This paper has appeared in "Optimal control and piecewise analyticity of the distance function," in *Optimization and Nonlinear Analysis,* A. Ioffe and S. Reich Eds., Pitman Research Notes in Mathematics, Longman Publishers (1992), pp. 298-310.

Optimal control theory and piecewise analyticity of the distance function for some real-analytic sub-Riemannian metrics

by H J SUSSMANN

§1. Introduction .

In this paper we will apply techniques of optimal control theory (in particular, the method of establishing piecewise regularity of the value function by means of the theory of subanalytic sets and stratifications) to prove the following theorem: for a sub-Riemannian structure of class C^ω on a nonholonomic two-dimensional C^ω subbundle of the tangent bundle of a three-dimensional C^ω manifold, of degree of non-holonomy equal to two, all sufficiently small spheres are subanalytic sets, and the distance function to a given point p is subanalytic on any annulus $\Sigma(p,\varepsilon,T)=\{x:\varepsilon\leq \mathrm{dist}(x,p)\leq T\}$, provided that the closed ball $B(p,T)=\{x:\mathrm{dist}(x,p)\leq T\}$ is compact. This implies in particular that the spheres $\{x:\mathrm{dist}(x,p)=R\}$ are analytically stratified sets for $R\leq T$.

In recent years, various results about piecewise smoothness of value functions have been proved (cf. [11], [15], [17], [18], [21], [22], [23], [28], [29], [30], [31], [32], [33], [34]) by means of a method that involves a combination of "trajectory analysis" (cf. [11], [14], [16], [17], [19], [20], [21], [22], [23], [24], [25], [26], [27], [35]) and the theory of subanalytic sets and stratifications. The main idea is that the value function V of an optimal control problem is a solution of the associated Hamilton-Jacobi-Bellman equation, and this equation can be solved by the method of characteristics. The characteristics of interest are none other than the optimal trajectories of the system, so we can view the graph of V as constructed by putting together characteristics. If we manage to establish good regularity properties of the optimal trajectories, then it is to be expected that the graph of V will also exhibit piecewise regularity properties of some kind. One way to make this precise is to find a family of optimal trajectories $t\rightarrow\gamma(t,\alpha)$, depending on a parameter α with values in some set A, such that the graph of V is the union of the graphs of these curves. If A is a subset of some real-analytic manifold N, and γ depends real-analytically on t and α, then this will express the graph of V as the image of $\mathbb{R}\times A$ under a real-analytic map μ. If A is a subanalytic subset of N, then this will imply that the graph of V is subanalytic, provided that the map μ is proper on the closure of $\mathbb{R}\times A$.

(The method works under weaker conditions as well: (i) It is not necessary that all the trajectories $t \to \gamma(t,\alpha)$ be optimal. It suffices that among the trajectories in this family there exist a subset consisting of optimal trajectories and large enough to generate the full graph of V. For example, on a complete Riemannian manifold, if we are interested in the function V_p given by $V_p(x) = \mathrm{dist}(x,p)$, we could consider the family of *all* geodesic arcs γ that start at p, are defined on some interval $[0,T_\gamma]$, and are parametrized by arc-length. In this case the parameter α would be the pair (v,T_γ), where $v = \gamma'(0)$ is the initial direction of the geodesic, so v is a unit tangent vector at p. (ii) The map γ need not be real-analytic; it suffices that it be subanalytic, which in particular allows γ to be "piecewise analytic." For example, for a linear time-optimal control problem with polyhedral control set, we know that bang-bang trajectories are a class sufficiently rich to contain optimal trajectories for every possible initial and target point. In this case, the parameter α would be the sequence of switching times characterizing a particular bang-bang trajectory.) To apply the method, the crucial part is to find a suitable family γ. This is done by "trajectory analysis": we study the structure of the optimal trajectories using the Maximum Principle of optimal control theory, and then try to restrict the trajectories even further, by applying other considerations, such as high-order necessary conditions for optimality or geometric methods, in order to end up with a γ that satisfies the required conditions.

The purpose of this note is to illustrate this general technique by proving a result on the piecewise regularity of the distance function associated to certain Carnot-Carathéodory metrics (i.e. to certain "sub-Riemannian structures"). The proof will provide a particularly transparent illustration of the use of the technique. In particular, the role of the properness condition will be significant. Properness is important because the class of subanalytic sets is not closed under the operation of taking images under general real-analytic maps, but is closed under images by *proper* maps. In our case, to satisfy the properness condition we will have to establish an *a priori* estimate on the size of the initial value $\lambda(0)$ of the adjoint vector of the Maximum Principle. Precisely, we will have to show that, if $\lambda(0)$ is too large, then the corresponding extremal cannot be optimal. It then turns out that this corresponds to a simple geometric fact: if $\lambda(0)$ is too large, then the corresponding control is highly oscillatory, and the oscillations almost cancel out, leading to an almost constant trajectory, which cannot be time-optimal.

We now give the precise definitions and state the main results.

Consider a *sub-Riemannian structure* (cf. [4], [5], [12], [37], [38]) on a smooth n-dimensional connected manifold M, i.e. a pair (E,g), where E is a nonholonomic smooth subbundle of the tangent bundle TM, and g is a smooth Riemannian metric on E. ("Nonholonomic" means that the Lie algebra L generated by the smooth sections of E satisfies $\{X(x) : X \in L\} = T_xM$ for all $x \in M$.) The metric $d_{E,g}$ associated to this structure is the function $d_{E,g}(p,q) = \inf\{|\gamma|\}$, where the infimum is taken over all absolutely continuous E-admissible paths $\gamma : [a_\gamma, b_\gamma] \to M$ that go from p to q, and $|\gamma| = \int_{a_\gamma}^{b_\gamma} g(\gamma'(t),\gamma'(t))^{1/2}dt$. (A path γ is *E-admissible* if $\gamma'(t) \in E(\gamma(t))$ for almost all

t.) It is well known that $d_{E,g}(p,q)$ is finite for all p, q and is in fact a metric that gives rise to the topology of M.

