Some Recent Results on the Maximum Principle of Optimal Control Theory[1]

Héctor J. Sussmann

Department of Mathematics
Rutgers University
New Brunswick, NJ 08903

sussmann@hamilton.rutgers.edu
http://math.rutgers.edu/˜sussmann

1 Introduction

The maximum principle of optimal control theory (also known as the "Pontryagin maximum principle") was announced in the 1958 International Congress of Mathematicians, and presented in detail in the book [17]. The results of [17], however, are only a *version* of the principle, i.e. one of several possible ways of translating the statement of the principle into rigorous, precise mathematics.

For us, a "principle" is a somewhat vague assertion —such as, for example, the statement that "if the derivative of a function is positive then the function is increasing"— that one feels ought to become true once the terms involved are defined with care and precise technical hypotheses are specified. Since this can usually be done in more than one way, a principle typically has several different *versions*. In modern mathematics, it is customary to be careful about precision and to state technical hypotheses in full, so the discoverer of a new principle will often present it in the form of a rigorous first version, thereby failing to do full justice to the true power of the principle as a potential source of many other, equally rigorous, theorems.

This is what happened to the maximum principle. The rigorous version of [17] required that the right-hand side of the differential equation $\dot{x} = f(x, u, t)$ defining the control system be of class C^1 with respect to the state variables, with appropriate conditions on the dependence on u and t, and yielded as a conclusion the existence of a nontrivial "adjoint vector" satisfying the Hamiltonian minimization condition. Since then, new and more general versions have been proved by various techniques. These

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

versions strengthen the original one by either weakening the technical hypotheses or yielding stronger conclusions. Examples of the former type of generalization are (a) the "nonsmooth analysis" version (cf. [9, 10]), in which the right-hand side is only required to be Lipschitz continuous with respect to x, and the adjoint equation is replaced by an adjoint inclusion, (b) the maximum principle for differential inclusions, in which seemingly more general dynamical behaviors of the form $\dot{x} \in F(x,t)$ are considered, and (iii) the recent version due to S. Lojasiewicz Jr. ([16]), in which only the reference vector field is required to be Lipschitz with respect to x, and the other ones are only asked to be continuous. As an example of the latter kind of extension, we mention various versions of the "high-order maximum principle" (cf. [1, 2, 3, 13, 14, 15]), in which f is usually assumed to be very smooth with respect to x, and this is used to construct "high-order variations" involving Lie brackets of the vector fields of the given system, in addition to the "needle variations" used in the classical proof of the maximum principle. This results in extra necessary conditions for the adjoint vector ζ, in the form of a new set of inequalities $\langle \zeta(t), v \rangle \geq 0$, for vectors v other than the ones arising from the Hamiltonian minimization condition.

The purpose of this note is to announce a new result that contains, generalizes and combines most of the existing versions. We use the word "combines" because the new version applies to "hybrid systems," i.e. systems made of pieces of different kinds, each one of which falls under the scope of one of the existing versions. For example, suppose we have a system whose evolution is governed by a differential inclusion in some region $\mathcal{R}_1$ of the state space X, by a very smooth control system in some other region $\mathcal{R}_2$, and by a controlled O.D.E. with a Lipschitz right-hand in some other region $\mathcal{R}_3$. Suppose a trajectory $\xi_* : [a,b] \mapsto X$ visits all three regions, and we want to test ξ_* for optimality. If ξ_* is optimal, then the maximum principle for differential inclusions will give a nontrivial, Hamiltonian minimizing adjoint vector on any subinterval I of $[a,b]$ such that $\xi_*(I) \subseteq \mathcal{R}_1$, the high-order maximum principle will yield another nontrivial adjoint vector, satisfying Hamiltonian minimization as well as the high-order conditions, on any interval I such that $\xi_*(I) \subseteq \mathcal{R}_2$, and, finally, the nonsmooth maximum principle will produce a nontrivial Hamiltonian minimizing adjoint vector on any interval I such that $\xi_*(I) \subseteq \mathcal{R}_3$. But none of these three results will give a nontrivial adjoint vector ζ defined on the whole interval $[a,b]$ and satisfying Hamiltonian minimization everywhere as well as the high-order conditions at those times t such that $\xi_*(t) \in \mathcal{R}_2$. The reason for this is that the adjoint vectors obtained from the various versions of the maximum principle, by applying each version to a subtrajectory of ξ_* that stays in the appropriate region, need not match. Since the proofs of the different versions are carried out by different techniques, it does not seem possible to combine them into a single construction yielding a global adjoint vector.

Our version takes care of the combination problem and yields a single, global adjoint vector in hybrid situations. In addition, this version

(I) applies to problems where none of the other versions can be used,

(II) gives stronger results even in some cases where one of the older versions is applicable.

To illustrate the first point, consider the following three examples.

Example 1.1 Find the minimum-time trajectories for the control system in $\mathbb{R}^2$ given by $\dot{x} = u\varphi(x,y)$, $\dot{y} = v\varphi(x,y)$, where the control vector (u,v) is subject to the constraint $u^2+v^2 = 1$, and $\varphi : \mathbb{R}^2 \to \mathbb{R}$ is the function given by $\varphi(x,y) = c_1$ if $y \geq 0$, $\varphi(x,y) = c_2$ if $y < 0$, where c_1, c_2 are positive constants such that $c_1 \neq c_2$.

Example 1.2 Same question as in Example 1.1, except that now $\varphi(x,y) = \sqrt{|y|}$.

Example 1.3 Let α be such that $1 < \alpha < 2$. Consider the optimal control problem in $\mathbb{R}$, with the control u subject to $|u| \leq 1$, in which the state variable x evolves according to $\dot{x} = u + |x|^\alpha \sin\frac{1}{x}$. Suppose we want to minimize the integral $\int_0^1 u(t)e^t\,dt$, among all control-trajectory pairs $(u(\cdot), x(\cdot))$ such that $x(0) = x(1) = 0$. Is $u(t) \equiv 0$, $x(t) \equiv 0$ a minimizer?

In Examples 1.1 and 1.2, we are dealing with the problem of finding the "light rays" in a "medium" where the speed of light φ is variable. In Example 1.1, the speed of light is discontinuous. It is clear that an optimal path will be a straight line as long as it is contained in one of the half-planes $y \geq 0$, $y \leq 0$. A path joining points A, B in different half-planes will be a broken line, with a corner at the point C where it crosses the x axis. This still leaves a one-parameter family of paths as possible candidates for the solution. To find the true optimal path we need an extra condition that will determine C. Naturally, C can be found by elementary calculus, and the result is the well known Snell law of refraction. However, *this problem does not fit within the framework of the existing versions of the maximum principle, because the right-hand side is discontinuous as a function of the state.*

In Example 1.2, if we restrict ourselves to one of the half-planes $y \geq 0$, $y \leq 0$, we find the famous *brachystochrone problem*, proposed by John Bernoulli in 1696 as a challenge to the mathematical community, and solved the following year, separately, by several eminent mathematicians. (The May 1697 issue of *Acta Eruditorum* contains solutions by John Bernoulli himself, his brother Jacob, Leibniz, Newton, Tschirnhaus, and l'Hôpital.) The solutions can be found by classical calculus of variations or optimal control methods, and turn out to be cycloids. If, however, we want to join points A, B lying in different half-planes, we run into a difficulty similar to that of Example 1.1: one can easily prove that an optimal path will consist of a cycloid going from A to a point C in the axis, followed by another cycloid going from C to B, but none of the existing versions of the maximum principle will narrow down the choice further by determining C. (Notice that the right-hand side of the system is continuous but not Lipschitz with respect to the state variable.)

In Example 1.3, the right-hand side is not Lipschitz with respect to x, even for the "reference control" $u(t) \equiv 0$, so none of the nonsmooth versions apply. Notice, however, that if we proceed naïvely, and observe that the right-hand side is differentiable with respect to x at $x = 0$, and its derivative is 0, we get the "adjoint equation" $\dot{z} = 0$ for the multiplier $t \mapsto z(t)$. On the other hand, the Hamiltonian is

$$H(x, z, z_0, u, t) = z.\Big(u + |x|^\alpha \sin \frac{1}{x}\Big) + z_0 u.e^t \,, \tag{1.1}$$

where z_0 is the abnormal multiplier. If $z_0 = 0$, then the only way that the minimization condition could hold would be to have $z(t) \equiv 0$, which would contradict the nontriviality of the pair (z, z_0). So $z_0 \neq 0$, and we may assume that $z_0 = 1$. Then, for the Hamiltonian minimization condition to hold, it is necessary that $z(t) + e^t = 0$, i.e. that $z(t) = -e^t$, which contradicts the equation $\dot{z} = 0$. So the necessary conditions of the maximum principle are not satisfied, and the trajectory-control pair $(0, 0)$ is not optimal, *provided that we can rigorously justify the application of the "maximum principle" with the "classical" adjoint equation $\dot{z} = 0$, even though the right-hand side is not of class C^1 or even Lipschitz.*

It will turn out that *our version of the maximum principle gives extra necessary conditions that determine C in Examples 1.1 and 1.2, and provides a rigorous justification for the argument used above to exclude the optimality of the pair $(0, 0)$ for Example 1.3.*

To illustrate Point (II), we consider

Example 1.4 Same as Example 1.3, but with $\alpha = 2$.

