Submitted for presentation at the 39th IEEE Conference on Decision and Control (CDC) , to be held in Sydney, Australia, December 12-15, 2000, and for publication in the conference proceedings. The date of this version is March 12, 2000. The version submitted to the CDC is an earlier one. In the present version, some typos that were found in the submitted version have been corrected.

New theories of set-valued differentials and new versions of the maximum principle of optimal control theory *

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

The purpose of this note is to announce a new theory of generalized differentials—the "generalized differential quotients," abbr. GDQs— which has good open mapping properties, and to use this theory to state—in Theorem 9.4— a version of the maximum principle for hybrid optimal control problems under weak regularity conditions. For single-valued maps, our GDQ theory essentially coincides with the one proposed by H. Halkin in [4], but GDQ theory applies as well to multivalued maps, thus making it possible to deal with non-Lipschitz vector fields, whose flow maps are in general set-valued.

The results presented here are much weaker than what can actually be proved by our methods. More general versions, involving systems of differential inclusions, are discussed in a detailed paper currently in preparation.

The GDQ concept contains several other notions of generalized differential, but does not include some important theories such as J. Warga's "derivate containers" (cf. [9]) and the "semidifferentials" and "multidifferentials" proposed by us in previous work (cf. [7]).

Recently, we have found another theory—the "path-integral generalized differentials", abbr. PIGDs—that contains that of GDQs as well as the other theories mentioned above. We conclude the paper by giving, in §10, a brief sketch of the definition of this concept.

2. Notational preliminaries

A *set-valued map* is a triple $F = (A, B, G)$ such that A and B are sets and G is a subset of $A \times B$. If $F = (A, B, G)$ is a set-valued map, we say that F *is a set-valued map from* A *to* B. In that case, we refer to the sets A, B, G as the *source*, *target*, and *graph* of F, respectively, and write $A = \mathrm{So}(F)$, $B = \mathrm{Ta}(F)$, $G = \mathrm{Gr}(F)$. If $x \in \mathrm{So}(F)$, we write $F(x) = \{y : (x, y) \in \mathrm{Gr}(F)\}$. The set $\mathrm{Do}(F) = \{x \in \mathrm{So}(F) : F(x) \neq \emptyset\}$ is the *domain* of F. If A, B are sets, we use $SVM(A, B)$ to denote the set of all set-valued maps from A to B, and write $F : A \longrightarrow\!\!\!\!\rightarrow B$ to indicate that $F \in SVM(A, B)$.

The expression "ppd map" stands for "possibly partially defined (that is, not necessarily everywhere defined) ordinary (that is, single-valued) map," and we write $f : A \dashrightarrow B$ to indicate that f is a ppd map from A to B. A *time-varying ppd map* from a set A to a set B is a ppd map from $A \times \mathbb{R}$ to B.

If $k \in \mathbb{N}$ and M is a manifold of class C^k, then TM, T^*M denote the tangent and cotangent bundle of M. If $x \in M$, then T_xM, T^*_xM denote the tangent and cotangent spaces of M at x.

*Research supported in part by NSF Grant DMS98-03411-00798 and AFOSR Grant 0923.

3. Regular maps

If X, Y are metric spaces, then $SVM_{comp}(X, Y)$ will denote the subset of $SVM(X, Y)$ whose members are the set-valued maps from X to Y that have a compact graph. We say that a sequence $\{F_j\}_{j\in\mathbb{N}}$ of members of $SVM_{comp}(X, Y)$ *inward graph-converges* to an $F \in SVM_{comp}(X, Y)$—and write $F_j \xrightarrow{\text{igr}} F$—if for every open subset Ω of $X \times Y$ such that $\mathrm{Gr}(F) \subseteq \Omega$ there exists a $j_\Omega \in \mathbb{N}$ such that $\mathrm{Gr}(F_j) \subseteq \Omega$ whenever $j \geq j_\Omega$.

Definition 3.1 Assume that X, Y are metric spaces. A *regular set-valued map* from X to Y is a set-valued map $F \in SVM(X, Y)$ such that

- for every compact subset K of X, the restriction $F \lceil K$ of F to K belongs to $SVM_{comp}(K, Y)$ and is a limit—in the sense of inward graph-convergence—of a sequence of continuous single-valued maps from K to Y.

We use $\mathrm{REG}(X; Y)$ to denote the set of all regular set-valued maps from X to Y. ♢

It is easy to see that if $F : X \to Y$ is an ordinary (that is, single-valued and everywhere defined) map, then F belogns to $\mathrm{REG}(X; Y)$ if and only if F is continuous.

It is not hard to prove the following

Theorem 3.2 *Let X, Y, Z be metric spaces, and suppose that $F \in \mathrm{REG}(X; Y)$ and $G \in \mathrm{REG}(Y; Z)$. Then the composite map $G \circ F$ belongs to* $\mathrm{REG}(X; Z)$. ♢

4. Generalized differential quotients (GDQs)

Definition 4.1 Let m, n be nonnegative integers, let $F : \mathbb{R}^m \longrightarrow\!\!\!\!\rightarrow \mathbb{R}^n$ be a set-valued map, and let Λ be a nonempty compact subset of $\mathbb{R}^{n\times m}$. Let S be a subset of $\mathbb{R}^m$. We say that Λ is a *generalized differential quotient* (abbreviated "GDQ") *of F at $(0, 0)$ in the direction of S*, and write $\Lambda \in GDQ(F; 0, 0; S)$, if for every positive real number δ there exist U, G such that

1. U is a compact neighborhood of 0 in $\mathbb{R}^m$ and $U \cap S$ is compact;
2. G is a regular set-valued map from $U \cap S$ to the δ-neighborhood Λ^δ of Λ in $\mathbb{R}^{n\times m}$;
3. $G(x) \cdot x \subseteq F(x)$ for every $x \in U \cap S$. ♢

If M, N are C^1 manifolds, $\bar{x} \in M$, $\bar{y} \in N$, $S \subseteq M$, and $F : M \longrightarrow\!\!\!\!\rightarrow N$, then we can define a set $GDQ(F; \bar{x}, \bar{y}; S)$ of compact nonempty subsets of the space $\mathrm{Lin}(T_{\bar{x}}M, T_{\bar{y}}N)$ of linear maps from $T_{\bar{x}}M$ to $T_{\bar{y}}N$ by picking coordinate charts $M \ni x \to \xi(x) \in \mathbb{R}^m$,

$N \ni y \to \eta(y) \in \mathbb{R}^n$—where $m = \dim M$, $n = \dim N$—defined near $\bar{x}$, $\bar{y}$ and such that $\xi(x) = 0$, $\eta(y) = 0$, and declaring a subset Λ of $\mathrm{Lin}(T_{\bar{x}}M, T_{\bar{y}}N)$ to belong to $GDQ(F; \bar{x}, \bar{y}; S)$ if $D\eta(\bar{y}) \circ \Lambda \circ D\xi(\bar{x})^{-1}$ is in $GDQ(\eta \circ F \circ \xi^{-1}; 0, 0; \xi(S))$. It turns out that, with this definition, *the set $GDQ(F; \bar{x}, \bar{y}; S)$ does not depend on the choice of the charts ξ, η.* In other words, *the notion of a GDQ is invariant under C^1 diffeomorphisms and makes sense intrinsically on C^1 manifolds.*

The following facts about GDQs can be verified.

