Sub-Riemannian Geometry and Optimal Transport

Ludovic Rifford

May 18, 2013

Preface

The main goal of these lectures is to give an introduction to sub-Riemannian geometry and optimal transport, and to present some of the recent progress in these two fields. This set of notes is divided into three chapters and two appendices. Chapter 1 is concerned with the notions of totally nonholonomic distributions and sub-Riemannian structures. The concepts of End-Point mappings and singular horizontal paths which play a major role through these lectures are introduced here. Chapter 2 deals with sub-Riemannian geodesics. We study first and second-order variations of the End-Point mapping to derive necessary and sufficient conditions for an horizontal path to be minimizing. We provide several examples, including the Montgomery counter-example of singular minimizing curve. In Chapter 3, we study the Monge problem for sub-Riemannian quadratic costs. We give a crash-course in optimal transport theory and explain how the sub-TWIST condition together with the Lipschitz regularity of a "variational" cost implies the well-posedness of Monge's problem. Then we study the fine regularity properties of sub-Riemannian distances to obtain existence and uniqueness of optimal transport maps in the sub-Riemannian context. We recall basic facts on ordinary differential equations in Appendix 1 and less classical results of differential calculus in normed vector spaces in Appendix 2. The latter plays a key role in Chapter 2.

The reader of these notes should be familiar with the basics in differential geometry and measure theory. Possible references in these fields include the textbooks by Lee [Lee03] and Evans-Gariepy [EG92]. For further reading, we strongly encourage the reader to look at other texts in sub-Riemannian geometry and optimal transport. Multiple viewpoints always lead to deeper understanding and may open new directions for research. Among them, we may suggest the textbooks by Montgomery [Mon02], Agrachev, Barilari and Boscain [ABB12], and Villani [Vil08].

This set of notes grew from a series of lectures that I gave during a CIMPA school in Beyrouth, Lebanon, on the invitation of Fernand Pelletier. I take the opportunity of this preface to warmly thank Ali Fardoun, Mohamad Mehdi and Fernand Pelletier who organized the school, Ahmed El Soufi for his support and friendship, and through him the "Centre International de Mathématiques Pures et Appliquées". My gratitude goes also to all faculties and students who attended this sub-Riemannian CIMPA school in making it a success.

Contents

Preface **i**

1 Sub-Riemannian structures **1**
1.1 Totally nonholonomic distributions . . . 1
1.2 Horizontal paths and End-Point mappings . . . 12
1.3 Regular and singular horizontal paths . . . 24
1.4 The Chow-Rashevsky Theorem . . . 37
1.5 Sub-Riemannian structures . . . 41
1.6 Notes and comments . . . 44

2 Sub-Riemannian geodesics **45**
2.1 Minimizing horizontal paths and geodesics . . . 45
2.2 The Hamiltonian geodesic equation . . . 51
2.3 The sub-Riemannian exponential map . . . 58
2.4 The Goh condition . . . 65
2.5 Examples of SR geodesics . . . 74
2.6 Notes and comments . . . 80

3 Introduction to optimal transport **83**
3.1 The Monge and Kantorovitch problems . . . 83
3.2 Optimal plans and Kantorovitch potentials . . . 86
3.3 A generalized Brenier-McCann Theorem . . . 96
3.4 Optimal transport on ideal and Lipschitz SR structures . . . 101
3.5 Other examples . . . 113
3.6 Notes and comments . . . 117

A Ordinary differential equations **121**
A.1 Preliminaries . . . 121
A.2 Existence and uniqueness results . . . 123
A.3 Linear systems . . . 127

B Elements of differential calculus **129**
B.1 First order calculus . . . 129
B.2 Second order study . . . 133

Bibliography **143**

Chapter 1

Sub-Riemannian structures

Throughout all the chapter, M denotes a smooth connected manifold without boundary of dimension $n \geq 2$.

1.1 Totally nonholonomic distributions

Distributions

A smooth *distribution* Δ of rank $m \leq n$ ($m \geq 1$) on M is a rank m subbundle of the tangent bundle TM, that is a smooth map that assigns to each point x of M a linear subspace $\Delta(x)$ of the tangent space T_xM of dimension m. In other terms, for every $x \in M$, there are an open neighborhood $\mathcal{V}_x$ of x in M and m smooth vector fields $X_x^1, \cdots, X_x^m$ linearly independent on $\mathcal{V}_x$ such that

$$\Delta(y) = \mathrm{Span}\Big\{X_x^1(y), \cdots, X_x^m(y)\Big\} \qquad \forall y \in \mathcal{V}_x.$$

Such a family of smooth vector fields is called a *local frame* in $\mathcal{V}_x$ for the distribution Δ. All the distributions which will be considered later will be smooth with constant rank $m \in [1, n]$. Thus, from now on, "distribution" always means "smooth distribution with constant rank". A co-rank k distribution on M is a distribution of rank $m = n - k$ and any smooth vector field X on M such that $X(x) \in \Delta(x)$ for any $x \in M$ is called a section of Δ.

Example 1.1.1. *We call trivial distribution on M the rank n distribution Δ defined by $\Delta(x) = T_xM$ for all $x \in M$. For topological reasons, such a distribution may not admit non-vanishing sections (for example, by the hairy ball theorem, there is no non-vanishing continuous vector fields on any even dimensional sphere).*

Example 1.1.2. *In $\mathbb{R}^3$ with coordinates (x, y, z), the distribution Δ defined by*

$$\Delta(x, y, z) = \mathit{Span}\Big\{X(x, y, z), Y(x, y, z)\Big\} \qquad \forall (x, y, z) \in \mathbb{R}^3$$

with

$$X = \partial_x - \frac{y}{2}\partial_z \quad \textit{and} \quad Y = \partial_y + \frac{x}{2}\partial_z,$$

is a rank 2 (or co-rank 1) distribution on $\mathbb{R}^3$.

Example 1.1.3. *More generally, if $x = (x_1, \ldots, x_n, y_1, \ldots, y_n, z)$ denotes the coordinates in $\mathbb{R}^{2n+1}$ and the $2n$ smooth vector fields $X^1, \ldots, X^n, Y^1, \ldots, Y^n$ are defined by*

$$X^i = \partial_{x_i} - \frac{y_i}{2}\partial_z, \quad Y^i = \partial_{y_i} + \frac{x_i}{2}\partial_z \qquad \forall i = 1, \ldots, n,$$

then the distribution Δ defined by

$$\Delta(x) = \mathit{Span}\Big\{X^1(x), , \ldots, X^n(x), Y^1(x), \ldots, Y^n(x)\Big\} \qquad \forall x \in \mathbb{R}^{2n+1},$$

is a co-rank 1 distribution on $\mathbb{R}^{2n+1}$.

Example 1.1.4. *Let α be a smooth non-degenerate 1-form on M, that is a 1-form which does not vanish ($\alpha_x \neq 0$ for any $x \in M$). The distribution Δ defined as*

$$\Delta(x) = \mathit{Ker}(\alpha_x) \qquad \forall x \in M,$$

is a co-rank 1 distribution on M.

Example 1.1.5. *As an example, consider the unit 3-sphere $\mathbb{S}^3$ in $\mathbb{R}^4$ with coordinates (x_1, y_1, x_2, y_2), that is*

$$\mathbb{S}^3 = \Big\{(x_1, y_1, x_2, y_2) \in \mathbb{R}^4 \,|\, x_1^2 + y_1^2 + x_2^2 + y_2^2 = 1\Big\}.$$

Let α be the smooth non-degenerate 1-form on $\mathbb{S}^3$ defined by

$$\alpha = \Big(x_1 dy_1 - y_1 dx_1 + x_2 dy_2 - y_2 dx_2\Big)_{|_{\mathbb{S}^3}},$$

then $\Delta = \mathit{Ker}(\alpha)$ is a co-rank 1 distribution on $\mathbb{S}^3$.

We say that a given distribution Δ on M admits a *global frame* if there are m smooth vector fields $X^1, \cdots, X^m$ on M such that

$$\Delta(x) = \mathrm{Span}\Big\{X^1(x), \cdots, X^m(x)\Big\} \qquad \forall x \in M.$$

In general, distributions do not admit global frames (see Example 1.1.1). It is worth noticing that in the particular case of $\mathbb{R}^n$ all distributions are trivial.

Proposition 1.1.6. *Any distribution in $\mathbb{R}^n$ admits a global frame.*

Proof. Let us first show how to construct a non-vanishing section of a given distribution in $\mathbb{R}^n$.

Lemma 1.1.7. *Let Δ be a distribution of rank m in $\mathbb{R}^n$. Then there is a non-vanishing smooth vector field X such that $X(x) \in \Delta(x)$, for any $x \in \mathbb{R}^n$.*

Proof of Lemma 1.1.7. Define the multivalued mapping $\delta : \mathbb{R}^n \to 2^{\mathbb{R}^n}$ by

$$\delta(x) = \Big\{v \in \Delta(x) \,|\, |v| = 1\Big\} \qquad \forall x \in \mathbb{R}^n.$$

By construction, δ is locally Lipschitz with respect to the Hausdorff distance on compact subsets of $\mathbb{R}^n$. By compactness of $\bar{B}(0_n, 2)$, there is $\epsilon \in (0, 1)$

such that for any $x, y \in \bar{B}(0_n, 2)$ with $|x-y| < \epsilon$, and any $v \in \delta(x)$, there is $w \in \delta(y)$ such that $|v-w| < 1$. Let $N \geq 2$ be an integer such that the increasing sequence of balls $\mathcal{B}_1, \dots, \mathcal{B}_N$ defined by

$$\mathcal{B}_i = B\left(0_n, i\epsilon\right) \qquad \forall i = 1, \dots, N,$$

satisfies $\bar{B}(0_n, 1) \subset \mathcal{B}_N$. For every $x \in \mathbb{R}^n$, we denote by $\mathrm{Proj}_{\delta(x)}$ the projection onto the $(m-1)$- dimensional sphere $\delta(x)$. Note that the mapping $\mathrm{Proj}_{\delta(x)}$ is well-defined and "smooth" on the open set

$$\mathcal{O}_x = \Big\{ w \in \mathbb{R}^n \mid \langle v, w \rangle \neq 0, \forall v \in \delta(x) \Big\}.$$

For every $i \in \{1, \dots, N-1\}$, consider a smooth mapping $P_i : \mathcal{B}_{i+1} \to \mathcal{B}_i$ such that

$$|P_i(x) - x| < \epsilon \qquad \forall x \in \mathcal{B}_{i+1}. \tag{1.1}$$

Note that such a smooth function exists because $\mathcal{B}_i$ is a ball and $\mathcal{B}_{i+1}$ is contained in the ϵ-neighborhood of $\mathcal{B}_i$. Let $\bar{w} \in \delta(0)$ be fixed. We define the vector field $X : \bar{B}(0_n, 1) \to \mathbb{R}^n$ as follows:
We first set

$$X_1(x) = \mathrm{Proj}_{\delta(x)}(\bar{w}) \qquad \forall x \in \mathcal{B}_1.$$

Then, given $X_i : \mathcal{B}_i \to \mathbb{R}^n$, we define $X_{i+1} : \mathcal{B}_{i+1} \to \mathbb{R}^n$ as

$$X_{i+1}(x) = \mathrm{Proj}_{\delta(x)}\Big(X_i\big(P_i(x)\big)\Big) \qquad \forall x \in \mathcal{B}_{i+1}.$$

By construction (by (1.1) and the definition of ϵ), $X_i\big(P_i(x)\big)$ belongs to $\mathcal{O}_x$ for any $x \in \mathcal{B}_{i+1}$. In conclusion, $X = X_N$ is smooth on $\bar{B}(0_n, 1)$ and satisfies $0_n \neq X(x) \in \delta(x)$ for any $x \in B(0_n, 1)$. Repeating the construction on the annuli $B(0_n, 2) \setminus B(0_n, 1), B(0_n, 3) \setminus B(0_n, 2), \dots$, we obtain a non-vanishing section of Δ on $\mathbb{R}^n$. □

We now prove Proposition 1.1.6 by induction on m. Let Δ be a rank $(m+1)$ distribution on $\mathbb{R}^n$. By Lemma 1.1.7, it admits a non-vanishing section X on $\mathbb{R}^n$. The multivalued mapping $\tilde{\Delta} : \mathbb{R}^n \to 2^{\mathbb{R}^n}$ defined by

$$\tilde{\Delta}(x) = \Delta(x) \cap \Big\{ X(x) \Big\}^{\perp} \qquad \forall x \in \mathbb{R}^n,$$

is a smooth rank m distribution (here $\{X(x)\}^{\perp}$ denotes the space which is orthogonal to $X(x)$ with respect to the Euclidean scalar product). Thus by induction, there are smooth vector fields $X^1, \dots, X^m$ on $\mathbb{R}^n$ such that

$$\tilde{\Delta}(x) = \mathrm{Span}\Big\{ X^1(x), \dots, X^m(x) \Big\} \qquad \forall x \in \mathbb{R}^n.$$

The family $\{X^1, \dots, X^m, X\}$ is a global frame for Δ. □

A finite family of smooth vector fields $\{X^1, \dots, X^k\}$ is called a *generating family* for Δ on M if there holds

$$\Delta(x) = \mathrm{Span}\Big\{ X^1(x), \cdots, X^k(x) \Big\} \qquad \forall x \in M.$$

Any distribution can be represented by a generating family.

Proposition 1.1.8. *Let Δ be a distribution of rank $m \leq n$ on M. Then there are $k = m(n+1)$ smooth vector fields $X^1, \cdots, X^k$ such that $\{X^1, \cdots, X^k\}$ is a generating family for Δ.*

Proof. By definition, for every $x \in M$, there is an open neighborhood $\mathcal{V}_x$ of x in M and m smooth vector fields $X_x^1, \cdots, X_x^m$ linearly independent on $\mathcal{V}_x$ such that

$$\Delta(y) = \mathrm{Span}\Big\{X_x^1(y), \cdots, X_x^m(y)\Big\} \qquad \forall y \in \mathcal{V}_x.$$

Since M is paracompact, there is a locally finite covering $\mathcal{V} = \{\mathcal{V}_i\}_{i\in I}$ where each open set $\mathcal{V}_i$ equals $\mathcal{V}_{x_i}$ for some $x_i \in M$.

Lemma 1.1.9. *There are a locally finite open covering $\{\mathcal{U}_j\}_{j\in J}$ of M and a partition $\cup_{l=1}^{n+1} J_l$ of J such that the following properties are satisfied:*

(a) for every $j \in J$, there is $i = i(j) \in I$ such that $\mathcal{U}_j \subset \mathcal{V}_i$,

(b) for every $l \in \{1, \ldots, n+1\}$ and any $j \neq j' \in J_l$, $\mathcal{U}_j \cap \mathcal{U}_{j'} = \emptyset$.

Proof of Lemma 1.1.9. Recall that every smooth manifold is triangulable. Let $\mathcal{T} = \{\mathcal{T}_t\}_{t\in T}$ be a triangulation of M that refines the covering $\{\mathcal{V}_i\}_{i\in I}$, in the sense that the closure of each face F of $\mathcal{T}$ is a subset of some $\mathcal{V}_i$. For every $\alpha \in \{0, \ldots, n\}$, denote by $\mathcal{T}^\alpha = \{\mathcal{T}_t^\alpha\}_{t\in T_\alpha}$ the family of α-dimensional faces in $\mathcal{T}$. For every $\alpha \in \{0, \ldots, n\}$, we can construct easily a collection of open sets $\mathcal{W}^\alpha = \{\mathcal{W}_s^\alpha\}_{s\in S_\alpha}$ satisfying the following properties:

- $\mathcal{W}^\alpha$ is a refinement of $\{\mathcal{V}_i\}_{i\in I}$,
- $\cup_{t\in T_\alpha} \mathcal{T}_t^\alpha \subset \cup_{s\in S_\alpha} \mathcal{W}_s^\alpha$,
- each $\mathcal{W}_s^\alpha$ is an open neighborhood of some α-dimensional face of $\mathcal{T}^\alpha$,
- for any $s \neq s' \in S_\alpha, \mathcal{W}_s^\alpha \cap \mathcal{W}_{s'}^\alpha = \emptyset$,
- for any $s \neq s' \in S_0, \overline{\mathcal{W}_s^\alpha} \cap \overline{\mathcal{W}_{s'}^\alpha} = \emptyset$,
- for any $\alpha \in \{1, \ldots, n\}$ and any $s \neq s' \in S_\alpha, \overline{\mathcal{W}_s^\alpha} \cap \overline{\mathcal{W}_{s'}^\alpha} \subset \cup_{t\in T_{\alpha-1}} \mathcal{T}_t^{\alpha-1}$.

For that, it suffices to proceed by induction on α and to make use of the properties of a triangulation. We conclude easily. □

Let us now show how to construct for every $r \in \{1, \ldots, m\}$ a family of sections $\{X_1^j, \ldots, X_{n+1}^j \,|\, 1 \leq j \leq r\}$ of Δ such that $\mathrm{Span}\{X_l^j(x) \,|\, 1 \leq j \leq r, 1 \leq l \leq n+1\}$ has dimension $\geq r$ for any $x \in M$. We proceed by induction on r.

First, for each $l \in \{1, \ldots, n+1\}$ and each $j \in J_l$, there is $i = i(j) \in I$ such that $\mathcal{U}_j \subset \mathcal{V}_i = \mathcal{V}_{x_i}$. Modifying $X_i^1 = X_{x_i}^1$ outside $\mathcal{U}_j$ if necessary, we may assume that X_i^1 is defined on M, does not vanish on $\mathcal{U}_j$, and vanishes outside $\mathcal{U}_j$. Define $X_1^1, \ldots, X_{n+1}^1$ by

$$X_l^1 = \sum_{j\in J_l} X_{i(j)}^1 \qquad \forall l = 1, \ldots, n+1.$$

By construction (Lemma 1.1.9 (b)), the interior of the supports of the $X_{i(j)}^1$'s are always disjoint. Therefore, each X_l^1 is a non-vanishing section of Δ on

$\cup_{j\in J_l}\mathcal{U}_j$. This shows that $\mathrm{Span}\{X_l^1(x)\,|\,1\leq l\leq n+1\}$ has dimension ≥ 1 for any $x\in M$.
Assume now that we have constructed a family of smooth vector fields $\{X_i^j,\,|\,1\leq j\leq r, 1\leq i\leq n+1\}$ such that

$$\mathrm{Span}\Big\{X_l^j(x)\,|\,1\leq j\leq r, 1\leq l\leq n+1\Big\}$$

has dimension $\geq r$ for any $x\in M$ (with $r<m$). For every $j\in J$, there is $s=s(j)\in\{1,\ldots,m\}$ such that

$$\mathrm{Span}\Big\{X^s_{x_i(j)}(x), X_l^j(x)\,|\,1\leq j\leq r, 1\leq l\leq n+1\Big\}$$

has dimension $\geq r+1$ for any $x\in\mathcal{U}_j$. Define $X_1^{r+1},\ldots,X_{n+1}^{r+1}$ by

$$X_l^{r+1}=\sum_{j\in J_l}X^{s(j)}_{i(j)}\qquad\forall l=1,\ldots,n+1.$$

We leave the reader to check that by construction (modifying the $X^{s(j)}_{x_i(j)}$'s if necessary as above), the vector space

$$\mathrm{Span}\Big\{X_l^j(x)\,|\,1\leq j\leq r+1, 1\leq l\leq n+1\Big\}$$

has dimension $\geq r+1$ for any $x\in M$. The proof is complete. □

The Hörmander condition

Recall that for any smooth vector fields X,Y on M given by

$$X(x)=\sum_{i=1}^n a_i(x)\partial_{x_i},\quad Y(x)=\sum_{i=1}^n b_i(x)\partial_{x_i},$$

in local coordinates $x=(x_1,\ldots,x_n)$, the Lie bracket $[X,Y]$ is the smooth vector field defined as

$$[X,Y](x)=\sum_{i=1}^n c_i(x)\partial_{x_i},$$

where $c_1,\ldots,c_n$ are the smooth scalar function given by

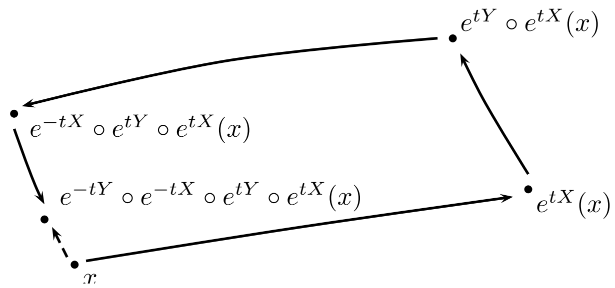
$$c_i=\sum_{j=1}^n\left(\partial_{x_j}b_i\right)a_j-\left(\partial_{x_j}a_i\right)b_j\qquad\forall i=1,\cdots,n.$$

For the upcoming controllability results (like the Chow-Rashesvky Theorem), it is important to keep in mind the following dynamical characterization of the Lie bracket.

Proposition 1.1.10. *Let X,Y be two smooth vector fields in an neighborhood of $x\in\mathbb{R}^n$. Then we have*

$$\begin{aligned}[X,Y](x) &:= D_xY\cdot X(x)-D_xX\cdot Y(x)\\ &= \lim_{t\to 0}\frac{\left(e^{-tY}\circ e^{-tX}\circ e^{tY}\circ e^{tX}\right)(x)-x}{t^2},\end{aligned}\tag{1.2}$$

where e^{tX} and e^{tY} denote respectively the flows of X and Y.



Proof. All the functions appearing in the proof will be defined locally for t close to 0 and/or in a neighborhood of x. Define the smooth function h_4 by

$$h_4(t) := \left(e^{-tY}\circ e^{-tX}\circ e^{tY}\circ e^{tX}\right)(x) \qquad \forall t.$$

We have $h_4'(0)=0$. As a matter of fact, we have for any t,

$$h_4'(t) = -Y(h_4(t)) + \left(\frac{\partial}{\partial x}e^{-tY}\right)_{(t,h_3(t))}\cdot h_3'(t)$$

where h_3 is defined by $h_3(t) := \left(e^{-tX}\circ e^{tY}\circ e^{tX}\right)(x)$. Then we have

$$h_3'(t) = -X(h_3(t)) + \left(\frac{\partial}{\partial x}e^{-tX}\right)_{(t,h_2(t))}\cdot h_2'(t),$$

where $h_2(t) := \left(e^{tY}\circ e^{tX}\right)(x)$ and

$$h_2'(t) = Y(h_2(t)) + \left(\frac{\partial}{\partial x}e^{tY}\right)_{(t,h_1(t))}\cdot h_1'(t),$$

with $h_1(t) := e^{tX}(x)$ and $h_1'(t) = X(e^{tX}(x))$. Since partial derivatives of the form $\frac{\partial}{\partial x}e^{tX}$ at $t=0$ are equal to Id, we get $h_1'(0)=X(x), h_2'(0) = X(x)+Y(x), h_3'(0)=Y(x)$ and $h_4'(0)=0$. Therefore, the left-hand side of (1.2) is equal to $\frac{1}{2}h_4''(0)$. By derivating the above formulas, we get

$$h_1''(0) = dX(h_1(0))\cdot h_1'(0) = dX(x)\cdot X(x),$$

and

$$h_2''(0) = dY(h_2(0))\cdot h_2'(0) + \left[\frac{d}{dt}\left[\left(\frac{\partial}{\partial x}e^{tY}\right)_{(t,h_1(t))}\cdot h_1'(t)\right]\right]_{t=0}.$$

But $dY(h_2(0))\cdot h_2'(0) = dY(x)\cdot(X(x)+Y(x))$ and

$$\begin{aligned}
&\left[\frac{d}{dt}\left[\left(\frac{\partial}{\partial x}e^{tY}\right)_{(t,h_1(t))}\cdot h_1'(t)\right]\right]_{t=0}\\
&= \left[\frac{d}{dt}\left(\frac{\partial}{\partial x}e^{tY}\right)_{(t,h_1(t))}\right]_{t=0}\cdot h_1'(0) + \left(\frac{\partial}{\partial x}e^{tY}\right)_{(0,h_1(0))}\cdot h_1''(0)\\
&= \left[\left(\frac{\partial^2}{\partial t\partial x}\left(e^{tY}\right)\right)_{(0,x)} + \left(\frac{\partial^2}{\partial x^2}\left(e^{tY}\right)\right)_{(0,x)}\cdot h_1'(0)\right]\cdot X(x) + dX(x)\cdot X(x)\\
&= \left(\frac{\partial}{\partial x}\left(\frac{\partial}{\partial t}e^{tY}\right)\right)_{(0,x)}\cdot X(x) + dX(x)\cdot X(x)\\
&= dY(x)\cdot X(x) + dX(x)\cdot X(x).
\end{aligned}$$

We infer that $h_2''(0) = dY(x)\cdot(2X(x)+Y(x)) + dX(x)\cdot X(x)$. In the same way, we have

$$h_3''(0) = -dX(h_3(0))\cdot h_3'(0) + \left[\frac{d}{dt}\left[\left(\frac{\partial}{\partial x}e^{-tX}\right)_{(t,h_2(t))}\cdot h_2'(t)\right]\right]_{t=0},$$

$-dX(h_3(0))\cdot h_3'(0) = -dX(x)\cdot Y(x)$ and

$$\begin{aligned}
&\left[\frac{d}{dt}\left[\left(\frac{\partial}{\partial x}e^{-tX}\right)_{(t,h_2(t))}\cdot h_2'(t)\right]\right]_{t=0}\\
&= \left[\frac{d}{dt}\left(\frac{\partial}{\partial x}e^{-tX}\right)_{(t,h_2(t))}\right]_{t=0}\cdot h_2'(0) + \left(\frac{\partial}{\partial x}e^{-tX}\right)_{(0,h_2(0))}\cdot h_2''(0)\\
&= -dX(x)\cdot(X(x)+Y(x)) + dY(x)\cdot(2X(x)+Y(x)) + dX(x)\cdot X(x)\\
&= -dX(x)\cdot Y(x) + dY(x)\cdot(2X(x)+Y(x)).
\end{aligned}$$

Which implies $h_3''(0) = -2dX(x)\cdot Y(x) + dY(x)\cdot(2X(x)+Y(x))$. Finally

$$\begin{aligned}
h_4''(0) &= -dY(h_4(0))\cdot h_4'(0) + \left[\frac{d}{dt}\left[\left(\frac{\partial}{\partial x}e^{-tY}\right)_{(t,h_3(t))}\cdot h_3'(t)\right]\right]_{t=0}\\
&= \left[\frac{d}{dt}\left(\frac{\partial}{\partial x}e^{-tY}\right)_{(t,h_3(t))}\right]_{t=0}\cdot h_3'(0) + \left(\frac{\partial}{\partial x}e^{-tY}\right)_{(0,h_3(0))}\cdot h_3''(0)\\
&= -dY(x)\cdot Y(x) - 2dX(x)\cdot Y(x) + dY(x)\cdot(2X(x)+Y(x))\\
&= 2(dY(x)\cdot X(x) - dX(x)\cdot Y(x))\\
&= 2[X,Y](x),
\end{aligned}$$

which concludes the proof. □

Remark 1.1.11. *We check easily that the following properties are satisfied:*

(i) Given smooth vector fields X_1, X_2, Y_1, Y_2 and $a_1, a_2 \in \mathbb{R}$, we have

$$\begin{aligned}
[a_1X_1 + a_2X_2, Y_1] &= a_1[X_1,Y_1] + a_2[X_2,Y_1]\\
[X_1, a_1Y_1 + a_2Y_2] &= a_1[X_1,Y_1] + a_2[X_1,Y_2].
\end{aligned}$$

(ii) Given smooth vector fields X and Y, we have $[X,Y] = -[Y,X]$.

(iii) Given three smooth vector fields X, Y, Z, the Jacobi identity is satisfied:

$$\big[X,[Y,Z]\big] + \big[Y,[Z,X]\big] + \big[Z,[X,Y]\big] = 0.$$

Remark 1.1.12. *Given a smooth diffeomorphism ϕ from a smooth manifold $\mathcal{U}$ to a smooth manifold $\mathcal{V}$ and X a smooth vector field on $\mathcal{U}$, we recall that the push-forward $\phi_*(X)$ of X is defined by*

$$\phi_*(X)(y) := D_{\phi^{-1}(y)}\phi\big(X(\phi^{-1}(y))\big) \qquad \forall y \in \mathcal{V}.$$

We have

$$[\phi_*(X), \phi_*(Y)] = \phi_*([X,Y]).$$

For any family $\mathcal{F}$ of smooth vector fields on an open set $\mathcal{O} \subset M$, we denote by Lie($\mathcal{F}$) the Lie algebra of vector fields generated by $\mathcal{F}$. It is the smallest vector subspace S of $\mathcal{X}^\infty(M)$ (the space of smooth vector fields on M) containing $\mathcal{F}$ that also satisfies

$$[X, Y] \in S \qquad \forall X \in \mathcal{F}, \forall Y \in S.$$

It can be constructed as follows: Denote by $\mathrm{Lie}^1(\mathcal{F})$ the space spanned by $\mathcal{F}$ in $\mathcal{X}^\infty(M)$ and define recursively the spaces $\mathrm{Lie}^k(\mathcal{F})$ $(k = 1, 2, \ldots)$ by

$$\mathrm{Lie}^{k+1}(\mathcal{F}) = \mathrm{Span}\left(\mathrm{Lie}^k(\mathcal{F}) \cup \left\{[X, Y] \,|\, X \in \mathcal{F}, Y \in \mathrm{Lie}^k(\mathcal{F})\right\}\right) \qquad \forall k \geq 0.$$

This defines an increasing sequence of vector spaces in $\mathcal{X}^\infty(M)$ satisfying

$$\mathrm{Lie}(\mathcal{F}) = \bigcup_{k \geq 1} \mathrm{Lie}^k(\mathcal{F}).$$

In general, Lie($\mathcal{F}$) is an infinite-dimensional subspace of $\mathcal{X}^\infty(M)$.

Example 1.1.13. *Let A be a $n \times n$ real matrix, b be a vector in $\mathbb{R}^n$, and X, Y be the smooth vector fields in $\mathbb{R}^n$ defined by*

$$X(x) = Ax, \quad Y(x) = b \qquad \forall x \in \mathbb{R}^n.$$

The non-zero Lie brackets of X and Y are always constant vector fields of the form

$$ad_X^1(Y) := [X, Y] = -Ab, \quad ad_X^2(Y) := \left[X, ad_X^1(Y)\right] = A^2 b,$$

and

$$ad_X^{k+1}(Y) := \left[X, ad_X^k(Y)\right] = (-1)^{k+1} A^{k+1} b \qquad \forall k \geq 0.$$

By the Cayley-Hamilton Theorem, A^n can be expressed as a linear combination of $A^0, \ldots, A^{n-1}$. Therefore, Lie(X, Y) is the set of vector fields Z in $\mathbb{R}^n$ of the form

$$Z(x) = \lambda A x + \sum_{i=0}^{n-1} \lambda_i A^i b \qquad \forall x \in \mathbb{R}^n,$$

with $\lambda, \lambda_0, \ldots, \lambda_{n-1} \in \mathbb{R}$. It is a finite-dimensional Lie algebra.

Example 1.1.14. *Let X, Y be the two smooth vector fields in $\mathbb{R}^2$ (with coordinates $x = (x_1, x_2)$) defined by*

$$X(x) = \partial_{x_1}, \quad Y(x) = f(x_1)\partial_{x_2} \qquad \forall x \in \mathbb{R}^2,$$

where f is a smooth scalar function. Then, Lie(X, Y) is the space of smooth vector fields spanned by X and

$$ad_Y^k(X) = f^{(k)} \partial_{x_2} \quad \text{for } k \geq 0.$$

Thus, Lie(X, Y) is infinite-dimensional whenever the derivatives of f span an infinite-dimensional space of functions.

For any point $x \in M$, $\mathrm{Lie}(\mathcal{F})(x)$ denotes the set of all tangent vectors $X(x)$ with $X \in \mathrm{Lie}(\mathcal{F})$. It follows that $\mathrm{Lie}(\mathcal{F})(x)$ is always a linear subspace of T_xM, hence finite-dimensional.

Example 1.1.15. *Returning to Example 1.1.14 and denoting by (e_1, e_2) the canonical basis of $\mathbb{R}^2$, we check that*

$$\mathit{Lie}(X,Y)(x) = \mathit{Span}\Big\{e_1, f^{(k)}(x_1)e_2 \,|\, k = 0,1,2,\ldots\Big\} \qquad \forall x \in \mathbb{R}^2.$$

In particular, $\mathit{Lie}(X,Y)(x) = \mathbb{R}e_1$ if $f(x)$ and all its derivatives at x vanish and $\mathit{Lie}(X,Y)(x) = \mathbb{R}^2$ otherwise.

We say that the smooth vector fields $X^1, \ldots, X^m$ satisfy the *Hörmander condition* on some open set $\mathcal{O} \subset M$ if and only if

$$\mathrm{Lie}\Big\{X^1, \cdots, X^m\Big\}(x) = T_xM \qquad \forall x \in \mathcal{O}.$$

A distribution Δ on M is called *totally nonholonomic* on M if for every $x \in M$, there are an open neighborhood $\mathcal{V}_x$ of x in M and a local frame $X_x^1, \cdots, X_x^m$ on $\mathcal{V}_x$ which satisfies the Hörmander condition on $\mathcal{V}_x$. This definition is intrinsic, it does not depend upon the choice of the local frame $X_x^1, \ldots, X_x^m$. This is a consequence of the following result:

Proposition 1.1.16. *Let $\{X^1, \ldots, X^m\}, \{Y^1, \ldots, Y^m\}$ be two families of linearly independent smooth vector fields on an open set $\mathcal{O} \subset M$ such that*

$$\mathit{Span}\Big\{X^1(x), \ldots, X^m(x)\Big\} = \mathit{Span}\Big\{Y^1(x), \ldots, Y^m(x)\Big\} \qquad \forall x \in \mathcal{O}.$$

Then there holds for any integer $k \geq 1$,

$$\mathit{Lie}^k\Big\{X^1, \ldots, X^m\Big\}(x) = \mathit{Lie}^k\Big\{Y^1, \ldots, Y^m\Big\}(x) \qquad \forall x \in \mathcal{O}.$$

Proof. It is sufficient to show that the following inclusion holds for any integer $k \geq 2$,

$$\mathrm{Lie}^k\Big\{X^1, \ldots, X^m\Big\}(x) \subset \mathrm{Lie}^k\Big\{Y^1, \ldots, Y^m\Big\}(x) \qquad \forall x \in \mathcal{O}.$$

Since the $Y^j(x)$ are always linearly independent, there are smooth functions $\alpha_i^j : \mathcal{O} \to \mathbb{R}$ with $i, j = 1, \ldots, m$, such that

$$X^i(x) = \sum_{j=1}^m \alpha_i^j(x) Y^j(x) \qquad \forall x \in \mathcal{O}, \forall i = 1, \ldots, m.$$

Then for every $i = 1, \ldots, m$ and every smooth vector field Z, there holds

$$[X^i, Z] = \left[\sum_{j=1}^m \alpha_i^j Y^j, Z\right] = \sum_{j=1}^m \alpha_i^j [Y^j, Z] - \sum_{j=1}^m d\alpha_i^j(Z) Y^j.$$

Since $\mathrm{Span}\,\big\{X^1(x), \ldots, X^m(x)\big\} \subset \mathrm{Span}\,\big\{Y^1(x), \ldots, Y^m(x)\big\}$ for any x, this shows that

$$\mathrm{Lie}^2\Big\{X^1, \ldots, X^m\Big\}(x) \subset \mathrm{Lie}^2\Big\{Y^1, \ldots, Y^m\Big\}(x) \qquad \forall x \in \mathcal{O}.$$

We conclude easily by an inductive argument. □

We also observe that any generating family for Δ does satisfy the Hörmander condition provided Δ is totally nonholonomic.

Proposition 1.1.17. *Let Δ be a totally nonholonomic distribution on M and $\{X^1, \ldots, X^k\}$ be a generating family for Δ. Then $X^1, \ldots, X^k$ satisfy the Hörmander condition on M.*

Proof. We need to show that

$$\mathrm{Lie}\Big\{X^1, \cdots, X^k\Big\}(x) = T_xM \qquad \forall x \in M.$$

Let $x \in M$ be fixed. By assumption, there is an open neighborhood $\mathcal{V}_x$ and a local frame $Y_x^1, \cdots, Y_x^m$ on $\mathcal{V}_x$ which satisfies the Hörmander condition on $\mathcal{V}_x$. Proceeding as in the proof of Proposition 1.1.16, we show that

$$\mathrm{Lie}^k\Big\{X^1, \ldots, X^k\Big\}(x) \subset \mathrm{Lie}^k\Big\{Y_x^1, \ldots, Y_x^m\Big\}(x),$$

for every integer $k \geq 1$. This proves that $X^1, \ldots, X^k$ satisfy the Hörmander condition on M. $\square$

Remark 1.1.18. *Since for any smooth vector field X, there holds $[X, X] = 0$, a one dimensional distribution cannot be totally nonholonomic.*

Degree of nonholonomy

If Δ is a rank m totally nonholonomic distribution on M, then for every $x \in M$, there are an open neighborhood $\mathcal{V}_x$ of x and m smooth vector fields $X_x^1, \ldots, X_x^m$ which satisfy the Hörmander condition on $\mathcal{V}_x$. We call *degree of nonholonomy* of Δ at x the smallest integer $r = r(x) \geq 1$ such that

$$\mathrm{Lie}^r\Big\{X^1, \ldots, X^m\Big\}(x) = T_xM.$$

Thanks to Proposition 1.1.16, this definition does not depend upon the choice of the local frame. Moreoever, we shall say that Δ is totally nonholonomic of degree r if the nonholonomy degree of any point in M is $\leq r$.

Example 1.1.19. *The distribution given in Example 1.1.2 is totally nonholonomic. We check easily that*

$$[X, Y] = \partial_z \qquad \forall i, j = 1, \ldots, n,$$

which means that Δ has degree 2.

Example 1.1.20. *More generally, the distribution given in Example 1.1.3 is totally nonholonomic of degree 2. We check easily that*

$$[X^i, Y^j] = \delta_{ij}\partial_z \qquad \forall i, j = 1, \ldots, n.$$

Example 1.1.21. *The Martinet distribution in $\mathbb{R}^3$ (with coordinates (x, y, z)) is defined as*

$$\Delta(x, y, z) = \mathit{Span}\Big\{X(x, y, z), Y(x, y, z)\Big\} \qquad \forall x \in \mathbb{R}^3,$$

where

$$X = \partial_x, \quad Y = \partial_y + \frac{x^2}{2}\partial_z.$$

The first Lie bracket of X, Y is given by

$$[X, Y] = x\partial_z.$$

For any $(x, y, z) \in \mathbb{R}^3$ with $x \neq 0$, the three vectors

$$X(x, y, z), Y(x, y, z), [X, Y](x, y, z)$$

are linearly independent. Hence, Δ is a totally nonholonomic distribution of degree 2 on $\mathbb{R}^3 \setminus \{x = 0\}$. The Lie bracket $[[X, Y], Y]$ is given by

$$[[X, Y], Y] = \partial_z.$$

Then, Δ is a totally nonholonomic distribution of degree 3 on $\mathbb{R}^3$.

Example 1.1.22. *More generally, if X, Y are given by*

$$X = \partial_x, \quad Y = \partial_y + x^l\partial_z,$$

with $l \in \mathbb{N}^*$, we check easily that the distribution spanned by X and Y is a totally nonholonomic distribution of degree $l + 1$.

Example 1.1.23. *Assume that M has dimension $n = 2p + 1$ and let α be a 1-form on M satisfying*

$$\alpha \wedge (d\alpha)^p \neq 0$$

then the distribution given by $\Delta = Ker(\alpha)$ is totally nonholonomic of degree 2. Such a 1-form is called a contact form and the associated distribution is called a contact distribution. As a matter of fact, given $\bar{x} \in M$, there is a local set of coordinates $(x_1, \ldots, x_n)$ in an open neighborhood $\bar{\mathcal{V}}$ of $\bar{x}$ such that α has the form

$$\alpha = \left(\sum_{i=1}^{2p} a_i dx_i\right) + dx_n,$$

where $a_1, \ldots, a_{2p}$ are smooth scalar function on $\bar{\mathcal{V}}$ such that

$$a_i(\bar{x}) = 0 \qquad \forall i = 1, \ldots, 2p.$$

Hence, the family of smooth vector fields $\bar{X}^1, \ldots, \bar{X}^{2p}$ given by

$$\bar{X}^i = \partial_{x_i} - a_i\partial_{x_n} \qquad \forall i = 1, \ldots, 2p,$$

defines a local frame for $\Delta = Ker(\alpha)$ in $\bar{\mathcal{V}}$. On the one hand, the $n = 2p + 1$-form $\alpha \wedge (d\alpha)^p$ at $\bar{x}$ reads

$$(\alpha \wedge (d\alpha)^p)_{\bar{x}} = \sum_{\sigma \in \mathcal{P}_{2p}} \left[\prod_{l=1,\ldots,p} \left(\frac{\partial a_{j_l}}{\partial x_{i_l}} - \frac{\partial a_{i_l}}{\partial x_{j_l}}\right)\right] dx_n \wedge (dx_{i_1} \wedge dx_{j_1}) \ldots \wedge \left(dx_{i_p} \wedge dx_{j_p}\right)|_{\bar{x}}, \tag{1.3}$$

where $\mathcal{P}_{2p}$ denotes the set of p-tuples of the form $\sigma = ((i_1, j_1), \ldots, (i_p, j_p))$ with $\{i_1, j_1, \ldots, i_p, j_p\} = \{1, \ldots, 2p\}$ and $i_l < j_l$ for all $l = 1, \ldots, p$. On the other hand, we check easily that

$$\left[\bar{X}^i, \bar{X}^j\right](\bar{x}) = \left(\partial_{x_i} a_j - \partial_{x_j} a_i\right) \partial_{x_n}(\bar{x}) \qquad \forall i, j = 1, \ldots, 2p.$$

Therefore, if there is $\bar{i} \in \{1, \ldots, 2p\}$ such that $[\bar{X}^{\bar{i}}, \bar{X}^j](\bar{x}) = 0$ for any $j \neq i$, then all the products appearing in (1.3) vanish, which implies that $(\alpha \wedge (d\alpha)^p)_{\bar{x}} = 0$, contradiction. We deduce that for every $i \in \{1, \ldots, n\}$, there holds

$$Span\Big\{\bar{X}^1(\bar{x}), \ldots, \bar{X}^{2p}(\bar{x}), \left[\bar{X}^i, \bar{X}^1\right](\bar{x}), \ldots, \left[\bar{X}^i, \bar{X}^{2p}\right](\bar{x})\Big\} = T_{\bar{x}}M. \qquad (1.4)$$

This means that $\Delta = Ker(\alpha)$ is totally nonholonomic of degree 2.

Example 1.1.24. *As an example, the 1-form given in Example 1.1.5 is a contact form on $\mathbb{S}^3$. There holds*

$$\begin{aligned}
\alpha \wedge d\alpha &= (x_1 dy_1 - y_1 dx_1 + x_2 dy_2 - y_2 dx_2) \wedge (2dx_1 \wedge dy_1 + 2dx_2 \wedge dy_2) \\
&= 2x_1 dy_1 \wedge dx_2 \wedge dy_2 - 2y_1 dx_1 \wedge dx_2 \wedge dy_2 \\
&\qquad\qquad +2x_2 dx_1 \wedge dy_1 \wedge dy_2 - 2y_2 dx_1 \wedge dy_1 \wedge dx_2.
\end{aligned}$$

A basis of the tangent space to $\mathbb{S}^3$ at $x = (x_1, y_1, x_2, y_2) \in \mathbb{S}^3$ is given by (V_1, V_2, V_3) with

$$\left\{\begin{array}{lcl}
V_1 &=& -y_1 e_1 + x_1 e_2 - y_2 e_3 + x_2 e_4 \\
V_2 &=& -x_2 e_1 + y_2 e_2 + x_1 e_3 - y_1 e_4 \\
V_3 &=& -y_2 e_1 - x_2 e_2 + y_1 e_3 + x_1 e_4.
\end{array}\right.$$

Then

$$\begin{aligned}
&(\alpha \wedge d\alpha)_x (V_1, V_2, V_3) = \\
&\qquad 2x_1^2 \left(x_1^2 + y_1^2 + x_2^2 + y_2^2\right) - 2y_1^2 \left(-x_1^2 - y_1^2 - x_2^2 - y_2^2\right) \\
&\qquad + 2x_2^2 \left(x_1^2 + y_1^2 + x_2^2 + y_2^2\right) - 2y_2^2 \left(-x_1^2 - y_1^2 - x_2^2 - y_2^2\right) \\
&\qquad\qquad = 2\left(x_1^2 + y_1^2 + x_2^2 + y_2^2\right)^2 = 2.
\end{aligned}$$

This means that the restriction of the 3-form $\alpha \wedge d\alpha$ to the tangents spaces to $\mathbb{S}^3$ does not vanish.

1.2 Horizontal paths and End-Point mappings

Horizontal paths

Let Δ be a distribution of rank $m \leq n$ in M. A continuous path $\gamma : [0, T] \to \mathbb{R}^n$ is said to be *horizontal* with respect to Δ if it is absolutely continuous with square integrable derivative (see Appendix A) and satisfies

$$\dot{\gamma}(t) \in \Delta\big(\gamma(t)\big) \quad \text{a.e. } t \in [0, T].$$

For every $x \in M$ and every $T > 0$, we denote by $\Omega_\Delta^{x,T}$ the set of horizontal paths $\gamma : [0, T] \to M$ starting at x. If Δ admits a global frame $X^1, \ldots, X^m$, then there is a one-to-one correspondence between $\Omega_\Delta^{x,T}$ and an open subset of $L^2([0, T]; \mathbb{R}^m)$.

Proposition 1.2.1. *Let $\mathcal{F} = \{X^1, \ldots, X^m\}$ be a global frame for Δ. Then for every $x \in M$ and every $T > 0$, there is an open subset $U_{\mathcal{F}}^{x,T}$ of $L^2([0,T];\mathbb{R}^m)$ such that the mapping*

$$u \in U_{\mathcal{F}}^{x,T} \longmapsto \gamma_u \in \Omega_{\Delta}^{x,T},$$

(where $\gamma_u : [0,T] \to M$ is the unique solution to the Cauchy problem

$$\dot{\gamma}_u(t) = \sum_{i=1}^{m} u_i(t) X^i(\gamma_u(t)) \quad a.e.\ t \in [0,T], \qquad \gamma_u(0) = x,) \tag{1.5}$$

is one-to-one.

Proof. The set of controls $u \in L^2([0,T];\mathbb{R}^m)$ such that the solution γ_u of (1.5) is well-defined on $[0,T]$ is a non-empty open set. Moreover, by construction, any path γ_u is absolutely continuous with square integrable derivative and almost everywhere tangent to Δ. This proves that the map under study is well-defined. Let $\gamma \in \Omega_{\Delta,x,T}$ be such that there are $u, v \in L^2([0,T];\mathbb{R}^m)$ such that

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

Since the tangent vectors $X^1(\gamma(t)), \ldots, X^m(\gamma(t))$ are always linearly independent in $T_{\gamma(t)}M$, we infer that $u(t) = v(t)$ for almost every $t \in [0,T]$, which proves that our map is injective. Furthermore, given $\gamma \in \Omega_{\Delta}^{x,T}$, for almost every $t \in [0,T]$, the path γ is differentiable at t and there is a unique $u(t) \in \mathbb{R}^m$ such that $\dot{\gamma}(t) = \sum_{i=1}^m u_i(t) X^i(\gamma(t))$. By construction, the function $u : [0,T] \to \mathbb{R}^m$ belongs to $L^2([0,T];\mathbb{R}^m)$. $\square$

As seen before, a general distribution may have no global frame, but it can be represented by $k = m(n+1)$ vector fields (see Proposition 1.1.8).

Proposition 1.2.2. *Let $\mathcal{F} = \{X^1, \cdots, X^k\}$ be a generating family for Δ on M. Then, for every $x \in M$ and every $T > 0$, there is an open subset $U_{\mathcal{F}}^{x,T}$ of $L^2([0,T];\mathbb{R}^k)$ such that the mapping*

$$u \in U_{\mathcal{F}}^{x,T} \longmapsto \gamma_u \in \Omega_{\Delta}^{x,T},$$

(where $\gamma_u : [0,T] \to M$ is the unique solution to the Cauchy problem

$$\dot{\gamma}_u(t) = \sum_{i=1}^{k} u_i(t) X^i(\gamma_u(t)) \quad a.e.\ t \in [0,T], \qquad \gamma_u(0) = x,) \tag{1.6}$$

is onto.

Proof. Let $\gamma \in \Omega_{\Delta}^{x,T}$ be fixed. For every $t \in [0,T]$, there is an open set $\mathcal{O}_t$ of $\gamma(t)$ in M and m integers $i_1^t, \ldots, i_m^t \in \{1, \ldots, k\}$ such that

$$\text{Span}\ \left\{X^{i_1^t}(x), \ldots, X^{i_m^t}(x)\right\} = \Delta(x) \qquad \forall x \in \mathcal{O}_t.$$

The curve $\gamma([0,T])$ is compact and is contained in $\cup_{t\in[0,T]}\mathcal{O}_t$. Hence, there are N times $t_1,\ldots,t_N \in [0,T]$ together with a partition of unity $\{\psi_j\}$ such that

$$[0,T] \subset \bigcup_{j=1}^{N} \mathcal{O}_{t_j}, \quad \text{Supp } (\psi_j) \subset \mathcal{O}_{t_j}, \quad \sum_{j=1}^{N} \psi_j = 1.$$

For every j, there is a smooth mapping $U_j : TM \to \mathbb{R}^m$ such that

$$v = \sum_{l=1}^{m} U_j(v) X^{i_l^{t_j}}(x),$$

for every $(x,v) \in TM$ with $x \in \mathcal{O}_{t_j}$ and $v \in \Delta(x)$. Then, there holds for almost every $t \in [0,T]$ and any $j \in \{1,\ldots,N\}$,

$$\gamma(t) \in \mathcal{O}_{t_j} \quad \Longrightarrow \quad \dot{\gamma}(t) = \sum_{l=1}^{m} U_j\left(\dot{\gamma}(t)\right) X^{i_l^{t_j}}\big(\gamma(t)\big).$$

By the properties satisfied by $\{\psi_j\}$, we infer that

$$\begin{aligned}
\dot{\gamma}(t) &= \sum_{j=1}^{N} \psi_j\big(\gamma(t)\big) \left[\sum_{l=1}^{m} U_j\left(\dot{\gamma}(t)\right) X^{i_l^{t_j}}\big(\gamma(t)\big)\right] \\
&= \sum_{j=1}^{N} \sum_{l=1}^{m} \Big(\psi_j\big(\gamma(t)\big) U_j\left(\dot{\gamma}(t)\right)\Big) X^{i_l^{t_j}}\big(\gamma(t)\big),
\end{aligned}$$

for almost every $t \in [0,T]$. Each mapping $t \mapsto \psi_j\big(\gamma(t)\big)U_j\left(\dot{\gamma}(t)\right)$ belongs to $L^2\big([0,T];\mathbb{R}\big)$. We infer easily the existence of $u \in L^2\big([0,T];\mathbb{R}^k\big)$ such that $\gamma = \gamma_u$. □

Remark 1.2.3. *If M is compact, then solutions to (1.5) (resp. (1.6)) are defined for any $u \in L^2([0,T];\mathbb{R}^m)$ (resp. $u \in L^2([0,T];\mathbb{R}^k)$).*

Given a family of smooth vector fields $\mathcal{F} = \left\{X^1,\cdots,X^k\right\}$ on M and $x \in M, T > 0$, a function $u \in U_{\mathcal{F}}^{x,T} \subset L^2\big([0,T];\mathbb{R}^k\big)$ is called a *control* and the corresponding solution of (1.6) is called the *trajectory* starting at x and associated with the control u. Since any horizontal path can be viewed as a trajectory associated to a control system like (1.6), we restrict in the next paragraph our attention to End-Point mappings associated with finite families of smooth vector fields.

End-Point mappings

Let $\mathcal{F} = \left\{X^1,\ldots,X^k\right\}$ be a family of $k \geq 1$ smooth vector fields on M. As before, given x and $T > 0$, there is a maximal open subset $U_{\mathcal{F}}^{x,T} \subset L^2\left([0,T];\mathbb{R}^k\right)$ such that for every $u \in U_{\mathcal{F}}^{x,T}$, there is a unique solution to the Cauchy problem

$$\dot{\gamma}_u(t) = \sum_{i=1}^{k} u_i(t) X^i\left(\gamma_u(t)\right) \quad \text{a.e. } t \in [0,T], \qquad \gamma_u(0) = x. \tag{1.7}$$

The End-Point mapping associated to $\mathcal{F}$ at x in time $T > 0$ is defined as follows,

$$\begin{array}{rcl} E_{\mathcal{F}}^{x,T} \, : \, U_{\mathcal{F}}^{x,T} & \longrightarrow & M \\ u & \longmapsto & \gamma_u(T). \end{array}$$

Given $u \in U_{\mathcal{F}}^{x,T}$, we denote by $X_{\mathcal{F}}^u$ the time-dependent vector field defined by

$$X_{\mathcal{F}}^u(t,x) := \sum_{i=1}^m u_i(t) X^i(x) \qquad \text{a.e. } t \in [0,T], \, \forall x \in M.$$

Its flow $\Phi_{\mathcal{F}}^u(t,x)$ is well-defined and smooth on a neighbourhood of x; we denote by $D_x \Phi_{\mathcal{F}}^u(t,x)$ its differential at (t,x) with respect to the x variable. The following result holds. (We refer the reader to Appendix A for reminders in differential equations and to Appendix B for reminders in differential calculus in infinite dimension.)

Proposition 1.2.4. *The End-Point mapping $E_{\mathcal{F}}^{x,T}$ is of class C^1 on $U_{\mathcal{F}}^{x,T}$ and for every control $u \in U_{\mathcal{F}}^{x,T}$, its differentiable at u,*

$$D_u E_{\mathcal{F}}^{x,T} \, : \, L^2([0,T];\mathbb{R}^k) \longrightarrow T_{E_{\mathcal{F}}^{x,T}(u)} M$$

is given by

$$D_u E_{\mathcal{F}}^{x,T}(v) = D_x \Phi_{\mathcal{F}}^u(T,x) \cdot \int_0^T \left(D_x \Phi_{\mathcal{F}}^u(t,x) \right)^{-1} \cdot X_{\mathcal{F}}^v \left(t, E_{\mathcal{F}}^{x,t}(u) \right) dt \tag{1.8}$$

for every $v \in L^2([0,T];\mathbb{R}^k)$. Moreover, the mapping

$$u \in U_{\mathcal{F}}^{x,T} \longmapsto D_u E_{\mathcal{F}}^{x,T} \tag{1.9}$$

is locally Lipschitz.

