University of Crete
Department of Mathematics and Applied Mathematics

Notes on Sub-Riemannian Geodesics and Volume

Konstantin Athanassopoulos

Iraklion, 2019

These notes are based on a series of unofficial lectures given in the academic year 2016-17 within a small group of colleagues at the Department of Mathematics of the University of Athens.

Contents

Chapter 1

Sub-Riemannian manifolds

1.1 The isoperimetric problem in the plane

Queen Dido of Tyre had to flee across the Mediterranean in order to escape from her brother Pygmalion, also King of Tyre, who had murdered her husband Acerbas, an allegedly wealthy priest of Hercules and second in power to Pygmalion. Eventually, Dido and her company arrived at the north african cost where Dido asked the Berber King Burdas for a piece of land, only as much as could be encompassed by an oxhide. Dido cut the oxhide into a long fine strip so that she could encircle a nearby hill. In doing this Dido faced the following problem. Given a string of fixed length and a fixed line, the coast, place the ends of the string on the line and determine the shape of the curve enclosing the maximum area. This is the form of the mathematical problem which nowdays is called the isoperimetric problem. Dido found that the solution to her problem is the semicircle and the city she founded in the semicircular region was Carthage.

We shall see a mathematical formulation of Dido's problem. Consider the differential 1-form

$$a = \frac{1}{2}(-ydx + xdy)$$

which satisfies $da = dx \wedge dy$. If $L : \mathbb{R} \to \mathbb{R}^2$ is any line through the origin, say $L(t) = (tv_1, tv_2)$ for some $(v_1, v_2) \in \mathbb{R}^2 \setminus \{(0,0)\}$, then

$$(L^*a)_t(s) = a_{(tv_1,tv_2)}(sv_1, sv_2) = \frac{1}{2}(-tv_2sv_1 + tv_1sv_2) = 0$$

for every $t \in \mathbb{R}$ and $s \in \mathbb{R}$. So $L^*a = 0$.

According to Stokes' formula, the area enclosed by a piecewise smooth simple closed curve γ in $\mathbb{R}^2$ is

$$A(\gamma) = \int_\gamma a.$$

If γ is not closed and has initial point at the origin, then the integral $A(\gamma)$ represents the area enclosed by γ and the line segment from the origin to the endpoint of γ, because of the preceding property of a.

The isoperimetric problem is the following constrained variational problem: Maximize the area enclosed by a rectifiable simple closed curve subject to the constraint that the lenght of the curve is a fixed constant. Dually, minimize the length of a rectifiable simple closed curve subject to the constraint that the area enclosed by the curve is a fixed constant.

The dual isoperimetric problem can be restated as a 3-dimensional geometric problem. Let $\gamma : [0,T] \to \mathbb{R}^2$ be a piecewise smooth curve and let $\delta_c : [0,T] \to \mathbb{R}^3$, $c \in \mathbb{R}$, be the family of curves defined by

$$\delta_c(t) = (x(t), y(t), \int_{\gamma|_{[0,t]}} a + c)$$

where $\gamma(t) = (x(t), y(t))$. We define the length of δ_c to be the length of γ. Each curve δ_c is called a horizontal lift of γ and

$$\dot{\delta}_c(t) = (x'(t), y'(t), a_{\gamma(t)}(\dot{\gamma}(t))).$$

Also, $\delta_c(T) = (x(T), y(T), A(\gamma))$. Let now

$$\omega = dz - \frac{1}{2}(-ydx + xdy)$$

and $\mathcal{H} = \mathrm{Ker}\omega$. Then $\mathcal{H}$ is a vector subbundle of the tangent bundle of $\mathbb{R}^3$ with fibre

$$\mathcal{H}_{(x,y,z)} = \left\{ \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} \in T_{(x,y,z)}\mathbb{R}^3 : v_3 - \frac{1}{2}(-yv_1 + xv_2) = 0 \right\}$$

over $(x,y,z) \in \mathbb{R}^3$. The differential 1-form ω is the standard contact 1-form of $\mathbb{R}^3$ and $\mathcal{H}$ is its standard contact structure. Note that $\mathcal{H}$ is generated by the smooth vector fields

$$X = \frac{\partial}{\partial x} - \frac{1}{2}y\frac{\partial}{\partial z}, \quad Y = \frac{\partial}{\partial y} + \frac{1}{2}x\frac{\partial}{\partial z}$$

which together with $Z = \dfrac{\partial}{\partial z}$ form a basis of the corresponding tangent space of $\mathbb{R}^3$ at each point. More precisely, if $u = (u_1, u_2, u_3) \in T_{(x,y,z)}\mathbb{R}^3$, then

$$u = u_1 X + u_2 Y + (\frac{1}{2}yu_1 - \frac{1}{2}xu_2 + u_3)Z.$$

Thus, if for $u = (u_1, u_2, u_3)$, $v = (v_1, v_2, v_3) \in \mathcal{H}_{(x,y,z)}$ we set

$$\langle u, v \rangle = u_1 v_1 + u_2 v_2$$

then $\langle \cdot, \cdot \rangle$ is a Riemannian metric on $\mathcal{H}$ with respect to which $\{X, Y\}$ becomes an orthonormal frame of $\mathcal{H}$. Now the isoperimetric problem can be formulated as follows: Find the horizontal lift of γ of minimum length that joins two fixed points of $\mathbb{R}^3$.

In the terminology of contact geometry a curve δ is called Legendrian, with respect to the contact structure $\mathcal{H}$, if $\dot{\delta}(t) \in \mathcal{H}_{\delta(t)}$ for all t. The length $L(\delta)$ of a Legendrian curve δ with respect to the Riemannian metric $\langle \cdot, \cdot \rangle$ on $\mathcal{H}$ is the same with that considered before, that is the euclidean length of the projection of δ onto its first two coordinates. For p, $q \in \mathbb{R}^3$ the formula

$$d(p, q) = \inf\{L(\delta) : \delta \text{ is Legendrian from } p \text{ to } q\}$$

defines a new distance on $\mathbb{R}^3$, the contact distance, because there exist Legendrian curves from p to q. Indeed, if $p = (0,0,0)$ and $q = (q_1, q_2, q_3)$, let $\gamma(t) = (x(t), y(t))$, $0 \le t \le T$, be a curve in $\mathbb{R}^2$ from $(0,0)$ to (q_1, q_2) such that $q_3 = \displaystyle\int_\gamma a$, as in Dido's problem. The lifted curve

$$\delta(t) = (x(t), y(t), \int_{\gamma|_{[0,t]}} a), \quad 0 \le t \le T$$

is then a Legendrian curve from p to q. Also, since the Riemannian length of δ is equal to the euclidean length of γ, there is a correspondence between the geodesics of d and the solutions of the dual isoperimetric problem.

The metric space $(\mathbb{R}^3, d)$ is isometrically homogeneous. This can be seen by introducing a group structure on $\mathbb{R}^3$, different from the usual one, such that the left translations are d-isometries. The group law defined by

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

makes $(\mathbb{R}^3, \cdot)$ a non-abelian Lie group. The left translation by (x, y, z) has Jacobian matrix at each point

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -\frac{1}{2}y & \frac{1}{2}x & 1 \end{pmatrix}$$

and $AX_{(0,0,0)} = X_{(x,y,z)}$, $AY_{(0,0,0)} = Y_{(x,y,z)}$ and $AZ_{(0,0,0)} = Z_{(x,y,z)}$. Hence $\mathcal{H}$ is invariant under left translations and since each left translation leaves the orthonormal frame $\{X, Y\}$ invariant, it is a d-isometry. The group $(\mathbb{R}^3, \cdot)$ is called the Heisenberg group and has also a matrix model as a subgroup of $GL(3, \mathbb{R})$.

Let

$$G = \left\{ \begin{pmatrix} 1 & a & c \\ 0 & 1 & b \\ 0 & 0 & 1 \end{pmatrix} : a, b, c \in \mathbb{R} \right\}.$$

This is a closed subgroup of $GL(3, \mathbb{R})$ with Lie algebra

$$\mathfrak{g} = \left\{ \begin{pmatrix} 0 & a & c \\ 0 & 0 & b \\ 0 & 0 & 0 \end{pmatrix} : a, b, c \in \mathbb{R} \right\}$$

which has basis consisting of

$$\tilde{X} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \tilde{Y} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}, \quad \tilde{Z} = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

It is easy to see that the map $\phi : \mathbb{R}^3 \to G$ defined by

$$\phi(x, y, z) = \begin{pmatrix} 1 & x & z + \frac{1}{2}xy \\ 0 & 1 & y \\ 0 & 0 & 1 \end{pmatrix}$$

is a Lie group isomorphism from $(\mathbb{R}^3, \cdot)$ onto G. Moreover, $\phi_{*(0,0,0)}(\frac{\partial}{\partial x}) = \tilde{X}$, $\phi_{*(0,0,0)}(\frac{\partial}{\partial y}) = \tilde{Y}$ and $\phi_{*(0,0,0)}(\frac{\partial}{\partial z}) = \tilde{Z}$. Note that $[\tilde{X}, \tilde{Y}] = \tilde{X}\tilde{Y} - \tilde{Y}\tilde{X} = \tilde{Z} - 0 = \tilde{Z}$ and $[\tilde{X}, \tilde{Z}] = [\tilde{Y}, \tilde{Z}] = 0$. This implies that the Heisenberg group is nilpotent of step 2.

1.2 Sub-Riemannian structures

The most general definition of a Sub-Riemannian structure on a smooth manifold M is the following.

Definition 1.2.1. A Sub-Riemannian structure on M is a triple consisting of
(i) a smooth real vector bundle $p : E \to M$ over M,
(ii) a Riemannian metric g on this vector bundle fiberwise and
(iii) a smooth vector bundle morphism $f : E \to TM$ into the tangent bundle TM of M.

$$\begin{tikzcd} E \arrow[rr, "f"] \arrow[dr, "p"'] & & TM \arrow[dl] \\ & M & \end{tikzcd}$$

A Sub-Riemannian manifold is a smooth manifold endowed with a Sub-Riemannian structure. Its horizontal distribution is the family of linear spaces $D_x = f(E_x)$, $x \in M$, where $E_x = p^{-1}(x)$ is the fibre over the point x. The non-negative integer $k(x) = \dim D_x$ is called the rank of the Sub-Riemannian structure at $x \in M$. If k is constant, the Sub-Riemannian structure is said to be of constant rank. The Sub-Riemannian structure is a classical Riemannian structure on M in case the vector bundle map f is surjective.

A smooth local vector field X defined on an open set $U \subset M$ is called horizontal if $X(x) \in D_x$ for every $x \in M$. The set of all horizontal vector fields on U has the structure of a finitely generated $C^\infty(U)$-module.

If $f_x : E_x \to T_xM$ denotes the restriction of f to the fibre E_x, the chain of obvious isomorphisms $(\mathrm{Ker} f_x)^\perp \cong E_x/\mathrm{Ker} f_x \cong D_x$ permits us to define an inner product on D_x which varies smoothly with respect to x. The corresponding norm of a vector $v \in D_x$ is given by the formula

$$\|v\| = \min\{g(u,u)^{1/2} : u \in E_x \text{ and } f_x(u) = v\}.$$

Two Sub-Riemannian structures on the same smooth manifold M consisting of smooth real vector bundles $p_i : E_i \to M$ with Riemannian metrics g_i and smooth vector bundle morphisms $f_i : E_i \to TM$, $i = 1, 2$, are called equivalent if there exist a smooth real vector bundle $p : E \to M$ endowed with a Riemannian metric g and surjective smooth vector bundle morphisms $q_i : E \to E_i$, which are compatible with the Riemannian metrics on E_i, $i = 1, 2$, such that $f_1 \circ q_1 = f_2 \circ q_2$, that is the following diagram commutes.

$$\begin{tikzcd} & E_1 \arrow[dr, "f_1"] & \\ E \arrow[ur, "q_1"] \arrow[dr, "q_2"'] & & TM \\ & E_2 \arrow[ur, "f_2"'] & \end{tikzcd}$$

The compatibility condition of the Riemannian metrics implies that

$$g_i(v_i, v_i)^{1/2} = \min\{g(u,u)^{1/2} : u \in E \text{ and } q_i(u) = v_i\}$$

for every $v_i \in E_i$, $i = 1, 2$. An immediate consequence is that the corresponding distributions coincide as well as the corresponding induced norms. It is rather obvious that equivalence of Sub-Riemannian structures on a given smooth manifold M is an equivalence relation. For the transitivity, it suffices from the above commutative diagram and a similar commutative diagram

$$\begin{tikzcd} & E_2 \arrow[dr, "f_2"] & \\ E' \arrow[ur, "q_2'"] \arrow[dr, "q_3'"'] & & TM \\ & E_3 \arrow[ur, "f_3"'] & \end{tikzcd}$$

to consider the obvious commutative diagram

$$\begin{tikzcd} & E_1 \arrow[dr, "f_1"] & \\ E \oplus E' \arrow[ur] \arrow[dr] & & TM \\ & E_3 \arrow[ur, "f_3"'] & \end{tikzcd}$$

and take the product Riemannian metric on $E \oplus E'$.

Examples 1.2.2. (a) Classically, a constant rank Sub-Riemannian structure on a smooth manifold M is defined by a smooth vector subbundle of the tangent bundle of M endowed with a Riemannian metric. This case is included in Definition 1.2.1 by taking the vector bundle morphism f to be the inclusion.

(b) If the vector bundle morphism f in Definition 1.2.1 is surjective, we have a classical Riemannian structure on M. In this case the commutative diagram

$$\begin{tikzcd} & E \arrow[dr, "f"] & \\ E \oplus TM \arrow[ur, "\text{projection}"] \arrow[dr, "f\oplus 0"'] & & TM \\ & TM \arrow[ur, "\mathrm{id}"'] & \end{tikzcd}$$

says that the Sub-Riemannian structure is equivalent to a Riemannian structure in the classical sense. On $E \oplus TM$ we consider the product Riemannian metric, where on TM we take the Riemannian metric induced by f.

(c) Let G be a Lie group and let V be a m-dimensional linear subspace of its Lie algebra $\mathfrak{g}$. Then V endowed with any inner product defines a Sub-Riemannian structure on G. In this case, the vector bundle $p : E \to G$ is the trivial vector bundle over G with total space $E = G \times \mathbb{R}^m$. If $\{X_1, ..., X_m\}$ is an orthonormal basis of V, the vector bundle morphism $f : G \times \mathbb{R}^m \to TG$ is defined by

$$f(g, u_1, ..., u_m) = \sum_{k=1}^{m} u_k X_k(g).$$

The action of G onto itself by left translations is isometric.

As a special case let G be a graded nilpotent Lie group. This means that

$$\mathfrak{g} = V_1 \oplus V_2 \oplus \cdots \oplus V_r$$

where $[V_i, V_j] = V_{i+j}$ and $V_i = 0$ for $i > r$. Hence G is nilpotent of step r. Taking $V = V_1$, we get a Sub-Riemannian structure as above. Such a Lie group G is called a Carnot group.

The Heisenberg group $\mathcal{H}$ we defined in section 1.1 is the simplest non-euclidean example of a Carnot group. It has step 2. Indeed, there are only two 3-dimensional, simply connected, nilpotent Lie groups, namely the euclidean space $\mathbb{R}^3$ and the Heisenberg group $\mathcal{H}$. To see this, let $\mathfrak{g}$ be the Lie algebra of such a Lie group G. Since G is nilpotent, $\mathfrak{g}$ has non-trivial center. Let $Z \neq 0$ be an element of the center of $\mathfrak{g}$. We complete Z to a basis $\{X, Y, Z\}$ of $\mathfrak{g}$. If $[X, Y] = 0$, then $\mathfrak{g}$ is abelian and $G \cong \mathbb{R}^3$. If $[X, Y] \neq 0$, there are a, b, $c \in \mathbb{R}$, not all three zero, such that $[X, Y] = aX + bY + cZ$. Now $[[X, Y], Y] = a[X, Y]$ and therefore $a = 0$ since G is nilpotent. Similarly $b = 0$. Thus, necessarily $c \neq 0$ and replacing Z with cZ we get a basis of $\mathfrak{g}$ consisting of elements X, Y, Z such that $[X, Y] = Z$ and $[X, Z] = [Y, Z] = 0$. Recalling that a fixed Lie algebra integrates to a unique simply connected Lie group, we conclude that $G \cong \mathcal{H}$.

(d) A higher dimensional analogue of the Heisenberg group can be defined as follows. Let (V, ω) be a $2n$-dimensional real symplectic vector space and let $\mathfrak{g} = V \oplus \mathbb{R}$, On $\mathfrak{g}$ we define a Lie bracket by

$$[(v, t), (u, s)] = (0, \omega(v, u)].$$

This makes $\mathfrak{g}$ a 2-step nilpotent graded Lie algebra. Let $\{X_1, Y_1, ..., X_n, Y_n\}$ be a symplectic basis of V. If $Z = (0, 1)$, then $\{X_1, Y_1, ..., X_n, Y_n, Z\}$ is a basis of $\mathfrak{g}$ with $[X_i, Y_j] = Z$, $i \neq j$, as the only non-trivial brackets. Now $\mathfrak{g}$ integrates to a unique simply connected Lie group of dimension $2n + 1$, This is the $(2n + 1)$-dimensional Heisenberg group.

(e) Let M be a smooth manifold and ξ be a smooth vector subbundle of its tangent bundle TM of codimension 1. Then, locally ξ is the kernel of a differential

1-form a. Indeed, if we choose any Riemannian metric g on M, then $TM \cong \xi \oplus \xi^{\perp}$. Locally, the line bundle $\xi^{\perp}$ is trivial and generated by a non-zero vector field X. If $a = g(.,X)$, then $\xi = \mathrm{Ker}a$. If $TM/\xi \cong \xi^{\perp}$ is orientable, then X is globally defined, because $\xi^{\perp}$ is globally trivial and $\xi = \mathrm{Ker}a$ throughout M.

A contact structure on a $(2n+1)$-dimensional smooth manifold M is a smooth vector subbundle ξ of TM of rank $2n$ which is the kernel of a maximally non-integrable differential 1-form a on M, meaning that $a \wedge (da)^n \neq 0$ all over M. If we put a Riemannian metric on ξ, we get a Sub-Riemannian structure, which is said to be of contact type. A contact manifold is an odd dimensional smooth manifold endowed with a contact structure. A concrete example is the Heisenberg group. Another important class of examples of contact manifolds are the pre-quantizations of symplectic manifolds. In such a case the contact manifold is the total space of a circle bundle over the given symplectic manifold.

(f) Let M, N be Riemannian manifolds and let $p : M \to N$ be a Riemannian submersion. This means that the tangent bundle of M splits as a Whitney direct sum $TM = \mathrm{Ker}p_* \oplus H$, where $H = (\mathrm{Ker}p_*)^{\perp}$ and $p_{*x}|_{H_x} : H_x \to T_{p(x)}N$ is a linear isometry of inner product vector spaces for every $x \in M$. The smooth vector subbundle H of TM defines a Sub-Riemannian structure of constant rank on M in the obvious way. This is usually called a Sub-Riemannian structure of bundle type. The most well known example of Riemannian submersion is the Hopf map $p : S^{2n+1} \to \mathbb{C}P^n$, if on $\mathbb{C}P^n$ we consider the Fubini-Study metric. Actually, this is a convenient way to define the Fubini-Study metric.

As a special case, let G be a Lie group and $p : Q \to N$ be a smooth principal G-bundle. Then p is a Riemannian submersion if and only if G acts on Q by isometries of a Riemannian metric, in which case we have an induced Riemannian metric on N making p Riemannian submersion. The corresponding horizontal distribution H is that of a connection on the bundle. This is the case in the example of the Hopf map.

A Sub-Riemannian structure on a smooth manifold M as defined in Definition 1.2.1 is called free if the vector bundle $p : E \to M$ is trivial, that is $E = M \times \mathbb{R}^m$ for some $m \in \mathbb{N}$ and p is the projection, and the Riemannian metric on E is the usual euclidean metric on $\mathbb{R}^m$. In this case, if we define $X_k = f(., e_k)$, $1 \leq k \leq m$, then the horizontal distribution $D = \{D_x : x \in M\}$ is globally generated by the set of smooth vector fields $\{X_1, X_2, ..., X_m\}$ of M. This means that for every vector $u \in D_x$ there exist u_1, u_2,..., $u_m \in \mathbb{R}$ such that

$$u = \sum_{k=1}^{m} u_k X_k(x).$$

This expansion of u may not be unique because the vectors $X_1(x)$, $X_2(x)$,..., $X_k(x)$ may be linearly dependent. The Sub-Riemannian structure described in Example 1.2.2(c) is free.