1. If M, N are C^1 manifolds, $\bar{x} \in M$, U is a neighborhood of $\bar{x}$ in M, $F : U \to N$ is a continuous map, F is differentiable at $\bar{x}$, $\bar{y} = F(\bar{x})$, and $L = DF(\bar{x})$, then $\{L\} \in GDQ(F; \bar{x}, \bar{y}; M)$.
2. If M, N are C^1 manifolds, $\bar{x} \in M$, U is a neighborhood of $\bar{x}$ in M, $F : U \to N$ is a Lipschitz continuous map, $\bar{y} = F(\bar{x})$, and Λ is the Clarke generalized Jacobian of F at $\bar{x}$, then $\Lambda \in GDQ(F; \bar{x}, \bar{y}; M)$.
3. *(The chain rule.)* If M_i is a C^1 manifold and $\bar{x}_i \in M_i$ for $i = 1, 2, 3$, $S_i \subseteq M_i$, $F_i : M_i \longrightarrow\!\!\!\!\!\rightarrow M_{i+1}$, and $\Lambda_i \in GDQ(F_i; \bar{x}_i, \bar{x}_{i+1}; S_i)$ for $i = 1, 2$, and either $S_2 \cap U$ is a retract of U for some comnpact neighborhood U of x_2 in M_2 or F_1 is single-valued, then $\Lambda_2 \circ \Lambda_1 \in GDQ(F_2 \circ F_1; \bar{x}_1, \bar{x}_3; S_1)$.
4. *(The product rule.)* If M_1, M_2, N_1, N_2, are C^1 manifolds, and, for $i = 1, 2$, $\bar{x}_i \in M_i$, $\bar{y}_i \in N_i$, $S_i \subseteq M_i$, $F_i : M_i \longrightarrow\!\!\!\!\!\rightarrow N_i$, and $\Lambda_i \in GDQ(F_i; \bar{x}_i, \bar{y}_i; S_i)$, then $\Lambda_1 \times \Lambda_2 \in GDQ(F_1 \times F_2; (\bar{x}_1, \bar{x}_2), (\bar{y}_1, \bar{y}_2); S_1 \times S_2)$.
5. *(Locality.)* If M, N, are C^1 manifolds, $\bar{x} \in M$, $\bar{y} \in N$, and, for $i = 1, 2$, $S_i \subseteq M$, $F_i : M \longrightarrow\!\!\!\!\!\rightarrow N$, and there exist neighborhoods U, V of $\bar{x}$, $\bar{y}$, in M, N, respectively, such that $U \cap S_1 = U \cap S_2$ and $(U \times V) \cap \mathrm{Gr}(F_1) = (U \times V) \cap \mathrm{Gr}(F_2)$, then $GDQ(F_1; \bar{x}, \bar{y}; S_1) = GDQ(F_2; \bar{x}, \bar{y}; S_2)$.

It is easy to exhibit maps that have GDQs at a point $\bar{x}$ but are not classically differentiable at $\bar{x}$ and do not have differentials at $\bar{x}$ in the sense of other theories such as Clarke's generalized Jacobians, Warga's derivate containers, or our "semidifferentials" and "multidifferentials". (A simple example is provided by the function $f : \mathbb{R} \to \mathbb{R}$ given by $f(x) = x \sin 1/x$ if $x \neq 0$, and $f(0) = 0$. The set $[-1, 1]$ belongs to $GDQ(f; 0, 0, \mathbb{R})$, but is not a differential of f at 0 in the sense of any of the other theories.)

In addition, GDQs have the following *directional open mapping property.*

Theorem 4.2 *Let m, n be nonnegative integers, and let C be a convex cone in $\mathbb{R}^m$. Let $F : \mathbb{R}^m \longrightarrow\!\!\!\!\!\rightarrow \mathbb{R}^n$ be a set-valued map, and let $\Lambda \in GDQ(F; 0, 0; C)$. Let D be a closed convex cone in $\mathbb{R}^n$ such that $D \subseteq \mathrm{Int}(LC) \cup \{0\}$ for every $L \in \Lambda$. Then there exist a convex cone Δ in $\mathbb{R}^n$ such that $D \subseteq \mathrm{Int}(\Delta)$, and positive constants $\bar{\varepsilon}$, κ, having the property that*

(I) if $y \in \Delta$ and $\|y\| \leq \bar{\varepsilon}$, then there exists an $x \in C$ such that $\|x\| \leq \kappa\|y\|$ and $y \in F(x)$.

Moreover, the cone Δ and the constants $\bar{\varepsilon}$, κ can be chosen so that the following stronger conclusion holds:

(II) if $y \in \Delta$ and $\|y\| = \varepsilon \leq \bar{\varepsilon}$ then there exists a compact connected subset Z_y of $(C \cap \bar{\mathbb{B}}^n(\kappa\varepsilon)) \times [0, 1]$ such that $(0, 0) \in Z_y$, $(x, 1) \in Z_y$ for some x belonging to $C \cap \bar{\mathbb{B}}^n(\kappa\varepsilon)$, and $ry \in F(x)$ whenever $0 \leq r \leq 1$ and $(x, r) \in Z_y$. ♢

Associated to concept of a GDQ there is a notion of "GDQ approximating multicone to a set at a point."

Definition 4.3 If X is a finite-dimensional real linear space, a *convex multicone* in X is a nonempty set of convex cones in X.

If M is a manifold of class C^1, $S \subseteq M$ and $x \in S$, a *GDQ approximating multicone to S at x* is a convex multicone $\mathcal{C}$ in T_xM such that there exist $m \in \mathbb{Z}_+$, $F : \mathbb{R}^m \longrightarrow\!\!\!\!\!\rightarrow M$, a closed convex cone D in $\mathbb{R}^m$, and a $\Lambda \in GDQ(F; 0, x; D)$, such that $F(D) \subseteq S$ and $\mathcal{C} = \{LD : L \in \Lambda\}$. ♢

If X is a finite-dimensional real linear space, then $X^\perp$ denotes the dual of X. If $C \subseteq X$, the *polar* of C is the set $C^\perp = \{y \in X^\perp : y(x) \leq 0 \text{ whenever } x \in C\}$, so $C^\perp$ is a closed convex cone in $X^\perp$. If $\mathcal{C}$ is a convex multicone in X, the *polar* of $\mathcal{C}$ is the set $\mathcal{C}^\perp = \mathrm{Clos}\Big(\bigcup\{C^\perp : C \in \mathcal{C}\}\Big)$, so $\mathcal{C}^\perp$ is a (not necessarily convex) closed cone in $X^\perp$.

5. Discontinuous vector fields and their flows

If $n \in \mathbb{N}$, we use $\mathcal{B}(\mathbb{R}^n)$, $\mathcal{BL}(\mathbb{R}^n, \mathbb{R})$, to denote, respectively, the σ-algebra of Borel subsets of $\mathbb{R}^n$ and the product σ-algebra $\mathcal{B}(\mathbb{R}^n) \otimes \mathrm{Lebesgue}(\mathbb{R})$. We let $\mathcal{N}(\mathbb{R}^n, \mathbb{R})$ denote the set of all subsets S of $\mathbb{R}^n \times \mathbb{R}$ such that $\Pi_n(S)$ is a Lebesgue-null subset of $\mathbb{R}$, where Π_n is the canonical projection $\mathbb{R}^n \times \mathbb{R} \ni (x, t) \to t \in \mathbb{R}$. Finally, we use $\mathcal{BL}_e(\mathbb{R}^n, \mathbb{R})$ to denote the σ-algebra of subsets of $\mathbb{R}^n \times \mathbb{R}$ generated by $\mathcal{BL}(X, \mathbb{R}) \cup \mathcal{N}(X, \mathbb{R})$. It is then clear that $\mathcal{B}(\mathbb{R}^n \times \mathbb{R}) \subset \mathcal{BL}(\mathbb{R}^n, \mathbb{R}) \subset \mathcal{BL}_e(\mathbb{R}^n, \mathbb{R})$, and both inclusions are strict.