Proof. Any smooth manifold can be smoothly embedded in an Euclidean space. Then without loss of generality we can assume that M is a smooth submanifold of some $\mathbb{R}^N$ and consequently that the X^i's are the restrictions of smooth vector fields $\tilde{X}^1, \ldots, \tilde{X}^k$ which are defined in an open neighborhood of M in $\mathbb{R}^N$. Given $u \in U_{\mathcal{F}}^{x,T}$ and $v \in L^2([0,T];\mathbb{R}^k)$ let us look at

$$\lim_{\epsilon \to 0} \frac{1}{\epsilon} \Big(E_{\mathcal{F}}^{x,T}(u + \epsilon v) - E_{\mathcal{F}}^{x,T}(u) \Big).$$

Using the previous notations, we have

$$\begin{aligned} \gamma_{u+\epsilon v}(T) &= \int_0^T \sum_{i=1}^k \left(u_i(t) + \epsilon v_i(t) \right) X^i\big(\gamma_{u+\epsilon v}(t)\big) \, dt \\ &= \int_0^T \sum_{i=1}^k \left(u_i(t) + \epsilon v_i(t) \right) \tilde{X}^i\big(\gamma_{u+\epsilon v}(t)\big) \, dt, \end{aligned} \tag{1.10}$$

with $\gamma_{u+\epsilon v}(0) = x$. For every $i = 1, \ldots, k$ and every $t \in [0,T]$, the Taylor expansion of each $\tilde{X}^i$ at $\gamma_u(t)$ gives

$$\begin{aligned} \tilde{X}^i\big(\gamma_{u+\epsilon v}(t)\big) = \tilde{X}^i\big(\gamma_u(t)\big) + D_{\gamma_u(t)} \tilde{X}^i \cdot \big(\gamma_{u+\epsilon v}(t) - \gamma_u(t)\big) \\ + \left| \gamma_{u+\epsilon v}(t) - \gamma_u(t) \right| o(1). \end{aligned} \tag{1.11}$$

Setting $\delta_x(t) := \gamma_{u+\epsilon v}(t) - \gamma_u(t)$ for any t, we may assume that δ_x has size ϵ, then (1.10) yields formally

$$\delta_x(T) \;=\; \int_0^T \sum_{i=1}^k u_i(t) D_{\gamma_u(t)} \tilde{X}^i \cdot \delta_x(t)\, dt \,+\, \epsilon \sum_{i=1}^m v_i(t) \tilde{X}^i\big(\gamma_u(t)\big) \,+\, o(\epsilon).$$

This suggests that the function $t \in [0,T] \mapsto \delta_x(t)$ should be solution to the Cauchy problem

$$\dot{\delta}_x(t) = \left[\sum_{i=1}^k u_i(t) D_{\gamma_u(t)} \tilde{X}^i\right] \delta_x(t) + \left[\sum_{i=1}^k v_i(t) \tilde{X}^i\big(\gamma_u(t)\big)\right] \qquad \text{a.e. } t \in [0,T], \tag{1.12}$$

with $\delta_x(0) = 0$. By (1.10)-(1.12) together with Gronwall's Lemma (see Appendix A) we check easily that for every $v \in L^2\big([0,T];\mathbb{R}^k\big)$, the quantity

$$\frac{1}{\epsilon}\Big(E_{\mathcal{F}}^{x,T}\big(u+\epsilon v\big) - E_{\mathcal{F}}^{x,T}\big(u\big) - \epsilon \delta_x(T)\Big)$$

tends to zero as ϵ tends to zero. For almost every $t \in [0,T]$, denote by $A_u(t)$ the matrix in $M_N(\mathbb{R})$ representing the linear operator $\sum_{i=1}^k u_i(t) D_{\gamma_u(t)} \tilde{X}^i$ in the canonical basis of $\mathbb{R}^N$ and for every $t \in [0,T]$, denote by $B_u(t)$ the matrix in $M_{N,k}(\mathbb{R})$ whose the columns are the $\tilde{X}^i(\gamma_u(t))$'s. Denote by $S_u : [0,T] \to M_N(\mathbb{R})$ the solution to the Cauchy problem

$$\dot{S}_u(t) = A_u(t) S_u(t) \quad \text{a.e. } t \in [0,T], \quad S_u(0) = I_n.$$

Note that $S_u(t)$ is exactly the Jacobian of the flow $\Phi_{\tilde{\mathcal{F}}}^u$ (with $\tilde{\mathcal{F}} = \{\tilde{X}^1, \ldots, \tilde{X}^k\}$) at $(t, \gamma_u(t))$ with respect to the x variable. The solution of (1.12) at time T is given by (see Appendix A)

$$\delta_x(T) = D_u E_{\mathcal{F}}^{x,T}(v) = S_u(T) \int_0^T S_u(t)^{-1} B_u(t) v(t) dt.$$

Thus we check that (1.8) is satisfied. Let us now prove the local Lipschitzness of $u \mapsto D_u E_{\mathcal{F}}^{x,T}$ and indeed give more details on the estimates that were needed in the above proof. Let $\bar{u}$ a control be fixed in $U_{\mathcal{F}}^{x,T} \subset L^2([0,T];\mathbb{R}^k)$. The curve $\gamma_{\bar{u}}\left([0,T]\right) \subset M \subset \mathbb{R}^N$ is compact. Let $\epsilon > 0$ be fixed, the set $\mathcal{V} \subset \mathbb{R}^N$ defined by

$$\mathcal{V} := \Big\{ \gamma_{\bar{u}}(t) + z \,|\, t \in [0,T],\, z \in B(0,\epsilon) \Big\}$$

is relatively compact. Then there is $K > 0$ such that all the $\tilde{X}^i$'s are bounded by K on $\mathcal{V}$ and all the $\tilde{X}^i$'s are K-Lipschitz on $\mathcal{V}$. Set

$$\delta := \frac{\epsilon}{KTe^{KT\|\bar{u}\|_{L^2}}}$$

and pick a control $u \in L^2([0,T];\mathbb{R}^k)$ with $\|u - \bar{u}\|_{L^2} < \delta$. We claim that u belongs to $U_{\mathcal{F}}^{x,T}$ and that the trajectory $\gamma_u : [0,T] \to M \subset \mathbb{R}^N$ (which is associated with u) is contained in $\mathcal{V}$. Argue by contradiction and assume that there is $\bar{t} \in [0,T]$ such that $\gamma_u(t)$ is on the boundary of $\mathcal{V}$. Taking $\bar{t} > 0$ smaller if necessary, we may assume that $\gamma_u(t)$ belongs to $\mathcal{V}$ for any $t \in [0,\bar{t})$. Set

$$f(t) := |\gamma_u(t) - \gamma_{\bar{u}}(t)| \qquad \forall t \in [0,\bar{t}].$$

Then we have for every $t \in [0,\bar{t})$,

$$\begin{aligned} f(t) &= \left| \int_0^t \sum_{i=1}^k u_i(s)\tilde{X}^i\big(\gamma_u(s)\big) - \sum_{i=1}^k \bar{u}_i(s)\tilde{X}^i\big(\gamma_{\bar{u}}(s)\big)\, ds \right| \\ &\leq \int_0^t \left| \sum_{i=1}^k \left(u_i(s) - \bar{u}_i(s)\right) \tilde{X}^i\big(\gamma_u(s)\big) \right| ds \\ &\qquad + \int_0^t \left| \sum_{i=1}^k \bar{u}_i(s) \left(\tilde{X}^i\big(\gamma_u(s)\big) - \tilde{X}^i\big(\gamma_{\bar{u}}(s)\big) \right) \right| ds \\ &\leq Kt\big\|u - \bar{u}\big\|_{L^2} + \int_0^t \left| \sum_{i=1}^k \bar{u}_i(s) \right| Kf(s)\, ds. \end{aligned}$$

By Gronwall's Lemma (see Appendix A) and definition of δ, we infer that

$$f\big(\bar{t}\big) \leq KT\big\|u - \bar{u}\big\|_{L^2} e^{KT\|\bar{u}\|_{L^2}} < \epsilon.$$

Thus we get a contradiction and the claimed is proved. Let $u, u' \in L^2([0,T];\mathbb{R}^k)$ with $\|u - \bar{u}\|_{L^2}, \|u' - \bar{u}\|_{L^2} < \delta$, by repeating the same argument we get

$$|\gamma_{u'}(t) - \gamma_u(t)| \leq Kt\big\|u' - u\big\|_{L^2} e^{KT\left(\|\bar{u}\|_{L^2} + \delta\right)} \qquad \forall t \in [0,T].$$

(This shows that End-Point mappings are locally Lipschitz.) Denote by $S_u, S_{u'} : [0,T] \to M_N(\mathbb{R})$ the solutions to the Cauchy problems

$$\begin{aligned} \dot{S}_u(t) = A_u(t)S_u(t) \quad \text{a.e. } t \in [0,T], \quad S_u(0) = I_n, \\ \dot{S}_{u'}(t) = A_{u'}(t)S_{u'}(t) \quad \text{a.e. } t \in [0,T], \quad S_{u'}(0) = I_n, \end{aligned}$$

where $A_u, A_{u'}$ are defined by

$$A_u(t) := \sum_{i=1}^k u_i(t) J_{\tilde{X}^i}\big(\gamma_u(t)\big), \quad A_u(t) := \sum_{i=1}^k u'_i(t) J_{\tilde{X}^i}\big(\gamma_{u'}(t)\big),$$

for almost every $t \in [0,T]$ ($J_{\tilde{X}^i}(\gamma_u(t))$ (resp. $J_{\tilde{X}^i}(\gamma_{u'}(t))$ denotes the Jacobian matrix of $\tilde{X}^i$ at $\gamma_u(t)$ (resp. at $\gamma_{u'}(t)$)). Taking $K > 0$ larger if necessary, we may assume that it is an upper bound for the $J_{\tilde{X}^i}$'s on $\mathcal{V}$ and a Lipschitz constant for the $J_{\tilde{X}^i}$'s on $\mathcal{V}$. Then we have for every $t \in [0,T]$,

$$\begin{aligned} \|S_u(t)\| &= \left\| I_n + \int_0^t A_u(s)S_u(s)\, ds \right\| \\ &\leq 1 + \int_0^t \sum_{i=1}^k |u_i(s)| \left\| J_{\tilde{X}^i}\big(\gamma_u(t)\big) \right\| \|S_u(s)\|\, ds \\ &\leq 1 + \int_0^t K \sum_{i=1}^k |u_i(s)|\, \|S_u(s)\|\, ds. \end{aligned}$$

By Gronwall's Lemma (see Appendix A), we infer that

$$\|S_u(t)\| \leq e^{KT\left(\|\bar{u}\|_{L^2}+\delta\right)} \qquad \forall t \in [0,T].$$

Set

$$g(t) := \|S_{u'}(t) - S_u(t)\| \qquad \forall t \in [0,T].$$

Then all in all, we have for every $t \in [0,T]$,

$$\begin{aligned}
g(t) &= \left\| \int_0^t A_{u'}(s)S_{u'}(s) - A_u(s)S_u(s)\, ds \right\| \\
&\leq \int_0^t \left\| \big(A_{u'}(s) - A_u(s)\big) S_{u'}(s) \right\| \, ds \\
&\qquad + \int_0^t \left\| A_u(s)\big(S_{u'}(s) - S_u(s)\big) \right\| \, ds \\
&\leq \int_0^t \sum_{i=1}^k |u_i'(s) - u_i(s)| \left\| J_{\tilde{X}^i}\big(\gamma_{u'}(t)\big) \right\| \|S_{u'}(s)\| \, ds \\
&\qquad + \int_0^t \sum_{i=1}^k |u_i(s)| \left\| J_{\tilde{X}^i}\big(\gamma_{u'}(t)\big) - J_{\tilde{X}^i}\big(\gamma_u(t)\big) \right\| \|S_{u'}(s)\| \, ds \\
&\qquad + \int_0^t \sum_{i=1}^k |u_i(s)| \left\| J_{\tilde{X}^i}\big(\gamma_u(t)\big) \right\| \|S_{u'}(s) - S_u(s)\| \, ds \\
&\leq \int_0^t K e^{KT\left(\|\bar{u}\|_{L^2}+\delta\right)} \sum_{i=1}^k |u_i'(s) - u_i(s)| \, ds \\
&\qquad + \int_0^t K^2 T e^{2KT\left(\|\bar{u}\|_{L^2}+\delta\right)} \|u' - u\|_{L^2} \sum_{i=1}^k |u_i(s)| \, ds \\
&\qquad + \int_0^t K \sum_{i=1}^k |u_i(s)| \, \|S_{u'}(s) - S_u(s)\| \, ds \\
&\leq KT\left[1 + KT\left(\left\|\bar{u}\right\|_{L^2} + \delta\right)\right] e^{2KT\left(\|\bar{u}\|_{L^2}+\delta\right)} \|u' - u\|_{L^2} \\
&\qquad + \int_0^t K \sum_{i=1}^k |u_i(s)| \, g(s)\, ds,
\end{aligned}$$

which, by Gronwall's Lemma, yields

$$\|S_{u'}(t) - S_u(t)\| \leq C \left\|u' - u\right\|_{L^2} \qquad \forall t \in [0,T],$$

for some constant $C > 0$. The functions $t \in [0,T] \mapsto S_u(t)^{-1}, S_{u'}(t)^{-1}$ are respectively solutions to the Cauchy problems

$$\begin{aligned}
&\tfrac{d}{dt}\left(S_u(t)^{-1}\right) = -S_u(t)^{-1} A_u(t) \quad \text{a.e. } t \in [0,T], \quad S_u(0)^{-1} = I_n, \\
&\tfrac{d}{dt}\left(S_{u'}(t)^{-1}\right) = -S_{u'}(t)^{-1} A_{u'}(t) \quad \text{a.e. } t \in [0,T], \quad S_{u'}(0)^{-1} = I_n.
\end{aligned}$$

Then with the same arguments as before, we may assume that the constant $C > 0$ is such that

$$\|S_u(t)\|, \left\|S_u(t)^{-1}\right\|, \|S_{u'}(t)\|, \left\|S_{u'}(t)^{-1}\right\| \leq C \qquad \forall t \in [0,T]$$

and

$$\|S_{u'}(t) - S_u(t)\|, \left\|S_{u'}(t)^{-1} - S_u(t)^{-1}\right\| \leq C\left\|u' - u\right\|_{L^2} \qquad \forall t \in [0,T].$$

Then we have

$$\left\|S_{u'}(T)S_{u'}(t)^{-1} - S_u(T)S_u(t)^{-1}\right\| \leq 2C^2\left\|u' - u\right\|_{L^2} \qquad \forall t \in [0,T].$$

Fix now $v \in L^2([0,T];\mathbb{R}^k)$, then we have

$$\begin{aligned} D_{u'}E_{\mathcal{F}}^{x,T}(v) - D_u E_{\mathcal{F}}^{x,T}(v) \\ = \int_0^T \Big[S_{u'}(T)S_{u'}(t)^{-1}B_{u'}(t) - S_u(T)S_u(t)^{-1}B_u(t)\Big]v(t)\,dt. \end{aligned}$$

And in turn,

$$\begin{aligned} \left|D_{u'}E_{\mathcal{F}}^{x,T}(v) - D_u E_{\mathcal{F}}^{x,T}(v)\right| \leq \\ \left|\int_0^T S_{u'}(T)S_{u'}(t)^{-1}\left(B_{u'}(t) - B_u(t)\right)v(t)\,dt\right| \\ + \left|\int_0^T \left(S_{u'}(T)S_{u'}(t)^{-1} - S_u(T)S_u(t)^{-1}\right)B_u(t)v(t)\,dt\right|. \end{aligned}$$

By the above estimates, we obtain a constant $D > 0$ such that

$$\left|D_{u'}E_{\mathcal{F}}^{x,T}(v) - D_u E_{\mathcal{F}}^{x,T}(v)\right| \leq D\left\|u' - u\right\|_{L^2}\|v\|_{L^2},$$

which shows that the mapping given by (1.9) is locally Lipschitz on $U_{\mathcal{F}}^{x,T}$. □

Remark 1.2.5. *If $M = \mathbb{R}^n$, the derivative of $E_{\mathcal{F}}^{x,T}$ at u is given by*

$$D_u E_{\mathcal{F}}^{x,T}(v) = S(T)\int_0^T S(t)^{-1}B(t)v(t)dt,$$

where $S : [0,T] \to M_n(\mathbb{R})$ is the solution to the Cauchy problem

$$\dot{S}(t) = A(t)S(t) \quad a.e.\ t \in [0,T], \quad S(0) = I_n.$$

and where the matrices $A(t) \in M_n(\mathbb{R}), B(t) \in M_{n,k}(\mathbb{R})$ are defined by

$$A(t) := \sum_{i=1}^k u_i(t)J_{X^i}\big(\gamma_u(t)\big) \qquad a.e.\ t \in [0,T]$$

($\gamma_u(t) = E_{\mathcal{F}}^{x,t}(u)$ and J_{X^i} denotes the Jacobian matrix of X^i at $\gamma_u(t)$) and

$$B(t) := \Big(X^1(\gamma_u(t)), \cdots, X^k(\gamma_u(t))\Big) \qquad a.e.\ t \in [0,T].$$

Properties of End-Point mappings

Given $u \in U_{\mathcal{F}}^{x,T}$, we set

$$\mathrm{Im}_{\mathcal{F}}^{x,T}(u) := D_u E_{\mathcal{F}}^{x,T}\big(L^2([0,T];\mathbb{R}^k)\big).$$

Defining $y = E_{\mathcal{F}}^{x,T}(u)$, we observe that $\mathrm{Im}_{\mathcal{F}}^{x,T}(u)$ is a vector space contained in T_yM, hence of dimension $\leq n$. We call *rank* of $u \in U_{\mathcal{F}}^{x,T}$ with respect to $E_{\mathcal{F}}^{x,T}$, denoted by $\mathrm{rank}_{\mathcal{F}}^{x,T}(u)$, the dimension of $\mathrm{Im}_{\mathcal{F}}^{x,T}(u)$.

For any $u \in L^2([0,T];\mathbb{R}^k)$ and $\lambda > 0$, we denote by u_λ the control in $L^2\big([0;\lambda^{-1}T],\mathbb{R}^k\big)$ defined by

$$u_\lambda(t) := \lambda u(\lambda t) \qquad \text{a.e. } t \in [0,\lambda^{-1}T].$$

Proposition 1.2.6. *For every $u \in U_{\mathcal{F}}^{x,T}$ and every $\lambda > 0$, u_λ belongs to $U_{\mathcal{F}}^{x,\lambda^{-1}T}$ and*

$$D_{u_\lambda} E_{\mathcal{F}}^{x,\lambda^{-1}T}(v_\lambda) = D_u E_{\mathcal{F}}^{x,T}(v) \qquad \forall v \in L^2\big([0,T];\mathbb{R}^k\big).$$

In particular, $\mathrm{rank}_{\mathcal{F}}^{x,T}(u) = \mathrm{rank}_{\mathcal{F}}^{x,\lambda^{-1}T}(u_\lambda)$.

Proof. We just notice that if $\gamma_u : [0,T] \to M$ is a solution of (1.7), then the path $\gamma_{u,\lambda} : [0,\lambda^{-1}T] \to M$ defined by

$$\gamma_{u,\lambda}(t) = \gamma_u(\lambda t) \qquad \forall t \in \left[0,\lambda^{-1}T\right],$$

satisfies for a.e. $t \in [0,\lambda^{-1}T]$,

$$\begin{aligned}\dot\gamma_{u,\lambda}(t) = \lambda\dot\gamma_u(\lambda t) &= \sum_{i=1}^k \lambda u_i(\lambda t) X^i\left(\gamma_u(\lambda t)\right)\\ &= \sum_{i=1}^k u_\lambda(t) X^i\left(\gamma_{u,\lambda}(t)\right).\end{aligned}$$

The remaining part of the result follows easily. □

For every $u \in L^2([0,T];\mathbb{R}^k)$, we denote by $\breve{u}$ the control in $L^2\big([0,T];\mathbb{R}^k\big)$ defined by

$$\breve{u}(t) := -u(T-t) \qquad \text{a.e. } t \in [0,T].$$

Proposition 1.2.7. *For every $u \in U_{\mathcal{F}}^{x,T}$, $\breve{u}$ belongs to $U_{\mathcal{F}}^{y,T}$ with $y := E_{\mathcal{F}}^{x,T}(u)$ and*

$$\left(D_x\Phi_{\mathcal{F}}\right)^u(T,x)^{-1} \cdot D_u E_{\mathcal{F}}^{x,T}(v) + D_{\breve{u}} E_{\mathcal{F}}^{y,T}(\breve{v}) = 0 \qquad \forall v \in L^2\big([0,T];\mathbb{R}^k\big).$$

In particular, $\mathrm{rank}_{\mathcal{F}}^{x,T}(u) = \mathrm{rank}_{\mathcal{F}}^{y,T}(\breve{u})$.

Proof. First, we note that

$$E_{\mathcal{F}}^{E_{\mathcal{F}}^{x,T}(u),T}(\breve{u}) = x \qquad \forall u \in U_{\mathcal{F}}^{x,T}.$$

The mapping $(z,v) \mapsto E_{\mathcal{F}}^{z,T}(v)$ is smooth and its derivative with respect to the z variable at $\left(y = E_{\mathcal{F}}^{x,T}(u), \breve{u}\right)$ is given by $D_x\Phi_{\mathcal{F}}^u(T,x)^{-1}$. Derivating the above equality at u yields the result. □

For any $u \in L^2([0,T];\mathbb{R}^k)$ and $u' \in L^2[0,T'];\mathbb{R}^k)$, we denote by $u * u'$ the *concatenation* of u and u', that is the control in $L^2[0,T+T'];\mathbb{R}^k)$ defined by

$$u * u'(t) = \begin{cases} u(t) & \text{if} \quad 0 \le t \le T; \\ u'(t-T) & \text{if} \quad T < t \le T+T' \end{cases}$$

for a.e. $t \in [0, T+T']$.

Proposition 1.2.8. *For every $u \in U_{\mathcal{F}}^{x,T}$ and $u' \in U_{\mathcal{F}}^{y,T'}$ with $y = E_{\mathcal{F}}^{x,T}(u)$, there holds $u * u' \in U_{\mathcal{F}}^{x,T+T'}$ and*

$$D_{u*u'} E_{\mathcal{F}}^{x,T+T'}\big(v * v'\big) \;=\; D_x \Phi_{\mathcal{F}}^{u'}(T', y) \cdot dE_{\mathcal{F}}^{x,T}(v) + D_{u'} E_{\mathcal{F}}^{y,T'}(v'), \quad (1.13)$$

for any $v \in L^2([0,T];\mathbb{R}^k)$ and $v' \in L^2([0,T'];\mathbb{R}^k)$. In particular,

$$D_{u*u'} E_{\mathcal{F}}^{x,T+T'}\big(v * 0\big) = D_x \Phi_{\mathcal{F}}^{u'}(T', y) \cdot D_u E_{\mathcal{F}}^{x,T}(v) \qquad \forall v \in L^2([0,T];\mathbb{R}^k), \quad (1.14)$$

$$D_{u*u'} E_{\mathcal{F}}^{x,T+T'}\big(0 * v'\big) = D_{u'} E_{y,T'}(v') \qquad \forall v' \in L^2([0,T];\mathbb{R}^k), \quad (1.15)$$

and

$$rank_{\mathcal{F}}^{x,T+T'}\big(u * u'\big) \ge \max\Big\{ rank_{\mathcal{F}}^{x,T}(u), rank_{\mathcal{F}}^{y,T'}(u') \Big\}. \quad (1.16)$$

Proof. We note that

$$E_{\mathcal{F}}^{x,T+T'}(u * u') = E_{\mathcal{F}}^{E_{\mathcal{F}}^{x,T}(u),T'}(u'),$$

for any $u \in U_{\mathcal{F}}^{x,T}$ and $u' \in U_{\mathcal{F}}^{y,T'}$ with $y = E_{\mathcal{F}}^{x,T}(u)$. The mapping $(z,v) \mapsto E_{\mathcal{F}}^{z,T'}(v)$ is smooth and its derivative with respect to the z variable at (y,u') is given by $D_x \Phi^{u'}(T,x)$. Derivating the above equality at u yields the result. □

Example 1.2.9. *Let $\mathcal{F} = \{X^1, X^2\}$ be the family of smooth vectors fields on $\mathbb{R}^4$ (with coordinates $x = (x_1,x_2,x_3,x_4)$ and canonical basis (e_1,e_2,e_3,e_4)) defined by*

$$X^1 = \partial_{x_1} \quad \textit{and} \quad X^2 = \partial_{x_2} + x_1^2 \partial_{x_3} + x_1 x_2 \partial_{x_4}.$$

Set $x = (-1,0,0,0)$, $y = (0,0,0,0)$, and define the controls $u, u' \in L^2\big([0,1];\mathbb{R}^2\big)$ by

$$u(t) = (1,0) \quad \textit{and} \quad u'(t) = (0,1) \qquad \forall t \in [0,1].$$

The control u has rank 3 with respect to $E_{\mathcal{F}}^{x,1}$. As a matter of fact, the trajectory $\gamma_u : [0,1] \to \mathbb{R}^4$ starting at x and associated with u equals

$$\gamma_u(t) = (-1+t, 0,0,0) \qquad \forall t \in [0,1],$$

and using the representation formula given in Remark 1.2.5, we have

$$D_u E_{\mathcal{F}}^{x,1}(v) = \int_0^T B(t) v(t) dt \qquad \forall v \in L^2\big([0,1];\mathbb{R}^2\big),$$

where $B(t) = \big(X^1(\gamma_u(t)), X^2(\gamma_u(t))\big)$ *for any* $t \in [0,1]$. *Then,*

$$\begin{aligned} Im_{\mathcal{F}}^{x,1}(u) &= Span\left\{\int_0^1 v_1(t)dt\, e_1 \mid v_1 \in L^2([0,1];\mathbb{R})\right\} \\ &+ Span\left\{\int_0^1 v_2(t)dt\, e_2 + \int_0^1 (1-t)^2 v_2(t)dt\, e_3 \mid v_2 \in L^2([0,1];\mathbb{R})\right\} \\ &= Span\{e_1, e_2, e_3\}. \end{aligned}$$

The trajectory $\gamma_{u'} : [0,1] \to \mathbb{R}^4$ *starting at* y *and associated with* u' *equals*

$$\gamma_{u'}(t) = (0,t,0,0) \qquad \forall t \in [0,1],$$

and there holds

$$D_{u'} E_{\mathcal{F}}^{y,1}(v) = S(T) \int_0^T S(t)^{-1} B(t) v(t) dt,$$

where

$$S(t) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \frac{t^2}{2} & 0 & 0 & 1 \end{pmatrix} \quad \text{and} \quad B(t) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{pmatrix} \qquad \forall t \in [0,1].$$

We infer that

$$\begin{aligned} Im_{\mathcal{F}}^{y,1}(u') &= Span\left\{\int_0^1 v_2(t)dt\, e_2 \mid v_2 \in L^2([0,1];\mathbb{R})\right\} \\ &+ Span\left\{\int_0^1 v_1(t)dt\, e_1 + \int_0^1 \left(1 - \frac{t^2}{2}\right) v_1(t)dt\, e_4 \mid v_1 \in L^2([0,1];\mathbb{R})\right\} \\ &= Span\{e_1, e_2, e_4\}. \end{aligned}$$

Finally, we note that

$$D_x \Phi_{\mathcal{F}}^{u'}(1,y)(e_3) = e_3.$$

Therefore, by (1.13)-(1.14), this implies

$$Im_{\mathcal{F}}^{x,2}(u * u') = \mathbb{R}^4,$$

which means that

$$4 = rank_{\mathcal{F}}^{x,2}\big(u * u'\big) > \max\left\{rank_{\mathcal{F}}^{x,1}(u), rank_{\mathcal{F}}^{y,1}(u')\right\} = 3.$$

The following proposition implies that the rank of a control is always larger or equal than the dimension of the family $\{X^1, \ldots, X^k\}$ at the end-point.

Proposition 1.2.10. *We have for every* $u \in U_{\mathcal{F}}^{x,T}$,

$$X^i\left(E_{\mathcal{F}}^{x,T}(u)\right) \in Im_{\mathcal{F}}^{x,T}(u) \qquad \forall i = 1, \cdots, k.$$

Proof. Let us first assume that we work in $\mathbb{R}^n$. In this case (see Remark 1.2.5), the derivative of $E_{\mathcal{F}}^{x,T}$ at u is given by

$$D_u E_{\mathcal{F}}^{x,T}(v) = S(T) \int_0^T S(t)^{-1} B(t) v(t) dt \qquad \forall v \in L^2\big([0,T];\mathbb{R}^k\big), \tag{1.17}$$

where $S(\cdot)$ is the solution to the Cauchy problem

$$\dot{S}(t) = A(t)S(t), \quad \text{a.e. } t \in [0,T], \quad S(0) = I_n.$$

and where the matrices $A(t) \in M_n(\mathbb{R}), B(t) \in M_{n,k}(\mathbb{R})$ are defined by

$$A(t) := \sum_{i=1}^k u_i(t) J_{X^i}\big(\gamma_u(t)\big) \qquad \text{a.e. } t \in [0,T]$$

and

$$B(t) := \big(X^1(\gamma_u(t)), \cdots, X^k(\gamma_u(t))\big) \qquad \text{a.e. } t \in [0,T].$$

Fix $i \in \{1, \cdots, k\}$ and denote by e_i the i-th vector of the canonical basis in $\mathbb{R}^k$. Define, for every $\epsilon \in (0,T)$, the control $v_\epsilon \in L^2\big([0,T];\mathbb{R}^k\big)$ by

$$v_\epsilon(t) = \begin{cases} 0 & \text{if } \ 0 \leq t \leq T-\epsilon; \\ (1/\epsilon)e_i & \text{if } \ T-\epsilon < t \leq T. \end{cases}$$

We have

$$D_u E_{\mathcal{F}}^{x,T}(v_\epsilon) = S(T) \int_{T-\epsilon}^T S(t)^{-1} \Big((1/\epsilon) X^i(\gamma_u(t)\Big)\, dt.$$

Hence

$$\begin{aligned}
&\Big|D_u E_{\mathcal{F}}^{x,T}(v_\epsilon) - X^i(\gamma_u(T))\Big| \\
&= \left|(1/\epsilon)S(T) \int_{T-\epsilon}^T S(t)^{-1} X^i(\gamma_u(t)\, dt - (1/\epsilon)S(T) \int_{T-\epsilon}^T S(T)^{-1} X^i(\gamma_u(T))\, dt\right| \\
&\qquad \leq (1/\epsilon)\, |S(T)| \int_{T-\epsilon}^T \big|S(t)^{-1} X^i(\gamma_u(t)) - S(T)^{-1} X^i(\gamma_u(T))\big|\, dt \\
&\qquad \leq (1/\epsilon)\, |S(T)| \int_{T-\epsilon}^T \big\|S(t)^{-1}\big\| \, \big|X^i(\gamma_u(t)) - X^i(\gamma_u(T))\big|\, dt \\
&\qquad\qquad + (1/\epsilon)\, |S(T)| \int_{T-\epsilon}^T \big\|S(t)^{-1} - S(T)^{-1}\big\| \, \big|X^i(\gamma_u(T))\big|\, dt.
\end{aligned}$$

Both mappings $t \mapsto X^i\left(x_u(t)\right)$ and $t \mapsto S(t)^{-1}$ are continuous at $t = T$. Therefore, there holds

$$\lim_{\epsilon \downarrow 0} D_u E_{\mathcal{F}}^{x,T}(v_\epsilon) = X^i(\gamma_u(T)).$$

Since $\operatorname{Im}_{\mathcal{F}}^{x,T}(u) = D_u E_{\mathcal{F}}^{x,T}\left(L^2([0,T];\mathbb{R}^k)\right)$ is a closed subset of $\mathbb{R}^k$, we infer that $X^i(\gamma_u(T))$ belongs to $\operatorname{Im}_{\mathcal{F}}^{x,T}(u)$.
If we are now in M, then there exists a local chart around x and $\bar{t} \in (0,T)$ such that $\gamma_u(\bar{t}) \in \mathcal{O}$. Set $T' := T - \bar{t}$ and define $u^1 \in L^2([0,\bar{t}];\mathbb{R}^k)$ and $u^2 : L^2([0,T'];\mathbb{R}^k)$ by

$$\begin{cases} u^1(t) = u(t) \quad \forall t \in [0,\bar{t}] \\ u^2(t) = u(t+\bar{t}) \quad \forall t \in [0,T'] \end{cases}$$

We conclude easily by the above proof in $\mathbb{R}^n$ together with (1.15). $\square$

1.3 Regular and singular horizontal paths

Regular and singular controls

Let $\mathcal{F} = \{X^1, \ldots, X^k\}$ be a family of $k \geq 1$ smooth vector fields on M. Given $x \in M$ and $T > 0$, we say that the control $u \in U_{\mathcal{F}}^{x,T}$ is *regular* with respect to x and $\mathcal{F}$ if $\mathrm{rank}_{\mathcal{F}}^{x,T}(u) = n$ (recall that M has dimension n). Otherwise, we shall say that u is *singular*. In other terms, u is singular if and only if it is a critical point of the End-Point mapping $E_{\mathcal{F}}^{x,T}$, that is if $E_{\mathcal{F}}^{x,T}$ is not a submersion at u.

Remark 1.3.1. *Proposition 1.2.10 shows that if $\mathcal{F} = \{X^1, \ldots, X^k\}$ is a family of smooth vector fields on M such that*

$$Span\Big\{X^1(x), \ldots, X^k(x)\Big\} = T_xM \qquad \forall x \in M,$$

then every non-trivial admissible control (that is u in some $U_{\mathcal{F}}^{x,T}$ with $u \neq 0, T > 0$) is regular.

By Propositions 1.2.6, 1.2.7, we observe that a given control $u \in U_{\mathcal{F}}^{x,T}$ is singular with respect to x and $\mathcal{F}$ if and only if any control of the form $u_\lambda \in U_{\mathcal{F}}^{x,\lambda^{-1}T}$ (with $\lambda \neq 0$) is singular with respect to x and $\mathcal{F}$ and if and only if $\breve{u} \in U_{\mathcal{F}}^{y,T}$ (with $y = E_{\mathcal{F}}^{x,T}$) is singular with respect to y and $\mathcal{F}$. Furthermore, Proposition 1.2.8 yields immediately the following result.

Proposition 1.3.2. *Let $u \in U_{\mathcal{F}}^{x,T}$ be a control which is singular with respect to x and $\mathcal{F}$. Let $T^1, T^2, T^3 > 0$ be such that $T^1 + T^2 + T^3 = T$ and $x^1, x^2 \in M$, and $u^1 \in U_{\mathcal{F}}^{x,T^2}, u^2 \in U_{\mathcal{F}}^{x^1,T^2}, u^3 \in U_{\mathcal{F}}^{x^2,T^3}$ be defined as*

$$\left\{\begin{array}{ll} u^1(t) = u(t) & \forall t \in [0, T^1] \\ u^2(t) = u(t + T^1) & \forall t \in [0, T^2] \\ u^3(t) = u(t + T^2) & \forall t \in [0, T^3] \end{array}\right. \quad \textit{and} \quad \left\{\begin{array}{l} x^1 = E_{\mathcal{F}}^{x,T^1}(u^1) \\ x^2 = E_{\mathcal{F}}^{x^1,T^2}(u^2). \end{array}\right.$$

Then all the controls u^1, u^2, u^3 are singular with respect to x and $\mathcal{F}$.

Define k Hamiltonians $h^1, \cdots, h^k : T^*M \to \mathbb{R}$ by

$$h^i = h_{X^i}, \quad \forall i = 1, \cdots, k,$$

that is

$$h^i(\psi) = p \cdot X^i(x) \qquad \forall \psi = (x, p) \in T^*M, \quad \forall i = 1, \ldots, k.$$

For every $i = 1, \ldots, m$, $\overrightarrow{h}_i$ denote the Hamiltonian vector field on T^*M associated to h_i, that is satisfying $\iota_{\overrightarrow{H}}\omega = -dH$, where ω denotes the canonical symplectic form on $T * M$. In local coordinates on T^*M, the Hamiltonian vector field $\overrightarrow{h}_i$ reads

$$\overrightarrow{h}_i(x, p) = \left(\frac{\partial h_i}{\partial p}(x, p), -\frac{\partial h_i}{\partial x}(x, p)\right).$$

Singular controls can be characterized as follows.

Proposition 1.3.3. *The control $u \in U_{\mathcal{F}}^{x,T}$ is singular with respect to x and $\mathcal{F}$ if and only if there exists an absolutely continuous arc $\psi : [0,T] \to T^*M$ that never intersects the zero section of T^*M, such that*

$$\dot{\psi}(t) = \sum_{i=1}^{k} u_i(t) \overrightarrow{h}^i(\psi(t)) \qquad a.e.\ t \in [0,T] \tag{1.18}$$

and

$$h^i(\psi(t)) = 0, \quad \forall t \in [0,T] \qquad \forall i = 1, \cdots, k. \tag{1.19}$$

We say that ψ is an abnormal extremal lift of $\gamma_u : [0,T] \to M$ (defined by (1.7)).

Proof. Let us first assume that we work in $\mathbb{R}^n$. If $D_u E_{\mathcal{F}}^{x,T} : L^2([0,T];\mathbb{R}^k) \to \mathbb{R}^n$ is not surjective, there exists $p \in (\mathbb{R}^n)^* \setminus \{0\}$ such that

$$p \cdot D_u E_{\mathcal{F}}^{x,T}(v) = 0 \qquad \forall v \in L^2\big([0,T];\mathbb{R}^k\big).$$

Remembering Remark 1.2.5, the above identity can be written as

$$\int_0^T pS(T)S(t)^{-1}B(t)v(t)dt = 0 \qquad \forall v \in L^2([0,T];\mathbb{R}^k).$$

Taking $v \in L^2([0,T];\mathbb{R}^k)$ defined as

$$v(t) = \big(pS(t)S(t)^{-1}B(t)\big)^* \qquad \forall t \in [0,T],$$

we deduce that

$$\int_0^T \left| \big(pS(T)S(t)^{-1}B(t)\big)^* \right|^2 ds = 0,$$

which implies that $pS(T)S(t)^{-1}B(t) = 0$ for any $t \in [0,T]$ (note that $t \mapsto pS(T)S(t)^{-1}B(t)$ is continuous). Let us now define, for each $t \in [0,T]$,

$$p(t) := pS(T)S(t)^{-1}.$$

By construction, $p : [0,T] \to (\mathbb{R}^n)^*$ is an absolutely continuous arc. Since $p \neq 0$ and $S(t)$ is invertible for all $t \in [0,T]$, $p(t)$ does not vanish on $[0,T]$. Moreover, recalling that, by definition of S,

$$\frac{d}{dt}S(t)^{-1} = -S(t)^{-1}A(t) \qquad \text{a.e. } t \in [0,T],$$

we conclude that p satisfies the following properties:

$$\dot{p}(t) = -p(t)A(t) \qquad \text{a.e. } t \in [0,T]$$

and

$$p(t)B(t) = 0 \qquad \forall t \in [0,T].$$

Which shows that (1.18)-(1.19) are satisfied with $\psi(t) = (\gamma_u(t), p(t))$ for any $t \in [0,T)$. By the way, we note that by construction, we have for every $t \in (0,T]$,

$$p(t) \cdot D_{u_t} E_{\mathcal{F}}^{x,t}(v) = 0 \qquad \forall v \in L^2\big([0,t];\mathbb{R}^k\big), \tag{1.20}$$

where u_t denotes the restriction of u to $[0,t]$.

Conversely, let us assume that there exists an absolutely continuous arc $p:[0,T]\to(\mathbb{R}^n)^*\setminus\{0\}$ such that (1.18) and (1.19) are satisfied with $\psi=(\gamma_u,p)$. This means that

$$-\dot{p}(t)=p(t)A(t) \qquad \text{a.e. } t\in[0,T]$$

and

$$p(t)^*B(t)=0 \qquad \forall t\in[0,T].$$

Setting $p:=p(T)\neq 0$, we have, for any $t\in[0,T]$,

$$p(t)=pS(T)S(t)^{-1}.$$

Hence, we obtain

$$pS(T)S(t)^{-1}B(t)=0 \qquad \forall t\in[0,T],$$

which in turn implies

$$p\cdot D_uE_{\mathcal{F}}^{x,T}(v)=0, \quad \forall v\in L^2([0,T];\mathbb{R}^k).$$

This concludes the proof. Again, the above proof shows indeed that (1.20) holds for any $t\in(0,T]$.

Assume now that we work on M. By Proposition 1.3.2, we can cut the path $\gamma_u:[0,T]\to M$ associated with u and u itself into a finite number of peaces $\gamma^1,\ldots,\gamma^l$ and $u^1,\ldots,u^l$ such that each control u^l is singular and each path γ^l is valued in a chart of M. Then we can apply the previous arguments on each chart and thanks to (1.20) obtain a non-vanishing absolutely continuous arc ψ satisfying (1.18)-(1.19) on $[0,T]$. □

Remark 1.3.4. *We keep in mind that if $\psi:[0,T]\to T^*M$ is an absolutely continuous arc satisfying (1.18)-(1.19), then*

$$p(t)\cdot D_{u_t}E_{\mathcal{F}}^{x,t}(v)=0 \qquad \forall v\in L^2\big([0,t];\mathbb{R}^k\big),$$

where $\psi(t)=(\gamma_u(t),p(t))$ and u_t denotes the restriction of u to $[0,t]$. (In the sequel, $\psi\cdot v$ or $p\cdot v$ with $\psi=(x,v)$ in local coordinates denotes the evaluation of the form ψ at $v\in T_xM$.)

Remark 1.3.5. *In local coordinates, Proposition 1.3.3 means that there exists an absolutely continuous arc $p:[0,T]\to\big(\mathbb{R}^n\big)^*\setminus\{0\}$ satisfying*

$$\dot{p}(t)=-\sum_{i=1}^{k}u_i(t)\,p(t)\cdot D_{\gamma_u(t)}X^i \qquad a.e.\ t\in[0,T] \tag{1.21}$$

and

$$p(t)\cdot X^i\big(\gamma_u(t)\big)=0 \qquad \forall t\in[0,T],\ \forall i=1,\ldots k. \tag{1.22}$$

Regular and singular paths

Let Δ be a distribution of rank $m \leq n$ in M. As seen before, it can be represented by a generating family $\mathcal{F} = \{X^1, \ldots, X^k\}$ of smooth vector fields (see Proposition 1.1.8). Given a point $x \in M$, a time $T > 0$, and an horizontal path $\gamma \in \Omega_\Delta^{x,T}$, we set

$$\mathrm{Im}_\Delta(\gamma) := D_u E_{\mathcal{F}}^{x,T}\left(L^2([0,T];\mathbb{R}^k)\right) \subset T_{E_{\mathcal{F}}^{x,T}(u)}M,$$

where $u \in U_{\mathcal{F}}^{x,T}$ is any control such that $\gamma = \gamma_u$ (see Proposition 1.2.2). The definition does not depend on the frame.

Proposition 1.3.6. *Let $\mathcal{F} = \{X^1, \ldots, X^k\}, \mathcal{F}' = \{Y^1, \ldots, Y^{k'}\}$ be two generating families for Δ and $x \in M, T > 0$ be fixed. If $u \in U_{\mathcal{F}}^{x,T}$ and $u' \in U_{\mathcal{F}'}^{x,T}$ satisfy*

$$\gamma_u^{\mathcal{F}}(t) = \gamma_{u'}^{\mathcal{F}'}(t) \qquad \forall t \in [0,T],$$

where $\gamma_u^{\mathcal{F}}$ (resp. $\gamma_{u'}^{\mathcal{F}'}$) denotes the solution to the Cauchy problem (1.7) associated with $\mathcal{F}$ (resp. $\mathcal{F}'$), then

$$Im_{\mathcal{F}}^{x,T}(u) = Im_{\mathcal{F}'}^{x,T}(u').$$

Proof. It is sufficient to prove that $\mathrm{Im}_{\mathcal{F}'}^{x,T}(u') \subset \mathrm{Im}_{\mathcal{F}}^{x,T}(u)$. For every $t \in [0,T]$, there is an open set $\mathcal{O}_t$ of $\gamma_u^{\mathcal{F}}(t)$ in M and m integers $i_1^t, \ldots, i_m^t \in \{1, \ldots, k\}$ such that

$$\mathrm{Span}\,\left\{X^{i_1^t}(x), \ldots, X^{i_m^t}(x)\right\} = \Delta(x) \qquad \forall x \in \mathcal{O}_t.$$

The curve $\gamma_u^{\mathcal{F}}([0,T])$ is compact and is contained in $\cup_{t\in[0,T]}\mathcal{O}_t$. Hence, there are N times $t_1, \ldots, t_N \in [0,T]$ together with a partition of unity $\{\psi_j\}$ such that

$$[0,T] \subset \bigcup_{j=1}^N \mathcal{O}_{t_j}, \quad \mathrm{Supp}\,(\psi_j) \subset \mathcal{O}_{t_j}, \quad \sum_{j=1}^N \psi_j = 1.$$

For every j, there is a smooth mapping $U^j : TM \rightarrow \mathbb{R}^m$ with $U_j(0) = 0$ such that

$$v = \sum_{l=1}^m U_l^j(v) X^{i_l^{t_j}}(x),$$

for every $(x,v) \in TM$ with $x \in \mathcal{O}_{t_j}$ and $v \in \Delta(x)$. Then, there holds for every $x \in \mathcal{O}_{t_j}$, every $w' \in \mathbb{R}^{k'}$, and every $v \in \Delta(v)$,

$$\sum_{i=1}^{k'} w_i' Y^i(x) = v + \sum_{l=1}^m U_l^j \left(\sum_{j=1}^{k'} w_j' Y^j(x) - v\right) X^{i_l^{t_j}}(x).$$

By Gronwall's Lemma (see Appendix A), there is a neighborhood $\mathcal{U}' \subset U_{\mathcal{F}'}^{x,T}$ of u' in $L^2([0,T];\mathbb{R}^{k'})$ such that for every $w' \in \mathcal{U}'$, the trajectory $\gamma_{w'}^{\mathcal{F}'}$ starting at x associated with w' is contained in $\cup_{j=1}^N \mathcal{O}_{t_j}$. We infer that for every $w' \in \mathcal{U}'$, there holds

$$
\begin{aligned}
&\sum_{i=1}^{k'} w_i'(t) Y^i\big(\gamma_{w'}^{\mathcal{F}'}(t)\big) - \sum_{i=1}^{k} u_i(t) X^i\big(\gamma_{w'}^{\mathcal{F}'}(t)\big) \\
&= \sum_{j=1}^{N}\sum_{l=1}^{m} \left[\psi_j\big(\gamma_{w'}^{\mathcal{F}'}(t)\big) U_l^j \left(\sum_{i=1}^{k'} w_i'(t) Y^i\big(\gamma_{w'}^{\mathcal{F}'}(t)\big) - \sum_{i=1}^{k} u_i(t) X^i\big(\gamma_{w'}^{\mathcal{F}'}(t)\big) \right) \right] \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad X^{i_l^{t_j}}\big(\gamma_{w'}^{\mathcal{F}'}(t)\big),
\end{aligned}
$$

for almost every $t \in [0,T]$. Each mapping

$$
\begin{aligned}
& w' \in \mathcal{U}' \\
&\qquad \longmapsto \quad \psi_j\big(\gamma_{w'}^{\mathcal{F}'}(t)\big) U^j \left(\sum_{j=1}^{k'} w_j'(\cdot) Y^j\big(\gamma_{w'}^{\mathcal{F}'}(\cdot)\big) - \sum_{i=1}^{k} u_i(\cdot) X^i\big(\gamma_{w'}^{\mathcal{F}'}(\cdot)\big) \right) \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \in L^2\big([0,T];\mathbb{R}^m\big)
\end{aligned}
$$

is at least of class C^1. Therefore, since $\gamma_{u'}^{\mathcal{F}'}(t) = \gamma_u^{\mathcal{F}}(t)$ for any $t \in [0,T]$ and $U^j(0) = 0$ for all j, there is a C^1 mapping $G' : \mathcal{U}' \to U_{\mathcal{F}}^{x,T}$ with $G'(u') = u$ such that

$$
E_{\mathcal{F}'}^{x,T}(w') = E_{\mathcal{F}}^{x,T}\big(G'(w')\big) \qquad \forall w' \in \mathcal{U}'.
$$

We infer that $\mathrm{Im}_{\mathcal{F}'}^{x,T}(u') \subset \mathrm{Im}_{\mathcal{F}}^{x,T}(u)$. $\square$

We call *rank* of $\gamma \in \Omega_\Delta^{x,T}$, denoted by $\mathrm{rank}_\Delta(\gamma)$, the dimension of $\mathrm{Im}_\Delta(u)$. We shall say that γ is *singular* (with respect to Δ) if $\mathrm{rank}_\Delta(\gamma) < n$ and *regular* otherwise.

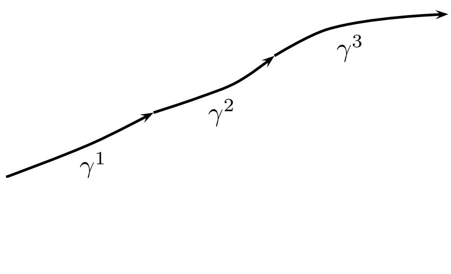
Remark 1.3.7. *By Remark 1.3.1, if Δ has rank $m = n$ then any non-trivial horizontal path is regular.*

Propositions 1.2.6, 1.2.7, 1.2.8, 1.2.10, 1.3.2 do apply to horizontal paths. The rank of an horizontal path depends only on the curve drawn by the path in M, it does not depend upon its parametrization. Proposition 1.2.8 yields the following result (the concatenation of paths is defined in the same way as the concatenation of controls):

Proposition 1.3.8. *Let $\gamma \in \Omega_\Delta^{x,T}$ be a singular horizontal path, $T^1, T^2, T^3 > 0$ be such that $T^1 + T^2 + T^3 = T$ and $\gamma^1 \in \Omega_\Delta^{x,T^1}, \gamma^2 \in \Omega_\Delta^{\gamma^1(T^1),T}, \gamma^3 \in \Omega_\Delta^{\gamma^2(T^2),T}$ be such that*

$$
\gamma = \gamma^1 * \gamma^2 * \gamma^3.
$$

Then the horizontal paths $\gamma^1, \gamma^2, \gamma^3$ are singular.



In a more geometric way, singular horizontal paths can be characterized as follows. Recall that T^*M denotes the cotangent bundle of M, $\pi : T^*M \to M$ the canonical projection, and ω the canonical symplectic form on T^*M. We denote by $\Delta^\perp \subset T^*M$ the annihilator of Δ in T^*M, that is

$$\Delta^\perp(x) = \Big\{ p \in T_x^*M \,|\, p \cdot v = 0, \, \forall v \in \Delta(x) \Big\} \qquad \forall x \in M.$$

It is a rank $n - m$ subbundle of the cotangent bundle T^*M, that is a smooth map that assigns to each point x of M a linear subspace $\Delta^\perp(x)$ of T_x^*M of dimension $n - m$ (or co-dimension m). In particular, the subbundle $\Delta^\perp$ is a submanifold of T^*M of dimension $2n - m$. Let $\bar{\omega}$ denote the restriction of ω to $\Delta^\perp$. This restriction needs not be symplectic, and hence it might admits characteristics subspaces Ker $\bar{w}(\psi)$ at $\psi \in \Delta^\perp$. We recall that the kernel of a bilinear form $\sigma = \bar{\omega}_\psi$ on $T_\psi \Delta^\perp$ is defined as

$$\text{Ker } \sigma = \Big\{ \xi \in T_\psi \Delta^\perp \,|\, \sigma\big(\xi, \xi'\big) = 0, \, \forall \xi' \in T_\psi \Delta^\perp \Big\}.$$

Definition 1.3.9. *A characteristic curve of $\bar{\omega}$ on $[0, T]$ is an absolutely continuous curve $\psi : [0, T] \to T^*M$ that never intersects the zero section of T^*M such that*

$$\psi(t) \in \Delta^\perp \qquad \forall t \in [0, T]$$

and

$$\dot{\psi}(t) \in \mathit{Ker}\ \bar{\omega}\big(\psi(t)\big) \qquad a.e. \ t \in [0, T].$$

Proposition 1.3.10. *The horizontal path $\gamma \in \Omega_\Delta^{x,T}$ is singular if and only if it is the projection of a characteristic curve of $\bar{\omega}$ on $[0, T]$.*

Proof. Let $\gamma \in \Omega_\Delta^{x,T}$ and a k generating family $\mathcal{F} = \{X^1, \ldots, X^k\}$ for Δ be fixed.

Lemma 1.3.11. *The tangent space to $\Delta^\perp$ at some $\psi \in \Delta^\perp$ satisfies*

$$T_\psi \Delta^\perp = \Big\{ \xi \in T_\psi T^*M \,|\, \omega_\psi\big(\overrightarrow{h}^i(\psi), \xi\big) = 0, \, \forall i = 1, \ldots, k \Big\} \tag{1.23}$$

and

$$\overrightarrow{h}^i(\psi) \in T_\psi \Delta^\perp \qquad \forall i = 1, \ldots, k. \tag{1.24}$$

Proof of Lemma 1.3.11. Let $\psi = (x, p) : (-\epsilon, \epsilon) \to T^*M$ be a smooth curve in $\Delta^\perp$ such that $\psi(0) = \psi$ and $\dot{\psi}(0) = \xi$ in $T_\psi \Delta^\perp$. There holds for every $i = 1, \ldots, k$,

$$p(t) \cdot X^i\big(x(t)\big) = 0 \qquad \forall t \in (-\epsilon, \epsilon),$$

which means that $h^i(\psi(t)) = 0$ for any $t \in (-\epsilon, \epsilon)$. Derivating yields

$$D_{\psi(t)} h^i \cdot \dot{\psi}(t) = 0 \qquad \forall t \in (-\epsilon, \epsilon),$$

which by definition of the $\overrightarrow{h}^i$'s means that

$$\omega_{\psi(t)} \left(\overrightarrow{h}^i(\psi(t)), \dot{\psi}(t) \right) = 0 \qquad \forall t \in (-\epsilon, \epsilon), \, \forall i = 1, \ldots, k.$$

Taking the above equality at $t=0$, we infer that $\omega_\psi\big(\overrightarrow{h}^i(\psi),\xi\big)=0$, which in turn shows that

$$T_\psi\Delta^\perp \subset \Big\{\xi \in T_\psi T^*M \mid \omega_\psi\big(\overrightarrow{h}^i(\psi),\xi\big)=0,\ \forall i=1,\ldots,k\Big\}.$$

By definition of the $\overrightarrow{h}^i$'s, the vector space appearing in the right-hand side can be seen as

$$\cap_{i=1}^k \mathrm{Ker}\ D_\psi h^i$$

Since $\mathcal{F}=\{X^1,\ldots,X^k\}$ is a k generating family for Δ of rank m, the linear forms $D_\psi h^1,\ldots,D_\psi h^k$ span a space of dimension m in the dual of $T_\psi T^*M$. This shows that the intersection of the Ker $D_\psi h^i$'s has dimension $2n-m$. The equality (1.23) follows. Consider now for every $i=1,\ldots,k$, $\psi^i:[0,\epsilon]\to T^*M$ a local solution to the Cauchy problem

$$\dot\psi^i(t)=\overrightarrow{h}^i\big(\psi^i(t)\big) \quad \forall t\in[0,\epsilon], \quad \psi^i(0)=\psi.$$

Since h^i is constant along the integral curves of $\overrightarrow{h}^i$ and $\psi\in\Delta^\perp$, there holds

$$h^i\big(\psi^i(t)\big)=0 \qquad \forall t\in[0,\epsilon],\ \forall i=1,\ldots,k.$$

This implies that $\psi^i(t)$ always remains in $\Delta^\perp$ and in turn gives (1.24). $\square$

Let us first assume that γ is singular. By definition, this means that there exists a control $u\in U_{\mathcal{F}}^{x,T}$ which is singular and such that

$$\dot\gamma(t)=\sum_{i=1}^k u_i(t)X^i\left(\gamma(t)\right) \quad \text{a.e. } t\in[0,T].$$

By Proposition 1.3.3, there exists an absolutely continuous arc $\psi:[0,T]\to T^*M$ that never intersects the zero section of T^*M, such that

$$\dot\psi(t)=\sum_{i=1}^k u_i(t)\overrightarrow{h}^i(\psi(t)) \qquad \text{a.e. } t\in[0,T]$$

and

$$h^i(\psi(t))=0, \quad \forall t\in[0,T] \qquad \forall i=1,\cdots,k.$$

The first property together with Lemma 1.3.11 shows that

$$\omega_{\psi(t)}\left(\dot\psi(t),\xi\right)=0 \qquad\qquad \text{a.e. } t\in[0,T],\ \forall\xi\in T_{\psi(t)}\Delta^\perp,$$

while the second property means that $\psi(t)$ belongs to $\Delta^\perp$ for all $t\in[0,T]$, which implies

$$\dot\psi(t)\in T_{\psi(t)}\Delta^\perp \qquad \text{a.e. } t\in[0,T].$$

Therefore, we deduce that $\dot\psi(t)$ belongs to Ker $\bar\omega\big(\psi(t)\big)$ for a.e. $t\in[0,T]$, which shows that γ is the projection of a characteristic curve of $\bar\omega$ on $[0,T]$.

Conversely, assume now that γ is the projection of a characteristic curve of $\bar\omega$ on $[0,T]$, that is that there exists an absolutely continuous curve $\psi:[0,T]\to$

T^*M that never intersects the zero section of T^*M whose projection is γ and such that

$$\psi(t) \in \Delta^\perp \qquad \forall t \in [0,T]$$

and

$$\dot{\psi}(t) \in \text{Ker } \bar{\omega}\big(\psi(t)\big) \qquad \text{a.e. } t \in [0,T].$$

Let $t \in [0,T]$ be fixed such that ψ is differentiable at t. Thanks to Lemma 1.3.11, there holds for every $i = 1, \ldots, k$,

$$\overrightarrow{h}^i\big(\psi(t)\big) \in T_\psi \Delta^\perp \quad \text{ and } \quad \omega_{\psi(t)}\big(\overrightarrow{h}^i(\psi), \xi\big) = 0\, \forall \xi \in T_\psi\Delta^\perp.$$

This means that $\overrightarrow{h}^i\big(\psi(t)\big)$ belongs to Ker $\bar{\omega}\big(\psi(t)\big)$. Hence

$$\xi(t) := \dot{\psi}(t) - \sum_{i=1}^k u_i(t)\overrightarrow{h}^i\big(\psi(t)\big) \in \text{Ker } \bar{\omega}\big(\psi(t)\big).$$

Since γ is the projection of ψ, $\psi(t)$ and $\xi(t)$ have the form (in local coordinates):

$$\psi(t) = \big(\gamma(t), p(t)\big) \quad \text{ and } \quad \xi(t) = \big(0, \theta(t)\big).$$

Therefore, there holds

$$0 = \omega_{\psi(t)}\big(\xi(t), \xi\big) = -\theta(t) \cdot v \qquad \forall \xi = (v,\theta) \in T_\psi \Delta^\perp.$$

Since $\Delta^\perp$ can be seen as the graph of the mapping $x \mapsto \Delta(x) \subset T_x^*M$, there holds

$$D_\psi \pi \left(T_\psi \Delta^\perp\right) = T_x M.$$

Therefore, we infer that $\theta(t) = 0$, which proves that $\dot{\psi}(t) = \sum_{i=1}^k u_i(t)$. We conclude easily by Proposition 1.3.3. □

Examples

Example 1.3.12. *Returning to Examples 1.1.2 and 1.1.19, consider in $\mathbb{R}^3$ with coordinates $x = (x_1, x_2, x_3)$, the totally nonholonomic rank two distribution Δ defined by*

$$\Delta(x) = Span\Big\{X^1(x), X^2(x)\Big\} \qquad \forall x \in \mathbb{R}^3$$

with

$$X^1 = \partial_{x_1} - \frac{x_2}{2}\partial_{x_3} \quad \text{ and } \quad X^2 = \partial_{x_2} + \frac{x_1}{2}\partial_{x_3}.$$

We claim that the singular horizontal paths are the constant curves or equivalently that the only singular control with respect to $\mathcal{F} = \{X^1, X^2\}$ is the control $u \equiv 0$. Let us prove this claim. Let $x \in \mathbb{R}^3, T > 0$ be fixed and $u \in U_{\mathcal{F}}^{x,T}$ be a singular control. Denote by $x : [0,T] \to \mathbb{R}^3$ the solution to the Cauchy problem

$$\dot{x}(t) = u_1(t)X^1(x(t)) + u_2(t)X^2(x(t)) \quad a.e.\ t \in [0,T], \quad x(0) = x. \qquad (1.25)$$

From Proposition 1.3.3, there exists an absolutely continuous arc $p : [0,T] \to \left(\mathbb{R}^3\right)^* \setminus \{0\}$ such that

$$\dot{p}(t) = -u_1(t)\, p(t) \cdot D_{x(t)}X^1 - u_2(t)\, p(t) \cdot D_{x(t)}X^2 \qquad (1.26)$$

for a.e. $t \in [0,T]$ *and*

$$p(t) \cdot X^1(x(t)) = p(t) \cdot X^2(x(t)) = 0 \qquad \forall t \in [0,T]. \tag{1.27}$$

Taking the derivatives in (1.27) gives

$$\dot{p}(t) \cdot X^i(x(t)) + p(t) \cdot D_{x(t)} X^i\big(\dot{x}(t)\big) = 0 \qquad a.e.\ t \in [0,T],\ \forall i = 1,2.$$

Which implies, by (1.25)-(1.26),

$$u_1(t)\, p(t) \cdot [X^1, X^i](x(t)) + u_2(t)\, p(t) \cdot [X^2, X^i](x(t)) = 0 \qquad a.e.\ t \in [0,T].$$

Taking $i = 1$ *and* $i = 2$, *we obtain that for almost every* $t \in [0,T]$,

$$u_1(t)\, p(t) \cdot [X^1, X^2](x(t)) = u_2(t)\, p(t) \cdot [X^1, X^2](x(t)) = 0.$$

This can be written as

$$|u(t)|^2 \left(p(t) \cdot [X^1, X^2](x(t))\right)^2 = 0 \qquad a.e.\ t \in [0,T],$$

Since $[X^1, X^2] = -\frac{\partial}{\partial x_3}$ *and (1.27) is satisfied with* $p(t) \neq 0$, *we deduce that* $u \equiv 0$.

Example 1.3.13. *The property of the previous example is satisfied by much more general distributions. A distribution* Δ *on* M *is called fat if, for every* $x \in M$ *and every section* X *of* Δ *with* $X(x) \neq 0$, *there holds*

$$T_xM = \Delta(x) + \big[X, \Delta\big](x), \tag{1.28}$$

where

$$\big[X, \Delta\big](x) := \Big\{[X, Z](x) \,|\, Z \text{ section of } \Delta\Big\}.$$

The condition above being very restrictive, there are very few fat distributions. Fat distributions on three-dimensional manifolds are the rank-two distributions Δ *satisfying*

$$T_xM = Span\Big\{X^1(x), X^2(x), \big[X^1, X^2\big](x)\Big\} \qquad \forall x \in \mathcal{V},$$

where (X^1, X^2) *is a local frame for* Δ *in* $\mathcal{V}$. *Another example of co-rank one fat distributions in odd dimension is given by contact distributions which were introduced in Example 1.1.23. In this case property (3.44) is an easy consequence of (1.4). Let us now prove that fat distributions do not admit non-trivial singular horizontal paths. By Proposition 1.3.8, we just need to show that non-constant short horizontal paths cannot be singular. Taking a local chart if necessary we can work in* $\mathbb{R}^n$ *and assume that* Δ *has a local frame* $X^1, \ldots, X^m$. *Let* $x \in \mathbb{R}^n, T > 0$ *be fixed and* $u \in U_{\mathcal{F}}^{x,T}$ *be a singular control. By Remark 1.3.5, there exists an absolutely continuous arc* $p : [0,T] \to \big(\mathbb{R}^n\big)^* \setminus \{0\}$ *satisfying (1.21)-(1.22). For almost every fixed* $t \in [0,T]$ *and every* $i = 1, \ldots, m$, *derivating (1.22) yields*

$$\sum_{j=1}^m u_j(t)\, p(t) \cdot \big[X^j, X^i\big]\left(\gamma_u(t)\right) = p(t) \cdot \left[\sum_{j=1}^m u_j(t) X^j, X^i\right] \left(\gamma_u(t)\right) = 0.$$

Setting the autonomous vector field $X(\cdot) := \sum_{j=1}^m u_j(t) X^j(\cdot)$, *we deduce that* $p(t)$ *annihilates all the* $X^i\big(\gamma_u(t)\big)$*'s and all the* $\big[X, X^i\big]\left(\gamma_u(t)\right)$*'s. This contradicts (3.44).*

Example 1.3.14. *Returning to Example 1.1.21, we consider in $\mathbb{R}^3$ with coordinates $x = (x_1, x_2, x_3)$, the totally nonholonomic rank two distribution Δ defined by*

$$\Delta(x) = Span\Big\{X^1(x), X^2(x)\Big\} \qquad \forall x \in \mathbb{R}^3$$

with

$$X^1 = \partial_{x_1} \quad \text{and} \quad X^2 = \partial_{x_2} + \frac{x_1^2}{2}\partial_{x_3}.$$

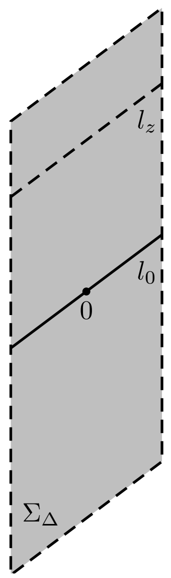
We claim that the singular horizontal curves are exactly the "traces of the distribution" on the surface

$$\Sigma_\Delta := \Big\{x \in \mathbb{R}^3 \,|\, x_1 = 0\Big\},$$

which in other terms means that the singular horizontal paths are either constant curves or are contained in a line l_z of the form

$$l_z = \Big\{x = (x_1, x_2, x_3) \in \mathbb{R}^3 \,|\, x_1 = 0 \text{ and } x_3 = z\Big\}$$

for some $z \in \mathbb{R}$.



