André:

Here is the product of my latest efforts on the paper. There still are some missing things:

1. The estimate of Proposition 3.12 is crucial for the proof of our main theorem. I think that in the references I give there —to work by Gershkovich and Vershik— the estimate is proved only for the case of a nonsingular distribution. (I confess that I have *not* looked at the Gershkovich-Vershik papers.) Of course, the estimate is also valid for singular distributions, and the proof is fairly trivial —though a little bit long— and I am sure there are references where it is done. Could you include such a reference? (For example, it's probably done in one of your papers. If so, then we should put that in.)
2. Add references on subanalyticity of the distance function in the Riemannian case, if you can find any.
3. Add mention of stuff by Agrachev and Sarychev on subanalyticity for the 2-generating non SBG case.
4. Also, I guess there is work by Sebastien Jacquet. His e-mail address is

   Sebastien.Jacquet@univ-savoie.fr

   You should try to get a copy and include a mention of it.
5. Include an example of a nonsubanalytic distance function for a nonconvex Riemannian manifold (the plane minus a sequence of holes) if you think this can be done without making it too long.
6. Give reference (e.g. Dixmier) on metric spaces where d can have infinite values.
7. Put your complete address in the front page.
8. If you want to put a footnote with some acknowledgement, do it.
9. Add other bibliographical references that you might think are needed, for example to your own papers.

In addition, you should **proofread the whole paper several times**, to make sure we get rid of mathematical mistakes and typographical errors.

Do you think you could finish the work, say, within one month from today, by **August 17** or so?

Also, would you be willing to take care of all the work from now on (sending the paper to the journal, dealing with the editors, etc.)?

Cordially,

—Hector

ADJOINT VECTOR ESTIMATES, CONVEXITY, AND SUBANALYTICITY OF THE DISTANCE FUNCTION FOR STRONGLY BRACKET-GENERATING SUB-RIEMANNIAN METRICS

Héctor J. Sussmann*

Department of Mathematics
Rutgers University
Piscataway, NJ 08854, USA

sussmann@hamilton.rutgers.edu
http://www.math.rutgers.edu/~sussmann

André Bellaïche

Université de Paris VII
Paris, France

abellaic@math.jussieu.fr

July 1998

ABSTRACT

We prove that, for a smooth strongly bracket-generating subriemannian structure on a smooth manifold M, the set $\mathcal{G}_{nc,min}(K)$ of all curves $\Xi : [0,1] \to T^*M$ in the cotangent bundle T^*M that are solutions of the "geodesic equation" and project down to a nonconstant length minimizer contained in a given compact set K is itself relatively compact. As a consequence, we show that, if $\bar{K}$ and $\hat{K}$ are compact subsets of M, and U is a neighborhood of $\bar{K}$ such that for every pair $(x,y) \in U \times \hat{K}$ there exists a length minimizer joining x to y, then there exists a compact set $\mathcal{K}$ of solutions $\Xi : [0,1] \to T^*M$ of the geodesic equation such that every length minimizer $\xi : [0,1] \to M$ for which $\partial\xi = (\xi(0), \xi(1)) \in \bar{K} \times \hat{K}$ is the projection of a $\Xi \in \mathcal{K}$. In the real-analytic case, this implies that the distance function $d : M \times M \to \mathbb{R}$ is subanalytic if M is convex (in the sense that any two points of M can be joined by a length minimizer) and, more generally, that the restriction $d \lceil \Omega$ of d to Ω is subanalytic in Ω —that is, the graph of $d \lceil \Omega$ is a subanalytic subset of $\Omega \times \mathbb{R}$— if Ω is an open subset of $M \times M$ such that for every $(\bar{x}, \hat{x}) \in \Omega$ there exists a length minimizer from $\bar{x}$ to $\hat{x}$. In particular, d is subanalytic on a neighborhood of the diagonal, and d is subanalytic on $U \times U$ if U is any open subset of M such that any two points of U can be joined in M by a length minimizer.

* Supported in part by NSF Grant DMS95-00798 and AFOSR Grant 0923. The final part of the work was carried out between January and June 1998, while this author was a visitor at the Institut Henri Poincaré (IHP), in Paris. He is grateful to the IHP for its kind hospitality during this period.

§1. Introduction.

The concept of a "strongly bracket-generating" (SBG) subriemannian structure was introduced by R. Strichartz in [29] and [30], in order to single out a natural class of subriemannian manifolds for which nonconstant abnormal minimizers do not occur.

The main purpose of this paper is to prove the following estimate on the fields of covectors corresponding to energy-minimizing arcs in a strongly bracket-generating subriemannian manifold $\mathcal{M}$: if K is a compact set of points of $\mathcal{M}$, and $\mathcal{G}_{nc,min}(K)$ is the set of all solutions Ξ of the geodesic equation for $\mathcal{M}$ that are defined on $[0,1]$ and project down to a nonconstant energy minimizer contained in K, then $\mathcal{G}_{nc,min}(K)$ is relatively compact. This has a number of important consequences, one of which is the fact that, if an open set U is "relatively convex" (that is, any two points of U can be joined by a minimizer, which need not be contained in U) then for every compact subset K of U the set of solutions of the geodesic equation that are defined on $[0,1]$ and project down to a nonconstant energy minimizer whose endpoints are in K is also relatively compact. This yields as a simple corollary the subanalyticity of the distance function for real-analytic subriemannian manifolds that satisfy a geodesic convexity condition. (Global subanalyticity of the distance to a point $\bar{x}$, for the 3-dimensional contact case, on any punctured ball $\{x : 0 < d(x,\bar{x}) \leq R\}$ such that the closed ball $\{x : 0 \leq d(x,\bar{x}) \leq R\}$ is compact, was proved in Sussmann [33]. A local result in the general SBG case, giving subanalyticity on some small neighborhood of a point, can be found in Ge [11]. More recently, an example of a non-SBG real-analytic subriemannian manifold for which there are points x such that the r-ball centered at x fails to be subanalytic even for small r has been exhibited by Agrachev, Bonnard, Chyba and Kupka in [2].)

As in Bellaïche [4], we work in a slightly more general setting than that usually considered in the literature (cf. for example Agrachev-Sarychev [1], Bär [3], Brockett [6,7], Bryant-Hsu [8], Gaveau [10], Ge [11], Gershkovich [12,13], Hamenstädt [14], Kupka [21], Liu-Sussmann [23,24], Montgomery [26,27], Strichartz [29,30], Sussmann [34], Taylor [36,37], Vershik-Gershkovich [38,39]). Instead of defining a "subriemannian manifold" to be a triple (M,E,G) such that M is a manifold, E is a subbundle of the tangent bundle TM, and G is a Riemannian metric on E, we allow E to be a "singular distribution," i. e. a pair (F,σ) such that F is a smooth vector bundle over M and $\sigma = \{\sigma_x\}_{x\in M}$ is a smooth vector-bundle homomorphism from F to TM, without requiring all the maps σ_x to have the same rank. On the other hand, we follow the customary practice of incorporating the "bracket-generating" property —also known as the "Hörmander condition"— into the definition.

Exactly as in the Riemannian case, the length minimization problem in a subriemannian manifold $\mathcal{M} = (M,F,\sigma,G)$ is closely related to that of energy minimization for arcs $\xi : [0,1] \to M$. Precisely, ξ is an energy minimizer if and only if it is a length minimizer parametrized by constant times arc-length. Since every admissible arc $\xi : [a,b] \to M$ admits a "canonical reparametrization" —i. e. a reparametrization $\xi_1 : [0,1] \to M$ which is parametrized by constant times arc-length— it follows that one can find all the length minimizers by finding instead all the energy minimizers.

The Maximum Principle of optimal control theory (cf. Berkovitz [5], Lee-Markus [22], Pontryagin *et al.* [28], and also Sussmann [35] for a coordinate-free version particularly adapted to the subriemannian minimization problem) implies that every energy minimizer is an "extremal." Extremals are of one of two (not mutually exclusive) kinds, namely, "geodesics" (also called "normal extremals") or "abnormal extremals." Geodesics are projections of curves in the cotangent bundle T^*M that are "solutions of the geodesic equation," i.e. integral curves of a Hamilton vector field $\vec{H}_{\mathcal{M}}$ on T^*M arising from a function $H_{\mathcal{M}}$ —the "Hamiltonian" of $\mathcal{M}$— which is smooth if $\mathcal{M}$ is smooth, and analytic if $\mathcal{M}$ is analytic. In general, it is possible for a length minimizer to be "strictly abnormal" (i. e. not a normal extremal, which implies in particular that it is an abnormal extremal). This was a controversial issue for more than ten years —during which a number of erroneous proofs were given of the nonexistence of strictly abnormal minimizers, as well as incorrect examples of such minimizers— but the question is now definitively settled. Examples of strictly abnormal minimizers are given in Montgomery [26,27], Liu-Sussmann [23], [24], Sussmann [34], and Agrachev-Sarychev [1]. (For earlier discussions of this issue, cf. Bär [3], Brockett [6,7], Gaveau [10], Hamenstädt [14], Strichartz [29,30], Taylor [36,37].)

If $\mathcal{M}$ is SBG, then nonconstant abnormal extremals do not occur. So the necessary condition for an arc $t \to \xi(t)$ in M to be a minimizer is that there exist a field $t \to \psi(t)$ of covectors along ξ —known in control theory as an "adjoint vector"— such that the curve $t \to \Xi(t) = (\xi(t), \psi(t))$ is a solution of the geodesic equation. Our main result is really an *a priori* bound for the adjoint vector ψ: if $K \subseteq M$ is compact, then there exists a compact subset J_K of T^*M such that, if $\xi : [0,1] \to M$ is any nonconstant energy minimizer contained in K, parametrized by constant times arc-length, then the solution $\Xi = (\xi, \psi)$ of the geodesic equation that projects down to ξ (which happens to be unique, cf. Theorem 6.7 below) is in fact contained in J_K.

Our estimate is trivial for Riemannian metrics, because in that case $||\psi(t)||$ is just the length of ξ. In the subriemannian case, one automatically gets a bound in terms of the length for the components of $\psi(t)$ in the direction of the vectors in $E(\xi(t))$, but this is not enough, in principle, to yield an estimate for the magnitude of the whole covector $\psi(t)$. To get such an estimate, one needs a more delicate argument, based on a detailed analysis of the geodesics and integration by parts.

The paper is organized as follows: in §2 we review some basic definitions and introduce notations about manifolds, charts, curves, momentum functions and Hamiltonian vector fields; we also define the notion of a singular distribution, and review some of the basic properties of these objects and of the natural classes of curves associated to them. In §3 we review the definitions of subriemannian manifold, the distance function, and length and energy minimizers. In §4 we discuss the basic convexity theorems for subriemannian manifolds, and in particular the subriemannian version of the Hopf-Rinow theorem. In §5 we review the theory of extremals and geodesics, and explain the basic facts relating these curves to minimizers, namely, that every energy minimizer is an extremal and every geodesic is a local minimizer. In §6 we review the definition of an SBG structure, and study some of its elementary properties, such as (a) the nonexistence of nonconstant abnormal extremals, (b) the uniqueness of the adjoint vector, (c) the fact that the local minimizers are exactly the geodesics, and (d) the particularly simpler form that the Hopf-Rinow theorem has in the SBG case. In §7 we state and prove our adjoint vector estimate. Finally, in §8 we use the results of §7 to prove theorems on subanalyticity of the subriemannian distance in the SBG case. A number of proofs —especially those dealing with relatively less interesting technical details such as measurability of various maps, and the precise properties of reparametrization by arc-length— are relegated to an appendix.

§2. Manifolds, singular distributions, charts, curves.

We begin by reviewing some basic concepts and well known facts, and introducing some notations that will be used throughout the paper.

The word *interval* will always mean "connected subset of $\mathbb{R}$." An interval I is *nontrivial* if it contains at least two points, i. e. if it has positive measure. In that case, every nonempty relatively open subinterval of I is also nontrivial.

We define a *metric space* to be a pair (M,d) where M is a set and $d: M\times M \to [0,+\infty]$ is an extended-real-valued function satisfying the usual properties —symmetry, the triangle inequality, and $d(x,y)=0 \iff x=y$ — of a distance. This only differs from the most widely used definition in that we allow d to take the value $+\infty$. If (M,d) is a metric space in our sense, then we can define $\tilde{d}(x,y) = \min\Big(d(x,y),1\Big)$, and $(M,\tilde{d})$ is another metric space having the same topology and uniform structure as (M,d). So, for example, the notion of metrizable space in our sense agrees with the usual one. In particular, metrizable spaces are always paracompact (cf. Kelley [19]).

If M is any metric space, and I is a compact subinterval of $\mathbb{R}$, we use $C^0(I,M)$ to denote the space of all continuous maps from I to M, with the topology of uniform convergence. A sequence $\{\xi_j\}$ in $C^0(I,M)$ converges in $C^0(I,M)$ —i. e. uniformly— if and only if $\xi_j(t_j) \to \xi(t)$ whenever $t_j \to t$. In particular, the topology of $C^0(I,M)$ depends only on the topology of M, and not on the metric, and is therefore well defined if M is just a metrizable topological space.

The word "smooth" will mean "of class C^∞." The word "manifold" will always stand for "smooth finite-dimensional Hausdorff paracompact manifold without boundary." (In any case, we will only be interested in subriemannian manifolds, which are always metrizable, and hence paracompact.)

If M is a manifold, we use $C^\infty(M)$ to denote the ring of smooth real-valued functions on M. A *vector bundle* is a smooth, finite-dimensional vector bundle. If F is a vector bundle over a manifold M, and $x\in M$, we use $F(x)$ to denote the fiber of F at x. The *dimension* $\dim F$ of F is its fiber dimension. We use F^* to denote the dual of F, and $\Gamma(F)$ to denote the space of smooth sections of F, so $\Gamma(F)$ is a $C^\infty(M)$-module. We write π_F to denote the canonical projection from F to M. We use $<\cdot,\cdot>$ —omitting the explicit mention of F— to indicate the canonical pairing of F^* and F, so $<v,w>$ is defined iff $v\in F^*$, $w\in F$, and $\pi_{F^*}(v)=\pi_F(w)$. If U is an open subset of M, then $F\lceil U$ denotes the restriction of F to U. A *local basis of sections* of F is a basis $\mathbf{f}=(f_1,\ldots,f_m)$ of smooth sections of $F\lceil U$ for some nonempty open subset U of M. If $\mathbf{f}$ is a local basis of sections of F, we write $U_{\mathbf{f}}$ to denote the domain of $\mathbf{f}$. If $x\in M$, then a *local basis of sections of F near x* is a local basis $\mathbf{f}$ of sections of F such that $x\in U_{\mathbf{f}}$. A *Riemannian metric* on F is a smooth section G of the bundle $\mathrm{Hom}(F,F^*)$ such that for each x the bilinear form

$$F(x)\times F(x)\ni (v,w)\to G_x(v,w)\stackrel{\mathrm{def}}{=}<G(x)v,w>\in\mathbb{R}$$

is symmetric and strictly positive definite. If G is a Riemannian metric on F, then a local basis $\mathbf{f}=(f_1,\ldots,f_m)$ of sections of F is *orthonormal* if $G_x(f_i(x),f_j(x))=\delta_{ij}$ for all $x \in U_{\mathbf{f}}$ and all $i, j \in \{1, \ldots, m\}$, where δ_{ij} is Kronecker's delta. It is clear that, if G is a Riemannian metric on the vector bundle F over the manifold M, and $x \in M$, then a local orthonormal basis of sections of F near x always exists, since it is always possible to choose a local basis of sections of F near x and then construct from it an orthonormal basis by means of the Gramm-Schmidt orthonormalization procedure.

Suppose M is a manifold. We use TM, T^*M to denote, respectively, the tangent and cotangent bundles of M. Then the expressions π_{TM}, π_{T^*M}, $\Gamma(TM)$, $\Gamma(T^*M)$ denote, respectively, the canonical projections from TM and T^*M to M, the space of all smooth vector fields on M, and the space of all smooth 1-forms on M. If $x \in M$, we write T_xM and T_x^*M —rather than $TM(x)$ and $T^*M(x)$— to denote, respectively, the fibers of TM and T^*M at x, i. e. the tangent and cotangent spaces of M at x.

2.1. Definition. A *singular subbundle* of a vector bundle B on a manifold M is a pair $E = (F, \sigma)$, where F is a vector bundle over M and $\sigma : \Gamma(F) \to \Gamma(B)$ is a homomorphism of $C^\infty(M)$-modules. □

The following observation is trivial, but we prove it in the appendix for completeness:

2.2. Proposition. *If (F, σ) is a singular subbundle of the vector bundle B over M, then σ induces, for each $x \in M$, a linear map $\sigma_x : F(x) \to B(x)$, such that, if $v \in F(x)$, and $V \in \Gamma(F)$ satisfies $V(x) = v$, then $\sigma_x(v) = \sigma(V)(x)$. The maps σ_x depend smoothly on x, in the sense that for each $V \in \Gamma(F)$ the map $x \to \sigma_x(V(x))$ is smooth. Conversely, any family $\{\sigma_x : x \in M\}$ of linear maps $F(x) \to B(x)$ that depends smoothly on x arises from a module homomorphism $\sigma : \Gamma(F) \to \Gamma(B)$.* □

If $E = (F, \sigma)$ is a singular subbundle of the bundle B over M, and $x \in M$, we write $E(x) = \sigma_x(F(x))$. We can then think of E as a "vector bundle whose fibers are allowed to be of variable dimension."

If (F, σ) is a singular subbundle of a vector bundle B over M, then Proposition 2.2 implies, in particular, that $\sigma(V)$ vanishes near x if $V \in \Gamma(F)$ and $V \equiv 0$ near x. This implies that $\sigma(V)$ is a well defined member of $\Gamma(B \lceil \Omega)$ if Ω is an open subset of M and $V \in \Gamma(F \lceil \Omega)$.

2.3. Definition. A *singular distribution* on a manifold M is a singular subbundle of the tangent bundle TM. □

If $E = (F, \sigma)$ is a singular distribution on M, then the space $\sigma(\Gamma(F))$ is a $C^\infty(M)$-submodule of the Lie algebra $\Gamma(TM)$ of all smooth vector fields on M. We use $\Gamma(E)$ to denote this submodule, and $\Lambda(E)$ to denote the Lie subalgebra of $\Gamma(TM)$ generated by $\Gamma(E)$, and write $\Lambda(E)(x) \stackrel{\text{def}}{=} \{X(x) : X \in \Lambda(E)\}$.

If $E^i = (F^i, \sigma^i)$, $i = 1, 2$, are singular distributions on M, we can define singular distributions $E^1 + E^2$, $[E^1, E^2]$ as follows. Let $F = F^1 \oplus F^2$ be the direct sum of the bundles F^1 and F^2, so the fiber $F(x)$ is $F^1(x) \oplus F^2(x)$ for each x. Let $\sigma_x : F(x) \to T_xM$ be given by $\sigma_x(v_1, v_2) = \sigma_x^1(v_1) + \sigma_x^2(v_2)$. Then $E^1 + E^2$ is the singular distribution (F, σ). Let $\tilde{F}^i$, $i = 1, 2$, be the bundles of 1-jets of sections of F^1, F^2, and let F' be the bundle $\tilde{F}^1 \otimes \tilde{F}^2$, so that $F'(x) = \tilde{F}^1(x) \otimes \tilde{F}^2(x)$ for each x. If $v^1 \in \tilde{F}^1(x)$, $v^2 \in \tilde{F}^2(x)$, then we can find smooth sections V^i of F^i such that the 1-jet of V^i at x is v^i. The value $[\sigma^1(V^1), \sigma^2(V^2)](x)$

of the Lie bracket $[\sigma^1(V^1), \sigma^2(V^2)]$ depends only on v^1 and v^2, and is independent of the choice of the sections V^1, V^2. We define $\tilde{\sigma}_x(v^1, v^2) = [\sigma^1(V^1), \sigma^2(V^2)](x)$. It is then clear that $\tilde{\sigma}_x : \tilde{F}^1(x) \times \tilde{F}^2(x) \to T_xM$ is a bilinear map. Therefore $\tilde{\sigma}_x$ induces a linear map $\sigma'_x : F'(x) \to T_xM$. We then define $[E^1, E^2]$ to be the singular distribution (F', σ'). It then follows that the fiber $[E^1, E^2](x)$ is the linear span of all the brackets $[X_1, X_2](x)$, for all $X_1 \in \Gamma(E^1)$, $X_2 \in \Gamma(E^2)$. Moreover,

$$E^1(x) \subseteq [E^1, E^2](x) \quad \text{if} \quad E^2(x) \neq \{0\}.$$

(*Proof*: Pick $Y \in \Gamma(E^2)$ such that $Y(x) \neq 0$. Let $v \in E^1(x)$. Find $X \in \Gamma(E^1)$ such that $X(x) = v$. Find a function $\rho \in C^\infty(M)$ such that $\rho(x) = (Y\rho)(x) = 1$. Then $v = [X, Y](x) - [\rho X, Y](x)$, so $v \in [E^1, E^2](x)$.) Similarly, $E^2(x) \subseteq [E^1, E^2](x)$ if $E^1(x) \neq \{0\}$.

In particular, if E is a singular distribution, we can define singular distributions E^k inductively by letting

$$E^1 \stackrel{\text{def}}{=} E, \qquad E^k \stackrel{\text{def}}{=} E + [E, E^{k-1}] \quad \text{if} \quad k > 1.$$

If $E(x) \neq \{0\}$ for all x, then E^k is simply equal to $[E, E^{k-1}]$ whenever $k > 1$.

2.4. Definition. If k is a positive integer, and E is a singular distribution on the manifold M, we say that E is *k-generating at a point x* if $E^k(x) = T_xM$, i.e. if T_xM is the linear span of all the iterated brackets $[X_1, [\ldots, [X_{i-1}, X_i]\ldots]](x)$, $i \in \{1, \ldots, k\}$, $X_1, \ldots, X_i \in \Gamma(E)$. We say that E is *k-generating on a subset S of M* if it is k-generating at every $x \in S$. We call E *bracket-generating* if $\Lambda(E)(x) = T_xM$ for all $x \in M$. □

Equivalently, E is bracket-generating iff for every x there exists a positive integer k such that E is k-generating at x. For a subset S of M, the smallest k for which E is k-generating on S is called the *degree of nonholonomy* of E on S. Then E is bracket-generating iff for every compact subset S of M the degree of nonholonomy of E on S is finite.

A *chart* of a manifold M is a smooth diffeomorphism $\mathbf{x}$ from an open subset $\mathrm{Dom}(\mathbf{x})$ of M onto an open subset $\mathrm{Im}(\mathbf{x})$ of $\mathbb{R}^n$. To a chart $\mathbf{x} = (x^1, \ldots, x^n)$ of M we associate local bases of sections $\partial^{\mathbf{x}}$, $d\mathbf{x}$ of TM, T^*M, with domain $\mathrm{Dom}(\mathbf{x})$, by letting $\partial^{\mathbf{x}}$ be the n-tuple $(\partial^{\mathbf{x}}_1, \ldots, \partial^{\mathbf{x}}_n)$, where $\partial^{\mathbf{x}}_i$ is the vector field usually labeled $\frac{\partial}{\partial x^i}$, and defining $d\mathbf{x} = (dx^1, \ldots, dx^n)$. We also associate charts $\mathbf{x}^{tan}$, $\mathbf{x}^{cot}$ of TM, T^*M, whose domains are the sets $\pi^{-1}_{TM}(\mathrm{Dom}(\mathbf{x}))$, $\pi^{-1}_{T^*M}(\mathrm{Dom}(\mathbf{x}))$, by letting $\mathbf{x}^{tan} = (x^1, \ldots, x^n, y^1, \ldots, y^n)$ and $\mathbf{x}^{cot} = (x^1, \ldots, x^n, p_1, \ldots, p_n)$, where the functions $y^i : \pi^{-1}_{TM}(\mathrm{Dom}(\mathbf{x})) \to \mathbb{R}$ and $p_i : \pi^{-1}_{T^*M}(\mathrm{Dom}(\mathbf{x})) \to \mathbb{R}$ are given by $y^i(x, y) = \langle dx^i, y \rangle$, $p_i(x, p) = \langle p, \partial^{\mathbf{x}}_i \rangle$, and "$x^i$" really means $x^i \circ \pi_{TM}$ or $x^i \circ \pi_{T^*M}$.

The manifold T^*M has a canonical symplectic form Ω_M, given by $\Omega_M = d\omega_M$, where $\omega_M(v) = \langle p, d\pi_{T^*M}(v) \rangle$ whenever $(x, p) \in T^*M$, $v \in T_{(x,p)}T^*M$. To every smooth function $H \in C^\infty(T^*M)$ we associate the *Hamilton vector field* $\vec{H}$ characterized by the identity

$$\Omega_M(X, \vec{H}) = \langle dH, X \rangle \quad \text{for every} \quad X \in \Gamma(T(T^*M)). \tag{2.1}$$

To each vector field $g \in \Gamma(TM)$ we associate a function $\mu_g \in C^\infty(T^*M)$ —the *momentum function* of g— by letting

$$\mu_g(x,p) \stackrel{\text{def}}{=} \langle p, g(x)\rangle \quad \text{for} \quad (x,p) \in T^*M\,. \tag{2.2}$$

The vector field $\vec{\mu}_g$ is the *Hamiltonian lift* of g. It is clear that $d\pi_{T^*M}(\vec{\mu}_g) = g$.

Suppose that $\mathbf{x} = (x^1, \ldots, x^n)$ is a chart of the manifold M with domain U, and let $\mathbf{x}^{cot} = (x^1, \ldots, x^n, p_1, \ldots, p_n) : \pi^{-1}_{T^*M}(U) \to \mathbb{R}^{2n}$ be the corresponding chart of T^*M. Then the identities

$$\omega_M = \sum_{i=1}^{n} p_i\, dx^i\,, \quad \Omega_M = \sum_{i=1}^{n} dp_i \wedge dx^i\,,$$

hold on $\pi^{-1}_{T^*M}(U)$. Relative to $\mathbf{x}^{cot}$, the Hamilton vector field $\vec{H}$ corresponding to a function $H \in C^\infty(T^*M)$ has the expression

$$\vec{H} = \sum_i \Big(\frac{\partial H}{\partial p_i}\frac{\partial}{\partial x^i} - \frac{\partial H}{\partial x^i}\frac{\partial}{\partial p_i}\Big)\,, \tag{2.3}$$

and the integral curves of $\vec{H}$ satisfy the system of $2n$ ordinary differential equations

$$\dot{x}^i = \frac{\partial H}{\partial p_i}\,, \qquad \dot{p}_i = -\frac{\partial H}{\partial x^i}\,, \qquad i = 1, \ldots, n \tag{2.4}$$

known as *Hamilton's equations.* It follows easily from (2.4) that

$$H \equiv \text{constant} \quad \text{along any trajectory of} \quad \vec{H}\,. \tag{2.5}$$

2.5. Definition. A *curve* in a topological space X is a continuous map ξ from a nonempty interval $I(\xi)$ —which is allowed to be open, closed or half-open, and can be bounded or unbounded, and even trivial— to X. We say that ξ is an *arc in M* if $I(\xi)$ is compact. We write $\mathcal{C}(M)$, $\mathcal{A}(M)$, to denote, respectively, the set of all curves and the set of all arcs in X. If X is a manifold, we say that an arc ξ in X is an *AC arc* if it is absolutely continuous. We write $\mathcal{C}_{AC,loc}(X)$, $\mathcal{A}_{AC}(X)$ to denote, respectively, the set of all locally absolutely continuous curves in X and the set of all AC arcs in M. $\square$

If M is a manifold, $\xi \in \mathcal{A}(M)$, and $I(\xi) = [a,b]$, then we write

$$\tau_-\xi \stackrel{\text{def}}{=} a\,, \quad \tau_+\xi \stackrel{\text{def}}{=} b\,, \quad \partial_-\xi \stackrel{\text{def}}{=} \xi(a)\,, \quad \partial_+\xi \stackrel{\text{def}}{=} \xi(b)\,, \quad \partial\xi \stackrel{\text{def}}{=} (\partial_-\xi, \partial_+\xi)\,,$$

and refer to $\tau_-\xi$, $\tau_+\xi$, $\partial_-\xi$ and $\partial_+\xi$, respectively, as the *initial and terminal times* and the *initial and terminal points* of ξ, and to $\partial\xi$ as the *boundary value*, or *boundary condition*, or *endpoint condition*, of ξ. If $\partial\xi = (\bar{x}, \hat{x})$, we say that ξ *goes from* $\bar{x}$ *to* $\hat{x}$.