Definition 5.1 Let $n, m \in \mathbb{Z}_+$, and let f be a ppd map from $\mathbb{R}^n \times \mathbb{R}$ to $\mathbb{R}^m$.

1. We say that f is *locally essentially Borel×Lebesgue* measurable, or *locally $\mathcal{BL}_e(\mathbb{R}^n, \mathbb{R})$-measurable*, if $f^{-1}(U) \cap K \in \mathcal{BL}_e(\mathbb{R}^n, \mathbb{R})$ for all open subsets U of $\mathbb{R}^m$ and all compact subsets K of $\mathrm{Do}(f)$.
2. We use $\mathcal{M}^{e,loc}_{\mathcal{BL}}(\mathbb{R}^n \times \mathbb{R}; \mathbb{R}^m)$ to denote the set of all locally $\mathcal{BL}_e(\mathbb{R}^n, \mathbb{R})$-measurable ppd maps from $\mathbb{R}^n \times \mathbb{R}$ to $\mathbb{R}^m$.
3. We call f *locally integrably bounded* (LIB) if for every compact subset K of $\mathrm{Do}(f)$ there exists an integrable funtion $\mathbb{R} \ni t \to \varphi(t) \in [0, +\infty]$ such that $||f(x, t)|| \leq \varphi(t)$ for all $(x, t) \in K$. ♢

If $n, m \in \mathbb{Z}_+$, $f : \mathbb{R}^n \times \mathbb{R} \dashrightarrow \mathbb{R}^m$, I is a compact interval, and $S \subseteq \mathrm{Do}(f)$, we write $\Xi_S(f, I)$ to denote the set of all curves $\xi \in C^0(I; \mathbb{R}^n)$ such that $(\xi(t), t) \in S$ for all $t \in I$. We write $\Xi(f, I)$ for $\Xi_{\mathrm{Do}(f)}(f, I)$,

If $\psi : \mathbb{R} \to \mathbb{R}_+ \cup \{+\infty\}$ is a function, we use $\Xi^\psi_S(f, I)$ to denote the set of all curves $\xi \in \Xi_S(f, I)$ such that

$$\limsup_{t \downarrow \bar{t}} \frac{||\xi(t) - \xi(\bar{t})||}{t - \bar{t}} \leq \psi(\bar{t}) \tag{1}$$

for almost every $\bar{t} \in I$.

Fact 5.2 *If $n, m \in \mathbb{Z}_+$, $f \in \mathcal{M}^{e,loc}_{\mathcal{BL}}(\mathbb{R}^n \times \mathbb{R}; \mathbb{R}^m)$, I is a compact interval, and $\xi \in \Xi(f, I)$, then $f \circ \xi$ is measurable.* ♢

If $n, m \in \mathbb{Z}_+$, $f \in \mathcal{M}^{e,loc}_{\mathcal{BL}}(\mathbb{R}^n \times \mathbb{R}; \mathbb{R}^m)$, I is a compact interval, and f is locally integrably bounded, then we can define a map $\mathcal{T}_{f,I} : I \times I \times \Xi(f, I) \to \mathbb{R}^m$ by letting

$$\mathcal{T}_{f,I}(a, t, \xi) = \int_a^t f(\xi(s), s)\, ds \tag{2}$$

if $a \in I$, $t \in I$, $\xi \in \Xi(f, I)$.

Definition 5.3 If $n, m \in \mathbb{Z}_+$ and $f : \mathbb{R}^n \times \mathbb{R} \dashrightarrow \mathbb{R}^m$, we call f *locally integrally continuous* if

1. $f \in \mathcal{M}^{e,loc}_{\mathcal{BL}}(\mathbb{R}^n \times \mathbb{R}; \mathbb{R}^m)$,
2. For every compact subset K of $\mathrm{Do}(f)$ and every compact interval I there exists an integrable funtion $\mathbb{R} \ni t \to \psi(t) \in [0, +\infty]$ such that $||f(x,t)|| \le \psi(t)$ for all $(x,t) \in K$ and the restriction to $\Xi^{\psi}_K(f, I)$ of the map $\mathcal{T}_{f,I}$ is continuous. $\diamondsuit$

By taking coordinate charts, it is easy to see that the concept of a "locally integrally continuous time-varying ppd section $f : M \times \mathbb{R} \dashrightarrow E$" is well defined if M is a manifold and E is a vector bundle over M.

Definition 5.4 Let M be a manifold of class C^1.

1. A *time-varying vector field* (abbreviated "TVVF") on M is a ppd map $f : M \times \mathbb{R} \dashrightarrow TM$ such that $f(x,t) \in T_xM$ whenever $(x,t) \in \mathrm{Do}(f)$.
2. We use $TVVF(M)$ to denote the set of all TVVFs on M.
3. If $f \in TVVF(M)$, a *trajectory* (or *integral curve*) of f is a locally absolutely continuous map $\xi : I \to M$, defined on a nonempty subinterval I of $\mathbb{R}$, such that the conditions $(\xi(t), t) \in \mathrm{Do}(f)$ and $\dot{\xi}(t) = f(\xi(t), t)$ are satisfied for a.e. $t \in I$.
4. We use $\mathrm{Traj}\,(f)$ to denote the set of all trajectories of f, and $\mathrm{Traj}_c(f)$ to denote the set of all $\xi \in \mathrm{Traj}\,(f)$ whose domain is a compact interval.
5. For $(x, b, a) \in M \times \mathbb{R} \times \mathbb{R}$, we define
$$\Phi^f(x, b, a) = \{\, \xi(b) : \xi \in \mathrm{Traj}\,(f) : \xi(a) = x \,\}\,. \quad (3)$$
The map $\Phi^f : M \times \mathbb{R} \times \mathbb{R} \longrightarrow\!\!\!\!\to M$ is the *flow* of f.
6. For each $(b, a) \in \mathbb{R} \times \mathbb{R}$, we define $\Phi^f_{b,a} : M \longrightarrow\!\!\!\!\to M$ by letting $\Phi^f_{b,a}(x) = \Phi^f(x, b, a)$ for $x \in M$. The maps $\Phi^f_{b,a} : M \longrightarrow\!\!\!\!\to M$ are the *flow maps* of f. $\diamondsuit$

Fact 5.5 *Let M be a manifold of class C^1, and assume $f \in TVVF(M)$. Then the flow maps $\Phi^f_{b,a}$ satisfy the identities $\Phi^f_{a,a} = \mathbb{1}_M$ and $\Phi^f_{c,b} \circ \Phi^f_{b,a} = \Phi^f_{c,a}$, if $a, b, c \in \mathbb{R}$ and $a \le b \le c$.* $\diamondsuit$

6. Approximate limits

If $n, m \in \mathbb{Z}_+$, f is a ppd map from $\mathbb{R}^n \times \mathbb{R}$ to $\mathbb{R}^m$, $\beta > 0$, $\bar{x} \in \mathbb{R}^n$, and $\mathbf{v}$ is a nonempty subset of $\mathbb{R}^m$, we define
$$\sigma_{f,\bar{x},\mathbf{v},\beta}(t) = \sup\{\, \mathrm{dist}(f(x,t), \mathbf{v}) : x \in \mathbb{R}^n\,,\ \|x - \bar{x}\| \le \beta \,\}\,,$$
so $\sigma_{f,\bar{x},\mathbf{v},\beta}$ is a function on $\mathbb{R}$ with values in $[0, \infty]$. (We take the value of the right-hand side to be zero if the set of those $x \in \mathbb{R}^n$ such that $\|x - \bar{x}\| \le \beta$ and $f(x,t)$ is defined is empty.)