Let us prove this claim. Let $x \in \mathbb{R}^3, T > 0$ be fixed and $u \in U_{\mathcal{F}}^{x,T}$ be a nontrivial singular control. Denote by $x : [0,T] \to \mathbb{R}^3$ the solution to the Cauchy problem

$$\dot{x}(t) = u_1(t)X^1(x(t)) + u_2(t)X^2(x(t)) \quad \text{a.e. } t \in [0,T], \quad x(0) = x.$$

As in the previous example, from Proposition 1.3.3, there exists an absolutely continuous arc $p : [0,T] \to \left(\mathbb{R}^3\right)^* \setminus \{0\}$ such that

$$\dot{p}(t) = -u_1(t)\, p(t) \cdot D_{x(t)}X^1 - u_2(t)\, p(t) \cdot D_{x(t)}X^2 \tag{1.29}$$

for a.e. $t \in [0,T]$ and

$$p(t) \cdot X^1(x(t)) = p(t) \cdot X^2(x(t)) = 0 \qquad \forall t \in [0,T]. \tag{1.30}$$

We deduce that

$$|u(t)|^2 \left(p(t) \cdot [X^1, X^2](x(t))\right)^2 = 0 \qquad a.e.\ t \in [0, T].$$

Since the three vectors $X^1(x), X^2(x), [X^1, X^2](x)$ *span* $\mathbb{R}^3$ *for every* x *with* $x_1 \neq 0$*, this shows that* $x_1(t) = 0$ *for all* $t \in [0, T]$*, which in turn implies that* $u_1 \equiv 0$*. We deduce that* x *has the form*

$$x(t) = \left(0, x_2(0) + \int_0^t u_2(s)ds, 0, x_3(0)\right),$$

which shows that it is contained in $l_{x_3(0)}$*. Conversely, if an horizontal path* $x \in \Omega_\Delta^{x,T}$ *has the form*

$$x(t) = (0, x_2(t), z) \qquad \forall t \in [0, T],$$

with $z \in \mathbb{R}$*, then any absolutely continuous arc* $p : [0, T] \to \mathbb{R}^3 \setminus \{0\}$ *of the form*

$$p(t) = (0, 0, p_3) \qquad \forall t \in [0, T],$$

with $p_3 \neq 0$ *satisfies (1.29) and (1.30). This shows that any horizontal path which is contained in a line* l_z *for some* $z \in \mathbb{R}$ *is singular.*

Example 1.3.15. *More generally, consider a totally nonholonomic distribution* Δ *of rank two in a manifold* M *of dimension three. We define the Martinet surface of* Δ *as the set defined by*

$$\Sigma_\Delta := \Big\{x \in M \mid \Delta(x) + [\Delta, \Delta](x) \neq T_x M\Big\},$$

where

$$\big[\Delta, \Delta\big](x) := \Big\{[X, Y](x) \mid X, Y \text{ sections of } \Delta\Big\}.$$

In other terms, a point $x \in M$ *belongs to* Σ_Δ *if and only if* Δ *is not a contact distribution at* x*, that is if for any (or for only one) local frame* $\{X^1, X^2\}$ *in a neighborhood of* x *the three vectors* $X^1(x), X^2(x), [X^1, X^2](x)$ *do not span* $T_x M$*. The singular paths with respect to* Δ *are exactly the horizontal paths which are contained in* Σ_Δ*. Let us prove this claim. The fact that singular curves are necessary included in* Σ_Δ *follows by the same argument an in Example 1.3.12. Let us now prove that any horizontal path which is included in* Σ_Δ *is singular. Let* $\gamma : [0, T] \to M$ *such a path be fixed, set* $\gamma(0) = x$*, and consider a local frame* $\{X^1, X^2\}$ *for* Δ *in a neighborhood* $\mathcal{V}$ *of* x*. Let* $\delta > 0$ *be small enough so that* $\gamma(t) \in \mathcal{V}$ *for any* $t \in [0, \delta]$*, in such a way that there is* $u \in L^2([0, \delta]; \mathbb{R}^2)$ *satisfying*

$$\dot\gamma(t) = u_1(t) X^1(\gamma(t)) + u_2(t) X^2(\gamma(t)) \qquad a.e.\ t \in [0, \delta].$$

Taking a change of coordinates if necessary, we can assume that we work in $\mathbb{R}^3$*. Let* $p_0 \in (\mathbb{R}^3)^* \setminus \{0\}$ *be such that* $p_0 \cdot X_1(x) = p_0 \cdot X_2(x) = 0$*, and let* $p : [0, \delta] \to (\mathbb{R}^3)^*$ *be the solution to the Cauchy problem*

$$\dot p(t) = -\sum_{i=1,2} u_i(t)\, p(t) \cdot D_{\gamma(t)} X^i \qquad a.e.\ t \in [0, \delta], \quad p(0) = p_0.$$

Define two absolutely continuous function $h_1, h_2 : [0, \delta] \to \mathbb{R}$ by

$$h_i(t) = p(t) \cdot X^i(\gamma(t)) \qquad \forall t \in [0, \delta], \quad \forall i = 1, 2.$$

As above, for every $t \in [0, \delta]$ we have

$$\dot{h}_1(t) = \frac{d}{dt}\left[p(t) \cdot X^1(\gamma(t))\right] = -u_2(t)\, p(t) \cdot [X^1, X^2]\big(\gamma(t)\big)$$

and

$$\dot{h}_2(t) = u_1(t)\, p(t) \cdot \left[X^1, X^2\right](\gamma(t)).$$

But since $\gamma(t) \in \Sigma_\Delta$ for every t, there are two continuous functions $\lambda_1, \lambda_2 : [0, \delta] \to \mathbb{R}$ such that

$$\left[X^1, X^2\right](\gamma(t)) = \lambda_1(t) X^1(\gamma(t)) + \lambda_2(t) X^2(\gamma(t)) \qquad \forall t \in [0, \delta].$$

This implies that the pair (h_1, h_2) is a solution of the linear differential system

$$\left\{\begin{array}{lcl} \dot{h}_1(t) & = & -u_2(t)\lambda_1(t)h_1(t) - u_2(t)\lambda_2(t)h_2(t) \\ \dot{h}_2(t) & = & u_1(t)\lambda_1(t)h_1(t) + u_1(t)\lambda_2(t)h_2(t). \end{array}\right.$$

Since $h_1(0) = h_2(0) = 0$ by construction, we deduce by the Cauchy-Lipschitz Theorem that $h_1(t) = h_2(t) = 0$ for any $t \in [0, \delta]$. In that way, we have constructed an absolutely continuous arc $p : [0, \delta] \to \left(\mathbb{R}^3\right)^* \setminus \{0\}$ satisfying (1.3.5)-(1.3.5) (with $\gamma_u = \gamma$). We can repeat this construction on a new interval of the form $[\delta, 2\delta]$ (with initial condition $p(\delta)$) and finally obtain an absolutely continuous arc satisfying (1.3.5)-(1.3.5) on $[0, T]$. By Proposition 1.3.3, we conclude that γ is singular.

Example 1.3.16. *Consider in $\mathbb{R}^4$ the two smooth vector fields X^1, X^2 given by*

$$X^1 = \partial_{x_1}, \quad X^2 = \partial_{x_2} + x_1 \partial_{x_3} + x_3 \partial_{x_4}.$$

These two vector fields are always linearly independent in $\mathbb{R}^4$. Moreover we have

$$[X^1, X^2] = \partial_{x_3}, \quad \left[X^2, [X^1, X^2]\right] = -\partial_{x_4}.$$

Therefore the family $\mathcal{F} = \{X^1, X^2\}$ spans a totally nonholonomic distribution Δ of rank two in $\mathbb{R}^4$. Let us look at singular horizontal paths of Δ or equivalently at singular controls with respect to End-Point mapping $E_{\mathcal{F}}^{x,T}$ with $x \in \mathbb{R}^4$ and $T > 0$. Let $u \in U_{\mathcal{F}}^{x,T}$ be a control satisfying $|u(t)| = 1$ for a.e. $t \in [0, T]$. This control is singular if and only if there is an arc $p = (p_1, p_2, p_3, p_4) : [0, T] \to \left(\mathbb{R}^4\right)^* \setminus \{0\}$ which satisfies (1.21) and (1.22). Denoting by $x = (x_1, x_2, x_3, x_4) : [0, T] \to \mathbb{R}^4$ the trajectory uniquely associated to x and u, (1.21) yields

$$\left\{\begin{array}{lcl} \dot{x}_1(t) & = & u_1(t) \\ \dot{x}_2(t) & = & u_2(t) \\ \dot{x}_3(t) & = & u_2(t)x_1(t) \\ \dot{x}_4(t) & = & u_2(t)x_3(t), \end{array}\right. \qquad \left\{\begin{array}{lcl} \dot{p}_1(t) & = & -u_2(t)p_3(t) \\ \dot{p}_2(t) & = & 0 \\ \dot{p}_3(t) & = & -u_2(t)p_4(t) \\ \dot{p}_4(t) & = & 0, \end{array}\right. \tag{1.31}$$

for a.e. $t \in [0, T]$, while (1.22) yields

$$p_1(t) = p_2(t) + x_1(t)p_3(t) + x_3(t)p_4(t) = 0 \qquad \forall t \in [0, T]. \tag{1.32}$$

System (1.31) implies that p_2 and p_4 are constant on $[0,T]$. If $p_4=0$, then (1.31) also implies that p_3 is constant on $[0,T]$. Hence we obtain that $p_2+x_1(t)p_3=0$ for every $t\in[0,T]$. Which means that either x_1 is constant or $p_2=p_3=0$. Since p does not vanish on $[0,T]$, we deduce that x_1 is constant, which means that $u_1\equiv 0$. But $u_2(t)p_3=0$ for almost every t, hence $p_3=0$ (remember that $|u(t)|=1$ a.e. $t\in[0,T]$). We obtain a contradiction. Therefore, $p_4\neq 0$, hence we deduce easily that

$$0=u_2(t)p_3(t)=\left(-\frac{\dot p_3(t)}{p_4}\right)p_3(t)=0 \qquad a.e.\ t\in[0,T].$$

Since p_3 is absolutely continuous, this means that it is constant on $[0,T]$. This implies that $u_2(t)=0$ for all $t\in[0,T]$. Then, the curve x has the form

$$x(t)=\big(x_1(t),x_2(0),x_3(0),x_4(0)\big) \qquad \forall t\in[0,T].$$

In conclusion, a singular curve passes through each point in $\mathbb{R}^4$.

Example 1.3.17. *The previous phenomena happens for more general rank two distributions in dimension four. Let Δ be a rank two distribution on a four-dimensional manifold M such that for every $x\in M$, there holds*

$$\Delta(x)+[\Delta,\Delta](x) \text{ has dimension three} \tag{1.33}$$

and

$$T_xM=\Delta(x)+[\Delta,\Delta](x)+\big[\Delta,[\Delta,\Delta]\big](x) \qquad \forall x\in M, \tag{1.34}$$

where

$$\big[\Delta,\Delta\big](x):=\Big\{[X,Y](x)\,|\,X,Y \text{ sections of } \Delta\Big\}$$

$$\text{and} \quad \big[\Delta,[\Delta,\Delta]\big](x):=\Big\{\big[X,[Y,Z]\big](x)\,|\,X,Y,Z \text{ sections of } \Delta\Big\}.$$

As above, we can work locally. so let us consider a frame $\{X^1,X^2\}$ and a trajectory $x:[0,T]\to\mathbb{R}^4$ associated to some control $u\in L^2([0,T];\mathbb{R}^2)$. If x is singular (with respect to Δ), there is $p:[0,T]\to\left(\mathbb{R}^4\right)^*\setminus\{0\}$ satisfying (1.21) and (1.22). Derivative two times yields for almost every $t\in[0,T]$ such that $u(t)\neq 0$,

$$p(t)\cdot\big[X^1,X^2\big]\big(x(t)\big)=0 \tag{1.35}$$

and

$$\begin{aligned} u_1(t)\,p(t)\cdot\big[X^1,\big[X^1,X^2\big]\big]\,\big(x(t)\big)& \\ +\,u_2(t)\,p(t)\cdot\big[X^2,\big[X^1,X^2\big]\big]\,\big(x(t)\big)&=0. \end{aligned} \tag{1.36}$$

Since M has dimension four and $[\Delta,[\Delta,\Delta]]$ has dimension three, there is (locally) a smooth non-vanishing 1-form α such that

$$\alpha_x\cdot v=0 \qquad \forall v\in\Delta(x)+[\Delta,\Delta](x),\ \forall x.$$

The, by (1.22) and (1.35)-(1.36), we infer that for almost every $t\in[0,T]$ such that $u(t)\neq 0$,

$$u_1(t)\,\alpha_{x(t)}\cdot\big[X^1,\big[X^1,X^2\big]\big]\,\big(x(t)\big)+u_2(t)\,\alpha_{x(t)}\ \cdot\big[X^2,\big[X^1,X^2\big]\big]\,\big(x(t)\big)=0.$$

By the above assumptions, for every x, the linear form

$$(\lambda_1, \lambda_2) \in \mathbb{R}^2 \longmapsto \Big(\alpha_x \cdot \big[X^1, \big[X^1, X^2\big]\big](x)\Big)\lambda_1 + \Big(\alpha_x \cdot \big[X^2, \big[X^1, X^2\big]\big](x)\Big)\lambda_2$$

has a kernel of dimension one. This shows that there is a smooth line field (a distribution of rank one) $L \subset \Delta$ on M such that the singular curves are exactly the integral curves of L.

1.4 The Chow-Rashevsky Theorem

Openness of End-Point mappings

The following result will imply easily the Chow-Rashevsky Theorem. We recall that a map is said to be open if the image of any open set is open.

Proposition 1.4.1. *Let $\mathcal{F} = \{X^1, \cdots, X^k\}$ be a family of smooth vector fields on M satisfying the Hörmander condition on M. Then for every $x \in M$ and every $T > 0$, the End-Point mapping $E_{\mathcal{F}}^{x,T} : U_{\mathcal{F}}^{x,T} \to M$ is open.*

Proof. Let $x \in M$ and $T > 0$ be fixed. Set for every $\epsilon > 0$,

$$d(\epsilon) = \max\Big\{\mathrm{rank}_{\mathcal{F}}^{x,\epsilon}(u) \,|\, u \in U_{\mathcal{F}}^{x,\epsilon} \text{ s.t. } \|u\|_{L^2} < \epsilon\Big\}.$$

By Proposition 1.2.8, the function $\epsilon \in (0, +\infty) \mapsto d(\epsilon)$ is nondecreasing with values in $\mathbb{N}$. So, there is ϵ_0 and $d_0 \in \mathbb{N}$ such that $d(\epsilon) = d_0$ for any $\epsilon \in (0, \epsilon_0)$. Since $\mathcal{F}$ satisfies the Hörmander condition at x, the space $\mathrm{Span}\{X^1(x), \ldots, X^k(x)\}$ has dimension ≥ 1. Then, thanks to Proposition 1.2.10, there holds

$$d(\epsilon) = d_0 \geq 1 \qquad \forall \epsilon \in [0, \epsilon_0].$$

Let $\epsilon \in (0, \epsilon_0)$ and $u^\epsilon \in U_{\mathcal{F}}^{x,\epsilon}$ such that $\|u^\epsilon\|_{L^2} < \epsilon$ and $\mathrm{rank}_{\mathcal{F}}^{x,\epsilon}(u^\epsilon) = d_0$ be fixed. There are d_0 controls $v^1, \cdots, v^{d_0}$ in $L^2([0,\epsilon]; \mathbb{R}^k)$ such that the linear map

$$\begin{array}{rcl} \mathcal{L} : \mathbb{R}^{d_0} & \longrightarrow & T_x M \\ \lambda = \left(\lambda^1, \cdots, \lambda^{d_0}\right) & \longmapsto & D_{u^\epsilon} E_{\mathcal{F}}^{x,\epsilon}\left(\sum_{j=1}^{d_0} \lambda^j v^j\right) \end{array}$$

is injective. By construction and the fact that the mapping $u \mapsto \mathrm{rank}_{\mathcal{F}}^{x,\epsilon}(u)$ is lower semicontinuous, the rank of any control u is equal to $\mathrm{rank}_{\mathcal{F}}^{x,\epsilon}(u^\epsilon) = d_0$ as soon as u is close enough to u^ϵ in $L^2([0,\epsilon]; \mathbb{R}^k)$. Hence, there is an open neighborhood $\mathcal{V}$ of $0 \in \mathbb{R}^{d_0}$ where the mapping

$$\begin{array}{rcl} \mathcal{E} : \mathcal{V} & \longrightarrow & M \\ \lambda & \longmapsto & E_{\mathcal{F}}^{x,\epsilon}\left(u^\epsilon + \sum_{j=1}^{d_0} \lambda^j v^j\right) \end{array}$$

is an embedding, whose the image is a submanifold N of class C^1 in M of dimension d_0. Moreover by construction again, there holds for every small $\lambda \in \mathbb{R}^{d_0}$,

$$\mathrm{Im}_{\mathcal{F}}^{x,\epsilon}\left(u^\epsilon + \sum_{j=1}^{d_0} \lambda^j v^j\right) = D_\lambda \mathcal{E}\big(\mathbb{R}^{d_0}\big) = T_{\mathcal{E}(\lambda)} N.$$

By Proposition 1.2.10, we infer that

$$X^i(y) \in T_y N \qquad \forall i = 1, \ldots, k, \; \forall y \in N.$$

Lemma 1.4.2. *Let Ω be an open subset of $\mathbb{R}^l$ ($l \geq 2$) and $\mathcal{S}$ be a submanifold of Ω of class C^1. Let X, Y be two smooth vector fields on Ω such that*

$$X(x),\, Y(x) \in T_x\mathcal{S} \qquad \forall x \in \mathcal{S}.$$

Then $[X,Y](x) \in T_x\mathcal{S}$ for any $x \in \mathcal{S}$.

Proof of Lemma 1.4.2. As in Proposition 1.1.10, we denote respectively by e^{tX} and e^{tY} the flows of X and Y. Since by assumption X and Y is always tangent to $\mathcal{S}$, we have for every $x \in \mathcal{S}$,

$$e^{tX}(x),\, e^{tY}(x) \in \mathcal{S} \qquad \forall t \text{ small.}$$

Then

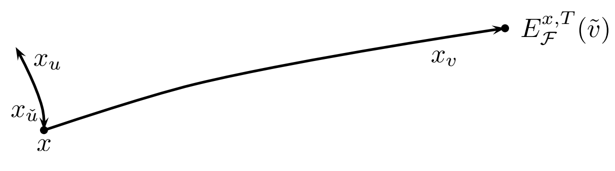
$$\left(e^{-tY} \circ e^{-tX} \circ e^{tY} \circ e^{tX}\right)(x) \in \mathcal{S} \qquad \forall x \in \mathcal{S} \text{ and } t \text{ small.}$$

By Proposition 1.1.10, we infer that $[X,Y](x) \in T_x\mathcal{S}$ for any $x \in \mathcal{S}$. $\square$

It follows from the above lemma that all the brackets involving $X^1, \ldots, X^k$ at $y \in N$ belong to T_yN. Since $\mathcal{F}$ satisfies the Hörmander condition, this shows that $d_0 = n$ and indeed that $d(\epsilon) = n$ for any $\epsilon > 0$.
Let $\mathcal{O}$ be an open subset of $U_{\mathcal{F}}^{x,T}$, $v \in \mathcal{O}$ and $\epsilon > 0$ to be chosen later. Since $d(\epsilon) = n$, there is $u \in U_{\mathcal{F}}^{x,\epsilon}$ such that $\|u\|_{L^2} < \epsilon$ and $\mathrm{rank}_{\mathcal{F}}^{x,\epsilon}(u) = n$. Define the control $\tilde{v} \in L^2([0,T];\mathbb{R}^k)$ by

$$\tilde{v} = u * \breve{u} * v_{\frac{T}{T-2\epsilon}}.$$

The trajectory associated with $\tilde{v}$ is the concatenation of the curve x_u starting at x and associated with u, $x_{\breve{u}}$ starting at $x_u(\epsilon)$ and associated with $\breve{u}$, and a reparametrization of x_v starting at x and associated with v.



By construction (thanks to Propositions 1.3.2), $\tilde{v}$ belongs to $U_{\mathcal{F}}^{x,T}$, is regular and satisfies $E_{\mathcal{F}}^{x,T}(\tilde{v}) = E_{\mathcal{F}}^{x,T}(v)$. Then, as above, the image of a small ball centered at the origin in $L^2([0,T];\mathbb{R}^k)$ by the mapping $w \mapsto E_{\mathcal{F}}^{x,T}(\tilde{v} + w)$ is an open neighboorhood of $E_{\mathcal{F}}^{x,T}(\tilde{v}) = E_{\mathcal{F}}^{x,T}(v)$. Furthermore, we have

$$\begin{aligned}
\left\|\tilde{v}-v\right\|_{L^2}^2 &= \int_0^{\epsilon} |u(t)-v(t)|^2\,dt + \int_{\epsilon}^{2\epsilon} |-u(2\epsilon-t)-v(t)|^2\,dt \\
&\quad + \int_{2\epsilon}^{T} \left| \left(\frac{T}{T-2\epsilon}\right) v\left(\frac{T\big(t-2\epsilon\big)}{T-2\epsilon}\right) - v(t)\right|^2 dt \\
&\leq 2\int_0^{\epsilon} |u(t)|^2\,dt + \int_0^{2\epsilon} |v(t)|^2\,dt \\
&\quad -2\int_0^{\epsilon} \langle u(t), v(t)\rangle\,dt + 2\int_{\epsilon}^{2\epsilon} \langle u(2\epsilon-t), v(t)\rangle\,dt \\
&\quad + \int_{2\epsilon}^{T} \left| \left(\frac{T}{T-2\epsilon}\right) u(t) - u(t)\right|^2 dt \\
&\quad + \left(\frac{T}{T-2\epsilon}\right)^2 \int_{2\epsilon}^{T} \left| u\left(\frac{T\big(t-2\epsilon\big)}{T-2\epsilon}\right) - u(t)\right|^2 dt \\
&\quad +2\left(\frac{T}{T-2\epsilon}\right) \int_{2\epsilon}^{T} \Bigg\langle u\left(\frac{T\big(t-2\epsilon\big)}{T-2\epsilon}\right) - u(t), \\
&\qquad\qquad \left(\frac{T}{T-2\epsilon}\right) u(t) - u(t) \Bigg\rangle\,dt.
\end{aligned}$$

Then since $\|u\|_{L^2} < \epsilon$ and both functions

$$t \in [2\epsilon, T] \quad \longmapsto \quad \left(\frac{T}{T-2\epsilon}\right) u(t) - u(t)$$

$$\text{and} \quad t \in [2\epsilon, T] \quad \longmapsto \quad u\left(\frac{T\big(t-2\epsilon\big)}{T-2\epsilon}\right) - u(t)$$

tend to zero in L^2, we infer that $\tilde{v}$ belong to $\mathcal{O}$ if ϵ is small enough. This shows that $E_{\mathcal{F}}^{x,T}(\mathcal{O})$ contains a neighborhood of $E_{\mathcal{F}}^{x,T}(\tilde{v}) = E_{\mathcal{F}}^{x,T}(v)$. □

Statement and proof

The aim of the present section is to prove the following result.

Theorem 1.4.3 (Chow-Rashevsky's Theorem)**.** *Let Δ be a totally nonholonomic distribution on M (assumed to be connected). Then, for every $x, y \in M$ and every $T > 0$, there is an horizontal path $\gamma \in \Omega_{\Delta}^{x,T}$ such that $\gamma(T) = y$.*

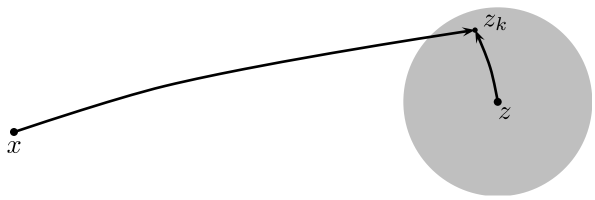
Thanks to the above discussion, the Chow-Rashevsky Theorem will be a straightforward consequence of the following result.

Theorem 1.4.4. *Let $\mathcal{F} = \left\{X^1, \cdots, X^k\right\}$ be a family of smooth vector fields on M. Assume that M is connected and that $\mathcal{F}$ satisfies the Hörmander condition on M. Then for every $x, y \in M$ and every $T > 0$, there is a control $u \in U_{\mathcal{F}}^{x,T}$ such that the solution of (1.7) satisfies $\gamma_u(T) = y$.*

Proof. Let x and $T > 0$ be fixed. Denote by $\mathcal{A}_{\mathcal{F}}(x,T)$ the set of points in M which can be joined from x by a control in $U_{\mathcal{F}}^{x,T}$, that is

$$\mathcal{A}_{\mathcal{F}}(x,T) = E_{\mathcal{F}}^{x,T}\left(U_{\mathcal{F}}^{x,T}\right).$$

By Proposition 1.4.1, $\mathcal{A}_{\mathcal{F}}(x,T)$ is an open set in M. Let us show that this set is closed as well. Let $\{z_k\}_k$ be a sequence of points in M converging to some $z \in M$. By openness of the mapping $E_{\mathcal{F}}^{z,1}$ and the fact that $E_{\mathcal{F}}^{z,1}(0) = z$, the set $E_{\mathcal{F}}^{z,1}\left(U_{\mathcal{F}}^{z,1}\right)$ is an neighborhood of z. Then, there is k large enough such that z_k belongs to that set.



The concatenation of u_k together with $\check{u}$ steers x to z. This shows that $\mathcal{A}_{\mathcal{F}}(x,T)$ is closed in M. In conclusion $\mathcal{A}_{\mathcal{F}}(x,T)$ is open, closed and nonempty (it contains x). By connectedness of M, we infer that $\mathcal{A}_{\mathcal{F}}(x,T) = M$. $\square$

Remark 1.4.5. *The Chow-Rashevsky may be of course obtained in different ways. For instance, consider in $\mathbb{R}^3$ a totally nonholonomic rank two distribution Δ generated by two smooth vector fields X^1, X^2 such that*

$$\mathit{Span}\Big\{X^1(x), X^2(x), \big[X^1, X^2\big](x)\Big\} = \mathbb{R}^3 \qquad \forall x \in \mathbb{R}^3.$$

Let $x \in \mathbb{R}^3$ and $\lambda > 0$ be fixed, define the function $\Phi_\lambda : \mathbb{R}^3 \to \mathbb{R}^3$ by

$$\Phi_\lambda\big(t_1, t_2, t_3\big) := \left(e^{\lambda X^1} \circ e^{t_3 X^2} \circ e^{-\lambda X^1} \circ e^{t_2 X^2} \circ e^{t_1 X^1}\right)(x),$$

for every $(t_1, t_2, t_3) \in \mathbb{R}^3$. It can be shown that Φ_λ is a local diffeomorphism in a neighborhood of the origin provided λ is small enough. This implies easily the Chow-Rashevsky Theorem for contact distributions in dimension three.

On the set of regular controls

The proof of Proposition 1.4.1 implies indeed the following result.

Proposition 1.4.6. *Let $\mathcal{F} = \big\{X^1, \cdots, X^k\big\}$ be a family of smooth vector fields on M satisfying the Hörmander condition on M (assumed to be connected). Then for every $x \in M$ and $T > 0$, the set of regular controls $\mathcal{R}_{\mathcal{F}}^{x,T} \subset U_{\mathcal{F}}^{x,T}$ is open and dense in $U_{\mathcal{F}}^{x,T}$. Moreover, for every $y \in M$, there is a smooth control $u \in \mathcal{R}_{\mathcal{F}}^{x,T}$ such that the solution of (1.7) satisfies $\gamma_u(T) = y$.*

Proof. The density part follows from the last part of the proof of Proposition 1.4.1. The openness of $\mathcal{R}_{\mathcal{F}}^{x,T}$ is a straightforward consequence of the C^1 regularity of $E_{\mathcal{F}}^{x,T}$ and the fact that regular controls are nothing but regular points of $E_{\mathcal{F}}^{x,T}$. Let $x, y \in M$ and $T > 0$ be fixed. Let us show that x can be steered to y by a smooth regular control. From Theorem 1.4.4, we know that there is $u \in U_{\mathcal{F}}^{x,T}$ such that the solution of (1.7) satisfies $\gamma_u(T) = y$. Replacing u by a control of the form

$$v * \check{v} * u_{\frac{T}{T-2\epsilon}}$$

with $v \in \mathcal{R}_{\mathcal{F}}^{x,\epsilon}$ and $\epsilon > 0$ small enough, we may assume that u is indeed a regular control joining x to y. Therefore, there are n controls $v^1, \cdots, v^n$ in $L^2([0,T];\mathbb{R}^k)$ such that the linear map

$$\begin{array}{rcl} \mathcal{L} : \mathbb{R}^n & \longrightarrow & T_xM \\ \lambda = \left(\lambda^1, \cdots, \lambda^n\right) & \longmapsto & D_u E_{\mathcal{F}}^{x,T}\left(\sum_{j=1}^n \lambda^j v^j\right) \end{array}$$

is bijective. In fact, since the set of smooth functions is dense in $L^2\left([0,T];\mathbb{R}^k\right)$, we may assume that all the v^j's are smooth. Therefore, by the Inverse Function Theorem (and up to take a set of local coordinates), there is an open neighborhood $\mathcal{V}$ of $0 \in \mathbb{R}^n$ such that the mapping

$$\begin{array}{rcl} \mathcal{E}_u : \mathcal{V} & \longrightarrow & M \\ \lambda & \longmapsto & E_{\mathcal{F}}^{x,T}\left(u + \sum_{j=1}^n \lambda^j v^j\right) \end{array}$$

is a diffeomorphism from $\mathcal{V}$ to $\mathcal{E}(\mathcal{V})$. The open set $\mathcal{E}(\mathcal{V})$ contains a ball centered at $\mathcal{E}_u(0) = E_{\mathcal{F}}^{x,T}(u) = y$ whose the radius $r > 0$ is controlled by $D_0\mathcal{E}_u$ and the local Lipschitz constant of $\lambda \mapsto D_\lambda \mathcal{E}_u$ (see Theorem B.1.4). Therefore, there is $\nu > 0$ such that if $\tilde{u} \in U_{\mathcal{F}}^{x,T}$ is such that $\|\tilde{u} - u\|_{L^2} < \nu$, then the image of the mapping

$$\begin{array}{rcl} \mathcal{E}_{\tilde{u}} : \mathcal{V} & \longrightarrow & M \\ \lambda & \longmapsto & E_{\mathcal{F}}^{x,T}\left(\tilde{u} + \sum_{j=1}^n \lambda^j v^j\right) \end{array}$$

contains a ball centered at $\mathcal{E}_{\tilde{u}}(0)$ with radius $r/2$. Taking $\tilde{u}$ smooth with $\|\tilde{u} - u\|_{L^2}$ small enough implies that

$$|\mathcal{E}_{\tilde{u}}(0) - y| = |\mathcal{E}_{\tilde{u}}(0) - \mathcal{E}_u(0)| < \frac{r}{2},$$

which shows that y can be steered from x by a smooth control. □

Given a totally nonholonomic distribution Δ on M, $x \in M$ and $T > 0$, we denote by $\mathcal{R}_{\Delta}^{x,T}$ the set of regular horizontal paths in $\Omega_{\Delta}^{x,T}$. The following result is an immediate corollary of Proposition 1.4.6.

Theorem 1.4.7. *Let Δ be a totally nonhonomic distribution on M (assumed to be connected). Then, for every $x, y \in M$ and every $T > 0$, there is a smooth path $\gamma \in \mathcal{R}_{\Delta}^{x,T}$ such that $\gamma(T) = y$.*

1.5 Sub-Riemannian structures

Definition

Let M be a smooth connected manifold of dimension n. A *sub-Riemannian structure* on M is given by a pair (Δ, g) where Δ is a totally nonholonomic distribution on M and g is a smooth Riemannian metric on Δ, that is for every $x \in M$, $g(\cdot,\cdot)$ is a scalar product on Δ_x. A simple way to construct a sub-Riemannian structure is to take a smooth connected Riemannian manifold (M,g), to consider a totally nonholonomic distribution Δ on M, and to take as sub-Riemannian metric the restriction of g to the distribution. In fact, any sub-Riemannian structure can be obtained in this way.

Example 1.5.1. *The space $\mathbb{R}^3$ (with coordinates (x,y,z)) equipped with the rank two distribution Δ given in Example 1.1.2 and with the metric $g = dx^2 + dy^2$ is the most simple sub-Riemannian structure we can imagine. The length of a vector $v = (v_1, v_2, v_3) \in \Delta(x,y,z)$ is given by*

$$|v|^g = \sqrt{v_1^2 + v_2^2}.$$

Since v is an horizontal vector, the latter quantity does not vanishes unless $v = 0$.

If the distribution Δ admits a frame $X^1, \ldots, X^m$ on an open set $\mathcal{O} \subset M$, then the family $\mathcal{F} = \left\{X^1, \ldots, X^m\right\}$ is called an *orthonormal family of vector fields* or an *orthonormal frame* for (Δ, g) in $\mathcal{O}$ if there holds

$$g_x\left(X^i(x), X^j(x)\right) = \delta_{ij} \qquad \forall i,j = 1,\ldots,m, \quad \forall x \in \mathcal{O},$$

where δ_{ij} denotes the Kronecker symbol (that is $\delta_{ij} = 1$ if $i = j$ and $\delta_{ij} = 0$ if $i \neq j$). Sub-Riemannian structures admit local orthonormal frames in a neighborhood of each point of M.

The sub-Riemannian distance

From now on, for every $x \in M$ we denote by $|\cdot|_x^g$ the sub-Riemannian norm on $\Delta(x)$, that is

$$|v|_x^g = \sqrt{g_x(v,v)} \qquad \forall v \in \Delta(x).$$

The *length* of an horizontal path $\gamma \in \Omega_\Delta^{x,T}$ is defined by

$$\text{length}^g(\gamma) := \int_0^T \left|\dot\gamma(t)\right|_{\gamma(t)}^g dt. \tag{1.37}$$

Note that since any horizontal path is absolutely continuous with square integrable derivative, the length of any horizontal path is finite.

Thanks to the Chow-Rashevsky Theorem, for every $x,y \in M$, there is at least one horizontal path joining x to y in time 1. For every $x,y \in M$, the *sub-Riemannian distance* between x and y, denoted by $d_{SR}(x,y)$, is defined as the infimum of lengths of horizontal paths joining x to y, that is,

$$d_{SR}(x,y) := \inf\Big\{\text{length}^g(\gamma) \mid \gamma \in \Omega_\Delta^{x,1} \text{ s.t. } \gamma(1) = y\Big\}.$$

The function d_{SR} defines a distance on $M \times M$ (the triangular inequality is easy, the fact that $d_{SR}(x,y) \Rightarrow x = y$ follows from the proof of Proposition 1.5.2) and makes M a metric space. Given $x \in M$ and $r \geq 0$, we call *sub-Riemannian ball* centered at x with radius r the set defined as

$$B_{SR}(x,r) = \Big\{y \in M \,|\, d_{SR}(x,y) < r\Big\}.$$

In fact, the Chow-Rashevsky Theorem yields the following result.

Proposition 1.5.2. *Let (Δ, g) be a sub-Riemannian structure on M, then the topology defined by d_{SR} coincides with the original topology of M. In particular, the sub-Riemannian distance d_{SR} is continuous on $M \times M$.*

Proof. We need to show that for every $x \in M$, the family of sub-Riemannian balls $\{B_{SR}(x,r)\}_{r>0}$ is a basis of neighborhoods for x with respect to the original topology. Let $\mathcal{F} = \{X^1, \ldots, X^m\}$ be an orthonormal frame for Δ on an open neighborhood $\mathcal{V}_x$ of some $x \in M$ Let $\mathcal{V} \subset \mathcal{V}_x$ be an open and relatively compact neighborhood of x with respect to the initial topology. Let us show that there is $r > 0$ small enough such that $B_{SR}(x,r) \subset \mathcal{V}$. Let $\mathcal{W}$ be an open neighborhood of x such that $\overline{\mathcal{W}} \subset \mathcal{V}$. Define the compact annulus $\mathcal{A}$ by

$$\mathcal{A} = \overline{\mathcal{V}} \setminus \mathcal{W}.$$

Any continuous path joining x to a point outside $\mathcal{V}$ has to cross $\mathcal{A}$. Hence since $X^1, \ldots, X^m$ are bounded on $\mathcal{A}$, there is $\delta > 0$ such that any solution $\gamma_u : [0,1] \to M$ to the Cauchy problem

$$\dot{\gamma}_u(t) = \sum_{i=1}^m u_i(t) X^i\left(\gamma_u(t)\right) \quad \text{a.e. } t \in [0,1], \qquad \gamma_u(0) = x, \tag{1.38}$$

with $u \in U_{\mathcal{F}}^{x,1}$ and $\gamma_u(1) \notin \mathcal{V}$ satisfies

$$\int_0^1 \left| \sum_{i=1}^m u_i(t) X^i\left(\gamma_u(t)\right) \right|_{\gamma_u(t)}^g dt > \delta.$$

By Proposition 1.2.1, this means that the sub-Riemannian ball $B_{SR}(x, \delta/2)$ is included in $\mathcal{V}$. Let us now show that any sub-Riemannian ball $B_{SR}(x,r)$ contains an open neighborhood of x with respect to the initial topology. The set $U_{\mathcal{F}}^{x,1}$ is open in $L^2\big([0,1];\mathbb{R}^m\big)$ and contains the control $u \equiv 0$. Thus there is $\nu > 0$ such that the L^2-ball $B_{L^2}(0,\nu)$ is contained in $U_{\mathcal{F}}^{x,1}$. Moreover, since $\mathcal{F}$ is orthonormal with respect to g, there holds for every $u \in B_{L^2}(0,\nu)$

$$\text{length}^g(\gamma_u) = \int_0^1 |u(t)| dt.$$

Thanks to the Cauchy-Schwarz inequality, we infer that

$$E_{\mathcal{F}}^{x,1}\left(B_{L^2}(0,\nu)\right) \subset B_{SR}(x,\nu).$$

Proposition 1.4.1 together with $E_{\mathcal{F}}^{x,1} = x$ concludes the proof. □

Thanks to the above result, the sub-Riemannian balls $B_{SR}(x,r)$ are always open with respect to the original topology of M and the closed sub-Riemannian balls $B_{SR}(x,r)$ defined as

$$\bar{B}_{SR}(x,r) = \Big\{ y \in M \,|\, d_{SR}(x,y) \leq r \Big\}$$

are always closed sets.

1.6 Notes and comments

Proposition 1.1.8 is taken from a paper by Sussmann [Sus10]; we note that it could also be proven by transversality arguments (see [GG73]).

The Hörmander condition introduced in Section 1.1 is also refered as bracket generating condition. The term comes from the analysis literature; it is named after Hörmander who obtained hypoellipticity results for linear operators associated with families of vector fields [Hör67]. Several other terms may be used to refer to totally nonholonomic distributions. They are called bracket generating by Montgomery [Mon02], nonholonomic by Bellaiche [Bel96], and they refer to completely nonholonomic families of vector fields by Agrachev and Sachkov [AS04].

The notion of singular curves play a major role in this monograph. Most of the examples of singular horizontal paths given in Section 1.3 are classical. The most valuable (Example 1.3.17) is taken from [Sus96].

Theorem 1.4.3 has been proved independently by Chow [Cho39] and Rashevsky [Ras38] in the 30s, see [Cho39, Ras38]. The proof that we present here is an adaptation of the one given by Bellaiche [Bel96] to prove the so-called Orbit Theorem (see also [AS04, Jur97]). Other proofs of the Chow-Rashevsky Theorem can be found in the texts of Bismut [Bis84], Gromov [Gro96], or Montgomery [Mon02].

Chapter 2

Sub-Riemannian geodesics

Throughout all the chapter, M denotes a smooth connected manifold without boundary of dimension $n \geq 2$ equipped with a sub-Riemannian structure (Δ, g) of rank $m \leq n$.

2.1 Minimizing horizontal paths and geodesics

Definitions

Given $x, y \in M$, we call *minimizing horizontal path* between x and y any path $\gamma \in \Omega_\Delta^{x,T}$ with $T \geq 0$ such that

$$d_{SR}(x,y) = \text{length}^g(\gamma).$$

Like in the Riemannian case, minimizing paths with constant speed minimize the so-called sub-Riemannian energy. Given $x, y \in M$, we define the *sub-Riemannian energy* between x and y by

$$e_{SR}(x,y) := \inf\Big\{\text{energy}^g(\gamma) \mid \gamma \in \Omega_\Delta^{x,1} \text{ s.t. } \gamma(1) = y\Big\},$$

where the *energy* of a path $\gamma \in \Omega_\Delta^{x,1}$ is defined as

$$\text{energy}^g(\gamma) := \int_0^T \left(|\dot\gamma(t)|^g_{\gamma(t)}\right)^2 dt. \tag{2.1}$$

The following result whose the proof is based on Cauchy-Schwarz's inequality, is fundamental.

Proposition 2.1.1. *For any $x, y \in M$, $e_{SR}(x,y) = d_{SR}(x,y)^2$.*

Proof. Let $x, y \in M$ be fixed. First, we observe that, for every horizontal path $\gamma : [0,1] \to M$ satisfying $\gamma(0) = x$ and $\gamma(1) = y$, the Cauchy-Schwarz inequality yields

$$\left(\int_0^1 |\dot\gamma(t)|^g_{\gamma(t)}\, dt\right)^2 \leq \int_0^1 \left(|\dot\gamma(t)|^g_{\gamma(t)}\right)^2 dt. \tag{2.2}$$

Taking the infimum over the set of $\gamma \in \Omega_\Delta^{x,1}$ such that $\gamma(1) = y$ yields $d_{SR}(x,y)^2 \leq e_{SR}(x,y)$. On the other hand, for every $\epsilon > 0$, there exists an horizontal path $\gamma \in \Omega_\Delta^{x,1}$, with $\gamma(1) = y$, such that

$$\text{length}^g(\gamma) = \int_0^1 |\dot\gamma(t)|^g_{\gamma(t)}\, dt \leq d_{SR}(x,y) + \epsilon.$$

Reparametrizing γ by arc-length, we get a new path $\xi \in \Omega_\Delta^{x,1}$ with $\gamma(1) = y$ satisfying

$$\left|\dot\xi(t)\right|^g_{\xi(t)} = \text{length}^g(\gamma) \qquad \text{a.e. } t \in [0,1].$$

Consequently,

$$e_{SR}(x,y) \leq \int_0^1 \left(\left|\dot\xi(t)\right|^g_{\xi(t)} \right)^2 dt = \text{length}^g(\gamma)^2 \leq \left(d_{SR}(x,y) + \epsilon\right)^2.$$

Letting ϵ tend to 0 completes the proof of the result. □

Given $x, y \in M$, we call *minimizing geodesic* between x and y any path $\gamma \in \Omega_\Delta^{x,1}$ joining x to y such that

$$e_{SR}(x,y) = \text{energy}^g(\gamma).$$

Thanks to the above proof and the fact that equality holds in the Cauchy-Schwarz inequality (2.2) if and only γ has constant speed (that is $|\dot\gamma(t)|^g_{\gamma(t)}$ is constant), we obtain the following result.

Proposition 2.1.2. *Given $x, y \in M$, a path $\gamma \in \Omega_\Delta^{x,1}$ is a minimizing geodesic between x and y if and only if it is a minimizing horizontal path between x and y with constant speed.*

Sufficiently near points can be joined by minimizing geodesics and a fortiori by minimizing horizontal paths.

Proposition 2.1.3. *Let $x \in M$, then there is $\rho > 0$ such that the following property is satisfied:*
For every $y, z \in B_{SR}(x,\rho)$ and any minimizing sequence $\{\gamma^k\}_k : [0,1] \to M$ of horizontal paths with constant speed such that

$$\lim_{k\to+\infty} \gamma^k(0) = y, \quad \lim_{k\to+\infty} \gamma^k(1) = z, \quad \lim_{k\to+\infty} \mathit{length}^g\left(\gamma^k\right) = d_{SR}(y,z), \quad (2.3)$$

up to taking a subsequence, $\{\gamma^k\}_k$ converges uniformly to some minimizing geodesic $\bar\gamma \in \Omega_\Delta^{y,1}$ joining y to z.
In particular, for every $y, z \in B_{SR}(x,\rho)$, there is a minimizing geodesic between y and z.

Proof. Fix $x \in M$ and $\mathcal{F} = \{X^1, \ldots, X^m\}$ an orthonormal frame for Δ on an open and relatively compact neighborhood $\mathcal{V}_x$ of x. From Proposition 1.5.2, there is $r > 0$ small enough such that $B_{SR}(x,r) \subset \mathcal{V}_x$. For any $y \in B_{SR}(x, r/4)$ and any horizontal path $\gamma : [0,1] \to M$ with constant speed satisfying

$$d_{SR}\left(\gamma(0), y\right) < \frac{r}{24} \quad \text{and} \quad \text{length}^g(\gamma) \leq \frac{2r}{3},$$

we have for every $t \in [0,1]$,

$$\begin{aligned} d_{SR}\big(x,\gamma(t)\big) &\leq d_{SR}(x,y) + d_{SR}\big(y,\gamma(t)\big) \\ &\leq r/4 + d_{SR}\big(y,\gamma(0)\big) + \operatorname{length}^g(\gamma) \\ &\leq r/4 + r/24 + 2r/3 = 23r/24 < r. \end{aligned}$$

Which means that γ is contained in $B_{SR}(x,r)$. Furthermore, for every such horizontal path, there is $u \in L^2\big([0,1];\mathbb{R}^m\big)$ such that

$$\dot\gamma(t) = \sum_{i=1}^m u_i(t) X^i\big(\gamma(t)\big) \quad \text{and} \quad |\dot\gamma(t)|^g_{\gamma(t)} = \|u\|_{L^2} = \operatorname{length}^g(\gamma),$$

for a.e. $t \in [0,1]$. Let $y,z \in B_{SR}(x,r/4)$ be fixed and $\{\gamma^k\}_k : [0,1] \to M$ be a sequence of horizontal paths with constant speed verifying (2.3). By the above discussion, we may assume without loss of generality that all the paths $\gamma^k : [0,1] \to M$ are valued in the compact set $\overline{\mathcal{V}_x}$ with derivatives bounded by r and associated with a sequence of controls $\{u^k\}_k$ in $L^2\big([0,1];\mathbb{R}^m\big)$ such that $\|u^k\|_{L^2} = \operatorname{length}^g(\gamma^k)$. Then by Arzela-Ascoli's theorem taking a subsequence if necessary the sequence $\{\gamma^k\}_k$ converges to some $\bar\gamma : [0,1] \to M$. Moreover, the sequence $\{u^k\}_k$ is bounded in L^2 so it weakly converges up to a subsequence to some $\bar v \in L^2\big([0,1];\mathbb{R}^m\big)$. We obtain easily that $\bar\gamma(0) = y, \bar\gamma(1) = z$,

$$\dot{\bar\gamma}(t) = \sum_{i=1}^m \bar v_i(t) X^i\big(\bar\gamma(t)\big) \qquad \text{a.e. } t \in [0,1],$$

and by lower semicontinuity of the L^2-norm under weak convergence we immediately deduce that

$$\|\bar v\|_{L^2} \leq \lim_k \left\|u^k\right\|_{L^2} = d_{SR}(y,z).$$

Furthermore, since $\bar\gamma$ is an horizontal path joining y to z, there holds

$$d_{SR}(y,z) \leq \operatorname{length}^g(\bar\gamma).$$

By Cauchy-Schwarz's inequality, we have $\operatorname{length}^g(\bar\gamma) \leq \|\bar v\|_{L^2}$. Then we infer that

$$\operatorname{energy}^g\big(\bar\gamma\big) = \|\bar v\|^2_{L^2} = d_{SR}(y,z)^2 = e_{SR}(y,z).$$

Which shows that $\bar\gamma$ is a minimizing geodesic joining y to z. □

Remark 2.1.4. *The above proof shows indeed that up to taking a subsequence, the sequence $\{u^k\}_k$ converges strongly to $\bar v$ in $L^2([0,1];\mathbb{R}^m)$. As a matter of fact, it converges weakly to $\bar v$ and satisfies*

$$\lim_k \left\|u^k\right\|_{L^2} = \|\bar v\|_{L^2}.$$

The SR Hopf-Rinow Theorem

The following sub-Riemannian version of the classical Riemannian Hopf-Rinow Theorem holds.

Theorem 2.1.5 (Hopf-Rinow Theorem)**.** *Let (Δ, g) be a sub-Riemannian structure on M. Assume that (M, d_{SR}) is a complete metric space. Then the following properties hold:*

(i) The balls $\bar{B}_{SR}(x,r)$ are compact,

(ii) for every $x, y \in M$ there exists at least one minimizing geodesic joining x to y.

Proof. Let us first recall that thanks to Proposition 1.5.2, the metric space (M, d_{SR}) is locally compact. That is for every $x \in M$, there is $r > 0$ such that the ball $\bar{B}_{SR}(x,r)$ is compact. Let $x \in M$ be fixed. We first show that all the balls $\bar{B}_{SR}(x,r)$ with $r \geq 0$ are compact. Denote by I_x the set of $r \geq 0$ such that $\bar{B}_{SR}(x,r)$ is compact. By inclusion of the balls $\bar{B}_{SR}(x,r') \subset \bar{B}_{SR}(x,r)$ if $r' \leq r$ and local compactness of (M, d_{SR}), I_x is an interval whose the supremum R_x is strictly positive. We claim that I is both closed and open in $[0, +\infty)$.

Lemma 2.1.6. *The interval I_x is closed in* $[0, +\infty)$.

Proof of Lemma 2.1.6. We need to show that R_x belongs to I_x, that is that $\bar{B}_{SR}(x, R_x)$ is compact. Let $\{y_k\}_k$ be a sequence of points in $\bar{B}_{SR}(x, R_x)$, we need to show that it has a convergent subsequence. We construct a Cauchy subsequence of $\{y_k\}_k$ as follows. For every integer $l \geq 1$, we set

$$K^l = \bar{B}_{SR}\left(x, R_x(1 - 2^{-l})\right).$$

By assumption, $\{K^l\}_l$ is an increasing sequence of compact sets in $\bar{B}_{SR}(x, R_x)$. For every $k \in \mathbb{N}$, there is $y_k^1 \in K^1$ such that

$$d_{SR}\left(y_k, y_k^1\right) = \inf\left\{ d_{SR}(y_k, z) \,|\, z \in K^1 \right\} \leq \frac{R_x}{2}.$$

By compactness of K^1, there is a strictly increasing mapping $\varphi^1 : \mathbb{N} \to \mathbb{N}$ such that the sequence $\{y^1_{\varphi^1(k)}\}_k$ converges to some $\bar{y}^1 \in K^1$. Thus there exists $k_1 \geq 0$ such that

$$d_{SR}\left(y^1_{\varphi^1(k)}, \bar{y}^1\right) \leq \frac{R_x}{2} \qquad \forall k \geq k_1.$$

Set $z_1 := y_{\varphi^1(k_1)}$. Now for every $k \in \mathbb{N}$, there is $y_k^2 \in K^2$ such that

$$d_{SR}\left(y_{\varphi^1(k)}, y_k^2\right) = \inf\left\{ d_{SR}(y_{\varphi^1(k)}, z) \,|\, z \in K^2 \right\} \leq \frac{R_x}{4}.$$

Again, by compactness of K^2 there exists a strictly increasing mapping $\varphi^2 : \mathbb{N} \to \mathbb{N}$ such that the sequence $\{y^2_{\varphi^2(k)}\}_k$ converges to some $\bar{y}^2 \in K^2$ and then there is $k_2 \geq k_1$ such that

$$d_{SR}\left(y^2_{\varphi^2(k)}, \bar{y}^2\right) \leq \frac{R_x}{4} \qquad \forall k \geq k_2.$$

Set $z_2 := y_{(\varphi^1\circ\varphi^2)(k_2)}$. By construction, there holds

$$\begin{aligned}
& d_{SR}(z_1, z_2) \\
\leq\ & d_{SR}\left(z_1, y^1_{\varphi^1(k_1)}\right) + d_{SR}\left(y^1_{\varphi^1(k_1)}, z_2\right) \\
=\ & d_{SR}\left(y_{\varphi^1(k_1)}, y^1_{\varphi^1(k_1)}\right) + d_{SR}\left(y^1_{\varphi^1(k_1)}, y_{(\varphi^1\circ\varphi^2)(k_2)}\right) \\
\leq\ & \frac{R_x}{2} + d_{SR}\left(y^1_{\varphi^1(k_1)}, \bar{y}^1\right) + d_{SR}\left(\bar{y}^1, y^1_{(\varphi^1\circ\varphi^2)(k_2)}\right) \\
& \qquad\qquad\qquad\qquad + d_{SR}\left(y^1_{(\varphi^1\circ\varphi^2)(k_2)}, z_2\right) \\
\leq\ & \frac{R_x}{2} + \frac{R_x}{2} + \frac{R_x}{2} + \frac{R_x}{2} \leq 2R_x.
\end{aligned}$$

Repeating this construction yields a sequence of strictly increasing mappings $\{\varphi^l\}_l$, a sequence (with two indices) $\{y^l_k\}_{k,l}$, a sequence of limits $\{\bar{y}^l\}_l$, and a nondecreasing sequence of integers $\{k_l\}_l$ such that

$$d_{SR}\left(y_k, y^l_k\right) = \inf\left\{d_{SR}(y_k, z) \mid z \in K^l\right\} \leq \frac{R_x}{2^l}$$

and

$$d_{SR}\left(y^l_{\varphi^l(k)}, \bar{y}^l\right) \leq \frac{R_x}{2^l} \qquad \forall k \geq k_l.$$

Define the sequence $\{z_l\}_l$ by

$$z_l := y_{(\varphi^1\circ\varphi^2\circ\cdots\circ\varphi^l)(k_l)}$$

Then proceeding as above shows that for every $l \geq 1$, one has

$$d_{SR}\big(z_l, z_{l+1}\big) \leq \frac{4R_x}{2^l}.$$

Hence $\{z_k\}_k$ is a Cauchy sequence in $\bar{B}_{SR}(x, R_x)$. Since (M, d_{SR}) is complete, it converges to some $z \in \bar{B}_{SR}(x, R_x)$. □

Lemma 2.1.7. *The interval I_x is open in* $[0, +\infty)$.

Proof of Lemma 2.1.7. We need to show that if $R \in I_x$, then there is $\delta > 0$ such that $R+\delta$ belongs to I_x. Let $R > 0$ in I_x be fixed. Denote by $\partial B_{SR}(x, R)$ the boundary of $\bar{B}_{SR}(x, R)$, that is $\partial B_{SR}(x, R) = \bar{B}_{SR}(x, R) \setminus B_{SR}(x, R)$. Since $\bar{B}_{SR}(x, R)$ is assumed to be compact, its boundary is compact too. From Proposition 2.1.3, we know that for every $y \in \partial B_{SR}(x, R)$, there is $\delta_y > 0$ such that $\bar{B}_{SR}(y, 2\delta_y)$ is compact. Since

$$\partial B_{SR}(x, R) \subset \cup_{y\in\partial B_{SR}(x,R)} B_{SR}\left(y, \delta_y\right),$$

there is a finite number of points $y_1, \ldots, y_N$ in $\partial B_{SR}(x, R)$ such that

$$\partial B_{SR}(x, R) \subset \cup_{i=1}^N B_{SR}\left(y_i, \delta_{y_i}\right).$$

Set

$$\delta = \min\left\{\frac{\delta_{y_i}}{2} \mid i = 1, \ldots, N\right\}.$$

We prove easily that

$$\bar{B}_{SR}(x, R+\delta) \subset \left(\bar{B}_{SR}(x, R) \cup \cup_{i=1}^{N} \bar{B}_{SR}\left(y_i, 2\delta_{y_i}\right)\right)$$

which is a finite union of compact sets, hence compact as well. This shows that $\bar{B}_{SR}(x, R+\delta)$ is compact. □

In conclusion, I_x is both open and closed in $[0, +\infty)$. Hence $I_x = [0, +\infty)$ which concludes the proof of (i). Let us now prove assertion (ii). We note that since Δ does not necessarily admit a global orthonormal frame on M, we cannot repeat verbatim the proof of Proposition 2.1.3. Let $x, y \in M$ be fixed, set $R := \max\{2d_{SR}(x, y), 1\}$. By (i), we know that $\bar{B}_{SR}(x, R)$ is compact. Let $\{\gamma^k\}_k$ be a sequence of horizontal paths with constant speed in $\Omega_{\Delta}^{x,1}$ joining x to y such that

$$d_{SR}(x, y) = \lim_{k \to +\infty} \operatorname{length}(\gamma^k).$$

Without loss of generality we may assume that

$$\operatorname{length}(\gamma^k) < R \qquad \forall k,$$

which means that all the curves γ^k remain in $\bar{B}_{SR}(x, R)$. By Proposition 2.1.3, for every $z \in \bar{B}_{SR}(x, R)$ there is $\rho_z > 0$ such that any minimizing sequence of horizontal paths with constant speed contained in $B_{SR}(z, \rho_z)$ converges uniformly (up to taking a subsequence) to some minimizing geodesic. By compactness, there are $z_1, \ldots, z_L \in \bar{B}_{SR}(x, R)$ and an integer $N > 1$ with $R/N < \min\{\rho_1, \ldots, \rho_L\}/4$ such that

$$B_{SR}(x, R) \subset \bigcup_{l=1}^{L} B_{SR}\big(z_l, 1/N\big).$$

Set for every $j = 0, \ldots, N$, $t_j := j/N$, for every $j = 0, \ldots, N-1$, $I_j := [t_j, t_{j+1}]$, and denote by γ_j^k the restriction of γ^k to the interval I_j. Fix $j \in \{0, \ldots, N-1\}$. For every k, there is $l \in \{1, \ldots, L\}$ (which may depend on k) such that $d_{SR}(\gamma^k(t_j), z_l) < 1/N$, then

$$d_{SR}\big(\gamma^k(t), z_l\big) \leq d_{SR}(\gamma^k(t_j), z_l) + \frac{\operatorname{length}^g(\gamma^k)}{N} < \frac{1}{N} + \frac{R}{N} < \rho_l,$$

for every $t \in I_j$. This shows that each piece of horizontal path γ_j^k with length $\operatorname{length}^g(\gamma^k)/N$ is contained in some $B_{SR}(z_l, \rho_l)$. Therefore, up to taking a subsequence, the sequence $\{\gamma_j^k\}_k$ converges to some minimizing geodesic with length $d_{SR}(x, y)/N$. We deduce easily the existence of a subsequence of $\{\gamma^k\}_k$ converging to some minimizing geodesic between x and y. □

Remark 2.1.8. *In fact, we proved a global version of Proposition 2.1.3. If (M, d_{SR}) is a complete metric space, then for every $x, y \in M$ and every minimizing sequence $\{\gamma^k\}_k$ of horizontal paths with constant speed in $\Omega_{\Delta}^{x,1}$ joining x to y such that*

$$d_{SR}(x, y) = \lim_{k \to +\infty} \mathit{length}(\gamma^k),$$

up to taking a subsequence, $\{\gamma^k\}_k$ converges uniformly to some minimizing geodesic joining x to y.

We shall say that the sub-Riemannian structure (Δ, g) on M is *complete* if the metric space (M, d_{SR}) is complete. The following result holds.

Proposition 2.1.9. *Let (Δ, g) be a sub-Riemannian structure on M (assumed to be connected). Assume that (M, g) is a complete Riemannian manifold. Then for any totally nonholonomic distribution Δ, the SR structure (Δ, g) on M is complete.*

Proof. Denote by d_g the Riemannian geodesic distance on M with respect to g. Since the set of paths joinging x to y contains the set of horizontal paths joining x to y, there holds

$$d_g(x, y) \leq d_{SR}(x, y) \qquad \forall x, y \in M.$$

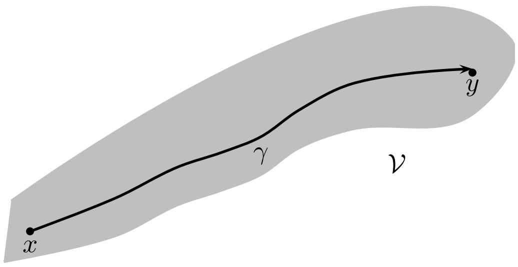
Therefore, any Cauchy sequence with respect to d_{SR} is a Cauchy sequence with respect to d_g. Hence it is convergent. Since both topology coincide, it is convergent with respect to d_{SR} as well. □

2.2 The Hamiltonian geodesic equation

Throughout all the section, we assume that the SR structure (Δ, g) is complete. Thanks to Theorem 2.1.5, minimizing geodesics exist between any pair of points in M.

Normal and abnormal geodesics

Let $x, y \in M$ and a minimizing geodesic $\gamma \in \Omega_{\Delta}^{x,1}$ joining x to y be fixed. Since γ minimizes the distance between x and y it cannot have self-intersection. Hence (Δ, g) admits an orthonormal frame along γ.



There is an open neighborhood $\mathcal{V}$ of $\gamma([0,1])$ in M and an orthonormal family $\mathcal{F}$ (with respect to the metric g) of m smooth vector fields $X^1, \ldots, X^m$ such that

$$\Delta(z) = \mathrm{Span}\Big\{X^1(z), \ldots, X^m(z)\Big\} \qquad \forall z \in \mathcal{V}.$$

Moreover, there is a control $u^\gamma \in L^2([0,1]; \mathbb{R}^m)$ (which indeed belong to the open set $\mathcal{U}_{\mathcal{F}}^{x,1}$ which was defined in Proposition 1.2.1) such that

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

Since γ is a minimizing geodesic between x and y, it minimizes the energy among all horizontal paths joining x to y. Since there is a local one-to-one correspondence between the set of horizontal paths starting at x and the set of trajectories of some control system (see Proposition 1.2.1), the control u^γ minimizes the quantity

$$\int_0^1 g_{\gamma_u(t)}\left(\sum_{i=1}^m u_i(t)X^i(\gamma_{x,u}(t)), \sum_{i=1}^m u_i(t)X^i(\gamma_u(t))\right)dt$$
$$= \int_0^1 \sum_{i=1}^m u_i(t)^2 dt =: C(u),$$

among all controls $u \in L^2([0,1];\mathbb{R}^m)$ such that the solution $\gamma_u : [0,1] \to M$ of the Cauchy problem

$$\dot{\gamma}_u(t) = \sum_{i=1}^k u_i(t)X^i\left(\gamma_u(t)\right) \quad \text{a.e. } t \in [0,1], \qquad \gamma_u(0) = x, \tag{2.4}$$

is well-defined on $[0,1]$ and satisfies (the End-Point mapping $E_{\mathcal{F}}^{x,1}$ has been defined in Chapter 1)

$$E_{\mathcal{F}}^{x,1}(u) = y.$$

In other terms, there is an open set $\mathcal{U} \subset L^2([0,1];\mathbb{R}^m)$ such that u^γ is solution to the following optimization problem:

$$u^\gamma \text{ minimizes } C(u) \text{ among all } u \in \mathcal{U} \text{ with } E_{\mathcal{F}}^{x,1}(u) = 1. \tag{2.5}$$

By the Lagrange Multiplier Theorem (see Theorem B.1.5), there is $p \in T_y^*M \simeq (\mathbb{R}^n)^*$ and $\lambda_0 \in \{0,1\}$ with $(\lambda_0, p) \neq (0,0)$ such that

$$p \cdot D_{u^\gamma} E_{\mathcal{F}}^{x,1}(v) = \lambda_0 D_{u^\gamma} C(v) \qquad \forall v \in L^2\big([0,1];\mathbb{R}^m\big). \tag{2.6}$$

Two cases may appear, either $\lambda_0 = 0$ or $\lambda_0 = 1$. By restricting $\mathcal{V}$ if necessary, we can assume that the cotangent bundle T^*M is trivializable with coordinates $(x,p) \in \mathcal{V} \times (\mathbb{R}^n)^*$ over $\mathcal{V}$.

First case: $\lambda_0 = 0$.
Then we have $p \in T_y^*M \setminus \{0\} \simeq (\mathbb{R}^n)^* \setminus \{0\}$ satisfying

$$p \cdot D_{u^\gamma} E_{\mathcal{F}}^{x,1}(v) = 0 \qquad \forall v \in L^2\big([0,1];\mathbb{R}^m\big).$$

This means that some nonzero linear form annihilates the image of $E_{\mathcal{F}}^{x,1}$. Then u^γ is singular with respect to x and $\mathcal{F}$ or equivalently the path γ is singular with respect to Δ. By Proposition 1.3.3 and Remark 1.3.5 (see also Proposition 1.3.10), γ admits an abnormal extremal lift, that is there is an absolutely continuous arc $p : [0,1] \to (\mathbb{R}^n)^* \setminus \{0\}$ with $p(1) = p$ which satisfies

$$\dot{p}(t) = -\sum_{i=1}^k u_i(t)\, p(t) \cdot D_{\gamma(t)} X^i \qquad \text{a.e. } t \in [0,1]$$

and

$$p(t) \cdot X^i\big(\gamma(t)\big) = 0, \quad \forall t \in [0,1] \qquad \forall i = 1, \cdots, m.$$

In other terms, γ is a *singular minimizing geodesic*.

