A SURVEY OF SINGULAR CURVES IN SUBRIEMANNIAN GEOMETRY

Richard Montgomery
Mathematics Dept. UCSC, Santa Cruz, CA
email: rmont@cats.ucsc.edu

1995

Contents

1 INTRODUCTION

SubRiemannian geometry is the geometry of a distribution of k-planes on an n-dimensional manifold with a smoothly varying inner product on the k-planes. When $k = n$ we recover Riemannian geometry. Much of the work in subRiemannian geometry has concentrated on developing similarities with Riemannian geometry. However, there are major differences between the two geometries. The following phenomena occur in subRiemannian geometry but not in Riemannian geometry.

- The Hausdorff dimension is larger than the manifold dimension Mitchell [33] , Pansu [44].
- The conjugate locus of a point contains that point. That is to say, the exponential map is never a local diffeomorphism in a neighborhood of the point at which it is based [46].
- The space of paths tangent to the distribution and joining two fixed points can have singularities. These singular curves can be minimizing geodesics, independent of the choice of inner product [34], [8], [31].

In this review we concentrate on this last phenomenon.

In the next section we give basic examples, definitions, and theorems of the subject, and set up our notation. We also illustrate the first two phenomena. §3 describes subRiemannian geodesics and singular curves and presents the first example of singular minimizers. In §4 we present computational tools for finding and understanding singular curves. §5 is devoted to rank 2 distributions where the singular minimizers are ubiquitious. In §6 we describe some generic properties of distributions: the rareness of stability and symmetry. In §6 we describe fat distributions which are distributions admitting no singular curves. In §8 and §9 new results are presented. The first states that for a large class of distributions the distribution is **determined** by its singular curves. The second concerns the impact of singular geodesics on the spectrum of subLaplacians. We give examples in which the singular curves dominate the spectral asymptotics of the subLaplacian. Except for these last two sections this article is of a survey nature, but with open problems interspersed along the way.

ACKNOWLEDGEMENTS

I am happy to have the oppurtunity to thank Robert Bryant and Boris Shapiro for several important conversations which made their way into this paper. I am also pleased to acknowledge valuable conversations with B. Jakubczyk, Ursula Hamenstadt, Lucas Hsu, Wen-Sheng Liu, Hector Sussmann, Richard Murray Shankar Sastry, and Andrei Schnirelman.

2 Preliminary examples, definitions and notation

A subRiemannian structure on a n-dimensional manifold is a smoothly varying distribution of k-planes together with a smoothly varying inner product on these planes. The dimension of the subRiemannian manifold is the pair (k, n). The manifold is denoted by Q, and the distribution by $\mathcal{D}$, $\mathcal{D} \subset TQ$. The inner product will be written $\langle \cdot, \cdot \rangle$. A path will be called horizontal if it is absolutely continuous and its derivatives lie in $\mathcal{D}$ wherever they exist. We define the length of such a path in the usual Riemannian manner:

$$\ell(\gamma) = \int \sqrt{\langle \dot{\gamma}(t), \dot{\gamma}(t) \rangle} dt.$$

The subRiemannian distance $d(x, y)$ between two points x and y is also defined as in Riemannian geometry:

$$d(x, y) = inf(\ell(\gamma))$$

where the infimum is taken over all horizontal paths which connect x and y. The distance is taken to be infinite if there is no such path. In this manner, every subRiemannian manifold is a metric space.

Definition 1 *A path which realizes the distance between its endpoints is called a* minimizing geodesic *or simply a* minimizer.

The example which motivated my interest in the subject is:

example 1: Q is the total space of a principal bundle over a Riemannian base space X. The distribution $\mathcal{D}$ is the horizontal distribution for a connection on the principal bundle and the metric is obtained by using the connection to lift the metric from B. That is to say, for each $q \in X$ the differential $d\pi_q$ of the bundle projection $\pi : Q \to X$, is a linear isomorphism between the horizontal space at q and the tangent space to B at $\pi(q)$. We declare that this differential is an isometry, thus defining the metric on the horizontal space.

Suppose that both endpoints q_0, q_1 of a geodesic lie in the same fiber. Then there is a $g \in G$, the structure group of the bundle, such that $q_0 g = q_1$. The problem of finding a minimizer is then the isoholonomic problem: Among all loops based at $\pi(q_0)$ whose holonomy (with respect to q_0) is g find the shortest one. If the group G is the circle group, if X is a two-dimenisonal Riemannian surface and if the curvature of the connection is a nonzero constant multiple of the area form on S, then the isoholonomic problem is the (dual of the) isoperimetric problem: Among all loops in S of a given area find the shortest one. The solutions to this problem have a very nice description in terms of the Kaluza-Klein type models from high energy physics and will be described in §3.3.

At least three questions of interest in physics can be phrased as isoholonomic problems. What is the most efficient way to flip the phase of a quantum mechanical sample using an imposed magnetic field? (See [63]), What is the most efficient way for a microorganism to swim? ([48]) What is the most efficient way for a falling upside-down cat to move so as to right itself? ([38].) In addition to the references given, [35] and [49] discuss these examples and others in this context.

2.1 PRESENTATIONS OF SUBRIEMANNIAN STRUCTURES

There are various natural presentations of a subRiemannian structure, each having its own advantages.

PRESENTATION 1: Q is a Riemannian manifold with distribution $\mathcal{D}$. The inner product $\langle \cdot, \cdot \rangle$ on $\mathcal{D}$ is the restriction of the Riemannian inner product.

Conversely, we can consider a family of Riemannian metrics of the form $ds^2_\epsilon = \langle \cdot, \cdot \rangle \oplus \frac{1}{\epsilon^2}(\cdot, \cdot)$ with respect to a splitting, $TQ = \mathcal{D} \oplus V$ of the tangent bundle, where V is a distribution complementary to $\mathcal{D}$ and $(\cdot, \cdot)$ is a fiber inner product on it. Let d_ϵ be the corresponding distance function. Then

$$lim_{\epsilon \to 0} d_\epsilon(x, y) = d(x, y),$$

the subRiemannian distance function. Thus we can think of subRiemannian structures as the limits of Riemannian ones.

In the case of example 1, there is a natural choice of such a splitting. Take $V = ker(d\pi)$ to be the vertical bundle. Each fiber V_q is naturally isomorphic to the Lie algebra of G. Fix an adjoint invariant inner product on the Lie algebra. This defines the metric $(\cdot, \cdot)$ on V. WE CALL SUCH METRICS KALUZA-KLEIN METRICS. If G is semi-simple then, **up to scale**, there is only one adjoint invariant inner product on its Lie algebra. So in this case we arrive at a canonical one-parameter family of Riemannian metrics converging to our subRiemannian structure.

PRESENTATION 2. Let θ^a, $a = 1, \ldots, c$ be a collection of linearly independent one-forms. Here $c = n - k$. Their vanishing defines a distribution

$$\mathcal{D} = \{v \in TQ : \theta^a(v) = 0, a = 1, \ldots, c\}$$

of rank k. The inner product on the distribution can be expressed in the form

$$ds^2 = \Sigma(\omega^\mu)^2 \text{ mod } \{\theta^a : a = 1, \ldots c\}$$

where the $\omega^\mu, \mu = 1, \ldots, k$ form a complementary set of one-forms so that together the θ and ω form a coframe field on Q.

If $\hat{\theta}, \hat{\omega}$ are another such collection of forms then they are related to the original collection by

$$\begin{pmatrix} \hat{\omega} \\ \hat{\theta} \end{pmatrix} = \begin{pmatrix} R & * \\ 0 & g \end{pmatrix} \begin{pmatrix} \omega \\ \theta \end{pmatrix},$$

where R is an orthogonal $n-k \times n-k$ matrix , $*$ denotes and arbitrary $k \times n-k$ matrix and g is an invertible $k \times k$ matrix. This illustrates the G-structure associated to a subRiemannian manifold.

In the case of example 1 we take the θ to be the components of the connection form, and the ω to be the pull-backs to Q of a coframing of the base.

PRESENTATION 3. A control system:

$$\dot{q} = \Sigma_{a=1}^{k} u^a(t) X_a(q) \tag{1}$$

linear in the controls defines a subRiemannian structure. Its solution curves $q(t)$ are the horizontal curves. The X_a form a basis for the distribution $\mathcal{D}$. If we declare them to be orthonormal then the minimizing geodesics are those solution curves which minimize the square of the L_2 norm $\int \Sigma(u^a(t))^2 dt$ of the controls.

This exhibits the manifold structure of the space $\Omega_{\mathcal{D}}(q_0)$ of all horizontal paths starting at q_0. Namely, it is coordinatized by the controls $u \in L_2([0,1], \mathbb{R}^k)$ by solving eq (1) with initial conditions $q(0) = q_0$.

In the case of example 1 the X_a are the horizontal lifts of an orthonormal frame on the base.

2.2 THE HEISENBERG GROUP

The simplest nontrivial example of a subRiemannian structure lives of the three-dimensional Heisenberg group, denoted H_3. The Heisenberg algebra is the three-dimensional Lie algebra with basis $\{X, Y, Z\}$ and with the only nonzero bracket between the basis elements being $[X, Y] = Z$. Think of X, Y, Z as left-invariant vector fields on the corresponding simply connected Lie group H_3. H_3 is diffeomorphic to $\mathbb{R}^3$. The exponential map provides the diffeomorphism. We define the distribution $\mathcal{D}$ on H_3 to be the span of X and Y which we declare to be orthonormal.

Z generates the center of the group. By modding out by the center we obtain a principal $\mathbb{R}$-bundle: $\mathbb{R} \to H_3 \to \mathbb{R}^2$. $\mathcal{D}$ is the horizontal distribution for a connection whose curvature is the area form on the plane and so is a particular case of example 1.

This distribution can be easily visualized. In appropriate exponential coordinates $\{x, y, z\}$ on H_3 the distribution is the kernel of the form $dz - \frac{1}{2}(xdy - ydx) = 0$ which in cylindrical coordinates is $dz - \frac{1}{2}r^2 d\theta = 0$. We visualize this distribution as a kind of continuous family of propellor blades. Along the z axis the distribution is parallel to the xy plane. As we move out radially from the axis in any direction the distribution 2-planes tilts in a monotonic way. They continue to contain the radial vector field $\frac{\partial}{\partial r}$ and the angle between their normal vector and the unit z-vector increases monotonically from zero to ninety degrees.

The metric on the distribution is $dx^2 + dy^2|_{\mathcal{D}}$. The projections of the Heisenberg geodesics to the xy plane are circles and straight lines. Given our realization, described above in example 1, that the problem of finding Heisenberg geodesics is a restatement of the isoperimetric problem of Dido, these geodesics have been known for over 2,000 years.

2.3 BRACKET GENERATION AND CHOW'S THEOREM

If the distribution is involutive, then the distance between two points will be infinite if they do not lie on the same leaf of the distribution. For in this case they cannot be connected by a horizontal path. At the opposite extreme of involutivity is the condition of bracket generation which is the situation of interest to us.

By abuse of notation, let $\mathcal{D}$ also denote the space of smooth horizontal vector fields. Form

$$\begin{aligned} \mathcal{D}^2 &= \mathcal{D} + [\mathcal{D}, \mathcal{D}] \\ \mathcal{D}^3 &= \mathcal{D}^2 + [\mathcal{D}, \mathcal{D}^2] \\ &\vdots \end{aligned}$$

where the brackets denote Lie brackets. Now evaluate these spaces of vector fields at a point $q \in Q$ thus obtaining a flag of subspaces:

$$\mathcal{D}_q \subset \mathcal{D}_q^2 \subset \mathcal{D}_q^3 \subset \ldots \subset T_q Q \tag{2}$$

Definition 2 *The distribution is said to be* BRACKET-GENERATING *if for each q there is a positive integer h such that $\mathcal{D}_q^h = T_q Q$. The first such h is called the degree of nonholonomy.*

REMARK Many authors use the phrase "satisfies Hormander's condition" instead of "bracket-generating".

REMARK Bracket generation is a generic property for distributions. On the other hand, the set of involutive distributions has infinite codimension within the space of all distributions.

Theorem 1 (CHOW) *If the distribution is bracket generating and if x and y lie in the same connected component of the underlying manifold then there exists a smooth horizontal path connecting them.*

Chow was led to his theorem by a result of Caratheodory, which is Chow's theorem for the case of a three-dimensional contact distribution. Caratheodory was in turn inspired by work of Carnot in thermodynamics. Consequently, subRiemannian manifolds are also called CARNOT-CARATHEODORY METRICS.