If I is a subinterval of $\mathbb{R}$, $\bar{x} \in \mathbb{R}^n$, and $\beta > 0$, we define
$$S_{\bar{x},I,\beta} = \{(x,t) : x \in \mathbb{R}^n, \|x - \bar{x}\| \le \beta, t \in I\}\,. \quad (4)$$

Fact 6.1 *Assume that $n, m \in \mathbb{Z}_+$, f belongs to $\mathcal{M}^{e,loc}_{\mathcal{BL}}(\mathbb{R}^n \times \mathbb{R}; \mathbb{R}^m)$, $x \in \mathbb{R}^n$, $\mathbf{v} \subseteq \mathbb{R}^m$, and $\mathbf{v} \ne \emptyset$. Let I be a nonempty subinterval of $\mathbb{R}$, and assume that $\bar{\beta} > 0$ and $S_{x,I,\bar{\beta}} \subseteq \mathrm{Do}(f)$. Then $\sigma_{f,\bar{x},\mathbf{v},\beta}$ is measurable on I whenever $0 < \beta \le \bar{\beta}$.* $\diamondsuit$

We let $\mathcal{C}(1)$ be the set of all cones in $\mathbb{R}$. Then $\mathcal{C}(1)$ has exactly four members, and "$C \in \mathcal{C}(1)$" is an alternative way of saying that C is one of the four sets $\{0\}$, $[0, +\infty\,[$, $]-\infty, 0]$, $\mathbb{R}$. If $C \in \mathcal{C}(1)$, $\bar{t} \in \mathbb{R}$, and $h > 0$, we use $\bar{t} + C(h)$, $\bar{t} + C$, to denote the sets $\{\bar{t} + r : r \in C, |r| \le h\}$, $\{\bar{t} + r : r \in C\}$, respectively.

Definition 6.2 Assume that $n, m \in \mathbb{Z}_+$, f belongs to $\mathcal{M}^{e,loc}_{\mathcal{BL}}(\mathbb{R}^n \times \mathbb{R}; \mathbb{R}^m)$, $\mathbf{v} \subseteq \mathbb{R}^m$, $\mathbf{v} \ne \emptyset$, $(\bar{x}, \bar{t}) \in \mathbb{R}^n \times \mathbb{R}$. Let $C \in \mathcal{C}(1)$.

1. We say that $\mathbf{v}$ is an *approximate limit set of f at $(\bar{x}, \bar{t})$ along C*, and write
$$\mathbf{v} \in \text{App–lim}_{x \to \bar{x}\,,\, t \to \bar{t}, t - \bar{t} \in C} f(x,t)\,,$$
if $(\bar{x}, \bar{t}) \in \mathrm{Int}_{\mathbb{R}^n \times (\bar{t} + C)}(\mathrm{Do}(f))$ and
$$\lim_{(h,\beta) \to (0,0)\,,\ h > 0,\ \beta \ge 0} \frac{1}{h} \int_{\bar{t} + C(h)} \sigma_{f,\bar{x},\mathbf{v},\beta}(t)\, dt = 0\,.$$
2. We say that a vector v is an *approximate limit value of f at $(\bar{x}, \bar{t})$ along C*, and write
$$v = \text{app–lim}_{x \to \bar{x}\,,\, t \to \bar{t}, t - \bar{t} \in C} f(x,t)\,,$$
if $\{v\} \in \text{App–lim}_{x \to \bar{x}\,,\, t \to \bar{t}, t - \bar{t} \in C} f(x,t)$. $\diamondsuit$

We say that $(\bar{x}, \bar{t})$ is a *point of approximate continuity along C* if $f(\bar{x}, \bar{t}) = \text{app–lim}_{x \to \bar{x}, t \to \bar{t}, t - \bar{t} \in C} f(x,t)$.

We use the expressions "approximate right limit," "approximate left limit," "approximate limit," "point of approximate right continuity," "point of approximate left continuity," "point of approximate continuity," respectively, as alternative names of "approximate limit along $[0, +\infty\,[$," "approximate limit along $]-\infty, 0]$," "approximate limit along $\mathbb{R}$," "point of approximate continuity along $[0, +\infty\,[$," "point of approximate continuity along $]-\infty, 0]$," and "point of approximate continuity along $\mathbb{R}$." By taking coordinate charts, it is easy to see that *all these concepts are well defined if M is a manifold, E is a vector bundle over M, and f is a time-varying ppd section $f : M \times \mathbb{R} \dashrightarrow E$.*

The following lemma gives a useful sufficient condition for a point (x,t) to be a point of approximate continuity of a time-varying vector field. To state the lemma, we first introduce the obvious one-sided analogues of the usual notions of a Lebesgue point and a point density.

- if $I \subseteq \mathbb{R}$ is an interval, and $C \in \mathcal{C}(1)$, then a *C-Lebesgue point* of a locally integrable function $\varphi : I \to \mathbb{R} \cup \{+\infty\}$ is a point $\bar{t} \in I$ such that $|\varphi(\bar{t})| < \infty$, $\bar{t} + C(\bar{h}) \subseteq I$ for some positive number $\bar{h}$, and
$$\lim_{h \to 0+} \frac{1}{h} \int_{\bar{t} + C(h)} \Big|\varphi(t) - \varphi(\bar{t})\Big|\, dt = 0\,.$$
- If $E \subseteq \mathbb{R}$ is a measurable set, a *point of C-density* of E is a point $\bar{t} \in E$ such that
$$\lim_{h \to 0+} h^{-1} \mathrm{meas}\Big((\bar{t} + C(h)) \backslash E\Big) = 0\,.$$

Lemma 6.3 Assume that $n, m \in \mathbb{Z}_+$, $C \in \mathcal{C}(1)$, $C \ne \{0\}$, $(\bar{x}, \bar{t}) \in \mathbb{R}^n \times \mathbb{R}$, and $f \in \mathcal{M}^{e,loc}_{\mathcal{BL}}(\mathbb{R}^n \times \mathbb{R}; \mathbb{R}^m)$. Assume, moreover, that $(\bar{x}, \bar{t})$ belongs to $\mathrm{Int}_{\mathbb{R}^n \times (\bar{t} + C)}(\mathrm{Do}(f))$, and there exist positive numbers $\bar{h}, \bar{\beta}$, an integrable function $\varphi : \bar{t} + C(\bar{h}) \to \mathbb{R} \cup \{+\infty\}$, and a measurable subset E of $\bar{t} + C(\bar{h})$, such that

(a) the set $S = S_{\bar{x}, \bar{t} + C(\bar{h}), \bar{\beta}}$ is contained in $\mathrm{Do}(f)$,
(b) $||f(x,t)|| \le \varphi(t)$ for all $(x,t) \in S$,
(c) $\bar{t}$ is a C-Lebesgue point of φ,
(d) $\bar{t}$ is a point of C-density of E,
(e) $\lim_{x \to \bar{x}, t \to \bar{t}, t \in E \cap (\bar{t} + C)} f(x,t) = f(\bar{x}, \bar{t})$.