Second case: $\lambda_0 = 1$.
Define in local coordinates, the Hamiltonian $H : \mathcal{V} \times (\mathbb{R}^n)^* \to \mathbb{R}$ by

$$H(x,p) := \frac{1}{2}\sum_{i=1}^{m}\big(p \cdot X^i(x)\big)^2 = \max_{u \in \mathbb{R}^m}\left\{\sum_{i=1}^{m} u_i\, p \cdot X^i(x) - \frac{1}{2}\sum_{i=1}^{m} u_i^2\right\} \tag{2.7}$$

for all $(x,p) \in \mathcal{V} \times (\mathbb{R}^n)^*$. Then the following result holds.

Proposition 2.2.1. *Equality (2.6) with $\lambda_0 = 1$ yields the existence of a smooth arc $p : [0,1] \longrightarrow (\mathbb{R}^n)^*$ with $p(1) = \frac{p}{2}$, such that the pair (γ, p) satisfies*

$$\left\{\begin{array}{rcl} \dot\gamma(t) &=& \frac{\partial H}{\partial p}(\gamma(t),p(t)) = \sum_{i=1}^m \left[p(t)\cdot X^i(\gamma(t))\right] X^i(\gamma(t)) \\ \dot p(t) &=& -\frac{\partial H}{\partial x}(\gamma(t),p(t)) = -\sum_{i=1}^m \left[p(t)\cdot X^i(\gamma(t))\right] p(t) \cdot D_{\gamma(t)}X^i \end{array}\right. \tag{2.8}$$

for a.e. $t \in [0,1]$ and

$$u_i^\gamma(t) = p(t) \cdot X^i(\gamma(t)) \qquad \textit{for a.e. } t \in [0,1], \quad \forall i = 1, \dots, m. \tag{2.9}$$

In particular, the path γ is smooth on $[0,1]$.

Proof. The differential of $C : L^2([0,1];\mathbb{R}^m) \to \mathbb{R}$ at u^γ is given by

$$D_{u^\gamma}C(v) = 2\langle u^\gamma, v\rangle_{L^2} \qquad \forall v \in L^2([0,1];\mathbb{R}^m).$$

Moreover by Remark 1.2.5, the differential of $E_{\mathcal{F}}^{x,1}$ at u^γ is given by

$$D_{u^\gamma}E_{\mathcal{F}}^{x,1}(v) = S(1)\int_0^1 S(t)^{-1}B(t)v(t)dt \qquad \forall v \in L^2([0,1];\mathbb{R}^m), \tag{2.10}$$

where the functions A, B, S were defined in Remark 1.2.5. Hence (2.6) yields

$$\int_0^1 \left[p \cdot S(1)S(t)^{-1}B(t) - 2u^\gamma(t)^*\right] v(t)dt = 0 \qquad \forall v \in L^2([0,1];\mathbb{R}^m).$$

Which implies

$$u^\gamma(t) = \frac{1}{2}\left(p \cdot S(1)S(t)^{-1}B(t)\right)^* \qquad \text{a.e. } t \in [0,1].$$

Let us define $p : [0,1] \to (\mathbb{R}^n)^*$ by

$$p(t) := \frac{1}{2} p \cdot S(1)S(t)^{-1} \qquad \forall t \in [0,1].$$

By construction, for a.e. $t \in [0,1]$ we have $u^\gamma(t)^* = p(t) \cdot B(t)$, which means that (2.9) is satisfied. Furthermore, as in the proof of Proposition 1.3.3, we have $\dot p(t) = -p(t) \cdot A(t)$ for a.e. $t \in [0,1]$. This means that (2.8) is satisfied for a.e. $t \in [0,1]$. The pair (γ, p) is solution to a smooth autonomous differential equation, hence it is smooth. $\square$

The curve $\psi : [0,1] \to T^*M$ given by $\psi(t) = (\gamma(t), p(t))$ for every $t \in [0,1]$ is a *normal extremal* whose the projection is γ and which satisfies $\psi(1) = (y, \frac{p}{2})$. We say that ψ is a *normal extremal lift* of γ. We also say that γ is a *normal minimizing geodesic*.

Define the Hamiltonian $H : T^*M \to \mathbb{R}$ as follows. For every $x \in M$, the restriction of H to the fiber T^*_xM is given by the nonnegative quadratic form

$$p \longmapsto \frac{1}{2} \max \left\{ \frac{p(v)^2}{g_x(v,v)} \mid v \in \Delta(x) \setminus \{0\} \right\}. \tag{2.11}$$

Let $\overrightarrow{H}$ denote the Hamiltonian vector field on T^*M associated to H, that is, $\iota_{\overrightarrow{H}}\omega = -dH$, or in local coordinates

$$\overrightarrow{H}(x,p) = \left(\frac{\partial H}{\partial p}(x,p), -\frac{\partial H}{\partial x}(x,p) \right).$$

A *normal extremal* is an integral curve of $\overrightarrow{H}$ defined on some interval $[0,T]$, i.e., a curve $\psi : [0,T] \to T^*M$ such that $\dot{\psi}(t) = \overrightarrow{H}(\psi(t))$, for $t \in [0,T]$. The projection of a normal extremal $\psi : [0,T] \to T^*M$ is a smooth horizontal path $\gamma := \pi \circ \psi : [0,T] \to M$ with constant speed given by

$$\left| \dot{\gamma}(t) \right|^g_{\gamma(t)} = \sqrt{2H\big(\psi(t)\big)} \qquad \forall t \in [0,T].$$

We check easily that the Hamiltonian defined by (2.11) reads as (2.7) in local coordinates. Then the previous study yields the following result.

Theorem 2.2.2. *Let $\gamma : [0,1] \to M$ be a minimizing geodesic between x and y in M. One of the two following non-exclusive cases may occur:*

- *γ is singular;*
- *γ admits a normal extremal lift in T^*M.*

Be careful, a minimizing geodesic could be both singular and the projection of a normal extremal. In Section 2.5, we shall see several examples of minimizing geodesics, including the cases of singular normal minimizing geodesics and strictly abnormal minimizing geodesic, that is abnormal geodesics admitting no normal extremal lift.

Remark 2.2.3. *In the Riemannian case, that is if Δ has rank $m = n$, any path is horizontal and regular (see Remark 1.3.7). As a consequence any minimizing geodesic is normal.*

Short normal geodesics are minimizing

Projections of normal extremals are minimizing for short times.

Proposition 2.2.4. *Let $\bar{x} \in M$ and $\bar{p} \in T^*_{\bar{x}}M$ with $H(\bar{x}, \bar{p}) \neq 0$ be fixed. Then there is a neighborhood $\mathcal{W}$ of $\bar{p}$ in $T^*_{\bar{x}}M$ and $\epsilon > 0$ such that every normal extremal so that $\psi(0) = (\bar{x}, p)$ (in local coordinates) belongs to $\mathcal{W}$ minimizes* *the SR energy on the interval* $[0,\epsilon]$. *That is if we set* $\gamma := \pi \circ \psi : [0,\epsilon] \to M$, *then we have*

$$e_{SR}\big(\gamma(0),\gamma(\epsilon)\big) = 2H(x,p)\epsilon^2.$$

In particular, γ *minimizes the length between* $\bar{x}$ *and* $\gamma(\epsilon)$.

Proof. Since the result is local, we can assume that we work in $\mathbb{R}^n$. Then we can assume that (Δ, g) admits an orthonormal frame $\mathcal{F} = \{X^1,\ldots,X^m\}$. For sake of simplicity, we identify $(\mathbb{R}^n)^*$ with $\mathbb{R}^n$. Then the Hamiltonian $H : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ which were defined in (2.7)-(2.11) is given by

$$H(x,p) := \max_{u\in\mathbb{R}^m}\left\{\langle p, \sum_{i=1}^m u_i X^i(x)\rangle - \frac{1}{2}\sum_{i=1}^m u_i^2\right\} = \frac{1}{2}\sum_{i=1}^m \langle p, X^i(x)\rangle^2,$$

for every $(x,p) \in R^n \times \mathbb{R}^n$.
Our aim is now to prove the following result: for every $p_0 \in \mathbb{R}^n$ such that $H(\bar{x},p_0) \neq 0$, there exist a neighborhood $\mathcal{W}$ of p_0 in $\mathbb{R}^n$ and $\epsilon > 0$ such that every solution $(x,p) : [0,\epsilon] \to \mathbb{R}^n \times \mathbb{R}^n$ of the Hamiltonian system

$$\left\{\begin{array}{rcl} \dot{x}(t) & = & \dfrac{\partial H}{\partial p}(x(t),p(t)) = \displaystyle\sum_{i=1}^m \langle p(t), X^i(x(t))\rangle X^i(x(t)) \\ \dot{p}(t) & = & -\dfrac{\partial H}{\partial x}(x(t),p(t)) = -\displaystyle\sum_{i=1}^m \langle p(t), X^i(x(t))\rangle \big(D_{x(t)}X^i\big)^*\big(p(t)\big), \end{array}\right. \tag{2.12}$$

with $x(0) = \bar{x}$ and $p(0) \in \mathcal{W}$, satisfies

$$2\epsilon H\big(\bar{x},p_0\big) = \int_0^\epsilon \sum_{i=1}^m \langle p(t), X^i(x(t))\rangle^2 dt \leq \int_0^\epsilon \sum_{i=1}^m u_i(t)^2 dt, \tag{2.13}$$

for every control $u \in L^2([0,\epsilon];\mathbb{R}^m)$ such that the solution of

$$\dot{y}(t) = \sum_{i=1}^m u_i(t) X^i(y(t)), \quad y(0) = \bar{x}, \tag{2.14}$$

satisfies $y(\epsilon) = x(\epsilon)$. Let $p_0 \in \mathbb{R}^n$ with $H(\bar{x},p_0) \neq 0$ be fixed, we need the following lemma.

Lemma 2.2.5. *There exist a neighborhood* $\mathcal{W}$ *of* p_0 *and* $\rho > 0$ *such that, for every* $p \in \mathcal{W}$*, there exists a function* $S : B(\bar{x},\rho) \to \mathbb{R}$ *of class* C^1 *which satisfies*

$$H(x,\nabla S(x)) = H(\bar{x},p), \quad \forall x \in B(\bar{x},\rho), \tag{2.15}$$

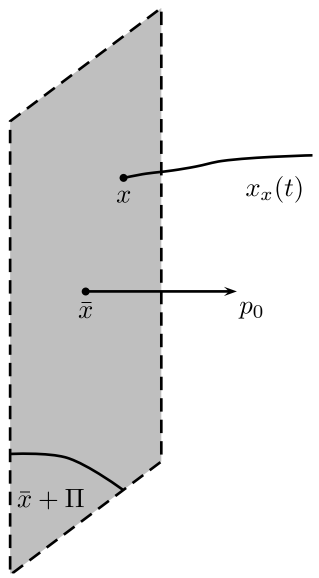
and such that, if $(x^p,p^p) : [-\rho,\rho] \to \mathbb{R}^n \times \mathbb{R}^n$ *denotes the solution of (2.12) satisfying* $x^p(0) = \bar{x}$ *and* $p^p(0) = p$*, then*

$$\nabla S(x^p(t)) = p^p(t), \quad \forall t \in (-\rho,\rho). \tag{2.16}$$

Proof of Lemma 2.2.5. The proof consists in applying the method of characteristics. Let Π be the linear hyperplane such that $\langle p_0, v\rangle = 0$ for every $v \in \Pi$. We first show how to construct locally S as the solution of the Hamilton-Jacobi equation (2.15) which vanishes on $\bar{x} + \Pi$ and such that $\nabla S(\bar{x}) = p_0$. Up to considering a smaller neighborhood $\mathcal{V}$, we assume that $H(x, p_0) \neq 0$ for every $x \in \mathcal{V}'$. For every $x \in (\bar{x} + \Pi) \cap \mathcal{V}$, set

$$\bar{p}(x) := \sqrt{\frac{H(\bar{x}, p_0)}{H(x, p_0)}}\, p_0.$$

Then, $H(x, \bar{p}(x)) = H(\bar{x}, p_0)$ and $\bar{p}(x) \perp \Pi$, for every $x \in \mathcal{V}'$. There exists $\mu > 0$ such that, for every $x \in (\bar{x} + \Pi) \cap \mathcal{V}$, the solution (x_x, p_x) of (2.12), satisfying $x_x(0) = x$ and $p_x(0) = \bar{p}(x)$, is defined on the interval $(-\mu, \mu)$.



For every $x \in (\bar{x} + \Pi) \cap \mathcal{V}$ and every $t \in (-\mu, \mu)$, set $\theta(t, x) := x_x(t)$. The mapping $(t, x) \mapsto \theta(t, x)$ is smooth. Moreover, $\theta(0, x) = x$ for every $x \in (\bar{x} + \Pi) \cup \mathcal{V}$ and $\dot{\theta}(0, \bar{x}) = \sum_{i=1}^m \langle \bar{p}(x), X^i(\bar{x}) \rangle X^i(\bar{x})$ does not belong to Π. Hence there exists $\rho \in (0, \mu)$ with $B(\bar{x}, \rho) \subset \mathcal{V}$ such that the mapping θ is a smooth diffeomorphism from $(-\rho, \rho) \times ((\bar{x} + \Pi) \cap B(\bar{x}, \rho))$ into a neighborhood $\mathcal{V}'$ of $\bar{x}$. Denote by $\varphi = (\tau, \pi)$ the inverse function of θ, that is the function such that $(\theta \circ \varphi)(x) = (\tau(x), \pi(x)) = x$ for every $x \in \mathcal{V}'$. Define the two vector fields X and P by

$$X(x) := \dot{\theta}(\tau(x), \pi(x)) \quad \text{and} \quad P(x) := p_{\pi(x)}(\tau(x)), \quad \forall x \in \mathcal{V}'.$$

Then,

$$\begin{aligned} X(\theta(t, x)) = \dot{\theta}(t, x) = \dot{x}_x(t) &= \sum_{i=1}^m \langle p_x(t), X^i(x_x(t)) \rangle X^i(x_x(t)) \\ &= \sum_{i=1}^m \langle P(\theta(t, x)), X^i(\theta(t, x)) \rangle X^i(\theta(t, x)), \end{aligned}$$

and

$$\begin{aligned} \sum_{i=1}^m \langle P(\theta(t, x)), X^i(x_x(t)) \rangle^2 = \sum_{i=1}^m \langle p_x(t), X^i(x_x(t)) \rangle^2 \\ = 2H(x, \bar{p}(x)) = 2H(\bar{x}, p_0), \end{aligned}$$

for every $t \in (-\rho, \rho)$ and every $x \in (\bar{x} + \Pi) \cap B(\bar{x}, \rho)$. For every $x \in \mathcal{V}'$, set $\alpha_i(x) := \langle P(x), X^i(x) \rangle$. Hence,

$$X(x) = \sum_{i=1}^{m} \alpha_i(x) X^i(x) \quad \text{and} \quad \sum_{i=1}^{m} \alpha_i(x)^2 = 2H(\bar{x}, p_0),$$

for every $x \in \mathcal{V}'$. Define the function $S : \mathcal{V}' \mapsto \mathbb{R}$ by

$$S(x) := 2H(\bar{x}, p_0)\tau(x), \quad \forall x \in \mathcal{V}'.$$

We next prove that $\nabla S(x) = P(x)$ for every $x \in \mathcal{V}'$. For every $t \in (-\rho, \rho)$, denote by $W_t := \{y \in \mathcal{V}' \mid \tau(y) = t\}$. In fact, W_t coincides with the set of $y \in \mathcal{V}'$ such that $S(y) = 2H(\bar{x}, p_0)t$. It is a smooth hypersurface which satisfies $\nabla S(y) \perp T_y W_t$ for every $y \in W_t$. Let $y \in W_t$ be fixed, there exists $x \in (\bar{x} + \Pi) \cup B(\bar{x}, \rho)$ such that $y = \theta(t, x) = x_x(t)$. Let us first prove that $P(y) = p_x(t)$ is orthogonal to $T_y W_t$. To this aim, without loss of generality we assume that $t > 0$. Let $w \in T_y W_t$, there exists $v \in \Pi$ such that $w = D_x \theta_t(v)$. For every $s \in [0, t]$, set $z(s) := D_x \theta(s, x)(v)$. We have

$$\dot{z}(s) = \frac{d}{ds} D_x \theta(s, x) v = \frac{d}{dx} \dot{\theta}(t, x) v = \frac{d}{dx} X(\theta(t, x)) v = D_{\theta(t,x)} X(z(s)).$$

Hence,

$$\begin{aligned}
\frac{d}{ds} \langle z(s), p_x(s) \rangle &= \langle \dot{z}(s), p_x(s) \rangle + \langle z(s), \dot{p}_x(s) \rangle \\
&= \langle dX(\theta(s, x)) z(s), p_x(s) \rangle \\
&\quad - \langle z(s), \sum_{i=1}^{m} \langle p_x(s), X^i(x_x(s)) \rangle \big(D_{x_x(s)} X^i\big)^* \big(p_x(s)\big) \rangle.
\end{aligned}$$

Since $X(x) = \sum_{i=1}^{m} \alpha_i(x) X^i(x)$ and $\sum_{i=1}^{m} \alpha_i(x)^2 = 2H(\bar{x}, p_0)$ for every $x \in \mathcal{V}'$, there holds

$$\begin{aligned}
& \big(D_{x_x(s)} X\big)^* \big(p_x(s)\big) \\
= \; & \sum_{i=1}^{m} \alpha_i(x_x(s)) \big(D_{x_x(s)} X^i\big)^* \big(p_x(s)\big) + \sum_{i=1}^{m} \langle X^i(x_x(s)), p_x(s) \rangle \nabla \alpha_i(x_x(s)) \\
= \; & \sum_{i=1}^{m} \alpha_i(x_x(s)) \big(D_{x_x(s)} X^i\big)^* \big(p_x(s)\big) + \sum_{i=1}^{m} \alpha_i(x_x(s)) \nabla \alpha_i(x_x(s)) \\
= \; & \sum_{i=1}^{m} \alpha_i(x_x(s)) \big(D_{x_x(s)} X^i\big)^* \big(p_x(s)\big).
\end{aligned}$$

We deduce that $\frac{d}{ds} \langle z(s), p_x(s) \rangle = 0$ for every $s \in [0, t]$. Hence,

$$\langle w, P(y) \rangle = \langle w, p_x(t) \rangle = \langle z(t), p_x(t) \rangle = \langle z(0), \bar{p}(x) \rangle = 0.$$

This proves that $P(y)$ is orthogonal to $T_y W_t$, which implies that $P(y)$ and $\nabla S(y)$ are colinear. Furthermore, since $S(x_x(s)) = 2H(\bar{x}, p_0)s$ for every $s \in [0, t]$, one gets

$$\langle \nabla S(x_x(t)), \dot{x}_x(t) \rangle = 2H(\bar{x}, p_0) = \langle p_x(t), \dot{x}_x(t) \rangle.$$

Since $\dot{x}_x(t) = X(y)$ does not belong to $T_y W_t$, we deduce that $\nabla S(x_x(t)) = p_x(t)$. In consequence, we proved that $\nabla S(x) = P(x)$ for every $x \in \mathcal{V}'$. $\square$

Let us now conclude the proof of Proposition 2.2.4. Clearly, there exists $\epsilon > 0$ such that every solution $(x,p) : [0,\epsilon] \to \mathbb{R}^n \times \mathbb{R}^n$ of (2.12), with $x(0) = \bar{x}$ and $p(0) \in \mathcal{W}$, satisfies

$$x(t) \in B(\bar{x},\rho), \quad \forall t \in [0,\epsilon].$$

Moreover, we have

$$S(x(\epsilon)) - S(\bar{x}) = 2\epsilon H(\bar{x},p).$$

Let $u \in L^1([0,\epsilon];\mathbb{R}^m)$ be a control such that the solution $y : [0,\epsilon] \to \mathcal{W}$ of (2.14) starting at $\bar{x}$ satisfies $y(\epsilon) = x(\epsilon)$. We have

$$\begin{aligned} S(x(\epsilon)) - S(\bar{x}) &= S(y(\epsilon)) - S(y(0)) \\ &= \int_0^\epsilon \frac{d}{dt}\left(S(y(t))\right) dt \\ &= \int_0^\epsilon \langle \nabla S(y(t)), \dot{y}(t)\rangle dt \\ &\leq \int_0^\epsilon H(y(t), dS(y(t))) + \frac{1}{2}\sum_{i=1}^m u_i(t)^2 dt \\ &= \epsilon H(\bar{x},p) + \frac{1}{2}\int_0^\epsilon \sum_{i=1}^m u_i(t)^2 dt. \end{aligned}$$

The conclusion follows. □

2.3 The sub-Riemannian exponential map

Definition

Recall that the Hamiltonian $H : T^*M \to \mathbb{R}$ which is canonically associated with our SR structure (Δ, g) is defined by

$$H(x,p) = \frac{1}{2}\max\left\{\frac{p(v)^2}{g_x(v,v)} \mid v \in \Delta(x) \setminus \{0\}\right\} \qquad \forall (x,p) \in T^*M. \qquad (2.17)$$

We recall that a normal extremal is a curve $\psi : [0,T] \to T^*M$ satisfying

$$\dot{\psi}(t) = \overrightarrow{H}(\psi(t)) \qquad \forall t \in [0,T].$$

Let $x \in M$ be fixed. We first define the domain $\mathcal{E}_x \subset T_x^*M$ of the SR exponential map by,

$$\mathcal{E}_x := \Big\{ p \in T_x^*M \mid \psi_{x,p} \text{ is defined on the interval } [0,1] \Big\},$$

where $\psi_{x,p}$ is the normal extremal so that $\psi_{x,p}(0) = (x,p)$ in local coordinates. The set $\mathcal{E}_x$ is an open subset of T_x^*M containing the origin and star-shaped with respect to 0.

Definition 2.3.1. *The sub-Riemannian exponential map from x is defined by*

$$\begin{aligned} \exp_x : \mathcal{E}_x \subset T_x^*M &\longrightarrow M \\ p &\longmapsto \pi\left(\psi_{x,p}(1)\right). \end{aligned}$$

By rescaling, if $(x_p, p_p) : [0, T] \to T^*M$ is the trajectory of the Hamiltonian vector field $\overrightarrow{H}$ with $x(0) = x, p(0) = p$, then we have

$$\big(x_p(\lambda t), \lambda p_p(\lambda t)\big) = \big(x_{\lambda p}(t), p_{\lambda p}(t)\big) \qquad \forall t \in [0, T/\lambda], \forall \lambda > 0.$$

Then, for every $p \in T_x^*M$, the curve $\gamma_p : [0, 1] \to M$ defined by

$$\gamma_p(t) := \exp_x(tp) = \pi\left(\psi_{x,p}(t)\right) \qquad \forall t \in [0, 1],$$

is an horizontal path with constant speed satisfying

$$\text{energy}^g(\pi(\psi_{x,p})) = \left(\text{length}^g(\pi(\psi_{x,p}))\right)^2 = 2H\big(\psi_{x,p}(0)\big) = 2H(x, p).$$

Proposition 2.3.2. *Assume that (Δ, g) is complete. Then*

$$\mathcal{E}_x = T_x^*M \qquad \forall x \in M.$$

Proof. We argue by contradiction. Let $\bar{x} \in M$ and $\psi = (\bar{\gamma}, p_{\bar{p}}) : [0, T) \to T^*M$ be a normal extremal starting at $(\bar{x}, \bar{p}) \in T_{\bar{x}}^*M$ that extends to no interval $[0, T + \epsilon)$ for $\epsilon > 0$. Let $\{t_k\}_k$ be any increasing sequence that approaches T, and set $y_k := \bar{\gamma}(t_k)$. Since $\bar{\gamma}$ is an horizontal path with constant speed $V = \sqrt{2H(\bar{x}, \bar{p})}$, we have

$$d_{SR}\big(y_k, y_l\big) \le V\big|t_k - t_l\big| \qquad \forall k, l.$$

Then $\{y_k\}_k$ is a Cauchy sequence in M. By completeness $\{y_k\}_k$ converges to some point $y \in M$. Let $\{X^1, \dots, X^m\}$ be a local orthonormal frame in a small ball $B_{SR}(y, r)$. In local coordinates near y, H reads

$$H(x, p) = \frac{1}{2} \sum_{i=1}^{m} \big(p \cdot X^i(x)\big)^2$$

and $(\bar{\gamma}, p_{\bar{p}})$ satisfies the differential system

$$\begin{cases} \dot{\bar{\gamma}}(t) &= \frac{\partial H}{\partial p}\big(\bar{\gamma}(t), p_{\bar{p}}(t)\big) = \sum_{i=1}^m \left[p_{\bar{p}}(t) \cdot X^i\big(\bar{\gamma}(t)\big)\right] X^i\big(\bar{\gamma}(t)\big) \\ \dot{p}_{\bar{p}}(t) &= -\frac{\partial H}{\partial x}\big(\bar{\gamma}(t), p_{\bar{p}}(t)\big) = -\sum_{i=1}^m \left[p_{\bar{p}}(t) \cdot X^i\big(\bar{\gamma}(t)\big)\right] p_{\bar{p}}(t) \cdot D_{\bar{\gamma}(t)} X^i, \end{cases}$$

for $t \in [T - \delta, T)$ with $\delta > 0$ small enough. Since H is constant along $(\bar{\gamma}, p_{\bar{p}})$, we have

$$\big|p_{\bar{p}}(t) \cdot X^i\big(\bar{\gamma}(t)\big)\big| \le \sqrt{m} V \qquad \forall i = 1, \dots, m,$$

and by compactness the $X'i$'s and the dX^i's are bounded in $\bar{B}_{SR}(y, r)$. Thus there is a constant $K > 0$ such that

$$|\dot{p}_{\bar{p}}(t)| \le K\, |p_{\bar{p}}(t)| \qquad \forall t \in [T - \delta, T).$$

By Gronwall's Lemma (see Lemma A.1.1), we infer that both $\bar{\gamma}$ and $p_{\bar{p}}$ are uniformly bounded near T. This means that the extremal ψ can be extended beyong T, which gives a contradiction. □

Remark 2.3.3. *Let $(x, p) \in T * M$ such that γ_p is singular be fixed. By Proposition 1.3.10, γ_p is the projection of a characteristic curve $\psi : [0, 1] \to T * M$ (written as (γ_p, q) in local coordinates). Then taking local coordinates,* *for every $\lambda \in \mathbb{R}$ the curve (here $\psi_{x,p} = (\gamma_p, p_p)$ denotes the normal extremal starting at (x,p))*

$$t \in [0,1] \longmapsto \psi_{x,p}(t) + \lambda\psi(t) = \big(\gamma_p(t), p_p(t) + \lambda q(t)\big)$$

is a normal extremal starting at $(x, p+\lambda q)$. Then we have

$$\exp_x\big(p+\lambda q\big) = \exp_x(p) \qquad \forall \lambda \in \mathbb{R}.$$

Remark 2.3.4. *Let (Δ, g) be a complete sub-Riemannian structure on M and $x \in M$ be fixed. If (Δ, g) does not admit singular minimizing curves from x, then the exponential map from x is onto. As a matter of fact, for every $y \in M$, there is a minimizing geodesic $\gamma : [0,1] \to M$ joining x to y. Since γ is not singular, it is the projection of a normal extremal (see Theorem 2.2.2), which means that there is $p \in T_x^*M$ such that $\exp_x(p) = y$.*

On the image of $\exp_x$

Sub-Riemannian exponential maps are "almost" onto.

Theorem 2.3.5. *Assume that (Δ, g) is complete and let $x \in M$ be fixed. There is an open and dense set $\mathcal{D} \subset M$ such that for every $y \in \mathcal{D}$ there is $p_y \in T_x^*M$ satisfying*

$$\exp_x\big(p_y\big) = y \quad \textit{and} \quad d_{SR}(x,y) = \sqrt{2H\,(x,p_y)}. \tag{2.18}$$

In particular, the set $\exp_x(T_x^*M)$ contains an open dense subset of M.

Proof. Let us begin with a preparatory lemma.

Lemma 2.3.6. *Let $y \neq x$ in M be such that there is a function $\phi : M \to \mathbb{R}$ differentiable at y such that*

$$\phi(y) = d_{SR}^2(x,y) \quad \textit{and} \quad d_{SR}^2(x,z) \geq \phi(z) \quad \forall z \in M. \tag{2.19}$$

Then there is a unique minimizing geodesic $\gamma : [0,1] \to M$ between x and y. It is the projection of a normal extremal $\psi : [0,1] \to T^*M$ satisfying $\psi(1) = (y, \frac{1}{2}D_y\phi)$. In particular $x = \exp_y(-\frac{1}{2}D_y\phi)$.

Proof of Lemma 2.3.6. Since $e_{SR}(x,z) = d_{SR}^2(x,z)$ for any $z \in M$, the assumption of the proposition implies that there is a neighborhood $\mathcal{U}$ of y in M such that

$$e_{SR}(x,z) \geq \phi(z) \quad \forall z \in \mathcal{U} \qquad \text{and} \qquad e_{SR}(x,y) = \phi(y). \tag{2.20}$$

Since (M, d_{SR}) is complete, there exists a minimizing geodesic $\gamma : [0,1] \to M$ between x and y. As before, we can parametrize the distribution Δ by a orthonormal family $\mathcal{F}$ of smooth vector fields $X^1, \ldots, X^m$ in a neighborhood $\mathcal{V}$ of $\gamma([0,1])$, and we denote by u^γ the control corresponding to γ. By construction, it minimizes the quantity

$$C(u) = \int_0^1 \sum_{i=1}^m u_i(t)^2 dt,$$

among all the controls $u \in L^2([0,1];\mathbb{R}^m)$ which are admissible with respect to $x, \mathcal{F}$ and $\mathcal{V}$ and which satisfy the constraint $E_{\mathcal{F}}^{x,1}(u) = y$. Let $u \in L^2([0,1];\mathbb{R}^m)$ be a control admissible with respect to $x, \mathcal{F}$ and $\mathcal{V}$ such that $E_{\mathcal{F}}^{x,1}(u) \in \mathcal{U}$. By (2.20) one has

$$C(u) \geq e_{SR}\left(x, E_{\mathcal{F}}^{x,1}(u)\right) \geq \phi\left(E_{\mathcal{F}}^{x,1}(u)\right).$$

Moreover

$$C(u^\gamma) = e_{SR}(x,y) = \phi(y) = \phi\left(E_{\mathcal{F}}^{x,1}(u^\gamma)\right).$$

Hence u^γ minimizes the functional $D : L^2([0,1];\mathbb{R}^m) \to \mathbb{R}$ defined as

$$D(u) := C(u) - \phi\left(E_{\mathcal{F}}^{x,1}(u)\right),$$

over the set of controls $u \in L^2([0,1];\mathbb{R}^m)$ such that $E_{\mathcal{F}}^{x,1}(u) \in \mathcal{U}$. This means that u^γ is a critical point of D. Setting $\lambda = D_y\phi$, we obtain

$$\lambda \cdot D_{u^\gamma} E_{\mathcal{F}}^{x,1} - D_{u^\gamma} C = 0.$$

By Proposition 2.2.1, the path γ admits a normal extremal lift $\psi : [0,1] \to T^*M$ satisfying $\psi(1) = (y, \frac{1}{2} D_y \phi)$. By the Cauchy-Lipschitz Theorem, such a normal extremal is unique. $\square$

Denote by $\mathcal{P}_x$ the set of points in M such that there is a unique normal minimizing geodesic γ_y from x to y. The previous lemma yields easily the following result.

Lemma 2.3.7. *The set $\mathcal{P}_x$ is dense in M.*

Proof of Lemma 2.3.7. Let $y \in M$ and $r > 0$ be fixed. Let $\varphi : M \to \mathbb{R}$ be a smooth function such that

$$\varphi(y) = 0 \quad \text{and} \quad \varphi(z) \geq 2r \quad \forall z \in \partial B_{SR}(y,r).$$

The continuous function

$$z \in \bar{B}_{SR}(x,r) \longmapsto d_{SR}(x,z) + \varphi(z)$$

is equal to $d_{SR}(x,y)$ at $z = y$ and by the triangle inequality it is larger than $d_{SR}(x,y) + r$ for $z \in \partial B_{SR}(y,r)$. Then there is $\bar{z} \in B_{SR}(y,r)$ such that

$$d_{SR}(x,z) \geq d_{SR}\big(x,\bar{z}\big) + \varphi\big(\bar{z}\big) - \varphi(z) \qquad \forall z \in B_{SR}(y,r).$$

We conclude easily by Lemma 2.3.6. $\square$

For every $y \in M$ denote by $\operatorname{rank}(y)$ the rank of the horizontal path γ_y (see Section 1.3).

Lemma 2.3.8. *The set of $y \in \mathcal{P}_x$ with rank$(y) = n$ is dense in M.*

Proof of Lemma 2.3.8. We argue by contradiction. Assume that there is an open set $\mathcal{O} \subset M$ such that any point $y \in \mathcal{P}_x \cap \mathcal{O}$ has rank $< n$. Set

$$\hat{r} := \max\Big\{\mathrm{rank}(y) \,|\, y \in \mathcal{P}_x \cap \mathcal{O}\Big\}.$$

Fix $\hat{y} \in \mathcal{P}_x \cap \mathcal{O}$ such that $\mathrm{rank}(\hat{y}) = \hat{r}$ and set $\hat{\gamma} := \gamma_{\hat{y}}$. For every $y \in \mathcal{P}_x \cap \mathcal{O}$ denote by Π_y the affine subspace of T_x^*M such that $\gamma_p = \gamma_y$, that is the space of $p \in T_x^*M$ such that

$$\gamma_p(t) = \exp_x(tp) = \pi\left(\psi_{x,p}(t)\right) = \gamma_y(t) \qquad \forall t \in [0,1].$$

Remembering Remark 2.3.3, we observe that the dimension of Π_y is exactly equal to $n - \mathrm{rank}(y)$. As a matter of fact, given $y \in \mathcal{P}_x \cap \mathcal{O}$ and an orthonormal family $\mathcal{F} = \{X^1, \ldots, X^m\}$ in a neighborhood $\mathcal{V}$ along γ_y, remembering the arguments given in Proposition 2.2.1 we check that $p \in T_x^*M$ belongs to Π_y if and only if $\psi_{x,p}(1) = (y, p_p(1))$ satisfies

$$2p_p(1) \cdot D_{u^{\gamma_y}} E_{\mathcal{F}}^{x,1}(v) = D_{u^{\gamma_y}} C(v) \qquad \forall v \in L^2\big([0,1];\mathbb{R}^m\big). \tag{2.21}$$

Let $\{y_k\}_k$ be a sequence in $\mathcal{P}_x \cap \mathcal{O}$ converging to $\hat{y}$, $\mathcal{F} = \{X^1, \ldots, X^m\}$ be an orthonormal family in a neighborhood $\mathcal{V}$ along $\hat{\gamma}$, and $\hat{u}$ the control associated with $\hat{\gamma}$ through $\mathcal{F}$. The End-Point mapping $E_{\mathcal{F}}^{x,1}$ is valued in $\mathbb{R}^n$; denote by $E_1, \ldots, E_n$ its n coordinates. The vector space (we identify $L^2([0,1];\mathbb{R}^m)$ with its dual)

$$\mathrm{Span}\Big\{D_{\hat{u}}E_1, \ldots, D_{\hat{u}}E_n\Big\}$$

has dimension $\mathrm{rank}(\hat{y}) = \hat{r}$. Let $i_1, \ldots, i_{\hat{r}} \in \{1, \ldots, n\}$ be such that

$$\mathrm{Span}\Big\{D_{\hat{u}}E_{i_1}, \ldots, D_{\hat{u}}E_{i_{\hat{r}}}\Big\} = \mathrm{Span}\Big\{D_{\hat{u}}E_1, \ldots, D_{\hat{u}}E_n\Big\}. \tag{2.22}$$

Proceeding as in the proof of Proposition 2.1.3 and using completeness of (Δ, g), we show that taking a subsequence if necessary, $\{\gamma_k := \gamma_{y_k}\}_k$ converges uniformly to some minimizing geodesic joining x to y. By uniqueness, we infer that $\lim_k \gamma_k = \hat{\gamma}$. Furthermore, the proof also shows that the controls $u_k := u_{\gamma_k}$ which are associated to the γ_k's through the orthonormal family $\mathcal{F}$ converges strongly to $\hat{u}$ in $L^2([0,1];\mathbb{R}^m)$ (see Remark 2.1.4). Then by regularity of $E_{\mathcal{F}}^{x,1}$ and the fact that $\mathrm{rank}(y_k) \leq \hat{r}$, we deduce that $\mathrm{rank}(y_k) = \mathrm{rank}(y)$ for k large enough and that

$$\mathrm{Span}\Big\{D_{u_k}E_{i_1}, \ldots, D_{u_k}E_{i_{\hat{r}}}\Big\} = \mathrm{Span}\Big\{D_{u_k}E_1, \ldots, D_{u_k}E_n\Big\}.$$

By (2.21) and (2.22), there is $\hat{\lambda} = (\hat{\lambda}_1, \ldots, \hat{\lambda}_{\hat{r}}) \in T_{\hat{y}}^*M \simeq (\mathbb{R}^{\hat{r}})^*$ such that (remember that we identify $L^2([0,1];\mathbb{R}^m)$ with its dual)

$$\sum_{j=1}^{\hat{r}} \hat{\lambda}_j D_{u^{\gamma_y}} E_{i_j} = u_{\gamma_y},$$

and more generally for every k there is $\lambda^k = (\lambda_1^k, \ldots, \lambda_{\hat{r}}^k) \in T_{y_k}^*M \simeq (\mathbb{R}^{\hat{r}})^*$ such that

$$\sum_{j=1}^{\hat{r}} \lambda_j^k D_{u_k} E_{i_j} = u_{\gamma_k},$$

Since $\{u_k\}_k$ converges to $\hat{u}$ in $L^2([0,1];\mathbb{R}^m)$, $\{D_{u_k}E_{\mathcal{F}}^{x,1}\}_k$ converges to $D_{\hat{u}}E_{\mathcal{F}}^{x,1}$ and $D_{\hat{u}}E_{i_1},\ldots,D_{\hat{u}}E_{i_{\hat{r}}}$ are linearly independant, we infer that $\{\lambda^k\}_k$ tends to $\hat{\lambda}$ as k tends to $+\infty$. Define $\{p_k\}_k$ and $\hat{p}$ in T_x^*M by

$$\psi_{x,p_k}(1) = (y_k, \lambda^k/2)\,\forall k \quad \text{and} \quad \psi_{x,\hat{p}}(1) = \big(\hat{y}, \hat{\lambda}/2\big).$$

By regularity of the Hamiltonian flow, $\{p_k\}_k$ tends to $\hat{p}$ and if a bounded sequence $\{p_k+q_k\}_k$ is contained in Π_k then it converges (up to a subsequence) to some point in $\Pi_{\hat{y}}$. This shows that Π_k tends to $\Pi_{\hat{y}}$. All in all we proved that the mapping $y\in\mathcal{P}_x\mapsto\Pi_y$ is continuous at $\hat{y}$. Let $\mathcal{S}$ be a smooth compact submanifold of dimension $\hat{r}$ in T_x^*M which is transverse to $\Pi_{\hat{y}}$ at $\hat{p}$, that is such that

$$\Pi_{\hat{y}}\cap\mathcal{S}=\{\hat{p}\} \quad \text{and} \quad T_{\hat{p}}\mathcal{S}\cap T_{\hat{p}}\Pi_{\hat{y}}=\{0\}.$$

By regularity of $y\mapsto\Pi_y$, there is an open neighrborhood $\mathcal{O}'\subset\mathcal{O}$ of $\hat{y}$ such that $\mathcal{S}$ is transverse to any Π_y with $y\in\mathcal{P}_x\cap\mathcal{O}'$. We infer that

$$\{y\} = \exp_x(\Pi_y) = \exp_x(\Pi_y\cap\mathcal{S}) \subset \exp_x(\mathcal{S}) \qquad \forall y\in\mathcal{P}_x\cap\mathcal{O}'.$$

But since $\mathcal{S}$ has dimension strictly less than n, the set $\exp_x(\mathcal{S})$ is a compact set of measure zero in M. Then $\mathcal{P}_x\cap\mathcal{O}'$ cannot be dense in $\mathcal{O}'$. Which gives a contradiction. $\square$

Returning to the proof of Theorem 2.3.5, we fix $\bar{y}\in\mathcal{P}_x$ with $\mathrm{rank}(\bar{y})=n$. Given an open set $\Omega\subset M$, we call a function $f:\Omega\to\mathbb{R}$ *Lipschitz in charts* if it is Lipschitz in a set of local coordinates in a neighborhood of any point of Ω. This is equivalent to saying that f is locally Lipschitz with respect to a (complete) Riemannian distance on M.

Lemma 2.3.9. *There is an open set $\mathcal{O}_{\bar{y}}$ of $\bar{y}$ in M such that the function*

$$y\in\mathcal{O}_{\bar{y}}\longmapsto d_{SR}\big(x,y\big)$$

is Lipschitz in charts.

Proof of Lemma 2.3.9. As before we fix an orthonormal family of vector fields $\mathcal{F}$ in an open neighborhood $\mathcal{V}$ along $\bar{\gamma}:=\gamma_{\bar{y}}$ which is associated with $\bar{u}\in L^2([0,1];\mathbb{R}^m)$ through $\mathcal{F}$. By a uniqueness-compactness argument, if $\{y_k\}_k$ converges to $\bar{y}$ and $\{\gamma_k\}_k$ is a sequence of minimizing geodesics between x and y_k then it converges (up to a subsequence) to $\bar{\gamma}$ and is associated with a sequence of controls $\{u_k\}_k$ which converges to $\bar{u}$ in $L^2([0,1];\mathbb{R}^m)$ (see Proposition 2.1.3 and Remark 2.1.4). Then there is a neighborhood $\mathcal{O}$ of $\bar{y}$ such that for every $y\in\mathcal{O}$ every minimizing geodesic between x and y is contained in $\mathcal{V}$ with rank n. Let $v^1,\ldots v^n$ in $L^2([0,1],\mathbb{R}^m)$ be such that the linear operator

$$\begin{array}{ccc} \mathbb{R}^n & \longrightarrow & T_{\bar{y}}M \\ \alpha & \longmapsto & \sum_{i=1}^m \alpha_i D_{\bar{u}}E_{\mathcal{F}}^{x,1}\left(v^i\right) \end{array}$$

is invertible. By continuity of $u\mapsto D_uE_{\mathcal{F}}^{x,1}$, taking $\mathcal{O}$ smaller if necessary, we may assume that for every $y\in\mathcal{O}$ and for every minimizing geodesic γ_y from x to y associated with a control u^y, the linear operator

$$\begin{array}{ccc} \mathbb{R}^n & \longrightarrow & T_yM \\ \alpha & \longmapsto & \sum_{i=1}^m \alpha_i D_{u^y}E_{\mathcal{F}}^{x,1}\left(v^i\right) \end{array}$$

is invertible. For every $y \in \mathcal{O}$, define $\mathcal{F}^y : \mathbb{R}^n \to M$ by

$$\mathcal{F}^y(\alpha) := E_{\mathcal{F}}^{x,1}\left(u^y + \sum_{i=1}^m \alpha_i v^i\right) \qquad \forall \alpha \in \mathbb{R}^n.$$

This mapping is well-defined and smooth in a neighborhood of the origin, satisfies

$$\mathcal{F}^y(0) = y,$$

and its differential at 0 is invertible. Hence by the Inverse Function Theorem, there are an open neighborhood $\mathcal{B}^y$ of y in M and a function $\mathcal{G}^y : \mathcal{B}^y \to \mathbb{R}^n$ with $\mathcal{G}^y(y) = 0$ such that

$$\mathcal{F}^y \circ \mathcal{G}^y(z) = z \qquad \forall z \in \mathcal{B}^y.$$

From the definition of the sub-Riemannian distance between two points, we infer that for any $z \in \mathcal{B}^y$ we have

$$d_{SR}(x,z) = \sqrt{e_{SR}(x,z)} \leq \left\| u^y + \sum_{i=1}^m \left(\mathcal{G}^y(z)\right)_i v^i \right\|_{L^2} =: \phi^y(z).$$

We conclude that, for every $y \in \mathcal{O}$, there are a open set $\mathcal{B}^y$ containing y and a C^1 function $\phi^y : \mathcal{B}^y \to \mathbb{R}^n$ such that

$$d_{SR}(x,y) = \phi^y(y) \quad \text{and} \quad d_{SR}(x,z) \leq \phi^y(z) \quad \forall z \in \mathcal{B}^y.$$

The C^1 norms of the ϕ^y's are uniformly bounded. This proves the lemma. □

To conclude the proof of Theorem 2.3.5, we note that by the Rademacher Theorem, the function $y \in \mathcal{O}_{\bar{y}} \mapsto d_{SR}(x,y)$ is differentiable almost everywhere in $\mathcal{O}_{\bar{y}}$. By Lemma 2.3.6, for every $y \in \mathcal{O}_{\bar{y}}$ where the function is differentiable, there is $p_y \in T_x^*M$ such that

$$y = \exp_x\left(p_y\right), \quad d_{SR}(x,y) = \sqrt{2H\left(x,p_y\right)}, \quad \psi_{x,p_y}(1) = \left(y, \frac{1}{2} D_y d_{SR}^2(x,\cdot)\right).$$

Since $d_{SR}(x,\cdot)$ is Lipschitz in $\mathcal{O}_{\bar{y}}$, there is some constant $K > 0$ such that all the p_y's remain in a compact subset of T_x^*M. Now every $y \in \mathcal{O}_{\bar{y}}$ can be approximated by a sequence $\{y_k\}_k$ of points in $\mathcal{O}_{\bar{y}}$ where $d_{SR}(x,\cdot)$ is differentiable. By compactness, up to taking a subsequence, the normal extremals starting at (x, p_{y_k}) will converge to a normal extremal starting whose the projection is a minimizing geodesic from x to y. □

Remark 2.3.10. *We already know that the sub-Riemannian distance is continuous on $M \times M$ (see Proposition 1.5.2). The proof of Theorem 2.3.5 shows that if (Δ, g) is complete and $x \in M$ be fixed, then the function $y \in M \to d_{SR}(x,y) \in \mathbb{R}$ is locally Lipschitz (in charts) on an open and dense subset of M.*

2.4 The Goh condition

Theorem 2.2.2 provides firt-order conditions for a given horizontal path to be a minimizing geodesic. The aim of the present section is to present a second-order necessary condition for a given singular path to be minimizing. For sake of simplicity, we fix an orthonormal family $\mathcal{F} = \{X^1, \ldots, X^m\}$ of smooth vector fields in some open chart $\mathcal{V}$ which contains a minimizing geodesic $\bar{\gamma} : [0,1] \to M$ from x to y (with $x \neq y$). As before, we denote by $\bar{u} = u^{\bar{\gamma}}$ the control which is associated with $\bar{\gamma}$ through $\mathcal{F}$. Recall that $C : L^2([0,1];\mathbb{R}^m)$ is defined by

$$C(u) := \|u\|_{L^2}^2 \qquad \forall u \in L^2\big([0,1];\mathbb{R}^m\big).$$

Define $F : L^2([0,1];\mathbb{R}^m) \to \mathbb{R}^n \times \mathbb{R}$ by

$$F(u) := \Big(E_{\mathcal{F}}^{x,1}(u), C(u)\Big) \qquad \forall u \in L^2\big([0,1];\mathbb{R}^m\big).$$

The Lagrange Multiplier Theorem asserts that if $\bar{u}$ minimizes $C(u)$ under the constraint $E_{\mathcal{F}}^{x,1}(u) = y$, then there are $\lambda \in (\mathbb{R}^n)^*$ and $\lambda_0 \in \{0,1\}$ with $(\lambda, \lambda_0) \neq (0,0)$ such that

$$\lambda \cdot D_{\bar{u}} E_{\mathcal{F}}^{x,1} = \lambda_0 D_{\bar{u}} C.$$

In Section 2.2, we saw that whenever $\lambda_0 = 0$ we cannot deduce that $\bar{\gamma}$ satisfies the geodesic equation, that is that it is the projection of a normal extremal. In the case $\lambda_0 = 0$, the control $\bar{u} \in \mathcal{U}_{\mathcal{F}}^{x,1}$ is necessarily singular which means that it is a critical point of $E_{\mathcal{F}}^{x,1}$. Thus we have to study what happens at second order.

Let $\mathcal{U}$ be an open set in $L^2 = L^2([0,1];\mathbb{R}^m)$ and $F : \mathcal{U} \to \mathbb{R}^N$ be a function of class C^2 with respect to the L^2-norm. We recall that we call critical point of F any $u \in \mathcal{U}$ such that $D_u F : \mathcal{U} \to \mathbb{R}^N$ is not surjective. Given a critical point u, we call corank of u, the quantity

$$\mathrm{corank}_F(u) := N - \dim\big(\mathrm{Im}\big(D_u F\big)\big).$$

For every $u \in \mathcal{U}$ the second differential of F at u is the quadratic mapping on $D_u^2 F : L^2 \to \mathbb{R}^N$ satisfying

$$F(u+v) = F(u) + D_u F(v) + \frac{1}{2} D_u^2 F \cdot (v,v) + \|v\|_{L^2}^2\, o(1).$$

If $Q : L^2 \to \mathbb{R}$ is a quadratic form, we define its negative index by

$$\mathrm{ind}_-(Q) := \max\Big\{\dim(L) \mid Q_{|L\setminus\{0\}} < 0\Big\}.$$

We are now ready to state the result whose the proof is given in Appendix B.

Theorem 2.4.1. *Let $F : \mathcal{U} \to \mathbb{R}^N$ be a mapping of class C^2 in an open set $\mathcal{U} \subset L^2$ and $\bar{u} \in \mathcal{U}$ be a critical point of F of corank r. If*

$$ind_-\Big(\lambda^*\left(D_{\bar{u}}^2 F\right)_{|Ker(D_{\bar{u}}F)}\Big) \geq r \qquad \forall \lambda \in \big(Im\big(D_{\bar{u}}F\big)\big)^{\perp} \setminus \{0\}, \tag{2.23}$$

then the mapping F is locally open at $\bar{u}$, that is the image of any neighborhood of $\bar{u}$ is an neighborhood of $F(\bar{u})$.

From Proposition 1.3.3 and Remark 1.3.5, we know that for every non-zero form $\bar{p} \in (\mathbb{R}^n)^*$ with

$$\bar{p} \cdot D_{\bar{u}} E_{\mathcal{F}}^{x,1} = 0,$$

the absolutely continuous arc $\bar{p} : [0,1] \to (\mathbb{R}^n)^* \setminus \{0\}$ defined by

$$\bar{p}(t) := \bar{p} \cdot \bar{S}(1)\bar{S}(t)^{-1} \qquad \forall t \in [0,1], \tag{2.24}$$

satisfies $\bar{p}(1) = \bar{p}$,

$$\dot{\bar{p}}(t) = -\sum_{i=1}^{k} \bar{u}_i(t)\, \bar{p}(t) \cdot D_{\gamma_{\bar{u}}(t)} X^i \qquad \text{a.e. } t \in [0,T], \tag{2.25}$$

and

$$\bar{p}(t) \cdot X^i\big(\gamma_{\bar{u}}(t)\big) = 0 \qquad \forall t \in [0,T], \, \forall i = 1, \dots m, \tag{2.26}$$

where $\bar{S} : [0,T] \to M_n(\mathbb{R})$ is the solution to the Cauchy problem

$$\dot{\bar{S}}(t) = \bar{A}(t)\bar{S}(t) \quad \text{a.e. } t \in [0,T], \quad \bar{S}(0) = I_n, \tag{2.27}$$

and the matrices $\bar{A}(t) \in M_n(\mathbb{R}), \bar{B}(t) \in M_{n,k}(\mathbb{R})$ are defined by

$$\bar{A}(t) := \sum_{i=1}^{m} \bar{u}_i(t) J_{X^i}\big(\gamma_{\bar{u}}(t)\big) \qquad \text{a.e. } t \in [0,T] \tag{2.28}$$

and

$$\bar{B}(t) := \Big(X^1(\gamma_{\bar{u}}(t)), \cdots, X^m(\gamma_{\bar{u}}(t))\Big) \qquad \forall t \in [0,T]. \tag{2.29}$$

The following result combined with Theorem 2.4.1 will yield a necessary condition for a minimizing horizontal path to be strictly abnormal. It holds in the general case of a control $\bar{u}$ which belongs to $\mathcal{U}_{\mathcal{F}}^{x,1} \cap L^\infty\big([0,1];\mathbb{R}^m\big)$. We do not need $\bar{u}$ to be minimizing.

Theorem 2.4.2. *Let $\bar{u} \in \mathcal{U}_{\mathcal{F}}^{x,1} \cap L^\infty\big([0,1];\mathbb{R}^m\big)$ and $\bar{p} \in (\mathbb{R}^n)^* \setminus \{0\}$ be such that*

$$\bar{p} \cdot D_{\bar{u}} E_{\mathcal{F}}^{x,1} = 0. \tag{2.30}$$

Assume that

$$ind_- \left(\bar{p} \cdot \Big(D_{\bar{u}}^2 E_{\mathcal{F}}^{x,1}\Big)_{|Ker(D_{\bar{u}} E_{\mathcal{F}}^{x,1})}\right) < +\infty. \tag{2.31}$$

Then the absolutely continuous arc $\bar{p} : [0,1] \to (\mathbb{R}^n)^* \setminus \{0\}$ defined by (2.24) satisfies

$$\bar{p}(t) \cdot \big[X^i, X^j\big]\big(\gamma_{\bar{u}}(t)\big) = 0 \qquad \forall t \in [0,1], \, \forall i,j = 1, \dots, m. \tag{2.32}$$

Proof. Let us first check that F is of class C^2 on $\mathcal{U}_{\mathcal{F}}^{x,1}$ (we refer the reader to Appendix B for basics in differential calculus in infinite dimension). Given $u \in U_{\mathcal{F}}^{x,1}$ and $v \in L^2([0,T];\mathbb{R}^m)$ we need to study the quantity

$$\gamma_{u+\epsilon v}(1) - \gamma_u(1) = E_{\mathcal{F}}^{x,1}\big(u+\epsilon v\big) - E_{\mathcal{F}}^{x,1}\big(u\big),$$

at second order when ϵ is small. We have

$$\gamma_{u+\epsilon v}(1) = \int_0^1 \sum_{i=1}^k \left(u_i(t)+\epsilon v_i(t)\right) X^i\big(\gamma_{u+\epsilon v}(t)\big)dt, \tag{2.33}$$

with $\gamma_{u+\epsilon v}(0) = x$. For every $i = 1,\dots,m$ and every $t \in [0,1]$, the Taylor expansion of each X^i at $\gamma_u(t)$ at second order gives

$$\begin{aligned} &X^i\big(\gamma_{u+\epsilon v}(t)\big) = X^i\big(\gamma_u(t)\big) + D_{\gamma_u(t)}X^i \cdot \big(\gamma_{u+\epsilon v}(t) - \gamma_u(t)\big) \\ &+ \frac{1}{2} D^2_{\gamma_u(t)}X^i \cdot \big(\gamma_{u+\epsilon v}(t) - \gamma_u(t), \gamma_{u+\epsilon v}(t) - \gamma_u(t)\big) + \left|\gamma_{u+\epsilon v}(t) - \gamma_u(t)\right|^2 o(1). \end{aligned}$$

Setting $\delta_x(t) := \gamma_{u+\epsilon v}(t) - \gamma_u(t)$ for any t, (2.33) yields formally (δ_x has size ϵ, see the proof of Proposition 1.2.4)

$$\begin{aligned} \delta_x(1) = &\int_0^1 \left[\sum_{i=1}^m u_i(t) D_{\gamma_u(t)}X^i \cdot \delta_x(t) + \epsilon \sum_{i=1}^m v_i(t) X^i\big(\gamma_u(t)\big)\right] dt \\ &+ \int_0^1 \left[\epsilon \sum_{i=1}^m v_i(t) D_{\gamma_u(t)}X^i \cdot \delta_x(t) + \frac{1}{2}\sum_{i=1}^m u_i(t) D^2_{\gamma_u(t)}X^i \cdot \left(\delta_x(t), \delta_x(t)\right)\right] dt \\ &\qquad + \|v\|_\infty^2 o(1). \end{aligned}$$

Write $\delta_x(t)$ as

$$\delta_x(t) = \delta_x^1(t) + \delta_x^2(t) + o(\epsilon^2),$$

where δ_x^1 is linear in ϵ and δ_x^2 is quadratic in ϵ. Then δ_x^1 and δ_x^2 must satisfy

$$\dot{\delta}_x^1(t) = \left[\sum_{i=1}^m u_i(t) D_{\gamma_u(t)}X^i\right] \cdot \delta_x^1(t) + \left[\epsilon \sum_{i=1}^k v_i(t) X^i\big(\gamma_u(t)\big)\right] \qquad \text{a.e. } t \in [0,1]$$

and

$$\begin{aligned} \dot{\delta}_x^2(t) = &\left[\sum_{i=1}^m u_i(t) D_{\gamma_u(t)}X^i\right] \cdot \delta_x^2(t) + \left[\epsilon \sum_{i=1}^m v_i(t) D_{\gamma_u(t)}X^i\right] \cdot \delta_x^1(t) \\ &+ \frac{1}{2}\sum_{i=1}^m u_i(t) D^2_{\gamma_u(t)}X^i \cdot \big(\delta_x^1(t), \delta_x^1(t)\big) \qquad \text{a.e. } t \in [0,1]. \end{aligned}$$

Then using the notations of the proof of Proposition 1.3.3, we get for every $t \in [0,1]$,

$$\delta_x^1(t) = S(t)\int_0^t S(s)^{-1}B(s)v(s)\,ds \tag{2.34}$$

and

$$\delta_x^2(t) = S(t)\int_0^t S(s)^{-1}\left[C(s)+D(s)\right]\,ds \tag{2.35}$$

where

$$C(t) = \sum_{i=1}^m v_i(t) D_{\gamma_u(t)}X^i \cdot \delta_x^1(t) \tag{2.36}$$

and

$$D(t) = \frac{1}{2}\sum_{i=1}^m u_i(t) D^2_{\gamma_u(t)}X^i \cdot \big(\delta_x^1(t),\delta_x^1(t)\big). \tag{2.37}$$

Proceeding as in the proof of Proposition 1.2.4 (that is using Gronwall-type estimates), we have the following expansion:

$$\gamma_{u+\epsilon v}(1) = \gamma_u(1) + \delta_x^1(1) + \delta_x^2(1) + o\big(\epsilon^2\big).$$

Then we have

$$D_u^2 E_{\mathcal{F}}^{x,1}\cdot(v,v) = 2\int_0^1 S(1)S(t)^{-1}\left[C(t)+D(t)\right]\,dt \qquad \forall v\in L^2\big([0,1];\mathbb{R}^m\big)$$

and $E_{\mathcal{F}}^{x,1}$ is C^2 on $\mathcal{U}_{\mathcal{F}}^{x,1}$.
Let us now fix $\bar{u}\in\mathcal{U}_{\mathcal{F}}^{x,1}\cap L^\infty\big([0,1];\mathbb{R}^m\big)$ and $\bar{p}\in(\mathbb{R}^n)^*\setminus\{0\}$ such that (2.30) and (2.31) are satisfied and prove that (2.44) holds. Note that we have for every $v\in L^2\left([0,1];\mathbb{R}^m\right)$,

$$D_{\bar{u}}^2 E_{\mathcal{F}}^{x,1}\cdot(v,v) = 2\int_0^1 \bar{S}(1)\bar{S}(t)^{-1}\left[\bar{C}(t)+\bar{D}(t)\right]\,dt, \tag{2.38}$$

where $\bar{C},\bar{D}$ are obtained by replacing u by $\bar{u}$ in (2.34)-(2.37) and the definitions of $\bar{S},\bar{A},\bar{B}$ (see (2.27)-(2.29)).

Lemma 2.4.3. *There is $K>0$ such that for any $\bar{t},\delta>0$ with $[\bar{t},\bar{t}+\delta]\subset[0,1]$, there holds for every $v\in Ker\left(D_{\bar{u}}E_{\mathcal{F}}^{x,1}\right)$ with $Supp(v)\in[\bar{t},\bar{t}+\delta]$,*

$$\left|D_{\bar{u}}^2 E_{\mathcal{F}}^{x,1}\cdot(v,v) - \bar{Q}_{\bar{t},\delta}(v)\right| \le K\,\|v\|_{L^2}^2\,\delta^2, \tag{2.39}$$

where $\bar{Q}_{\bar{t},\delta}: L^2\left([0,1];\mathbb{R}^m\right)\to\mathbb{R}^n$ is defined by

$$\bar{Q}_{\bar{t},\delta}(v) := \int_{\bar{t}}^{\bar{t}+\delta} \bar{p}\big(\bar{t}\big)\cdot\sum_{i=1}^m v_i(t) D_{\bar{\gamma}(\bar{t})}X^i\left[\int_{\bar{t}}^t \sum_{j=1}^m v_j(s)X^j\big(\bar{\gamma}(\bar{t})\big)\,ds\right]dt, \tag{2.40}$$

for every $v\in L^2\left([0,1];\mathbb{R}^m\right)$.

