in S$ such that

$$v = \sum_{i=0}^{\ell} \lambda_i v_i, \quad \lambda_i \geq 0, \quad \sum_{i=0}^{n} \lambda_i = 1.$$

If $\varphi : M \to \mathbb{R}$ is a function defined on a smooth manifold $M$, we say that $\varphi$ is locally Lipschitz if $\varphi$ is locally Lipschitz in any coordinate chart, as a function defined on $\mathbb{R}^n$.

The classical Rademacher theorem implies that a locally Lipschitz function $\varphi : M \to \mathbb{R}$ is differentiable almost everywhere. Still we can introduce a weak notion of differential that is defined at every point.

If $\varphi : M \to \mathbb{R}$ is locally Lipschitz, any point $q \in M$ is the limit of differentiability points. In what follows, whenever we write $d_q \varphi$, it is implicitly understood that $q \in M$ is a differentiability point of $\varphi$.

**Definition 11.22.** Let $\varphi : M \to \mathbb{R}$ be a locally Lipschitz function. The (Clarke) generalized differential of $\varphi$ at the point $q \in M$ is the set

$$\partial_q \varphi := \text{conv}\{\xi \in T^*_q M | \xi = \lim_{q_n \to q} d_{q_n} \varphi\} \quad (11.8)$$

Notice that, by definition, $\partial_q \varphi$ is a subset of $T^*_q M$. It is closed by definition and bounded since the function is locally Lipschitz, hence compact.

**Exercise 11.23.** (i). Show that the mapping $q \mapsto \partial_q \varphi$ is upper semicontinuous in the following sense: if $q_n \to q$ in $M$ and $\xi_n \to \xi$ in $T^*_q M$ where $\xi_n \in \partial_{q_n} \varphi$, then $\xi \in \partial_q \varphi$.

(ii). We say that $q$ is regular for $\varphi$ if $0 \notin \partial_q \varphi$. Prove that the set of regular point for $\varphi$ is open in $M$.
From the very definition of generalized differential we have the following result.

**Lemma 11.24.** Let \( \varphi : M \to \mathbb{R} \) be a locally Lipschitz function and \( q \in M \). The following are equivalent:

1. \( \partial_q \varphi = \{ \xi \} \) is a singleton,
2. \( d_q \varphi = \xi \) and the map \( x \mapsto d_x \varphi \) is continuous at \( q \), i.e., for every sequence of differentiability point \( q_n \to q \) we have \( d_{q_n} \varphi \to d_q \varphi \).

**Remark 11.25.** Let \( A \) be a subset of \( \mathbb{R}^n \) of measure zero and consider the set of half-lines \( L_v = \{ q + tv, t \geq 0 \} \) emanating from \( q \) and parametrized by \( v \in S^{n-1} \). It follows from Fubini’s theorem that for almost every \( v \in S^{n-1} \) the one-dimensional measure of the intersection \( A \cap L_v \) is zero.

If we apply this fact to the case when \( A \) is the set at which a locally Lipschitz function \( \varphi : \mathbb{R}^n \to \mathbb{R} \) fails to be differentiable, we deduce that for almost all \( v \in S^{n-1} \), the function \( t \mapsto \varphi(q + tv) \) is differentiable for a.e. \( t \geq 0 \).

**Example 11.26.** Let \( \varphi : \mathbb{R} \to \mathbb{R} \) defined by

1. \( \varphi(x) = |x| \). Then \( \partial_0 \varphi = [-1, 1] \),
2. \( \varphi(x) = x \) if \( x < 0 \) and \( \varphi(x) = 2x \) if \( x \geq 0 \). In this case \( \partial_0 \varphi = [1, 2] \).

In particular in the first example \( 0 \) is a minimum for \( \varphi \) and \( 0 \in \partial_0 \varphi \). In the second case the function is locally invertible near the origin and \( \partial_0 \varphi \) is separated from zero. In what follows we will prove that these fact corresponds to general results (cf. Proposition 11.30 and Theorem 11.34).

The following is a classical hyperplane separation theorem for closed convex sets in \( \mathbb{R}^n \).

**Lemma 11.27.** Let \( K \) and \( C \) be two disjoint, closed, convex sets in \( \mathbb{R}^n \), and suppose that \( K \) is compact. Then there exists \( \varepsilon > 0 \) and a vector \( v \in S^{n-1} \) such that

\[
\langle x, v \rangle > \langle y, v \rangle + \varepsilon, \quad \forall x \in K, \forall y \in C.
\]

We also recall here another useful result from convex analysis.

**Lemma 11.28** (Carathéodory). Let \( S \subset \mathbb{R}^n \) and \( x \in \text{conv}(S) \). Then there exists \( x_0, \ldots, x_n \in S \) such that \( x \in \text{conv}\{x_0, \ldots, x_n\} \).

The notion of generalized gradient permits to extend some classical properties of critical points of smooth functions.

**Proposition 11.29.** Let \( \varphi : M \to \mathbb{R} \) be locally Lipschitz and \( q \) be a local minimum for \( \varphi \). Then \( 0 \in \partial_q \varphi \).

**Proof.** Since the claim is a local property we can assume without loss of generality that \( M = \mathbb{R}^n \). As usual we will identify vectors and covectors with elements of \( \mathbb{R}^n \) and the duality covectors-vectors is given by the Euclidean scalar product, that we still denote \( \langle \cdot, \cdot \rangle \).

Assume by contradiction that \( 0 \notin \partial_q \varphi \) and let us show that \( q \) cannot be a minimum for \( \varphi \). To this aim, we prove that there exists a direction \( w \) in \( S^{n-1} \) such that the scalar map \( t \mapsto \varphi(q + tw) \) has no minimum at \( t = 0 \).
The set $\partial q \varphi$ is a compact convex set that does not contain the origin, hence by Lemma 11.27, there exist $\varepsilon > 0$ and $v \in S^{n-1}$ such that

$$\langle \xi, v \rangle < -\varepsilon, \quad \forall \xi \in \partial q \varphi.$$  

By definition of generalized differential, one can find open neighborhoods $O_q$ of $q$ in $\mathbb{R}^n$ and $V_v$ of $v$ in $S^{n-1}$ such that for all differentiability point $q' \in O_q$ of $\varphi$ one has

$$\langle d_{q'} \varphi, v' \rangle \leq -\varepsilon/2, \quad \forall v' \in V_v.$$  

Fix $q' \in O_q$ where $\varphi$ is differentiable and a vector $w \in V_v$ such that the set of differentiable points with the line $\{q + tw\}$ has full measure (cf. Remark 11.25). Then we can compute for $t > 0$

$$\varphi(q + tw) - \varphi(q) = \int_0^t \langle d_{q+sw} \varphi, w \rangle \, ds \leq -\varepsilon t/2.$$  

Thus $\varphi$ cannot have a minimum at $q$. \hfill \Box

The following proposition gives an estimate for the generalized differential of some special class of function.

**Proposition 11.30.** Let $\varphi_\omega : M \to \mathbb{R}$ be a family of $C^1$ functions, with $\omega \in \Omega$ a compact set. Assume that the following maps are continuous:

$$(\omega, q) \mapsto \varphi_\omega(q), \quad (\omega, q) \mapsto d_q \varphi_\omega$$  

Then the function $a(q) := \min_{\omega \in \Omega} \varphi_\omega(q)$ is locally Lipschitz on $M$ and

$$\partial_q a \subset \text{conv}\{d_q \varphi_\omega | \forall \omega \in \Omega \text{ s.t. } \varphi_\omega(q) = a(q)\}. \quad (11.10)$$  

**Proof.** As in the proof of Proposition 11.29 we can assume that $M = \mathbb{R}^n$. Notice that, if we denote by $\Omega_q = \{\omega \in \Omega, \varphi_\omega(q) = a(q)\}$ we have by compactness of $\Omega$ that $\Omega_q$ is non empty for every $q \in M$ and we can rewrite the claim as follows

$$\partial_q a \subset \text{conv}\{d_q \varphi_\omega | \omega \in \Omega_q\}. \quad (11.11)$$  

We divide the proof into two steps. In step (i) we prove that $a$ is locally Lipschitz and then in (ii) we show the estimate (11.11).

(i). Fix a compact $K \subset M$. Since every $\varphi_\omega$ is Lipschitz on $K$ and $\Omega$ is compact, there exists a common Lipschitz constant $C_K > 0$, i.e. the following inequality holds

$$\varphi_\omega(q) - \varphi_\omega(q') \leq C_K |q - q'|, \quad \forall q, q' \in K, \quad \omega \in \Omega,$$  

Clearly we have

$$\min_{\omega \in \Omega} \varphi_\omega(q) - \varphi_\omega(q') \leq C_K |q - q'|, \quad \forall q, q' \in K, \quad \omega \in \Omega,$$  

and since the last inequality holds for all $\omega \in \Omega$ we can pass to the min with respect to $\omega$ in the left hand side and

$$a(q) - a(q') \leq C_K |q - q'|, \quad \forall q, q' \in K.$$  

234
Since the constant $C_K$ depends only on the compact set $K$ we can exchange in the previous reasoning the role of $q$ and $q'$, that gives
\[ |a(q) - a(q')| \leq C_K|q - q'|, \quad \forall q, q' \in K. \]

(ii) Define $D_q := \text{conv}\{d_q \varphi_\omega|\forall \omega \in \Omega_q\}$. Let us first prove prove that $d_q a \in D_q$ for every differentiability point $q$ of $a$.

Fix any $\xi \notin D_q$. By Lemma [11.27] applied to the pair $D_q$ and $\{\xi\}$, there exist $\varepsilon > 0$ and $v \in S^{n-1}$ such that
\[ \langle d_q \varphi_\omega, v \rangle > \langle \xi, v \rangle + \varepsilon, \quad \forall \omega \in \Omega_q, \]
By continuity of the map $(\omega, q) \mapsto d_q \varphi_\omega$, there exists a neighborhood $O_q$ of $q$ and $V$ neighborhood of $\Omega_q$ such that
\[ \langle d_q' \varphi_\omega', v \rangle > \langle \xi, v \rangle + \varepsilon/2, \quad \forall q' \in O_q, \forall \omega' \in V, \]
An integration argument let us to prove that there exists $\delta > 0$ such that for $\omega \in V$
\[ \frac{1}{t}(\varphi_\omega(q + tv) - \varphi_\omega(q)) > \langle \xi, v \rangle + \varepsilon/4, \quad \forall 0 < t < \delta. \]
Clearly we have
\[ \frac{1}{t}(\varphi_\omega(q + tv) - a(q)) \geq \langle \xi, v \rangle + \varepsilon/4, \quad \forall 0 < t < \delta. \]
and since the minimum in $a(q + tv) = \min_{\omega \in \Omega} \varphi_\omega(q + tv)$ is attained for $\omega \in \Omega_{q+tv} \subset V$ for $t$ small enough, we can pass to the minimum w.r.t. $\omega \in V$ in the left hand side, proving that there exists $t_0 > 0$ such that
\[ \frac{1}{t}(a(q + tv) - a(q)) \geq \langle \xi, v \rangle + \varepsilon/4, \quad \forall 0 < t < t_0. \]
Passing to the limit for $t \to 0$ we get
\[ \langle d_q a, v \rangle \geq \langle \xi, v \rangle + \varepsilon/4 \quad (11.12) \]
If $d_q a \notin D_q$ we can choose $\xi = d_q a$ in the above reasoning and (11.12) gives the contradiction $\langle d_q a, v \rangle \geq \langle d_q a, v \rangle + \varepsilon/4$. Hence $d_q a \in D_q$ for every differentiability point $q$ of $a$.

Now suppose that one has a sequence $q_n \to q$, where $q_n$ are differentiability points of $a$. Then $d_{q_n} a \in D_{q_n}$ for all $n$ from the first part of the proof. We want to show that, whenever the limit $\xi = \lim_{n \to \infty} d_{q_n} a$ exists, then $\xi \in D_q$. This is a consequence of the fact that the map $(\omega, q) \mapsto d_q \varphi_\omega$ is continuous (in particular upper semicontinuous in the sense of Exercise [11.23]) and the fact that $\Omega$ is compact. \hfill \Box

**Exercise 11.31.** Complete the second part of the proof of Proposition [11.30]. Hint: use Carathéodory lemma.

### 11.3.1 Locally Lipschitz map and Lipschitz submanifolds

As for scalar functions, a map $f : M \to N$ between smooth manifolds is said to be locally Lipschitz if for any coordinate chart in $M$ and $N$ the corresponding function from $\mathbb{R}^n$ to $\mathbb{R}^n$ is locally Lipschitz.

For a locally Lipschitz map between manifolds $f : M \to N$ the (Clarke) generalized differential is defined as follows
\[ \partial_q f := \text{conv}\{L \in \text{Hom}(T_q M, T_f(q) N)| L = \lim_{q_n \to q} D_{q_n} f, \text{ } q_n \text{ diff. point of } f\}, \]
The following lemma shows how the standard chain rule extends to the Lipschitz case.

235
Lemma 11.32. Let $M$ be a smooth manifold and $f : M \to N$ be a locally Lipschitz map.

(a) If $\phi : M \to M$ is a diffeomorphism and $q \in M$ we have

$$\partial_q(f \circ \phi) = \partial_{\phi(q)} f \cdot D_q \phi.$$  \hfill (11.13)

(b) If $\varphi : N \to W$ is a $C^1$ map, and $q \in M$ we have

$$\partial_q(\varphi \circ f) = D_{f(q)} \varphi \cdot \partial_q f.$$  \hfill (11.14)

Moreover the generalized differential, as a set, is upper semicontinuous. More precisely for every neighborhood $\Omega \in \text{Hom}(T_q M, T_{f(q)} N)$ of $\partial_q f$ there exists a neighborhood $O_q$ of $q$ such that $\partial_q' f \in \Omega$, for every $q' \in O_q$.

Sketch of the proof. For a detailed proof of this result see ?? Here we only give the main ideas.

(a). Since $\phi$ is a diffeomorphism, it sends every differentiability point $q$ of $f \circ \phi$ to a differentiability point $\phi(q)$ for $f$. Then (11.13) is true at differentiability point and passing to the limit it is also valid for sub-differential (one proves both inclusions using $\phi$ and $\phi^{-1}$). Part (b) can be proved along the same lines. The semicontinuity can be proved by using the hyperplane separation theorem and the Carathéodory Lemma. \hfill $\square$

Definition 11.33. Let $f : M \to N$ be a locally Lipschitz map. A point $q \in M$ is said critical for $f$ if $\partial_q f$ contains a non-surjective map. If $q \in M$ is not critical it is said regular.

Notice that by the semicontinuity property of Lemma 11.32 it follows that the set of regular point of a locally Lipschitz map $f$ is open.

Theorem 11.34. Let $f : \mathbb{R}^n \to \mathbb{R}^n$ be a locally Lipschitz map and $q \in M$ be a regular point. Then there exists neighborhood $O_{f(q)}$ and a locally Lipschitz map $g : O_{f(q)} \subset \mathbb{R}^n \to \mathbb{R}^n$ such that $f \circ g = g \circ f = \text{Id}$.

Remark 11.35. The classical $C^1$ version of the inverse function theorem (cf. Theorem ??) can be proved from Theorem 11.34 and the chain rule (Lemma 11.32). Indeed Theorem 11.34 implies that there exists a locally Lipschitz inverse $g$ and using the chain rule it is easy to show that the sub-differential of $g$ contains only one element (this implies that it is differentiable at that point) and the differential of $g$ is the inverse of the differential of $f$.

Before proving Theorem 11.34 we need the following technical lemma.

Lemma 11.36. Let $f : \mathbb{R}^n \to \mathbb{R}^n$ be a locally Lipschitz map and $q \in M$ be a regular point. Then there exists a neighborhood $O_q$ of $q$ and $\varepsilon > 0$ such that

$$\forall v \in S^{n-1}, \exists \xi_v \in S^{n-1} \quad \text{s.t.} \quad (\xi_v, \partial_x f(v)) > \varepsilon, \quad \forall x \in O_q.$$  \hfill (11.15)

Moreover $|f(x) - f(y)| \geq \varepsilon |x - y|$, for all $x, y \in O_q$.

We stress that (11.15) means that the inequality $(\xi_v, L(v)) > \varepsilon$ holds for every $x \in O_q$ and every element $L \in \partial_x f$.  

236
Proof. Notice that, since \( q \) is a regular point, the set \( \partial_q f \) contains only invertible linear maps. For every \( v \in S^{n-1} \), the set \( \partial_q f(v) \) is compact and convex, and does not contain the zero linear map. By the hyperplane separation theorem we can find \( \xi_v \) such that \( \langle \xi_v, \partial_q f(v) \rangle > \varepsilon(v) \). The map \( x \mapsto \partial_x f \) is upper semicontinuous, hence there exists a neighborhood \( O_q \) of \( q \) such that \( \langle \xi_v, \partial_x f(v) \rangle > \varepsilon(v) \) for all \( x \in O_q \). Since \( S^{n-1} \) is compact, there exists a uniform \( \varepsilon = \min \{ \varepsilon(v), v \in S^{n-1} \} \) that satisfies (11.15).

To prove the second statement of the Lemma, write \( y = x + sv \), where \( s = |x - y| \) and \( v \in S^{n-1} \). Consider a vector \( v' \in S^{n-1} \) close to \( v \) such that almost every point in the direction of \( v' \) is a point of differentiability (cf. Remark 11.25), and set \( y' = x + sv' \) and \( \xi_{v'} \) the vector associated to \( v' \) defined by (11.15). Then we can write

\[
f(y') - f(x) = \int_0^s (D_{x + t v'} f)v' \, dt.\]

and we have the inequality

\[
|f(y') - f(x)| \geq \langle \xi_{v'}, f(y') - f(x) \rangle \\
= \int_0^s \langle \xi_{v'}, (D_{x + t v'} f)v' \rangle \, dt \\
\geq \varepsilon |y' - x|
\]

Since \( \varepsilon \) does not depend on \( v \), we can pass to the limit for \( v' \to v \) in the above inequality (in particular \( y' \to y \)) and the Lemma is proved.

Proof of Theorem 11.34. The inequality proved in Lemma 11.36 implies that \( f \) is injective in the neighborhood \( O_q \) of the point \( q \). If we show that \( f(O_q) \) covers a neighborhood \( O_{f(q)} \) of the point \( f(q) \), then the inverse function \( g : O_{f(q)} \to \mathbb{R}^n \) is well defined and locally Lipschitz.

Without loss of generality, up to restricting the neighborhood \( O_q \), we can assume that every point in \( O_q \) is regular for \( f \) and moreover that the estimate of the Lemma 11.36 holds also on the topological boundary \( \partial O_q \). Lemma 11.36 also implies that

\[
\text{dist}(f(q), \partial f(O_q)) \geq \varepsilon \text{dist}(q, \partial O_q) > 0,
\]

where \( \text{dist}(x, A) = \inf_{y \in A} |x - y| \) denotes the Euclidean distance from \( x \) to the set \( A \). Then consider a neighborhood \( W \subset f(O_q) \) of \( f(q) \) such that \( |y - f(q)| < \text{dist}(y, \partial f(O_q)) \), for every \( y \in W \). Fix an arbitrary \( \bar{y} \in W \) and let us show that the equation \( f(x) = \bar{y} \) has a solution. Define the function

\[
\psi : \overline{O_q} \to \mathbb{R}, \quad \psi(x) = |f(x) - \bar{y}|^2.
\]

By construction \( \psi(q) < \psi(z) \), for all \( z \in \partial O_q \), hence by continuity \( \psi \) attains the minimum on some point \( \bar{x} \in O_q \). By Proposition 11.29 we have \( 0 \in \partial_{\bar{x}} \psi \). Moreover, using the chain rule

\[
\partial_{\bar{x}} \psi = (f(\bar{x}) - \bar{y})^T \cdot \partial_{\bar{x}} f
\]

Since \( \bar{x} \) is a regular point of \( f \), the linear map \( \partial_{\bar{x}} f \) is invertible. Thus \( 0 \in \partial_{\bar{x}} \psi \) implies \( f(\bar{x}) = \bar{y} \).

We say that \( c \in \mathbb{R} \) is a regular value of a locally Lipschitz function \( \varphi : M \to \mathbb{R} \) if \( \varphi^{-1}(c) \neq \emptyset \) and every \( x \in \varphi^{-1}(c) \) is a regular point.

237
Corollary 11.37. Let \( \varphi : M \to \mathbb{R} \) be locally Lipschitz and assume that \( c \in \mathbb{R} \) is a regular value for \( \varphi \). Then \( \varphi^{-1}(c) \) is a Lipschitz submanifold of \( M \) of codimension 1.

Proof. We show that in any small neighborhood \( O_x \) of every \( x \in \varphi^{-1}(c) \) the set \( O_x \cap \varphi^{-1}(c) \) can be described as the zero locus of a locally Lipschitz function. Since \( \partial_x \varphi \) does not contain 0, by the hyperplane separation theorem there exists \( v_1 \in S^{n-1} \), such that \( \langle \partial_x \varphi, v_1 \rangle > 0 \) for every \( x \) in the compact neighborhood \( O_x \cap \varphi^{-1}(y) \).

Let us complete \( v_1 \) to an orthonormal basis \( \{v_1, v_2, \ldots, v_n\} \) of \( \mathbb{R}^n \) and consider the map

\[
\begin{align*}
f : O_x & \to \mathbb{R}^n, \\
f(x') & = \begin{pmatrix}
\varphi(x') - c \\
\langle v_2, x' \rangle \\
\vdots \\
\langle v_n, x' \rangle
\end{pmatrix}
\end{align*}
\]

By construction \( f \) is locally Lipschitz and \( x \) is a regular point of \( f \). Hence there exists, by Theorem 11.34 a Lipschitz inverse \( g \) of \( f \). In particular the inverse map is a Lipschitz function that transforms the hyperplane \( \{y_1 = 0\} \) into \( \varphi^{-1}(c) \). Hence the level set \( \varphi^{-1}(c) \) is a Lipschitz submanifold. \( \square \)

11.3.2 A non-smooth version of Sard Lemma

In this section we prove a Sard-type result for the special class of Lipschitz functions we considered in the previous section.

We first recall the statement of the classical Sard lemma. We denote by \( C_f \) the critical point of a smooth map \( f : M \to N \), i.e. the set of points \( x \) in \( M \) at which the differential of \( f \) is not surjective.

Theorem 11.38 (Sard lemma). Let \( f : \mathbb{R}^n \to \mathbb{R}^m \) be a \( C^k \) function, with \( k \geq \max\{n - m + 1, 1\} \). Then the set \( f(C_f) \) of critical values of \( f \) has measure zero in \( \mathbb{R}^m \).

Notice that the classical Sard Lemma does not apply to \( C^1 \) functions \( \varphi : \mathbb{R}^n \to \mathbb{R} \), whenever \( n \geq 1 \).

Theorem 11.39. Let \( M \) be a smooth manifold and \( \varphi_\omega : M \to \mathbb{R} \) a family of smooth functions, with \( \omega \in \Omega \). Assume that

(i) \( \Omega = \bigcup_{i \in \mathbb{N}} N_i \) is the union of smooth submanifold, and is compact,

(ii) the maps \( (\omega, q) \mapsto \varphi_\omega(q) \) and \( (\omega, q) \mapsto d_q \varphi_\omega \) are continuous on \( \Omega \times M \),

(iii) the maps \( \psi_i : N_i \times M \to \mathbb{R}, (\omega, q) \mapsto \varphi_\omega(q) \) are smooth.

Then the set of critical values of the function \( a(q) = \min_{\omega \in \Omega} \varphi_\omega(q) \) has measure zero in \( \mathbb{R} \).

Proof. We are going to define a countable set of smooth functions \( \Phi_\alpha \) indexed by \( \alpha = (\alpha_0, \ldots, \alpha_n) \in \mathbb{N}^{n+1} \), where \( n = \dim M \), such that to every critical point \( q \) of \( a \) there corresponds a critical point \( z_q \) of some \( \Phi_\alpha \). Moreover we have \( \Phi_\alpha(z_q) = a(q) \).
Denote by $\Lambda_n = \{ (\lambda_0, \ldots, \lambda_n) | \lambda_i \geq 0, \sum \lambda_i = 1 \}$. For every $\alpha = (\alpha_0, \ldots, \alpha_n) \in \mathbb{N}^{n+1}$ let us consider the map

$$\Phi_{\alpha} : N_{\alpha_0} \times \ldots \times N_{\alpha_n} \times \Lambda_n \times M \to \mathbb{R}$$

$$\Phi_{\alpha}(\omega_0, \ldots, \omega_n, \lambda_0, \ldots, \lambda_n, q) = \sum_{i=0}^{n} \lambda_i \varphi_{\omega_i}(q). \quad (11.16)$$

By computing partial derivatives, it is easy to see that a point $z = (\omega_0, \ldots, \omega_n, \lambda_0, \ldots, \lambda_n, q)$ is critical for $\Phi_{\alpha}$ if and only if it satisfies the following relations:

$$\begin{cases}
\sum_{i=0}^{n} \lambda_i \frac{\partial \psi_{\alpha_i}}{\partial \omega}(\omega_i, q) = 0, & i = 0, \ldots, n, \\
\sum_{i=0}^{n} \lambda_i d_q \varphi_{\omega_i} = 0 & i = 0, \ldots, n, \\
\varphi_{\omega_0}(q) = \ldots = \varphi_{\omega_n}(q)
\end{cases} \quad (11.17)$$

Recall that $\psi_i$ is simply the restriction of the map $(\omega, q) \mapsto \varphi_{\omega}(q)$ for $\omega \in N_i$.

Let us now show that every critical point $q$ of $a$ can be associated to a critical point $z_q$ of some $\Phi_{\alpha}$. By Proposition 11.30, the function $a$ is locally Lipschitz. Assume that $q$ is a critical point of $a$, then we have

$$0 \in \partial_q a \subset \text{conv} \{ d_q \varphi_{\omega} | \forall \omega \in \Omega \text{ s.t. } \varphi_{\omega}(q) = a(q) \}.$$ 

By Carathéodory lemma there exist $n + 1$ element $\tilde{\omega}_0, \ldots, \tilde{\omega}_n$ and $n + 1$ scalars $\tilde{\lambda}_0, \ldots, \tilde{\lambda}_n$ such that $\tilde{\lambda}_i \geq 0, \sum_{i=0}^{n} \tilde{\lambda}_i = 1$ and

$$0 = \sum_{i=0}^{n} \tilde{\lambda}_i d_q \varphi_{\tilde{\omega}_i}, \quad \varphi_{\tilde{\omega}_i}(q) = a(q), \quad \forall i = 0, \ldots, n.$$ 

Moreover, let us choose for every $i = 0, \ldots, n$ an index $\tilde{\alpha}_i \in \mathbb{N}$ such that $\tilde{\omega}_i \in N_{\tilde{\alpha}_i}$. Since $\varphi_{\tilde{\omega}_i}(q) = a(q) = \min_{\Omega} \varphi_{\omega}(q)$, $\tilde{\omega}_i$ is critical for the map $\psi_{\alpha_i}$, namely we have

$$\frac{\partial \psi_{\alpha_i}}{\partial \omega}(\tilde{\omega}_i, q) = 0.$$ 

This implies that $z_q = (\tilde{\omega}_0, \ldots, \tilde{\omega}_n, \tilde{\lambda}_0, \ldots, \tilde{\lambda}_n, q)$ satisfies the relations (11.17) for the function $\Phi_{\alpha}$, with $\tilde{\alpha} = (\tilde{\alpha}_0, \ldots, \tilde{\alpha}_n)$. Moreover it is easy to check that $\Phi_{\tilde{\alpha}}(z_q) = a(q)$ since

$$\Phi_{\tilde{\alpha}}(z_q) = \sum_{i=0}^{n} \tilde{\lambda}_i \varphi_{\tilde{\omega}_i}(q) = \left( \sum_{i=0}^{n} \tilde{\lambda}_i \right) a(q) = a(q).$$

Then if $C_{\alpha}$ denotes the set of critical points of $a$ and $C_{\alpha}$ the set of critical point of $\Phi_{\alpha}$ we have

$$\text{meas}(a(C_{\alpha})) \leq \text{meas} \left( \bigcup_{\alpha \in \mathbb{N}^{n+1}} \Phi_{\alpha}(C_{\alpha}) \right) \leq \sum_{\alpha \in \mathbb{N}^{n+1}} \text{meas}(\Phi_{\alpha}(C_{\alpha})) = 0,$$

since $\text{meas}(\Phi_{\alpha}(C_{\alpha})) = 0$ for all $\alpha$ by classical Sard lemma. \qed
We want to apply the previous result in the case of functions that are infimum of smooth functions on level sets of a submersion.

**Theorem 11.40.** Let \( F : N \rightarrow M \) be a smooth map between finite dimensional manifolds and \( \varphi : N \rightarrow \mathbb{R} \) be a smooth function. Assume that

(i) \( F \) is a submersion

(ii) for all \( q \in M \) the set \( N_q = \{ x \in N, \varphi(x) = \min_{y \in F^{-1}(q)} \varphi(y) \} \) is a non empty compact set.

Then the set of critical values of the function \( a(q) = \min_{x \in F^{-1}(q)} \varphi(x) \) has measure zero in \( \mathbb{R} \).

**Proof.** Denote by \( C_a \) the set of critical points of \( a \) and \( a(C_a) \) is the set of its critical values. Let us first show that for every point \( q \in M \) there exist an open neighborhood \( O_q \) of \( q \) such that \( \text{meas}(a(C_a) \cap O_q) = 0 \).

From assumption (i), it follows that for every \( q \in M \) the set \( F^{-1}(q) \) is a smooth submanifold in \( N \). Let us now consider an auxiliary non-negative function \( \psi : N \rightarrow \mathbb{R} \) such that

(A0) \( A_\alpha := \psi^{-1}([0, \alpha]) \) is compact for every \( \alpha > 0 \).

and select moreover a constant \( c > 0 \) such that the following assumptions are satisfied:

(A1) \( N_q \subset \text{int} A_c \),

(A2) \( c \) is a regular level of \( \psi|_{F^{-1}(q)} \).

The existence of such a \( c > 0 \) is guaranteed by the fact that (A1) is satisfied for all \( c \) big enough since \( N_q \) is compact and \( A_c \) contains any compact as \( c \to +\infty \). Moreover, by classical Sard lemma (cf. Theorem 11.38), almost every \( c \) is a regular value for the smooth function \( \psi|_{F^{-1}(q)} \).

By continuity, there exists a neighborhood \( O_q \) of the point \( q \) such that assumptions (A0)-(A2) are satisfied for every \( q' \in O_q \), for \( c > 0 \) and \( \psi \) fixed. We observe that (A2) is equivalent to require that level set of \( F \) are transversal to level of \( \psi \). We can infer that \( F^{-1}(O_q) \cap A_c \) is a smooth manifold with boundary that has the structure of locally trivial bundle. Maybe restricting the neighborhood of \( q \) then we can assume

\[
F^{-1}(q) \cap A_c = \Omega, \quad F^{-1}(O_q) \cap A_c \simeq O_q \times \Omega,
\]

where \( \Omega \) is a smooth manifold with boundary. In this neighborhood we can split variables in \( N \) as follows \( x = (\omega, q) \) with \( \omega \in \Omega \) and \( q \in M \) and the restriction \( a|_{O_q} \) is written as

\[
a|_{O_q} : O_q \rightarrow \mathbb{R}, \quad a(q) = \min_{\omega \in \Omega} \varphi(\omega, q).
\]

Notice that \( \Omega \) is compact and is the union of its interior and its boundary, which are smooth by assumptions (A0)-(A2). We can then apply the Theorem 11.39 to \( a|_{O_q} \), that gives \( \text{meas}(a(C_a \cap O_q) = 0 \) for every \( q \in M \).

We have built a covering of \( M = \bigcup_{q \in M} O_q \). Since \( M \) is a smooth manifold, from every covering it is possible to extract a countable covering, i.e. there exists a sequence \( q_n \) of points in \( M \) such that

\[
M = \bigcup_{n \in \mathbb{N}} O_{q_n}
\]

[240]
In particular this implies that
\[ \text{meas}(a(C_a)) \leq \sum_{n \in \mathbb{N}} \text{meas}(a(C_a) \cap O_{q_n}) = 0 \]
since \( \text{meas}(a(C_a) \cap O_q) = 0 \) for every \( q \).

\[ \square \]

Remark 11.41. Notice that we do not assume that \( N \) is compact. In that case the proof is easier since every submersion \( F : N \to M \) with \( N \) compact automatically endows \( N \) with a locally trivial bundle structure.

We end this section by applying the previous theory to get information about the regularity of sub-Riemannian spheres. Before proving the main result we need two lemmas.

**Lemma 11.42.** Fix \( q_0 \in M \) and let \( K \subset T^*_q M \setminus (H^{-1}(0) \cap T^*_q M) \) be a compact set such that all normal extremals associated with \( \lambda_0 \in K \) are not abnormal. Then there exists \( \varepsilon = \varepsilon(K) \) such that \( t\lambda_0 \) is a regular point for the \( \mathcal{E}_{q_0} \) for all \( 0 < t \leq \varepsilon \).

**Proof.** By Corollary 8.46 for every strongly normal extremal \( \gamma(t) = \mathcal{E}(t\lambda_0) \), with \( \lambda_0 \in T^*_q M \), there exists \( \varepsilon = \varepsilon(\lambda_0) > 0 \) such that \( \gamma|_{[0,\varepsilon]} \) does not contain points conjugate to \( q_0 \), or equivalently, \( t\lambda_0 \) is a regular point for the \( \mathcal{E}_{q_0} \) for all \( 0 < t \leq \varepsilon \). Since \( K \) is compact, it follows that there exists \( \varepsilon = \varepsilon(K) \) such that the above property holds uniformly on \( K \).

**Lemma 11.43.** Let \( q_0 \in M \) and \( K \subset M \) be a compact set such that every point of \( K \) is reached from \( q_0 \) by only strictly normal minimizers. Define the set
\[ C = \{ \lambda_0 \in T^*_q M \mid \lambda_0 \text{ minimizer, } \mathcal{E}(\lambda_0) \in K \}. \]

Then \( \overline{C} \) is compact.

**Proof.** It is enough to show that \( C \) is bounded. Assume by contradiction that there exists a sequence \( \lambda_n \in C \) of covectors (and the associate sequence of minimizing trajectories \( \gamma_n \), associated with controls \( u_n \)) such that \( |\lambda_n| \to +\infty \), where \( |\cdot| \) is some norm in \( T^*_q M \). Since these minimizers are normal they satisfy the relation
\[ \lambda_n D_{u_n} F = u_n, \quad \forall n \in \mathbb{N}. \tag{11.18} \]
and dividing by \( |\lambda_n| \) one obtain the identity
\[ \frac{\lambda_n}{|\lambda_n|} D_{u_n} F = \frac{u_n}{|\lambda_n|}, \quad \forall n \in \mathbb{N}. \tag{11.19} \]
Using compactness of minimizers whose endpoints stay in a compact region, we can assume that \( u_n \to u \). Moreover the sequence \( \lambda_n / |\lambda_n| \) is bounded and we can assume that \( \lambda_n / |\lambda_n| \to \lambda \) for some final covector \( \lambda \). Using that \( D_{u_n} F \to D_{\lambda} F \) and the fact that \( |\lambda_n| \to +\infty \), passing to the limit for \( n \to \infty \) in (11.19) we obtain \( \lambda D_{\lambda} F = 0 \). This implies in particular that the minimizers \( \gamma_n \) converge to a minimizer \( \gamma \) (associated to \( \lambda \)) that is abnormal and reaches a point of \( K \) that is a contradiction. \[ \square \]
Theorem 11.44. Let $M$ be a sub-Riemannian manifold, $q_0 \in M$ and $r_0 > 0$ such that every point different from $q_0$ in the compact ball $B_{r_0}(q_0)$ is not reached by abnormal minimizers. Then the sphere $S_{q_0}(r)$ is a Lipschitz submanifold of $M$ for almost every $r \leq r_0$.

Proof. Let us fix $\delta > 0$ and consider the annulus $A_\delta = B_{r_0}(q_0) \setminus B_{\delta}(q_0)$. Define the set

$$C = \{ \lambda_0 \in T^*_q M \mid \lambda_0 \text{ minimizer}, \mathcal{E}(\lambda_0) \in \overline{A}_\delta \}$$

By Lemma 11.43 the set $C_0 := \overline{C}$ is compact. Moreover define

$$C_1 := \{ \lambda_0 \in C_0 \cap H^{-1}([0, \varepsilon_0]) \},$$

for some $\varepsilon_0 > 0$ that is chosen later. Notice that $C_1$ is compact. For every $\lambda_0 \in T^* M$ let us consider the control $u$ associated with $\gamma(t) = \mathcal{E}(t\lambda_0)$ and denote by

$$\Phi_{\lambda_0} := (P_{0,t})_*: T_{q_0} M \to T^*_{\mathcal{E}_{q_0}(\lambda_0)} M,$$

the pullback of the flow defined by the control $u$, computed at $q_0$. For a fixed $\lambda_0 \in C_0$, using that $C_1$ is compact, let us choose $\varepsilon = \varepsilon(\lambda_0)$ satisfying the following property: for every $\lambda_1 \in C_1$, the covector $\Phi_{\lambda_0}(\lambda_1) \in T^*_q M$, is a regular point of $\mathcal{E}_{q_0}(\lambda_0)$. Being $C_0$ also compact, we can define $\varepsilon_0 = \min\{\varepsilon(\lambda_0), \lambda_0 \in C_0\}$. Define the map

$$\Psi : C_0 \times C_1 \to D_\delta \subset M, \quad \Psi(\lambda_0, \lambda_1) = \mathcal{E}_{q_0}(\lambda_0)(\Phi_{\lambda_0}(\lambda_1)).$$

By construction $\Psi$ is a submersion. We want to apply Theorem 11.40 to the submersion $\Psi$ and the scalar function

$$\mathcal{H} : C_0 \times C_1 \to \mathbb{R}, \quad \mathcal{H}(\lambda_0, \lambda_1) = H(\lambda_0) + H(\lambda_1).$$

Let us show that the assumption of Theorem 11.40 are satisfied. Indeed we have to show that the set

$$N_q = \{ (\lambda_0, \lambda_1) \in C_0 \times C_1 \mid \mathcal{H}(\lambda_0, \lambda_1) = \min_{\Psi(\lambda_0, \lambda_1) = q} \mathcal{H}(\lambda_0, \lambda_1) \}, \quad \forall q \in \overline{A}_\delta,$$

is non empty and compact. Let us first notice that

$$\Psi(\lambda_0, s\lambda_0) = \mathcal{E}_{q_0}((1 + s)\lambda_0), \quad \mathcal{H}(\lambda_0, s\lambda_0) = (1 + s^2)H(\lambda_0).$$

By definition of $C_0$, for each $q \in \overline{A}_\delta$ there exists $\lambda_0 \in C_0$ such that $\mathcal{E}_{q_0}(\lambda_0) = q$ and such that the corresponding trajectory is a minimizer. Moreover we can always write this unique minimizer as the union of two minimizers. It follows that

$$\min_{\Psi(\lambda_0, \lambda_1) = q} \mathcal{H}(\lambda_0, \lambda_1) = \min_{\mathcal{E}_{q_0}(\lambda_0) = q} H(\lambda_0) = f(q), \quad \forall q \in \overline{A}_\delta.$$

This implies that $N_q$ is non empty for every $q$. Moreover one can show that $N_q$ is compact. By applying Theorem 11.40 one gets that the function

$$a(q) = \min_{\Psi(\lambda_0, \lambda_1) = q} \mathcal{H}(\lambda_0, \lambda_1) = f(q),$$

is locally Lipschitz in $\overline{A}_\delta$ and the set of its critical values has measure zero in $\overline{A}_\delta$. Since $\delta > 0$ is arbitrary we let $\delta \to 0$ and we have that $f$ is locally Lipschitz in $B_{r_0}(r_0) \setminus \{ q_0 \}$ and the set of its critical values has measure zero. In particular almost every $r \leq r_0$ is a regular value for $f$. Then, applying Corollary 11.37, the sphere $f^{-1}(r^2/2)$ is a Lipschitz submanifold for almost every $r \leq r_0$. \qed
11.4 Geodesic completeness and Hopf-Rinow theorem

We start by proving a technical lemma that is needed later.

Lemma 11.45. For every $\varepsilon > 0$ and $x \in M$ we have

$$B(x, r + \varepsilon) = \bigcup_{y \in B(x, r)} B(y, \varepsilon).$$

(11.20)

Proof. The inclusion $\supseteq$ is a direct consequence of the triangle inequality. Let us prove the converse inclusion $\subseteq$.

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

Theorem 11.46. Let $M$ be a sub-Riemannian manifold. Then $(M, d)$ is complete if and only if all sub-Riemannian closed balls are compact.

Proof. (i). Assume that all closed balls are compact and let $\{x_j\}$ be a Cauchy sequence in $M$. We have to prove that $\{x_j\}$ admits a convergent subsequence. By assumption, if we fix $\varepsilon > 0$ there exists $n_0 \in \mathbb{N}$ such that $d(x_j, x_k) < \varepsilon$ for all $j, k \geq n_0$. Let us define $R := \max_{j \leq n_0} d(x_j, x_{n_0}) + \varepsilon > 0$. By construction $x_j \in B(x_{n_0}, R)$ for every $j$, and $B(x_{n_0}, R)$ has compact closure by assumption. Hence the sequence admits a converging subsequence.

(ii). Assume now that $(M, d)$ is complete. Fix $x \in M$ and define

$$A := \{r > 0 \mid \overline{B}(x, r) \text{ is compact} \}, \quad R := \sup A.$$ (11.21)

Since the topology of $(M, d)$ is locally compact then $A \neq \emptyset$ and $R > 0$. First we prove that $A$ is open and then we prove that $R = +\infty$. Notice in particular that this proves that $A = \{0, +\infty\}$ since, by Remark 3.41, $r \in A$ implies $[0, r] \subseteq A$.

(ii.a) It is enough to show that, if $r \in A$, then there exists $\delta > 0$ such that $r + \delta \in A$. For each $y \in B(x, r)$ there exists $r(y) < \varepsilon$ small enough such that $\overline{B}(y, r(y))$ is compact. We have

$$\overline{B}(x, r) \subseteq \bigcup_{y \in B(x, r)} \overline{B}(y, r(y)).$$

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

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

Moreover, there exists $\delta > 0$ such that the set of points $\overline{B}(x, r + \delta) = \{y \in M \mid \text{dist}(y, B(x, r)) \leq \delta\}$, where the equality is given by Lemma 11.45 satisfies

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

This proves that $r + \delta \in A$, since a finite union of compact sets is compact.
(ii.b) Assume by contradiction that $R < +\infty$ and let us prove that $B := \overline{B}(x, R)$ is compact. Since $B$ is a closed set, it is enough to show that it is totally bounded, i.e., it admits an $\varepsilon$-net\textsuperscript{2} for every $\varepsilon > 0$. Fix $\varepsilon > 0$ and consider an $(\varepsilon/3)$-net $S$ for the ball $B' = B(x, R - \varepsilon/3)$, that exists by compactness. By Lemma 11.45 one has for every $y \in B$ that $\text{dist}(y, B') < \varepsilon/3$. Then it is easy to show that
\[
\text{dist}(y, S) < \text{dist}(y, B') + \varepsilon/3 < \varepsilon,
\]
that is $S$ is an $\varepsilon$-net for $B$ and $B$ is compact.

This shows that if $R < +\infty$, then $R \in A$. Hence (ii.a) implies that $R + \delta \in A$ for some $\delta > 0$, contradicting the fact that $R$ is a sup. Hence $R = +\infty$. \hfill \Box

The next result implies that the geodesic completeness of $M$, i.e., the completeness of $\tilde{H}$, implies the completeness of $M$ as a metric space.

**Theorem 11.47** (sub-Riemannian Hopf-Rinow). Let $M$ be a sub-Riemannian manifold that does not admit abnormal length minimizers. If there exists a point $x \in M$ such that the exponential map $E_x$ is defined on the whole $T_x M$, then $M$ is complete with respect to the sub-Riemannian distance.

**Proof.** For the fixed $x \in M$, let us consider
\[
A = \{r > 0 \mid B(x, r) \text{ is compact}\}, \quad R := \sup A.
\]
As in the proof of Theorem 11.46 one can show that $A \neq \emptyset$ and that $A$ is open (by using the local compactness of the topology and repeating the proof of (ii.a)). Assume now that $R < +\infty$ and let us show that $R \in A$. By openness of $A$ this will give a contradiction and $A = [0, +\infty[$.

We have to show that $B(x, R)$ is compact, i.e., for every sequence $y_i \in \overline{B}(x, R)$ we can extract a convergent subsequence. Define $r_i := d(y_i, x)$. It is not restrictive to assume that $r_i \to R$ (if it is not the case, the sequence stays in a compact ball and the existence of a convergent subsequence is clear). Since the ball $B(x, r_i)$ is compact, by Theorem 3.40 there exists a length minimizing trajectory $\gamma_i : [0, r_i] \to M$ joining $x$ and $y_i$, parametrized by unit speed.

Due to the completeness of the vector field $\tilde{H}$, we can extend each curve $\gamma_i$, parametrized by length, to the common interval $[0, R]$. By construction this sequence of trajectory is normal
\[
\gamma_i(t) = E(t \lambda_i) = \pi \circ e^{\tilde{H}}(\lambda_i),
\]
for some $\lambda_i \in T_x M$, and is contained in the compact set $\overline{B}(x, R)$. Since there is no abnormal minimizer, by Lemma 11.43 the sequence $\{\lambda_i\}$ is bounded in $T_x^* M$, thus there exists a subsequence $\lambda_{i_n}$ converging to $\lambda$. Then $r_{i_n} \lambda_{i_n} \to R \lambda$ and by continuity of $E$ we have that $\{y_i\}$ has a convergent subsequence
\[
y_{i_n} = \gamma_{i_n}(r_{i_n}) = E(r_{i_n} \lambda_{i_n}) \to E(R \lambda) =: y
\]
To end the proof, one should just notice that an arbitrary Cauchy sequence in $M$ is bounded, hence contained in a suitable ball centered at $x$, which is compact since $R = +\infty$. Thus it admits a convergent subsequence. \hfill \Box

As an immediate corollary we have the following version of geodesic completeness theorem.

**Corollary 11.48.** Let $M$ be a sub-Riemannian manifold that does not admit abnormal length minimizers. If the vector field $\tilde{H}$ is complete on $T^* M$, then $M$ is complete with respect to the sub-Riemannian distance.

\textsuperscript{2}An $\varepsilon$-net $S$ for a set $B$ in a metric space is a finite set of points $S = \{z_i\}_{i=1}^N$ such that for every $y \in B$ one has $\text{dist}(y, S) < \varepsilon$ (or, equivalently, for every $y \in B$ there exists $i$ such that $d(y, z_i) < \varepsilon$).
Chapter 12

Abnormal extremals and second variation

In this chapter we are going to discuss in more details abnormal extremals and how the regularity of the sub-Riemannian distance is affected by the presence of these extremals.

12.1 Second variation

We want to introduce the notion of Hessian (and second derivative) for smooth maps between manifolds. We first discuss the case of the second differential of a map between linear spaces.

Let $F : V \to M$ be a smooth map from a linear space $V$ on a smooth manifold $M$. As we know, the first differential of $F$ at a point $x \in V$

$$D_x F : V \to T_{F(x)} M, \quad D_x F(v) = \left. \frac{d}{dt} \right|_{t=0} F(x + tv), \quad v \in V,$$

and is a well defined linear map independent on the linear structure on $V$. This is not the case for the second differential. Indeed it is easy to see that the second order derivative

$$D_x^2 F(v) = \left. \frac{d^2}{dt^2} \right|_{t=0} F(x + tv)$$

has not invariant meaning if $D_x F(v) \neq 0$. Indeed in this case the curve $\gamma : t \mapsto F(x + tv)$ is a smooth curve in $M$ with nonzero tangent vector. Then there exists some local coordinates on $M$ such that the curve $\gamma$ is a straight line. Hence the second derivative $D_x^2 F(v)$ vanish in these coordinates.

In general, the linear structure on $V$ let us to define the second differential of $F$ as a quadratic map

$$D_x^2 F : \text{Ker} \, D_x F \to T_{F(x)} M$$

On the other hand the map (12.2) is not independent on the choice of the linear structure on $V$ and this construction cannot be used if the source of $F$ is a smooth manifold.

Assume now that $F : N \to M$ is a map between smooth manifolds. The first differential is a linear map between the tangent spaces

$$D_x F : T_x N \to T_{F(x)} M, \quad x \in N.$$
and the definition of second order derivative should be modified using smooth curves with fixed tangent vector (that belong to the kernel of $D_x F$):

$$D_x^2 F(v) = \frac{d^2}{dt^2} \bigg|_{t=0} F(\gamma(t)), \quad \gamma(0) = x, \quad \dot{\gamma}(0) = v \in \text{Ker} \, D_x F; \quad (12.3)$$

Computing in coordinates we find that

$$\frac{d^2}{dt^2} \bigg|_{t=0} F(\gamma(t)) = \frac{d^2}{dt^2} \frac{dF}{dx}(\dot{\gamma}(0), \ddot{\gamma}(0)) + \frac{dF}{dx}\dddot{\gamma}(0) \quad (12.4)$$

that shows that term (12.4) is defined only up to $\text{Im} \, D_x F$.

Thus is intrinsically defined only a certain part of the second differential, which is called the *Hessian of $F$*, i.e. the quadratic map

$$\text{Hess}_x F : \text{Ker} \, D_x F \to \mathcal{T}_{F(x)} M / \text{Im} \, D_x F$$

### 12.2 Abnormal extremals and regularity of the distance

In the previous chapter we proved that if we have abnormal minimizer that reach some point $q$, then the sub-Riemannian distance is not smooth at $q$. If we also have that no normal minimizers reach $q$ we can say that it is not even Lipschitz.

**Proposition 12.1.** Assume that there are no normal minimizers that join $q_0$ to $\hat{q}$. Then $f$ is not Lipschitz in a neighborhood of $\hat{q}$. Moreover

$$\lim_{q \to \hat{q}} |d_q f| = +\infty. \quad (12.5)$$

In the previous theorem $| \cdot |$ is an arbitrary norm of the fibers of $T^* M$.

**Proof.** Consider a sequence of smooth points $q_n \in \Sigma$ such that $q_n \to \hat{q}$. Since $q_n$ are smooth we know that there exists unique controls $u_n$ and covectors $\lambda_n$ such that

$$\lambda_n D_{u_n} F = u_n, \quad \lambda_n = d_{q_n} f.$$ 

Assume by contradiction that $|d_{q_n} f| \leq M$ then, using compactness we find that $u_n \to u$, $\lambda_n \to \lambda$ with $\lambda D_u F = u$, that means that the associate geodesic reach $\hat{q}$. In other words, there exists a normal minimizer that goes at $\hat{q}$, that is a contradiction. \qed

Let us now consider the end-point map $F : \mathcal{U} \to M$. As we explained in the previous section, its Hessian at a point $u \in \mathcal{U}$ is the quadratic vector function

$$\text{Hess}_u F : \text{Ker} \, D_u F \to \text{Coker} \, D_u F = T_{F(u)} M / \text{Im} \, D_u F.$$ 

**Remark 12.2.** Recall that $\lambda D_u F = 0$ if and only if $\lambda \in (\text{Im} \, D_u F)^\perp$. In other words, for every abnormal extremal there is a well defined scalar quadratic form

$$\lambda \text{Hess}_u F : \text{Ker} \, D_u F \to \mathbb{R}$$

Notice that the dimension of the space $\text{Im} \, D_u F^\perp$ of such covectors coincide with $\dim \text{Coker} \, D_u F$. 

246
Definition 12.3. Let $Q : V \to \mathbb{R}$ be a quadratic form defined on a vector space $V$. The index of $Q$ is the maximal dimension of a negative subspace of $Q$:

$$\text{ind}^- Q = \sup \{ \dim W \mid Q|_{W\setminus\{0\}} < 0 \}. \quad (12.6)$$

Recall that in the finite-dimensional case this number coincides with the number of negative eigenvalues in the diagonal form of $Q$.

The following notion of index of the map $F$ will be also useful:

Definition 12.4. Let $F : U \to M$ and $u \in U$ be a critical point for $F$. The index of $F$ at $u$ is

$$\text{Ind}_u F = \min_{\lambda \in \text{Im} D_u F^\perp} \text{ind}^- (\lambda \text{Hess}_u F) - \text{codim} \text{Im} D_u F$$

Remark 12.5. If $\text{codim} \text{Im} D_u F = 1$, then there exists a unique (up to scalar multiplication) non-zero $\lambda \perp \text{Im} D_u F$, hence $\text{Ind}_u F = \text{ind}^- (\lambda \text{Hess}_u F) - 1$.

Theorem 12.6. If $\text{Ind}_u F \geq 1$, then $u$ is not a strictly abnormal minimizer.

We state without proof the following result (see Lemma 20.8 of [3])

Lemma 12.7. Let $Q : \mathbb{R}^N \to \mathbb{R}^n$ be a vector valued quadratic form. Assume that $\text{Ind}_0 Q \geq 0$. Then there exists a regular point $x \in \mathbb{R}^n$ of $Q$ such that $Q(x) = 0$.

Definition 12.8. Let $\Phi : E \to \mathbb{R}^n$ be a smooth map defined on a linear space $E$ and $r > 0$. We say that $\Phi$ is $r$-solid at a point $x \in E$ if there exists a constant $C > 0$, $\bar{\epsilon} > 0$ and a neighborhood $U$ of $x$ such that for all $\epsilon < \bar{\epsilon}$ there exists $\delta(\epsilon) > 0$ satisfying

$$B_{\hat{\Phi}(x)} (C\epsilon^r) \subset \hat{\Phi}(B_x(\epsilon)), \quad (12.7)$$

for all maps $\hat{\Phi} \in C^0(E, \mathbb{R}^n)$ such that $\|\hat{\Phi} - \Phi\|_{C^0(U, \mathbb{R}^n)} < \delta$.

Exercise 12.9. Prove that if $x$ is a regular point of $\Phi$, then $\Phi$ is 1-solid at $x$.

(Hint: Use implicit function theorem to prove that $\Phi$ satisfies (12.7) and Brower theorem to show that the same holds for some small perturbation)

Proposition 12.10. Assume that $\text{Ind}_x \Phi \geq 0$. Then $\Phi$ is 2-solid at $x$.

Proof. We can assume that $x = 0$ and that $\Phi(0) = 0$. We divide the proof in two steps: first we prove that there exists a finite dimensional subspace $E' \subset E$ such that the restriction $\Phi|_{E'}$ satisfies the assumptions of the theorem. Then we prove the proposition under the assumption that $\dim E < +\infty$.