Proposition 1.2.3. *Every Sub-Riemannian structure on a smooth manifold M is equivalent to a free one.*

Proof. Suppose that we are given a Sub-Riemannian structure on M as in Definition 1.2.1. There exists a smooth vector bundle $\tilde{p} : \tilde{E} \to M$ whose Whitney direct sum with $p : E \to M$ is a trivial vector bundle over M. So there exists some $m \in \mathbb{N}$ such that $E \oplus \tilde{E} = M \times \mathbb{R}^m$. On $\tilde{E}$ we choose any fiberwise Riemannian metric and on $E \oplus \tilde{E}$ we consider the product metric, so that E and $\tilde{E}$ become orthogonal vector bundles. Let $p_1 : E \oplus \tilde{E} \to E$ denote the projection. On M we have now a free Sub-Riemannian structure consisting of the trivial smooth vector bundle $E \oplus \tilde{E} = M \times \mathbb{R}^m \to M$ endowed with the above product metric and the smooth vector bundle morphism $f \circ p_1 : E \oplus \tilde{E} \to TM$. The commutative diagram

$$\begin{tikzcd}
 & E \oplus \tilde{E} \arrow[dr, "f \circ p_1"] & \\
E \oplus \tilde{E} \arrow[ur, "\mathrm{id}"] \arrow[dr, "p_1"'] & & TM \\
 & E \arrow[ur, "f"'] &
\end{tikzcd}$$

says that the initial Sub-Riemannian structure is equivalent to the constructed free Sub-Riemannian structure, because the involved surjective vector bundle morphisms $\mathrm{id} : E \oplus \tilde{E} \to E \oplus \tilde{E}$ and $p_1 : E \oplus \tilde{E} \to E$ are trivially compatible with the Riemannian metrics. □

It follows from Proposition 1.2.3 that for every Sub-Riemannian structure on a smooth manifold M there exists a sufficiently large number of globally defined smooth vector fields on M which generate the corresponding horizontal distribution.

1.3 Horizontal curves

Let M be a Sub-Riemannian manifold whose Sub-Riemannian structure consists of the smooth real vector bundle $p : E \to M$ endowed with the Riemannian metric g and the smooth vector bundle morphism $f : E \to TM$ with corresponding horizontal distribution $D = \{D_x : x \in M\}$. An absolutely continuous curve $\gamma : I \to M$, where $I \subset \mathbb{R}$ is an interval, is called horizontal (or admissible) if there exists a measurable map $u : I \to E$ such that $u(t) \in E_{\gamma(t)}$ and $\dot{\gamma}(t) = f(\gamma(t), u(t)) \in D_{\gamma(t)}$ a.e. on I. Such a measurable map u is called a control function for γ and may not be unique. By Proposition 1.2.3, the Sub-Riemannian structure is equivalent to a free one and there exist globally defined horizontal smooth vector fields X_1, X_2,..., X_m, for some large enough $m \in \mathbb{N}$, which generate D. As the proof of Proposition 1.2.3 shows, the control function u for γ lifts to a measurable function $(u_1, u_2, ..., u_m) : I \to \mathbb{R}^m$ whose pointwise norm is equal to $|u| = (g(u,u))^{1/2}$ and such that

$$\dot{\gamma}(t) = \sum_{k=1}^{m} u_k(t) X_k(\gamma(t))$$

for almost all $t \in I$.

Example 1.3.1. On $\mathbb{R}^2$ we consider the free Sub-Riemannian structure defined by the vector bundle morphism $f : \mathbb{R}^2 \times \mathbb{R}^2 \to T\mathbb{R}^2 \cong \mathbb{R}^2 \times \mathbb{R}^2$ with

$$f((x,y),(v_1,v_2)) = ((x,y),(v_1,x^2v_2)).$$

The rank of this Sub-Riemannian structure is 2 on $\mathbb{R}^2 \setminus \{(0,y) : y \in \mathbb{R}\}$ and 1 at the points on the line $\{(0,y) : y \in \mathbb{R}\}$. Its horizontal distribution is generated by the smooth vector fields

$$X = \frac{\partial}{\partial x} \quad \text{and} \quad Y = x^2\frac{\partial}{\partial y}.$$

Let $\gamma : [-1,1] \to \mathbb{R}^2$ be the smooth curve $\gamma(t) = (t,t^2)$ whose velocity is

$$\dot\gamma(t) = X(\gamma(t)) + \frac{2}{t}Y(\gamma(t))$$

for $t \neq 0$ and $\dot\gamma(0) = X(\gamma(0))$. So, we have a control function $u = (u_1,u_2)$ where $u_1(t) = 1$ for every $-1 \leq t \leq 1$ and $u_2 : [-1,1] \to \mathbb{R}$ is defined by

$$u_2(t) = \begin{cases} \frac{2}{t}, & \text{for } t \in [-1,1] \setminus \{0\}, \\ 0, & \text{for } t \neq 0. \end{cases}$$

Note that although u_2 is measurable, it is not continuous, not even L^1. On the trivial vector bundle $\mathbb{R}^2 \times \mathbb{R}^2$ we consider the euclidean metric. Then, for the induced norm on the horizontal distribution we have $\|\dot\gamma(t)\|^2 = (u_1(t))^2 + (u_2(t))^2$. This leads to the notion of minimal control.

If $\gamma : I \to M$ is a horizontal curve on the Sub-Riemannian manifold M, then for almost every $t \in I$ there exists a unique element $u^*(t) \in (\mathrm{Ker} f_{\gamma(t)})^\perp \leq E_{\gamma(t)}$ such that $\|\dot\gamma(t)\| = (g(u^*(t),u^*(t)))^{1/2} = |u^*(t)|$. This almost everywhere defined control function u^* is called the minimal control associated to γ and may not be continuous even in case γ is smooth, as the previous Example 1.3.1 shows. The fact that u^* is indeed measurable is not so obvious and follows from the following.

Lemma 1.3.2. *Let $I \subset \mathbb{R}$ be an interval and let $K \subset \mathbb{R}^m$ be a compact set. Let $h : I \times K \to \mathbb{R}^m$ and $u : I \to \mathbb{R}^m$ be two functions with the following properties:*

(P1) *$h(.,x) : I \to \mathbb{R}^m$ is measurable for every $x \in K$ and $h(t,.) : K \to \mathbb{R}^m$ is continuous for every $t \in I$.*

(P2) *u is measurable.*

(P3) *For every $t \in I$ there exists a unique $u^*(t) \in \mathbb{R}^m$ such that*

$$\|u^*(t)\| = \min\{\|x\| : h(t,x) = u(t), \quad x \in K\}.$$

Then $u^* : I \to \mathbb{R}^m$ is measurable.

Proof. As a first step we show that the function $\phi = \|u^*\| : I \to \mathbb{R}$ is measurable. For this it suffices to prove that the set $\phi^{-1}([0,c])$ is measurable for every $c > 0$. The definition of u^* implies that

$$\phi^{-1}([0,c]) = \{t \in I : h(t,x) = u(t) \text{ and } \|x\| \leq c \text{ for some } x \in K\}.$$

If A is a countable dense subset of K and

$$C_{a,l} = \{t \in I : \|h(t,a) - u(t)\| < \frac{1}{l}\}, \quad a \in A, l \in \mathbb{N},$$

it suffices to prove that $\phi^{-1}([0,c]) = \bigcap_{l=1}^{\infty} \bigcup_{a \in A \cap \overline{B(0,c)}} C_{a,l}$.

Indeed, if $t \in \phi^{-1}([0,c])$, there exists some $x \in K$ such that $h(t,x) = u(t)$ and $\|x\| \leq c$. Since $h(t,.)$ is continuous, for every $l \in \mathbb{N}$ there exists $a \in A$ such that $\|a\| \leq c$ and $\|h(t,a) - u(t)\| < \frac{1}{l}$.

Conversely, let $t \in \bigcap_{l=1}^{\infty} \bigcup_{a \in A \cap \overline{B(0,c)}} C_{a,l}$, that is for every $l \in \mathbb{N}$ there exists $a_l \in A$ such that $\|a_l\| \leq c$ and $\|h(t,a_l) - u(t)\| < \frac{1}{l}$. Since $K \cap \overline{B(0,c)}$ is compact, the sequence $(a_l)_{l \in \mathbb{N}}$ has some limit point $x \in K \cap \overline{B(0,c)}$. The continuity of $h(t,.)$ implies that $\|h(t,x) - u(t)\| = 0$.

We proceed now to show that u^* is measurable. Let $F \subset K$ be a closed set. Arguing as above, we consider the sets

$$G_{a,l} = \{t \in I : \|h(t,a) - u(t)\| < \frac{1}{l} \text{ and } \|a\| < \|u^*(t)\| + \frac{1}{l}\}, \quad a \in A, l \in \mathbb{N}.$$

Since $\|u^*\|$ is measurable, each $G_{a,l}$ is measurable and so it suffices to prove that

$$(u^*)^{-1}(F) = \bigcap_{l=1}^{\infty} \bigcup_{a \in A} G_{a,l}.$$

Indeed, if $u^*(t) \in F$, then the definition of u^* and the continuity of $h(t,.)$ imply that for every $l \in \mathbb{N}$ there exists some $a \in A$ such that $0 \leq \|a\| - \|u^*(t)\| < \frac{1}{l}$ and $\|h(t,a) - u(t)\| < \frac{1}{l}$.

Conversely, if $t \in \bigcap_{l=1}^{\infty} \bigcup_{a \in A} G_{a,l}$, then for every $l \in \mathbb{N}$ there exists $a_l \in A$ such that $\|a_l\| < \|u^*(t)\| + \frac{1}{l}$ and $\|h(t,a_l) - u(t)\| < \frac{1}{l}$. By compactness, the sequence $(a_l)_{l \in \mathbb{N}}$ has some limit point $x \in F$ and then $\|x\| \leq \|u^*(t)\|$ and $\|h(t,x) - u(t)\| = 0$. By the definition of $u^*(t)$ we must necessarily have $u^*(t) = x \in F$. This concludes the proof. □

If now u^* is the minimal control of a horizontal curve $\gamma : I \to M$ we apply Lemma 1.3.2 setting $h(t,x) = f(\gamma(t),x)$. The assumptions of the lemma are obviously satisfied and thus u^* is measurable. If in addition there exists a control u for γ such that $|u| \in L^1$, then also $|u^*| \in L^1$.

Definition 1.3.3. Let a, $b \in \mathbb{R}$, with $a < b$ and let $\gamma : [a,b] \to M$ be a horizontal curve with L^1 controls. The Sub-Riemannian length of γ is

$$L(\gamma) = \int_a^b \|\dot{\gamma}(t)\| dt = \int_a^b |u^*(t)| dt$$

where u^* is the minimal control associated to γ.

The Sub-Riemannian length of horizontal curves with L^1 controls remains invariant under strictly monotone absolutely continuous reparametrizations. Let c, $d \in \mathbb{R}$ with $c < d$ and let $\phi : [c, d] \to [a, b]$ be increasing, absolutely continuous and surjective. We assume further that $\gamma \circ \phi$ is absolutely continuous, which is true in case ϕ is strictly increasing. If u^* is the minimal control associated with γ, from the chain rule follows that the minimal control associated to the horizontal curve $\gamma \circ \phi$ is $\phi'(u^* \circ \phi)$. Since $|u^*| \in L^1$, from the change of variables formula (for Lebesgue integrals) we have $\phi'|u^* \circ \phi| \in L^1$ and

$$L(\gamma \circ \phi) = \int_c^d |u^*(\phi(t))|\phi'(t)dt = \int_a^b |u^*(t)|dt = L(\gamma).$$

A horizontal curve $\gamma : [0, T] \to M$, $T > 0$, is said to be parametrized by arclength if $\|\dot{\gamma}(t)\| = 1$ for almost all $t \in [0, T]$. A horizontal curve γ parametrized by arclength has automatically L^∞ controls and $L(\gamma) = T$.

Proposition 1.3.4. *Every horizontal curve $\gamma : [0, T] \to M$, $T > 0$, with L^1 controls and with positive length is an absolutely continuous, monotone reparametrization of a horizontal curve which is parametrized by arc length.*

Proof. Let u^* be the minimal control associated with γ and let $\phi : [0, T] \to [0, L(\gamma)]$ be the length function of γ defined by

$$\phi(t) = \int_0^t |u^*(s)|ds$$

which is increasing and absolutely continuous, because $|u^*| \in L^1$. If $\phi(t_1) = \phi(t_2)$ and $t_1 < t_2$, then $\dot{\gamma} = 0$ a.e. on $[t_1, t_2]$ and so γ is constant on $[t_1, t_2]$. Thus, there exists a well defined map $\delta : [0, L(\gamma)] \to M$ such that $\gamma = \delta \circ \phi$.

$$\begin{tikzcd} {[0,T]} \arrow[r, "\gamma"] \arrow[d, "\phi"'] & M \\ {[0,L(\gamma)]} \arrow[ur, dashed, "\delta"'] & \end{tikzcd}$$

In order to show that δ is absolutely continuous, let U be the domain of a chart on M contained in a lager compact set over which $p : E \to M$ is trivial. Then, by equivalence of norms in finite dimensional vector spaces and continuity, there exists a constant $c > 0$ such that

$$\|\gamma(t_2) - \gamma(t_1)\| \leq c \int_{t_1}^{t_2} |u^*(s)|ds$$

for $0 \leq t_1 < t_2 \leq T$ with $\gamma([t_1, t_2]) \subset U$. In the left hand side of the above inequality by abuse of notation we have denoted again with γ the local presentation of the curve with respect to the chart on U and by $\|.\|$ the euclidean norm. From this we get

$$\|\delta(\phi(t_2)) - \delta(\phi(t_1))\| \leq c|\phi(t_2) - \phi(t_1)|.$$

This shows that δ is Lipschitz and hence absolutely continuous.

Finally, in order to prove that δ is horizontal and parametrized by arclength, we need to find a control for δ of norm equal to 1 pointwise. For this it suffices to define

$$v(s) = \frac{1}{|u^*(t)|} u^*(t), \quad \text{for } s = \phi(t).$$

This is well defined a.e. as the following Lemma 1.3.5 shows. From the chain rule we have

$$f(\gamma(t), u^*(t)) = \dot{\gamma}(t) = \phi'(t)\dot{\delta}(s)$$

and from fiberwise linearity $f(\delta(s), v(s)) = \dot{\delta}(s)$, since $\phi' = |u^*|$ a.e. $\square$

In the proof of Proposition 1.3.4 we have used the following measure theoretic fact which is a consequence of Vitali's covering theorem.

Lemma 1.3.5. *Let $\phi : [0,1] \to \mathbb{R}$ be an increasing absolutely continuous function. If $A = \{t \in [0,1] : \phi'(t) \text{ exists and } \phi'(t) = 0\}$, then $\phi(A)$ has Lebesgue measure zero.*

Proof. Let $\epsilon > 0$ and let $\delta > 0$ correspond to ϵ as in the definition of absolute continuity for ϕ. For every $t \in A$ there exists $\delta_t > 0$ such that $(t - \delta_t, t + \delta_t) \subset (0, 1)$ and

$$\frac{|\phi(t+h) - \phi(t)|}{|h|} < \epsilon$$

whenever $0 < |h| < \delta_t$.

The family $\mathcal{U} = \{[t, t+h] : t \in A \text{ and } 0 < h < \delta_t\}$ is a Vitali cover. By Vitali's covering theorem, there exists a countable subfamily of $\mathcal{U}$ consisting of pairwise disjoint intervals $I_k = [t_k, t_k + h_k]$, $k \in \mathbb{N}$, such that

$$\lambda\Big(A \cap \Big(\bigcup_{k=1}^{\infty} I_k\Big)^c\Big) = 0$$

where λ denotes the Lebesgue measure. Thus, there exists a countable family of closed intervals J_k, $k \in \mathbb{N}$, such that

$$A \cap \Big(\bigcup_{k=1}^{\infty} I_k\Big)^c \subset \bigcup_{k=1}^{\infty} J_k \quad \text{and} \quad \sum_{k=1}^{\infty} \lambda(J_k) < \delta.$$

Now

$$\phi(A) \subset \bigcup_{k=1}^{\infty} \phi(I_k) \cup \bigcup_{k=1}^{\infty} \phi(J_k)$$

and $\phi(I_k)$, $\phi(J_k)$, $k \in \mathbb{N}$, are intervals. Since ϕ is assumed to be increasing, we have

$$\sum_{k=1}^{\infty} \lambda(\phi(I_k)) < \epsilon \sum_{k=1}^{\infty} h_k \leq \epsilon.$$

On the other hand, let $F_m = \bigcup_{k=1}^{m} J_k$ and let $F_{m,1}$,..., $F_{m,l}$ be the connected components of F_m. Since $\sum_{j=1}^{l} \lambda(F_{m,j}) < \delta$, the absolute continuity of ϕ gives

$$\lambda(\phi(F_m)) = \lambda\left(\bigcup_{j=1}^{m} \phi(F_{m,j})\right) \leq \sum_{j=1}^{m} \lambda(\phi(F_{m,j})) < \epsilon.$$

This holds for every $m \in \mathbb{N}$, which means that

$$\sum_{k=1}^{\infty} \lambda(\phi(J_k)) = \sup\{\lambda(\phi(F_m)) : m \in \mathbb{N}\} \leq \epsilon.$$

Therefore,

$$\sum_{k=1}^{\infty} \lambda(\phi(I_k)) + \sum_{k=1}^{\infty} \lambda(\phi(J_k)) \leq 2\epsilon.$$

This proves that $\phi(A)$ is a set of Lebesgue measure zero. $\square$

Chapter 2

Sub-Riemannian manifolds as length metric spaces

2.1 The Sub-Riemannian distance

As in the case of a Riemannian manifold, if M is a connected Sub-Riemannian manifold, we can define the "distance" of two points x, $y \in M$ in a similar way. More precisely, the Sub-Riemannian distance (also known as Carnot-Caratheodory distance) of x and y is defined to be

$$d(x,y) = \inf\{L(\gamma) : \gamma \text{ is a horizontal curve in } M \text{ with } L^1 \text{ controls from } x \text{ to } y\}.$$

It is obvious that d is symmetric and satisfies the triangle inequality. Also $d(x,x) = 0$ for every $x \in M$. However, $d(x,y)$ may be infinite, because there may not exist any horizontal curve from x to y. For example, if $E \subset TM$ is (the total space of) an integrable subbundle of the tangent bundle of M, then there exists a horizontal curve in M from x to y if and only if x and y belong to the same leaf of the foliation to which E integrates. So we have to answer the following:

Question 2.1.1. *Under what sufficient conditions d is a distance function? In particular, under what conditions on the Sub-Riemannian structure of M any two points x, $y \in M$ can be connected with a horizontal curve?*

In the Riemannian case, the topology induced by d coincides with the original manifold topology.

Question 2.1.2. *If d is a distance, does it induce the original manifold topology?*

This chapter is devoted to giving satisfactory answers to these two questions. Recall however that in Riemannian Geometry the distance can be realized by minimizing geodesics.

Definition 2.1.3. A horizontal curve $\gamma : [0,T] \to M$, $T > 0$, with L^1 controls is called a length minimizer if $L(\gamma) = d(\gamma(0), \gamma(T))$.

In Riemannian Geometry geodesics are smooth curves.

Question 2.1.4. *Are all Sub-Riemannian length minimizers smooth?*

This question will be answered in chapter 3, which is devoted to the study of Sub-Riemannian geodesics. As in Riemannian Geometry, it is sometimes more convenient to work with the energy of curves instead of the length.

Definition 2.1.5. The energy of a horizontal curve $\gamma : [0, T] \to M$, $T > 0$, with L^2 controls is

$$J(\gamma) = \frac{1}{2}\int_0^T \|\dot{\gamma}(t)\|^2 dt$$

and is not invariant under reparametrizations.

Proposition 2.1.6. *For any x, $y \in M$ let*

$$e(x, y) = \inf\{2J(\gamma) : \gamma : [0, 1] \to M \text{ is horizontal with } L^2 \text{ controls from } x \text{ to } y\}.$$

Then, $(d(x, y))^2 = e(x, y)$. Moreover, a horizontal curve $\gamma : [0, 1] \to M$ with L^2 controls from x to y is an energy minimizer if and only if it is a length minimizer and $\|\dot{\gamma}\|$ is constant a.e.