Now the situation is different, because the right-hand side is Lipschitz, so the nonsmooth version of Clarke [9] *does* apply. It turns out, however, that this version gives the adjoint differential inclusion $-\dot{z} \in z.[-1, 1]$, since the Clarke generalized gradient of the function $x \to x^2 \sin \frac{1}{x}$ at $x = 0$ is the interval $[-1, 1]$. The function $z(t) = -e^t$ is a solution of this equation, so *the nonsmooth maximum principle fails to exclude the possibility that the control $u(t) \equiv 0$ might be optimal.* By contrast, the formal argument that we used above for the case when $1 < \alpha < 2$ applies without change when $\alpha = 2$, and establishes the nonoptimality of $u(t) \equiv 0$, provided that we can justify it rigorously. It will turn out that *our version of the maximum principle provides the necessary justification*, so in particular *this version* *yields a strictly stronger conclusion than the nonsmooth one in some cases where both versions apply.*

Since our version contains the nonsmooth one as a special case, it is clear that in general *our version cannot possibly give rise to a unique adjoint equation.* For instance, in Example 1.4 it is essential that our version implies the conclusion derived from the nonsmooth maximum principle — with the adjoint inclusion $-\dot{z} \in z[-1,1]$— as well as the different conclusion involving the adjoint equation $\dot{z} = 0$. This nonuniqueness is not a disadvantage of the new version but, quite to the contrary, it makes it even better, since the possibility of choosing among different adjoint equations means that several necessary conditions have to hold, making the new conclusions stronger, as we saw in our analysis of Example 1.4.

We now proceed to outline the statement of our version of the maximum principle, and some of the ideas going into its proof. The sketch given here will of necessity be rather incomplete, due to space limitations. For outlines covering other aspects of the proof, the reader is referred to our papers [19, 20, 21]. A complete account [22], with full details, is in preparation.

The basic strategy for the proof is exactly the same as that of the classical argument given in [17]. One looks at the flow defined by the reference vector field, and constructs "point variations" using the other vector fields of the system. These variations can be combined to yield "endpoint maps," and one applies a suitable open mapping theorem to these maps. Open mapping theorems usually say that if the differential of a map at a point is onto, then the map itself is open at the point. To apply such theorems to our situation, one needs to compute the differential of the endpoint maps, and this is done using the chain rule, since the endpoint map is a composite of simpler maps arising from the reference flow and the variations.

In the classical proof, all this is done in the class of continuously differentiable maps. The main point of our generalization is that exactly the same argument works within much more general classes of possibly set-valued maps (or *map germs*, to be precise), provided only that for these maps there is a concept of "generalized differential" for which the chain rule and a reasonable open mapping theorem hold. In our previous papers [19, 20, 21] we introduced one such class, and called its members "semidifferentiable maps." This class suffices to handle Lojasiewicz-type extensions of the nonsmooth maximum principle. Here we propose to use a still larger class, whose members will be called "multidifferentiable maps." This class still has a good chain rule as well as a suitable open mapping theorem, and in addition contains many maps that are not semidifferentiable. For example, if $F : \mathbb{R}^n \mapsto 2^{\mathbb{R}^n}$ is a locally Lipschitz set-valued map with compact (but not necessarily convex) values, then the flow of the differential inclusion $\dot{x} \in F(x)$ is multidifferentiable along every trajectory. (This follows by looking at the flows of suitable Lipschitz selections of the convexification of

F, and approximating their trajectories by trajectories of the nonconvexified inclusion, using the uniform approximation results of [11, 12]. More generally, for a differential inclusion $\dot{x} \in F(x,t)$, and a trajectory ξ such that F is integrably pseudolipschitz along ξ, the flow is multidifferentiable along ξ, as will be explained below.)

It turns out that, in the setting of multidifferentiable set-valued maps, one can rewrite the classical arguments of the book [17] for general *systems* $\{F_u : u \in U\}$ *of differential inclusions*, thereby obtaining a maximum principle that includes the case of a system of vector fields as well as that of a single differential inclusion. All that is needed is for the F_u to be "almost lower semicontinuous" and "locally integrably lower bounded," as defined below, and for the reference trajectory ξ_* to be generated by a reference inclusion $\dot{x} \in F_{u_*}(x,t)$ such that F_{u_*} is integrably pseudolipschitz along ξ_*. We handle this general situation by turning upside-down the conventional wisdom that says that "differential inclusions are more general than systems of vector fields, because a system of vector fields always gives rise to a differential inclusion, but the converse is not true, because set-valued maps, even when they are very smooth, typically do not admit nice single-valued selections." Our rejoinder to this statement is that it is true if "nice" is taken to mean "continuous," but there is no good reason for interpreting the word "nice" so narrowly. It has been shown by A. Bressan, in a series of papers (cf. [4, 5, 6, 7, 8]) that lower semicontinuous set-valued maps admit "conically continuous" selections. By slightly modifying and extending Bressan's results, one can show that, for a general differential inclusion with an almost lower semicontinuous right-hand side $(x,t) \mapsto F(x,t)$, there exist sufficiently many "nice" single-valued selections f of F. Here "nice" is taken to mean, essentially, that the indefinite integral $t \mapsto \int_a^t f(\xi(s),s)ds$ depends continuously on the continuous function ξ. (A precise definition will be given later.) This property suffices for the ordinary differential equation $\dot{x} = f(x,t)$ to be as good as an O.D.E. whose right-hand side is continuous with respect to x. Using these "nice" selections, we can carry out the entire program of the classical proof in the differential inclusions setting. As a byproduct of this work, we feel entitled to claim that *the "vector field approach" has been vindicated, and is in fact* **more general** *than that based on a single differential inclusion.* (We contend that the first part of our summary of the "conventional wisdom," according to which every system of vector fields gives rise to a differential inclusion, is also misleading. It is true that, given a control system $\dot{x} = f(x,u)$, $u \in U$, one can always define a differential inclusion $\dot{x} \in F(x)$ by letting $F(x) = \{f(x,u) : u \in U\}$. But *the passage from the control system to the differential inclusion often involves a loss of information.* For example, the control system $\dot{x} = ux$, with $x \in \mathbb{R}$, $u \in \mathbb{R}$, gives rise to the differential inclusion $\dot{x} \in F(x)$, where $F(x) = \{0\}$ if $x = 0$, $F(x) = \mathbb{R}$ of $x \neq 0$. However, if we specify that an admissible control is an integrable function $t \mapsto u(t)$, then the reachable set from 0 for the control system is just $\{0\}$, whereas the one for the differential inclusion is $\mathbb{R}$. This happens because not every trajectory of $\dot{x} \in F(x)$ is an admissible trajectory of $\dot{x} = ux$.)

2 Multidifferentials of set-valued maps

If X, Y are sets, we use $SVM(X,Y)$ to denote the set of all set-valued maps from X to Y, so the elements of $SVM(X,Y)$ are exactly the maps from X to the set 2^Y of all subsets of Y. If $F \in SVM(X,Y)$, then the *domain* of F is the set $\mathrm{Dom}(F) = \{x \in X : F(x) \neq \emptyset\}$, and the *graph* of F is the set $\mathrm{Graph}(F) = \{(x,y) \in X \times Y : y \in F(x)\}$. If $F \in SVM(X,Y)$ is a set-valued map, and S a set, the *restriction* of F to S is the map $F\lceil S$ whose graph is $\mathrm{Graph}(F) \cap (S \times Y)$.

If X is a metric space, we use d_X to denote the distance function on X. Then d_X is a function from $X \times X$ to $[0,\infty[$, but we will also use d_X to refer to the distance between subsets of X, given by

$$d_X(S,T) = \inf\{d_X(s,t) : s \in S, t \in T\}\,, \tag{2.2}$$

so d_X takes values in $[0,\infty]$. (The equality $d_X(S,T) = \infty$ holds iff one of the sets S, T is empty.) We use Δ_X to denote the function on $2^X \times 2^X$ given by

$$\Delta_X(S,T) = \max\Big(0, \sup\{d_X(s,T) : s \in S\}\Big)\,. \tag{2.3}$$

Then Δ_X takes values in $[0,+\infty]$. (We have $\Delta_X(S,T) = 0$ iff $S \subseteq \mathrm{Clos}(T)$, and $\Delta_X(S,T) = +\infty$ iff $T = \emptyset \neq S$.) The Hausdorff distance $d_X^h(S,T)$ between two subsets S, T of X is given by $d_X^h(S,T) = \Delta_X(S,T) + \Delta_X(T,S)$.

If X, Y are metric spaces, then the product $X \times Y$ is a metric space with distance defined by $d_{X\times Y}\Big((x,y),(x',y')\Big) = d_X(x,x') + d_Y(y,y')$.

