Mass Transportation on Sub-Riemannian Manifolds

A. Figalli* L. Rifford†

Abstract

We study the optimal transport problem in sub-Riemannian manifolds where the cost function is given by the square of the sub-Riemannian distance. Under appropriate assumptions, we generalize Brenier-McCann's Theorem proving existence and uniqueness of the optimal transport map. We show the absolute continuity property of Wassertein geodesics, and we address the regularity issue of the optimal map. In particular, we are able to show its approximate differentiability a.e. in the Heisenberg group (and under some weak assumptions on the measures the differentiability a.e.), which allows to write a weak form of the Monge-Ampère equation.

1 Introduction

The optimal transport problem can be stated as follows: given two probability measures μ and ν, defined on measurable spaces X and Y respectively, find a measurable map $T : X \to Y$ with

$$T_\sharp \mu = \nu \qquad \text{(i.e. } \nu(A) = \mu\left(T^{-1}(A)\right) \text{ for all } A \subset Y \text{ measurable),}$$

and in such a way that T minimizes the transportation cost. This last condition means

$$\int_X c(x, T(x))\, d\mu(x) = \min_{S_\sharp \mu = \nu} \left\{ \int_X c(x, S(x))\, d\mu(x) \right\},$$

where $c : X \times Y \to \mathbb{R}$ is some given cost function, and the minimum is taken over all measurable maps $S : X \to Y$ with $S_\sharp \mu = \nu$. When the transport condition $T_\sharp \mu = \nu$ is satisfied, we say that T is a *transport map*, and if T minimizes also the cost we call it an *optimal transport map*. Up to now the optimal transport problem has been intensively studied in a Euclidean or a Riemannian setting by many authors, and it turns out that the particular choice $c(x, y) = d^2(x, y)$ (here d denotes a Riemannian distance) is suitable for studying some partial differential equations (like the semi-geostrophic or porous medium equations), for studying functional inequalities (like Sobolev and Poincaré-type inequalities) and for applications in geometry (for example, in the study of lower bound on the Ricci curvature of the manifolds). We refer to the books [6, 37, 38] for an excellent presentation.

*Université de Nice-Sophia Antipolis, Labo. J.-A. Dieudonné, UMR 6621, Parc Valrose, 06108 Nice Cedex 02, France (`figalli@unice.fr`)

†Université de Nice-Sophia Antipolis, Labo. J.-A. Dieudonné, UMR 6621, Parc Valrose, 06108 Nice Cedex 02, France (`rifford@unice.fr`)

After the existence and uniqueness results of Brenier for the Euclidean case [11] and McCann for the Riemannian case [27], people tried to extend the theory in a sub-Riemannian setting. In [7] Ambrosio and Rigot studied the optimal transport problem in the Heisenberg group, and recently Agrachev and Lee were able to extend their result to more general situations such as sub-Riemannian structures corresponding to 2-generating distributions [2].

Two key properties of the optimal transport map result to be useful for many applications: the first one is the fact that the transport map is differentiable a.e. (this for example allows to write the Jacobian of the transport map a.e.), and the second one is that, if μ and ν are absolutely continuous with respect to the volume measure, so are all the measure belonging to the (unique) Wasserstein geodesic between them. Both these properties are true in the Euclidean case (see for example [6]) or on compact Riemannian manifolds (see [18, 9]). If the manifold is noncompact, the second property still remains true (see [20, Section 5]), while the first one holds in a weaker form. Indeed, although one cannot hope for its differentiability in the non-compact case, as it is done in [22, Section 3] the transport map can be shown to be approximately differentiable a.e., which turns out to be enough for extending many results from the compact to the non-compact case. Up to now, the only available results in these directions in a sub-Riemannian setting were proved in [23], where the authors show that the absolute continuity property along Wassertein geodesics holds in the Heisenberg group.

The aim of this paper is twofold: on the one hand, we prove new existence and uniqueness results for the optimal transport map on sub-Riemannian manifolds. In particular, we show that the structure of the optimal transport map is more or less the same as in the Riemannian case (see [27]). On the other hand, in a still large class of cases, we prove that the transport map is (approximately) differentiable almost everywhere, and that the absolute continuity property along Wasserstein geodesics holds. This settles several open problems raised in [7, Section 7]: first of all, regarding problem [7, Section 7 (a)], we are able to extend the results of Ambrosio and Rigot [7] and of Agrachev and Lee [2] to a large class of sub-Riemannian manifolds, not necessarily two-generating. Concerning question [7, Section 7 (b)], we can prove a regularity result on optimal transport maps, showing that under appropriate assumptions (including the Heisenberg group) they are approximately differentiable a.e. Moreover, under some weak assumptions on the measures, the transport map is shown to be truly differentiable a.e. (see Theorem 3.7 and Remark 3.8). This allows for the first time in this setting to apply the area formula, and to write a weak formulation of the Monge-Ampère equation (see Remark 3.9). Finally, Theorem 3.5 answers to problem [7, Section 7 (c)] not only in the Heisenberg group (which was already solved in [23]) but also in more general cases.

The structure of the paper is the following:

In Section 2, we introduce some concepts of sub-Riemannian geometry and optimal transport appearing in the statements of the results.

In Section 3, we present our results on the mass transportation problem in sub-Riemannian geometry: existence and uniqueness theorems on optimal transport maps (Theorems 3.2 and 3.3), absolute continuity property along Wassertein geodesics (Theorem 3.5) , and finally regularity of the optimal transport map and its consequences (Theorem 3.7 and Remarks 3.8, 3.9). For sake of simplicity, all the measures appearing in these results are assumed to have compact supports. In the last paragraph of Section 3, we discuss the possible extensions of our results to the non-compact case.

In Section 4, we give a list of sub-Riemannian structures for which our different results may be applied. These cases include fat distributions, two-generating distributions, generic distribution of rank ≥ 3, nonholonomic distributions on three-dimensional manifolds, medium-fat distributions, codimension-one nonholonomic distributions, and rank-two distributions in four-dimensional manifolds.

Since the proofs of the theorems require lots of tools and results from sub-Riemannian geometry, we provide in Section 5 a short course in sub-Riemannian geometry. First, for sake of completeness, we recall and give the proofs of basic facts in sub-Riemannian geometry, such as the characterization of singular horizontal paths, the description of sub-Riemannian minimizing geodesics, or the properties of the sub-Riemannian exponential mapping. Then, we present some results concerning the regularity of the sub-Riemannian distance function and its cut locus. These latter results are the key tools in the proofs of the our transport theorems; some of them were already known (Theorem 5.9 has almost been proved in this form in [12]) while others are completely new under our assumptions.

In Section 6, taking advantage of the regularity properties obtained in the previous section, we provide all the proofs of the results stated in Section 3.

Finally, in Appendix A, we recall some classical facts in nonsmooth analysis, while in Appendix B we prove auxiliary results needed in Section 4.

2 Preliminaries

2.1 Sub-Riemannian manifolds

A sub-Riemannian manifold is given by a triple (M, Δ, g) where M denotes a smooth connected manifold of dimension n, Δ is a smooth nonholonomic distribution of rank $m < n$ on M, and g is a Riemannian metric on M[1]. We recall that a smooth distribution of rank m on M is a rank m subbundle of TM. This means that, for every $x \in M$, there exist a neighborhood $\mathcal{V}_x$ of x in M, and a m-tuple $(f_1^x, \ldots, f_m^x)$ of smooth vector fields on $\mathcal{V}_x$, linearly independent on $\mathcal{V}_x$, such that

$$\Delta(z) = \operatorname{Span}\{f_1^x(z), \ldots, f_m^x(z)\} \qquad \forall z \in \mathcal{V}_x.$$

One says that the m-tuple of vector fields $(f_1^x, \ldots, f_m^x)$ represents locally the distribution Δ. The distribution Δ is said to be *nonholonomic* (also called totally nonholonomic *e.g.* in [3]) if, for every $x \in M$, there is a m-tuple $(f_1^x, \ldots, f_m^x)$ of smooth vector fields on $\mathcal{V}_x$ which represents locally the distribution and such that

$$\operatorname{Lie}\{f_1^x, \ldots, f_m^x\}(z) = T_zM \qquad \forall z \in \mathcal{V}_x,$$

[1] Note that in general, the definition of a sub-Riemannian structure only involves a Riemannian metric on the distribution. However, since in the sequel we need a global Riemannian distance on the ambient manifold and we need to use Hessians, we prefer to work with a metric defined globally on TM.

that is, such that the Lie algebra[2] spanned by $f_1^x, \ldots, f_m^x$, is equal to the whole tangent space T_zM at every point $z \in \mathcal{V}_x$. This Lie algebra property is often called *Hörmander's condition.*

A curve $\gamma : [0,1] \to M$ is called a *horizontal path* with respect to Δ if it belongs to $W^{1,2}([0,1], M)$ and satisfies

$$\dot\gamma(t) \in \Delta(\gamma(t)) \qquad \text{for a.e. } t \in [0,1].$$

According to the classical Chow-Rashevsky Theorem (see [8, 16, 29, 31, 32]), since the distribution is nonholonomic on M, any two points of M can be joined by a horizontal path. That is, for every $x, y \in M$, there is a horizontal path $\gamma : [0,1] \to M$ such that $\gamma(0) = x$ and $\gamma(1) = y$. For $x \in M$, let $\Omega_\Delta(x)$ denote the set of horizontal paths $\gamma : [0,1] \to M$ such that $\gamma(0) = x$. The set $\Omega_\Delta(x)$, endowed with the $W^{1,2}$-topology, inherits of a Hilbert manifold structure (see [29]). The end-point mapping from x is defined by

$$\begin{array}{rccc} \mathrm{E}_x : \Omega_\Delta(x) & \longrightarrow & M \\ \gamma & \longmapsto & \gamma(1) \end{array}$$

It is a smooth mapping. A path γ is said to be *singular* if it is horizontal and if it is a critical point for the end-point mapping E_x, that is if the differential of E_x at γ is singular (i.e. not onto). A horizontal path which is not singular is called *nonsingular* or *regular.* Note that the regularity or singularity property of a given horizontal path depends only on the distribution, not on the metric g.

The *length* of a path $\gamma \in \Omega_\Delta(x)$ is defined by

$$\mathrm{length}_g(\gamma) := \int_0^1 \sqrt{g_{\gamma(t)}(\dot\gamma(t), \dot\gamma(t))}dt. \tag{2.1}$$

The *sub-Riemannian distance* $d_{SR}(x, y)$ (also called Carnot-Carathéodory distance) between two points x, y of M is the infimum over the lengths of the horizontal paths joining x and y. According to the Chow-Rashevsky Theorem (see [8, 16, 29, 31, 32]), since the distribution is nonholonomic on M, the sub-Riemannian distance is finite and continuous[3] on $M \times M$. Moreover, if the manifold M is a complete metric space[4] for the sub-Riemannian distance d_{SR}, then, since M is connected, for every pair x, y of points of M there exists a horizontal path γ joining x to y such that

$$d_{SR}(x, y) = \mathrm{length}_g(\gamma).$$

[2]We recall that, for any family $\mathcal{F}$ of smooth vector fields on M, the Lie algebra of vector fields generated by $\mathcal{F}$, denoted by $\mathrm{Lie}(\mathcal{F})$, is the smallest vector space S satisfying

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

where $[X, Y]$ is the Lie bracket of X and Y.

[3]In fact, thanks to the so-called Mitchell's ball-box Theorem (see [29]), the sub-Riemannian can be shown to be locally Hölder continuous on $M \times M$.

[4]Note that, since the distribution Δ is nonholonomic on M, the topology defined by the sub-Riemannian distance d_{SR} coincides with the original topology of M (see [8, 29]). Moreover, it can be shown that if the Riemannian manifold (M, g) is complete then, for any nonholonomic distribution Δ on M, the sub-Riemannian manifold (M, Δ, g) equipped with its sub-Riemannian distance is complete.

Such a horizontal path is called a *sub-Riemannian minimizing geodesic* between x and y.

Assuming that (M, d_{SR}) is complete, denote by T^*M the cotangent bundle of M, by ω the canonical symplectic form on T^*M, and by $\pi : T^*M \to M$ the canonical projection. The *sub-Riemannian Hamiltonian* $H : T^*M \to \mathbb{R}$ which is canonically associated with the sub-Riemannian structure is defined as follows: for every $x \in M$, the restriction of H to the fiber T_x^*M is given by the nonnegative quadratic form

$$p \longmapsto \frac{1}{2} \max \left\{ \frac{p(v)^2}{g_x(v,v)} \mid v \in \Delta(x) \setminus \{0\} \right\}. \tag{2.2}$$

Let $\overrightarrow{H}$ denote the Hamiltonian vector field on T^*M associated to H, that is, $\iota_{\overrightarrow{H}}\omega = -dH$. A *normal extremal* is an integral curve of $\overrightarrow{H}$ defined on $[0,1]$, i.e., a curve $\psi(\cdot) : [0,1] \to T^*M$ satisfying

$$\dot{\psi}(t) = \overrightarrow{H}(\psi(t)), \qquad \forall t \in [0,1].$$

Note that the projection of a normal extremal is a horizontal path with respect to Δ. For every $x \in M$, the *exponential mapping* with respect to x is defined by

$$\begin{array}{rcl} \exp_x : T_x^*M & \longrightarrow & M \\ p & \longmapsto & \psi(1), \end{array}$$

where ψ is the normal extremal such that $\psi(0) = (x,p)$ in local coordinates. We stress that, unlike the Riemannian setting, the sub-Riemannian exponential mapping with respect to x is defined on the cotangent space at x.

Remark: from now on, all sub-Riemannian manifolds appearing in the paper are assumed to be complete with respect to the sub-Riemannian distance.

2.2 Preliminaries in optimal transport theory

As we already said in the introduction, we recall that, given a cost function $c : X \times Y \to \mathbb{R}$, we are looking for a transport map $T : X \to Y$ which minimizes the transportation cost $\int c(x, T(x))\, d\mu$. The constraint $T_\# \mu = \nu$ being highly non-linear, the optimal transport problem is quite difficult from the viewpoint of calculus of variation. The major advance on this problem was due to Kantorovich, who proposed in [24, 25] a notion of weak solution of the optimal transport problem. He suggested to look for *plans* instead of transport maps, that is probability measures γ in $X \times Y$ whose marginals are μ and ν, i.e.

$$(\pi_X)_\sharp \gamma = \mu \qquad \text{and} \qquad (\pi_Y)_\sharp \gamma = \nu,$$

where $\pi_X : X \times Y \to X$ and $\pi_Y : X \times Y \to Y$ are the canonical projections. Denoting by $\Pi(\mu,\nu)$ the set of plans, the new minimization problem becomes the following:

$$C(\mu,\nu) = \min_{\gamma \in \Pi(\mu,\nu)} \left\{ \int_{M \times M} c(x,y)\, d\gamma(x,y) \right\}. \tag{2.3}$$

If γ is a minimizer for the Kantorovich formulation, we say that it is an *optimal plan*. Due to the linearity of the constraint $\gamma \in \Pi(\mu, \nu)$, it is simple, using weak topologies, to prove existence of solutions to (2.3): this happens for instance whenever X and Y are Polish spaces, and c is lower semicontinuous and bounded from below (see for instance [37, 38]). The connection between the formulation of Kantorovich and that of Monge can be seen by noticing that any transport map T induces the plan defined by $(Id \times T)_\sharp \mu$ which is concentrated on the graph of T. Thus, the problem of showing existence of optimal transport maps can be reduced to prove that an optimal transport plan is concentrated on a graph. Moreover, if one can show that *any* optimal plan in concentrated on a graph, since $\frac{\gamma_1+\gamma_2}{2}$ is optimal if so are γ_1 and γ_2, uniqueness of the transport map easily follows.

Definition 2.1. A function $\phi : X \to \mathbb{R}$ is said *c-concave* if there exists a function $\phi^c : Y \to \mathbb{R} \cup \{-\infty\}$, with $\phi^c \not\equiv -\infty$, such that

$$\phi(x) = \inf_{y \in Y} \{c(x, y) - \phi^c(y)\}.$$

If ϕ is c-concave, we define the *c-superdifferential of ϕ at x* as

$$\partial^c \phi(x) := \{y \in Y \mid \phi(x) + \phi^c(y) = c(x, y)\}.$$

Moreover we define the *c-superdifferential of ϕ* as

$$\partial^c \phi := \{(x, y) \in X \times Y \mid y \in \partial^c \phi(x)\}.$$

As we already said in the introduction, we are interested in studying the optimal transport problem on $M \times M$ (M being a complete sub-Riemannian manifold) with the cost function given by $c(x, y) = d^2_{SR}(x, y)$.

Definition 2.2. Denote by $P_c(M)$ the set of compactly supported probability measures in M and by $P_2(M)$ the set of Borel probability measures on M with finite 2-order moment, that is the set of μ satisfying

$$\int_M d^2_{SR}(x, x_0)\, d\mu(x) < +\infty \qquad \text{for a certain } x_0 \in M.$$

Moreover, we denote by $P_c^{ac}(M)$ (resp. $P_2^{ac}(M)$) the subset of $P_c(M)$ (resp. $P_2(M)$) that consists of the probability measures on M which are absolutely continuous with respect to the volume measure.

Obviously $P_c(M) \subset P_2(M)$. Moreover we remark that, by the triangle inequality for d_{SR}, the definition of $P_2(M)$ does not depends on x_0. The space $P_2(M)$ can be endowed with the so-called *Wasserstein distance* W_2:

$$W_2^2(\mu, \nu) := \min_{\gamma \in \Pi(\mu,\nu)} \left\{ \int_{M \times M} d^2(x, y)\, d\gamma(x, y) \right\}.$$

(note that W_2^2 is nothing else than the infimum in the Kantorovich problem). As W_2 defines a finite metric on $P_2(M)$, one can speak about geodesic in the metric space

(P_2, W_2). This space turns out, indeed, to be a length space (see for example [6, 37, 38]).

From now on, $\text{supp}(\mu)$ and $\text{supp}(\nu)$ will denote the supports of μ and ν respectively, i.e. the smallest closed sets on which μ and ν are respectively concentrated.