(i). Denote $k := \dim \text{Coker} D_0 \Phi$ and consider the Hessian

$$\text{Hess}_0 \Phi : \text{Ker} D_0 \Phi \to \text{Coker} D_0 \Phi$$

We can rewrite the assumption on the index of $\Phi$ as follows

$$\text{ind}^- \lambda \text{Hess}_0 \Phi \geq k, \quad \forall \lambda \in \text{Im} D_0 \Phi^\perp \setminus \{0\}. \quad (12.8)$$
Since property (12.8) is invariant by multiplication of the covector by a positive scalar we are reduced to the sphere

\[ \lambda \in S^{k-1} = \{ \lambda \in \text{Im } D_0 \Phi^\perp, |\lambda| = 1 \}. \]

By definition of index, for every \( \lambda \in S^{k-1} \), there exists a subspace \( E_\lambda \subset E \), \( \dim E_\lambda = k \) such that

\[ \lambda \text{ Hess}_u \Phi|_{E_\lambda \setminus \{0\}} < 0 \]

By the continuity of the form with respect to \( \lambda \), there exists a neighborhood \( O_\lambda \) of \( \lambda \) such that

\[ E_\lambda' = E_\lambda \text{ for every } \lambda' \in O_\lambda. \]

By compactness we can choose a finite covering of \( S^{k-1} \) made by open subsets

\[ S^{k-1} = O_{\lambda_1} \cup \ldots \cup O_{\lambda_N} \]

Then it is sufficient to consider the finite-dimensional subspace

\[ E' = \bigoplus_{j=1}^N E_{\lambda_j} \]

(ii). Assume \( \dim E < \infty \) and split

\[ E = E_1 \oplus E_2 \quad E_2 := \text{Ker } D_0 \Phi \]

The Hessian is a map

\[ \text{Hess}_0 \Phi : E_2 \to \mathbb{R}^n / D_0 \Phi(E_1) \]

According to Lemma 12.7 there exists \( e_2 \in E_2 \), regular point of \( \text{Hess}_0 \Phi \), such that

\[ \text{Hess}_0 \Phi(e_2) = 0 \quad \implies \quad D_0^2 \Phi(e_2) = D_0 \Phi(e_1), \quad \text{for some } e_1 \in E_1. \]

Define the map \( Q : E \to \mathbb{R}^n \) by the formula

\[ Q(v_1 + v_2) := D_0 \Phi(v_1) + \frac{1}{2} D_0^2 \Phi(v_2), \quad v = v_1 + v_2 \in E = E_1 \oplus E_2. \]

and the vector \( e := -e_1/2 + e_2 \). From our assumptions it follows that \( e \) is a regular point of \( Q \) and \( Q(e) = 0 \). In particular there exists \( c > 0 \) such that

\[ B_0(c) \subset Q(B_0(1)) \]

and the same holds for some perturbation of the map \( Q \) (see Exercise 12.9). Consider then the map

\[ \Phi_\varepsilon : v_1 + v_2 \mapsto \frac{1}{\varepsilon^2} \Phi(\varepsilon^2 v_1 + \varepsilon v_2) \quad (12.9) \]

Using that \( v_2 \in \text{Ker } D_0 \Phi \) we compute the Taylor expansion with respect to \( \varepsilon \)

\[ \Phi_\varepsilon(v_1 + v_2) = Q(v_1 + v_2) + O(\varepsilon) \quad (12.10) \]

hence for small \( \varepsilon \) the image of \( \Phi_\varepsilon \) contain a ball around 0 from which it follows that

\[ B_{\delta(0)}(c\varepsilon^2) \subset \Phi(B_0(\varepsilon)) \quad (12.11) \]

Moreover as soon as \( \varepsilon \) is fixed we can perturb the map \( \Phi \) and still the estimate (12.11) holds. \( \square \)
Actually we proved the following statement, that is stronger than 2-solideness of $\Phi$:

**Lemma 12.11.** Under the assumptions of the Theorem [12.10] there exists $C > 0$ such that for every $\varepsilon$ small enough

$$B_{\Phi(0)}(C\varepsilon^2) \subset \Phi(B'_0(\varepsilon^2) \times B''_0(\varepsilon))$$

(12.12)

where $B'$ and $B''$ denotes the balls in $E_1$ and $E_2$ respectively.

The key point is that, in the subspace where the differential of $\Phi$ vanish, the ball of radius $\varepsilon$ is mapped into a ball of radius $\varepsilon^2$, while the restriction on the other subspace “preserves” the order, as the estimates (12.9) and (12.10) show.\footnote{If $0 < \varepsilon < 1$, then $B_0(c) \subset B_0(c\varepsilon) \subset B_0(c\varepsilon^2) \subset B_0(c\varepsilon^3)$ and $B_0(c) \subset B'_0(c\varepsilon^2) \subset B'_0(c\varepsilon^3)$}

**Proof of Theorem 12.6.** We prove that if $\text{Ind}_u F \geq 1$, where $u$ is a strictly abnormal geodesic, then $u$ cannot be a minimizer. It is sufficient to show that the “extended” endpoint map

$$\Phi : \mathcal{U} \to \mathbb{R} \times M, \quad \Phi(u) = \left( J(u), F(u) \right),$$

is locally open at $u$. Recall that $d_u J = \lambda D_u F$, for some $\lambda \in T_{F(u)} M$, if and only if $d_u J |_{\text{Ker} D_u F} = 0$ (see also Proposition 8.11). Since $u$ is strictly abnormal, it follows that $d_u J |_{\text{Ker} D_u F} \neq 0$.\footnote{If $0 < \varepsilon < 1$, then $B_0(c) \subset B_0(c\varepsilon) \subset B_0(c\varepsilon^2) \subset B_0(c\varepsilon^3)$ and $B_0(c) \subset B'_0(c\varepsilon^2) \subset B'_0(c\varepsilon^3)$}

Moreover from the definition of $\Phi$ and (12.13) one has

$$\text{Ker} D_u \Phi = \text{Ker} d_u J \cap \text{Ker} D_u F, \quad \dim d_u J = 1.$$

Moreover, a covector $\bar{\lambda} = (\alpha, \lambda)$ in $\mathbb{R} \times T^*_F(u) M$ annihilates the image of $D_u \Phi$ if and only if $\alpha = 0$ and $\lambda \in \text{Im} D_u F^\perp$, indeed if

$$0 = \bar{\lambda} D_u \Phi = \alpha d_u J + \lambda D_u F$$

with $\alpha \neq 0$, this would imply that $u$ is also normal. In other words we proved the equality

$$\text{Im} D_u \Phi^\perp = \{ (0, \lambda) \in \mathbb{R} \times T^*_F(u) M \mid \lambda \in \text{Im} D_u F^\perp \}$$

(12.14)

Combining (12.13) and (12.14) one obtains for every $\bar{\lambda} = (0, \lambda) \in \text{Im} D_u \Phi^\perp$

$$\bar{\lambda} \text{Hess}_u \Phi = \lambda \text{Hess}_u F|_{\text{Ker} d_u J \cap \text{Ker} D_u F}$$

(12.15)

Moreover codim $\text{Im} D_u \Phi = \text{codim} \text{Im} D_u F$ since $\dim \text{Im} D_u F = \dim \text{Im} D_u F + 1$ by (12.13) and $D_u \Phi$ takes values in $\mathbb{R} \times T^*_F(u) M$. Then for every $\bar{\lambda} = (0, \lambda) \in \text{Im} D_u \Phi^\perp$

$$\text{ind}^{-} (\bar{\lambda} \text{Hess}_u \Phi) - \text{codim} \text{Im} D_u \Phi = \text{ind}^{-} (\lambda \text{Hess}_u F|_{\text{Ker} d_u J \cap \text{Ker} D_u F}) - \text{codim} \text{Im} D_u F$$

$$\geq \text{ind}^{-} (\lambda \text{Hess}_u F) - 1 - \text{codim} \text{Im} D_u F$$

and passing to the infimum with respect to $\bar{\lambda}$ we get

$$\text{Ind}_u \Phi \geq \text{Ind}_u F - 1 \geq 0.$$ 

By Proposition [12.10] this implies that $\Phi$ is locally open at $u$. Hence $u$ cannot be a minimizer.\qed
Now we prove that, under the same assumptions on the index of the endpoint map given in Theorem \[12.6\] the sub-Riemannian is Lipschitz even if some abnormal minimizers are present.

**Theorem 12.12.** Let \( K \subset B_{q_0}(r_0) \) be a compact and assume that \( \text{Ind}_u F \geq 1 \) for every abnormal minimizer \( u \) such that \( F(u) \in K \). Then \( \bar{f} \) is Lipschitz on \( K \).

**Proof.** Recall that if there are no abnormal minimizers reaching \( K \), Theorem \[12.11\] ensures that \( \bar{f} \) is Lipschitz on \( K \). Then, using compactness of the set of all minimizers, it is sufficient to prove the estimate in neighborhood of a point \( q = F(u) \), where \( u \) is abnormal.

Since \( \text{Ind}_u F \geq 1 \) by assumption, Theorem \[12.6\] implies that every abnormal minimizer \( u \) is not strictly abnormal, i.e., has also a normal lift. We have

\[
\text{Hess}_u F : \text{Ker} D_u F \rightarrow \text{Coker} D_u F, \quad \text{with} \quad \text{Ind}_u F \geq 1.
\]

and, since \( u \) is also normal, it follows that \( d_u J = \lambda D_u F \) for some \( \lambda \in T^*_F(u) M \), hence \( \text{Ker} D_u F \subset \text{Ker} d_u J \). The assumption of Lemma \[12.11\] are satisfied, hence splitting the the space of controls

\[
L^2_k([0, 1]) = E_1 \oplus E_2, \quad E_2 := \text{Ker} D_u F
\]

we have that there exists \( C_0 > 0 \) and \( R > 0 \) such that for \( 0 \leq \varepsilon < R \) we have

\[
B_q(C_0 \varepsilon^2) \subset F(B_\varepsilon), \quad B_\varepsilon := B'_u(\varepsilon^2) \times B''_u(\varepsilon), \quad q = F(u), \quad (12.16)
\]

where \( B'_u(r) \) and \( B''_u(r) \) are the ball of radius \( r \) in \( E_1 \) and \( E_2 \) respectively, and \( B_q(r) \) is the ball of radius \( r \) in coordinates on \( M \).

Let us also observe that, since \( J \) is smooth on \( B'_u(\varepsilon^2) \times B''_u(\varepsilon) \), with \( d_u J = 0 \) on \( E_2 \), by Taylor expansion we can find constants \( C_1, C_2 > 0 \) such that for every \( u' = (u'_1, u'_2) \in B_\varepsilon \) one has (we write \( u = (u_1, u_2) \))

\[
J(u') - J(u) \leq C_1 \|u'_1 - u_1\| + C_2 \|u'_2 - u_2\|^2
\]

Pick then any point \( q' \in K \) such that \( |q' - q| = C_0 \varepsilon^2 \), with \( 0 \leq \varepsilon < R \). Then \( (12.16) \) implies that there exists \( u' = (u'_1, u'_2) \in B_\varepsilon \) such that \( F(u') = q' \). Using that \( f(q') \leq J(u') \) and \( f(q) = J(u) \), since \( u \) is a minimizer, we have

\[
f(q') - f(q) \leq J(u') - J(u) \leq C_1 \|u'_1 - u_1\| + C_2 \|u'_2 - u_2\|^2 \leq C \varepsilon^2 = C'|q' - q| \quad (12.17)
\]

where we can choose \( C = \max\{C_1, C_2\} \) and \( C' = C/C_0 \).

Since \( K \) is compact, and the set of control \( u \) associated with minimizers that reach the compact set \( K \) is also compact, the constants \( R > 0 \) and \( C_0, C_1, C_2 \) can be chosen uniformly with respect to \( q \in K \). Hence we can exchange the role of \( q' \) and \( q \) in the above reasoning and get

\[
|f(q') - f(q)| \leq C'|q' - q|,
\]

for every pair of points \( q, q' \) such that \( |q' - q| \leq C_0 R^2 \). \( \square \)
12.3 Goh and generalized Legendre conditions

In this section we present some necessary conditions for the index of the quadratic form along an abnormal extremal to be finite.

**Theorem 12.13.** Let $u$ be an abnormal minimizer and let $\lambda_1 \in T^*_F(u)M$ satisfy $\lambda_1 D_u F = 0$. Assume that $\text{ind}^{-}\lambda_1 \text{Hess}_u F < +\infty$. Then the following condition are satisfied:

(i) $\langle \lambda(t), [f_i, f_j](\gamma(t)) \rangle \equiv 0$, for a.e. $t$, $\forall i, j = 1, \ldots, k$, (Goh condition)

(ii) $\langle \lambda(t), [[f_u(t), f_v], f_v](\gamma(t)) \rangle \geq 0$, for a.e. $t$, $\forall v \in \mathbb{R}^k$, (Generalized Legendre condition)

where $\lambda(t)$ and $\gamma(t) = \pi(\lambda(t))$ are respectively the extremal and the trajectory associated to $\lambda_1$.

**Remark 12.14.** Notice that, in the statement of the previous theorem, if $\lambda_1$ satisfies the assumption $\lambda_1 D_u F = 0$, then also $-\lambda_1$ satisfies the same assumptions. Since $\text{ind}^{-}(-\lambda_1 \text{Hess}_u F) = \text{ind}^{+}\lambda_1 \text{Hess}_u F$ this implies that the statement holds under the assumption $\text{ind}^{+}\lambda_1 \text{Hess}_u F < +\infty$. Indeed the proof shows that as soon as the Goh condition is not satisfied, both the positive and the negative index of this form are infinity.

Notice that these condition are related to the properties of the distribution of the sub-Riemannian structure and not to the metric. Indeed recall that the extremal $\lambda(t)$ is abnormal if and only if it satisfies

$$\dot{\lambda}(t) = \sum_{i=1}^{k} u_i(t) \tilde{h}_i(\lambda(t)), \quad \langle \lambda(t), f_i(\gamma(t)) \rangle = 0, \quad \forall i = 1, \ldots, k,$$

i.e. $\lambda(t)$ satisfies the Hamiltonian equation and belongs to $\mathcal{D}^\perp_{\gamma(t)}$. Goh condition are equivalent to require that $\lambda(t) \in (\mathcal{D}^2_{\gamma(t)})^\perp$.

**Corollary 12.15.** Assume that the sub-Riemannian structure is 2-generating, i.e. $\mathcal{D}^2_q = T_q M$ for all $q \in M$. Then there are no strictly abnormal minimizers. In particular $\tilde{f}$ is locally Lipschitz on $M$.

**Proof.** Since $\mathcal{D}^2_q = T_q M$ implies $(\mathcal{D}^2_{\gamma(t)})^\perp = 0$ for every $q \in M$, no abnormal extremal can satisfy the Goh condition. Hence by Theorem 12.13 it follows that $\text{Ind}_u F = +\infty$, for any abnormal minimizer $u$. In particular, from Theorem 12.6 it follows that the minimizer cannot be strictly abnormal. Hence $\tilde{f}$ is globally Lipschitz by Theorem 12.12.

**Remark 12.16.** Notice that $\tilde{f}$ is locally Lipschitz on $M$ if and only if the sub-Riemannian structure is 2-generating. Indeed if the structure is not 2-generating at a point $q$, then from Ball-Box Theorem (Corollary 10.53) it follows that the squared distance $\tilde{f}$ is not Lipschitz at the base point $q_0$.

On the other hand, on the set where $\tilde{f}$ is positive, we have that $\tilde{f}$ is Lipschitz if and only if the sub-Riemannian distance $d(q_0, \cdot)$ is.

Before going into the proof of the Goh conditions (Theorem 12.13) we discuss an important corollary.
**Theorem 12.17.** Assume that $D_{q_0} \neq D_{q_0}^2$. Then for every $\varepsilon > 0$ there exists a normal extremal path $\gamma$ starting from $q_0$ such that $\ell(\gamma) = \varepsilon$ and $\gamma$ is not a length-minimizer.

Before the proof, this is the idea: fix an element $\xi \in D_{q_0}^+ \setminus \langle D_{q_0}^2 \rangle^\perp$ which is non empty by assumptions. We want to build an abnormal minimizing trajectory that has $\xi$ as initial covector and that is the limit of a sequence of strictly normal length-minimizers. In this way this abnormal will have finite index (the abnormal quadratic form will be the limit of positive ones) and then by Goh condition $\xi \cdot D_{q_0}^2 = 0$, which is a contradiction.

**Proof.** Assume by contradiction that there exists $T > 0$ such that all normal extremal paths $\gamma_\lambda$ associated with initial covector $\lambda \in H^{-1}(1/2) \cap T_{q_0}^* M$ minimize on the segment $[0, T]$. Since restriction of length-minimizers are still length-minimizers, by suitably reducing $T > 0$, we can assume, thanks to Lemma 3.34 that there exists a compact set $K$ such that $\{\gamma_\lambda(T) \mid \lambda \in H^{-1}(1/2)\} \subset K$.

Fix an element $\xi \in D_{q_0}^+ \setminus \langle D_{q_0}^2 \rangle^\perp$, which is non empty by assumptions. Then consider, given any $\lambda_0 \in H^{-1}(1/2) \cap T_{q_0}^* M$, the family of normal extremal paths (and corresponding normal trajectories)

$$
\lambda_s(t) = e^{tH}(\lambda_0 + s\xi), \quad \gamma_s(t) = \pi(\lambda_s(t)), \quad t \in [0, T],
$$

and let $u_s$ be the control associated with $\gamma_s$, and defined on $[0, T]$. Due to Theorem 11.9 there exists a positive sequence $s_n \to +\infty$ such that $q_n := \gamma_{s_n}(T)$ is a smooth point for the squared distance from $q_0$, for every $n \in \mathbb{N}$. By compactness of minimizers reaching $K$, there exists a subsequence of $s_n$ that we still denote by the same symbol, and a minimizing control $\bar{u}$ such that $u_{s_n} \to \bar{u}$, when $n \to \infty$. In particular $\gamma_{s_n}$ is a strictly normal length-minimizer for every $n \in \mathbb{N}$.

Denote $\Phi^n_t = P^n_{0,t}$ the non autonomous flow generated by the control $u_{s_n}$. The family $\lambda_{s_n}(t)$ satisfies

$$
\lambda_{s_n}(t) = e^{tH}(\lambda_0 + s_n\xi) = (\Phi^n_t)^* (\lambda_0 + s_n\xi).
$$

Moreover, by continuity of the flow with respect to convergence of controls, we have that $\Phi^n_t \to \Phi_t$ for $n \to \infty$, where $\Phi_t$ denotes the flow associated with the control $\bar{u}$. Hence we have that the rescaled family

$$
\frac{1}{s_n} \lambda_{s_n}(t) = (\Phi^n_t)^* \left( \frac{1}{s_n} \lambda_0 + \xi \right)
$$

converges for $n \to \infty$ to the limit extremal $\lambda(\bar{t}) = \Phi_{\bar{t}}^* \xi$. Notice that $\lambda(t)$ is, by construction, an abnormal extremal associated to the minimizing control $\bar{u}$, and with initial covector $\xi$.

The fact that $u_{s_n}$ is a strictly normal minimizer says that the Hessian of the energy $J$ restricted to the level set $F^{-1}(q_n)$ is non negative. Recall that

$$
\text{Hess}_u J|_{F^{-1}(q)} = I - \lambda_1 D_u^2 F,
$$

where $\lambda_1 \in T_{F(u)} M$ is the final covector of the extremal lift. In particular we have for every $n \in \mathbb{N}$ and every control $v$ the following inequality

$$
\|v\|^2 - \lambda_{s_n}(T) D_{u_{s_n}}^2 F(v, v) \geq 0.
$$

This immediately implies

$$
\frac{1}{s_n} \|v\|^2 - \frac{1}{s_n} \lambda_{s_n}(T) D_{u_{s_n}}^2 F(v, v) \geq 0,
$$

indeed it is enough to fix an arbitrary compact $K$ with $q_0 \in \text{int}(K)$ such that the corresponding $\delta_K$ defined by Lemma 3.34 is smaller than $T$.
and passing to the limit for \( n \to \infty \) one gets

\[-\bar{\lambda}(T)D_u^2 F(v, v) \geq 0.\]

In particular one has that

\[\text{ind}^+ \bar{\lambda}(T)\text{Hess}_u F = \text{ind}^- (-\bar{\lambda}(T)D_u^2 F) = 0.\]

Hence the abnormal extremal has finite (positive) index and we can apply Goh conditions (see Theorem 12.13 and Remark 12.14). Thus \( \xi \) is orthogonal to \( D_u^2 q_0 \), which is a contradiction since \( \xi \in D_{q_0}^\perp \setminus (D_u^2 q_0)^\perp \).

Remark 12.18 (About the assumptions of Theorem 12.17). Assume that the sub-Riemannian structure is bracket-generating and is not Riemannian in an open set \( O \subset M \), i.e., \( D_{q_0} \neq T_{q_0} M \) for every \( q \in O \). Then there exists a dense set \( D \subset O \) such that \( D_{q_0} \neq D_u^2 q_0 \) for every \( q \in D \).

Indeed assume that \( D_q \neq D_u^2 q \) for all \( q \) in an open set \( A \), then it is easy to see that \( D_i q \neq T_q M \) for all \( q \in A \), since the structure is not Riemannian. Hence the structure is not bracket-generating in \( A \), which gives a contradiction.

12.3.1 Proof of Goh condition - (i) of Theorem 12.13

Proof of Theorem 12.13. Denote by \( u \) the abnormal control and by \( P_t = \exp \int_0^t f_{u(s)} ds \) the nonautonomous flow generated by \( u \). Following the argument used in the proof of Proposition 8.4 we can write the end-point map as the composition

\[E(u + v) = P_1 (G(v)), \quad D_u E = P_1 D_0 G,\]

and reduced the problem to the expansion of \( G \), which is easier. Indeed denoting \( g^i_t := P^{-1}_t f_i \), the map \( G \) can be interpreted as the end-point map for the system

\[\dot{q}(t) = g^i_{v(t)}(q(t)) = \sum_{i=1}^k v_i(t) g^i_t(q(t))\]

and the Hessian of \( F \) can be computed easily starting from the Hessian of \( G \) at \( v = 0 \)

\[\text{Hess}_u F = P_1 \text{Hess}_0 G\]

from which we get, using that \( \lambda_0 = P^*_1 \lambda_1 \),

\[\lambda_1 \text{Hess}_u F = \lambda_1 P_1 \text{Hess}_0 G = \lambda_0 \text{Hess}_0 G\]

Moreover computing

\[\langle \lambda(t), [f_i, f_j]\gamma(t) \rangle = \langle \lambda_0, P^{-1}_t [f_i, f_j](\gamma(t)) \rangle = \langle \lambda_0, [g^i_t, g^j_t]\gamma(0) \rangle\]

the Goh and generalized Legendre conditions can also be rewritten as

\[\langle \lambda_0, [g^i_t, g^j_t]\gamma(0) \rangle \equiv 0, \quad \text{for a.e. } t \in [0, 1], \quad \forall i, j = 1, \ldots, k, \quad (G.1)\]

\[\langle \lambda_0, [[g^i_{v(t)}, g^i_t], g^j_t]\gamma(0) \rangle \geq 0, \quad \text{for a.e. } t \in [0, 1], \quad \forall i = 1, \ldots, k. \quad (L.1)\]

253
Now we want to compute the Hessian of the map $G$. Using the Volterra expansion computed in Chapter 6 we have

$$G(v(\cdot)) \simeq q_0 \circ \left( \text{Id} + \int_0^1 g^t v(t) dt + \int_{0 \leq \tau \leq t \leq 1} g^\tau v(t) \circ g^t v(t) d\tau dt \right) + O(\|v\|^3)$$

where we used that $g^t v$ is linear with respect to $v$ to estimate the remainder.

This expansion let us to recover immediately the linear part, i.e. the expressions for the first differential, which can be interpreted geometrically as the integral mean

$$D_0 G(v) = \int_0^1 g^t v(q_0) dt,$$

On the other hand the expression for the quadratic part, i.e. the second differential

$$D^2_0 G(v) = 2 q_0 \circ \int_{0 \leq \tau \leq t \leq 1} g^\tau v(t) \circ g^t v(t) d\tau dt.$$ 

has not an immediate geometrical interpretation. Recall that the second differential $D^2_0 G$ is defined on the set

$$\text{Ker } D_0 G = \{ v \in L^2_k[0, 1], \int_0^1 g^t v(q_0) dt = 0 \} \quad (12.19)$$

and, for such a $v$, $D^2_0 G(v)$ belong to the tangent space $T_{q_0}M$. Indeed, using Lemma 8.27 and that $v$ belong to the set (12.19), we can symmetrize the second derivative, getting the formula

$$D^2_0 G(v) = \int_{0 \leq \tau \leq t \leq 1} [g^\tau v(t), g^t v(t)](q_0) d\tau dt,$$

which shows that the second differential is computed by the integral mean of the commutator of the vector field $g^t v(t)$ for different times.

Now consider an element $\lambda_0 \in \text{Im } D_0 G^\perp$, i.e. that satisfies

$$\langle \lambda_0, g^t v(q_0) \rangle = 0, \quad \text{for a.e. } t \in [0, 1], \forall v \in \mathbb{R}^k.$$

Then we can compute the Hessian

$$\lambda_0 \text{Hess}_0 G(v) = \int_{0 \leq \tau \leq t \leq 1} \langle \lambda_0, [g^\tau v(t), g^t v(t)](q_0) \rangle d\tau dt \quad (12.20)$$

Remark 12.19. Denoting by $K$ the bilinear form

$$K(\tau, t)(v, w) = \langle \lambda_0, [g^\tau v(t), g^t w(t)](q_0) \rangle,$$

the Goh and generalized Legendre conditions are rewritten as follows

$$K(t, t)(v, v) = 0, \quad \forall v, w \in \mathbb{R}^k, \quad \text{for a.e. } t \in [0, 1], \quad (G.2)$$

$$\frac{\partial K}{\partial \tau}(\tau, t) \bigg|_{\tau=t} (v, v) \geq 0, \quad \forall v \in \mathbb{R}^k, \quad \text{for a.e. } t \in [0, 1]. \quad (L.2)$$
Indeed, the first one easily follows from (G.1). Moreover recall that \( g^t_v = P_t^{-1} f_v \), hence the map \( t \mapsto g^t_v \) is Lipschitz for every fixed \( v \). By definition of \( P_t = \exp \int_t^1 f_{w(t)} dt \) it follows that
\[
\frac{\partial}{\partial t} g^t_v = [g^t_{u(t)}, g^t_v]
\]
which shows that (L.2) is equivalent to (L.1).

Finally we want to express the Hessian of \( G \) in Hamiltonian terms. To this end, consider the family of functions on \( T^*M \) which are linear on fibers, associated to the vector fields \( g^t_v \):
\[
h^t_v(\lambda) := \langle \lambda, g^t_v(q) \rangle, \quad \lambda \in T^*M, \quad q = \pi(\lambda).
\]
and define, for a fixed element \( \lambda_0 \in \text{Im} \, D_0^+G \):
\[
\eta^t_v := \tilde{h}^t_v(\lambda_0) \in T_{\lambda_0} T^*M
\] (12.21)

Using the identities
\[
\sigma_\lambda(\tilde{h}^t_v, \tilde{h}^t_w) = \{h^t_v, h^t_w\}(\lambda) = \langle \lambda, [g^t_v, g^t_w](q) \rangle, \quad q = \pi(\lambda)
\]
and computing at the point \( \lambda_0 \in T_{\lambda_0} T^*M \) we find
\[
\sigma_{\lambda_0}(\eta^t_v, \eta^t_w) = \langle \lambda_0, [g^t_v, g^t_w](q_0) \rangle
\]
and we get the final expression for the Hessian
\[
\lambda_0 \text{Hess}_0G(v(\cdot)) = \int_0^1 \int_0^1 \sigma_{\lambda_0}(\eta^t_v(\tau), \eta^t_v(t)) dt d\tau.
\] (12.22)

where the control \( v \in \text{Ker} \, D_0G \) satisfies the relation (notice that \( \pi_* \eta^t_v = g^t_v(q_0) \))
\[
\pi_* \int_0^1 \eta^t_v(\tau) dt = \int_0^1 \pi_* \eta^t_v(t) dt = 0
\]

Moreover the “Hamiltonian” version of Goh and Legendre conditions is expressed as follows:
\[
\sigma_{\lambda_0}(\eta^t_v, \eta^t_w) = 0, \quad \forall \, v, w \in \mathbb{R}^k, \quad \text{for a.e. } t \in [0, 1], \quad (G.3)
\]
\[
\sigma_{\lambda_0}(\eta^t_v, \eta^t_v) \geq 0, \quad \forall \, v \in \mathbb{R}^k, \quad \text{for a.e. } t \in [0, 1]. \quad (L.3)
\]

We are reduced to prove, under the assumption \( \text{ind}^{-} \lambda_0 \text{Hess}_0G < +\infty \), that (G.3) and (L.3) hold. Actually we will prove that Goh and generalized Legendre conditions are necessary conditions for the restriction of the quadratic form to the subspace of controls in \( \text{Ker} \, D_0G \) that are concentrated on small segments \([t, t+s]\).

In what follows we fix once for all \( t \in [0, 1] \). Consider an arbitrary vector control function \( v : [0, 1] \rightarrow \mathbb{R}^k \) with compact support in \([0, 1]\) and build, for \( s > 0 \) small enough, the control
\[
v_s(\tau) = v \left( \frac{\tau - t}{s} \right), \quad \text{supp } v_s \subset [t, t+s]. \quad (12.23)
\]
The idea is to apply the Hessian to this particular control functions and then compute the asymptotics for \( s \to 0 \).

Notice that this space depend on the choice of \( t \), while \( \text{codim} E_s \) does not depend on \( s \).

**Remark 12.20.** We will use the following identity (writing \( \sigma \) for \( \sigma_{\lambda_0} \)), which holds for arbitrary control functions \( v, w : [0, 1] \to \mathbb{R}^k \)

\[
\int_{\alpha \leq \tau \leq t \leq \beta} \sigma(\eta^r_{v(t)}, \eta^t_{w(t)}) dt d\tau = \int_{\alpha \leq \tau \leq t \leq \beta} \sigma(\int_{\alpha}^{\tau} \eta^r_{v(t)} dt, \eta^t_{w(t)}) dt = \int_{\alpha \leq \tau \leq t \leq \beta} \sigma(\eta^r_{v}(\tau), \int_{\tau}^{\beta} \eta^t_{w(t)} dt) d\tau. \tag{12.24}
\]

For the specific choice \( w(t) = \int_{0}^{t} v(\tau) d\tau \) we have also the integration by parts formula

\[
\int_{\alpha \leq \tau \leq t \leq \beta} \eta^t_{w(t)} dt = \eta^\beta_{w(\beta)} - \eta^\alpha_{w(\alpha)} - \int_{\alpha}^{\beta} \dot{\eta}^t_{w(t)} dt. \tag{12.25}
\]

Combining (12.22) and (12.24), we rewrite the Hessian applied to \( v_s \) as follows

\[
\lambda_0 \text{Hess}_0 G(v_s(\cdot)) = \int_{t}^{t+s} \sigma(\int_{t}^{\tau} \eta^{\theta}_{v_s(\theta)} d\theta, \eta^{\tau}_{v_s(\tau)}) d\tau. \tag{12.26}
\]

Notice that the control \( v_s \) is concentrated on the segment \([t, t+s]\), thus we have restricted the extrema of the integral. The integration by parts formula (12.25), using our boundary conditions, gives

\[
\int_{t}^{\tau} \eta^{\theta}_{v_s(\theta)} d\theta = \eta^{\tau}_{w_s(\tau)} - \int_{t}^{\tau} \eta^{\theta}_{w_s(\theta)} d\theta. \tag{12.27}
\]

where we defined

\[
w_s(\theta) = \int_{t}^{\theta} v_s(\tau) d\tau, \quad \theta \in [t, t+s].
\]

Combining (12.26) and (12.27) one has

\[
\lambda_0 \text{Hess}_0 G(v_s(\cdot)) = \int_{t}^{t+s} \sigma(\eta^{\tau}_{w_s(\tau)}, \eta^{\tau}_{v_s(\tau)}) d\tau - \int_{t}^{t+s} \sigma(\int_{t}^{\tau} \dot{\eta}^{\theta}_{w_s(\theta)} d\theta, \eta^{\tau}_{v_s(\tau)}) d\tau
\]

\[
= \int_{t}^{t+s} \sigma(\eta^{\tau}_{w_s(\tau)}, \eta^{\tau}_{v_s(\tau)}) d\tau - \int_{t}^{t+s} \sigma(\dot{\eta}^{\tau}_{w_s(\tau)}, \int_{\tau}^{t+s} \eta^{\theta}_{v_s(\theta)} d\theta) d\tau - \int_{t}^{t+s} \sigma(\dot{\eta}^{\tau}_{w_s(\tau)}, \int_{\tau}^{t+s} \eta^{\theta}_{w_s(\theta)} d\theta) d\tau. \tag{12.28}
\]

where the second equality uses (12.24).

Next consider the second term in (12.28) and apply again the integration by part formula (recall that \( w_s(t+s) = 0 \))

\[
\int_{t}^{t+s} \sigma(\dot{\eta}^{\tau}_{w_s(\tau)}, \int_{\tau}^{t+s} \eta^{\theta}_{v_s(\theta)} d\theta) d\tau = -\int_{t}^{t+s} \sigma(\dot{\eta}^{\tau}_{w_s(\theta)}, \eta^{\tau}_{v_s(\tau)}) d\tau - \int_{t}^{t+s} \sigma(\dot{\eta}^{\tau}_{w_s(\tau)}, \int_{\tau}^{t+s} \eta^{\theta}_{w_s(\theta)} d\theta) d\tau.
\]
Collecting together all these results one obtains

\[
\lambda_0 \text{Hess}_0 G(v_s(\cdot)) = \int_t^{t+s} \sigma(\eta_{w_s}(\tau), \eta_{w_s}(\tau)) d\tau + \int_t^{t+s} \sigma(\dot{\eta}_{w_s}(\tau), \eta_{w_s}(\tau)) d\tau + \int_t^{t+s} \sigma(\dot{\eta}_{w_s}(\tau), \int_\tau^{t+s} \eta_{w_s}(\theta) d\theta) d\tau.
\]

This is indeed a homogeneous decomposition of \(\lambda_0 \text{Hess}_0 G(v_s(\cdot))\) with respect to \(s\), in the following sense. Since

\[
w_s(\theta) = s w\left(\frac{\theta - t}{s}\right),
\]

we can perform the change of variable

\[
\zeta = \frac{\tau - t}{s}, \quad \tau \in [t, t+s],
\]

and obtain the following expression for the Hessian:

\[
\lambda_0 \text{Hess}_0 G(v_s(\cdot)) = s^2 \int_0^1 \sigma(\eta_{w(\theta)}^{t+s\theta}, \eta_{w(\theta)}^{t+s\theta}) d\theta + s^3 \int_0^1 \sigma(\dot{\eta}_{w(\theta)}^{t+s\theta}, \dot{\eta}_{w(\theta)}^{t+s\theta}) d\theta + s^4 \int_0^1 \sigma(\dot{\eta}_{w(\theta)}^{t+s\theta}, \int_0^1 \dot{\eta}_{w(\zeta)}^{t+s\zeta} d\zeta) d\theta
\]

We recall that here \(v_s\) is defined through a control \(v\) compactly supported in \([0,1]\) by (12.23) and \(w\) is the primitive of \(v\), that is also compactly supported on \([0,1]\).

In particular we can write

\[
\lambda_0 \text{Hess}_0 G(v_s(\cdot)) = s^2 \int_0^1 \sigma(\eta_{w(\theta)}^{t+s\theta}, \eta_{w(\theta)}^{t+s\theta}) d\theta + O(s^3).
\]

By assumption \(\text{ind}^- \lambda_0 \text{Hess}_0 G < +\infty\). This implies that the quadratic form given by its principal part

\[
w(\cdot) \mapsto \int_0^1 \sigma(\eta_{w(\theta)}^t, \eta_{w(\theta)}^t) d\theta,
\]

has also finite index. Indeed, assume that (12.31) has infinite negative index. Then by continuity every sufficiently small perturbation of (12.31) would have infinite index too. Hence, for \(s\) small enough, the quadratic form \(\lambda_0 \text{Hess}_0 G\) would also have infinite index, contradicting our assumption on (12.30).

To prove Goh condition, it is then sufficient to show that if (12.31) has finite index then the integrand is zero, which is guaranteed by the following

**Lemma 12.21.** Let \(A : \mathbb{R}^k \times \mathbb{R}^k \to \mathbb{R}\) be a skew-symmetric bilinear form and define the quadratic form

\[
Q : \mathcal{U} \to \mathbb{R}, \quad Q(w(\cdot)) = \int_0^1 A(w(t), \dot{w}(t)) dt,
\]

where \(\mathcal{U} := \{w(\cdot) \in \text{Lip}[0,1], w(0) = w(1) = 0\}\). Then \(\text{ind}^- Q < +\infty\) if and only if \(A \equiv 0\).

257
Proof. Clearly if $A = 0$, then $Q = 0$ and $\text{ind}^{-}Q = 0$. Assume then that $A \neq 0$ and we prove that $\text{ind}^{-}Q = +\infty$. We divide the proof into steps

(i). The bilinear form $B : \mathcal{U} \times \mathcal{U} \to \mathbb{R}$ defined by

$$B(w_{1}(\cdot), w_{2}(\cdot)) = \int_{0}^{1} A(w_{1}(t), \dot{w}_{2}(t)) dt$$

is symmetric. Indeed, integrating by parts and using the boundary conditions we get

$$B(w_{1}, w_{2}) = \int_{0}^{1} A(w_{1}(t), \dot{w}_{2}(t)) dt = -\int_{0}^{1} A(\dot{w}_{1}(t), w_{2}(t)) dt = \int_{0}^{1} A(w_{2}(t), \dot{w}_{1}(t)) dt = B(w_{2}, w_{1})$$

(ii). $Q$ is not identically zero. Since $Q$ is the quadratic form associated to $B$ and from the polarization formula

$$B(w_{1}, w_{2}) = \frac{1}{4}(Q(w_{1} + w_{2}) - Q(w_{1} - w_{2}))$$

it easily follows that $Q \equiv 0$ if and only if $B \equiv 0$. Then it is sufficient to prove that $B$ is not zero.

Assume that there exists $x, y \in \mathbb{R}^{k}$ such that $A(x, y) \neq 0$, and consider a smooth nonconstant function

$$\alpha : \mathbb{R} \to \mathbb{R}, \quad \text{s.t.} \quad \alpha(0) = \alpha(1) = \dot{\alpha}(0) = \dot{\alpha}(1) = 0.$$ 

Then $\dot{\alpha}(t) z, \alpha(t) z \in \mathcal{U}$ for every $z \in \mathbb{R}^{k}$ and we can compute

$$B(\dot{\alpha}(\cdot)x, \alpha(\cdot)y) = \int_{0}^{1} A(\dot{\alpha}(t)x, \dot{\alpha}(t)y) dt = A(x, y) \int_{0}^{1} \dot{\alpha}(t)^{2} dt \neq 0.$$ 

(iii). $Q$ has the same number of positive and negative eigenvalues. Indeed it is easy to see that $Q$ satisfies the identity

$$Q(w(1 - \cdot)) = -Q(w(\cdot))$$

from which (iii) follows.

(iv). $Q$ is non zero on an infinite dimensional subspace.

Consider some $w \in \mathcal{U}$ such that $Q(w) = \alpha \neq 0$. For every $x = (x_{1}, \ldots, x_{N}) \in \mathbb{R}^{N}$ one can built the function

$$w_{x}(t) = x_{i} w(NT - i), \quad t \in \left[\frac{i}{N}, \frac{i+1}{N}\right], \quad i = 1, \ldots, N.$$ 

An easy computations shows that

$$Q(w_{x}) = \alpha \sum_{i=1}^{N} x_{i}^{2}$$

In particular there exists a subspace of arbitrary large dimension where $Q$ is nondegenerate. \qed
12.3.2 Proof of generalized Legendre condition - (ii) of Theorem 12.13

Applying Lemma 12.21 for any \( t \) we prove that the \( s^2 \) order term in (12.29) vanish and we get to

\[
\lambda_0 \text{Hess}_0 G(v(\cdot)) = s^3 \int_0^1 \sigma(\hat{\eta}_w(\theta), \hat{\eta}_w(\theta)) d\theta + O(s^4)
\]

where the last equality follows from the fact that \( \eta_t^s \) is Lipschitz with respect to \( t \) (see also (12.21)), i.e.

\[
\eta_t^{t+s\theta} = \eta_t^s + O(s)
\]

On the other hand \( \dot{\eta}_u^t \) is only measurable bounded, but the Lebesgue points of \( u \) are the same of \( \dot{\eta} \). In particular if \( t \) is a Lebesgue point of \( \dot{\eta} \), the quantity \( \dot{\eta}_w(\cdot) \) is well defined and we can write

\[
\lambda_0 \text{Hess}_0 G(v(\cdot)) = s^3 \int_0^1 \sigma(\dot{\eta}_w(\theta), \dot{\eta}_w(\theta)) d\theta
\]

Using the linearity of \( \sigma \) and the boundedness of the vector fields we can estimate

\[
\left| \int_0^1 \sigma(\dot{\eta}_w(\theta), \dot{\eta}_w(\theta)) d\theta - \sigma(\dot{\eta}_w(\theta), \dot{\eta}_w(\theta)) d\theta \right| \leq C \int_0^1 |\dot{\eta}_w(\theta) - \dot{\eta}_w(\theta)| d\theta
\]

\[
\leq C \sup_{|\tau| \leq 1} \frac{1}{s} \int_0^s |\dot{\eta}_v^{t+\tau} - \dot{\eta}_v^{t}| d\tau \to 0 \quad s \to 0
\]

where the last term tends to zero by definition of Lebesgue point. Hence we come to

\[
\lambda_0 \text{Hess}_0 G(v(\cdot)) = s^3 \int_0^1 \sigma(\dot{\eta}_w(\theta), \dot{\eta}_w(\theta)) d\theta + o(s^3)
\]

To prove the generalized Legendre condition we have to prove that the integrand is a non negative quadratic form. This follows from the following Lemma, which can be proved similarly to Lemma 12.21.

**Lemma 12.22.** Let \( Q : \mathbb{R}^k \to \mathbb{R} \) be a quadratic form on \( \mathbb{R}^k \) and

\[
\mathcal{U} := \{ w(\cdot) \in \text{Lip}[0,1], w(0) = w(1) = 0 \}.
\]

The quadratic form

\[
Q : \mathcal{U} \to \mathbb{R}, \quad Q(w(\cdot)) = \int_0^1 Q(w(t)) dt
\]

has finite index if and only if \( Q \) is non negative.
12.3.3 More on Goh and generalized Legendre conditions

If Goh condition is satisfied, the generalized Legendre condition can also be characterized as an intrinsic property of the module. Indeed one can see that the quadratic map

\[ U_{\gamma(t)} \to \mathbb{R}, \quad v \mapsto \langle \lambda(t), [[f_{u(t)}, f_v], f_v](\gamma(t)) \rangle \]

is well defined and does not depend on the extension of \( f_v \) to a vector field \( f_{v(t)} \) on \( U \).

Notice that, using the notation \( h_v(\lambda) = \langle \lambda, f_v(q) \rangle \) an abnormal extremal satisfies

\[ h_v(\lambda_t) \equiv 0, \quad \forall v \in \mathbb{R}^k \]

Recalling that the Poisson bracket between linear functions on \( T^*M \) is computed by the Lie bracket

\[ \{ h_v, h_w \}(\lambda) = \langle \lambda, [f_v, f_w](q) \rangle \]

we can rewrite the Goh condition as follows

\[ \{ h_v, h_w \}(\lambda(t)) \equiv 0, \quad \forall v, w \in \mathbb{R}^k \] (12.33)

while strong Legendre conditions reads

\[ \{ \{ h_{u(t)}, h_v \}, h_v \} \geq 0, \quad \forall v \in \mathbb{R}^k \] (12.34)

Taking derivative of (12.33) with respect to \( t \) we find

\[ \{ h_{u(t)}, \{ h_v, h_w \} \}(\lambda(t)) \equiv 0, \quad \forall v, w \in \mathbb{R}^k \]

and using Jacobi identity of the Poisson bracket we get that the bilinear form

\[ (v, w) \mapsto \{ \{ h_{u(t)}, h_v \}, h_w \}(\lambda) \] (12.35)

is symmetric. Hence the generalized Legendre condition says that the quadratic form associated to (12.35) is nonnegative.

Now we want to characterize the trajectories that satisfy these conditions. Recall that, if \( \lambda(t) \) is an abnormal geodesic, we have

\[ \dot{\lambda}(t) = h_{\tilde{u}(t)}(\lambda(t)), \quad h_i(\lambda(t)) \equiv 0, \quad 0 \leq t \leq 1. \] (12.36)

where \( \tilde{h}_{u(t)} = \sum_{i=1}^{k} u_i(t)\tilde{h}_i(t) \). Moreover for any smooth function \( a : T^*M \to \mathbb{R} \)

\[ \frac{d}{dt}a(\lambda(t)) = \{ h_{u(t)}, a \}(\lambda(t)) = \sum_{i=1}^{k} u_i(t)\{ h_i, a \}(\lambda(t)) \]

**Notation.** We will denote the iterated Poisson brackets

\[ h_{i_1 \ldots i_k}(\lambda) = \{ h_{i_1}, \ldots, \{ h_{i_{k-1}}, h_{i_k} \} \}(\lambda) \]

\[ = \langle \lambda, [f_{i_1}, \ldots, [f_{i_{k-1}}, f_{i_k}]](q) \rangle, \quad q = \pi(\lambda) \] (12.37, 12.38)
Differentiating the identities in (12.36), using (12.37), we get

\[ h_i(\lambda(t)) = 0 \Rightarrow \sum_{j=1}^{k} u_j(t)h_{ji}(\lambda(t)) = 0, \quad \forall t. \]  

(12.39)

If \( k \) is odd we always have a nontrivial solution of the system, if \( k \) is even is possible only for those \( \lambda \) that satisfy \( \det\{h_{ij}(\lambda)\} = 0 \). But we want to characterize only those controls that satisfy Goh conditions, i.e. such that

\[ h_{ij}(\lambda(t)) \equiv 0. \]  

(12.40)

Hence you cannot recover the control \( u \) from the linear system (12.39). We differentiate again equations (12.40) and we find

\[ \sum_{l=1}^{k} u_l(t)h_{lj}(\lambda(t)) \equiv 0. \]  

(12.41)

For every fixed \( t \), these are \( k(k-1)/2 \) equations in \( k \) variables \( u_1, \ldots, u_k \). Hence

(i) If \( k = 2 \), we have 1 equation in 2 variables and we can recover the control \( u_1, u_2 \) up to a scalar multiplier, if at least one of the coefficients does not vanish. Since we can always deal with length-parametrized curve this uniquely determine the control \( u \).

(ii) If \( k \geq 3 \), we have that the system is overdetermined.

Remark 12.23. For generic systems it is proved that, when \( k \geq 3 \), Goh conditions are not satisfied. On the other hand, in the case of Carnot groups, for big codimension of the distribution, abnormal minimizers always appear.

### 12.4 Rank 2 distributions and nice abnormal extremals

Consider a rank 2 distribution generated by a local frame \( f_1, f_2 \) and let \( h_1, h_2 \) be the associated linear Hamiltonian. An abnormal extremal \( \lambda(t) \) associated with a control \( u(t) \) satisfies the system of equations

\[ \dot{\lambda}(t) = u_1(t)h_1(\lambda(t)) + u_2(t)h_2(\lambda(t)), \]

\[ h_1(\lambda(t)) = h_2(\lambda(t)) = 0. \]  

(12.42)

Define the linear Hamiltonian associated with the \( h_{12}(\lambda(t)) = \langle \lambda, [f_1, f_2](q) \rangle \). Notice that in this special framework the Goh condition is rewritten as \( h_{12}(\lambda(t)) = 0 \) for a.e. \( t \).

Equivalently, every abnormal extremal satisfies Goh conditions if and only if \( \lambda(t) \in (D^2)^\perp \).

Lemma 12.24. Every nontrivial abnormal extremal on a rank 2 sub-Riemannian structure satisfies the Goh condition.

**Proof.** Indeed differentiating the identity (12.42) one gets (we omit \( t \) in the notation for simplicity)

\[ u_2\{h_2, h_1\} = u_2h_{21}(\lambda) = 0, \]

\[ u_1\{h_1, h_2\} = -u_1h_{21}(\lambda) = 0, \]

Since at least one among \( u_1 \) and \( u_2 \) is not identically zero, we have that \( h_{12}(\lambda(t)) \equiv 0 \), that is Goh condition. \( \square \)
From now on we focus on a special class of abnormal extremals.

**Definition 12.25.** An abnormal extremal $\lambda(t)$ is called *nice abnormal* if, for every $t \in [0, 1]$, it satisfies

$$\lambda(t) \in (D^2)^\perp \setminus (D^3)^\perp.$$  

**Remark 12.26.** Assume that $\lambda(t)$ is a nice abnormal extremal. The system (12.41) obtained by differentiating twice the equations (12.42) reads

$$u_1 h_{112}(\lambda) = u_2 h_{221}(\lambda).$$  

(12.43)

Under our assumption, at least one coefficient in (12.43) is nonzero and we can uniquely recover the control $u = (u_1, u_2)$ up to a scalar as follows

$$u_1(t) = h_{221}(\lambda(t)), \quad u_2(t) = h_{112}(\lambda(t)).$$  

(12.44)

If we plug this control into the original equation we find that $\lambda(t)$ is a solution of

$$\dot{\lambda} = h_{221}(\lambda)\vec{h}_1(\lambda) + h_{112}(\lambda)\vec{h}_2(\lambda).$$  

(12.45)

Let us now introduce the quadratic Hamiltonian

$$H_0 = h_{221} h_1 + h_{112} h_2.$$  

(12.46)

**Theorem 12.27.** Any abnormal extremal belong to $(D^2)^\perp$. Moreover we have that $\lambda(t) \in (D^2)^\perp \setminus (D^3)^\perp$ for all $t \in [0, 1]$ if and only if $\lambda(t)$ satisfies

$$\dot{\lambda}(t) = \vec{H}_0(\lambda(t))$$  

(12.47)

with initial condition $\lambda_0 \in (D^2_q)^\perp \setminus (D^3_q)^\perp$.

**Remark 12.28.** Notice that, as soon as $n > 3$, the set $(D^2_q)^\perp \setminus (D^3_q)^\perp$ is nonempty for an open dense set of $q \in M$. Indeed assume that we have $D^2_q = D^3_q$ for any $q$ in a open neighborhood $O_{q_0}$ of a point $q_0$ in $M$. Then it follows that

$$D^2_{q_0} = D^3_{q_0} = D^4_{q_0} = \ldots$$

and the structure cannot be bracket generating, since $\dim D^i_{q_0} < \dim M$ for every $i > 1$. The case $n = 3$ will be treated separately.

**Proof.** Using that any abnormal extremal belong to the subset $\{h_1(\lambda(t)) = h_2(\lambda(t)) = 0\}$, it is easy to show that an abnormal extremal $\lambda(t)$ satisfies (12.45) if and only if it is an integral curve of the Hamiltonian vector field $\vec{H}_0$.

It remains to prove that a solution of the system

$$\dot{\lambda}(t) = \vec{H}(\lambda(t)), \quad \lambda_0 \in (D^2)^\perp \setminus (D^3)^\perp,$$  

(12.48)

satisfies $\lambda(t) \in (D^2)^\perp \setminus (D^3)^\perp$ for every $t$. First notice that the solution cannot intersect the set $(D^3)^\perp$ since these are equilibrium points of the system (12.48) (since at these points the Hamiltonian has a root of order two).
We are reduced to prove that \((D^2)^\perp\) is an invariant subset for \(\vec{H}\). Hence we prove that the functions \(h_1, h_2, h_{12}\) are constantly zero when computed on the extremal.

To do this we find the differential equation satisfied by these Hamiltonians. Recall that, for any smooth function \(a : T^*M \to \mathbb{R}\) and any solution of the Hamiltonian system \(\lambda(t) = e^{t\vec{H}}\lambda_0\), we have \(\dot{a} = \{H, a\}\). Hence we get

\[
\dot{h}_{12} = \{h_{221}h_1 + h_{112}h_2, h_{12}\}
\]

\[
= \{h_{221}, h_{12}\}h_1 + \{h_{112}, h_{12}\}h_2 + h_{112}h_{221} + h_{212}h_{112} = 0
\]

for some smooth coefficients \(c_1\) and \(c_2\). We see that there exists smooth functions \(a_1, a_2, a_{12}\) and \(b_1, b_2, b_{12}\) such that

\[
\begin{cases}
\dot{h}_1 = a_1h_1 + a_2h_2 + a_{12}h_{12} \\
\dot{h}_2 = b_1h_1 + b_2h_2 + b_{12}h_{12} \\
\dot{h}_{12} = c_1h_1 + c_2h_2
\end{cases}
\]

(12.49)

If we plug the solution \(\lambda(t)\) into the equation of (12.48), i.e. if we consider it as a system of differential equations for the scalar functions \(h_i(t) := h_i(\lambda(t))\), with variable coefficients \(a_i(\lambda(t)), b_i(\lambda(t)), c_i(\lambda(t))\), we find that \(h_1(t), h_2(t), h_{12}(t)\) satisfy a nonautonomous homogeneous linear system of differential equation with zero initial condition, since \(\lambda_0 \in (D^2)^\perp\), i.e.

\[
h_1(\lambda_0) = h_2(\lambda_0) = h_{12}(\lambda_0) = 0.
\]

(12.50)

Hence

\[
h_1(\lambda(t)) = h_2(\lambda(t)) = h_{12}(\lambda(t)) = 0, \quad \forall t.
\]

\[\square\]

We also can prove easily that nice abnormals satisfy the generalized Legendre condition. Recall that if \(\lambda(t)\) is an abnormal extremal, then \(-\lambda(t)\) is also an abnormal extremal.

**Lemma 12.29.** Let \(\lambda(t)\) be a nice abnormal. Then \(\lambda(t)\) or \(-\lambda(t)\) satisfy the generalized Legendre condition.

**Proof.** Let \(u(t)\) be the control associated with the extremal \(\lambda(t)\). It is sufficient to prove that the quadratic form

\[
Q_t : v \mapsto \langle \lambda(t), [f_{u(t)}, f_v] \rangle, \quad v \in \mathbb{R}^2
\]

is non negative definite. We already proved (cf. ??) that the bilinear form

\[
B_t : (v, w) \mapsto \langle \lambda(t), [f_{u(t)}, f_v, f_w] \rangle, \quad v, w \in \mathbb{R}^2
\]

is symmetric. From (12.52) it is easy to see that \(u(t) \in \text{Ker} B_t\) for every \(t\). Hence \(Q_t\) is degenerate for every \(t\). On the other hand if the quadratic form is identically zero we have \(\lambda(t) \in (D^3)^\perp\), which is a contradiction.

Hence the quadratic form has rank 1 and is semi-definite and we can choose \(\pm \lambda_0\) in such a way that (12.51) is positive at \(t = 0\). Since the sign of the quadratic form does not change along the curve (it is continuous and it cannot vanish) we have that it is positive for all \(t\).

\[\square\]
12.5 Optimality of nice abnormal in rank 2 structures

Up to now we proved that every nice abnormal extremal in a rank 2 sub-Riemannian structure automatically satisfies the necessary condition for optimality. Now we prove that actually they are strict local minimizers.

**Theorem 12.30.** Let \( \lambda(t) \) be a nice abnormal extremal and let \( \gamma(t) \) be corresponding abnormal trajectory. Then there exists \( s > 0 \) such that \( \gamma|_{[0,s]} \) is a strict local length minimizer in the \( L^2 \)-topology for the controls (equivalently the \( H^1 \)-topology for trajectories).

**Remark 12.31.** Notice that this property of \( \gamma \) does not depend on the metric but only on the distribution. In particular the value of \( s \) will be independent on the metric structure defined on the distribution.

It follows that, as soon as the metric is fixed, small pieces of nice abnormal are also global minimizers.

Before proving Theorem 12.30 we prove the following technical result.

**Lemma 12.32.** Let \( \Phi : E \to \mathbb{R}^n \) be a smooth map defined on a Hilbert space \( E \) such that \( \Phi(0) = 0 \), where 0 is a critical point for \( \Phi \)

\[
\lambda D_0 \Phi = 0, \quad \lambda \in \mathbb{R}^{n*}, \quad \lambda \neq 0.
\]

Assume that \( \lambda \text{Hess}_0 \phi \) is a positive definite quadratic form. Then for every \( v \) such that \( \langle \lambda, v \rangle < 0 \), there exists a neighborhood of zero \( O \subset E \) such that

\[
\Phi(x) \notin \mathbb{R}^+ v, \quad \forall x \in O, x \neq 0, \quad \mathbb{R}^+ = \{ \alpha \in \mathbb{R}, \alpha > 0 \}.
\]

In particular the map \( \Phi \) is not locally open and \( x = 0 \) is an isolated point on its level set.

**Proof.** In the first part of the proof we build some particular set of coordinates that simplifies the proof, exploiting the fact that the Hessian is well defined independently on the coordinates.

Split the domain and the range of the map as follows

\[
E = E_1 \oplus E_2, \quad E_2 = \text{Ker} \, D_0 \Phi, \quad \mathbb{R}^n = \mathbb{R}^{k_1} \oplus \mathbb{R}^{k_2}, \quad \mathbb{R}^{k_1} = \text{Im} \, D_0 \Phi, \tag{12.53}
\]

where we select the complement \( \mathbb{R}^{k_2} \) in such a way that \( v \in \mathbb{R}^{k_2} \) (notice that by our assumption \( v \notin \mathbb{R}^{k_1} \)). Accordingly to the notation introduced, let us write

\[
\Phi(x_1, x_2) = (\Phi_1(x_1, x_2), \Phi_2(x_1, x_2)), \quad x_i \in E_i, \ i = 1, 2.
\]

Since \( \Phi_1 \) is a submersion by construction, the Implicit function theorem implies that by a smooth change of coordinates we can linearize \( \Phi_1 \) and assume that \( \Phi \) has the form

\[
\Phi(x_1, x_2) = (D_0 \Phi_1(x_1), \Phi_2(x_1, x_2)),
\]

since \( x_2 \in E_2 = \text{Ker} \, D_0 \Phi \). Notice that, by construction of the coordinate set, the function \( x_2 \mapsto \Phi_2(0, x_2) \) coincides with the restriction of \( \Phi \) to the kernel of its differential, modulo its image.
Hence for every scalar function \( a : \mathbb{R}^{k_2} \to \mathbb{R} \) such that \( d_0 a = \lambda \) we have the equality

\[
\lambda \text{Hess}_0 \Phi = \text{Hess}_0 (a \circ \Phi_2(0, \cdot)) > 0
\]

In particular the function \( a \circ \Phi_2(0, y) \) is non negative in a neighborhood of 0.

