This paper has appeared in the Proceedings of the 33rd IEEE CDC, Orlando, FL, Dec. 1994, pp. 1950-1956

A strong version of the Maximum Principle under weak hypotheses*

Héctor J. Sussmann
Department of Mathematics, Rutgers University, New Brunswick, NJ 08903
E-mail: sussmann@hamilton.rutgers.edu

Abstract

We present a version of the finite-dimensional Maximum Principle that incorporates high-order point variations and is valid under minimal technical assumptions, weaker than those of the usual classical and nonsmooth versions. The proof follows the classical idea of using needle variations and applying an appropriate open mapping theorem to a multiparameter variation, whose effect is computed in terms of those of the needle variatipons by means of the chain rule. However, this has to be carried out in a new setting, namely, the class of "semidifferentiable maps," that contains all maps arising in the optimal control problem and has a concept of generalized differential with all the right properties.

1. Introduction

The Maximum Principle (MP) of optimal control theory deals with a trajectory $[a, b] \ni t \to x_*(t) \in M$ of a control system Σ with state space M, and gives a necessary condition for x_* to be a "boundary trajectory" (i.e. such that $x_*(b)$ lies on the boundary of the set of points reachable from $x_*(a)$) or a minimizer of some cost functional of the Lagrange, Mayer or Bolza type, subject to various constraints. We will only be concerned here with the finite-dimensional versions of the MP, in which M is an open subset of $\mathbb{R}^n$ or a smooth manifold. In most such versions, the trajectories are supposed to be generated by a dynamical equation $\dot{x} = f(x, t, u)$, together with the specification of a class $\mathcal{U}$ of "admissible open-loop controls" $t \to u(t)$. The "classical" version of the MP (cf., e.g. [6]) requires f to be of class C^1 with respect to x, with an appropriate integral bound on $\frac{\partial f}{\partial x}$. The "nonsmooth analysis" version (cf., e.g., [2], [3]) requires that f be Lipschitz with respect to x, on some tube $\{(x, t, u) : ||x - x_*(t)|| \leq \delta\}$ about the reference trajectory x_*, with a Lipschitz constant $k(t, u)$ such that the function $t \to k(t, u_*(t))$ is integrable, where u_* is a control that generates x_*. More recently, S. Lojasiewicz Jr. has announced a sufficient condition for local controllability along x_* (i.e., equivalently, a necessary condition for x_* to be a boundary trajectory), in which it is only required that the time-varying vector field $(x, t) \to f_*(x, t)$ of which x_* is a trajectory be locally Lipschitz with respect to x for each t on some tube containing x_* —with a Lipschitz constant $\kappa(t)$ which is integrable as a function of t— but the other vector fields of the system are not required to be Lipschitz in x. (For a simple illustration of the extra power of Lojasiewicz's result, cf. Example 1 below.)

In this note we describe an extension of the Lojasiewicz result with the following features: (1) We allow the inclusion of the effect of *high-order point variations*, thereby obtaining a stronger necessary condition. (Our version is a result for general "point variational vectors," which contains the ordinary MP as a special case, corresponding to the "needle variations.") (2) We formulate the MP as a general necessary condition for a *local separation property*, defined below in Section 2. (Most other problems usually considered in optimal control can easily be reduced to separation questions, cf. Example 2.) (3) We allow fairly general control spaces, with no structure whatsoever. (This follows the ideas of B. Kaskosz and S. Lojasiewicz Jr. on "generalized control systems," cf. [4], and those of Lojasiewicz in [5].) It turns out that the MP can be proved for general parametrized families $F = \{F^\eta\}_{\eta \in \mathcal{U}}$ of time-varying vector fields F^η, where $\mathcal{U}$ is just a set. This general setting suffices to get the "weak" form of the MP, and then, under mild extra conditions —some of which are described below in Section 8— one can pass from the "weak" form to the "strong" one, and get the full force of the MP. (4) In particular, even for "classical" systems $\dot{x} = f(x, t, u)$, we are allowed to consider fairly general classes of open-loop controls —e.g. piecewise continuous, or piecewise constant— and need not work with the class of all measurable controls[†]. (5) Finally —but in our view most importantly— *the proof is surprisingly simple,* and very much in the spirit of the classical approach of [6], although a minimal amount of "nonsmooth analysis" is required, in the form of the definition and basic properties of the Clarke generalized Jacobian.

The proof essentially relies on two basic facts from Advanced Calculus, namely, the chain rule and the open mapping theorem (used, respectively, to compute the differential of the endpoint map, and to infer, from the surjectivity —or transversality to some cone— of the differential, that the reachable set is itself a neighborhood[‡] of $x_*(b)$, or has a suitable transversality property). The crucial technical point is to redo the relevant Advanced Calculus —and, in particular the chain rule and open mapping properties— in a nontraditional setting, working within a class SD large enough to contain all the maps that occur in the proof of the MP —some of which are neither Lipschitz, nor differentiable in the ordinary sense, or even single-valued— and using an appropriate definition of "differential." We take SD to be the class of "semidifferentiable maps" (cf. Section 4), and use as our concept of generalized differential an extension of J. Warga's theory of "derivate containers."

*This research was supported in part by NSF Grants DMS-8902994 and DMS92-02554.

[†]We never use the Ekeland variational principle, so we do not need the space of controls to be complete for the Ekeland metric.

[‡]We always use "neighborhood" as in point set topology: N is a neighborhood of p iff $p \in \text{Interior}(N)$.

We now proceed to substantiate our earlier assertions that (a) other optimal control questions can be reduced to separation problems, and (b) the Lojasiewicz theorem can give stronger information than the nonsmooth MP.

Example 1 Consider the Lagrange problem of minimizing $\int_a^b L(x(t),t,u(t))dt$ among all trajectory-control pairs[§] $(x(\cdot),u(\cdot))$ for $\dot{x}=f(x,t,u)$ such that $x(a)=p_*$, $x(b)\in S$, where the numbers a,b, the point p_* and the set S are given. Let (x_*,u_*) be a solution. Let $q_*=x_*(b)$. Pick a smooth function $x\to\varphi(x)$ such that $\varphi(q_*)=0$ and $\varphi(x)>0$ if $x\neq q_*$. Let $\tilde{\Sigma}$ be the system with dynamics $\dot{X}=F(X,t,u)$, where $X=(x,y)$, y is a new scalar variable, and $F(X,t,u)=(f(x,t,u),L(x,t,u))$. Let $y_*(t)=\int_a^t L(x_*(s),s,u_*(s))ds$, $X_*(t)=(x_*(t),y_*(t))$. Write $Q_*=(q_*,y_*(b))$. Let $\tilde{S}=\{(x,y)\in S\times\mathbb{R}, y\leq y_*(b)-\varphi(x)\}$, $P_*=(p_*,0)$. Then $\mathcal{R}^{\tilde{\Sigma}}_{[a,b]}(P_*)\cap\tilde{S}=\{Q_*\}$, so the MP applies. ∎

Example 2 Consider a system $\dot{x}=y+ug(x,y)$, $\dot{y}=uh(x,y)$, $|u|\leq 1$, where $g,h:\mathbb{R}^2\to\mathbb{R}$ are continuous, $g(0,0)=0$, and $h(0,0)=1$. By formally applying the MP we find that this system is locally controllable from $(0,0)$. However, the usual forms of the MP do not rigorously imply this conclusion, because they require the right-hand side to be Lipschitz in x and y for all values of u. The Lojasiewicz result makes the application of the MP completely rigorous. ∎

2. The Maximum Principle

We first introduce some notations and definitions.