The following result is well-known (see for instance [38, Chapter 5]):

Theorem 2.3. *Let us assume that $\mu, \nu \in P_2(M)$. Then there exists a c-concave function ϕ such that the following holds: a transport plan $\gamma \in \Pi(\mu, \nu)$ is optimal if and only if $\gamma(\partial^c \phi) = 1$ (that is, γ is concentrated on the c-superdifferential of ϕ). Moreover one can assume that the following holds:*

$$\phi(x) = \inf_{y \in \text{supp}(\nu)} \left\{ d_{SR}^2(x, y) - \phi^c(y) \right\} \qquad \forall x \in M,$$

$$\phi^c(y) = \inf_{x \in \text{supp}(\mu)} \left\{ d_{SR}^2(x, y) - \phi(x) \right\} \qquad \forall y \in M.$$

In addition, if $\mu, \nu \in P_c(M)$, then both infima are indeed minima (so that $\partial^c \phi(x) \cap \text{supp}(\nu) \neq \emptyset$ for μ-a.e. x), and the functions ϕ and ϕ^c are continuous.

By the above theorem we see that, in order to prove existence and uniqueness of optimal transport maps, it suffices to prove that there exist two Borel sets $Z_1, Z_2 \subset M$, with $\mu(Z_1) = \nu(Z_2) = 1$, such that and $\partial^c \phi$ is a graph inside $Z_1 \times Z_2$ (or equivalently that $\partial^c \phi(x) \cap Z_2$ is a singleton for all $x \in Z_1$).

3 Statement of the results

3.1 Sub-Riemannian versions of Brenier-McCann's Theorems

The main difficulty appearing in the sub-Riemannian setting (unlike the Riemannian situation) is that, in general, the squared distance function is not locally Lipschitz on the diagonal. This gives rise to difficulties which make the proofs more technical than in the Riemannian case (and some new ideas are also needed). In order to avoid technicalities which would obscure the main ideas of the proof, we will state our results under some simplifying assumptions on the measures, and in Paragraph 3.4 we will explain how to remove them.

Before stating our first existence and uniqueness result, we introduce a definition:

Definition 3.1. Given a c-concave function $\phi : M \to \mathbb{R}$, we call "moving" set $\mathcal{M}^\phi$ and "static" set $\mathcal{S}^\phi$ respectively the sets defined as follows:

$$\mathcal{M}^\phi := \{x \in M \mid x \notin \partial^c \phi(x)\}, \tag{3.1}$$

$$\mathcal{S}^\phi := M \setminus \mathcal{M}^\phi = \{x \in M \mid x \in \partial^c \phi(x)\}. \tag{3.2}$$

We will also denote by $\pi_1 : M \times M \to M$ and $\pi_2 : M \times M \to M$ the canonical projection on the first and on the second factor, respectively. In the sequel, D denotes the diagonal in $M \times M$, that is

$$D := \{(x, y) \in M \times M \mid x = y\}.$$

Furthermore, we refer the reader to Appendix A for the definition of a locally semiconcave function.

Theorem 3.2 (**Optimal transport map for absolutely continuous measures).** *Let $\mu \in P_c^{ac}(M)$, $\nu \in P_c(M)$. Assume that there exists an open set $\Omega \subset M \times M$ such that $\mathrm{supp}(\mu \times \nu) \subset \Omega$, and d_{SR}^2 is locally semiconcave (resp. locally Lipschitz) on $\Omega \setminus D$. Let ϕ be the c-concave function provided by Theorem 2.3. Then:*

(i) $\mathcal{M}^\phi$ is open, and ϕ is locally semiconcave (resp. locally Lipschitz) in a neighborhood of $\mathcal{M}^\phi \cap \mathrm{supp}(\mu)$. In particular ϕ is differentiable μ-a.e. in $\mathcal{M}^\phi$.

(ii) For μ-a.e. $x \in \mathcal{S}^\phi$, $\partial^c \phi(x) = \{x\}$.

In particular, there exists a unique optimal transport map defined μ-a.e. by[5]

$$T(x) := \begin{cases} \exp_x(-\frac{1}{2}\, d\phi(x)) & \text{if } x \in \mathcal{M}^\phi \cap \mathrm{supp}(\mu), \\ x & \text{if } x \in \mathcal{S}^\phi \cap \mathrm{supp}(\mu), \end{cases}$$

and for μ-a.e. x there exists a unique minimizing geodesic between x and $T(x)$.

As can be seen from the proof (given in Section 6), assertion (ii) in Theorem 3.2 always holds without any assumption on the sub-Riemannian distance. That is, for any optimal transport problem on a complete sub-Riemannian manifold between two measures $\mu \in P_c^{ac}(M)$ and $\nu \in P_c(M)$, we always have

$$T(x) = x \qquad \text{for } \mu\text{-a.e. } x \in \mathcal{S}^\phi,$$

where ϕ is the c-concave function provided by Theorem 2.3.

Theorem 3.2 above can be refined if the sub-Riemannian distance is assumed to be locally Lipschitz on the diagonal. In that way, we obtain the sub-Riemannian version of McCann's Theorem on Riemannian manifolds (see [27]), improving the result of Agrachev and Lee (see [2]).

Theorem 3.3 (**Optimal transport map for more general measures).** *Let $\mu, \nu \in P_c(M)$, and suppose that μ gives no measure to countably $(n-1)$-rectifiable sets. Assume that there exists an open set $\Omega \subset M \times M$ such that $\mathrm{supp}(\mu \times \nu) \subset \Omega$, and d_{SR}^2 is locally semiconcave on $\Omega \setminus D$. Suppose further that d_{SR}^2 is locally Lipschitz on Ω, and let ϕ be the c-concave function provided by Theorem 2.3. Then:*

(i) $\mathcal{M}^\phi$ is open, and ϕ is locally semiconcave in a neighborhood of $\mathcal{M}^\phi \cap \mathrm{supp}(\mu)$. In particular ϕ is differentiable μ-a.e. in $\mathcal{M}^\phi$.

(ii) For μ-a.e. $x \in \mathcal{S}^\phi$, $\partial^c \phi(x) = \{x\}$.

In particular, there exists a unique optimal transport map defined μ-a.e. by

$$T(x) := \begin{cases} \exp_x(-\frac{1}{2}\, d\phi(x)) & \text{if } x \in \mathcal{M}^\phi \cap \mathrm{supp}(\mu), \\ x & \text{if } x \in \mathcal{S}^\phi \cap \mathrm{supp}(\mu), \end{cases}$$

and for μ-a.e. x there exists a unique minimizing geodesic between x and $T(x)$.

[5] The factor $\frac{1}{2}$ appearing in front of $d\phi(x)$ is due to the fact that we are considering the cost function $d_{SR}^2(x, y)$ instead of the (equivalent) cost $\frac{1}{2} d_{SR}^2(x, y)$

The regularity properties of the sub-Riemannian distance functions required in the two results above are satisfied by many sub-Riemannian manifolds. In particular, Theorem 3.2 holds as soon as there are no singular sub-Riemannian minimizing geodesic between two distinct points in Ω. In Section 4, we provide a list of sub-Riemannian manifolds which satisfy the assumptions of our different results.

3.2 Wasserstein geodesics

By Theorem 3.2, it is not difficult to deduce the uniqueness of the Wasserstein geodesic between μ and ν. Moreover the structure of the transport map allows to prove, as in the Riemannian case, that all the measures inside the geodesic are absolutely continuous if μ is. This last property requires however that, if $(x, y) \in \Omega$, then all geodesics from x to y do not "exit from Ω":

Definition 3.4. Let $\Omega \subset M \times M$ be an open set. We say that Ω is *totally geodesically convex* if for every $(x, y) \in \Omega$ and every geodesic $\gamma : [0,1] \to M$ from x to y, one has

$$(x, \gamma(t)),\ (\gamma(t), y) \in \Omega \qquad \forall t \in [0,1].$$

Observe that, if $\Omega = U \times U$ with $U \subset M$, then the above definition reduces to say that U is totally geodesically convex in the classical sense.

Theorem 3.5 (Absolute continuity of Wasserstein geodesics). *Let $\mu \in P_c^{ac}(M)$, $\nu \in P_c(M)$. Assume that there exists an open set $\Omega \subset M \times M$ such that $\operatorname{supp}(\mu \times \nu) \subset \Omega$, and d_{SR}^2 is locally semiconcave on $\Omega \setminus D$. Let ϕ be the c-concave function provided by Theorem 2.3. Then there exists a unique Wasserstein geodesic $(\mu_t)_{t\in[0,1]}$ joining $\mu = \mu_0$ to $\nu = \mu_1$, which is given by $\mu_t := (T_t)_\# \mu$ for $t \in [0,1]$, with*

$$T_t(x) := \begin{cases} \exp_x(-\frac{t}{2}\, d\phi(x)) & \text{if } x \in \mathcal{M}^\phi \cap \operatorname{supp}(\mu), \\ x & \text{if } x \in \mathcal{S}^\phi \cap \operatorname{supp}(\mu). \end{cases}$$

Moreover, if Ω is totally geodesically convex, then $\mu_t \in P_c^{ac}(M)$ for all $t \in [0,1)$.

3.3 Regularity of the transport map and the Monge-Ampère equation

The structure of the transport map provided by Theorem 3.2 allows also to prove in certain cases the approximate differentiability of the optimal transport map, and a useful Jacobian identity. Let us first recall the notion of approximate differential:

Definition 3.6 (Approximate differential). We say that $f : M \to \mathbb{R}$ has an *approximate differential* at $x \in M$ if there exists a function $h : M \to \mathbb{R}$ differentiable at x such that the set $\{f = h\}$ has density 1 at x with respect to the volume measure. In this case, the approximate value of f at x is defined as $\tilde{f}(x) = h(x)$, and the approximate differential of f at x is defined as $\tilde{d}f(x) = dh(x)$.

It is not difficult to show that the above definitions make sense. In fact, $h(x)$ and $dh(x)$ do not depend on the choice of h, provided x is a density point of the set $\{f = h\}$.

To write the formula of the Jacobian of T, we will need to use the notion of Hessian. We recall that the Hessian of a function $f : M \to \mathbb{R}$ is defined as the covariant derivative of df: $\mathrm{Hess}\, f(x) = \nabla df(x) : T_xM \times T_xM \to M$. Observe that the notion of the Hessian depends on the Riemannian metric on TM. However, since the transport map depends only on d_{SR}, which in turn depends only on the restriction of metric to the distribution, a priori it may seem strange that the Jacobian of T is expressed in terms of Hessians. However, as we will see below, the Jacobian of T depends on the Hessian of the function $z \mapsto \phi(z) - d_{SR}^2(z, T(x))$ computed at $z = x$. But since $\phi(z) - d_{SR}^2(z, T(x))$ attains a maximum at x, x is a critical point for the above function, and so its Hessian at x is indeed independent on the choice of the metric.

The following result is the sub-Riemannian version of the properties of the transport map in the Riemannian case. It was proved on compact manifolds in [18], and extended to the noncompact case in [22]. The problem in our case is that the structure of the sub-Riemannian cut-locus is different from the Riemannian case (see for example Proposition 5.10), and so many complications arise when one tries to generalize the Riemannian argument to our setting. Trying to extend the differentiability of the transport map in great generality would need some new results on the sub-Riemannian cut-locus which go behind the scope of this paper (see the Open Problem in Paragraph 5.8). For this reason, we prefer to state the result under some simplifying assumptions, which however holds in the important case of the Heisenberg group (see [29]), or for example for the standard sub-Riemannian structure on the three-sphere (see [10]).

We refer the reader to Paragraph 5.8 for the definitions of the global cut-locus $\mathrm{Cut}_{SR}(M)$.

Theorem 3.7 (Approximate differentiability and jacobian identity). *Let $\mu \in P_c^{ac}(M)$, $\nu \in P_c(M)$. Assume that there exists a totally geodesically convex open set $\Omega \subset M \times M$ such that $\mathrm{supp}(\mu \times \nu) \subset \Omega$, d_{SR}^2 is locally semiconcave on $\Omega \setminus D$, and for every $(x, y) \in Cut_{SR}(M) \cap (\Omega \setminus D)$, there are at least two distinct sub-Riemannian minimizing geodesics joining x to y. Let ϕ be the c-concave function provided by Theorem 2.3. Then the optimal transport map is differentiable μ-a.e. inside $\mathcal{M}^\phi \cap \mathrm{supp}(\mu)$, and it is approximately differentiable at μ-a.e. x. Moreover*

$$Y(x) := d(\exp_x)_{-\frac{1}{2}d\phi(x)} \quad \textit{and} \quad H(x) := \frac{1}{2}\mathrm{Hess}\, d_{SR}^2(\cdot, T(x))|_{z=x}$$

exists for μ-a.e. $x \in \mathcal{M}^\phi \cap \mathrm{supp}(\mu)$, and the approximate differential of T is given by the formula

$$\tilde{d}T(x) = \begin{cases} Y(x)\big(H(x) - \frac{1}{2}\mathrm{Hess}\,\phi(x)\big) & \textit{if } x \in \mathcal{M}^\phi \cap \mathrm{supp}(\mu), \\ Id & \textit{if } x \in \mathcal{S}^\phi \cap \mathrm{supp}(\mu), \end{cases}$$

where $Id : T_xM \to T_xM$ denotes the identity map.

Finally, assuming both μ and ν absolutely continuous with respect to the volume measure, and denoting by f and g their respective density, the following Jacobian identity holds:

$$\big|\det\big(\tilde{d}T(x)\big)\big| = \frac{f(x)}{g(T(x))} \neq 0 \qquad \mu\textit{-a.e.} \tag{3.3}$$

In particular, $f(x) = g(x)$ for μ-a.e. $x \in \mathcal{S}^\phi \cap \mathrm{supp}(\mu)$.

Remark 3.8 (Differentiability a.e. of the transport map). If we assume that $f \neq g$ μ-a.e., then by the above theorem we deduce that $T(x) \neq x$ μ-a.e. (or equivalently $x \notin \partial^c\phi(x)$ μ-a.e.). Therefore the optimal transport is given by

$$T(x) = \exp_x(-\tfrac{1}{2}d\phi(x)) \qquad \mu\text{-a.e.},$$

and in particular T is differentiable (and not only approximate differentiable) μ-a.e.

Remark 3.9 (The Monge-Ampère equation). Since the function $z \mapsto \phi(z) - d_{SR}^2(z, T(x))$ attains a maximum at $T(x)$ for μ-a.e. x, it is not difficult to see that the matrix $H(x) - \frac{1}{2}\mathrm{Hess}\,\phi(x)$ (defined in Theorem 3.7) is nonnegative definite μ-a.e. This fact, together with (3.3), implies that the function ϕ satisfies the *Monge-Ampère* type equation

$$\det\big(H(x) - \tfrac{1}{2}\mathrm{Hess}\,\phi(x)\big) = \tfrac{f(x)}{|\det(Y(x))| g(T(x))} \qquad \text{for } \mu\text{-a.e. } x \in \mathcal{M}^\phi.$$

In particular, thanks to Remark 3.8,

$$\det\big(H(x) - \tfrac{1}{2}\mathrm{Hess}\,\phi(x)\big) = \tfrac{f(x)}{|\det(Y(x))| g(T(x))} \qquad \mu\text{-a.e.}$$

provided that $f \neq g$ μ-a.e.

3.4 The non-compact case

Let us briefly show how to remove the compactness assumption on μ and ν, and how to relax the hypothesis $\mathrm{supp}(\mu \times \nu) \subset \Omega$. We assume $\mu, \nu \in P_2(M)$ (so that Theorem 2.3 applies), and that $\mu \times \nu(\Omega) = 1$. Take an increasing sequence of compact set $K_l \subset \Omega$ such that $\cup_l K_l = \Omega$. We consider

$$\psi_l(x) := \inf\left\{d_{SR}^2(x, y) - \phi^c(y) \mid y \text{ s.t. } (x, y) \in K_l\right\}.$$

Since now ϕ^c is not a priori continuous (and so $\partial^c\psi_l$ is not necessarily closed), we first define

$$\phi_l^c(y) := \inf\left\{d_{SR}^2(x, y) - \psi_l(x) \mid x \text{ s.t. } (x, y) \in K_l\right\},$$

and then consider

$$\phi_l(x) := \inf\left\{d_{SR}^2(x, y) - \phi_l^c(y) \mid y \text{ s.t. } (x, y) \in K_l\right\}.$$

In this way the following properties holds (see for example the argument in the proof of [38, Proposition 5.8]):

- ϕ_l and ϕ_l^c are both continuous;
- $\psi_l(x) \geq \phi(x)$ for all $x \in M$;
- $\phi^c(y) \leq \phi_l^c(y)$ for all $y \in \pi_2(K_l)$;
- $\phi_l(x) = \psi_l(x)$ for all $x \in \pi_1(K_l)$.

This implies that $\partial^c\phi \cap K_l \subset \partial^c\phi_l$, and so

$$\partial^c\phi \cap \Omega \subset \cup_l \partial^c\phi_l.$$

One can therefore prove (i) and (ii) in Theorem 3.2 with ϕ_l in place of ϕ, and from this and the hypothesis $\mu \times \nu(\Omega) = 1$ it is not difficult to deduce that $(x, \partial^c\phi(x)) \cap \Omega$ is a singleton for μ-a.e. x (see the argument in the proof of Theorem 3.2). This proves existence and uniqueness of the optimal transport map.

Although in this case we cannot hope for any semiconcavity result for ϕ (since, as in the non-compact Riemannian case, ϕ is just a Borel function), the above argument shows that the graph of the optimal transport map is contained in the union of $\partial^c\phi_l$. Thus, as in [20, Section 5], one can use $\partial^c\phi_l$ to construct the (unique) Wasserstein geodesic between μ and ν, and in this way obtain the absolutely continuity of all measures belonging to the geodesic follows as in the compactly supported case.

Finally, the fact that the graph of the optimal transport map is contained in $\cup_l \partial^c\phi_l$ allows also to prove the approximate differentiability of the transport map and the Jacobian identity, provided that one replaces the hessian of ϕ with the approximate hessian (see [22, Section 3] for see how this argument works in the Riemannian case).

