This paper will be presented at the 37th IEEE Conference on Decision and Control, to be held in Tampa, FL, in December 1998, and will appear in the conference proceedings. The date of this version is August 10, 1998.

Transversality conditions and a strong maximum principle for systems of differential inclusions [1,2]

Héctor J. Sussmann
Department of Mathematics; Rutgers, the State University of New Jersey
Hill Center—Busch Campus, 110 Frelinghuysen Road, Piscataway, NJ 08854-8019, USA
E-mail: sussmann@hamilton.rutgers.edu Web site: http://www.math.rutgers.edu/~sussmann

1. Introduction

In a series of recent papers (cf. Sussmann [10, 11, 12, 13]), we presented a general version of the finite-dimensional Pontryagin Maximum Principle (PMP) that extends and brings together into a single unified framework the classical smooth version (cf. Pontryagin *et al.* [9], Berkovitz [3]) —including high-order necessary conditions for optimality, cf. Knobloch [7], Krener [8] — and F. Clarke's nonsmooth formulation (cf. Clarke [4, 5]), as well as nonsmooth versions for systems of vector fields (as in Kaskosz-Lojasiewicz [6]) and of differential inclusions (as in Tuan [16], Zhu [17]), including the "Lojasiewicz improvement" (i.e., for a control system $\dot{q} = f(q, u, t)$, the possibility of allowing the maps $q \to f(q, u, t)$ to be just continuous for general u, t, and only requiring Lipschitz continuity for the maps $q \to f(q, \eta_*(t), t)$, where η_* is the reference control), and many situations with non-Lipschitz reference vector fields and even discontinuous ones.

The purpose of this note is to announce a further extension of this result, also described in [15], in which the classical transversality condition is replaced by an improved form, involving "weakly approximating cones."

Our version of the PMP is similar in spirit to the classical statement of [9], though much more general, and is proved using the same strategy of (a) reducing the optimal control problem to a geometric separation problem, (b) constructing "needle variations," (c) using a topological argument, based on a result closely related to the Brouwer fixed-point theorem, to construct an "adjoint vector" that satisfies a finite subset of the collection of inequalities that occur in the Hamiltonian maximization condition of the PMP, and (d) concluding with a compactness argument to end up with an adjoint vector that satisfies all the inequalities of the Hamiltonian maximization condition.

Remark 1 Work on a version of the PMP that incorporates state space constraints, using similar techniques, is now in progress. However, the proof of this version is not yet complete as of this writing (August 1998), so the discussion here will be limited to the problem without state space constraints. □

[1] Research supported in part by NSF Grant DMS95-00798 and AFOSR Grant 0923.

[2] Most of this work was done while the author was a visitor at the Centre Émile Borel of the Institut Henri Poincaré (IHP), in Paris, France. The author is very grateful for the hospitality provided by the IHP.

2. Notations and definitions

If A, B are sets, we use $SVM(A, B)$ to denote the set of all set-valued maps from A to B, so $F \in SVM(A, B)$ iff F is an ordinary single-valued map from A to the set 2^B of subsets of B. We also write $F : A \twoheadrightarrow B$ to indicate that $F \in SVM(A, B)$. If $F : A \twoheadrightarrow B$, the *graph* of F is the set $Gr(F) = \{(x, y) : y \in F(x)\}$. All linear spaces are over $\mathbb{R}$, and no other field of scalars will be considered. If V, W are linear spaces, we use $L(V, W)$ to denote the set of linear maps from V to W, and write $V^* \stackrel{\text{def}}{=} L(V, \mathbb{R})$, so V^* is the dual of V. If V is normed, then $\bar{B}_V$ denotes the closed unit ball of V. The abbreviation "FDRLS" stands for "finite-dimensional real linear space." If X is a linear space, then a *cone* in X is a nonempty subset of X which is closed under multiplication by nonnegative scalars. If C is a cone in X, then the *polar* of C is the set $C^\dagger$ of all $w \in X^*$ such that $\langle w, v\rangle \le 0$ for all $v \in S$.

2.1. Reduction to a separation problem

It is well known that necessary conditions for a reference trajectory-control pair (ξ_*, η_*) to be a solution of a Bolza optimal control problem

$$\begin{cases} \text{minimize} & \varphi(\xi(a), \xi(b)) + \int_a^b L(\xi(t), \eta(t), t)\, dt, \\ \text{subject to} & \dot{\xi}(t) = f(\xi(t), \eta(t), t) \text{ for a.e. } t\,, \\ \text{and} & (\xi(a), \xi(b)) \in S\,, \end{cases}$$

with state variable $q \in Q$ and control variable $u \in U$ —where U is a general set, and Q is an open subset of a FDRLS X or, more generally, a smooth manifold— are special cases of necessary conditions for the *separation property*:

$$\mathcal{R}^{\Sigma}_{[a,b]}(\bar{q}) \cap S = \{\hat{q}\}\,, \tag{1}$$

where $\bar{q} = \xi_*(a)$, $\hat{q} = \xi_*(b)$, S is some given subset of the state space Q, and the *reachable set* $\mathcal{R}^{\Sigma}_{[a,b]}(\bar{q})$, for a system Σ, is defined by

$$\mathcal{R}^{\Sigma}_{[a,b]}(\bar{q}) \stackrel{\text{def}}{=} \{\xi(b) : \xi \in \mathrm{TRAJ}(\Sigma)\,, \xi(a) = \bar{q}\}\,, \tag{2}$$

where $\mathrm{TRAJ}(\Sigma)$ is the set of all trajectories of Σ.

