Singular Extremals on Lie Groups

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

May 24, 1994

1 Introduction

We investigate the space of abnormal or singular extremals of nonlinear deterministic control systems linear in the controls. This article began as an attempt to understand some unpublished work of U. Hamenstädt's [7] on the space of singular curves for left-invariant systems on Lie groups. We found that a "symplectic" approach espoused by Hsu [8] clarified the situation and simplified calculations.

Fix an initial state. By the ENDPOINT MAP we will mean the map that assigns to each control the corresponding final state. The critical "points" of this map are the controls corresponding to abnormal extremals. (See, for example, Sontag's text [14] p. 56, or that of Pontrjagin et al [12].) We will call the corresponding curves in state space the SINGULAR CURVES.

Introduce a running cost function which is quadratic positive-definite in the controls. The corresponding optimal control problem is known as the "subRiemannian" geodesic problem and is a special case of the problem of Lagrange in the calculus of variations. A basic, unanswered question is "are all subRiemannian minimizers smooth?" Since normal extremals are automatically smooth, this is equivalent to the question "are all singular subRiemannian minimizers smooth?" This question motivated our work. By the way, for this particular optimal control problem, the class of singular and abnormal extremals coincide, so there will be no confusion by our use of the term singular curve or singular extremal. (See an appendix to [10] for more on this.)

Our main result concerns the singular curves for a class of systems whose state spaces are compact Lie groups. We show that each such extremal lies in a lower dimensional subgroup within which it is regular. We use this result to prove that all subRiemannian minimizers for such a system are smooth.

A central ingredient of our proof is a symplectic geometric characterization of the singular curves. Lucas Hsu [8] found this characterization in the case of SMOOTH singular curves. In order to make it useful, we must extend his characterization to nonsmooth singular curves, which we do in propositions 1 and 2 of §3. We have found the symplectic point of view quite useful in performing calculations and believe that it is essential for further understanding of singular curves.

ACKNOWLEDGEMENTS. In addition to essential conversations with Ursula Hamenstädt and Lucas Hsu we would like to acknowledge a crucial conversation with Viktor Ginzburg. We would also like to acknowledge useful conversations with Hector Sussmann, Robert Bryant, Ge Zhong, Eugene Lerman, and Andrey Todorov, as well as the receipt of a handwritten manuscript from Hamenstädt. We thank the reviewer for a thorough job and a question regarding the open problem (§3.6.1) concerning distributions of odd rank. This work was supported by National Science Foundation grant DMS-912708 and a Faculty Research Grant from the University of California.

2 SubRiemannian Geometry and Characteristic Submanifolds

Consider a system of the form

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

with cost functional

$$E[u] = \int_0^T \frac{1}{2} \Sigma (u^a(t))^2 dt \tag{2}$$

Here q is the state variable, a point on an n-dimensional manifold Q. The X_a are a family of smooth vector fields on Q. We write u for the control vector $(u^1, \ldots, u^k)$. The goal is to find controls $t \to u(t)$ which steer between two given points, say q_0 to q_1, in time T, and in such a way as to minimize $E[u]$ over all such controls.

We prefer to rephrase this problem in the language of differential geometry. We will suppose that the vector fields X_a are linearly independent at each point q and so define a field of k-planes $D_q \subset T_q Q$, where TQ denotes the tangent bundle to Q. (Assume $k < n$ so that the problem is different from Riemannian geometry.) The union $D = \cup_{q \in Q} D_q \subset TQ$ of these k-planes forms a distribution on Q. Declaring the vector fields X_a to be an orthonormal frame for D defines a fiber inner product $\langle \cdot, \cdot \rangle$ on D. A geometric structure consisting of such a pair $(D, \langle \cdot, \cdot \rangle)$ is called a *subRiemannian* metric [15], or alternatively a *singular Riemannian* [5], or *Carnot-Caratheodory* [6] metric.

In these terms our optimal control problem becomes *the subRiemannian geodesic problem:* Find a path $\gamma : [0,1] \to Q$ which minimizes the integral

$$E(\gamma) = \frac{1}{2}\int \langle \dot{\gamma}, \dot{\gamma} \rangle dt$$

subject to the constraints

- (a) $\dot{\gamma}(t) \in D$ whenever this derivative exists
- (b) $\dot{\gamma}$ is square integrable (and in particular the derivative $\dot{\gamma}(t)$ exists for almost all t)
- (c) $\gamma(0) = q_0$
- (d) $\gamma(1) = q_1$.

We will call a solution to this problem a *minimizing subRiemannian geodesic* or simply a *minimizer*. A basic open question is, "are all minimizers smooth?"

Let Ω_D, $\Omega_D(q_0)$, and $\Omega_D(q_0, q_1)$ be the space of all curves satisfying the constraints (a)-(b), respectively (a)-(c), and (a)-(d). These path spaces do not depend on the choice of inner product $\langle \cdot, \cdot \rangle$ on D for if $\dot{\gamma}$ is integrable with respect to one smooth metric on D then it is integrable with respect to any other. Bismut [3] shows how to give $\Omega_D(q_0)$ the structure of a Hilbert manifold.

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

Thus $\Omega_D(q_0, q_1) = end_{q_0}^{-1}(q_1)$. Bismut observed that *end* is a smooth map and calculated its derivative $d(end)$. This derivative is well-known and can be found in many places, for example in Bismut's book, or at the top of p. 57 of Sontag's text. The formula for the derivative is repeated later on in this paper.

Definition 2 *A* **singular curve** $\gamma \in \Omega_D$ *is a singular point of* $end_{\gamma(0)}$. *A curve* $\gamma \in \Omega_D$ *is* **regular** *if it is not singular, i.e. if end is a submersion at* γ.

REMARKS.

• The condition that a curve is singular depends only on the distribution and not at all on the inner product on the distribution.