We use $\mathbb{R}$, $\mathbb{R}^{n\times m}$, to denote the real line and the space of n by m real matrices. We write $\mathbb{R}^n=\mathbb{R}^{n\times 1}$, $\mathbb{R}_n=\mathbb{R}^{1\times n}$, $\mathbb{R}_+=[0,+\infty)$, $\bar{\mathbb{R}}_+=\mathbb{R}_+\cup\{+\infty\}$. "Interval" means "nonempty connected subset of $\mathbb{R}$." If μ is a map, then $\mathrm{Dom}(\mu)$ is the domain of μ.

If $M\subseteq\mathbb{R}^n$ is open, and I is an interval, then a *control vector field* (abbr. CVF) with *state space* M and *time domain* I, is a map $M\times I\ni(x,t)\to F(x,t)\in\mathbb{R}^n$ which is continuous in x for each fixed t and measurable in t for each fixed x. In that case, F_t will denote, for $t\in I$, the partial map $x\to F(x,t)$, and $\mathrm{TDom}(F)$ will denote the time domain of F. We use $CVF(M,I)$ to denote the set of CVF's with state space M and time domain I, and write $CVF(M)=\cup_I CVF(M,I)$. An $F\in CVF(M,I)$ is *locally integrably bounded* (abbr. LIB) if for every compact subset K of M there exists a locally Lebesgue integrable function $\psi_K^F:I\to\bar{\mathbb{R}}_+$ such that $||F(x,t)||\leq\psi_K^F(t)$ for all $(x,t)\in K\times I$. We call F *locally integrably Lipschitz* (abbr. LIL) if it is LIB and for every compact subset K of A there exists a locally Lebesgue integrable function $\hat{\psi}_K^F:I\to\bar{\mathbb{R}}_+$ such that $||F(x,t)-F(y,t)||\leq\hat{\psi}_K^F(t)||x-y||$ for all $(x,y,t)\in K\times K\times I$.

We use $\partial F(x)$ to denote the Clarke generalized Jacobian of a locally Lipschitz map F at a point x. If F is a CVF such that each F_t is locally Lipschitz, then $\partial_x F$ denotes the map $(x,t)\to\partial F_t(x)$, i.e. the Clarke generalized Jacobian of F with respect to the state variable.

A *control system* in $\mathbb{R}^n$ is a parametrized family of CVF's with a common state space, i.e., more precisely, a triple $\Sigma=(M,\mathcal{U},F)$ where M —the *state space* of Σ— is open in $\mathbb{R}^n$, $\mathcal{U}$ —the *control space*— is a set, and F —the *dynamics*— is a family $\{F^\eta\}_{\eta\in\mathcal{U}}$ of CVF's on M. The elements of $\mathcal{U}$ are the (*open-loop*) *controls.*

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

A *controlled trajectory* of a control system $\Sigma=(M,\mathcal{U},F)$ is a pair $\gamma=(\xi,\eta)$ such that $\eta\in\mathcal{U}$ and $\xi\in\mathrm{Traj}(F^\eta)$. A *trajectory* of Σ is a ξ such that (ξ,η) is a controlled trajectory of Σ for some $\eta\in\mathcal{U}$.

Let $\Sigma=(M,\mathcal{U},f)$ be a control system. If $p,x\in M$, $a\in\mathbb{R}$, $b\in\mathbb{R}$, $a\leq b$, we say that x is Σ-*reachable from* p *over* $[a,b]$ if there exists a controlled trajectory $\gamma=(\xi,\eta)$ of Σ such that $\xi(a)=p$ and $\xi(b)=x$. We define $\mathcal{R}^\Sigma_{[a,b]}(p)$, the Σ-*reachable set from* p *over* $[a,b]$, to be the set of all $x\in M$ that are Σ-reachable from p over $[a,b]$, and write $\mathcal{R}^\Sigma(p)\stackrel{\mathrm{def}}{=}\cup_{a,b}\,\mathcal{R}^\Sigma_{[a,b]}(p)$. More generally, we define reachable sets from subsets of M, 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 $\mathbb{R}$. The system $\Sigma=(M,\mathcal{U},F)$ will be said to have the *substitution property for* $\mathcal{S}$ if, whenever $\eta_1,\eta_2\in\mathcal{U}$, $E\in\mathcal{S}$, $E\subseteq\mathrm{TDom}(F^{\eta_1})\cap\mathrm{TDom}(F^{\eta_2})$, it follows that there exist an $\eta_3\in\mathcal{U}$ such that $\mathrm{TDom}(F^{\eta_1})\subseteq\mathrm{TDom}(F^{\eta_3})$ and $F^{\eta_3}(x,t)=\chi_E(t)F^{\eta_2}(x,t)+(1-\chi_E(t))F^{\eta_1}(x,t)$ for $(x,t)\in\mathrm{Dom}(F^{\eta_1})$. (Here χ_E is, for a set E, the indicator function of E.) If Σ has the substitution property for the class of all measurable subsets (resp. compact subsets, compact subintervals) of $\mathbb{R}$, then Σ will be said to have the *measurable (*resp. *compact, interval) substitution property* (abbr. MSP, CSP, ISP).

If T is a topological space, $A\subseteq T$, $B\subseteq T$, and $q_*\in A\cap B$, we say that A and B are *locally separated* at q_* if $U\cap A\cap B=\{q_*\}$ for some neighborhood U of q_*.

If $\Sigma=(M,\mathcal{U},F)$ is a control system, $x_*:[a,b]\to M$ is a trajectory of Σ, $S_0,S\subseteq M$, $x_*(a)\in S_0$, and $x_*(b)\in S$, our version of the MP will give necessary conditions for $\mathcal{R}^\Sigma_{[a,b]}(S_0)$ and S to be locally separated at $x_*(b)$ or, more generally —as in Clarke [3]— for $g(\mathcal{R}^\Sigma_{[a,b]}(S_0))$ and S to be separated at $g(x_*(b))$, where $S\subseteq\mathbb{R}^k$ for some k, and g maps some neighborhood of $x_*(b)$ to $\mathbb{R}^k$.

The statement of the MP deals with:

(D1) a control system $\Sigma=(M,\mathcal{U},f)$ in $\mathbb{R}^n$;

(D2) an interval $[a,b]$, points $p_*,q_*\in M$, a (possibly set-valued) map g, with values in $\mathbb{R}^m$, defined on a neighborhood U of q_* in $\mathbb{R}^n$, and subsets S_0, S of $\mathbb{R}^n$, $\mathbb{R}^m$ such that $p_*\in S_0$ and $g(q_*)\in S$;

(D3) a controlled trajectory $\gamma_*=(\xi_*,\eta_*)$ of Σ, defined on $[a,b]$ and such that $p_*=\xi_*(a)$ and $q_*=\xi_*(b)$;

[§] Such objects are sometimes called "admissible pairs" in the control literature. We will call them "controlled trajectories."

