NNT :

THESE DE DOCTORAT DE L' UNIVERSITE DE BOURGOGNE FRANCHE-COMTE

PREPAREE A L'UNIVERSITÉ DE DIJON

Ecole doctorale n° 553

Carnot-Pasteur

Doctorat de mathématiques

Par

Mme BADREDDINE Zeinab

MASS TRANSPORTATION IN SUB-RIEMANNIAN STRUCTURES ADMITTING SINGULAR MINIMIZING GEODESICS

Soutenance prévue à Dijon, le 04 décembre 2017

Composition du Jury:

M, JEAN, Frédéric	Professeur, Ensta Paris-Tech	Rapporteur
M, LE DONNE, Enrico	Professeur, Université de Jyväskylä	Rapporteur
Mme, RIGOT, Séverine	Professeur, Université de Nice	Examinatrice
M, BONNARD, Bernard	Professeur, Université de Bourgogne	Directeur de thèse
M, RIFFORD, Ludovic	Professeur, Université de Nice	Co-directeur de thèse

Résumé

Cette thèse est consacrée à l'étude des problèmes de transport optimal en géométrie sous-Riemannienne. Plus précisément, notre but est d'étendre le caractère bien-posé du problème de Monge aux cas des structures sous-Riemanniennes admettant des géodésiques minimisantes singulières. Dans une première partie, on traite le cas des distributions analytiques de rang 2 en dimension 4. On montre l'existence et l'unicité de solutions pour le coût quadratique sous-Riemannien, tant que la distribution satisfait une certaine condition de croissance. La stratégie de la preuve combine la technique de Figalli-Rifford basée sur la régularité de la distance sous Riemannienne en dehors de la diagonale en absence de géodésiques minimisantes singulières, avec une propriété de contraction de la mesure pour les courbes singulières dans l'esprit d'un résultat de Cavalletti et Huesmann. Dans une deuxième partie, on s'intéresse aux problèmes de régularité de la distance sous-Riemannienne et on définit sur les groupes de Carnot, les structures sous-Riemanniennes h-idéales sur lesquelles la distance sous-Riemannienne est h-semiconcave. Sous une certaine hypothèse de la distribution, on prouve heuristiquement la propriété MCP sur les groupes de Carnot. On prévoit que cette propriété MCP est une condition suffisante pour appliquer la méthode de Cavalletti-Huesmann afin de prouver que le problème de Monge est bien-posé.

Mots clés : problème de Monge, géométrie sous-Riemannienne, singulières minimisantes.

Abstract

This thesis is devoted to the study of mass transportation on sub-Riemannian geometry. More precisely, our aim is to extend previous results on the well-posedness of the Monge problem to cases of sub-Riemannian structures admitting singular minimizing geodesics. In the first part, we show that, in the case of rank-two analytic distributions in dimension four, we have existence and uniqueness of solutions for the sub-Riemannian quadratic cost, as soon as the distribution satisfies some growth condition. Our strategy to prove it, combines the technique used by Figalli-Rifford which is based on the regularity of the sub-Riemannian distance outside the diagonal in absence of singular minimizing curves, together with a localized contraction property for singular curves in the spirit of the previous work by Cavalletti and Huesmann. In the second part, we deal with regularity issues of the sub-Riemannian distance and we define a class of sub-Riemannian structures on Carnot groups, called h-ideal sub-Riemannian structures, on which the sub-Riemannian distance is h-semiconcave. Together with an assumption on the distribution, we prove heuristically the MCP property on Carnot groups. Anyway, we attempt to prove that MCP property defined on this class of Carnot groups is sufficient to apply the Cavalletti-Huesmann method to prove the well-posedness of the Monge problem.

Keywords : Monge problem, sub-Riemannien geometry, singular minimizers.

Acknowledgement

The completion of this doctoral thesis would not have been possible without the support of several people for whom I would like to express my sincere gratitude.

Firstly, I am extremely grateful to my advisor Prof. Bernard BONNARD for the continuous support to make this thesis possible. I acknowledge you for all the patience and kindness. I would like to express my sincere gratitude to my co-advisor Prof. Ludovic RIFFORD. Thank you for your patience, motivation and consistent encouragement. Your valuable guidance helped me in all the time of research and writing this thesis. It is and it will always be an honor to work with you.

My deepest appreciation goes to my thesis committee members, Prof. Frédéric JEAN and Prof. Enrico LE DONNE and Prof. Séverine RIGOT. Thank you for spending time reading, reviewing and commenting my work. It is my honor and privilege to present my work to you.

A very special gratitude goes to all down at IMB in Dijon and LJAD in Nice. The thesis would not have come to a successful completion, without the help I received from the Inria team, McTAO, in Sophia-Antipolis. All of you have extended his support in a very special way, and I gained a lot from you, through your personal and scholarly interactions, and suggestions.

Lastly and most importantly, a special thanks to my family. Words cannot express how grateful I am to my parents for all their unconditional love and encouragement. My brother, I can always count on you to cheer me up and to make things better. To you I dedicate this thesis.

Contents

Chapter 1

Introduction

This thesis, running from 2014 to 2017, was an occasion to discover the Theory of Optimal Transportation and sub-Riemannian Geometry. The mass transportation theory found its roots through the French geometer Monge and became a popular subject in various areas of sciences including economics, optic such as the reflector problem, meteorology and oceanography. On the other hand, important motivations from the field of thermodynamics involving hyperbolic geometry, analysis of hyper elliptic operators etc, made the first steps towards sub-Riemannian geometry. This is a natural geometry in mechanics to study systems with nonholonomic constraints like skates, wheels, rolling balls etc, for reference see the book [Mont02]. The aim of this thesis is to present some of the recent progress in solving the Monge problem on sub-Riemannian manifolds.

Optimal mass transportation theory was born in the $XVIII^e$ century. It was raised by the French mathematician and physicist Gaspard Monge through one of his remarkable writings [Mon81], *Mémoire sur la théorie des déblais et des remblais*, published in 1781. He has formulated a mathematical problem of "Excavations and enbankements", concerning with the transport of a pile of soil to completely fill up an excavation with minimal transport expenses.

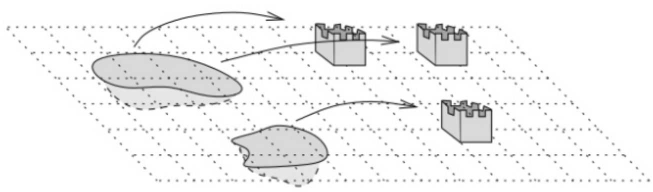


Figure 1.1

We shall model the pile of soil by a measure μ and the excavation by a measure ν, both defined on some measured space X. Obviously, the transport involves a conservation of mass, that is why in the sequel, we only consider probability measures. We also model the transportation by a measurable cost function $c(x, y)$, which denotes the cost of transporting a unit of mass from a position x to some position y. Thus, the Monge problem can be formulated as follows:

Definition 1. *(The Monge Problem) Let X be a measured space, and $c : X \times X \to [0, +\infty[$ be the cost function. Let μ, ν be two probability measures on X. Then, the Monge problem consists in minimizing the transportation cost*

$$\int_X c(x, T(x)) \, \mathrm{d}\mu(x),$$

among all the measurable maps $T : X \to X$ pushing forward μ to ν (we denote it by $T_\sharp\mu = \nu$).

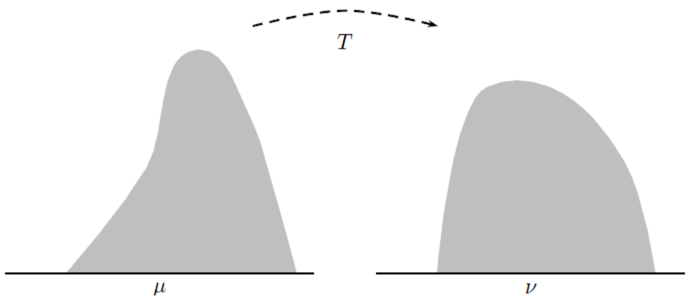


Figure 1.2 – The Monge transportation problem

When T satisfies the transport condition $T_\sharp\mu = \nu$ and minimizes the cost, we call it an optimal transport map. Monge proposed a method to construct an optimal transport map without proving it. And the Monge problem has stayed with no solution for centuries. For this purpose, the Academy of Paris offered the Bordin prize [Dar85] in 1884. However, a rigorous proof of the subject was claimed by Appell [App87] in 1887. Studied by many other researchers, Sudakov [Sud79] showed in 1979 solutions for the Monge problem as mappings from $\mathbb{R}^n$ to $\mathbb{R}^n$.

The Russian mathematician and economic Leonid Kantorovich, who received a Nobel prize in 1975 in economics, drew an attention to the Monge problem and saw a way to connect it to his work which gave the possibility to find solutions and to study them. In particular, we turn to his work [Ka42] in 1942. He introduced a relaxation form of the Monge problem by representing the transportation as a probability measure.

Definition 2. *(The Kantorovich Problem) Let X be a measured space, and $c : X \times X \to [0, +\infty[$ be the cost function. Let μ, ν be two probability measures on X. Then, the Kantorovich problem consists in minimizing the transportation cost*

$$\int_{X \times X} c(x, y)\ \mathrm{d}\alpha(x, y),$$

among all the probability measures α on $X \times X$ such that $P^1_\sharp(\alpha) = \mu$ and $P^2_\sharp(\alpha) = \nu$ (with $P^i : X \times X \to X$ the projection map into the i-th component).

Moreover, Kantorovich shows that the minimization problem introduced in Definition 2 is equivalent to the following maximization problem:

Definition 3. *(The Dual Problem) Let X be a measured space, and $c : X \times X \to [0, +\infty[$ be the cost function. Let μ, ν be two probability measures on X. Then, the dual problem consists in maximizing the following quantity*

$$\int_X \psi(y)\mathrm{d}\nu(y) - \int_X \varphi(x)\mathrm{d}\mu(x),$$

among all the functions $(\varphi, \psi) \in L^1(\mu) \times L^1(\nu)$ such that

$$\psi(y) - \varphi(x) \leq c(x, y).$$

Since that time, the transport problem is known as the "Monge-Kantorovich transport problem". This formulation allows to highlight the concept of c-convexity, where c is the given cost function. A way to see that a function ϕ is c-convex, is to show that for each point we can put under the function ϕ a support function of the form $x \mapsto -c(x, y) + constant$.

Again, the work of Yann Brenier [Br91] in 1991, gave a new birthdate for the Monge problem. He showed existence and uniqueness of solution for the quadratic Monge problem for a quadratic Euclidean cost. Then, McCann [Mc01] provided the solution for the Riemannian case where the cost is given by the quadratic geodesic distance. These types of results are based on the local Lipschitzness of the cost function. In case the probability measures are compactly supported, regularity properties of the solutions for the dual problem are obtained from the regularity of the cost. Recently, Cavalletti and Huesmann [CH15] develop a new technic to solve the Monge problem, based on a contraction property on proper non-branching measured metric spaces. This contraction property concerns the behavior of the measure under the shrinking of sets to points. The non-branching condition is necessary to get some consequences of the contraction property. In particular, their proof covers spaces enjoying the (k,n)-measure contraction property, abbreviated MCP(k,n). The notion of MCP has been introduced by Ohta [Oh07], and Sturm [St06]. For sake of simplicity, we proceed to define the notion of measure contraction property on measured metric spaces with negligible cut locus.

Definition 4. *A measured metric space (X, d, μ) is said to be with negligible cut locus if for every $x \in X$, there exists a measurable set $C(x) \subset X$ with*

$$\mu(C(x)) = 0$$

such that $\forall y \in X \backslash C(x)$, there is a unique minimizing geodesic $\gamma_x : [0,1] \to X$ connecting x to y.

According to the Ohta definition [Oh07] (see also [Rif13]), we have the following definition:

Definition 5. *Let (X, d, μ) be a measured metric space with negligible cut locus and $k \in \mathbb{R}$, $n > 1$ be fixed. We say that (X, d, μ) satisfies MCP(k, n) if for every $x \in X$ and every measured set $A \subset X$ (provided that $A \subset B(x, \pi\sqrt{n-1/k})$ if $k > 0$) with $0 < \mu(A) < \infty$,*

$$\mu\big(A_{s,x}\big) \geq \int_A s\Big[\frac{s_k(sd(x,z)/\sqrt{n-1})}{s_k(d(x,z)/\sqrt{n-1})}\Big]^{n-1} \mathrm{d}\mu(z)$$

where $A_{s,x}$ is the s-interpolation of A from x defined by

$$A_{s,x} := \Big\{\gamma_x(s) |\ \gamma_x(0) = x, \gamma_x(1) \in A \backslash C(x)\Big\}, \forall s \in [0,1].$$

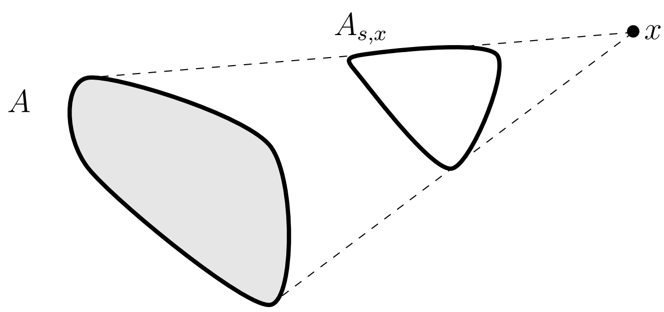


Figure 1.3

and the function $s_k : [0,+\infty[\to [0,+\infty[$ ($s_k : [0,\pi/\sqrt{k}[\to [0,+\infty[$ if $k>0$) is defined by

$$s_k(t) := \begin{cases} \dfrac{sin(\sqrt{k}t)}{\sqrt{k}} & \text{if } k>0 \\ t & \text{if } k=0 \\ \dfrac{sinh(\sqrt{-k}t)}{\sqrt{-k}} & \text{if } k<0 \end{cases}.$$

For example, $\mathbb{R}^n$ equipped with a constant Riemannian metric satisfies MCP(0,n). The MCP(k,n) can be regarded as a generalized notion of the lower curvature bound on Riemannian manifolds. Let (X, d, μ) be a complete Riemannian manifold, the condition MCP(k,n) is equivalent to require that X has Ricci curvature bounded from below by $k \in \mathbb{R}$ (see Theorem 3.2 in [Oh07]), and $n \geq 1$ is the upper bound of the dimension of the space. Recently, it has been studied on Heisenberg groups by Juillet [Ju09] (see also Rifford [Rif13] and Rizzi [Riz16] for other MCP type results on Carnot groups).

This thesis is concerned with the study of the Monge problem in sub-Riemannian cases. The sub-Riemannian geometry is defined on a manifold M, on which every trajectory has velocity contained in a subbundle Δ of the tangent bundle TM, called "horizontal" distribution. Such trajectories are called "horizontal" paths. Riemannian manifolds appear as a special case of sub-Riemannian geometry on which $\Delta = TM$. Roughly speaking, in sub-Riemannian geometry, we cannot move in all directions. We always have a kind of "hidden" or "forbidden" directions. Let us describe a simple model. We consider in $\mathbb{R}^3$ with coordinates (x, y, z), the distribution Δ generated by the two smooth vector fields $X^1 := \partial_x - y\partial_z$ and $X^2 := \partial_y + x\partial_z$. In this case, horizontal paths $t \mapsto (x(t), y(t), z(t))$ are those paths satisfying $\dot{z}(t) = x(t)\dot{y}(t) - y(t)\dot{x}(t)$. The study of horizontal curves is useful in determining the sub-Riemannian distance between two points. The latter is defined as the infimum of the lengths of all horizontal paths joining these two points. In the bracket generating case, the Chow-Rashevsky Theorem (see [Cho39],[Ras38]) guarantees us that the sub-Riemannian distance between two points is finite. One can ask what is the shortest path one should consider to transport a mass from one position to another. It is relevant to concentrate on developing similarities between Riemannian and sub-Riemannian geometries. However, there are major differences. In particular, the space of horizontal curves joining two points can have singularities. A minimizing geodesic is defined as a horizontal curve which minimizes the distance between its endpoints. The existence of "singular" paths is of central importance to sub-Riemannian geometry because singular paths can be minimizers.

The study of the Monge problem in the sub-Riemannian geometry has been concerned with the sub-Riemannian quadratic cost (given by the square of the sub-Riemannian distance). It began with a paper by Ambrosio and Rigot [AR04] about the transportation problem in Heisenberg groups, seen as the prototype of the sub-Riemannian geometry. However, Agrachev and Lee [AL09] proved that the local Lipschitzness of the squared sub-Riemannian distance is sufficient to guarantee existence and uniqueness of solutions for the Monge problem. Then, Figalli and Rifford [FR10] removed the assumption of Lipschitzness on the diagonal. Their proof considered on the diagonal was based on a Pansu-Rademacher Theorem. Furthermore, Rifford [Rif14] proved the local semiconcavity of the sub-Riemannian distance in absence of singular minimizing curves. The semiconcavity brings us closer to a smooth regularity: it can be seen locally as the sum of a smooth function and a concave function. Such result shows that, in absence of singular minimizing paths, sub-Riemannian distances enjoy the same kind of regularity as Riemannian distances. For example, in the case of a two-rank distribution Δ on a three-dimensional manifold M, we have existence and uniqueness of solutions for the sub-Riemannian quadratic cost because non-trivial singular horizontal paths are included in the Martinet surface given by $\Sigma_\Delta := \{x \in M | \Delta(x) + [\Delta,\Delta](x) \neq T_xM\}$ which has Lebesgue measure zero. In general, we do not know if the Monge problem (for the sub-Riemannian quadratic cost) has solutions.

Our aim is to extend previous results on existence and uniqueness of optimal transport maps to cases of sub-Riemannian structures which admits many singular minimizing geodesics. The first relevant case to consider is the one of rank-two distributions in dimension four. In this case, as shown by Sussman [Sus96], singular horizontal paths can be seen (locally) as the orbits of a smooth vector field, at least, outside a set of Lebesgue measure zero. Our aim is to show that, in the case of rank-two analytic distribution in dimension four, we have existence and uniqueness of solutions for the sub-Riemannian quadratic cost, as soon as the distribution satisfies some growth condition.

Theorem 1. *Let M be a real analytic manifold of dimension 4 and (Δ, g) be a complete analytic sub-Riemannian structure of rank 2 on M such that*

$$\forall x \in M,\ \Delta(x) + [\Delta,\Delta](x) \text{ has dimension 3}, \tag{1.1}$$

where

$$[\Delta,\Delta] := Span\big\{[X,Y] \mid X, Y \text{ sections of } \Delta\big\}.$$

Let μ, ν be two probability measures with compact support on M such that μ is absolutely continuous with respect to the Lebesgue measure.

Then, there is existence and uniqueness of an optimal transport map from μ to ν for the sub-Riemannian quadratic cost $c : M \times M \to [0, +\infty[$ defined by:

$$c(x, y) := d_{SR}^2(x, y), \ \forall (x, y) \in M \times M.$$

This theorem is proved in chapter 3. Our strategy to prove it, is twofold. It combines the technique used by Figalli-Rifford [FR10] (see also the paper by Agrachev-Lee [AL09]) which is based on the regularity of the distance function outside the diagonal in absence of singular minimizing curves, together with a localized contraction property for singular curves in the spirit of the previous work by Cavalletti and Huesmann [CH15].

As we saw before, in order to obtain existence and uniqueness for optimal transport maps, it is convenient to be able to show that MCP is satisfied to apply Cavalletti-Huesmann's method. So we deal with regularity issues of the sub-Riemannian distance and we define a class of sub-Riemannian structures on Carnot groups, called h-ideal sub-Riemannian structures, on which the sub-Riemannian distance d_{SR} is h-semiconcave. Such regularity is fundamental. Together with an assumption on the distribution (see ASSUMPTION 1 (6.10)), we prove the MCP property on Carnot groups as a consequence of the upper bound of the horizontal symmetrical Hessian of d_{SR}.

Proposition 1. *Let $\mathcal{G}$ be a Carnot group whose first layer is h-ideal and satisfies ASSUMPTION 1. Then , there is $N > 0$ such that $(\mathcal{G}, d_{SR}, \mathcal{L}^n)$ satisfies $MCP(0, N)$.*

The differentiability of an h-semiconcave function is the consequence of a sub-Riemannian version of the famous theorem of Alexandrov [Mag05] (see also [DGNT04]) which states that an h-semiconcave function is two times differentiable a.e. with respect to the horizontal directions whenever its second order horizontal derivatives are Radon measures.

After the detailed presentation and explanation of our research work, we now proceed to the structure of this thesis.

Chapter 2 can be seen as a general introduction of the optimal mass transportation problem. It concerns the study of the Monge-Kantorovich problem. We investigate a powerful duality formulation due to Kantorovich. The main purpose is to prove existence and uniqueness of an optimal transport map solution for this mass transportation problem. The chapter ends with the statement of preliminary results of existence and uniqueness of solutions for the Monge problem, in the case of quadratic Euclidean cost and the quadratic Riemannian cost which refer respectively to Theorems by Brenier [Br91] and McCann [Mc01].

Chapter 3 presents the basics of the sub-Riemannian geometry. We refer to the distribution as the horizontal space, and objects tangent to it as horizontal. We introduce the Hörmander condition as a bracket generating condition under which the Chow-Rashevsky is true. The Chow-Rashevsky Theorem gives us a license to search for minimizing geodesics, i.e. shortest horizontal curves. We study the End-point map on a sub-Riemannian manifold and its singularities. This End-point map yields a horizontal path passing through a fixed point to its endpoint. Its critical points are called singular paths for the distribution. They play a major role in this thesis.

In chapter 4, we turn our attention to the optimal transport problem on sub-Riemannian manifolds where the cost function is given by the square of the sub-Riemannian distance. Under regularity assumptions for the sub-Riemannian distance, Figalli and Rifford generalized the Brenier-McCann Theorem. We also give an introduction to Cavalletti- Huesmann's method to prove existence and uniqueness of the optimal transport map, using a measure contraction assumption.

Chapter 5 is devoted entirely to the proof of the theorem 1. This section is the subject of an article to appear [Bad17].

Then, we define in chapter 6 the class of h-ideal sub-Riemannian structures on Carnot groups. We present some analytic tools necessary to the understanding of the h-semiconcavity. The chapter ends by establishing the MCP on Carnot groups under suitable regularity assumptions. Unfortunately, this chapter is prospective. Until now, there are no obvious examples of Carnot groups satisfying these hypotheses.

Finally, in chapter 7, we make some comments about this work and try to sketch some research perspectives that may lead to some interesting results.

Chapter 2

Optimal Transport Theory

In the sequel, M denotes a smooth connected manifold without boundary of dimension $n \geq 2$.

2.1 The Monge Problem

The transport problem considered by Monge, was to transport some mass from one place to another with minimal cost. A current formulation of the Monge problem is the following:

Definition 6 (**Transport map**). *Let μ, ν be two probability measures on M, and $c : M \times M \to [0, +\infty[$ be the cost function. We call transport map from μ to ν, any μ-measurable application $T : M \to M$ such that $T_\sharp \mu = \nu$.*

The condition $T_\sharp \mu = \nu$ means that T is pushing forward μ to ν, i.e. for any ν-measurable set B in the target space M,

$$\nu(B) = \mu(T^{-1}(B)).$$

Therefore, ***the Monge problem*** *is modeled as an optimal transport problem minimizing the transportation cost*

$$\int_M c(x, T(x)) \mathrm{d}\mu(x), \text{ among all the transport maps } T : M \to M.$$

We check that $T_\sharp \mu = \nu$ is equivalent to a change of variables formula. In fact, consider $M = \mathbb{R}^n$ and μ, ν two probability measures on M absolutely continuous with respect to the Lebesgue measure. We take $\mu = f \mathrm{d}x$ and $\nu = g \mathrm{d}y$, with

$f, g \in \mathcal{L}^1(\mathbb{R}^n, \mathbb{R})$. Then, for any ν-measurable set B in the target space, the condition $T_\sharp\mu = \nu$ yields

$$\int_{T^{-1}(B)} f(x)\mathrm{d}x = \int_B g(y)\mathrm{d}y.$$

If T is a diffeomorphism, we perform the change of variable $y = T(x)$, that leads to

$$\int_{T^{-1}(B)} f(x)\mathrm{d}x = \int_{T^{-1}(B)} g(T(x))|det(D_xT)|\mathrm{d}x,$$

we deduce

$$|det(D_xT)| = \frac{f(x)}{g(T(x))}, \ \mu - a.e.\ x \in \mathbb{R}^n,$$

called the Monge-Ampère equation.

Several difficulties arise in solving the Monge problem.
First of all, transport maps may not exist.

Example 1. *We consider in $\mathbb{R}^n$ the two probability measures μ, ν given by*

$$\mu = \delta_x, \ \nu = \frac{1}{2}\delta_{y_1} + \frac{1}{2}\delta_{y_2}$$

where $x, y_1 \neq y_2 \in R^n$ and δ_a denotes the Dirac mass at point $a \in \mathbb{R}^n$. There are no transport maps from μ to ν. If such a map T exists, then

$$\frac{1}{2} = \nu(\{y_1\}) = \mu(T^{-1}(\{y_1\})) = 0 \text{ or } 1$$

which is impossible.

Secondly, minimizers of the Monge problem may not be unique.

Example 2. *Let μ,ν be two probability measures given by*

$$\mu = \chi_{[0,1]}\mathcal{L}^1, \quad \nu = \chi_{[1,2]}\mathcal{L}^1,$$

the restrictions of the Lebesgue measure $\mathcal{L}^1$ on the intervals $[0,1]$ and $[1,2]$ respectively. There are two maps $T_1(x) = x + 1$ and $T_2(x) = 2 - x$, pushing forward μ to ν for the cost function $c(x, y) := |x - y|, \ \forall x, y \in \mathbb{R}$.

The fact that the constraint $T_\sharp\mu = \nu$ is highly non linear with respect to T is the main difficulty to deal with the Monge problem. That is why, Kantorovich proposed a relaxed form of the problem.

2.2 The Kantorovich Problem

We denote by $P^i : M \times M \to M$ the projection map into the i-th component.

Definition 7 (**Transport plan**). *Let μ, ν be two probability measures on M. We denote by $\Pi(\mu,\nu)$ the set of probability measures α in the product space $M \times M$ with*

$$P^1_\sharp(\alpha) = \mu \quad \textit{and} \quad P^2_\sharp(\alpha) = \nu.$$

Any measure $\alpha \in \Pi(\mu,\nu)$ is called transport plan between μ and ν.

The set $\Pi(\mu,\nu)$ of transport plans between μ and ν is a convex set which can not be empty (it always contains the product $\mu \times \nu$). The property $P^1_\sharp(\alpha) = \mu$ means that the first marginal of α is equal to μ, i.e.

$$\text{for any } \mu\text{-measurable set } A \subseteq M, \text{we have } \alpha(A \times M) = \mu(A).$$

The definition $P^2_\sharp(\alpha) = \nu$ is similar with the second marginal of α, i.e.

$$\text{for any } \nu\text{-measurable set } B \subseteq M, \text{we have } \alpha(M \times B) = \nu(B).$$

This is also equivalent to have for every $(\varphi,\psi) \in L^1(\mu) \times L^1(\nu)$,

$$\int_{M\times M} \Big[\varphi(x) + \psi(y)\Big] \mathrm{d}\alpha(x,y) = \int_M \varphi(x)\mathrm{d}\mu(x) + \int_M \psi(y)\mathrm{d}\nu(y).$$

The Kantorovich problem consists in minimizing the transportation cost

$$\int_{M\times M} c(x,y)\mathrm{d}\alpha(x,y), \text{ among all transport plans } \alpha \in \Pi(\mu,\nu).$$

We notice that by considering a transport map $T : M \to M$ from μ and ν, we can define a transport plan $\alpha \in \Pi(\mu,\nu)$ as follows

$$\alpha := (Id \times T)_\sharp \mu.$$

We say that the transport plan α is induced by a transport map T. This shows that the Kantorovich problem is more general that the Monge problem: it is a relaxation form of the Monge problem.

Practically, the Monge problem consists in transporting each mass as it is, while the Kantorovich problem allows to separate the starting mass and send the different parts to different places. The difference between the Monge problem and the Kantorovich problem can be seen through the following example:

Example 3. *Returning to Example 1, consider the two probability measures on $\mathbb{R}^n$ given by*

$$\mu = \delta_x, \ \ \nu = \frac{1}{2}\delta_{y_1} + \frac{1}{2}\delta_{y_2}.$$

In contrary to transport maps, there is a transport plan α between μ and ν, solution of the Kantorovich problem, given by

$$\alpha = \frac{1}{2}\delta_{(x,y_1)} + \frac{1}{2}\delta_{(x,y_2)}.$$

The Kantorovich problem comes down to a linear minimization problem (with respect to α) on a set of constraints $\Pi(\mu, \nu)$ which is convex and weakly compact. The existence of optimal transport plans becomes easy.

We recall that the support of a measure μ, denoted by *supp* μ, refers to the smallest closed set $F \subset M$ of full mass $\mu(F) = \mu(M) = 1$.

Theorem 2 (Existence of optimal transport plans). *Let μ, ν be two probability measures compactly supported on M. Assume that the cost function $c : M \times M \to [0, +\infty[$ is continuous. Then, there is at least one optimal transport plan $\alpha \in \Pi(\mu, \nu)$ solving the Kantorovich problem.*

Proof of Theorem 2. We notice easily that the product $\mu \times \nu$ is a transport plan. Moreover, all the transport plans are concentrated on *supp* $\mu \times$ *supp* ν which is compact (because by assumption, *supp* μ , *supp* ν are compact). Without loss of generality, we can assume that M is compact. The existence of optimal transport plans is a consequence of the weak closure of $\Pi(\mu, \nu)$ together with the continuity of the cost function c. $\square$

We now introduce the concept of c-cyclically monotonicity.

Definition 8. ***(c-cyclically monotone)*** *A subset $X \subset M \times M$ is said to be c-cyclically monotone if for every $N \in \mathbb{N}$ and every $(x_1, y_1), \ldots, (x_N, y_N) \in X$ it holds*

$$\sum_{i=1}^{N} c(x_i, y_i) \leq \sum_{i=1}^{N} c(x_{i+1}, y_i)$$

with $x_{N+1} = x_1$.

The following proposition shows a specific property of optimal transport plans. We refer the reader to Theorem 3.2.5 in [Rif14] and Chapter 5 in [Vil08].

Proposition 2. *Let μ,ν be two probability measures compactly supported on M, and let the cost function $c : M \times M \to [0, +\infty[$ be continuous. Then, there is a c-cyclically monotone set $S \subset supp\ \mu \times supp\ \nu$ such that for any optimal transport plan $\alpha \in \Pi(\mu, \nu)$,*

$$supp\ \alpha \subset S.$$

Optimal conditions to establish the equality between the infimum of the Monge problem and the infimum of the Kantorovich have been proved by Pratelli [Pra07]: transport maps do exist as soon as the initial measure is assumed to be non-atomic. It can be seen through examples 1 and 3. Since transport plans can be approximated by plans coming from transport maps, it is predicted that the infimum of the Kantorovich problem coincides with the infimum of the Monge problem.

2.3 The Dual Problem

The dual problem is a basic concept in the optimal transport theory, considered as another face of the original Kantorovich problem.

In the textbook [Vil08], Villani explains the concept of Kantorovich duality in an informal way and illustrate how the Monge problem can be reformulated from an economic viewpoint. Consider a large consortium of bakeries and cafés, there is a company which has in charge of the transportation of productions, by buying bread at the bakeries and selling them to the cafés. The original Monge-Kantorovich problem starts with the notion of cost, while in the dual problem, the central notion is the price. Let us denote by $\varphi(x)$ the price at which bread is bought at bakery x and $\psi(y)$ the price at which it is sold at café y. So the transportation cost becomes $\psi(y) - \varphi(x)$ instead of the original cost $c(x, y)$. As to be competitive, the company needs to set up prices in such a way that

$$\psi(y) - \varphi(x) \leq c(x, y),\ \forall x, y$$

and the problem becomes to maximize the profits.

This approach leads to a dual formulation (see chapter 5 [Vil08]) given by

$$\inf_{\alpha \in \Pi(\mu,\nu)} \left\{ \int_{M \times M} c(x, y) \mathrm{d}\alpha(x, y) \right\} = \sup_{\substack{(\varphi, \psi) \in L^1(\mu) \times L^1(\nu) \\ \psi(y) - \varphi(x) \leq c(x,y)}} \left\{ \int_M \psi(y) \mathrm{d}\nu(y) - \int_M \varphi(x) \mathrm{d}\mu(x) \right\}. \tag{2.1}$$

This leads to find a pair of integrable functions (φ, ψ) optimal on the right-hand side, and a transport plan α optimal on the left-hand side. The pair of functions (φ, ψ) should satisfy $\psi(y) - \varphi(x) \leq c(x, y)$. Then, for a given y, $\psi(y)$ will be the infimum of $\varphi(x) + c(x, y)$ among all x. For a given x, $\varphi(x)$ will be the supremum of $\psi(y) - c(x, y)$ among all y. So it makes sense to describe a pair of integrable functions (φ, ψ) as follows

$$\varphi(x) = \sup_{y \in M} \Big\{ \psi(y) - c(x, y) \Big\}, \ \forall x \in M$$

and

$$\psi(y) = \inf_{x \in M} \Big\{ \varphi(x) + c(x, y) \Big\}, \ \forall y \in M.$$

We may now introduce the concept of c-convexity which turns out later to be an indispensable tool for existence of optimal transport maps.

Definition 9 (**c-convexity**). *We say that a function* $\varphi : M \to \mathbb{R} \cup \{+\infty\}$ *not identically* $+\infty$*, is c-convex if there exists a function* $\psi : M \to \mathbb{R} \cup \{\pm\infty\}$ *such that*

$$\varphi(x) = \sup_{y \in M} \left\{ \psi(y) - c(x, y) \right\}, \ \forall x \in M.$$

The c-transform of φ*, denoted by* φ^c*, is the function given by*

$$\varphi^c(y) = \inf_{x \in M} \left\{ \varphi(x) + c(x, y) \right\}, \ \forall y \in M.$$

Definition 10. *Let* $\varphi : M \to \mathbb{R} \cup \{+\infty\}$ *be a c-convex function. We call contact set of the pair* (φ, φ^c) *the set defined by*

$$\Gamma_\varphi := \Big\{ (x, y) \in M \times M | \ \varphi^c(y) - \varphi(x) = c(x, y) \Big\},$$

which is a closed convex set.

For every $x \in M$*, we define the c-subdifferential of* φ *at* x *by*

$$\Gamma_\varphi(x) := \Big\{ y \in M | (x, y) \in \Gamma_\varphi \Big\}.$$

We may indeed assume that the dual problem can be reduced to find a c-convex function $\varphi \in L^1(\mu)$ such that

$$\inf_{\alpha \in \Pi(\mu, \nu)} \left\{ \int_{M \times M} c(x, y) \mathrm{d}\alpha(x, y) \right\} = \int_M \varphi^c(y) \mathrm{d}\nu(y) - \int_M \varphi(x) \mathrm{d}\mu(x) \quad (2.2)$$

where, by definition of c-convex functions, the constraint of the dual problem $\varphi^c(y) - \varphi(x) \leq c(x, y)$ is satisfied.

The pair of functions (φ, φ^c) solution of the dual problem (2.2), is called the Kantorovich potentials.

We give here a characterization of the supports of optimal transport plans which are c-cyclically monotone sets (see Theorem 3.2.13 in [Rif14]).

Proposition 3. *Let $S \subset M \times M$ be a c-cyclically monotone compact set. Then, there is a c-concave function φ valued in $\mathbb{R}$, such that*

$$\varphi(x) = \sup_{y \in M} \Big\{ \varphi^c(y) - c(x, y) \Big\}, \ \forall x \in M$$

$$\varphi^c(y) = \inf_{x \in M} \Big\{ \varphi(x) + c(x, y) \Big\}, \ \forall y \in M$$

and

$$S \subset \Gamma_\varphi.$$

The fact that the infimum and supremum are attained is straigthforward from the continuity of φ, φ^c and the compactness of S.

Theorem 3. *Let μ,ν be two probability measures compactly supported on M, and let the cost function $c : M \times M \to [0, +\infty[$ be continuous. Let (φ, φ^c) be the Kantorovich potentials, solution of the dual problem (2.2). Then,*

any transport plan $\alpha \in \Pi(\mu, \nu)$ is optimal if and only if supp $\alpha \subseteq \Gamma_\varphi$.

We say that α is concentrated on Γ_φ.

Proof of Theorem 3. Let (φ, φ^c) be the Kantorovich potentials, solution of the dual problem and let $\alpha \in \Pi(\mu, \nu)$ be an optimal transport plan. So, we have

$$\int_{M \times M} c(x, y) \mathrm{d}\alpha(x, y) = \int_M \varphi^c(y) \mathrm{d}\nu(y) - \int_M \varphi(x) \mathrm{d}\mu(x).$$

Since $\alpha \in \Pi(\mu, \nu)$,

$$\int_{M \times M} c(x, y) \mathrm{d}\alpha(x, y) = \int_{M \times M} \big[\varphi^c(y) - \varphi(x)\big] \mathrm{d}\alpha(x, y)$$

$$\Rightarrow \int_{M\times M} \Big[c(x,y) - \varphi^c(y) + \varphi(x)\Big] \mathrm{d}\alpha(x,y) = 0.$$

As $c(x,y) \geq \varphi^c(y) - \varphi(x), \forall x, y \in M$, then

$$c(x,y) = \varphi^c(y) - \varphi(x) \text{ for almost every } (x,y) \in supp\ \alpha.$$

This shows that

$$supp\ \alpha \subseteq \Gamma_\alpha.$$

Reciprocally, let $\alpha \in \Pi(\mu,\nu)$ such that $supp\ \alpha \subseteq \Gamma_\varphi$. Then, for any transport plan $\beta \in \Pi(\mu,\nu)$, we have

$$\begin{aligned}\int_{M\times M} c(x,y)\mathrm{d}\alpha(x,y) &= \int_M \varphi^c(y)\mathrm{d}\nu(y) - \int_M \varphi(x)\mathrm{d}\mu(x)\\ &\leq \int_{M\times M} c(x,y)\mathrm{d}\beta(x,y)\end{aligned}$$

which implies that α is optimal. □

Theorem 4. *Let μ,ν be two probability measures compactly supported on M, and let the cost function $c : M\times M \to [0,+\infty[$ be continuous. Let (φ,φ^c) be the Kantorovich potentials, solution of the dual problem (2.2). Assume that for μ-a.e. $x \in M$, $\Gamma_\varphi(x)$ is a singleton. Then, there is a unique transport map $T : M \to M$ from μ to ν such that*

$$\Gamma_\varphi(x) = \big\{T(x)\big\},\ \ \mu - a.e.\ x \in M.$$

Proof of Theorem 4. By assumption, there is a Borel set N such that $\mu(N) = 0$ and for every $x \notin N$, there is $y_x \in M$ such that $\Gamma_\varphi(x) = \{y_x\}$. Hence, for every $x \in M\backslash N$, and every $y \in \Gamma_\varphi(x)$, we have $y = y_x$. We set $T(x) := y_x$ for μ-a.e. $x \in M$ and the conclusion follows. □

In other words, the problem of existence and uniqueness of optimal transport maps can be reduced to prove that Γ_φ is concentrated on a graph, that is to show that for μ-a.e. $x \in M$, the set $\Gamma_\varphi(x)$ is a singleton.

2.4 Previous results for the Monge problem

2.4.1 Euclidean case

Brenier [Br91] showed solutions for the Monge problem for the quadratic Euclidean cost.

Theorem 5 (**Brenier Theorem**). *Let $M = \mathbb{R}^n$. Let μ,ν be the two probability measures compactly supported on $\mathbb{R}^n$ such that μ is absolutely continuous with respect to the Lebesgue measure. Let $c : \mathbb{R}^n \times \mathbb{R}^n \to [0, +\infty[$ be the quadratic cost function given by*

$$c(x, y) := \frac{|x-y|^2}{2}, \forall (x, y) \in \mathbb{R}^n \times \mathbb{R}^n.$$

Then, there is a unique transport map $T : \mathbb{R}^n \to \mathbb{R}^n$. In fact, there is a convex function $\varphi \in L^1(\mathbb{R}^n)$ such that

$$T(x) = x + \nabla\varphi(x), \ \mu\text{-a.e. } x \in \mathbb{R}^n.$$

Proof of Theorem 5. We recall that, by the dual problem (2.2) and by Theorem 2, there is a convex function $\varphi \in L^1(\mathbb{R}^n)$ such that

$$\inf_{\alpha \in \Pi(\mu,\nu)} \left\{ \int_{M\times M} \frac{|x-y|^2}{2} \mathrm{d}\alpha(x, y) \right\} = \int_M \varphi^c(y)\mathrm{d}\nu(y) - \int_M \varphi(x)\mathrm{d}\mu(x).$$

Note that φ is convex on $\mathbb{R}^n$. Thanks to the Rademacher Theorem (see Appendix B.1), since μ is absolutely continuous with respect to the Lebesgue measure, φ is differentiable almost everywhere on $\mathbb{R}^n$.

Expanding $|x - y|^2/2$ into $|x|^2/2 + |y|^2/2 - x.y$ yields

$$x.y \leq \left[\varphi(x) + \frac{|x|^2}{2}\right] + \left[\frac{|y|^2}{2} - \varphi^c(y)\right], \forall x, y \in \mathbb{R}^n.$$

Let $(x, y) \in$ *supp* α be fixed. Thanks to Theorem 3, for any optimal plan $\alpha \in \Pi(\mu, \nu)$, we have *supp* $\alpha \subseteq \Gamma_\varphi$. Hence,

$$\begin{aligned}
(x, y) \in \Gamma_\varphi \ &\Leftrightarrow \ \varphi^c(y) - \varphi(x) = \frac{|x-y|^2}{2} \\
&\Leftrightarrow \ \varphi^c(y) = \varphi(x) + \frac{|x-y|^2}{2} \leq \varphi(z) + \frac{|z-y|^2}{2}, \ \forall z \in \mathbb{R}^n \\
&\Leftrightarrow \ \left[\varphi(x) + \frac{|x|^2}{2}\right] - x.y \leq \left[\varphi(z) + \frac{|z|^2}{2}\right] - z.y, \ \forall z \in \mathbb{R}^n
\end{aligned}$$

This means that the mapping $z \mapsto \varphi(z) + |z|^2/2 - z.y$ admits a minimum at x. Hence, its differential at x is equal to zero, that is

$$\nabla\varphi(x) + x - y = 0 \Rightarrow y = x + \nabla\varphi(x).$$

Thus, there is a unique transport map $T : M \to M$ such that

$$T(x) = x + \nabla\varphi(x), \ \mu\text{-a.e. } x \in \mathbb{R}^n.$$

$\square$

2.4.2 Riemannian case

McCann [Mc01] proved solutions for the Monge problem in the Riemannian case with the cost given by the quadratic geodesic distance. We start with some basic definitions and preliminaries in the Riemannian settings.

Let (M, g) be a Riemannian manifold where M is a real smooth manifold of dimension n equipped with the inner product g_x on the tangent space T_xM at each point $x \in M$.

We define the geodesic distance, denoted d_g, between two points x, y in M by:

$$d_g(x,y) := \inf\Big\{ l^g(\gamma) |\ \gamma : [0,1] \to M \ s.t.\ \gamma(0) = x, \gamma(1) = y \Big\}$$

where $l^g(\gamma)$ is the length of the curve γ given by

$$l^g(\gamma) := \int_0^1 \sqrt{g_{\gamma(t)}(\dot\gamma(t), \dot\gamma(t))}\ \mathrm{d}t.$$

For any $x \in M$, we define the exponential map at x by

$$\begin{array}{rccl} \exp_x : & T_xM & \longrightarrow & M \\ & v & \longmapsto & \exp_x(v) := \gamma(1) \end{array}$$

where $\gamma : [0,1] \to M$ is the unique minimizing geodesic with initial conditions $\gamma(0) = x$ and $\dot\gamma(0) = v$.

Assume that the Riemannian manifold (M, g) is equipped with the Levi-Civita connection and for every smooth curve $\gamma : [0,1] \to M$, we denote by $\nabla_{\dot\gamma}$ the associated covariant derivative along γ. We recall that γ is said to be a geodesic if and only if $\nabla_{\dot\gamma(t)}\dot\gamma(t)$ vanishes. Here, we study geodesics using the Lagrangian approach. It seems easier to apply the Euler-Lagrange equations than calculate the coefficients of the Levi-Civita connection.

Definition 11 (**Lagrangian action**)**.** *Let $L : TM \to \mathbb{R}$ be a Lagrangian function. The cost $c : M \times M \to [0, +\infty[$ associated to L is given by*

$$c(x,y) := \min\Big\{ \int_0^1 L(\gamma(t), \dot\gamma(t))\mathrm{d}t |\ \gamma : [0,1] \to M \\ \textit{with } \gamma(0) = x, \gamma(1) = y \Big\}, \forall x, y \in M.$$

The Lagrangian which associates $L(x,v) = g_x(v,v), \forall (x,v) \in TM$ leads to the quadratic geodesic cost $c = d_g^2$. Let $x, y \in M$ be fixed, the goal is to find a path $\gamma : [0,1] \to M$ from x to y that minimizes the functional

$$F(\gamma) := \int_0^1 L\big(\gamma(t), \dot{\gamma}(t)\big)\mathrm{d}t.$$

We recall the Euler-Lagrange equation.