• (fatness.) Most of the work on subRiemannian geometry assumes that the underlying distribution is "fat" which is synonymous with "strong bracket generating". (See [17].) Fat distributions are ones for which there are no singular curves besides the trivial constant curves. They are rare objects. For example, Rayner [13] showed that if $n - m \neq 1$ then the rank m of a fat distribution must be a multiple of 4 and also must be less than $n(n-1)/2$.

Lemma 1 *If a minimizer is regular then it is smooth.*

This is proved by using the implicit function theorem, together with the method of Lagrange multipliers. See for example Hamenstädt [6].

In view of this lemma, the question of the smoothness of minimizers reduces to the question of the smoothness of the singular minimizers. Hamenstädt suggested the following idea in order to investigate this last question. Try to associate to every singular curve γ some smooth submanifold $M = M(\gamma) \subset Q$ containing it and such that γ is regular as an integral curve for the restriction $D_M := D \cap TM$ of D to M. (For simplicity, assume also that D_M has constant rank and so forms a smooth distribution on M.)

Definition 3 *We will call such a submanifold $M(\gamma)$ a "characteristic submanifold" for the singular curve γ.*

Now (M, D_M) is a subRiemannian manifold in its own right, with inner product inherited from D. If $\gamma \subset M$ is a minimizer for the subRiemannian geodesic problem on Q it is automatically a minimizer for the problem on M. So we can apply Lemma 1 to conclude

Corollary 1 *If every singular curve has a characteristic submanifold then every minimizing geodesic is smooth.*

To prove our main result we use this corollary.

Sussmann [16] gives an example of a singular curve which admits no characteristic submanifolds. (It is a nonsmooth characteristic curve for a smooth rank 2 distribution on $\mathbf{R}^3$.) Thus characteristic submanifolds don't always exist.

If a singular curve is a smooth embedded curve then it is a characteristic submanifold for itself. In order to try to make the characteristic submanifold M of a curve unique, at least as a germ of a submanifold in a neighborhood of the curve, one should add the condition that M be maximal among the class of all connected characteristic submanifolds of the curve. If G is a Lie group and D is invariant under left multiplication, then it is sensible to look for characteristic submanifolds which are Lie subgroups.

Definition 4 (Hamenstädt) *Suppose $Q = G$ is a Lie group and $D \subset TG$ is a left-invariant distribution. Let $\gamma \subset G$ be a D-curve passing through the identity. A closed connected subgroup K of G is called* **a characteristic subgroup** *of γ if it is a characteristic submanifold and is the largest connected characteristic submanifold which is also a Lie group.*

In the examples presented at the end of this paper every singular curve through the identity lies in a characteristic subgroup. However Bryant and Hsu [4] give an example of a distribution on the group G of rigid motions of three-space for which the points of a general singular curve through the identity generate the entire group. Thus characteristic subgroups do not exist for these curves. (Their curves are smooth.)

A SPECULATION. If D is analytic or appropriately generic, then every every singular curve lies in a characteristic submanifold.

3 Singular Curves

The goal of this section is to describe Hsu's characterization [8] of the smooth singular curves and extend it to all singular curves.

3.1 Microlocal Characterization

Let

$$D^0 \subset T^*Q$$

be the annihilator of the distribution D. It is a smooth subbundle of the cotangent bundle and its sections consist of one-forms which annihilate the control vector fields X_i. Let ω be the restriction of the canonical two-form $(\Sigma dp_i \wedge dq^i)$ to D^0.

Definition 5 *A characteristic for D^0 is an absolutely continuous curve in D^0 which never intersects the zero section and whose derivative lies in the kernel of ω whenever it exists. The characteristic will be called an H_1^p characteristic if its derivative is in L^p.*

To be explicit, an absolutely continuous curve $\zeta : [0,1] \to D^0$ is a characteristic if

- $\zeta(t)$ is never zero; i.e. if when we write it in standard cotangent coordinates $(q(t), p(t))$ on T^*Q, we have $p(t) \neq 0$ for all t between 0 and 1, and if
- $\omega(\zeta(t))(\dot{\zeta}(t), v) = 0$ for each t for which the derivative $\dot{\zeta}(t)$ exists and for each $v \in T_{\zeta(t)}D^0$ (Recall that absolutely continuous curves are differentiable almost everywhere and their derivatives are measurable functions.)

3.2 Pontrjagin's abnormal extremals

Define the *momentum functions* P_i on T^*Q associated to our control vector fields X_i by

$$P_i(q,p) = p(X_i(q)), i = 1, \ldots, k.$$

Here $p \in T_q^*Q$ is a costate. In other words, the P_i are the given vector fields viewed as fiber-linear functions on the cotangent bundle. Choose k measurable functions $u^i(t)$, $0 \le t \le 1$ and define the time-dependent Hamiltonian

$$H_u(q,p,t) = \Sigma u^i(t) P_i(q,p).$$

Introduce coordinates $q^\mu, \mu = 1, \ldots, n$ and dual coordinates $p_\mu, \mu = 1, \ldots, n$ so that together they form canonical coordinates on the cotangent bundle with typical covector being written $\Sigma p_\mu dq^\mu$. Then the Hamiltonian equations for H_h take the canonical form:

$$\begin{aligned} \dot{q}^\mu &= \Sigma u^i(t) X_i^\mu(q) \\ \dot{p}_\mu &= -\Sigma u^i(t) \frac{\partial X_i^\nu}{\partial q^\mu} p_\nu. \end{aligned} \quad (3)$$

