This paper has appeared in the book "Optimal Control of differential equations," N.H. Pavel Ed., M. Dekker Inc. New York (1994). In this version, some misprints that appeared in the book version have been corrected.

A strong version of the Lojasiewicz Maximum Principle

H. J. SUSSMANN*, Department of Mathematics, Rutgers University, New Brunswick, New Jersey 08903.

1 INTRODUCTION

The purpose of this note is to announce and outline the proof of a version of the Maximum Principle of finite-dimensional Control Theory that requires weaker technical assumptions than previous "classical" and "nonsmooth" formulations —such as those of the book [7] by Pontryagin et al. or the "maximum Principle under minimal hypotheses" due to F.H. Clarke (cf. [2])— and at the same time yields a much stronger conclusion. (Full details of the proof will be given in [10].) The result presented here follows and generalizes a remarkable idea of S. Lojasiewicz Jr. [6], who recently proved a new version of the Maximum Principle, in the form of a necessary condition for the terminal point $q_* = \xi_*(b)$ of a reference trajectory $\xi_* : [a,b] \to \Omega$ of a system Σ to belong to the boundary of $\mathcal{R}^{\Sigma}_{[a,b]}(p_*)$ (where Ω is the state space of Σ, $p_* \stackrel{\mathrm{def}}{=} \xi_*(a)$, and $\mathcal{R}^{\Sigma}_{[a,b]}(p_*)$ is the Σ-reachable set from p_* over $[a,b]$) *without requiring the right-hand side of the system equations to be locally Lipschitz with respect to the state variable x.* In Lojasiewicz's result, only the reference vector field (i.e. the vector field $(x,t) \to f_{\eta_*}(x,t)$ giving rise to the reference trajectory) has to be locally Lipschitz with respect to the state variable x —with the usual integral bounds of Carathédory type for the Lipschitz constants— while the other vector fields f_η are only assumed to be continuous with respect to x. We use the same weak hypotheses as Lojasiewicz and, by generalizing his method of proof, are able to incorporate a number of extra features into the result. Specifically, we prove a general "separation theorem," giving a necessary condition for $\mathcal{R}^{\Sigma}_{[a,b]}(p_*)$ to be separated at q_* from some other set S. (We say that two sets S_1, S_2 are *separated at q* if $S_1 \cap S_2 = \{q\}$.) This includes the Lojasiewicz result on controllability along a reference trajectory, and also easily implies the necessary conditions for all the other properties usually considered in the various versions of the Maximum Principle (such as

* Partially supported by NSF grant DMS92-02554.

optimal control problems of the Lagrange, Meyer and Bolza forms, with various kinds of constraints on the endpoints, and with variable as well as fixed time), all of which follow by simple transformations (such as adding the cost as an extra state variable) that recasts all these as separation conditions.

Most importantly, we incorporate *high-order point variations*, thereby making our necessary conditions much stronger than those of the usual "smooth" and "nonsmooth" versions. We do this by abstractly defining a class of "variations" much larger than that of the usual needle variations of the classical Maximum Principle. High-order variations make a significant difference whenever a candidate trajectory ξ_* for a property $\mathcal{P}$ —such as noncontrollability or optimality— passes the test of the usual versions of Maximum Principle but can be proved to violate some other "high-order condition." Two simple examples of this phenomenon are given in §6. As these examples show, the situations where high-order conditions matter are by no means exceptional or pathological.

High-order conditions gave rise to a vast classical body of literature in the 1960's and 70's (cf., e.g., the book [5]). More recently, they have attracted renewed interest due to the development of methods for generating high-order variations expressed in terms of Lie brackets (cf., e.g., Sussmann [8] and Kawski [4] for high-order conditions for local controllability from a point, and Bianchini-Stefani [1] for controllability about a reference trajectory). All this work usually assumes a considerable amount of smoothness, and this is undoubtedly one of the reasons why it has by and large failed to establish contact with the parallel developments based on nonsmooth methods, whose primary concern was precisely the elimination of unnecessary requirements such as continuous differentiability. However, it is our belief that the main reason for this failure is not so much the difference in technical assumptions but the difference in *methods of proof.* Whereas most of the work on high-order conditions was based on the classical approach —needle variations and approximating cones— of [7], and sought to extend it by bringing in new classes of variations, the nonsmooth proofs —especially those based on some application of Ekeland's variational principle— follow a different route which does not lend itself so easily to the incorporation of high-order variations.

The proof presented here follows the classical approach of [7]. As explained in §5, the main ingredient of any proof based on this method is the choice of an appropriate "generalized differential" **GD**, i.e., roughly speaking, a way of defining, for certain pairs (M,p) consisting of a (possibly multivalued) map M and a point in the domain of M, the "differential of M at p." If **GD** satisfies a good chain rule and gives rise to a reasonable separation theorem, then we can consider control systems Σ and reference trajectories ξ_* generated by reference controls η_* such that (I) the flow maps $\Phi^{t,s}$ corresponding to η_* are **GD**-differentiable at every point, and (II) the variations in some class $\mathbf{V}$ are also **GD**-differentiable. For any such Σ, ξ_*, η_*, the chain rule for **GD** together with the open mapping and separation theorems imply a version of the Maximum Principle. The classical result of [7] corresponds to the ordinary differentiability theory, in which (M,p) is differentiable iff M is continuous near p and differentiable at p in the ordinary sense, and the differential of (M,p) is the ordinary differential of M at p. If the reference control η_* gives rise to a vector field $(x,t) \to f_{\eta_*}(x,t)$ which is of class C^1 with respect to x and satisfies the usual Carathéodory measurability conditions and integral bounds, then the corresponding flow maps are of class C^1, so all the pairs $(\Phi^{t,s}, x)$ are differentiable. If the vector fields f_η corresponding to the other controls are continuous in x, bounded by integrable functions of t, and have uniqueness of trajectories, then the needle variations are also differentiable at almost all times t, by standard arguments involving the Lusin and Scorza-Dragoni theorems, and these facts together imply the classical version of the Maximum Principle.

Our argument is essentially that of the classical proof, except that we use, instead of the ordinary differential, a generalized differential, that we define in §5 and call "semidifferential." (Our definition of is an extension of J. Warga's theory of "derivate containers." An alternative candidate, at least for Lipschitz maps, would be F. Clarke's generalized Jacobian, but this object does not satisfy the correct chain rule and is therefore inadequate for our purposes.)

The class of semidifferentiable pairs is much larger than that of pairs that are differentiable in the ordinary sense. For example, all (M, p) such that M is Lipschitz near p are semidifferentiable, as are the needle variations and high-order variations that occur in our control problems. (Since our needle variations are set-valued, because they are built from continuous vector fields whose trajectories need not be unique, it is crucial for our proof that we are able to differentiate set-valued maps.) Semidifferentials turn out to satisfy the chain rule as well as a separation theorem, and this suffices to imply our results.

The geometric formulation of the Maximum Principle as a separation theorem includes a transversality condition, and this requires the choice of an appropriate concept of "tangent cone" to a set at a point. We choose the type of tangent cone that fits most naturally within our general framework, namely, "generalized approximating cones" (GAC's), defined in §3. For convex sets or submanifolds, where there is a natural object that deserves to be called "tangent cone," this object is a GAC in our sense. In other cases, GAC's fare rather well compared with other cones such as the Clarke tangent cone. (It is easy to give examples of sets that have very large GAC's but very small Clarke cones, so that for such sets a transversality condition formulated in terms of GAC's is stronger than one involving a Clarke cone. Also, GAC's have the right behavior with respect to inclusion —in the sense that any GAC of a set S at a point q is also a GAC at q of any set S' containing q— while this property fails for Clarke cones. It should be pointed out, however, that there are also examples of sets with a large Clarke cone and a small GAC, so neither concept of tangent cone is better that the other one in all cases.) Moreover, it can be shown by simple examples that GAC's are more general than the approximating cones considered in [7].

Our results can be extended to incorporate state space constraints as well, but this requires a more complicated argument, so this extension will not be discussed here. Several other extensions are briefly outlined in §2. In particular, we remark that the extension to systems with jumps is completely straightforward, and yields as a byproduct a simple proof of the Maximum Principle on manifolds.

2 STATEMENT OF THE MAIN RESULTS

We use $\mathbb{R}$, $\mathbb{N}$ to denote, respectively, the real line and the set of strictly positive integers. We write $\mathbb{R}_+ = [0, +\infty)$, $\bar{\mathbb{R}} = \mathbb{R} \cup \{-\infty, +\infty\}$, $\bar{\mathbb{R}}_+ = \mathbb{R}_+ \cup \{+\infty\}$. An *interval* is an arbitrary nonempty connected subset of $\mathbb{R}$.

The abbreviation FDRLS stands for "finite-dimensional real linear space." When working with FDRLS's, we will freely endow them with norms. All the concepts defined using norms will in fact be independent of the choice of norm. If X, Y are FDRLS's, then $L(X, Y)$ is the set of all linear maps from X to Y. If X is a FDRLS, then X^* denotes the dual $L(X, \mathbb{R})$ of X. If $x \in X$, $A \in L(X, Y)$, we write Ax rather than $A(x)$. In particular, if $z \in X^*$, $x \in X$, then zx is an alternative notation for $z(x)$, the value of z at x. If $z \in Y^*$, $A \in L(X, Y)$, we write zA rather than $z \circ A$ for the composite map $x \to z(Ax)$ from X to $\mathbb{R}$, so that $zA = A^*(z)$, where $A^* : Y^* \to X^*$ is the adjoint of A. (Notice that with these conventions the identity $z(Ax) = (zA)x$ holds, so the expression zAx is well defined.)