Lemma 1. *Let x, y be two distinct points of M and let $\gamma : [0,1] \to M$ be a minimizing path such that $\gamma(0) = x$ and $\gamma(1) = y$. Then, it satisfies*

$$\nabla_{\gamma(0)} L = \frac{d}{dt}\Big(\nabla_{\dot{\gamma}(0)} L\Big)$$

Proof of Lemma 1. Let $\epsilon > 0$, we define

$$\gamma_\epsilon(t) = \gamma(t) + \varepsilon h(t), \forall t \in [0,1]$$

where $h : [0,1] \to M$ is a random funtion such that $h(0) = h(1) = 0$.

We set

$$F(\gamma_\varepsilon) := \int_0^1 L\big(\gamma_\varepsilon(t), \dot{\gamma_\varepsilon}(t)\big)\mathrm{d}t. \tag{2.3}$$

At the extremum $\epsilon = 0$, that is $\gamma_0(t) = \gamma(t), \forall t \in [0,1]$, we have

$$\frac{\partial F}{\partial \epsilon}|_{\epsilon=0} = 0.$$

Derivating (2.3) with respect to ε yields

$$\begin{aligned}
\frac{\partial F}{\partial \epsilon} &= \int_0^1 \frac{\partial}{\partial \epsilon} L\big(\gamma_\epsilon(t), \dot{\gamma}_\epsilon(t)\big)\mathrm{d}t \\
&= \int_0^1 (\nabla_{\gamma_\epsilon(t)} L \cdot \frac{\partial \gamma_\epsilon}{\partial \epsilon} + \nabla_{\dot{\gamma}_\epsilon(t)} L \cdot \frac{\partial \dot{\gamma}_\epsilon}{\partial \epsilon})\mathrm{d}t \\
&= \int_0^1 \big(\nabla_{\gamma_\epsilon(t)} L \cdot h(t) + \nabla_{\dot{\gamma}_\epsilon(t)} L \cdot \dot{h}(t)\big)\mathrm{d}t
\end{aligned}$$

By the integration by parts formula, we obtain:

$$\frac{\partial F}{\partial \epsilon} = \int_0^1 \nabla_{\gamma_\epsilon(t)} L \cdot h(t)\mathrm{d}t - \int_0^1 \frac{d}{dt}\big(\nabla_{\dot{\gamma}_\epsilon(t)} L\big) \cdot h(t)\mathrm{d}t + \Big[\nabla_{\dot{\gamma}_\epsilon(t)} L \cdot h(t)\Big]_0^1$$

The fact that $h(0) = h(1) = 0$ gives

$$\frac{\partial F}{\partial \epsilon} = \int_0^1 \left[\nabla_{\gamma_\epsilon(t)} L - \frac{d}{dt}\big(\nabla_{\dot{\gamma}_\epsilon(t)} L\big)\right] \cdot h(t)\mathrm{d}t$$

Since the derivative of the functional F with respect to ϵ is equal to zero at $\epsilon = 0$, we obtain

$$0 = \int_0^1 \left[\nabla_{\gamma_\epsilon(0)} L - \frac{d}{dt}\big(\nabla_{\dot{\gamma}_\epsilon(0)} L\big)\right] \cdot h(t)\mathrm{d}t$$

Since h is an arbitrary function, we get

$$\nabla_{\gamma_\epsilon(0)} L = \frac{d}{dt}\big(\nabla_{\dot{\gamma}_\epsilon(0)} L\big).$$

□

Another lemma is needed.

Lemma 2. *For every $(x, v) \in TM$, we have*

$$\frac{1}{2}\nabla_v L = v.$$

Proof of Lemma 2. Let $x \in M$. In a system of local coordinates on M given by n-real valued functions $x_1, \dots, x_n$, the vector fields $\partial_{x_1}, \dots, \partial_{x_n}$ form an orthonormal basis of T_xM. A vector $v \in T_xM$ is given by

$$v = \sum_{i=1}^{n} v^i \partial_{x_i}$$

and the components of the metric tensor at a point $x \in M$ are of the form

$$g_{ij}(x) = g_x(\partial_{x_i}, \partial_{x_j}).$$

Therefore, we obtain

$$\begin{aligned}
\frac{1}{2}\nabla_v L &= \frac{1}{2}\sum_{i=1}^{n} \frac{\partial L}{\partial v^i}(x,v)\ \partial_{x_i} \\
&= \frac{1}{2}\sum_{i=1}^{n} \frac{\partial g_x}{\partial v^i}(v,v)\ \partial_{x_i}(x) \\
&= \frac{1}{2}\sum_{i=1}^{n} \frac{\partial}{\partial v^i}\Big(g_{ij}(x)v^i(x)v^j(x)\Big)\ \partial_{x_i}(x) \\
&= \sum_{i=1}^{n} g_{ij}(x)\ v^j(x)\ \partial_{x_i}(x) \\
&= \sum_{i=1}^{n} v^i(x)\ \partial_{x_i}(x) \\
&= v.
\end{aligned}$$

$\square$

We state now the McCann Theorem.

Theorem 6 (**McCann Theorem**)**.** *Let (M,g) be a complete Riemannian manifold. Let μ,ν be two probability measures compactly supported on M such that μ is absolutely continuous with respect to the Lebesgue measure. Let $c : M \times M \to [0,+\infty[$ be the quadratic cost function given by*

$$c(x,y) := \frac{1}{2}d_g^2(x,y),\ \ \forall x,y \in M.$$

Then, there is a unique transport map $T : M \to M$ solution for the Monge problem. Moreover, there is a Lipschitz function $\varphi \in L^1(\mu)$ such that

$$T(x) = exp_x(-\nabla_x \varphi),\ \ \mu\text{-}a.e.\ x \in M.$$

Proof of Theorem 6. By the dual formulation (2.2) of the Monge problem and by Theorem 3, there is a c-convex function $\varphi \in L^1(\mu)$ such that for any optimal transport plan $\alpha \in \Pi(\mu,\nu)$, we have *supp* $\alpha \subseteq \Gamma_\varphi$. By the Kantorovich potentials definition, the function φ is the supremum of a family of functions $x \mapsto \varphi^c(y) - c(x,y)$ with $y \in M$, which are Lipschitz with respect to the variable x (because the Riemannian distance d_g is Lipschitz). Therefore, φ is Lipschitz on M. Since μ is absolutely continuous with respect to the Lebesgue measure, the Rademacher theorem (see Appendix B.1) implies that φ is differentiable almost everywhere on M.

Let $\alpha \in \Pi(\mu,\nu)$ be optimal and $(x,y) \in supp\ \alpha$ be fixed. Then,

$$\begin{aligned}(x,y)\in\Gamma_\varphi \Leftrightarrow\ & \varphi^c(y)-\varphi(x)=c(x,y)\\ &\Leftrightarrow \varphi^c(y)=\varphi(x)+c(x,y)\leq\varphi(z)+c(z,y),\ \forall z\in M\\ &\Leftrightarrow c(z,y)\geq\varphi(x)-\varphi(z)+c(x,y),\ \forall z\in M. \qquad (2.4)\end{aligned}$$

Define the function $\psi : M \longrightarrow \mathbb{R}$ by

$$\psi(z)=\varphi(x)-\varphi(z)+c(x,y),\ \forall z\in M.$$

Since ψ depends on z, it is differentiable at almost every $z \in M$. Moreover, using inequality (2.4), we get

$$\psi(z)\leq c(z,y),\forall z\in M,\ \text{and } c(x,y)=\psi(x). \qquad (2.5)$$

The following lemma is crucial to conclude.

Lemma 3. *Let $\bar{x} \neq \bar{y}$ in M and let $\psi : M \to \mathbb{R}$ be a differentiable function at $\bar{x}$ such that*

$$\psi(z)\leq\frac{1}{2}d_g^2(z,\bar{y}),\forall z\in M\ \textit{and equality for } z=\bar{x}.$$

Then, there is a unique minimizing geodesic $\gamma : [0,1] \to M$ between $\bar{x}$ and $\bar{y}$. Moreover, $\bar{y} = exp_{\bar{x}}(\nabla_{\bar{x}}\psi)$, where $exp_{\bar{x}}$ is the exponential map at $\bar{x}$.

Proof of Lemma 3. Let $\bar{x} \neq \bar{y}$ be fixed in M and $\gamma : [0,1] \to M$ be a path such that $\gamma(0) = \bar{x}$ and $\gamma(1) = \bar{y}$. By hypothesis, for $\varepsilon > 0$ given, and $h : [0,1] \to M$ such that $h(1) = 0$, we have

$$\psi\Big(\gamma(0)+\varepsilon h(0)\Big)\leq\frac{1}{2}d_g^2\Big(\gamma(0)+\varepsilon h(0),\gamma(1)+\varepsilon h(1)\Big). \qquad (2.6)$$

By the definition of the Lagrangian action, we have

$$\begin{aligned}\frac{1}{2}d_g^2&\Big(\gamma(0)+\varepsilon h(0),\gamma(1)+\varepsilon h(1)\Big)\\ &\leq\frac{1}{2}\int_0^1 g_{\gamma(t)}\Big(\dot\gamma(t)+\varepsilon\dot h(t),\dot\gamma(t)+\varepsilon\dot h(t)\Big)\mathrm{d}t\\ &\leq\frac{1}{2}\int_0^1 L\Big(\gamma(t)+\varepsilon h(t),\dot\gamma(t)+\varepsilon\dot h(t)\Big)\mathrm{d}t \qquad (2.7)\end{aligned}$$

Thus, by (2.6) and (2.7),

$$\frac{1}{2}\int_0^1 L\Big(\gamma(t)+\varepsilon h(t),\dot{\gamma}(t)+\varepsilon\dot{h}(t)\Big)\mathrm{d}t \;-\; \psi\Big(\gamma(0)+\varepsilon h(0)\Big)\geq 0. \tag{2.8}$$

The derivative of (2.8) with respect to ε yields

$$\frac{1}{2}\int_0^1 \nabla_{\gamma(0)}L\cdot h(t)\mathrm{d}t+\frac{1}{2}\Big[\nabla_{\dot{\gamma}(0)}L\cdot h(t)\Big]_0^1$$
$$-\frac{1}{2}\int_0^1\frac{d}{dt}\big(\nabla_{\dot{\gamma}(0)}L\big)\cdot h(t)\mathrm{d}t-\nabla_{\bar{x}}\psi\cdot h(0)=0$$

$$\Rightarrow \frac{1}{2}\int_0^1\Big[\nabla_{\gamma(0)}L-\frac{d}{dt}\big(\nabla_{\dot{\gamma}(0)}L\big)\Big]\cdot h(t)\mathrm{d}t$$
$$-\Big[\frac{1}{2}\nabla_{\dot{\gamma}(0)}L-\nabla_{\bar{x}}\psi\Big]\cdot h(0)=0.$$

Using the Euler-Lagrange equation (see Lemma 1) and by Lemma 2, we obtain

$$\nabla_{\bar{x}}\psi=\frac{1}{2}\nabla_{\dot{\gamma}(0)}L=\dot{\gamma}(0)$$

This implies that

$$\bar{y}=\exp_{\bar{x}}(\nabla_{\bar{x}}\psi).$$

□

Returning to (2.5) and thanks to the above Lemma, there is a unique optimal transport map $T:M\to M$ such that

$$T(x)=\exp_x(\nabla_x\psi)=\exp_x(-\nabla_x\varphi),\;\mu\text{-a.e. } x\in M.$$

□

We refer the reader to the result of Bernard and Buffoni [BB05] who proved existence of an optimal transport map, solution for the Monge problem when the cost is the action associated to a Lagrangian function on a compact manifold.

Chapter 3

Sub-Riemannian Geometry

3.1 Sub-Riemannian structure

3.1.1 Horizontal distribution

Let M be a smooth connected manifold without boundary of dimension $n \geq 2$. A vector field X on M is a smooth map from M into TM that assigns a vector $X(x)$ at the point $x \in M$. We denote by $\chi^\infty(M)$ the set of all smooth vector fields.

Definition 12 (Horizontal Distribution). *A smooth distribution Δ of rank $m \leq n$ ($m \geq 1$) on M is a rank m subbundle of the tangent bundle TM. In other terms, for each $x \in M$, we assign a m-dimensional linear subspace $\Delta(x)$ of the tangent space T_xM in such a way that for an open neighborhood $\mathcal{V}_x$ of x in M, there is m smooth vector fields $X_x^1, \dots, X_x^m$ linearly independent on $\mathcal{V}_x$ such that*

$$\Delta(y) = Span\big\{X_x^1(y), \dots, X_x^m(y)\big\}, \forall y \in \mathcal{V}_x.$$

Such family of smooth vector fields $\{X_x^1, \dots, X_x^m\}$ is called a local frame for the distribution Δ in $\mathcal{V}_x$.

Given a smooth vector field X on M, it is said to be "horizontal" with respect to Δ if it is a section of the distribution Δ, that is

$$\forall x \in M, \;\; X(x) \in \Delta(x).$$

A set of smooth vector fields $\{X^1, \dots, X^m\}$ is said to be a global generating family for the distribution Δ on M if

$$\forall x \in M, \Delta(x) = Span\big\{X^1(x), \dots, X^m(x)\big\}.$$

In [Sus08], Sussmann proved that we can always construct a global generating family for Δ (see also Proposition 1.1.8 [Rif14]).

Proposition 4. *Let Δ be a smooth distribution of rank m on the n-dimensional manifold M $(m \leq n)$. Then, there are $k = m(n+1)$ smooth vector fields $X^1, \dots, X^k$ such that*

$$\forall x \in M, \Delta(x) = Span\{X^1(x), \dots, X^k(x)\}.$$

Remark 1. *If $m = n$, then the distribution Δ will be tangent to the manifold M, that is $\Delta(x) = T_x M$, $\forall x \in M$.*

Example 4. ***(Heisenberg group in $\mathbb{R}^3$)*** *In $\mathbb{R}^3$ with coordinates (x, y, z), the distribution Δ given by*

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

with

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

is a distribution of rank 2 on $\mathbb{R}^3$.

Example 5. ***(Heisenberg group in $\mathbb{R}^{2n+1}$)*** *More generally, in $\mathbb{R}^{2n+1}$ with coordinates $x = (x_1, \dots, x_n, y_1, \dots, y_n, z)$, we consider the n smooth vector fields $X^1, \dots, X^n, Y^1, \dots, Y^n$ given by*

$$X^i = \partial_{x_i} - \frac{y_i}{2}\,\partial_z \text{ and } Y^i = \partial_{y_i} + \frac{x_i}{2}\,\partial_z, \forall i = 1, \dots, n.$$

The distribution Δ given by

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

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

3.1.2 Totally nonholonomic distribution

Definition 13. *Given two smooth vector fields X, Y on M, the Lie bracket assigns to X and Y a third vector field , denoted by $[X, Y]$, such that*

$$[X, Y](x) = DY(x)X(x) - DX(x)Y(x), \ \forall x \in M \tag{3.1}$$

where DX and DY are the Jacobian matrices of X and Y, respectively.

In local coordinates $x = (x_1, \ldots, x_n)$, for any smooth vector fields X, Y on M given by

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

with $a_i, b_i : M \to \mathbb{R}$ smooth functions, the Lie Bracket of X, Y defined by the formula above (3.1) is given by

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

where $c_i = \displaystyle\sum_{j=1}^{n} (\partial_{x_j} b_i)a_j - (\partial_{x_j} a_i)b_j, \quad \forall i = 1, \ldots, n.$

Let $\mathcal{O}$ be an open set in M. For any family $\{X^1, \ldots, X^m\}$ of smooth vector fields defined on $\mathcal{O}$, we denote by $Lie(X^1, \ldots, X^m)$ the Lie algebra of vector fields generated by $\{X^1, \ldots, X^m\}$. It is the smallest vector subspace of $\chi^\infty(M)$ containing $\{X^1, \ldots, X^m\}$ and that also satisfies

$$[X^i, Y] \in Lie(X^1, \ldots, X^m), \;\; \forall i = 1, \ldots, m, \forall Y \in Lie(X^1, \ldots, X^m).$$

It can be construct as follows. We denote by

$$Lie^1(X^1, \ldots, X^m) = Span\big\{X^1, \ldots, X^m\big\}$$

the space spanned by $\big\{X^1, \ldots, X^m\big\}$ in $\chi^\infty(M)$.
Then, for $k \geq 1$, we define recursively the spaces $Lie^{k+1}(X^1, \ldots, X^m)$ by

$$\begin{aligned} Lie^{k+1}(X^1, \ldots, X^m) = Span\Big\{ & Lie^k(X^1, \ldots, X^m) \cup \\ & \{[X^i, X] \; | i = 1, \ldots, m, \;\; X \in Lie^k(X^1, \ldots, X^m)\}\Big\}. \end{aligned}$$

This defines an increasing sequence of subspaces of $\chi^\infty(M)$ given by

$$Lie(X^1, \ldots, X^m) = \bigcup_{k \geq 1} Lie^k(X^1, \ldots, X^m).$$

In general, $Lie\big(X^1, \ldots, X^m\big)$ is an infinite-dimensional subspace in $\chi^\infty(M)$.

For any point $x \in \mathcal{O}$, $Lie(X^1, \ldots, X^m)(x) = \{X(x) | X \in Lie(X^1, \ldots, X^m)\}$. It follows that $Lie(X^1, \ldots, X^m)(x)$ is a linear subspace of $T_x M$, hence of finite dimension.

Definition 14 (Hörmander condition). *Consider m smooth vector fields $X^1, \ldots, X^m$ on an open subset $\mathcal{O}$ of M. We say that $X^1, \ldots, X^m$ satisfy the Hörmander condition if and only if*

$$Lie(X^1, \ldots, X^m)(x) = T_x M, \ \forall x \in \mathcal{O}.$$

Definition 15 (Totally nonholonomic distribution). *Any distribution Δ on M is called totally nonholonomic or bracket generating on M, if for each $x \in M$, there are an open neighborhood $\mathcal{V}_x$ of x in M and a local frame $\left\{X_x^1, \ldots, X_x^m\right\}$ on $\mathcal{V}_x$ which satisfies the Hörmander condition on $\mathcal{V}_x$.*

Example 6. *Consider the distribution given in example 4 by*

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

where

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

Since $[X, Y] = \partial_z$, it follows that X, Y and $[X, Y]$ are linearly independent at every point of $\mathbb{R}^3$. Hence, Δ is totally nonholonomic on $\mathbb{R}^3$.

Example 7. *Consider in $\mathbb{R}^3$ with coordinates (x, y, z) the distribution given by*

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

where

$$X = \partial_x \ \text{and} \ \ Y = \partial_y + zx\ \partial_z.$$

Computing the Lie brackets of X, Y, we get

$$[X, Y] = [\partial_x, \partial_y + zx\partial_z] = z\partial_z.$$

$$\big[X, [X, Y]\big] = [\partial_x, z\partial_z] = 0.$$

$$\big[Y, [X, Y]\big] = [\partial_y + zx\partial_z, z\partial_z] = 0.$$

The distribution Δ is totally nonholonomic on $\mathbb{R}^3 \backslash \{z = 0\}$.

Example 8. *Not all distributions are totally nonholonomic. Consider in $\mathbb{R}^4$ with coordinates (x, y, z, t) the distribution given by*

$$\Delta(x,y,z,t) = Span\Big\{X(x,y,z,t), Y(x,y,z,t)\Big\}, \forall (x,y,z,t) \in \mathbb{R}^4$$

where

$$X = \partial_x - y\partial_y + t\partial_t \text{ and } Y = \partial_y + z\partial_z - t\partial_t.$$

A computation shows

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

$$\big[X, [X, Y]\big] = \partial_y.$$

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

All the iterated brackets of X, Y provide either 0 or ∂_y, so the brackets of X, Y do not generate $\mathbb{R}^4$.

This definition does not depend of the choice of the local frame. It is a consequence of the following result whose proof is taken from Proposition 1.1.16 in [Rif14].

Proposition 5. *Let $\{X^1, \ldots, X^m\}$ and $\{Y^1, \ldots, Y^m\}$ be two family of m smooth vector fields on an open subset $\mathcal{O}$ of M such that*

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

Then, for every $x \in \mathcal{O}$,

$$Lie(X^1, \ldots, X^m)(x) = Lie(Y^1, \ldots, Y^m)(x).$$

Definition 16 (Sub-Riemannian structure). *Let M be a smooth connected manifold of dimension n. A sub-Riemannian structure on M is a pair (Δ, g) where Δ is a totally nonholonomic distribution of rank m ($m \leq n$) endowed with a smooth Riemannian metric g; that is for every $x \in M$, $g_x(.,.)$ is a scalar product on $\Delta(x)$.*

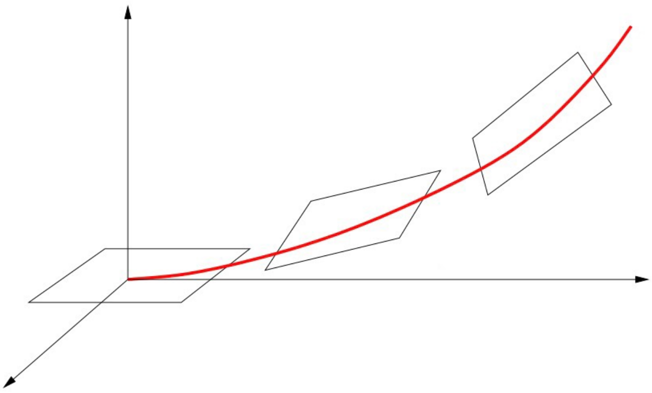


Figure 3.1 – Horizontal path

3.2 Horizontal path and End-Point mapping

Definition 17 (Horizontal path). *An absolutely continuous path $\gamma : [0,1] \to M$ is said to be horizontal with respect to Δ if its derivative is square-integrable on the interval $[0,1]$ and satisfies*

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

We recall now the Chow-Rashevsky Theorem ([Cho39], [Ras38]).

Theorem 7 (Chow-Rashevsky Theorem). *Let M be a smooth connected manifold equipped with a sub-Riemannian structure (Δ, g). Then, for every two points $x, y \in M$, there is an horizontal path joining x and y.*

Given $k = m(n+1)$, let $\left\{X^1, \dots, X^k\right\}$ be a global generating family for Δ on M. There is a correspondence between horizontal paths and an open subset of $L^2([0,1], \mathbb{R}^k)$.

Proposition 6. *Given a point $x \in M$, there exists an open subset $\mathcal{U}^x \subset L^2([0,1], \mathbb{R}^k)$ such that for every function $u \in \mathcal{U}^x$, the solution to the following Cauchy problem:*

$$\left\{ \begin{array}{rcl} \dot{\gamma_u}(t) & = & \displaystyle\sum_{i=1}^{k} u_i(t) X^i(\gamma_u(t)), \ \text{a.e. } t \in [0,1] \\ \gamma_u(0) & = & x \end{array} \right. \tag{3.2}$$

is well-defined on $[0,1]$.

The function u is called a control and the corresponding solution of the system (3.2) is called the trajectory starting at x and associated with the control u. Any horizontal path can be viewed as a trajectory associated to a control system like (3.2).

We refer the reader to the textbook [Rif14] for further details and proofs.

Definition 18 (**End-Point mapping**). *Given a point $x \in M$, the End-point mapping at x assigns to each curve starting at x its endpoint. It is defined by*

$$\begin{array}{rccl} End^x: & \mathcal{U}^x \subset L^2([0,1],\mathbb{R}^k) & \longrightarrow & M \\ & u & \longmapsto & \gamma_u(1) \end{array}$$

where $\gamma_u : [0,1] \to M$ is the unique solution to (3.2) associated to the control $u \in \mathcal{U}^x \subset L^2([0,1],\mathbb{R}^k)$.

Proposition 7. *Given a point $x \in M$, the End-point mapping End^x is of class C^1 on the open subset $\mathcal{U}^x \subset L^2([0,1],\mathbb{R}^k)$.*

Given $x \in M$ and an open subset $\mathcal{U}^x \subset L^2([0,1],\mathbb{R}^k)$. For every control $u \in \mathcal{U}^x$, we denote by

$$D_u\mathrm{End}^x : L^2([0,1],\mathbb{R}^k) \to T_{\mathrm{End}^x(u)}M$$

the differential of the End-point mapping End^x at u.

Remark 2. *If $M = \mathbb{R}^n$, the differential of End^x at u is given by*

$$D_u End^x(v) = S(1)\int_0^1 S(t)^{-1}B(t)v(t)\mathrm{d}t,$$

where $S : [0,1] \to M_n(\mathbb{R})$ is the solution to the Cauchy problem

$$\dot{S}(t) = A(t)S(t), \ \text{a.e. } t \in [0,1], \text{and } S(0) = I_n$$

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

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

with $\gamma_u(t)$ given by (3.2) and J_{X^i} the Jacobian matrix of X^i at $\gamma_u(t)$ and

$$B(t) = \big(X^1(\gamma_u(t)), \ldots, X^k(\gamma_u(t))\big).$$

We set

$$\mathrm{Im}^x(u) := D_u\mathrm{End}^x(L^2([0,1],\mathbb{R}^k)) \tag{3.3}$$

which is a vector space contained in $T_{\mathrm{End}^x(u)}M$, hence of dimension smaller than or equal to n.

The following proposition shows that the dimension of $\mathrm{Im}^x(u)$ is larger or equal to k (k is the dimension of the global frame generating Δ.)

Proposition 8. *Given $x \in M$ and an open subset $\mathcal{U}^x$ of $L^2([0,1],\mathbb{R}^k)$. For every $u \in \mathcal{U}^x$, we have*

$$X^i(End^x(u)) \in Im^x(u),\ \forall i = 1,\dots,k.$$

3.3 Regular and singular horizontal paths

Definition 19. *Given $x \in M$, we say that a control u is singular with respect to x if and only if it is a critical point of the End-point mapping End^x, that is, if End^x is not a submersion at u.*
Otherwise, we shall say that u is regular.

Definition 20. *A horizontal path $\gamma : [0,1] \to M$ is said to be singular (resp. regular) if and only if any control u associated to γ (i.e. $\gamma = \gamma_u$) is singular (resp. regular).*

The property of being singular does not depend upon the choice of the frame $\{X^1,\dots,X^m\}$ of the distribution.

Singular controls can be characterized as follows (see section 1.3 in [Rif14]).

Proposition 9. *In local coordinates, a control $u \in L^2([0,1],\mathbb{R}^k)$ is singular if and only if there exists an absolutely continuous arc $p : [0,1] \to (\mathbb{R}^n)^*\backslash\{0\}$ satisfying*

$$\left\{\begin{array}{rcll} \dot{p}(t) & = & -\sum_{i=1}^{k} u_i(t)p(t)\cdot D_{\gamma_u(t)}X^i, & a.e.\ t \in [0,1] \\ p(t)\cdot X^i(\gamma_u(t)) & = & 0 & \forall t \in [0,1],\ \forall i = 1,\dots,k \end{array}\right. \tag{3.4}$$

Proof of Proposition 9. Given $x \in M$, let u be a singular control with respect to x. There exists an open subset $\mathcal{U}^x \subset L^2([0,1],\mathbb{R}^k)$ such that End^x is not a submersion at u. It means that the differential $D_u\text{End}^x$ of the End-point mapping at u is not surjective. So there exists $p \in (\mathbb{R}^n)^* \backslash \{ 0\}$ such that

$$p \cdot D_u\text{End}^x(v) = 0, \ \forall v \in L^2([0,1], \mathbb{R}^k). \tag{3.5}$$

By remark 2, the identity (3.5) can be written as

$$\int_0^1 p \cdot S(1)S(t)^{-1}B(t)v(t)\mathrm{d}t = 0, \forall v \in L^2([0,1], \mathbb{R}^k).$$

By choosing $v \in L^2([0,1], \mathbb{R}^k)$ defined as

$$v(t) = \Big(p \cdot S(1)S(t)^{-1}B(t)\Big)^*, \forall t \in [0,1],$$

we obtain

$$\int_0^1 | \big(p \cdot S(1)S(t)^{-1}B(t)\big)^* |^2 \mathrm{d}t = 0.$$

Note that $t \mapsto p \cdot S(1)S(t)^{-1}B(t)$ is continuous, then

$$p \cdot S(1)S(t)^{-1}B(t) = 0, \forall t \in [0,1].$$

We define

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

By construction, $p : [0,1] \to (\mathbb{R}^n)^*$ is an absolutely continuous arc such that $p(t)$ does not vanish on $[0,1]$ (because $p \neq 0$ and $S(t)$ is invertible for any $t \in [0,1]$). Moreover, by the definition of $S(t)$ (see remark 2), we have

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

Recalling the definition of A and B in remark 2, it follows that p satisfies

$$\begin{cases} \dot{p}(t) = -p(t)A(t) & a.e. \ t \in [0,1] \\ p(t)B(t) = 0 & \forall t \in [0,1] \end{cases}.$$

Conversely, we assume that there is an absolutely continuous arc $p : [0,1] \to (\mathbb{R}^n)^* \backslash \{0\}$ satisfying (3.4). It implies

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

and

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

We put $p := p(1) \neq 0$, and we get

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

Hence,

$$p.S(1)S(t)^{-1}B(t) = 0$$

which implies

$$p.D_u\mathrm{End}^x(v) = 0, \forall v \in L^2([0,1], \mathbb{R}^k).$$

$\square$

Remark 3. *If $k = n$ (Riemannian case), then any non-trivial path is horizontal and regular. In fact, given any $x \in M$ and an open subset $\mathcal{U}^x$ of $L^2([0,1], \mathbb{R}^k)$, it means that for any $u \in \mathcal{U}^x$, $Im^x(u)$ is of dimension n (see (3.3) and Proposition 8). This implies that the End-Point map End^x is a submersion.*

Example 9. ***(The Heisenberg group in $\mathbb{R}^3$)***
Returning to examples 4 and 6, consider the totally nonholonomic distribution Δ in $\mathbb{R}^3$ given by

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

where

$$X^1 = \partial_x - \frac{y}{2}\,\partial_z \; and \;\; X^2 = \partial_y + \frac{x}{2}\,\partial_z.$$

We claim that there are no non-trivial singular paths. Let $x \in \mathbb{R}^3$ and $u = (u_1, u_2) \in L^2([0,1], \mathbb{R}^2)$ be a singular control. We denote by $\gamma : [0,1] \to \mathbb{R}^3$ the horizontal path associated to u such that

$$\dot{\gamma}(t) = u_1(t)X^1\big(\gamma(t)\big) + u_2(t)X^2\big(\gamma(t)\big), a.e.\ t \in [0,1], \;\; and\ \gamma(0) = x. \tag{3.6}$$

By Proposition 9, there exists an absolutely continuous arc $p : [0,1] \to (\mathbb{R}^3)^* \backslash \{0\}$ satisfying

$$\dot{p}(t) = -u_1(t)p(t) \cdot D_{\gamma(t)}X^1 - u_2(t)p(t) \cdot D_{\gamma(t)}X^2, \;\; a.e.\ t \in [0,1] \tag{3.7}$$

and

$$p(t) \cdot X^1(\gamma(t)) = p(t) \cdot X^2(\gamma(t)) = 0, \;\; \forall t \in [0,1]. \tag{3.8}$$

Derivating (3.8) yields

$$\dot{p}(t)\cdot X^i(\gamma(t)) + p(t)\cdot D_{\gamma(t)}X^i(\dot{\gamma}(t)) = 0.$$

By (3.6) and (3.7), we get for a.e $t \in [0,1]$

$$\begin{aligned} &- u_1(t)p(t)\cdot D_{\gamma(t)}X^1X^i(\gamma(t)) - u_2(t)p(t)\cdot D_{\gamma(t)}X^2X^i(\gamma(t)) \\ &\qquad + p(t)\cdot D_{\gamma(t)}X^i\Big(u_1(t)X^1(\gamma(t)) + u_2(t)X^2(\gamma(t))\Big) = 0 \end{aligned}$$

$$\begin{aligned} &\Rightarrow u_1(t)p(t)\cdot\Big(D_{\gamma(t)}X^1X^i(\gamma(t)) - D_{\gamma(t)}X^iX^1(\gamma(t))\Big) \\ &\qquad + u_2(t)p(t)\cdot\Big(D_{\gamma(t)}X^2X^i(\gamma(t)) - D_{\gamma(t)}X^iX^2(\gamma(t))\Big) = 0 \end{aligned}$$

$$\Rightarrow u_1(t)p(t)\cdot[X^i,X^1](\gamma(t)) + u_2(t)p(t)\cdot[X^i,X^2](\gamma(t)) = 0.$$

Taking $i=1$ *and* $i=2$, *we obtain*

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

$$\Rightarrow |u(t)|^2\Big(p(t)\cdot[X^1,X^2](\gamma(t))\Big)^2 = 0,\ \ a.e.\ t\in[0,1].$$

Note that $[X^1,X^2] = \partial_z$. *Since* X^1, X^2 *and* $[X^1,X^2]$ *generate* $\mathbb{R}^3$ *and,* $p(t) \not\equiv 0, \forall t \in [0,1]$, *it follows that* $u \equiv 0$.

Example 10. ***(The Martinet distribution in*** $\mathbb{R}^3$***)***
Consider in $\mathbb{R}^3$ *with coordinates* (x,y,z), *the distribution* Δ *given by*

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

with

$$X^1 = \partial_x,\ \ and\ X^2 = \partial_y + \frac{x^2}{2}\partial_z.$$

The first Lie bracket of X, Y *gives* $[X^1, X^2] = x\partial_z$. *Then, the distribution* Δ *is totally nonholonomic on* $\mathbb{R}^3 \backslash \{\ x = 0\}$. *We claim that non-trivial singular paths are contained in the Martinet surface given by*

$$\Sigma_\Delta := \Big\{(x, y, z) \in \mathbb{R}^3 |\ x = 0\Big\}.$$

Let $\bar{x} = (x, y, z) \in \mathbb{R}^3$ *and* $u = (u_1, u_2) \in L^2([0, 1], \mathbb{R}^2)$ *be a control. We denote by* $\gamma : [0, 1] \to \mathbb{R}^3$ *the horizontal path associated to* u *given by*

$$\gamma(t) = \big(x(t), y(t), z(t)\big), \forall t \in [0, 1]$$

such that

$$\dot{\gamma}(t) = u_1(t)X^1\big(\gamma(t)\big) + u_2(t)X^2\big(\gamma(t)\big), a.e.\ t \in [0, 1],\ \text{ and } \gamma(0) = \bar{x}. \tag{3.9}$$

Assume that u *is singular. By Proposition 9, there exists an absolutely continuous arc* $p : [0, 1] \to (\mathbb{R}^3)^* \backslash \{0\}$ *satisfying*

$$\dot{p}(t) = -u_1(t)p(t).D_{\gamma(t)}X^1 - u_2(t)p(t).D_{\gamma(t)}X^2,\ a.e.\ t \in [0, 1] \tag{3.10}$$

and

$$p(t) \cdot X^1(\gamma(t)) = p(t) \cdot X^2(\gamma(t)) = 0,\ \forall t \in [0, 1]. \tag{3.11}$$

By derivating (3.11) and by (3.9) and (3.10), we obtain

$$|u(t)|^2 \Big(p(t).[X^1, X^2](\gamma(t))\Big)^2 = 0,\ a.e.\ t \in [0, 1].$$

Since $u(t) \not\equiv 0$ *and* $p(t) \not\equiv 0$, *for a.e.* $t \in [0, 1]$, *we deduce that*

$$x(t) = 0, \forall t \in [0, 1].$$

Hence, for every $t \in [0, 1], \gamma(t) \in \Sigma_\Delta$.

3.4 Minimizing geodesics

Given two points $x, y \in M$, we introduce the sub-Riemannian distance between x and y as the infimum of the length of horizontal curves joining them.

Definition 21 (**Sub-Riemannian distance**). *Let* x, y *be two points on* M. *The sub-Riemannian distance between* x *and* y *is defined by*

$$d_{SR}(x, y) := \inf \Big\{l(\gamma) |\ \gamma : [0, 1] \to M \text{ horizontal curve}, \gamma(0) = x, \gamma(1) = y\Big\}$$

where the length of an horizontal path $\gamma : [0,1] \to M$ is given by

$$l(\gamma) := \int_0^1 \sqrt{g_{\gamma(t)}(\dot{\gamma}(t), \dot{\gamma}(t))} \mathrm{d}t.$$

We also call d_{SR} the Carnot-Carathéodory distance of the sub-Riemannian structure.

An horizontal path $\gamma : [0,1] \to M$ between x and y is said to be minimizing if it minimizes the sub-Riemannian distance between x and y.

We introduce the sub-Riemannian energy between x and y by

$$e_{SR}(x,y) := \inf \Big\{ e(\gamma) | \ \gamma : [0,1] \to M \text{ horizontal path } \gamma(0) = x, \gamma(1) = y \Big\}$$

where the energy of the curve γ is given by

$$e(\gamma) := \int_0^1 g_{\gamma(t)}(\dot{\gamma}(t), \dot{\gamma}(t)) \mathrm{d}t.$$

Like in the Riemannian case, an horizontal path between x and y is called minimizing geodesic if it minimizes the sub-Riemannian energy between x and y.

The following result whose proof is based on the Cauchy-Schwartz inequality (see Proposition 2.1.1 in [Rif14]), is fundamental.

Proposition 10. *For any $x, y \in M$, $e_{SR}(x,y) = d^2_{SR}(x,y)$.*

We recall that we have equality in the Cauchy-Schwartz inequality if and only if γ has constant speed. Thanks to the proposition 10, we obtain the following result.

Proposition 11. *Let x, y be two points in M. A path γ between x and y is a minimizing geodesic if and only if it is an horizontal path minimizing the sub-Riemannian distance between x and y with constant speed.*

We introduce now the sub-Riemannian version of the classical Riemannian Hopf-Rinow Theorem (see Theorem 2.1.5 in [Rif14] and Theorem 7.1 in [Stri86], see also [HR31], [C-V59] for the classical Theorem).

Theorem 8 (**Sub-Riemannian Hopf-Rinow Theorem**)**.** *Assume that the metric space (M, d_{SR}) is complete. Then, there is at least a minimizing geodesic between any pair of points in M.*

3.5 Normal and abnormal extremals

Throughout all this section, we will assume that (M, d_{SR}) is a complete metric space.

Let $\bar{x}, \bar{y} \in M$ and $\bar{\gamma} : [0,1] \to M$ be a minimizing geodesic joining $\bar{x}$ and $\bar{y}$ be fixed. Since $\bar{\gamma}$ minimizes the distance between $\bar{x}$ and $\bar{y}$, it can not have self-intersection. Hence, there are an open neighborhood $\mathcal{V}$ of $\bar{\gamma}\big([0,1]\big)$ in M and an orthonormal family of m smooth vector fields $X^1, \ldots, X^m$ such that

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

Moreover, by Proposition 6, there is an open subset $\mathcal{U}^x \subset L^2([0,1], \mathbb{R}^m)$ and a control $\bar{u} \in \mathcal{U}^x$ such that

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

Since $\bar{\gamma}$ is a minimizing geodesic between $\bar{x}$ and $\bar{y}$, it minimizes the energy among all horizontal paths joining $\bar{x}$ to $\bar{y}$. This means that $\bar{\gamma}$ minimizes the following quantity

$$\int_0^1 g_{\gamma_u(t)}\big(\dot{\gamma}_u(t), \dot{\gamma}_u(t)\big) \mathrm{d}t = \int_0^1 g_{\gamma_u(t)}\big(\sum_{i=1}^{m} u_i(t) X^i(\gamma_u(t)), \sum_{i=1}^{m} u_i(t) X^i(\gamma_u(t))\big) \mathrm{d}t$$
$$= \int_0^1 \sum_{i=1}^{m} \big(u_i(t)\big)^2 \mathrm{d}t,$$

among all controls $u \in \mathcal{U}^x \subset L^2([0,1], \mathbb{R}^m)$.

Considering the End-point mapping $\mathrm{End}^{\bar{x}}$ given by

$$\begin{array}{rrcl} \mathrm{End}^{\bar{x}} : & \mathcal{U}^{\bar{x}} \subset L^2([0,1], \mathbb{R}^m) & \longrightarrow & M \\ & u & \longmapsto & \mathrm{End}^{\bar{x}}(u) = \gamma_u(1) \end{array}$$

and setting

$$C(u) := \int_0^1 \sum_{i=1}^{m} \big(u_i(t)\big)^2 \mathrm{d}t,$$

we note that $\bar{u}$ is a solution of the following optimization problem:

$$\bar{u} \text{ minimizes } C(u) \text{ among all } u \in \mathcal{U}^{\bar{x}} \text{ with } \mathrm{End}^{\bar{x}}(u) = \bar{y}.$$

Thanks to the Lagrange Multiplier Theorem (see Theorem B.1.5 in [Rif14]), there exist $\lambda_0 \in \{0,1\}$ and $p \in T^*_{\bar{y}}M$ with $(\lambda_0, p) \neq (0,0)$ such that

$$p \cdot D_{\bar{u}}\mathrm{End}^{\bar{x}} = \lambda_0 D_{\bar{u}} C. \tag{3.12}$$

3.5.1 Case $\lambda_0 = 0$

Minimizers that arise with $\lambda_0 = 0$ correspond to singular minimizers. In fact, when $\lambda_0 = 0$, the Lagrange Multiplier equation is reduced to

$$p \cdot D_{\bar{u}}\mathrm{End}^{\bar{x}}(v) = 0, \forall v \in L^2([0,1], \mathbb{R}^m). \tag{3.13}$$

This means that some linear form $p \neq 0$ annihilates the image of the differential of $\mathrm{End}^{\bar{x}}$. Then, $\bar{u}$ is a critical of $\mathrm{End}^{\bar{x}}$ and equivalently, the curve $\bar{\gamma}$ associated to $\bar{u}$ is a singular minimizing geodesic.

Proposition 12. *The fact that $\lambda_0 = 0$ yields the existence of an absolutely continuous arc $p : [0,1] \to (\mathbb{R}^n)^* \setminus \{0\}$ with $p(1) = p$ which satisfies*

$$\begin{cases} \dot{p}(t) & = & -\displaystyle\sum_{i=1}^{m} \bar{u}_i(t) p(t) \cdot D_{\gamma(t)} X^i & , a.e.\ t \in [0,1] \\ p(t).X^i(\gamma(t)) & = & 0 & , \forall t \in [0,1] \end{cases} \tag{3.14}$$

Such curve $\psi : [0,1] \to T^*M$ given by $\psi(t) = \big(\bar{\gamma}(t), p(t)\big), \forall t \in [0,1]$, with $\psi(1) = (\bar{y}, p)$ is an abnormal extremal lift of $\bar{\gamma}$.

Definition 22. *We say that a $\gamma : [0,1] \to M$ is a normal minimizing geodesic if it admits a normal extremal lift.*

3.5.2 Case $\lambda_0 = 1$

Definition 23. *The sub-Riemannian Hamiltonian is a function on T^*M given by*

$$\begin{aligned} H : \ T^*M & \longrightarrow \mathbb{R} \\ (x,p) & \longmapsto H(x,p) := \max_{u \in \mathbb{R}^m} \Big(\sum_{i=1}^{m} u_i \ p \cdot X^i(x) - \frac{1}{2} \sum_{i=1}^{m} u_i^2 \Big). \end{aligned} \tag{3.15}$$

Moreover, the Hamiltonian H can be written as follows

$$H(x,p) := \frac{1}{2} \sum_{i=1}^{m} \big(p \cdot X^i(x)\big)^2. \tag{3.16}$$

In fact, differentiating (3.15) with respect to u_i yields

$$p \cdot X^i(x) - u_i = 0, \forall i = 1, \dots, k,$$

that is, the Hamiltonian defined in (3.15) attains its maximum for $p \cdot X^i(x) = u_i$, from which formula (3.16) is obtained.

Proposition 13. *The fact that $\lambda_0 = 1$ yields, in the local coordinates, the existence of a smooth arc $p : [0,1] \to T^*_{\bar{y}}M$ with $p(1) = p/2$ such that*

$$\begin{cases} \dot{\bar{\gamma}}(t) & = & \dfrac{\partial H}{\partial p}(\bar{\gamma}(t), p(t)) & = & \displaystyle\sum_{i=1}^{m} \big(p(t) \cdot X^i(\bar{\gamma}(t))\big) X^i(\bar{\gamma}(t)) \\ \dot{p}(t) & = & -\dfrac{\partial H}{\partial x}(\bar{\gamma}(t), p(t)) & = & -\displaystyle\sum_{i=1}^{m} \big(p(t) \cdot X^i(\bar{\gamma}(t))\big)\big(p(t) \cdot D_{\bar{\gamma}(t)} X^i\big) \end{cases} \tag{3.17}$$

with

$$\bar{u}_i(t) = p(t) \cdot X^i(\bar{\gamma}(t)), \forall t \in [0,1], \forall i = 1, \ldots, m.$$

Proof of Proposition 13. Thanks to the Lagrange Multiplier Theorem with $\lambda_0 = 1$ (see (3.12)), there exists $p \in T^*_{\bar{y}}M$ satisfying

$$p \cdot D_{\bar{u}}\mathrm{End}^{\bar{x}}(v) = D_{\bar{u}}C(v), \ \forall v \in L^2([0,1], \mathbb{R}^m) \tag{3.18}$$

where $C(v) = \displaystyle\int_0^1 \sum_{i=1}^{m} (v_i(t))^2 \mathrm{d}t$, and the differential of C at $\bar{u}$ is given by

$$D_{\bar{u}}C(v) = 2 < \bar{u}, v >_{L^2([0,1],\mathbb{R}^m)}, \forall v \in L^2([0,1], \mathbb{R}^m).$$

By remark 2, the differential of $End^{\bar{x}}$ at $\bar{u}$ is given by

$$D_{\bar{u}}End^{\bar{x}}(v) = S(1) \int_0^1 S(t)^{-1} B(t) v(t) \mathrm{d}t, \forall v \in L^2([0,1], \mathbb{R}^m),$$

where A, B, S were defined in remark 2. Hence, (3.18) yields

$$\int_0^1 \big[p \cdot S(1) S(t)^{-1} B(t) - 2\bar{u}(t)\big] v(t) \mathrm{d}t = 0, \forall v \in L^2([0,1], \mathbb{R}^m)$$

which implies

$$\bar{u}(t) = \frac{1}{2}\big[p \cdot S(1) S(t)^{-1} B(t)\big], a.e.\ t \in [0,1].$$

We define $p : [0,1] \to T^*_{\bar{y}}M$ by

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

Then, by construction, we have

$$\bar{u}(t) = p(t)B(t), a.e.\ t \in [0,1]$$

with $B(t) = \Big(X^1(\bar{\gamma}(t)), \ldots, X^m(\bar{\gamma}(t))\Big)$.