The first equation simply says that $q(t) = \Phi_t(q(0))$ where Φ_t is the flow of the time-dependent vector field $\Sigma u^i(t) X_i$. **We will call any such time-dependent flow a** *horizontal flow*. (This terminology arises because we think of the distribution planes as the choice of "horizontal planes" and so such a flow is one for which all integral curves are horizontal.) The second equation says that $p(t)$ is the push-forward of $p(0)$ with respect to this flow:

$$p(t) = d\Phi_t(q_0)^{-1T} p(0).$$

Definition 6 *An H^2_1-abnormal Pontrjagin extremal is a solution $t \to (q(t), p(t))$ of the above equations (3) which also satisfies the constraints*

$$P_i(q(t), p(t)) = 0, i = 1, \dots k$$

and for which the u defining the equations (3) is in $L^2([0,1], \mathbf{R}^k)$

3.3 The Characterization

Proposition 1 *Let $\zeta : [0,1] \to T^*Q$ be an absolutely continuous path which never intersects the zero section and whose derivative is square integrable. Let $\gamma = \pi \circ \zeta : [0,1] \to Q$ be the curve over which ζ lies. Let q_0 and q_1 be the starting and ending points of γ. Then the following are equivalent.*

- *(i) $\zeta(t)$ annihilates the image of the differential , $R(t) = d(end(\gamma_t))$ for $0 \leq t \leq 1$, where γ_t is the restriction of γ to the interval $[0, t]$.*
- *(ii) $\gamma \in \Omega_D$, $\zeta(t) \in D^0_{\gamma(t)}$, and $\zeta(t) = d\Phi_t(q_0)^{-1T}\zeta(0)$, for $0 \leq t \leq 1$ where Φ_t is any t-dependent horizontal flow which generates the curve γ.*
- *(iii) ζ is an H^2_1 abnormal Pontrjagin extremal.*
- *(iv) ζ is an H^2_1 characteristic for D^0.*

As an immediate corollary we have

Proposition 2 *A D-curve $\gamma \in \Omega_D$ is singular if and only if there is a continuous everywhere nonvanishing H^2_1-characteristic curve $\zeta(t) \in D^0$ which projects onto γ. Let $R(\gamma)$ denote the image of the differential of the endpoint map at γ. For fixed γ , the set of all such characteristic curves with the zero curve included forms a vector space Γ_γ whose dimension is $n - dim(R(\gamma))$. The annihilator of $R(\gamma)$ is $\{\zeta(1), \zeta \in \Gamma_\gamma\}$.*

REMARKS.

• If, in the definition of Ω_D, we had used curves with derivative in L^p, $p \geq 1$, instead of L^2, i.e. L^p as opposed to L^2 controls, the same theorem would hold, the only change being that the characteristics would be in the Sobolev space H^p_1 instead of H^2_1.

• Parts of these propositions can be found in standard control literature, almost always for L^∞ instead of L^2 controls. For example see Sontag's text, [14], p. 56-57.

• If a singular curve is smooth then so is any characteristic projecting onto it. In this case Proposition 2 is due to Hsu [8]. Of course Hsu's result is of no help in regards to the question of whether or not every singular minimizer is smooth.

3.4 The differential of the endpoint map

The derivative of curve $\gamma \in \Omega_D$ can be expanded as

$$\dot{\gamma} = \Sigma h^i(t) X_i(\gamma(t)). \tag{4}$$

Let $\Phi_t = \Phi_t(\cdot; h) : Q \to Q$ denote the time-dependent (local) diffeomorphism defined by the time-dependent vector field $\Sigma X_i(q) h^i(t)$. Then

$$\gamma(t) = \Phi_t(q_0)$$

where $q_0 = \gamma(0)$, and $end(\gamma) = \gamma(1) = \Phi_1(q_0)$. The differential, or Jacobian matrix, $d\Phi_s(q_0) : T_{q_0}Q \to T_{\gamma(s)}Q$ of Φ_s is an invertible linear map. Then the derivative of end in the direction u at γ is

$$d(end(\gamma))(u) = d\Phi_1(q_0) \int_0^1 d\Phi_s^{-1}(q_0)(\Sigma X_i(\gamma(s)) u^i(s)) ds. \tag{5}$$

Another way to express this derivative is

$$d(end(\gamma))(u) = Z(1)$$

where $Z(t) = Z(t, u(\cdot))$ is a vector field along γ which satisfies a certain 1st order inhomogeneous linear ordinary differential equation defined by u. In terms of coordinates q_μ on Q this equation is

$$\frac{dZ^\mu}{dt} = \Sigma \frac{\partial X_i^\mu}{\partial q^\alpha}|_{\gamma(t)} h^i(t) Z^\alpha(t) + \Sigma X_i^\mu(\gamma(t) u^i(t) \tag{6}$$

$$\text{with } Z(0) = 0.$$

Observe that the curve $Z(t)$ is an H_1^2 vector field along γ. (More generally, if the u^i are in L^p then $Z(t)$ is a continuous vector field along γ whose derivative is in L^p.)

REMARKS.

• Versions of this formula for the derivative of the endpoint map can be found in many places. See for example [2], [12], the first chapter of Bismut [3], or chapter 2 of the text by Sontag [14].

• The square-integrable functions h^i are the coordinates of the curve γ with respect to a chart for the Hilbert manifold structure on $\Omega_D(q_0)$ defined by Bismut.

• The image $R(\gamma)$ of the differential of the endpoint map is independent of the parameterization of γ. The formula for the differential of the endpoint map **does** depend on the choice of frame X_i. However Bismut shows (p. 23-24) that its image is independent of frame. Bismut's formula on the top of his p. 24 is essentially the formula for the Jacobian of the coordinate transformation of $L^2([0,1], \mathbf{R}^k)$ which is induced by taking a new frame for D.

3.5 Proof of Proposition 1

A proof of the equivalence of items (i)-(ii)-(iii) is basically contained in Sontag [14], ch. 2.8, the only real difference being that he works with L^∞ as opposed to L^2 controls.

We will prove the equivalence of (iii) and (iv) which is Hsu's characterization.