The reduction of the optimal control problem to the separation property is well known, and dates back to the book [9]. It is carried out in two steps. First, we transform the endpoint condition $(\xi(a), \xi(b)) \in S$ into one of the form $\xi(a) = \bar{q}$, $\xi(b) \in \hat{S}$, i.e. a restriction in which the initial state is fixed and a set $\hat{S}$ of possible terminal states is specified. This is done by considering a new system with state space $Q \times Q$, whose state consists of two copies (q_1, q_2) of the state q. On the interval $[a, b]$, the first copy is required to obey the law $\dot{q}_1 = f(q_1, u, t)$, while the second one satisfies $\dot{q}_2 = 0$. We supplement this by adding the requirements that $\dot{q}_1 = \sum_{i=1}^{\nu} v^i X_i(q_1)$ and $\dot{q}_2 = \sum_{i=1}^{\nu} v^i X_i(q_2)$, on the interval $[a-1, a\,[$, where $(X_1, \ldots, X_\nu)$ is a finite sequence of vector fields on Q such that $X_1(q), \ldots, X_\nu(q)$ span the tangent space T_qQ for every $q \in Q$, and v is a control taking completely arbitrary values in $\mathbb{R}^\nu$. In this way, if we start the new system at time $a-1$ at a point $(\bar{q}_0, \bar{q}_0)$, then we can produce an arbitrary point $(\bar{q}, \bar{q})$ at time a (assuming, as we clearly may without loss of generality, that Q is connected), and then the condition $(\xi(a), \xi(b)) \in S$ is exactly equivalent to $\Xi(b) \in S$, where $\Xi : [a-1, b] \to Q \times Q$ is a trajectory of the new system given by $\Xi(t) = (\xi(t), \xi(a))$ for $a \le t \le b$, and $\Xi(t) = (\gamma(t), \gamma(t))$ for $a-1 \le t \le a$, with the curve $\gamma : [a-1, a] \to Q$ chosen so that $\gamma(a-1) = \bar{q}_0$ and $\gamma(a) = \xi(a)$.

Next, for a Bolza problem whose endpoint condition is of the form $\xi(a) = \bar{q}$, $\xi(b) \in S$, we may clearly assume that the first term of the cost functional is of the form $\varphi(\xi(b))$. We then introduce a new "augmented control system" $\Sigma^\#$, with state space $Q^\# = \mathbb{R} \times Q$ and state variable $q^\# = (q^0, q)$, where q obeys the law $\dot{q} = f(q, u, t)$, and the new scalar variable q^0 —the "running cost"— is subject to the dynamical equation $\dot{q}^0 = f^0(q, u, t)$, where $f^0 \stackrel{\text{def}}{=} L$. If the trajectory-control pair (ξ, η) is a solution of our original optimal control problem, and c is its cost, then the pair $(\xi^\#, \eta)$ —where $\xi^\#$ is the curve in $\mathbb{R} \times Q$ given by $\xi^\#(t) = (\xi^0(t), \xi(t))$ for $a \le t \le b$, and $\xi^0(t) = \int_a^t f^0(\xi(s), \eta(s), s)\, ds$— satisfies the separation condition $\mathcal{R}^{\Sigma^\#}_{[a,b]}(\bar{q}^\#) \cap S^\# = \{\hat{q}^\#\}$, where $\bar{q}^\# = (0, \bar{q})$, $\hat{q}^\# = (c - \varphi(\xi(b)), \xi(b))$, and

$$S^\# = \left\{(q^0, q) : q \in S\,,\ q^0 \le c - \varphi(q) - \|q - \xi(b)\|^2\right\}.$$

So a necessary condition for separation, applied to this situation, yields a necessary condition for optimality for the Bolza problem. (These reductions are explained in more detail, for example, in [15], where we only consider Lagrange problems, i.e. we take $\varphi = 0$.)

The separation problem is in fact more general, since it also includes as special cases:

(I) the question of *necessary conditions for failure of local controllability* along a reference trajectory, i.e., equivalently, the problem of *sufficient conditions for local controllability* along a reference trajectory,

(II) the Pareto optimality problem.

Indeed, the local controllability problem is a special case of the separation question, since, if a system is not locally controllable along a reference trajectory $\xi_* : [a, b] \to Q$, then there exists a sequence $\{q_j\}$ of points such that $q_j \to \xi_*(b)$ and the q_j are not reachable from $\xi_*(a)$. By passing to a subsequence, one may assume that $\{q_j\}$ has a limiting direction v, and then our version of the separation condition will yield a nontrivial adjoint vector ψ along ξ_*, as explained below.

As for the Pareto optimality question, it is identical to the optimal control problem, except that instead of one criterion functional J there is a finite collection $\{J^1, \ldots, J^m\}$ of such functionals, given by

$$J^i = \varphi^i(\xi(a), \xi(b)) + \int_a^b L^i(\xi(t), \eta(t), t)\, dt\,, \quad (3)$$

and the goal is to find a necessary condition for a trajectory $\xi_* \in \mathrm{TRAJ}(\Sigma)$ and corresponding control η_* to have the property that, whenever (ξ, η) is a trajectory-control pair for which $\xi \in \mathrm{TRAJ}(\Sigma)$ and $(\xi(a), \xi(b)) \in S$, it follows that the vector

$$\mathcal{J}(\xi, \eta) \stackrel{\text{def}}{=} (J^1(\xi, \eta), \ldots, J^m(\xi, \eta))$$