The technical assumptions are
(A1) either: (i) Σ has the CSP, or (ii) Σ has the ISP and every F^η, $\eta \in \mathcal{U}$ is LIB;
(A2) The control $\eta_* \in \mathcal{U}$ is such that F^{η_*} is LIL;
(A3) S_0 has a GAC —i.e. a "generalized approximating cone"¶— C_0 at p_*, and S has a GAC C at $g(q_*)$;
(A4) either (i) C is not a linear subspace of $\mathbb{R}^m$ or (ii) $C = \{0\}$ and $g(q_*)$ belongs to $\mathrm{Clos}(S - \{g(q_*)\})$;
(A5) g is semidifferentiable at q_*.

Theorem 1 (The Strong Maximum Principle.) *Assume that data (D1)...(D3) are given, such that (A1)...(A5) hold. Let $G \subseteq \mathbb{R}^{m\times n}$ be a semidifferential of g at q_*. Let* $\mathbf{v}$ *be a compatible set of point variational vectors for γ_*. Then, if $g(\mathcal{R}^{\Sigma}_{[a,b]}(S_0))$ is locally separated from S at $g(q_*)$, it follows that there exist: (I) a measurable matrix function $A : [a,b] \to \mathbb{R}^{n\times n}$ such that $A(t) \in \partial_x F^{\eta_*}(\xi_*(t),t)$ for a.e. $t \in [a,b]$, (II) a $B \in G$, (III) a $c^* \in \mathbb{R}_m - \{0\}$, such that, if $\zeta : [a,b] \to \mathbb{R}_n$ is the unique absolutely continuous solution of $\dot\zeta(t) = -\zeta(t)A(t)$ with terminal condition $\zeta(b) = c^*B$, then (i) for every $\eta \in \mathcal{U}$ the inequality $\zeta(t).F^{\eta_*}(\xi_*(t),t) \leq \zeta(t).F^{\eta}(\xi_*(t),t)$ holds for almost all t in $[a,b] \cap \mathrm{TDom}(\eta)$, (ii) $c^*c \leq 0$ for all $c \in C$, (iii) $\zeta(a)c_0 \geq 0$ for all $c_0 \in C_0$, and (iv) $\zeta(t)v \geq 0$ for all $(v,t) \in \mathbf{v}$.* ▮

Remark 1 If g is locally Lipschitz near q_*, then g is semidifferentiable at q_*, and in that case we can take G to be any Warga derivate container (cf. Section 3) of g at q_*, and in particular we can choose $G = \partial g(q_*)$. ▮

Remark 2 Theorem 1 can be strengthened in various ways. For example:
1. Conclusion only (i) gives a *weakly* minimizing ζ, i.e. a ζ such that for each η the inequality $\zeta(t)F^{\eta_*}(\xi_*(t),t) \leq \zeta(t)F^{\eta}(\xi_*(t),t)$ holds for all t in the complement of a null set $E(\eta)$ that may depend on η. Call ζ *strongly minimizing* if $\zeta(t)F^{\eta_*}(\xi_*(t),t) = \min\{\zeta(t)F^{\eta}(\xi_*(t),t) : \eta \in \mathcal{U}\}$ for a.e. $t \in [a,b]$, i.e. if $E(\eta)$ can be chosen not to depend on η. Then under mild conditions one can take ζ to be strongly minimizing (cf. Section 8.)
2. If $\Sigma = (M,\mathcal{U},F)$ is invariant under deletions —that is, for every $\eta \in \mathcal{U}$, $t_* \in \mathrm{TDom}(F^\eta)$, $\alpha > 0$, the CVF $(x,t) \to \chi_{(-\infty,t_*)}(t)F^\eta(x,t) + \chi_{[t_*,\infty)}(t)F^\eta(x,t+\alpha)$ occurs in F— and insertions —that is, for every $\eta, \tilde\eta \in \mathcal{U}$, $t_* \in \mathbb{R}$, $\alpha > 0$ such that $[t_*, t_*+\alpha] \subseteq \mathrm{TDom}(F^\eta) \cap \mathrm{TDom}(F^{\tilde\eta})$, the CVF $(x,t) \to \chi_{(-\infty,t_*)}(t)F^\eta(x,t) + \chi_{[t_*,t_*+\alpha]}(t)F^{\tilde\eta}(x,t) + \chi_{(t_*+\alpha,\infty)}(t)F^\eta(x,t-\alpha)$ occurs in F— then one gets the stronger conclusion that the Hamiltonian $t \to \zeta(t).F^{\eta_*}(\xi_*(t),t)$ can be taken to be a.e. equal to a constant ρ. If in addition one assumes that $\cup_{0\leq\varepsilon\leq\bar\varepsilon}\mathcal{R}^{\Sigma}_{[a,b-\varepsilon]}(S_0)$ (resp. $\cup_{0\leq\varepsilon\leq\bar\varepsilon}\mathcal{R}^{\Sigma}_{[a,b+\varepsilon]}(S_0)$, $\cup_{-\bar\varepsilon\leq\varepsilon\leq\bar\varepsilon}\mathcal{R}^{\Sigma}_{[a,b+\varepsilon]}(S_0)$) is locally separated from S_0 at q_*, then ρ can be taken to be ≤ 0 (resp. ≥ 0, $= 0$).
3. The substitution condition —which is required for needle variations to be "admissible" in the sense of Section 6— excludes problems $\dot x = f(x,t,u)$ where the controls are assumed to be continuous, or smooth. To incorporate these cases, one can develop a more general theory, in which distributed variations are allowed.
4. Call two control systems $\Sigma = (M,\mathcal{U},F)$, $\hat\Sigma = (M,\hat{\mathcal{U}},\hat F)$ "trajectory-equivalent" (TE) if their sets of trajectories are identical. If we are given data (D1)...(D3) for which (A1)...(A5) hold, and $\hat\Sigma = (M,\hat{\mathcal{U}},\hat F)$ is TE to $\hat\Sigma$, then we may apply the MP with $\hat\Sigma$ instead of Σ, using instead of η_* any control $\hat\eta_* \in \hat{\mathcal{U}}$ such that $\xi_* \in \mathrm{Traj}(\hat F^{\hat\eta_*})$ and $\hat F^{\hat\eta_*}$ is LIL. This adds extra power to the MP because, as pointed out by Kaskosz and Lojasiewicz in [4], a pair (ξ_*,η_*) could satisfy the conditions of the conclusion of the MP, while the conclusion fails for $(\xi_*,\hat\eta_*)$ for some other control $\hat\eta_*$ that also generates ξ_*. Clearly, there exists a system $\hat\Sigma_{max}$ which is maximal among all systems that are TE to Σ. (The dynamics $\hat F$ of $\hat\Sigma$ consists of all the CVF's G such that $\mathrm{Traj}(G) \subseteq \cup_\eta \mathrm{Traj}(F^\eta)$.) The MP clearly yields maximum information when applied to $\hat\Sigma_{max}$. Since $\hat\Sigma_{max}$ is not of the form $\dot x = f(x,t,u)$, with $u \in \mathbb{R}^n$, even if Σ is, this gives an extra reason for looking at general classes of vector fields, as we have done here following [4].
5. We can allow *systems with jumps.* For instance, one could consider a finite sequence $(\Sigma_0,\ldots,\Sigma_m)$ of systems $\Sigma_i = (M_i,\mathcal{U}_i,F^i)$ on spaces $X_i = \mathbb{R}^{n_i}$, together with "jump times" $t_1 < t_2 < \ldots < t_m$ in $[a,b]$ and possibly set-valued "jump maps" J_i from M_{i-1} to M_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 each ξ_i is a trajectory of Σ_i for η_i, defined on $[t_{i-1},t_i]$ (where $t_0 = a$, $t_{m+1} = b$), such that $\xi_i(t_i) = J_i(\xi_{i-1}(t_i))$ for $i = 1,\ldots,m$. 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 $\mathbf{J}_i$, and their domains should be neighborhoods of the x_i^*. (In particular, the J_i could be Lipschitz, in which case we could take $\mathbf{J}_i$ to be $\partial J_i(x_i^*)$, or any other Warga derivate container of J_i at x_i^*.) The definition of the reachable set is obvious, and the conclusion is the same as that of Theorem 1, except that we now get finite sequences $A = (A_0,\ldots,A_m)$, $\zeta = (\zeta_0,\ldots,\zeta_m)$ such that $A_i : [t_{i-1},t_i] \to \mathbb{R}^{n_i\times n_i}$, $\zeta_i : [t_{i-1},t_i] \to \mathbb{R}_{n_i}$ with the obvious properties, plus the jump condition $(\exists D_i \in \mathbf{J}_i)(\zeta_{i-1}(t_i) = \zeta_i(t_i)D_i)$.
6. The *Maximum Principle on manifolds* follows trivially from the MP 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.)
7. Instead of systems with fixed jumps at fixed times, one can also allow optional jumps at optional times, in which case one obtains extra conditions on ζ.
8. State space constraints can be incorporated as well. ▮