Proof. If there is no horizontal curve with L^1 controls from x to y, both $d(x, y)$ and $e(x, y)$ are equal to infinity. So we assume that there are horizontal curves with L^2 controls from x to y. From the Cauchy-Schwarz inequality we have

$$(L(\gamma))^2 = \left(\int_0^1 \|\dot{\gamma}(t)\|dt\right)^2 \leq \int_0^1 \|\dot{\gamma}(t)\|^2 dt = 2J(\gamma)$$

and the equality holds if and only if $\|\dot{\gamma}\|$ is constant a.e. Hence $(d(x, y))^2 \leq e(x, y)$. To prove the reverse inequality, for every $\epsilon > 0$ there exists a horizontal curve $\gamma : [0, 1] \to M$ with L^1 controls from x to y such that $L(\gamma) \leq d(x, y) + \epsilon$. By Proposition 1.3.4, γ is the absolutely continuous, increasing reparametrization of a horizontal curve $\delta : [0, 1] \to M$ from x to y with $\|\dot{\delta}\| = L(\gamma)$ a.e. Consequently,

$$2J(\delta) = (L(\gamma))^2 \leq (d(x, y) + \epsilon)^2.$$

This shows that $e(x, y) \leq (d(x, y)+\epsilon)^2$ for every $\epsilon > 0$ and hence $e(x, y) = (d(x, y))^2$. The second assertion is obvious from the above. □

2.2 The Lie bracket generating condition

Let $\mathcal{A}$ be a real Lie algebra and let $F \subset \mathcal{A}$. The Lie subalgebra of $\mathcal{A}$ generated by F is the smallest Lie subalgebra of $\mathcal{A}$ which contains F and will be denoted by $Lie(F)$. We shall give a description of $Lie(F)$. Let $F^1 =< F >$, where $< \cdot >$ denotes span as a vector space. We define inductively

$$F^{j+1} = F^j + < [F^1, F^j] >, \quad j \in \mathbb{N},$$

where $[F^1, F^j] = \{X, Y : X \in F^1, Y \in F^j\}$. Obviously,

$$F^1 \subset F^2 \subset \cdots \subset F^j \subset F^{j+1} \subset \cdots \subset \mathcal{A}.$$

Lemma 2.2.1. *$[F^i, F^j] \subset F^{i+j}$ for every i, $j \in \mathbb{N}$.*

Proof. This holds by definition for $i = 1$ and $j \in \mathbb{N}$. For $i = 2$ and all $j \in \mathbb{N}$ we have (suppressing the spans $< \cdot >$)

$$[F^2, F^j] = [F^1 + [F^1, F^1], F^j] = [F^1, F^j] + [[F^1, F^1], F^j]$$

$$\subset F^{j+1} + [[F^1, F^j], F^1] + [[F^j, F^1], F^1] = F^{j+1} + [[F^1, F^j], F^1]$$

$$\subset F^{j+1} + [F^{j+1}, F^1] = F^{j+2}.$$

For the proof of the general formula we use induction with repect to i. Suppose that $[F^{i-1}, F^j] \subset F^{i-1+j}$ for all $j \in \mathbb{N}$. Then,

$$[F^i, F^j] = [F^{i-1}, F^j] + [[F^1, F^{i-1}], F^j] \subset F^{i-1+j} + [[F^{i-1}, F^j], F^1] + [[F^j, F^1], F^{i-1}]$$

$$\subset F^{i-1+j} + [F^{i-1+j}, F^1] + [F^{j+1}, F^{i-1}] \subset F^{i+j} + [F^{i-1}, F^{j+1}] \subset F^{i+j}. \quad \square$$

If $F^\infty = \bigcup_{j=1}^{\infty} F^j$, then F^∞ is a Lie subalgebra of $\mathcal{A}$, by Lemma 2.2.1, and thus $Lie(F) \subset F^\infty$. On the other hand, by definition $F^1 =< F >\subset Lie(F)$ and inductively we have $F^{j+1} = F^j + < [F^1, F^J] >\subset Lie(F)$. Consequently,

$$Lie(F) = \bigcup_{j=1}^{\infty} F^j.$$

Note that if $F^{j_0+1} = F^{j_0}$ for some $j_0 \in \mathbb{N}$, then $Lie(F) = F^{j_0}$. So, if $\mathcal{A}$ is finite dimensional, there always exists some $j_0 \in \mathbb{N}$ such that

$$< F >= F^1 \subset F^2 \subset \cdots \subset F^{j_0} = Lie(F).$$

Let now M be a connected Sub-Riemannian manifold with corresponding horizontal distribution D. Let $\mathcal{D}$ denote the subsheaf of the tangent sheaf $\mathcal{TM}$ of M consisting of germs of horizontal local smooth vector fields. As in the case of Lie algebras we have a flag of subsheaves

$$\mathcal{D} = \mathcal{D}^1 \subset \mathcal{D}^2 \subset \cdots \subset \mathcal{D}^j \subset \mathcal{D}^{j+1} \subset \cdots \subset \mathcal{TM},$$

where $\mathcal{D}^{j+1} = \mathcal{D}^j + < [\mathcal{D}^1, \mathcal{D}^j] >$, the span taken over the germs of smooth functions defined on open subsets of M.

Definition 2.2.2. The Sub-Riemannian structure of the connected Sub-Riemannian manifold M is said to be Lie bracket generating if for every $x \in M$ there exists $r(x) \in \mathbb{N}$ such that $\mathcal{D}_x^{r(x)-1} \neq \mathcal{D}_x^{r(x)} = \mathcal{TM}_x$. We call $r(x)$ the step (or degree of non-holonomy) of the Sub-Riemannian structure of M at the point $x \in M$. If $ev_x : \mathcal{D}^i \to T_xM$ denotes the evaluation at $x \in M$ and $D_x^i = ev_x(\mathcal{D}_x^i)$,

$1 \leq i \leq r(x)$, we also put $n_j(x) = \dim D_x^j$ and call $(n_1(x), n_2(x), ..., n_{r(x)}(x))$ the growth vector at x.

The step may vary from point to point and in this case $\mathcal{D}^j$ is a sheaf which does not arise as a sheaf of germs of smooth sections of a vector subbundle of the tangent bundle TM of M.

Definition 2.2.3. The Sub-Riemannian structure of M is called regular at the point $x \in M$ if the growth vector is constant on an open neighbourhood of x. If the growth vector is constant throughout M, then the Sub-Riemannian structure is called equiregular.

The continuity of linear independence implies that if the Sub-Riemannian structure is Lie bracket generating, then the step r is an upper semicontinuous function on M and each n_j, is a lower semicontinuous function, because every point $x \in M$ has an open neighbourhood V such that $r(y) \leq r(x)$ for every $y \in V$. Moreover, the set $W_j \subset V$ of points at which n_j takes its maximal value on V is not empty and open, since n_j is lower semicontinuous. Thus, the set $W_1 \cap W_2 \cap \cdots \cap W_{r(x)}$ is open and consists of regular points. Therefore, the set of regular points is an open and dense subset of M.

Examples 2.2.4. (a) The Sub-Riemannian structure of the Heisenberg group is free and of contact type and defined by the subbundle of the tangent bundle $T\mathbb{R}^3$ of $\mathbb{R}^3$ which is the kernel of the differential 1-form

$$\omega = dz - \frac{1}{2}(xdy - ydx).$$

It is generated by the globally defined smooth vector fields

$$X = \frac{\partial}{\partial x} - \frac{1}{2}y\frac{\partial}{\partial z}, \quad Y = \frac{\partial}{\partial y} + \frac{1}{2}x\frac{\partial}{\partial z}.$$

Since $[X, Y] = \dfrac{\partial}{\partial z}$, it follows that the Sub-Riemannian structure of the Heisenberg group is Lie bracket generating of constant step 2 and growth vector $(2, 3)$. In particular, it is equiregular.

(b) The Martinet distribution on $\mathbb{R}^3$ is the kernel of the differential 1-form

$$\omega = dz - y^2 dx$$

and is a Sub-Riemannian structure of constant rank 2. The kernel of ω is generated by the globally defined smooth vector fields

$$X = \frac{\partial}{\partial x} + y^2\frac{\partial}{\partial z} \quad \text{and} \quad Y = \frac{\partial}{\partial y}.$$

Since $[X, Y] = -2y\dfrac{\partial}{\partial z}$ and $[[X, Y], Y] = 2y\dfrac{\partial}{\partial z}$, the Martinet distribution is Lie bracket generating. The growth vector outside the plane $y = 0$ is $(2, 3)$ and on that plane it is $(2, 2, 3)$. Thus, this distribution is not equiregular, as no point of the plane $y = 0$ is regular.

(c) On $\mathbb{R}^3$ we consider the Sub-Riemannian structure of constant rank 2 defined by the subbundle of $T\mathbb{R}^3$ which is generated by the globally defined smooth vector fields

$$X = \frac{\partial}{\partial x} \quad \text{and} \quad Y = \frac{\partial}{\partial y} + xz\frac{\partial}{\partial z}.$$

Then we have

$$[X, Y] = z\frac{\partial}{\partial z}, \quad [X, [X, Y]] = 0, \quad [Y, [X, Y]] = 0.$$

Hence off the plane $z = 0$ this Sub-Riemannian structure is Lie bracket generating of step 2, but is not Lie bracket generating at the points of this plane.

2.3 The Chow-Rashevskii Theorem

This section is devoted to the proof of the following theorem which gives answers to Questions 2.1.1 and 2.1.2.

Theorem 2.3.1. (W-L. Chow and P.K. Rashevskii) *Let M be a connected smooth Sub-Riemannian n-manifold with corresponding Sub-Riemannian "distance" d. If the Sub-Riemannian structure of M satisfies the Lie bracket generating condition, then (M, d) is a metric space and the topology induced by d coincides with the manifold topology.*

The proof will be divided into steps. The most crucial and informative step is the following.

Proposition 2.3.2. *Let M be a smooth Sub-Riemannian n-manifold. If the Sub-Riemannian structure of M satisfies the Lie bracket generating condition, then for every $x \in M$ and every $\epsilon > 0$ there exists an open neighbourhood U of x in M such that every point of U can be joined in M to x with a piecewise smooth horizontal curve with L^∞ controls of length at most ϵ.*

Proof. There exists an open neighbourhood of the point $x \in M$ over which the horizontal distribution is generated (over C^∞ functions) be a finite set of horizontal smooth vector fields $X_1, X_2, ..., X_m$ for some $m \in \mathbb{N}$. Let Φ^i be the (local) flow of X_i, $1 \le i \le m$. We shall show first that for every open neighbourhood V of $0 \in \mathbb{R}^n$ there exist $(s_1, s_2, ..., s_n) \in V$ and $1 \le i_1, i_2, ..., i_n \le m$ such that $(s_1, s_2, ..., s_n)$ is a regular point of the smooth map $\psi : V \to M$ defined by

$$\psi(t_1, t_2, ..., t_n) = (\Phi^{i_n}_{t_n} \circ \cdots \Phi^{i_1}_{t_1})(x).$$

There exists $1 \le i_1 \le m$ such that $X_{i_1}(x) \ne 0$, because otherwise the Sub-Riemannian structure is not Lie bracket generating at x. So there exists an open interval I_1 containing $0 \in \mathbb{R}$ such that the map $\phi_1 : I_1 \to V$ with $\phi_1(t) = \Phi^1_t(x)$

is a smooth embedding. Thus, $S_1 = \phi_1(I_1)$ is a regular 1-dimensional submanifold of M. If $n = 1$, we have finished. If $n > 1$, there exists some $t_1^1 \in I_1$ and some $1 \leq i_2 \leq m$ such that $X_{i_2}(\Phi^1_{t_1^1}(x))$ is not tangent at S_1, because otherwise all X_1, X_2,..., X_m are tangent to S_1 and the Sub-Riemannian structure is not Lie bracket generating. The map $\phi_2(t_1, t_2) = (\Phi^{i_2}_{t_2} \circ \Phi^{i_1}_{t_1})(x)$ is defined and is smooth on some open neighbourhood of $(t_1^1, 0)$ in $\mathbb{R}^2$ with values in V and $\dfrac{\partial \phi_2}{\partial t_1}(t_1^1, 0) = X_{i_1}(\Phi^{i_1}_{t_1^1}(x))$ is tangent to S_1, while $\dfrac{\partial \phi_2}{\partial t_2}(t_1^1, 0) = X_{i_2}(\Phi^{i_1}_{t_1^1}(x))$ is transverse to S_1. Hence, ϕ_2 maps some smaller open neighbourhood of $(t_1^1, 0)$ diffeomorphically onto a regular 2-dimensional submanifold S_2 of M. In particular, the conclusion is proved in case $n = 2$. If $n > 2$, there exist (t_1^2, t_2^2) close enough to $(t_1^1, 0)$ and some $1 \leq i_3 \leq m$ such that $X_{i_3}(\phi_2(t_1^2, t_2^2))$ is not tangent to S_2, because the Sub-Riemannian structure is assumed to be Lie bracket generating. The map $\phi_3(t_1, t_2, t_3) = (\Phi^{i_3}_{t_3} \circ \Phi^{i_2}_{t_2} \circ \Phi^{i_1}_{t_1})(x)$ is defined and is smooth on an open neighbourhood of $(t_1^2, t_2^2, 0)$ in $\mathbb{R}^3$ with values in V and

$$\frac{\partial \phi_3}{\partial t_1}(t_1^2, t_2^2, 0), \quad \frac{\partial \phi_3}{\partial t_2}(t_1^2, t_2^2, 0) \in T_{\phi_2(t_1^2, t_2^2)} S_2$$

while $\dfrac{\partial \phi_3}{\partial t_3}(t_1^2, t_2^2, 0) = X_{i_3}(\phi_2(t_1^2, t_2^2))$. Again since $X_{i_3}(\phi_2(t_1^2, t_2^2))$ is not tangent to S_2, it follows that ϕ_3 maps some open neighbourhood of $(t_1^2, t_2^2, 0)$ diffeomorphically onto a regular 3-dimensional submanifold S_3 of M. If $n = 3$, we have finished. If $n > 3$ we repeat the same argument a finite of times to reach the assertion.

Since $(s_1, s_2, ..., s_n)$ is a regular point of ψ, there exists an open neighbourhood $W \subset V$ of $(s_1, s_2, ..., s_n)$ such that $\psi(W)$ is an open subset of M and ψ maps W diffeomorphically onto $\psi(W)$. Note that $x \notin \psi(W)$. In order to get a local parametrization of M around x, we consider the smooth embedding $\hat{\psi} : W \to M$ defined by

$$\hat{\psi}(t_1, t_2, ..., t_n) = (\Phi^{i_1}_{-s_1} \circ \Phi^{i_2}_{-s_2} \circ \cdots \circ \Phi^{i_n}_{-s_n})(\psi(t_1, t_2, ..., t_n)).$$

Then, $\hat{\psi}(s_1, s_2, ..., s_n) = x$ and $\hat{\psi}(W)$ is an open subset of M.

If now we are given $\epsilon > 0$, we apply the above taking

$$V = \{(t_1, t_2, ..., t_n) \in \mathbb{R}^n : |t_1| + |t_2| + \cdots + |t_n| < \frac{\epsilon}{2}\}.$$

If $y = \hat{\psi}(t_1, t_2, ..., t_n) \in \hat{\psi}(W)$, then y can be joined to x with a piecewise smooth curve γ (not necessarily in $\hat{\psi}(W)$) consisting of pieces of integral curves of the vector fields X_{i_1}, X_{i_2},..., X_{i_n} of successive time lengths $|t_1|$, $|t_2|$,..., $|t_n|$, $|s_1|$, $|s_2|$,..., $|s_n|$. Therefore, γ is horizontal with L^∞ controls and

$$L(\gamma) \leq |t_1| + |t_2| + \cdots + |t_n| + |s_1| + |s_2| + \cdots + |s_n| < \epsilon.$$

Consequently, it suffices to take $U = \hat{\psi}(W)$. $\square$

Rephrasing Proposition 2.3.2, if the Sub-Riemannian structure is Lie bracket generating, then for every $x \in M$ and every $\epsilon > 0$ there exists an open set $U \subset M$ such that $x \in U \subset \{y \in M : d(x, y) < \epsilon\}$. This means that each "open ball"

$B(x,\epsilon) = \{y \in M : d(x,y) < \epsilon\}$ is an open subset of M. Moreover, the function $d(x,.)$ is continuous on U.

Going one step further, in order to prove that d is actually finite on all of $M \times M$, always assuming the Lie bracket generating condition, we consider the binary relation

$$x \sim y \text{ if and only if } d(x,y) \text{ is finite.}$$

This is an equivalence relation, because d is symmetric and satisfies the triangle inequality. Also, Proposition 2.3.2 implies that the corresponding equivalence classes are open subsets of M. Thus, if M is connected, there must be only one equivalence class and d must be finite on $M \times M$.

Proof of Theorem 2.3.1. So far we have proved that d is a pseudo-distance and that the manifold topology is finer than the d-topology. To conclude the proof of Theorem 2.3.1 it remains to prove that the d-topology is finer than the manifold topology.

Let $x \in M$ and let U be an open neighbourhood of x in M for which there exists a smooth diffeomorphism $\phi : U \to \mathbb{R}^n$. Let $K \subset U$ be a compact neighbourhood of x. There exists $\delta > 0$ (depending on K) such that if $\gamma : [0,T] \to U$, $T > 0$, is a horizontal curve with L^1 controls, $\gamma(0) = x$ and $L(\gamma) \leq \delta$, then $\gamma([0,T]) \subset K$. Indeed, by the equivalence of norms in finite dimensional real vector spaces and the compactness of K, there exists a constant $c > 0$ such that

$$\|\phi(\gamma(t)) - \phi(\gamma(0))\| \leq c \int_0^t |\dot{\gamma}(s)| ds$$

for every horizontal curve $\gamma : [0,T_0] \to K$ with L^1 controls and $\gamma(0) = x$, where the norm in the left hand side is euclidean. There exists a small enough $\delta > 0$ such that

$$0 < c\delta < \inf\{\|\phi(x) - y\| : y \in \partial\phi(K)\}.$$

If there exists a horizontal curve $\gamma : [0,T] \to U$, $T > 0$, with L^1 controls such that $\gamma(0) = x$ and $\gamma([0,T])$ is not contained in K, then $T_0 = \sup\{t \in [0,T] : \gamma(t) \in K\} < T$ and $\gamma(T_0) \in \partial K$. Therefore

$$L(\gamma) \geq \frac{1}{c}\|\phi(\gamma(T_0)) - \phi(x)\| \geq \delta.$$

It follows immediately from the assertion we have just proved that

$$\{y \in M : d(x,y) < \delta\} \subset K \subset U$$

which means that U is open in the d-topology. This concludes the proof of Theorem 2.3.1. □

Example 2.3.3. Let M be a smooth 3-manifold and let a be a contact differential 1-form, that is $a \wedge da \neq 0$ everywhere on M. Since a is nowhere vanishing, $D = \mathrm{Ker}a$ is a smooth subbundle of TM which is the horizontal distribution of a Sub-Riemannian structure on M, having chosen a Riemannian metric g on M and restricting it on

D. In particular, M is orientable and $a \wedge da$ is a volume element. If $\psi : M \to \mathbb{R}$ is a nowhere vanishing smooth function, then ψa is another contact differential 1-form which defines the same contact distribution D and the volume element $\psi^2 a \wedge da$. Thus the induced orientation of M depends only on the contact distribution D. The orthogonal line bundle $D^{\perp}$ of D is trivial and generated by a smooth vector field on M such that $a = g(X, \cdot)$ and $g(X, X) = 1$. If Y, Z are two pointwise linearly independent horizontal smooth (local) vector fields then we have

$$(a \wedge da)(X, Y, Z) = 2da(Y, Z) \quad \text{and}$$

$$da(Y, Z) = Ya(Z) - Za(Y) - a([Y, Z]) = -a([Y, Z]).$$

It follows that the contact condition $a \wedge da \neq 0$ is equivalent to $da|_D \neq 0$ and also equivalent to $[Y, Z]_x \notin D_x$ for every point x in the domain of definition of any two pointwise linearly independent horizontal smooth (local) vector fields Y, Z. This last form of the contact condition implies that a Sub-Riamannian structure of contact type on a smooth 3-manifold M satisfies the Lie bracket generating condition and is equiregular with constant growth vector $(2, 3)$. From Proposition 2.3.2, if M is connected, any two points of M can be joined with a piecewise smooth horizontal curve with L^{∞} controls. The origins of this fact can be traced back to C. Caratheodory's theory on the mathematical foundations of thermodynamics.

