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
sussmann@hamilton.rutgers.edu,
http://www.math.rutgers.edu/~sussmann

1 Introduction

The purpose of this note is to announce two new theories of generalized differentials—the "generalized differential quotients," abbr. GDQs, and the "path-integral generalized differentials", abbr. PIGDs—which have good open mapping properties and lead to general versions of the maximum principle. In particular, we use GDQ theory to state—in Theorem 5— 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 other detailed papers 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]).

For this reason, we conclude the paper by giving, in §11, a brief sketch of the definition of our second theory—the PIGDs—that contains that of GDQs as well as the other theories mentioned above.

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

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 $F \in SVM(A, B)$.

If F_1 and F_2 are set-valued maps, then the *composite* $F_2 \circ F_1$ is defined iff $\mathrm{Ta}(F_1) = \mathrm{So}(F_2)$ and in that case:

$$\mathrm{So}(F_2 \circ F_1) \stackrel{\mathrm{def}}{=} \mathrm{So}(F_1)$$

$$\mathrm{Ta}(F_2 \circ F_1) \stackrel{\mathrm{def}}{=} \mathrm{Ta}(F_2)$$

$$\mathrm{Gr}(F_2 \circ F_1) \stackrel{\mathrm{def}}{=} \{(x, z) : \quad (\exists y) \quad ((x, y) \in \mathrm{Gr}(F_1), (y, z) \in \mathrm{Gr}(F_2))\}.$$

If A is a set, then $\mathbb{I}_A$ denotes the *identity map* of A, that is, the triple (A, A, Δ_A), where $\Delta_A = \{(x, x) : \quad x \in A\}$.

Throughout this paper, the word "map" always stands for "set-valued map." 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 set A to a set 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.

A *cone* in a real linear space X is a nonempty subset C of X such that $r \cdot c \in C$ whenever $c \in C, r \in \mathbb{R}$ and $r \geq 0$.

We use $\mathbb{N}$ to denote the set of strictly positive integers, and write $\mathbb{Z}_+ \stackrel{\mathrm{def}}{=} \mathbb{N} \cup \{0\}$.

If $n \in \mathbb{Z}_+, r \in \mathbb{R}$, and $r > 0$, we use $\bar{\mathbb{B}}^n(r), \mathbb{B}^n(r)$ to denote, respectively, the closed and open balls in $\mathbb{R}^n$ with radius r. We write $\bar{\mathbb{B}}^n, \mathbb{B}^n$ for $\bar{\mathbb{B}}^n(1), \mathbb{B}^n(1)$.

If $k \in \mathbb{N}$ and M is a manifold of class C^k, then TM and T^*M denote the tangent and cotangent bundles of M, so TM and T^*M are manifolds of class C^{k-1}. If $x \in M$, then T_xM and T_x^*M denote the tangent and cotangent spaces of M at x.

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

Definition 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 belongs to $\mathrm{REG}(X;Y)$ if and only if F is continuous.

It is not hard to prove the following

Theorem 1. *Let X, Y, Z be metric spaces, and suppose that F belongs to* $\mathrm{REG}(X;Y)$ *and G belongs to* $\mathrm{REG}(Y;Z)$. *Then the composite map $G \circ F$ belongs to* $\mathrm{REG}(X;Z)$. ♢

4 Generalized differential quotients (GDQs)

Definition 2. 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 $i = 1, 2, 3$, $S_i \subseteq M_i$, $F_i : M_i \longrightarrow\!\!\!\!\to M_{i+1}$, 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 compact neighborhood U of $\bar{x}_2$ 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 $i = 1, 2$, $\bar{x}_i \in M_i$, $\bar{y}_i \in N_i$, $S_i \subseteq M_i$, $F_i : M_i \longrightarrow\!\!\!\!\to N_i$, and Λ_i belongs to $GDQ(F_i; \bar{x}_i, \bar{y}_i; S_i)$, 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 $i = 1, 2$, $S_i \subseteq M$, $F_i : M \longrightarrow\!\!\!\!\to N$, and there exist neighborhoods U, V of $\bar{x}$, $\bar{y}$, in M, N, respectively, such that $(U \times V) \cap \mathrm{Gr}(F_1) = (U \times V) \cap \mathrm{Gr}(F_2)$ and $U \cap S_1 = U \cap S_2$, 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 2. *Let m, n be nonnegative integers, and let C be a convex cone in $\mathbb{R}^m$. Let $F : \mathbb{R}^m \longrightarrow\!\!\!\!\to \mathbb{R}^n$ 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 $D \subseteq \mathrm{Int}(\Delta) \cup \{0\}$, 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) belongs to Z_y. ♢

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

Definition 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 an $m \in \mathbb{Z}_+$, a map $F : \mathbb{R}^m \longrightarrow\!\!\!\!\!\rightarrow M$, 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 $X^\dagger$ denotes the dual of X. If S is a subset of X, the *polar* of S in X is the set

$$S_X^\perp = \{y \in X^\dagger : y(x) \leq 1 \text{ whenever } x \in S\}.$$

If C is a cone in X, then $C_X^\perp$ is a closed convex cone in $X^\dagger$, and

$$C_X^\perp = \{y \in X^\dagger : y(x) \leq 0 \text{ whenever } x \in C\}.$$

When it is clear from the context what the space X is, we will write $C^\perp$ rather than $C_X^\perp$. We remark, however, that if C is a cone in a linear subspace Y of a linear space X, then $C_Y^\perp$ and $C_X^\perp$ are different objects, and this distinction will be crucial in the statement of our main result (cf. the definition of "multiplier," Def. 10).

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 $X^\dagger$.

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 the real line, 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 $\mathcal{BL}(\mathbb{R}^n, \mathbb{R}) \cup \mathcal{N}(\mathbb{R}^n, \mathbb{R})$. 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 4. 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 $\mathcal{M}_{\mathcal{BL}}^{e,loc}(\mathbb{R}^n \times \mathbb{R}; \mathbb{R}^m)$ 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 1 *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 the function $I \ni t \mapsto f(\xi(t), t) \in \mathbb{R}^m$ is measurable.* $\diamondsuit$

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. 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 $\mathbb{R} \ni t \to \psi(t) \in [0, +\infty]$ with the property that $||f(x, t)|| \leq \psi(t)$ 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 6. 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 $f(x, t) \in T_x M$ 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 $\text{Traj}\,(f)$ to denote the set of all trajectories of f, and $\text{Traj}_{\,c}(f)$ to denote the set of all $\xi \in \text{Traj}\,(f)$ whose domain is a compact interval.
5. For $(x, b, a) \in M \times \mathbb{R} \times \mathbb{R}$, we define
$$\Phi^f(x, b, a) = \{\, \xi(b) : \xi \in \text{Traj}\,(f) : \xi(a) = x \,\}\,. \tag{3}$$
The map $\Phi^f : M \times \mathbb{R} \times \mathbb{R} \longrightarrow\!\!\!\!\rightarrow M$ is the *flow* of f.
6. For each $(b, a) \in \mathbb{R} \times \mathbb{R}$, we define a map $\Phi^f_{b,a} : M \longrightarrow\!\!\!\!\rightarrow M$ by letting $\Phi^f_{b,a}(x) = \Phi^f(x, b, a)$ for $x \in M$. The maps $\Phi^f_{b,a} : M \longrightarrow\!\!\!\!\rightarrow M$ are the *flow maps* of f. $\diamondsuit$

Fact 2 *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 $\Phi^f_{a,a} = \mathbb{I}_M$ and $\Phi^f_{c,b} \circ \Phi^f_{b,a} = \Phi^f_{c,a}$, if $a, b, c \in \mathbb{R}$ and $a \leq b \leq c$.* $\diamondsuit$

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
$$\sigma_{f,\bar{x},\mathbf{v},\beta}(t) = \sup\{\, \text{dist}(f(x,t), \mathbf{v}) : x \in \mathbb{R}^n\,,\ \|x - \bar{x}\| \leq \beta \,\}\,,$$
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 $\|x - \bar{x}\| \leq \beta$ 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
$$S_{\bar{x},I,\beta} = \{(x,t) : x \in \mathbb{R}^n, \|x - \bar{x}\| \leq \beta, t \in I\}\,. \tag{4}$$

Fact 3 *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 $\mathbf{v} \neq \emptyset$. Let I be a nonempty subinterval of $\mathbb{R}$, and assume that $\bar{\beta} > 0$ and $S_{x,I,\bar{\beta}} \subseteq \text{Do}(f)$. Then $\sigma_{f,\bar{x},\mathbf{v},\beta}$ is measurable on I whenever $0 < \beta \leq \bar{\beta}$.* $\diamondsuit$

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 $\{\bar{t} + r : r \in C, |r| \leq h\}$, $\{\bar{t} + r : r \in C\}$, respectively.

Definition 7. Assume that $n, m \in \mathbb{Z}_+$, f belongs to $\mathcal{M}^{e,loc}_{\mathcal{BL}}(\mathbb{R}^n \times \mathbb{R}; \mathbb{R}^m)$, $\mathbf{v} \subseteq \mathbb{R}^m$, $\mathbf{v} \neq \emptyset$, $(\bar{x}, \bar{t}) \in \mathbb{R}^n \times \mathbb{R}$. 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

$$\lim_{(h,\beta)\to(0,0)\,,\ h>0,\ \beta\geq 0} \frac{1}{h}\int_{\bar{t}+C(h)} \sigma_{f,\bar{x},\mathbf{v},\beta}(t)\,dt = 0\ .$$

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 $\{v\} \in \text{App–lim}_{x\to\bar{x}\,,\,t\to\bar{t},t-\bar{t}\in C} f(x,t)$. ♢

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 1. *Assume that $n, m \in \mathbb{Z}_+$, $C \in \mathcal{C}(1)$, $C \neq \{0\}$, $(\bar{x}, \bar{t}) \in \mathbb{R}^n \times \mathbb{R}$, 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) $||f(x,t)|| \leq \varphi(t)$ 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 8. 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
$$k^\alpha : [a,b] \to [0,+\infty], \quad \text{for} \quad 0<\alpha\le\bar{\alpha},$$
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 the curve ξ. $\diamondsuit$

If n, m, f, a, b, ξ, Λ are as in Definition 8, 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 4 *If $\mathcal{B} \subseteq L^1([a,b],\mathbb{R}^{n\times n})$ and $\mathcal{B}$ is bounded, then the map*

$$\mathcal{B}_{weak} \ni L \mapsto M_L \in C^0(\,[a,b]\times[a,b]\,;\mathbb{R}^{n\times n}\,)$$

is continuous. ♢

If n, m, f, a, b, ξ, Λ are as in Definition 8, 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 5 *$\mathcal{M}(\Lambda)$ 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 $\mathbb{R}^{n\times n}$. ♢

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

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

Then the set $[\mathcal{M}_{b,a}(\Lambda);\mathbf{v}_+,\mathbf{v}_-]$ belongs to

$$GDQ\Big(\Phi^f;(\xi(a),b,a),\xi(b);Q\times C^{b,a}\Big).$$ ♢

Theorem 3, combined with the chain rule, yields a result for products of flows. To state this result, we first define, if $\mathbf{M}^1, \ldots, \mathbf{M}^m \subseteq \mathbb{R}^{n\times n}$ and $\mathbf{v}_+^1, \ldots, \mathbf{v}_+^m, \mathbf{v}_-^0, \ldots, \mathbf{v}_-^{m-1} \subseteq \mathbb{R}^n$, a set

$$[\mathbf{M}^1, \ldots, \mathbf{M}^m; \mathbf{v}_+^1, \ldots, \mathbf{v}_+^m, \mathbf{v}_-^0, \ldots, \mathbf{v}_-^{m-1}]$$

of linear maps from $\mathbb{R}^n \times \mathbb{R}^{m+1}$ to $\mathbb{R}^n$. Precisely, this set consists of all the maps $[M^1, \ldots, M^m; v_+^1, \ldots, v_+^m, v_-^0, \ldots, v_-^{m-1}]$, for all $M^1 \in \mathbf{M}^1, \ldots,$ $M^m \in \mathbf{M}^m$, $v_+^1 \in \mathbf{v}_+^1, \ldots, v_+^m \in \mathbf{v}_+^m$, $v_-^0 \in \mathbf{v}_-^0, \ldots, v_-^{m-1} \in \mathbf{v}_-^{m-1}$, given by

$$\begin{aligned}&[M^1, \ldots, M^m; v_+^1, \ldots, v_+^m, v_-^0, \ldots, v_-^{m-1}](y, \alpha^m, \ldots, \alpha^0)\\ &\stackrel{\text{def}}{=} P^{m,0}\cdot y + \alpha^m v_+^m - \alpha^0 P^{m,0} v_-^0 + \sum_{k=1}^{m-1} \alpha^k P^{m,k}\cdot (v_+^k - v_-^k)\end{aligned}$$

where $\quad P^{\ell,k} = M^\ell \cdot M^{\ell-1} \cdot \ldots \cdot M^{k+1}$ if $k, \ell \in \mathbb{N}, 1 \leq k \leq \ell \leq m$.

