

## SOME ASPECTS OF DIFFERENTIAL GEOMETRY ASSOCIATED WITH HYPOELLIPTIC SECOND ORDER OPERATORS

THOMAS J. S. TAYLOR

**I investigate some aspects of the geometry of the bicharacteristics of a class of hypoelliptic second order partial differential operators. The resulting geometry looks quite a bit like Riemannian geometry, although with interesting differences.**

**1. Introduction.** In this paper, we investigate the relationship between certain problems in the calculus of variations and the geometry of the bicharacteristics of a second order hypoelliptic operator.

Let  $M$  denote a connected  $C^\infty$  manifold of dimension  $m$ . Let  $\Delta$  denote a second order hypoelliptic partial differential operator on  $M$ . We assume that the set of second order zeros of the principal symbol of  $\Delta$  is a smooth submanifold of  $T^*M$ . Since  $\Delta$  is second order it is always possible, at least locally, to find a function  $V(x)$  and vector fields  $\{g_i\}_{i=0}^n$  such that  $\Delta = \sum_{i=1}^n c_i g_i^2 + g_0 + V(x)$  for some constants  $c_i = \pm 1$ . Hörmander [19] has shown that a sufficient condition that  $\Delta$  be hypoelliptic is that the sign of the  $c_i$  doesn't change and that the evaluation map on vector fields is at each point  $x \in M$  onto  $T_x M$  when restricted to the Lie algebra of vector fields (over  $\mathbf{R}$ ) generated by the vector fields  $\{g_i\}_{i=0}^n$ . In this paper I will be interested in hypoelliptic operators which satisfy the stronger condition that the Lie algebra generated by  $\{g_i\}_{i=1}^n$  is onto  $TM$ .

Now, because  $\Delta$  is of second order,  $\Delta$  defines a quadratic form  $G^*$  on  $T^*M$ : if  $f_1, f_2 \in C^\infty(M)$  and  $f_1(x) = f_2(x) = 0$ , then

$$G_x^*(df_1(x), df_2(x)) = [\Delta(f_1 f_2)]|_x.$$

Because  $\Delta$  is hypoelliptic,  $G^*$  is a nonnegative quadratic form, and, in fact, locally  $G^* = \sum_{i=1}^n g_i \otimes g_i$ . Note that when  $n < m$ ,  $G^*$  is degenerate, i.e.,  $\ker G^* \neq \{0\}$ .

Let  $D \subset TM$  be the distribution spanned locally by the vector fields  $\{g_i\}_{i=1}^n$ . We assume that  $D$  is of constant rank, or equivalently, that  $\ker G^*$  is of constant dimension.

It has been pointed out by Brockett [3] that the quadratic form  $G^*$  endows  $M$  with a structure that is in many ways similar to Riemannian structures, so much so that he uses the phrase “singular Riemannian geometry”. In fact, the singular Riemannian dual metric tensor  $G^*$  and various properties of the singular Riemannian geometry can be obtained as the limits of properties of a parameterized family of dual Riemannian metric tensors,  $G_\varepsilon^*$ , as  $\varepsilon \rightarrow 0$ ; see Gunther [10]. The associated Riemann metrics  $G_\varepsilon$  have some components which go to  $\infty$  as  $\varepsilon \rightarrow 0$ , hence the adjective “singular”. The term “singular Riemannian geometry” has also been applied to Riemannian manifolds with conical singularities, but for lack of a better expression, I will continue to use the terminology of Brockett.

In this paper I will discuss some previously known properties of the singular Riemannian geodesic flow (or equivalently, the bicharacteristic flow of the operator  $\Delta$ ) as well as prove some new results. Now, Gaveau [8] has discussed this flow. Certain errors of [8] are corrected in [3]. Nagel, Stein and Wainger [16], and Fefferman and Phong [7] have discussed properties of pseudometrics which are related to the distance function associated to a singular Riemannian structure. Gromov [9] has discussed some of the metric properties of singular Riemannian geometry; he calls the associated distance function a “Carnot-Carathéodory metric”. Mitchell [18] discusses Hausdorff dimension properties of Carnot-Carathéodory metrics. Strichartz [19] has discussed in some detail properties of the exponential map.

In §2, I develop some aspects of singular Riemannian geometry, both intrinsic and in relation to Riemannian geometry, and I also develop some connections between the geodesic problem for singular Riemannian geometry and certain techniques and problems of optimal control theory. In §3, I discuss results concerning certain “complete” singular Riemannian exponential maps, in particular that every two points are connected by a minimizing geodesic and that geodesics do not minimize past the first conjugate point. Some of these results are due to Gromov [9] and other authors and some are original. I use these results in §4 to investigate singular Riemannian geodesic coordinates connected with a point  $x \in M$ ; these coordinates may be defined on open geodesic cones with  $x$  as the base point.

The work in this paper was in part motivated by a desire to apply the techniques of Kannai [14] to the study of the asymptotics of hypoelliptic diffusion equations. These applications will appear in a sequel.

**2. Background.** Assume that our second order operator  $\Delta$  is locally of the form  $\Delta = \sum_{i=1}^n g_i^2 + \text{lower order}$ , where  $\{g_i\}_{i=1}^n$  are vector fields on  $M$ , which are nonvanishing. We assume also that the Lie algebra generated by  $\{g_i\}_{i=1}^n$  has the property that it spans the tangent space of  $M$ , so that  $\Delta$  is hypoelliptic. Then, by our earlier assumptions, the distribution  $D$  spanned by  $\{g_i\}_{i=1}^n$  is of constant rank, and satisfies the property that the smallest integrable subbundle of  $TM$  which contains  $D$  is  $TM$  itself.

Now consider the quadratic form  $G^*$  induced by  $\Delta$  on  $T^*M$ . Because  $G^* = \sum_{i=1}^n g_i \otimes g_i$ , the kernel of  $G^*$  is  $D^\perp$ , (= the annihilator of  $D$  in  $T^*M$ ). Thus, although  $G^*$  is degenerate,  $G^*$  factors to a positive definite form  $\tilde{G}^*$  on  $T^*M/D^\perp$ . Also,  $G^*$  determines a mapping  $\Gamma: T^*M \rightarrow TM$ , an element  $\xi \in T_x^*M$  is mapped into the element

$$\Gamma(\xi) = G^*(\xi, \circ) = \sum_{i=1}^n \xi(g_i(x))g_i(x) \in T_xM.$$

Clearly the image of  $\Gamma$  is  $D$  and the kernel of  $\Gamma$  is  $D^\perp$ . Thus,  $\Gamma$  factors to an isomorphism  $\tilde{\Gamma}$  from  $T^*M/D^\perp$  to  $D$ , indeed for  $\tilde{\xi} \in T^*M/D^\perp$ ,  $\tilde{\Gamma}(\tilde{\xi}) = \tilde{G}^*(\tilde{\xi}, \circ)$ . (Note that  $T^*M/D^\perp$  is  $D^*$ , the dual space of  $D$ .)

$G^*$  also induces a positive definite quadratic form on  $D$ , call it  $G$ , defined for  $v, w \in D$  by  $G(v, w) = \tilde{G}^*(\tilde{\Gamma}^{-1}v, \tilde{\Gamma}^{-1}w)$ . One may verify that the vector fields  $\{g_i\}_{i=1}^n$  form an orthonormal frame for  $D$  with respect to  $G$ . Now, suppose that  $c: [0, 1] \rightarrow M$  is a smooth curve in  $M$  such that  $\dot{c}(t) \in D$  for  $t \in [0, 1]$ . We may, in analogy with Riemannian geometry, assign an arc length to this curve by the formula  $\text{length}(c) = \int_0^1 \sqrt{G(\dot{c}, \dot{c})} dt$ .

At this juncture, it is fruitful to apply a controllability result to learn that, under the conditions we have imposed on the distribution  $D$ , given any two points  $x, y \in M$  there exists a smooth curve  $c: [0, 1] \rightarrow M$  such that  $c(0) = x, c(1) = y$  and  $\dot{c}(t) \in D$  for all  $t$ . Indeed, because the Lie algebra generated by  $\{g_i\}_{i=1}^n$  spans  $TM$ , Chow's Theorem (as discussed in Ballieul [1], for example) gives us the existence of a curve of the required type.

Therefore, for every two points  $x, y \in M$ , we can define the distance,  $r(x, y)$ , between  $x$  and  $y$  by

$$r(x, y) = \inf \int_0^1 \sqrt{G(\dot{c}, \dot{c})} dt,$$

where the infimum is to be taken over all absolutely continuous curves connecting  $x$  and  $y$  with  $\dot{c}(t) \in D$  a.e. with respect to  $t$ .

We call  $r(x, y)$  “the singular Riemannian distance” between  $x$  and  $y$ , for it is easy to see that  $r(x, y) = r(y, x)$  and  $r(x, y) = 0 \Rightarrow x = y$ , while the triangle inequality for  $r(x, y)$  follows from the fact that  $r(x, y)$  is independent of parameterization of the curves  $c(t)$ : for  $c_i(t): [0, 1] \rightarrow M$  for  $i = 1, 2$  with  $\dot{c}_i(t) \in D$  a.e. such that  $c_1(0) = x$ ,  $c_1(1) = y = c_2(0)$ ,  $c_2(1) = z$  we can define

$$c(t) = \begin{cases} c_1(2t), & t \in [0, 1/2], \\ c_2(2t - 1), & t \in [1/2, 1], \end{cases}$$

and  $c(t)$  will have a length with respect to  $G$  which is no smaller than the infimum of the lengths of all curves tangent to  $D$  which go from  $x$  to  $y$ . Thus  $r: M \times M \rightarrow \mathbf{R}^+$  is a globally defined distance function.