Assume now that \( \Phi(x_1, x_2) = sv \) for some \( s \geq 0 \). Since \( v \in \mathbb{R}^{k_2} \) it follows that

\[
D_0 \Phi(x_1) = 0 \implies x_1 = 0, \quad \text{and} \quad \Phi_2(0, x_2) = sv.
\]

In particular we have

\[
\frac{d}{ds} \bigg|_{s=0} a(\Phi_2(0, x_2)) = \frac{d}{ds} \bigg|_{s=0} a(sv) = \langle \lambda, v \rangle \leq 0 \quad \implies \quad a(sv) \leq 0 \quad \text{for} \quad s \geq 0
\]

which is a contradiction. \( \square \)

Let \( \lambda(t) \) be an abnormal extremal and let \( \gamma(t) \) be corresponding abnormal trajectory.

\[
\dot{\gamma} = u_1 f_1(\gamma) + u_2 f_2(\gamma). \quad (12.55)
\]

In what follows we always assume that \( \tilde{\gamma} \equiv \{\gamma(t) : t \in [0, 1]\} \) is a smooth one-dimensional submanifold of \( M \), with or without border. Then either the curve \( \gamma \) has no self-intersection or \( \tilde{\gamma} \) is diffeomorphic to \( S^1 \). In both cases we can chose a basis \( f_1, f_2 \) in a neighborhood of \( \tilde{\gamma} \) in such a way that \( \gamma \) is the integral curve of \( f_1 \)

\[
\dot{\gamma} = f_1(\gamma)
\]

Then \( \gamma \) is the solution of \( (12.55) \) with associated control \( \tilde{u} = (1, 0) \). Notice that a change of the frame on \( M \) corresponds to a smooth change of coordinates on the end-point map. With analogous reasoning as in the previous section, we describe the end point map

\[
F : (u_1, u_2) \mapsto \gamma(1)
\]

as the composition

\[
F = e^{f_1} \circ G
\]

where \( G \) is the end point map for the system

\[
\dot{q} = (u_1 - 1)e^{-t f_1} f_1 + u_2 e^{-t f_1} f_2. \quad (12.56)
\]

Since \( e^{-t f_1} f_1 = f_1 \), denoting \( g_t := e^{-t f_1} f_2 \) and defining the primitives

\[
w(t) = \int_0^t (1 - u_1(\tau)) d\tau, \quad v(t) = \int_0^t u_2(\tau) d\tau, \quad (12.57)
\]

we can rewrite the system, whose endpoint map is \( G \), as follows

\[
\dot{q} = -\dot{w} f_1(q) + \dot{v} g_t(q).
\]

The Hessian of \( G \) is computed

\[
\lambda_0 \text{Hess}_0 G(u_1, \dot{v}) = \int_0^1 \langle \lambda_0, [\int_0^t -\dot{w}(\tau) f_1 + \dot{v}(\tau) g_t d\tau, -\dot{w}(t) f_1 + \dot{v}(t) g_t](q_0) \rangle dt. \quad (12.58)
\]

265
Recall that
\[ D_0 G(u_1, \dot{v}) = \int_0^1 -\dot{w}(t)f_1(q_0) + \dot{v}(t)g_t(q_0)dt \]
\[ = -w(1)f_1(q_0) + \int_0^1 \dot{v}(t)g_t(q_0)dt \]
and the condition \( \lambda_0 \in \text{Im} D_0 G^\perp \) is rewritten as
\[ \langle \lambda_0, f_1(q_0) \rangle = \langle \lambda_0, g_t(q_0) \rangle = 0, \quad \forall \ t. \quad (12.59) \]
Notice that since equality (12.59) is valid for all \( t \) then we have that
\[ \langle \lambda_0, \dot{g}_t(q_0) \rangle = \langle \lambda_0, [f_1, g_t](q_0) \rangle = 0, \quad (12.60) \]
Then we can rewrite our quadratic form only as a function of \( \dot{v} \), since all terms containing \( \dot{w} \) disappear
\[ \lambda_0 \text{Hess}_0 G(\dot{v}) = \int_0^1 \langle \lambda_0, [\int_0^t \dot{v}(\tau)g_\tau d\tau, \dot{v}(t)g_t](q_0) \rangle dt \quad (12.61) \]
with the extra condition
\[ \int_0^1 \dot{v}(t)g_t(q_0)dt = w(1)f_1(q_0). \quad (12.62) \]
Now we rearrange these formulas, using integration by parts, rewriting the Hessian as a quadratic form on the space of primitives
\[ v(t) = \int_0^t \dot{v}(\tau)d\tau \]
Using the equality
\[ \int_0^t \dot{v}(\tau)g_\tau d\tau = v(t)g_t - \int_0^t v(\tau)\dot{g}_\tau d\tau \quad (12.63) \]
we have
\[ \lambda_0 \text{Hess}_0 G(\dot{v}) = \int_0^1 \langle \lambda_0, [v(t)g_t, \dot{v}(t)g_t](q_0) \rangle dt \]
\[ - \int_0^1 \langle \lambda_0, [\int_0^t v(\tau)\dot{g}_\tau d\tau, \dot{v}(t)g_t](q_0) \rangle dt \]
The first addend is zero since \([g_t, g_t] = 0\). Exchanging the order of integration in the second term
\[ \int_0^1 \langle \lambda_0, [\int_0^t v(\tau)\dot{g}_\tau d\tau, \dot{v}(t)g_t](q_0) \rangle dt = \int_0^1 \langle \lambda_0, [v(t)\dot{g}_t, \int_t^1 \dot{v}(\tau)g_\tau d\tau](q_0) \rangle dt \]
and then integrating by parts
\[ \int_t^1 \dot{v}(\tau)g_\tau d\tau = v(1)g_1 - v(t)g_t - \int_t^1 v(\tau)\dot{g}_\tau d\tau \]
we get to
\[ \lambda \text{Hess}_0 G(\dot{v}) = \int_0^1 \langle \lambda_0, [\dot{g}_t, g_t](q_0) \rangle v(t)^2 dt \]
\[ + \int_0^1 \langle \lambda_0, [\int_0^t v(\tau)\dot{g}_\tau, v(t)\dot{g}_t - v(1)g_1](q_0) \rangle dt \]  \hspace{1cm} (12.64)

which can also be rewritten as follows
\[ \lambda \text{Hess}_0 G(\dot{v}) = \int_0^1 \langle \lambda_0, [\dot{g}_t, g_t](q_0) \rangle v(t)^2 dt \]
\[ + \int_0^1 \langle \lambda_0, [\int_0^t v(\tau)\dot{g}_\tau, d\tau + v(1)g_1, v(t)\dot{g}_t](q_0) \rangle dt. \]  \hspace{1cm} (12.65)

Moreover, again integrating by parts the extra condition (12.62), we find
\[ \int_0^1 v(t)\dot{g}_t(q_0) dt = -w(1)f_1(q_0) + v(1)g_1(q_0) \]  \hspace{1cm} (12.66)

Remark 12.33. Notice that we cannot plug in the expression (12.66) directly into the formula since this equality is valid only at the point \( q_0 \), while in (12.64) we have to compute the bracket.

Notice that the vectors \( f_1(q_1) \) and \( f_2(q_1) \) are linearly independent, then also
\[ f_1(q_0) = e^{-f_1}(f_1(q_1)), \quad \text{and} \quad g_1(q_0) = e^{-f_1}(f_2(q_1)), \]
are linearly independent. From (12.66) it follows that for every pair \( (w, v) \) in the kernel the following estimates are valid
\[ |w(1)| \leq C\|v\|_{L^2}, \quad |v(1)| \leq C\|v\|_{L^2}. \]  \hspace{1cm} (12.67)

**Theorem 12.34.** Let \( \gamma : [0, 1] \to M \) be an abnormal trajectory and assume that the quadratic form (12.64) satisfies
\[ \lambda_0 \text{Hess}_0 G(\dot{v}) \geq \alpha\|v\|_{L^2}^2. \]  \hspace{1cm} (12.68)

Then the curve is locally minimizer in the \( L^2 \) topology of controls.

Remark 12.35. Notice that the estimate (12.68) depends only on \( v \), while the map \( G \) is a smooth map of \( \dot{v} \) and \( \dot{w} \). Hence Lemma 12.32 does not apply.

Moreover, the statement of Lemma 12.32 violates for the endpoint map, since it is locally open as soon as the bracket generating condition is satisfied (this is equivalent to the Chow-Rashevsky Theorem). Moreover the final point of the trajectory is never isolated in the level set.

What we are going to use is part of the proof of this Lemma, to show that the statements holds for the restriction of the endpoint map to some subset of controls.

**Proof of Theorem 12.34.** Our goal is to prove that there are no curves shorter than \( \gamma \) that join \( q_0 \) to \( q_1 = \gamma(1) \).

To this extent we consider the restriction of the endpoint map to the set of curves that are shorter or have the same length than the original curve. Hence we need to fix some sub-Riemannian structure on \( M \).
We can then assume the orthonormal frame $f_1, f_2$ to be fixed and that the length of our curve is exactly 1 (we can always dilate all the distances on our manifold and the local optimality of the curve is not affected).

The set of curves of length less or equal than 1 can be parametrized, using Lemma 3.15, by the set
\[
\{(u_1, u_2) | u_1^2 + u_2^2 \leq 1\}
\]
Following the notation (12.57), notice that
\[
\{(u_1, u_2) | u_1^2 + u_2^2 \leq 1\} \subset \{(w, v) | \dot{w} \geq 0\}.
\]
We want to show that, for some function $a \in C^\infty(M)$ such that $d_q a = \lambda \in \text{Im} D_0 F^\perp$, we have
\[
a \circ F|_{D_0(\dot{w}, \dot{v})} = \lambda \text{Hess}_0 F(\dot{w}, \dot{v}) + R(w, v),
\]
where
\[
\frac{R(w, v)}{\|v\|^2} \rightarrow 0 \quad (12.69)
\]
in the domain
\[
D = \{(\dot{w}, \dot{v}) \in \text{Ker} D_0 F, \dot{w} \geq 0\}
\]
Indeed if we prove (12.69), we have that the point $(\dot{w}, \dot{v}) = (0, 0)$ is locally optimal for $F$. This means that the curve $\gamma$, i.e. the curve associated to controls $u_1 = 1, u_2 = 0$, is also locally optimal.

Using the identity
\[
\exp \int_0^t \dot{v}(\tau) f_2 d\tau = e^{v(t)f_2}
\]
and applying the variations formula (6.22) to the endpoint map $F$ we get
\[
F(\dot{w}, \dot{v}) = q_0 \circ \exp \int_0^1 (1 - \dot{w}(t)) f_1 + \dot{v}(t) f_2 dt
\]
\[
= q_0 \circ \exp \int_0^1 (1 - \dot{w}(t)) e^{-v(t)f_2} f_1 dt \circ e^{v(1)f_2}
\]
Hence we can express the endpoint map as a smooth function of the pair $(\dot{w}, \dot{v})$.

Now, to compute (12.69), we can assume that the function $a$ is constant on the trajectories of $f_2$ (since we only fix its differential at one point) so that
\[
e^{v(1)f_2} \circ a = a
\]
which simplifies our estimates:
\[
a \circ F(\dot{w}, \dot{v}) = q_0 \circ \exp \int_0^1 (1 - \dot{w}(t)) e^{-v(t)f_2} f_1 dt a
\]
Writing
\[
(1 - \dot{w}(t)) e^{-v(t)f_2} f_1 = f_1 + X^0(v(t)) + \dot{w}(t) X^1(v(t)) \quad (12.70)
\]
and using the variation formula (6.23), setting $Y^i_t = e^{(t-1)f_1} X^i$ for $i = 0, 1$, we get (recall that $q_1 = q_0 \circ e^{f_1(q_0)}$)
\[
a \circ F(\dot{w}, \dot{v}) = q_1 \circ \exp \int_0^1 Y^0_t(v(t)) + \dot{w}(t) Y^1_t(v(t)) dt a, \quad Y^0_t(0) = Y^1_t(0) = 0
\]
Expanding the chronological exponential we find that
\[
268
\]
(a) the zero order term vanish since $Y_0^0(0) = Y_1^1(0) = 0$,

(b) all first order terms vanish since the vector fields $f_1$ and $[f_1, f_2]$ spans the image of the differential (hence are orthogonal to $\lambda = d_qa$)

(c) the second order terms are in the Hessian, since our domain $D$ is contained in the kernel of the differential

In other words it remains to show that every term in $v, w$ of order greater or equal than 3 in the expansion can be estimated with $o(\|v\|^2)$.

Let us prove first the claim for monomial of order three:

\[
\int_0^1 \dot{w}(t)v^2(t)dt = o(\|v\|^2), \quad \int_0^1 \dot{w}(t)\int_0^t \dot{w}(\tau)v(\tau)d\tau dt = o(\|v\|^2)
\]

\[
\int_0^1 \int_0^t \dot{w}(\tau)v(\tau)d\tau d\tau dt = o(\|v\|^2)
\]

Using that $\dot{w} \geq 0$, which is the key assumption, and the fact that $(\dot{w}, \dot{v}) \in \text{Ker} D_0F$, which gives the estimates (12.67), we compute

\[
\left| \int_0^1 \dot{w}(t)v^2(t)dt \right| \leq \int_0^1 |\dot{w}(t)|v^2(t)dt \\
= \int_0^1 \dot{w}(t)v^2(t)dt \\
= w(1)v^2(1) - \int_0^1 w(t)v(t)\dot{v}(t)dt \\
\leq \|v\|^3 + \varepsilon\|v\|^2,
\]

where we estimate for the second term follows from

\[
\left| \int_0^1 w(t)v(t)\dot{v}(t)dt \right| \leq \max w(t) \left| \int_0^1 v(t)\dot{v}(t)dt \right| \\
\leq w(1)\|v\|\|\dot{v}\| \\
\leq C\|\dot{v}\|\|v\|^2
\]

The second integral can be rewritten

\[
\int_0^1 \dot{w}(t)\int_0^t \dot{w}(\tau)v(\tau)d\tau dt = w(1)\int_0^1 \dot{w}(t)v(t)dt - \int_0^1 w(t)v(t)\dot{w}(t)dt
\]

and then we estimate

\[
\left| \int_0^1 \dot{w}(t)\int_0^t \dot{w}(\tau)v(\tau)d\tau dt \right| \leq 2|w(1)|\int_0^1 v(t)\dot{w}(t)dt \\
\leq C\|\dot{w}\|\|v\|^2
\]

\footnote{where $o(\|v\|^2)$ have the same meaning as in (12.69).}
Finally, the last integral is very easy to estimate using the equality
\[
\int_0^1 \dot{w}(t) \int_0^t \dot{w}(\tau) \int_0^\tau \dot{w}(s) ds d\tau dt = \frac{1}{6} \int_0^1 \dot{w}(t)^3 dt \leq C\|\dot{w}\|\|v\|^2
\]

Starting from these estimates it is easy to show that any mixed monomial of order greater than three satisfies these estimates as well.

Applying these results to a small piece of abnormal trajectory we can prove that small pieces of nice abnormalities are minimizers.

**Proof of Theorem 12.30.** If we apply the arguments above to a small piece \(\gamma_s = \gamma|_{[0,s]}\) of the curve \(\gamma\) it is easy to see that the Hessian rescale as follows,
\[
\lambda_0 \text{Hess}_0 G_s(v) = \int_0^s \langle \lambda_0, [g_t, \dot{g}_t](q_0) \rangle v(t)^2 dt \\
+ \int_0^s \langle \lambda_0, \left[ \int_0^t v(\tau) \dot{g}_\tau d\tau, v(t) \dot{g}_t - v(s) g_s(q_0) \right] \rangle dt
\]

Since the generalized Legendre condition ensures\(^3\) that (see also Lemma 12.29)
\[
\langle \lambda_0, [g_t, \dot{g}_t](q_0) \rangle \geq C > 0
\]
then the norm
\[
\|v\|_g = \left( \int_0^s \langle \lambda_0, [g_t, \dot{g}_t](q_0) \rangle v(t)^2 dt \right)^{1/2}
\]
(12.71)
is equivalent to the standard \(L^2\)-norm. Hence the Hessian can be rewritten as
\[
\lambda \text{Hess}_0 G_s(v) = \|v\|_g + \langle Tv, v \rangle
\]
where \(T\) is a compact operator in \(L^2\) of the form
\[
(Tv)(t) = \int_0^s K(t, \tau) v(\tau) d\tau
\]
Since \(\|T\|^2 = \|K\|^2_{L^2} \to 0\) for \(s \to 0\), it follows that the Hessian is positive definite for small \(s > 0\).

\[\square\]

**12.6 Conjugate points along abnormalities**

In this section, we give an effective way to check the inequality (12.68) that implies local minimality of the nice abnormal geodesic according to Theorem 12.34.

\(^3\) it is semidefinite and we already know that \(f_1\) is in the kernel
We define \( Q_1(v) := \lambda \text{Hess}_0 G(\hat{v}) \). Quadratic form \( Q_1 \) is continuous in the topology defined by the norm \( \|v\|_{L_2} \). The closure of the domain of \( Q_1 \) in this topology is the space
\[
D(Q_1) = \left\{ v \in L_2[0,1] : \int_0^1 v(t)\hat{g}_t(q_0)\ dt \in \text{span}\{f_1(q_0), g_1(q_0)\} \right\}.
\]
The extension of \( Q_1 \) to this closure is denoted by the same symbol \( Q_1 \). We set:
\[
l(t) = \langle \lambda_0, [\hat{g}_t, g_t](q_0) \rangle, \quad X_t = v_1 g_1 + \int_0^t v(\tau)\hat{g}_\tau\ d\tau\]
and we rewrite the form \( Q_1 \) in these more compact notations:
\[
Q_1(v) = \int_0^1 l(t)v(t)^2\ dt + \int_0^1 \langle \lambda_0, [X_t, \dot{X}_t](q_0) \rangle\ dt,
\]
\[
\dot{X}_t = v(t)\hat{g}_t, \quad X_1 \wedge g_1 = 0, \quad X_0(q_0) \wedge f_1(q_0) = 0.
\]
Moreover, we introduce the family of quadratic forms \( Q_s \), for \( 0 < s \leq 1 \), as follows
\[
Q_s(v) := \int_0^s l(t)v(t)^2\ dt + \int_0^s \langle \lambda_0, [X_t, \dot{X}_t](q_0) \rangle\ dt,
\]
\[
\dot{X}_t = v(t)\hat{g}_t, \quad X_s \wedge g_s = 0, \quad X_0(q_0) \wedge f_1(q_0) = 0.
\]
(1)
Recall that \( l(t) \) is a strictly positive continuous function. In particular, \( \int_0^1 l(t)v(t)^2\ dt \) is the square of a norm of \( v \) that is equivalent to the standard \( L_2 \)-norm. Next statement is proved by the same arguments as Proposition 8.40. We leave details to the reader.

**Proposition 12.36.** The form \( Q_1 \) is positive definite if and only if \( \ker Q_s = 0, \ \forall s \in (0,1] \).

**Definition 12.37.** A time moment \( s \in (0,1] \) is called conjugate to 0 for the abnormal geodesic \( \gamma \) if \( \ker Q_s \neq 0 \).

We are going to characterize conjugate times in terms of an appropriate “Jacobi equation”.

Let \( \xi_t \in T_{\lambda_0}(T^*M) \) and \( \zeta_t \in T_{\lambda_0}(T^*M) \) be the values at \( \lambda_0 \) of the Hamiltonian lifts of the vector fields \( f_1 \) and \( g_t \). Recall that the Hamiltonian lift of a field \( f \in \text{Vec}M \) is the Hamiltonian vector field associated to the Hamiltonian function \( \lambda \mapsto \langle \lambda, f(q) \rangle, \ \lambda \in T^*_q M, \ q \in M \). We have:
\[
Q_s(v) = \int_0^s l(t)v(t)^2\ dt + \int_0^s \sigma(x(t), \dot{x}(t))\ dt,
\]
\[
\dot{x}(t) = v(t)\hat{\xi}_t, \quad x(s) \wedge \zeta_s = 0, \quad \pi_s x(0) \wedge \pi_s \xi_1 = 0,
\]
where \( \sigma \) is the standard symplectic product on \( T_{\lambda_0}(T^*M) \) and \( \pi : T^*M \rightarrow M \) is the standard projection. Moreover
\[
l(t) = \sigma(\hat{\zeta}_t, \zeta_t), \quad 0 \leq t \leq 1.
\]
(12.73)
Let \( E = \text{span}\{\xi_t, \zeta_t, 0 \leq t \leq 1\} \). We use only the restriction of \( \sigma \) to \( E \) in the expression of \( Q_s \) and we are going to get rid of unnecessary variables. Namely, we set: \( \Sigma \triangleq E/(\ker \sigma|_E) \).
Lemma 12.38. \( \dim \Sigma \leq 2 (\dim \text{span}\{f_1(q_0), g_t(q_0), 0 \leq t \leq 1\} - 1) \).

Proof. Dimension of \( \Sigma \) is equal to twice the codimension of a maximal isotropic subspace of \( \sigma |E \). We have: \( \sigma(\xi_1, \zeta_t) = \langle \lambda_0, [f_1, g_t](q_0) \rangle = 0 \), \( \forall t \in [0,1] \), hence \( \xi_1 \in \ker \sigma|E \). Moreover, \( \pi_s(E) = \text{span}\{f_1(q_0), g_t(q_0), 0 \leq t \leq 1\} \) and \( E \cap \ker \pi_s \) is an isotropic subspace of \( \sigma|E \).

We denote by \( \zeta_t \in \Sigma \) the projection of \( \zeta_t \) to \( \Sigma \) and by \( \Pi \subset \Sigma \) the projection of \( E \cap \ker \pi_s \). Note that the projection of \( \xi_1 \) to \( \Sigma \) is 0; moreover, equality (12.73) implies that \( \zeta_s \neq 0, \forall t \in [0,1] \). The final expression of \( Q_s \) is as follows:

\[
Q_s(v) = \int_0^s l(t)v(t)^2\,dt + \int_0^s \sigma(x(t), \dot{x}(t))\,dt,
\]

\[
\dot{x}(t) = v(t)\zeta_s, \quad x(s) \wedge \zeta_s = 0, \quad x(0) \in \Pi.
\]

We have: \( v \in \ker Q_s \) if and only if

\[
\int_0^s \left( l(t)v(t) + \sigma(x(t), \dot{x}(t)) \right) w(t)\,dt = 0,
\]

for any \( w(\cdot) \) such that

\[
\int_0^s \zeta_s w(t)\,dt \in \Pi + \mathbb{R}\zeta_s.
\]

We obtain that \( v \in \ker Q_s \) if and only if there exists \( \nu \in \Pi^\perp \cap \zeta_s^\perp \) such that

\[
l(t)v(t) + \sigma(x(t), \dot{x}(t)) = \sigma(\nu, \dot{x}(t)), \quad 0 \leq t \leq s.
\]

We set \( y(t) = x(t) - \nu \) and obtain the following:

Theorem 12.39. A time moment \( s \in (0,1] \) is conjugate to 0 if and only if there exists a nontrivial solution of the equation

\[
l(t)\dot{y} = \sigma(\zeta_s, y)\dot{\zeta}_s
\]

that satisfy the following boundary conditions:

\[\exists \nu \in \Pi^\perp \cap \zeta_s^\perp \text{ such that } (y(s) + \nu) \wedge \zeta_s = 0, \quad (y(0) + \nu) \in \Pi.\]

Remark 12.40. Notice that identity (12.73) implies that \( y(t) = \zeta_s \) for \( t \in [0,1] \) is a solution to the equation (12.74). However this solution may violate the boundary conditions.

Let us consider the special case: \( \dim \text{span}\{f_1(q_0), g_t(q_0), 0 \leq t \leq 1\} = 2 \); this is what we automatically have for abnormal geodesics in a 3-dimensional sub-Riemannian manifold. In this case, \( \dim E = 2, \dim \Pi = 1 \); hence \( \Pi^\perp = \Pi, \zeta_s^\perp = \mathbb{R}\zeta_s \) and \( \Pi^\perp \cap \zeta_s^\perp = 0 \). Then \( \nu \) in the boundary conditions (12.75) must be 0 and \( y(s) = c\zeta_s \), where \( c \) is a nonzero constant. Hence \( y(t) = c\zeta_s \) for \( 0 \leq t \leq 1 \) and \( y(0) = c\zeta_0 = \Pi \). We obtain:

Corollary 12.41. If \( \dim \text{span}\{f_1(q_0), g_t(q_0), 0 \leq t \leq 1\} = 2 \), then the segment \([0,1]\) does not contain conjugate time moments and assumption of Theorem 12.34 is satisfied.

We can apply this corollary to the isoperimetric problem studied in Section 4.4.2. Abnormal geodesics correspond to connected components of the zero locus of the function \( b \) (see notations in Sec. 4.1.2). All these abnormal geodesics are nice if and only if zero is a regular value of \( b \). Take a compact connected component of \( b^{-1}(0) \); this is a smooth closed curve. Our corollary together with Theorem 12.34 implies that this closed curve passed once, twice, three times or arbitrary number of times is a locally optimal solution of the isoperimetric problem. Moreover, this is true for any Riemannian metric on the surface \( M \)!
12.6.1 Abnormals in dimension 3

Nice abnormals for the isoperimetric problem on surfaces

Recall the isoperimetric problem: given two points $x_0, x_1$ on a 2-dimensional Riemannian manifold $N$, a 1-form $\nu \in \Lambda^1 N$ and $c \in \mathbb{R}$, we have to find (if it exists) the minimum:

$$\min \{ \ell(\gamma), \gamma(0) = x_0, \gamma(T) = x_1, \int_{\gamma} \nu = c \}$$  \hspace{1cm} (12.76)

As shown in Section 4.4.2, this problem can be reformulated as a sub-Riemannian problem on the extended manifold

$$M = N \times \mathbb{R} = \{(x,y), x \in N, y \in \mathbb{R}\},$$

where the sub-Riemannian structure is defined by the contact form

$$\mathcal{D} = \text{Ker} (dy - \nu)$$

and the sub-Riemannian length of a curve coincides with the Riemannian length of its projection on $N$. If we write $d\nu = b \, dV$, where $b$ is a smooth function and $dV$ denote the Riemannian volume on $N$, we have that the Martinet surface is defined by the cylinder

$$\mathcal{M} = \mathbb{R} \times b^{-1}(0),$$

where, generically, the set $b^{-1}(0)$ is a regular level of $b$.

Since the distribution is well behaved with the projection on $N$ by construction, it follows that the distribution is always transversal to the Martinet surface and all abnormal are nice, since $\mathcal{D}_q^3 = T_q M$ for all $q$.

Thus the projection of abnormal geodesics on $N$ are the connected components of the set $b^{-1}(0)$ and we can recover the whole abnormal extremal integrating the 1-form $\nu$ to find the missing component. In other words the abnormal extremals are spirals on $\mathcal{M}$ with step equal to $\int_A d\nu$, (if $d\nu$ is the volume form on $N$, it coincide with the area of the region $A$ inside the curve defined on $N$ by the connected component of $b^{-1}(0)$).

**Corollary 12.42.** Let $M$ be a sub-Riemannian manifold, dim $M = 3$, and let $\gamma : [0,1] \to M$ be a nice abnormal geodesic. Then $\gamma$ is a strict local minimizer for the $L^2$ control topology, for any metric.

**Remark 12.43.** Notice that we do not require that the curve does not self-intersect since in the 3D case this is automatically guaranteed by the fact that nice abnormal are integral curves of a smooth vector fields on $M$.

A non nice abnormal extremal

In this section we give an example of non nice (and indeed not smooth) abnormal extremal.

Consider the isoperimetric problem on $\mathbb{R}^2 = \{(x_1, x_2), x_i \in \mathbb{R}\}$ defined by the 1-form $\nu$ such that

$$d\nu = x_1 x_2 \, dx_1 \, dx_2.$$
Here \( b(x_1, x_2) = x_1x_2 \) and the set \( b^{-1}(0) \) consists of the union of the two axes, with moreover \( dB|_0 = 0 \).

Let us fix \( \bar{x}_1, \bar{x}_2 > 0 \) and consider the curve joining \((0, \bar{x}_2)\) and \((\bar{x}_1, 0)\) that is the union of two segment contained in the coordinate axes

\[
\gamma: [-\bar{x}_2, \bar{x}_1] \to \mathbb{R}^2, \quad \gamma(t) = \begin{cases} (0, -t), & t \in [-\bar{x}_2, 0], \\ (t, 0), & t \in [0, \bar{x}_1]. \end{cases}
\]

**Proposition 12.44.** The curve \( \gamma \) is a projection of an abnormal extremal that is not a length minimizer.

**Proof of Proposition 12.44.** Let us built a family of “variations” \( \gamma_{\varepsilon, \delta} \) of the curve \( \gamma \) defined as in Figure 12.1. Namely in \( \gamma_{\varepsilon, \delta} \) we cut a corner of size \( \varepsilon \) at the origin and we turn around a small circle of radius \( \delta \) before reaching the endpoint. Denoting by \( D_\varepsilon \) and \( D_\delta \) the two region enclosed by the curve it is easy to see that the isoperimetric condition rewrites as follows

\[
0 = \int_{\gamma_{\varepsilon, \delta}} d\nu = \int_{D_\varepsilon} d\nu - \int_{D_\delta} d\nu
\]

It is then easy using that \( d\nu = x_1x_2dx_1dx_2 \) to show that there exists \( c_1, c_2 > 0 \) such that

\[
\int_{D_\varepsilon} d\nu = c_1\varepsilon^4, \quad \int_{D_\delta} d\nu = c_2\delta^3
\]

while

\[
\ell(\gamma_{\varepsilon, \delta}) - \ell(\gamma) = 2\pi\delta - (2 - \sqrt{2})\varepsilon \quad (12.77)
\]

Choosing \( \varepsilon \) in such a way that \( c_1\varepsilon^4 = c_2\delta^3 \) it is an easy exercise to show that the quantity \( (12.77) \) is negative when \( \delta > 0 \) is very small.

**Remark 12.45.** If you consider some plane curve \( \tilde{\gamma} \) that is a projection of a normal extremal having the same endpoint \( \gamma \) and contained in the set \( \{(x_1, x_2) \in \mathbb{R}^2, x_1 > 0, x_2 > 0\} \), then \( \tilde{\gamma} \) must have self intersections. Indeed it is easy to see that if it is not the case then the isoperimetric condition

\[
\int_{\tilde{\gamma}} d\nu = 0
\]

cannot be satisfied.

It is still an open problem to find which is the length minimizer joining these two points. We know that it should be a projection of a normal extremal (hence smooth) but for instance we do not know how many self-intersection it has.

### 12.6.2 Higher dimension

Now consider another important special case that is typical if dimension of the ambient manifold is greater than 3. Namely, assume that, for some \( k \geq 2 \), the vector fields

\[
f_1, f_2, (\text{ad}f_1)f_2, \ldots, (\text{ad}f_1)^{k-1}f_2 \quad (12.78)
\]
are linearly independent in any point of a neighborhood of our nice abnormal geodesic \( \gamma \), while \((\text{adj}_1)^k f_2\) is a linear combination of the vector fields \((12.78)\) in any point of this neighborhood; in other words,

\[
(\text{adj}_1)^k f_2 = \sum_{i=0}^{k-1} a_i \text{adj}_1^i f_2 + \alpha f_1,
\]

where \(a_i, \alpha\) are smooth functions. In this case, all closed to \( \gamma \) solutions of the equation \( \dot{q} = f_1(q) \) are abnormal geodesics.

A direct calculation based on the fact that \( \langle \lambda_t, (\text{adj}_1^i) f_2(\gamma(t)) \rangle = 0, \) \( 0 \leq t \leq 1 \), gives the identity:

\[
\zeta_t^{(k)} = \sum_{i=0}^{k-1} a_i(\gamma(t)) \zeta^{(i)} + \alpha(\gamma(t)) \xi_1, \quad 0 \leq t \leq 1. \tag{12.79}
\]

Identity \((12.79)\) implies that \( \dim E = k \) and \( \Pi = 0 \). The boundary conditions \((12.75)\) take the form:

\[
y(0) \in \zeta_0, \quad (y(s) - y(0)) \wedge \zeta_s = 0. \tag{12.80}
\]

The characterization of conjugate points is especially simple and geometrically clear if the ambient manifold has dimension 4. Let \( \Delta \) be a rank 2 equiregular distribution in a 4-dimensional manifold (the Engel distribution). Then abnormal geodesics form a 1-foliation of the manifold and condition \((12.78)\) is satisfied with \( k = 2 \). Moreover, \( \dim E = 3 \), \( \dim \Sigma = 2 \) and \( \zeta_s = \mathbb{R} \zeta_0 \). Recall that \( y(t) = \zeta_s \), \( 0 \leq t \leq s \), is a solution to \((12.74)\). Hence boundary conditions \((12.80)\) are equivalent to the condition

\[
\zeta_s \wedge \zeta_0 = 0. \tag{12.81}
\]

It is easy to re-write relation \((12.81)\) in the intrinsic way without special notations we used to simplify calculations. We have the following characterization of conjugate times.

**Lemma 12.46.** A time moment \( t \) is conjugate to \( 0 \) for the abnormal geodesic \( \gamma \) if and only if

\[
e^t f_1 D_{\gamma(0)} = D_{\gamma(t)}.
\]

The flow \( e^t f_1 \) preserves \( D^2 \) and \( f_1 \) but it does not preserve \( D \). The plane \( e^t f_1 D \) rotates around the line \( \mathbb{R} f_1 \) inside \( D^2 \) with a nonvanishing angular velocity. Conjugate moment is a moment when the plane makes a complete revolution. Collecting all the information we obtain:
Lemma 12.49. We need the following technical lemma.

\[ \Phi^* \text{length one.} \]

Proof.

Theorem 12.48. Let \( D \) be the Engel distribution, \( f_1 \) be a horizontal vector field such that \( [f_1, D^2] = D^2 \) and \( \dot{\gamma} = f_1(\gamma) \). Then \( \gamma \) is an abnormal geodesic. Moreover

(i) if \( e_t^{f_1} D\gamma(0) \neq D\gamma(t), \forall t \in (0, 1), \) then \( \gamma \) is a local length minimizer for any sub-Riemannian structure on \( D \)

(ii) If \( e_t^{f_1} D\gamma(0) = D\gamma(t) \) for some \( t \in (0, 1) \) and \( \gamma \) is not a normal geodesic, then \( \gamma \) is not a local length minimizer.

12.7 Equivalence of local minimality

Now we prove that, under the assumption that our trajectory is smooth, it is equivalent to be locally optimal in the \( H^1 \)-topology or in the uniform topology for the trajectories.

Recall that a curve \( \tilde{\gamma} \) is called a \( C^0 \)-local length-minimizer if \( \ell(\tilde{\gamma}) \leq \ell(\gamma) \) for every curve \( \gamma \) that is \( C^0 \)-close to \( \gamma \) satisfying the same boundary conditions, while it is called a \( H^1 \)-local length-minimizer if \( \ell(\tilde{\gamma}) \leq \ell(\gamma) \) for every curve \( \gamma \) such that the control \( u \) corresponding to \( \gamma \) is close in the \( L^2 \) topology to the control \( \tilde{u} \) associated with \( \tilde{\gamma} \) and \( \gamma \) satisfies the same boundary conditions.

Any \( C^0 \)-local minimizer is automatically a \( H^1 \)-local minimizer. Indeed it is possible to show that for every \( v, w \) in a neighborhood of a fixed control \( u \) there exists a constant \( C > 0 \) such that

\[ |\gamma_v(t) - \gamma_w(t)| \leq C\|v - w\|_{L^2}, \forall t \in [0, T], \]

where \( \gamma_v \) and \( \gamma_w \) are the trajectories associated to controls \( v, w \) respectively.

Theorem 12.48. Let \( M \) be a sub-Riemannian structure that is the restriction to \( D \) of a Riemannian structure \((M, g)\). Assume \( \tilde{\gamma} \) is of class \( C^1 \) and has no self intersections. If \( \tilde{\gamma} \) is a (strict) local minimizer in the \( L^2 \) topology for the controls then \( \tilde{\gamma} \) is also a (strict) local minimizer in the \( C^0 \) topology for the trajectories.

Proof. Since \( \tilde{\gamma} \) has no self intersections, we can look for a preferred system of coordinates on an open neighborhood \( \Omega \) in \( M \) of the set \( V = \{\tilde{\gamma}(t) : t \in [0, 1]\} \). For every \( \varepsilon > 0 \), define the cylinder in \( \mathbb{R}^n = \{(x, y) \in \mathbb{R}^n : x \in \mathbb{R}, y \in \mathbb{R}^{n-1}\} \) as follows

\[ I_\varepsilon \times B_\varepsilon^{n-1} = \{(x, y) \in \mathbb{R}^n : x \in [-\varepsilon, 1 + \varepsilon], y \in \mathbb{R}^{n-1}, |y| < \varepsilon\}, \quad (12.82) \]

We need the following technical lemma.

Lemma 12.49. There exists \( \varepsilon > 0 \) and a coordinate map \( \Phi : I_\varepsilon \times B_\varepsilon^{n-1} \to \Omega \) such that for all \( t \in [0, 1] \)

(a) \( \Phi(t, 0) = \tilde{\gamma}(t) \),

(b) the Riemannian metric \( \Phi^* g \) is the identity matrix at \((t, 0)\), i.e., along \( \tilde{\gamma} \).

Proof of the Lemma. As in the proof of Theorem ??, for every \( \varepsilon > 0 \) we can find coordinates in the cylinder \( I_\varepsilon \times B_\varepsilon^{n-1} \) such that, in these coordinates, our curve \( \tilde{\gamma} \) is rectified \( \tilde{\gamma}(t) = (t, 0) \) and has length one.

Our normalization of the curve \( \tilde{\gamma} \) implies that for the matrix representing the Riemannian metric \( \Phi^* g \) in these coordinates satisfies

\[ \Phi^* g = \begin{pmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \end{pmatrix}, \quad \text{with} \quad G_{11}(x, 0) = 1 \]
where $G_{ij}$, for $i, j = 1, 2$, are the blocks of $\Phi^*g$ corresponding to the splitting $\mathbb{R}^n = \mathbb{R} \times \mathbb{R}^{n-1}$ defined in (12.82). For every point $(x, 0)$ let us consider the orthogonal complement $T(x, 0)$ of the tangent vector $e_1 = \partial_x$ to $\bar{\gamma}$ with respect to $G$. It can be written as follows (in this proof $\langle \cdot, \cdot \rangle$ is the Euclidean product in $\mathbb{R}^n$)

$$T(x, 0) = \{(v_x, y), y \in \mathbb{R}^{n-1}\}$$

for some family of vectors $v_x \in \mathbb{R}^{n-1}$, depending smoothly with respect to $x$. Let us consider now the smooth change of coordinates

$$\Psi : \mathbb{R}^n \to \mathbb{R}^n, \quad \Psi(x, y) = (x - \langle v_x, y \rangle, y)$$

Fix $\varepsilon > 0$ small enough such that the restriction of $\Psi$ to $I_\varepsilon \times B^1_{\varepsilon}$ is invertible. Notice that this is possible since

$$\det D\Psi(x, y) = 1 - \left\langle \frac{\partial v_x}{\partial x}, y \right\rangle.$$ 

It is not difficult to check that, in the new variables (that we still denote by the same symbol), one has

$$G(x, 0) = \begin{pmatrix} 1 & 0 \\ 0 & M(x, 0) \end{pmatrix},$$

where $M(x, 0)$ is a positive definite matrix for all $x \in I_\varepsilon$. With a linear change of coordinates in the $y$ space

$$(x, y) \mapsto (x, M(x, 0)^{1/2}y)$$

we can finally normalize the matrix in such a way that $G(x, 0) = \text{Id}$ for all $x \in I_\varepsilon$. \hfill \Box

We are now ready to prove the theorem. We check the equivalence between the two notions of local minimality in the coordinate set, denoted $(x, y)$, defined by the previous lemma. Notice that the notion of local minimality is independent on the coordinates.

Given an admissible curve $\gamma(t) = (x(t), y(t))$ contained in the cylinder $I_\varepsilon \times B^1_{\varepsilon}$ and satisfying $\gamma(0) = (0, 0)$ and $\gamma(1) = (1, 0)$ and denoting the reference trajectory $\bar{\gamma}(t) = (t, 0)$ we have that

$$\|\gamma - \bar{\gamma}\|^2_{H^1} = \int_0^1 |\dot{x}(t) - 1|^2 + |\dot{y}(t)|^2 dt$$

$$= \int_0^1 |\dot{x}(t)|^2 + |\dot{y}(t)|^2 dt - 2 \int_0^1 \dot{x}(t) dt + 1$$

$$= \int_0^1 |\dot{x}(t)|^2 + |\dot{y}(t)|^2 dt - 1$$

where we used that $x(0) = 0$ and $x(1) = 1$ since $\gamma$ satisfies the boundary conditions. If we denote by

$$J(\gamma) = \int_0^1 \langle G(\gamma(t))\dot{\gamma}(t), \dot{\gamma}(t) \rangle dt, \quad J_e(\gamma) = \int_0^1 |\dot{x}(t)|^2 + |\dot{y}(t)|^2 dt$$

(12.83) respectively the energy of $\gamma$ and the “Euclidean” energy, we have $\|\gamma - \bar{\gamma}\|^2_{H^1} = J_e(\gamma) - 1$ and the $H^1$-local minimality can be rewritten as follows:

Indeed it is easily checked that $v_x = -G^2_{21}(x, 0)$, where $G^2_{21}$ denotes the first column of the $(n - 1) \times (n - 1)$ matrix $G_{21}$.
there exists $\varepsilon > 0$ such that for every $\gamma$ admissible and $J_e(\gamma) \leq 1 + \varepsilon$ one has $J(\gamma) \geq 1$.

Next we build the following neighborhood of $\bar{\gamma}$: for every $\delta > 0$ define $A_\delta$ as the set of admissible curves $\gamma(t) = (x(t), y(t))$ in $I_\varepsilon \times B^n_{\delta \varepsilon}^{-1}$ such that the dilated curve $\gamma_\delta(t) = (x(t), \frac{1}{\delta}y(t))$ is still contained in the cylinder. This implies that in particular that $\gamma$ is contained in $I_\varepsilon \times B^n_{\delta \varepsilon}^{-1}$. Notice that $A_\delta \subset A_{\delta'}$ whenever $\delta < \delta'$. Moreover, every curve that is $\varepsilon \delta$ close to $\bar{\gamma}$ in the $C^0$-topology is contained in $A_\delta$.

It is then sufficient to prove that, for $\delta > 0$ small enough, for every $\gamma \in A_\delta$ one has $\ell(\gamma) \geq \ell(\bar{\gamma})$. Indeed it is enough to check that $J(\gamma) \geq J(\bar{\gamma})$. Let us consider two cases

(i) $\gamma \in A_\delta$ and $J_e(\gamma) \leq 1 + \varepsilon$. In this case $(\ast)$ implies that $J(\gamma) \geq 1$.

(ii) $\gamma \in A_\delta$ and $J_e(\gamma) > 1 + \varepsilon$. In this case we have $G(x,0) = \text{Id}$ and, by smoothness of $G$, we can write for $(x,y) \in I_\varepsilon \times B^n_{\delta \varepsilon}^{-1}$ and $\delta \to 0$

$$\langle G(x,y)v,v \rangle = (1 + O(\delta)) \langle v,v \rangle,$$

where $O(\delta)$ is uniform with respect to $(x,y)$. Since $\gamma \in A_\delta$ implies that $\gamma$ is contained in $I_\varepsilon \times B^n_{\delta \varepsilon}^{-1}$ we can deduce for $\delta \to 0$

$$J(\gamma) = J_e(\gamma)(1 + O(\delta)) \geq (1 + \varepsilon)(1 + O(\delta))$$

and one can choose $\tilde{\delta} > 0$ small enough such that the last quantity is strictly bigger than one.

This proves that there exists $\tilde{\delta} > 0$ such every admissible curve $\gamma \in A_\tilde{\delta}$ is longer than $\bar{\gamma}$.

Remark 12.50. Notice that this result implies in particular Theorem 4.57 since normal extremals are always smooth. Nevertheless, the argument of Theorem 4.57 can be adapted for more general coercive functional (see [3]), while this proof use specific estimates that hold only for our explicit cost (i.e., the distance).

We proved in Theorem 12.27 that nice abnormals are smooth and cannot have self-intersections, being solution of a smooth Hamiltonian system. Thus we can combine Theorem 12.30 and 12.48 and obtain the following result.

**Corollary 12.51.** Let $\gamma(t)$ be a nice abnormal trajectory. Then there exists $s > 0$ such that $\gamma|_{[0,s]}$ is a strict local length minimizer in the $C^0$-topology.
Chapter 13

Some model spaces

13.1 Carnot groups of step 2

13.1.1 Heisenberg
13.1.2 (3, 6)
13.1.3 \((k, k(k + 1)/2)\)

13.2 Other nilpotent structures

13.2.1 Grushin
13.2.2 Martinet

13.3 Left invariant structures

13.3.1 \(SU(2), SO(3), SL(2)\)
13.3.2 \(SE(2)\)
13.3.3 \((3, 5)\) - Rolling sphere with twist
Chapter 14

Curves in the Lagrange Grassmannian

In this chapter we introduce the manifold of Lagrangian subspaces of a symplectic vector space. After a description of its geometric properties, we discuss how to define the curvature for regular curves in the Lagrange Grassmannian, that are curves with non-degenerate derivative. Then we discuss the non-regular case, where a reduction procedure let us to reduce to a regular curve in a reduced symplectic space.

14.1 The geometry of the Lagrange Grassmannian

In this section we recall some basic facts about Grassmanians of $k$-dimensional subspaces of an $n$-dimensional vector space and then we consider, for a vector space endowed with a symplectic structure, the submanifold of its Lagrangian subspaces.

Definition 14.1. Let $V$ be an $n$-dimensional vector space. The Grassmanian of $k$-planes on $V$ is the set

$$G_k(V) := \{ W | W \subset V \text{ is a subspace, } \dim(W) = k \}.$$  

It is a standard fact that $G_k(V)$ is a compact manifold of dimension $k(n-k)$.

Now we describe the tangent space to this manifold.

Proposition 14.2. Let $W \in G_k(V)$. We have a canonical isomorphism

$$T_W G_k(V) \simeq \text{Hom}(W, V/W).$$

Proof. Consider a smooth curve on $G_k(V)$ which starts from $W$, i.e. a smooth family of $k$-dimensional subspaces defined by a moving frame

$$W(t) = \text{span}\{e_1(t), \ldots, e_k(t)\}, \quad W(0) = W.$$  

We want to associate in a canonical way with the tangent vector $\dot{W}(0)$ a linear operator from $W$ to the quotient $V/W$. Fix $w \in W$ and consider any smooth extension $w(t) \in W(t)$, with $w(0) = w$. Then define the map

$$W \to V/W, \quad w \mapsto \dot{w}(0) \text{ (mod } W).$$  

(14.1)
We are left to prove that the map (14.1) is well defined, i.e. independent on the choices of representatives. Indeed if we consider another extension \( w_1(t) \) of \( w \) satisfying \( w_1(t) \in W(t) \) we can write

\[
w_1(t) = w(t) + \sum_{i=1}^{k} \alpha_i(t)e_i(t),
\]

for some smooth coefficients \( \alpha_i(t) \) such that \( \alpha_i(0) = 0 \) for every \( i \). It follows that

\[
\dot{w}_1(t) = \dot{w}(t) + \sum_{i=1}^{k} \dot{\alpha}_i(t)e_i(t) + \sum_{i=1}^{k} \alpha_i(t)\dot{e}_i(t), \tag{14.2}
\]

and evaluating (14.2) at \( t = 0 \) one has

\[
\dot{w}_1(0) = \dot{w}(0) + \sum_{i=1}^{k} \dot{\alpha}_i(0)e_i(0).
\]

This shows that \( \dot{w}_1(0) = \dot{w}(0) \pmod W \), hence the map (14.1) is well defined. In the same way one can prove that the map does not depend on the moving frame defining \( W(t) \).

Finally, it is easy to show that the map that associates the tangent vector to the curve \( W(t) \) with the linear operator \( W \to V/W \) is surjective, hence it is an isomorphism since the two space have the same dimension. \( \square \)

Let us now consider a symplectic vector space \((\Sigma, \sigma)\), i.e. a \( 2n \)-dimensional vector space \( \Sigma \) endowed with a non degenerate symplectic form \( \sigma \in \Lambda^2(\Sigma) \).

**Definition 14.3.** A vector subspace \( \Pi \subset \Sigma \) of a symplectic space is called

(i) **symplectic** if \( \sigma|_{\Pi} \) is nondegenerate,

(ii) **isotropic** if \( \sigma|_{\Pi} \equiv 0 \),

(iii) **Lagrangian** if \( \sigma|_{\Pi} \equiv 0 \) and \( \dim \Pi = n \).

Notice that in general for every subspace \( \Pi \subset \Sigma \), by nondegeneracy of the symplectic form \( \sigma \), one has

\[
\dim \Pi + \dim \Pi^\perp = \dim \Sigma. \tag{14.3}
\]

where as usual we denote the symplectic orthogonal by \( \Pi^\perp = \{ x \in \Sigma \mid \sigma(x, y) = 0, \forall y \in \Pi \} \).

**Exercise 14.4.** Prove the following properties for a vector subspace \( \Pi \subset \Sigma \):

(i) \( \Pi \) is symplectic iff \( \Pi \cap \Pi^\perp = \{0\} \),

(ii) \( \Pi \) is isotropic iff \( \Pi \subset \Pi^\perp \),

(iii) \( \Pi \) is Lagrangian iff \( \Pi = \Pi^\perp \).

**Exercise 14.5.** Prove that, given two subspaces \( A, B \subset \Sigma \), one has the identities \( (A + B)^\perp = A^\perp \cap B^\perp \) and \( (A \cap B)^\perp = A^\perp + B^\perp \).
Example 14.6. Any symplectic vector space admits Lagrangian subspaces. Indeed fix any non-zero element $e_1 := e \neq 0$ in $\Sigma$. Choose iteratively

$$e_i \in \text{span}\{e_1, \ldots, e_{i-1}\} \setminus \text{span}\{e_1, \ldots, e_{i-1}\}, \quad i = 2, \ldots, n.$$  \hfill (14.4)

Then $\Pi := \text{span}\{e_1, \ldots, e_n\}$ is a Lagrangian subspace by construction. Notice that the choice (14.4) is possible by (14.3).

**Lemma 14.7.** Let $\Pi = \text{span}\{e_1, \ldots, e_n\}$ be a Lagrangian subspace of $\Sigma$. Then there exists vectors $f_1, \ldots, f_n \in \Sigma$ such that

(i) $\Sigma = \Pi \oplus \Delta$, \quad $\Delta := \text{span}\{f_1, \ldots, f_n\}$,

(ii) $\sigma(e_i, f_j) = \delta_{ij}$, \quad $\sigma(e_i, e_j) = \sigma(f_i, f_j) = 0$, \quad $\forall i, j = 1, \ldots, n$.

*Proof.* We prove the lemma by induction. By nondegeneracy of $\sigma$ there exists a non-zero $x \in \Sigma$ such that $\sigma(e_n, x) \neq 0$. Then we define the vector

$$f_n := \frac{x}{\sigma(e_n, x)}, \quad \Rightarrow \quad \sigma(e_n, f_n) = 1.$$  \hfill (14.5)

The last equality implies that $\sigma$ restricted to $\text{span}\{e_n, f_n\}$ is nondegenerate, hence by (a) of Exercise 14.4

$$\text{span}\{e_n, f_n\} \cap \text{span}\{e_n, f_n\} = 0,$$  \hfill (14.6)

And we can apply induction on the $2(n - 1)$ subspace $\Sigma' := \text{span}\{e_n, f_n\}$. Notice that (14.5) implies that $\sigma$ is non degenerate also on $\Sigma'$.

**Remark 14.8.** In particular the complementary subspace $\Delta = \text{span}\{f_1, \ldots, f_n\}$ defined in Lemma 14.7 is Lagrangian and transversal to $\Pi$.

Considering coordinates induced from the basis chosen for this splitting we can write $\Sigma = \mathbb{R}^{n*} \oplus \mathbb{R}^n$, (denoting $\mathbb{R}^{n*}$ denotes the set of row vectors). More precisely $x = (\zeta, z)$ if

$$x = \sum_{i=1}^{n} \zeta^i e_i + z^i f_i, \quad \zeta = (\zeta^1 \cdots \zeta^n), \quad z = \begin{pmatrix} z^1 \\ \vdots \\ z^n \end{pmatrix},$$

and using canonical form of $\sigma$ on our basis (see Lemma 14.7) we find that in coordinates, if $x_1 = (\zeta_1, z_1), x_2 = (\zeta_2, z_2)$ we get

$$\sigma(x_1, x_2) = \zeta_1 z_2 - \zeta_2 z_1,$$  \hfill (14.6)

where we denote with $\zeta z$ the standard rows by columns product.

Lemma 14.7 shows that the group of symplectomorphisms acts transitively on pairs of transversal Lagrangian subspaces. The next exercise, whose proof is an adaptation of the previous one, describes all the orbits of the action of the group of symplectomorphisms on pairs of subspaces of a symplectic vector spaces.

**Exercise 14.9.** Let $\Lambda_1, \Lambda_2$ be two subspaces in a symplectic vector space $\Sigma$, and assume that $\dim \Lambda_1 \cap \Lambda_2 = k$. Show that there exists Darboux coordinates $(p, q)$ in $\Sigma$ such that

$$\Lambda_1 = \{(p, 0)\}, \quad \Lambda_2 = \{((p_1, \ldots, p_k, 0, \ldots, 0), (0, \ldots, 0, q_{k+1}, \ldots, q_n))\}.$$
14.1.1 The Lagrange Grassmannian

**Definition 14.10.** The Lagrange Grassmannian $L(\Sigma)$ of a symplectic vector space $\Sigma$ is the set of its $n$-dimensional Lagrangian subspaces.

**Proposition 14.11.** $L(\Sigma)$ is a compact submanifold of the Grassmannian $G_n(\Sigma)$ of $n$-dimensional subspaces. Moreover

$$\dim L(\Sigma) = \frac{n(n+1)}{2}. \quad (14.7)$$

**Proof.** Recall that $G_n(\Sigma)$ is a $n^2$-dimensional compact manifold. Clearly $L(\Sigma) \subset G_n(\Sigma)$ as a subset. Consider the set of all Lagrangian subspaces that are transversal to a given one

$$\Delta^h = \{ \Lambda \in L(\Sigma) : \Lambda \cap \Delta = 0 \}.$$

Clearly $\Delta^h \subset L(\Sigma)$ is an open subset and since by Lemma [14.7] every Lagrangian subspace admits a Lagrangian complement

$$L(\Sigma) = \bigcup_{\Delta \in L(\Sigma)} \Delta^h.$$

It is then sufficient to find some coordinates on these open subsets. Every $n$-dimensional subspace $\Lambda \subset \Sigma$ which is transversal to $\Delta$ is the graph of a linear map from $\Pi$ to $\Delta$. More precisely there exists a matrix $S_\Lambda$ such that

$$\Lambda \cap \Delta = 0 \Leftrightarrow \Lambda = \{(z^T, S_\Lambda z), z \in \mathbb{R}^n\}.$$

(Here we used the coordinates induced by the splitting $\Sigma = \Pi \oplus \Delta$.) Moreover it is easily seen that

$$\Lambda \in L(\Sigma) \Leftrightarrow S_\Lambda = (S_\Lambda)^T.$$

Indeed we have that $\Lambda \in L(\Sigma)$ if and only if $\sigma|_{\Lambda} = 0$ and using (14.6) this is rewritten as

$$\sigma((z_1^T, S_\Lambda z_1), (z_2^T, S_\Lambda z_2)) = z_1^T S_\Lambda z_2 - z_2^T S_\Lambda z_1 = 0,$$

which means exactly $S_\Lambda$ symmetric. Hence the open set of all subspaces that are transversal to $\Lambda$ is parametrized by the set of symmetric matrices, that gives coordinates in this open set. This also proves that the dimension of $L(\Sigma)$ coincide with the dimension of the space of symmetric matrices, hence (14.7). Notice also that, being $L(\Sigma)$ a closed set in a compact manifold, it is compact. $\square$

Now we describe the tangent space to the Lagrange Grassmannian.