2.6. Definition. If $E = (F, \sigma)$ is a singular distribution on the manifold M, an *E-curve* is a curve $\xi \in \mathcal{C}_{AC,loc}(M)$ such that $\dot{\xi}(t) \in E(\xi(t))$ for almost every $t \in I(\xi)$. An *E-arc* is an E-curve which is an arc. We use $\mathcal{C}_E(M)$, $\mathcal{A}_E(M)$ to denote, respectively, the set of all E-curves and the set of all E-arcs in M. $\square$

If ξ is an E-curve in M, then there exists a section ζ of F along ξ —i. e. a map $I(\xi) \ni t \to \zeta(t) \in F(\xi(t))$— such that $\sigma_{\xi(t)}(\zeta(t)) = \dot{\xi}(t)$ for almost all $t \in I(\xi)$. In general, ζ will not be unique, since the maps σ_x need not be injective. It can be shown, however, that ζ can always be chosen to be measurable:

2.7. Proposition. *Let $E = (F, \sigma)$ be a smooth singular distribution on a manifold M, and let ξ be an E-curve. Then there exists a measurable map $I(\xi) \ni t \to \zeta(t) \in F(\xi(t))$ such that $\dot{\xi}(t) = \sigma_{\xi(t)}(\zeta(t))$ for almost all $t \in I(\xi)$.*

PROOF. This follows from standard measurable selection theorems. In addition, for our specific situation there is a simple direct proof, as shown in the appendix. ☐

If ξ is an E-curve in M, J is a subinterval of $I(\xi)$, $\mathbf{f} = (f_1, \ldots, f_m)$ is a local basis of sections of F such that $\xi(J) \subseteq U_{\mathbf{f}}$, and $g_i = \sigma(f_i)$ for $i = 1, \ldots, m$, then Proposition 2.7 tells us that there exist measurable functions $\eta^i : J \to \mathbb{R}$, $i = 1, \ldots, m$, such that

$$\dot{\xi}(t) = \sum_{i=1}^{m} \eta^i(t) g_i(\xi(t)) \quad \text{for a.e.} \quad t \in J\,. \tag{2.6}$$

2.8. Definition. A vector-valued function $\eta = (\eta^1, \ldots, \eta^m)$ such that (2.6) holds is said to be a *local control representation* of ξ with respect to $\mathbf{f}$ on J. The functions η^i are the *controls* corresponding to the representation η. ☐

2.9. Remark. Local control representations are not unique, first because the local basis $\mathbf{f}$ is not unique, and second because, for a given $\mathbf{f}$, the functions η^i need not be unique. Once the local basis $\mathbf{f}$ is specified, choosing a control representation is of course tantamount to choosing a measurable map ζ such that $\zeta(t) \in F(\xi(t))$ and $\sigma_{\xi(t)}(\zeta(t)) = \dot{\xi}(t)$ for a. e. t.

On the other hand, if $\mathbf{f}^\ell = (f_1^\ell, \ldots, f_m^\ell)$, $\ell = 1, 2$, are two local bases of sections, then $f_i^1 = \sum_{j=1}^m \alpha_i^j f_j^2$ on $U_{\mathbf{f}^1} \cap U_{\mathbf{f}^2}$, where the α_i^j are smooth functions. So, if $J \subseteq I(\xi)$, $\xi(J) \subseteq U_{\mathbf{f}^1} \cap U_{\mathbf{f}^2}$, and $\eta_1 = (\eta_1^1, \ldots, \eta_1^m)$ is a local control representation of ξ on J with respect to $\mathbf{f}^1$, then we can get a local control representation $\eta_2 = (\eta_2^1, \ldots, \eta_2^m)$ of ξ on J with respect to $\mathbf{f}^2$ by letting $\eta_2^j = \sum_{i=1}^m \eta_1^i \alpha_i^j$, and it is then clear that η_2 is locally integrable, continuous, or smooth on J if and only if η_1 is. ☐

2.10. Definition. We say that an E-curve ξ on M is *E-admissible* if, whenever $\mathbf{f}$ is a local basis of sections of F and J is a compact subinterval of $I(\xi)$ such that $\xi(J) \subseteq U_{\mathbf{f}}$, it follows that ξ has an integrable local control representation η on J with respect to $\mathbf{f}$. We use $\mathcal{C}_E^a(M)$, $\mathcal{A}_E^a(M)$ to denote, respectively, the set of all E-admissible E-curves and the set of all E-admissible E-arcs in M. ☐

2.11. Remark. In view of Remark 2.9, to prove that an E-curve ξ is E-admissible it suffices to prove that for every $t \in I$ there exist a compact subinterval J_t of $I(\xi)$ which is a relative neighborhood of t in $I(\xi)$, a local basis $\mathbf{f}^t$ of sections of F such that $\xi(J_t) \subseteq U_{\mathbf{f}^t}$, and an integrable local control representation η_t on J_t with respect to $\mathbf{f}^t$. ☐

2.12. Remark. If $E(x)$ has constant dimension, independent of x, then every E-curve is E-admissible. On the other hand, for general singular distributions it is easy to give examples of absolutely continuous —and even smooth— E-arcs that are not E-admissible (cf. Example A.1 in the appendix). □

2.13. Proposition. *If a singular distribution E on M is bracket-generating and M is connected then for every $\bar{x}, \hat{x} \in M$ there exists a $\xi \in \mathcal{A}^a_E(M)$ such that $\partial\xi = (\bar{x}, \hat{x})$.*

PROOF. This is a well-known consequence of "Chow's Theorem," cf. Chow [9], Hermann [16], Isidori [17], Jurdjevic [18], Krener [20], Sussmann [31]. □

§3. Subriemannian manifolds .

3.1. Definition. A *subriemannian manifold* is a quadruple $\mathcal{M} = (M, F, \sigma, G)$ such that M is a manifold, (F, σ) is a bracket-generating smooth singular distribution on M, and G is a Riemannian metric on F. The *dimension* of $\mathcal{M}$ is the dimension of M. □

From now on, we assume throughout the paper that $\mathcal{M} = (M, F, \sigma, G)$ is an n-dimensional subriemannian manifold, and let $m = \dim F$. We write $E = (F, \sigma)$.

If $\mathbf{f}$ is a local basis of sections of F, then we write $U^*_{\mathbf{f}} \stackrel{\text{def}}{=} \pi^{-1}_{T^*M}(U_{\mathbf{f}})$.

3.2. Definition. If $x \in M$, then the *G-length* $||v||_G$ of a vector $v \in T_xM$ is defined to be the infimum of the numbers $\sqrt{<G(x)w, w>}$, ranging over all the vectors $w \in F(x)$ such that $\sigma_x(w) = v$. □

Since the infimum of the empty set is equal to $+\infty$, it is clear that

$$\text{if} \quad v \in T_xM\,, \quad \text{then} \quad ||v||_G < \infty \iff v \in E(x)\,.$$

3.3. Remark. *If $v \in E(x)$, then there exists a unique $w \in F(x)$ such that $\sigma_x(w) = v$ and* $<G(x)w, w> = \|v\|^2_G$. (Indeed, the set $S(v)$ of those $w \in F(x)$ such that $\sigma_x(w) = v$ is an affine-linear subspace of $F(x)$, so the positive-definite quadratic form $w \to <G(x)w, w>$ has a unique minimum on $S(v)$.) It follows in particular, that $||v||_G = 0$ iff $v = 0$. □

We will use $\sigma^{-1}_{x,c}(v)$ to denote, for $v \in E(x)$, the "canonical choice" of a $w \in \sigma^{-1}_x(v)$ provided by Remark 3.3. That is, $\sigma^{-1}_{x,c}(v)$ is the unique $w \in F(x)$ such that $\sigma_x(w) = v$ and $\|w\|_G = \|v\|_G$.

If $\xi \in \mathcal{C}_E(M)$, then $\sigma^{-1}_{\xi(t),c}(\dot{\xi}(t))$ is a well defined member of $F(\xi(t))$ for almost every $t \in I(\xi)$. We write

$$\omega_\xi(t) \stackrel{\text{def}}{=} \sigma^{-1}_{\xi(t),c}(\dot{\xi}(t))\,, \tag{3.1}$$

so $\omega_\xi(t) \in F(\xi(t))$, $\|\omega_\xi(t)\|_G = \|\dot{\xi}(t)\|_G$, and $\dot{\xi}(t) = \sigma_{\xi(t)}(\omega_\xi(t))$ for almost every $t \in I(\xi)$.

If J is a subinterval J of $I(\xi)$ and $\mathbf{f}$ is a local basis of sections of F such that $\xi(J) \subseteq U_{\mathbf{f}}$, then we can write $\omega_\xi(t)$ uniquely as $\sum_{i=1}^m \eta^i_{\mathbf{f},\xi,J}(t) f_i(\xi(t))$ for a. e. $t \in J$.

3.4. Definition. If ξ, J, $\mathbf{f}$ are as above, then the control $\eta_{\mathbf{f},\xi,J}$ is the *canonical control representation* of ξ on J with respect to $\mathbf{f}$. □

The following two elementary technical observations are proved in the appendix:

3.5. Proposition. *If ξ is a locally absolutely continuous curve in M, then the set $I_0(\xi) \stackrel{\rm def}{=} \Big\{t \in I(\xi) : ||\dot\xi(t)||_G < \infty\Big\}$ is measurable, and the map ω_ξ is measurable on $I_0(\xi)$. In particular, (i) the a. e. defined function $I(\xi) \ni t \to ||\dot\xi(t)||_G \in \mathbb{R} \cup \{+\infty\}$ is measurable, (ii) if ξ is an E-curve, then ω_ξ is a. e. defined and measurable on $I(\xi)$.* □

3.6. Proposition. *A locally absolutely continuous curve ξ in M is an E-curve if and only if $||\dot\xi(t)||_G < \infty$ for almost all $t \in I(\xi)$. Moreover, ξ is E-admissible if and only if the function $t \to ||\dot\xi(t)||_G$ is locally integrable.* □

3.7. Definition. The *G-length* $||\xi||_G$ of an absolutely continuous arc ξ in M is the number $||\xi||_G \stackrel{\rm def}{=} \int_{I(\xi)} ||\dot\xi(t)||_G \, dt$, which is a well defined member of $[0, +\infty]$ by Proposition 3.5. If ξ is not absolutely continuous we let $||\xi||_G \stackrel{\rm def}{=} +\infty$. □

It follows from Proposition 3.6 that

3.8. Proposition. *If $\xi \in \mathcal{A}(M)$, then ξ is an E-admissible E-arc iff $||\xi||_G < \infty$.* □

3.9. Definition. The *distance* $d_{\mathcal{M}}(\bar x, \hat x)$ between two points $\bar x$, $\hat x$ of M is given by

$$\begin{aligned} d_{\mathcal{M}}(\bar x, \hat x) &\stackrel{\rm def}{=} \inf \Big\{ ||\xi||_G : \xi \in \mathcal{A}(M)\ ,\ \ \partial\xi = (\bar x, \hat x)\Big\} \\ &= \inf \Big\{ ||\xi||_G : \xi \in \mathcal{A}_E(M),\ \partial\xi = (\bar x, \hat x)\Big\} \\ &= \inf \Big\{ ||\xi||_G : \xi \in \mathcal{A}_E^a(M),\ \partial\xi = (\bar x, \hat x)\Big\}. \end{aligned}$$

(The equality of the the three right-hand sides follows easily from Proposition 3.8, together with the fact that $\inf \emptyset = +\infty$.) □

3.10. Remark. *$(M, d_{\mathcal{M}})$ is a metric space, whose topology coincides with the original manifold topology of M.* This follows easily from the fact that the definition of a subriemannian manifold used here includes the requirement that E be bracket-generating. In particular, *the metric space $(M, d_{\mathcal{M}})$ is locally compact.* □

3.11. Remark. If M is connected then $d_{\mathcal{M}}(\bar x, \hat x) < +\infty$ for all $\bar x, \hat x \in M$. □

We now state a well known estimate relating the subriemannian distance to the distance arising from an ordinary Riemannian metric. (For a proof, cf. for example Gershkovich [12,13], Vershik-Gershkovich [38,39].)

3.12. Proposition. *Let g be a Riemannian metric on M, and let d_g be the corresponding distance function. Let K be a compact subset of M, and let $k > 0$ be an integer such that E is k-generating on K. Then there exist positive constants Δ_1, Δ_2 such that the inequalities*

$$\Delta_1 d_g(x_1, x_2) \le d_{\mathcal{M}}(x_1, x_2) \le \Delta_2 d_g(x_1, x_2)^{1/k} \tag{3.2}$$

hold for all $x_1, x_2 \in K$. □

3.13. Definition. If $\xi_i : I_i \to M$, $i = 1, 2$, are arcs in M, then we say that ξ_1 is a *time-reparametrization* of ξ_2 if there exists an absolutely continuous monotonically nondecreasing function $\theta : I_2 \to I_1$ such that $\xi_1 \circ \theta = \xi_2$. We write $TR(\xi)$ to denote the set of all time-reparametrizations of ξ. □

3.14. Remark. The formula "$\xi_1 \in TR(\xi_2)$" does not define an equivalence relation, because (i) the map θ need not be invertible, and (ii) even when θ is invertible, its compositional inverse need not be absolutely continuous. □

3.15. Definition. We use $\mathcal{C}^a(\mathcal{M})$ (resp. $\mathcal{A}^a(\mathcal{M})$) as an alternative notation for the set $\mathcal{C}^a_E(M)$ (resp. $\mathcal{A}^a_E(M)$) of E-admissible E-curves (resp. E-arcs) in M. If $\xi \in \mathcal{C}^a(\mathcal{M})$, then we say that ξ is *parametrized by a constant times arc-length* if there exists a constant c such that $||\dot{\xi}(t)||_G = c$ for almost all $t \in I(\xi)$. We say that ξ is *parametrized by arc-length* if the constant c is equal to 1. We call ξ *canonically parametrized* if it is parametrized by constant times arc-length and $I(\xi)$ is the interval $[0, 1]$. □

We use $\mathcal{C}^{PAL}(\mathcal{M})$ (resp. $\mathcal{C}^{PCAL}(\mathcal{M})$, $\mathcal{C}^{CAN}(\mathcal{M})$) to denote the set of all curves $\xi \in \mathcal{C}^a(\mathcal{M})$ that are parametrized by arc-length (resp. parametrized by constant times arc-length, canonically parametrized). We use $\mathcal{A}^{PAL}(\mathcal{M})$, $\mathcal{A}^{PCAL}(\mathcal{M})$, $\mathcal{A}^{CAN}(\mathcal{M})$ to denote the corresponding sets of arcs, and write $\mathcal{A}^{PAL,0}(\mathcal{M})$ to denote the set of all $\xi \in \mathcal{A}^{PAL}(\mathcal{M})$ such that $\tau_-\xi = 0$. Naturally, $\mathcal{C}^{CAN}(\mathcal{M}) = \mathcal{A}^{CAN}(\mathcal{M})$.

It is clear that

$$\xi \in \mathcal{A}^{PAL,0}(\mathcal{M}) \implies I(\xi) = [0, ||\xi||_G]\,, \tag{3.3}$$

and

$$\xi \in \mathcal{A}^{CAN}(\mathcal{M}) \implies ||\dot{\xi}(t)||_G = ||\xi||_G \quad \text{for a. e.} \quad t \in [0, 1]\,. \tag{3.4}$$

The following two technical propositions are proved in the appendix.

3.16. Proposition. *Arc-length is invariant under time-reparametrizations, in the sense that, if ξ_1 and ξ_2 are locally absolutely continuous arcs in M, and $\xi_1 \in TR(\xi_2)$, then $||\xi_2||_G = ||\xi_1||_G$.* □

3.17. Proposition. *If $\xi \in \mathcal{A}^a(\mathcal{M})$, then the set $TR(\xi) \cap \mathcal{A}^{PAL,0}(\mathcal{M})$ consists of a unique element ξ_1. If ξ is nontrivial (i. e. if $I(\xi)$ is nontrivial), then there exists a unique $\xi_2 \in TR(\xi) \cap \mathcal{A}^{CAN}(\mathcal{M})$.* □

3.18. Definition. If $\xi \in \mathcal{A}^a(\mathcal{M})$, then the unique member of $TR(\xi) \cap \mathcal{A}^{PAL,0}(\mathcal{M})$ is called the *arc-length reparametrization* of ξ. If ξ is nontrivial, then the unique member of $TR(\xi) \cap \mathcal{A}^{CAN}(\mathcal{M})$ is the *canonical reparametrization* of ξ. □

3.19. Remark. If $\xi : \{a\} \to M$ is a trivial arc, then there is no "canonical reparametrization" in the sense of our definition, but in that case we agree to call the constant arc $\xi_2 : [0, 1] \to M$ given by $\xi_2(t) = \xi(a)$ for $0 \le t \le 1$ the *canonical reparametrization* of ξ. □

If $c \in [0, \infty[$, we use $\mathcal{C}^{a,\le c}(\mathcal{M})$ —resp. $\mathcal{A}^{a,\le c}(\mathcal{M})$— to denote the set of all curves —resp. arcs— $\xi \in \mathcal{C}^a(\mathcal{M})$ such that $||\dot{\xi}||_G \le c$ for almost all $t \in I(\xi)$. We prove in the appendix that:

3.20. Proposition. *If $\xi : [a,b] \to M$ is an arc, and $c > 0$, then the following two conditions are equivalent:*

(a) $d_{\mathcal{M}}(\xi(t_1),\xi(t_2)) \le c|t_1 - t_2|$ *for all* t_1, t_2 *in* $[a,b]$,

(b) $\xi \in \mathcal{A}^{a,\le c}(\mathcal{M})$. □

An important consequence of Proposition 3.20 is the following

3.21. Proposition. *If $[a,b]$ is a compact interval, and $c \in \mathbb{R}$, $c > 0$, then the set $C^0([a,b],M) \cap \mathcal{A}^{a,\le c}(\mathcal{M})$ is closed in $C^0([a,b],M)$.*

PROOF. Since d is continuous, the set of all arcs ξ that satisfy Condition (a) of the statement of Proposition 3.20 is obviously closed. The conclusion then follows immediately from Proposition 3.20. □

3.22. Definition. A *length minimizer* of $\mathcal{M}$ is an arc $\xi \in \mathcal{A}^a(\mathcal{M})$ with the property that $||\xi||_G = d_{\mathcal{M}}(\partial\xi)$. A *local length minimizer* of $\mathcal{M}$ is an E-curve $\xi \in \mathcal{C}^a(\mathcal{M})$ such that for every $t \in I(\xi)$ there exists a compact subinterval J of $I(\xi)$ with the property that t belongs to the relative interior of J in $I(\xi)$ and the restriction of ξ to J is a length minimizer. □

We use $Min(\mathcal{M})$ —resp. $LMin(\mathcal{M})$— to denote the set of all length minimizers —resp. local length minimizers— of $\mathcal{M}$. In addition, we use superscripts to denote various classes of members of $Min(\mathcal{M})$ or $LMin(\mathcal{M})$, in a self-explanatory way. For example, $LMin^{PAL}(\mathcal{M})$ is the set of all local length minimizers that are parametrized by arc-length, and $Min^{CAN}(\mathcal{M})$ is the set of all canonically parametrized length minimizers.

3.23. Remark. If ξ is a length minimizer, then Proposition 3.17 tells us that there exist $\xi_1 \in TR(\xi) \cap Min^{PAL,0}(\mathcal{M})$, and $\xi_2 \in TR(\xi) \cap Min^{CAN}(\mathcal{M})$. Then $\partial\xi_1 = \partial\xi_2 = \partial\xi$, and Proposition 3.16 implies that $||\xi_1||_G = ||\xi_2||_G = ||\xi||_G$, so ξ_1 and ξ_2 are length minimizers as well. In particular, *if there exists a length minimizer ξ such that $\partial\xi = (\bar{x},\hat{x})$, then the set of all such minimizers contains at least one canonically parametrized arc, and one arc which is parametrized by arc-length and defined on an interval of the form $[0,L]$.* □

3.24. Proposition. *$Min^{CAN}(\mathcal{M})$ is a closed subset of $C^0([0,1],M)$.*

PROOF. Suppose $\{\xi_j\}_{j=1}^{\infty}$ is a sequence of members of $Min^{CAN}(\mathcal{M})$ that converge uniformly to an arc $\xi : [0,1] \to M$. Let $L_j = ||\xi_j||_G$. Then $L_j = d_{\mathcal{M}}(\xi_j(0),\xi_j(1))$ since ξ_j is a length minimizer. So $L_j \to L = d_{\mathcal{M}}(\xi(0),\xi(1))$ as $j \to \infty$. In particular, the sequence $\{L_j\}$ is bounded. If $\Lambda_k = \sup\{L_j : j \in \mathbb{N}, j \ge k\}$, then the fact that ξ_j is canonically parametrized implies that $||\dot{\xi}_j(t)||_G = L_j$ for almost all $t \in [0,1]$. Therefore $d_{\mathcal{M}}(\xi_j(t),\xi_j(t')) \le \Lambda_k|t-t'|$ whenever $t,t' \in [0,1]$ and $j \ge k$. So, for every k, $d_{\mathcal{M}}(\xi(t),\xi(t')) \le \Lambda_k|t-t'|$ whenever $t,t' \in [0,1]$. Therefore $d_{\mathcal{M}}(\xi(t),\xi(t')) \le L|t-t'|$ whenever $t,t' \in [0,1]$. Proposition 3.20 then implies that $\xi \in \mathcal{A}^{a,\le L}(\mathcal{M})$. So $||\dot{\xi}(t)||_G \le L$ for almost all t, and then $L = d_{\mathcal{M}}(\xi(0),\xi(1)) \le ||\xi||_G = \int_0^1 ||\dot{\xi}(t)||_G dt \le L$. Therefore $||\xi||_G = d_{\mathcal{M}}(\xi(0),\xi(1))$, showing that ξ is a minimizer, and $\int_0^1 ||\dot{\xi}(t)||_G dt \le L$, so $||\dot{\xi}(t)||_G = L$ for almost all t, and then ξ is canonically parametrized. □

If $a,b \in \mathbb{R}$ and $a \le b$, we use $\mathcal{A}_{[a,b]}(\mathcal{M})$ to denote the set of all absolutely continuous arcs in M whose interval of definition is $[a,b]$.

3.25. Definition. The extended-real-valued functional $\mathcal{E}_{\mathcal{M},a,b} : \mathcal{A}_{[a,b]}(\mathcal{M}) \to \mathbb{R} \cup \{\infty\}$ given by $\mathcal{E}_{\mathcal{M},a,b}(\xi) = \int_a^b ||\dot{\xi}(t)||_G^2 \, dt$ is the *energy functional* of $\mathcal{M}$ on the interval $[a,b]$. $\square$

(It is clear that the above integral is defined —but need not be finite— for all absolutely continuous arcs $\xi : [a,b] \to M$.) We let

$$d^{\mathcal{E},a,b}(\bar{x},\hat{x}) = \inf\left\{\mathcal{E}_{\mathcal{M},a,b}(\xi) : \xi \in \mathcal{A}_{[a,b]}(\mathcal{M}),\ \partial\xi = (\bar{x},\hat{x})\right\}. \tag{3.5}$$

3.26. Definition. An arc $\xi \in \mathcal{A}_{[a,b]}(\mathcal{M})$ is an *energy minimizer* if $\mathcal{E}_{\mathcal{M},a,b}(\xi) = d^{\mathcal{E},a,b}(\partial\xi)$. A *local energy minimizer* of $\mathcal{M}$ is an E-curve $\xi \in \mathcal{C}^a(\mathcal{M})$ such that for every $t \in I(\xi)$ there exists a compact subinterval J of $I(\xi)$ with the property that t belongs to the relative interior of J in $I(\xi)$ and the restriction of ξ to J is an energy minimizer. $\square$

The following proposition is proved in the appendix:

3.27. Proposition. *If $a, b \in \mathbb{R}$, $a < b$, then:*

(A) $d^{\mathcal{E},a,b}(\bar{x},\hat{x}) = (b-a).d_{\mathcal{M}}(\bar{x},\hat{x})^2$ for all $\bar{x},\hat{x}$ in M;
(B) if $\xi \in \mathcal{A}_{[a,b]}(\mathcal{M})$, then the following are equivalent:

(B.1) ξ is an energy minimizer;
(B.2) ξ is a length minimizer parametrized by constant times arc-length. $\square$

3.28. Remark. *(The Riemannian case).* A Riemannian manifold is a subriemannian manifold $\mathcal{M} = (M, F, \sigma, G)$ such that $F = TM$ and σ is the identity map. More generally, if a subriemannian manifold $\mathcal{M} = (M, F, \sigma, G)$ is such that σ is surjective —i. e., equivalently, σ_x is surjective for each $x \in M$— then there is a Riemannian metric $G^{\#}$ on M that gives rise to the same length and energy functionals, and the same distance function. Indeed, F splits as the direct sum $F_1 \oplus F_2$, where F_2 is the kernel of σ, and F_1 is such that $\sigma_1 = \sigma \lceil F_1$ is a bundle isomorphism onto TM. Then the restriction G_1 of the metric G to F_1 induces, via the isomorphism σ_1, an ordinary Riemannian metric $G^{\#}$ on TM. It is then clear that in this case every locally absolutely continuous curve in M is an E-admissible E-curve. Moreover, the definition of the G-length $||v||_G$ of a tangent vector is such that $||v||_G$ is precisely the Riemannian length $||v||_{G^{\#}}$. $\square$

§4. Completeness, convexity, and the Hopf-Rinow theorem.

4.1. Definition. Let $\mathcal{C}$ be a set of curves in a manifold M. A member ξ of $\mathcal{C}$ is *maximal within* $\mathcal{C}$ if there does not exist a $\xi' \in \mathcal{C}$ such that $I(\xi) \subseteq I(\xi')$, $I(\xi) \neq I(\xi')$, and $\xi'(t) = \xi(t)$ for all $t \in I(\xi)$. We use $\mathrm{Max}(\mathcal{C})$ to denote the set of all $\xi \in \mathcal{C}$ that are maximal within $\mathcal{C}$. We say that $\mathcal{C}$ has the *extension property* if, whenever $\xi \in \mathrm{Max}(\mathcal{C})$ and τ is a limit point of $I(\xi)$ such that $\lim_{t \to_{I(\xi)} \tau} \xi(t)$ exists, it follows that τ is an interior point of $I(\xi)$. $\square$

4.2. Example. *The class $\mathcal{C}^{PAL}(\mathcal{M})$ of all admissible curves parametrized by arc-length has the extension property. If $c > 0$, then the class $\mathcal{C}^{a,\leq c}(\mathcal{M})$ also has the extension property.* (See the appendix for a proof.) $\square$

4.3. Example. *In a riemannian manifold $\mathcal{M} = (M, G)$, the class $\mathcal{G}^{PAL}(\mathcal{M})$ of all geodesic arcs parametrized by arc-length has the extension property.* This follows easily from the fact that an arc ξ is in $\mathcal{G}^{PAL}(\mathcal{M})$ iff it is parametrized by arc-length and is a solution of the geodesic equation. We will see below that a similar conclusion is true for strongly bracket-generating subriemannian manifolds. □

4.4. Definition. Let $\bar{\xi} = \{\xi_j\}_{j=1}^{\infty}$ be a sequence of curves in a metric space X, and let ξ be a curve in X. We say that $\bar{\xi}$ *converges to ξ on compact intervals* if, whenever J is a compact subinterval of $I(\xi)$, it follows that $J \subseteq I(\xi_j)$ for sufficiently large j, and the ξ_j converge uniformly to ξ on J. □

We will need the following fact —proved in the appendix— which is a special case of a general result on "arc systems," namely, Proposition 2 of §2 of Liu-Sussmann [25]:

4.5. Proposition. *If $c>0$, then $\mathcal{C}^{a,\leq c}(\mathcal{M})$ has the following compactness property:*

(C) If $\{\xi_j\}_{j=1}^{\infty}$ is a sequence in $\mathrm{Max}(\mathcal{C}^{a,\leq c}(\mathcal{M}))$ and $\{t_j\}_{j=1}^{\infty}$ is a sequence in $\mathbb{R}$ such that $t_j \in I(\xi_j)$ for each j, and the limits $\bar{t} = \lim_{j\to\infty} t_j$ and $\bar{x} = \lim_{j\to\infty} \xi_j(t_j)$ exist, then there exist a subsequence $\{\xi_{j(k)}\}_{k=1}^{\infty}$ of $\{\xi_j\}_{j=1}^{\infty}$ and a curve $\xi \in \mathrm{Max}(\mathcal{C}^{a,\leq c}(\mathcal{M}))$ such that $\bar{t} \in I(\xi)$ and $\{\xi_{j(k)}\}_{k=1}^{\infty}$ converges to ξ on compact intervals. □

4.6. Definition. We call $\mathcal{M}$ *complete* if the metric space $(M, d_{\mathcal{M}})$ is complete. □

We define

$$B_{\mathcal{M}}(x, r) \stackrel{\text{def}}{=} \left\{ x' \in M : d_{\mathcal{M}}(x', x) \leq r \right\}. \tag{4.1}$$

The following statement is then the subriemannian analogue of the well-known Hopf-Rinow theorem of Riemannian geometry. The proof is given in the appendix.