Also, if $\mathbf{f} = (f^1, \ldots, f^m)$ is an m-tuple of time-varying vector fields on $\mathbb{R}^n$, we define the *product flow map*

$$\Phi^{\mathbf{f}} : \mathbb{R}^n \times \mathbb{R}^{m+1} \longrightarrow\!\!\!\!\rightarrow \mathbb{R}^n$$

by letting

$$\Phi^{\mathbf{f}}(x, a^m, \ldots, a^1, a^0) \stackrel{\text{def}}{=} \left(\Phi^{f^m}_{a^m, a^{m-1}} \circ \Phi^{f^{m-1}}_{a^{m-1}, a^{m-2}} \circ \ldots \circ \Phi^{f^1}_{a^1, a^0}\right)(x)$$

for $(x, a^m, \ldots, a^1, a^0) \in \mathbb{R}^n \times \mathbb{R}^{m+1}$.

Remark 1. 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 $J_x^1(VF(M)) \times T_x^* M$, where $J_x^1(VF(M))$ is the space of 1-jets at x of smooth vector fields on M. Then, if a, b, ξ are such that $a, b \in \mathbb{R}$, $a \leq b$, 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 8.

With this definition of variational generator Theorem 3 remains true on manifolds.

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

Theorem 4. *Let* n, m, $\mathbf{a}_*$, $a_*^0,\ldots,a_*^m$, $f^1,\ldots,f^m$, $\mathbf{f}$, $x_*^0,\ldots,x_*^m$, $\xi_*^1,\ldots,\xi_*^m$, $\mathbf{C}$, C, $C^0,\ldots,C^m$, *be such that* $n\in\mathbb{Z}_+$, $m\in\mathbb{N}$,

$$\mathbf{a}_*=(a_*^m,\ldots,a_*^0)\in\mathbb{R}^{m+1},\qquad a_*^0\le a_*^1\le\cdots\le a_*^m\,,$$

$C^0,\ldots,C^m\in\mathcal{C}(1)$, $C=C^m\times C^{m-1}\times\ldots\times C^1\times C^0$,

$$\mathbf{C}=C^{\mathbf{a}_*}=\mathbf{a}_*+C=(a_*^m+C^m)\times\ldots\times(a_*^0+C^0)\,,$$

and $\mathbf{f}=(f^1,\ldots,f^m)\in\Big(TVVF(\mathbb{R}^n)\Big)^m$. *Assume that, for* $i=1,\ldots,m$,

1. f^i *is locally integrally continuous;*
2. $\xi_*^i\in C^0(\,[a_*^{i-1},a_*^i]\,;\mathbb{R}^n\,)\cap\mathrm{Traj}\,_c(f^i)$,
3. $\xi_*^i(a_*^{i-1})=x_*^{i-1}$, *and* $\xi_*^i(a_*^i)=x_*^i$;
4. Λ^i *is a variational generator for* f^i *along* ξ_*^i,
5. $\mathbf{v}_+^i$, $\mathbf{v}_-^{i-1}$ *are nonempty compact convex subsets of* $\mathbb{R}^n$ *such that*

$$\mathbf{v}_+^i\in\text{App–lim}_{x\to x_*^i\,,\,t\to a_*^i,t-a_*^i\in C^i}f^i(x,t)\,,$$
$$\mathbf{v}_-^{i-1}\in\text{App–lim}_{x\to x_*^{i-1}\,,\,t\to a_*^{i-1},t-a_*^{i-1}\in C^{i-1}}f^i(x,t)\,.$$

Let $\mathbf{M}^i=\mathcal{M}_{a_*^i,a_*^{i-1}}(\Lambda^i)$, *for* $i=1,\ldots,m$. *Then the set*

$$[\mathbf{M}^1,\ldots,\mathbf{M}^m;\mathbf{v}_+^1,\ldots,\mathbf{v}_+^m,\mathbf{v}_-^0,\ldots,\mathbf{v}_-^{m-1}]$$

belongs to $GDQ\Big(\Phi^{\mathbf{f}};(x_*^0,a_*^m,\ldots,a_*^0),x_*^m;\mathbb{R}^n\times\mathbf{C}\Big)$. ◊

9 A GDQ maximum principle

Theorem 3, together with the directional open mapping theorem 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 $F^i=\{F_\eta^i\}_{\eta\in\mathcal{U}^i}$ such that, for each $\eta\in\mathcal{U}^i$, F_η^i is a ppd time-varying vector field on Q^i. Each L^i is a family $\{L_\eta^i\}_{\eta\in\mathcal{U}^i}$ of ppd functions $L_\eta^i:Q^i\times\mathbb{R}\dashrightarrow\mathbb{R}$. For each $i\in\{1,\ldots,\mu\}$, the switching constraint $\mathcal{S}^i$ is a subset of the product $Q^i\times\mathbb{R}\times Q^{i\tilde{+}1}\times\mathbb{R}$, 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

$$\boldsymbol{\eta} = (\eta^1, \dots, \eta^\mu) \in \mathcal{U}^1 \times \cdots \times \mathcal{U}^\mu .$$

A *trajectory* for a controller $\boldsymbol{\eta} = (\eta^1, \dots, \eta^\mu)$ is a μ-tuple $\boldsymbol{\xi} = (\xi^1, \dots, \xi^\mu)$ 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

$$\hat{\xi}^i(t) \stackrel{\mathrm{def}}{=} (\xi^i(t), t) \quad \text{and} \quad \sigma^i(\boldsymbol{\xi}) \stackrel{\mathrm{def}}{=} (\hat{\xi}^i(a_+(\xi^i)), \hat{\xi}^{i\tilde{+}1}(a_-(\xi^{i\tilde{+}1}))).$$

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 index 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

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

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

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

H3. C^i_σ are, for $i = 1, \dots, \mu$, $\sigma \in \{+, -\}$, 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 $\sigma \in \{+,-\}$,

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

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

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

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

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

$$\Big(T_{x^i_+}Q^i \times \mathbb{R} \times T_{x^{i\tilde{+}1}}Q^{i\tilde{+}1} \times \mathbb{R}\Big)^\dagger$$

(that is, of $\Big(T_{\Xi^i_+}(Q^i \times \mathbb{R}) \times T_{\Xi^{i\tilde{+}1}_-}(Q^{i\tilde{+}1} \times \mathbb{R})\Big)^\dagger$, or, equivalently, of $\Big(T_{(\Xi^i_+,\Xi^{i\tilde{+}1}_-)}(Q^i \times \mathbb{R} \times Q^{i\tilde{+}1} \times \mathbb{R})\Big)^\dagger$), and Ω^i belongs to the generalized differential quotient $GDQ(\varphi^i; P^i, \varphi^i(P^i); \mathcal{Q}^i_+ \times \mathcal{Q}^{i\tilde{+}1}_-)$.

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. 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 belongs to $\{1, \ldots, \mu\}$ and $\sigma \in \{+,-\}$;

- each ψ^i is absolutely continuous and satisfies the *adjoint differential inclusion*

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

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

$$\check{\Psi}_i \in \psi_0 \Omega^i + (\mathcal{C}^i)^{\perp}_{\check{\Xi}_i}, \qquad \kappa^i_\sigma \in (C^i_\sigma)^{\perp}_{R^i_\sigma}$$

hold, where

$$\check{\Psi}_i = (-\psi^i(a^i_+), -\kappa^i_+ + h^i_+, \psi^{i\tilde{+}1}(a^{i\tilde{+}1}_-), -\kappa^i_- - h^{i\tilde{+}1}_-)\,,$$
$$\check{\Xi}^{\perp}_i = T_{(\Xi^i_+, \Xi^{i\tilde{+}1}_-)}(Q^i \times \mathbb{R} \times Q^{i\tilde{+}1}_- \times \mathbb{R})\,,$$

and

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

Remark 2. The switching conditions take a more familiar form in the case of "fixed switching times" (that is, when the sets T^i_σ consist of the single points a^i_σ) or of "totally free switching times," that is, when the T^i_σ are equal to the whole real line $\mathbb{R}$ or, more generally, are neighborhoods of the a^i_σ. Indeed, in both cases we can take $C^i_\sigma = R^i_\sigma$, and then $(C^i_\sigma)^{\perp}_{R^i_\sigma} = \{0\}$. It follows that the κ^i_σ vanish, and the switching condition becomes

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

Suppose, in addition, that either

I. we are in the fixed switching times case and the switching conditions are "$(x^i_+, x^{i\tilde{+}1}_-) \in S^i_0$," where each S^i_0 is a subset of $Q^i \times Q^{i\tilde{+}1}$,

or

II. we are in the free switching times case, the switching conditions are of the form $(x^i_+, x^{i\tilde{+}1}_-) \in S^i_0$, where each S^i_0 is a subset of $Q^i \times Q^{i\tilde{+}1}$, the switching cost functions φ_i do not depend on the times, and the times a^i_σ are required to satisfy $a^i_+ = a^{i\tilde{+}1}_-$.

Then, in case I, each $\mathcal{C}_i$ will be a multicone in the product

$$T_{x^i_+} Q^i \times \{0\} \times T_{x^{i\tilde{+}1}_-} Q^{i\tilde{+}1} \times \{0\},$$

so the switching condition will not impose any restriction on the h^i_σ. On the other hand, in Case II each $\mathcal{C}_i$ will be a multicone in the set

$$\{(v, r, w, s) \in T_{x^i_+} Q^i \times \mathbb{R} \times T_{x^{i\tilde{+}1}_-} Q^{i\tilde{+}1} \times \mathbb{R} : r = s\}.$$

Hence all the members $(\hat{v}, \hat{r}, \hat{w}, \hat{s})$ of $\mathcal{C}_i^{\perp}$ will satisfy $\hat{r} + \hat{s} = 0$. Moreover, the fact that the φ_i do not depend on the times implies that we can choose the Ω_i to have vanishing time components. It then follows that the time-part of the switching condition becomes the familiar requirement that

$$h_+^i = h_{i+1}^- ,$$

that is, the condition that the Hamiltonian should not jump at the switchings.

Definition 10. 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

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

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 $\tilde{F}_{\eta^i}^i$ and $\tilde{F}_\zeta^i$. ♢

Definition 11. 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 5. *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 5 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 becomes the transversality condition. When the augmented vector fields $\tilde{F}_{\eta^i}^i$ are of class C^1, one can take

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

and the adjoint differential inclusion becomes the classical adjoint equation. On the other hand, if the function $x \to \tilde{F}_{\eta^i}^i(x, t)$ 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 $x \to \tilde{F}_{\eta^i}^i(x, t)$ at $\xi^i(t)$ has an obvious integral uniformity property with respect to t. So Theorem 5 is in fact stronger than the classical versions, even in the setting of single-valued differentials.