A *set-valued map* (abbreviated SVM) is a set M all whose members are ordered pairs. If M is a SVM, we write $M(x) = \{y : (x, y) \in M\}$, $\mathrm{Dom}(M) = \{x : M(x) \neq \emptyset\}$, $\mathrm{Im}(M) = \cup_x M(x)$. We call M *single-valued at* x if $M(x)$ consists of a single point. (In that case, if $M(x) = \{y\}$, we also write $M(x) = y$.) We call M *single-valued* if it is single-valued at every $x \in \mathrm{Dom}(M)$. If M is a SVM and A is a set, the *restriction* $M\lceil A$ of M to A is the SVM $M \cap (A \times \mathrm{Im}(M))$, and we define $M(A) \stackrel{\mathrm{def}}{=} \mathrm{Im}(M\lceil A)$. If A, B are sets, a *set-valued map from* A *to* B is a SVM M such that $\mathrm{Dom}(M) \subseteq A$ and $\mathrm{Im}(M) \subseteq B$. We use $SVM(A, B)$ to denote the set of all SVM's from A to B. An *ordinary map* (or just *map*, or *mapping*) from A to B is a single-valued $M \in SVM(A, B)$. The notation $M : A \to B$ will always mean "M is an ordinary map from A to B which is everyhwhere defined, i.e. such that $\mathrm{Dom}(M) = A$." If A is any set, then id_A denotes the identity map of A, i.e. $\mathrm{id}_A = \{(x, x) : x \in A\}$.

If A is a topological space, I is an interval, L is a FDRLS, $\varphi : A \times I \to L$ is a mapping, and $(x, t) \in A \times I$, we use φ_t, φ^x to denote, respectively, the partial maps $y \to \varphi(y, t)$ and $s \to \varphi(x, s)$. An *L-valued Carathéodory function* (or *map*) *on* A *with time domain* I is a map $\varphi : A \times I \to L$ such that: (i) the function $\varphi_t : A \to L$ is continuous for every $t \in \mathbb{R}$, and (ii) for every $x \in A$, the function $\varphi^x : I \to L$ is Lebesgue measurable. (If A is separable metric, it then follows easily that φ must be jointly Borel$\times$Lebesgue-measurable.) A Carathéodory function $\varphi : A \times I \to L$ will be called *locally integrably bounded* (LIB) if for every compact subset K of A there exists a locally Lebesgue integrable function $\psi_K^B : I \to \mathbb{R}_+$ such that $||\varphi(x, t)|| \leq \psi_K^B(t)$ for all $(x, t) \in K \times I$. If A is metric, then φ is *locally integrably Lipschitz* (LIL) if it is LIB and for every compact subset K of A there exists a locally Lebesgue integrable function $\psi_K^L : I \to \bar{\mathbb{R}}_+$ such that $||\varphi(x, t) - \varphi(y, t)|| \leq \psi_K^L(t)\mathrm{dist}(x, y)$ for all $(x, y, t) \in K \times K \times I$.

If Ω is an open subset of a FDRLS X, and I is an interval, then a *control vector field* (CVF) on Ω with *time domain* I is a Carathéodory map $f : \Omega \times I \to X$. We use $\mathrm{TDom}(f)$ to denote the time domain of the CVF f, and $CVF(\Omega, I)$ to denote the set of all CVF's on Ω with time domain I. (We reserve the words "vector field," or "ordinary vector field" to mean "continuous map from Ω to X," so vector fields are vector-valued maps $x \to f(x)$ that do not depend on t, whereas CVF's are allowed to depend on t.) The concepts of LIB and LIL control vector field are therefore well defined. We call a CVF *locally integrably of class* C^1 (LIC^1) if f is LIL and each partial map f_t is of class C^1. We use $CVF_{LIB}(\Omega, I)$ (resp. $CVF_{LIL}(\Omega, I)$, $CVF_{LIC^1}(\Omega, I)$) to denote the set of all $f \in CVF(\Omega, I)$ that are LIB (resp. LIL, LIC^1).

A *curve* in Ω is a continuous map $\xi : I \to \Omega$, where I is an interval. The *graph* of a curve ξ is the subset $\mathrm{Graph}(\xi) \stackrel{\mathrm{def}}{=} \{(\xi(t), t) : t \in \mathrm{Dom}(\xi)\}$ of $\Omega \times \mathbb{R}$. An *arc* in Ω is an absolutely continuous curve in Ω whose domain is a compact interval. A *trajectory* of an $f \in CVF(\Omega, I)$ is an arc ξ such that $\mathrm{Graph}(\xi) \subseteq \Omega \times I$ and $\dot{\xi}(t) = f(\xi(t), t)$ for almost all $t \in \mathrm{Dom}(\xi)$. If $f \in CVF(\Omega, I)$, we use $\mathrm{Traj}(f)$ to denote the set of all trajectories of f.

A *control system* on a FDRLS X is a triple $\Sigma = (\Omega, \mathcal{U}, f)$, where Ω is an open subset of X, $\mathcal{U}$ is a set, and $f = \{f_\eta : \eta \in \mathcal{U}\}$ is a parametrized family of CVF's on Ω. The set Ω is the *state space* of the system, $\mathcal{U}$ is the *open-loop control space*, the elements of $\mathcal{U}$ are the *open-loop controls* or, simply, *controls*, and f is the *dynamics*.

A *controlled trajectory*, or *controlled arc* of a control system $\Sigma = (\Omega, \mathcal{U}, f)$ is a pair $\gamma = (\xi, \eta)$ such that $\eta \in \mathcal{U}$ and $\xi \in \mathrm{Traj}(f_\eta)$. A *trajectory* of Σ is an arc ξ such that (ξ, η) is a controlled arc of Σ for some $\eta \in \mathcal{U}$. We use $\mathrm{Traj}(\Sigma)$ (resp. $\mathrm{CTraj}(\Sigma)$) to denote the set of all trajectories (resp. controlled trajectories) of Σ. If $\gamma = (\xi, \eta) \in \mathrm{CTraj}(\Sigma)$ then we call ξ (resp. γ) a trajectory (resp. controlled trajectory) Σ-*generated by* η or, simply *generated by* η if the context makes it clear which system Σ is being considered. We call a $\gamma = (\xi, \eta) \in \mathrm{CTraj}(\Sigma)$ *LIB-controlled* (resp. *LIL-controlled*, *IC^1-controlled*) if the CVF f_η is LIB (resp. LIL, IC^1). The *domain* $\mathrm{Dom}(\gamma)$ of a controlled trajectory $\gamma = (\xi, \eta)$ is, by definition, $\mathrm{Dom}(\xi)$. The *starting time* (resp. *terminal time, starting point, terminal point*) of an arc ξ, or of a controlled trajectory $\gamma = (\xi, \eta)$, with domain $[a, b]$ is a (resp. b, $\xi(a)$, $\xi(b)$). If p, q are the starting and terminal points of ξ, or of $\gamma = (\xi, \eta)$, we will say that ξ (or γ) *goes from p to q*, and that η *steers p to q*.

Let $\Sigma = (\Omega, \mathcal{U}, f)$ be a control system on X. If $p, x \in \Omega$, we say that x is Σ-*reachable from p* if there exists a $\gamma = (\xi, \eta) \in \mathrm{CTraj}(\Sigma)$ that goes from p to x. If, in addition, γ can be chosen so that $\mathrm{Dom}(\gamma)$ is a given interval $[a, b]$, then we say that x is Σ-*reachable from p over* $[a, b]$. We define $\mathcal{R}^\Sigma(p)$ (resp. $\mathcal{R}^\Sigma_{[a,b]}(p)$), the Σ-*reachable set from p* (resp. the Σ-*reachable set from p over* $[a, b]$), to be the set of all points $x \in M$ that are Σ-reachable from p (resp. Σ-reachable from p over $[a, b]$). More generally, we define reachable sets from subsets of Ω, by letting $\mathcal{R}^\Sigma(S) = \cup_{p \in S} \mathcal{R}^\Sigma(p)$ and $\mathcal{R}^\Sigma_{[a,b]}(S) = \cup_{p \in S} \mathcal{R}^\Sigma_{[a,b]}(p)$.

