Set-valued differentials and the hybrid maximum principle[*]

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

1. Introduction

In recent years, it has become clear that most (smooth, nonsmooth, high-order, and hybrid) versions of the maximum principle (abbr. MP) for finite-dimensional, deterministic optimal control problems without state space constraints can be derived in a unified way, by using a modified version of the approach of the classical book [4] by Pontryagin *et al.*—that is, by constructing "packets of needle variations," linearly approximating (that is, "differentiating") these packets at the base value of the variation parameter, and propagating the resulting linear approximations to the terminal point of the trajectory by means of the differentials of the reference flow maps, in order to construct an "approximating cone" to a suitable reachable set. The classical approach must be modified in three basic ways. *First*, the classical differential (which is used twice, first to obtain linear approximations to the variations and then to propagate these approximations to the terminal point) must be replaced by other objects, called "generalized differentials" (abbr. GDs). Examples of GDs are: J. Warga's "derivate containers" (cf. [9]), H. Halkin's "screens" (cf. [3]), the "semidifferentials" and "multidifferentials" proposed by us in previous work (cf. [6]), and our more recent "generalized differential quotients" (abbr. GQDs) and "path-integral generalized differentials" (abbr. PIGDs), presented in [7] and [8]. *Second,* the time-varying vector fields that occur in the classical MP must be replaced by *flows*. *Third,* the needle variations must be replaced by abstract variations. A notion of GD will yield a version of the MP provided it satisfies some natural properties such as the chain rule and an appropriate "directional open mapping property" (abbr. DOMP). In the GD setting the "differentials" of a map at a point are sets of linear maps rather than single linear maps. It follows that these differentials cannot quite be used to propagate cones in a natural way, since the image of a cone under a set of linear maps is a set of cones rather than a single cone. Hence, if we agree to call a set of cones a "multicone," the natural class of objects that must be used as linear approximations to sets is that of *multicones*. In particular, the "Pontryagin cone" of the classical theory is now replaced by a "Pontryagin multicone."

In this note, we propose an axiomatic definition of the concept of a "generalized differentiation theory" (abbr. GDT) and a precise statement of the DOMP, and we outline the definitions of our two most recent GDTs, namely, the GDQs and PIGDs. In addition, we give a complete statement of a "hybrid MP for general GDTs,", which now amounts to saying that "to every GDT that has the DOMP is associated a version of the hybrid MP." (We find it convenient not to include the DOMP among the GDT axioms, thus allowing the possibility that a GDT might fail to have the DOMP and consequently fail to give rise to a MP. On the other hand, we choose to include a form of the chain rule, thereby disqualifying concepts such as Clarke's "generalized Jacobian" (cf. [2]), and Halkin's "shields" (cf. [3]). Finally, we limit ourselves to theories in which the GDs of a set-valued map F from a C^1 manifold M to a C^1 manifold N at a point $(x,y) \in M \times N$ are *nonempty compact sets of linear maps* from T_xM to T_yN—where T_qQ denotes the tangent space of Q at q—thereby excluding theories due to Ioffe, Mordukhovich and others, where the GDs are different kinds of objects. We are fully aware that these choices are somewhat arbitrary, and that, when a truly definitive version is achieved, the details of the definitions might have to be modified.)

2. Preliminaries

By a *set-valued map* (abbr. SVM) we mean a triple $F = (A, B, G)$ such that A and B are sets and G is a subset of $A \times B$. The sets A, B, G are, respectively, the *source*, *target*, and *graph* of F, and we write $A = \mathrm{So}(F)$, $B = \mathrm{Ta}(F)$, $G = \mathrm{Gr}(F)$. If x is any object, we write $F(x) = \{y : (x, y) \in \mathrm{Gr}(F)\}$. (Hence $F(x) = \emptyset$ unless $x \in \mathrm{So}(F)$.) The sets $\mathrm{Do}(F) = \{x \in \mathrm{So}(F) : F(x) \neq \emptyset\}$, $\mathrm{Im}(F) = \bigcup_{x\in \mathrm{So}(F)} F(x)$, are, respectively, the *domain* and *image* of F. If $F = (A, B, G)$ is an SVM, we say that F is an *SVM from A to B*, and write $F : A \longrightarrow\!\!\!\!\rightarrow B$. We use $SVM(A, B)$ to denote the set of all SVMs from A to 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.

3. GDTs

We use $\mathbb{Z}_+$ to denote the set of all nonnegative integers. If $k \in \mathbb{Z}_+$, then $\mathcal{M}^k$, $\mathcal{SVM}(\mathcal{M}^k)$ will denote, respectively, the class of all finite-dimensional Hausdorff manifolds of class C^k without boundary, and the class of all SVMs F such that $\mathrm{So}(F)\in\mathcal{M}^k$ and $\mathrm{Ta}(F)\in\mathcal{M}^k$. If $k \geq 1$ and $x\in M\in\mathcal{M}^k$, then T_xM, T^*_xM, TM, T^*M will denote, respectively, the tangent and cotangent spaces to M at x, and the tangent and cotangent bundles of M. Clearly, TM and T^*M belong to $\mathcal{SVM}(\mathcal{M}^{k-1})$.

If $M \in \mathcal{M}^1$, $x \in M$, and $S \subseteq M$, then T^B_xS will denote the *Bouligand tangent cone* to S at x. By definition, T^B_xS is the set of all tangent vectors $v \in T_xM$ such that there exist a sequence $\{(x_j, h_j)\}_{j\in\mathbb{N}}$ of points of $S\times\mathbb{R}$ having the property that $h_j > 0$ for all j, $h_j \downarrow 0$ as $j \to \infty$, and $\lim_{j\to\infty} h_j^{-1}\Big(\varphi(x_j) - \varphi(x) - h_j(v\varphi)\Big) = 0$ for all functions $\varphi \in C^1(M, \mathbb{R})$.