¶The concept of a "generalized approximating cone," as well as others used in the statement of Theorem 1 —such as "semidifferentiable map," "semidifferential," "variational vectors" and "compatibility"— will be defined in later sections.

3. Warga derivate containers

The theory of "derivate containers" was developed by J. Warga in [10], [11], [12], [13].

Definition 1 Assume $M \subseteq \mathbb{R}^n$ is open, $F : M \to \mathbb{R}^m$ is a map, and $p \in M$. A *(Warga) derivate container* of F at p is a compact subset $\mathbf{W}$ of $\mathbb{R}^{m\times n}$ such that, if U is any neighborhood of $\mathbf{W}$ in $\mathbb{R}^{m\times n}$, then there exist a neighborhood V_U of p in M and a sequence $\{F_j\}$ of maps of class C^1 from V_U to $\mathbb{R}^m$ such that (i) $F_j \to F$ uniformly on V_U as $j \to \infty$, and (ii) for every j, and every $x \in V_U$, the differential $DF_j(x)$ is in U. ∎

Remark 3 *A map F that has a Warga derivate container at p has to be Lipschitz near p,* since it is a uniform limit of a sequence of maps that are uniformly Lipschitz on a neighborhood V of p. The Clarke generalized Jacobian $\partial F(p)$ is a Warga derivate container of F at p, and can be characterized as the smallest of all convex derivate containers of F at p. ∎

If F is a LIL CVF, and $a, b \in \mathrm{TDom}(F)$, then $\Phi_F^{b,a}$ denotes the flow map corresponding to F and the times a, b. That is, $\Phi_F^{b,a}(x) = y$ iff $(\exists \xi \in \mathrm{Traj}(F))(\xi(a) = x \wedge \xi(b) = y)$. The following result, due to J. Warga, plays a crucial role in the proof of the MP:

Theorem 2 *(Warga's Theorem on differentiation of Lipschitz flows.) Assume that $a \leq b$, $M \subseteq \mathbb{R}^n$ is open, $F : M \times [a, b] \to \mathbb{R}^n$ is a LIL CVF on M, and $\xi : [a, b] \to M$ is a trajectory of F. Let $\mathbf{A}$ be the set of all measurable selections of $t \to \partial F_t(\xi(t))$. For each $A \in \mathbf{A}$, let $W_A : [a, b] \to \mathbb{R}^{n\times n}$ be the solution of $\dot{W}(t) = A(t)W(t)$ such that $W(a) =$ identity. Let $\mathbf{W} = \{W_A(b) : A \in \mathbf{A}\}$. Then $\mathbf{W}$ is a derivate container of $\Phi_F^{b,a}$ at $\xi(a)$.* ∎

4. Semidifferentiable maps

We now define semidifferentiable maps, and review their main properties, without proof.

A *set-valued map* (abbr. SVM) (or *multifunction*) 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 (and, if $M(x) = \{y\}$, we 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 $M \in SVM(A, B)$ 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 X and Y are topological spaces, an $F \in SVM(X, Y)$ is *compactly valued* if $F(x)$ is compact for every $x \in \mathrm{Dom}(F)$, and *upper semicontinuous* (abbr. 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 call F *compact* if F has compact domain and is compactly-valued and USC, i.e. if F is compact as a subset of $X \times Y$.

Definition 2 Let $F \in SVM(\mathbb{R}^n, \mathbb{R}^m)$, and assume F is compact. A *single-valued approximating sequence* (SVAS) for F is a sequence $\{F_j\}_{j=1}^{\infty}$ of continuous $\mathbb{R}^m$-valued maps with domain $\mathrm{Dom}(F)$, such that

(SV1) the maps F_j, $j = 1, 2, \ldots$ are uniformly bounded;

(SV2) 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 $(x, y) = \lim_{\ell\to\infty}\big(x_{j(\ell)}, F_{j(\ell)}(x_{j(\ell)})\big)$ exists, then $y \in F(x)$.

We say that F *has single-valued approximations* if there exists a SVAS for F. ∎

Definition 3 An $F \in SVM(\mathbb{R}^n, \mathbb{R}^m)$ is *regular* near $p \in \mathrm{Dom}(F)$ if there is a neighborhood N of p such that $F\lceil N$ is compact and has a SVAS. ∎

Definition 4 Let $F, \hat{F} \in SVM(\mathbb{R}^n, \mathbb{R}^m)$, $p \in \mathbb{R}^n$. We say that $\hat{F}$ is a *first-order approximation* (abbr. FOA) *to F near p*, if $||x - p||^{-1} \sup\{||y - \hat{y}|| : (y, \hat{y}) \in F(x) \times \hat{F}(x)\}$ goes to 0 as $x \to p$ via $\mathrm{Dom}(F)\backslash\{p\}$. We write $FOA(F, p)$ to denote the set of all FOA's of F near p. ∎

Since $\sup\, \emptyset = -\infty$, a necessary condition for $FOA(F, p)$ to be $\neq \emptyset$ is that $\mathrm{Dom}(F) \cap U \subseteq \mathrm{Dom}(\hat{F}) \cap U$ for some neighborhood U of p.