In addition, when the $\tilde{F}_{\eta^i}^i$ 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 $\partial \tilde{F}_{\eta^i, t}^i(\xi^i(t))$, where $\tilde{F}_{\eta^i, t}^i$ is the map $x \to \tilde{F}_{\eta^i}^i(x, t)$, 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 5 often yields a stronger conclusion than the usual nonsmooth results.

Theorem 5 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 Proof of Theorem 5

It is clear that we can assume, without loss of generality, that

$$\varphi^i(x_+^i, a_+^i, x_-^{i\tilde{+}1}, a_-^{i\tilde{+}1}) = 0 \quad \text{for} \quad i = 1, \ldots, \mu\,. \tag{8}$$

We make this assumption throughout our proof.

For each $i \in \{1, \ldots, \mu\}$, we let $\mathcal{X}^i$ denote the space of all continuous fields of covectors along ξ^i, so the members of $\mathcal{X}^i$ are the maps

$$[a_-^i, a_+^i] \ni t \to \psi(t) \in T^*_{\xi^i(t)} Q^i$$

such that ψ is continuous as a map from $[a_-^i, a_+^i]$ to T^*Q^i. Then $\mathcal{X}^i$ is a Banach space.

If $i \in \{1, \ldots, \mu\}$, we use $\mathcal{V}^i$ to denote the set of all pairs (ζ, t) such that $a_-^i < t < a_+^i$, $\zeta \in \mathcal{U}^i$ and $(\xi(t), t)$ is a point of approximate continuity of $\tilde{F}^i_{\eta^i}$ and $\tilde{F}^i_\zeta$. We then write $\mathcal{V}$ to denote the set of all triples (i, ζ, t) such that $i \in \{1, \ldots, m\}$ and $(\zeta, t) \in \mathcal{V}^i$.

If $\mathcal{W}$ is a subset of $\mathcal{V}$, define $\Psi_{\mathcal{W}}$ to be the set of all multipliers

$$(\boldsymbol{\psi}, \psi_0, \boldsymbol{\kappa}) = (\psi^1, \ldots, \psi^\mu, \psi_0, \kappa_-^1, \kappa_+^1, \ldots, \kappa_-^\mu, \kappa_+^\mu)$$

along $(\boldsymbol{\xi}, \boldsymbol{\eta})$ such that

$$\psi_0 + \sum_{i=1}^{\mu} (\|\psi^i(a_+^i)\| + |\kappa_-^i| + |\kappa_+^i|) = 1 \tag{9}$$

and

(&) *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 $(i, \zeta, t) \in \mathcal{W}$.

Then $\Psi_{\mathcal{W}}$ is a compact subset of the product space

$$\mathcal{X} \stackrel{\text{def}}{=} \mathcal{X}^1 \times \mathcal{X}^2 \times \cdots \times \mathcal{X}^\mu \times \mathbb{R}^{2\mu+1}\,.$$

Our goal is to prove that $\Psi_{\mathcal{V}} \neq \emptyset$. It is clear that, if $\mathcal{W}_1, \ldots, \mathcal{W}_k$ are subsets of $\mathcal{V}$, then

$$\Psi_{\mathcal{W}_1 \cup \ldots \cup \mathcal{W}_k} = \Psi_{\mathcal{W}_1} \cap \ldots \Psi_{\mathcal{W}_k}\,.$$

Therefore, if we prove that

(*) *$\Psi_{\mathcal{W}} \neq \emptyset$ whenever $\mathcal{W}$ is finite,*

then we will have shown that

$$\{\Psi_{\mathcal{W}}\}_{\mathcal{W} \subseteq \mathcal{V},\, \mathcal{W} \text{ finite}}$$

is a family of nonempty compact subsets of $\mathcal{X}$ that has the finite intersection property. Since

$$\Psi_{\mathcal{V}} = \bigcap \Big\{\Psi_{\mathcal{W}} : \mathcal{W} \subseteq \mathcal{V},\ \mathcal{W} \text{ finite}\Big\}$$

it will follow that $\Psi_{\mathcal{V}} \neq \emptyset$, proving our conclusion.

So it suffices to prove (*). For this purpose, we fix a finite subset $\mathcal{W}$ of $\mathcal{V}$, and write $\mathcal{W} = \cup_{i=1}^{\mu} \hat{\mathcal{W}}^i$, where $\hat{\mathcal{W}}^i \subseteq \{i\} \times \mathcal{V}^i$. Write $\hat{\mathcal{W}}^i = \{i\} \times \mathcal{W}^i$, so $\mathcal{W}^i \subseteq \mathcal{V}^i$.

We introduce the *cost-augmented state spaces*

$$Q^i_{\mathrm{c}} \stackrel{\text{def}}{=} Q^i \times \mathbb{R}\,,$$

together with the *cost-augmented time-varying vector fields*

$$F^i_{\zeta,\mathrm{c}} \in TVVF(Q^i_{\mathrm{c}}),$$

defined by

$$\mathrm{Do}(\hat{F}^i_{\zeta,\mathrm{c}}) = \{\, (q,r,t) \in Q^i_{\mathrm{c}} : (q,t) \in \mathrm{Do}(F^i_\zeta) \cap \mathrm{Do}(L^i_\zeta)\,\},$$

$$F^i_{\zeta,\mathrm{c}}(q,r,t) = \begin{bmatrix} F^i_\zeta(q,t) \\ L^i_\zeta(q,t) \end{bmatrix} \ \text{if}\ (q,r,t) \in \mathrm{Do}(F^i_\zeta).$$

(Here we are using the canonical identification of $T_{(q,r)}(Q^i \times \mathbb{R})$ with $T_q Q^i \times \mathbb{R}$ and writing the members of $T_q Q^i \times \mathbb{R}$ as column pairs. The above formula defines $F^i_{\zeta,\mathrm{c}}(q,r,t)$ as a member of $T_q Q^i \times \mathbb{R}$, so $F^i_{\zeta,\mathrm{c}}(q,r,t)$ belongs to $T_{(q,r)}(Q^i \times \mathbb{R})$. Therefore $F^i_{\zeta,\mathrm{c}}$ is indeed a time-varying vector field on Q^i_{c}.)

Then an integral curve of $F^i_{\zeta,\mathrm{c}}$ is a locally absolutely continuous curve

$$I \in t \to \xi_{\mathrm{c}}(t) = (\xi(t), \lambda(t)) \in Q^i_{\mathrm{c}},$$

defined on an interval I, such that

(a) ξ is an integral curve of F^i_ζ,
(b) the function $I \in t \to L^i_\zeta(\xi(t), t)$ is a.e. defined and locally integrable,

and

(c) $\lambda(t) = \lambda(s) + \int_s^t L^i_\zeta(\xi(u), u)\,du$ for all $s, t \in I$.

(In other words, ξ_c consists of an integral curve ξ of F^i_ζ together with a "running cost" function λ along ξ.)

We also introduce the *cost-augmented variational generators* Λ^i_c, defined by

$$\Lambda^i_c(t) = \tilde{\Lambda}^i(t) \times \{0\}\,.$$

Precisely:

- If Q^i is $\mathbb{R}^n$ or an open subset of $\mathbb{R}^n$, so that Q^i_c is an open subset of $\mathbb{R}^n \times \mathbb{R} = \mathbb{R}^{n+1}$, then $\tilde{\Lambda}^i(t)$ is a subset of $\mathbb{R}^{(n+1)\times n}$, whose members are $(n+1) \times n$ block matrices

$$\mathbf{L} = \begin{bmatrix} L \\ \ell \end{bmatrix}, \;\; L \in \mathbb{R}^{n\times n}, \;\; \ell \in \mathbb{R}^{1\times n},$$

so the adjoint differential inclusion is equivalent to the assertion that

$$-\dot{\psi}(t) = [\psi(t), -\psi_0] \cdot \begin{bmatrix} L(t) \\ \ell(t) \end{bmatrix} \qquad \text{a.e.}$$

for some measurable selection

$$t \to \mathbf{L}(t) = \begin{bmatrix} L(t) \\ \ell(t) \end{bmatrix}$$

of $\tilde{\Lambda}^i$. In that case, the set $\Lambda^i_c(t)$ is a subset of $\mathbb{R}^{(n+1)\times(n+1)}$, whose members are the square $(n+1) \times (n+1)$ block matrices

$$\mathbf{L}_c = \begin{bmatrix} L & 0 \\ \ell & 0 \end{bmatrix}$$

such that $\mathbf{L}_c = \begin{bmatrix} L \\ \ell \end{bmatrix} \in \tilde{\Lambda}^i(t)$. The adjoint differential inclusion, together with the statement that ψ_0 is constant, is equivalent to the assertion that

$$-\dot{\psi}_c(t) = \psi_c(t) \cdot \mathbf{L}_c(t)$$

for some measurable selection

$$t \to \mathbf{L}_c(t) = \begin{bmatrix} L(t) & 0 \\ \ell(t) & 0 \end{bmatrix}.$$

of $\tilde{\Lambda}^i$, where

$$\psi_c(t) = [\psi(t), -\psi_0]\,.$$

- If Q^i is a manifold, then the above description of the nature of $\tilde{\Lambda}^i$ and $\Lambda^i_{\rm c}$ and their relation to the adjoint equation remains true locally, in coordinates, and can be made valid globally, in an intrinsic way, as explained in Remark 1.

We now let $\xi^i_{\rm c}$ be the cost-augmented version of ξ^i, obtained by initializing the running cost to the value 0 at time a^i_-. That is, $\xi^i_{\rm c} : [a^i_-, a^i_+] \to Q^i_{\rm c}$ is the curve given by

$$\xi^i_{\rm c}(t) = \begin{bmatrix} \xi^i(t) \\ \lambda^i(t) \end{bmatrix},$$

where

$$\lambda^i(t) = \int_{a^i_-}^{t} L^i_{\eta^i}(\xi^i(s), s)\, ds\,.$$

If $\nu \in \mathbb{N}$, we use $\mathbb{R}_{+,\nu}$ to denote the nonnegative orthant of $\mathbb{R}_\nu$, that is, the set of all row vectors $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_\nu) \in \mathbb{R}_\nu$ such that $\varepsilon_j \geq 0$ for $j = 1, \ldots, \nu$. For $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_\nu) \in \mathbb{R}_\mu$, we write

$$|\boldsymbol{\varepsilon}| \stackrel{\rm def}{=} |\varepsilon_1| + \cdots + |\varepsilon_\nu|\,.$$

In particular, if $\boldsymbol{\varepsilon} \in \mathbb{R}_{+,\nu}$, then $|\boldsymbol{\varepsilon}| = \varepsilon_1 + \cdots + \varepsilon_\nu$.