**Proposition 14.12.** Let $\Lambda \in L(\Sigma)$. Then we have a canonical isomorphism

$$T_\Lambda L(\Sigma) \simeq Q(\Lambda),$$

where $Q(\Lambda)$ denote the set of quadratic forms on $\Lambda$.

**Proof.** Consider a smooth curve $\Lambda(t)$ in $L(\Sigma)$ such that $\Lambda(0) = \Lambda$ and $\dot{\Lambda}(0) \in T_\Lambda L(\Sigma)$ its tangent vector. As before consider a point $x \in \Lambda$ and a smooth extension $x(t) \in \Lambda(t)$ and denote with $\dot{x} := \dot{x}(0)$. We define the map

$$\dot{\Lambda} : x \mapsto \sigma(x, \dot{x}), \quad (14.8)$$

284
that is nothing else but the quadratic map associated to the self-adjoint map $x \mapsto \dot{x}$ by the symplectic structure. We show that in coordinates $\dot{\Lambda}$ is a well defined quadratic map, independent on all choices. Indeed
\[
\Lambda(t) = \{(z^T, S_{\Lambda(t)} z), z \in \mathbb{R}^n \},
\]
and the curve $x(t)$ can be written
\[
x(t) = (z(t)^T, S_{\Lambda(t)} z(t)), \quad x(0) = (z^T, S_{\Lambda} z),
\]
for some curve $z(t)$ where $z = z(0)$. Taking derivative we get
\[
\dot{x}(t) = (\dot{z}(t)^T, \dot{S}_{\Lambda(t)} z(t) + S_{\Lambda(t)} \dot{z}(t)),
\]
and evaluating at $t = 0$ (we simply omit $t$ when we evaluate at $t = 0$) we have
\[
x = (z^T, S_{\Lambda} z), \quad \dot{x} = (\dot{z}^T, \dot{S}_{\Lambda} z + S_{\Lambda} \dot{z}),
\]
and finally get, using the simmetry of $S_{\Lambda}$, that
\[
\sigma(x, \dot{x}) = z^T (\dot{S}_{\Lambda} z + S_{\Lambda} \dot{z}) - \dot{z}^T S_{\Lambda} z = z^T \dot{S}_{\Lambda} z + \dot{z}^T S_{\Lambda} z - \dot{z}^T S_{\Lambda} z = z^T \dot{S}_{\Lambda} z.
\] (14.9)

Exercise 14.13. Let $\Lambda(t) \in L(\Sigma)$ such that $\Lambda = \Lambda(0)$ and $\sigma$ be the symplectic form. Prove that the map $S : \Lambda \times \Lambda \to \mathbb{R}$ defined by $S(x, y) = \sigma(x, \dot{y})$, where $\dot{y} = \dot{y}(0)$ is the tangent vector to a smooth extension $y(t) \in \Lambda(t)$ of $y$, is a symmetric bilinear map.

Remark 14.14. We have the following natural interpretation of this result: since $L(\Sigma)$ is a submanifold of the Grassmanian $G_n(\Sigma)$, its tangent space $T_{\Lambda}L(\Sigma)$ is naturally identified by the inclusion with a subspace of the Grassmannian
\[
i : L(\Sigma) \hookrightarrow G_n(\Sigma), \quad i_* : T_{\Lambda}L(\Sigma) \hookrightarrow T_{\Lambda}G_n(\Sigma) \simeq \text{Hom}(\Lambda, \Sigma/\Lambda),
\]
where the last isomorphism is Proposition 14.2. Being $\Lambda$ a Lagrangian subspace of $\Sigma$, the symplectic structure identifies in a canonical way the factor space $\Sigma/\Lambda$ with the dual space $\Lambda^*$ defining
\[
\Sigma/\Lambda \simeq \Lambda^*, \quad \langle [z]_\Lambda, x \rangle = \sigma(z, x).
\] (14.10)

Hence the tangent space to the Lagrange Grassmanian consist of those linear maps in the space $\text{Hom}(\Lambda, \Lambda^*)$ that are self-adjoint, which are naturally identified with quadratic forms on $\Lambda$ itself. \footnote{Any quadratic form on a vector space $q \in Q(V)$ can be identified with a self-adjoint linear map $L : V \to V^*$, $L(v) = B(v, \cdot)$ where $B$ is the symmetric bilinear map such that $q(v) = B(v, v)$.}

Remark 14.15. Given a curve $\Lambda(t)$ in $L(\Sigma)$, the above procedure associates to the tangent vector $\dot{\Lambda}(t)$ a family of quadratic forms $\dot{\Lambda}(t)$, for every $t$.

We end this section by computing the tangent vector to a special class of curves that will play a major role in the sequel, i.e. the curve on $L(\Sigma)$ induced by the action on $\Lambda$ by the flow of the linear Hamiltonian vector field $\vec{h}$ associated with a quadratic Hamiltonian $h \in C^\infty(\Sigma)$. (Recall that a Hamiltonian vector field transform Lagrangian subspaces into Lagrangian subspaces.)
Proposition 14.16. Let $\Lambda \in L(\Sigma)$ and define $\Lambda(t) = e^{t\bar{h}}(\Lambda)$. Then $\dot{\Lambda} = 2h|_{\Lambda}$.

Proof. Consider $x \in \Lambda$ and the smooth extension $x(t) = e^{t\bar{h}}(x)$. Then $\dot{x} = \bar{h}(x)$ and by definition of Hamiltonian vector field we find
\[
\sigma(x, \dot{x}) = \sigma(x, \bar{h}(x)) = \langle d_x h, x \rangle = 2h(x),
\]
where in the last equality we used that $h$ is quadratic on fibers. \hfill \square

14.2 Regular curves in Lagrange Grassmannian

The isomorphism between tangent vector to the Lagrange Grassmannian with quadratic forms makes sense to the following definition (we denote by $\dot{\Lambda}$ the tangent vector to the curve at the point $\Lambda$ as a quadratic map)

Definition 14.17. Let $\Lambda(t) \in L(\Sigma)$ be a smooth curve in the Lagrange Grassmannian. We say that the curve is

(i) monotone increasing (decreasing) if $\dot{\Lambda}(t) \geq 0$ ($\dot{\Lambda}(t) \leq 0$).

(ii) strictly monotone increasing (decreasing) if the inequality in (i) is strict.

(iii) regular if its derivative $\dot{\Lambda}(t)$ is a non degenerate quadratic form.

Remark 14.18. Notice that if $\Lambda(t) = \{(p, S(t)p), p \in \mathbb{R}^n\}$ in some coordinate set, then it follows from the proof of Proposition 14.12 that the quadratic form $\dot{\Lambda}(t)$ is represented by the matrix $\dot{S}_\Lambda(t)$ (see also (14.9)). In particular the curve is regular if and only if det $\dot{S}_\Lambda(t) \neq 0$.

The main goal of this section is the construction of a canonical Lagrangian complement. (i.e. another curve $\Lambda^0(t)$ in the Lagrange Grassmannian defined by $\Lambda(t)$ and such that $\Sigma = \Lambda(t) \oplus \Lambda^0(t)$.)

Consider an arbitrary Lagrangian splitting $\Sigma = \Lambda(0) \oplus \Delta$ defined by a complement $\Delta$ to $\Lambda(0)$ (see Lemma 14.7) and fix coordinates in such a way that that
\[
\Sigma = \{(p, q), p, q \in \mathbb{R}^n\}, \quad \Lambda(0) = \{(p, 0), p \in \mathbb{R}^n\}, \quad \Delta = \{(0, q), q \in \mathbb{R}^n\}.
\]
In these coordinates our regular curve is described by a one parametric family of symmetric matrices $S(t)$
\[
\Lambda(t) = \{(p, S(t)p), p \in \mathbb{R}^n\},
\]
such that $S(0) = 0$ and $\dot{S}(0)$ is invertible. All Lagrangian complement to $\Lambda(0)$ are parametrized by a symmetric matrix $B$ as follows
\[
\Delta_B = \{(Bq, q) \in \mathbb{R}^n\}, \quad B = B^T.
\]
The following lemma shows how the coordinate expression of our curve $\Lambda(t)$ change in the new coordinate set defined by the splitting $\Sigma = \Lambda(0) \oplus \Delta_B$. 

286
Lemma 14.19. Let \( S_B(t) \) the one parametric family of symmetric matrices defining \( \Lambda(t) \) in coordinates w.r.t. the splitting \( \Lambda(0) \oplus \Delta_B \). Then the following identity holds

\[
S_B(t) = (S(t)^{-1} - B)^{-1}.
\]  

(14.11)

Proof. It is easy to show that, if \( (p,q) \) and \( (p',q') \) denotes coordinates with respect to the splitting defined by the subspaces \( \Delta \) and \( \Delta_B \) we have

\[
\begin{align*}
\begin{cases}
p' = p - Bq \\
q' = q
\end{cases}
\end{align*}
\]

(14.12)

The matrix \( S_B(t) \) by definition is the matrix that satisfies the identity \( q' = S_B(t)p' \). Using that \( q = S(t)p \) by definition of \( \Lambda(t) \), from (14.12) we find

\[
q' = q = S(t)p = S(t)(p' + Bq'),
\]

and with straightforward computations we finally get

\[
S_B(t) = (I - S(t)B)^{-1}S(t) = (S(t)^{-1} - B)^{-1}.
\]

\[\Box\]

Since \( \dot{S}(t) \) represents the tangent vectors to the regular curve \( \Lambda(t) \), its properties are invariant with respect to change of coordinates. Hence it is natural to look for a change of coordinates (i.e. a choice of the matrix \( B \)) that simplifies the second derivative our curve.

Corollary 14.20. There exists a unique symmetric matrix \( B \) such that \( \ddot{S}_B(0) = 0 \).

Proof. Recall that for a one parametric family of matrices \( X(t) \) we have

\[
\frac{d}{dt}X(t)^{-1} = -X(t)^{-1}\dot{X}(t)X(t)^{-1}.
\]

Applying twice this identity to (14.11) (we omit \( t \) to denote the value at \( t = 0 \)) we get

\[
\frac{d}{dt} \Big|_{t=0} S_B(t) = -(S^{-1} - B)^{-1} \left( \frac{d}{dt} \Big|_{t=0} S^{-1}(t) \right) (S^{-1} - B)^{-1}
\]

\[
= (S^{-1} - B)^{-1}S^{-1}\dot{S}S^{-1}(S^{-1} - B)^{-1}
\]

\[
= (I - SB)^{-1}\dot{S}(I - BS)^{-1}.
\]

Hence for the second derivative evaluated at \( t = 0 \) (remember that in our coordinates \( S(0) = 0 \)) one gets

\[
\ddot{S}_B = \dot{S} + 2\dot{S}BS,
\]

and using that \( \dot{S} \) is non degenerate, we can choose \( B = -\frac{1}{2}\dot{S}^{-1}\dot{S}\dot{S}^{-1} \).

\[\Box\]

We set \( \Lambda^\circ(0) := \Delta_B \), where \( B \) is determined by (14.13). Notice that by construction \( \Lambda^\circ(0) \) is a Lagrangian subspace and it is transversal to \( \Lambda(0) \). The same argument can be applied to define \( \Lambda^\circ(t) \) for every \( t \).
**Definition 14.21.** Let $\Lambda(t)$ be a regular curve, the curve $\Lambda^o(t)$ defined by the condition above is called **derivative curve** of $\Lambda(t)$.

**Exercise 14.22.** Prove that, if $\Lambda(t) = \{(p, S(t)p), p \in \mathbb{R}^n\}$ (without the condition $S(0) = 0$), then the derivative curve $\Lambda^o(t) = \{(p, S^o(t)p), p \in \mathbb{R}^n\}$, satisfies

$$ S^o(t) = B(t)^{-1} + S(t), \quad \text{where} \quad B(t) := -\frac{1}{2} \dot{S}(t)^{-1} \ddot{S}(t) S(t)^{-1}, \quad (14.13) $$

provided $\Lambda^o(t)$ is transversal to the subspace $\Delta = \{(0, q), q \in \mathbb{R}^n\}$. (Actually this condition is equivalent to the invertibility of $B(t)$.) Notice that if $S(0) = 0$ then $S^o(0) = B(0)^{-1}$.

**Remark 14.23.** The set $\Lambda^t$ of all $n$-dimensional spaces transversal to a fixed subspace $\Lambda$ is an affine space over $\text{Hom}(\Sigma/\Lambda, \Lambda)$. Indeed given two elements $\Delta_1, \Delta_2 \in \Lambda^t$ we can associate with their difference the operator

$$ \Delta_2 - \Delta_1 \mapsto A \in \text{Hom}(\Sigma/\Lambda, \Lambda), \quad A([z]_\Lambda) = z_2 - z_1 \in \Lambda, \quad (14.14) $$

where $z_i \in \Delta_i \cap [z]_\Lambda$ are uniquely identified.

If $\Lambda$ is Lagrangian, we have identification $\Sigma/\Lambda \simeq \Lambda^*$ given by the symplectic structure (see (14.10)) that $\Lambda^h$, that coincide by definition with the intersection $\Lambda^t \cap L(\Sigma)$ is an affine space over $\text{Hom}^S(\Lambda^*, \Lambda)$, the space of selfadjoint maps between $\Lambda^*$ and $\Lambda$, that it isomorphic to $Q(\Lambda^*)$.

Notice that if we fix a distinguished complement of $\Lambda$, i.e. $\Sigma = \Lambda \oplus \Delta$, then we have also the identification $\Sigma/\Lambda \simeq \Delta$ and $\Lambda^h \simeq Q(\Lambda^*) \simeq Q(\Delta)$.

**Exercise 14.24.** Prove that the operator $A$ defined by (14.14), in the case when $\Lambda$ is Lagrangian, is a self-adjoint operator.

**Remark 14.25.** Assume that the splitting $\Sigma = \Lambda \oplus \Delta$ is fixed. Then our curve $\Lambda(t)$ in $L(\Sigma)$, such that $\Lambda(0) = \Lambda$, is characterized by a family of symmetric matrices $S(t)$ satisfying $\Lambda(t) = \{(p, S(t)p), p \in \mathbb{R}^n\}$, with $S(0) = 0$.

By regularity of the curve, $\Lambda(t) \in \Lambda^h$ for $t > 0$ small enough, hence we can consider its coordinate presentation in the affine space on the vector space of quadratic forms defined on $\Delta$ (see Remark 14.23) that is given by $S^{-1}(t)$ and write the Laurent expansion of this curve in the affine space

$$ S(t)^{-1} = \left( t \dot{S} + \frac{t^2}{2} \ddot{S} + O(t^3) \right)^{-1} \quad \begin{array}{l}
= \frac{1}{t} S^{-1} \left( I + \frac{t}{2} \dot{S} \dot{S}^{-1} + O(t^2) \right)^{-1} \\
= \frac{1}{t} \dot{S}^{-1} \left( \frac{1}{2} \ddot{S} \dot{S}^{-1} + O(t) \right) \\
\end{array} $$

It is not occasional that the matrix $B$ coincides with the free term of this expansion. Indeed the formula (14.11) for the change of coordinates can be rewritten as follows

$$ S_B(t)^{-1} = S^{-1}(t) - B, \quad (14.15) $$

and the choice of $B$ corresponds exactly to the choice of a coordinate set where the curve $\Lambda(t)$ has no free term in this expansion (i.e. $S_B(t)^{-1}$ has no term of order zero). This is equivalent to say that a regular curve let us to choose a privileged origin in the affine space of Lagrangian subspaces that are transversal to the curve itself.
14.3 Curvature of a regular curve

Now we want to define the curvature of a regular curve in the Lagrange Grassmannian. Let $\Lambda(t)$ be a regular curve and consider its derivative curve $\Lambda^o(t)$.

The tangent vectors to $\Lambda(t)$ and $\Lambda^o(t)$, as explained in Section 14.1, can be interpreted in a canonical way as a quadratic form on the space $\Lambda(t)$ and $\Lambda^o(t)$ respectively

$$\dot{\Lambda}(t) \in Q(\Lambda(t)), \quad \dot{\Lambda}^o(t) \in Q(\Lambda^o(t)).$$

Being $\Lambda^o(t)$ a canonical Lagrangian complement to $\Lambda(t)$ we have the identifications through the symplectic form

$$\Lambda(t)^* \simeq \Lambda^o(t), \quad \Lambda^o(t)^* \simeq \Lambda(t),$$

and the quadratic forms $\dot{\Lambda}(t), \dot{\Lambda}^o(t)$ can be treated as (self-adjoint) mappings:

$$\dot{\Lambda}(t) : \Lambda(t) \to \Lambda^o(t), \quad \dot{\Lambda}^o(t) : \Lambda^o(t) \to \Lambda(t). \quad (14.16)$$

**Definition 14.26.** The operator $R_{\Lambda}(t) := \dot{\Lambda}^o(t) \circ \dot{\Lambda}(t) : \Lambda(t) \to \Lambda(t)$ is called the *curvature* operator of the regular curve $\Lambda(t)$.

**Remark 14.27.** In the monotonic case, when $|\dot{\Lambda}(t)|$ defines a scalar product on $\Lambda(t)$, the operator $R(t)$ is, by definition, symmetric with respect to this scalar product. Moreover $R(t)$, as quadratic form, has the same signature and rank as $\dot{\Lambda}^o(t) \text{sign}(\dot{\Lambda}^o(t))$.

**Definition 14.28.** Let $\Lambda_1, \Lambda_2$ be two transversal Lagrangian subspaces of $\Sigma$. We denote

$$\pi_{\Lambda_1 \Lambda_2} : \Sigma \to \Lambda_2,$$

the projection on $\Lambda_2$ parallel to $\Lambda_1$, i.e. the linear operator such that

$$\pi_{\Lambda_1 \Lambda_2}|_{\Lambda_1} = 0, \quad \pi_{\Lambda_1 \Lambda_2}|_{\Lambda_2} = Id.$$

**Exercise 14.29.** Assume $\Lambda_1$ and $\Lambda_2$ be two Lagrangian subspaces in $\Sigma$ and assume that, in some coordinate set, $\Lambda_i = \{(x, S_i x), x \in \mathbb{R}^n\}$ for $i = 1, 2$. Prove that $\Sigma = \Lambda_1 \oplus \Lambda_2$ if and only if $\ker(S_1 - S_2) = \{0\}$. In this case show that the following matrix expression for $\pi_{\Lambda_1 \Lambda_2}$:

$$\pi_{\Lambda_1 \Lambda_2} = \begin{pmatrix} S_{12}^{-1}S_1 & -S_{12}^{-1} \\ S_2S_{12}^{-1}S_1 & -S_2S_{12}^{-1} \end{pmatrix}, \quad S_{12} := S_1 - S_2. \quad (14.18)$$

From the very definition of the derivative of our curve we can get the following geometric characterization of the curvature of a curve.

**Proposition 14.30.** Let $\Lambda(t)$ a regular curve in $L(\Sigma)$ and $\Lambda^o(t)$ its derivative curve. Then

$$\dot{\Lambda}(t)(x_t) = \pi_{\Lambda(t)\Lambda^o(t)}(\dot{x}_t), \quad \dot{\Lambda}^o(t)(x_t) = -\pi_{\Lambda^o(t)\Lambda(t)}(\dot{x}_t).$$

In particular the curvature is the composition $R_{\Lambda}(t) = \dot{\Lambda}^o(t) \circ \dot{\Lambda}(t)$.
Proof. Recall that, by definition, the linear operator $\hat{\Lambda} : \Lambda \to \Sigma/\Lambda$ associated with the quadratic form is the map $x \mapsto \hat{x}$ (mod $\Lambda$). Hence to build the map $\Lambda \to \Lambda^o$ it is enough to compute the projection of $\hat{x}$ onto the complement $\Lambda^o$, that is exactly $\pi_{\Lambda\Lambda^o}(\hat{x})$. Notice that the minus sign in equation (14.30) is a consequence of the skew symmetry of the symplectic product. More precisely, the sign in the identification $\Lambda^o \simeq \Lambda^*$ depends on the position of the argument. 

The curvature $R_\Lambda(t)$ of the curve $\Lambda(t)$ is a kind of relative velocity between the two curves $\Lambda(t)$ and $\Lambda^o(t)$. In particular notice that if the two curves moves in the same direction we have $R_\Lambda(t) > 0$.

Now we compute the expression of the curvature $R_\Lambda(t)$ in coordinates.

**Proposition 14.31.** Assume that $\Lambda(t) = \{(p, S(t)p)\}$ is a regular curve in $L(\Sigma)$. Then we have the following coordinate expression for the curvature of $\Lambda$ (we omit $t$ in the formula)

$$R_\Lambda = ((2\dot{S})^{-1}\ddot{S}) - ((2\dot{S})^{-1}\ddot{S})^2$$

$$= \frac{1}{2}\ddot{S}_{\cdot}^{-1}\dddot{S} - \frac{3}{4}(\dddot{S}_{\cdot}^{-1})^2.$$  \hfill (14.19)

$$= \frac{1}{2}\ddot{S}_{\cdot}^{-1}\dddot{S} - \frac{3}{4}(\dddot{S}_{\cdot}^{-1})^2.$$ \hfill (14.20)

**Proof.** Assume that both $\Lambda(t)$ and $\Lambda^o(t)$ are contained in the same coordinate chart with $\Lambda(t) = \{(p, S(t)p)\}, \quad \Lambda^o(t) = \{(p, S^o(t)p)\}.$

We start the proof by computing the expression of the linear operator associated with the derivative $\dot{\Lambda} : \Lambda \to \Lambda^o$ (we omit $t$ when we compute at $t = 0$). For each element $(p, Sp) \in \Lambda$ and any extension $(p(t), S(t)p(t))$ one can apply the matrix representing the operator $\pi_{\Lambda\Lambda^o}$ (see (14.18)) to the derivative at $t = 0$ and find $\pi_{\Lambda\Lambda^o}(p, Sp) = (p', S^o p'), \quad p' = -(S - S^o)^{-1}\dot{S}p.$

Exchanging the role of $\Lambda$ and $\Lambda^o$, and taking into account of the minus sign one finds that the coordinate representation of $R$ is given by $R = (S^o - S)^{-1}\dot{S}^o(S^o - S)^{-1}\dot{S}.$ \hfill (14.21)

We prove formula (14.20) under the extra assumption that $S(0) = 0$. Notice that this is equivalent to the choice of a particular coordinate set in $L(\Sigma)$ and, being the expression of $R$ coordinate independent by construction, this is not restrictive.

Under this extra assumption, it follows from (14.19) that $\Lambda(t) = \{(p, S(t)p)\}, \quad \Lambda^o(t) = \{(p, S^o(t)p)\},$

where $S^o(t) = B(t)^{-1} + S(t)$ and we denote by $B(t) := -\frac{1}{2}\dot{S}(t)^{-1}\dot{S}(t)\dot{S}(t)^{-1}$. Hence we have, assuming $S(0) = 0$ and omitting $t$ when $t = 0$

$$R = (S^o - S)^{-1}\dot{S}^o(S^o - S)^{-1}\dot{S}$$

$$= B \left( \frac{d}{dt} \bigg|_{t=0} B(t)^{-1} + S(t) \right) B\dot{S}$$

$$= (BS)^2 - B\dot{S}.$$ \hfill (14.20)

Plugging $B = -\frac{1}{2}\dot{S}^{-1}\dddot{S}\dot{S}^{-1}$ into the last formula, after some computations one gets to (14.20). \hfill \box
Remark 14.32. The formula for the curvature \( R_\Lambda(t) \) of a curve \( \Lambda(t) \) in \( L(\Sigma) \) takes a very simple form in a particular coordinate set given by the splitting \( \Sigma = \Lambda(0) \oplus \Lambda^\circ(0) \), i.e. such that 
\[
\Lambda(0) = \{(p,0), p \in \mathbb{R}^n\}, \quad \Lambda^\circ(0) = \{(0,q), q \in \mathbb{R}^n\}.
\]

Indeed using a symplectic change of coordinates in \( \Sigma \) that preserves both \( \Lambda \) and \( \Lambda^\circ \) (i.e. of the kind \( p' = Ap, \ q' = (A^{-1})^* q \) we can choose the matrix \( A \) in such a way that \( \dot{S}(0) = I \). Moreover we know from Proposition that the fact that \( \Lambda^\circ = \{(0,q), q \in \mathbb{R}^n\} \) is equivalent to \( \ddot{S}(0) = 0 \). Hence one finds from (14.20) that 
\[
R = \frac{1}{2} \ddot{S}
\]
When the curve \( \Lambda(t) \) is strictly monotone, the curvature \( R \) represents a well defined operator on \( \Lambda(0), \) naturally endowed with the sign definite quadratic for \( \dot{\Lambda}(0) \). Hence in these coordinates the eigenvalues of \( \ddot{S} \) (and not only the trace and the determinant) are invariants of the curve.

Exercise 14.33. Let \( f : \mathbb{R} \to \mathbb{R} \) be a smooth function. The Schwartzian derivative of \( f \) is defined as 
\[
Sf := \left( \frac{f''}{2f'} \right)' - \left( \frac{f''}{2f'} \right)^2
\]  
(14.22)
Prove that \( Sf = 0 \) if and only if \( f(t) = \frac{at+b}{ct+d} \) for some \( a, b, c, d \in \mathbb{R} \).

Remark 14.34. The previous proposition says that the curvature \( R \) is the matrix version of the Schwartzian derivative of the matrix \( S \) (cfr. (14.19) and (14.22)).

Example 14.35. Let \( \Sigma \) be a 2-dimensional symplectic space. In this case \( L(\Sigma) \simeq \mathbb{P}^1(\mathbb{R}) \) is the real projective line. Let us compute the curvature of a curve in \( L(\Sigma) \) with constant (angular) velocity \( \alpha > 0 \). We have
\[
\Lambda(t) = \{(p, S(t)p), p \in \mathbb{R}\}, \quad S(t) = \tan(\alpha t) \in \mathbb{R}.
\]
From the explicit expression it easy to find the relation 
\[
\dot{S}(t) = \alpha(1 + S^2(t)), \quad \Rightarrow \quad \frac{\ddot{S}(t)}{2\dot{S}(t)} = \alpha S(t),
\]
from which one gets that \( R(t) = \alpha \dot{S}(t) - \alpha^2 S^2(t) = \alpha^2 \), i.e. the curve has constant curvature.

We end this section with a useful formula on the curvature of a reparametrized curve.

Proposition 14.36. Let \( \varphi : \mathbb{R} \to \mathbb{R} \) a diffeomorphism and define the curve \( \Lambda_\varphi(t) := \Lambda(\varphi(t)) \). Then
\[
R_{\Lambda_\varphi}(t) = \varphi^2(t)R_{\Lambda}(\varphi(t)) + R_\varphi(t)\text{Id}.
\]  
(14.23)
Proof. It is a simple check that the Schwartzian derivative of the composition of two function \( f \) and \( g \) satisfies 
\[
S(f \circ g) = (Sf \circ g)(g')^2 + Sg.
\]
Notice that \( R_\varphi(t) \) makes sense as the curvature of the regular curve \( \varphi : \mathbb{R} \to \mathbb{R} \subset \mathbb{P}^1 \) in the Lagrange Grassmannian \( L(\mathbb{R}^2) \).
Exercise 14.37. (Another formula for the curvature). Let $\Lambda_0, \Lambda_1 \in L(\Sigma)$ be such that $\Sigma = \Lambda_0 \oplus \Lambda_1$ and fix two tangent vectors $\xi_0 \in T_{\Lambda_0} L(\Sigma)$ and $\xi_1 \in T_{\Lambda_1} L(\Sigma)$. As in (14.16) we can treat each tangent vector as a linear operator

$$\xi_0 : \Lambda_0 \rightarrow \Lambda_1, \quad \xi_1 : \Lambda_1 \rightarrow \Lambda_0, \quad (14.24)$$

and define the cross-ratio $[\xi_1, \xi_0] = -\xi_1 \circ \xi_0$. If in some coordinates $\Lambda_i = \{(p, S_i p)\}$ for $i = 0, 1$ we have

$$[\xi_1, \xi_0] = (S_1 - S_0)^{-1} \hat{S}_1 (S_1 - S_0)^{-1} \hat{S}_0.$$  

Let now $\Lambda(t)$ a regular curve in $L(\Sigma)$. By regularity $\Sigma = \Lambda(0) \oplus \Lambda(t)$ for all $t > 0$ small enough, hence the cross ratio

$$[\dot{\Lambda}(t), \dot{\Lambda}(0)] : \Lambda(0) \rightarrow \Lambda(0),$$

is well defined. Prove the following expansion for $t \rightarrow 0$

$$[\dot{\Lambda}(t), \dot{\Lambda}(0)] \simeq \frac{1}{t^2} Id + \frac{1}{3} R_\Lambda(0) + O(t). \quad (14.25)$$

14.4 Reduction of non-regular curves in Lagrange Grassmannian

In this section we want to extend the notion of curvature to non-regular curves. As we will see in the next chapter, it is always possible to associate with an extremal a family of Lagrangian subspaces in a symplectic space, i.e. a curve in a Lagrangian Grassmannian. This curve turns out to be regular if and only if the extremal is an extremal of a Riemannian structure. Hence, if we want to apply this theory for a genuine sub-Riemannian case we need some tools to deal with non-regular curves in the Lagrangian Grassmannian.

Let $(\Sigma, \sigma)$ be a symplectic vector space and $L(\Sigma)$ denote the Lagrange Grassmannian. We start by describing a natural subspace of $L(\Sigma)$ associated with an isotropic subspace $\Gamma$ of $\Sigma$. This will allow us to define a reduction procedure for a non regular curve.

Let $\Gamma$ be a $k$-dimensional isotropic subspace of $\Sigma$, i.e. $\sigma|\Gamma = 0$. This means that $\Gamma \subset \Gamma^\perp$. In particular $\Gamma^\perp/\Gamma$ is a $2(n - k)$ dimensional symplectic space with the restriction of $\sigma$.

Lemma 14.38. There is a natural identification of $L(\Gamma^\perp/\Gamma)$ as a subspace of $L(\Sigma)$:

$$L(\Gamma^\perp/\Gamma) \simeq \{\Lambda \in L(\Sigma), \Gamma \subset \Lambda\} \subset L(\Sigma). \quad (14.26)$$

Moreover we have a natural projection

$$\pi^\Gamma : L(\Sigma) \rightarrow L(\Gamma^\perp/\Gamma), \quad \Lambda \mapsto \Lambda^\Gamma,$$

where $\Lambda^\Gamma := (\Lambda \cap \Gamma^\perp) + \Gamma = (\Lambda + \Gamma) \cap \Gamma^\perp$.

Proof. Assume that $\Lambda \in L(\Sigma)$ and $\Gamma \subset \Lambda$. Then, since $\Lambda$ is Lagrangian, $\Lambda = \Lambda^\perp \subset \Gamma^\perp$, hence the identification (14.26).

Assume now that $\Lambda \in L(\Gamma^\perp/\Gamma)$ and let us show that $\pi^\Gamma(\Lambda) = \Lambda$, i.e. $\pi^\Gamma$ is a projection. Indeed from the inclusions $\Gamma \subset \Lambda \subset \Gamma^\perp$ one has $\pi^\Gamma(\Lambda) = \Lambda^\Gamma = (\Lambda \cap \Gamma^\perp) + \Gamma = \Lambda + \Gamma = \Lambda$.

3Here $\hat{S}_i$ denotes the matrix associated with $\xi_i$. 

292
We are left to check that \( \Lambda^\Gamma \) is Lagrangian, i.e. \((\Lambda^\Gamma)^\cdot = \Lambda^\Gamma\).

\[
(\Lambda^\Gamma)^\cdot = ((\Lambda \cap \Gamma^\cdot) + \Gamma)^\cdot \\
= (\Lambda \cap \Gamma^\cdot)^\cdot \cap \Gamma^\cdot \\
= (\Lambda + \Gamma) \cap \Gamma^\cdot = \Lambda^\Gamma,
\]

where we repeatedly used Exercise 14.5. (The identity \((\Lambda \cap \Gamma^\cdot) + \Gamma = (\Lambda + \Gamma) \cap \Gamma^\cdot \) is also a consequence of the same exercise.)

**Remark 14.39.** Let \( \Gamma^0 = \{ \Lambda \in L(\Sigma), \Lambda \cap \Gamma = \{0\} \} \). The restriction \( \pi^\Gamma|_{\Gamma^0} \) is smooth. Indeed it can be shown that \( \pi^\Gamma \) is defined by a rational function, since it is expressed via the solution of a linear system.

The following example shows that the projection \( \pi^\Gamma \) is not globally continuous on \( L(\Sigma) \).

**Example 14.40.** Consider the symplectic structure \( \sigma \) on \( \mathbb{R}^4 \), with Darboux basis \( \{e_1, e_2, f_1, f_2\} \), i.e. \( \sigma(e_i, f_j) = \delta_{ij} \). Let \( \Gamma = \text{span}\{e_1\} \) be a one dimensional isotropic subspace and define

\[
\Lambda_\varepsilon = \text{span}\{e_1 + \varepsilon f_2, e_2 + \varepsilon f_1\}, \quad \forall \varepsilon > 0.
\]

It is easy to see that \( \Lambda_\varepsilon \) is Lagrangian for every \( \varepsilon \) and that

\[
\Lambda_\varepsilon^\Gamma = \text{span}\{e_1, f_2\}, \quad \forall \varepsilon > 0, \\
\Lambda_0^\Gamma = \text{span}\{e_1, e_2\}.
\]

Indeed \( f_2 \in \Gamma^\cdot \), that implies \( e_1 + \varepsilon f_2 \in \Lambda_\varepsilon \cap \Gamma^\cdot \), therefore \( f_2 \in \Lambda_\varepsilon \cap \Gamma^\cdot \). By definition of reduced curve \( f_2 \in \Lambda_\varepsilon^\Gamma \) and (14.27) holds. The case \( \varepsilon = 0 \) is trivial.

### 14.5 Ample curves

In this section we introduce ample curves.

**Definition 14.41.** Let \( \Lambda(t) \in L(\Sigma) \) be a smooth curve in the Lagrange Grassmannian. The curve \( \Lambda(t) \) is **ample** at \( t = t_0 \) if there exists \( N \in \mathbb{N} \) such that

\[
\Sigma = \text{span}\{\lambda^{(i)}(t_0) | \lambda(t) \in \Lambda(t), \lambda(t) \text{ smooth}, 0 \leq i \leq N\}.
\]

(14.28)

In other words we require that all derivatives up to order \( N \) of all smooth sections of our curve in \( L(\Sigma) \) span all the possible directions.

As usual, we can choose coordinates in such a way that, for some family of symmetric matrices \( S(t) \), one has

\[
\Sigma = \{(p, q) | p, q \in \mathbb{R}^n\}, \quad \Lambda(t) = \{(p, S(t)p) | p \in \mathbb{R}^n\}.
\]

**Exercise 14.42.** Assume that \( \Lambda(t) = \{(p, S(t)p), p \in \mathbb{R}^n\} \) with \( S(0) = 0 \). Prove that the curve is ample at \( t = 0 \) if and only if there exists \( N \in \mathbb{N} \) such that all the columns of the derivative of \( S(t) \) up to order \( N \) (and computed at \( t = 0 \)) span a maximal subspace:

\[
\text{rank}\{\dot{S}(0), \ddot{S}(0), \ldots, S^{(N)}(0)\} = n.
\]

(14.29)

In particular, a curve \( \Lambda(t) \) is regular at \( t_0 \) if and only if is ample at \( t_0 \) with \( N = 1 \).
An important property of ample and monotone curves is described in the following lemma.

Lemma 14.43. Let \( \Lambda(t) \in L(\Sigma) \) a monotone, ample curve at \( t_0 \). Then, there exists \( \varepsilon > 0 \) such that \( \Lambda(t) \cap \Lambda(t_0) = \{0\} \) for \( 0 < |t - t_0| < \varepsilon \).

Proof. Without loss of generality, assume \( t_0 = 0 \). Choose a Lagrangian splitting \( \Sigma = \Lambda \oplus \Pi \), with \( \Lambda = J_0 \). For \( |t| < \varepsilon \), the curve is contained in the chart defined by such a splitting. In coordinates, \( \Lambda(t) = \{(p, S(t)p) \mid p \in \mathbb{R}^n\} \), with \( S(t) \) symmetric and \( S(0) = 0 \). The curve is monotone, then \( S(t) \) is a semidefinite symmetric matrix. It follows that \( S(t) \) is semidefinite too.

Suppose that, for some \( t \), \( \Lambda(t) \cap \Lambda(0) \neq \{0\} \) (assume \( t > 0 \)). This means that \( \exists v \in \mathbb{R}^n \) such that \( S(t)v = 0 \). Indeed also \( v^*S(t)v = 0 \). The function \( \tau \mapsto v^*S(\tau)v \) is monotone, vanishing at \( \tau = 0 \) and \( \tau = t \). Therefore \( v^*S(\tau)v = 0 \) for all \( 0 \leq \tau \leq t \). Being a semidefinite, symmetric matrix, \( v^*S(\tau)v = 0 \) if and only if \( S(\tau)v = 0 \). Therefore, we conclude that \( v \in \ker S(\tau) \) for \( 0 \leq \tau \leq t \). This implies that, for any \( i \in \mathbb{N} \), \( v \in \ker S^{(i)}(0) \), which is a contradiction, since the curve is ample at \( 0 \).

Exercise 14.44. Prove that a monotone curve \( \Lambda(t) \) is ample at \( t_0 \) if and only if one of the equivalent conditions is satisfied

(i) the family of matrices \( S(t) - S(t_0) \) is nondegenerate for \( t \neq t_0 \) close enough, and the same remains true if we replace \( S(t) \) by its \( N \)-th Taylor polynomial, for some \( N \) in \( \mathbb{N} \).

(ii) the map \( t \mapsto \det(S(t) - S(t_0)) \) has a finite order root at \( t = t_0 \).

Let us now consider an analytic monotone curve on \( L(\Sigma) \). Without loss of generality we can assume the curve to be non increasing, i.e. \( \Lambda(t) \geq 0 \). By monotonicity

\[
\Lambda(0) \cap \Lambda(t) = \bigcap_{0 \leq \tau \leq t} \Lambda(\tau) =: \Upsilon_t 
\]

Clearly \( \Upsilon_t \) is a decreasing family of subspaces, i.e. \( \Upsilon_t \subset \Upsilon_{t'} \) if \( t \leq t' \). Hence the family \( \Upsilon_t \) for \( t \to 0 \) stabilizes and the limit subspace \( \Upsilon \) is well defined

\[
\Upsilon := \lim_{t \to 0} \Upsilon_t
\]

The symplectic reduction of the curve by the isotropic subspace \( \Upsilon \) defines a new curve \( \Lambda(t) := \Lambda(t)^\Upsilon \in L(\Upsilon^\perp/\Upsilon) \).

Proposition 14.45. If \( \Lambda(t) \) is analytic and monotone in \( L(\Sigma) \), then \( \Lambda(t) \) is ample \( L(\Upsilon^\perp/\Upsilon) \).

Proof. By construction, in the reduced space \( \Upsilon^\perp/\Upsilon \) we removed the intersection of \( \Lambda(t) \) with \( \Lambda(0) \). Hence

\[
\Lambda(0) \cap \Lambda(t) = \{0\}, \quad \text{in} \quad L(\Upsilon^\perp/\Upsilon) \quad (14.30)
\]

In particular, if \( \tilde{S}(t) \) denotes the symmetric matrix representing \( \Lambda(t) \) such that \( \tilde{S}(0) = \Lambda(t_0) \), it follows that \( \tilde{S}(t) \) is non degenerate for \( 0 < |t| < \varepsilon \). The analyticity of the curve guarantees that the Taylor polynomial (of a suitable order \( N \)) is also non degenerate.
14.6 From ample to regular

In this section we prove the main result of this chapter, i.e. that any ample monotone curve can be reduced to a regular one.

**Theorem 14.46.** Let $\Lambda(t)$ be a smooth ample monotone curve and set $\Gamma := \text{Ker } \dot{\Lambda}(0)$. Then the reduced curve $t \mapsto \Lambda^\Gamma(t)$ is a smooth regular curve. In particular $\dot{\Lambda}^\Gamma(0) > 0$.

Before proving Theorem 14.46 let us discuss two useful lemmas.

**Lemma 14.47.** Let $v_1(t), \ldots, v_k(t) \in \mathbb{R}^n$ and define $V(t)$ as the $n \times k$ matrix whose columns are the vectors $v_i(t)$. Define the matrix $S(t) := \int_0^t V(\tau)V(\tau)^*d\tau$. Then the following are equivalent:

(i) $S(t)$ is invertible (and positive definite),

(ii) $\text{span}\{v_i(\tau) | i = 1, \ldots, k; \tau \in [0, t]\} = \mathbb{R}^n$.

**Proof.** Fix $t > 0$ and let us assume $S(t)$ is not invertible. Since $S(t)$ is non negative then there exists a nonzero $x \in \mathbb{R}^n$ such that $\langle S(t)x, x \rangle = 0$. On the other hand

$$\langle S(t)x, x \rangle = \int_0^t \langle V(\tau)V(\tau)^*x, x \rangle d\tau = \int_0^t \|V(\tau)^*x\|^2d\tau$$

This implies that $V(\tau)^*x = 0$ (or equivalently $x^*V(\tau) = 0$) for $\tau \in [0, t]$, i.e. the nonzero vector $x^*$ is orthogonal to $\text{Im}_{\tau \in [0, t]} V(\tau) = \text{span}\{v_i(\tau) | i = 1, \ldots, k, \tau \in [0, t]\} = \mathbb{R}^n$, that is a contradiction. The converse is similar. □

**Lemma 14.48.** Let $A, B$ two positive and symmetric matrices such that $0 < A < B$. Then we have also $0 < B^{-1} < A^{-1}$.

**Proof.** Assume first that $A$ and $B$ commute. Then $A$ and $B$ can be simultaneously diagonalized and the statement is trivial for diagonal matrices.

In the general case, since $A$ is symmetric and positive, we can consider its square root $A^{1/2}$, which is also symmetric and positive. We can write

$$0 < \langle Av, v \rangle < \langle Bv, v \rangle$$

By setting $w = A^{1/2}v$ in the above inequality and using $\langle Av, v \rangle = \langle A^{1/2}v, A^{1/2}v \rangle$ one gets

$$0 < \langle w, w \rangle < \langle A^{-1/2}BA^{-1/2}w, w \rangle,$$

which is equivalent to $I < A^{-1/2}BA^{-1/2}$.

Since the identity matrix commutes with every other matrix, we obtain

$$0 < A^{1/2}B^{-1}A^{1/2} = (A^{-1/2}BA^{-1/2})^{-1} < I$$

which is equivalent to $0 < B^{-1} < A^{-1}$ reasoning as before. □

**Proof of Theorem 14.46.** By assumption the curve $t \mapsto \Lambda(t)$ is ample, hence $\Lambda(t) \cap \Gamma = \{0\}$ and $t \mapsto \Lambda^\Gamma(t)$ is smooth for $t > 0$ small enough. We divide the proof into three parts: (i) we compute the coordinate presentation of the reduced curve. (ii) we show that the reduced curve, extended by continuity at $t = 0$, is smooth. (iii) we prove that the reduced curve is regular.
(i). Let us consider Darboux coordinates in the symplectic space \( \Sigma \) such that
\[ \Sigma = \{(p, q) : p, q \in \mathbb{R}^n\}, \quad \Lambda(t) = \{(p, S(t)p) : p \in \mathbb{R}^n\}, \quad S(0) = 0. \]
Moreover we can assume also \( \mathbb{R}^n = \mathbb{R}^k \oplus \mathbb{R}^{n-k} \), where \( \Gamma = \{0\} \oplus \mathbb{R}^{n-k} \). According to this splitting we have the decomposition \( p = (p_1, p_2) \) and \( q = (q_1, q_2) \). The subspaces \( \Gamma \) and \( \Gamma^\perp \) are described by the equations
\[ \Gamma = \{(p, q) : p_1 = 0, q = 0\}, \quad \Gamma^\perp = \{(p, q) : q_2 = 0\} \]
and \((p_1, q_1)\) are natural coordinates for the reduced space \( \Gamma^\perp/\Gamma \). Up to a symplectic change of coordinates preserving the splitting \( \mathbb{R}^n = \mathbb{R}^k \oplus \mathbb{R}^{n-k} \) we can assume that
\[ S(t) = \begin{pmatrix} S_{11}(t) & S_{12}(t) \\ S_{12}^*(t) & S_{22}(t) \end{pmatrix}, \quad \text{with} \quad S(0) = \begin{pmatrix} I_k & 0 \\ 0 & 0 \end{pmatrix}. \tag{14.31} \]
where \( I_k \) is the \( k \times k \) identity matrix. Finally, from the fact that \( S \) is monotone and ample, that implies \( S(t) > 0 \) for each \( t > 0 \), it follows
\[ S_{11}(t) > 0, \quad S_{22}(t) > 0, \quad \forall \ t > 0. \tag{14.32} \]
Then we can compute the coordinate expression of the reduced curve, i.e. the matrix \( S^\Gamma(t) \) such that
\[ \Lambda^\Gamma(t) = \{(p_1, S^\Gamma(t)p_1) : p_1 \in \mathbb{R}^k\}. \]
From the identity
\[ \Lambda(t) \cap \Gamma^\perp = \{(p, S(t)p), S(t)p \in \mathbb{R}^k\} = \left\{ \begin{pmatrix} S^{-1}(t)(q_1) \\ 0 \end{pmatrix}, q_1 \in \mathbb{R}^k \right\} \tag{14.33} \]
one gets the key relation \( S^\Gamma(t)^{-1} = (S(t)^{-1})_{11} \).
Thus the matrix expression of the reduced curve \( \Lambda^\Gamma(t) \) in \( L(\Gamma^\perp/\Gamma) \) is recovered simply by considering it as a map of \((p_1, q_1)\) only, i.e.
\[ S(t)p = \begin{pmatrix} S_{11} & S_{12} \\ S_{12}^* & S_{22} \end{pmatrix} \begin{pmatrix} p_1 \\ p_2 \end{pmatrix} = \begin{pmatrix} S_{11}p_1 + S_{12}p_2 \\ S_{12}^*p_1 + S_{22}p_2 \end{pmatrix} \]
from which we get \( S(t)p \in \mathbb{R}^k \) if and only if \( S_{12}^*(t)p_1 + S_{22}(t)p_2 = 0 \). Then
\[ \Lambda^\Gamma(t) = \{(p_1, S_{11}p_1 + S_{12}p_2) : S_{12}^*(t)p_1 + S_{22}(t)p_2 = 0\} = \{(p_1, (S_{11} - S_{12}S_{22}^{-1}S_{12}^*)p_1)\} \]
that means
\[ S^\Gamma = S_{11} - S_{12}S_{22}^{-1}S_{12}^*. \tag{14.34} \]
(ii). By the coordinate presentation of \( S^\Gamma(t) \) the only term that can give rise to singularities is the inverse matrix \( S_{22}^{-1}(t) \). In particular, since by assumption \( t \mapsto \det S_{22}(t) \) has a finite order zero at \( t = 0 \), the a priori singularity can be only a finite order pole.
To prove that the curve is smooth it is enough to show that \( S^\Gamma(t) \to 0 \) for \( t \to 0 \), i.e. the curve remains bounded. This follows from the following

Claim I. As quadratic forms on \( \mathbb{R}^k \), we have the inequality \( 0 \leq S^\Gamma(t) \leq S_{11}(t) \).
Indeed \( S(t) \) symmetric and positive one has that its inverse \( S(t)^{-1} \) is symmetric and positive also. This implies that \( S^{\Gamma}(t)^{-1} = (S(t)^{-1})_{11} > 0 \) and so is \( S^{\Gamma}(t) \). This proves the left inequality of the Claim I.

Moreover using (14.34) and the fact that \( S_{22} \) is positive definite (and so \( S_{22}^{-1} \)) one gets

\[
\langle (S_{11} - S^{\Gamma})p_1, p_1 \rangle = \langle S_{12}S_{22}^{-1}S_{12}^{\ast}p_1, p_1 \rangle = \langle S_{22}^{-1}(S_{12}^{\ast}p_1), (S_{12}p_1) \rangle \geq 0.
\]

Since \( S(t) \rightarrow 0 \) for \( t \rightarrow 0 \), clearly \( S_{11}(t) \rightarrow 0 \) when \( t \rightarrow 0 \), that proves that \( S^{\Gamma}(t) \rightarrow 0 \) also.

(iii). We are reduced to show that the derivative of \( t \mapsto S^{\Gamma}(t) \) at 0 is non degenerate matrix, which is equivalent to show that \( t \mapsto S^{\Gamma}(t)^{-1} \) has a simple pole at \( t = 0 \).

We need the following lemma, whose proof is postponed at the end of the proof of Theorem 14.46.

**Lemma 14.49.** Let \( A(t) \) be a smooth family of symmetric nonnegative \( n \times n \) matrices. If the condition \( \text{rank}(A, \dot{A}, \ldots, A^{(N)})|_{t=0} = n \) is satisfied for some \( N \), then there exists \( \varepsilon_0 > 0 \) such that \( \varepsilon t A(0) < \int_0^t A(\tau) d\tau \) for all \( \varepsilon < \varepsilon_0 \) and \( t > 0 \) small enough.

Applying the Lemma to the family \( A(t) = \dot{S}(t) \) one obtains (see also (14.31))

\[
\langle S(t)p, p \rangle > \varepsilon t |p_1|^2
\]

for all \( 0 < \varepsilon < \varepsilon_0 \), any \( p \in \mathbb{R}^n \) and any small time \( t > 0 \).

Now let \( p_1 \in \mathbb{R}^k \) be arbitrary and extend it to a vector \( p = (p_1, p_2) \in \mathbb{R}^n \) such that \( (p, S(t)p) \in \Lambda(t) \cap \Gamma^C \) (i.e. \( S(t)p = (q_1, 0)^T \) or equivalently \( S(t)^{-1}(q_1, 0) = (p_1, p_2) \)). This implies in particular that \( S^{\Gamma}(t)p_1 = q_1 \) and

\[
\langle S^{\Gamma}(t)p_1, p_1 \rangle = \langle S(t)p, p \rangle \geq \varepsilon t |p_1|^2,
\]

This identity can be rewritten as \( S^{\Gamma}(t) > \varepsilon t \mathbb{1}_k > 0 \) and implies by Lemma 14.48

\[
0 < S^{\Gamma}(t)^{-1} < \frac{1}{\varepsilon t} \mathbb{1}_k
\]

which completes the proof. \( \square \)

**Proof of Lemma 14.49.** We reduce the proof of the Lemma to the following statement:

**Claim II.** There exists \( c, \tilde{N} > 0 \) such that for any sufficiently small \( \varepsilon, t > 0 \)

\[
\det \left( \int_0^t A(\tau) - \varepsilon A(0) \, d\tau \right) > c t^{\tilde{N}}.
\]

Moreover \( c, \tilde{N} \) depends only on the \( 2N \)-th Taylor polynomial of \( A(t) \).

Indeed fix \( t_0 > 0 \). Since \( A(t) \geq 0 \) and \( A(t) \) is not the zero family, then \( \int_0^{t_0} A(\tau) d\tau > 0 \). Hence, for a fixed \( t_0 \), there exists \( \varepsilon \) small enough such that \( \int_0^{t_0} A(\tau) - \varepsilon A(0) \, d\tau > 0 \). Assume now that the matrix \( S_t = \int_0^t A(\tau) - \varepsilon A(0) \, d\tau > 0 \) is not strictly positive for some \( 0 < t < t_0 \), then det \( S(\tau) = 0 \) for some \( \tau \in [t, t_0] \), that is a contradiction.

We now prove Claim II. We may assume that \( t \mapsto A(t) \) is analytic. Indeed, by continuity of the determinant, the statement remains true if we substitute \( A(t) \) by its Taylor polynomial of sufficiently big order.
An analytic one parameter family of symmetric matrices $t \mapsto A(t)$ can be simultaneously diagonalized (see ??), in the sense that there exists an analytic (with respect to $t$) family of vectors $v_i(t)$, with $i = 1, \ldots, n$, such that

$$\langle A(t)x, x \rangle = \sum_{i=1}^{n} \langle v_i(t), x \rangle^2.$$ 

In other words $A(t) = V(t)V(t)^*$, where $V(t)$ is the $n \times n$ matrix whose columns are the vectors $v_i(t)$. (Notice that some of these vectors can vanish at 0 or even vanish identically.)

Let us now consider the flag $E_1 \subset E_2 \subset \ldots \subset E_N = \mathbb{R}^n$ defined as follows

$$E_i = \text{span}\{v^{(l)}_j, 1 \leq j \leq n, 0 \leq l \leq i \}.$$ 

Notice that this flag is finite by our assumption on the rank of the consecutive derivatives of $A(t)$ and $N$ is the same as in the statement of the Lemma. We then choose coordinates in $\mathbb{R}^n$ adapted to this flag (i.e. the spaces $E_i$ are coordinate subspaces) and define the following integers (here $e_1, \ldots, e_n$ is the standard basis of $\mathbb{R}^n$)

$$m_i = \min\{j : e_i \in E_j\}, \quad i = 1, \ldots, n.$$ 

In other words, when written in this new coordinate set, $m_i$ is the order of the first nonzero term in the Taylor expansion of the $i$-th row of the matrix $V(t)$. Then we introduce a quasi-homogeneous family of matrices $\tilde{V}(t)$: the $i$-th row of $\tilde{V}(t)$ is the $m_i$-homogeneous part of the $i$-thed row of $V(t)$. Then we define $\hat{A}(t) := \tilde{V}(t)\tilde{V}(t)^*$. The columns of the matrix $\hat{A}(t)$ satisfies the assumption of Lemma 14.47 then \( \int_0^t \hat{A}(\tau)d\tau > 0 \) for every $t > 0$.

If we denote the entries $A(t) = \{a_{ij}(t)\}^n_{i,j=1}$ and $\hat{A}(t) = \{\hat{a}_{ij}(t)\}^n_{i,j=1}$ we obtain

$$\hat{a}_{ij}(t) = c_{ij} \tau^{m_i+m_j}, \quad a_{ij}(t) = \hat{a}_{ij}(t) + O(\tau^{m_i+m_j+1}),$$

for suitable constants $c_{ij}$ (some of them may be zero).

Then we let $A^\varepsilon(t) := A(t) - \varepsilon A(0) = \{a^\varepsilon_{ij}(t)\}^n_{i,j=1}$. Of course $a^\varepsilon_{ij}(t) = \tau^{m_i+m_j} + O(\tau^{m_i+m_j+1})$ where

$$c^\varepsilon_{ij} = \begin{cases} (1 - \varepsilon)c_{ij}, & \text{if } m_i + m_j = 0, \\ c_{ij}, & \text{if } m_i + m_j > 0. \end{cases}$$

From the equality

$$\int_0^t a^\varepsilon_{ij}(\tau)d\tau = t^{m_i+m_j+1} \left( \frac{c^\varepsilon_{ij}}{m_i + m_j + 1} + O(t) \right)$$

one gets

$$\det \left( \int_0^t A^\varepsilon(\tau)d\tau \right) = t^{n+2} \sum_{i=1}^N m_i \left( \det \left( \frac{c^\varepsilon_{ij}}{m_i + m_j + 1} \right) + O(t) \right)$$

On the other hand

$$\det \left( \int_0^t \hat{A}(\tau)d\tau \right) = t^{n+2} \sum_{i=1}^N m_i \left( \det \left( \frac{c_{ij}}{m_i + m_j + 1} \right) + O(t) \right) > 0$$

hence $\det \left( \frac{c^\varepsilon_{ij}}{m_i + m_j + 1} \right) > 0$ for small $\varepsilon$. The proof is completed by setting

$$c := \det \left( \frac{c_{ij}}{m_i + m_j + 1} \right), \quad \hat{N} := n + 2 \sum_{i=1}^N m_i.$$
14.7 Conjugate points in $L(\Sigma)$

In this section we introduce the notion of conjugate point for a curve in the Lagrange Grassmannian. In the next chapter we explain why this notion coincide with the one given for extremal paths in sub-Riemannian geometry.

**Definition 14.50.** Let $\Lambda(t)$ be a monotone curve in $L(\Sigma)$. We say that $\Lambda(t)$ is *conjugate to* $\Lambda(0)$ if $\Lambda(t) \cap \Lambda(0) \neq \{0\}$.

As a consequence of Lemma [14.43] we have the following immediate corollary.

