Introduction to sub-Riemannian Geometry

John Darges

May 4, 2020

1 Introduction

Sub-Riemannian geometry is a natural generalization of Riemannian geometry in which a Riemannian structure is imposed only partially onto a manifold. As such, one can actually think of Riemannian geometry as a special case of Riemannian geometry. In simple terms, we can think of imposing a Riemannian structure on a smooth manifold M as defining scalar product onto the tangent bundle TM. We call this product a Riemannian metric and it is defined with vector fields, which can be thought as functions existing within the tangent bundle, as its inputs. Thus, the length of a vector field defines a function on the manifold M. Crucially, the metric is represented as a very specific class of square matrix, that is symmetric and positive definite. This matrix is always invertible since the metric is defined to positive definite everywhere on the tangent bundle. In sub-Riemannian geometry, we have a metric that only satisfies the characteristics of being a metric on a portion of the tangent bundle. This portion forms a sub-bundle of the tangent bundle and is referred to as the collection of distributions since the fibers of this sub-bundle are called distributions. Therefore, unless we are studying the Riemannian case, the bundle rank of the set of distributions is always less than that of the tangent bundle. This sub-bundle depends on the choice of metric. The representation of the metric as a matrix is key here. We choose symmetric matrices which are not invertible and thus not everywhere positive definite. Therefore, the set of distributions always excludes the kernel of this matrix.

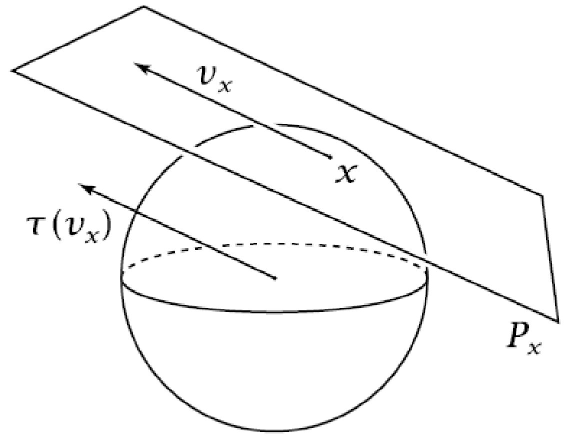

Figure 3.0: Tangent to a sphere

There has developed a close correspondence between the areas of sub-Riemannian geometry and optimal control theory. We will take a look at optimal control theory in a later portion of this paper. In brief, optimal control theory studies optimization problems of systems of differential equations where only certain solution curves in the target space are deemed admissible. We can see an analogy here with sub-Riemannian geometry in that only certain velocity vectors in our tangent bundle are contained within distributions. From this, we much of the language of sub-Riemannian geometry will follow the language of control theory. The correspondence exists in that geometric control problems are studied within the context of sub-Riemannian geometry. Additionally, problems in sub-Riemannian geometry, particularly that of characterizing length-minimizing curves can be expressed in the manner of control theory.

While we have talked about the core of sub-Riemannian structure existing as sub-bundles of the tangent bundle, there exist generalizations of sub-Riemannian structure that define it through arbitrary Euclidean bundles. We will first define sub-Riemannian structure in the most general context, though, for practical purposes, the cases we will think of will be sub-bundles of the tangent bundle, as this is the easiest case to conceptualize. After defining sub-Riemannian structure, we will see that most concepts of Riemannian geometry readily generalize and that the same core theorems also hold in the more general case. Finally, we will explore one of the relationships between control theory and sub-Riemannian geometry.

2 Sub-Riemannian Structure

2.1 Defining sub-Riemannian Structure

Before defining sub-Riemannian structure, we must first cover the important concept of being bracket-generating. Given a family of vector fields $\mathcal{F}$ on a smooth manifold M, the Lie algebra generated by $\mathcal{F}$ is denoted $Lie_{\mathcal{F}}$. $\mathcal{F}$ is called **bracket-generating** if $Lie^p_{\mathcal{F}} = T_pM$ for every $p \in M$, where $Lie^p_{\mathcal{F}}$ denotes the set of vector fields of $Lie_{\mathcal{F}}$

fixed at p. $Lie_{\mathcal{F}}$ has step k if $\text{span}\{[X_1(p), ...[X_{j-1}(p), X_j(p)]] : X_i \in \mathcal{F}, j \leq k\} = T_pM$ for all $p \in M$. Note that the step can depend on our choice of p.