2.4 Existence of length minimizers

In this section we shall be concerned with the existence, locally and globally, of length minimizers and the Sub-Riemannian version of the Hopf-Rinow theorem. If M is a Sub-Riemannian manifold with a free Sub-Riemannian structure and its horizontal distribution is generated by globally defined smooth vector fields X_1, X_2,..., X_m, for some $m \in \mathbb{N}$, then the length minimizers from $x \in M$ to $y \in M$ are the solutions of the constrained optimal control problem

$$\dot{\gamma}(t) = \sum_{k=1}^{m} u_k(t) X_k(\gamma(t))$$

$$\min \int_0^T \|\dot{\gamma}(t)\| dt$$

where $\gamma : [0, T] \to M$, $T > 0$, with u_1, u_2,..., u_m in L^1 and $\gamma(0) = x$, $\gamma(T) = y$.

Throughout this section we shall assume that M is a Sub-Riemannian n-manifold carrying a Sub-Riemannian structure consisting of a smooth real vector bundle $p : E \to M$ endowed with a Riemannian metric g and a smooth vector bundle morphism $f : E \to TM$ with horizontal distribution $D = \{D_x : x \in M\}$. Moreover, we assume that the Lie bracket generating condition is satisfied, so that if d is the Sub-Riemannian distance, then (M, d) is a metric space and d induces the manifold topology on M, by Theorem 2.3.1.

We shall need the weak semicontinuity of the length functional.

Proposition 2.4.1. *Let γ, $\gamma_k : [0,1] \to M$, $k \in \mathbb{N}$ be absolutely continuous curves with the following properties:*

(i) Each γ_k is horizontal with a.e. constant $\|\dot\gamma\| = L(\gamma_k)$.

(ii) $\gamma = \lim\limits_{k\to+\infty} \gamma_k$ uniformly.

(iii) $\liminf\limits_{k\to+\infty} L(\gamma_k)$ is finite.

Then γ is horizontal with L^∞ controls and $L(\gamma) \le \liminf\limits_{k\to+\infty} L(\gamma_k)$.

Proof. Let us denote $L = \liminf\limits_{k\to+\infty} L(\gamma_k)$. Passing to a subsequence, we may assume that $L = \lim\limits_{k\to+\infty} L(\gamma_k)$ and we have to prove that γ is horizontal with L^∞ controls and $L(\gamma) \le L$. Let $\epsilon > 0$. If K is a compact neighbourhood of $\gamma([0,1])$, there exists $k_0 \in \mathbb{N}$ such that $L(\gamma_k) < L + \epsilon$ and $\gamma_k([0,1]) \subset K$ for every $k \ge k_0$, by uniform convergence. The set

$$V_x = \{f(x,u) : |u| \le L + \epsilon\} \subset T_xM$$

is convex, because f is fiberwise linear. Thus, $\dot\gamma_k(t) \in V_{\gamma_k(t)}$, if $\dot\gamma_k(t)$ exists, for $k \ge k_0$.

Let $0 < t < 1$ and $h > 0$ be so small that $\gamma([t, t+h])$ is contained in an open neighbourhood $U \subset K$ of $\gamma(t)$ in M for which there exists a diffeomorphism $\phi : U \to \mathbb{R}^n$. If u_k^* is the minimal control of γ_k, then

$$\frac{1}{h}[(\phi\circ\gamma_k)(t+h) - (\phi\circ\gamma_k)(t)] = \frac{1}{h}\int_t^{t+h} \phi_{*\gamma_k(s)}(f(\gamma_k(s), u_k^*(s)))ds$$

which belongs to $\overline{\mathrm{conv} \bigcup\limits_{t\le s\le t+h} \phi_{*\gamma_k(s)}(V_{\gamma_k(s)})}$, where conv means convex hull.

On the other hand, taking larger k_0, if necessary, for $k \ge k_0$ and $t \le s \le t+h$ we have $\|\phi(\gamma_k(t)) - \phi(\gamma(t))\| < h$, by uniform convergence, and there exists a constant $c > 0$, depending on K and the metric g, such that

$$\|\phi(\gamma_k(s)) - \phi(\gamma_k(t))\| \le c\int_t^s |u_k^*(\tau)|d\tau \le c(L+\epsilon)h$$

where the norm on the left hand sides are euclidean. It follows that there is a constant $C > 1$ independent of k and h such that

$$\|\phi(\gamma_k(s)) - \phi(\gamma(t))\| \le Ch$$

for every $t \le s \le t+h$. Since the Sub-Riemannian metric d induces the manifold topology, we conclude that $\gamma_k([t, t+h]) \subset B(\gamma(t), r(h))$, where the radii of the Sub-Riamannian balls $B(\gamma(t), r(h))$ satisfy $\lim\limits_{h\to 0^+} r(h) = 0$. Thus, eventually we have $B(\gamma(t), r(h)) \subset U$ for small enough h. Since

$$\frac{1}{h}[(\phi\circ\gamma_k)(t+h) - (\phi\circ\gamma_k)(t)] \in \overline{\mathrm{conv}\bigcup_{t\le s\le t+h} \phi_{*\gamma_k(s)}(V_{\gamma_k(s)})} \subset \overline{\mathrm{conv}\bigcup_{x\in B(\gamma(t),r(h))} \phi_{*x}(V_x)}$$

eventually for all k, taking the limit for $k \to +\infty$ we get

$$\frac{1}{h}[(\phi \circ \gamma)(t+h) - (\phi \circ \gamma)(t)] \in \overline{\mathrm{conv} \bigcup_{x \in B(\gamma(t), r(h))} \phi_{*x}(V_x)}.$$

If now γ is differentiable at t, taking the limit for $h \to 0^+$ we arrive at $\dot{\gamma}(t) \in V_{\gamma(t)}$. So, there exists a unique $u^*(t)$ such that $\dot{\gamma}(t) = f(\gamma(t), u^*(t))$ and

$$\|\dot{\gamma}(t)\| = |u^*(t)| \leq L + \epsilon.$$

Evidently, u^* is the minimal control of γ and γ is horizontal with L^∞ controls. Moreover, $L(\gamma) \leq L + \epsilon$ for every $\epsilon > 0$. $\square$

Corollary 2.4.2. *Let $x \in M$ and $r > 0$ be such that the Sub-Riemannian closed ball $\overline{B(x,r)}$ is compact. Then, for every $y \in B(x,r)$ there exists a length minimizer in $\overline{B(x,r)}$ from x to y.*

Proof. Let $y \in B(x,r)$. There exists a sequence of horizontal curves with L^∞ controls $\gamma_k : [0,1] \to M$ from x to y such that $\|\dot{\gamma}\| = L(\gamma_k)$ a.e. and $\lim_{k \to +\infty} L(\gamma_k) = d(x,y)$. Eventually, $L(\gamma_k) < r$ and $\gamma_k([0,1]) \subset \overline{B(x,r)}$. There exists a constant $c > 0$ depending on $\overline{B(x,r)}$ and the metric g such that if $0 < t < 1$ and $U \subset M$ is an open neighbourhood of $\gamma(t)$ for which there exists a diffeomorphism $\phi : U \to \mathbb{R}^n$, then

$$\|(\phi \circ \gamma)(t) - (\phi \circ \gamma)(s)\| \leq cL(\gamma_k)|t-s| < Cr|t-s|$$

for s close enough to t. This means that γ_k is locally Lipschitz with Lipschitz constant independent of k. Also,

$$d(\gamma_k(t), \gamma_k(s)) \leq L(\gamma_k)|t-s| < r|t-s|$$

for every t, $s \in [0,1]$. It follows from the Arzala-Ascoli theorem that there exists a subsequence $(\gamma_{k_l})_{l \in \mathbb{N}}$ which converges uniformly to some locally Lipschitz curve $\gamma : [0,1] \to M$, hence absolutely continuous. Thus, the assumptions of Proposition 2.4.1 are satisfied and we conclude that γ is horizontal from x to y with L^∞ controls. Moreover, it has length

$$L(\gamma) \leq \liminf_{k \to +\infty} L(\gamma_k) = d(x,y). \quad \square$$

As in a general length metric space, we have now a Sub-Riemannian version of the Hopf-Rinow-Cohn Vossen theorem.

Theorem 2.4.3. *Let M be a connected Sub-Riemannian manifold whose Sub-Riemannian structure satisfies the Lie bracket generating condition with Sub-Riemannian distance d. If the metric space (M,d) is complete, then*

(a) for every $x \in M$ and $r > 0$ the closed Sub-Riemannian ball $\overline{B(x,r)}$ is compact and

(b) for every x, $y \in M$ there exists a length minimizer from x to y.

Proof. The second assertion (b) follows from (a) and Corollary 2.4.2. We proceed to the proof of (a). We first observe that if there exists some point $x \in M$ such that $\overline{B(x,r)}$ is compact for every $r > 0$, then this holds for every point of M. Suppose that this is not true. Then,

$$\rho(x) = \sup\{r > 0 : \overline{B(x,r)} \text{ is compact}\}$$

is finite for every $x \in M$. Since M is locally compact and d induces the manifold topology, we have a well defined function $\rho : M \to (0, +\infty)$. We shall prove that ρ is (uniformly) continuous. Let x, $y \in M$. If $\rho(x) \leq d(x,y)$, then $\rho(x) - \rho(y) \leq d(x,y)$. If $\rho(x) > d(x,y)$, we pick $0 < \delta < \rho(x) - d(x,y)$ and $\rho(x) - \delta < r \leq \rho(x)$. Then, $r - d(x,y) > \rho(x) - \delta - d(x,y) > 0$ and $\overline{B(y, r - d(x,y))} \subset \overline{B(x,r)}$. This implies that $r \leq \rho(y) + d(x,y)$ for all $\rho(x) - \delta < r \leq \rho(x)$ and so $\rho(x) \leq \rho(y) + d(x,y)$. By symmetry, we conclude $|\rho(x) - \rho(y)| \leq d(x,y)$ and ρ is uniformly continuous.

Now we observe that for every $\epsilon > 0$ the set $\overline{B(x, \rho(x) - \epsilon)}$ is a compact ϵ-net in $\overline{B(x,\rho(x))}$. Indeed, let $y \in \overline{B(x,\rho(x))}$. There exists a sequence $(y_k)_{k\in\mathbb{N}}$ in $B(x,\rho(x))$ converging to y and so eventually $d(y_k, y) < \epsilon$. By Corollary 2.4.2, there exists a length minimizer $\gamma_k : [0,1] \to B(x,\rho(x))$ with L^∞ controls from x to y_k. If $y_k \in \overline{B(x,\rho(x) - \epsilon)}$ for some large k, we are done. If not, there exists some $0 < t_k < 1$ such that $d(\gamma_k(t_k), x) = \rho(x) - \epsilon$ and then eventually

$$d(y_k, \gamma_k(t_k)) = d(x, y_k) - d(x, \gamma_k(t_k)) < \rho(x) - (\rho(x) - \epsilon) = \epsilon.$$

Taking the limit for $k \to +\infty$ we get $d(y, \overline{B(x,\rho(x) - \epsilon)}) \leq \epsilon$. This shows that $\overline{B(x,\rho(x) - \epsilon)}$ is a ϵ-net in $\overline{B(x,\rho(x))}$.

Since $\overline{B(x,\rho(x))}$ is d-complete, we conclude that it is compact. Therefore ρ takes on a minimum value η on $\overline{B(x,\rho(x))}$. Let $\{z_1, ..., z_l\}$ be a finite η-net in $\overline{B(x,\rho(x))}$ and

$$A = \bigcup_{i=1}^{l} B(z_i, \eta).$$

The set $\overline{A}$ is compact by the choice of η and of course A contains $\overline{B(x,\rho(x))}$. By the compactness of $\overline{B(x,\rho(x))}$, there exists some $\epsilon > 0$ such that $d(y,z) \geq 2\epsilon$ for every $y \in M \setminus A$ and $z \in \overline{B(x,\rho(x))}$. If $y \in M \setminus A$ and $\gamma : [0,1] \to M$ is any horizontal curve with L^1 controls from x to y, there exists some $0 < t_0 < 1$ such that $\gamma(t_0) \in \partial B(x,\rho(x))$ and

$$L(\gamma) = L(\gamma|_{[0,t_0]}) + L(\gamma|_{[t_0,1]}) \geq \rho(x) + 2\epsilon.$$

Therefore, $d(x,y) > \rho(x) + \epsilon$. This shows that $B(x,\rho(x) + \epsilon) \subset A$. Hence $\overline{B(x,\rho(x) + \epsilon)}$ is compact, which contradicts the definition of $\rho(x)$. This concludes the proof. □

Chapter 3

Sub-Riemannian geodesics

3.1 Normal and abnormal length minimizers

In Riemannian Geometry geodesics locally minimize length and are solutions of a second order (in general non-linear) system of differential equations. Hence they are always smooth curves that can be determined uniquely by initial position and initial velocity. In Sub-Riemannian Geometry this is not possible, because the initial velocities of geodesics emanating from a given point belong to a proper linear subspace of the tangent space at this point. It is however possible to determine the geodesic from its initial point and an element of the cotangent space. This motivates a Hamiltonian approach for the study of Sub-Riemannian geodesics. The starting point is the Pontryagin Maximum Principle.

Let M be a connected Sub-Riemannian n-manifold whose Sub-Riemannian structure consists of a smooth real vector bundle $p : E \to M$ endowed with a Riemannian metric g, a smooth vector bundle morphism $f : E \to TM$ with horizontal distribution $D = \{D_x : x \in M\}$ and satisfies the Lie bracket generating condition. Let X_1, X_2,..., X_m be globally defined horizontal smooth vector fields which generate the horizontal distribution coming from an orthonormal frame of the bundle.

Theorem 3.1.1. *Let $\gamma : [0, L] \to M$, $L > 0$, be a horizontal curve with L^∞ controls parametrized by arclength and with minimal control $u^* = (u_1^*, u_2^*, ..., u_m^*)$, that is*

$$\dot{\gamma}(t) = \sum_{k=1}^{m} u_k^*(t) X_k(\gamma(t))$$

for a.e. $t \in [0, L]$. If γ is a length minimizer, then there exists a (not necessarily unique) continuous lift $\lambda : [0, L] \to T^*M$ of γ such that one of the following is satisfied:

(N) $\lambda(t)(X_k(\gamma(t)) = u_k^*(t)$ *a.e. for every* $1 \leq k \leq m$ *or*

(A) $\lambda(t)(X_k(\gamma(t)) = 0$ *for every* $1 \leq k \leq m$ *and* $\lambda \neq 0$.

$$\begin{tikzcd} & T^*M \arrow[d] \\ {[0,L]} \arrow[ur, dashed, "\lambda"] \arrow[r, "\gamma"] & M \end{tikzcd}$$

In Rimannian Geometry the case (A) does not occur, because then X_1, X_2,..., X_m generate TM (or rather the tangent sheaf $\mathcal{TM}$) and (A) would imply that $\lambda = 0$.

In the proof of Theorem 3.1.1 we shall use flows of non-autonomous vector fields. Recall that a non-autonomous smooth vector field on M is a family of smooth vector fields $(X_t)_{t\in\mathbb{R}}$ of M such that for every smooth function $\phi : M \to \mathbb{R}$ the function $X_.(\phi) : \mathbb{R} \to \mathbb{R}$ is L^∞. An integral curve is by definition a solution of the ODE $\dot{\gamma}(t) = X_t(\gamma(t))$. Locally, it is a solution of an ODE

$$x' = F(t, x)$$

where $F : \Omega \to \mathbb{R}^n$ is a map which is defined on an open set $\Omega \subset \mathbb{R} \times \mathbb{R}^n$ and has the following properties:

(i) $F(., x)$ is measurable and locally bounded for fixed x.

(ii) $F(t, .)$ is smooth for fixed t.

(iii) $F(t, x)$ has locally bounded derivatives with respect to x, meaning that if $I \subset \mathbb{R}$ and $K \subset \mathbb{R}^n$ are compact such that $I \times K \subset \Omega$, then there exists a constant $C > 0$ (depending on I and K) such that

$$\left\| \frac{\partial F}{\partial x} \right\| \leq C \quad \text{on } I \times K.$$

The existence and uniqueness of integral curves are provided by Caratheodory's theorem which we recall.

Theorem 3.1.2. (C. Caratheodory) *If $F : \Omega \to \mathbb{R}^n$ satisfies (i), (ii) and (iii), then for every $(t_0, x_0) \in \Omega$ there exists a unique local solution $x(t; t_0, x_0)$ of the non-autonomous differential equation $x' = F(t, x)$, that is $x'(t; t_0, x_0) = F(t, x(t; t_0, x_0))$ for a.e. t in an open interval with center t_0 and $x(t_0; t_0, x_0) = x_0$. Moreover, $x(t; t_0, x_0)$ is Lipschitz with respect to t and smooth with respect to x_0.*

If we assume that the solutions are defined for every $t \in \mathbb{R}$ and denote $\Phi_{t_0,t}(x_0) = x(t; t_0, x_0)$, then the family of maps $\Phi_{t_0,t}$ is the flow of the non-autonomous vector field. Obviously, $\Phi_{t,t} = \text{id}$, $\Phi_{t_2,t_3} \circ \Phi_{t_1,t_2} = \Phi_{t_1.t_3}$ and $(\Phi_{t_1,t_2})^{-1} = \Phi_{t_2,t_1}$. If the solutions are not defined for all $t \in \mathbb{R}$, the above hold locally. In the case of an autonomous smooth vector field, the corresponding flow is $\Phi_{0,t}$ and satisfies $\Phi_{0,t} \circ \Phi_{0,s} = \Phi_{0,t+s}$.

Proof of Theorem 3.1.1. Since we assume that γ is parametrized by arclength and is a length minimizer, by Proposition 2.1.6, its minimal control u^* minimizes the energy functional $J : L^\infty([0, L], \mathbb{R}^m) \to \mathbb{R}$ defined by

$$J(u) = \frac{1}{2} \int_0^L |u(t)|^2 dt$$

under the constraint that the corresponding to the control u horizontal curve joins $\gamma(0)$ and $\gamma(L)$.

A variation $u = u^* + v$ of u^*, for some $v \in L^\infty([0, L], \mathbb{R}^m)$, defines a horizontal curve $\delta : [0, L] \to M$, which is the solution of $\dot{\delta}(t) = f(\delta(t), u(t))$ a.e. in $[0, L]$ with $\delta(0) = \gamma(0)$. Let $\Phi_{0,t}$ denote the flow of the non-autonomous vector field

$$f(., u^*(t)) = \sum_{k=1}^{m} u_k^*(t) X_k$$

whose solution is γ, so that $\gamma(t) = \Phi_{0,t}(\gamma(0))$ for $t \in [0, L]$. We define

$$x(t) = (\Phi_{0,t})^{-1}(\delta(t)), \quad t \in [0, L].$$

Note that in case $v = 0$ we have $x(t) = \gamma(0)$. In general, $x(t)$ is a solution of an ODE which is obtained by differentiating the formula $\Phi_{0,t}(x(t)) = \delta(t)$. More precisely, we have

$$\begin{aligned} f(\Phi_{0,t}(x(t)), u(t)) = f(\delta(t), u(t)) = \dot{\delta}(t) &= \frac{\partial \Phi_{0,t}}{\partial t}(x(t)) + (\Phi_{0,t})_{*x(t)}(\dot{x}(t)) \\ &= f(\Phi_{0,t}(x(t)), u^*(t)) + (\Phi_{0,t})_{*x(t)}(\dot{x}(t)) \end{aligned}$$

Therefore

$$\dot{x}(t) = (\Phi_{0,t})^{-1}_{*x(t)}(f(\Phi_{0,t}(x(t)), v(t)))$$

and $x(0) = \gamma(0)$. The left hand side of this ODE depends linearly on v. In the sequel, we shall write

$$g_v(t, x) = (\Phi_{0,t})^{-1}_{*x(t)}(f(\Phi_{0,t}(x), v(t)))$$

and $x(t; u^* + v)$ instead of $x(t)$, because $x(t)$ comes from the variation $u^* + v$ of u^*.

For every $v \in L^\infty([0, L], \mathbb{R}^m)$ we consider now $h_v : \mathbb{R} \to \mathbb{R} \times M$ which is defined by

$$h_v(s) = \begin{pmatrix} J(u^* + sv) \\ x(L; u^* + sv) \end{pmatrix} \in \mathbb{R} \times M.$$

Claim. There exists some $\tilde{\lambda} \in T^*(\mathbb{R} \times M) \cong \mathbb{R} \oplus T^*M$, $\tilde{\lambda} \neq 0$, such that $\tilde{\lambda}(\dot{h}_v(0)) = 0$ for every $v \in L^\infty([0, L], \mathbb{R}^m)$.