4 Examples

The aim of the present section is to provide a list of examples where some of our theorems apply. For each kind of sub-Riemannian manifold that we present, we provide a regularity result for the associated squared sub-Riemannian distance function. We let the reader to see in each case what theorem holds under that regularity property. Before giving examples, we recall that if Δ is a smooth distribution on M, we call section of Δ any smooth vector field X satisfying $X(x) \in \Delta(x)$ for any $x \in M$. For any smooth vector field Z on M and every $x \in M$, we shall denote by $[Z, \Delta](x)$, $[\Delta, \Delta](x)$, and $[Z, [\Delta, \Delta]]$ the subspaces of T_xM given by

$$[Z, \Delta](x) := \{[Z, X](x) \mid X \text{ section of } \Delta\},$$

$$[\Delta, \Delta](x) := \mathrm{Span}\{[X, Y](x) \mid X, Y \text{ sections of } \Delta\},$$

and

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

4.1 Fat distributions

The distribution Δ is called *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

$$T_xM = \Delta(x) + [X, \Delta](x).$$

The condition above being very restrictive, there are very few fat distributions (see [29]). Fat distributions on three-dimensional manifolds are the rank-two distributions Δ satisfying

$$T_xM = \mathrm{Span}\{f_1(x), f_2(x), [f_1, f_2](x)\} \qquad \forall x \in M,$$

where (f_1, f_2) is a 2-tuple of vector fields representing locally the distribution Δ. A classical example of fat distribution in $\mathbb{R}^3$ is given by the distribution spanned by the vector fields

$$X_1 = \frac{\partial}{\partial x_1}, \qquad X_2 = \frac{\partial}{\partial x_2} + x_1 \frac{\partial}{\partial x_3}.$$

This is the distribution appearing in the Heisenberg group (see [7, 8, 23]). It can be shown that, if Δ is a fat distribution, then any nontrivial (i.e. not constant) horizontal path with respect to Δ is nonsingular (see [12, 29, 32]). As a consequence, Theorems 5.9 and 5.16 yield the following result.

Proposition 4.1. *If Δ is fat on M, then the squared sub-Riemannian distance function is locally Lipschitz on $M \times M$ and locally semiconcave on $M \times M \setminus D$.*

4.2 Two-generating distributions

A distribution Δ is called *two-generating* if

$$T_x M = \Delta(x) + [\Delta, \Delta](x) \qquad \forall x \in M.$$

Any fat distribution is two-generating. Moreover, if the ambient manifold M has dimension three, then any two-generating distribution is fat. The distribution Δ in $\mathbb{R}^4$ which is spanned by the vector fields

$$X_1 = \frac{\partial}{\partial x_1}, \qquad X_2 = \frac{\partial}{\partial x_2}, \qquad X_3 = \frac{\partial}{\partial x_3} + x_1 \frac{\partial}{\partial x_4},$$

provides an example of distribution which is two-generating but not fat. It is easy to see that, if the distribution is two-generating, then there are no Goh paths (see Paragraph 5.9 for the definition of Goh path). As a consequence, by Theorem 5.16, we have:

Proposition 4.2. *If Δ is two-generating on M, then the squared sub-Riemannian distance function is locally Lipschitz on $M \times M$.*

The above result and its consequences in optimal transport are due to Agrachev and Lee (see [2]).

4.3 Generic sub-Riemannian structures

Let (M, g) be a complete Riemannian manifold of dimension ≥ 4, and $m \geq 3$ be a positive integer. Denote by $\mathcal{D}_m$ the space of rank m distributions on M endowed with the Whitney C^∞ topology. Chitour, Jean and Trélat proved that there exists an open dense subset $\mathcal{O}_m$ of $\mathcal{D}_m$ such that every element of $\mathcal{O}_m$ does not admit nontrivial minimizing singular paths (see [14, 15]). As a consequence, we have

Proposition 4.3. *Let (M, g) be a complete Riemannian manifold of dimension ≥ 4. Then for any generic distribution of rank ≥ 3, the squared sub-Riemannian distance function is locally semiconcave on $M \times M \setminus D$.*

This result implies in particular that, for generic sub-Riemannian manifolds, we have existence and uniqueness of optimal transport maps, and absolute continuity of Wasserstein geodesics.

4.4 Nonholonomic distributions on three-dimensional manifolds

Assume that M has dimension 3 and that Δ is a nonholonomic rank-two distribution on M, and define

$$\Sigma_\Delta := \left\{ x \in M \mid \Delta(x) + [\Delta, \Delta](x) \neq \mathbb{R}^3 \right\}.$$

The set Σ_Δ is called the *singular set* or the *Martinet set* of Δ. As an example, take the nonholonomic distribution Δ in $\mathbb{R}^3$ which is spanned by the vector fields

$$f_1 = \frac{\partial}{\partial x_1}, \qquad f_2 = \frac{\partial}{\partial x_2} + x_1^2 \frac{\partial}{\partial x_3}.$$

It is easy to show that the singular set of Δ is the plane $\{x_1 = 0\}$. This distribution is often called the *Martinet distribution*, and Σ_Δ the *Martinet surface*. The singular horizontal paths of Δ correspond to the horizontal paths which are included in Σ_Δ. This means that necessarily any singular horizontal path is, up to reparameterization, a restriction of an arc of the form $t \mapsto (0, t, \bar{x}_3) \in \mathbb{R}^3$ with $\bar{x}_3 \in \mathbb{R}$. This kind of result holds for any rank-two distribution in dimension three (we postpone its proof to Appendix B):

Proposition 4.4. *Let Δ be a 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 nontrivial horizontal path $\gamma : [0,1] \to M$ is singular if and only if it is included in Σ_Δ.*

Proposition 4.4 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. As a consequence, thanks to Theorems 5.9 and 5.16, the following result holds:

Proposition 4.5. *Let Δ be a nonholonomic distribution on a three-dimensional manifold. The squared sub-Riemannian distance function is locally Lipschitz on $M \times M \setminus (\Sigma_\Delta \times \Sigma_\Delta)$ and locally semiconcave on $M \times M \setminus (D \cup \Sigma_\Delta \times \Sigma_\Delta)$.*

We observe that, since Σ_Δ is countably 2-rectifiable, for any pair of measures $\mu, \nu \in P_c(M)$ such that μ gives no measure to countably 2-rectifiable sets, the conclusions of Theorem 3.3 hold.

4.5 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

$$T_x M = \Delta(x) + [\Delta, \Delta](x) + [X, [\Delta, \Delta]](x).$$

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 $\mathbb{R}^4$ which is spanned by the vector vector fields

$$f_1 = \frac{\partial}{\partial x_1}, \qquad f_2 = \frac{\partial}{\partial x_2}, \qquad f_3 = \frac{\partial}{\partial x_3} + (x_1 + x_2 + x_3)^2 \frac{\partial}{\partial x_4}.$$

Medium-fat distribution were introduced by Agrachev and Sarychev in [4] (we refer the interested reader to that paper for a detailed study of this kind of distributions). It can easily be shown that medium-fat distributions do not admit nontrivial Goh paths. As a consequence, Theorem 5.16 yields:

Proposition 4.6. *Assume that Δ is medium-fat. Then the squared sub-Riemannian distance function is locally Lipschitz on $M \times M \setminus D$.*

Let us moreover observe that, given a medium-fat distribution, it can be shown that, for a generic smooth complete Riemannian metric on M, the distribution does not admit nontrivial singular sub-Riemannian minimizing geodesics (see [14, 15]). As a consequence, we have:

Proposition 4.7. *Let Δ be a medium-fat distribution on M. Then, for "generic" Riemannian metrics, the squared sub-Riemannian distance function is locally semiconcave on $M \times M \setminus D$.*

Notice that, since two-generating distributions are medium-fat, the latter result holds for two-generating distributions.

4.6 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

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

The following result holds (we postpone its proof to Appendix B):

Proposition 4.8. *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.*

From Theorem 5.16, we have:

Proposition 4.9. *The squared sub-Riemannian distance function is locally Lipschitz on $M \times M \setminus (\Sigma_\Delta \times \Sigma_\Delta)$.*

Note that, as for medium-fat distributions, for generic metrics the function d_{SR}^2 is locally semiconcave on $M \times M \setminus (D \cup \Sigma_\Delta \times \Sigma_\Delta)$.

4.7 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

$$T_x M = \mathrm{Span}\{f_1(x), f_2(x), [f_1, f_2](x), [f_1, [f_1, f_2]](x), [f_2, [f_1, f_2]](x)\}$$

for any local parametrization of the distribution. In [36], Sussmann shows that there is a smooth horizontal vector field X on M such that the singular horizontal curves

γ parametrized by arc-length are exactly the integral curves of X, i.e. the curves satisfying

$$\dot\gamma(t) = X(\gamma(t)).$$

By the way, it can also be shown that those curves are locally minimizing between their end-points (see [26, 36]). For every $x \in M$, denote by $\mathcal{O}(x)$ the orbit of x by the flow of X and set

$$\Omega := \{(x,y) \in M \times M \mid y \notin \mathcal{O}(x)\}.$$

Sussmann's Theorem, together with Theorem 5.9, yields the following result:

Proposition 4.10. *Under the assumption above, the function d_{SR}^2 is locally semiconcave in the interior of Ω.*

As an example, consider the distribution Δ in $\mathbb{R}^4$ spanned by the two vector fields

$$f_1 = \frac{\partial}{\partial x_1}, \qquad f_2 = \frac{\partial}{\partial x_2} + x_1 \frac{\partial}{\partial x_3} + x_3 \frac{\partial}{\partial x_4}.$$

It is easy to show that a horizontal path $\gamma : [0,1] \to \mathbb{R}^4$ is singular if and only if it satisfies, up to reparameterization by arc-length,

$$\dot\gamma(t) = f_1(\gamma(t)), \qquad \forall t \in [0,1].$$

By Proposition 4.10, we deduce that, for any complete metric g on $\mathbb{R}^4$, the function d_{SR}^2 is locally semiconcave on the set

$$\Omega = \left\{(x,y) \in \mathbb{R}^4 \times \mathbb{R}^4 \mid (y-x) \notin \mathrm{Span}\{e_1\}\right\},$$

where e_1 denotes the first vector in the canonical basis of $\mathbb{R}^4$. Consequently, for any pair of measures $\mu \in P_c^{ac}(M)$, $\nu \in P_c(M)$ satisfying $\mathrm{supp}(\mu \times \nu) \subset \Omega$, Theorem 3.2 applies (or more in general, if $\mu \times \nu(\Omega) = 1$, we can apply the argument in Paragraph 3.4).

5 A short course in sub-Riemannian geometry

Throughout this section, (M, Δ, g) denotes a sub-Riemannian manifold of rank $m < n$ which is assumed to be complete with respect to the sub-Riemannian distance. As in the Riemannian case, the Hopf-Rinow Theorem holds. In particular, any two points in M can be joined by a minimizing geodesics, and any sub-Riemannian ball of finite radius is a compact subset of M. We refer the reader to [29, Appendix D] for the proofs of those results.

5.1 Nonholonomic distributions vs. nonholonomic control systems

Any nonholonomic distribution can be parametrized locally by a nonholonomic control system, that is by a smooth dynamical system with parameters called controls. Indeed, assume that $\mathcal{V}$ is an open subset of M such that there are m smooth vector fields

$f_1, \ldots, f_m$ on $\mathcal{V}$ which parametrize the nonholonomic distribution Δ on $\mathcal{V}$, that is which satisfy

$$\Delta(x) = \mathrm{Span}\left\{f_1(x), \ldots, f_m(x)\right\} \qquad \forall x \in \mathcal{V},$$

and

$$\mathrm{Lie}\left\{f_1, \ldots, f_m\right\}(x) = T_x M \qquad \forall x \in \mathcal{V}.$$

Given $x \in \mathcal{V}$, there is a correspondence between the set of horizontal paths in $\Omega_\Delta(x)$ which remain in $\mathcal{V}$ and the set of admissible controls of the control system

$$\dot{x} = \sum_{i=1}^{m} u_i f_i(x).$$

A *control* $u \in L^2([0,1], \mathbb{R}^m)$ is called *admissible* with respect to x and $\mathcal{V}$ if the solution $\gamma_{x,u}$ to the Cauchy problem

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

is well-defined on $[0,1]$ and remains in $\mathcal{V}$. The set $\mathcal{U}_x$ of admissible controls is an open subset of $L^2([0,1], \mathbb{R}^m)$.

Proposition 5.1. *Given $x \in M$, the mapping*

$$\begin{array}{ccc} \mathcal{U}_x & \longrightarrow & \Omega_\Delta(x) \\ u & \longmapsto & \gamma_{x,u} \end{array}$$

is one-to-one.

Given $x \in M$, the end-point-mapping from x, from the control viewpoint, takes the following form

$$\begin{array}{rcl} \mathrm{E}_x : \mathcal{U}_x & \longrightarrow & M \\ u & \longmapsto & \gamma_{x,u}(1) \end{array}$$

This mapping is smooth. The derivative of the end-point mapping from x at $u \in \mathcal{U}_x$, that we shall denote by $d\mathrm{E}_x(u)$, is given by

$$d\mathrm{E}_x(u)(v) = d\Phi^u(1,x) \int_0^1 \left(d\Phi^u(t,\bar{x})\right)^{-1} \left(\sum_{i=1}^{m} v_i(t) f_i(\gamma_{x,u}(t))\right) dt \quad \forall v \in L^2([0,1], \mathbb{R}^m),$$

where $\Phi^u(t,x)$ denotes the flow of the time-dependent vector field X^u defined by

$$X^u(t,x) := \sum_{i=1}^{m} u_i(t) f_i(x) \qquad \text{for a.e. } t \in [0,1], \quad \forall x \in \mathcal{V},$$

(note that the flow is well-defined in a neighborhood of x). We say that an admissible control u is *singular* with respect to x if $d\mathrm{E}_x$ is singular at u. Observe that this is equivalent to say that its associated horizontal path is singular (see the definition of singular path given in Section 2). It is important to notice that the singularity of a given horizontal path does not depend on the metric but only on the distribution.

5.2 Characterization of singular horizontal paths

Denote by ω the canonical symplectic form on T^*M and by $\Delta^\perp$ the annihilator of Δ in T^*M minus its zero section. Define $\overline{\omega}$ as the restriction of ω to $\Delta^\perp$. An absolutely continuous curve

$$\psi : [0,1] \to \Delta^\perp \tag{5.1}$$

$$\text{such that} \qquad \dot{\psi}(t) \in \ker \overline{\omega}(\psi(t)) \qquad \text{for a.e. } t \in [0,1] \tag{5.2}$$

is called an *abnormal extremal* of Δ.

Proposition 5.2. *A horizontal path $\gamma : [0,1] \to M$ is singular if and only if it is the projection of an abnormal extremal ψ of Δ. The curve ψ is said to be an abnormal extremal lift of γ.*

If the distribution is parametrized by a family of m smooth vector fields $f_1, \ldots, f_m$ on some open set $\mathcal{V} \subset M$, and if in addition the cotangent bundle T^*M is trivializable over $\mathcal{V}$, then the singular controls, or equivalently the singular horizontal paths which are contained in $\mathcal{V}$, can be characterized as follows. Define the pseudo-Hamiltonian $H_0 : \mathcal{V} \times (\mathbb{R}^n)^* \times (\mathbb{R}^m) \longmapsto \mathbb{R}$ by

$$H_0(x,p,u) = \sum_{i=1}^m u_i p(f_i(x)). \tag{5.3}$$

Proposition 5.3. *Let $x \in \mathcal{V}$ and u be an admissible control with respect to x and $\mathcal{V}$. Then, the control u is singular (with respect to x) if and only if there is an arc $p : [0,1] \longrightarrow (\mathbb{R}^n)^* \setminus \{0\}$ in $W^{1,2}$ such that the pair $(x = \gamma_{x,u}, p)$ satisfies*

$$\left\{ \begin{array}{rcl} \dot{x}(t) & = & \frac{\partial H_0}{\partial p}(x(t),p(t),u(t)) = \sum_{i=1}^m u_i(t) f_i(x(t)) \\ \dot{p}(t) & = & -\frac{\partial H_0}{\partial x}(x(t),p(t),u(t)) = -\sum_{i=1}^m u_i(t) p(t) \cdot df_i(x(t)) \end{array} \right. \tag{5.4}$$

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

$$p(t) \cdot f_i(x(t)) = 0 \qquad \forall t \in [0,1], \quad \forall i = 1, \ldots, m. \tag{5.5}$$

Note that properties (5.4) and (5.5) are nothing more than the conditions (5.2) and (5.1) written in local coordinates (with $\psi(t) = (\gamma(t) = \gamma_{x,u}(t), p(t))$). As a consequence, by a gluing process [6] along the horizontal path, Proposition 5.2 can be seen as a corollary of Proposition 5.3.

[6] For every $t \in (0,1]$, denote by $\mathrm{E}_x^t : \Omega_\Delta(x) \to M$ the end point-mapping from x at time t. If we parametrized γ in a neighborhood of $\gamma(1)$ (so that the cotangent bundle T^*M is trivializable in such neighborhood) and rewrite the proof of Proposition 5.3 in that neighborhood, then we can construct for some $t \in (0,1)$ a covector $p : [t,1] \to (\mathbb{R}^n)^* \setminus \{0\}$ satisfying (5.4)-(5.5) and such that $p(t) \cdot d\mathrm{E}_x^t(\gamma)(v) = 0$ for every $v \in T_\gamma \Omega_\Delta(x)$. Repeating the contruction of p in a neighborhood of $\gamma(t)$ and using the compactness of the set $\gamma([0,1])$, we obtain an abnormal extremal lift of γ on the whole interval $[0,1]$.