Let  $I = [A, B]$  be a closed interval. Define a locally minimizing singular Riemannian geodesic to be a curve  $c: I \rightarrow M$  such that for  $a, b \in I$  sufficiently close  $r(c(a), c(b)) = \text{length}(c([a, b]))$ , and define a minimizing geodesic between  $x = c(A)$  and  $y = c(B)$  to be a geodesic such that  $r(x, y) = \text{length}(c([A, B]))$ .

Assume, for the moment, that the distribution  $D$  is trivial, so that there are globally defined vector fields  $g_a$  such that  $G^* = \sum g_a \otimes g_a$ . Also, assume that the  $\mathbf{R}$ -linear span of  $\{g_a\}_{a=1}^n$  consists of complete vector fields. Given a curve  $c: [0, 1] \rightarrow M$  such that  $\dot{c} \in D$  for all  $t$  we can define an  $n$ -tuple of functions  $u_a: [0, 1] \rightarrow \mathbf{R}$  by  $u_a(t) = G(g_a, \dot{c}(t))$ . Then,  $c(t)$  is a solution of the differential equation

$$(1) \quad \dot{c}(t) = \sum_{a=1}^n g_a(c(t))u_a(t).$$

A family of such differential equations, parameterized by the functions  $\{u_1, \dots, u_n\}$  is called a control system; the  $n$ -tuple  $u$  is called a control. It follows from the above discussion that the control system (1) has the property that for any  $x, y \in M$  there is a control  $u = (u_1, \dots, u_n)$  such that the solution  $c$  has the property that  $c(0) = x$ ,  $c(1) = y$ . A control system with this property is said to be “controllable”, and the control  $u$  is said to “steer”  $x$  to  $y$ .

Now the control system (1) may be solved for any control  $u$  with the property that  $u_1, \dots, u_n$  are each in  $L^1[0, 1]$ . If we make  $\mathbf{R}^n = \mathbf{R} \times \mathbf{R} \times \dots \times \mathbf{R}$  ( $n$  times) into an inner product space with the standard quadratic form  $\langle \cdot, \cdot \rangle = (1$ 's on the diagonal) then the set of controls  $u$  with this property is  $L^1([0, 1]; \mathbf{R}^n) = \{u: \int_0^1 \sqrt{\langle u, u \rangle} dt < \infty\}$ . The completeness conditions on  $\{g_i\}_{i=1}^n$  imply that control system (1) defines a map  $h: L^1([0, 1]; \mathbf{R}^n) \times M \rightarrow M$ ; if  $c(t, u, x_0)$  is the solution

of (1) with initial condition  $c(0) = x_0$  for a given  $u \in L^1([0, 1]; \mathbf{R}^n)$ , we define  $h(u, x_0) = c(1, u, x_0)$ .

The following lemma is a simple consequence of Gronwall's lemma.

LEMMA 1. *The mapping  $h(\circ, x_0) : L^1([0, 1]; \mathbf{R}^n) \rightarrow M$  is continuous.*

Now, since  $\{g_a\}_{a=1}^n$  is an orthonormal frame for  $D$ ,

$$\begin{aligned} G(\dot{c}, \dot{c}) &= G\left(\sum_{a=1}^n g_a u_a(t), \sum_{b=1}^n g_b u_b(t)\right) \\ &= \sum_{a,b=1}^n u_a(t)u_b(t)G(g_a, g_b) \\ &= \sum_{a,b=1}^n u_a u_b \delta_{ab} = \sum_{a=1}^n u_a(t)^2. \end{aligned}$$

Thus,

$$r(x, y) = \inf \int_0^1 \sqrt{\sum u_a(t)^2} dt = \inf \|u\|_{L^1}.$$

where the infimum is taken over all  $u$  which steer  $x$  to  $y$ .

We then have:

LEMMA 2. *Let  $r$  be the singular Riemannian distance function. Then  $r$  is a continuous map  $M \times M \rightarrow \mathbf{R}$ .*

*Proof.* Suppose that  $r$  is not continuous. Then, there exists  $z, y \in M$ ,  $\varepsilon > 0$  and a sequence of points  $(z_k, y_k) \in M \times M$  which converges to  $(z, y)$  and such that  $|r(z, y) - r(z_k, y_k)| > \varepsilon$ . Then, the triangle inequality implies that  $r(z, z_k) + r(y, y_k) > \varepsilon$ . By Lemma 1 the solution  $\gamma(t, y)$  of the control system (1) is a continuous mapping from  $L^1([0, 1]; \mathbf{R}^n)$  to  $M$ . Since (1) is controllable for every sequence  $y_k$  (or  $z_k$ ) there is a sequence of  $u_k$  in  $L^1$  such that  $y_k = h(u_k, y)$ . Now let  $\Omega_N \subset L^1$  be the subspace of all functions such that  $u$  is constant on each open interval

$$\left(\frac{j-1}{N}T, \frac{j}{N}T\right), \quad \text{for } j = 1, 2, \dots, N.$$

The dimension of  $\Omega_N$  is  $nN$ , so that  $\Omega_N$  may be identified with  $\mathbf{R}^{nN}$ . It is then a consequence of the definition of  $h(0, y)$  that  $h(0, y)|_{\Omega_N}$  is a  $C^\infty$  mapping of  $\Omega_N$  into  $M$ . But, it is established in the literature of control theory (see [4], the proof of theorem one, for example),

that for  $N$  sufficiently large  $h(t, y)$  maps neighborhoods of zero in  $\Omega_N$  onto neighborhoods of  $y$  in  $M$ . Thus,  $h(0, y)$  maps neighborhoods of zero in  $L^1$  into neighborhoods of  $y$  in  $M$ . But this implies that the  $u_k$  may be chosen in such a way as to approach zero as  $k \rightarrow \infty$ . But the length of the path  $s \rightarrow c(s, y, u_k)$  for  $s \in [0, t]$  is just the  $L^1$  norm of  $u_k$ , while we have assumed that the optimal distance of the  $y_k$  (or  $z_k$ ) is bounded below. This is a contradiction. Thus  $r$  is jointly continuous.  $\square$

An easy localization argument gives:

**COROLLARY 1.** *A singular Riemannian distance  $r(x, y)$  is continuous.*

**3. The singular Riemannian exponential map.** In Riemannian geometry, there are several different but equivalent notions of a complete Riemannian manifold,  $M$ , for example, that  $M$  is a complete metric space with respect to the Riemannian distance function or that the geodesic flow is complete. For a complete Riemannian manifold, every two points can be joined by a minimizing geodesic. In the singular Riemannian case, the issue is not so clear. We have, however, the following which is a particular case of a theorem of Gromov [9].

**THEOREM 1.** *If the metric space  $(M, r)$ , with  $r$  the singular Riemannian metric, is complete then:*

- (a) *the closed geodesic balls are compact.*
- (b) *any two points  $x, y \in M$  may be joined by a minimizing geodesic.*

Suppose that  $\hat{G}$  is a Riemannian metric on  $M$ . One method of obtaining a singular Riemannian metric  $G$  is simply to restrict  $\hat{G}$  to the distribution  $D$ . The following theorem follows from a statement in Gromov's book.

**THEOREM 2.** *Suppose that  $G$  is the restriction to  $D$  of a complete Riemannian metric  $\hat{G}$ , and that  $r$  is the singular Riemannian distance associated to  $G$ . Then  $(M, r)$  is complete.*

*Proof (outline).* Suppose  $\hat{r}$  is the Riemannian distance. Then  $r \geq \hat{r}$  because the length of a curve  $c$  which is tangent to  $D$  is the same with respect to  $\hat{G}$  as to  $G$ , but for  $r$  one can minimize only over such curves,

while for  $\hat{r}$  one can minimize over all curves. Thus, any sequence Cauchy for  $r$  is also Cauchy for  $\hat{r}$ . Conversely, suppose  $\{x_n\} \subset M$  is Cauchy for  $\hat{r}$ , then there is an  $x \in M$  such that  $x_n$  converges to  $x$ . But, since  $r$  is continuous  $r(x_n, x)$  converges to zero, and  $r(x_n, x_{n+m}) \leq r(x_n, x) + r(x_{n+m}, x)$ .  $\square$

On the other hand, the converse of this theorem, that every complete singular Riemannian metric is the restriction of a complete Riemannian metric is not so clear (to the author). The next theorem gives sufficient conditions, of a different type than Gromov's, for two arbitrary points to be connected by a minimizing singular Riemannian geodesic.