It is an important open problem to understand in detail the structure of the spheres relative to this metric, and the regularity properties of the distance function V_p, defined by $V_p(x) = d_{E,g}(x,p)$ for a fixed point $p \in M$. As opposed to the Riemannian case, if $\dim E < n$ then the function V_p cannot be smooth on any punctured neighborhood of p, no matter how small. (This follows from the general necessary condition for stabilizability given in [2].) The spheres $\{x : V_p(x) = R\}$ have been computed explicitly for the Heisenberg group, and have turned out to be stratified sets (cf. [3], [4]).

It is natural to conjecture that, under suitable analyticity conditions, V_p might at least be subanalytic, which would imply in particular that the spheres are stratified sets. (Here we call a function *subanalytic* if its graph is subanalytic. Other definitions of "subanalytic function" have been proposed in the literature, e.g. including the exta requirement that the function be locally bounded, or continuous, but in our case V_p *is* continuous and our definition is equivalent to all the others.) Such a result would be of importance in a number of applications such as non-holonomic motion planning (cf. [6], [10]), where the piecewise smoothness of V_p would correspond to the property that one can use optimal control to construct a piecewise smooth feedback steering the whole space to p.

In this note we announce such a result, and outline the proof, for the case of a general real-analytic sub-Riemannian structure in dimension three, with $\dim E = 2$ and degree of non-holonomy 2. (The *degree of non-holonomy* is the smallest δ such that the sections of TM can be locally spanned by sections of E and Lie brackets of such sections of degree $\leq \delta$.) This will be established by using results from optimal control theory (in particular the "Maximum Principle", cf. [1], [7], [9]) in order to prove certain facts that then make it possible to apply the general machinery of subanalytic sets and stratifications, as presented, e.g., in [36]. It will be clear from the proof outlined here that the main difficulty in extending the result to higher dimensions lies in the control theoretic part. (Such an extension would follow immediately if one could prove higher-dimensional analogues of our bound on λ_3, cf. below.)

Our result is:

Theorem 1.1. *Let M be a connected three-dimensional real-analytic manifold, and let E be a two-dimensional real-analytic subbundle of the tangent bundle TM of M, such that, locally, the sections of TM are spanned by two linearly independent sections Y_1 and Y_2 of E together with their Lie bracket $[Y_1, Y_2]$. Assume E is endowed with a real-analytic metric g, and let $d_{E,g} : M \times M \to \mathbb{R}$ be the corresponding distance function. Fix $p \in M$, and let $V_p(q) = d_{E,g}(q,p)$. Let $T > 0$ be such that the ball $B(p,T) = \{x : d_{E,g}(x,p) \leq T\}$ is compact. Then for every ε such that $0 < \varepsilon \leq T$ the graph $G_{\varepsilon,T}(V_p) = \{(x,t) : x \in M, V_p(x) = t, \varepsilon \leq t \leq T\}$ of the restriction of V_p to the annulus $\Sigma(p,\varepsilon,T) = \{x : \varepsilon \leq V_p(x) \leq T\}$ is a subanalytic subset of $M \times \mathbb{R}$.* ∎

Remark 1.1. It follows in particular that the function V_p is "piecewise analytic" on $\Sigma(p,\varepsilon,T)$, in the sense that there exists a finite stratification $\mathcal{S}$ of $\Sigma(p,\varepsilon,T)$ (with real-analytic, subanalytic strata), such that V_p is analytic on each stratum of $\mathcal{S}$. ▮

Remark 1.2. Theorem 1.1 implies, in particular, that every sphere $\{x : d_{E,g}(x,p) = t\}$, for $0 < t \leq T$, is subanalytic and compact, so these spheres have finite real-analytic subanalytic stratifications. ▮

Remark 1.3. Our proof does not allow us to conclude that V_p is subanalytic on a neighborhood of p. ▮

The basic idea of the proof is as follows. The compactness of the ball $B(p,T)$ enables us to conclude that every point of $B(p,T)$ can be attained from p by an *optimal* (i.e. length-minimizing) E-admissible arc. Optimal arcs can be analyzed by means of the Maximum Principle of Optimal Control Theory (cf. [1], [7], [9]). In our case, this leads to the construction of a function H and corresponding Hamiltonian vector field X on the cotangent bundle T^*M with the property that every minimizing E-admissible arc is the projection of an integral curve of X along which H has the constant value $-\frac{1}{2}$. In control theory language, this corresponds to writing the equations of motion locally as $x' = u_1 Y_1(x) + u_2 Y_2(x)$, where u_1 and u_2 are controls, subject to the restriction $u_1^2 + u_2^2 \leq 1$, and (Y_1, Y_2) is an orthonormal basis of sections of E. The Control Theory Hamiltonian is then the real-valued function on $T^*M \times \mathbb{R} \times \mathbb{R}^2$ given by $\mathcal{H}(\lambda, x, \lambda_0, u) = u_1\langle\lambda, Y_1(x)\rangle + u_2\langle\lambda, Y_2(x)\rangle + \lambda_0$. Here λ is a nonzero covector at x, and λ_0 is a real number. The Maximum Principle says that along an extremal λ must satisfy the "adjoint equation," u must mimimize $\mathcal{H}$, $\mathcal{H}$ must vanish, and the constant λ_0 must be ≥ 0. "Abnormal" trajectories, for which $\lambda_0 = 0$, can occur in general (cf. [8], [13]), but in the special case considered here they do not occur, because the "strong bracket-generating" condition of [13] holds. So $\lambda_0 \neq 0$, and we may normalize so that $\lambda_0 = 1$. Then the Hamiltonian minimization condition entails that $u_i = -\langle\lambda, Y_i(x)\rangle$, which means that the extremals in the control theory sense are precisely the integral curves of the Hamilton vector field X of $H = -\frac{1}{2}\langle\lambda, Y_1(x)\rangle^2 - \frac{1}{2}\langle\lambda, Y_2(x)\rangle^2$. (It also follows that along a minimizing trajectory the constraint $u_1^2 + u_2^2 \leq 1$ is actually an equality, i.e. $u_1^2 + u_2^2 = 1$.)

