THE GEOMETRY OF SUB-RIEMANNIAN THREE-MANIFOLDS

KEENER HUGHEN

Nov 95

ABSTRACT. The local equivalence problem for sub-Riemannian structures on three-manifolds is solved. In the course of the solution, it is shown how to attach a canonical Riemannian metric and connection to the given sub-Riemannian metric and it is shown how all of the differential invariants of the sub-Riemannian structure can be calculated. The relation between the completeness of the sub-Riemannian metric, the associated Riemannian metric, and geodesic completeness is investigated, and an example is given of a manifold that is complete in the sub-Riemannian metric but not complete in the canonical associated Riemannian metric. It is shown that the Jacobi equations for sub-Riemannian geodesics can be interpreted as a scalar, fourth-order, self-adjoint linear operator along each geodesic. The influence of the differential invariants of the sub-Riemannian structure on the conjugate points is investigated, and the results are used to prove a Bonnet-Myers-type theorem for complete sub-Riemannian 3-manifolds.

CONTENTS

1991 *Mathematics Subject Classification.* 53A55, 53C10, 53C22, 58E25, 53A40.

Key words and phrases. sub-Riemannian geometry, contact 3-manifolds, G-structures, second variation.

This paper is in final form and no version of it will be submitted for publication elsewhere.

1. Introduction

Sub-Riemannian geometry is the study of a smooth manifold equipped with a positive definite inner product on a sub-bundle of the tangent bundle (see §1.1 below for the precise definition). When the sub-bundle is equal to the tangent bundle we have the case of Riemannian geometry. Sub-Riemannian geometry is the natural setting for control theory (Brockett [1], Hermann [12]). CR geometry can be thought of as a special case of sub-Riemannian geometry (Webster [23], Chern and Hamilton [5]). Also, classical isoperimetric problems on surfaces, such as that of Pappus (finding the shortest curve which encloses a given area) can be considered as problems in sub-Riemannian geometry.

We now outline the contents of this manuscript. In the remainder of this section we define what a sub-Riemannian structure on a manifold is and give some examples. We also show how to associate a Riemannian metric to a given sub-Riemannian manifold and we compare some of the properties of the manifold with the two metrics. In particular, we discuss the notion of completeness of a sub-Riemannian manifold and show by example that a complete sub-Riemannian manifold may have an associated Riemannian metric that is not complete.

In §2 we solve the local equivalence problem for sub-Riemannian metrics on three-manifolds. We obtain the differential invariants distinguishing sub-Riemannian structures and we interpret these invariants geometrically. We show how to attach a canonical Riemannian metric and connection to the sub-Riemannian three-manifold. We also compute all of the homogeneous examples.

In §3 we derive the sub-Riemannian geodesic equations and we compute the second variation of the length functional and obtain the Jacobi operator. We interpret this operator as a scalar, fourth-order self-adjoint linear operator along each geodesic. We show that the index of the Hessian of the length functional is equal to the number of conjugate points; in particular, a geodesic is not length-minimizing beyond its first conjugate point. We show how the differential invariants of the sub-Riemannian structure influence the conjugate points and use this to prove a Bonnet-Myers-type theorem for complete sub-Riemannian three-manifolds. We then define the exponential map and explore some of its properties; it is shown that this exponential map is never a local diffeomorphism. Finally, we compute the geodesics for some of the homogeneous manifolds.

In §4 we discuss possible generalizations to higher dimensions and the problems encountered there due to the presence of rigid curves and abnormal extremals.

1.1. Definitions and examples. Let M be a smooth manifold. A *distribution* D on M is a sub-bundle of the tangent bundle TM, and a *D-curve* on M is a smooth immersed curve $\gamma : [a,b] \to M$ tangent to D, i.e., $\dot\gamma(t) \in D_{\gamma(t)}$ for all $t \in [a,b]$. A distribution D is said to be *bracket generating* if for every $p \in M$ the sections of D near p together with all their commutators span the tangent space T_pM of M at p. This condition is equivalent to there being no completely integrable subsystem of the corresponding differential system $I = D^\perp$. In this case, by a well known theorem of Chow (see [10] for the smooth version of this theorem) there is a D-curve joining any two points of M. A *sub-Riemannian metric* is a smoothly varying positive definite inner product $\langle\ ,\ \rangle$ on D; in the special case where D is equal to the tangent bundle, $\langle\ ,\ \rangle$ gives a Riemannian metric.

A *sub-Riemannian manifold*, denoted by the triple $(M, D, \langle\ ,\ \rangle)$, is a smooth n-dimensional manifold M equipped with a sub-Riemannian metric $\langle\ ,\ \rangle$ on a bracket generating distribution D of rank $m > 0$. The length of a D-curve $\gamma : [a,b] \to M$ is defined to be

$$L(\gamma) = \int_a^b \sqrt{\langle \dot\gamma(t), \dot\gamma(t)\rangle}\, dt.$$

Because D is bracket generating we may endow M with a distance d: the distance $d(p,q)$ between any two points p and q of M is

$$d(p,q) = \inf_\gamma \left\{ L(\gamma) \mid \gamma \text{ is a } D\text{-curve joining } p \text{ to } q \right\}.$$

This distance d is often referred to as the *Carnot-Carathéodory distance*.

The prototypical example of a sub-Riemannian manifold is given by the following sub-Riemannian structure on the Heisenberg group H^3, the group of upper triangular 3×3 matrices

$$H^3 = \left\{ \begin{bmatrix} 1 & y & z + \frac{1}{2}xy \\ 0 & 1 & x \\ 0 & 0 & 1 \end{bmatrix} \quad : \quad x, y, z \in \mathbb{R} \right\} \cong \mathbb{R}^3.$$

We take the distribution D to be the kernel of the left-invariant one-form $\omega^3 = dz + \frac{1}{2}(xdy - ydx)$. This distribution is spanned by the vector fields $\frac{\partial}{\partial x} + \frac{y}{2}\frac{\partial}{\partial z}$ and $\frac{\partial}{\partial y} - \frac{x}{2}\frac{\partial}{\partial z}$, and we take the inner product $\langle\ ,\ \rangle$ on D to be the one for which these two vector fields are everywhere orthonormal.

Let $\gamma(t) = (x(t), y(t), z(t))$, $a \le t \le b$ be a D-curve with $(x(a), y(a)) = (0,0)$. The length of the tangent vector $\dot\gamma$ is equal to $\dot x^2 + \dot y^2$ and so the length of γ is equal to the length of the projection of γ in the plane. The difference $z(b) - z(a)$ in the z-coordinate of γ is equal to the integral of one-half of $ydx - xdy$, and by Green's Theorem this is equal to the algebraic area enclosed by the curve (x,y) and the line segment connecting the endpoints of (x,y).

We see then that the D-curves that realize the distance $d(p,q)$ between two points $p, q \in H^3$ are the lifts of curves in the plane that minimize the length subject to the constraint that they "enclose" a fixed area, in the sense discussed above. This is the classical problem of Pappus, and it is well known that such curves are given by straight lines and circular arcs. Thus the "sub-Riemannian geodesics" of this example are straight lines in the plane and certain helices.

Another example of a sub-Riemannian manifold is given by taking M to be $\mathbb{R}^2 \times S^1$ with the distribution $D = \ker(\sin\phi\, dx - \cos\phi\, dy)$, where x and y are the coordinates on $\mathbb{R}^2$ and ϕ is the coordinate on S^1. This distribution is spanned by the vector fields $\cos\phi\,\partial/\partial x+\sin\phi\,\partial/\partial y$ and $\partial/\partial\phi$, and we take the inner product $\langle\ ,\ \rangle$ on D to be the one for which these two vectors are everywhere orthonormal.

Every D-curve $\gamma(t)=(x(t),y(t),\phi(t))$ with $\gamma^*(\cos\phi\,dx+\sin\phi\,dy)\neq 0$ is the lift of a regular curve $\alpha(t)=(x(t),y(t))$ in the plane whose tangent vector $\dot\alpha(t)$ forms the angle $\phi(t)$ with the x-axis, i.e.,

$$\dot\alpha(t)=v(t)\cos\phi(t)\frac{\partial}{\partial x}+v(t)\sin\phi(t)\frac{\partial}{\partial y},$$

where $v(t)$ is the speed of $\alpha(t)$. Conversely, every regular curve α in the plane may be lifted to a D-curve $\gamma(t)=(x(t),y(t),\phi(t))$ by setting $\phi(t)$ equal to the angle between $\dot\alpha(t)$ and the x-axis. The tangent vector $\dot\gamma(t)$ of the D-curve γ has squared length

$$\langle\dot\gamma(t),\dot\gamma(t)\rangle=v^2(t)+\dot\phi^2(t)=v^2(t)(1+\left(\frac{\dot\phi(t)}{v(t)}\right)^2)=v^2(t)(1+\kappa^2(t))$$

where $\kappa(t)$ is the curvature of α, and so the length of γ is equal to the integral of $\sqrt{1+\kappa^2(t)}\,v(t)$ along α.

Thus the D-curves with $\cos\phi\,dx+\sin\phi\,dy\neq 0$ that realize the distance between two points (x_0,y_0,ϕ_0) and (x_1,y_1,ϕ_1) of M are the lifts of curves α in the plane joining (x_0,y_0) to (x_1,y_1) with initial angle ϕ_0 and final angle ϕ_1 that minimize the functional

$$L(\kappa)=\int\sqrt{1+\kappa^2}\,ds$$

among all such curves in the plane.

As another example, consider $M=\mathbb{R}^4$ with coordinates (x,y,z,w) and let D be the bracket generating Engel distribution spanned by the two vector fields $\partial/\partial x$ and $\partial/\partial y+x\,\partial/\partial z+z\,\partial/\partial w$. Define the sub-Riemannian metric $\langle\ ,\ \rangle$ by declaring these vector fields to be everywhere orthonormal.

Let $\gamma_0(t)$ be the D-curve described by $(x(t),y(t),z(t),w(t))=(t,0,0,0)$, where, say, $a\le t\le b$. The length of $\gamma_0(t)$ is equal to $b-a$. Let $\gamma(t)=(x(t),y(t),z(t),w(t))$ be any other D-curve joining $(a,0,0,0)$ to $(b,0,0,0)$. The derivative of γ is

$$\dot\gamma=\dot x\frac{\partial}{\partial x}+\dot y\left(\frac{\partial}{\partial y}+\frac{\dot z}{\dot y}\frac{\partial}{\partial z}+\frac{\dot w}{\dot y}\frac{\partial}{\partial w}\right),$$

and so γ has length

$$L(\gamma)=\int_a^b\sqrt{\dot x^2+\dot y^2}\,dt\ge\int_a^b|\dot x|\,dt\ge b-a=L(\gamma_0).$$

Therefore γ_0 is a length-minimizing curve.

It can be shown [3] that γ_0 is *rigid*, i.e., it has a C^1-neighborhood $\mathcal{U}$ in the space of D-curves that join $(a,0,0,0)$ to $(b,0,0,0)$ with the property that every other D-curve in $\mathcal{U}$ is just a reparametrization of γ_0. Rigid curves need not satisfy the geodesic equations; see [15] and [18] and §4.1 below for discussions on this issue.

1.2. Metric space properties. The sub-Riemannian manifold $(M, D, \langle\ ,\ \rangle)$ with the Carnot-Carathéodory distance d is a metric space (M, d). We discuss some of the properties of this metric space below. We also show how to associate a Riemannian metric to the sub-Riemannian structure and we show by example that a complete sub-Riemannian manifold may have an associated Riemannian metric that is not complete.

If the rank of D is strictly less than the dimension of M, then the Hausdorff dimension of (M, d) is strictly larger than the dimension of M as manifold. More precisely, Mitchell [16] shows the Hausdorff dimension of (M, d) is equal to the sum $\sum_i i(\dim(D_i) - \dim(D_{i-1}))$, where D_i denotes the span of the sections of D together with their commutators of order less than or equal to i. For example, if D is a contact distribution on a 3-manifold M then the Hausdorff dimension of $(M, D, \langle\ ,\ \rangle)$ is equal to 4.

Suppose $\gamma : [a, b] \to M$ is any continuous curve, not necessarily tangent to D. The arc length $L_A(\gamma)$ of γ is the supremum over all partitions of $[a, b]$ of the sum $\sum_k d(\gamma(t_k), \gamma(t_{k+1}))$. We define the distance $d_A(p, q)$ between any two points p and q of M to be the infimum of all arc lengths of continuous curves joining p to q. It is not too hard to show that the metric d_A is equal to the Carnot-Carathéodory metric d. Therefore (M, d) is a "length space" in the sense of Gromov [8]. Consequently, every point p in M has a neighborhood U with the property that for every $q \in U$ there is a continuous curve γ joining p to q with arc length equal to $d(p, q)$. In fact, this curve may be assumed to be Lipschitz, and so its derivative exists almost everywhere, and where it exists, $\dot{\gamma} \in D$ (see Strichartz [21]). Furthermore, if the distribution D satisfies the *strong bracket generating hypothesis*, i.e., if TM is generated by D and $[X, D]$ for every nonzero local section X of D, then Strichartz shows, by using a theorem of Pontryagin, that this Lipschitz curve γ is actually smooth. Therefore in the strong bracket generating case, any two points within a sufficiently small neighborhood U may be joined by a length-minimizing D-curve. The only requirement for this neighborhood U is that it be small enough so that its closure is compact. If (M, d) has the property that every neighborhood has compact closure, then the same argument would show that every two points in M may be joined by a length-minimizing D-curve.

The sub-Riemannian manifold $(M, D, \langle\ ,\ \rangle)$ will be said to be *complete* if M is complete with respect to the metric d. The next proposition gives useful equivalent criterion for completeness. Its proof follows from general topology and will be omitted.

Proposition 1.1. *The following are equivalent:*

1. $(M, D, \langle\ ,\ \rangle)$ *is complete.*
2. *The closed balls are compact.*
3. *There is a nested sequence* $\{K_n\}$ *of compact sets with* $M = \bigcup K_n$ *such that if* $q_n \notin K_n$*, then* $d(p, q_n) \to \infty$*.*
4. *Every* D*-curve that leaves every compact set has infinite length.*

Corollary 1.1. *If* $(M, D, \langle\ ,\ \rangle)$ *is complete and* D *satisfies the strong bracket generating hypothesis, then every two points in* M *may be joined by a length-minimizing* D*-curve.*

Given a sub-Riemannian manifold $(M, D, \langle\ ,\ \rangle)$, it is always possible to obtain a Riemannian metric $\langle\ ,\ \rangle_R$ on M by choosing a sub-bundle $E \subset TM$ complementary to D, putting a positive definite inner product on E, and declaring E to be orthogonal to D. We will say a Riemannian metric $\langle\ ,\ \rangle_R$ is *associated* to $(M, D, \langle\ ,\ \rangle)$ if $\langle V, W\rangle_R = \langle V, W\rangle$ for all $V, W \in D$. If γ is a D-curve, then $\langle\dot\gamma, \dot\gamma\rangle_R = \langle\dot\gamma, \dot\gamma\rangle$, and hence the Riemannian length of γ is equal to the sub-Riemannian length of γ. It follows that the Riemannian distance is no greater than the Carnot-Carathéodory distance. Therefore, if $(M, D, \langle\ ,\ \rangle)$ has a complete associated Riemannian metric, then $(M, D, \langle\ ,\ \rangle)$ is complete.

Proposition 1.2. *There exists a complete sub-Riemannian manifold that has a non-complete associated Riemannian metric.*

Proof. We will construct an example. Consider $M = \mathbb{R}^3$ with the sub-Riemannian structure defined by

$$D = \ker(dz - \frac{1}{2}r^2 d\theta)$$
$$\langle\ ,\ \rangle = \frac{1}{1+z^2}(dr^2 + r^2 d\theta^2),$$

where (r, θ, z) are cylindrical coordinates on $\mathbb{R}^3$. Set $\omega = dz - \frac{1}{2}r^2 d\theta$. Observe that $\omega \wedge d\omega = -r\, dz \wedge dr \wedge d\theta \neq 0$; therefore D is a contact distribution and in particular is bracket generating.

The associated Riemannian metric $\langle\ ,\ \rangle + \omega^2$ is not complete: Let γ_0 be the curve described by $(r, \theta, z) = (0, 0, t)$. Its length in the metric $\langle\ ,\ \rangle + \omega^2$ is equal to the integral over the positive t-axis of $1/(1+t^2)$. Therefore γ_0 is a curve that leaves every compact set but has finite length.