Now we give the general definition of a sub-Riemannian structure on a smooth manifold M.

A **sub-Riemannian structure** on M is a triple (U, f, g) where

i) U is a Euclidean bundle such that a scalar product $g(\cdot,\cdot)_p$, the **sub-Riemannian metric**, is defined on each fiber U_p, for all $p \in M$. Naturally, since our bundle is Euclidean, we can associate a matrix g^{ij} to our metric.

ii) f is a function $U \to TM$ that is linear on each fiber U_p and $\pi \circ f = \pi_U$, where $\pi : TM \to M$ and $\pi_U : U \to M$ are the canonical projections.

iii) Vector fields $f(\sigma)$, where $\sigma : M \to U$ is a smooth section, are referred to as **horizontal vector fields**. The set of all horizontal vector fields, $\mathcal{D} \subset TM$, of a sub-Riemannian structure is, by definition, bracket-generating. The step of the sub-Riemannian structure is the step of $\mathcal{D}$.

iv) The **distribution** of the structure is the collection of subspaces $\{\mathcal{D}_p\}_{p\in M}$ where $\mathcal{D}_p = f(U_p) \subset T_pM$. The sub-Riemannian length of a vector $V \in \mathcal{D}_p$ is given by $\|V\| = \min\{g(X, X)_p : V = f(X), X \in U_p\}$.

This generalized definition is quite intimidating, so we will consider a specific case of sub-Riemannian structure to give a clearer picture. From our definition, let U be a sub-bundle of TM such that the smooth sections of U (which are vector fields), are bracket-generating. We let f be the inclusion map $\iota : U \to TM$, which clearly satisfies the conditions above. When $U = TM$, this results in a Riemannian structure as, after all, sub-Riemannian geometry is a generalization. To have interesting cases, we will take U to be a proper sub-bundle. So, our metric will only be positive-definite on U, meaning that, as a matrix, g^{ij} will have non-trivial kernel and thus will *not* be invertible. Finally, we can see that the set of horizontal vector fields is just U itself while the distribution is the collection of fibers of U.

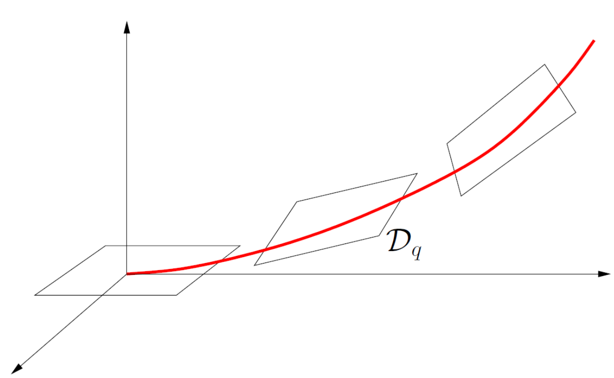

Figure 3.1: A horizontal curve

In essence, we can think of imposing a sub-Riemannian structure as being able to impose a Riemannian structure on only part of the tangent bundle of M. The purpose of having the generalized definition is, however, to study cases of sub-Riemannian manifolds with singularities where our horizontal vector fields may have singularities. In this case, we must introduce a measure for well-defined integration and volume. This results in an intersection with an area of study known as *geometric measure theory*. An example of measures we can define are *Hausdorff measures*: for a subset $E \subset M$,

i) the $\alpha-$**dimensional Hausdorff measure** is defined as $\mathcal{H}^\alpha(E) = \lim_{\epsilon \to 0^+} \mathcal{H}^\alpha_\epsilon(E)$ where

$$\mathcal{H}^\alpha_\epsilon(E) = \inf \left\{ \sum_{i=1}^{\infty} \mathrm{diam}(H_i)^\alpha : E \subset \bigcup_{i=1}^{\infty} H_i,\ H_i \text{ nonempty s.t. } \mathrm{diam}(H_i) < \epsilon \right\}$$