We can then try to apply the method of [36] to prove that the graph $G_{\varepsilon,T}(V_p)$ is subanalytic in $M \times \mathbb{R}$. We write the statement that $(x,t) \in G_{\varepsilon,T}(V_p)$ in the equivalent form:

$$\varepsilon \leq t \leq T \ \wedge \ (\exists\lambda)(\zeta(\lambda,p,t) = x) \ \wedge \ \Big(\sim (\exists\bar{\lambda})(\exists\bar{t})(\zeta(\bar{\lambda},p,\bar{t}) = x \ \wedge \ 0 \leq \bar{t} < t)\Big) \ ,$$

where $\zeta : T^*M \times \mathbb{R} \to M$ is the flow of X.

The above statement is clearly constructed starting from formulae that define real-analytic or semianalytic sets and applying the various logical connectives of first-order predicate calculus, namely, "and," "or," negation, and the existential quantifier. These connectives correspond to set-theoretic operations (intersection, union, complementation, projection) that preserve the class of subanalytic sets, provided only that every time a projection is taken this projection happens to be proper on the closure of the set being projected. In our case, this amounts to ensuring that the quantifier $(\exists\lambda)$ is bounded, in the sense that we can replace $(\exists\lambda)$ by $(\exists\lambda \in K)$, K being a compact set. Now, λ basically has three components $\lambda_i = \langle\lambda, Y_i\rangle$, where Y_1 and Y_2 are, as above, an orthonormal basis of sections of E, and $Y_3 = [Y_1, Y_2]$. As indicated above, for a minimizing arc the controls u_i satisfy $u_1^2 + u_2^2 = 1$ and $u_i = -\langle\lambda, Y_i\rangle$, which gives an *a priori* bound on the components λ_1 and λ_2. The main problem is then to get an *a priori* bound for λ_3. This cannot be done by just invoking the Maximum Principle, for this result allows extremals with arbitrarily large λ_3. However, we show that, if λ_3 is too large, then the corresponding trajectory cannot be optimal on a sufficiently large time-interval. (The heuristic reason for this is that, as we show below, the control vector $u = (u_1, u_2)$ rotates with angular velocity λ_3. If λ_3 is too large, then the control oscillates very fast, so that the resulting trajectory is close to a constant, and therefore not optimal. A rigorous argument is outlined below.) As ε becomes small, the bound on λ_3 increases. This is the reason why we cannot prove subanalyticity of V_p all the way up to p.

Remark 1.4. Naturally, the same idea works for general real-analytic Riemannian metrics, where everything is easier, because the equations for the extremals yield a bound for all the components of λ. In that case, there is no extra component such as λ_3 to worry about, and one gets subanalyticity up to p (a fact which in any case is obvious directly) and global subanalyticity of V_p if the metric is complete. ∎

Remark 1.5. The technical obstructions that prevent an extension of our results to more general real-analytic sub-Riemannian structures are: (1) the fact that, for more general metrics, abnormal trajectories exist, and (2) the difficulty with getting bounds for the components of the covector λ in directions transversal to those of E. ∎

§2. Outline of the proof .

We first remark that the infimum defining $d_{E,g}(p, q)$ remains the same if we just limit ourselves to E-admissible trajectories that are γ *parametrized by arc-length*, i.e. are such that $g(\gamma'(t), \gamma'(t)) \equiv 1$. For such a γ, the length $|\gamma|$ is just the *time along* γ.

We then show that, if $q \in B(p, T)$, then there is an optimal (i.e time-minimizing) trajectory from p to q, i.e. an E-admissible $\gamma : [0, V_p(q)] \to M$, parametrized by arc-length, such that $\gamma(0) = p$, $\gamma(V_p(q)) = q$. First extend g to a true Riemannian metric $\hat{g}$ on M. Next, take a minimizing sequence, i.e. a sequence $\{\gamma_\nu\}$ of E-admissible curves $\gamma_\nu : [0, V_p(q) + \varepsilon_\nu] \to M$, parametrized by arc-length, such that $\gamma(0) = p$, $\varepsilon_\nu > 0$, $\varepsilon_\nu \to 0$ as $\nu \to \infty$, and $\gamma(V_p(q) + \varepsilon_\nu) = q$. If we let $\tilde{\gamma}_\nu$ denote the restriction of γ_ν to the interval $[0, V_p(q)]$, then $\tilde{\gamma}_\nu$ is entirely contained in $B(p, T)$. The arcs γ_ν satisfy $||\gamma_\nu'(t)|| = g(\gamma_\nu'(t), \gamma_\nu'(t))^{1/2} = 1$, where $||v||$ denotes the Riemannian length of a tangent vector v. So the sequence $\tilde{\gamma}_\nu$ is equi-Lipschitz and therefore, after passing to a subsequence, we may assume that $\tilde{\gamma}_\nu$ converges uniformly to a Lipschitz curve $\gamma : [0, V_p(q)] \to M$. Clearly, $\gamma(V_p(q)) = q$, so all we need is to show that γ is E-admissible and parametrized by arc-length. Since γ is Lipschitz, it suffices to show that, at every point $t \in [0, V_p(q)]$ where γ is differentiable, the vector $\gamma'(t)$ is in $E(\gamma(t))$ and satisfies $||\gamma'(t)|| = 1$. This follows by a standard convexity argument. Let $\tilde{p} = \gamma(t)$. Choose coordinates on a neighborhood U of $\tilde{p}$, so that we can regard all the spaces $E(x)$, $x \in U$, as subspaces of $\mathbb{R}^3 \sim T_{\tilde{p}}M$. Now forget about $\hat{g}$, and use $||v||$ to denote the usual Euclidean length of v, if $v \in E(x)$, $x \in U$, and $E(x)$ is embedded in $\mathbb{R}^3$ as above. By shrinking U, if necessary, we may assume that there is a constant $C > 0$ such that $||v|| \le C|||v|||$ whenever $v \in E(x)$, $x \in U$, where $|||v||| = g(v,v)^{1/2}$. Assume that, with respect to our coordinates, U is the unit cube $]-1,1[^3$. Let $B(x)$ denote the unit ball in $E(x)$ (relative to g), so that, for $x \in U$, we can regard $B(x)$ as a subset of $T_{\tilde{p}}M$. Let B^ε denote the closed convex hull of the union of the $B(x)$, taken over all $x \in U$ such that $||x - \tilde{p}|| \le \varepsilon$. Then B^ε is compact if $0 < \varepsilon < 1$, and $\bigcap_{\varepsilon>0} B^\varepsilon = B(\tilde{p})$. Fix ε such that $0 < \varepsilon < 1$, and then choose $h < \frac{\varepsilon}{2C}$. Choose $\bar{\nu}$ so large that $||\gamma_\nu(t) - \tilde{p}|| \le \frac{\varepsilon}{2}$ for $\nu \ge \bar{\nu}$. Then, if $|s - t| \le h$ and $\nu \ge \bar{\nu}$, we have $\gamma_\nu(s) - \gamma_\nu(t) = \int_t^s \gamma_\nu'(\tau)d\tau$, so that $||\gamma_\nu(s) - \gamma_\nu(t)|| \le C|s-t| \le Ch \le \frac{\varepsilon}{2}$, provided that $\gamma_\nu(\tau) \in U$ for all τ between t and s. In particular, this gives $||\gamma_\nu(s) - \tilde{p}|| \le \varepsilon$ if $\gamma_\nu(\tau) \in U$ for τ between t and s. Since $\varepsilon < 1$, this implies that in fact $\gamma_\nu(s) \in U$, and $||\gamma_\nu(s) - \tilde{p}|| \le \varepsilon$, for all s such that $|s - t| \le h$. Therefore, for all such s we have $\gamma_\nu'(s) \in B^\varepsilon$. Integrating with respect to s from t to $t+h$ and dividing by h we conclude that $h^{-1}(\gamma_\nu(t+h) - \gamma_\nu(t)) \in B^\varepsilon$. Letting $\nu \to \infty$ we get $h^{-1}(\gamma(t+h) - \gamma(t)) \in B^\varepsilon$. Letting $h \to 0$ we get $\gamma'(t) \in B^\varepsilon$. Since this is true for all $\varepsilon > 0$, we get $\gamma'(t) \in B(\tilde{p})$. So, for every t where γ is differentiable, we have shown that $\gamma'(t) \in E(\gamma(t))$ and $|||\gamma'(t)||| \le 1$. In particular, γ is E-admissible. On the other hand, if the inequality $|||\gamma'(t)||| \le 1$ was strict on a subset of $[0, V_p(q)]$ of positive measure, it would follow that $|\gamma| < V_p(q)$, contradicting the fact that $V_p(q)$ is the infimum of the lengths of all the arcs from p to q. So $|||\gamma'(t)||| = 1$ for almost all t, and γ is parametrized by arc-length.