Suppose that ζ is an H^2_1 abnormal Pontrjagin extremal. Then it must satisfy the constraint $P_i(\zeta(t)) = 0$. Also, ζ must satisfy Hamilton's equation for H_u. This equation can be written $\omega(\dot{\zeta}, v) = dH_u(v) = \Sigma u^i(t) dP_i(v)$ (a.e.) for any tangent vector v. Now a vector $v \in T(T^*Q)$ is tangent to $\mathcal{D}^0$ if and only if $dP_i(v) = 0, i = 1, 2, \ldots, n-k$. It follows that $\omega(\dot{\zeta}, v) = 0$ for all such v; i.e. ζ is an H^2_1 characteristic curve of $\mathcal{D}_0$. This logic is easily turned around: suppose ζ is an H^2_1 characteristic curve, that is, its derivative is in L^2 and $\omega(\dot{\zeta}, v) = 0$ (a.e) for all v tangent to $\mathcal{D}^0$. Now the one-forms $dP_i \in T^*(T^*Q)$ span the annihilator of $T\mathcal{D}^0 \subset T(T^*Q)$. It follows that we must have $\omega(\dot{\zeta}, \cdot) = \Sigma u^i dP_i$ (a.e) for some functions u^i of t. But this is Hamilton's equation for the time-dependent Hamiltonian H_u. Finally note that the u_i are in L^2 (or for that matter L^p) if and only if $\dot{\zeta}$ is in L^2 (resp. L^p).

QED.

3.6 Contact and Symplectic Considerations

3.6.1 The case of odd rank

The dimension of D^0 is $2dim(Q) - rank(D)$. Skew-symmetric bilinear forms on odd-dimensional spaces always have kernels. Consequently, if the rank of D is odd, then there is a characteristic vector passing through every point of D^0. If we knew that these vectors were tangent to actual characteristic curves we would know that a characteristic passed through every point of D^0, 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 2 *Suppose the rank of the distribution is odd. Then smooth characteristics pass through an open dense set of D^0 and consequently smooth singular curves pass through an open dense subset of Q.*

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

To illustrate the difficulties here we must recall that the set of two-forms is stratified by rank (see Martinet [11]). For typical distributions this stratification induces a stratification of D^0 according to the rank of ω. The big open stratum is the one on which the kernel of ω is one-dimensional and so we have a smooth line field on this stratum. The central difficulty is that this field typically DOES NOT extend in a continuous manner to the lower dimensional strata. For numerous examples of this phenomenon, see the book [18] by Zhitomirskii. We now present an example not found there (it is not a generic degeneration) which illustrates the phenomenon in a striking way.

Example.

The manifold is $\mathbf{R}^4$ with coordinates (x_1, x_2, y_1, y_2). The distribution is the three-plane field annihilated by the one-form:

$$\theta = dy_1 + S dy_2$$

where

$$S = \frac{1}{2}(x_1^2 + x_2^2).$$

The rank of $d\theta$ restricted to $\{\theta = 0\}$ is two AWAY FROM the plane $\{S = 0\}$. It follows that there is a unique singular curve passing through every point with $S \neq 0$. (For distributions of corank 1 we can directly study the singular curves through points as opposed to the characteristics in D^0, because the characteristics through θ and through $t\theta \in D^0$ are related by dilation and project to the same singular curves.) The kernel of this restricted two-form is spanned by the vector field

$$x_2 \frac{\partial}{\partial x_1} - x_1 \frac{\partial}{\partial x_2}.$$

Every characteristic in $\{S \neq 0\}$ must be tangent to this field. This vector field cannot be extended to the locus $\{S = 0\}$. In fact, if γ is any curve not contained in $\{S = 0\}$ but passing through it, then γ cannot be tangent to the direction field.

At any point of $\{S = 0\}$ the restriction of $d\theta$ to the distribution is zero. Now the distribution 3-planes intersect the tangent space to $\{S = 0\}$ along the line spanned by the vector $\frac{\partial}{\partial y_2}$. It follows that the singular locus $\{S = 0\}$, is also foliated by singular curves, namely the lines within it which are parallel to the y_2-axis. So in this example there is a unique singular curve passing through every point, but the corresponding direction field is discontinuous.

It may be possible to create a singularity of elliptic type for the direction field on the surface $\{S = 0\}$ by changing S (and perhaps adding dimensions and more forms). No singular curve could pass through such a point.

3.6.2 The case of even rank

In the remainder of this paper we will restrict ourselves to the case where 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 7 *The characteristic variety,* $\Sigma \subset D^0 \setminus 0$ *is the set of nonzero covectors in* D^0 *at which the two-form* ω *is not symplectic.*

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 $f d^N x$, where f is a function and $d^N x$ is a local volume form on D^0. Thus Σ is defined by the single scalar equation $f = 0$ and we expect it to be either empty or a hypersurface.

WARNINGS.

1. Typically Σ is not a smooth submanifold. However it appears that generically its a smooth stratified subvariety, the strata being points where the rank of ω is constant.

2. The intersection of the kernel of ω with the tangent space to Σ, or to one of its strata, may be smaller than the kernel of the restriction of ω to Σ. Consequently, even when this intersection $ker\omega \cup T\Sigma$ defines a distribution on Σ it can be a noninvolutive distribution. This is what happens in our examples at the end. We make this warning to deter the reader from making the following error. In the next section we show that at each point $\sigma = (q, p)$ the kernel of ω projects in a 1-to-1 manner onto a subspace of D_q. Consequently, if $ker\omega \cup T\Sigma$ were involutive, the projection of its leaves would provide us with characteristic submanifolds for every singular curve.

4 Calculating Characteristics

We follow Hsu's notation. Set $c = n - k = corank(D)$ and pick a local framing θ^a, $a = 1, 2, \ldots, c$ for D^0. Thus an arbitrary element of D^0 can be written uniquely as

$$\theta = \Sigma \lambda_a \theta^a. \tag{7}$$

This defines fiber coordinates $\lambda_a, a = 1, \ldots, c$ on D^0 and shows that D^0 has dimension $2n - k = n + c$. On the other hand, we can think of the θ^a as one-forms on D^0 by pulling them back from Q by the projection $\pi : D^0 \to Q$. Then this equation (7) is the expression for the restriction of T^*Q's canonical one-form to D^0. Thus

$$\omega = d\theta = \Sigma d\lambda_a \wedge \theta^a + \lambda_a d\theta^a \tag{8}$$

is the restriction of the canonical two-form to D^0.