We use $FOA_{LIP}(F, p)$ to denote the set of all $\hat{F} \in FOA(F, p)$ that are Lipschitz.

We say that a subset K of $\mathbb{R}^n$ is *closed* (resp. *convex*) *near* $p \in \mathbb{R}^n$ if there is a neighborhood U of p such that $U \cap K$ is a closed (resp. convex) subset of $\mathbb{R}^n$. If K is convex near p, U is a neighborhood of p such that $K \cap U$ is convex, and $K_p^U = \{x - p : x \in K \cap U\}$, then the set $T_pK \stackrel{\mathrm{def}}{=} \mathrm{Clos} \cup_{r\geq 0} rK_p^U$ —which is independent of U— is the *tangent cone* of K at p.

Definition 5 If $F \in SVM(\mathbb{R}^n, \mathbb{R}^m)$, and $p \in \mathbb{R}^n$, we say that F is *semidifferentiable at p* if

(SD1) F is regular near p and single-valued at p;

(SD2) $\mathrm{Dom}(F)$ is closed near p and convex near p;

(SD3) $FOA_{LIP}(F, p) \neq \emptyset$.

If F is semidifferentiable at p, then a *semidifferential* of F at p is a subset $\mathbf{W}$ of $\mathbb{R}^{m\times n}$ which is a Warga derivate container of $\hat{F}$ at p for some $\hat{F} \in FOA_{LIP}(F, p)$. We use $SD(F, p)$ to denote the set of all semidifferentials of F at p, so $SD(F, p) \neq \emptyset$ iff F is semidifferentiable at p. ∎

Remark 4 *The semidifferential $\mathbf{W}$, when it exists, is never unique,* because (a) $\hat{F}$ is not unique, and (b) for a given $\hat{F}$, the derivate container of $\hat{F}$ at p is not unique. In fact, *if $\mathbf{W} \in SD(F, p)$, $\mathbf{W} \subseteq \tilde{\mathbf{W}} \subseteq \mathbb{R}^{m\times n}$, and $\tilde{\mathbf{W}}$ is compact, then $\tilde{\mathbf{W}} \in SD(F, p)$ as well.* In applications, it is always desirable to pick $\hat{F}$ so that $\mathbf{W}$ can be taken to be as small as possible. ∎

‖So in particular the graph of a SVM M is just M itself.

Remark 5 *If $\hat{F} \in FOA_{LIP}(F,p)$, then $\hat{F}(p) = F(p)$, and F is continuous at p, i.e.* $\sup\{||y - F(p)|| : y \in F(x)\}$ goes to 0 as $x \to p$, $x \in \mathrm{Dom}(F)$. ∎

Remark 6 *If $\mathbf{W} \in SD(F,p)$ and $\mathbf{W} = \{W\}$, $W \in \mathbb{R}^{m\times n}$, then F is differentiable at p along $K = \mathrm{Dom}(F)$ in the ordinary sense, with differential W.* That is, if $\theta_W(x) \stackrel{\mathrm{def}}{=} ||x-p||^{-1} \sup\Big\{||y-F(p)-W(x-p)|| : y \in F(x)\Big\}$, then $\lim_{x\to p, x\in K, x\neq p} \theta(x) = 0$. The converse is also true: *if F is differentiable at p along K in the ordinary sense, that is, if there is a linear map $W \in \mathbb{R}^{m\times n}$ such that $\theta_W(x) \to 0$, then $\{W\} \in SD(F,p)$.* ∎

Remark 7 Remark 3 shows that (i) the theory of derivate containers does not allow us to assign a generalized differential at p to a continuous map F which is differentiable at p in the ordinary sense but is not Lipschitz near p, (ii) even when F is both Lipschitz near p and differentiable at p, it may happen that $\{F'(p)\}$ is not a derivate container of F at p. For example, the function $f : \mathbb{R} \to \mathbb{R}$ given by $f(x) = x + 2x^2 \sin\frac{1}{x}$ is locally Lipschitz, and every derivate container $\mathbf{W}$ of f at 0 has to contain the set $[-1, 3]$. On the other hand, $\{1\} \in SD(f, 0)$, since $f'(0) = 1$ (cf. Remark 6). Therefore that *the concept of semidifferential is more general than that of a Warga derivate container, or of the Clarke generalized Jacobian, even for Lipschitz maps.* ∎

Theorem 3 (The Chain Rule for semidifferentials.) *Let $F^1 \in SVM(\mathbb{R}^{n_1}, \mathbb{R}^{n_2})$, $F^2 \in SVM(\mathbb{R}^{n_2}, \mathbb{R}^{n_3})$, have domains K^1, K^2. Assume that $F^1(K^1) \subseteq K^2$. Define $F = F^2 \circ F^1 \in SVM(\mathbb{R}^{n_1}, \mathbb{R}^{n_3})$ by*

$$x^3 \in (F^2 \circ F^1)(x^1) \Leftrightarrow (\exists x^2 \in F^1(x^1))(x^3 \in F^2(x^2)).$$

Let $p^1 \in K^1$, $\mathbf{W}^1 \in SD(F^1, p^1)$. Let $p^2 = F^1(p^1)$, $\mathbf{W}^2 \in SD(F^2, p^2)$. Then (a) F is semidifferentiable at p^1 and (b) if $\mathbf{W} = \mathbf{W}^2 \circ \mathbf{W}^1 = \{W^2 \circ W^1 : W^1 \in \mathbf{W}^1, W^2 \in \mathbf{W}^2\}$, then $\mathbf{W} \in SD(F, p^1)$. ∎

5. Approximating cones and separation theorems

If X, Y are topological spaces, $F \in SVM(X,Y)$, $p \in \mathrm{Dom}(F)$, F is single-valued at p, and $q = F(p)$, we call F *open at p* if $F(U)$ is a neighborhood of $q = F(p)$ in Y whenever $U \subseteq X$ is a neighborhood of p.

Definition 6 Let $S \subseteq \mathbb{R}^n$, $q \in S$, $v \in \mathbb{R}^n$. We say that S is a *v-directional semineighborhood of q* if there exists a convex set $C \subseteq Y$ such that $C \subseteq S$, $q \in C$, and $q + rv \in \mathrm{Int}(C)$ for some $r > 0$. ∎

It is clear that if $v = 0$ then a v-directional semineighborhood of q is just an ordinary neighborhood of q. A simple compactness argument shows that *S is a neighborhood of q if and only if S is a v-directional semineighborhood of q for every $v \in \mathbb{R}^n\backslash\{0\}$.*

Definition 7 If $F \in SVM(\mathbb{R}^n, \mathbb{R}^m)$, $p \in \mathbb{R}^n$, $F(p) = q$, and $v \in \mathbb{R}^m$, we say that F is *v-directionally open at p* if for every neighborhood V of p the image $F(V)$ is a v-directional semineighborhood of q. ∎

Clearly, F is open at p if and only if it is v-directionally open at p for every $v \in Y - \{0\}$.