Moreover,

$$\dot{p}(t) = -p(t)A(t), \forall t \in [0,1]$$

with $A(t) = \displaystyle\sum_{i=1}^{m} \bar{u}_i(t) D_{\bar{\gamma}(t)} X^i$.

$\square$

A curve ψ given by $\psi(t) = \big(\bar{\gamma}(t), p(t)\big), \forall t \in [0,1]$, which is solution of the Hamiltonian system

$$\dot{\psi}(t) = \Big(\frac{\partial H}{\partial p}(\psi(t)), -\frac{\partial H}{\partial x}(\psi(t))\Big), \forall t \in [0,1] \text{ with } \psi(1) = \big(\bar{\gamma}(1), p/2\big) \tag{3.19}$$

is a normal extremal lift of $\bar{\gamma}$.

Definition 24. *We say that γ is a normal minimizing geodesic if it admits a normal extremal lift.*

Proposition 14. *Let $\gamma : [0,1] \to M$ be a minimizing geodesic joining two points of M. Then, γ verifies one of the two following properties:*

1. *γ is singular*
2. *γ is normal*

Note that γ can be singular and normal at the same time.

Chapter 4

Optimal transport problem on sub-Riemannian structures

4.1 Statement of the problem

Let M be a smooth connected manifold without boundary of dimension $n \geq 2$. Let (Δ, g) be a complete sub-Riemannian structure on M of rank m $(m < n)$. We will be concerned with the study of the Monge problem for the quadratic geodesic sub-Riemannian cost.

Monge quadratic sub-Riemannian Formulation

Let μ,ν be two probability measures compactly supported on M.

Minimize

$$T \longmapsto \int_M c(x, T(x)) \mathrm{d}\mu(x)$$

over all transport maps $T : M \to M$ from μ to ν,

where

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

The following result is taken from problem (2.2) and Theorem 3 in Chapter 2.

Theorem 9. *Let μ,ν be two probability measures compactly supported on M, and the cost function $c : M \times M \to [0, +\infty[$ be continuous. Then, there is a c-convex function $\varphi : M \to \mathbb{R}$ such that*

$$\varphi(x) := \sup_{y \in M} \left\{ \varphi^c(y) - c(x,y) \right\}, \ \forall x \in M,$$

$$\varphi^c(y) := \inf_{x \in M} \left\{ \varphi(x) + c(x,y) \right\}, \ \forall y \in M,$$

and

$$\forall \alpha \in \Pi(\mu,\nu), \alpha \text{ is optimal } \Leftrightarrow \alpha(\Gamma_\varphi) = 1$$

where

$$\Gamma_\varphi := \Big\{ (x,y) \in M \times M; \ \varphi^c(y) - \varphi(x) = c(x,y) \Big\}.$$

Unlike the Riemannian case, the quadratic sub-Riemannian cost is not locally Lipschitz. Here appears the main difficulty in solving the Monge quadratic sub-Riemannian problem. Under regularity properties for d_{SR}, Figalli and Rifford [FR10] generalize the Brenier-McCann theorem ([Br91], [Mc01]) proving existence and uniqueness of an optimal transport map.

4.2 The sub-Riemannian version of the Brenier-McCann Theorem

The main issue in the Brenier-McCann result is the regularity of the c-convex function φ provided by Theorem 9. In particular, the regularity properties of φ are consequences of regularity assumptions made on the cost function. The method developed by Figalli and Rifford [FR10], for the sub-Riemannian case, requires local semiconcavity property for the sub-Riemannian distance outside the diagonal. We will see later that this regularity property made on d_{SR} holds as soon as there is no singular minimizing geodesic joining two distinct points in M. On the diagonal, the existence of a unique optimal transport map is a consequence of a Pansu-Rademacher Theorem [MS01] (see Appendix B.2) without any assumption on the sub-Riemannian distance.

In the sequel, we denote by D the diagonal of $M \times M$, that is the set of all pairs of the form (x, x) with $x \in M$.

The above discussion motivates the following definition.

Definition 25. *Let $\varphi : M \to \mathbb{R}$ be the c-convex function provided by Theorem 9. We define the "static" set S^φ and "moving" set $\mathcal{M}^\varphi$ as follows*

$$S^\varphi := \{x \in M |\ x \in \Gamma_\varphi(x)\},$$

$$\mathcal{M}^\varphi := \{x \in M |\ x \notin \Gamma_\varphi(x)\}.$$

Note that we can easily check that $\mathcal{M}^\varphi$ coincides with the set

$$\{x \in M |\ \varphi(x) \neq \varphi^c(x)\ \} = \{x \in M |\ \varphi(x) > \varphi^c(x)\ \}$$

which is open by the continuity of φ and φ^c.

We state now the result of Figalli and Rifford [FR10].

Theorem 10. *Let μ,ν be two probability measures compactly supported on M such that μ is absolutely continuous with respect to the Lebesgue measure. Let $\varphi : M \to \mathbb{R}$ be the c-convex function provided by Theorem 9. Assume that the cost function d^2_{SR} is locally semiconcave on $(M \times M)\backslash D$.*
Then, there is a unique optimal transport map $T : M \to M$ from μ to ν such that for μ-almost every $x \in M$,

$$T(x) = \begin{cases} exp_x\left(\frac{1}{2}d_x\varphi\right) & , \quad x \in \mathcal{M}^\varphi \\ x & , \quad x \in S^\varphi \end{cases}.$$

Proof of Theorem 10. 1. In a first step, we prove that

$$\mu\text{-a.e. } x \in S^\varphi, \Gamma_\varphi(x) = \{x\}.$$

It is sufficient to prove the result for x contained in an open set $\mathcal{V} \subseteq M$ such that there is an orthonormal family of m vector fields $X^1, \dots, X^m$ generating $\Delta(z)$, $\forall z \in \mathcal{V}$. Let $x \in S^\varphi$ be fixed. By a change of coordinates if necessary, we can write the vector fields as follows

$$X^i = \frac{\partial}{\partial x_i} + \sum_{j=1}^{n} a_{ij}\frac{\partial}{\partial x_j},\ \forall i = 1, \dots, m.$$

We remark that the function $z \in M \mapsto \varphi^c(y) - d^2_{SR}(z, y)$ is locally Lipschitz with respect to the sub-Riemannian distance when y varies on a compact set. Then, φ is also locally Lipschitz with respect to the sub-Riemannian distance. By the Pansu-Rademacher theorem (see Appendix B.2), since μ is absolutely continuous with respect to the Lebesgue measure, φ is differentiable with respect to the vector fields $X^1, \dots, X^m$ μ-almost everywhere on $\mathcal{V}$. Hence, we have:

$$\varphi(y) - \varphi(x) = \sum_{i=1}^{m} X^i \varphi(x)(y_i - x_i) + o(d_{SR}(x, y)), \ \forall y \in \mathcal{V}.$$

Let $\gamma_i^x : [0,1] \to M$, $i = 1, \dots, m$ be the integral flow associated to X^i starting at x. Then, we denote by

$$l_i = \lim_{t \to 0} \frac{\varphi(\gamma_i^x(t)) - \varphi(x)}{t}, \forall i = 1, \dots, m.$$

Recall that $g(\gamma_i^x(t), \gamma_i^x(t)) = g(X^i(\gamma_i^x(t)), X^i(\gamma_i^x(t))) = 1$, $\forall t \in [0,1]$.

It follows

$$d_{SR}(x, \gamma_i^x(t)) \leq |t|, \forall t \in [0,1].$$

Then,

$$x \in \Gamma_\varphi(x) \Rightarrow \varphi(x) - \varphi(z) \leq d_{SR}^2(x, z), \forall z \in \mathcal{V}.$$

In particular,

$$\varphi(x) - \varphi(\gamma_i^x(t)) \leq d_{SR}^2(x, \gamma_i^x(t)) \leq t^2.$$

This implies that $l_i = 0$. Hence,

$$X^i \varphi(x) = 0, \forall i = 1, \dots, m. \tag{4.1}$$

Assume now that there exists $y \in \Gamma_\varphi(x)$ such that $y \neq x$. So we have

$$\varphi(x) - \varphi(z) \leq d_{SR}^2(x, z) - d_{SR}^2(x, y), \forall z \in \mathcal{V}.$$

Let $\gamma_{x,y} : [0,1] \to M$ be a minimizing geodesic joining x to y. Then, $\forall t \in [0,1]$,

$$\varphi(x) - \varphi(\gamma_{x,y}(t)) \leq d_{SR}^2(x, \gamma_{x,y}(t)) - d_{SR}^2(x, y),$$

$$\Rightarrow -o(d_{SR}(x, \gamma_{x,y}(t))) \leq d_{SR}^2(x, \gamma_{x,y}(t)) - d_{SR}^2(x, y),$$

$$\Rightarrow -o(t \ d_{SR}(x, y)) \leq (1-t)^2 d_{SR}^2(x, y) - d_{SR}^2(x, y),$$

$$\Rightarrow -o(t \ d_{SR}(x, y)) \leq -2t \ d_{SR}^2(x, y) + t^2 \ d_{SR}^2(x, y),$$

$$\Rightarrow o(t \ d_{SR}(x, y)) \geq 2t \ d_{SR}^2(x, y) - o(t \ d_{SR}(x, y)),$$

$$\Rightarrow o(t\ d_{SR}(x,y)) \geq t\ d^2_{SR}(x,y).$$

For t small enough, $\dfrac{o(t\ d_{SR}(x,y))}{t}$ tends to zero which implies that $d^2_{SR}(x,y) = 0$.

This contradicts the fact that $x \neq y$.

2. Let us now prove that

$$\mu\text{-a.e. } x \in \mathcal{M}^\varphi, \Gamma_\varphi(x) = \big\{\exp_x(\frac{1}{2}d_x\varphi)\big\}.$$

Fix $\bar{x} \in \mathcal{M}^\varphi$. It follows that there is $k \in \mathbb{N}$ such that

$$d_{SR}(\bar{x},y) \geq 1/k,\ \forall y \in \Gamma_\varphi(\bar{x}).$$

Since Γ_φ is a closed set in $M \times M$, there exists an open neighborhood $\mathcal{V}_{\bar{x}}$ of $\bar{x}$ in $\mathcal{M}^\varphi$ such that

$$d_{SR}(z,w) \geq 1/k,\ \forall z \in \mathcal{V}_{\bar{x}},\ \forall w \in \Gamma_\varphi(z).$$

We define the function $\tilde{\varphi} : M \to \mathbb{R}$ as follows

$$\tilde{\varphi}(z) := \sup_{y\in M}\big\{\varphi^c(y) - d^2_{SR}(z,y)|\ d_{SR}(z,y) \geq 1/k\big\}.$$

Since d^2_{SR} is locally semiconcave on $M \times M\backslash D$, the function $\tilde{\varphi}$ is also locally semiconcave on M. The fact that, by construction, φ and $\tilde{\varphi}$ coincide on $\mathcal{V}_{\bar{x}}$ implies that φ is locally semiconcave on $\mathcal{M}^\varphi$. Thanks to the Rademacher Theorem (see Appendix B.1), φ is almost everywhere differentiable on $\mathcal{M}^\varphi$.

Let $\bar{y} \in \Gamma_\varphi(\bar{x})$ be given. By the definition of the Kantorovich potentials (see Theorem 9), we have

$$\varphi^c(\bar{y}) = \inf_{z\in M}\big\{\varphi(z) + d^2_{SR}(z,\bar{y})\big\} = \varphi(\bar{x}) + d^2_{SR}(\bar{x},\bar{y})$$

$$\Rightarrow \varphi(\bar{x}) + d^2_{SR}(\bar{x},\bar{y}) \leq \varphi(z) + d^2_{SR}(z,\bar{y}),\ \forall z \in M.$$

We define the function

$$\begin{array}{rccl} \psi : & M & \longrightarrow & \mathbb{R} \\ & z & \longmapsto & \psi(z) := \varphi(\bar{x}) + d^2_{SR}(\bar{x},\bar{y}) - \varphi(z) \end{array} \tag{4.2}$$

such that

$$\psi(z) \leq d^2_{SR}(z,\bar{y}),\ \forall z \in M \text{ and equality for } z = \bar{x}.$$

To conclude, we need the following lemma.

Lemma 4. *Let $x \neq y \in M$ be fixed and $\psi : M \to \mathbb{R}$ be a differentiable function at x such that*

$$\psi(z) \leq d_{SR}^2(z,y),\ \forall z \in M \text{ and equality for } z = x.$$

Then, there exists a unique minimizing geodesic $\gamma : [0,1] \to M$ joining x to y such that $y = exp_x(-d_x\psi/2)$.

Proof of Lemma 4. Let $x \neq y\ \in M$. Since $e_{SR}(z,y) = d_{SR}^2(z,y)$, $\forall z \in M$, there is a neighborhood $\mathcal{V}_x$ of x on M such that

$$\psi(z) \leq e_{SR}(z,y), \forall z \in \mathcal{V}_x \text{ and } \psi(x) = e_{SR}(x,y).$$

Without loss of generality, we can assume that there are m smooth vector fields $X^1, \dots, X^m$ on $\mathcal{V}_x$ such that

$$\Delta(z) = Span\big\{X^1(z), \dots, X^m(z)\big\},\ \forall z \in \mathcal{V}_x.$$

By the sub-Riemannian version of the Hopf-Rinow Theorem (see Theorem 8), there exists a minimizing geodesic $\gamma : [0,1] \to M$ joining y to x, associated to control $u^\gamma \in L^2([0,1], \mathbb{R}^m)$. By construction, u^γ minimizes the following quantity

$$C(u) = \int_0^1 \sum_{i=1}^m (u_i(t))^2 \mathrm{d}t,\ \forall u \in L^2([0,1], \mathbb{R}^m) \text{ such that } \mathrm{End}^y(u^\gamma) = x.$$

Let u be a control in $L^2([0,1], \mathbb{R}^m)$ such that $\mathrm{End}^y(u) \in \mathcal{V}_x$. Hence,

$$C(u) \geq e_{SR}(\mathrm{End}^y(u), y) \geq \psi(\mathrm{End}^y(u)),$$

and

$$C(u^\gamma) = e_{SR}(x,y) = \psi(x) = \psi(\mathrm{End}^y(u^\gamma)).$$

It means that u^γ minimizes

$$\begin{array}{rccl} D: & L^2([0,1], \mathbb{R}^m) & \longrightarrow & \mathbb{R} \\ & u & \longmapsto & D(u) = C(u) - \psi(\mathrm{End}^y(u)) \end{array}.$$

Then,

$$d_{u^\gamma} C - d_x\psi \cdot d_{u^\gamma}\mathrm{End}^y = 0$$

$$\Rightarrow d_{u^\gamma} C = d_x\psi \cdot d_{u^\gamma}\mathrm{End}^y.$$

Setting $p = d_x\psi$ and by the Lagrange Multiplicators Theorem (see Theorem B.1.5 in [Rif14]) with $\lambda_0 = 1$, there exists a normal extremal $\psi : [0,1] \to T^*M$ satisfying

$$\psi(1) = \big(x, d_x\psi/2\big).$$

Hence, there is a unique minimizing geodesic $\gamma : [0,1]$, projection of the normal extremal $\psi : [0,1] \to T^*M$, such that

$$y = \exp_x(-d_x\psi/2).$$

$\square$

Returning to our proof, the function ψ defined in (4.2) depends of z. As φ is almost everywhere differentiable on $\mathcal{M}^\varphi$, then ψ is also differentiable a.e. on $\mathcal{M}^\varphi$, in particular, at $\bar{x} \in \mathcal{M}^\varphi$. Thanks to the Lemma 4, there exists a unique minimizing geodesic joining $\bar{x}$ to $\bar{y}$. Moreover,

$$\bar{y} = \exp_{\bar{x}}(-d_{\bar{x}}\psi/2) = \exp_{\bar{x}}(d_{\bar{x}}\varphi/2).$$

In conclusion, there is a unique transport map $T : M \to M$ from μ to ν such that

$$\text{for } \mu\text{-a.e. } x \in \mathcal{M}^\varphi,\ T(x) = \exp_x(d_x\varphi/2).$$

$\square$

Proposition 15. *Assume that the distribution Δ has no non-trivial singular minimizing geodesics. Then, the sub-Riemannian distance d_{SR} is locally semiconcave on $(M \times M)\backslash D$.*

Proof of Proposition 15 . Fix $(x, y) \in (M \times M)\backslash D$. For sake of simplicity, let us first assume that there is a unique minimizing geodesic $\gamma : [0,1] \to M$ steering x to y. There exist an open neighborhood $\mathcal{V}$ of $\gamma([0,1])$ on M and an orthonormal family (with respect to g) of m vector fields $X^1, \dots, X^m$ such that

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

According to a change of coordinates if necessary, we can assume that $\mathcal{V}$ is an open subset of $\mathbb{R}^n$. Moreover, there is a control $u^\gamma \in L^2([0,1], \mathbb{R}^m)$ associated to γ, ie.

$$\dot{\gamma}(t) = \sum_{i=1}^{m} u_i^\gamma(t) X^i(\gamma(t)),\ \forall t \in [0,1]$$

and

$$||u^\gamma||_{L^2} = \text{length}(\gamma) = d_{SR}(x, y).$$

Assume now that there is a sequence $\{\gamma^k\}_k$ of minimizing geodesics between x and y. We denote by $\mathcal{K}$ the set of minimizing geodesics between x and y. By [FR10] (see also [Rif14]), $\mathcal{K}$ is compact with respect to the uniform topology. So we can repeat the above by covering $\mathcal{K}$ by a finite number of open tubes admitting orthonormal frames. Up to taking a subsequence, the bounded sequence of controls $\{u^k\}$ in $L^2([0,1],\mathbb{R}^m)$ associated to γ^k such that

$$||u^k||_{L^2} = \text{length}(\gamma^k)$$

converges to some u^γ in $L^2([0,1],\mathbb{R}^m)$. By the lower semicontinuity of the norm, we have

$$||u^\gamma||_{L^2} = \text{length}(\gamma).$$

Since by assumption, γ is regular, there exist n controls $v^1,\dots,v^n \in L^2([0,1],\mathbb{R}^m)$ such that the linear operator

$$\begin{array}{ccc} \mathbb{R}^n & \longrightarrow & T_yM \\ \alpha & \longmapsto & \displaystyle\sum_{i=1}^n \alpha_i D_{u^\gamma}\text{End}^x(v^i) \end{array} \tag{4.3}$$

is invertible.

Recall that $C^\infty(\mathbb{R}^m)$ is dense in $L^2([0,1],\mathbb{R}^m)$, we can assume that we have $\tilde{v}^1,\dots,\tilde{v}^n$ in $C^\infty(\mathbb{R}^m)$ close to $v^1,\dots,v^n$ satisfying the property (4.3). By abuse of notation, we set $v^i = \tilde{v^i}, \forall i = 1,\dots,n$.

Define locally

$$\begin{array}{cccc} \mathcal{H}: & M\times\mathbb{R}^n & \to & M\times M \\ & (z,\alpha) & \mapsto & (z, \text{End}^x(u^\gamma + \displaystyle\sum_{i=1}^n \alpha_i v^i)). \end{array}$$

This mapping is well-defined and of class C^2 in the neighborhood of $(x,0)$. It satisfies $\mathcal{H}(x,0) = (x,y)$ and its differential at $(x,0)$ is invertible.
By the Local Inverse Function Theorem, there exist an open set $\mathcal{B}$ of $M\times M$ centered at (x,y) and a function $\mathcal{G}:\mathcal{B}\to M\times\mathbb{R}^n$ of class C^2 such that

$$\mathcal{H}\circ\mathcal{G}(z,w) = (z,w), \forall (z,w)\in\mathcal{B}.$$

Denote by ζ the second component of $\mathcal{G}$. For any $(z,w)\in\mathcal{B}$,

$$d_{SR}(x,z) \le ||u^\gamma + \sum_{i=1}^n (\zeta(z,w))_i v^i||_{L^2}.$$

Define

$$\phi^{x,y}(z,w) := ||u^\gamma + \sum_{i=1}^n (\zeta(z,w))_i v^i||_{L^2}, \forall z\in\mathcal{B}.$$

Then,
$$\phi^{x,y}(z,w) \geq d_{SR}(z,w), \forall (z,w) \in \mathcal{B} \text{ and } \phi^{x,y}(x,y) = d_{SR}(x,y).$$
For every $(z,w) \in \mathcal{B}$, we put a C^2 function $\phi^{x,y}$ on the graph of d_{SR} at (z,w) with a uniform control of C^2 norm of $\phi^{x,y}$. Hence, d_{SR} is locally semiconcave on $(M \times M) \backslash D$.

$\square$

Example 11. ***(Rank two in dimension three)***
Consider a totally nonholonomic distribution Δ of rank 2 on a manifold M of dimension 3. We define the Martinet surface of Δ as the set defined by

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

where
$$[\Delta, \Delta](x) := Span\big\{ [X,Y](x) |\ X, Y \ sections\ of\ \ \Delta \big\}.$$
By the same argument as in example 9, we prove that singular horizontal paths are contained in Σ_Δ. We claim that Σ_Δ is a closed subset of M which is countably 2-rectifiable. Let us prove our claim. By a change of coordinates if necessary, we assume that we work in $\mathbb{R}^3$ with coordinates (x_1, x_2, x_3). Let X^1, X^2 be two smooth vector fields generating the distribution, that is

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

By a change of coordinates if necessary, we can assume that for $i = 1, 2$

$$X^i = \partial_{x_i} + \alpha_i(x) \partial_{x_3}$$

with $\alpha_i : \mathbb{R}^3 \to \mathbb{R}$ smooth functions.
We set $I = (i_1, \ldots, i_k) \in \{1, 2\}$ and we denote by X^I the vector field given by

$$X^I = \Big[X^{i_1}, \big[X^{i_2}, \ldots [X^{i_k}, X^{i_{k-1}}] \ldots \big] \Big].$$

Since Δ is totally nonholonomic, there exists a positive integer r such that

$$\mathbb{R}^3 = Span\big\{ X^I(x); length(I) \leq r \big\}, \forall x \in \mathbb{R}^3.$$

For any I of $length(I) \geq 2$, there exists a function $g_I : \mathbb{R}^3 \to \mathbb{R}$ such that

$$X^I(x) = g_I(x) \partial_{x_3}.$$

We set
$$\mathcal{A}_k := \big\{ x \in \mathbb{R}^3 |\ g_I(x) = 0, \forall I, length(I) \leq k \big\}$$

and

$$\Sigma_\Delta := \bigcup_{k=2}^{r-1} \left(\mathcal{A}_k \backslash \mathcal{A}_{k+1}\right).$$

By the Implicit Function Theorem, each set $\mathcal{A}_k \backslash \mathcal{A}_{k+1}$ *I can be covered by a countable union of smooth hyper surfaces. Assume that there is* $x \in \mathcal{A}_k \backslash \mathcal{A}_{k+1}$*. There is* $J = (j_1, \ldots, j_{k+1})$ *of length* $k+1$ *such that*

$$g_J(x) \neq 0, \forall x \in \mathbb{R}^3.$$

We set $I = (j_2, \ldots, j_{k+1})$*. Since* $g_I(x) = 0, \forall x \in \mathbb{R}^3$*, we have*

$$\begin{aligned} X^J &= \left[X^{j_1}, X^I\right] \\ &= \left[\partial_{x_{j_1}} + \alpha_{x_{j_1}} \partial_{x_3}, g_I(x)\partial_{x_3}\right] \\ &= \left(\partial_{x_{j_1}} g_I(x) + \alpha_{x_{j_1}}(x)\partial_{x_3} g_I(x)\right)\partial_{x_3}. \end{aligned}$$

So we have

$$\partial_{x_{j_1}} g_I(x) \neq 0 \text{ or } \partial_{x_3} g_I(x) \neq 0.$$

Hence, we deduce that

$$\mathcal{A}_k \backslash \mathcal{A}_{k+1} \subset \bigcup_{length(I)=k} \left\{x \in \mathbb{R}^3;\ \exists\, i \in \{1,2,3\} \ s.t.\ \partial_{x_i} g_I(x) \neq 0\right\}.$$

It follows that Σ_Δ *has Lebesgue measure zero. For any* $x \neq y \in M$ *such that* $x \notin \Sigma_\Delta$ *or* $y \notin \Sigma_\Delta$*, any minimizing geodesic joining* x *and* y *is nonsingular. By Theorem 15, the sub-Riemannian distance is locally semiconcave. Set* $\Omega := M \backslash \Sigma_\Delta$ *a subset of full Lebesgue measure. By Theorem 10, for any two probability measure* μ, ν *compactly supported on* M *such that*

$$supp(\mu) \subset \Omega \text{ or } supp(\nu) \subset \Omega$$

there is existence and uniqueness of optimal transport maps from μ *to* ν*.*

Example 12. ***(Rank two in dimension four)***
In $\mathbb{R}^4$ *with coordinates* (x_1, x_2, x_3, x_4)*, we consider the distribution given by*

$$\Delta(x) = Span\left\{X^1(x), X^2(x)\right\}, \forall x \in \mathbb{R}^4$$

with

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

Let $\gamma : [0,1] \to M$ be a singular horizontal curve associated to a control $u \in L^2([0,1], \mathbb{R}^2)$. By Proposition 9, there exists an absolutely continuous arc $p : [0,1] \to (\mathbb{R}^4)^* \backslash \{0\}$ satisfying

$$\dot{p}(t) = -u_1(t)p(t) \cdot D_{\gamma(t)}X^1 - u_2(t)p(t) \cdot D_{\gamma(t)}X^2, a.e.t \in [0,1] \tag{4.4}$$

and

$$p(t) \cdot X^1(\gamma(t)) = p(t) \cdot X^2(\gamma(t)) = 0, \forall t \in [0,1]. \tag{4.5}$$

A computation gives

$$[X^1, X^2] = \partial_{x_3}, \quad \big[X^1, [X^1, X^2]\big] = 0, \quad \big[X^2, [X^1, X^2]\big] = -\partial_{x_4}.$$

From (4.5), we get

$$p_1(t) = p_2(t) + x_1(t)p_3(t) + x_3(t)p_4(t) = 0, \forall t \in [0,1].$$

From (4.4), we get

$$\dot{p}(t) = -u_2(t)p_3(t)\partial_{x_1} - u_2(t)p_4(t)\partial_{x_3}.$$

Hence,

$$\begin{cases} \dot{p_1}(t) &= -u_2(t)p_3(t) \\ \dot{p_2}(t) &= 0 \\ \dot{p_3}(t) &= -u_2(t)p_4(t) \\ \dot{p_4}(t) &= 0 \end{cases} \tag{4.6}$$

It implies p_1 and p_3 are constants on $[0,1]$. Assume that $p_4 = 0$. Then, p_3 is a constant on $[0,1]$ which means that x_1 is constant or $p_2 = p_3 = 0$. As $p \not\equiv 0$, then x_1 is constant on $[0,1]$ and $u_1 \equiv 0$. Or, we have $u_2(t)p_3(t) = 0$ then $p_3 = 0$ (because $|u(t)| = 1$ a.e. $t \in [0,1]$) which contradicts the fact that $p \not\equiv 0$. Hence, $p_4 \neq 0$ and we deduce that

$$0 = u_2(t)p_3(t) = \Big(-\frac{\dot{p_3}(t)}{p_4(t)}\Big)p_3(t), \text{ a.e. } t \in [0,1].$$

It follows that p_3 is a constant on $[0,1]$ and then, $u_2 \equiv 0$. Thus, γ satisfies

$$\dot{\gamma}(t) = u_1(t)X^1(\gamma(t))$$

and is of the form

$$\gamma(t) = \Big(\gamma_1(t), \gamma_2(0), \gamma_3(0), \gamma_4(0)\Big), a.e. \ t \in [0,1].$$

Up to a parameterization by arc-length, singular horizontal curves with respect to Δ *satisfy*

$$\dot{\gamma}(t) = X^1(\gamma(t)), \ \textit{a.e. } t \in [0,1].$$

We denote by Ω *the subset of* M *given by*

$$\Omega := \Big\{(x,y) \in \mathbb{R}^4 \times \mathbb{R}^4 | \ (y-x) \notin Span\{e_1\}\Big\},$$

where e_1 *denotes the first vector in the canonical basis of* $\mathbb{R}^4$*. The sub-Riemannian distance function* d_{SR} *is locally semiconcave on the interior of* Ω*. For any two probability measure* μ, ν *compactly supported on* M *such that*

$$supp(\mu \times \nu) \subset \Omega,$$

there is existence and uniqueness of optimal transport maps from μ *to* ν*.*

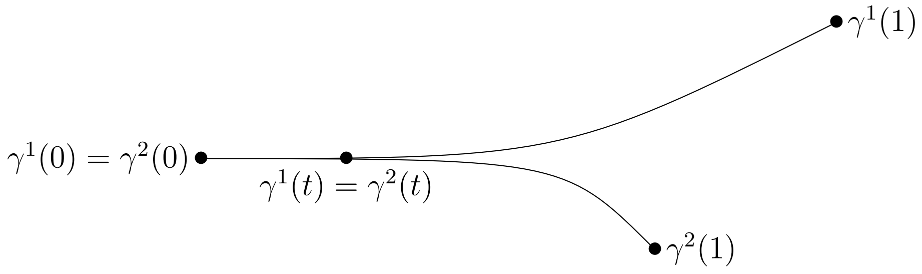
4.3 Method of Cavalletti and Huesmann

In [CH15], Cavalletti and Huesmann proved existence and uniqueness of the optimal transport maps on metric measured non-branching spaces using a new technique based on a localized contraction property.

The non-branching condition plays a crucial role. We recall its definition.

Definition 26. *Two distinct geodesics* $\gamma^1, \gamma^2 : [0,1] \to M$ *branch if there exists* $t \in]0,1[$ *such that*

$$\gamma^1(s) = \gamma^2(s) \textit{ for all } s \in [0,t].$$



A space where there are no branching geodesics is called non-branching.

In our setting, the assumption of Cavalletti and Huesmann amounts to ensure that M is equipped with a complete sub-Riemannian structure (Δ, g) together with a measure η on M such that the metric measured space (M, d_{SR}, η) is non-branching and satisfies the following property:

For every compact set $K \subset M$, there exists a measurable function $f : [0, 1] \to [0, 1]$ with

$$\limsup_{t \to 0} f(t) > \frac{1}{2}$$

and a positive constant $\delta \leq 1$ such that

$$\eta(A_{t,x}) \geq f(t)\eta(A), \ \forall 0 \leq t \leq \delta \tag{4.7}$$

for any compact set $A \subset K$ and any base point $x \in K$ with

$$A_{t,x} := \Big\{ \gamma(t) | \ \gamma : [0, 1] \to M \text{ minimizing geodesic s.t. } \gamma(0) \in A, \gamma(1) = x \Big\}.$$

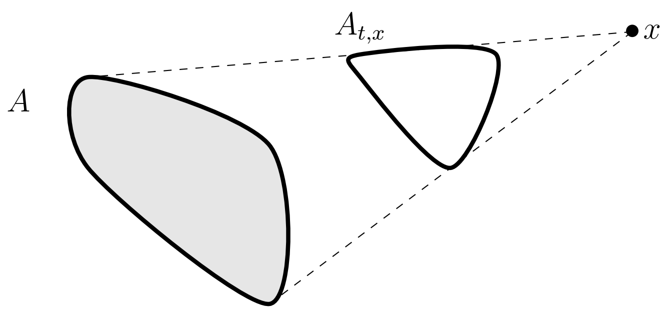


Figure 4.1

The condition (4.7) on measure η means that when we contract a set towards a point, its measure does not shrink too fast.

We recall that there exists a c-convex function φ provided by Theorem 9 such that any optimal transport plan α is concentrated on

$$\Gamma_\varphi := \Big\{ (x, y) \in M \times M | \ \varphi^c(y) - \varphi(x) = c(x, y) \Big\}.$$

Moreover, Γ_φ is a c-cyclically monotone set on $M \times M$ (see Proposition 2).

We start by showing that branching at starting points belonging to the support of optimal transport plans does not happen almost everywhere.

Lemma 5. *Let (M, d_{SR}, η) be a non-branching metric measured space. Let (x_0, y_0) and (x_1, y_1) be two distinct points of Γ_φ. Then, for any $i = 1, 2$ and every minimizing geodesics $\gamma_i : [0, 1] \to M$ joining x_i to y_i, we have*

$$\gamma_0(t) \neq \gamma_1(t), \forall t \in]0, 1[.$$

Proof of Lemma 5. Assume by contradiction that there is $\bar{t} \in]0, 1[$ such that

$$\gamma_0(\bar{t}) = \gamma_1(\bar{t}).$$

For $i = 0, 1$, we have

$$d_{SR}\big(x_i, \gamma_i(\bar{t})\big) = \bar{t}\ d_{SR}\big(x_i, y_i\big) \quad \text{and} \quad d_{SR}\big(\gamma_i(\bar{t}), y_i\big) = (1 - \bar{t})\ d_{SR}\big(x_i, y_i\big).$$

Case1: $d_{SR}(x_0, y_0) \neq d_{SR}(x_1, y_1)$

$$\begin{aligned}
& d_{SR}^2(x_0, y_1) + d_{SR}^2(x_1, y_0) \\
\leq\ & \Big(d_{SR}(x_0, \gamma_0(\bar{t}))\ +\ d_{SR}(\gamma_1(\bar{t}), y_1)\Big)^2\ +\ \Big(d_{SR}(x_1, \gamma_1(\bar{t}))\ +\ d_{SR}(\gamma_0(\bar{t}), y_0)\Big)^2 \\
\leq\ & \Big(\bar{t}\ d_{SR}(x_0, y_0) + (1 - \bar{t}) d_{SR}(x_1, y_1)\Big)^2 + \Big(\bar{t} d_{SR}(x_1, y_1) + (1 - \bar{t}) d_{SR}(x_0, y_0)\Big)^2 \\
\leq\ & \Big(\bar{t}^2 + (1 - \bar{t})^2\Big) d_{SR}^2(x_0, y_0) \\
& \qquad + \Big(\bar{t}^2 + (1 - \bar{t})^2\Big) d_{SR}^2(x_1, y_1) + 4\bar{t}(1 - \bar{t}) d_{SR}(x_0, y_0) d_{SR}(x_1, y_1) \\
\leq\ & d_{SR}^2(x_0, y_0) + d_{SR}^2(x_1, y_1) - 2\bar{t} d_{SR}^2(x_0, y_0) + 2\bar{t}^2 d_{SR}^2(x_0, y_0) - 2\bar{t} d_{SR}^2(x_1, y_1) \\
& \qquad + 2\bar{t}^2 d_{SR}^2(x_1, y_1) + 4\bar{t} d_{SR}(x_0, y_0) d_{SR}(x_1, y_1) - 4\bar{t}^2 d_{SR}(x_0, y_0) d_{SR}(x_1, y_1) \\
<\ & d_{SR}^2(x_0, y_0) + d_{SR}^2(x_1, y_1).
\end{aligned}$$

The last inequality is obtained from $0 < \bar{t} < 1$ which contradicts the c-cyclically monotonicity of Γ_φ.

Case2: $d_{SR}(x_0, y_0) = d_{SR}(x_1, y_1)$

We define the curve $\gamma : [0,1] \to M$ by

$$\gamma(s) = \begin{cases} \gamma_0(s) & s \in [0, \bar{t}\,] \\ \gamma_1(s) & s \in [\,\bar{t}, 1] \end{cases}$$

Then, γ coincides with γ_0 on the interval $[0, \bar{t}]$ which contradicts the fact that M is non-branching. $\square$

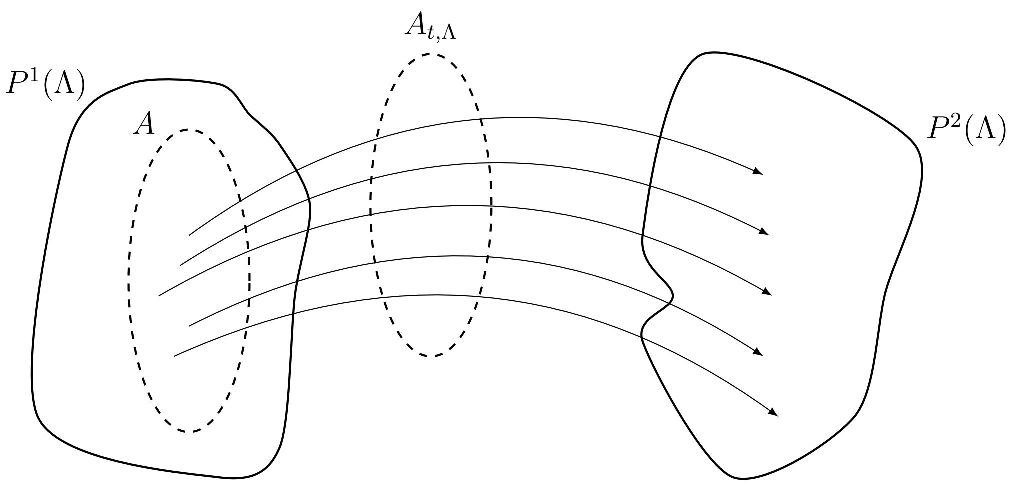
We denote by $P^1 : M \times M \to M$ the projection map into the first component. The following proposition is a main consequence of Assumption (4.7).

Proposition 16. *For any compact set Λ of Γ_φ, the following inequality holds*

$$\eta(A_{t,\Lambda}) \geq f(t)\eta(A), \ \forall t \in [0,1], \ \textit{for any } A \subset P^1(\Lambda)$$

where

$$A_{t,\Lambda} := \big\{\gamma(t) |\ \gamma : [0,1] \to M \ \textit{minimizing geodesic s.t. } \gamma(0) \in A, \gamma(1) \in P^2(\Lambda)\big\}.$$



Proof of Proposition 16. We will proceed in two steps.
1. Let $\{y_i\}_{i\in\mathbb{N}}$ be a dense set in $P^2(\Lambda)$. For $n \in \mathbb{N}$ and $i \leq n$, we consider the following family of sets

$$E_n(i) := \Big\{x \in P^1(\Lambda) |\ \varphi^c(y_i) - \varphi(x) = c(x, y_i)\Big\}$$
$$= \Big\{x \in P^1(\Lambda) |\ y_i \in \Gamma_\varphi(x)\Big\}.$$

We set for $i \in \mathbb{N}$, $\Lambda_n := \bigcup_{i=1}^{n} E_n(i) \times \{y_i\}$ such that

$$P^1(\Lambda_n) = P^1(\Lambda).$$

By Assumption (4.7), it holds that for any compact $A \subset P^1(\Lambda)$,

$$\eta\Big((A \cap E_n(i))_{t,y_i}\Big) \geq f(t)\ \eta\Big(A \cap E_n(i)\Big), \forall t \in [0, \delta]$$

where $f : [0, 1] \to [0, 1]$ is independent of $\{y_i\}_{i\in\mathbb{N}}$ and of n, and satisfies

$$\limsup_{t\to 0} f(t) > 1/2.$$

Since $A = \bigcup_{i=1}^{n} A \cap E_n(i)$, it follows

$$A_{t,\Lambda_n} := \Big\{\gamma(t)|\gamma : [0, 1] \to M \text{ minimizing geodesic s.t. } \gamma(0) \in A, \gamma(1) \in P^2(\Lambda_n)\Big\}$$
$$= \bigcup_{i=1}^{n} \{\gamma(t)|\gamma : [0, 1] \to M \text{ minimizing geodesic s.t.}$$
$$\gamma(0) \in A \cap E_n(i), \gamma(1) = y_i\Big\}$$
$$= \bigcup_{i=1}^{n} \big(A \cap E_n(i)\big)_{t,y_i}.$$

Thanks to Lemma 5, we have for all $t \in [0, 1]$,

$$\Big(A \cap E_n(i)\Big)_{t,y_i} \cap \Big(A \cap E_n(j)\Big)_{t,y_j} = \emptyset.$$

Then it holds for all $t \in [0, \delta]$:

$$\eta\big(A_{t,\Lambda_n}\big) \geq \eta\Big(\bigcup_{i=1}^{n}(A \cap E_n(i))_{t,y_i}\Big)$$
$$\geq \sum_{i=1}^{n} \eta\Big((A \cap E_n(i))_{t,y_i}\Big)$$

$$\geq f(t) \sum_{i=1}^{n} \eta\Big(A \cap E_n(i)\Big)$$

$$\geq f(t)\ \eta\Big(\bigcup_{i=1}^{n} A \cap E_n(i)\Big)$$

$$\geq f(t)\ \eta(A).$$

2. For all $n \in \mathbb{N}$, we have $\Lambda_n \subset supp(\mu) \times supp(\nu)$ a compact set . Then, there exists a subsequence $\{\Lambda_{n_k}\}_{k\in\mathbb{N}}$ of $\{\Lambda_n\}_{n\in\mathbb{N}}$ converging to a compact space K with the Hausdorff metric. Let $(x, y) \in K$. By the definition of $E_{n_k}(i)$, we get

$$\varphi^c(y) - \varphi(x) = c(x, y),\ x \in P^1(\Lambda) \text{ and } y \in P^2(\Lambda).$$

So we have $K \subset \Lambda$. Hence,

$$\eta(A_{t,\Lambda}) \geq \eta(A_{t,K}) \geq \lim_{k\to+\infty} \sup\ \eta(A_{t,\Lambda_{n_k}}) \geq f(t)\ \eta(A).$$

$\square$

Lemma 6. *Let $\Lambda_1, \Lambda_2 \subset \Gamma_\varphi$ be two compact sets such that*

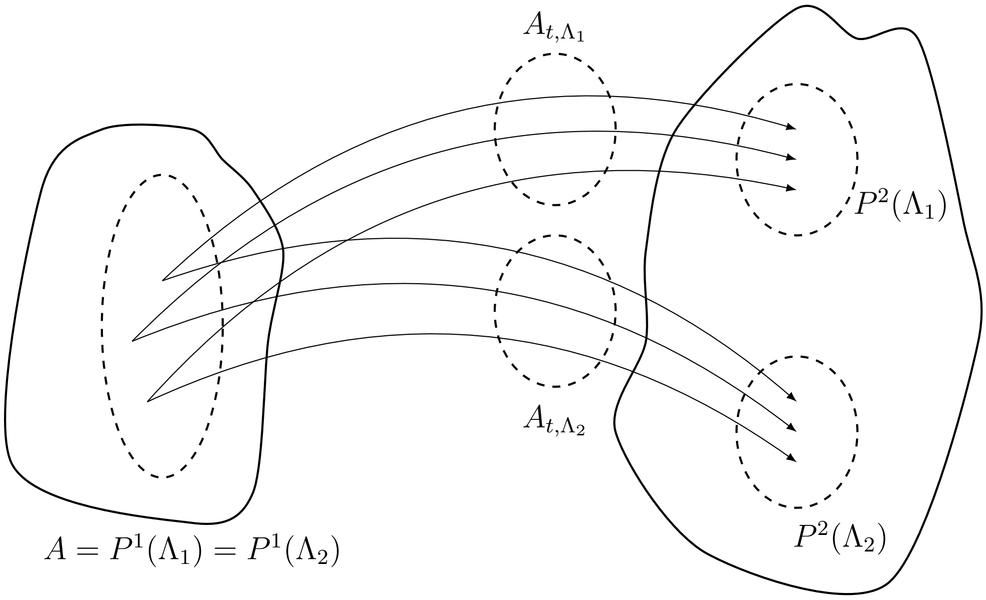
(i) $P^1(\Lambda_1) = P^1(\Lambda_2)$

(ii) $P^2(\Lambda_1) \cap P^2(\Lambda_2) = \emptyset$

Then, $\eta(P^1(\Lambda_1)) = \eta(P^1(\Lambda_2)) = 0$.

Proof of Lemma 6. We set $A = P^1(\Lambda_1) = P^1(\Lambda_2)$ and we define the following sets for $i = 1, 2$:

$$A_{t,\Lambda_i} := \big\{\gamma(t)|\gamma : [0, 1] \to M \text{ minimizing geodesic s.t.} \gamma(0) \in A, \gamma(1) \in P^2(\Lambda_i)\big\}.$$



Since $P^2(\Lambda_1) \cap P^2(\Lambda_2) = \emptyset$, by lemma 5, we have

$$A_{t,\Lambda_1} \cap A_{t,\Lambda_2} = \emptyset, \ \forall t \in [0,1].$$

For $\delta > 0$ fixed, we define $A^\delta := \left\{x | \ d_{SR}(A,x) \leq \delta\right\}$.