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 $d\lambda_a$. Since the θ^a are nonzero, this implies that the intersection $ker d\pi \cap ker(\omega)$ is zero. This equation can be read in another way. The spaces $ker\omega$ can be defined by the Pfaffian system $i_Y \omega = 0$, as Y varies over all vector fields tangent to D_0. (A solution to this Pfaffian system is, by definition, a curve ζ such that $\zeta^*(i_Y \omega) = 0$ for all such Y.) Taking $Y = \frac{\partial}{\partial \lambda_a}$ we find that

$$\theta^a = 0$$

along any characteristic. This simply says that the projection of a characteristic must be a D-curve. Together these two facts imply

Lemma 2 *Let $(q, \lambda) \in D^0$ and let $d\pi = d\pi_{(q,\lambda)}$ denote the differential at (q, λ) of the canonical projection $\pi : D^0 \to Q$. Then $d\pi$ maps the kernel of ω at the point (q, λ) isomorphically onto some subspace of D_q.*

In order to obtain more detailed information we now complete the frame θ^a to form a (local) coframe θ^a, ω^μ , $\mu = 1, \ldots, k$ of Q, i.e. local trivialization of T^*Q as a vector bundle. 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 D^0. Relative to this frame, ω has the block form

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

where

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

is the restriction to $D \subset TQ$ of the form $\Sigma \lambda_a d\theta^a$ to $D \subset TQ$ and where "$*$" and its negative transpose "$-*^T$" denote matrices whose precise form do not matter for us. **It follows from linear algebra that the kernel of ω projects isomorphically, as in the above lemma, to the kernel of** $w(\lambda)$. Define the structure functions $c^a_{bc}, c^a_{\mu c}, c^a_{\mu\nu}$ relative to our frame by

$$d\theta^a = \Sigma c^a_{\mu\nu} \omega^\mu \omega^\nu + \Sigma c^a_{\mu c} \omega^\mu \theta^c + \Sigma c^a_{bc} \theta^b \theta^c.$$

(We suppress the $\wedge$s.) Then, since $w(\lambda) = \Sigma \lambda_a d\theta^a \text{mod}\{\theta^a\}$ we have

$$w(\lambda) = \Sigma \lambda_a c^a_{\mu\nu} \omega^\mu \omega^\nu. \tag{9}$$

REMARKS.

• There is a simple basis-independent description of $w(\lambda)$. Recall that the fiber coordinates λ represent a point θ of D^0 sitting in some fiber D^0_q. Extend this covector to form a local section of D^0, still denoted by the same symbol. Then $w(\lambda) = d\theta$ restricted to D_q. One easily sees that the bilinear form is independent of the choice of extension. See, for example, Rayner [13] for this calculation.

• We can express the characteristic variety Σ as the solution variety to the equation $w(\lambda)^l = 0$, $2l = dim(D_q)$. For fixed q this is a single homogeneous polynomial equation of degree l for the variable $\lambda \in D_q^0 \cong \mathbf{R}^{n-k}$. This shows that

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

is a real algebraic variety. (It may be desirable to projectivize thus obtaining $P\Sigma_q \subset PD_q^0$.)

As just described, the equations for the kernel of ω can be written as a Pfaffian system $\{i_Y \omega = 0\}$. By linearity, it suffices (locally) to restrict the set of vector fields Y to our framing $(\frac{\partial}{\partial \lambda_a}, e_a, e_\mu)$. The resulting Pfaffian system is

$$\theta^a = 0$$

$$d\lambda_a + \lambda_b c^b_{\mu a} \omega^\mu = 0$$

$$\lambda_a c^a_{\mu\nu} \omega^\nu = 0.$$

Here the summation convention is in force and in deriving the second equation we used the first equation to get rid of an additional term. We can now easily write down the characteristic equations for a curve ζ. Write $\dot{\zeta} = \dot{\lambda}_a \frac{\partial}{\partial \lambda_a} + \dot{\gamma}^a e_a + \dot{\gamma}^\mu e_\mu$. Then the first set of Pfaffian equations simply say $\dot{\gamma}^a = 0$,i.e. again, that the projection of ζ to Q is tangent to D. The next two sets become

$$\dot{\lambda}_a + \Sigma \lambda_b c^b_{\mu a} \dot{\gamma}^\mu = 0 \tag{10}$$

and

$$\Sigma \lambda_a c^a_{\mu\nu} \dot{\gamma}^\mu = 0 \tag{11}$$

which are the characteristic equations.

The second of these equations says that $\dot{\gamma}$ is in the kernel of the skew symmetric form $w(\lambda)$ on $D_{\gamma(t)}$. It is an algebraic condition which is necessary, but not sufficient, for the existence of a characteristic tangent to the direction $\dot{\gamma}$.

5 Lie Group Examples.

Many of the calculations of this section can be found in Lerman [9] who did them for reasons internal to symplectic geometry. (See also Weinstein [17].)

Let G be a Lie group with left-invariant distribution $D \subset TG \cong G \times \mathcal{G}$. D is defined by choosing a linear subspace of the Lie algebra of G. More specifically,

$$D \cong G \times D_e$$

and its annihilator is

$$D^0 \cong G \times D_e^0.$$

The identifications "≅" are by left translations. e denotes the identity element of G and we identify the Lie algebra $\mathcal{G}$ of G with the tangent space to G at e. $D_e \subset \mathcal{G}$ is a fixed linear subspace of the Lie algebra and $D_e^0 \subset \mathcal{G}^*$ is its annihilator.