On the other hand, we claim that $\langle\ ,\ \rangle$ is complete. First observe that if $\gamma(t) = (r(t), \theta(t), z(t))$ is a D-curve that leaves every compact set but stays bounded in z, i.e., $|z(t)| \leq z_0$, then the length of γ is

$$\begin{aligned} L(\gamma) &= \lim_{h\to\infty} \int_0^h \frac{1}{\sqrt{1+z^2(t)}} \sqrt{\dot r^2(t) + r^2(t)\dot\theta^2(t)}\,dt \\ &\geq \frac{1}{\sqrt{1+z_0^2}} \lim_{h\to\infty} \int_0^h |\dot r(t)|\,dt \\ &\geq \frac{1}{\sqrt{1+z_0^2}} \lim_{h\to\infty} (r(h) - r(0)). \end{aligned}$$

Because γ leaves every compact set, $\lim_{h\to\infty} r(h) = \infty$, and therefore γ has infinite length.

To show that the D-curves that escape every z-bound have infinite length, we will find another associated Riemannian metric $\langle\ ,\ \rangle_R$ for which $\langle\ ,\ \rangle_R \geq dz^2$. Set

$$\langle\ ,\ \rangle_R = \langle\ ,\ \rangle + \frac{\omega}{1+z^2}\left(\mu_1 dr + \mu_2 d\theta + (1+z^2)dz\right),$$

where μ_1 and μ_2 are the continuous functions

$$\mu_1 = \begin{cases} 0 & \text{if } z^2 \leq \frac{4}{r^2} - 1, \\ 2\sqrt{z^2 - \frac{4}{r^2} + 1} & \text{if } z^2 \geq \frac{4}{r^2} - 1 \end{cases}$$

and

$$\mu_2 = \begin{cases} \frac{1}{2}r^2(1+z^2) & \text{if } z^2 \leq \frac{4}{r^2} - 1, \\ 4 - \frac{1}{2}r^2(1+z^2) & \text{if } z^2 \geq \frac{4}{r^2} - 1. \end{cases}$$

The reader may easily verify that the quadratic form $\langle\ ,\ \rangle_R - dz^2$ is non-negative.

Therefore, every curve that escapes every z-bound has infinite length with respect to $\langle\ ,\ \rangle_R$, and thus $(M, D, \langle\ ,\ \rangle)$ is complete. $\square$

2. Local Equivalence of Sub-Riemannian Structures

In this section, we solve the local equivalence problem for sub-Riemannian metrics on three-manifolds. We show that sub-Riemannian structures on three-manifolds locally depend on two functions φ_1 and K of three variables and we investigate how these differential invariants influence the geometry. We show for example that if φ_1 vanishes identically, then the three-manifold naturally fibers over a surface with Gauss curvature K. We show how to attach a canonical Riemannian metric and connection to the sub-Riemannian three-manifold. We also compute and classify all of the homogeneous examples.

2.1. The G_0-structure. We define the G-structure that completely characterizes the sub-Riemannian structure.

Given an n-dimensional manifold M, every local coframing $\eta = (\eta^1, \dots, \eta^n)$ on $U \subset M$ determines a sub-Riemannian structure $(D, \langle\ ,\ \rangle)$ on U by setting $D = \{\eta^{m+1}, \dots, \eta^n\}^\perp$ and $\langle\ ,\ \rangle = (\eta^1)^2 + \cdots + (\eta^m)^2|_D$. Conversely, given a sub-Riemannian structure $(D, \langle\ ,\ \rangle)$ on U, we can always choose a local coframing η that satisfies $D = \{\eta^{m+1}, \dots, \eta^n\}^\perp$ and $\langle\ ,\ \rangle = (\eta^1)^2 + \cdots + (\eta^m)^2|_D$. Now such a choice of coframing η is not unique, for D determines $\eta^{m+1}, \dots, \eta^n$ only up to a $Gl(n-m, \mathbb{R})$ action, the quadratic form $(\eta^1)^2 + \cdots + (\eta^m)^2$ determines $\eta^1, \dots, \eta^m$ only up to an $O(m)$ action and, furthermore, since $\langle\ ,\ \rangle = (\eta^1)^2 + \cdots + (\eta^m)^2|_D$, we may add arbitrary multiples of the forms $\eta^{m+1}, \dots, \eta^n$ to each η^i for $1 \le i \le m$.

Let us say a coframing $\eta = (\eta^1, \dots, \eta^n)$ is *0-adapted* to $(M, D, \langle\ ,\ \rangle)$ if $D = \{\eta^{m+1}, \dots, \eta^n\}^\perp$ and $\langle\ ,\ \rangle = (\eta^1)^2 + \cdots + (\eta^m)^2|_D$. The set of 0-adapted coframes of $(M, D, \langle\ ,\ \rangle)$ forms a G_0-structure $\mathcal{B}_0 \to M$, where the structure group is

$$G_0 = \left\{ \begin{pmatrix} A & B \\ 0 & C \end{pmatrix} \ :\ A \in O(m),\ B \in \mathcal{M}(m, n-m),\ C \in Gl(n-m, \mathbb{R}) \right\}.$$

Now specifying the sub-Riemannian structure $(D, \langle\ ,\ \rangle)$ on M is equivalent to specifying the G_0-structure. In particular, sub-Riemannian structures are locally equivalent if and only if their corresponding G_0-structures are locally equivalent. We will solve this equivalence problem using the equivalence method of Cartan (see [7] for a nice description of this method).

Given a G-structure $\mathcal{B} \to M$, let R_g denote the right action of G on $\mathcal{B}$ defined by $R_g\eta = g^{-1}\eta$, and let ω denote the $\mathbb{R}^n$-valued tautological one-form on $\mathcal{B}$ (in terms of a local section η of $\mathcal{B}$ we have $\omega = g^{-1}\eta$). Note that ω is semi-basic, and that $R_g^*\omega = g^{-1}\omega$. A pseudo-connection is given by a $\mathfrak{g}$-valued one-form Θ on $\mathcal{B}$ whose restriction to the fiber is the Maurer-Cartan form. In terms of Θ, the exterior derivative of ω can always be expressed in the form

$$d\omega = -\Theta \wedge \omega + T\omega \wedge \omega,$$

where T is the torsion associated to Θ.

2.2. Reduction of the G_0-structure for the 3-dimensional case. For sub-Riemannian three-manifolds, we show how to reduce the structure group to the group $O(2)$. As a consequence, we see that two sub-Riemannian structures on a three-manifold M^3 are equivalent if and only if a canonical coframing on a certain S^1-bundle $\mathcal{B}_2 \to M^3$ is preserved.

If $(M^3, D, \langle\ ,\ \rangle)$ is a 3-dimensional sub-Riemannian manifold, then the bracket generating distribution D is necessarily a contact distribution, and so every 0-adapted coframing $\eta = (\eta^1, \eta^2, \eta^3)$ satisfies $\eta^3 \wedge d\eta^3 \neq 0$. The structure group for the 0-adapted coframes is

$$G_0 = \left\{ \begin{bmatrix} A & b \\ 0 & c \end{bmatrix} \ :\ A \in O(2),\ \ b \in \mathbb{R}^2,\ \ c \in \mathbb{R} - \{0\} \right\},$$

and the structure equations $d\omega = -\Theta \wedge \omega + T\omega \wedge \omega$ are in this case

$$d \begin{pmatrix} \omega^1 \\ \omega^2 \\ \omega^3 \end{pmatrix} = - \begin{pmatrix} 0 & \alpha & \beta \\ -\alpha & 0 & \gamma \\ 0 & 0 & \delta \end{pmatrix} \wedge \begin{pmatrix} \omega^1 \\ \omega^2 \\ \omega^3 \end{pmatrix} + \begin{pmatrix} T^1_{23} & T^1_{31} & T^1_{12} \\ T^2_{23} & T^2_{31} & T^2_{12} \\ T^3_{23} & T^3_{31} & T^3_{12} \end{pmatrix} \begin{pmatrix} \omega^2\wedge\omega^3 \\ \omega^3\wedge\omega^1 \\ \omega^1\wedge\omega^2 \end{pmatrix}.$$

We can modify β, γ, and δ so as to absorb the first two columns of T, and we can modify α so as to absorb the top two entries T^1_{12} and T^2_{12} of the last column of T. Of course, T^3_{12} cannot possibly be absorbed since D is a contact distribution. Thus the structure equations can be written

$$d \begin{pmatrix} \omega^1 \\ \omega^2 \\ \omega^3 \end{pmatrix} = - \begin{pmatrix} 0 & \alpha & \beta \\ -\alpha & 0 & \gamma \\ 0 & 0 & \delta \end{pmatrix} \wedge \begin{pmatrix} \omega^1 \\ \omega^2 \\ \omega^3 \end{pmatrix} + \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & T^3_{12} \end{pmatrix} \begin{pmatrix} \omega^2\wedge\omega^3 \\ \omega^3\wedge\omega^1 \\ \omega^1\wedge\omega^2 \end{pmatrix}.$$

The choice of pseudo-connection

$$\Theta = \begin{pmatrix} 0 & \alpha & \beta \\ -\alpha & 0 & \gamma \\ 0 & 0 & \delta \end{pmatrix}$$

for which the structure equations are of the above form is not unique, for we may add arbitrary multiples of ω^3 to β, γ, and δ.

The equation $R^*_g\omega = g^{-1}\omega$ implies $R^*_g\omega^3 = c^{-1}\omega^3$, and taking the exterior derivative of both sides gives

$$-c^{-1}\delta \wedge \omega^3 + c^{-1}T^3_{12}\,\omega^1 \wedge \omega^2 = -c^{-1}R^*_g\delta \wedge \omega^3 + R^*_g T^3_{12}(\det A^{-1}\omega^1 \wedge \omega^2 + \Phi \wedge \omega^3),$$

where Φ is a non-zero combination of ω^1 and ω^2. In particular,

$$R^*_g T^3_{12} = c^{-1} \det A\ T^3_{12}.$$

By considering the set $\mathcal{B}_1$ of coframes $\eta \in \mathcal{B}_0$ satisfying $T^3_{12}(\eta) = 1$, we may reduce to the G_1-structure $\mathcal{B}_1 \to M^3$ where

$$G_1 = \left\{ \begin{bmatrix} A & b \\ 0 & \det A \end{bmatrix} \ :\ A \in O(2), b \in \mathbb{R}^2 \right\} \subset G_0.$$

Restricting to $\mathcal{B}_1$, the structure equations become

$$d\begin{pmatrix}\omega^1\\ \omega^2\\ \omega^3\end{pmatrix} = -\begin{pmatrix}0 & \alpha & \beta\\ -\alpha & 0 & \gamma\\ 0 & 0 & 0\end{pmatrix}\wedge\begin{pmatrix}\omega^1\\ \omega^2\\ \omega^3\end{pmatrix} + \begin{pmatrix}0 & 0 & 0\\ 0 & 0 & 0\\ T^3_{23} & T^3_{31} & 1\end{pmatrix}\begin{pmatrix}\omega^2\wedge\omega^3\\ \omega^3\wedge\omega^1\\ \omega^1\wedge\omega^2\end{pmatrix}.$$

Again, we may add arbitrary multiples of ω^3 to β and γ, so the choice of pseudo-connection is not unique.

If we now differentiate the equation $R_g^*\omega^3 = \det A^{-1}\omega^3$, we find that the induced action of G_1 on the torsion space $\{(T^3_{23}, T^3_{31})\}$ is

$$R_g^*\begin{pmatrix}T^3_{23}\\ T^3_{31}\end{pmatrix} = A^{-1}\left(\frac{1}{\det A}\begin{pmatrix}T^3_{23}\\ T^3_{31}\end{pmatrix} - b\right).$$

In particular, G_1 acts transitively on the reduced torsion space, thus we may consider the G_2-structure $\mathcal{B}_2 \to M^3$ where

$$\mathcal{B}_2 = \{\eta \in B_1 \mid T^3_{23}(\eta) = T^3_{31}(\eta) = 0\}$$

and

$$G_2 = \left\{\begin{bmatrix}A & 0\\ 0 & \det A\end{bmatrix} \;:\; A \in O(2)\right\}.$$

The local sections of the bundle $\mathcal{B}_2 \to M^3$ are the local coframings $\eta = (\eta^1, \eta^2, \eta^3)$ of M^3 that satisfy $d\eta^3 = \eta^1\wedge\eta^2$ and for which the sub-Riemannian structure is given by $D = (\eta^3)^\perp$, $\langle\, ,\,\rangle = (\eta^1)^2 + (\eta^2)^2|_D$. We will say such a coframing is *2-adapted* to the sub-Riemannian structure.

Restricting to $\mathcal{B}_2$, the structure equations are now

$$d\begin{pmatrix}\omega^1\\ \omega^2\\ \omega^3\end{pmatrix} = -\begin{pmatrix}0 & \alpha & 0\\ -\alpha & 0 & 0\\ 0 & 0 & 0\end{pmatrix}\wedge\begin{pmatrix}\omega^1\\ \omega^2\\ \omega^3\end{pmatrix} + \begin{pmatrix}T^1_{23} & T^1_{31} & 0\\ T^2_{23} & T^2_{31} & 0\\ 0 & 0 & 1\end{pmatrix}\begin{pmatrix}\omega^2\wedge\omega^3\\ \omega^3\wedge\omega^1\\ \omega^1\wedge\omega^2\end{pmatrix}.$$

By replacing α with $\alpha + 1/2(T^2_{31} + T^1_{23})\omega^3$, we may assume that $T^1_{23} = -T^2_{31}$. Furthermore, differentiation of $d\omega^3 = \omega^1\wedge\omega^2$ shows that $(T^1_{31} - T^2_{23})\,\omega^1\wedge\omega^2\wedge\omega^3 = 0$.

Thus the structure equations can be written

$$d\begin{pmatrix}\omega^1\\ \omega^2\\ \omega^3\end{pmatrix} = -\begin{pmatrix}0 & \alpha & 0\\ -\alpha & 0 & 0\\ 0 & 0 & 0\end{pmatrix}\wedge\begin{pmatrix}\omega^1\\ \omega^2\\ \omega^3\end{pmatrix} + \begin{pmatrix}a_1 & a_2 & 0\\ a_2 & -a_1 & 0\\ 0 & 0 & 1\end{pmatrix}\begin{pmatrix}\omega^2\wedge\omega^3\\ \omega^3\wedge\omega^1\\ \omega^1\wedge\omega^2\end{pmatrix}, \tag{1}$$

where we have set $a_1 = T^1_{23} = -T^2_{31}$ and $a_2 = T^1_{31} = T^2_{23}$. Now the choice of pseudo-connection is unique, for $\Theta\wedge\omega = 0$ if and only if $\alpha = 0$.

Let A_i, B_i be the covariant derivatives in the ω^i direction of a_1, a_2, respectively. By differentiating the structure equations (1), we find

$$da_1 = -2a_2\alpha + \sum_{i=1}^{3} A_i\omega^i, \tag{2}$$

$$da_2 = 2a_1\alpha + \sum_{i=1}^{3} B_i\omega^i, \tag{3}$$

$$d\alpha = (A_2 - B_1)\omega^2\wedge\omega^3 + (A_1 + B_2)\omega^3\wedge\omega^1 + K\omega^1\wedge\omega^2. \tag{4}$$

Because the choice of pseudo-connection Θ is unique, the automorphisms of the G_2-structure $\mathcal{B}_2 \to M^3$ must preserve the coframing $(\omega^1, \omega^2, \omega^3, \alpha)$ on $\mathcal{B}_2$. In particular, every automorphism must preserve the functions a_1, a_2, K, and all of their covariant derivatives. It follows from the general theory of $\{e\}$-structures that a_1, a_2 and K form a complete set of differential invariants of the G_2-structure $\mathcal{B}_2 \to M^3$.

Note that if a_1 and a_2 are both identically zero, then K is the only invariant. If a_1 and a_2 are not both identically zero, then the G_2-structure may be reduced to a discrete G-structure: It is not too hard to show that G_2 acts by rotation on the torsion space $\{(a_1, a_2)\}$, thus we may consider the subset $\mathcal{B}_3 \subset \mathcal{B}_2$ of coframes for which a_2 is zero and a_1 is positive. The stabilizer group $G_3 \subset G_2$ of $\mathcal{B}_3$ is isomorphic to $\mathbb{Z}_2 \times \mathbb{Z}_2$. Restricting to $\mathcal{B}_3$, the forms ω^1, ω^2, and ω^3 are now basic and α is some combination $\alpha = \sum \lambda_i\omega^i$, thus the structure equations for the G_3-structure $\mathcal{B}_3 \to M^3$ are

$$d\begin{pmatrix}\omega^1\\ \omega^2\\ \omega^3\end{pmatrix} = \begin{pmatrix}\lambda_3 + \varphi_1 & 0 & -\lambda_1\\ 0 & \lambda_3 - \varphi_1 & -\lambda_2\\ 0 & 0 & 1\end{pmatrix}\begin{pmatrix}\omega^2\wedge\omega^3\\ \omega^3\wedge\omega^1\\ \omega^1\wedge\omega^2\end{pmatrix}, \tag{5}$$

where $\varphi_1 = \sqrt{a_1^2 + a_2^2}$.