Hence,

$$\begin{aligned}
\eta(A) &= \limsup_{\delta \to 0} \ \eta(A^\delta) \\
&\geq \limsup_{t \to 0} \ \eta(A_{t,\Lambda_1} \cap A_{t,\Lambda_2}) \\
&= \limsup_{t \to 0} \ \Big[\eta(A_{t,\Lambda_1}) + \eta(A_{t,\Lambda_2})\Big] \\
&\geq 2 \ \limsup_{t \to 0} \ f(t) \ \eta(A).
\end{aligned}$$

By hypothesis, we have

$$\limsup_{t \to 0} \ f(t) > 1/2.$$

Then, it follows

$$\eta(A) = 0.$$

□

Theorem 11. *Let (M, d_{SR}, η) be a non-branching metric space verifying assumption (4.7). Let μ,ν be two probability measures compactly supported on M such that μ is absolutely continuous with respect to η. Then, there is existence and uniqueness of an optimal transport map $T : M \to M$ solution for the Monge problem.*

Proof of Theorem 11. We consider the set

$$E := \{x \in M | \; \Gamma_\varphi(x) \text{ is not a singleton}\}$$

and we assume by contradiction that

$$\eta(E) > 0.$$

It follows that there is $k \in \mathbb{N}$ such that the set given by

$$E_k := \{x \in E | \; diam \; \Gamma_\varphi(x) > 1/k\}$$

has positive measure with respect to η. Without loss of generality, we can assume that the manifold M can be covered finitely by many open balls $(\mathcal{U}_i)_{i\in I}$ of diameter less or equal to $1/k$. From $(\mathcal{U}_i)_{i\in I}$, we construct a finite family of open sets $(\mathcal{V}_i)_{i\in I}$ pairwise disjoint covering M by proceeding as follows

$$\begin{cases} \mathcal{V}_1 &= \mathcal{U}_1 \\ \mathcal{V}_2 &= \mathcal{U}_2 \backslash \mathcal{U}_1 \\ \mathcal{V}_3 &= \mathcal{U}_3 \backslash (\mathcal{U}_1 \cup \mathcal{U}_2) \\ & \vdots \\ \mathcal{V}_n &= \mathcal{U}_n \backslash (\mathcal{U}_1 \cup \mathcal{U}_2 \cup \dots \mathcal{U}_{n-1}) \\ & \vdots \end{cases}$$

such that

$$\bigcup_{i\in I} \mathcal{U}_i = \bigcup_{i\in I} \mathcal{V}_i.$$

Therefore, for any $x \in E_k$, there are $i_x, j_x \in I$ with $i_x \neq j_x$ such that

$$\Gamma_\varphi(x) \cap \mathcal{V}_{i_x} \neq \emptyset \text{ and } \Gamma_\varphi(x) \cap \mathcal{V}_{j_x} \neq \emptyset.$$

Denote by

$$E_{k,i} := \bigcup_{x\in E_k} \{x\} \times \left(\Gamma_\varphi(x) \cap \mathcal{V}_{i_x}\right)$$

and

$$E_{k,j} := \bigcup_{x \in E_k} \{x\} \times \big(\Gamma_\varphi(x) \cap \mathcal{V}_{j_x}\big).$$

We notice that $P^1(E_{k,i}) = P^1(E_{k,j}) = E_k$ such that

$$\eta(E_k) > 0. \tag{4.8}$$

We also have $P^2(E_{k,i}) \cap P^2(E_{k,j}) = \emptyset$ since for any $x \in E_k$, $\mathcal{V}_{i_x} \cap \mathcal{V}_{j_x} = \emptyset$, for $i_x \neq j_x$. Using Lemma 6,

$$\eta(E_k) = 0$$

which contradicts (4.8).

We conclude that for a.e. $x \in M$, $\Gamma_\varphi(x)$ is a singleton. Thus, any optimal transport plan $\alpha \in \Pi(\mu, \nu)$ such that *supp* $\alpha \subset \Gamma$, is concentrated on a graph. $\square$

Chapter 5

Mass Transportation on sub-Riemannian structures of rank 2 in dimension 4

5.1 Introduction and main result

For a two-rank distribution Δ on a three-dimensional manifold M (see Example 11), we have existence and uniqueness of optimal transport maps for the sub-Riemannian quadratic cost because non-trivial singular horizontal paths are included in the Martinet surface Σ_Δ given by $\Sigma_\Delta := \{x \in M| \ \Delta(x) + [\Delta, \Delta](x) \neq T_x M\}$ which has Lebesgue measure zero. The first relevant case to consider is the one of rank-two distributions in dimension four. In this case, as shown by Sussmann [Sus96], singular horizontal paths can be seen (locally) as the orbits of a smooth vector field, at least, outside a set of Lebesgue measure zero.

The definition of a real analytic manifold is similar to that of a smooth manifold. We begin by recalling that an analytic function f is an infinitely differentiable function such that the Taylor series at any point x_0 in its domain, converges to $f(x)$ for x in a neighborhood of x_0. We say that a manifold M of dimension n is real analytic if transition maps are analytic. We provide M with a real analytic distribution Δ of rank m ($m < n$), that is for each $x \in M$, there is an open neighborhood $\mathcal{U}$ containing x and m analytic vector fields $X^1, \ldots, X^m$ on $\mathcal{U}$ such that

$$\Delta(y) = Span\{X^1(y), \ldots, X^m(y)\}, \ \forall y \in \mathcal{U}.$$

In this case, for analytic functions $u_i : [0,1] \to \mathbb{R}, i = 1, \ldots, m$, the Cauchy problem given by

$$\begin{cases} \dot{\gamma_u}(t) &= \displaystyle\sum_{i=1}^{m} u_i(t)X^i(\gamma_u(t)), \ a.e.\ t \in [0,1] \\ \gamma_u(0) &= x \end{cases}$$

has a real analytic solution on M for $t \in [0,1]$.

Our main result is the following:

Theorem 12. *Let M be a real analytic manifold of dimension 4 and (Δ, g) be a complete analytic sub-Riemannian structure of rank 2 on M such that*

$$\forall x \in M, \ \Delta(x) + [\Delta,\Delta](x) \ \text{has dimension 3}, \tag{5.1}$$

where

$$[\Delta,\Delta] := Span\{[X,Y] \mid X, Y \ \text{sections of}\ \Delta\}.$$

Let μ, ν be two probability measures compactly supported on M such that μ is absolutely continuous with respect to the Lebesgue measure. Then, there is existence and uniqueness of an optimal transport map from μ to ν for the sub-Riemmannian quadratic cost $c : M \times M \to [0, +\infty[$ defined by:

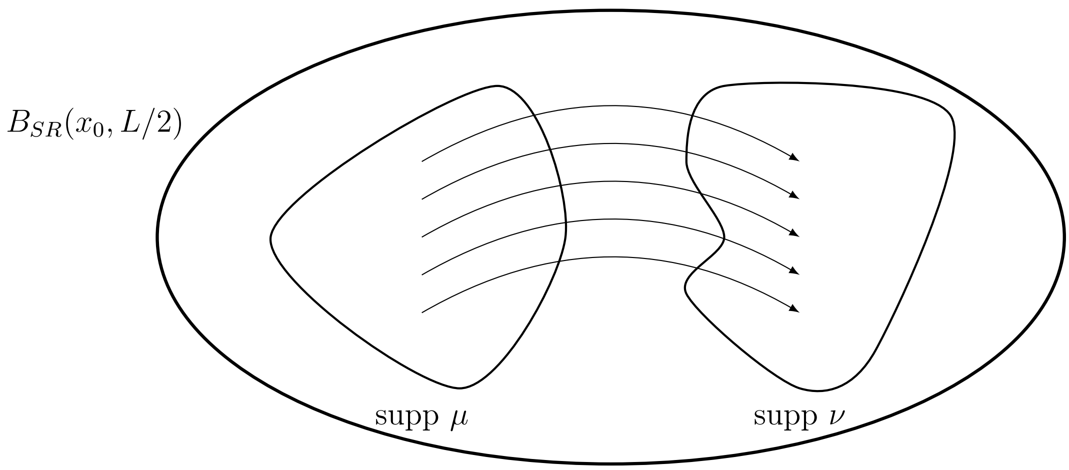
$$c(x,y) := d_{SR}^2(x,y), \ \forall (x,y) \in M \times M.$$

Since both $supp(\mu)$ and $supp(\nu)$ are compact and the metric space (M, d_{SR}) is complete, there are $x_0 \in M$ and a constant $L > 0$ such that

$$supp\ \mu \ \cup supp\ \nu \subset B_{SR}(x_0, L/4) \tag{5.2}$$

where $B_{SR}(x_0, L/4)$ is the sub-Riemannian ball in $\mathbb{R}^4$ centered at x_0 of radius $L/4$.

As a consequence, any minimizing geodesic $\gamma : [0,1] \to M$ from $x \in supp\ \mu$ to $y \in supp\ \nu$ is contained in $B_{SR}(x_0, L/2)$.



From now on, we work in the compact set $B_{SR}(x_0, L/2)$ of diameter L and so, we proceed as if M were a compact manifold.

We recall that there exists a c-convex function $\varphi : M \to \mathbb{R}$ provided by Theorem 9 such that any optimal transport plan $\alpha \in \Pi(\mu, \nu)$ is concentrated on

$$\Gamma_\varphi := \Big\{(x, y) \in M \times M |\ \varphi^c(y) - \varphi(x) = c(x, y)\Big\}.$$

Following [FR10], let us consider the following definition:

Definition 27. *We call "static" set $\mathcal{S}$ and "moving" set $\mathcal{M}$ respectively the sets defined as follows:*

$$\mathcal{S} := \Big\{x \in M |\ x \in \Gamma_\varphi(x)\Big\},$$

$$\mathcal{M} := \Big\{x \in M |\ x \notin \Gamma_\varphi(x)\Big\}.$$

As in [FR10], we shall show that "static" points do not move, i.e. almost every $x \in \mathcal{S}$ is transported to itself. For sake of completeness, the proof of the following lemma is given in Theorem 10.

Lemma 7. *For μ-a.e. $x \in \mathcal{S}$, we have $\Gamma_\varphi(x) = \{x\}$.*

We need now to show that almost every moving point is sent to a singleton. For this aim, we need to distinguish between two types of moving points. For every

$x \in M$ and every $T > 0$, we denote by $\Omega^R_{x,T}$ the set of regular minimizing geodesics $\gamma : [0,T] \to M$ starting at x. We also denote by $\Omega^S_{x,T}$ the set of singular minimizing geodesics $\gamma : [0,T] \to M$ starting at x.

Definition 28. *Let $T > 0$. For every $x \in \mathcal{M}$, we set*

$$\Gamma^S(x) := \left\{ y \in \Gamma_\varphi(x) \,|\, \exists \gamma \in \Omega^S_{x,T} \ s.t.\ \gamma(T) = y \right\}$$

and

$$\Gamma^R(x) := \left\{ y \in \Gamma_\varphi(x) \,|\, \exists \gamma \in \Omega^R_{x,T} \ s.t.\ \gamma(T) = y \right\}.$$

Moreover, we let

$$\mathcal{M}^S := \left\{ x \in \mathcal{M} |\ \Gamma^S(x) \neq \emptyset \right\} \quad and \quad \mathcal{M}^R := \left\{ x \in \mathcal{M} |\ \Gamma^R(x) \neq \emptyset \right\}.$$

Note that, by construction, for every $x \in \mathcal{M}$, $\Gamma_\varphi(x) = \Gamma^R(x) \cup \Gamma^S(x)$. Furthermore, if there are no non-trivial singular minimizing curves then $\mathcal{M}^S = \emptyset$.

First, using techniques reminiscent to the previous works by Agrachev-Lee [AL09] and Figalli-Rifford [FR10], we prove that

Proposition 17. *For $\mathcal{L}^4$-a.e. $x \in \mathcal{M}^R$, $\Gamma^R(x)$ is a singleton.*

Then, using a localized contraction property for singular curves which holds thanks to (5.1), the technique developed by Cavalletti and Huesmann [CH15] allows to show that

Proposition 18. *For $\mathcal{L}^4$-a.e. $x \in \mathcal{M}^S$, $\Gamma^S(x)$ is a singleton.*

It remains to show that for almost every $x \in M$, $\Gamma_\varphi(x)$ is a singleton. Again this will follow from a local contraction property together with the approach of Cavalletti and Huesmann [CH15], see Section 5.4.

5.2 Proof of Proposition 17

Argue by contradiction, by assuming that there is a compact set $A \subset \mathcal{M}^R$ of positive Lebesgue measure such that

$$\forall x \in A,\ \ \Gamma^R(x) \text{ is not a singleton.} \tag{5.3}$$

Without loss of generality, we may assume that we work in $\mathbb{R}^4$.

For every $k \in \mathbb{N}$, we define the set

$$W_k := \Big\{ x \in \mathcal{M} \,|\, \exists p_x \in \mathbb{R}^4; |p_x| \leq k \text{ and}$$
$$\varphi(x) \leq \varphi(z) - \langle p_x, x - z\rangle + k\, |x-z|^2,\ \forall z \in \bar{B}(x, 1/k) \Big\}, \quad (5.4)$$

where $\bar{B}(x, 1/k)$ denotes the closed ball in $\mathbb{R}^4$ centered at x with radius $1/k$. The set W_k is well-defined, up to a change of coordinates, for k large enough.

Lemma 8. $\mathcal{M}^R \subset \bigcup_{k\in\mathbb{N}} W_k$.

Proof of Lemma 8. Let $\bar{x} \in \mathcal{M}^R$ and $\bar{y} \in \Gamma^R(\bar{x})$. By the same argument used in the proof of Proposition 15, we may assume that there are a regular minimizing geodesic $\bar{\gamma} : [0,1] \to M$ steering $\bar{y}$ to $\bar{x}$, and an open neighborhood $\mathcal{V}$ of $\bar{\gamma}([0,1])$ admitting an orthonormal family (with respect to g) $\mathcal{F}$ of two vector fields X^1, X^2 such that

$$\Delta(z) = Span\Big\{ X^1(z), X^2(z) \Big\},\ \forall z \in \mathcal{V}.$$

According to a change of coordinates if necessary, we can assume that $\mathcal{V}$ is an open subset of $\mathbb{R}^4$. Moreover, there is a control $\bar{u} \in L^2([0,1], \mathbb{R}^2)$ associated to $\bar{\gamma}$, ie.

$$\dot{\bar{\gamma}}(t) = \sum_{i=1}^{2} \bar{u}_i(t) X^i(\bar{\gamma}(t)),\ \forall t \in [0,1].$$

Since $\bar{\gamma}$ is regular, there exist $v^1, v^2, v^3, v^4 \in L^2([0,1], \mathbb{R}^2)$ such that the linear operator

$$\begin{array}{ccc} \mathbb{R}^4 & \to & \mathbb{R}^4 \\ \alpha & \mapsto & \sum_{i=1}^{4} \alpha_i D_{\bar{u}} End^{\bar{y}}(v^i) \end{array} \quad (5.5)$$

is invertible. Recall that $C^\infty([0,1], \mathbb{R}^2)$ is dense in $L^2([0,1], \mathbb{R}^2)$, we can assume that we have v^1, v^2, v^3, v^4 in $C^\infty([0,1], \mathbb{R}^2)$.

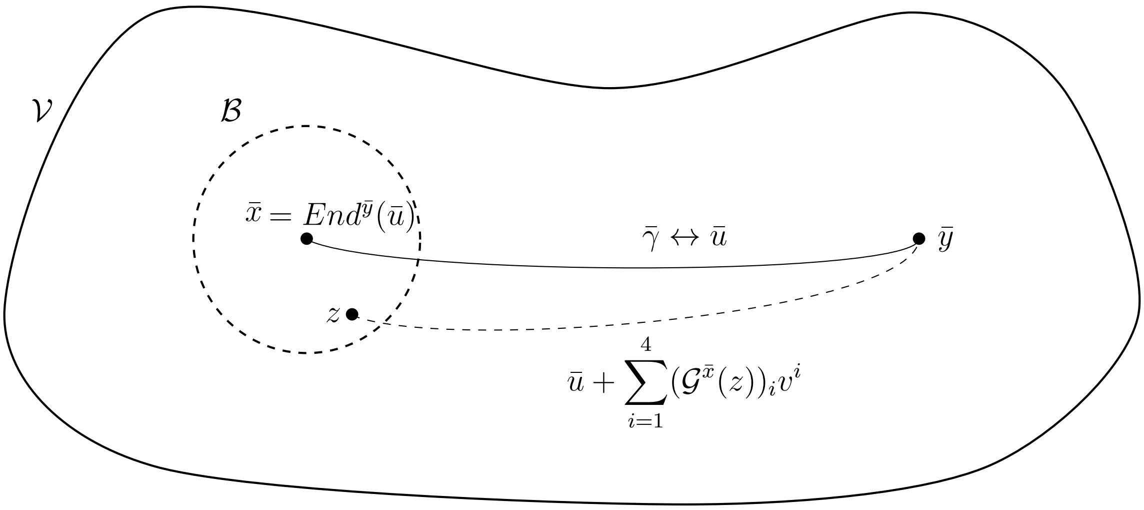
Define locally

$$\begin{array}{cccc} \mathcal{F}^{\bar{x}} : & \mathbb{R}^4 & \to & \mathbb{R}^4 \\ & \alpha & \mapsto & End^{\bar{y}}(\bar{u} + \sum_{i=1}^{4} \alpha_i v^i) \end{array} .$$

This mapping is well-defined and of class C^2 in the neighborhood of zero. It satisfies $\mathcal{F}^{\bar{x}}(0) = \bar{x}$ and its differential at 0 is invertible.
By the Local Inverse Function Theorem, there exist an open ball $\mathcal{B}$ of $\mathbb{R}^4$ centered at $\bar{x}$ and a function $\mathcal{G}^{\bar{x}} : \mathcal{B} \to \mathbb{R}^4$ of class C^2 such that

$$\mathcal{F}^{\bar{x}} \circ \mathcal{G}^{\bar{x}}(z) = z, \ \forall z \in \mathcal{B}.$$

$$\forall z \in \mathcal{B}, \ d_{SR}^2(z, \bar{y}) \leq ||\bar{u} + \sum_{i=1}^{4} (\mathcal{G}^{\bar{x}}(z))_i v^i||_{L^2}^2.$$



Define

$$\phi^{\bar{x},\bar{y}}(z) := ||\bar{u} + \sum_{i=1}^{4} (\mathcal{G}^{\bar{x}}(z))_i v^i||_{L^2}^2, \forall z \in \mathcal{B}.$$

Then, we conclude that there is a C^2 function $\phi^{\bar{x},\bar{y}} : \mathcal{B} \to \mathbb{R}^4$ such that

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

Recall that, by the definition of the Kantorovitch potentials, for every $z \in M$, we have

$$\begin{cases} \varphi(z) & \geq & \varphi^c(\bar{y}) - d_{SR}^2(z, \bar{y}) \\ \varphi(\bar{x}) & = & \varphi^c(\bar{y}) - d_{SR}^2(\bar{x}, \bar{y}) \end{cases}.$$

Then, $\forall z \in \mathcal{B}$,

$$\begin{cases} \varphi(z) & \geq & \varphi^c(\bar{y}) - \phi^{\bar{x},\bar{y}}(z) \\ \varphi(\bar{x}) & = & \varphi^c(\bar{y}) - \phi^{\bar{x},\bar{y}}(\bar{x}) \end{cases}.$$

Define
$$\psi^{\bar{x},\bar{y}}(z) := \varphi^c(\bar{y}) - \phi^{\bar{x},\bar{y}}(z), \forall z \in \mathcal{B}.$$
Hence, we put locally a C^2 function under the graph of φ with a uniform control on the C^2 norm of $\psi^{\bar{x},\bar{y}}$. Then, for $\bar{x} \in \mathcal{M}^R$, we can find $k \in \mathbb{N}$ such that there is $p_{\bar{x}} \in \mathbb{R}^4$ with $|p_{\bar{x}}| \leq k$ verifying
$$\varphi(\bar{x}) \leq \varphi(y) - \langle p_{\bar{x}}, \bar{x} - y\rangle + k\ |\bar{x} - y|^2, \quad \forall y \in \bar{B}(\bar{x}, 1/k).$$
$\square$

We are ready to complete the proof of Proposition 17.

Since $\mathcal{M}^R \subset \bigcup_{k\in\mathbb{N}} W_k$ (by Lemma 8), there exists $k \in \mathbb{N}$ such that
$$A_k := A \cap W_k \text{ is of positive Lebesgue measure.}$$
Let $\bar{x}$ be a density point of A_k and $\bar{y} \in \Gamma^R(\bar{x})$. By the definition of the Kantorovitch potentials, we have that
$$\varphi(\bar{x}) + d^2_{SR}(\bar{x}, \bar{y}) \leq \varphi(z) + d^2_{SR}(z, \bar{y}), \forall z \in M$$
$$\Rightarrow \varphi(\bar{x}) + d^2_{SR}(\bar{x}, \bar{y}) - \varphi(z) \leq d^2_{SR}(z, \bar{y}), \forall z \in M.$$

We define the function
$$\begin{array}{rccl} \rho^{\bar{x}}: & M & \to & \mathbb{R} \\ & z & \mapsto & \rho^{\bar{x}}(z) := \varphi(\bar{x}) + d^2_{SR}(\bar{x}, \bar{y}) - \varphi(z) \end{array}$$
verifying
$$\rho^{\bar{x}}(z) \leq d^2_{SR}(z, \bar{y}), \forall z \in M \text{ and equality for } z = \bar{x}. \tag{5.6}$$

Let $\tilde{A}_k := A_k \cap B(\bar{x}, 1/2k)$. For every $y \in \tilde{A}_k$, there is $p_y \in \mathbb{R}^4$, $|p_y| \leq k$ such that
$$\varphi(y) \leq \varphi(z) - \langle p_y, y - z\rangle + k\ |y - z|^2,\ \forall z \in B(y, 1/k).$$

We define the function $\tilde{\varphi} : B(\bar{x}, 1/2k) \to \mathbb{R}$ as follows

$$\tilde{\varphi}(x) = \sup_{y \in \tilde{A}_k} \Psi_y(x), \ \forall y \in B(\bar{x}, 1/2k)$$

where

$$\forall y \in \tilde{A}_k, \ \Psi_y(x) := \varphi(y) + \langle p_y, y - x \rangle - k \ |y - x|^2.$$

We claim that for every $x \in \tilde{A}_k$, $\tilde{\varphi}(x) = \varphi(x)$. Let us prove our claim. In fact, for every $x \in \tilde{A}_k$,we have

$$\tilde{\varphi}(x) \geq \Psi_y(x), \ \forall y \in \tilde{A}_k,$$

that is

$$\tilde{\varphi}(x) \geq \varphi(y) + \langle p_y, y - x \rangle - k \ |y - x|^2, \ \forall y \in \tilde{A}_k.$$

In particular, for $y = x \in \tilde{A}_k$, we obtain

$$\varphi(x) \leq \tilde{\varphi}(x).$$

Assume that there is $x \in \tilde{A}_k$ such that $\varphi(x) < \tilde{\varphi}(x)$.
Then, there is $y \in \tilde{A}_k$, $y \neq x$ such that

$$\varphi(x) < \ \Psi_y(x)$$

that is

$$\varphi(x) < \ \varphi(y) + \langle p_y, y - x \rangle - k \ |y - x|^2. \tag{5.7}$$

Since $x, y \in \tilde{A}_k$, we have $x \in B(y, 1/k)$.
So,

$$\varphi(y) \leq \varphi(x) - \langle p_y, y - x \rangle + k|x - y|^2$$

$$\Rightarrow \varphi(y) + \langle p_y, y - x \rangle - k \ |x - y|^2 \leq \varphi(x)$$

which contradicts inequality (5.7). And the conclusion follows.

Moreover, let $y \in \tilde{A}_k$ be fixed. There exists a neighborhood $B(y, 1/k)$ of y contained in $B(\bar{x}, 1/2k)$ such that for every $x \in B(y, 1/k)$, there is $\tilde{p}_x \in \mathbb{R}^4$ such that $\forall x' \in B(y, 1/k)$, we have

$$\begin{aligned}\Psi_y(x) - \Psi_y(x') &= \langle p_y, x' - x\rangle + k(|x' - y|^2 - |x - y|^2) \\ &\leq \langle p_y, x' - x\rangle + k|x' - x|^2 - 2k\langle y - x, x' - x\rangle \\ &\leq \langle p_y - 2k(y - x), x' - x\rangle + k|x' - x|^2\end{aligned}$$

Take $\tilde{p}_x := p_y - 2k(y - x)$, we obtain

$$\Psi_y(x) \leq \Psi_y(x') - \langle p_y - 2k(y - x), x' - x\rangle + k|x' - x|^2.$$

This means that for every $y \in \tilde{A}_k$, Ψ_y is locally semiconvex on $B(\bar{x}, 1/2k)$. According to Lemma 21 in Appendix A, since $\tilde{\varphi}$ is the supremum of local semiconvex functions Ψ_y among all $y \in \tilde{A}_k$, then $\tilde{\varphi}$ is locally semiconvex on $B(\bar{x}, 1/2k)$. By the Rademacher Theorem (see Appendix B.1), $\tilde{\varphi}$ is differentiable almost everywhere on $B(\bar{x}, 1/2k)$.

We also define the function

$$\begin{aligned}\tilde{\rho}^{\bar{x}}: \quad B(\bar{x}, 1/2k) &\to \mathbb{R} \\ z &\mapsto \tilde{\rho}^{\bar{x}}(z) := \tilde{\varphi}(\bar{x}) + d_{SR}^2(\bar{x}, \bar{y}) - \tilde{\varphi}(z)\end{aligned}$$

such that

$$\tilde{\rho}^{\bar{x}} = \rho^{\bar{x}} \text{ on } \tilde{A}_k. \tag{5.8}$$

Here, $\bar{x}$ is fixed and $\tilde{\rho}^{\bar{x}}$ is a function of z. By the definition of $\tilde{\rho}^{\bar{x}}$, as $\tilde{\varphi}$ is differentiable at almost every $z \in B(\bar{x}, 1/2k)$, $\tilde{\rho}^{\bar{x}}$ is also differentiable almost everywhere on $B(\bar{x}, 1/2k)$.

On the other hand, following the proof of Lemma 8, for $\bar{x} \in \mathcal{M}^R$ and $\bar{y} \in \Gamma^R(\bar{x})$, there are an open set $\mathcal{B}_{\bar{x}}$ in $\mathbb{R}^4$ containing $\bar{x}$ and a C^2 function $\phi^{\bar{x},\bar{y}} : \mathcal{B}_{\bar{x}} \to \mathbb{R}$ such that

$$\phi^{\bar{x},\bar{y}}(z) \geq d_{SR}^2(z, \bar{y}), \forall z \in \mathcal{B}_{\bar{x}} \text{ and equality for } z = \bar{x}. \tag{5.9}$$

Consequently, by (5.6), (5.8), (5.9), we obtain

$$\tilde{\rho}^{\bar{x}}(z) \leq d_{SR}^2(z, \bar{y}) \leq \phi^{\bar{x},\bar{y}}(z), \ \forall z \in \mathcal{B}_{\bar{x}} \cap \tilde{A}_k$$

and

equality for $z = \bar{x}$.

Note that $\phi^{\bar{x},\bar{y}}$ is a C^2 function and $\tilde{\rho}^{\bar{x}}$ is differentiable almost everywhere on $B(\bar{x}, 1/2k)$. Then,

$$d_{\bar{x}}\phi^{\bar{x},\bar{y}} = d_{\bar{x}}\tilde{\rho}^{\bar{x}}.$$

It means that there is a unique $\bar{y} \in \Gamma^R(\bar{x})$ such that $\bar{y}$ is characterized by

$$\bar{y} = exp_{\bar{x}}(d_{\bar{x}}\tilde{\rho}^{\bar{x}}) = exp_{\bar{x}}(-d_{\bar{x}}\tilde{\varphi}),$$

with $exp_{\bar{x}} : T^*_{\bar{x}}M \to M$ the sub-Riemannian exponential map from $\bar{x}$. This contradicts assumption (5.3) and the conclusion follows.

Remark 4. *The above argument can be used to prove the required result in the general case, with M a smooth connected manifold of dimension n equipped with a complete sub-Riemannian structure (Δ, g) of rank $m (m < n)$.*

5.3 Proof of Proposition 18

Our aim is to prove that

$$\text{for almost every } x \in \mathcal{M}^S,\ \Gamma^S(x) \text{ is a singleton.}$$

First, we need to construct a line field, defined on a set of full Lebesgue measure, whose orbits correspond to the singular curves.

The following holds (see [Sus96], [Rif14], [LS95]) :

Lemma 9. *There is an open set $\mathcal{H}$ of full Lebesgue measure on M such that:*

$$\forall x \in \mathcal{H},\ T_xM = \Delta(x) + [\Delta, \Delta](x) + [\Delta, [\Delta, \Delta]](x). \tag{5.10}$$

Proof of Lemma 9. We denote by $\mathscr{S}$ the set given by

$$\mathscr{S} = \Big\{x \in M | \Delta(x) + [\Delta, \Delta](x) + [\Delta(x), [\Delta, \Delta]](x) \neq T_xM\Big\}.$$

Assume by contradiction that $\mathscr{S}$ is of positive Lebesgue measure on M. It is sufficient to work locally. Taking a sufficiently small open neighborhood $\mathcal{V}$ of the origin in M and doing a change of coordinates if necessary we may assume that there are a set of coordinates (x_1, x_2, x_3, x_4) and two vector fields X^1, X^2 on $\mathcal{V}$ of the form

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

where $A, B : M \to \mathbb{R}$ are smooth functions such that $A(0) = B(0) = 0$ and

$$\Delta(x) = Span\Big\{X^1(x), X^2(x)\Big\}, \ \forall x \in \mathcal{V}.$$

So we have

$$[X^1, X^2] = A_{x_1}\partial_{x_3} + B_{x_1}\partial_{x_4} \text{ on } \mathcal{V}.$$

By hypothesis (5.1) in Theorem 12, we have

$$\forall x \in M, \Delta(x) + [\Delta, \Delta](x) \text{ has dimension } 3.$$

We may assume

$$A_{x_1} \neq 0 \ \text{ on } \mathcal{V}.$$

We denote by X^3 the vector field given by

$$X^3 := \frac{1}{A_{x_1}}[X^1, X^2] = \partial_{x_3} + C\partial_{x_4}$$

where $C := B_{x_1}/A_{x_1}$ is smooth.

A computation gives

$$\big[X^1, X^3\big] = [\partial_{x_1}, \partial_{x_3} + C\partial_{x_4}] = C_{x_1}\partial_{x_4} \tag{5.11}$$

and

$$\begin{aligned} \big[X^2, X^3\big] &= [\partial_{x_2} + A\partial_{x_3} + B\partial_{x_4}, \partial_{x_3} + C\partial_{x_4}] \\ &= \big(-A_{x_3} - CA_{x_4}\big)\partial_{x_3} \\ &\quad + \big(C_{x_2} + AC_{x_3} + BC_{x_4} - B_{x_3} - CB_{x_4}\big)\partial_{x_4} \end{aligned} \tag{5.12}$$

Let $x \in \mathscr{S} \cap \mathcal{V}$. It follows

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

Since $\Delta + [\Delta, \Delta]$ is of dimension 3, it means that

$$det\Big(X^1, X^2, [X^1, X^2], \big[X^1, [X^1, X^2]\big]\Big) = 0 \tag{5.13}$$

and,

$$det\Big(X^1, X^2, [X^1, X^2], \big[X^2, [X^1, X^2]\big]\Big) = 0 \tag{5.14}$$

which is equivalent to

$$det\Big(X^1, X^2, [X^1, X^2], \big[X^1, X^3\big]\Big) = 0 \tag{5.15}$$

and,

$$det\Big(X^1, X^2, [X^1, X^2], \big[X^2, X^3\big]\Big) = 0 \tag{5.16}$$

that is,

$$C_{x_1} = 0 \tag{5.17}$$

and

$$C_{x_2} + AC_{x_3} + BC_{x_4} - B_{x_3} - CB_{x_4} + CA_{x_3} + C^2 A_{x_4} = 0. \tag{5.18}$$

For every k-tuple $I = (i_1, \dots, i_{k-1}, 3)$ s.t. $(i_1, \dots, i_{k-1}) \in \{1, 2\}^{k-1}$, we denote by Z^I the smooth vector field constructed by the Lie brackets of X^1, X^2 as follows

$$Z^I = \Big[X^{i_1}, \big[X^{i_2}, \dots, [X^{i_{k-1}}, X^3]\big] \dots\Big].$$

Note $l(I)$ the length of the Lie brackets Z^I. By totally nonholonomicity, for every $x \in \mathcal{V}$, there exists an integer $r(x) \geq 2$ such that

$$T_x M = Span\Big\{Z^I(x) \mid \;\; l(I) \leq r(x)\Big\}.$$

For every I of $l(I) \geq 2$,

$$Z^I(x) = Z^I_3(x)\ \partial_{x_3} + Z^I_4(x)\ \partial_{x_4}.$$

We define the following set

$$\mathcal{A}_k := \Big\{x \in \mathcal{V} \mid \ Z^I_4(x) - C(x) Z^I_3(x) = 0 \;\; \forall I \ s.t. \ l(I) \leq k\Big\}$$

and

$$\mathscr{S} \cap \mathcal{V} = \bigcup_{k=2}^{r-1} \mathcal{A}_k \backslash \mathcal{A}_{k+1} \text{ where } r = \max_{x \in \mathcal{V}} \; r(x). \tag{5.19}$$

Recall that $\mathscr{S}$ is supposed to be of positive Lebesgue measure. By (5.19), there is $2 \le \bar{k} \le r-1$ such that $\mathcal{A}_{\bar{k}} \backslash \mathcal{A}_{\bar{k}+1}$ has positive Lebesgue measure. Fix $\bar{x}$ a density point in $\mathcal{A}_{\bar{k}} \backslash \mathcal{A}_{\bar{k}+1}$. There exists some $J' = (i_1, \dots, i_{\bar{k}}, 3)$ of length $\bar{k}+1$ such that on a neighborhood $\mathcal{V}_{\bar{x}}$ of $\bar{x}$,

$$Z_4^{J'} - CZ_3^{J'} \neq 0 \text{ on } \mathcal{V}_{\bar{x}}. \tag{5.20}$$

From J', we take $J = (i_2, \dots, i_{\bar{k}}, 3)$ of length $\bar{k}$, so that $Z_4^J - CZ_3^J = 0$. And, we compute $Z^{J'}$ in terms of Z^J:

$$Z^{J'} = \begin{cases} [X^1, Z^J] &= (Z_3^J)_{x_1} \partial_{x_3} + (Z_4^J)_{x_1} \partial_{x_4} \\ [X^2, Z^J] &= \Big((Z_3^J)_{x_2} + A(\bar{x})(Z_3^J)_{x_3} + B(Z_3^J)_{x_4} - Z_3^J A_{x_3} - Z_4^J A_{x_4} \Big) \partial_{x_3} \\ & \quad + \Big((Z_4^J)_{x_2} + A(Z_4^J)_{x_3} + B(Z_4^J)_{x_4} - Z_3^J B_{x_3} - Z_4^J B_{x_4} \Big) \partial_{x_4} \end{cases}$$

Replacing $Z_3^{J'}$ and $Z_4^{J'}$ in (5.20), it follows that on $\mathcal{V}$, we have

$$(Z_4^J)_{x_1} - C(Z_3^J)_{x_1} \neq 0 \tag{5.21}$$

or

$$\begin{aligned} &(Z_4^J)_{x_2} + A(Z_4^J)_{x_3} + B(Z_4^J)_{x_4} - Z_3^J B_{x_3} - Z_4^J B_{x_4} \\ &\qquad - C\big((Z_3^J)_{x_2} + A(Z_3^J)_{x_3} + B(Z_3^J)_{x_4} - Z_3^J A_{x_3} - Z_4^J A_{x_4} \big) \neq 0, \end{aligned} \tag{5.22}$$

We recall that $Z_4^J - CZ_3^J$ is a smooth function such that

$$Z_4^J - CZ_3^J = 0 \text{ on } \mathcal{A}_{\bar{k}} \backslash \mathcal{A}_{\bar{k}+1}. \tag{5.23}$$

Since $\bar{x}$ is a density point on $\mathcal{A}_{\bar{k}} \backslash \mathcal{A}_{\bar{k}+1}$, we have

$$\big(Z_4^J - CZ_3^J \big)_{x_i} (\bar{x}) = 0, \forall i = 1, 2, 3, 4.$$

Note that by (5.17), $C_{x_1} = 0$. And, by computing the partial derivatives of (5.23), we obtain

$$\big(Z_4^J\big)_{x_1}(\bar{x}) - C(\bar{x})\big(Z_3^J\big)_{x_1}(\bar{x}) = 0 \tag{5.24}$$

$$\big(Z_4^J\big)_{x_i}(\bar{x}) - C(\bar{x})\big(Z_3^J\big)_{x_i}(\bar{x}) = C_{x_i}(\bar{x})Z_3^J(\bar{x}), \quad \forall i = 2,3,4 \tag{5.25}$$

Using (5.25), we can check that the left-hand side of (5.22) evaluated at the point $\bar{x}$ is equal to

$$\begin{aligned}
&(Z_4^J)_{x_2} + A(Z_4^J)_{x_3} + B(Z_4^J)_{x_4} - Z_3^J B_{x_3} - Z_4^J B_{x_4} \\
&\qquad\qquad - C\Big((Z_3^J)_{x_2} + A(Z_3^J)_{x_3} + B(Z_3^J)_{x_4} - Z_3^J A_{x_3} - Z_4^J A_{x_4}\Big) \\
&= \Big((Z_4^J)_{x_2} - C(Z_3^J)_{x_2}\Big) + A\Big((Z_4^J)_{x_3} - C(Z_3^J)_{x_3}\Big) + B\Big((Z_4^J)_{x_4} - CZ_3^J)_{x_4}\Big) \\
&\qquad\qquad - Z_3^J\Big(B_{x_3} - CA_{x_3}\Big) - Z_4^J\Big(B_{x_4} - CA_{x_4}\Big) \\
&= C_{x_2}Z_3^J + AC_{x_3}Z_3^J + BC_{x_4}Z_3^J - Z_3^J\Big(B_{x_3} - CA_{x_3}\Big) - Z_4^J\Big(B_{x_4} - CA_{x_4}\Big) \\
&= \Big(C_{x_2} + AC_{x_3} + BC_{x_4} - B_{x_3} + CA_{x_3}\Big)Z_3^J - Z_4^J\Big(B_{x_4} - CA_{x_4}\Big) \\
&\underset{by(5.18)}{=} \Big(CB_{x_4} - C^2A_{x_4}\Big)Z_3^J - Z_4^J\Big(B_{x_4} - CA_{x_4}\Big) \\
&= -\Big(B_{x_4} - CA_{x_4}\Big)\Big(Z_4^J - CZ_3^J\Big) \\
&= 0
\end{aligned}$$

This and (5.24) imply that

$$Z_4^{J'}(\bar{x}) - C(\bar{x})Z_3^{J'}(\bar{x}) = 0$$

which contradicts (5.21) and (5.22),i.e. the fact that $\bar{x} \notin \mathcal{A}_{\bar{k}+1}$. □

We need another lemma.

Lemma 10. *There exists a line subbundle L of Δ such that the singular horizontal curves defined on $\mathcal{H}$ are exactly the trajectories described on L.*

Proof of Lemma 10. It is sufficient to prove the result in a neighborhood of each point in $\mathcal{H}$. So, let us consider a local frame $\{X^1, X^2\}$ such that

$$\Delta(z) = Span\{X^1(z), X^2(z)\}, \ \forall z \in M.$$

Let $\gamma : [0,1] \to M$ be a trajectory associated to some control $u \in L^2([0,1], \mathbb{R}^2)$. In local coordinates, singular curves can be characterized as follows (see Proposition 1.3.3 [Rif14]):

γ is singular with respect to Δ if there is $p : [0,1] \to (\mathbb{R}^4)^* \backslash \{0\}$ satisfying :

$$\dot{p}(t) = -\sum_{i=1}^{2} u_i(t) p(t).D_{\gamma(t)} X^i, \ a.e. \ t \in [0,1] \tag{5.26}$$

$$p(t).X^i(\gamma(t)) = 0, \forall t \in [0,1], \ \forall i = 1,2 \tag{5.27}$$

Derivative (5.27) two times yields for almost every $t \in [0,1]$ such that $u(t) \neq 0$

$$p(t).\Big[X^1(t), X^2(t)\Big](\gamma(t)) = 0, \tag{5.28}$$

and

$$u_1(t)p(t).\Big[X^1, [X^1, X^2]\Big](\gamma(t)) + u_2(t)p(t).\Big[X^2, [X^1, X^2]\Big](\gamma(t)) = 0. \tag{5.29}$$

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

$$\alpha_x.v = 0, \ \forall v \in \Delta(x) + \Big[\Delta, \Delta\Big](x), \ \forall x \in \mathcal{H}.$$

Then, by (5.27), (5.28)-(5.29), we infer that for almost every $t \in [0,1]$ such that $u(t) \neq 0$, we have:

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

By above assumption, for every $x \in \mathcal{H}$, the linear form

$$(\lambda_1, \lambda_2) \mapsto \Big(\alpha_x.[X^1, [X^1, X^2]](x)\Big)\lambda_1 + \Big(\alpha_x.[X^2, [X^1, X^2]](x)\Big)\lambda_2$$

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

We are ready now to prove Proposition 18. Without loss of generality, it is sufficient to prove the result locally. We can assume that (x_1, x_2, x_3, x_4) denotes the coordinates in an open neighborhood $\mathcal{V}$ in M and consider $\{X^1, X^2\}$ a local frame of Δ such that

$$\Delta(x) = Span\{X^1(x), X^2(x)\}, \forall x \in \mathcal{V}.$$

Doing a change of coordinates if necessary, we can assume that

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

where $A, B : \mathcal{V} \to \mathbb{R}$ are smooth functions.

For the upcoming results, it is important to keep in mind the following notations.

Notation 1. *We denote by A_{x_i}, B_{x_i} the partial derivative with respect to the variable x_i, and $A_{x_i x_j}, B_{x_i x_j}$ the second partial derivative with respect to the variable x_i and x_j, of A and B respectively.*

We compute the Lie brackets of X^1 and X^2 :

$$\Big[X^1, X^2\Big] = A_{x_1}\partial_{x_3} + B_{x_1}\partial_{x_4} \tag{5.30}$$

$$\Big[X^1, [X^1, X^2]\Big] = A_{x_1x_1}\partial_{x_3} + B_{x_1x_1}\partial_{x_4}$$

$$\Big[X^2, [X^1, X^2]\Big] = E\partial_{x_3} + F\partial_{x_4}$$

$$\textit{with} \begin{cases} E &= A_{x_2x_1} + AA_{x_3x_1} + BA_{x_1x_4} - A_{x_1}A_{x_3} - B_{x_1}A_{x_4}, \\ F &= B_{x_2x_1} + AB_{x_3x_1} + BB_{x_1x_4} - A_{x_1}B_{x_3} - B_{x_1}B_{x_4}. \end{cases}$$

By hypothesis (5.1) and (5.30), we can assume that

$$A_{x_1}(x) \neq 0, \ \forall x \in \mathcal{V}. \tag{5.31}$$

We denote by $\mathcal{H}^c$ the complementary set of $\mathcal{H}$ on M given by

$$\mathcal{H}^c = \Big\{x \in M| \ \Delta(x) + \Big[\Delta, \Delta\Big](x) + \Big[\Delta, [\Delta, \Delta]\Big](x) \neq T_xM\Big\}.$$

Thus, $\mathcal{H}^c$ is a closed set of Lebesgue measure zero on M.

The above discussion implies indeed the following lemma.

Lemma 11. *There exists an analytic horizontal vector field X given by*

$$X = \alpha_1 X^1 + \alpha_2 X^2$$

with $\alpha_1, \alpha_2 : \mathcal{V} \to \mathbb{R}$ smooth functions given by

$$\begin{cases} \alpha_1 &= & EB_{x_1} & - & FA_{x_1} \\ \alpha_2 &= & B_{x_1x_1}A_{x_1} & - & A_{x_1x_1}B_{x_1} \end{cases}$$

(E and $F : \mathcal{V} \to \mathbb{R}$ smooth functions defined in Notation 1).

The vector field X vanishes on $\mathcal{H}^c$ and any solution of the Cauchy problem $\dot{x}(t) = X(x(t))$ is analytic and singular.