Then $(\bar{x}, \bar{t})$ is a point of approximate continuity of f. $\diamondsuit$

7. Variational generators

If $\xi : [a,b] \to \mathbb{R}^n$ is a continuous curve, and $\alpha > 0$, we use $\mathcal{T}^n(\xi,\alpha)$ to denote the "α-tube about ξ in $\mathbb{R}^n$," that is, the set

$$\mathcal{T}^n(\xi,\alpha) \stackrel{\text{def}}{=} \{(x,t) : x \in \mathbb{R}^n, a \le t \le b, \|x - \xi(t)\| \le \alpha\}\,.$$

Definition 7.1 Let $n, m \in \mathbb{Z}_+$, and let f be a ppd map from $\mathbb{R}^n \times \mathbb{R}$ to $\mathbb{R}^m$. Let a, b, ξ be such that $a, b \in \mathbb{R}$, $a \le b$, and $\xi \in C^0(\,[a,b]\,;\mathbb{R}^n\,)$. A *variational generator for f about ξ* is a measurable set-valued map $\Lambda : [a,b] \longrightarrow\!\!\!\!\!\rightarrow \mathbb{R}^{m\times n}$ with compact convex nonempty values such that there exist k_Λ, $\bar\alpha$, $\mathbf{k}$ having the following three properties:

1. $k_\Lambda : [a,b] \to [0,+\infty]$ is integrable and such that $\sup\Big\{\|L\| : L \in \Lambda(t)\Big\} \le k_\Lambda(t)$ for all $t \in [a,b]$;
2. $\bar\alpha > 0$ and $\mathcal{T}^n(\xi,\bar\alpha) \subseteq \mathrm{Do}(f)$;
3. $\mathbf{k} = \{k^\alpha\}_{0<\alpha\le\bar\alpha}$ is a family of Lebesgue-integrable functions $k^\alpha : [a,b] \to [0,+\infty]$, for $0 < \alpha \le \bar\alpha$, such that $\lim_{\alpha\downarrow 0}\int_a^b k^\alpha(t)dt = 0$ and

$$\sup\Big\{\min\Big\{\|\Delta^f_\xi(x,t,L)\| : L\in\Lambda(t)\Big\} : \|x-\xi(t)\| \le \alpha\Big\} \le \alpha k^\alpha(t)$$

for all $t \in [a,b]$ and all $\alpha \in]\,0,\bar\alpha]$, where

$$\Delta^f_\xi(x,t,L) \stackrel{\text{def}}{=} f(x,t) - f(\xi(t),t) - L\cdot(x-\xi(t))\,.$$

We use $VG(f,\xi)$ to denote the set of all variational generators of f about ξ. ♢

If n, m, f, a, b, ξ, Λ are as in Definition 7.1, we use $\Gamma(\Lambda)$ to denote the set of all measurable single-valued selections of Λ. Then $\Gamma(\Lambda)$ is a nonempty convex weakly compact subset of $L^1([a,b];\mathbb{R}^{m\times n})$.

We now specialize to the case when $m = n$. If L belongs to $L^1([a,b];\mathbb{R}^{n\times n})$, we let M_L be the fundamental matrix solution of the linear time-varying equation $\dot M = L(t)\cdot M$. That is, M_L is a continuous map from $[a,b]\times[a,b]$ to $\mathbb{R}^{n\times n}$ that satisfies

$$M_L(t,s) = \mathbb{1}_{\mathbb{R}^n} + \int_s^t L(r)\cdot M_L(r,s)\,dr\,. \tag{5}$$

Fact 7.2 *If $\mathcal{B} \subseteq L^1([a,b],\mathbb{R}^{n\times n})$ and $\mathcal{B}$ is bounded, then the map $\mathcal{B}_{weak} \ni L \to M_L \in C^0(\,[a,b]\times[a,b]\,;\mathbb{R}^{n\times n}\,)$ is continuous.* ♢

If n, m, f, a, b, ξ, Λ are as in Definition 7.1, and $m = n$, we define

$$\mathcal{M}(\Lambda) \stackrel{\text{def}}{=} \{\,M_L : L \in \Gamma(\Lambda)\,\} \subseteq C^0(\,[a,b]\times[a,b]\,;\mathbb{R}^{n\times n}\,)\,.$$

Fact 7.3 *$\mathcal{M}(\Lambda)$ is nonempty and compact. In particular, if $t, s \in [a,b]$ then the set*

$$\mathcal{M}_{t,s}(\Lambda) \stackrel{\text{def}}{=} \{\,M_L(t,s) : L \in \Gamma(\Lambda)\,\} \tag{6}$$

is a nonempty compact subset of $\mathbb{R}^{n\times n}$. ♢

Remark 7.4 The concept of a variational generator also makes intrinsic sense on manifolds. The important new point is that now Λ has to be taken to be a section along ξ of an appropriate bundle. We make this precise in the only case that will be used here, namely, when f is an "augmented ppd vector field" on a manifold M, that is, a time-varying ppd section of the bundle $TM\times\mathbb{R}$. In that case, we let E_M be the vector bundle over M whose fiber $E_M(x)$ at each $x \in M$ is the product $J^1_x(VF(M)) \times T^*_xM$, where $J^1_x(VF(M))$ is the space of 1-jets at x of smooth vector fields on M. Then, if a, b, ξ are such that $a, b \in \mathbb{R}$, $a \le b$, and $\xi \in C^0(\,[a,b]\,;M\,)$, a *variational generator for f about ξ* is a measurable set-valued map $\Lambda : [a,b] \longrightarrow\!\!\!\!\!\rightarrow E_M$ such that

- $\Lambda(t)$ is, for a.e. t, a compact convex nonempty subset of $E_M(\xi(t))$ such that $\pi_{1,0}(v) = f_1(\xi(t),t)$ for every $v \in \Lambda(t)$, where $\pi_{1,0}$ is the canonical projection from $J^1(VF(M))$ to $J^0(VF(M))$ and f_1 is the TM component of f;
- locally, relatively to suitable coordinate charts, there exist k_Λ, $\bar\alpha$, $\mathbf{k}$ having the three properties of Definition 7.1.

With this definition of variational generator Theorem 8.1 remains true on manifolds.

A more detailed discussion of the invariant definition of variational generators is given in Sussmann [6]. ♢

8. Differentiation of flows

We let $\mathcal{C}(2)$ be the set of all cones in $\mathbb{R}^2$ that are products $C_+ \times C_-$, where $C_+ \in \mathcal{C}(1)$ and $C_- \in \mathcal{C}(1)$. Then $\mathcal{C}(2)$ has exactly sixteen members.

If X_-, X_+ are finite-dimensional real linear spaces, $\mathbf{M}$ is a set of linear maps from X_- to X_+, and $\mathbf{v}_+$, $\mathbf{v}_-$, are nonempty subsets of X_-, X_+, we define a set $[\mathbf{M};\mathbf{v}_+,\mathbf{v}_-]$ of linear maps from $X_-\times\mathbb{R}^2$ to X_+ by letting $[\mathbf{M};\mathbf{v}_+,\mathbf{v}_-]$ be the set of all maps $[M;v_+,v_-]$, for all $M \in \mathbf{M}$, $v_+ \in \mathbf{v}_+$, $v_- \in \mathbf{v}_-$, where $[M;v_+,v_-]$ is the linear map from $X_-\times\mathbb{R}^2$ to X_+ given by

$$[M;v_+,v_-](y,\beta,\alpha) \stackrel{\text{def}}{=} M\cdot y + \beta\cdot v_+ - \alpha\cdot M\cdot v_-\,.$$

It is then clear that *if $\mathbf{M}$, $\mathbf{v}_+$, $\mathbf{v}_-$ are compact, then $[\mathbf{M};\mathbf{v}_+,\mathbf{v}_-]$ is compact.*

The following result is the general theorem on GDQ differentiation with respect to the state and the endpoints, along a trajectory $\xi \in \mathrm{Traj}_c(f)$ of the flow Φ^f generated by a time-varying vector field f. The basic requirements are local integral continuity, the existence of a variational generator, and the existence of approximate limits of f at the endpoints of ξ.