Let $\mathcal{S}$ be a class of subsets of the real line. The system $\Sigma = (\Omega, \mathcal{U}, f)$ will be said to have the *exchange property with respect to* $\mathcal{S}$ if, whenever η_1, η_2 belong to $\mathcal{U}$, $E \in \mathcal{S}$, $E \subseteq \mathrm{TDom}(f_{\eta_1}) \cap \mathrm{TDom}(f_{\eta_2})$, it follows that there exist an $\eta_3 \in \mathcal{U}$ such that (i) $\mathrm{TDom}(f_{\eta_1}) \subseteq \mathrm{TDom}(f_{\eta_3})$ and (ii) the identity

$$f_{\eta_3}(x, t) = \chi_E(t) f_{\eta_2}(x, t) + \Big(1 - \chi_E(t)\Big) f_{\eta_1}(x, t) \tag{2.1}$$

holds for $(x, t) \in \Omega \times \mathrm{TDom}(f_{\eta_1})$. (Here χ_E is, for a set E, the indicator function of E.) If Σ has the exchange property with respect to the class of all compact subsets (resp.

compact subintervals) of $\mathbb{R}$, then Σ will be said to have the *compact* (resp. *interval*) *exchange property.*

The following result is the Lojasiewicz Maximum Principle. In the statement, the standard notation $\partial F(x)$ is used for the Clarke generalized Jacobian of a locally Lipschitz map F at a point x. According to our general conventions, $(f_{\eta_*})_t$ is the map $x \to f_{\eta_*}(x,t)$, which is locally Lipschitz for a.e. t if $f_{\eta_*} \in CVF_{LIL}(\Omega, I)$, so $\partial(f_{\eta_*})_t(x)$ is well defined for all $x \in \Omega$, for a.e. $t \in I$. The basic assumptions are

(A1) $\Sigma = (\Omega, \mathcal{U}, f)$ is a control system on a FDRLS X, such that either: (i) Σ has the compact exchange property, or (ii) Σ has the interval exchange property and every f_η, $\eta \in \mathcal{U}$ is locally integrably bounded;

(A2) $\xi_* : [a,b] \to \Omega$ is a trajectory of Σ, $p_* = \xi_*(a)$, and $q_* = \xi_*(b)$.

Theorem 2.1. (The Lojasiewicz Maximum Principle.) *Assume that (A1) and (A2) hold. Then, if q_* is not an interior point of the reachable set $\mathcal{R}^{\Sigma}_{[a,b]}(p_*)$, it follows that for every $\eta_* \in \mathcal{U}$ such that $\gamma_* = (\xi_*, \eta_*)$ is a LIL-controlled trajectory of Σ there exists a pair (A, ζ) such that A is a measurable function from $[a,b]$ to $L(X,X)$ that satisfies*

$$A(t) \in \partial(f_{\eta_*})_t(\xi_*(t)) \text{ for almost all } t \in [a,b]\,, \tag{2.2}$$

and ζ is an absolutely continuous map from $[a,b]$ to X^* that satisfies the three conditions

(AE) $\dot{\zeta}(t) = -\zeta(t)A(t)$ *for a.e.* $t \in [a,b]$,

(NT) $\zeta(t) \neq 0$ *for all* $t \in [a,b]$,

(WHM) $\Big(\forall \eta \in \mathcal{U}\Big)\Big(\zeta(t) f_{\eta_*}(\xi_*(t),t) \leq \zeta(t) f_\eta(\xi_*(t),t)$ *for* a.e. $t \in [a,b]\Big)$.

We will refer to (AE) as the *adjoint equation* corresponding to $A(\cdot)$, and to (NT) and (WHM) as the *nontriviality* and *weak Hamiltonian minimization* conditions. The word "weak" refers to the fact that the exceptional set of measure zero where the inequality is allowed to fail is allowed to depend on η. Naturally, it is much more desirable to have instead the stronger condition

$$\text{(SHM)}\quad \zeta(t) f_{\eta_*}(\xi_*(t),t) = \min\Big\{\zeta(t) f_\eta(\xi_*(t),t)\ :\ \eta \in \mathcal{U}\Big\} \text{ for a.e. } t \in [a,b],$$

which is the version of (WHM) where the exceptional set can be taken not to depend on η. The question whether one can substitute (SHM) for (WHM) is a separate issue, related to the "separability" of Σ, that will be taken up in [10].

We note the following features of Theorem 2.1:

(L1) Only the "reference CVF" f_{η_*} is assumed to be LIL; the other CVF's f_η are only required to be continuous in x and measurable in t.

(L2) A pair (A, ζ) that satisfies all the conclusions has to exist for every choice of $\eta_* \in \mathcal{U}$ such that ξ_* is a trajectory of Σ generated by η_*. This means, in particular, as pointed out by Kaskosz and Lojasiewicz in [3], that one can often get stronger information by adding new CVF's f_η parametrized by η in some set $\mathcal{U}_{new}$, in such a way that no new trajectories are introduced, and then applying the necessary condition to some η_* in the new class.

We now state the simplest version of our Strong Maximum Principle, and do it directly in the form of a separation theorem. Recall from §1 that two sets S_1, S_2 are *separated at* q if $S_1 \cap S_2 = \{q\}$.

Our statement involves two new kinds of objects, namely, "generalized approximating cones" and "variational vectors," and one new property, namely, "compatibility" of a set of variational vectors. These three concepts are defined in §3 below. In addition to (A1) and (A2), we now assume

- (A3) Y is a FDRLS, U is an open subset of X, $g : U \to Y$ is a Lipschitz map, $q_* \in U$, S_0, S are subsets of X, Y such that $p_* \in S_0$ and $g(q_*) \in S$;
- (A4) S_0 has a generalized approximating cone C_0 at p_*, and S has a generalized approximating cone C at $g(q_*)$;
- (A5) either (i) C is not a linear subspace of Y or (ii) $C = \{0\}$ and $g(q_*)$ belongs to $\mathrm{Clos}(S - \{g(q_*)\})$.

Theorem 2.2. (The Strong Maximum Principle.) *Assume that (A1), (A2), (A3), (A4), (A5) hold. Let $\eta_* \in \mathcal{U}$ be such that $\gamma_* = (\xi_*, \eta_*)$ is a LIL-controlled trajectory of Σ. Let $\mathbf{v}$ be a compatible set of variational vectors for γ_*. Then, if $g(\mathcal{R}^{\Sigma}_{[a,b]}(S_0))$ is separated from S at $g(q_*)$, it follows that there exist: (I) a measurable function $A : [a,b] \to L(X,X)$ that satisfies (2.2), (II) a $B \in L(X,Y)$ such that $B \in \partial g(q_*)$, (III) a $w \in Y^* - \{0\}$, and (IV) an absolutely continuous map $\zeta : [a,b] \to X^*$, such that (AE) and (WHM) hold, and in addition (i) $\zeta(b) = wB$, (ii) $wv \le 0$ for all $v \in C$, (iii) $\zeta(a)v \ge 0$ for all $v \in C_0$, and (iv) $\zeta(t)v \ge 0$ for all $(v,t) \in \mathbf{v}$.*

Theorem 2.2 implies Theorem 2.1 as can be seen by taking $Y = X$, $g = \mathrm{id}_X$, $C = \{0\}$, $S_0 = \{p_*\}$, and S=any subset of X such that $S \cap \mathcal{R}^{\Sigma}_{[a,b]}(p_*) = \{q_*\}$, and $q_* \in \mathrm{Clos}(S \backslash \{q_*\})$.

Theorem 2.2 can be strengthened in a number of ways. For example:

1. The set $\partial g(q_*)$ can be replaced by an arbitrary derivate container of g at q_*. Actually, g can be assumed to be set-valued and semidifferentiable at q_*, provided that $\mathrm{Dom}(g)$ is a neighborhood of q_*.
2. We can consider *systems with jumps*. For instance, one could consider a finite sequence $(\Sigma_0, \ldots, \Sigma_m)$ of systems $\Sigma_i = (\Omega_i, \mathcal{U}_i, f^i)$ on FDRLS's X_i, together with jump times $t_1 < t_2 < \ldots < t_m$ in $[a,b]$ and jump maps $J_i \in SVM(\Omega_{i-1}, \Omega_i)$. A control would now be an $m+1$-tuple $\eta = (\eta_0, \ldots, \eta_m) \in \mathcal{U}_0 \times \ldots \mathcal{U}_m$, and a trajectory for such an η would be an $m+1$-tuple $\xi = (\xi_0, \ldots, \xi_m)$ such that $\xi_i : [t_{i-1}, t_i] \to \Omega_i$ (where we let $t_0 = a$, $t_{m+1} = b$), $(\xi_i, \eta_i) \in \mathrm{CTraj}(\Sigma_i)$, and $\xi_i(t_i) = J_i(\xi_{i-1}(t_i))$. The maps J_i should be assumed to be semidifferentiable at the jump points $x_i = \xi^*_{i-1}(t_i)$ of the reference trajectory ξ^*, with semidifferentials $\mathbf{D}_i$, and their domains should be neighborhoods of the x_i. (In particular, the J_i could be Lipschitz, in which case we can take $\mathbf{D}_i = \partial J_i(x_i)$.) The definition of the reachable set is obvious, and the conclusion is the same, except that now we get finite sequences $A = (A_0, \ldots, A_m)$, $\zeta = (\zeta_0, \ldots, \zeta_m)$ such that $A_i : [t_{i-1}, t_i] \to L(X_i, X_i)$, $\zeta_i : [t_{i-1}, t_i] \to X^*_i$ with the obvious properties, plus the jump condition $(\exists D_i \in \mathbf{D}_i)(\zeta_{i-1}(t_i) = \zeta_i D_i)$. It is clear that the proof sketched in §4 and §5 applies with no change to this situation.