Theorem 4 *Let $F \in SVM(\mathbb{R}^n, \mathbb{R}^m)$, $p \in \mathrm{Dom}(F)$, $v \in \mathbb{R}^m$. Assume that F is semidifferentiable at p and $\mathbf{W} \in SD(F,p)$. Let $C = T_p\mathrm{Dom}(F)$. Assume that*

$$v \in \bigcap \Big\{ \mathrm{Int}(WC) : W \in \mathbf{W} \Big\}.$$

Then F is v-directionally open at p. ∎

Definition 8 Let $S \subseteq \mathbb{R}^n$, $q \in S$. A *generalized approximating cone* (GAC) for S at q is a closed convex cone C in $\mathbb{R}^n$ 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 map $F_i \in SVM(\mathbb{R}^n, \mathbb{R}^n)$ whose domain is 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 is semidifferentiable at 0 with semidifferential $\{\mathrm{id}_{\mathbb{R}^n}\}$ (i.e. F_i is differentiable at 0 with differential $\mathrm{id}_{\mathbb{R}^n}$, cf. Remark 6). If the F_i can be taken to be single-valued (in which case they have to be continuous near 0) then C is an *approximating cone* for S at p. ∎

Theorem 5 *Let $F \in SVM(\mathbb{R}^n, \mathbb{R}^m)$, let $K = \mathrm{Dom}(F)$, and let $p \in K$ be such that F is single-valued at p. Let $q = F(p)$. Let $S \subseteq \mathbb{R}^m$ be such that $q \in S$. Assume that S has a GAC C at q such that C is not a linear subspace of $\mathbb{R}^m$, and F is semidifferentiable at p. Let $\mathbf{W} \in SD(F,p)$. Let $C_0 = T_p\mathrm{Dom}(F)$. Then, either $F(K) \cap S \neq \{q\}$, or*
$(*)$ there exists $(W, \lambda) \in \mathbf{W} \times (\mathbb{R}_m\backslash\{0\})$ such that $\lambda W c_0 \geq 0$ for all $c_0 \in C_0$ and $\lambda c \leq 0$ for all $c \in C$. ∎

6. Point Variations

If $X = \mathbb{R}^n$, a *k-parameter semidifferentiable* (resp. *differentiable*) *point variation* (abbr. SDPV, DPV) in X is a pair $\mathcal{V} = (x_*, V)$ such that (i) $x_* \in X$, (ii) $V \in SVM(X \times \mathbb{R}^k_+, X)$, (iii) $\mathrm{Dom}(V) \cap U = (X \times \mathbb{R}^k_+) \cap U$ for some neighborhood U of $(x_*, 0)$ in $X \times \mathbb{R}^k$, (iv) $V(x, 0) = x$ for $x \in X$, x near x_*, and (v) V is semidifferentiable (resp. differentiable) at $(x_*, 0)$. The point x_* is the *base point* of $\mathcal{V}$. If V is a SDPV, then V is a DPV iff there is a linear map $L : X \times \mathbb{R}^k$ —the *differential* of $\mathcal{V}$— such that $\sup\{||y - x_* - L(x - x_*, \alpha)|| : y \in V(x, \alpha)\} = o(||x - x_*|| + ||\alpha||)$ as $x \to x_*$, $\alpha \to 0$, $\alpha \in \mathbb{R}^k_+$. The map L is obviously unique, and has to be of the form $(z, \alpha) \to z + \alpha_1 v_1 + \ldots + \alpha_k v_k$. We write $\mathcal{DV} \stackrel{\mathrm{def}}{=} (v_1, \ldots, v_k)$.

If $\Sigma = (M, \mathcal{U}, F)$ is a control system on X, $\gamma_* = (\xi_*, \eta_*)$ is a controlled trajectory, $\mathrm{Dom}(\xi_*) = [a, b]$, and $t_* \in [a, b]$, then a *Σ-admissible k-parameter SDPV of γ_* at time t_** is a k-parameter SDPV $\mathcal{V} = (x_*, V)$ in X with base point $x_* = \xi_*(t_*)$ such that
(ADM) there exist functions $\tau_-, \tau, \tau_+ : (0, \infty) \to [a, b]$ such that $\tau_-(\varepsilon) \leq \tau(\varepsilon) \leq \tau_+(\varepsilon)$ for all ε and $\lim_{\varepsilon\to 0+} \tau_-(\varepsilon) = \lim_{\varepsilon\to 0+} \tau_+(\varepsilon) = t_*$, with the property that for every $\varepsilon > 0$ there exist a neighborhood $K(\varepsilon)$ of $(\xi_*(t_*), 0)$ in $M \times \mathbb{R}^k_+$ of such that

(I) *if* $\alpha \in \mathbb{R}_+^k$, *and* $\xi_1 : [\tau_-(\varepsilon), \tau(\varepsilon)] \to M$, $\xi_2 : [\tau(\varepsilon), \tau_+(\varepsilon)] \to M$ *are trajectories of* η_* *such that* $(\xi_1(\tau(\varepsilon)), \alpha) \in K(\varepsilon)$ *and* $\xi_2(\tau(\varepsilon)) \in V(\xi_1(\tau(\varepsilon)), \alpha)$, *then* $\xi_2(\tau_+(\varepsilon)) \in \mathcal{R}^{\Sigma}_{[\tau_-(\varepsilon), \tau_+(\varepsilon)]}(\xi_1(\tau_-(\varepsilon))$.

Definition 9 A *point variational vector* (PVV) for a controlled trajectory γ_* of a control system Σ is a pair (v, t_*) such that there is a Σ-admissible differentiable 1-parameter point variation $\mathcal{V}$ of γ_* at time t_* such that $\mathcal{DV} = v$. A set $\mathbf{v}$ of PVV'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 and a Σ-admissible differentiable k-parameter point variation $\mathcal{V}$ of γ_* at t_* such that, if $\mathcal{DV} = (v_1, \ldots, v_k)$, then $\{v : (v, t_*) \in \mathbf{w}\} \subseteq \{v_1, \ldots, v_k\}$. ▮

7. Outline of the proof of the MP

The simplest examples of PVV's are the "needle variational vectors," defined as follows. Let $\eta, \eta_* \in \mathcal{U}$ be 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 $x \in M$, $\alpha \geq 0$, define $V_{\eta,\eta_*,t_*}(x, \alpha)$ to be the set of all $y \in M$ 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, using standard techniques such as the Lusin and Scorza-Dragoni theorems, 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 $G(\eta, \eta_*)$ of $[a, b]$ such that $[a, b] \backslash G(\eta, \eta_*)$ has measure zero, with the property that $(x_*, V_{\eta,\eta_*,t_*})$ is a 1-parameter DPV for every $(x_*, t_*) \in M \times G(\eta, \eta_*)$, and $\mathcal{D}(x_*, V_{\eta,\eta_*,t_*}) = \{v_{x_*,t_*,\eta,\eta_*}\}$, where $v_{x_*,t_*,\eta,\eta_*} = F^\eta(x_*, t_*) - F^{\eta_*}(x_*, t_*)$. (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 SVAS's follows by regularizing F^η and F^{η_*}.)