is such that $\mathcal{J}(\xi, \eta) - \mathcal{J}(\xi_*, \eta_*)$ does not belong to the closed negative orthant $]-\infty, 0]^m$ unless $\mathcal{J}(\xi, \eta)$ is equal to $\mathcal{J}(\xi_*, \eta_*)$. This problem is transformed into one about separation of sets exactly as the optimal control problem was, by first reducing it to the case of an endpoint constraint $\xi(a) = \bar{q}$, $\xi(b) \in S$ (so that the expressions $\varphi^i(\xi(a), \xi(b))$ in (3) can be replaced by $\varphi^i(\xi(b))$), and then using a "running cost" m-vector $q^0 = (q^{0,1}, \ldots, q^{0,m})$, whose components obey equations $\dot{q}^{0,i} = L^i(q, u, t)$, taking $Q^\# = \mathbb{R}^m \times Q$, with state variable $q^\# = (q^0, q)$, and letting

$$S^\# = \left\{(q^0, q) : q \in S\,,\ q^0 \le c^0 - \varphi(q) - \|q - \xi(b)\|^2 \mathbf{1}_m\right\},$$

where $\varphi(q) \stackrel{\text{def}}{=} (\varphi^1(q), \ldots, \varphi^m(q))^\dagger$, $\mathbf{1}_m$ is the column m-vector $(1, \ldots, 1)^\dagger$, c^0 is the vector $(c^{0,1}, \ldots, c^{0,m})^\dagger$, the $c^{0,i}$ are defined by $c^{0,i} = J^i(\xi_*, \eta_*)$, and, for two m-vectors $v = (v^1, \ldots, v^m)^\dagger$, $w = (w^1, \ldots, w^m)^\dagger$, the inequality $v \le w$ stands for "$v^i \le w^i$ for $i = 1, \ldots, m$."

2.2. Weakly approximating cones

For reasons that are explained in detail in [14] and [15], the nontriviality condition for the separation problem holds *if the cone C is not a linear subspace*. So, if we want the local controllability problem to be included in our general Maximum Principle, then we need a theory of "tangent cones" in which the half-line $\{rv : r \ge 0\}$ can be treated as a "tangent cone" at s to a set S consisting of the point s together with a sequence $\{s_j\}$ that converges to s and has limiting direction v. The definition of "weakly approximating cone" meets this requirement, as we now explain.

Definition 1 Assume that Y is a FDRLS, $S \subseteq Y$, and $s \in S$. Let C be a closed convex cone in Y. We say that C is a *Boltyanskii approximating cone* to S at s if there exist a neighborhood U of 0 in Y and a continuous map $F : C \cap U \to S$ such that $F(v) = s + v + o(\|v\|)$ as $v \to 0$ via values in C. $\square$

Definition 2 If S is a closed subset of a FDRLS X, $s \in S$, and S^* is a closed subset of S, we say that S^* *separates s from its complement in S* if the following condition holds:

(#) *for every neighborhood U of s in X there is a neighborhood V of s in X such that, if $K \subseteq S$ is connected, $K \cap V \neq \emptyset$ and $K \cap (S \backslash U) \neq \emptyset$, then it follows that $K \cap S^* \neq \emptyset$.* □

(In simpler terms, (#) says, roughly, that one cannot go in S from points close to s to points far from s without crossing S^*.)

Definition 3 Suppose Y is a FDRLS and $s \in S \subseteq Y$. Let C be a closed convex cone in Y. We say that C is a *weakly approximating cone* to S at s if there exist a compact neighborhood U of 0 in Y, a closed subset C^* of C that separates 0 from its complement in C, and a continuous map $F : C^* \cap U \to Y$ such that

$$F(v) = s + v + o(\|v\|) \text{ as } v \to_{C^*} 0\,, \tag{4}$$

and $F(C^* \cap U) \subseteq S$. □

Remark 2 In the version of Definition 3 given in [14], the map F is required to be defined and continuous on a full neighborhood of 0 in C, and the differentiability condition (4) is required to hold as $v \to_C 0$, while F has to map $C^* \cap U$ into S. This form of the definition is equivalent to the one given here because, if F is as in our Definition 3, then the Tietze extension theorem, applied to the continuous map $G : C^* \cap U \to Y$ that sends v to $\frac{F(v)-s-v}{\|v\|}$ if $v \neq 0$, and maps 0 to 0, yields a continuous extension $\tilde{G}$ of G to C, and then, if we define $\tilde{F}(v) = s + v + \|v\|.\tilde{G}(v)$, it is easy to see that $\tilde{F}$ satisfies the conditions of the definition of [14]. □

2.3. The improved transversality condition

As usual, for a reference trajectory ξ_* defined on an interval $[a, b]$, the conclusion of the maximum principle —in all its finite-dimensional versions— asserts the existence of an adjoint vector ψ, also defined on $[a, b]$, having certain properties, one of which is the "transversality condition," relating the terminal value $\psi(b)$ of ψ to the terminal set S. Our stronger form of the transversality condition says that $\boxed{\psi(b) \neq 0 \text{ and } -\psi(b) \in C^\dagger}$, where C is a *weakly approximating cone* to the terminal set S. This makes the result applicable in situations where other notions of a tangent or normal cone —such as the Clarke or Mordukhovich normal cones— would yield a weaker conclusion.