If K is a compact metric space, and X is a metric space, a *regular set-valued map* from K to X is a set-valued map $F \in SVM(K,X)$ with a compact graph, having the property that there exist a sequence $\{F_j\}$ of continuous (single-valued) maps $F_j : K \mapsto X$ and a compact subset K' of X such that $F_j(K) \subseteq K'$ for all j, and

$$\lim_{j\to\infty} \Delta_{K\times X}(\mathrm{Graph}(F_j), \mathrm{Graph}(F)) = 0\,. \tag{2.4}$$

We use $\mathrm{REG}(K,X)$ to denote the set of all $F \in SVM(K,X)$ that are regular, and identify the space $C^0(K,X)$ in an obvious way with a subset of $\mathrm{REG}(K,X)$. Clearly, if $F \in \mathrm{REG}(K,X)$ then (i) $\mathrm{Dom}(F) = K$, and (ii) $F \in C^0(K,X)$ iff F is single-valued.

If S is a closed subset of $\mathbb{R}^n$, a *map of class C^1 from S to $\mathbb{R}^m$ in the Whitney sense* is a pair $F = (F^0, F^1)$ such that $F^0 \in C^0(S,\mathbb{R}^m)$,

$F^1 \in C^0(S, \mathbb{R}^{m\times n})$, and, for each $x \in S$,

$$^0(y) - F^0(x) = F^1(x).(y-x) + o(||y-x||) \text{ as } y \to x \text{ via values in } S\,. \quad (2.5)$$

We use $C^1(S, \mathbb{R}^m)$ to denote the set of all maps of class C^1 in the Whitney sense from S to $\mathbb{R}^m$. The well known *Whitney extension theorem* (cf. [23, 28]) says that a pair (F^0, F^1) belongs to $C^1(S, \mathbb{R}^m)$ iff there exists a map $\tilde{F} : \mathbb{R}^n \mapsto \mathbb{R}^m$ of class C^1 in the usual sense such that $\tilde{F}\lceil S = F^0$ and $D\tilde{F}\lceil S = F^1$. In many cases —e.g. if S is a closed convex subset of $\mathbb{R}^n$ with nonempty interior— one can identify $C^1(S, \mathbb{R}^m)$ with the space of maps $G \in C^1(\mathrm{Int}(S), \mathbb{R}^m)$ such that both G and DG have continuous extensions to S, but this identification is not valid for general closed sets S, even if $S = \mathrm{Clos}(\mathrm{Int}(S))$. If $F = (F^0, F^1) \in C^1(S, \mathbb{R}^m)$, we will write $F(x)$ for $F^0(x)$, and define $DF = F^1$.

If $\mathbf{L}$ is a subset of $\mathbb{R}^{m\times n}$, and $S \subseteq \mathbb{R}^n$ is closed, we use $C^1_{\mathbf{L}}(S, \mathbb{R}^m)$ to denote the set of all maps $F \in C^1(S, \mathbb{R}^m)$ such that $DF(S) \subseteq \mathbf{L}$ (i.e. $DF(x) \in \mathbf{L}$ for each $x \in S$).

If S is a subset of a metric space X, and $p \in X$, we say that S is *closed near* p if the set

$$S_{\varepsilon,p} = \{x \in S : d_X(x,p) \le \varepsilon\} \quad (2.6)$$

is closed in X for some $\varepsilon > 0$.

Let $F \in SVM(\mathbb{R}^n, \mathbb{R}^m)$ and let $p \in \mathbb{R}^n$, $q \in \mathbb{R}^m$. Let $S \subseteq \mathbb{R}^n$ be such that $p \in S$. Let $\mathbf{L}$ be a compact subset of $\mathbb{R}^{m\times n}$. We say that $\mathbf{L}$ is a *multidifferential of* F *at* (p,q) *along* S if for every neighborhood $\mathbf{L}'$ of $\mathbf{L}$ in $\mathbb{R}^{m\times n}$ there exist an $\bar{\varepsilon} > 0$, families $\{G_\varepsilon\}_{\varepsilon\in]0,\bar{\varepsilon}]}$, $\{H_\varepsilon\}_{\varepsilon\in]0,\bar{\varepsilon}]}$, and a function $\theta :]0,\bar{\varepsilon}] \to]0,\infty[$ such that $\lim_{\varepsilon\downarrow 0}\theta(\varepsilon) = 0$, with the property that, for each $\varepsilon \in]0,\bar{\varepsilon}]$, if we define $S_{\varepsilon,p}$ by (2.6), then

- (MD.1) $S_{\varepsilon,p}$ is compact,
- (MD.2) $G_\varepsilon \in \mathrm{REG}(S_{\varepsilon,p}, \mathbb{R}^m)$,
- (MD.3) $\mathrm{Graph}(G_\varepsilon) \subseteq \mathrm{Graph}(F)$,
- (MD.4) $H_\varepsilon \in C^1_{\mathbf{L}'}(S_{\varepsilon,p}, \mathbb{R}^m)$,
- (MD.5) $H_\varepsilon(p) = q$,
- (MD.6) $\Delta_{\mathbb{R}^n}(G_\varepsilon(x), H_\varepsilon(x)) \le \varepsilon\theta(\varepsilon)$ for all $x \in S_{\varepsilon,p}$.

We use $MD_{p,q,S}(F)$ to denote the set of all multidifferentials of F at (p,q) along S, and call F *multidifferentiable at* (p,q) *along* S if $MD_{p,q,S}(F) \neq \emptyset$. It is easy to see that, if F and F' are two set-valued maps in $SVM(\mathbb{R}^n, \mathbb{R}^m)$ such that $F = F'$ near (p,q) (that is, there exists a neighborhood U of (p,q) in $\mathbb{R}^n \times \mathbb{R}^m$ such that $\mathrm{Graph}(F) \cap U = \mathrm{Graph}(F') \cap U$), and S, S' are such that $S = S'$ near p, then $MD_{p,q,S}(F) = MD_{p,q,S'}(F')$. This means that multidifferentiability at (p,q) along S is really a property of germs $\mathbf{F}$ at (p,q) of set-valued maps F from $\mathbb{R}^n$ to $\mathbb{R}^m$ and germs $\mathbf{S}$ of sets S at p, and the multidifferential $M_{p,q,\mathbf{S}}(\mathbf{F})$ of a germ $\mathbf{F}$ at (p,q) along a germ $\mathbf{S}$ at p is well defined.

If F is multidifferentiable at (p,q) along S, then S is necessarily closed near p, and the germ of S at p is contained in the germ of $\mathrm{Dom}(F)$ at p (that is, there exists $\varepsilon > 0$ such that $x \in S$, $d_X(x,p) \le \varepsilon$ implies $x \in \mathrm{Dom}(F)$).

If $F \in SVM(\mathbb{R}^n, \mathbb{R}^m)$, we will write $MD_{p,q}(F)$ for $MD_{p,q,\mathbb{R}^n}(F)$.

Multidifferentials are never unique. (For example, if $\mathbf{L} \in M_{p,q}(F)$, $\mathbf{L}'$ is compact, and $\mathbf{L} \subseteq \mathbf{L}'$, then obviously $\mathbf{L}' \in MD_{p,q}(F)$ as well.) It would be nice if among all the members of $MD_{p,q}(F)$ there was one which is a subset of all the others —in which case it would be reasonable to call this set *the* multidifferential of F at p, q. It turns out, however, that in general such a "canonical" multidifferential does not exist.

Example 2.1 Consider the map $\mathbb{R}^2 \ni (x,y) \mapsto F(x,y) = (0,0) \in \mathbb{R}^2$. For $\alpha \ge 0$, let $\mathbf{L}_\alpha$ be the set of all 2×2 matrices (a_{ij}) such that $\sum_{ij} a_{ij}^2 = \alpha^2$. Let $G_\varepsilon(x,y) = (0,0)$, and

$$H_\varepsilon(x,y) = \alpha\varepsilon^2(\sin\frac{x}{\varepsilon^2} + \sin\frac{y}{\varepsilon^2}, \cos\frac{x}{\varepsilon^2} + \cos\frac{y}{\varepsilon^2}). \qquad (2.7)$$

Then $DH_\varepsilon(x,y) \in \mathbf{L}_\alpha$ for all (x,y). If we let $p = q = (0,0)$, it follows easily that $\mathbf{L}_\alpha \in MD_{p,q}(F)$. Clearly, the sets $\mathbf{L}_\alpha$ are pairwise disjoint.

So we have to accept the fact that multidifferentials in general are not unique, and there is no canonical way to single out a distinguished member of $MD_{p,q}(F)$.

An important example of multidifferentials is provided by the *semidifferentials*, introduced in our previous papers [19, 20, 21]. A set-valued map $F \in SVM(\mathbb{R}^n, \mathbb{R}^m)$ is *semidifferentiable* at $p \in \mathbb{R}^n$ if there is a compact neighborhood U of p such that (a) $\mathrm{Dom}(F) \cap U$ is compact, (b) $F\lceil U \in \mathrm{REG}(\mathrm{Dom}(F) \cap U, \mathbb{R}^m)$, (c) $p \in \mathrm{Dom}(F)$, and (d) there exist a Lipschitz map $G : \mathbb{R}^n \mapsto \mathbb{R}^m$ and a function $\theta :]0,\infty[\mapsto [0,\infty]$ such that $\lim_{s\downarrow 0}\theta(s) = 0$ and

$$\Delta_{\mathbb{R}^n}(F(x), G(x)) \le ||x-p||\,\theta(||x-p||) \ \text{ whenever } \ x \in \mathrm{Dom}(F). \qquad (2.8)$$

(It then follows that $F(p) = \{G(p)\}$, since $\Delta_{\mathbb{R}^n}(F(p), G(p)) = 0$.) Any G for which the above conditions hold is called a *first-order Lipschitz approximation of F at p*. A *semidifferential* of F at p is any subset $\mathbf{L}$ of $\mathbb{R}^{m\times n}$ which is a Warga derivate container of G at p for some first-order Lipschitz approximation G of F at p. (A *Warga derivate container* of G at p is a compact subset $\mathbf{L}$ of $\mathbb{R}^{m\times n}$ such that for every neighborhood $\mathbf{L}'$ of $\mathbf{L}$ in $\mathbb{R}^{m\times n}$ there exist a neighborhood V of p and a sequence $\{H_j\}$ of C^1 maps $V \mapsto \mathbb{R}^m$ such that $H_j \to G$ uniformly on V and $DH_j(x) \in \mathbf{L}'$ for all j and all $x \in V$.) We use $SD_p(F)$ to denote the set of all semidifferentials of F at p.