Now assume that $\gamma_* = (\xi_*, \eta_*)$ is a controlled trajectory of Σ with domain $[a, b]$ such that F^{η_*} is LIL, and η is such that $[a, b] \subseteq \mathrm{TDom}(\eta)$. If we pick $t_* \in G(\eta, \eta_*)$ and let $x_* = \xi_*(t_*)$, then it is easy to show that the variation $(x_*, V_{\eta,\eta_*,t_*})$ is Σ-admissible. (It suffices to pick $\tau_-(\varepsilon) = \tau(\varepsilon) = t_*$, $\tau_+(\varepsilon) = t_* + \varepsilon$. Actually, at this point it is not necessary for F^{η_*} to be LIL, and all one needs is for F^{η_*} to have uniqueness of trajectories.) If we let $\mathbf{v}_{\eta,\gamma_*}$ be the set $\{v_{\xi_*(t),t,\eta,\eta_*} : t \in G(\eta, \eta_*)\}$, then $\mathbf{v}_{\eta,\gamma_*}$ is a set of PVV's for γ_*, which is obviously compatible since no two different members of $\mathbf{v}_{\eta,\gamma_*}$ correspond to the same time t_*.

Using these facts, it is easy to see that Theorem 1 will follow if we prove that, under the same assumptions, there exist A and c^* such that (ii), (iii) and (iv) hold. (That is, we can forget about the Hamiltonian minimization condition (i).) The proof is a rather simple compactness argument, spelled out in detail in [8]. Moreover, it suffices to assume that there exists a neighborhood Z of 0 in $\mathbb{R}^n$ and an $H_0 \in SVM(\mathbb{R}^n, \mathbb{R}^n)$ such that $\mathrm{Dom}(H_0) = Z$, $H_0(Z) \subseteq C_0$, and $\{\mathrm{id}_{\mathbb{R}^n}\} \in SD(H_0, 0)$.

To complete the proof of Theorem 1, we avoid notational complications and assume, for simplicity, that only two variation times are involved. That is, we assume that $\mathbf{v} = \mathbf{v}_1 \cup \mathbf{v}_2$, where $\mathbf{v}_i$ consists of k_i PVV's $(v_{i,j}, t_i)$, all of them having the same variation time t_i, and $t_1 < t_2$. Using the compatibility hypothesis, we find, for $i = 1, 2$, κ_i-parameter Σ-admissible DPV's $\mathcal{V}_i = (\xi_*(t_i), V_i)$ such that $\mathcal{DV}_i = (\tilde{v}_{i,1}, \ldots, \tilde{v}_{i,\kappa_i})$ and every $v_{i,j}$ is equal to $\tilde{v}_{i,\tilde{\jmath}}$ for some $\tilde{\jmath} = \tilde{\jmath}(i, j)$.

Let $Q \in SVM(Z \times \mathbb{R}_+^{\kappa_1} \times \mathbb{R}_+^{\kappa_2}, M)$ be given by

$$Q(z, \alpha) = g(\Phi_F^{b,t_2}(V_2(\Phi_F^{t_2,t_1}(V_1(\Phi_F^{t_1,a}(H_0(z)), \alpha^1)), \alpha^2)))$$

for $\alpha^i \in \mathbb{R}_+^{\kappa_i}$, $i = 1, 2$, $\alpha = (\alpha^1, \alpha^2)$. (That is, $y \in Q(z, \alpha)$ iff there exist z_1, z_2, z_3 such that $z_1 \in H_0(z)$, $z_2 \in V_1(\Phi_F^{t_1,a}(z_1), \alpha^1)$, $z_3 \in V_2(\Phi_F^{t_2,t_1}(z_2), \alpha^2)$, $y \in g(\Phi_F^{b,t_2}(z_3))$.) Let $\mathbf{W}_1$, $\mathbf{W}_2$, $\mathbf{W}_3$, be, respectively, derivate containers of $\Phi_F^{t_1,a}$ at p_*, of $\Phi_F^{t_2,t_1}$ at $\xi_*(t_1)$, and of Φ_F^{b,t_2} at $\xi_*(t_2)$. Then the chain rule implies that, if we write $\langle \alpha^i, \mathcal{DV}_i \rangle = \sum_{j=1}^{\kappa_i} \alpha_j^i \tilde{v}_{i,j}$, and define $\hat{\mathbf{W}}$ to be the set of all maps $\hat{\mathbf{W}}_{W_1,W_2,W_3,B} : (z, \alpha^1, \alpha^2) \to B(W_3(\langle \alpha^2, \mathcal{DV}_2 \rangle + W_2(\langle \alpha^1, \mathcal{DV}_1 \rangle + W_1 z)))$, for $W_i \in \mathbf{W}_i$, $i = 1, 2, 3$, then $\hat{\mathbf{W}} \in SD(Q, (0, 0))$. If we now choose the $\mathbf{W}_i$ to be the derivate containers given by Theorem 2, and apply Theorem 5, the conclusion follows.

8. Classical systems; separability

We now show how the MP applies to systems of the form $\dot{x} = f(x, t, u)$. We select a setting similar to that of F. Clarke's version of the MP, presented in [2], [3].

Recall that (a) the *Borel σ-algebra* $\mathcal{B}(X)$ of a metric space X is the σ-algebra generated by the open sets, and (b) a *Polish space* is a separable topological space X whose topology arises from a metric d such that (X, d) is complete. We use $\mathcal{L}(\mathbb{R}^n)$ to denote the σ-algebra of Lebesgue measurable subsets of $\mathbb{R}^n$, and write $\mathcal{L}(E) = \{X \in \mathcal{L}(\mathbb{R}^n) : X \subseteq E\}$ for $E \in \mathcal{L}(\mathbb{R}^n)$. We will use the following selection theorem (cf. Aumann [1], and Schwartz [7], Ch. I, §4, Lemma 7):

Proposition 1 *If P is a Polish space, and $A \subseteq P \times \mathbb{R}^n$ is $\mathcal{B}(P) \times \mathcal{L}(\mathbb{R}^n)$-measurable, then (a) the projection $\pi(A)$ of A on $\mathbb{R}^n$ is Lebesgue measurable, and (b) there exists a Lebesgue measurable selection of A over $\pi(A)$, i.e. a Lebesgue measurable map $\sigma : \pi(A) \to P$ such that $(\sigma(t), t) \in A$ for every $t \in \pi(A)$.* ▮

If P is a Polish space, and M is a smooth manifold, we define —using $T_x M$, TM to denote the tangent space of M at x and the tangent bundle of M— a *classical control dynamics* (abbr. CCD) *on M with ambient control space P* to be a pair (U, f) such that

(I) U is a SVM such that $\mathrm{Im}(U) \subseteq P$, $I = \mathrm{Dom}(U)$ is an interval, and U is $\mathcal{L}(I) \times \mathcal{B}(P)$-measurable as a subset of $I \times P$,

(II) $(x, t, u) \to f(x, t, u)$ is a map from $M \times U$ to TM, such that $f(x, t, u) \in T_x M$ for all $(x, t, u) \in M \times U$,

(III) every map $f(\cdot, t, u)$, $(t, u) \in U$, is continuous,

(IV) every map $f(x,\cdot,\cdot)$, $x \in M$, is $\mathcal{L}(I) \times \mathcal{B}(P)$-measurable.
(In particular, (I), (III) and (IV) imply that f is $\mathcal{B}(M) \times \mathcal{L}(I) \times \mathcal{B}(P)$-measurable.)