For example, *if S is a set consisting of a sequence $\{q_j\}$ of points converging to a limit $\hat{q}$, and having a limiting direction v at $\hat{q}$, then our version of the maximum principle includes the usual necessary conditions for local controllability along a reference trajectory directly, without the need to resort to a special argument.* This is possible because the concept of a "weakly approximating cone" is such that, if S is as in our case, then the half-line $\{r.v : r \geq 0\}$ is a weakly approximating cone to S in the sense of our theory. On the other hand, the Mordukhovich normal cone to S at $\hat{q}$ is the whole space, so any version of the transversality condition involving this cone —or the even larger Clarke normal cone— will not yield a nontrivial adjoint vector.

2.4. Multidifferentials

In this subsection, X and Y are normed FDRLS's and $Z = L(X, Y)$.

If $F : X \twoheadrightarrow Y$, the concept of a *multidifferential* of F at a point $(x, y) \in X \times Y$ in the direction of a closed convex cone C in X was defined in [14]. We now review the definition, in a slightly more general setting.

Recall that, if S is a closed subset of X, a *Whitney map of class C^1* from S to Y is a pair (h, H), where h, H are continuous maps from S to Y and to Z, respectively, such that

$$\lim_{x' \to x, x' \in S, x' \neq x} \left(\frac{h(x') - h(x) - H(x).(x' - x)}{\|x' - x\|} \right) = 0$$

for all $x \in S$.

The *Whitney extension theorem* says that, *if h, H are maps from a closed subset S of X to Y and Z, respectively, then there exists a map $\tilde{h} : X \to Y$ of class C^1 such that $\tilde{h}(x) = h(x)$ and $D\tilde{h}(x) = H(x)$ for all $x \in S$ if and only iff (h, H) is a Whitney map of class C^1 from S to Y.*

If $\mathbf{D} \subseteq Z$, we define $C^1_{\mathbf{D}}(S, Y)$ to be the set of all Whitney maps (h, H) of class C^1 from S to Y such that $H(x) \in \mathbf{D}$ for all $x \in S$.

If K is a compact metric space and M is a metric space, a *regular set-valued map* from K to M is a set-valued map $F : K \twoheadrightarrow M$ such that there exists a sequence $\{F_j\}_{j=1}^{\infty}$ of continuous single-valued maps $F_j : K \to M$ with the property that

$$\lim_{j \to \infty} \sup \Big\{ \text{dist}\Big(q, Gr(F)\Big) : q \in Gr(F_j) \Big\} = 0\,. \tag{5}$$

We use $\text{REG}(K, M)$ to denote the set of all regular set-valued maps from K to M. (Regular maps are studied in detail in Sussmann [14].)

Definition 4 Suppose $F : X \twoheadrightarrow Y$, $(x_*, y_*) \in X \times Y$, and $\mathbf{D}$ is a compact subset of Z. Let S be a closed subset of X such that $x_* \in S$. We say that $\mathbf{D}$ is a *multidifferential of F at (x_*, y_*) in the direction of S*, and write $\mathbf{D} \in MD_S(F; x_*, y_*)$, if for every neighborhood $\mathbf{D}'$ of $\mathbf{D}$ in Z there exists a pair (ρ, Θ) such that $\rho > 0$, Θ is a function on the interval $]\,0, \rho]$ with values in $[0, \infty[$ and such that $\lim_{s \downarrow 0} \Theta(s) = 0$, and

(MD) for every $\varepsilon > 0$ there exist f, h, H such that

- (MD.1) $f \in \text{REG}(S \cap (x_* + \rho \bar{B}_X), Y)$;
- (MD.2) $Gr(f) \subseteq Gr(F)$;
- (MD.3) $(h, H) \in C^1_{\mathbf{D}'}(S \cap (x_* + \rho \bar{B}_X), Y)$;
- (MD.4) $h(x_*) = y_*$;
- (MD.5) the inequality

$$\sup \Big\{ \|y - h(x)\| : y \in f(x) \Big\} \leq \varepsilon + \Theta(\|x - x_*\|)\|x - x_*\| \tag{6}$$

holds whenever $x \in S \cap (x_* + \rho \bar{B}_X)$. □

2.5. The transversal intersection theorem

It has been shown in our series of papers [10, 11, 12, 13, 14, 15] that general necessary conditions for separation —and corresponding versions of the maximum principle— can be derived from suitable *transversal intersection theorems*, saying, roughly, that if two set-valued maps $F_i : X_i \to\!\!\!\!\to Y$ —where X_1, X_2, Y are FDRLS's— are such that each F_i is "differentiable at $(0,0)$ in the direction of a closed convex cone C_i, with differential $\mathbf{D}_i$," the "cones" $\mathbf{D}_1 C_1$ and $\mathbf{D}_2 C_2$ are "strongly transversal," and C_2^* is a closed subset of C_2 that separates 0 from its complement in C_2, then the set $F_1(C_1) \cap F_2(C_2^*)$ contains a sequence $\{y_j\}$ of points such that $y_j \neq 0$ but $\lim_{j\to\infty} y_j = 0$. The main result of [14] is precisely, one such separation theorem, in which the concept of "differentiability" is the one provided by the theory of multidifferentials. The following is a special case of the general result of [14]:

Theorem 1 *Let X, Y be FDRLS's, and let C_1, C_2 be closed convex cones in X, Y. Let $y_* \in Y$, $F : C_1 \to\!\!\!\!\to Y$, $\mathbf{D} \subseteq L(X,Y)$, be such that $\mathbf{D} \in MD_{C_1}(F; 0, y_*)$. Let S be a subset of Y having C_2 as a weakly approximating cone at y_*. Suppose that C_2 is not a linear susbpace of Y. Then either (a) there exist a nontrivial linear functional w on Y and a $D \in \mathbf{D}$ such that $-w \in C_2^\dagger$ and $wD \in C_1^\dagger$, or (b) the intersection $F(C_1) \cap S$ contains a sequence $\{y_j\}$ of points such that $y_j \neq y_*$ and $y_j \to y_*$.* □

2.6. ALSC set-valued maps

If X and Y are metric spaces and $F : X \to\!\!\!\!\to Y$, then F is *lower semicontinuous* (LSC) if $\{x : F(x) \cap \Omega \neq \emptyset\}$ is an open subset of X whenever Ω is open in Y.