Clearly, if F is semidifferentiable at p, and $q = F(p)$, then $\mathrm{Dom}(F)$ is closed near p, F is multidifferentiable at (p,q) along $\mathrm{Dom}(F)$, and the inclusion $SD_p(F) \subseteq MD_{p,q,\mathrm{Dom}(F)}(F)$ holds.

Multidifferentiable maps are of interest to us because (i) they have a good calculus, and in particular a chain rule, (ii) they satisfy good open mapping and separation theorems, and (iii) many maps naturally associated to controls systems and differential inclusions are multidifferentiable.

The *chain rule for multidifferentiable maps* says that,

Theorem 2.2 *If $F_1 \in SVM(\mathbb{R}^n, \mathbb{R}^m)$, $F_2 \in SVM(\mathbb{R}^m, \mathbb{R}^\nu)$, $S_1 \subseteq \mathbb{R}^n$, $S_2 \subseteq \mathbb{R}^m$, $p \in S_1$, $q \in F_1(p)$, $q \in S_2$, $r \in F_2(q)$, $F_1(S_1) \subseteq S_2$, $\mathbf{L_1} \in MD_{p,q,S_1}(F_1)$, and $\mathbf{L_2} \in MD_{q,r,S_2}(F_2)$, then $\mathbf{L_2} \circ \mathbf{L_1} \in MD_{p,r,S_1}(F_2 \circ F_1)$, if one of the following conditions holds:*

(CR.1) S_1 is convex near p and S_2 is convex near q,

(CR.2) F_1 is a single-valued map of class C^1 on S_1 and $\mathbf{L_1} = \{DF_1(p)\}$.

If M is a manifold of class C^1, and $p \in M$, we use T_pM to denote the tangent space of M at p. If $S \subseteq M$ and $p \in S$, the *Bouligand tangent cone* (or *contingent tangent cone*) to S at p is the set $T_p^B(S)$ of all vectors $v \in T_pM$ such that either $v = 0$ or there exist a sequence $\{p_j\}$ in S and real numbers $h_j > 0$ such that $h_j \to 0$ and $f(p_j) - f(p) = h_j.vf + o(h_j)$ as $j \to \infty$ for every $f \in C^1(M, \mathbb{R})$. A subset S of M is said to be *convex near* p if there exists a C^1 coordinate chart $\kappa : U \to \mathbb{R}^m$, defined on a neighborhood U of p, such that $\kappa(S \cap U)$ is convex.

It follows easily from the chain rule that, if M, N are manifolds of class C^1, $p \in M$, $q \in N$, and $S \subseteq M$ is such that $p \in S$, S is closed near p and convex near p, and $F \in SVM(M,N)$, then $MD_{p,q,S}(F)$ is intrinsically defined as a collection of compact subsets of $L(T_pM, T_qN)$, the space of linear maps from T_pM to T_qN.

The *open mapping theorem* gives a sufficient condition for a set-valued map to be open at a point. If $F \in SVM(X,Y)$, where X and Y are topological spaces, we call F *open* at a point $(p,q) \in \mathrm{Graph}(F)$ if for every neighborhood U of p the image $F(U)$ is a neighborhood of q. The theorem then says that

Theorem 2.3 *If M, N are C^1 manifolds, $F \in SVM(M,N)$, $S \subseteq M$, $p \in S$, S is closed and convex near p, $q \in N$, $\mathbf{L} \in MD_{p,q,S}(F)$, and*

(2.3.) every $L \in \mathbf{L}$ maps $T_p^B(S)$ onto T_qN,*

then F is open at (p,q).

More generally, there is a *directional open mapping theorem*: if M, N are C^1 manifolds, $p \in M$, $q \in F(p)$, $v \in T_qN$, we say that F is *directionally* *open at* (p, q) *in the direction of* v if for every neighborhood U of p the image $F(U)$ is a v-directional neighborhood of q. (If $q \in N$ and $v \in T_qN$, a *v-directional neighborhood* of q is a subset V of N with the property that, for some chart κ mapping a neighborhood U of q to $\mathbb{R}^n$, the set $\kappa(V \cap U)$ contains a convex set W such that $\kappa(q) \in W$ and $\kappa(q)+r.d\kappa(q).v \in \mathrm{Int}(W)$ for some $r > 0$. It is easy to see that if this property holds for some chart κ defined near p then it holds for every such chart.) Then Theorem 2.3 has the following generalization:

Theorem 2.4 *If* M, N, F, S, p, q, $\mathbf{L}$ *are as in Theorem 2.3,* $v \in T_qN$, *and instead of (2.4.*) we assume*

(2.4.)* $v \in \mathrm{Int}(LT_p^B(S))$ *for every* $L \in \mathbf{L}$,

then F *is directionally open at* (p, q) *in the direction of* v.

Clearly, Theorem 2.3 is a special case of Theorem 2.4, corresponding to $v = 0$.

3 A separation theorem

Theorem 2.4, together with standard separation properties of convex sets, imply the following rudimentary version of the maximum principle.

Theorem 3.1 *Suppose that* M_1, M_2, N *are* C^1 *manifolds,* $q \in N$, *and, for* $i = 1, 2$, $p_i \in S_i \subseteq M_i$, $F_i \in SVM(M_i, N)$, $q \in F_i(p_i)$, S_i *is closed and convex near* p_i, *and* $\mathbf{L_i} \in MD_{p_i,q,S_i}(F_i)$. *Suppose that the sets* $F_1(S_1)$ *and* $F_2(S_2)$ *are separated near* q, *in the sense that*

$$F_1(S_1) \cap F_2(S_2) = \{q\}\,. \tag{3.9}$$

Assume, in addition, that

(i) $T_{q_2}^B S_2$ *is not a linear subspace of* $T_{q_2}M_2$,

(ii) $q \notin F_2(x)$ *whenever* $x \in M_2$ *is close enough to* p_2 *and* $x \neq p_2$.

Then there exist a nontrivial linear functional $z : T_qN \mapsto \mathbb{R}$ *and* $L_i \in \mathbf{L_i}$ *for* $i = 1, 2$ *such that* $z \circ L_1$ *is* ≥ 0 *on* $T_{p_1}^B S_1$ *and* $z \circ L_2$ *is* ≤ 0 *on* $T_{p_2}^B S_2$.

Theorem 3.1 is proved as follows. We apply Theorem 2.4 to the set-valued map $Q \in SVM(M_1 \times M_2, N \times \mathbb{R})$ such that

$$Q(x_1, x_2) = (F_1(x_1) - F_2(x_2), \mu(x_2))\,, \tag{3.10}$$

where we identify a neighborhood W of p_2 in M_2 with a neighborhood of 0 in $T_{p_2}M_2$ using a coordinate chart, and $\mu : T_{p_2}M_2 \mapsto \mathbb{R}$ is a linear functional which is ≥ 0 but not identically 0 on $T_{p_2}^B S_2$. Here

$$F_1(x_1) - F_2(x_2) \stackrel{\mathrm{def}}{=} \{y_1 - y_2 : y_1 \in F_1(x_1), y_2 \in F_2(x_2)\}\,. \tag{3.11}$$

The existence of μ is guaranteed by the fact that $T^B_{p_2}S_2$ is a closed convex cone in $T_{p_2}M_2$ —since S_2 is convex near p_2— which is not a linear subspace. Letting $\mathcal{N} = N \times \mathbb{R}$, $\bar{q} = (q, 0)$, we pick $v = (0, 1) \in T_{\bar{q}}\mathcal{N} \sim T_q N \times \mathbb{R}$. For $L_i \in MD_{p_i,q,S_i}(F_i)$, $i = 1, 2$, let

$$C(L_1, L_2) = \left\{ \Big(L_1u_1 - L_2u_2, \mu(u_2)\Big) : (u_1, u_2) \in T^B_{p_1}S_1 \times T^B_{p_2}S_2 \right\}. \quad (3.12)$$