Now let γ be an optimal trajectory from p to q. We will apply the Maximum Principle (MP) to obtain information about γ. The principle is usually stated (cf. [1], [7], [9]) as a necessary condition for optimality of a trajectory $\gamma : [a,b] \to \Omega$ of a control system of the form $x' = f(x,u)$ corresponding to a control $u(\cdot) : [a,b] \to U$. Here x takes values in an open subset Ω of $\mathbb{R}^n$, U is a closed subset of $\mathbb{R}^m$, and $f : \Omega \times U \to \mathbb{R}^n$ is smooth. "Optimality" means that the 4-tuple $(a, b, \gamma, u(\cdot))$ minimizes the integral $J(\bar{a}, \bar{b}, \bar{\gamma}, \bar{u}(\cdot)) = \int_{\bar{a}}^{\bar{b}} L(\bar{\gamma}(s), \bar{u}(s))ds$ among all admissible 4-tuples $(\bar{a}, \bar{b}, \bar{\gamma}, \bar{u}(\cdot))$ such that $\bar{\gamma}(\bar{a}) = \gamma(a)$ and $\bar{\gamma}(\bar{b}) = \gamma(b)$. An "admissible 4-tuple" is a 4-tuple $(a, b, \gamma, u(\cdot))$ such that $a \in \mathbb{R}$, $b \in \mathbb{R}$, $a \le b$, $u(\cdot)$ is a bounded U-valued measurable function on $[a,b]$ and $\gamma : [a,b] \to M$ is absolutely continuous and satisfies $\gamma'(t) = f(\gamma(t), u(t))$ for almost every $t \in [a,b]$. The "Lagrangian" $(x,u) \to L(x,u)$ is supposed to be of class C^1 in x and u. The MP says that, if $(a, b, \gamma, u(\cdot))$ is optimal, then there exist an absolutely continuous function $\lambda : [a,b] \to \mathbb{R}^n$ and a real number $\lambda_0 \geq 0$ such that (i) $(\lambda(t), \lambda_0) \neq (0,0)$ for $t \in [a,b]$, (ii) $\lambda'(t) = -D_\lambda \mathcal{H}(\lambda(t), \gamma(t), \lambda_0, u(t))$ for almost all $t \in [a,b]$ (the "adjoint equation"), and (iii) $\mathcal{H}(\lambda(t), \gamma(t), \lambda_0, u(t)) = \min_{v \in U} \mathcal{H}(\lambda(t), \gamma(t), \lambda_0, v) = 0$ for almost all $t \in [a,b]$ (the "hamiltonian minimization condition"). Here $\mathcal{H} : \mathbb{R}^n \times \Omega \times \mathbb{R} \times U \to \mathbb{R}^n$ is the *control theory Hamiltonian*, defined by $\mathcal{H}(\lambda, x, \lambda_0, v) = \langle \lambda, f(x,u) \rangle + \lambda_0 L(x,u)$. (An intrinsic formulation on manifolds is possible, cf. [21], with f being a family of vector fields on a manifold M depending on a parameter $u \in U$. Then λ must be regarded as a covector at x, so $\mathcal{H}$ is defined on $T^*M \times \mathbb{R} \times U$. We will not need this formulation here.)