We write $\mathcal{RLS}$, $\mathcal{FDRLS}$, to denote, respectively, the

---

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

class of all linear spaces over $\mathbb{R}$, and the class of all $X \in \mathcal{RLS}$ that are finite-dimensional. If $X, Y \in \mathcal{RLS}$, then $\mathrm{Lin}(X,Y)$ denotes the set of all $\mathbb{R}$-linear maps from X to Y. A subset of $\mathrm{Lin}(X,Y)$ will be called a *linear multimap* from X to Y. A set of linear multimaps from X to Y will be called a *linear multimultimap* from X to Y. We use $\mathrm{LinM}(X,Y)$, $\mathrm{LinMM}(X,Y)$, to denote, respectively, the set of all linear multimaps from X to Y and the set of all linear multimultimaps from X to Y. If $X, Y \in \mathcal{FDRLS}$, then we use $\mathrm{LinM}_c(X,Y)$, $\mathrm{LinMM}_c(X,Y)$, respectively, to denote the set of all nonempty compact linear multimaps from X to Y, and the set of all subsets of $\mathrm{LinM}_c(X,Y)$. The members of $\mathrm{LinMM}_c(X,Y)$ are the *compact linear multimultimaps* from X to Y.

A *generalized differentiation theory* (abbr. GDT) is a correspondence $\mathcal{D}$ that assigns to every $F \in \mathcal{SVM}(\mathcal{M}^1)$, every point $(x,y) \in \mathrm{So}(F) \times \mathrm{Ta}(F)$, and every subset S of the source $\mathrm{So}(F)$, a compact linear multimultimap $\mathcal{D}(F;x,y;S) \in \mathrm{LinMM}_c(T_xM, T_yY)$, in such a way that the following axioms are satisfied:

1. If $M, N \in \mathcal{M}^1$, $f : M \mapsto N$ is a map of class C^1, $x \in M$, and $y = f(x)$, then $\{Df(x)\} \in \mathcal{D}(f;x,y;M)$.
2. (*Monotonicity*) Whenever $F_1, F_2 \in \mathcal{SVM}(\mathcal{M}^1)$, $M = \mathrm{So}(F_1) = \mathrm{So}(F_2)$, $N = \mathrm{Ta}(F_1) = \mathrm{Ta}(F_2)$, $x \in M$, $y \in N$, $S \subseteq M$, $\Lambda \in \mathcal{D}(F_1;x,y;S)$, $\Lambda' \in \mathrm{LinM}_c(T_xM, T_yN)$, $\mathrm{Gr}(F_1) \subseteq \mathrm{Gr}(F_2)$ and $\Lambda \subseteq \Lambda'$, it follows that $\Lambda' \in \mathcal{D}(F_2;x,y;S)$.
3. *(The chain rule)* If $F_i \in \mathcal{SVM}(\mathcal{M}^1)$ and $S_i \subseteq M_i$ for $i = 1,2$, $M_1 = \mathrm{So}(F_1)$, $M_2 = \mathrm{Ta}(F_1) = \mathrm{So}(F_2)$, $M_3 = \mathrm{Ta}(F_2)$, $x_i \in M_i$ for $i = 1,2,3$, $F_1(S_1) \subseteq S_2$, and $\Lambda_2 \in \mathcal{D}(F_2;x_2,x_3;S_2)$, then $\Lambda_2 \circ \Lambda_1 \in \mathcal{D}(F_2 \circ F_1;x_1,x_3;S_1)$, if either (a) F_1 is a diffeomorphism of class C^1 from a neighborhood U_1 of x_1 onto a neighborhood U_2 of x_2, $F_1(x_1) = x_2$, $F_1(S_1 \cap U_1) = S_2 \cap U_2$, and $\Lambda_1 = \{DF_1(x_1)\}$, or (b) F_1 is a single-valued map of class C^1 defined on a neighborhood U of x_1, $F_1(x_1) = x_2$, $\Lambda_1 \in \mathcal{D}(F_1;x_1,x_2;S_1)$, and $\Lambda_1 = \{DF_1(x_1)\}$, or (c) $\Lambda_1 \in \mathcal{D}(F_1;x_1,x_2;S_1)$, $\Lambda_1 \cdot T^B_{x_1}S_1 \subseteq T^B_{x_2}S_2$ and, for some coordinate chart $\kappa : U \to \mathbb{R}^n$ on M_2 such that $x_2 \in U$ and $\kappa(x_2) = 0$, there is a polyhedral cone C in $\mathbb{R}^n$ such that $\kappa(S_2 \cap U) = C \cap \kappa(U)$.
4. *(The product rule)* If, for $i = 1,2$, $F_i \in \mathcal{SVM}(\mathcal{M}^1)$, $x_i \in M_i = \mathrm{So}(F_i)$, $y_i \in N_i = \mathrm{Ta}(F_i)$, $S_i \subseteq M_i$, and $\Lambda_i \in \mathcal{D}(F_i;x_i,y_i;S_i)$, then $\Lambda_1 \times \Lambda_2$ belongs to $\mathcal{D}(F_1 \times F_2;(x_1,x_2),(y_1,y_2);S_1 \times S_2)$.

4. The directional open mapping property

We say that a GDT $\mathcal{D}$ has the *directional open mapping property* if the following statement is true:.

(DOMP) *Assume that* $n, m \in \mathbb{Z}_+$, $F \in \mathcal{SVM}(\mathbb{R}^n, \mathbb{R}^m)$, $v \in \mathbb{R}^m$, *C is a closed convex cone in* $\mathbb{R}^n$, *Λ belongs to* $\mathcal{D}(F;0,0;C)$, *and* $v \in \bigcap_{L\in\Lambda} \mathrm{Int}(LC)$. *Then there exists a closed convex cone D in $\mathbb{R}^n$ such that* $v \in \mathrm{Int}(D)$, *having the property that for every* $\delta \in]0,\infty[$ *there exists an* $\varepsilon(\delta) \in]0,\infty[$ *such that* $\boxed{D \cap \{y \in \mathbb{R}^m : \|y\| \le \varepsilon(\delta)\} \subseteq F(C \cap \{x \in \mathbb{R}^n : \|x\| \le \delta\})}$. ◇