ii) the $\alpha-$**dimensional spherical Hausdorff measure** is defined as $\mathcal{S}^\alpha(E) = \lim_{\epsilon \to 0^+} \mathcal{S}^\alpha_\epsilon(E)$, where

$$\mathcal{S}^\alpha_\epsilon(E) = \inf \left\{ \sum_{i=1}^{\infty} \mathrm{diam}(S_i)^\alpha : E \subset \bigcup_{i=1}^{\infty} S_i,\ S_i \text{ open ball s.t. } \mathrm{diam}(S_i) < \epsilon \right\}$$

However, it is only important here to keep in mind that we have a concept of measurability for functions mapping from or onto M.

2.2 Curves in sub-Riemannian Manifolds

Now, we discuss the types of curves we find interesting in sub-Riemannian geometry. A **horizontal curve** is a Lipschitz curve $\gamma : [0,1] \to M$ for which there exists some measurable and essentially bounded function $u : [0,1] \to U$ such that $t \mapsto u(t) \in U_{\gamma(t)}$ such that $\gamma'(t) = u(t)$ a.e. on $[0,1]$. Moreover, u is called a **control** with respect to γ and may not be unique. An important remark is this: even if, for some curve γ, $\gamma'(t)$ lies in the distribution for almost every $t \in [0,1]$, the function $u(t) = \gamma'(t)$ may not be essentially bounded and thus γ possibly not horizontal.

Now, given a horizontal curve γ, define pointwise for every point of differentiability of γ a function, u_m by

$$u_m(t) = \mathrm{argmin}\{g(u,u)_{\gamma(t)} : u \in U_{\gamma(t)}, \gamma'(t) = f(u)\}$$

By Lemma 3.12 in [1], u_m is essentially bounded and measurable and therefore a control of γ. We call this the **minimal control** of γ. This, along with our notion of the length of a vector, allows us to have a definition for the length of a curve that satisfies our Riemannian instincts:

$$L(\gamma) = \int_0^1 \|\gamma'(t)\| dt = \int_0^1 g(u_m(t), u_m(t))_{\gamma(t)} dt$$

This definition gives us length which is invariant under reparametrization and gives us a notion of arc-length parametrization.

3 Sub-Riemannian Manifolds as Metric Spaces

3.1 Distance and Rashevskii-Chow Theorem

Finally, we now have the tools needed to define a distance, known as the **Caratheodory-Carnot distance** to endow the sub-Riemannian structure with the structure of a metric space. Given points $p, q \in M$, the distance between them is defined as

$$d(p,q) = \inf\{L(\gamma) : \gamma \text{ is horizontal }, \gamma(0) = p, \gamma(1) = q\}$$

This is clearly another generalization of the Riemannian distance, which leads us to our first important result:

Theorem 1 (Rashevskii-Chow) Let M be a sub-Riemannian manifold and d be the Caratheodory-Carnot distance. Then

i) (M, d) is a metric space

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

To prove Theorem 1, we require three lemmas which we will not prove:

Lemma 1.2 ([1]] 3.33) Let M be an $n-$dimensional sub-Riemannian manifold with generating family $\mathcal{F} = \{Y^1, ..., Y^k\}$. Then for every $p \in M$ and every neighborhood $V \subset \mathbb{R}^n$ of the origin, there exists $s \in V$ and n vector fields, $F_1, ..., F_n$, in $\mathcal{F}$ such that s is a regular point of the map

$$\psi : \mathbb{R}^n \to M),\ \psi(x) = \exp(x_n F_n(... \exp(x_1 F_1)p)$$

Lemma 1.3 ([1]] 3.36) Let $p \in M$ and $K \subset M$ be a compact subset such that $p \in \text{int}(K)$. Then there exists a number $\delta_K > 0$ such that every horizontal curve γ, with base-point p and length $L(\gamma) < \delta_K$, has its image contained in K.