Proof. Doing a change of coordinates if necessary, we can assume that $\mathcal{V}$ is an open subset of $\mathbb{R}^n$. In that case, the differential of E_x at u is given by

$$d\mathrm{E}_x(u)(v) = S(1)\int_0^1 S(t)^{-1}B(t)v(t)dt \qquad \forall v \in L^2([0,1],\mathbb{R}^m), \tag{5.6}$$

where for every $t \in [0,1]$, $B(t)$ is the $n \times m$ matrix given by

$$B(t) := (f_1(x_u(t)), \dots, f_m(x_u(t)))$$

and $S(\cdot)$ is the solution of the Cauchy problem

$$\dot{S}(t) = A(t)S(t) \qquad \text{for a.e. } t \in [0,1], \quad S(0) = I_n, \tag{5.7}$$

with

$$A(t) := \sum_{i=1}^m u_i(t) df_i(x(t)) \qquad \text{for a.e. } t \in [0,1].$$

If $d\mathrm{E}_x(u)$ is not surjective, then there exists $p \in (\mathbb{R}^n)^* \setminus \{0\}$ such that

$$p \cdot d\mathrm{E}_x(u)(v) = 0 \qquad \forall v \in L^2([0,1],\mathbb{R}^m).$$

Owing to (5.6), the above identity can be written as

$$\int_0^1 p \cdot S(1)S(t)^{-1}B(t)v(t)dt = 0$$

for any $v \in L^2([0,1],\mathbb{R}^m)$. By the arbitrariness of v and the continuity of $S(t)$ and $B(t)$, we deduce that $p \cdot S(1)S(t)^{-1}B(t) = 0$ for any $t \in [0,1]$. Let us now define, for each $t \in [0,1]$,

$$p(t) := p \cdot S(1)S(t)^{-1}.$$

By construction, $p : [0,1] \to \mathbb{R}^n$ is a $W^{1,2}$ arc for which (5.5) is satisfied. Since $p \neq 0$ and $S(t)$ is invertible for all $t \in [0,1]$, $p(t)$ does not vanish on $[0,1]$. By (5.7) we conclude that the pair (x,p) satisfies also (5.4).
Conversely, let us assume that there exists an arc $p : [0,1] \to (\mathbb{R}^n)^* \setminus \{0\}$ in $W^{1,2}$ which satisfies (5.4) and (5.5). This implies that

$$\dot{p}(t) = -p(t) \cdot A(t) \qquad \text{for a.e. } t \in [0,1],$$

and

$$p(t) \cdot B(t) = 0 \qquad \forall t \in [0,1].$$

Setting $p := p(1) \neq 0$, we have, for any $t \in [0,1]$,

$$p(t) = p \cdot S(1)^{-1}S(t).$$

Hence, we obtain

$$p \cdot S(1)S(t)^{-1}B(t) = 0,$$

which gives

$$p \cdot d\mathrm{E}_x(u)(v) = 0 \qquad \forall v \in L^2([0,1],\mathbb{R}^m).$$

This concludes the proof. □

A control or a horizontal path which is singular is sometimes called *abnormal.* If it is not singular, we call it *nonsingular* or *regular.*

5.3 Sub-Riemannian minimizing geodesics

As we said in Section 2, since the metric space (M, d_{SR}) is assumed to be complete, for every pair $x, y \in M$ there is a horizontal path γ joining x to y such that

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

If γ is parametrized by arc-length, then using Cauchy-Schwarz inequality it is easy to show that γ minimizes the quantity

$$\int_0^1 g_{\gamma(t)}(\dot\gamma(t), \dot\gamma(t))dt =: \text{energy}_g(\gamma),$$

over the horizontal paths joining x to y. This infimum, denoted by $e_{SR}(x,y)$, is called the *sub-Riemannian energy* between x and y. Since M is assumed to be complete, the infimum is always attained, and the horizontal paths which minimize the sub-Riemannian energy are those which minimize the sub-Riemannian distance and which are parametrized by arc-length. In particular, one has

$$e_{SR}(x,y) = d_{SR}^2(x,y) \qquad \forall x, y \in M.$$

Assume from now that γ is a given horizontal path minimizing $\text{energy}_g(\gamma)$ between x and y. Such a path is called a *sub-Riemannian minimizing geodesic*. Since γ minimizes also the distance, it has no self intersection. Hence we can parametrize the distribution along γ: there is an open neighborhood $\mathcal{V}$ of $\gamma([0,1])$ in M and an orthonormal family (with respect to the metric g) of m smooth vector fields $f_1, \ldots, f_m$ such that

$$\Delta(z) = \text{Span}\{f_1(z), \ldots, f_m(z)\} \qquad \forall z \in \mathcal{V}.$$

Moreover, since γ belongs to $W^{1,2}([0,1], M)$, there is a control $u^\gamma \in L^2([0,1], \mathbb{R}^m)$ (in fact, $|u^\gamma(t)|^2$ is constant), which is admissible with respect to x and $\mathcal{V}$, such that

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

By the discussion above, we know that u^γ minimizes the quantity

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

among all controls $u \in L^2([0,1], \mathbb{R}^m)$ which are admissible with respect to x and $\mathcal{V}$, and which satisfy the constraint

$$\text{E}_x(u) = y.$$

By the Lagrange Multiplier Theorem, there is $\lambda \in (\mathbb{R}^n)^*$ and $\lambda_0 \in \{0,1\}$ such that

$$\lambda \cdot d\text{E}_x(u^\gamma) - \lambda_0 dC(u^\gamma) = 0. \tag{5.8}$$

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 $(x,p) \in$

$\mathbb{R}^n \times (\mathbb{R}^n)^*$ over $\mathcal{V}$.

First case: $\lambda_0 = 0$. The linear operator $d\mathrm{E}_x(u^\gamma) : L^2([0,1], \mathbb{R}^m) \to T_yM$ cannot be onto, which means that the control u is necessarily singular. Hence there is an arc $p : [0,1] \longrightarrow (\mathbb{R}^n)^* \setminus \{0\}$ in $W^{1,2}$ satisfying (5.4) and (5.5). In other terms, $\gamma = \gamma_{x,u^\gamma}$ admits an abnormal extremal lift in T^*M. We also says that γ is an *abnormal minimizing geodesic.*

Second case: $\lambda_0 = 1$. In local coordinates, the Hamiltonian H (defined in (2.2)) takes the following form:

$$H(x,p) = \frac{1}{2}\sum_{i=1}^m \big(p \cdot f_i(x)\big)^2 = \max_{u \in \mathbb{R}^m} \left\{ \sum_{i=1}^m u_i p \cdot f_i(x) - \frac{1}{2}\sum_{i=1}^m u_i^2 \right\} \tag{5.9}$$

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

Proposition 5.4. *Equality (5.8) with $\lambda_0 = 1$ yields the existence of an arc $p : [0,1] \longrightarrow (\mathbb{R}^n)^*$ in $W^{1,2}$, with $p(1) = \frac{\lambda}{2}$, such that the pair $(\gamma = \gamma_{x,u^\gamma}, p)$ satisfies*

$$\begin{cases} \dot\gamma(t) &= \frac{\partial H}{\partial p}(\gamma(t),p(t)) = \sum_{i=1}^m \left[p(t)\cdot f_i(\gamma(t))\right] f_i(\gamma(t)) \\ \dot p(t) &= -\frac{\partial H}{\partial x}(\gamma(t),p(t)) = -\sum_{i=1}^m \left[p(t)\cdot f_i(\gamma(t))\right] p(t)\cdot df_i(\gamma(t)) \end{cases} \tag{5.10}$$

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

$$u_i^\gamma(t) = p(t)\cdot f_i(\gamma(t)) \qquad \textit{for a.e. } t \in [0,1], \quad \forall i = 1,\dots,m. \tag{5.11}$$

In particular, the path γ is smooth on $[0,1]$. The curve γ and the control u^γ are called normal.

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

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

Moreover, the differential of E_x at u^γ is given by

$$d\mathrm{E}_x(u^\gamma)(v) = S(1)\int_0^1 S(t)^{-1}B(t)v(t)dt \qquad \forall v \in L^2([0,1],\mathbb{R}^m), \tag{5.12}$$

where the functions A, B, S were defined in the proof of Proposition 5.3. Hence (5.8) yields

$$\int_0^1 \left[\lambda \cdot S(1)S(t)^{-1}B(t) - 2u^\gamma(t)^*\right] v(t)dt = 0 \qquad \forall v \in L^2([0,1],\mathbb{R}^m).$$

which implies

$$u^\gamma(t) = \frac{1}{2}\left(\lambda\cdot S(1)S(t)^{-1}B(t)\right)^* \qquad \text{for a.e. } t \in [0,1].$$

Let us define $p : [0,1] \to (\mathbb{R}^n)^*$ by

$$p(t) := \frac{1}{2}\lambda \cdot S(1)S(t)^{-1} \qquad \forall t \in [0,1].$$

By construction, for a.e. $t \in [0,1]$ we have $u^\gamma(t)^* = p(t)\cdot B(t)$, which means that (5.11) is satisfied. Furthermore, as in the proof of Proposition 5.3, we have $\dot p(t) = -p(t)\cdot A(t)$ for a.e. $t \in [0,1]$. This means that (5.10) is satisfied for a.e. t. □

The curve $\psi : [0,1] \rightarrow T^*M$ 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 $\psi(1) = (y, \frac{\lambda}{2})$. We say that ψ is a *normal extremal lift* of γ. We also say that γ is a *normal minimizing geodesic.*

To summarize, we proved that the minimizing geodesic (or equivalently the minimizing control u^γ) is either abnormal or normal. Note that it could be both normal and abnormal. 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 found the first counterexample to this assertion (see [28, 29]).

5.4 The sub-Riemannian exponential mapping

Let $x \in M$ be fixed. The *sub-Riemannian exponential mapping* from x is defined by

$$\begin{array}{rcl} \exp_x : T_x^*M & \longrightarrow & M \\ p & \longmapsto & \psi(1), \end{array}$$

where ψ is the normal extremal so that $\psi(0) = (x,p)$ in local coordinates. Note that $H(\psi(t))$ is constant along a normal extremal ψ, hence we have

$$\text{energy}_g(\pi(\psi)) = \left(\text{length}_g(\pi(\psi))\right)^2 = 2H(\psi(0)).$$

The exponential mapping is not necessarily onto. However, since (M, d_{SR}) is complete, the following result holds (see [33]).

Proposition 5.5. *For every $x \in M$, the set $\exp_x(T_x^*M)$ is dense in M.*

The above result is a straightforward consequence of Proposition A.1 together with the following:

Proposition 5.6. *Let $y \in M$ be such that there is a function $\phi : M \rightarrow I\!R$ differentiable at y such that*

$$\phi(y) = d_{SR}^2(x,y) \quad \textit{and} \quad d_{SR}^2(x,z) \geq \phi(z) \quad \forall z \in M.$$

Then there is a unique minimizing geodesic between x and y, and it is the projection of normal extremal $\psi : [0,1] \rightarrow T^*M$ satisfying $\psi(1) = (y, \frac{1}{2}d\phi(y))$. In particular $x = \exp_y(-\frac{1}{2}d\phi(y))$.

Proof. 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

$$e_{SR}(x,z) \geq \phi(z) \quad \forall z \in \mathcal{U} \qquad \text{and} \qquad e_{SR}(x,y) = \phi(y). \tag{5.13}$$

Since (M, d_{SR}) is complete, there exists a minimizing geodesic $\gamma : [0,1] \rightarrow M$ between x and y. As before, we can parametrize the distribution Δ by a orthonormal family of smooth vector fields $f_1, \ldots, f_m$ 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 and $\mathcal{V}$ and which satisfy the constraint $\mathrm{E}_x(u) = y$. Let $u \in L^2([0,1], \mathbb{R}^m)$ be a control admissible with respect to x and $\mathcal{V}$ such that $\mathrm{E}_x(u) \in \mathcal{U}$. By (5.13) one has

$$C(u) \geq e_{SR}(x, \mathrm{E}_x(u)) \geq \phi(\mathrm{E}_x(u)).$$

Moreover

$$C(u^\gamma) = e_{SR}(x, y) = \phi(y) = \phi(\mathrm{E}_x(u^\gamma)).$$

Hence we deduce that the control u^γ minimizes the functional $D : L^2([0,1], \mathbb{R}^m) \to \mathbb{R}$ defined as

$$D(u) := C(u) - \phi(\mathrm{E}_x(u))$$

over the set of controls $u \in L^2([0,1], \mathbb{R}^m)$ such that $\mathrm{E}_x(u) \in \mathcal{U}$. This means that u^γ is a critical point of D. Setting $\lambda = d\phi(y)$, we obtain

$$\lambda \cdot d\mathrm{E}_x(u^\gamma) - dC(u^\gamma) = 0.$$

By Proposition 5.4, the path γ admits a normal extremal lift $\psi : [0,1] \to T^*M$ satisfying $\psi(1) = (y, \frac{1}{2} d\phi(y))$. By the Cauchy-Lipschitz Theorem, such a normal extremal is unique. □

5.5 The horizontal eikonal equation

As in the Riemannian case, the sub-Riemannian distance function from a given point satisfies a Hamilton-Jacobi equation. For the definition of viscosity solution, see Paragraph A.1.

Proposition 5.7. *For every $x \in M$, the function $f(\cdot) = d_{SR}(x, \cdot)$ is a viscosity solution of the Hamilton-Jacobi equation*

$$H(y, df(y)) = \frac{1}{2} \qquad \forall y \in M \setminus \{x\}. \tag{5.14}$$

Proof. Recall that the sub-Riemannian distance is continuous on $M \times M$. Hence, the function f is continuous on M. Let us first prove that f is a viscosity subsolution of (5.14) on $M \setminus \{x\}$. Let $\phi : M \to \mathbb{R}$ be a C^1 function satisfying $\phi \geq f$ and such that $\phi(y) = f(y)$ for some $y \in M \setminus \{x\}$. Let $\gamma : [0, L] \to M$ be a piecewise C^1 horizontal path joining x to y such that $g_{\gamma(t)}(\dot\gamma(t), \dot\gamma(t)) = 1$ for a.e. $t \in [0, L]$. Since d_{SR} satisfies the triangle inequality, we have for every $t \in [0, L]$,

$$f(y) \leq f(\gamma(t)) + t \leq \phi(\gamma(t)) + t.$$

Hence letting t tend to L, we obtain

$$d\phi(y) \cdot \dot\gamma(L) = \lim_{t \uparrow L} \frac{\phi(\gamma(L)) - \phi(\gamma(t))}{t} \leq 1.$$

But for any $v \in \Delta(y)$ with $g_y(v, v) = 1$ there is a piecewise C^1 horizontal path with unit speed joining x to y such that $\dot\gamma(L) = v$. Hence we deduce that

$$H(y, d\phi(y)) \leq \frac{1}{2}.$$

To prove now that f is a viscosity supersolution, let $\psi : M \to \mathbb{R}$ be a C^1 function satisfying $\psi \leq f$ and such that $\psi(y) = f(y)$ for some $y \in M \setminus \{x\}$. Since (M, d_{SR}) is assumed to be complete, there is a minimizing geodesic $\gamma : [0,1] \to M$ between x and y. Up to reparameterize γ, we can assume that γ is defined on the interval $[0, L = \text{length}_g(\gamma)]$ and has unit speed. For every $t \in (0, L)$, the horizontal curve γ minimizes the length between the points x and $\gamma(t)$. Hence we have

$$\psi(\gamma(t)) \leq f(\gamma(t)) = f(y) - t = \psi(y) - t.$$

Hence letting t tend to L, we obtain

$$d\psi(y) \cdot \dot{\gamma}(L) = \lim_{t \uparrow L} \frac{\psi(\gamma(L)) - \psi(\gamma(t))}{t} \geq 1.$$

The conclusion follows. □

5.6 Compactness of minimizing geodesics

The compactness of minimizing curves is crucial to prove regularity properties of the sub-Riemannian distance. Let us denote by $W^{1,2}_{\Delta}([0,1], M)$ the set of horizontal paths $\gamma : [0,1] \to M$ endowed with the $W^{1,2}$-topology. For every $\gamma \in W^{1,2}_{\Delta}([0,1], M)$, the energy of γ with respect to g, denoted by $\text{energy}_g(\gamma)$. is well-defined. The classical compactness result taken from Agrachev [1] reads as follows:

Proposition 5.8. *For every compact $K \subset M$, the set*

$$\mathcal{K} := \left\{ \gamma \in W^{1,2}_{\Delta}([0,1], M) \mid \exists\, x, y \in K \text{ with } e_{SR}(x,y) = \text{energy}_g(\gamma) \right\}$$

is a compact subset of $W^{1,2}([0,1], M)$.

Proof. First we note that, since (M, d_{SR}) is complete, the set $\mathcal{K}$ is a bounded subset of $W^{1,2}([0,1], M)$. Thus we can find a sequence $\{\gamma_k\} \subset \mathcal{K}$ that weakly converges to $\gamma \in W^{1,2}([0,1], M)$, and such that the sequence $\{\|\gamma_k\|_{W^{1,2}}\}$ converges to some $N \geq 0$. By the lower semicontinuity of the norm under weak convergence, we immediately deduce that.

$$\|\gamma\|_{W^{1,2}} \leq N. \tag{5.15}$$

Moreover it is not difficult to see that the constraint to be a horizontal path is closed under weak convergence, and so γ is horizontal with respect to Δ. Hence, by continuity of the energy on $M \times M$, we obtain

$$N^2 = \lim_{k \to \infty} \|\gamma_k\|^2_{W^{1,2}} = \lim_{k \to \infty} e_{SR}(\gamma_k(0), \gamma_k(1)) = e_{SR}(\gamma(0), \gamma(1)) \leq \|\gamma\|^2_{W^{1,2}}.$$

Combining this with (5.15), we deduce that $\|\gamma\|_{W^{1,2}} = N$. This implies that the sequence $\{\gamma_k\}$ converges to γ in the strong topology of $W^{1,2}$. □

5.7 Local semiconcavity of the sub-Riemannian distance

As we said in Section 2, the sub-Riemannian distance can be shown to be locally Hölder continuous on $M \times M$. But in general, it has no reason to be more regular. Within the next sections, we are going to show that, under appropriate assumptions on the sub-Riemannian structure, d_{SR} enjoyes more regularity properties such as local semiconcavity or locally Lipschitz regularity.