A *control* is a Lebesgue measurable map $\eta : J \to P$ such that J is a subinterval of I and $\eta(t) \in U(t)$ for each $t \in J$. Proposition 1 implies that every control can be extended to I.

If we let $\mathcal{U}$ denote the class of all controls, and write $F^\eta(x,t) = f(x,t,\eta(t))$, $F = \{F^\eta\}_{\eta\in\mathcal{U}}$, then $\Sigma = (M,\mathcal{U},F)$ is a control system. A Σ arising in this way from a CCD (U,f) will be called a *classical control system.* The crucial fact about such systems is that they are *separable*, in the following sense:

Definition 10 A control system $\Sigma = (M,\mathcal{U},F)$ is *separable along a curve* $\xi : [a,b] \to M$ if there exists a countable subset $\mathcal{U}_0$ of $\mathcal{U}$, such that (i) $[a,b] \subseteq \mathrm{TDom}(F^\eta)$ for every $\eta \in \mathcal{U}_0$, and (ii) for every $t \in [a,b]$, the set of values $\{F^\eta(\xi(t),t) : \eta \in \mathcal{U}_0\}$ is dense in $\{F^\eta(\xi(t),t) : \eta \in \mathcal{U}\ ,\ G(\xi) \subseteq \mathrm{Dom}(F^\eta)\}$. A control system is *separable* if it is separable along every curve. ▮

The following observation is then completely trivial:

Proposition 2 *For a separable system, weak and strong minimization are equivalent.* ▮

Theorem 6 *A classical control system is separable and has the measurable substitution property.*

PROOF. The substitution property is trivial. To prove separability, we fix a curve $\xi : J \to M$. After replacing M with a relatively compact neighborhood of ξ, if necessary, we assume that M is second countable. The map $U \ni (t,u) \to f(\xi(t),t,u) \in TM$ is $\mathcal{L}(I) \times \mathcal{B}(P)$-measurable. Let $\mathcal{B}$ be a countable family of open subsets of TM which is a basis of the topology of TM. Let $B \in \mathcal{B}$. Let $Q_B = \{(t,u) : u \in U(t), f(\xi(t),t,u) \in B\}$. Then Q_B is $\mathcal{L}(I) \times \mathcal{B}(P)$-measurable. Let $I_B = \{t : (\exists u)((t,u) \in Q_B)\}$. Then Proposition 1 implies that I_B is measurable and there is a measurable map $\tilde\eta_B : I_B \to P$ such that $(t,\tilde\eta_B(t)) \in Q_B$ for every $t \in I_B$. Let $\eta_B : I \to P$ be a control that extends $\tilde\eta_B$. Let $\mathcal{U}_0 = \{\eta_B : B \in \mathcal{B}\}$. For $t \in I$, let $V(t) = \{F^\eta(\xi(t),t) : \eta \in \mathcal{U}\}$, $V_0(t) = \{F^\eta(\xi(t),t) : \eta \in \mathcal{U}_0\}$. We have to show that $V_0(t)$ is dense in $V(t)$. Let $v \in V(t)$. Let $\eta \in \mathcal{U}$ be such that $v = F^\eta(\xi(t),t)$. Let Z be a neighborhood of v. Pick $B \in \mathcal{B}$ such that $v \in B$ and $B \subseteq Z$. Then $v = f(\xi(t),t,\eta(t))$, so $(t,\eta(t)) \in Q_B$, and then $t \in I_B$. Therefore $\eta_B(t) = \tilde\eta_B(t)$, and $(t,\tilde\eta_B(t)) \in Q_B$, so $f(\xi(t),t,\tilde\eta_B(t)) \in B$. Clearly, $f(\xi(t),t,\tilde\eta_B(t)) \in V_0(t)$. So $Z \cap V_0(t) \neq \emptyset$. So $V_0(t)$ is dense in $V(t)$. ▮

It follows from Theorem 6 that for a classical control system the MP applies, with the extra conclusion that the adjoint vector is *strongly minimizing.* (For the optimality results, we have to require that the augmented system Σ^L, obtained by adding the running cost as a new state variable, satisfy the same technical assumptions. That is, the Lagrangian $(x,t,u) \to L(x,t,u)$ has to be continuous in x for each (t,u), and $\mathcal{L}(I) \times \mathcal{B}(P)$-measurable in (t,u) for each x, exactly as in[3].)

For Theorem 1 to apply, the reference CVF $f_* : (x,t) \to f(x,t,u_*(t))$ has to be LIL. This is a weaker assumption than those of [3], Theorems 5.1.2 and 5.2.1, where (a) $f(x,t,u)$ has to be Lipschitz with respect to x for each t,u, with a Lipschitz constant $k(t,u)$, and (b) the function $t \to k(t,u_*(t))$ has to be integrable. Condition (b) is our requirement that f_* be LIL —and this hypothesis is essential, as can be shown by simple examples— but *Condition (a) is no longer needed.*

Remark 8 Another important example of a class where separability is always guaranteed is that of systems generated by a dynamics $\dot x = f(x,u)$, $x \in M$, $u \in U$, where M is a manifold, U is an arbitrary set, and $f(x,u)$ is continuous in x for each fixed u. (A *control* is now a function $t \to u(t)$ such that the map $(x,t) \to f(x,u(t))$ is a CVF.) The separability of such a system follows from the fact that the space of continuous vector fields on a second countable manifold is separable. ▮

References

[1] Aumann, R.J., "Measurable utility and the measurable choice theorem," *La Décision*, Actes Colloque Int. du CNRS, Aix-en-Provence (1967), 15-26.

[2] Clarke, F.H., "The Maximum Principle under minimal hypotheses," *S.I.A.M. J. Control Optim.* **14** (1976), 1078-1091.

[3] Clarke, F.H., *Optimization and Nonsmooth Analysis,* Wiley Interscience, New York, 1983.

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

[5] Lojasiewicz Jr., S., "Local controllability of parametrized differential equations," in preparation.

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

[7] Schwartz, J.T., *W-* algebras,* Gordon & Breach, New York (1967).

[8] Sussmann, H.J., "A strong version of the Lojasiewicz Maximum Principle," in *Optimal Control of Differential Equations,* Nicolai H. Pavel Ed., M. Dekker Inc., New York (1994).

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

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

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

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

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