4.7. Theorem. *The following four conditions are equivalent:*

(i) $\mathcal{M}$ is complete;
(ii) every closed ball $B_{\mathcal{M}}(x, r)$ is compact;
(iii) if $\{x_j\}_{j=1}^{\infty}$ is any sequence of points of M whose distances to a given point $\bar{x}$ are bounded above by a finite constant, then there exist a subsequence of $\{x_j\}_{j=1}^{\infty}$ that converges to a limit x, and a length minimizer ξ such that $\partial\xi = (\bar{x}, x)$;
(iv) every curve $\xi \in \mathrm{Max}(\mathcal{C}^{a,\leq 1}(\mathcal{M}))$ satisfies $I(\xi) = \mathbb{R}$;

If the class $LMin^{PAL}(\mathcal{M})$ of arc-length parametrized locally length-minimizing curves of $\mathcal{M}$ has the extension property, then (i), (ii), (iii) and (iv) are equivalent to

(v) every $\xi \in \mathrm{Max}(LMin^{PAL}(\mathcal{M}))$ satisfies $I(\xi) = \mathbb{R}$. □

The following proposition is also proved in the appendix:

4.8. Proposition. *If $\bar{x}, \hat{x} \in M$, $r \geq 0$, are such that $d_{\mathcal{M}}(\bar{x}, \hat{x}) \leq r$ and $B_{\mathcal{M}}(\hat{x}, r)$ is compact, then there exists a length minimizer ξ such that $\partial\xi = (\bar{x}, \hat{x})$.* □

4.9. Definition. Let S be a subset of M. We call S *relatively convex in $\mathcal{M}$* if any two points of S can be joined by a length minimizer ξ in $\mathcal{M}$. □

4.10. Theorem. *Every point of M has a neighborhood which is relatively convex in $\mathcal{M}$. If $\mathcal{M}$ is complete and connected, then M is relatively convex in $\mathcal{M}$.*

PROOF. The first assertion follows from Proposition 4.8: given $x \in M$, we can always find $r > 0$ such that $B_{\mathcal{M}}(x,r)$ is compact; then, if $\bar{x}$, $\hat{x}$ belong to $U = B_{\mathcal{M}}(x,\frac{r}{2})$, the ball $B_{\mathcal{M}}(\bar{x},\frac{r}{2})$ is closed and entirely contained in $B_{\mathcal{M}}(x,r)$. So $B_{\mathcal{M}}(\bar{x},\frac{r}{2})$ is compact. Therefore there exists a length minimizer joining $\bar{x}$ to $\hat{x}$.

The second assertion follows from Theorem 4.7: if $\mathcal{M}$ is complete, then (iii) of that theorem holds; given $\bar{x}$, $\hat{x}$, we can choose $x_j = \hat{x}$ for all j, and then (iii) yields a length minimizer ξ such that $\partial\xi = (\bar{x},\hat{x})$. ∎

§5. Extremals and geodesics.

The *Pontryagin Maximum Principle* (cf. Pontryagin *et al.* [28], Berkovitz [5], Lee-Markus [22] for the Maximum Principle in $\mathbb{R}^n$, Sussmann [35] for coordinate-free versions on manifolds, Strichartz [29,30] and Liu-Sussmann [24] for the application to subriemannian minimizers) gives a necessary condition for a curve $\xi : [a,b] \to M$ to be an energy minimizer. The condition says that ξ has to be an "extremal," in a sense that we now define.

5.1. Definition. If ξ is a curve in a manifold M, then a *cotangent lift* of ξ is a curve $\Xi : I(\xi) \to T^*M$ such that $\pi_{T^*M} \circ \Xi = \xi$. ∎

Clearly, a cotangent lift Ξ of a curve ξ is given by $\Xi(t) = (\xi(t),\psi(t))$ for $t \in I(\xi)$, where ψ is a *field of covectors along* ξ, i. e. a map such that $\psi(t) \in T^*_{\xi(t)}M$ for each $t \in I(\xi)$.

5.2. Definition. An *extremal* is an E-admissible E-curve ξ such that there exist a locally absolutely continuous cotangent lift $\Xi : I(\xi) \to T^*M$ and a constant ψ_0 that satisfy, if $\Xi(t) = (\xi(t),\psi(t))$, the conditions:

- *(E.1)* *$\psi_0 \geq 0$, and $(\psi(t),\psi_0) \neq (0,0)$ for all $t \in I(\xi)$,*
- *(E.2)* *for almost every $t \in I(\xi)$, the function $F\Big(\xi(t)\Big) \ni v \to \Big\langle \psi(t), \sigma_{\xi(t)}(v)\Big\rangle - \frac{\psi_0}{2}\Big\|v\Big\|^2$ is maximized by the value $v = \omega_\xi(t)$;*
- *(E.3)* *for every local orthonormal basis $\mathbf{f} = (f_1,\ldots,f_m)$ of sections of F, every coordinate chart $\mathbf{x} = (x^1,\ldots,x^n)$ of M, and every compact subinterval $[a,b]$ of $I(\xi)$ with the property that $\xi([a,b]) \subseteq U_{\mathbf{f}} \cap \mathrm{Dom}(\mathbf{x})$, if we write $g_i = \sigma(f_i)$, then the field of covectors ψ satisfies the "adjoint equation"*

$$\dot{\psi}(t) = -\sum_{i=1}^{m} \eta^i_{\mathbf{f},\xi,[a,b]}(t).\psi(t).\frac{\partial g_i}{\partial x}(\xi(t)) \quad \text{for a. e.} \quad t \in [a,b]\,, \tag{5.1}$$

relative to the chart $\mathbf{x}$. ∎

As stated, Condition (E.3) of Definition 5.2 appears to be highly nonintrinsic, because it seems to depend on the choice of $\mathbf{f}$ and $\mathbf{x}$. This, however, is not so, as we now show.

To state (E.3) in an intrinsic way, we first introduce the vector bundle $\hat{F}$ over T^*M obtained by pulling back F via the map $\pi_{T^*M} : T^*M \to M$. That is, $\hat{F}$ is the bundle over T^*M whose fiber $\hat{F}(x,p)$ over a point $(x,p) \in T^*M$ is just $F(x)$. The points of $\hat{F}$ are then the triples (x,p,v) such that $x \in M$, $p \in T_x^*M$, and $v \in F(x)$.

We then define the "control Hamiltonian" $H_{\mathcal{M}}^c : \hat{F} \times \mathbb{R} \to \mathbb{R}$ by letting

$$H_{\mathcal{M}}^c(x,p,v,p_0) \stackrel{\mathrm{def}}{=} \Big\langle p, \sigma_x(v) \Big\rangle - \frac{p_0}{2} ||v||_G^2 . \tag{5.2}$$

The *critical set* $Crit(\mathcal{M})$ is then the set of all $(x,p,v,p_0) \in \hat{F} \times \mathbb{R}$ that satisfy the equation $\frac{\partial H_{\mathcal{M}}^c}{\partial v}(x,p,v,p_0) = 0$. Then $Crit(\mathcal{M})$ is an intrinsic object, because, for fixed (x,p,p_0), v varies in the linear space $F(x)$, so the gradient of $H_{\mathcal{M}}^c$ with respect to v is well defined.

Now, if $(x,p,v,p_0) \in Crit(\mathcal{M})$, we define a tangent vector $\vec{H}_{\mathcal{M}}^c(x,p,v,p_0)$ as follows:

(Step 1) We choose, in an arbitrary fashion, a smooth section V of $\hat{F}$.
(Step 2) Define a real-valued function $H_{\mathcal{M}}^{c,V,p_0}$ on the cotangent bundle T^*M, by letting $H_{\mathcal{M}}^{c,V,p_0}(x',p') = H_{\mathcal{M}}^c(x',p',V(x',p'),p_0)$ for $(x',p') \in T^*M$.
(Step 3) By definition, then,

$$\vec{H}_{\mathcal{M}}^c(x,p,v,p_0) \stackrel{\mathrm{def}}{=} \vec{H}_{\mathcal{M}}^{c,V,p_0}(x,p) \tag{5.3}$$

where $\vec{H}_{\mathcal{M}}^{c,V,p_0}$ is the Hamiltonian vector field arising from the function $H_{\mathcal{M}}^{c,V,p_0}$.

To justify the definition, we need the following fact, proved in the appendix:

(∗) *the vector $\vec{H}_{\mathcal{M}}^c(x,p,v,p_0) \in T_{x,p}T^*M$ defined by (5.3) does not depend on the choice of the section V.*

The following proposition —proved in the appendix— then tells us that (E.3) is an intrinsic condition provided that (E.2) holds:

5.3. Proposition. *Let ξ be an E-admissible E-curve. Let $\psi : I(\xi) \to T^*M$ be a field of covectors along ξ and let ψ_0 be a constant such that, if $\Xi(t) = (\xi(t), \psi(t))$, then Condition (E.2) of Definition 5.2 holds. Then:*

(A) $(\xi(t), \psi(t), \omega_\xi(t), \psi_0) \in Crit(\mathcal{M})$ for almost every $t \in I(\xi)$;
(B) (E.3) holds if and only if $\dot{\Xi}(t) = \vec{H}_{\mathcal{M}}^c(\Xi(t), \omega_\xi(t), \psi_0)$ for a. e. $t \in I(\xi)$. □

The maximum principle now implies the following necessary condition for an arc to be an energy minimizer. The details of the derivation are given in the appendix.

5.4. Theorem. *If $\xi_* : [a,b] \to M$ is an energy minimizer, then ξ_* is an extremal.* □

5.5. Definition. A *normal extremal* of $\mathcal{M}$ is an extremal that satisfies the conditions of Definition 5.2 with a nonvanishing value of ψ_0. An *abnormal extremal* of $\mathcal{M}$ is an extremal that satisfies the conditions of Definition 5.2 with $\psi_0 = 0$. □

5.6. Remark. An extremal ξ can be both normal and abnormal, since it is possible for ξ to satisfy the conditions of Definition 5.2 with different choices of the pair (ψ, ψ_0) and, in particular, it can happen that $\psi_0 = 0$ for some of these pairs but not for all of them. □

5.7. Remark. *A normal extremal satisfies the conditions of Definition 5.2 with* $\psi_0 = 1$. Indeed, it is easy to see that, if a pair (ψ, ψ_0) is such that these conditions hold, and $r \in \mathbb{R}$ $r > 0$, then the pair $(r\psi, r\psi_0)$ also satisfies the conditions. □

It turns out that the normal extremals are projections on M of curves Ξ in T^*M that satisfy a Hamiltonian equation called the "geodesic equation," as we now explain.

If $p \in T^*_x M$, then the composite map $p \circ \sigma_x : F(x) \to \mathbb{R}$ is a linear functional on the inner product space $F(x)$, so $p \circ \sigma_x$ has a well-defined operator norm $||p||_G$, given by the formula $||p||_G = \sup\Big\{ <p, \sigma_x(v)> : v \in F(x), ||v||_G \le 1 \Big\}$.

5.8. Definition. The real-valued function $H_{\mathcal{M}}$ on T^*M given by $H_{\mathcal{M}}(x,p) = \frac{1}{2}||p||_G^2$ for $(x,p) \in T^*M$ is the *Hamiltonian* of $\mathcal{M}$. The integral curves of $\vec{H}_{\mathcal{M}}$ are the *solutions of the geodesic equation* of $\mathcal{M}$. The curves on M that are projections of solutions of the geodesic equation are the *geodesics*. We write $GEO(\mathcal{M})$, $\widehat{GEO}(\mathcal{M})$ to denote, respectively, the set of all geodesics and the set of all solutions of the geodesic equation of $\mathcal{M}$. □

It follows immediately from (2.5) that

$$H_{\mathcal{M}} \equiv \text{constant along every } \Xi \in \widehat{GEO}(\mathcal{M}). \tag{5.4}$$

If $\mathbf{f} = (f_1, \ldots, f_m)$ is a local orthonormal local basis of sections of F, and $(x,p) \in U^*_{\mathbf{f}}$, then $||p||_G$ is given by the formula $||p||_G^2 = \sum_{i=1}^m <p, g_i(x)>^2$, where $g_i = \sigma(f_i)$. Therefore

$$H_{\mathcal{M}}(x,p) = \frac{1}{2}\sum_{i=1}^m <p, g_i(x)>^2 . \tag{5.5}$$

Equivalently, if we use μ_X to denote, for a vector field X on M, the momentum function $(x,p) \to \mu_X(x,p) \stackrel{\text{def}}{=} <p, X(x)>$, defined earlier (cf. Formula (2.2)), then

$$H_{\mathcal{M}} = \frac{1}{2}\sum_{i=1}^m (\mu_{g_i})^2 \text{ on } U^*_{\mathbf{f}}. \tag{5.6}$$

It is clear that, relative to the chart $\mathbf{x}^{cot}$ of T^*M induced by a coordinate chart $\mathbf{x}$ of M,

$$\frac{\partial H_{\mathcal{M}}}{\partial p}(x,p) = \sum_{i=1}^m <p, g_i(x)>.g_i(x) \quad \text{and} \quad \frac{\partial H_{\mathcal{M}}}{\partial x}(x,p) = \sum_{i=1}^m <p, g_i(x)>.p.\frac{\partial g_i}{\partial x}(x), \tag{5.7}$$

where we are treating p as a row vector, and each $\frac{\partial g_i}{\partial x}(x)$ is a square matrix.

Then the geodesic equation, in coordinates, becomes the system:

$$\dot{x} = \sum_{i=1}^m <p, g_i(x)>.g_i(x), \qquad \dot{p} = -\sum_{i=1}^m <p, g_i(x)>.p.\frac{\partial g_i}{\partial x}(x). \tag{5.8}$$

This is a smooth system of first order ordinary differential equations on T^*M, so

5.9. Proposition. *Every solution of the geodesic equation is smooth.* $\square$

5.10. Remark. The geodesic equations (5.8) should really have been written in the form $\dot{x} = \sum_{i,j=1}^{m} \delta^{ij} <p, g_i(x)>.g_j(x)$, $\dot{p} = -\sum_{i,j=1}^{m} \delta^{ij} <p, g_i(x)>.p.\frac{\partial g_i}{\partial x}(x)$ —where δ^{ij} is Kronecker's delta— so as to preserve the usual summation conventions. $\square$

Suppose that $\Xi \in \widehat{GEO}(\mathcal{M})$, and write $\Xi(t) = (\xi(t), \psi(t))$. Then ξ is an E-admissible E-curve, having a control representation with control η given by

$$\eta(t) = \Big(\eta^1(t), \ldots, \eta^m(t)\Big) = \tilde{\varphi}_{\mathbf{f}}\Big(\Xi(t)\Big) = \Big(\tilde{\varphi}^1_{\mathbf{f}}\Big(\Xi(t)\Big), \ldots, \tilde{\varphi}^m_{\mathbf{f}}\Big(\Xi(t)\Big)\Big),$$

where $\tilde{\varphi}^i_{\mathbf{f}} = \varphi_{\mathbf{f},i}$ (that is, $\tilde{\varphi}^i_{\mathbf{f}} = \sum_{j=1}^{m} \delta^{ij} \varphi_{\mathbf{f},j}$, cf. Remark 5.10), and the $\varphi_{\mathbf{f},i}$ are just the momentum functions μ_{g_i}, so that

$$\varphi_{\mathbf{f},i}(x,p) \stackrel{\text{def}}{=} <p, g_i(x)>. \tag{5.9}$$

We prove in the appendix that

5.11. Proposition. *If* $\mathbf{x}$ *is a chart of* M, $\mathbf{f}$ *is a local orthonormal basis of sections of* F, $\xi \in \mathcal{C}_{AC,loc}(M)$, $I(\xi) \subseteq U_{\mathbf{f}} \cap \mathrm{Dom}(\mathbf{x})$, ψ *is a field of covectors along* ξ, *and* $\dot{\xi}(t) = \sum_{i=1}^{m} \tilde{\varphi}^i_{\mathbf{f}}(\xi(t), \psi(t)) g_i(\xi(t))$ *a. e., then the control* $t \to \tilde{\varphi}_{\mathbf{f}}(\xi(t), \psi(t))$ *is the canonical control representation* $\eta_{\mathbf{f},\xi,I(\xi)}$ *of* ξ *with respect to* $\mathbf{f}$ *on* $I(\xi)$. $\square$

The field of covectors ψ is then a solution of $\dot{\psi}(t) = -\psi(t).\sum_{i=1}^{m} \eta^i_{\mathbf{f},\xi,I(\xi)}(t).\frac{\partial g_i}{\partial x}(\xi(t))$, which is precisely the adjoint equation (5.1). So, if we let $\psi_0 \equiv 1$, then Conditions (E.1) and (E.3) of Definition 5.2 are satisfied. Moreover, it is clear that the quadratic form $\mathbb{R}^m \ni (u^1, \ldots, u^m) \to \sum_{i=1}^{m} u^i <p, g_i(\xi(t))> - \frac{\psi_0}{2}||u||^2 \in \mathbb{R}$ is maximized for the values $u^i = <p, g_i(\xi(t))> = \eta^i_{\mathbf{f},\xi,I(\xi)}(t)$ of the u variables, so (E.2) holds as well.

Conversely, suppose ξ is a normal extremal, and let ψ be such that Conditions (E.1,2,3) of Definition 5.2 hold with $\psi_0 = 1$. Write $\Xi(t) = (\xi(t), \psi(t))$. Then the minimization condition implies that $\eta_{\mathbf{f},\xi,I(\xi)}(t) = \tilde{\varphi}_{\mathbf{f}}(\Xi(t))$, so Ξ is a solution of the geodesic equation, and ξ is a geodesic. We have therefore proved:

5.12. Proposition. *A curve* ξ *is a geodesic if and only if it is a normal extremal.* $\square$

So far, we have found a necessary condition for an arc to be a minimizer, namely, that it be an extremal. This condition is not in general sufficient, even for local minimization, but the following weaker fact is true:

5.13. Theorem. *Every geodesic curve is a local energy minimizer.*

PROOF. This is a standard fact in optimal control theory, proved by the technique of "fields of extremals," i. e. by using the method of characteristics to solve the Hamilton-Jacobi equation. For a detailed proof, see Hamenstädt [14], or Liu-Sussmann [24], where the proof, although it is only given for the case of a nonsingular distribution E, works equally well for the general case of possibly singular distributions. $\square$

5.14. Theorem. $GEO(\mathcal{M})$ *has the extension property (cf. Definition 4.1).*

PROOF. Suppose that $\xi \in \mathrm{Max}(GEO(\mathcal{M}))$, $I = I(\xi)$, and τ is a limit point of I such that the limit $x = \lim_{t\to\tau, t\in I}\xi(t)$ exists. Let $\Xi \in \widehat{GEO}(\mathcal{M})$ be such that $\pi_{T^*M}\circ\Xi = \xi$, and write $\Xi(t) = (\xi(t), \psi(t))$ for $t \in I$. We have to prove that t is an interior point of I. Pick a neighborhood U of x which is contained in the domain of a coordinate chart $\mathbf{x}$ and in that of an orthonormal basis of sections $\mathbf{f} = (f_1, \ldots, f_m)$ of F. Then the right-hand side of (5.8) is bounded along Ξ for $t \in]\tau-\varepsilon, \tau+\varepsilon[\cap I$, for some $\varepsilon > 0$, because of Gronwall's inequality and the fact that $\sum_{i=1}^m <\psi(t), g_i(\xi(t))>^2 =$ constant. Therefore $\lim_{t\to\tau,t\in I}\Xi(t)$ exists. So, by standard properties of ordinary differential equations, Ξ extends to a full neighborhood $]\tau-\varepsilon, \tau+\varepsilon[$ of τ, and then ξ extends to a geodesic defined on $]\tau-\varepsilon, \tau+\varepsilon[\cup I$. Since $\xi \in \mathrm{Max}(GEO(\mathcal{M}))$, we have $]\tau-\varepsilon, \tau+\varepsilon[\subseteq I$, so τ is an interior point of I, as desired.□

§6. Strongly bracket generating subriemannian structures.

If $x \in M$, we let $E(x)^\perp \stackrel{\text{def}}{=} E^\perp(x) \stackrel{\text{def}}{=} \Big\{p \in T_x^*M : <p, v> = 0 \text{ for all } v \in E(x)\Big\}$. For each point $(x,p) \in T^*M$ such that $p \in E(x)^\perp$, we define a bilinear form $Z(x,p)$ on $F(x)$ as follows. Let $v, w \in F(x)$, and let V, W be smooth sections of F such that $V(x) = v$ and $W(x) = w$. Then $\sigma(V)$ and $\sigma(W)$ are smooth vector fields on M, so the Lie bracket $[\sigma(V), \sigma(W)]$ is well defined, and we can let $Z(x,p)(v,w) = \Big\langle p, [\sigma(V), \sigma(W)]\Big\rangle$. The number $Z(x,p)(v,w)$ does not depend on the choice of the sections V, W. (If, for example, V_1 is another section of F such that $V_1(x) = v$, then we can choose a local basis $\mathbf{f} = (f_1, \ldots, f_m)$ of sections of F near x, and write $V_1 - V = \sum_i \zeta^i f_i$ near x, where the ζ^i are smooth functions that vanish at x. If $g_i = \sigma(f_i)$, then $\sigma(V_1) - \sigma(V) = \sum_i \zeta^i g_i$, so

$$[\sigma(V_1) - \sigma(V), \sigma(W)] = \sum_i \zeta^i [g_i, \sigma(W)] - \sum_i (\sigma(W)\zeta^i).g_i\,. \tag{6.1}$$

Since $<p, g_i(x)> = 0$ —because $p \in E(x)^\perp$— and $\zeta^i(x) = 0$, it follows from (6.1) that $<p, [\sigma(V_1), \sigma(W)](x)> = <p, [\sigma(V), \sigma(W)](x)>$, as desired.)

It is then clear that $Z(x,p)$ is always skew-symmetric.

6.1. Definition. We call E —and $\mathcal{M}$— *strongly bracket-generating* (SBG) at a point $x \in M$ if the bilinear form $Z(x,p)$ is nonsingular for every $p \in E(x)^\perp\backslash\{0\}$. We call E —and $\mathcal{M}$— *strongly bracket-generating* if it is SBG at every point of M. □

6.2. Remark. *A Riemannian manifold —or, more generally, a subriemannian manifold for which σ is surjective, cf. Remark 3.28— is always SBG*, because in the Riemannian case the condition of Definition 6.1 holds vacuously, since $E(x)^\perp = \{0\}$. □

The definition given above of the SBG condition is manifestly intrinsic. We now give a local characterization of the SBG condition in terms of a choice of a basis of sections.

Suppose $\mathbf{f} = (f_1, \ldots, f_m)$ is a local basis of smooth sections of F. Write $g_i = \sigma(f_i)$. Define a vector field $\mathbf{g}\theta$, for each $\theta = (\theta^1, \ldots, \theta^m) \in \mathbb{R}^m$, by letting $\mathbf{g}\theta = \theta^1 g_1 + \ldots + \theta^m g_m$. We prove in the appendix that

6.3. Proposition. *A singular distribution $E = (F, \sigma)$ is strongly bracket-generating at a point x in the domain $U_{\mathbf{f}}$ of a local basis of sections of F if and only if*

(SBG$_{loc}$) *for every $\theta = (\theta^1, \ldots, \theta^m) \in \mathbb{R}^m \backslash \{0\}$, the tangent space $T_x M$ is spanned by the $2m$ vectors $g_1(x)$, $\ldots$, $g_m(x)$, $[\mathbf{g}\theta, g_1](x)$, $\ldots$, $[\mathbf{g}\theta, g_m](x)$.* □

6.4. Corollary. *If a singular distribution $E = (F, \sigma)$ on a manifold M is strongly bracket-generating at a point x then E is 2-generating at x.* □

The following elementary observation explains the importance of the SBG condition:

6.5. Theorem. *If $\mathcal{M}$ is SBG, then a nonconstant extremal cannot be abnormal.*

PROOF. We first prove the following "coordinate free version of the adjoint equation" (cf. Sussmann [35]):

6.6. Proposition. *Let $\mathbf{f} = (f_1, \ldots, f_m)$ be a local orthonormal basis of sections of F, and let $g_i = \sigma(f_i)$. Let $\xi : [a, b] \to U_{\mathbf{f}}$ be an E-admissible E-arc. Let η be the canonical control representation $\eta_{\mathbf{f},\xi,[a,b]}$ of ξ on $[a, b]$ relative to $\mathbf{f}$, so $\eta(t) = (\eta^1(t), \ldots, \eta^m(t))$, and $\sum_{i=1}^m \eta^i(t) f_i(\xi(t)) = \omega_\xi(t)$. Let ψ be an absolutely continuous field of covectors along ξ that satisfies the adjoint equation (5.1). Let $\Xi(t) = (\xi(t), \psi(t))$. Then*

$$\frac{d}{dt}\Big\langle \psi(t), X(\xi(t)) \Big\rangle = \sum_{i=1} \eta^i(t) . \Big\langle \psi(t), [g_i, X](\xi(t)) \Big\rangle \quad \text{a. e.} \tag{6.2}$$

—that is, $\frac{d}{dt}\Big(\mu_X(\Xi(t))\Big) = \sum_{i=1} \eta^i(t) . \mu_{[g_i, X]}(\Xi(t))$— for every smooth vector field X on M. In particular,

$$\frac{d}{dt}\Big(\varphi_{\mathbf{f},j}(\Xi(t))\Big) = \sum_{i=1}^m \eta^i(t) . \zeta_{\mathbf{f},ij}(\Xi(t)) \ \text{a.e.} \tag{6.3}$$

for $j = 1, \ldots, m$, where

$$\zeta_{\mathbf{f},ij}(x, p) \stackrel{\text{def}}{=} \langle p, [g_i, g_j](x) \rangle . \tag{6.4}$$

Proof of Proposition 6.6. In coordinates, the adjoint equation says that (6.2) holds when $X = \frac{\partial}{\partial x^i}$, $i = 1, \ldots,$. An elementary computation shows that the set of smooth vector fields X for which (6.2) holds is a $C^\infty(M)$-module. So (6.2) holds for all X. □

Proof of Theorem 6.5. Let ξ be an abnormal extremal. We want to prove that $\xi(t)$ is constant. It suffices to work locally, so we may assume that ξ is contained in the domain $U_{\mathbf{f}}$ of an orthonormal basis $\mathbf{f} = (f_1, \ldots, f_m)$ of sections of F. Let ψ be such that the conditions of Definition 5.2 hold with $\psi_0 = 0$, and let $\Xi(t) = (\xi(t), \psi(t))$. Then the Hamiltonian maximization condition implies that the functions $t \to \varphi_{\mathbf{f},i}(\Xi(t))$ vanish identically, so the left-hand side of (6.3) vanishes identically. The fact that $\psi(t) \neq 0$ (which is a consequence of (E.1), since $\psi_0 = 0$) implies then, since $\mathcal{M}$ is SBG, that the matrix $\Big(\zeta_{\mathbf{f},ij}(\Xi(t))\Big)_{ij}$ is nonsingular for each t. So $\eta \equiv 0$, and then ξ is a constant curve. □

6.7. Theorem. *Assume $\mathcal{M}$ is SBF. Let ξ be a nonconstant geodesic. Then the cotangent lift $\Xi \in \widehat{GEO}(\mathcal{M})$ of ξ is unique.*

PROOF. Clearly, we may work locally, so we will assume that $M = \mathbb{R}^n$ and there exists a a global orthonormal basis $\mathbf{f} = (f_1, \ldots, f_m)$ of sections of F. Write $g_i = \sigma(f_i)$, and let $\Xi_j = (\xi, \psi_j(t))$, $j = 1, 2$, be two distinct solutions of the geodesic equation. Then $\psi_1(t) \neq \psi_2(t)$ for every t because if $\psi_1(t) = \psi_2(t)$ for some t then $\Xi_1 \equiv \Xi_2$ by the uniqueness of solutions of ordinary differential equations.

Then, if $\ell \in \{1, 2\}$, $\dot{\xi}(t) = \sum_{i=1}^m \varphi_{\mathbf{f}}^i(\Xi_\ell(t)) g_i(\xi(t))$. Proposition 5.11 then tells us that both vector functions $t \to \tilde{\varphi}_{\mathbf{f}}(\Xi_j(t))$ are equal to the canonical control representation $\eta_{\mathbf{f},\xi,I(\xi)}$. So $\varphi_{\mathbf{f}}(\Xi_1(t)) = \varphi_{\mathbf{f}}(\Xi_2(t))$ for a. e. t. We then use Equation (6.3) and conclude that $\frac{d}{dt}\Big(\varphi_{\mathbf{f},j}(\Xi_\ell(t))\Big) = \sum_{i=1}^m \eta_{\mathbf{f},\xi}^i(t) \zeta_{\mathbf{f},ij}(\xi, \psi_\ell(t)))$ a. e. for $j = 1, \ldots, m$, $\ell = 1, 2$. So

$$\sum_{i=1}^{m} \eta_{\mathbf{f},\xi}^i(t) \Big(\zeta_{\mathbf{f},ij}(\xi(t), \psi_1(t) - \psi_2(t))\Big) = 0 \tag{6.5}$$

for $j = 1, \ldots, m$. For each t, $\psi_1(t) \neq \psi_2(t)$ but $\varphi_{\mathbf{f}}(\Xi_1(t)) = \varphi_{\mathbf{f}}(\Xi_2(t))$. This implies that $\psi_1(t) - \psi_2(t) \neq 0$ but $\varphi_{\mathbf{f}}(\xi, \psi_1(t) - \psi_2(t)) = 0$. It then follows from the SBG condition that the matrix $\Big(\zeta_{\mathbf{f},ij}(\xi(t), \psi_1(t) - \psi_2(t))\Big)_{ij}$ is nonsingular. But then (6.5) implies that $\eta_{\mathbf{f},\xi,I(\xi)}(t) \equiv 0$. So $\xi(t) \equiv$ constant. □

6.8. Remark. The conclusion of Theorem 6.7 is obviously false if ξ is a constant curve, unless σ is surjective. Indeed, if $x \in M$ is such that σ_x is not surjective, then we can choose I arbitrarily, and let $\xi : I \to M$ be the constant curve such that $\xi(t) \equiv x$. Pick a p such that $<p, \sigma_x(v)> = 0$ for all $v \in F(x)$, and define $\psi(t) \equiv p$. Then $t \to (\xi(t), \psi(t))$ is in $\widehat{GEO}(\mathcal{M})$, and it is clear that p is not unique. □

6.9. Theorem. *Let $\mathcal{M}$ be SBG, and let ξ be a curve in M. Then ξ is a local energy minimizer if and only if it is a geodesic curve.*

PROOF. The conclusion is evidently true if $I(\xi)$ is trivial, since in this case ξ is both a geodesic and an energy minimizer.