Proof. Theorem 1.

i) To show d is a metric, symmetry follows from the fact that reversing orientation of an admissible curve preserves length.

   For the triangle inequality, consider three points $p, q, r \in M$ such that an admissible curve γ_1 connects p and q and an admissible curve γ_2 connects q and r. Then the concatenation of the two curves also results in an admissible curve which will have length equal to the sum of the other two lengths. Naturally, q may not fall along the trajectory from p to r, showing the triangle inequality. This shows that d is indeed a metric.

   To show $d(p,q) = 0$ if and only if $p = q$, Let $p \in K$ compact such that q is not in K. By Lemma 1.3, each admissible curve connecting p and q has length greater than δ_K, so the distance is strictly greater than zero.

   This shows d is a metric.

ii) To show that the two induced topologies are equivalent, we must show, for all $p \in M$:

1. For every $\epsilon > 0$, there exists a neighborhood O_p such that $O_p \subset B_\epsilon(p)$.
2. For every neighborhood O_p there exists $\delta > 0$ such that $B_\delta(p) \subset O_p$.

where O_p is a neighborhood under the manifold topology.

Let us start by proving the first statement: By Lemma 1.2, there exists a neighborhood $V' \subset V$ of s such that ψ is a diffeomorphism $V' \to \psi(V')$. To build a local diffeomorphism that contains p, consider the map

$$\psi'(s) = \exp(-s_n F_n(\dots \exp(-s_1 F_1)(s_1, \dots, s_n))$$

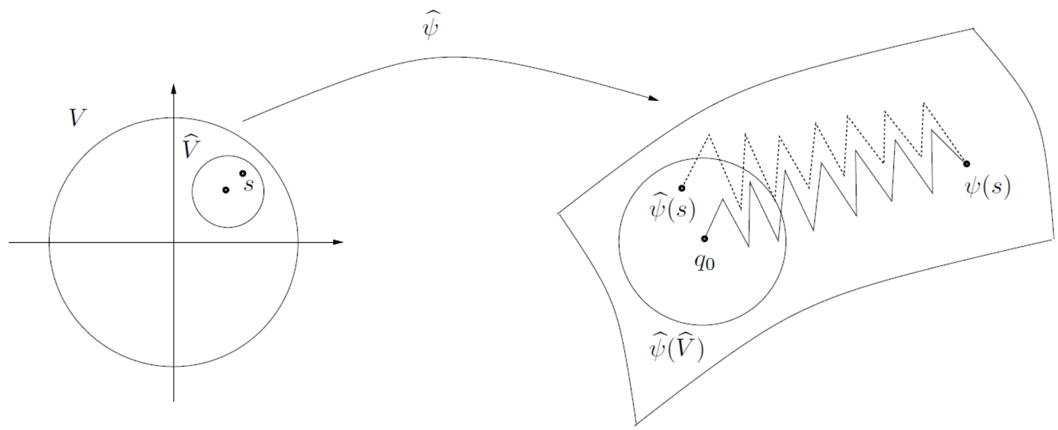

Figure 3.4: The map $\widehat{\psi}$

Then ψ' is a diffeomorphism on a neighborhood V' s' such that $\psi'(s') = p$. Now fix $\epsilon > 0$ and apply this to where V is the neighborhood $V = \{s : \|s\| < \epsilon\}$ of the origin in $\mathbb{R}^n$. Then statement 1 holds by letting $O_p = \psi'(V')$. In fact, for q in this set, let $s \in \mathbb{R}^n$ be the vector such that $q = \psi'(s)$ and let γ be an admissible curve connecting p and q constructed by concatenating the integral curves of the vector fields $F_1, \dots, F_n$. Then, it follows that

$$d(p,q) \leq L(\gamma) \leq \sum_{i=1}^{n}(|s_i| + |s_i'|) \leq 2\epsilon$$

Therefore, $O_p \subset B_\epsilon(p)$.