Definition 5 If X, Y are metric spaces, $I \subseteq \mathbb{R}$ is an interval, and $F : X \times I \to\!\!\!\!\to Y$, then F is said to be *almost lower semicontinuous* (ALSC) if for every $\varepsilon > 0$ there exists a closed subset J of I such that $\text{meas}(I \backslash J) \leq \varepsilon$ and the restriction F_J of F to $X \times J$ is LSC. □

2.7. The pseudolipschitz condition

Suppose that X is a normed FDRLS, $a, b \in \mathbb{R}$, $a < b$, $F : X \times [a,b] \to\!\!\!\!\to X$, and $\xi : [a,b] \to X$ is a trajectory of F (i.e. ξ is absolutely continuous and $\dot{\xi}(t) \in F(\xi(t), t)$ for almost all $t \in [a,b]$). We say that F is *integrably pseudolipschitz along* ξ if there exist an $\varepsilon > 0$ and an integrable function $[a,b] \ni t \to C(t) \in [0, \infty]$ such that $F(x,t) \cap (\dot{\xi}(t) + \varepsilon \bar{B}_X)$ is closed whenever $\|x - \xi(t)\| \leq \varepsilon$ and $F(x,t) \cap (\dot{\xi}(t) + \varepsilon \bar{B}_X) \subseteq F(x',t) + C(t).\|x - x'\|.\bar{B}_X$ whenever $x, x' \in \xi(t) + \varepsilon \bar{B}_X$ and $a \leq t \leq b$.

2.8. Generalized Jacobians

If X, Y are FDRLS's, $\Omega \subseteq X$ is open, $G : \Omega \to Y$ is a Lipschitz-continuous map, and $x \in \Omega$, we define $\partial G(x)$ —the "Clarke generalized Jacobian of G at x"— to be the closed convex hull of the set of all $D \in L(X,Y)$ such that $D = \lim_{j\to\infty} D_j$ for some sequence $\{D_j\}_{j=1}^\infty$ for which there are points $x_j \in \Omega$ converging to x and such that G is differentiable at each x_j and $D_j = DG(x_j)$.

2.9. Linear multiflows

A *linear multiflow* on a FDRLS Q, over a time interval I, is a family $\Lambda = \{\Lambda_{t,s} : s, t \in I, s \leq t\}$ such that

a. *$\Lambda_{t,s}$ is a nonempty compact set of linear maps from Q to Q whenever $s, t \in I$ and $s \leq t$,*

b. *$\Lambda_{t,t} = \{\mathrm{id}_Q\}$ for every $t \in I$,*

c. *the identity $\Lambda_{t_3,t_2} \circ \Lambda_{t_2,t_1} = \Lambda_{t_3,t_1}$ holds whenever $t_1 \leq t_2 \leq t_3$ and the t_i are in I.*

We use $LMFLOW(Q, I)$ to denote the set of all linear multiflows on Q over the time interval I.

If $\Lambda \in LMFLOW(Q, I)$, a *compatible selection* of Λ is a family $L = \{L_{t,s} : s, t \in I, s \leq t\}$ such that $L_{t,s} \in \Lambda_{t,s}$ whenever $s, t \in I$ and $s \leq t$, and $L_{t_3,t_2} L_{t_2,t_1} = L_{t_3,t_1}$ whenever $t_1 \leq t_2 \leq t_3$ and the t_i are in I. We use $CS(\Lambda)$ to denote the set of all compatible selections of Λ.

2.10. Flows and their multidifferentials

A *flow* on a set Q, over a time interval I, is a family $\Phi = \{\Phi_{t,s} : s, t \in I, s \leq t\}$ such that

a. *$\Phi_{t,s} : Q \to\!\!\!\!\to Q$ whenever $s, t \in I$ and $s \leq t$,*

b. *$\Phi_{t,t} = \mathrm{id}_Q$ for every $t \in I$,*

c. *the identity $\Phi_{t_3,t_2} \circ \Phi_{t_2,t_1} = \Phi_{t_3,t_1}$ holds whenever $t_1 \leq t_2 \leq t_3$ and the t_i are in I.*

We use $FLOW(Q, I)$ to denote the set of all flows on Q over the time interval I.

If $\Phi \in FLOW(Q, I)$, then a *trajectory* of Φ is a map $\xi : I \to Q$ such that $\xi(t) \in \Phi_{t,s}(\xi(s))$ whenever $s, t \in I$ and $s \leq t$. We use $\mathrm{TRAJ}(\Phi)$ to denote the set of all trajectories of Φ.

If Q is a FDRLS, $\Phi \in FLOW(Q, I)$, $\xi \in \mathrm{TRAJ}(\Phi)$ and $\Lambda \in LMFLOW(Q, I)$, we say that Λ is a *multidifferential of Φ along ξ*, and write $\Lambda \in MD(\Phi; \xi)$, if $\Lambda_{t,s} \in MD_Q(\Phi_{t,s}; \xi(s), \xi(t))$ whenever $s, t \in I$ and $s \leq t$.