Proof of the claim. If the claim is not true, there exist v_0, v_1,..., $v_n \in L^\infty([0, L], \mathbb{R}^m)$ such that the vectors

$$\begin{pmatrix} \frac{\partial J(u^*+sv_0)}{\partial s}\big|_{s=0} \\ \frac{\partial x(L;u^*+sv_0)}{\partial s}\big|_{s=0} \end{pmatrix}, \begin{pmatrix} \frac{\partial J(u^*+sv_1)}{\partial s}\big|_{s=0} \\ \frac{\partial x(L;u^*+sv_1)}{\partial s}\big|_{s=0} \end{pmatrix}, \ldots, \begin{pmatrix} \frac{\partial J(u^*+sv_n)}{\partial s}\big|_{s=0} \\ \frac{\partial x(L;u^*+sv_n)}{\partial s}\big|_{s=0} \end{pmatrix}$$

are linearly independent. The map with

$$F(s_0, s_1, ..., s_n) = \begin{pmatrix} J(u^* + s_0 v_0 + s_1 v_1 + \cdots + s_n v_n) \\ x(L; u^* + s_0 v_0 + s_1 v_1 + \cdots + s_n v_n) \end{pmatrix} \in \mathbb{R} \times M$$

is defined and is smooth on some open neighbourhood of 0 in $\mathbb{R}^{n+1}$, by the smooth dependence of the solutions of smooth ODE's from parameters, and the above vectors are the columns of the Jacobian matrix F_{*0}. From the inverse map theorem,

F sends some open neighbourhood of 0 in $\mathbb{R}^{n+1}$ diffeomorphically onto an open neighbourhood of

$$F(0) = \begin{pmatrix} J(u^*) \\ x(L; u^*)) \end{pmatrix} = \begin{pmatrix} J(u^*) \\ \gamma(0)) \end{pmatrix}.$$

This implies that there exist s_0, s_1,..., $s_n \in \mathbb{R}$ such that $x(L; u^* + v) = p$ and $J(u^* + v) < J(u^*)$ for $v = s_0 v_0 + s_1 v_1 + \cdots + s_n v_n$. But since

$$\delta(L) = \Phi_{0,L}(x(L; u^* + v)) = \Phi_{0,L}(\gamma(0)) = \gamma(L),$$

this contradicts the assumption that γ is a (constrained) minimizer of J. This proves the claim.

Normalizing, there exists some $\lambda_0 \in T^*M$ such that $\tilde{\lambda} = (-1, \lambda_0)$ or $\tilde{\lambda} = (0, \lambda_0)$ and in the second case necessarily $\lambda_0 \neq 0$. Now the claim becomes

$$\lambda_0\Big(\frac{\partial x(L; u^* + sv)}{\partial s}\big|_{s=0}\Big) = \frac{\partial J(u^* + sv)}{\partial s}\big|_{s=0}$$

or

$$\lambda_0\Big(\frac{\partial x(L; u^* + sv)}{\partial s}\big|_{s=0}\Big) = 0 \quad \text{and } \lambda_0 \neq 0.$$

On the one hand we have

$$\frac{\partial J(u^* + sv)}{\partial s}\bigg|_{s=0} = \frac{1}{2}\int_0^L \frac{\partial}{\partial s}\bigg|_{s=0} |u^*(t) + sv(t)|^2 dt = \int_0^L \left(\sum_{k=1}^m u_k^*(t) v_k(t)\right) dt$$

for every $v = (v_1, v_2, ..., v_m) \in L^\infty([0, L], \mathbb{R}^m)$. Since

$$x(L; u^* + sv) = x(0; u^* + sv) + \int_0^L g_{sv}(t, x(t; u^* + sv))dt$$

$$= \gamma(0) + s\int_0^L g_v(t, x(t; u^* + sv))dt,$$

on the other hand we have

$$\frac{\partial x(L; u^* + sv)}{\partial s}\bigg|_{s=0} = \int_0^L g_v(t, x(t; u^*))dt = \int_0^L (\Phi_{0,t})^{-1}_{*x(t;u^*)}(f(\Phi_{0,t}(x(t; u^*), v(t)))dt$$

$$= \int_0^L (\Phi_{0,t})^{-1}_{*\gamma(0)}(f(\gamma(t), v(t)))dt$$

and therefore

$$\lambda_0\Big(\frac{\partial x(L; u^* + sv)}{\partial s}\bigg|_{s=0}\Big) = \int_0^L ((\Phi_{0,t})^{-1})^*\lambda_0)(f(\gamma(t), v(t)))dt$$

$$= \int_0^L \left(\sum_{k=1}^m v_k(t)((\Phi_{0,t})^{-1})^*\lambda_0)(X_k(\gamma(t)))\right) dt.$$

Thus, if we put $\lambda(t) = ((\Phi_{0,t})^{-1})^*\lambda_0$, then $\lambda : [0, L] \to T^*M$ is a continuous lift of γ and

$$\int_0^L \left(\sum_{k=1}^m \lambda(t)(X_k(\gamma(t))v_k(t)\right)dt = \int_0^L \left(\sum_{k=1}^m u_k^*(t)v_k(t)\right)dt$$

or

$$\int_0^L \left(\sum_{k=1}^m \lambda(t)(X_k(\gamma(t))v_k(t)\right)dt = 0 \quad \text{and } \lambda \neq 0$$

for every $(v_1, v_2, ..., v_m) \in L^\infty([0, L], \mathbb{R}^m)$. It follows that $\lambda(t)(X_k(\gamma(t))) = u^*(t)$ a.e. on $[0, L]$ for all $1 \le k \le m$ or $\lambda(t)(X_k(\gamma(t))) = 0$ for all $1 \le k \le m$ and $\lambda \neq 0$. □

A length minimizer $\gamma : [0, L] \to M$ parametrized by arclength with minimal control $u^* = (u_1^*, u_2^*, ..., u_m^*)$ is called normal if it satisfies conclusion (N) of Theorem 3.1.1, that is it has a continuous lift $\lambda : [0, L] \to T^*M$ such that $\lambda(t)(X_k(\gamma(t)) = u_k^*(t)$ a.e. for every $1 \le k \le m$. If γ satisfies (A) it will be called abnormal. A non-constant length minimizer can be normal and abnormal at the same time due to the non-uniqueness of the lift λ.

In Riemannian Geometry all minimal geodesics are normal. The question now arises whether there do exist abnormal Sub-Riemannian length minimizers. We shall describe examples of abnormal length minimizers later in this chapter. For the time being we make the observation that normal length minimizers are always smooth. Indeed, if (N) holds, then the minimal control u^* is continuous and therefore $\dot{\gamma}$ is defined everywhere and is continuous. As the definition of λ in the last part of the proof of Theorem 3.1.1 shows, if γ is C^1, then its lift λ is also C^1 and so u^* is C^1. Continuing, this shows inductively that γ and its minimal control u^* are smooth.

Theorem 3.1.1 motivates the introduction of Hamiltonian methods to the study of Sub-Riemannian length minimizers. The Sub-Riemannian Hamiltonian of the Sub-Riemannian manifold M is defined as in the Riemannian case to be the smooth function $H : T^*M \to \mathbb{R}$ with

$$H(q, p) = \frac{1}{2}|p \circ f_q|^2 = \frac{1}{2}\sum_{k=1}^m (p(X_k(q)))^2$$

where $|p \circ f_q|$ denotes the norm of the element

$$E_q \xrightarrow{f_q} T_qM \xrightarrow{p} \mathbb{R}$$

of the dual space E_q^*. Equivalent Sub-Riemannian structures on M define the same Sub-Riemannian Hamiltonian function on T^*M.

Proposition 3.1.2. *A normal length minimizer on M is the projection of an integral curve of the Hamiltonian vector field on T^*M corresponding to the Sub-Riemannian Hamiltonian H.*

Proof. Let $H_k : T^*M \to \mathbb{R}$ be the smooth function $H_k(q, p) = p(X_k(q))$, $1 \le k \le m$. Then,

$$H = \frac{1}{2}\sum_{k=1}^m H_k^2 \quad \text{and} \quad dH = \sum_{k=1}^m H_k dH_k.$$

If we denote by X_H the Hamiltonian vector field of H and X_{H_k} the Hamiltonian vector field of H_k, then

$$X_H = \sum_{k=1}^{m} H_k \cdot X_{H_k}.$$

If $\gamma : [0, L] \to M$ is a normal length minimizer with minimal control u^*, then

$$\dot{\gamma}(t) = \sum_{k=1}^{m} u_k^*(t) X_k(\gamma(t))$$

and therefore

$$\dot{\lambda}(t) = \sum_{k=1}^{m} u_k^*(t) X_{H_k}(\lambda(t))$$

for every $t \in [0, L]$, by the following Remark 3.1.3. The condition (N) of normality now gives

$$\dot{\lambda}(t) = \sum_{k=1}^{m} u_k^*(t) X_{H_k}(\lambda(t)) = \sum_{k=1}^{m} H_k(\lambda(t)) X_{K_k}(\lambda(t)) = X_H(\lambda(t)). \quad \square$$

Remark 3.1.3. Let M be a smooth n-manifold and let Y be a smooth vector field on M with flow ϕ_t. Then the infinitesimal generator Y^* of the flow $(\phi_t^{-1})^*$ on T^*M is the Hamiltonian vector field with corresponding Hamiltonian $H(q, p) = p(Y(q))$. Indeed, if we work locally and

$$Y = \sum_{k=1}^{n} Y_k \frac{\partial}{\partial q^k}, \quad p = \sum_{k=1}^{n} p_k dq^k,$$

then

$$H(q^1, ..., q^k, p_1, ..., p_k) = \sum_{k=1}^{n} Y_k(q^1, ..., q^n) p_k$$

and so

$$dH = \sum_{k=1}^{n} p_k dY_k + \sum_{k=1}^{n} Y_k dp_k = \sum_{k=1}^{n} \left(\sum_{l=1}^{n} p_l \frac{\partial Y_l}{\partial q^k} \right) dq^k + \sum_{k=1}^{n} Y_k dp_k.$$

On the other hand, the value of the infinitesimal generator Y^* at (q, p) is

$$Y^*(q, p) = \left.\frac{d}{dt}\right|_{t=0} (\phi_t(q), (\phi_t^{-1})^*_{*\phi_t(q)}(p)) = (Y(q), \left.\frac{d}{dt}\right|_{t=0} (\phi_t^{-1})^*_{*\phi_t(q)}(p)).$$

Recall that the standard symplectic 2-form ω on T^*M is given locally by the formula

$$\omega = \sum_{k=1}^{n} dq^k \wedge dp_k.$$

Now we compute

$$(i_{Y^*}\omega)(q,p) = \sum_{k=1}^{n} Y_k dp_k - \sum_{k=1}^{n} dp_k\left(\frac{d}{dt}\Big|_{t=0} (\phi_t^{-1})^*_{*\phi_t(q)}(p)\right) dq^k$$

and

$$dp_k\left(\frac{d}{dt}\Big|_{t=0} (\phi_t^{-1})^*_{*\phi_t(q)}(p)\right) = dp_k\left(\sum_{i=1}^{n} p_i \frac{d}{dt}\Big|_{t=0} (\phi_t^{-1})^*_{*\phi_t(q)}(dq^i)\right)$$

$$= dp_k\left(\sum_{i=1}^{n} p_i \frac{d}{dt}\Big|_{t=0} (q^i \circ \phi_{-t})_{*\phi_t(q)}\right) = dp_k\left(\sum_{i=1}^{n} p_i \left(\sum_{j=1}^{n}\left(-\frac{\partial Y_i}{\partial q^j}\right)\frac{\partial}{\partial p^j}\right)\right)$$

$$= -\sum_{i=1}^{n} p_i \frac{\partial Y_i}{\partial q^k}.$$

Substituting we arrive at

$$(i_{Y^*}\omega) = \sum_{k=1}^{n} Y_k dp_k + \sum_{k=1}^{n}\left(\sum_{i=1}^{n} p_i \frac{\partial Y_i}{\partial q^k}\right) dq^k = dH.$$

In Darboux local coordintes on T^*M the integral curves of the Hamiltonian vector field X_H of the Sub-Riemannian Hamiltonian H are the solutions of Hamilton's equations

$$\dot{q} = \frac{\partial H}{\partial p} = \sum_{k=1}^{m} p(X_k(q)) \cdot X_k(q)$$

$$\dot{p} = -\frac{\partial H}{\partial q} = -\sum_{k=1}^{m} p(X_k(q)) \cdot p(\frac{\partial X_k}{\partial q}(q)).$$

A curve on M which is the projection of an integral curve of X_H will be called a normal geodesic of the Sub-Riemannian manifold M. Sub-Riemannian normal geodesics have two basic similar properties like Riemannian geodesics do. Firstly, it follows immediately from the above equations that if $\lambda(t) = (q(t), p(t))$ is an integral curve of X_H, then the rescaled curve $\lambda_a(t) = (q(at), p(at))$ is also an integral curve of X_H for $a \neq 0$. Secondly, normal geodesics are always parametrized proportionally to arclength. Indeed, if γ is a normal geodesic with minimal control $u^* = (u_1^*, u_2^*, ..., u_m^*)$ and lift λ, then the first one of Hamilton's equations becomes the already known

$$\dot{\gamma}(t) = \sum_{k=1}^{m} \lambda(t)(X_k(\gamma(t)) \cdot X_k(\gamma(t)) = \sum_{k=1}^{m} u^*(t) \cdot X_k(\gamma(t))$$

and

$$\|\dot{\gamma}(t)\|^2 = \sum_{k=1}^{m} (u_k^*(t))^2 = 2H(\lambda(t)) = 2H(\lambda(0))$$

since the Sub-Riemannian Hamiltonian is constant along the integral curves of the associated Hamiltonian vector field X_H.

We turn now to abnormal length minimizers. From the definition of the Sub-Riemannian Hamiltonian and the characteristic property (A) follows immediately that if $\gamma : [0, L] \to M$ is an abnormal length minimizer with lift $\lambda : [0, L] \to T^*M$, then $\lambda(t) \in H^{-1}(0)$ for every $t \in [0, L]$. However, $H^{-1}(0) = D^\circ$ is the annihilator of the horizontal distribution, again by the very definition of H. This implies that the abnormal length minimizers do not depend on the Sub-Riemannian metric but only on the horizontal distribution itself.

Before we describe an explicit example, we make a helpful final remark. Let $\sigma : I \to H^{-1}(0) = D^\circ \subset T^*M$ be any smooth curve defined on some open interval I. Using the notations of the proof of Proposition 3.1.2, we have $H_k(\sigma(t)) = 0$ for all $t \in I$ and $1 \leq k \leq m$. Differentiating we get

$$\omega(X_{H_k}(\sigma(t)), \dot{\sigma}(t)) = dH_k(\sigma(t))(\dot{\sigma}(t)) = 0,$$

where ω is the standard symplectic 2-form on T^*M. This implies that in case D° is a smooth submanifold of T^*M, then $X_{H_k} \in \mathrm{Ker}(\omega|_{D^\circ})$ for every $1 \leq k \leq m$.

Example 3.1.4. Let M be a connected, orientable, smooth 3-manifold and a a contact differential 1-form on M. The annihilator $D^\circ \subset T^*M$ of the contact distribution $D = \mathrm{Ker} a$ is a smooth 4-dimensional submanifold of T^*M and is actually the total space of an orientable real line bundle over M, hence trivial, whose fibre is generated by a. In other words, we have an isomorphism of vector bundles $F : M \times \mathbb{R} \to D^\circ$ given by the formula $F(q, t) = (q, ta)$.

$$\begin{tikzcd} M \times \mathbb{R} \arrow[rr, "F"] \arrow[dr, "\text{projection}"'] & & D^\circ \arrow[dl] \\ & M & \end{tikzcd}$$

Suppose that locally $a = h_1 dq^1 + h_2 dq^2 + h_3 dq^3$. Then,

$$F^*(\omega|_{D^\circ}) = \sum_{k=1}^{3} d(q^k \circ F) \wedge d(p_k \circ F) = \sum_{k=1}^{3} dq^k \wedge d(th_k) = \sum_{k=1}^{3} dq^k \wedge (h_k dt + t dh_k)$$
$$= \left(\sum_{k=1}^{3} h_k dq^k \right) \wedge dt + t \sum_{k=1}^{3} dq^k \wedge dh_k = a \wedge dt - t da$$

and so

$$F^*((\omega|_{D^\circ}) \wedge (\omega|_{D^\circ})) = 2t dt \wedge a \wedge da.$$

Since a is a contact form, that is $a \wedge da \neq 0$ on M, it follows that the standard symplectic 2-form ω on T^*M is non-degenerate on $D^\circ \setminus M \times \{0\}$. Actually, the two connected components of $D^\circ \setminus M \times \{0\}$ endowed with the restriction of ω on them are copies of the symplectization of the contact manifold M. Thus, $\mathrm{Ker}(\omega|_{D^\circ \setminus M \times \{0\}}) = \{0\}$, which implies that there are no abnormal length minimizers.

In particular, there are no abnormal length minimizers in the Sub-Riemannian Geometry of the Heisenberg group. If $d_{\mathcal{H}}$ denotes the Sub-Riemannian distance on the Heisenberg group $\mathcal{H}$, then trivially $d_{\mathcal{H}}(x, y) \geq \|x - y\|$ for every x, $y \in \mathcal{H}$, where $\| \cdot \|$ is the euclidean norm of $\mathbb{R}^3$. Hence $d_{\mathcal{H}}$ is complete and from Theorem 2.4.3 we have that any two points of $\mathcal{H}$ can be joined with a length minimizer which is a normal geodesic, that is the projection of a solution of Hamilton's equations of the Sub-Riemannian Hamiltonian H.

Recall that the Heisenberg horizontal distribution is generated by the vector fields

$$X = \frac{\partial}{\partial x} - \frac{1}{2}y\frac{\partial}{\partial z}, \quad Y = \frac{\partial}{\partial y} + \frac{1}{2}x\frac{\partial}{\partial z}.$$

If we denote the coordinates (q,p) on $T^*\mathbb{R}^3$ by $q = (x,y,z)$ and $p = (p_x, p_y, p_z)$, the Sub-Riemannian Hamiltonian is

$$H = \frac{1}{2}(H_X^2 + H_Y^2)$$

where the functions H_X, H_Y are defined by the formulas

$$H_X(q,p) = p_x - \frac{1}{2}yp_z \quad \text{and} \quad H_Y(q,p) = p_y + \frac{1}{2}xp_z.$$

Differentiating we find $DH = H_X \cdot DH_X + H_Y \cdot DH_Y$ and

$$DH_X(x,y,x,p_x,p_y,p_z) = (0, -\frac{1}{2}p_z, 0, 1, 0, -\frac{1}{2}y),$$

$$DH_Y(x,y,x,p_x,p_y,p_z) = (\frac{1}{2}p_z, 0, 0, 0, 1, \frac{1}{2}x).$$

Hence

$$DH(x,y,x,p_x,p_y,p_z) = (\frac{1}{2}p_zH_Y, -\frac{1}{2}p_zH_X, 0, H_X, H_Y, -\frac{1}{2}yH_X + \frac{1}{2}xH_Y)$$

and the corresponding Hamiltonian system of differential equations is

$$\begin{aligned}
\dot{x} &= H_X \\
\dot{y} &= H_Y \\
\dot{z} &= -\frac{1}{2}yH_X + \frac{1}{2}xH_Y \\
\dot{p}_x &= -\frac{1}{2}p_zH_Y \\
\dot{p}_y &= \frac{1}{2}p_zH_X \\
\dot{p}_z &= 0.
\end{aligned}$$

Note that the third equation is just $\dot{z} = -\frac{1}{2}y\dot{x} + \frac{1}{2}x\dot{y}$, which is the differential constraint of the third coordinate of the lifts of solution curves of Dido's isoperimetric problem, as we saw in section 1.1. Also p_z is constant by the sixth equation and so the system can be considered as a parametrized dynamical system on $\mathbb{R}^4$, the parameter being p_z. Setting $u = x + iy$, the first two equations become

$$\dot{u} = H_X + iH_Y = p_x + ip_y + \frac{1}{2}p_ziu.$$

Differentiating and using the fourth and fifth equations we arrive at

$$\ddot{u} = -\frac{1}{2}p_xH_Y + \frac{1}{2}p_ziH_X + \frac{1}{2}p_zi(H_X + iH_Y) = p_zi(H_X + iH_Y) = ip_z\dot{u}.$$

Integrating once we get $\dot{u}(t) = \dot{u}(0)e^{ip_z t}$. A second integration gives

$$u(t) = u(0) + \frac{\dot{u}(0)}{ip_z}(e^{ip_z t} - 1).$$

Thus, the geodesics in the Sub-Riemannian Geometry of the Heisenberg group are given by the general formulas

$$x(t) + iy(t) = \frac{H_X(0) + iH_Y(0)}{ip_z}(e^{ip_z t} - 1) + x(0) + iy(0),$$

$$z(t) = z(0) + \int_0^t [x(s)\dot{y}(s) - \dot{x}(s)y(s)]ds.$$

3.2 Local optimality of normal geodesics

Let M be a Sub-Riemannian smooth n-manifold which satisfies the Lie bracket generating condition, as in the previous section 3.1, the notations of which we adopt in the present section. Let $\gamma : (a, b) \to M$, $a < b$, be a normal geodesic. By definition, γ is the projection on M of an integral curve $(\gamma, \lambda) : (a, b) \to T^*M$ of the Hamiltonian vector field X_H of the Sub-Riemannian Hamiltonian $H : T^*M \to \mathbb{R}$. Without loss of generality we may assume that $H(\gamma(t), \lambda(t)) = \dfrac{1}{2}$ for every $a < t < b$, so that γ is parametrized by arclength. The goal of this section is to prove that γ locally minimizes length, that is to prove that for every $a < c < b$ there exists some $\delta > 0$ such that $[c - \delta, c + \delta] \subset (a, b)$ and $\gamma|_{[c-\delta,c+\delta]}$ is a length minimizer.