EXAMPLE 1, REVISITED The Ambrose-Singer theorem (cf. Kobayashi-Nomizu, ch. 2) is Chow's theorem applied to example 1. The bracket generating condition is equivalent to the condition that the values $F(q)(X, Y)$ of the curvature F

at a point q, together with all of its covariant derivatives $D_Z F(X,Y), D_Z D_W F(X,Y), \ldots$ span the Lie algebra of the structure group.

REMARK/WARNING In the theory of exterior differential systems there is a flag called the derived flag which is closely related BUT NOT THE SAME AS the flag just discussed. It is defined recursively as opposed to iteratively. Set $E^1 = \mathcal{D}$, $E^2 = \mathcal{D}^2$, $E^3 = [D^2, D^2]$, and generally, $E^{j+1} = [E^j, E^j]$. Note $\mathcal{D}^j \subset E^j$.

Definition 3 *The jth derived flag is the ideal of differential forms generated by the annihilator of E^{j+1}.*

2.4 Hausdorff dimension

Definition 4 *Consider the flag of eq 2 and let $n_i(q) = dim(\mathcal{D}^i_q)$ so that $n_1 = k$. The list $(n_1, n_2, \ldots, n_h)$ is called the* GROWTH VECTOR *at q. The distribution $\mathcal{D}$ of k-planes is called* REGULAR *if the growth vector is independent of the point q.*

The integers $n_i(q)$ are the most basic numerical invariants associated with a distribution.

Let $Gr(T_qQ)$ denote the graded vector space corresponding to this filtration (2) of T_qQ:

$$Gr(T_qQ) = V_1 \oplus V_2 \oplus \ldots V_h$$

where

$$V_j = \mathcal{D}^j(q)/\mathcal{D}^{j-1}(q).$$

We assume that $\mathcal{D}$ is bracket generating so that $n_h = n$ where h is the degree of nonholonomy. Then $Gr(T_qQ)$ is a vector space of the same dimension as Q. Set

$$dimV_i = d_i = n_i - n_{i-1}.$$

Then $k = d_1$ and

$$n = d_1 + d_2 + \ldots + d_h.$$

Theorem 2 (Mitchell [33]; Pansu [44]) *Let Q be a subRiemannian manifold whose underlying distribution is bracket generating and regular. Then the Hausdorff dimension of Q with respect to the metric induced by the subRiemannian structure is*

$$|d| = d_1 + 2d_2 + 3d_3 + \ldots + hd_h.$$

And the $|d|$-dimensional Hausdorff measure $d\mu_{Haus}$ is absolutely continuous with respect to Lebesgue measure $d^n x$ on Q (as defined by any smooth coordinate system or Riemannian metric on Q): there is a positive Lebesgue-L^1 function f such that $d\mu_{Haus} = f d^n x$.

EXAMPLE The three-dimensional Heisenberg group has $(d_1, d_2) = (2, 1)$ and so its Hausdorff dimension is $4 = 2 + 2 \times 1$. The corresponding 4-dimensional Hausdorff measure is a constant multiple of the usual three-dimensional Lebesgue measure $dxdydz$ since both are Haar measures on this group.

OPEN PROBLEM Calculate the Hausdorff measure for the three-dimensional Heisenberg group. In other words, calculate the constant of proportionality relating it to Lebesgue. (For a nice description and introduction to Hausdorff measure see the book by Falconer [12].)

3 GEODESICS

In the bracket generating case Chow's theorem guarantees us that the subRiemannian distance between two points is finite. A MINIMIZING GEODESIC or simply MINIMIZER is a horizontal curve whose arclength realizes the distance between its endpoints.

"Most" minimizing geodesic are characterized as solutions to a differential equation of Hamiltonian type. This equation can be derived either by an application of the method of Lagrange multipliers or by the maximum principle. In order to describe the Hamiltonian let X_i, $i = 1, 2, \ldots, k$ be a local frame for the distribution and define the matrix-valued function $g_{ij}(q) = \langle X_i(q), X_j(q) \rangle$ and let g^{ij} be the inverse matrix. Think of the X_i as fiber-linear functions on the cotangent bundle according to

$$X_i(q, p) = p(X_i(q)) \quad q \in Q, p \in T_q Q^*$$

Set

$$H = \frac{1}{2} \Sigma g^{ij} X_i X_j. \tag{3}$$

H is a fiber-quadratic positive semi-definite form $T^*Q \to I\!R$ whose rank is k. One easily sees that it is independent of the choice of frame.

Theorem 3 *Let $\gamma \subset Q$ be the projection of an integral curve $\zeta \subset T^*Q$ for the Hamiltonian vector field with Hamiltonian H. Then every sufficiently short subarc of γ is a minimizing geodesic.*

Definition 5 *The curves γ of this theorem will be called* NORMAL *or* REGULAR *geodesics.*

We now come to the third major difference between Riemannian and sub-Riemannian geometry. If the distribution is the entire tangent space then sub-Riemannian geometry becomes Riemannian and theorem 2 is well-known. It is a basic fact of Riemannian geometry that all geodesics are characterized by this theorem. In subRiemannian geometry this is false: there are minimizers which are not the projections of integral curves for the Hamiltonian vector field of H. These extraneous minimizers will be described momentarily.

3.1 COMETRIC AND SUBLAPLACIAN

By polarization, the Hamiltonian H defines a symmetric covariant two-tensor $g : T^*Q \times T^*Q \to I\!R$. This contains all the information of the subRiemannian structure. Think of such a tensor as a vector bundle map $g : T^*Q \to TQ$. Then $im(g) = \mathcal{D}$, the distribution, and $ker(g) = \mathcal{D}^\perp$, the bundle of covectors annhilating the distribution. The fiber-inner product on the distribution is regained by setting $\langle v_1, v_2 \rangle = p_1(g(p_2))$ whenever $v_1 = g(p_1), v_2 = g(p_2) \in \mathcal{D}$.

Our formula for the Hamiltonian H can also be thought of as defining a 2nd order differential operator. As an operator **it does depend** on the choice of framing X_i, since $X_i X_j \neq X_j X_i$ as operators. Just as in Riemannian geometry, first order terms will have to be added in order to have a well-defined subLaplacian. To date no such operator has been defined. (For more see our final section.) But whatever the "subLaplacian" of a subRiemannian metric is, its principal symbol will have to be the Hamiltonian $H : T^*Q \to I\!R$.

3.2 THE GEODESICS OF EXAMPLE 1: KALUZA KLEIN THEORY

If the bundle is a circle bundle with connection form A then H is the Hamiltonian governing the motion of a charged particle in the magnetic field dA. The charge λ corresponds to the momentum in the fiber direction which in turn has the interpretation as the Lagrange multiplier corresponding to the constraint that the paths be horizontal. It is a constant of the motion. If the base is a two-dimensional Riemannian surface X^2 then we can write $dA = B(areaform)$, thus defining the scalar magnetic field $B : X \to I\!R$ as the Hodge dual of the curvature two-form. Hamilton's equations are then equivalent to

$$k_g(x(s)) = cB(x(s))$$

where $x(s)$ is the projection to the base space of an extremal curve, s is arc length, and k_g is the geodesic curvature of $x(s)$, and the constant $c = \lambda/|v|$. Here $|v|$ is the particle's speed and is easily seen to be a constant of the motion. ($H = \frac{1}{2}|v|^2$.) **The normal geodesics are the horizontal lifts of the trajectories of charged particles in the magnetic field B.**

This characterization of the normal geodesics can be seen nicely by using the method of Lagrange multipliers. Consider the functional

$$S = \int \{\frac{1}{2}\|\dot{\gamma}\|^2 dt - \lambda \gamma^* A\}$$

where $\lambda = \lambda(t)$ is a Lagrange multiplier introduced to impose the constraint that the curve γ be horizontal. The norm $\|\dot{\gamma}\|$ of the bundle velocity is taken with respect to a Kaluza-Klein metric on the circle bundle. (See presentation 1, §2.1.) The functional is invariant under the circle action and the Noether-conserved quantity corresponding to this action is the multiplier λ. This means that $\frac{d\lambda}{dt} = 0$

along any solution curve. We interpret λ as electric charge. Set λ equal to a fixed constant. Then S defines a Lagrangian dynamics on the two-dimensional base space X. (We may ignore the exact differential term λdz coming from $A = dz - A_1 dx - A_2 dy$.) This is the standard Lagrangian description found in almost any physics books for the motion of charged particles in (electro-) magnetic fields. This method of obtaining the motion in space(-time) by adding an extra dimension is attributed to Kaluza [28] and Klein, hence the name.

If B is constant, then fixing the holonomy is the same as fixing the area enclosed by the loop. So the minimization problem is the isoperimetric problem (actually its dual): find the shortest loop enclosing a fixed area. The extrema are well-known to be the curves of constant geodesic curvature.

HEISENBERG GROUP REVISITED (See §2.2) The base space X^2 is the Euclidean plane with its standard metric. The constant curvature curves on the plane are circles and straight lines and these are the projections of the Heisenberg geodesics. These are well-known to be the trajectories of nonrelativistic charged particles in a constant planar magnetic fields. The bundle is the circle bundle (or $I\!R$ bundle) whose curvature is the area form. If we choose a gauge (section) so that the connection form is $A = dz - \frac{1}{2}(xdy - ydx)$ where x, y are coordinates on the plane and z is the fiber coordinate then as the planar circles are traversed the height z changes according to the area swept out by a line from the origin to the moving point on the circle.

For a general principal bundles with connection, the Hamiltonian of theorem 2 has the form $H = \frac{1}{2}\Sigma g^{ij} X_i X_j$ where the X_i are the horizontal lifts of coordinate vector fields $\frac{\partial}{\partial x^i}$ on the base space X and the matrix g^{ij} is the inverse of the base metric g_{ij} in these coordinates. If we locally trivialize the bundle: $Q \sim X \times G$ (locally) then

$$X_i(b, g; p, \lambda) = p_i - \Sigma \lambda_a A_i^a(b)$$

as a fiber-linear function on T^*Q. Here $(b, g) \in B \times G$, $(p, \lambda) \in T_b^* X \times (Lie(G)^*$, and $A = \Sigma A_i^a(b) E_a dx^i$ is the pullback of the connection one-form with respect to the local section which defines the given trivialization. The E_a form a basis for the Lie algebra $Lie(G)$. $p = p_i dx^i$ and $\lambda = \Sigma \lambda_a \theta^a$ where θ^a is the dual basis to E_a. This Hamiltonian is the Hamiltonian which describes the motion of particle with non-Abelian charge λ traveling over the Riemannian base space X under the influence of the gauge field A. If $x(t)$ denotes the projection of such a minimizer to the base X, and if ∇ is the Levi-Civita connection on X, and F the curvature of the connection A then these equations are equivalent to WONG'S EQUATIONS:

$$\nabla_{\dot{x}} \dot{x} = (\lambda, F(\dot{x}, \cdot))^{\#}$$
$$\frac{D\lambda}{dt} = 0$$

The curvature F can be thought of as a two-form on X with values in the adjoint bundle of Lie algebras over X. The multipliers or "color charges" λ are sections along $x(t)$ of the dual vector bundle, the co-adjoint bundle. Thus $(\lambda, F(\dot{x}, \cdot))$ defines a one-form, or force, along $x(t)$. The superscript $^{\#}$ indicates that we turn it into a vector field along $x(t)$ by raising indices with respect to the metric. The second equation says that the color charge is covariantly constant.

These normal subRiemannian geodesics have an appealing description in terms of Riemannian geometry. Recall from Presentation 1, §2.1, that a Kaluza-Klein metric on Q is a Riemannian metric which agrees with the subRiemannian structure on the horizontal planes, and yields a fixed bi-invariant metric on the group fibers of Q.

Proposition 1 *The normal subRiemannian geodesics are the horizontal lifts of the projections to X of the Kaluza-Klein geodesics. The charge λ measures the angle of the corresponding Riemannian geodesic with the vertical.*

REMARK We called the above subRiemannian geodesic equations Wong's equations, in honor of the physicist [58] who wrote them down as a classical limit of the quantum dynamics of a Yang-Mills particle. They were written down earlier by Kerner [29], and in various forms by Sternberg [51], Weinstein, [57] and probably many others.