To apply the MP to our situation, we first define a function $H : T^*M \to R$ by letting, for $\lambda \in T^*_x M$,

$$H(\lambda, x) = -\frac{1}{2}\langle \lambda, v_1 \rangle^2 - \frac{1}{2}\langle \lambda, v_2 \rangle^2 \ ,$$

where $v = (v_1, v_2)$ is an arbitrary orthonormal basis of $T_x M$. (It is clear that $H(\lambda, x)$ does not depend on the choice of v.) Then H is real-analytic. Let X be the Hamilton vector field associated to H. Then the MP implies:

Lemma 2.1. *Every minimizing trajectory of our sub-Riemannian metric is the projection on M of an integral curve of X along which $H = -\frac{1}{2}$.*

PROOF. First proceed locally. If γ is contained in an open set Ω which is the domain of a cubic coordinate chart and is such that on Ω the bundle E admits two linearly independent real-analytic sections Y_1 and Y_2, then we can in fact choose Y_1 and Y_2 so that $Y_1(x)$ and $Y_2(x)$ form an orthonormal basis at each point x. An optimal trajectory $\gamma : [a,b] \to \Omega$ is then a minimum-time trajectory (i.e. an optimal trajectory for $L \equiv 1$) for the control problem $x' = u_1 Y_1(x) + u_2 Y_2(x)$, where the control $u = (u_1, u_2)$ takes values in the disc $\Delta = \{(u_1, u_2) : u_1^2 + u^2 \leq 1\}$. Let $u(\cdot) = (u_1(\cdot), u_2(\cdot))$ be the corresponding control. The MP then implies that there exists a $\lambda : [a,b] \to \mathbb{R}^3 - \{0\}$ that satisfies for almost all $t \in [a,b]$, the adjoint equation

$$\lambda'(t) = -u_1(t)\langle \lambda(t), D_x Y_1(\gamma(t)) \rangle - u_2(t)\langle \lambda(t), D_x Y_2(\gamma(t)) \rangle \ ,$$

as well as the condition

$$u_1(t)\varphi_1(t) + u_2(t)\varphi_2(t) = \min\{v_1\varphi_1(t) + v_2\varphi_2(t) : (v_1, v_2) \in \Delta\} \ ,$$

$u_1(t)\varphi_1(t) + u_2(t)\varphi_2(t) = \min\{v_1\varphi_1(t) + v_2\varphi_2(t) : (v_1, v_2) \in \Delta\}$, ("hamiltonian minimization") where $\varphi_i(t) = \langle \lambda(t), Y_i(\gamma(t)) \rangle$ for $i = 1, 2$. If for some t the numbers $\varphi_i(t)$ do not both vanish, then the minimum is attained at only one value of v, namely, $v_i = -\rho(t)^{-1}\varphi_i(t)$, where $\rho(t) = (\varphi_1(t)^2 + \varphi_2(t)^2)^{1/2}$. So γ and λ satisfy

$$\gamma' = -\rho^{-1}\sum_{i=1}^{2}\langle \lambda, Y_i \rangle Y_i \ ,$$

$$\lambda' = \rho^{-1}\sum_{i=1}^{2}\langle \lambda, Y_i \rangle \langle \lambda, D_x Y_i \rangle \ .$$

This shows in particular that, modulo a reparametrization, the curve $t \to (\lambda(t), \gamma(t))$ is an integral curve of X. In particular, H is constant on this curve, so $\rho = (2H)^{1/2}$ is constant as well. So ρ is constant in a neighborhood of every t such that $\rho(t) \neq 0$. Since ρ is continuous, it follows that ρ is constant on the whole interval $[a, b]$. Multiplying $\lambda(t)$ by an appropriate constant, we can then make sure that $\rho = 1$. Then (λ, γ) is a trajectory of X, along which $H \equiv -\frac{1}{2}$.

The preceding argument also shows that, if the $\varphi_i(t)$ vanish for one value of t, then they vanish identically. A trajectory for which $\varphi_1(t) \equiv \varphi_2(t) \equiv 0$ will be called "abnormal." If γ is abnormal, then a simple computation shows that the derivatives $\psi_i = \varphi_i'$ are given by $\psi_1(t) = u_2(t)\langle\lambda(t), [Y_2, Y_1](\gamma(t))\rangle$, $\psi_2(t) = u_1(t)\langle\lambda(t), [Y_1, Y_2](\gamma(t))\rangle$. Since the three vectors $Y_1(\gamma(t))$, $Y_2(\gamma(t))$, $[Y_1, Y_2](\gamma(t))$ span $T_{\gamma(t)}M$, and the covector $\lambda(t)$ satisfies $\langle\lambda(t), Y_1(\gamma(t))\rangle = \langle\lambda(t), Y_2(\gamma(t))\rangle = 0$ but $\lambda(t) \neq 0$, it follows that $\langle\lambda(t), [Y_1, Y_2](\gamma(t))\rangle \neq 0$. But then $u_1(t) \equiv u_2(t) \equiv 0$. So $t \to \gamma(t)$ is in fact a constant curve, and it is obvious that such a curve cannot be time-minimizing. So there are no abnormal minimizing trajectories.

Hence, locally, every minimizing trajectory is a projection of an integral curve of X along which $H \equiv -\frac{1}{2}$, as stated. Moreover, we claim that the covector-valued function $t \to \lambda(t)$ is in fact unique. Indeed, as shown before, the numbers $\varphi_i(t) = \langle\lambda(t), Y_i(\gamma(t))\rangle$ are uniquely determined as the negatives of the controls $u_i(t)$, which are the coefficients of the expression of $\gamma'(t)$ as a linear combination of $Y_1(\gamma(t))$ and $Y_2(\gamma(t))$. And then the derivatives $\psi_1(t) = u_2(t)\langle\lambda(t), [Y_2, Y_1](\gamma(t))\rangle$ and $\psi_2(t) = u_1(t)\langle\lambda(t), [Y_1, Y_2](\gamma(t))\rangle$ are determined as well. Since $u_1(t)$ and $u_2(t)$ cannot both vanish, it follows the quantity $\langle\lambda(t), [Y_1, Y_2](\gamma(t))\rangle$ is determined as well. So $\lambda(t)$ is uniquely determined. In view of the uniqueness, the covector-valued functions $\lambda(t)$ that exist for sufficiently small t-intervals can be patched together and the result is true globally. ∎