2.11. Variations of maps and flows

If Q is a set, $k \in \mathbb{N}$, and $F : Q \to\!\!\!\!\to Q$, a *k-parameter variation* of F is a map $\nu : Q \times [0, +\infty[^k \to\!\!\!\!\to Q$ with the property that $\nu(q, 0) = F(q)$ for all $q \in Q$. If I is an interval, $a, b \in I$, $a \leq b$, and $\Phi \in FLOW(Q, I)$, then a *k-parameter variation of Φ carried by the interval* $[a,b]$ is a variation of $\Phi_{b,a}$. We write $VAR^k(\Phi; a, b)$ to denote the set of all k-parameter variations of Φ carried by $[a,b]$.

Given a FDRLS Q, a 1-*parameter impulse infinitesimal variation (IIV) in* Q is just a member of $Q \times \mathbb{R}$, i.e. a pair $V = (v, t)$ consisting of a vector $v \in Q$ and a time $t \in \mathbb{R}$. If $k \in \mathbb{N}$, then a *k-parameter IIV in* Q is a k-tuple $V = (V_1, \ldots, V_k)$ of 1-parameter IIV's in Q. If $V_i = (v_i, t_i)$, we write $\alpha(V) = \min(t_1, \ldots, t_k)$, $\beta(V) = \max(t_1, \ldots, t_k)$. If $a \in I$, $b \in I$ are such that $a \leq \alpha(V)$ and $\beta(V) \leq b$, we say that V is *carried by* $[a,b]$. We use $IIV^k(Q)$ to denote the set of all k-parameter IIV's in Q, and $IIV^k(Q; a, b)$ to denote the set of all $V \in IIV^k(Q)$ that are carried by the interval $[a,b]$.

If $\Phi \in FLOW(Q,I)$, we will also refer to the members of $IIV^k(Q)$ as *k-parameter IIV's of* Φ, and we will write $IIV^k(\Phi)$, $IIV^k(\Phi;a,b)$ instead of $IIV^k(Q)$, $IIV^k(Q;a,b)$.

If $\Lambda \in LMFLOW(Q,I)$, $V=(V_1,\ldots,V_k)$ belongs to $IIV^k(Q;a,b)$, and $V_i=(v_i,t_i)$ for $i=1,\ldots,k$, we use $\mathcal{E}(V,\Lambda,a,b)$ to denote the set of all linear maps $E(V,L,a,b): Q\times\mathbb{R}^k \to Q$ given by

$$E(V,L,a,b)(y,\rho_1,\ldots,\rho_k)=L_{b,a}y+\sum_{i=1}^{k}\rho_i L_{b,t_i}v_i, \qquad (7)$$

as L ranges over $CS(\Lambda)$.

We should think of each V_i as an "infinitesimal variation $(q,\varepsilon)\to q+\varepsilon v_i$ of the state at time t_i." Then this variation produces an "infinitesimal variation" of the map $q\to\Phi_{b,a}(q)=(\Phi_{b,t_i}\circ\Phi_{t_i,a})(q)$, in the form of the map $(\delta q,\delta\varepsilon)\to L_{b,a}(q)\delta q+\delta\varepsilon.L_{b,t_i}v_i$, which is in fact a *set* of linear maps, since L is not unique. Formula (7) gives the combined infinitesimal effect of the V_i on the map $\Phi_{b,a}$.

With the above interpretation, it is natural to regard the set $\mathcal{E}(V,\Lambda,a,b)\subseteq L(Q\times\mathbb{R}^k,Q)$ as a reasonable candidate for a multidifferential of a true variation $\nu\in VAR^k(\Phi;a,b)$. The following definition makes this precise:

Definition 6 Suppose Q is a FDRLS, I is an interval, $\Phi\in FLOW(Q,I)$, $\xi\in\mathrm{TRAJ}(\Phi)$, $\Lambda\in MD(\Phi;\xi)$, $k\in\mathbb{N}$, $\nu\in VAR^k(\Phi;a,b)$ and $V=(V_1,\ldots,V_k)\in IIV^k(\Phi;a,b)$. We say that V is *generated by* $(\nu,\Phi,\xi,\Lambda,a,b)$ if $\mathcal{E}(V,\Lambda,a,b)\in MD_{Q\times[0,+\infty[^k}(\nu;(\xi(a),0),\xi(b))$. □

2.12. DI control systems

A *DI control system* on a FDRLS Q, with time-interval I, is an indexed family $\mathcal{F}=\{F_u\}_{u\in U}$ of members F_u of $SVM(Q\times I,Q)$. (The abbreviation "DI" stands for "differential inclusion.")

If $\mathcal{E}$ is a class of subsets of I, we say that $\mathcal{F}$ has the *substitution property with respect to* $\mathcal{E}$ if, whenever $u_1,u_2\in U$, and $E\in\mathcal{E}$, it follows that there exists a $u_3\in U$ such that, for every $q\in Q$, $F_{u_3}(q,t)=F_{u_1}(q,t)$ if $t\in E$, and $F_{u_3}(q,t)=F_{u_2}(q,t)$ if $t\notin E$.

We say that $\mathcal{F}$ is *closed under interval* (resp. *measurable*) *substitutions* if $\mathcal{F}$ has the substitution property with respect to $\mathcal{E}$, where $\mathcal{E}$ is the class of all subintervals (resp. the class of all measurable subsets) of I.