3. The *Maximum Principle on manifolds* follows easily from the result on linear spaces for systems with jumps. (It suffices to partition the domain $[a, b]$ of the reference trajectory ξ_* into intervals $[t_{i-1}, t_i]$ such that there are coordinate patches Ω_i for which $\xi_*([t_{i-1}, t_i]) \subseteq \Omega_i$. The J_i can then be taken to be the coordinate changes.)
4. The systems with fixed jumps at fixed times considered above can be replaced by systems with optional jumps at optional times, in which case one obtains extra conditions on ζ.
5. State space constraints can be incorporated as well.

3 GENERALIZED APPROXIMATING CONES AND VARIATIONS

If X and Y are FDRLS's, a set-valued map $F \in SVM(X, Y)$ is said to be *compactly valued* if $F(x)$ is compact for every $x \in \mathrm{Dom}(F)$, and *upper semicontinuous* (USC) if, whenever $x \in X$, $U \subseteq Y$ is open, and $F(x) \subseteq U$, there exists a neighborhood V of x such that $F(y) \subseteq U$ for all $y \in V$. (In particular, if F is USC then $\mathrm{Dom}(F)$ must be closed.) We will call F *compact* if Graph(F) is compact, i.e. if F has compact domain and is compactly-valued and USC. If F is compact, then a *single-valued approximating sequence* for F is a sequence $\{F_j\}_{j=1}^{\infty}$ that satisfies the following three conditions:

(SV1) each F_j is a continuous Y-valued map with domain $\mathrm{Dom}(F)$,
(SV2) the maps F_j, $j = 1, 2, \ldots$ are uniformly bounded,
(SV3) if $\{x_\ell\}_{\ell=1}^{\infty}$ is any sequence of points in $\mathrm{Dom}(F)$, $j(\ell)$ are integers going to ∞ as $\ell \to \infty$, and the limit $(x, y) = \lim_{\ell\to\infty}\big(x_{j(\ell)}, F_{j(\ell)}(x_{j(\ell)})\big)$ exists, then $y \in F(x)$.

We will say that F *has single-valued approximations* if there exists a single-valued approximating sequence for F.

Definition 3.1. *(Regular maps)* If X and Y are FDRLS's, a set-valued map $F \in SVM(X, Y)$ is *regular* if it is compact and has single-valued approximations. ∎

The next definition involves the condition "$F_i(x) = q + x + o(\|x\|)$ as $x \to 0$ through values in N_i," which should be interpreted at stating that

$$\lim_{x\to 0\,,\, x\in N_i\,,\, x\neq 0}\left(\sup\left\{\frac{\|y - x - q\|}{\|x\|} : y \in F_i(x)\right\}\right) = 0\ . \tag{3.1}$$

Definition 3.2. *(Generalized approximating cones)* Let X be a FDRLS, and let $S \subseteq X$, $q \in S$. A *generalized approximating cone* (GAC) for S at q is a closed convex cone C in X which is the closure of the union of an increasing sequence $C_1 \subseteq C_2 \subseteq C_3 \subseteq \ldots$ of closed convex cones such that for each i there exists a regular map $F_i \in SVM(N_i, X)$, defined on a relative neighborhood N_i of 0 in C_i, such that $F_i(N_i) \subseteq S$, $F_i(0) = q$, and $F_i(x) = q + x + o(\|x\|)$ as $x \to 0$ through values in N_i. ∎

If $k \in \mathbb{N}$ and $r > 0$, we use $\sigma_k(r)$ to denote the closed simplex $\{(\alpha_1, \ldots, \alpha_k) \in \mathbb{R}^k : \alpha_1 \geq 0, \ldots, \alpha_k \geq 0, \alpha_1 + \ldots + \alpha_k \leq r\}$. If X is a FDRLS, a *k-parameter variation* in X is a pair $\mathcal{V} = (x_*, V)$ such that $x_* \in X$ and $V \in SVM(X \times \mathbb{R}^k_+, X)$ with the property that there exist $r > 0$ and a neighborhood K of x_* for which (i) $K \times \sigma_k(r) \subseteq \mathrm{Dom}(V)$, (ii) the restriction $V \lceil K \times \sigma_k(r)$ is regular, and (iii) $V(x, 0) = x$ for $x \in K$. The point x_* is the *base point* of the variation $\mathcal{V}$. If in addition there exists a linear map $L : X \times \mathbb{R}^k$ such that

$$\lim_{x \to x_*, \alpha \to 0, \alpha \in \mathbb{R}^k_+} \frac{\sup\{||y - x_* - L(x - x_*, \alpha)|| : y \in V(x, \alpha)\}}{||x - x_*|| + ||\alpha||} = 0\,, \tag{3.2}$$

then we call $\mathcal{V}$ *differentiable*, and the map L is the *differential* of $\mathcal{V}$. Clearly, L has to be of the form $L(z, \alpha) = z + \alpha_1 v_1 + \ldots + \alpha_k v_k$. We write $\mathcal{D}\mathcal{V} \stackrel{\mathrm{def}}{=} (v_1, \ldots, v_k)$.

If $\Sigma = (\Omega, \mathcal{U}, f)$ is a control system on X, $\gamma_* = (\xi_*, \eta_*) \in \mathrm{CTraj}(\Sigma)$, $\mathrm{Dom}(\gamma_*) = [a, b]$, and $t_* \in [a, b]$, then a *Σ-admissible k-parameter variation of γ_* at time t_** is a k-parameter variation $\mathcal{V} = (x_*, V)$ in X with base point $x_* = \xi_*(t_*)$ such that

(ADM) there exist functions $\tau_{,}\tau, \tau+, r : (0, \infty) \to (0, \infty)$ such that $a \leq \tau_-(\varepsilon) \leq \tau(\varepsilon) \leq \tau_+(\varepsilon) \leq b$ for every $\varepsilon > 0$, $\lim_{\varepsilon \to 0+} \tau_-(\varepsilon) = \lim_{\varepsilon \to 0+} \tau_+(\varepsilon) = t^*$, $\lim_{\varepsilon \to 0+} r(\varepsilon) = 0$, such that for each $\varepsilon > 0$ the exists a neighborhood $K(\varepsilon) \subseteq \Omega$ of $\xi_*(t_*)$ with the property that $K(\varepsilon) \times \sigma_k(r(\varepsilon)) \subseteq \mathrm{Dom}(V)$ and

(I) if $\xi_1 : [\tau_-(\varepsilon), \tau(\varepsilon)] \to \Omega$ and $\xi_2 : [\tau(\varepsilon), \tau_+(\varepsilon)] \to \Omega$ are trajectories of η_* such that $\xi_1(\tau(\varepsilon)) \in K(\varepsilon)$ and $\xi_2(\tau(ep)) \in V(\xi_1(\tau(\varepsilon)), \alpha)$ for some $\alpha \in \sigma_k(r(\varepsilon))$, if follows that $\xi_2(b) \in \mathcal{R}^{\Sigma}_{[\tau_-(\varepsilon), \tau_+(\varepsilon)]}(\xi_1(a))$.

Definition 3.3. A *variational vector* (VV) for a controlled trajectory γ_* of a control system Σ is a pair (v, t_*) such that there exists a Σ-admissible differentiable 1-parameter variation $\mathcal{V}$ of γ_* at time t_* such that $\mathcal{D}\mathcal{V} = v$. A set $\mathbf{v}$ of VV's for γ_* is *compatible* if, whenever $\mathbf{w}$ is a finite subset of $\mathbf{v}$ all whose elements have the same variation time t_*, then there exist a $k \in \mathbb{N}$ and a Σ-admissible differentiable k-parameter variation $\mathcal{V}$ of γ_* at t_* such that, if $\mathcal{D}\mathcal{V} = (v_1, \ldots, v_k)$, then $\{v : (v, t_*) \in \mathbf{w}\} \subseteq \{v_1, \ldots, v_k\}$. ∎

4 SKETCH OF THE PROOF: REDUCTION TO THE FINITE CASE

The simplest examples of VV's are the "needle variational vectors," defined as follows. Let η, η_* be open-loop controls such that f_η and f_{η_*} are LIB. Let $t_* \in \mathbb{R}$ be such that $[t_*, t_* + \varepsilon] \subseteq \mathrm{TDom}(f_\eta) \cap \mathrm{TDom}(f_{\eta_*})$ for some $\varepsilon > 0$. For $x \in \Omega$, $\alpha \geq 0$, define $V_{\eta, \eta_*, t_*}(x, \alpha)$ to be the set of all $y \in \Omega$ such that there exist $\delta \in \mathrm{Traj}(f_\eta)$, $\delta_* \in \mathrm{Traj}(f_{\eta_*})$ that satisfy (i) $\mathrm{Dom}(\delta) = \mathrm{Dom}(\delta_*) = [t_*, t_* + \alpha]$, (ii) $\delta(t_*) = x$, (iii) $\delta_*(t_*) = y$, and (iv) $\delta_*(t_* + \alpha) = \delta(t_* + \alpha)$. Then given any $x_* \in \Omega$ the pair $(x_*, V_{\eta, \eta_*, t_*})$ is a 1-parameter variation at x_*. (The compactness of the sets $V_{\eta, \eta_*, t_*}(x, \alpha)$ for (x, α) close to $(x_*, 0)$ and the upper semicontinuity of the SVM V_{η, η_*, t_*} near $(x_*, 0)$ follow from standard properties of solutions of ordinary differential equations. The existence of single-valued approximations follows by regularizing f_η and f_{η_*}.) Moreover, using standard techniques such as the Scorza-Dragoni theorem, one can prove that, if an interval $[a,b]$ is such that $[a,b] \subseteq \mathrm{TDom}(f_\eta) \cap \mathrm{TDom}(f_{\eta_*})$, then there exists a subset $E(\eta,\eta_*)$ of $[a,b]$ such that $[a,b]\backslash E(\eta,\eta_*)$ has measure zero, with the property that $(x_*, V_{\eta,\eta_*,t_*})$ is differentiable for every $t_* \in E(\eta,\eta_*)$ and every $x_* \in \Omega$, and the differential $\mathcal{D}(x_*, V_{\eta,\eta_*,t_*})$ is the vector $f_\eta(x_*,t_*) - f_{\eta_*}(x_*,t_*)$.