Theorem 8.1 *Assume that Q is a manifold of class C^1, f is a time-varying pdd vector field on Q, $a, b \in \mathbb{R}$, $a \le b$, $C \in \mathcal{C}(2)$, $C_+, C_- \in \mathcal{C}(1)$, $C = C_+ \times C_- \in \mathcal{C}(2)$, and $C^{b,a} = (b,a) + C = (b + C_+)\times(a + C_-)$.*

Assume, in addition, that

1. *f is locally integrally continuous;*
2. *Λ is a variational generator for f along ξ,*
3. *$\mathbf{v}_+$, $\mathbf{v}_-$ are nonempty compact convex subsets of $T_{\xi(a)}Q$, $T_{\xi(b)}Q$, such that*

$$\mathbf{v}_+ \in \mathrm{App\text{–}lim}_{x\to\xi(b)\,,\,t\to b,\,t-b\in C_+} f(x,t)\,,$$

$$\mathbf{v}_- \in \mathrm{App\text{–}lim}_{x\to\xi(a)\,,\,t\to a,\,t-a\in C_-} f(x,t)\,.$$

Then the set $[\mathcal{M}_{b,a}(\Lambda);\mathbf{v}_+,\mathbf{v}_-]$ belongs to $GDQ\Big(\Phi^f;(\xi(a),b,a),\xi(b);Q\times C^{b,a}\Big)$. ♢

9. A GDQ maximum principle

Theorem 8.1, together with the directional open mapping theorem 4.2, imply a version of the maximum principle that contains and improves upon several previous

smooth and nonsmooth versions, for vector field systems as well as for differential inclusions and systems of differential inclusions. Moreover, one can also allow "jump maps," and obtain a "hybrid" version. We state this more general version directly but, for simplicity, we only discuss the vector field case.

For our restricted purposes, let us define a *hybrid optimal control problem* to consist of the specification of a finite sequence $(\Sigma^1, \ldots, \Sigma^\mu)$ of "ordinary control systems," together with "Lagrangians" $L^1, \ldots, L^\mu$ for $\Sigma^1, \ldots, \Sigma^\mu$, "switching constraints" $\mathcal{S}^1, \ldots, \mathcal{S}^\mu$, "switching cost functions" $\varphi^1, \ldots, \varphi^\mu$, and "time sets" $T^1_-, T^1_+, T^2_-, T^2_+, \ldots, T^\mu_-, T^\mu_+$.

Precisely, each Σ^i is a triple $\Sigma^i = (Q^i, \mathcal{U}^i, F^i)$ consisting of a *state space* Q^i, a *controller space* $\mathcal{U}^i$, and a *controlled dynamics*, that is, a parametrized family $F^i = \{F^i_\eta\}_{\eta \in \mathcal{U}^i}$ such that, for each $\eta \in \mathcal{U}^i$, F^i_η is a ppd time-varying vector field on Q^i. Each L^i is a family $\{L^i_\eta\}_{\eta \in \mathcal{U}^i}$ of ppd functions $L^i_\eta : Q^i \times \mathbb{R} \dashrightarrow \mathbb{R}$. For each $i \in \{1, \ldots, \mu\}$, the switching constraint $\mathcal{S}^i$ is a subset of $Q^i \times \mathbb{R} \times Q^{i\tilde{+}1} \times \mathbb{R}$, where "$i\tilde{+}1$" means "$i+1$" if $i < \mu$, and "1" if $i = \mu$. The switching cost functions are functions $\varphi^i : Q^i \times \mathbb{R} \times Q^{i\tilde{+}1} \times \mathbb{R} \to \mathbb{R}$. The time sets T^i_-, T^i_+ are subsets of $\mathbb{R}$. A *controller* is a μ-tuple $\boldsymbol{\eta} = (\eta^1, \ldots, \eta^\mu) \in \mathcal{U}^1 \times \cdots \times \mathcal{U}^\mu$. A *trajectory* for a controller $\boldsymbol{\eta} = (\eta^1, \ldots, \eta^\mu)$ is a μ-tuple $\boldsymbol{\xi} = (\xi^1, \ldots, \xi^\mu)$ with the property that, for each i, ξ^i is an absolutely continuous curve in Q^i, defined on a compact interval $I(\xi^i) = [a_-(\xi^i), a_+(\xi^i)]$, that satisfies the conditions $\hat{\xi}^i(t) \in \mathrm{Do}(F^i_{\eta^i})$ and $\dot{\xi}^i(t) = F^i_{\eta^i}(\hat{\xi}^i(t))$ for a.e. $t \in I(\xi^i)$, and is such that $\sigma^i(\boldsymbol{\xi}) \in \mathcal{S}^i$, where $\hat{\xi}^i(t) \stackrel{\mathrm{def}}{=} (\xi^i(t), t)$ and $\sigma^i(\boldsymbol{\xi}) \stackrel{\mathrm{def}}{=} (\hat{\xi}^i(a_+(\xi^i)), \hat{\xi}^{i\tilde{+}1}(a_-(\xi^{i\tilde{+}1})))$. A *trajectory-control pair* (abbr. TCP) is a pair $(\boldsymbol{\xi}, \boldsymbol{\eta})$ such that $\boldsymbol{\eta}$ is a controller and $\boldsymbol{\xi}$ is a trajectory for $\boldsymbol{\eta}$. A TCP $(\boldsymbol{\xi}, \boldsymbol{\eta})$ is *admissible* if, for each i, $a_-(\xi^i) \in T^i_-$, $a_+(\xi^i) \in T^i_+$, and the functions $I(\xi^i) \ni t \to L^i_{\eta^i}(\hat{\xi}^i(t)) \in \mathbb{R}$ are a.e. defined—that is, $\hat{\xi}^i(t) \in \mathrm{Do}(L^i_{\eta^i})$ for a.e. $t \in I(\xi^i)$—and Lebesgue integrable. The *cost* of an admissible TCP $(\boldsymbol{\xi}, \boldsymbol{\eta})$ is the number

$$J(\boldsymbol{\xi}, \boldsymbol{\eta}) \stackrel{\mathrm{def}}{=} \sum_{i=1}^{\mu} \int_{a_-(\xi^i)}^{a_+(\xi^i)} L^i_{\eta^i}(\hat{\xi}^i(t))\, dt + \sum_{i=1}^{\mu} \varphi\Big(\sigma^i(\boldsymbol{\xi})\Big)\,.$$

An *optimal TCP* is an admissible TCP $(\boldsymbol{\xi}, \boldsymbol{\eta})$ such that $J(\boldsymbol{\xi}, \boldsymbol{\eta}) \leq J(\boldsymbol{\xi}', \boldsymbol{\eta}')$ for every admissible TCP $(\boldsymbol{\xi}', \boldsymbol{\eta}')$.

For each i, we define the *L-augmented dynamics* to be the family $\tilde{F}^i = \{\tilde{F}^i_\eta\}_{\eta \in \mathcal{U}^i}$ of ppd maps from $Q^i \times \mathbb{R}$ to $TQ^i \times \mathbb{R}$ given by

$$\tilde{F}^i_\eta(q,t) = \begin{bmatrix} F^i_\eta(q,t) \\ L^i_\eta(q,t) \end{bmatrix} \quad \text{for} \quad (q,t) \in Q^i \times \mathbb{R}\,.$$

Now assume that

H1. $(\boldsymbol{\xi}, \boldsymbol{\eta}) = ((\xi^1, \ldots, \xi^\mu), (\eta^1, \ldots, \eta^\mu))$ is an admissible TCP, and $a^i_- = a(\xi^i)$, $a^i_+ = a_+(\xi^i)$, $x^i_- = \xi^i(a^i_-)$, $x^i_+ = \xi^i(a^i_+)$, for $i = 1, \ldots, \mu$.