Recall that D denotes the diagonal of $M \times M$, that is, the set of all pairs of the form (x, x) with $x \in M$. Thanks to Proposition 5.8, the following result holds:

Theorem 5.9. *Let Ω be an open subset of $M \times M$ such that for every pair $(x, y) \in \Omega$ with $x \neq y$, any minimizing geodesic between x and y is nonsingular. Then, the distance function d_{SR} (or equivalently d_{SR}^2) is locally semiconcave on $\Omega \setminus D$.*

Proof. For sake of simplicity, we are just going to sketch the proof, and we refer the reader to [12, 32] for more details. Let us fix $(x, y) \in \Omega \setminus D$ and show that d_{SR} is semiconcave in a neighborhood of (x, y) in $M \times M \setminus D$. Let $\mathcal{U}_x$ and $\mathcal{U}_y$ be two compact neighborhoods of x and y such that $\mathcal{U}_x \times \mathcal{U}_y \subset \Omega \setminus D$. Denote by $\mathcal{K}$ the set of minimizing horizontal paths γ in $W^{1,2}_{\Delta}([0,1], \mathbb{R}^m)$ such that $\gamma(0) \in \mathcal{U}_x$ and $\gamma(1) \in \mathcal{U}_y$. Thanks to Proposition 5.8, $\mathcal{K}$ is a compact subset of $W^{1,2}([0,1], M)$. Let $(x', y') \in \mathcal{U}_x \times \mathcal{U}_y$ be fixed. Since (M, d_{SR}) is assumed to be complete, there exists a sub-Riemannian minimizing geodesic $\gamma_{x',y'}$ between x' and y'. Moreover by assumption, it is nonsingular. As before, we can parametrize Δ b y a family of smooth orthonormal vector fields along $\gamma_{x',y'}$, and we denote by $u^{x',y'}$ the control in $L^2([0,1], \mathbb{R}^m)$ corresponding to $\gamma_{x',y'}$. Since $u^{x',y'}$ is nonsingular, there are n linearly independent controls $v_1^{x',y'}, \ldots v_n^{x',y'}$ in $L^2([0,1], \mathbb{R}^m)$ such that the linear operator

$$\begin{array}{rcl} \mathcal{E}^{x',y'} : \mathbb{R}^n & \longrightarrow & \mathbb{R}^n \\ \alpha & \longmapsto & \sum_{i=1}^m \alpha_i dE_{x'}\left(u^{x',y'}\right)\left(v_i^{x',y'}\right) \end{array}$$

is invertible. Set

$$\begin{array}{rcl} \mathcal{F}^{x',y'} : \mathbb{R}^n \times \mathbb{R}^n & \longrightarrow & \mathbb{R}^n \times \mathbb{R}^n \\ (z, \alpha) & \longmapsto & \left(z, E_z\left(u^{x',y'} + \sum_{i=1}^m \alpha_i v_i^{x',y'}\right)\right) \end{array}$$

This mapping is well-defined and smooth in a neighborhood of $(x', 0)$, satisfies

$$\mathcal{F}^{x',y'}(x', 0) = (x', y'),$$

and its differential at $(x', 0)$ is invertible. Hence by the Inverse Function Theorem, there are an open ball $\mathcal{B}^{x',y'}$ centered at (x', y') in $\mathbb{R}^n \times \mathbb{R}^n$ and a function $\mathcal{G}^{x',y'} : \mathcal{B}^{x',y'} \to \mathbb{R}^n \times \mathbb{R}^n$ such that

$$\mathcal{F}^{x',y'} \circ \mathcal{G}^{x',y'}(z, w) = (z, w) \qquad \forall (z, w) \in \mathcal{B}^{x',y'}.$$

Denote by $\left(\alpha^{x',y'}\right)^{-1}$ the second component of $\mathcal{G}_{x',y'}$. From the definition of the sub-Riemannian energy between two points, we infer that for any $(z, w) \in \mathcal{B}^{x',y'}$ we have

$$e_{SR}(z, w) \leq \left\| u^{x',y'} + \sum_{i=1}^m \left(\left(\alpha^{x',y'}\right)^{-1}(z, w)\right)_i v_i^{x',y'} \right\|_{L^2}^2 .$$

Set

$$\phi^{x',y'}(z,w) := \left\| u^{x',y'} + \sum_{i=1}^{m} \left(\left(\alpha^{x',y'} \right)^{-1} (z,w) \right)_i \right\|_{L^2} \qquad \forall (z,w) \in \mathcal{B}^{x',y'}.$$

We conclude that, for every $(x',y') \in \mathcal{U}_x \times \mathcal{U}_y$, there is a smooth function $\phi^{x',y'}$ such that $d_{SR}(z,w) \leq \phi^{x',y'}(z,w)$ for any (z,w) in $\mathcal{B}^{x',y'}$. By compactness of $\mathcal{K}$ and thanks to a quantitative version of the Inverse Function Theorem, the $C^{1,1}$ norms of the functions $\phi^{x',y'}$ are uniformly bounded and the radii of the balls $\mathcal{B}^{x',y'}$ are uniformly bounded from below by a positive constant for x', y' in $\mathcal{U}_x \times \mathcal{U}_y$. The result follows from Lemma A.2. □

5.8 Sub-Riemannian cut locus

For every $x \in M$, the *singular set* of $d_{SR}(x,\cdot)$, denoted by $\Sigma\left(d_{SR}(x,\cdot)\right)$, is defined as the set of points $y \neq x \in M$ where $d_{SR}(x,\cdot)$ (or equivalently d_{SR}^2) is not continuously differentiable. The *cut-locus* of x is defined as

$$\mathrm{Cut}_{SR}(x) := \overline{\Sigma\left(d_{SR}(x,\cdot)\right)}$$

and the *global cut-locus* of M as

$$\mathrm{Cut}_{SR}(M) := \left\{(x,y) \in M \mid y \in \mathrm{Cut}_{SR}(x)\right\}.$$

The next result highlights a major difference between Riemannian and sub-Riemannian geometries (we refer the reader to [1] for its proof):

Proposition 5.10. *For every $x \in M$, $x \in Cut_{SR}(x)$, or equivalently $D \subset Cut_{SR}(M)$.*

A covector $p \in T_x^*M$ is said to be *conjugate* with respect to $x \in M$, if the mapping $\exp_x$ is singular at p, that is if $d\exp_x(p)$ is singular. For every $x \in M$, we denote by $\mathrm{Conj}_{min}(x)$ the set of points $y \in M \setminus \{x\}$ for which there is $p \in T_x^*M$ which is conjugate with respect to x and such that

$$\exp_x(p) = y \qquad \text{and} \qquad e_{SR}(x,y) = 2H(x,p).$$

The following result holds:

Proposition 5.11. *Let Ω be an open subset of $M \times M$. Assume that Ω is totally geodesically convex and that the sub-Riemannian distance is locally semiconcave on $\Omega \setminus D$. Then, for every $x \in M$, we have*

$$\left(\{x\} \times Cut_{SR}(x)\right) \cap \Omega = \left(\{x\} \times \left(\Sigma\left(d_{SR}(x,\cdot)\right) \cup Conj_{min}(x) \cup \{x\}\right)\right) \cap \Omega.$$

Moreover, the set $(\{x\} \times Cut_{SR}(x)) \cap \Omega$ has Hausdorff dimension $\leq n-1$, and the function d_{SR} is of class C^∞ on the open set $\Omega \setminus Cut_{SR}(M)$.

Proof. We already know that x belongs to $\mathrm{Cut}_{SR}(x)$. We are first going to prove that the inclusion

$$\left(\{x\}\times\left(\mathrm{Cut}_{SR}(x)\setminus\{x\}\right)\right)\cap\Omega\subset\left(\{x\}\times\left(\Sigma\left(d_{SR}(x,\cdot)\right)\cup\mathrm{Conj}_{min}(x)\right)\right)\cap\Omega$$

holds. In fact, we are going to show that

$$\Big(\{x\}\times\Big(\overline{\Sigma\left(d_{SR}(x,\cdot)\right)}\setminus\left(\Sigma\left(d_{SR}(x,\cdot)\right)\cup\{x\}\right)\Big)\Big)\cap\Omega\subset\left(\{x\}\times\mathrm{Conj}_{min}(x)\right)\cap\Omega,$$

which gives the result. Set $f(\cdot):=d_{SR}(x,\cdot)$. We need the following lemma (see Paragraph A.2 for the definition of $\partial_L f$).

Lemma 5.12. *For every $y\in M\setminus\{x\}$ and every $\zeta\in\partial_L f(y)$, there exists a normal extremal $\psi(\cdot):[0,1]\to T^*M$ whose projection $\gamma=\pi(\psi)$ is minimizing between x and y, and such that $\psi(1)=(x,f(y)\zeta)$ in local coordinates.*

Proof of Lemma 5.12. By definition of the limiting subdifferential, there exists a sequence $\{y_k\}$ of points in M converging to y and a sequence $\{\zeta_k\}\in D^-f(y_k)$ such that $\lim\zeta_k=\zeta$. For each integer k, denote by $\gamma_k:[0,1]\to M$ a minimizing geodesic joining x to y_k. From Proposition 5.6, for each k , the horizontal path γ_k admits a normal extremal lift $\psi_k:[0,1]\to T^*M$ satisfying $\psi_k(1)=(y_k,f(y_k)\zeta_k)$. Since the sequence $\{\psi_k(1)\}$ is bounded, up to a subsequence the sequence $\{\psi_k\}$ converges uniformly towards a normal extremal $\psi:[0,1]\to T^*M$. The projection of ψ given by $\gamma=\pi(\psi)$ is a sub-Riemannian minimizing geodesic between x and y as the limit in $W^{1,2}([0,1],M)$ of a sequence of minimizing horizontal paths between x and y_k. □

Let us return to the proof of Proposition 5.11 and show that

$$\Big(\{x\}\times\Big(\overline{\Sigma(f)}\setminus\left(\Sigma(f)\cup\{x\}\right)\Big)\Big)\cap\Omega\subset\left(\{x\}\times\mathrm{Conj}_{min}(x)\right)\cap\Omega.$$

Let us fix $y\in M$ such that $(x,y)\in\Big(\{x\}\times\Big(\overline{\Sigma(f)}\setminus\left(\Sigma(f)\cup\{x\}\right)\Big)\Big)\cap\Omega$. Since y does not belong to the singular set of f, and f is locally semiconcave in a neighborhood of (x,y), there is $\zeta\in T_y^*M$ such that $\partial_L f(y)=D^-f(y)=\{\zeta\}$. By Proposition 5.6, there is a normal extremal $\psi:[0,1]\to T^*M$ whose projection is minimizing between x and y and such that $\psi(1)=(y,f(y)\zeta)$. On the other hand, since y belongs to the closure of $\Sigma(f)$, there is a sequence $\{y_k\}$ of points in $\Sigma(f)$ which converges to y. For every integer k, the limiting subdifferential $\partial_L f(y_k)$ admits at least two elements $\zeta_k^1\neq\zeta_k^2$. Hence, by the lemma above, for each k there are two normal extremals $\psi_k^1,\psi_k^2:[0,1]\to T^*M$ whose projections $\gamma_k^1=\pi(\psi_k^1),\gamma_k^2=\pi(\psi_k^2)$ are minimizing between x and y_k, and such that $\psi_k^1(1)=(y_k,f(y_k)\zeta_k^1)$ and $\psi_k^2(1)=(y_k,f(y_k)\zeta_k^2)$. Since $\partial_L f(y)=\{\zeta\}$, the sequences $\{\psi_k^1(1)\}$ and $\{\psi_k^2(1)\}$ both converge to $(y,f(y)\zeta)=\psi(1)$. Hence, the two sequences $\{\psi_k^1\},\{\psi_k^2\}$ converge uniformly to ψ. This proves that for k large, $\psi_k^1(0)=(x,p_k^1)$ and $\psi_k^2(0)=(x,p_k^2)$ are close to $\psi(0)=(x,p)$ in T_x^*M and satisfy $\exp_x(p_k^1)=\exp_x(p_k^2)$. This shows that the exponential mapping from x cannot be injective in a neighborhood of p (which satisfies $\exp_x(p)=y$ and $e_{SR}(x,y)=2H(x,p)$). This concludes the proof of the inclusion.

Let us now show the other inclusion. We need the following result.

Lemma 5.13. *The function f is of class $C^{1,1}_{loc}$ on the open set*

$$\mathcal{O}_x := \{y \in M \mid (x,y) \in \Omega\} \setminus \overline{\Sigma(f)}.$$

Proof of Lemma 5.13. Let ω_x be the open subset of M defined by

$$\omega_x := \{y \in M \mid (x,y) \in \Omega\}.$$

Since by assumption we already know that f is locally semiconcave on $\omega_x \setminus \{x\}$, we are going to show that $-f$ is locally semiconcave on $\omega_x \setminus \overline{\Sigma(f)}$. Then Proposition A.6 will yield the result. Let $y \in \omega_x \setminus \overline{\Sigma(f)}$ be fixed. Since $\omega_x \setminus \overline{\Sigma(f)}$ is open, there is an open neighborhood $\mathcal{V} \subset \omega_x \setminus \overline{\Sigma(f)}$ of y where f is C^1. For every $z \in \mathcal{V}$, denote by $v(z)$ the unique tangent vector in T_zM satisfying

$$g_z(v(z), v(z)) = 1 \qquad \text{and} \qquad H(z, df(z)) = \frac{1}{2}\left(df(z)\cdot v(z)\right)^2 = \frac{1}{2}$$

(it exists thanks to Proposition 5.7). Since f is C^1 on $\mathcal{V}$, the function $z \mapsto v(z)$ is a continuous vector field on $\mathcal{V}$. For every $y' \in \mathcal{V}$, denote by $z_{y'}$ a solution to the Cauchy problem $\dot{z}(t) = v(z(t)), z(0) = y'$ (by the Cauchy-Peano Theorem, there exists such a solution defined on the interval $[0,\varepsilon]$ for a certain $\varepsilon > 0$). By construction, we have for every $y' \in \mathcal{V}$, $f(z_{y'}(t)) = f(y') + t$. Moreover, by the triangle inequality, we also have

$$f(z) \geq f(z_{y'}(\varepsilon)) - d_{SR}(z, z_{y'}(\varepsilon)) \qquad \forall z \in M$$

and

$$f(y') = f(z_{y'}(\varepsilon)) - \varepsilon = f(z_{y'}(\varepsilon)) - d_{SR}(y', z_{y'}(\varepsilon)).$$

Thus, for every $y' \in \mathcal{V}$, the function $z \mapsto f(z_{y'}(\varepsilon)) - d_{SR}(z, z_{y'}(\varepsilon))$ touches f at y' from below. Furthermore, by assumption (Ω is assumed to be totally geodesically convex), we know that the function $z \mapsto d_{SR}(z, z_{y'}(\varepsilon))$ is locally semiconcave in a neighborhood of y'. By Lemma A.2 applied to $-f$ the result follows easily. $\square$

We now return to the proof of Proposition 5.11 and we show that

$$\left(\{x\} \times \mathrm{Conj}_{min}(x)\right) \subset \left(\{x\} \times \overline{\Sigma(f)}\right) \cap \Omega.$$

Let $y \in \mathrm{Conj}_{min}(x)$ be such that $(x,y) \in \Omega$. We argue by contradiction. If y does not belong to $\overline{\Sigma(f)}$, then f is $C^{1,1}_{loc}$ in a neighborhood $\mathcal{V}$ of y. Define

$$\begin{array}{rccc} \Psi : & \mathcal{V} & \longrightarrow & T^*_{\bar{x}}M \\ & y & \longmapsto & \psi(0), \end{array}$$

where $\psi : [0,1] \to TM$ is the normal extremal satisfying $\psi(1) = (y, f(y)df(y))$. This mapping is locally Lipschitz on $\mathcal{V}$. Moreover by construction, Ψ is an inverse of the exponential mapping. This proves that $p_0 := \Psi(x)$ is not conjugate with respect to x. We obtain a contradiction.

The second part of Proposition 5.11 follows from the fact that the set $\Sigma(f) \cap \omega_x$ is of Hausdorff dimension lower than or equal to $n-1$ (see Theorem A.8), and the fact that the set $\mathrm{Conj}_{min}(x)$ is contained in the set

$$\mathrm{Conj}(x) := \{y \in M \mid \exists\, p \in T^*_xM \text{ conjugate w.r.t. } x \text{ s.t. } \exp_x(p) = y\},$$

which has Hausdorff dimension lower than or equal to $n-1$ thanks to Sard's Theorem (see [21, Theorem 3.4.3]).

It remains to prove that d_{SR} is of class C^∞ on the open set $\hat{\Omega} := \Omega \setminus \mathrm{Cut}_{SR}(M)$. Let us first show that $\hat{\Omega}$ is open or equivalently that $\mathrm{Cut}_{SR}(M) \cap \Omega$ is a closed subset of Ω. Let $(x, y) \in \Omega$ be such that there is a sequence $\{(x_k, y_k)\} \in \mathrm{Cut}_{SR}(M) \cap \Omega$ converging to (x, y) as k tends to infinity. If $x = y$, then we know that $(x, y) \in \mathrm{Cut}_{SR}(M)$, so we may assume that $x \neq y$. Moreover, by Proposition 5.11 together with a diagonal process, we may as well assume that for each k, there are two elements $\zeta_k^1 \neq \zeta_k^2 \in T_{y_k}^* M$ such that

$$\zeta_1^k, \zeta_2^k \in \partial_L f_k(y_k),$$

where f_k is defined as $f_k(z) := d_{SR}(x_k, z)$ for any $z \in M$. By Lemma 5.12 for each k there are two normal extremals $\psi_k^1, \psi_k^2 : [0,1] \to T^*M$ whose projections $\gamma_k^1 = \pi(\psi_k^1), \gamma_k^2 = \pi(\psi_k^2)$ are minimizing between x_k and y_k and such that $\psi_k^1(1) = (y_k, f_k(y_k)\zeta_k^1)$ and $\psi_k^2(1) = (y_k, f_k(y_k)\zeta_k^2)$. Without loss of generality, we can assume that both sequences $\{\psi_k^1\}, \{\psi_k^2\}$ converge respectively to ψ^1 and ψ^2. Two cases appear. As before, we denote by f the function $d_{SR}(x, \cdot)$.
First case: $\psi^1(1) \neq \psi^2(1)$. Then, this means that the limiting subdifferential of f at y contains two elements. This proves that $(x, y) \in \mathrm{Cut}_{SR}(M)$.
Second case: $\psi^1(1) = \psi^2(1)$. This implies that for k large, $\psi_k^1(0) = (x_k, p_k^1)$ and $\psi_k^2(0) = (x_k, p_k^2)$ are close to $\psi(0) = (x, p)$ in T^*M and satisfy $\exp_{x_k}(p_k^1) = \exp_{x_k}(p_k^2)$. This show that the exponential mapping from x cannot be injective in a neighborhood of p (which, by construction, satisfies $\exp_x(p) = y$ and $e_{SR}(x, y) = 2H(x, p)$). By the first part of the proposition, we deduce that (x, y) belongs to the global cut-locus of M.
So, we proved that the set $\hat{\Omega}$ is an open subset of Ω. Let us now explain why d_{SR} is smooth on that set.
Let $(x, y) \in \Omega \setminus \mathrm{Cut}_{SR}(M)$, and let p be such that $\exp_x(p) = y$ and $e_{SR}(x, y) = 2H(x, p)$. Since p is not conjugate with respect to x, the exponential mapping is a (smooth) local diffeomorphism from a neighborhood of p into a neighborhood of y. In fact, since the sub-Riemannian distance function is C^1 in a neighborhood of (x, y), for every x' in a neighborhood of x and every y' in a neighborhood of y, we have

$$d_{SR}(x', y') = \sqrt{2H\left(x', (\exp_{x'})^{-1}(y')\right)}.$$

This proves that d_{SR} is of class C^∞ in $\Omega \setminus \mathrm{Cut}_{SR}(M)$. $\square$

From the proof of Lemma 5.13, it follows easily that for every $y \in M \setminus \{x\}$ such that $(x, y) \in \Omega$, if we denote by $\gamma : [0,1] \to M$ a sub-Riemannian geodesic between x and y, then

$$\gamma(t) \in M \setminus \Sigma\left(d_{SR}(x, \cdot)\right) \qquad \forall t \in (0,1),$$

which means that, for every $t \in (0,1)$, there is only one sub-Riemannian minimizing geodesic between x and $\gamma(t)$. A priori, it could happen that $\gamma(t)$ belongs to $\mathrm{Cut}_{SR}(x)$ on a subinterval of the form $[t, 1)$. In fact, under an additional assumption on the sub-Riemannian structure, we can show that this situation cannot occur.