Now suppose that $I(\xi)$ is nontrivial. The fact that if ξ is a geodesic curve then ξ is a local energy minimizer is a general property of subriemannian manifolds, as shown earlier in Theorem 5.13. To prove the converse, we assume that ξ is a local energy minimizer. This means that for every $t \in I$ there exists a compact subinterval $[a_t, b_t]$ of I such that the restriction $\xi_t \stackrel{\text{def}}{=} \xi \lceil [a_t, b_t]$ is an energy minimizer, and t belongs to the interior J_t of $[a_t, b_t]$ relative to I. Then each ξ_t is parametrized by constant times arc-length, so for each t there exists a unique $L_t \in \mathbb{R}$ such that $||\dot{\xi}(t)||_G = L_t$ for a. e. $t \in J_t$. (The uniqueness follows because J_t is nontrivial, since I is.) It then follows that, if $J_t \cap J_s \neq \emptyset$, then $L_t = L_s$. (Indeed, $J_t \cap J_s$ is relatively open in I, and nonempty, so it is nontrivial.) Since I is connected, all the L_t are equal, so there exists $L \in \mathbb{R}$ such that $||\dot{\xi}(t)||_G = L$ for a. e. $t \in I$. Therefore ξ is parametrized by constant times arc-length.

Now, if $L = 0$ then $\xi \equiv$ constant, so $\xi = \pi_{T^*M} \circ \Xi$, where $\Xi(t) = (\xi(t), 0)$ for all $t \in I(\xi)$. Clearly, $\Xi \in \widehat{GEO}(\mathcal{M})$, so ξ is a geodesic.

If $L \neq 0$ then every curve ξ_t is nonconstant, so it is a geodesic. Moreover, the adjoint vector $J_t \ni s \to \psi_t(s) \in T^*_{\xi(s)}M$ such that the curve $[a_t, b_t] \ni s \to \Xi_t(s) = (\xi(s), \psi_t(s))$ is a geodesic is unique by Theorem 6.7. If $J_t \cap J_s \neq \emptyset$, then ψ_t and ψ_s must agree on $J_t \cap J_s \neq \emptyset$, using Theorem 6.7 once again. So there exists a map $I \ni t \to \psi(t) \in T^*_{\xi(t)}M$ such that $\psi(s) = \psi_t(s)$ whenever $s \in J_t$. Then ψ is locally absolutely continuous, and the curve $I \ni t \to \Xi(t) = (\xi(t), \psi(t))$ is in $\widehat{GEO}(\mathcal{M})$. So ξ is a geodesic. ☐

6.10. Corollary. *If $\mathcal{M}$ is SBG, then $LMin^{PAL}(\mathcal{M})$ has the extension property.*

PROOF. Let $\xi \in \mathrm{Max}(LMin^{PAL}(\mathcal{M}))$. Let τ be a limit point of $I(\xi)$ such that the limit $x = \lim_{t\to\tau, t\in I} \xi(t)$ exists. Then ξ is a local energy minimizer, so it is a geodesic by Theorem 6.9. If $\tilde{\xi} : \tilde{I} \to M$ was a geodesic that extends ξ to a strictly larger interval, then $\tilde{\xi}$ would also be a local energy minimizer, by Theorem 6.9 (or by Theorem 5.13), contradicting the maximality of ξ within $LMin^{PAL}(\mathcal{M})$. So $\tilde{\xi}$ cannot exist, and then $\xi \in \mathrm{Max}(GEO(\mathcal{M}))$. By Theorem 5.14, $GEO(\mathcal{M})$ has the extension property. Since $\xi \in \mathrm{Max}(GEO(\mathcal{M}))$ and τ is a limit point of $I(\xi)$ such that $\lim_{t\to\tau, t\in I} \xi(t)$ exists, it follows that τ is an interior point of $I(\xi)$. ☐

Thanks to Theorem 6.9 and Corollary 6.10, the subriemannian Hopf-Rinow theorem takes a particularly form in the SBG case:

6.11. Theorem. *If $\mathcal{M}$ is SBG then the following six conditions are equivalent:*

- *(i)* *$\mathcal{M}$ is complete;*
- *(ii)* *every closed ball $B_{\mathcal{M}}(x, r)$ is compact;*
- *(iii)* *if $\{x_j\}_{j=1}^{\infty}$ is any sequence of points of M whose distances to a given point $\bar{x}$ are bounded above by a finite constant, then there exist a subsequence of $\{x_j\}_{j=1}^{\infty}$ that converges to a limit x, and a length minimizer ξ such that $\partial\xi = (\bar{x}, x)$;*
- *(iv)* *every $\xi \in \mathrm{Max}(\mathcal{C}^{a,\leq 1}(\mathcal{M}))$ satisfies $I(\xi) = \mathbb{R}$;*
- *(v)* *every $\xi \in \mathrm{Max}(LMin^{PAL}(\mathcal{M}))$ satisfies $I(\xi) = \mathbb{R}$;*
- *(vi)* *every $\xi \in \mathrm{Max}(GEO(\mathcal{M}))$ satisfies $I(\xi) = \mathbb{R}$.* ☐

§7. The main estimate.

We are now ready to prove our main adjoint vector estimate, and derive some consequences. If $S \subseteq M$, we let $\mathcal{G}_{nc,min}(S)$ denote the set of all $\Xi \in C^0([0,1], \pi^{-1}_{T^*M}(K)) \cap \widehat{GEO}(\mathcal{M})$ such that the projection $\xi = \pi_{T^*M} \circ \Xi$ is a nonconstant energy minimizer.

The following result is then our adjoint vector estimate:

7.1. Theorem. *Let $\mathcal{M} = (M, F, \sigma, G)$ be an SBG subriemannian manifold, and let K be a compact subset of M. Then $\mathcal{G}_{nc,min}(K)$ is a relatively compact subset of $C^0([0,1], T^*M)$.*

PROOF. It suffices to show that

(A) *for every compact subset K of M there exists a compact subset J_K of T^*M such that $\mathcal{G}_{nc,min}(K) \subseteq C^0([0,1], J_K)$.*

Indeed, if (A) holds, then $\mathcal{G}_{nc,min}(K)$ is uniformly bounded. In addition $\mathcal{G}_{nc,min}(K)$ is equicontinuous, because all its members are integral arcs of the vector field $\vec{H}_{\mathcal{M}}$. So $\mathcal{G}_{nc,min}(K)$ is relatively compact in $C^0([0,1],T^*M)$ by the Ascoli-Arzelà theorem.

We now prove (A). First of all, we show that the desired result is local:

(LOC) if $\bar{K}$ is a compact subset of M such that every point $x \in \bar{K}$ has a compact neighborhood K_x for which there exists a compact subset J_x of T^*M with the property that $\mathcal{G}_{nc,min}(K_x) \subseteq C^0([0,1],J_x)$, then there exists a compact subset $\bar{J}$ of T^*M such that $\mathcal{G}_{nc,min}(\bar{K}) \subseteq C^0([0,1],\bar{J})$.

To prove (LOC), we pick a covering of $\bar{K}$ by finitely many open sets $U_1,\ldots,U_\nu$ whose closures $K_1,\ldots,K_\nu$ are compact and such that there are compact subsets J_i of T^*M with the property that $\mathcal{G}_{nc,min}(K_i) \subseteq C^0([0,1],J_i)$. We then let $\delta > 0$ be a Lebesgue number for the covering $\{U_1,\ldots,U_\nu\}$ of $\bar{K}$ relative to the metric $d_{\mathcal{M}}$, i.e. a number such that every subset of $\bar{K}$ of $d_{\mathcal{M}}$-diameter $\leq \delta$ is contained in one of the U_i. We let Λ be the $d_{\mathcal{M}}$-diameter of $\bar{K}$, and pick a positive integer N such that $\Lambda \leq N\delta$.

Let $\Xi \in \mathcal{G}_{nc,min}(\bar{K})$, and write $\xi = \pi_{T^*M} \circ \Xi$. Then ξ is parametrized by constant times arc-length, so $||\dot{\xi}(t)||_G = ||\xi||_G$ for all $t \in [0,1]$. Clearly, $0 < ||\xi||_G \leq \Lambda$. So, if we let I_k be the closed interval $\left[\frac{k-1}{N},\frac{k}{N}\right]$, for $k = 1,\ldots,N$, and use ξ_k to denote the restriction of ξ to I_k, then each ξ_k is entirely contained in one of the sets U_i, since $d_{\mathcal{M}}(\xi(t),\xi(t')) \leq \Lambda N^{-1} \leq \delta$ whenever $t,t' \in I_k$. Let $\theta_k(t) = \xi\left(\frac{k-1+t}{N}\right)$ for $t \in [0,1]$. Then θ_k is the canonical reparametrization of ξ_k. So

$$||\dot{\theta}_k(t)||_G = \frac{||\xi||_G}{N} \quad \text{for a.e.} \quad t\,. \tag{7.1}$$

Let $\Xi(t) = (\xi(t),\psi(t))$ for $t \in [0,1]$. Then, if we define $\psi_k(t) = \frac{1}{N}\cdot\psi\left(\frac{k-1+t}{N}\right)$ for $t \in [0,1]$, an easy computation shows that each of the curves $t \to (\theta_k(t),\psi_k(t))$ is in $\widehat{GEO}(\mathcal{M})$. Moreover, θ_k is nonconstant, because (7.1) implies that $||\dot{\theta}_k(t)||_G \neq 0$ for a. e. t. So $(\theta_k(t),\psi_k(t)) \in J_k$ for $t \in [0,1]$, $k = 1,\ldots,N$. Let $\bar{J} = \bigcup_{k=1}^{\nu}\left\{(x,p) : \left(x,\frac{p}{N}\right) \in J_k\right\}$. Then Ξ is entirely contained in $\bar{J}$. So (LOC) is proved.

We now assume that M is in fact $\mathbb{R}^n$, K is the closed unit ball $\{x \in \mathbb{R}^n : ||x|| \leq 1\}$, and F has a global orthonormal basis $\mathbf{f} = (f_1,\ldots,f_m)$ of sections. From now on, we fix the local basis $\mathbf{f}$, and write $g_i = \sigma(f_i)$. We write $\tilde{\varphi}^i$, φ_i, ζ_{ij}, rather than $\tilde{\varphi}^i_{\mathbf{f}}$, $\varphi_{\mathbf{f},i}$, $\zeta_{\mathbf{f},ij}$, so $\varphi_i(x,p) = \langle p, g_i(x)\rangle$ and $\zeta_{ij}(x,p) = \langle p,[g_i,g_j](x)\rangle$. Also, we use $\Phi(x,p)$, $Z(x,p)$ to denote, respectively, the column vector with components $\tilde{\varphi}^1(x,p),\ldots,\tilde{\varphi}^m(x,p)$ and the $m \times m$ matrix $\Big(\zeta_{ij}(x,p)\Big)_{1\leq i,j\leq m}$.

Now T^*M is just the product $\mathbb{R}^n \times \mathbb{R}^n$. For vectors $v \in \mathbb{R}^n$, $||v||$ will always mean Euclidean norm. For matrices A in $\mathbb{R}^{\mu\times\nu}$, $||A||$ will denote operator norm, corresponding to regarding A as linear map from $\mathbb{R}^\nu$ to $\mathbb{R}^\mu$. We let $\tilde{K}$ be the ball $\{x \in \mathbb{R}^n : ||x|| \leq 2\}$. Using Proposition 3.12, we pick positive constants Δ_1, Δ_2, such that

$$\Delta_1||x-x'|| \leq d_{\mathcal{M}}(x,x') \leq \Delta_2||x-x'||^{1/2} \quad \text{for all} \quad x,x' \in \tilde{K}\,. \tag{7.2}$$

Our goal is to prove that an *priori* bound

$$\kappa_+(\Xi) \stackrel{\text{def}}{=} \sup\Big\{||\psi(t)|| : 0 \le 1\Big\} \le C \tag{7.3}$$

holds for all $\Xi = (\xi, \psi) \in \mathcal{G}_{nc,min}(K)$, with a constant C that does not depend on Ξ.

So from now on we fix a $\Xi \in \mathcal{G}_{nc,min}(K)$, write $\Xi(t) = (\xi(t), \psi(t))$, and try to prove (7.3), making sure that C that does not depend on Ξ. We let $\eta : [0,1] \to \mathbb{R}^m$ be the canonical control representation $\eta_{\mathbf{f},\xi,[0,1]}$ of ξ with respect to $\mathbf{f}$. We let $L(\xi) = ||\xi||_G$, and write $\partial\xi = (\bar{x}, \hat{x})$, so $\bar{x} = \xi(0)$, $\hat{x} = \xi(1)$, and $d_{\mathcal{M}}(\bar{x}, \hat{x}) = L(\xi)$.

We let Λ denote the diameter of K relative to the metric $d_{\mathcal{M}}$. We then know that $\eta(t) = \Phi(\Xi(t))$, i. e. that $\eta^i(t) = \tilde{\varphi}^i(\Xi(t))$ for $i = 1, \ldots, m$, and that $||\eta(t)||$ is a. e. constant and equal to $L(\xi)$, so $||\eta(t)|| \le \Lambda$ for a. e. t. Then

$$\Big|<\psi(t), g_i(\xi(t)>\Big| = |\varphi_i(\Xi(t))| = |\eta^i(t)| \le L(\xi) \le \Lambda \text{ for all } t \in [0,1]\,, i \in \{1, \ldots, m\}\,. \tag{7.4}$$

This gives us an *a priori* bound on the components of $\psi(t)$ in the directions of the vector fields g_i. To get (7.3), we need —since $\mathcal{M}$ is 2-generating (cf. Corollary 6.4)— an *a priori* bound on the components of $\psi(t)$ in the direction of the brackets $[g_i, g_j]$, i. e. on the entries $\zeta_{ij}(\Xi(t))$ of the matrix $Z(\Xi(t))$.

The adjoint equation (5.1), together with the bound (7.4), implies that

$$||\dot{\psi}(t)|| \le \Lambda.\Theta_1.||\psi(t)||\,, \quad \text{where} \quad \Theta_1 \stackrel{\text{def}}{=} \sup\left\{\sum_{i=1}^{m} \left\|\frac{\partial g_i}{\partial x}(x)\right\| : x \in K\right\}. \tag{7.5}$$

So $||\psi(t)|| \le e^{\Lambda.\Theta_1}||\psi(t')||$ whenever $t, t' \in [0,1]$. It follows that

$$\kappa_+(\Xi) \le e^{\Lambda.\Theta_1}.\kappa_-(\Xi)\,, \quad \text{where} \quad \kappa_-(\Xi) \stackrel{\text{def}}{=} \min\Big\{||\psi(t)|| : 0 \le 1\Big\}\,. \tag{7.6}$$

Therefore, it suffices to derive an *a priori* bound for $\kappa_-(\Xi)$.

Formula (6.3) says that

$$\frac{d}{dt}\Bigg(\Phi(\Xi(t))\Bigg) = -Z(\Xi(t)).\Phi(\Xi(t)) \quad \text{for} \quad \text{a.e.} \quad t\,. \tag{7.7}$$

We now write

$$\hat{x} - \bar{x} = \xi(1) - \xi(0) = \int_0^1 \mathbf{g}(\xi(t)).\Phi(\Xi(t))\,dt \stackrel{\text{def}}{=} I \tag{7.8}$$

where, if $x \in \mathbb{R}^n$, $\mathbf{g}(x)$ is the $n \times m$ matrix whose columns are the vectors $g_1(x), \ldots, g_m(x)$, so that, if $u = (u^1, \ldots, u^m)$ is any column m-vector, then $\mathbf{g}(x).u$ stands for $\sum_{i=1}^m u^i g_i(x)$.

It follows from (7.2) that

$$L(\xi)^2 = d_{\mathcal{M}}(\bar{x}, \hat{x})^2 \le \Delta_2^2||\hat{x} - \bar{x}||\,. \tag{7.9}$$

We would like to use (7.8) to prove a bound of the form $||\hat{x}-\bar{x}|| \leq \text{constant}.L(\xi)^2\kappa_-(\xi)^{-1}$. Since $L(\xi) \neq 0$, this bound, together with (7.9), would imply an upper bound for $\kappa_-(\xi)$.

First of all, it is clear that

$$||\hat{x}-\bar{x}|| \leq \Theta_2.L(\xi)\,, \quad \text{where} \quad \Theta_2 = \sup\left\{\sum_{i=1}^{m} ||g_i(x)|| : x \in \tilde{K}\right\}. \tag{7.10}$$

This is not yet the $L(\xi)^2$ estimate we are seeking, but will be useful later.

In order to get a better bound, we use integration by parts to rewrite the integral of (7.8). The idea is to write $\Phi = Z^{-1}.Z.\Phi = -Z^{-1}.\dot{\Phi}$, so that the integral will become $\int_0^1 \frac{d}{dt}\Big(Z(\Xi(t))^{-1}\Big)\Phi(\Xi(t))\,dt$ + boundary terms. The derivative of $Z(\Xi(t))^{-1}$ will involve the control —i. e. $\Phi(\Xi(t))$— again, and this will give an $L(\xi)^2$-bound for $||\hat{x}-\bar{x}||$.

This, however, does not work directly as explained, because $Z(\Xi(t))$ may fail to be invertible. So we need a better argument, for which some preliminary work is required.

We let $S = \mathbb{R}^n \times (\mathbb{R}^n \backslash \{0\})$, and use Σ_k, for any integer k, to denote the set of all smooth real-valued functions h on S that are positively homogeneous of degree k with respect to p, i. e. such that $h(x,rp) = r^k h(x,p)$ whenever $(x,p) \in S$ and $r > 0$.

In order to prove our bound, we will assume, to begin with, that

$$\kappa_-(\Xi) \geq 1\,. \tag{7.11}$$

This implies, in particular, that Ξ is entirely contained in S.

We let $D(x,p)$ denote the determinant of the matrix $Z(x,p)$, so $D \in \Sigma_m$. We then define $A_1(x,p) = ||p||^{-2m}\Big(||\Phi(x,p)||^{2m} + D(x,p)^2\Big)$ for $p \neq 0$. Then $A_1 \in \Sigma_0$. The SBG condition implies that $A_1(x,p)$ never vanishes. So we can define $A_2(x,p) = A_1(x,p)^{-1}$, and then $A_2 \in \Sigma_0$ as well. We then rewrite the integral of (7.8) as

$$\int_0^1 \mathbf{g}(\xi(t)).\Phi(\Xi(t))\,dt = I_1 + I_2\,,$$

where

$$I_1 = \int_0^1 \frac{||\Phi(\Xi(t))||^{2m}}{||\psi(t)||^{2m}}.A_2(\Xi(t)).\mathbf{g}(\xi(t)).\Phi(\Xi(t))\,dt\,,$$

and

$$I_2 = \int_0^1 \frac{D(\Xi(t))^2}{||\psi(t)||^{2m}}.A_2(\Xi(t)).\mathbf{g}(\xi(t)).\Phi(\Xi(t))\,dt\,.$$

Since $A_2 \in \Sigma_0$, there is a constant Θ_3 such that the bound $|A_2(x,p)| \leq \Theta_3$ holds for all $(x,p) \in K \times (\mathbb{R}^n \backslash \{0\})$. Then

$$||I_1|| \leq \Theta_2\Theta_3\frac{L(\xi)^{2m+1}}{\kappa_-(\Xi)^{2m}} \leq \Theta_2\Theta_3\Lambda^{2m-1}\frac{L(\xi)^2}{\kappa_-(\Xi)}\,,$$

where we have used the facts that $L(\xi) \leq \Lambda$, $m \geq 1$, and $\kappa_-(\Xi) \geq 1$.

We now need to estimate I_2. For this purpose, we let $\tilde{Z}(x,p)$ denote the matrix of complementary minors of $Z(x,p)$, so $\tilde{Z}(x,p)$ is an $m \times m$ matrix whose components belong to Σ_{m-1}, and $\tilde{Z}(x,p).Z(x,p) = D(x,p).\mathbf{1}_m$, where $\mathbf{1}_m$ is the $m \times m$ identity matrix. Then

$$\begin{aligned} I_2 &= \int_0^1 \frac{D(\Xi(t))}{||\psi(t)||^{2m}}.A_2(\Xi(t)).\mathbf{g}(\xi(t)).\Big(D(\Xi(t))\mathbf{1}_m\Big).\Phi(\Xi(t))\,dt \\ &= \int_0^1 \frac{D(\Xi(t))}{||\psi(t)||^{2m}}.A_2(\Xi(t)).\mathbf{g}(\xi(t)).\tilde{Z}(\Xi(t)).Z(\Xi(t)).\Phi(\Xi(t))\,dt \\ &= -\int_0^1 \mathbf{B}(\Xi(t)).\frac{d}{dt}\Big(\Phi(\Xi(t))\Big)\,dt\,, \end{aligned}$$

where $\mathbf{B}$ is an $n \times m$ matrix of smooth functions on S, given by

$$\mathbf{B}(x,p) = D(x,p).||p||^{-2m}.A_2(x,p).\mathbf{g}(x).\tilde{Z}(x,p)\,.$$

It is then clear that the components of $\mathbf{B}$ are in Σ_{-1}.

Integration by parts then yields $I_2 = I_3 + \bar{\beta} - \hat{\beta}$, where

$$I_3 = \int_0^1 \frac{d}{dt}\Big(\mathbf{B}(\Xi(t))\Big).\Phi(\Xi(t))\,dt\,, \quad \bar{\beta} = \mathbf{B}(\Xi(0)).\Phi(\Xi(0))\,, \quad \text{and} \quad \hat{\beta} = \mathbf{B}(\Xi(1)).\Phi(\Xi(1))\,.$$

The derivative $\frac{d}{dt}\Big(\mathbf{B}(\Xi(t))\Big)$ is equal to $\vec{H}_{\mathcal{M}}\mathbf{B}(\Xi(t))$, i. e. to $\sum_{i=1}^m \tilde{\varphi}^i(\Xi(t)).(G_i\mathbf{B})(\Xi(t))$, where G_i is the Hamilton vector field $\vec{\mu}_{g_i}$ arising from the momentum function μ_{g_i}. Since $\mu_{g_i} \in \Sigma_1$, the differential operators G_i map Σ_k into Σ_k for every k. So the components of the matrices $(G_i\mathbf{B})(x,p)$ are in Σ_1. Therefore there is a constant Θ_4, depending only on K, such that $||\vec{H}_{\mathcal{M}}\mathbf{B}(\Xi(t))|| \le \Theta_4.L(\xi).\kappa_-(\Xi)^{-1}$, and then

$$||I_3|| \le \Theta_4.\frac{L(\xi)^2}{\kappa_-(\Xi)}\,. \tag{7.12}$$

We have therefore shown that

$$\hat{x} - \bar{x} = I_1 + I_3 + \bar{\beta} - \hat{\beta}\,, \tag{7.13}$$

where $I_1 + I_3$ satisfies the bound

$$||I_1 + I_3|| \le \Theta_5.\frac{L(\xi)^2}{\kappa_-(\Xi)}\,, \qquad \Theta_5 = \Theta_4 + \Theta_2\Theta_3\Lambda^{2m-1}\,. \tag{7.14}$$

We now have to find estimates for the boundary terms $\bar{\beta}$, $\hat{\beta}$ of the integration by parts. Let $\mathbf{B}(x,p) = (b_1(x,p),\ldots,b_m(x,p))$. Then the vectors $b_i(x,p)$ are linear combinations $\sum_{j=1}^m \rho_i^j(x,p)g_j(x)$, with coefficient functions $\rho_i^j \in \Sigma_{-1}$. Then

$$\bar{\beta} = \sum_{i=1}^m \sum_{j=1}^m \rho_i^j(\Xi(0)).\tilde{\varphi}^i(\Xi(0)).g_j(\bar{x}) = \sum_{j=1}^m \bar{v}^j.g_j(\bar{x})\,,$$

where $\bar{v}^j = \sum_{i=1}^m \rho_i^j(\Xi(0))\tilde{\varphi}^i(\Xi(0))$, so that the vector $\bar{v} = (\bar{v}^1, \ldots, \bar{v}^m)$ satisfies

$$||\bar{v}|| \leq \frac{\Theta_6.L(\xi)}{\kappa_-(\Xi)}, \tag{7.15}$$

for a constant Θ_6 that does not depend on Ξ. This, of course, gives the bound

$$||\bar{\beta}|| \leq \frac{\Theta_2.\Theta_6.L(\xi)}{\kappa_-(\Xi)}, \tag{7.16}$$

which is not yet sufficient for our purposes, but will be used later.

We now let $t \to \gamma(t)$, $0 \leq t \leq 1$, be the solution of the initial value problem $\dot{x} = \sum_{i=1}^m \bar{v}^i g_i(x)$, $x(0) = \bar{x}$, and write $\bar{y} = \gamma(1)$, so $\bar{y}$ is the point of $\mathbb{R}^n$ that is reached from $\bar{x}$ in time 1 following a trajectory of the control system $\dot{x} = \sum_{i=1}^m u^i g_i(x)$ with the constant control $\bar{v}$. Then $||\dot{\gamma}(t)||_G \leq ||\bar{v}||$ for a. e. t, so $||\gamma||_G \leq ||\bar{v}||$, and then (7.15) yields $d_{\mathcal{M}}(\bar{y}, \bar{x}) \leq \frac{\Theta_6.L(\xi)}{\kappa_-(\Xi)}$. So, if we assume that

$$\kappa_-(\Xi) \geq 6.\Theta_6, \tag{7.17}$$

we find that

$$d_{\mathcal{M}}(\bar{y}, \bar{x}) \leq \frac{L(\xi)}{6}. \tag{7.18}$$

Next, we observe that the vector $\dot{\gamma}(t)$ satisfies $||\dot{\gamma}(t)|| \leq \Theta_2.\Theta_6.L(\xi).\kappa_-(\Xi)^{-1}$ as long as $\gamma(t) \in \tilde{K}$. Let us assume that

$$\kappa_-(\Xi) \geq \Theta_2.\Theta_6.\Lambda. \tag{7.19}$$

Then $||\dot{\gamma}(t)|| \leq 1$ as long as $||\gamma(t)|| \leq 2$. Since $\gamma(0) = \bar{x} \in K$, we have $||\gamma(0)|| \leq 1$, and then $\gamma(t) \in \tilde{K}$ for all $t \in [0,1]$.