Proof of Lemma 11. Let $T > 0$ and let $u \in L^2([0,1], \mathbb{R}^2)$ be a singular control and
$x : [0,T] \to M$ be a solution to the Cauchy problem

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

There exists an absolutely continuous arc $p : [0,T] \to (\mathbb{R}^4)^* \backslash \{0\}$ such that

$$\dot{p}(t) = -u_1(t)p(t).D_{x(t)}X^1 - u_2(t)p(t).D_{x(t)}X^2, \textit{a.e. } t \in [0,T] \tag{5.32}$$

$$p(t).X^1(x(t)) = p(t).X^2(x(t)) = 0, \forall t \in [0,T] \tag{5.33}$$

Taking the derivatives in (5.33) gives

$$p(t).[X^1, X^2](x(t)) = 0, \ \ \forall t \in [0,T] \tag{5.34}$$

which implies that $\forall t \in [0,T]$,

$$\begin{cases} p_1(t) = 0 \\ p_2(t) + A(x(t))p_3(t) + B(x(t))p_4(t) = 0 \\ A_{x_1}(x(t))p_3(t) + B_{x_1}(x(t))p_4(t) = 0 \end{cases}$$

Assume that condition (5.31) is true, then we obtain

$$p(t) = \Big(0, [A(x(t))\frac{B_{x_1}}{A_{x_1}}(x(t)) - B(x(t))]p_4(t), -\frac{B_{x_1}}{A_{x_1}}(x(t))p_4(t), p_4(t)\Big), \quad \forall t \in [0,T].$$

By taking the derivatives in (5.34), we obtain for every $t \in [0,T]$

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

$$\Rightarrow u_1(t)(p_3(t)A_{x_1x_1} + p_4(t)B_{x_1x_1}) + u_2(t)(p_3(t)E + p_4(t)F) = 0.$$

We can write

$$\begin{cases} u_1(t) &= & -(p_3(t)E + p_4(t)F) &= & -p_4(t)(F - \frac{B_{x_1}}{A_{x_1}}E) \\ u_2(t) &= & p_3(t)A_{x_1x_1} + p_4(t)B_{x_1x_1}) &= & p_4(t)(B_{x_1x_1} - A_{x_1x_1}\frac{B_{x_1}}{A_{x_1}}) \end{cases}.$$

Assume that $p_4(t) = 1, \forall t \in [0,1]$, we obtain

$$\begin{cases} \alpha_1(x) &= & EB_{x_1} - FA_{x_1} \\ \alpha_2(x) &= & A_{x_1}B_{x_1x_1} - B_{x_1}A_{x_1x_1} \end{cases} \tag{5.35}$$

□

Lemma 12. *There is a positive constant $C > 0$ such that*

$$div_x X \geq -C|X(x)|, \ \forall x \in \mathcal{V}.$$

Proof of Lemma 12. Let us compute the divergence of X. For every $x \in \mathcal{V}$,

$$\begin{aligned} div_x X &= \alpha_1(x) div_x X^1 + \alpha_2(x) div_x X^2 + X^1(\alpha_1) + X^2(\alpha_2) \\ &= \alpha_2(x) div_x X^2 + B_{x_1}(A_{x_1x_2x_1} + A_{x_1}A_{x_3x_1} + AA_{x_1x_3x_1} + B_{x_1}A_{x_1x_4} \\ &\quad +BA_{x_1x_1x_4} - A_{x_3}A_{x_1x_1} - A_{x_1}A_{x_1x_3} - B_{x_1x_1}A_{x_4} - B_{x_1}A_{x_1x_4}) \\ &\quad -A_{x_1}(B_{x_1x_2x_1} + A_{x_1}B_{x_3x_1} + AB_{x_1x_3x_1} + B_{x_1}B_{x_1x_4} + BB_{x_1x_1x_4} \\ &\quad -B_{x_3}A_{x_1x_1} - A_{x_1}B_{x_1x_3} - B_{x_1x_1}B_{x_4} - B_{x_1}B_{x_1x_4}) + EB_{x_1x_1} \\ &\quad -FA_{x_1x_1} + A_{x_2x_1}B_{x_1x_1} + A_{x_1}B_{x_2x_1x_1} - B_{x_2x_1}A_{x_1x_1} - B_{x_1}A_{x_2x_1x_1} \end{aligned}$$

$$+AA_{x_3x_1}B_{x_1x_1} + AA_{x_1}B_{x_3x_1x_1} - AB_{x_3x_1}A_{x_1x_1} - AB_{x_1}A_{x_3x_1x_1}$$

$$+BA_{x_4x_1}B_{x_1x_1} + BA_{x_1}B_{x_4x_1x_1} - BB_{x_4x_1}A_{x_1x_1} - BB_{x_1}A_{x_4x_1x_1}$$

$$= \alpha_2(x)div_xX^2 + EB_{x_1x_1} - FA_{x_1x_1}$$

$$+B_{x_1x_1}(BA_{x_4x_1} + AA_{x_3x_1} + A_{x_2x_1} + A_{x_1}B_{x_4} - B_{x_1}A_{x_4})$$

$$+A_{x_1x_1}(-BB_{x_4x_1} - AB_{x_3x_1} - B_{x_2x_1} + A_{x_1}B_{x_3} - B_{x_1}A_{x_3})$$

$$= \alpha_2(x)div_xX^2 + EB_{x_1x_1} - FA_{x_1x_1}$$

$$+B_{x_1x_1}A_{x_1}B_{x_4} + B_{x_1x_1}(E + A_{x_1}A_{x_3}) - A_{x_1x_1}B_{x_1}A_{x_3} - A_{x_1x_1}(F + B_{x_1}B_{x_4})$$

$$= \alpha_2(x)div_xX^2 + 2EB_{x_1x_1} - 2FA_{x_1x_1}$$

$$+B_{x_1x_1}(A_{x_1}B_{x_4} + A_{x_1}A_{x_3}) - A_{x_1x_1}(B_{x_1}A_{x_3} + B_{x_1}B_{x_4})$$

$$= \alpha_2(x)div_xX^2 + 2EB_{x_1x_1} - 2FA_{x_1x_1}$$

$$+(B_{x_1x_1}A_{x_1} - A_{x_1x_1}B_{x_1})(A_{x_3} + B_{x_4})$$

$$= 2\ B_{x_1x_1}E - 2\ A_{x_1x_1}F + 2\ \alpha_2(x)div_xX^2.$$

By (5.35), we can write $B_{x_1x_1} = \dfrac{\alpha_2 + B_{x_1}A_{x_1x_1}}{A_{x_1}}$ and $F = \dfrac{EB_{x_1} - \alpha_1}{A_{x_1}}$.

Hence, $div_xX = 2\ \alpha_2\dfrac{E}{A_{x_1}} + 2\ \alpha_1\dfrac{A_{x_1x_1}}{A_{x_1}} + 2\ \alpha_2 div_xX^2$

$$= 2\ \alpha_2\left(\frac{E}{A_{x_1}} + div_xX^2\right) + 2\ \alpha_1\frac{A_{x_1x_1}}{A_{x_1}}$$

As we noticed before, without loss of generality, we proceed as if M is a compact manifold. Then, the functions $\left(E/A_{x_1} + div_xX^2\right)$ and $\left(A_{x_1x_1}/A_{x_1}\right)$ are bounded on M. There exist $c_1, c_2 > 0$ such that

$$\left|\frac{A_{x_1x_1}}{A_{x_1}(x)}\right| \leq c_1 \text{ and } \left|\frac{E}{A_{x_1}}(x) + div_xX^2\right| \leq c_2,\ \forall x \in \mathcal{V}.$$

Thus,

$$div_xX \geq -c_1|\alpha_1| - c_2|\alpha_2|,\ \forall x \in \mathcal{V}$$

$$\geq -C|X(x)|, \forall x \in \mathcal{V}$$

with $C = \max\{c_1, c_2\} > 0$ positive constant. $\square$

The following process is equivalent to the process introduced by Belotto and Rifford [BR16] to set the contraction property.

Let $\varepsilon \in \{1, +1\}$ and $T > 0$, we denote by $(\varphi^X_{\varepsilon t})_{0\leq t\leq T}$ the analytic flow of the vector field X generating locally singular minimizing geodesics.

For every subset A in $\mathcal{V}$, we set

$$A^S_t = \varphi^X_{\varepsilon t}(A), \;\; \forall t \in [0, T] \text{ and } A^S_0 = A.$$

We denote by $l(A, A^S_t) := \sup_{x\in A} \; \mathit{length}\; \varphi^X_{\varepsilon t}(A) = \sup_{x\in A} \int_0^t |X(\varphi^X_{\varepsilon s}(x))| \mathrm{d}s,$

where $|X(\varphi^X_{\varepsilon s}(x))|$ stands for the norm of $X(\varphi^X_{\varepsilon s}(x))$ with respect to g.

We recall that there is $L > 0$, by (5.2), such that for every $x \in A$, we have

$$\int_0^t |X(\varphi^X_{\varepsilon s}(x))| \mathrm{d}s \leq L, \;\; \forall t \in [0, T]. \tag{5.36}$$

We state now divergence formulas, one of the main tool of the present paper (see [BR16], Proposition B.1).

Lemma 13. *For every compact A in M, there is a smooth function $\mathcal{J} : [0, T] \times A \to [0, +\infty[$ such that for every $t \in [0, T]$, we have:*

$$\mathcal{J}(0, z) = 1 \quad \mathit{and} \quad \frac{\partial \mathcal{J}}{\partial t}(t, z) = \mathit{div}\; X(\varphi^X_{\varepsilon t}(z))\; \mathcal{J}(t, z) \tag{5.37}$$

$$\forall x \in A, \;\; \mathcal{L}^4(A^S_t) = \int_{A^S_t} dz = \int_A \mathcal{J}(t, z)\; dz \tag{5.38}$$

and

$$\mathcal{L}^4(A^S_t) = \int_A \mathit{exp}\Big(\int_0^t \mathit{div}\; X(\varphi^X_{\varepsilon s}(z)) \;\; ds\Big)\; dz \tag{5.39}$$

The following result is an immediate corollary of Lemma 13.

Lemma 14. *Let $T > 0$. For every subset A in $\mathcal{V}$, we have*

$$\mathcal{L}^4(A_t^S) \;\geq\; exp(-C\ l(A, A_t^S))\ \mathcal{L}^4(A),\ \ \forall t \in [0, T]. \tag{5.40}$$

Proof of Lemma 14. Let A be a subset in $\mathcal{V}$. By Lemma 12, there is a constant $C > 0$ such that

$$div\ X(z) \geq -C|X(z)|,\ \ \forall z \in A.$$

Therefore, by (5.39), we infer that, $\forall t \in [0, T]$,

$$\begin{aligned}
\mathcal{L}^4(A_t^S) \;&\geq\; \int_A exp\Big(-C\int_0^t |X(\varphi_{\varepsilon s}^X(z))|\ \ ds\Big)\ dz \\
&\geq\; \int_A exp\Big(-C\ l(A, A_t^S)\Big)\ dz \\
&\geq\; exp\Big(-C\ l(A, A_t^S)\Big)\mathcal{L}^4(A).
\end{aligned}$$

□

The following result whose proof is based on the local contraction property (5.40), is fundamental.

Lemma 15. *Let $T > 0$. The closed set given by*

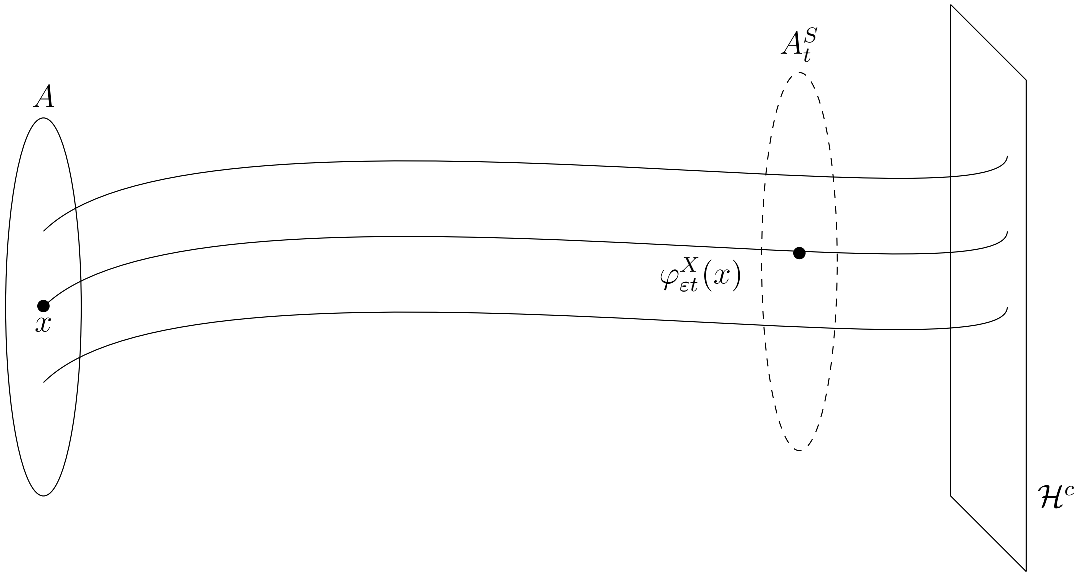
$$\Big\{x \in \mathcal{M}|\ \exists\gamma \in \Omega_{x,T}^S\ such\ that\ \gamma(T) \in \mathcal{H}^c\Big\}$$

is of Lebesgue measure zero on M.

Proof of Lemma 15. Let A be a subset of $\mathcal{M}$ of positive Lebesgue measure. Without loss of generality, we can assume that A is contained in an open set $\mathcal{V}$ in M. We argue by contradiction by assuming that

$$\mathcal{L}^4\Big(\big\{x \in A|\ \exists\gamma \in \Omega_{x,T}^S \text{ such that } \gamma(T) \in \mathcal{H}^c\big\}\Big) > 0.$$

By Lemma 11, there is an analytic horizontal vector field X defined on $\mathcal{V}$ generating singular minimizing geodesic defined on $\mathcal{V}$.



Moreover, X vanishes on $\mathcal{H}^c$. Then, for every $x \in A$, the flow of X starting at x requires an infinite time to reach $\mathcal{H}^c$, that is

$$A_t^S = \varphi_{\varepsilon t}^X(A) \underset{t\to\infty}{\longrightarrow} S \subset \mathcal{H}^c.$$

By Lemma (14), we have

$$\mathcal{L}^4(A_t^S) \geq \; exp(-C\; l(A, A_t^S))\mathcal{L}^4(A), \;\; \forall t \in [0, T].$$

By (5.36), we obtain

$$l(A, A_t^S) \leq L, \forall t \in [0, T].$$

Hence,

$$\mathcal{L}^4(A_t^S) \geq \; exp(-CL)\mathcal{L}^4(A), \;\; \forall t \in [0, T].$$

Since M is assumed to be compact and all the orbits $\varphi_{\varepsilon t}^Z(x)$ with $x \in A$ tends to S as t tends to ∞, by the Dominated Convergence Theorem, we deduce that

$$\lim_{t\to\infty} \mathcal{L}^4(A_t^S) = 0.$$

So we obtain

$$\mathcal{L}^4(A) = 0,$$

which implies the contradiction. □

In the spirit of [CH15], we have the following result.

Lemma 16. *Let Λ_1, Λ_2 be two subsets of Γ_φ such that* *(i)* $P^1(\Lambda_1) = P^1(\Lambda_2)$ *and* $P^1(\Lambda_i) \subset \mathcal{M}^S, \forall i = 1, 2.$

(ii) $P^2(\Lambda_1) \cap P^2(\Lambda_2) = \emptyset.$

Then, $\mathcal{L}^4(P^1(\Lambda_1)) = \mathcal{L}^4(P^1(\Lambda_2)) = 0.$

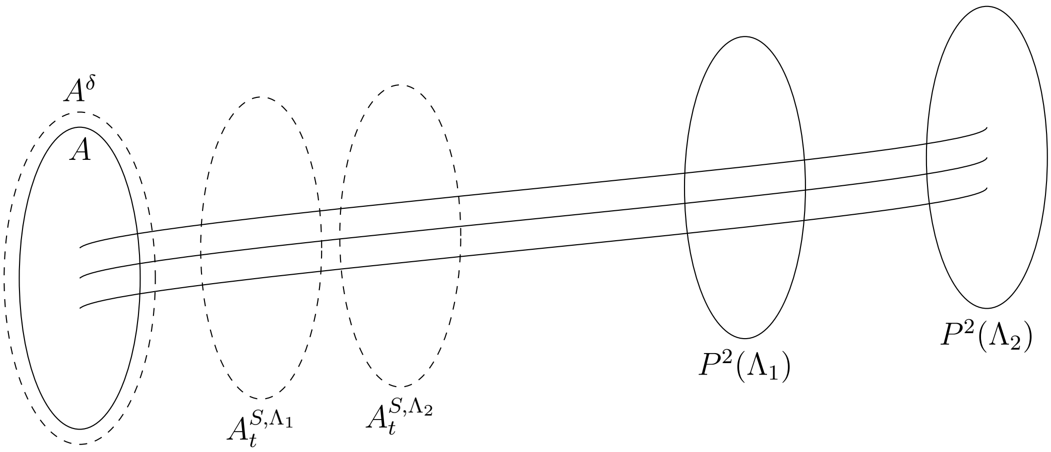
Proof of Lemma 16. Set $A = P^1(\Lambda_1) = P^1(\Lambda_2)$. We can assume that A is contained in an open set $\mathcal{V}$ in M. Let $T > 0$. For every $i = 1, 2$, we define

$$A_t^{S,\Lambda_i} := \big\{\varphi_{\varepsilon t}^X(x) |\ \varphi_0^X(x) \in A \text{ and } \varphi_{\varepsilon T}^X(x) \in P^2(\Lambda_i)\big\},\ \forall t \in [0, T].$$

Since $P^2(\Lambda_1) \cap P^2(\Lambda_2) = \emptyset$, we have

$$A_t^{S,\Lambda_1} \cap A_t^{S,\Lambda_2} = \emptyset, \forall t \in [0, T].$$

For $\delta > 0$ fixed, we define $A^\delta = \{x : d_{SR}(x, A) \leq \delta\}$.



$$\begin{aligned}
\mathcal{L}^4(A) &= \limsup_{\delta\to 0} \mathcal{L}^4(A^\delta) \\
&\geq \limsup_{t\to 0} \mathcal{L}^4(A_t^{S,\Lambda_1} \cup A_t^{S,\Lambda_2}) \\
&= \limsup_{t\to 0} [\mathcal{L}^4(A_t^{S,\Lambda_1}) + \mathcal{L}^4(A_t^{S,\Lambda_2})] \\
&\geq exp\Big(-C\ l(A, A_t^{S,\Lambda_1})) + exp\Big(-C\ l(A, A_t^{S,\Lambda_2}))\Big)\mathcal{L}^4(A).
\end{aligned}$$

Since $t \to 0$, we have A_t^{S,Λ_i} very close to A. So we can choose $l(A, A_t^{S,\Lambda_i})) > 0$ sufficiently small, that is

$$exp\Big(-C\ l(A, A_t^{S,\Lambda_i}))\Big) > \frac{1}{2}.$$

Hence, we obtain $\mathcal{L}^4(A) = 0$. □

We are ready to complete the proof of Proposition 18.

Consider the following set

$$E := \{x \in \mathcal{M}^S : \Gamma^S(x) \text{ is not a singleton}\}$$

and assume that E has positive measure.

It follows that there is $k \in \mathbb{N}$ such that the set given by

$$E_k := \{x \in E : \, diam \; \Gamma^S(x) > \frac{1}{k}\}$$

has positive Lebesgue measure.

Without loss of generality, we can assume that the manifold M can be covered by finitely many open balls $(\mathcal{U}_i)_{i\in I}$ of diameter less or equal to $1/k$. From $(\mathcal{U}_i)_{i\in I}$, we construct a finite family of open sets $(\mathcal{V}_i)_{i\in I}$ pairwise disjoint covering M by proceeding as follows

$$\begin{cases} \mathcal{V}_1 &= \mathcal{U}_1 \\ \mathcal{V}_2 &= \mathcal{U}_2 \backslash \mathcal{U}_1 \\ &\vdots \\ \mathcal{V}_n &= \mathcal{U}_n \backslash (\mathcal{U}_1 \cup \mathcal{U}_2 \cup \cdots \cup \mathcal{U}_{n-1}) \\ &\vdots \end{cases}$$

such that

$$\bigcup_{i\in I} \mathcal{U}_i = \bigcup_{i\in I} \mathcal{V}_i.$$

Therefore, for any $x \in E_k$, there are $i_x, j_x \in I$ with $i_x \neq j_x$ such that

$$\Gamma^S(x) \cap \mathcal{V}_{i_x} \neq \emptyset \text{ and } \Gamma^S(x) \cap \mathcal{V}_{j_x} \neq \emptyset.$$

Denote by

$$E_{k,i} := \bigcup_{x\in E_k} \{x\} \times (\Gamma^S(x) \cap \mathcal{V}_{i_x})$$

and

$$E_{k,j} := \bigcup_{x\in E_k} \{x\} \times (\Gamma^S(x) \cap \mathcal{V}_{j_x}).$$

We notice that $P^1(E_{k,i}) = P^1(E_{k,j}) = E$ such that

$$\mathcal{L}^4(E) > 0. \tag{5.41}$$

We also have $P^2(E_{k,i}) \cap P^2(E_{k,j}) = \emptyset$ since for any $x \in E_k$, $\mathcal{V}_{i_x} \cap \mathcal{V}_{j_x} = \emptyset$, for $i_x \neq j_x$. Using lemma 16, we obtain

$$\mathcal{L}^4(P^1(E_{k,i})) = \mathcal{L}^4(P^1(E_{k,j})) = \mathcal{L}^4(E) = 0,$$

which contradicts assumption (5.41).

We conclude that for a.e. $x \in \mathcal{M}^S, \Gamma^S(x)$ is a singleton.

5.4 End of the proof of Theorem 12

In the previous sections, we have shown that

$$a.e.\ x \in \mathcal{M}^R,\ \Gamma^R(x) \text{ is a singleton (see section 5.2)},$$

and

$$a.e.\ x \in \mathcal{M}^S,\ \Gamma^S(x) \text{ is a singleton (see section 5.3)}.$$

To complete the proof of Theorem 12, it remains to prove that

$$\mathcal{M}^S \cap \mathcal{M}^R \text{ has Lebsgue measure zero.}$$

For this purpose, we will use again the technique introduced by Cavalletti and Huesmann [CH15]. First, we will show a localized contraction property for regular horizontal curves.

Lemma 17. *There is a positive constant $\tilde{C}$ such that for $T > 0$ and for every set A in $\mathcal{M}^R$,*

$$\mathcal{L}^4(A_t^R) \geq \tilde{C}\mathcal{L}^4(A),\ \forall t \in [0, T] \tag{5.42}$$

with

$$A_t^R := \left\{\gamma(t) |\ \gamma \in \Omega_{x,T}^R;\ x \in A \text{ and } \gamma(T) \in \Gamma^R(x)\right\}.$$

Proof of Lemma 17. Let A be a compact set of $\mathcal{M}^R$ of positive measure. Since $\mathcal{M}^R \subset \bigcup_{k\in\mathbb{N}} W_k$ (by Lemma 8), for every point x of A, there exists $k = k(x) \in \mathbb{N}$ such that

$$x \in A_k := A \cap W_k,$$

so there is $p_x \in \mathbb{R}^4$ with $|p_x| \leq k$ verifying

$$\varphi(x) \leq \varphi(z) - \langle p_x, x - z\rangle + k|x-z|^2, \ \forall z \in B(x, 1/k).$$

Let $\tilde{A}_k := A_k \cap B(x, 1/2k)$. As in section 5.2, we define the function

$$\tilde{\varphi}(z) = \begin{cases} \varphi(z) & \text{if } z \in \tilde{A}_k \\ \\ \sup\limits_{y \in \tilde{A}_k} \{\varphi(y) + \langle p_y, y - z\rangle - k\ |y-z|^2\} & \text{if not} \end{cases}$$

For any $x \in A$, $\tilde{\varphi}$ is locally semiconvex on $B(x, 1/2k)$. By the Alexandrov Theorem, $\tilde{\varphi}$ is twice differentiable at a.e. $z \in B(x, 1/2k)$. Moreover, there exists a constant $C_k > 0$ such that

$$Hess_z\tilde{\varphi} \geq -C_k I_4, \ a.e.\ z \in B(x, 1/2k) \tag{5.43}$$

where I_4 is the 4×4 identity matrix.

We notice that $A = \bigcup\limits_{k \in \mathbb{N}} \tilde{A}_k$. Denote by $\tilde{C} > 0$ the constant given by

$$\tilde{C} := \sup_{k \in \mathbb{N}} C_k.$$

Then,

$$Hess_x\tilde{\varphi} \geq -\tilde{C} I_4, \ a.e.\ x \in A.$$

By section 5.2, for almost every $x \in A \subset \mathcal{M}^R$, there exists a unique $y \in \Gamma^R(x)$ given by

$$y := exp_x(-d_x\tilde{\varphi}).$$

Then, the curve $\gamma_x(t) : [0, T] \to M$ defined by

$$\gamma_x(t) := exp_x(-td_x\tilde{\varphi}), \ a.e.\ x \in A$$

is the unique regular minimizing geodesic joining x to y.

For every $t \in [0, T]$, we define the function

$$\begin{array}{rccl} T_t : & M & \to & M \\ & x & \mapsto & T_t(x) = \gamma_x(t) = exp_x(-td_x\tilde{\varphi}) \end{array}.$$

Note that, $\forall t \in [0, T]$, $A_t^R = \{T_t(z) : z \in A\}$ then we have

$$\mathcal{L}^4(A_t^R) = \int_{A_t^R} \mathrm{d}x = \int_{\{T_t(z); z \in A\}} \mathrm{d}x = \int_A det(Jac\ T_t(x))\mathrm{d}x. \tag{5.44}$$

However, the function T_t results from the composition of the two following functions

$$f : x \in M \to d_x\tilde{\varphi} \in T^*_x M, \text{ and } g : p \in T^*M \to exp_x(-tp) \in M.$$

By computing the Jacobien of T_t, we obtain

$$Jac\ T_t(x) = Jac\ g(f(x)) \times Hess_x\tilde{\varphi}\ .$$

Here, g is smooth on T^*M and by (5.43), there is a constant $\tilde{C} > 0$ such that

$$Jac\ T_t(x) \geq -\tilde{C}\ I_4,\ a.e.\ x \in A.$$

By (5.44), this implies

$$\mathcal{L}^4(A_t^R) \geq \tilde{C}\mathcal{L}^4(A), \forall t \in [0, T].$$

□

We conclude with the following lemma.

Lemma 18. *$\mathcal{M}^R \cap \mathcal{M}^S$ has Lebesgue measure zero on M.*

Proof of Lemma 18. Assume that there is a set A of $\mathcal{M}^R \cap \mathcal{M}^S$ such that

$$\mathcal{L}^4(A) > 0. \tag{5.45}$$

Let $T > 0$ and $\varepsilon \in \{-1, +1\}$. For every $t \in [0, T]$, we define the two following intermediate subsets by

$$A_t^R := \{\gamma_x(t) |\ \gamma_x \in \Omega^R_{x,T} \text{ with } x \in A \text{ and } \gamma^R_x(T) \in \Gamma^R(x)\},$$

and

$$A_t^S := \varphi^X_{\varepsilon t}(A).$$

We claim that for every $x \in A$, there is $t = t(x) \in]0, T[$ such that

$$\varphi^X_{\varepsilon s}(x) \neq \gamma_x(s), \forall s \in]t, T[.$$

As a matter of fact, regular minimizing geodesics are analytic as projections of the analytic sub-Riemannian Hamiltonian system and singular minimizing geodesic are analytic as the analytic flow of X. Assume that $\varphi^X_{\varepsilon T}(x) = \gamma_x(T)$. By the principle of isolated zeros for analytic functions, there is $t = t(x) \in]0, T[$ such that

$$\varphi^X_{\varepsilon s}(x) \neq \gamma_x(s), \forall s \in]t, T[.$$

Without loss of generality, we can assume that there is $\bar{t} \in]0, T[$ such that for every $x \in A$

$$t = t(x) \leq \bar{t} \text{ and } A_s^R \cap A_s^S = \emptyset, \ \forall s \in]\bar{t}, T[$$

and

$$A_{\bar{t}}^R \cap A_{\bar{t}}^S \neq \emptyset.$$

We denote by

$$\bar{A} := A_{\bar{t}}^R \cup A_{\bar{t}}^S.$$

We may assume that $\bar{A}$ has positive Lebesgue measure. Notice that for $s \geq \bar{t}$, when $s \to \bar{t}$, A_s^R and A_s^S converge to $\bar{A}$, then one has

$$\begin{aligned} \mathcal{L}^4(\bar{A}) = \lim_{\delta \to 0} \sup \mathcal{L}^4(\bar{A}^\delta) &\geq \lim_{s \to \bar{t}+} \sup \mathcal{L}^4(A_s^{\Lambda_1} \cup A_s^{\Lambda_2}) \\ &= \lim_{s \to \bar{t}+} \sup \mathcal{L}^4(A_s^R \cup A_s^S) \\ &= \lim_{s \to \bar{t}+} \sup [\mathcal{L}^4(A_s^R) + \mathcal{L}^4(A_s^S)] \\ &\geq \lim_{s \to \bar{t}+} \Big(\tilde{C} + exp\Big(-C \ l(\bar{A}, A_t^S)\Big)\Big) \mathcal{L}^4(\bar{A}). \end{aligned} \tag{5.46}$$

where $\bar{A}^\delta := \{x; d_{SR}(x, \bar{A}) \leq \delta\}$, for a given $\delta > 0$.

The inequality (5.46) follows from Lemmas 14 and 17 according to which we have

$$\mathcal{L}^4(A_s^R) \geq \tilde{C}\mathcal{L}^4(\bar{A}) \text{ and } \mathcal{L}^4(A_s^S) \geq exp\Big(-Cl(\bar{A}, A_t^S)\Big)\mathcal{L}^4(\bar{A}), \forall s \in]\bar{t}, T[.$$

As $s \to \bar{t}$, $l(\bar{A}, A_t^S)$ tends to zero. So we can choose $l(\bar{A}, A_t^S) > 0$ sufficiently small such that

$$\tilde{C} + exp\Big(-C \ l(\bar{A}, A_t^S)\Big) + \tilde{C} > 1.$$

It implies that $\mathcal{L}^4(\bar{A}) = 0$. And the conclusion follows. $\square$

Chapter 6

The study of h-concavity, h-semiconcavity and MCP on Carnot groups

A method introduced by Cavalletti and Huesmann [CH15] shows that we are able to prove existence and uniqueness of optimal transport maps on spaces satisfying the MCP. We recall that a sub-Riemannian structure is said to be ideal if it is complete and has no non-trivial singular minimizing curves. In [Rif13], Rifford proved that ideal sub-Riemannian structures on Carnot groups satisfy such property and this follows from the semiconcavity of the sub-Riemannian distance outside the diagonal. The aim of this section is to study suitable regularity assumptions guaranteeing the validity of the Cavalletti-Huesmann method for more general Carnot groups. Unfortunately, the content is prospective. We showed the MCP property on Carnot groups when the sub-Riemannian distance is assumed to be h-semiconcave. But until now, we have no examples of Carnot groups which are h-semiconcave.

6.1 Preliminaries on Carnot groups

We recall some basic facts on Carnot groups. For further details on Carnot groups, we refer the reader to [LeDo17].

A Carnot group $\mathcal{G}$ of step r is a simply connected Lie group whose Lie algebra $\mathfrak{g}$ admits a nilpotent stratification of step r. It means that $\mathfrak{g} = V_1 + \cdots + V_r$ with

$$[V_1, V_j] = V_{j+1}, \forall 1 \leq j \ \leq r, \quad V_r \neq \{0\}, \quad V_{r+1} = \{0\}.$$

We assume that a scalar product $<.,.>$ is given on $\mathfrak{g}$ for which the V_j's are mutually orthogonal. The assumption that $\mathcal{G}$ is simply connected and nilpotent ensures that the exponential map $\exp : \mathfrak{g} \to \mathcal{G}$ is a global diffeomorphism (see [Var74]). This allows to define the inverse of the exponential map given by the mapping

$$\begin{array}{rccl} \xi : & \mathcal{G} & \longrightarrow & \mathfrak{g} \\ & g & \longmapsto & \xi(g) = \xi_1(g) + \cdots + \xi_r(g) \end{array}$$

such that $\xi_i : \mathcal{G} \to V_i$, for $i = 1, \ldots, r$.
The identification of $\mathcal{G}$ and its Lie algebra $\mathfrak{g}$ is justified by the Baker-Campbell-Hausdorff formula

$$\exp(Z)\exp(Z') = \exp(Z + Z' + \frac{1}{2}[Z, Z'] + \dots), \ Z, Z' \in \mathfrak{g} \tag{6.1}$$

where the dots indicate a finite linear combination of terms containing commutators of order two and higher.

A Carnot group of step r is naturally equipped with a family of dilations defined by

$$\delta_\lambda(g) = \exp \circ \Delta_\lambda \circ \exp^{-1}(g), \ \forall g \in \mathcal{G}$$

where $exp : \mathfrak{g} \to \mathcal{G}$ is the exponential map and $\Delta_\lambda : \mathfrak{g} \to \mathfrak{g}$ is defined by
$\Delta_\lambda(v_1 + \cdots + v_r) = \lambda v_1 + \cdots + \lambda^r v_r$.

The first layer V_1 plays a key role. We denote $\exp(V_1) = \mathcal{H}_e$, where e is the unit element of the group $\mathcal{G}$. Assume that V_1 is of dimension m, we fix $\{X^1, \ldots, X^m\}$ an orthonormal basis of V_1. The first layer V_1 behaves as a sub-Riemannian structure on $\mathcal{G}$: we call horizontal directions its elements, and any metric on it provides a sub-Riemannian metric by translation. The homogeneity of the first layer implies the homogeneity of the sb-Riemannian distance, that is for every $\lambda > 0$,

$$d_{SR}(0, \delta_\lambda(g)) = \lambda d_{SR}(0, g), \ \forall g \in \mathcal{G}.$$

In particular, this yields the invarinace of the sub-Riemannian balls by dilations, that is for every $\lambda > 0$,

$$\delta_\lambda(B_{SR}(0, r)) = B_{SR}(0, \lambda r), \forall r > 0$$

where $B_{SR}(0, r)$ denotes the sub-Riemannian ball centered at the origin with radius r.

We say that an absolutely continuous curve $\gamma : [0, 1] \to \mathcal{G}$ is horizontal if $\dot{\gamma}(t) \in T\mathcal{H}_{\gamma(t)}, \ a.e. \ t \in [0, 1]$.

Given $g \in \mathcal{G}$, we denote the horizontal plane $\mathcal{H}_g$ by the m-dimensional submanifold of $\mathcal{G}$ passing through g given by

$$\mathcal{H}_g = \big\{ g' \in \mathcal{G} : \; g' = gh \text{ with } h \in \exp(V_1) \big\}.$$

We define another kind of curve joining two points $g, g' \in \mathcal{G}$.

Definition 29. *Given $g, g' \in \mathcal{G}$, for $\lambda \in [0,1]$, we denote by*

$$g_\lambda := g\delta_\lambda(g^{-1}g')$$

the twisted "convex combination" of g and g' based at g.

Given $g \in \mathcal{G}$, and $g' \in \mathcal{H}_g$, the map given by

$$\lambda \in [0,1] \longmapsto g_\lambda \in \mathcal{G}$$

is said to be a horizontal segment, and in particular, a geodesic.

Proposition 19. *Given $g, g' \in \mathcal{G}$, one has*

1. $g' \in \mathcal{H}_g \Leftrightarrow g^{-1}g' \in \mathcal{H}_e$
2. $g' \in \mathcal{H}_g \Leftrightarrow g \in \mathcal{H}_{g'}$
3. $g' \in \mathcal{H}_g \Rightarrow g_\lambda \in \mathcal{H}_g, \forall \lambda \in [0,1]$

Proof of Proposition 19.

1. $g' \in \mathcal{H}_g \Leftrightarrow g' = gh$ with $h \in exp(V_1) = \mathcal{H}_e$ which means that $g^{-1}g' = h \in \mathcal{H}_e$.
2. $g' \in \mathcal{H}_g \Leftrightarrow g^{-1}g' \in \mathcal{H}_e \Leftrightarrow g'^{-1}g \in \mathcal{H}_e \Leftrightarrow g \in \mathcal{H}_{g'}$.
3. For $\lambda \in [0,1]$, $g_\lambda = g\delta_\lambda(g^{-1}g') \Leftrightarrow g^{-1}g' = \delta_\lambda(g^{-1}g')$.
   If $g' \in \mathcal{H}_g$, then $g^{-1}g' \in \mathcal{H}_e \Rightarrow \delta_\lambda(g^{-1}g') \in \mathcal{H}_e \Rightarrow g_\lambda \in \mathcal{H}_g$.

□

We denote by Ω an open subset of $\mathcal{G}$. Given $i = 1, \dots, m$, the action of X^i on a function $f : \Omega \to \mathbb{R}$ is given by

$$X^i f(g) = \lim_{t \to 0} \frac{f(g \exp(tX^i)) - f(g)}{t} = \frac{d}{dt} f\big(g\exp(tX^i)\big)|_{t=0}.$$

Let k be a positive integer. We denote by $C^k_h(\Omega)$ the space of functions $f : \Omega \to \mathbb{R}$ which have continuous derivatives up to order k with respect to the horizontal vector fields $X^1, \dots, X^m$.

Definition 30. *Let $f : \Omega \to \mathbb{R}$ be a function of class C_h^2 on Ω.*

1. *The horizontal gradient of f at a point $g \in \Omega$ is the horizontal vector given by*

$$\nabla_h f(g) = \sum_{i=1}^{m} X^i f(g) X^i.$$

2. *The horizontal symmetrical hessian of f at a point $g \in \Omega$ is the matrix given by*

$$\left(\nabla_h^2 f\right)^*(g) = \frac{1}{2}\Big[X^i X^j f(g) + X^j X^i f(g)\Big]_{i,j=1,\dots,m}.$$

According to [CP10] and [DGS03], we introduce the following:

Definition 31. *Let $f : \Omega \to \mathbb{R}$ be a function. We call the Pansu differential of f at $g \in \Omega$ the map $Df(g) : \Omega \to \mathbb{R}$ given by*

$$Df(g)(h) = \lim_{\lambda \to 0^+} \frac{f\big(g\delta_\lambda(h)\big) - f(g)}{\lambda}, \forall h \in \Omega.$$

If $f \in C_h^1(\Omega)$, then the Pansu differential $Df(g)$ is given by

$$Df(g)(h) =< \nabla_h f(g), \xi_1(h) >, \forall h \in \Omega.$$

6.2 h-concavity on Carnot groups

Several notions of convexity on Heisenberg groups, and more generally in Carnot groups, have been introduced and compared as the horizontal convexity (see [DGN03], [CP10]) and the viscosity convexity (see [LMS04]). These definitions are proved to be equivalent on Carnot groups (see [BR03], [JLMS07], [Mag05] and [Wan05]).

Definition 32. *We say that a function $f : \mathcal{G} \to \mathbb{R}$ is h-concave on $\mathcal{G}$ if it is concave on every horizontal segment, that is,*

$$\lambda f(g') + (1-\lambda) f(g) \leq f\big(g\delta_\lambda(g^{-1}g')\big),$$

$\forall g \in \mathcal{G}, \forall g' \in \mathcal{H}_g, \forall \lambda \in [0,1]$.

Proposition 20. *Let $f : \mathcal{G} \to \mathbb{R}$ be a function of class C_h^2 on $\mathcal{G}$.*

1. *f is h-concave on $\mathcal{G}$ if and only if*

$$f(g') \leq f(g) + < \nabla_h f(g), \xi_1(g') - \xi_1(g) >, \ \forall g \in \mathcal{G}, \forall g' \in \mathcal{H}_g. \tag{6.2}$$

2. *f is h-concave on $\mathcal{G}$ if and only if*

$$\left(\nabla_h^2 f\right)^*(g) \leq 0, \ \forall g \in \mathcal{G}. \tag{6.3}$$

Proof of Proposition 20. 1. Since f is h-concave on $\mathcal{G}$, we have

$$\lambda f(g') + (1-\lambda) f(g) \leq f\big(g\delta_\lambda(g^{-1}g')\big),$$

$\forall g \in \mathcal{G}, \forall g' \in \mathcal{H}_g, \forall \lambda \in [0,1]$.

It follows that

$$f(g') - f(g) \ \leq \frac{f\big(g\delta_\lambda(g^{-1}g')\big) - f(g)}{\lambda}.$$

By making λ tends to zero, we obtain

$$f(g') - f(g) \leq Df(g)(g^{-1}g')$$

$$\Rightarrow f(g') \leq f(g) + < \nabla_h f(g), \xi_1(g^{-1}g') >$$

$$\Rightarrow f(g') \leq f(g) + < \nabla_h f(g), \xi_1(g') - \xi_1(g) >,$$

$\forall g \in \mathcal{G}, \forall g' \in \mathcal{H}_g$.

2. If f is twice differentiable with respect to $X^i, i, \ldots, m$, then for any $g \in \mathcal{G}$, $g' \in \mathcal{H}_g$, we have

$$f(g') = f(g) + < \nabla_h f(g), \xi_1(g') - \xi_1(g) > +$$
$$\frac{1}{2} < \left(\nabla_h^2 f\right)^*(g).(\xi_1(g') - \xi_1(g)), \xi_1(g') - \xi_1(g) > + o(|\xi_1(g') - \xi_1(g)|^2).$$

Since f is h-concave, we have by (6.2) for every $g \in \mathcal{G}$, $g' \in \mathcal{H}_g$

$$f(g') \leq f(g) + < \nabla_h f(g), \xi_1(g') - \xi_1(g) > .$$

Hence, $\forall g \in \mathcal{G}, \forall g' \in \mathcal{H}_g$,

$$< \left(\nabla_h^2 f\right)^*(g) \cdot (\xi_1(g') - \xi_1(g)), \xi_1(g') - \xi_1(g) > \leq 0$$

□

Remark 5. *Let $f : \mathcal{G} \to \mathbb{R}$ be an h-concave function on $\mathcal{G}$. We claim that the convolution of f by a mollifier sequence $(\varphi_\varepsilon)_{\varepsilon>0}$, defined by*

$$f_\varepsilon(g) := \varphi_\varepsilon * f(g) = \int_{\mathcal{G}} f(z^{-1}g)\varphi_\varepsilon(z)\mathrm{d}z, \forall g \in \mathcal{G}$$

is a sequence of smooth and h-concave functions on $\mathcal{G}$. In fact, for any $g \in \mathcal{G}$, $g' \in \mathcal{H}_g$, and any $\lambda \in [0,1]$, we have:

$$\begin{aligned}
f_\varepsilon\big(g\delta_\lambda(g^{-1}g')\big) &= \varphi_\varepsilon * f\big(g\delta_\lambda(g^{-1}g')\big)\\
&= \int_{\mathcal{G}} f\big(z^{-1}g\delta_\lambda(g^{-1}g')\big)\varphi_\varepsilon(z)\mathrm{d}z\\
&= \int_{\mathcal{G}} f\Big(z^{-1}g\delta_\lambda\big((z^{-1}g)^{-1}z^{-1}g'\big)\Big)\varphi_\varepsilon(z)\mathrm{d}z\\
&\geq \int_{\mathcal{G}} \varphi_\varepsilon(z)\big[\lambda f(z^{-1}g) + (1-\lambda)f(z^{-1}g')\big]\mathrm{d}z\\
&\geq \lambda \int_{\mathcal{G}} \varphi_\varepsilon(z)f(z^{-1}g)\mathrm{d}z + (1-\lambda)\int_{\mathcal{G}} \varphi_\varepsilon(z)f(z^{-1}g')\mathrm{d}z\\
&\geq \lambda f_\varepsilon(g') + (1-\lambda)f_\varepsilon(g).
\end{aligned}$$

6.2.1 First-order horizontal derivative of h-concave functions

Theorem 13. *An h-concave function $f : \mathcal{G} \to \mathbb{R}$ is Lipschitz with respect to the sub-Riemannian distance.*

$$\nabla_h f \in L^\infty_{loc}(\mathcal{G}).$$

Proof of Theorem 13. We denote by $f_\varepsilon := \varphi_\varepsilon * f$ the convolution of f by the mollifier sequence $(\varphi_\varepsilon)_{\varepsilon>0}$. By remark 5, $(f_\varepsilon)_\varepsilon$ is a sequence of smooth functions on $\mathcal{G}$ which are h-concave. Moreover, $(f_\varepsilon)_\varepsilon$ converges uniformly to f on every compact subset of $\mathcal{G}$. By inequality (6.2), we have for any $g \in \mathcal{G}$, $g' \in \mathcal{H}_g$,

$$f_\varepsilon(g') \leq f_\varepsilon(g) + < \nabla_h f_\varepsilon(g), \xi_1(g') - \xi_1(g) > . \tag{6.4}$$

Let $g_0 \in \mathcal{G}$. Fix $B_{SR}(g_0, R) \subset B_{SR}(g_0, 3R)$.
For every $g \in B_{SR}(g_0, R)$, and $g' \in \mathcal{H}_g \backslash \{0\}$, we have

$$\frac{f_\varepsilon(g') - f_\varepsilon(g)}{d_{SR}(g,g')} \leq \frac{< \nabla_h f_\varepsilon(g), \xi_1(g') - \xi_1(g) >}{d_{SR}(g,g')}.$$

Let us take now $g' \in \partial B_{SR}(g,R) \cap \mathcal{H}_g$ such that $0 < \varepsilon < R/2$:

$$\begin{aligned}
\frac{< \nabla_h f_\varepsilon(g), \xi_1(g) - \xi_1(g') >}{d_{SR}(g,g')} &\leq \frac{f_\varepsilon(g) - f_\varepsilon(g')}{R} \\
&\leq \frac{2}{R}||f_\varepsilon||_{L^\infty(B_{SR}(g_0,2R))} \\
&\leq \frac{2}{R}||f||_{L^\infty(B_{SR}(g_0,3R))}.
\end{aligned}$$

By taking the supremum over all $g' \in \partial B_{SR}(g,R) \cap \mathcal{H}_g$, we have

$$|\nabla_h f_\varepsilon(g)| \leq \frac{2}{R}||f||_{L^\infty(B_{SR}(g_0,3R))}.$$

Since g is arbitrary,

$$||\nabla_h f_\varepsilon \,||_{L^\infty(B_{SR}(g_0,R))} \leq \frac{2}{R}||f||_{L^\infty(B_{SR}(g_0,3R))}.$$

We denote by $Lip(f) = \frac{2}{R}||f||_{L^\infty(B_{SR}(g_0,3R))}$. Hence, for any $g, g' \in \bar{B}_{SR}(g_0,R)$, we have

$$|f_\varepsilon(g') - f_\varepsilon(g)| \leq Lip(f) d_{SR}(g,g').$$

Let ε tends to zero,

$$|f(g') - f(g)| \leq Lip(f)\ d_{SR}(g,g'),\ \forall g, g' \in \bar{B}_{SR}(g_0,R).$$

$\square$

Thanks to the Pansu-Rademacher Theorem (see Appendix B.2), an h-concave function is differentiable a.e. with respect to $X^i, i = 1, \ldots, m$.