Let $x \in M$. A point $y \in M$ is not a *cut point* with respect to x if there exists a horizontal path $\gamma : [0, L] \to M$ with unit speed such that $d_{SR}(x, \gamma(L)) = \mathrm{length}_{SR}(\gamma)$ and $y = \gamma(t)$ for some $t \in (0, L)$, that is there exists a sub-Riemannian minimizing path joining x to y which is the strict restriction of a minimizing path starting from x. Denote by $\mathcal{L}(x)$ the set of cut points with respect to x. The following result is proved in [35].

Lemma 5.14. *Assume that any nontrivial sub-Riemannian geodesic is nonsingular. Then, for every $x \in M$, one has*

$$\mathrm{Conj}_{min}(x) \subset \mathcal{L}(x).$$

Consequently, like in the Riemannian case, the following result holds.

Proposition 5.15. *Assume that any nontrivial sub-Riemannian geodesic is nonsingular. If γ is a minimizing horizontal curve between $x \neq y$, then we have*

$$\gamma(t) \notin Cut_{SR}(x) \cup Cut_{SR}(y) \qquad \forall t \in (0, 1).$$

An important property of the Riemannian distance function is that it fails to be semiconvex at the cut locus (see [18, Proposition 2.5]. This property plays a key role in the proof of the differentiability of the transport map. We do not know if that property holds in the sub-Riemannian case:

Open problem. Assume that d_{SR} is locally semiconcave on $M \times M \setminus D$. Let $x, y \in M$ be such that there is a function $\phi : M \to \mathbb{R}$ twice differentiable at y such that

$$\phi(y) = d_{SR}(x, y) \quad \text{and} \quad d_{SR}^2(x, z) \geq \phi(z) \quad \forall z \in M.$$

Is it true that $y \notin \mathrm{Cut}_{SR}(x)$?

5.9 Locally lipschitz regularity of the sub-Riemannian distance

Since any locally semiconcave function is locally Lipschitz, Theorem 5.9 above gives a sufficient condition that insures the Lipschitz regularity of d_{SR}^2 out of the diagonal. In [2], Agrachev and Lee demonstrates that, under some stronger assumption, one can prove global Lipschitz regularity. 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, for every $t \in [0, 1]$ and every local parametrization of Δ by smooth vector fields $f_1, \ldots, f_m$ in a neighborhood of $\gamma(t)$, we have

$$\psi(t) \cdot \big([f_i, f_j](\gamma(t))\big) = 0 \qquad \forall i, j = 1, \ldots, m.$$

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

$$p \cdot f_i(\gamma(0)) = p \cdot [f_i, f_j](\gamma(0)) = 0 \qquad \forall i, j = 1, \ldots, m,$$

where $f_1, \ldots, f_m$ is as above a parametrization of Δ in a neighborhood of $\gamma(0)$. Agrachev and Lee proved the following result (see [2, Theorem 5.5]):

Theorem 5.16. *Let Ω be an open subset of $M \times M$ such that any sub-Riemannian minimizing geodesic joining two points of Ω is not a Goh path. Then, the function d_{SR}^2 is locally Lipschitz on $\Omega \times \Omega$.*

6 Proofs of the results

6.1 Proof of Theorem 3.2

Let us first prove (i). We easily see that $\mathcal{M}^\phi$ coincides with the set

$$\{x \in M \mid \phi(x) + \phi^c(x) < 0\}.$$

Thus, since both ϕ and ϕ^c are continuous, $\mathcal{M}^\phi$ is open. Let us now prove that ϕ is locally semiconcave (resp. locally Lipschitz) in an open neighborhood of $\mathcal{M}^\phi \cap \mathrm{supp}(\mu)$. Let $x \in \mathcal{M}^\phi \cap \mathrm{supp}(\mu)$ be fixed. Since $x \notin \partial^c\phi(x)$, there is $r > 0$ such that $d_{SR}(x,y) > r$ for any $y \in \partial^c\phi(x)$. In addition, since the set $\partial^c\phi$ is closed in $M \times M$ and $\mathrm{supp}(\mu \times \nu) \subset \Omega$, there exists a neighborhood $\mathcal{V}_x$ of x which is included in $\mathcal{M}^\phi \cap \pi_1(\Omega)$ and such that

$$d_{SR}(x,w) > r \qquad \forall z \in \mathcal{V}_x, \quad \forall w \in \partial^c\phi(z).$$

Let $\phi_{x,r} : M \to \mathbb{R}$ be the function defined by

$$\phi_{x,r}(z) := \inf \left\{ d_{SR}^2(z,y) - \phi^c(y) \mid y \in \mathrm{supp}(\nu),\, d_{SR}(z,y) > r \right\}.$$

We recall that $\mathrm{supp}(\mu \times \nu) \subset \Omega$ and that d_{SR}^2 is locally semiconcave (resp. locally Lipschitz) in $\Omega \setminus D$. Thus, up to considering a smaller $\mathcal{V}_x$, we easily get that the function $\phi_{x,r}$ is locally semiconcave (resp. locally Lipschitz) in $\mathcal{V}_x$. Since $\phi = \phi_{x,r}$ in $\mathcal{V}_x$, (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 $\{f_1, \ldots, f_m\}$. Moreover, up to working in charts, we can assume that $\mathcal{V}$ is a subset of $\mathbb{R}^n$.

First of all we remark that, since all functions $z \mapsto d_{SR}^2(z,y) - \phi^c(y)$ are locally uniformly Lipschitz with respect to the sub-Riemannian distance when y varies in a compact set, also ϕ is locally Lipschitz with respect to d_{SR}. Up to a change of coordinates in $\mathbb{R}^n$, we can assume that the vector fields f_i are of the form

$$f_i = \frac{\partial}{\partial x_i} + \sum_{j=m+1}^{n} a_{ij}(x) \frac{\partial}{\partial x_j} \qquad \forall i = 1, \ldots, m,$$

with $a_{ij} \in C^\infty(\mathbb{R}^n)$. Therefore, thanks to [30, Theorem 3.2], for a.e. $x \in \mathcal{V}$, ϕ is differentiable with respect to all vector fields f_i, and

$$\phi(y) - \phi(x) - \sum_{i=1}^{m} f_i\phi(x)(y_i - x_i) = o\big(d_{SR}(x,y)\big) \qquad \forall y \in \mathcal{V}. \tag{6.1}$$

Recalling that μ is absolutely continuous, we get that (6.1) holds at μ-a.e. $x \in \mathcal{V}$. Thus it suffices to prove that $\partial^c\phi(x) = \{x\}$ for all such points.

Let us fix such an x. We claim that

$$f_i\phi(x) = 0 \qquad \forall i = 1, \cdots m. \tag{6.2}$$

Indeed, fix $i \in \{1, \cdots, m\}$ and denote by $\gamma_i^x(t) : (-\varepsilon, \varepsilon) \to M$ the integral curve of the vector field f_i starting from x, i.e.

$$\begin{cases} \dot{\gamma}_i^x(t) = f_i(\gamma_i^x(t)) & \forall t \in (-\epsilon, \epsilon) \\ \gamma_i^x(0) = x. \end{cases}$$

By the assumption on x, there is a real number l_i such that

$$\lim_{t \to 0} \frac{\phi(\gamma_i^x(t)) - \phi(x)}{t} = l_i.$$

By construction, the curve γ_i^x is horizontal with respect to Δ. Thus, since $g(\dot{\gamma}_i^x(t), \dot{\gamma}_i^x(t)) = 1$ for any t, we have

$$d_{SR}(x, \gamma_i^x(t)) \leq |t| \qquad \forall t \in (-\varepsilon, \varepsilon).$$

This gives

$$\phi(\gamma_i^x(t)) \leq \phi(x) + d_{SR}^2(\gamma_i^x(t), x) \leq \phi(x) + t^2,$$

which implies that $l_i = 0$ and proves the claim.

Assume now by contradiction that there exists a point $y \in \partial^c\phi(x) \setminus \{x\}$, with $(x, y) \in \Omega$. Then the function

$$z \mapsto \phi(z) - d_{SR}^2(z, y) \leq \phi^c(x)$$

attains a maximum at x. Let $\gamma_{x,y} : [0, 1] \to M$ denotes a minimizing geodesic from x to y. Then

$$\phi(\gamma_{x,y}(t)) - d_{SR}^2(\gamma_{x,y}(t), y) \leq \phi(x) - d_{SR}^2(x, y) \qquad \forall t \in [0, 1],$$

or equivalently

$$\phi(\gamma_{x,y}(t)) - \phi(x) \leq d_{SR}^2(\gamma_{x,y}(t), y) - d_{SR}^2(x, y) \qquad \forall t \in [0, 1].$$

Observe now that, by (6.1) together with (6.2), we have

$$\phi(\gamma_{x,y}(t)) - \phi(x) = o\big(d_{SR}(\gamma_{x,y}(t), x)\big) = o\big(t d_{SR}(x, y)\big).$$

On the other hand, $d_{SR}^2(\gamma_{x,y}(t), y) = (1 - t)^2 d_{SR}^2(x, y)$. Combining all together, for all $t \in [0, 1]$ we have

$$\begin{aligned} o\big(t d_{SR}(x, y)\big) = \phi(\gamma_{x,y}(t)) - \phi(x) &\leq d_{SR}^2(\gamma_{x,y}(t), y) - d_{SR}^2(x, y) \\ &= -2t d_{SR}^2(x, y) + o\big(t d_{SR}(x, y)\big), \end{aligned}$$

that is

$$2t d_{SR}^2(x, y) \leq o\big(t d_{SR}(x, y)\big) \qquad \forall t \in [0, 1].$$

As $x \neq y$, this is absurd for t small enough, and the proof of (ii) is completed.

Since $\mathrm{supp}(\mu \times \nu) \subset \Omega$, we immediately have that any optimal plan γ is concentrated on $\partial^c\phi \cap \Omega$. Moreover, combining (i) and (ii), we obtain that $\partial^c\phi(x)) \cap \mathrm{supp}(\nu)$ is a singleton for μ-a.e. x. This easily gives existence and uniqueness of the optimal transport map.

To prove the formula for $T(x)$, we have to show that

$$\partial^c \phi(x) \cap \operatorname{supp}(\nu) = \exp_x \left(-\frac{1}{2} d\phi(x) \right)$$

for all $x \in \mathcal{M}^\phi \cap \operatorname{supp}(\mu)$ where ϕ is differentiable. This is a consequence of Proposition 5.6 applied to the function $z \mapsto \phi(z) + \phi^c(y)$ at the point x. Moreover, again by Proposition 5.6, the geodesic from x to $T(x)$ is unique for μ-a.e. $x \in \mathcal{M}^\phi \cap \operatorname{supp}(\mu)$. Since $T(x) = x$ for $x \in \mathcal{S}^\phi \cap \operatorname{supp}(\mu)$, the geodesic is clearly unique also in this case.

6.2 Proof of Theorem 3.3

We will prove only (ii), as all the rest follows as in the proof of Theorem 3.2.

Let us consider the "bad" set defined by

$$\mathcal{B} := \left\{ x \in \mathcal{S}^\phi \cap \operatorname{supp}(\mu) \mid (\partial^c \phi(x) \setminus \{x\}) \cap \operatorname{supp}(\nu) \neq \emptyset \right\}.$$

We have to show that $\mathcal{B}$ is μ-negligible. For each $k \in \mathbb{N}$, we consider the sequence of function constructed as follows:

$$\phi_k(x) := \inf \left\{ d_{SR}^2(x,y) - \phi^c(y) \mid y \in \operatorname{supp}(\nu),\, d_{SR}(x,y) > 1/k \right\}.$$

Since $\operatorname{supp}(\mu \times \nu) \subset \Omega$ and d_{SR}^2 is locally semiconcave in $\Omega \setminus D$, the functions ϕ_k are locally semiconcave in a neighborhood of $\mathcal{B}$.

Thus, by Theorem A.8 and the assumptions on μ, there exists a Borel set G, with $\mu(G) = 1$, such that all ϕ_k are differentiable in G. Since for any $x \in \mathcal{B}$ there exists $y \in \partial^c \phi(x) \setminus \{x\}$ such that $d_{SR}(y,x) > 1/k$ for some k, we deduce that

$$\bigcup_{k \in \mathbb{N}} \{\phi = \phi_k\} = \mathcal{B}.$$

This gives that, up to set of μ-measure zero, $\mathcal{B}$ coincides with $\cup_{k \in \mathbb{N}} A_k$, where

$$A_k := \mathcal{B} \cap \{\phi = \phi_k\} \cap G.$$

Thus, to conclude the proof, it suffices to show that $\mu(A_k) = 0$ for all $k \in \mathbb{N}$.

Let $x \in A_k$. Then, if $y \in \partial^c \phi(x)$ and $d_{SR}(x,y) > 1/k$, the function

$$z \mapsto \phi_k(z) - d_{SR}^2(z,y) \leq \phi^c(x) \tag{6.3}$$

attains a maximum at x. Therefore, if we show that $d\phi_k(x) = 0$ for μ-a.e. $x \in A_k$, equation (6.3) together with the semiconcavity of $d_{SR}^2(z,y)$ for z close to x would imply that $d_{SR}^2(\cdot, y)$ is differentiable at x, and its differential is equal to 0. This would contradict Proposition 5.7, concluding the proof. Therefore we just need to show that $d\phi_k(x) = 0$ μ-a.e. in A_k.

Let X be a smooth section of Δ such that $g_x(X(x), X(x)) = 1$ for any $x \in M$. We claim the following:
Claim 1: for μ-a.e. $x \in A_k$, $d\phi_k(x) \cdot X(x) \leq 0$.

Since we can apply Claim 1 with a countable set of vector fields $\{X_l\}$ so that $\{X_l(x)\}$ is dense in $\Delta(x)$ for all $x \in \mathrm{supp}(\mu)$, Claim 1 clearly implies that $d\phi_k(x) = 0$ μ-a.e. in A_k. Let us prove the claim.

Let d_g denote the Riemannian distance associated to the Riemannian metric g, and $\theta(x,t)$ denote the flow of X, that is the function $\theta : M \times \mathbb{R} \to M$ satisfying

$$\frac{d}{dt}\theta(x,t) = X(\theta(x,t)), \qquad \theta(x,0) = x.$$

Fix $\varepsilon > 0$ small, and consider the "cone" around the curve $t \mapsto \theta(x,t)$ given by

$$C_x^\varepsilon := \Big\{ y \in \Omega \mid \exists\, t \in [0,\varepsilon] \text{ such that } d_g\big(\theta(x,t), y\big) \leq \varepsilon t \Big\}.$$

Moreover we define

$$R_\varepsilon := \big\{ x \in \mathrm{supp}(\mu) \cap A_k \mid A_k \cap C_x^\varepsilon = \{x\} \big\}.$$

Claim 2: R_ε is countably $(n-1)$-rectifiable for any $\varepsilon > 0$.
Indeed, since the statement is local, we can assume that we are in $\mathbb{R}^n$, Moreover, since X is smooth, we can assume that there exists $\bar{v} \in \mathbb{R}^n$ such that C_x^ε contains the "euclidean cone"

$$\bar{C}_x^{\varepsilon/2} := \Big\{ y \in \Omega \mid \exists\, t \in [0, \frac{\varepsilon}{2}] \text{ such that } |x + t\bar{v} - y| \leq \frac{\varepsilon}{2} t \Big\}.$$

Thus it suffices to prove that

$$\bar{R}_{\varepsilon/2} := \big\{ x \in \mathrm{supp}(\mu) \cap A_k \mid A_k \cap \bar{C}_x^{\varepsilon/2} = \{x\} \big\}$$

is $(n-1)$-rectifiable for any $\epsilon > 0$.
Assume now that $z, z' \in \bar{R}_{\varepsilon/2}$, with $z \neq z'$. Then, since $z \notin \bar{C}_{z'}^{\varepsilon/2}$, we have

$$|z' + t\bar{v} - z| > \frac{\varepsilon}{2} t \qquad \forall t \in [0, \varepsilon/2],$$

or equivalently

$$|z - t\bar{v} - z'| > \frac{\varepsilon}{2} t \qquad \forall t \in [0, \varepsilon/2].$$

This implies that

$$z' \notin \bar{C}_z^{\varepsilon/2,-} := \Big\{ y \in \Omega \mid \exists\, t \in [0, \frac{\varepsilon}{2}] \text{ such that } |x - t\bar{v} - y| \leq \frac{\varepsilon}{2} t \Big\}.$$

Since $z, z' \in \bar{R}_{\varepsilon/2}$ where arbitrary, we have proved that for all $z \in \bar{R}_{\varepsilon/2}$,

$$\bar{R}_{\varepsilon/2} \cap \big(\bar{C}_z^{\varepsilon/2} \cup \bar{C}_z^{\varepsilon/2,-} \big) = \{z\}.$$

By [13, Theorem 4.1.6] $\bar{R}_\varepsilon$ is countably $(n-1)$-rectifiable for any $\varepsilon > 0$, and this concludes the proof of Claim 2.