2.3. The homogeneous manifolds. The homogeneous sub-Riemannian manifolds are those for which the the group of automorphisms of the reduced G-structure acts transitively. We now classify all of the homogeneous examples of sub-Riemannian three-manifolds.

The group of automorphisms of the G_2-structure $\mathcal{B}_2 \to M^3$ are identified with the group of diffeomorphisms of $\mathcal{B}_2$ that preserve the coframing $(\omega^1, \omega^2, \omega^3, \alpha)$. This group has dimension less than or equal to four with equality if and only if the group acts transitively on $\mathcal{B}_2$. In this case a_1, a_2, and K are constant; equations (2) and (3) then imply that a_1 and a_2 are both zero. Thus the homogeneous manifolds with the largest group of symmetries are those for which a_1 and a_2 are both zero. In this case, $K \equiv K_0$ is the only invariant.

There are three distinct homogeneous examples with a_1 and a_2 both identically zero corresponding to the sign of K_0: If $K_0 > 0$, then locally $(M^3, D, \langle\ ,\ \rangle)$ is equivalent to $SO(3)$; if $K_0 = 0$, then locally $(M^3, D, \langle\ ,\ \rangle)$ is equivalent to the Heisenberg group; and if $K_0 < 0$, then locally $(M^3, D, \langle\ ,\ \rangle)$ is equivalent to $Sl(2, \mathbb{R})$. In each case, $D = (\eta^3)^\perp$ and $\langle\ ,\ \rangle = (\eta^1)^2 + (\eta^2)^2|_D$, where (η^1, η^2, η^3) is the standard left invariant coframing satisfying

$$d\begin{pmatrix} \eta^1 \\ \eta^2 \\ \eta^3 \end{pmatrix} = \begin{pmatrix} K_0 & 0 & 0 \\ 0 & K_0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \eta^2 \wedge \eta^3 \\ \eta^3 \wedge \eta^1 \\ \eta^1 \wedge \eta^2 \end{pmatrix}.$$

The homogeneous examples which have a_1 and a_2 not both identically zero are those for which the group of automorphisms of the G_3-structure $\mathcal{B}_3 \to M^3$ acts transitively; in this case, the symmetry group has dimension three and the functions λ_i and φ_1 are constant. Differentiating the structure equations (5) we find that if λ_i and φ_1 are constant then $\lambda_1(\lambda_3-\varphi_1) = \lambda_2(\lambda_3+\varphi_1) = 0$. Thus the homogeneous manifolds with a_1 and a_2 not both identically zero can be classified according to three cases: case 1 has $\lambda_1 = \lambda_2 = 0$, case 2 has $\lambda_2 = 0, \lambda_3 = \varphi_1$, and case 3 has $\lambda_1 = 0, \lambda_3 = -\varphi_1$. To compute these examples, we compute the Lie algebras corresponding to the different cases. We assume that φ_1 is positive.

<u>case 1</u>: $\lambda_1 = \lambda_2 = 0$.

Here the structure equations are

$$d\begin{pmatrix} \omega^1 \\ \omega^2 \\ \omega^3 \end{pmatrix} = \begin{pmatrix} \lambda_3+\varphi_1 & 0 & 0 \\ 0 & \lambda_3-\varphi_1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \omega^2\wedge\omega^3 \\ \omega^3\wedge\omega^1 \\ \omega^1\wedge\omega^2 \end{pmatrix},$$

and so the Lie algebra $\mathfrak{g}$ of M is determined by the signs of the diagonal entries of the matrix C of the structure constants

$$C = \begin{pmatrix} \lambda_3+\varphi_1 & 0 & 0 \\ 0 & \lambda_3-\varphi_1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

a) If $\lambda_3 > \varphi_1$, so that the diagonal entries are all positive, then $\mathfrak{g} = \mathfrak{so}(3)$. Here the manifold is $SO(3)$ with the sub-Riemannian structure $D = (\eta^3)^\perp$ and $\langle\ ,\ \rangle = \frac{1}{\lambda_3-\varphi_1}(\eta^1)^2 + \frac{1}{\lambda_3+\varphi_1}(\eta^2)^2\Big|_D$, where η is the standard coframing satisfying

$$d\begin{pmatrix} \eta^1 \\ \eta^2 \\ \eta^3 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \eta^2 \wedge \eta^3 \\ \eta^3 \wedge \eta^1 \\ \eta^1 \wedge \eta^2 \end{pmatrix}.$$

b) If $\lambda_3 = \varphi_1$, then $\mathfrak{g}$ is the Lie algebra of $E(2)$, the group of rigid motions of the Euclidean plane. A specific example here is given by the sub-Riemannian structure on $\mathbb{R}^2 \times S^1$, with coordinates (x, y, ϕ), induced by the coframing

$$\begin{aligned} \eta^1 &= \sqrt{2\varphi_1}\,(\cos\phi\, dx + \sin\phi\, dy) \\ \eta^2 &= -d\phi/\sqrt{2\varphi_1} \\ \eta^3 &= \sin\phi\, dx - \cos\phi\, dy. \end{aligned}$$

c) If $-\varphi_1<\lambda_3<\varphi_1$, then $\mathfrak{g}=\mathfrak{sl}(2,\mathbb{R})$. Here $M=Sl(2,\mathbb{R})$ with the sub-Riemannian structure $D=(\eta^3)^\perp$ and $\langle\ ,\ \rangle = \frac{1}{-\lambda_3+\varphi_1}(\eta^1)^2+\frac{1}{\lambda_3+\varphi_1}(\eta^2)^2\Big|_D$, where η is the coframing satisfying

$$d\begin{pmatrix}\eta^1\\ \eta^2\\ \eta^3\end{pmatrix}=\begin{pmatrix}1&0&0\\0&-1&0\\0&0&1\end{pmatrix}\begin{pmatrix}\eta^2\wedge\eta^3\\ \eta^3\wedge\eta^1\\ \eta^1\wedge\eta^2\end{pmatrix}.$$

d) If $\lambda_3=-\varphi_1$, then $\mathfrak{g}$ is the Lie algebra of $E(1,1)$, the group of rigid motions of the Lorentzian plane. A specific example here is given by the sub-Riemannian structure induced by the following coframing on $\mathbb{R}^3$ with coordinates (x,y,z):

$$\begin{aligned}\eta^1&=dz/\sqrt{2\varphi_1}\\ \eta^2&=\sqrt{2\varphi_1}\,(\cosh z\,dx-\sinh z\,dy)\\ \eta^3&=\sinh z\,dx-\cosh z\,dy.\end{aligned}$$

e) If $\lambda_3<-\varphi_1$, then $\mathfrak{g}=\mathfrak{sl}(2,\mathbb{R})$. Here $M=Sl(2,\mathbb{R})$ again, but with the sub-Riemannian structure $D=(\eta^3)^\perp$ and $\langle\ ,\ \rangle = \frac{1}{-\lambda_3+\varphi_1}(\eta^1)^2-\frac{1}{\lambda_3+\varphi_1}(\eta^2)^2\Big|_D$, where η is the coframing satisfying

$$d\begin{pmatrix}\eta^1\\ \eta^2\\ \eta^3\end{pmatrix}=\begin{pmatrix}-1&0&0\\0&-1&0\\0&0&1\end{pmatrix}\begin{pmatrix}\eta^2\wedge\eta^3\\ \eta^3\wedge\eta^1\\ \eta^1\wedge\eta^2\end{pmatrix}.$$

<u>case 2</u>: $\lambda_2=0, \lambda_3=\varphi_1$.

In this case, the structure equations are

$$d\begin{pmatrix}\omega^1\\ \omega^2\\ \omega^3\end{pmatrix}=\begin{pmatrix}2\varphi_1&0&-\lambda_1\\0&0&0\\0&0&1\end{pmatrix}\begin{pmatrix}\omega^2\wedge\omega^3\\ \omega^3\wedge\omega^1\\ \omega^1\wedge\omega^2\end{pmatrix}.$$

The Lie algebra is determined by the sign of the determinant $2\varphi_1-\lambda_1^2/4$ of a certain 2×2 block of the symmetric matrix $1/2(C+C^T)$.

a) If $2\varphi_1-\lambda_1^2/4>0$, then it is easily verified that this case is given by the sub-Riemannian structure on $\mathbb{R}^3$ induced by the coframing

$$\begin{aligned}\eta^1&=e^{-\frac{\lambda_1}{2}z}\left((-\sigma\sin\sigma z-\frac{\lambda_1}{2}\cos\sigma z)dx+(\sigma\cos\sigma z-\frac{\lambda_1}{2}\sin\sigma z)dy\right)\\ \eta^2&=-dz\\ \eta^3&=e^{-\frac{\lambda_1}{2}z}\left(\cos\sigma z\,dx+\sin\sigma z\,dy\right),\end{aligned}$$

where $\sigma=\sqrt{2\varphi_1-\lambda_1^2/4}$.

b) If $2\varphi_1{-}\lambda_1^2/4{=}0$, then it is easily verified that this case is given by the sub-Riemannian structure on $\mathbb{R}^3$ induced by the coframing

$$\begin{aligned}\eta^1 &= -\frac{\lambda_1^2}{4}e^{\frac{\lambda_1}{2}z}(dx - zdy)\\ \eta^2 &= dz\\ \eta^3 &= \frac{\lambda_1}{2}e^{\frac{\lambda_1}{2}z}(dx + (1-z)dy).\end{aligned}$$

c) If $2\varphi_1{-}\lambda_1^2/4{<}0$, then it is easily verified that this case is given by the sub-Riemannian structure on $\mathbb{R}^3$ induced by the coframing

$$\begin{aligned}\eta^1 &= e^{-\frac{\lambda_2}{2}z}\left((\frac{\lambda_2}{2}\sinh\sigma z - \sigma\cosh\sigma z)dx + (\sigma\sinh\sigma z - \frac{\lambda_2}{2}\cosh\sigma z)dy\right)\\ \eta^2 &= -dz\\ \eta^3 &= e^{-\frac{\lambda_2}{2}z}\left(-\sinh\sigma z\, dx + \cosh\sigma z\, dy\right),\end{aligned}$$

where $\sigma = \sqrt{\lambda_1^2/4{-}2\varphi_1}$.

<u>case 3</u>: $\lambda_1 = 0, \lambda_3 = -\varphi_1$.

Here the structure equations are

$$d\begin{pmatrix}\omega^1\\ \omega^2\\ \omega^3\end{pmatrix} = \begin{pmatrix}0 & 0 & 0\\ 0 & -2\varphi_1 & -\lambda_2\\ 0 & 0 & 1\end{pmatrix}\begin{pmatrix}\omega^2\wedge\omega^3\\ \omega^3\wedge\omega^1\\ \omega^1\wedge\omega^2\end{pmatrix}.$$

Now the determinant of the certain 2×2 block of the symmetric matrix $1/2(C{+}C^T)$ is equal to $-2\varphi_1{-}\lambda_2^2/4$, which is always negative. Therefore the Lie algebra in this case is determined and this situation is essentially the same situation as case 2 c). It is easily verified that this case is in fact given by the sub-Riemannian structure on $\mathbb{R}^3$ induced by the coframing

$$\begin{aligned}\eta^1 &= dz\\ \eta^2 &= e^{-\frac{\lambda_2}{2}z}\left((\frac{\lambda_2}{2}\sinh\sigma z - \sigma\cosh\sigma z)dx + (\sigma\sinh\sigma z - \frac{\lambda_2}{2}\cosh\sigma z)dy\right)\\ \eta^3 &= e^{-\frac{\lambda_2}{2}z}\left(-\sinh\sigma z\, dx + \cosh\sigma z\, dy\right),\end{aligned}$$

where $\sigma = \sqrt{2\varphi_1{+}\lambda_2^2/4}$.

2.4. The meaning of the invariants. The functions a_1, a_2, and K form a generating set of differential invariants for sub-Riemannian structures on 3-manifolds. We now give geometric interpretations of these invariants. We show for example that if $a_1^2{+}a_2^2$ is identically zero, then M^3 naturally fibers over a surface with Gauss curvature K.

Proposition 2.1. *The functions $a_1^2 + a_2^2$ and K are well defined on M^3.*

Proof. We make use of equations (2) and (3) to compute that the exterior derivative of $a_1^2 + a_2^2$ is equal to zero modulo the semi-basic one forms ω^i, from which it follows that $a_1^2 + a_2^2$ is well defined on M^3. Similarly, if we differentiate both sides of equation (4) and then wedge with ω^3, we find that dK is equal to zero modulo ω^i. □

Recall that a local coframing η is 2-adapted to the sub-Riemannian manifold $(M^3, D, \langle\ ,\ \rangle)$ if η is the image of a local section of $\mathcal{B}_2$. A sub-Riemannian manifold will be said to be *amenable* if the leaf space of the codimension 2 foliation defined by the equations $\eta^1 = \eta^2 = 0$ can be given the structure of a smooth surface N in such a way that the natural projection $\pi_N : M^3 \to N$ is a smooth submersion, for any (and hence every) 2-adapted coframing $\eta = (\eta^1, \eta^2, \eta^3)$.

Given an amenable sub-Riemannian manifold and any 2-adapted coframing η, the vector field V defined by $\eta^3(V) = 1$ and $V \lrcorner\, d\eta^3 = 0$ is vertical for $\pi_N : M^3 \to N$. At each point $p \in M^3$, the vector V_p and the plane D_p span the tangent space T_pM; thus the derivative map $\pi_N'(p) : T_pM \to T_{[p]}N$ restricts to be an isomorphism $\pi_N'(p) : D_p \to T_{[p]}N$. Let $(\langle\ ,\ \rangle_N)_p$ denote the induced inner product on the tangent space $T_{[p]}N$.

Now as p varies in the fiber over $[p]$, the inner product $(\langle\ ,\ \rangle_N)_p$ varies in the space of positive definite inner products on $T_{[p]}N$. There is a canonical area form, namely $\omega^1\wedge\omega^2 = \eta^1\wedge\eta^2$, on N, and by definition, for each p the inner product $(\langle\ ,\ \rangle_N)_p$ has unit volume with respect to this area form. So in fact as p moves along the fiber, $(\langle\ ,\ \rangle_N)_p$ will define a curve in the space $S_{[p]}$ of unit-volume positive definite inner products on $T_{[p]}N$. This space $S_{[p]}$ is a two-dimensional disk and has a natural metric on it that is isometric to the Poincaré metric on the disk. Now, the fiber $\pi_N^{-1}[p]$ also has a natural metric, namely η^3. We will show in Theorem 2.1 below that if p moves along the fiber with unit speed, then $(\langle\ ,\ \rangle_N)_p$ defines a curve in $S_{[p]}$ whose speed is $2\sqrt{a_1^2+a_2^2}$.

We first define a coordinate system adapted to the sub-Riemannian structure. Fix a point $q \in M^3$, and choose a neighborhood $U \subset M^3$ of q and coordinates $(x, y, z) : U \to \mathbb{R}^3$ centered at q such that the vector field V is given by $\partial/\partial z$. Then $V\lrcorner(\omega^1\wedge\omega^2)=0$ implies that $\omega^1\wedge\omega^2=f(x,y)\,dx\wedge dy$, and by absorbing in x and y we may assume that $f(x,y)\equiv 1$ on U. In these coordinates the quadratic form $(\omega^1)^2+(\omega^2)^2$ is equal to $Edx^2 + 2Fdxdy + Gdy^2$, where E, F, G are functions on U that satisfy $EG-F^2 = 1$.