6.2.2 Second-order horizontal derivative of h-concave functions

We deal with the existence almost everywhere of the second order horizontal derivative of h-concave functions (see [Mag05] and [GM04]).

Definition 33. (BV_h^2 ***functions***) *We say that a function $f \in L^1(\mathcal{G})$ has h-bounded second variation in $\mathcal{G}$ and we denote $f \in BV_h^2(\mathcal{G})$ if for any* $i = 1, \dots, m$,

$$\sup \left\{ \int_{\mathcal{G}} X^i f \; div \; \varphi \; \mathrm{d}x \mid \varphi \in C_c^1(\mathcal{G}), |\varphi| < 1 \right\} < \infty.$$

The following theorem extends a well-known property of concave functions. For sake of completeness, its proof is given in Appendix B.4.

Theorem 14. *Let $f : \mathcal{G} \to \mathbb{R}$ be an h-concave function. Then, f belongs to the class $BV_h^2(\mathcal{G})$.*

The following theorem corresponds to Theorem 3.9 in [AM03].

Theorem 15. *Let $f \in BV_h^2(\mathcal{G})$. Then for a.e. $g \in \mathcal{G}$, there exists a polynomial P_g of homogeneous degree less than or equal to 2 such that*

$$\lim_{r \to 0^+} \frac{1}{r^2 |B_{SR}(g, r)|} \int_{B_{SR}(g,r)} |f(y) - P_g(y)| \mathrm{d}y = 0.$$

The following theorem has been proved in [DGN03], [JLMS07] and [LMS04].

Theorem 16. *Let $f : \mathcal{G} \to \mathbb{R}$ be an h-concave function. Then, for every $g \in \mathcal{G}$ there exist $\delta > 0$ with $\bar{B}_{SR}(g, \delta)$ and a constant $C = C(g) > 0$ such that for every $r < \delta/15$ the following estimates hold*

$$\sup_{y \in B_{SR}(g,r)} |f(y)| \leq \frac{C}{|B_{SR}(g, r)|} \int_{B_{SR}(g,r)} |f(y)| \mathrm{d}y$$

and

$$||\nabla_h f||_{L^\infty(B_{SR}(g,r))} \leq \frac{C}{r |B_{SR}(g, r)|} \int_{B_{SR}(g,r)} |f(y)| \mathrm{d}y.$$

We give now the sub-Riemannian version of the Alexandrov-Busemann-Feller Theorem ([BF36], [Alex39]).

Theorem 17. *Let $f : \mathcal{G} \to \mathbb{R}$ be an h-concave function. Then for a.e. $g \in \mathcal{G}$ there exists a unique polynomial P_g of homogeneous degree less than or equal to 2 such that the following holds*

$$\lim_{y \to g} \frac{|f(y) - P_g(y)|}{d_{SR}(g, y)^2} = 0. \tag{6.5}$$

Proof of theorem 17. We recall by Theorem 14, that the h-concave function f is $BV_h^2(\mathcal{G})$. Then, by Theorem 15, for a.e. $g \in \mathcal{G}$, there exists a unique polynomial P_g of homogeneous degree ≤ 2 such that

$$\lim_{r \to o^+} \frac{1}{r^2} \frac{1}{|B_{SR}(g,r)|} \int_{B_{SR}(g,r)} |f - P_g| \mathrm{d}y = 0. \tag{6.6}$$

Let $g_0 \in \mathcal{G}$ be fixed such that (6.6) is satisfied. We set $v(y) = f(y) - P_{g_0}(y)$.
The polynomial P_{g_0} can be of the form $L + R$, that is the sum of a polynomial L of homogeneous degree ≤ 1 and a polynomial R of homogeneous degree equal to 2 or $R \equiv 0$. Moreover, we can write L of the form

$$L(g) = c + \sum_{j=1}^{m} \alpha_j g_j, \text{ with } c, \alpha_j \in \mathbb{R}, j = 1, \ldots, m.$$

We note that L et $-L$ are both h-concave. Hence, $w = f - L$ is also h-concave. We now have $v = f - P_{g_0} = f - L - R = w - R$.
There exist $c_1 > 0$ such that $\forall r > 0$,

$$\sup_{B_{SR}(g_0,r)} |\nabla_h R| \leq c_1 r \text{ and } \sup_{B_{SR}(g_0,r)} |R| \leq c_1 r^2$$

As w is h-concave, then $v + R$ is h-concave.
By Theorem 16, there exists $r_0 > 0$ such that

$$||\nabla_h v||_{L^\infty(B_{SR}(g_0,r))} \leq \frac{C}{r} \frac{1}{|B_{SR}(g_0, 15r)|} \int_{B_{SR}(g_0,15r)} |w(y)| \mathrm{d}y + \sup_{B_{SR}(g_0,r)} |\nabla_h R|$$

for any $0 < r < r_0$ such that $\overline{B_{SR}(g_0, 15r_0)} \subset \mathcal{G}$.

It follows that

$$\begin{aligned} ||\nabla_h v||_{L^\infty(B_{SR}(g_0,r))} \leq & \frac{C}{r} \frac{1}{|B_{SR}(g_0, 15r)|} \int_{B_{SR}(g_0,15r)} |v(y)| \mathrm{d}y + \sup_{B_{SR}(g_0,r)} |\nabla_h R| \\ & + \frac{C}{r} \frac{1}{|B_{SR}(g_0, 15r)|} \int_{B_{SR}(g_0,15r)} |R(y)| \mathrm{d}y \\ \leq & \frac{C}{r} \frac{1}{|B_{SR}(g_0, 15r)|} \int_{B_{SR}(g_0,15r)} |v(y)| \mathrm{d}y + c_1 r + C c_1 r \\ = & \frac{C}{r} \frac{1}{|B_{SR}(g_0, 15r)|} \int_{B_{SR}(g_0,15r)} |v(y)| \mathrm{d}y + c_1 r(1 + C). \end{aligned}$$

Let $\mathcal{Q}$ be the Hausdorff dimension of the Carnot groups $\mathcal{G}$. We may choose an arbitrary $\varepsilon \in]0, 1/2[$ and $\tau \in]0, \varepsilon^Q[$ such that

$$|\{y \in B(g_0, r) : |v(y)| \geq \varepsilon r^2\}| \leq (\varepsilon r^2)^{-1} \int_{B(g_0,r)} |v(y)| \mathrm{d}y \underset{r \to 0^+}{\longrightarrow} \varepsilon^{-1} o(r^Q).$$

Fix $r_1 < r_0$ depending on ϵ and τ such that

$$|\{y \in B_{SR}(g_0, r) : |v(y)| \geq \varepsilon r^2\}| < \tau |B(x, r)|, \forall \;\; 0 < r < r_1.$$

We take $y \in B_{SR}(g_0, \frac{r}{2})$ and $B_{SR}(y, \tau^{\frac{1}{Q}} r) \subset B_{SR}(g_0, \frac{r}{2})$. There is

$$z_r \in B(y, \tau^{\frac{1}{Q}} r) \text{ such that } |v(z_r)| < \varepsilon r^2, \forall \; r < r_1.$$

It implies for $y \in B(g_0, \frac{r}{2})$ and $z_r \in B(y, \tau^{\frac{1}{Q}} r)$

$$|v(y)| < \varepsilon r^2 + |v(z_r) - v(y)|.$$

Hence, for $r_2 < r_1/3$ such that

$$||\nabla_h v||_{L^\infty(B_{SR}(g_0,3r))} \leq Cr + 3(1 + C)c_1 r = c_2 r, \forall \; r < r_2.$$

We obtain for $y \in B(g_0, \frac{r}{2})$,

$$|v(y)| < \varepsilon r^2 + c_2 r d(z_r, y) < \varepsilon r^2 + c_2 \tau^{\frac{1}{Q}} r^2 < \varepsilon(1 + c_2) r^2.$$

As ε is arbitrary,

$$\lim_{r \to 0} \frac{|v(y)|}{r^2} = 0.$$

And the conclusion follows. □

6.3 h-semiconcavity on Carnot groups

Definition 34. *We say that a function $f : \mathcal{G} \to \mathbb{R}$ is h-semiconcave on $\mathcal{G}$ if it is semiconcave on every horizontal segment, that is, there exists $C > 0$ such that*

$$\lambda f(g') + (1 - \lambda) f(g) \leq f\big(g \delta_\lambda (g^{-1} g')\big) + \lambda(1 - \lambda) C |\xi_1(g') - \xi_1(g)|^2,$$

$\forall g \in \mathcal{G}, \forall g' \; \in \mathcal{H}_g, \forall \lambda \in [0, 1]$.

The constant C is called h-semiconcavity constant for f in Ω.

The following proposition is fundamental (see [DGN03]).

Proposition 21. *Let $g \in \mathcal{G}$. Then for any $g' \in \mathcal{H}_g$, one has*

$$\xi_1(g_\lambda) = (1-\lambda)\xi_1(g) + \lambda\xi_1(g') \tag{6.7}$$

with $g_\lambda := g\delta_\lambda(g^{-1}g')$, for every $\lambda \in [0,1]$.

Proof of Proposition 21. Via the Baker-Campbell-Hausdorff formula (see (6.1)), one has

$$\begin{aligned} g_\lambda &= g\delta_\lambda(g^{-1}g') \\ &= exp\big(\xi(g)\big)\ exp\big(\xi(\delta_\lambda(g^{-1}g'))\big) \\ &= exp\big(\xi(g) + \xi(\delta_\lambda(g^{-1}g')) + \frac{1}{2}\big[\xi(g), \xi(\delta_\lambda(g^{-1}g'))\big] + \dots\big). \end{aligned}$$

Since $g_\lambda \in \mathcal{H}_g$, we have $\delta_\lambda(g^{-1}g') \in \mathcal{H}_e$ which means that

$$\xi_i(\delta_\lambda(g^{-1}g')) = 0, \forall i = 2, \dots, r.$$

Then,

$$g_\lambda = exp\big(\xi_1(g) + \dots + \xi_r(g) + \xi_1(\delta_\lambda(g^{-1}g')) + \frac{1}{2}\big[\xi(g), \xi_1(\delta_\lambda(g^{-1}g'))\big] + \dots\big).$$

Hence,

$$\begin{aligned} \xi_1(g_\lambda) &= \xi_1(g) + \xi_1(\delta_\lambda(g^{-1}g')) \\ &= \xi_1(g) + \lambda\xi_1(g^{-1}g') \\ &= \xi_1(g) + \lambda\big(\xi_1(g') - \xi_1(g)\big). \end{aligned}$$

$\square$

From Proposition 21, we remark that an h-semiconcave function as given in Definition 34 can be regarded as a smooth perturbation of an h-concave function, that is it can be written as the sum of an h-concave function and a smooth one. More precisely,

$$f(g) = \Big(f(g) - C|\xi_1(g)|^2\Big) + C|\xi_1(g)|^2$$

with $g \mapsto f(g) - C|\xi_1(g)|^2$ an h-concave function.

Therefore, the h-semiconcave functions share all the regularity enjoyed by the h-concave functions.

Theorem 18. *Let $f : \mathcal{G} \to \mathbb{R}$ be an h-semiconcave function.*

(i) The function f is Lipschitz with respect to the sub-Riemannian distance. Thanks to the Pansu-Rademacher Theorem, f is differentiable a.e. with respect to $X^i, i = 1, \dots, m$.

(ii) ***(The sub-Riemannian version of the Alexandrov Theorem)*** *The function f is twice differentiable almost everywhere on $\mathcal{G}$ with respect to X^i, $i = 1, \dots, m$.*

We also have the following properties that relate the h-semiconcavity property of a function to its derivatives.

Proposition 22. *Let $f : \mathcal{G} \to \mathbb{R}$ be an h-semiconcave function with C as h-semiconcavity constant. Then, f satisfies the following properties:*

1. *For any $g \in \mathcal{G}$, $g' \in \mathcal{H}_g$,*

$$f(g') \leq f(g) + < \nabla_h f(g), \xi_1(g') - \xi_1(g) > +C|\xi_1(g') - \xi_1(g)|^2. \tag{6.8}$$

2. *For any $g \in \mathcal{G}$,*

$$\left(\nabla_h^2 f\right)^*(g) \leq C\ I_m \tag{6.9}$$

where I_m denotes the $m \times m$ identity matrix.

6.4 MCP on Carnot groups

Let $\mathcal{G}$ be a Carnot group of dimension n whose first layer V_1 has dimension m.

We define a class of sub-Riemannian structures, called h-ideal sub-Riemannian structures on Carnot groups.

Definition 35. *We say that a sub-Riemannian structure is h-ideal if it is complete and the sub-Riemannian distance d_{SR} is h-semiconcave on $(\mathcal{G} \times \mathcal{G}) \backslash D$, where D denotes the diagonal of $\mathcal{G} \times \mathcal{G}$.*

As in [Rif13], we define the horizontal cut-locus at a given $g \in \mathcal{G}$ as

$$cut_h(g) := \overline{\Sigma_h(d_{SR}(g, .))}$$

where $\Sigma_h(d_{SR}(g, .))$ denotes the set of points $g' \in \mathcal{G}$ such that the pointed distance $d_{SR}(g, .)$ is not differentiable at g' with respect to $X^i, i = 1, \dots, m$.

Without loss of generality, we proceed as if we work in $\mathbb{R}^n$ where $(x_1, \dots, x_n)$ denotes the local coordinates. Moreover, we fix $\{X^1, \dots, X^m\}$ an orthonormal basis of V_1. Up to a change of coordinates, we can assume that the vector fields X^i are of the form

$$X^1 = \partial_{x_1}, \text{ and } X^i = \partial_{x_i} + \sum_{j=m+1}^{n} \alpha_i^j \partial_{x_j}, \forall i = 2, \dots, m$$

with $\alpha_i^j \in C^\infty(M)$.

For any horizontal vector field $X := \sum_{i=1}^{m} a_i X^i$, we define the horizontal divergence of X, denoted by $div_h X$, as follows

$$div_h X := \sum_{i=1}^{m} X^i(a_i).$$

We make the following assumption.

<u>**ASSUMPTION 1**</u> For every $i = 2, \dots, m$,

$$X^i(\alpha_i^j) = 0, \forall j = m+1, \dots, n. \tag{6.10}$$

As the sub-Riemannian structure is invariant by translation, it is sufficiant to prove the result at the origin 0.

Proposition 23. *Let $\mathcal{G}$ be a Carnot group whose first layer is h-ideal and satisfies ASSUMPTION 1. Then, there is $N > 0$ such that for every measurable set*

$$A \subseteq B_{SR}(0,1) \backslash B_{SR}(0,1/2)$$

with $0 < \mathcal{L}^n(A) < +\infty$, we have

$$\mathcal{L}^n(A_s) \geq s^N \mathcal{L}^n(A), \ \forall s \in [1/2, 1]$$

where

$$A_s := \Big\{ \gamma(s) | \ \gamma : [0,1] \to \mathcal{G} \text{ minimizing geodesic with } \gamma(0) = 0, \gamma(1) \in A \backslash cut_h(0) \Big\}.$$

Proof of Proposition 23. Without loss of generality, we may assume that we work in $\mathbb{R}^n$. We denote by f the sub-Riemannian distance pointed at the origin 0, such that

$$\begin{array}{rccl} f: & \mathcal{G} & \to & [0, +\infty[\\ & g & \mapsto & f(g) := d_{SR}(0, g). \end{array}$$

Let $f_\varepsilon = \phi_\varepsilon * f$ be the convolution of f and the mollifier sequence $(\phi_\varepsilon)_\varepsilon$. We may note that f is h-semiconcave on $\mathcal{G}$ outside $cut_h\{0\}$. By remark 5, for $\epsilon > 0$ given, f_ε is smooth and h-semiconcave on $\mathcal{G}$ outside $cut_h\{0\}$. It follows that, by Proposition 22, there is a constant $C > 0$ such that

$$\big(\nabla_h^2 f_\varepsilon\big)^*(g) \leq C I_m, \text{ for a.e. } g \in \Big(\mathcal{G}\backslash cut_h\{0\}\Big) \cap \Big(B_{SR}(0,1)\backslash B_{SR}(0,1/2)\Big) \quad (6.11)$$

where I_m denotes the $m \times m$ identity matrix.

We denote by $\mathcal{Z}_\varepsilon$ the horizontal vector field defined by

$$\mathcal{Z}_\varepsilon(g) := -\nabla_h f_\varepsilon(g), \quad (6.12)$$

for a.e. $g \in \Big(\mathcal{G}\backslash cut_h\{0\}\Big) \cap \Big(B_{SR}(0,1)\backslash B_{SR}(0,1/2)\Big)$.

Let $g \in \mathcal{G}$ be fixed, and $\varphi_t^{\mathcal{Z}_\varepsilon}$ be the flow of $\mathcal{Z}_\varepsilon$ from g. For every measurable set $A \subseteq B_{SR}(0,1)\backslash B_{SR}(0,1/2)$, we denote by

$$A_t^\varepsilon := \varphi_{1-t}^{\mathcal{Z}_\varepsilon}(A_1), \ \forall t \in [1/2, 1]$$

where $A_1 = A\backslash cut_h\{0\}$.

To measure the variation of the volume along the trajectories of the flow $(\varphi_t^{\mathcal{Z}_\varepsilon})_t$, we have by the definition of the divergence

$$\frac{d}{dt}\Big\{\mathcal{L}^n\big(\varphi_t^{\mathcal{Z}_\varepsilon}(A_1)\big)\Big\} = \int_{\varphi_t^{\mathcal{Z}_\varepsilon}(A_1)} div\mathcal{Z}_\varepsilon(g)\, \mathrm{d}g, \quad \forall t \in [0, 1/2]. \quad (6.13)$$

We compute now the divergence of $\mathcal{Z}_\varepsilon$, for a.e. $g \in A_1$

$$\begin{aligned} div\mathcal{Z}_\varepsilon(g) &= div\ \big(\nabla_h f_\varepsilon(g)\big) \\ &= -div\ (\sum_{i=1}^m X^i f_\varepsilon(g)\ X^i(g) \\ &= -\sum_{i=1}^m X^i\big(X^i f_\varepsilon\big)(g) - \sum_{i=1}^m X^i f_\varepsilon(g)\ div\ X^i(g) \\ &= -div_h \nabla_h f_\varepsilon(g) - \sum_{i=1}^m X^i f_\varepsilon(g)\ div\ X^i(g) \end{aligned}$$

We claim that for a.e. $g \in A_1$, $div_h\nabla_h f_\varepsilon(g)$ is bounded from below. In fact, we have

$$\nabla_h f_\varepsilon = \sum_{i=1}^{m} \big(X^i f_\varepsilon\big) X^i = \big(\partial_{x_1} f_\varepsilon\big) X^1 + \sum_{i=2}^{m} \big(X^i f_\varepsilon\big) X^i$$
$$= \big(\partial_{x_1} f_\varepsilon\big) X^1 + \sum_{i=2}^{m} \big(\partial_{x_i} f_\varepsilon + \sum_{i=m+1}^{n} \alpha_i^j \partial_{x_j} f_\varepsilon\big) X^i$$

Then,

$$\begin{aligned}
div_h \nabla_h f_\varepsilon &= \sum_{i=1}^{m} X^i \big(X^i f_\varepsilon\big) \\
&= \partial^2_{x_1 x_1} f_\varepsilon + \sum_{i=2}^{m} X^i \big(X^i f_\varepsilon\big) \\
&= \partial^2_{x_1 x_1} f_\varepsilon + \sum_{i=2}^{m} \big(\partial_{x_i} + \sum_{j=m+1}^{n} \alpha_i^j \partial_{x_j}\big)\big(\partial_{x_i} f_\varepsilon + \sum_{l=m+1}^{n} \alpha_i^l \partial_{x_l} f_\varepsilon\big) \\
&= \partial^2_{x_1 x_1} f_\varepsilon + \sum_{i=2}^{m} \partial^2_{x_i x_i} f_\varepsilon + \sum_{i=2}^{m} \sum_{l=m+1}^{n} (\partial_{x_i} \alpha_i^l) \partial_{x_l} f_\varepsilon + 2 \sum_{i=2}^{m} \sum_{l=m+1}^{n} \alpha_i^l \partial^2_{x_i x_l} f_\varepsilon \\
&\quad + \sum_{i=2}^{m} \sum_{j=m+1}^{n} \sum_{l=m+1}^{n} \alpha_i^j (\partial_{x_j} \alpha_i^l) \partial_{x_l} f_\varepsilon + \sum_{i=2}^{m} \sum_{j=m+1}^{n} \sum_{l=m+1}^{n} \alpha_i^j \alpha_i^l \partial^2_{x_j x_l} f_\varepsilon \\
&= \mathrm{E} \ + \ \mathrm{F}
\end{aligned}$$

where

$$E = \partial^2_{x_1 x_1} f_\varepsilon + \sum_{i=2}^{m} \partial^2_{x_i x_i} f_\varepsilon + 2 \sum_{i=2}^{m} \sum_{l=m+1}^{n} \alpha_i^l \partial^2_{x_i x_l} f_\varepsilon + \sum_{i=2}^{m} \sum_{j=m+1}^{n} \sum_{l=m+1}^{n} \alpha_i^j \alpha_i^l \partial^2_{x_j x_l} f_\varepsilon$$

and

$$F = \sum_{i=2}^{m} \sum_{l=m+1}^{n} (\partial_{x_i} \alpha_i^l) \partial_{x_l} f_\varepsilon + \sum_{i=2}^{m} \sum_{j=m+1}^{n} \sum_{l=m+1}^{n} \alpha_i^j (\partial_{x_j} \alpha_i^l) \partial_{x_l} f_\varepsilon.$$

By (6.11), there is a constant $C > 0$ such that

$$E(g) \leq C, \ a.e. \ g \in A_1.$$

On the other hand, we have

$$
\begin{aligned}
F &= \sum_{i=2}^{m} \sum_{j=m+1}^{n} (\partial_{x_i} \alpha_i^j) \partial_{x_j} f_\varepsilon + \sum_{i=2}^{m} \sum_{j=m+1}^{n} \sum_{l=m+1}^{n} \alpha_i^j (\partial_{x_j} \alpha_i^l) \partial_{x_l} f_\varepsilon \\
&= \sum_{i=2}^{m} \Big(\sum_{l=m+1}^{n} (\partial_{x_i} \alpha_i^l) \partial_{x_l} f_\varepsilon + \sum_{l=m+1}^{n} \sum_{j=m+1}^{n} \alpha_i^j (\partial_{x_j} \alpha_i^l) \Big) \partial_{x_l} f_\varepsilon \\
&= \sum_{i=2}^{m} \sum_{l=m+1}^{n} \Big((\partial_{x_i} \alpha_i^l) + \sum_{j=m+1}^{n} \alpha_j^i (\partial_{x_j} \alpha_i^l) \Big) \partial_{x_l} f_\varepsilon \\
&= \sum_{i=2}^{m} \sum_{l=m+1}^{n} X^i (\alpha_i^l) \partial_{x_l} f_\varepsilon .
\end{aligned}
$$

By ASSUMPTION 1 (6.10), we have

$$\forall i = 2, \ldots, m, \;\; X^i(\alpha_i^l) = 0, \;\; \forall l = m+1, \ldots, n.$$

It means that $F = 0$.

Thus, we get

$$div_h \; \nabla_h f_\varepsilon(g) \leq C, \;\; a.e. \; g \in A_1.$$

And the claim follows.

Furthermore, f_ε is Lipschitz with respect to the sub-Riemannian distance. So there is a constant $C' > 0$ such that

$$\sum_{i=1}^{m} X^i f_\varepsilon(g) \; div \; X^i(x) \leq C', a.e. \; g \in A_1.$$

It follows that there is a constant $\bar{C} > 0$ such that

$$div \mathcal{Z}_\varepsilon(g) \geq -\bar{C}, a.e. \; g \in A_1.$$

Thanks to (6.13), we obtain

$$\frac{d}{dt} \Big\{ \mathcal{L}^n(A_{1-t}^\varepsilon) \Big\} \geq - \int_{\varphi_t^{\mathcal{Z}_\varepsilon}(A_1)} \bar{C} \mathrm{d}x = -\bar{C} \mathcal{L}^n(A_{1-t}^\varepsilon).$$

Using the Gronwall Lemma, it follows that there is $N > 0$ such that

$$\mathcal{L}^n(A_{1-t}^\varepsilon) \geq t^N \mathcal{L}^n(A_1), \;\; \forall t \in [0, 1/2].$$

Making ε tends to zero yields

$$\mathcal{L}^n(A_{1-t}) \geq t^N \mathcal{L}^n(A_1), \ \forall t \in [0, 1/2]$$

with

$$A_t = \varphi^{\mathcal{Z}}_{1-t}(A_1).$$

$\square$

Lemma 19. *Let $\mathcal{G}$ be a Carnot group whose first layer is h-ideal and satisfies ASSUMPTION 1. Then, there is $N > 0$ such that for every $k \in \mathbb{N}$ and for every measurable set*

$$A \subset B_{SR}(0, \frac{1}{2^k}) \backslash B_{SR}(0, \frac{1}{2^{k+1}})$$

with $0 < \mathcal{L}^n(A) < +\infty$, we have

$$\mathcal{L}^n(A_s) \geq s^N \mathcal{L}^n(A), \ \forall s \in [0, 1]$$

where

$$A_s := \Big\{ \gamma(s) | \ \gamma : [0,1] \to \mathcal{G} \text{ minimizing geodesic with } \gamma(0) = 0, \gamma(1) \in A \backslash cut_h(0) \Big\}.$$

Proof of Lemma 19. Let us take a measurable set $A \subset B_{SR}(0, 1/2^k) \backslash B_{SR}(0, 1/2^{k+1})$. By dilations properties, for every $k \in \mathbb{N}$, we have

$$\delta_{2^k}(A) \subset B_{SR}(0, 1) \backslash B_{SR}(0, 1/2)$$

and

$$\delta_{2^k}(A_s) = \big(\delta_{2^k}(A)\big)_s, \ \forall s \in [0, 1].$$

So It is sufficient to prove our property for a measurable set A such that

$$A \subset B_{SR}(0, 1) \backslash B_{SR}(0, 1/2).$$

Given $s \in]\frac{1}{4}, \frac{1}{2}[$, we set

$$B := \delta_{(2s)^{-1}}(A_{2s}) \subset B_{SR}(0, 1) \backslash B_{SR}(0, 1/2)$$

where $\delta_{(2s)^{-1}}$ is the dilation of factor $1/2s$.
Hence,

$$B_{\frac{1}{2}} := \big(\delta_{(2s)^{-1}}(A_{2s})\big)_{\frac{1}{2}}.$$

And we note that

$$B_{\frac{1}{2}} = \delta_{(2s)^{-1}}\big((A_{2s})_{\frac{1}{2}}\big) = \delta_{(2s)^{-1}}(A_s).$$

By Proposition 23, there is $N > 0$ such that for any $s \in]\frac{1}{4}, \frac{1}{2}[$

$$\begin{array}{rrcl} & \mathcal{L}^n(B_{\frac{1}{2}}) & \geq & \big(\frac{1}{2}\big)^N \mathcal{L}^n(B) \\ \Rightarrow & \mathcal{L}^n\big(\delta_{(2s)^{-1}}(A_s)\big) & \geq & \big(\frac{1}{2}\big)^N \mathcal{L}^n\big(\delta_{(2s)^{-1}}(A_{2s})\big) \\ \Rightarrow & (2s)^{-Q}\mathcal{L}^n(A_s) & \geq & \frac{1}{2^N}(2s)^{-Q}\mathcal{L}^n(A_{2s}) \\ \Rightarrow & \mathcal{L}^n(A_s) & \geq & \frac{1}{2^N}\ \mathcal{L}^n(A_{2s}) \end{array}$$

where Q is the homogeneous dimension of $\mathcal{G}$.

We repeat the above recursively and we obtain for any $k \in \mathbb{N}$

$$\mathcal{L}^n(A_s) \geq \big(\frac{1}{2^N}\big)^k\ \mathcal{L}^n(A_{2^k s}), \qquad \forall s \in]\frac{1}{2^{k+1}}, \frac{1}{2^k}[.$$

Let $s \in [0, 1]$, there is $k \in \mathbb{N}$ such that

$$s \in]\frac{1}{2^{k+1}}, \frac{1}{2^k}[\quad \text{and} \quad 2^k s \in]\frac{1}{2}, 1[.$$

Then, by Proposition 23, for any measurable set $A \subset B_{SR}(0, 1)\backslash B_{SR}(0, 1/2)$, we get

$$\mathcal{L}^n(A_s) \geq \big(\frac{1}{2^N}\big)^k\ \mathcal{L}^n(A_{2^k s}) \geq \big(\frac{1}{2^N}\big)^k (2^k s)^N \mathcal{L}^n(A) = s^N \mathcal{L}^n(A).$$

□

We claim that MCP defined on Carnot groups provides the non-branching condition so we can apply the Cavalletti-Huesmann method to prove existence and uniqueness of optimal transport maps.

Chapter 7

Conclusion and Perspectives

In this thesis, we were interested in the study of the Monge problem on sub-Riemannian structures, that is to prove existence and uniqueness for optimal transport maps. We restricted our attention to transportation problems between compactly supported probability measures from a smooth manifold into itself where the cost is given by the square of the sub-Riemannian distance. Two different methods enable to prove existence and uniqueness of optimal transport maps in sub-Riemannian geometry: the sub-Riemannian version of the Brenier-McCann theorems which requires regularity properties for d_{SR} and, the Cavalletti-Huesmann method covering, in particular, spaces satisfying the measure contraction property (MCP). Combining these two methods leads to prove existence and uniqueness of optimal transport maps on some sub-Riemannian structures admitting many singular minimizing geodesics. As seen in chapter 5, we treated the case of sub-Riemannian structures of rank two in dimension four. In chapter 6, we studied the concept of h-semiconcavity and MCP on Carnot groups. This study makes possible to apply the Cavalletti-Huesmann method on h-ideal sub-Riemannian structures on Carnot Groups.

Our framework raises many questions and perspectives, let us present them.

7.1 Influence of the cost

The regularity of the sub-Riemannian distance is central. In particular, in the proof of Proposition 23, the h-semiconcavity of d_{SR} plays a crucial role to establish the MCP on Carnot groups. The fact that d_{SR} is h-semiconcave provides a lower bound on its horizontal symmetrical hessian. We suggested to study such kind of regularity for d_{SR} on more general sub-Riemannian structures.

Theorem 19. *Let M be a manifold of dimension n equipped with a sub-Riemannian structure (Δ, g) of dimension m. Let $f : M \to \mathbb{R}$ be the squared sub-Riemannian distance pointed at the origin 0, given by*

$$f(x) := d_{SR}^2(0, x), \ \forall x \in M.$$

For any $x \in M$, there is a submanifold S_x tangent to the distribution at x such that f is semiconcave on S_x.

Proof of Theorem 19. Let $x \in M$ and $\gamma : [0,1] \to M$ be a minimizing geodesic joining o and x. There exists an open neighborhood $\mathcal{V}$ of $\gamma([0,1])$ in M. Without loss of generality, we can assume that $\mathcal{V}$ is an open subset of $\mathbb{R}^n$ and that there is an orthonormal family $\mathcal{F}$ of m smooth vector fields $X^1, \dots, X^m$ such that

$$\Delta(z) = Span\big\{X^1(z), \dots, X^m(z)\big\}, \quad \forall z \in \mathcal{V}.$$

Moreover, there is a control function $u^\gamma \in L^2([0,1], \mathbb{R}^m)$ such that

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

The End-point map associated to $\mathcal{F}$ at the origin is given by

$$\begin{array}{rccl} End^o : & L^2([0,1], \mathbb{R}^m) & \longrightarrow & \mathcal{G} \\ & u & \longmapsto & End^o(u) = \gamma_u(1). \end{array}$$

The End-point mapping End^o is of class C^1.

By Proposition 8, we have

$$X^i\big(End^0(u^\gamma)\big) \in D_{u^\gamma} End^0\big(L^2([0,1], \mathbb{R}^m)\big).$$

There exit $v^1, \dots, v^m \in L^2([0,1], \mathbb{R}^m)$ such that

$$D_{u^\gamma} End^o(v^i) = X^i(End^o(u^\gamma)), \ \forall i = 1, \dots, m.$$

We define the application $\mathcal{L} : \mathbb{R}^m \to M$ by

$$\begin{array}{rccl} \mathcal{L}_x : & \mathbb{R}^m & \to & M \\ & \alpha & \mapsto & \mathcal{L}_x(\alpha) := End^o\big(u^\gamma + \sum_{i=1}^m \alpha_i v^i\big). \end{array} \tag{7.1}$$

Here, $\mathcal{L}_x$ is of class C^1 in a neighborhood of the origin and $\mathcal{L}_x(0) = x$. Its differential at o is given by :

$$\begin{array}{rccl} D\mathcal{L}_{x\ |_{\alpha=0}}: & \mathbb{R}^m & \longrightarrow & T_{End^0(u^\gamma)}M \\ & \beta & \longmapsto & \sum_{i=1}^m \beta_i D_{u^\gamma} End^o(v^i) = \sum_{i=1}^m \beta_i X^i(End^o(u^\gamma)). \end{array}$$

As its differential at 0 is injective, then $\mathcal{L}_x$ is an immersion at $\alpha = 0$. Hence, the rank of the linear application $D\mathcal{L}_{x\ |_{\alpha=0}}$ is m, equal to the dimension of $\mathbb{R}^m$. It means that the image of a ball in the neighborhood of $\alpha = 0$ by the application $\mathcal{L}_x$ is a submanifold S_x of $\mathbb{R}^n$ of dimension m. Moreover, the tangent space to this submanifold at the point $x = \mathcal{L}_x(\alpha = 0)$ is the image of the differential of $D\mathcal{L}_{x\ |_{\alpha=0}}$. Thus, we obtain a submanifold S_x contained in $End^0(L^2([0,1], \mathbb{R}^m)$ and tangent to the distribution at x.

For every z in S_x, there is $\alpha \in \mathbb{R}^m$ such that

$$\mathcal{L}_x(\alpha) = z. \tag{7.2}$$

Since $\{X^i(End^o(u^\gamma))\}_{i=1}^m$ form an orthonormal basis of the distribution $\Delta\big(End^o(u^\gamma)\big)$, we may assume that $D\mathcal{L}_{x\ |_{\alpha=0}}$ is an invertible linear application. Thanks to the Local Inverse Theorem, there are a ball $\mathcal{B}$ centered at x in S_x and an application $\mathcal{J}_x : \mathcal{B} \to \mathbb{R}^m$ of class C^2 such that $\mathcal{L}_x \circ \mathcal{J}_x(z) = z$, $\forall z \in \mathcal{B}$.

Hence, for any $z \in \mathcal{B}$,

$$d^2_{SR}(o,z) = e_{SR}(o,z) \leq ||u^\gamma + \sum_{i=1}^m (\mathcal{J}_x(z))_i v^i||^2_{L^2}$$

and

$$d^2_{SR}(o,x) = e_{SR}(o,x) = ||u^\gamma||^2_{L^2}.$$

We set

$$\phi^{o,x}(z) := ||u^\gamma + \sum_{i=1}^m (\mathcal{J}_x(z))_i v^i||^2_{L^2},\ \forall z \in \mathcal{B}.$$

Then, there exists a function $\phi^{o,x}$ of class C^2 such that

$$f(z) \leq \phi^{o,x}(z), \forall z \in \mathcal{B} \text{ and } f(x) = \phi^{o,x}(x). \tag{7.3}$$

In fact, for any $x \in M$, we can construct a submanifold S_x tangent to the distribution at x such that for any point $y \in S_x$, we can put a support function $\phi^{0,x}$ of class C^2 on the graph of the function f. It means that f is semiconcave on the submanifold S_x tangent to $\Delta(x)$. □

With this type of regularity, it might be possible to have informations on the horizontal symmetrical Hessian of d_{SR} in order to obtain a contraction measure property similarly to the proof of Proposition 23 (see also [Rif13]).

7.2 The Cauchy problem for BV functions

The concept of the Cauchy problem for BV functions appears in the proof of Proposition 23. It seems unlikely that an h-semiconcave function is BV_h^2 (see Definition 33 for the definition of BV_h^2).

Let $\mathcal{G}$ be a Carnot group and $f : \mathcal{G} \to \mathbb{R}$ an h-semiconcave function. If we consider the horizontal vector field $Z := -\nabla_h f$, we don't know if the flow of Z exists. That is why, we proceed by creating a subsequence $(f_\varepsilon)_\varepsilon$ of smooth and h-semiconcave functions approximating f.

For instance, we thought it would be interesting to extend to the case of BV_h vector fields the method of Ambrosio [Amb04] (see Appendix C). For this purpose, an interesting work would be to extend the Diperna-Lions theory [DL89] to the case of BV_h vector fields.

Appendix A

Local semiconvexity

Let (Δ, g) be a sub-Riemannian structure of rank $m \leq n$ on the manifold M.

We recall here the definition of local semiconvexity of a given function.

Definition 36. *A function $f : \Omega \to \mathbb{R}$, defined on the open set $\Omega \subset M$, is called locally semiconvex on Ω if for every $x \in \Omega$ there exist a neighborhood Ω_x of x and a smooth diffeomorphism $\varphi_x : \Omega_x \to \varphi_x(\Omega_x)$ such that $f \circ \varphi_x^{-1}$ is locally semiconvex on the open subset $\tilde{\Omega}_x = \varphi_x(\Omega_x) \subset \mathbb{R}^n$.*

By the way, we recall that the function $\tilde{f} : \tilde{\Omega} \to \mathbb{R}$ is locally semiconvex on the open subset $\tilde{\Omega} \subset \mathbb{R}^n$ if for every $\bar{x} \in \tilde{\Omega}$, there exist $C, \delta > 0$ such that

$$f\Big(\lambda y + (1-\lambda)x\Big) \leq \lambda f(y) + (1-\lambda) f(x) + \lambda(1-\lambda)C|x-y|^2,$$

$$\forall \lambda \in [0,1], \forall x, y \in B(\bar{x}, \delta)$$

where $B(\bar{x}, \delta)$ is the open ball in $\mathbb{R}^n$ centered at $\bar{x}$ with radius δ.

The following result is useful to prove the local semiconvexity of a given function.

Lemma 20. *Let $f : \Omega \to \mathbb{R}$ be a function defined on an open set $\Omega \subset \mathbb{R}^n$. Assume that for every $\bar{x} \in \Omega$, there exist a neighborhood $\mathcal{V} \subset \Omega$ of $\bar{x}$ and a positive real number σ such that, for every $x \in \mathcal{V}$, there is $p_x \in \mathbb{R}^n$ such that*

$$f(x) \leq f(y) - \langle p_x, x - y\rangle + \sigma|x-y|^2, \ \forall y \in \mathcal{V}.$$

Then, the function f is locally semiconvex on Ω.

Proof of Lemma 20. Let $\bar{x} \in \Omega$ be fixed and $\mathcal{V}$ be the neighborhood given by assumption. Without loss of generality, we can assume that $\mathcal{V}$ is an open ball $\mathcal{B}$. Let $x, y \in \mathcal{B}$ and $\lambda \in [0,1]$. The point $\hat{x} := \lambda y + (1-\lambda)x$ belongs to $\mathcal{B}$. By assumption, there exists $\hat{p} \in \mathbb{R}^n$ such that

$$f(\hat{x}) \leq f(z) - \langle \hat{p}, \hat{x} - z \rangle + \sigma |\hat{x} - z|^2, \ \forall z \in \mathcal{B}.$$

Hence, we easily get

$$\begin{cases} f(\hat{x}) & \leq & f(x) - \lambda \langle \hat{p}, y - x \rangle + \sigma \lambda |x - y|^2 \\ f(\hat{x}) & \leq & f(y) - (1-\lambda) \langle \hat{p}, x - y \rangle + \sigma (1-\lambda) |x - y|^2 \end{cases}$$

$$\Rightarrow \begin{cases} (1-\lambda) f(\hat{x}) & \leq & (1-\lambda) f(x) - \lambda(1-\lambda) \langle \hat{p}, y - x \rangle + \sigma \lambda (1-\lambda) |x - y|^2 \\ \lambda f(\hat{x}) & \leq & \lambda f(y) + \lambda(1-\lambda) \langle \hat{p}, y - x \rangle + \sigma \lambda (1-\lambda) |x - y|^2 \end{cases}$$

$$\Rightarrow f(\hat{x}) \leq \lambda f(x) + (1-\lambda) f(y) + 2\sigma\lambda(1-\lambda)|x-y|^2$$

and the conclusion follows. □

Remark 6. *Thanks to Lemma 20, a way to prove that a given function $f : \Omega \to \mathbb{R}$ is locally semiconvex on Ω is to show that for every $x \in \Omega$, we can put a support function ϕ of class C^2 under the graph of f at x with a uniform control of C^2 norm of ϕ.*

Let us derive another important consequence of the definition of semiconvexity.

Lemma 21. *Let Ω be a subset of $\mathbb{R}^n$ and $\{u_\alpha\}_{\alpha \in \mathcal{A}}$ be a family of functions defined on Ω and locally semiconvex with C_α the semiconvexity constant. Then, the function $u := \sup_{\alpha \in \mathcal{A}} u_\alpha$ is also locally semiconvex on Ω.*

Proof of Lemma 21. Take $x, y \in \Omega$ and $\lambda \in [0,1]$ such that $\lambda y + (1-\lambda)x \in \Omega$. Given any $\varepsilon > 0$, we can find α such that

$$u(\lambda y + (1-\lambda)x) \leq u_\alpha(\lambda y + (1-\lambda)x) + \varepsilon.$$

Then we have $\delta_\alpha > 0$ such that $\forall y \in B(x, \delta_\alpha)$

$$u(\lambda y + (1-\lambda)x) - \lambda u(y) - (1-\lambda) u(x)$$

$$\leq u_\alpha(\lambda y + (1-\lambda)x) + \varepsilon - \lambda u_\alpha(y) - (1-\lambda)u_\alpha(x)$$

$$\leq \lambda(1-\lambda)C_\alpha|x-y|^2 + \varepsilon.$$

Since $\varepsilon > 0$ is arbitrary, we obtain the assertion. $\square$

More details of local semiconvexity of a given function are given in [CS04].

Appendix B

Geometric analysis

B.1 The Rademacher Theorem

The Rademacher Theorem [Rad20] states that real valued Lipschitz functions on $\mathbb{R}^n$ are differentiable almost everywhere with respect to the Lebesgue measure.

Theorem 20. *Let $f : \Omega \to \mathbb{R}$ be a Lipschitz function, where $\Omega \subseteq \mathbb{R}^n$ be open. Then, f is differentiable at a.e. $x \in \Omega$. That is the partial derivatives exist a.e. and*

$$\nabla f(x) = \Big(\frac{\partial f}{\partial x_1}(x), \dots, \frac{\partial f}{\partial x_n}(x) \Big)$$

satisfies

$$\lim_{y \to x} \frac{f(y) - f(x) - \nabla f(x).(y - x)}{|y - x|} = 0, \;\; a.e. \; x \in \Omega.$$

Lemma 22. *Let $g : \mathbb{R}^n \to \mathbb{R}$ be a smooth function such that*

$$\int_{\mathbb{R}^n} g(x)\varphi(x)\mathrm{d}x = 0, \;\; \forall \varphi \in C_c^\infty(\mathbb{R}^n).$$

Then, $g(x) = 0$, a.e. $x \in \mathbb{R}^n$.

Proof of Lemma 22. We set

$$E = \{x \in \mathbb{R}^n; g(x) \neq 0\}$$

and we assume by contradiction that E has positive Lebesgue measure.
For $l \in \mathbb{N}$, we define

$$E_l^+ = \{x \in \mathbb{R}^n; g(x) \geq \frac{1}{l}\}$$

and
$$E_l^- = \{x \in \mathbb{R}^n; g(x) \leq -\frac{1}{l}\}$$
such that
$$E = \bigcup_{l\in\mathbb{N}} E_l^+ \cup E_l^-.$$
Since $\mathcal{L}^n(E) > 0$, it follows that either $\mathcal{L}^n(E_l^+) > 0$ or $\mathcal{L}^n(E_l^-) > 0$. Assume that
$$\mathcal{L}^n(E_l^-) > 0.$$
Let $a \in \mathbb{R}^n$ be fixed. Thanks to the Lebesgue density Theorem, we have
$$\lim_{r\to 0} \frac{\mathcal{L}^n\Big(B(a,r)\cap E_l^-\Big)}{\mathcal{L}^n\Big(B(a,r)\Big)} = 1.$$
We may assume that there is $\bar{r} > 0$ such that for any $\bar{r} < r$,
$$\Big|1 - \frac{\mathcal{L}^n\Big(B(a,r)\cap E_l^-\Big)}{\mathcal{L}^n\Big(B(a,r)\Big)}\Big| < \frac{1}{100}.$$
Let $0 \leq \mu \leq \bar{r}$ and $\varphi \in C_c^\infty(\mathbb{R}^n)$ be the test function given by
$$\varphi(x) := \begin{cases} 1 & x \in B(a,\mu) \\ 0 \leq \varphi(x) \leq 1 & x \in B(a,\bar{r}) \\ 0 & \text{else} \end{cases}$$

Hence, we get

$$\begin{aligned}\int_{\mathbb{R}^n} g(x)\varphi(x)\mathrm{d}x = \int_{B(a,\mu)} g(x)\varphi(x)\mathrm{d}x + \int_{\mathbb{R}^n\backslash B(a,\bar{r})} g(x)\varphi(x)\mathrm{d}x \\ + \int_{B(a,\bar{r})\backslash B(a,\mu)} g(x)\varphi(x)\mathrm{d}x\end{aligned}$$

$$\begin{aligned}&\leq \int_{B(a,\mu)} g(x)\mathrm{d}x + \int_{B(a,\bar{r})\backslash B(a,\mu)} g(x)\mathrm{d}x \\ &\leq -\frac{1}{l}\mathcal{L}^n(B(a,\mu)) - \frac{1}{l}\mathcal{L}^n(B(a,\bar{r})\backslash B(a,\mu)).\end{aligned}$$

When μ tends to 0, we obtain

$$\int_{\mathbb{R}^n} g(x)\varphi(x)\mathrm{d}x \leq -\frac{1}{l}\mathcal{L}^n(B(a,\bar{r})).$$

Making l tends to ∞ yields

$$\int_{\mathbb{R}^n} g(x)\varphi(x)\mathrm{d}x = 0.$$

$\square$

We are now ready to prove Theorem 20.