Let θ denote the canonical one-form on T^*G and let $\Theta : TG \to \mathcal{G}$ denote the (left) Maurer-Cartan form. (For matrix groups, $\Theta = g^{-1}dg$.) Then

$$\theta(g, \mu) = \mu(\Theta(g))$$

It follows that the canonical two-form, $\omega = d\theta$ satisfies

$$\omega = d\mu \wedge \Theta - \mu([\Theta, \Theta])$$

where we have used Cartan's structure equation $d\Theta = -[\Theta, \Theta]$. Applying this two-form to vectors $(\xi_1, \mu_1), (\xi_2, \mu_2) \in T_(g, \mu)(T^*G) = \mathcal{G} \times \mathcal{G}^*$ we obtain

$$\omega(g, \mu)((\xi_1, \mu_1), (\xi_2, \mu_2)) = \mu_1(\xi_2) - \mu_2(\xi_1) - \mu([\xi_1, \xi_2]).$$

Restricting the form to D^0 means that $\mu, \mu_1, \mu_2 \in D_e^0$.

If (ξ_1, μ_1) represents a characteristic direction then for all $(\xi_2, \mu_2) \in \mathcal{G} \times D_e^0$ the above expression is zero. Setting $\xi_2 = 0$ we find that $-\mu_2(\xi_1) = 0$ for all $\mu_2 \in D_e^0$; i.e.

$$\xi_1 \in D_e. \tag{12}$$

which is in accord with §4, lemma 2. Using this information we now have

$$\omega = \mu_1(\xi_2) - \mu([\xi_1, \xi_2]) = \langle \mu_1 + ad^*_{\xi_1}(\mu), \xi_2 \rangle$$

Or

$$\mu_1 + ad^*_{\xi_1}(\mu) = 0. \tag{13}$$

Equations (12) and (13) describe the kernel of $\omega(g, \mu)$. Replacing (ξ_1, μ_1) with the tangent $(\dot{g}, \dot{\mu}) = (\xi, \dot{\mu})$ to a curve $(g(t), \mu(t))$ in D^0 we find that such a curve is characteristic if and only if

$$\xi(t) \in D_e$$

and

$$\dot{\mu}(t) + ad^*_{\xi(t)}\mu(t) = 0 \tag{14}$$

where

$$g(t)^{-1}\dot{g}(t) = \xi(t).$$

Now suppose that G admits a bi-invariant inner product, for example a nondegenerate Killing form. We will then identify D_e^0 with $D_e^{\perp}$, the orthogonal complement to D_e relative to the inner product. Then ad^* is identified with ad and so equation (14) for the evolution of μ becomes

$$\dot{\mu}(t) + [\xi(t), \mu(t)] = 0.$$

If, moreover, $D_e^\perp$ is a subalgebra, say $\mathcal{H}$, of $\mathcal{G}$ then the two terms in this equation are orthogonal to each other, and hence individually zero. Thus the characteristic equations become

$$\dot{\mu} = 0 \tag{15}$$

$$\xi \in \mathcal{G}_\mu \cap D_e \tag{16}$$

where

$$\mathcal{G}_\mu = \{\xi : [\xi, \mu] = 0\} \tag{17}$$

is the isotropy algebra of μ under the adjoint action. (To see that the two terms are orthogonal, observe that $\dot{\mu} \in \mathcal{H}$, and $[\xi, \mu] \in \mathcal{H}^\perp = D_e$ since if $\mu_2 \in \mathcal{H}$ then

$$\langle [\xi, \mu], \mu_2 \rangle = -\langle \xi, [\mu, \mu_2] \rangle$$

and the last term is zero since $\xi \in D_e$.)

Now suppose that G is a compact connected Lie group and that $\mathcal{H} = \mathcal{T}$ is the Lie algebra of its maximal torus T. (See Adams [1] for example.) For example, if G is the group of all $n \times n$ unitary matrices then T is the subgroup of diagonal ones. We take $D_e = \mathcal{T}^\perp$. Now

$$\mathcal{G} = \mathcal{T} \oplus D_e$$

is an orthogonal decomposition of $\mathcal{G}$ which is invariant under the action of $\mathcal{T}$ by Lie bracket. Upon decomposing D_e into irreducibles under this action we obtain

$$D_e = \oplus_{a \in \Delta_+} V_a.$$

This is called the root space decomposition and the a's are called the positive roots. They are elements of $\mathcal{T}^*$. Each V_a is a two (real) dimensional subspace and admits an orthonormal basis X_a, Y_a such that for each $h \in \mathcal{T}$ we have $[h, X_a] = a(h) Y_a$, $[h, Y_a] = -a(h) X_a$. In other words ad_h acts on V_a by $a(h)J$ where J is rotation by 90 degrees. If a is a root, then so is $-a$. Making this change amounts to reversing the orientation on V_a. Thus, ignoring orientations, $V_a = V_{-a}$. The set of all roots is denoted Δ and is the disjoint union of Δ_+, $-\Delta_+$ and the 0 covector (whose corresponding root space is $\mathcal{T}$). We have the bracket relations

$$[V_a, V_b] \subset V_{a+b} \oplus V_{a-b}$$

and

$$[V_a, V_a] = Span(a)$$

where we use the bi-invariant inner product to identify $a \in Lie(T)^*$ with an element of $Lie(T)$ which we also denote by a. If there is no two-dimensional subspace corresponding to $a \pm b$, i.e. if $a \pm b$ is not a root, then $V_{a \pm b} = 0$. It follows from the second relation that

$$su(2)_a := V_a \oplus Span(a)$$

is a subalgebra of $\mathcal{G}$. It is isomorphic to the Lie algebra $su(2)$ and is called the root-$su(2)$ corresponding to the root a.