**THEOREM 3.** *Assume that  $G^*$  is globally of the form  $\sum_{a=1}^n g_a \otimes g_a$  with  $\{g_a\}_{a=1}^n$  globally defined vector fields with  $\mathbf{R}$ -span consisting of complete vector fields. Then we have that every two points are connected by a minimizing singular Riemannian geodesic.*

*Proof.* Let  $c: [0, 1] \rightarrow M$  be a smooth curve in  $M$  which is tangent to  $D$  for all  $t$  in  $[0, 1]$ , so that  $c$  is a solution of the differential equation

$$(1) \quad \dot{c}(t) = \sum_{a=1}^n g_a(c(t))u_a(t),$$

and conversely, any solution  $c$  of such a differential equation is a curve in  $M$  which is tangent to  $D$ . Now, exactly as in the Riemannian case, the singular Riemannian geodesic problem may be phrased: Find a curve  $c: [0, 1] \rightarrow M$  which is tangent to  $D$  such that  $c(0) = x$ ,  $c(1) = y$  and which minimizes the functional

$$l^2(c) = \int_0^1 G(\dot{c}, \dot{c}) dt.$$

Such an extremal  $c$  will also minimize the length  $\int_0^1 \sqrt{G(\dot{c}, \dot{c})} dt$  and will have the property that  $G(\dot{c}, \dot{c})$  is independent of  $t$  and that  $l^2(c) = r(x, y)^2$ .

Now, recall that the vector fields  $\{g_a\}_{a=1}^n$  are an orthonormal frame for  $D$ , so that  $G(g_a, g_b) = \delta_{ab}$ . If we substitute the differential equation (1) in the functional  $l^2(c)$  we have that  $l^2(c)$  is equal to the functional of  $u$

$$\eta(u) = \int_0^1 \sum_{a=1}^n u_a(t)^2 dt.$$

Thus, the singular Riemannian geodesic problem may be rephrased: Find an  $n$ -tuple of functions  $u(t)$  such that the solution  $c(t)$  of the initial value problem

$$(2) \quad \dot{c}(t) = \sum_{a=1}^n g_a(c(t))u_a(t); \quad c(0) = x$$

satisfies the condition  $c(1) = y$ , i.e.,  $u$  steers the differential equation (1) from  $x$  to  $y$  and such that, among all  $n$ -tuples which steer  $x$  to  $y$ ,  $u$  minimizes  $\eta(u)$ .

This way of considering the matter is the archetype of the (nonlinear) “optimal control problem” so that the deep results of Cesari [6] concerning the existence of optimal controls may be applied.

Now, all the results of Cesari’s paper are in the context of nonlinear optimal control systems on Euclidean spaces, so we will have to modify our situation. Consider  $M$  to be an embedded submanifold of  $\mathbf{R}^{2m+1}$ , by the Whitney embedding theorem. Since  $M$  has a tubular neighborhood in  $\mathbf{R}^{2m+1}$ , it’s easy to see that we may extend the vector fields  $g_a$  to smooth vector fields  $\tilde{g}_a$  on  $\mathbf{R}^{2m+1}$ , in such a way that the extended vector fields are all tangent to  $M$ . Thus, any solution of the differential equation

$$(3) \quad \dot{c}(t) = \sum_{a=1}^n \tilde{g}_a u_a(t), \quad c(0) = x \in M$$

satisfies the property that  $c(t) \in M$  for all  $t$ . Now, let  $x, y \in M$ , and suppose that  $\psi_\varepsilon \in C_0^\infty(\mathbf{R}^{2m+1})$  satisfies the property that  $0 \leq \psi_\varepsilon \leq 1$  and that  $\psi_\varepsilon = 1$  on some open neighborhood in  $\mathbf{R}^{2m+1}$  of  $E_\varepsilon = \{z: r(x, z) < r(x, y) + \varepsilon\} \subset M$ . Then, if  $u = (u_1, \dots, u_n)$  minimizes  $\eta(u)$  among all  $u$  which steer the solutions  $c_\varepsilon$  of the differential equation

$$(4) \quad \dot{c}_\varepsilon(t) = \sum_{a=1}^n \psi_\varepsilon(c_\varepsilon(t)) \tilde{g}_a(c_\varepsilon(t)) u_a(t)$$

from  $x$  to  $y$ , then  $u$  will also steer the solutions of equation (3) from  $x$  to  $y$  and will minimize  $\eta(u)$  among all such  $u$ .

Indeed, we can see this as follows. Since  $r(x, y) = \inf \int_0^1 \sqrt{G(\dot{c}, \dot{c})} dt$  over all curves  $c$  tangent to  $D$  such that  $c(0) = x$ ,  $c(1) = y$ , we have that for every  $\delta > 0$  there is a curve  $c^\delta$  tangent to  $D$  and connecting  $x$  to  $y$  which has a length  $< r(x, y) + \delta$ , and which therefore is contained in  $E_\varepsilon$  for all  $t$  if  $\delta < \varepsilon$ .

I claim that this means that a solution  $c_\epsilon$  of (4) which connects  $x$  to  $y$  and which minimizes  $\eta(u)$  has the property that  $r(x, c_\epsilon(t)) < r(x, y)$  for all  $t < 1$ . Suppose the contrary. Then there exists a  $t_0 < 1$  such that  $r(x, c_\epsilon(t_0)) = r(x, y)$  and  $r(x, c_\epsilon(t)) < r(x, y)$  for  $t < t_0$ , because  $r$  is continuous. Then

$$\int_0^1 \sum_{a=1}^n u_a(t)^2 dt = \int_0^{t_0} \sum_{a=1}^n u_a(t)^2 dt + \int_{t_0}^1 \sum_{a=1}^n u_a(t)^2 dt.$$

But, for  $t < t_0$ ,  $\psi_\epsilon(c_\epsilon(t)) = 1$ , so that for  $t < t_0$ ,  $c_\epsilon(t)$  also solves (3) and is minimizing for  $t < t_0$ , so that

$$\int_0^{t_0} \sum_{a=1}^n u_a(t)^2 dt = r(x, y)^2 / t_0.$$

However, the fact that  $u$  minimizes  $\eta(u)$  among those  $u$  which steer (4) from  $x$  to  $y$  implies that the corresponding solution  $c_\epsilon(t)$  of (4) is an extremal of the functional

$$\int_0^1 \psi_\epsilon^{-2}(c_\epsilon(t)) G(\dot{c}_\epsilon, \dot{c}_\epsilon) dt = \int_0^1 \sum_{a=1}^n u_a(t)^2 dt,$$

so that the standard arguments give us that  $\psi_\epsilon^{-2} G(\dot{c}_\epsilon, \dot{c}_\epsilon)$  is independent of time, hence that  $\sum u_a(t)^2$  is independent of time, hence that

$$\eta(u) = r(x, y)^2 / t_0 + ([1 - t_0] / t_0) r(x, y)^2 = (1 / t_0^2) r(x, y)^2 > r(x, y)^2.$$

Also

$$\int_0^1 \psi_\epsilon^{-1}(c_\epsilon(t)) \sqrt{G(\dot{c}_\epsilon, \dot{c}_\epsilon)} dt = (1 / t_0) r(x, y) > r(x, y).$$

But, if we define  $u_\delta(t)$  by  $u_{a\delta}(t) = G(g_a, \dot{c}^\delta(t))$ , then, for  $\delta$  sufficiently small  $u_\delta(t)$ , steers both (3) and (4) from  $x$  to  $y$ ; the solution in both cases is  $c^\delta$  because  $c^\delta(t)$  is contained in  $E_\epsilon$  for all  $t$  in  $[0, 1]$ . We have, therefore that

$$\begin{aligned} & \int_0^1 \psi_\epsilon^{-1}(c^\delta(t)) \sqrt{G(\dot{c}^\delta, \dot{c}^\delta)} dt \\ &= \int_0^1 \sqrt{G(\dot{c}^\delta, \dot{c}^\delta)} dt < r(x, y) + \delta < (1 / t_0) r(x, y), \end{aligned}$$

for  $\delta$  sufficiently small. But, this contradicts the fact that  $c_\epsilon(t)$  is minimizing, which proves my claim.

Thus,  $c_\epsilon(t)$  is minimizing implies that  $r(x, c_\epsilon(t)) < r(x, y)$ , which implies that  $u(t)$  steers  $x$  to  $y$  for both (3) and (4) and that  $\eta(u)$  is

the same in both cases. Therefore, in order to show the existence of a  $u$  which steers (3) from  $x$  to  $y$  it is sufficient to show the existence of a  $u$  which does the job for (4). But, (4) satisfies the conditions of Existence Theorem III of Cesari [6] if, instead of  $\eta(u)$ , we consider the problem of minimizing the equivalent functional

$$\eta(u) + 1 = \int_0^1 \left[ \sum_{a=1}^n u_a(t)^2 + 1 \right] dt.$$

Thus, the theorem of Cesari implies that there exists a  $u$  which steers (4) from  $x$  to  $y$  and which minimizes  $\eta(u) + 1$  among all such  $u$  (here, we have that Cesari's  $g = g_0 = \mu = 1$ ,  $g_i = 0$  for  $i = 1, 2, \dots, m$ , and  $C = \sup \sum_{a,i} |\tilde{g}_a^i|$  where  $\tilde{g}_a^i$  is the  $i$ th component of the vector field  $\tilde{g}_a$  with respect to the  $i$ th coordinate of some choice of linear coordinate on  $\mathbf{R}^{2m+1}$ ).  $\square$

**REMARK.** For various reasons, it seems obvious that the singular Riemannian structures of the last theorem must give rise to a complete metric on  $M$ , although I haven't yet been able to prove it. In this eventuality, the last theorem will follow from Theorem 1.