Let $x = \gamma(c)$ and $\lambda = \lambda(c)$. There exist a smooth submanifold $S \subset M$ of codimension 1 such that $x \in S$ and $T_xS = \mathrm{Ker}\lambda$ and a differential 1-form $\tilde{\lambda}$ defined on an open neighbourhood Ω of x (with $S \subset \Omega$) such that $\tilde{\lambda}(x) = \lambda$, $T_yS = \mathrm{Ker}\tilde{\lambda}(y)$ and $H(y, \tilde{\lambda}(y)) = \dfrac{1}{2}$ for every $y \in S$. This can be constructed as follows. Let Ω be an open neighbourhood of x contained in the domain of a chart on which we have a generating frame X_1, X_2,..., X_m of the horizontal distribution. Let $h : \Omega \to \mathbb{R}$ be a smooth function with $dh(x) = \lambda$. We assume that γ is non-constant, otherwise there is nothing to prove, and so $dh(x) \neq 0$. Since $H(x, dh(x)) = \dfrac{1}{2}$, we may choose Ω so that $H(y, dh(y)) > 0$ for all $y \in \Omega$. Taking a smaller open neighbourhood if necessary, $S = h^{-1}(h(x))$ is a smooth submanifold of M of codimension 1 and $T_yS = \mathrm{Ker}dh(y)$ for every $y \in S$. If now

$$\phi(y) = \sum_{k=1}^{m} (dh(y)(X_k(y)))^2$$

then

$$H(y, \frac{1}{\sqrt{\phi(y}}dh(y)) = \frac{1}{2}\sum_{k=1}^{m} \frac{1}{\phi(y)}(dh(y)(X_k(y)))^2 = \frac{1}{2}.$$

Thus, it suffices to take $\tilde{\lambda} = \dfrac{1}{\sqrt{\phi}}dh$.

For each $y \in S$ we have a unique integral curve (γ_y, λ_y) of X_H with $\gamma_y(c) = y$ and $\lambda_y(c) = \tilde{\lambda}(y)$. Since

$$\tilde{\lambda}(x)(\dot{\gamma}(c)) = \lambda(c)(\dot{\gamma}(c)) = \lambda(c)\Big(\sum_{k=1}^{m} \lambda(c)(X_k(\gamma(c)))X_k(\gamma(c))\Big) = 1,$$

we have $\dot{\gamma}(c) \notin T_xS$. It follows from the inverse map theorem that there exist $\epsilon > 0$ and some open neighbourhood W of x in S such that the map $\mu : (c-\epsilon, c+\epsilon) \times W \to M$ defined by $\mu(t, y) = \gamma_y(t)$ maps $(c-\epsilon, c+\epsilon) \times W$ diffeomorphically onto an open neighbourhood $U \subset \Omega$ of x. Let $V : U \to \mathbb{R}$ be the smooth function defined by $V(\gamma_y(t) = t$ and ω be the differential 1-form on U with $\omega(\gamma_y(t)) = \lambda_y(t)$.

Lemma 3.2.1. *$\omega = dV$.*

Taking Lemma 3.2.1 for granted, we have $H(y, dV(y)) = \dfrac{1}{2}$ for every $y \in U$. Let $c - \epsilon < t_1 < t_2 < c + \epsilon$ and let $\sigma : [0, L] \to U$, $L > 0$ be a horizontal curve with L^1 controls parametrized by arclength such that $\sigma(0) = \gamma(t_1)$ and $\sigma(L) = \gamma(t_2)$. Then

$$\begin{aligned} L(\gamma|_{[t_1,t_2]}) &= t_2 - t_1 = V(\gamma(t_2)) - V(\gamma(t_1)) = V(\sigma(L)) - V(\sigma(0)) \\ &= \int_0^L dV(\sigma(t))(\dot{\sigma}(t))dt \le \int_0^L \|\dot{\sigma}(t)\|dt = L(\sigma), \end{aligned}$$

because

$$\sum_{k=1}^{m} (dV(y)(X_k(y)))^2 = 2H(y, dV(y)) = 1$$

and so $|dV(\sigma(t))(\dot{\sigma}(t))| \le \|\dot{\sigma}(t)\|$, from the Cauchy-Schwartz inequality.

Thus, γ restricted to any closed subinterval of $(c-\epsilon, c+\epsilon)$ minimizes length within the class of horizontal curves with L^1 controls and with values in U. However, as in the proof of Theorem 2.3.1, there exists some sufficiently small $0 < \delta < \epsilon$ such that for every horizontal curve $\sigma : [0, L] \to U$, $L > 0$, with L^1 controls parametrized by arclength such that $\sigma(0) = x$ and $L(\sigma) \le \delta$ we have $\sigma([0, L]) \subset U$. Then, $\gamma|_{[c-\delta,c+\delta]}$ is a length minimizer.

To conclude the proof of the local optimality of the normal geodesics, it remains to prove Lemma 3.2.1.

Proof of Lemma 3.2.1. The flow ψ of the smooth vector field $X = \mu_*(\dfrac{\partial}{\partial t})$ is defined on an open subset of $\mathbb{R} \times U$. From the definition of V we have $V(\psi_s(y)) = s + V(y)$ and differentiating with respect to y we get $dV(\psi_s(y)) \circ (\psi_s)_{*y} = dV(y)$, or equivalently $(\psi_s)^*(dV) = dV$. In other words, dV is invariant under the flow of X. Note that $V|_S = 0$ and

$$dV(\mu(c, y))(X(\mu(c, y))) = (V \circ \mu)_{*(c,y)}(\frac{\partial}{\partial t}) = 1.$$

On the other hand,

$$\omega(y))(X(y)) = \lambda_y(c)(X(y)) = \lambda_y(c)(\dot{\gamma}_y(c)) = 1$$

for every $y \in W$. Thus, the differential 1-forms ω and dV coincide on W. If we show that ω is also invariant under the flow of X, then $\omega = dV$ on U will follow.

For the invariance of ω it suffices to prove that if $y \in W$, $v_0 \in T_yM$ and $v(t) = (\psi_{t-c})_{*y}(v_0)$, then $(\omega(\mu(t,y))(v(t))$ is constant with respect to t. Locally on M, we have

$$\frac{dv}{dt} = \frac{\partial X}{\partial y}(v)$$

and

$$X(\mu(t,y)) = \dot{\gamma}_y(t) = \sum_{k=1}^{m} \lambda_y(t)(X_k(\gamma_y(t))X_k(\gamma_y(t)) = \sum_{k=1}^{m} \omega(\gamma_y(t))(X_k(\gamma_y(t))X_k(\gamma_y(t)),$$

that is $X = \sum_{k=1}^{m} \omega(X_k)X_k$. Since $\sum_{k=1}^{m} (\omega(X_k))^2 = 1$, differentiating we get

$$\sum_{k=1}^{m} \omega(X_k) \cdot \frac{\partial}{\partial y}\omega(X_k) = 0.$$

Hence

$$\begin{aligned} \omega\big(\frac{\partial X}{\partial y}\big) &= \sum_{k=1}^{m} \omega(X_k)\omega\big(\frac{\partial X_k}{\partial y}\big) + \sum_{k=1}^{m} \omega(X_k)\frac{\partial}{\partial y}\omega(X_k)) \\ &= \sum_{k=1}^{m} \omega(X_k)\omega\big(\frac{\partial X_k}{\partial y}\big) + 0 = -\frac{d}{dt}\omega(\mu(t,y)) \end{aligned}$$

from the second equation of the Hamiltonian system of differential equations. Therefore,

$$\frac{d}{dt}(\omega(\mu(t,y))(u(t)) = \frac{d}{dt}\omega(\mu(t,y)) \cdot v(t) + \omega\big(\frac{\partial X}{\partial y}\big)(v(t)) = 0.$$

This concludes the proof. $\square$.

3.3 The abnormal length minimizer of W. Liu and H.J. Sussmann

In this section we shall present an example of an abnormal length minimizer due to W. Liu and and H.J. Sussmann. On $\mathbb{R}^3$ we consider the nowhere vanishing differential 1-form

$$\theta = x^2 dy - (1-x)dz$$

whose kernel E is a smooth subbundle of $T\mathbb{R}^3$ of rank 2. If

$$X = \frac{\partial}{\partial x} \quad \text{and} \quad Y = (1-x)\frac{\partial}{\partial y} + x^2\frac{\partial}{\partial z},$$

then $\theta(X) = \theta(Y) = 0$ and $\{X, Y\}$ is a global basis of sections of E. On E we consider the metric g which makes $\{X, Y\}$ a global orthonormal frame of E and is defined by formula

$$g_{(x,y,z)}\left(\begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix}, \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix} \right) = v_1u_1 + \frac{1}{(1-x)^2 + x^4}(v_2u_2 + v_3u_3).$$

We have

$$[X,Y] = -\frac{\partial}{\partial y} + 2x\frac{\partial}{\partial z},$$

$$[X,[X,Y]] = 2\frac{\partial}{\partial z}, \quad [Y,[X,Y]] = 0,$$

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

Obviously, X, Y and $[X,Y]$ are linearly independent everywhere except at the planes $x=0$ or $x=2$ and X, Y and $[X,[X,Y]]$ are linearly independent everywhere except at the plane $x=1$. Thus, the Lie bracket generating condition is satisfied and by Theorem 2.3.1 any two points of $\mathbb{R}^3$ can be joined with a piecewise smooth horizontal curve with L^∞ controls. This Sub-Riemannian structure of $\mathbb{R}^3$ is not equiregular, as outside the planes $x=0$ or $x=2$ the step is 2 and the growth vector is $(2,3)$ and on these planes the step is 3 and the growth vector is $(2,2,3)$.

The Sub-Riemannian Hamiltonian $H : T^*\mathbb{R}^3 \to \mathbb{R}$ is given by the formula

$$H(x,y,z,\xi,\eta,\zeta) = \frac{1}{2}\left[\xi^2 + \left((1-x)\eta + x^2\zeta\right)^2\right]$$

and the associated Hamiltonian system of differential equations is

$\dot{x} = \xi$
$\dot{y} = (1-x)[(1-x)\eta + x^2\zeta]$
$\dot{z} = x^2[(1-x)\eta + x^2\zeta]$
$\dot{\xi} = (2x\zeta - \eta)[(1-x)\eta + x^2\zeta]$
$\dot{\eta} = 0$
$\dot{\zeta} = 0.$

For any a, $b \in \mathbb{R}$, $a < b$, the smooth curve $\gamma : [a,b] \to \mathbb{R}^3$ with $\gamma(t) = (0,t,0)$ is parametrized by arclength and is horizontal with controls $u_X^*(t) = 0$ and $u_Y^*(t) = 1$. For any continuous function $\zeta : [a,b] \to \mathbb{R}$, we have a continuous lift $\lambda : [a,b] \to T^*\mathbb{R}^3$ of γ defined by $(\gamma(t),\lambda(t)) = (0,t,0,0,0,\zeta(t))$ which satisfies the abnormality condition (A) of Theorem 3.1.1 and $H(\gamma(t),\lambda(t)) = 0$ for every $a \le t \le b$.

We observe that γ is not a normal geodesic. Indeed, if it were, there would exist smooth functions ξ, η, $\zeta : [a,b] \to \mathbb{R}$ such that $(0,t,0,\xi(t),\eta(t),\zeta(t))$, $a \le t \le b$, is a solution of the above system of differential equations, that is

$$0 = \xi, \quad 1 = \eta, \quad 0 = 0, \quad \dot{\xi} = -\eta^2, \quad \dot{\eta} = 0, \quad \dot{\zeta} = 0$$

from which follows the contradiction $1 = \eta = 0$. We shall prove however that γ is a length minimizer, if $b-a$ is small enough.

Let $\tilde{\gamma} : [\tilde{a},\tilde{b}] \to \mathbb{R}^3$ be any other horizontal curve with L^1 controls from $(0,a,0)$ to $(0,b,0)$ and let $\tau = L(\tilde{\gamma})$. We shall prove that $\tau \ge b-a$, if $b-a$ is sufficiently small. There is no loss of generality if we assume that $\tilde{a} = 0$. Let u, v be the L^1 controls of $\tilde{\gamma}$, which are unique and if $\tilde{\gamma}(t) = (\tilde{x}(t),\tilde{y}(t),\tilde{z}(t))$, then

$$u(t) = \tilde{x}'(t), \quad v(t)(1-\tilde{x}(t)) = \tilde{y}'(t), \quad v(t)(\tilde{x}(t))^2 = \tilde{z}'(t)$$

a.e. on $[0,\tilde{b}]$. Therefore,

$$v(t) = \frac{(1-\tilde{x}(t))\tilde{y}'(t) + (\tilde{x}'(t))^2\tilde{z}'(t)}{(1-\tilde{x}(t))^2 + (\tilde{x}(t))^4}$$

a.e. on $[0, \tilde{b}]$. We also have the additional formulas $\tilde{x}(\tilde{b}) = 0$ and

$$\tilde{x}(t) = \int_0^t u(s)ds,$$

$$\int_0^{\tilde{b}} v(t)(\tilde{x}(t))^2 dt = \tilde{z}(\tilde{b}) - \tilde{z}(0) = 0.$$

We shall use the following technical lemma, whose proof will be postponed.

Lemma 3.3.1. *Let $0 < \tau < \dfrac{2}{3}$ and u, $v \in L^1([0, \tau], \mathbb{R})$ with $|u(t)| \leq 1$ and $|v(t)| \leq 1$ a.e. on $[0, \tau]$. We define the function $x : [0, \tau] \to \mathbb{R}$ by*

$$x(t) = \int_0^t u(s)ds$$

and we assume that $x(\tau) = 0$ and $\displaystyle\int_0^\tau v(t)(x(t))^2 dt = 0$.

Then,

$$\int_0^\tau v(t)(1 - x(t))dt \leq \tau$$

and the equality holds if and only if $u = 0$ and $v = 1$ a.e.

We assume first the $\tilde{\gamma}$ is parametrized by arclength, that is $\tau = \tilde{b}$ and $u^2 + v^2 = 1$. In this case Lemma 3.3.1 gives that if $0 < \tau < \dfrac{2}{3}$, then

$$b - a = \tilde{y}(\tilde{b}) - \tilde{y}(0) = \int_0^{\tilde{b}} v(t)(1 - x(t))dt \leq \tau.$$

Thus, if $0 < b - a < \dfrac{2}{3}$, then either $0 < \tau < \dfrac{2}{3}$, and so $b - a \leq \tau$, from the above, or $\tau \geq \dfrac{2}{3} > b - a$. Hence, if $0 < b - a < \dfrac{2}{3}$, then $b - a \leq \tau$. Moreover, $b - a = \tau$ if and only if $u = 0$ and $v = 1$ a.e., which means that $\tilde{x}(t) = 0$, $\tilde{y}(t) = a + t$ and $\tilde{z}(t) = 0$. In other words, $\tilde{\gamma}(t) = \gamma(a + t)$.

If $\tilde{\gamma}$ is not parametrized by arclength, according to Proposition 1.3.4, there exist an absolutely continuous, surjective and increasing function $h : [0, \tilde{b}] \to [0, \tau]$ and a horizontal curve $\delta : [0, \tau] \to \mathbb{R}^3$ parametrized by arclength such that $\tilde{\gamma} = \delta \circ h$. From the previous part we have $b - a \leq \tau$, if $0 < b - a < \dfrac{2}{3}$, and the equality hods if and only if $\delta(t) = \gamma(a + t)$ or equivalently $\tilde{\gamma}$ is a reparametrization of γ.

Thus, in any case if $0 < b - a < \dfrac{2}{3}$, then γ is a length minimizer and it remains to prove the technical Lemma 3.3.1.

Proof of Lemma 3.3.1. Let $V(t) = \displaystyle\int_0^t v(s)ds$ and $\alpha = \tau - V(\tau)$. Obviously, we have $|V(t)| \leq t$ and $\alpha \geq 0$. If $\beta = \sup\{|x(t)| : 0 \leq t \leq \tau\}$, then

$$\left| \int_0^\tau x(s)v(s)ds \right| \leq \beta\tau$$

and therefore

$$\int_0^\tau v(t)(1-x(t))dt \leq \int_0^\tau v(t)dt + \left|\int_0^\tau x(s)v(s)ds\right| \leq V(\tau)+\beta\tau = \tau-\alpha+\beta\tau.$$

So, it suffices to prove that $\beta\tau \leq \alpha$. This is trivial if $\beta = 0$. Suppose that $\beta > 0$. If $\beta \leq \frac{3}{2}\alpha$, then certainly $\beta\tau \leq \frac{3}{2}\tau\alpha < a$, from our assumption $0 < \tau < \frac{2}{3}$. Hence it suffices to prove that

$$\frac{2}{3}\beta^3 \leq \int_0^\tau (x(t))^2 dt \leq \beta^2\alpha.$$

For the right inequality we have

$$\int_0^\tau (x(t))^2 dt \leq \int_0^\tau (x(t))^2(1-|v(t)|)dt \leq \beta^2\int_0^\tau (1-|v(t)|)dt \leq \beta^2(\tau-|V(\tau)|) \leq \beta^2\alpha.$$

In order to prove the left inequality, let $0 \leq t_0 \leq \tau$ be such that $\beta = |x(t_0)|$. Then, $t_0 \geq \beta$, since $|u| \leq 1$ a.e., and

$$\beta = \left|-\int_{t_0}^\tau u(s)ds\right| \leq \tau - t_0.$$

Thus, $0 \leq t_0-\beta \leq t_0 \leq t_0+\beta \leq \tau$. For $t \in [t_0-\beta, t_0]$ we have

$$\beta \leq \left|\int_0^t u(s)ds\right| + \int_t^{t_0} |u(s)|ds \leq |x(t)| + t_0 - t$$

and similarly for $t \in [t_0, t_0+\beta]$ we have $\beta \leq |x(t)| + t - t_0$. This implies that

$$\int_{t_0-\beta}^{t_0} (x(t))^2 dt \geq \int_{t_0-\beta}^{t_0} (t+\beta-t_0)^2 dt = \frac{\beta^3}{3}$$

and

$$\int_{t_0}^{t_0+\beta} (x(t))^2 dt \geq \int_{t_0}^{t_0+\beta} (-t+\beta+t_0)^2 dt = \frac{\beta^3}{3}.$$

It follows that

$$\int_0^\tau (x(t))^2 dt \geq \frac{2\beta^3}{3}.$$

The above reasoning shows that if

$$\int_0^\tau (1-x(t))v(t)dt = \tau$$

then $\beta\tau = \alpha$ and so necessarily $\beta = \alpha = 0$. This means that $x = 0$ and therefore $u = 0$ a.e. Also, $V(\tau) = \tau$ and this can happen only if $v = 1$ a.e., because $v(t) \leq 1$ a.e. $\square$

This concludes the description of the example of W. Liu and H.J. Sussmann which appeared in 1995. The first example of an abnormal length minimizer had been constructed by R. Montgomery in 1994. It was simplified later by I. Kupka on the one hand and by F. Pelletier and L-V. Bouche on the other.