2.13. Point variations of DI control systems

Let $\mathcal{F}=\{F_u\}_{u\in U}$ be a DI control system with state space Q and time-interval I, and let $\Phi\in FLOW(Q,I)$. We say that Φ is $\mathcal{F}$*-admissible* if $\Phi_{b,a}(q)\subseteq\mathcal{R}^{\mathcal{F},0}_{[a,b]}(q)$ (cf. Formula (2)) whenever $q\in Q$, $a,b\in I$, $a\le b$.

If Φ is $\mathcal{F}$-admissible, $a,b\in I$, $a\le b$, $k\in\mathbb{N}$, and ν belongs to $VAR^k(\Phi;a,b)$, we call ν $\mathcal{F}$*-admissible over the interval* $[a,b]$ if $\nu(q,r_1,\ldots,r_k)\subseteq\mathcal{R}^{\mathcal{F}}_{[a,b]}(q)$ for all $q,r_1,\ldots,r_k$ such that $q\in Q$ and $0\le r_i<+\infty$ for $i=1,\ldots,k$.

Now, let $\mathcal{V}$ be a set of one-parameter IIV's of Q (i.e. a subset of $Q\times\mathbb{R}$). Suppose that $\mathcal{F}$ is a DI control system on Q with time-interval I, $\Phi\in FLOW(Q,I)$, $\xi\in\mathrm{TRAJ}(\Phi)$, and $\Lambda\in MD(\Phi;\xi)$. We say that $\mathcal{V}$ is *$\mathcal{F}$-admissible as a set of infinitesimal point variations along ξ, relative to Φ and Λ*, if

(IA) *Whenever $k\in\mathbb{N}$, $t\in I$, and $V=(V_1,\ldots,V_k)$ belongs to $\Big(\mathcal{V}\cap(Q\times\{t\})\Big)^k$, there exists sequences $\{V^j\}$ in $IIV^k(Q)$, $\{\nu^j\}$ in $VAR^k(\Phi)$, $\{a^j\}$, $\{b^j\}$ in I, such that $a^j\le b^j$, $V^j\in IIV^k(Q;a^j,b^j)$, $V^j\to V$ (in $(Q\times\mathbb{R})^k$), V^j is generated by $(\nu^j,\Phi,\xi,\Lambda,a^j,b^j)$, ν^j is $\mathcal{F}$-admissible over $[a^j,b^j]$, $a^j\to t$ and $b^j\to t$.*

2.14. Multidifferentials of Lipschitz flows

Let Q be a normed FDRLS, and let $\xi:[a,b]\to Q$ be a curve. If $\varepsilon>0$, we use $\mathcal{T}(\xi,\varepsilon)$ to denote the tube

$$\mathcal{T}(\xi,\varepsilon)\stackrel{\mathrm{def}}{=}\{(q,t): a\le t\le b\,,\ \|q-\xi(t)\|\le\varepsilon\}\,. \qquad (8)$$

A time-varying vector field $Q\times\mathbb{R}\ni(q,t)\to f(q,t)\in Q$ is said to be *integrably Lipschitz near* ξ if there exist $\varepsilon>0$ and an integrable function $k:[a,b]\to[0,+\infty]$ such that, on the tube $\mathcal{T}(\xi,\varepsilon)$, $f(q,t)$ is measurable with respect to t for each q and Lipschitz in q for almost every t, with Lipschitz constant $k(t)$.

If f is integrably Lipschitz near ξ, then we define a set-valued map $\mathcal{M}^{f,\xi}:[a,b]\twoheadrightarrow L(Q,Q)$ by letting $\mathcal{M}^{f,\xi}(t)$ be the Clarke generalized Jacobian at $\xi(t)$ of the map $q\to f(q,t)$. We then use $\Gamma(\mathcal{M}^{f,\xi})$ to denote the set of all single-valued measurable selections $[a,b]\ni t\to M(t)\in\mathcal{M}^{f,\xi}(t)$ of $\mathcal{M}^{f,\xi}$. For each $M\in\Gamma(\mathcal{M}^{f,\xi})$, we let $L^M:[a,b]\times[a,b]\to L(Q,Q)$ denote the fundamental solution of the matrix equation $\dot X(t)=M(t)X(t)$ (so that $\frac{\partial L^M}{\partial t}(t,s)=M(t)L^M(t,s)$ and $L^M(t,t)=\mathrm{id}_Q$). We then let $\Lambda^{f,\xi}_{t,s}$ denote the set $\Big\{L^M(t,s): M\in\Gamma(\mathcal{M}^{f,\xi})\Big\}$, and define

$$\Lambda^{f,\xi}\stackrel{\mathrm{def}}{=}\{\Lambda^{f,\xi}_{t,s}: s,t\in[a,b],\, s\le t\}.$$

J. Warga's theorem on differentiation of Lipschitz flows says that, *if $\xi:[a,b]\to Q$ is an integral curve of f, and $f:Q\times\mathbb{R}\to Q$ is integrably Lipschitz near ξ, then the flow Φ^f of f is multidifferentiable along ξ and $\Lambda^{f,\xi}\in MD(\Phi^f,\xi)$.*

3 The main theorem

We assume that

- *(A1) Q is a FDRLS;*
- *(A2) $\mathcal{F}=\{F_u\}_{u\in U}$ is a DI control system on Q;*
- *(A3) for every $u\in U$, the set-valued map F_u is ALSC and has closed values;*
- *(A4) $\mathcal{F}$ is closed under measurable substitutions;*