Now, recall that a necessary condition that a control  $u$  be an optimal control is that it satisfy the Pontryagin maximum principle. But, for the Pontryagin maximum principle to be satisfied is equivalent to the statement that a geodesic is the projection onto  $M$  of a solution in local coordinates of Hamilton's equations on  $T^*M$ , with Hamiltonian function  $G^*(p, p)$ . On the other hand, simple localization arguments lend themselves nicely to give

**COROLLARY 2.** *For an arbitrary singular Riemannian metric, a geodesic is a projection of an orbit in  $T^*M$  of the Hamiltonian vector field associated to the function  $G^*(p, p)$ .*

In Riemannian geometry, there is a converse to this corollary, that the projection of the geodesic flow is locally minimizing. In our case, the situation is not so clear.

Now, let  $H$  be the Hamiltonian vector field on  $T^*M$  which is associated with the Hamiltonian function  $G^*(p, p)$ . Recall that the exponential map  $\exp_x: T_x^*M \rightarrow M$  is defined by  $\exp_x(p) = \pi \exp(H)(x, p)$  for  $p \in T_x^*M$ , where  $\pi$  is the canonical projection of  $T^*M$  on  $M$ . We have the following corollary.

**COROLLARY 3.** *Assume the conditions of one of Theorems 1, 2, or 3. The exponential map  $\exp_x: T_x^*M \rightarrow M$  is onto.*

*Proof.* Every two points are connected by a minimizing geodesic, hence every two points are connected by a geodesic. But, a geodesic is the projection of an integral curve of  $H$ , so that there is for every  $x, y \in M$  an integral curve  $\xi: [0, 1] \rightarrow T^*M$  such that  $\pi\xi(0) = x$  and  $\pi\xi(1) = y$ . Then, if  $\xi(0) = (x, p)$  we have  $\exp_x p = \pi \exp(H)(x, p) = \pi\xi(1) = y$ . □

Let  $S_x$  be the “unit sphere” with respect to  $G_x^*$  in the cotangent space  $T_x^*M$ , i.e.,  $S_x = \{P = G_x^*(P, P) = 1\}$  (of course,  $S_x$  is actually a cylinder).

**DEFINITION.** Let  $\mu_x: S_x \rightarrow \mathbf{R}^+$  be the function on  $S_x$  which assigns to each  $p \in S_x$  the largest non-negative extended real number  $T \in [0, \infty]$  such that the geodesic  $t \rightarrow \exp_x(tp)$  is minimizing for all  $t \in [0, T)$ .

The properties of the analogous function in Riemannian geometry are well known. Now, for any  $p \in T_x^*M$ , the length of the geodesic  $t \rightarrow \exp_x(tp)$  between  $t = 0$  and  $t = 1$  is  $\sqrt{G_x^*(p, p)}$ . Thus, for  $t \leq \mu_x(p)$ ,  $p \in S_x$ , we have that  $r(x, \exp_x(tp)) = t$ , and, for all  $t \geq 0$ ,  $r(x, \exp_x(tp)) \leq t$ . Thus, for each  $t > 0$ ,  $\exp_x(t, \cdot)$  maps  $S_x$  into  $B_x(t) = \{y: r(x, y) < t\}$ . For  $r$  complete,  $B_x(t)$  is compact, by Theorem 1.

We use this compactness to develop a method of considering the uniform limits of geodesics  $t \rightarrow \exp_x(tp_n)$ ,  $t \in [0, a]$  for some  $a > 0$ , and  $p_n$  a sequence in  $S_x$ . We have that, for each  $t$ ,  $\exp_x(t, \cdot)$  maps  $S_x$  into  $B_x(t)$ ;  $S_x$  is a completely regular topological space,  $B_x(t)$  is compact. It follows that  $\exp_x(t, \cdot)$  extends to a continuous mapping  $\widetilde{\exp}_x(t, \cdot)$  of the Stone-Ćech compactification  $\beta(S_x)$  into  $B_x(t)$ . We have that  $\widetilde{\exp}_x(t, p) = \exp_x(tp)$  for  $p \in S_x$ . Consider now the function  $t \rightarrow \widetilde{\exp}_x(t, p)$ , for  $p \in \beta(S_x)$ . Since  $S_x$  is open and dense in  $\beta(S_x)$ , there always exists a sequence  $\{p_n\} \subset S_x$  such that any neighborhood of  $p$  contains all but a finite number of the  $p_n$ 's. By continuity,  $\exp_x(tp_n)$  converges to  $\widetilde{\exp}_x(t, p)$  for each fixed  $t$ . Thus,

$$r(\widetilde{\exp}_x(t, p), \widetilde{\exp}_x(t + \varepsilon, p)) = \lim_{n \rightarrow \infty} r(\exp_x(tp_n), \exp_x([t + \varepsilon]p_n))$$

by continuity of  $r$ , so that  $r(\widetilde{\exp}_x(t, p), \widetilde{\exp}_x(t + \varepsilon, p)) \leq \varepsilon$ . Thus, the function  $t \rightarrow \widetilde{\exp}_x(t, p)$  is a continuous curve in  $M$  which is uniformly continuous with respect to  $r$ .

DEFINITION. In the case that  $p_x \in \beta(S_x) - S$  we call  $t \rightarrow \widetilde{\text{exp}}_x(t, p)$  a generalized geodesic.

Note that  $\text{exp}_x(tp_n)$  converges pointwise in  $t$  to  $\widetilde{\text{exp}}_x(t, p)$ ; it's an easy exercise to show that it converges uniformly for  $t$  in compacts by using the uniform continuity and an  $\varepsilon/3$  argument.

If there is a generalized geodesic such that  $r(x, \widetilde{\text{exp}}_x(t, p)) = t$  for  $t \in [0, T]$ , we can take this as the definition of a minimizing generalized geodesic.

REMARK.  $\widetilde{\text{exp}}_x$  defines a continuous map  $\beta(S_x) \times [0, T] \rightarrow M$ , but it need not be the case that  $p_1, p_2 \in \beta(S_x)$ ,  $p_1 \neq p_2$ , implies that  $\widetilde{\text{exp}}_x(t, p_1)$ ,  $\widetilde{\text{exp}}_x(t, p_2)$  are different curves: a generalized geodesic may be equal to a geodesic. Indeed, we have:

LEMMA 3. *The only minimizing generalized geodesics are equal to geodesics.*

The proof of this lemma requires the introduction of certain concepts. Suppose that  $t \rightarrow \widetilde{\text{exp}}_x(t, p)$  is a minimizing generalized geodesic for  $t \in [0, T]$ . A repetition of an argument given above implies that

$$r(\widetilde{\text{exp}}_x(t, p), \widetilde{\text{exp}}_x(s, p)) \leq |t - s| \quad \text{for all } t, s > 0.$$

For  $0 \leq s < t \leq T$  we have

$$\begin{aligned} t = r(x, \widetilde{\text{exp}}_x(t, p)) &\leq r(x, \widetilde{\text{exp}}_x(s, p)) \\ &\quad + r(\widetilde{\text{exp}}_x(s, p), \widetilde{\text{exp}}_x(t, p)); \end{aligned}$$

the first term is equal  $s$ , the second less than or equal  $t - s$ , so that  $r(\widetilde{\text{exp}}_x(s, p), \widetilde{\text{exp}}_x(t, p)) = t - s$  for a minimizing geodesic. It follows that the curve  $c(t) = \widetilde{\text{exp}}_x(t, p)$  satisfies

$$t = \sum_{m=0}^{n-1} r\left(c\left(\frac{mt}{n}\right), c\left(\frac{m+1}{n}t\right)\right) \quad \text{for all } n > 0.$$

This motivates the following. Let  $c: [a, b] \rightarrow M$  be an arbitrary absolutely continuous curve. Let  $\mathcal{P} = \{a = t_0 < t_1 < \cdots < t_{n+1} = b\}$  denote a partition of  $[a, b]$ . Define the length functional

$$\tilde{l}^{(a,b)}(c) = \sup_{\mathcal{P}} \sum_i r(c(t_i), c(t_{i+1})).$$

In general, this length might not be the same as the length with respect to  $G$ , see [9] for discussion. We use the convention that the length of a curve  $c$  with respect to  $G$  is infinity if  $c$  is not in  $D$ , a.e. Let  $l^{(a,b)}(c)$  denote the length with respect to  $G$  of  $c$  between  $c(a)$  and  $c(b)$ . We have

LEMMA 4.  $\tilde{l}^{(a,b)}(c) = l^{(a,b)}(c)$ .

*Proof.* Both lengths are local functionals, i.e.,  $\tilde{l}^{(a,b)}(c) = \tilde{l}^{(a,d)}(c) + \tilde{l}^{(d,b)}(c)$ , so that it suffices to work locally. Suppose that  $U_2 \subset U_1 \subset U_0 \subset M$  are open neighborhoods of  $c(a)$ , such that  $TU_0$  is trivial, that  $\bar{U}_2, \bar{U}_1$  are compact, that  $\partial U_1$  is a smooth submanifold and that  $U_2 = \{y : r(c(a), y) < T\}$  for some  $T$  sufficiently small. Let  $\phi \in C_0^\infty(U_0)$  satisfy  $0 \leq \phi \leq 1$ ,  $\phi|_{U_2} \equiv 1$ ,  $\phi|_{U_1} > 0$ ,  $\text{supp}(\phi) = \bar{U}_1$ .