We say that $\mathcal{D}$ has the *strong directional open mapping property* if

(SDOMP) *Whenever n, m, F, v, C, Λ satisfy the hypotheses of* **(DOMP)**, *it follows that D and the function $\delta \mapsto \varepsilon(\delta)$ can be chosen so that the following stronger conclusion holds:*

(*) if $y \in D$ and $\|y\| \le \varepsilon(\delta)$ then there exists a compact connected subset Z_y of $\{x \in C : \|x\| \le \delta\} \times [0,1]$ such that $(0,0) \in Z_y$, $(x,1) \in Z_y$ for some $x \in C$ such that $\|x\| \le \delta$, and $ry \in F(x)$ whenever $0 \le r \le 1$ and $(x,r) \in Z_y$. ◇

If, in statements (DOMP) (resp. (SDOMP)), the function $\delta \mapsto \varepsilon(\delta)$ can be chosen to be linear for small enough δ, then we say that $\mathcal{D}$ has the DOMP (resp. SDOMP) *with linear rate.*

5. Approximating multicones

Associated to each GDT $\mathcal{D}$ there is a notion of "$\mathcal{D}$-approximating multicone to a set at a point." Recall that if $X \in \mathcal{RLS}$ then a *convex multicone* in X is a nonempty set of convex cones in X.

Definition 5.1 If $\mathcal{D}$ is a GDT, $M \in \mathcal{M}^1$, $S \subseteq M$ and $x \in S$, a *$\mathcal{D}$-approximating multicone to S at x* is a convex multicone $\mathcal{C}$ in T_xM such that there exist $X \in \mathcal{FDRLS}$, $F : X \longrightarrow\!\!\!\!\to M$, a closed convex cone K in X, and a $\Lambda \in \mathcal{D}(F;0,x;K)$, such that $F(K) \subseteq S$ and $\mathcal{C} = \{LK : L \in \Lambda\}$. ◇

If $X \in \mathcal{RLS}$, then $X^\dagger$ denotes the dual space of X. If $X \in \mathcal{FDRLS}$, then we identify $X^{\dagger\dagger}$ to X in the usual way. If $\mathcal{C}$ is a convex multicone in X, the *polar* of $\mathcal{C}$ is the (not necessarily convex) cone $\mathcal{C}^\perp = \bigcup\{C^\perp : C \in \mathcal{C}\}$, where $C^\perp$ denotes the usual polar cone of C, that is, $C^\perp = \{y \in X^\dagger : \langle y, x\rangle \le 0 \text{ whenever } x \in C\}$.

6. Generalized differential quotients (GDQs)

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 $F_j \xrightarrow{\mathrm{igr}} F$—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 $j \ge j_\Omega$.

Definition 6.1 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 $F : X \mapsto Y$ is an ordinary (that is, single-valued and everywhere defined) map, then F belongs to $\mathrm{REG}(X;Y)$ if and only if F is continuous.

It is not hard to prove the following.

Theorem 6.2 *Assume that X, Y, Z are metric spaces. Let $F \in \mathrm{REG}(X;Y)$, $G \in \mathrm{REG}(Y;Z)$. Then the composite map $G \circ F$ belongs to* $\mathrm{REG}(X;Z)$. ◇

Definition 6.3 Let $m, n \in \mathbb{Z}_+$, let $F : \mathbb{R}^m \longrightarrow\!\!\!\!\to \mathbb{R}^n$ 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 $x \in U \cap S$. $\diamondsuit$

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 linear multimultimap $GDQ(F;\bar{x},\bar{y};S) \in \mathrm{LinMM}_c(T_{\bar{x}}M, T_{\bar{y}}N)$ 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 $\bar{x}$, $\bar{y}$ such that $\xi(\bar{x}) = 0$, $\eta(\bar{y}) = 0$, 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 the multimap $D\eta(\bar{y}) \circ \Lambda \circ D\xi(\bar{x})^{-1}$ 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. Moreover, the following four results can be proved.

Theorem 6.4 *GDQ is a GDT.* $\diamondsuit$

Theorem 6.5 *GDQ has the strong directional open mapping property with linear rate.* $\diamondsuit$

Theorem 6.6 *If* $F : \mathbb{R}^n \mapsto \mathbb{R}^m$ *is a continuous map,* $x \in \mathbb{R}^n$, *and* F *is classically differentiable at* x, *then* $\{DF(x)\} \in GDQ(F;x,F(x);\mathbb{R}^n)$. $\diamondsuit$

Theorem 6.7 *If* $F : \mathbb{R}^n \mapsto \mathbb{R}^m$ *is Lipschitz-continuous, and* $x \in \mathbb{R}^n$, *then the Clarke generalized Jacobian* $\partial F(x)$ *belongs to* $GDQ(F;x,F(x);\mathbb{R}^n)$. $\diamondsuit$

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.)

7. Vector fields, flows, variational generators

Definition 7.1 If $M \in \mathcal{M}^1$, then 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 $(x,t) \in \mathrm{Do}(f)$. More generally, if $E \in \mathcal{FDRLS}$, an *E-augmented time-varying vector field* on M is a ppd map

$$M \times \mathbb{R} \ni (x,t) \dashrightarrow f^@(x,t) = (f^0(x,t), f(x,t)) \in E \times TM$$

such that $f(x,t) \in T_xM$ whenever $(x,t) \in \mathrm{Do}(F)$.

We use $TVVF(M)$, $TVVF(M;E)$, to denote, respectively, the set of all TVVFs on M and the set of all E-augmented TVVFs on M.

The abbreviation "LAC" will stand for "locally absolutely continuous." If $f \in TVVF(M)$, then a *trajectory* (or *integral curve*) of f is a LAC map $\xi : I \mapsto M$, defined on a nonempty subinterval I of $\mathbb{R}$, such that $(\xi(t),t) \in \mathrm{Do}(f)$ and $\dot{\xi}(t) = f(\xi(t),t)$ for almost all $t \in I$. If $f^@ \in TVVF(M;E)$, and $f^@ = (f^0, f)$, then a *trajectory* (or *integral curve*) of $f^@$ is a trajectory of f, and an *augmented trajectory* (or *augmented integral curve*) of $f^@$ is a LAC map $I \ni t \to \xi^@(t) = (\xi^0(t), \xi(t)) \in E \times M$, defined on a nonempty subinterval I of $\mathbb{R}$, such that $(\xi(t),t)$ belongs to $\mathrm{Do}(F)$, $\dot{\xi}(t) = f(\xi(t),t)$ and $\dot{\xi}^0(t) = f^0(\xi(t),t)$ for a.e. $t \in I$.