**Corollary 14.51.** Conjugate points on a monotone and ample curve in $L(\Sigma)$ are isolated.

The following two results describe general properties of conjugate points

**Theorem 14.52.** Let $\Lambda(t), \Delta(t)$ two ample monotone curves in $L(\Sigma)$ defined on $\mathbb{R}$ such that

(i) $\Sigma = \Lambda(t) \oplus \Delta(t)$ for every $t \geq 0$,

(ii) $\dot{\Lambda}(t) \leq 0, \dot{\Delta}(t) \geq 0$, as quadratic forms.

Then there exists no $\tau > 0$ such that $\Lambda(\tau)$ is conjugate to $\Lambda(0)$. Moreover $\exists \lim_{t \to +\infty} \Lambda(t) = \Lambda(\infty)$.

**Proof.** Fix coordinates induced by some Lagrangian splitting of $\Sigma$ in such a way that $S\Lambda(0) = 0$ and $S\Delta(0) = I$. The monotonicity assumption implies that $t \mapsto S\Lambda(t)$ (resp. $t \mapsto S\Delta(t)$) is a monotone increasing (resp. decreasing) curve in the space of symmetric matrices. Moreover the transversality of $\Lambda(t)$ and $\Delta(t)$ implies that $S\Delta(t) - S\Lambda(t)$ is a non degenerate matrix for all $t$. Hence

$$0 < S\Lambda(t) < S\Delta(t) < I, \quad \text{for all } t > 0.$$ 

In particular $\Lambda(t)$ never leaves the coordinate neighborhood under consideration, the subspace $\Lambda(t)$ is always traversal to $\Lambda(0)$ for $t > 0$ and has a limit $\Lambda(\infty)$ whose coordinate representation is $S\Lambda(\infty) = \lim_{t \to +\infty} S\Lambda(t)$. 

**Theorem 14.53.** Let $\Lambda_s(t)$, for $t,s \in [0,1]$ be an homotopy of curves in $L(\Sigma)$ such that $\Lambda_s(0) = \Lambda$ for $s \in [0,1]$. Assume that

(i) $\Lambda_s(\cdot)$ is monotone and ample for every $s \in [0,1]$,

(ii) $\Lambda_0(\cdot), \Lambda_1(\cdot)$ and $\Lambda_s(1)$, for $s \in [0,1]$, contains no conjugate points to $\Lambda$.

Then no curve $t \mapsto \Lambda_s(t)$ contains conjugate points to $\Lambda$.

**Proof.** Let us consider the open chart $\Lambda^h$ defined by all the Lagrangian subspaces traversal to $\Lambda$. The statement is equivalent to prove that $\Lambda_s(t) \in \Lambda^h$ for all $t > 0$ and $s \in [0,1]$. Let us fix coordinates induced by some Lagrangian splitting $\Sigma = \Lambda \oplus \Delta$ in such a way that $\Lambda = \{(p,0)\}$ and

$$\Lambda_s(t) = \{(B_s(t)q,q)\}$$

for all $s$ and $t > 0$ (at least for $t$ small enough, indeed by ampleness $\Lambda_s(t) \in \Lambda^h$ for $t$ small). Moreover we can assume that $B_s(t)$ is a monotone increasing family of symmetric matrices.
Notice that $x^TB_s(\tau)x \to -\infty$ for every $x \in \mathbb{R}^n$ when $\tau \to 0^+$, due to the fact that $\Lambda_s(0) = \Lambda$ is out of the coordinate chart. Moreover, a necessary condition for $\Lambda_s(t)$ to be conjugate to $\Lambda$ is that there exists a nonzero $x$ such that $x^TB_s(\tau)x \to \infty$ for $\tau \to t$.

It is then enough to show that, for all $x \in \mathbb{R}^n$ the function $(t,s) \mapsto x^TB_s(t)x$ is bounded. Indeed by assumptions $t \mapsto x^TB_0(t)x$ and $t \mapsto x^TB_1(t)x$ are monotone increasing and bounded up to $t = 1$. Hence the continuous family of values $M_s := x^TB_s(1)x$ is well defined and bounded for all $s$. The monotonicity implies that actually $x^TB_s(t)x < +\infty$ for all values of $t,s \in [0,1]$. (See also Figure 14.7).

![Figure 14.1: Proof of Theorem 14.53](image)

### 14.8 Comparison theorems for regular curves

In this last section we prove two comparison theorems for regular monotone curves in the Lagrange Grassmannian.

**Corollary 14.54.** Let $\Lambda(t)$ be a monotone and regular curve in the Lagrange Grassmannian such that $R_\Lambda(t) \leq 0$. Then $\Lambda(t)$ contains no conjugate points to $\Lambda(0)$.

*Proof.* This is a direct consequence of Theorem 14.52.

**Theorem 14.55.** Let $\Lambda(t)$ be a monotone and regular curve in the Lagrange Grassmannian. Assume that there exists $k \geq 0$ such that for all $t \geq 0$

(i) $R_\Lambda(t) \leq k \text{Id}$. Then, if $\Lambda(t)$ is conjugate to $\Lambda(0)$, we have $t \geq \frac{\sqrt{k}}{\sqrt{n}}$.

(ii) $\frac{1}{n}\text{trace }R_\Lambda(t) \geq k$. Then for every $t \geq 0$ there exists $\tau \in [t, t + \frac{\pi}{\sqrt{k}}]$ such that $\Lambda(\tau)$ is conjugate to $\Lambda(0)$.

300
We stress that assumption (i) means that all the eigenvalues of \( R_\Lambda(t) \) are smaller or equal than \( k \), while (ii) requires only that the average of the eigenvalues is bigger or equal than \( k \).

**Remark 14.56.** Notice that the estimates of Theorem 14.55 are sharp, as it is immediately seen by considering the example of a 1-dimensional curve of constant velocity (see Example 14.35).

**Proof.** (i). Consider the real function

\[
\varphi : \mathbb{R} \to ]0, \frac{\pi}{\sqrt{k}}[, \quad \varphi(t) = \frac{1}{\sqrt{k}}(\arctan \sqrt{kt} + \frac{\pi}{2})
\]

Using that \( \dot{\varphi}(t) = (1 + kt^2)^{-1} \) it is easy to show that the Schwarzian derivative of \( \varphi \) is

\[
R_\varphi(t) = -\frac{k}{(1 + kt^2)^2}.
\]

Thus using \( \varphi \) as a reparametrization we find, by Proposition 14.36

\[
R_\Lambda(\varphi(t)) = \varphi'^2 R_\Lambda(\varphi(t)) + R_\varphi(t) \text{Id}
\]

\[
= \frac{1}{(1 + kt^2)^2} (R_\Lambda(\varphi(t)) - k \text{Id}) \leq 0.
\]

By Corollary 14.54 the curve \( \Lambda \circ \varphi \) has no conjugate points, i.e. \( \Lambda \) has no conjugate points in the interval \( ]0, \frac{\pi}{\sqrt{k}}[ \).

(ii). We prove the claim by showing that the curve \( \Lambda(t) \), on every interval of length \( \pi/\sqrt{k} \) has non trivial intersection with every subspace (hence in particular with \( \Lambda(0) \)). This is equivalent to prove that \( \Lambda(t) \) is not contained in a single coordinate chart for a whole interval of length \( \pi/\sqrt{k} \).

Assume by contradiction that \( \Lambda(t) \) is contained in one coordinate chart. Then there exists coordinates such that \( \Lambda(t) = \{(p, S(t)p)\} \) and we can write the coordinate expression for the curvature:

\[
R_\Lambda(t) = \hat{B}(t) - B(t)^2,
\]

where \( B(t) = (2S(t))^{-1} \tilde{S}(t) \).

Let now \( b(t) := \text{trace } B(t) \). Computing the trace in both sides of equality

\[
\dot{B}(t) = B^2(t) + R_\Lambda(t),
\]

we get

\[
\dot{b}(t) = \text{trace}(B^2(t)) + \text{trace } R_\Lambda(t).
\]

**Lemma 14.57.** For every \( n \times n \) symmetric matrix \( S \) the following inequality holds true

\[
\text{trace}(S^2) \geq \frac{1}{n}(\text{trace } S)^2.
\]

**Proof.** For every symmetric matrix \( S \) there exists a matrix \( M \) such that \( MSM = D \) is diagonal. Since \( \text{trace}(MAM^{-1}) = \text{trace}(A) \) for every matrix \( A \), it is enough to prove the inequality (14.36) for a diagonal matrix \( D = \text{diag}(\lambda_1, \ldots, \lambda_n) \). In this case (14.36) reduces to the Cauchy-Schwartz inequality

\[
\sum_{i=1}^{n} \lambda_i^2 \geq \frac{1}{n} \left( \sum_{i=1}^{n} \lambda_i \right)^2.
\]
Applying Lemma 14.57 to (14.35) and using the assumption (ii) one gets

\[ \dot{b}(t) \geq \frac{1}{n} b^2(t) + nk, \]  

(14.37)

By standard results in ODE theory we have \( b(t) \geq \varphi(t) \), where \( \varphi(t) \) is the solution of the differential equation

\[ \dot{\varphi}(t) = \frac{1}{n} \varphi^2(t) + nk \]  

(14.38)

The solution for (14.38), with initial datum \( \varphi(t_0) = 0 \), is explicit and given by

\[ \varphi(t) = n\sqrt{k} \tan(\sqrt{k}(t - t_0)). \]

This solution is defined on an interval of measure \( \pi/\sqrt{k} \). Thus the inequality \( b(t) \geq \varphi(t) \) completes the proof.
Chapter 15

Jacobi curves

Now we are ready to introduce the main object of this part of the book, i.e. the Jacobi curve associated with a normal extremal. Heuristically, we would like to extract geometric properties of the sub-Riemannian structure by studying the symplectic invariants of its geodesic flow, that is the flow of $\tilde{H}$. The simplest idea is to look for invariants in its linearization.

As we explain in the next sections, this object is naturally related to geodesic variations, and generalizes the notion of Jacobi fields in Riemannian geometry to more general geometric structures.

In this chapter we consider a sub-Riemannian structure $(M, U, f)$ on a smooth $n$-dimensional manifold $M$ and we denote as usual by $H : T^*M \to \mathbb{R}$ its sub-Riemannian Hamiltonian.

15.1 From Jacobi fields to Jacobi curves

Fix a covector $\lambda \in T^*M$, with $\pi(\lambda) = q$, and consider the normal extremal starting from $q$ and associated with $\lambda$, i.e.

$$\lambda(t) = e^{t\tilde{H}}(\lambda), \quad \gamma(t) = \pi(\lambda(t)). \quad \text{(i.e. } \lambda(t) \in T^*_\gamma(t)M).$$

For any $\xi \in T_\lambda(T^*M)$ we can define a vector field along the extremal $\lambda(t)$ as follows

$$X(t) := e^{t\tilde{H}} \xi \in T_{\lambda(t)}(T^*M)$$

The set of vector fields obtained in this way is a $2n$-dimensional vector space which is the space of Jacobi fields along the extremal. For an Hamiltonian $H$ corresponding to a Riemannian structure, the projection $\pi_*$ gives an isomorphisms between the space of Jacobi fields along the extremal and the classical space of Jacobi fields along the geodesic $\gamma(t) = \pi(\lambda(t))$.

Notice that this definition, equivalent to the standard one in Riemannian geometry, does not need curvature or connection, and can be extended naturally for any strongly normal sub-Riemannian geodesic.

In Riemannian geometry, the study of one half of this vector space, namely the subspace of classical Jacobi fields vanishing at zero, carries informations about conjugate points along the given geodesic. By the aforementioned isomorphism, this corresponds to the subspace of Jacobi fields along the extremal such that $\pi_*X(0) = 0$. This motivates the following construction: For
any \( \lambda \in T^*M \), we denote \( V_\lambda := \ker \pi_\lambda \) the vertical subspace. We could study the whole family of (classical) Jacobi fields (vanishing at zero) by means of the family of subspaces along the extremal

\[
L(t) := e^{t\tilde{H}} V_\lambda \subset T_{\lambda(t)}(T^*M).
\]

Notice that actually, being \( e^{t\tilde{H}} \) a symplectic transformation and \( V_\lambda \) a Lagrangian subspace, the subspace \( L(t) \) is a Lagrangian subspace of \( T_{\lambda(t)}(T^*M) \).

15.1.1 Jacobi curves

The theory of curves in the Lagrange Grassmannian developed in Chapter ?? is an efficient tool to study family of Lagrangian subspaces contained in a single symplectic vector space. It is then convenient to modify the construction of the previous section in order to collect the informations about the linearization of the Hamiltonian flow into a family of Lagrangian subspaces at a fixed tangent space.

By definition, the pushforward of the flow of \( \tilde{H} \) maps the tangent space to \( T^*M \) at the point \( \lambda(t) \) back to the tangent space to \( T^*M \) at \( \lambda \):

\[
e^{-t\tilde{H}} : T_{\lambda(t)}(T^*M) \to T_{\lambda}(T^*M).
\]

If we then restrict the action of the pushforward \( e^{-t\tilde{H}} \) to the vertical subspace at \( \lambda(t) \), i.e. the tangent space \( T_{\lambda(t)}(T^*_\gamma M) \) at the point \( \lambda(t) \) to the fiber \( T^*_\gamma M \), we define a one parameter family of \( n \)-dimensional subspaces in the \( 2n \)-dimensional vector space \( T_{\lambda}(T^*M) \). This family of subspaces is a curve in the Lagrangian Grassmannian \( L(T_{\lambda}(T^*M)) \).

**Notation.** In the following we use the notation \( V_\lambda := T_{\lambda}(T^*_q M) \) for the vertical subspace at the point \( \lambda \in T^*M \), i.e. the tangent space at \( \lambda \) to the fiber \( T^*_q M \), where \( q = \pi(\lambda) \). Being the tangent space to a vector space, sometimes it will be useful to identify the vertical space \( V_\lambda \) with the vector space itself, namely \( V_\lambda \simeq T^*_q M \).

**Definition 15.1.** Let \( \lambda \in T^*M \). The Jacobi curve at the point \( \lambda \) is defined as follows

\[
J_{\lambda}(t) := e^{-t\tilde{H}} V_{\lambda(t)},
\]

where \( \lambda(t) := e^{t\tilde{H}}(\lambda) \) and \( \gamma(t) = \pi(\lambda(t)) \). Notice that \( J_{\lambda}(t) \subset T_{\lambda}(T^*M) \) and \( J_{\lambda}(0) = V_\lambda = T_{\lambda}(T^*_q M) \) is vertical.

As discussed in Chapter [14], the tangent vector to a curve in the Lagrange Grassmannian can be interpreted as a quadratic form. In the case of a Jacobi curve \( J_{\lambda}(t) \) its tangent vector is a quadratic form \( \dot{J}_{\lambda}(t) : J_{\lambda}(t) \to \mathbb{R} \).

**Proposition 15.2.** The Jacobi curve \( J_{\lambda}(t) \) satisfies the following properties:

(i) \( J_{\lambda}(t + s) = e^{-t\tilde{H}} J_{\lambda(t)}(s) \), for all \( t, s \geq 0 \),

(ii) \( \dot{J}_{\lambda}(0) = -2H|_{T^*_q M} \) as quadratic forms on \( V_\lambda \simeq T^*_q M \).

(iii) \( \text{rank } \dot{J}_{\lambda}(t) = \text{rank } H|_{T^*_\gamma M} \)
Proof. Claim (i) is a consequence of the semigroup property of the family \( \{ e^{-t\vec{H}} \}_{t \geq 0} \).

To prove (ii), introduce canonical coordinates \((p, x)\) in the cotangent bundle. Fix \( \xi \in \mathcal{V}_\lambda \). The smooth family of vectors defined by \( \xi(t) = e^{-t\vec{H}} \xi \) (considering \( \xi \) as a constant vertical vector field) is a smooth extension of \( \xi \), i.e., it satisfies \( \xi(0) = \xi \) and \( \xi(t) \in J_\lambda(t) \). Therefore, by (14.8)

\[
J_\lambda(0)\xi = \sigma(\xi, \dot{\xi}) = \sigma \left( \xi, \left. \frac{d}{dt} \right|_{t=0} e^{-t\vec{H}} \xi \right) = \sigma(\xi, [\vec{H}, \xi]).
\] (15.2)

To compute the last quantity we use the following elementary, although very useful, property of the symplectic form \( \sigma \).

**Lemma 15.3.** Let \( \xi \in \mathcal{V}_\lambda \) a vertical vector. Then, for any \( \eta \in T_\lambda(T^*M) \)

\[
\sigma(\xi, \eta) = \langle \xi, \pi^*\eta \rangle,
\] (15.3)

where we used the canonical identification \( \mathcal{V}_\lambda = T^*_qM \).

**Proof.** In any Darboux basis induced by canonical local coordinates \((p, x)\) on \( T^*M \), we have \( \sigma = \sum_{i=1}^n dp_i \wedge dx_i \) and \( \xi = \sum_{i=1}^n \xi^i \partial_p^i \). The result follows immediately. \( \square \)

To complete the proof of point (ii) it is enough to compute in coordinates

\[
\pi^*_\eta[\vec{H}, \xi] = \pi^*_\eta \left[ \frac{\partial H}{\partial p} \frac{\partial}{\partial x} - \frac{\partial H}{\partial x} \frac{\partial}{\partial p} \xi \frac{\partial}{\partial p} \right] = -\frac{\partial^2 H}{\partial p^2} \xi \frac{\partial}{\partial x},
\]

Hence by Lemma 15.3 and the fact that \( H \) is quadratic on fibers one gets

\[
\sigma(\xi, [\vec{H}, \xi]) = -\left\langle \xi, \frac{\partial^2 H}{\partial p^2} \xi \right\rangle = -2H(\xi).
\]

(iii). The statement for \( t = 0 \) is a direct consequence of (ii). Using property (i) it is easily seen that the quadratic forms associated with the derivatives at different times are related by the formula

\[
\dot{J}_\lambda(t) \circ e^{t\vec{H}} = \dot{J}_\lambda(t)(0).
\] (15.4)

Since \( e^{-t\vec{H}} \) is a symplectic transformation, it preserves the sign and the rank of the quadratic form. \( \square \)

**Remark 15.4.** Notice that claim (iii) of Proposition 15.2 implies that rank of the derivative of the Jacobi curve is equal to the rank of the sub-Riemannian structure. Hence the curve is regular if and only if it is associated with a Riemannian structure. In this case of course it is strictly monotone, namely \( \dot{J}_\lambda(t) < 0 \) for all \( t \).

**Corollary 15.5.** The Jacobi curve \( J_\lambda(t) \) associated with a sub-Riemannian extremal is monotone nonincreasing for every \( \lambda \in T^*M \).

---

1Notice that \( \dot{J}_\lambda(t), \dot{J}_\lambda(t)(0) \) are defined on \( J_\lambda(t), J_\lambda(t)(0) \) respectively, and \( J_\lambda(t) = e^{-t\vec{H}} J_\lambda(t)(0) \).
15.2 Conjugate points and optimality

At this stage we have two possible definition for conjugate points along normal geodesics. On one hand we have singular points of the exponential map along the extremal path, on the other hand we can consider conjugate points of the associated Jacobi curve. The next result show that actually the two definition coincide.

**Proposition 15.6.** Let $\gamma(t) = E_q(t\lambda)$ be a normal geodesic starting from $q$ with initial covector $\lambda$. Denote by $J_\lambda(t)$ its Jacobi curve. Then for $s > 0$

$$\gamma(s) \text{ is conjugate to } \gamma(0) \iff J_\lambda(s) \text{ is conjugate to } J_\lambda(0).$$

**Proof.** By Definition 8.43, $\gamma(s) \text{ is conjugate to } \gamma(0)$ if $s\lambda$ is a critical point of the exponential map $E_q$. This is equivalent to say that the differential of the map from $T^*_qM$ to $M$ defined by $\lambda \mapsto \pi \circ e^{s\tilde{H}}(\lambda)$ is not surjective at the point $\lambda$, i.e. the image of the differential $e^{s\tilde{H}}$ has a nontrivial intersection with the kernel of the projection $\pi$.

$$e_s^{\tilde{H}}J_\lambda(0) \cap T_{\lambda(s)}T^{*}_{\gamma(s)}M \neq \{0\}. \quad (15.5)$$

Applying the linear invertible transformation $e^{-s\tilde{H}}$ to both subspaces one gets that $(15.5)$ is equivalent to

$$J_\lambda(0) \cap J_\lambda(s) \neq \{0\}$$

which means by definition that $J_\lambda(s)$ is conjugate to $J_\lambda(0)$. \qed

The next result shows that, as soon as we have a segment of points that are conjugate to the initial one, the segment is also abnormal.

**Theorem 15.7.** Let $\gamma: [0,1] \to M$ be a normal extremal path such that $\gamma|_{[0,s]}$ is not abnormal for all $0 < s \leq 1$. Assume $\gamma|_{[t_0,t_1]}$ is a curve of conjugate points to $\gamma(0)$. Then the restriction $\gamma|_{[t_0,t_1]}$ is also abnormal.

**Remark 15.8.** Recall that if a curve $\gamma: [0,T] \to M$ is a strictly normal trajectory, it can happen that a piece of it is abnormal as well. If the trajectory is strongly normal, then if $t_0, t_1$ satisfy the assumptions of Theorem 15.7 necessarily $t_0 > 0$.

**Proof.** Let us denote by $J_\lambda(t)$ the Jacobi curve associated with $\gamma(t)$. From Proposition 15.6 it follows that $J_\lambda(t) \cap J_\lambda(0) \neq \{0\}$ for each $t \in [t_0, t_1]$. We now show that actually this implies

$$J_\lambda(0) \cap \bigcap_{t \in [t_0, t_1]} J_\lambda(t) \neq \{0\}. \quad (15.6)$$

We can assume that the whole piece of the Jacobi curve $J_\lambda(t)$, with $t_0 \leq t \leq t_1$, is contained in a single coordinate chart. Otherwise we can cover $[t_0, t_1]$ with such intervals and repeat the argument on each of them. Let us fix coordinates given by a Lagrangian splitting in such a way that

$$J_\lambda(t) = \{(p,S(t)p), p \in \mathbb{R}^n\}, \quad J_\lambda(0) = \{(p,0), p \in \mathbb{R}^n\}$$
Moreover we can assume that \( S(t) \leq 0 \) for every \( t_0 \leq t \leq t_1 \), i.e. is non positive definite and monotone decreasing. In particular \( J_\lambda(t_1) \cap J_\lambda(0) \neq \{0\} \) if and only if there exists a vector \( v \) such that \( S(t_1)v = 0 \). Since the map \( t \mapsto v^T S(t)v \) is nonpositive and decreasing this means that \( S(t)v = 0 \) for all \( t \in [t_0,t_1] \), thus

\[
J_\lambda(0) \cap J_\lambda(t_1) \subset J_\lambda(0) \cap \bigcap_{t \in [t_0,t_1]} J_\lambda(t)
\]

that implies that actually we have the equality in (15.7).

We are left to show that if a Jacobi curve \( J_\lambda(t) \) is such that every \( t \) is a conjugate point for \( 0 \leq \tau \leq \tau \), then the corresponding extremal is also abnormal. Indeed let us fix an element \( \xi \neq 0 \) such that \( \xi \in \bigcap_{t \in [0,\tau]} J_\lambda(t) \) which is non-empty by the above discussion. Then we consider the vertical vector field

\[
\xi(t) = e^{t\tilde{H}} \xi \in T_\lambda(t)(T^*_\gamma(t)M), \quad 0 \leq t \leq \tau.
\]

By construction, the vector field \( \xi \) is preserved by the Hamiltonian field, i.e. \( e^{t\tilde{H}} \xi = \xi \), that implies \( [\tilde{H}, \xi](\lambda(t)) = 0 \). Then the statement is proved by the following

**Exercise 15.9.** Define \( \eta(t) = \xi(\lambda(t)) \in T^*_\gamma(t)M \) (by canonical identification \( T_\lambda(T^*_\gamma M) \simeq T^*_\gamma M \)). Show that the identity \( [\tilde{H}, \xi](\lambda(t)) = 0 \) rewrites in coordinates as follows

\[
\sum_{i=1}^k h_i(\eta(t))^2 = 0, \quad \dot{\eta}(t) = \sum_{i=1}^k h_i(\lambda(t))\tilde{h}_i(\eta(t)).
\]

Exercise 15.9 shows that \( \eta(t) \) is a family of covectors associated with the extremal path corresponding to controls \( u_i(t) = h_i(\lambda(t)) \) and such that \( h_i(\eta(t)) = 0 \), that means that it is abnormal.

**Corollary 15.10.** Let \( J_\lambda(t) \) be the Jacobi curve associated with \( \lambda \in T^*M \) and \( \gamma(t) = \pi(\lambda(t)) \) the associated sub-Riemannian extremal path. Then \( \gamma|_{[0,\tau]} \) is not abnormal for all \( 0 \leq \tau \leq t \) if and only if \( J_\lambda(\tau) \cap J_\lambda(0) = \{0\} \) for all \( 0 \leq \tau \leq t \).

### 15.3 Reduction of the Jacobi curves by homogeneity

The Jacobi curve at point \( \lambda \in T^*M \) parametrizes all the possible geodesic variations of the geodesic associated with an initial covector \( \lambda \). Since the variations in the direction of the motion are always trivial, i.e. the trajectory remains the same up to parametrizations, one can reduce the space of variation to an \((n-1)\)-dimensional one.

This idea is formalized by considering a reduction of the Jacobi curve in a smaller symplectic space. As we show in the next section, this is a natural consequence of the homogeneity of the sub-Riemannian Hamiltonian.

---

\(^2\)Indeed it is proved that the only invariant of a pair of two Lagrangian subspaces in a symplectic space is the dimension of the intersection, i.e. the rank of the difference \( \text{rank}(S(t) - S(0)) \). Add exercise
Remark 15.11. This procedure was already exploited in Section 8.9, obtained by a direct argument via Proposition 8.37. Indeed one can recognize that the procedure that reduced the equation for conjugate points of one dimension corresponds exactly to the reduction by homogeneity of the Jacobi curve associated to the problem.

We start with a technical lemma, whose proof is left as an exercise.

Lemma 15.12. Let $\Sigma = \Sigma_1 \oplus \Sigma_2$ be a splitting of the symplectic space, with $\sigma = \sigma_1 \oplus \sigma_2$. Let $\Lambda_i \in L(\Sigma_i)$ and define the curve $\Lambda(t) := \Lambda_1(t) \oplus \Lambda_2(t) \in L(\Sigma)$. Then one has the splittings:

$$\dot{\Lambda}(t) = \dot{\Lambda}_1(t) \oplus \dot{\Lambda}_2(t),$$
$$R_\Lambda(t) = R_{\Lambda_1}(t) \oplus R_{\Lambda_2}(t).$$

Consider now a Jacobi curve associated with $\lambda \in T^*M$:

$$J_\lambda(t) = e^{-t\vec{H}}V_\lambda(t), \quad V_\lambda = T_{T^*\pi(\lambda)}M.$$ 

Denote by $\delta_\alpha : T^*M \to T^*M$ the fiberwise dilation $\delta_\alpha(\lambda) = \alpha \lambda$, where $\alpha > 0$.

Definition 15.13. The Euler vector field $\vec{E} \in \text{Vec}(T^*M)$ is the vertical vector field defined by

$$\vec{E}(\lambda) = \frac{d}{ds} \bigg|_{s=1} \delta_s(\lambda), \quad \lambda \in T^*M.$$ 

It is easy to see that in canonical coordinates $(x, \xi)$ it satisfies $\vec{E} = \sum_{i=1}^n \xi_i \frac{\partial}{\partial \xi_i}$ and the following identity holds

$$e^{t\vec{E}} \lambda = e^t \lambda, \quad \text{i.e.} \ e^{t\vec{E}}(\xi, x) = (e^t \xi, x).$$

Exercise 15.14. Prove that the Euler vector field is characterized by the identity

$$i_{\vec{E}} \sigma = s, \quad s = \text{Liouville 1-form in } T^*M.$$ 

Lemma 15.15. We have the identity $e^{-t\vec{H}} \vec{E} = \vec{E} - t\vec{H}$. In particular $[\vec{H}, \vec{E}] = -\vec{H}$.

Proof. The homogeneity property (8.30) of the Hamiltonian can be rewritten as follows

$$e^{t\vec{H}}(\delta_s \lambda) = \delta_s(e^{st\vec{H}}(\lambda)), \quad \forall s, t > 0.$$ 

Applying $\delta_{-s}$ to both sides and changing $t$ into $-t$ one gets the identity

$$\delta_{-s} \circ e^{-t\vec{H}} \circ \delta_s = e^{-st\vec{H}}. \quad (15.9)$$

Computing the 2nd order mixed partial derivative at $(t, s) = (0, 1)$ in (15.9) one gets, by (2.27), that $[\vec{H}, \vec{E}] = -\vec{H}$. Thus, by (??) we have $e^{-t\vec{H}} \vec{E} = \vec{E} - t\vec{H}$, since every higher order commutator vanishes.

\qed

Proposition 15.16. The subspace $\tilde{\Sigma} = \text{span}\{\vec{E}, \vec{H}\}$ is invariant under the action of the Hamiltonian flow. Moreover $\{\vec{E}, \vec{H}\}$ is a Darboux basis on $\Sigma \cap H^{-1}(1/2)$. 

308
Proof. The fact that $\tilde{\Sigma}$ is an invariant subspace is a consequence of the identities
$$e^{-t\tilde{H}}\tilde{E} = \tilde{E} - t\tilde{H}, \quad e^{-t\tilde{H}}\tilde{H} = 0.$$ Moreover, on the level set $H^{-1}(1/2)$, we have by homogeneity of $H$ w.r.t. $p$:
$$\sigma(\tilde{E}, \tilde{H}) = \tilde{E}(H) = \frac{d}{dt} \bigg|_{t=0} H(e^{t\tilde{E}}(p, x)) = p \frac{\partial H}{\partial p} = 2H = 1. \quad (15.10)$$

It follows that $\{\tilde{E}, \tilde{H}\}$ is a Darboux basis for $\tilde{\Sigma}$.

In particular we can consider the the symplectic splitting $\Sigma = \tilde{\Sigma} \oplus \tilde{\Sigma}^\perp$. 

**Exercise 15.17.** Prove the following intrinsic characterization of the skew-orthogonal to $\tilde{\Sigma}$:
$$\tilde{\Sigma}^\perp = \{ \xi \in T^*_\lambda(T^*M) : \langle d\lambda H, \xi \rangle = \langle s_\lambda, \xi \rangle = 0 \}.$$ The assumptions of Lemma 15.12 are satisfied and we could split our Jacobi curve.

**Definition 15.18.** The *reduced Jacobi curve* is defined as follows
$$\hat{J}_\lambda(t) := J_\lambda(t) \cap \tilde{\Sigma}^\perp. \quad (15.11)$$

Notice that, if we put $\hat{\mathcal{V}}_\lambda := \mathcal{V}_\lambda \cap T\lambda H^{-1}(1/2)$, we get
$$\hat{J}_\lambda(0) = \hat{\mathcal{V}}_\lambda, \quad \hat{J}_\lambda(t) = e^{-t\tilde{H}}\hat{\mathcal{V}}_\lambda.$$ Moreover we have the splitting
$$J_\lambda(t) = \hat{J}_\lambda(t) \oplus \mathbb{R}(\tilde{E} - t\tilde{H}).$$

We stress again that $\hat{J}_\lambda(t)$ is a curve of $(n-1)$-dimensional Lagrangian subspaces in the $(2n-2)$-dimensional vector space $\tilde{\Sigma}^\perp$.

**Exercise 15.19.** With the notation above
(i) Show that the curvature of the curve $J_\lambda(t) \cap \tilde{\Sigma}$ in $\tilde{\Sigma}$ is always zero.
(ii) Prove that $J_\lambda(0) \cap J_\lambda(s) \neq \{0\}$ if and only if $\hat{J}_\lambda(0) \cap \hat{J}_\lambda(s) \neq \{0\}$. 

309
Chapter 16

Riemannian curvature

On a manifold, in general there is no canonical method for identifying tangent spaces at different points, (or more generally fibers of a vector bundle at different points). Thus, we have to expect that a notion of derivative for vector fields (or sections of a vector bundle), has to depend on certain choices.

In our presentation we introduce the general notion of *Ehresmann connection* and we then we discuss how this notion is related with the notion of parallel transport and covariant derivative usually introduced in classical Riemannian geometry.

16.1 Ehresmann connection

Given a smooth fiber bundle $E$, with base $M$ and canonical projection $\pi : E \rightarrow M$, we denote by $E_q = \pi^{-1}(q)$ the fiber at the point $q \in M$. The *vertical distribution* is by definition the collection of subspaces in $TE$ that are tangent to the fibers

$$V = \{V_z\}_{z \in E}, \quad V_z := \ker \pi_*, z = T_z E_{\pi(z)} \subset T_z E.$$ 

**Definition 16.1.** Let $E$ be a smooth fiber bundle. An *Ehresmann connection* on $E$ is a smooth vector distribution $H$ in $E$ satisfying

$$H = \{H_z\}_{z \in E}, \quad T_z E = V_z \oplus H_z.$$ 

Notice that $V$, being the kernel of the pushforward $\pi_*$, is canonically associated with the fiber bundle. Defining a connection means exactly to define a canonical complement to this distribution. For this reason $H$ is also called *horizontal distribution*.

**Definition 16.2.** Let $X \in \text{Vec}(M)$. The *horizontal lift* of $X$ is the unique vector field $\nabla_X \in \text{Vec}(E)$ such that

$$\nabla_X(z) \in H_z, \quad \pi_* \nabla_X = X, \quad \forall z \in E.$$ 

The uniqueness follows from the fact that $\pi_* : T_z E \rightarrow T_{\pi(z)} M$ is an isomorphism when restricted to $H_z$. Indeed $\pi_* : T_z E \rightarrow T_{\pi(z)} M$ is a surjective linear map with $\ker \pi_* = V_z$.

**Notation.** In the following we will refer also at $\nabla$ as the connection on $E$.  
Given a smooth curve $\gamma : [0, T] \to M$ on the manifold $M$, the connection lets us to define the parallel transport along $\gamma$, i.e. a way to identify tangent vectors belonging to tangent spaces at different points of the curve. Let $X_t$ be a nonautonomous smooth vector field defined on a neighborhood of $\gamma$, that is an extension of the velocity vector field of the curve, i.e. such that

$$\dot{\gamma}(t) = X_t(\gamma(t)), \quad \forall t \in [0, T].$$

Then consider the non autonomous vector field $\nabla_{X_t} \in \text{Vec}(E)$ obtained by its lift.

**Definition 16.3.** Let $\gamma : [0, T] \to M$ be a smooth curve. The parallel transport along $\gamma$ is the map $\Phi$ defined by the flow of $\nabla_{X_t}$

$$\Phi_{t_0, t_1} := \exp \int_{t_0}^{t_1} \nabla_{X_s} ds : E_{\gamma(t_0)} \to E_{\gamma(t_1)}, \quad \text{for } 0 < t_0 < t_1 < T. \quad (16.2)$$

In the general case we need some extra assumptions on the vector field to ensure that (16.2) exists (even for small time $t > 0$) since the existence time of a solution also depend on the point on the fiber. For instance if we the fibers are compact, then it is possible to find such $t > 0$.

**Exercise 16.4.** Show that the parallel transport map sends fibers to fibers and does not depend on the extension of the vector field $X_t$. (Hint: consider two extensions and use the existence and uniqueness of the flow.)

### 16.1.1 Curvature of an Ehresmann connection

Assume that $\pi : E \to M$ is a smooth fiber bundle and let $\nabla$ be a connection on $E$, defining the splitting $E = V \oplus \mathcal{H}$. Given an element $z \in E$ we will also denote by $z_{\text{hor}}$ (resp. $z_{\text{ver}}$) its projection on the horizontal (resp. vertical) subspace at that point.

The commutator of two vertical vector field is always vertical. The curvature operator associated with the connection computes if the same holds true for two horizontal vector fields.

**Definition 16.5.** Let $E$ be a smooth fiber bundle and $\nabla$ a connection on $E$. Let $X, Y \in \text{Vec}(M)$ and define

$$R(X, Y) := [\nabla_X, \nabla_Y]_{\text{ver}} \quad (16.3)$$

The operator $R$ is called the curvature of the connection.

Notice that, given a vector field on $E$, its horizontal part coincide, by definition, with the lift of its projection. In particular

$$[\nabla_X, \nabla_Y]_{\text{hor}} = \nabla_{[X,Y]}, \quad \text{(i.e. } \pi_*[\nabla_X, \nabla_Y] = [X,Y])$$

Hence $R(X, Y)$ computes the nontrivial part of the bracket between the lift of $X$ and $Y$ and $R \equiv 0$ if and only if the horizontal distribution $\mathcal{H}$ is involutive.

The curvature $R(X, Y)$ is also rewritten in the following more classical way

$$R(X, Y) = [\nabla_X, \nabla_Y] - \nabla_{[X,Y]},$$

$$= \nabla_X \nabla_Y - \nabla_Y \nabla_X - \nabla_{[X,Y]}.$$

Next we show that $R$ is actually a tensor on $T_q M$, i.e. the value of $R(X, Y)$ at $q$ depends only on the value of $X$ and $Y$ at the point $q$.

---

1 this is always possible with a (maybe non autonomous) vector field.
Proposition 16.6. \( R \) is a skew symmetric tensor on \( M \).

Proof. The skew-symmetry is immediate. To prove that the value of \( R(X, Y) \) at \( q \) depends only on the value of \( X \) and \( Y \) at the point \( q \), it is sufficient to prove that \( R \) is linear on functions. By skew-symmetry, we are reduced to prove that \( R \) is linear in the first argument, namely

\[
R(aX, Y) = aR(X, Y), \quad \text{where} \quad a \in C^\infty(M).
\]

Notice that the symbol \( a \) in the right hand side stands for the function \( \pi^* a = a \circ \pi \) in \( C^\infty(E) \), that is constant on fibers.

By definition of lift of a vector field it is easy to prove the identities \( \nabla aX = a\nabla X \) and \( \nabla X a = Xa \) for every \( a \in C^\infty(M) \). Applying the definition of \( \nabla \) and the Leibnitz rule for the Lie bracket one gets

\[
R(aX, Y) = [\nabla aX, \nabla Y] - [\nabla aX, \nabla Y] = a[\nabla X, \nabla Y] - (\nabla Y a)\nabla X - a[\nabla X, \nabla Y] - (Y a)\nabla X = a(\nabla X, \nabla Y) - (Y a)\nabla X - a\nabla [X, Y] + (Y a)\nabla X = aR(X, Y).
\]

\[\square\]

16.1.2 Linear Ehresmann connections

Assume now that \( E \) is a vector bundle on \( M \) (i.e. each fiber \( E_q = \pi^{-1}(q) \) has a natural structure of vector space). In this case it is natural to introduce a notion of linear Ehresmann connection \( \nabla \) on \( E \).

Given a vector bundle \( \pi : E \to M \), we denote by \( C^\infty_L(E) \) the set of smooth functions on \( E \) that are linear on fibers.

Remark 16.7. For a vector bundle \( \pi : E \to M \), the base manifold \( M \) can be considered immersed in \( E \) as the zero section (see also Example 2.41). The “dual” version of this identification is the inclusion \( i : C^\infty(M) \to C^\infty(E) \). Indeed any function in \( C^\infty(M) \) can be considered as a functions in \( C^\infty(E) \) which is constant on fibers, i.e. more precisely \( a \in C^\infty(M) \mapsto \pi^* a \in C^\infty(E) \).

Exercise 16.8. Show that a vector field on \( E \) is the lift of a vector field on \( M \) if and only if, as a differential operator on \( C^\infty(E) \), it maps the subspace \( C^\infty(M) \) into itself.

After this discussion it is natural to give the following definition.

Definition 16.9. A linear connection on a vector bundle \( E \) on the base \( M \) is an Ehresmann connection \( \nabla \) such that the lift \( \nabla_X \) of a vector field \( X \in \text{Vec}(M) \) satisfies the following property: for every \( a \in C^\infty_L(E) \) it holds \( \nabla_X a \in C^\infty_L(E) \).

Remark 16.10. Given a local basis of vector fields \( X_1, \ldots, X_n \) on \( M \) we can build dual coordinates \( (u_1, \ldots, u_n) \) on the fibers of \( E \) defining the functions \( u_i(z) = \langle z, X_i(q) \rangle \) where \( q = \pi(z) \). In this way

\[
E = \{(u, q), q \in M, u \in \mathbb{R}^n\},
\]

313
and the tangent space to $E$ is splitted in $T_zE \simeq T_qM \oplus T_zE_q$. A connection on $E$ is determined by the lift of the vector fields $X_i, i = 1, \ldots, n$ on the base manifold (recall that $\pi_*\nabla X_i = X_i$)

$$\nabla_{X_i} X_i = X_i + \sum_{j=1}^n a_{ij}(u,q)\partial_{u_j}, \quad i = 1, \ldots, n, \quad (16.4)$$

where $a_{ij} \in C^\infty(E)$ are suitable smooth functions. Then $\nabla$ is linear if and only if for every $i, j$ the function $a_{ij}(u,q) = \sum_{k=1}^n \Gamma^k_{ij}(q)u_k$ is linear with respect to $u$.

The smooth functions $\Gamma^k_{ij}$ are also called the Christoffel symbols of the linear connection.

**Exercise 16.11.** Let $\gamma$ be a smooth curve on the manifold such that $\dot{\gamma}(t) = \sum_{i=1}^n v_i(t)X_i(\gamma(t))$. Show that the differential equation $\dot{\xi}(t) = \nabla_{\dot{\gamma}(t)}\xi(t)$ for the parallel transport along $\gamma$ are written as $\dot{u_j} = \sum_{i,k} \Gamma^k_{ij}v_i u_k$ where $(u_1, \ldots, u_n)$ are the vertical coordinates of $\xi$.

Notice that, for a linear connection, the parallel transport is defined by a first order linear (nonautonomous) ODE. The existence of the flow is then guaranteed from standard results from ODE theory. Moreover, when it exists, the map $\Phi_{t_0, t_1}$ is a linear transformation between fibers.

### 16.1.3 Covariant derivative and torsion for linear connections

Once a connection on a linear vector bundle $E$ is given, we have a well defined linear parallel transport map

$$\Phi_{t_0, t_1} := \exp \int_{t_0}^{t_1} \nabla_{X_s} ds : E_{\gamma(t_0)} \to E_{\gamma(t_1)}, \quad \text{for } 0 < t_0 < t_1 < T. \quad (16.5)$$

If we consider the dual map of the parallel transport one can naturally introduce a non autonomous linear flow on the dual bundle (notice the exchange of $t_0, t_1$ in the integral)

$$\Phi^*_{t_0, t_1} := \left(\exp \int_{t_1}^{t_0} \nabla_{X_s} ds \right)^* : E^*_{\gamma(t_0)} \to E^*_{\gamma(t_1)}, \quad \text{for } 0 < t_0 < t_1 < T. \quad (16.6)$$

The infinitesimal generator of this “adjoint” flow defines a linear parallel transport, hence a linear connection, on the dual bundle $E^*$.

In what follows we will restrict our attention to the case of the vector bundle $E = T^*M$ and we assume that a linear connection $\nabla$ on $T^*M$ is given. Notice that, by the above discussion, all the constructions can be equivalently performed on the dual bundle $E^* = TM$.

For every vector field $Y \in \text{Vec}(M)$ let us denote with $Y^* \in C^\infty(T^*M)$ the function

$$Y^*(\lambda) = \langle \lambda, Y(q) \rangle, \quad q = \pi(z),$$

namely the smooth function on $E$ associated with $Y$ that is linear on fibers. This identification between vector fields on $M$ and linear functions on $T^*M$ permits us to define the **covariant derivative** of vector fields.

**Definition 16.12.** Let $X, Y \in \text{Vec}(M)$. We define $\nabla_X Y = Z$ if and only if $\nabla_X Y^* = Z^*$ with $Z \in \text{Vec}(M)$.  

314
Notice that the definition is well-posed since $\nabla$ is linear, hence $\nabla X Y^*$ is a linear function and there exists $Z \in \text{Vec}(M)$ such that $\nabla X Y^* = Z^2$. 

**Lemma 16.13.** Let $\{X_1, \ldots, X_n\}$ be a local frame on $M$. Then $\nabla X_i X_j = \Gamma^k_{ij} X_k$, where $\Gamma^k_{ij}$ are the Christoffel symbols of the connection $\nabla$.

**Proof.** Let us prove this in the coordinates dual to our frame. In these coordinates, the linear connection is specified by the lifts 

$$\nabla X_i = X_i + \Gamma^k_{ij} u_k \partial_{u_j}, \quad \text{where} \quad u_j(\lambda) = \langle \lambda, X_j \rangle.$$ 

Moreover $X_j^* = u_j$. Hence it is immediate to show $\nabla X_i X_j^* = \Gamma^k_{ij} X_k^*$, and the lemma is proved. \( \square \)

We now introduce the torsion tensor of a linear connection on $T^* M$. As usual, $\sigma$ denotes the canonical symplectic structure on $T^* M$.

**Definition 16.14.** The **torsion** of a linear connection $\nabla$ is the map $T: \text{Vec}(M)^2 \to \text{Vec}(M)$ defined by the identity

$$T(X,Y)^* := \sigma(\nabla X, \nabla Y), \quad \forall X,Y \in \text{Vec}(M).$$

(16.7)

It is easy to check that $T$ is actually a tensor, i.e. the value of $T(X,Y)$ at a point $q$ depends only on the values of $X,Y$ at the point. The torsion computes how much the horizontal distribution $\mathcal{H}$ is far from being Lagrangian. In particular $\mathcal{H}$ is Lagrangian if and only if $T \equiv 0$.

The classical formula for the torsion tensor, in terms of the covariant derivative, is recovered in the following lemma.

**Lemma 16.15.** The torsion tensor satisfies the identity

$$T(X,Y) = \nabla X Y - \nabla Y X - [X,Y].$$

(16.8)

**Proof.** We have to prove that $T(X,Y)^* = \nabla X Y^* - \nabla Y X^* - [X,Y]^*$. Notice that by definition of the Liouville 1-form $s \in \Lambda^1(T^*M)$, $s_\lambda = \lambda \circ \pi^*$ we have $X^*(\lambda) = \langle \lambda, X \rangle = \langle s_\lambda, \nabla X \rangle$. Then we have, using that $\sigma = ds$ and the Cartan formula [4.74]

$$T(X,Y)^* = ds(\nabla X, \nabla Y) = \nabla X \langle s, \nabla Y \rangle - \nabla Y \langle s, \nabla X \rangle - \langle s, [\nabla X, \nabla Y] \rangle = \nabla X Y^* - \nabla Y X^* - [X,Y]^*,$$

where in the second equality we used that $\langle s, [\nabla X, \nabla Y] \rangle = \langle s, [\nabla X, \nabla Y]_{\text{hor}} \rangle = \langle s, \nabla [X,Y] \rangle$ since the Liouville form by definition depends only on the horizontal part of the vector. \( \square \)

**Exercise 16.16.** Show that a linear connection $\nabla$ on a vector bundle $E$ satisfies the following Leibnitz rule

$$\nabla_X (aY) = a \nabla_X Y + (Xa) Y, \quad \text{for each} \ a \in \mathcal{C}^\infty(M).$$

\(^{2}\)There is no confusion in the notation above since, by definition, $\nabla_X$ it is well defined when applied to smooth functions on $T^* M$. Whenever it is applied to a vector field we follow the aforementioned convention.
16.2 Riemannian connection

In this section we want to introduce the Levi-Civita connection on a Riemannian manifold $M$ by defining an Ehresmann connection on $T^*M$ via the Jacobi curve approach.

Recall that every Jacobi curve associated with a trajectory on a Riemannian manifold is regular. Moreover, as showed in Chapter 14, every regular curve in the Lagrangian Grassmannian admits a derivative curve, which defines a canonical complement to the curve itself. Hence, following this approach, it is natural to introduce the Riemannian connection at $\lambda \in T^*M$ as the canonical complement to the Jacobi curve defined at $\lambda$.

**Definition 16.17.** The **Levi-Civita connection** on $T^*M$ is the Ehresmann connection $\mathcal{H}$ defined by

$$\mathcal{H}_\lambda = J_\lambda'(0), \quad \lambda \in T^*M,$$

where as usual $J_\lambda(t)$ denotes the Jacobi curve defined at the point $\lambda \in T^*M$ and $J_\lambda'$ denotes its derivative curve.

The next proposition characterizes the Levi-Civita connection as the unique linear connection on $T^*M$ that is linear, metric preserving and torsion free.

**Proposition 16.18.** The Levi-Civita connection satisfies the following properties:

(i) is a linear connection,

(ii) is torsion free,

(iii) is metric preserving, i.e. $\nabla_X H = 0$ for each vector field $X \in \text{Vec}(M)$.

**Proof.** (i). It is enough to prove that the connection $\mathcal{H}_\lambda$ is 1-homogeneous, i.e.

$$\mathcal{H}_{c\lambda} = \delta_c \circ \mathcal{H}_\lambda, \quad \forall c > 0.$$  \hspace{1cm} (16.9)

Indeed in this case the functions $a_{ij} \in C^\infty(T^*M)$ defining the connection (see (16.4)) are 1-homogeneous, hence linear as a consequence of Exercise 16.19.

Let us prove (16.9). The differential of the dilation on the fibers $\delta_c : T^*M \to T^*M$ satisfies the property $\delta_{c*}(T_\lambda(T^*_qM)) = T_{c\lambda}(T^*_qM)$. From this identity and differentiating the identity

$$e^{t\vec{H}} \circ \delta_c = \delta_c \circ e^{t\vec{H}}, \quad \forall c > 0,$$  \hspace{1cm} (16.10)

one easily gets that

$$J_{c\lambda}(t) = \delta_{c*} J_\lambda(ct), \quad \forall t \geq 0, \lambda \in T^*M.$$  \hspace{1cm} (16.11)

Indeed one has the following chain of identities

$$J_{c\lambda}(t) = e^{-t\vec{H}}(T_{c\lambda}(T^*_qM))$$
$$= e^{-t\vec{H}} \circ \delta_{c*}(T_\lambda(T^*_qM)) \quad \text{(by (16.10))}$$
$$= \delta_{c*} \circ e^{-t\vec{H}}(T_\lambda(T^*_qM))$$
$$= \delta_{c*} J_\lambda(ct).$$
Now we show that the same relation holds true also for the derivative curve, i.e.

\[ J^c_\lambda(t) = \delta_c J^\lambda_t, \quad \forall t \geq 0, \lambda \in T^*M. \] (16.12)

Indeed one can check in coordinates (we denote as usual \( J^\lambda_t = \{ (p, S^\lambda_t(p)), p \in \mathbb{R}^n \}) that the identity (16.11) is written as \( S^\lambda_t = \frac{1}{c} S^\lambda(ct) \) thus \( S^\lambda_t = c S^\lambda_t(t)^{-1}. \) From here one also gets \( B^\lambda(t) = c B^\lambda(ct) \) and (16.12) follows from the identity \( S^\lambda_t = B^{-1}(t) + S(t). \) (See also Exercise 14.22). In particular at \( t = 0 \) the identity (16.12) says that \( H^\lambda = \delta_c H^\lambda. \)

(ii). It is a direct consequence of the fact that \( J^\lambda_0 \) is a Lagrangian subspace of \( T^\lambda(T^*M) \) for every \( \lambda \in T^*M, \) hence the symplectic form vanishes when applied to two horizontal vectors.

(iii). Again, for every \( X \in \text{Vec}(M), \) both \( \nabla X \) and \( \vec{H} \) are horizontal vector field. Since the horizontal space is Lagrangian \( \nabla X = 0. \)

Exercise 16.19. Let \( f : \mathbb{R}^n \to \mathbb{R} \) be a smooth function that satisfies \( f(\alpha x) = \alpha f(x) \) for every \( x \in \mathbb{R}^n \) and \( \alpha \geq 0. \) Then \( f \) is linear.

The following theorem says that a connection satisfying the three properties above is unique. Then it characterize the Levi-Civita connection in terms of the structure constants of the Lie algebra defined by an orthonormal frame.

Theorem 16.20. There is a unique Ehresmann connection \( \nabla \) satisfying the properties (i), (ii), and (iii) of Proposition 16.18, that is the Levi-Civita connection. Its Christoffel symbols are computed by

\[ \Gamma^k_{ij} = \frac{1}{2} (c^k_{ij} - c^k_{j i} + c^k_{i j}), \] (16.13)

where \( c^k_{ij} \) are the smooth functions defined by the identity \( [X_i, X_j] = \sum_{k=1}^n c^k_{ij} X_k. \)

Proof. Let \( X_1, \ldots, X_n \) be a local orthonormal frame for the Riemannian structure and let us consider coordinates \((q, u)\) in \( T^*M, \) where the fiberwise coordinates \( u = (u_1, \ldots, u_n) \) are dual to the orthonormal frame. From the linearity of the connection it follows that there exist smooth functions \( \Gamma^k_{ij} : M \to \mathbb{R} \) (depending on \( q \) only) such that

\[ \nabla_{X_i} = X_i + \sum_{j=1}^n \Gamma^k_{ij} u_k \partial_{u_j}, \quad i = 1, \ldots, n. \]

In particular \( \nabla_{X_i} X_j = \Gamma^k_{ij} X_k. \) In these coordinates the Hamiltonian vector field associated with the Riemannian Hamiltonian \( H = \frac{1}{2} \sum_{i=1}^n u_i^2 \) reads (see also Exercise ??)

\[ \vec{H} = \sum_{i,j,k=1}^n u_i X_i + c^k_{ij} u_i u_k \partial_{u_j}, \]

while the symplectic form \( \sigma \) is written \((\nu_1, \ldots, \nu_n \) denotes the dual basis to \( X_1, \ldots, X_n)\)

\[ \sigma = \sum_{i,j,k=1}^n du_k \wedge \nu_k - c^k_{ij} u_k \nu_i \wedge \nu_k. \]

\footnote{Recall that \( B \) is the zero order term of the expansion of \( S^{-1}. \)}
Since the horizontal space is Lagrangian, one has the relations
\[
0 = \sigma(\nabla X_i, \nabla X_j) = \sum_{k=1}^{n} (\Gamma^k_{ij} - \Gamma^k_{ji} - c^k_{ij}) u_k, \quad \forall i, j = 1, \ldots, n,
\]
hence \(c^k_{ij} = \Gamma^k_{ij} - \Gamma^k_{ji}\) for all \(i, j, k\). Moreover the connection is metric, i.e. it satisfies
\[
0 = \nabla X_i H = \sum_{j,k=1}^{n} \Gamma^k_{ij} u_k u_j, \quad \forall i = 1, \ldots, n.
\]
The last identity implies that \(\Gamma^k_{ij}\) is skew-symmetric with respect to the pair \((j, k)\), i.e. \(\Gamma^k_{ij} = -\Gamma^j_{ik}\). Thus combining the two identities one gets
\[
c^k_{ij} - c^j_{ik} + c^i_{jk} = (\Gamma^k_{ij} - \Gamma^k_{ji}) - (\Gamma^i_{jk} + \Gamma^j_{ki}) + (\Gamma^j_{ik} - \Gamma^j_{ik})
= \Gamma^k_{ij} - \Gamma^j_{ik} = 2\Gamma^k_{ij}.
\]

Remark 16.21. Notice that in the classical approach one can recover formula [16.13] from the following particular case of the Koszul formula
\[
\Gamma^k_{ij} = g(\nabla X_i, X_j, X_k) = \frac{1}{2} (g([X_i, X_j], X_k) - g([X_j, X_k], X_i) + g([X_k, X_i], X_j)),
\]
that holds for every orthonormal basis \(X_1, \ldots, X_n\). Notice also that the Hamiltonian vector field is written in coordinates \(\vec{H} = \sum_{i=1}^{n} u_i \nabla X_i\), which gives another proof of the fact that it is horizontal.

Let \(X, Y, Z, W \in \text{Vec}(M)\). We define \(R(X, Y)Z = W\) if \(R(X, Y)Z^* = W^*\).