Now let $v = (v_1, v_2)$ be an orthonormal basis of $E(p)$, and let Y_1, Y_2 be real-analytic sections of E on a neighborhood Ω of p such that $Y_i(p) = v_i$ for $i = 1, 2$, and $Y_1(x), Y_2(x)$ form an orthonormal basis of $E(x)$ for each $x \in \Omega$. Let $Y_3 = [Y_1, Y_2]$, $v_3 = Y_3(p)$. Then an optimal trajectory $\gamma : [0, b] \to M$ such that $\gamma(0) = p$ is uniquely specified by the three numbers $\lambda_i = \langle\lambda(0), v_i\rangle$, $i = 1, 2, 3$. These numbers must satisfy $\lambda_1^2 + \lambda_2^2 = 1$. The main technical observation is the following

Lemma 2.2. *For every $\varepsilon >$ there exists an $A(\varepsilon) > 0$ such that, if $|\lambda_3| > A(\varepsilon)$, then the corresponding trajectory γ is not optimal on the interval $[0, \varepsilon]$.*

PROOF. Is clearly suffices to prove the result for ε sufficiently small. So we may assume that ε is so small that all the trajectories under consideration stay in a compact subset $\mathcal{K}$ of Ω up to time ε. Moreover, we will only consider the case $\lambda_3 > 0$, the other case being similar.

Define the functions φ_i as above, for $i = 1, 2$. Define $\varphi_3(t) = \langle\lambda(t), Y_3(\gamma(t))\rangle$. Then we can write the vector fields $[Y_1, Y_3]$, $[Y_2, Y_3]$ as linear combinations of Y_1, Y_2, Y_3 with analytic coefficients: $[Y_i, Y_3] = \sum_{j=1}^{3} \alpha_i^j Y_j$. Then the φ_i satisfy the system of differential equations

$$\varphi_1' = -\varphi_2\varphi_3 \ , \ \varphi_2' = \varphi_1\varphi_3 \ , \ \varphi_3' = \sum_{i=1}^{2}\sum_{j=1}^{3} \alpha_i^j \varphi_i\varphi_j \ .$$

Since the α_i^j are bounded on $\mathcal{K}$, and $|\varphi_i| \leq 1$ for $i = 1, 2$, there exists a fixed constant K such that the bound $|\varphi_3'(t)| \leq K(1 + |\varphi_3(t)|)$ holds for all optimal trajectories in $\mathcal{K}$. Gronwall's inequality then implies the bound $|\varphi_3(t) - \varphi_3(0)| \leq (1 + |\varphi_3(0)|)(e^{Kt} - 1)$, from which it follows that, if $A > 0$, then $|\varphi_3(t)| > A$ throughout the interval $[0, \varepsilon]$, if $|\varphi_3(0)| > 2A$, provided that ε is so small that $(1 + 2A)(e^{K\varepsilon} - 1) < A$. (If ε is so small that $e^{K\varepsilon} - 1 < \frac{1}{3}$, then it suffices to take $A > 1$.) If $\varphi_3(0) > 2A$, then the equations $\varphi_1' = -\varphi_2\varphi_3$, $\varphi_2' = \varphi_1\varphi_3$, imply that the column vector $\Phi(t) = (\varphi_1(t), \varphi_2(t))^\dagger$ satisfies $\Phi(t) = R(\sigma(t))\Phi(0)$, where $\sigma(t) = \int_0^t \varphi_3(s)ds$, and $R(s)$ is the matrix of a counterclockwise rotation in the plane by an angle s. It then follows that

$$\int_0^t \Phi(s)ds = \int_0^t R(\sigma(s))\Phi(0)ds = \int_0^{\sigma(t)} R(\sigma)\Phi(0)\theta(\sigma)d\sigma \ ,$$

where $\theta(\sigma(s))\varphi_3(s) = 1$. From this we get a bound $||\int_0^t \Phi(s)ds|| \leq \frac{C}{A}$, where C is a fixed constant. (Indeed, we can compute $\int_0^{\sigma(t)} R(\sigma)\theta(\sigma)d\sigma$ by parts. The result is $\theta(\sigma(t))\int_0^{\sigma(t)} R(\sigma)d\sigma - \int_0^{\sigma(t)} \frac{d\theta}{d\sigma}(\sigma)\int_0^\sigma R(s)dsd\sigma$. The function $\sigma \to \int_0^\sigma R(s)ds$ is uniformly bounded for $\sigma \in \mathbb{R}$. On the other hand, $\theta(\sigma(t)) = \frac{1}{\varphi_3(t)} < \frac{1}{A}$. And $\frac{d\theta}{d\sigma}$ is $O(A^{-2})$, because $\frac{d\theta}{ds} = \frac{d}{ds}(\varphi_3(s)^{-1}) = -\frac{\varphi_3'(s)}{\varphi_3(s)^2} = O(A^{-1})$ (recall that $\varphi_3' = O(1 + |\varphi_3|)$), and $\frac{ds}{d\sigma} = \varphi_3(s)^{-1} = O(A^{-1})$. So $\int_0^{\sigma(t)} \frac{d\theta}{d\sigma}(\sigma)\int_0^\sigma R(s)dsd\sigma = O(\sigma(t)A^{-2}) = O(A^{-1})$.)