If $f \in TVVF(M)$, $f^@ = (f^0, f) \in TVVF(M;E)$, we use $\mathrm{Traj}\,(f)$, $\mathrm{Traj}_c(f)$, $\mathrm{Traj}\,(f^@)$, $\mathrm{Traj}_c(f^@)$, $\mathrm{Traj}^@(f^@)$, $\mathrm{Traj}_c^@(f^@)$, to denote, respectively, the set of all trajectories of f, the set of all $\xi \in \mathrm{Traj}\,(f)$ whose domain is a compact interval, the sets $\mathrm{Traj}\,(f)$, $\mathrm{Traj}_c(f)$, the set of all augmented trajectories of $f^@$, and the set of all $\xi^@ \in \mathrm{Traj}^@(f^@)$ whose domain is a compact interval. $\diamondsuit$

Definition 7.2 Assume that $M \in \mathcal{M}^1$, $E \in \mathcal{FDRLS}$, $f \in TVVF(M)$, and $f^@ = (f^0, f) \in TVVF(M;E)$. If $(x,b,a) \in M \times \mathbb{R} \times \mathbb{R}$, we define

$$\Phi^f(x,b,a) = \{\xi(b) : \xi \in \mathrm{Traj}\,(f) : \xi(a) = x\}\,,$$
$$\Phi^{f^@}(x,b,a) = \Phi^f(x,b,a)\,,$$
$$\Phi^{@,f^@}(x,b,a) = \{\xi^@(b) : \xi^@ \in \mathrm{Traj}^@(f^@), \xi^@(a) = (x,0)\}.$$

The set-valued maps $\Phi^f : M \times \mathbb{R} \times \mathbb{R} \longrightarrow\!\!\!\!\rightarrow M$, $\Phi^{@,f^@} : M \times \mathbb{R} \times \mathbb{R} \longrightarrow\!\!\!\!\rightarrow E \times M$, are called the *flow* of f (or of $f^@$) and the *augmented flow* of $f^@$.

For each $(b,a) \in \mathbb{R} \times \mathbb{R}$, we define set-valued maps $\Phi^f_{b,a} : M \longrightarrow\!\!\!\!\rightarrow M$ and $\Phi^{@,f^@}_{b,a} : M \longrightarrow\!\!\!\!\rightarrow E \times M$, by letting $\Phi^f_{b,a}(x) = \Phi^f(x,b,a)$, $\Phi^{@,f^@}_{b,a}(x) = \Phi^{@,f^@} f(x,b,a)$ for $x \in M$. The set-valued maps $\Phi^f_{b,a} : M \longrightarrow\!\!\!\!\rightarrow M$, $\Phi^{@,f^@}_{b,a} : M \longrightarrow\!\!\!\!\rightarrow M$, are the *flow maps* of f (or of $f^@$) and the *augmented flow maps* of $f^@$. $\diamondsuit$

If $E \in \mathcal{FDRLS}$ and $A, B \in SVM(M, E \times M)$, then the "augmented composite" $A\#B \in SVM(M, E \times M)$ is defined by letting $(A\#B)(x)$ be, for each $x \in M$, the set $\{(\alpha+\beta, z) : (\exists y)((\alpha, y) \in B(x) \wedge (\beta, z) \in A(y)\}$,

Fact 7.3 *Let* $M \in \mathcal{M}^1$, $E \in \mathcal{FDRLS}$, $f \in TVVF(M)$, $f^@ = (f^0, f) \in TVVF(M;E)$. *Then the flow maps* $\Phi^f_{b,a}$ *satisfy the* flow identities $\boxed{\Phi^f_{a,a} = \mathbb{1}_M}$ *and* $\boxed{\Phi^f_{c,b} \circ \Phi^f_{b,a} = \Phi^f_{c,a}}$, *if* $a, b, c \in \mathbb{R}$ *and* $a \leq b \leq c$. *The augmented flow maps satisfy* $\boxed{\Phi^{@,f^@}_{a,a}(x) = \{(0,x)\}}$ *and* $\boxed{\Phi^{@,f^@}_{c,a} = \Phi^{@,f^@}_{c,b} \# \Phi^{@,f^@}_{b,a}}$. $\diamondsuit$

If $k, \ell \in \mathbb{Z}_+$, $M \in \mathcal{M}^k$, and $0 \leq \ell < k$, we let $J^\ell_{VF}(M)$ be the space of all ℓ-jets of vector fields on M. In particular, $J^0_{VF}(M)$ is canonically identified with TM. We also let $J^{-1}_{VF}(M) \stackrel{\mathrm{def}}{=} M$. If $\ell, \ell', k \in \mathbb{Z}_+ \cup \{-1\}$ and $\ell \leq \ell' < k$, then there is a natural projection $\Pi_{M,\ell',\ell} : J^{\ell'}_{VF}(M) \mapsto J^\ell_{VF}(M)$. If $x \in M$, we write $J^\ell_{VF}(M;x) \stackrel{\mathrm{def}}{=} \Pi^{-1}_{M,\ell,-1}(x)$, so $J^\ell_{VF}(M;x)$ is the set of all ℓ-jets at x of vector fields on M. The sets $J^\ell_{VF}(M;x)$ are linear spaces, and each $\Pi_{M,\ell',\ell}$ maps $J^{\ell'}_{VF}(M;x)$ linearly to $J^{\ell}_{VF}(M;x)$. In particular, if $x \in M$ and $v \in T_xM$, the space $\tilde{J}^1_{VF}(M;x,v) \stackrel{\text{def}}{=} \Pi^{-1}_{M,1,0}(x,v)$ is affine. Clearly, $\dim \tilde{J}^1_{VF}(M;x,v) = n^2$, if $n = \dim M$.