If $r > 0$, we use $\mathcal{P}_\nu(r)$ to denote the ν-dimensional simplex

$$\mathcal{P}_\nu(r) \stackrel{\rm def}{=} \{\, \boldsymbol{\varepsilon} \in \mathbb{R}_{+,\nu} : |\boldsymbol{\varepsilon}| \leq r \,\}\,.$$

For each i, we let ν^i be the cardinality of $\mathcal{W}^i$. We choose once and for all an ordered ν^i-tuple

$$\mathbf{W}^i = (\,(\zeta^i_1, t^i_1)\,, (\zeta^i_2, t^i_2)\,, \ldots\,, (\zeta^i_{\nu^i}, t^i_{\nu^i})\,) \tag{10}$$

such that the times t^i_j satisfy

$$t^i_1 \leq t^i_2 \leq \ldots \leq t^i_{\nu^i}\,,$$

and $\mathcal{W}^i$ is the set $\{\,(\zeta^i_1, t^i_1)\,, (\zeta^i_2, t^i_2)\,, \ldots\,, (\zeta^i_{\nu^i}, t^i_{\nu^i})\,\}$. (The ordered ν^i-tuple $\mathbf{W}^i$ is of course uniquely determined by the set $\mathcal{W}^i$ in the special case when $\mathcal{W}^i$ has no "repeated times"—i.e., if $(\zeta, t) \in \mathcal{W}^i$, $(\zeta', t) \in \mathcal{W}^i$ implies $\zeta = \zeta'$.) Also, we write $t^i_0 = a^i_-$, $t^i_{\nu^i+1} = a^i_+$.

We let $\hat{r}$ be the minimum of all the nonzero members of the set

$$\Big\{\, t^i_{j+1} - t^i_j : j = 0, \ldots, \nu^i\,,\ i = 1, \ldots, \mu \Big\}\,.$$

We then define, for each i, affine functions

$$\mathbb{R}_{\nu^i}(\hat{r}) \ni \boldsymbol{\varepsilon}^i \;\longrightarrow\; \tau^i_j(\boldsymbol{\varepsilon}^i) \in \mathbb{R}$$

inductively for $j = 1, \ldots, \nu^i + 1$, and prove inductively that the inequality

$$t^i_j \leq \tau^i_j(\boldsymbol{\varepsilon}^i) \leq t^i_j + \varepsilon^i_1 + \ldots + \varepsilon^i_{j-1} \quad \text{if} \quad \boldsymbol{\varepsilon}^i = (\varepsilon^i_1, \ldots, \varepsilon^i_{\nu^i}) \in \mathcal{P}_{\nu^i}(\hat{r}) \tag{11}$$

holds if $j \in \{1, \ldots, \nu^i\}$.

The construction is as follows. First, we define $\tau_1^i(\boldsymbol{\varepsilon}^i) = t_1^i$, so (11) is trivially true when $j = 1$. Next, assume that $\tau_j^i(\boldsymbol{\varepsilon}^i)$ has been defined for some $j \in \mathbb{N}$ such that $1 \leq j < \nu^i$, and (11) holds. If t_{j+1}^i is equal to t_j^i, then we let $\tau_{j+1}^i(\boldsymbol{\varepsilon}^i) = \tau_j^i(\boldsymbol{\varepsilon}^i) + \varepsilon_j^i$. If $t_{j+1}^i > t_j^i$, then we define $\tau_{j+1}^i(\boldsymbol{\varepsilon}^i) = t_{j+1}^i$. It is then clear that, in both cases, $\boldsymbol{\varepsilon}^i \rightarrow \tau_{j+1}^i(\boldsymbol{\varepsilon}^i)$ is an affine function, and (11) holds if j is replaced by $j+1$. We complete the definition by letting $\tau_{\nu^i+1}^i(\boldsymbol{\varepsilon}^i) = a_+^i$. It is clear that $\tau_j^i(0) = t_j^i$ for all i, j.

It follows from the construction that the inequalities

$$\tau_j^i(\boldsymbol{\varepsilon}^i) + \varepsilon_j \leq \tau_{j+1}^i(\boldsymbol{\varepsilon}^i) \tag{12}$$

hold for $j = 1, \ldots, \nu^i$ and $\boldsymbol{\varepsilon}^i \in \mathcal{P}_{\nu^i}(\hat{r})$. Indeed, (12) follows clearly from the definition of $\tau_{j+1}^i(\boldsymbol{\varepsilon}^i)$ if $t_{j+1}^i = t_j^i$. If $t_{j+1}^i > t_j^i$, then (12) follows because (11) implies that

$$\begin{aligned}
\tau_j^i(\boldsymbol{\varepsilon}^i) + \varepsilon_j^i &\leq t_j^i + \varepsilon_1^i + \ldots + \varepsilon_{j-1}^i + \varepsilon_j^i \\
&\leq t_j^i + |\boldsymbol{\varepsilon}^i| \\
&\leq t_j^i + \hat{r} \leq t_{j+1}^i = \tau_{j+1}^i(\boldsymbol{\varepsilon}^i)\,.
\end{aligned}$$

The inequality (12) implies that, if we write

$$I_j^i(\boldsymbol{\varepsilon}^i) \stackrel{\text{def}}{=} [\,\tau_j^i(\boldsymbol{\varepsilon}^i)\,,\ \tau_j^i(\boldsymbol{\varepsilon}^i) + \varepsilon_j^i\,[\ ,$$

then *for each* i, $\{I_j^i(\boldsymbol{\varepsilon}^i)\}_{j\in\{1,\ldots,\nu^i\}}$ *is a familiy of pairwise disjoint subintervals of* $[a_-^i, a_+^i]$, *such that* $I_j^i(\boldsymbol{\varepsilon}^i)$ *has length* ε_j^i *and* $I_j^i(\boldsymbol{\varepsilon}^i) \subseteq [t_j^i, t_j^i + \hat{r}]$ *for* $j = 1, \ldots, \nu^i$. We let

$$\mathcal{I}^i(\boldsymbol{\varepsilon}^i) \stackrel{\text{def}}{=} \bigcup_{j=1}^{\nu^i} I_j(\boldsymbol{\varepsilon}^i)\,,$$

so $\mathcal{I}^i(\boldsymbol{\varepsilon}^i)$ has measure $|\boldsymbol{\varepsilon}^i|$. Write

$$\mathcal{Q}_*^i = Q^i \times R_+^i \times R_-^i \times \mathbb{R}_{\nu^i}\,.$$

Fix i, and define set-valued maps

$$\Theta_j^i : \mathcal{Q}_*^i \longrightarrow\!\!\!\!\rightarrow Q_{\mathrm{c}}^i$$

inductively, for $j = 1, \ldots, \nu^i + 1$, as follows. First of all, we let

$$\Theta_1^i(z, \alpha_+, \alpha_-, \boldsymbol{\varepsilon}) = \Phi^{F^i_{\eta^i,\mathrm{c}}}_{t_i^1, a_-^i + \alpha_-}(z, 0)\,.$$

We then define

$$\Theta_{j+1}^i(z, \alpha_+, \alpha_-, \boldsymbol{\varepsilon}) = \left(\Phi^{F^i_{\eta^i,\mathrm{c}}}_{\tau_{j+1}^i(\boldsymbol{\varepsilon}), \tau_j^i(\boldsymbol{\varepsilon}) + \varepsilon_j} \circ \Phi^{F^i_{\varsigma_j^i,\mathrm{c}}}_{\tau_j^i(\boldsymbol{\varepsilon}) + \varepsilon_j, \tau_j^i(\boldsymbol{\varepsilon})}\right)\big(\Theta_j^i(z, \alpha_+, \alpha_-, \boldsymbol{\varepsilon})\big)$$

for $j = 1, \ldots, \nu^i - 1$, and

$$\begin{aligned}
&\Theta_{\nu^i+1}^i(z, \alpha_+, \alpha_-, \boldsymbol{\varepsilon}) = \\
&\qquad \left(\Phi^{F^i_{\eta^i,\mathrm{c}}}_{a_+^i + \alpha_+, \tau_{\nu^i}^i(\boldsymbol{\varepsilon}) + \varepsilon_{\nu^i}} \circ \Phi^{F^i_{\varsigma_{\nu^i}^i,\mathrm{c}}}_{\tau_{\nu^i}^i(\boldsymbol{\varepsilon}) + \varepsilon_{\nu^i}, \tau_{\nu^i}^i(\boldsymbol{\varepsilon})}\right)\big(\Theta_{\nu^i}^i(z, \alpha_+, \alpha_-, \boldsymbol{\varepsilon})\big)\,.
\end{aligned}$$

We let

$$\Theta^i_{\#} \stackrel{\text{def}}{=} \Theta^i_{\nu^i+1} \,.$$

(In other words: we make a "packet of needle variations of η^i," by substituting the controller ζ^i_j for η^i on the interval $[\tau^i_j(\boldsymbol{\varepsilon}), \tau^i_j(\boldsymbol{\varepsilon}) + \varepsilon_j]$ for $j = 1, \ldots, \nu^i$; then, using the new control—which depends on $\boldsymbol{\varepsilon}$ as a parameter—we move in the cost-augmented state space Q^i_{c} by initializing the state component at z and the running cost component at 0 at time $a^i_- + \alpha_-$, and then following integral curves of the new dynamics up to time $a^i_+ + \alpha_+$, thus obtaining, for each value of z, α_+, α_- and $\boldsymbol{\varepsilon}$, a not necessarily unique point in Q^i_{c}; then $\Theta^i_{\#}(z, \alpha_+, \alpha_-, \boldsymbol{\varepsilon})$ is the set of all points that can be obtained in this way. The fact that the endtimes are $a^i_+ + \alpha_+$ and $a^i_- + \alpha_-$ rather than a^i_+, a^i_- means that we are also making "variations of the endtimes;" the fact that the initial condition is z rather than x^i_- means that we are making "variations of the initial state" as well.)

Write

$$\begin{aligned}
X^i_- &= \xi^i_{\mathrm{c}}(a^i_-) = (x^i_-, 0)\,,\\
X^{i,1}_- &= (X^i_-, a^i_-)\,,\\
Y^i_- &= F^i_{\eta^i,\mathrm{c}}(X^{i,1}_-)\,,\\
X^i_+ &= \xi^i_{\mathrm{c}}(a^i_+) = (\xi^i(a^i_+), \lambda^i(a^i_+))\,,\\
X^{i,1}_+ &= (X^i_+, a^i_+)\,,\\
Y^i_+ &= F^i_{\eta^i,\mathrm{c}}(X^{i,1}_+)\,,\\
x^i_j &= \xi^i(t^i_j)\,,\\
X^i_j &= \xi^i_{\mathrm{c}}(t^i_j) = (\xi^i(t^i_j), \lambda^i(t^i_j))\,,\\
X^{i,1}_j &= (X^i_j, t^i_j)\,,\\
Y^i_j &= F^i_{\eta^i,\mathrm{c}}(X^{i,1}_j)\,,\\
\hat{Y}^i_j &= F^i_{\zeta^i_j,\mathrm{c}}(X^{i,1}_j)\,,\\
Z^i_j &= \hat{Y}^i_j - Y^i_j\,,\\
X^{i,2}_j &= (X^i_j, t^i_{j+1}, t^i_j)\,,\\
X^{i,3}_j &= (X^i_j, t^i_j, t^i_j, t^i_{j+1})\,.
\end{aligned}$$

It is then clear that $\Theta^i_{\#}(x^i_-, 0, 0, 0) = \{X^i_+\}$ and $\Theta^i_j(x^i_-, 0, 0, 0) = \{X^i_j\}$ for $j = 1, \ldots, \nu^i$. We now let

$$K^i = Q^i \times C^i_+ \times C^i_- \times \mathcal{P}_{\nu^i}(\hat{r})\,, \tag{13}$$

and compute GDQs

$$D^i_j \in GDQ(\Theta^i_j; (x^i_-, 0, 0, 0), X^i_j; K^i) \tag{14}$$

inductively, by applying Theorem 3 and the chain rule.