We claim that  $\phi^{-1}G$  gives rise to a distance  $r_\phi$  on  $U_1$  such that  $(U_1, r_\phi)$  is a complete metric space. Indeed, pick a smooth complement  $\delta$  of  $D$  over  $U_0$ :  $TU_0 = D \oplus \delta$ . Pick a smooth frame  $\{\xi_1, \dots, \xi_{m-n}\}$  of  $\delta$ ; since  $TU_0$  is trivial these exist. Declare  $\delta$  to be orthogonal to  $D$  and  $\{\xi_1, \dots, \xi_{m-n}\}$  to be orthonormal. This gives rise to a Riemannian metric tensor  $\mathcal{G} = G + H$ , such that  $\mathcal{G}|_D = G$ .  $\phi^{-1}\mathcal{G}$  is a complete Riemannian metric on  $U_1$ : the dual tensor of  $\phi^{-1}\mathcal{G}$  is  $\phi\mathcal{G}^*$ , and  $\phi\mathcal{G}^*$  gives rise to a complete geodesic flow on the unit co-sphere bundle in  $T^*U_1$ . We may see this because the unit co-sphere bundle over any compact in  $U_1$  is compact so that the Hamiltonian vector field  $\eta$  associated to  $\phi\mathcal{G}^*$ , which restricts to the co-sphere bundle, is complete unless the geodesics of  $\phi^{-1}\mathcal{G}$  can reach  $\partial U_1$  from an interior point with a finite length. But any such curve will also have a finite length with respect to  $\mathcal{G}$ , and  $\phi^{-1}(c(t))$  will blow up as  $c(t)$  approaches  $\partial U_1$  faster than any inverse power of the distance from  $\partial U_1$ , so that the length with respect to  $\phi^{-1}\mathcal{G}$  is infinity.

Thus our claim follows from Theorem 2. Now, let  $\mathcal{E}_\varepsilon = G + \varepsilon^{-1}H$ . By the above  $\phi^{-1}\mathcal{E}_\varepsilon$  is complete on  $U_1$  for each  $\varepsilon > 0$ . Let  $r_\phi^\varepsilon$  denote the associated distance. By arguments, in the proof of Theorem 2  $r_\phi^\varepsilon \leq r_\phi$ , also  $r_\phi^\varepsilon \leq r_\phi^{\varepsilon_1}$  if  $\varepsilon_1 < \varepsilon$ . Let  $r_\phi^0 = \lim_{\varepsilon \rightarrow 0} r_\phi^\varepsilon$ . It follows that  $r_\phi^0$  is a continuous metric,  $r_\phi \geq r_\phi^0 \geq r_\phi^\varepsilon$  for  $\varepsilon > 0$ .

Indeed, the defining properties of a metric are preserved in the limit of an increasing family of such metrics, continuity follows from the triangle inequality and the fact that  $r_\phi(x, y) \geq r_\phi^0(x, y)$  while  $r_\phi$  is continuous. Let  $\Pi_\delta$  denote the orthogonal projection, with respect to  $\mathcal{E}_\varepsilon$ , on  $\delta$  (this is independent of  $\varepsilon > 0$ ). Suppose that  $c$  is a curve in  $U$ , with  $\Pi_\delta \dot{c} \neq 0$  on a subset of finite measure in  $[a, b]$  for every  $b > a$ ,

so that

$$\int \sqrt{\phi^{-1}\mathcal{E}_\varepsilon(\Pi_\delta\dot{c}, \Pi_\delta\dot{c})} dt > 0.$$

For  $\varepsilon \geq 0$ , define

$$\begin{aligned} \tilde{l}_{\varepsilon,\phi}^{(a,b)}(c) &= \sup_{\mathcal{P}} \sum r_\phi^\varepsilon(c(t_i), c(t_{i+1})), \quad \text{and} \\ \tilde{l}_\phi^{(a,b)}(c) &= \sup_{\mathcal{P}} \sum r_\phi(c(t_i), c(t_{i+1})). \end{aligned}$$

It follows from well known results of Riemannian geometry that, for  $\varepsilon > 0$ ,  $[r_\phi^\varepsilon(x, y)]^2$  is  $C^\infty$  for  $x$  and  $y$  sufficiently close, so that  $\tilde{l}_{\varepsilon,\phi}^{(a,b)}(c)$  is equal to the Riemannian length of  $c$  with respect to  $\phi^{-1}\mathcal{E}_\varepsilon$ . Indeed, the triangle inequality implies that refining a partition does not decrease the sum, so that any such sum is bounded by

$$\int_a^b \left[ \frac{d}{ds} r_\phi^\varepsilon(c(t), c(t+s)) \right]_{s=0} dt;$$

and one may find a sequence of partitions which converges to this integral. But,

$$\frac{d}{ds} r_\phi^\varepsilon(c(t), c(t+s))|_{s=0} = \sqrt{\phi^{-1}\mathcal{E}_\varepsilon(\dot{c}(t), \dot{c}(t))}$$

(use geodesic coordinates at  $c(t)$ ).

Thus,

$$\tilde{l}_{\varepsilon,\phi}^{(a,b)}(c) \geq \varepsilon^{-1/2} \int_a^b \sqrt{\phi^{-1}H(\Pi_\delta\dot{c}, \Pi_\delta\dot{c})} dt.$$

The inequalities  $r_\phi^\varepsilon \leq r_\phi^0 \leq r_\phi$  imply that  $\tilde{l}_{\varepsilon,\phi}^{(a,b)}(c) \leq \tilde{l}_{0,\phi}^{(a,b)}(c) \leq \tilde{l}_\phi^{(a,b)}(c)$ . Thus,  $\tilde{l}_{0,\phi}^{(a,b)}(c) = \infty$  and  $\tilde{l}_\phi^{(a,b)}(c) = \infty$ . Therefore,  $\tilde{l}_{\varepsilon,\phi}^{(a,b)}(c) < \infty$  implies that  $\dot{c} \in D$ , a.e. But,  $\dot{c} \in D$  a.e. implies  $\mathcal{E}_\varepsilon(\dot{c}, \dot{c}) = G(\dot{c}, \dot{c})$  a.e., so that  $\tilde{l}_{\varepsilon,\phi}^{(a,b)}(c)$  is the length of  $c$  with respect to  $\phi^{-1}G$  for  $\varepsilon > 0$  and also for  $\varepsilon = 0$ . It follows that  $r_\phi(x, y) = \inf \tilde{l}_{0,\phi}^{(a,b)}(c)$  where the infimum is over  $c$  such that  $c(a) = x, c(b) = y$ . Thus, by the remark following Proposition 1.6 of [9], it follows that  $\tilde{l}_\phi^{(a,b)}(c) = \tilde{l}_{0,\phi}^{(a,b)}(c)$ .

But, one may see that  $r_\phi(x, y) = r(x, y)$  for  $x, y \in U_2$  such that  $r_\phi(c(a), x) + r_\phi(c(a), y) < \frac{2}{3}T$ , since in this case a minimizing geodesic  $\gamma(t)$  between  $x$  and  $y$  is contained entirely in  $U_2$  (the triangle inequality applied to  $r(c(a), \gamma(t))$  plus the fact that  $\gamma(t)$  is minimizing does the trick). Thus, for  $b$  sufficiently small  $\tilde{l}_{0,\phi}^{(a,b)}(c) = l^{(a,b)}(c)$  and  $\tilde{l}_\phi^{(a,b)}(c) = \tilde{l}_\phi^{(a,b)}(c)$ . □

*Proof of Lemma 3.* Suppose that there is a minimizing generalized geodesic:  $r(x, \widetilde{\exp}_x(t, p)) = t$ ,  $t \in [0, T]$ ,  $T > 0$ . Thus,  $\tilde{l}^{(0,t)}(\widetilde{\exp}_x(\cdot, p)) = t$  for  $t \in [0, T]$  hence  $l^{(0,t)}(\widetilde{\exp}_x(\cdot, p)) = t$  so that  $\widetilde{\exp}_x(\cdot, p)$  is minimizing, hence a geodesic.  $\square$

**LEMMA 5.** *Suppose that  $p \in S_x$  and that  $\mu_x(p) = T < \infty$ . Then either  $\exp_x(Tp)$  is the first critical value of  $\exp_x$  along  $t \rightarrow \exp_x(tp)$  or there exists another minimizing geodesic or generalized geodesic connecting  $x$  to  $\exp_x(Tp)$ .*

**REMARK.** Although by Lemma 3, a minimizing generalized geodesic is equal to a geodesic, it may happen that it is equal to  $t \rightarrow \exp_x(tp)$ . Thus, the converse of this lemma is *not* implied: that the existence of a minimizing generalized geodesic connecting  $r$  and  $\exp_x(Tp)$  implies that  $t \rightarrow \exp_x(tp)$  is not minimizing for  $t > T$ .