Theorem 2.1. *Let $(M^3, D, \langle\ ,\ \rangle)$ be an anemable sub-Riemannian manifold, and let $[p]$ be a point in the leaf space N. As p moves along the fiber $\pi_N^{-1}[p]$ with unit speed (with respect to the metric ω^3) the corresponding inner product $(\langle\ ,\ \rangle_N)_p$ on $T_{[p]}N$ defines a curve in $S_{[p]}$ whose speed is $2\sqrt{a_1^2 + a_2^2}$ (with respect to the Poincaré metric).*

Proof. Choose coordinates as above so that $[p] = (0, 0)$. Now the fiber is the set $\{(0, 0, z) \mid z \in \mathbb{R}\}$ and the metric ω^3 on the fiber is just dz. The curve in $S_{[p]}$ corresponding to the unit speed curve $\gamma(t) = (0, 0, t)$ in the fiber is

$$t \longmapsto \begin{bmatrix} E(0,0,t) & F(0,0,t) \\ F(0,0,t) & G(0,0,t) \end{bmatrix}.$$

The squared speed of this curve with respect to the Poincaré metric is equal to $F_z^2 - E_zG_z$.

The Lie derivative of $(\omega^1)^2+(\omega^2)^2$ in the V direction is easily seen to be equal to

$$a_2(\omega^1)^2 - a_2(\omega^2)^2 - 2a_1\omega^1\omega^2.$$

On the other hand, in adapted coordinates, the Lie derivative in the direction $V{=}\partial/\partial z$ is equal to $E_z dx^2{+}2F_z dxdy{+}G_z dy^2$. Because the area form $\omega^1{\wedge}\omega^2$ is equal to $dx{\wedge}dy$, it follows that

$$F_z^2 - E_z G_z = (-2a_1)^2 - (2a_2)(-2a_2) = 4(a_1^2 + a_2^2),$$

and therefore the speed of the curve is equal to $2\sqrt{a_1^2{+}a_2^2}$. □

As a consequence of this theorem, if $a_1^2{+}a_2^2{=}0$ then the induced inner product $(\langle\ ,\ \rangle_N)_p$ gives a well defined metric on N. In this case the structure equations (1) become

$$\begin{aligned} d\omega^1 &= -\alpha \wedge \omega^2 \\ d\omega^2 &= \alpha \wedge \omega^1 \\ d\omega^3 &= \omega^1 \wedge \omega^2 \\ d\alpha &= K\omega^1 \wedge \omega^2. \end{aligned} \tag{6}$$

Differentiating the last equation, we find that dK is equal to zero modulo ω^1, ω^2 and therefore K is well defined on N.

Proposition 2.2. *Let $(M^3, D, \langle\ ,\ \rangle)$ be amenable. The induced inner product $(\langle\ ,\ \rangle_N)_p$ on $T_{[p]}N$ gives a well defined metric on N if and only if $a_1^2 + a_2^2 = 0$. In this case, the Gauss curvature of N is equal to K.*

Proof. The first statement follows from Theorem 2.1. To compute the curvature of N with the metric $\langle\ ,\ \rangle_N$, we choose a local section $\sigma : W \to M^3$ for some suitably small neighborhood $W \subseteq N$, and consider the pull-back bundle $\sigma^{-1}\mathcal{B}_2$ over W. If $\eta = (\eta^1, \eta^2, \eta^3)$ is any 2-adapted coframing then $(\sigma^*\eta^1, \sigma^*\eta^2)$ is an orthonormal coframing of W. Because η^3 is uniquely determined, the bundle $\mathcal{F}_0$ of orthonormal coframes for W can be identified with $\sigma^{-1}\mathcal{B}_2$.

With this identification, we let $(\sigma, \mathbf{1})$ be the map from $\mathcal{F}_0$ to $\mathcal{B}_2$ that sends a point $([p], \eta_p)$ in $\mathcal{F}_0$ to the point $(\sigma([p]), \eta_p)$ in $\sigma^{-1}\mathcal{B}_2$. Unwinding definitions, it is clear that the pull-backs $(\sigma, \mathbf{1})^*\omega^1$ and $(\sigma, \mathbf{1})^*\omega^2$ are the tautological one-forms on $\mathcal{F}_0$. The structure equations (6) now imply that $(\sigma, \mathbf{1})^*\alpha$ is the connection form on $\mathcal{F}_0$ and therefore $(\sigma, \mathbf{1})^*K = \sigma^*K$ is the Gauss curvature of N. □

If $a_1^2{+}a_2^2$ vanishes identically, then in adapted coordinates the sub-Riemannian metric has the expression $E(x,y)dx^2{+}2F(x,y)dxdy{+}G(x,y)dy^2$ where $EG{-}F^2{=}1$. In fact we can always choose geodesic normal coordinates on N to assume the metric has the expression

$$dx^2 + dy^2 - H(x,y)(xdy - ydx)^2.$$

The contact form ω^3 is then $\omega^3{=}dz{-}e_1dx{-}e_2dy$, where $(e_2)_x{-}(e_1)_y{=}1$. Note that if $\gamma : [a,b] \to U$ is a D-curve that lies in an adapted coordinate neighborhood U with the property that the projection $\pi_N{\circ}\gamma$ is a closed curve on N, then Stokes' Theorem implies $z(b){-}z(a)$ is equal to the area enclosed by $\pi_N{\circ}\gamma$. Furthermore, the length of the D-curve γ is equal to the length of $\pi_N{\circ}\gamma$. We can thus describe sub-Riemannian manifolds with $a_1^2{+}a_2^2$ identically zero as sub-Riemannian manifolds of "areal type."

If $(M^3, D, \langle\ ,\ \rangle)$ is not of areal type then K is not even well defined on N. To interpret K in this case, we associate to the sub-Riemannian manifold the canonical Riemannian metric $(\omega^1)^2+(\omega^2)^2+(\omega^3)^2$. Let $\bar{\Theta}_{\mathrm{LC}}$ denote the Levi-Civita connection of this Riemannian metric restricted to the bundle $\mathcal{B}_2$: since a 2-adapted coframing (η^1, η^2, η^3) is an orthonormal coframing of M^3 in this metric, we think of $\mathcal{B}_2$ as a sub-bundle of the $SO(3)$-bundle of orthonormal coframes over M^3, and then $\bar{\Theta}_{\mathrm{LC}}$ is the pull-back by inclusion of the Levi-Civita connection.

Setting

$$\bar{\Theta}_{\mathrm{LC}} = \begin{pmatrix} 0 & \alpha_3 & -\alpha_2 \\ -\alpha_3 & 0 & \alpha_1 \\ \alpha_2 & -\alpha_1 & 0 \end{pmatrix}$$

and comparing the structure equations $d\omega = -\bar{\Theta}_{\mathrm{LC}} \wedge \omega$ with (1), we find $\alpha_1 = (1/2-a_1)\omega^1 - a_2\omega^2$, $\alpha_2 = -a_2\omega^1 + (a_1+1/2)\omega^2$, and $\alpha_3 = -1/2\omega^3 + \alpha$. We compute the entries of the curvature $\Omega = d\bar{\Theta}_{\mathrm{LC}} + \bar{\Theta}_{\mathrm{LC}} \wedge \bar{\Theta}_{\mathrm{LC}}$ to be

$$\begin{aligned}
\Omega_{12} &= (A_2 - B_1)\,\omega^2 \wedge \omega^3 + (A_1 + B_2)\,\omega^3 \wedge \omega^1 \\
&\quad + (a_1^2 + a_2^2 + K - \frac{3}{4})\,\omega^1 \wedge \omega^2, \\
\Omega_{13} &= (A_3 - a_2)\,\omega^2 \wedge \omega^3 + (B_3 + a_1^2 + a_2^2 + a_1 - \frac{1}{4})\,\omega^3 \wedge \omega^1 \\
&\quad - (A_1 + B_2)\,\omega^1 \wedge \omega^2, \\
\Omega_{23} &= (B_3 + a_1 - a_1^2 - a_2^2 + \frac{1}{4})\,\omega^2 \wedge \omega^3 + (-A_3 + a_2)\,\omega^3 \wedge \omega^1 \\
&\quad + (A_2 - B_1)\,\omega^1 \wedge \omega^2.
\end{aligned}$$

Thus, $K + a_1^2 + a_2^2 - \frac{3}{4}$ is the sectional curvature of the plane D.

Remark 1. *The sub-Riemannian structure $(D, \langle\ ,\ \rangle)$ produces a CR structure on the three-manifold M^3. The Webster curvature W (*[23], [5]*) of a CR manifold is defined to be a certain combination of the components of the curvature tensor of the (unique) Riemannian metric inducing the CR structure (see* [5] *for the definition of a metric "inducing the CR structure" and for the explicit formula for W. For a CR structure induced by a sub-Riemannian structure, this Riemannian metric inducing the CR structure turns out to be $(\omega^1)^2+(\omega^2)^2+4(\omega^3)^2$. Computing the curvature tensor of this metric and substituting into the formula for W, we find*

$$W = \frac{K}{4}.$$

2.5. Spin calculations. The geometric significance of some of the higher order derivatives of the invariants will be illuminated in the next section. It will be necessary to determine the combinations of the derivatives of the invariants that are well defined on M^3. We now discuss a procedure for doing this that makes use of the complex structure on D.

On an open set $U \subseteq M^3$, we choose an orientation of D and consider the coframes in $\mathcal{B}_2$ that preserve this orientation. The set of such coframes forms an $SO(2)$-structure and we will continue to use "the G_2-structure $\mathcal{B}_2 \to M^3$" to denote this $SO(2)$-structure. We identify an element $g \in G_2$ with the complex number $e^{i\theta}$, and set $\Omega = \omega^1 + i\omega^2$. The equation $R_g^*\,\omega = g^{-1}\omega$ implies $R_{e^{i\theta}}^*\,\Omega = e^{-i\theta}\,\Omega$, $R_{e^{i\theta}}^*\,\omega^3 = \omega^3$, and $R_{e^{i\theta}}^*\,\alpha = \alpha$. Furthermore, $R_{e^{i\theta}}^*\,\Omega = e^{-i\theta}\,\Omega$ implies $R_{e^{i\theta}}^*\,\bar{\Omega} = e^{i\theta}\,\bar{\Omega}$. It follows that the pullback by $R_{e^{i\theta}}$ of any function φ appearing in the covariant derivatives of any of these forms is equal to $e^{in\theta}\varphi$ for some integer n. This integer n is called the *spin* of φ.

Note that a function φ on $\mathcal{B}_2$ is well defined on M^3 if and only if it has spin zero, for φ is constant on the fibers if and only if $R_{e^{i\theta}}^*\,\varphi = \varphi$. It is clear that $\text{spin}(d\varphi) = \text{spin}(\varphi)$, $\text{spin}(\bar{\varphi}) = -\text{spin}(\varphi)$, and $\text{spin}(\varphi \wedge \rho) = \text{spin}(\varphi) + \text{spin}(\rho)$. To find the spin of derivatives of the invariants, we simply compute the structure equations in the basis $\{\Omega, \bar{\Omega}, \omega^3, \alpha\}$.

Set $a = a_1 + ia_2$. The structure equations (1) are now

$$d\Omega = i\alpha \wedge \Omega + ia\bar{\Omega} \wedge \omega^3$$
$$d\omega^3 = \frac{i}{2}\Omega \wedge \bar{\Omega},$$

and differentiating these equations gives

$$da = 2ia\alpha + b^1\Omega + b^2\bar{\Omega} + b^3\omega^3 \tag{7}$$

$$d\alpha = \frac{i}{2}K\Omega \wedge \bar{\Omega} - (b^1\bar{\Omega} + \bar{b^1}\Omega) \wedge \omega^3. \tag{8}$$

Differentiating again, we get (after some simplification) in particular that

$$db^1 = ib^1\alpha + b^{11}\Omega + b^{12}\bar{\Omega} + b^{13}\omega^3 \tag{9}$$

$$db^2 = 3ib^2\alpha + b^{21}\Omega + b^{22}\bar{\Omega} + b^{23}\omega^3 \tag{10}$$

where $b^{21} - b^{12} = aK - \frac{i}{2}b^3$.

These equations allow us to easily read off the spin of each function. For example, because Ω has spin 1 and ω^3 has spin 0, the equation $d\Omega = i\alpha\wedge\Omega + ia\bar{\Omega}\wedge\omega^3$ implies the spin of a is equal to -2. This in turn allows us to read off the spin of the b^i's. In this way, we construct the following table:

function	a	K	b^1	b^2	b^3	b^{11}	b^{12}	b^{13}	b^{21}	b^{22}	b^{23}
spin	-2	0	-3	-1	-2	-4	-2	-3	-2	0	-1

For future reference we will now express each b^i and b^{ij} in terms of the functions A_i, B_i, A_{ij}, and B_{ij}, where

$$dA_i = A_{i0}\,\alpha + \sum A_{ij}\omega^j \quad \text{and} \quad dB_i = B_{i0}\,\alpha + \sum B_{ij}\omega^j.$$

Comparing (7) with (2) and (3), we find

$$2b^1 = A_1 + B_2 + i(B_1 - A_2) \tag{11}$$

$$2b^2 = A_1 - B_2 + i(B_1 + A_2) \tag{12}$$

$$b^3 = A_3 + iB_3. \tag{13}$$

Now differentiating these equations and comparing with (9) and (10), we find

$$(14)\qquad \begin{aligned} 4b^{11} &= (A_{11} - A_{22} + B_{12} + B_{21}) + i(-A_{12} - A_{21} + B_{11} - B_{22}) \\ 4b^{12} &= (A_{11} + A_{22} + B_{21} - B_{12}) + i(A_{12} - A_{21} + B_{11} + B_{22}) \\ 4b^{21} &= (A_{11} + A_{22} + B_{12} - B_{21}) + i(A_{21} - A_{12} + B_{11} + B_{22}) \\ 4b^{22} &= (A_{11} - A_{22} - B_{12} - B_{21}) + i(A_{12} + A_{21} + B_{11} - B_{22}). \end{aligned}$$

Furthermore, we have the following relations:

$$(15)\qquad A_{10} = A_2 + 2B_1, \qquad A_{20} = A_1 - 2B_2,$$

$$(16)\qquad B_{10} = 2A_1 - B_2, \qquad B_{20} = 2A_2 + B_1,$$

$$(17)\qquad B_{12} - B_{21} = 2a_1K + B_3, \qquad A_{21} - A_{12} = 2a_2K - A_3.$$

3. The Geodesics

We now turn to the problem of computing the geodesics of $(M^3, D, \langle\ ,\ \rangle)$. We use the Griffiths formalism to compute the geodesic equations and we show that the curves that satisfy these equations are precisely the extremals among D-curves of the length functional. We compute the second variation of the length functional and show that the index is zero if and only if there are no solutions of a certain fourth-order self-adjoint equation, defined in §3.2, that vanish to first order. In particular, a geodesic is not length-minimizing beyond its first conjugate point. We then obtain criterion on the invariants that allows us to estimate the maximum distance L to the first conjugate point. As a consequence, if $(M^3, D, \langle\ ,\ \rangle)$ is complete and the invariants satisfy the criterion, then M^3 is compact with diameter (with respect to the Carnot-Carathéodory distance d) no greater than L. We define the sub-Riemannian exponential map $\exp_p$ and show that unlike the Riemannian exponential map, $\exp_p$ is never a local diffeomorphism. We also show that $(M^3, D, \langle\ ,\ \rangle)$ is complete if and only if every geodesic can be extended indefinitely. Finally, we discuss the geodesics of the homogeneous examples in some detail.

3.1. The geodesic equations. As in the last section, let us choose an orientation of D and consider the $SO(2)$-structure consisting of the set of coframes in $\mathcal{B}_2$ that preserve the orientation. We continue to use the notation "the G_2-structure $\mathcal{B}_2 \to M^3$" to denote this $SO(2)$-structure. Every D-curve has a canonical lift as an integral curve on $\mathcal{B}_2$ of the system $S = \{\omega^2, \omega^3\}$ with transversality condition $\omega^1 \neq 0$. This lift just corresponds to choosing a 2-adapted coframing η along the curve so that the framing (e_1, e_2, e_3) dual to η has e_1 in the same direction as the tangent vector. Now the length of the D-curve is equal to the integral of ω^1 along the lifted integral curve of S. Thus our problem of finding the extremals of the length functional L among D-curves is equivalent to finding the extremals of

$$\mathcal{L}(\gamma) = \int_\gamma \omega^1$$

among integral curves γ of S on $\mathcal{B}_2$.