Proof of Lemma 2.4.3. Let $\bar{t},\delta>0$ with $[\bar{t},\bar{t}+\delta]\subset[0,1]$ and $v\in\mathrm{Ker}\left(D_{\bar{u}}E_{\mathcal{F}}^{x,1}\right)$ with $\mathrm{Supp}(v)\in[\bar{t},\bar{t}+\delta]$ be fixed. By Remark 1.2.5, we have

$$\bar{S}(1)\int_0^1 \bar{S}(t)^{-1}\bar{B}(t)v(t)\,dt = 0.$$

Then (2.38) yields

$$\bar{p} \cdot \left(D_{\bar{u}}^2 E_{\mathcal{F}}^{x,1}\right)_{|\mathrm{Ker}\left(d_{\bar{u}} E_{\mathcal{F}}^{x,1}\right)}(v) = 2 \int_0^1 \bar{p}(t) \cdot \left[\bar{C}(t) + \bar{D}(t)\right] \, dt.$$

Setting

$$\bar{\delta}_x^1(t) := \bar{S}(t) \int_0^t \bar{S}(s)^{-1} \bar{B}(s) v(s) \, ds \qquad \forall t \in [0,1],$$

we have $\bar{\delta}_x^1(t) = 0$ for every $t \in \left[0, \bar{t}\right] \cup \left[\bar{t} + \delta, 1\right]$ and by Cauchy-Schwarz's inequality, we have for every $t \in \left[\bar{t}, \bar{t} + \delta\right]$,

$$\begin{aligned}
\left|\bar{\delta}_x^1(t)\right| &= \left| \int_{\bar{t}}^t \bar{S}(t) \bar{S}(s)^{-1} \bar{B}(s) v(s) \, ds \right| \\
&\leq \sup_{s \in [0,1]} \left\{ \left\| \bar{S}(t) \bar{S}(s)^{-1} \bar{B}(s) \right\| \right\} \sqrt{t - \bar{t}} \, \|v\|_{L^2} \\
&\leq K_1 \sqrt{\delta} \, \|v\|_{L^2},
\end{aligned}$$

where K_1 is a constant depending only upon the sizes of $\bar{S}, \bar{S}^{-1}, \bar{B}$ in a neighborhood of the curve $\gamma_{\bar{u}}([0,1])$. Then we have

$$\bar{D}(t) = 0 \qquad \forall t \in t \in \left[0, \bar{t}\right] \cup \left[\bar{t} + \delta, 1\right],$$

and

$$\left|\bar{D}(t)\right| \leq K_3 \, \delta \, \|v\|_{L^2}^2 \, \left\|\bar{u}\right\|_{L^\infty} \qquad \forall t \in \left[\bar{t}, \bar{t} + \delta\right],$$

which gives

$$\left| \int_0^1 \bar{p}(t) \cdot \bar{D}(t) \, dt \right| \leq K_4 \, \|v\|_{L^2}^2 \, \delta^2,$$

where K_3, K_4 are some constants depending on K_1, on the size of the $D^2 X^j$'s, $\bar{p}$ and $\|u\|_{L^\infty}$. Note that since we can write ($\bar{\gamma} = \gamma_{\bar{u}}$)

$$\begin{aligned}
& \bar{\delta}_x^1(t) - \int_{\bar{t}}^t \sum_{j=1}^m v_j(s) X^j\left(\bar{\gamma}(\bar{t})\right) ds \\
= \; & \bar{\delta}_x^1(t) - \int_{\bar{t}}^t \sum_{j=1}^m v_j(s) X^j\left(\bar{\gamma}(s)\right) ds \\
& \qquad + \int_{\bar{t}}^t \sum_{j=1}^m v_j(s) \left[X^j\left(\bar{\gamma}(s)\right) - X^j\left(\bar{\gamma}(\bar{t})\right)\right] \, ds \\
= \; & \int_0^t \bar{S}(t) \bar{S}(s)^{-1} \bar{B}(s) v(s) - \bar{B}(s) v(s) \, ds \\
& \qquad + \int_{\bar{t}}^t \sum_{j=1}^m v_j(s) \left[X^j\left(\bar{\gamma}(s)\right) - X^j\left(\bar{\gamma}(\bar{t})\right)\right] \, ds \\
= \; & \int_{\bar{t}}^t \left(\bar{S}(t) - \bar{S}(s)\right) \bar{S}(s)^{-1} \bar{B}(s) v(s) \, ds \\
& \qquad + \int_{\bar{t}}^t \sum_{j=1}^m v_j(s) \left[X^j\left(\bar{\gamma}(s)\right) - X^j\left(\bar{\gamma}(\bar{t})\right)\right] \, ds,
\end{aligned}$$

we have (since $\bar{u}$ belongs to $L^\infty\big([0,1];\mathbb{R}^m\big)$, $\bar{S}$ and $\bar{\gamma}$ are both Lipschitz)

$$\left|\bar{\delta}^1_x(t) - \int_{\bar{t}}^t \sum_{j=1}^m v_j(s) X^j\big(\bar{\gamma}(\bar{t})\big)\, ds\right| \leq K_2 \|v\|_{L^2}\, \delta^{\frac{3}{2}}, \quad \forall t \in t \in \big[\bar{t}, \bar{t}+\delta\big], \quad (2.41)$$

where K_1 is a constant depending only upon the sizes of $\bar{S}, \bar{S}^{-1}, \bar{B}$ and the Lipschitz constants of the X^j's in a neighborhood of the curve $\gamma_{\bar{u}}\left([0,1]\right)$. By (2.40), we have

$$\begin{aligned}
\int_0^1 \bar{p}(t)\cdot\bar{C}(t)\,dt - \bar{Q}_{\bar{t},\delta}(v) &= \int_{\bar{t}}^{\bar{t}+\delta} \bar{p}(t)\cdot\bar{C}(t)\,dt - \bar{Q}_{\bar{t},\delta}(v) \\
= \int_{\bar{t}}^{\bar{t}+\delta} \bar{p}(t)\cdot &\left(\sum_{i=1}^m v_i(t) D_{\bar{\gamma}(t)}X^i \cdot \bar{\delta}^1_x(t)\right. \\
&\left. - \sum_{i=1}^m v_i(t) D_{\bar{\gamma}(\bar{t})}X^i \cdot \left[\int_{\bar{t}}^t \sum_{j=1}^m v_j(s) X^j\big(\bar{\gamma}(\bar{t})\big)\,ds\right]\right) dt \\
= \int_{\bar{t}}^{\bar{t}+\delta} \bar{p}(t)\cdot &\left(\sum_{i=1}^m v_i(t) D_{\bar{\gamma}(t)}X^i\right)\cdot\left[\bar{\delta}^1_x(t) - \int_{\bar{t}}^t \sum_{j=1}^m v_j(s) X^j\big(\bar{\gamma}(\bar{t})\big)\,ds\right] dt.
\end{aligned}$$

By (2.41), we infer that

$$\left|\int_{\bar{t}}^{\bar{t}+\delta} \bar{p}(t)\cdot\bar{C}(t)\,dt - \bar{Q}_{\bar{t},\delta}(v)\right| \leq K_5\, \|v\|_{L^2}^2 \delta^2,$$

for some constant K_5 depending on the datas. All in all, we get

$$\left|\int_0^1 \bar{p}(t)\cdot\left[\bar{C}(t)+\bar{D}(t)\right]\,dt - \bar{Q}_{\bar{t},\delta}(v)\right| \leq K_6\, \|v\|_{L^2}^2 \delta^2,$$

for some constant K_6 depending on the datas. We conclude easily. $\square$

Returning to the proof of Theorem 2.4.2, we argue by contradiction and assume that (2.44) does not hold. Hence we assume that there are $\bar{t} \in (0,1)$ and $\bar{i} \neq \bar{j} \in \{1,\cdots,m\}$ such that

$$\begin{aligned}
N_{\bar{i},\bar{j}}(\bar{t}) &:= \bar{p}(\bar{t})\cdot\left[X^{\bar{i}}, X^{\bar{j}}\right]\big(\bar{\gamma}(\bar{t})\big) > 0 \\
&= \bar{p}(\bar{t})\cdot\Big(D_{\bar{\gamma}(\bar{t})}X^{\bar{j}}\cdot X^{\bar{i}}\big(\bar{\gamma}(\bar{t})\big) - D_{\bar{\gamma}(\bar{t})}X^{\bar{i}}\cdot X^{\bar{j}}\big(\bar{\gamma}(\bar{t})\big)\Big). \quad (2.42)
\end{aligned}$$

Let $\delta > 0$ such that $[\bar{t}, \bar{t}+\delta] \subset [0,1]$ and $\bar{Q}_{\bar{t},\delta} : L^2([0,1] \to \mathbb{R}^n$ be the mapping defined by (2.40). We observe that there holds for every $v \in L^2\big([0,1];\mathbb{R}^m\big)$,

$$\begin{aligned}
\bar{Q}_{\bar{t},\delta}(v) &= \int_{\bar{t}}^{\bar{t}+\delta}\int_{\bar{t}}^t \left[\sum_{i,j=1}^m v_i(t)v_j(s)\left(\bar{p}(\bar{t})\cdot D_{\bar{\gamma}(\bar{t})}X^i\cdot X^j\big(\bar{\gamma}(\bar{t})\big)\right)\right] ds\,dt \\
&= \int_{\bar{t}}^{\bar{t}+\delta}\int_{\bar{t}}^t \langle v(s), \bar{M}v(t)\rangle\, ds\,dt \\
&= \int_{\bar{t}}^{\bar{t}+\delta} \langle w(t), \bar{M}v(t)\rangle\, dt, \quad (2.43)
\end{aligned}$$

where $\bar{M}$ is the $m \times m$ matrix defined by

$$\bar{M}_{i,j} = \bar{p}(\bar{t}) \cdot D_{\bar{\gamma}(\bar{t})} X^i \cdot X^j \big(\bar{\gamma}(\bar{t})\big),$$

and

$$w(t) := \int_{\bar{t}}^{t} v(s)\, ds \qquad \forall t \in [0,1].$$

Thanks to Lemma 2.4.3, in order to get a contradiction, we need to show that for every integer $N > 0$, there are $\delta > 0$ and a subspace $L_\delta \subset L^2\big([0,1];\mathbb{R}^m\big)$ of dimension larger than N such that the restriction of $\bar{Q}_{\bar{t},\delta}$ to $L \setminus \{0\}$ satisfies the following property:

$$\bar{Q}_{\bar{t},\delta}(v) < -K\, \|v\|_{L^2}^2\, \delta^2 \qquad \forall v \in L \setminus \{0\}.$$

As a matter of fact, given $N \in \mathbb{N}$ strictly larger than n, if L is a vector subspace of dimension N, then the linear operator

$$\Big(D_{\bar{u}} E_{\mathcal{F}}^{x,1}\Big)_{|L} \,:\, L \longrightarrow \mathbb{R}^n$$

has a kernel of dimension at least $N - n$, which means that

$$\mathrm{Ker}\,\Big(D_{\bar{u}} E_{\mathcal{F}}^{x,1}\Big) \cap L$$

has dimension at least $N - n$.
Let N an integer strictly larger than n be fixed and $\delta > 0$ with $\big[\bar{t}, \bar{t}+\delta\big] \subset [0,1]$ to be chosen later. Denote by $L = L_{\bar{t},\delta,N}$ the vector space in $L^2\big([0,1];\mathbb{R}^m\big)$ of all the controls v such that there is a sequence $\{a_1, \ldots, a_N\}$ such that

$$\begin{cases} v_{\bar{i}}(t) &=& \sum_{k=1}^{N} a_k \cos\left(k\frac{(t-\bar{t})2\pi}{\delta}\right) \\ v_{\bar{j}}(t) &=& \sum_{k=1}^{N} a_k \sin\left(k\frac{(t-\bar{t})2\pi}{\delta}\right) \end{cases} \qquad \forall t \in \big[\bar{t}, \bar{t}+\delta\big],$$

$$v_{\bar{i}}(t) = v_{\bar{j}}(t) = 0 \qquad \forall t \notin \big[\bar{t}, \bar{t}+\delta\big],$$

and

$$v_i(t) = 0, \quad \forall i \neq \bar{i}, \bar{j} \qquad \forall t \in [0,1].$$

Let $v \in L \setminus \{0\}$, taking as before $w(t) := \int_{\bar{t}}^{t} v(s) ds$, we have

$$\begin{cases} w_{\bar{i}}(t) &=& \frac{\delta}{2\pi} \sum_{k=1}^{N} \frac{a_k}{k} \sin\left(k\frac{(t-\bar{t})2\pi}{\delta}\right) \\ w_{\bar{j}}(t) &=& \frac{\delta}{2\pi} \sum_{k=1}^{N} \frac{a_k}{k} \left(1 - \cos\left(k\frac{(t-\bar{t})2\pi}{\delta}\right)\right), \end{cases} \qquad \forall t \in \big[\bar{t}, \bar{t}+\delta\big],$$

$$w_{\bar{i}}(t) = w_{\bar{j}}(t) = 0 \qquad \forall t \notin \big[\bar{t}, \bar{t}+\delta\big],$$

and

$$w_i(t) = 0, \quad \forall i \neq \bar{i}, \bar{j} \qquad \forall t \in [0,1].$$

Then we have

$$\int_0^1 w_{\bar{i}}(t) v_{\bar{j}}(t)\, dt = \sum_{k=1}^{+\infty} \frac{\delta^2 a_k^2}{4\pi k}$$

and

$$\int_0^1 w_{\bar{j}}(t) v_{\bar{i}}(t)\, dt = -\sum_{k=1}^{+\infty} \frac{\delta^2 a_k^2}{4\pi k}.$$

We have for every $t \in [0,1]$

$$\begin{aligned}\langle w(t), \bar{M} v(t)\rangle = w_{\bar{i}}(t)\bar{M}_{\bar{i}\bar{i}} v_{\bar{i}}(t) + w_{\bar{i}}(t)\bar{M}_{\bar{i}\bar{j}} v_{\bar{j}}(t) \\ + w_{\bar{j}}(t)\bar{M}_{\bar{j}\bar{i}} v_{\bar{i}}(t) + w_{\bar{j}}(t)\bar{M}_{\bar{j}\bar{j}} v_{\bar{j}}(t).\end{aligned}$$

But

$$\int_0^1 w_{\bar{i}}(t)\bar{M}_{\bar{i}\bar{i}} v_{\bar{i}}(t)\, dt = \bar{M}_{\bar{i}\bar{i}} \int_0^1 w_{\bar{i}}(t)\dot{w}_{\bar{i}}(t)\, dt = 0 = \int_0^1 w_{\bar{j}}(t)\bar{M}_{\bar{j}\bar{j}} v_{\bar{j}}(t)\, dt.$$

In conclusion, we have

$$\bar{Q}_{\bar{t},\delta}(v) = \int_0^1 \langle w(t), \bar{M} v(t)\rangle\, dt = -N_{\bar{i},\bar{j}}\left(\bar{t}\right) \sum_{k=1}^{N} \frac{\delta^2 a_k^2}{4\pi k} = -\frac{\delta^2 N_{\bar{i},\bar{j}}\left(\bar{t}\right)}{4\pi} \sum_{k=1}^{N} \frac{a_k^2}{k}.$$

Since $N_{\bar{i},\bar{j}}\left(\bar{t}\right) > 0$, $\bar{Q}_{\bar{t},\delta}(v)$ is negative. Moreover, we observe that

$$\|v\|_{L^2}^2 = \delta \sum_{k=1}^{N} a_k^2,$$

which yields

$$\frac{\left|\bar{Q}_{\bar{t},\delta}(v)\right|}{\|v\|_{L^2}^2 \delta^2} = \frac{N_{\bar{i},\bar{j}}\left(\bar{t}\right)}{4\pi} \frac{\sum_{k=1}^{N} \frac{a_k^2}{k}}{\delta \sum_{k=1}^{N} a_k^2} \geq \frac{1}{\delta}\left(\frac{N_{\bar{i},\bar{j}}\left(\bar{t}\right)}{4\pi}\right).$$

We conclude easily by taking $\delta > 0$ small enough. □

A minimizing geodesic is called *strictly abnormal* if it is singular and admits no normal extremal lift. A control is called *strictly abnormal* if its associated horizontal path is strictly abnormal.

Theorem 2.4.4. *Let $\bar{\gamma} : [0,1] \to M$ be a minimizing geodesic from x to y (with $x \neq y$) which is strictly abnormal. Then there is an abnormal lift $\bar{\psi} = \left(\bar{\gamma}, \bar{p}\right) : [0,1] \to T^*M$ of $\bar{\gamma}$ such that*

$$\bar{p}(t) \cdot \left[X^i, X^j\right]\left(\bar{\gamma}(t)\right) = 0 \qquad \forall t \in [0,1],\, \forall i,j = 1,\ldots,m. \tag{2.44}$$

The latter property is called the Goh condition.

Proof. According to the previous notations, we define the mapping $F : \mathcal{U}_{\mathcal{F}}^{x,1} \to \mathbb{R}^n \times \mathbb{R}$ by

$$F(u) := \left(E_{\mathcal{F}}^{x,1}(u), \|u\|_{L^2}^2\right) \qquad \forall u \in \mathcal{U}_{\mathcal{F}}^{x,1}.$$

This function, which is of class C^2, cannot be open at $\bar{u}$. As a matter of fact, if the image of a neighborhood of $\bar{u}$ contains a neighborhood of $F(\bar{u})$ then it contains a control $u \in \mathcal{U}_{\mathcal{F}}^{x,1}$ with

$$E_{\mathcal{F}}^{x,1}(u) = y \quad \text{and} \quad \|u\|_{L^2}^2 \leq \left\|\bar{u}\right\|_{L^2}^2,$$

which contradicts the minimality of $\bar{u}$ from x to y. Therefore by Theorem 2.4.1 we infer that there is $\lambda \in \big(\mathrm{Im}(D_{\bar{u}}F)\big)^{\perp} \setminus \{0\}$ such that

$$\mathrm{ind}_{-}\Big(\lambda^{*}\left(D^2_{\bar{u}}F\right)_{|\mathrm{Ker}(d_{\bar{u}}F)}\Big) < r := n - \mathrm{rank}(\bar{u}). \tag{2.45}$$

Since the control $\bar{u}$ is strictly abnormal, the last coordinates of λ is zero. Denote by $\bar{p}$ the dual of the first n coordinates of λ. Then we have

$$\bar{p} \cdot D_{\bar{u}}E^{x,1}_{\mathcal{F}}(v) = 0 \qquad \forall v \in L^2\big([0,1];\mathbb{R}^m\big).$$

Since $\bar{u}$ is minimizing, $|u(t)|$ is constant and $\bar{u}$ belongs to $L^{\infty}\big([0,1];\mathbb{R}^m\big)$. Theorem 2.4.2 concludes the proof. $\square$

Example 2.4.5. *A distribution Δ is called medium-fat if, for every $x \in M$ and every section X of Δ with $X(x) \neq 0$, there holds*

$$T_xM = \Delta(x) + [\Delta,\Delta](x) + \big[X,[\Delta,\Delta]\big](x), \tag{2.46}$$

where

$$\big[\Delta,\Delta\big](x) := \Big\{[X,Y](x) \,|\, X, Y \text{ sections of } \Delta\Big\}$$

$$\text{and} \quad \big[X,[\Delta,\Delta]\big](x) := \Big\{\big[X,[Y,Z]\big](x) \,|\, Y, Z \text{ sections of } \Delta\Big\}.$$

Any two-generating distribution is medium-fat. An example of medium-fat distribution which is not two-generating is given by the rank-three distribution in $\mathbb{R}^4$ with coordinates $x = (x_1,x_2,x_3,x_4)$ defined by

$$\Delta(x) = \Big\{X^1(x), X^2(x), X^3(x)\Big\} \qquad \forall x \in \mathbb{R}^4,$$

with

$$X^1 = \partial_{x_1}, \quad X^2 = \partial_{x_2}, \quad X^3 = \partial_{x_3} + (x_1+x_2+x_3)^2\partial_{x_4}.$$

Medium-fat distribution do not admit non-trivial Goh paths. As a matter of fact, if $\gamma : [0,T] \to M$ is an horizontal path which admits an abnormal lift $\psi = (\gamma,p) : [0,T] \to T^*M$ satisfying the Goh condition, then we have

$$p(t) \cdot \big[X^i, X^j\big]\big(\gamma(t)\big) = 0 \qquad \forall i,j = 1,\ldots,m, \tag{2.47}$$

for every t in a small interval $I \subset [0,T]$ such that $\gamma(t)$ is in a local chart of M and Δ is parametrized by a family $\{X^1,\ldots,X^m\}$ of smooth vector fields. Then if we denote by u the control which is associated to γ through $\mathcal{F}$, derivating the previous equality yields for any $i,j = 1,\ldots,m$,

$$p(t) \cdot \left[\sum_{k=1}^{m} u_k(t)X^k, \big[X^i,X^j\big]\right]\big(\gamma(t)\big) = 0 \qquad \forall t \in I. \tag{2.48}$$

Since $\psi = (\gamma,p)$ is an abnormal lift, we also have $p \cdot X^i = 0$ along γ, then by (2.46), (2.47)-(2.48) we get a contradiction.

2.5 Examples of SR geodesics

Geodesics in the Heisenberg group

The Heisenberg group $\mathbb{H}^1$ is the sub-Riemannian structure (Δ, g) in $\mathbb{R}^3$ where Δ is the totally nonholonomic rank 2 distribution (see Example 1.1.2) spanned by the vector fields

$$X = \partial_x - \frac{y}{2}\partial_z \quad \text{and} \quad Y = \partial_y + \frac{x}{2}\partial_z,$$

and g is the metric making the family $\{X, Y\}$ orthonormal, that is defined by

$$g = dx^2 + dy^2.$$

The above structure can be shown to be left-invariant under the group law

$$(x, y, z) \star (x', y', z') = \left(x + x', y + y', z + z' + \frac{1}{2}\big(xy' - x'y\big) \right).$$

Thanks to Proposition 1.2.1, any horizontal path on $[0, T]$ has the form $\gamma_u = (x, y, z) : [0, T] \to \mathbb{R}^3$ where

$$\left\{ \begin{array}{lcl} \dot{x}(t) & = & u_1(t) \\ \dot{y}(t) & = & u_2(t) \\ \dot{z}(t) & = & \frac{1}{2}\left(u_2(t)x(t) - u_1(t)y(t)\right), \end{array} \right. \tag{2.49}$$

for some $u \in L^2\big([0, T]; \mathbb{R}^2\big)$. This means that

$$z(T) - z(0) = \int_0^T \frac{1}{2}\left(x(t)\dot{y}(t) - y(t)\dot{x}(t)\right) dt = \int_c \frac{1}{2}\left(xdy - ydx\right),$$

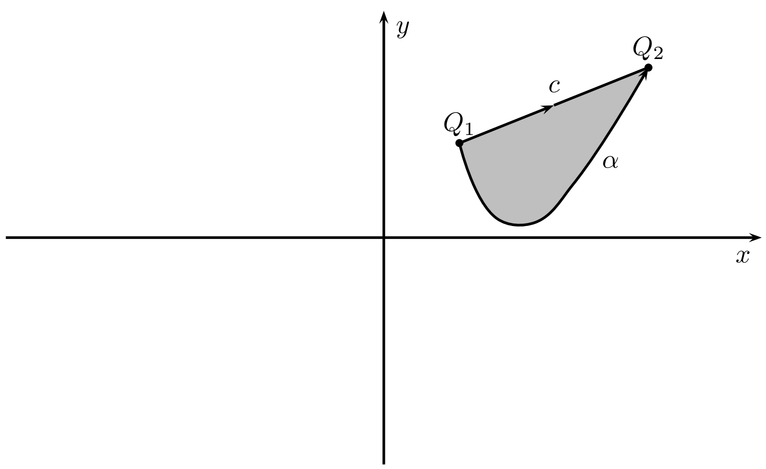
where $\alpha(t) = (x(t), y(t))$ is the projection of the curve γ to the plane. According to the Stockes Theorem, we have

$$\int_\alpha \frac{1}{2}\left(xdy - ydx\right) = \int_{\mathcal{D}} dx \wedge dy + \int_c \frac{1}{2}\left(xdy - ydx\right),$$

where $\mathcal{D}$ denotes the domain which is enclosed by the curve α and the segment

$$c := \big[Q_1, Q_2\big] := \big[(x(0, y(0)), (x(T), y(T)\big]$$

from Q_1 to Q_2.



Therefore, given two points $P_1 = (x_1, y_1, z_1), P_2 = (x_2, y_2, z_2)$ in $\mathbb{R}^3$, the horizontal paths which minimizes the length from P_1 to P_2 are the curves $\gamma : [0,1] \to \mathbb{R}^3$ whose the signed aread of $\mathcal{D}$ satisfies

$$\int_{\mathcal{D}} dx \wedge dy = (z_2 - z_1) - \int_c \frac{1}{2} (xdy - ydx),$$

with minimal length. According to the isoperimetric inequality, the curves in the plane sweeping the same area and which minimize the length are given by circles. This fact can be easily recovered by Theorem 2.2.2 and Proposition 2.2.1 (we saw in Example 1.3.12 that Δ admits no non-trivial singular horizontal paths). Assume that $\gamma_u = (x, y, z) : [0,1] \to \mathbb{R}^3$ is a minimizing geodesic from $P_1 := \gamma_u(0)$ to $P_2 := \gamma_u(1) \neq P_1$. Then according to Proposition 2.2.1, there is a smooth arc $p = (p_1, p_2, p_3) : [0,1] \to (\mathbb{R}^3)^*$ such that the following system of differential equations holds

$$\left\{ \begin{array}{lll} \dot{x} & = & p_x - \frac{y}{2} p_z \\ \dot{y} & = & p_y + \frac{x}{2} p_z \\ \dot{z} & = & \frac{1}{2} \left(\left(p_y + \frac{x}{2} p_z \right) x - \left(p_x - \frac{y}{2} p_z \right) y \right), \end{array} \right. \qquad \left\{ \begin{array}{lll} \dot{p}_x & = & - \left(p_y + \frac{x}{2} p_z \right) \frac{p_z}{2} \\ \dot{p}_y & = & \left(p_x - \frac{y}{2} p_z \right) \frac{p_z}{2} \\ \dot{p}_z & = & 0. \end{array} \right.$$

Hence $p_z = \bar{p}_z$ for every t. Which implies that

$$\ddot{x} = -\bar{p}_z \dot{y} \quad \text{and} \quad \ddot{y} = \bar{p}_z \dot{x}.$$

If $\bar{p}_z = 0$, then the geodesic from P_1 to P_2 is a segment with constant speed. If $\bar{p}_z \neq 0$, we have or

$$\dddot{x} = -\bar{p}_z^2 \dot{x} \quad \text{and} \quad \dddot{y} = -\bar{p}_z^2 \dot{y}.$$

Which means that the curve $t \mapsto (x(t), y(t))$ is a circle.

A singular minimizing geodesic

As we said above, minimizing geodesics do not necessarily satisfy the Hamiltonian geodesic equation. As an example, consider the Martinet distribution (which already appeared in Examples 1.1.21 and 1.3.14) in $\mathbb{R}^3$ with coordinates (x_1, x_2, x_3) defined by

$$\Delta(x) = \operatorname{Span}\Big\{ X(x), Y(x) \Big\} \qquad \forall x \in \mathbb{R}^3,$$

with

$$X = \partial_{x_1}, \quad Y = \partial_{x_2} + \frac{x_1^2}{2} \partial_{x_3},$$

and equipped with a smooth metric g. In Example 1.3.14, we saw that singular curves are given by the horizontal paths which are contained in the Martinet set

$$\Sigma_\Delta = \Big\{ x = 0 \Big\},$$

that is of the form

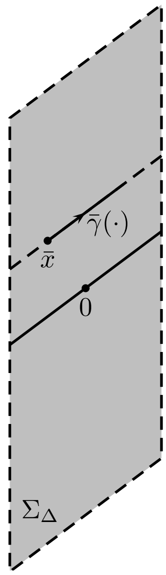
$$x(t) = \left(0, x_2(0) + \int_0^t u_2(s) ds, 0, x_3(0) \right),$$

with $u_2 \in L^2([0,T]; \mathbb{R})$. Such curves are locally minimizing.

Theorem 2.5.1. *For any $\bar{x} = (0, \bar{x}_2, \bar{x}_3) \in \mathbb{R}^3$, there is $\bar{\epsilon} > 0$ such that for every $\epsilon \in (0, \bar{\epsilon})$ the horizontal path given by*

$$\bar{\gamma}(t) = \big(0, \bar{x}_2 + t, \bar{x}_3\big) \qquad \forall t \in [0, \epsilon],$$

minimizes the length among all horizontal paths joining $(0, \bar{x}_2, \bar{x}_3)$ to $(0, \bar{x}_2 + \epsilon, \bar{x}_3)$.



Proof. Let $\bar{x} = (0, \bar{x}_2, \bar{x}_3) \in \mathbb{R}^3$ be fixed. It is more convenient to work with an orthonormal frame that we now construct. In the sequel all the constructions are performed in an open neighborhood of $\bar{x}$ that we always denote by $\mathcal{V}$. First, there is a smooth function $\lambda : \mathcal{V} \to (0, +\infty)$ such that

$$|\lambda(x) Y(x)|_x^g = 1 \qquad \forall x \in \mathcal{V}. \tag{2.50}$$

Set $\tilde{Y} := \lambda Y$ and pick a smooth section $\tilde{X}$ of Δ such that

$$\left|\tilde{X}(x)\right|_x^g = 1 \quad \text{and} \quad g_x\left(\tilde{X}(x), Y(x)\right) = 0 \qquad \forall x \in \mathcal{V}. \tag{2.51}$$

Since for any $x \in \Sigma_\Delta \cap \mathcal{V}$ the vector $Y(x)$ is tangent to Σ_Δ and $\Delta(x)$ is transverse to Σ_Δ, the vector $\tilde{X}(x)$ is necessarily transverse to Σ_Δ. Let us perform a change of coordinates. For this, consider the diffeomorphism given by (it is indeed defined in an open neighborhood of $\bar{x}$)

$$\begin{array}{rcl} \Phi : \mathbb{R} \times \Sigma_\Delta & \longrightarrow & \mathbb{R}^3 \\ \big(s, x_2, x_3\big) & \longmapsto & e^{s\tilde{X}}\big(0, x_2, x_3\big), \end{array}$$

and set

$$\hat{X} := \Phi^* \tilde{X} \quad \text{and} \quad \hat{Y} = \Phi^* \tilde{Y}$$

and $\hat{g}$ the metric g in the new set of coordinates, that is

$$\hat{g}_x(v, w) = g_{\Phi(x)}\big(D_x\Phi(v), D_x\Phi(w)\big).$$

By construction (by (2.50)-(2.51)), $\{\hat{X},\hat{Y}\}$ is a local orthonormal frame with respect to g and there are smooth functions $\phi_1,\phi_2,\phi_3:\mathcal{V}\to\mathbb{R}$ such that

$$\hat{X}=\partial_{x_1} \tag{2.52}$$

$$\text{and}\qquad \hat{Y}=x_1\phi_1(x)\partial_{x_1}+(1+x_1\phi_2(x))\,\partial_{x_2}+\left(\frac{x_1^2}{2}+x_1\phi_3(x)\right)\partial_{x_3}.$$

The Martinet set Σ_Δ is invariant under this change of coordinates and we check that

$$\left[\hat{X},\hat{Y}\right](x)=\left(\phi_1(x)+x_1\frac{\partial\phi_1}{\partial x_1}(x)\right)\partial_{x_1}+\left(\phi_2(x)+x_1\frac{\partial\phi_2}{\partial x_1}(x)\right)\partial_{x_2}$$
$$+\left(x_1+\phi_3(x)+x_1\frac{\partial\phi_3}{\partial x_1}(x)\right)\partial_{x_3}.$$

Then $\phi_3(x)$ vanishes on Σ_Δ with a non-vanishing derivative. Morever the derivative of $\det(\hat{X},\hat{Y},[\hat{X},\hat{Y}])$ on Σ_Δ with respect to the x_1 variable is equal to

$$\frac{\det(\hat{X},\hat{Y},[\hat{X},\hat{Y}])}{\partial x_1}(x)=1+2\frac{\partial\phi_3}{\partial x_1}(x)\neq 0.$$

This means that $\hat{Y}$ has indeed the form

$$\hat{Y}=x_1\phi_1(x)\partial_{x_1}+(1+x_1\phi_2(x))\,\partial_{x_2}+x_1^2\varphi(x)\partial_{x_3}, \tag{2.53}$$

with $\varphi(0)\neq 0$. Up to dilate, we may assume that $\varphi(0)=1$. We need to show that among all controls $u=(u_1,u_2):[0,\tau]\to\mathbb{R}^2$ with $u_1^2+u_2^2\leq 1$ steering $\bar{x}=\big(0,\bar{x}_2,\bar{x}_3\big)$ to $\big(0,\bar{x}_2+\epsilon,\bar{x}_3\big)$, we have $\epsilon<\tau$. It is sufficient to prove the result for $\bar{x}=0$, we set $P:=(0,\epsilon,0)$. There is $r>0$ such that $\bar{B}_{SR}(0,r)$ is included in $\mathcal{V}$. If $\epsilon\in(0,r)$, then any minimizing geodesic joining 0 to $(0,\epsilon,0)$ is contained in $\bar{B}_{SR}(0,r)$. As a matter of fact, we know that $d_{SR}(0,P)\leq\epsilon<r$. Let $C_1,C_2>0$ be upper bounds for ϕ_1,ϕ_2 on $\bar{B}_{SR}(0,r)$ and $\delta>0$ be such that

$$|\varphi(x)-1|\leq\delta\qquad\forall x\in\bar{B}_{SR}(0,r). \tag{2.54}$$

Let $\gamma_u=x:[0,\tau]\to\mathbb{R}^3$ be a competitor for $\bar{\gamma}$. Note that thanks to (2.52)-(2.53), then end-point conditions give

$$\left\{\begin{array}{lcl} x_1(\tau) & = & \int_0^\tau u_1(s)\,ds+\int_0^\tau u_2(s)x_1(s)\phi_1\big(x(s)\big)\,ds=0\\ x_2(\tau) & = & \int_0^\tau u_2(s)\,\big(1+x_1(s)\phi_2\big(x(s)\big)\big)\,ds=\epsilon\\ x_3(\tau) & = & \int_0^\tau u_2(s)x_1(s)^2\varphi\big(x(s)\big)\,ds=0.\end{array}\right. \tag{2.55}$$

Set

$$\beta:=\max\Big\{\big|x_1(s)\big|\,|\,s\in[0,\tau]\Big\}. \tag{2.56}$$

Note that if $\gamma_u\neq\bar{\gamma}$, then β is necessarily positive. Taking $r>0$ smaller if necessary (and a fortiori $\epsilon>0$ smaller), we may assume that

$$\beta\leq\frac{1}{2C_2},\qquad\delta\leq\frac{1}{2},\qquad\sqrt{1+\beta^2\delta^2}\leq 2. \tag{2.57}$$

The last equation in (2.55) yields (by (2.54)-(2.56))

$$\begin{aligned}
\int_0^\tau x_1(s)^2\,ds &\leq \int_0^\tau x_1(s)^2\big(1-u_2(s)\big)\,ds + \int_0^\tau x_1(s)^2 u_2(s)\,ds \\
&\leq \beta^2\left(\tau - \int_0^\tau u_2(s)\,ds\right) + \int_0^\tau x_1(s)^2\left(1-\varphi\big(x(s)\big)\right)u_2(s)\,ds \\
&\leq \beta^2\left(\tau - \int_0^\tau u_2(s)\,ds\right) + \delta\int_0^\tau x_1(s)^2\,ds.
\end{aligned}$$

Therefore (by (2.57))

$$\int_0^\tau x_1(s)^2\,ds \leq \frac{\beta^2}{1-\delta}\left(\tau - \int_0^\tau u_2(s)\,ds\right) \leq 2\beta^2\left(\tau - \int_0^\tau u_2(s)\,ds\right). \quad (2.58)$$

Let $\bar{s} \in [0,\tau]$ be such that $|x_1(\bar{s})| = \beta$. Since

$$\left|\dot{x}_1(s)\right| \leq |u_1(s) + u_2(s)x_1(s)\phi_1(x(s))| \leq \sqrt{1+\beta^2\delta^2} \leq 2$$

for almost every $s \in [0,\tau]$ and $x_1(0) = x_1(\tau) = 0$, we have

$$\bar{s}, \tau - \bar{s} \geq \frac{\beta}{\sqrt{1+\beta^2\delta^2}} \geq \beta/2.$$

Which means that the interval $[\bar{s}-\beta/2, \bar{s}+\beta/2]$ is included in $[0,\tau]$ and

$$\left|x_1(s)\right| \geq \frac{\beta}{2} \qquad \forall s \in \big[\bar{s}-\beta/4, \bar{s}+\beta/4\big].$$

Therefore we have

$$\int_0^\tau \left|x_1(s)\right|^2 ds \geq \int_{\bar{s}-\beta/4}^{\bar{s}+\beta/4} \left|x_1(s)\right|^2 ds \geq \frac{\beta^3}{8}.$$

By (2.58), we deduce that

$$\frac{\beta^3}{8} \leq 2\beta^2\left(\tau - \int_0^\tau u_2(s)\,ds\right)$$

which implies

$$\int_0^\tau u_2(s)\,ds \leq \tau - \frac{\beta}{16}. \quad (2.59)$$

Then by the second line in (2.55) and the definitions of β and C_2, we have

$$\begin{aligned}
\epsilon &= \int_0^\tau u_2(s)\,ds + \int_0^\tau u_2(s)x_1(s)\phi_2\big(x(s)\big)\,ds \\
&\leq \int_0^\tau u_2(s)\,ds + \int_0^\tau \left|u_2(s)\right|\left|x_1(s)\right|\left|\phi_2\big(x(s)\big)\right|\,ds \\
&\leq \int_0^\tau u_2(s)\,ds + \beta C_2\tau.
\end{aligned}$$

Consequently by (2.59), we get

$$\epsilon \leq \tau + \beta\left(C_2\tau - \frac{1}{16}\right).$$

In conclusion, if $\beta > 0$ and $\tau < 1/(16C_2)$ (that is $\bar{\epsilon} > 0$ small enough), then τ cannot be smaller than ϵ. This shows the result. $\square$

Remark 2.5.2. *Theorem 2.5.1 only asserts that $\bar{\gamma}$ minimizes the length between its end-points. It is not necessarily a geodesic, or equivalently it has not necessarily constant speed. This latter property depends upon the metric g.*

From the proof of Theorem 2.5.1, we have local coordinates (x_1, x_2, x_3) in an open neighborhood of $0 \in \mathbb{R}^3$ such that the Martinet distribution equipped with a smooth metric g admits a local orthonormal frame of the form

$$\Delta(x) = \mathrm{Span}\Big\{\hat{X}(x), \hat{Y}(x)\Big\} \qquad \forall x \in \mathbb{R}^3,$$

with

$$\hat{X} = \partial_{x_1}, \quad \hat{Y} = x_1\phi_1(x)\partial_{x_1} + (1 + x_1\phi_2(x))\,\partial_{x_2} + x_1^2\varphi(x)\partial_{x_3},$$

and $\varphi(0) = 1$. According to Proposition 2.2.1, for every $\bar{p} = (\bar{p}_1, \bar{p}_2, \bar{p}_3) \in (\mathbb{R}^3)^*$, the normal extremal (with respect to g) on $[0,1]$ starting at $(0,p)$ is the trajectory $(x,p) : [0,1] \to \mathbb{R}^3 \times (\mathbb{R}^3)^*$ satisfying

$$\left\{\begin{array}{rcl} \dot{x}_1 &=& p_1 + \Big(p \cdot \hat{Y}(x)\Big)\, x_1\phi_1(x) \\ \dot{x}_2 &=& \Big(p \cdot \hat{Y}(x)\Big)\,(1 + x_1\phi_2(x)) \\ \dot{x}_3 &=& \Big(p \cdot \hat{Y}(x)\Big)\, x_1^2\varphi(x), \end{array}\right. \tag{2.60}$$

$$\left\{\begin{array}{rcl} \dot{p}_1 &=& -\Big(p \cdot \hat{Y}(x)\Big)\,\Big[p_1\big(\phi_1(x) + x_1\frac{\partial\phi_1}{\partial x_1}(x)\big) \\ && \quad \ldots + p_2\big(\phi_2(x) + x_1\frac{\partial\phi_2}{\partial x_1}(x)\big) + p_3\big(2x_1\varphi(x) + x_1^2\frac{\partial\varphi}{\partial x_1}(x)\big)\Big] \\ \dot{p}_2 &=& -\Big(p \cdot \hat{Y}(x)\Big)\,\Big[p_1x_1\frac{\partial\phi_1}{\partial x_2}(x) + p_2x_1\frac{\partial\phi_2}{\partial x_2}(x) + p_3x_1^2\frac{\partial\varphi}{\partial x_2}(x)\Big] \\ \dot{p}_3 &=& -\Big(p \cdot \hat{Y}(x)\Big)\,\Big[p_1x_1\frac{\partial\phi_1}{\partial x_3}(x) + p_2x_1\frac{\partial\phi_2}{\partial x_3}(x) + p_3x_1^2\frac{\partial\varphi}{\partial x_3}(x)\Big], \end{array}\right. \tag{2.61}$$

with

$$p \cdot \hat{Y}(x) = p_1x_1\phi_1(s) + p_2\big(1 + x_1\phi_2(x)\big) + p_3x_1^2\varphi(x)$$

and

$$x_1(0) = x_2(0) = x_3(0) = 0, \quad p_1(0) = \bar{p}_1, \quad p_2(0) = \bar{p}_2, \quad p_3(0) = \bar{p}_3. \tag{2.62}$$

Note that if $\phi_1 \equiv \phi_2 \equiv 0$ and $\varphi \equiv 1$, that is whenever $g = dx_1^2 + dx_2^2$, then the horizontal path given by

$$\bar{\gamma}(t) = \big(0, t, 0\big) \qquad \forall t \in [0, \epsilon], \tag{2.63}$$

is the projection of the normal extremal starting at $(0, \bar{p})$ with $\bar{p} = (0, 1, 0)$. Then whenever $\phi_1 \equiv \phi_2 \equiv 0$ and $\varphi \equiv 1$ and for $\epsilon > 0$ small enough, a reparametrization of $\bar{\gamma}$ is a singular normal minimizing geodesic between its end-points (see Example 1.3.14 and Theorem 2.2.2). Different choices of metrics can provide examples of strictly abnormal minimizing geodesics.

Proposition 2.5.3. *If $\phi_2(0) \neq 0$, then any reparametrization of $\bar{\gamma}$ given by (2.63) is not the projection of a normal extremal.*

Proof. We argue by contradiction and assume that there is $\bar{p} = (\bar{p}_1, \bar{p}_2, \bar{p}_3) \in (\mathbb{R}^3)^*$ and $\hat{\gamma} : [0,1] \to \mathbb{R}^3$ a reparametrization of $\bar{\gamma}$ such that the systems differential equations (2.60)-(2.61) are satisfied with $x = \hat{\gamma}$ and initial conditions (2.62). The system (2.60)-(2.61) is the Hamiltonian system which is associated with the Hamiltonian given by

$$H(x,p) = p \cdot \hat{X}(x) + p \cdot \hat{Y}.$$

Since H is constant along its extremals and $x_1(t) = x_3(t) = 0$ for any $t \in [0,\epsilon]$, we have (see (2.7))

$$\begin{aligned} \Big(p(t) \cdot \hat{X}\big(\bar{\gamma}(t)\big)\Big)^2 + \Big(p(t) \cdot \hat{Y}\big(\bar{\gamma}(t)\big)\Big)^2 &= p_1(t)^2 + p_2(t)^2 \\ &= \bar{p}_1^2 + \bar{p}_2^2 \qquad \forall t \in [0,1]. \end{aligned}$$

On the other hand, since $x_1(t) = 0$ for every $t \in [0,1]$, the second and third equations in (2.61) yield

$$\dot{p}_2 = \dot{p}_3 = 0 \quad \Longrightarrow \quad p_2(t) = \bar{p}_2 \quad \forall t \in [0,1].$$

Moreover, (2.60) also gives $\dot{x}_2 = \bar{p}_2$ that is $\bar{p}_2 \neq 0$ ($\hat{\gamma}$ has constant speed). Since p_1 is smooth and both p_2 and $p_1^2 + p_2^2$ are constant, p_1 is necessarily constant. The first equation in (2.60) and $x_1 = 0$ give $\dot{x}_1 = p_1$. Hence $p_1 = \bar{p}_1 = 0$. Then, using that $p_1 = 0, p_2 = \bar{p}_2 \neq 0$, the first equation in (2.61) gives

$$\bar{p}_2 \phi_2\big(\hat{\gamma}(t)\big) = 0 \qquad \forall t \in [0,1].$$

By assumption on $\phi_2(0)$, we deduce that $\bar{p}_2 = 0$. Since we know that $\hat{\gamma}$ joins 0 to $(0,\epsilon,0)$ with $\epsilon \neq 0$, this contradicts the equality $\dot{x}_2 = \bar{p}_2$. $\square$

2.6 Notes and comments

Theorem 2.2.2 may be seen as a weak form of the Pontryagin maximum principle which has been developed by the russian school of control in the 60s. In the general context of optimal control theory, the strong form of the Pontryagin maximum principle provides necessary conditions for a control to be optimal. For further details on this topics, we refer the reader to the seminal book by Pontryagin and its collaborators [PBGM] and to the more recent textbooks by Agrachev and Sachkov [AS04], Clarke [Cla83], or Vinter [Vin00]. The material presented in Sections 2.1 and 2.2 is by now classical. It can be found in the Montgomery textbook [Mon02] which also provide many references.

Theorem 2.3.5 about the image of the sub-Riemannian exponential map has been proven by Agrachev and the author, see [Agr09]. It extends a previous density result, based on Lemma 2.3.6, which was obtained by Trélat and the author in [RT05]. Given a complete sub-Riemannian structure (Δ, g) on a smooth manifold M and $x \in M$, we do not know if the image of $\exp_x$ has full Lebesgue measure in M. This open problem is indeed "contained" in the sub-Riemannian Sard conjecture. Given $x \in M$ (which is equipped with a SR structure), denote by $\mathcal{S}_\Delta^{x,1}$ the set of singular horizontal paths in $\Omega_\Delta^{x,1}$ (that is $\mathcal{S}_\Delta^{x,1} := \Omega_\Delta^{x,1} \setminus \mathcal{R}_\Delta^{x,1}$ with the notations of Chapter 1). The SR Sard conjecture states that the image of $\mathcal{S}_\Delta^{x,1}$ by the End-Point mapping

$$\begin{array}{rcl} E_\Delta^{x,T} : \Omega_\Delta^{x,1} & \longrightarrow & M \\ \gamma & \longmapsto & \gamma(1), \end{array}$$

has Lebesgue measure zero in M. We even do not know if $E_{\Delta}^{x,T}$ can have a non-empty interior in M. We refer the reader to Montgomery's book [Mon02] for further details on the SR Sard Conjecture and to the paper [RT05] for various sub-Riemannian Sard-like conjectures.

The theory of second variation for singular geodesics in sub-Riemannian geometry has been developed by Agrachev and Sarychev [AS96]. The results and proofs that we present in Section 2.4 are taken from Agrachev-Sarychev's paper [AS99]. Example 2.4.5 (medium-fat distributions) is taken from [AS99] as well.

For decades the prevailing wisdom was that every sub-Riemannian minimizing geodesic is normal, meaning that it admits a normal extremal lift. In 1991, Montgomery [Mon94] found the first counter-example to this assertion. We refer the reader to Montgomery's book [Mon02] for an historical account on the existence of strictly abnormal minimizing geodesics. The second example which is presented in Section 2.5 is the Montgomery counter-example. The proof of local minimality of characteristic lines in the Martinet surface (Theorem 2.5.1) is taken from the monograph by Liu and Sussmann [LS95]. Note that the Montgomery counter-example as well as all other known counter-examples exhibit smooth singular minimizing curves. The existence of non-smooth sub-Riemannian geodesics is open.

In the first example of Section 2.5, we briefly explained that the sub-Riemannian structure under study was indeed left-invariant under some group law. This additional structure makes $\mathbb{H}^1$ a Carnot group. We refer the reader to the Montgomery textbook [Mon02] or to the Jean monograph [Jea12] in the present volume for further details on Carnot groups.

Chapter 3

Introduction to optimal transport

Throughout all the chapter, M denotes a smooth connected manifold without boundary of dimension $n \geq 2$.

3.1 The Monge and Kantorovich problems

The Monge problem

Let

$$c : M \times M \to [0, +\infty)$$

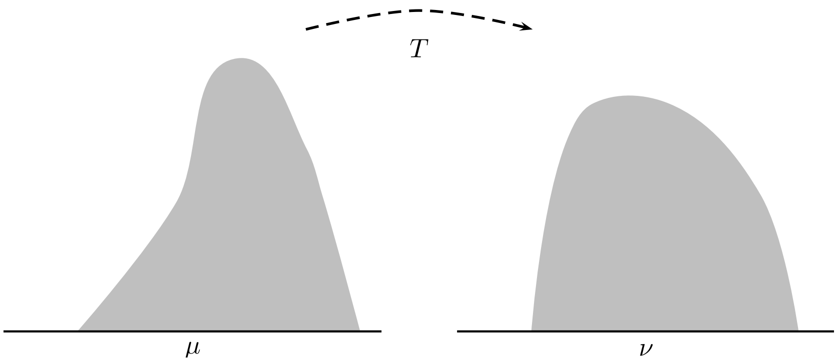
be a cost function and μ, ν be two probability measures on M. We recall that a probability measure on M is a Borel measure with total mass 1. The *Monge optimal transport problem* from μ to ν with respect to the cost c consists in minimizing the transportation cost

$$\int_M c\big(x, T(x)\big)\, d\mu(x), \tag{3.1}$$

among all the measurable maps $T : M \to M$ pushing forward μ to ν (we denote it by $T_\sharp \mu = \nu$) that is satisfying

$$\mu\big(T^{-1}(B)\big) = \nu(B) \qquad \forall B \text{ measurable set in } M. \tag{3.2}$$

Such maps are called *transport maps* from μ to ν.



We set

$$C_{\mathcal{M}}(\mu,\nu) := \inf \left\{ \int_M c\big(x,T(x)\big)\, d\mu \,|\, T_\sharp \mu = \nu \right\}, \tag{3.3}$$

where $T_\sharp \mu = \nu$ means implicitely that T is a measurable map from M to itself which pushes forward μ to ν.

Remark 3.1.1. *The property (3.2) is equivalent to*

$$\int_M \varphi(T(x))\, d\mu(x) = \int_M \varphi(y)\, d\nu(y),$$

for all ν-integrable function φ. If $M = \mathbb{R}^n$ and μ and ν are absolutely continuous with respect to the Lebesgue measure respectively with densities f and g in $L^1(\mathbb{R}^n;[0,+\infty))$, the latter property can be written as

$$\int_M \varphi(T(x)) f(x)\, dx = \int_M \varphi(y) g(y)\, dy,$$

for any $\varphi \in L^\infty(\mathbb{R}^n;\mathbb{R})$. Therefore, if T is a diffeomorphism, then the change of variable $y = T(x)$ yields the Monge-Ampère equation

$$\big|\det\big(D_x T\big)\big| = \frac{f(x)}{g(T(x))} \qquad \mu - a.e.\ x \in \mathbb{R}^n. \tag{3.4}$$

Example 3.1.2. *Transport maps may not exist. For example, consider in $\mathbb{R}^n$ the probability measures μ, ν given by*

$$\mu = \delta_x \quad \text{and} \quad \nu = \frac{1}{2}\delta_{y_1} + \frac{1}{2}\delta_{y_2},$$

where $x, y_1, y_2 \in \mathbb{R}^n$, $y_1 \neq y_2$ and δ_a denotes the Dirac mass at some point $a \in \mathbb{R}^n$. There are no transport maps from μ to ν. If such a map T exists, then

$$\frac{1}{2} = \nu\big(\{y_1\}\big) = \mu\left(T^{-1}\big(\{y_1\}\big)\right) = 0 \text{ or } 1,$$

which is impossible.

Example 3.1.3. *Minimizers of Monge's problem may not be unique. On the real line $\mathbb{R}$, consider the probability measures μ and ν given by*

$$\mu = 1_{[0,1]}\,\mathcal{L}^1 \quad \text{and} \quad \nu = 1_{[1,2]}\,\mathcal{L}^1,$$

where $\mathcal{L}^1$ denotes the Lebesgue measure in $\mathbb{R}$. In other terms, μ and ν are respectively the restriction of the Lebesgue measure on the intervals $[0,1]$ and $[1,2]$. The two maps $T_1, T_2 : \mathbb{R} \to \mathbb{R}$ given by

$$T_1(x) = x + 1 \quad \text{and} \quad T_2(x) = 2 - x \qquad \forall x \in \mathbb{R},$$

push forward μ to ν. This is a straightforward consequence of the fact that both T_1 and T_2 are affine maps which are bijective from $[0,1]$ to $[1,2]$ with determinant 1 together with a change of variable (see Remark 3.1.1). Consider the Monge cost $c : \mathbb{R} \times \mathbb{R} \to [0,+\infty)$ given by

$$c(x,y) := |y - x| \qquad \forall x, y \in \mathbb{R}.$$

We check easily that the transportation cost for T_1 and T_2 are given by

$$\int_{\mathbb{R}} c\big(x, T_i(x)\big)\, d\mu(x) = \int_0^1 |T_i(x) - x|\ dx = 1 \qquad i = 1, 2.$$

Furthermore, we also check that if T is a map which pushes forward μ to ν, then

$$\begin{aligned}
\int_{\mathbb{R}} c\big(x, T(x)\big)\, d\mu(x) &= \int_0^1 |T(x) - x|\ dx \\
&= \int_0^1 [T(x) - x]\ dx = \int_0^1 T(x)\, dx - \int_0^1 x\, dx \\
&= \int_1^2 y\, dy - \int_0^1 x\, dx = 1.
\end{aligned}$$

This shows that the infimum in the definition of $C_{\mathcal{M}}(\mu, \nu)$ is attained by all transport maps from μ to ν. So, it is not unique.

The constraint $T_\sharp \mu = \nu$ being highly non-linear, the Monge optimal transport problem is quite difficult from the viewpoint of optimization. That is why we will study a notion of weak solution for this problem.

The Kantorovitch relaxation

Given two probability measures μ, ν on M, we denote by $\Pi(\mu, \nu)$ the set of probability measures α in the product $M \times M$ with first and second marginals μ and ν, that is such that

$$\pi^1_\sharp \alpha = \mu \quad \text{and} \quad \pi^2_\sharp \alpha = \nu, \tag{3.5}$$

where $\pi^i : M \times M \to M$ denotes respectively the projection on the first and second variable in $M \times M$. The *Kantorovitch optimal transport problem* with respect to the cost $c : M \times M \to [0, +\infty)$ consists in minimizing the quantity

$$C(\alpha) := \int_{M \times M} c(x, y)\, d\alpha(x, y), \tag{3.6}$$

among all the $\alpha \in \Pi(\mu, \nu)$. Any measure in $\alpha \in \Pi(\mu, \nu)$ is called a *transport plan* between μ and ν. We set

$$C_{\mathcal{K}}(\mu, \nu) := \inf \Big\{ C(\alpha) \,|\, \alpha \in \Gamma(\mu, \nu) \Big\}. \tag{3.7}$$

Remark 3.1.4. *The property (3.5) is equivalent to*

$$\mu(B) = \alpha\big(B \times M\big) \quad \textit{and} \quad \nu(B) = \alpha\big(M \times B\big),$$

for any measurable set B in M, which is also equivalent to

$$\int_{M \times M} [\varphi_1(x) + \varphi_2(y)]\ d\alpha(x, y) = \int_M \varphi_1(x)\, d\mu(x) + \int_M \varphi_2(y)\, d\nu(y),$$

for all μ-integrable function φ_1 and ν-integrable function φ_2. In particular, the set $\Pi(\mu, \nu)$ is a convex set which always contains the product measure $\mu \times \nu$.

Remark 3.1.5. *If $T : M \to M$ is a transport map from μ to ν then the measure α on $M \times M$ given by*

$$\alpha := (Id \times T)_\sharp \, \mu,$$

is a transport plan between μ and ν. This means that the Kantorovitch optimization problem is more general than the Monge optimization problem, or

$$C_{\mathcal{K}}(\mu, \nu) \leq C_{\mathcal{M}}(\mu, \nu),$$

for all probability measures μ, ν on M.

Example 3.1.6. *Returning to Example 3.1.2, we note that the product measure*

$$\alpha = \frac{1}{2}\delta_{(x,y_1)} + \frac{1}{2}\delta_{(x,y_2)},$$

is a transport plan between μ and ν. In contrary to Monge's transport maps, Kantorovitch's transport plans allow splitting of mass.

The Kantorovitch optimal transport problem is an infinite-dimensional optimization problem which involves a functional C which is linear in α and a set of constraints $\Pi(\mu, \nu)$ which is convex and weakly compact. The existence of optimal transport plans becomes easy.

3.2 Optimal plans and Kantorovitch potentials

Optimal plans

Throughout this section, we fix a cost $c : M \times M \to [0, +\infty)$. We recall that the support $\mathrm{spt}(\mu)$ of a measure μ refers to the smallest closed set $F \subset M$ of full mass $\mu(F) = \mu(M) = 1$.

Theorem 3.2.1. *Let μ, ν be two probability measures on M. Assume that c is continuous and that Supp(μ) and Supp(ν) are compact. Then the Kantorovitch optimal transport problem admits at least one solution, that is there is $\bar{\alpha} \in \Pi(\mu, \nu)$ such that*

$$C(\bar{\alpha}) = C_{\mathcal{K}}(\mu, \nu) := \inf \Big\{ C(\alpha) \, | \, \alpha \in \Gamma(\mu, \nu) \Big\}. \tag{3.8}$$

Proof. We first note that $C_{\mathcal{K}}(\mu, \nu)$ is finite. As a matter of fact, since the product measure $\mu \times \nu$ belongs to $\Pi(\mu, \nu)$ and c is bounded on $\mathrm{Supp}(\mu) \times \mathrm{Supp}(\nu)$ (by assumption c is continuous and $\mathrm{Supp}(\mu), \mathrm{Supp}(\nu)$ are compact), we have $C_{\mathcal{K}}(\mu, \nu) \leq C(\mu \times \nu) < +\infty$. In fact, the supports of all transport plans between μ and ν are contained in the set $\mathrm{Supp}(\mu) \times \mathrm{Supp}(\nu) \subset M \times M$ which is compact by assumption on $\mathrm{Supp}(\mu)$ and $\mathrm{Supp}(\nu)$. Then we can assume without loss of generality that M is compact. Denote by $\mathcal{P}(M \times M)$ the set of probability measures on $M \times M$ and define $F : \mathcal{P}(M \times M) \to \mathbb{R}$ by

$$F(\alpha) := \int_{M \times M} c(x, y) \, d\alpha(x, y) \qquad \forall \alpha \in \mathcal{P}(M \times M).$$

The functional F is continuous on $\mathcal{P}(M \times M)$ equipped with the topology of weak convergence, that is for any sequence $\{\alpha_k\}_k$ and any α in $\mathcal{P}(M \times M)$ satisfying

$$\int_{M\times M} \varphi(x,y)\, d\alpha_k(x,y) \longrightarrow_{k\to+\infty} \int_{M\times M} \varphi(x,y)\, d\alpha(x,y),$$

for any measurable function $\varphi : M \to \mathbb{R}$ which is bounded, we have

$$\lim_{k\to+\infty} F(\alpha_k) = F(\alpha).$$

This fact is a straigthforward consequence of the continuity of c together with the compactness of $M \times M$. By Prokhorov's Theorem, the set of probability measures on $M\times M$ is compact with respect to weak convergence. We conclude easily. Let $\{\alpha_k\}_k$ be a sequence in $\Pi(\mu,\nu)$ such that

$$C_{\mathcal{K}}(\mu,\nu) = \lim_{k\to+\infty} C(\alpha_k).$$

By Prokhorov's Theorem, up to taking a subsequence, we may assume that $\{\alpha_k\}$ converges to some probability measure $\bar{\alpha}$. By Remark 3.1.4, $\bar{\alpha}$ belongs to $\Pi(\mu,\nu)$. Moreover it satisfies $C(\bar{\alpha}) = C_{\mathcal{K}}(\mu,\nu)$ by continuity of F. $\square$

The supports of optimal transport plans have specific properties. Let us introduce the notion of c-cyclically monotone sets.

Definition 3.2.2. *A subset $S \subset M \times M$ is called c-cyclically monotone if for any finite number of points $(x_j,y_j) \in S, j = 1,\ldots,J$, and σ a permutation on the set $\{1,\ldots,J\}$,*

$$\sum_{j=1}^{J} c(x_j,y_j) \leq \sum_{j=1}^{J} c\left(x_{\sigma(j)},y_j\right).$$

Remark 3.2.3. *The definition given above is equivalent to the following one: for any finite number of points $(x_j,y_j) \in S, j = 1,\ldots,J$,*

$$\sum_{j=1}^{J} c(x_j,y_j) \leq \sum_{j=1}^{J} c\left(x_j,y_{j+1}\right),$$

with $y_{J+1} = y_1$. The equivalence is a straightforward consequence of the decomposition of a permutation into disjoint commuting cycles.

Remark 3.2.4. *If c is assumed to be continuous, the c-cyclical monotonocity is stable under closure. The closure of a c-cyclically monotone set is c-cyclically monotone.*

Given two probability measures μ,ν on M, we call *optimal transport plan* between μ and ν any $\alpha \in \Pi(\mu,\nu)$ satisfying $C_{\mathcal{K}}(\mu,\nu) = C(\alpha)$. Optimal transport plans always have c-cyclically monotone supports.