Chapter 4

Popp's Sub-Riemannian Volume

4.1 Construction of Popp's volume

Let M be a Sub-Riemannian smooth n-manifold whose Sub-Riemannian structure satisfies the Lie bracket generating condition and let

$$\mathcal{D} = \mathcal{D}^1 \subset \mathcal{D}^2 \subset \cdots \subset \mathcal{D}^j \subset \mathcal{D}^{j+1} \subset \cdots \subset \mathcal{TM}$$

be the associated flag of subsheaves of the tangent sheaf $\mathcal{TM}$ of M. Recall that $\mathcal{D}$ is the subsheaf of $\mathcal{TM}$ of horizontal vector fields and $\mathcal{D}^{j+1} = \mathcal{D}^j + < [\mathcal{D}^1, \mathcal{D}^j] >$, where the span is taken over the germs of smooth functions defined on open subsets of M. The Lie bracket generating condition ensures, by definition, that for every $x \in M$ there exists some $r(x) \in \mathbb{N}$ such that $\mathcal{D}_x^{r(x)-1} \neq \mathcal{D}_x^{r(x)} = \mathcal{TM}_x$.

Let $x \in M$ be a regular point, that is the step $r(x)$ and the growth vector $(n_1(x), n_2(x), ..., n_{r(x)}(x))$ are constant on an open neighbourhood U of x, where $n_j(x) = \dim \mathcal{D}_x^j$. Let $gr(\mathcal{D})$ denote the graded sheaf

$$gr(\mathcal{D}) = \mathcal{D}^1 \oplus \mathcal{D}^2/\mathcal{D}^1 \oplus \cdots \oplus \mathcal{D}^r/\mathcal{D}^{r-1}$$

over U. The Lie bracket of local vector fields induces a bilinear map $[\cdot, \cdot]$ on $gr(\mathcal{D})$ which respects the grading, meaning that

$$[\mathcal{D}^i/\mathcal{D}^{i-1}, \mathcal{D}^j/\mathcal{D}^{j-1}] \subset \mathcal{D}^{i+j}/\mathcal{D}^{i+j-1}.$$

Indeed, if X, $X' \in \mathcal{D}^i$, Y, $Y' \in \mathcal{D}^j$ and $X' - X = V \in \mathcal{D}^{i-1}$, $Y' - Y = W \in \mathcal{D}^{j-1}$, then

$$[X', Y'] - [X, Y] = [X, W] + [V, Y] + [V, W]$$

and $[X, W]$, $[V, Y]$, $[V, W] \in \mathcal{D}^{i+j-1}$. The stalk of $gr(\mathcal{D})$ at x is called the nilpotentization of $\mathcal{D}$ at x. In general the nilpotentizations at different regular points are not isomorphic Lie algebras.

The following lemma is crucial for the rest of this section.

Lemma 4.1.1. (O. Popp) *Let E be a real vector space of finite dimension n with a flag of subspaces*

$$\{0\} = E^0 \leq E^1 \leq \cdots \leq E^{r-1} \leq E^r = E$$

for some $r \in \mathbb{N}$. *Let* $gr(E) = E^1 \oplus (E^2/E^1) \oplus \cdots \oplus (E^r/E^{r-1})$ *be the associated graded vector space. Then, there exists a natural isomorphism*

$$\theta : \bigwedge{}^n E^* \to \bigwedge{}^n gr(E)^*.$$

Proof. Let $[X_1, X_2, ..., X_n]$ be an ordered basis of E adapted to the given flag of subspaces. This means that if $k_i = \dim E_i$, then $[X_1, X_2, ..., X_{k_i}]$ is a basis of E_i. We define $\hat{\theta} : E \to gr(E)$ by

$$\hat{\theta}(X_{k_i+l} = X_{k_i+l} + E^i \in E^{i+1}/E^i, \quad 1 \leq l \leq k_{i+1} - k_i.$$

This is a linear isomorphism, which is not natural since it depends on the initial choice of the adapted basis. However, the induced linear isomorphism $\theta : \bigwedge^n E \to \bigwedge^n gr(E)$ does not depend on the choice of the adapted basis. If $[X_1^*, X_2^*, ..., X_n^*]$ is the dual basis on E^*, then θ is defined by

$$\theta(X_1^* \wedge X_2^* \wedge \cdots \wedge X_n^*) = \hat{\theta}(X_1)^* \wedge \hat{\theta}(X_2)^* \wedge \cdots \wedge \hat{\theta}(X_n)^*.$$

We shall show that this holds for any other choice of adapted basis. Let $[Y_1, Y_2, ..., Y_n]$ be another adapted basis of E. The change of basis matrix is of the form

$$\begin{pmatrix} A_1 & * & \cdots & * \\ 0 & A_2 & \cdots & * \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & A_r \end{pmatrix}$$

where A_i is $(k_i - k_{i-1}) \times (k_i - k_{i-1})$ matrix. So, if $[Y_1^*, Y_2^*, ..., Y_n^*]$ is the dual basis we have

$$Y_1^* \wedge Y_2^* \wedge \cdots \wedge Y_n^* = \left(\prod_{i=1}^r \det A_i\right) X_1^* \wedge X_2^* \wedge \cdots \wedge X_n^*$$

On the other hand, the corresponding change of basis matrix on $gr(E)$ is

$$\begin{pmatrix} A_1 & 0 & \cdots & 0 \\ 0 & A_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & A_r \end{pmatrix}$$

and therefore

$$\hat{\theta}(Y_1)^* \wedge \hat{\theta}(Y_2)^* \wedge \cdots \wedge \hat{\theta}(Y_n)^* = \left(\prod_{i=1}^r \det A_i\right) \hat{\theta}(X_1)^* \wedge \hat{\theta}(X_2)^* \wedge \cdots \wedge \hat{\theta}(X_n)^*$$

$$= \left(\prod_{i=1}^r \det A_i\right) \theta(X_1^* \wedge X_2^* \wedge \cdots \wedge X_n^*) = \theta\left(\left(\prod_{i=1}^r \det A_i\right) X_1^* \wedge X_2^* \wedge \cdots \wedge X_n^*\right)$$

$$= \theta(Y_1^* \wedge Y_2^* \wedge \cdots \wedge Y_n^*). \quad \square$$

Let now M be an equiregular Sub-Riemannian n-manifold with growth vector $(n_1, n_2, ..., n_r)$ and associated flag $\mathcal{D} = \mathcal{D}^1 \subset \mathcal{D}^2 \subset \cdots \subset \mathcal{D}^r \subset \mathcal{TM}$. For each $x \in M$ let $ev_x : \mathcal{D}^i \to T_xM$ denote the evaluation at x and $D^i_x = ev_x(\mathcal{D}^i_x)$, $1 \le i \le r$. So, we have a family D^i, $1 \le i \le r$, of distributions, where $D^1 = D$ is the horizontal distribution of the Sub-Riemannian structure, and a flag of vector subspaces

$$D_x = D^1_x \le D^2_x \le \cdots \le D^r_x = T_xM$$

for each $x \in M$.

Let $x \in M$ be fixed and v, $w \in D^1_x$. Let V and W be arbitrary horizontal extensions of v and w, respectively, to smooth local vector fields defined on an open neighbourhood of x. Then, the element $[V, W](x) + D^1_x \in D^2_x/D^1_x$ does not depend on the choices of the extensions V, W. Indeed, let $\{X_1, X_2, ..., X_m\}$ be a generating frame of horizontal vector fields on a perhaps smaller open neighbourhood of x. If $\tilde{V}$ and $\tilde{W}$ is another choice of local horizontal extensions of v and w, respectively, then

$$\tilde{V} - V = \sum_{k=1}^{m} f_k X_k \quad \text{and} \quad \tilde{W} - W = \sum_{k=1}^{m} g_k X_k$$

for some smooth functions f_k, g_k, $1 \le k \le m$, defined on some open neighbourhood of x such that $f_k(x) = g_k(x) = 0$ for every $1 \le k \le m$. We have

$$\begin{aligned} &[\tilde{V}, \tilde{W}] - [V, W] \\ &= \sum_{k=1}^{m} \big(V(g_k)X_k + g_k[V, X_k]\big) + \sum_{k=1}^{m} \big(W(f_k)X_k - f_k[X_k, W]\big) \\ &\quad + \sum_{k,ll=1}^{m} \big(f_k g_l [X_k, X_l] - g_l X_l(f_k) X_k - f_k X_k(g_l) X_l\big) \end{aligned}$$

and evaluating at x we get

$$[\tilde{V}, \tilde{W}](x) - [V, W](x) = \sum_{k=1}^{m} \big(V(g_k)(x) - W(f_k)(x)\big) X_k(x).$$

Therefore, $[\tilde{V}, \tilde{W}](x) - [V, W](x) \in D^1_x$.

The above implies that the linear epimorphism $\tilde{\pi} : \mathcal{D}^1_x \otimes \mathcal{D}^1_x \to (\mathcal{D}^2/D^1)_x$ defined by

$$\tilde{\pi}([V]_x \otimes [W]_x) = \big[[V, W]\big]_x + \mathcal{D}^1_x$$

induces a linear epimorphism $\pi : D^1_x \otimes D^1_x \to D^2_x/D^1_x$ such that the following diagram commutes. Here $\big[\cdot\big]_x$ denotes the germ at x.

$$\begin{tikzcd} \mathcal{D}^1_x \otimes \mathcal{D}^1_x \arrow[rr, "\tilde{\pi}"] \arrow[d, "ev_x \otimes ev_x"'] & & (\mathcal{D}^2/\mathcal{D}^1)_x \arrow[d, "\pi"] \\ D^1_x \otimes D^1_x \arrow[rr, "\pi"] & & D^2_x/D^1_x \end{tikzcd}$$

Similarly, for every $2 \leq k \leq r$ we have a well defined linear epimorphism $\tilde{\pi}_k : \bigotimes_{i=1}^{k} \mathcal{D}_x^1 \to (\mathcal{D}^k/\mathcal{D}^{k-1})_x$ by

$$\tilde{\pi}_k([V_1]_x \otimes \cdots \otimes [V]_k) = \big[[V_1, [V_2[\cdots, [V_{k-1}, V_k]\cdots]\big]_x + \mathcal{D}_x^{k-1}$$

which induces a linear epimorphism $\pi_k : \bigotimes_{i=1}^{k} D_x^1 \to D_x^k/D_x^{k-1}$.

Since D_x^1 carries an inner product, π_k induces an inner product on D_x^k/D_x^{k-1}. Further, we get an inner product on

$$gr(T_xM) = D_x^1 \oplus (D^2/D_x^1) \oplus \cdots \oplus (D_x^r/D_x^{r-1})$$

which makes this direct sum an orthogonal direct sum. Using the natural isomorphism

$$\theta : \bigwedge{}^n (T_xM)^* \to \bigwedge{}^n (gr(T_xM))^*$$

of Lemma 4.1.1 we get a canonical volume on T_xM, defined from any orthonormal basis of $gr(T_xM)$.

The above can be carried out in an open neighbourhood of the point x. So, if M is orientable, gluing in the usual way we get a smooth volume element on M. Its smoothness will follow from the local formula which will be proved in the sequel. This is Popp's volume on M.

First we consider the case where the equiregular Sub-Riemannian structure is of step 2. The growth vector is then (k, n) where $k = \dim D_x$ for all $x \in M$ and $\mathcal{D}+ <[\mathcal{D}, \mathcal{D}]>= \mathcal{TM}$. Let $\{X_1, X_2, ..., X_k\}$ be a local orthonormal frame of the horizontal distribution D and let $X_{k+1}, ..., X_n$ be any choice of pointwise linearly independent local vector fields such that $\{X_1, X_2, ..., X_k, X_{k+1}, ..., X_n\}$ is a local frame of TM. According to the construction of Popp's volume we need to compute the inner product on the orthogonal direct sum

$$gr(T_xM) = D_x \oplus (T_xM/D_x).$$

By construction, this inner product is induced from the linear epimorphism

$$\pi : D_x \otimes D_x \to T_xM/D_x$$

with $\pi(X_i \otimes X_j) = [X_i, X_j] + D_x$ all evaluated at x. Recall that if $\langle \cdot, \cdot \rangle$ is the Sub-Riemannian inner product on D_x, the inner product of $D_x \otimes D_x$ is given by the formula $\langle V_1 \otimes W_1, V_2 \otimes W_2 \rangle = \langle V_1, V_2 \rangle \cdot \langle W_1, W_2 \rangle$. We shall calculate explicitly the norm on T_xM/D_x from which the inner product will follow by polarization.

There exist smooth locally defined functions b_{ij}^l, $k+1 \leq l \leq n$, $1 \leq i, j \leq k$, on an open neighbourhood of x such that

$$[X_i, X_j](x) + D_x = \sum_{l=k+1}^{n} b_{ij}^l X_l(x) + D_x.$$

Let $v + D_x = \sum_{l=k+1}^{n} c^l X_l(x) + D_x \in T_xM/D_x$ and $w = \sum_{i,j=1}^{k} a_{ij} X_i(x) \otimes X_j(x)$ with $\pi(w) = v + D_x$ or equivalently

$$\sum_{l=k+1}^{n} \Big(\sum_{i,j=1}^{k} a_{ij} b_{ij}^l \Big) X_l(x) = \sum_{l=k+1}^{n} c^l X_l(x).$$

We can consider $b^l = (b_{ij}^l)_{1 \leq i,j \leq k}$ and $a = (a_{ij})_{1 \leq i,j \leq k}$ as vectors in $\mathbb{R}^{k^2}$ and then the norm of $v + D_x$ is

$$\min\{\|a\| : \langle a, b^l \rangle = c^l, \quad k+1 \leq l \leq n\},$$

where the inner product $\langle \cdot, \cdot \rangle$ and norm $\| \cdot \|$ are euclidean. Note that the vectors $b^l = (b_{ij}^l)_{1 \leq i,j \leq k}$, $k+1 \leq l \leq n$, are linear independent, because they are the rows of the matrix of the epimorphism π with respect to the ordered basis $[X_i(x) \otimes X_j(x) : 1 \leq i,j \leq k]$ of $D_x \otimes D_x$ and $[X_{k+1}(x) + D_x, ..., X_n(x) + D_x]$ of T_xM/D_x.

Lemma 4.1.2. *Let $1 \leq q \leq d$ be integers and b_1,..., $b_q \in \mathbb{R}^d$ be linearly independent vectors. Let also c_1,..., $c_q \in \mathbb{R}$. If $c = (c_1, ..., c_q) \in \mathbb{R}^q$ and $B = (\langle b_i, b_j \rangle)_{1 \leq i,j \leq q}$, then*

$$\min\{\|a\|^2 : a \in \mathbb{R}^d, \quad \langle a, b_j \rangle = c_j, \quad 1 \leq j \leq q\} = \langle B^{-1}(c), c \rangle.$$

Proof. If $g_j : \mathbb{R}^d \to \mathbb{R}$ is the function $g_j(a) = \langle a, b_j \rangle - c_j$, then $\nabla g_j(a) = b_j$ for every $x \in \mathbb{R}^d$, $1 \leq j \leq q$. Since b_1,..., b_q are assumed to be linearly independent, the minimum is attained at a solution of the system of equations

$$2a = \sum_{i=1}^{q} \lambda_i b_i,$$

$$\langle a, b_j \rangle = c_j, \quad 1 \leq j \leq q,$$

by Lagrange's theorem, where λ_1,..., λ_q are the Lagrange multipliers. Substituting we get the linear system

$$\frac{1}{2} \sum_{i=1}^{q} \lambda_i \langle b_i, b_j \rangle = c_j, \quad 1 \leq j \leq q,$$

and therefore

$$\lambda_i = 2 \sum_{j=1}^{q} \tilde{B}_{ij} c_j, \quad 1 \leq i \leq q,$$

where $B^{-1} = (\tilde{B}_{ij})_{1 \leq i,j \leq q}$. Note that B is symmetric and positive definite, in particular invertible, because b_1,..., b_q are linearly independent. It follows now that

$$a = \sum_{i=1}^{q} \Big(\sum_{j=1}^{q} \tilde{B}_{ij} c_j \Big) b_i$$

and so the minimum value we seek is

$$\|a\|^2 = \sum_{i,j,\rho,s=1}^{q} \tilde{B}_{ij}\tilde{B}_{\rho s}c_j c_s\langle b_i, b_\rho\rangle = \sum_{i,j,s=1}^{q} \tilde{B}_{ij}c_s c_j \sum_{\rho=1}^{q}\langle b_i, b_\rho\rangle \tilde{B}_{\rho s}$$

$$= \sum_{i,j=1}^{q} \tilde{B}_{ij}c_i c_j = \langle B^{-1}(c), c\rangle. \quad \square$$

From the preceding Lemma 4.1.2 we get that if $B = (\langle b^s, b^l\rangle)_{k+1\leq s,l\leq n}$, then B^{-1} is the Gram matrix of the inner product on T_xM/D_x with respect to the ordered basis $[X_{k+1}(x) + D_x, ..., X_n(x) + D_x]$ of T_xM/D_x. From its definition follows now that if $\{X_1^*, ..., X_k^*, X_{k+1}^*, ..., X_n^*\}$ is the dual basis, then Popp's volume is

$$\frac{1}{\sqrt{\det B}} X_1^* \wedge \cdots \wedge X_k^* \wedge X_{k+1}^* \wedge \cdots \wedge X_n^*.$$

This concludes the local description of Popp's volume in the case of an equiregular Sub-Riemannian structure of step 2.

The local formula for Popp's volume of a general equiregular Sub-Riemannian structure of step r and growth vector $(n_1, n_2, ..., n_r)$ is analogous. Let $\{X_1, X_2, ..., X_n\}$ be an adapted local frame on an open neighbourhood of a given point $x \in M$ of the associated flag $\mathcal{D} = \mathcal{D}^1 \subset \mathcal{D}^2 \subset \cdots \subset \mathcal{D}^r \subset \mathcal{TM}$. There exist locally defined smooth functions $b^l_{i_1 i_2\cdots i_k}$, $2 \leq k \leq r$, such that

$$[X_{i_1}, [X_{i_2}, [\cdots, [X_{i_{k-1}}, X_{i_k}]\cdots] + D_x^{k-1} = \sum_{l=n_{k-1}+1}^{k_k} b^l_{i_1 i_2\cdots i_k} X_l + D_x^{k-1}$$

evaluated at x, for $1 \leq i_1, ..., i_k \leq n_1$. Again we consider $b^l = (b^l_{i_1 i_2\cdots i_k})_{1\leq i_1,...,i_k\leq n_1}$ as a vector in $\mathbb{R}^{n_1^k}$. Recall the linear epimorphism $\pi_k : \bigotimes_{i=1}^{k} D_x^1 \to D_x^k/D_x^{k-1}$ with

$$\pi_k(X_{i_1} \otimes \cdots \otimes X_{i_k}) = [X_{i_1}, [X_{i_2}, [\cdots, [X_{i_{k-1}}, X_{i_k}]\cdots] + D_x^{k-1}.$$

The norm in D_x^k/D_x^{k-1} is induced by π_k and can be computed as before. More precisely, if $v + D_x^{k-1} = \sum_{l=n_{k-1}+1}^{n_k} c^l X_l + D_x^{k-1}$, then

$$\|v + D_x^{k-1}\| = \min\{\|a\| : \langle a, b^l\rangle = c^l, \quad n_{k-1} \leq l \leq n_k\} = \langle B_k^{-1}(c), c\rangle$$

where $c = (c_{n_{k-1}+1}, ..., c_{n_k})$, $B_k = (\langle b^i, b^j\rangle)_{n_{k-1}\leq i,j\leq n_k}$ and the inner products are euclidean, Thus, B_k^{-1} is the Gram matrix of the inner product in D_x^k/D_x^{k-1}. If $\{X_1^*, X_2^*, ..., X_n^*\}$ is the dual frame, then Popp's volume is locally

$$\left(\prod_{k=2}^{r} \det B_k\right)^{-1/2} X_1^* \wedge \cdots \wedge X_n^*.$$

4.2 Examples of Popp's volume

Let M be a contact 3-manifold with contact 1-form a and contact distribution $D = \mathrm{Ker} a$. There exists a unique smooth vector field Z on M such that $i_Z a = 1$ and $i_Z da = 0$, which is called the Reeb vector field of the contact structure. Recall that the contact Sub-Riemannian structure (with a chosen Riemannian metric $\langle \cdot, \cdot \rangle$ on D) satisfies the Lie bracket generating condition and is equiregular of step 2 with growth vector $(2, 3)$. If $\{X_1, X_2\}$ is a local orthonormal frame of D, then $[X_1, X_2, Z]$ is an adapted local frame, since Z is transverse to D. Corresponding to the Sub-Riemannian metric there is a skew-symmetric vector bundle morphism $J : D \to D$ defined by the property

$$da(v, w) = \langle J(v), w \rangle$$

for every v, $w \in D_x$, $x \in M$. The quantity

$$\|J\| = \left(\sum_{i,j=1}^{2} \langle X_i, J(X_j) \rangle^2 \right)^{1/2}$$

is known as the Hilbert-Schmidt norm of J.