3.3 SINGULAR GEODESICS

We finally come to the third and final major difference between Riemannian and subRiemannian geometry. A basic fact of Riemannian geometry that all geodesics can be characterized as solutions to the geodesic equations. In subRiemannian geometry this is false: there are minimizers which are not the projections of integral curves for the Hamiltonian vector field of H. Such extraneous minimizers will be described in detail §2.8.

We begin by describing the possible candidates for these extraneous or SINGULAR minimizers. They depend only on the distribution, and not at all on the inner product on it. Let $\Omega_{\mathcal{D}}$ be the space of all absolutely continuous paths $\gamma : [0,1] \to Q$ which are horizontal ($\dot{\gamma} \in \mathcal{D}$) and square integrable. For $q_0 \in Q$ let $\Omega_{\mathcal{D}}(q_0) \subset \Omega_{\mathcal{D}}$ denote the subset consisting of the paths starting at q_0. (Strictly speaking the derivative need only lie in $\mathcal{D}$ almost everywhere. These path spaces do not depend on the choice of inner product $\langle \cdot, \cdot \rangle$ on $\mathcal{D}$ for if $\dot{\gamma}$ is square integrable with respect to one smooth metric on $\mathcal{D}$ then it is square integrable with respect to any other.) $\Omega_{\mathcal{D}}(q_0)$ is a Hilbert manifold with charts as described in Presentation 3 above. (See for example, Bismut [4], ch. 1 for details concerning this Hilbert space structure. Also see Ge Zhong [16].)

Definition 6 *The* **endpoint map** *is the map $end = end_{q_0} : \Omega_{\mathcal{D}}(q_0) \to Q$ which assigns to each curve its endpoint: $end_{q_0}(\gamma) = \gamma(1)$.*

Thus

$$\Omega_{\mathcal{D}}(q_0, q_1) = end_{q_0}^{-1}(q_1)$$

consists of those horizontal paths beginning at q_0 and ending at q_1. It is the space over which we minimize when we define the subRiemannian distance and minimizing geodesics. *end* is a smooth map and its derivative $d(end)$ can be calculated by using the variation of parameters formula for ODEs. ([4], ch. 1, or p. 57 of [50], or below.)

Definition 7 *A* SINGULAR CURVE $\gamma \in \Omega_{\mathcal{D}}$ *is a singular point of* $end_{\gamma(0)}$*; that is to say, it is a horizontal curve for which the image of the differential* $dend_{\gamma}$ *is not all of* $T_{\gamma(1)}Q$*. A curve* $\gamma \in \Omega_{\mathcal{D}}$ *is* REGULAR *if it is not singular; that is to say, if end is a submersion at* γ*.*

REMARK: The condition that a curve is singular depends only on the distribution, not on the inner product on the distribution. An inspection of the proof in §3.6 yields the fact that the singular curves correspond to those curves for which the linearization of the control system (eq 1, §2.1) is not controllable. Thus the singular curves are intimately related to the critical points of the input-output map.

Anyone who has taught multivariable calculus should know to be careful in applying the method of Lagrange multipliers in the presence of singular points for the constraining function. If F is the constraining function, and E is the function to be minimized then the correct statement of the method is that every constrained minimizer is a critical point of the function $\lambda_0 E + \lambda F$ for some choice of multipliers $(\lambda_0, \lambda) \neq (0, 0)$. The singular extremals are the those for which $\lambda_0 = 0$ and they correspond to singular points of the constraint functional. They may or may not minimize E. In our case E is the length or integrated kinetic energy, and F summarizes the constraints $\dot{\gamma} \in \mathcal{D}$, $\gamma(0) = q_0, \gamma(1) = q_1$.

Theorem 4 *Every minimizing geodesic is either a singular curve or one of the normal geodesics described by theorem 2. The two possibilities need not be mutually exclusive.*

This theorem is a particular instance of Pontrjagin's maximum principle. It can also be proved using standard methods of the calculus of variations, and was well-known to Bliss, Caratheodory, Morse and others. In this earlier literature and in the control literature the singular curves are often called abnormal extremals.

3.4 A SINGULAR MINIMIZER

We now describe the first and simplest example of a strictly singular subRiemannian minimizer. This is a curve which, in addition to being a minimizing geodesic and a singular curve, is not a normal geodesic. That is, it is not the projection any integral curve for the Hamiltonian H. This example is due to the author [34]. Let Q be $\mathbb{R}^3$ with rank 2 distribution defined by the Pfaffian equation

$$dz - y^2 dx = 0.$$

This is a bracket generating distribution which is of contact type away from the plane $y = 0$ and which has growth vector $(2, 2, 3)$ on this plane. The singular curves of this distribution lie on the plane and are lines parallel to the x axis.

Theorem 5 (MONTGOMERY;see also Kupka; Liu-Sussmann) *Let $\langle \cdot, \cdot \rangle$ be any metric on the above distribution and let γ be any of the singular curves. Then every sufficiently short subarc of γ is a minimizing geodesic. For almost all choices of metric, these are not projections of normal extremals.*

To see what is so special about these singular curves, notice that

$$z(t) = z(0) + \int_0^t (y(t))^2 \frac{dx}{dt} dt$$

for any horizontal curve. The singular curves are those for which $y(t)$ is identically zero. Let $\gamma(t) = (t, 0, z_0)$ be such a curve and $\alpha(t) = (t+a(t), b(t), z_0+c(t))$ be a perturbation of γ with $\alpha(0) = \gamma(0)$ and α a horizontal curve. If $|\frac{da}{dt}| < 1$ then $\frac{dx}{dt} > 0$ along α so that $z(t) \geq z_0$ with equality holding if and only if the second coordinate $b(\cdot)$ is identically zero. **It follows that such an α cannot have the same ending height as γ unless it is a reparameterization of γ.** This shows that each singular curve is C^1-rigid, where we use the definition:

Definition 8 *(C^1 rigidity; definition introduced by Bryant-Hsu [8]) A horizontal curve γ for a distribution $\mathcal{D}$ is called C^1-RIGID or C^1-ISOLATED if it is isolated with respect to the C^1-topology on the space $\Omega_{\mathcal{D}}(\gamma(0), \gamma(1))$ (modulo reparameterization) of all horizontal curves sharing its endpoints.*

The usual calculus of variations breaks down for C^1-rigid curves since they admit no smooth variations.

The fact that these curves are C^1-isolated does not in itself prove minimality. For it is a general fact that for any distribution of rank greater than 1, there are no C^0-isolated or even Sobolev H^1-isolated curves in $\Omega_{\mathcal{D}}(q_0, q_1)$.

Lest this example seem too special, we observe that it is stable in the sense of singularity theory: there is a Whitney-open set of distributions diffeomorphic to it. This was proved by Martinet. More precisely he proved:

Theorem 6 ((Martinet)) *. Let $\mathcal{D}$ be a distribution on a three-manifold which is defined in the neighborhood of some point P as the kernel of a nonzero one-form θ. Suppose that in this neighborhood we have*

$$\theta \wedge d\theta = f d^3 x$$

where d^3x is a volume form, and where

$$f(P) = 0$$

$$df(P) \neq 0$$

Also suppose that the distribution is transverse to the degeneration surface

$$\Sigma = \{f = 0\}$$

which is equivalent to the assumption that $\theta \wedge df \neq 0$. Then there exist coordinates (x, y, z) centered at P such that $\theta = dz - y^2 dx$.

3.5 PROOF: MOTIONS IN MAGNETIC FIELDS

My proof of the length minimality of the preceding singular curve is a rather brute force proof. By theorem 4 it suffices to show that every normal geodesic having the same endpoints is longer. So I "sort through " all normal geodesics.

To do this, I assume that the subRiemannian structure has some symmetry. Suppose that translations in z are isometries of the subRiemannian structure. In effect, this is assuming that the structure is of the bundle type of example 1. So, think of Q as a circle bundle over the two-dimensional Riemannian base space X, with X coordinatized by x and y, and the fiber being coordinatized by z. The distribution is then

$$\mathcal{D}_{(x,y,z)} = ker(dz - A_1(x, y)dx - A_2(x, y)dy)$$

If $\sqrt{g}dxdy$ is the surface area element, then the scalar magnetic field is given by

$$B = \frac{1}{\sqrt{g}}(\frac{\partial A_2}{\partial x} - \frac{\partial A_1}{\partial y})$$

which is negative of the Hodge dual of the curvature of the connection form $dz - A_1(x, y)dx - A_2(x, y)dy$. We assume, then, that there is a point P for which $B(P) = 0$ and $dB(P) \neq 0$. This is equivalent to the Martinet conditions. **The singular curves are precisely the horizontal lifts $\hat{C}$ of the zero locus**

$$C = \{(x, y) : B(x, y) = 0\}$$

of the magnetic field. As discussed earlier, the normal geodesics satisfy the equations of motion of a particle with charge λ traveling on the base in the magnetic field B. Recall that these equations are equivalent to

$$k_g(x, y) = \lambda B(x, y)$$

where k_g is the geodesic curvature of the curve $(x(t), y(t)$ with respect to the metric on the base space. Thus the curves $\hat{C}$ are the projections of a normal geodesic iff the zero locus C is a geodesic on the base X. Typically this is not the case.

To prove minimality, I argue by contradiction. Assume that every sufficiently short subarc of a singular curve $\hat{C}$ is **not** minimizing. Let the two endpoints approach each other along $\hat{C}$, thus obtaining a sequence γ_i of normal geodesics with endpoints along a fixed $\hat{C}$ and each shorter than the corresponding arc of $\hat{C}$. **I first show that the corresponding sequences of charges λ_i must tend to ∞ in order for these regular curves to be shorter, or even of the same order of length (in powers of $1/\lambda$) as the corresponding arc of $\hat{C}$.** The above proof of C^1-rigidity shows that in order for such a sequence of curves to have the correct initial and final z values, their projections to the xy plane must have double points or "kinks". The projected curves are divided into arcs, where the endpoints of each arc lies on the zero locus. The crux of our argument involves showing that in order for the boundary conditions to be satisfied, the kinks of each arc must "take up enough length" to force arcs to be longer than the corresponding arcs of C. The analysis is based on analyzing the normal geodesic equations in the limit $\lambda \to \infty$.

Liu and Sussmann [31] have found a short real analysis proof based on an inequality which allows one to easily prove the general case. It also allows them to prove a generalization of this theorem to rank 2 distributions in arbitrary dimensions which we will discuss later.

3.6 SINGULAR CURVES AND CHARACTERISTICS

The singular curves have a nice "microlocal" description. discovered by Lucas Hsu. Let

$$\mathcal{D}^{\perp} \subset T^*Q$$

be the annihilator of the distribution D and ω be the restriction of the canonical two-form $(\Sigma dp_i \wedge dq^i)$ to $\mathcal{D}^{\perp}$.

Definition 9 *A characteristic for $\mathcal{D}^{\perp}$ is an absolutely continuous curve in $\mathcal{D}^{\perp}$ which never intersects the zero section and whose derivative lies in the kernel of ω whenever it exists.*

(Recall that absolutely continuous curves are differentiable almost everywhere and their derivatives are measurable functions.) To be explicit, an absolutely continuous curve $\zeta : [0,1] \to \mathcal{D}^{\perp}$ is a characteristic if

- $\zeta(t)$ is never the zero covector
- $\omega(\zeta(t))(\dot{\zeta}(t), v) = 0$ for each t for which the derivative $\dot{\zeta}(t)$ exists and for each $v \in T_{\zeta(t)}\mathcal{D}^{\perp}$.

Theorem 7 (Hsu**)** *A horizontal curve is singular if and only if there is an absolutely continuous everywhere nonvanishing characteristic curve $\zeta(t) \in \mathcal{D}^{\perp} \subset T^*Q$ which projects onto it.*

REMARK. Hsu proved this under the assumption that the horizontal curve was smooth. I extended his theorem [36] to the case where the derivative of the curve is square integrable, (or more generally in some L^p, $p \geq 1$). This generalization is essential in order to understand analytic properties of minimizing geodesics. In some rough sense this theorem is well-known to the nonlinear control community (cf. Sontag's text) and is contained in Bismut's book. The characteristics are precisely the abnormal extremals of Pontrjagin's maximum principle.