Now assume that $\gamma_* = (\xi_*,\eta_*)$ is a LIL-controlled trajectory of Σ with domain $[a,b]$, and η is such that $[a,b] \subseteq \mathrm{TDom}(\eta)$ and $f_\eta \in CVF_{LIB}(\Omega, I)$. If we pick $t_* \in E(\eta,\eta_*)$ and let $x_* = \xi_*(t_*)$, then it is easy to show that $(x_*, V_{\eta,\eta_*,t_*})$ is a Σ-admissible differentiable 1-parameter variation of γ_* at time t_*. (It suffices to pick $\tau_-(\varepsilon) = t_*$, $\tau_+(\varepsilon) = t_* + \varepsilon$.) Since $\mathcal{D}(x_*, V_{\eta,\eta_*,t_*}) = f_\eta(\xi_*(t_*),t_*) - f_{\eta_*}(\xi_*(t_*),t_*)$, we see that

$$\mathbf{v}_{\eta,\gamma_*} = \{(f_\eta(\xi_*(t_*),t_*) - f_{\eta_*}(\xi_*(t_*),t_*),t_*) : t_* \in E(\eta,\eta_*)\}$$

is a set of VV's for γ_*, which is obviously compatible since no two different members of $\mathbf{v}_{\eta,\gamma_*}$ can correspond to the same time t_*.

Using the above observations, it is easy to see that Theorem 2.2 will follow if we prove

Proposition 4.1. *Assume that (A1), (A2), (A3), (A4), (A5) hold. Let $\eta_* \in \mathcal{U}$ be such that $\gamma_* = (\xi_*,\eta_*)$ is a LIL-controlled trajectory of Σ. Let $\mathbf{v}$ be a finite compatible set of variational vectors for γ_*. Then, if $g(\mathcal{R}^\Sigma_{[a,b]}(S_0))$ is separated from S at $g(q_*)$, it follows that there exist: (I) a measurable function $A : [a,b] \to L(X,X)$ that satisfies (2.2), (II) a $B \in L(X,Y)$ such that $B \in \partial g(q_*)$, (III) a $w \in Y^*\backslash\{0\}$, and (IV) an absolutely continuous map $\zeta : [a,b] \to X^*$, such that (AE) holds, and in addition (i) $\zeta(b) = wB$, (ii) $wv \le 0$ for all $v \in C$, (iii) $\zeta(a)v \ge 0$ for all $v \in C_0$, and (iv) $\zeta(t)v \ge 0$ for all $(v,t) \in \mathbf{v}$.*

The main difference between Proposition 4.1 and Theorem 2.2 is that in Proposition 4.1 the Hamiltonian minimization condition does not occur, and the set $\mathbf{v}$ of VV's is finite.

To prove that Proposition 4.1 implies Theorem 2.2, we consider first the case when all the f_η, $\eta \in \mathcal{U}$ are LIB. Fix a —not necessarily finite— compatible set $\mathbf{v}$ of VV's for γ_*. Let Q be the set of all triples (A,B,w) such that

- (Q1) A is a measurable function from $[a,b]$ to $L(X,X)$ that satisfies (2.2),
- (Q2) $B \in \partial g(q_*)$ (so in particular $B \in L(X,Y)$),
- (Q3) $w \in Y^*$, $||w|| = 1$, and $wv \le 0$ for all $v \in C$,
- (Q4) if $\zeta_{A,B,w}$ is the unique absolutely continuous map $\zeta : [a,b] \to X^*$ such that (AE) holds and $\zeta(b) = wB$, then $\zeta_{A,B,w}(a)v \ge 0$ for all $v \in C_0$.

Then it is clear that Q is compact, if the set $\mathcal{A}$ of measurable selections A of $t \to \partial(f_{\eta_*})_t(\xi_*(t))$ is given the weak topology of $L^1([a,b], L(X,X))$. Moreover, $\zeta_{A,B,w}$ (regarded as a member of $L^\infty([a,b],X^*)$) depends continuously on (A,B,w). For each subset $\mathcal{S}$ of $\mathcal{U}$ and each subset $\mathbf{w}$ of $\mathbf{v}$, let $Q(\mathcal{S},\mathbf{w})$ be the set of those $(A,B,w) \in Q$ for which (i) the inequality

$$\zeta_{A,B,w}(t) f_{\eta_*}(\xi_*(t),t) \le \zeta_{A,B,w}(t) f_\eta(\xi_*(t),t) \ \text{ for a.e. } t \in [a,b]\,, \tag{4.1}$$

holds for all $\eta \in \mathcal{S}$, and (ii) $\zeta_{A,B,w}(t)v \geq 0$ for all $(v,t) \in \mathbf{w}$. Our goal is to prove that $Q(\mathcal{U},\mathbf{v}) \neq \emptyset$. Suppose we show that $Q(\mathcal{S},\mathbf{w}) \neq \emptyset$ whenever $\mathcal{S}$, $\mathbf{w}$ are finite subsets of $\mathcal{U}$, $\mathbf{v}$. Then the conclusion that $Q(\mathcal{U},\mathbf{v}) \neq \emptyset$ follows by an elementary compactness argument.

Now let $\mathcal{S} \subseteq \mathcal{U}$, $\mathbf{w} \subseteq \mathbf{v}$, and assume that $\mathcal{S}$, $\mathbf{w}$ are finite. Let $\mathcal{S} = \{\eta_1,\ldots,\eta_m\}$. Set $\eta_0 = \eta_*$. Let $\mathbf{w} = \{(w_1,t_1),\ldots,(w_\ell,t_\ell)\}$. Let $k \in \mathbb{N}$. Using Lusin's theorem, find a compact subset F_k of $\cap_{i=1}^m E(\eta_i,\eta_*)$ such that $\mathrm{meas}([a,b]\backslash F_k) < 2^{-k}$, with the property that the functions $t \to f_{\eta_i}(\xi_*(t),t)$ are continuous on F_k for $i = 0,1,\ldots,m$. If $N \in \mathbb{N}$, let $\mathcal{I}_N$ be the set of intervals $[a+\frac{j-1}{N}(b-a), a+\frac{j}{N}(b-a)]$, for $j = 1,\ldots,N$. Let $\mathcal{I}_N^+$ be the set of those $J \in \mathcal{I}_N$ such that $\mathrm{meas}(F_k \cap J) > 0$. Let $F_{k,N} = \cup_{J\in\mathcal{I}_N^+}(F_k \cap J)$. Then $F_{k,N} \subseteq F_k$ and $\mathrm{meas}(F_k \backslash F_{k,N}) = 0$. For fixed k and N, and $J \in \mathcal{I}_N^+$, find m different points $s^1_{k,N,J},\ldots,s^m_{k,N,J}$ in $F_k \cap J$, making sure that none of the $s^i_{k,N,J}$ is an endpoint of J and that $s^i_{k,N,J} \neq t_j$ for all J, i, j. Let $\mathbf{z}(k,N)$ be the union of $\mathbf{w}$ and the set of all pairs $W_{k,N,J,i} = (f_{\eta_i}(\xi_*(s^i_{k,N,J}),s^i_{k,N,J}) - f_{\eta_*}(\xi_*(s^i_{k,N,J}),s^i_{k,N,J}), s^i_{k,N,J})$, for all $J \in \mathcal{I}_N^+$ and all i. Then $\mathbf{z}(k,N)$ is a compatible set of VV's for γ_*. (The compatibility follows from the fact that the only way a set V of more that one VV in $\mathbf{z}(k,N)$ can correspond to the same time is if $V \subseteq \mathbf{w}$.) So Proposition 4.1 implies that there exists a triple $(A^{k,N},B^{k,N},w^{k,N}) \in Q(\emptyset,\mathbf{z}(k,N))$. Let $\tilde{F}_k = \cap_N F_{k,N}$, so $\mathrm{meas}(F_k\backslash\tilde{F}_k) = 0$. Find a sequence $\{N(j)\}$ going to ∞ such that $\{(A^{k,N(j)},B^{k,N(j)},w^{k,N(j)})\}_{j=1}^\infty$ converges to a limit (A^k,B^k,w^k). Then $\zeta_{A^{k,N(j)},B^{k,N(j)},w^{k,N(j)}} \to \zeta_{A^k,B^k,w^k}$ uniformly. If $t \in \tilde{F}_k$, and $i \in \{1,\ldots,m\}$, then $t = \lim s^i_{k,N(j),J(j)}$ for some sequence $\{J(j)\}$. Since $W_{k,N(j),J(j),i} \in \mathbf{z}(k,N(j))$, we have

$$\zeta^{k,j}(s^{k,j,i})\Big(f_{\eta_i}(\xi_*(s^{k,j,i}),s^{k,j,i}) - f_{\eta_*}(\xi_*(s^{k,j,i}),s^{k,j,i})\Big) \geq 0\,, \tag{4.2}$$

where we have written $\zeta^{k,j} \stackrel{\mathrm{def}}{=} \zeta_{A^{k,N(j)},B^{k,N(j)},w^{k,N(j)}}$, $s^{k,j,i} = s^i_{k,N(j),J(j)}$. Letting $j \to \infty$, and using the fact that the functions $t \to f_{\eta_i}(\xi_*(t),t)$ and $t \to f_{\eta_*}(\xi_*(t),t)$ are continuous on F_k, we get

$$\zeta_{A^k,B^k,w^k}(t)\Big(f_{\eta_i}(\xi_*(t),t) - f_{\eta_*}(\xi_*(t),t)\Big) \geq 0 \ \text{ for } \ t \in \tilde{F}_k\,. \tag{4.3}$$