H2. $\tilde{\Lambda}^1, \ldots, \tilde{\Lambda}^\mu$ are variational generators for $\tilde{F}^1_{\eta^1}, \ldots, \tilde{F}^\mu_{\eta^\mu}$ along $\xi^1, \ldots, \xi^\mu$.

H3. C^i_σ are, for $i = 1, \ldots, \mu$, $\sigma \in \{+,-\}$, cones in $\mathbb{R}$ such that $a^i_\sigma \in \mathrm{Int}_{a^i_\sigma + C^i_\sigma} T^i_\sigma$ and $\Xi^i_\sigma \stackrel{\mathrm{def}}{=} (x^i_\sigma, a^i_\sigma)$ is a point of approximate continuity of $\tilde{F}^i_{\eta^i}$ along C^i_σ.

H4. For each $i \in \{1, \ldots, \mu\}$, $\mathcal{C}^i$ is a convex multicone in $T_{\Xi^i_+} \mathcal{Q}^i_+ \times T_{\Xi^{i\tilde{+}1}_-} \mathcal{Q}^{i\tilde{+}1}_-$, where, for $\sigma \in \{+,-\}$,

$$R^i_\sigma \stackrel{\mathrm{def}}{=} \begin{cases} \{0\} & \text{if} \quad C^i_\sigma = \{0\}\,, \\ \mathbb{R} & \text{if} \quad C^i_\sigma \neq \{0\}\,, \end{cases}$$

$$\mathcal{Q}^i_\sigma \stackrel{\mathrm{def}}{=} Q^i \times (a^i_\sigma + R^i_\sigma)\,.$$

H5. $\mathcal{C}^i$ is a GDQ approximating multicone for the restricted switching set

$$\mathcal{S}^i_{rest} \stackrel{\mathrm{def}}{=} \mathcal{S}^i \cap (\mathcal{Q}^i_+ \times \mathcal{Q}^{i\tilde{+}1}_-)\,.$$

at the switching point $P^i = \sigma^i(\boldsymbol{\xi}) = (\Xi^i_+, \Xi^{i\tilde{+}1}_-)$.

H6. For each $i \in \{1, \ldots, \mu\}$, Ω^i is a subset of the dual space $\Big(T_{x^i_+} Q^i \times R^i_+ \times T_{x^{i\tilde{+}1}} Q^{i\tilde{+}1} \times R^{i\tilde{+}1}_-\Big)^\dagger$ (that is, of $\Big(T_{\Xi^i_+} \mathcal{Q}^i_+ \times T_{\Xi^{i\tilde{+}1}_-} \mathcal{Q}^{i\tilde{+}1}_-\Big)^\dagger$), and Ω^i belongs to $GDQ(\varphi^i; P^i, \varphi^i(P^i); \mathcal{Q}^i_+ \times \mathcal{Q}^{i\tilde{+}1}_-)$.

H7. For each $i \in \{1, \ldots, \mu\}$ and each $\eta \in \mathcal{U}^i$, the time-varying map $\tilde{F}^i_\eta$ is locally integrally continuous.

H8. Each control system Σ^i is invariant under time-interval substitutions. (That is, if $\eta, \zeta \in \mathcal{U}^i$, $\xi \in C^0(\,[a,b]\,;\, Q^i\,) \cap \mathrm{Traj}_{\,c}(F^i_\eta)$, and J is a compact subinterval of $[a,b]$ such that $(\xi(t), t) \in \mathrm{Do}(F^i_\zeta)$ for $t \in J$, then there exists a controller $\theta \in \mathcal{U}^i$ such that $F^i_\theta(q,t) = F^i_\eta(q,t)$ whenever $q \in Q^i$, $t \in [a,b]$, $t \notin J$, and $F^i_\theta(q,t) = F^i_\zeta(q,t)$ whenever $q \in Q^i$, $t \in J$.)

We now define the notion of a "multiplier" along $(\boldsymbol{\xi}, \boldsymbol{\eta})$, and what it means for a multiplier to be "Hamiltonian-maximizing."

For $i = 1, \ldots, \mu$, and $\zeta \in \mathcal{U}^i$, we define the *Hamiltonian* $H^i_\zeta : T^*Q^i \times \mathbb{R} \times \mathbb{R} \dashrightarrow \mathbb{R}$ by letting

$$H^i_\zeta(q, \lambda, t, \lambda_0) = \lambda \cdot F^i_\zeta(q,t) - \lambda_0 L^i_\zeta(q,t)\,.$$

Definition 9.1 If H1-H8 hold, then a *multiplier along* $(\boldsymbol{\xi}, \boldsymbol{\eta})$ is a triple $(\boldsymbol{\psi}, \psi_0, \boldsymbol{\kappa})$ with the property that:

- $\boldsymbol{\psi}$ is a μ-tuple $(\psi^1, \ldots, \psi^\mu)$ such that each ψ^i is a field of covectors along ξ^i (that is, ψ^i is a map from the interval $[a^i_-, a^i_+]$ to T^*Q^i, such that $\psi^i(t)$ belongs to $T^*_{\xi^i(t)} Q^i$ for every $t \in [a^i_-, a^i_+]$);
- $\boldsymbol{\kappa}$ is a 2μ-tuple $(\kappa^1_-, \kappa^1_+, \ldots, \kappa^\mu_-, \kappa^\mu_+)$ such that $\kappa^i_\sigma \in R^i_\sigma$ whenever $i \in \{1, \ldots, \mu\}$ and $\sigma \in \{+,-\}$;
- each ψ^i is absolutely continuous and satisfies the *adjoint differential inclusion*

$$-\dot{\psi}^i(t) \in [\psi^i(t), -\psi_0] \cdot \tilde{\Lambda}(t) \text{ for a.e. } t \in [a^i_-, a^i_+];$$

- $\psi_0 \in \mathbb{R}$ and $\psi_0 \geq 0$;
- for each $i \in \{1, \ldots, \mu\}$, $\sigma \in \{+,-\}$, the *switching conditions*

$$(-\psi^i(a^i_+), -\kappa^i_+ + h^i_+, \psi^{i\tilde{+}1}(a^{i\tilde{+}1}_-), -\kappa^i_- - h^{i\tilde{+}1}_-) \in \psi_0 \Omega^i + (\mathcal{C}^i)^\perp,$$

$$\kappa^i_\sigma \in (C^i_\sigma)^\perp$$

hold, where

$$h^i_\sigma = H^i_{\eta^i}(x^i_\sigma, \psi^i(a^i_\sigma), a^i_\sigma, \psi_0)\,.$$

Definition 9.2 If H1-H8 hold, and $(\boldsymbol{\psi}, \psi_0, \boldsymbol{\kappa})$ is a multiplier along $(\boldsymbol{\xi}, \boldsymbol{\eta})$, we say that $(\boldsymbol{\psi}, \psi_0, \boldsymbol{\kappa})$ is *Hamiltonian-maximizing* if, for every $i \in \{1, \ldots, \mu\}$, the inequality

$$H^i_\zeta(\xi^i(t), \psi^i(t), t, \psi_0) \leq H^i_{\eta^i}(\xi^i(t), \psi^i(t), t, \psi_0)$$

holds whenever $\zeta \in \mathcal{U}^i$, $t \in]a^i_-, a^i_+[$, and $(\xi^i(t), t)$ is a point of approximate continuity of $\tilde{F}^i_{\eta^i}$ and $\tilde{F}^i_\zeta$. ♢