For $j = 1, \ldots, \nu^i$, we let D^i_j be the set of all linear maps Δ^i_{j,L_c}, for all measurable selections L_c of Λ^i_c, where

$$\Delta^i_{j,L_c}(v, \alpha_+, \alpha_-, \boldsymbol{\varepsilon}) \stackrel{\text{def}}{=} M^i_{j,-} \cdot \left(\tilde{v} - \alpha_- Y^i_-\right) + \sum_{k=1}^{j-1} \varepsilon_k M^i_{j,k} \cdot Z^i_k + (\tau^i_j(\boldsymbol{\varepsilon}_j) - t^i_j) Y^i_j\,,$$

$$M^i_{j,-} \stackrel{\text{def}}{=} M_{L_c}(t^i_j, a^i_-)\,,$$

$$M^i_{j,k} \stackrel{\text{def}}{=} M_{L_c}(t^i_j, t^i_k)\,, \qquad \text{and}$$

$$\tilde{v} \stackrel{\text{def}}{=} \begin{bmatrix} v \\ 0 \end{bmatrix},$$

and prove by induction on j that (14) holds for every $j \in \{1, \ldots, \nu^i\}$.

First of all, let $A^i_1 : Q^i \times \mathbb{R} \times \mathbb{R} \times \mathbb{R}_{\nu^i} \to Q^i_c \times \mathbb{R} \times \mathbb{R}$ be the map

$$(z, \alpha_+, \alpha_-, \boldsymbol{\varepsilon}) \to ((z, 0), t^1, a^i_- + \alpha_-).$$

Then

$$\Theta^i_1 = \Phi^{F^i_{\eta^i,c}} \circ A^i_1\,.$$

The map A^i_1 is of class C^1. Therefore, if B^i_1 is the linear map

$$B^i_1 : T_{x^i_-} Q^i \times \mathbb{R} \times \mathbb{R} \times \mathbb{R}_{\nu^i} \to T_{X^i_-} Q^i_c \times \mathbb{R} \times \mathbb{R}$$

given by

$$B^i_1(v, \alpha_+, \alpha_-, \boldsymbol{\varepsilon}) = (\tilde{v}, 0, \alpha_-)\,,$$

then

$$\{B^i_1\} \in GDQ(A^i_1; (x^i_-, 0, 0, 0), (X^i_-, t^i_1, a^i_-); K^i)\,.$$

Let

$$K^i_1 = Q^i \times (t^i_1 + C^i) \times (a^i_- + C_-)\,.$$

Theorem 3 tells us that a member Δ^i_1 of $GDQ(\Phi^{F^i_{\eta^i,c}}; (X^i_-, t^i_1, a^i_-), X^i_1; K^i_1)$ is given by

$$\Delta^i_1 = [\mathcal{M}_{t^i_1, a^i_-}(\Lambda^i_c); \{F^i_{\eta^i,c}(X^{i,1}_1)\}, \{F^i_{\eta^i,c}(X^{i,1}_-)\}]\,.$$

Clearly, $A^i_1 K^i \subseteq K^i_1$, so the chain rule applies, and we can conclude that

$$D^i_1 = \Delta^i_1 \circ B^i_! \in GDQ(\Theta^i_1; (x^i_-, 0, 0, 0), X^i_1; K^i_1)\,.$$

Now assume that $j \in \{2, \ldots, \nu^i\}$, and we have shown that D^i_{j-1} belongs to $GDQ(\Theta^i_{j-1}; (x^i_-, 0, 0, 0), X^i_{j-1}; K^i)$. Let

$$A^i_j : Q^i \times \mathbb{R} \times \mathbb{R} \times \mathbb{R}_{\nu^i} \longrightarrow\!\!\!\!\!\rightarrow Q^i_c \times \mathbb{R} \times \mathbb{R} \times \mathbb{R}$$

be the set-valued map that sends $(z, \alpha_+, \alpha_-, \boldsymbol{\varepsilon})$ to the set

$$\Theta^i_{j-1}(z, \alpha_+, \alpha_-, \boldsymbol{\varepsilon}) \times \{\tau^i_{j-1}(\boldsymbol{\varepsilon}) + \varepsilon_{j-1}\} \times \{\tau^i_{j-1}(\boldsymbol{\varepsilon})\} \times \{\tau^i_j(\boldsymbol{\varepsilon})\}.$$

Let $\tilde{A}^i_j : Q^i_{\rm c} \times \mathbb{R} \times \mathbb{R} \times \mathbb{R} \longrightarrow\!\!\!\!\to Q^i_{\rm c} \times \mathbb{R} \times \mathbb{R}$ be the set-valued map that sends the point (Z, r_1, r_2, r_3) to the set

$$\Phi^{F^i_{\zeta^i,{\rm c}}}(Z, r_1, r_2) \times \{r_3\} \times \{r_1\}.$$

Then

$$\Theta^i_j = \Phi^{F^i_{\eta^i,{\rm c}}} \circ \tilde{A}^i_j \circ A^i_j \,.$$

It follows from the inductive hypothesis that, if $\mathbf{B}^i_j$ is the set of all linear maps

$$B^i_{j,L_{\rm c}} : T_{x^i_-} Q^i \times \mathbb{R} \times \mathbb{R} \times \mathbb{R}_{\nu^i} \to T_{X^i_{j-1}} Q^i_{\rm c} \times \mathbb{R} \times \mathbb{R} \times \mathbb{R} \,,$$

for all measurable selections $L_{\rm c}$ of $\Lambda^i_{\rm c}$, where

$$\begin{aligned} B^i_{j,L_{\rm c}}(v, \alpha_+, \alpha_-, \boldsymbol{\varepsilon}) &= \Big(\Delta^i_{j-1,L_{\rm c}}(v, \alpha_+, \alpha_-, \boldsymbol{\varepsilon}) \,,\, B^i_{j,2}(\boldsymbol{\varepsilon}) \,,\, B^i_{j,3}(\boldsymbol{\varepsilon}) \,,\, B^i_{j,4}(\boldsymbol{\varepsilon}) \Big) \,, \\ B^i_{j,2}(\boldsymbol{\varepsilon}) &= \tau^i_{j-1}(\boldsymbol{\varepsilon}) + \varepsilon_{j-1} - t^i_{j-1} \,, \\ B^i_{j,3}(\boldsymbol{\varepsilon}) &= \tau^i_{j-1}(\boldsymbol{\varepsilon}) - t^i_{j-1} \,, \\ B^i_{j,4}(\boldsymbol{\varepsilon}) &= \tau^i_j(\boldsymbol{\varepsilon}) - t^i_j \,, \end{aligned}$$

then

$$\mathbf{B}^i_j \in GDQ(A^i_j; (x^i_-, 0, 0, 0), X^{i,3}_{j-1}; K^i) \,.$$

Theorem 3 implies that, if $\tilde{B}^i_j$ is the linear map

$$\tilde{B}^i_j : T_{X^i_{j-1}} Q^i_{\rm c} \times \mathbb{R} \times \mathbb{R} \times \mathbb{R} \to T_{X^i_{j-1}} Q^i_{\rm c} \times \mathbb{R} \times \mathbb{R}$$

given by

$$\tilde{B}^i_j(V, \rho_1, \rho_2, \rho_3) = (V + (\rho_1 - \rho_2)\hat{Y}^i_{j-1}, \rho_3, \rho_1) \,,$$

then

$$\{\tilde{B}^i_j\} \in GDQ(\tilde{A}^i_j; X^{i,3}_{j-1}, X^{i,2}_{j-1}; Q^i_{\rm c} \times \mathbb{R}^3) \,.$$

Finally, Theorem 3 also implies that, if $\hat{B}^i_{j,L_{\rm c}}$ is the linear map

$$\hat{B}^i_{j,L_{\rm c}} : T_{X^i_{j-1}} Q^i_{\rm c} \times \mathbb{R} \times \mathbb{R} \to T_{X^i_j} Q_{\rm c}$$

given by

$$\hat{B}^i_{j,L_{\rm c}}(V, \sigma_1, \sigma_2) = \sigma_1 Y^j_j + M^i_{j,j-1} \cdot (V - \sigma_2 Y^i_{j-1}) \,,$$

and $\hat{\mathbf{B}}^i_j$ is the set of all $\hat{B}_{j,L_{\rm c}}$ for all measurable selections $L_{\rm c}$ of $\Lambda^i_{\rm c}$, then

$$\hat{\mathbf{B}}^i_j \in GDQ(\Phi^{F^i_{\eta^i,{\rm c}}}; X^{i,2}_{j-1}, X^i_j; Q^i_{\rm c} \times \mathbb{R}^2) \,.$$

Then the chain rule implies that

$$\hat{\mathbf{B}}^i_j \circ \tilde{B}^i_j \circ \mathbf{B}^i_j \in GDQ(\Theta^i_j; (x^i_-, 0, 0, 0), X^{i,3}_j; K^i) \,.$$

Clearly, $\hat{\mathbf{B}}^i_j \circ \tilde{B}^i_j \circ \mathbf{B}^i_j$ is the set of all maps $\hat{B}^i_{j,L_c} \circ \tilde{B}^i_j \circ B^i_{j,L_c}$, for all $L_c \in \Gamma(\Lambda^i_c)$. Given a point $(v, \alpha_+, \alpha_-, \boldsymbol{\varepsilon}) \in T_{X^i_-} Q^i_c \times \mathbb{R} \times \mathbb{R} \times \mathbb{R}_{\nu^i}$, we have

$$B^i_{j,L_c}(v, \alpha_+, \alpha_-, \boldsymbol{\varepsilon}) = (V, \rho_1, \rho_2, \rho_3)\,,$$

and

$$\tilde{B}^i_j(V, \rho_1, \rho_2, \rho_3) = (V + (\rho_1 - \rho_2)\hat{Y}^i_{j-1}, \rho_3, \rho_1)\,,$$

where

$$V = \Delta^i_{j-1,L_c}(v, \alpha_+, \alpha_-, \boldsymbol{\varepsilon})\,, \quad \rho_1 = B^i_{j,2}(\boldsymbol{\varepsilon})\,, \quad \rho_2 = B^i_{j,3}(\boldsymbol{\varepsilon}))\,, \quad \rho_3 = B^i_{j,4}(\boldsymbol{\varepsilon})\,.$$

It follows that

$$\begin{aligned}
(\hat{\mathbf{B}}^i_{j,L_c} \circ \tilde{B}^i_j \circ \mathbf{B}^i_{j,L_c})(v, \alpha_+, \alpha_-, \boldsymbol{\varepsilon}) &= \hat{\mathbf{B}}^i_{j,L_c}\Big(\tilde{B}^i_j(V, \rho_1, \rho_2, \rho_3)\Big) \\
&= \mathbf{B}^i_{j,L_c}\Big((V + (\rho_1 - \rho_2)\hat{Y}^i_{j-1}, \rho_3, \rho_1)\Big) \\
&= \rho_3 Y^i_j + M^i_{j,j-1} \cdot W\,,
\end{aligned}$$

where

$$W = V + (\rho_1 - \rho_2)\hat{Y}^i_{j-1} - \rho_1 Y^i_{j-1}\,.$$

Since $\rho_1 - \rho_2 = \varepsilon_{j-1}$ and $\rho_1 = \tau^i_{j-1}(\boldsymbol{\varepsilon}) - t^i_{j-1} + \varepsilon_{j-1}$, we find

$$\begin{aligned}
W &= V + \varepsilon_{j-1}\hat{Y}^i_{j-1} - \rho_1 Y^i_{j-1} \\
&= \Delta^i_{j-1,L_c}(v, \alpha_+, \alpha_-, \boldsymbol{\varepsilon}) + \varepsilon_{j-1}\hat{Y}^i_{j-1} - \rho_1 Y^i_{j-1} \\
&= M^i_{j-1,-} \cdot \Big(\tilde{v} - \alpha_- Y^i_-\Big) + \sum_{k=1}^{j-2} \varepsilon_k M^i_{j-1,k} \cdot Z^i_k \\
&\quad + (\tau^i_{j-1}(\boldsymbol{\varepsilon}) - t^i_{j-1})Y^i_{j-1} + \varepsilon_{j-1}\hat{Y}^i_{j-1} - \rho_1 Y^i_{j-1} \\
&= M^i_{j-1,-} \cdot \Big(\tilde{v} - \alpha_- Y^i_-\Big) + \sum_{k=1}^{j-2} \varepsilon_k M^i_{j-1,k} \cdot Z^i_k \\
&\quad + (\tau^i_{j-1}(\boldsymbol{\varepsilon}) - t^i_{j-1} + \varepsilon_{j-1})Y^i_{j-1} + \varepsilon_{j-1}(\hat{Y}^i_{j-1} - Y^i_{j-1}) - \rho_1 Y^i_{j-1} \\
&= M^i_{j-1,-} \cdot \Big(\tilde{v} - \alpha_- Y^i_-\Big) + \sum_{k=1}^{j-2} \varepsilon_k M^i_{j-1,k} \cdot Z^i_k \\
&\quad + (\tau^i_{j-1}(\boldsymbol{\varepsilon}) - t^i_{j-1} + \varepsilon_{j-1} - \rho_1)Y^=_{j-1} i + \varepsilon_{j-1} Z^i_{j-1} \\
&= M^i_{j-1,-} \cdot \Big(\tilde{v} - \alpha_- Y^i_-\Big) + \sum_{k=1}^{j-1} \varepsilon_k M^i_{j-1,k} \cdot Z^i_k\,.
\end{aligned}$$