PROOF A 1-parameter family of horizontal curves

$$\gamma_s(t), -\epsilon \leq s \leq \epsilon, 0 \leq t \leq L$$

satisfies the system

$$\frac{d\gamma_s}{dt}(t) = \Sigma u_a(s,t) X_a(\gamma_s(t))$$

following "Presentation 1" above. Let

$$\delta\gamma(t) = \frac{\partial}{\partial s}\gamma_s(t)|_{s=0}.$$

Using $\frac{\partial}{\partial s}\frac{\partial}{\partial t} = \frac{\partial}{\partial t}\frac{\partial}{\partial s}$ and introducing coordinates q^μ so that $X_a = X_a^\mu \frac{\partial}{\partial q^\mu}$ we derive the FIRST VARIATION OF ENDPOINT FORMULA

$$\frac{d}{dt}\delta\gamma^m = \Sigma u_a(t)\frac{\partial X_a^m}{\partial q^\beta}\delta\gamma^\beta + \Sigma X_a^m \delta u_a,$$

an inhomogeneous linear differential equation. Here

$$\delta u_a(t) = \frac{\partial}{\partial s} u_a(s,t)|_{s=0}.$$

This equation can be solved in integral form by using the variation of parameters formula from ODEs and introduing the 1-parameters family $\phi_t : Q \to Q$ of diffeomorphisms defined by integrating the time dependent vector field $\Sigma u_a(t) X_a(q)$. Note $\phi_t(q_0) = \gamma(t)$. We obtain:

$$dend(\gamma)\cdot\delta\gamma = d\phi_1(q_0)\int_0^1 d\phi_t(q_0)^{-1}(\Sigma\delta u_a(t) X_a(\gamma(t)))dt.$$

Now suppose that the image of the differential of the endpoint map is not all of $T_q Q$. Then there is a covector $p_1 \epsilon T_q Q$ which annihilates all the $dend(\gamma)\delta\gamma$'s. Write $p_1 = d\phi_1(q_0)^T p_0$ where ^{T} denotes the transpose (dual) of the differential $d\phi_1(q_0) : T_q Q \to T_q Q$ and

$$p(t) = d\phi_t(q_0)^T p_0, 0 \leq t \leq L. \tag{4}$$

One easily calculates that:

$$< p_1, dend(\gamma)\delta\gamma >= \int_0^1 < p(s), \Sigma X_a(\gamma(s))\delta u^a(s) > ds.$$

It follows from the fundamental lemma of the calculus of variations that that

$$p(s) \in \mathcal{D}^{\perp}{}_{\gamma(s)} \tag{5}$$

One checks directly that the two conditions, eq 4 and eq 5 are equivalent to the statement that the curve $(\gamma(s), p(s))$ is a characteristic for $\mathcal{D}^{\perp}$. For more details see [36].

THE SINGULAR GEODESICS OF EXAMPLE 1

Let F be the curvature two-form. The singular curves are precisely the horizontal lifts to Q of the curves $x(t)$ on X for which there exists a nonzero section λ of the coadjoint bundle along x for which

$$\lambda \cdot F(\dot{x}, \cdot) = 0$$

$$\frac{D\lambda}{dt} = 0$$

We can think of these as the limit of the Wong equations as the multiplier λ tends to infinity.

For the case of circle bundles λ is a constant (the co-adjoint bundle is canonically trivial.) And the equations are simply

$$B = 0$$

That is, the singular curves lie in the zero locus of the magnetic field. Assuming that 0 is a regular value for B this is a collection of smooth curves. In a neighborhood of each one we can choose a gauge so that the connection form A is the normal form of Martinet:

$$A = dz - (A_1 + y^2)dx$$

Then the zero locus of the magnetic field is given by $y = 0$ and x parameterizes the points of the singular curves. A_1 is a constant. This is the case of theorem 5.

4 FINDING CHARACTERISTICS

The dimension of $\mathcal{D}^{\perp}$ is $2n - k = 2c + k$ where $c = n - k$ is the corank of $\mathcal{D}$. So this dimension has the same parity as the rank k of $\mathcal{D}$. Bilinear skew-symmetric forms on odd-dimensional spaces always have kernels and on even spaces they generically have no kernel (they are symplectic). Hence the case of even k is markedly different from odd k.

If k is odd then there is a characteristic vector passing through every point of $\mathcal{D}^{\perp}$. If we knew that these vectors were tangent to actual characteristic curves we would know that a characteristic passed through every point of $\mathcal{D}^{\perp}$, and hence a singular curve through every point of Q. But we do not know this.

However, by using Darboux's theorem (for closed two-forms of constant rank), and the fact that the locus of points on which a two-forms attains its maximal rank is open we can easily show:

Corollary 1 *Suppose the rank of the distribution is odd. Then smooth characteristics pass through an open dense set of* $\mathcal{D}^{\perp}$ *and consequently smooth singular curves pass through an open dense subset of* Q*.*

OPEN PROBLEM: Is this open dense set all of Q?

To proceed further, we will need an efficient tool for finding the characteristic curves. This is provided by exterior differential systems [7].

We follow Hsu's notation [[27]]. Pick a local framing θ^a, $a = 1, 2, \ldots, c$ for $\mathcal{D}^{\perp}$. Thus an arbitrary element of $\mathcal{D}^{\perp}$ can be written uniquely as

$$\theta = \Sigma \lambda_a \theta^a. \tag{6}$$

where the $\lambda_a, a = 1, \ldots, c$ are fiber coordinates for $\mathcal{D}^{\perp}$. We can also think of the θ^a as one-forms on $\mathcal{D}^{\perp}$ by pulling them back from Q by the projection $\pi : \mathcal{D}^{\perp} \to Q$. Then eq (6) can also be viewed as the formula for the restriction of T^*Q's canonical one-form to $\mathcal{D}^{\perp}$. Thus

$$\omega = d\theta = \Sigma d\lambda_a \wedge \theta^a + \lambda_a d\theta^a \tag{7}$$

is the restriction of the canonical two-form to $\mathcal{D}^{\perp}$.

In particular

$$i_{\frac{\partial}{\partial \lambda_a}} \omega = \theta^a$$

where $\frac{\partial}{\partial \lambda_a}$ are the vertical vector fields dual to the $d\lambda_a$. This says that the intersection of the space of vertical vectors, $ker d\pi$, with $ker(\omega)$ is zero. We can read this equation in another way. The space of characteristic vectors ($ker\omega$) are defined by the Pfaffian system $i_Y \omega = 0$, as Y varies over all vector fields tangent to $\mathcal{D}^{\perp}$, or equivalently, as it varies over any (local) framing $Y_1, \ldots, Y_{2n-k}$ of $\mathcal{D}^{\perp}$. Taking $Y = \frac{\partial}{\partial \lambda_a}$ we find that

$$\theta^a = 0$$

along any characteristic which simply says that the projections of characteristics must be horizontal curves. It follows that the differential of the projection $\pi : \mathcal{D}^{\perp} \to Q$ maps $ker(\omega)$ linearly isomorphically onto a subspace of D_q.

Now complete the frame θ^a to form a (local) coframe θ^a, w^μ , $\mu = 1, \ldots, k$ of Q. Let e_a, e_μ be the corresponding dual frame of TQ. Then the e_μ form a basis for D and, $(\frac{\partial}{\partial \lambda_a}, e_a, e_\mu)$ forms a local basis of vector fields on $\mathcal{D}^{\perp}$. Relative to this frame, ω has the block form

$$\begin{pmatrix} 0 & -I & 0 \\ I & 0 & -*^T \\ 0 & * & w(\lambda) \end{pmatrix}$$

where

$$w(\lambda) \quad = \quad \Sigma \lambda_a d\theta^a|_D$$

(The off-diagonal block marked "$*$" and its negative transpose "$-*^T$" are matrices whose precise form do not matter now.) **The following proposition follows directly from this matrix expression and linear algebra**

Proposition 2 *Let $(q,\lambda) \in \mathcal{D}^\perp$ and let $d\pi = d\pi_{(q,\lambda)}$ denote the differential at (q,λ) of the canonical projection $\pi : \mathcal{D}^\perp \to Q$. Then $d\pi$ maps the kernel of ω at (q,λ) isomorphically onto the kernel of the two-form $w(\lambda)$ on D_q.*

The two-form $w(\lambda)$ is obviously crucial. It has the following intrinsic descriptions. Define a linear map

$$w : \mathcal{D}^\perp{}_q \to \Lambda^2(D_q^*)$$

as follows. Extend the element $\lambda \in \mathcal{D}^\perp{}_q$ to form a local section $\tilde{\lambda}$ of $\mathcal{D}$ defined in a neighborhood of q. Then

$$w(\lambda) = d\tilde{\lambda}(q)|\mathcal{D}_q. \tag{8}$$

This is the map whose kernel defines the first derived system in the theory of exterior differential systems. Dually, we have

$$w = \delta^*$$

where the map

$$\delta : \Lambda^2 D_q \to T_qQ/D_q$$

is given by

$$X \wedge Y \mapsto [\tilde{X}, \tilde{Y}](q) mod D_q$$

where $\tilde{X}$, $\tilde{Y}$ are any extensions of the vectors $X, Y \in D_q$ to sections of D. (δ is the first term of the nilpotentization – the canonical graded Lie bracket structure on $Gr(T_qQ)$. See §2.6.)

The case of even rank

Suppose that the rank $k = 2l$ is even. Recall that a two-form ω is called symplectic at a point if it has no kernel there.

Definition 10 *The characteristic variety, $\Sigma \subset \mathcal{D}^\perp \setminus 0$ is the set of nonzero covectors in $\mathcal{D}^\perp$ at which the two-form ω is not symplectic.*

Any characteristic must lie completely in the characteristic variety.

Σ is defined by the equation $\omega^{n-l} = 0$ where $2l$ is the rank of D. Now ω^{n-l} is a form of top dimension and so has the form $fd^N x$, where f is a function, $N = 2(n-l)$ is the dimension of $\mathcal{D}^\perp$ and $d^N x$ is a local volume form on $\mathcal{D}^\perp$. Thus Σ is defined by the single scalar equation $f = 0$ and we expect it to be either empty or a hypersurface. By the above analysis, we can also express Σ as the solution variety

$$\Sigma = \{(q, \lambda) \in \mathcal{D}^{\perp} : w(\lambda)^l = 0\}$$

For fixed q the equation $w(\lambda) = 0$ is a single homogeneous polynomial equation of degree l for the variable $\lambda \in \mathcal{D}^{\perp}{}_q \cong \mathbf{R}^{n-k}$. This shows that

$$\Sigma_q = \Sigma \cap T_q^* Q$$

is a real algebraic variety.

Typically Σ is not a smooth submanifold. For generic $\mathcal{D}$ it will be a smooth stratified set in the sense of Whitney. The strata are determined by the rank of ω and the relative positions of $ker(\omega)$ with the rank strata.

5 RANK TWO EXAMPLES

If $\mathcal{D}$ is rank 2, then $\mathcal{D}^{\perp}$ has dimension $2n - 2$. The charactersitic subvariety $\Sigma \subset \mathcal{D}^{\perp}$ is defined by the fiber-linear condition $\Sigma \lambda_a d\theta^a = 0(\ mod \mathcal{D}^{\perp})$. Using eq 8 we see that

$$\Sigma = (\mathcal{D}^2)^{\perp}.$$

The kernel of the form ω on $\mathcal{D}^{\perp}$ is two-dimensional along Σ. This two-plane is transverse to Σ at λ if and only if $\lambda \notin (\mathcal{D}^3)^{\perp}$. In this case the intersection of the kernel with the tangent space to Σ defines a line field on Σ in a neighborhood of such a point. The integral curves of this line field will be characteristics. (One can check that these are also the characteristics for the restriction of ω to Σ.)

Definition 11 *The projections of the just-described characteristics are called the* GENERIC SINGULAR CURVES. *(Liu and Sussmann call them* REGULAR ABNORMAL CURVES.*)*

Theorem 8 (Bryant-Hsu (for rigidity) Liu-Sussmann, (for minimality)) *Every sufficiently short subarc of a generic singular curve is a C^1-rigid curve and a minimizing geodesic.*

GROWTH VECTOR (2,3,4)

There is only one such distribution up to local diffeomorphism. It is called the Engel distribution

Definition 12 *An Engel distribution is a rank 2 distribution on a 4-manifold which is regular and bracket generating.*

The growth vector of an Engel distribution is necessarily $(2, 3, 4)$. So any point of an Engel manifold is contained in a neighborhood for which the distribution admits a local framing by vector fields X, W such that $\{X, W, [X, W], [X, [X, W]]\}$ span the entire tangent space TQ. Write $Y = [X, W]$, $Z = [X, [X, W]]$.