Theorem 3.2.5. *Let μ,ν be two probability measures on M. Assume that c is continuous and that Supp(μ) and Supp(ν) are compact. Then there is a c-cyclically monotone compact set $\mathcal{S} \subset$ Supp$(\mu) \times$ Supp(ν) such that the support of any optimal transport plan between μ and ν is contained in $\mathcal{S}$.*

Proof. Let us first show that the supports of optimal transport plans are always c-cyclically monotone. We argue by contradiction and assume that there is an optimal transport plan $\alpha \in \Pi(\mu,\nu)$ whose the support is not c-cyclically monotone. Then there is an integer $J>1$, J points $(x_1,y_1),\dots,(x_J,y_J)$ in $\mathrm{Supp}(\alpha)$ and a permutation σ on the set $\{1,\dots,J\}$ such that

$$\sum_{j=1}^J c(x_j,y_j) > \sum_{j=1}^J c\left(x_{\sigma(j)},y_j\right).$$

By continuity of c, there are open sets U_j,V_j for $j=1,\dots,J$ which contain respectively x_j,y_j such that

$$\sum_{j=1}^J c(u_j,v_j) > \sum_{j=1}^J c\left(u_{\sigma(j)},v_j\right) \qquad \forall \big((u_j,v_j)\big)_{j=1,\dots,J} \in \Pi_{j=1}^J \big(U_j\times V_j\big). \tag{3.9}$$

Each (x_j,y_j) belongs to the support of α, then we have $\alpha(U_j\times V_j)>0$. Define the probability measure P on $\Pi_{j=1}^J\big(U_j\times V_j\big)$ by

$$P=\Pi_{j=1}^J\left[\frac{1}{\alpha\left(U_j\times V_j\right)}1_{U_j\times V_j}\alpha\right].$$

It is a product of probability measures, hence it is a probability measure as well. Set

$$\bar m := \min\Big\{\alpha\big(U_j\times V_j\big) \,|\, j=1,\dots,J\Big\},$$

denote by π^{U_j} (resp. π^{V_j}) the projection from $\Pi_{j=1}^J\big(U_j\times V_j\big)$ to U_j (resp. to V_j) and define the measure $\tilde\alpha$ on $M\times M$ by

$$\tilde\alpha = \alpha + \frac{\bar m}{J}\left[\sum_{j=1}^J\Big(\big(\pi^{U_{\sigma(j)}},\pi^{V_j}\big)_\sharp P - \big(\pi^{U_j},\pi^{V_j}\big)_\sharp P\Big)\right].$$

We have

$$\begin{aligned}
\tilde\alpha &\geq \alpha - \frac{\bar m}{J}\sum_{j=1}^J\big(\pi^{U_j},\pi^{V_j}\big)_\sharp P\\
&= \alpha - \frac{1}{J}\sum_{j=1}^J\frac{\bar m}{\alpha\left(U_j\times V_j\right)}1_{U_j\times V_j}\alpha\\
&\geq \alpha - \frac{1}{J}\sum_{j=1}^J 1_{U_j\times V_j}\alpha \geq \alpha-\alpha=0.
\end{aligned}$$

Moreover

$$\pi^1_\sharp\left(\sum_{j=1}^J\big(\pi^{U_j},\pi^{V_j}\big)_\sharp P\right) = \sum_{j=1}^J \pi^{U_j}_\sharp P = \sum_{j=1}^J \pi^{U_{\sigma(j)}}_\sharp P = \pi^1_\sharp\left(\sum_{j=1}^J\big(\pi^{U_{\sigma(j)}},\pi^{V_j}\big)_\sharp P\right)$$

and

$$\pi^2_\sharp\left(\sum_{j=1}^J\big(\pi^{U_j},\pi^{V_j}\big)_\sharp P\right) = \sum_{j=1}^J \pi^{V_j}_\sharp P = \pi^2_\sharp\left(\sum_{j=1}^J\big(\pi^{U_{\sigma(j)}},\pi^{V_j}\big)_\sharp P\right).$$

Therefore α is a non-negative measure which belongs to $\Pi(\mu,\nu)$. But by construction, we have

$$\int_{M\times M} c(x,y)\, d\tilde{\alpha}(x,y) = \int_{M\times M} c(x,y)\, d\alpha(x,y) + \\ \frac{\bar{m}}{J}\int \sum_{j=1}^{J}\Big(c\big(u_{\sigma(j)},v_j\big) - c\big(u_j,v_j\big)\Big)\, dP\left(\left(u_1,v_1,\right),\dots,\left(u_J,v_J\right)\right),$$

and the last term is negative (by (3.9)). This means that α cannot be optimal and gives a contradiction. Then we know that the supports of any optimal transport plan between μ and ν is c-cyclically monotone. Denote by $\Pi^{opt}(\mu,\nu)$ the set of optimal transport plans in $\Pi(\mu,\nu)$ and set

$$\mathcal{S} := \bigcup_{\alpha\in\Pi^{opt}(\mu,\nu)} \mathrm{Supp}(\alpha).$$

By construction, $\mathcal{S}$ is a subset of $\mathrm{Supp}(\mu)\times\mathrm{Supp}(\nu)\subset M\times M$ which contains the supports of all optimal transport plans. It remains to show that $\mathcal{S}$ is c-cyclically monotone. Let $(x_1,y_1),\dots,(x_J,y_J)$ be J points in $\mathcal{S}$ and σ be a permutation on the set $\{1,\dots,J\}$. For each $j=1,\dots,J$ the point (x_j,y_j) belongs to the support of an optimal transport plan α_j. Let $\bar{\alpha}$ be the convex combination of the α_j's, that is

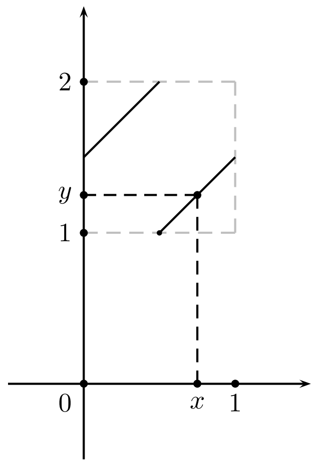
$$\alpha := \frac{1}{J}\sum_{j=1}^{J}\alpha_j.$$

Since $\Pi(\mu,\nu)$ is convex and the mapping $\alpha\mapsto C(\alpha)$ is linear, $\bar{\alpha}$ belongs to $\Pi^{opt}(\mu,\nu)$. Then its support is c-cyclically monotone and contains all the (x_j,y_j)'s. We infer that

$$\sum_{j=1}^{J} c(x_j,y_j) \le \sum_{j=1}^{J} c\left(x_{\sigma(j)},y_j\right).$$

We conclude by Remark 3.2.4. □

Example 3.2.6. *Returning to Example 3.1.3, we can show that the set provided by Theorem 3.2.5 has to be $\mathcal{S}=[0,1]\times[1,2]=Supp(\mu)\times Supp(\nu)$. As a matter of fact, for every $(x,y)\in[0,1]\times[1,2]$ there is a bijective function $T:[0,1]\to[1,2]$ which is lower semicontinuous, increasing and piecewise affine with slope 1, and whose the graph contains (x,y). Thanks to the observation we did in Example 3.1.3, such a function is a transport map from $\mu=1_{[0,1]}\,\mathcal{L}^1$ to $\nu=1_{[1,2]}\,\mathcal{L}^1$, hence it is optimal.*



Kantorovitch potentials

The aim of this section is to characterize c-cyclically monotone sets in a more analytic way.

Definition 3.2.7. *A function $\psi : M \to \mathbb{R} \cup \{+\infty\}$, not identically $+\infty$, is said to be c-convex if there is a non-empty set $\mathcal{A} \subset M \times \mathbb{R}$ such that*

$$\psi(x) := \sup\Big\{\lambda - c(x,y) \,|\, (y,\lambda) \in \mathcal{A}\Big\} \qquad \forall x \in M. \tag{3.10}$$

The c-transform of ψ, denoted by ψ^c is the function $\psi^c : M \to \mathbb{R} \cup \{-\infty\}$ defined by

$$\psi^c(y) := \inf\Big\{\psi(x) + c(x,y) \,|\, x \in M\Big\} \qquad \forall y \in M. \tag{3.11}$$

The pair (ψ, ψ^c) is called a c-pair of potentials.

The following result shows that the opposite of a c-convex function is the c-transform of the opposite of its c-transform.

Proposition 3.2.8. *Given a c-convex function ψ, the function $-\psi^c$ is c-convex and we have*

$$\psi(x) = \sup\Big\{\psi^c(y) - c(x,y) \,|\, y \in M\Big\} \qquad \forall x \in M. \tag{3.12}$$

Proof. By definition of ψ^c we have

$$\psi^c(y) - c(x,y) \leq \psi(x) \qquad \forall x \in M, \forall y \in M.$$

Which implies that $\psi(x) \geq \sup_{y\in M}\{\psi^c(y) - c(x,y)\}$ for any $x \in M$. Let us show that $\psi(x) \leq \sup_{y\in M}\{\psi^c(y) - c(x,y)\}$ for any $x \in M$. Argue by contradiction and assume that there is $\bar{x} \in M$ such that

$$\psi\big(\bar{x}\big) > \sup\Big\{\psi^c(y) - c\big(\bar{x}, y\big)\, y \in M\Big\}.$$

Since ψ is c-convex, there are a set $\mathcal{A} \subset M \times \mathbb{R}$, $(\bar{y}, \bar{\lambda}) \in \mathcal{A}$ and $\delta > 0$ such that

$$\bar{\lambda} - c(\bar{x}, \bar{y}) + \delta \geq \psi(\bar{x}) \geq \sup\Big\{\psi^c(y) - c(\bar{x}, y) \,|\, y \in M\Big\} + 3\delta.$$

Then we get

$$\psi^c(\bar{y}) \leq \bar{\lambda} - 2\delta.$$

which by definition of $\psi^c(\bar{y})$ implies that there is $x \in M$ such that

$$\psi(x) + c(x, \bar{y}) \leq \bar{\lambda} - \delta.$$

This contradicts (3.10). □

Example 3.2.9. *If $M = \mathbb{R}^n$ and c is given by $c(x, y) = |y - x|$, then the c-convex functions are exactly the functions which are 1-Lipschitz on $\mathbb{R}^n$. As a matter of fact, if $f : \mathbb{R}^n \to \mathbb{R}$ is 1-Lipschitz then for every $x \in \mathbb{R}^n$,*

$$f(x) \geq f(y) - |y - x| \qquad \forall y \in \mathbb{R}^n,$$

which yields

$$f(x) = \sup\Big\{f(y) - c(x, y) \,|\, y \in \mathbb{R}^n\Big\}.$$

Moreover, f is its own c-transform. Conversely, any c-convex function is a supremum of 1-Lipschitz function which is not identically $+\infty$. Then it is finite everywhere and 1-Lipschitz.

Example 3.2.10. *If $M = \mathbb{R}^n$ and c is given by $c(x, y) = |y - x|^2/2$, then the c-convex functions are the functions $\psi : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ such that the function*

$$x \in \mathbb{R}^n \longmapsto \psi(x) + \frac{1}{2}|x|^2$$

is convex. As a matter of fact, any c-convex function can be written as

$$\psi(x) = \sup\Big\{\lambda - \frac{|y|^2}{2} - \langle x, y\rangle \,|\, (y, \lambda) \in \mathcal{A}\Big\} - \frac{|x|^2}{2} \qquad \forall x \in \mathbb{R}^n.$$

which shows that $\psi + |\cdot|^2/2$ is convex as a supremum of affine functions. Conversely, any convex function on $\mathbb{R}^n$ can be expressed as the supremum of affine functions. That is given a convex function $\varphi : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$, there is a set $\mathcal{B} \subset \mathbb{R}^n \times \mathbb{R}$ such that

$$\varphi(x) = \sup\Big\{\langle x, y\rangle + \beta \,|\, (y, \beta) \in \mathcal{B}\Big\} \qquad \forall x \in \mathbb{R}^n.$$

Then for every $x \in \mathbb{R}^n$,

$$\varphi(x) - \frac{1}{2}|x|^2 = \sup\left\{\left(\beta + \frac{|y|^2}{2}\right) - \frac{|y - x|^2}{2} \,|\, (y, \beta) \in \mathcal{B}\right\},$$

which shows that $\psi := \varphi - |\cdot|^2/2$ is c-convex.

The c-cyclically monotone sets are the sets which are contained in the c-subdifferential of c-convex functions.

Definition 3.2.11. *Let $\psi : M \to \mathbb{R} \cup \{+\infty\}$ be a c-convex function. For every $x \in M$, the c-subdifferential of ψ at x is defined by*

$$\partial_c \psi(x) := \Big\{ y \in M \,|\, \psi^c(y) = \psi(x) + c(x,y) \Big\}.$$

We call contact set of the pair (ψ, ψ^c) the set defined by

$$\partial_c \psi := \Big\{ (x,y) \in M \times M \,|\, y \in \partial_c \psi(x) \Big\}.$$

Remark 3.2.12. *By the above definitions, a pair (x,y) in $M \times M$ belongs to $\partial_c \psi$ if and only if*

$$\psi(x) + c(x,y) \leq \psi(z) + c(z,y) \qquad \forall z \in M,$$

which is also equivalent to

$$\psi^c(y) - c(x,y) \geq \psi^c(z) - c(x,z) \qquad \forall z \in M.$$

In particular, both $\psi(x)$ and $\psi^c(y)$ are finite.

The following result is the cornerstone of the results of existence and uniqueness of optimal transport maps that we will present in the next sections.

Theorem 3.2.13. *For $S \subset M \times M$ to be c-cyclically monotone, it is necessary and sufficient that $S \subset \partial_c \psi$ for some c-convex $\psi : M \to \mathbb{R} \cup \{+\infty\}$. In fact, for every c-cyclically monotone set $S \subset M \times M$, there is a c-pair of potentials (ψ, ψ^c) with $S \subset \partial_c \psi$ satisfying*

$$\psi(x) = \sup \Big\{ \psi^c(y) - c(x,y) \,|\, y \in \pi^2(S) \Big\} \qquad \forall x \in M, \tag{3.13}$$

$$\psi^c(y) = \inf \Big\{ \psi(x) + c(x,y) \,|\, x \in \pi^1(S) \Big\} \qquad \forall y \in M. \tag{3.14}$$

If c is continuous and S is compact, then both ψ, ψ^c are valued in $\mathbb{R}$ and continuous, and the infimum and supremum in (3.13)-(3.14) are attained.

Proof. First, given a c convex function $\psi : X \to \mathbb{R} \cup \{+\infty\}$ the contact set of (ψ, ψ^c) is c-cyclically monotone. As a matter of fact, given $(x_j, y_j) \in \partial_c \psi, j = 1, \ldots, J$, and σ a permutation on the set $\{1, \ldots, J\}$, we have

$$\psi^c\big(y_j\big) = \psi\big(x_j\big) + c\big(x_j, y_j\big) \quad \text{and} \quad \psi^c\big(y_j\big) \leq \psi\big(x_{\sigma(j)}\big) + c\big(x_{\sigma(j)}, y_j\big),$$

for every $j = 1, \ldots, J$. Hence

$$\begin{aligned} \sum_{j=1}^J c\big(x_j, y_j\big) &= \sum_{j=1}^J \psi^c\big(y_j\big) - \sum_{j=1}^J \psi\big(x_j\big) \\ &= \sum_{j=1}^J \psi^c\big(y_j\big) - \sum_{j=1}^J \psi\big(x_{\sigma(j)}\big) \\ &\leq \sum_{j=1}^J c\big(x_{\sigma(j)}, y_j\big). \end{aligned}$$

Let us now show that a c-cyclically monotone set $S \subset M \times M$ is necessarily included in the contact set of some c-convex function. Fix $(\bar{x}, \bar{y})$ in the c-cyclically monotone set $S \subset M \times M$ and define $\psi : M \to \mathbb{R} \cup \{+\infty\}$ by

$$\begin{aligned}\psi(x) := \sup\Big\{ & \big[c(\bar{x},\bar{y}) - c(x_1,\bar{y})\big] \\ & + \sum_{j=1}^{J-1} \big[c(x_j,y_j) - c(x_{j+1},y_j)\big] + \big[c(x_J,y_J) - c(x,y_J)\big] \\ & \qquad \mid J \in \mathbb{N}, J \geq 2, (x_j,y_j) \in S, \forall j = 1,\ldots,J\Big\},\end{aligned}$$

for every $x \in M$. We claim that ψ is a c-convex function whose the contact set contains S. First taking $J = 2$, $x = x_1 = x_2 = \bar{x}$ and $y_1, y_2 = \bar{y}$, we check easily that $\psi(\bar{x}) \geq 0$. Furthermore, by c-cyclical monotonicity of S, we have

$$\begin{aligned}& \big[c(\bar{x},\bar{y}) - c(x_1,\bar{y})\big] \\ & \quad + \sum_{j=1}^{J-1} \big[c(x_j,y_j) - c(x_{j+1},y_j)\big] + \big[c(x_J,y_J) - c(\bar{x},y_J)\big] \leq 0,\end{aligned}$$

for any pairs $(x_1,y_1),\ldots,(x_J,y_J)$ belonging to S. Thus we have $\psi(\bar{x}) \leq 0$ and in turn $\psi(\bar{x}) = 0$. This shows that ψ is not identically $+\infty$. Define $\phi : M \to \mathbb{R} \cup \{-\infty\}$ by

$$\begin{aligned}\phi(y) := \sup\Big\{ & \big[c(\bar{x},\bar{y}) - c(x_1,\bar{y})\big] \\ & + \sum_{j=1}^{J-1} \big[c(x_j,y_j) - c(x_{j+1},y_j)\big] + c(x_J,y) \\ & \mid J \in \mathbb{N}, J \geq 2, (x_j,y_j) \in S, \forall j = 1,\ldots,J-1, (x_J,y) \in S\Big\} \qquad \forall y \in M.\end{aligned}$$

Note that if $y \in M$ is such that there are no $x \in M$ with $(x,y) \in S$, then $\phi(y) = -\infty$. However, as above we check easily that $\phi(\bar{y}) = 0$ which shows that ϕ is not identically $-\infty$. Therefore, by construction we have for every $x \in M$,

$$\psi(x) = \sup\Big\{\phi(y) - c(x,y) \,|\, y \in \pi^2(S)\Big\} = \sup\Big\{\phi(y) - c(x,y) \,|\, y \in M\Big\}, \tag{3.15}$$

which shows that ψ is c-convex. It remains to check that $S \subset \partial_c\psi$. Let $(x,y) \in S$ be fixed, we need to show that

$$\psi(x) + c(x,y) \leq \psi(z) + c(z,y) \qquad \forall z \in M.$$

By construction of ψ, we have for every $z \in M$,

$$\begin{aligned}\psi(z) \geq \sup\Big\{ & \big[c(\bar{x},\bar{y}) - c(x_1,\bar{y})\big] \\ & + \sum_{j=1}^{J-1} \big[c(x_j,y_j) - c(x_{j+1},y_j)\big] + [c(x,y) - c(z,y)] \\ & \mid J \in \mathbb{N}, J \geq 2, (x_j,y_j) \in S, \forall j = 1,\ldots,J-1, x_J = x\Big\} \\ & \qquad = \psi(x) + c(x,y) - c(z,y).\end{aligned}$$

We get the necessary and sufficient condition for a set to be c-cyclically monotone. Let us now turn to the second part of the result, that is let us prove that for any c-cyclically monotone set $S \subset M \times M$, there is a c-pair of potentials ψ, ψ^c with $S \subset \partial_c \psi$ which in addition satisfies (3.13)-(3.14).
Let S be a c-cyclically monotone set. We already know that there is $\phi : M \to \mathbb{R} \cup \{-\infty\}$ which is not identically $-\infty$ such that the function $\psi : M \to \mathbb{R} \cup \{+\infty\}$ defined by

$$\psi(x) := \sup\Big\{\phi(y) - c(x,y) \,|\, y \in \pi^2(S)\Big\} \qquad \forall x \in M, \tag{3.16}$$

is c-convex with $S \subset \partial_c \psi$ (remember (3.15)). Let $\phi_1 = \psi^c : M \to \mathbb{R} \cup \{-\infty\}$ be the c-transform of ψ, that is the function defined by

$$\phi_1(y) := \inf\Big\{\psi(x) + c(x,y) \,|\, x \in M\Big\} \qquad \forall y \in M. \tag{3.17}$$

If $y \in \pi^2(S)$, then there is $x \in M$ with $\psi(x) = \phi(y) - c(x,y)$ and $(x,y) \in S \subset \partial_c \psi$, that is

$$\phi(y) = \psi(x) + c(x,y) \leq \psi(z) + c(z,y) \qquad \forall z \in M.$$

Then we get $\phi(y) \leq \phi_1(y)$ for all $y \in \pi^2(S)$. On the other hand, by construction of ψ, we have $\psi(x) \geq \phi(y) - c(x,y)$ for any $x \in M$ and any $y \in \pi^2(S)$. Therefore

$$\phi_1(y) = \phi(y) \qquad \forall y \in \pi^2(S). \tag{3.18}$$

By Proposition 3.2.8, we have

$$\psi(x) = \sup\Big\{\phi_1(y) - c(x,y) \,|\, y \in M\Big\} \qquad \forall x \in M,$$

and by (3.16) and (3.18), we also have

$$\psi(x) = \sup\Big\{\phi_1(y) - c(x,y) \,|\, y \in \pi^2(S)\Big\} \qquad \forall x \in M.$$

We claim that ϕ_1 defined by (3.17) satisfies

$$\phi_1(y) = \inf\Big\{\psi(x) + c(x,y) \,|\, x \in \pi^1(S)\Big\} \qquad \forall y \in M.$$

If not, there are $\bar{x}, \bar{y} \in M$ and $\delta > 0$ such that

$$\psi(\bar{x}) + c(\bar{x}, \bar{y}) \leq \psi(z) + c(z, \bar{y}) - \delta \qquad \forall z \in \pi^1(S).$$

Taking the infimum in the right-hand side we get

$$\psi(\bar{x}) + c(\bar{x}, \bar{y}) \leq \phi_1(\bar{y}) - \delta.$$

But by construction of ϕ_1, we have $\phi_1(\bar{y}) \leq \psi(\bar{x}) + c(\bar{x}, \bar{y})$. We get a contradiction. It remains to show that both ψ, ψ^c are finite valued and continuous provided c is continuous and S is compact. We claim that under those assumptions, ψ^c is bounded from above on $\pi^2(S)$. Since ψ is not identically $+\infty$, there is $\bar{x} \in M$ with $\psi(\bar{x}) < +\infty$. Since c is continuous and $\pi^2(S)$ is compact, the function $y \mapsto c(\bar{x}, y)$ is bounded on $\pi^2(S)$. Then we deduce that ψ^c is bounded on $\pi^2(y)$. By (3.13), we infer that $\psi(x)$ is finite for any $x \in M$. Let $x \in M$ be fixed and $\{x_k\}_k$ be a sequence converging to x. For every $k > 0$, there is $y_k \in \pi^2(S)$ such that

$$\psi(x_k) \leq \psi^c(y_k) - c\big(x_k, y_k\big) + \frac{1}{k}.$$

Then we have for every $k > 0$,

$$\begin{aligned} \psi(x) \geq \psi^c(y_k) - c\big(x, y_k\big) &= \psi^c(y_k) - c\big(x_k, y_k\big) + c\big(x_k, y_k\big) - c\big(x, y_k\big) \\ &\geq \psi(x_k) - \frac{1}{k} + c\big(x_k, y_k\big) - c\big(x, y_k\big). \end{aligned} \tag{3.19}$$

For every $k > 0$, there is $z_k \in \pi^2(S)$ such that

$$\psi(x) \leq \psi^c(z_k) - c\big(x, z_k\big) + \frac{1}{k}.$$

Then we also have for every $k > 0$,

$$\begin{aligned} \psi(x_k) \geq \psi^c(z_k) - c\big(x_k, z_k\big) &= \psi^c(z_k) - c\big(x, z_k\big) + c\big(x, z_k\big) - c\big(x_k, z_k\big) \\ &\geq \psi(x) - \frac{1}{k} + c\big(x, z_k\big) - c\big(x_k, z_k\big). \end{aligned} \tag{3.20}$$

Let $\mathcal{V}$ be a compact neighborhood of x. The function c is continuous on the compact set $\mathcal{V} \times \pi^2(S)$, hence it is uniformly continuous. We conclude easily from (3.19)-(3.20) that $\psi(x_k)$ tends to $\psi(x)$ as k tends to $+\infty$. In the same way, we can show that ψ is bounded on $\pi^1(S)$ and ψ^c if always valued in $\mathbb{R}$ and continuous. The fact that the infimum and supremum in (3.13)-(3.14) are attained is straigthforward from the continuity of ψ, ψ^c and the compactness of S. □

Corollary 3.2.14. *Let μ, ν be two probability measures on M. Assume that c is continuous and that Supp(μ) and Supp(ν) are compact. Then there is a c-cyclically monotone compact set $\mathcal{S} \subset$ Supp$(\mu) \times$ Supp(ν) such that for every $\alpha \in \Pi(\mu, \nu)$ the following properties are equivalent:*

(i) α is optimal,

(ii) Supp$(\alpha) \subset \mathcal{S}$.

Proof. By Theorem 3.2.5, there is a c-cyclically monotone compact set $\mathcal{S} \subset \mathrm{Supp}(\mu) \times \mathrm{Supp}(\nu)$ such that the support of any optimal transport in $\Pi(\mu, \nu)$ is contained in $\mathcal{S}$. Let us show that $\mathcal{S}$ satisfies the equivalence given in the statement of the theorem. First, by construction we have (i) $\Rightarrow$ (ii). By Theorem 3.2.13, there is a c-pair of potentials with $\mathcal{S} \subset \partial_c \psi$. Then we have

$$\psi^c(y) - \psi(x) = c(x, y) \qquad \forall (x, y) \in \mathcal{S}. \tag{3.21}$$

Furthermore we have

$$\psi^c(y) - \psi(x) \leq c(x, y) \qquad c(x, y) \qquad \forall x, y \in M. \tag{3.22}$$

Let us show that (ii) $\Rightarrow$ (i). Let $\alpha \in \Pi(\mu,\nu)$ be such that $\mathrm{Supp}(\alpha) \subset \mathcal{S}$. On the one hand, by (3.21), we have

$$\begin{aligned}\int_M \psi^c(y)\,d\nu(y) - \int_M \psi(x)\,d\mu(x) &= \int_{M\times M} \left(\psi^c(y) - \psi(x)\right)\,d\alpha(x,y)\\ &= \int_{M\times M} c(x,y)\,d\alpha(x,y) = C(\alpha).\end{aligned}$$

On the other hand, (3.22) yields for every $\alpha' \in \Pi(\mu,\nu)$,

$$\begin{aligned}\int_M \psi^c(y)\,d\nu(y) - \int_M \psi(x)\,d\mu(x) &= \int_{M\times M} \left(\psi^c(y) - \psi(x)\right)\,d\alpha'(x,y)\\ &\leq \int_{M\times M} c(x,y)\,d\alpha(x,y) = C(\alpha').\end{aligned}$$

This shows that α is optimal. $\square$

Remark 3.2.15. *Let μ,ν be two compactly supported probability measures on M and $c : M\times M \to [0,+\infty)$ be a continuous cost. Actually, the proof of Corollary 3.2.14 shows that if (ψ,ψ^c) is a c-pair of potentials and α is a transport plan between μ and ν with Supp$(\alpha) \subset \partial_c\psi$, then α is optimal, that is $C_{\mathcal{K}}(\mu,\nu) = C(\alpha)$.*

3.3 A generalized Brenier-McCann Theorem

Throughout this section, we fix a cost $c : M\times M \to [0,+\infty)$ which is assumed to be continuous. Given two compactly supported probability measures μ,ν on M, we know by Theorems 3.2.5 and 3.2.13 that there is a c-cyclically monotone compact set $\mathcal{S} \subset \mathrm{Supp}(\mu)\times\mathrm{Supp}(\nu)$ which contains the supports of all optimal plans between μ and ν and a c-pair of real-valued continuous potentials (ψ,ψ^c) satisfying

$$\psi(x) = \max\Big\{\psi^c(y) - c(x,y)\,|\, y\in \pi^2(\mathcal{S})\Big\} \qquad \forall x\in M, \tag{3.23}$$

$$\psi^c(y) = \min\Big\{\psi(x) + c(x,y)\,|\, x\in \pi^1(\mathcal{S})\Big\} \qquad \forall y\in M, \tag{3.24}$$

and

$$\mathcal{S} \subset \partial_c\psi. \tag{3.25}$$

To prove the existence and uniqueness of an optimal transport map, we will show that $\mathcal{S}$ is concentrated on a graph. More precisely, we will prove that for every x outside a μ-negligible set $N\subset M$, the set $\partial_c\psi(x)$ is a singleton.

Theorem 3.3.1. *Let μ,ν be two probability measures on M. Assume that c is continuous and that Supp(μ) and Supp(ν) are compact. Let $\mathcal{S}$ and (ψ,ψ^c) given by Theorems 3.2.5 and 3.2.13 as above. Moreover assume that for μ-a.e. $x\in M$, the set $\partial_c\psi(x)$ is a singleton. Then there is a unique optimal transport map from μ to ν. It satisfies*

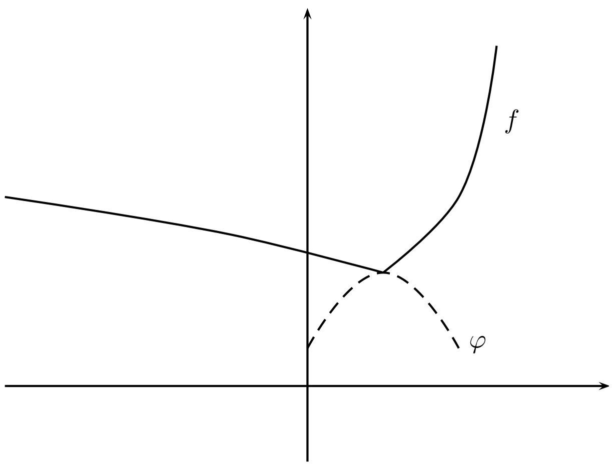
$$\partial_c\psi(x) = \big\{T(x)\big\} \qquad \mu - a.e.\ x\in M. \tag{3.26}$$

Proof. By Theorem 3.2.1, there is an optimal transport plan α between μ and ν. By assumption, there is a Borel set N such that $\mu(N) = 0$ and for every $x \notin N$, $\partial_c\psi(x)$ is a singleton $\{y_x\}$. Then for every $(x, y) \in \mathrm{Supp}(\alpha) \setminus (N \times M)$, we have $(x, y) \in \partial_c\psi$, that is $y = y_x$. Setting $T(x) := y_x$ for μ-a.e. $x \in M$, we get (3.26) and in turn the uniqueness. $\square$

Remark 3.3.2. *We maybe need to make clear what me mean by uniqueness of an optimal transport map. We say that there is a unique optimal transport map from μ to ν if there is uniqueness up to a set of μ-measure zero. That is if T_1 and T_2 are two optimal transport maps from μ to ν, there is a set N with $\mu(N) = 0$ such that $T_1(x) = T_2(x)$ for every $x \notin N$.*

We now introduce an assumption on the cost c. For this we need to define the notion of sub-differential. Given an open set $\Omega \subset M$ and a function $f : \Omega \to \mathbb{R}$, we say that $p \in T_x^*M$ is a *sub-differential* for f at $x \in \Omega$ if there is a function $\varphi : \Omega \to \mathbb{R}$ which is differentiable at x with $D_x\varphi = p$ such that

$$f(x) = \varphi(x) \quad \text{and} \quad f(y) \geq \varphi(y) \qquad \forall y \in \Omega.$$



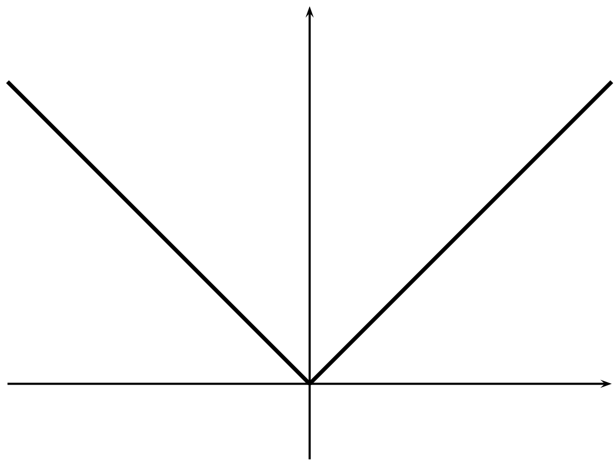
We denote by $D_x^- f$ the set of sub-differentials of f at x. In the same way, we say that $p \in T_x^*M$ is a *super-differential* for f at $x \in \Omega$ if there is a function $\varphi : \Omega \to \mathbb{R}$ which is differentiable at x with $D_x\varphi = p$ such that

$$f(x) = \varphi(x) \quad \text{and} \quad f(y) \leq \varphi(y) \qquad \forall y \in \Omega.$$

We denote by $D_x^+ f$ the set of super-differentials of f at x.

Remark 3.3.3. *If $f : \Omega \to \mathbb{R}$ is differentiable at $x \in \Omega$, then $D_x^- f = D_x^+ f = \{D_x f\}$.*

Remark 3.3.4. *The sub-differential and/or the super-differential may not be a singleton. It could be empty or contain several sub-differentials.*



For example, the sub-differential of the function $x \mapsto |x|$ at the origin is the interval $[-1, 1]$ while its super-differential is empty.

By (3.23), for every $(x, y) \in \partial_c \psi$ there is a link between the super-differentials of ψ at x and the sub-differentials of the cost c at (x, y). This lead us to the following definition which will be satisfied by variational costs.

Definition 3.3.5. *We say that the cost c satisfies the sub-TWIST condition if*

$$D^-_x c(\cdot, y_1) \cap D^-_x c(\cdot, y_2) = \emptyset \qquad \forall y_1 \neq y_2 \in M, \forall x \in M, \tag{3.27}$$

where $D^-_x(\cdot, y_i)$ denotes the sub-differential of the function $x \mapsto c(x, y_i)$ at x.

The following result makes the sub-TWIST condition relevant.

Lemma 3.3.6. *Assume that the cost c satisfies the sub-TWIST condition. Let (ψ, ψ^c) be a c-pair of potentials and $x \in M$ be such that ψ has a non-empty super-differential at x. Then $\partial_c \psi(x)$ is a singleton.*

Proof. Argue by contradiction and assume that $y_1 \neq y_2$ both belong to $\partial_c \psi(x)$. Then we have

$$\psi^c(y_i) = \psi(x) + c(x, y_i) \leq \psi(z) + c(z, y_i) \qquad \forall z \in M.$$

Thus, for every $i = 1, 2$,

$$c(z, y_i) \geq -\psi(z) + \psi(x) + c(x, y_i),$$

with equality at $z = x$. Since ψ is super-differentiable at x, we infer that both functions $z \mapsto c(z, y_1)$ and $z \mapsto c(z, y_2)$ share a common sub-differentiable at x. This contradicts the sub-TWIST condition. $\square$

By Theorem 3.3.1 and Lemma 3.3.6, in order to prove the existence and uniqueness of optimal transport maps from a compactly supported probability measure μ to another one ν, it is sufficient to show that the super-differential of the potential ψ is non-empty for μ-almost every point in M. Such a property can be obtained thanks to Rademacher's Theorem. We recall that a function defined on a smooth manifold is called *Lipschitz in charts* if it is Lipschitz in a set of local coordinates in a neighborhood of any point. The Rademacher Theorem asserts that any function which is Lipschitz in charts on an open subset Ω of M is differentiable almost everywhere in Ω.

Theorem 3.3.7. *Let $c : M \times M \to [0, +\infty)$ be a cost which is Lipschitz in charts and satisfies the sub-TWIST condition. Let μ, ν be two probability measures with compact support on M. Assume that μ is absolutely continuous with respect to the Lebesgue measure. Then there is existence and uniqueness of an optimal transport map from μ to ν. In fact, there is a c-convex function $\psi : M \to \mathbb{R}$ which is Lipschitz in charts such that*

$$\partial_c \psi(x) = \{T(x)\} \qquad \mu - a.e.\ x \in M. \tag{3.28}$$

Proof. By Theorems 3.2.5 and 3.2.13 there is a c-cyclically monotone compact set $\mathcal{S} \subset \mathrm{Supp}(\mu) \times \mathrm{Supp}(\nu)$ which contains the supports of all optimal plans between μ and ν together with a c-pair of real-valued continuous potentials (ψ, ψ^c) such that (3.23)-(3.25) are satisfied. In a neighborhood of each $x \in M$, the function ψ is the maximum of a family of functions $x \in \pi^2(S) \mapsto \psi^c(y) - c(x, y)$ with $y \in \pi^2(\mathcal{S})$ which are uniformly Lipschitz (in charts) in the x variable . Therefore, ψ is Lipschitz in charts on M. Since μ is assumed to be absolutely continuous with respect to the Lebesgue measure, Rademacher's Theorem implies that ψ is differentiable and a fortiori super-differentiable μ-a.e. We conclude easily by Theorem 3.3.1 and Lemma 3.3.6. □

Example 3.3.8. *Let $M = \mathbb{R}^n$ and $c : \mathbb{R}^n \times \mathbb{R}^n \to [0, +\infty)$ be the quadratic Euclidean cost or Brenier cost defined by $c(x, y) = |y - x|^2/2$ for any $x, y \in \mathbb{R}^n$. Remembering Example 3.2.10, we know that c-convex functions are the functions $\psi : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ such that the function $\psi + |\cdot|^2/2$ is convex. Furthermore, c satisfies the sub-TWIST condition. As a matter of fact, it is smooth and its partial derivative with respect to the x variable is given by*

$$\frac{\partial c}{\partial x}(x, y) = x - y \qquad \forall x, y \in \mathbb{R}^n.$$

Therefore $y_1 \neq y_2 \Rightarrow D_x c(\cdot, y_1) \neq D_x c(\cdot, y_2)$. By Theorem 3.3.7, given a pair of compactly supported probability measures μ, ν in $\mathbb{R}^n$ with μ absolutely continuous with respect to the Lebesgue measure, there is a unique optimal transport map $T : M \to M$ from μ to ν satisfying (3.28) where $\psi : \mathbb{R}^n \to \mathbb{R}$ is a locally Lipschitz c-convex function. Note that for every $x \in \mathbb{R}^n$ where ψ is differentiable at x, we have

$$y \in \partial_c \psi(x) \implies \psi(x) + c(x, y) \leq \psi(z) + c(z, y) \quad \forall z \in \mathbb{R}^n,$$

which means that the derivative of the function $z \mapsto \psi(z) + c(z, y)$ vanishes at $z = x$, that is $y = x + \nabla_x \psi$. Setting $\varphi(x) := \psi(x) + |x|^2/2$ for every $x \in M$, we obtain a convex function such that

$$T(x) = \nabla_x \varphi \qquad \mu - a.e\ x \in \mathbb{R}^n.$$

In other terms, the unique optimal transport map from μ to ν is given by the gradient of a convex function.

Example 3.3.9. *Let $M = \mathbb{R}^n$, note that the Monge cost $c : \mathbb{R}^n \times \mathbb{R}^n \to [0, +\infty)$ is Lipschitz but does not satisfy the sub-TWIST condition. As a matter of fact, we have*

$$\frac{\partial c}{\partial x}(x, y) = \frac{x - y}{|x - y|} \qquad \forall x \neq y \in \mathbb{R}^n.$$

This means that $D_x c(\cdot, y_1) = D_x c(\cdot, y_2)$ for any y_1, y_2 such that $y_1 - x$ and $y_2 - x$ are positively colinear. Hence Theorem 3.3.8 do not apply. In fact, we already saw through Example 3.1.3 that uniqueness of optimal transport maps does not hold in this context.

Example 3.3.10. *Let (M, g) be a complete Riemannian manifold. The geodesic distance d_g is Lipschitz in charts on $M \times M$. Define the quadratic geodesic cost or McCann's cost $c : M \times M \to [0, +\infty)$ by*

$$c(x, y) := \frac{1}{2} d_g^2(x, y) \qquad \forall x, y \in M.$$

Then c is Lipschitz in charts on $M \times M$ and satisfies the sub-TWIST condition. As a matter of fact, given $x \in M$ and $p \in T_x^*M$ in $D_x^- c(\cdot, y)$ for some $y \in M$, there is a function $\varphi : M \to \mathbb{R}$ which is differentiable at x with $D_x \varphi = p$ such that

$$\frac{1}{2} d_g^2(x, y) = \varphi(x) \quad \text{and} \quad \frac{1}{2} d_g^2(z, y) \geq \varphi(z) \qquad \forall z \in M.$$

Then we argue as in the proof of Lemma 2.3.6. If we denote by $\bar{\gamma} : [0, 1] \to M$ a minimizing geodesic from y to x, then we obtain that for every curve $\gamma : [0, 1] \to M$ with $\gamma(0) = y$,

$$\frac{1}{2} \mathit{energy}^g(\gamma) - \varphi\big(\gamma(1)\big) \geq 0,$$

with equality for $\gamma = \bar{\gamma}$. As in Lemma 2.3.6, we infer that there is a unique minimizing geodesic between x and y and that

$$y = \exp_x\big(-D_x \varphi\big) = \exp_x(-p),$$

where $\exp_x : T_x^*M \to M$ stands for the exponential map which was defined in Section 2.3 (if we use the Riemannian exponential map, we have $y = \exp_x\big(-\nabla_x^g \varphi\big)$). The point y is uniquely determined by p, then c satisfies the sub-TWIST condition. Arguing as above, we deduce that for every pair of compactly supported probability measures μ, ν on M with μ absolutely continuous with respect to the Lebesgue measure, there is a unique optimal transport map T from μ to ν satisfying (3.26) where $\psi : M \to \mathbb{R}$ is a c-convex function which is Lipschitz in charts. By the above discussion, we have

$$T(x) = \exp_x\big(D_x \psi\big) \qquad \mu - a.e\ x \in M \tag{3.29}$$

and for μ-a.e. $x \in M$ there is a unique minimizing geodesic from x to $T(x)$.

Let M be a smooth connected manifold equipped with a complete sub-Riemannian structure (Δ, g) and whose the sub-Riemannian distance is denoted by d_{SR}. In the next section, our purpose is now to study the Monge problem for the sub-Riemannian quadratic cost, that is for the cost $c : M \times M \to [0, +\infty)$ defined by

$$c(x, y) := \frac{1}{2} d_{SR}(x, y)^2 \qquad \forall x, y \in M.$$

As we saw before, in order to obtain existence and uniqueness results for optimal transport maps, it is convenient to be able to show that super-differentials of potentials are non-empty almost everywhere and that some sub-TWIST condition is satisfied by the cost function. The sub-TWIST condition follows immediately from Lemma 2.3.6. So we just have to deal with regularity issues of c-convex functions. In the case of compactly supported probability measures, regularity properties of Kantorovitch potentials can be obtained from the regularity of the cost. We develop this approach in the next section by showing that under additional assumptions the sub-Riemannian distance is Lipschitz and even locally semiconcave outside the diagonal.

Remark 3.3.11. *As explained above, if M equipped with a SR structure for which the cost $c = d_{SR}^2$ is Lipschitz on $M \times M$, then for every pair of compactly supported probability measures μ, ν on M with μ absolutely continuous with respect to the Lebesgue measure, there is a unique optimal transport map T from μ to ν which can be expressed as*

$$T(x) = \exp_x\big(D_x\psi\big) \qquad \mu - a.e \; x \in M, \tag{3.30}$$

where $\psi : M \to \mathbb{R}$ is a c-convex function which is Lipschitz in charts.

3.4 Optimal transport on ideal and Lipschitz SR structures

Ideal SR structures

Let (Δ, g) be a sub-Riemannian structure of rank $m \le n$ on M. We call it *ideal* if it is complete and has no non-trivial minimizing singular curves. We recall that this implies that for every $x \neq y \in M$, any minimizing geodesic $\gamma : [0,1] \to M$ joining x to y is regular. By the results of the previous chapter, all minimizing geodesics are smooth and projections of normal extremals of the Hamiltonian geodesic equation. We recall that D denotes the diagonal of $M \times M$, that is, the set of all pairs of the form (x,x) with $x \in M$. Sub-Riemannian distances of ideal SR structures are locally semiconcave outside the diagonal.

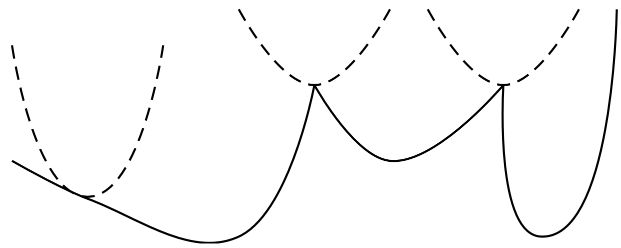
A function $f : \Omega \to \mathbb{R}$, defined on the open set $\Omega \subset M$, is called *locally semiconcave* on Ω if for every $x \in \Omega$ there exist a neighborhood Ω_x of x and a smooth diffeomorphism $\varphi_x : \Omega_x \to \varphi_x(\Omega_x) \subset \mathbb{R}^n$ such that $f \circ \varphi_x^{-1}$ is locally semiconcave on the open subset $\tilde{\Omega}_x = \varphi_x(\Omega_x) \subset \mathbb{R}^n$. By the way, we recall that the function $\tilde{f} : \tilde{\Omega} \to \mathbb{R}$, defined on the open set $\tilde{\Omega} \subset \mathbb{R}^n$, is locally semiconcave on $\tilde{\Omega}$ if for every $\bar{x} \in \tilde{\Omega}$ there exist $C, \delta > 0$ such that

$$\begin{aligned} &\mu f(y) + (1-\mu) f(x) - f\big(\mu x + (1-\mu) y\big) \\ &\qquad \le \mu(1-\mu) C |x-y|^2 \qquad \forall \mu \in [0,1],\ \forall x, y \in B\big(\bar{x}, \delta\big). \end{aligned} \tag{3.31}$$

This is equivalent to say that the function $\tilde{f}$ can be written locally as

$$\tilde{f}(x) = \Big(\tilde{f}(x) - C|x|^2\Big) + C|x|^2 \qquad \forall x \in B\big(\bar{x}, \delta\big),$$

with $\tilde{f}(x) - C|x|^2$ concave, that is as the sum of a concave function and a smooth function. Note that every locally semiconcave function is locally Lipschitz on its domain, and thus, by Rademacher's Theorem, it is differentiable almost everywhere on its domain.



The following result is useful to prove the local semiconcavity of a given function.

Lemma 3.4.1. *Let $f : \Omega \to \mathbb{R}$ be a function defined on an open set $\Omega \subset \mathbb{R}^n$. Assume that for every $\bar{x} \in \Omega$ there exist a neighborhood $\mathcal{V} \subset \Omega$ of $\bar{x}$ and a positive real number σ such that, for every $x \in \mathcal{V}$, there is $p_x \in \mathbb{R}^n$ such that*

$$u(y) \leq u(x) + \langle p_x, y - x \rangle + \sigma |y - x|^2 \qquad \forall y \in \mathcal{V}. \tag{3.32}$$

Then the function u is locally semiconcave on Ω.

Proof of Lemma 3.4.1. Let $\bar{x} \in \Omega$ be fixed and $\mathcal{V}$ be the neighborhood given by assumption. Without loss of generality, we can assume that $\mathcal{V}$ is an open ball $\mathcal{B}$. Let $x, y \in \mathcal{B}$ and $\mu \in [0, 1]$. The point $\hat{x} := \mu x + (1 - \mu) y$ belongs to $\mathcal{B}$. By assumption, there exists $\hat{p} \in \mathbb{R}^n$ such that

$$u(z) \leq u(\hat{x}) + \langle \hat{p}, z - \hat{x} \rangle + \sigma |z - \hat{x}|^2 \qquad \forall z \in \mathcal{B}.$$

Hence we easily get

$$\begin{aligned} \mu u(y) + (1 - \mu) u(x) &\leq u(\hat{x}) + \mu\sigma |x - \hat{x}|^2 + (1 - \mu)\sigma |y - \hat{x}|^2 \\ &\leq u(\hat{x}) + \left(\mu(1 - \mu)^2 \sigma + (1 - \mu)\mu^2 \sigma\right) |x - y|^2 \\ &\leq u(\hat{x}) + 2\mu(1 - \mu)\sigma |x - y|^2, \end{aligned}$$

and the conclusion follows. □

Remark 3.4.2. *Thanks to Lemma 3.4.1, a way to prove that a given function $f : \Omega \to \mathbb{R}$ is locally semiconcave on Ω is to show that for every $x \in \Omega$ we can put a C^2 support function φ on the graph of u at x with a uniform control of the C^2 norm of φ.*

Outside the diagonal, sub-Riemannian distances of ideal SR structures enjoy the same kind of regularity as Riemannian distances.

Theorem 3.4.3. *Let (Δ, g) be an ideal sub-Riemannian structure on M. Then the SR distance is continuous on $M \times M$ and locally semiconcave on $M \times M \setminus D$. In particular, d_{SR} is Lipschitz in charts on $M \times M \setminus D$.*

Proof. The continuity of d_{SR} follows from Proposition 1.5.2. To prove the local semiconcavity, we proceed as explained in Remark 3.4.2. Let us fix $(x,y) \in M \times M \setminus D$ and $\gamma \in \Omega_{\Delta}^{x,1}$ be a minimizing geodesic joining x to y. There is an open neighborhood $\mathcal{V}$ of $\gamma([0,1])$ in M and an orthonormal family $\mathcal{F}$ (with respect to the metric g) of m smooth vector fields $X^1, \ldots, X^m$ such that

$$\Delta(z) = \mathrm{Span}\Big\{X^1(z), \ldots, X^m(z)\Big\} \qquad \forall z \in \mathcal{V}.$$

Taking a change of coordinates if necessary, we may assume that $\mathcal{V}$ is an open subset of $\mathbb{R}^n$. Furthermore, there is a control $u^\gamma \in L^2([0,1];\mathbb{R}^m)$ such that

$$\dot{\gamma}(t) = \sum_{i=1}^m u_i^\gamma(t) X^i(\gamma(t))dt \qquad \text{a.e. } t \in [0,1].$$

Since u^γ is regular, there are $v^1, \ldots v^n$ in $L^2([0,1];\mathbb{R}^m)$ such that the linear operator

$$\begin{array}{ccc} \mathbb{R}^n & \longrightarrow & \mathbb{R}^n \\ \alpha & \longmapsto & \sum_{i=1}^m \alpha_i D_{u^\gamma} E_{\mathcal{F}}^{x,1}\left(v^i\right) \end{array}$$

is invertible. Define locally $\mathcal{F} : \mathbb{R}^n \to \mathbb{R}^n$ by

$$\begin{array}{rcl} \mathcal{F} : \mathbb{R}^n \times \mathbb{R}^n & \longrightarrow & \mathbb{R}^n \times \mathbb{R}^n \\ (z,\alpha) & \longmapsto & \left(z, E_{\mathcal{F}}^{z,1}\left(u^\gamma + \sum_{i=1}^m \alpha_i v^i\right)\right) \end{array}$$

This mapping is well-defined and C^2 in a neighborhood of $(x,0)$. Moreover it satisfies

$$\mathcal{F}(x,0) = (x,y),$$

and its differential at $(x,0)$ is invertible. Hence by the Inverse Function Theorem, there are an open ball $\mathcal{B}$ centered at (x,y) in $\mathbb{R}^n \times \mathbb{R}^n$ and a function $\mathcal{G} : \mathcal{B} \to \mathbb{R}^n \times \mathbb{R}^n$ of class C^2 such that

$$\mathcal{F} \circ \mathcal{G}(z,w) = (z,w) \qquad \forall (z,w) \in \mathcal{B}.$$

Denote by α^{-1} the second component of $\mathcal{G}$. From the definition of the sub-Riemannian energy between two points, we infer that for any $(z,w) \in \mathcal{B}$ we have

$$e_{SR}(z,w) \leq \left\| u^\gamma + \sum_{i=1}^m \left(\alpha^{-1}(z,w)\right)_i v_i \right\|_{L^2}^2.$$

Set

$$\phi^{x,y}(z,w) := \left\| u^\gamma + \sum_{i=1}^m \left(\alpha^{-1}(z,w)\right)_i \right\|_{L^2} \qquad \forall (z,w) \in \mathcal{B}.$$

We conclude that, there is a function $\phi^{x,y}$ of class C^2 such that $d_{SR}(z,w) \leq \phi^{x,y}(z,w)$ for any (z,w) in a neighborhood of (x,y) in $M \times M$, and $d_{SR}(x,y) = \phi^{x,y}(x,y)$. By compactness, the C^2 norms of the functions $\phi^{x,y}$ are uniformly bounded. As a matter of fact, from Remark 2.1.8 we know that the set of minimizing geodesics from x to y is compact with respect to the uniform topology; any sequence of minimizing geodesics $\{\gamma_k\}_k$ from x_k to y_k converges uniformly to a minimizing geodesic from x to y. We also know (see Remark 2.1.4) that if we cover the set of minimizing curves from x to y by a finite number of open tubes admitting orthonormal frames, then minimizing control converge in L^2. We conclude easily. $\square$

Remark 3.4.4. *The above arguments can be used to prove the following result. Let (Δ, g) be a sub-Riemannian structure of rank $m < n$ on M. Assume that it is complete and that there is an open set $\Omega \subset M \times M$ such that for every $(x, y) \in \Omega$ with $x \neq y$, any minimizing geodesic between x and y is regular. Then d_{SR} is locally semiconcave on $\Omega \setminus D$.*

Remark 3.4.5. *Any SR structure of rank $m = n$, that is any Riemannian structure on M is ideal, see Remarks 1.3.7, 2.2.3.*

Lipschitz SR structures

Let (Δ, g) be a sub-Riemannian structure of rank $m < n$ on M. We call it *Lipschitz* if it is complete and if the sub-Riemannian distance function is Lipschitz in charts on $M \times M$ outside the diagonal (or equivalently if the sub-Riemannian energy is Lipschitz in charts on $M \times M \setminus D$). A particular case of Lipschitz SR structures is given by ideal SR structures. The aim of the present section is to provide a weaker sufficient condition for a complete SR structure to be Lipschitz. According to Theorem 2.44, a horizontal path $\gamma : [0, 1] \to M$ will be called a *Goh path* if it admits an abnormal lift $\psi : [0, 1] \to \Delta^\perp$ which annihilates $[\Delta, \Delta]$, that is, an abnormal lift $\psi = (\gamma, p) : [0, 1] \to T^*M$ (in local coordinates, see Proposition 1.3.3 and the subsequent remarks) such that for every every local parametrization of Δ by smooth vector fields $X^1, \ldots, X^m$ in a neighborhood of $\gamma([0, 1])$, we have

$$p(t) \cdot \big[X^i, X^j\big]\big(\gamma(t)\big) = 0 \qquad \forall t \in [0, 1], \, \forall i, j = 1, \ldots, m.$$

Of course, the above definition does not depend upon the parametrization.

Theorem 3.4.6. *Let (Δ, g) be a complete sub-Riemannian structure on M, assume that any sub-Riemannian minimizing geodesic joining two distinct points in M is not a Goh path. Then, the SR structure (Δ, g) is Lipschitz.*

Proof. Let us fix $(x, y) \in M \times M \setminus D$ and $\gamma \in \Omega_\Delta^{x,1}$ a minimizing geodesic joining x to y. As before, denote by $\mathcal{F} = \{X^1, \ldots, X^m\}$ an orthonormal family of vector fields along $\gamma([0, 1])$ and by u^γ the control associated with γ. Two cases may appear:

First case: $\bar{u} := u^\gamma$ is not singular.
Then by the arguments given in the proofs of Lemma 2.3.9 and Theorem 3.4.3, there are $\delta, K > 0$ such that

$$e_{SR}(x, z) \leq e_{SR}(x, y) + K|z - y| \qquad \forall z \in B(y, \delta).$$

Since any control which is close enough to $\bar{u}$ is regular, there is $\bar{\epsilon} > 0$ such that for every $u \in L^2\big([0, 1]; \mathbb{R}^m\big)$ satisfying

$$\big\|u - \bar{u}\big\|_{L^2} < \bar{\epsilon}, \quad e_{SR}\left(x, E_{\mathcal{F}}^{x,1}(u)\right) = \|u\|_{L^2},$$

there holds

$$e_{SR}(x, z) \leq e_{SR}\left(x, E_{\mathcal{F}}^{x,1}(u)\right) + 2K \left|z - E_{\mathcal{F}}^{x,1}(u)\right|, \tag{3.33}$$

for every $z \in B\left(E_{\mathcal{F}}^{x,1}(u), \delta/2\right)$.

Second case: $\bar{u} = u^{\gamma}$ is singular.
By Theorem 2.4.2, we have necessarily

$$\mathrm{ind}_{-}\left(\lambda^{*}\left(D_{\bar{u}}^{2}E_{\mathcal{F}}^{x,1}\right)_{|\mathrm{Ker}\left(D_{\bar{u}}E_{\mathcal{F}}^{x,1}\right)}\right) = +\infty, \tag{3.34}$$

for all $\lambda \in \mathrm{Im}\left(D_{\bar{u}}E_{\mathcal{F}}^{x,1}\right)^{\perp} \setminus \{0\}$. Recall that $C : L^2([0,1];\mathbb{R}^m)$ is defined by

$$C(u) := \|u\|_{L^2}^2 \qquad \forall u \in L^2\big([0,1];\mathbb{R}^m\big).$$

Let $E_0 \subset L^2([0,1];\mathbb{R}^m)$ be a vector space such that

$$E_0 + \mathrm{Ker}\left(D_{\bar{u}}E_{\mathcal{F}}^{x,1}\right) = L^2\big([0,1];\mathbb{R}^m\big).$$

Set

$$E := E_0 \oplus \left(\mathrm{Ker}\left(D_{\bar{u}}E_{\mathcal{F}}^{x,1}\right) \cap \mathrm{Ker}\left(D_{\bar{u}}C\right)\right) \quad \text{and} \quad F := \left(E_{\mathcal{F}}^{x,1}\right)_{|\{\bar{u}\}+E}.$$

By construction, $D_{\bar{u}}E_{\mathcal{F}}^{x,1}$ and $D_{\bar{u}}F$ have the same image in $\mathbb{R}^n$ and E_0 has finite dimension. Then by (3.34), we have

$$\mathrm{ind}_{-}\left(\lambda^{*}\left(D_{\bar{u}}^{2}F\right)_{|\mathrm{Ker}(D_{\bar{u}}F)}\right) = +\infty,$$

for all $\lambda \in \mathrm{Im}\left(D_{\bar{u}}F\right)^{\perp} \setminus \{0\}$. We can apply Theorem B.2.2 to the function F. Hence there are $c > 0, \bar{\epsilon} \in (0,1)$ such that for every $\epsilon \in (0,\bar{\epsilon})$ and every $z \in B\left(F(\bar{u}), c\epsilon^2\right)$, there are $w_1, w_2 \in L^2([0,1];\mathbb{R}^m)$ such that

$$z = F\big(\bar{u} + w_1 + w_2\big) \tag{3.35}$$

and

$$w_1 \in \mathrm{Ker}\big(D_{\bar{u}}F\big), \quad \big\|w_1\big\|_{L^2} < \epsilon, \quad \big\|w_2\big\|_{L^2} < \epsilon^2. \tag{3.36}$$

Let $z \in B(y, c\epsilon^2)$ with $|z-y| = c\epsilon^2/2$. Then there are $w_1, w_2 \in L^2\big([0,1];\mathbb{R}^m\big)$ such that (3.35)-(3.36) are satisfied. Set $u := \bar{u} + w_1 + w_2$. Then we have

$$z = E_{\mathcal{F}}^{x,1}(u),$$

and (note that $\mathrm{Ker}(D_{\bar{u}}F) \subset \mathrm{Ker}(D_{\bar{u}}C)$),

$$\begin{aligned}
e_{SR}(x,z) \le C(u) \quad &\le \quad C\big(\bar{u}\big) + D_{\bar{u}}C \cdot \big(w_1 + w_2\big) + \big\|w_1 + w_2\big\|_{L^2}^2 \\
&= \quad e_{SR}(x,y) + D_{\bar{u}}C \cdot w_2 + \big\|w_1 + w_2\big\|_{L^2}^2 \\
&\le \quad e_{SR}(x,y) + 2\big\|\bar{u}\big\|_{L^2}\,\epsilon^2 + \big(\epsilon + \epsilon^2\big)^2 \\
&\le \quad e_{SR}(x,y) + \left(\frac{4\big\|\bar{u}\big\|_{L^2} + 8}{c}\right)|z-y|.
\end{aligned}$$

Proceeding as in the proof of Theorem B.2.2, we can show that the above estimate holds in a neighborhood of $\bar{u}$, that is (taking $c > 0, \bar{\epsilon} \in (0,1)$ smaller if necessary) for every $\epsilon \in (0,\bar{\epsilon})$, for every $u \in L^2\big([0,1];\mathbb{R}^m\big)$, and every $z \in \mathbb{R}^n$ with

$$\big\|u-\bar{u}\big\|_{L^2} < \epsilon, \quad \Big|z - E_{\mathcal{F}}^{x,1}(u)\Big| < c\,\epsilon^2,$$

there are $w_1, w_2 \in L^2([0,1];\mathbb{R}^m)$ such that

$$z = E_{\mathcal{F}}^{x,1}\big(u + w_1 + w_2\big)$$

and

$$w_1 \in \mathrm{Ker}\big(D_u C\big), \quad \big\|w_1\big\|_{L^2} < \epsilon, \quad \big\|w_2\big\|_{L^2} < \epsilon^2.$$

This shows that for every $u \in L^2\big([0,1];\mathbb{R}^m\big)$ satisfying

$$\big\|u-\bar{u}\big\|_{L^2} < \bar{\epsilon}, \quad e_{SR}\left(x, E_{\mathcal{F}}^{x,1}(u)\right) = \|u\|_{L^2},$$

there holds

$$e_{SR}(x,z) \le e_{SR}\left(x, E_{\mathcal{F}}^{x,1}(u)\right) + \left(\frac{4\|u\|_{L^2}+8}{c}\right)\left|z - E_{\mathcal{F}}^{x,1}(u)\right|, \tag{3.37}$$

for every $z \in B\left(E_{\mathcal{F}}^{x,1}(u), c\bar{\epsilon}/4\right)$.

Let us explain how to conclude by compactness. Let $x \in M$ and $\mathcal{B}$ a compact set in M such that $\{x\}\times\mathcal{B}\cap D = \emptyset$ be fixed. Denote by $\mathcal{S}$ the set of all $y \in \mathcal{B}$ such that there is at least one singular minimizing geodesic between x and y. The set $\mathcal{S}$ is a compact subset of $\mathcal{B}$, and the set of singular minimizing geodesic between x and a point in $\mathcal{S}$ is compact with respect to the uniform topology. Then by the previous observation (second case) together with a compactness argument (see Remarks 2.1.4, 2.1.8), we infer that an inequality of the form (3.37) holds for any minimizing control u which is close enough to a control corresponding to a singular minimizing geodesic joining x to a point in $\mathcal{S}$. Denote by $\mathcal{S}'$ the set of y in $\mathcal{B}$ corresponding to such controls. By construction, any minimizing geodesic from x to a point in $\mathcal{B}\setminus\mathcal{S}'$ is regular. Actually, it is far from being singular. Then by the arguments given in the first case together with compactness arguments, an inequality of the form (3.33) holds for any y $(= E_{\mathcal{F}}^{x,1}(u))$ in $\overline{\mathcal{B}\setminus\mathcal{S}'}$. In that way, we prove that $e_{SR}(x,\cdot)$ (or equivalently $d_{SR}(x,\cdot)$) is locally Lipschitz in $M\setminus\{x\}$. The same proof shows that e_{SR} is indeed uniformly locally Lipschitz with respect to one variable. We conclude easily. □

Remark 3.4.7. *The above arguments can be used to prove the following result. Let (Δ, g) be a sub-Riemannian structure of rank $m<n$ on M. Assume that it is complete and that there is an open set $\Omega \subset M\times M$ such that for every $(x,y)\in\Omega$ with $x\neq y$, no minimizing geodesic between x and y is a Goh path. Then d_{SR} is Lipschitz in charts on $\Omega\setminus D$.*

Remark 3.4.8. *Note that if the path γ is constant on $[0,1]$, it is a Goh path if and only if there is a differential form $p \in T^*_{\gamma(0)}M$ satisfying*

$$p\cdot X^i(\gamma(0)) = p\cdot\big[X^i, X^j\big](\gamma(0)) = 0 \qquad \forall i,j = 1,\ldots,m,$$

where $X^1,\ldots,X^m$ is as above a parametrization of Δ in a neighborhood of $\gamma(0)$. The above proof shows that if Δ is 2-generating then e_{SR} is Lipschitz in charts on $M\times M$.