Let us come back to the proof of Claim 1. Thanks to Claim 2 we just need to show that

$$x \in \big(\mathrm{supp}(\mu) \cap A_k\big) \setminus \big(\cup_j R_{1/j}\big) \quad \Longrightarrow \quad d\phi_k(x) \cdot X(x) \leq 0.$$

Let $x \in \big(\mathrm{supp}(\mu) \cap A_k\big) \setminus \big(\cup_j R_{1/j}\big)$. Then $\phi(x) = \phi_k(x)$, and there exists a sequence of points $\{x_j\}$ such that $x_j \neq x$ and $x_j \in A_k \cap C_x^{1/j}$ for all $j \in \mathbb{N}$. In particular $\phi(x_j) = \phi_k(x_j)$ for all $j \in \mathbb{N}$. Since $x \in \mathcal{S}^\phi$, we have $x \in \partial^c\phi(x)$, and so

$$\phi(z) - \phi(x) \leq d_{SR}^2(z,x) \qquad \forall z \in M.$$

Let $t_j \in [0, \frac{1}{j}]$ be such that $d_g\big(\theta(x,t_j), x_j\big) \leq \frac{1}{j} t_j$. Then, since d_{SR}^2 is locally Lipschitz, we get

$$\begin{aligned}
\phi_k(x_j) - \phi_k(x) = \phi(x_j) - \phi(x) &\leq d_{SR}^2(x_j,x) \\
&\leq 2d_{SR}^2\big(\theta(x,t_j),x_j\big) + 2d_{SR}^2\big(\theta(x,t_j),x\big) \\
&\leq C d_g\big(\theta(x,t_j),x_j\big) + 2d_{SR}^2\big(\theta(x,t_j),x\big) \\
&\leq \frac{C}{j} t_j + 2d_{SR}^2\big(\theta(x,t_j),x\big).
\end{aligned}$$

We now observe that, since X is a unitary horizontal vector field, $d_{SR}\big(\theta(x,t_j),x\big) \leq t_j$. Moreover $t_j = d_g(x_j,x) + o\big(d_g(x_j,x)\big)$ as $j \to \infty$. Therefore, up to subsequences, one easily gets (looking everything in charts)

$$\lim_{j \to +\infty} \frac{x_j - x}{d_g(x_j,x)} = X(x),$$

which implies

$$d\phi_k(x) \cdot X(x) \leq 0,$$

as wanted.

6.3 Proof of Theorem 3.5

Let us first prove the uniqueness of the Wasserstein geodesic. A basic representation theorem (see [38, Corollary 7.22]) states that any Wasserstein geodesic necessarily takes the form $\mu_t = (e_t)_\# \Pi$, where Π is a probability measure on the set Γ of minimizing geodesics $[0,1] \to M$, and $e_t : \Gamma \to M$ is the evaluation at time t: $e_t(\gamma) := \gamma(t)$. Thus uniqueness follows easily from Theorem 3.2.

The proof of the absolute continuity of μ_t is done as follows. Fix $t \in (0,1)$, and define the function

$$\phi_{1-t}(x) := \inf_{y \in \mathrm{supp}(\nu)} \left\{ \frac{d_{SR}^2(x,y)}{1-t} - \phi^c(y) \right\},$$

$$\phi_t^c(y) := \inf_{x \in \mathrm{supp}(\mu)} \left\{ \frac{d_{SR}^2(x,y)}{t} - \phi(x) \right\}.$$

It is not difficult to see that

$$\frac{d_{SR}^2(x,z)}{t} + \frac{d_{SR}(z,y)^2}{1-t} \geq d_{SR}^2(x,y) \qquad \forall x,y,z \in M. \tag{6.4}$$

Indeed, for all $\varepsilon > 0$,

$$d_{SR}^2(x,y) \leq \big(d_{SR}(x,z) + d_{SR}(z,y)\big)^2 \leq (1+\varepsilon) d_{SR}^2(x,z) + \Big(1 + \frac{1}{\varepsilon}\Big) d_{SR}^2(z,y).$$

Choosing $\varepsilon > 0$ so that $1 + \varepsilon = 1/t$, (6.4) follows. Since $\phi(x) + \phi^c(y) \le d_{SR}^2(x, y)$ for all $x \in \mathrm{supp}(\mu)$ and $y \in \mathrm{supp}(\nu)$, by (6.4) we get

$$\left[\frac{d_{SR}(z,y)^2}{1-t} - \phi^c(y)\right] + \left[\frac{d_{SR}^2(x,z)}{t} - \phi(x)\right] \ge 0 \qquad \forall x \in \mathrm{supp}(\mu),\ y \in \mathrm{supp}(\nu),\ z \in M.$$

This implies

$$\phi_{1-t}(z) + \phi_t^c(z) \ge 0 \qquad \forall z \in M. \tag{6.5}$$

We now remark that (6.4) becomes an equality if and only if there exists a geodesic $\gamma : [0,1] \to M$ joining x to y such that $z = \gamma(t)$. Thus, by the definition of $T_t(x)$ we get

$$\frac{d_{SR}(x, T_t(x))^2}{t} + \frac{d_{SR}(T_t(x), T(x))^2}{1-t} = d_{SR}^2(x, T(x)) \qquad \text{for } \mu\text{-a.e. } x. \tag{6.6}$$

Moreover, since

$$\phi(x) + \phi^c(T(x)) = d_{SR}^2(x, T(x)) \qquad \text{for } \mu\text{-a.e. } x,$$

we obtain

$$\phi_{1-t}(T_t(x)) + \phi_t^c(T_t(x)) = 0 \qquad \text{for } \mu\text{-a.e. } x,$$

or equivalently

$$\phi_{1-t}(z) + \phi_t^c(z) = 0 \qquad \text{for } \mu_t\text{-a.e. } z. \tag{6.7}$$

Let us now decompose the set $\mathcal{M}^\phi \cap \mathrm{supp}(\mu)$ as

$$A_k := \{x \in \mathcal{M}^\phi \cap \mathrm{supp}(\mu) \mid d_{SR}(x,y) > 1/k \quad \forall y \in \partial^c\phi(x)\}.$$

Since $T_t(x) = x$ on $\mathcal{S}^\phi \cap \mathrm{supp}(\mu)$, defining $\mu_t^k := \mu_t \lfloor T_t(A_k)$ we have

$$\mu_t = \big(\cup_k \mu_t^k\big) \cup \mu\lfloor\big(\mathcal{S}^\phi \cap \mathrm{supp}(\mu)\big) \qquad \forall t \in [0,1].$$

Thus it suffices to prove that μ_t^k is absolutely continuous for each $k \in \mathbb{N}$.

We consider the functions

$$\phi_{k,1-t}(x) := \inf\left\{\frac{d_{SR}^2(x,y)}{1-t} - \phi^c(y) \mid y \in \mathrm{supp}(\nu),\ d_{SR}(x,y) > (1-t)/k\right\}.$$

$$\phi_{k,t}^c(y) := \inf\left\{\frac{d_{SR}^2(x,y)}{t} - \phi(x) \mid y \in \mathrm{supp}(\nu),\ d_{SR}(x,y) > t/k\right\}.$$

Since $d_{SR}(x, T(x)) > 1/k$ for $x \in A_k$, they coincide respectively with ϕ_{1-t} and ϕ_t^c inside $T_t(A_k)$. Thus, thanks to (6.5) and (6.7), we have

$$\phi_{k,1-t}(z) + \phi_{k,t}^c(z) \ge \phi_{1-t}(z) + \phi_t^c(z) \ge 0 \qquad \forall z \in M,$$

with equality μ_t-a.e. on $T_t(A_k)$.

Observe now that, by the compactness of the supports of μ and ν, and the fact that Ω is totally geodesically convex, $\mathrm{supp}(\mu \times \mu_t)$ and $\mathrm{supp}(\mu_t \times \nu)$ are compact and contained in Ω. Thus, since d_{SR}^2 is locally semiconcave on $\Omega \setminus D$, both functions $\phi_{k,1-t}$

and $\phi^c_{k,t}$ are locally semiconcave in a neighborhood of $T_t(A_k)$. It follows from [20, Theorem A.19] that both differentials $d\phi_{k,t}(z)$, $d\phi^c_{k,1-t}(z)$ exist and are equal for μ-a.e. $z \in T_s(A_k)$. Moreover, again by [20, Theorem A.19], the map $z \mapsto d\phi_{k,t}(z) = d\phi^c_{k,1-t}(z)$ is locally Lipschitz on $T_s(A_k)$. Since for $x \in A_k$ we have

$$\phi_{k,t}(\cdot) \leq \frac{d_{SR}(x,\cdot)^2}{t} - \phi(x) \qquad \text{on } \{z \mid d_{SR}(x,z) > t/k\}$$

with equality at $T_t(x)$ for μ-a.e. $x \in A_k$, by Proposition 5.7 we get

$$x = \exp_{T_t(x)}(-\tfrac{1}{2} d\phi_{k,t}(T_t(x))) \qquad \text{for } \mu\text{-a.e. } x \in A_k.$$

Denoting by $\Phi_t : T^*M \to T^*M$ the Euler-Lagrange flow (i.e. the flow of the Hamiltonian vector field $\overrightarrow{H}$), we see that the map

$$F_{t,k}(z) := \exp_z(-\tfrac{1}{2} d\phi_{k,t}(z)) = \Phi_t(z, -\tfrac{1}{2} d\phi_{k,t}(z))$$

is locally Lipschitz on $\mathrm{supp}(\mu_t) \cap T_t(A_k)$. Therefore it is clear that μ_t^k cannot have a singular part with respect to the volume measure, since otherwise the same would be true for $(F_{t,k})_\#(\mu_t^k) = \mu \lfloor A_k$. This concludes the proof of the absolute continuity property.

6.4 Proof of Theorem 3.7

We recall that, by Theorem 3.2, the function ϕ is locally semiconcave in a neighborhood of $\mathcal{M}^\phi \cap \mathrm{supp}(\mu)$. Thus, since μ is absolutely continuous with respect to the volume measure, by Theorem A.9 $d\phi(x)$ is differentiable for μ-a.e. $x \in \mathcal{M}^\phi \cap \mathrm{supp}(\mu)$. By Theorem 3.2, for μ-a.e. x there exists a unique minimizing geodesic between x and $T(x)$. Thanks to our assumptions this implies that $T(x) = \exp_x(-\frac{1}{2} d\phi(x))$ do not belongs to $\mathrm{Cut}_{SR}(x)$ for μ-a.e. $x \in \mathcal{M}^\phi \cap \mathrm{supp}(\mu)$. Thus, by Proposition 5.11, the function

$$(z,w) \mapsto d^2_{SR}(z,w)$$

is smooth near $(x, T(x))$. Exactly as in the Riemannian case, this implies that the map $x \mapsto \exp_x(-\frac{1}{2} d\phi(x))$ is differentiable for μ-a.e. x, and its differential is given by $Y(x)\big(H(x) - \frac{1}{2}\mathrm{Hess}^{\,2}_x\phi\big)$ (see [18, Proposition 4.1]). On the other hand, since $T(x) = x$ for $x \in \mathcal{S}^\phi \cap \mathrm{supp}(\mu)$, it is clear by Definition 3.6 that T is approximately differentiable μ-a.e. in $\mathcal{S}^\phi \cap \mathrm{supp}(\mu)$, and that its approximate differential is given by the identity matrix I. This proves the first part of the theorem.

To prove the change of variable formula, we first remark that, since both μ and ν are absolutely continuous, there exists also an optimal transport map S from ν to μ, and it is well-known that S is an inverse for T a.e., that is

$$S \circ T = Id \quad \mu\text{-a.e.}, \qquad T \circ S = Id \quad \nu\text{-a.e.}$$

(see for instance [6, Remark 6.2.11]). This gives in particular that T is a.e. injective. Applying [6, Lemma 5.5.3] (whose proof is in the Euclidean case, but still works on a manifold) we deduce that $|\det(\tilde{d}T(x))| > 0$ μ-a.e., and that the Jacobian identity holds.

A Elements of nonsmooth analysis

The aim of this section is to recall some classical tools of nonsmooth analysis. Recall that throughout this section, M denotes a smooth connected manifold of dimension n.

A.1 Viscosity solutions of Hamilton-Jacobi equations

Let $F : T^*M \times \mathbb{R} \to \mathbb{R}$ be a given continuous function, and let U an open subset of M. A continuous function $u : U \to \mathbb{R}$ is said to be a *viscosity subsolution* on U of the Hamilton-Jacobi equation

$$F(x, du(x), u(x)) = 0 \tag{A.1}$$

if and only if, for every C^1 function $\phi : U \to \mathbb{R}$ satisfying $\phi \geq u$ we have

$$\forall x \in U, \qquad \phi(x) = u(x) \quad \Longrightarrow \quad F(x, d\phi(x), u(x)) \leq 0.$$

Similarly, a continuous function $u : U \to \mathbb{R}$ is said to be a *viscosity supersolution* of (A.1) on U if and only if, for every C^1 function $\psi : U \to \mathbb{R}$ satisfying $\psi \leq u$ we have,

$$\forall x \in U, \qquad \psi(x) = u(x) \quad \Longrightarrow \quad F(x, d\psi(x), u(x)) \geq 0.$$

A continuous function $u : U \to \mathbb{R}$ is called a *viscosity solution* of (A.1) on U if it is both a viscosity subsolution and a viscosity supersolution of (A.1) on U.

A.2 Generalized differentials

Let $u : U \to \mathbb{R}$ be a continuous function on an open set $U \subset M$. The *viscosity subdifferential* of u at $x \in U$ is the convex subset of T_x^*M defined by

$$D^-u(x) := \left\{ d\psi(x) \mid \psi \in C^1(U) \text{ and } u - \psi \text{ attains a global minimum at } x \right\}.$$

Similarly, the *viscosity superdifferential* of u at x is the convex subset of T_x^*M defined by

$$D^+u(x) := \left\{ d\phi(x) \mid \phi \in C^1(U) \text{ and } u - \phi \text{ attains a global maximum at } x \right\}.$$

Note that if u is differentiable at $x \in U$, then $D^-u(x) = D^+u(x) = \{du(x)\}$. The notions of sub and superdifferentials allow to give another characterization of the viscosity sub and supersolutions. A continuous function $u : U \to \mathbb{R}$ is a viscosity subsolution of (A.1) on U if and only if the following property is satisfied:

$$F(x, \zeta, u(x)) \leq 0 \qquad \forall x \in U, \quad \forall \zeta \in D^+u(x).$$

In the same way, a continuous function $u : U \to \mathbb{R}$ is a viscosity supersolution of (A.1) on U if and only if

$$F(x, \zeta, u(x)) \geq 0 \qquad \forall x \in U, \quad \forall \zeta \in D^-u(x).$$

The following result is classical (see [13, 32]).

Proposition A.1. *Let $u : U \to I\!R$ be a continuous function on an open set $U \subset M$. The viscosity subdifferential (resp. superdifferential) of u is nonempty on a dense subset of U.*

The *limiting subdifferential* of u at $x \in U$ is the subset of $T_x^* M$ defined by

$$\partial_L u(x) := \left\{ \lim_{k\to\infty} \zeta_k \mid \zeta_k \in D^- u(x_k),\, x_k \to x \right\}.$$

By construction, the graph of the limiting subdifferential is closed in T^*M. Moreover, the function u is locally Lipschitz on its domain if and only if the graph of the limiting subdifferential of u is locally bounded (see [17, 32]).

Let $u : U \to \mathbb{R}$ be a locally Lipschitz function. The *Clarke generalized subdifferential* of u at the point $x \in U$ is the nonempty compact convex subset of $T_x^* M$ defined by

$$\partial u(x) := \mathrm{conv}\left(\partial_L u(x)\right),$$

that is, the convex hull of the limiting differential of u at x. Notice that, for every $x \in U$,

$$D^- u(x) \subset \partial_L u(x) \subset \partial u(x) \qquad \text{and} \qquad D^+ u(x) \subset \partial u(x).$$

It can be shown that, if $\partial u(x)$ is a singleton, then u is differentiable at x and $\partial u(x) = \{du(x)\}$. The converse result is false.

A.3 Locally semiconcave functions

For an introduction to semiconcavity, we refer the reader to [13] and [20, Appendix A]. A function $u : U \to \mathbb{R}$, defined on the open set $U \subset M$, is called *locally semiconcave* on U if for every $x \in U$ there exist a neighborhood U_x of x and a smooth diffeomorphism $\varphi_x : U_x \to \varphi_x(U_x) \subset \mathbb{R}^n$ such that $f \circ \varphi_x^{-1}$ is locally semiconcave on the open subset $\tilde{U}_x = \varphi_x(U_x) \subset \mathbb{R}^n$. We recall that the function $u : U \to \mathbb{R}$, defined on the open set $U \subset \mathbb{R}^n$, is locally semiconcave on U if for every $\bar{x} \in U$ there exist $C, \delta > 0$ such that

$$\mu u(y) + (1-\mu)u(x) - u(\mu x + (1-\mu)y) \le \mu(1-\mu)C|x-y|^2, \tag{A.2}$$

for all x, y in the ball $B_\delta(\bar{x})$ and every $\mu \in [0,1]$. This is equivalent to say that the function u can be written locally as

$$u(x) = \left(u(x) - C|x|^2\right) + C|x|^2 \qquad \forall x \in B_\delta(\bar{x}),$$

with $u(x) - C|x|^2$ concave. 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 (in fact a better result holds, see Theorem A.8). The following result will be useful in the proof of our theorems.