For $\mu \in \mathcal{T}$ set

$$\Delta(\mu) = \{a \in \Delta_+ : a(\mu) = 0\}$$

and

$$W_\mu = \oplus_{a \in \Delta(\mu)} V_a$$

It follows directly from the definitions and the bracket relations that

$$\mathcal{G}_\mu = W_\mu \oplus \mathcal{T}$$

(compare with eq. (17)) so that

$$\mathcal{G}_\mu \cap D_e = W_\mu$$

It follows immediately from this, equation (16), and a remark in §4 that

$$ker(w(\mu)) = W_\mu.$$

Combined with equation (15) this shows that $ker(\omega(g, \mu)) = W_\mu \oplus \{0\}$. In particular this means that $(g, \mu) \in \Sigma$ if and only if there is an $a \in \Delta_+$ such that $a(\mu) = 0$. (Σ is the characteristic variety defined at the end of §3.) Now the root hyperplanes are, *by definiton*, the hyperplanes $ker(a) \subset \mathcal{T}$. They form the walls of the Weyl chambers and their union is sometimes called the infinitesimal diagram of G and will be denoted

$$\Sigma_e = \cup_{a \in \Delta_+} \{\mu : a(\mu) = 0\}.$$

It follows that $\Sigma = G \times \Sigma_e$ is the characteristic variety.

Now, let

$$\mathcal{G}^{(\mu)} = \text{Lie algebra generated by } W_\mu.$$

It follows from the bracket relations for the V_a that

$$\mathcal{G}^{(\mu)} = W_\mu \oplus Span(\Delta(\mu)) = \oplus_{a \in \Delta(\mu)} su(2)_a.$$

(It follows that $\mathcal{G}^{(\mu)}$ is contained in the isotropy algebra of μ.) Let $G^{(\mu)}$ denote the closed connected Lie subgroup of G whose Lie algebra is $\mathcal{G}^{(\mu)}$. From what we have just said and the above calculations it follows that every characteristic passing through the fiber over e has the form $(g(t), \mu)$ where $\mu \in \Sigma_e$ is constant and where $g(t)$ is an H^2_1-curve in $G^{(\mu)}$ which is tangent to

$$D^{(\mu)} := D \cap TG^{(\mu)}.$$

We have proved

Theorem 1 *The characteristic variety of D is $G \times \Sigma_e$ where Σ_e is the infinitesimal diagram of G described above. Every characteristic passing through the fiber over e has the form $(g(t), \mu)$ where $\mu \in \Sigma_e$ is constant and $g(t)$ is an H^2_1-curve in $G^{(\mu)}$ which is tangent to*

$$D^{(\mu)} := D \cap TG^{(\mu)}.$$

Conversely, every such curve is a characteristic. Consequently a D-curve $\gamma(t)$ through e is singular if and only if it lies in one of the closed proper subgroups $G^{(\mu)}$ of G.

As a corollary to this theorem we have the following regularity theorem. Recall from §2 that if $\gamma \in \Omega_D(e)$ then a characteristic subgroup for γ is a closed connected subgroup containing γ within which γ is regular.

Theorem 2 *Let G be a compact connected Lie group and D the left-invariant distribution generated by the orthogonal complement to G's maximal torus. Then every singular D-curve is contained in a characteristic subgroup which has dimension less than that of G. These characteristic subgroups are of the form $G^{(\mu)}$ for some $\mu \in \Sigma_e \setminus 0$.*

Proof. By induction on the dimension of G. The first nontrivial ($D \neq TG$) case is dimension 3 which occurs when $G = SU(2)$ or $SO(3)$. The distribution is of rank 2 and of contact type (it is the canonical connection for the Hopf fibration $G \to S^2 = G/T$) thus every nonconstant curve in Ω_D is regular so there is nothing to check.

Now suppose we have proved the statement for all connected compact Lie groups of dimension less than n and let G be such a group of dimension n. Let γ be a singular curve through $e \in G$. By theorem 1, γ lies in some characteristic subgroup G^{μ} of G. Now G^{μ} is compact, connected, and has dimension less than n. Its maximal torus is $Lie(T^{\mu}) = Lie(G^{\mu}) \cap Lie(T)$ and hence $D^{(\mu)} = D \cap TG^{\mu}$ is the distribution generated by $G^{(\mu)}$s maximal torus. By the inductive hypothesis, γ lies in a characteristic subgroup $K = (G^{(\mu)})^{(\nu)}$ of $G^{(\mu)}$. $K \subset G$ is the characteristic subgroup of γ.

Finally, we must show that $K = G^{(\beta)}$ for some $\beta \in \mathcal{T}$. To do this observe that the roots of $G^{(\mu)}$ are $\pm\Delta(\mu)$, restricted to $\mathcal{T}^{\mu}$. We are to think of ν as an element of $\mathcal{T}^{(\mu)*}$. Extend ν to a linear functional $\beta : \mathfrak{T} \to \mathbf{R}$ in such a way that $a(\beta) \neq 0$ for any root a not in $\Delta(\mu)$. One checks from the definitions that $\mathcal{G}^{(\beta)} = (\mathcal{G}^{\mu)})^{(\nu)}$. QED.

Using corollary 1 of §2 we have as an immediate corollary

Theorem 3 *Let G, D be as in the previous theorem. Then, no matter what the inner product on this distribution D, all subRiemannian minimizing geodesics are smooth.*

Example. $G = SU(n)$. The nontrivial G^μ are all of the form $SU(n_1) \times \ldots \times SU(n_j)$ where $n_i \geq 2$ and $n_1 + \ldots n_j \leq n$. The induced distribution on such a G^μ is a product of the distributions $Lie(T_j)^\perp$ on each $SU(n_i)$.($Lie(T_i)$ denotes the diagonal matrices in $Lie(SU(n_i))$.) Now on a product of manifolds with distributions the singular paths $\gamma = (\gamma_1(t), \gamma_2(t), \ldots, \gamma_j(t))$ are those paths for which at least one component γ_i is singular. And if each of the distributions has an inner product so that the product has a subRiemannian structure, then γ is a minimizing geodesic only if each of its components γ_i are minimizing geodesics, as the reader can easily check.

So in this case the last two theorems can be proved by induction on the n of $SU(n)$, beginning with $n = 2$ as before. The case $SU(3)$ is the first interesting one.