For the second statement, fix $\epsilon > 0$ and let K be a compact set containing p. Define C_K and δ_K to be as in Lemma 1.3 and fix $\delta = \min\{\delta_K, \epsilon/C_K\}$. Consider a sequence $\gamma_n : [0,1] \to M$ of admissible trajectories joining p and q. such that $L(\gamma_n)$ converges to $d(p,q)$ as $n \to \infty$. Without loss of generality, we have that $L(\gamma_n)$ is bounded by δ for all n. Then, by Lemma 1.3, $\gamma([0,1]) \subset K$ for all n. Therefore, taking $n \to \infty$, $q \in B_{C_K\delta}(p) \subset B_\epsilon(p)$. Therefore, when $d(p,q) \leq \delta$, then q is contained in an ϵ-neighborhood of p under the manifold topology.

□

3.2 Length Minimizers and Completeness

Here, we will discuss length-minimizers on sub-Riemannian manifolds and their existence. A **length-minimizer** of two points $p, q \in M$ is a horizontal curve γ such that $L(\gamma) = d(p,q)$. Previously, we defined the Carathedory-Carnot distance as an infimum and there is generally no way to guarantee that can be achieved as a minimum. However, we can characterize when length-minimizers exist. First, we give a lemma without proof.

Lemma 2.1 Let $\{\gamma_n : [0,1] \to M\}_{n\geq 1}$ be a sequence of horizontal curves parametrized by arc length such that the sequence converges to a curve γ uniformly and $\liminf\limits_{n\to\infty} L(\gamma_n)$ is bounded. Then $\gamma : [0,1] \to M$ is a horizontal curve and $L(\gamma)$ is bounded by $\liminf\limits_{n\to\infty} L(\gamma_n)$.

It follows that, since $[0,1]$ is compact, uniformly convergence sequences of length-minimizers converge to length-minimizers. Things brings us to our next major theorem:

Theorem 2 (Existence of minimizers) Let M be a sub-Riemannian manifold and $p \in M$. Assuming the closure of $B(p,r)$, for radius $r > 0$, is compact, then, for all $q \in B(p,r)$, there exists a length minimizer connecting p and q such that the Caratheodory-Carnot distance, $d(p,q)$, attains a minimum.

Proof. Fix $q \in B_r(p)$ and let $\gamma_n : [0,1] \to M$ be a minimizing sequence of admissible curves parametrized by constant speed which connect p and q such that $L(\gamma_n)$ converges to $d(p,q)$. Since $d(p,q) < r$ we have that $L(\gamma_n) \leq r$ for large enough n. Therefore, we assume without loss of generality that $\gamma_n([0,1])$ is contained in the closure of $B_r(p)$ for all n. In particular, this means that

$$|\gamma_n(t) - \gamma_n(\tau)| \leq \int_\tau^t |\dot{\gamma}_n(s)| ds \leq C_r r |t - \tau|,\ t, \tau \in [0,1]$$

for sufficiently large n where the constant C_r depends on the closure of $B_r(p)$. In other words, all curves in the sequence are Lipschitz with a common Lipschitz constant. This implies that the sequence is equicontinuous and uniformly bounded. Then, by Arzela-Ascoli theorem, there exists a subsequence of curves γ_{n_k} such that the subsequence uniformly converges to a Lipschitz curve $\gamma : [0,1] \to M$. Then, by Lemma 2.1, γ satisfies $L(\gamma) \leq \liminf L(\gamma_n) = d(p,q)$, implying that $L(\gamma) = d(p,q)$. $\square$

With this, we conclude our parallels to Riemannian structure by exploring completeness. Before, the statement on completeness, we provide a lemma:

Lemma 3.1 Let M be a sub-Riemannian manifold. For every $\epsilon > 0$ and $p \in M$, we have

$$B(p, r+\epsilon) = \bigcup_{q \in B(p,r)} B(q, \epsilon)$$

Theorem 3 (Completeness Criteria) Let M be a sub-Riemannian manifold. Then the following are equivalent:

i) (M,d) is complete.

ii) The closure of $B(p,r)$ is compact for every $p \in M$ and $r > 0$.

iii) There exists some $\epsilon > 0$ such that the closure of $B(p,\epsilon)$ is compact for every $p \in M$.