We now compare the points $\bar{z} = \bar{x} + \bar{\beta}$ and $\bar{y}$. We let $\gamma^\#(t) = \bar{x} + t\bar{\beta}$. Also, we let $(x, x') \to G_j(x, x')$ be smooth matrix-valued functions with the property that $g_j(x) - g_j(x') = G_j(x, x').(x - x')$ for $x, x' \in \mathbb{R}^n$. (For example, if Dg_j is the Jacobian matrix of g_j, then we can take $G(x, x') = \int_0^1 Dg_j(x + s(x' - x))\,ds$.) Then, if we define $B_1(t) = \sum_{j=1}^m \bar{v}^j G_j(\gamma(t), \gamma^\#(t))$ and $B_2(t) = \sum_{j=1}^m \bar{v}^j G_j(\gamma^\#(t), \bar{x})$, we have

$$\begin{aligned}
\gamma(t) - \gamma^\#(t) &= \int_0^t \sum_{i=1}^m \bar{v}^j \Big(g_j(\gamma(t)) - g_j(\bar{x})\Big)\,dt \\
&= \int_0^t \sum_{i=1}^m \bar{v}^j \Big(g_j(\gamma(t)) - g_j(\gamma^\#(t))\Big)\,dt + \int_0^t \sum_{i=1}^m \bar{v}^j \Big(g_j(\gamma^\#(t))\,dt - g_j(\bar{x})\Big)\,dt \\
&= \int_0^t B_1(t).\Big(\gamma(t) - \gamma^\#(t)\Big)\,dt + \int_0^t B_2(t).\Big(\gamma^\#(t) - \bar{x}\Big)\,dt\,.
\end{aligned}$$

The vector $\gamma^{\#}(t)-\bar{x}$ is equal to $t\bar{\beta}$, which is bounded in norm by $\Theta_2.\Theta_6.L(\xi).\kappa_-(\Xi)^{-1}$ thanks to (7.16). So, if we let $\Theta_7=\sup\left\{\sum_{j=1}^m ||G_j(x,x')||: x,x'\in\tilde{K}\right\}$, then

$$\left\|\int_0^t B_2(t).\left(\gamma^{\#}(t)-\bar{x}\right)dt\right\|\le\Theta_2.\Theta_6.\Theta_7.L(\xi)^2.\kappa_-(\Xi)^{-1}.$$

Therefore Gronwall's inequality yields

$$||\gamma(t)-\gamma^{\#}(t)||\le\Theta_8\frac{L(\xi)^2}{\kappa_-(\Xi)}, \quad \text{where} \quad \Theta_8=e^{\Lambda.\Theta_7}\Theta_2.\Theta_6.\Theta_7.$$

In particular,

$$||\bar{y}-\bar{z}||\le\Theta_8\frac{L(\xi)^2}{\kappa_-(\Xi)}. \tag{7.20}$$

On the other hand, (7.2) implies that $d_{\mathcal{M}}(\bar{y},\bar{z})\le\Delta_2||\bar{y}-\bar{z}||^{1/2}$. So (7.20) tells us that

$$d_{\mathcal{M}}(\bar{y},\bar{z})\le\Delta_2\Theta_8^{1/2}\kappa_-(\Xi)^{-1/2}L(\xi). \tag{7.21}$$

If we now assume that

$$\kappa_-(\Xi)\ge 36\Delta_2^2\Theta_8, \tag{7.22}$$

then (7.21) becomes $d_{\mathcal{M}}(\bar{y},\bar{z})\le\frac{L(\xi)}{6}$. Combining this and (7.18), we get $d_{\mathcal{M}}(\bar{z},\bar{x})\le\frac{L(\xi)}{3}$. A similar computation yields $d_{\mathcal{M}}(\hat{z},\hat{x})\le\frac{L(\xi)}{3}$, where $\hat{z}=\hat{x}+\hat{\beta}$. Then

$$L(\xi)=d_{\mathcal{M}}(\bar{x},\hat{x})\le d_{\mathcal{M}}(\bar{x},\bar{z})+d_{\mathcal{M}}(\bar{z},\hat{z})+d_{\mathcal{M}}(\hat{z},\hat{x})\le\frac{2L(\xi)}{3}+d_{\mathcal{M}}(\bar{z},\hat{z}),$$

so

$$d_{\mathcal{M}}(\bar{z},\hat{z})\ge\frac{L(\xi)}{3}. \tag{7.23}$$

On the other hand, (7.13) says that

$$\hat{z}-\bar{z}=I_1+I_3. \tag{7.24}$$

Then (7.14) implies that

$$||\hat{z}-\bar{z}||\le\Theta_5.\frac{L(\xi)^2}{\kappa_-(\Xi)}. \tag{7.25}$$

We now use (7.2) to conclude that

$$d_{\mathcal{M}}(\hat{z},\bar{z})\le\Delta_2||\hat{z}-\bar{z}||^{1/2}\le\Delta_2\Theta_5^{1/2}.L(\xi).\kappa_-(\Xi)^{-1/2}. \tag{7.26}$$

If we combine (7.23) and (7.26), we get $\frac{L(\xi)}{3}\le\Delta_2\Theta_5^{1/2}.L(\xi).\kappa_-(\Xi)^{-1/2}$. Since $L(\xi)\neq 0$, this implies the inequality

$$\kappa_-(\Xi)\le 9\Delta_2^2\Theta_5. \tag{7.27}$$

The bound (7.27) has been established under the assumptions that $\Xi\in\mathcal{G}_{nc,min}(K)$ and (7.11), (7.17), (7.19), (7.22) hold. It then follows that the bound $\kappa_-(\Xi)\le C$ holds for all $\Xi\in\mathcal{G}_{nc,min}(K)$, if we let

$$C=\max\Big(1\ ,\ 6.\Theta_6\ ,\ \Theta_2.\Theta_6.\Lambda\ ,\ 36\Delta_2^2\Theta_8\ ,\ 9\Delta_2^2\Theta_5\Big). \tag{7.28}$$

Therefore $\kappa_+(\Xi)\le C.e^{\Lambda.\Theta_1}$. So, if we let $J=\{(x,p): x\in K\ ,\ ||p||\le C.e^{\Lambda.\Theta_1}\}$, we see that every $\Xi\in\mathcal{G}_{nc,min}(K)$ is entirely contained in J. $\square$

So far, we have studied the sets $\mathcal{G}_{nc,min}(K)$, whose members Ξ are such that $\pi_{T^*M} \circ \Xi$ is entirely contained in K. We would now like to discuss sets Σ of arcs $\Xi \in C^0([0,1], T^*M)$ defined by the conditions that (a) $\xi = \pi_{T^*M} \circ \Xi$ is a nonconstant energy minimizer, and (b) the endpoint condition $\partial\xi$ belongs to a given compact subset of $M \times M$. In the next section, these sets will play a crucial role in the proof of the subanalyticity of the distance function.

It is now no longer true in general that all our arcs will be contained in a compact subset of T^*M. On the other hand, Theorem 7.1 trivially implies that

7.2. Proposition. *If Σ is a subset of $\mathcal{G}_{nc,min}(M)$ with the property that all the projections $\xi = \pi_{T^*M} \circ \Xi$, $\Xi \in \Sigma$, are contained in a fixed compact subset of M, then Σ is relatively compact in $C^0([0,1], T^*M)$.* $\square$

So it suffices to investigate the relative compactness of the set of projections.

If $\bar{S}, \hat{S}$ are subsets of M, we use $\mathcal{EM}(\bar{S}, \hat{S})$ to denote the set of all energy minimizers $\xi : [0,1] \to M$ such that $\partial\xi \in \bar{S} \times \hat{S}$. We would like to know whether $\mathcal{EM}(\bar{K}, \hat{K})$ is a compact subset of $C^0([0,1], M)$ if $\bar{K}$ and $\hat{K}$ are compact.

The answer is clearly "no," even in the Riemannian case, unless some extra conditions are imposed, as can be seen by taking M to be the plane $\mathbb{R}^2$ minus one point P, and choosing $\bar{K} \subseteq M$ and $\hat{K} \subseteq M$ to be convex, but such that the convex hull of $\bar{K} \cup \hat{K}$ contains P. This suggests that the result may be true if one requires that whenever $\bar{x} \in \bar{K}$ and $\hat{x} \in \hat{K}$ it follows that there is an energy minimizer $\xi : [0,1] \to M$ such that $\partial\xi = (\bar{x}, \hat{x})$. Remark 7.4 below shows that this condition is not yet sufficient, even for Riemannian metrics.

It turns out that, for SBG metrics, a slightly stronger hypothesis suffices, as we now show. Moreover, Theorem 7.1 turns out to play a crucial role in the proof.

7.3. Theorem. *Assume that $\mathcal{M}$ is SBG. Let $\bar{K}, \hat{K}$ be compact subsets of M such that*

(#) there exists a neighborhood U of $\bar{K}$ in M with the property that, whenever $\bar{x} \in U$ and $\hat{x} \in \hat{K}$, it follows that there exists a length minimizer ξ such that $\partial\xi = (\bar{x}, \hat{x})$.

Then

(A) There exists a compact subset $\mathcal{K}$ of $C^0([0,1], T^*M) \cap \widehat{GEO}(\mathcal{M})$ such that

$$\mathcal{EM}(\bar{K}, \hat{K}) = \left\{ \pi_{T^*M} \circ \Xi : \Xi \in \mathcal{K} \right\}. \tag{7.29}$$

(B) $\mathcal{EM}(\bar{K}, \hat{K})$ is a compact subset of $C^0([0,1], M)$.

(C) The set $\left\{ \Xi \in \mathcal{G}_{nc,min}(M) : \pi_{T^*M} \circ \Xi \in \mathcal{EM}(\bar{K}, \hat{K}) \right\}$ is relatively compact in the space $C^0([0,1], T^*M)$.

PROOF. It is clear that (B) is a trivial corollary of (A), since the map

$$C^0([0,1], T^*M) \ni \Xi \to \pi_{T^*M} \circ \Xi \in C^0([0,1], M)$$

is continuous. Also, Proposition 7.2 implies that (C) follows from (B). So it suffices to prove (A).

Clearly, (A) will follow if we prove:

(A') *There exists a compact subset J of T^*M with the property that, if $\xi:[0,1]\to M$ is any nonconstant energy minimizer such that $\xi(0)\in\bar{K}$ and $\xi(1)\in\hat{K}$, and $\Xi\in\widehat{GEO}(\mathcal{M})$ is such that $\pi_{T^*M}\circ\Xi=\xi$, then Ξ is entirely contained in J.*

Indeed, suppose (A') holds. Let $\mathcal{K}_0$ be the set of all $\Xi\in C^0([0,1],T^*M)\cap\widehat{GEO}(\mathcal{M})$ such that Ξ is contained in J and $\pi_{T^*M}\circ\Xi\in\mathcal{EM}(\bar{K},\hat{K})$. Then $\mathcal{K}_0$ is compact. (*Proof*: let $\{\Xi_j\}_{j=1}^{\infty}$ be a sequence in $\mathcal{K}_0$. We want a subsequence that converges to a member of $\mathcal{K}_0$. Since $C^0([0,1],J)\cap\widehat{GEO}(\mathcal{M})$ is compact, it suffices to assume that the Ξ_j converge in $C^0([0,1],T^*M)$ to an arc $\Xi\in C^0([0,1],J)\cap\widehat{GEO}(\mathcal{M})$. If $\xi_j=\pi_{T^*M}\circ\Xi_j$, and $\xi=\pi_{T^*M}\circ\Xi$, then $\partial\xi\in\bar{K}\times\hat{K}$, because $\partial\xi_j\in\bar{K}\times\hat{K}$. Moreover, ξ is a canonically parametrized minimizer, because of Proposition 3.24. So $\Xi\in\mathcal{K}_0$.) Let $\mathcal{K}_1$ be the set of all constant curves $\Xi_x^{cons}:[0,1]\to T^*M$, for all $x\in\bar{K}\cap\hat{K}$, where $\Xi_x^{cons}(t)\stackrel{\text{def}}{=}(x,0)$ for $t\in[0,1]$. Then $\mathcal{K}_1$ is obviously compact. Let $\mathcal{K}=\mathcal{K}_0\cup\mathcal{K}_1$. Then $\mathcal{K}$ is compact, and (7.29) holds.

So it suffices to prove (A'). We take J to be the closure of the set J_0 of all points that are of the form $\Xi(t)$ for some $\Xi\in C^0([0,1],T^*M)\cap\widehat{GEO}(\mathcal{M})$ such that $\xi=\pi_{T^*M}\circ\Xi$ is a nonconstant length minimizer and $\partial\xi\in\bar{K}\times\hat{K}$. We have to prove that J is compact. It suffices to show that every sequence $\{(x_j,p_j)\}$ of points of J_0 has a convergent subsequence. So we pick one such sequence, and find curves $\Xi_j\in C^0([0,1],T^*M)\cap\widehat{GEO}(\mathcal{M})$ such that $\xi_j=\pi_{T^*M}\circ\Xi_j$ is a nonconstant length minimizer and $\partial\xi_j\in\bar{K}\times\hat{K}$, and times $\tau_j\in[0,1]$ such that $(x_j,p_j)=\Xi_j(\tau_j)$. Write $\partial\xi_j=(\bar{x}_j,\hat{x}_j)$. Then all the $\bar{x}_j$ are in $\bar{K}$, and all the $\hat{x}_j$ are in $\hat{K}$. We let L_j denote the length of ξ_j. Then $L_j=d_{\mathcal{M}}(\bar{x}_j,\hat{x}_j)$, because ξ_j is a length minimizer. So $L_j\le\Lambda$, where $\Lambda=\sup\{d_{\mathcal{M}}(x,x'):x\in\bar{K},x'\in\hat{K}\}$. (It is clear that $\Lambda<\infty$, because $d_{\mathcal{M}}$ is bounded on the compact set $\bar{K}\times\hat{K}$.) Therefore, by passing to a subsequence, we may assume that the limits $\bar{x}=\lim_{j\to\infty}\bar{x}_j$, $\hat{x}=\lim_{j\to\infty}\hat{x}_j$, $L=\lim_{j\to\infty}L_j$, and $\tau=\lim_{j\to\infty}\tau_j$ exist. Clearly, then, $\bar{x}\in\bar{K}$, $\hat{x}\in\hat{K}$, $0\le\tau\le1$, and $0\le L\le\Lambda$.

Pick $\delta>0$ such that the closed ball $B=\{x\in M:d_{\mathcal{M}}(x,\bar{x})\le2\delta\}$ is compact and contained in U. Then pick α such that $0<\alpha<1$ and $\alpha.\Lambda\le\delta$. By passing to a subsequence, we may assume that $d_{\mathcal{M}}(\bar{x}_j,\bar{x})\le\delta$ for all j. Then, since ξ_j is parametrized by constant times arc-length, and $||\xi_j||_G=L_j$, we have $||\dot{\xi}_j(t)||_G=L_j$ for a. e. t, so $d_{\mathcal{M}}(\xi_j(t),\bar{x}_j)\le tL_j\le t\Lambda$ for all j and all t. Therefore $d_{\mathcal{M}}(\xi_j(t),\bar{x}_j)\le\delta$ for all j, if $t\in[0,\alpha]$. It follows that $\xi_j(t)\in B$ for all j and all $t\in[0,\alpha]$. Moreover,

$$d_{\mathcal{M}}(\xi_j(t),\xi_j(t'))\le L_j|t-t'|\le\Lambda|t-t'|\quad\text{whenever}\quad t,t'\in[0,\alpha]\,. \tag{7.30}$$

So, if we let ζ_j denote the restriction of ξ_j to the interval $[0,\alpha]$, we see that $\{\zeta_j\}$ is an equicontinuous sequence of curves in the compact metric space $(B,d_{\mathcal{M}})$. Therefore, by passing to a subsequence we may assume that the ζ_j converge uniformly on $[0,\alpha]$ to a curve $\zeta:[0,\alpha]\to B$. It then follows from (7.30) that

$$d_{\mathcal{M}}(\zeta(t),\zeta(t'))\le L|t-t'|\quad\text{whenever}\quad t,t'\in[0,\alpha]\,. \tag{7.31}$$

Then Proposition 3.20 implies that ζ is an E-admissible E-arc, and $||\dot{\zeta}(t)||_G\le L$ for almost all $t\in[0,\alpha]$. Therefore $||\zeta||_G\le L\alpha$.

On the other hand, since each ζ_j is a length minimizer, the distance $d_{\mathcal{M}}(\bar{x}_j, \zeta_j(\alpha))$ is exactly equal to $L_j\alpha$. So $d_{\mathcal{M}}(\bar{x}, \zeta(\alpha)) = L\alpha$, because $d_{\mathcal{M}}$ is continuous. It follows that $||\zeta||_G = L\alpha$, and ζ is a length minimizer. Since $||\dot{\zeta}(t)||_G \leq L$ for almost all $t \in [0, \alpha]$, we must in fact have $||\dot{\zeta}(t)||_G = L$ for almost all $t \in [0, \alpha]$. So ζ is a length minimizer parametrized by constant times arc-length.

We now use the fact that $\zeta(\alpha) \in U$, and let θ be a length minimizer joining $\zeta(\alpha)$ to $\hat{x}$. We parametrize θ so that its interval of definition is $[\alpha, 1]$, and θ is parametrized by constant times arc-length. Since each ξ_j is a minimizer, and $||\dot{\xi}_j(t)||_G = L_j$ a. e., we have $d_{\mathcal{M}}(\xi_j(\alpha), \hat{x}_j) = (1-\alpha)L_j$. So $d_{\mathcal{M}}(\zeta(\alpha), \hat{x}) = (1-\alpha)L$, and then $||\dot{\theta}(t)||_G = L$ for a. e. t.

Now define $\xi_* : [0,1] \to M$ by letting $\xi_*(t) = \zeta(t)$ if $t \in [0, \alpha]$ and $\xi_*(t) = \theta(t)$ if $t \in [\alpha, 1]$. Then ξ_* is an E-admissible E-arc such that $||\dot{\xi}_*(t)||_G = L$ for a. e. $t \in [0,1]$. So $||\xi_*||_G = L$. On the other hand,

$$d_{\mathcal{M}}(\xi_*(0), \xi_*(1)) = d_{\mathcal{M}}(\bar{x}, \hat{x}) = \lim_{j\to\infty} d_{\mathcal{M}}(\bar{x}_j, \hat{x}_j) = \lim_{j\to\infty} L_j = L\,.$$

So ξ_* is a length minimizer parametrized by constant timees arc-length. Therefore ξ_* is a geodesic arc.

Now write $\Xi_j = (\xi_j, \psi_j)$, and let Z_j be the restriction of Ξ_j to the interval $[0, \alpha]$. Then the Z_j are in $C^0([0,\alpha], \pi_{T^*M}^{-1}(B)) \cap \widehat{GEO}(\mathcal{M})$, and the projections $\zeta_j = \pi_{T^*M} \circ Z_j$ are nonconstant length minimizers. By Theorem 7.1, all the Z_j are contained in a fixed compact subset J_B of T^*M. It then follows that we may assume, after passing once more to a subsequence, that the limit $\bar{p} = \lim_{j\to\infty} \psi_j(0)$ exists.

Let $\Xi \in \mathrm{Max}(\widehat{GEO}(\mathcal{M}))$ be such that $\Xi(0) = (\bar{x}, \bar{p})$. Then the domain of Ξ is an interval I. We will show that $[0,1] \subseteq I$.

Write $\Xi(t) = (\xi(t), \psi(t))$ for $t \in I$. Pick a β such that $0 < \beta \leq \alpha$ and $[0, \beta] \subseteq I$. Then $\Xi_j(t) \to \Xi(t)$ for all $t \in [0, \beta]$, by the usual theorem on continuous dependence on the initial conditions of the solutions of an ordinary differential equation. It follows that $\xi_j(t) \to \xi(t)$ for $t \in [0, \beta]$. Since $\xi_j(t) \to \xi_*(t)$ for $t \in [0, \alpha]$, we conclude that ξ and ξ_* coincide on $[0, \beta]$.

Since ξ_* is a length minimizer parametrized by arc-length, it is the projection of an arc $\Xi_* \in C^0([0,1], T^*M) \cap \widehat{GEO}(\mathcal{M})$. If $L \neq 0$, then Theorem 6.7 implies that Ξ and Ξ_* coincide on $[0, \beta]$, since ξ and ξ_* do. So Ξ_* is a solution of the geodesic equation which is defined on $[0,1]$ and satisfies $\Xi_*(0) = \Xi(0)$. Since Ξ is maximally defined, it follows that $[0,1] \subseteq I$. We now consider the remaining possibility, namely, that $L = 0$. In this case $L_j \to 0$, and $L_j^2 = \sum_{i=1}^m \langle \psi_j(0), g_i(\xi_j(0)) \rangle^2$. It follows that $\langle \bar{p}, g_i(\bar{x}) \rangle = 0$ for $i = 1 \ldots, m$. It is then easy to see that $(\bar{x}, \bar{p})$ is an equilibrium point for the geodesic equation, so $I = \mathbb{R}$.

We have therefore proved that the domain of Ξ contains $[0,1]$. It then follows from the theorem on continuous dependence on the initial conditions of the solutions of an ordinary differential equation that $\Xi_j(t) \to \Xi(t)$ uniformly for $t \in [0,1]$. Therefore $(x_j, p_j) = \Xi_j(\tau_j)$ converges as $j \to \infty$ to the limit $\Xi(\tau)$. This completes the proof of (A'). $\square$

7.4. Remark. Theorem 7.3 is *not* valid in general —even in the Riemannian case— if (#) is replaced by the weaker requirement that

($\#_w$) *whenever $\bar{x} \in \bar{K}$ and $\hat{x} \in \hat{K}$, it follows that there exists a length minimizer ξ such that $\partial\xi = (\bar{x}, \hat{x})$.*

Indeed, we may take M to be the two-dimensional unit sphere from which one point Q belonging to the Equator has been removed. Let N, S be the North and South poles, respectively. Choose $\bar{K} = \{S\}$, $\hat{K} = \{N\}$. Then ($\#_w$) holds, but (#) does not. Clearly, the conclusion of Theorem 7.3 is not true. □

7.5. Remark. In the Riemannian case, we can choose the set $\mathcal{K}$ of Theorem 7.3 to be the set of *all* $\Xi \in C^0([0,1], T^*M) \cap \widehat{GEO}(\mathcal{M})$ whose projection $\pi_{T^*M} \circ \Xi$ is in $\mathcal{EM}(\bar{K}, \hat{K})$. *This does not work in the general subriemannian SBG case.* Indeed, if $\mathcal{M}$ is not Riemannian, then we can pick a point $x \in M$ such that σ_x is not surjective, and choose $\bar{K} = \hat{K} = \{x\}$. Then, if $p \in E(x)^\perp$ is arbitrary, we can define $\Xi_{x,p}(t) = (x, p)$ for $t \in [0,1]$, and the arc $\Xi_{x,p}$ is in $\mathcal{K}$. Since $E(x)^\perp$ is not compact, we see that $\mathcal{K}$ is not compact either. □

§8. Subanalyticity of the distance function.

We are now ready to prove the result on subanalyticity of the distance function $d_{\mathcal{M}}$. Recall that, if N_1, N_2 are real-analytic manifolds, and $\nu : N_1 \to N_2$ is a mapping, then ν is said to be *subanalytic* on a subset S of N_1 if the set $\{(x, \nu(x)) : x \in S\}$ (i. e. the graph of the restriction of ν to S) is a subanalytic subset of $N_1 \times N_2$. We say that the subriemannian manifold $\mathcal{M} = (M, F, \sigma, G)$ is *real-analytic* if M is a real-analytic manifold, F is a real-analytic vector bundle over M, and both the map σ and the metric G are real-analytic.

8.1. Theorem. *Let $\mathcal{M} = (M, F, \sigma, G)$ be a real-analytic strongly bracket-generating subriemannian manifold. Let $\bar{K}$, $\hat{K}$ be compact subanalytic subsets of M such that Condition (#) of the statement of Theorem 7.3 holds. Then the distance function $d_{\mathcal{M}}$ is subanalytic on $\bar{K} \times \hat{K}$.*

PROOF. Using Theorem 7.3, we let $\mathcal{K}$ be a compact subset of $C^0([0,1], T^*M) \cap \widehat{GEO}(\mathcal{M})$ such that every $\xi \in \mathcal{EM}(\bar{K}, \hat{K})$ is the projection $\pi_{T^*M} \circ \Xi$ of some $\Xi \in \mathcal{K}$. We let $\mathcal{F}$ be the "geodesic flow" on T^*M, that is, the flow of the vector field $\vec{H}_{\mathcal{M}}$. By definition,

(1) $\mathcal{F}$ is a T^*M-valued partially defined map on $T^*M \times \mathbb{R}$;
(2) a point $(x, p, t) \in T^*M \times \mathbb{R}$ is in the domain $\mathrm{Dom}(\mathcal{F})$ of $\mathcal{F}$ iff the maximally defined solution $\Xi_{x,p}$ of the initial value problem $\dot{\Xi}(s) = \vec{H}_{\mathcal{M}}(\Xi(s))$, $\Xi(0) = (x, p)$, is defined for $s = t$;

and

(3) $\mathcal{F}(x, p, t) = \Xi_{x,p}(t)$ if $(x, p, t) \in \mathrm{Dom}(\mathcal{F})$.

We then use $\mathcal{F}_1$ to denote the "exponential map," i. e. the T^*M-valued partially defined map on T^*M, whith domain $\mathrm{Dom}(\mathcal{F}_1) \stackrel{\mathrm{def}}{=} \Big\{(x,p) \in T^*M : (x,p,1) \in \mathrm{Dom}(\mathcal{F})\Big\}$, defined by letting $\mathcal{F}_1(x,p) \stackrel{\mathrm{def}}{=} \mathcal{F}(x,p,1)$ for $(x,p) \in \mathrm{Dom}(\mathcal{F}_1)$. Then $\mathrm{Dom}(\mathcal{F}_1)$ is open in T^*M.

Let $J = \{\Xi(0) : \Xi \in \mathcal{K}\}$. Then J is a compact subset of $\mathrm{Dom}(\mathcal{F}_1)$. Let $\tilde{J}$ be a compact subanalytic subset of $\mathrm{Dom}(\mathcal{F}_1)$ such that $J \subseteq \tilde{J}$. Let $A : \mathrm{Dom}(\mathcal{F}_1) \to M \times M \times \mathbb{R}$ be the map given by $A(x,p) = \Big(x\,,\, \mathcal{F}_1(x,p)\,,\, 2H_{\mathcal{M}}(x,p)\Big)$. Then A is a real-analytic map. Since $\tilde{J}$ is a compact subanalytic subset of $\mathrm{Dom}(\mathcal{F}_1)$, the image $Q = A(\tilde{J})$ is a compact subanalytic subset of $M \times M \times \mathbb{R}$.

Let

$$P = \Big\{(\bar{x},\hat{x},r) \in Q \cap (\bar{K} \times \hat{K} \times \mathbb{R}) : \sim (\exists s \in \mathbb{R})(0 \le s < r \wedge (\bar{x},\hat{x},s) \in Q\Big\}\,.$$

Then P is a subanalytic subset of $M \times M \times \mathbb{R}$. (*Proof*: This follows from general properties of subanalytic sets, as explained, for example, in Sussmann [32]: any formula of a first-order predicate calculus constructed out of atomic formulae that define subanalytic sets, combined by means of the usual logical connectives $\wedge$, $\vee$, $\sim$, $\exists$ and $\forall$, automatically defines a subanalytic set, provided only that all the quantifiers are proper. In our case,

$$(\bar{x},\hat{x},r) \in P \Longleftrightarrow \Bigg(\bar{x} \in \bar{K} \wedge \hat{x} \in \hat{K} \wedge (\bar{x},\hat{x},r) \in Q \wedge \sim (\exists s \in \mathbb{R})(0 \le s < r \wedge (\bar{x},\hat{x},s) \in Q)\Bigg)\,,$$

and the existential quantifier is proper because of the bound $0 \le s < r$.)

It is then easy to see that P is precisely the graph of $d_{\mathcal{M}} \lceil (\bar{K} \times \hat{K})$. So this graph is subanalytic, and our proof is complete. ☐

Theorem 8.1 implies a number of subanalyticity results about the distance function:

8.2. Theorem. *Assume that $\mathcal{M}$ is SBG and real-analytic. Then*

(1) If U is any open subset of $M \times M$ having the property that for every pair $(\bar{x},\hat{x}) \in U$ there exists a length minimizer ξ such that $\partial\xi = (\bar{x},\hat{x})$, then $d_{\mathcal{M}}$ is subanalytic on U (that is, the graph of the restriction of $d_{\mathcal{M}}$ to U is a subanalytic subset of $U \times \mathbb{R}$);
(2) $d_{\mathcal{M}}$ is subanalytic on a neighborhood of the diagonal of $M \times M$;
(3) if V is an open subset of M which is relatively convex in $\mathcal{M}$ (cf. Definition 4.9), then $d_{\mathcal{M}}$ is subanalytic on $V \times V$;
(4) if any two points of M can be joined by a length minimizer, then $d_{\mathcal{M}}$ is subanalytic on $M \times M$;
(5) if M is complete, then $d_{\mathcal{M}}$ is subanalytic on $M \times M$;
(6) if $\bar{x} \in M$ and $R > 0$ are such that the closed ball $B_R(\bar{x}) = \{x \in M : d_{\mathcal{M}}(x,\bar{x})\}$ is compact, then the function $x \to d_{\mathcal{M}}(x,\bar{x})$ is subanalytic on $B_R(\bar{x})$.

PROOF. Statement (1) follows from Theorem 8.1 because subanalyticity is a local property: to prove that the graph $A = \{(\bar{x}, \hat{x}, r) \in U \times \mathbb{R} : r = d_{\mathcal{M}}(\bar{x}, \hat{x})\}$ is subanalytic in $U \times \mathbb{R}$, it suffices to find, for each $(\bar{x}, \hat{x}, r) \in U \times \mathbb{R}$, an open neighborhood W of $(\bar{x}, \hat{x}, r)$ in $U \times \mathbb{R}$ such that $A \cap W$ is subanalytic in W. If $r \neq d_{\mathcal{M}}(\bar{x}, \hat{x})$, then $A \cap W = \emptyset$ if W is a sufficiently small neighborhood of $(\bar{x}, \hat{x}, r)$, since $d_{\mathcal{M}}$ is continuous. If $r = d_{\mathcal{M}}(\bar{x}, \hat{x})$, then we can pick compact subanalytic subsets $\bar{K}$, $\hat{K}$, of M such that $\bar{K} \times \hat{K} \subseteq U$ and $\bar{K} \times \hat{K}$ is a neighborhood of $(\bar{x}, \hat{x})$. It is then clear that $\bar{K}$ has an open neighborhood N such that every pair $(x, x') \in N \times \hat{K}$ is the endpoint condition $\partial\xi$ of some length minimizer ξ. So $\bar{K}$ and $\hat{K}$ satisfy the hypotheses of Theorem 8.1, and then $A \cap (\bar{K} \times \hat{K} \times \mathbb{R})$ is subanalytic in $M \times M \times \mathbb{R}$. If $\bar{W}$, $\hat{W}$ denote the interiors of $\bar{K}$, $\hat{K}$, and $W = \bar{W} \times \hat{W} \times \mathbb{R}$, then W is a neighborhood of $(\bar{x}, \hat{x}, r)$ in $U \times \mathbb{R}$, and $A \cap W$ is subanalytic in W. So a W with the desired property exists, and then A is subanalytic in $U \times \mathbb{R}$, completing the proof of (1).