Proposition 3.1. *The extremals of* $\mathcal{L}$ *among integral curves of* $S = \{\omega^2, \omega^3\}$ *on* $\mathcal{B}_2$ *are precisely the projections of integral curves, with transversality condition* $\omega^1 \neq 0$*, of the differential system* $I = \{\omega^2, \omega^3, \alpha - \lambda\omega^1, d\lambda - a_2\omega^1\}$ *on the bundle* $Y = \mathcal{B}_2 \times \mathbb{R}$*, where* λ *is the coordinate of the fiber* $\mathbb{R}$*.*

Proof. Following the Griffiths algorithm [8], we let Z be the affine subbundle $Z = S + \omega^1$ of the cotangent bundle $T^*\mathcal{B}_2$ and let ζ be the restriction to Z of the canonical one-form on $T^*\mathcal{B}_2$:

$$\zeta = \omega^1 + \lambda_1\omega^2 + \lambda_2\omega^3.$$

By the general theory, the extremals of the functional

$$\widetilde{\mathcal{L}}(\gamma) = \int_\gamma \zeta$$

among unconstrained curves on Z project to be extremals of $\mathcal{L}$ among integral curves of S on $\mathcal{B}_2$. The curve γ is an extremal of $\widetilde{\mathcal{L}}$ among unconstrained curves on Z if and only if $\gamma'(t) \lrcorner d\zeta|_{\gamma(t)} = 0$.

We compute $d\zeta$ modulo $\omega^2 \wedge \omega^3$:

$$d\zeta \equiv -\alpha \wedge \omega^2 + \lambda_1\alpha \wedge \omega^1 + (a_2 - \lambda_1 a_1)\omega^3 \wedge \omega^1 + \lambda_2\omega^1 \wedge \omega^2 + d\lambda_1 \wedge \omega^2 + d\lambda_2 \wedge \omega^3.$$

We contract $d\zeta$ with the vector fields dual to the coframing $(\omega^1, \omega^2, \omega^3, \alpha, d\lambda_1, d\lambda_2)$ on Z and find that subject to $\gamma^*\omega^1 \neq 0$, the condition $\gamma' \lrcorner d\zeta = 0$ is equivalent to the condition that γ is an integral curve of the system $\{\omega^2, \omega^3, \alpha - \lambda_2\omega^1, d\lambda_2 - a_2\omega^1\}$, and that γ lies on the locus $Y \subset Z$ defined by $\lambda_1 = 0$. Thus the integral curves of I on Y project to be extremals of $\mathcal{L}$ among integral curves of S.

To show that all the extremals of $\mathcal{L}$ arise this way, we need to show that every extremal among integral curves of S on $\mathcal{B}_2$ can be lifted to an integral curve of I on Y. Because D is a contact distribution every integral curve of S on $\mathcal{B}_2$ is regular. By a result of Hsu [13] it follows that every extremal of $(\mathcal{B}_2, S, \mathcal{L})$ lifts to a unique extremal of $(Y, \{0\}, \tilde{\mathcal{L}})$, i.e., it lifts to an integral curve of I. □

We will say that a D-curve $\gamma : [a, b] \to M^3$ parametrized with unit speed is a *geodesic* if it lifts to an integral curve of the system I on Y. By the previous proposition, a geodesic γ is necessarily an extremal of the length functional $L(\gamma)$ and conversely every extremal is necessarily a geodesic. Consequently, every length-minimizing D-curve is in fact a geodesic.

Under the assumption that γ has unit speed, the condition that it lifts to an integral curve of I is

$$\omega^1 = ds, \quad \omega^2 = 0, \quad \omega^3 = 0, \quad \alpha = \lambda\, ds, \quad d\lambda = a_2\, ds. \tag{18}$$

We will refer to these equations as the *geodesic equations*.

Remark 2. *The I-lift $\bar{\gamma}$ of the geodesic γ pulls back the forms $\alpha_1=(\frac{1}{2}-a_1)\,\omega^1-a_2\omega^2$, $\alpha_2=-a_2\omega^1+(a_1+\frac{1}{2})\,\omega^2$, and $\alpha_3=-\frac{1}{2}\omega^3+\alpha$ in the Levi-Civita connection $\bar{\Theta}_{LC}=\left[\begin{smallmatrix}0 & \alpha_3 & -\alpha_2\\ -\alpha_3 & 0 & \alpha_1\\ \alpha_2 & -\alpha_1 & 0\end{smallmatrix}\right]$ to be*

$$\begin{aligned}\bar{\gamma}^*\alpha_1 &= (\frac{1}{2}-a_1)\,ds\\ \bar{\gamma}^*\alpha_2 &= -a_2\,ds\\ \bar{\gamma}^*\alpha_3 &= \lambda\,ds.\end{aligned}$$

The curvature of γ is equal to $\sqrt{\lambda^2+a_2^2}$. Thus, if $a_1^2+a_2^2=0$, the curvature of the geodesic γ is equal to λ and is constant, and the torsion τ defined by $\bar{\gamma}^*\alpha_1=\tau ds$ is equal to 1/2.

As a consequence of this remark, we shall refer to the function λ along the lifted geodesic $\bar{\gamma}$ as the *sub-Riemannian curvature* of γ.

3.2. The Jacobi operator. We compute the second variation of the length functional and interpret the Jacobi operator as a scalar, fourth-order self-adjoint linear operator along each geodesic. We show that the index of the Hessian of the length functional is equal to the number of conjugate points.

Because the geodesic equations are defined there, it is convenient to work up on the bundle Y. Set $\omega^4=\alpha-\lambda\omega^1$ and $\omega^5=d\lambda-a_2\omega^1$. Every integral curve of the system $\{\omega^2,\omega^3\}$ on $\mathcal{B}_2$ has a canonical lift as an integral curve of the system $I'=\{\omega^2,\omega^3,\omega^4\}\subset I$ on Y. Therefore we identify D-curves with integral curves of I', with transversality condition $\omega^1\neq 0$, on Y. The length of the D-curve is equal to the integral of ω^1 along the lifted integral curve of I'.

Let $\gamma:[0,l]\to M^3$ be a D-curve joining p to q and let $\bar{\gamma}$ be its lift as integral curve of I'. If γ_t is a fixed endpoint variation of γ through D-curves, then γ_t lifts to be a variation $\bar{\gamma}_t$ of $\bar{\gamma}$ through integral curves of I' which does not necessarily fix the endpoints, but satisfies $\pi\circ\bar{\gamma}_t(0)=p$, $\pi\circ\bar{\gamma}_t(l)=q$ where π is the standard projection $\pi:Y\to M$. A variation $\bar{\gamma}_t$ through integral curves of I' on Y which satisfies $\pi\circ\bar{\gamma}_t(0)=p$, $\pi\circ\bar{\gamma}_t(l)=q$ will be said to be an *admissable variation* of $\bar{\gamma}$ and its variational vector field $\partial\bar{\gamma}_t/\partial t$ at $t=0$ will be said to be an *infinitesimal admissable variation* along $\bar{\gamma}$.

Now suppose the I'-curve $\bar{\gamma}$ is the lift of a geodesic, so that $\bar{\gamma}$ is an integral curve of I with $\bar{\gamma}^*\omega^1\neq 0$. Let $\bar{\gamma}_{t,u}$ be a 2-parameter variation of $\bar{\gamma}$ with infinitesimal admissable variations

$$X=\frac{\partial}{\partial t}\bar{\gamma}_{t,u}\Big|_{t=u=0}\qquad\text{and}\qquad W=\frac{\partial}{\partial u}\bar{\gamma}_{t,u}\Big|_{t=u=0}.$$

The hessian $\mathcal{L}_{**}(X,W)$ of the length functional $\mathcal{L}$ is

$$\mathcal{L}_{**}(X,W)=\frac{\partial^2}{\partial t\partial u}\mathcal{L}(\bar{\gamma}_{t,u})\Big|_{t=u=0}.$$

Let $(e_1,\ldots,e_5)$ be the framing dual to the coframing $(\omega^1,\ldots,\omega^5)$ on Y, and write $X=\sum_i X_ie_i$ and $W=\sum_i W_ie_i$.

We define the *Jacobi operator* along the lifted geodesic $\bar\gamma$ to be the following fourth-order differential operator on the space of smooth functions on $[0, l]$:

$$Ju = \ddot{\ddot{u}} + \frac{d}{ds}(p\dot u) + q,$$

where p and q are given by the following expressions along $\bar\gamma$:

$$p = -3a_1 + K + \lambda^2 \tag{19}$$

$$q = 2a_1^2 + 4a_2^2 - 2a_1K + B_{12} - 2A_{11} - 2B_{21} + 6a_1\lambda^2 + 6B_1\lambda. \tag{20}$$

This definition is motivated by the following proposition.

Proposition 3.2. *Let $\bar\gamma : [0, l] \to Y$ be a lifted geodesic with 2-parameter admissable variation $\bar\gamma_{t,u}$. The hessian of the length functional evaluated at the infinitesimal admissable variations $X = \sum_i X_i e_i$ and $W = \sum_i W_i e_i$ of $\bar\gamma_{t,u}$ is*

$$\mathcal{L}_{**}(X, W) = \int_0^l W_3\, JX_3\, ds.$$

Proof. A straightforward computation shows the hessian $\mathcal{L}_{**}(X, W)$ is equal to the integral $\int_{\bar\gamma} W \lrcorner\, d(X \lrcorner\, d\zeta)$, where $\zeta = \omega^1 + \lambda\omega^3$ (see [8]). Now we compute that $W \lrcorner\, d(X \lrcorner\, d\zeta)$ is equivalent, modulo I, to

$$\begin{aligned}&\Big[W_2\Big(\dot X_4 + \frac{d}{ds}(a_1X_3) - X_5 - X_2(a_1 + K + \lambda^2) + X_3(B_2 + a_2\lambda)\Big)\\ &+ W_3\Big(a_1X_4 - \frac{d}{ds}(a_1X_2 + X_5) + X_2(A_1 + B_2 - a_2\lambda) + X_3(B_3 + a_1^2 + a_2^2)\Big)\\ &+ W_4\Big(-X_4 + a_1X_3 - \dot X_2\Big) + W_5\Big(\dot X_3 - X_2\Big)\Big]\omega^1.\end{aligned} \tag{21}$$

Set $\Gamma(t, u, s) = \bar\gamma_{t,u}(s)$. We have

$$\Gamma^*\left\{\begin{array}{c}\omega^1\\ \omega^2\\ \omega^3\\ \omega^4\\ \omega^5\end{array}\right\} = \left\{\begin{array}{l} X_1(t,u,s)dt + W_1(t,u,s)du + Y_1(t,u,s)ds\\ X_2(t,u,s)dt + W_2(t,u,s)du\\ X_3(t,u,s)dt + W_3(t,u,s)du\\ X_4(t,u,s)dt + W_4(t,u,s)du\\ X_5(t,u,s)dt + W_5(t,u,s)du + Y_5(t,u,s)ds\end{array}\right\}$$

where $X_i(0, 0, s) = X_i$ and $W_i(0, 0, s) = W_i$. Because $\bar\gamma$ is the lift of a geodesic, we also have $Y_1(0, 0, s) = 1$, $\partial Y_1/\partial t(0, 0, s) = \partial Y_1/\partial u(0, 0, s) = 0$, and $Y_5(0, 0, s) = 0$. Now the structure equations for the coframing $(\omega^1, \dots, \omega^5)$ on Y are easily computed to be

$$\begin{aligned}
d\omega^1 &= \omega^2\wedge\omega^4 - \lambda\omega^1\wedge\omega^2 + a_1\omega^2\wedge\omega^3 + a_2\omega^3\wedge\omega^1\\
d\omega^2 &= -\omega^1\wedge\omega^4 + a_2\omega^2\wedge\omega^3 - a_1\omega^3\wedge\omega^1\\
d\omega^3 &= \omega^1\wedge\omega^2\\
d\omega^4 &= \omega^1\wedge\omega^5 - \lambda\omega^2\wedge\omega^4 + (A_2 - B_1 - a_1\lambda)\omega^2\wedge\omega^3\\
&\quad + (A_1 + B_2 - a_2\lambda)\omega^3\wedge\omega^1 + (K+\lambda^2)\omega^1\wedge\omega^2\\
d\omega^5 &= 2a_1\omega^1\wedge\omega^4 - a_2\omega^2\wedge\omega^4 - a_1a_2\omega^2\wedge\omega^3\\
&\quad - (B_3 + a_2^2)\omega^3\wedge\omega^1 + (B_2 + a_2\lambda)\omega^1\wedge\omega^2.
\end{aligned} \tag{22}$$

These equations, when pulled back by Γ, imply that X_i and W_i satisfy the following system of differential equations:

$$\begin{aligned}
\dot X_1 &= -\lambda X_2 - a_2X_3\\
\dot X_2 &= -X_4 + a_1X_3\\
\dot X_3 &= X_2\\
\dot X_4 &= X_5 - (A_1 + B_2 - a_2\lambda)X_3 + (K+\lambda^2)X_2,
\end{aligned} \tag{23}$$

where $\dot X_i$ denotes the derivative of X_i with respect to s. The first equation may be integrated to give $X_1 = -\lambda X_3 + X_1(0)$; because $X = \sum_i X_ie_i$ is an infinitesimal admissable variation, X_1, X_2, X_3 must vanish at the endpoints, so that $X_1 = -\lambda X_3$. Thus, we may rewrite the above equations as

$$\begin{aligned}
X_1 &= -\lambda X_3\\
X_2 &= \dot X_3\\
X_4 &= a_1X_3 - \ddot X_3\\
X_5 &= -\dddot X_3 + (a_1 - K - \lambda^2)\dot X_3 + (\dot a_1 + A_1 + B_2 - a_2\lambda)X_3.
\end{aligned} \tag{24}$$

Now if we substitute these formulas into the expression (21) and integrate by parts several times, we find that the integral $\int_{\bar\gamma} W \lrcorner d(X \lrcorner d\zeta)$ is equal to

$$\begin{aligned}
\int_0^l W_3\Big(&\ddddot X_3 + (-3a_1 + K + \lambda^2)\ddot X_3 + (4a_2\lambda + \dot K - 2\dot a_1 - A_1)\dot X_3\\
&+ (2a_1^2 + 2a_2^2 + B_3 + \dot a_2\lambda - \ddot a_1 - \dot A_1 - \dot B_2)X_3\Big)ds.
\end{aligned} \tag{25}$$

The geodesic $\bar\gamma$ pulls back α to be λds; thus equation (2) implies that along $\bar\gamma$, $\dot a_1 = -2a_2\lambda + A_1$. This together with $a_2 = \dot\lambda$ implies that the derivative of $p = -3a_1 + K + \lambda^2$ with respect to s is

$$\dot p = -3\dot a_1 + \dot K + 2\lambda\dot\lambda = 4a_2\lambda + \dot K - 2\dot a_1 - A_1.$$

Finally, we need to show that we can rewrite the expression

$$2a_1^2 + 2a_2^2 + B_3 + \dot a_2\lambda - \ddot a_1 - \dot A_1 - \dot B_2 \tag{26}$$

as q, where q is given by equation (20). We first use equations (2), (3), (15), and (16) to rewrite (26) without the "dots." Equations (2) and (3) imply $\dot{a}_1 = -2a_2\lambda + A_1$ and $\dot{a}_2 = 2a_1\lambda + B_1$, and equations (15) and (16) imply $\dot{A}_1 = -\lambda(A_2 + 2B_1) + A_{11}$ and $\dot{B}_2 = \lambda(2A_2 + B_1) + B_{21}$. Using these equations, the expression (26) becomes

$$2a_1^2 + 4a_2^2 + B_3 - 2A_{11} - B_{21} + 6a_1\lambda^2 + 6B_1\lambda.$$

Now we use (17) to get rid of B_3, and this expression becomes the right-hand side of equation (20).

Therefore (25) is equal to the integral along $\bar{\gamma}$ of the product of W_3 and JX_3, where J is the Jacobi operator, and the proposition is proved. □

We see in the proof of Proposition 3.2 that any infinitesimal admissable variation $X = \sum_i X_i e_i$ satisfies $\dot{X}_3 = X_2$ and that X_1, X_2, X_3 vanish at the endpoints 0 and l. In particular, X_3 vanishes to first order at 0 and l. Let $C_0^\infty[0, l]$ denote the space of smooth functions on the interval $[0, l]$ that vanish to first order at 0 and l. Note that the Jacobi operator J is formally self-adjoint on $C_0^\infty[0, l]$.

We denote the quadratic form $\mathcal{L}_{**}(u, u)$ by $Q(u)$. Recall the index of $Q(u)$ is the dimension of the largest subspace of $C_0^\infty[0, l]$ on which $Q(u)$ is negative definite. Because J is self-adjoint on $C_0^\infty[0, l]$ the eigenvalues λ_i are real and constitute a countable set with the point at infinity the only possible cluster point. Thus the eigenvalues can be arranged

$$\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n < 0 \leq \lambda_{n+1} \leq \cdots$$

in ascending order. Furthermore, the index of $Q(u)$ is equal to the number of negative eigenvalues $\lambda_1 \leq \cdots \leq \lambda_n < 0$ of J. In particular, the index of $Q(u)$ is finite.