Therefore

$$M^i_{j,j-1} \cdot W = M^i_{j,-} \cdot \Big(\tilde{v} - \alpha_- Y^i_-\Big) + \sum_{k=1}^{j-1} \varepsilon_k M^i_{j,k} \cdot Z^i_k\,.$$

Then

$$\begin{aligned}(\hat{\mathbf{B}}^i_{j,L_c} \circ \tilde{B}^i_j \circ \mathbf{B}^i_{j,L_c})(v, \alpha_+, \alpha_-, \boldsymbol{\varepsilon}) &= \rho_3 Y^i_j + M^i_{j,j-1} \cdot W \\ &= M^i_{j,-} \cdot \left(\tilde{v} - \alpha_- Y^i_-\right) + \sum_{k=1}^{j-1} \varepsilon_k M^i_{j,k} \cdot Z^i_k + (\tau^i_j(\boldsymbol{\varepsilon}) - t^i_j) Y^i_j \\ &= \Delta^i_{j,L_c}(v, \alpha_+, \alpha_-, \boldsymbol{\varepsilon}) .\end{aligned}$$

It follows that

$$\hat{\mathbf{B}}^i_{j,L_c} \circ \tilde{B}^i_j \circ \mathbf{B}^i_{j,L_c} = D^i_j ,$$

so (14) holds.

Now that we have proved that (14) holds for all indices $j \in \{1, \dots, \nu^i\}$, we know in particular that

$$D^i_{\nu^i} \in GDQ(\Theta^i_{\nu^i}; (x^i_-, 0, 0, 0), X^i_{\nu^i}; K^i) . \tag{15}$$

We let $D^i_\#$ be the set of all linear maps $\Delta^i_{\#,L_c}$, for all measurable selections L_c of Λ^i_c, where

$$\Delta^i_{\#,L_c}(v, \alpha_+, \alpha_-, \boldsymbol{\varepsilon}) \stackrel{\text{def}}{=} \alpha_+ Y^i_+ + M^i_{+,-} \cdot \left(\tilde{v} - \alpha_- Y^i_-\right) + \sum_{k=1}^{\nu^i} \varepsilon_k M^i_{+,k} \cdot Z^i_k , \tag{16}$$

and we let

$$M^i_{+,-} \stackrel{\text{def}}{=} M_{L_c}(a^i_+, a^i_-) \quad \text{and} \quad M^i_{+,k} \stackrel{\text{def}}{=} M_{L_c}(a^i_+, t^i_k) .$$

We will prove that

$$D^i_\# \in GDQ(\Theta^i_\#; (x^i_-, 0, 0, 0), X^i_+; K^i) . \tag{17}$$

Let $A^i_\# : Q^i \times \mathbb{R} \times \mathbb{R} \times \mathbb{R}_{\nu^i} \longrightarrow\!\!\!\!\!\rightarrow Q^i_c \times \mathbb{R} \times \mathbb{R} \times \mathbb{R}$ be the set-valued map that sends $(z, \alpha_+, \alpha_-, \boldsymbol{\varepsilon})$ to the set

$$\Theta^i_{\nu^i}(z, \alpha_+, \alpha_-, \boldsymbol{\varepsilon}) \times \{\tau^i_{\nu^i}(\boldsymbol{\varepsilon}) + \varepsilon_{\nu^i}\} \times \{\tau^i_{\nu^i}(\boldsymbol{\varepsilon})\} \times \{a^i_+ + \alpha_+\}.$$

Let $\tilde{A}^i_\# : Q^i_c \times \mathbb{R} \times \mathbb{R} \times \mathbb{R} \longrightarrow\!\!\!\!\!\rightarrow Q^i_c \times \mathbb{R} \times \mathbb{R}$ be the set-valued map that sends (Z, r_1, r_2, r_3) to the set $\Phi^{F^i_{\varsigma^i,c}}(Z, r_1, r_2) \times \{r_3\} \times \{r_1\}$. Then

$$\Theta^i_\# = \Phi^{F^i_{\eta^i,c}} \circ \tilde{A}^i_\# \circ A^i_\# .$$

It follows from (15) that, if $\mathbf{B}^i_\#$ is the set of all linear maps

$$B^i_{\#,L_c} : T_{x^i_-} Q^i \times \mathbb{R} \times \mathbb{R} \times \mathbb{R}_{\nu^i} \to T_{X^i_{\nu^i}} Q^i_c \times \mathbb{R} \times \mathbb{R} \times \mathbb{R} ,$$

for all measurable selections $L_{\rm c}$ of $\Lambda_{\rm c}^i$, where

$$B^i_{\#,L_{\rm c}}(v,\alpha_+,\alpha_-,\boldsymbol{\varepsilon}) = \Big(\,\Delta^i_{\nu^i,L_{\rm c}}(v,\alpha_+,\alpha_-,\boldsymbol{\varepsilon})\,,\, B^i_{\#,2}(\boldsymbol{\varepsilon})\,,\, B^i_{\#,3}(\boldsymbol{\varepsilon})\,,\,\alpha_+\Big)\,,$$
$$B^i_{\#,2}(\boldsymbol{\varepsilon}) = \tau^i_{\nu^i}(\boldsymbol{\varepsilon}) + \varepsilon_{\nu^i} - t^i_{\nu^i}\,,$$
$$B^i_{\#,3}(\boldsymbol{\varepsilon}) = \tau^i_{\nu^i}(\boldsymbol{\varepsilon}) - t^i_{\nu^i}\,,$$

then

$$\mathbf{B}^i_{\#} \in GDQ(A^i_{\#};(x^i_-,0,0,0),X^{i,3}_{\nu^i};K^i)\,.$$

Theorem 3 then implies that, if $\tilde{B}^i_{\#}$ is the linear map

$$\tilde{B}^i_{\#}: T_{X^i_{\nu^i}}Q^i_{\rm c}\times\mathbb{R}\times\mathbb{R}\times\mathbb{R}\to T_{X^i_{\nu^i}}Q^i_{\rm c}\times\mathbb{R}\times\mathbb{R}$$

given by

$$\tilde{B}^i_{\#}(V,\rho_1,\rho_2,\rho_3) = (V+(\rho_1-\rho_2)\hat{Y}^i_{\nu^i},\rho_3,\rho_1)\,,$$

then

$$\{\tilde{B}^i_{\#}\}\in GDQ(\tilde{A}^i_{\#};X^{i,3}_{\nu^i},X^{i,2}_{\nu^i};Q^i_{\rm c}\times\mathbb{R}^3)\,.$$

Finally, Theorem 3 also implies that, if $\hat{B}^i_{\#,L_{\rm c}}$ is the linear map

$$\hat{B}^i_{\#,L_{\rm c}}: T_{X^i_{\nu^i}}Q^i_{\rm c}\times\mathbb{R}\times\mathbb{R}\to T_{x^i_+}Q_{\rm c}$$

given by

$$\hat{B}^i_{\#,L_{\rm c}}(V,\sigma_1,\sigma_2) = \sigma_1 Y^i_+ + M^i_{+,\nu^i}\cdot(V-\sigma_2 Y^i_{\nu^i})\,,$$

and $\hat{\mathbf{B}}^i_{\#}$ is the set of all $\hat{B}_{\#,L_{\rm c}}$ for all measurable selections $L_{\rm c}$ of $\Lambda^i_{\rm c}$, then

$$\hat{\mathbf{B}}^i_{\#}\in GDQ(\Phi^{F^i_{\eta^i,{\rm c}}};X^{i,2}_{\nu^i},X^i_+;Q^i_{\rm c}\times(a^i_++C^i_+)\times\mathbb{R})\,.$$

It is easy to verify that

$$(\tilde{A}^i_{\#}\circ A^i_{\#})(K^i)\subseteq Q^i_{\rm c}\times(a^i_++C^i_+)\times\mathbb{R})\,.$$

Therefore the chain rule implies that

$$\hat{\mathbf{B}}^i_{\#}\circ\tilde{B}^i_{\#}\circ\mathbf{B}^i_{\#}\in GDQ(\Theta^i_{\#};(x^i_-,0,0,0),X^{i,3}_{\#};K^i)\,.$$

Clearly, $\hat{\mathbf{B}}^i_{\#}\circ\tilde{B}^i_{\#}\circ\mathbf{B}^i_{\#}$ is the set of all maps

$$\hat{B}^i_{\#,L_{\rm c}}\circ\tilde{B}^i_{\#}\circ B^i_{\#,L_{\rm c}}\,,$$

for all $L_{\rm c}\in\Gamma(\Lambda^i_{\rm c})$.

Given a point $(v,\alpha_+,\alpha_-,\boldsymbol{\varepsilon})$ belonging to the product $T_{X^i_-}Q^i_{\rm c}\times\mathbb{R}\times\mathbb{R}\times\mathbb{R}_{\nu^i}$, we have

$$B^i_{\#,L_{\rm c}}(v,\alpha_+,\alpha_-,\boldsymbol{\varepsilon}) = (V,\rho_1,\rho_2,\rho_3)\,,$$

and

$$\tilde{B}^i_{\#}(V,\rho_1,\rho_2,\rho_3) = (V + (\rho_1-\rho_2)\hat{Y}^i_{\nu^i}, \rho_3, \rho_1)\,,$$

where

$$V = \Delta^i_{\nu^i,L_c}(v,\alpha_+,\alpha_-,\boldsymbol{\varepsilon})\,, \quad \rho_1 = B^i_{\#,2}(\boldsymbol{\varepsilon})\,, \quad \rho_2 = B^i_{\#,3}(\boldsymbol{\varepsilon}))\,, \quad \rho_3 = \alpha_+\,.$$