Theorem 9 ((Engel)) *The framing $\{X, W\}$ can be chosen so that the defining relations for Y and Z are the only bracket relations among these four vector fields. In fact, centered about any point of an Engel manifold we can find local coordinates x, y, z, w so that $W = \frac{\partial}{\partial w}$ and $X = \frac{\partial}{\partial x} + w\frac{\partial}{\partial y} + y\frac{\partial}{\partial z}$.*

This assertion is the Darboux theorem of Engel manifolds. See, for example, the text [7] for a proof.

EXAMPLE: Engel manifolds can be constructed out of any three-manifold Y as follows. Every three-manifold Y admits a contact structure $E \subset TY$. The four-manifold Q consists of all contact directions. In other words Q^4 is the total space of the projectivization of the 2-plane bundle E which is a circle bundle over Y. The Engel distribution on Q is defined by declaring that a curve is tangent to this plane field if its projection to the three-manifold is tangent to the corresponding contact direction there.

FOLIATION BY SINGULAR CURVES

An Engel distribution $\mathcal{D}$ admits an unique line field $L \subset \mathcal{D}$ with the property that $[L, [\mathcal{D}, \mathcal{D}]] \equiv 0 \ (mod[\mathcal{D}, \mathcal{D}])$. L is the span of the vector field W of the standard frame $\{X, W\}$ above. The integral curves of L are exactly the singular curves of $\mathcal{D}$. According to theorem 8 these are C^1-rigid locally minimizing geodesics. In the above example these curves are the circle fibers.

Bryant and Hsu [8] have shown that each such curve admits an intrinsic real projective structure. Here is a geometric re-interpretation of the structure which they discovered. Let Y be the quotient of M by the foliation of singular curves. (The analysis is local in a neighborhood of an arc of a singular curve, so the quotient need only be local.) Since $[L, \mathcal{D}^2] \subset \mathcal{D}^2$, and $L \subset \mathcal{D}^2$ the rank 3 distribution $\mathcal{D}^2$ descends to a rank 2 distribution, say $\pi_*\mathcal{D}^2$, on Y. $\mathcal{D}$ itself does not descend. Let $y \in Y$ and write $C = \pi^{-1}(y)$ for the corresponding singular curve. For each point $q \in C$, the subspace $d\pi_q(\mathcal{D}_q)$ forms a line in the two dimensional vector space $\pi_*\mathcal{D}^2_y$. In this way we get an intrinsically defined development map

$$\delta : C \to \mathbb{R}P^1$$

where the $\mathbb{R}P^1$ is the set of lines in $\pi_*\mathcal{D}^2_y$. Now observe that such a development map δ for a curve is equivalent to a real projective structure on it. Using the local normal form (Theorem 9) one checks that δ is monotonic: its derivative never vanishes. Bryant and Hsu proved the following remarkable theorem relating the real projective structure to C^1-rigidity.

Theorem 10 ((Bryant-Hsu)) *Let $A \subset C$ be an arc of a singular curve C. Then A is C^1 rigid if and only if the restriction of the development map to A is one-to-one.*

In other words, as soon as the curve C begins to wrap more than one time around the projective line, it fails to be C^1 rigid.

EXISTENCE OF ENGEL STRUCTURES The existence of L implies that we have a well-defined full flag

$$L \subset \mathcal{D} \subset \mathcal{D}^2 \subset TQ$$

of subbundles of the tangent space. Modulo problems of orienting these subbundles, this implies that every Engel manifold must be parallelizable. (Put a metric on Q and use Graham-Schmidt to make a frame field $\{e_1, e_2, e_3, e_4\}$ with $L = Span(e_1)$, $\mathcal{D} = Span\{e_1, e_2\}$, etc..)

Conversely, Gershkovich [17] has claimed to have shown that every parallelizable 4-manifold admits an Engel structure. This should be compared to the result that every 3-manifold admits a contact structure.

DIMENSION (2,3,5)

In this case there is a unique generic singular curve tangent to every horizontal vector $v_q \in \mathcal{D}_q$. We have a kind of "singular exponential map" and can get from any one point to any other via concatenations of singular geodesics.

This case was studied in detail by E. Cartan in his "five variables paper" [9]. Among other things, he showed how to construct the complete diffeomorphism invariants of such a structure. It is a 4th order symmetric covariant tensor on the distribution; that is, a section of the bundle $S^4(\mathcal{D}^*)$.

GOURSAT CASE: (2,3,4,5, ...) The space $Q = J^k(\mathbb{R})$ of k-jets of a real function $y = y(x)$ of a real variable x inherits a natural rank 2 distribution whose integral curves are the k-jets of a function, together with the "vertical curves" described below. Let y_j stand for the jth derivative so that $x, y, y_1, \dots, y_j$ form coordinates on Q. A curve $(x, y(x), y_1(x), \dots, y_k(x))$ is the k-jet of a function iff

$$dy - y_1 dx = 0$$

$$dy_1 - y_2 dx = 0$$

$$\dots$$

$$dy_{k-1} - y_k dx = 0.$$

These j one-forms define a rank 2 distribution in the $j+2$ dimensional jet space.

Theorem 11 (Goursat normal form; see: [7], [42] Thm 3) *Suppose that a distribution has growth vector $(2, 3, 4, 5, \dots, k+2)$. Also suppose that the rank vector of its* RECURSIVELY *defined flag : $E^{j+1} = [E^j, E^j]$ with $E^1 = \mathcal{D}$ is also $(2, 3, 4, 5, \dots, k+2)$; or, what is the same thing, that $E^j = \mathcal{D}^j$. Then the distribution is locally diffeomorphic to the the above distribution on the jet space.*

For $k = 1$ this distribution is the standard 3 dimensional contact distribution and for $k = 2$ it is the Engel distribution. For $k > 1$ the Goursat distribution has the property that exactly one singular curve passes thru each point. These are the integral curve of the vector field $\frac{\partial}{\partial y_k}$. They are the vertical curves referred to above and correspond to varying only the kth derivative.

THE GOURSAT CASE IN INFINITE DIMENSIONS The Goursat distribution is spanned by the vector fields

$$X = \frac{\partial}{\partial x} + y_1 \frac{\partial}{\partial y} + y_2 \frac{\partial}{\partial y_1} + \dots y_k \frac{\partial}{\partial y_{k-1}}$$

and the "vertical vector field"

$$\frac{\partial}{\partial y_k}$$

This last vector field disappears upon letting $k \to \infty$. Formally, in the infinite variable limit the rank of the distribution drops from two to one: it becomes a line field!. The integral curve thru the point with coordinates $(0, y_0, y_1, y_2, \dots)$ is represented by the function whose Taylor coefficients at 0 are the y_j. Note the relevance of whether or not a curve is analytic here.

This phenomenon was pointed out to us by A. Shnirelman. He argues that it can be related to the problem of free will in philosophy.

The Goursat system describes the kinematics of a truck-trailer system consisting of the front truck containing the 2 control actuators: "steer" and "drive", together with $k-2$ (passive) trailers attached. When we let $k \to \infty$ it appears that we can read off the entire past history of the system from its current state, (assuming the control inputs to be analytic) and that this state in turn determines the future motion. A truck driver pulling an infinite string of trailers can no longer steer his vehicle!

6 NECESSARY AND SUFFICIENT CONDITIONS FOR RIGIDITY

Agrachev and Sarychev [1] have found simple necessary conditions and sufficient conditions for a horizontal curve γ to be C^1- rigid. In order to state their conditions observe that if $\lambda(t) \in (\mathcal{D}^2_{\lambda(t)})^\perp$ then the quadratic form

$$Q_{\lambda(t)}(v) := \lambda(t)([v, [\dot{\gamma(t)}, v]]) \; ; \; v \in \mathcal{D}_{\lambda(t)}$$

is well-defined. This is because the Lie-bracket induces a well defined map $\mathcal{D}^2_q \times \mathcal{D}_q \to \mathcal{D}^3_q / \mathcal{D}^2_q$. (See § 2.5 on nilpotentization.)

Theorem 12 (Agrachev-Sarychev [1]: theorems 4.3 and 5.5) *Let γ be a singular C^1-path. If γ is C^1 rigid then there exists a nonzero characteristic $\lambda(t)$ along $\gamma(t)$ such that for each t*

$$\lambda(t) \in (\mathcal{D}^2)^\perp \tag{9}$$

$$Q_{\lambda(t)} \geq 0 \tag{10}$$

Conversely if there exists a nonzero characteristic λ along γ with $\lambda(t) \in (D^2)^\perp$ and if for each t the quadratic form $Q_{\lambda(t)}$ is positive-definite **transverse to** $\dot\gamma$ *(which is to say it is positive-definite on some, and hence any, $k-1$ dimensional subspace transverse to $\dot\gamma(t)$) then every sufficiently short subarc of γ is C^1-rigid.*

REMARK: These conditions hold trivially for the characteristics of the rank 2 case governed by the theorems of Liu-Sussmann and Bryant-Hsu.

REMARK: There theorem is actually slightly stronger, as it is stated purely for Lipshitz curves

EXAMPLE: For a typical distribution their will be no rigid curves if the corank $n-k \leq \binom{k}{2}$. For in this case D^2 typically spans.

SKETCH OF PROOF

If a smooth function $F : \mathcal{H} \to I\!R^n$, where $\mathcal{H}$ is Hilbert space, has an isolated zero at $0 \in \mathcal{H}$, then 0 must be a critical point for F. Let $\lambda \in (I\!R^n)^*$ be a corresponding non-zero Lagrange multiplier: $\lambda \cdot dF(0) = 0$. We may think of λ as reading off the first component F_1 of F relative to some basis for $I\!R^N$. Thus $\lambda d^2F(0) = d^2F_1(0)$. If this is a positive-definite quadratic form then $F_1(h) \geq c||h||^2$ for some positive constant c, so that 0 is an isolated zero for F_1 and hence for F. A similar argument shows that if λd^2F is of mixed sign for all $\lambda \in (im(d^2F(0)))^\perp$ then 0 is not an isolated zero.

We apply these considerations with F being the endpoint map and $\mathcal{H}$ parameterizing a Lipshitz neighborhood of the given curve γ. The parameterization is by bounded measurable controls $u_a(t)$ and is defined by a frame field X_a for the distribution near γ as in presentation 3, §2.1. The multiplier $\lambda = \lambda_1$ is a covector attached at $\gamma(1)$ which can be propagated backwards along γ as in as in § 3.7, eq (4), using a horizontal flow ϕ_t with $\phi_t(\gamma(0)) = \gamma(t)$. The resulting covector $\lambda \in \mathcal{D}^\perp$ is a characteristic along γ. The second variation formula can be written:

$$\lambda_1 d^2 end(\gamma)(v) = \int\int_{0\leq s\leq t\leq 1} \lambda(s)([v(s), \phi^*_{t-s}v(t)]|_{\gamma(s)})dsdt.$$

In order to obtain the two conditions of the theorem expand $v(s) = \Sigma f_a(s)X_a(\gamma(s))$ in terms of the frame field to obtain the second variation in the form

$$\int\int_{0\leq s\leq t\leq 1} K_{ab}(t,s)f_a(t)f_b(s)dsdt$$

where the kernel K is $K_{ab}(t,s) = \lambda(s)([X_a, \phi^*_{t-s}X_b]|_{\gamma(s)})$, $t \geq s$.

Lemma 1 *Consider a quadratic form of the above type where the matrix kernel K is continuous in (t,s) and is skew-symmetric on the diagonal: $K_{ab}(t,t) = -K_{ab}(t,t)$. If this quadratic form has finite index then K is identically zero on the diagonal. If this condition holds and if the quadratic form is also positive then the "interior derivative" of the kernel on the diagonal must be positive :* $\frac{\partial K_{ab}}{\partial t}|_{t,t} \geq 0$.

PROOF OF LEMMA: Suppose $K_{12}(t,t) = -K_{21}(t,t) \neq 0$. Take f of the form $\delta(t-t_1)e_1 + \delta(t-t_2)e_2$ where $e_1, e_2, \ldots$ is the standard basis for $\mathbb{R}^k$ and with $t_1 < t_2$. One calculates formally that the corresponding second variation is $K_{12}(t_2, t_1)$. Switching e_1 and e_2 we get $K_{21}(t_2, t_1)$. For t_2, t_1 close to t one of these variations will be postive and the other negative. Letting t vary, we see that the span of the set of such f, or more precisely, of approximate delta function families for such f, is infinite-dimensional. Hence the index is infinite.