It turns out that the number of negative eigenvalues of J is intimately related to the number of zeroes of solutions of the *Jacobi equation*

$$Ju = 0, \qquad u(0) = \dot{u}(0) = 0. \tag{27}$$

This will be made more precise in the following theorem. A point $c \in (0, l)$ is a *conjugate point of J with multiplicity m* if the space of solutions of (27) which vanish to first order at c has dimension $m > 0$. Of course, the multiplicity of a conjugate point of J is either 1 or 2 since J is a fourth-order operator.

Theorem 3.1. *The index of $Q(u)$ is equal to the number of conjugate points of J counted with multiplicity.*

Proof. Suppose the index of $Q(u)$ is n, and let

$$\lambda_1 \leq \cdots \leq \lambda_n < 0$$

be the negative eigenvalues of J. For each $x \in (0, l]$, let $\Lambda_1(x) \leq \Lambda_2(x) \leq \cdots$ denote the eigenvalues, in ascending order, of the operator J on the space $C_0^\infty[0, x]$. Note that $\Lambda_i(l) = \lambda_i$ for each $1 \leq i \leq n$. It follows from general theory that each $\Lambda_i(x)$ is a strictly decreasing continuous function on $(0,l]$ with $\lim_{x\to 0^+}\Lambda_i(x)=\infty$ (see [14], for example). Therefore, each $\Lambda_i(x)$ has exactly one root c_i. We claim these roots

$$0 < c_1 \leq c_2 \leq \cdots \leq c_n < l$$

are precisely the conjugate points of J between 0 and l. To see this, note that $\Lambda_i(c_i)=0$ by definition implies there exists a $u_i \in C_0^\infty[0,c_i]$ satisfying $Ju_i=0$. By extending u_i to a solution of $Ju=0$ on $[0,l]$ we get a solution of (27) vanishing to first order at c_i. Conversely, if $c\in(0,l)$ is a conjugate point, then 0 is an eigenvalue of J on $C_0^\infty[0,c]$ and so $\Lambda_j(c)=0$ for some positive integer j. Because $\Lambda_j(x)$ strictly decreases, it follows that $\Lambda_j(l)$ is equal to one of the negative eigenvalues, say λ_j, of J, and therefore $c=c_j$. □

The point $\gamma(c)$ along a geodesic γ is a *conjugate point of* γ if c is a conjugate point of the corresponding Jacobi operator J.

Corollary 3.1. *A geodesic no longer minimizes length beyond its first conjugate point.*

Proof. Suppose $\gamma(c)$ is the first conjugate point of the geodesic γ. By Theroem 3.1, for every $l>c$ the index of $Q(u)$ on the space $C_0^\infty[0,l]$ is positive and so there is a function $u\in C_0^\infty[0,l]$ for which $Q(u)<0$. Setting $X_3=u$ and defining X_1,X_2,X_4,X_5 by equations (24) gives a direction $X=\sum_i X_i e_i$ along which the length functional $\mathcal{L}$ decreases. □

3.3. A Bonnet-Myers-type theorem.

We now seek sufficient conditions on the invariants to guarantee every geodesic reaches a conjugate point in finite time. We first formulate a comparison result that will allow us to estimate the maximum distance to the first conjugate point.

Proposition 3.3. *If the operator* $J_1=\frac{d^4}{ds^4}+\frac{d}{ds}(p_1\frac{d}{ds})+q_1$ *has a conjugate point on the interval* $(0,l]$ *and if* $p_2\geq p_1$ *and* $q_2\leq q_1$, *then the operator* $J_2=\frac{d^4}{ds^4}+\frac{d}{ds}(p_2\frac{d}{ds})+q_2$ *also has a conjugate point on the interval* $(0,l]$.

Proof. It follows from Theorem 3.1 that J has a conjugate point in the interval $(0,l]$ if and only if there is a function $u\in C_0^\infty[0,l]$ for which $Q(u)\leq 0$. The proposition now follows immediately from the observation

$$Q(u)=\int_0^l u\ddot{\ddot{u}}+u\frac{d}{ds}(p\dot{u})+uqu\,ds=\int_0^l \ddot{u}^2-p\dot{u}^2+qu^2\,ds.$$

□

In the case where p and q are constant, the solution to the Jacobi equation (27) may be computed explicitly. The proof of the following proposition makes use of the fact that c is a conjugate point of the Jacobi operator if and only if the wronskian of two independent solutions of the Jacobi equation vanishes at c.

Proposition 3.4. *Suppose $p = p_0 > 0$ and $q = q_0 \geq 0$ are constant. The Jacobi operator J has a conjugate point if and only if $4q_0 < p_0^2$. In this case, the first conjugate point occurs no later than $2\pi/\sqrt{p_0 - 2\sqrt{q_0}}$. Furthermore, if $q_0 = 0$, then the first conjugate point is precisely $2\pi/\sqrt{p_0}$.*

Proof. If $4q_0{=}p_0^2$, then two independent solutions of the Jacobi equation (27) are $\mu(s) = -\kappa s\cos(\kappa s){+}\sin(\kappa s)$ and $\nu{=}s\sin(\kappa s)$, where $\kappa{=}\sqrt{p_0/2}$. We compute the wronskian to be $\kappa^2 s^2 + \cos^2(\kappa s) - 1$, which is equal to 0 if and only if $\kappa s{=}0$, i.e., only at $s{=}0$.

It follows from Proposition 3.3 that there is also no conjugate point if $4q_0{>}p_0^2$.

However, if $q_0{=}0$, then two independent solutions of (27) are $\mu(s){=}1{-}\cos(\sqrt{p_0}s)$ and $\nu(s) = \sin(\sqrt{p_0}s){-}\sqrt{p_0}s$. We compute the wronskian to be

$$-\sqrt{p_0}(\sqrt{p_0}s\,\sin(\sqrt{p_0}s) + 2\cos(\sqrt{p_0}s) - 2).$$

The first positive root is seen to be $2\pi/\sqrt{p_0}$.

Finally, if $4q_0{<}p_0^2$, but $q_0{\neq}0$, then two independent solutions of (27) are $\mu(s) = \cos(\kappa_1 s) - \cos(\kappa_2 s)$ and $\nu(s){=}\kappa_2\sin(\kappa_1 s){-}\kappa_1\sin(\kappa_2 s)$, where $p_0{=}\kappa_1^2{+}\kappa_2^2$ and $q_0{=}\kappa_1^2\kappa_2^2$. The wronskian is now

$$w(s) = (\kappa_1^2 + \kappa_2^2)\sin(\kappa_1 a)\sin(\kappa_2 s) + 2\kappa_1\kappa_2\cos(\kappa_1 s)\cos(\kappa_2 s) - 2\kappa_1\kappa_2.$$

Using a few trig identities, it can be shown that w vanishes at c if and only if the function

$$W(s) = (\kappa_2 - \kappa_1)^2\sin^2\left((\kappa_2 + \kappa_1)\frac{s}{2}\right) - (\kappa_2 + \kappa_1)^2\sin^2\left((\kappa_2 - \kappa_1)\frac{s}{2}\right)$$

vanishes at c. We may assume $\kappa_2{>}\kappa_1$, and let $\tau{=}(\kappa_1{+}\kappa_2)c/2$ and $\rho{=}(\kappa_2{-}\kappa_1)/(\kappa_2{+}\kappa_1)$. Now W vanishes at c if and only if $\sin(\tau){=}\pm\sin(\rho\tau)/\rho$. Because $0{<}\rho{<}1$ it is clear that the first positive τ for which $\sin(\tau){=}\pm\sin(\rho\tau)/\rho$ is less than or equal to the first root π/ρ of the function $\sin(\rho\tau)$. It follows that the first conjugate point c is less than or equal to $2\pi/\sqrt{p_0{-}2\sqrt{q_0}}$. □

As a consequence of the previous two propositions, if $p_0{>}0$ and $q_0{\geq}0$ are any pair of constants that satisfy $p_0{-}2\sqrt{q_0}{>}0$, then every geodesic for which $p{\geq}p_0$ and $q{\leq}q_0$ reaches a conjugate point no later than $2\pi/\sqrt{p_0{-}2\sqrt{q_0}}$.

For instance, if $(M, D, \langle\ ,\ \rangle)$ is a sub-Riemannian manifold for which $a_1^2{+}a_2^2 = 0$, then $\lambda = \lambda_0$ is constant and the functions p and q are $p = K{+}\lambda_0^2$ and $q = 0$. If K is bounded below by a positive constant, say $K \geq K_0 > 0$, then p is bounded below by $K_0{+}\lambda_0^2$. It follows that every geodesic whose I-lift has initial curvature λ_0 has a conjugate point no later than $2\pi/\sqrt{K_0{+}\lambda_0^2}$ and therefore every geodesic reaches a conjugate point no later than $2\pi/\sqrt{K_0}$.

Unfortunately, in the case $a_1^2 + a_2^2 > 0$, q is quadratic in λ and will grow larger than any fixed constant as λ gets larger. What we need is an estimate for λ which will allow us to find a lower bound on p and an upper bound on q.

A somewhat crude estimate is obtained by noting that $\dot{\lambda} = a_2$; if the function $\varphi_1 = \sqrt{a_1^2{+}a_2^2}$ is bounded above by a positive constant c_1, then $|\dot{\lambda}| \leq c_1$ implies $\lambda_0{-}c_1 l{\leq}\lambda{\leq}\lambda_0{+}c_1 l$ on the interval $[0, l]$. Let Ψ be the function

$$\Psi = \begin{cases} (\lambda_0 - c_1 l)^2 & \text{if } \lambda_0 \geq c_1 l \\ 0 & \text{if } \lambda_0 \leq c_1 l \end{cases},$$

and assume K is bounded below, say $K \geq K_0 > 0$. We have the following bound on p along every geodesic of length l whose I-lift has initial curvature λ_0:

$$p \geq K_0 - 3c_1 + \Psi. \tag{28}$$

If K is at least as big as $2\varphi_1$, then $K(a_1+\varphi_1)-a_2^2$ is at least as big as $2\varphi_1(a_1+\varphi_1)-(\varphi_1^2-a_1^2)$, which is equal to $(a_1+\varphi_1)^2$ and therefore is non-negative. In particular, if $K_0 \geq 2c_1$, then the term $2a_1^2+4a_2^2-2a_1K$ in q is bounded above by $2c_1^2+2c_1K_0$. If also $|B_1|$ and $|B_{12}-2A_{11}-2B_{21}|$ are bounded above by positive constants c_2 and c_3, respectively, then we have the following bound on q along every geodesic of length l whose I-lift has initial curvature λ_0:

$$q \leq 2c_1^2 + 2c_1K_0 + c_3 + 6c_1(\lambda_0 + c_1 l)^2 + 6c_2(\lambda_0 + c_1 l). \tag{29}$$

Now the functions B_1 and $B_{12}-2A_1-2B_{21}$ are functions on the bundle $\mathcal{B}_2$. A quick inspection of the structure equations $da_2=2a_1+\sum B_i\omega^i$, $dA_i=A_{i0}\alpha+\sum A_{ij}\omega^j$, and $dB_i=B_{i0}\alpha+\sum B_{ij}\omega^j$ reveals that these functions are not well defined on the base manifold M. Referring back to the spin calculations of §2.5, equations (11 and (12) imply that B_1 is the sum $B_1=b_I^1+b_I^2$ of the imaginary parts of b^1 and b^2. It follows that $|B_1|$ is bounded from above by the well defined function on M

$$\varphi_2 = \sqrt{2(b^1\bar{b}^1 + b^2\bar{b}^2)}. \tag{30}$$

Similarly, equations (14) imply $B_{12}-2A_{11}-2B_{21}= - 3b_R^{11}-5b_R^{12}+b_R^{21}-b_R^{22}$, and therefore $|B_{12} - 2A_{11} - 2B_{21}|$ is bounded from above by the well defined function on M

$$\varphi_3 = -3b_R^{11} + \sqrt{2(25b^{12}\bar{b}^{12} + b^{21}\bar{b}^{21} + b^{22}\bar{b}^{22} - b^{21}\bar{b}^{22} - \bar{b}^{21}b^{22})}. \tag{31}$$

Thus the inequality (29) holds if $\varphi_2 \leq c_2$ and $\varphi \leq c_3$.

Set

$$p_0 = K_0 - 3c_1 + \Psi$$

and

$$q_0 = 2c_1^2 + 2c_1K_0 + c_3 + 6c_1(\lambda_0 + c_1 l)^2 + 6c_2(\lambda_0 + c_1 l).$$

If there is a value for l, say L, for which the function $p_0-2\sqrt{q_0}$ is bounded below by the constant $4\pi^2/L^2$, for all positive λ_0, then Propositions 3.3 and 3.4 imply that every geodesic has a conjugate point on the interval $[0, L]$.

Theorem 3.2. *Let $(M^3, D, \langle\ ,\ \rangle)$ be a sub-Riemannian manifold whose invariants φ_1, φ_2, and φ_3 are bounded above by positive constants c_1, c_2, and c_3, respectively, and K is bounded below by the constant K_0, where*

$$K_0 \geq \max\left\{38c_1 + 2\sqrt{74c_1^2 + c_3}\ ,\quad \frac{3}{4}\left(\frac{c_2}{c_1}\right)^2 - \frac{c_3}{2c_1} - c_1\right\}.$$

Then every geodesic has a conjugate point on the interval $[0, L]$, where

$$L = \frac{3\pi}{\sqrt{K_0 - 9c_1 - 2\sqrt{2c_1^2 + 2c_1K_0 + c_3}}}.$$

Proof. By the discussion immediately preceding the theorem, it suffices to show that the hypothesis implies that for all positive λ_0 the function $p_0 - 2\sqrt{q_0}$ is bounded below by $4\pi^2/L^2$.

For notational convenience, set $y = c_1 l,\ x = \lambda_0,\ p_0' = K_0 - 3c_1,\ q_0' = 2c_1^2 + 2c_1 K_0 + c_3$, and

$$\begin{aligned} F(x,y) &= 4\pi^2 c_1^2 - p_0 y^2 + 2\sqrt{q_0}\, y^2 \\ &= 4\pi^2 c_1^2 + 2y^2\sqrt{q_0' + 6c_1(x+y)^2 + 6c_2(x+y)} - y^2 p_0' - \Psi y^2. \end{aligned}$$

Now $L = 3\pi/\sqrt{p_0' - 6c_1 - 2\sqrt{q_0'}}$ and we need to show that $F(x, c_1 L) \leq 0$ for all x.

Along each ray $x = \mu y$, $0 \leq \mu \leq 1$, the function F is

$$F = 4\pi^2 c_1^2 + 2y^2\sqrt{q_0' + 6c_1(\mu+1)^2 y^2 + 6c_2(\mu+1)y} - y^2 p_0'.$$

This function strictly increases in μ; thus it suffices to consider F on the region $x \geq y$.

On the region $x \geq y$, we have

$$F(x,y) = 4\pi^2 c_1^2 + 2y^2\sqrt{q_0' + 6c_1(x+y)^2 + 6c_2(x+y)} - y^2 p_0' - (x-y)^2 y^2.$$

The inequality

$$K_0 \geq \frac{3}{4}\left(\frac{c_2}{c_1}\right)^2 - \frac{c_3}{2c_1} - c_1$$

is equivalent to the condition $9c_2^2 \leq 6c_1 q_0'$. Under this condition,

$$F(x,y) \leq 4\pi^2 c_1^2 + 2y^2\left(\sqrt{6c_1}(x+y) + \sqrt{q_0'}\right) - y^2 p_0' - (x-y)^2 y^2.$$

For each fixed y this function has a maximum at $x = y + \sqrt{6c_1}$. Along the line $x = y + \sqrt{6c_1}$, the function F is

$$F = 3\sqrt{6c_1}\, y^3 - (p_0' - 2\sqrt{q_0'} - 6c_1)y^2 + 4\pi^2 c_1^2.$$

We find this cubic is negative at $y = c_1 L$ if and only if

$$p_0' - 2\sqrt{q_0'} \geq 6c_1 + \left(\frac{81\sqrt{6}\pi}{5}\right)^{2/3} c_1.$$

It remains only to show that this condition is guaranteed by the inequality

$$K_0 \geq 38c_1 + 2\sqrt{74c_1^2 + c_3}$$

in the hypothesis of the theorem.