Definition 7.4 Let $M \in \mathcal{M}^2$, let $I \subseteq \mathbb{R}$ be a nontrivial interval, and let $\xi : I \mapsto M$ be a LAC curve. A *variational bundle* over ξ is a measurable set-valued map $I \ni t \mapsto \mathcal{V}(t) \subseteq \tilde{J}^1_{VF}(M;\xi(t),\dot{\xi}(t))$ having compact convex nonempty values. We use $VB(\xi)$ to denote the set of all variational bundles over ξ. If $\mathcal{V} \in VB(\xi)$, we call $\mathcal{V}$ *locally integrable* (abbr. LI) if, for every C^2 coordinate chart $\kappa : U \to \mathbb{R}^n$ on M, and every compact subinterval J of I such that $\xi(J) \subseteq U$, the absolute values of all the components relative to κ of all the members of $\mathcal{V}(t)$, for $t \in J$, are bounded by integrable functions of t. We use $VB_{LI}(\xi)$ to denote the set of all LI $\mathcal{V} \in VB(\xi)$. ♢

If $\mathcal{V} \in VB_{LI}(\xi)$, we write $\Gamma(\mathcal{V})$ to denote the set of all measurable selections of $\mathcal{V}$. Using coordinate charts, it easy to embed $\Gamma(\mathcal{V})$ as a convex weakly compact subset of a Banach space. If $\gamma \in \Gamma(\mathcal{V})$, then γ gives rise to a *connection* ∇^{γ} *along* ξ, in the manner described in detail in [5]. (In local coordinates, if we write $\partial_i \stackrel{\text{def}}{=} \frac{\partial}{\partial x_i}$, pick for each t a C^1 vector field X_t near $\xi(t)$ such that the 1-jet of X_t at $\xi(t)$ is $\gamma(t)$, and write $X_t(x) = \sum_{i=1}^n X^i_t(x)\partial_i$, we can define the "covariant derivative" $\nabla^{\gamma}v$ of a LAC vector field $t \mapsto v(t) = \sum_{i=1}^n v^i(t)\partial_i \in T_{\xi(t)}M$ along ξ by letting $\nabla^{\gamma}v(t) = \sum_{i=1}^n \Big(\dot{v}^i(t) - \sum_{j=1}^n (\partial_j X^i_t)(\xi(t))v^j(t)\Big)\partial_i$. To see that this makes intrinsic sense fix a t such that ξ and v are differentiable at t, and pick, for s near t, a C^1 vector field V_s such that $V_s(\xi(s)) = v(s)$ and the map $(s,x) \mapsto V_s(x)$ is differentiable at $(t,\xi(t))$. Write $V_s(x) = \sum_{i=1}^n V^i(s,x)\partial_i$. Then $v^i(s) = V^i(s,\xi(s))$, so

$$\begin{aligned}\dot{v}^i(t) &= \frac{\partial V^i}{\partial s}(t,\xi(t)) + \sum_{j=1}^n \frac{\partial V^i}{\partial x_j}(t,\xi(t)) \cdot \dot{\xi}^j(t) \\ &= \frac{\partial V^i}{\partial s}(t,\xi(t)) + \sum_{j=1}^n \frac{\partial V^i}{\partial x_j}(t,\xi(t)) \cdot X^j_t(\xi(t)) .\end{aligned}$$

Then $\nabla^{\gamma}v(t) = \frac{\partial V^i}{\partial s}(t,\xi(t)) + [X_t, V_t](\xi(t))$.) We can then define a *linear flow* $L^{\gamma} = \{L^{\gamma}_{t,s} : s,t \in I, s \leq t\}$, consisting of maps $L^{\gamma}_{t,s} : T_{\xi(s)}M \mapsto T_{\xi(t)}M$, by letting $L^{\gamma}_{t,s}(v)$ be, for $v \in T_{\xi(s)}M$, the "parallel translate of v to $T_{\xi(t)}M$ by means of ∇^{γ}," that is, the value at time t of the unique LAC vector field V along ξ such that $V(s) = v$ and $\nabla^{\gamma}V \equiv 0$, we can then define a *linear multiflow* $\Lambda^{\mathcal{V}} = \{\Lambda^{\mathcal{V}}_{t,s} : s,t \in I, s \leq t\}$, consisting of compact nonempty linear multimaps $\Lambda^{\mathcal{V}}_{t,s}$ from $T_{\xi(s)}M$ to $T_{\xi(t)}M$, by letting $\Lambda^{\mathcal{V}}_{t,s}$ be the set of all maps $L^{\gamma}_{t,s}$, for all $\gamma \in \Gamma(\mathcal{V})$.

If $E \in \mathcal{FDRLS}$, it is obvious how to define the concept of *E-augmented variational bundle* along a LAC curve ξ on M. If $\mathcal{V}^{@}$ is such a bundle, and $\mathcal{V}^{@}$ is LI, then each measurable selection $\gamma^{@} = (\gamma^0, \gamma) \in \Gamma(\mathcal{V}^{@})$ gives rise to an *augmented linear flow* $L^{@,\gamma^{@}} = \{L^{@,\gamma^{@}}_{t,s} : s,t \in I, s \leq t\}$, consisting of linear maps $L^{@,\gamma^{@}}_{t,s} : T_{\xi(s)}M \mapsto E \times T_{\xi(t)}M$, and this enables us to define an augmented linear multiflow $\Lambda^{@,\mathcal{V}^{@}} = \{\Lambda^{@,\mathcal{V}^{@}}_{t,s} : s,t \in I, s \leq t\}$, consisting of compact nonempty linear multimaps $\Lambda^{@,\mathcal{V}^{@}}_{t,s}$ from $T_{\xi(s)}M$ to $E \times T_{\xi(t)}M$.