Proof. (iii) ⇒ i)) Assuming iii), we will show every Cauchy sequence converges. Let $\epsilon > 0$ be as in iii). Let $\{x_n\}_{n \geq 1}$ be such a Cauchy sequence such that $d(x_n, x_m) \leq \epsilon$ where $n, m > N$ for some N. Then x_n is in $B_\epsilon(x_m)$ for $n \geq m > N$. Since, by assumption, this is compact, then the sequence admits a convergent subsequence which implies that the sequence converges in M.

(ii) ⇒ iii)) Clearly there exists ϵ such that the closure of $B_\epsilon(p)$ is compact if we are assuming that it is compact for any $r > 0$. Simply set $\epsilon = r$.

(i) ⇒ ii)) Now we assume (M,d) is complete. Fix $p \in M$ and define a set A to be the set of radii r such that the closure of $B_r(p)$ is compact. Let R be the supremum of A. The set A is nonempty since the topology of (M,d) is locally compact. Furthermore, this implies that R is nonzero. Then, we will show that A is open and that $R = \infty$. Since a closed subset of a compact set is compact in (M,d), this means that if the closure of $B_r(p)$ is compact and $r > \rho$, then $B_\rho(p)$ has compact closure. Therefore, we will have shown that $A = (0, \infty)$.

It is enough to show that, if $r \in A$, then there exists $\delta > 0$ such that $r + \delta \in A$. For each $q \in B_r(p)$, there exists $\rho(q)$ small enough such that the closure of $B_{\rho(q)}(q)$ is compact. We have that

$$\bar{B}_r(p) \subset \textstyle\bigcup_{q \in \bar{B}_r(x)} \bar{B}_{\rho(q)}(q)$$

By compactness of the closure of $B_r(p)$, there exist a finitely many points $q_1, \dots, q_N$ in this set such that

$$\bar{B}_r(p) \subset \bigcup_{i=1}^{N} \bar{B}_{\rho(q_i)}(q_i)$$

Moreover, since the closure of $B_{r+\delta}(p)$ coincides with the set of points $q \in M$ such that the distance of q to $B_r(p)$ is less than δ, then by Lemma 3.1, there exists $\delta > 0$ such that

$$\bar{B}_{r+\delta}(p) \subset \bigcup_{i=1}^{n} \bar{B}_{\rho(q_i)}(q_i)$$

So, this shows that $r+\delta \in A$ since the union of finitely many compact sets is also compact. Therefore, for any $r \in A$, there exists δ such that $(r-\delta, r+\delta) \subset A$, showing that A is in fact open.

To show that A is an unbounded set, assume for a contradiction that R is a finite number. This implies that the closure of any $B_r(p) \subset B_R(p)$ is compact. Therefore, every sequence contained in the closure of $B_r(p)$ is contained in some compact set and therefore has a convergent subsequence. Consequently, this means that the closure of $B_R(p)$ is compact and hence $R \in A$. However, we showed above that there exists some δ_R such that $R + \delta_R \in A$. This contradicts that R is the supremum, so $R = \infty$. Since $A = (0, \infty)$ for any p, this implies that the closure of $B_r(p)$ is compact for any $p \in M$ and any $r > 0$.

□

4 Pontryagin Extremals

4.1 The Maximal Principle

This section gives a small background on optimal control theory so that we can derive an analogy between optimal control problems and problems on sub-Riemannian manifolds. Consider a system of ordinary differential equations

$$\dot{x} = f(x, u)$$

where $x \in \mathbb{R}^n$ is our state variable, $u \in \mathbb{R}^p$ is our control parameter, and f is a continuous vector field which is differentiable over x. A certain set of values $U \subset \mathbb{R}^p$ are deemed to be admissible. A starting point $x_0 \in \mathbb{R}^n$ and ending point $x_1 \in \mathbb{R}^n$ are give. With a fixed time interval, a function $\nu : [0, T] \to U$ is an admissible control. When subjected to a cost functional, the optimal control $\nu^*(t)$ is the control such that the corresponding solution trajectory (which has endpoints x_0 and x_1) $\chi^*([0, T]) \in \mathbb{R}^n$ minimizes the cost functional. The Pontryagin maximum principle provides a way to characterize solutions to these types of problems.