5.1 The case of $SU(3)$

Let $Q = SU(3) = G$ be the group of 3×3 unitary matrices with determinant one. Let $T \subset G$ denote its usual maximal torus, the diagonal matrices. Define the left-invariant distribution D on G by left-translating $Lie(T)^\perp \subset Lie(G)$ about G. Here the superscript $\perp$ denotes the orthogonal complement relative to the Killing form $\langle X, Y \rangle = Re(tr(XY^*))$. Thus a typical element of D lying over the identity $e \in G$ is of the form

$$\begin{pmatrix} 0 & -\bar{a} & -\bar{b} \\ a & 0 & -\bar{c} \\ b & c & 0 \end{pmatrix}.$$

Let $\{e_1, e_2, e_3\}$ denote the standard basis for $\mathbf{C}^3$ so that T is diagonal with respect to this basis. Define subgroups $G_i \subset G, i = 1, 2, 3$ by $g \in G_i$ iff $ge_i = e_i$. These are the "root $SU(2)$'s" of G. They are isomorphically embedded copies of $SU(2)$ in G. Under this isomorphism $D_i = TG_i \cap D$ is mapped to the canonical left-invariant rank 2 distribution E on $SU(2)$.

Using the Killing form and left translation, identify D^0 with $G \times Lie(T)$. Corresponding to each $i, i = 1, 2, 3$, define the ith root $\alpha_i \in Lie(T)$ by $\alpha_i(e_i) = -2e_i$, $\alpha_i(e_j) = e_j, j \neq i$. Thus, for example $\alpha_3 = diag(1, 1, -2)$. And define the root hyperplanes to be the three lines $L_i = \mathbf{R}\alpha_i$. Then, according to theorem 1 of the previous subsection the characteristic variety is

$$\Sigma = G \times (L_1 \cup L_2 \cup L_3) \setminus 0.$$

The characteristics passing through a point $(e, \mu) \in \Sigma_e$, with $\mu \in L_i \setminus 0$ are precisely the curves of the form $(\gamma(t), \mu)$ with γ a D_i-curve in G_i. Consequently, a D-curve passing through e is singular relative to D if and only if it is contained in one of the Lie subgroups G_i. In this case it must be a D_i-curve. But for $(SU(2), E)$, every nonconstant E-curve is regular. (As mentioned in the proof of theorem 2, E is of contact type.) The following theorem now follows immediately from Corollary 1, and is a special case of theorem 3.

Theorem 4 *Every minimizer on $(SU(3), D)$ is smooth. This is true regardless of the choice of the inner product D, in particular, it need not be left-invariant.*

REMARKS.

• The set of endpoints of singular curves passing through the identity is

$$G_1 \cup G_2 \cup G_3$$

which is the wedge of three three-spheres, since the pairwise intersection of the G_i's is the base point e. It is curious that the projections of these three S^3 generate the homology of the flag manifold.

• The tangent space to D^0 at (g, μ) can be canonically identified with $Lie(G) \oplus Lie(T)$. If $\mu \in L_i \setminus 0$ as above then the kernel of ω at (g, μ) is $(Lie(G_i) \cap D) \oplus \{0\}$. For example when $\mu = \alpha_3$ the vectors in this have the form $(\xi, 0)$ where the first component ξ has the form

$$\begin{pmatrix} 0 & -\bar{a} & 0 \\ a & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

(a varying over **C**). The subspace of matrices ξ of this form is **not** a Lie subalgebra of $Lie(G)$. At first glance this might seem to contradict the basic result that the kernel of a closed two form is an involutive distribution. Compare this with the warning at the end of §3. But it is not a contradiction, for the result is only true when the kernel has constant rank. The basic result is proved using the formula $d\omega(X, Y, Z) = \omega([X, Y], Z) + \ldots$ and to reach the desired conclusion X , Y, and Z must be vector fields which are defined in an open set and are in the kernel of ω. In our case any such vector field is identically zero since ω is symplectic off of Σ.

The subspace generated by the space of matrices $Lie(G_i) \cap D$ is precisely G_i. An alternative way to construct G_i is to recognize that if ω_i denotes the restriction of ω to $G \times L_i$, then the kernel of ω_i at any nonzero point is $Lie(G_i) \oplus \{0\}$. This has constant rank and so the hypothetical construction of the warning at the end of §3 can be applied. Every leaf of the distribution $ker(\omega_i)$ through $\{e\} \times L_i$ projects onto G_i.

References

[1] J. F. Adams, **Lectures on Lie Groups**, U. of Chicago Press, 1969.

[2] A. A. Agrachev, and R. V. Gamkrelidze, *The exponential representation of flows and chronological calculus*, Matem. Sbornik, **109** (149) 467-532, (1978).

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

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

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

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

[7] U. Hamenstädt, *Left Invariant Bracket Generating Distributions on Lie Groups* handwritten notes, (1991).

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

[9] E. Lerman, *How Fat is a Fat Bundle?* Lett. Math. Phys., v. 15, 335-339 (1988).

[10] W-S. Liu and H.J. Sussmann, *Abnormal subRiemannian minimizers*, IMA preprint (1993).

[11] J. Martinet, *Sur les Singularites Des Formes Differentialles*, Ann. Inst. Fourier, v. 20, no. 1, 95-198, (1970).

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

[13] 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).

[14] E.D. Sontag, **Mathematical Control Theory** Springer-Verlag, 1990.

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

[16] H.J. Sussmann, *A weak regularity theorem for real analytic optimal control problems*, Revista Matematica Iberoamericana, Vol. 2, No. 3, pp. 307-317,(1986).

[17] A. Weinstein, *Fat Bundles and Symplectic Manifolds*, Adv. in Math, v. 37 , 239-250, (1980).

[18] , M. Ya. Zhitomirskii, **Typical Singularities of Differential 1-Forms and Pfaffian Equations** Translations of Math. Monographs series, vol. 113, AMS, (1991).