The condition $p_0' - 2\sqrt{q_0} \geq 6c_1 + (81\sqrt{6}\pi/5)^{2/3} c_1$ is implied by $p_0' - 2\sqrt{q_0'} \geq 31c_1$, which is equivalent to $2\sqrt{2c_1^2 + 2c_1 K_0 + c_3} \leq K_0 - 34c_1$. Squaring both sides and collecting terms, this inequality becomes $0 \leq K_0^2 - 76c_1 K_0 + 1148c_1^2 - 4c_3$, and finally by the quadratic formula, this is implied by $K_0 \geq 38c_1 + 2\sqrt{74c_1^2 + c_3}$. □

If the sub-Riemannian manifold $(M^3, D, \langle\ ,\ \rangle)$ is complete, so that every pair of points can be joined by a minimizing sub-Riemannian geodesic, then we have the following corollary:

Corollary 3.2. *Let $(M^3, D, \langle\ ,\ \rangle)$ be a complete sub-Riemannian manifold. Under the hypothesis of the Theorem 3.2, M is compact, and in fact the sub-Riemannian diameter is no greater than*

$$L = \frac{3\pi}{\sqrt{K_0 - 9c_1 - 2\sqrt{2c_1^2 + 2c_1 K_0 + c_3}}}.$$

Proof. Every pair of points can be joined by a minimizing geodesic; on the other hand, by the theorem, no geodesic of length greater than L is length-minimizing. Therefore every pair of points must be within this distance of each other. □

In the case $a_1^2 + a_2^2 = 0$, this result may be slightly improved by the discussion following Proposition 3.4:

Proposition 3.5. *If $(M^3, D, \langle\ ,\ \rangle)$ is a complete sub-Riemannian manifold with $a_1^2 + a_2^2 = 0$ and $K \geq K_0 > 0$, then M is compact with sub-Riemannian diameter no greater than $2\pi/\sqrt{K_0}$.*

3.4. The exponential map. The solutions to the geodesic equations (18) are completely determined by specifying an initial point $p \in M^3$ and an initial value (θ_0, λ_0) in the fiber $S^1 \times \mathbb{R}$ over p. We denote by $\gamma_{\theta_0,\lambda_0}(s)$ the unique geodesic whose I-lift has initial point $(p, \theta_0, \lambda_0) \in Y$. The (sub-Riemannian) *exponential map*

$$\exp_p : \mathbb{R} \times S^1 \times \mathbb{R} \longrightarrow M^3$$

is defined by $\exp_p(s, \theta_0, \lambda_0) = \gamma_{\theta_0,\lambda_0}(s)$.

The next proposition shows that $\exp_p$ is never a local diffeomorphism at any point p.

Proposition 3.6. *For every $p \in M^3$, the differential $d\exp_p$ is singular at each of the points $(0, \theta_0, \lambda_0)$ in the cylinder $\{0\} \times S^1 \times \mathbb{R}$.*

Proof. The exponential map $\exp_p$ lifts in an obvious way to a map $\mathrm{Exp}_p : \mathbb{R} \times S^1 \times \mathbb{R} \longrightarrow Y$ defined by $\mathrm{Exp}_p(s, \theta_0, \lambda_0) = (\gamma_{\theta_0,\lambda_0}(s), \theta(s), \lambda(s))$, the unique solution to the geodesic equations (18) with initial conditions $\theta(0) = \theta_0$, $\lambda(0) = \lambda_0$, and $\gamma_{\theta_0,\lambda_0}(0) = p$. Let X_i and W_i be the functions on $\mathbb{R} \times S^1 \times \mathbb{R}$ defined by

$$\mathrm{Exp}_p^* \begin{Bmatrix} \omega^1 \\ \omega^2 \\ \omega^3 \\ \omega^4 \\ \omega^5 \end{Bmatrix} = \begin{Bmatrix} ds + X_1 d\theta_0 + W_1 d\lambda_0 \\ X_2 d\theta_0 + W_2 d\lambda_0 \\ X_3 d\theta_0 + W_3 d\lambda_0 \\ X_4 d\theta_0 + W_4 d\lambda_0 \\ X_5 d\theta_0 + W_5 d\lambda_0 \end{Bmatrix}. \tag{32}$$

By definition of Exp_p, these functions satisfy the initial conditions

$$\begin{aligned} X_i(0, \theta_0, \lambda_0) &= W_i(0, \theta_0, \lambda_0) = 0 \qquad \text{for } i = 1, 2, 3 \\ -X_4(0, \theta_0, \lambda_0) &= W_5(0, \theta_0, \lambda_0) = 1 \\ X_5(0, \theta_0, \lambda_0) &= W_4(0, \theta_0, \lambda_0) = 0. \end{aligned} \tag{33}$$

Note that Exp_p pulls back the form $\omega^1\wedge\omega^2\wedge\omega^3$ to be $(X_2W_3-X_3W_2)\,ds\wedge d\theta_0\wedge d\lambda_0$. On the other hand, this is equal to the pull-back by $\exp_p$ of the volume form $\eta^1\wedge\eta^2\wedge\eta^3$ on M. Therefore $d(\exp_p)$ is singular at (c,θ_0,λ_0) if and only if $X_2W_3-X_3W_2$ vanishes at (c,θ_0,λ_0).

By a similar computation to the one in the proof of Proposition 3.2, the structure equations (22) imply that X_i and W_i satisfy the system of differential equations (23) together with the additional equation $\dot{X}_5=2a_1X_4+(B_3+a_2^2)X_3+(B_2+a_2\lambda)X_2$. Expanding the solution to this system of o.d.e.'s with the initial conditions (33) in a power series, we find

$$X_2W_3-X_3W_2=-\frac{1}{12}s^4+\frac{1}{180}(-3a_1(0)+K(0)+\lambda_0^2)s^6+\mathrm{O}(s^7),$$

and therefore $d\exp_p$ is singular at every point $(0,\theta_0,\lambda_0)$. □

In analogy with Riemannian geometry, we say the point $(c,\theta_0,\lambda_0)\in\mathbb{R}\times S^1\times\mathbb{R}$ is a *conjugate point of p* if the differential $d(\exp_p)$ is singular there.

By varying θ_0 and λ_0, Exp_p can be thought of as a two-parameter admissable variation of the lifted geodesic $\mathrm{Exp}_p(s,\theta_0,\lambda_0)$ with infinitesimal admissable variations, say, $X=\sum X_ie_i$ and $W=\sum W_ie_i$. This variation Exp_p is actually a variation through integral curves of I, and thus the hessian $\mathcal{L}_{**}(X,W)$ is *a priori* zero. Therefore $X_3(s,\theta_0,\lambda_0)$ and $W_3(s,\theta_0,\lambda_0)$ are solutions of the Jacobi equation (27). Furthermore, the initial conditions (33) imply they are independent solutions. Because $X_2=\dot{X}_3$ and $W_2=\dot{W}_3$, the function $X_2W_3-X_3W_2$ is the wronskian of the independent solutions X_3 and W_3. Thus, the point (c,θ_0,λ_0) is a conjugate point of p if and only if $\gamma_{\theta_0,\lambda_0}(c)$ is a conjugate point of the geodesic $\gamma_{\theta_0,\lambda_0}$.

Because the exponential map is never a local diffeomorphism there is no "totally normal neighborhood" around any point in M. Consequently, we shouldn't expect there to be a neighborhood U of p with the property that for every point q in U there is a unique length-minimizing geodesic joining p to q. Indeed, we will show such a neighborhood U does not exist. This result is a consequence of the following proposition.

Proposition 3.7. *For each $r>0$, there is a positive Λ_0 with the property that for all θ_0 and all $|\lambda_0|\geq\Lambda_0$ the geodesic $\gamma_{\theta_0,\lambda_0}$ does not minimize length beyond r.*

Proof. We will show for each r, there is a Λ_0 with the property that if $|\lambda_0|\geq\Lambda_0$ then the Jacobi operator along the geodesic $\gamma_{\theta_0,\lambda_0}$ has a conjugate point on the interval $(0,r]$.

It suffices to prove this proposition for small r, for if $\gamma_{\theta_0,\lambda_0}$ does not minimize length on the interval $[0,r]$, then it obviously does not minimize length on $[0,r']$ for any $r'\geq r$. Assume r is small enough so that the closed ball $B_r(p)$ is compact. There exists positive constants K_0,c_1,c_2, and c_3 so that $|K|\leq K_0$ and $\varphi_i\leq c_i$ on $B_r(p)$. Furthermore, the estimate $\lambda_0-c_1r\leq\lambda\leq\lambda_0+c_1r$ holds on the interval $[0,r]$. Set

$$\begin{aligned}p_0&=-K_0-3c_1+(\lambda_0-c_1r)^2\\q_0&=4c_1^2+2c_1K_0+c_3+6c_1(\lambda_0+c_1r)^2+6c_2(\lambda_0+c_1r).\end{aligned}$$

If $\lambda_0\geq c_1r$, then equations (19) and (20) imply that along the interval $[0,r]$,

$$p\geq p_0\quad\text{and}\quad q\leq q_0.$$

Because p_0 and q_0 are quadratic in λ_0, it is clear that for each r there is a Λ_0 with the property that if $\lambda_0 \geq \Lambda_0$ then $p_0 - 2\sqrt{q_0}$ is bounded below by $4\pi^2/r^2$ on the interval $[0, r]$. Propositions 3.3 and 3.4 imply that if $\lambda_0 \geq \Lambda_0$, then $\gamma_{\theta_0,\lambda_0}$ reaches a conjugate point by r, and hence does not minimize length beyond r. □

Fix a point p in M^3. For each positive r, let $A_r(p)$ be the subset of the cylinder $\{r\} \times S^1 \times \mathbb{R}$

$$A_r(p) = \{(r, \theta_0, \lambda_0) \mid \gamma_{\theta_0,\lambda_0} \text{ is length-minimizing on the interval } [0, r]\,\}\,.$$

This set $A_r(p)$ is clearly closed, and by the proposition above it is also bounded and is therefore compact.

Let $S_r(p)$ denote the sphere of radius r about p. It is clear that the image $\exp_p(A_r(p))$ is contained in $S_r(p)$ for every r. Choose r small enough so that $S_r(p)$ is contained in a coordinate neighborhood of M. Then $S_r(p)$ is compact, and every point in $S_r(p)$ can be joined to p by a length-minimizing geodesic. It follows that the map $\exp_p : A_r(p) \to S_r(p)$ is surjective. Now if this map were also injective, then it would in fact be a homeomorphism, which is impossible. Consequently, for small enough r there is at least one point q on the sphere $S_r(p)$ with the property that there is more than one length-minimizing geodesic joining p to q.

As in Riemannian geometry, we define the *cut locus C(p) of* p to be the set of points q for which there exists a geodesic starting at p and passing through q with the property that q is the first point along the geodesic where the geodesic ceases to minimize length. By the definition of $A_r(p)$, it is clear that the exponential image of the boundary $\partial A_r(p)$ is contained in the cut locus of p. In particular, every sphere $S_r(p)$ contains at least one cut point of p.

We say $(M^3, D, \langle\ ,\ \rangle)$ is *geodesically complete* if every geodesic can be extended indefinitely, i.e., if the exponential map is defined on all of $\mathbb{R} \times S^1 \times \mathbb{R}$. The following theorem is the analog of the Hopf-Rinow Theorem in Riemannian geometry.

Theorem 3.3. *The sub-Riemannian manifold* $(M^3, D, \langle\ ,\ \rangle)$ *is complete if and only if it is geodesically complete. In this case, every pair of points may be joined by a length-minimizing geodesic.*

Proof. Assume $(M^3, D, \langle\ ,\ \rangle)$ is complete, and let γ be a geodesic that is defined for $0 \leq s < \sigma$. Let $\{s_n\}$ be a convergent sequence, converging to σ, with $s_n < \sigma$. Since $d(\gamma(s_n), \gamma(s_m)) \leq |s_n - s_m|$, it follows that the sequence $\{\gamma(s_n)\}$ is Cauchy, and hence converges to some $p \in M^3$. We need only show the curvature λ remains bounded as $s \to \sigma$. Consider the closed ball $B_\sigma(p)$ of radius σ centered at p. By the completeness assumption, closed balls are compact. Thus the function $\sqrt{a_1^2 + a_2^2}$ is bounded on $B_\sigma(p)$ by a non-negative constant c_1. The equation $\dot{\lambda} = a_2$ now implies

$$\lambda_0 - c_1\sigma \leq \lambda \leq \lambda_0 + c_1\sigma$$

on the interval $[0, \sigma]$, and so $\lambda(\sigma)$ is bounded.

Conversely, if $(M^3, D, \langle\ ,\ \rangle)$ is geodesically complete, then every pair of points may be joined by a length-minimizing geodesic. The proof of this statement is essentially the same as de Rahm's proof of the statement in the Riemannian case (see for example [4]), and will be omitted. Consider the closed ball $B_r(p)$ of radius r. If $B_r(p)$ is compact for every r, then by Proposition

1.1 $(M^3, D, \langle\ ,\ \rangle)$ is complete. Since every point in $B_r(p)$ may be joined to p by a length-minimizing geodesic, there is a set $A \subset [0,r]\times S^1\times\mathbb{R}$ satisfying $B_r(p) \subset \exp_p(A)$. By Proposition 3.7 we may assume A is bounded and hence has compact closure. Therefore the closed ball $B_r(p)$, being contained in the compact set $\exp_p(\bar{A})$, is compact. □

3.5. Geodesics on homogeneous manifolds. We now describe the geodesics of the homogeneous manifolds $(M^3, D, \langle\ ,\ \rangle)$ in some detail. Here M is a Lie group and $\langle\ ,\ \rangle$ is a positive definite inner product on a Lie subalgebra D.

For homogeneous manifolds with $a_1^2+a_2^2 = 0$, we may choose a basis $\{\xi_1, \xi_2, \xi_3\}$ of the Lie algebra $\mathfrak{g}$ with bracket relations

$$([\xi_2,\xi_3] \quad [\xi_3,\xi_1] \quad [\xi_1,\xi_2]) = (\xi_1 \quad \xi_2 \quad \xi_3)\begin{pmatrix} K_0 & 0 & 0\\ 0 & K_0 & 0\\ 0 & 0 & 1\end{pmatrix}$$

that induces the sub-Riemannian structure, i.e., $\{\xi_1, \xi_2\}$ forms an orthonormal basis of D. With this basis the geodesic equations (18) can be written

$$\dot\gamma = L'_{\gamma(s)}(\cos\theta\,\xi_1 + \sin\theta\,\xi_2), \quad \dot\theta = -\lambda, \quad \dot\lambda = 0,$$

where L_p denotes left multiplication by p. Note that $\cos\theta\xi_1 + \sin\theta\xi_2$ is a periodic curve on $\mathfrak{g}$ with period $2\pi/\lambda_0$.

Consider the Heisenberg group equipped with the sub-Riemannian structure for which $a_1^2+a_2^2$ and K_0 are both zero. The geodesic equations may be explicitly integrated, and we find the geodesic $\gamma_{\theta_0,\lambda_0}(s)$ starting at $(0,0,0)$ is given in coordinates by

$$x(s) = \frac{\sin\theta_0 - \sin(\theta_0 - \lambda_0 s)}{\lambda_0}, \quad y(s) = \frac{\cos(\theta_0-\lambda_0 s) - \cos\theta_0}{\lambda_0}, \quad z(s) = \frac{\lambda_0 s - \sin\lambda_0 s}{2\lambda_0^2}.$$

The Jacobian of the exponential map $\exp_p$ is computed to be

$$(\lambda_0 s\sin\lambda_0 s + 2\cos\lambda_0 s - 2)/\lambda_0^4.$$

Thus the first conjugate point along $\gamma_{\theta_0,\lambda_0}(s)$ occurs at $2\pi/\lambda_0$. Since $x(2\pi/\lambda_0) = y(2\pi/\lambda_0) = 0$ and $z(2\pi/\lambda_0) = \pi/\lambda_0^2$, we see that the first conjugate locus of $(0,0,0)$ is the entire z-axis. This is precisely the cut locus of (0,0,0); after all, every circle in the plane ceases to minimize length among curves enclosing a fixed area immediately after it first closes up.