Theorem 2.4 implies that if $v \in \mathrm{Int}(C(L_1, L_2))$ for all L_1, L_2, then the pair $(0, r)$ belongs to $Q(S_1 \times S_2)$ for small enough r. Pick a small $r > 0$, and write

$$(0, r) = (y_1 - y_2, \mu(x_2)), \quad (3.13)$$

with $x_1 \in S_1$, $x_2 \in S_2$. Then $y_1 = y_2$. Also, $\mu(x_2) = r \neq 0$, and $y_2 \in F_2(x_2)$. Therefore $y_2 \neq q$ by Assumption (ii). So $F_1(S_1) \cap F_2(S_2)$ contains a point of N other than q, contradicting the assumption that $F_1(S_1) \cap F_2(S_2) = \{q\}$. Therefore there must exist $L_1 \in \mathbf{L}_1$, $L_2 \in \mathbf{L}_2$, such that $v \notin \mathrm{Int}(C(L_1, L_2))$. But then there exists a nontrivial linear functional $\bar{z}$ on $T_{\bar{q}}\mathcal{N}$ such that $\bar{z}(v) \leq 0$ and $\bar{z}$ is ≥ 0 on $C(L_1, L_2)$. Write $\bar{z}(x, r) = z(x) + \rho r$, where $z : T_qN \mapsto \mathbb{R}$ is linear. Taking $x = 0$, $r = 1$, we get $\rho \leq 0$. If $u_1 \in T^B_{p_1}S_1$, $u_2 \in T^B_{p_2}S_2$, we have $(L_1u_1 - L_2u_2, \mu(u_2)) \in C(L_1, L_2)$, so $z(L_1u_1) - z(L_2u_2) + \rho\mu(u_2) \geq 0$. Taking $u_2 = 0$ we find $z(L_1u_1) \geq 0$. Taking $u_1 = 0$ we find $z(L_2u_2) \leq \rho\mu(u_2)$. Since $\rho \leq 0$ and $\mu(u_2) \geq 0$, we have $z(L_2u_2) \leq 0$. Finally, if $z \equiv 0$ we would get $\rho\mu(u_2)) \geq 0$ for all $u_2 \in T^B_{p_2}S_2$. Since $\rho \leq 0$, and μ is ≥ 0 but not $\equiv 0$ on $T^B_{p_2}S_2$, we must have $\rho = 0$. But then $\bar{z} = 0$, which is a contradiction. This completes the proof of Theorem 3.1.

4 Multidifferentials of flows

A *flow* is a triple $\mathcal{F} = (T, \mathbf{X}, \mathbf{F})$, where T is a totally ordered set, $\mathbf{X} = \{X_t\}_{t\in T}$ is a family of sets parametrized by T, and —using $\preceq$ to denote the order relation in T— $\mathbf{F} = \{F_{t,s}\}_{s,t\in T, s\preceq t}$ is a family of set-valued maps such that

(FL.1) $F_{t,s} \in SVM(X_s, X_t)$ whenever $s, t \in T$ and $s \preceq t$,

(FL.2) $F_{t_3,t_2} \circ F_{t_2,t_1} = F_{t_3,t_1}$ whenever $t_1, t_2, t_3 \in T$ and $t_1 \preceq t_2 \preceq t_3$,

(FL.3) $F_{t,t} = \mathrm{id}_{X_t}$ for every $t \in T$.

A *trajectory* of a flow $\mathcal{F} = (T, \mathbf{X}, \mathbf{F})$ is a single-valued map ξ such that $\mathrm{Dom}(\xi) = T$, $\xi(t) \in X_t$ for every $t \in T$, and $\xi(t) \in F_{t,s}(\xi(s))$ whenever $s, t \in T$ and $s \preceq t$.

If $\mathcal{F} = (T, \mathbf{X}, \mathbf{F})$ is a flow such that each X_t is a C^1 manifold, and ξ is a trajectory of $\mathcal{F}$, a *multidifferential of $\mathcal{F}$ along ξ* is a family $\mathcal{L} = \{\mathbf{L}_{t,s}\}_{s,t\in T, s\preceq t}$ such that

(MDF.1) $\mathbf{L}_{t,s} \in MD_{\xi(s),\xi(t)}(F_{t,s})$ whenever $s,t \in T$ and $s \preceq t$,

(MDF.2) $\mathbf{L}_{t,t} = \{\mathrm{id}_{T_{\xi(t)}X_t}\}$ for all $t \in T$,

(MDF.3) $\mathbf{L}_{t_3,t_2} \circ \mathbf{L}_{t_2,t_1} = \mathbf{L}_{t_3,t_1}$ whenever $t_1, t_2, t_3 \in T$ and $t_1 \preceq t_2 \preceq t_3$.

(Naturally, if $\mathbf{L}$, $\mathbf{L}'$ are subsets of $L(X,Y)$ and $L(Y,Z)$, respectively, where X, Y, Z are linear spaces, $\mathbf{L}' \circ \mathbf{L}$ denotes the set $\{L' \circ L : L \in \mathbf{L}, L' \in \mathbf{L}'\}$.)

We call $\mathcal{F}$ *multidifferentiable along* ξ if there exists a multidifferential of $\mathcal{F}$ along ξ. We remark that multidifferentiability of a flow $\mathcal{F} = (T, \mathbf{X}, \mathbf{F})$ along a trajectory ξ of $\mathcal{F}$ implies in particular that the domain of each map $F_{t,s}$ is a full neighborhood of $\xi(s)$. (Indeed, according to our general conventions, $MD_{\xi(s),\xi(t)}(F_{t,s})$ stands for $MD_{\xi(s),\xi(t),X_s}(F_{t,s})$, which cannot be nonempty unless $\xi(s) \in \mathrm{Int}_{X_s}(\mathrm{Dom}(F_{t,s}))$.) Moreover, even if $\xi(s) \in \mathrm{Int}_{X_s}(\mathrm{Dom}(F_{t,s}))$ for all t, s, the requirement that $\mathcal{F}$ be multidifferentiable along ξ is in general stronger than the condition that $F_{t,s}$ be multidifferentiable at $(\xi(s), \xi(t))$ for all (s,t), since it is possible to construct examples where $MD_{\xi(s),\xi(t)}(F_{t,s}) \neq \emptyset$ whenever $s, t \in T$ and $s \preceq t$, but a familiy $\mathcal{L}$ satisfying (MDF.1) and (MD.2) as well as the compatibility condition (MDF.3) does not exist.

Given finite-dimensional linear spaces Y_t for t in a totally ordered set T, and a family $\mathbf{L} = \{\mathbf{L}_{t,s}\}_{t,s\in T, s\preceq t}$ such that each $L_{t,s}$ is a subset of the space $L(Y_s, Y_t)$ of linear maps from Y_s to Y_t, a *compatible selection* of $\mathbf{L}$ is a family $L = \{L_{t,s}\}_{t,s\in T,s\preceq t}$ such that $L_{t,s}$ belongs to $\mathbf{L}_{t,s}$ for each t, s, and $L_{t_3,t_2} L_{t_2,t_1} = L_{t_3,t_1}$ whenever $t_1, t_2, t_3 \in T$ and $t_1 \preceq t_2 \preceq t_3$.

The following result is proved by an elementary argument using Zorn's Lemma.

Theorem 4.1 *If all the $\mathbf{L}_{t,s}$ are compact and nonempty, and the compatibility condition (MDF.3) holds, then $\mathbf{L}$ has a compatible selection. More generally, any compatible selection $\tilde{L}$ defined on a subset $\tilde{T}$ of T can be extended to a compatible selection on T.*

5 Point variations of a flow

Let M be a manifold of class C^1, let $\bar{x} \in M$, and let k be a positive integer. A *k-parameter point variation on M with base point $\bar{x}$* is a set-valued map $V \in SVM(\mathbb{R}_+^k \times M, M)$ such that

(i) $\mathrm{Dom}(V)$ is a neighborhood of $(0, \bar{x})$ in $\mathbb{R}_+^k \times M$,

(ii) $V(0, x) = x$ whenever $(0, x) \in \mathrm{Dom}(V)$.

We use $\mathrm{VAR}_k(M, \bar{x})$ to denote the set of all k-parameter point variations on M with base point $\bar{x}$, and write $\mathrm{VAR}(M, \bar{x}) \stackrel{\mathrm{def}}{=} \cup_{k=1}^{\infty} \mathrm{VAR}_k(M, \bar{x})$.

We call a $V \in \mathrm{VAR}_k(M,\bar{x})$ *nicely multidifferentiable* if there exists a k-tuple $(v_1,\ldots,v_k)$ of vectors in $T_{\bar{x}}M$ such that, if $L:\mathbb{R}^k \times T_{\bar{x}}M \mapsto T_{\bar{x}}M$ is the linear map given by $L(r_1,\ldots,r_k,u)=r_1v_1+\ldots+r_kv_k+u$, then $\{L\}$ belongs to $MD_{(0,\bar{x}),\bar{x},\mathbb{R}^k_+\times M}(V)$. We use $NMD(V)$ to denote the set of all such k-tuples $(v_1,\ldots,v_k)$.