For each $v \in \mathbb{R}^n$ with $|v| = 1$, we define the directional derivative in the direction v at $x \in \Omega$ by

$$D_v f(x) = \lim_{t\to 0} \frac{f(x+tv) - f(x)}{t}.$$

Let us prove that $D_v f(x)$ exists at a.e. $x \in \Omega$. We consider

$$\overline{D}_v f(x) = \limsup_{t\to 0} \frac{f(x+tv) - f(x)}{t}$$

and

$$\underline{D}_v f(x) = \liminf_{t\to 0} \frac{f(x+tv) - f(x)}{t}.$$

Let A_v be the set of points $x \in \Omega$ such that $D_v f(x)$ does not exist

$$A_v = \{x \in \mathbb{R}^n | \overline{D}_v f(x) \neq \underline{D}_v f(x)\}.$$

Since f is continuous, A_v is a measurable set. We claim that A_v has Lebesgue measure zero. Let us prove our claim. Let $v \in \mathbb{R}^n$ with $|v| = 1$. For any $x \in \Omega$, we define the function

$$\begin{array}{rccl} \varphi: & \mathbb{R} & \to & \mathbb{R} \\ & t & \to & \varphi(t) = f(x+tv). \end{array}$$

As f is Lipschitz on Ω, then φ is also Lipschitz on $\mathbb{R}$. It follows that φ is differentiable a.e. $t \in \mathbb{R}$. Hence, $\mathcal{L}^1(A_v \cap L) = 0$, for any line L parallel to the direction v.
Thanks to the Fubini Theorem, we obtain

$$\mathcal{L}^n(A_v) = 0$$

which implies that f est differentiable a.e. $x \in \Omega$ in the direction v.
We denote by

$$\mathcal{B}_v = \{x \in \Omega | D_v f(x) \text{ exists } \}.$$

Let $\{e_1, \dots, e_n\}$ be the canonical basis of $\mathbb{R}^n$, and let

$$\mathfrak{B}_v = \mathcal{B}_v \cap \mathcal{B}_{e_1} \cap \cdots \cap \mathcal{B}_{e_n}.$$

It is easy to check that $\mathfrak{B}_v$ is of full Lebesgue measure in $\mathbb{R}^n$. Let us show that

$$\forall x \in \mathfrak{B}_v, D_v f(x) = v.\nabla f(x).$$

Let $\varphi \in C_c^\infty(\mathbb{R}^n)$. For any $x \in \mathfrak{B}_v$, we have

$$\begin{aligned}\int_\Omega D_v f(x)\varphi(x)\mathrm{d}x &= -\int_\Omega f(x) D_v\varphi(x)\mathrm{d}x \\ &= -\sum_{i=1}^n v_i \int_\Omega f(x)\frac{\partial\varphi}{\partial x^i}(x)\mathrm{d}x \\ &= \sum_{i=1}^n v_i \int_\Omega \frac{\partial f}{\partial x^i}(x)\varphi(x)\mathrm{d}x \\ &= \int_{\mathbb{R}^n} (v.\nabla f(x))\xi(x)\mathrm{d}x.\end{aligned}$$

This and Lemma 22 imply that

$$D_v f(x) = v.\nabla f(x) \ a.e. \ x \in \Omega.$$

We choose now a countable dense subset in $\mathbb{R}^n$ such that $|v_k| = 1$, and let

$$A_k = \{x \in \Omega; \nabla f(x) \text{ exists and } D_{v_k} f(x) = v_k.\nabla f(x)\}.$$

For any $k \in \mathbb{N}$, each subset $\Omega \backslash A_k$ has Lebesgue measure zero. It means that

$$A = \bigcap_{k=1}^\infty A_k \text{ satisfies } \mathcal{L}^n(\Omega\backslash A) = 0.$$

Let us prove that f is differentiable a.e. $x \in A$. For any $x \in A$ and any $v \in \mathbb{R}^n$ with $|v| = 1$, we set

$$Q(x, v, t) := \frac{f(x + tv) - f(x)}{t} - v.\nabla f(x).$$

By a density argument, for $\varepsilon > 0$, there is $N \geq 0$ such that

$$\forall k \geq N, |v - v_k| \leq \varepsilon.$$

We recall that $Q(x, v_k, t) \underset{t\to 0}{\longrightarrow} 0$, that is $\exists\delta > 0$ such that for $0 < |t| < \delta$,

$$|Q(x, v_k, t)| \leq \frac{\varepsilon}{2}.$$

Assume that f is C-Lipschitz. Then, we have

$$|\frac{\partial f}{\partial x_i}| \leq C \text{ which means that } |\nabla f(x)| \leq \sqrt{n}C \; a.e.$$

Hence, we get for $x \in A$

$$|Q(x, v, t)| \leq |Q(x, vk, t)| + |Q(x, v, t) - Q(x, v_k, t)|$$

$$\leq \frac{\varepsilon}{2} + \left|\frac{f(x + vt) - f(x + v_k t)}{t} - (v - v_k).\nabla f(x)\right|$$

$$\leq \frac{\varepsilon}{2} + C|v - v_k| + |(v - v_k).\nabla f(x)|$$

$$\leq \frac{\varepsilon}{2} + C(1 + \sqrt{n})|v - v_k|$$

We can choose k sufficiently large such that

$$|v - v_k| < \frac{2}{2(1 + \sqrt{n})C}.$$

Then, $|Q(x, v, t)| < \varepsilon$ and the conclusion follows.

B.2 The Pansu-Rademacher Theorem

[MS01] gave an extension of the Rademacher Theorem.

Theorem 21. ***(The Pansu-Rademacher Theorem)***
Let $X_1, \ldots, X_m$ be m smooth vector fields satisfying the Hörmander condition and of the following form

$$X_j = \partial_j + \sum_{i=m+1}^{n} a_{ij}(x)\partial_i, j = 1, \ldots, m$$

where $a_{ij} \in C^{\infty}(\mathbb{R}^n)$. Let $(\mathbb{R}^n, d)$ be a Carnot-Caratheodory space induced by $X_1, \dots, X_m$. Assume that $f : \mathbb{R}^n \to \mathbb{R}$ is a Lipschitz function. Then, for a.e. $x \in \mathbb{R}^n$

$$\lim_{y \to x} \frac{f(y) - f(x) - \sum_{j=1}^{m} X_j f(x)(y_j - x_j)}{d(x,y)} = 0.$$

Proof of Theorem 21. For any $j = 1, \dots, m$, let $x \in \mathbb{R}^n$ such that $X_j(x) \neq 0$. We denote by $\mathcal{O}_j$ the orbit of x under X_j, that is

$$\mathcal{O}_j := \{\varphi_x^{X_j}(t) |\ t \in [0,1]\}$$

where $\varphi_x^{X_j}(.)$ is solution of the Cauchy problem

$$\dot{x}(t) = X_j(x(t)); \quad x(0) = x.$$

Let $f_{/\mathcal{O}_j} : \mathbb{R} \to \mathbb{R}$ be the restriction of f to $\mathcal{O}_j$. The function $f_{/\mathcal{O}_j}$ is Lipschitz, then it is differentiable at a.e. $x \in \mathbb{R}$. This means that $X_j f(x)$ exists a.e. $x \in \mathbb{R}^n$, $\forall j = 1, \dots, m$. Moreover, assume that f is L-Lipschitz, we get

$$|\nabla_h f(x)| \leq L, \ a.e. \ x \in \mathbb{R}^n.$$

Hence, $|\nabla_h f| \in L^p_{loc}(\mathbb{R}^n), \forall p \geq 1$ and thanks to the Lebesgue differentiation Theorem, we have:

$$\lim_{r \to o^+} \frac{1}{|B(x,r)|} \int_{B(x,r)} |\ |\nabla_h f|^p(x) - |\nabla_h f|^p(y)\ | \mathrm{d}y = 0$$

$$\Rightarrow \lim_{r \to o^+} \frac{1}{|B(x,r)|} \int_{B(x,r)} |\ \nabla_h f(x) - \nabla_h f(y)\ |^p \mathrm{d}y = 0. \tag{B.1}$$

Let $\Omega \subset \mathbb{R}^n$ be a bounded open set with homogeneous dimension Q and fix $p > Q$. Let $x \in \Omega$, we set

$$g(y) = f(y) - \sum_{j=1}^{m} X_j f(x)(y_j - x_j) \text{ such that } g(x) = f(x)$$

and

$$\forall j = 1, \dots, m, \quad X_j g(y) = X_j f(y) - X_j f(x).$$

By the Morrey inequality, $\exists C = C(\Omega, X, Q, p) > 0$ such that

$$|\ g(y) - g(x)\ | \leq Cr(\frac{1}{B(x,r)} \int_{B(x,r)} |\ \nabla_h g(z)\ |^p|)^{\frac{1}{p}}, \forall y \in B(x,r).$$

We choose $r = 2d(x,y)$ such that

$$\frac{| f(y) - f(x) - \sum_{j=1}^{m} X_j f(x)(y_j - x_j) |}{d(x,y)} \\ \leq 2C(\frac{1}{B(x,2d(x,y))}\int_{B(x,2d(x,y))} | \nabla_h f(z) - \nabla_h f(x) |^p|)^{\frac{1}{p}}.$$

By (B.1), the conclusion follows. ◻

B.3 The Alexandrov Theorem

The classical thorem of Alexandrov ([BF36], [Alex39], see also [How98]) states that a concave function in $\mathbb{R}^n$ admits a second-order derivative almost everywhere.

Theorem 22. ***(The Alexandrov Theorem)*** *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex function. Then, f is twice differentiable a.e. $x \in \mathbb{R}^n$.*

Proof of Theorem 22. We recall that the convex function f is locally Lipschitz. By the Rademacher Theorem (see Theorem 20), f is differentiable a.e. $x \in \mathbb{R}^n$. We denote the subdifferential of f by:

$$\partial f(x) = \{D_x f\}, \ a.e. \ x \in \mathbb{R}^n$$

where $D_x f$ is the classical differential of f at x.

We define the function

$$\begin{array}{cccc} F: & \mathbb{R}^n & \to & \mathbb{R}^n \\ & x & \mapsto & F(x) = x + \partial f(x). \end{array}$$

Lemma 23. *F is onto.*

Proof of Lemma 23. Let $y \in \mathbb{R}^n$ be fixed. We define

$$\varphi : \left\{ \begin{array}{ccc} \mathbb{R}^n & \longrightarrow & \mathbb{R} \\ x & \longmapsto & \varphi(x) = \frac{1}{2}||x||^2 + f(x) - x.y \end{array} \right.$$

We note that φ is convex and satisfies

$$\lim_{||x||\to+\infty} \varphi(x) = +\infty.$$

Then, φ admis a global minimum at some point x_0. It follows that

$$0 \in \partial\varphi(x_0) = x_0 + \partial f(x_0) - y = F(x_0) - y$$

$\Rightarrow y \in F(x_0)$. ☐

Lemma 24. *Let $y_0 \in F(x_0)$ and $y_1 \in F(x_1)$ such that*

$$y_0 = x_0 + b_0,\ y_1 = x_1 + b_1$$

with $b_0 \in \partial f(x_0), b_1 \in \partial f(x_1)$. Then, we have

$$||y_1 - y_0|| \geq ||x_1 - x_0|| \tag{B.2}$$

Proof of Lemma 24. We recall that

$$(y_1 - y_0).(x_1 - x_0) = ||x_1 - x_0||^2 + (b_1 - b_0).(x_1 - x_0).$$

Since f is convex, we have

$$(b_1 - b_0).(x_1 - x_0) \geq 0.$$

Then,

$$(y_1 - y_0).(x_1 - x_0) \geq ||x_1 - x_0||^2.$$

By the Cauchy-Schwartz inequality, we get

$$||y_1 - y_0||\ ||x_1 - x_0|| \geq (y_1 - y_0).(x_1 - x_0) \geq ||x_1 - x_0||^2$$

$\Rightarrow ||y_1 - y_0| \geq ||x_1 - x_0||$. ☐

The inequality (B.2) shows that F is injective. This and Lemma 23 imply that F est bijective. Thus, F is invertible. We define $G : \mathbb{R}^n \to \mathbb{R}^n$ such that

$$G(y) = x \Leftrightarrow y \in F(x).$$

By (B.2), we have for $y_0 \in F(x_0), y_1 \in F(x_1)$

$$||y_1 - y_0| \geq ||x_1 - x_0||.$$

$$\Rightarrow ||G(y_1) - G(y_0)| \leq ||y_1 - y_0||$$

G is 1-Lipschitz, then by the Rademacher Theorem, G is differentiable a.e.

Lemma 25. *Let $G : \mathbb{R}^n \to \mathbb{R}^n$ be a Lipschitz function. Then, the set*

$$E = \{G(y) \in \mathbb{R}^n | \ D_y G \textit{ exists and is invertible}\}$$

is of full Lebesgue measure.

Proof of Lemma 25. By the Approximation of Lipschitz functions (see Theorem 6.6.1 in [EG92]), for $\varepsilon < 0$, there is a function $h : \mathbb{R}^n \to \mathbb{R}$ of class C^1 such that

$$\mathcal{L}^n\{y \in \mathbb{R}^n | \ h(y) \neq G(y) \text{ or } D_y h \neq D_y G\} \leq \varepsilon.$$

Thanks to the Sard Lemma, the image by h of the set

$$H = \{y \in \mathbb{R}^n | \ G \text{ is not differentiable at } y \textit{ or } D_y G \text{ is not invertible}\}$$

has Lebesgue measure zero. We denote by $E = H^c$, the complementary set of H. And the conclusion follows. □

Fix $y_0 \in \mathbb{R}^n$ with $G(y_0) = x_0$ such that G is differentiable at y_0 and $D_{y_0} G$ is inversible. Hence,

$$G(y) - G(y_0) = D_{y_0} G.(y - y_0) + o(||y - y_0)$$

$$\Rightarrow x - x_0 = D_{y_0} G.(y - y_0) + o(||y - y_0||)$$

$$\Rightarrow y - y_0 = (D_{y_0} G)^{-1}.(x - x_0) + o(||y - y_0||).$$

As $o(||y - y_0||) = o(||x - x_0||)$, we get

$$\Rightarrow y - y_0 = (D_{y_0} G)^{-1}.(x - x_0) + o(||x - x_0||). \tag{B.3}$$

We have $y \in F(x)$ and $y_0 \in F(x_0)$, so (B.3) shows that F is differentiable a.e. We recall that by definition we have

$$F(x) = x + \partial f(x) \Rightarrow \partial f(x) = F(x) - x.$$

Thus, f is twice differentiable almost everywhere.

□

B.4 The sub-Riemannian version of the Alexandrov Theorem

In this section, our aim is to prove Theorem 14. In the spirit of Gutierrez and Montanari [GM04] and [GM05], we show that any locally h-concave function is $BV^2_{h,loc}$.

B.4.1 Tools from matrix analysis

Let $A = (a_{ij})_{ij}$ be an $m \times m$ symmetric matrix with eigenvalues $\lambda_1, \dots, \lambda_m$.

The second elementary symmetric function of A is defined by

$$\sigma_2(A) = s(\lambda) = \sum_{j<k} \lambda_j \lambda_k, \text{ with } \lambda = (\lambda_1, \dots, \lambda_m).$$

We can check that

$$s(\lambda) = \frac{1}{2}\Big\{(\sum_{j=1}^m \lambda_j)^2 - \sum_{j=1}^m \lambda_j^2\Big\} \tag{B.4}$$

and

$$\frac{\partial s}{\partial \lambda_j}(\lambda) = \sum_{k \neq j} \lambda_k.$$

Lemma 26. *If $\sigma_2(A) \geq 0$ and $trace(A) \geq 0$, then*

$$\frac{\partial s}{\partial \lambda_j}(\lambda) \geq 0, \ \forall j = 1, \dots, m.$$

Proof of Lemma 26. We recall that

$$trace(A) = \sum_{j=1}^m \lambda_j.$$

Hence,

$$trace(A) = \sum_{j \neq k} \lambda_k + \lambda_j = \frac{\partial s}{\partial \lambda_j}(\lambda) + \lambda_j. \tag{B.5}$$

This implies that

$$\lambda_j \geq 0 \quad \text{or} \quad \frac{\partial s}{\partial \lambda_j}(\lambda) + \lambda_j \geq 0.$$

Assume that $\lambda_j \geq 0$. We have, by hypothesis, $\sigma_2(A) = s(\lambda) \geq 0$.

Then, by (B.4)

$$\sum_{k=1}^{m} \lambda_k \geq (\sum_{k=1}^{m} \lambda_k^2)^{1/2} \geq \lambda_j, \tag{B.6}$$

By (B.5) and (B.6), we get

$$\frac{\partial s}{\partial \lambda_j}(\lambda) = \sum_{k \neq j} \lambda_k = \sum_{k=1}^{m} \lambda_k - \lambda_j \geq 0.$$

$\square$

Proposition 24. *If $\sigma_2(A) \geq 0$ and trace$(A) \geq 0$, then*

$$\sum_{i,j=1}^{m} \frac{\partial \sigma_2}{\partial a_{ij}}(A) x_i x_j \geq 0, \ \forall x \in \mathbb{R}^m.$$

Proof of Proposition 24. Let C be a non-negative definite Hermitian matrix. Let $\nu = (\nu_1, \ldots, \nu_{m)}$ be the eigenvalues of the matrix $A + C$ such that

$$\sigma_2(A + C) - \sigma_2(A) = s(\nu) - s(\lambda).$$

Since $C \geq 0$, we get $\nu_j \geq \lambda_j$, for any $j = 1, \ldots, m$.

Moreover, by lemma 26, we have

$$\delta = \frac{1}{2} \min \left\{ \frac{\partial s}{\partial \lambda_j}(\lambda_1, \ldots, \lambda_m), j = 1, \ldots, m \right\} \geq 0.$$

Choose C sufficiently small such that

$$\sigma_2(A + C) - \sigma(A) = \int_0^1 \frac{d}{d\tau} s(\lambda + \tau(\nu - \lambda)) d\tau$$

$$= \sum_{j=1}^{m} \int_0^1 \frac{\partial s}{\partial \lambda_j}(\lambda + \tau(\nu - \lambda)) d\tau (\nu_j - \lambda_j)$$

$$\geq \delta \sum_{j=1}^{m} (\nu_j - \lambda_j)$$

$$= \delta \, \big(trace(A+C) - trace(A)\big)$$

$$= \delta \; trace(C) \geq 0 \qquad \text{since } C \geq 0.$$

We take now $C = tx.x^T = t(x_i x_j)$, $x \in \mathbb{R}^m$ and $t > 0$ sufficiently small. We obtain

$$\sigma_2(A + tx.x^T) - \sigma_2(A) \geq \delta trace(C) = \delta t|x|^2.$$

Thus, for every $x \in \mathbb{R}^m$

$$\frac{d}{dt}\sigma_2(A + tx.x^T)|_{t=0} = \sum_{i,j=1}^{m} \frac{\partial \sigma_2}{\partial a_{ij}}(A) x_i x_j \geq \delta |x|^2 \geq 0.$$

□

B.4.2 Proof of Theorem 14

Let Ω be an open subset of the Carnot group $\mathcal{G}$.

Theorem 23. *Let $u : \Omega \to \mathbb{R}$ be an h-concave function. Then,*

$$\frac{X^i X^j u + X^j X^i u}{2} \quad \textit{is a Radon measure for} \quad i,j = 1, \ldots, m.$$

Proof of Theorem 23. We denote by $u_\varepsilon = \phi_\epsilon * u$ the convolution of u by a mollifier sequence $(\phi_\varepsilon)_{\varepsilon>0}$. We recall, by remark 5, $(u_\varepsilon)_\varepsilon$ is a sequence of smooth functions on M which are h-concave on M. Moreover, by Proposition 20, the $m \times m$ symmetric matrix

$$(\nabla_h^2 u(x))^* \text{ is negative semidefinite on } \Omega. \tag{B.7}$$

For some $\rho = (\rho_1, \ldots, \rho_m) \in \mathbb{R}^m$, we define L_ε by

$$L_\varepsilon(\psi) = -\sum_{i,j=1}^{m} \int_\Omega u_\varepsilon(x) \big(\frac{X^i X^j + X^j X^i}{2}\big) \psi(x) \rho_i(x) \rho_j(x) dx, \; \forall \psi \in C_c^\infty(\Omega).$$

Integrating by parts yields

$$L_\varepsilon(\psi) = -\sum_{i,j=1}^{m} \int_\Omega \psi(x) \big(\frac{X^i X^j + X^j X^i}{2}\big) u_\epsilon(x) \rho_i(x) \rho_j(x) dx.$$

We deduce by (B.7) that

$$L_\epsilon(\psi) \geq 0, \quad \forall \psi \geq 0.$$

We define now

$$L(\psi) := -\sum_{i,j=1}^{m} \int_\Omega u(x)(\frac{X_iX_j + X_jX_i}{2})\psi(x)\rho_i(x)\rho_j(x)dx, \ \psi \in C_c^\infty(\Omega).$$

Since $(u_\varepsilon)_\varepsilon$ converges uniformly to u, we get

$$L(\psi) = \lim_{\varepsilon \to 0} L_\varepsilon(\psi) \geq 0.$$

Thanks to the Riesz Representation Theorem (see section 1.8, [EG92]), there exists a Radon Theorem μ^ρ on Ω such that

$$L(\psi) = \int_\Omega \psi \mathrm{d}\mu^\rho, \forall \psi \in C_c^\infty(\Omega).$$

If we take $\rho = e_i$ then

$$\int_\Omega u(p)X_i^2\psi(x)dx = \int_\Omega \psi d\mu^{ii}.$$

Let us choose now $\rho = \dfrac{e_i + e_j}{\sqrt{2}}$, we obtain

$$-\int_\Omega u(x)(\frac{X_iX_j + X_jX_i + X_i^2 + X_j^2}{2})\psi(x)dx = \int_\Omega \psi d\mu^{ij}$$

$$\Rightarrow \quad -\int_\Omega u(x)(\frac{X_iX_j + X_jX_i}{2})\psi(x)dx = \int_\Omega \psi(d\mu^{ij} - d\mu^{ii} - d\mu^{jj}).$$

This implies

$$\frac{X_iX_ju + X_jX_iu}{2} = -\big(d\mu^{ij} - d\mu^{ii} - d\mu^{jj}\big) = d\nu^{ij}.$$

□

Let $u : \Omega \to \mathbb{R}$ a function of class C_h^2 on Ω. We denote by

$$\mathcal{H}(u) = \frac{X_iX_ju + X_jX_iu}{2}.$$

Definition 37. *A function $u \in C_h^2(\Omega)$ is said to be $\sigma_2(h)$-concave on Ω if*

1. *the trace of the symmetric matrix $\mathcal{H}(u)$ is non negative,*
2. *the second elementary symmetric function in the eigenvalues of $\mathcal{H}(u)$ given by*

$$\sigma_2\big(\mathcal{H}(u)\big) = \sum_{i<j}\Big\{X^iX^iuX^jX^ju - (\frac{X^iX^ju + X^jX^iu}{2})^2\Big\}$$

is non negative.

We pick a local frame $\{Y^1, \ldots, Y^n\}$ of the tangent space of M such that

$$Y^i = X^i, \forall i = 1, \ldots, m.$$

We denote by

$$[X^i, X^j] = \sum_{k=1}^{n} \alpha_{ij,k} Y^k, \forall i, j = 1, \ldots, m$$

where $\alpha_{ij,k}$ are constants.

Theorem 24. *Let u, v be two functions in $C^2(\Omega)$ such that $u + v$ is $\sigma_2(h)$-convex in Ω satisfying $u = v$ on $\partial\Omega$ et $v < u$ sur Ω. Then,*

$$\int_\Omega \Big\{\sigma_2(\mathcal{H}(u)) + \frac{3}{4}\sum_{i,j=1}^{m}\sum_{k=1}^{n}(\alpha_{ij,k})^2(Y^ku)^2\Big\}dz$$
$$\leq \int_\Omega \Big\{\sigma_2(\mathcal{H}(v)) + \frac{3}{4}\sum_{i,j=1}^{m}\sum_{k=1}^{n}(\alpha_{ij,k})^2(Y^kv)^2\Big\}dz.$$

Proof of Theorem 24. We set

$$S(u) = \sigma_2\big(\mathcal{H}(u)\big) = \sum_{i<j}\Big\{X^iX^iuX^jX^ju - (\frac{X^iX^ju + X^jX^iu}{2})^2\Big\}.$$

Setting $r_{ij} = \dfrac{X^iX^ju + X^jX^iu}{2}$, we obtain:

$$\frac{\partial S}{\partial r_{ii}}(u) = \sum_{j\neq i} X^jX^ju, \qquad \frac{\partial S}{\partial r_{ij}}(u) = -\frac{X^iX^ju + X^jX^iu}{2}.$$

We recall that, by section B.4.1, since u is $\sigma_2(h)-$convex,

$$\frac{\partial S}{\partial r_{ij}}(u) \text{ is non-negative definite.}$$

(we apply the result of Proposition 24 with $A = \mathcal{H}(u)$, where u is a $\sigma_2(h)$-convex function.)

Let $0 \le s \le 1$ and $\varphi(s) = S(v + sw)$ where $w = u - v$. We get

$$\int_\Omega \{S(u) - S(v)\}\ \mathrm{d}z = \int_0^1 \int_\Omega \dot{\varphi}(s)\ \mathrm{d}z\ \mathrm{d}s$$

$$= \int_0^1 \int_\Omega \Big\{ \sum_{i,j=1}^m \frac{\partial S}{\partial r_{ij}}(v + sw)(z)(X^i X^j) w(z) \Big\}\ \mathrm{d}z\ \mathrm{d}s$$

$$= \int_0^1 \int_\Omega \Big\{ \sum_{i,j=1}^m X^i (\frac{\partial S}{\partial r_{ij}}(v + sw)(z) X^j w(z)) - X^i (\frac{\partial S}{\partial r_{ij}}(v + sw)(z)) X j w(z) \Big\}\ \mathrm{d}z\ \mathrm{d}s.$$

As $w = 0$ on $\partial\Omega$ and $w > 0$ on Ω, then the normal to $\partial\Omega$ is given by $\nu_X = -\dfrac{Xw}{|Dw|}$, with $Xw = (X^1 w, \ldots, X^m w)$. By an integration par parts, we get

$$A = \int_0^1 \int_\Omega \sum_{i,j=1}^m X^i \big(\frac{\partial S}{\partial r_{ij}}(v + sw)(z) X^j w(z) \big) dz ds$$

$$= \int_0^1 \int_{\partial\Omega} \sum_{i,j=1}^m \frac{\partial S}{\partial r_{ij}}(v + sw)(z) X^j w(z) \nu_{X^i} d\sigma(z) ds$$

$$= -\int_0^1 \int_{\partial\Omega} \sum_{i,j=1}^m \frac{\partial S}{\partial r_{ij}}(v + sw)(z) X^j w(z) \frac{X^i w}{|Dw|} d\sigma(z) ds$$

$$= -\int_0^1 \int_{\partial\Omega} \sum_{i,j=1}^m \{ \frac{\partial S}{\partial r_{ij}}(v) + s \frac{\partial S}{\partial r_{ij}}(w)(z) \} X^j w(z) \frac{X^i w}{|Dw|} d\sigma(z) ds$$

$$= -\int_0^1 \int_{\partial\Omega} \sum_{i,j=1}^m \{ \frac{\partial S}{\partial r_{ij}}(v) + s \frac{\partial S}{\partial r_{ij}}(u)(z) - s \frac{\partial S}{\partial r_{ij}}(v)(z) \} X^j w(z) \frac{X^i w}{|Dw|} d\sigma(z) ds$$

$$= -\frac{1}{2}\int_{\partial\Omega}\sum_{i,j=1}^{m}\{\frac{\partial S}{\partial r_{ij}}(v) + \frac{\partial S}{\partial r_{ij}}(u)(z)\}X^j w(z)\frac{X^i w}{|Dw|}d\sigma(z)$$

$$= -\frac{1}{2}\int_{\partial\Omega}\sum_{i,j=1}^{m}\frac{\partial S}{\partial r_{ij}}(u+v)X^j w(z)\frac{X^i w}{|Dw|}d\sigma(z) \leq 0.$$

On the other hand, we remark that for $j = 1, \ldots, m$ fixed, we have:

$$\sum_{i=1}^{m} X^i(\frac{\partial S}{\partial r_{ij}}w) = X^j(\frac{\partial S}{\partial r_{jj}}w) + \sum_{i\neq j} X^i(\frac{\partial S}{\partial r_{ij}}w)$$

$$= X^j\big(\sum_{k\neq j} X^k X^k w\big) - \sum_{i\neq j} X^i\big(\frac{X^i X^j w + X^j X^i w}{2}\big)$$

$$= \sum_{i\neq j}\Big(X^j X^i X^i w - X^i(\frac{X^i X^j w + X^j X^i w}{2})\Big)$$

$$= \sum_{i\neq j}\Big(\frac{[X^j, X^i]X^i w}{2} + \frac{[X^j, X^i]X^i w}{2} + \frac{X^i[X^j, X^i]w}{2}\Big)$$

$$= 3\sum_{i\neq j}(\frac{X^i[X^j, X^i]w}{2})$$

$$= \frac{3}{2}\sum_{i\neq j} X^i[X^j, X^i]w.$$

Hence,

$$B = \int_0^1\int_\Omega\sum_{i,j=1}^{m} X^i(\frac{\partial S}{\partial r_{ij}}(v+sw)(z))X^j w(z)dzds$$

$$= \frac{3}{2}\int_0^1\int_\Omega\sum_{i\neq j}^{m} X^i[X^j, X^i](v+sw)(z)X^j w(z)dzds$$

$$= \frac{3}{2}\int_0^1 \int_{\partial\Omega} \sum_{i\neq j}^{m} [X^j, X^i](v+sw)(z) X^j w(z) \nu_{X^i} d\sigma(z) ds$$

$$-\frac{3}{2}\int_0^1 \int_{\Omega} \sum_{i\neq j}^{m} [X^j, X^i](v+sw)(z) X^i X^j w(z) dz\ ds$$

$$= \frac{3}{2}\int_0^1 \int_{\partial\Omega} \sum_{j\neq 1}^{m} [X^j, X^1](v+sw)(z) X^j w(z) \nu_{X^1} d\sigma(z) ds$$

$$+ \cdots + \frac{3}{2}\int_0^1 \int_{\partial\Omega} \sum_{j\neq m}^{m} [X^j, X^m](v+sw)(z) X^j w(z) \nu_{X^m} d\sigma(z) ds$$

$$-\frac{3}{2}\int_0^1 \int_{\Omega} \sum_{i\neq j} [X^j, X^i](v+sw)(z) X^i X^j w(z) dz ds$$

$$= -\frac{3}{2}\int_0^1 \int_{\Omega} \sum_{i\neq j} [X^j, X^i](v+sw)(z) X^i X^j w(z) dz ds$$

$$= -\frac{3}{2}\int_0^1 \int_{\Omega} \sum_{j\neq 1}^{m} [X^j, X^1](v+sw)(z) X^1 X^j w(z) dz ds$$

$$-\frac{3}{2}\int_0^1 \int_{\Omega} \sum_{j\neq 2}^{m} [X^j, X^2](v+sw)(z) X^2 X^j w(z) dz ds$$

$$- \ldots -\frac{3}{2}\int_0^1 \int_{\Omega} \sum_{j\neq m}^{m} [X^j, X^m](v+sw)(z) X^m X^j w(z) dz ds$$

$$\begin{aligned} = -\frac{3}{2}\int_0^1 \int_{\Omega} \Big\{ & [X^2, X^1](v+sw)(z) X^1 X^2 w(z) + \cdots + \\ & [X^m, X^1](v+sw)(z) X^1 X^m w(z) + [X^1, X^2](v+sw)(z) X^2 X^1 w(z) + \cdots + \\ & [X^1, X^m](v+sw)(z) X^m X^1 w(z) + \cdots + [X^1, X^m](v+sw)(z) X^m X^1 w(z) \\ & + \cdots + [X^{m-1}, X^m](v+sw)(z) X^m X^{m-1} w(z) \Big\} dz ds \end{aligned}$$

$$= -\frac{3}{2}\int_0^1 \int_{\Omega} \sum_{j\neq i} [X^j, X^i](v+sw)(z) [X^i, X^j] w(z) dz ds$$

$$= \frac{3}{2}\int_0^1 \int_\Omega \sum_{j\neq i}^m [X^i, X^j](v+sw)(z)[X^i, X^j]w(z)dzds$$

$$= \frac{3}{2}\int_0^1 \int_\Omega \sum_{j\neq i}^m \sum_{k=1}^n \alpha_{ij,k}Y^k(v+sw)(z)\alpha_{ij,k}Y^k w(z)dzds$$

$$= \frac{3}{2}\int_0^1 \int_\Omega \sum_{j\neq i}^m \sum_{k=1}^n (\alpha_{ij,k})^2\Big(Y^k v(z) + sY^k w(z)\Big)Y^k w(z)dzds$$

$$= \frac{3}{2}\int_\Omega \sum_{j\neq i}^m \sum_{k=1}^n (\alpha_{ij,k})^2\Big(\int_0^1 (Y^k v(z) + sY^k w(z))\mathrm{d}s\Big)Y^k w(z)dz$$

$$= \frac{3}{2}\int_\Omega \sum_{j\neq i}^m \sum_{k=1}^n (\alpha_{ij,k})^2\Big(sY^k v(z) + \frac{s^2}{2}Y^k w(z)\Big)|_0^1 \; Y^k w(z)dz$$

$$= \frac{3}{2}\int_\Omega \sum_{j\neq i}^m \sum_{k=1}^n (\alpha_{ij,k})^2\Big(Y^k v(z) + \frac{1}{2}(Y^k u(z) - Y^k v(z))\Big)\Big(Y^k u(z) - Y^k v(z)\Big)dz$$

$$= \frac{3}{2}\int_\Omega \sum_{j\neq i}^m \sum_{k=1}^n (\alpha_{ij,k})^2\frac{1}{2}\Big(Y^k u(z) + Y^k v(z)\Big)\Big(Y^k u(z) - Y^k v(z)\Big)dz$$

$$= \frac{3}{4}\int_\Omega \sum_{j\neq i}^m \sum_{k=1}^n (\alpha_{ij,k})^2\Big((Y^k u)^2(z) - (Y^k v)^2(z)\Big)dz.$$

And the conclusion follows.

□

Lemma 27. *Let u and $v \in C^2(\Omega)$ be two $\sigma_2(h)$-convex functions. Let $f : \mathbb{R}^2 \to \mathbb{R}$ be a convex function such that f is non-decreasing with respect to each variable. Then,*

$$w = f(u,v) \textit{ is a } \sigma_2(h) - \textit{convex function.}$$

Proof of Lemma 27. Assume that f is of class C^2 on $\mathbb{R}^2$. We have

$$X_j w = \frac{\partial f}{\partial u}X_j u + \frac{\partial f}{\partial v}X_j v$$

and

$$X_iX_jw = \frac{\partial f}{\partial u}X_iX_ju + \frac{\partial^2 f}{\partial v\partial u}X_ivX_ju + \frac{\partial^2 f}{\partial u^2}X_iuX_ju$$
$$+ \frac{\partial f}{\partial v}X_iX_jv + \frac{\partial^2 f}{\partial v^2}X_ivX_jv + \frac{\partial^2 f}{\partial u\partial v}X_iuX_jv.$$

For any $h = (h_1, h_2) \in \mathbb{R}^2$, we have

$$< \mathcal{H}(w)h, h >= \sum_{i,j=1}^{m} X_iX_jwh_ih_j$$

$$= \frac{\partial f}{\partial u} < \mathcal{H}(u)h, h > + \frac{\partial f}{\partial v} <H(v)h, h>$$
$$+ \frac{\partial^2 f}{\partial u^2}(\sum_{i=1}^{m} X_iuh_i)(\sum_{j=1}^{m} X_juh_j) + \frac{\partial^2 f}{\partial v^2}(\sum_{i=1}^{m} X_ivh_i)(\sum_{j=1}^{m} X_jvh_j)$$
$$+ \frac{\partial^2 f}{\partial u\partial v}(\sum_{i=1}^{m} X_iuh_i)(\sum_{j=1}^{m} X_jvh_j) + \frac{\partial^2 f}{\partial v\partial u}(\sum_{i=1}^{m} X_ivh_i)(\sum_{j=1}^{m} X_juh_j)$$

Since u and v are $\sigma_2(h)$-convex,

$$\frac{\partial f}{\partial u} \geq 0 \text{ and } \frac{\partial f}{\partial v} \geq 0.$$

Moreover, the matrix

$$\begin{pmatrix} \dfrac{\partial^2 f}{\partial u^2} & \dfrac{\partial^2 f}{\partial u\partial v} \\ \dfrac{\partial^2 f}{\partial v\partial u} & \dfrac{\partial^2 f}{\partial v^2} \end{pmatrix}$$

est non negative definite. Then, w is $\sigma_2(h)$-convex.

Assume now that f is a continuous function. We consider $f_\epsilon = \phi_\epsilon * f$ the convolution of f by the mollifier sequence ϕ_ϵ. Since f is convex, f_ϵ is also convex. From the above , $w_\epsilon = f_\epsilon(u, v)$ is $\sigma_2(h)$-convex such that

$$w_\epsilon \underset{\epsilon \to 0}{\longrightarrow} w.$$

Hence, we conclude that w is $\sigma_2(h)$-convex. □

Proposition 25. *Let $u \in C^2(\Omega)$ be a $\sigma_2(h)$-convex function. Then, for every compact $K \subset \Omega$, there exists a constant $C = C(K, \Omega) > 0$, independant of u such that*

$$\int_\Omega \Big\{\sigma_2(\mathcal{H}(u)) + \frac{3}{4}\sum_{i,j=1}^{m}\sum_{k=1}^{n}(\alpha_{ij,k})^2(Y_k u)^2\Big\}dz \le C(osc_\Omega u)^2.$$

with $osc_\Omega u$ the oscillation of u on Ω.

Proof of Proposition 25. We proof the result on the Heisenberg group in $\mathbb{R}^{2n+1}$ with coordinates $(x_1, \ldots, x_n, y_1, \ldots, y_n, t)$ where

$$\Delta = Span\{X^1, \ldots, X^n, Y^1, \ldots, Y^n\}$$

with

$$X^i = \partial_{x_i} - 2y_i\partial_t, \quad Y^i = \partial_{y_i} + 2x_i\partial_t$$

and

$$[X^i, Y^j] = 4\partial_t, \forall i, j = 1, \ldots, n.$$

Let $\bar{x} \in \Omega$ and $B_R = B_R(\bar{x})$ be the ball centered at $\bar{x}$ of radius R such that $B_R \subset \Omega$. For $0 < \sigma < 1$, let $B_{\sigma R}$ be the concentric ball centered at $\bar{x}$ of radius σR. Note that the sub-Riemannian structure is invariant by translation, so we may assume that $\bar{x} = 0$.

Set $M = \max\limits_{x \in B_R} u(x)$, we get

$$u - M \le 0 \text{ on } B_R.$$

Choose $\varepsilon > 0$ such that $u - M - \varepsilon < -\varepsilon$. By subtracting a constant, we may assume that

$$u < -\varepsilon \text{ on } B_R.$$

Put

$$m_0 = \inf_{x \in B_R} u(x) \text{ and } v(x) = \frac{m_0}{(1-\sigma^4)R^4}(R^4 - ||x||^4).$$

We check easily that

$$v = 0 \ \text{ on } \partial B_R \text{ and } \ v = m_0 \text{ on } \partial B_{\sigma R}.$$

Following the calculations in the proof of Proposition 6.2 in [GM05], we get

$$\sigma_2(\mathcal{H}(v)) = c_n(|x|^2 + |y|^2)^2\left(\frac{m_0}{(1-\sigma^4)R^4}\right)^2 \ge 0,$$

with c_n a positive constant and, since $m_0 \leq 0$

$$trace(\mathcal{H}(v)) = -(8n+4)(|x|^2+|y|^2)\frac{m_0}{(1-\sigma^4)R^4} \geq 0.$$

This implies that v is $\sigma_2(h)$-convex.

Since $v - m_0 = 0$ sur $\partial B_{\sigma R}$, we have

$$v \leq m_0 \text{ on } B_{\sigma R}.$$

In particular,

$$v \leq u \text{ on } B_{\sigma R}.$$

Let $\rho \in C_0^\infty(\mathbb{R}^2)$ such that $\int_{R^2} \rho(x)\mathrm{d}x = 1$ and

$$f_h(x_1,x_2) = h^{-2}\int_{\mathbb{R}^2} \rho(\frac{x-y}{h}) \max\{y_1,y_2\}\mathrm{d}y_1\mathrm{d}y_2.$$

Put

$$w_h = f_h(u,v).$$

By Lemma 27, w_h is $\sigma_2(h)$-convex.

If $y \in B_R$, we have

$$v(y) \leq u(y).$$

Then, for h sufficiently smell, we have

- if $v(y) < u(y)$ then, $f_h(u,v)(y) = u(y)$.
- if $v(y) = u(y)$ then, $f_h(u,v)(y) = u(y) + \alpha h$.

Hence,

$$\begin{aligned}\int_{B_{\sigma R}} \{\sigma_2(\mathcal{H}(u)) + 12n(\partial_t u)^2\}\mathrm{d}z &= \int_{B_{\sigma R}} \{\sigma_2(\mathcal{H}(w_h)) + 12n(\partial_t w_h)^2\}\mathrm{d}z \\ &\leq \int_{B_R} \{\sigma_2(\mathcal{H}(w_h)) + 12n(\partial_t w_h)^2\}\mathrm{d}z \quad \text{(B.8)}\end{aligned}$$

Since $f_h(u,v) \geq v$ on B_R and $f_h(u,v) = 0$ on ∂B_R, we apply Theorem 24:

$$\begin{aligned}\int_{B_R} \{\sigma_2(\mathcal{H}(w_h)) + 12n(\partial_t w_h)^2\}\mathrm{d}z &\leq \int_{B_R} \{\sigma_2(\mathcal{H}(v)) + 12n(\partial_t v)^2\}\mathrm{d}z \\ &= (\frac{m_0}{(1-\sigma)})^2 R^{2n-2}\int_{B_1} (c_n(|x|^2+|y|^2)^2 + 48nt^2)\mathrm{d}z.\end{aligned} \quad \text{(B.9)}$$

By (B.8) and (B.9), we obtain

$$\int_{B_{\sigma R}} \{\sigma_2(\mathcal{H}(u)) + 12n(\partial_t u)^2\}\mathrm{d}z \le (\frac{m_0}{(1-\sigma)})^2 R^{2n-2} \int_{B_1} (c_n(|x|^2+|y|^2)^2 + 48nt^2)\mathrm{d}z.$$

$$\le C(m_0)^2 R^{2n-2}$$

$$\le CR^{2n-2}(osc_{B_R}u + \varepsilon)^2.$$

Make $\varepsilon \to 0$. And, the conclusion follows.

□

Corollary 1. *Let $u \in C^2(\Omega)$ be a $\sigma_2(h)$-convex function. Then, there is a positive constant C independent of u such that*

$$\int_\Omega (Y^k u)^2(z)\mathrm{d}z \le C(osc_\Omega u)^2, \forall k = 1,\dots,m_2.$$

In other words, $\forall k = 1,\dots,m_2$, $Y^k u$ is $L^2(\Omega)$.