Next let $\{k(j)\}$ be a sequence going to ∞ such that $\{(A^{k(j)},B^{k(j)},w^{k(j)})\}_{j=1}^\infty$ converges to a limit (A,B,w). Let $F = \cup_K \cap_{k\geq K} \tilde{F}_k$. Then $\mathrm{meas}([a,b]\backslash\tilde{F}) = 0$. If $t \in \tilde{F}$, then $t \in \tilde{F}_{k(j)}$ for sufficiently large j, so $\zeta_{A^{k(j)},B^{k(j)},w^{k(j)}}(t)(f_{\eta_i}(\xi_*(t),t) - f_{\eta_*}(\xi_*(t),t)) \geq 0$ for large j, and then $\zeta_{A,B,w}(t)\Big(f_{\eta_i}(\xi_*(t),t) - f_{\eta_*}(\xi_*(t),t)\Big) \geq 0$. This shows that $(A,B,w) \in Q(\mathcal{S},\mathbf{w})$, so $Q(\mathcal{S},\mathbf{w}) \neq \emptyset$.

This completes the proof that Proposition 4.1 implies the conclusion of Theorem 2.2 for the case of a system Σ all whose CVF's are LIB. The other conclusion can then be proved by applying the compact exchange property, as we now show. Fix a relatively compact open subset $\hat{\Omega}$ of Ω such that ξ_* is entirely contained in $\hat{\Omega}$. Let $\hat{\Sigma} = (\hat{\Omega},\mathcal{U},\hat{f})$, where $\hat{f} = \{\hat{f}_\eta\}_{\eta\in\mathcal{U}}$ and $\hat{f}_\eta$ is the restriction of f_η to $\hat{\Omega} \times \mathrm{TDom}(f_\eta)$. Then let $\tilde{\Sigma} = (\hat{\Omega},\tilde{\mathcal{U}},\tilde{f})$, where $\tilde{\mathcal{U}} = \{\eta \in \mathcal{U} : f_\eta \in CVF_{LIB}(\Omega,I)\}$ and $\tilde{f} = \{\hat{f}_\eta\}_{\eta\in\tilde{\mathcal{U}}}$. If we apply the result for LIB systems to the system $\tilde{\Sigma}$, we get the desired conclusion, except that

(WHM) is only known to hold for $\eta \in \tilde{\mathcal{U}}$. If $\eta \in \mathcal{U}$, let $\psi(t) = \sup\{||f_\eta(x,t)|| : x \in K\}$, where $K = \mathrm{Clos}(\hat{\Omega})$. Then ψ is measurable and finite. Let $L_N = \{t \in [a,b] : \psi(t) > N\}$. Fix $\varepsilon > 0$. Find N such that $\mathrm{meas}(L_N) < \frac{\varepsilon}{2}$. Find a compact subset F of $[a,b]\backslash L_N$ such that $\mathrm{meas}([a,b]\backslash F) < \varepsilon$. Using the compact exchange property, find a control $\eta_\#$ such that $f_{\eta_\#}$ agrees with f_η on $\Omega \times F$ and with f_{η_*} on $\Omega \times ([a,b]\backslash F)$. Then $\hat{f}_{\eta_\#} \in CVF_{LIB}(\hat{\Omega}, I)$, and therefore $\zeta(t)\Big(f_{\eta_\#}(\xi_*(t),t) - f_{\eta_*}(\xi_*(t),t)\Big) \geq 0$ for almost all $t \in [a,b]$. But then $\zeta(t)\Big(f_\eta(\xi_*(t),t) - f_{\eta_*}(\xi_*(t),t)\Big) \geq 0$ for a. e. $t \in F$, i.e. for all t in a set whose complement has measure $< \varepsilon$. Since ε is arbitrary, the conclusion follows.

5 SKETCH OF THE PROOF: SEMIDIFFERENTIABLE MAPS

To prove Proposition 4.1, it suffices to assume that there are regular maps $F_0 \in SVM(N_0, S_0)$, $F \in SVM(N, S)$, where N_0, N are relative neighborhoods of 0 in C_0, C, such that $F_0(0) = p_*$, $F(0) = g(q_*)$, and $F_0(x) = p_* + x + o(\|x\|)$, $F(y) = g(q_*) + y + o(\|y\|)$ as $x, y \to 0$ through values in N_0, N, respectively. Let $t_1 < t_2 < \ldots < t_m$ be the times of the VV's in $\mathbf{v}$. For each i, using the compatibility assumption, find a Σ-admissible differentiable variation $\mathcal{V}_i = (\xi_*(t_i), V_i)$ of γ_*, depending on k_i parameters, with the property that all the v's such that $(v, t_i) \in \mathbf{v}$ occur in $\mathcal{DV}_i$. Let Φ be the flow of the vector field f_{η_*}, so $t \to \Phi^{t,s}(x)$ is the maximal trajectory of f_{η_*} that goes through x at time s.

Let us define a set-valued map $M \in SVM(N_0 \times \mathbb{R}^{k_1} \times \ldots \times \mathbb{R}^{k_m}, Y)$ by letting $y \in M(z, \alpha^1, \ldots, \alpha^m)$ if there exist points $x_0, x_1, \hat{x}_1, x_2, \hat{x}_2, \ldots, x_m, \hat{x}_m, x_{m+1}$ such that $x_0 \in F_0(z)$, $x_1 = \Phi^{t_1,a}(x_0)$, $\hat{x}_1 \in V_1(x_1, \alpha^1)$, $x_2 = \Phi^{t_2,t_1}(\hat{x}^1)$, $\hat{x}_2 \in V_2(x_2, \alpha^2)$, $\ldots$, $x_{m+1} = \Phi^{b,t_m}(\hat{x}_m)$, $y = g(x_{m+1})$.

Then the hypothesis that $g(\mathcal{R}^\Sigma_{[a,b]}(S_0)) \cap S = \{g(q_*)\}$ implies that the images of the maps M and F are separated at $g(q_*)$. We want to show that this fact implies our conclusion. This requires using a "separation theorem," i.e. a necessary condition for two set-valued maps $\mu_i \in SVM(Q_i, Y)$ that map 0 to a point q to have the property that $\mathrm{Im}(\mu_1)$ and $\mathrm{Im}(\mu_2)$ are separated at q.

Notice that, if the μ_i are linear, the sets Q_i are convex and compact, and $q = 0$, then a necessary condition for $\mathrm{Im}(\mu_1) \cap \mathrm{Im}(\mu_2) = \{q\}$ is the existence of a nontrivial $w \in Y^*$ such that $w\mu_1 x \geq 0$ for all $x \in Q_1$ and $w\mu_2 x \leq 0$ for all $x \in Q_2$, *provided that at least one of the sets $\mu_i Q_i$ does not contain 0 in its interior relative to its linear span.* What we need is a generalization of this fact to the case when the μ_i are nonlinear and multivalued. Naturally, such a generalization should involve the "differentials" of the μ_i, so our first task is to single out a class of maps that can be differentiated in a reasonable sense and give rise to a good separation theorem. Moreover, since our map M is constructed by composing maps of two kinds —variations and Lipschitz flow maps— we should make sure that (I) our class contains all Lipschitz maps as well as our variations and (II) the "differential" of M can be computed in terms of the differentials of the variations and those of the flow maps, by means of some version of the chain rule.

For Lipschitz maps, one possible theory of generalized differentials is given by the Clarke generalized Jacobian. This theory, however, is not adequate here, because the chain rule for generalized Jacobians is not good enough, since all that can be said about the generalized gradient of a composite map is that it is contained in the convex hull of the composite of the generalized gradients. For Lipschitz flows, which are essentially iterates of a large number of Lipschitz maps, this gives a generalized gradient which is too large to be useful.

The correct theory for Lipschitz maps turns out to be that of *derivate containers*, introduced by J. Warga (cf. [11], [12], [13], [14]).