Proposition 1 (Pontryagin Maximal Principle) Let $\nu^*(t)$ be an optimal control and $\chi^*(t)$ be an optimal trajectory corresponding to the above problem. Denote the vector field $p^* : [0,T] \to \mathbb{R}^n$ to be the solution to the adjoint equation:

$$\dot{p}^*(t) = \langle p^*(t), D_x f(x^*(t), u^*(t)) \rangle \text{ where } p^*(T) = \dot{\chi}^*(T)$$

Then the following holds:

$$\langle p^*(t), f(x^*(t), u^*(t)\rangle = \max_{w \in U} \langle p^*(t), f(x^*(t), w)\rangle \text{ a.e. for } t \in [0,T]$$

4.2 Characterization in the sub-Riemannian Case

We can use ideas from the above in order to characterize length-minimizers on sub-Riemannian manifolds. Consider how this compares to the language developed in section 2. Instead of some subset of $\mathbb{R}^p$ being our admissible values, we are taking a sub-bundle of the tangent bundle, namely our set of distributions. Moreover, optimization requires that we are minimizing the length of our trajectory. In the Riemannian setting, length-minimizers satisfy conditions of a system of ordinary differential equations. However, in the sub-Riemannian case, we cannot parametrize a length-minimizer by an initial velocity vector since the bundle rank of our sub-Riemannian structure will be strictly less than the dimension of the manifold. Instead, we parametrize length-minimizers by their initial point p_0 and an initial covector $\lambda_0 \in T_{p_0}M$. The following proposition gives conditions of length-minimizers that must be satisfied. A trajectory satisfying either condition is called a **Pontryagin extremal**.

Proposition 2 (Characterization of Pontryagin Extremals) Let $\gamma : [0,1] \to M$ be an admissible curve with is a length-minimizer parametrized by constant speed. Let $\nu(t)$ be the corresponding minimal control such that

$$L(\gamma) = \int_0^1 g(\nu(t), \nu(t))_{\gamma(t)} dt = d(\gamma(0), \gamma(1))$$

with $g(\nu(t), \nu(t))_{\gamma(t)}$ constant a.e. on $[0,1]$. Let $P_{0,t}$ be the flow of ν. Then there exists $\lambda_0 \in T^*_{\gamma(0)}M$ such that, defining $\lambda(t) = (P_{0,t}^{-1})^* \lambda_0$ so that $\lambda(t) \in T^*_{\gamma(t)}M$, then one of the following conditions is satisfied:

(N) In coordinates, we have $\nu_i(t) = \langle \gamma(t), f_i(\gamma(t)) \rangle$

(A) $0 = \langle \gamma(t), f_i(\gamma(t)) \rangle$

If $\lambda(t)$ satisfies condition (N), we say it is a normal extremal. On the other, hand if it satisfies (A), then it is an abnormal extremal. $\gamma(t)$ is called a normal or abnormal extremal trajectory, respectively.

It requires close examination, however this characterization of Pontryagin extremals is indeed a generalization to sub-Riemannian manifolds of the Pontryagin Maximal Principle. This this way, we can see the connection to control theory where theorems and techniques for solving problems can be utilized to solve the problem of characterizing length-minimizing curves on a sub-Riemannian manifold.

5 References

1. Andrei Agrachev, Davide Barilari, Ugo Boscain. A Comprehensive Introduction to sub-RiemannianGeometry. A Comprehensive Introduction to sub-Riemannian Geometry, In press.

2. L.S. Pontryagin, V.G. Boltayanskii, R.V. Gamkrelidze, E.F. Mishchenko, "The mathematical theory of optimal processes" , Wiley, 1962.

3. J. M. Lee. Riemannian Manifolds: An Introduction to Curvature. Springer-Verlag, 1997.

4. T. Khai Nguyen. Viscosity Solutions and Optimal Control Problems. Lecture Notes. North Carolina State University. Raleigh, North Carolina. Autumn 2019.