Now suppose $\mathcal{F}=(T,\mathbf{X},\mathbf{F})$ is a flow such that T has a minimum and a maximum. Let ξ be a trajectory of $\mathcal{F}$. A *combined point variation* of $\mathcal{F}$ along ξ is a map $\mathcal{V}$, defined on a finite subset $|\mathcal{V}|$ of T, such that $\mathcal{V}(t) \in \mathrm{VAR}(X_t,\xi(t))$ for each $t \in |\mathcal{V}|$. (The set $|\mathcal{V}|$ is the *carrier* of $\mathcal{V}$.) Given a combined point variation $\mathcal{V}$ of $\mathcal{F}$ along ξ, the *endpoint map* $\mathcal{E}_{\mathcal{V}}$ is defined as follows. Write $|\mathcal{V}|=\{t_1,\ldots,t_m\}$, where $t_1 \prec t_2 \prec \ldots \prec t_m$. Let $a=\min T$, $b=\max T$. Let $\mathcal{V}(t_j)=V_j$, and let k_j be such that $V_j \in \mathrm{VAR}_{k_j}(X_{t_j},\xi(t_j))$. Let $k=k_1+\ldots+k_m$, and write $\kappa_j=k_1+\ldots+k_j$ for $j=1,\ldots,m$. For $a \preceq t \prec t_1$, let $E_t(x)=F_{t,a}(x)$. Then define

$$E_{t_1}(\varepsilon_1,\ldots,\varepsilon_{k_1},x)=V_1\Big(\varepsilon_1,\ldots,\varepsilon_{k_1},F_{t_1,a}(x)\Big)\,. \tag{5.14}$$

For $j=1,\ldots,m-1$, $t_j \prec t \prec t_{j+1}$, define

$$E_t(\varepsilon_1,\ldots,\varepsilon_{\kappa_j},x)=F_{t,t_j}\Big(E_{t_j}(\varepsilon_1,\ldots,\varepsilon_{\kappa_j},x)\Big)\,. \tag{5.15}$$

Then let

$$\begin{aligned}&E_{t_{j+1}}(\varepsilon_1,\ldots,\varepsilon_{\kappa_{j+1}},x)=\\&\quad V_{j+1}\Big(\varepsilon_1,\ldots,\varepsilon_{k_{j+1}},F_{t_{j+1},t_j}(E_{t_j}(\varepsilon_{k_j+1},\ldots,\varepsilon_{\kappa_{j+1}},x))\Big)\,.\end{aligned} \tag{5.16}$$

We then let

$$E_t(\varepsilon_1,\ldots,\varepsilon_k,x)=F_{t,t_m}\Big(E_{t_m}(\varepsilon_1,\ldots,\varepsilon_k,x)\Big)\ \text{ for } t_m \preceq t \preceq b\,. \tag{5.17}$$

Finally, $\mathcal{E}_{\mathcal{V}}$ is, by definition, the map E_b. Then $\mathcal{E}_{\mathcal{V}} \in SVM(\mathbb{R}^k_+ \times X_a, X_b)$.

If $\mathcal{F}=(T,\mathbf{X},\mathbf{F})$ is a flow such that each X_t is a C^1 manifold, a *point variational vector* for $\mathcal{F}$ along a trajectory ξ of $\mathcal{F}$ is a pair (t,v) such that $t \in T$ and $v \in T_{\xi(t)}X_t$. Let W be a set of variational vectors for $\mathcal{F}$ along ξ. Let $\mathcal{R}$ be a subset of X_b, where $b=\max T$. Let $a=\min T$. We say that W is *$\mathcal{R}$-compatible* if, whenever $t_1 \prec t_2 \prec \ldots \prec t_m$, and $W_1,\ldots,W_m$ are finite subsets of W such that $W_j \subseteq \{t_j\}\times T_{\xi(t_j)}X_{t_j}$ for $j=1,\ldots,m$, it follows that there exists a combined variation $\mathcal{V}$ of $\mathcal{F}$ along ξ with carrier $\{t_1,\ldots,t_m\}$, such that

- (i) each $\mathcal{V}(t_j)$ is nicely multidifferentiable, and there is a $\mathbf{v}_j \in NMD(V(t_j))$ in which all the vectors v for which $(t_j,v)\in W_j$ occur,
- (ii) the image of $\mathcal{E}_{\mathcal{V}}$ is contained in $\mathcal{R}$.

6 The maximum principle

If M is a C^1 manifold, $S \subseteq M$ and $q \in S$, an *approximating cone* for S at q is a closed convex cone C in the tangent space T_qM such that there exist a neighborhood U of 0 in T_qM and a continuous map $\varphi : C \cap U \mapsto S$ with the property that $\varphi(0) = q$ and $\varphi(v) = q + v + o(||v||)$ as $v \to 0$, $v \in C$. (The condition that $\varphi(v) = q + v + o(||v||)$ is in principle coordinate-dependent, but one can easily show that it is in fact independent of the choice of coordinates. It can be stated invariantly as follows: for every $\psi \in C^1(M, \mathbb{R})$, $\psi(\varphi(v)) - \psi(q) - v\psi = o(||v||)$ as $v \to 0$.)

If X, Y are linear spaces and $A : X \mapsto Y$ is a linear map, we use $A^\dagger$ to denote the transpose of A, so $A^\dagger$ maps the dual space of Y to the dual space of X.

Theorem 6.1 *Let $\mathcal{F} = (T, \mathbf{X}, \mathbf{F})$ be a flow such that T has a minimum a and a maximum b, and each X_t is a C^1 manifold. Let ξ be a trajectory of $\mathcal{F}$, and let $\mathbf{L}$ be a multidifferential of $\mathcal{F}$ along ξ. Let $\mathcal{R}$ be a subset of X_b, and let W be an $\mathcal{R}$-compatible set of variational vectors for $\mathcal{F}$ along ξ. Let S be a subset of X_b such that $\xi(b) \in S$. Assume that $\mathcal{R} \cap S = \{\xi(b)\}$. Let C be an approximating cone for S at $\xi(b)$, such that C is not a linear subspace of $T_{\xi(b)}X_b$. Then there exist a nontrivial linear functional $z : T_{\xi(b)}X_b \mapsto \mathbb{R}$ and a compatible selection $\{L_{t,s}\}_{s,t \in T, s \preceq t}$ of $\mathbf{l}$ such that, if we define*

$$\zeta(t) = L_{b,t}^\dagger z \ \text{ for } \ t \in T\,, \tag{6.18}$$

then

$$z(v) \leq 0 \ \text{ for all } \ v \in C\,, \tag{6.19}$$

and

$$\zeta(t)(w) \geq 0 \ \text{ whenever } \ t \in T, (t, v) \in W\,. \tag{6.20}$$

7 Applications

Having stated the maximum principle in a rather abstract form, we now have to show how all the well known versions, as well as many new results, follow from our result. A detailed account is now in preparation. Here we shall limit ourselves to discussing a few examples.

To begin with, we observe that minimization problems of the kind usually studied in optimal control theory can be reduced to separation problems of the type discussed in Theorem 6.1. For example, suppose the variable x takes values in $\mathbb{R}^n$, and we want to minimize an integral $\int_a^b L(x(t), u(t), t)dt$, subject to a dynamical equation $\dot{x} = f(x, u, t)$, and boundary conditions $x(a) = \bar{x}$, $x(b) = \hat{x}$. We can then consider the new control system in $\mathbb{R}^{n+1}$ with dynamics $\dot{x} = f(x, u, t)$, $\dot{y} = L(x, u, t)$, in which the "running cost" has been added as an extra state variable. A trajectory-control pair $(\xi, u(\cdot))$ of the original system is a minimizer iff the "augmented" trajectory $\tilde{\xi} : [a,b] \mapsto \mathbb{R}^{n+1}$ given by $\tilde{\xi}(t) = (\xi(t), \int_a^t L(\xi(s), u(s))ds)$ satisfies a separation condition. Precisely, let $\mathcal{R}$ be the reachable set over the interval $[a,b]$ for the new system from the point $(\bar{x}, 0)$, and let S be the set $\{\hat{x}\}\times]-\infty, c]$, where $c = \int_a^b L(\xi(t), u(t))dt$. Then $(\xi, u(\cdot))$ is a minimizer iff $\mathcal{R} \cap S = \{(\hat{x}, c)\}$.

A similar reduction is possible for more general optimal control problems, with endpoint constraints of the form $x(a) \in A$, $x(b) \in B$, and even for constraints of the type $(x(a), x(b)) \in Q$, where Q is a subset of $\mathbb{R}^n \times \mathbb{R}^n$. As in the simple example of fixed endpoints. this reduction *always* leads to a set S whose approximating cone —if it exists— is not a linear subspace.

From now on, we take the reduction for granted, and concentrate on separation problems.