**Proposition 16.22** (Bianchi identity). For every \(X, Y, Z \in \text{Vec}(M)\) the following identity holds
\[
R(X, Y)Z + R(Y, Z)X + R(Z, X)Y = 0.
\] (16.14)

**Proof.** We will show that [16.14] is a consequence of the Jacobi identity [2.30]. Using that \(\nabla\) is a torsion free connection we can write
\[
[X, [Y, Z]] = \nabla_X [Y, Z] - \nabla_{[Y, Z]} X
= \nabla_X \nabla_Y Z - \nabla_Y \nabla_Z X - \nabla_{[Y, Z]} X,
\]
\[
[Z, [X, Y]] = \nabla_Z \nabla_X Y - \nabla_Z \nabla_Y X - \nabla_{[X, Y]} Z,
\]
\[
[Y, [Z, X]] = \nabla_Y \nabla_Z X - \nabla_Y \nabla_X Z - \nabla_{[Z, X]} Y,
\]
Then
\[
0 = [X, [Y, Z]] + [Y, [Z, X]] + [Z, [X, Y]]
= \nabla_X \nabla_Y Z - \nabla_X \nabla_Z Y - \nabla_{[Y, Z]} X
+ \nabla_Z \nabla_X Y - \nabla_Z \nabla_Y X - \nabla_{[X, Y]} Z
+ \nabla_Y \nabla_Z X - \nabla_Y \nabla_X Z - \nabla_{[Z, X]} Y
= R(X, Y)Z + R(Y, Z)X + R(Z, X)Y.
\]
\[\square\]
Exercise 16.23. Prove the second Bianchi identity


(Hint: Expand the identity \(\nabla_{[X,[Y,Z]]} + [Y,[Z,X]] + [Z,[X,Y]] W = 0\).

Let us denote \((X, Y, Z, W) := \langle R(X, Y)Z, W \rangle\). Following this notation, the first Bianchi identity can be rewritten as follows:

\[(X, Y, Z, W) + (Z, X, Y, W) + (Y, Z, X, W) = 0, \quad \forall X, Y, Z, W \in \text{Vec}(M). \quad (16.15)\]

Remark 16.24. The property of the Riemann tensor can be reformulated as follows


Proposition 16.25. For every \(X, Y, Z, W \in \text{Vec}(M)\) we have \((X, Y, Z, W) = (Z, W, X, Y)\).

Proof. Using (16.15) four times we can write the identities

\[(Y, Z, W, X) + (W, Y, Z, X) + (Z, W, Y, X) = 0,\]
\[(W, X, Y, Z) + (Y, W, X, Z) + (X, Y, W, Z) = 0.\]

Summing all together and using the skew symmetry (16.16), one gets \((X, Z, W, Y) = (W, Y, X, Z)\).

Proposition 16.26. Assume that \((X, Y, X, W) = 0\) for every \(X, Y, W \in \text{Vec}(M)\). Then

\[(X, Y, Z, W) = 0 \quad \forall X, Y, Z, W \in \text{Vec}(M). \]

Proof. By assumptions and the skew-simmetry properties (16.16) of the Riemann tensor we have that \((X, Y, Z, W) = 0\) whenever any two of the vector fields coincide. In particular

\[0 = (X, Y + W, Z, Y + W) = (X, Y, Z, W) + (X, W, Z, Y). \quad (16.17)\]

since the two extra terms that should appear in the expansion vanish by assumptions. Then (16.17) can be rewritten as

\[(X, Y, Z, W) = (Z, X, Y, W),\]

i.e. the quantity \((X, Y, Z, W)\) is invariant by ciclic permutations of \(X, Y, Z\). But the cyclic sum of terms is zero by (16.15), hence \((X, Y, Z, W) = 0. \quad \square\)

We end this section by summarizing the symmetry property of the Riemann curvature as follows

Corollary 16.27. There is a well defined map

\[\overline{R} : \wedge^2 T_qM \to \wedge^2 T_qM, \quad \overline{R}(X \wedge Y) := R(X, Y).\]

Moreover \(\overline{R}\) is skew-adjoint with respect to the induced scalar product on \(\wedge^2 T_qM\), that means

\[\langle \overline{R}(X \wedge Y), Z \wedge W \rangle = \langle X \wedge Y, \overline{R}(Z \wedge W) \rangle.\]
16.3 Relation with Hamiltonian curvature

In this section we compute the curvature of the Jacobi curve associated with a Riemannian geodesic and we describe the relation with the Riemann curvature discussed in the previous section. As we show, the curvature associated to a geodesic is a kind of sectional curvature operator in the direction of the geodesic itself.

Definition 16.28. The Hamiltonian curvature tensor at $\lambda \in T^*M$ is the operator

$$ R_\lambda := R_{J_\lambda(0)} : \mathcal{V}_\lambda \to \mathcal{V}_\lambda. $$

In other words $R_\lambda$ is the curvature of the Jacobi curve associated with $\lambda$ at $t = 0$.

Proposition 16.29. Let $\xi \in \mathcal{V}_\lambda$ and $V$ be a smooth vertical vector field extending $\xi$. Then

$$ R_\lambda(\xi) = -[\vec{H}, [\vec{H}, V]_{\text{hor}}]_{\text{ver}}(\lambda) $$

Proof. This is a direct consequence of Proposition 14.30. Indeed recall that the curvature of the Jacobi curve is expressed through the composition

$$ R_\lambda = \dot{J}_\lambda(0) \circ J_\lambda(0). $$

Moreover, being $J_\lambda(0) = \mathcal{V}_\lambda$ and $J_\lambda^\circ(0) = \mathcal{H}_\lambda$ we have that

$$ \pi_{J(0),J^\circ(0)}(\xi) = \xi_{\text{hor}}, \quad \pi_{J^\circ(0),J(0)}(\eta) = \eta_{\text{ver}}. $$

Finally we can extend vectors in $J_\lambda(0)$ (resp. $J_\lambda^\circ(0)$) by applying the Hamiltonian vector field since $J_\lambda(t) = e^{t\vec{H}} J_\lambda(0)$ (resp. $J_\lambda^\circ(t) = e^{t\vec{H}} J_\lambda^\circ(0)$). From these remarks we obtain the following formulas

$$ \dot{J}_\lambda(0) \xi = [\vec{H}, V]_{\text{hor}}, \quad \dot{J}_\lambda^\circ(0) \eta = -[\vec{H}, W]_{\text{ver}} $$

for some $V$ vertical (resp. $W$ horizontal) extension of the vector $\xi \in \mathcal{V}_\lambda$ (reps. $\eta \in \mathcal{H}_\lambda$). \hfill \Box

Another immediate property of the curvature tensor is the homogeneity with respect to the rescaling of the covector (that corresponds to reparametrization of the trajectory). Indeed by choosing $\varphi(t) = ct$, with $c > 0$, in Proposition 14.30 one gets

Corollary 16.30. For every $c > 0$ we have $R_{c\lambda} = c^2 R_\lambda$.

If we use the Riemannian product to identify the tangent and the cotangent space at a point, we recognize that $R_\lambda$ is nothing but the sectional curvature operator where one entry is the tangent vector $\dot{\gamma}$ of the geodesic.

Let us denote by $I : TM \to T^*M$ the isomorphism defined by the Riemannian scalar product $\langle \cdot | \cdot \rangle$. In particular $I(v) = \lambda$ for $\lambda \in T^*_qM$ and $v \in T_qM$ if $\langle \lambda, w \rangle = \langle v, w \rangle$ for all $w \in T_qM$.

Let denote $H_q = H|_{T^*_qM}$. Recall that the differential of $H_q$ can be interpreted as a linear map $DH_q : T^*_qM \to T_qM$ that sends $\lambda \in T^*_qM$ into $D\lambda H_q$ seen as a linear functional on $T_q^*M$, i.e. a tangent vector. This map is actually the inverse of the isomorphism $I$.

Lemma 16.31. $D\lambda H_q = I^{-1}(\lambda)$.

Proof. It is a simple consequence of the formula $H(\lambda) = \frac{1}{2} \langle \lambda, I^{-1}(\lambda) \rangle$. \hfill \Box
Corollary 16.32. Assume $I(v) = \lambda$, then $\vec{H}(\lambda) = \nabla v$.

Proof. Indeed, since $\vec{H}$ is an horizontal vector field, it is sufficient to show that $\pi_* \vec{H}(\lambda) = v$, which is a consequence of Lemma 16.31. Indeed for every vertical vector $\xi \in T_\lambda(T_q^* M)$ one has

$$
\langle \xi, v \rangle = \langle \xi, I^{-1}(\lambda) \rangle = D_\lambda H(\xi) = \sigma(\xi, \vec{H}(\lambda)) = \left\langle \xi, \pi_* \vec{H}(\lambda) \right\rangle.
$$

By arbitrary of $\xi \in T_\lambda(T_q^* M)$ one has the equality $v = \pi_* \vec{H}(\lambda)$. \qed

Theorem 16.33. We have the following identity

$$
R(I(X))(I(Y)) = R(X,Y)X, \quad \forall X,Y \in T_q M.
$$

(16.18)

Proof. We have to compute the quantity

$$
R(I(X))(I(Y)) = -[\vec{H}, [\vec{H}, I(Y)]_{\text{hor}}]_{\text{ver}}(I(X))
$$

First notice that $\pi_* [\vec{H}, I(Y)] = -Y$ hence $[\vec{H}, I(Y)]_{\text{hor}} = -\nabla Y$. Then

$$
-[\vec{H}, [\vec{H}, I(Y)]_{\text{hor}}]_{\text{ver}}(I(X)) = [\nabla X, \nabla Y]_{\text{ver}}(I(X)) = R(X,Y)(X).
$$

\qed

Definition 16.34. The Ricci tensor at $\lambda$ is defined as the trace of the curvature operator at $\lambda$,

$$
\text{Ric}(\lambda) := \text{trace } R_{\lambda}.
$$

Exercise 16.35. Prove the following expression for the Ricci tensor, where $X_1, \ldots, X_n$ is a local orthonormal frame and $\dot{\gamma}(0) = v = I^{-1}(\lambda)$ is the tangent vector to the geodesic:

$$
\text{Ric}(\lambda) = \sum_{i=1}^{n} \langle R(v, X_i)v | X_i \rangle
$$

$$
= \sum_{i=1}^{n} \sigma_{\lambda}([\vec{H}, \nabla X_i], \nabla X_i).
$$

This shows that $\text{Ric}(\lambda) = \text{Ric}(v)$ coincide with the classical Riemannian Ricci tensor.

16.4 Locally flat spaces

In this section we want to show that the Riemannian curvature is the only obstruction for a Riemannian manifold to be locally Euclidean. Finally we show that the Riemannian curvature is also completely recovered by the Hamiltonian curvature $R_{\lambda}$.

A Riemannian manifold $M$ is called flat if $R(X,Y) = 0$ for every $X,Y \in \text{Vec}(M)$.

Theorem 16.36. $M$ is flat if and only if $M$ is locally isometric to $\mathbb{R}^n$. 

321
Proof. If \( M \) is locally isometric to \( \mathbb{R}^n \), then its curvature tensor at every point in a neighborhood is identically zero.

Then let us assume that the Riemann tensor \( R \) vanishes identically and prove that \( M \) is locally Euclidean. We will do that by showing that there exists coordinate such that the Hamiltonian, in these set of coordinates, is written as the Hamiltonian of the Euclidean \( \mathbb{R}^n \).

Since \( R \) is identically zero the horizontal distribution (defined by the Levi Civita connection) is involutive. Hence, by Frobenius theorem, there exists a horizontal Lagrangian foliation of \( T^*M \), i.e. for each \( \lambda \in T^*M \), there exists a leaf \( \mathcal{L}_\lambda \) of the foliation passing through this point that is tangent to the horizontal space \( \mathcal{H}_\lambda \). In particular each leaf is transversal to the fiber \( T^*_qM \), where \( q = \pi(\lambda) \).

Fix a point \( q_0 \in M \) and a neighborhood \( O_{q_0} \) where \( R \) is identically zero. Define the map

\[
\Psi : \pi^{-1}(O_{q_0}) \rightarrow T^*_{q_0}M, \quad \lambda \in \pi^{-1}(O_{q_0}) \mapsto \mathcal{L}_\lambda \cap T^*_{q_0}M
\]

that assigns to each \( \lambda \) the intersection of the leaf passing through this point and \( T^*_{q_0}M \).

Exercise 16.37. Show that \( \Psi \) is a linear, orthogonal transformation, i.e. \( H(\Psi(\lambda)) = H(\lambda) \) for all \( \lambda \in \pi^{-1}(O_{q_0}) \). (Hint: use the linearity of the connection and the fact that \( \ddot{H} \) is horizontal).

Fix now a basis \( \{\nu_1, \ldots, \nu_n\} \) in \( T^*_{q_0}M \) that is orthonormal (with respect to the dual metric). Being \( \Psi \) linear on fibers, we can write

\[
\Psi(\lambda) = \sum_{i=1}^n \psi_i(\lambda)\nu_i, \quad \text{where} \quad \psi_i(\lambda) = \langle \lambda, X_i(q) \rangle
\]

for a suitable basis of vector fields \( X_1, \ldots, X_n \) in the neighborhood \( O_{q_0} \). Moreover \( X_1, \ldots, X_n \) is an orthonormal basis since \( \Psi \) is an orthogonal map.

We want to show that \( \{X_1, \ldots, X_n\} \) is an orthonormal basis of vector fields that commutes everywhere.

Let us show that the fact that the foliation is Lagrangian implies \( [X_i, X_j] = 0 \) for all \( i,j = 1, \ldots, n \).

Indeed the tautological 1-form is written in these coordinates as \( s = \sum_{i=1}^n \psi_i \nu_i \) and

\[
\sigma = ds = \sum_{i=1}^n d\psi_i \wedge \nu_i + \psi_i d\nu_i. \tag{16.19}
\]

Since on each leaf the function \( \psi_i \) is constant by definition (hence \( d\psi_i|_\mathcal{L} = 0 \)), we have that \( \sigma|_\mathcal{L} = \sum_{i} \psi_i d\nu_i \). In particular each leaf is Lagrangian if and only if \( d\nu_i = 0 \) for \( i = 1, \ldots, n \). Then, from the Cartan formula, one gets

\[
0 = d\nu_i(X_j, X_k) = -\nu_i([X_j, X_k]), \quad \forall i,j,k.
\]

This proves that \( [X_i, X_j] = 0 \) for each \( i,j = 1, \ldots, n \). Hence, in the coordinate set \( (\psi, q) \), we have \( H(\psi, q) = \frac{1}{2} |\psi|^2 \). \( \square \)

The next result shows that the Hamiltonian curvature can detect if a manifold is flat or not.

Corollary 16.38. \( M \) is flat if and only if \( R_\lambda = 0 \) for every \( \lambda \in T^*M \).
Proof. Assume that $M$ is flat. Then $R$ is identically zero and a fortiori $R_\lambda = 0$ from (16.18).

Let us prove the converse. Recall that $R_\lambda = 0$ implies, again by (16.18), that

$$(X,Y,X,W) = 0, \quad \forall X,Y,W \in \text{Vec}(M).$$

Then the statement is a consequence of Proposition 16.26.

Exercise 16.39. Prove that actually the Riemann tensor $R$ is completely determined by $\mathcal{R}$.

16.5 Example: curvature of the 2D Riemannian case

In this section we apply the definition of curvature discussed in this chapter to a two dimensional Riemannian surface. As we explain, we recover that the Riemannian curvature tensor is determined by the Gauss curvature of the manifold.

Let $M$ be a 2-dimensional surface and $f_1, f_2 \in \text{Vec}(M)$ be a local orthonormal frame for the Riemannian metric. The Riemannian Hamiltonian $H$ is written as follows (we use canonical coordinates $\lambda = (p, x)$ on $T^*M$

$$H(p, x) = \frac{1}{2}((p, f_1(x))^2 + (p, f_2(x))^2) \quad (16.20)$$

Here, for a covector $\lambda = (p, x) \in T^*M$, the symplectic vector space $\Sigma_\lambda = T_{\lambda}(T^*_x M)$ is 4-dimensional.

Recall that, being $M$ 2-dimensional, the level set $H^{-1}(1/2) \cap T^*_x M$ is a circle. Hence, there is a well defined vector field that produces rotation on the reduced fiber. Let us define the angle $\theta$ on the level $H^{-1}(1/2) \cap T^*_x M$ by setting

$$\langle p, f_1(x) \rangle = \cos \theta, \quad \langle p, f_2(x) \rangle = \sin \theta,$$

in such a way that $\theta = 0$ corresponds to the direction of $f_1$. Denote by $\partial_\theta$ the rotation in the fiber of the unit tangent bundle and by $\vec{E}$, the Euler vector field. Denote finally by $\vec{H}' := [\partial_\theta, \vec{H}]$.

Notice that $\Sigma_\lambda = \mathcal{V}_\lambda \oplus \mathcal{H}_\lambda$ where $\mathcal{V}_\lambda = \text{span}\{\vec{E}, \partial_\theta\}$ and $\mathcal{H}_\lambda = \text{span}\{\vec{H}, \vec{H}'\}$.

Lemma 16.40. The vector fields $\{\vec{E}, \partial_\theta, \vec{H}, \vec{H}'\}$ at $\lambda$ form a Darboux basis for $\Sigma_\lambda$.

Proof. We want to compute the following symplectic products of the vector fields:

$$\sigma(\partial_\theta, \vec{E}) = 0, \quad \sigma(\partial_\theta, \vec{H}) = 0, \quad \sigma(\vec{E}, \vec{H}) = 1. \quad (16.21)$$

$$\sigma(\partial_\theta, \vec{H}') = 1, \quad \sigma(\vec{E}, \vec{H}') = 0, \quad \sigma(\vec{H}, \vec{H}') = 0. \quad (16.22)$$

Indeed, let us prove first (16.21). The first equality follows from the fact that both vectors belong to the vertical subspace, that is Lagrangian. The second one is a consequence of the fact that, by construction, $\partial_\theta$ is tangent to the level set of $H$, i.e. $\sigma(\partial_\theta, \vec{H}) = \partial_\theta(\vec{H}) = \langle dH, \partial_\theta \rangle = 0$. The last identity is (15.10).

As a preliminary step for the proof of (16.22) notice that, if $s = i_{\vec{E}} \sigma$ denotes the tautological Liouville form, one has

$$\langle s, \vec{H} \rangle = 1, \quad \langle s, \vec{H}' \rangle = 0. \quad (16.23)$$
These two identities follows from
\[ \langle s, \vec{H} \rangle = \sigma(\vec{E}, \vec{H}) = 1, \]  
(16.24)\[ \langle s, \vec{H}' \rangle = \langle s, [\partial_\theta, \vec{H}] \rangle = ds(\partial_\theta, \vec{H}) = \sigma(\partial_\theta, \vec{H}) = 0, \]  
(16.25)where in the second line we used the Cartan formula (4.74) and the fact that \( \partial_\theta \) is vertical.

Let us now prove (16.22). Being \( [\partial_\theta, \vec{H}'] = [\partial_\theta, [\partial_\theta, \vec{H}]] = -\vec{H} \), we have again by Cartan formula and (16.23)\[ \sigma(\partial_\theta, \vec{H}') = ds(\partial_\theta, \vec{H}') = \langle s, [\partial_\theta, \vec{H}'] \rangle = \langle s, \vec{H} \rangle = \sigma(\vec{E}, \vec{H}) = 1 \]
Moreover by (16.23)\[ \sigma(\vec{E}, \vec{H}') = \langle s, \vec{H}' \rangle = 0. \]
The last computation is similar. Let us write\[ \sigma(\vec{H}, \vec{H}') = \langle d\vec{H}, \vec{H}' \rangle = \langle d\vec{H}, [\partial_\theta, \vec{H}] \rangle, \]
and apply the Cartan formula to the last term (with \( d\vec{H} \) as 1-form).\[ d\vec{H}([\partial_\theta, \vec{H}]) = d^2\vec{H}(\partial_\theta, \vec{H}) - \partial_\theta(d\vec{H}, \vec{H}) + \vec{H}(d\vec{H}, \partial_\theta) = 0 \]
since the three terms are all equal to zero.

Now we compute the curvature via the Jacobi curve, reduced by homogeneity. Notice that by Lemma 16.40 we can remove the symplectic space spanned by \( \{ \vec{E}, \vec{H} \} \) and, being \( \{ \vec{E}, \vec{H} \} = \{ \partial_\theta, \vec{H}' \} \), we have\[ \hat{J}_\lambda(t) = \text{span}\{ e^{-t\vec{H}} \partial_\theta \}. \]
Then we define the generator of the Jacobi curve\[ V_t = e^{-t\vec{H}} \partial_\theta, \quad \dot{V}_t = e^{-t\vec{H}} [\vec{H}, \partial_\theta] = -e^{-t\vec{H}} \vec{H}' \]
Notice that\[ \sigma(V_t, \dot{V}_t) = -1, \quad \text{for every } t \geq 0. \]  
(16.26)\[ \text{Indeed it is true for } t = 0 \text{ and the equality is valid for all } t \text{ since the transformation } e^{t\vec{H}} \text{ is symplectic.} \]
To compute the curvature of the Jacobi curve let us write\[ V_t = \alpha(t)V_0 - \beta(t)\dot{V}_0 \]  
(16.27)We claim that the matrix \( S(t) \) representing the 1-dimensional Jacobi curve (that actually is a scalar), is given in these coordinates by\[ S(t) = \frac{\beta(t)}{\alpha(t)} = \frac{\sigma(V_0, V_t)}{\sigma(\dot{V}_0, V_t)}. \]
Indeed the identity\[ V_t = \alpha(t)V_0 - \beta(t)\dot{V}_0 = \alpha(t) \left( V_0 - \frac{\beta(t)}{\alpha(t)} \dot{V}_0 \right), \]  
(16.28)
tells us that the matrix representing the vector space spanned by $V_t$ is the graph of the linear map $V_0 \mapsto -\frac{\beta(t)}{\alpha(t)} \tilde{V}_0$. Moreover, using that $V_0$ and $\tilde{V}_0$ is a Darboux basis, it is easy to compute

$$
\sigma(V_0, V_t) = \alpha(t) \sigma(V_0, V_0) - \beta(t) \sigma(V_0, \tilde{V}_0) = \beta(t),
$$

$$
\sigma(\tilde{V}_0, V_t) = \alpha(t) \sigma(\tilde{V}_0, V_0) - \beta(t) \sigma(\tilde{V}_0, \tilde{V}_0) = \alpha(t).
$$

Differentiating the identity (16.26) with respect to $t$ one gets the relations

$$
\sigma(V_t, \ddot{V}_t) = 0, \quad \sigma(V_t, V_t^{(3)}) = -\sigma(\dot{V}_t, \dot{V}_t)
$$

Notice that these quantities are constant with respect to $t$. Collecting the above results one can compute the asymptotic expansion of $S(t)$ with respect to $t$

$$
S(t) = -t + \frac{t^3}{6} \sigma(V_0, \tilde{V}_0) + O(t^5)
$$

$$
= \left( -t + \frac{t^3}{6} \sigma(V_0, \tilde{V}_0) + O(t^5) \right) \left( 1 - \frac{t^2}{2} \sigma(\tilde{V}_0, V_0) + O(t^4) \right)
$$

and one gets for the derivative of $S(t)$ at $t = 0$

$$
\dot{S}(0) = -1, \quad \ddot{S}(0) = 0, \quad \overline{S}(0) = 2\sigma(\tilde{V}_0, \tilde{V}_0).
$$

The formula for the curvature $\mathcal{R}$ is finally computed in terms of $S(t)$ as follows:

$$
\mathcal{R} = -\frac{1}{2} \bar{S}(0) = \sigma(\tilde{V}_0, \tilde{V}_0)
$$

Using that $V_t = e^{-t\dot{H}} \partial_{\theta}$ we can expand $V_t$ as follows

$$
V_t = \partial_{\theta} + t[\dot{H}, \partial_{\theta}] + \frac{t^2}{2} [\dot{H}, [\dot{H}, \partial_{\theta}]] + O(t^3)
$$

hence (16.33) is rewritten as

$$
\mathcal{R} = \sigma([\dot{H}, [\dot{H}, \partial_{\theta}]], [\dot{H}, \partial_{\theta}])
$$

$$
= \sigma([\dot{H}, \dot{H}'], \dot{H}').
$$

To end this section, we compute the curvature $\mathcal{R}$ with respect to the orthonormal frame $f_1, f_2$.

Denote the Hamiltonians

$$
h_i(p, x) = \langle p, f_i(x) \rangle, \quad i = 1, 2.
$$

The PMP reads

$$
\begin{align*}
\dot{x} &= h_1 f_1(x) + h_2 f_2(x) \\
\dot{h}_1 &= \{H, h_1\} = \{h_2, h_1\} h_2 \\
\dot{h}_2 &= \{H, h_2\} = -\{h_2, h_1\} h_1
\end{align*}
$$

325
Moreover \( \{h_2, h_1\}(p, x) = \langle p, [f_2, f_1](x) \rangle \). Assume that
\[
[f_1, f_2] = a_1 f_1 + a_2 f_2, \quad a_i \in C^\infty(M).
\]
Then
\[
\{h_2, h_1\} = -a_1 h_1 - a_2 h_2.
\]
If we restrict to \( h_1 = \cos \theta \) and \( h_2 = \sin \theta \) equations (16.36) become
\[
\begin{aligned}
\dot{x} &= \cos \theta f_1 + \sin \theta f_2 \\
\dot{\theta} &= a_1 \cos \theta + a_2 \sin \theta
\end{aligned}
\]
and it is easy to compute the following expression for \( \vec{H} \) and commutators
\[
\vec{H} = h_1 f_1 + h_2 f_2 + (a_1 h_1 + a_2 h_2) \partial \theta,
\]
\[
\vec{H}' = -h_2 f_1 + h_1 f_2 + (-a_1 h_2 + a_2 h_1) \partial \theta,
\]
\[
[\vec{H}, \vec{H}'] = (f_1 a_2 - f_2 a_1 - a_1^2 - a_2^2) \partial \theta.
\]
Recall that
\[
\kappa = f_1 a_2 - f_2 a_1 - a_1^2 - a_2^2,
\]
is the Gaussian curvature of the surface \( M \) (see also Chapter 4). Since \( \sigma(\partial \theta, \vec{H}') = 1 \) one gets
\[
\mathcal{R} = \sigma([\vec{H}, \vec{H}'], \vec{H}') = \sigma(\kappa \partial \theta, \vec{H}') = \kappa.
\]

**Exercise 16.41.** In this exercise we recover the previous computations introducing dual coordinates to our frame. Let \( \nu_1, \nu_2 \) be the dual basis to \( f_1, f_2 \) and set
\[
f_\theta := h_1 f_1 + h_2 f_2, \quad \nu_\theta := h_1 \nu_1 + h_2 \nu_2.
\]
Define the smooth function \( b := a_1 h_1 + a_2 h_2 \) on \( T^*M \). In these notation
\[
\vec{H} = f_\theta + b \partial \theta, \quad \vec{H}' = f_\theta' + b' \partial \theta,
\]
where \( ' \) denotes the derivative with respect to \( \theta \). Then, using that in these coordinates the tautological form is \( s = \nu_\theta \), show that the symplectic form is written as
\[
\sigma = ds = d\theta \wedge \nu_\theta - b \nu_1 \wedge \nu_2,
\]
and compute the following expressions
\[
i_{\vec{H}} \sigma = (b' - b) \nu_\theta - d\theta,
\]
\[
[\vec{H}, \vec{H}'] = (f_\theta b' - f_\theta' b - b^2 - b^2) \partial \theta,
\]
showing that this gives an alternative proof of the above computation of the curvature.

\[\text{here we still use the notation } h_1, h_2 \text{ as functions of } \theta \text{ satisfying } \partial_\theta h_1 = -h_2, \partial_\theta h_2 = h_1\]
Chapter 17

Curvature in 3D contact sub-Riemannian geometry

The main goal of this chapter is to compute the curvature of the three dimensional contact sub-Riemannian case. Then we will discuss some general features of the curvature in sub-Riemannian geometry.

17.1 3D contact sub-Riemannian manifolds

In this section we consider a sub-Riemannian manifold $M$ of dimension 3 whose distribution is defined as the kernel of a contact 1-form $\omega \in \Lambda^1(M)$, i.e. $D_q = \ker \omega_q$ for all $q \in M$. Let us also fix a local orthonormal frame $f_1, f_2$ such that

$$D_q = \ker \omega_q = \text{span}\{f_1(q), f_2(q)\}$$

Recall that the 1-form $\omega \in \Lambda^1(M)$ defines a contact distribution if and only if $\omega \wedge d\omega \neq 0$ is never vanishing.

**Exercise 17.1.** Let $M$ be a 3D manifold, $\omega \in \Lambda^1 M$ and $D = \ker \omega$. The following are equivalent:

(i) $\omega$ is a contact 1-form,

(ii) $d\omega|_D \neq 0$,

(iii) $\forall f_1, f_2 \in \overline{D}$ linearly independent, then $[f_1, f_2] \notin \overline{D}$.

**Remark 17.2.** The contact form $\omega$ is defined up to a smooth function, i.e. if $\omega$ is a contact form, $a\omega$ is a contact form for every $a \in C^\infty(M)$. This let us to normalize the contact form by requiring that

$$d\omega|_D = \nu_1 \wedge \nu_2, \quad \text{(i.e. } d\omega(f_1, f_2) = 1.)$$

where $\nu_1, \nu_2$ is the dual basis to $f_1, f_2$. This is equivalent to say that $d\omega$ is equal to the area form induced on the distribution by the sub-Riemannian scalar product.

**Definition 17.3.** The *Reeb vector field* of the contact structure is the unique vector field $f_0 \in \text{Vec}(M)$ that satisfies

$$d\omega(f_0, \cdot) = 0, \quad \omega(f_0) = 1$$
In particular $f_0$ is transversal to the distribution and the triple $\{f_0, f_1, f_2\}$ defines a basis of $T_qM$ at every point $q \in M$. Notice that $\omega, \nu_1, \nu_2$ is the dual basis to this frame.

**Remark 17.4.** The flow generated by the Reeb vector field $e^{t f_0} : M \to M$ is a group of diffeomorphisms that satisfy $(e^{t f_0})^* \omega = \omega$. Indeed

$$\mathcal{L}_{f_0} \omega = d(i_{f_0} \omega) + i_{f_0} d\omega = 0$$

since $i_{f_0} \omega = \omega(f_0) = 1$ is constant and $i_{f_0} d\omega = d\omega(f_0, \cdot) = 0$.

In what follows, to simplify the notation, we will replace the contact form $\omega$ by $\nu_0$, as the dual element to the vector field $f_0$. We can write the structure equations of this basis of 1-forms

$$\begin{cases}
d\nu_0 = \nu_1 \wedge \nu_2 \\
d\nu_1 = c_{01}^1 \nu_0 \wedge \nu_1 + c_{02}^1 \nu_0 \wedge \nu_2 + c_{12}^1 \nu_1 \wedge \nu_2 \\
d\nu_2 = c_{01}^2 \nu_0 \wedge \nu_1 + c_{02}^2 \nu_0 \wedge \nu_2 + c_{12}^2 \nu_1 \wedge \nu_2
\end{cases} \quad (17.1)$$

The structure constants $c_{ij}^k$ are smooth functions on the manifold. Recall that the equation

$$d\nu_k = \sum_{i,j=0}^2 c_{ij}^k \nu_i \wedge \nu_j \quad \text{if and only if} \quad [f_j, f_i] = \sum_{k=0}^2 c_{ij}^k f_k.$$

Introduce the coordinates $(h_0, h_1, h_2)$ in each fiber of $T^*M$ induced by the dual frame

$$\lambda = h_0 \nu_0 + h_1 \nu_1 + h_2 \nu_2$$

where $h_i(\lambda) = \langle \lambda, f_i(q) \rangle$ are the Hamiltonians linear on fibers associated to $f_i$, for $i = 0, 1, 2$. The sub-Riemannian Hamiltonian is written as follows

$$H = \frac{1}{2}(h_1^2 + h_2^2).$$

We now compute the Poisson bracket $\{H, h_0\}$, denoting with $\{H, h_0\}_q$ its restriction to the fiber $T^*_qM$.

**Proposition 17.5.** The Poisson bracket $\{H, h_0\}_q$ is a quadratic form. Moreover we have

$$\{H, h_0\} = c_{01}^1 h_1^2 + (c_{01}^2 + c_{02}^1) h_1 h_2 + c_{02}^2 h_2^2, \quad (17.2)$$

$$c_{01}^1 + c_{02}^2 = 0. \quad (17.3)$$

Notice that $\Delta^1_q \subset \ker \{H, h_0\}_q$ and $\{H, h_0\}_q$ can be treated as a quadratic form on $T^*_qM/\Delta^1_q = \Delta^2_q$.

**Proof.** Using the equality $\{h_i, h_j\}(\lambda) = \langle \lambda, [f_i, f_j](q) \rangle$ we get

$$\{H, h_0\} = \frac{1}{2}\{h_1^2 + h_2^2, h_0\} = h_1 \{h_1, h_0\} + h_2 \{h_2, h_0\}$$

$$= h_1(c_{01}^1 h_1 + c_{02}^1 h_2) + h_2(c_{01}^2 h_1 + c_{02}^2 h_2)$$

$$= c_{01}^1 h_1^2 + (c_{01}^2 + c_{02}^1) h_1 h_2 + c_{02}^2 h_2^2.$$
Differentiating the first equation in (17.1) one gets:

\[ 0 = d^2
\nu_0 = d\nu_1 \wedge \nu_2 - \nu_1 \wedge \nu_2 \\
= (c^2_0 \nu_0 \wedge \nu_1) \wedge \nu_2 - \nu_1 \wedge (c^2_0 \nu_0 \wedge \nu_2) \\
= (c^1_0 + c^2_0)\nu_0 \wedge \nu_1 \wedge \nu_2 \\
\]

which proves (17.3).

**Remark 17.6.** Being \(\{H, h_0\}\) a quadratic form on the Euclidean plane \(\mathcal{D}_q\) (using the canonical identification of the vector space \(\mathcal{D}_q\) with its dual \(\mathcal{D}^*_q\) given by the scalar product), it can be interpreted as a symmetric operator on the plane itself. In particular its determinant and its trace are well defined. From (17.3) we get

\[ \text{trace } \{H, h_0\}_q = c^1_0 + c^2_0 = 0. \]

This identity is a consequence of the fact that the flow defined by the normalized Reeb \(f_0\) preserves not only the distribution but also the area form on it.

It is natural then to define our *first invariant* as the positive eigenvalue of this operator, namely:

\[ \chi(q) = \sqrt{-\det \{H, h_0\}_q}. \] (17.4)

Notice that the function \(\chi\) measures an intrinsic quantity since both \(H\) and \(h_0\) are defined only by the sub-Riemannian structure and are independent by the choice of the orthonormal frame. Indeed the quantity \(\{H, h_0\}\) compute the derivative of \(H\) along the flow of \(h_0\), i.e. the obstruction to the fact that the flow of the Reeb field \(f_0\) (which preserves the distribution and the volume form on it) to preserve the metric. Notice that, by definition \(\chi \geq 0\).

**Corollary 17.7.** Assume that the vector field \(f_0\) is complete. Then \(\{e^{tf_0}\}_{t \in \mathbb{R}}\) is a group of sub-Riemannian isometries if and only if \(\chi \equiv 0\).

In the case when \(\chi \equiv 0\) one can consider (locally) the quotient of \(M\) with respect to the action of this group, i.e. the space of trajectories described by \(f_0\). The two dimensional surface defined by the quotient structure is endowed with a well defined Riemannian metric.

The sub-Riemannian structure on \(M\) coincide with the isoperimetric Dido problem constructed on this surface. The Heisenberg case corresponds with the case when the surface has zero Gaussian curvature.

### 17.1.1 Curvature of a 3D contact structure

In this section we compute the sub-Riemannian curvature of a 3D contact structure with a technique similar to that used in Section 16.5 for the 2D Riemannian case. Let us consider the level set \(\{H = 1/2\}\) and define the coordinate \(\theta\) in such a way that

\[ h_1 = \cos \theta, \quad h_2 = \sin \theta. \]

On the bundle \(T^*M \cap H^{-1}(1/2)\) we introduce coordinates \((x, \theta, h_0)\). Notice that each fiber is topologically a cylinder \(S^1 \times \mathbb{R}\).
The sub-Riemannian Hamiltonian equation written in these coordinates are

$$\begin{aligned}
\dot{x} &= h_1 f_1(x) + h_2 f_2(x) \\
\dot{h}_1 &= \{H, h_1\} = \{h_2, h_1\} h_2 \\
\dot{h}_2 &= \{H, h_2\} = -\{h_2, h_1\} h_1 \\
\dot{h}_0 &= \{H, h_0\}
\end{aligned} \tag{17.5}$$

Computing the Poisson bracket $\{h_2, h_1\} = h_0 + c^1_{12} h_1 + c^2_{12} h_2$ and introducing the two functions $a, b : T^*M \to \mathbb{R}$ given by

$$a = \{H, h_0\} = \sum_{i,j=1}^2 c^i_{0j} k_i h_j, \quad b := c^1_{12} h_1 + c^2_{12} h_2.$$ we can rewrite the system, when restricted to $H^{-1}(1/2)$, as follows

$$\begin{aligned}
\dot{x} &= \cos \theta f_1 + \sin \theta f_2 \\
\dot{\theta} &= -h_0 - b \\
\dot{h}_0 &= a
\end{aligned} \tag{17.6}$$

Notice that, while $a$ is intrinsic, the function $b$ depends on the choice of the orthonormal frame.

In particular we have for the Hamiltonian vector field in the coordinates $(q, \theta, h_0)$ (where we use $h_1, h_2$ as a shorthand for $\cos \theta$ and $\sin \theta$):

$$\tilde{H} = h_1 f_1 + h_2 f_2 - (h_0 + b) \partial_\theta + a \partial_{h_0} \tag{17.7}$$

$$[\partial_\theta, \tilde{H}] = \tilde{H}' = -h_2 f_1 + h_1 f_2 + a' \partial_{h_0} - b' \partial_\theta \tag{17.8}$$

where we denoted by $'$ the derivative with respect to $\theta$, e.g. $h_1' = -h_2$ and $h_2' = h_1$.

Now consider the symplectic vector space $\Sigma_\lambda = T_\lambda(T^*M)$. The vertical subspace $\mathcal{V}_\lambda$ is generated by the vectors $\partial_\theta, \partial_{h_0}, \tilde{E}$. Hence the Jacobi curve is

$$J_\lambda(t) = \text{span}\{e^{-t\tilde{H}} \partial_\theta, e^{-t\tilde{H}} \partial_{h_0}, e^{-t\tilde{H}} \tilde{E}\}$$

The first reduction, by homogeneity, let us to split the space $\Sigma_\lambda = \text{span}\{\tilde{E}, \tilde{H}\} \oplus \text{span}\{\tilde{E}, \tilde{H}\}^\perp$ and consider the reduced Jacobi curve $\Lambda(t) := \tilde{J}_\lambda(t)$ in the 4-dimensional symplectic space

$$\Lambda(t) := e^{-t\tilde{H}} \mathcal{V}_\lambda / \mathbb{R}\tilde{H} = \text{span}\{e^{-t\tilde{H}} \partial_\theta, e^{-t\tilde{H}} \partial_{h_0}\} / \mathbb{R}\tilde{H}$$

Next we describe the second reduction of the Jacobi curve, the one related with the fact that the curve is non-regular. Indeed notice that the rank of $\tilde{J}_\lambda(t)$ is 1. To find the new reduced curve, we need to compute the kernel of the derivative of the curve at $t = 0$

$$\Gamma := \text{Ker} \dot{\Lambda}(0)$$

From the definition of $\dot{\Lambda} := \dot{\Lambda}(0)$ it follows that

$$\dot{\Lambda}(\partial_\theta) = \pi_* [\tilde{H}, \partial_\theta] = h_2 f_1 - h_1 f_2$$

$$\dot{\Lambda}(\partial_{h_0}) = \pi_* [\tilde{H}, \partial_{h_0}] = \pi_* (\partial_\theta) = 0$$

Hence $\Gamma = \mathbb{R}\partial_{h_0}$ and $\Gamma^\perp$ is 3-dimensional in $\mathcal{V}_\lambda / \mathbb{R}\tilde{H}$. 330
Proposition 17.8. We have the following characterizations:

(i) \( \Gamma^\prec = \text{span}\{\partial_{h_0}, \partial_{\theta}, \vec{H}'\} \) in \( \tilde{V}_\lambda/\mathbb{R}\vec{H} \),

(ii) \( \{\partial_{\theta}, \vec{H}'\} \) is a Darboux basis for \( \Gamma^\prec/\Gamma \).

Proof. Since \( \partial_{h_0} \) and \( \partial_{\theta} \) are vertical to prove (i) it is enough to show that \( \vec{H}' \) is skew-orthogonal to \( \partial_{h_0} \). It is easy to compute, by Cartan formula

\[
\sigma(\partial_{h_0}, \vec{H}') = \partial_{h_0} \langle s, \vec{H}' \rangle - \vec{H}' \langle s, \partial_{h_0} \rangle - \langle s, [\partial_{h_0}, \vec{H}'] \rangle = 0,
\]

since all the three terms vanish. Indeed \( \langle s, \vec{H}' \rangle = \sigma(\vec{E}, \vec{H}') = 0 \) and \( \langle s, \partial_{h_0} \rangle = \langle s, [\partial_{h_0}, \vec{H}'] \rangle = 0 \) since \( \partial_{h_0} \) and \([\partial_{h_0}, \vec{H}']\) are both vertical, as can be computed from (17.8).

To complete the proof of (ii) it is enough to show, using \([\partial_{\theta}, \vec{H}'] = -\vec{H} \), that

\[
\sigma(\partial_{\theta}, \vec{H}') = \partial_{\theta} \langle s, \vec{H}' \rangle - \vec{H}' \langle s, \partial_{\theta} \rangle - \langle s, [\partial_{\theta}, \vec{H}'] \rangle = \langle s, \vec{H} \rangle = 1.
\]

\[\square\]

Next we compute the curvature in terms of the Hamiltonian vector field and its commutators. For a vector field \( W \) we use the notations

\[ \dot{W} := [\vec{H}, W], \quad W' := [\partial_{\theta}, W]. \]

Let us consider the vector field \( V_t = e^{-t\vec{H}} \partial_{h_0} \). Notice that

\[ \dot{V}_0 = \partial_{\theta}, \quad \ddot{V}_0 = -\vec{H}'. \]

The fact that \( \partial_{\theta} \) and \( \partial_{h_0} \) are vertical implies that

\[ \sigma(V_t, \dot{V}_t) = 0, \quad \forall t \geq 0 \]

Differentiating the above identity at \( t = 0 \) we get (from now on, we omit \( t \) when we evaluate at \( t = 0 \))

\[ \sigma(\dot{V}, \dot{V}) + \sigma(V, \ddot{V}) = 0 \quad \implies \quad \sigma(V, \ddot{V}) = 0. \]

Differentiating once more the last identity and using \( \sigma(\dot{V}, \ddot{V}) = -\sigma(\partial_{\theta}, \vec{H}') = -1 \) one gets

\[ \sigma(\dot{V}, \ddot{V}) + \sigma(V, V^{(3)}) = 0 \quad \implies \quad \sigma(V, V^{(3)}) = 1. \]

With similar computations one can show that \( \sigma(\dot{V}, V^{(3)}) = \sigma(V, V^{(4)}) = 0 \). Evaluating all derivatives of order 4 one can see that

\[ r := \sigma(\dot{V}, V^{(3)}) = -\sigma(\dot{V}, V^{(4)}) = \sigma(V, V^{(5)}). \]

Proposition 17.9. The sub-Riemannian curvature is

\[ \mathcal{R} = \frac{1}{10} \sigma([\vec{H}, \vec{H}'], \vec{H}') = -\frac{r}{10}. \]
Proof. The second equality follows from the definition of $r$ and the fact that $\tilde{V} = -\tilde{H}'$ and $V^{(3)} = [\tilde{H}, \tilde{H}']$.

To prove the first identity we have to compute the Schwartzian derivative of the bi-reduced curve, in the symplectic basis $(\dot{V}, -\ddot{V})$ of the space $\Gamma^\perp / \Gamma$ (notice the minus sign).

Recall that $\Lambda(t) = \text{span}\{V_t, \dot{V}_t\}$. To compute the 1-dimensional reduced curve $\Lambda^\perp(t)$ in the symplectic space $\Gamma^\perp / \Gamma$ we need to compute the intersection of $\Lambda(t)$ with $\Gamma^\perp$ (for all $t$). In other words we look for $x(t)$ such that

$$\sigma(\dot{V}_t + x(t)V_t, V_0) = 0 \implies x(t) = -\frac{\sigma(\dot{V}_t, V_0)}{\sigma(V_t, V_0)}. \quad (17.9)$$

Then we write this vector as a linear combination of the Darboux basis (cf. (16.28) for the 2D Riemannian case)

$$\dot{V}_t + x(t)V_t = \alpha(t)\dot{V}_0 - \beta(t)\ddot{V}_0 + \xi(t)V_0 \quad (17.10)$$

To see it as a curve in the space $\Gamma / \Gamma^\perp$ we simply ignore the coefficient along $V_0$. In these coordinates the matrix $S(t)$, which is a scalar, representing the curve is

$$S(t) = \frac{\beta(t)}{\alpha(t)} \quad (17.11)$$

Notice that this is a one-dimensional non-degenerate curve. These coefficients are computed by the symplectic products

$$\alpha(t) = -\sigma(\dot{V}_t + x(t)V_t, \ddot{V}_0) \quad (17.12)$$
$$\beta(t) = -\sigma(\dot{V}_t + x(t)V_t, \dot{V}_0) \quad (17.13)$$

Combining (17.12), (17.13) with (17.11) and (17.9) one gets

$$S(t) = \frac{\sigma(V_t, V_0)\sigma(V_t, V_0) - \sigma(V_t, V_0)\sigma(\dot{V}_t, V_0)}{\sigma(V_t, V_0)\sigma(V_t, V_0) - \sigma(V_t, V_0)\sigma(\dot{V}_t, V_0)} \quad (17.14)$$

After some computations, by Taylor expansion one gets

$$S(t) = \frac{t}{4} - \frac{t^3}{120}r + O(t^4) \quad (17.15)$$

Since $\ddot{S}_0 = 0$ the curvature is computer by

$$\mathcal{R} = \frac{\ddot{S}_0}{2S_0} = -\frac{r}{10}$$

We end this section by computing the expression of the curvature in terms of the orthonormal frame for the distribution and the Reeb vector field. As usual we restrict to the level set $H^{-1}(1/2)$ where

$$h_1^2 + h_2^2 = 1, \quad h_1 = \cos \theta, \quad h_2 = \sin \theta.$$ 

In the following we use the notation

$$f_\theta = h_1 f_1 + h_2 f_2, \quad v_\theta = h_1 v_1 + h_2 v_2.$$
If \( h = (h_1, h_2) = (\cos \theta, \sin \theta) \) we denote by \( h' = (-h_2, h_1) = (-\sin \theta, \cos \theta) \) its derivative with respect to \( \theta \) and, more in general, we denote \( F' := \partial_\theta F \) for a smooth function \( F \) on \( T^*M \).

To express the quantity \( r = \sigma([\vec{H}, \vec{H}'], \vec{H}') \) we start by computing the commutator \( [\vec{H}, \vec{H}'] \).

From (17.7) and (17.8) one gets
\[
[\vec{H}, \vec{H}'] = -f_0 + h_0 f_\theta + (f_2 c_1^2 - f_1 c_2^2) - (h_0 + b) b - (b')^2 + a' \partial_\theta.
\]

Next we write, following this notation, the symplectic form \( \sigma = ds \). The Liouville form \( s \) is expressed, in the dual basis \( \nu_0, \nu_1, \nu_2 \) to the basis of vector fields \( f_1, f_2, f_0 \) as follows
\[
s = h_0 \nu_0 + \nu_\theta
\]
hence the symplectic form \( \sigma \) is written as follows
\[
\sigma = dh_0 \wedge \nu_0 + h_0 \nu_\theta \wedge \nu_\theta' + d\nu_\theta
\]
where we used that \( d\nu_0 = \nu_1 \wedge \nu_2 = \nu_\theta \wedge \nu_\theta' \). Computing the symplectic product then one finds the value of
\[
10R = h_0^2 + \frac{3}{2} a' + \kappa
\]
where
\[
\kappa = f_2 c_1^2 - f_1 c_2^2 - (c_1^2)^2 - (c_2^2)^2 + \frac{c_0^2 - c_{02}}{2}
\]
(17.16)

By homogeneity, the function \( R \) is defined on the whole \( T^*M \), and not only for \( \lambda \in H^{-1}(1/2) \).

For every \( \lambda = (h_0, h_1, h_2) \in T^*_xM \)
\[
10R = h_0^2 + \frac{3}{2} a' + \kappa(h_1^2 + h_2^2)
\]

Remark 17.10. The restriction of \( R \) to the 1-dimensional subspace \( \lambda \in D^\perp \) (that corresponds to \( \lambda = (h_0, 0, 0) \)), is a strictly positive quadratic form. Moreover it is equal to 1/10 when evaluated on the Reeb vector field. Hence the curvature \( R \) encodes both the contact form \( \omega \) and its normalization.

On the orthogonal complement (with respect to \( R \)) \( \{h_0 = 0\} \) we have that \( R \) is treated as a quadratic form
\[
R = \frac{3}{2} a' + \kappa(h_1^2 + h_2^2).
\]

Remark 17.11. (i). If \( a \neq 0 \) there always exists a frame such that
\[
a = 2\chi h_1 h_2
\]
and in this frame we can express \( R \) as a quadratic form on the whole \( T^*M \)
\[
R = h_0^2 + (\kappa + 3\chi) h_1^2 + (\kappa - 3\chi) h_2^2.
\]
It is easily seen from this formulas that we can recover the two invariants \( \chi, \kappa \) considering
\[
\text{trace}(10R|_{h_0=0}) = 2\kappa, \quad \text{discr}(10R|_{h_0=0}) = 36\chi.
\]
(ii). When \( a = 0 \) the eigenvalues of \( R \) coincide and \( \chi = 0 \). In this case \( \kappa \) represents the Riemannian curvature of the surface defined by the quotient of \( M \) with respect to the flow of the Reeb vector field.
Indeed the flow $e^{t f_0}_\ast$ preserves the metric and it is easy to see that the identities

$$e^{t f_0}_\ast f_i = f_i, \quad i = 1, 2.$$ 

implies $[f_0, f_1] = [f_0, f_2] = 0$. Hence $c^2_{01}, c^1_{02} = 0$ and the expression of $\kappa$ reduces to the Riemannian curvature of a surface whose orthonormal frame is $f_1, f_2$.

**Exercise 17.12.** Let $f_1, f_2$ be an orthonormal frame for $M$ and denote by $\hat{f}_1, \hat{f}_2$ the frame obtained rotating $f_1, f_2$ by an angle $\theta = \theta(q)$. Show that the structure constants $\hat{c}^k_{ij}$ of rotated frame satisfies

$$\hat{c}^1_{12} = \cos \theta(c^1_{12} - f_1(\theta)) - \sin \theta(c^2_{12} - f_2(\theta)),$$

$$\hat{c}^2_{12} = \sin \theta(c^1_{12} - f_1(\theta)) + \cos \theta(c^2_{12} - f_2(\theta)).$$

**Exercise 17.13.** Show that the expression (17.16) for $\kappa$ does not depend on the choice of an orthonormal frame $f_1, f_2$ for the sub-Riemannian structure.
Chapter 18

Asymptotic expansion of the 3D contact exponential map

In this chapter we study the small time asymptotics of the exponential map in the three-dimensional contact case and see how the structure of the cut and the conjugate locus is encoded in the curvature.

Let us consider the sub-Riemannian Hamiltonian of a 3D contact structure (cf. Section 17.1.1)

$$\vec{H} = h_1 f_1 + h_2 f_2 - (h_0 + b)\partial_\theta + a \partial_{h_0}$$

written in the dual coordinates \((h_0, h_1, h_2)\) of a local frame \(f_0, f_1, f_2\), where \(\nu_0\) is the normalized contact form, \(f_0\) is the Reeb vector field and \(f_1, f_2\) is a local orthonormal frame for the sub-Riemannian structure. As usual the coordinate \(\theta\) on the level set \(H^{-1}(1/2)\) is defined such a way that \(h_1 = \cos \theta\) and \(h_2 = \sin \theta\).

In this chapter it will be convenient to introduce the notation \(\rho := -h_0\) for the function linear on fibers of \(T^*M\) associated with the opposite of the Reeb vector field. The Hamiltonian system (18.1) on the level set \(H^{-1}(1/2)\) is rewritten in the following form:

$$\begin{cases}
\dot{q} = \cos \theta f_1 + \sin \theta f_2 \\
\dot{\theta} = \rho - b \\
\dot{\rho} = -a
\end{cases}$$

(18.2)

The exponential map starting from the initial point \(q_0 \in M\) is the map that to each time \(t > 0\) and every initial covector \((\theta_0, \rho_0) \in T^*_0 M \cap H^{-1}(1/2)\) assigns the first component of the solution at time \(t\) of the system (18.2), denoted by \(E_{q_0}(t, \theta_0, \rho_0)\), or simply \(E(t, \theta_0, \rho_0)\).

Conjugate points are points where the differential of the exponential map is not surjective, i.e. solutions to the equation

$$\frac{\partial E}{\partial \theta_0} \wedge \frac{\partial E}{\partial \rho_0} \wedge \frac{\partial E}{\partial t} = 0.$$  

(18.3)

The variation of the exponential map along time is always nonzero and independent with respect to variations of the covectors in the set \(H^{-1}(1/2)\) (see also Section 8.9 and Proposition 8.37). This implies that (18.3) is equivalent to

$$\frac{\partial E}{\partial \theta_0} \wedge \frac{\partial E}{\partial \rho_0} = 0.$$  

(18.4)
18.1 Nilpotent case

The nilpotent case, i.e. the Heisenberg group, corresponds to the case when the functions $a$ and $b$ vanish identically, i.e. the system

\[
\begin{aligned}
\dot{q} &= \cos \theta f_1 + \sin \theta f_2 \\
\dot{\theta} &= \rho \\
\dot{\rho} &= 0
\end{aligned}
\]  

(18.5)

Let us first recover, in this notation, the conjugate locus in the case of the Heisenberg group. Let us denote coordinates on the manifold $\mathbb{R}^3$ as follows

\[ q = (x, y), \quad x = (x_1, x_2) \in \mathbb{R}^2, y \in \mathbb{R}. \]  

(18.6)

Notice moreover that in this case the Reeb vector field is proportional to $\partial_y$ and its dual coordinate $\rho$ is constant along trajectories. There are two possible cases:

(i) $\rho = 0$. Then the solution is a straight line contained in the plane $y = 0$ and is optimal for all time.

(ii) $\rho \neq 0$. In this case we claim that the equation (18.4) is equivalent to the following

\[ \frac{\partial x}{\partial \theta_0} \wedge \frac{\partial x}{\partial \rho_0} = 0. \]  

(18.7)

By the Gauss’ Lemma (Proposition 8.37), the covector $p = (p_x, \rho)$ at the final point annihilates the differential of the exponential map restricted to the level set, i.e.

\[ \left\langle p, \frac{\partial \mathcal{E}}{\partial \theta_0} \right\rangle = \left\langle p_x, \frac{\partial x}{\partial \theta_0} \right\rangle + \rho \frac{\partial y}{\partial \theta_0} = 0 \]  

(18.8)

\[ \left\langle p, \frac{\partial \mathcal{E}}{\partial \rho_0} \right\rangle = \left\langle p_x, \frac{\partial x}{\partial \rho_0} \right\rangle + \rho \frac{\partial y}{\partial \rho_0} = 0 \]  

(18.9)

and since $\rho \neq 0$ it follows that among the three vectors

\[
\begin{pmatrix}
\frac{\partial x_1}{\partial \theta_0} & \frac{\partial x_1}{\partial \rho_0} \\
\frac{\partial x_2}{\partial \theta_0} & \frac{\partial x_2}{\partial \rho_0} \\
\frac{\partial y}{\partial \theta_0} & \frac{\partial y}{\partial \rho_0}
\end{pmatrix}
\]  

(18.10)

the third one is always a linear combination of the first two.