If the symmetry condition holds on the diagonal then we can expand the kernel K in a Taylor's expansion near the diagonal to get the second condition. QED

A simple calculation shows that the two conditions of this lemma correspond to the two conditions of the theorem of Agrachev and Sarychev.

7 STABILITY AND ASYMMETRY OF DISTRIBUTIONS

CLASSIFICATION OF THE STABLE REGULAR DISTRIBUTIONS

Definition 13 *A distribution is called* STABLE *if any nearby distribution is locally diffeomorphic to it.*

Here "nearby" is in the sense of the Whitney C^k-topology, k at least 1, on the space of all distributions.

Theorem 13 ((Gershkovich and Vershik) [55]) *The only stable distributions occur in dimensions $(1,n)$, $(n-1,n)$, and $(2,4)$. In each of these dimensions there is exactly one stable regular distribution and these are the line fields, contact or odd-contact distributions, and the Engel distribution described in the previous section.*

The idea behind this theorem is a simple and fundamental one which goes back at least to Riemann in his famous first lecture on differential geometry: the idea of counting functional dimension. Locally a distribution of k-planes on an n-manifold is a map from $\mathbb{R}^n$ to the Grassmannian $G_k(\mathbb{R}^n)$ of k-planes in n-space. The Grassmannian has dimension $k(n-k)$ and so a distribution is specified by this many functions of n variables. A diffeomorphism is specified by n functions of n variables. So the quotient space, which is the space of distributions up to diffeomorphism equivalence, is specified by $k(n-k)-n$ functions of the n variables. In order for there to be any stable distributions this number has to be zero or negative. For if the number were positive, then the space of nearby inequivalent distributions is parameterized by an infinite-dimensional function space. $k(n-k)-n \leq 0$ exactly for the (k,n) in the range of the theorem.

The proof can be made rigorous by working on finite jet bundles of distributions. The space of polynomials of degree r in n variables has dimension $\binom{n+r}{n}$.

It follows that the space $J^r(G_k(\mathbb{R}^n))_0$ of r-jets of rank k-distributions at 0 has dimension

$$g(r;k,n) = k(n-k)\binom{n+r}{n}$$

and that the space $Diff^{r+1}(\mathbb{R}^n)_0$ of r+1 -jets at 0 of diffeomorphisms of $\mathbb{R}^n$ taking 0 to 0 has dimension

$$d(r,n) = n[\binom{n+r+1}{n} - 1].$$

The group $Diff_0^{r+1}$ acts on $J^r(G_k(\mathbb{R}^n))_0$. If a distribution $\mathcal{D}$ is stable then the orbit of its r-jet $j^r(\mathcal{D})_0$ at 0 must be open. But the dimension of at orbit cannot be larger than that of the group so this is possible only if

$$d(r;n) \geq g(r;k,n).$$

A simple calculation shows that this is possible for all r only if $k(n-k) \leq n$. See [54] [39].

Asymmetry of distributions

It is well-known that contact distributions admit infinite-dimensional symmetry groups. So do Engel distributions as is easily seen from their relation to contact distributions. As soon as we are away from the stable range, the situation changes drastically. An argument along the lines of the previous dimension count strongly suggests the validity of the following.

[folktheorem] The typical bracket generating distribution whose dimension (k,n) is outside the stable range $k(n-k) \leq n$ admits no local symmetries.

heuristic proof:[following Bryant]

A distribution on M is a section of the bundle $G_k(TM) \to M$ whose typical fiber $G_k(T_xM)$ is the Grassmannian of k-planes in the tangent space T_xM. The r-jet of a distribution then defines a section $j^r(\mathcal{D}) : M \to J^r(G_k(TM))$ or the corresponding jet bundle. Now the Lie group $Diff_x^{r+1}$ of $r+1$ jets of diffeomorphisms fixing x acts on the fiber thus defining a singular foliation of the jet bundle. **We expect that the Thom transversality theorem holds in this context.** In other words, we expect that a Whitney open and dense set of distributions have the property that their r-jets are transverse to the leaves of this foliation.

Coordinates in a neighborhood of a point on M induce a local fiber preserving diffeomorphism between $J^r(G_k(TM))$ and $J^r(G_k(\mathbb{R}^n))_0 \times \mathbb{R}^n$ which takes the leaves of this foliation to the orbits of the $Diff_0^{r+1}$-action. (The group acts trivially on the second factor.) The r-jet of a distribution then becomes a map $j^r(\mathcal{D}) : \mathbb{R}^n \to J^r(G_k(\mathbb{R}^n))_0$. The distribution cannot have local symmetries if there is a neighborhood of 0 such that in this neighborhood $j^r(\mathcal{D})(x) \neq \mathcal{O}$ for $x \neq 0$ in this neighborhood, where $\mathcal{O}$ denotes the orbit of $j^r(\mathcal{D})(0)$. This will be true if the orbit has codimension n or greater and if $j^r(\mathcal{D})$ is transverse to the orbit at 0. But the codimension of an orbit is at least $g(r;k,n) - d(r;n)$. An easy calculation shows that for r large enough we have $g(r;k,n) - d(r;n) \geq n$ provided $k(n-k) > n$. (This inequality holds as soon as $r \geq \frac{n(n+1)-k(n-k)}{k(n-k)-n}$.)

No complete rigorous proof of this folk theorem has been written down.

8 Fat distributions

Fatness is the only simple condition which rules out the existence of singular curves. Weinstein [56] coined the use of the adjective "fat" in the bundle context of example 1 where it is the strongest opposite to a connection being flat. Many authors assume fatness to obtain their results .

Definition 14 *A distribution is called* FAT *or* STRONG-BRACKET GENERATING *if the restriction of the canonical two-form to its annihilator $\mathcal{D}^{\perp} \subset T^*Q$ is a symplectic form.*

Proposition 3 *The following are equivalent*

- *(i) $\mathcal{D}$ is fat*
- *(ii) For each $q \in Q$ and nonzero $\lambda \in {\mathcal{D}^{\perp}}_q$ the two-form $w(\lambda)$ is symplectic on $\mathcal{D}_q$*
- *(iii) For each $q \in Q$ and each nonzero vector $v \in \mathcal{D}_q$ we have*

$$\mathcal{D}_q + [V, \mathcal{D}]_q = T_q Q$$

where V is any horizontal extension of v.

Item (iii) is the origin of the phrase STRONG BRACKET GENERATING as a synonym for fat. The equivalence is proved using Cartan's magic formula $d\tilde{\lambda}(V,X) = V[\tilde{\lambda}(X)] - X[\tilde{\lambda}(V)] + \tilde{\lambda}([V,X])$.

Obviously we have

Proposition 4 *If the distribution is fat then it admits no singular curves.*

THE CONVERSE IS FALSE. The simplest counterexample is of the type of example 1. Consider a planar magnetic field $B(x,y) = x^2 + y^2$, with corresponding distribution

$$\mathcal{D} = ker(dz - A)$$

in xyz space, where $A = A_1(x,y)dx + A_2(x,y)dy$ satisfies $dA = Bdx \wedge dy$. $\mathcal{D}^{\perp}$ fails to be symplectic exactly over the z-axis $\{x = y = 0\}$. There can be no characteristics because any such curve would have to have its projection lie in the z-axis and also be horizontal, which is impossible.

NECESSARY CONDITIONS FOR FATNESS The condition of fatness persists under C^1-perturbations of the distribution. However the existence of fat distributions is an extremely restrictive condition on the dimension (k,n) of a nonholonomic maniold.

Proposition 5 (Rayner [46]) *Suppose that $\mathcal{D}$ is fat at q and has codimension 2 or greater. Then*

- *i) $n_2(q) = n$; i.e. $\mathcal{D}$ is two-bracket generating*
- *ii) k is a multiple of 4*
- *iii) $k \geq (n-k)+1$*
- *iv) the sphere $S^{k-1} \subset \mathcal{D}_q$ admits $n-k$ linearly independent vector fields*

Conversely, for any dimension (k, n) satisfying these numerical restrictions (i)-(iv) there exists a fat distribution with these dimensions on $I\!R^n$.

PROOF

i) The strong bracket generating condition (iii) above implies that $T_qQ = \mathcal{D}_q + [\mathcal{D}, \mathcal{D}]_q$.

ii) The condition that $w(\lambda)$ be symplectic can be written $w(\lambda)^l \neq 0$ where $2l = k$. Relative to a local frame θ^a for $\mathcal{D}^\perp$, $w(\lambda) = \Sigma_a \lambda_a d\theta^a | \mathcal{D}_q$. So the equation $w(\lambda)^l = 0$ is a single homogeneous polynomial equation of degree l in the λ_a. If l is odd such a polynomial equation has nontrivial solutions.

iii) Define skew-symmetric operators $J^a : \mathcal{D}_q \to \mathcal{D}_q$, $a = 1, 2, \ldots, n-k$ by

$$J_a(v) = g(d\theta^a(v, \cdot))$$

where g is the cometric. Then fatness implies that for each nonzero covector $(\lambda_1, \lambda_2, \cdot, \lambda_{n-k})$ the operator $\Sigma \lambda_a J_a$ is invertible. Thus, for each nonzero vector v the collection $\{v, J_1(v), \ldots, J_{n-k}(v)$ is linearly independent in D_q.

iv) These are the vector fields $v \to J_a(v)$.

The converse follows from a theorem of Adams which states that condition (iv) implies that the Clifford algebra C_{n-k} has a representation on $I\!R^k$. This means that there exist $n-k$ linear maps $J_a : I\!R^k \to I\!R^k$, $a = 1, 2, \ldots, n-k$ which are skew-symmetric and orthogonal and skew-commute with each other: $J_a J_b + J_b J_a = -2\delta_{ab} I$. It follows that any linear combination $J(\lambda) = \Sigma \lambda_a J_a$ of them satisfies $J(\lambda)^2 = -\|\lambda\|^2 I$. Now put coordinates $(x, y) = (x_1, \ldots, x_k, y^1, \ldots, y_{n-k})$ on $I\!R^n$ and define the distribution $\mathcal{D}$ to be the one annhilated by the $n-k$ forms

$$\theta_a = dy_a - \Sigma_{ij} J_a^{ij} x_i dx_j.$$

Using the standard inner product on $I\!R^k$ and the linear projections of the $\mathcal{D}_{(x,y)}$ onto $I\!R^k$ we dentify the form $w(\lambda)$ with $J(\lambda)$. Hence all the $w(\lambda)$ are are symplectic for $\lambda \neq 0$.

EXAMPLES The only examples of fat distributions which are not contact **with which I am familiar** relate to the quaternions. Distributions of quaternionic, or more generally, Clifford type, **are not** the only types of non-contact fat distributions. This can be seen by a dimension counting argument.

To define a quaternionic version of a CR manifold let $Q \subset \mathbb{H}^{k+1}$ be a real hypersurface of a quaternionic vector space. The distribution on Q is defined by letting $\mathcal{D}_q$ be the maximal dimensional quaternionic subspace of T_qQ. Thus

$$\mathcal{D}_q = T_qQ \cap iT_qQ \cap jT_qQ \cap kT_qQ.$$

The dimension of $(\mathcal{D}, Q)$ is $(8k, 8k+3)$. For generic hypersurface this will be a fat distribution. For more on these manifolds see §V. 5 of [3] and references therein where they are called QR manifolds.

Another class of examples is realized by self-dual connections. The typical instanton, or self-dual SU(2) connection, over a four-manifold will be fat. In other words it defines a fat distribution on its principal bundle.

The standard seven-sphere admits a rank four fat distribution which can be obtained in both ways. To realize it as a QR structure embed S^7 in $\mathbb{R}^8 = \mathbb{H}^2$. To realize this distribution as coming from a connection we realize S^7 as the total space for the quaternionic Hopf fibration $S^7 \to S^4$ and put the standard instanton gauge field on this bundle. The symmetry group of this distribution is the 21-dimensional Lie group $Sp(2,1;\mathbb{H})$ of quaternionic linear maps of $\mathbb{H}^3$ which preserve the split quaternionic form $|q_0|^2 + |q_1|^2 - |q_2|^2$. S^7 is isomorphic to the quadric $\{|q_0|^2 + |q_1|^2 - |q_2|^2 = 0\}$ in the 2-dimensional quaternionic projective plane $\mathbb{P}\mathbb{H}^2$ and this defines the action of $Sp(2,1;\mathbb{H})$ on S^7. The usual symmetry group $Sp(2;\mathbb{H}) \times Sp(1;\mathbb{H})$ of the standard instanton sits inside $Sp(2,1;\mathbb{H})$ in the obvious way.