It follows that

$$\begin{aligned}(\hat{\mathbf{B}}^i_{\#,L_c} \circ \tilde{B}^i_{\#} \circ \mathbf{B}^i_{\#,L_c})(v,\alpha_+,\alpha_-,\boldsymbol{\varepsilon}) &= \hat{\mathbf{B}}^i_{\#,L_c}\Big(\tilde{B}^i_{\#}(V,\rho_1,\rho_2,\rho_3)\Big)\\ &= \mathbf{B}^i_{\#,L_c}\Big((V + (\rho_1-\rho_2)\hat{Y}^i_{\nu^i}, \rho_3, \rho_1)\Big)\\ &= \rho_3 Y^i_- + M^i_{+,\nu^i}\cdot W\,,\end{aligned}$$

where

$$W = V + (\rho_1-\rho_2)\hat{Y}^i_{\nu^i} - \rho_1 Y^i_{\nu^i}\,.$$

Since $\rho_1-\rho_2 = \varepsilon_{\nu^i}$ and $\rho_1 = \tau^i_{\nu^i}(\boldsymbol{\varepsilon}) - t^i_{\nu^i} + \varepsilon_{\nu^i}$, we find

$$\begin{aligned}W &= V + \varepsilon_{\nu^i}\hat{Y}^i_{\nu^i} - \rho_1 Y^i_{\nu^i}\\ &= \Delta^i_{\nu^i,L_c}(v,\alpha_+,\alpha_-,\boldsymbol{\varepsilon}) + \varepsilon_{\nu^i}\hat{Y}^i_{\nu^i} - \rho_1 Y^i_{\nu^i}\\ &= M^i_{\nu^i,-}\cdot\Big(\tilde{v}-\alpha_- Y^i_-\Big) + \sum_{k=1}^{\nu^i-1}\varepsilon_k M_{\nu^i,k}\cdot Z^i_k - \rho_1 Y^i_{\nu^i}\\ &\quad +(\tau^i_{\nu^i}(\boldsymbol{\varepsilon}) - t^i_{\nu^i})Y^i_{\nu^i} + \varepsilon_{\nu^i}\hat{Y}^i_{\nu^i}\\ &= M^i_{\nu^i,-}\cdot\Big(\tilde{v}-\alpha_- Y^i_-\Big) + \sum_{k=1}^{\nu^i-1}\varepsilon_k M^i_{\nu^i,k}\cdot Z^i_k - \rho_1 Y^i_{\nu^i}\\ &\quad +(\tau^i_{j-1}(\boldsymbol{\varepsilon}) - t^i_{j-1} + \varepsilon_j)Y^i_{j-1} + \varepsilon_{\nu^i}(\hat{Y}^i_{\nu^i} - Y^i_{\nu^i})\\ &= M^i_{\nu^i,-}\cdot\Big(\tilde{v}-\alpha_- Y^i_-\Big) + \sum_{k=1}^{\nu^i-1}\varepsilon_k M^i_{\nu^i,k}\cdot Z^i_k\\ &\quad +(\tau^i_{\nu^i}(\boldsymbol{\varepsilon}) - t^i_{\nu^i} + \varepsilon_{\nu^i} - \rho_1)Y^i_{\nu^i} + \varepsilon_{\nu^i} Z^i_k\\ &= M^i_{\nu^i,-}\cdot\Big(\tilde{v}-\alpha_- Y^i_-\Big) + \sum_{k=1}^{\nu^i}\varepsilon_k M^i_{\nu^i,k}\cdot Z^i_k.\end{aligned}$$

Therefore

$$\begin{aligned}M^i_{+,\nu^i}\cdot W &= M^i_{+,\nu^i}\cdot\Big(M^i_{\nu^i,-}\cdot(\tilde{v}-\alpha_- Y^i_-) + \sum_{k=1}^{\nu^i}\varepsilon_k M^i_{\nu^i,k}\cdot Z^i_k\Big)\\ &= M^i_{+,-}\cdot(\tilde{v}-\alpha_- Y^i_-) + \sum_{k=1}^{\nu^i}\varepsilon_k M^i_{+,k}\cdot Z^i_k\,.\end{aligned}$$

Then

$$
\begin{aligned}
(\hat{\mathbf{B}}^i_{\#,L_c} \circ \tilde{B}^i_{\#} \circ \mathbf{B}^i_{\#,L_c})(v,\alpha_+,\alpha_-,\boldsymbol{\varepsilon}) &= \rho_3 Y^i_+ + M^i_{+,\nu^i} \cdot W \\
&= \alpha_+ Y^i_+ + M^i_{+,-} \cdot (\tilde{v} - \alpha_- Y^i_-) + \sum_{k=1}^{\nu^i} \varepsilon_k M^i_{+,k} \cdot Z^i_k \\
&= \Delta^i_{\#,L_c}(v,\alpha_+,\alpha_-,\boldsymbol{\varepsilon}) \,.
\end{aligned}
$$

It follows that

$$
\hat{\mathbf{B}}^i_{\#,L_c} \circ \tilde{B}^i_{\#} \circ \mathbf{B}^i_{\#,L_c} = D^i_{\#} \,,
$$

so (17) holds.

We have thus shown that *for each $i \in \{1,\ldots,\mu\}$, the set $D^i_{\#}$ of all linear maps $\Delta^i_{\#,L_c}$ defined by (16), for all measurable selections L_c of Λ^i_c, is a GDQ of $\Theta^i_{\#}$ at $((x^i_-,0,0,0),x^i_+)$ along the set K^i defined by (13).*

We now combine all the $\Theta^i_{\#}$ into a single "grand map"

$$
\boldsymbol{\Theta} : \mathcal{Q}_* \longrightarrow\!\!\!\!\to \mathcal{Q}_{\#} \times \mathbb{R} \,,
$$

where

$$
\mathcal{Q}_* = \mathcal{Q}^1_* \times \ldots \times \mathcal{Q}^\mu_* \,, \quad \text{and} \quad \mathcal{Q}_{\#} = \mathcal{Q}^1_{\#} \times \ldots \times \mathcal{Q}^\mu_{\#} \,.
$$

Roughly speaking, if $p \in \mathcal{Q}_*$, then $p = (p^1,\ldots,p^\mu)$, where

$$
p^i = (z^i, \alpha^i_+, \alpha^i_-, \boldsymbol{\varepsilon}^i) \quad \text{for} \quad i = 1 \ldots, \mu \,. \tag{18}
$$

Then each p^i gives rise to one or several points $\Theta^i_{\#}(p^i)$. So to each p^i there correspond one or several terminal points w^i and terminal Lagrangian costs ℓ^i, as well as a terminal time $a^i_+ + \alpha^i_+$, an initial time $a^i_- + \alpha^i_-$, and an initial state z^i. In particular, this gives rise to "switching points" $\sigma^i \in \mathcal{Q}^i_{\#}$, defined by

$$
\sigma^i = (w^i, a^i_+ + \alpha^i_+, z^{i\tilde{+}1}, a^{i\tilde{+}1}_- + \alpha^{i\tilde{+}1}_-) \,. \tag{19}
$$

Moreover, p also gives rise to a cost σ_0, given by

$$
\sigma_0 = \sum_{i=1}^{\mu} \ell^i + \sum_{i=1}^{\mu} \varphi^i(w^i, a^i_+ + \alpha^i_+, z^{i\tilde{+}1}, a^{i\tilde{+}1}_- + \alpha^{i\tilde{+}1}_-) \,. \tag{20}
$$

We will define $\boldsymbol{\Theta}(p)$ to be the set of all $\mu + 1$-tuples

$$
(\sigma^1,\ldots,\sigma^\mu,\sigma_0) \in \mathcal{Q}_{\#} \times \mathbb{R} \tag{21}
$$

obtained from p in this way.

The precise definition is as follows. Let

$$
p = (p^1,\ldots,p^\mu) \in \mathcal{Q}_* = \mathcal{Q}^1_* \times \ldots \mathcal{Q}^\mu_* \,.
$$

Define z^i, α_+^i, α_-^i, $\boldsymbol{\varepsilon}^i$ by means of (18), so

$$z^i \in Q^i\,, \quad \alpha_+^i \in R_+\,, \quad \alpha_-^i \in R_-\,, \quad \boldsymbol{\varepsilon}^i \in \mathbb{R}_{\nu^i}\,.$$

Then $\boldsymbol{\Theta}(p)$ is the set of all $\mu+1$-tuples (21) such that, for some

$$w^1 \in Q^1, \ldots, w^\mu \in Q^\mu, \ell^1 \in \mathbb{R}, \ldots, \ell^\mu \in \mathbb{R},$$

the conditions (19) and $(w^i, \ell^i) \in \Theta^i_\#(p^i)$ hold whenever $i = 1, \ldots, \mu$, and (20) is satisfied.

This completes the definition of of $\boldsymbol{\Theta}$. Let

$$\bar{p} = (\bar{p}^1, \ldots, \bar{p}^\mu)\,,$$

where $\bar{p}^i = (x_-^i, 0, 0, 0)$ for $i = 1, \ldots, \mu$. Then

$$\boldsymbol{\Theta}(\bar{p}) = \{(\bar{\boldsymbol{\sigma}}, \bar{\sigma}_0)\}\,,$$

where

$$\begin{aligned}
\bar{\boldsymbol{\sigma}} &= (\bar{\sigma}^1, \ldots, \bar{\sigma}^\mu)\,, \\
\bar{\sigma}^i &= (x_+^i, a_+^i, x_-^{i\tilde{+}1}, a_-^{i\tilde{+}1}) \quad \text{for} \quad i = 1, \ldots, \mu\,, \\
\bar{\sigma}_0 &= \sum_{i=1}^{\mu} \lambda^i(a_+^i)\,.
\end{aligned}$$

(Recall that we are assuming that (8) holds.)

Let

$$\mathbf{K} = K^1 \times \ldots \times K^\mu\,.$$

We now write down a GDQ

$$\mathbf{D} \in GDQ(\boldsymbol{\Theta}; \bar{p}, (\bar{\boldsymbol{\sigma}}, \bar{\sigma}_0); \mathbf{K})\,. \tag{22}$$

For this purpose, we first define a linear map $\boldsymbol{\Delta}_{\mathbf{L}_c, \boldsymbol{\omega}}$, for each μ-tuple $\mathbf{L}_c = (L_c^1, \ldots, L_c^\mu)$ of measurable selections L_c^i of Λ_c^i and each μ-tuple $\boldsymbol{\omega} = (\omega^1, \ldots, \omega^\mu)$ such that $\omega^i \in \Omega^i$ for $i = 1, \ldots, \mu$.

We then let $\mathbf{D}$ denote the set of all maps $\boldsymbol{\Delta}_{\mathbf{L}_c, \boldsymbol{\omega}}$, for all possible pairs $(\mathbf{L}_c, \boldsymbol{\omega})$ of μ-tuples.

Let

$$\mathbf{u} = (u^1, , \ldots, u^\mu) \in T_{\bar{p}}\mathcal{Q}_*\,.$$

Write

$$u^i = (v^i, \alpha_+^i, \alpha_-^i, \boldsymbol{\varepsilon}^i)\,,$$

so $u^i \in T_{\bar{p}^i}\mathcal{Q}_*^i$ for each i. Let $\boldsymbol{\varepsilon}^i = (\varepsilon_1^i, \ldots, \varepsilon_{\nu^i}^i)$ for each i.

Let $\mathbf{L}_c = (L_c^1, \ldots, L_c^\mu)$ be a μ-tuple of measurable selections L_c^i of Λ_c^i. Write

$$L_c^i(t) = \begin{bmatrix} L^i(t) & 0 \\ \ell^i(t) & 0 \end{bmatrix}\,.$$

Then define

$$\boldsymbol{\Delta}_{\mathbf{L}_c,\boldsymbol{\omega}}(\mathbf{u}) = (s^1, \ldots, s^\mu, s_0)\,,$$

where

$$\begin{aligned} s^i &= (q^i, \alpha^i_+, v^{i\tilde{+}1}, \alpha_-^{i\tilde{+}1})\,, \\ \tilde{q}^i &= \begin{bmatrix} q^i \\ q^i_c \end{bmatrix}\,, \\ \tilde{q}^i &= \sum_{j=1}^{\nu^i} \varepsilon^i_j M^i_{+,j} \cdot Z^i_j + M^i_{+,-} \cdot (\tilde{v}^i - \alpha^i_- Y^i_-) + \alpha^i_+ Y^i_+\,, \\ s_0 &= \sum_{i=1}^{\mu} (q^i_c + \langle \omega^i, s^i \rangle)\,. \end{aligned}$$

Then it is easily verified that (22) holds.