Theorem 25. *Let $u : \Omega \to \mathbb{R}$ be an h-convex function. Then, $u \in BV_h^2(\Omega)$.*

Proof of Theorem 25. Let $u : \Omega \to \mathbb{R}$ be an h-convex function. Then, u is Lipschitz with respect to the sub-Riemannian distance and $X^i u$ exists a.e. on Ω, $i = 1,\dots,m$. Moreover, there is a Radon measure $d\nu^{ij}$ such that

$$\frac{X^i X^j u + X^j X^i u}{2} = d\nu^{ij}, i,j = 1.$$

We recall that for $i,j = 1,\dots,m$,

$$X^i X^j = \frac{X^i X^j + X^j X^i + [X^i, X^j]}{2} = \frac{X^i X^j + X^j X^i}{2} + \sum_{k=1}^n \alpha_{ij,k} Y^k.$$

Let $\phi = \sum_{j=1}^m \phi_j X^j$ be a function of class C^2 with a compact support K and $||\phi|| < 1$. We get

$$\int_\Omega X^i u(z)\ div_X \phi(z) dz$$

$$= -\int_\Omega u(z)\ X^i div_X \phi(z) dz$$

$$= -\sum_{j=1}^{m} \int_\Omega u(z)\ X^i X^j \phi_j(z) dz$$

$$= -\sum_{j=1}^{m} \int_\Omega u(z) \Big(\frac{X^i X^j \phi_j(z) + X^j X^i \phi_j(z)}{2} + \sum_{k=1}^{n} \alpha_{ij,k} Y^k \phi_j(z) \Big) dz$$

$$= \sum_{j=1}^{m} \Big(\int_\Omega \phi_j(z) d\nu^{ij}(z) + \sum_{k=1}^{n} \alpha_{ij,k} \int_\Omega u(z) Y^k(z) \phi_j(z) dz \Big)$$

$$\leq \sum_{j=1}^{m} \nu^{ij}(K) + \sum_{k=1}^{n} \alpha_{ij,k} \int_\Omega u(z) Y^k \phi_j(z) dz.$$

Let $u_\epsilon = \varphi_\epsilon * u$ be a sequence of smooth functions which are h-convex. Then,

$$u_\epsilon \text{ is } \sigma_2(h) - \text{convex and } Y^k u_\epsilon \in L^2(\Omega), \forall k = 1, \ldots, n-m.$$

$$|\int_\Omega u_\epsilon Y^k \phi_j dz| = |\int_\Omega Y^k u_\epsilon \phi_j dz| \leq ||Y^k u_\epsilon||_{L^2(K)} \leq C.$$

By making ϵ tends to 0, we get

$$|\int_\Omega u(z) Y^k \phi_j(z) dz| \leq C.$$

Hence,

$$\int_\Omega X^i u(z)\ div_X \phi(z) dz < +\infty$$

which implies that $u \in BV_h^2(\Omega)$. $\square$

Appendix C

Cauchy Problem for BV vector fields

Here, we study the Cauchy problem for BV functions.

Let $b : [0, 1] \times \mathbb{R}^n \to \mathbb{R}^n$ be a bounded vector field such that

1. $b_t(x) := b(t, x) \in BV_{loc}(\mathbb{R}^n), a.e.\ t \in [0, 1]$
2. $D.b_t << \mathcal{L}^n$ such that $D.b_t = divb_t\mathcal{L}^n$
3. $|D.b_t|(B_R) \in L^1_{loc}([0, 1]),\ \forall R > 0$
4. $\displaystyle\int_0^1 ||[divb_t]^-||_{L^\infty(B_R)} < +\infty$

In [Amb04] (see also [Amb03]), Ambrosio defined a class of Lagrangian flows, solutions to the Cauchy problem $\dot{\gamma}(t) = b(t, \gamma(t))$ for $\mathcal{L}^n$-almost every initial conndition x and proved for them existence and uniqueness.

We introduce now our main notations and definitions. By abuse of notation, we set $\mathscr{C}$ the space of continuous $\mathbb{R}^n$-valued maps in $[0, 1]$, i.e. $\mathscr{C} = C\big([0, 1], \mathbb{R}^n\big)$.

Definition 38. *Let $x \in \mathbb{R}^n$. We denote by $\mathscr{C}^b_x$ the subspace of $\mathscr{C}$ given by*

$$\mathscr{C}^b_x := \Big\{\gamma(x) \in \mathscr{C}; \gamma(x)(t) = x + \int_0^t b\big(\tau, \gamma(x)(\tau)\big)\mathrm{d}\tau, \forall t \in [0, 1]\Big\}.$$

It is clear that the subspace $\mathscr{C}^b_x$ is made up by solutions of the ODE $\dot{\gamma}(t) = b(t, \gamma(t))$ starting at x.

Definition 39. *Let $A \subseteq \mathbb{R}^n$ and $\gamma :\ A \to \mathscr{C}$ be a $\mathcal{L}^n$-measurable map. We say that γ is a Lagrangian flow from A, relative to b if for a.e. $x \in A$,*

$$\gamma(x) \in \mathscr{C}^b_x.$$

The main result introduced by Ambrosio [Amb04] is the following:

Theorem 26. *Let $b : [0,1] \times \mathbb{R}^n \to \mathbb{R}^n$ be a bounded vector field such that*

1. $b_t(x) := b(t,x) \in BV_{loc}(\mathbb{R}^n), a.e. t \in [0,1]$
2. $D.b_t << \mathcal{L}^n$ *such that* $D.b_t = divb_t\mathcal{L}^n$
3. $|D.b_t|(B_R) \in L^1_{loc}([0,1]),\ \forall R > 0$
4. $\displaystyle\int_0^1 ||[divb_t]^-||_{L^\infty(B_R)} < +\infty$

Then, for any $\mathcal{L}^n$-measurable set $A \subseteq \mathbb{R}^n$, there is a unique Lagrangian flow starting from A relative to b.

Its proof is based on tools borrowed from Probability and calculus of variations together with a renormalization property.

C.1 Probability measures on $\mathscr{C}$

Let $\mathcal{P}(\mathscr{C})$ be the space of probability measures on $\mathscr{C}$.

Let A be a subset of $\mathbb{R}^n$ and let $\eta : A \to \mathcal{P}(\mathscr{C})$ be a $\mathcal{L}^n$-measurable application. For $t \in [0,1]$, we define η_{tx} the image of η_x by the application $\gamma \mapsto \gamma(t)$ such that for any subset K of $\mathbb{R}^n$,

$$\eta_{tx}(K) = \eta_x\big(\gamma : \gamma(t) \in K\big).$$

Lemma 28. *Let $x \in A$. Assume that for each $t \in [0,1]$, η_{tx} is a Dirac measure. Then, η_x is a Dirac measure.*

Proof of Lemma 28. Let γ, γ' be two distinct functions in $\mathscr{C}$. Then, there is $t_0 \in [0,1]$ such that $\gamma(t_0) \neq \gamma'(t_0)$. By a continuity argument, we can choose an open interval I containing t_0 such that

$$\gamma(t) \neq \gamma'(t), \forall t \in I.$$

We can construct two disjoint neighborhoods K and K' of $\gamma(t_0)$ and $\gamma'(t_0)$ respectively. Let $t \in I$ be fixed such that, by definition, we have

$$\eta_{tx}(K) = \eta\big(\gamma :\ \gamma(t) \in K\big)$$

and

$$\eta_{tx}(K') = \eta\big(\gamma' :\ \gamma'(t) \in K'\big).$$

Without loss of generality, since η_{tx} is a Dirac measure, we may assume that η_{tx} is concentrated at a point k_0 in K. This means that η_x is concentrated at an application $\gamma \in \mathscr{C}$ such that $\gamma(t) = k_0 \in K$. Hence, η_x is a Dirac measure. $\square$

Theorem 27. *Assume that*

$$\forall t \in [0,1], \bigcup_{x\in A} \textit{supp}\ \eta_{tx} \subset \mathbb{R}^n \tag{C.1}$$

and

$\forall t \in [0,1]$*, there is a negligeable subset* A_t *of* A *such that*

$$\eta_x\big(\gamma : \gamma(t) \in K\big)\ \eta_x\big(\gamma : \gamma(t) \in K'\big) = 0, \forall x \in A\backslash A_t \tag{C.2}$$

where K*,*K' *are two disjoint subsets of* $\mathbb{R}^n$*.*

Then, there is a negligible subset $\bar{A}$ *of* A *such that*

$$\forall x \in A\backslash\bar{A}, \eta_x \textit{ is a Dirac measure.}$$

***Proof of Theorem** 27.* By Lemma 28, it is sufficient to prove that there is a negligeable suset $\bar{A}$ of A such that $\forall x \in A\backslash\bar{A}$,

$$\forall t \in [0,1],\ \ \eta_{tx} \text{ is a Dirac measure.}$$

Without loss of generality, we may assume that

$$\bigcup_{x\in A} \textit{supp}\ \eta_{tx}\ \subset Q, \text{where } Q \text{ is a cube.}$$

We divide Q into two disjoint parts Q_1, Q_2 such that $Q = Q_1 \cup Q_2$. Let $t \in [0,1]$ be fixed. By (C.2), there is a negligeable set $A_t \subset A$ such that

$$\eta_{tx}(Q_1)\eta_{tx}(Q_2) = 0, \forall x \in A\backslash A_t.$$

This involves two possibilities :

- $\textit{supp}\ \eta_{tx} \not\subset Q_1 \Rightarrow x \notin A_1\backslash A_t := \big\{y \in A\backslash A_t : \eta_{ty}(Q_1) \neq 0\big\}$
- $\textit{supp}\ \eta_{tx} \not\subset Q_2 \Rightarrow x \notin A_2\backslash A_t := \big\{y \in A\backslash A_t : \eta_{ty}(Q_2) \neq 0\big\}$.

For $h \in \mathbb{N}$ fixed, we consider the canonical decomposition of Q into 2^{nh} cubes Q_i^h of side 2^{-h}. From $\{Q_i^h\}_i$, we can construct a family of sets $\{\bar{Q}_i^h\}_i$ pairwise disjoint such that

$$\bar{Q}_1^h = Q_1^h,\ \ \bar{Q}_2^h = Q_2^h\backslash Q_1^h,\ \ldots\ \ \bar{Q}_i^h = Q_i^h\backslash(Q_1^h \cup \cdots \cup Q_{i-1}^h)$$

and

$$Q = \bigcup_i \bar{Q}_i^h.$$

By assumption (C.2), $\forall t \in [0,1]$, there is a negligeable set A_t of A such that

$$\eta_{tx}(\bar{Q}_i^h)\ \eta_{tx}(\bar{Q}_j^h) = 0, \forall x \in A \backslash A_t, i \neq j.$$

For every i, we construct the set $A_i^h \backslash A_t$ such that

$$\forall x \in A_i^h \backslash A_t,\ supp\ \eta_{tx} \subset \bar{Q}_i^h.$$

Hence, $diam\ supp\ \eta_{tx} \leq \sqrt{n} 2^{-h}$. Make h tends to ∞,

$$diam\ supp\ \eta_{tx} = 0$$

which implies that η_{tx} is a Dirac measure. It follows from Lemma 28 that there is $\bar{A} = \bigcup_{t \in [0,1]} A_t$ a negligeable set such that

$$\eta_x \text{ is a Dirac measure.}$$

□

C.2 Renormalization property

In this section, we proceed by extending the Diperna-Lions theory [DL89] to obtain a renormalization property in the case of BV_{loc} functions.

Let us recall the result introduced by Diperna-Lions in the case of Sobolev functions.

Theorem 28. *Let $B \in W^{1,1}_{loc}(\mathbb{R}^n, \mathbb{R}^n)$ and $w \in L^{\infty}_{loc}(\mathbb{R}^n)$ satisfying the transport equation*

$$D.\nabla w = c\mathcal{L}^n, \textit{for some } c \in \mathcal{L}^1_{loc}(\mathbb{R}^n). \tag{C.3}$$

Then, for any $h \in C^1(\mathbb{R}^n)$, we have

$$B.\nabla\big(h(w)\big) = c\dot{h}(w)\mathcal{L}^n.$$

Proof of Theorem 28. Let $(\rho_\varepsilon)_\varepsilon$ be a mollifier sequence in $\mathbb{R}^n$. From (C.3), we get

$$\big(D.\nabla w\big) * \rho_\varepsilon = c * \varepsilon\mathcal{L}^n.$$

$$\Rightarrow B.\nabla(w * \rho_\varepsilon = c * \rho_\varepsilon \mathcal{L}^n + r_\varepsilon \tag{C.4}$$

where

$$r_\varepsilon := B.\nabla(w * \rho_\varepsilon) - (B.\nabla w) * \rho_\varepsilon.$$

We multiply both sides of (C.4) by $\dot{h}(w * \ \rho_\varepsilon)$, we obtain

$$B.\nabla\big(h(w) * \rho_\varepsilon\big) = \dot{h}(w * \ \rho_\varepsilon)\big[c * \rho_\varepsilon \mathcal{L}^n + r_\varepsilon\big].$$

We note that

$$\begin{aligned}
r_\varepsilon &= \int_{\mathbb{R}^n} w(x - \varepsilon y)\frac{B(x-\varepsilon y) - B(x)}{\varepsilon}.\nabla\rho(y)\mathrm{d}y - (x.divB) * \rho_\varepsilon(x) \\
&\underset{\varepsilon\to 0}{\approx} -w(x)\int_{\mathbb{R}^n}\sum_{i,j=1}^{n}\frac{\partial B^i}{\partial x_j}(x)y_j\frac{\partial \rho}{\partial y_i}(y)\mathrm{d}y - w(x)divB(x) \\
&\underset{\varepsilon\to 0}{\approx} w(x)\int_{\mathbb{R}^n}\sum_{i,j=1}^{n}\frac{\partial B^i}{\partial x_j}(x)\rho(y)\frac{\partial y_j}{\partial y_i}\mathrm{d}y - w(x)divB(x) \\
&= 0 \qquad \text{because} \int_{\mathbb{R}^n}\rho(y)\frac{\partial y_j}{\partial_i}\mathrm{d}y = \delta_{ij}.
\end{aligned} \tag{C.5}$$

Therefore, when ε tends to zero, using the fact that $r_\varepsilon \underset{\varepsilon\to 0}{\longrightarrow} 0$ (by (C.5)), we obtain

$$B.\nabla\big(h(w)\big) = c\dot{h}(w)\mathcal{L}^n.$$

□

Before we introduce the normalization property for BV_{loc} functions, we recall some definitions and properties for BV_{loc} functions.

Definition 40. *Let $\Omega \subseteq \mathbb{R}^n$. We say that $B \in L^1(\Omega)$ has bounded variation in Ω, and we denote $B \in BV(\Omega)$, if*

$$\sup\Big\{\int_\Omega B\ div\varphi\ \mathrm{d}x \mid \varphi \in C_c^1(\Omega, \mathbb{R}^n), |\varphi| \le 1\Big\} < \infty.$$

We also define the local version of the above concept.

Definition 41. *Let $\Omega \subseteq \mathbb{R}^n$. We say that $B \in L^1_{loc}(\Omega)$ has locally bounded variation in Ω, and we denote $B \in BV_{loc}(\Omega)$, if for every open set $U \subset \Omega$,*

$$\sup\Big\{\int_U B\ div\varphi\ \mathrm{d}x \mid \varphi \in C_c^1(U, \mathbb{R}^n), |\varphi| \le 1\Big\} < \infty.$$

Let $B \in BV_{loc}(\Omega)$. The structure Theorem for BV_{loc} functions (see Theorem 1, section 5.1 in [EG92]) asserts that the first derivative DB of B is a $\mathbb{R}^n$-valued measure of finite total variation $|DB|$ with

$$|DB|(\Omega) := \sup\Big\{\int_U B\ div\varphi\ \mathrm{d}x \mid \varphi \in C^1_c(U, \mathbb{R}^n), |\varphi| \leq 1, \forall U \subset \Omega\Big\}.$$

Hence, DB admits a polar decomposition

$$DB = N|DB| \text{with } |N(x)| = 1, a.e.\ x \in \Omega.$$

By The Lebesgue Decomposition Theorem (Theorem 3, section 1.6.2 in [EG92]), we may set

$$D.B = D^a B + D^s B$$

such that

$$|D^a B| << \mathcal{L}^n \text{ and } |D^s B| \perp \mathcal{L}^n.$$

We may obtain $DB = D^a B + trace(N)|D^s B|$. In particular,

$$DB << \mathcal{L}^n \text{ is equivalent to } trace(N) = 0.$$

To establish the renormalization property for BV_{loc} functions, we need the following property.

Definition 42. *Let ρ be a convolution kernel and N be a $n \times n$-matrix, we define*

1. *the anisotropic energy of ρ given by*

$$\Lambda(N, \rho) := \int_{R^n} | < Nz, \nabla\rho(z) > |\mathrm{d}z.$$

2. *the isotropic energy of ρ given by*

$$I(\rho) := \int_{\mathbb{R}^n} |z|\ |\nabla\rho(z)|\mathrm{d}z.$$

Proposition 26. ***(Optimal commutator estimates)***
Let $B \in BV_{loc}(\mathbb{R}^n)$ and $w \in L^\infty_{loc}(\mathbb{R}^n, \mathbb{R}^k)$. Let ρ be a convolution kernel such that for some $\varepsilon > 0$, we set

$$r_\varepsilon := B \cdot \nabla(w * \rho_\varepsilon) - (B \cdot \nabla w) * \rho_\varepsilon.$$

Then, for any compact set $K \subset \mathbb{R}^n$,

$$\limsup_{\varepsilon \to 0} \int_K |r_\varepsilon| \mathrm{d}x \leq ||w||_\infty I(\rho) |D^s B|(K) \tag{C.6}$$

and

$$\limsup_{\varepsilon \to 0} \int_K |r_\varepsilon| \mathrm{d}x \leq ||w||_\infty \int_K I(N(x), \rho) d|D^s B|(x) + ||w||_\infty (n + I(\rho)) |D^a B|(K). \tag{C.7}$$

We show that the weak solutions of the transport equations verify a renormalization property. We present the renormalization result in the case where we suppress the time dependance.

Proposition 27. *We assume that $B \in BV_{loc}(\mathbb{R}^n)$ such that $DB << \mathcal{L}^n$. Let $w \in L^\infty_{loc}(\mathbb{R}^n, \mathbb{R}^k)$ satisfying*

$$B.\nabla w_i = c_i \mathcal{L}^n, i = 1, \ldots, n$$

with $c \in L^1_{loc}(\mathbb{R}^n, \mathbb{R}^k)$.

Then,

$$B.\nabla h(w) = \sum_{i=1}^{k} \frac{\partial h}{\partial z_i}(w) c_i \mathcal{L}^n, \forall h \in C^1(\mathbb{R}^k).$$

Proof of Proposition 27. We set

$$\sigma := B.\nabla h(w) - \sum_{i=1}^{k} \frac{\partial h}{\partial z_i}(w) c_i \mathcal{L}^n.$$

By (C.6), σ is a measure absolutely continuous with respect to $|D^s B|$.
By (C.7), we have

$$|\sigma| \leq C(h, w) \Lambda(N, \rho) |D^s B| + C(h, w, \rho) |D^a B|$$

then,

$$|\sigma| \leq C(h, w) \Lambda(N, \rho) |D^s B|.$$

As σ is independant of ρ , then

$$|\sigma| \leq C(h, w) \inf_\rho \Lambda(N, \rho) |D^s B|.$$

Thanks to the Alberti Lemma (see Theorem 2.2 in [Amb04]), we have

$$\inf\left\{\Lambda(N,\rho) : \rho \in C_c^\infty(\mathbb{R}^n), \rho \geq 0, \int_{\mathbb{R}^n} \rho = 1\right\} = |trace(M)|.$$

Since $DB << \mathcal{L}^n$, we have $trace(N) = 0$ and the conclusion follows. □

The proposition 27 is necessary to obtain the following comparison principle.

Theorem 29. *Assume that $b_t(x) = b(t,x) \in BV_{loc}(\mathbb{R}^n)$ such that*

1. $D.b_t = divb_t\mathcal{L}^n$,
2. $|b_t| + |divb_t| \in L^1_{loc}([0,1] \times \mathbb{R}^n)$,
3. $\int_0^1 ||[divb_t]^-||_{L^\infty(B_R)}\mathrm{d}t < +\infty, \forall R > 0.$

For $i = 1, 2$, let $w_t^i(x) = w^i(t,x) \in L^\infty_{loc}([0,1] \times \mathbb{R}^n)$ verifying the transport equation

$$\frac{\partial w_t^i}{\partial t} + b_t \nabla w_t^i = -w_t^i \ divb_t \ in \ [0,1] \times \mathbb{R}^n.$$

Then, $w_0^1 \leq w_0^2$ implies $w_t^1 \leq w_t^2, \forall t \in [0,1]$.

Proof of Theorem 29. We set $w_t = w_t^1 - w_t^2, \forall t \in [0,1]$ such that, by hypothesis, it verifies

$$\frac{\partial w_t}{\partial t} + b_t.\nabla w_t = -w_t \ div \ b_t \quad \text{in } [0,1] \times \mathbb{R}^d.$$

For a given $\varepsilon > 0$, we define $\beta_\varepsilon(t) = \sqrt{\varepsilon^2 + t^2} - \varepsilon \in C^1(\mathbb{R})$ such that

$$\lim_{\varepsilon \to 0} \beta_\varepsilon(t) = t.$$

Using the renormalization property (see Proposition 27), we get

$$\frac{\partial \beta_\varepsilon}{\partial t}(w_t) + b_t.\nabla\big(\beta_\varepsilon(w_t)\big) = -div \ b_t \ \dot{\beta}_\varepsilon(w_t)w_t.$$

Thanks to the inequality $-\beta_\varepsilon(t) \leq \beta_\varepsilon(t) - t\dot{\beta}_\varepsilon(t) \leq 0$, and since $w_t \in L^\infty_{loc}([0,1] \times \mathbb{R}^n)$, there is $R > 0$ such that

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\int_{B_R} \beta_\varepsilon(w_t)\mathrm{d}x &= \int_{B_R} div \ b_t \ \beta_\varepsilon(w_t)\mathrm{d}x - \int_{B_R} divb_t \ \dot{\beta}_\varepsilon(w_t)w_t\mathrm{d}x \\ &= \int_{B_R} div \ b_t \left[\beta_\varepsilon(w_t) - \dot{\beta}_\varepsilon(w_t)w_t\right]\mathrm{d}x\end{aligned}$$

$$= -\int_{B_R} div\ b_t[\dot{\beta}_\varepsilon(w_t)w_t - \beta_\varepsilon(w_t)]\mathrm{d}x$$

$$\leq \int_{B_R} [div\ b_t]^-\ \beta_\varepsilon(w_t)\mathrm{d}x$$

$$\leq ||[div\ b_t]^-||_{L^\infty(B_R)} \int_{B_R} \beta_\varepsilon(w_t)\mathrm{d}x$$

Hence, for a.e. $x \in B_R$,

$$\frac{\mathrm{d}}{\mathrm{d}t}\beta_\varepsilon(w_t) \leq ||[div\ b_t]^-||_{L^\infty(B_R)}\beta_\varepsilon(w_t).$$

Let $\varepsilon \to 0$, we obtain

$$\frac{\partial}{\partial t}w_t \leq ||[div\ b_t]^-||_{L^\infty(B_R)}w_t.$$

Applying the Gronwall Lemma, we get $\forall t \in [0, 1]$

$$w_t \leq exp\Big(\int_0^t ||[div\ b_s]^-||_{L^\infty(B_R)}\mathrm{d}s\Big)\ w_0. \tag{C.8}$$

Since $w_0 = w_0^1 - w_0^2 \leq 0$, we deduce by (C.8)

$$w_t = w_t^1 - w_t^2 \leq 0.$$

□

C.3 Proof of Theorem 26

Our aim is to prove, under suitable assumptions, existence and uniqueness of solution for the Cauchy problem fro BV vector fields.

Let $A \subset \mathbb{R}^n$ be fixed.

Proposition 28. *There exists a $\mathcal{L}^n$-measurable application $\eta : A \to \mathcal{P}(\mathscr{C})$ such that*

$$\textit{for a.e. } x \in A,\ \eta_x\big(\mathscr{C}\backslash\mathscr{C}_x^b\big) = 0.$$

Proof of Proposition 28. Let $\gamma : A \to \mathscr{C}$ be a function. By a continuity argument, it is sufficient to prove that for a.e. $x \in A$,

$$\gamma(x)(t) := x + \int_0^t b(\tau, \gamma(\tau))\mathrm{d}\tau, \ \forall t \in [0,1].$$

We denote by $b_h := b * \rho_h$ the smooth approximation of b by convolution such that $b_h : [0,1] \times \mathbb{R}^n \to \mathbb{R}^n$ verifies the following conditions:

(i) $b_h \in L^\infty([0,1] \times \mathbb{R}^n, \mathbb{R}^n)$ and $b_h \underset{h\to\infty}{\longrightarrow} b$ in $L^1_{loc}([0,1] \times \mathbb{R}^n, \mathbb{R}^n)$.

(ii) $\nabla b_h(t,x) \in L^\infty([0,1] \times B_R, \mathbb{R}^n)$, $\forall r > 0$.

Let $\gamma_h(x)(.)$ be the unique solution of the Cauchy problem $\dot\gamma(t) = b_h(t, \gamma(t))$ with initial condition $\gamma(0) = x$. We recall that (see Proposition B.1 [BR16]), for every $R > 0$, there is a smooth function $J_h : [0,1] \times B_R \to [0, \infty[$, which is the Jacobian of the application $x \longmapsto \gamma(x)(t)$ such that for every $t \in [0,1]$, and for every $x \in B_R$, we have

$$\frac{\partial J_h}{\partial t}(t,x) = div\ b_h\big(t, \gamma_h(x)(t)\big)\ J_h(t,x).$$

Integration in time t yields:

$$e^{-C_R(t)} \le J_h(t,x) \tag{C.9}$$

where

$$C_R(t) := \int_0^1 ||[div\ b_t]^-||_{L^\infty(B_{M_R})}\mathrm{d}t$$

and

$$M_R = R + ||b_t||_\infty.$$

In particular, for any $\varphi \in C_c^\infty(\mathbb{R}^n)$, the changement of variables $y = \gamma_h(x)(t)$ and (C.9) give

$$\begin{aligned} |\int_{B_R} \varphi\big(\gamma_h(x)(t)\big)\mathrm{d}x| &\le e^{C_R(t)} \int_{B_R} |\varphi\big(\gamma_h(x)(t)\big)| J_h(t,x)\mathrm{d}x \\ &\le e^{C_R(t)} \int_{B_{M_R}} |\varphi(y)|\mathrm{d}y. \end{aligned} \tag{C.10}$$

Let $\bar t \in [0,1]$ be fixed. We define $\forall x \in B_R$, $\forall \gamma(x) \in \mathscr{C}$

$$\phi(x, \gamma(x)) := |\gamma(x)(\bar t) - x - \int_0^{\bar t} b(\tau, \gamma(x)(\tau))\mathrm{d}\tau|. \tag{C.11}$$

We define now a family of measures $(\eta_h) : B_R \to \mathcal{P}(\mathscr{C})$ by

$$\int_{B_R \times \mathscr{C}} \phi(x, \gamma(x)) \mathrm{d}\eta_{hx}(\gamma) \; \mathrm{d}x := \int_{B_R} \phi(x, \gamma_h(x)) \; \mathrm{d}x. \tag{C.12}$$

The following theorem is crucial. We refer the reader to see [Youn69] and section 2 in [AlMu01] for its proof.

Theorem 30. ***(Fundamental Theorem on Young measures)*** *Let K be a compact metric space. Let $A \subset \mathbb{R}^n$ be a bounded $\mathcal{L}^n$-measurable set and (η_h) be a sequence of $\mathcal{L}^n$-measurable measure-valued maps from A to $\mathcal{P}(K)$. Then, there exist a $\mathcal{L}^n$-measurable measure-valued map $\eta : A \to \mathcal{P}(K)$ and a subsequence $h(k)$ such that*

$$\lim_{k \to \infty} \int_A \int_K \phi(x, u) \mathrm{d}\eta_{h(k)x}(u) \; \mathrm{d}x = \int_A \int_K \phi(x, u) \; \mathrm{d}\eta_x(u) \mathrm{d}x$$

for any bounded function $\phi(x, u) : A \times K \to [0, +\infty[$ continuous with respect to u and $\mathcal{L}^n$-measurable with respect to x.

Hence, there exist a $\mathcal{L}^n$-measurable application $\eta : A \to \mathcal{P}(\mathscr{C})$ and a subsequence $h(k)$ such that

$$\lim_{k \to \infty} \int_{B_R} \int_{\mathscr{C}} \phi(x, \gamma(x)) \mathrm{d}\eta_{h(k)x}(\gamma(x)) \mathrm{d}x = \int_{B_R} \int_{\mathscr{C}} \phi(x, \gamma(x)) \mathrm{d}\eta_x(\gamma(x)) \mathrm{d}x. \tag{C.13}$$

Then, we obtain

$$\begin{aligned}
&\int_{B_R} \int_{\mathscr{C}} |\gamma(x)(\bar{t}) - x - \int_0^{\bar{t}} b(\tau, \gamma(x)(\tau)) \mathrm{d}\tau| \mathrm{d}\eta_x(\gamma(x)) \mathrm{d}x \\
&\qquad = \int_{B_R} \int_{\mathscr{C}} \phi(x, \gamma(x)) \mathrm{d}\eta_x(\gamma(x)) \mathrm{d}x, && \text{by (C.11)} \\
&\qquad = \lim_{k \to \infty} \int_{B_R} \int_{\mathscr{C}} \phi(x, \gamma(x)) \mathrm{d}\eta_{h(k)x}(\gamma(x)) \mathrm{d}x, && \text{by (C.13)} \\
&\qquad = \lim_{k \to \infty} \int_{B_R} \phi(x, \gamma_{h(k)}(x)) \mathrm{d}x, && \text{by (C.12)} \\
&\qquad = \lim_{k \to \infty} \int_{B_R} |\gamma_{h(k)}(x)(\bar{t}) - x - \int_0^{\bar{t}} b(\tau, \gamma_{h(k)}(x)(\tau)) \mathrm{d}\tau| \mathrm{d}x, && \text{by (C.11)}
\end{aligned}$$

$$= \lim_{k\to\infty} \int_{B_R} |\int_0^{\bar{t}} (b_{h(k)} - b)(\tau, \gamma_{h(k)}(x)(\tau))\mathrm{d}\tau|\mathrm{d}x, \qquad \text{(because } \gamma_h \in \mathscr{C}_x^{b_h}\text{)}$$

$$\leq \limsup_{k\to\infty} \int_{B_R} \int_0^1 |b_{h(k)} - b|(\tau, \gamma_{h(k)}(x)(\tau))\mathrm{d}\tau|\mathrm{d}x$$

$$\leq \limsup_{k\to\infty} \int_0^1 e^{C_R(\tau)} \int_{B_{M_R}} |b_{h(k)} - b|(\tau, y)\ \mathrm{d}y\ \mathrm{d}\tau, \qquad \text{by (C.10)}$$

$$= 0, \qquad \text{because } b_h \underset{h\to\infty}{\to} b \text{ in } L^1_{loc}([0,1]\times\mathbb{R}^n, \mathbb{R}^n).$$

And the conclusion follows. □

Proposition 29. *For a.e. $x \in A$, η_x is a Dirac measure.*

Proof of Proposition 29. By Lemma 28 and Theorem 27, the problem can be reduced to prove the following condition:

$\forall t \in [0,1]$, there exists a negligeable set A_t of A such that

$$\eta_x\big(\gamma(x) : \gamma(x)(t) \in K\big)\ \eta_x\big(\gamma(x) : \gamma(x)(t) \in K'\big) = 0, \forall x \in A\backslash A_t$$

where K, K' are two disjoint subsets of $\mathbb{R}^n$.

We assume by contradiction that there are $t_0 \in [0,1]$ and a negligeable set A_{t_0} of A such that there are two disjoint subsets K and K' of $\mathbb{R}^n$ verifying

$$\eta_x\big(\gamma(x) : \gamma(x)(t_0) \in K\big)\ \eta_x\big(\gamma(x) : \gamma(x)(t_0) \in K'\big) > 0, \forall x \in A\backslash A_{t_0}.$$

Without loss of generality, we can assume that there is a constant $C > 0$ such that $\forall x \in A\backslash A_{t_0}$

$$0 < \eta_x\big(\gamma(x) : \gamma(x)(t_0) \in K\big) \leq C\ \eta_x\big(\gamma(x) : \gamma(x)(t_0) \in K'\big). \tag{C.14}$$

Therefore, we define the following measures

$$\eta_x^1 := \eta_x\ \llcorner\{\gamma(x) : \gamma(x)(t_0) \in K\}$$

and

$$\eta_x^2 := C\ \eta_x\ \llcorner\{\gamma(x) : \gamma(x)(t_0) \in K'\}.$$

For $i = 1, 2$, and $\forall t \in [0,1]$, we associate the measure μ_t^i given by

$$< \mu_t^i, \varphi > := \int_A \int_{\mathscr{C}} \varphi(\gamma(t)) \mathrm{d}\eta_x^i(\gamma) \mathrm{d}x, \ \forall \varphi \in C_c^\infty(\mathbb{R}^n, \mathbb{R}).$$

By Proposition 28, we have for a.e. $x \in A$, $\eta_x(\mathscr{C} \backslash \mathscr{C}_x^b) = 0$.
Hence, for a.e. $x \in A$ and for any $\gamma(x) \in \mathscr{C}$,

$$\dot{\gamma}(x)(t) = b(t, \gamma(x)(t)) = b_t(\gamma(x)(t)), \forall t \in [0,1].$$

Therefore, we obtain for $i = 1, 2$, for any $\varphi \in C_c^\infty(\mathbb{R}^n, \mathbb{R})$

$$\begin{aligned}
\frac{d}{dt} < \mu_t^i, \varphi > &= \frac{d}{dt} \int_A \int_{\mathscr{C}} \varphi\big(\gamma(x)(t)\big) \ \mathrm{d}\eta_x^i(\gamma(x)) \ \mathrm{d}x \\
&= \int_A \int_{\mathscr{C}} \nabla\varphi\big(\gamma(x)(t)\big).\dot{\gamma}(x)(t) \ \mathrm{d}\eta_x^i(\gamma(x)) \ \mathrm{d}x \\
&= \int_A \int_{\mathscr{C}} \nabla\varphi\big(\gamma(x)(t)\big).b_t\big(\gamma(x)(t)\big) \ \mathrm{d}\eta_x^i(\gamma(x)) \ \mathrm{d}x \\
&= < \mu_t^i \ , \ b_t \nabla\varphi > \\
&= < b_t \ \mu_t^i \ , \ \nabla\varphi > .
\end{aligned}$$

This means that the measures $\mu_t^i, i = 1, 2$ are solutions of the transport equation

$$\begin{cases}
\dfrac{\partial \mu_t^i}{\partial t} & + \quad D.(b_t \mu_t^i) = 0 \\
\mu_0^1 & = \quad \eta_{t_0 x}(K)\mathcal{L}^n \lfloor A \\
\mu_0^2 & = \quad C \ \eta_{t_0 x}(K')\mathcal{L}^n \lfloor A
\end{cases}$$

By (C.14), we note that

$$\mu_0^1 \leq \mu_0^2.$$

Thanks to the comparison principle (see Theorem 29), we have

$$\mu_t^1 \leq \mu_t^2, \forall t \in [0,1].$$

On the other hand, since K and K' are disjoint, we have for any $t \in [0,1]$

$$\mu_t^1 = \int_A \eta_{t_0 x} \lfloor K \mathrm{d}x \quad \perp \quad C \int_A \eta_{t_0 x} \lfloor K' \mathrm{d}x = \mu_t^2$$

which implies the contradiction. $\square$

Bibliography

[ABB13] A.Agrachev, Davide Barilari and U.Boscain, *Introduction to Riemannian and sub-Riemannian geometry (from Hamiltonian viewpoint)*, Preprint SISSA 09/2012/M, 2013.

[AL09] A.Agrachev and P. Lee, *Optimal transportation under nonholonomic constraints*, Trans. Amer. Math. Soc., 361(11), 6019- 6047, 2009.

[Alex39] A.D.Alexandrov, *Almost everywhere existence of the second differential of a convex function and some properties of convex surfaces connected with it*, Leningrad state Univ. Annals [Uchenye Zapiski] Math. See. 6, 3-36, 1939.

[AlMu01] G.Alberti and S.Müller, *A new approach to variational problems with scales*, Commun. Pure Appl.Math. 54, 761-825, 2001.

[AM03] L.Ambrosio and V.Magnani, *Weak differentiability of BV functions on stratified groups*, Math. Z.245, 123-153, 2003.

[Amb03] L.Ambrosio, *Lecture notes on transport equation and Cauchy problem for BV vector fields and applications*, Available at http://cvgmt.sns.it, 2003.

[Amb04] L.Ambrosio, *Transport equation and Cauchy problem for BV vector fields*, Invent. Math. 158, 227-260, 2004.

[App87] P.Appell, *Mémoire sur les déblais et les remblais des systèmes continus ou discontinus*, Mémoires présentés par divers Savants à l'Académie des Sciences de l'Institut de France, Paris 29, 1-208, 1887.

[AR04] L. Ambrosio and D. Rigot, *Optimal transportation on the Heisenberg group*, J. Funct. Anal. 208(2), 261-301, 2004.

[Bad17] Z.Badreddine, *Mass transportation on sub-Riemannian structures of rank 2 in dimension 4*, to appear in Annales de l'IHP(C) Nonlinear Analysis.

[BB05] P.Bernard and B.Buffoni, *Optimal mass transportation and Mather theory*, Journal of the European Mathematical Society, 9:85-121, 2005.

[BF36] H.Buesmann and W.Feller, *Krümmungsindikatritizen Konvexer Flchen*, Acta. Math. 66, 1-47, 1936.

[Br91] Y. Brenier, *Polar factorization and monotone rearrangement of vector- valued functions*, Comm.Pure Appl. Math., 44: 375-417, 1991.

[BR03] Z.M.Balogh and M.Rickly, *Regularity of convex functions on Heisenberg groups*, Ann.Scuola Norm. Sup. PisaCl. Sci., II(5):847-868, 2003.

[BR16] A. Belotto da Silva and L. Rifford, *The sub-Riemannian Sard conjecture on Martinet surfaces* , 2016.

[CH15] F. Cavalletti and M. Huesmann, *Existence and uniqueness of optimal transport maps*, Annales of IHP(C) Nonlinear Analysis, vol. 32, Issue 6, Pages 1367-1377, 2015.

[Cho39] C.-L. Chow, *Über systeme von linearen partiellen Differentialgleichungen ester ordnung*, Math. Ann., 117, 98-105, 1939.

[CP10] A.Calogero and R.Pini, *c horizontal convexity on Carnot groups*, arXiv:1005.0975, 2010.

[CS04] P. Cannarsa and C. Sinestrari, *Semiconcave Functions, Hamilton- Jacobi Equations and Optimal Control*, Progress in Nonlinear Differential Equations and Their Applications , vol. 58 (Birkhäuser), 2004.

[C-V59] S.E. Cohn-Vossen, *Some questions of differential geometry in the large*, Fizmatgiz Moscow (in Russian), 1959.

[Dar85] Darboux, G. Prix Bordin (géométrie), C.R.Accad.Sci.Paris 101, 1312-1316, 1885.

[DGN03] D.Danielli, N.Garofalo and D.M.Nhieu, *Notions of convexity in Carnot groups*, Comm. Anal. Geom.11(263-341), 2003.

[DGNT04] D.Danielli, N.Garofalo, D.M.Nhieu and F.Tournier, *The theorem of Busemann-Feller-Alexandrov in Carnot groups*, Comm.Anal.Geom.12(4),853-886, 2004.

[DGS03] D.Danielli, N.Garofalo and S.Salsa, *Variational inequalities with lack of ellipticity. I. optimal interior regularity and non-degeneracy of the free boundary*, Indiana Univ. Math. J., 52:361:398, 2003.

[DL89] R.J.Diperna and P.L.Lions, *Ordinary differential equation, transport theory and Sobolev spaces*, Invent.Math. 98,511-547, 1989.

[EG92] L.Evans and R.Gariepy, *Measure Theory and Fine Properties of Functions*, CRC PRESS LLC, Studies in Advanced Mathematics, 1992.

[FR10] A. Figalli and L. Rifford, *Mass transportation on sub-Riemannian manifolds*, Geom. Funct. Anal., 20(1), 124-159, 2010.

[GM04] C.E.Guitérrez and A.Montanari, *On the second derivatives of convex functions on the Heisenberg group*, ANN.Scuola Norm. Sup. Pisa Cl. Sic.(5), Vol. III, pp. 349-366, 2004.

[GM05] C.E.Gutiérrez and A.Montanari, *Maximum and comparison principles for convex functions on the Heisenberg group*, Comm. in Partial Differential Equations, Issue 9-10, volume 29, 2005.

[GN98] N.Garofalo and D.M.Nhieu, *Lipschitz continuity, global smooth approximation and extension theorems for Sobolev functions in Carnot-Caratheodory spaces*, Jour.Anak.Math.74, 67-97, 1998.

[How98] R.Howard, *Alexandrov's theorem on the second derivatives of convex functions via Rademacher's theorem on the first derivatives of Lipschitz Functions*, Lectures given at the functional analysis seminar at the University of South Carolina, 1998.

[HR31] H.Hopf and W.Rinow, *Ueber den Begriff der vollständigen differentialgeometrischen Flächen*, Comm.Math.Helv. 3 pp. 209-225, 1931.

[JLMS07] P.Juutinen, G. Lu, J.Manfredia and B.Stroffolini, *Convex functions in Carnot groups*, REV. MAT. IBEROAMERICANA 23, no.1, 191-200, 2007.

[Ju09] N.Juillet, *Geometric inequalities and generalized Ricci bounds on the Heisenberg group*, Int. Math. Res. Not. IMRN, (13): 2347-2373, 2009.

[Ka42] L. Kantorovitch, *On the translocation of masses*, C.R. (Docklady) Acad. Sa. URSS, 37: 199-201, 1942.

[LeDo17] E. Le Donne, *Lecture notes on sub-Riemannian geometry*, enrico.ledonne.googlepages.com, 2017.

[LMS04] G.Lu, J.Manfredi and B.Stroffollini, *Convex functions on the Heisenberg group*, Calc.Var. PDE, 2004.

[LS95] W. Liu, H.J. Sussmann, *Shortest paths for sub-Riemannian metrics on rank-two distributions*, Mem. Amer. MAth. Soc., 118:564, 1995.

[Mag05] V.Magnani, *Lipschitz continuity, Alexandrov theorem and characterizations for H-convex functions*, Springer-Verlag, 2005.

[Mc01] R. McCann, *Polar factorization of maps in Riemannian manifolds* , Geom. Funct. Anal. 11: 589-608, 2001.

[Mont02] R.Montgomery, *A tour of sub-Riemannian geometries, their geodesics and applications*, Mathematical Surveys and Monographs, vol.91, American Mathematical Society, Providence, RI, 2002.

[Mon81] G. Monge, *Mémoire sur la théorie des déblais et des remblais*, Histoire de l'Académie Royale des Sciences de Paris: 666-704, 1781.

[MS01] R. Monti, F. Serra Cassano, *Surface measures in Carnot-Caratheodory spaces*, Calc. Var. Partial Differential Equations 13:3, 339-376, 2001.

[Oh07] S.-I. Ohta, *On the measure contraction property of metric measure spaces*, Comm. Math. Helv., 82(4): 805-828, 2007.

[Pan89] P.Pansu, *Métriques de Carnot-Carathéodory et quasiiométriques des espaces symétriques de rang un*, Annals of Mathematics, second series, 129(1):1-60, 1989.

[Pra07] A.Pratelli, *On the equality between Monge's infimum and Kantorovich's minimum in optimal mass transportation*, Ann. Inst. H. Poincaré Probab. Statist. 43(1):1-13, 2007.

[Rad20] H.Rademacher, *Über partielle und totale Differenzierbarkeit von Funktionen mehrerer Variabeln II.*, Math. Anna. 81(1), 52-63, 1920.

[Ras38] P.K. Rashevsky, *About connecting two points of a completely nonholonomic space by admissible curve*, Uch. Zapiski Ped. Inst. Libkenchta, 2, 83-94, 1938.

[Rif13] L. Rifford, *Ricci curvature in Carnot groups*, Math. Control Relat. Fields 3, no. 4, 467-487. MR 3110060, 2013.

[Rif14] L. Rifford, *Sub-Riemannian Geometry and optimal transport*, Springer Briefs in Mathematics,140 pp., 2014.

[Riz16] L. Rizzi, *Measure contraction property of Carnot groups*, arXiv:1510.05960v2 [math.MG], 2016.

[Sr98] A.-M. Srivastava, *A course on Borel sets*, Springer, 1998.

[Sud79] V.N. Sudakov, *Geometric problems in the theory of infinite dimensional distributions*, Proceedings of Steklov Institute, 141, 1-178, 1979.

[Sus08] H.J. Sussmann, *Smooth distributions are globally finitely spanned*, In Analysis and design of nonlinear control systems, Springer, Berlin 3-8, 2008.

[Sus96] H.J. Sussmann, *A cornucopia of abnormal sub-Riemannian minimizers*, Birkhäusser 341-364, 1996.

[St06] K.-T. Sturm, *On the geometry of metric measures spaces II*, Acta. Math., 196(1): 133-177, 2006.

[Stri86] R.S. Strichartz, *Sub-Riemannian geometry*, J. Differential Geom., 24(2):221-263, 1986.

[Var74] V.S. Varadarajan, *Lie groups, Lie Algebras, and Their Representations*, Springer-Verlag, New-York, Berlin, Heidelberg, Tokyo, 1974.

[Vil03] C. Villani, *Topics in Mass Transportation*, Graduate Studies in Mathematics Surveys, Vol. 58, American Mathematical Society, Providence, RI, 2003.

[Vil08] C. Villani, *Optimal transport, Old and New*, Springer, Berlin, 2008.

[Wan05] C.Y.Wang, *Viscosity convex functions on Carnot groups*, Proceedings of the American Mathematical Society, Vol.133, no.4, pp. 1247-1253, 2005.

[Youn69] L.C.Young, *Lectures on the calculus of variations and optimal control theory*, Saunders, 1969.