Statement (2) is an immediate consequence of (1), because Proposition 4.8 implies that there exists a neighborhood U of the diagonal having the property of (1). Statement (3) also follows from (1), because if V is relatively convex then the set $U = V \times V$ satisfies the condition of (1). Statement (4) is a special case of (3). Statement (5) follows from (4) and the subriemannian Hopf-Rinow theorem (Theorem 4.7).

Finally, to prove (6), we first observe that there exists $R' > R$ such that $B_{R'}(\bar{x})$ is compact. (Find finitely many compact balls $B_{r_\alpha}(x_\alpha)$, $\alpha \in A$, such that the balls $B_{\frac{r_\alpha}{2}}(x_\alpha)$ cover $B_R(\bar{x})$. Let $\rho = \min\left(\frac{R}{2}, \min\left\{\frac{r_\alpha}{4} : \alpha \in A\right\}\right)$, $R' = R + \rho$. If $x \in B_{R'}(\bar{x})$, then there is an E-admissible E-curve $\xi : [0, L] \to M$, parametrized by arc-length, such that $\xi(0) = \bar{x}$, $\xi(L) = x$, and $||\xi||_G = L \leq R' + \rho$. If $L \geq 2\rho$, then $x' = \xi(L - 2\rho) \in B_R(\bar{x})$, since $L - 2\rho \leq R$. So $x' \in B_{\frac{r_\alpha}{2}}(x_\alpha)$ for some α, and then $x \in B_{r_\alpha}(x_\alpha)$, since $4\rho \leq r_\alpha$. If $L < 2\rho$, then $x \in B_R(\bar{x})$. So $B_{R'}(\bar{x}) \subseteq \bigcup_{\alpha \in A} B_{r_\alpha}(x_\alpha)$. Therefore $B_{R'}(\bar{x})$ is compact.)

Now let $\beta = \frac{R'-R}{2}$. If $d_{\mathcal{M}}(x, \bar{x}) < \beta$, then the closed ball $B_R(\bar{x})$ is contained in $B_{R+\beta}(x)$, which is a closed subset of $B_{R'}(\bar{x})$. So $B_{R+\beta}(x)$ is compact, and then Proposition 4.8 tells us that every point of $B_{R+\beta}(x)$ can be joined to x by a minimizer. Therefore every point of $B_R(\bar{x})$ can be joined to every point of $B_\beta(\bar{x})$ by a minimizer. Therefore the hypothesis of Theorem 8.1 is satisfied with $\bar{K} = \{\bar{x}\}$ and $\hat{K} = B_R(\bar{x})$. □

Appendix.

Proof of Proposition 2.2. First, we show that, if we define $\sigma_x(v)$ by $\sigma_x(v) = \sigma(V)(x)$, then $\sigma_x(v)$ does not depend on the choice of V. If $V_1(x) = V_2(x) = v$, then we can pick a local basis $\mathbf{f} = (f_1, \ldots, f_m)$ of sections of F near x, and write $W = V_1 - V_2 = \sum_{i=1}^m w^i f_i$ on $U_{\mathbf{f}}$, where the w^i belong to $C^\infty(U_{\mathbf{f}})$ and vanish at x. If $\rho \in C^\infty(M)$ is such that $\rho(x) = 1$ and support(ρ) is a compact subset of $U_{\mathbf{f}}$, then

$$\sigma(W)(x) = \rho(x)^2\sigma(W)(x) = \Big(\rho^2\sigma(W)\Big)(x) = \sigma(\rho^2 W)(x) = \left(\sigma\Big(\sum_i(\rho w^i)(\rho f_i)\Big)\right)(x)$$

$$= \Big(\sum_i(\rho w^i)(\sigma(\rho f_i))\Big)(x) = \sum_i(\rho w^i)(x)\sigma(\rho f_i)(x) = 0\,.$$

So $\sigma(V_1)(x) = \sigma(V_2)(x)$, as desired. The last two statements are trivial. □

Proof of Proposition 2.7. We let $n = \dim(M)$, $m = \dim(F)$, $I = I(\xi)$. It clearly suffices to work locally, i.e. to assume that $\xi(I)$ is entirely contained in an open subset U of M which is the domain of a smooth coordinate chart $\mathbf{x}$ and of a local basis $\mathbf{f} = (f_1, \ldots, f_m)$ of smooth sections of F. Let $H \subseteq I$ be such that $\dot{\xi}(t)$ exists and belongs to $E(\xi(t))$ for all $t \in H$, and $I \backslash H$ has measure zero. Let $g_i(x) = \sigma_x(f_i(x))$. For $t \in I$, define $\mathcal{G}(t)$ to be the linear span of the vectors $g_1(\xi(t)), \ldots, g_m(\xi(t))$. For $k \in \{1, \ldots, m\}$, and a subset J of $\{1, \ldots, m\}$ of cardinality k, let $I_{k,J}$ be the set of those $t \in I$ such that $\mathcal{G}(t)$ has dimension k and is spanned by the vectors $g_j(\xi(t))$, $j \in J$. Then $I_{k,J}$ is defined by a finite collection of equalities and inequalities involving continuous functions of t, namely, the $k \times k$ and $(k+1) \times (k+1)$ minors of the $m \times n$ matrix $\mathbf{g}(\xi(t))$ whose columns are the component vectors of $g_1(\xi(t)), \ldots, g_m(\xi(t))$ relative to the chart $\mathbf{x}$. So every $I_{k,J}$ is measurable. For $t \in I_{k,J} \cap H$, the vector $\dot{\xi}(t)$ can be expressed uniquely as a linear combination $\dot{\xi}(t) = \sum_{j \in J} \eta^j_{k,J}(t) g_j(\xi(t))$. It then follows easily that the functions $\eta^j_{k,J}$ are measurable, since they can be obtained from the components of the vectors $g_j(\xi(t))$ and $\dot{\xi}(t)$ by means of Cramer's rule. So, if we take $\eta_{k,J}(t) = \sum_{j \in J} \eta^j_{k,J}(t) f_j(\xi(t))$ for $t \in I_{k,J} \cap H$, we see that $\eta_{k,J}$ is measurable on $I_{k,J} \cap H$. Since $H = \cup_{k,J}(I_{k,J} \cap H)$, and $I \backslash H$ has measure zero, the existence of a measurable ζ follows immediately. □

A.1. Example. *(A smooth E-arc that is not E-admissible.)* Let $M = \mathbb{R}^2$, and let E be the singular distribution such that, if $(x, y) \in M$, then $E(x, y)$ is the linear span of $X(x, y)$ and $Y(x, y)$, where $X = y\frac{\partial}{\partial x}$ and $Y = \frac{\partial}{\partial y}$. (To realize E as a singular distribution, take F to be the trivial bundle on M with fiber $\mathbb{R}^2$, and let $\sigma_{(x,y)}(\alpha, \beta) = \alpha X(x, y) + \beta Y(x, y)$.) Let $\xi(t) = (t^2, t^2)$ for $t \in [-1, 1]$. Then ξ is a smooth E-arc, and in fact ξ is such that $\dot{\xi}(t) \in E(\xi(t))$ for *all* (and not just *almost* all) $t \in [-1, 1]$. Let $\mathbf{f} = (f_1, f_2)$ be the canonical basis of sections of F. Then $\sigma(f_1) = X$, $\sigma(f_2) = Y$, and the only expression of $\dot{\xi}(t)$ as a linear combination of $X(\xi(t))$ and $Y(\xi(t))$, for $t \neq 0$, is

$$\dot{\xi}(t) = \frac{2}{t}.X(\xi(t)) + 2t.Y(\xi(t)) \quad \text{for} \quad t \in \mathbb{R}.$$

This shows that ξ is not E-admissible, since the function $t \to \frac{2}{t}$ is not integrable. □

Proof of Proposition 3.5. We work locally, as in the proof of Proposition 2.7. So we assume that ξ is contained in the domain U of a chart $\mathbf{x}$ and that a local orthonormal basis $\mathbf{f} = (f_1, \ldots, f_m)$ of sections of F exists on U. We let Q be the set of those $t \in I(\xi)$ such that $\dot{\xi}(t)$ exists, so Q is measurable and $I \backslash Q$ is a null set. We define $\mathcal{G}(t)$, the g_i and the sets $I_{k,J}$ exactly as in the proof of Proposition 2.7, and we let $I_{k,J,0}$ be the set of those $t \in I_{k,J}$ such that $\dot{\xi}(t)$ exists and belongs to $\mathcal{G}(t)$. It is then clear that $I_0(\xi)$ is the union of the $I_{k,J,0}$ for all k, J.

Fix k and J, and choose a $t \in I_{k,J,0}$. Then we can find smooth vector fields $Y_1, \ldots, Y_{n-k}$ on a neighborhood V of $\xi(t)$ such that $Y_1(x), \ldots, Y_{n-k}(x)$, together with the $g_j(x)$ for $j \in J$, form a basis of $T_x M$ for all $x \in V$. We can then find $\varepsilon > 0$ such that $\xi(]\,t - \varepsilon, t + \varepsilon\,[) \subseteq V$. Then, if $t - \varepsilon < s < t + \varepsilon$, we can write $\dot{\xi}(s)$ in a unique way as a linear combination $\dot{\xi}(s) = \sum_{i=1}^{n-k} \alpha^i(s) Y_i(\xi(s)) + \sum_{j \in J} \beta^j(s) g_j(\xi(s))$, and the coefficient functions α^i, β^j are measurable, because they are given, in terms of the components of

$\dot{\xi}(s)$, the $Y_i(\xi(s))$, and the $g_j(\xi(s))$, by Cramer's rule. Clearly, $s \in I_{k,J,0}$ if and only if $s \in I_{k,J}$ and $\alpha_1(s) = \ldots = \alpha_{n-k}(s) = 0$. Since $I_{k,J}$ is measurable, we conclude that $]t-\varepsilon, t+\varepsilon[\cap I_{k,J,0}$ is measurable.

So we have shown that for every $t \in I_{k,J,0}$ there is an $\varepsilon_t > 0$ with the property that $]\,t-\varepsilon_t, t+\varepsilon_t\,[\cap I_{k,J,0}$ is measurable. Then the set

$$Z = \bigcup_{t \in I_{k,J,0}} \Big(\,]\,t-\varepsilon_t, t+\varepsilon_t\,[\cap I_{k,J,0}\Big)$$

is open in $\mathbb{R}$, and $I_{k,J,0}$ is a locally measurable subset of Z. So $I_{k,J,0}$ is a measurable subset of $\mathbb{R}$.

It then follows that $I_0(\xi)$ is measurable. Next, we must show that ω_ξ is measurable on $I_0(\xi)$. It clearly suffices to fix k and J, and show that ω_ξ is measurable on $I_{k,J,0}$. Since k and J are now fixed, we may relabel the f_i's so as to assume that $J = \{m-k+1, \ldots, m\}$. Then for every $t \in I_{k,J,0}$ and $i \in \{1, \ldots, m-k\}$ we can write $g_i(\xi(t)) = \sum_{j=m-k+1}^{m} v_i^j(t) g_j(\xi(t))$. This expression is unique (because the vectors $g_j(\xi(t))$, for $m-k+1 \leq j \leq m$, are linearly independent), and the functions v_i^j are continuous on $I_{k,J,0}$.

If we define $\tilde{h}_i(t) = f_i(\xi(t)) - \sum_{j=m-k+1}^{m} v_i^j(t) f_j(\xi(t))$ for $i = 1, \ldots, m-k$, and $\tilde{h}_i(t) = f_i(\xi(t))$ for $i = m-k+1, \ldots, m$, then the vectors $\tilde{h}_i(t)$ belong to $F(\xi(t))$ for every $i \in \{1, \ldots, m\}$ and $t \in I_{k,J,0}$, and $(\tilde{h}_1(t), \ldots, \tilde{h}_m(t))$ is a basis of $F(\xi(t))$ such that $(\tilde{h}_1(t), \ldots, \tilde{h}_{m-k}(t))$ is a basis of the kernel $K(t)$ of $\sigma_{\xi(t)}$. Moreover, the $\tilde{h}_i$ depend continuously on t for $t \in I_{k,J,0}$. If we apply the Gramm-Schmidt orthogonalization procedure (with respect to the inner product $G(\xi(t))$ on $F(\xi(t))$) we find an orthonormal basis $(h_1(t), \ldots, h_m(t))$ of $F(\xi(t))$ such that the maps $t \to h_i(t)$ are continuous on $I_{k,J,0}$, and $\mathrm{span}(h_1(t), \ldots, h_i(t)) = \mathrm{span}(\tilde{h}_1(t), \ldots, \tilde{h}_i(t))$ for every $t \in I_{k,J,0}$ and every $i \in \{1, \ldots, m\}$. This implies, in particular, that $(h_1(t), \ldots, h_{m-k}(t))$ is an orthonormal basis of $K(t)$ and $(h_{m-k+1}(t), \ldots, h_m(t))$ is an orthonormal basis of the orthogonal complement $K(t)^\perp$ of $K(t)$ in $F(\xi(t))$.

Let $h_i^\#(t) = \sigma_{\xi(t)}(h_i(t))$. Then $h_1^\#(t) = \ldots = h_{m-k}^\#(t) = 0$, and the vectors $h_{m-k+1}^\#(t), \ldots, h_m^\#(t)$ form a basis of $\mathcal{G}(t)$.

For $t \in I_{k,J,0} \cap Q$, the vector $\dot{\xi}(t)$ belongs to $\mathcal{G}(t)$, so it can be expressed in a unique way as a linear combination $\dot{\xi}(t) = \sum_{j=m-k+1}^{m} \alpha^j(t) h_j^\#(t)$, and the coefficient functions α^j are obviously measurable. If we let $\tilde{\omega}(t) = \sum_{j=m-k+1}^{m} \alpha^j(t) h_j(t)$, then $\tilde{\omega}$ is measurable on $I_{k,J,0} \cap Q$ and $\sigma_{\xi(t)}(\tilde{\omega}(t)) = \dot{\xi}(t)$. If $t \in I_{k,J,0} \cap Q$, and $w \in F(\xi(t))$ is such that $\sigma_{\xi(t)}(w) = \dot{\xi}(t)$, then $w - \tilde{\omega}(t) \in K(t)^\perp$ and $\tilde{\omega}(t) \in K(t)$, so $||w||_{G} \geq ||\tilde{\omega}(t)||_{G}$. Therefore $||\tilde{\omega}(t)||_{G} = ||\dot{\xi}(t)||_{G}$. This shows that $\tilde{\omega} \equiv \omega_\xi$ on $I_{k,J,0} \cap Q$. So ω_ξ is measurable on $I_{k,J,0} \cap Q$, and our proof is complete. $\square$

Proof of Proposition 3.6. The first assertion is trivial. To prove the second one, we observe that, if $J \subseteq I(\xi)$ is a compact interval, then $\int_J ||\omega_\xi(t)||_G \, dt = \int_J ||\dot\xi(t)||_G \, dt$. If this integral is finite for every choice of J, then $\eta_{\mathbf{f},\xi,J}$ is integrable whenever $\mathbf{f}$ is an orthonormal local basis of sections of F such that $\xi(J) \subseteq U_{\mathbf{f}}$. Therefore ξ is E-admissible. The converse is trivial, since the existence of some integrable control representation η of ξ on a compact subinterval J of $I(\xi)$ implies that $||\omega_\xi||_G$ is integrable on J, because ω_ξ is measurable and $||\omega_\xi(t)||_G \leq ||\eta(t)||$. □

A.1. Proposition. *Let I_1, I_2 be compact subintervals of $\mathbb{R}$, and let $\theta : I_2 \to I_1$ be absolutely continuous, monotonically nondecreasing, and surjective. Let $h : I_1 \to [0,\infty]$ be a Lebesgue-measurable nonnegative extended-real-valued function. Then the function $I_2 \ni t \to h\big(\theta(t)\big).\dot\theta(t) \in [0,\infty]$ is Lebesgue measurable, and*

$$\int_{I_1} h(s)\, ds = \int_{I_2} h\big(\theta(t)\big).\dot\theta(t)\, dt\,, \tag{A.1}$$

using, in the right-hand side, the convention that $\infty.0 = 0$.

PROOF. We can assume without loss of generality that h is bounded. Indeed, we can truncate h by defining $h_N(t) = \min(h(t), N)$, so the h_N are bounded. Then $0 \leq h_N \uparrow h$ and $0 \leq (h_N \circ \theta).\dot\theta \uparrow (h \circ \theta).\dot\theta$ pointwise as $N \uparrow \infty$, implying that $\int_{I_1} h_N \uparrow \int_{I_1} h$ and $\int_{I_2} (h_N \circ \theta).\dot\theta \uparrow \int_{I_2} (h \circ \theta).\dot\theta$. (Notice that the statement that $(h_N \circ \theta).\dot\theta \uparrow (h \circ \theta).\dot\theta$ pointwise is valid *provided that we use the convention that $\infty.0 = 0$.*)

Now, any bounded measurable function h is a uniform limit of finite linear combinations of indicator functions of measurable sets, and the set of measurable functions for which (A.1) holds is obviously closed under uniform limits, since $\dot\theta$ is integrable. So it suffices to prove that (A.1) holds when h is the indicator function χ_Q of a Lebesgue-measurable subset Q of I_1.

If Q is a closed interval $[\alpha,\beta]$, then (A.1) holds. (Indeed, in this case $\int_{I_1} h = \beta - \alpha$, and $\int_{I_2} (h \circ \theta).\dot\theta = \int_{\theta^{-1}([\alpha,\beta])} \dot\theta(t)\, dt$. Since θ is monotonically nondecreasing and surjective, the set $Z = \theta^{-1}([\alpha,\beta])$ is a closed interval of the form $[\alpha',\beta']$, where α' and β' are such that $\theta(\alpha') = \alpha$ and $\theta(\beta') = \beta$. But then

$$\int_{\theta^{-1}([\alpha,\beta])} \dot\theta(t)\, dt = \int_{\alpha'}^{\beta'} \dot\theta(t)\, dt = \theta(\beta') - \theta(\alpha') = \beta - \alpha = \int_{I_1} h\,,$$

so (A.1) holds, as stated.)

It is clear that the set S of all measurable Q such that (A.1) holds for $h = \chi_Q$ is a monotone class, that is, a family of sets closed under increasing unions and under decreasing unions. Since S contains all compact subintervals of I_1, it follows easily that S contains all subintervals of I_1. Moreover, S is clearly closed under finite disjoint unions. So S contains all finite unions of intervals. It then follows from the Monotone Class Theorem (cf. [15]) that S contains all Borel measurable subsets of I_1. If Q is an arbitrary Lebesgue measurable subset of I_1, then we can write $Q = Q_1 \cup Q_2$, with Q_1 Borel measurable, Q_2 a null set, and $Q_1 \cap Q_2 = \emptyset$. Since $Q_1 \in S$, it suffices to prove that $Q_2 \in S$. We can find a Borel measurable subset P of I_2 such that $Q_2 \subseteq P$ and P has measure zero. If h is the indicator function of P, then we know that $\int_{I_1} h = \int_{I_2} (h \circ \theta).\dot{\theta}$. So $\int_{I_2} (h \circ \theta).\dot{\theta} = 0$. Therefore $(h \circ \theta).\dot{\theta}$ vanishes almost everywhere. Let h' be the indicator function of Q_2. Then $0 \leq h' \leq h$ pointwise. So $0 \leq (h' \circ \theta).\dot{\theta} \leq (h \circ \theta).\dot{\theta}$ pointwise. Therefore $(h' \circ \theta).\dot{\theta}$ vanishes almost everywhere. So $\int_{I_2} (h' \circ \theta).\dot{\theta} = 0 = \int_{I_1} h'$. Therefore $Q_2 \in S$, and our proof is complete. □

A.2. Proposition. *Let I_1, I_2 be compact subintervals of $\mathbb{R}$, and let $\theta : I_2 \to I_1$ be absolutely continuous, monotonically nondecreasing, and surjective. Then (i) if $S \subseteq I_2$ is Lebesgue measurable, it follows that $\theta(S)$ is Lebesgue measurable; (ii) if S has measure zero, then $\theta(S)$ has measure zero; (iii) if $B(\theta)$ is the set of all $t \in I_2$ such that $\dot{\theta}(t)$ exists and is $\neq 0$. Then $I_1 \backslash \theta\Big(B(\theta)\Big)$ has measure zero.*

PROOF. Suppose S has measure zero. Let $\varepsilon > 0$. Find $\delta > 0$ such that $\int_Q \dot{\theta} < \varepsilon$ whenever Q is measurable and such that $\mathrm{meas}(Q) < \delta$. Then find a relatively open subset U of I_2 such that $\mathrm{meas}(U) < \delta$ and $S \subseteq U$. Express U as the disjoint union of a sequence $\{I_j\}$ of relatively open subintervals of I_2. If a_j, b_j are the left and right endpoints of I_j, then $\theta(S) \subseteq \theta(U) \subseteq \bigcup_j [\theta(a_j), \theta(b_j)]$. So

$$\mathrm{meas}(\theta(S)) \leq \mathrm{meas}(\theta(U)) \leq \sum_j (\theta(b_j) - \theta(a_j)) = \int_U \dot{\theta} < \varepsilon\,.$$

Since ε was arbitrary, $\theta(S)$ has measure zero, and Part (ii) is proved.

Now, if $S \subseteq I_2$ is measurable, then for each positive integer N we can find a compact subset K_N of S such that $\mathrm{meas}(S \backslash K_N) < 2^{-N}$. If $K = \bigcup_N K_N$, then $\mathrm{meas}(S \backslash K) = 0$. So $\mathrm{meas}\Big(\theta(S \backslash K)\Big) = 0$. Moreover, $\theta(S) = \theta(S \backslash K) \cup \theta(K) = \theta(S \backslash K) \bigcup \Big(\bigcup_N \theta(K_N)\Big)$. Since the sets $\theta(K_N)$ are compact, we see that $\theta(S)$ is the union of a set of measure zero and a countable family of compact sets. So S is measurable, and Part (i) is proved.

To prove Part (iii), we let $Q = I_1 \backslash \theta\Big(B(\theta)\Big)$, and observe that Q is measurable by Part (i). Let h be the indicator function of Q. Then Proposition A.1 tells us that

$$\int_{I_1} h(s)\, ds = \int_{I_2} h(\theta(t)).\dot{\theta}(t)\, dt\,. \tag{A.2}$$

Now, if $t \in I_1$ is such that $\dot{\theta}(t)$ exists and is > 0, then $t \in B(\theta)$, so $\theta(t) \in \theta\Big(B(\theta)\Big)$, and then $\theta(t) \notin Q$, so $h(\theta(t)) = 0$, and $h(\theta(t))\dot{\theta}(t) = 0$. Obviously, this last equality also holds if $\dot{\theta}(t)$ exists and is equal to 0. So $h(\theta(t))\dot{\theta}(t) = 0$ for all t for which $\dot{\theta}(t) = 0$ exists, i. e. for almost all $t \in I_2$. Therefore the right-hand side of (A.2) vanishes. So $\mathrm{meas}(P) = \int_{I_1} h(s)\, ds = 0$, and the proof of Part (iii) is complete. □

A.3. Proposition. *Let I_1, I_2 be compact subintervals of $\mathbb{R}$, and let $\theta : I_2 \to I_1$ be absolutely continuous, monotonically nondecreasing, and surjective. Let M be a manifold, and let $\xi : I_1 \to M$ be an absolutely continuous arc. Let $\xi' = \xi \circ \theta$. Then ξ' is absolutely continuous and*

$$\dot{\xi}'(t) = \dot{\xi}(\theta(t)).\dot{\theta}(t) \tag{A.3}$$

for almost all $t \in I_2$, in the sense that, for almost every $t \in I_2$, $\dot{\xi}'(t)$ and $\dot{\theta}(t)$ exist, and either (a) $\dot{\theta}(t) \neq 0$, $\dot{\xi}(\theta(t))$ exists, and (A.3) holds, or (b) $\dot{\theta}(t) = 0$ and $\dot{\xi}'(t) = 0$. □

PROOF. It is clear the result is local, so it suffices to assume that $M = \mathbb{R}^n$. Then (A.3) is really a finite set of identities, one for each scalar component of the vector-valued functions ξ and ξ'. So it suffices to assume that $n = 1$. Let $I_i = [a_i, b_i]$ for $i = 1, 2$.

The function $\xi : I_1 \to \mathbb{R}$ is absolutely continuous, so we can write $\xi = \xi_+ - \xi_-$, where ξ_+ and ξ_- are absolutely continuous and monotonically nondecreasing. (For example, we may take $\xi_+(s) = \int_{a_1}^s |\dot{\xi}(r)|\, dr$.) Let $\xi'_- = \xi_- \circ \theta$, $\xi'_+ = \xi_+ \circ \theta$. Then $\xi' = \xi'_+ - \xi'_-$.

For $s \in I_1$, let

$$h_+(s) = \dot{\xi}_+(s) \text{ if } \dot{\xi}_+(s) \text{ exists}, \qquad h_+(s) = +\infty \text{ if } \dot{\xi}_+(s) \text{ does not exist},$$

$$h_-(s) = \dot{\xi}_-(s) \text{ if } \dot{\xi}_-(s) \text{ exists}, \qquad h_-(s) = +\infty \text{ if } \dot{\xi}_-(s) \text{ does not exist}.$$

Then h_+ and h_- are Lebesgue measurable $[0, \infty]$-valued functions defined everywhere on I_1, so we can conclude from Proposition A.1 that the nonnegative functions $t \to h_+(\theta(t)).\dot{\theta}(t)$ and $t \to h_-(\theta(t)).\dot{\theta}(t)$ are Lebesgue-measurable on I_2. Moreover, Proposition A.1 also yields, when applied to the interval $[a_2, t]$ for any $t \in [a_2, b_2]$, the identities

$$\int_{a_1}^{\theta(t)} h_+(s)\, ds = \int_{a_2}^{t} h_+(\theta(\tau)).\dot{\theta}(\tau)\, d\tau\,, \tag{A.4}$$

$$\int_{a_1}^{\theta(t)} h_-(s)\, ds = \int_{a_2}^{t} h_-(\theta(\tau)).\dot{\theta}(\tau)\, d\tau\,. \tag{A.5}$$

In particular, (A.4) and (A.5) imply that $\int_{a_2}^{b_2} h_+(\theta(\tau)).\dot{\theta}(\tau)\, d\tau = \xi_+(b_1) - \xi_+(a_1) < \infty$ and $\int_{a_2}^{b_2} h_-(\theta(\tau)).\dot{\theta}(\tau)\, d\tau = \xi_-(b_1) - \xi_-(a_1) < \infty$. Therefore the nonnegative measurable functions $t \to h_+(\theta(t)).\dot{\theta}(t)$ and $t \to h_-(\theta(t)).\dot{\theta}(t)$ are integrable. Then (A.4) implies that

$$\xi'_+(t) - \xi'_+(a_2) = \xi_+(\theta(t)) - \xi_+(\theta(a_2)) = \xi_+(\theta(t)) - \xi_+(a_1) = \int_{a_1}^{\theta(t)} \dot{\xi}_+(s)\, ds$$
$$= \int_{a_1}^{\theta(t)} h_+(s)\, ds = \int_{a_2}^{t} h_+(\theta(\tau)).\dot{\theta}(\tau)\, d\tau\,.$$

So ξ'_+ is the indefinite integral of an integrable function, and therefore ξ'_+ is absolutely continuous. A similar argument, using (A.5), shows that

$$\xi'_-(t) - \xi'_-(a_2) = \int_{a_2}^{t} h_-(\theta(\tau)).\dot{\theta}(\tau)\, d\tau$$

for all $t \in I_2$, so ξ'_- is absolutely continuous as well. It then follows that ξ' is absolutely continuous.

The Lebesgue differentiation theorem then implies that there is a subset Z of I_2 such that $I_2 \backslash Z$ has zero Lebesgue measure, having the properties that, for all $t \in Z$:

(i) $\dot{\xi}'_+(t)$ *and* $\dot{\xi}'_-(t)$ *exist and are equal to* $h_+(\theta(t)).\dot{\theta}(t)$ *and* $h_-(\theta(t)).\dot{\theta}(t)$*, respectively,*

(ii) $\dot{\theta}(t)$ *exists.*

Now let $t \in Z$. Then the products $h_+(\theta(t)).\dot{\theta}(t)$ and $h_-(\theta(t)).\dot{\theta}(t)$ are finite. Suppose first that $\dot{\theta}(t) \neq 0$. Then $h_+(\theta(t))$ and $h_-(\theta(t))$ must themselves be finite. It then follows from our definition of h_+ and h_- that $\dot{\xi}(t)$ exists and is equal to $h_+(\theta(t)) - h_-(\theta(t))$. Moreover, (i) implies that $\dot{\xi}'(t)$ exists and is equal to $\Big(h_+(\theta(t)) - h_-(\theta(t))\Big).\dot{\theta}(t)$, i. e. to $\dot{\xi}(t).\dot{\theta}(t)$. So (A.3) holds.