This 7-sphere can be identified with the set of points at infinity for the quaternionic hyperbolic plane $\{|q_0|^2 + |q_1|^2 - |q_2|^2 < 1\} \subset \mathbb{P}\mathbb{H}$. The isometry group of this space is again $Sp(2,1;\mathbb{H})$ and extends to the action on the sphere at infinity. The Riemannian metric on this plane kinduces a conformal subRiemannian structure on the 7-sphere at infinity. This structure has as its underlying distribution the one just described. Examples of this type underlie connections between subRiemannian geometry and rigidity phenomena in Riemannian geometry which has been explored by Pansu and Hamenstädt.

CONFORMAL STRUCTURES ON $\mathcal{D}^\perp$ [I am indebted to R. Bryant for the observations here.]

Given a distribution $\mathcal{D}$ of dimension $(4,7)$ we can define a fiberwise conformal structure $(\cdot,\cdot)$ on its annihilator $\mathcal{D}^\perp$ by

$$(\lambda, \nu) = d\lambda \wedge d\nu|_{\mathcal{D}}$$

(A choice d^4x of section of $\Lambda^4\mathcal{D}^*$ defines a bona-fide fiber bilinear form by: $(\lambda,\nu) = d\lambda \wedge d\nu|_{\mathcal{D}}/d^4x$.) The "light-cone" $\{\lambda \in \mathcal{D}^\perp : (\lambda,\lambda) = 0\}$ is the characteristic variety Σ. Consequently, the distribution is fat precisely when this bilinear form is definite.

Bryant (unpublished) has analyzed the possible symmetry groups for nondegenerate distributions of type $(4,7)$ and has shown that it is always a finite dimensional Lie group of dimension **at most** 21. This maximum symmetry is attained for only one elliptic distribution, namely the previous example on S^7, and for only one hyperbolic example. The underlying manifold for the hyperbolic example is the Grassmannian of isotropic 2-planes in a 6-dimensional symplectic vector space. The symplectic orthogonal complement of such an isotropic 2-plane E is a coisotropic 4-plane $E^\perp \supset E$. The distribution plane $\mathcal{D}_E$ at E is spanned by the tangents to the curves obtained by keeping $E^\perp$ fixed and spinning E within it. The group $Sp(3;\mathbb{R})$ of linear symplectic isomorphisms of the 6-dimensional vector space is the 21-dimensional symmetry group.

9 MOST DISTRIBUTIONS ARE DETERMINED BY THEIR SINGULAR CURVES

In this section we announce a new result inspired by the the question:

IS A NONHOLONOMIC DISTRIBUTION DETERMINED BY ITS SINGULAR CURVES?

posed to us in 1992 by Jakubczyck. Details will be provided in a subsequent publication.

In the generality stated, the answer to Jackubcyk's question is "no" due to the existence of moduli of fat distributions. In other words, there exist continuous families of distributions each of which admit no singular curves and no two of which are locally diffeomorphic. However the rank 2 examples and the hyperbolic case in dimension $(4, 7)$ just discussed indicate that for a large class of dimensions (k, n) the answer may generically be "yes".

Let us make Jacubcyck's question more precise.

Definition 15 *A distribution germ is* STRONGLY DETERMINED BY ITS SINGULAR CURVES *if every diffeomorphism germ which takes its singular curves to the singular curves of another distribution of the same dimension is the germ of a diffeomorphism which takes this distribution to the other.*

Theorem 14 *Consider distributions of dimension (k, n) for which the corank $n-k$ satisfies $n-k \geq 3$. There is a Whitney-open set of such distribution which are strongly determined by their singular curves. These distributions have the property that the k-plane through any point is the linear hull of the tangents to the singular curves passing through that point. If (k, n) is outside of the fat range (see proposition 6 above) then this set of distributions is dense.*

REMARK Zhitomirskii [60] [62] has conjectured that in dimension $(n-1, n)$ and $(2, 4)$ distributions are WEAKLY DETERMINED by their singular curves. By this we mean that if two such distributions have orbit-equivalent classes of singular curves then they are in fact diffeomorphic. But a diffeomorphism which takes one set of curves to the other need not be a diffeomorphism between the distributions. (Consider the Engel case or odd-rank contact case.) In these papers Zhitomirskii classified the stable nonregular distributions and his conjecture holds for them.

SKETCH OF PROOF OF THE THEOREM

[$k = 2l + 1$ odd.] If k is odd then the annihilator $\mathcal{D}^{\perp}$ has odd dimenison so that the restricted canonical two-form, ω, has a nontrivial kernel at every point $\lambda \in \mathcal{D}^{\perp}$. For a generic $\mathcal{D}$ and a generic point λ of $\mathcal{D}^{\perp}$ the dimension of this kernel will be 1. This kernel will vary smoothly in a neighborhood of λ, thus defining a line-field whose integral curves are characteristics. Recall that the differential of the projection maps this kernel linearly isomorphically onto the kernel of the two-form $w(\lambda)$ on $\mathcal{D}_q$, $\pi(\lambda) = q$. The map

$$\lambda \mapsto ker(w(\lambda))$$

thus associates to each generic covector a direction in $\mathcal{D}_q$ which is the tangent to the singular curve corresponding to the characteristic passing through λ. Since $w(t\lambda) = tw(\lambda)$ for $t \in I\!R$ this map is defined projectively as a map

$$(I\!P\mathcal{D}^{\perp}{}_q)^{reg} \subset I\!P\mathcal{D}^{\perp}{}_q \to I\!PD_q$$

where the superscript "^{reg}" indicates that we must restrict to the open subset of such generic (or "regular") λ. This map is the projectivization of a homogeneous degree l vector valued polynomial. The theorem is proved by showing that for generic distribution germs this map is FREE in the following sense.

Definition 16 *A subset of a vector space is called* FREE *if its linear span is the entire vector space. A map into a vector space is called* FREE *if its image is free. A map into a projective space is called free if the corresponding map into the unprojectivized vector space is free.*

EXAMPLE If $V = I\!R^k$ then the curve $t \to (1, t, t^2, \ldots, t^{k-1})$ is free.

EXAMPLE A representation of a group G on a vector space V is irreducible if and only if for each nonzero vector v the orbit map $G \to V$ given by $g \mapsto gv$ is free.

REMARK Freeness is an open condition. The components of a nonfree map must satisfy a linear dependence condition, hence the word "free".

REMARK For $c \geq k$ the map is a submersion at typical λ.

[THE CASE $k = 2l$ ODD] The main ideas are the same. The proof is significantly complicated by the fact that at typical points λ the form $w(\lambda)$ has no kernel. So we must restrict the entire discussion to the characteristic variety

$$\Sigma = \{\lambda \in \mathcal{D}^{\perp} : w(\lambda)^l = 0\}$$

which is the set of λ for which w has a kernel. We expect Σ to be a hypersurface (typically singular). For generic distribution, and for generic point $\lambda \in \Sigma$ we expect:

- 1) Σ is smooth near λ
- 2) the kernel of the restricted two-form at λ is a two-plane,
- 3) this two-plane intersects $T_\lambda \Sigma$ transversely.

The set of such λ will then form an open dense subset, say $\Sigma^{reg} \subset \Sigma$, typically the complement of some closed subvariety. And on this subvariety the intersection of the kernel with the tangent space defines a line field.

In this manner we obtain a map

$$I\!P\Sigma^{reg} \to I\!P D_q.$$

It is defined by homogeneous polynomials of degree $2l - 1$. This map is generically free provided the corank is greater than or equal to 3 and the proof is finished as in the odd rank case. ($I\!P\Sigma^{reg}$ has dimension $c - 2$ and hence is a curve when the corank c is 3.) Details of the argument will be provided in a future publication.

10 SINGULAR CURVES DOMINATING SPECTRAL ASYMPTOTICS

This section is a summary and discussion of a new result which appears in [40].

If subRiemannian manifolds inherited a geometrically natural, **smooth**, volume measure $d\mu$ then they would have a natural subLaplacian, whose eigenvalues would be defined by the Dirichlet (Rayleigh-Ritz) principle for the "horizontal Dirichlet energy":

$$\int g(df, df) d\mu.$$

Here g is the cometric (§3.1), that is, the bilinear form on the cotangent bundle whose associated quadratic form is the geodesic Hamiltonian H.

When $(\mathcal{D}, Q)$ is a principal bundle with connection (example 1, §2) there is such a measure. It is locally the product of the Riemannian measure on the base with Haar measure on the fiber. (See §2.5) Its subLaplacian is the covariant (or horizontal) Laplacian $\hat{H}$ which acts on functions ψ on the principal bundle according to the coordinate formula:

$$\hat{H}\psi = -\frac{1}{\sqrt{g}} \Sigma D_\mu (g^{\mu\nu} \sqrt{g} D_\nu \psi).$$

Here the D_μ are the covariant differentials in the coordinate directions on the base, which is to say they are the directional derivative operators in the direction of the horizontal lifts of the $\frac{\partial}{\partial x^\mu}$. The $g^{\mu\nu}$ are the inverse matrix coefficients on the Riemannian base space. (If we replace the D_μ by the $\frac{\partial}{\partial x^\mu}$ we would have the regular Laplacian acting on functions on the base space.) A change in the Haar measure merely multiplies the entire measure by a constant, and thus leaves the subLaplacian fixed up to an overall multiplicative constant.

If Q is a circle bundle then the covariant Laplacian is the quantum Hamiltonian $\hat{H}$, or Schrodinger operator, for a nonrelativistic charged particle on the Riemannian base space travelling under the influence of the magnetic field given by the curvature of the (given) connection. The charge $\lambda \in \mathbb{Z}$ indicates what representation of the circle the quantum wavefunction ψ is in. More specifically, its charge is λ if

$$\psi(e^{i\theta} q) = e^{i\lambda\theta}\psi(q)$$

where $q \to e^{i\theta} q$ indicates the circle action on Q. Let $\mathcal{L}^* = \mathcal{L}^{-1}$ denote the Hermitian line bundle whose unit vectors form Q. The set of ψ which transform in this manner is naturally isomorphic to the space of square integrable sections of its λth power, $\mathcal{L}^\lambda$. We will denote this space of sections by $\Gamma(\lambda) = \Gamma(\mathcal{L}^\lambda)$;

$$L_2(Q) = \bigoplus_{\lambda \in \mathbb{Z}} \Gamma(\lambda)$$

Since $\hat{H}$ commutes with the S^1 action it commutes with orthogonal projection $I\!P_\lambda$ onto $\Gamma(\lambda)$. Thus

$$\hat{H} = \bigoplus_{\lambda} \hat{H}(\lambda)$$

where

$$\hat{H}(\lambda) = I\!P_\lambda \hat{H} I\!P_\lambda = \hat{H} I\!P_\lambda.$$

$\hat{H}(\lambda)$ is the standard quantum Hamiltonian for a nonrelativistic particle in the magnetic field $B = *dA$ corresponding to the curvature of this bundle, provided the particle has spin zero and charge λ . (Actually, there are various units we have suppressed, eg. the mass, and Planck's constant. λ corresponds to (charge)/(Planck's constant).)

In theorem 4 we saw considered the case $Q \to X$ where X is a Riemannian surface. We saw that the singular geodesics corresponded to the zero locus of a magnetic field.

QUESTION: DO SINGULAR GEODESICS PERSIST UPON QUANTIZATION?

In our sketch of the proof of theorem 4 we saw that there is a sense in which the singular geodesics correspond to the limit of infinite charge of regular sub-Riemannian geodesics. This suggests that we should investigate the spectral asymptotics of $\hat{H}(\lambda)$ as $\lambda \to \infty$ in order to see the quantum effects of the singular geodesic.

We will make the following assumptions.

- (A1)The zero locus C is a compact connected curve
- (A2)If the surface X has a boundary then C does not intersect the boundary. Use Dirichlet conditions on the boundary.
- (A3)If the surface is noncompact then the magnetic field B is bounded away from zero at infinity.