Definition 7.5 Assume that $M \in \mathcal{M}^2$, $E \in \mathcal{FDRLS}$, $f \in TVVF(M)$, $f^{@} = (f^0, f) \in TVVF(M;E)$, $I \subseteq \mathbb{R}$ is a nontrivial interval, and $\xi : I \mapsto M$ is an integral curve of f. Let $\mathcal{D}$ be a GDT. An *f-variational* (or *$f^{@}$-variational*) *generator of $\mathcal{D}$-differentials over* ξ is a LI variational bundle $\mathcal{V}$ over ξ such that $\Lambda^{\mathcal{V}}_{t,s}$ belongs to $\mathcal{D}(\Phi^f_{t,s};\xi(s),\xi(t);M)$ for all $s,t \in I$ such that $s \leq t$. An *augmented $f^{@}$-variational generator of $\mathcal{D}$-differentials over* ξ is an E-augmented LI variational bundle $\mathcal{V}^{@}$ over ξ such that $\Lambda^{@,\mathcal{V}^{@}}_{t,s} \in \mathcal{D}(\Phi^{@,f^{@}}_{t,s};\xi(s),(\theta(s,t),\xi(t));M)$ for all $s,t \in I$ such that $s \leq t$, where we let $\theta(s,t) = \int_s^t f^0(\xi(r),r)\,dr$. ♢

Example 7.6 If f is defined on an open subset of $M \times I$ containing the graph of ξ, and $f(x,t)$ is measurable in t for each x, and Lipschitz in x for each t with Lipschitz constants $k(t)$ such that k is LI, then the Clarke generalized Jacobians $\partial f_t(\xi(t))$—where $f_t(x) \stackrel{\text{def}}{=} f(x,t)$—define an f-variational generator of $\mathcal{D}$-differentials over ξ, if $\mathcal{D}$ is the GDT of Warga derivate containers. ♢

Example 7.7 Sufficient conditions for a variational bundle $\mathcal{V}$ to be a an f-variational generator of GDQ-differentials are given in [8]. ♢

8. Points of regular $\mathcal{D}$-differentiability

Definition 8.1 Let $\mathcal{D}$ be a GDT. Assume that $M \in \mathcal{M}^1$, $E \in \mathcal{FDRLS}$, $f^{@} = (f^0, f) \in TVVF(M;E)$. A point $(x,t) \in M \times \mathbb{R}$ is a *point of regular $\mathcal{D}$-differentiability* of $f^{@}$ if $\boxed{\{L\} \in \mathcal{D}(\Phi^{@,f^{@}};(x,t,t),(0,x);M \times \mathbb{R} \times \mathbb{R})}$, where $L : T_xM \times \mathbb{R} \times \mathbb{R} \mapsto E \times T_xM$ is the linear map given by $\boxed{L(v,\tau,\sigma) = (0,v) + (\tau - \sigma)f^{@}(x,t)}$ for $v \in T_xM$, $\tau \in \mathbb{R}$, $\sigma \in \mathbb{R}$. ♢

9. A hybrid maximum principle

In [8], we stated and proved a hybrid maximum principle for GDQ theory. In [7], we stated and proved a general, abstract version of the maximum principle for arbitrary GDTs having the DOMP. Here we present an intermediate result, which can be regarded as a special "hybrid systems" case of the result of [7], or as an extension of the theorem of [8] to an arbitrary GDT $\mathcal{D}$. The proof is exactly as in [8], except that GDQs have to replaced throughout by $\mathcal{D}$-differentials.

From now on, the augmentation space E considered earlier will be $\mathbb{R}$, so all "augmented" objects will be $\mathbb{R}$-augmented. (By taking E to be $\mathbb{R}^d$ for $d > 1$ one gets, in a similar way, a maximum principle for Pareto optimality, but we will not discuss that here.)

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 $F^{0,1}, \ldots, F^{0,\mu}$ for $\Sigma^1, \ldots, \Sigma^{\mu}$, "switching constraints" $\mathcal{S}^1, \ldots, \mathcal{S}^{\mu}$, and "switching cost functions" $\varphi^1, \ldots, \varphi^{\mu}$.

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 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 $F^{0,i}$ is a family $\{F^{0,i}_\eta\}_{\eta\in\mathcal{U}^i}$ of ppd functions $F^{0,i}_\eta : Q^i\times\mathbb{R} \dashrightarrow \mathbb{R}$. 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 $\varphi^i : Q^i \times \mathbb{R} \times Q^{i\tilde{+}1} \times \mathbb{R} \to \mathbb{R}$. 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 integral curve of $F^i_{\eta^i}$, defined on a compact interval $I(\xi^i) = [a(\xi^i), b(\xi^i)]$. 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, $\sigma^i(\boldsymbol{\xi}) \in \mathcal{S}^i$, where $\hat{\xi}^i(t) \overset{\text{def}}{=} (\xi^i(t), t)$ and $\sigma^i(\boldsymbol{\xi}) \overset{\text{def}}{=} (\hat{\xi}^i(b(\xi^i)), \hat{\xi}^{i\tilde{+}1}(a(\xi^{i\tilde{+}1})))$, and the functions $I(\xi^i) \ni t \mapsto F^{0,i}_{\eta^i}(\hat{\xi}^i(t)) \in \mathbb{R}$ are a.e. defined—that is, $\hat{\xi}^i(t) \in \mathrm{Do}(F^{0,i}_{\eta^i})$ for a.e. $t \in I(\xi^i)$—and Lebesgue integrable. The *cost* of an admissible TCP $(\boldsymbol{\xi},\boldsymbol{\eta})$ is the number

$$J(\boldsymbol{\xi},\boldsymbol{\eta}) \overset{\text{def}}{=} \sum_{i=1}^{\mu} \int_{a(\xi^i)}^{b(\xi^i)} F^{0,i}_{\eta^i}(\hat{\xi}^i(t))\,dt + \sum_{i=1}^{\mu} \varphi^i\Big(\sigma^i(\boldsymbol{\xi})\Big)\,.$$

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

For each i, we define the *augmented dynamics* to be the family $F^{@,i} = \{F^{@,i}_\eta\}_{\eta\in\mathcal{U}^i}$ of ppd maps from $Q^i \times \mathbb{R}$ to $\mathbb{R} \times TQ^i$ given by $F^{@,i}_\eta(q,t) = (F^{0,i}_\eta(q,t), F^i_\eta(q,t))$ for $(q,t) \in Q^i \times \mathbb{R}$.