Definition 5.1. Let X, Y be FDRLS's, let Ω be an open subset of X, and let $F:\Omega\to Y$ be a map. A *Warga derivate container* of F at a point $p\in\Omega$ is a compact subset $\mathbf{D}$ of $L(X,Y)$ with the property that, if U is any neighborhood of $\mathbf{D}$ in $L(X,Y)$, then there exist a neighborhood V of p in Ω and a sequence $\{F_j\}$ of maps of class C^1 from V to Y such that (i) $F_j\to F$ uniformly on V as $j\to\infty$, and (ii) for every j, and every $x\in V$, the differential $DF_j(x)$ belongs to U. ∎

A function that has a Warga derivate container at p in the sense of Definition 5.1 is automatically Lipschitz near p, since it is a uniform limit of a sequence of functions that are uniformly Lipschitz on a neighborhood V of p. So the theory of derivate containers does not yet meet all our requirements, since we have to be able to differentiate maps such as the V_i, that need not be Lipschitz or even single-valued.

To produce a theory with all the desirable features we have to go beyond Lipschitz maps, and define the class of *semidifferentiable set-valued maps*:

Definition 5.2. If X, Y are FDRLS's, a set-valued map $F\in SVM(X,Y)$ is *semidifferentiable* at a point $p\in X$ if (i) F is regular, (ii) Dom(F) is a compact convex subset of X with nonempty interior, (iii) $p\in\mathrm{Dom}(F)$, (iv) F is single-valued at p, and (v) there exists a Lipschitz map $\Psi:V\to Y$, defined on a neighborhood V of p, such that $F(x)=\Psi(x)+o(||x-p||)$ as $x\to p$ through values in $V\cap\mathrm{Dom}(F)$ (i.e. $\lim_{x\to p,x\in\mathrm{Dom}(F),x\neq p}\frac{\sup\{||y-\Psi(x)||:y\in F(x)\|\}}{||x-p||}=0$). In that case, if $\mathbf{D}$ is a derivate container of Ψ at p, then $\mathbf{D}$ is a *semidifferential* of F at p. ∎

The following two results are the chain rule and the separation theorem for semidifferentiable maps. (The chain rule is trivial, but we call it a theorem because of the crucial role it plays in our proof of the Maximum Principle. The separation theorem is proved in [9].)

Theorem 5.1. *Let X^1, X^2, X^3 be FDRLS's, and let K^1, K^2 be compact convex subsets of X^1, X^2 with nonempty interior. Let $F^1\in SVM(X^1,X^2)$, $F^2\in SVM(X^2,X^3)$ be SVM's with domains K^1, K^2. Assume that $\mathrm{Im}(F^1)\subseteq K^2$. Define the composite map $F=F^2\circ F^1\in SVM(X_1,X_3)$ by*

$$(F^2\circ F^1)(x^1)\stackrel{\mathrm{def}}{=}\{x^3\in X^3:(\exists x^2\in F^1(x^1))(x^3\in F^2(x^2))\}\ . \tag{5.1}$$

Then: (i) if F^1 and F^2 are compactly-valued and upper semicontinuous, it follows that F is compactly-valued and upper semicontinuous as well, (ii) if F^1 and F^2 are regular, then

F is regular, (iii) if $p^1 \in K^1$, $p^2 \in K^2$, $F^1(p^1) = p^2$, and F^1, F^2 are semidifferentiable at p^1, p^2 with semidifferentials $\mathbf{D}^1$, $\mathbf{D}^2$, then (a) F is semidifferentiable at p^1 and (b) if

$$\mathbf{D} = \mathbf{D}^2 \circ \mathbf{D}^1 = \{D^2 \circ D^1 : D^1 \in \mathbf{D}^1 \,,\, D^2 \in \mathbf{D}^2\} \,, \tag{5.2}$$

then $\mathbf{D}$ is a semidifferential of F at p^1. ▮

Theorem 5.2. *Let X, Y be FDRLS's, let K be a compact convex subset of X with nonempty interior, and let $F \in SVM(X, Y)$ be such that $\mathrm{Dom}(F) = K$. Let $p \in K$ be such that F is single-valued at p, and write $q = F(p)$. Let $K_0 = \{x - p : x \in K\}$. Assume that F is semidifferentiable at p with semidifferential $\mathbf{D}$. Let $S \subseteq Y$ be such that $q \in S$, and assume that S has a generalized approximating cone C at q, such that either (i) C is not a linear subspace of Y, or (ii) $C = \{0\}$ and $q \in \mathrm{Clos}(S - \{q\})$. Finally, assume that $\mathrm{Im}(F)$ is separated from S at q. Then there exist $D \in \mathbf{D}$ and $w \in Y^*\backslash\{0\}$ such that $wDx \geq 0$ for every $x \in K_0$ and $wv \leq 0$ for every $v \in C$.* ▮

Notice that, if $X, Y, K, F, p, q, K, K_0, \mathbf{D}$ are as in the hypotheses of Theorem 5.2, and $q \notin \mathrm{Int}(\mathrm{Im}(F))$, then we can satisfy the remaining hypotheses of the theorem by taking $C = \{0\}$ and letting S be a sequence $\{q_j\}$ such that $q_j \to q$ but $q_j \notin \mathrm{Im}(F)$. We can then conclude that there is a $w \in Y^*\backslash\{0\}$ and a $D \in \mathbf{D}$ such that $wy \geq 0$ for every $y \in DK_0$, so $0 \notin \mathrm{Int}(DK_0)$. So Theorem 5.2 implies the *open mapping theorem*: if $X, Y, K, F, p, q, K, K_0, \mathbf{D}$ are as in the hypotheses of Theorem 5.2, and $0 \in \mathrm{Int}(DK_0)$ for every $D \in \mathbf{D}$, then $q \in \mathrm{Int}(\mathrm{Im}(F))$.

To prove Theorem 2.2 we use Theorems 5.1 and 5.2. The maps F_0, V_i, F, g and the flow maps are all semidifferentiable at the appropriate points, so if one knows how to compute semidifferentials for each of them, then the chain rule tells us how to get a semidifferential for M. Semidifferentials of F_0, the V_i, F and g are easily evaluated. For the flow maps, we use the following fact:

Proposition 5.1. *Let $a \leq b$, let X be a FDRLS, and let $\Omega \subseteq X$ be open. Let $f \in CVF_{LIL}(\Omega, [a, b])$. Let $\xi : [a, b] \to \Omega$ be a trajectory of f, and let $\mathcal{A}$ be the set of all measurable selections of $t \to \partial f_t(\xi(t))$. For each $A \in \mathcal{A}$, let $D_A : [a, b] \to L(X, X)$ be the solution of $\dot{D}(t) = A(t)D(t)$ such that $D(a) = \mathrm{id}_X$. Let Φ be the flow map corresponding to f over $[a, b]$ (that is, $\Phi(x) = y$ iff $(\exists \xi \in \mathrm{Traj}(f))(\xi(a) = x \wedge \xi(b) = y)$). Let $\mathbf{D} = \{D_A(b) : A \in \mathcal{A}\}$. Then $\mathbf{D}$ is a derivate container of Φ at $\xi(a)$.* ▮

We can now complete the proof of Theorem 2.2. For simplicity, assume that the number m of variation times is just 1. Write $k = k_1$, $\alpha = \alpha^1$, $V = V_1$, $\mathcal{V} = \mathcal{V}_1$, $\tau = t_1$, so M is the set-valud map given by

$$M(z, \alpha) = g(\Phi^{b,\tau}(V(\Phi^{\tau,a}(F_0(z)), \alpha))) \,. \tag{5.3}$$

Let $\mathcal{DV} = (v_1, \ldots, v_k)$. Let $\mathbf{D_1}$, $\mathbf{D_2}$ be, respectively, derivate containers of $\Phi^{\tau,a}$ at p_* and of $\Phi^{b,\tau}$ at $\xi_*(\tau)$. Use the fact that $\partial g(q_*)$ is a derivate container of g at q_*. The chain rule implies that, if we define $\hat{\mathbf{D}}$ to be the set of all maps

$$\hat{\mathbf{D}}_{D_1,D_2,B} : (z, \alpha) \to B(D_2(D_1 z + \sum_{i=1}^{k} \alpha_i v_i)) \,, \tag{5.4}$$

for $D_1 \in \mathbf{D}_1$, $D_2 \in \mathbf{D}_2$, $B \in \partial g(q_*)$, then $\hat{\mathbf{D}}$ is a derivate container of M at $(0,0)$. If we now choose $\mathbf{D}_1$ and $\mathbf{D}_2$ to be the derivate containers given by Proposition 5.3, and apply Theorem 5.2, the conclusion follows.

6 HIGH ORDER VARIATIONS AND NONLIPSCHITZ RIGHT-HAND SIDES

We now give three examples, to illustrate the power of Theorem 2.2. A systematic discussion of high-order conditions in relation to Theorem 2.2 will be included in [10].