Theorem 15 *Consider the family $\hat{H}(\lambda)$ of covariant Laplacians in the situation of theorem 4 where the magnetic field vanishes in a nondegenerate manner along a closed curve C. Make the above assumptions (A1), (A2), (A3). Then the normalized ground state (= lowest eigensection) for $\hat{H}(\lambda)$ tends to a probability measure concentrated on C as $\lambda \to \infty$. Its energy (= lowest eigenvalue) is $O(\lambda^{2/3})$. The same is true for all eigenfunctions and values, with the level j of the eigenvalue $E_j(\lambda)$ fixed as we let $\lambda \to \infty$. If the gradient of the magnetic field has constant magnitude, say b_0, along C then we can be more precise about the energy asymptotics:*

$$E_j(\lambda) = \lambda^{2/3} b_0^{2/3} E_* + O(1)$$

where E_* is a universal constant; $E_* \simeq .5698$.

In physical terms, the fact that the eigensections concentrate on the zero locus C means that as the charge becomes very large (or Planck's constant very small) it becomes almost certain that the particle is very close to C. The form of the limiting distribution on C is not known.

This eigenvalue growth specified by this theorem should be contrasted with:

Theorem 16 *Suppose that the magnetic field on the surface satisfies $|B| \geq B_0$ for some positive constant B_0. Then*

$$E_1(\lambda) \geq B_0 |\lambda|.$$

which presumably holds in spirit for a general 3-dimensional contact manifold.

We view the phenomenon of eigenfunction concentration and eigenvalue growth governed by singular geodesics C as the first instance of a general relation between the spectral asymptotics of a class of second order subelliptic operators and the singular geodesics of their corresponding symbols – the underlying subRiemannian metric. The charge λ is a measure of the vertical or transverse part of the gradient of a function. So in a more general treatment the large asymptotic parameter will be a measure of the size of the components of the gradients transverse to the distribution.

RELATED RESULTS AND CONJECTURES Guillemin and Uribe [21] [22] [23] have related the spectral asymptotics for covariant Laplacians **on fat bundles** to the subRiemannian geodesic flow corresponding to the normal Hamiltonian H. Christ [11] has conjectured that the presence of the singular curves of the type occuring in Martinet's normal form (that is, the singular minimizers of theorem 4 above) signals the breakdown of analytic hypoellipticity. Their are hints (Sussmann, private communication; Bismut [4]) that the small-time asymptotics for the subelliptic heat kernel is dominated by contributions from singular curves.

11 Bibliography

References

[1] A.A. Agrachev and A.V. Sarychev, *Abnormal Sub-Riemannian Geodesics: Morse Index and Rigidity*, preprint, 1993.

[2] W. Ambrose and I.M. Singer, "A theorem on holonomy", Trans. AMS,75, 428-453, (1953).

[3] Bejancu, A. **Geometry of CR-Submanifolds**, D. Reidel, Dordrecht, Holland, 1986.

[4] J-M. Bismut, **Large Deviations and the Malliavin Calculus**, Birkhauser, 1984.

[5] G.A. Bliss, **Lectures on Calculus of Variations**, Univ. of Chicago Press, 1946.

[6] R.W. Brockett,*Nonlinear Control Theory and Differential Geometry* , Proc. of the Int. Congress of Mathematicians, Warszawa, (1983).

[7] R.L. Bryant, S.S. Chern, R.B. Gardner, H.L. Goldschmidt, and P.A. Griffiths **Exterior Differential Systems** v. 18, MSRI Publications, Springer-Verlag, 1991.

[8] R. Bryant and L. Hsu, *Rigidity of Integral Curves of Rank Two Distributions*, Invent. Math., **114**, 435-461, (1993).

[9] E. Cartan, *Les systèmes de Pfaff à cinque variables et lès équations aux dérivées partielles du second ordre*, Ann. Sci. Ècole Normale, no. 3, **27**, 109-192, (1910).

[10] W.L. Chow ,*Uber Systeme van Linearen partiellen Differentialgleichungen erster Ordnung*, Math. Ann., vol. 117, 98-105, (1939).

[11] M. Christ, *Analytic Hypoelliptic Breaks Down for Weakly PseudoConvex Reinhardt Domains* preprint, UCLA, (1991).

[12] K.J. Falconer, **The geometry of fractal sets**, Cambridge Univ. Press, (1985).

[13] Ge Zhong, *Horizontal path space and Carnot-Caratheodory Metrics*, Pac. J. Math, **161**, 255-286, (1993).

[14] Ge Zhong, *Caustics in Optimal Control: An example when the symmetry is broken* Lect. in Applied Math., **29**, 203-212, (1993).

[15] Ge Zhong, *On the cut points and conjugate points in a constrained variational problem*, Field Inst. Comm., **1**, 1993.

[16] Ge Zhong, *On a constrained variational problem and the space of horizontal paths*, Pac. J. Math., **149**, 61-94.

[17] V. Gershkovich, Engel Structures on Four-Dimensional Manifolds, preprint, U. of Melbourne, (1993).

[18] M. Gromov, J. Lafontaine, and P. Pansu, **Structures metriques pour les varietes Riemanniennes**, Paris, Cedic, [1981].

[19] M. Gromov, Carnot-Caratheodory spaces seen from within, IHES preprint M/94/6, (221 pages), (1994).

[20] A. Guichardet *On rotation and vibration motions of molecules* Ann. Inst. H. Poincare, Phys. Theor., **40** no.3 329-342 (1984).

[21] V. Guillemin and A. Uribe, *The Trace Formula for Vector Bundles*, Bulletin of the A.M.S., **14** no. 2, 222- 224, (1986).

[22] , V. Guillemin and A. Uribe, *Some Spectral Properties of Periodic Potentials* in Lect. Notes in Math ., No. 1256, 192-214, Springer.

[23] , V. Guillemin and A. Uribe, *The Laplace Operator on the nth Tensor Power of a Line Bundle: Eigenvalues Which are Uniformly Bounded in n* Asymptotic Analysis, v. 1, 105-113 (1988).

[24] U. Hamenstädt , *Some Regularity Theorems for Carnot-Caratheodory Metrics* , J.D.G. , v. 32, 819-850 (1990).

[25] U. Hamenstädt , *Zur Theorie von for Carnot-Caratheodory Metriken und ihren Anwedungen* , Bonner Math. Schriften, Univ. Bonn, No. 180, (1987).

[26] R. Hermannn, *Some Differential Geometric Aspects of the Lagrange Variational Problem* , Indiana Math. J., 634-673, (1962).

[27] L. Hsu, *Calculus of Variations via the Griffiths Formalism*, J.D.G., vol 36, no. 3, 551-591, (1991).

[28] Th. Kaluza, *Zum unitatsproblem der physik*, Berlin Berichte, p. 966, (1921).

[29] , R. Kerner, *Generalization of the Kaluza-Klein theory for an arbitrary non-Abelian group*, Ann. Inst. Henri Poincare, **9** 143-152, (1968).

[30] I. Kupka,*Abnormal Extremals*, preprint, (1992).

[31] W-S. Liu and H.J. Sussmann, *Shortest paths for subRiemannian metrics on rank two distributions*, preprint, to appear, Trans. A.M.S., (1994).

[32] J. Martinet, *Sur les Singularites Des Formes Differentialles*, Ana. Insp. (Grenoble), vol. 20, no. 1, 95-198, (1970).

[33] J. Mitchell, *On Carnot-Caratheodory Metrics*, J. Diff. Geom. v. 21, 34-45, (1985).

[34] R. Montgomery, *Abnormal Minimizers*, to appear, SIAM J. Control and Opt., (1994).

[35] R. Montgomery, *The Isoholonomic Problem and Some Applications*, Comm. Math. Phys., vol. 128, 565-592, (1990).

[36] R. Montgomery, *Singular Extremals on Lie Groups*, to appear, MCSS, (1995).

[37] R. Montgomery, *Gauge Theory and Control Theory*,pp. 343-378, in **Nonholonomic Motion Planning** J. Canny and Z. Li ed., Kluwer Acad. Press,Norwell, MA., USA, (1993).

[38] R. Montgomery, *Gauge Theory and the Falling Cat*, pp 193-218, in **Dynamics and Control of Mechanical Systems**, M. Enos, ed., Fields Inst. Communications series, vol 1 , AMS pub., Providence, RI. (1993).

[39] R. Montgomery, *Generic Distributions and Finite Dimensional Lie Algebras* , Journal of Differential Equations, **103** pp. 387-393 (1993).

[40] R. Montgomery, *Hearing the zero locus of a magnetic field*, preprint, (1994).

[41] M. Morse and S. Myers, *The Problems of Lagrange and Mayer with Variable Endpoints*, Proc. of the Am. Acad. v. 66, no. 6, 236-253, (1931).

[42] R. Murray *Nilpotent Bases for a Class of Non-Integrable Distributions with Applications to Trajectory Generation for Nonholonomic Systems*, CDS Technical Memorandum no. CIT-CDS 92-002, Calif. Inst. of Tech., (1993).

[43] A. Nagel, E. M. Stein, and S. Wainger, *Balls and metrics defined by vector fields*, Acta Math. **155**, 103-147, (1985).

[44] P. Pansu, *Metriques de Carnot-Caratheodory et quasiisometries des espaces symetriques de rang un*, Ann. of Math., **129**, 1-60, (1989).

[45] L. S. Pontrjagin, V. G. Boltyanskii, R. V. Gamkrelidze, and E. F. Mishchenko, **The Mathematical Theory of Optimal Processes**, Wiley, Interscience, [1962].

[46] C.B. Rayner *The Exponential Map for the Lagrange Problem on Differentiable Manifolds*,Phil. Trans. of the Royal Soc. of London, ser. A, Math. and Phys. Sci. no. 1127, v. 262, 299-344, (1967).

[47] L.P. Rothschild and E.M. Stein, *Hypoelliptic differential operators and nilpotent groups*, Acta Math, **137**. 247-320, (1976).

[48] A. Shapere and F. Wilczek, *Self-propulsion at low Reynolds number* , Physical Review Letters, vol. 58, 2051-2054, (1987).

[49] A. Shapere and F. Wilczek, *Geometric Phases in Physics*, World Sci., Singapore, (1989).

[50] E.D. Sontag, **Mathematical Control Theory**, Springer-Verlag, New York, (1990).

[51] S. Sternberg, On minimal coupling and the symplectic mechanics of a classical particle in the presence of a Yang-Mills field, Proc Nat Acad Sci, **74**, 5253-5254, (1977).

[52] R. Strichartz, *Sub-Riemannian Geometry*, J. Diff. Geom. 24, 221-263, (1983).

[53] R. Strichartz, *Corrections to "Sub-Riemannian Geometry"*, J. Diff. Geom. 30, no. 2, 595-596, (1989).

[54] A.M. Vershik and V. Ya Gerhskovich *An estimate of the functional dimension of the space of orbits of germs of generic distributions* Math. USSR, Zametki, **44:45** 596-603, (1988).

[55] A.M. Vershik and V. Ya Gerhskovich *Nonholonomic Dynamical Systems, Geometry of Distributions and Variational Problems* in **Dynamical Systems VII** ed. V.I. Arnol'd and S.P. Novikov, vol. 16 of the Encyclopaedia of Mathematical Sciences series, Springer-Verlag, NY, 1994. (Russian original 1987.)

[56] A. Weinstein, *Fat bundles and symplectic manifolds*, Adv. in Math, **37**, 239-250, (1980).

[57] A. Weinstein, *A universal phase space for a particle in a Yang-Mills field*, Lett. Math. Phys., **2**, 417-420, (1978).

[58] S.K. Wong [1970], "Field and Particle Equations for the Classical Yang-Mills Field and Particles with Isotopic Spin", Nuovo Cimento, 65a, 689-693.

[59] L.C. Young, **Lectures on the Calculus of Variations and Optimal Control Theory**, Chelsea, 1980.

[60] M. Ya. Zhitomirskii, **Typical Singularities of Differential 1-forms and Pfaffian Equations** Transl. of Math. Monographs series, vol. 113, A.M.S., Providence, RI, (1992).

[61] M. Ya. Zhitomirskii, *Singularties and Normal Forms of Odd-Dimensional Pfaff Equations* Funktsional. Anal. I Ego Prilozhen. **23**, no. 1, 70-71, (1989).

[62] M. Ya. Zhitomirskii, *Normal Forms of germs of 2-dimensional distributions in* $I\!R^4$ Funktsional. Anal. I Ego Prilozhen., **24**, no. 2, 81-82, English translation: Funct. Anal. Appl. **24**, (1990).

[63] J.W. Zwanziger, M. Koenig, and A. Pines, *Berry's Phase*, in the Annual Review of Chemistry, (1990).