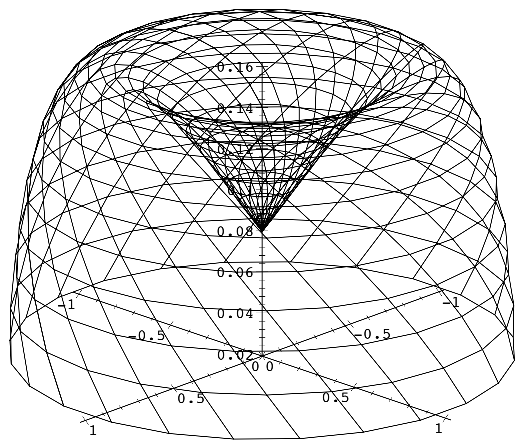
Thus in the Heisenberg case the conjugate locus and the cut locus coincide. Furthermore, there are precisely two cut points of (0,0,0) on every sphere $S_r(0,0,0)$, at $(0,0,\pm r)$.

We now look more closely at the sphere $S_r(0,0,0)$ of radius r:

$$S_r(0,0,0) = \left\{\gamma_{\theta_0,\lambda_0}(r) \mid 0 \le \theta_0 \le 2\pi,\ -\frac{2\pi}{r} \le \lambda_0 \le \frac{2\pi}{r}\right\}.$$

The distance from (0,0,0) to a point $(x,y,0)$ in the xy-plane is equal to $\sqrt{x^2+y^2}$, however the distance from (0,0,0) to a point $(0,0,z)$ along the z-axis is equal to $\sqrt{4\pi z}$. Therefore $S_r(0,0,0)$ intersects the xy-plane in a circle of radius r and it intersects the z-axis at $\pm r^2/4\pi$.

The following is a picture of the upper half of the sphere $S_1(0,0,0)$. Note the singularity at the z-axis:



Consider now the group $SO(3)$ equipped with the sub-Riemannian structure for which $a_1^2+a_2^2=0$ and $K_0=1$. We identify $SO(3)$ with the bundle $\mathcal{F}_{S^2}$ of oriented orthonormal frames of S^2. The sub-Riemannian structure is the one induced by the unique coframing (ω^1,ω^2,ρ) on $\mathcal{F}_{S^2}$ where ρ is the connection form for $\mathcal{F}_{S^2}\to S^2$ and (ω^1,ω^2) is the tautological one-form.

Every D-curve γ on $SO(3)$ is the lift of a regular curve $\bar\gamma$ on S^2 by parallel transport: $\gamma(s)=(\bar\gamma(s);e_1(s))$, where $e_1(s)$ is the parallel transport of the initial vector $e_1(0)$ along $\bar\gamma$. Note that the geodesics on $SO(3)$ are the lifts of circular arcs on S^2.

Because $K\equiv 1$ and $a_1^2+a_2^2=0$, we see the Jacobi operator along a geodesic γ with curvature λ_0 has its first conjugate point at $2\pi/\sqrt{1+\lambda_0^2}$. This is precisely where the projected curve $\bar\gamma$ first closes up. Therefore the conjugate locus of $(\bar p,e_1)$ consists of the S^1 fiber over $\bar p$. It can also be shown that the cut locus of $(\bar p;\,e_1)$ consists of the S^1 fiber over $\bar p$, the S^1 fiber over the antipodal point of $\bar p$, and the set of $(\bar q;\,e_1')$ such that the angle formed by the parallel transport of e_1 along the great circle joining $\bar p$ to $\bar q$ and the vector e_1' is equal to π.

For homogeneous manifolds with $a_1^2+a_2^2\neq 0$, we may choose a basis $\{\xi_1,\xi_2,\xi_3\}$ of the Lie algebra $\mathfrak{g}$ with bracket relations

$$\left([\xi_2,\xi_3]\quad [\xi_3,\xi_1]\quad [\xi_1,\xi_2]\right)=\left(\xi_1\quad \xi_2\quad \xi_3\right)\begin{pmatrix}\lambda_3+\varphi_1 & 0 & -\lambda_1\\ 0 & \lambda_3-\varphi_1 & -\lambda_2\\ 0 & 0 & 1\end{pmatrix}$$

that induces the sub-Riemannian structure. With this basis the geodesic equations (18) can be written

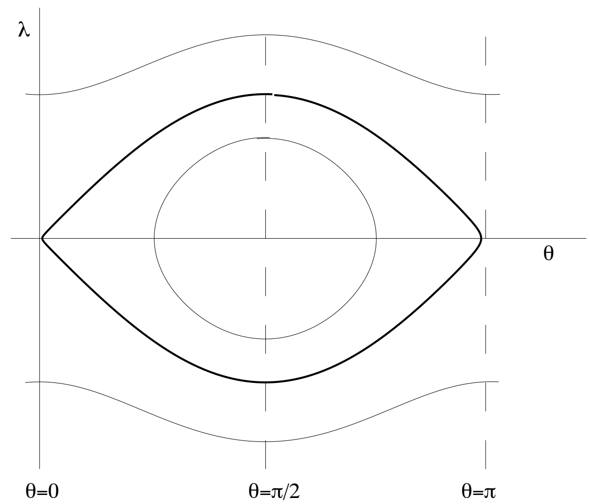
$$\dot\gamma=L'_{\gamma(s)}(\cos\theta\,\xi_1+\sin\theta\,\xi_2),\quad \dot\theta=\lambda_1\cos\theta+\lambda_2\sin\theta-\lambda,\quad \dot\lambda=-\varphi_1\sin 2\theta.$$

Consider the group $E(2)$ of rigid motions of the Euclidean plane equipped with the sub-Riemannian structure for which $\varphi_1=\lambda_3=1/2$ and $\lambda_1=\lambda_2=0$. This is case 1 b) of §2.3; identifying

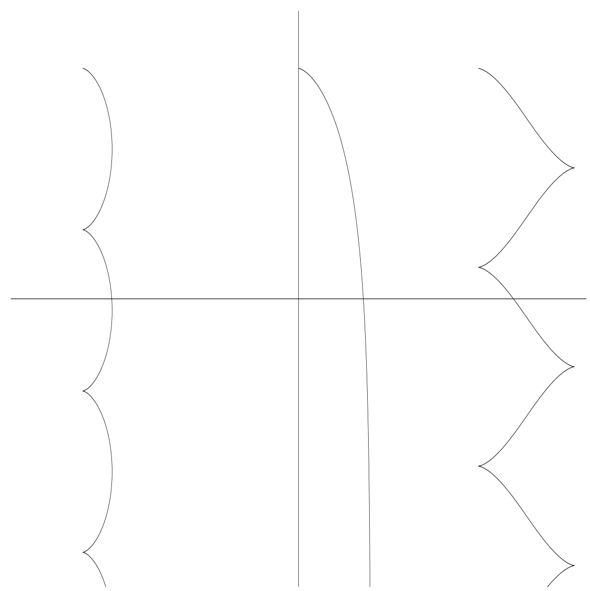
$E(2)$ with $\mathbb{R}^2 \times S^1$, with coordinates (x, y, ϕ), the geodesic equations can be written

$$\dot{x} = \cos\theta\cos\phi, \quad \dot{y} = \cos\theta\sin\phi, \quad \dot{\phi} = -\sin\theta, \quad \dot{\theta} = -\lambda, \quad \dot{\lambda} = -\frac{1}{2}\sin 2\theta.$$

The last two equations imply $\lambda^2 = c - \frac{1}{2}\cos 2\theta$, where $c \geq -1/2$; the phase portrait is shown in the following diagram:



Plotted below are three solutions $(x(t), y(t))$ of the geodesic equations with $\theta_0 = \pi/2$. The curve on the left has $0<\lambda_0<1$ (so that (θ, λ) lies inside the separatrix), the middle curve has $\lambda_0=1$ (so that (θ, λ) lies on the separatrix), and the curve on the right has $\lambda_0>1$ (so that (θ, λ) lies outside the separatrix):



Remark 3. *Note that these curves have kinks; this should be compared with the minimizing curves of the functional* $\int \kappa^2$*. Curves that minimize this functional are known to be regular* [2]*. It is surprising that the minimizing curves for these two functionals behave so differently.*

By examining the Jacobi operator, it can be shown that for λ_0 very near zero, the geodesic $\gamma_{\pi/2,\lambda_0}$ has no conjugate points. It can also be shown that for λ_0 bigger than about 16, the geodesic $\gamma_{\pi/2,\lambda_0}$ has a conjugate point no later than $2\pi/(\sqrt{\lambda_0 - 2 - 2\sqrt{3\lambda_0}})$.

4. Some Remarks on the Higher Dimensional Case

In this section, we discuss possible generalizations to higher dimensions and the difficulties encountered there due to the presence of non-regular curves and abnormal minimizers. Let $(M, D, \langle\ ,\ \rangle)$ be a sub-Riemannian manifold, where M has dimension n and D has rank $m > 0$. Denote by $\Omega_D(p, q)$ the space of D-curves joining p to q.

4.1. Non-regular curves. It is a fundamental problem to determine the topology of $\Omega_D(p, q)$. This problem is complicated by the presence of so-called "non-regular" curves. These are the curves at which "the natural candidate" for the tangent space $T_\gamma\Omega_D(p, q)$ fails to be the true tangent space [3]. Consequently, one needs to be careful when applying the techniques of the calculus of variations to find the extremals of the length functional $\mathcal{L}$ defined on $\Omega_D(p, q)$.

Recall as in the proof of Proposition 3.1 that associated to the variational problem $(M, D, \mathcal{L})$ there is a variational problem $(Z, \{0\}, \widetilde{\mathcal{L}})$ with the property that extremals of $(Z, \{0\}, \widetilde{\mathcal{L}})$ project to be extremals of $(M, D, \mathcal{L})$ [8]. The extremals of $(Z, \{0\}, \widetilde{\mathcal{L}})$ are precisely the integral curves of the Euler-Lagrange system on Z. The D-curves that are the projections of integral curves of the Euler-Lagrange system will be said to be *geodesics*. Every geodesic is an extremal, but the converse need not be true. In [13], Hsu shows that every *regular* extremal is a geodesic. For strongly bracket generating distributions (defined in §1.2 above), it can be shown that every D-curve is regular. For instance, contact distributions are strongly bracket generating. Distributions that are strongly bracket generating are rare, however: if D is strongly bracket generating of rank $m \neq n-1$, then m must be a multiple of 4 and also must be less than $n(n-1)/2$. Furthermore, if D is not strongly bracket generating, then *there are always non-regular curves* [3]. If one of these non-regular curves happens to be an extremal of $(M, D, \mathcal{L})$, there is no guarantee that it is a geodesic.

In [15], Liu and Sussman give an example of a length-minimizing curve on a sub-Riemannian Engel manifold that is not a geodesic. This curve is rigid, i.e., it is isolated in the space $\Omega_D(p, q)$ with the C^1-topology. Because rigid curves have no compact variations they are trivially local extremals of $(M, D, \mathcal{L})$; it is somewhat surprising that rigid curves can be global extremals. Generalizations of Theorem 3.2 to the higher dimensional case will thus require an estimate of how far one can take one of these abnormal minimizers before it is no longer minimizing. This is an interesting problem in itself, and is closely related to the problem of determining how far one can take a rigid curve and still have a rigid curve (see Bryant and Hsu [3]). Note that the length-minimizing D-curve given in the third example of §1.1 is also rigid. However, it is possible to show this rigid curve does satisfy the geodesic equations.

4.2. Smoothness of minimizers and completeness issues. Because the length-minimizing curves need not lift to solutions of the Euler-Lagrange equations, it is natural to ask if all the extremals of $(M, D, \mathcal{L})$ are smooth. If γ is a regular extremal, then it is a consequence of Pontryagin's Maximum Principle (see Strichartz [21], for example) that γ is smooth. If γ is a non-regular extremal however, then it need not be smooth. In fact, in [15] Liu and Sussman give an example of an abnormal extremal that is not even C^1. It is not known at this time whether this curve is locally optimal. See Montgomery [17] for an interesting discussion on a sufficient condition to guarantee smoothness of sub-Riemannian minimizers.

It may be more natural in the higher dimensional case to define a D-curve as a Lipschitz curve, rather than smooth curve, whose derivative lies in D (wherever it exists). Now the Carnot-Carathéodory distance $d(p, q)$ between two points p and q is the infimum of the lengths of Lipschitz curves tangent to D joining p to q. If p and q are contained in a neighborhood that has compact closure, then it follows that there is a length-minimizing D-curve joining p to q [21]. Thus, if $(M, D, \langle\ ,\ \rangle)$ is complete, then Proposition 1.1 implies that every pair of points may be joined by a length-minimizing D-curve. This length-minimizing D-curve need not be a geodesic. Referring to the proof of Theorem 3.3 it seems likely then that $(M, D, \langle\ ,\ \rangle)$ geodesically complete does not imply every pair of points may be joined by a length-minimizing geodesic or even by a length-minimizing D-curve. Consequently, it seems likely that geodesic completeness need not imply completeness.

Finally, it will be interesting to investigate the relation between cut points and conjugate points in the higher dimensional case. It is well known in Riemannian geometry that if $\gamma(t_0)$ is the cut point of $p = \gamma(0)$ along γ, then either a) $\gamma(t_0)$ is the first conjugate point of $\gamma(0)$ along γ or b) there exists a different geodesic α from p to $\gamma(t_0)$ such that the length of α equals the length of γ; conversely, if a) or b) is satisfied, then there exists t_1 in $(0, t_0]$ such that $\gamma(t_1)$ is the cut point of p along γ. This is probably not true in sub-Riemannian geometry. Certainly a) implies that γ reaches a cut point before t_0, but the proofs of the other implications in Riemannian geometry won't carry over to sub-Riemannian geometry because of the presence of abnormal minimizers.

References

1. Brockett, R.: *Control theory and singular Riemannian geometry*, New Directions in Applied Mathematics (Hilton and Young, eds.) Springer, Berlin, 1981.
2. Bryant, R. and Griffiths, P.: *Reduction for constrained variational problems and* $\int \frac{1}{2}\kappa^2\, ds$, American Journal of Math **108** (1986), 525-570.
3. Bryant, R. and Hsu, L.: *Rigidity of integral curves of rank two distributions*, Invent. Math. **114** (1993), 435-461.
4. Carmo, M. do: "Riemannian Geometry," Birkhäuser, Boston, 1993.
5. Chern, S.S. and Hamilton, R.: *On Riemannian metrics adapted to three-dimensional contact manifolds*, Lecture Notes in Mathematics **1111**, 1985.
6. Coppel, W.A.: *Disconjugacy*, Lecture Notes in Mathematics **220**, 1971.
7. Gardner, R.: "The Method of Equivalence and Its Applications," CBMS **58**, 1989.

8. Griffiths, P.: "Exterior Differential Systems and the Calculus of Variations," Progress in Mathematics **25**, Birkhäuser, 1983.
9. Gromov, M.: "Structures Métrique pour les Variétés Riemannienes," Cedic, Paris, 1981.
10. Gromov, M.: *Carnot-Carathéodory spaces seen from within*, preprint.
11. Hamenstädt, U.: *Some regularity theorems for Carnot-Carathéodory metrics*, J. Diff. Geom. **32** (1990), 819-850.
12. Hermann R.: *Geodesics of singular Riemannian metrics*, Bull. AMS **79** (1973), 780-782.
13. Hsu, L.: *Calculus of Variations via the Griffiths formalism*, J. Diff. Geom. **36** (1992), 551-589.
14. Kato, T.: "Perturbation Theory for Linear Operators," Springer-Verlag, Berlin, 1966.
15. Liu, W-S. and Sussman, H.: *Abnormal sub-Riemannian minimizers*, preprint.
16. Mitchell, J.: *On Carnot-Carathéodory metrics*, J. Diff. Geom. **21** (1985), 35-45.
17. Montgomery, R.: *Singular Extremals on Lie Groups*, preprint.
18. Montgomery, R.: *A survey of singular curves in sub-Riemannian geometry*, Journal of Dynamical and Control Systems **1** (1995), 49-90.
19. Ratcliffe, J.: "Foundations of Hyperbolic Manifolds," Graduate Texts in Math. **149**, Springer-Verlag, New York, 1994.
20. Rumin, M.: *Formes differentielles sur les varietes de contact*, J. Diff. Geom. **39** (1994), 281-330.
21. Strichartz, R.: *Sub-Riemannian geometry*, J. Diff. Geom. **24** (1986), 221-263.
22. Strichartz, R.: *Corrections to "Sub-Riemannian geometry"*, J. Diff. Geom. **30** (1989), 595-596.
23. Webster, S.: *Pseudo-Hermitian structures on a real hypersurface*, J. Diff. Geom. **13** (1978), 25-31.

Duke University
E-mail address: keen@math.duke.edu