Lemma A.2. *Let $u : U \to I\!R$ be a function defined on an open set $U \subset I\!R^n$. Assume that for every $\bar{x} \in U$ there exist a neighborhood $\mathcal{V} \subset U$ of $\bar{x}$ and a positive real number σ such that, for every $x \in \mathcal{V}$, there is $p_x \in I\!R^n$ such that*

$$u(y) \le u(x) + \langle p_x, y - x\rangle + \sigma|y-x|^2 \qquad \forall y \in \mathcal{V}. \tag{A.3}$$

Then the function u is locally semiconcave on U.

Proof. Let $\bar{x} \in U$ 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. □

The converse result can be stated as follows (its proof is left to the reader).

Proposition A.3. *Let U be an open and convex subset of $\mathbb{R}^n$ and $u : U \to \mathbb{R}$ be a function which is C-semiconcave on U, that is, which satisfies*

$$\mu u(y) + (1-\mu)u(x) - u(\mu x + (1-\mu)y) \leq \mu(1-\mu)C|x-y|^2, \tag{A.4}$$

for every $x, y \in U$. Then, for every $x \in U$ and every $p \in D^+u(x)$, we have

$$u(y) \leq u(x) + \langle p, y - x \rangle + \frac{C}{2}|y-x|^2 \qquad \forall y \in U, \tag{A.5}$$

In particular, $D^+u(x) = \partial u(x)$ for every $x \in U$.

Remark A.4. As a consequence (see [13, 32]) we obtain that, if a function $u : U \to \mathbb{R}$ is locally semiconcave on an open set $U \subset M$, then, for every $x \in U$,

$$\partial_L u(x) = \left\{ \lim_{k\to\infty} du(x_k) \mid x_k \in \mathcal{D}_u, x_k \to x \right\},$$

where $\mathcal{D}_u$ denotes the set of points of U at which u is differentiable.

The following result is useful to obtain several characterization of the singular set of a given locally semiconcave function (we refer the reader to [13, 32] for its proof):

Proposition A.5. *Let U be an open subset of M and $u : U \to \mathbb{R}$ be a function which is locally semiconcave on U. Then, for every $x \in U$, u is differentiable at x if and only if $\partial u(x)$ is a singleton.*

Another useful result is the following (see [13, Corollary 3.3.8]):

Proposition A.6. *Let $u : U \to \mathbb{R}$ be a function defined on an open set $U \subset M$. If both functions u and $-u$ are locally semiconcave on U, then u is of class $C^{1,1}_{loc}$ on U.*

Fathi generalized the proposition above as follows (see [19] or [20, Theorem A.19]):

Proposition A.7. *Let U be an open subset of M and $u_1, u_2 : U \to \mathbb{R}$ be two functions with u_1 and $-u_2$ locally semiconcave on U. Assume that $u_1(x) \leq u_2(x)$ for any $x \in U$. If we define $\mathcal{E} = \{x \in U \mid u_1(x) = u_2(x)\}$, then both u_1 and u_2 are differentiable at each $x \in \mathcal{E}$ with $du_1(x) = du_2(x)$ at such a point. Moreover, the map $x \mapsto du_1(x) = du_2(x)$ is locally Lipschitz on $\mathcal{E}$.*

A.4 Singular sets of semiconcave functions

Let $u : U \to \mathbb{R}$ be a function which is locally semiconcave on the open set $U \subset M$. We recall that, since such a function is locally Lipschitz on U, its limiting subdifferential is always nonempty on U. We define the *singular set* of u as the subset of U

$$\begin{aligned}\Sigma(u) &:= \{x \in U \mid u \text{ is not differentiable at } x\} \\ &= \{x \in U \mid \partial u(x) \text{ is not a singleton}\} \\ &= \{x \in U \mid \partial_L u(x) \text{ is not a singleton}\}.\end{aligned}$$

From Rademacher's theorem, $\Sigma(u)$ has Lebesgue measure zero. In fact, the following result holds (see [13, 32]):

Theorem A.8. *Let U be an open subset of M. The singular set of a locally semiconcave function $u : U \to \mathbb{R}$ is countably $(n-1)$-rectifiable, i.e., is contained in a countable union of locally Lipschitz hypersurfaces of M.*

A.5 Alexandrov's second differentiability theorem

As shown by Alexandrov (see [38]), locally semiconcave functions are two times differentiable almost everywhere.

Theorem A.9. *Let U be an open subset of $\mathbb{R}^n$ and $u : U \to \mathbb{R}$ be a function which is locally semiconcave on U. Then, for a.e. $x \in U$, u is differentiable at x and there exists a symmetric operator $A(x) : \mathbb{R}^n \to \mathbb{R}^n$ such that the following property is satisfied:*

$$\lim_{t \downarrow 0} \frac{u(x+tv) - u(x) - t\,du(x)\cdot v - \frac{t^2}{2}\langle A(x)\cdot v, v\rangle}{t^2} = 0 \qquad \forall v \in \mathbb{R}^n.$$

Moreover, $du(x)$ is differentiable a.e. in U, and its differential is given by $A(x)$.

B Proofs of auxiliary results

B.1 Proof of Proposition 4.4

The first part of the proposition is just a corollary of Proposition 4.8 for $n = 3$. Let us prove the second part of the proposition. Let $\gamma : [0,1] \to M$ be a nontrivial singular horizontal path. Our aim is to show that, for every $t \in [0,1]$, the point $\gamma(t)$ belongs to Σ_Δ. Fix $\bar{t} \in [0,1]$ and parametrize the distribution by two smooth vector fields f_1, f_2 in an open neighborhood $\mathcal{V}$ of $\gamma(\bar{t})$. Let $u \in L^2([0,1], \mathbb{R}^2)$, and let I be an open subinterval of $[0,1]$ containing $\bar{t}$ such that

$$\dot{\gamma}(t) = u_1(t) f_1(\gamma(t)) + u_2(t) f_2(\gamma(t)) \qquad \text{for a.e. } t \in I.$$

Note that since γ is assumed to be nontrivial, we can assume that u is not identically zero in any neighborhood of $\bar{t}$. From Proposition 5.3, there is an arc $p : [0,1] \longrightarrow (\mathbb{R}^3)^* \setminus \{0\}$ in $W^{1,2}$ such that

$$\dot{p}(t) = -u_1(t) p(t) \cdot df_1(\gamma(t)) - u_2(t) p(t) \cdot df_2(\gamma(t)),$$

for almost every $t \in I$ and

$$p(t) \cdot f_1(\gamma(t)) = p(t) \cdot f_2(\gamma(t)) = 0 \qquad \forall t \in I.$$

Let us take the derivative of the quantity $p(t) \cdot f_1(\gamma(t))$ (which is absolutely continuous). We have for almost every $t \in I$,

$$\begin{aligned}
0 &= \frac{d}{dt}\left[p(t) \cdot f_1(\gamma(t))\right] \\
&= \dot{p}(t) \cdot f_1(\gamma(t)) + p(t) \cdot df_1(\gamma(t)) \cdot \dot{\gamma}(t) \\
&= -\sum_{i=1,2} u_i(t) p(t) \cdot df_i(\gamma(t)) \cdot f_1(\gamma(t)) + \sum_{i=1,2} u_i(t) p(t) \cdot df_1(\gamma(t)) \cdot f_i(\gamma(t)) \\
&= -u_2(t) p(t) \cdot [f_1, f_2](\gamma(t)).
\end{aligned}$$

In the same way, if we differentiate the quantity $p(t) \cdot f_2(\gamma(t))$, we obtain

$$0 = \frac{d}{dt}\left[p(t) \cdot f_2(\gamma(t))\right] = u_1(t) \cdot [f_1, f_2](\gamma(t)).$$

Therefore, since u is not identically zero in any neighborhood of $\bar{t}$, thanks to the continuity of the mapping $t \mapsto p(t) \cdot [f_1, f_2](\gamma(t))$, we deduce that

$$p(\bar{t}) \cdot [f_1, f_2](\gamma(\bar{t}) = 0.$$

But we already know that $p(t) \cdot f_1(\gamma(\bar{t})) = p(t) \cdot f_2(\gamma(\bar{t})) = 0$ where the two vectors $f_1(\gamma(\bar{t})), f_2(\gamma(\bar{t}))$ are linearly independent. Therefore, since $p(\bar{t}) \neq 0$, we conclude that the Lie bracket $[f_1, f_2](\gamma(\bar{t}))$ belongs to the linear subspace spanned by $f_1(\gamma(\bar{t})), f_2(\gamma(\bar{t}))$, which means that $\gamma(\bar{t})$ belongs to Σ_Δ. Let us now prove that any horizontal path included in Σ_Δ is singular. Let γ such a path be fixed, set $\gamma(0) = x$, and consider a parametrization of Δ by two vector fields f_1, f_2 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

$$\dot{\gamma}(t) = u_1(t) f_1(\gamma(t)) + u_2(t) f_2(\gamma(t)) \qquad \text{for a.e. } t \in [0, \delta].$$

Let $p \in (\mathbb{R}^3)^*$ be such that $p_0 \cdot f_1(x) = p \cdot f_2(x) = 0$ and $p : [0, \delta] \to (\mathbb{R}^3)^*$ be the solution to the Cauchy problem

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

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

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

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

$$\dot{h}_1(t) = \frac{d}{dt}\left[p(t) \cdot f_1(\gamma(t))\right] = -u_2(t) p(t) \cdot [f_1, f_2](\gamma(t))$$

and

$$\dot{h}_2(t) = u_1(t)p(t) \cdot [f_1, f_2](\gamma(t)).$$

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

$$[f_1, f_2](\gamma(t)) = \lambda_1(t) f_1(\gamma(t)) + \lambda_2(t) f_2(\gamma(t)), \qquad \forall t \in [0, \delta].$$

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

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

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 abnormal lift of γ on the interval $[0, \delta]$. We can in fact repeat this construction on a new interval of the form $[\delta, 2\delta]$ (with initial condition $p(\delta)$) and finally obtain an abnormal lift of γ on $[0, 1]$. By Proposition 5.2, we conclude that γ is singular.

B.2 Proof of Proposition 4.8

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

$$\Delta(x) = \mathrm{Span}\{f_1(x), \ldots, f_{n-1}(x)\} \qquad \forall x \in \mathcal{V},$$

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

$$f_i = \frac{\partial}{\partial x_i} + \alpha_i(x) \frac{\partial}{\partial x_n} \qquad \forall i = 1, \ldots, n-1,$$

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

$$[f_i, f_j] = \left[\left(\frac{\partial \alpha_j}{\partial x_i} - \frac{\partial \alpha_i}{\partial x_j} \right) + \left(\frac{\partial \alpha_j}{\partial x_n} \alpha_i - \frac{\partial \alpha_i}{\partial x_n} \alpha_j \right) \right] \frac{\partial}{\partial x_n},$$

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 f_I the C^∞ vector field constructed by Lie brackets of $f_1, f_2, \ldots, f_{n-1}$ as follows,

$$f_I = [f_{i_1}, [f_{i_2}, \ldots, [f_{i_{k-1}}, f_{i_k}] \ldots]].$$

We call $k = \mathrm{length}(I)$ the length of the Lie bracket f_I. Since Δ is nonholonomic, there is some positive integer r such that

$$\mathbb{R}^n = \mathrm{Span}\left\{ f_I(x) \mid \mathrm{length}(I) \le r \right\} \qquad \forall x \in \mathcal{V}.$$

It is easy to see that, for every I such that length$(I) \geq 2$, there is a C^∞ function $g_I : \mathcal{V} \to \mathbb{R}$ such that

$$f_I(x) = g_I(x)\frac{\partial}{\partial x_n} \qquad \forall x \in \mathcal{V}.$$

Defining the sets A_k as

$$A_k := \{x \in \mathcal{V} \mid g_I(x) = 0 \quad \forall I \text{ such that length}(I) \leq k\},$$

we have

$$\Sigma_\Delta = \bigcup_{k=2}^{r} (A_k \setminus A_{k+1}).$$

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

$$g_J(x) = \left(\frac{\partial g_I}{\partial x_{j_1}}(x) + \frac{\partial g_I}{\partial x_n}(x)\alpha_{j_1}(x)\right)\frac{\partial}{\partial x_n} \neq 0.$$

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.

References

[1] A. Agrachev. Compactness for Sub-Riemannian length-minimizers and subanalyticity. Control theory and its applications (Grado, 1998), *Rend. Sem. Mat. Univ. Politec. Torino*, 56(4):1–12, 2001.

[2] A. Agrachev and P. Lee. Optimal transportation under nonholonomic constraints. Preprint, 2007.

[3] A. Agrachev and Y. Sachkov. *Control theory from the geometric viewpoint.* Encyclopaedia of Mathematical Sciences, 87, Control Theory and Optimization, II, Springer-Verlag, Berlin, 2004.

[4] A. Agrachev and A. Sarychev. Sub-Riemannian metrics: minimality of singular geodesics versus subanalycity. *ESAIM Control Optim. Calc. Var.*, 4:377–403, 1999.

[5] G. Alberti and L. Ambrosio: A geometrical approach to monotone functions in $\mathbb{R}^n$. *Math. Z.*, 230(2):259–316, 1999.

[6] L. Ambrosio, N. Gigli and G. Savaré. Gradient flows in metric spaces and in the Wasserstein space of probability measures. *Lectures in Mathematics*, ETH Zurich, Birkhäuser, 2005.

[7] L. Ambrosio and S. Rigot. Optimal mass transportation in the Heisenberg group. *J. Funct. Anal.*, 208(2):261–301, 2004.

[8] A. Bellaïche. The tangent space in sub-Riemannian geometry. In *Sub-Riemannian Geometry*, Birkhäuser, 1–78, 1996.

[9] P. Bernard and B. Buffoni. Optimal mass transportation and Mather theory. *J. Eur. Math. Soc.*, 9(1):85–121, 2007.

[10] U. Boscain and F. Rossi. Invariant Carnot-Carathéodory metrics on $S^3, SO(3), SL(2)$ and Lens spaces. Preprint, 2007.

[11] Y. Brenier. Polar factorization and monotone rearrangement of vector-valued functions. *Comm. Pure Appl. Math.*, 44:375–417, 1991.

[12] P. Cannarsa and L. Rifford. Semiconcavity results for optimal control problems admitting no singular minimizing controls. *Ann. Inst. H. Poincaré Non Linéaire*, to appear.

[13] P. Cannarsa and C. Sinestrari. *Semiconcave functions, Hamilton-Jacobi equations, and optimal control.* Progress in Nonlinear Differential Equations and their Applications, 58. Birkhäuser Boston Inc., Boston, MA, 2004.

[14] Y. Chitour, F. Jean and E. Trélat. Genericity results for singular curves. *J. Diff. Geom.*, 73(1):45–73, 2006.

[15] Y. Chitour, F. Jean and E. Trélat. Singular trajectories of control-affine systems. *SIAM J. Control Optim.*, to appear.

[16] W. L. Chow. Über Systeme von linearen partiellen Differentialgleichungen ester Ordnung. *Math. Ann.*, 117: 98–105, 1939.

[17] F. H. Clarke, Yu. S. Ledyaev, R. J. Stern and P. R. Wolenski. *Nonsmooth Analysis and Control Theory.* Graduate Texts in Mathematics, vol. 178. Springer-Verlag, New York, 1998.

[18] D. Cordero-Erausquin, R. McCann and M. Schmuckenschlaeger. A Riemannian interpolation inequality a la Borell, Brascamp and Lieb. *Invent. Math.*, 146:219–257, 2001.

[19] A. Fathi. Weak KAM Theorem and Lagrangian Dynamics. Cambridge University Press, to appear.

[20] A. Fathi and A. Figalli. Optimal transportation on non-compact manifolds. *Israel J. Math.*, to appear.

[21] H. Federer. *Geometric measure theory.* Die Grundlehren des mathematischen Wissenschaften, Band 153. Springer-Verlag, New York, 1969.

[22] A. Figalli. Existence, Uniqueness, and Regularity of Optimal Transport Maps. *SIAM Journal of Math. Anal.*, 39(1):126–137, 2007.

[23] A. Figalli and N. Juillet. Absolute continuity of Wasserstein geodesics in the Heisenberg group. Preprint, 2007.

[24] L. V. Kantorovich. On the transfer of masses. *Dokl. Akad. Nauk. SSSR*, 37:227–229, 1942.

[25] L. V. Kantorovich. On a problem of Monge. *Uspekhi Mat. Nauk.*, 3:225–226, 1948.

[26] W. Liu and H. J. Sussmann. Shortest paths for sub-Riemannian metrics on rank-2 distributions. *Mem. Amer. Math. Soc.*, 118(564), 1995.

[27] R. McCann. Polar factorization of maps in Riemannian manifolds. *Geom. Funct. Anal.*, 11:589–608, 2001.

[28] R. Montgomery. Abnormal minimizers. *SIAM J. Control Optim.*, 32(6):1605–1620, 1994.

[29] R. Montgomery. *A tour of sub-Riemannian geometries, their geodesics and applications. Mathematical Surveys and Monographs*, Vol. 91. American Mathematical Society, Providence, RI, 2002.

[30] R. Monti and F. Serra Cassano. Surface measures in Carnot-Carathéodory spaces. *Calc. Var. Partial Differential Equations*, 13(3):339–376, 2001.

[31] P. K. Rashevsky. About connecting two points of a completely nonholonomic space by admissible curve. *Uch. Zapiski Ped. Inst. Libknechta*, 2:83–94, 1938.

[32] L. Rifford. *Nonholonomic Variations: An Introduction to Subriemannian Geometry.* Monograph, in progress.

[33] L. Rifford and E. Trélat. Morse-Sard type results in sub-Riemannian geometry. *Math. Ann.*, 332(1):145–159, 2005.

[34] L. Rifford and E. Trélat. On the stabilization problem for nonholonomic distributions. *J. Eur. Math. Soc.*, to appear.

[35] A. Sarychev. The index of the second variation of a control system. *Math. USSR Sbornik*, 41(3):383–401, 1982.

[36] H. J. Sussmann. A cornucopia of four-dimensional abnormal sub-Riemannian minimizers. In *Sub-Riemannian Geometry*, Birkhäuser, 341–364, 1996.

[37] C. Villani. Topics in Mass Transportation. *Graduate Studies in Mathematics Surveys*, Vol. 58. American Mathematical Society, Providence, RI, 2003.

[38] C. Villani: Optimal transport, old and new. Lecture notes, 2005 Saint-Flour summer school, available online at *http://www.umpa.ens-lyon.fr/˜cvillani.*