**Proposition 18.1.** The first conjugate time is $t_c(\theta_0, \rho_0) = 2\pi/|\rho_0|$.  

**Proof.** In the standard coordinates $(x_1, x_2, y)$ the two vector fields $f_1$ and $f_2$ defining the orthonormal frame are

\[ f_1 = \partial_{x_1} - \frac{x_2}{2} \partial_y, \quad f_2 = \partial_{x_2} + \frac{x_1}{2} \partial_y \]

Thus, the first two coordinates of the horizontal part of the Hamiltonian system satisfy

\[
\begin{aligned}
\dot{x}_1 &= \cos \theta \\
\dot{x}_2 &= \sin \theta
\end{aligned}
\]  

(18.11)
It is then easy to integrate the $x$-part of the exponential map being $\theta(t) = \theta_0 + \rho t$ (recall that $\rho \equiv \rho_0$ and, without loss of generality we can assume $\rho > 0$)

$$x(t; \theta_0, \rho_0) = \int_0^t \begin{pmatrix} \cos(\theta_0 + \rho s) \\ \sin(\theta_0 + \rho s) \end{pmatrix} ds = \int_{\theta_0}^{\theta_0 + \rho t} \begin{pmatrix} \cos \rho s \\ \sin \rho s \end{pmatrix} ds \quad (18.12)$$

Due to the symmetry of the Heisenberg group, the determinant of the Jacobian map will not depend on $\theta_0$. Hence to compute the determinant of the Jacobian it is enough to compute partial derivatives at $\theta_0 = 0$

\[
    \frac{\partial x}{\partial \theta_0} = \begin{pmatrix} \cos \rho t - 1 \\ \sin \rho t \end{pmatrix},
    \frac{\partial x}{\partial \rho_0} = -\frac{1}{\rho^2} \begin{pmatrix} \sin \rho t \\ 1 - \cos \rho t \end{pmatrix} + \frac{t}{\rho} \begin{pmatrix} \cos \rho t \\ \sin \rho t \end{pmatrix}
\]

and denoting by $\tau := \rho t$ one can compute

\[
    \frac{\partial x}{\partial \theta_0} \wedge \frac{\partial x}{\partial \rho_0} = \frac{1}{\rho^2} \det \begin{pmatrix} \cos \tau - 1 & \tau \cos - \sin \tau \\ \sin \tau & 1 + \tau \sin \tau \cos \tau \end{pmatrix},
    \frac{1}{\rho^2} (\tau \sin \tau + 2 \cos \tau - 2).
\]

The fact that $t_c = 2\pi/|\rho|$ follows from Exercise 18.2.

**Exercise 18.2.** Prove that $\tau_c = 2\pi$ is the first positive root of the equation $\tau \sin \tau + 2 \cos \tau - 2 = 0$. Moreover show that $\tau_c$ is a simple root.

### 18.2 General case: second order asymptotic expansion

Let us consider the Hamiltonian system for the general 3D contact case

$$\begin{cases}
    \dot{q} = f_\theta := \cos \theta f_1 + \sin \theta f_2 \\
    \dot{\theta} = \rho - b \\
    \dot{\rho} = -a
\end{cases} \quad (18.13)$$

We are going to study the asymptotic expansion for our system for the initial parameter $\rho_0 \to \pm \infty$. To this aim, it is convenient to introduce the change of variables $r := 1/\rho$ and denote by $\nu := r(0) = 1/\rho_0$ its initial value. Notice that $\rho$ is no more constant in the general case and $\rho_0 \to \infty$ implies $\nu \to 0$.

The main result of this section says that the conjugate time for the perturbed system is a perturbation of the conjugate time of the nilpotent case, where the perturbation has no term of order 2.

**Proposition 18.3.** The conjugate time $t_c(\theta_0, \nu)$ is a smooth function of the parameter $\nu$ for $\nu > 0$. Moreover for $\nu \to 0$

$$t_c(\theta_0, \nu) = 2\pi|\nu| + O(|\nu|^3).$$

337
Proof. Let us introduce a new time variable $\tau$ such that $\frac{d\tau}{d\tau} = r$. If we now denote by $\hat{F}$ the derivative of a function $F$ with respect to the new time $\tau$, the system (18.13) is rewritten in the new coordinate system $(q, \theta, r)$ (where we recall $r = 1/\rho$), as follows

$$\begin{align*}
\dot{q} &= r f_\theta \\
\dot{\theta} &= 1 - rb \\
\dot{r} &= r^3 a \\
\dot{t} &= r
\end{align*} \quad (18.14)
$$

To compute the asymptotics of the conjugate time, it is also convenient to consider a system of coordinates, depending on a parameter $\varepsilon$, corresponding to the quasi-homogeneous blow up of the sub-Riemannian structure at $q_0$ and converging to the nilpotent approximation. In other words we consider the change of coordinates $\Phi_\varepsilon$ such that $f_\theta \mapsto f_\theta^{\varepsilon} + \varepsilon f^{(0)} + \varepsilon^2 f^{(1)} + \ldots$

Accordingly to this change of coordinates we have the equalities

$$f_i = \frac{1}{\varepsilon} f_i^{\varepsilon}, \quad f_0 = \frac{1}{\varepsilon} f_0^{\varepsilon}, \quad b = \frac{1}{\varepsilon} b^{\varepsilon}, \quad a = \frac{1}{\varepsilon^2} a^{\varepsilon}$$

where $f_0^{\varepsilon}$ is the Reeb vector field defined by the orthonormal frame $f_1^{\varepsilon}, f_2^{\varepsilon}$ (and analogously for $a^{\varepsilon}, b^{\varepsilon}$).

Let us now define, for fixed $\varepsilon$, the variable $w$ such that $r = \varepsilon w$. The system (18.14) is finally rewritten in the following form

$$\begin{align*}
\dot{q} &= w f_\theta^{\varepsilon} \\
\dot{\theta} &= 1 - wb^{\varepsilon} \\
\dot{w} &= \varepsilon w^3 a^{\varepsilon} \\
\dot{t} &= \varepsilon w
\end{align*} \quad (18.15)$$

Notice that the dynamical system is written in a coordinate system that depends on $\varepsilon$. Moreover the initial asymptotic for $\rho_0 \to \infty$, corresponding to $r \to 0$, is now reduced to fix an initial value $w(0) = 1$ and send $\varepsilon \to 0$.

Consider some linearly adapted coordinates $(x, y)$, with $x \in \mathbb{R}^2$ and $y \in \mathbb{R}$ (cf. Definition 10.24). If we denote by $q^{\varepsilon} = (x^{\varepsilon}, y^{\varepsilon})$ the solution of the horizontal part of the $\varepsilon$-system (18.15), conjugate points are solutions of the equation

$$\left. \frac{\partial q^{\varepsilon}}{\partial \theta} \wedge \frac{\partial q^{\varepsilon}}{\partial w} \right|_{w_0=1} = 0.$$

As in Section 18.1 one can check that this condition is equivalent to

$$\left. \frac{\partial x^{\varepsilon}}{\partial \theta} \wedge \frac{\partial x^{\varepsilon}}{\partial w} \right|_{w_0=1} = 0.$$

Notice that the original parameters $(t, \theta_0, \rho_0)$ parametrizing the trajectories in the exponential map correspond to a conjugate point if the corresponding parameters $(\tau, \theta_0, \varepsilon)$ satisfy

$$\varphi(\tau, \varepsilon, \theta_0) := \left. \frac{\partial x^{\varepsilon}}{\partial \theta_0} \wedge \frac{\partial x^{\varepsilon}}{\partial w} \right|_{w_0=1} = 0 \quad (18.16)$$
For $\epsilon = 0$, i.e. the nilpotent approximation, the first conjugate time is $\tau_c = 2\pi$, and moreover it is a simple root. Thus one gets
\[
\varphi(2\pi, 0, \theta_0) = 0, \quad \frac{\partial \varphi}{\partial \tau}(2\pi, 0, \theta_0) \neq 0. \tag{18.17}
\]
Hence the implicit function theorem guarantees that there exists a smooth function $\tau_c(\epsilon, \theta_0)$ such that $\tau_c(0, \theta_0) = 2\pi$ and
\[
\varphi(\tau_c(\epsilon, \theta_0), \epsilon, \theta_0) = 0. \tag{18.18}
\]
In other words $\tau_c(\epsilon, \theta_0)$ computes the conjugate time $\tau$ associated with parameters $\epsilon, \theta_0$. By smoothness of $\tau_c$ one immediately has the expansion for $\epsilon \to 0$
\[
\tau_c(\epsilon, \theta_0) = 2\pi + O(\epsilon). \tag{18.19}
\]
Now the statement of the proposition is rewritten in terms of the function $\tau_c$ as follows
\[
\tau_c(\epsilon, \theta_0) = 2\pi + O(\epsilon^2). \tag{18.19}
\]
Differentiating the identity (18.18) with respect to $\epsilon$ one has
\[
\frac{\partial \varphi}{\partial \tau} \frac{\partial \tau_c}{\partial \epsilon} + \frac{\partial \varphi}{\partial \epsilon} = 0,
\]
and, thanks to (18.17), the expansion (18.19) holds if and only if $\frac{\partial \varphi}{\partial \epsilon}(2\pi, 0, \theta_0) = 0$.

Moreover differentiating the expression (18.16) with respect to $\epsilon$ one has
\[
\frac{\partial \varphi}{\partial \epsilon}(2\pi, 0, \theta_0) = \frac{\partial^2 x^\epsilon}{\partial \epsilon \partial \theta_0} \wedge \frac{\partial^2 x^\epsilon}{\partial \epsilon \partial w_0} - \frac{\partial^2 x^\epsilon}{\partial \epsilon \partial \theta_0} \wedge \frac{\partial x^\epsilon}{\partial \theta_0} \bigg|_{w_0=1, \epsilon=0, \tau=2\pi}
\]
The second one vanishes since at $\epsilon = 0$ is the Heisenberg case, whose horizontal part at $\tau = 2\pi$ does not depend on $\theta_0$. Hence we are reduced to prove that
\[
\frac{\partial^2 x^\epsilon}{\partial \epsilon \partial \theta_0} \bigg|_{\epsilon=0, \tau=2\pi} = 0. \tag{18.20}
\]
which is a consequence of the following lemma.

**Lemma 18.4.** The quantity $\frac{\partial x^\epsilon}{\partial \epsilon} \bigg|_{\epsilon=0, \tau=2\pi}$ does not depend on $\theta_0$.

**Proof of Lemma.** To prove the lemma it will be enough to find the first order expansion in $\epsilon$ of the solution of the system (18.15).

Recall that when $\epsilon = 0$ the system corresponds to the Heisenberg case, i.e. we have $a^\epsilon|_{\epsilon=0} = 0, b^\epsilon|_{\epsilon=0} = 0$. This gives the expansion of $w$ (recall that $w(0) = w_0 = 1$)
\[
w(t) = w(0) + \int_0^t \varepsilon a^\epsilon(\tau) w^3(\tau) d\tau \quad \Rightarrow \quad w = 1 + O(\epsilon^2)
\]
Analogously we have $b^\epsilon = \varepsilon \langle \beta, u \rangle + O(\epsilon^2)$, where $\langle \beta, u \rangle = \beta_1 u_1 + \beta_2 u_2$ and $\beta$ denotes the (constant) coefficient of weight zero in the expansion of $b$ with respect to $\epsilon$. 339
Denoting \( u(\theta) = (\cos \theta, \sin \theta) \), the equation for \( \theta \) then is reduced to

\[
\dot{\theta} = 1 - \varepsilon \langle \beta, u(\theta) \rangle + O(\varepsilon^2), \quad \theta(0) = \theta_0.
\]

This equation can be integrated and one gets

\[
\frac{\partial \theta}{\partial \varepsilon} \bigg|_{\varepsilon=0} = -\int_0^t \langle \beta, u(\theta(\tau)) \rangle \, d\tau = \langle \beta, u'(\theta_0 + t) - u'(\theta_0) \rangle
\]

(18.21)

where \( u'(\theta) = (-\sin \theta, \cos \theta) \).

Next we are going to use (18.21) to compute the derivative of \( x^\varepsilon \) wrt \( \varepsilon \). The equation for the horizontal part of (18.15) can be expanded in \( \varepsilon \) as follows

\[
\dot{x}^\varepsilon = u(\theta) + \varepsilon f_{u(\theta)}(x) + O(\varepsilon^2)
\]

where the first term is Heisenberg, and \( f_{u(\theta)}(x) \) is the term of weight zero of \( f_u \), which is linear with respect to \( x_1 \) and \( x_2 \) because of the weight.\(^1\) To compute the derivative of the solution with respect to parameter we use the following general fact

**Lemma 18.5.** Let \( \phi(\varepsilon, t) \) denote the solution of the differential equation \( \dot{y} = F(\varepsilon, y) \) with fixed initial condition \( y(0) = y_0 \). Then the derivative \( \frac{\partial \phi}{\partial \varepsilon} \) satisfies the following linear ODE

\[
\frac{d}{dt} \frac{\partial \phi}{\partial \varepsilon}(\varepsilon, t) = \frac{\partial F}{\partial y}(\varepsilon, \phi(\varepsilon, t)) \frac{\partial \phi}{\partial \varepsilon}(\varepsilon, t) + \frac{\partial F}{\partial \varepsilon}(\varepsilon, \phi(\varepsilon, t))
\]

We apply the above lemma when \( y = (x, \theta) \) and \( F = (F^x, F^\theta) \) and we compute at \( \varepsilon = 0 \). In particular we need the solution of the original system at \( \varepsilon = 0 \)

\[
\phi(0, t) = (\bar{x}(t), \bar{\theta}(t)), \quad \bar{\theta}(t) = \theta_0 + t, \quad \bar{x}(t) = u'(\theta_0) - u'(\theta_0 + t).
\]

Then by Lemma [18.5] we have

\[
\frac{d}{dt} \frac{\partial x}{\partial \varepsilon} = \frac{\partial F^x}{\partial x} \frac{\partial x}{\partial \varepsilon} + \frac{\partial F^x}{\partial \theta} \frac{\partial \theta}{\partial \varepsilon} + \frac{\partial F^x}{\partial \varepsilon}
\]

Computing the derivatives at \( \varepsilon = 0 \) gives

\[
\frac{\partial F^x}{\partial x} \bigg|_{\varepsilon=0} = 0, \quad \frac{\partial F^x}{\partial \theta} \bigg|_{\varepsilon=0} = u'(\bar{\theta}(t)), \quad \frac{\partial F^x}{\partial \varepsilon} \bigg|_{\varepsilon=0} = f_{u(\bar{\theta}(t))}^{(0)}(\bar{x}(t))
\]

and we obtain the equation for \( \frac{\partial x}{\partial \varepsilon} \)

\[
\frac{d}{dt} \frac{\partial x}{\partial \varepsilon} \bigg|_{\varepsilon=0} = \frac{\partial \theta}{\partial \varepsilon} \bigg|_{\varepsilon=0} u'(\theta_0 + t) + f_{u(\theta_0 + t)}^{(0)}(u'(\theta_0) - u'(\theta_0 + t))
\]

\(^1\)Recall that this is the zero order part of the vector field \( f_u \) along \( \partial_x \), hence only \( x \) variables appear and have order 1.
If we set $s = \theta_0 + t$ we can rewrite this equation

$$
\frac{d}{ds} \frac{\partial x}{\partial \epsilon} \bigg|_{\epsilon=0} = \frac{\partial \theta}{\partial \epsilon} u'(s) + f^{(0)}_{u'(s)}(u' - u'(s))
$$

and integrating one has

$$
\frac{\partial x}{\partial \epsilon} \bigg|_{(2\pi,0)} = \int_{\theta_0}^{\theta_0+2\pi} \langle \beta, u'(s) - u'(\theta_0) \rangle u'(s) ds
+ \int_{\theta_0}^{\theta_0+2\pi} f^{(0)}_{u'(s)}(u' - u'(s)) ds
$$

In the last expression it is easy to see that all terms where $\theta_0$ appears are zero, while the others vanish since we compute integrals of periodic functions over a period (which does not dep end on $\theta_0$). This finishes the proof of Lemma 18.4, hence the proof of the Proposition 18.3.

18.3 General case: higher order asymptotic expansion

Next we continue our analysis about the structure of the conjugate locus for a 3D contact structure by studying the higher order asymptotic. In this section we determine the coefficient of order 3 in the asymptotic expansion of the conjugate locus. Namely we have the following result, whose proof is postponed to Section 18.3.1.

**Theorem 18.6.** In a system of local coordinates around $q_0 \in M$ one has the expansion

$$
\text{Con}_{q_0}(\theta_0, \nu) = q_0 \pm \pi f_0 |\nu|^2 \pm \pi (a' f_{\theta_0} - a f_{\theta_0}') |\nu|^3 + O(|\nu|^4), \quad \nu \to 0^\pm. \quad (18.22)
$$

If we choose coordinates such that $a = 2\chi h_1 h_2$ one gets

$$
\text{Con}_{q_0}(\theta_0, \nu) = q_0 \pm \pi f_0 |\nu|^2 \pm 2\pi \chi(q_0)(\cos^3 \theta f_2 - \sin^3 \theta f_1) |\nu|^3 + O(|\nu|^4), \quad \nu \to 0^\pm. \quad (18.23)
$$

Moreover for the conjugate length we have the expansion

$$
\ell_c(\theta_0, \nu) = 2\pi |\nu| - \pi \kappa |\nu|^3 + O(|\nu|^4), \quad \nu \to 0^\pm. \quad (18.24)
$$

Analogous formulas can be obtained for the asymptotics of the cut locus at a point $q_0$ where the invariant $\chi$ is non vanishing.

**Theorem 18.7.** Assume $\chi(q_0) \neq 0$. In a system of local coordinates around $q_0 \in M$ such that $a = 2\chi u_1 u_2$ one gets

$$
\text{Cut}_{q_0}(\theta, \nu) = q_0 \pm \pi \nu^2 f_0(q_0) \pm 2\pi \chi(q_0) \cos \theta f_1(q_0) \nu^3 + O(\nu^4), \quad \nu \to 0^\pm
$$

Moreover the cut length satisfies

$$
\ell_{cut}(\theta, \nu) = 2\pi |\nu| - \pi (\kappa + 2\chi \sin^2 \theta) |\nu|^3 + O(\nu^4), \quad \nu \to 0^\pm \quad (18.25)
$$
We can collect the information given by the asymptotics of the conjugate and the cut loci in Figure 18.1.

All geometrical information about the structure of these sets is encoded in a pair of quadratic forms defined on the fiber at the base point $q_0$, namely the curvature $R$ and the sub-Riemannian Hamiltonian $H$.

Recall that the sub-Riemannian Hamiltonian encodes the information about the distribution and about the metric defined on it (see Exercise 4.3).

Let us consider the kernel of the sub-Riemannian Hamiltonian

$$\ker H = \{ \lambda \in T^*_q M : \langle \lambda, v \rangle = 0, \forall v \in D_q \} = D_q^\perp.$$  \hspace{1cm} (18.26)

The restriction of $R$ to the 1-dimensional subspace $D_q^\perp$ for every $q \in M$, is a strictly positive quadratic form. Moreover it is equal to $1/10$ when evaluated on the Reeb vector field. Hence the curvature $R$ encodes both the contact form $\omega$ and its normalization.

If we denote by $D_q^*$ the orthogonal complement of $D_q^\perp$ in the fiber with respect to $R^k$, we have that $R$ is a quadratic form on $D_q^*$ and, by using the Euclidean metric defined by $H$ on $D_q$, as a symmetric operator.

As we explained in the previous chapter, at each $q_0$ where $\chi(q_0) \neq 0$ there always exists a frame such that

$$\{ H, h_0 \} = 2\chi h_1 h_2$$

\footnote{this is indeed isomorphic to the space of linear functionals defined on $D_q$.}
and in this frame we can express the restriction of $\mathcal{R}$ to $D_q^*$ (corresponding to the set $\{h_0 = 0\}$) on this subspace as follows (see Section 17.1.1)

$$10\mathcal{R} = (\kappa + 3\chi)h_1^2 + (\kappa - 3\chi)h_2^2.$$ 

From this formulae it is easy to recover the two invariants $\chi, \kappa$ considering

$$\text{trace}(10\mathcal{R}|_{h_0=0}) = 2\kappa, \quad \text{discr}(10\mathcal{R}|_{h_0=0}) = 36\chi^2,$$

where the discriminant of an operator $Q$, defined on a two-dimensional space, is defined as the square of the difference of its eigenvalues, and can be compute by the formula $\text{discr}(Q) = \text{trace}^2(Q) - 4\det(Q)$.

The cubic term of the conjugate locus (for a fixed value of $\nu$) parametrizes an astroid. The cuspidal directions of the astroid are given by the eigenvectors of $R$, and the cut locus intersect the conjugate locus exactly at the cuspidal points in the direction of the eigenvector of $R$ corresponding to the larger eigenvalue.

Finally the “size” of the cut locus increases for bigger values of $\chi$, while $\kappa$ is involved in the length of curves arriving at cut/conjugate locus.

**Remark 18.8.** The expression of the cut locus given in Theorem 18.7 gives the truncation up to order 3 of the asymptotics of the cut locus of the exponential map. It is possible to show that this is actually the exact cut locus corresponding to the truncated exponential map at order 3, which is the object of the next sections (see Section 18.3.4).

### 18.3.1 Proof of Theorem 18.6: asymptotics of the exponential map

The proof of Theorem 18.6 requires a careful analysis of the asymptotic of the exponential map. Let us consider again our Hamiltonian system in the form (18.14)

$$\begin{align*}
\dot{q} &= rf_\theta \\
\dot{\theta} &= 1 - rb \\
\dot{r} &= r^3a \\
\dot{\tau} &= r
\end{align*}$$

(18.27)

where we recall that equations are written with respect to the time $\tau$. In particular, since we restrict on the level set $H^{-1}(1/2)$, the trajectories are parametrized by length and the time $t$ coincides with the length of the curve. Thus in what follows we replace the variable $t$ by $\ell$.

Next, we consider a last change of the time variable. Namely we parametrize trajectories by the coordinate $\theta$. In other words we rewrite again the equations in such a way that $\dot{\theta} = 1$ and the dot will denote derivative with respect to $\theta$. The equations are rewritten in the following form:

$$\begin{align*}
\dot{q} &= \frac{r}{1 - rb}f_\theta \\
\dot{\theta} &= 1 \\
\dot{r} &= \frac{r^3a}{1 - rb} \\
\dot{\ell} &= \frac{1}{1 - rb}
\end{align*}$$

(18.28)
where we recall that \( f_\theta = \cos \theta f_1 + \sin \theta f_2 \). Moreover we define \( F(t; \theta_0, \nu) := q(t + \theta_0; \theta_0, \nu) \), where \( q(\theta_0; \theta_0, \nu) = q_0 \). This means that the curve that corresponds to initial parameter \( \theta_0 \) start from \( q_0 \) at time equal to \( \theta_0 \).

Notice that in (18.28) we can solve the equation for \( r = r(\tau) \) and substitute it in the first equation. In this way we can write the trajectory as an integral curve of the nonautonomous vector field

\[
F(t; \theta_0, \nu) = q_0 \circ Q^\theta_0,\nu, \quad Q^\theta_0,\nu = \exp \int_{\theta_0}^{t_{\theta_0} + t} \frac{r(\tau)}{1 - r(\tau)b(\tau)} f_\tau d\tau.
\]

To simplify the notation in what follows we denote the flow \( Q^\theta_0,\nu \) simply by \( Q_t \) and by \( V_t \) the nonautonomous vector field defined by this flow

\[
Q_t = \exp \int_{\theta_0}^{\theta_0 + t} V_\tau d\tau, \quad V_\tau := \frac{r(\tau)}{1 - r(\tau)b(\tau)} f_\tau.
\]

(18.29)

We start by analyzing the asymptotics of the end point map after time \( t = 2\pi \).

**Lemma 18.9.** \( F(2\pi; \theta_0, \nu) = q_0 - \pi f_0(q_0)\nu^2 + O(\nu^3) \)

**Proof.** From (18.28), recalling that \( r(0) = \nu \), it is easy to see that \( r \) satisfies the identity

\[
r(t) = \nu + \tilde{r}(t)\nu^3 = \nu + O(\nu^3)
\]

for some smooth function \( \tilde{r}(t) \). Thus, to find the second order term in \( \nu \) of the endpoint map \( F(2\pi; \theta, \nu) \), we can then assume that \( r \) is constantly equal to \( \nu = r(0) \).

Using the Volterra expansion (cf. (6.9))

\[
\exp \int_{\theta_0}^{\theta_0 + 2\pi} V_\tau d\tau = \left( \text{Id} + \int_{\theta_0}^{\theta_0 + 2\pi} V_\tau d\tau + \int_{\theta_0}^{\theta_0 + 2\pi} \int_{\theta_0}^{\theta_0 + 2\pi} V_{\tau_2} \circ V_{\tau_1} d\tau_1 d\tau_2 + \ldots \right)
\]

(18.30)

and substituting \( r(\tau) \equiv \nu \) we have the following expansion for the first term in (18.30):

\[
\int_{\theta_0}^{\theta_0 + 2\pi} V_\tau d\tau = \int_{\theta_0}^{\theta_0 + 2\pi} \frac{\nu}{1 - \nu b(\tau)} f_\tau d\tau = \int_{\theta_0}^{\theta_0 + 2\pi} \nu(1 + \nu b(\tau) + O(\nu^2)) f_\tau d\tau, \\
= \nu \int_{\theta_0}^{\theta_0 + 2\pi} f_\tau d\tau + \nu^2 \int_{\theta_0}^{\theta_0 + 2\pi} b(\tau) f_\tau d\tau + O(\nu^3), \\
= \nu^2 \int_{\theta_0}^{\theta_0 + 2\pi} b(\tau) f_\tau d\tau + O(\nu^3)
\]

Notice that the first order term in \( \nu \) vanishes since we integrate over a period and \( \int_{\theta_0}^{\theta_0 + 2\pi} f_\tau d\tau = 0. \)
The second term in (18.30) can be rewritten using Lemma 8.27
\[ \int_0^{\tau_2} \int_0^{\tau_1} V_{r_2} \circ V_{r_1} d\tau_1 d\tau_2 = \frac{1}{2} \int_{\theta_0}^{\theta_0+2\pi} V_r d\tau \circ \int_{\theta_0}^{\theta_0+2\pi} V_r d\tau + \int_0^{\tau_2} \int_0^{\tau_1} [V_{r_2}, V_{r_1}] d\tau_1 d\tau_2 \]
\[ = \frac{\nu^2}{2} \left( \int_{\theta_0}^{\theta_0+2\pi} f_r d\tau \circ \int_{\theta_0}^{\theta_0+2\pi} f_r d\tau + \int_0^{\tau_2} \int_0^{\tau_1} [f_{r_2}, f_{r_1}] d\tau_1 d\tau_2 \right) \]
\[ = \frac{\nu^2}{2} \int_{\theta_0}^{\theta_0+2\pi} \int_{\theta_0}^{\theta_0+2\pi} [f_{r_2}, f_{r_1}] d\tau_1 d\tau_2 \]
where we used again \( \int_{\theta_0}^{\theta_0+2\pi} f_r d\tau = 0 \). Notice that higher order terms in the Volterra expansions are \( O(\nu^3) \). Collecting together the two expansions and recalling that
\[ [f_2, f_1] = f_0 + \alpha_1 f_1 + \alpha_2 f_2 \]
one easily obtains
\[ F(2\pi; \theta_0, \nu) = q_0 + \nu^2 \left( \int_{\theta_0}^{\theta_0+2\pi} b(t) f_t dt + \frac{1}{2} \left[ \int_{\theta_0}^{\tau} f_r d\tau, f_t \right] dt \right) + O(\nu^3) \]
\[ = q_0 - \pi \nu^2 f_0(q_0) + O(\nu^3) \] (18.31)
Notice that the factor \( \pi \) in (18.31) comes out from the evaluation of integrals of kind \( \int_{\theta_0}^{\theta_0+2\pi} \cos^2 \tau d\tau \) and \( \int_{\theta_0}^{\theta_0+2\pi} \sin^2 \tau d\tau \).

Next we prove a symmetry of the exponential map

**Lemma 18.10.** \( F(t; \theta_0, \nu) = F(t; \theta_0 + \pi, -\nu) \)

**Proof.** It is a direct consequence of our geodesic equation. Recall that \( F(t; \theta_0, \nu) = q(t + \theta_0; \theta_0, \nu) \), is the solution of the system, with initial condition \( q(\theta_0; \theta_0, \nu) = q_0 \).

Applying the transformation \( t \mapsto t + \pi \) and \( \nu \mapsto -\nu \) we see that the right hand side of \( \dot{q} \) in (18.28) is preserved while the right hand side of \( \dot{r} \) change sign (we use that \( u_i(t + \pi) = -u_i(t) \), hence \( a(t + \pi) = a(t) \) and \( b(t + \pi) = -b(t) \)). Then, if \( (q(t), r(t)) \) is a solution of the system then \( (q(t + \pi), -r(t + \pi)) \) is also a solution. The lemma follows.

The symmetry property just proved permits to characterize all odd terms in the expansion in \( \nu \) of the exponential map at \( t = 2\pi \), as follows.

**Corollary 18.11.** Consider the expansion
\[ F(2\pi; \theta, \nu) \simeq \sum_{n=0}^{\infty} q_n(\theta)\nu^n. \]
We have the following identities
(i) \( q_n(\theta + \pi) = (-1)^n q_n(\theta) \),
\( q_{2n+1}(\theta) = \frac{1}{2} \int_{\theta}^{\theta + \pi} d\theta \frac{dq_{2n+1}}{d\theta}(\tau) d\tau \).

**Proof.** This is an immediate consequence of Lemma 18.10 and the identity

\[
2q_{2n+1}(\theta) = q_{2n+1}(\theta) - q_{2n+1}(\theta + \pi) = -\int_{\theta}^{\theta + \pi} d\theta \frac{dq_{2n+1}}{d\theta}(\tau) d\tau.
\]

We already computed the terms \( q_1(\theta) \) and \( q_2(\theta) \). To find \( q_3(\theta) \) we start by computing the derivative of the map \( F \) with respect to \( \theta \).

**Lemma 18.12.** \( \frac{\partial F}{\partial \theta_0}(2\pi; \theta_0, \nu) = -\pi [f_0, f_{\theta_0}] q_0 \nu^3 + O(\nu^4) \)

**Proof.** We stress that, since we are now interested to third order term in \( \nu \), we can no more assume that \( r(\tau) \) is constant. Differentiating (3.59) with respect to \( \theta \) gives two terms as follows:

\[
\frac{\partial F}{\partial \theta_0} = \frac{\partial}{\partial \theta_0} (q_0 \circ Q_t) = q_0 \circ \frac{\partial}{\partial \theta_0} \left( \exp \int_{\theta}^{\theta + 2\pi} V_{\tau} d\tau \right)
\]

\[
= q_0 \circ (Q_{2\pi} \circ V_{\theta_0+2\pi} - V_{\theta_0} \circ Q_{2\pi})
\]

Next let us rewrite

\[
Q_{2\pi} \circ V_{\theta_0+2\pi} = Q_{2\pi} \circ V_{\theta_0+2\pi} \circ Q_{2\pi}^{-1} \circ Q_{2\pi}
\]

so that (18.32) can be rewritten as

\[
\frac{\partial F}{\partial \theta_0} = q_0 \circ (\text{Ad} Q_{2\pi} \circ V_{\theta_0+2\pi} - V_{\theta_0} \circ Q_{2\pi})
\]

Thanks to Lemma 18.9 we can write

\[
Q_{2\pi} = \text{Id} - \pi \nu^2 f_0 + O(\nu^3)
\]

that implies the following asymptotics for the action of its adjoint by (6.18)

\[
\text{Ad} Q_{2\pi} = \text{Id} - \pi \nu^2 \text{ad} f_0 + O(\nu^3)
\]

We are left to compute the asymptotic expansion of (18.33). To this goal, recall that \( r = r(\tau) \) satisfies

\[
\dot{r} = \frac{r^3}{1 - r^b} a = r^3 a + O(r^4)
\]

hence we can compute its term of order 3 with respect to \( \nu \)

\[
r(t) = \nu + \nu^3 \int_{\theta_0}^{t} a(\tau) d\tau + O(\nu^4)
\]

This in particular implies that \( r(\theta_0 + 2\pi) = \nu + O(\nu^4) \) since \( \int_{\theta_0}^{\theta_0 + 2\pi} a(t) dt = 0. \)
This allows us to replace \( r(\cdot) \) with \( \nu \) in the term \( V_{\theta_0 + 2\pi} \) since \( r(\theta + 2\pi) = \nu + O(\nu^4) \). Moreover using that \( b(\theta_0 + 2\pi) = b(\theta_0) \) and \( f_{\theta_0 + 2\pi} = f_{\theta_0} \) we get

\[
\text{Ad} Q_{2\pi} \circ V_{\theta_0 + 2\pi} - V_{\theta_0} = (\text{Id} - \pi \nu^2 \text{ad} f_0 + O(\nu^3)) \left( \frac{\nu}{1 - \nu b} f_{\theta_0} \right) - \left( \frac{\nu}{1 - \nu b} f_{\theta_0} \right) + O(\nu^4)
\]

\[
= -\pi \nu^2 \text{ad} f_0(\nu f_{\theta_0}) + O(\nu^4)
\]

(18.36)

and finally plugging (18.34) and (18.36) into (18.33) one obtains

\[
\partial F \partial \theta = q_0 \circ (-\pi \nu^2 \text{ad} f_0(\nu f_{\theta_0}) + O(\nu^4)) \circ (\text{Id} + O(\nu))
\]

\[
= q_0 \circ (-\pi \nu^3 [f_0, f_{\theta_0}] + O(\nu^4))
\]

18.3.2 Asymptotics of the conjugate locus

In this section we finally prove Theorem 18.6 by computing the expansion of the conjugate time \( t_c(\theta_0, \nu) \). We know from Proposition 18.3 that

\[
\tau_c(\theta_0, \nu) = 2\pi + \nu^2 s(\theta_0) + O(\nu^3)
\]

By definition of conjugate point, the function \( s = s(\theta_0) \) is characterized as the solution of the equation

\[
\frac{\partial F}{\partial s} \wedge \frac{\partial F}{\partial \theta} \wedge \frac{\partial F}{\partial \nu} \bigg|_{(2\pi + \nu^2 s, \theta, \nu)} = 0,
\]

(18.37)

where \( s \) is considered as a parameter. Notice that the derivative with respect to \( s \) is computed by

\[
\frac{\partial F}{\partial s} = \frac{\partial F}{\partial t} \frac{\partial t}{\partial s} = (\nu f_\theta + O(\nu^2))\nu^2 \simeq \nu^3 f_\theta + O(\nu^4)
\]

Moreover, from the expansion of \( F \) with respect to \( \nu \) one has

\[
\frac{\partial F}{\partial \nu} = -2\pi \nu f_0 + O(\nu^2)
\]

Thus

\[
F(2\pi + \nu^2 s; \theta, \nu) = F(2\pi, \theta, \nu) + \nu^3 s f_\theta + O(\nu^4)
\]

and differentiation with respect to \( \theta_0 \) together with Lemma 18.12 gives

\[
\frac{\partial F}{\partial \theta} (2\pi + \nu^2 s; \theta, \nu) = \nu^3 (\pi [f_\theta, f_0] + s f_{\theta'} + O(\nu^4)
\]

where as usual \( f_{\theta'} \) denotes the derivative with respect to \( \theta \).

Then, collecting together all these computations, the equation for conjugate points (18.37) can be rewritten as

\[
f_\theta \wedge (s f_{\theta'} + \pi [f_\theta, f_0]) \wedge f_0 = O(\nu)
\]

(18.38)

Since \( f_\theta, f_{\theta'} \) are an orthonormal frame on \( D \) and \( f_0 \) is transversal to the distribution, (18.38) is equivalent to

\[
f_\theta \wedge (s f_{\theta'} + \pi [f_\theta, f_0]) = O(\nu)
\]

347
that implies
\[ s(\theta) = \pi \langle [f_0, f_\theta], f_{\theta'} \rangle + O(\nu) \]
where \( \langle \cdot, \cdot \rangle \) denotes the the scalar product on the distribution. Hence
\[ t_c(\theta, \nu) = 2\pi + \pi\nu^2 \langle [f_0, f_\theta], f_{\theta'} \rangle_{q_0} + O(\nu^3) \]
To find the expression of conjugate locus, we evaluate the exponential map at time \( t_c(\theta, \nu) \).

We first consider the asymptotic of the conjugate locus. Using again that the first order term with respect to \( \nu \) of \( \partial F \) is \( \nu f_\bar{\theta} \) we have
\[ F(2\pi + \nu^2 s(\theta_0), \theta_0, \nu) = F(2\pi; \theta_0, \nu) + \nu^3 s(\theta_0) f_{\theta_0} + O(\nu^4) \]
Hence, by Corollary [18.11] and Lemma [18.9] one gets
\[ \text{Con}_{q_0}(\theta_0, \nu) = q_0 - \nu^2 f_0(q_0) - \nu^3 \frac{\nu^3}{2} \int_{\theta_0}^{\theta_0 + \pi} dq_3 \frac{d}{d\tau} + \nu^3 s(\theta_0) f_{\theta_0} + O(\nu^4) \]
Moreover, since
\[ \frac{\partial F}{\partial \theta_0}(2\pi, \nu, \theta_0) = \nu^3 [f_{\theta_0}, f_0] + O(\nu^4) \]
we have by definition that \( q_3(\theta) = [f_\bar{\theta}, f_0] \) and
\[ \text{Con}_{q_0}(\theta_0, \nu) = q_0 - \nu^2 f_0(q_0) - \nu^3 \frac{\nu^3}{2} \int_{\theta_0}^{\theta_0 + \pi} \pi [f_{\theta_0}, f_0] d\tau + \nu^3 s(\theta_0) f_{\theta_0} \]
\[ = q_0 - \nu^2 f_0(q_0) - \nu^3 \frac{\nu^3}{2} \int_{\theta_0}^{\theta_0 + \pi} \pi [f_{\theta_0}, f_0] + s'(t) f_{\theta_0} + s(t) f_{\theta_0} dt \] (18.39)
where the last identify follows by writing \( f_{\theta'} = -f_\theta \) and integrating by parts. Using that
\[ s(\theta) = \pi \langle [f_0, f_\theta], f_{\theta'} \rangle \]
\[ s'(\theta) = \pi \langle [f_0, f_{\theta'}], f_{\theta'} \rangle - \pi \langle [f_0, f_\theta], f_\theta \rangle = 2\pi a \]
we can rewrite (18.39) as follows
\[ \pi [f_{\theta_0}, f_0] + s'(t) f_{\theta_0} + s(t) f_{\theta_0} = \pi [f_{\theta_0}, f_0] + 2\pi a f_{\theta_0} + \pi \langle [f_0, f_{\theta_0}], f_{\theta_0} \rangle f_{\theta_0} \]
\[ = \pi \langle [f_{\theta_0}, f_0], f_{\theta_0} \rangle f_{\theta_0} + 2\pi a f_{\theta_0} \]
\[ = 3\pi a f_{\theta_0} \]
Finally
\[ \text{Con}_{q_0}(\theta_0, \nu) = q_0 - \nu^2 f_0(q_0) - \frac{3\nu^3}{2} \int_{\theta_0}^{\theta_0 + \pi} a(\tau) f_\tau d\tau + O(\nu^4) \]
\[ = q_0 - \nu^2 f_0(q_0) + \nu^3 \pi (a' f_{\theta_0} - a_{\theta_0}) + O(\nu^4) \]
18.3.3 Asymptotics of the conjugate length

Similarly, we consider conjugate length. Recall that

\[ \ell_c(\theta, \nu) = \int_{\theta_0}^{\theta_0 + b(\nu) \cdot t} \frac{r(t)}{1 - r(t) Q_t^{b(\nu)} b(t)} \, dt \]

where we replaced \( b(t) \) by its value along the flow \( Q_t^{b(\nu)} b(t) \).

As a first step, notice that we can reduce to an integral over a period, up to higher order terms with respect to \( \nu \). Namely

\[ \ell_c(\theta_0, \nu) = \int_{\theta_0}^{\theta_0 + 2\pi} \frac{r(t)}{1 - r(t) Q_t^{\nu} b(t)} \, dt + \nu^3 s(\theta_0) + O(\nu^4) \quad (18.40) \]

Indeed \( t_c(\theta_0, \nu) = 2\pi + \nu^2 s(\theta) + O(\nu^3) \) and the first order term w.r.t. \( \nu \) in the integrand is exactly \( \nu \) by \( \text{[18.35]} \). In what follows we use again the notation \( Q_t := Q_t^{b(\nu)} \), and we compute the expansion in \( \nu \) of the integral appearing in \( \text{[18.40]} \).

First notice that

\[ \frac{r(t)}{1 - r(t) Q_t^{\nu} b(t)} = r(t) \left( 1 + r(t) Q_t^{b(\nu)} b(t) + r^2(t) [Q_t^{b(\nu)} b(t) \circ Q_t^{b(\nu)} b(t)] + O(r(t)^3) \right) \]

Using that \( r(t) = \nu + O(\nu^3) \) and \( Q_t^{b(\nu)} b(t) = b(t) + O(\nu) \) we have that

\[ \frac{r(t)}{1 - r(t) Q_t^{\nu} b(t)} = r(t) + r^2(t) Q_t^{b(\nu)} b(t) + r^3(t) b(t)^2 + O(\nu^4) \]

Now each addend of the sum expands as follows

\[ r(t) = \nu + \nu^3 \int_0^t a(t) \, dt + O(\nu^4) \quad (18.41) \]

\[ r^2(t) Q_t^{(\nu)} b(t) = (\nu^2 + O(\nu^4)) \left( \text{Id} + \nu \int_0^t f_\tau \, d\tau + O(\nu) \right) b(t) \quad (18.42) \]

\[ = \nu^2 b(t) + \nu^3 \int_0^t f_\tau \, d\tau b(t) + O(\nu^4) \]

\[ r^3(t) b(t)^2 = \nu^3 b(t)^2 + O(\nu^4) \quad (18.43) \]

\[ r^3(t) b(t)^2 = \nu^3 b(t)^2 + O(\nu^4) \quad (18.44) \]

Integrating the sum over the interval \([\theta_0, \theta_0 + 2\pi]\) and considering terms only up to \( O(\nu^4) \) we have

\[ \ell_c(\theta_0, \nu) = 2\pi \nu + \left( \int_{\theta_0}^{\theta_0 + 2\pi} \left[ \int_0^t a(\tau) \, d\tau + \int_0^t f_\tau \, d\tau \right] b(t) + b^2(t) \, dt \right) \nu^3 + O(\nu^4) \]

where the coefficient in \( \nu^2 \) vanishes since \( \int_{\theta_0}^{\theta_0 + 2\pi} b(\tau) \, d\tau = 0 \). A straightforward computation of the integrals ends the proof of the theorem.
18.3.4 Stability of the conjugate locus

In this section we want to prove that the third order Taylor polynomial of the exponential map corresponds to a stable map in the sense of singularity theory. More precisely it can be treated as a one parameter family of maps between 2-dimensional manifolds that has only singular points of “cusp” and “fold” type. As a consequence the original exponential map can be treated as a perturbation of the (truncated) stable one.

The classic Whitney theorem on the stability of maps between 2-dimensional manifolds then implies that the structure of their singularity will be the same, and actually the singular set of the perturbed one is the image under an homeomorphism of the singular set of the truncated map.

Fix some local coordinates \((x_0, x_1, x_2)\) around the point \(q_0\) such that

\[ q_0 = (0, 0, 0), \quad f_i(q_0) = \partial x_i, \quad \forall i = 0, 1, 2. \]

**Lemma 18.13.** In these coordinates we have

\[
\frac{1}{\pi} F(2\pi + \pi \eta^2 \tau, \theta, \nu) = (x_0(\tau, \theta, \nu), x_1(\tau, \theta, \nu), x_2(\tau, \theta, \nu))
= (-\nu^2, (\tau - c_{02}) \cos(\theta) \nu^3, (\tau + c_{01}^2) \sin(\theta) \nu^3) + O(\nu^4)
\] (18.45)

Let us define the new variable \(\zeta = \sqrt{-x_0(\tau, \theta, \nu)} = \sqrt{\nu^2 + O(\nu^4)} = \nu + O(\nu^3)\) and apply the smooth change of variables \((\tau, \theta, \nu) \mapsto (\tau, \theta, \zeta)\). The map (18.45) is rewritten as follows

\[
\frac{1}{\pi} F(2\pi + \pi \eta^2 \tau, \theta, \nu) = (-\zeta^2, (\tau - c_{02}) \cos(\theta) \zeta^3 + O(\zeta^4), (\tau + c_{01}^2) \sin(\theta) \zeta^3 + O(\zeta^4))
\] (18.46)

Notice that the first coordinate function of this map is constant in the new variables, when \(\zeta\) is constant. The map (18.46) can be interpreted as a family of maps, parametrized by \(\zeta\), depending on two variables

\[
\frac{1}{\pi} F(2\pi + \pi \eta^2 \tau, \theta, \nu) = (-\zeta^2, \zeta^3 \Phi_\zeta(\tau, \theta))
\] (18.47)

where we have defined

\[
\Phi_\zeta(\tau, \theta) = ((\tau - c_{02}) \cos(\theta), (\tau + c_{01}^2) \sin(\theta)) + O(\zeta)
\] (18.48)

The critical set of the map \(\Phi_0(\tau, \theta)\) is a smooth closed curve in \(\mathbb{R} \times S^1\) defined by the equation

\[
\tau = c_{02} \sin^2(\theta) - c_{01}^2 \cos^2(\theta).
\] (18.49)

The critical values of this map, that is the image under the map \(\Phi_0\) of the set defined by (18.49), is the astroid

\[
\mathcal{A}_0 = \{2\chi(-\sin^3(\theta), \cos^3(\theta)), \theta \in S^1\}
\] (18.50)

The restriction to \(\Phi_0\) to the set \(\mathcal{A}_0\) is a one-to-one map. Moreover every critical point of \(\Phi_0\) is a fold or a cusp. This implies that \(\Phi_0\) is a Whitney map. Hence it is stable, in the sense of Thom-Mather theory, see [?, ?].

In other words, for any compact \(K \subset \mathbb{R} \times S^1\) big enough, there exists \(\varepsilon > 0\) such that for all \(\zeta \in [0, \varepsilon]\), the map \(\Phi_\zeta|_K\) is equivalent to \(\Phi_0|_K\), under a smooth family of change of coordinates in the source and in the image. Moreover, this family can be chosen to be smooth with respect to the parameter \(\zeta\).

Collecting these results, we have proved that the shape of the conjugate locus described in Figure [18.1] obtained via third order approximation of the end-point map is indeed a picture of the true shape.
Theorem 18.14. Suppose $M$ is a 3D contact sub-Riemannian structure and $\chi(q_0) \neq 0$. Then there exists $\varepsilon > 0$ such that for every closed ball $B = B(q_0, r)$ with $r \leq \varepsilon$ there exists an open set $U \subset B \setminus \{q_0\}$ and a diffeomorphism $\Psi : U \to \mathbb{R}^3 \times \{\pm 1\}$ such that $B \cap \text{Con}_{q_0} \subset U$ and

$$
\Psi(B \cap \text{Con}_{q_0}) = \{(\zeta^2, \cos^3(\theta)\zeta^3, -\sin^3(\theta)\zeta^3) : \zeta > 0, \theta \in S^1 \times \{\pm 1\}\}.
$$

In particular, each of the two connected components of $B \cap \text{Con}_{q_0}$ contains 4 cuspidal edges.

A similar statement concerning the stability of the cut locus can be found in [1].
Chapter 19

The volume in sub-Riemannian geometry

19.1 The Popp volume

For an equiregular sub-Riemannian manifold $M$, Popp’s volume is a smooth volume which is canonically associated with the sub-Riemannian structure, and it is a natural generalization of the Riemannian one. In this chapter we define the Popp volume and we prove a general formula for its expression, written in terms of a frame adapted to the sub-Riemannian distribution.

As a first application of this result, we prove an explicit formula for the canonical sub-Laplacian, namely the one associated with Popp’s volume. Finally, we discuss sub-Riemannian isometries, and we prove that they preserve Popp’s volume.

19.2 Popp volume for equiregular sub-Riemannian manifolds

Recall that a distribution $\mathcal{D}$ is equiregular if the growth vector is constant, i.e. for each $i = 1, 2, \ldots, m$, $k_i(q) = \dim(\mathcal{D}_q^i)$ does not depend on $q \in M$. In this case the subspaces $\mathcal{D}_q^i$ are fibres of the higher order distributions $\mathcal{D}_q^i \subset TM$.

For equiregular distributions we will simply talk about growth vector and step of the distribution, without any reference to the point $q$.

Next, we introduce the nilpotentization of the distribution at the point $q$, which is fundamental for the definition of Popp’s volume.

**Definition 19.1.** Let $\mathcal{D}$ be an equiregular distribution of step $m$. The nilpotentization of $\mathcal{D}$ at the point $q \in M$ is the graded vector space

$$\text{gr}_q(\mathcal{D}) = \mathcal{D}_q \oplus \mathcal{D}_q^2 / \mathcal{D}_q \oplus \cdots \oplus \mathcal{D}_q^m / \mathcal{D}_q^{m-1}.$$  

The vector space $\text{gr}_q(\mathcal{D})$ can be endowed with a Lie algebra structure, which respects the grading. Then, there is a unique connected, simply connected group, $\text{Gr}_q(\mathcal{D})$, such that its Lie algebra is $\text{gr}_q(\mathcal{D})$. The global, left-invariant vector fields obtained by the group action on any orthonormal basis of $\mathcal{D}_q \subset \text{gr}_q(\mathcal{D})$ define a sub-Riemannian structure on $\text{Gr}_q(\mathcal{D})$, which is called the nilpotent approximation of the sub-Riemannian structure at the point $q$.

In what follows, we provide the definition of Popp’s volume. Our presentation follows closely the one that can be found in [7]. (See also [27]). The definition rests on the following lemmas.
Lemma 19.2. Let $E$ be an inner product space and $V$ a vector space. Let $\pi : E \to V$ be a surjective linear map. Then $\pi$ induces an inner product on $V$ such that the norm of $v \in V$ is

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

(19.1)

Proof. It is easy to check that Eq. (19.1) defines a norm on $V$. Moreover, since $\| \cdot \|_E$ is induced by an inner product, i.e. it satisfies the parallelogram identity, it follows that $\| \cdot \|_V$ satisfies the parallelogram identity too. Notice that this is equivalent to consider the inner product on $V$ defined by the linear isomorphism $\pi : (\ker \pi)^\perp \to V$. Indeed the norm of $v \in V$ is the norm of the shortest element $e \in \pi^{-1}(v)$.

Lemma 19.3. Let $E$ be a vector space of dimension $n$ with a flag of linear subspaces $\{0\} = F^0 \subset F^1 \subset F^2 \subset \cdots \subset F^m = E$. Let $\text{gr}(F) = F^1 \oplus F^2/F^1 \oplus \cdots \oplus F^m/F^{m-1}$ be the associated graded vector space. Then there is a canonical isomorphism $\theta : \wedge^n E \to \wedge^n \text{gr}(F)$.

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 basis for $E$, i.e. $X_1, \ldots, X_{k_i}$ is a basis for $F^i$. We define the linear map $\theta : E \to \text{gr}(F)$ 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, $\hat{\theta}$ induces a map $\hat{\theta} : \wedge^n E \to \wedge^n \text{gr}(F)$, which sends $X_1 \wedge \ldots \wedge X_n$ to $\hat{\theta}(X_1) \wedge \ldots \wedge \hat{\theta}(X_n)$. The proof that $\theta$ does not depend on the choice of the adapted basis is “dual” to the proof of [27, Lemma 10.4].

The idea behind Popp’s volume is to define an inner product on each $D^i_q/D^{i-1}_q$ which, in turn, induces an inner product on the orthogonal direct sum $\text{gr}_q(D)$. The latter has a natural volume form, which is the canonical volume of an inner product space obtained by wedging the elements an orthonormal dual basis. Then, we employ Lemma 19.3 to define an element of $(\wedge^n T_q M)^* \simeq \wedge^n T^*_q M$, which is Popp’s volume form computed at $q$.

Fix $q \in M$. Then, let $v, w \in D_q$, and let $V, W$ be any horizontal extensions of $v, w$. Namely, $V, W \in \Gamma(D)$ and $V(q) = v, W(q) = w$. The linear map $\pi : D_q \otimes D_q \to D^2_q/D_q$

$$
\pi(v \otimes w) := [V, W]_q \mod D_q,
$$

(19.2)
is well defined, and does not depend on the choice the horizontal extensions. Indeed let $\tilde{V}$ and $\tilde{W}$ be two different horizontal extensions of $v$ and $w$ respectively. Then, in terms of a local frame $X_1, \ldots, X_k$ of $D$

$$
\tilde{V} = V + \sum_{i=1}^k f_i X_i, \quad \tilde{W} = W + \sum_{i=1}^k g_i X_i,
$$

(19.3)

where, for $1 \leq i \leq k$, $f_i, g_i \in \mathcal{C}^\infty(M)$ and $f_i(q) = g_i(q) = 0$. Therefore

$$
[V, W]_q = [V, W] + \sum_{i=1}^k (V(g_i) - W(f_i)) X_i + \sum_{i,j=1}^k f_i g_j [X_i, X_j].
$$

(19.4)

Thus, evaluating at $q$, $[\tilde{V}, \tilde{W}]_q = [V, W]_q \mod D_q$, as claimed. Similarly, let $1 \leq i \leq m$. The linear maps $\pi_i : \otimes^i D_q \to D^i_q/D^{i-1}_q$

$$
\pi_i(v_1 \otimes \cdots \otimes v_i) = [V_1, [V_2, \ldots, [V_{i-1}, V_i]]]_q \mod D^{i-1}_q,
$$

(19.5)
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 $\pi_i$ are surjective and, by Lemma [19.2], they induce an inner product space structure on $D_q/D_q^{-1}$. Therefore, the nilpotentization of the distribution at $q$, namely

\[ \text{gr}_q(D) = D_q \oplus D_q^2 / D_q \oplus \cdots \oplus D_q^m / D_q^{m-1}, \]

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) $\mu_q \in \wedge^n \text{gr}_q(D)^*$, 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 $\text{gr}_q(D)$ to $T_qM$ through the map $\theta_q : \wedge^n T_qM \to \wedge^n \text{gr}_q(D)$ defined in Lemma [19.3]. Namely

\[ P_q = \theta_q^*(\mu_q) = \mu_q \circ \theta_q, \tag{19.7} \]

where $\theta_q^*$ denotes the dual map and we employ the canonical identification $(\wedge^n T_qM)^* \simeq \wedge^n T_q^*M$. Eq. (19.7) is defined only in the domain of the chosen local frame. Since $M$ is orientable, with a standard argument, these $n$-forms can be glued together to obtain Popp’s volume $P \in \Omega^n(M)$. The smoothness of $P$ follows directly from Theorem [19.5].

**Remark 19.4.** The definition of Popp’s volume can be restated as follows. Let $(M, D)$ be an oriented sub-Riemannian manifold. Popp’s volume is the unique volume $P \in \Omega^n(M)$ such that, for all $q \in M$, the following diagram is commutative:

\[
\begin{array}{ccc}
(M, D) & \xrightarrow{P} & (\wedge^n T_qM)^* \\
\text{gr}_q & \downarrow & \downarrow \theta_q^* \\
\text{gr}_q(D) & \xrightarrow{\mu} & (\wedge^n \text{gr}_q(D))^*
\end{array}
\]

where $\mu$ associates the inner product space $\text{gr}_q(D)$ with its canonical volume $\mu_q$, and $\theta_q^*$ is the dual of the map defined in Lemma [19.3].

### 19.3 A formula for Popp volume

In this section we prove an explicit formula for the Popp volume.

We say that a local frame $X_1, \ldots, X_n$ is adapted if $X_1, \ldots, X_{k_i}$ is a local frame for $D_i$, where $k_i := \dim D_i$, and $X_1, \ldots, X_k$ are orthonormal. It is useful to define the functions $c^i_{ij} \in C^\infty(M)$ by

\[ [X_i, X_j] = \sum_{l=1}^n c^i_{lj} X_l. \tag{19.8} \]

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

\[ [X_{i_1}, [X_{i_2}, \ldots, [X_{i_{j-1}}, X_{i_j}]]] = \sum_{l=k_{j-1}+1}^{k_j} b^i_{i_1 i_2 \cdots i_j} X_l \mod D^{j-1}, \tag{19.9} \]
where $1 \leq i_1, \ldots, i_j \leq k$. These are a generalization of the $c^l_{ij}$, with an important difference: the structure constants of Eq. (19.8) are obtained by considering the Lie bracket of all the fields of the local frame, namely $1 \leq i, j, l \leq n$. On the other hand, the adapted structure constants of Eq. (19.9) are obtained by taking the iterated Lie brackets of the first $k$ elements of the adapted frame only (i.e. the local orthonormal frame for $\mathcal{D}$), and considering the appropriate equivalence class. For $j = 2$, the adapted structure constants can be directly compared to the standard ones. Namely $b^l_{ij} = c^l_{ij}$ 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:

$$[B_j]^{hl} = \sum_{i_1, i_2, \ldots, i_j = 1}^k b^h_{i_1 i_2 \ldots i_j} b^l_{i_1 i_2 \ldots i_j}, \quad j = 1, \ldots, m,$$

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

**Theorem 19.5.** Let $X_1, \ldots, X_n$ be a local adapted frame, and let $\nu^1, \ldots, \nu^n$ be the dual frame. Then Popp’s volume $\mathcal{P}$ satisfies

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

where $B_j$ is defined by (19.10) in terms of the adapted structure constants (19.9).