*Proof.* The proof follows closely that of Kobayashi ([15], Theorem 4.2), with exceptions which we describe below. We work in the cotangent space rather than the tangent space. We use Kobayashi's notation: a cotangent vector is denoted by  $X$  instead of  $p$ . Kobayashi's vectors  $b_k X_k$  are now not necessarily contained in a compact subset of the cotangent space. However, as  $\widetilde{\exp}_x$  maps  $\beta(S_x) \times \mathbf{R}^+$  into  $M$  and extends  $\exp_x$ , we see that the pairs  $(b_k, X_k)$  are contained in a compact subset of  $\beta(S_x) \times \mathbf{R}^+$ , so that we may take  $(b_k, X_k)$  to converge to a point  $(b, Y)$ . Then  $\widetilde{\exp}_x(t, Y)$  is a minimizing geodesic or generalized geodesic with  $\widetilde{\exp}_x(b, Y) = \exp_x(bX)$ . Indeed,

$$\begin{aligned} r(\exp_x(bX), \widetilde{\exp}_x(b, Y)) &\leq r(\exp_x(bX), \exp_x(a_k X)) \\ &\quad + r(\exp_x(b_k X_k), \exp_x(bX_k)) \\ &\quad + r(\exp_x(bX_k), \widetilde{\exp}_x(b, Y)), \end{aligned}$$

since  $\exp_x(a_k X) = \exp_x(b_k X_k)$ . But, the first term on the right of the inequality is less than  $\varepsilon/3$  for  $k$  sufficiently large since  $a_k \searrow b$ , likewise the second term for  $|b_k - b| < \varepsilon/3$ , and the third term as well since  $\exp_x(tX_k)$  converges to  $\widetilde{\exp}_x(t, Y)$ . Thus, if  $Y$  is not equal  $X$ , there is an additional minimizing geodesic or generalized geodesic. If  $Y = X$ , then as Kobayashi shows  $\exp_x(TX)$  is a critical value of  $\exp_x$ , by Theorem 4 (below) the first critical value.  $\square$

Now, we have remarked that the property of being locally minimizing is not so clear for singular Riemannian geodesics as for Riemannian geodesics. In fact, in the Riemannian situation, geodesics are

(locally) uniformly locally minimizing; this amounts to the statement that  $\mu_x$  is bounded away from zero. A result which follows easily from Kobayashi ([14], Theorem 4.3) is

**LEMMA 6.** *Suppose that there are no minimizing generalized geodesics and that  $\mu_x$  is locally bounded away from zero on  $S_x$ , then  $\mu_x$  is continuous.*

*Proof.* The proof follows directly from that of Kobayashi, with the exception that, now his vector  $Y$  may be in  $\beta(S_x) - S_x$ . The minimizing generalized geodesic  $t \rightarrow \exp_x(t, Y)$  is equal to some geodesic  $t \rightarrow \exp_x(tp)$  for some  $p$  in  $S_x$ . The only problem with this is that it may happen that  $p$  is Kobayashi's  $X$ . We exclude this possibility by hypotheses which exclude all minimizing generalized geodesic.  $\square$

**THEOREM 4.** *Singular Riemannian geodesics do not minimize past the first critical value of the exponential map.*

*Proof.* The discussion in Sternberg [15], on page 184 through the first sentence of page 187, follows as well in the singular Riemannian case as it does for "regular Lagrangians". This is also the case from the middle of page 192 to the bottom of page 193. The proof of Sternberg's Lemma 5.3 may be rewritten as follows (in the rest of the proof, we follow the notation of Sternberg). Since  $F$  is locally a diffeomorphism of  $U \subset M$ , it induces a map  $F_*: TU \rightarrow TM$ , which is full rank as a vector bundle morphism (and hence itself a local diffeomorphism). Thus, there is a sub-bundle  $D' \subset TU$  such that the biggest integrable sub-bundle of  $TU$  containing  $D'$  is  $TU$  itself, and such that  $F_*$  is locally an injection of  $D'$  onto  $D$ . Then  $(D', G' = G \circ F)$  is a singular Riemannian structure on  $U$ . Moreover, the length, with respect to  $G'$ , of any curve  $B$  in  $U$  such that  $\dot{B} \in D'$  almost always, is equal to the length of  $F \circ B$  with respect to  $G$ . Thus, if  $B$  does not minimize with respect to  $G'$  then  $F \circ B$  does not minimize with respect to  $G$ . Also, since  $F$  is a local diffeomorphism,  $U$  is covered by a collection of sets  $\{V_\alpha\}$  such that  $F_\alpha = F|_{V_\alpha}$  is a diffeomorphism onto the image of  $F_\alpha$ , so that  $F_\alpha^*$  is a diffeomorphism of  $T^*(\text{image}(F_\alpha))$  onto  $T^*V_\alpha$ .

It follows then, that the Hamiltonian vector field  $H$  associated to the dual of  $G$  is pushed forward along  $F^*$  to the Hamiltonian vector field  $H'$  associated to the dual of  $G'$ ,  $H'|_{V_\alpha} = (F_\alpha^*)_*H|_{\text{image}(F_\alpha)}$ . Thus,

we get that the following diagram commutes:

$$\begin{array}{ccc}
 T_z^*U & \xrightarrow{\exp'_z} & U \\
 (F^*)^{-1} \uparrow \downarrow F^* & & \downarrow F \\
 T_x^*M & \xrightarrow{\exp_x} & M
 \end{array}$$

where  $F(z) = x$ .

Thus, since  $F$  is a local diffeomorphism,  $\exp_x$  is singular at  $p \in T_x^*M$  if and only if  $\exp'_z$  is singular at  $F^*(p)$ , so that if the geodesic  $t \rightarrow \exp_x(tp)$  does not minimize for  $t > 1$  then neither does  $t \rightarrow \exp'_z(tF^*(p))$ .

Now, associated to the vector field  $H'$  on  $T^*U$  there is a vector field  $T(H')$  on  $T(T^*U)$  (Sternberg, p. 185). Consider the coordinates  $x^i, y^i$  on  $T^*U$  and coordinates  $x^i, y^i, \xi^i, \zeta^i$  on  $T(T^*U)$ , where  $\{x^i\}$  are coordinates pulled back from  $U$ ,  $\{y^i\}$  are coordinates on the fibers of  $T^*U$ ,  $\{\xi^i\}$  are coordinates on the subspace of  $T(T^*U)$  which are tangent to the fibers of  $T^*U$  and  $\{\zeta^i\}$  are coordinates corresponding to the tangent spaces of  $U$ . In these coordinates the vector field  $T(H')$  has a representation:

$$\begin{aligned}
 T(H') = & \sum_{i,j} \left[ (G^*)^{ij} y_j \frac{\partial}{\partial x^i} - \sum_b \frac{\partial (G^*)^{ij}}{\partial x^k} y_i y_j \frac{\partial}{\partial y_k} \right] \\
 & + \sum_{i,j} \left[ \sum_k \frac{\partial (G^*)^{ij}}{\partial x^k} y_i \xi^k + (G^*)^{ij} \zeta_i \right] \frac{\partial}{\partial \xi^j} \\
 & - \sum_{i,j,k} \left[ \frac{\partial (G^*)^{ij}}{\partial x^k} y_i \zeta_j + \sum_l \frac{\partial^2 (G^*)^{ij}}{\partial x^k \partial x^l} y_i y_j \xi^l \right] \frac{\partial}{\partial \zeta_k}.
 \end{aligned}$$

Note that the integral curves of  $T(H')$  in  $T(T^*U)$  project down to integral curves of  $H'$  in  $T^*U$ . Along an integral curve of  $H'$ , an integral curve of  $T(H')$  amounts to a solution of the following linear differential equation:

$$\begin{aligned}
 \dot{\xi}^i &= \sum_{j=1} [A_j^i \xi^j + B^{ij} \zeta_j], \\
 \dot{\zeta}_i &= - \sum_j [C_{ij} \xi^j + D_i^j \zeta_j],
 \end{aligned}
 \tag{4}$$

where the matrices  $A_j^i$ ,  $B^{ij}$ ,  $C_{ij}$ ,  $D_i^j$  are smooth functions of  $t$ , and

$$A_j^i = \sum_k \frac{\partial(G^*)^{ik}}{\partial x^j} y_k; \quad B^{ij} = (G^*)^{ij};$$

$$C_{ij} = \sum_{kl} \frac{\partial^2(G^*)^{kl}}{\partial x^i \partial x^j} y_k y_l; \quad D_i^j = \sum_k \frac{\partial(G^*)^{kj}}{\partial x^i} y_k.$$

But, the differential equation (4) is in Hamiltonian form and, in fact, are the Hamilton equations corresponding to the following optimal control problem: Find a control  $v = (v_1, v_2, \dots, v_n)$  such that the differential equation

$$(5) \quad \dot{\xi}^i = \sum_{\alpha=1}^n a_\alpha^i v_\alpha + \sum_{j=1}^m b_j^i \xi^j$$

steers the point  $\xi_1$  to the point  $\xi_2$ , and such that, among all controls which steer  $\xi_1$  to  $\xi_2$ ,  $v$  is an extremal of the functional:

$$(6) \quad \delta^2 \eta'(v) = \int_0^1 \left[ \sum_{\alpha=1}^n (v_\alpha(t))^2 + \sum_{i,j=1}^m c_{ik}(t) \xi^i(t) \xi^j(t) \right] dt$$

where the above matrices  $a_\alpha^i$ ,  $b_j^i$ , and  $c_{ij}$  are defined as follows (note that they are smooth functions of  $t$ ):

$$a_\alpha^i = g_\alpha^i; \quad b_j^i = A_j^i = D_j^i; \quad c_{ij} = C_{ij}.$$

But, the functional (6) is the second variation (see Bryson and Ho [5]) of the functional  $\eta'(u)$  defining the length of a curve in  $U$  with derivative in  $D' \subset TU$ . Thus, if the function  $t \rightarrow x(t)$  is a minimizing geodesic for all of  $t < T$  it follows that  $\delta^2 \eta'(v)$  is non-negative.