Remark 3.4.9. *If a SR structure (Δ, g) on M is Lipschitz, then for every $x \in M$, the exponential mapping $\exp_x$ is onto. In fact, for every y there is a minimizing geodesic joining x to y which is normal. This can be shown by the arguments which were given at the end of the proof of Theorem 2.3.5.*

A Brenier-McCann Theorem on Lipschitz SR structures

Before stating our existence and uniqueness result for Lipschitz SR structures, we introduce a definition.

Definition 3.4.10. *Given a c-convex function $\psi : M \to \mathbb{R}$, we call "moving" set $\mathcal{M}^\psi$ and "static" set $\mathcal{S}^\psi$ respectively the sets defined as follows:*

$$\mathcal{M}^\psi := \Big\{ x \in M \,|\, x \notin \partial_c \psi(x) \Big\},$$

$$\mathcal{S}^\psi := M \setminus \mathcal{M}^\phi = \Big\{ x \in M \,|\, x \in \partial_c \psi(x) \Big\}.$$

As shown by the following result, under classical assumptions on the measures and Lipschitzness of the sub-Riemannian structure, static points do not move while moving points obey a transportation law of the form (3.29)-(3.30).

Theorem 3.4.11. *Let (Δ, g) be a Lipschitz sub-Riemannian structure on M and μ, ν be two compactly supported probability measures on M. Assume that μ is absolutely continuous with respect to the Lebesgue measure. Then there is existence and uniqueness of an optimal transport map from μ to ν for the SR quadratic cost $c : M \times M \to [0, +\infty)$ defined by*

$$c(x,y) := \frac{1}{2} d_{SR}^2(x,y) \qquad \forall x, y \in M.$$

In fact, there is a continuous c-convex function $\psi : M \to \mathbb{R}$ such that the following holds:

(i) $\mathcal{M}^\psi$ is open, and ψ is Lipschitz in charts on $\mathcal{M}^\psi$. In particular ψ is differentiable μ-a.e. in $\mathcal{M}^\psi$.

(ii) For μ-a.e, $x \in \mathcal{S}^\psi$, $\partial_c \psi(x) = \{x\}$.

In particular, there exists a unique optimal transport map defined μ-a.e. by

$$T(x) := \begin{cases} \exp_x (D_x \psi) & \text{if } x \in \mathcal{M}^\psi, \\ x & \text{if } x \in \mathcal{S}^\psi, \end{cases}$$

and for μ-a.e. $x \in M$ there exists a unique minimizing geodesic between x and $T(x)$.

Proof. Let $\mathcal{S} \subset \mathrm{Supp}(\mu) \times \mathrm{Supp}(\nu)$ and (ψ, ψ^c) be respectively the c-cyclically monotone set and the c-pair of potentials satisfying (3.23)-(3.25). Since the sets $\mathrm{Supp}(\mu), \mathrm{Supp}(\nu)$ are assumed to be compact, both ψ, ψ^c are indeed continuous and the supremum and infimum in (3.23)-(3.24) are attained. We check easily that $x \in M$ belongs to $\mathcal{S}^\psi$ if and only if $\psi(x) = \psi^c(x)$. Then $\mathcal{M}^\psi$ coincides with the set

$$\Big\{ x \in M \,|\, \psi(x) \neq \psi^c(x) \Big\} = \Big\{ x \in M \,|\, \psi(x) > \psi^c(x) \Big\},$$

which is open by continuity of ψ and ψ^c. Let us now prove that ψ is Lipschitz in charts in an open neighborhood of $\mathcal{M}^\psi \cap \mathrm{Supp}(\mu)$. Let $x \in \mathcal{M}^\psi$ be fixed. Since $x \notin \partial_c\psi(x)$ and $\psi_c(x)$ is closed in M (by continuity of ψ, ψ_c and compactness of $\mathcal{S}$), there is $r > 0$ such that $d_{SR}(x,y) > 2r$ for any $y \in \partial_c\psi(x)$. In addition, since the set $\partial_c\psi$ is closed in $M \times M$ (again by continuity of ψ, ψ_c and compactness of $\mathcal{S}$), there exists a neighborhood $\mathcal{V}_x$ of x which is included in $\mathcal{M}^\psi$ such that

$$d_{SR}(z,w) \geq r \qquad \forall z \in \mathcal{V}_x, \quad \forall w \in \partial_c\psi(z).$$

Let $\psi_{x,r} : M \to \mathbb{R}$ be the function defined by

$$\psi_{x,r}(z) := \sup\Big\{\psi^c(y) - \frac{1}{2}d_{SR}^2(z,y) \,|\, y \in \pi^2(\mathcal{S}),\, d_{SR}(z,y) \geq r\Big\}.$$

By construction, ψ coincides with $\psi_{x,r}$ on $\mathcal{V}_x$. By assumption, d_{SR} is Lipschitz in charts outside the diagonal, then by compactness of $\mathcal{S}$ we deduce that $\psi_{x,r}$ is Lipschitz in charts. In conclusion Ψ is Lipschitz in charts on $\mathcal{M}^\psi$ and (i) is proved.

To prove (ii), we observe that it suffices to prove the result for x belonging to an open set $\mathcal{V} \subset M$ on which the horizontal distribution $\Delta(x)$ is parametrized by a orthonormal family a smooth vector fields $\mathcal{F} = \{X^1, \ldots, X^m\}$. In fact, up to working in charts, we can assume that $\mathcal{V}$ is a convex subset of $\mathbb{R}^n$ where the C^2-norms of the X^i's are bounded. Let us fix a compact ball $\mathcal{B}$ in $\mathcal{V}$ and show that (ii) holds for μ-a.e. $x \in \mathcal{B}$.

Recall that the Hamiltonian $H : \mathcal{V} \times (\mathbb{R}^n)^* \to \mathbb{R}$ which is associated to our sub-Riemannian structure is defined by (see Chapter 2)

$$H(x,p) := \frac{1}{2}\sum_{i=1}^m \big(p \cdot X^i(x)\big)^2 \qquad \forall (x,p) \in \mathcal{V} \times \big(\mathbb{R}^n\big)^*.$$

For every $p \in (\mathbb{R}^n)^* \setminus \{0\}$, denote by Π_p the linear hyperplane in $\mathbb{R}^n$ which is orthogonal to p, that is

$$\Pi_p := \Big\{v \in \mathbb{R}^n \,|\, p \cdot v = 0\Big\}.$$

From Lemma 2.2.5 and its proof, for every $\bar{x} \in \mathcal{V}$ and every $\bar{p} \in (\mathbb{R}^n)^*$ with $H(\bar{x},\bar{p}) \neq 0$, there is $\rho > 0$ such that the Dirichlet problem

$$\begin{cases} H(x, D_xS(x)) = H\big(\bar{x},\bar{p}\big), \\ S_{|\bar{x}+\Pi_{\bar{p}}} = 0, \end{cases} \tag{3.38}$$

admits a solution of class C^1 on the ball $B(\bar{x},\rho)$. We leave the reader to check that the radius ρ depends "continuously" on $\bar{x}$, $H(\bar{x},p)$ and $|p|$ ($|p|$ denotes the Euclidean norm of p). Then, by compactness of $\mathcal{B}$ there is a function

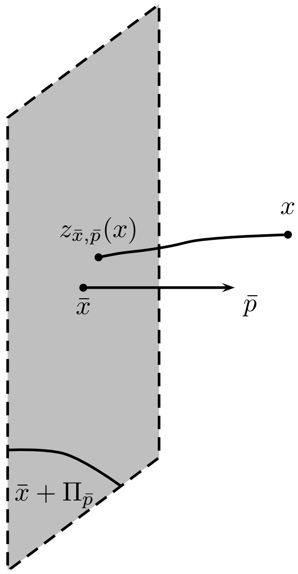
$$\rho \,:\, (0,+\infty) \times (0,+\infty) \longrightarrow (0,\infty)$$

which is decreasing in the first variable and increasing in the second variable such that for every $\bar{x} \in \mathcal{B}$ and every $\bar{p} \in (\mathbb{R}^n)^*$ with $H(\bar{x},\bar{p}) \neq 0$, the solution to

(3.38) is defined on the open ball $B\left(\bar{x}, \rho(H(\bar{x},\bar{p}), |\bar{p}|)\right)$. For any $\bar{x}, \bar{p}$ satisfying the previous assumptions, we denote by

$$S_{\bar{x},\bar{p}} \,:\, B\left(\bar{x}, \rho\big(H(\bar{x},\bar{p}), |\bar{p}|\big)\right) \longrightarrow \mathbb{R}$$

the solution to the Dirichlet problem (3.38), with $\rho_{\bar{x},\bar{p}} := \rho(H(\bar{x},\bar{p}), |\bar{p}|)$. The functions $S_{\bar{x},\bar{p}}$ being constructed by the method of characteristics (see Proof of Lemma 2.2.5), the following result holds (note that the parametrization of characteristics that we use in the statement of Lemma 3.4.12 differs from the one which is used to construct $S_{\bar{x},\bar{p}}$, see last statement).



Lemma 3.4.12. *There is a function*

$$\tau \,:\, (0,+\infty) \times (0,+\infty) \longrightarrow (0,+\infty)$$

which is increasing in the first variable and decreasing in the second variable such that the following property holds:
For every $\bar{x} \in \mathcal{B}$, for every $\bar{p} \in (\mathbb{R}^n)^*$ with $H(\bar{x},\bar{p}) \neq 0$, and every $x \in B(\bar{x}, \rho_{\bar{x},\bar{p}}/2)$ there are

$$z_{\bar{x},\bar{p}}(x) \in (\bar{x} + \Pi_{\bar{p}}) \cap B(\bar{x}, \rho_{\bar{x},\bar{p}})$$

$$\textit{and} \quad t_{\bar{x},\bar{p}}(x) \in \Big(-\tau\big(H(\bar{x},\bar{p}), |\bar{p}|\big), \tau\big(H(\bar{x},\bar{p}), |\bar{p}|\big)\Big)$$

such that

$$x = \gamma_{\bar{x},\bar{p}}\big(t_{\bar{x},\bar{p}}(x); z_{\bar{x},\bar{p}}(x)\big)$$

where (we set $\tau_{\bar{x},\bar{p}} := \tau\big(H(\bar{x},\bar{p}), |\bar{p}|\big)$)

$$\Big(\gamma_{\bar{x},\bar{p}}\big(\cdot; z_{\bar{x},\bar{p}}(x)\big), p_{\bar{x},\bar{p}}\big(\cdot; z_{\bar{x},\bar{p}}(x)\big)\Big) \,:\, (-\tau_{\bar{x},\bar{p}}, \tau_{\bar{x},\bar{p}}) \longrightarrow \mathcal{V} \times \left(\mathbb{R}^n\right)^*$$

is the solution to the Hamiltonian system

$$\left\{ \begin{array}{rcl} \dot{\gamma}_{\bar{x},\bar{p}}\big(t; z_{\bar{x},\bar{p}}(x)\big) & = & \frac{\partial H}{\partial p}\Big(\gamma_{\bar{x},\bar{p}}\big(t; z_{\bar{x},\bar{p}}(x)\big), p_{\bar{x},\bar{p}}\big(t; z_{\bar{x},\bar{p}}(x)\big)\Big) \\ \dot{p}_{\bar{x},\bar{p}}\big(t; z_{\bar{x},\bar{p}}(x)\big) & = & -\frac{\partial H}{\partial x}\Big(\gamma_{\bar{x},\bar{p}}\big(t; z_{\bar{x},\bar{p}}(x)\big), p_{\bar{x},\bar{p}}\big(t; z_{\bar{x},\bar{p}}(x)\big)\Big), \end{array} \right.$$

with

$$\gamma_{\bar{x},\bar{p}}\big(0;z_{\bar{x},\bar{p}}(x)\big)=z_{\bar{x},\bar{p}}(x) \quad \text{and} \quad p_{\bar{x},\bar{p}}\big(0;z_{\bar{x},\bar{p}}(x)\big)=\bar{p}.$$

In particular, $\gamma_{\bar{x},\bar{p}}$ is an horizontal path joining $z_{\bar{x},\bar{p}}(x)$ to x which satisfies

$$H\left(\gamma_{\bar{x},\bar{p}}\big(t;z_{\bar{x},\bar{p}}(x)\big),p_{\bar{x},\bar{p}}\big(t;z_{\bar{x},\bar{p}}(x)\big)\right)=H\left(z_{\bar{x},\bar{p}}(x),\bar{p}\right) \qquad \forall t\in\left(-\tau_{\bar{x},\bar{p}},\tau_{\bar{x},\bar{p}}\right).$$

For every $x\in\mathcal{V}$, we denote by $\Delta^{\perp}(x)$ the set of $p\in(\mathbb{R}^n)^*$ such that $H(x,p)\neq 0$. Pick a sequence $\{(x_k,p_k)\}_k$ of $\mathcal{B}\times(\mathbb{R}^n)^*$ which is a dense subset of

$$\Big\{(x,p)\in\mathcal{B}\times(\mathbb{R}^n)^* \mid p\in\Delta^{\perp}(x)\Big\}.$$

and set for every k,

$$\rho_k:=\rho_{x_k,p_k}, \quad \tau_k:=\tau_{x_k,p_k}, \quad t_k(\cdot):=t_{x_k,p_k}(\cdot),$$

$$z_k(\cdot):=z_{x_k,p_k}(\cdot), \quad \gamma_k(\cdot,\cdot):=\gamma_{x_k,p_k}(\cdot,\cdot), \quad p_k(\cdot,\cdot):=p_{x_k,p_k}(\cdot,\cdot).$$

The following result is a consequence of the Lipschitz regularity of the sub-Riemannian distance along horizontal paths together with Rademacher's theorem.

Lemma 3.4.13. *There is a set N of Lebesgue measure zero in $\mathcal{V}$ such that for every $x\in\mathcal{B}\setminus N$ and any k, the following property holds:*

$$x\in B\big(x_k,\rho_k/2\big) \text{ and } x=\gamma_k\big(t;z_k(x)\big) \\ \Longrightarrow s\mapsto\psi\left(\gamma_k\big(s;z_k(x)\big)\right) \text{ is differentiable at } t.$$

Proof of Lemma 3.4.13. Let k be fixed. By construction, all the curves $\gamma_k(\cdot;z)$ (with $z\in(\bar{x}+\Pi_{p_k})\cap B(x_k,\rho_k)$) are horizontal with respect to the distribution (we may assume without loss of generality that the curves $\gamma_k(\cdot;z)$ are defined on $(-\tau_k,\tau_k)$ for all $z\in(\bar{x}+\Pi_{p_k})\cap B(x_k,\rho_k)$). The potential ψ is expressed as

$$\psi(x)=\max\Big\{\psi^c(y)-\frac{1}{2}d_{SR}^2(x,y)\,|\,y\in\pi^2(\mathcal{S})\Big\} \qquad \forall x\in M,$$

with ψ^c continuous and $\pi^2(\mathcal{S})$ compact. Hence, given $\bar{s}\in(-\tau_k,\tau_k)$, there is $\bar{y}\in\pi^2(\mathcal{S})$ such that

$$\psi\left(\gamma_k\big(\bar{s};z\big)\right)=\psi^c\big(\bar{y}\big)-\frac{1}{2}d_{SR}^2\left(\gamma_k\big(\bar{s};z\big),\bar{y}\right).$$

Then we have for every $s\in(-\tau_k,\tau_k)$,

$$\begin{aligned}
\psi\left(\gamma_k(s;z)\right) &\geq \psi^c\big(\bar{y}\big)-\frac{1}{2}d_{SR}^2\left(\gamma_k(s;z),\bar{y}\right)\\
&\geq \psi^c\big(\bar{y}\big)-d_{SR}^2\left(\gamma_k(s;z),\gamma_{k,l}\big(\bar{s};z\big)\right)-d_{SR}^2\left(\gamma_k\big(\bar{s};z\big),\bar{y}\right)\\
&\geq \psi\left(\gamma_k\big(\bar{s};z\big)\right)-2H\big(z,p_k\big)\big|s-\bar{s}\big|^2\\
&\geq \psi\left(\gamma_k\big(\bar{s};z\big)\right)-4\tau_k\,H\big(z,p_k\big)\big|s-\bar{s}\big|.
\end{aligned}$$

This shows that each function $s\mapsto\psi\left(\gamma_k(s;z)\right)$ is locally Lipschitz on its domain. By Rademacher's theorem, we infer that it is differentiable almost everywhere on $(-\tau_k,\tau_k)$. Since the paths $\gamma_k(\cdot;z)$ with $z\in(x_k+\Pi_{p_k})\cap B(x_k,\rho_k)$ laminate a set bigger which is than the ball $B(x_k,\rho_k/2)$ in a continuous way, Fubini's theorem implies the existence of a negligeable set $N_{k,l}$ such that the property stated in the lemma holds for k. We conclude by setting $N=\cup N_k$. $\square$

Before starting the proof of (ii), we need a last result giving an estimates on the deviation of normal geodesics. For every (x,p), we set $(\gamma_{x,p}, p_{x,p}) := (\gamma_{x,p}(\cdot;x), p_{x,p}(\cdot;x))$, the solution of the Hamiltonian system starting at (x,p); it is defined on $(-\tau(H(x,p),|p|), \tau(H(x,p),|p|))$.

Lemma 3.4.14. *There is a function*

$$C \,:\, (0,+\infty)\times(0,+\infty) \longrightarrow (0,+\infty)$$

which is decreasing in the first variable and increasing in the second variable such that the following property holds:
For every $h,R>0$, every k, and every $(x,p)\in \mathcal{B}\times(\mathbb{R}^n)^*$ satisfying

$$H\big(x_k,p_k\big), H(x,p) > h, \quad \big|p_k\big|, |p| < R, \quad x \in B\big(x_k, \rho_k/2\big), \tag{3.39}$$

one has

$$\big|\gamma_k\big(t_k(x)+s; z_k(x)\big) - \gamma_{x,p}(s)\big| \le C(h,R)\,\big|p_k\big(t_k(x); z_k(x)\big) - p\big|\; s, \tag{3.40}$$

for every $s \in (-\tau(h,R), \tau(h,R)) \cap (-t_k(x)-\tau(h,R), -t_k(x)+\tau(h,R))$.

Proof of Lemma 3.4.14. Since the C^1-norms of the X^i's are bounded on $\mathcal{V}$, there is an increasing function $P:(0,+\infty)\to(0,+\infty)$ such that the solutions to our Hamiltonian system starting from a pair (x,p) with $x\in\mathcal{B}, H(x,p)>h$ and $|p|<R$ remains in the set $\mathcal{V}\times B(0,P(R))$ on the interval $(-\tau(h,R),\tau(h,R))$ (note that since H is constant along the Hamiltonian trajectories, the solutions remains in the set $\{H(x,p)>h\}$). Now, considering Lipschitz constants of the Hamiltonian vector field on the "cylinder" $\mathcal{V}\times B(0,P(R))$ (the C^2-norms of the X^i's are bounded on $\mathcal{V}$) and using Gronwall's Lemma (see Appendix A), we prove easily the existence of an increasing function $C:[0,+\infty)\to[0,+\infty)$ such that

$$\begin{aligned}\big|\gamma_k\big(t_k(x)+s; z_k(x)\big) - \gamma_{x,p}(s)\big| + \big|p_k\big(t_k(x)+s; z_k(x)\big) - p_{x,p}(s)\big| \\ \le C(R)\,\big|p_k\big(t_k(x); z_k(x)\big) - p\big|,\end{aligned} \tag{3.41}$$

for every $h,R>0$, every k, and every $(x,p)\in\mathcal{B}\times(\mathbb{R}^n)^*$ satisfying (3.39), and every $s\in(-\tau(h,R),\tau(h,R))\cap(-t_k(x)-\tau(h,R),-t_k(x)+\tau(h,R))$. Let us denote by I the latter interval and set

$$u(s) := \big|\gamma_k\big(t_k(x)+s; z_k(x)\big) - \gamma_{x,p}(s)\big| \qquad \forall s\in I.$$

Considering again the Lipschitz constants of the Hamiltonian vector field that we always denote by K, we obtain formally for every s,

$$\begin{aligned} u(s) &= \left|\int_0^s \frac{\partial H}{\partial p}\left(\gamma_k\big(t_k(x)+r; z_k(x)\big), p_k\big(t_k(x)+r; z_k(x)\big)\right)\right. \\ &\qquad\qquad \left. -\frac{\partial H}{\partial p}\left(\gamma_{x,p}(r), p_{x,p}(r)\right)\,dr\right| \\ &\le K\int_0^s u(r)\,dr + K\int_0^s \big|p_k\big(t_k(x)+r; z_k(x) - p_{x,p}(r)\big|\,dr, \end{aligned}$$

which by (3.41) gives

$$u(s) \le K\int_0^s u(r)\,dr + K\int_0^s C(R)\,\big|p_k\big(t_k(x); z_k(x)\big) - p\big|\,dr.$$

Gronwall's Lemma (see Lemma A.1.1) concludes the proof. □

We are now ready to prove that for every $x \in \mathcal{B} \setminus N$, we have $\partial_c \psi(x) = \{x\}$. Fix $x \in \mathcal{B} \setminus N$ and argue by contradiction, that is assume that there is $\bar{y} \neq x$ such that $\bar{y} \in \partial^c \psi(x) \setminus \{x\}$. Then we have (remembering Remark 3.2.12)

$$\psi(x) + c\big(x, \bar{y}\big) \leq \psi(z) + c\big(z, \bar{y}\big) \qquad \forall z \in M,$$

which can be written as

$$\psi(x) - \psi(z) \leq \frac{1}{2} d_{SR}^2\big(z, \bar{y}\big) - \frac{1}{2} d_{SR}^2\big(x, \bar{y}\big) \qquad \forall z \in M. \tag{3.42}$$

Since d_{SR} is Lipschitz in charts outside the diagonal, there is a normal minimizing geodesic joining x to $\bar{y}$ (see Remark 3.4.9), that is there is $p \in T_x^* M$ such that $\exp_x(p) = \bar{y}$ and $d_{SR}(x, y)^2 = 2H(x, p) \neq 0$. Note that since x belongs to $\partial_c \psi(x)$, we have

$$\psi(x) = \psi(x) + c(x, x) \leq \psi(z) + c(z, x) \qquad \forall z \in \mathcal{V}.$$

Set $h := H(x, p)/2$, $R := 2|p|$ and pick k such that

$$H\left(x_k, p_k\right) > h, \quad \left|p_k\right|, |p| < R, \quad x \in B\left(x_k, \rho_k/2\right).$$

Applying the previous inequality with $z = \gamma_k\left(t_k(x) + s; z_k(x)\right)$ and s small yields

$$\begin{aligned}
& \psi\left(\gamma_k\big(t_k(x); z_k(x)\big)\right) = \psi(x) \\
\leq \quad & \psi\left(\gamma_k\big(t_k(x) + s; z_k(x)\big)\right) + \frac{1}{2} d_{SR}^2\left(\gamma_k\big(t_k(x) + s; z_k(x)\big), x\right) \\
\leq \quad & \psi\left(\gamma_k\big(t_k(x) + s; z_k(x)\big)\right) + H\left(z_k(x), p_k\right) s^2,
\end{aligned}$$

because $\gamma_k\left(\cdot; z_k(x)\right)$ is an horizontal path joining $x = \gamma_k\left(t_k(x); z_k(x)\right)$ to $\gamma_k\left(t_k(x) + s; z_k(x)\right)$ of length $s\sqrt{2H\left(z_k(x), p_k\right)}$. Since x does not belong to N, the function

$$s \longmapsto \psi\left(\gamma_k\big(t_k(x) + s; z_k(x)\big)\right)$$

is differentiable at $s = 0$. Then Lemma 3.4.13 together with the previous inequality allows us to write

$$\frac{d}{ds} \psi\left(\gamma_k\big(t_k(x) + s; z_k(x)\big)\right)_{|s=0} = 0. \tag{3.43}$$

Since d_{SR} is Lipschitz outside the diagonal and $\bar{y} \neq x$, there are $\rho, K > 0$ such that

$$\left|d_{SR}^2\big(z, \bar{y}\big) - d_{SR}^2\big(z', \bar{y}\big)\right| \leq K|z' - z| \qquad \forall z, z' \in B(x, \rho).$$

Then applying (3.42) with $z = \gamma_k\big(t_k(x) + s; z_k(x)\big)$ and s small and using (3.40) yields (as in Lemma 3.4.14, $\gamma_{x,p}$ denotes the geodesic starting at x with initial covector p, note that $\gamma_{x,p}(s)$ belongs to $\mathcal{V}$ for small s)

$$\begin{aligned}
& \psi(x) - \psi\left(\gamma_k\big(t_k(x) + s; z_k(x)\big)\right) \\
\leq \quad & \frac{1}{2} d_{SR}^2\left(\gamma_k\big(t_k(x) + s; z_k(x)\big), \bar{y}\right) - \frac{1}{2} d_{SR}^2\left(x, \bar{y}\right) \\
\leq \quad & \frac{K}{2}\left|\gamma_k\big(t_k(x) + s; z_k(x)\big) - \gamma_{x,p}(s)\right| + \frac{1}{2} d_{SR}^2\big(\gamma_{x,p}(s), \bar{y}\big) - \frac{1}{2} d_{SR}^2\left(x, \bar{y}\right) \\
\leq \quad & \frac{KC(h, R)}{2}\left|p_k\big(t_k(x); z_k(x)\big) - p\right| s + \frac{1}{2}(1 - s)^2 d_{SR}^2\big(x, \bar{y}\big) - \frac{1}{2} d_{SR}^2\big(x, \bar{y}\big) \\
= \quad & \left(\frac{KC(h, R)}{2}\left|p_k\big(t_k(x); z_k(x)\big) - p\right| - d_{SR}^2\big(x, \bar{y}\big)\right) s + \frac{d_{SR}^2(x, y)}{2} s^2.
\end{aligned}$$

The quantity

$$\frac{KC(h,R)}{2}\left|p_k\big(t_k(x);z_k(x)\big)-p\right|$$

tends to 0 as (x_k, p_k) tends to (x, p). We infer that for (x_k, p_k) close enough to (x, p), the derivative of the function $s \mapsto \psi\left(\gamma_k\big(t_k(x)+s;z_k(x)\big)\right)$ cannot be zero. This contradicts (3.43).

It remains to prove the formula for $T(x)$ and the uniqueness of minimizing geodesic between x and $T(x)$ μ-almost everywhere. We need to show that

$$\partial^c\psi(x)\cap \mathrm{Supp}(\nu)=\exp_x\left(\frac{1}{2}D_x\psi\right)$$

for all $x \in \mathcal{M}^\psi \cap \mathrm{Supp}(\mu)$ where ψ is differentiable, which is the case for μ-almost every $x \in \mathcal{M}^\psi$ by assertion (i) and Rademacher's theorem. This is a consequence of Lemma 2.3.6 applied to the function $z \mapsto -\psi(z)+\psi^c(y)$ at the point x with $y \in \partial\psi_c(x)$. Moreover, again by Lemma 2.3.6, the geodesic from x to $T(x)$ is unique for μ-a.e. $x \in \mathcal{M}^\psi \cap \mathrm{Supp}(\mu)$. Since $T(x)=x$ for $x \in \mathcal{S}^\psi \cap \mathrm{Supp}(\mu)$, the geodesic is clearly unique also in this case. □

Remark 3.4.15. *If the sub-Riemannian structure is assumed to be ideal, then the potential ψ can be shown to be locally semiconcave on the moving set.*

Remark 3.4.16. *The above arguments show that Theorem 3.4.11 remains true under more general assumptions. Let (Δ, g) be a complete sub-Riemannian structure on M and μ, ν be two compactly supported probability measures in M with μ is absolutely continuous with respect to the Lebesgue measure. Assume that there are two open sets $\Omega_1, \Omega_2 \subset M$ with*

$$\mu\left(M\setminus\Omega_1\right)=0 \quad \text{and} \quad \mathit{Supp}(\nu)\subset\Omega_2$$

such that the sub-Riemannian distance is Lipschitz in charts on $(\Omega_1\times\Omega_2)\setminus D$. Then there is existence and uniqueness of an optimal transport map with respect to the sub-Riemannian quadratic cost.

3.5 Other examples

We conclude the present chapter with a list of examples for which we have existence and uniqueness of optimal transport maps for the SR quadratic cost, that is the cost $c: M\times M \to [0,+\infty)$ defined by

$$c(x,y):=\frac{1}{2}d_{SR}^2(x,y) \qquad \forall x,y\in M.$$

Given a cost function, we shall say that the Monge problem is *well-posed*, if we have existence and uniqueness of optimal transport maps from an absolutely continuous compactly supported measure to a compactly supported measure. All the examples that we review below have already been encoutenred within the text.

Fat distributions

Recall (see Example 1.3.13) that a distribution Δ on M is called *fat* if, for every $x \in M$ and every section X of Δ with $X(x) \neq 0$, there holds

$$T_xM = \Delta(x) + \big[X, \Delta\big](x), \tag{3.44}$$

where

$$\big[X, \Delta\big](x) := \Big\{[X, Z](x) \,|\, Z \text{ section of } \Delta\Big\}.$$

We saw that fat distributions do not admit non-trivial singular horizontal paths. This means that any complete sub-Riemannian structure associated with a fat distribution is ideal. In conclusion, by Theorem 3.4.11, the Monge problem for any sub-Riemannian structure associated with a fat distributions is well-posed.

Two-generating distributions

A distribution Δ is called *two-generating* if

$$T_xM = \Delta(x) + [\Delta, \Delta](x) \qquad \forall x \in M.$$

Two-generating distributions do not admit Goh paths (see Example 2.4.5). By Theorem 3.4.11, the Monge problem for any sub-Riemannian structure associated with a two-generating distributions is well-posed.

Totally nonholonomic distributions on three-dimensional manifolds

Assume that M has dimension 3, that Δ is a nonholonomic rank-two distribution on M, and define

$$\Sigma_\Delta := \Big\{x \in M \mid \Delta(x) + [\Delta, \Delta](x) \neq \mathbb{R}^3\Big\}.$$

The set Σ_Δ is called the *singular set* or the *Martinet set* of Δ.

Proposition 3.5.1. *Let Δ be a totally nonholonomic distribution on a three-dimensional manifold. Then, the set Σ_Δ is a closed subset of M which is countably 2-rectifiable. Moreover, a non-trivial horizontal path $\gamma : [0, 1] \to M$ is singular if and only if it is included in Σ_Δ.*

Proof. The first part will follow from Proposition 3.5.2 while the second part has already been proved in Example 1.3.15. $\square$

Proposition 3.5.1 implies that for any pair $(x, y) \in M \times M$ (with $x \neq y$) such that x or y does not belong to Σ_Δ, any sub-Riemannian minimizing geodesic between x and y is nonsingular. Moreover Σ_Δ has Lebesgue measure zero. As a consequence, by Remarks 3.4.4 and 3.4.16, the Monge problem is well-posed.

Medium-fat distributions

The distribution Δ is called *medium-fat* if, for every $x \in M$ and every vector field X on M such that $X(x) \in \Delta(x) \setminus \{0\}$, there holds

$$T_xM = \Delta(x) + [\Delta, \Delta](x) + [X, [\Delta, \Delta]](x).$$

As shown in Example 2.4.5, medium-fat distributions do not admit non-trivial Goh paths. As a consequence, the Monge problem for sub-Riemannian structures involving medium-fat distributions is well-posed.

Codimension-one nonholonomic distributions

Let M have dimension n and Δ be a nonholonomic distribution of rank $n-1$. As in the case of nonholonomic distributions on three-dimensional manifolds, we can define the singular set associated to the distribution as

$$\Sigma_\Delta := \Big\{ x \in M \mid \Delta(x) + [\Delta, \Delta](x) \neq T_xM \Big\}.$$

The following result holds.

Proposition 3.5.2. *If Δ is a nonholonomic distribution of rank $n-1$, then the set Σ_Δ is a closed subset of M which is countably $(n-1)$-rectifiable. Moreover, any Goh path is contained in Σ_Δ.*

Proof. The fact that Σ_Δ is a closed subset of M is obvious. Let us prove that it is countably $(n-1)$-rectifiable. Since it suffices to prove the result locally, we can assume that we have

$$\Delta(x) = \mathrm{Span}\Big\{ X^1(x), \ldots, X^{n-1}(x) \Big\} \qquad \forall x \in \mathcal{V},$$

where $\mathcal{V}$ is an open neighborhood of the origin in $\mathbb{R}^n$. Moreover, doing a change of coordinates if necessary, we can also assume that (with coordinates $(x_1, \ldots, x_n)$)

$$X^i = \partial_{x_i} + \alpha_i(x)\, \partial x_n \qquad \forall i = 1, \ldots, n-1,$$

where each $\alpha_i : \mathcal{V} \longrightarrow \mathbb{R}$ is a C^∞ function satisfying $\alpha_i(0) = 0$. Hence, for any $i, j \in \{1, \ldots n-1\}$, we have

$$\big[X^i, X^j\big] = \left[\left(\frac{\partial \alpha_j}{\partial x_i} - \frac{\partial \alpha_i}{\partial x_j} \right) + \left(\frac{\partial \alpha_j}{\partial x_n}\alpha_i - \frac{\partial \alpha_i}{\partial x_n}\alpha_j \right) \right] \partial_{x_n},$$

and so

$$\Sigma_\Delta = \left\{ x \in \mathcal{V} \mid \left(\frac{\partial \alpha_j}{\partial x_i} - \frac{\partial \alpha_i}{\partial x_j} \right) + \left(\frac{\partial \alpha_j}{\partial x_n}\alpha_i - \frac{\partial \alpha_i}{\partial x_n}\alpha_j \right) = 0 \quad \forall i, j \in \{1, \ldots, n-1\} \right\}.$$

For every tuple $I = (i_1, \ldots, i_k) \in \{1, \ldots, n-1\}^k$ we denote by X^I the smooth vector field constructed by Lie brackets of $X^1, X^2, \ldots, X^{n-1}$ as follows,

$$X^I = \left[X^{i_1}, \left[X^{i_2}, \ldots, \left[X^{i_{k-1}}, X^{i_k}\right] \ldots\right]\right].$$

We call $k = \text{length}(I)$ the length of the Lie bracket X^I. Since Δ is totally nonholonomic, there is some positive integer r such that

$$\mathbb{R}^n = \text{Span}\Big\{X^I(x) \mid \text{length}(I) \leq r\Big\} \qquad \forall x \in \mathcal{V}.$$

It is easy to see that, for every I such that $\text{length}(I) \geq 2$, there is a smooth function $g_I : \mathcal{V} \to \mathbb{R}$ such that

$$X^I(x) = g_I(x)\partial_{x_n} \qquad \forall x \in \mathcal{V}.$$

Defining the sets A_k as

$$A_k := \Big\{x \in \mathcal{V} \mid g_I(x) = 0 \quad \forall I \text{ such that } \text{length}(I) \leq k\Big\},$$

we have

$$\Sigma_\Delta = \bigcup_{k=2}^{r} \left(A_k \setminus A_{k+1}\right).$$

By the Implicit Function Theorem, it is easy to see that each set $A^k \setminus A^{k+1}$ can be covered by a countable union of smooth hypersurfaces. Indeed assume that some given x belongs to $A_k \setminus A_{k+1}$. This implies that there is some $J = (j_1, \ldots, j_{k+1})$ of length $k+1$ such that $g_J(x) \neq 0$. Set $I = (j_2, \ldots, j_{k+1})$. Since $g_I(x) = 0$, we have

$$g_J(x) = \left(\frac{\partial g_I}{\partial x_{j_1}}(x) + \frac{\partial g_I}{\partial x_n}(x)\alpha_{j_1}(x)\right) \neq 0.$$

Hence, either $\frac{\partial g_I}{\partial x_{j_1}}(x) \neq 0$ or $\frac{\partial g_I}{\partial x_n}(x) \neq 0$.
Consequently, we deduce that we have the following inclusion

$$A^k \setminus A^{k+1} \subset \bigcup_{\text{length}(I)=k} \left\{x \in \mathcal{V} \mid \exists\, i \in \{1, \ldots, n\} \text{ such that } \frac{\partial g_I}{\partial x_i}(x) \neq 0\right\}.$$

We conclude easily.
The fact that any Goh path is contained in Σ_Δ is obvious. □

As a consequence by Remarks 3.4.4 and 3.4.16, the Monge problem for sub-Riemannian structures involving codimension one distributions is well-posed.

Rank-two distributions in dimension four

Let (M, Δ, g) be a complete sub-Riemannian manifold of dimension four, and let Δ be a *regular* rank-two distribution, that is satisfying

$$T_xM = \text{Span}\Big\{X^1(x), X^2(x), \left[X^1, X^2\right](x), \left[X^1, \left[X^1, X^2\right]\right](x), \left[X^2, \left[X^1, X^2\right]\right](x)\Big\}$$

for any local parametrization $\mathcal{F} = \{X^1, X^2\}$ of the distribution. In Example 1.3.17, we saw that there is a smooth horizontal vector field X on M such that the singular horizontal paths γ parametrized by arc-length are exactly the integral curves of X, i.e. the curves satisfying

$$\dot{\gamma}(t) = X(\gamma(t)).$$

For every $x \in M$, denote by $\mathcal{O}(x)$ the orbit of x by the flow of X and set

$$\Omega := \Big\{(x, y) \in M \times M \mid y \notin \mathcal{O}(x)\Big\}.$$

According to Remark 3.4.4, the following result holds:

Proposition 3.5.3. *Under the assumption above, the function d_{SR} is locally semiconcave in the interior of Ω.*

The above result allow us to obtain existence and uniqueness of optimal transport maps in certain cases. Let us consider the distribution given in Example 1.3.16, that is the distribution Δ in $\mathbb{R}^4$ spanned by the vector fields

$$X^1 = \partial x_1, \qquad X^2 = \partial x_2 + x_1 \partial x_3 + x_3 \partial x_4.$$

As shown in Example 1.3.16, an horizontal path $\gamma : [0, 1] \to \mathbb{R}^4$ is singular if and only if it satisfies, up to reparameterization by arc-length,

$$\dot{\gamma}(t) = X^1\big(\gamma(t)\big) \qquad \forall t \in [0, 1].$$

By the above proposition, we deduce that, for any complete metric g on $\mathbb{R}^4$, the sub-Riemannian distance function d_{SR} is locally semiconcave on the set

$$\Omega = \Big\{(x, y) \in \mathbb{R}^4 \times \mathbb{R}^4 \mid (y - x) \notin \mathrm{Span}\{e_1\}\Big\},$$

where e_1 denotes the first vector in the canonical basis of $\mathbb{R}^4$. Consequently, for any pair of compactly supported probability measures μ, ν on M such that μ is absolutely continuous with respect to the Lebesgue measure and

$$\mathrm{Supp}\big(\mu \times \nu\big) \subset \Omega,$$

the Monge problem is well-posed.

3.6 Notes and comments

In 1781, Monge's original work [Mon81] was concerned with the moving of soil that was modelized as an optimal transport problem consisting in minimizing the transportation cost

$$\int_{\mathbb{R}^3} |T(x) - x| \, d\mu(x),$$

between continuous distributions of mass. The Monge problem was rediscovered several decades later, in 1942, by Kantorovitch [Ka42] who proved a duality theorem to study the relaxed form of the problem (which is by now refered as Kantorovitch problem). We refer the reader to the textbooks [Vil03, Vil08] by Villani for historial accounts on the optimal transport theory.

The Kantorovitch duality theorem which is not precisely stated in the present monograph appears through Theorem 3.2.13 and Corollary 3.2.14. Actually, our presentation of the theory leading to existence and uniqueness of optimal transport maps closely follows the one of Gangbo and McCann in [GM96]. For sake of simplicity, we restrict our attention to transportation problems between compactly supported probability measures from a smooth manifold into itself with continuous costs. Most of the results of Sections 3.1-3.2 remain true in the more general context of lower semicontinuous costs on the product of two Polish spaces and non-compactly supported probability measures. We refer the reader to Villani's monograph [Vil08] for general statements.

As seen through Example 3.1.2, transport maps may not exist. In fact, Pratelli [Pra07] proved that transport maps do exist as soon as the initial measure is assumed to be non-atomic. The Prokhrorov Theorem which is used in the proof of Theorem 3.2.1 can be found in Billingsley's book [Bil99]. Theorem 3.2.13 extends a result by Rockafellar [Roc66] about the sub-differentials of convex functions. The sub-TWIST condition introduced in Section 3.3 is a natural extension of the classical TWIST condition (see [Vil08]). Thanks to Lemma 2.3.6, many costs obtained in a variational way do satisfy the sub-TWIST condition. This is the case of the quadratic Euclidean cost appearing in Example 3.3.8, or of the quadratic geodesic cost appearing in Example 3.3.10. In fact, Examples 3.3.8-3.3.10 refer respectively to theorems by Brenier [Bre91] and McCann [McC01]. This type of result can be developped further by considering locally Lipschitz costs associated with problems of calculus of variations involving Tonelli Lagrangians (see [BB07]) or even with some optimal control problems (see [AL09]). As seen in Example 3.1.3, minimizers of the original Monge problem with cost $c(x, y) = |y - x|$ in $\mathbb{R}^n$ may not be unique. However, existence of optimal transport maps can be proved, see [Vil08] and references therein.

The study of Monge-type problems in sub-Riemannian geometry began with a paper by Ambrosio and Rigot [AR04] about the transportation problem in the Heisenberg group (see also [Rig05]). Then, Agrachev and Lee [AL09] extended the well-posedness result of Ambrosio-Rigot to the case of sub-Riemannian quadratic costs which are Lipschitz in charts on $M \times M$ (see Remark 3.3.11). Then, Figalli and the author [FR10] removed the assumption of Lipschitzness on the diagonal; this is Theorem 3.4.11. We observe that our proof of assertion (ii) differs from the original proof in [FR10] which was based on a Pansu-Rademacher Theorem. All these results are concerned with SR quadratic costs (that is $c = d_{SR}^2$). As in the Euclidean case, the Monge problem for the non-quadratic cost $c = d_{SR}$ does not enjoy uniqueness. Using techniques developped by Champion and De Pascale [CDP11], De Pascale and Rigot [DPR11] obtained an existence result for the classical Monge problem in the Heisenberg group.

The local semiconcavity of some SR distances outside the diagonal is demonstrated in Theorem 3.4.3. Such regularity is fundamental and sometimes necessary. First, it shows that distances of ideal sub-Riemannian structures share the same type of properties as Riemannian distances, at least outside the diagonal. It can be useful to get Sard's theorems and as a consequence regularity properties of sub-Riemannian spheres, see [Rif04, Rif06]. Then, the semiconcavity of the cost allows to consider probability measures which do not charge rectifiable sets and hence not necessarily continuous, see [Vil08]. Finally, semiconcavity of the cost may be transfered to potentials (see Remark 3.4.15) and then permit to get a Monge-Ampère-like equation (see Remark 3.1.1). This latter consequence is due to a famous theorem by Alexandrov (see [EG92]) which states that locally semiconvace functions are two times differentiable almost everywhere. We refer the reader to our paper [FR10] for further details on sub-Riemannian Monge-Ampère equations, to [CR05, RT07] for further details on semiconcave SR distances, and to the Cannarsa-Sinestrari's book [CS04] for an detailed exposition on semiconcavity.

We do not know if the Monge problem (for the SR quadratic cost) is well-posed for general sub-Riemannian structures. The method presented in this chapter requires regularity properties for d_{SR}. According to the Mitchell ball-box theorem (see [Jea12, Mit85, Mon02]), the sub-Riemannian distance is always locally Hölder in charts. In Chapter 2, we saw that given a complete sub-Riemannian structure and $x \in M$ the function

$$y \in M \longmapsto d_{SR}(x, y)$$

is Lipschitz in charts on a dense subset of M. We do not know if this set has necessarily full Lebesgue measure in M (note that the Sard Conjecture that we mentioned in Section 2.6 would imply such a result). Anyway, such a result would not be sufficient to prove the well-posedness of Monge problem for general sub-Riemannian structures...

Appendix A

Ordinary differential equations

We recall here basic facts on ordinary differential equations. For further details, we refer the reader to the textbook [HS74].

A.1 Preliminaries

Absolutely continuous curves

A function $f : [a, b] \to \mathbb{R}^n$ is said to be *absolutely continuous*, if for each $\epsilon > 0$, there exists $\delta > 0$ such that for each family of disjoints intervals $\{]a_i, b_i[\}_{i\in\mathbb{N}}$ included in $[a, b]$, and satisfying

$$\sum_{i\in\mathbb{N}} b_i - a_i < \delta,$$

we have

$$\sum_{i\in\mathbb{N}} |f(b_i) - f(a_i)| < \epsilon.$$

Any absolutely continuous function is continuous. In fact, a function $f : [a, b] \to \mathbb{R}^n$ is absolutely continuous if and only if it is differentiable almost everywhere on $[a, b]$, its derivative $\dot{f}(t) := \frac{d}{dt} f(t)$ is integrable with respect to the Lebesgue measure on $[a, b]$, and we have for each $t \in [a, b]$,

$$f(t) = f(a) + \int_a^t \frac{d}{dt} f(s)ds \qquad \forall t \in [a, b].$$

A function $f : [a, b] \to \mathbb{R}^n$ is called *absolutely continuous with square integrable derivative* if it is absolutely continuous on $[a, b]$ and satisfies

$$\dot{f} \in L^2\big([a, b]; \mathbb{R}^n\big).$$

Let M be a smooth manifold without boundary of dimension $n \geq 2$. A function $f : [a, b] \to M$ is called absolutely continuous (resp. absolutely continuous with square integrable derivative) if it is absolutely continuous (resp. absolutely continuous with square integrable derivative) in charts. Such a notion does not depend on the atlas chosen to cover M.

The Gronwall Lemma

The Gronwall lemma is a key tool to obtain estimates involving solutions of differential equations.

Lemma A.1.1 (Gronwall's Lemma). *Let $\epsilon > 0$, $\alpha : [0,\epsilon] :\to \mathbb{R}$ be a continuous function, and $\beta \in L^1([0;\epsilon],\mathbb{R})$. Assume that $u : [0,\epsilon] \to \mathbb{R}$ is a continuous function satisfying*

$$u(t) \leq \alpha(t) + \int_0^t \beta(s)u(s)ds \qquad \forall t \in [0,\epsilon]. \tag{A.1}$$

Then there holds

$$u(t) \leq \alpha(t) + e^{\int_0^t \beta(s)ds} \int_0^t e^{-\int_0^s \beta(r)dr}\beta(s)\alpha(s)\,ds \qquad \forall t \in [0,\epsilon]. \tag{A.2}$$

If, in addition, α is nondecreasing, then

$$u(t) \leq \alpha(t)e^{\int_0^t \beta(s)ds} \qquad \forall t \in [0,\epsilon]. \tag{A.3}$$

Proof. Let us first assume that α is of class C^1 on $[0,\epsilon]$. Define the function $\mu : [0,\epsilon] \to \mathbb{R}$ by

$$\mu(t) := \alpha(t) + \int_0^t \beta(s)u(s)\,ds \qquad \forall t \in [0,\epsilon].$$

The function μ is absolutely continuous on $[0,\epsilon]$, and

$$\dot{\mu}(t) = \dot{\alpha}(t) + \beta(t)u(t) \qquad \text{a.e. } t \in [0,\epsilon].$$

By (A.1), we deduce that we have for almost every $t \in [0,\epsilon]$,

$$\dot{\mu}(t) \leq \dot{\alpha}(t) + \beta(t)\mu(t).$$

Which implies

$$\dot{\mu}(t) - \beta(t)\mu(t) \leq \dot{\alpha}(t) - \beta(t)\alpha(t) + \beta(t)\alpha(t) \qquad \text{a.e. } t \in [0,\epsilon].$$

Multiplying both sides by $e^{-\int_0^t \beta(s)ds}$, we obtain for almost every $t \in [0,\epsilon]$,

$$\frac{d}{dt}\left\{\mu(t)e^{-\int_0^t \beta(s)ds}\right\} \leq \frac{d}{dt}\left\{\alpha(t)e^{-\int_0^t \beta(s)ds}\right\} + e^{-\int_0^t \beta(s)ds}\beta(t)\alpha(t).$$

Integrating between 0 and $t \in [0,\epsilon]$, we obtain

$$\mu(t)e^{-\int_0^t \beta(s)ds} - \mu(0) \leq \alpha(t)e^{-\int_0^t \beta(s)ds} - \alpha(0) + \int_0^t e^{-\int_0^s \beta(r)dr}\beta(s)\alpha(s)ds.$$

Using the fact that $\mu(0) = \alpha(0)$ together with (A.1), we get (A.2). If α is merely continuous, we can find for each positive integer k, a C^1 function $\alpha_k : [0,\epsilon] \to \mathbb{R}$ satisfying

$$|\alpha(t) - \alpha_k(t)| \leq \frac{1}{k} \qquad \forall t \in [0,\epsilon].$$

Moreover, by assumption, the function u satisfies

$$u(t) \leq \left[\alpha_k(t) + \frac{1}{k}\right] + \int_0^t \beta(s)u(s)\,ds \qquad \forall t \in [0,\epsilon].$$

Applying the result proved in the first part of the proof and passing to the limit gives the result. If α is nondecreasing, the we have for every $t \in [0,\epsilon]$,

$$\begin{aligned} e^{\int_0^t \beta(s)ds} \int_0^t e^{-\int_0^s \beta(r)dr} \beta(s)\alpha(s)\,ds &\leq \alpha(t) e^{\int_0^t \beta(s)ds} \int_0^t e^{-\int_0^s \beta(r)dr} \beta(s)\,ds \\ &= \alpha(t)\left(e^{\int_0^t \beta(s)ds} - 1\right). \end{aligned}$$

We conclude easily. □

A.2 Existence and uniqueness results

The Cauchy-Peano Theorem

Let $I \subset \mathbb{R}$ be an open interval, Ω be an open subset of $\mathbb{R}^n$, and $f : I \times \Omega \to \mathbb{R}^n$ be a function satisfying the following property:

(H_{CP}) For every $x \in \Omega$, there exist $\delta > 0$, a locally integrable function $c : I \to [0,+\infty)$, and a nondecreasing function $\omega : [0,+\infty) \to [0,+\infty)$ with $\omega(h) \to 0$ as $h \to 0$ such that

$$|f(t,y) - f(t,z)| \leq c(t)\omega(|y-z|) \quad \text{and} \quad |f(t,y)| \leq c(t)$$

for almost all $t \in I$ and all $y, z \in B(x,\delta)$.

Given $(t_0, x_0) \in I \times \Omega$, our aim is to solve locally the following Cauchy problem

$$\dot{x}(t) = f(t, x(t)), \quad \text{a.e. } t, \qquad x(t_0) = x_0. \tag{A.4}$$

Theorem A.2.1 (Cauchy-Peano's Theorem)**.** *Assume that $f : I \times \Omega \to \mathbb{R}^n$ satisfies the property (H_{CP}). Then for every $(t_0, x_0) \in I \times \Omega$, there is $\epsilon > 0$ such that the Cauchy problem (A.4) admits a solution on $[t_0 - \epsilon, t_0 + \epsilon]$.*

Proof. The proof consists in applying the classical Euler iterative scheme. By assumption (H_{CP}), there are $\delta > 0$ and $c \in L^1_{loc}(I, [0,+\infty))$ such that

$$|f(t,y) - f(t,z)| \leq c(t)\omega(|y-z|) \quad \text{and} \quad |f(t,y)| \leq c(t) \tag{A.5}$$

for almost every $t \in I$ and all $y, z \in B(x_0, \delta)$. Since c is locally integrable on I, there is $\epsilon > 0$ such that $[t_0 - \epsilon, t_0 + \epsilon] \subset I$ and

$$\int_{t_0-\epsilon}^{t_0+\epsilon} M(t)\,dt < \delta. \tag{A.6}$$

We are going to prove that the Cauchy problem (A.4) admits a solution on $[t_0, t_0 + \epsilon]$. Let

$$\pi = \left\{t_0, t_1, \cdots, t_N\right\}$$

be a partition of $[t_0, t_0 + \epsilon]$; we recall that the diameter $\mu(\pi)$ of π is given by

$$\mu(\pi) := \max\Big\{t_{i+1} - t_i \mid 0 \le i \le N-1\Big\}.$$

We proceed by considering, on the interval $[t_0, t_1]$, the Cauchy problem

$$\dot{x}(t) = f(t, x_0), \quad x(t_0) = x_0.$$

It has a unique solution x on $[t_0, t_1]$ given by

$$x(t) = x_0 + \int_{t_0}^{t} f(s, x_0)\, ds \qquad \forall t \in [t_0, t_1].$$

Note that by (A.6), one has

$$|x(t) - x_0| \le \int_{t_0}^{t} |f(s, x_0)|\, ds \le \int_{t_0}^{t} M(s)\, ds < \delta,$$

for every $t \in [t_0, t_1]$; we set $x_1 := x(t_1)$. Next we iterate, by considering on $[t_1, t_2]$ the Cauchy problem

$$\dot{x}(t) = f(t, x_1), \quad x(t_1) = x_1.$$

The next so-called node of the scheme is $x_2 := x(t_2)$. We proceed in this manner until an arc x_π has been defined no all of $[t_0, t_0 + \epsilon]$. By construction, we have

$$|x_\pi(t) - x_0| \le \int_{t_0 - \epsilon}^{t_0 + \epsilon} M(t)\, dt < \delta \qquad \forall t \in [t_0, t_0 + \epsilon].$$

Moreover, we have for any $t, t' \in [t_0, t_0 + \epsilon]$ such that $t < t'$,

$$|x_\pi(t') - x_\pi(t)| \le \int_{t}^{t'} M(s)\, ds \le w(t' - t),$$

where the function $w : [0, +\infty) \to [0, +\infty)$ is defined by

$$w(h) := \max\left\{\int_{t}^{t'} M(s)\, ds \mid t, t' \in [t_0, t_0 + \epsilon] \text{ s.t. } t < t'\right\},$$

for every $h \ge 0$. On each interval $[t_i, t_{i+1}]$, we have

$$\int_{t_i}^{t_{i+1}} |f(s, x_i) - f(s, x_\pi(s))|\, ds \le \omega\left(w\big(\mu(\pi)\big)\right) \int_{t_i}^{t_{i+1}} c(t)\, dt.$$

Thus we have for any $t \in [t_0, t_0 + \epsilon]$,

$$\left| x_\pi(t) - x_0 - \int_{t_0}^{t} f(s, x_\pi(s))\, ds \right| \le \omega\left(w\big(\mu(\pi)\big)\right) \int_{t_0}^{t} c(s)\, ds. \tag{A.7}$$

Now let π_j be a sequence of partitions such that $\pi_j \to 0$, that is such that $\mu(\pi_j) \to 0$ as $j \to \infty$. Then the family $\{x_{\pi_j}\}_j$ is equicontinuous and uniformly bounded. Then by the Ascoli-Arzelà Theorem, some subsequence of it converges uniformly to a continuous function x. By the Lebesgue dominating convergence theorem, the function $t \mapsto f(t, x_{\pi_j}(t)$ converges to the function $t \mapsto f(t, x(t))$ in L^1. Moreover, passing to the limit in (A.7) yields

$$x(t) = x_0 + f(s, x(s))\, ds \qquad \forall t \in [t_0, t_0 + \epsilon].$$

We proceed in the same way on $[t_0 - \epsilon, t_0]$ backwards in time. □

Remark A.2.2. *The Cauchy-Peano is only an existence result. In the autonomous case, it says that if $f : \Omega \to \mathbb{R}^n$ is continuous then for every $x_0 \in \Omega$, the Cauchy problem*

$$\dot{x}(t) = f(x(t)), \quad x(0) = x_0,$$

admits at least one solution locally. A counterexample to uniqueness is for example given by $f : \mathbb{R} \to \mathbb{R}$ defined by

$$f(x) := \sqrt{|x|} \qquad \forall x \in \mathbb{R}.$$

The Cauchy problem $\dot{x}(t) = f(x(t)), x(0) = 0$ admits two smooth solutions:

$$x(t) = 0 \quad \textit{and} \quad x(t) = \frac{t^2}{4} \qquad \forall t \in \mathbb{R}.$$

The Cauchy-Carathéodory Theorem

Let $I \subset \mathbb{R}$ be an open interval, Ω be an open subset of $\mathbb{R}^n$, and $f : I \times \Omega \to \mathbb{R}^n$ be a function satisfying the following property:

(H_{CC}) For every $x \in \Omega$, there exist $\delta > 0$ and a locally integrable function $c : I \to [0, +\infty)$ such that

$$|f(t,y) - f(t,z)| \leq c(t)|y - z| \quad \text{and} \quad |f(t,y)| \leq c(t)$$

for almost every $t \in I$ and all $y, z \in B(x, \delta)$.

The following result provides existence and uniqueness for the Cauchy problem (A.4).

Theorem A.2.3 (Cauchy-Carathéodory's Theorem)**.** *Assume that $f : I \times \Omega \to \mathbb{R}^n$ satisfies the property (H_{CC}). Then for every $(t_0, x_0) \in I \times \Omega$, there is $\epsilon > 0$ such that the Cauchy problem (A.4) admits a solution $x : [t_0 - \epsilon, t_0 + \epsilon] \to \Omega$. If $y : [t_0, t_0 + \epsilon] \to \Omega$ (or $y : [t_0 - \epsilon, t_0] \to \Omega$) is an other solution of (A.4), then $x(t) = y(t)$ for all $t \in [t_0 - \epsilon, t_0 + \epsilon]$.*

Proof. The local existence is a consequence of Theorem A.2.1. Assume that x, y are two solutions of the Cauchy problem (A.4) on $[t_0, t_0 + \epsilon]$. Set $u(t) := |x(t) - y(t)|$ for every $t \in [t_0, t_0 + \epsilon]$. We have for every $t \in [t_0, t_0 + \epsilon]$,

$$\begin{aligned} u(t) &= \left| \int_{t_0}^t f(s, x(s))\, ds - \int_{t_0}^t f(s, y(s))\, ds \right| \\ &\leq \int_{t_0}^t |f(s, x(s)) - f(s, y(s))|\, ds \\ &\leq \int_{t_0}^t c(s) u(s)\, ds. \end{aligned}$$

By Gronwall's Lemma, we obtain that $u(t) = 0$ for every $t \in [t_0, t_0 + \epsilon]$. $\square$

Remark A.2.4. *In the autonomous case, the Cauchy-Carathéodory Theorem says that if $f : \Omega \to \mathbb{R}^n$ is locally Lipschitz then for every $x_0 \in \Omega$, the Cauchy problem*

$$\dot{x}(t) = f(x(t)), \quad x(0) = x_0,$$

admits a solution locally and this solution is unique.

Global existence theorems

By the Cauchy-Carathéodory Theorem, under assumption (H_{CC}), for every $(t_0, x_0) \in I \times \Omega$, the unique solution to the Cauchy problem (A.4) can be extended to a maximal interval of the form $I = (\alpha, \beta)$ with $\alpha < t_0 < \beta$ and $\alpha \in \mathbb{R} \cup \{-\infty\}$, $\beta \in \mathbb{R} \cup \{+\infty\}$. Under additional assumptions, we can sometimes insure that any solution can be extended to $\mathbb{R}$.

Theorem A.2.5. *Let $f : \mathbb{R} \times \mathbb{R}^m \to \mathbb{R}^n$ be a function satisfying the assumptions (H_{CC}) (with $\Omega = \mathbb{R}^n$) and such that there exist two functions K, M in $L^1_{loc}(\mathbb{R}, [0, +\infty))$ such that*

$$|f(t, x)| \le K(t)|x| + M(t) \qquad a.e.\ t \in \mathbb{R} \quad \forall x \in \mathbb{R}^n. \tag{A.8}$$

Then any solution of $\dot{x} = f(x(t))$ can be extended to $\mathbb{R}$.