Next consider the case when $\dot{\theta}(t) = 0$. In this case, we are not guaranteed that $h_+(\theta(t))$ and $h_-(\theta(t))$ are finite, but it is still true that $h_+(\theta(t)).\dot{\theta}(t) = h_-(\theta(t)).\dot{\theta}(t) = 0$. Then (i) implies that $\dot{\xi}'(t) = 0$ as well. $\square$

A.4. Remark. *It is not possible to strengthen the conclusion of Proposition A.3 to obtain a subset* Z *of* I_2 *of full measure, such that* $\dot{\xi}'(t)$*,* $\dot{\xi}(\theta(t))$ *and* $\dot{\theta}(t)$ *exist for all* $t \in Z$*.* For example, θ could be constant on a nontrivial subinterval J of I_2, and its value on J could be a point of I_1 where ξ is not differentiable. Then the set of points t such that $\dot{\xi}(\theta(t))$ does not exist will have positive measure. $\square$

Proof of Proposition 3.16. Let $I_i = I(\xi_i)$. Let $\theta : I_2 \to I_1$ be absolutely continuous, monotonically nondecreasing, surjective, and such that $\xi_1 \circ \theta = \xi_2$. Then Proposition A.3 tells us that there is a subset Z of I_2 of full measure such that, if $t \in Z$, then (a) $\dot{\xi}_2(t)$ and $\dot{\theta}(t)$ exist, (b) $\dot{\xi}_2(t)$ vanishes if $\dot{\theta}(t)$ does, and (c) if $\dot{\theta}(t) \neq 0$ then $\dot{\xi}_1(\theta(t))$ exists and satisfies $\dot{\xi}_2(t) = \dot{\xi}_1(\theta(t)).\dot{\theta}(t)$.

If we let $h(s) = ||\dot{\xi}_1(s)||_G$ when $\dot{\xi}_1(s)$ exists, and $h(s) = +\infty$ otherwise, then

$$||\dot{\xi}_2(t)||_G = h(\theta(t)).\dot{\theta}(t) \tag{8.6}$$

for all $t \in Z$, if we make the convention that $\infty.0 = 0$, as in the statement of Proposition A.1. So $||\xi_1||_G = \int_{I_1} h = \int_{I_2} (h \circ \theta).\dot{\theta} = \int_{I_2} ||\dot{\xi}_2(t)||_G \, dt = ||\xi_2||_G$, using Proposition A.1 and Formula (8.6). $\square$

Proof of Proposition 3.17. We let $I(\xi) = [a, b]$, and define $L = ||\xi||_G$.

We first prove uniqueness. Let ξ_1, ξ'_1 be two members of $TR(\xi) \cap \mathcal{A}^{PAL,0}(\mathcal{M})$. Then $||\xi_1||_G = ||\xi'_1||_G = L$ by Proposition 3.16, so the intervals $I(\xi_1)$ and $I(\xi'_1)$ are both equal to $[0, L]$ by (3.3).

Let $\theta : I(\xi) \to [0, L]$ and $\theta' : I(\xi) \to [0, L]$ be absolutely continuous, nondecreasing, surjective, and such that $\xi_1 \circ \theta = \xi'_1 \circ \theta' = \xi$. If $t \in I(\xi)$ is arbitrary, use ξ_t, $\xi_{1,t}$, $\xi'_{1,t}$ to denote, respectively, the restrictions of ξ, ξ_1 and ξ'_1 to the intervals $[a, t]$, $[0, \theta(t)]$ and $[0, \theta'(t)]$. Also, let θ_t and θ'_t be the restrictions of θ and θ' to $[a, t]$. Then $\xi_{1,t}$ and $\xi'_{1,t}$ are both members of $TR(\xi_t) \cap \mathcal{A}^{PAL,0}(\mathcal{M})$. So we already know that their domains $[0, \theta(t)]$ and $[0, \theta'(t)]$ coincide. Therefore $\theta(t) = \theta'(t)$.

We have thus shown that $\theta(t) = \theta'(t)$ for all $t \in [a, b]$. If $s \in [0, L]$, then we can find $t \in [a, b]$ such that $\theta(t) = s$. Then $\theta'(t) = s$ as well. So $\xi_1(s) = \xi_1(\theta(t)) = \xi(t)$, and $\xi'_1(s) = \xi'_1(\theta'(t)) = \xi(t)$. Therefore $\xi_1 \equiv \xi'_1$, proving the uniqueness of ξ_1. The uniqueness of ξ_2 then follows trivially.

The existence of ξ_2 follows easily from the corresponding assertion for ξ_1 if $||\xi||_G > 0$, and is proved directly if $||\xi||_G = 0$ but ξ is nontrivial.

To prove the existence of ξ_1, we first observe that the conclusion is valid for a finite concatenation of curves if it is valid for each of them. Therefore we can work locally, that is, we can assume that $M = \mathbb{R}^n$ and there exists a global orthonormal basis $\mathbf{f} = (f_1, \ldots, f_m)$ of sections of F. We let $g_i = \sigma(f_i)$.

We then define a map $\theta : [a,b] \to [0,L]$ by letting $\theta(t) = \int_a^t ||\dot\xi(\tau)||_G \, d\tau$. Then θ is absolutely continuous, because the function $t \to ||\dot\xi(t)||_G$ is integrable. Moreover, θ is obviously monotonically nondecreasing and surjective.

If $a \le t_1 < t_2 \le b$ and $\theta(t_2) = \theta(t_1)$, then $\int_{t_1}^{t_2} ||\dot\xi(s)||_G \, ds = 0$, so $||\dot\xi(t)||_G = 0$ for almost all $t \in [t_1, t_2]$, and then $\dot\xi(t) = 0$ for almost all $t \in [t_1, t_2]$ by Remark 3.3. Therefore $\xi(t_2) = \xi(t_1)$. This enables us to define $\xi_1 : [0,L] \to M$ by letting $\xi_1(s) = \xi(t)$ if $\theta(t) = s$. It is then clear that $\xi_1 \circ \theta = \xi$.

Let $C = \sup \Big\{ \Big(\sum_{i=1}^m ||g_i(\xi(t))||^2 \Big)^{\frac{1}{2}} : t \in [a,b] \Big\}$. Let $\eta = \eta_{\mathbf{f},\xi,[a,b]}$. If $0 \le s_1 < s_2 \le L$, and $s_i = \theta(t_i)$, then

$$||\,\xi_1(s_2) - \xi_1(s_1)\,|| = ||\,\xi(t_2) - \xi(t_1)\,|| = ||\int_{t_1}^{t_2} \dot\xi(t)\,dt|| = ||\int_{t_1}^{t_2} \Big(\sum_{i=1}^m \eta^i(t) g_i(\xi(t))\Big)\,dt\,||$$
$$\le C \int_{t_1}^{t_2} ||\eta(t)||\,dt = C\Big(\theta(t_2) - \theta(t_1)\Big) = C(s_2 - s_1)\,,$$

so ξ_1 is a Lipschitz arc.

To conclude the proof, we have to show that $\xi_1 \in \mathcal{A}_E(M)$, i.e. that $\dot\xi_1(s) \in E(\xi_1(s))$ for almost every $s \in I_1$. Let B_1 denote the set of all points $t \in [a,b]$ such that ξ and θ are differentiable at t and $\dot\xi(t) \in E(\xi(t))$. Then $[a,b]\backslash B_1$ has measure zero. Let B be the set of those $t \in [a,b]$ such that $\dot\theta(t)$ exists and is not equal to zero. Then Proposition A.2 tells us that $[0,L]\backslash\theta(B)$ has measure zero. In addition, $B\backslash B_1$ has measure zero, so Proposition A.2 also implies that $\theta(B\backslash B_1)$ has measure zero. Let H_1 be the set of those $s \in [0,L]$ such that the derivative $\dot\xi_1(s)$ does not exist. Then meas$(H_1) = 0$, because ξ_1 is Lipschitz. Define $H = H_1 \cup \Big([0,L]\backslash\theta(B)\Big) \cup \theta(B\backslash B_1)$. Then meas$(H) = 0$. Suppose $s \in [0,L]$ and $s \notin H$. Then $\dot\xi_1(s)$ exists, because $s \notin H_1$. Also, s belongs to $\theta(B)$, so $s = \theta(t)$ for some $t \in B$. Then t must belong to B_1, for otherwise s would be in $\theta(B\backslash B_1)$. Therefore ξ and θ are differentiable at t, and $\dot\xi(t) \in E(\xi(t))$. Since $\xi = \xi_1 \circ \theta$, and ξ_1 is differentiable at $s = \theta(t)$, the chain rule implies that $\dot\xi(t) = \dot\xi_1(s).\dot\theta(t)$. Since $\dot\xi(t) \in E(\xi(t))$, and $\dot\theta(t) \neq 0$ (because $t \in B$), it follows that $\dot\xi_1(s) \in E(\xi(t)) = E(\xi_1(s))$. Since this is true for all $s \in [0,L]\backslash H$, and H has measure zero, our proof is complete. $\square$

Proof of Proposition 3.20. It is clear that $(b) \Rightarrow (a)$. To prove the reverse implication, it suffices to work locally. So we assume that $M = \mathbb{R}^n$ and there exists a global orthonormal basis $\mathbf{f} = (f_1, \ldots, f_m)$ of sections of F. We let $I(\xi) = [a, b]$, and write $g_i = \sigma(f_i)$.

For each positive integer N, let $t_{k,N} = a + k2^{-N}(b-a)$, for $k = 0, 1, \ldots, 2^N$. Let $x_{k,N} = \xi(t_{k,N})$. Then for each $k \in \{1, \ldots, 2^N\}$ the distance $d_{\mathcal{M}}(x_{k-1,N}, x_{k,N})$ is less than $c(b-a)2^{-N}$, so we may find an E-admissible E-arc $\xi_{k,N} : [t_{k-1,N}, t_{k,N}] \to \mathbb{R}^n$ that satisfies

$$\xi_{k-1,N}(t_{k-1,N}) = x_{k-1,N}\,, \quad \xi_{k,N}(t_{k,N}) = x_{k,N}\,, \quad ||\xi_{k,N}||_{G} \leq c(b-a)2^{-N}(1 + 2^{-N})\,,$$

and is parametrized by constant times arc-length.

We then find a control representation $\eta_{k,N} : [t_{k-1,N}, t_{k,N}] \to \mathbb{R}^m$ of $\xi_{k,N}$, so

$$\dot{\xi}_{k,N}(t) = \sum_{i=1}^{m} \eta^i_{k,N}(t) g_i(\xi_{k,N}(t)) \quad \text{for a. e.} \quad t \in [t_{k-1,N}, t_{k,N}]\,,$$

and $||\eta_{k,N}(t)|| = c_{k,N}$ for a. e. $t \in [t_{k-1,N}, t_{k,N}]$, where the constant $c_{k,N}$ is such that $2^{-N}(b-a)c_{k,N} = ||\xi_{k,N}||_{G}$. Since $||\xi_{k,N}||_{G} \leq c(b-a)2^{-N}(1+2^{-N})$, it follows that $c_{k,N}$ satisfies $c_{k,N} \leq c(1 + 2^{-N})$.

Next, define $\xi_N : [a, b] \to \mathbb{R}^n$, $\eta_N : [a, b] \to \mathbb{R}^m$, by letting

$$\xi_N(t) = \xi_{k,N}(t) \quad \text{and} \quad \eta_N(t) = \eta_{k,N}(t) \quad \text{whenever} \quad t \in [t_{k-1,N}, t_{k,N}]\,.$$

Then each ξ_N is a trajectory for the control η_N, and η_N satisfies the pointwise bound $||\eta_N(t)|| \leq c(1 + 2^{-N})$ for a. e. t. It follows that the set $\{\eta_N : N \in \mathbb{N}\}$ is weak*-precompact in $L^\infty([a, b], \mathbb{R}^m)$. So there exists an infinite subset J of $\mathbb{N}$, such that the limit $\eta_J = \text{weak}^*\text{-}\lim_{N \to_J \infty} \eta_N$ exists. Then the curves ξ_N, $N \in J$, converge uniformly to a limit ξ_J, which satisfies $\xi_J(t) = \xi_J(a) + \int_a^t \left(\sum_{i=1}^m \eta^i_J(\tau) g_i(\xi_J(\tau)) \right) d\tau$ for $t \in [a, b]$.

On the other hand, $\xi_N(t_{k,N}) = x_{k,N} = \xi(t_{k,N})$ by construction. Given any $t \in [a, b]$ which is of the form $t_{k,N}$ for some N, k, we also have $t = t_{k2^{\nu-N},\nu}$ for all ν such that $\nu \geq N$. This happens, in particular, for infinitely many values of ν belonging to J. So $\xi_\nu(t) = \xi(t)$ for infinitely many ν's in J, and then $\xi_J(t) = \xi(t)$. Since this happens for all t's in a dense subset of $[a, b]$, we can conclude that $\xi_J \equiv \xi$. So ξ is in fact an E-admissible E-arc, having a control representation with control η_J. It is clear that $||\eta_J(t)|| \leq c$ a. e., so $||\dot{\xi}(t)||_{G} \leq c$ a. e., and our conclusion is proved. $\square$

Proof of Proposition 3.27. Let $\xi \in \mathcal{A}_{[a,b}(\mathcal{M})$, and write $(\bar{x}, \hat{x}) = \partial\xi$. If ξ is not an E-curve, then $||\dot{\xi}(t)||_{G} = \infty$ for t in a set of positive measure, so both $||\xi||_{G}$ and $\mathcal{E}_{\mathcal{M}}(\xi)$ are infinite. If ξ is an E-curve but isn't admissible, then $\int_a^b ||\dot{\xi}(t)||_{G} dt = +\infty$. Since $||\dot{\xi}(t)||^2_{G} \geq ||\dot{\xi}(t)||_{G} - 1$ for all t, we have $\int_a^b ||\dot{\xi}(t)||^2_{G} dt = +\infty$ as well.

Now suppose ξ is E-admissible. Let ξ_1 be its canonical reparametrization, and define $\xi_2(t) = \xi_1\left(\frac{t-a}{b-a}\right)$ for $a \leq t \leq b$. Then the function $t \to ||\dot{\xi}_2(t)||_{G}$ is equal a. e. to a constant c, and $||\xi||_{G} = ||\xi_2||_{G} = (b-a).c$. Therefore

$$\begin{aligned} d_{\mathcal{M}}(\bar{x}, \hat{x}) \leq ||\xi||_{G} = \int_a^b ||\dot{\xi}(t)||_{G} dt \leq (b-a)^{1/2}.\Big(\int_a^b ||\dot{\xi}(t)||^2_{G} dt \Big)^{1/2} \\ = (b-a)^{1/2} \mathcal{E}_{\mathcal{M},a,b}(\xi)^{1/2}\,, \end{aligned} \tag{A.7}$$

and

$$(b-a).d^{\mathcal{E},a,b}(\bar{x},\hat{x}) \leq (b-a).\mathcal{E}_{\mathcal{M},a,b}(\xi_2) = (b-a).\int_a^b ||\dot{\xi}_2(t)||_G^2\,dt \\ = (b-a)^2c^2 = ||\xi_2||_G^2 = ||\xi||_G^2\,. \tag{A.8}$$

In (A.7), we can take the infimum with respect to all ξ such that $\partial\xi = (\bar{x},\hat{x})$, and conclude that $d_{\mathcal{M}}(\bar{x},\hat{x}) \leq (b-a)^{1/2}d^{\mathcal{E},a,b}(\bar{x},\hat{x})^{1/2}$. If, on the other hand, we take the infimum in (A.8), we get $d_{\mathcal{M}}(\bar{x},\hat{x})^2 \geq (b-a)^{1/2}d^{\mathcal{E},a,b}(\bar{x},\hat{x})$. This proves Part (A).

To prove (B), we observe that, if ξ is an energy minimizer, then

$$(b-a)^{1/2}.\mathcal{E}_{\mathcal{M},a,b}(\xi)^{1/2} = (b-a)^{1/2}.d^{\mathcal{E},a,b}(\bar{x},\hat{x})^{1/2} = d_{\mathcal{M}}(\bar{x},\hat{x})\,,$$

so the inequalities of (A.7) are in fact equalities. Therefore $d_{\mathcal{M}}(\bar{x},\hat{x}) = ||\xi||_G$ —showing that ξ is a length minimizer— and $\int_a^b ||\dot{\xi}(t)||_G\,dt = (b-a)^{1/2}\Big(\int_a^b ||\dot{\xi}(t)||_G^2\,dt\Big)^{1/2}$, which, as is well known, implies that the function $t \to ||\dot{\xi}(t)||_G$ is a. e. constant. So ξ is parametrized by constant times arc-length, and the implication (B.1)$\Rightarrow$(B.2) has been proved.

Conversely, suppose ξ is a length minimizer parametrized by constant times arc-length. Then $\xi = \xi_2$, and $||\xi||_G^2 = d_{\mathcal{M}}(\bar{x},\hat{x})^2 = (b-a).d^{\mathcal{E},a,b}(\bar{x},\hat{x})$. So the inequality of (A.8) is in fact an equality. Therefore $d^{\mathcal{E},a,b}(\bar{x},\hat{x}) = \mathcal{E}_{\mathcal{M},a,b}(\xi_2) = \mathcal{E}_{\mathcal{M},a,b}(\xi)$, showing that ξ is an energy minimizer. □

Proof of the assertions of Example 4.2. Suppose $\xi \in \mathrm{Max}(\mathcal{C}^{PAL}(\mathcal{M}))$ and τ is a limit point of $I(\xi)$ such that $\bar{x} = \lim_{t\to_{I(\xi)}\tau}\xi(t)$ exists. Then $\tau \in I(\xi)$, for otherwise we would be able to extend ξ to $I(\xi)\cup\{\tau\}$ and obtain a new curve in $\mathcal{C}^{PAL}(\mathcal{M})$ strictly extending ξ, thereby contradicting the maximality of ξ.

Find an orthonormal basis $\mathbf{f} = (f_1,\ldots,f_m)$ of sections of F in a neighborhood of $\bar{x}$, and write $g_i = \sigma(f_i)$. Then the maximality of ξ implies that $I(\xi)$ is nontrivial. Since $\tau \in I(\xi)$, there is an $\varepsilon > 0$ such that at least one of the intervals $A_- =]\,\tau-\varepsilon,\tau\,]$, $A_+ = [\,\tau,\tau+\varepsilon\,[$, is contained in $I(\xi)$. By making ε smaller, if necessary, we may assume that $\xi\Big((A_-\cup A_+)\cap I(\xi)\Big) \subseteq U_{\mathbf{f}}$.

Suppose, for example, that $A_- \subseteq I(\xi)$, Choose an integrable control $\eta : A_- \to \mathbb{R}^m$ such that $\dot{\xi}(t) = \sum_i \eta^i(t)g_i(\xi(t))$ and $||\dot{\xi}(t)||_G = ||\eta(t)|| = 1$ for a. e. $t \in A_-$. Extend η to an integrable function $\tilde{\eta} : A_-\cup A_+ \to \mathbb{R}^m$ such that $||\tilde{\eta}(t)|| = 1$ for a. e. $t \in A_-\cup A_+$. By the existence theorem for solutions of ordinary differential equations, there exists —after making ε smaller, if necessary— a solution $\xi_1 : A_-\cup A_+ \to M$ of the initial value problem $\dot{x}(t) = \sum_i \tilde{\eta}^i(t)g_i(x(t))$, $x(\tau) = \xi(b)$. Then $\xi_1 = \xi$ on A_-. If we reparametrize $\xi_1 \lceil A_+$ by arc-length, and then concatenate the result with ξ, we obtain a curve $\xi' \in \mathcal{C}^{PAL}(\mathcal{M})$ that extends ξ to a strictly larger interval, contradicting the maximality of ξ.

The proof that $\mathcal{C}^{a,\leq c}(\mathcal{M})$ has the extension property is identical to the above argument, except that now η and $\tilde{\eta}$ are chosen so that $||\eta(t)|| \leq c$ and $||\tilde{\eta}(t)|| \leq c$ almost everywhere, and there is no need to reparametrize $\xi_1 \lceil A_+$. □

Proof of Proposition 4.5. The set $\mathcal{A}^{a,\leq c}(\mathcal{M})$ is an *arc system* on the metric space $(M, d_{\mathcal{M}})$, with time domain $\mathbb{R}$, in the sense of [25]. Then $\mathcal{C}^{a,\leq c}(\mathcal{M})$ is exactly the set of all curves ξ in M such that the restriction $\xi \lceil J$ of ξ to every compact subinterval of $I(\xi)$ is in $\mathcal{A}^{a,\leq c}(\mathcal{M})$. That is, $\mathcal{C}^{a,\leq c}(\mathcal{M}) = \left(\mathcal{A}^{a,\leq c}(\mathcal{M})\right)^{\#}$, using the notations of [25].

Moreover, $\mathcal{A}^{a,\leq c}(\mathcal{M})$ has the "compactness property" of [25]. Indeed, if $\{\xi_j\}_{j=1}^{\infty}$ is a sequence of arcs in $\mathcal{A}^{a,\leq c}(\mathcal{M})$, whose graphs $G(\xi_j) \stackrel{\text{def}}{=} \left\{ (\xi_j(t), t) : t \in I(\xi_j) \right\}$ are all contained in some fixed compact subset K of $M \times \mathbb{R}$, then there is an infinite subset S of $\mathbb{N}$ such that, if $I(\xi_j) = [a_j, b_j]$, then the limits $a = \lim_{j \to_S \infty} a_j$ and $b = \lim_{j \to_S \infty} b_j$ exist. Since the ξ_j belong to $\mathcal{A}^{a,\leq c}(\mathcal{M})$, the inequality $d_{\mathcal{M}}(\xi_j(t), \xi_j(t')) \leq c|t - t'|$ holds whenever $t, t' \in [a_j, b_j]$, so the sequence $\{\xi_j\}_{j=1}^{\infty}$ is equicontinuous. Since all the ξ_j are contained in a fixed compact subset of M, the Ascoli-Arzelà Theorem implies that there is an infinite subset S' of S such that the arcs ξ_j, $j \in S'$, converge uniformly (in the sense that, if $a_j \leq t_j \leq b_j$ and $t_j \to t$ then $\xi_j(t_j) \to \xi(t)$) to an arc $\xi : [a, b] \to M$. Then the inequality $d_{\mathcal{M}}(\xi(t), \xi(t')) \leq c|t - t'|$ holds whenever $t, t' \in [a, b]$. Therefore Proposition 3.20 implies that $\xi \in \mathcal{A}^{a,\leq c}(\mathcal{M})$.

Moreover, is t is clear that $\mathcal{A}^{a,\leq c}(\mathcal{M})$ also has the "local existence property" of [25]. So $\mathcal{A}^{a,\leq c}(\mathcal{M})$ satisfies the hypotheses of Proposition 2 of [25]. The conclusion then follows from that proposition. ◻

Proof of Theorem 4.7. The implications $(iii) \Rightarrow (ii) \Rightarrow (i) \Rightarrow (iv)$ are trivial. To see that $(iv) \Rightarrow (iii)$, we assume that (iv) holds, pick an $\bar{x} \in M$, and $r > 0$, and a sequence $\{x_j\}$ in $B_{\mathcal{M}}(\bar{x}, r)$, and try to find a convergent subsequence and a length minimizer joining $\bar{x}$ to the limit of this subsequence.

Pick arcs $\xi_j \in \mathcal{A}^{PAL,0}(\mathcal{M})$ such that $\partial\xi_j = (\bar{x}, x_j)$ and

$$L_j \stackrel{\text{def}}{=} ||\xi_j||_G \leq d_{\mathcal{M}}(\bar{x}, x_j) + \frac{1}{j}. \tag{A.9}$$

Then $L_j \leq r + 1$ for all j. Let $I(\xi_j) = [0, L_j]$. Since the sequence $\{L_j\}$ is bounded we may assume, after passing to a subsequence, that $L = \lim_{j \to \infty} L_j$ exists.

Let $\tilde{\xi}_j \in \text{Max}(\mathcal{C}^{a,\leq 1}(\mathcal{M}))$ be an extension of ξ_j. (The existence of $\tilde{\xi}_j$ follows easily from Zorn's Lemma.) Since (iv) holds, the domain $I(\tilde{\xi}_j)$ of each $\tilde{\xi}_j$ is the whole real line. By Proposition 4.5, there exist an infinite subset S of $\mathbb{N}$ and a curve $\tilde{\xi} \in \text{Max}(\mathcal{C}^{a,\leq 1}(\mathcal{M}))$ such that $0 \in I(\tilde{\xi})$ and $\tilde{\xi}_j \to \tilde{\xi}$ on compact intervals. Since (iii) holds, $I(\tilde{\xi})$ is in fact equal to $\mathbb{R}$. In particular, this implies that the sequence $\{\xi_j(L_j)\}_{j \in S}$ converges to $\tilde{\xi}(L)$, so the limit $x = \lim_{j \to_S \infty} x_j$ exists and equals $\tilde{\xi}(L)$.

Let ξ be the restriction of $\tilde{\xi}$ to $[0, L]$. If $t_1, t_2 \in [0, L]$, then $\xi(t_\ell) = \lim_{j \to_S \infty} \xi_j(t_\ell)$ for $\ell = 1, 2$. So $d_{\mathcal{M}}(\xi(t_1), \xi(t_2)) = \lim_{j \to_S \infty} d_{\mathcal{M}}(\tilde{\xi}_j(t_1), \tilde{\xi}_j(t_2))$. Since $\tilde{\xi}_j \in \mathcal{C}^{a,\leq 1}(\mathcal{M})$, Proposition 3.20 implies that $d_{\mathcal{M}}(\tilde{\xi}_j(t_1), \tilde{\xi}_j(t_2)) \leq |t - t'|$. So $d_{\mathcal{M}}(\xi(t_1), \xi(t_2) \leq |t_1 - t_2|$ for all $t_1, t_2 \in [0, L]$. Then Proposition 3.20 tells us that $\xi \in \mathcal{C}^{a,\leq 1}(\mathcal{M})$. This implies, in particular, that $||\xi||_G \leq L$, and $d_{\mathcal{M}}(\bar{x}, x) \leq ||\xi||_G$. On the other hand,

$$L = \lim_{j \to \infty} L_j \leq \lim_{j \to_S \infty} d_{\mathcal{M}}(\bar{x}, x_j) + \frac{1}{j} = d_{\mathcal{M}}(\bar{x}, x).$$

So $||\xi||_{G} \le d_{\mathcal{M}}(\bar{x}, x)$, and then $||\xi||_{G} = d_{\mathcal{M}}(\bar{x}, x)$. Therefore ξ is a length minimizer, and the implication $(iv) \Rightarrow (iii)$ is proved.

Now assume that $LMin^{PAL}(\mathcal{M})$ has the extension property. We prove that $(i) \Rightarrow (v)$. Let $\xi \in \mathrm{Max}(LMin^{PAL}(\mathcal{M}))$. Then the extension property implies, to begin with, that $I(\xi)$ is open, and in particular nontrivial. Let $b = \sup I(\xi)$, $a = \inf I(\xi)$. Then $a \notin I(\xi)$ and $b \notin I(\xi)$. If b was finite, then the completeness of $(M, d_{\mathcal{M}})$ would imply that $\lim_{t\uparrow b} \xi(t)$ exists, since $d_{\mathcal{M}}(\xi(t), \xi(t')) \le |t - t'|$ whenever $t, t' \in I(\xi)$, because ξ is parametrized by arc-length. But then the extension property would imply that $b \in \mathrm{Int}(I(\xi))$, which is a contradiction. So $b = +\infty$. Similarly, $a = -\infty$.

We now prove that $(v) \Longrightarrow (iii)$. We pick $\bar{x} \in M$, $r > 0$, and a sequence $\{x_j\}$ in $B_{\mathcal{M}}(\bar{x}, r)$, and try to find a convergent subsequence and a length minimizer joining $\bar{x}$ to the limit $\hat{x}$ of this subsequence. Pick arcs $\xi_j \in \mathcal{A}^{PAL,0}(\mathcal{M})$ such that $\partial\xi_j = (\bar{x}, x_j)$ and (A.9) holds. Then $L_j \le r + 1$ for all j. Let $I(\xi_j) = [0, L_j]$. Since the sequence $\{L_j\}$ is bounded we may assume, after passing to a subsequence, that $L = \lim_{j\to\infty} L_j$ exists.

Let $\tilde{\xi}_j \in \mathrm{Max}(\mathcal{C}^{a,\le 1}(\mathcal{M}))$ be an extension of ξ_j. (The existence of $\tilde{\xi}_j$ follows easily from Zorn's Lemma.) By Proposition 4.5, there exist an infinite subset S of $\mathbb{N}$ and a curve $\tilde{\xi} \in \mathrm{Max}(\mathcal{C}^{a,\le 1}(\mathcal{M}))$ such that $0 \in I(\tilde{\xi})$ and $\{\tilde{\xi}_j\}_{j\in S}$ converges to $\tilde{\xi}$ on compact intervals. It is then clear that $\tilde{\xi}(0) = \bar{x}$.