We write $\mathcal{Q}^i = Q^i \times \mathbb{R}$, $\mathcal{Q}^i_\# = \mathcal{Q}^i \times \mathcal{Q}^{i\tilde{+}1}$, and make the following assumptions.

H1. $\mathcal{D}$ is a GDT that has the DOMP.

H2. $(\boldsymbol{\xi},\boldsymbol{\eta}) = ((\xi^1,\ldots,\xi^\mu),(\eta^1,\ldots,\eta^\mu))$ is an admissible TCP, and $a^i = a(\xi^i)$, $b^i = b(\xi^i)$, $x^i = \xi^i(a^i)$, $y^i = \xi^i(b^i)$, for $i = 1,\ldots,\mu$.

H3. For $i = 1,\ldots,\mu$, $\Lambda^{@,i}$ is an augmented $F^{@,i}_{\eta^i}$-variational generator of $\mathcal{D}$-differentials over ξ^i.

H4. For $i = 1,\ldots,\mu$, $\Xi^i_- = (x^i, a^i)$ and $\Xi^i_+ = (y^i, b^i)$ are points of regular $\mathcal{D}$-differentiability of $F^{@,i}_{\eta^i}$.

H5. For each $i \in \{1,\ldots,\mu\}$, $\mathcal{C}^i$ is a convex multicone in $T_{\Xi^i_+}\mathcal{Q}^i \times T_{\Xi^{i\tilde{+}1}_-}\mathcal{Q}^{i\tilde{+}1}$ (that is, in $T_{(\Xi^i_+,\Xi^{i\tilde{+}1}_-)}\mathcal{Q}^i_\#$), and is a $\mathcal{D}$-approximating multicone for $\mathcal{S}^i$ 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 $\Big(T_{y^i}Q^i \times \mathbb{R} \times T_{x^{i\tilde{+}1}}Q^{i\tilde{+}1} \times \mathbb{R}\Big)^\dagger$ (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 $GDQ(\varphi^i; P^i, \varphi^i(P^i); Q^i \times \mathbb{R} \times Q^{i\tilde{+}1} \times \mathbb{R})$.

H7. Each system Σ^i is invariant under time-interval substitutions. (That is, if η,ζ belong to $\mathcal{U}^i$, $\xi \in C^0([a,b]\,;Q^i) \cap \mathrm{Traj}_c(F^i_\eta)$, 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 an "adjoint pair" along $(\boldsymbol{\xi},\boldsymbol{\eta})$, and what it means for such a pair 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

$$H^i_\zeta(q,\lambda,t,\lambda_0) = \lambda \cdot F^i_\zeta(q,t) - \lambda_0 F^{0,i}_\zeta(q,t)\,.$$

Definition 9.1 If H1-H7 hold, then an *adjoint pair along* $(\boldsymbol{\xi},\boldsymbol{\eta})$ is a pair $(\boldsymbol{\psi},\psi_0)$ with the property that:

- $\boldsymbol{\psi}$ is a μ-tuple $(\psi^1,\ldots,\psi^\mu)$ such that each ψ^i is an absolutely continuous field of covectors along ξ^i;
- $\psi_0 \in \mathbb{R}$ and $\psi_0 \ge 0$;
- each ψ^i satisfies the *adjoint differential inclusion*

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

- for each $i \in \{1,\ldots,\mu\}$, the *switching condition*

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

holds, where

$$\begin{aligned} h^i_+ &= H^i_{\eta^i}(y^i, \psi^i(b^i), b^i, \psi_0)\,, \\ h^i_- &= H^i_{\eta^i}(x^i, \psi^i(a^i), a^i, \psi_0)\,. \end{aligned} \qquad \diamondsuit$$

Definition 9.2 If H1-H7 hold, and $(\boldsymbol{\psi},\psi_0)$ is an adjoint pair along $(\boldsymbol{\xi},\boldsymbol{\eta})$, we say that $(\boldsymbol{\psi},\psi_0)$ satisfies the *Hamiltonian maximization condition* if, for every $i \in \{1,\ldots,\mu\}$, the inequality

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

holds whenever $\zeta \in \mathcal{U}^i$, $t \in]a^i, b^i[$, and $(\xi^i(t), t)$ is a point of regular $\mathcal{D}$-differentiability of $F^{@,i}_{\eta^i}$ and $F^{@,i}_\zeta$. $\diamondsuit$

Definition 9.3 If $(\boldsymbol{\psi},\psi_0)$ is an adjoint pair along $(\boldsymbol{\xi},\boldsymbol{\eta})$, we say that $(\boldsymbol{\psi},\psi_0)$ satisfies the *nontriviality condition* if either $\psi_0 \ne 0$ or at least one of the functions ψ^i is not identically zero. $\diamondsuit$

Theorem 9.4 *If H1-H7 hold, and the pair* $(\boldsymbol{\xi},\boldsymbol{\eta})$ *is optimal, then there exists an adjoint pair along* $(\boldsymbol{\xi},\boldsymbol{\eta})$ *that satisfies the Hamiltonian maximization and nontriviality conditions.* $\diamondsuit$

Remark 9.5 The switching condition of Definition 9.1 takes a particularly simple form in the three most common situations, namely,

(a) free switching times,
(b) fixed switching times,
(c) equal but otherwise free switching times.