Now, the Hamiltonian function associated with (4), or equivalently with (5) and (6) is:

$$\mathcal{H}' = \frac{1}{2} \sum_{i,j=1}^m \left[ \left( \sum_{\alpha=1}^n a_\alpha^i a_\alpha^j \right) \zeta_i \zeta_j + 2b_j^i \zeta_i \xi^j - c_{ij} \xi^i \xi^j \right].$$

If  $\xi(t)$  is the trajectory associated with an extremal control  $v$ , then:

$$\delta^2 \eta'(v) = \int_0^1 \left[ \sum_{i=1}^m \dot{\xi}^i \zeta_i - \mathcal{H}'(\xi, \zeta) \right] dt$$

because:

$$\dot{\xi}^i = \frac{\partial \mathcal{H}'}{\partial \zeta_i} \quad \text{and} \quad \dot{\zeta}_i = -\frac{\partial \mathcal{H}'}{\partial \xi^i}$$

so that Sternberg's Lemma 5.4 on p. 195 also holds in the singular Riemannian case. The consequence of this is that:

$$\delta^2 \eta'(v) = \frac{1}{2} \sum_i [\xi^i(1)\zeta_i(1) - \xi^i(0)\zeta_i(0)].$$

Now, assume that  $y_0 \in T_z^*U$  is a critical point of  $\exp'_z$ , but that the geodesic  $\gamma: t \rightarrow \exp'_z(ty_0)$  is minimizing for all  $t$  in  $[0, 1 + \varepsilon]$  for some  $\varepsilon > 0$ . Now, one solution of the equation (4) along  $\gamma$  is the zero solution:  $\xi(t) = \zeta(t) = 0$  for all  $t \in [0, 1 + \varepsilon]$ . Another solution may be constructed in the following way. Let  $\zeta_0$  be an element of  $\ker[d(\exp'_z)] \subset T_{y_0}(T_z^*U)$ . Note that  $\zeta_0$  is not tangent to the line spanned by  $y_0$  in  $T_z^*U$ , since, for all  $t$ , the length along  $\gamma$  between  $\gamma(0)$  and  $\gamma(t)$  is  $[G'_z(ty_0, ty_0)]^{(1/2)}$ , and this is strictly increasing if  $y_0$  is not in  $\ker[G^*]$ . Let  $s \rightarrow y^s$  be a smooth curve in  $T_x^*U$  such that  $y^0 = y_0$  and such that  $(d/ds)y^s|_{s=0} = \zeta_0$ . Then:

$$\frac{d}{ds}[\exp tH'](z, y^s)|_{s=0} = [\exp tT(H')](z, y_0, 0, \zeta_0)$$

for all  $t$  by definition of  $T(H')$ . Let

$$(s(t), y(t), \xi(t), \zeta(t)) = \exp[tT(H')](z, y_0, 0, \zeta_0).$$

Then,  $(\xi(t), \zeta(t))$  is a solution of (\*1) along  $(x(t), y(t))$ . Since  $\exp[tT(H')]$  is a diffeomorphism for all  $t$ , and  $\exp[tT(H')] = d(\exp[tH'])$ , the fact that  $d(\exp'_z)(y_0, \zeta_0) = (\exp'_z(y_0), 0)$  implies that, for  $\pi$  the canonical projection of  $T^*U$  on  $U$ , that

$$(d\pi)(x(1), y(1), \xi(1), \zeta(1)) = (x(1), \xi(1)) = (x(1), 0);$$

i.e., that  $\xi(1) = 0$ . Thus, the value of  $\delta^2 \eta'(v)$  associated with  $(\xi(t), \zeta(t))$  is:

$$\frac{1}{2} \sum \xi^i(1)\zeta_i(1) - \xi_i(0)\zeta_i(0) = 0.$$

Now, let the curve  $\xi: t \rightarrow \xi(t)$  be the projection of the integral curve of  $T(H')$  with initial condition  $\zeta_0$ . Then,  $\xi(t)$  is not tangent of infinite order at  $t = 1$  to the curve  $\xi(t) \equiv 0$ , for otherwise the two curves must be equal for all  $t$ . Define the curve  $\Xi: [0, 1 + \varepsilon] \rightarrow \mathbf{R}^m$  to be equal to  $\xi(t)$  when  $t < 1$ , and equal to 0 when  $t > 1$ . Thus, the curve  $\Xi(t)$  gives rise to a control  $v$  for which  $\delta^2 \eta'(v) = 0$ , but on the

other hand  $\Xi(t)$  does not come from a solution of (4) since  $\Xi(t)$  is not infinitely smooth. Thus, zero is not a minimum of  $\delta^2\eta'(v)$  near the curve  $(x(t), y(t), 0, 0)$ , so that  $(x(t), y(t))$  is not minimizing for  $t > 1$ .  $\square$

**4. Applications.** In this section we will assume throughout that the conditions of Lemma 6 hold. There are examples for which this is the case; in particular, if  $\Delta$  is locally diffeomorphic to the “subelliptic Laplacian” on the Heisenberg group. The Hamiltonian flow for  $G^*$  may then be computed explicitly. A somewhat more general situation than this is when  $D$  has the property that the commutator of any section of  $D$  with  $D$  yields  $TM$ , i.e., when the doubly characteristic set of  $\Delta$  is symplectic. Bismut [2] has discussed this condition. Strichartz [19] has shown in this circumstance that any geodesic is locally minimizing; the proof of his Theorem 5.4 gives an estimate of the distance for which the geodesic  $\exp_x(tp)$  is minimizing which is locally uniform in  $p$ . We can also show that there are no minimizing generalized geodesics in this case, however, the methods required are more appropriately introduced in the sequel to this paper.

**COROLLARY 4.** *For every  $x \in M$ , there is an open dense subset of  $M$ , call it  $U(x)$ , on which the function  $y \rightarrow r(x, y)$  is infinitely differentiable.*

*Proof.* We know that every two points in  $M$  are connected by a minimizing geodesic, and that geodesics do not minimize past the first critical point of  $\exp_x$ . Thus,  $\exp_x$  is an open map on the set  $V(x)$  of  $(tp) \in T_x^*M$  such that  $p$  is in  $S_x$  and such that  $0 < t < \mu_x(p)$ . Since  $\mu_x$  is continuous,  $V(x)$  is open. But,  $\bar{V}(x)$ , the closure of  $V(x)$ , is mapped onto  $M$  by  $\exp_x$ . Thus, the image,  $U(x)$ , of  $V(x)$  is open and dense.

Also, we have that for  $p \in S_x$  and  $t \leq \mu_x(p)$ ,  $r(x, \exp_x(tp)) = t$ . Thus, the pullback  $(\exp_x)^*r(x, y)^2 = G^*(tp, tp)$  for  $y = \exp_x(tp)$ . Thus, on  $U(x)$ ,  $r(x, \cdot)$  is smooth.  $\square$

Recall that, if  $\omega$  is the canonical symplectic form on  $T^*M$  and  $\omega^*$  is its dual, the Hamiltonian vector field  $H$  associated with the function  $h = G^*(p, p)$  is  $\omega^*(dh, \cdot)$ . In the standard local coordinates on  $T^*M$ ,  $H$  may be written as

$$H = G^* \left( p, \frac{\partial}{\partial x} \right) - \sum_{i=1}^m \left[ \frac{\partial}{\partial x_i} G^*(p, p) \right] \frac{\partial}{\partial \xi_i}.$$

Now recalling also that  $G^*$  defines a mapping from  $T^*M$  to its dual  $TM$ , we may see that  $G^*(p, \partial/\partial x)$  may also be written  $G^*(p, \cdot)$ . We have the following lemma.

**LEMMA 7.** *Suppose that  $y = \exp_x(sp_0) \in U(x)$ . Then  $G_y^*(dr, \cdot)$ , considered as an element of  $D_y$ , is equal to  $(d/ds)\exp_x(sp_0)$ .*

*Proof.* Now,  $\exp_x(sp_0)$  is the projection  $\pi[\exp(sH)](x, p_0)$ . Therefore, the tangent vector  $(d/ds)\exp_x(sp_0)$  is equal to the push forward

$$\pi_* \frac{d}{ds} [\exp(sH)](x, p_0) = \pi_* H([\exp(sH)](x, p_0)) = G_y^*(p(s), \cdot),$$

where  $p(s)$  is the cotangent vector in  $T_y^*M$  which is the image of  $p_0 \in T_x^*M$  under  $\exp(sH)$ . Let  $V(x) \subset T_x^*M$  be the open subset that is defined in the proof of the previous corollary, so that  $\exp_x|V(x)$  is a diffeomorphism onto its image,  $U(x)$ , in  $M$ .