To clarify the geometric meaning of Eq. (19.11), let us consider more closely the case $m = 2$. If $\mathcal{D}$ is a step 2 distribution, we can build a local adapted frame $\{X_1, \ldots, X_k, X_{k+1}, \ldots, X_n\}$ by completing any local orthonormal frame $\{X_1, \ldots, X_k\}$ of the distribution to a local frame of the whole tangent bundle. Even though it may not be evident, it turns out that $B_{2}^{-1}(q)$ is the Gram matrix of the vectors $X_{k+1}, \ldots, X_n$, seen as elements of $T_q M / D_q$. The latter has a natural structure of inner product space, induced by the surjective linear map $\langle \cdot, \cdot \rangle : D_q \otimes D_q \to T_q M / D_q$ (see Lemma 19.2). Therefore, the function appearing at the beginning of Eq. (19.11) is the volume of the parallelepiped whose edges are $X_1, \ldots, X_n$, seen as elements of the orthogonal direct sum $\text{gr}_q(\mathcal{D}) = D_q \oplus T_q M / D_q$.

**Proof of Theorem 19.5**

We are now ready to prove Theorem 19.5. For convenience, we first prove it for a distribution of step $m = 2$. Then, we discuss the general case. In the following subsections, everything is understood to be computed at a fixed point $q \in M$. Namely, by $\text{gr}(\mathcal{D})$ we mean the nilpotentization of $\mathcal{D}$ at the point $q$, and by $\mathcal{D}^i$ we mean the fibre $\mathcal{D}^i_q$ of the appropriate higher order distribution.

**Step 2 distribution**

If $\mathcal{D}$ is a step 2 distribution, then $\mathcal{D}^2 = TM$. The growth vector is $G = (k, n)$. 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

$$[X_i, X_j] = \sum_{l=k+1}^n b^l_{ij} Y_l \mod \mathcal{D}.$$  

(19.12)
For each \( l = k + 1, \ldots, n \), we can think to \( b^l_{ij} \) as the components of an Euclidean vector in \( \mathbb{R}^{k^2} \), which we denote by the symbol \( b^l \). According to the general construction of Popp's volume, we need first to compute the inner product on the orthogonal direct sum \( \text{gr}(D) = D \oplus D^2/D \). By Lemma [19.2], the norm on \( D^2/D \) is induced by the linear map \( \pi : \otimes^2 D \to D^2/D \)
\[
\pi(X_i \otimes X_j) = [X_i, X_j] \mod D. \tag{19.13}
\]
The vector space \( \otimes^2 D \) inherits an inner product from the one on \( D \), namely \( \forall X, Y, Z, W \in D, \langle X \otimes Y, Z \otimes W \rangle = \langle X, Z \rangle \langle Y, W \rangle \). \( \pi \) is surjective, then we identify the range \( D^2/D \) with \( \ker \pi^+ \subset \otimes^2 D \), and define an inner product on \( D^2/D \) by this identification. In order to compute explicitly the norm on \( D^2/D \) (and then, by polarization, the inner product), let \( Y \in D^2/D \). Then
\[
\|D^2/D\|_Y = \min\{\|\otimes^2 D\|_Z \text{ s.t. } \pi(Z) = Y\}. \tag{19.14}
\]
Let \( Y = \sum_{l=k+1}^n c^l Y^l \) and \( Z = \sum_{i,j=1}^k a_{ij} X_i \otimes X_j \in \otimes^2 D \). We can think to \( a_{ij} \) as the components of a vector \( a \in \mathbb{R}^{k^2} \). Then, Eq. (19.14) writes
\[
\|D^2/D\|_Y = \min\{|a| \text{ s.t. } a \cdot b^l = c^l, \ l = k + 1, \ldots, n\}, \tag{19.15}
\]
where \( |a| \) is the Euclidean norm of \( a \), and the dot denotes the Euclidean inner product. Indeed, \( \|D^2/D\|_Y \) is the Euclidean distance of the origin from the affine subspace of \( \mathbb{R}^{k^2} \) defined by the equations \( a \cdot b^l = c^l \) for \( l = k + 1, \ldots, n \). In order to find an explicit expression for \( \|D^2/D\|_Y^2 \) in terms of the \( b^l \), we employ the Lagrange multipliers technique. Then, we look for extremals of
\[
L(a, b^{k+1}, \ldots, b^n, \lambda_{k+1}, \ldots, \lambda_n) = |a|^2 - 2 \sum_{l=k+1}^n \lambda_l (a \cdot b^l - c^l). \tag{19.16}
\]
We obtain the following system
\[
\begin{cases}
\displaystyle{\sum_{l=k+1}^n \lambda_l b^l - a = 0,} \\
\displaystyle{\sum_{l=k+1}^n \lambda_l b^l \cdot b^r = c^r, \quad r = k + 1, \ldots, n.}
\end{cases} \tag{19.17}
\]
Let us define the \( n-k \) square matrix \( B \), with components \( B^{hl} = b^h \cdot b^l \). \( B \) is a Gram matrix, which is positive definite iff the \( b^l \) are \( n-k \) linearly independent vectors. These vectors are exactly the rows of the representative matrix of the linear map \( \pi : \otimes^2 D \to D^2/D \), which has rank \( n-k \). Therefore \( B \) is symmetric and positive definite, hence invertible. It is now easy to write the solution of system (19.17) by employing the matrix \( B^{-1} \), which has components \( B^{-1}_{hl} \). Indeed a straightforward computation leads to
\[
\|D^2/D\|^2_{\epsilon^* Y_s} = \epsilon^h B^{-1}_{hl} \epsilon^l. \tag{19.18}
\]
By polarization, the inner product on \( D^2/D \) is defined, in the basis \( Y_i \), by
\[
\langle Y_i, Y_h \rangle_{D^2/D} = B^{-1}_{ih}. \tag{19.19}
\]
Observe that \( B^{-1} \) is the Gram matrix of the vectors \( Y_{k+1}, \ldots, Y_n \) seen as elements of \( D^2/D \). Then, by the definition of Popp’s volume, if \( \nu^1, \ldots, \nu^k, \mu^{k+1}, \ldots, \mu^n \) is the dual basis associated with \( X_1, \ldots, X_k, Y_{k+1}, \ldots, Y_n \), the following formula holds true
\[
\mathcal{P} = \frac{1}{\sqrt{\det B}} \nu^1 \wedge \cdots \wedge \nu^k \wedge \mu^{k+1} \wedge \cdots \wedge \mu^n. \tag{19.20}
\]
357
General case

In the general case, the procedure above can be carried out with no difficulty. Let \( X_1, \ldots, X_n \) be a local adapted frame for the flag \( \mathcal{D}^0 \subset \mathcal{D} \subset \mathcal{D}^2 \subset \cdots \subset \mathcal{D}^m \). As usual, \( k_i = \dim(\mathcal{D}^i) \). For \( j = 2, \ldots, m \) we define the adapted structure constants \( b^j_{i_1 \ldots i_j} \in C^\infty(M) \) by

\[
[X_{i_1}, [X_{i_2}, \ldots, [X_{i_{j-1}}, X_{i_j}]]] = \sum_{l=k_{j-1}+1}^{k_j} b^l_{i_1 i_2 \ldots i_j} X_l \mod \mathcal{D}^{j-1},
\]

(19.21)

where \( 1 \leq i_1, \ldots, i_j \leq k \). Again, \( b^j_{i_1 \ldots i_j} \) can be seen as the components of a vector \( b^j \in \mathbb{R}^k \).

Recall that for each \( j \) we defined the surjective linear map \( \pi_j : \otimes^j \mathcal{D} \to \mathcal{D}^j / \mathcal{D}^{j-1} \)

\[
\pi_j(X_{i_1} \otimes X_{i_2} \otimes \cdots \otimes X_{i_j}) = [X_{i_1}, [X_{i_2}, \ldots, [X_{i_{j-1}}, X_{i_j}]]] \mod \mathcal{D}^{j-1}.
\]

(19.22)

Then, we compute the norm of an element of \( \mathcal{D}^j / \mathcal{D}^{j-1} \) exactly as in the previous case. It is convenient to define, for each \( 1 \leq j \leq m \), the \( k_j - k_{j-1} \) dimensional square matrix \( B_j \), of components

\[
[B_j]^{hl} = \sum_{i_1, i_2, \ldots, i_j = 1}^k b^h_{i_1 i_2 \ldots i_j} b^l_{i_1 i_2 \ldots i_j}.
\]

(19.23)

with the understanding that \( B_1 \) is the \( k \times k \) identity matrix. Each one of these matrices is symmetric and positive definite, hence invertible, due to the surjectivity of \( \pi_j \). The same computation of the previous case, applied to each \( \mathcal{D}^j / \mathcal{D}^{j-1} \) shows that the matrices \( B_j^{-1} \) are precisely the Gram matrices of the vectors \( X_{k_{j-1}+1}, \ldots, X_{k_j} \in \mathcal{D}^j / \mathcal{D}^{j-1} \), in other words

\[
\langle X_{k_{j-1}+l}, X_{k_{j-1}+h} \rangle_{\mathcal{D}^j / \mathcal{D}^{j-1}} = B_{lh}^{-1}.
\]

(19.24)

Therefore, if \( \nu^1, \ldots, \nu^n \) is the dual frame associated with \( X_1, \ldots, X_n \), Popp’s volume is

\[
\mathcal{P} = \frac{1}{\sqrt{\prod_{j=1}^m \det B_j}} \nu^1 \wedge \ldots \wedge \nu^n.
\]

(19.25)

### 19.4 Popp volume and isometries

In the last part of the paper we discuss the conditions under which a local isometry preserves Popp’s volume. In the Riemannian setting, an isometry is a diffeomorphism such that its differential is an isometry for the Riemannian metric. The concept is easily generalized to the sub-Riemannian case.

**Definition 19.6.** A (local) diffeomorphism \( \phi : M \to M \) is a (local) isometry if its differential \( \phi_* : T\mathcal{M} \to T\mathcal{M} \) preserves the sub-Riemannian structure \( (\mathcal{D}, \langle \cdot | \cdot \rangle) \), namely

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

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

**Remark 19.7.** Condition i), which is trivial in the Riemannian case, is necessary to define isometries in the sub-Riemannian case. Actually, it also implies that all the higher order distributions are preserved by \( \phi_* \), i.e. \( \phi_*(\mathcal{D}^i_q) = \mathcal{D}^i_{\phi(q)} \), for \( 1 \leq i \leq m \).
Definition 19.8. Let $M$ be a manifold equipped with a volume form $\mu \in \Omega^n(M)$. We say that a (local) diffeomorphism $\phi : M \to M$ is a (local) volume preserving transformation if $\phi^* \mu = \mu$.

In the Riemannian case, local isometries are also volume preserving transformations for the Riemannian volume. Then, it is natural to ask whether this is true also in the sub-Riemannian setting, for some choice of the volume. The next proposition states that the answer is positive if we choose Popp’s volume.

Proposition 19.9. Sub-Riemannian (local) isometries are volume preserving transformations for Popp’s volume.

Proposition 19.9 may be false for volumes different than Popp’s one. We have the following.

Proposition 19.10. Let $\text{Iso}(M)$ be the group of isometries of the sub-Riemannian manifold $M$. If $\text{Iso}(M)$ acts transitively on $M$, then Popp’s volume is the unique volume (up to multiplication by scalar constant) such that Proposition 19.9 holds true.

Definition 19.11. Let $M$ be a Lie group. A sub-Riemannian structure $(M, D, \langle \cdot | \cdot \rangle)$ is left invariant if $\forall g \in M$, the left action $L_g : M \to M$ is an isometry.

As a trivial consequence of Proposition 19.9 we recover a well-known result (see again [27]).

Corollary 19.12. Let $(M, D, \langle \cdot | \cdot \rangle)$ be a left-invariant sub-Riemannian structure. Then Popp’s volume is left invariant, i.e. $L_g^* P = P$ for every $g \in M$.

This section is devoted to the proof of Propositions 19.9 and 19.10.

Proof of Proposition 19.9

Let $\phi \in \text{Iso}(M)$ be a (local) isometry, and $1 \leq i \leq m$. The differential $\phi_*$ induces a linear map

$$\tilde{\phi}_* : \otimes^i D_q \to \otimes^i D_{\phi(q)}.$$ (19.26)

Moreover $\phi_*$ preserves the flag $D \subset \ldots \subset D^m$. Therefore, it induces a linear map

$$\hat{\phi}_* : D_q^i/D_q^{i-1} \to D_{\phi(q)}^i/D_{\phi(q)}^{i-1}.$$ (19.27)

The key to the proof of Proposition 19.9 is the following lemma.

Lemma 19.13. $\tilde{\phi}_*$ and $\hat{\phi}_*$ are isometries of inner product spaces.

Proof. The proof for $\tilde{\phi}_*$ is trivial. The proof for $\hat{\phi}_*$ is as follows. Remember that the inner product on $D_q^i/D_q^{i-1}$ is induced by the surjective maps $\pi_i : \otimes^i D \to D_q^i/D_q^{i-1}$ defined by Eq. (19.5). Namely, let $Y \in D_q^i/D_q^{i-1}$. Then

$$\|Y\|_{D_q^i/D_q^{i-1}} = \min \{\|Z\|_{\otimes D_q} \text{ s.t. } \pi_i(Z) = Y \}.$$ (19.28)

As a consequence of the properties of the Lie brackets, $\pi_i \circ \tilde{\phi}_* = \tilde{\phi}_* \circ \pi_i$. Therefore

$$\|Y\|_{D_q^i/D_q^{i-1}} = \min \{\|\tilde{\phi}_* Z\|_{\otimes D_{\phi(q)}} \text{ s.t. } \pi_i(\tilde{\phi}_* Z) = \tilde{\phi}_* Y = \hat{\phi}_* Y\} = \|\hat{\phi}_* Y\|_{D_{\phi(q)}^i/D_{\phi(q)}^{i-1}}.$$ (19.29)

By polarization, $\hat{\phi}_*$ is an isometry. □
Since \( \text{gr}_q(D) = \bigoplus_{i=1}^m D_q^i / D_q^{i-1} \) is an orthogonal direct sum, \( \hat{\phi}_* : \text{gr}_q(D) \rightarrow \text{gr}_{\phi(q)}(D) \) is also an isometry of inner product spaces.

Finally, Popp’s volume is the canonical volume of \( \text{gr}_q(D) \) when the latter is identified with \( T_qM \) through any choice of a local adapted frame. Since \( \phi_* \) is equal to \( \hat{\phi}_* \) under such an identification, and the latter is an isometry of inner product spaces, the result follows.

\[ \Box \]

**Proof of Proposition 19.10**

Let \( \mu \) be a volume form such that \( \phi^* \mu = \mu \) for any isometry \( \phi \in \text{Iso}(M) \). There exists \( f \in C^\infty(M) \), \( f \neq 0 \) such that \( \mathcal{P} = f \mu \). It follows that, for any \( \phi \in \text{Iso}(M) \)

\[
f \mu = \mathcal{P} = \phi^* \mathcal{P} = (f \circ \phi) \phi^* \mu = (f \circ \phi) \mu , \tag{19.30}
\]

where we used the \( \text{Iso}(M) \)-invariance of Popp’s volume. Then also \( f \) is \( \text{Iso}(M) \)-invariant, namely \( \phi^* f = f \) for any \( \phi \in \text{Iso}(M) \). By hypothesis, the action of \( \text{Iso}(M) \) is transitive, then \( f \) is constant.

\[ \Box \]

**Hausdorff dimension and Hausdorff volume**

**Density of the Hausdorff volume with respect to a smooth volume**

**Bibliographical notes**

family
Chapter 20

The sub-Riemannian heat equation

In this chapter we derive the sub-Riemannian heat equation and we discuss the strictly related question of how to define an intrinsic volume in sub-Riemannian geometry.

20.1 The heat equation

To write the heat equation in a sub-Riemannian manifold, let us recall how to write it in the Riemannian context and let us see which mathematical structures are missing in the sub-Riemannian one.

20.1.1 The heat equation in the Riemannian context

Let $(M, g)$ be an oriented Riemannian manifold of dimension $n$ and let $\omega$ the Riemannian volume defined by

$$\omega(X_1, \ldots, X_n) = 1, \text{ where } \{X_1, \ldots, X_n\} \text{ is a local orthonormal frame.}$$

In coordinates if $g$ is represented by a matrix $(g_{ij})$, we have

$$\omega = \sqrt{\det(g_{ij})} \, dx_1 \wedge \ldots \wedge dx_n.$$

Let $\phi$ be a quantity (depending on the position $q$ and the time $t$) subjects to a diffusion process e.g. the temperature of a body, the concentration of a chemical product, the noise etc..... Let $F$ be a time dependent vector field representing the flux of the quantity $\phi$, i.e., how much of $\phi$ is flowing through the unity of surface in unitary time.

Our purpose is to get a partial differential equation describing the evolution of $\phi$. The Riemannian heat equation is obtained by postulating the following two facts:

(R1) the flux is proportional to minus the gradient of $\phi$ i.e., normalizing the proportionality constant to one, we assume that

$$F = -\grad(\phi); \quad (20.1)$$
the quantity φ satisfies a conservation law, i.e. for every bounded open set V having a smooth boundary ∂V we have the following: the rate of decreasing of φ inside V is equal to the rate of flowing of φ via F, out of V, through ∂V. In formulas this is written as

$$-\frac{d}{dt} \int_V \phi \, \omega = \int_{\partial V} F \cdot \nu \, dS. \quad (20.2)$$

Here ν is the external (Riemannian) normal to ∂V and dS is the element of area induced by ω on M, thanks to the Riemannian structure, i.e., dS = ω(ν, ·). The quantity F · ν is a notation for $g_q(F(q,t), ν(q))$.

Applying the Riemannian divergence theorem to (20.2) and using (20.1) we have then

$$-\frac{d}{dt} \int_V \phi \, \omega = \int_{\partial V} F \cdot \nu \, dS = \int_V \text{div}_\omega(F) \, \omega = -\int_V \text{div}_\omega(\text{grad}(\phi)) \, \omega. \quad (20.3)$$

By the arbitrariness of V and defining the Riemannian Laplacian (usually called the Laplace-Beltrami operator) as

$$\Delta \phi = \text{div}_\omega(\text{grad}(\phi)), \quad (20.3)$$

we get the heat equation

$$\frac{\partial}{\partial t} \phi(q,t) = \Delta \phi(q,t). \quad (20.3)$$

Useful expressions for the Riemannian Laplacian

In this section we get some useful expressions for Δ. To this purpose we have to recall what are grad and divω in formula (20.13).

We recall that the gradient of a smooth function ϕ : M → R is a vector field pointing in the direction of the greatest rate of increase of ϕ and its magnitude is the derivative of ϕ in that direction. In formulas it is the unique vector field grad(ϕ) satisfying for every $q \in M$,

$$g_q(\text{grad}(\varphi),v) = d\varphi(v), \quad \text{for every } v \in T_qM. \quad (20.4)$$

In coordinates, if g is represented by a matrix $(g_{ij})$, and calling $(g^{ij})$ its inverse, we have

$$\text{grad}(\varphi)^i = \sum_{j=1}^n g^{ij} \partial_j \varphi. \quad (20.5)$$

If $\{X_1, \ldots, X_n\}$ is a local orthonormal frame for g, we have the useful formula

$$\text{grad}(\varphi) = \sum_{i=1}^n X_i(\varphi)X_i. \quad (20.6)$$
Exercise 20.1. Prove that if the Riemannian metric is defined globally via a generating family \( \{X_1, \ldots, X_m\} \) with \( m \geq n \), in the sense of Chapter 3, then \( \text{grad}(\varphi) = \sum_{i=1}^{m} X_i(\varphi) X_i \).

Recall that the divergence of a smooth vector field \( X \) says how much the flow of \( X \) is increasing or decreasing the volume. It is defined in the following way. The Lie derivative in the direction of \( X \) of the volume form is still a \( n \)-form and hence point-wise proportional to the volume form itself. The “point-wise” constant of proportionality is a smooth function that by definition is the divergence of \( X \). In formulas

\[
L_X \omega = \text{div}_\omega(X) \omega.
\]

Now using \( d\omega = 0 \) and the Cartan formula we have that \( L_X \omega = i_X d\omega + d(i_X \omega) = d(i_X \omega) \). Hence the divergence of a vector field \( X \) can be defined by

\[
d(i_X \omega) = \text{div}_\omega(X) \omega. \tag{20.7}
\]

In coordinates, if \( \omega = h(x) dx^1 \wedge \ldots dx^n \) we have

\[
\text{div}_\omega(X) = \frac{1}{h(x)} \sum_{i=1}^{n} \partial_i(h(x)X^i). \tag{20.8}
\]

Remark 20.2. Notice that to define the divergence of a vector field it is not necessary a Riemannian structure, but only a volume form.

If we put together formula \( \ref{20.5} \) and formula \( \ref{20.8} \) with \( X = \text{grad}(\varphi) \) we get the well known expression

\[
\Delta(\varphi) = \text{div}_\omega(\text{grad}(\varphi)) = \frac{1}{h(x)} \sum_{i,j=1}^{n} \partial_i(h(x)g^{ij} \partial_j \varphi). \tag{20.9}
\]

Combining formula \( \ref{20.6} \) with the property \( \text{div}(aX) = a \text{div}(X) + X(a) \) where \( X \) is a vector field and \( a \) a function, we get

\[
\Delta(\varphi) = \sum_{i=1}^{n} \left( X_i^2 \varphi + \text{div}_\omega(X_i) X_i(\varphi) \right) \text{ where } \{X_1, \ldots X_n\} \text{ is a local orthonormal frame.} \tag{20.10}
\]

Similarly, defining the Riemannian structure via a generating family we get

\[
\Delta(\varphi) = \sum_{i=1}^{m} \left( X_i^2 \varphi + \text{div}_\omega(X_i) X_i(\varphi) \right) \text{ where } \{X_1, \ldots X_m\}, \ m \geq n, \text{ is a generating family} \tag{20.11}
\]

Remark 20.3. Notice that the choice of the volume form does not affect the second order terms, but only the first order ones.

When \( \Delta \) is built with respect to the Riemannian volume form, it is called the Laplace-Beltrami operator.
20.1.2 The heat equation in the sub-Riemannian context

Let $M$ be a sub-Riemannian manifold of dimension $n$. Let $\mathcal{D}$ be the associated set of horizontal vector fields and $g_q$ the corresponding metric on the distribution $\mathcal{D}_q$.

As in the Riemannian case, we assume by simplicity that $M$ is oriented and we assume that a volume form $\omega$ has been assigned on $M$. In Chapter 19 we have seen that, in the equiregular case, the sub-Riemannian structure induces, canonically, a volume form on $M$. For the moment we assume that the volume form is assigned independently of the sub-Riemannian structure.

As in the previous section, we denote by $\phi$ the quantity subject to the diffusion process, by $F$ the corresponding flux, and we postulate that:

(SR1) the heat flows in the direction where $\phi$ is varying more but only among horizontal directions;

(SR2) the quantity $\phi$ satisfies a conservation law, i.e. for every bounded open set $V$ having a smooth and orientable boundary $\partial V$ we have the following: the rate of decreasing of $\phi$ inside $V$ is equal to the rate of flowing of $\phi$ via $F$, out of $V$, through $\partial V$.

To derive the heat equation in the Riemannian case, we have used the following ingredients that are not directly available in the sub-Riemannian context:

- the Riemannian gradient;
- the Riemannian normal to $\partial V$, and the inner product to define the conservation;
- the Riemannian divergence theorem.

Hence the standard Riemannian construction fails in the sub-Riemannian context and we have to reason in a different way to derive the heat equation. Let us analyze one by one the ingredients above and let us see how to generalize them in sub-Riemannian geometry.

The horizontal gradient

In sub-Riemannian geometry the gradient of a smooth function $\varphi : M \to \mathbb{R}$ is a horizontal vector field (called horizontal gradient) pointing in the horizontal direction of the greatest rate of increase of $\varphi$ and its magnitude is the derivative of $\varphi$ in that direction. In formulas it is the unique vector field $\text{grad}_H(\varphi)$ satisfying for every $q \in M$,

$$\langle \text{grad}_H(\varphi) \mid v \rangle_q = d\varphi(v), \text{ for every } v \in \mathcal{D}_q M. \quad (20.12)$$

Here $\langle \cdot \mid \cdot \rangle_q$ is the scalar product induced by the sub-Riemannian structure on $\mathcal{D}_q$ (see Exercise 3.8).

If $\{X_1, \ldots, X_m\}$ is a generating family then

$$\text{grad}_H(\varphi) = \sum_{i=1}^{m} X_i(\varphi) X_i.$$

The postulate (SR1) is then written as

$$F = -\text{grad}_H(\phi).$$
The conservation of the heat

The next step is to express the conservation of the heat without a Riemannian structure. This can be done thanks to the following Lemma, whose proof is left for exercise.

Lemma 20.4. Let $M$ be a smooth manifold provided with a smooth volume form $\omega$. Let $\Omega$ be an embedded bounded sub-manifold (possible with boundary) of codimension 1. Let $F$ be a (possible time dependent) complete smooth vector field and $P_{0,t}$ be the corresponding flow. Consider the cylinder formed by the images of $\Omega$ translated by the flow of $F$ for times between 0 and $t$ (see Figure 20.1):

$$\Pi_F(t, \Omega) = \{ P_{0,t}(\Omega) \mid s \in [0, t] \}.$$

Then

$$\frac{d}{dt} \int_{\Pi_F(t, \Omega)} i_F \omega = \int_\Omega i_F |_{t=0} \omega.$$

With the notation of this Lemma, the postulate (SR2) is written as

$$-\frac{d}{dt} \int_V \phi \omega = \frac{d}{dt} \int_{\Pi_{F(t, \partial V)}} \omega = \int_{\partial V} i_F \omega,$$

where in the last equality we have used the result of the lemma.

Now, using the Stokes theorem, the definition of divergence 20.7 and using that $F = -\text{grad}_H \phi$ we have

$$\int_{\partial V} i_F \omega = \int_V \text{div}_{\partial V} \omega = \int_V \text{div}_{\omega}(F) \omega = -\int_V \text{div}(\text{grad}_{H}(\phi)) \omega.$$

By the arbitrariness of $V$ and defining

$$\Delta_H \phi = \text{div}_{\omega}(\text{grad}_{H}(\phi)),$$

we get the sub-Riemannian heat equation

$$\frac{\partial}{\partial t} \phi(q, t) = \Delta_H \phi(q, t).$$

Definition 20.5. Let $M$ be a sub-Riemannian manifolds and let $\omega$ be a volume on $M$. The operator $\Delta_H \phi = \text{div}_{\omega}(\text{grad}_{H}(\phi))$ is called the sub-Riemannian Laplacian.
When it is possible to construct a volume from the sub-Riemannian structure, then the corresponding sub-Riemannian Laplacian is called the *intrinsic sub-Laplacian*. The construction of a canonical volume form in an equiregular sub-Riemannian manifold has been done in Chapter 19 and it is called the Popp volume. Here let us just remark that in the case of left-invariant structures on Lie groups, the Popp volume is proportional to the left Haar volume (that is a canonical volume that can be built on any Lie group).

**Remark 20.6.** Notice that the expression of the sub-Riemannian Laplacian does not change if we multiply the volume by a (non zero) constant.

20.1.3 Few properties of the sub-Riemannian Laplacian: the Hörmander theorem and the existence of the heat kernel

The same computation of the Riemannian case provides the following expression for the sub-Riemannian Laplacian,

$$\triangle_H(\phi) = \sum_{i=1}^{m} \left( X_i^2 \phi + \text{div}_\omega(X_i)X_i(\phi) \right) \quad \text{where} \quad \{X_1, \ldots, X_m\}, \text{is a generating family.} \quad (20.14)$$

In the Riemannian case, the operator $\triangle_H$ is elliptic, i.e., in coordinates it has the expression

$$\triangle_H = \sum_{i,j=0}^{n} a_{ij}(x)\partial_i\partial_j + \text{first order terms},$$

where the matrix $(a_{ij})$ is symmetric and positive definite for every $x$.

In the sub-Riemannian (and not-Riemannian) case, $\triangle_H$ it is not elliptic since the matrix $(a_{ij})$ can have several zero eigenvalues. However, a theorem of Hörmander says that thanks to the Lie bracket generating condition $\triangle_H$ is hypoelliptic. More precisely we have the following.

**Theorem 20.7** (Hörmander). Let $Y_0, Y_1 \ldots Y_k$ be a set of Lie bracket generating vector fields on a smooth manifold $M$. Then the operator $L = Y_0 + \sum_{i=1}^{k} Y_i^2$ is hypoelliptic which means that if $\varphi$ is a distribution defined on an open set $\Omega \subset M$, such that $L\varphi$ is $C^\infty$, then $\varphi$ is $C^\infty$ in $\Omega$.

Notice that:

- Elliptic operators with $C^\infty$ coefficients are hypoelliptic.

- The heat operator $\triangle - \partial_t$, where $\triangle$ is the Laplace-Beltrami operator on a Riemannian manifold $M$ is not elliptic (since the matrix of coefficients of the second order derivatives in $\mathbb{R} \times M$ has one zero eigenvalue), but it is hypoelliptic since if $\{X_1 \ldots X_n\}$ is an orthonormal frame, then $Y_0 = \sum_{i=1}^{n} \text{div}_\omega(X)X(X(\phi) - \partial_t$ and $Y_1 := X_1, \ldots Y_n := X_n$ are Lie Bracket generating in $\mathbb{R} \times M$.

- The sub-Riemannian heat operator $\triangle_H - \partial_t$ is hypoelliptic since if $\{X_1 \ldots X_m\}$ is a generating family, then $Y_0 = \sum_{i=1}^{m} \text{div}_\omega(X_i)X_i(X(\phi) - \partial_t$ and $Y_1 := X_1, \ldots Y_m := X_m$ are Lie Bracket generating in $\mathbb{R} \times M$. (The hypoellipticity of $\triangle_H$ alone is consequence of the fact that $\{X_1 \ldots X_m\}$ are Lie Bracket generating on $M$.)
One of the most important consequences of the Hörmander theorem is that the heat evolution smooths out immediately every initial condition. Indeed if one can guarantee that a solution of \((\Delta_H - \partial_t)\varphi = 0\) exists in distributional sense in an open set \(\Omega \subset \mathbb{R} \times M\), then, being \(0 \in C^\infty\), it follows that \(\varphi\) is \(C^\infty\) in \(\Omega\).

A standard result for the existence of a solution in \(L^2(M,\omega)\) is given by the following theorem. See for instance \([30, 37]\).

**Theorem 20.8** (Stone). Let \(M\) be a smooth manifold and \(\omega\) a volume on \(M\). If \(\Delta\) is a non negative and essentially self-adjoint operator on \(L^2(M,\omega)\), then, there exists a unique solution to the Cauchy problem

\[
\begin{cases}
(\partial_t - \Delta)\phi = 0 \\
\phi(q,0) = \phi_0(q) \in L^2(M,\omega),
\end{cases}
\tag{20.15}
\]
on \([0,\infty[\times M\). Moreover for each \(t \in [0,\infty[\) this solution belongs to \(L^2(M,\omega)\).

It is immediate to prove that \(\Delta_H\) is non-negative and symmetric on \(L^2(M,\omega)\). If in addition one can prove that \(\Delta_H\) is essentially self-adjoint, then thanks to the Hörmander theorem, one has that the solution of \((20.15)\) is indeed \(C^\infty\) in \([0,\infty[\times M\).

The discussion of the theory of self-adjoint operators is out of the purpose of this book. However the essential self-adjointness of \(\Delta_H\) is guaranteed by the completeness of the sub-Riemannian manifold as metric space.

**Theorem 20.9** (Strichartz, \([33, 34]\)). Consider a sub-Riemannian manifold that is complete as metric space. Let \(\omega\) be a volume on \(M\). Then \(\Delta_H\) defined on \(C^\infty_c(M)\) is essentially self-adjoint in \(L^2(M,\omega)\).

Typical cases in which the sub-Riemannian manifold is complete are let-invariant structure on Lie groups, sub-Riemannian manifold obtained as restriction of complete Riemannian manifolds, sub-Riemannian structures defined in \(\mathbb{R}^n\) having as generating family a set of sub-linear vector fields.

When the manifold is not complete as metric space (as for instance the standard Euclidean structure on the unitary disc in \(\mathbb{R}^2\)), then to study the Cauchy problem \((20.15)\) one need to specify more the problem (e.g., boundary conditions).

As a consequence of the hypoellipticity of \(\Delta_H - \partial_t\) of Therem 20.8 and of Theorem 20.9 we have

**Corollary 20.10.** Consider a sub-Riemannian manifold that is complete as metric space. Let \(\omega\) be a volume on \(M\). There exists a unique solution to the Cauchy problem \((20.15)\), that is \(C^\infty\) in \([0,\infty[\times M\).

Indeed under the hypothesis of competeness of the manifold one can also guarantee the existence of a convolution kernel.

**Theorem 20.11** (Strichartz, \([33, 34]\)). Consider a sub-Riemannian manifold that is complete as metric space. Let \(\omega\) be a volume on \(M\). Then the unique solution to the Cauchy problem \((20.15)\) on \([0,\infty[\times M\) can be written as

\[
\phi(q,t) = \int_M \phi_0(\bar{q})K_t(q,\bar{q})\omega(\bar{q})
\]

where \(K_t(q,\bar{q})\) is a positive function defined on \([0,\infty[\times M\times M\) which is smooth, symmetric for the exchange of \(q\) and \(\bar{q}\) and such that for every fixed \(t, q\), we have \(K_t(q,\cdot) \in L^2(M,\omega)\).
20.1.4 The heat equation in the non-Lie-bracket generating case

If the sub-Riemannian structure is not Lie-bracket generated, i.e., when we are dealing with a proto-sub-Riemannian structure in the sense of Section 3.1.5 then the operator $\Delta H$ can be defined as above, but in general it is not hypoelliptic and the heat evolution does not smooth the initial condition.

Consider for example the the proto-sub-Riemannian structure on $\mathbb{R}^3$ for which an orthonormal frame is given by $\{\partial_x, \partial_y\}$. Take as volume the Lebesgue volume on $\mathbb{R}^3$. Then $\Delta H = \partial_x^2 + \partial_y^2$ on $\mathbb{R}^3$. This operator is not obtained from Lie-bracket generating vector fields. Consider the corresponding heat operator $\Delta H - \partial_t$ on $]0, \infty[\times \mathbb{R}^3$. Since the $z$ direction is not appearing in this operator, any discontinuity in the $z$ variable is not smoothed by the evolution. For instance if $\psi(x,y,t)$ is a solution of the heat equation $\Delta H - \partial_t = 0$ on $]0, \infty[\times \mathbb{R}^3$, then $\psi(x,y,t)\theta(z)$ is a solution of the heat equation in $]0, \infty[\times \mathbb{R}^3$, where $\theta$ is the Heaviside function.

20.2 The heat-kernel on the Heisenberg group

In this section we construct the heat kernel on the Heisenberg sub-Riemannian structure. To this purpose it is convenient to see this structure as a left-invariant structure on a matrix representation of the Heisenberg group. This point of view is useful to build in a canonical way a volume form and hence the sub-Riemannian Laplacian. Moreover this point of view permits to look for a simplified version of the heat kernel using the group law.

20.2.1 The Heisenberg group as a group of matrices

The Heisenberg group $H_2$ can be seen as the 3-dimensional group of matrices

$$H_2 = \left\{ \begin{pmatrix} 1 & x & z + \frac{1}{2}xy \\ 0 & 1 & y \\ 0 & 0 & 1 \end{pmatrix} \mid x, y, z \in \mathbb{R} \right\}$$

endowed with the standard matrix product. $H_2$ is indeed $\mathbb{R}^3$, endowed with the group law

$$(x_1, y_1, z_1) \cdot (x_2, y_2, z_2) = (x_1 + x_2, y_1 + y_2, z_1 + z_2 + \frac{1}{2}(x_1y_2 - x_2y_1)).$$

This group law comes from the matrix product after making the identification

$$(x, y, z) \sim \begin{pmatrix} 1 & x & z + \frac{1}{2}xy \\ 0 & 1 & y \\ 0 & 0 & 1 \end{pmatrix}.$$ 

The identity of the group is the element $(0, 0, 0)$ and the inverse element is given by the formula

$$(x, y, z)^{-1} = (-x, -y, -z).$$

A basis of its Lie algebra of $H_2$ is $\{p_1, p_2, k\}$ where

$$p_1 = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad p_2 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} \quad k = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \quad (20.16)$$
They satisfy the following commutation rules: \([p_1, p_2] = k, [p_1, k] = [p_2, k] = 0\), hence \(H_2\) is a 2-step nilpotent group.

**Remark 20.12.** Notice that if one write an element of the algebra as \(xp_1 + yp_2 + zk\), one has that

\[
\exp(xp_1 + yp_2 + zk) = \begin{pmatrix} 1 & x & z + \frac{1}{2}xy \\ 0 & 1 & y \\ 0 & 0 & 1 \end{pmatrix}.
\] (20.17)

Hence the coordinates \((x, y, z)\) are the coordinates on the Lie algebra related to the basis \(\{p_1, p_2, k\}\), transported on the group via the exponential map. They are called coordinates of the “first type”.

As we will see later, coordinate \(x, y, w = z + \frac{1}{2}xy\), that are more adapted to the group, are also useful.

The standard sub-Riemannian structure on \(H_2\) is the one having as generating family:

\[
X_1(g) = gp_1, \quad X_2(g) = gp_2.
\]

With a straightforward computation one get the following coordinate expression for the generating family:

\[
X_1 = \partial_x - \frac{y}{2} \partial_z, \quad X_2 = \partial_y + \frac{x}{2} \partial_z,
\]

that we already met several times in the previous chapters.

Let \(L_g\) (resp. \(R_g\)) be the left (resp. right) multiplication on \(H_2\):

\[
L_g : H_2 \ni h \mapsto gh \quad (\text{resp. } R_g : H_2 \ni h \mapsto hg).
\]

**Exercise** Prove that, up to a multiplicative constant, there exist one and only one 3-form \(dh_L\) on \(H_2\) which is left-invariant, i.e. such that \(L_g^* dh = dh\) and that in coordinates coincide (up to a constant) with the Lebesgue measure \(dx \wedge dy \wedge dz\). Prove the same for a right-invariant 3-form \(dh_R\).

The left- and right-invariant forms built in the exercise above are called the left and right Haar measures. Since they coincide up to a constant the Heisenberg group is said to be “unimodular”. In the following we normalise the left and right Haar measures on the sub-Riemannian structure in such a way that

\[
dh_L(X_1, X_2, [X_1, X_2]) = dh_R(X_1, X_2, [X_1, X_2]) = 1.
\] (20.18)

The 3-form obtained in this way coincide with the Lebesgue measure and in the following we call it simply the “Haar measure”

\[
dh = dx \wedge dy \wedge dz.
\]

**Exercise** Prove that the two conditions (20.18) are invariant by change of the orthonormal frame.

### 20.2.2 The heat equation on the Heisenberg group

Given a volume form \(\omega\) on \(\mathbb{R}^3\), the sub-Riemannian Laplacian for the Heisenberg sub-Riemannian structure is given by the formula,

\[
\Delta_H(\phi) = (X_1^2 + X_2^2 + \text{div}_\omega(X_1)X_1 + \text{div}_\omega(X_2)X_2)\phi.
\] (20.19)

369
If we take as volume the Haar volume $dh$, and using the fact that $X_1$ and $X_2$ are divergence free with respect to $dh$, we get for the sub-Riemannian Laplacian

$$\Delta_H(\phi) = (X_1)^2 + (X_2)^2 = (\partial_x - \frac{y}{2}\partial_z)^2 + (\partial_y + \frac{x}{2}\partial_z)^2.$$  (20.20)

The heat equation on the Heisenberg group is then

$$\partial_t \phi(x, y, z, t) = \Delta_H(\phi) = ((\partial_x - \frac{y}{2}\partial_z)^2 + (\partial_y + \frac{x}{2}\partial_z)^2)\phi(x, y, z, t)$$

For this equation, we are looking for the heat kernel, namely a function $K_t(q, \bar{q})$ such that the solution to the Cauchy problem

$$\begin{cases}
(\Delta_H - \partial_t)\phi = 0 \\
\phi(q, 0) = \phi_0(q) \in L^2(\mathbb{R}^3, dh)
\end{cases}$$  (20.21)

can be expressed as

$$\phi(q, t) = \int_{\mathbb{R}^3} K_t(q, \bar{q})\phi_0(\bar{q})dh(\bar{q}).$$  (20.22)

The existence of a heat kernel that is smooth, positive and symmetric is guaranteed by Theorem 20.9 since the Heisenberg group (as sub-Riemannian structure) is complete.

The construction of the explicit expression of the heat kernel on the Heisenberg group was an important achievement of the end of the seventies [16, 21]. Here we propose an elementary direct method, divided in the following step:

**STEP 1.** We look for a special form for $K_t(q, \bar{q})$ using the group law.

**STEP 2.** We make a change of variables in such a way that the coefficients of the heat equation depend only on one variable instead than two.

**STEP 3.** By using the Fourier transform in two variables, we transform the heat equation (that was a PDE in 3 variable plus the time) in a heat equation with an harmonic potential in one variable plus the time.

**STEP 4.** We find the kernel for the heat equation with the harmonic potential, thanks to the Mehler formula for Hermite polynomials.

**STEP 5.** We come back to the original variables.

Let us make these steps one by one.

**STEP 1.** Due to invariance under the group law, we have that for $K_t(q, \bar{q}) = K_t(p \cdot q, p \cdot \bar{q})$ for every $p \in H_2$. Taking $p = q^{-1}$ we have that $K_t(q, \bar{q}) = K_t(0, q^{-1} \bar{q})$ hence we can write

$$K_t(q, \bar{q}) = p_t(q^{-1} \cdot \bar{q}) = p_t(\bar{x} - x, \bar{y} - y, \bar{z} - z) = p_t(x - \bar{x}, y - \bar{y}, z - \bar{z}),$$

for a suitable function $p_t(\cdot)$ called the *fundamental solution*. In the last equality we have used the symmetry of the heat kernel.
STEP 2 Let us make the change the variable $z \rightarrow w$, where
\[ w = z + \frac{1}{2}xy \]

(cf. Remark 20.12). In the new variables we have that the Haar measure is $dh = dx \wedge dy \wedge dw$. The generating family and the sub-Riemannian Laplacian become
\[
X_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \partial_x & 0 \\ 0 & 0 & \partial_x \end{pmatrix} = \partial_x \\
X_2 = \begin{pmatrix} 0 & 0 & 1 \\ 1 & \partial_y + x \partial_w & 0 \\ 0 & 0 & \partial_x \end{pmatrix} = \partial_y + x \partial_w
\]
\[
\Delta_H(\phi) = (X_1)^2 + (X_2)^2 = \partial_x^2 + (\partial_y + x \partial_w)^2.
\]

The new coordinates are very useful since now the coefficients of the different terms in $\Delta_H$ depend only on one variable. We are then looking for the solution to the Cauchy problem
\[
\begin{cases}
\partial_t \varphi(x, y, w, t) = \Delta_H(\varphi(x, y, w, t)) = (\partial_x^2 + (\partial_y + x \partial_w)^2) \varphi(x, y, w, t) \\
\varphi(x, y, w, 0) = \varphi_0(x, y, w) \in L^2(\mathbb{R}^3, dh)
\end{cases}
\]

STEP 3 By making the Fourier transform in $y$ and $w$, we have $\partial_y \rightarrow i\mu$, $\partial_w \rightarrow i\nu$ and the Cauchy problem become
\[
\begin{cases}
\partial_t \hat{\varphi}(x, \mu, \nu, t) = (\partial_x^2 - (\mu + \nu x)^2) \hat{\varphi}(x, \mu, \nu, t) \\
\hat{\varphi}(x, \mu, \nu, 0) = \hat{\varphi}_0(x, \mu, \nu).
\end{cases}
\]

By making the change of variable $x \rightarrow \theta$, where $\mu + \nu x = \nu \theta$, i.e., $\theta = x + \frac{\mu}{\nu}$, we get:
\[
\begin{cases}
\partial_t \tilde{\varphi}^{\mu, \nu}(\theta, t) = (\partial_\theta^2 - \nu^2 \theta^2) \tilde{\varphi}^{\mu, \nu}(\theta, t) \\
\tilde{\varphi}^{\mu, \nu}(\theta, 0) = \tilde{\varphi}_0^{\mu, \nu}(\theta),
\end{cases}
\]

where we set $\tilde{\varphi}^{\mu, \nu}(\theta, t) := \hat{\varphi}(\theta - \frac{\mu}{\nu}, \mu, \nu, t)$, and $\tilde{\varphi}_0^{\mu, \nu}(\theta) = \hat{\varphi}_0(\theta - \frac{\mu}{\nu}, \mu, \nu)$.

STEP 4. We have the following

**Theorem 20.13.** The solution of the Cauchy problem for the evolution of the heat in an harmonic potential, i.e.
\[
\begin{cases}
\partial_t \psi(\theta, t) = (\partial_\theta^2 - \nu^2 \theta^2) \psi(\theta, t) \\
\psi(\theta, 0) = \psi_0(\theta) \in L^2(\mathbb{R}, d\theta)
\end{cases}
\]

can be written in the form of a convolution kernel
\[
\psi(\theta, t) = \int_{\mathbb{R}} Q_t^{\nu}(\theta, \tilde{\theta}) \psi_0(\tilde{\theta}) d\tilde{\theta}.
\]

where
\[
Q_t^{\nu}(\theta, \tilde{\theta}) := \sqrt{\frac{\nu}{2\pi \sinh(2\nu t)}} \exp \left( -\frac{1}{2} \frac{\nu^2 \cosh(2\nu t)}{\sinh(2\nu t)} (\theta^2 + \tilde{\theta}^2) + \frac{\nu \theta \tilde{\theta}}{\sinh(2\nu t)} \right).
\]
Remark 20.14. In the case $\nu = 0$ we interpret $Q_0^\nu(\theta, \bar{\theta})$ as
\[
\lim_{\nu \to 0} Q_0^\nu(\theta, \bar{\theta}) = \frac{1}{\sqrt{4\pi t}} \exp[-\frac{(\theta - \bar{\theta})^2}{4t}] .
\] (20.31)

Proof. For $\nu = 0$, equation (20.29) is the standard heat equation on $\mathbb{R}$ and the heat kernel is given by formula (20.31). See for instance [71]. In the following we assume $\nu \neq 0$. The eigenvalues and the eigenfunctions of the operator $\frac{\partial^2}{\partial \theta^2} - \nu^2 \theta^2$ on $\mathbb{R}$ are (see Appendix ??)
\[
E_j = -2\nu(j + 1/2)
\]
\[
\phi_j^{\nu}(\theta) = \frac{1}{\sqrt{2^j j!}} \left( \frac{\nu}{\pi} \right)^{\frac{j}{2}} \exp(-\frac{\nu^2 \theta^2}{2}) H_j(\sqrt{\nu} \theta)
\] (20.32)
where $H_j$ are the Hermite polynomials
\[
H_j(\theta) = (-1)^j \exp(\theta^2) \frac{d^j}{d\theta^j} \exp(-\theta^2).
\]
Being $\{\phi_j^{\nu}\}_{j \in \mathbb{N}}$ an orthonormal frame of $L^2(\mathbb{R})$, we can write
\[
\psi(\theta, t) = \sum_j C_j(t) \phi_j^{\nu}(\theta).
\]
Using equation (20.29), we obtain that
\[
C_j(t) = C_j(0) \exp(tE_j)
\]
where $C_j(0) = \int_\mathbb{R} \phi_j^{\nu}(\bar{\theta}) \psi_0(\bar{\theta}) d\bar{\theta}$. Hence
\[
\psi(\theta, t) = \int_\mathbb{R} Q_t^{\nu}(\theta, \bar{\theta}) \psi_0(\bar{\theta}) d\bar{\theta}
\]
where
\[
Q_t^{\nu}(\theta, \bar{\theta}) = \sum_j \phi_j^{\nu}(\theta) \phi_j^{\nu}(\bar{\theta}) \exp(tE_j).
\]
After some algebraic manipulations and using the Mehler formula for Hermite polynomials
\[
\sum_j \frac{H_j(\theta)H_j(\bar{\theta})}{2^j j!} (w^j) = (1 - w^2)^{-\frac{1}{2}} \exp \left( \frac{2\theta \bar{\theta}\bar{\omega} - (\theta^2 + \bar{\theta}^2)w^2}{1 - w^2} \right), \quad \forall w \in \mathbb{R}
\]
with $\theta \to \sqrt{\nu} \theta$, $\bar{\theta} \to \sqrt{\nu} \bar{\theta}$, $w \to \exp(-2\nu t)$, one get formula (20.30). \square

Using Theorem 20.13 we can write the solution to 20.29 as
\[
\bar{\phi}_{\mu, \nu}(\theta, t) = \int_\mathbb{R} Q_t^{\nu}(\theta, \bar{\theta}) \psi_0^{\mu, \nu}(\bar{\theta}) d\bar{\theta}.
\]

STEP 5 We now come back to the original variables step by step. We have
\[
\hat{\phi}(x, \mu, \nu, t) = \hat{\phi}_{\mu, \nu}(x + \frac{\mu}{\nu} t) = \int_\mathbb{R} Q_t^{\nu}(x + \frac{\mu}{\nu} \bar{\theta}) \psi_0^{\mu, \nu}(\bar{\theta}) d\bar{\theta} = \int_\mathbb{R} Q_t^{\nu}(x + \frac{\mu}{\nu} \bar{x} + \frac{\mu}{\nu} \bar{\theta}) \hat{\psi}_0(\bar{x}, \mu, \nu) d\bar{x}.
\]
In the last equality we made the change of integration variable $\bar{\theta} \to \bar{x}$ with $\bar{\theta} = x + \frac{\mu}{\nu}$ and we used the fact that $\hat{\varphi}_0^\mu,\nu(x + \frac{\mu}{\nu}) = \hat{\varphi}_0(x, \mu, \nu)$.

Now, using the fact that $\hat{\varphi}_0(x, \mu, \nu)$ is the Fourier transform of the initial condition, i.e.

$$\hat{\varphi}_0(x, \mu, \nu) = \int_{\mathbb{R}} \int_{\mathbb{R}} \varphi(x, y, w)e^{-i\mu y}e^{-i\nu w}dydw,$$

and making the inverse Fourier transform we get

$$\varphi(x, y, w, t) = \frac{1}{(2\pi)^2} \int_{\mathbb{R}} \int_{\mathbb{R}} \hat{\varphi}(x, \mu, \nu, t)e^{i\mu y}e^{i\nu w}d\mu d\nu = \int_{\mathbb{R}^3} \left( \frac{1}{(2\pi)^2} \int_{\mathbb{R}} \int_{\mathbb{R}} Q_t^\nu(x + \frac{\mu}{\nu}, \bar{x} + \frac{\mu}{\nu})e^{i\mu(y-\bar{y})}e^{i\nu(w-\bar{w})}d\mu d\nu \right) \varphi_0(x, \bar{y}, \bar{w})dx dydw.$$

Coming back to the variable $x, y, z$, we have

$$\phi(x, y, z, t) = \varphi(x, y, z + \frac{1}{2}xy) = \int_{\mathbb{R}^3} K_t(x, y, z, \bar{x}, \bar{y}, \bar{z})\varphi_0(\bar{x}, \bar{y}, \bar{z})dx dydz.$$

where

$$K_t(x, y, z, \bar{x}, \bar{y}, \bar{z}) = \frac{1}{(2\pi)^2} \int_{\mathbb{R}} \int_{\mathbb{R}} Q_t^\nu(x + \frac{\mu}{\nu}, \bar{x} + \frac{\mu}{\nu})e^{i\mu(y-\bar{y})}e^{i\nu(z-\bar{z}+\frac{1}{2}(xy-\bar{x}\bar{y}))}d\mu d\nu.$$

Setting $\bar{x}, \bar{y}, \bar{z}$ to zero and after some algebraic manipulations we get for the fundamental solution

$$p_t(x, y, z) = \frac{1}{(2\pi t)^3} \int_{\mathbb{R}} \frac{2\tau}{\sinh(2\tau)} \exp \left( -\frac{\tau(x^2 + y^2)}{2t \tanh(2\tau)} \right) \cos(\frac{z\tau}{t})d\tau. \quad (20.33)$$

The integral representation (20.33) can be computed explicitly on the origin and on the $z$ axis. Indeed we have

$$K_t(0, 0, 0; 0, 0, 0) = p_t(0, 0, 0) = \frac{1}{16t^2} \quad (20.34)$$

$$K_t(0, 0, 0; 0, 0, z) = p_t(0, 0, z) = \frac{1}{8t^2 \left( 1 + \cosh \left( \frac{z}{2t} \right) \right)} = \frac{1}{4t^2} \exp \left( -\frac{d^2(0, 0, 0; 0, 0, z)}{4t} \right) f(t) \quad (20.35)$$

In the last equality we have used the fact that for the Heisenberg group $d(0, 0, 0; 0, 0, z) = \sqrt{4\pi z}$. Here $f(t)$ is a smooth function of $t$ such that $f(0) = 1$ (here $z \neq 0$ is fixed). A more detailed analysis permits to get for every fixed $(x, y, z)$ such that $x^2 + y^2 \neq 0$

$$K_t(0, 0, 0; x, y, z) = p_t(x, y, z) = \frac{C + O(t)}{t^{3/2}} \exp \left( -\frac{d^2(0, 0, 0; x, y, z)}{4t} \right). \quad (20.36)$$

Notice that the asymptotics (20.34), (20.35), (20.36) are deeply different with respect to those in the Euclidean case. Indeed the heat kernel for the standard heat equation in $\mathbb{R}^n$ is given by the formula

$$K_t(0, 0, 0; x, y, z) = \frac{1}{(4\pi t)^{n/2}} \exp \left( -\frac{x^2 + y^2 + z^2}{4t} \right). \quad (20.37)$$
Comparing (20.37) with (20.34), (20.35), (20.36), one has the impression that the heat diffusion on the Heisenberg group at the origin and on the points on the $z$ axis, is similar to the one in an Euclidean space of dimension 4. While on all the other points it is similar to the one in an Euclidean space of dimension 3. Indeed the difference of asymptotics between the Heisenberg and the Euclidean case at the origin is related to the fact that the Hausdorff dimension of the Heisenberg group is 4, while its topological dimension is 3 (See Chapter ??). While the difference of asymptotics on the $z$ axis (without the origin) is related to the fact that these are points reached a one parameter family of optimal geodesics starting from the origin and hence they are at the same time cut and conjugate points. For more details see [?]. It is interesting to remark that on a Riemannian manifold of dimension $n$ the asymptotics are similar to the Euclidean ones for points close enough. Indeed for every $\bar{q}$ close enough to $q$ we have $K_t(q, \bar{q}) = \frac{1+O(t)}{(4\pi t)^{n/2}} \exp \left(-\frac{d(q, \bar{q})^2}{4t}\right)$.
Appendix A

Hermite polynomials
Appendix B

Elliptic functions
Appendix C

Structural equations for curves in Lagrange Grassmannian
Index

2D Riemannian problem, 103
abnormal extremals, 246
AC admissible curve, 86
admissible curve, 62
bracket-generating, 61
bundle map, 57
Carnot-Caratheodory distance, 71
Cartan's formula, 110
characteristic curve, 99
chronological
calculus, 131
exponential
left, 135
right, 133
conjugate point, 164
contact
form, 102
sub-Riemannian structure, 102
cotangent
bundle, 56
cotangent bundle
canonical coordinates, 56
cotangent space, 54
critical point
constrained, 150
differential form, 55
differential of a map, 50
distribution, 62
dual, 101
end-point map, 145
differential, 146
energy functional, 80
Euler vector field, 308
exponential map, 160
extremal
abnormal, 80, 100
normal, 80, 97
path, 80
flag, 223
flow, 47
free
sub-Riemannian structure, 68
fundamental solution of the Heisenberg group,
370
Gauss’s Theorema Egregium, 33
Gauss-Bonnet, 26
global version, 31
local version, 27
Hamiltonian
ODE, 96
sub-Riemannian, 97
system, 96
vector field, 93
Heisenberg group
heat kernel, 368
Hessian, 151
index
of a map, 247
of a quadratic form, 247
induced bundle, 57
integral curve, 46
intrinsic sub-Laplacian, 366
isoperimetric problem, 105
Jacobi curve, 304
reduced, 309
Lagrange
multiplier, 148, 150
multipliers rule, 148, 150
Bibliography


384