The first example illustrates how the Lojasiewicz Maximum Principle can apply in situations where the classical and nonsmooth versions do not:

Example 6.1. Consider a control system in $\mathbb{R}^2$ of the form $\dot{x} = y + u\mu(x,y)$, $\dot{y} = u\nu(x,y)$, $|u| \leq 1$, where μ, ν are continuous functions such that $\mu(0,0) = 0$ and $\nu(0,0) \neq 0$. Is Σ locally controllable from $(0,0)$? A formal application of the Maximum Principle yields a positive answer, except for the fact that Σ does not satisfy the technical hypotheses of the classical and nonsmooth versions, because the right-hand side need not be locally Lipschitz in x and y. On the other hand, the Lojasiewicz Maximum Principle provides a rigorous justification of the formal argument, since the reference trajectory $t \to \xi(t) = (0,0)$ corresponds to the control $u(t) \equiv 0$, and the right-hand side of Σ is of class C^1 when $u = 0$. ∎

The next two examples are well known in the literature on high-order conditions, but perhaps less so in the nonsmooth analysis community. They illustrate how, in very simple situations, it can happen that high-order conditions settle controllability or optimality questions for which the Maximum Principle alone does not suffice, and how these conditions involve certain Lie brackets. We first describe in each case how the new Lie bracket information solves the problem, and then state two theorems showing that, in fact, these solutions are special cases of Theorem 2.2. The applicability of Theorem 2.2 to the second example —which is a time-optimal control problem rather than one of controllability— presupposes the optimal control version of Theorem 2.2 which, as explained in §1, is a simple corollary of Theorem 2.2. (As is well known, the conclusion in this case —for autonomous systems— is the existence of an adjoint variable with the usual properties plus the fact that the constant value of the Hamiltonian is ≤ 0.)

Example 6.2. Consider the control system $\dot{x}_1 = u_1$, $\dot{x}_2 = u_2$, $\dot{x}_3 = u_2 x_1$ in $\mathbb{R}^3$, with control constraints $|u_1| \leq 1$, $|u_2| \leq 1$. If $\xi_* : [0,1] \to \mathbb{R}^3$ is the constant trajectory given by $\xi_*(t) \equiv (0,0,0)$, then $\xi_*(1)$ is an interior point of the reachable set from $\xi_*(0)$ —i.e. our system is locally controllable from $(0,0,0)$— but this fact cannot be detected by applying the usual Maximum Principle, since ξ_* is a Pontryagin extremal. The controllability along ξ_* can, however, be easily detected by an elementary Lie bracket calculation. It suffices to observe that, if we write our system in vector form as $\dot{x} = u_1 f_1(x) + u_2 f_2(x)$, where the vector fields f_1, f_2 have components $(1,0,0)$ and $(0,1,x_1)$, respectively, then $[f_1, f_2]$ has components $(0,0,1)$, so $f_1(0,0,0)$, $f_2(0,0,0)$ and $[f_1, f_2](0,0,0)$ are linearly independent, and then Chow's Theorem implies local controllability from 0. ∎

Example 6.3. Consider the control system $\dot{x}_1 = 1 + x_2^2$, $\dot{x}_2 = u$ in $\mathbb{R}^2$, with control constraint $|u| \leq 1$. Suppose we want to go from $(0,0)$ to $(1,0)$ in minimum time. One possible candidate is the trajectory $\xi_* : [0,1] \to \mathbb{R}^2$ given by $\xi_*(t) = (t,0)$, corresponding to the "singular" (i.e. not bang-bang) control $u_*(t) \equiv 0$. This trajectory passes the test of the usual Maximum Principle, but it is easy to see by direct inspection that ξ_* is not optimal, since it is clear that motion from left to right is much faster if one is far away from the x_1 axis, so a "bang-bang" control of the form $u(t) = 1$ for $0 \leq t \leq \tau$, $u(t) = -1$ for $\tau < t \leq 2\tau$ ought to do better than u_*. (The actual value of τ turns out to be the solution of $6\tau + 2\tau^3 = 3$, which obviously satisfies $0 < \tau < \frac{1}{2}$, and the optimal time is $2\tau \sim 0.93$.) The nonoptimality of u_* can be detected by means of a high-order test: we write our system in the form $\dot{x} = f(x) + ug(x)$, where f, g have components $(1 + x_2^2, 0)$ and $(0,1)$, respectively, and verify that the bracket $[g,[f,g]]$ has components $(-2,0)$. The well known Legendre-Clebsch condition says that, for our trajectory to be optimal, there must exist an adjoint vector $t \to \zeta(t)$ that satisfies the usual conditions of the Maximum Principle, as well as the "high-order condition" $\zeta(t)[g,[f,g]](\xi_*(t)) \leq 0$. In our case, if we write $\zeta = (\zeta_1, \zeta_2)$, the Hamiltonian is $H = \zeta_1(1 + x_2^2) + \zeta_2 u$. The adjoint equation says that $\dot{\zeta}_1 = 0$, $\dot{\zeta}_2 = -2\zeta_1 x_2$, which reduces to $\dot{\zeta}_1 = \dot{\zeta}_2 = 0$ along our trajectory ξ_*. The value of the Hamiltonian along (ξ_*, u_*) is ζ_1. The fact that u_* must minimize the Hamiltonian says that ζ_1 must vanish. The Maximum Principle also requires that $H \leq 0$ along (ξ_*, u_*), and that ζ be $\neq 0$. All these conditions can be satisfied by taking $\zeta = (c, 0)$, where $c < 0$ is a constant, and these are the only choices of ζ that meet all the requirements. However, the extra condition that $\zeta[g,[f,g]] \leq 0$, implies $-2c \leq 0$, which is a contradiction. So ξ_* is not optimal. ∎

We now show how the application of Chow's Theorem in Example 6.2 and the use of the Legendre-Clebsh condition in Example 6.3 are particular cases of Theorem 2.2. This follows from the following two facts, proved in [10]:

Theorem 6.1. *Let Π be a convex polyhedron in $\mathbb{R}^m$, and let $\mathcal{F}(\Pi)$ be the set of open faces of Π. Let $g_1, \ldots, g_m$ be vector fields of class C^1 on an open subset Ω of $\mathbb{R}^n$, and write $ug = u_1 g_1 + \ldots + u_m g_m$ for $u = (u_1, \ldots, u_m) \in \mathbb{R}^m$. Let Σ be the control system $\dot{x} = ug(x)$, $u \in \Pi$, where the open-loop controls are all measurable Π-valued functions defined on compact intervals. Let $\gamma_* = (\xi_*, \eta_*)$ be a controlled trajectory of Σ, and let $[a,b] = \mathrm{Dom}(\gamma_*)$. For each $F \in \mathcal{F}(\Pi)$, let $E_F(\eta_*) = \{t \in [a,b] : \eta_*(t) \in F\}$. Then, if $F \in \mathcal{F}(\Pi)$ and $v, w \in \mathbb{R}^m$ are parallel to F, it follows that $([vg, wg](\xi_*(t)), t)$ is a variational vector for γ_* for almost every $t \in E_F(\eta_*)$.* ∎

(Recall that an open face of Π is, by definition, a maximal member —with respect to inclusion— of the set $\{F : (\exists S \in Aff(\mathbb{R}^m))(F = \mathrm{Int}_S(\Pi \cap S))\}$, where $Aff(\mathbb{R}^m)$ is the set of affine subspaces of $\mathbb{R}^m$.)

Theorem 6.2. *Let f, g be vector fields of class C^1 on an open subset Ω of $\mathbb{R}^n$, and let Σ be the control system $\dot{x} = f(x) + ug(x)$, $u \in [-1,1]$, where the open-loop controls are all measurable $[-1,1]$-valued functions defined on compact intervals. Let $\gamma_* = (\xi_*, \eta_*)$ be a controlled trajectory of Σ, and let $[a,b] = \mathrm{Dom}(\gamma_*)$. Let $E(\eta_*) = \{t \in [a,b] : -1 < \eta_*(t) < 1\}$. Then $(-[g,[f,g]](\xi_*(t)), t)$ is a variational vector for γ_* for almost every $t \in E(\eta_*)$.* ∎

REFERENCES

1 R. M. Bianchini and G. Stefani, Local controllability about a reference trajectory, in *Analysis and Optimization of Systems*, A. Bensoussan and J.L. Lions Eds., Lect. Notes on Control 83, Springer-Verlag, Berlin, 1986.

2 F. H. Clarke, *Optimization and Nonsmooth Analysis,* Wiley Interscience, New York, 1983.

3 B. Kaskosz and S. Lojasiewicz Jr., "A Maximum Principle for generalized control systems," *Nonlinear Anal. TMA*, **9** (1985), pp. 109-130.

4 M. Kawski, "High-order small-time local controllability," in *Nonlinear Controllability and Optimal Control*, H.J. Sussmann ed., M. Dekker Inc. (1990), pp. 431-467.

5 H. Knobloch, *High Order Necessary Conditions in Optimal Control Theory*, Springer Lect. Notes in Control and Information Sciences, No. 34, New York (1981).

6 S. Lojasiewicz Jr., Local controllability of parametrized differential equations, in preparation.

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

8 H. J. Sussmann, A general theorem on local controllability, *S.I.A.M. J. Control and Optimization* 25, No. 1: 158-194 (1987).

9 H.J. Sussmann, Open mapping and separation theorems for semidifferentiable maps, in preparation.

10 H.J. Sussmann, A strong Maximum Principle under weak hypotheses, in preparation.

11 J. Warga, Fat homeomorphisms and unbounded derivate containers, *J. Math. Anal. Appl.* **81** (1981), 545-560.

12 J. Warga, Controllability, extremality and abnormality in nonsmooth optimal control, *J. Optim. Theory Applic.* **41** (1983), 239-260.

13 J. Warga, Optimization and controllability without differentiability assumptions, *SIAM J. Control and Optimization* **21** (1983), 837-855.

14 J. Warga, Homeomorphisms and local C^1 approximations, *Nonlinear Anal. TMA* **12** (1988) 593-597.