Now, let  $(z_1, z_2, \dots, z_m)$  be a “spherical” coordinate system on  $T_x^*M$ , i.e., a coordinate system such that  $(z_1)^2(p) = G_x^*(p, p)$  and such that for  $i \geq 2$ ,  $z_i(sp) = z_i(p)$  for all  $s > 0$ . Then, since  $\exp_x|V(x)$  is a diffeomorphism onto  $U(x)$ , we may define coordinates  $\{x_i\}$  on  $U(x)$  by  $x_i = [(\exp_x)^{-1}]^* z_i$ . These coordinates have the properties that

- (a)  $x_1 = r(x, \cdot)$ .
- (b) The function  $s \rightarrow x_i(\exp_x(sp_0))$  is constant for  $s < \mu_x(p_0)$ , for  $i \geq 2$ .

Thus, we have that for  $s < \mu_x$ ,

$$\frac{d}{ds} x_1(\exp_x(sp_0)) = 1 \quad \text{and} \quad \frac{d}{ds} x_i(\exp_x(sp_0)) = 0 \quad \text{for } i > 2.$$

On the other hand, Hamilton’s equations give that

$$\dot{x}_i = (\partial/\partial \xi_i)G^*(p, p);$$

where the functions  $\xi_i$  are functions on  $T^*U(x)$  which on each fiber restrict to the dual basis  $\{\xi_1, \dots, \xi_m\}$  of the frame  $\{p_1, \dots, p_m\} = \{dx_1, \dots, dx_m\}$ . Thus,  $(\partial/\partial \xi_i)G^*(p, p) = 2 \sum_{j=1}^m (G^*)^{ij} \xi_j$ , so that with respect to the frame  $\{\partial/\partial x^i\}$ ,

$$G^*(dr, \cdot) = (2, 0, 0, \dots, 0) = 2 \frac{\partial}{\partial r}. \quad \square$$

**DEFINITION.** We call the coordinates  $\{x_1, \dots, x_m\}$  introduced in the preceding lemma a “singular Riemannian geodesic spherical coordinate system” on the set  $U(x)$ .

**COROLLARY 5.** For  $y \neq x$ , let  $v \in D_y$ , and suppose that  $y = \exp_x(s_0 p_0)$  for some  $p_0 \in S_x$ . Suppose also that

$$G(v, (d/ds) \exp_x(sp_0)|_{s=s_0}) = 0.$$

Then  $dr(x, \cdot)(v) = 0$ .

*Proof.* Recall that in the coordinates defined in the proof of the last lemma, we have that  $G^*(dr, \cdot) = \partial/\partial r$ . Note also that  $G^*(dr, \cdot)$  determines a linear functional on  $T^*U(x)$ , i.e., a tangent vector. In fact, this tangent vector is in  $D$ , since  $D$  is the dual space of  $T^*M/(\ker G^*)$ . Thus,  $G(G^*(dr, \cdot), \cdot)$  defines an element in  $D^* = T^*M/(\ker G^*)$ , in fact  $G(G^*(dr, \cdot), \cdot) = dr, \text{ mod}(\ker G^*)$ . But,  $G(v, \partial/\partial r) = G(G^*(dr, \cdot), v) = dr(v) = 0$ .

**COROLLARY 6.** With respect to the singular Riemannian geodesic spherical coordinate system on  $U(x)$ , we have that  $G^* = \partial/\partial r \otimes \partial/\partial r +$  (terms involving only  $\partial/\partial x_i \otimes \partial/\partial x_j$  for  $i, j \geq 2$ ).

*Proof.* See Hörmander [13], (Appendix C.5).

**LEMMA 8.** Suppose  $y \in U(x)$ . Then  $r(z_1, z_2)$  is a  $C^\infty$  function of  $(z_1, z_2) \in M \times M$  at  $(z_1, z_2) = (y, x)$ .

*Proof.* We have that there exists  $p_0 \in S_x$  and  $\lambda \in \mathbf{R}^+$  such that  $0 < \lambda < \mu_x(p_0)$  and  $y = \exp_x(\lambda p_0)$ . Thus,  $y = \pi \exp(H)(x, \lambda p_0)$ . Suppose  $p_1$  is the element of  $T_y^*M$  such that  $(y, p_1) = \exp(H)(x, \lambda p_0)$ . Now, the fact that  $y \in U(x)$  implies that  $\exp_x$  is full rank at  $\lambda p_0$  by Theorem 4. But, since  $\exp(H)$  is a diffeomorphism, this is equivalent to the fact that  $d \exp(H)$  maps  $T_{\lambda p_0}(T_x^*M) \subset T_{(x, \lambda p_0)}(T^*M)$  onto a subspace of  $T_{(y, p_1)}(T^*M)$  which is transversal to  $T_{p_1}(T_y^*M)$ . It follows then, since  $\exp(H)(y, -p_1) = \exp(-H)(y, p_1)$ , that  $d \exp(H)$  maps  $T_{-p_1}(T_y^*M)$  into a subspace of  $T_{(x, -\lambda p_0)}(T^*M)$  which is transversal to  $T_{-\lambda p_0}(T_x^*M)$ . Thus,  $\exp_y$  maps  $-p_1$  to  $x$ , and  $d \exp_y$  is full rank at  $-p_1$ . Moreover, since  $d \exp_x$  is also full rank at  $\lambda' p_0$  for  $0 < \lambda' < \lambda$  it follows that  $d \exp_y$  is full rank at  $-\lambda' p_1$  for  $0 < \lambda' < 1$ , so that  $x \in U(y)$ . Thus,  $r(z_1, z_2) \in C^\infty$ , separately in the variables  $z_1$  and  $z_2$  at  $(x, y)$ , hence jointly.  $\square$

We have the following easy consequence.

**COROLLARY 7.** *There is a locally finite cover of  $M$  of the form  $\{U(x_i)\}_{i=1}^\infty$  for some countable set of points  $\{x_i\}_{i=1}^\infty \subset M$ .*

**Acknowledgment.** Parts of this paper are in the author's 1983 Harvard Ph.D. Thesis.

The author wishes to acknowledge the guidance of Professor Roger Brockett and useful conversations with Peter Crouch and John Quigg, as well as the assistance of Chris Muñoz in obtaining relevant literature. The author wishes to express his gratitude to the referee for helpful suggestions as well as the information that reference [11], which unfortunately is unavailable to the author, is an important new paper in this area.

#### REFERENCES

- [1] J. Ballieul, *Geometric methods for nonlinear optimal control problems*, J. Optim. Theory Appl., **25** (1978), p. 519.
- [2] J. M. Bismut, *Large Deviations and the Malliavin Calculus*, Birkhäuser, Boston (1984).
- [3] R. Brockett, *Control Theory and Singular Riemannian Geometry*, in *New Directions in Applied Mathematics*, Springer-Verlag, New York (1981).
- [4] ———, *System theory on group manifolds and coset spaces*, SIAM J. Control, **10** (1972), 265.
- [5] A. E. Bryson and Y. C. Ho, *Applied Optimal Control; Optimization, Estimation and Control*, Blaisdell Publ., Waltham, Mass. (1969).
- [6] L. Cesari, *Existence theorems for weak and usual optimal solutions in Lagrange problems with unilateral constraints*, I, Trans. Amer. Math. Soc., **124** (1966), 369.
- [7] C. Fefferman and D. H. Phong, *Subelliptic Eigenvalue Problems*, in *Proceedings of the Conference on Harmonic Analysis in Honor of Antoni Zygmund*, Wadsworth Math. Series 1981, p. 590.
- [8] B. Gaveau, *Principe de moindre action, propagation de la chaleur et estimées sous elliptiques sur certains groupes nilpotents*, Acta Math., **139** (1977), 95.
- [9] M. Gromov, *Sur Structures Métriques pour les Variétés Riemanniennes*, Cedric-Nathan 1981.
- [10] N. Gunther, *Hamiltonian Mechanics and Optimal Control*, Ph.D. Thesis, Harvard University (1982).
- [11] U. Hamenstadt, *On the geometry of Carnot-Carathéodory metrics*, preprint.
- [12] L. Hörmander, *Hypoelliptic second order differential equations*, Acta Math., **119** (1967), 147.
- [13] ———, *The Analysis of Linear Partial Differential Operators*, III, Springer-Verlag (1985).
- [14] Y. Kannai, *Off diagonal short time asymptotics for fundamental solutions of diffusion equations*, Comm. in P.D.E., **2** (1977), 781.

- [15] S. Kobayashi, *On Conjugate and Cut Loci*, in *Studies in Global Geometry and Analysis*, Studies in Math., Vol. 4, MAA (1967).
- [16] A. Nagel, E. M. Stein and S. Wainger, *Boundary behavior of functions holomorphic in domains of finite type*, Proc. Natl. Acad. Sci. USA, **78** (1981), 6596.
- [17] S. Sternberg, *Lectures on Differential Geometry*, Chelsea Publishing Co., New York (1983).
- [18] J. Mitchell, *On Carnot-Carathéodory metrics*, J. Diff. Geom., **21** (1985), 35–45.
- [19] R. S. Strichartz, *Sub-Riemannian geometry*, J. Diff. Geom., **24** (1986), 221–263.

Received November 24, 1986 and in revised form July 28, 1987. This work was supported in part by the Office of Naval Research under JSEP contract number N0041-75-C-0648 and in part by Arizona State University through a Faculty Grant in Aid.

ARIZONA STATE UNIVERSITY  
TEMPE, AZ 85287