The trajectory γ satisfies

$$\gamma(t) = p - \sum_{i=1}^{2}\int_0^t \varphi_i(s)Y_i(\gamma(s))ds \ .$$

Write

$$Y_i(\gamma(s)) = Y_i(p) - \sum_{j=1}^{2}\int_0^s \varphi_j(\tau)(D_xY_i)Y_j(\gamma(\tau))d\tau \ .$$

Then

$$\gamma(t) = p - \sum_{i=1}^{2}\int_0^t \varphi_i(s)Y_i(p)ds + \sum_{i=1}^{2}\sum_{j=1}^{2}\int_0^t\int_0^s \varphi_i(s)\varphi_j(\tau)(D_xY_i)Y_j(\gamma(\tau))d\tau ds \ .$$

Clearly,

$$(D_xY_i)Y_j(\gamma(\tau)) = (D_xY_i)Y_j(p) + O(\varepsilon) \ .$$

So

$$\gamma(t) = p - \sum_{i=1}^{2}\int_0^t \varphi_i(s)Y_i(p)ds + \sum_{i=1}^{2}\sum_{j=1}^{2}\int_0^t\int_0^s \varphi_i(s)\varphi_j(\tau)(D_xY_i)Y_j(p)d\tau ds + O(\varepsilon^3)\ .$$

Since $\int_0^s \varphi_i(\tau)d\tau = O(A^{-1})$, we get the bound $||\gamma(t)-p|| = O(A^{-1}) + O(\varepsilon^3)$. If we then choose $A > c\varepsilon^{-3}$, where c is a suitable constant, we get $||\gamma(\varepsilon)-p|| = O(\varepsilon^3)$. On the other hand, it is well known (cf. [5], [38]) that the Euclidean distance and the metric $d_{E,g}$ satisfy, near p, an inequality $d_{E,g}(q,p) \le C||q-p||^{1/2}$. This implies in particular that $d_{E,g}(\gamma(\varepsilon),p) = O(\varepsilon^{3/2}) = o(\varepsilon)$. So, if ε is sufficiently small, $\gamma(\varepsilon)$ can be reached from p in time much smaller than ε, whereas γ itself goes from p to q in time ε. So γ is not optimal, and the proof of the lemma is complete. ∎

We now pick an $\varepsilon > 0$ and choose an $A > 0$ with the property that, if $t \to (\gamma(t),\lambda(t))$ is a trajectory of X that corresponds to initial conditions λ_i such that $\lambda_1^2 + \lambda_1^2 = 1$, $|\lambda_3| > A$, then γ is not optimal on $[0,\varepsilon]$. Let $S = G_{\varepsilon,T}(V_p)$. Let $K(\varepsilon)$ be the compact subset of T_p^*M consisting of those λ such that $\langle\lambda,v_1\rangle^2 + \langle\lambda,v_2\rangle^2 = 1$ and $|\lambda_3| \le A$. Then $(x,t) \in S$ if and only if (i) $\varepsilon \le t \le T$, and (ii) there exists $\lambda \in K(\varepsilon)$ such that $\pi(\zeta(\lambda,p,t)) = x$ but there do not exist an s and a λ such that $0 < s < t$, $\lambda \in K(\varepsilon)$, and $\pi(\zeta(\lambda,p,s)) = x$. Here ζ is the flow of X, which in particular is a real-analytic map on some open subset N of $T^*M \times \mathbb{R}$ containing $K(\varepsilon)\times\{p\}\times[0,T]$, and $\pi : T^*M \to M$ denotes the canonical projection. The set $S_1 = \{(\lambda,q,t,x) \in N \times M : \pi(\zeta(\lambda,q,t)) = x\}$ is an analytic subset of $N \times M$. The intersection S_2 of S_1 with $K(\varepsilon)\times\{p\}\times]0,T]$ is semianalytic in $N \times M$. Also, the set S_3 of those $(\lambda,q,t,\bar\lambda,\bar q,\bar t,x)$ such that $\pi(\zeta(\lambda,q,t)) = \pi(\zeta(\bar\lambda,\bar q,\bar t)) = x$ is analytic in $N\times N\times M$. Its intersection S_4 with the set $K(\varepsilon)\times\{p\}\times\mathbb{R}\times K(\varepsilon)\times\{p\}\times\mathbb{R}\times M$ is semianalytic in $N\times N\times M$. If we intersect with the set $\{(\lambda,q,t,\bar\lambda,\bar q,\bar t,x) : 0<\bar t<t\le T\}$ we get another semianalytic set S_5. Let μ be the map $(\lambda,q,t,\bar\lambda,\bar q,\bar t,x) \to (\lambda,q,t,x)$. Then μ maps S_5 to a subanalytic subset of S_2, because μ is analytic and proper on the closure of S_5. The set $S_6 = S_2 - \mu(S_5)$ is therefore subanalytic in $N \times M$. Let $\xi(\lambda,q,t,x) = (x,t)$. Then $\xi(S_6) = S$. Since ξ is analytic and proper on the closure of S_6, we see that S is a subanalytic subset of $M \times \mathbb{R}$. Our proof is now complete. ∎

REFERENCES

[1] L.D. Berkovitz, *Optimal Control Theory*, Springer-Verlag, New York, 1974.

[2] R.W. Brockett, "Asymptotic Stability and Feedback Stabilization," in *Differential Geometric Control Theory* (R.W. Brockett, R.S. Millman and H.J. Sussmann, eds.), Birkhäuser, Boston, 181-191, 1983.

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

[4] R.W. Brockett, "Control Theory and Singular Riemannian Geometry," in *New Directions in Applied Mathematics* (P.J. Hilton and G.S. Young, eds.), Springer-Verlag, (1981).

[5] V.Y. Gershkovich, "Two-sided Estimates of Metrics Generated by Absolutely Nonholonomic Distributions on Riemannian Manifolds," *Soviet Math. Dokl.*, 30 (2), 1984.

[6] G. Lafferriere and H.J Sussmann, "Motion Planning for Controllable Systems Without Drift: a Preliminary Report," *Technical Report Sycon 90-04*, Rutgers Univ., June 1990.

[7] E.B. Lee and L. Markus, *Foundations of Optimal Control Theory*, Wiley, New York, 1968.

[8] R. Montgomery, "Geodesics which do not Satisfy the Geodesic Equations," 1991 preprint.

[9] L.S. Pontryagin, V.G. Boltianskii, R.V. Gamkrelidze and E.F. Mischenko, *The Mathematical Theory of Optimal Processes*, Wiley, New York, 1962.

[10] S. Sastry and Z. Li, "Robot Motion Planning with Nonholonomic Constraints," *Proc. 28th IEEE CDC, Tampa, Fla.*, 211-216, IEEE Publications, 1989.

[11] H. Schättler and H.J. Sussmann, "On the regularity of optimal controls," *Journal of Applied Mathematics and Physics (ZAMP)*, **38**, March 1987, pp.292-301.