Consider first the "classical" case of a system $\dot{x} = f(x,u,t)$, $x \in \mathbb{R}^n$, $u \in U$, in which it is assumed that, for the reference control $u_* : [a,b] \mapsto U$ and corresponding trajectory $\xi_* : [a,b] \mapsto \mathbb{R}^n$, the time-varying vector field $(x,t) \mapsto f_*(x,t) \stackrel{\text{def}}{=} f(x, u_*(t), t))$ satisfies C^1-Carathéodory conditions in a tube $\mathcal{T} = \{(x,t) : a \leq t \leq b, ||x - \xi_*(t)|| \leq \varepsilon\}$. (This means that $f_*(x,t)$ is of class C^1 in x for each fixed t, measurable in t for each fixed x, and such that $||\frac{\partial f_*}{\partial x}(x,t)||$ is bounded by an integrable function of t.) Then f_* gives rise to a flow $\mathcal{F}$ consisting of C^1 maps, and its is then obvious that $\mathcal{F}$ is multidifferentiable along ξ_*, with a multidifferential $\mathbf{L} = \{\mathbf{L}_{t,s}\}$ consisting of singleton sets $\mathbf{L}_{t,s} = \{L_{t,s}\}$. Moreover, the maps $L_{t,s}$ are obtained by solving the variational equation

$$\dot{M}(t) = \frac{\partial f_*}{\partial x}(\xi_*(t), t).M(t)\,. \tag{7.21}$$

Precisely, $L_{t,s}$ is the value at t of the solution $M(\cdot)$ of (7.21) such that $M(s) =$ identity. The condition $z(t) = L_{b,t}^{\dagger} z(b)$, when differentiated, becomes the familiar adjoint equation. We can define a set of point variational vectors as follows: for each $(t_0, u_0) \in [a,b] \times U$, let

$$v(t_0, u_0) = f(\xi_*(t_0), u_0, t_0) - f_*(\xi_*(t_0), t_0)\,. \tag{7.22}$$

Then we would like to define a variation V_{t_0,u_0} by letting $V_{t_0,u_0}(\varepsilon, x)$ be the point obtained by starting at x at time t_0, following "the" trajectory of $\dot{\xi} = f(\xi, u_0, t)$ up to time $t_0 + \varepsilon$, and then going backwards along the trajectory of f_* up to time t. This will easily yield a nicely multidifferentiable variation with variational vector $(t_0, v(t_0, u_0))$, if f_* and the vector field $(x,t) \mapsto f(x, u_0, t)$ are continuous at (x_0, t_0). (Notice, however, that even in this case the map V_{t_0,u_0} can be set-valued, because we are not assuming that $f(x, u_0, t)$ is Lipschitz in x, so the trajectories need not be unique. The regularity of V_{t_0,x_0} follows from standard properties of solutions of ordinary differential equations, together with the fact that continuous vector fields can be approximated by smooth ones, so the corresponding flow maps can be approximated by single-valued ones.) So, in the special case when $f(x,u,t)$ is jointly continuous in (x,t) for each fixed u, and f_* is also jointly continuous, we get the Hamiltonian minimization condition in the strong form that $z(t).f(\xi_*(t),u,t) \geq z(t).f(\xi_*(t),u_*(t),t)$ for all (u,t). In the more general case when each vector field $(x,t) \mapsto f(x,u,t)$ is only assumed to be continuous in x, measurable in t, and bounded by an integrable function of t, then well known techniques, based on the Scorza-Dragoni theorem, make it possible to establish the nice multidifferentiability of V_{t_0,u_0} for $t_0 \in E \cap E(u_0)$, where E is a subset of $[a,b]$ of full measure, and $E(u_0)$ is a u_0-dependent subset of full measure. This yields the Hamiltonian minimization condition in the "weak" form:

$$(\forall u \in U)\Big(z(t).f(\xi_*(t),u,t) \geq z(t).f_*(\xi_*(t),t) \text{ for a.e. } t \in [a,b]\Big). \quad (7.23)$$

Under suitable separability conditions (e.g. if $f(x,u,t)$ is independent of t, or if there exists a countable subset U_0 of U such that $\{f(\xi_*(t),u,t) : u \in U_0\}$ is dense in $\{f(\xi_*(t),u,t) : u \in U\}$) we can go from this to the "strong" minimization condition

$$\min\{z(t).f(\xi_*(t),u,t) : u \in U\} = z(t).f_*(\xi_*(t),t) \text{ for a.e.} t \in [a,b]. \quad (7.24)$$

If the reference vector field f_* is only assumed to be Lipschitz with respect to x, with a Lipschitz constant bounded by an integrable function of t, then all the above considerations apply, except only that now the flow maps $F_{t,s}$ arising from f_* are just Lipschitz. A theorem of J. Warga ([24, 25, 26, 27]) allows us to find a multidifferential of this flow. All we have to do is to rewrite (7.21) in the form $\dot{M}(t) = A(t).M(t)$, where A is an arbitrary selection of the map $t \mapsto \partial f_*(\xi_*(t),t)$, and $\partial f_*(x,t)$ is the Clarke generalized Jacobian of $f_*(\cdot,t)$ at x. We then define $\mathbf{L}_{t,s}$, for $s \leq t$, to be the set of all $M(t)$, ranging over all measurable selections $A(\cdot)$, and all solutions $M(\cdot)$ for which $M(s) =$ identity. The resulting family $\mathbf{L}$ is a multidifferential of $\mathcal{F}$ along ξ_*. The needle variations are treated exactly as in the classical case, and we get the Lojasiewicz version of the nonsmooth maximum principle, in which the reference vector field is Lipschitz and the other ones are just continuous.

We now discuss Examples 1.3 and 1.4 in the light of Theorem 6.1. The reference vector field is given by $f(x) = |x|^\alpha \sin\frac{1}{x}$, where $\alpha > 1$, and the reference trajectory is $\xi_*(t) \equiv 0$. Given any trajectory $[0,1] \ni t \mapsto x(t)$ of f, Gronwall's inequality yields, to begin with, $|x(t)| \leq e^{t-s}|x(s)|$ for $0 \leq s \leq t \leq 1$, as long as $|x(t)| \leq 1$ for all $t \in [0,1]$. This implies that $|x(t)| \leq e^{t-s}|x(s)|$ whenever $0 \leq s \leq t \leq 1$, if $|x(s)| \leq e^{-1}$. But then $|x(t) - x(s)| \leq C|x(s)|^\alpha$, where $C > 0$ is a constant. This means that the flow maps $F_{t,s}$ satisfy $F_{t,s}(x) = x + o(|x|)$. In other words, the flow

$\mathcal{F}$ generated by our reference vector field is multidifferentiable along our reference trajectory, with multidifferential $\mathbf{L}$ given by $\mathbf{L}_{t,s} = \{\text{identity}\}$. This gives the "adjoint equation" $z(t) \equiv$ constant, providing a rigorous justification to the formal arguments of the introduction.

So far, we have only considered examples where the integrated form $z(s) = L^{\dagger}_{t,s}z(t)$ of the adjoint equation can be differentiated to yield an adjoint equation of the standard form. It should be clear from our discussion that in our approach there is no need for this differentiated form to exist. To illustrate this, we consider Examples 1.1 and 1.2. In Example 1.1, we can write the dynamics in control form by taking as control space the set U of all ordered pairs of unit vectors, and defining, for each $u = (u_1, u_2) \in U$, $f(x, y, u) = c_1u_1$ if $y \geq 0$, and $f(x, y, u) = c_2u_2$ if $y < 0$. Given A, B lying on opposite sides of the x axis, and a C on the x axis, the broken line ξ obtained by concatenating the segments from A to C and from C to B is an integral curve of $f(\cdot, u)$ for a unique constant control $u \in U$. Even though the vector field $f(\cdot, u)$ is discontinuous, the corresponding flow maps $F_{s,t}$ are well defined. If we choose the zero of the time axis so that $\xi(0) = A$, it is easy to see that, if τ denotes the time such that $\xi(\tau) = C$, then every map $F_{t,s}$ is affine linear near $\xi(s)$, as long as $s \neq \tau \neq t$. So the adjoint equation in integrated form makes perfect sense, and yields an adjoint vector $t \mapsto z(t)$ that is actually constant for $t < \tau$ and also for $t > \tau$, but jumps at $t = \tau$. A somewhat long but straightforward calculation yields the expected conclusion, namely, that for ξ to be optimal it has to satisfy the condition of Snell's law of refraction.

As for Example 1.4, the situation is similar, although the calculations are more complicated and the behavior of the adjoint vector is rather different. We omit the details, which are rather straightforward, and limit ourselves to pointing out that the main step is to choose a good way to realize the cycloids that are optimal trajectories in the upper and lower half planes in terms of flows of vector fields depending on a parameter u, so that the broken cycloids described in the introduction become trajectories of a vector field corresponding to a pair (u_1, u_2) of u-values, similarly to what we did in the case of the broken lines. It turns out that the adjoint equation in integrated form still makes sense, and has a solution $t \mapsto z(t)$ that goes to infinity as $t \to \tau$, where τ is the time when our broken cycloid ξ crosses the x axis. If, however, one takes the totally ordered set T of Theorem 6.1 to be the *complement* of $\{\tau\}$ in $\mathrm{Dom}(\xi)$, then z is perfectly well behaved there. The final result of the application of Theorem 6.1 is an extra necessary condition for an optimum: the optimal cycloids in the upper half plane, given as graphs of functions $x \mapsto y(x)$, satisfy $y(1 + \frac{dy}{dx}^2) = \mathit{constant} > 0$, and those in the lower half plane satisfy the same condition, except that now the constant is negative. The extra condition that links the two parts of an optimal broken cycloid $\xi : [0, b] \mapsto \mathbb{R}^2$ crossing the x axis is that the two constants have the same absolute value, i.e. that $|y|(1+\frac{dy}{dx}^2)$ remain constant throughout the interval $[0,b]$. (Notice that, as $t\to\tau$, $y(t)\to 0$ but $\frac{dy}{dx}^2\to\infty$, since ξ approaches the x axis vertically.)