Definition 9.3 If $(\boldsymbol{\psi}, \psi_0, \boldsymbol{\kappa})$ is a multiplier along $(\boldsymbol{\xi}, \boldsymbol{\eta})$, we say that $(\boldsymbol{\psi}, \psi_0, \boldsymbol{\kappa})$ is *nontrivial* if it is not true that $\psi_0 = \kappa^1_- = \kappa^1_+ = \ldots = \kappa^\mu_- = \kappa^\mu_+ = 0$ and all the functions ψ^i are identically zero. ♢

Theorem 9.4 *If H1-H8 hold, and the pair $(\boldsymbol{\xi}, \boldsymbol{\eta})$ is optimal, then there exists a nontrivial Hamiltonian-maximizing multiplier along $(\boldsymbol{\xi}, \boldsymbol{\eta})$.* ♢

By taking $\mu = 1$, Theorem 9.4 can be shown to include the classical "non-hybrid" smooth and nonmsooth versions of the maximum principle given, for example, in Pontryagin *et al.* [5], Berkovitz [1], Clarke [3, 2]. In that case, the switching condition of Definition 9.1 becomes the transversality condition. When the augmented vector fields $\tilde{F}^i_{\eta^i}$ are of class C^1, one can take

$$\tilde{\Lambda}^i(t) = \left\{ \frac{\partial \tilde{F}^i_{\eta^i}}{\partial x}(\xi^i(t), t) \right\}, \tag{7}$$

and the adjoint differential inclusion becomes the classical adjoint equation. On the other hand, if the function $x \to \tilde{F}^i_{\eta^i}(x, t)$ is differentiable at $\xi^i(t)$ for almost all t, then one can still take $\tilde{\Lambda}^i$ to be given by (7), and $\tilde{\Lambda}^i$ is a variational generator, provided that the differentiability of $x \to \tilde{F}^i_{\eta^i}(x, t)$ at $\xi^i(t)$ has an obvious integral uniformity property with respect to t. So Theorem 9.4 is in fact stronger than the classical versions, even in the setting of single-valued differentials.

In addition, when the $\tilde{F}^i_{\eta^i}$ are Lipschitz continuous on some tube about the reference trajectory, with an integrable Lipschitz constant, then one can take $\tilde{\Lambda}^i(t)$ to be $\partial \tilde{F}^i_{\eta^i,t}(\xi^i(t))$, where $\tilde{F}^i_{\eta^i,t}$ is the map $x \to \tilde{F}^i_{\eta^i}(x, t)$, and "$\partial$" stands for "Clarke generalized Jacobian." Moreover, in the Lipschitz case one can often take the $\tilde{\Lambda}^i$ to be smaller than the Clarke generalized Jacobian (for example, equal to the classical differential, when it exists), so even in the Lipschitz case Theorem 9.4 often yields a stronger conclusion than the usual nonsmooth results.

Theorem 9.4 also applies to problems where the vector fields are only continuous with respect to the state (in which case the flow maps are set-valued) and to problems with discontinuous vector fields. An important class of such problems arises from *differential inclusion systems*. As long as the inclusions under consideration are almost lower semicontinuous, then there exist sufficiently many integrally continuous selections to make our theorem applicable. All these applications will be discussed in a subsequent paper.

10. Path-integral generalized differentials

If $n, m \in \mathbb{Z}_+$, $\alpha : [0, 1] \to \mathbb{R}^n$ is a Lipschitz function, and $h : [0, 1] \to \mathbb{R}^{m \times n}$ is integrable, we use $h * \alpha$ to denote the "chronological product" of h and α, that is, the absolutely continous function $\beta : [0, 1] \to \mathbb{R}^m$ given by

$$\beta(t) = \int_0^t h(s) \cdot \dot{\alpha}(s)\, ds\,.$$

Let $n \in \mathbb{Z}_+$, and let S be a subset of $\mathbb{R}^n$. We write $\mathcal{A}(S)$ to denote the subset of $C^0(\,[0, 1]\,; \mathbb{R}^n\,)$ consisting of all absolutely continuous curves $\alpha : [0, 1] \to \mathbb{R}^n$ such that $\alpha(0) = 0$ and $\dot{\alpha}(t) \in S$ for almost all $t \in [0, 1]$.

If C is a convex cone in $\mathbb{R}^n$, and $r > 0$, we write $C(r) = \{v \in C : \|v\| \leq r\}$.

Definition 10.1 Let $n, m \in \mathbb{Z}_+$, let $F : \mathbb{R}^n \longrightarrow\!\!\!\!\to \mathbb{R}^m$, and let C be a closed convex cone in $\mathbb{R}^n$. We say that Λ is a *path-integral generalized differential of F at $(0, 0)$ in the direction of C*, and write $\Lambda \in PIGD(F, C)$, if Λ is a nonempty compact subset of $\mathbb{R}^{m \times n}$, and for every positive real number δ there exists a number $R \in]\,0, \infty\,[$ with the property that

(#) for every $r \in]\,0, R]$ there exists a map $G \in \mathrm{REG}(\mathcal{A}(C(r)); C^0(\,[0, 1]\,; \mathbb{R}^{m \times n}\,) \times \mathbb{R}^m)$ such that

(#.a) $h(t) \in \Lambda^\delta$ and $\|v\| \leq \delta r$ whenever $\alpha \in \mathcal{A}(C(r))$, $(h, v) \in G(\alpha)$, $t \in [0, 1]$,

(#.b) $\mathrm{Gr}(\Phi_G) \subseteq \mathrm{Gr}(F)$, where Φ_G is the set-valued map from $\mathcal{A}(C(r))$ to $\mathbb{R}^m$ such that, if x belongs to $\mathcal{A}(C(r))$, then $\Phi_G(x)$ is the set of all $y \in \mathbb{R}^m$ for which $y = (h * \alpha)(1) + v$ for some triple $(\alpha, h, v) \in \mathrm{Gr}(G)$ such that $\alpha(1) = x$. ♢

It then turns out that every GDQ is a PIGD, and every derivate container, semidifferential, and multidifferential is also a PIGD. Moreover, PIGDs are intirinsicall defined of manifolds, and satisfy all the desirable properties such as the chain rule (for polyhedral cones) and the directional open mapping theorem. The resulting version of the maximum principle is thus even more general than the one involving GDQs.

References

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

[2] Clarke, F. H., *The Maximum Principle under minimal hypotheses.* SIAM J. Control Optim. **14**, 1976, pp. 1078–1091.

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

[4] Halkin, H., *Necessary conditions for optimal control problems with differentiable or nondifferentiable data.* In *Mathematical Control Theory*, Lect. Notes in Math. **680**, Springer-Verlag, Berlin, 1978, pp. 77–118.

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

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

[7] Sussmann, H. J., *Multidifferential calculus: chain rule, open mapping and transversal intersection theorems.* In *Optimal Control: Theory, Algorithms, and Applications*, W. W. Hager and P. M. Pardalos Eds., Kluwer, 1998, pp. 436–487.

[8] Sussmann, H. J., *A maximum principle for hybrid optimal control problems.* In *Proc. 38th IEEE Conf. Decision and Control, Phoenix, AZ, Dec. 1999.* IEEE publications, New York, 1999, pp. 425–430.

[9] Warga, J., *Optimization and controllability without differentiability assumptions.* SIAM J. Control and Optimization **21**, 1983, pp. 837–855.