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

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

[14] H.J. Sussmann, "Piecewise analyticity of optimal cost functions and optimal feedback," in *Proc. Joint Automatic Control Conference*, Denver, Colo. (August 1979).

[15] H.J. Sussmann, "A bang-bang theorem with bounds on the number of switchings," *S.I.A.M. J. Control and Optimization* **17**, No. 5 (1979), pp. 629-651.

[16] H.J. Sussmann, "Les semigroupes sousanalytiques et la régularité des commandes en boucle fermée," *Astérisque (Soc. Math. de France)* **75-76** (1980), pp. 219-226.

[17] H.J. Sussmann, "Analytic stratifications and control theory," in *Proc. 1978. Int. Congress of Mathematicians*, Helsinki (1980), pp. 865-871.

[18] H.J. Sussmann, "Bounds on the number of switchings for trajectories of piecewise analytic vector fields," J. Diff. Equations **43** (1982), pp. 399-418.

[19] H.J. Sussmann, "Time-optimal control in the plane," in *Feedback Control of Linear and Nonlinear Systems*, Proceedings of the Joint Workshop on Feedback and Synthesis of Linear and Nonlinear Systems, Bielefeld-Rome 1981, D. Hinrichsen and A. Isidori Eds., Springer-Verlag Lect. Notes in Control and Information Sciences, No. 39, New York (1982), pp. 244-260.

[20] H.J. Sussmann, "Subanalytic sets and optimal control in the plane," in *Proc. 21st I.E.E.E. Conference on Decision and Control*, Orlando, Fla. (Dec 1982), pp. 295-299.

[21] H.J. Sussmann, *Lie brackets, Real Analyticity and Geometric Control*, in *Differential Geometric Control Theory* (R.W. Brockett, R.S. Millman and H.J. Sussmann, eds.), Birkhäuser, Boston, 1-116, 1983.

[22] H.J. Sussmann, "Lie Brackets and real analyticity in control theory," in *Mathematical Control Theory*, C. Olech Ed., Banach Center Publications, Volume 14, PWN-Polish Scientific Publishers,Warsaw, Poland, 1985, pp. 515-542.

[23] H.J. Sussmann, "Resolution of singularities and linear time-optimal control," in *Proceedings of the 23rd I.E.E.E. Conference on Decision and Control*, Las Vegas, Nevada (Dec. 1984), pp. 1043-1046.

[24] H.J. Sussmann, "Some recent results on the regularity of optimal cost functions," in *Proceedings of the Berkeley-Ames Conference on Nonlinear Problems in Control and Fluid Dynamics*, L.R. Hunt and C.F. Martin Eds., Math. Sci. Press, Brookline, Mass, 1984, pp.429-434.

[25] H.J. Sussmann, "The structure of time-optimal trajectories for single-input systems in the plane : the C^∞ nonsingular case," *S.I.A.M. J. Control and Optimization*, **25**, No. 2 (1987), pp. 433-465.

[26] H.J. Sussmann, "The structure of time-optimal trajectories for single-input systems in the plane: the general real-analytic case," *S.I.A.M. J. Control and Optimization* **25**, No. 4, 1987,pp. 868-904.

[27] H.J. Sussmann, "Regular synthesis for time-optimal control of single-input real-analytic systems in the plane," *S.I.A.M. J. Control and Optimization* **25**, No. 5, 1987, pp. 1145-1162.

[28] H.J. Sussmann, "Envelopes, conjugate points and optimal bang-bang extremals," in *Algebraic and Geometric Methods in Nonlinear Control Theory*, M. Fliess and M. Hazewinkel Eds., D. Reidel Publishing Co., Dordrecht, The Netherlands, 1986, pp.325-346.

[29] H.J. Sussmann, "A weak regularity theorem for real analytic optimal control problems," *Revista Matemática Iberoamericana* **2**, No. 3, 1986, pp. 307-317.

[30] H.J. Sussmann, "Recent developments in the regularity theory of optimal trajectories," in *Linear and Nonlinear Mathematical Control Theory*, Rendiconti del Seminario Matematico, Università e Politecnico di Torino, Fascicolo Speciale 1987, pp. 149-182.

[31] H.J. Sussmann, "Trajectory regularity and real analyticity: some recent results," in *Proceedings of the 25th I.E.E.E. Conference on Decision and Control*, Athens, Greece, Dec. 1986, pp. 592-595.

[32] H.J. Sussmann, "A new regularity theorem for bang-bang trajectories," *Proceedings of the Conference on Nonlinear Control*, Nantes, France, June 1988.

[33] H.J. Sussmann, "Hamiltonian lifts and optimal trajectories," *Proceedings of the 27th I.E.E.E. Conference on Decision and Control*, Austin, Texas, Dec. 1988, pp. 1182-1186.

[34] H.J. Sussmann, "Envelopes, Lie brackets, and high-order necessary conditions for optimality," *Proceedings of the 28th I.E.E.E. Conference on Decision and Control*, Tampa, Florida, Dec. 1989, pp. 1107-1112.

[35] H.J. Sussmann, "Synthesis, presynthesis, sufficient conditions for optimality, and subanalytic sets," in "Nonlinear Controllability and Optimal Control," H. J. Sussmann Ed., M. Dekker, Inc., New York, 1990, pp. 1-20.

[36] H.J. Sussmann, "Real Analytic Desingularization and Subanalytic Sets: an Elementary Approach," *Trans. Amer. Math. Soc.* **317** (1990), pp. 417-461.

[37] A.M. Vershik and V. Ya Gershkovich, "Non-holonomic Riemannian Manifolds," in *Dynamical Systems* v. 7, part of the Mathematical Encyclopaedia series, vol.16 (in Russian), MIR pub., 1988.

[38] A.M. Vershik and V. Ya Gershkovich, "Non-holonomic Manifolds and Nilpotent Analysis," *J. Geom. and Phys.* vol. 5, no. 3, (1988).

Héctor J. Sussmann
Department of Mathematics
Rutgers University
New Brunswick, NJ 08903

U.S.A.

E-mail: sussmann@hamilton.rutgers.edu