Proof. Let $(t_0, x_0) \in \mathbb{R} \times \mathbb{R}^n$ and $x : [\alpha, \beta] \to \mathbb{R}^n$ a maximal solution to the Cauchy problem (A.4) be fixed. We argue by contradiction and assume that $\beta < +\infty$ (the case $\alpha > -\infty$ is left to the reader). Set for every $t \in (\alpha, \beta)$, $u(t) := |x(t)|$. By (A.8), we have for every $t \in [t_0, \beta)$,

$$\begin{aligned} u(t) &= \left| x_0 + \int_{t_0}^t f(s, x(s))\, ds \right| \\ &\le |x_0| + \int_{t_0}^t K(s)u(s) + M(s)\, ds \\ &\le \left(|x_0| + \int_{t_0}^t M(s)\, ds \right) + \int_{t_0}^t K(s)u(s)\, ds. \end{aligned}$$

By Gronwall's Lemma, we infer that

$$|x(t)| \le \left(|x_0| + \int_{t_0}^{\beta} M(s)\, ds \right) e^{\int_{t_0}^{\beta} K(s) ds} =: C < +\infty \qquad \forall t \in [t_0, \beta).$$

Then for any $t < t' \in [t_0, \beta)$, we have

$$\begin{aligned} |x(t') - x(t)| &= \left| \int_t^{t'} f(s, x(s))\, ds \right| \\ &\le C \int_t^{t'} K(s)\, ds + \int_t^{t'} M(s)\, ds. \end{aligned}$$

Since both K, L are in $L^1_{loc}(\mathbb{R}, [0, +\infty))$, this shows that for each sequence $\{t_k\}_k$ converging to β from below, the sequence $\{x(t_k)\}_k$ is a Cauchy sequence. Then $x(t)$ has a limit $x(\beta)$ as t tends to β. By Theorem A.2.3, the Cauchy problem $\dot{y} = f(t, y(t)), y(\beta) = x(\beta)$ admits a local solution. This shows that x can be extended beyond β and yields a contradiction. □

Remark A.2.6. *For sake of simplicity, we stated Theorem A.2.5 in the case of a nonautonomous function defined on $\mathbb{R} \times \mathbb{R}^n$. The same results holds for a function defined on $I \times \mathbb{R}^n$ where I is an open interval in $\mathbb{R}$. Namely, any solution to the Cauchy problem can be extended to I.*

A.3 Linear systems

Let $I \subset \mathbb{R}$ be an interval and $A \in L^1(I; M_n(\mathbb{R}))$ be a function from I into the set of $n \times n$ matrices denoted by $M_n(\mathbb{R})$. By the above results, for every $t_0 \in I$, the Cauchy problem

$$\dot{S}(t) = A(t)S(t), \quad \text{a.e. } t \in I, \quad S(t_0) = I_n, \tag{A.9}$$

has a unique solution which is defined on I. In the same way, the Cauchy problem

$$\dot{Y}(t) = -Y(t)A(t), \quad \text{a.e. } t \in I, \quad Y(t_0) = I_n,$$

admits a solution defined on I. Hence, the function $Z : I \to M_n(\mathbb{R})$ defined as $Z(t) := Y(t)S(t)$ for every $t \in I$, satisfies for almost every $t \in I$,

$$\begin{aligned} \dot{Z}(t) &= \dot{Y}(t)S(t) + Y(t)\dot{S}(t) \\ &= -Y(t)A(t)S(t) + Y(t)A(t)S(t) = 0. \end{aligned}$$

Since $Z(t_0) = I_n$, we deduce by uniqueness, that $Z(t) = I_n$ for every $t \in I$. This shows that the matrix $S(t)$ is invertible for every $t \in I$.

Proposition A.3.1. *Let $C \in L^1_{loc}(I; \mathbb{R}^n)$, $t_0 \in I$, and $\xi_0 \in \mathbb{R}^n$. The solution to the Cauchy problem*

$$\dot{\xi}(t) = A(t)\xi(t) + C(t), \quad \textit{for a.e. } t \in I, \quad \xi(t_0) = \xi_0 \tag{A.10}$$

is given by

$$\xi(t) = S(t)\xi_0 + S(t)\int_{t_0}^{t} S(s)^{-1}C(s)ds, \quad \forall t \in I. \tag{A.11}$$

Proof. By uniqueness, it suffices to verify that the absolutely continuous function $\xi : I \to \mathbb{R}^n$ given by (A.11) satisfies the Cauchy problem (A.10). We have $\xi(t_0) = \xi_0$ and we verify that for almost every $t \in I$,

$$\begin{aligned} \dot{\xi}(t) &= \dot{S}(t)\xi_0 + \dot{S}(t)\int_{y_0}^{t} S(s)^{-1}C(s)ds + S(t)S(t)^{-1}C(t) \\ &= A(t)S(t)\xi_0 + A(t)S(t)\int_{y_0}^{t} S(s)^{-1}C(s)ds + C(t) \\ &= A(t)\xi(t) + C(t). \end{aligned}$$

□

Appendix B

Elements of differential calculus

We recall here basic facts of first order calculus in normed vector spaces and less basic facts of second order calculus. We refer the reader to textbook [AMR83] for further details on differential calculus in normed spaces. The results of second order calculus are taken from the textbook [AS04].

B.1 First order calculus

Differentials

Given two normed vector spaces $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$, we denote by $\mathcal{L}(X, Y)$ the space of continuous linear maps from X to Y. This space is equipped with the operator norm (we denote alternatively by $T \cdot u$ or $T(u)$ the image of u by the operator T)

$$\|T\| = \sup\Big\{\|T(u)\|_Y \mid u \in X, \|u\|_X = 1\Big\}.$$

Let $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$ be two normed vector spaces, $\mathcal{U}$ be an open subset of X and let $F : \mathcal{U} \subset X \to Y$ be a given mapping. Let $\bar{u} \in \mathcal{U}$. We say that F is differentiable at $\bar{u}$ provided there is a continuous linear map $D_{\bar{u}}F : X \to Y$ such that for every $\epsilon > 0$, there is $\delta > 0$ such that

$$0 < \big\|u - \bar{u}\big\|_X < \delta \quad \Longrightarrow \quad \frac{\big\|F(u) - F\big(\bar{u}\big) - D_{\bar{u}}F \cdot \big(u - \bar{u}\big)\big\|_Y}{\big\|u - \bar{u}\big\|_X} < \epsilon.$$

This property can also be written as

$$\lim_{u \to \bar{u}} \frac{\big\|F(u) - F\big(\bar{u}\big) - D_{\bar{u}}F \cdot \big(u - \bar{u}\big)\big\|_Y}{\big\|u - \bar{u}\big\|_X} = 0,$$

or

$$F(u) = F(\bar{u}) + D_{\bar{u}}F \cdot \big(u - \bar{u}\big) + \big\|u - \bar{u}\big\|_X\, o(1).$$

The map F is said to be differentiable in $\mathcal{U} \subset X$ if it is differentiable at every $u \in \mathcal{U}$. The map

$$\begin{array}{rcl} DF : \mathcal{U} & \longrightarrow & \mathcal{L}(X, Y) \\ u & \longmapsto & D_u F \end{array}$$

is called the derivative of F. If DF is a continuous map on $\mathcal{U}$ (where $\mathcal{L}(X,Y)$ has the norm topology) we say that F is of class C^1 on $\mathcal{U}$. Finally we recall that given a function F of class C^1 on an open set $\mathcal{U} \subset X$ and a point $u \in \mathcal{U}$, the derivative D_uF is called singular if it is not surjective and in that case u is called critical.

The Inverse Function Theorem

Here we provide the proof of the Inverse Function Theorem and of a quantitative version of it which are used through Chapters 1 to 3. Our proof is taken from Clarke's monograph [Cla83].

Theorem B.1.1 (Inverse Function Theorem)**.** *Let $\mathcal{U}$ be an open set of $\mathbb{R}^n$, $F : \mathcal{U} \to \mathbb{R}^n$ be a function of class C^1, and $x \in \mathcal{U}$ be such that D_xF is not singular. Then there exists neighborhoods $U \subset \mathcal{U}$ of x and V of $F(x)$ such that $F_{|U} : U \to V$ is a C^1 diffeomorphism.*

Proof. Note that since D_xF is linear from $\mathbb{R}^n$ to $\mathbb{R}^n$ and not singular, it is indeed invertible. Let $r > 0$ be such that for every $x' \in \bar{B}(x,r) \subset \mathcal{U}$, $D_{x'}F$ is nonsingular and

$$\|D_{x'}F - D_xF\| \leq \frac{1}{2\left\|\left(D_xF\right)^{-1}\right\|}. \tag{B.1}$$

Set

$$\delta := \frac{1}{2\left\|\left(D_xF\right)^{-1}\right\|}.$$

Lemma B.1.2. *We have for every $x', x'' \in \bar{B}(x,r)$,*

$$|F(x') - F(x'')| \geq \delta \, |x' - x''| \, .$$

Proof of Lemma B.1.2. Let $x', x'' \in \bar{B}(x,r)$ be fixed, we may suppose $x' \neq x''$. Set

$$v := \frac{x'' - x'}{|x'' - x'|}, \quad w := \frac{D_xF(v)}{|D_xF(v)|}, \quad \text{and } \lambda := |x'' - x'|.$$

We note that since D_xF is non-singular, w is well-defined. We have for every $t \in [0, \lambda]$,

$$\begin{aligned}
& \langle w, D_{x'+tv}F(v)\rangle \\
= \;& \langle w, D_xF(v)\rangle + \langle w, \left[D_{x'+tv}F - D_xF\right](v)\rangle \\
= \;& |D_xF(v)| + \langle w, \left[D_{x'+tv}F - D_xF\right](v)\rangle \quad \text{(by definition of } w) \\
\geq \;& |D_xF(v)| - \left|\left[D_{x'+tv}F - D_xF\right](v)\right| \quad \text{(by Cauchy-Schwarz)} \\
\geq \;& \frac{|D_xF(v)| \left\|\left(D_xF\right)^{-1}\right\|}{\left\|\left(D_xF\right)^{-1}\right\|} - \|D_{x'+tv}F - D_xF\| \quad (\text{since } |v| = 1) \\
\geq \;& \frac{1}{\left\|\left(D_xF\right)^{-1}\right\|} - \frac{1}{2\left\|\left(D_xF\right)^{-1}\right\|} \quad (\text{by } (B.1)) \\
= \;& \delta.
\end{aligned}$$

But since F is C^1 on $\mathcal{U}$, we have

$$F(x' + \lambda v) - F(x') = \int_0^\lambda D_{x'+tv}F(v)\,dt,$$

which together with the above calculus gives

$$\langle w, F(x' + \lambda v) - F(x')\rangle = \int_0^\lambda \langle w, D_{x'+tv}F(v)\rangle\,dt \geq \delta\lambda.$$

By the Cauchy-Schwarz inequality, we get $|F(x') - F(x'')| \geq \delta |x' - x''|$. $\square$

Lemma B.1.3. *The set $B\left(F(x), \frac{r\delta}{2}\right)$ is included in $F(B(x,r))$.*

Proof of Lemma B.1.3. Let $y \in B\left(F(x), \frac{r\delta}{2}\right)$ be fixed, xe define $\chi : \bar{B}(x,r) \to \mathbb{R}$ by

$$\forall x' \in \bar{B}(x,r), \quad \chi(x') := |y - F(x)|^2.$$

We need to show that there is $x' \in B(x,r)$ such that $\chi(x') = 0$. Since $\bar{B}(x,r)$ is compact, χ attains its minimum on that set at some point $\bar{x} \in \bar{B}(x,r)$. We claim first that $\bar{x}$ belongs to $B(x,r)$. Otherwise, by Lemma B.1.2 and the triangle inequality, one has

$$\begin{aligned}
\frac{r\delta}{2} > |y - F(x)| &\geq |F(\bar{x}) - F(x)| - |y - F(\bar{x})| \\
&\geq \delta|\bar{x} - x| - |y - F(\bar{x}| \\
&\geq \delta r - |y - F(x)| \quad \text{(by minimality of } \bar{x}) \\
&\geq \delta r - \frac{\delta r}{2} = \frac{\delta r}{2},
\end{aligned}$$

which is a contradiction. Thus $\bar{x}$ yields a local minimum for χ on $B(x,r)$, and consequently

$$D_{\bar{x}}\chi = 2(y - F(\bar{x})^* D_{\bar{x}}F = 0.$$

Since $D_{\bar{x}}F$ is non-singular (by construction of r), we obtain $y = F(\bar{x})$. $\square$

To prove the theorem, we now set $V = B\left(F(x), \frac{r\delta}{2}\right)$, and we define G on V as follows: for every $y \in V$, $G(y)$ is the unique $x' \in B(x,r)$ such that $F(x') = y$. We choose U as a neighborhood of x satisfying $U \subset \mathcal{U}$ and $F(U) \subset V$. By construction, $G \circ F = Id$ on U and $F \circ G = Id$ on V. Moreover, by Lemma B.1.2, the function G is $\frac{1}{\delta}$-Lipschitz. Let $y \in V$ and $h \neq 0$ be such that $y + h \in V$. There is $x' \in B(x,r)$ such that $F(x') = y$, moreover we have

$$\begin{aligned}
G(y+h) &= G(F(x') + h) \\
&= G\left(F\left(x' + \left(D_{x'}F\right)^{-1}(h)\right)\right) + o(h)) \\
&= G\left(F\left(x' + \left(D_{x'}F\right)^{-1}(h)\right)\right) + o(h) \\
&= x' + \left(D_{x'}F\right)^{-1}(h) + o(h).
\end{aligned}$$

Then

$$\lim_{h \to 0} \frac{\left|G(y+h) - G(y) - \left(D_{x'}F\right)^{-1}(h)\right|}{|h|} = 0.$$

Which proves that G is C^1 on V. $\square$

We now give a quantitative version of the Inverse Function Theorem which is useful in Section 1.4.

Theorem B.1.4. *Let $\rho, L, \mu, B > 0$ and $F : \Omega := B(0,\rho) \to \mathbb{R}^n$ be a function of class C^1 on Ω which satisfies the following properties:*

- $\|D_{x'}F - D_xF\| \leq L|x'-x|, \quad \forall x, x' \in \Omega,$
- $\|D_0F\| \leq B,$
- $\left|\det\left(D_0F\right)\right| \geq \mu.$

Then there is $r := r(n,\rho,L,\mu,B) > 0$ and a C^1 function $G : B(F(0),r) \to \Omega$ such that $F \circ G = Id$ and $G \circ F = Id$.

Proof. In the proof of Theorem B.1.1, we setted $V := B\left(F(x), \frac{r\delta}{2}\right)$ with $\delta := \frac{1}{2\left\|(D_0F)^{-1}\right\|}$ and $r > 0$ such that

$$\|D_{x'}F - D_xF\| \leq \delta, \quad \forall x, x' \in \bar{B}(0,r),$$

and D_xF non-singular for every $x \in \bar{B}(0,r)$.
There is a continuous and nondecreasing function $\omega : [0,+\infty) \to [0,+\infty)$ such that we have for any non-singular matrix A,

$$\left\|A^{-1}\right\| \leq \frac{\omega(\|A\|)}{|\det(A)|}.$$

For every $B' > 0$, there is a constant $K(B') > 0$ such that for any $n \times n$ matrices A, A' with $\|A\|, \|A'\| \leq B'$, we have

$$|\det(A) - \det(A')| \leq K(B')\,\|A - A'\|\,.$$

We have

$$\delta := \frac{1}{2\,\|D_0F^{-1}\|} \geq \frac{\left|\det\left(D_0F\right)\right|}{2\omega(\|D_0F\|)} \geq \frac{\mu}{2\omega(B)}.$$

Thus we can take

$$r := \min\left\{\rho, \frac{\mu}{2L\omega(B)}, \frac{\mu}{2}\Big(K(B+L\rho)L\Big)^{-1}\right\}.$$

□

The Lagrange Multiplier Theorem

The Lagrange Multiplier Theorem plays a major role in Chapter 2.

Theorem B.1.5 (Multipliers Lagrange Theorem)**.** *Let $(X, \|\cdot\|_X)$ be a normed vector space, $\mathcal{U}$ be an open subset of X, and $E : \mathcal{U} \to \mathbb{R}^n$ and $C : \mathcal{U} \to \mathbb{R}$ two mappings of class C^1 on $\mathcal{U}$. Assume that $\bar{u} \in \mathcal{U}$ satisfies the following property:*

$$C\left(\bar{u}\right) \leq C(u) \quad \text{for every} \quad u \in \mathcal{U} \quad \text{such that} \quad E(u) = E\left(\bar{u}\right). \tag{B.2}$$

Then there exist $\lambda_0 \in \mathbb{R}$ and $\lambda \in \mathbb{R}^n$ with $(\lambda_0, \lambda) \neq (0,0)$ such that

$$\lambda^* D_{\bar{u}}E = \lambda_0 D_{\bar{u}}C.$$

Proof. Define the mapping $\Phi : \mathcal{U} \subset X \to \mathbb{R} \times \mathbb{R}^n$ by

$$\Phi(u) := (C(u), E(u)), \quad \forall u \in \mathcal{U}.$$

The mapping Ψ is of class C^1 on $\mathcal{U}$. We claim that $\bar{u}$ is necessarily a critical point of Φ, that is $D_{\bar{u}}\Phi$ is singular. We argue by contradiction. If $\bar{u}$ is not a critical point, the continuous linear map $D_{\bar{u}}\Phi : X \to \mathbb{R} \times \mathbb{R}^n$ is surjective. Then there exists a linear subspace Y of X of dimension $n+1$ such that the restriction of $D_{\bar{u}}\Phi$ to Y is an isomorphism. Let $y_1, \ldots, y_{n+1}$ be a basis of Y and $\mathcal{B}$ be an open neighborhood of 0 in $\mathbb{R}^{n+1}$ such that

$$\bar{u} + \sum_{i=1}^{n+1} \beta_i \, y_i \in \mathcal{U} \qquad \forall \beta = \big(\beta_1, \ldots, \beta_{n+1}\big) \in \mathcal{B}.$$

The mapping

$$\begin{array}{rcl} \hat{\Phi} \, : \, \mathcal{B} & \longrightarrow & \mathbb{R}^{n+1} \\ \beta = \big(\beta_1, \ldots, \beta_{n+1}\big) & \longmapsto & \Phi\left(\bar{u} + \sum_{i=1}^{n+1} \beta_i \, y_i\right) \end{array}$$

is of class C^1 on $\mathcal{B}$ with a derivative which is invertible at $\beta = 0$. Hence, by the Inverse Function Theorem, the point $\Phi(\bar{u}) = (C(\bar{u}), E(\bar{u}))$ belongs to the interior of the image of $\hat{\Phi}(\mathcal{B})$. Thus for $\epsilon > 0$ small enough, there is $y \in Y$ with $\bar{u} + y \in \mathcal{U}$ such that

$$\Phi\big(\bar{u} + y\big) = \big(C\big(\bar{u}\big) - \epsilon, E\big(\bar{u}\big)\big),$$

which contradicts (B.2). In consequence, $\bar{u}$ is a critical point of Φ. Hence, there exists a non-zero $n+1$-tuple $p = (-\lambda_0, \lambda)$ (with $\lambda_0 \in \mathbb{R}$ and $\lambda \in \mathbb{R}^n$) which is orthogonal to the image of $D_{\bar{u}}\Phi$, that is such that

$$-\lambda_0 D_{\bar{u}} C + \lambda^* D_{\bar{u}} E = 0.$$

□

B.2 Second order study

Preliminaries

Let us denote by $\mathcal{L}^2(X, Y)$ the space of all continuous bilinear maps from $X \times X$ to Y. We can equip it with the operator norm

$$\|T\| = \sup\Big\{ \big\|T(u_1, u_2)\big\|_Y \,|\, u_1, u_2 \in X, \|u_1\|_X = \|u_2\|_X = 1 \Big\}.$$

Given an open set $\mathcal{U} \subset X$ and a mapping $F : \mathcal{U} \subset X \to Y$, we define

$$D^2 F := D\big(DF\big) \, : \mathcal{U} \subset X \longrightarrow \mathcal{L}^2(X, Y)$$

if it exists (where we identify $\mathcal{L}(X, \mathcal{L}(X, Y))$ with $\mathcal{L}^2(X, Y)$. If $D^2 F$ exists and is continuous on $\mathcal{U}$, we say that F is of class C^2 on $\mathcal{U}$. In this case, the second derivative $D_u^2 F$ is symmetric at any point, that is

$$D_u^2 F \cdot (v, w) = D_u^2 F \cdot (w, v) \qquad \forall v, w \in X, \, \forall u \in \mathcal{U}.$$

If $F:\mathcal{U}\subset X\rightarrow Y$ is a function of class C^2 then we have for every $u\in\mathcal{U}$ the second order Taylor formula

$$F(u+h)=F(u)+D_uF(h)+\frac{1}{2}D_u^2F\cdot(h,h)+\|h\|_X^2\,o(1),$$

which means that

$$\lim_{v\rightarrow u}\frac{\left\|F(v)-F(u)-D_uF\cdot(v-u)-\frac{1}{2}D_u^2F\cdot(v-u,v-u)\right\|_Y}{\|v-u\|_X^2}=0.$$

By the Inverse Function Theorem, any function of class C^1 is locally open around any point with an invertible derivative. We are going to provide a second-order sufficient condition for local openness around critical points.

A second-order sufficient condition for local openness

Let $(X,\|\cdot\|_X)$ be a normed vector space, N be a positive integer, $\mathcal{U}$ be an open subset of X and $F:\mathcal{U}\rightarrow\mathbb{R}^N$ be a mapping of class C^2 on $\mathcal{U}$. Given a critical point $u\in\mathcal{U}$, we call corank of u, the quantity

$$\mathrm{corank}_F(u):=N-\dim\left(\mathrm{Im}\big(D_uF\big)\right).$$

We also recall that if $Q:X\rightarrow\mathbb{R}$ is a quadratic form (that is Q is defined by $Q(v):=B(v,v)$ with $B:X\times X\rightarrow\mathbb{R}$ a symmetric bilinear form), we define its negative index by

$$\mathrm{ind}_-(Q):=\max\left\{\dim(L)\mid Q_{|L\setminus\{0\}}<0\right\},$$

where $Q_{|L\setminus\{0\}}<0$ means

$$Q(u)<0\qquad\forall u\in L\setminus\{0\}.$$

The following result provides a sufficient condition for local openness around a critical point.

Theorem B.2.1. *Let $F:\mathcal{U}\rightarrow\mathbb{R}^N$ be a mapping of class C^2 in an open set $\mathcal{U}\subset X$ and $\bar{u}\in\mathcal{U}$ be a critical point of F of corank r. If*

$$ind_-\left(\lambda^*\left(D_{\bar{u}}^2F\right)_{|Ker(D_{\bar{u}}F)}\right)\geq r\qquad\forall\lambda\in\left(Im\big(D_{\bar{u}}F\big)\right)^{\perp}\setminus\{0\},\tag{B.3}$$

then the mapping F is locally open at $\bar{u}$, that is the image of any neighborhood of $\bar{u}$ is an neighborhood of $F(\bar{u})$.

In the above statement, $\left(D_{\bar{u}}^2F\right)_{|\mathrm{Ker}(D_{\bar{u}}F)}$ refers to the quadratic mapping from $\mathrm{Ker}(D_{\bar{u}}F)$ to $\mathbb{R}^N$ defined by

$$\left(D_{\bar{u}}^2F\right)_{|\mathrm{Ker}(D_{\bar{u}}F)}(v):=D_{\bar{u}}^2F\cdot(v,v)\qquad\forall v\in\mathrm{Ker}(D_{\bar{u}}F).$$

The following result is a quantitative version of the previous theorem, it is useful in Section 3.4. (We denote by $B_X(\cdot,\cdot)$ the balls in X with respect to the norm $\|\cdot\|_X$.)

Theorem B.2.2. *Let $F : \mathcal{U} \to \mathbb{R}^N$ be a mapping of class C^2 in an open set $\mathcal{U} \subset X$ and $\bar{u} \in \mathcal{U}$ be a critical point of F of corank r. If*

$$ind_- \Big(\lambda^* \left(D^2_{\bar{u}}F\right)_{|Ker(D_{\bar{u}}F)}\Big) \geq r \qquad \forall \lambda \in \big(Im\big(D_{\bar{u}}F\big)\big)^{\perp} \setminus \{0\}, \tag{B.4}$$

then there exist $\bar{\epsilon}, c \in (0,1)$ such that for every $\epsilon \in (0, \bar{\epsilon})$ the following property holds: For every $u \in \mathcal{U}, z \in \mathbb{R}^N$ with

$$\|u - \bar{u}\|_X < \epsilon, \quad |z - F(u)| < c\,\epsilon^2, \tag{B.5}$$

there are $w_1, w_2 \in X$ such that $u + w_1 + w_2 \in \mathcal{U}$,

$$z = F\big(u + w_1 + w_2\big), \tag{B.6}$$

and

$$w_1 \in Ker(D_uF)\,, \quad \big\|w_1\big\|_X < \epsilon, \quad \big\|w_2\big\|_X < \epsilon^2. \tag{B.7}$$

The proof of Theorems B.2.1 and B.2.2 that we give in the next sections are taken from the Agrachev-Sachkov textbook [AS04] and the Agrachev-Lee article [AL09].

Proof of Theorem B.2.1

We need two preliminary lemmas.

Lemma B.2.3. *Let $G : \mathbb{R}^k \to \mathbb{R}^l$ be a mapping of class C^2 with $G(0) = 0$. Assume that there is*

$$\bar{v} \in Ker(D_0G) \quad with \quad D_0^2G \cdot \big(\bar{v}, \bar{v}\big) \in Im\big(D_0G\big), \tag{B.8}$$

such that the linear mapping

$$w \in Ker(D_0G) \longmapsto Proj_{\mathcal{K}}\left[D_0^2G \cdot \big(\bar{v}, w\big)\right] \in \mathcal{K} \tag{B.9}$$

is surjective, where $\mathcal{K} := Im(D_0G)^{\perp}$ and $Proj_{\mathcal{K}} : \mathbb{R}^l \to \mathcal{K}$ denotes the orthogonal projection onto $\mathcal{K}$. Then there is a sequence $\{u_i\}_i$ converging to 0 in $\mathbb{R}^k$ such that $G(u_i) = 0$ and $D_{u_i}G$ is surjective for any i.

Proof. Let E a vector space in $\mathbb{R}^k$ such that $\mathbb{R}^k = E \oplus \mathrm{Ker}(D_0G)$. Since $D_0^2G \cdot \big(\bar{v}, \bar{v}\big)$ belongs to $\mathrm{Im}\big(D_0G\big)$ there is $\hat{v} \in E$ such that

$$D_0G\big(\hat{v}\big) = -\frac{1}{2}D_0^2G \cdot \big(\bar{v}, \bar{v}\big).$$

Define the family of mappings $\{\Phi_\epsilon\}_{\epsilon>0} : E \times \mathrm{Ker}(D_0G) \to \mathbb{R}^l$ by

$$\Phi_\epsilon(z,t) := \frac{1}{\epsilon^5}G\left(\epsilon^2\bar{v} + \epsilon^3 t + \epsilon^4\hat{v} + \epsilon^5 z\right) \qquad \forall (z,t) \in E \times \mathrm{Ker}(D_0G), \forall \epsilon > 0.$$

For every $\epsilon > 0$, Φ_ϵ is of class C^2 on $E \times \mathrm{Ker}(D_0G) \to \mathbb{R}^l$ and its derivative at $(z,t) = (0,0)$ is given by

$$D_{(0,0)}\Phi_\epsilon(Z,T) = D_{\epsilon^2\bar{v}+\epsilon^4\hat{v}}G(Z) + \frac{1}{\epsilon^2}D_{\epsilon^2\bar{v}+\epsilon^4\hat{v}}G(T),$$

for any $(Z,T) \in E \times \mathrm{Ker}(D_0G)$. For every $(Z,T) \in E \times \mathrm{Ker}(D_0G)$, the first term of the right-hand side $D_{\epsilon^2\bar{v}+\epsilon^4\hat{v}}G(Z)$ tends to $D_0G(Z)$ as ϵ tends to 0 and since

$$\begin{aligned} \frac{1}{\epsilon^2} D_{\epsilon^2\bar{v}+\epsilon^4\hat{v}}G(T) &= \frac{1}{\epsilon^2}\Big[D_0G(T) + D_0^2G \cdot \big(\epsilon^2\bar{v}+\epsilon^4\hat{v},T\big) + \big|\epsilon^2\bar{v}+\epsilon^4\hat{v}\big|\, o(1)\Big] \\ &= \frac{1}{\epsilon^2}\Big[D_0^2G \cdot \big(\epsilon^2\bar{v}+\epsilon^4\hat{v},T\big) + \big|\epsilon^2\bar{v}+\epsilon^4\hat{v}\big|\, o(1)\Big], \end{aligned}$$

the second term tends to $D_0^2G(\bar{v},T)$ as ϵ tends to 0. By (B.9), the linear mapping

$$(Z,T) \in E \times \mathrm{Ker}(D_0G) \longmapsto D_0G(Z) + D_0^2G \cdot \big(\bar{v},T\big) \in \mathbb{R}^l$$

is surjective. Then there is $\bar{\epsilon} > 0$ such that $D_0\Phi_\epsilon$ is surjective for all $\epsilon \in (0,\bar{\epsilon})$. Therefore for every $\epsilon \in (0,\bar{\epsilon})$ the set

$$\Big\{(z,t) \in E \times \mathrm{Ker}(D_0G) \,|\, \hat{\Phi}_\epsilon(z,t) = 0\Big\}$$

is a submanifold of class C^2 of dimension $k-l>0$ which contains the origin. Then there is a sequence $\{(z_i,t_i)\}_i$ converging to the origin such that $\Psi_{1/i}(z_i,t_i) = 0$ and $D_{(z_i,t_i)}\Psi_{1/i}$ is surjective for all i large enough. Thus setting

$$u_i := \frac{1}{i^2}\bar{v} + \frac{1}{i^3}t_i + \frac{1}{i^4}\hat{v} + \frac{i^5}{z\ i} \qquad \forall i,$$

we get $G(u_i) = 0$ and $D_{u_i}G$ surjective for all i large enough. This proves the lemma. □

Lemma B.2.4. *Let $Q : \mathbb{R}^k \to \mathbb{R}^m$ be a quadratic mapping such that*

$$\mathit{ind}_-\,(\lambda^*Q) \geq m, \quad \forall \lambda \in (\mathbb{R}^m) \setminus \{0\}. \tag{B.10}$$

Then the mapping Q has a regular zero, that is there is $v \in \mathbb{R}^k$ such that $Q(v) = 0$ and D_vQ is surjective.

Proof. Since Q is a quadratic mapping, there is a symmetric bilinear map $B : \mathbb{R}^k \times \mathbb{R}^k \to \mathbb{R}^m$ such that

$$Q(v) = B(v,v) \qquad \forall v \in \mathbb{R}^k.$$

The kernel of Q, denoted by $\mathrm{Ker}(Q)$ is the set of $v \in \mathbb{R}^k$ such that

$$B(v,w) = 0 \qquad \forall w \in \mathbb{R}^k.$$

It is a vector subpace of $\mathbb{R}^k$. Up to considering the restriction of Q to a vector space E satisfying $E \oplus \mathrm{Ker}(Q) = \mathbb{R}^k$, we may assume that $\mathrm{Ker}(Q) = 0$. We now prove the result by induction on m.
In the case $m = 1$, we need to prove that there is $v \in \mathbb{R}^k$ with $Q(v) = 0$ and $D_vQ \neq 0$. By (B.10), we know that $\mathrm{ind}_-\,(Q) \geq 1$ and $\mathrm{ind}_-\,(-Q) \geq 1$, which means that there are two vector lines L^+, L^- in $\mathbb{R}^k$ such that $Q_{|L^+\setminus\{0\}} < 0$ and $Q_{|L^-\setminus\{0\}} > 0$. Then the restriction of Q to $L^+ \oplus L^-$ is a quadratic form which is sign-indefinite. Such a form has regular zeros.
Let us now prove the statement of the lemma for a fixed $m > 1$ under the

assumption that it has been proven for all values less than m. So we consider a quadratic mapping $Q : \mathbb{R}^k \to \mathbb{R}^m$ satisfying (B.10) and such that $\mathrm{Ker}(Q) = \{0\}$. We distinguish two cases:

First case: $Q^{-1}(0) \neq \{0\}$.
Take any $v \neq 0$ such that $Q(v) = 0$. If v is a regular point, then the statement of the lemma follows. Thus we assume that v is a critical point of Q. Since $D_vQ(w) = 2B(v,w)$ for all $w \in \mathbb{R}^k$ and $\mathrm{Ker}(Q) = \{0\}$, the derivative $D_vQ : \mathbb{R}^k \to \mathbb{R}^m$ cannot be zero. Then its kernel $E = \mathrm{Ker}(D_vQ)$ has dimension $k - r$ with $r := \mathrm{rank}(D_vQ) \in [1, m-1]$. Set $F := \mathrm{Im}(D_vQ)^\perp$ and define the quadratic form

$$\tilde{Q} \,:\, E \simeq \mathbb{R}^{k-r} \longrightarrow F \simeq \mathbb{R}^{m-r}$$

by

$$\tilde{Q}(w) := \mathrm{Proj}_F\big(Q(w)\big) \qquad \forall w \in E,$$

where $\mathrm{Proj}_F : \mathbb{R}^m \to F$ denotes the orthogonal projection to F. We have for every $\lambda \in F$ and every $w \in E$,

$$\lambda^* \tilde{Q}(w) = \lambda^* Q(w).$$

We claim that $\mathrm{ind}_-\left(\lambda^* Q\right) \geq m - r$, for every $\lambda \in F \setminus \{0\}$. As a matter of fact, by assumption, for every $\lambda \in F \setminus \{0\}$ there is a vector space L of dimension m such that $(\lambda^* Q)_{|L\setminus\{0\}} < 0$. The space $L \cap E$ has dimension at least $m - k$ as the intersection of L of dimension m and E of dimension $k - r$ in $\mathbb{R}^k$. By induction, we infer that $\tilde{Q}$ has a regular zero $\tilde{w} \in E = \mathrm{Ker}(D_vQ)$, that is $Q(\tilde{w}) \in \mathrm{Im}(D_vQ)$ and

$$w \in E = \mathrm{Ker}(D_vQ) \longmapsto \mathrm{Proj}_F\left(B\big((\tilde{w}, w)\big)\right) \in F$$

is surjective. Define $F : \mathbb{R}^k \to \mathbb{R}^m$ by

$$F(u) := Q\big(v + u\big) \qquad \forall u \in \mathbb{R}^k.$$

The function F is of class C^2 verifies $D_0F = D_vQ, D_0^2F = B$ and the assumptions of Lemma B.2.3 are satisfied with $\bar{v} = \tilde{w}$. We deduce that Q has a regular zero as well.

Second case: $Q^{-1}(0) = \{0\}$.
In fact, we are going to prove that this case cannot appear. First we claim that Q is surjective. Since Q is homogeneous ($Q(rv) = r^2 Q(v)$ for all $v \in \mathbb{R}^k$ and $r \in \mathbb{R}$), we have

$$Q(\mathbb{R}^k) = \Big\{ rQ(v) \,|\, r \geq 0, v \in \mathbb{S}^{k-1} \Big\}.$$

The set $Q(\mathbb{S}^{k-1})$ is compact, hence $Q(\mathbb{R}^k)$ is closed. Assume that $Q(\mathbb{R}^k) \neq \mathbb{R}^m$ and take $x = Q(v)$ on the boundary of $Q(\mathbb{R}^k)$. Then x is necessarily a critical point for Q. Proceeding as in the first case, we infer that $x = Q(w)$ for some non-critical point. This gives a contradiction. Then we have $Q(\mathbb{R}^k) = \mathbb{R}^m$. Consequently the mapping

$$\begin{array}{rrcl} \mathcal{Q} := \frac{Q}{|Q|} \,:\, & \mathbb{S}^{k-1} & \longrightarrow & \mathbb{S}^{m-1} \\ & v & \longmapsto & \frac{Q(v)}{|Q(v)|} \end{array}$$

is surjective. By Sard's Theorem (see [GG73]), it has a regular value x, that is $x \in \mathbb{S}^{m-1}$ such that $D_v\mathcal{Q}$ is surjective for all $v \in \mathbb{S}^{k-1}$ satisfying $\mathcal{Q}(v) = x$ for all $v \in \mathbb{S}^{k-1}$. Among the set of $v \in \mathbb{S}^{k-1}$ such that $\mathcal{Q}(v) = x$ take $\bar{v}$ for which $|Q(v)|$ is minimal, that is such that

$$Q(\bar{v}) = \bar{a}x$$

and

$$\forall a > 0, \forall v \in \mathbb{S}^{k-1}, \quad Q(v) = ax \Longrightarrow a \geq \bar{a}.$$

In other terms, if we define the smooth function $\Psi : (0, +\infty) \times \mathbb{S}^{k-1} \to \mathbb{R}^m$ as,

$$\Psi(a, v) := Q(v) - ax, \qquad \forall a > 0,\ \forall v \in \mathbb{S}^{k-1},$$

then the pair $(\bar{a}, \bar{v})$ satisfies

$$\bar{a} \leq a \quad \text{for every} \quad (a, v) \in (0, +\infty) \times \mathbb{S}^{k-1} \quad \text{with } \Psi(a, v) = 0.$$

By the Lagrange Multiplier Theorem (Theorem B.1.5), there is $\lambda_0 \in \mathbb{R}$ and $\lambda \in \mathbb{R}^m$ with $(\lambda_0, \lambda) \neq (0, 0)$ such that

$$\lambda^* D_{\bar{v}} Q = 0 \quad \text{and} \quad -\lambda^* x = \lambda_0.$$

Note that we have for every $h \in T_{\bar{v}}\mathbb{S}^{k-1} \subset \mathbb{R}^k$, we have

$$\begin{aligned} D_{\bar{v}}\mathcal{Q}(h) &= \frac{1}{|Q(\bar{v})|} D_{\bar{v}}Q(h) + [D_{\bar{v}}|Q|(h)]\, Q(\bar{v}) \\ &= \frac{1}{\bar{a}} D_{\bar{v}}Q(h) + \bar{a}\,[D_{\bar{v}}|Q|(h)]\, x. \end{aligned} \tag{B.11}$$

Consequently, if $\lambda_0 = 0$ (that is if $(\bar{a}, \bar{v})$ is a critical point of ψ), then $\lambda^* D_{\bar{v}}\mathcal{Q} = 0$ which contradicts the fact $D_{\bar{v}}\mathcal{Q}$ is surjective (because λ cannot be collinear with x by 2-homogeneity of Q). In conclusion, we can assume without loss of generality that $\lambda_0 = -1$. Since $(\bar{a}, \bar{v})$ is not a critical point of ψ, the set

$$\mathcal{C} = \Big\{(a, v) \in (0, +\infty) \times \mathbb{S}^{k-1} \mid \Psi(a, v) = 0\Big\}$$

is a smooth submanifold of $(0, +\infty) \times \mathbb{S}^{k-1}$ of dimension $k-m$ in a neighborhood of $(\bar{a}, \bar{v})$. Then for every $(h_a, h_v) \in \mathrm{Ker}(D_{\bar{a},\bar{v}}\Psi)$, which is equivalent to $h_a = 0$ and $D_{\bar{v}}Q(h_v) = 0$ with $h_v \in T_{\bar{v}}\mathbb{S}^{k-1}$, there is a smooth curve $\gamma = (\gamma_a, \gamma_v) : (-\epsilon, \epsilon) \to \mathcal{C}$ such that $\gamma(0) = (\bar{a}, \bar{v})$ and $\dot{\gamma}(0) = (h_a, h_v)$. Then differentiating two times the equality $\Psi(\gamma(t)) = 0$ and using that $\frac{\partial^2 \Psi}{\partial a^2} = 0$ and $\lambda^* \frac{\partial \Psi}{\partial v}(\bar{a}, \bar{v}) = \lambda^* D_{\bar{v}} Q = 0$, we get

$$\lambda^* \frac{\partial^2 \Psi}{\partial v^2}\big(\bar{a}, \bar{v}\big) = \lambda^* \ddot{\gamma}(0) \frac{\partial \Psi}{\partial a}\big(\bar{a}, \bar{v}\big) = \ddot{\gamma}(0)\lambda^* x = \ddot{\gamma}(0).$$

Note that $\frac{\partial^2 \Psi}{\partial v^2} = Q$. Furthermore, since $(\bar{a}, \bar{v})$ is solution to our minimization problem with constraine, we have $\gamma_a(t) \geq \bar{a} = \gamma_a(0)$ for all $t \in (-\epsilon, \epsilon)$. Then we have

$$\lambda^* Q(h) \geq 0 \qquad \forall h \in \mathrm{Ker}\big(D_{\bar{v}}Q\big) \cap T_{\bar{v}}\mathbb{S}^{k-1}.$$

Since $Q(\bar{v}) = \bar{a} > 0$ we have indeed

$$\lambda^* Q(h) \geq 0 \qquad \forall h \in \big(\mathrm{Ker}\big(D_{\bar{v}}Q\big) \cap T_{\bar{v}}\mathbb{S}^{k-1}\big) \oplus \mathbb{R}\bar{v} =: L. \tag{B.12}$$

Let us compute the dimension of the non-negative subspace L of the quadratic form $\lambda^* Q$. Since $D_{\bar{v}}\mathcal{Q}$ is surjective, we have

$$\dim\left(\mathrm{Im}\big(D_{\bar{v}}\mathcal{Q}\big)\right) = m-1.$$

Which means (remember (B.11)) that $\mathrm{Im}\left(D_{\bar{v}}Q_{|\mathbb{S}^{k-1}}\right)$ has dimension m or $m-1$. But $\lambda^* D_{\bar{v}}Q = 0$ with $\lambda \neq 0$, thus we have necessarily

$$\dim\left(\mathrm{Im}\big(D_{\bar{v}}Q_{|\mathbb{S}^{k-1}}\big)\right) = m-1$$

and

$$\begin{aligned}\dim\left(\mathrm{Ker}\big(D_{\bar{v}}Q\big)\cap T_{\bar{v}}\mathbb{S}^{k-1}\right) = \dim\left(\mathrm{Ker}\big(D_{\bar{v}}Q_{|\mathbb{S}^{k-1}}\big)\right) &= k-1-(m-1)\\ &= k-m.\end{aligned}$$

Consequently, $\dim(L) = k-m+1$, thus $\mathrm{ind}_-\left(\lambda^* Q\right)$ has to be $\leq m-1$, which contradicts the hypothesis of the lemma. This shows that $Q^{-1}(0) = \{0\}$ is impossible and concludes the proof of the lemma. □

We are ready to prove Theorem B.2.1. Set

$$S := \left\{\lambda \in \left(\mathrm{Im}\big(D_{\bar{u}}F\big)\right)^{\perp} \mid |\lambda| = 1\right\} \subset \mathbb{R}^N.$$

By assumption (B.3), for every $\lambda \in S$, there is a subspace $E_\lambda \subset \mathrm{Ker}\left(D_{\bar{u}}F\right)$ of dimension r such that

$$\lambda^*\left(D_{\bar{u}}^2 F\right)_{|E_\lambda\setminus\{0\}} < 0.$$

By continuity of the mapping $\nu \mapsto \nu^*\left(D_{\bar{u}}^2F\right)_{|E_\lambda}$, there is an open set $\mathcal{O}_\lambda \subset S$ such that

$$\nu^*\left(D_{\bar{u}}^2F\right)_{|E_\lambda\setminus\{0\}} < 0 \qquad \forall \nu \in \mathcal{O}_\lambda.$$

Choose a finite covering

$$S = \bigcup_{i=1}^{I} \mathcal{O}_{\lambda_i}$$

and a finite dimensional space $E \subset X$ such that

$$\mathrm{Im}\left(D_{\bar{u}}F_{|E}\right) = \mathrm{Im}\big(D_{\bar{u}}F\big).$$

Then the restriction $\tilde{F}$ of F to the finite dimensional subspace $E + \sum_{i=1}^{I} E_{\lambda_i} \subset X$ satisfies

$$\mathrm{ind}_-\left(\lambda^*\left(D_{\bar{u}}^2\tilde{F}\right)_{|\mathrm{Ker}(D_{\bar{u}}\tilde{F})}\right) \geq r \qquad \forall \lambda \in \left(\mathrm{Im}\big(D_{\bar{u}}\tilde{F}\big)\right)^{\perp}\setminus\{0\},$$

with

$$r = \mathrm{corank}_F(\bar{u}) := N - \dim\left(\mathrm{Im}\big(D_{\bar{u}}F\big)\right) = N - \dim\left(\mathrm{Im}\big(D_{\bar{u}}\tilde{F}\big)\right).$$

Set $\mathcal{K} := \left(\mathrm{Im}\left(D_{\bar{u}}\tilde{F}\right)\right)^{\perp}$ and define the quadratic mapping $Q : \mathrm{Ker}(D_{\bar{u}}\tilde{F}) \to \mathcal{K}$ by

$$Q(v) := \mathrm{Proj}_{\mathcal{K}}\left[\left(D_{\bar{u}}^2\tilde{F}\right)\cdot(v,v)\right] \qquad \forall v \in \mathrm{Ker}\left(D_{\bar{u}}\tilde{F}\right),$$

where $\mathrm{Proj}_{\mathcal{K}} : \mathbb{R}^N \to \mathcal{K}$ denotes the orthogonal projection onto $\mathcal{K}$. The assumption (B.10) of Lemma B.2.4 is satisfied. Then by Lemmas B.2.4, Q has a regular zero, that is $\bar{v} \in \mathrm{Ker}(D_{\bar{u}}\tilde{F})$ such that

$$Q(\bar{v}) = 0 \quad \Longleftrightarrow \quad D^2_{\bar{u}}\tilde{F} \cdot (\bar{v}, \bar{v}) \in \mathcal{K} = \mathrm{Im}\left(D_{\bar{u}}\tilde{F}\right)$$

and

$$\begin{aligned} &D_{\bar{v}}Q \text{ surjective} \\ &\qquad \Longleftrightarrow \quad w \in \mathrm{Ker}\big(D_{\bar{u}}\tilde{F}\big) \;\mapsto\; \mathrm{Proj}_{\mathcal{K}}\left[D^2_{\bar{u}}\tilde{F} \cdot (\bar{v}, w)\right] \in \mathcal{K} \text{ surjective.} \end{aligned}$$

Setting $G(v) := \tilde{F}(\bar{u}+v) - \tilde{F}(\bar{u})$ and applying Lemma B.2.3, we get a sequence $\{u_i\}_i$ converging to $\bar{u}$ such that $F(u_i) = F(\bar{u})$ and $D_{u_i}\tilde{F}$ is surjective for any i. By the Inverse Function Theorem, this implies that F is locally open at $\bar{u}$.

Proof of Theorem B.2.2

Proceeding as in the proof of Theorem B.2.1, we may assume that X is finite dimensional. We may also assume that $\bar{u} = 0$ and $F(\bar{u}) = 0$. As before, set $\mathcal{K} := (\mathrm{Im}\,(D_{\bar{u}}F))^{\perp}$ and define the quadratic mapping $Q : \mathrm{Ker}(D_0F) \to \mathcal{K}$ by

$$Q(v) := \mathrm{Proj}_{\mathcal{K}}\left[\left(D_0^2F\right) \cdot (v, v)\right] \qquad \forall v \in \mathrm{Ker}\,(D_0F)\,,$$

where $\mathrm{Proj}_{\mathcal{K}} : \mathbb{R}^N \to \mathcal{K}$ denotes the orthogonal projection onto $\mathcal{K}$. By (B.4) and Lemmas B.2.4, Q has a regular zero $\bar{v} \in \mathrm{Ker}(D_0F)$. Let E be a vector space in $\mathbb{R}^k$ such that $X = E \oplus \mathrm{Ker}(D_0F)$. Define $G : E \times \mathrm{Ker}(D_0F) \to \mathbb{R}^N$ by

$$G(z,t) := D_0F(z) + \frac{1}{2}\left(D_0^2F\right) \cdot (t,t) \qquad \forall (z,t) \in E \times \mathrm{Ker}(D_0F).$$

Then assumptions of Lemma B.2.3 are satisfied and there is a sequence $\{(z_i, t_i)\}_i$ converging to 0 such that $G(z_i, t_i) = 0$ and $D_{(z_i,t_i)}G$ is surjective for all i.

Lemma B.2.5. *There are $\mu, c > 0$ such that the image of any continuous mapping $\tilde{G} : B(0,1) \to \mathbb{R}^N$ with*

$$\sup\Big\{\Big|\tilde{G}(u) - G(u)\Big| \mid u = (z,t) \in B_X(0,1)\Big\} \leq \mu \tag{B.13}$$

contains the ball $\bar{B}(0,c)$.

Proof. This is a consequence of the Brouwer Theorem which asserts that any continuous mapping from $\bar{B}(0,1) \subset \mathbb{R}^n$ into itself has a fixed point, see [Bre93]. Let i large enough such that $u_i := (t_i, z_i)$ belongs to $B(0, 1/4)$. Since $D_{u_i}G$ is surjective, there is a affine space V of dimension N which contains u_i and such that $D_{u_i}G_{|V}$ is invertible. Then by the Inverse Function Theorem, there is a open ball $\mathcal{B} = B_X(u_i, \rho) \cap V$ of u_i in V such that the mapping

$$G_{|V} : \mathcal{B} \longrightarrow G_{|V}(\mathcal{B}) \subset \mathbb{R}^N$$

is a smooth diffeomophism. We denote by $\mathcal{G} : G_{|V}(\mathcal{B}) \to \mathcal{B}$ its inverse. The set $G_{|V}(\mathcal{B})$ contains some closed ball $\bar{B}(0, c)$. Taking $c > 0$ sufficiently small we may assume that

$$\mathcal{G}(y) \in B_X\big(u_i, \rho/4\big) \qquad \forall y \in \bar{B}(0, c).$$

There is $\mu > 0$ such that any continuous mapping $\tilde{G} : B_X(0, 1) \to \mathbb{R}^N$ verifying (B.13) satisfies

$$\tilde{G}(u) \in G_{|V}(\mathcal{B}) \qquad \forall u \in B_X(u_i, \rho/2) \cap V$$

and

$$\Big|\big(\mathcal{G} \circ \tilde{G}\big)(u) - u\Big| \leq \frac{\rho}{4} \qquad \forall u \in B_X(u_i, \rho/2) \cap V.$$

Let $\tilde{G} : B_X(0, 1) \to \mathbb{R}^N$ be a continuous mapping verifying (B.13) and $y \in \bar{B}(0, c)$ be fixed. By the above construction, the function

$$\Psi : B_X(\mathcal{G}(y), \rho/4) \longrightarrow B_X(\mathcal{G}(y), \rho/4)$$

defined by

$$\Psi(u) := u - \big(\mathcal{G} \circ \tilde{G}\big)(u) + \mathcal{G}(y) \qquad \forall u \in B_X(\mathcal{G}(y), \rho/4),$$

is continuous from $B_X(\mathcal{G}(y), \rho/4)$ into itself. Thus by Brouwer's Theorem, it has a fixed point, that is there is $u \in B_X(\mathcal{G}(y), \rho/4)$ such that

$$\Psi(u) = u \quad \Longleftrightarrow \quad \tilde{G}(u) = y.$$

This concludes the proof of the lemma. $\square$

Define the family of mappings $\{\Phi_\epsilon\}_{\epsilon>0} : E \times \mathrm{Ker}(D_0F) \to \mathbb{R}^N$ by

$$\Phi_\epsilon(z, t) := \frac{1}{\epsilon^2} F\left(\epsilon^2 z + \epsilon t\right) \qquad \forall (z, t) \in E \times \mathrm{Ker}(D_0F),\ \forall \epsilon > 0.$$

By Taylor's formula at second order for F at 0, we have

$$\Phi_\epsilon(z, t) = G(z, t) + o(1),$$

as ϵ tends to 0. Then there is $\bar{\epsilon} > 0$ (with $|(\bar{\epsilon}^2, \bar{\epsilon})| \leq 1/2$) such that for every $\epsilon \in (0, \bar{\epsilon})$,

$$|\Phi_\epsilon(z, t) - G(z, t)| \leq \frac{\mu}{2} \qquad \forall (z, t) \in \big(E \times \mathrm{Ker}(D_0F)\big) \cap B(0, 1).$$

By Lemma B.2.5 applied to $\tilde{G} = \Phi_\epsilon$, we infer that $\bar{B}(0, c)$ is contained in $\Phi_\epsilon\big(B(0, 1)\big)$, which in turn implies that for every $z \in \mathbb{R}^N$ such that $|z| = |z - F(\bar{u})| < c\epsilon^2$, there are w_1, w_2 in X such that

$$z = w_1 + w_2, \quad w_1 \in \mathrm{Ker}(D_{\bar{u}}F), \quad \|w_1\|_X < \epsilon, \quad \|w_2\|_X < \epsilon^2.$$

Let us now show that the above result holds uniformly for u close to $\bar{u} = 0$. Since F is C^1, the vector space $\mathrm{Ker}(D_uF)$ is transverse to E for u close to $\bar{u}$.

Moreover, again by C^1 regularity, for every $\delta > 0$, there is $\nu > 0$ such that for every $u \in B_X(\bar{u}, \nu)$,

$$\operatorname{Ker}(D_u F) \cap B(0,1) \subset \Big\{ y + z \in X \,|\, y \in \operatorname{Ker}(D_{\bar{u}} F) \cap B(0,1),\ \|z\|_X < \delta \Big\}.$$

Therefore, there is $\nu > 0$, such that for every $u \in B_X(\bar{u}, \nu)$, there is a vector space $W_u \subset X$ such that (W_u could be reduced to $\{0\}$)

$$X = E \oplus W_u \oplus \operatorname{Ker}\big(D_u F\big),$$

and there are linear mappings

$$\pi_1 : \operatorname{Ker}(D_0 F) \to W_u, \quad \pi_2 : \operatorname{Ker}(D_0 F) \to \operatorname{Ker}\big(D_u F\big)$$

such that for every $t \in \operatorname{Ker}(D_0 F)$, we have

$$t = \pi_1(t) + \pi_2(t), \quad \big|\pi_1(t)\big|_X \leq K|t|, \quad \big|\pi_1(t)\big|_X \leq K|t|,$$

for some constant $K > 0$ (which depends on $\operatorname{Ker}(D_0 F), E$, and $\|\cdot\|_X$). Given $u \in B_X(\bar{u}, \nu)$ and $\epsilon \in (0, \bar{\epsilon})$ we define $\tilde{G} : \big(E \times \operatorname{Ker}(D_0 F)\big) \cap B(0,1) \to \mathbb{R}^N$ by

$$\tilde{G}(z,t) := \frac{1}{\epsilon^2} \Big(F\big(u + \epsilon^2 z + \epsilon^2 \pi_1(t) + \epsilon \pi_2(t)\big) - F(u) \Big),$$

for every $(z,t) \in \big(E \times \operatorname{Ker}(D_0 F)\big) \cap B(0,1)$. Taking ν and $\bar{\epsilon} > 0$ smaller if necessary, by Taylor's formula for F at u at second order, by the above construction and by the fact that $D_u F$ and D_u^2 are respectively close to $D_0 F$ and $D_0^2 F$, we may assume that (B.13) is satisfied. We conclude easily.

Bibliography

[AMR83] Abraham, R., Marsden, J.E., Ratiu, T.: Manifolds, tensor analysis, and applications. Global Analysis Pure and Applied: Series B, 2. Addison-Wesley Publishing Co., Reading, Mass. (1983)

[Agr09] Agrachev, A.: Any sub-Riemannian metric has points of smoothness. Dokl. Akad. Nauk, **424**(3), 295–298 (2009), translation in Dokl. Math. **79**(1), 45–47 (2009)

[ABB12] Agrachev, A., Barilari, D., Boscain, U.: Introduction to Riemannian and sub-Riemannian geometry. To appear.

[AL09] Agrachev, A., Lee, P.: Optimal transportation under nonholonomic constraints. Trans. Amer. Math. Soc., **361**(11), 6019–6047 (2009)

[AS04] Agrachev, A.A., Sachkov, Yu.L.: Control Theory from the Geometric Viewpoint. Encyclopaedia of Mathematical Sciences, vol. 87, Springer-Verlag, Berlin (2004)

[AS96] Agrachev, A., Sarychev, A.: Abnormal sub-Riemannian geodesics: Morse index and rigidity. Ann. Inst. H. Poincaré, **13**, 635–690 (1996)

[AS99] Agrachev, A., Sarychev, A.: Sub-Riemannian metrics: minimality of singular geodesics versus subanalyticity. ESAIM Control Optim. Calc. Var., **4**, 377–403 (1999)

[AR04] Ambrosio, L., Rigot S.: Optimal transportation in the Heisenberg group. J. Funct. Anal. **208**(2), 261–301 (2004)

[Bel96] Bellaiche, A.: The tangent space in sub-Riemannian geometry. In Sub-Riemannian Geometry, Birkhäuser, 1–78 (1996)

[BB07] Bernard, P., Buffoni B.: Optimal mass transportation and Mather theory. J. Eur. Math. Soc., **9**(1), 85–121 (2007)

[Bil99] Billingsley, P.: Convergence of probability measures. Second ed. John Wiley & Sons Inc., New York (1999)

[Bis84] Bismut, J.-M.: Large deviations and the Malliavin calculus. Progress in Mathematics 45, Birkhäuser (1984)

[Bre93] Bredon, G.E.: Topology and Geometry. Graduate Texts in Mathematics, vol. 139. Springer-Verlag, New York (1993)

[Bre91] Brenier, Y.: Polar factorization and monotone rearrangement of vector-valued functions. Comm. Pure Appl. Appl. Math., **44**, 375–417 (1991)

[CR05] Cannarsa, P., Rifford, L.: Semiconcavity results for optimal control problems admitting no singular minimizing controls. Ann. Inst. H. Poincaré Anal. Non Linéaire, **25**(4), 773–802 (2008)

[CS04] Cannarsa, P., Sinestrari, C.: Semiconcave functions, Hamilton-Jacobi equations, and optimal control. Progress in Nonlinear Differential Equations and their Applications, 58. Birkhäuser, Boston (2004)

[CDP11] Champion, T., De Pascale, L.: The Monge problem in $\mathbb{R}^d$. Duke Math. J. **157**(3), 551–572 (2011)

[Cho39] Chow, C.-L.: Über systeme von linearen partiellen differentialgleichungen ester ordnung. Math. Ann., **117**, 98–105 (1939)

[Cla83] Clarke, F.H.: Optimization and Nonsmooth Analysis. Wiley-Interscience, New York (1983) Republished as vol. 5 of Classics in Applied Mathematics, SIAM (1990)

[DPR11] De Pascale, L., Rigot, S.: Monge's transport problem in the Heisenberg group. Adv. Calc. Var. **4**(2), 195–227 (2011)

[EG92] Evans, L.C., Gariepy, R.: Measure theory and fine properties of functions. CRC Press, Boca Raton, FL (1992)

[FR10] Figalli, A., Rifford, L.: Mass Transportation on sub-Riemannian Manifolds, Geom. Funct. Anal., **20**(1),124–159 (2010)

[GM96] Gangbo, W., McCann, R.J.: The geometry of optimal transportation. Acta Math., **177**(2), 113–161 (1996)

[GG73] Golubitsky, M., Guillemin, V.: Stable mappings and their singularities. Graduate Texts in Mathematics, vol. 14. Springer-Verlag, New York (1973)

[Gro96] Gromov, M.: Carnot-Carathéodory spaces seen from within. In Sub-Riemannian Geometry, Birkhäuser, 79–323 (1996)

[HS74] Hirsch, M.W., Smale, S.: Differential equations, dynamical systems, and linear algebra. Pure and Applied Mathematics, vol. 60. Academic Press, New York-London (1974)

[Hör67] Hörmander, L.: Hypoelliptic second order differential operators. Acta. Math., **119**: 147-171 (1967)

[Jea12] Jean, F.: Control of Nonholonomic Systems and Sub-Riemannian Geometry. Lectures given at the CIMPA School "Géométrie sous-riemannienne", Beirut, Lebanon (2012)

[Jur97] Jurdjevic, V.: Geometric Control Theory. Cambridge Studies in Advanced Mathematics, vol. 52, Cambridge University Press, Cambridge (1997)

[Ka42] Kantorovitch L.: On the translocation of masses. C.R. (Doklady) Acad. Sci. URSS, **37**:199-201 (1942)

[Lee03] Lee, J.M.: Introduction to Smooth Manifolds. Graduate Texts in Mathematics, vol. 218. Springer-Verlag, New York (2003)

[LS95] Liu, W., Sussmann, H.J.: Shortest paths for sub-Riemannian metrics on rank-2 distributions. Mem. Amer. Math. Soc., vol. 118, number 564 (1995)

[McC01] McCann R.J.: Polar factorization of maps on Riemannian manifolds. Geom. Funct. Anal., **11** (3): 589–608 (2001)

[Mit85] Mitchell, J.: On Carnot-Carathéodory metrics. J. Differential Geom., **21**(1):35-45 (1985)

[Mon81] Monge, G.: Mémoire sur la théorie des déblais et des remblais. Histoire de l'Académie Royale des Sciences de Paris: 666–704 (1781)

[Mon94] Montgomery, R.: Abnormal minimizers. SIAM J. Control Optim. **32**(6):1605–1620 (1994)

[Mon02] Montgomery, R.: A tour of subriemannian geometries, their geodesics and applications, volume 91 of Mathematical Surveys and Monographs. American Mathematical Society, Providence, RI (2002)

[PBGM] Pontryagin, L., Boltyanskii, V., Gamkrelidze, R., Mischenko, E.: The Mathematical Theory of Optimal Processes. Wiley-Interscience, New-York (1962)

[Pra07] Pratelli, A.: On the equality between Monge's infimum and Kantorovitch's minimum in optimal mass transportation. Ann. Inst. H. Poincaré Probab. Statist. **43**(1):1–13 (2007)

[Ras38] Rashevsky, P.K.: About connecting two points of a completely nonholonomic space by admissible curve. Uch. Zapiski Ped. Inst. Libknechta, **2**, 83-94 (1938)

[Rif04] Rifford, L.: A Morse-Sard theorem for the distance function on Riemannian manifolds. Manuscripta Math., **113**:251–265 (2004)

[Rif06] Rifford, L: À propos des sphères sous-riemanniennes. Bull. Belg. Math. Soc. Simon Stevin, **13**(3):521–526 (2006)

[RT05] Rifford, L., Trélat, E.: Morse-Sard type results in sub-riemannian geometry. Math. Ann., **332**(1), 145–159 (2005)

[RT07] Rifford, L., Trélat, E.: On the stabilization problem for nonholonomic distributions. J. Eur. Math. Soc. **11**(2):223-255 (2009)

[Rig05] Rigot, S.: Mass transportation in groups of type H. Commun. Contemp. Math. **7**(4), 509–537 (2005)

[Roc66] Rockafellar, R.T.: Characterization of the subdifferentials of convex functions. Pacific J. Math. **17**:497-510 (1966)

[Sus96] Sussmann, H.J.: A cornucopia of four-dimensional abnormal sub-Riemannian minimizers. In Sub-Riemannian Geometry, Birkhäuser, 341–364 (1996)

[Sus10] Sussmann, H.J.: Smooth distributions are globally finitely spanned. In Analysis and design of nonlinear control systems, Springer, Berlin, 3–8 (2008)

[Vil03] Villani, C.: Topics in Mass Transportation Graduate Studies in Mathematics Surveys, Vol. 58. American Mathematical Society, Providence, RI (2003)

[Vil08] Villani, C.: Optimal transport, Old and New. Springer, Berlin (2008)

[Vin00] Vinter, R.B.: Optimal Control. Birkhäuser, Boston (2000)