Case (a) arises when the switching constraint $\mathcal{S}^i$ is "independent of the times" (that is, a set of the form $\{(x,t,y,s) \in Q^i \times \mathbb{R} \times Q^{i\tilde{+}1} \times \mathbb{R} : (x,y) \in \mathcal{S}^i_0\}$, where $\mathcal{S}^i_0$ is a subset of $Q^i \times Q^{i\tilde{+}1}$), the times $b^i, a^{i\tilde{+}1}$ are totally free, and the function $(x,t,y,s) \mapsto \varphi^i(x,t,y,s)$

is independent of t and s. In that case, if $\mathcal{C}_0^i$ is a $\mathcal{D}$-approximating multicone to $\mathcal{S}_0^i$ at $(y^i, x^{i\tilde{+}1})$, we can take $\mathcal{C}^i$ to be the set of all cones $\hat{C}$, $C \in \mathcal{C}_0^i$, where, for a cone C in $T_{y^i}Q^i \times T_{x^{i\tilde{+}1}}Q^{i\tilde{+}1}$, $\hat{C}$ is the cone $\{(v,\tau,w,\sigma) \in T_{y^i}Q^i \times \mathbb{R} \times T_{x^{i\tilde{+}1}}Q^{i\tilde{+}1} \times \mathbb{R} : (v,w) \in C\}$. Then the polar $\hat{C}^\perp$ of $\hat{C}$ turns out to be the set $\{(\alpha,0,\beta,0) : (\alpha,\beta) \in C^\perp\}$. Also, we can take Ω^i to be a subset of $T^*_{y^i}Q^i \times \{0\} \times T^*_{x^{i\tilde{+}1}}Q^{i\tilde{+}1} \times \{0\}$. It then follows from the switching condition that $h_+^i = h_-^{i\tilde{+}1} = 0$, that is, "the Hamiltonian vanishes when the switching times are free."

Case (b) arises when the switching constraint $\mathcal{S}^i$ is a subset of $Q^i \times \{b^i\} \times Q^{i\tilde{+}1} \times \{a^{i\tilde{+}1}\}$. In that case we can assume that the multicone $\mathcal{C}^i$ is contained in $T_{y^i}Q^i \times \{0\} \times T_{x^{i\tilde{+}1}}Q^{i\tilde{+}1} \times \{0\}$. It follows that the switching condition imposes no restriction on h_+^i and $h_-^{i\tilde{+}1}$, so that "the values of the Hamiltonian are free when the switching times are fixed."

Finally, Case (c) arises when the switching cost function $(x,t,y,s) \mapsto \varphi^i(x,t,y,s)$ is independent of t and s, and the switching constraint $\mathcal{S}^i$ involves a restriction on the switching states only, plus the requirement that the times b^i and $a^{i\tilde{+}1}$ be equal but otherwise arbitrary. Once again, we can take Ω^i to be a subset of the product $T^*_{y^i}Q^i \times \{0\} \times T^*_{x^{i\tilde{+}1}}Q^{i\tilde{+}1} \times \{0\}$. Moreover, we can assume that $\mathcal{C}^i$ consists of cones $\check{C}$, where, for a cone C in $T_{y^i}Q^i \times T_{x^{i\tilde{+}1}}Q^{i\tilde{+}1}$, $\check{C}$ is the cone $\{(v,r,w,r) \in T_{y^i}Q^i \times \mathbb{R} \times T_{x^{i\tilde{+}1}}Q^{i\tilde{+}1} \times \mathbb{R} : (v,w) \in C\}$. Clearly, the polar $\check{C}^\perp$ of a cone $\check{C}$ is the set $\{(\alpha,\rho,\beta,-\rho) : (\alpha,\beta) \in C^\perp, \rho \in \mathbb{R}\}$. It then follows from the switching condition that $h_+^i = h_-^{i\tilde{+}1}$, that is, "the Hamiltonian does not jump when the switching times are equal but otherwise free."

The most common situation obtains when we are in Case (c) for $i = 1,\ldots,\mu-1$, that is, when "the true switching times are free but there is no clock-resetting at the switchings." $\diamondsuit$

10. Path-integral generalized differentials

GDQ theory appears to be broad enough to yield a very general maximum principle. On the other hand, Theorem 6.7 is not valid if we substitute "Warga derivate container" for "Clarke generalized Jacobian," because it is possible to exhibit Lipschitz maps that have derivate containers that are not GDQs. It is therefore reasonable to look for even more general GDTs that might contain both GDQ theory and other theories such as the derivate containers and the multidifferentials. The following is one example of such a theory.

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 continuous function $\beta : [0,1] \to \mathbb{R}^m$ given by $\beta(t) = \int_0^t h(s) \cdot \dot{\alpha}(s)\,ds$.

Let $n \in \mathbb{Z}_+$, $r \in \mathbb{R}$, and let S be a subset of $\mathbb{R}^n$. We write $\mathcal{A}(S(r))$ 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 (i) $\alpha(0) = 0$, and (ii) $\dot{\alpha}(t) \in S$ and $\|\dot{\alpha}(t)\| \le r$ for almost all $t \in [0,1]$.

Definition 10.1 Let $n, m \in \mathbb{Z}_+$, let $F : \mathbb{R}^n \longrightarrow\!\!\!\!\rightarrow \mathbb{R}^m$, 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 $\delta \in\,]0,\infty[$ there exists $R \in\,]0,\infty[$ such that

(#) for every $r \in\,]0,R]$ there exists a regular set-valued map $G : \mathcal{A}(C(r)) \longrightarrow\!\!\!\!\rightarrow C^0([0,1];\mathbb{R}^{m\times n}) \times \mathbb{R}^m$ with the property that
  (#.a) $h(t) \in \Lambda^\delta$ and $\|v\| \le \delta r$ 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 $x \in \mathcal{A}(C(r))$, then $\Phi_G(x)$ is the set of all $y \in \mathbb{R}^m$ for which the equality $y = (h*\alpha)(1) + v$ holds for some triple $(\alpha,h,v) \in \mathrm{Gr}(G)$ such that $\alpha(1) = x$. $\diamondsuit$

It then turns out that every GDQ is a PIGD, and every derivate container, semidifferential, and multidifferential is also a PIGD. Moreover, the definition of the PIGDs can easily be extended to manifolds, and one obtains a GDT that has the DOMP and contains all other theories proposed so far.

References

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

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

[3] 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.

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

[5] 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.

[6] 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.

[7] Sussmann, H. J., *Résultats récents sur les courbes optimales.* In *15^e Journée Annuelle de la Société Mathémathique de France (SMF)*, Publications de la SMF, Paris, 2000, pp. 1-52.

[8] Sussmann, H. J., *New theories of set-valued differentials and new versions of the maximum principle of optimal control theory.* In *Nonlinear Control in the year 2000*, A. Isidori, F. Lamnabhi-Lagarrigue and W. Respondek Eds., Springer-Verlag, London, 2000, pp. 487-526.

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