To conclude our list of examples, we outline the results, described in more detail in our 1996 CDC paper [21], about a maximum principle for systems of differential inclusions. The setting here is a family $\mathbf{F}=\{F_u\}_{u\in U}$ of set-valued maps from $\mathbb{R}^n\times[a,b]$ to $2^{\mathbb{R}^n}$. We assume that each F_u has closed nonempty values and is almost lower semicontinuous (ALSC) and locally integrably lower bounded (LILB). (A map $F:\mathbb{R}^n\times[a,b]\mapsto 2^{\mathbb{R}^n}$ is LILB if for every compact subset K of $\mathbb{R}^n$ there is an integrable function $\rho:[a,b]\mapsto[0,\infty]$ such that $d(0,F(x,t))\le\rho(t)$ whenever $x\in K$, $t\in[a,b]$. We call F ALSC if $[a,b]$ can be expressed as a union $\cup_{j=0}^{\infty}J_j$ of subsets such that each J_j for $j>0$ is compact, J_0 has measure zero, and for each $j>0$ the restriction $F\lceil\mathbb{R}^n\times J_j$ is lower semicontinuous. Recall that if X, Y are metric spaces and $F\in SVM(X,Y)$, then F is *lower semicontinuous* if $F^{-1}(\Omega)$ is open in X whenever Ω is open in Y.) Also, we assume that $\mathbf{F}$ is closed under interval substitutions, i.e. that if I is a subinterval of $[a,b]$, $u_1,u_2\in U$, and we define G by $G(x,t)=F_{u_1}(x,t)$ if $t\in I$, and $G(x,t)=F_{u_2}(x,t)$ if $t\notin I$, then $G=F_u$ for some $u\in U$.

Suppose $\xi_*:[a,b]\mapsto\mathbb{R}^n$ is a trajectory of F_{u_*} for some $u_*\in U$. Assume, moreover, that F_{u_*} is *integrably pseudolipschitz* along ξ_*. (This means that there exist an integrable function $k(\cdot):[a,b]\mapsto[0,\infty]$ and an $\varepsilon>0$ such that, for almost all t, the inclusion

$$F(x,t)\cap\bar{B}(\dot{\xi}_*(t),\varepsilon)\subseteq F(x',t)+\bar{B}(0,k(t)||x'-x||) \tag{7.25}$$

holds whenever $||x-\xi_*(t)||\le\varepsilon$ and $||x'-\xi_*(t)||\le\varepsilon$. Here $\bar{B}(x,r)$ denotes the closed ball with center x and radius r.)

Under these hypotheses, one can show that the flow generated by the differential inclusion $\dot{x}\in F_{u_*}(x,t)$ is multidifferentiable along ξ_*, and one can describe a way to compute a multidifferential. (This requires, to begin with, the construction of an integrably Lipschitz vector field $(x,t)\mapsto f(x,t)$ which is a selection of the convexification of F_{u_*} and is such that $\dot{\xi}_*(t)=f(\xi_*(t),t)$ for a.e. t. Once this is done, the flow of f is clearly semidifferentiable, and a semidifferential $\mathbf{L}$ of this flow is given by Warga's theorem quoted above. Finally, to prove that the flow of F_{u_*} is multidifferentiable with multidifferential $\mathbf{L}$, one approximates the trajectories of f by trajectories of F_{u_*}, using a variant of the results of [11, 12] on uniform approximation of relaxed trajectories by non-relaxed ones.)

Having shown the multidifferentiability of the reference flow, the next step is to construct needle variations. For this purpose, we use a modification of the theory of "conically continuous" selections of lower semicontinuous set-valued maps, due to A. Bressan (cf. [4, 5, 6, 7, 8] and also Pucci [18]). The main point is that, if F is ALSC, then for almost every t_0 the following is true: for each $x_0 \in \mathbb{R}^n$ and each $y_0 \in F(x_0, t_0)$, there exists a Borel×Lebesgue-measurable selection f of F such that $f(x_0, t_0) = y_0$, f satisfies integral bounds $||f(x,t)|| \leq \psi_K(t)$, with $\psi_K \in L^1([a,b])$, on each compact subset K of $\mathbb{R}^n$, and f has the following continuity property: for each K, the map $\xi \mapsto \int_a^{\cdot} f(\xi(s), s)ds$ is continuous from the space A_K of all absolutely continuous $\xi : [a,b] \mapsto K$ such that $||\dot{\xi}(t)|| \leq \psi_K(t)$ for a.e. t to the space $C^0([a,b], \mathbb{R}^n)$ of continuous maps from $[a,b]$ to $\mathbb{R}^n$. (Both spaces are endowed with the sup norm.) It turns out that this continuity property suffices for the O.D.E. $\dot{x} = f(x,t)$ to have the same properties as O.D.E.'s with a continuous right-hand side. In particular, this O.D.E. has local existence of solutions, the flow maps $\Phi_{t,s}$ are upper semicontinuous and compactly valued for small enough $|t-s|$, and they can be approximated by continuous single-valued maps, obtained by regularizing f. Finally, f can also be chosen so that f is "almost continuous" at (x_0, t_0), in the sense that there exists a measurable subset E of $[a,b]$ such that t_0 is a point of density of E and $f\lceil \mathbb{R}^n \times E$ is continuous at (x_0, t_0). This suffices to construct, for every $u \in U$, and for almost all $t \in [a,b]$, needle variations generating the variational vectors (t, y) for all $y \in F_u(\xi_*(t), t)$. (The order of the quantifiers matters here. The bad set of measure zero may depend on u.) The compatibility condition is easily verified. The conclusion is that, if $\mathcal{R}$ denotes the reachable set from $\xi_*(a)$ over $[a,b]$, and $S \subseteq \mathbb{R}^n$ is such that $\xi_*(b) \in S$ and S has an approximating cone C at $\xi_*(b)$ which is not a linear subspace, then $\mathcal{R} \cap S = \{\xi_*(b)\}$ implies that for every locally integrably Lipschitz selection f of the convexification of F_{u_*} such that ξ is an integral curve of f there exists a nontrivial solution $t \mapsto z(t)$ of the adjoint inclusion $-\dot{z}(t) \in z(t)\partial_x f(\xi_*(t), t)$ such that, for each $u \in U$, we have

$$\inf \left\{ z(t).y : y \in F_u(\xi_*(t), t) \right\} \geq z(t).\dot{\xi}_*(t) \;\text{ for a.e. } t \in [a,b]\,. \tag{7.26}$$

Under suitable separability conditions the exceptional set of measure zero for which the conclusion of (7.26) fails can be taken to be independent of u, and one then gets the stronger conclusion that

$$\inf \left\{ z(t).y : y \in \bigcup_{u \in U} F_u(\xi_*(t), t) \right\} \geq z(t).\dot{\xi}_*(t) \;\text{ for a.e. } t \in [a,b]\,. \tag{7.27}$$

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] Agrachev, A. A. and R. V. Gamkrelidze, "The Morse index and the Maslov index for smooth control systems," *Doklady Akad. Nauk USSR* **287** (1986).

[4] Bressan, A., "On differential relations with lower continuous right-hand side. An existence theorem." *J. Diff. Equations* **37** (1980), pp. 89-97.

[5] Bressan, A., "Solutions of lower semicontinuous differential inclusions on closed sets," *Rend. Sem. Mat. Univ. Padova* **69** (1983), pp. 99-107.

[6] Bressan, A., "Directionally continuous selections and differential inclusions," *Funk. Ekvac.* **31** (1988), pp. 459-470.

[7] Bressan, A., "On the qualitative theory of lower semicontinuous differential inclusions," *J. Diff. Equations* **77** (1989), pp. 379-391.

[8] Bressan, A., "Upper and lower semicontinuous differential inclusions: a unified approach," in *Nonlinear Controllability and Optimal Control*, H. J. Sussmann Ed., M. Dekker Inc., New York, 1990, pp. 21-31.

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

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

[11] Colombo, R.M., A. Fryszkowski, T. Rzežuchowski, and V. Staicu, "Continuous selections of solution sets of Lipschitzean differential inclusions," *Funkcialaj Ekvacioj* **34** (1991), 321-330.

[12] Fryszkowski, A., and T. Rzežuchowski, "Continuous version of the Filippov-Ważewski relaxation theorem," *J. Diff. Equations* **94** (1993), pp. 254-265.

[13] Kelley, H. J., R. E. Kopp and H. G. Moyer, "Singular extremals," in *Topics in Optimization*, G. Leitman ed., Academic Press (1967).

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

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

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

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

[18] Pucci, A., "Sistemi di equazione differenziale con secondo membro discontinuo rispetto all'incognita," *Rend. Ist. Mat. Univ. Trieste* **3** (1971), pp. 75-80

[19] 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.

[20] 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.

[21] Sussmann, H.J., "A strong maximum principle for systems of differential inclusions," to appear in the proceedings of the 35th IEEE Conference on Decision and Control, to be held in Kobe, Japan, Dec. 1996.

[22] Sussmann, H.J., "The maximum principle of optimal control theory," in preparation.

[23] Tougeron, J. C., *Idéaux de fonctions différentiables*, Springer-Verlag, New York, 1972.

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

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

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

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

[28] Whitney, H., "Analytic extensions of differentiable functions defined in closed sets," *Trans. Amer. Math. Soc.* **36** (1934), 63-89.