We claim that $L \in I(\tilde{\xi})$. To see this, we first observe that, since $\tilde{\xi} \in \mathrm{Max}(\mathcal{C}^{a,\le 1}(\mathcal{M}))$, the extension property implies that $I(\tilde{\xi})$ is open. Let $b = \sup I(\tilde{\xi})$, so $b > 0$, because $0 \in I(\tilde{\xi})$. Let τ be such that $0 < \tau \le L$ and $\tau < b$. Then the compact interval $[0, \tau]$ is contained in $I(\tilde{\xi})$. So $[0, \tau] \subseteq I(\tilde{\xi}_j)$ if $j \in S$ is sufficiently large, and $\tilde{\xi}_j \lceil [0, \tau] \to \tilde{\xi} \lceil [0, \tau]$ uniformly as $j \to_S \infty$. If t, t' belong to the interval $[0, \tau]$, then $\tilde{\xi}(t) = \lim_{j\to_S\infty} \xi_j(t)$, and $\tilde{\xi}(t') = \lim_{j\to_S\infty} \xi_j(t')$, so $d_{\mathcal{M}}(\tilde{\xi}(t'), \tilde{\xi}(t)) = \lim_{j\to_S\infty} d_{\mathcal{M}}(\xi_j(t'), \xi_j(t))$. Since the arc ξ_j is parametrized by arc-length, we have $d_{\mathcal{M}}(\xi_j(t'), \xi_j(t)) \le |t' - t|$, so

$$d_{\mathcal{M}}(\tilde{\xi}(t'), \tilde{\xi}(t)) \le |t' - t| \,. \tag{A.10}$$

On the other hand (assuming, say, that $t < t'$), the triangle inequality yields

$$d_{\mathcal{M}}(\xi_j(0), \xi_j(t)) + d_{\mathcal{M}}(\xi_j(t), \xi_j(t')) + d_{\mathcal{M}}(\xi_j(t'), \xi_j(L_j)) \ge d_{\mathcal{M}}(\xi_j(0), \xi_j(L_j)) \ge L_j - \frac{1}{j}\,.$$

Since $d_{\mathcal{M}}(\xi_j(0), \xi_j(t)) \le ||\xi_j \lceil [0, t]||_{G} = t$ and $d_{\mathcal{M}}(\xi_j(t'), \xi_j(L_j)) \le ||\xi_j \lceil [t', L_j]||_{G} = L_j - t'$, it follows that $d_{\mathcal{M}}(\xi_j(t), \xi_j(t')) \ge |t' - t| - \frac{1}{j}$. If we then let $j \to_S \infty$, we find that the inequality $d_{\mathcal{M}}(\tilde{\xi}(t), \tilde{\xi}(t')) \ge |t' - t|$ holds. This, together with (A.10), and the fact that τ was any number such that $0 \le \tau \le L$ and $\tau < b$, yields

$$d_{\mathcal{M}}(\tilde{\xi}(t), \tilde{\xi}(t')) = |t' - t| \quad \text{whenever} \quad t, t' \in [0, L] \cap [0, b[\,. \tag{A.11}$$

Now let ξ be the restriction of $\tilde{\xi}$ to the interval $[0, L] \cap [0, b[$. Then (A.11) implies that ξ is a local minimizer parametrized by arc-length. Let $\hat{\xi}$ be an extension of ξ such that $\hat{\xi} \in \mathrm{Max}(LMin^{PAL})$. (The existence of $\hat{\xi}$ follows from Zorn's lemma.) Then the fact that $LMin^{PAL}$ has the extension property implies that $I(\hat{\xi}) = \mathbb{R}$.

If $b \leq L$, then $\hat{\xi}(b) = \lim_{t \uparrow b} \xi(t)$, so $\lim_{j \to_S \infty} \tilde{\xi}(t) = \lim_{j \to_S \infty} \xi(t)$ exists. Then the extension property for $\mathcal{C}^{a,\leq 1}(\mathcal{M})$ implies that $b \in \mathrm{Int}(I(\tilde{\xi}))$, contradicting the fact that $b = \sup I(\tilde{\xi})$. So $b > L$. But then (A.11) says that

$$d_{\mathcal{M}}(\xi(t), \xi(t')) = |t' - t| \quad \text{whenever} \quad t, t' \in [0, L]\,. \tag{A.12}$$

Let $\hat{x} = \xi(L)$. Then $\xi_j(L) \to \hat{x}$ as $j \to_S \infty$, and

$$d_{\mathcal{M}}(\xi_j(L), x_j) = d_{\mathcal{M}}(\xi_j(L), \xi_j(L_j)) \leq |L - L_j| \to 0\,,$$

so $x_j \to \hat{x}$ as $j \to_S \infty$. Moreover, since $\xi(0) = \bar{x}$, (A.12) implies that $d_{\mathcal{M}}(\bar{x}, \hat{x}) = L$. Since $\partial\xi = (\bar{x}, \hat{x})$, and $||\xi||_G \leq L$, we see that ξ is a length minimizer joining $\bar{x}$ to $\hat{x}$. The proof is then complete. $\square$

Proof of Proposition 4.8. Let $\rho = d_{\mathcal{M}}(\bar{x}, \hat{x})$, and pick a sequence $\{\xi_j\}$ in $\mathcal{A}^{PAL,0}(\mathcal{M})$ such that $\partial\xi_j = (\bar{x}, \hat{x})$ and $\tau_+\xi_j \leq \rho + \frac{1}{j}$. Then $\rho \leq \tau_+\xi_j$. Let $\tilde{\xi}_j$ denote the restriction of ξ_j to $[0, \rho]$. Then $\tilde{\xi}_j$ is contained in $B_{\mathcal{M}}(\bar{x}, r)$.

Since the $\tilde{\xi}_j$ are parametrized by arc-length, the inequality $d_{\mathcal{M}}(\tilde{\xi}_j(t'), \tilde{\xi}_j(t)) \leq |t' - t|$ holds for all $t, t' \in [0, \rho]$. Since all the $\tilde{\xi}_j$ are contained in $B_{\mathcal{M}}(\bar{x}, r)$, which is compact, the Ascoli-Arzelà theorem implies that the sequence $\{\tilde{\xi}_j\}$ has a subsequence $\{\tilde{\xi}_{j(k)}\}$ that converges uniformly to an arc $\xi : [0, \rho] \to M$. Then $d_{\mathcal{M}}(\xi(t'), \xi_j(t)) \leq |t' - t|$ for all $t, t' \in [0, \rho]$. So Proposition 3.20 tells us that ξ is an E-admissible E-arc such that $||\dot{\xi}(t)||_G \leq 1$ for almost all t. On the other hand, $d_{\mathcal{M}}(\hat{x}, \tilde{\xi}_{j(k)}(\rho)) \leq \frac{1}{j}$, so $\tilde{\xi}_{j(k)}(\rho) \to \hat{x}$. Therefore $\xi(\rho) = \hat{x}$. Since $\rho = d_{\mathcal{M}}(\bar{x}, \hat{x})$, and $||\dot{\xi}(t)||_G \leq 1$ for almost all $t \in [0, \rho]$, we must have $||\xi||_G = \rho$, so ξ is a length minimizer. $\square$

Proof of Statement $(*)$ of §5. We work locally, in a neighborhood U of x where there is a coordinate chart $\mathbf{x}$ and an orthonormal basis $\mathbf{f} = (f_1, \ldots, f_m)$ of sections of F. Then the section V is given by $V(x', p') = \sum_{i=1}^m V^i(x', p') f_i(x')$, where the V^i are smooth functions in a neighborhood of (x, p). If W is another section of $\hat{F}$ such that $W(x, p) = v$, then W also has an expression $W(x', p') = \sum_{i=1}^m W^i(x', p') f_i(x')$, and the functions W^i satisfy $W^i(x, p) = V^i(x, p)$ for $i = 1, \ldots, m$. Then $H_{\mathcal{M}}^{c,W,p_0} - H_{\mathcal{M}}^{c,V,p_0}$ vanishes at (x, p), so

$$H_{\mathcal{M}}^{c,W,p_0}(x', p') - H_{\mathcal{M}}^{c,V,p_0}(x', p') = \sum_{i=1}^m \Big(W^i(x', p') - V^i(x', p') \Big) h_i(x', p')$$

near (x, p), where the h^i are smooth functions. Therefore

$$\vec{H}_{\mathcal{M}}^{c,W,p_0}(x, p) - \vec{H}_{\mathcal{M}}^{c,V,p_0}(x', p') = \sum_{i=1}^m \Big(W^i(x, p) - V^i(x, p) \Big) \vec{h}_i(x, p) + \sum_{i=1}^m h_i(x, p) z^i\,, \tag{A.13}$$

where $z^i = \vec{Z}^i(x, p)$, and $Z^i = W^i - V^i$.

The fact that (x, p, v, p_0) belongs to $Crit(\mathcal{M})$ implies that $h_i(x, p) = 0$. Since $W^i(x, p) - V^i(x, p) = 0$ as well, (A.13) says that

$$\vec{H}_{\mathcal{M}}^{c,W,p_0}(x, p) = \vec{H}_{\mathcal{M}}^{c,V,p_0}(x, p)\,,$$

and the proof of $(*)$ is complete. $\square$

Proof of Proposition 5.3. The fact that $(\xi(t),\psi(t),\omega_\xi(t),\psi_0)\in Crit(\mathcal{M})$ is an immediate consequence of (E.2). Fix a $\bar{t}\in I$, and let $\bar{x}=\xi(\bar{t})$. Let U be a neighborhood of $\bar{x}$ which is contained in the domain $U_{\mathbf{f}}$ of a local orthonormal basis $\mathbf{f}=(f_1,\ldots,f_m)$ of sections of F and also in the domain $\mathrm{Dom}(\mathbf{x})$ of a chart $\mathbf{x}$ near $\bar{x}$. Let $g_i=\sigma(f_i)$. Fix a $t\in I$ such that $\xi(t)\in U$. Write $\omega_\xi(t)=\sum_{i=1}^m v^i f_i(\xi(t))$, and define $V(x)=\sum_{i=1}^m v^i f_i(x)$, for $x\in U$. Then the Hamiltonian function $H_{\mathcal{M}}^{c,V,\psi_0}$ is given, near $(\xi(t),\psi(t))$, by

$$H_{\mathcal{M}}^{c,V,\psi_0}(x,p)=\sum_{i=1}^m v^i\langle p,g_i(x)\rangle-\frac{\psi_0}{2}\sum_{i=1}^m (v^i)^2\,.$$

This implies that

$$\begin{aligned}\vec{H}^c_{\mathcal{M}}(\Xi(t),\omega_\xi(t),\psi_0)&=\vec{H}_{\mathcal{M}}^{c,V,\psi_0}(\xi(t),\psi(t))\\&=\sum_{i=1}^m v^i.g_i^j(\xi(t)).\frac{\partial}{\partial x^j}-\sum_{i=1}^m v^i.\frac{\partial \mu_{g_i}}{\partial x^j}\Big(\xi(t),\psi(t)\Big).\frac{\partial}{\partial p_j}\,,\end{aligned}$$

where μ_{g_i} is the momentum function corresponding to g_i —defined by Formula (2.2)— and we have written $g_i=\sum_{i=1}^m g_i^j.\frac{\partial}{\partial x^j}$. So $\vec{H}_{\mathcal{M}}^{c,V,\psi_0}(\xi(t),\psi(t))$ is exactly equal to $\dot{\Xi}(t)$, and our proof is complete. □

Proof of Theorem 5.4. We let $\mathcal{U}$ denote the set $\Gamma(F)$ of all smooth sections of F. We define

$$w(x,u)=\sigma(u)(x)\,,\quad L(x,u)=\frac{1}{2}||u(x)||_G^2\quad\text{for}\quad u\in\mathcal{U}\,,x\in M\,.$$

We then consider the optimal control problem

$$(P)\qquad\begin{cases}\text{minimize} & \int_a^b L(\xi(t),\eta(t))\,dt\\ \text{subject to} & \dot{\xi}(t)\;=\;w(\xi(t),\eta(t))\ \text{a.e.}\\ \text{and} & \partial\xi\;=\;\partial\xi_*\ ,\end{cases}$$

where ξ ranges over the space of absolutely continuous M-valued curves on $[a,b]$, and η takes values in the class $Adm_{[a,b]}(\mathcal{U})$ of "admissible controls on $[a,b]$," that is, the set of all maps $[a,b]\ni t\to\eta(t)\in\mathcal{U}$ such that (i) the map $t\to w(x,\eta(t))$ is measurable for each $x\in M$, and (ii) for each chart $\mathbf{x}$ and each compact subset K of $\mathrm{Dom}(\mathbf{x})$ there exists an integrable function $\varphi:[a,b]\to[0,\infty]$ with the property that

$$||w^{\mathbf{x}}(x,\eta(t))||+\left\|\frac{\partial w^{\mathbf{x}}}{\partial x}(x,\eta(t))\right\|\leq\varphi(t)\quad\text{whenever}\quad x\in K,t\in[a,b]\,,$$

where $w^{\mathbf{x}}$ is the vector-valued function that expresses w in the chart $\mathbf{x}$.

Let $[a,b]\ni t\to\eta_*(t)\in\mathcal{U}$ be an admissible control such that $\eta_*(t)(\xi_*(t))=\omega_{\xi_*}(t)$ for almost all t. (The existence of η_* is easily proved: it suffices to work locally, i. e. to assume that ξ_* is contained in the domain $U_{\mathbf{f}}$ of an orthonormal basis $\mathbf{f}=(f_1,\ldots,f_m)$ of sections of F; we then write $\omega_{\xi_*}(t)=\sum_{i=1}^m v^i(t)f_i(\xi_*(t))$, where the v^i are integrable functions; we let ζ be a smooth compactly supported function on M which is equal to 1 on a neighborhood of $\xi_*([a,b])$ and whose support is contained in $U_{\mathbf{f}}$; we then define the smooth section $\eta_*(t)$ of F by letting $\eta_*(t)(x) = \sum_{i=1}^m v^i(t).\zeta(x).f_i(x)$.)

It is then clear that the pair (ξ_*, η_*) is a solution of (P). So the Pontryagin principle implies that there exist an absolutely continuous field of covectors $t \to \psi(t)$ along ξ and a real number ψ_0 such that (E.1) holds, as well as the Hamiltonian maximization condition (HMC) and the adjoint equation. The HMC says that, if we let

$$h(x,p,u,p_0) \stackrel{\text{def}}{=} <p, w(x,u)> - p_0 L(x,u)$$

for $(x,p) \in T^*M$, $p_0 \in \mathbb{R}$, $u \in \mathcal{U}$, then for almost all t the function

$$\mathcal{U} \ni u \to h(\xi_*(t), \psi(t), u, \psi_0)$$

is maximized by $u = \eta_*(t)$. Since $h(x,p,u,p_0)$ only depends, for a given (x,p,p_0), on the value $u(x)$ of u at x, the HMC for a given t is equivalent to the statement that the function $F(\xi_*(t)) \ni y \to <\psi(t), \sigma_{\xi_*(t)}(y)> - \frac{\psi_0}{2}||y||_G^2$ is maximized by $y = \eta_*(t)(\xi_*(t))$, i. e. by $y = \omega_{\xi_*}(t)$. So the HMC is exactly equivalent to (E.2). It then follows from Proposition 5.3 that $(\xi_*(t), \psi(t), \omega_{\xi_*}(t), \psi_0) \in Crit(\mathcal{M})$ for almost every $t \in [a,b]$.

Finally, the adjoint equation says, relative to a chart $\mathbf{x}$, that

$$\dot{\psi}(t) = -\frac{\partial h}{\partial x}(\xi_*(t), \psi(t), \eta_*(t), \psi_0) \quad \text{for a. e.} \quad t. \tag{A.14}$$

If we define

$$h_{u,p_0}(x,p) \stackrel{\text{def}}{=} h(x,p,u,p_0),$$

and let $\Xi_*(t) = (\xi_*(t), \psi(t))$, then A.14 implies that

$$\dot{\psi}(t) = -\frac{\partial h_{\eta_*(t),\psi_0}}{\partial x}(\Xi_*(t)) \quad \text{for a. e.} \quad t. \tag{A.15}$$

In addition, it is clear that $\frac{\partial h_{u,p_0}}{\partial p}(x,p) = w(x,u)$, so

$$\dot{\xi}_*(t) = \frac{\partial h_{\eta_*(t),\psi_0}}{\partial p}(\Xi_*(t)) \quad \text{for a. e.} \quad t. \tag{A.16}$$

It follows from (A.15) and (A.16) that

$$\dot{\Xi}_*(t) = \vec{h}_{\eta_*(t),\psi_0}(\Xi_*(t)) \quad \text{for a. e.} \quad t. \tag{A.17}$$

Now, if we write $\eta_*(t)(x) = \sum_{i=1}^m \theta^i(x,t) f_i(x)$, and define $V_t(x) = \sum_{i=1}^m \theta^i(x,t) f_i(x)$, then V_t is, for each t, a smooth section of F near $\xi_*(t)$, such that $V_t(\xi_*(t)) = \omega_{\xi_*}(t)$, and

$$H_{\mathcal{M}}^{c,V_t,\psi_0}(x,p) = h(x,p,\eta_*(t),\psi_0) = h_{\eta_*(t),\psi_0}(x,p) \quad \text{for} \quad (x,p) \text{ near } (\xi_*(t), \psi(t)). \tag{A.18}$$

Then (A.17) and (A.18) imply

$$\dot{\Xi}_*(t) = \vec{H}_{\mathcal{M}}^{c,V_t,\psi_0}(\Xi_*(t)) \quad \text{for a. e.} \quad t. \tag{A.19}$$

Since $(\xi_*(t), \psi(t), \omega_{\xi_*}(t), \psi_0) \in Crit(\mathcal{M})$ for almost every $t \in [a,b]$, and V_t is, for each t, a smooth section of F near $\xi_*(t)$ such that $V_t(\xi_*(t)) = \omega_{\xi_*}(t)$, the definition of $\vec{H}_{\mathcal{M}}^c$ implies that

$$\vec{H}_{\mathcal{M}}^{c,V_t,\psi_0}(\Xi_*(t)) = \vec{H}_{\mathcal{M}}^c(\Xi_*(t), \omega_{\xi_*}(t), \psi_0) \quad \text{for a. e.} \quad t. \tag{A.20}$$

Therefore

$$\dot{\Xi}_*(t) = \vec{H}_{\mathcal{M}}^c(\Xi_*(t), \omega_{\xi_*}(t), \psi_0) \quad \text{for a. e.} \quad t. \tag{A.21}$$

Then Proposition 5.3 implies that (E.3) holds, and ξ_* is an extremal. □

Proof of Proposition 5.11. Let $\eta^i(t) = \tilde{\varphi}^i_{\mathbf{f}}(\xi(t), \psi(t))$, $\eta(t) = (\eta^1(t), \ldots, \eta^m(t))$. We have to prove that $||\eta(t)|| = ||\dot{\xi}(t)||_G$ for a. e. t. To see this, fix a $t \in I(\xi)$ such that $\dot{\xi}(t) = \sum_{i=1}^m \eta^i(t) g_i(\xi(t))$. Then for any other representation $\dot{\xi}(t) = \sum_{i=1}^m (\eta^i(t)+\nu^i) g_i(\xi(t))$ of $\dot{\xi}(t)$ as a linear combination of the vectors $g_i(\xi(t))$, the sum $\sum_{i=1}^m \nu^i.g_i(\xi(t))$ vanishes. So $\sum_{i=1}^m \nu^i.\eta^i(t) = \sum_{i=1}^m \nu^i.\langle p, g_i(\xi(t))\rangle = \langle p, \sum_{i=1}^m \nu^i.g_i(\xi(t))\rangle = 0$. It then follows that $||\eta(t)+\nu||^2 = ||\eta(t)||^2 + ||\nu||^2 \geq ||\eta(t)||^2$. So $\eta(t) = \eta_{\mathbf{f},\xi,I(\xi)}(t)$. □

Proof of Proposition 6.3. Let $p \in T^*_x M \backslash \{0\}$ be such that $H_{\mathcal{M}}(x,p) = 0$. If v, w are any two vectors in $F(x)$ and $v = \sum_i v^i f_i(x)$, $w = \sum_i w^i f_i(x)$, then an easy calculation shows that $Z(x,p)(v,w) = \sum_{i,j} \zeta_{i,j}(x,p) v^i w^j$, where $\zeta_{i,j}(x,p) = \langle p, [g_i, g_j](x)\rangle$. Let $v \in F(x)$ be such that $v \neq 0$. Let $\theta = (v^1, \ldots, v^m)$. Then

$$Z(x,p)(v,w) = 0 \quad \text{for all} \quad w \in F(x) \tag{A.22}$$

if and only if $\sum_i \langle p, [v^i g_i, g_j](x)\rangle = 0$ for all j, i. e. iff $\langle p, [\mathbf{g}\theta, g_j](x)\rangle = 0$ for all j or, equivalently, iff p is orthogonal to all the vectors $[\mathbf{g}\theta, g_j](x)$, for $j = 1, \ldots, m$. Since p is also orthogonal to the vectors $g_j(x)$, for $j = 1, \ldots, m$, because $H_{\mathcal{M}}(x,p) = 0$, it follows that (A.22) is equivalent to

$$p \quad \text{is orthogonal to the linear span of the} \quad g_j(x) \quad \text{and the} \quad [\mathbf{g}\theta, g_j](x)\,. \tag{A.23}$$

If E is SBG at x, then a nonzero p and a nonzero θ for which (A.23) holds cannot exist, for any such pair (p, θ) would give rise to a nonzero vector $v \in F(x)$ belonging to the kernel of the bilinear form $Z(x,p)$, and this would contradict the SBG condition, since (A.23) implies that $H_{\mathcal{M}}(x,p) = 0$, so $Z(x,p)$ is nonsingular. So, if $\theta \neq 0$, then the span of the vectors $g_j(x)$ and $[\mathbf{g}\theta, g_j](x)$ must be the whole space $T_x M$, showing that (SBG$_{loc}$) holds.

Conversely, if (SBG$_{loc}$) holds, then for every nonzero θ the span of the vectors $g_j(x)$ and $[\mathbf{g}\theta, g_j](x)$ is the whole space $T_x M$. Therefore there cannot exist a nonzero p and a nonzero θ for which (A.23) holds. So, if $p \neq 0$ is such that $H_{\mathcal{M}}(x,p) = 0$, and $v \neq 0$, then (A.22) cannot hold either. Therefore $Z(x,p)$ is nonsingular, so E is SBG at x. □

REFERENCES

1. A. A. Agrachev and A. V. Sarychev, "Abnormal sub-Riemannian geodesics: Morse index and rigidity," *Ann. Inst. Poincaré Anal. Non Linéaire* **13** (1966), no. 6, pp. 635-690.

2. A. A. Agrachev, B. Bonnard, M. Chyba and I. Kupka, "Sub-Riemannian sphere in Martinet flat case," *ESAIM Control Optim. Calc. Var.* **2** (1997), pp. 377-448 (electronic).

3. C. Bär, *Carnot-Caratheodory Metriken,* Diplomarbeit, Bonn, 1988.

4. A. Bellaïche, "The tangent space in sub-Riemannian geometry," in *Sub-Riemannian geometry,* A. Bellaïche, J.-J. Risler Eds., Progr. Math. 144, Birkhäuser, Basel, 1996, pp. 4-78.

5. L. D. Berkovitz, *Optimal Control Theory*, Springer-Verlag, New York, 1974.

6. 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), pp. 11-27.
7. R. W. Brockett, "Nonlinear control theory and differential geometry," *Proc. of the Int. Congress of Mathematicians*, Vol. 2, North-Holland (1984), pp. 1357-1368.
8. R. Bryant, and L. Hsu, "Rigidity of integral curves of rank 2 distributions," *Inventiones Mathematicae* **114** (1993), pp. 435-461.
9. W. L. Chow, "Über Systeme von linearen partiellen Differentialgleichungen erster Ordnung," *Math. Ann.* **117** (1939), pp. 98–105.
10. B. Gaveau, "Principe de moindre action, propagation de la chaleur et estimées sousel-liptiques sur certains groupes nilpotents," *Acta Math.* **139** (1977), pp. 94-153.
11. Z. Ge, "Horizontal path spaces and Carnot-Carathéodory metrics," *Pacific Journal of Mathematics* **161**, No. 2 (1993), pp. 255-286.
12. V. Y. Gershkovich, "Two-sided estimates of metrics generated by absolutely nonholonomic distributions on Riemannian manifolds." (Russian) *Dokl. Akad. Nauk SSSR* **278** (1984), pp. 1040-1044.
13. V. Y. Gershkovich, "Estimates for ε-balls of nonholonomic metrics" [in *Global Analysis and Nonlinear Equations* (Russian), Voronezh Gos. Univ., Voronezh (1988), pp. 44-57], *Global Analysis–Studies and Applications IV,* Lect. Notes in Math. 1453, Springer, Berlin (1990), pp. 69-85.
14. U. Hamenstädt, "Some regularity theorems for Carnot-Carathéodory metrics," *J. of Diff. Geometry* **32** (1990), pp. 819-850.
15. P. R. Halmos, *Measure Theory*, Van Nostrand, New York, 1950.
16. R. Hermann, "On the accessibility problem in control theory," in *International Symposium on Nonlinear Differential Equations and Nonlinear Mechanics*, J.P. LaSalle and S. Lefschetz Eds., Academic Press, New York (1963), pp. 325-332.
17. A. Isidori, *Nonlinear Control Systems* (3rd ed.), Springer-Verlag, New York, 1995.
18. V. Jurdjevic, *Geometric Control Theory*, Cambridge Univ. Press, 1997.
19. J. L. Kelley, *General Topology*, Springer-Verlag, New York, 1975. (Reprint of the 1955 edition, Van Nostrand, Toronto, Ont.)
20. A. J. Krener, "A generalization of Chow's theorem and the bang-bang theorem to nonlinear control problems," *SIAM J. Control* **12** (1974), pp. 43-52.
21. I. Kupka, "Géometrie sous-riemannienne," Séminaire Bourbaki, Vol. 1995/96, *Astérisque* No. 241 (1997), Exp. No. 817,5, pp. 351-380.
22. E. B. Lee and L. Markus, *Foundations of Optimal Control Theory*, Wiley, New York, 1968.
23. W. S. Liu and H. J. Sussmann, "Abnormal sub-Riemannian minimizers," in *Differential Equations, Dynamical Systems and Control Science*, K.D. Elworthy, W.N. Everitt, and E.B. Lee Eds., Lecture Notes in Pure and Applied Mathematics, Vol. 152. M. Dekker, New York (1993), pp. 705-716.

24. W.S. Liu and H. J. Sussmann, "Shortest paths for sub-riemannian metrics on rank-2 distributions," *Memoirs of the American Math. Society*, No. 564, Vol. 118, November 1995.

25. W.S. Liu and H. J. Sussmann, "Continuous dependence with respect to the input of trajectories of control-affine systems," *SIAM Journal on Control and Optimization*, to apear.

26. R. Montgomery, "A survey of singular curves in subriemannian geometry," *J. Dynam. Control Systems* **1** (1995), no. 1, pp. 49-90.

27. R. Montgomery, "Survey of singular geodesics," in *Sub-Riemannian geometry,* A. Bellaïche, J.-J. Risler Eds., Progr. Math. 144, Birkhäuser, Basel, 1996, pp. 325-339.

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

29. R. Strichartz, "Sub-Riemannian geometry," *J. Diff. Geom.* 24, 221-263, (1983).

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

31. 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.

32. H. J. Sussmann, "Real analytic desingularization and subanalytic sets: an elementary approach," *Trans. Amer. Math. Soc.* **317** (1990), pp. 417-461.

33. H. J. Sussmann, "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.

34. H. J. Sussmann, "A cornucopia of abnormal sub-Riemannian minimizers," in *Sub-Riemannian geometry,* A. Bellaïche, J.-J. Risler Eds., Progr. Math. 144, Birkhäuser, Basel, 1996, pp. 341-364.

35. H. J. Sussmann, "An introduction to the coordinate-free maximum principle," in: *Geometry of Feedback and Optimal Control*, B. Jakubczyk and W. Respondek Eds., M. Dekker, Inc., New York, 1997, pp. 463-557.

36. T. J. Taylor, "Some aspects of differential geometry associated with hypoelliptic second order operators," *Pacific Journal of Mathematics* **136** No. 2 (1989), pp. 355-378.

37. T. J. Taylor, "Off-diagonal asymptotics of hypoelliptic diffusion equations and singular Riemannian geometry," *Pacific Journal of Mathematics* **136** No. 2 (1989), pp. 379-399.

38. A. M. Vershik and V. Ya. Gershkovich, "Non-holonomic problems and the theory of distributions," *Acta Appl. Math.* **12** (1988), pp. 189-209.

39. A. M. Vershik and V. Ya. Gershkovich, "Non-holonomic manifolds and nilpotent analysis," *J. Geom. and Phys.* **5** (1988), pp. 407-452.