We use $\mathrm{TRAJ}(\mathcal{F})$ to denote the set of all absolutely continuous curves $\xi:[a,b]\to Q$ such that ξ satisfies $\dot\xi(t)=F_u(\xi(t),t)$ a.e. for some $u\in U$. We then assume that

- *(A5) $u_*\in U$, $\xi_*:[a,b]\to Q$ is a trajectory of F_{u_*}, and F_{u_*} is integrably pseudolipschitz along ξ_*;*

(A6) $S \subseteq Q$*, C is a closed convex cone in Q, C is a weakly approximating cone to S at $\xi_*(b)$, and C is not a linear subspace of Q.*

We let $\bar{q} = \xi_*(a)$, $\hat{q} = \xi_*(b)$, and define

$$\mathcal{R}^{\mathcal{F}}_{[a,b]}(\bar{q}) \stackrel{\mathrm{def}}{=} \{\xi(b) : \xi \in \mathrm{TRAJ}(\mathcal{F})\,,\, \xi(a) = \bar{q}\}\,.$$

We let $f : Q \times [a, b] \to Q$ be such that,

(A7) *ξ_* is an integral curve of f;*

(A8) *f is integrably Lipschitz near ξ_*;*

(A9) *f is a single-valued selection of the convexified set-valued map $(q,t) \to \mathrm{co}(F_{u_*}(q,t))$, where "co" stands for "closed convex hull."*

(A10) *$\mathcal{V} \subseteq Q \times ([a,b]\backslash K)$, and $\mathcal{V}$ is $\mathcal{F}$-admissible as a set of infinitesimal point variations along ξ_*, relative to Φ^f and Λ^{f,ξ_*}.*

The following result is then our "Maximum Principle with improved transversality conditions":

Theorem 2 *Assume that (A1,...,10) hold, and that the separation condition $\mathcal{R}^{\mathcal{F}}_{[a,b]}(\bar{q}) \cap S = \{\hat{q}\}$ is satisfied. Then there exist a measurable selection $M \in \Gamma(\mathcal{M}^{f,\xi_*})$ and a nonzero covector $\hat{\psi} \in Q^*$, such that, if we define*

$$\psi(t) \stackrel{\mathrm{def}}{=} \hat{\psi} \circ L^M(b,t)\,, \qquad (9)$$

then

(1) $-\hat{\psi} \in C^\dagger$,

(2) for every $u \in U$ there exists a subset $B_u \subseteq [a, b]$ of measure zero with the property that

$$\langle \psi(t), \dot{\xi}_*(t)\rangle \geq \max\{\langle \psi(t), v\rangle : v \in F_u(\xi_*(t), t)\}$$

for all $t \in [a,b]\backslash B_u$,

(3) $\langle \psi(t), v\rangle \leq 0$ *whenever* $(v,t) \in \mathcal{V}$.

Remark 3 Formula (9) is the *integrated form of the adjoint equation.* It is equivalent to the differentiated form

$$-\dot{\psi}(t) = \psi(t)M(t) \quad \text{for a.e. } t\,, \qquad (10)$$

together with the terminal condition $\psi(b) = \hat{\psi}$.

Condition (1) is our stronger form of the transversality condition. Condition (2) is the usual Hamiltonian minimization property.

Finally, (3) is an additional set of inequalities that corresponds to "high-order conditions." □

References

[1] Agrachev, A. A. and R. V. Gamkrelidze, "A second-order optimality principle for a time-optimal problem," *Math. Sbornik* **100**, 142 (1976).

[2] Agrachev, A. A. and R. V. Gamkrelidze, "The exponential representation of flows and the chronological calculus," *Math. Sbornik* **109**, 149 (1978).

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

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

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

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

[7] Knobloch, H. W., *High Order Necessary Conditions in Optimal Control*, Springer-Verlag, 1975.

[8] Krener, A. J., "The higher order maximum principle and its application to singular extremals," *SIAM J. Control Opt.* **15** (1977), pp. 256-293.

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

[10] Sussmann, H.J., "A strong version of the Maximum Principle under weak hypotheses," *Proc. 33rd IEEE Conference on Decision and Control*, Orlando, FL, 1994, pp. 1950-1956.

[11] 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, pp. 293-309.

[12] Sussmann, H.J., "Some recent results on the maximum principle of optimal control theory." In *Systems and Control in the Twenty-First Century*, C.I. Byrnes, B.N. Datta, D.S. Gilliam & C.F. Martin Eds., Birkhäuser Boston, 1997, pp. 351-372.

[13] Sussmann, H.J., "A strong maximum principle for systems of differential inclusions." In *Proceedings of the 35th IEEE Conference on Decision and Control*, Kobe, Japan, Dec. 1996. IEEE publications, 1996, pp. 1809-1814.

[14] Sussmann, H. J., "Multidifferential calculus: chain rule, open mapping and transversal intersection theorems," to appear in *Optimal control: theory, algorithms, and applications*, W.W. Hager & P.M. Pardalos Eds., Kluwer Acad. Publishers, 1998.

[15] Sussmann, H. J., "Geometry and optimal control," to appear in *Perspectives in optimal control*, J. Baillieul & J.C. Willems Eds., Springer, 1998.

[16] Tuan, H. D., "Controllability and extremality in nonconvex differential inclusions," *J. Optim. Theory Appl.* **85** (1995), pp. 435-472.

[17] Zhu. Q. J., "Necessary optimality conditions for nonconvex differential inclusions with endpoint constraints," *J. Diff. Equations* **124** (1996), pp. 186-204.