For every $1 \leq i, j \leq 2$ there is a locally defined smooth function b_{ij} such that $[X_i, X_j](x) + D_x = b_{ij}(x) Z(x) + D_x$. According to the calculations of the previous section 4.1, Popp's volume is locally

$$\left(\sum_{i,j=1}^{2} b_{ij}^2 \right)^{-1/2} X_1^* \wedge X_2^* \wedge Z.$$

However,

$$\sum_{i,j=1}^{2} b_{ij}^2 = \sum_{i,j=1}^{2} (a([X_i, X_j]))^2 = \sum_{i,j=1}^{2} (da(X_i, X_j))^2 = \sum_{i,j=1}^{2} \langle X_i, J(X_j) \rangle^2 = \|J\|^2$$

and so Popp's volume is locally

$$\frac{1}{\|J\|} X_1^* \wedge X_2^* \wedge Z.$$

Example 4.2.1. The Heisenberg group $\mathcal{H}$ is a special case of a contact 3-manifold. Recall that the contact 1-form is the standard one $\omega = dz - \frac{1}{2}(xdy - ydx)$ whose kernel D is globally generated by the vector fields

$$X = \frac{\partial}{\partial x} - \frac{1}{2} y \frac{\partial}{\partial z}, \quad Y = \frac{\partial}{\partial y} + \frac{1}{2} x \frac{\partial}{\partial z}.$$

The metric on D is the one with respect to which $\{X, Y\}$ becomes an orthonormal frame. Here $Z = \frac{\partial}{\partial z}$ is the Reeb vector field and since $[X, Y] = Z$, it follows from the above that Popp's volume on the Heisenberg group is

$$(dx - \frac{1}{2} ydz) \wedge (dy + \frac{1}{2} xdz) \wedge dz = dx \wedge dy \wedge dz.$$

In other words, Popp's volume on the Heisenberg group $\mathcal{H}$ coincides with the 3-dimensional euclidean volume.

Example 4.2.2. The Martinet distribution D on $\mathbb{R}^3$ was defined in Example 2.3.4(b) and is the kernel of the differential 1-form $\omega = dz - y^2 dx$. It is globally generated by the vector fields

$$X = \frac{\partial}{\partial x} + y^2 \frac{\partial}{\partial z}, \quad Y = \frac{\partial}{\partial y}$$

and satisfies the Lie bracket generating condition but is not equiregular. The growth vector at points on the plane $y = 0$ is $(2, 2, 3)$. Its restriction on the subspace $M = \mathbb{R}^3 \setminus \{y = 0\}$ is equiregular with growth vector $(2, 3)$. On D we consider the restriction of the euclidean metric. Note that X and Y are orthogonal and

$$\left\{ \frac{1}{\sqrt{1+y^2}} X, Y, \frac{\partial}{\partial z} \right\}$$

is an adapted frame. Moreover,

$$\left[\frac{1}{\sqrt{1+y^2}} X, Y \right] + D = \frac{-2y}{\sqrt{1+y^2}} \frac{\partial}{\partial z} + D.$$

Therefore, Popp's volume on M is

$$\frac{1}{|2y|}(dx + y^2 dz) \wedge dy \wedge dz = \frac{1}{|2y|} dx \wedge dy \wedge dz.$$

The singularities of this form occur precisely on the plane $y = 0$, which consists of the non-regular points and the equiregularity of D fails.

4.3 Sub-Riemannian isometries and Popp's volume

Let M be a Sub-Riemannian manifold with corresponding horizontal distribution D satisfying the Lie bracket generating condition. A diffeomorphism $\phi : M \to M$ is called a Sub-Riemannian isometry if the following two conditions are satisfied:

(i) $\phi_{*x}(D_x) = D_{\phi(x)}$ for every $x \in M$ and

(ii) $\langle \phi_{*x}(v), \phi_{*x}(w) \rangle = \langle v, w \rangle$ for every v, $w \in D_x$ and $x \in M$.

Note that ϕ induces a map of sheaves $\phi_* : \mathcal{TM} \to \mathcal{TM}$ which makes the following diagram commutative

$$\begin{array}{ccc} \mathcal{TM} & \xrightarrow{\phi_*} & \mathcal{TM} \\ \downarrow & & \downarrow \\ M & \xrightarrow{\phi} & M \end{array}$$

and such that $\phi_*(\mathcal{D}^1) = \mathcal{D}^1$. Since ϕ_* preserves Lie brackets, we also have $\phi_*(\mathcal{D}^k) = \mathcal{D}^k$ for every $k \in \mathbb{N}$. Taking evaluations we conclude that $\phi_*(D^k_x) = D^k_{\phi(x)}$ for every $x \in M$, $k \in \mathbb{N}$.

It is obvious that the set of all Sub-Riemannian isometries of M is a subgroup of the group of diffeomorphisms of M.

Suppose now that the Sub-Riemannian structure of M is equiregular of step r and with growth vector $(n_1, n_2, ..., n_r)$. If $\phi : M \to M$ is a Sub-Riemannian isometry, then we have a linear isometry $\tilde{\phi}_{*x} : \bigotimes_{i=1}^{k} D_x \to \bigotimes_{i=1}^{k} D_{\phi(x)}$ for every $k \in \mathbb{N}$ and $x \in M$. Since ϕ_* preserves the flag

$$\mathcal{D} = \mathcal{D}^1 \subset \mathcal{D}^2 \subset \cdots \subset \mathcal{D}^r = \mathcal{TM},$$

it induces linear isomorphisms $\hat{\phi}_{*x} : D_x^k / D_x^{k-1} \to D_{\phi(x)}^k / D_{\phi(x)}^{k-1}$ for every $2 \le k \le r$ and $x \in M$. Actually, $\hat{\phi}_{*x}$ is a linear isometry of inner product vector spaces. To see this, we recall that the inner product on D_x^k / D_x^{k-1} is induced by the linear epimorphism $\pi_k : \bigotimes_{i=1}^{k} D_x \to D_x^k / D_x^{k-1}$ of section 4.1. Since ϕ_* preserves the Lie brackets, the following diagram commutes.

$$\begin{tikzcd}
\bigotimes_{i=1}^{k} D_x \arrow[r, "\tilde{\phi}_{*x}"] \arrow[d, "\pi_k"] & \bigotimes_{i=1}^{k} D_{\phi(x)} \arrow[d, "\pi_k"] \\
D_x^k / D_x^{k-1} \arrow[r, "\hat{\phi}_{*x}"] & D_{\phi(x)}^k / D_{\phi(x)}^{k-1}
\end{tikzcd}$$

If $v + D_x^{k-1} \in D_x^k / D_x^{k-1}$, we have

$$\|v + D_x^{k-1}\| = \min\{\|a\| : a \in \bigotimes_{i=1}^{k} D_x \quad \text{and} \quad \pi_k(a) = v + D_x^{k-1}\}$$

$$= \min\{\|\tilde{\phi}_{*x}(a)\| : a \in \bigotimes_{i=1}^{k} D_x \quad \text{and} \quad \pi_k(\tilde{\phi}_{*x}(a)) = \hat{\phi}_{*x}(v+D_x^{k-1})\} = \|\hat{\phi}_{*x}(v+D_x^{k-1})\|,$$

since $\tilde{\phi}_{*x}$ is a linear isometry and $\hat{\phi}_{*x}$ is a linear isomorphism.

Theorem 4.3.1. *The Sub-Riemannian isometries of an equiregular Sub-Riemannian orientable manifold M preserve Popp's volume. Moreover, if the group of the Sub-Riemannian isometries of M acts transitively on M, then Popp's volume is the unique volume element on M which is invariant by Sub-Riemannian isometries, up to multiplication by a scalar constant.*

Proof. It follows from the above that if $\phi : M \to M$ is a Sub-Riemannian isometry, then $\hat{\phi}_* : gr(T_xM) \to gr(T_{\phi(x)}M)$ is a linear isometry. The definition of Popp's volume implies now that it is preserved by ϕ. For the second assertion, let Ω be a volume form on M such that $\phi^*\Omega = \Omega$ for every Sub-Riemannian isometry ϕ. There exists a nowhere vanishing smooth function $f : M \to \mathbb{R}$ such that $f\Omega$ is Popp's volume. We have

$$f\Omega = \phi^*(f\Omega) = (f \circ \phi)\phi^*\Omega = (f \circ \phi)\Omega$$

and therefore $f = f \circ \phi$. The transitivity of the action of the Sub-Riemannian isometries implies that f must be constant. □

Example 4.3.2. As we saw in section 1.1 the left translations of the Heisenberg group $\mathcal{H}$ leaves invariant the two standard horizontal global smooth vector fields X, Y which generate its horizontal distribution and form an orthonormal frame, as well as $Z = [X, Y]$. It follows that the group of Sub-Riemannian isometries of $\mathcal{H}$ acts transitively on $\mathcal{H}$. Since Popp's volume on $\mathcal{H}$ is euclidean volume on $\mathbb{R}^3$, it follows that euclidean volume is invariant under left translations of $\mathcal{H}$. But it is easy to see that the euclidean volume is also invariant under the right translations of $\mathcal{H}$. This implies that $\mathcal{H}$ is a unimodular group and the euclidean volume is its Haar measure.

4.4 The Sub-Riemannian Laplacian

Let M be a Sub-Riemannian connected n-manifold whose Sub-Riemannian structure consists of a smooth real vector bundle $p : E \to M$ endowed with a Riemannian metric g, a smooth vector bundle morphism $f : E \to TM$ with horizontal distribution $D = \{D_x : x \in M\}$ and satisfies the Lie bracket generating condition. Let $\langle \cdot, \cdot \rangle$ be the induced Sub-Riemannian metric on D by g. If $\psi : M \to \mathbb{R}$ is a smooth function, then for every $x \in M$ there exists a unique vector $\nabla_h \psi(x) \in D_x$ such that

$$\psi_{*x}(v) = \langle \nabla_h \psi(x), v \rangle$$

for every $v \in D_x$. The vector field $\nabla_h \psi$ on M defined in this way is called the horizontal gradient of ψ. If $x \in M$ has an open neighbourhood on which the Sub-Riemannian structure is of constant rank k, there exists a horizontal orthhonormal local frame $\{X_1, X_2, ..., X_k\}$ on a possibly smaller open neighbourhood of x generating D and the horizontal gradient is locally given by the formula

$$\nabla_h \psi = \sum_{i=1}^{k} X_i(\psi) \cdot X_i.$$

It follows that if the Sub-Riemannian structure is of constant rank, the horizontal gradient of a smooth function is a smooth vector field.

Recall now that if M is orientable and Ω is a volume form on M, then for every smooth vector field X on M the Ω-divergence of X is defined to be the unique smooth function $\mathrm{div}_\Omega X : M \to \mathbb{R}$ such that $L_X \Omega = (\mathrm{div}_\Omega X) \cdot \Omega$. Note that $\mathrm{div}_\Omega X = \mathrm{div}_{-\Omega} X$. If we have a Sub-Riemannian structure of constant rank k, the Sub-Riemannian Laplacian associated to the volume form Ω of a smooth function $\psi : M \to \mathbb{R}$ is defined to be

$$\Delta_h \psi = \mathrm{div}_\Omega(\nabla_h \psi)$$

and is a second order differential operator. In terms of a generating horizontal orthonormal local frame $\{X_1, X_2, ..., X_k\}$ we have

$$\Delta_h \psi = \sum_{i=1}^{k} \mathrm{div}_\Omega(X_i(\psi) \cdot X_i) = \sum_{i=1}^{k} X_i(X_i(\psi)) + \sum_{i=1}^{k} (\mathrm{div}_\Omega X_i) \cdot X_i(\psi).$$

Thus, locally

$$\Delta_h = \sum_{i=1}^{k} X_i^2 + \sum_{i=1}^{k} (\mathrm{div}_\Omega X_i) \cdot X_i$$

as a second order differential operator. Therefore, only the first order terms of the Sub-Riemannian Laplacian depend on the choice of the volume form Ω.

If M is an orientable Sub-Riemannian manifold whose horizontal distribution has constant rank and Ω is a smooth volume form on M with associated Sub-Riemannian Laplacian Δ_h, then the Sub-Riemannian heat equation on M is the PDE

$$\frac{\partial}{\partial t}\phi(x,t) = \Delta_h \phi(t,x)$$

and governs the heat diffusion in the directions of the horizontal distribution.

Example 4.4.1. In the case of the Heisenberg group $\mathcal{H}$ we have the global generating orthonormal frame consisting of the vector fields

$$X = \frac{\partial}{\partial x} - \frac{1}{2} y \frac{\partial}{\partial z}, \quad Y = \frac{\partial}{\partial y} + \frac{1}{2} x \frac{\partial}{\partial z}.$$

According to Example 4.2.1, Popp's volume on $\mathcal{H}$ is the 3-dimensional euclidean volume. Therefore, X and Y have divergence zero and the Sub-Riemannian Laplacian on $\mathcal{H}$ is given by the formula

$$\Delta_h = X^2 + Y^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{1}{4}(x^2+y^2)\frac{\partial^2}{\partial z^2} - y\frac{\partial^2}{\partial x \partial z} + x\frac{\partial^2}{\partial y \partial z}.$$

Thus, the Sub-Riemannian heat equation in the Heisenberg group is

$$\frac{\partial \phi}{\partial t}(x,y,z,t) = \frac{\partial^2 \phi}{\partial x^2}(x,y,z,t) + \frac{\partial^2 \phi}{\partial y^2}(x,y,z,t) + \frac{1}{4}(x^2+y^2)\frac{\partial^2 \phi}{\partial z^2}(x,y,z,t)$$

$$-y\frac{\partial^2 \phi}{\partial x \partial z}(x,y,z,t) + x\frac{\partial^2 \phi}{\partial y \partial z}(x,y,z,t).$$

We assume now that our Sub-Riemannian structure of M of constant rank k is equiregular of step r and growth vector $(n_1, n_2, ..., n_r)$. Using the same notations as in section 4.1, let $\{X_1, X_2, ..., X_n\}$ be an adapted local frame. There exist locally defined smooth functions $b^l_{i_1 i_2 \cdots i_k}$, $2 \le k \le r$, such that

$$[X_{i_1}, [X_{i_2}, [\cdots, [X_{i_{k-1}}, X_{i_k}] \cdots] + D^{k-1} = \sum_{l=n_{k-1}+1}^{k_k} b^l_{i_1 i_2 \cdots i_k} X_l + D^{k-1}$$

for $1 \leq i_1, ..., i_k \leq n_1$. As in section 4.1, if we put $b^l = (b^l_{i_1 i_2 \cdots i_k})_{1 \leq i_1,...,i_k \leq n_1}$ and $B_k = (\langle b^i, b^j \rangle)_{n_{k-1} \leq i,j \leq n_k}$, where the inner products are euclidean, then Popp's volume is locally

$$\left(\prod_{j=2}^{r} \det B_j\right)^{-1/2} X_1^* \wedge \cdots \wedge X_n^*.$$

We shall find a local formula for the Sub-Riemannian Laplacian associated to Popp's volume P. This will be an immediate consequence of the local expression of divergence with respect to Popp's volume.

Recall that if ω is a differential 1-form and X, Y are smooth vector fields on M, then $X(\omega(Y)) = (L_X\omega)(Y) + \omega([X, Y])$. Therefore,

$$(L_{X_i} X_j^*)(X_s) = X_i(X_j^*(X_s)) - X_j^*([X_i, X_s]) = -X_j^*\left(\sum_{l=1}^{n} c_{is}^l X_l\right) = -c_{is}^j$$

where $[X_i, X_s] = \sum_{l=1}^{n} c_{is}^l X_l$. Hence,

$$L_{X_i} X_j^* = -\sum_{l=1}^{n} c_{il}^j X_l^*.$$

Since L_{X_i} is a derivation, we compute

$$L_{X_i}\left(\left(\prod_{j=2}^{r} \det B_j\right)^{-1/2} X_1^* \wedge \cdots \wedge X_n^*\right) = X_i\left(\left(\prod_{j=2}^{r} \det B_j\right)^{-1/2}\right) X_1^* \wedge \cdots \wedge X_n^*$$

$$+\left(\prod_{j=2}^{r} \det B_j\right)^{-1/2} \left(L_{X_i} X_1^* \wedge \cdots \wedge X_n^* + \cdots + X_1^* \wedge \cdots \wedge L_{X_i} X_n^*\right)$$

$$= \left[X_i\left(\left(\prod_{j=2}^{r} \det B_j\right)^{-1/2}\right) + \left(\prod_{j=2}^{r} \det B_j\right)^{-1/2} \left(-\sum_{j=1}^{n} c_{ij}^j\right)\right] X_1^* \wedge \cdots \wedge X_n^*.$$

This means that the divergence of X_i with respect to Popp's volume is

$$\mathrm{div}_P X_i = -\left(\prod_{j=2}^{r} \det B_j\right)^{-1/2} \cdot X_i\left(\left(\prod_{j=2}^{r} \det B_j\right)^{1/2}\right) - \sum_{j=1}^{n} c_{ij}^j.$$

But

$$\left(\prod_{j=2}^{r} \det B_j\right)^{-1/2} \cdot X_i\left(\left(\prod_{j=2}^{r} \det B_j\right)^{1/2}\right) = \frac{\displaystyle\sum_{j=2}^{r} (\det B_2) \cdots X_i(\det B_j) \cdots (\det B_r)}{\displaystyle 2\prod_{j=2}^{r} \det B_j}$$

$$= \frac{\sum_{j=2}^{r} (\det B_2) \cdots (\det B_j) \mathrm{Tr}(B_j^{-1} X_j(B_j)) \cdots (\det B_r)}{2 \prod_{j=2}^{r} \det B_j} = \frac{1}{2} \sum_{j=2}^{r} \mathrm{Tr}(B_j^{-1} X_j(B_j))$$

and so

$$\mathrm{div}_P X_i = -\frac{1}{2} \sum_{j=2}^{r} \mathrm{Tr}(B_j^{-1} X_j(B_j)) - \sum_{j=1}^{n} c_{ij}^j.$$

Substituting now we get the local expression

$$\Delta_h = \sum_{i=1}^{k} X_i^2 - \sum_{i=1}^{k} \left(\frac{1}{2} \sum_{j=2}^{r} \mathrm{Tr}(B_j^{-1} X_j(B_j)) + \sum_{j=1}^{n} c_{ij}^j \right) X_i$$

for the Sub-Riemannian Laplacian in the equiregular case of step r and rank k.

Bibliography

[1] A. Agrachev, D. Balibari, U. Boscain, A comprehensive introduction to Sub-Riemannian Geometry, Cambridge University Press, 2019.

[2] D. Balibari, L. Rizzi, A formula for Popp's volume in Sub-Riemannian Geometry, *Analysis and Geometry in Metric Spaces* 1 (2013), 42-57.

[3] A. Bressan, B. Piccoli, Introduction to the mathematical theory of control, AIMS Series on Applied Math., volume 2, Amer. Inst. Math. Series, 2007.

[4] F. Jean, Control of non-holonomic systems: From Sub-Riemannian Geometry to motion planning, Springer, 2014.

[5] W. Liu, H.J. Sussmann, Shortest paths for Sub-Riemannian metrics on rank-two distributions, Mem. Amer. Math. Soc. 118, 1995.

[6] R. Montgomery, A tour of Sub-Riemannian Geometries, their geodesics and applications, Math. Surveys and Monographs, volume 91, Amer. Math. Soc., 2002.

[7] L. Rifford, Sub-Riemannian Geometry and optimal transport, Springer, 2014.

[8] R.S. Strichartz, Sub-Riemannian Geometry, *J. Diff. Geom.* 24 (1986), 221-263. Corrections, *J. Diff. Geom.* 30 (1989), 595-596.