Now let γ be a smooth real-valued function on the manifold $\mathcal{Q}_\#$ such that $\gamma(\bar{\boldsymbol{\sigma}}) = 0$ and $\gamma(\boldsymbol{\sigma}) > 0$ whenever $\boldsymbol{\sigma} \neq \bar{\boldsymbol{\sigma}}$.
Let

$$\mathcal{C}_\# = \mathcal{C}^1 \times \ldots \times \mathcal{C}^\mu \times]-\infty, 0]\,.$$

Define a subset $\tilde{\mathcal{S}}_{rest,\#}$ of $\mathcal{Q}_\# \times \mathbb{R}$ by letting

$$\begin{aligned} &\tilde{\mathcal{S}}_{rest,\#} \stackrel{\text{def}}{=} \\ &\quad \{(\boldsymbol{\sigma}, \sigma_0) \in \mathcal{Q}_\# \times \mathbb{R} : \boldsymbol{\sigma} \in \mathcal{S}_{rest,\#}\,,\ \sigma_0 + \gamma(\boldsymbol{\sigma}) \leq \bar{\sigma}_0\}\,, \end{aligned}$$

where

$$\mathcal{S}_{rest,\#} \stackrel{\text{def}}{=} \mathcal{S}^1_{rest} \times \ldots \times \mathcal{S}^\mu_{rest}$$

Then $\mathcal{C}_\#$ is a GDQ approximating multicone to $\tilde{\mathcal{S}}_{rest,\#}$ at the point $(\bar{\boldsymbol{\sigma}}, \bar{\sigma}_0)$.
Clearly, if $p \in \mathbf{K}$ is such that $(\boldsymbol{\sigma}, \sigma_0) \in \tilde{\mathcal{S}}_{rest,\#}$ for some $(\boldsymbol{\sigma}, \sigma_0) \in \boldsymbol{\Theta}(p)$, then $p, \boldsymbol{\sigma}$ give rise to an admissible trajectory-control pair $(\check{\boldsymbol{\xi}}, \check{\boldsymbol{\eta}})$ with cost σ_0. If $\boldsymbol{\sigma} \neq \bar{\boldsymbol{\sigma}}$, then $\sigma_0 < \bar{\sigma}_0$, contradicting the optimality of $(\boldsymbol{\xi}, \boldsymbol{\eta})$. So $\boldsymbol{\sigma} = \bar{\boldsymbol{\sigma}}$. Moreover, the fact that $(\boldsymbol{\sigma}, \sigma_0) \in \tilde{\mathcal{S}}_{rest,\#}$ also implies that $\sigma_0 \leq \bar{\sigma}_0$, so the optimality of $(\boldsymbol{\xi}, \boldsymbol{\eta})$ tells us that $\sigma_0 = \bar{\sigma}_0$. Hence

$$\boldsymbol{\Theta}(\mathbf{K}) \cap \tilde{\mathcal{S}}_{rest,\#} = \{(\bar{\boldsymbol{\sigma}}, \bar{\sigma}_0)\}\,.$$

It then follows from the transversal intersection theorem that the multicones $\mathbf{D}(\mathbf{K})$ and $\mathcal{C}_\#$ are not transversal. Therefore there exists a nontrivial linear functional $\bar{\Psi} \in T^*_{(\bar{\boldsymbol{\sigma}}, \bar{\sigma}_0)}(\mathcal{Q}_\# \times \mathbb{R})$ such that

1. there exist $C^1 \in \mathcal{C}^1, \ldots, C^\mu \in \mathcal{C}^\mu$ such that

   $$\left\langle \bar{\Psi}, (s^1, \ldots, s^\mu, r) \right\rangle \geq 0$$

   whenever $s^1 \in C^1, \ldots, s^\mu \in C^\mu$, $r \leq 0$;

2. there exist μ-tuples

$$\mathbf{L}_{\mathrm{c}} = (L^1_{\mathrm{c}}, \ldots, L^\mu_{\mathrm{c}}) \in \Gamma(\Lambda^1_{\mathrm{c}} \times \ldots \times \Gamma(\Lambda^\mu_{\mathrm{c}}),$$
$$\bar{\boldsymbol{\omega}} = (\bar{\omega}^1, \ldots, \bar{\omega}^\mu) \in \Omega^1 \times \ldots \times \Omega^\mu,$$

   such that
$$\Big\langle \bar{\Psi}, \boldsymbol{\Delta}_{\mathbf{L}_{\mathrm{c}},\bar{\boldsymbol{\omega}}}(\mathbf{u})\Big\rangle \leq 0$$
   for all $\mathbf{u} \in \mathbf{K}$.

Now write

$$\bar{\Psi} = (\bar{\Psi}^1, \ldots, \bar{\Psi}^\mu, -\psi_0), \tag{23}$$

where $\bar{\Psi}^i \in T_{\bar{\sigma}^i}\mathcal{Q}^i_{\#}$ for $i = 1, \ldots, \mu$, and $\psi_0 \in \mathbb{R}$.
Then

$$\boxed{\psi_0 \geq 0} \tag{24}$$

and, for each i,

$$\Big\langle \bar{\Psi}^i, s^i \Big\rangle \geq 0 \quad \text{whenever} \quad s^i \in C^i. \tag{25}$$

Since $\mathcal{Q}^i_{\#} = \mathcal{Q}^i_{+} \times \mathcal{Q}^{i\tilde{+}1}_{-}$, we can write

$$\bar{\Psi}^i = (\bar{\psi}^i_1, \pi^i_1, \bar{\psi}^i_2, \pi^i_2), \tag{26}$$
and (27)
$$\bar{\omega}^i = (\omega^i_1, \omega^i_{1,0}, \omega^i_2, \omega^i_{2,0}),$$

where

$$\bar{\psi}^i_1 \in T^*_{x^i_+} Q^i, \ \pi^i_1 \in R^i_+, \quad \bar{\psi}^i_2 \in T^*_{x^{i\tilde{+}1}_-} Q^{i\tilde{+}1}, \quad \pi^i_2 \in R^{i\tilde{+}1}_-,$$

and

$$\omega^i_1 \in T^*_{x^i_+} Q^i, \quad \omega^i_{1,0} \in R^i_+, \quad \omega^i_2 \in T^*_{x^{i\tilde{+}1}_-} Q^{i\tilde{+}1}, \quad \omega^i_{2,0} \in R^{i\tilde{+}1}_-.$$

Now pick
$$\mathbf{u} = (u^1, \ldots, u^\mu) \in \mathbf{K},$$
and write $u^i = (v^i, \alpha^i_+, \alpha^i_-, \boldsymbol{\varepsilon}^i)$.
Define the s^i, q^i, q^i_c, and s_0 as above. Then

$$\sum_{i=1}^{\mu} \langle \bar{\Psi}^i, s^i \rangle - \psi_0 s_0 \leq 0,$$

so

$$\sum_{i=1}^{\mu}\Big(\langle\bar{\psi}_1^i,q^i\rangle+\langle\bar{\psi}_2^i,v^{i\tilde{+}1}\rangle+\pi_1^i\alpha_+^i+\pi_2^i\alpha_-^{i\tilde{+}1}\Big)\le\psi_0 s_0\,. \tag{28}$$

Let

$$\tilde{\psi}_1^i=[\bar{\psi}_1^i,-\psi_0]\,,\quad \tilde{\psi}_2^i=[\bar{\psi}_2^i,0]\,,\quad \tilde{\omega}_1^i=[\omega_1^i,0]\,,\quad \tilde{\omega}_2^i=[\omega_2^i,0]\,,\quad \tilde{v}^i=\begin{bmatrix}v^i\\0\end{bmatrix}\,,$$

$$\check{\psi}_1^i=\tilde{\psi}_1^i-\psi_0\tilde{\omega}_1^i\,, \tag{29}$$
$$\check{\psi}_2^i=\tilde{\psi}_2^i-\psi_0\tilde{\omega}_2^i\,, \tag{30}$$

$$\check{\pi}_1^i=\pi_1^i-\psi_0\omega_{1,0}^i\,, \tag{31}$$
$$\check{\pi}_2^i=\pi_2^i-\psi_0\omega_{2,0}^i\,. \tag{32}$$

Then (28) says that

$$\sum_{i=1}^{\mu}\Big(\langle\bar{\psi}_1^i,q^i\rangle+\langle\bar{\psi}_2^i,v^{i\tilde{+}1}\rangle+\pi_1^i\alpha_+^i+\pi_2^i\alpha_-^{i\tilde{+}1}\Big)\le\psi_0\Big(\sum_{i=1}^{\mu}q_c^i+\sum_{i=1}^{\mu}\langle\bar{\omega}^i,s^i\rangle\Big)\,,$$

that is

$$\sum_{i=1}^{\mu}\Big(\Big(\langle\bar{\psi}_1^i,q^i\rangle-\psi_0 q_c^i\Big)+\langle\bar{\psi}_2^i,v^{i\tilde{+}1}\rangle+\pi_1^i\alpha_+^i+\pi_2^i\alpha_-^{i\tilde{+}1}\Big)\le\psi_0\sum_{i=1}^{\mu}\langle\omega_1^i,s^i\rangle\,,$$

which can be rewritten as

$$\begin{aligned}&\sum_{i=1}^{\mu}\Big(\langle\tilde{\psi}_1^i,\tilde{q}^i\rangle+\langle\tilde{\psi}_2^i,\tilde{v}^{i\tilde{+}1}\rangle+\pi_1^i\alpha_+^i+\pi_2^i\alpha_-^{i\tilde{+}1}\Big)\\&\quad\le\psi_0\sum_{i=1}^{\mu}\langle\omega_1^i,q^i\rangle+\langle\omega_2^i,v^i\rangle+\omega_{1,0}^i\alpha_+^i+\omega_{2,0}^i\alpha_-^{i\tilde{+}1}\,,\end{aligned}$$

i.e., as

$$\sum_{i=1}^{\mu}\Big(\langle\check{\psi}_1^i,\tilde{q}^i\rangle+\langle\check{\psi}_2^i,\tilde{v}^{i\tilde{+}1}\rangle+\check{\pi}_1^i\alpha_+^i+\check{\pi}_2^i\alpha_-^{i\tilde{+}1}\Big)\le 0\,. \tag{33}$$

We now extract information from (33) by making special choices of $\mathbf{u}$, i.e., of the v^i, α_+^i, α_-^i, $\boldsymbol{\varepsilon}^i$. First, we write

$$L_c(t)=\begin{bmatrix}L(t)&0\\\ell(t)&0\end{bmatrix}\,,$$

define

$$\hat{\psi}^i(t) = \check{\psi}^i_1 \cdot M_{L^i_{\rm c}}(a^i_+, t) \tag{34}$$

for $i = 1, \dots, \mu$, and observe that $\hat{\psi}^i$ is absolutely continuous and

$$\frac{d}{dt}(\hat{\psi}^i(t)) = -\hat{\psi}^i(t) \cdot L_{\rm c}(t) \ \text{ c.e.}\,.$$

Therefore, if we write

$$\hat{\psi}^i(t) = [\psi^i(t), \sigma^i(t)]\,, \tag{35}$$

then

$$\dot{\psi}^i(t) = -\psi^i(t) \cdot L^i(t) - \sigma^i(t) \cdot \ell(t)$$

and $\dot{\sigma}^i(t) \equiv 0$. Since $\hat{\psi}^i(a^i_+) = \check{\psi}^i_1 = [\psi^i(a^i_+), -\psi_0]$, we can conclude that

$$\sigma^i(t) \equiv -\psi_0\,, \tag{36}$$

so

$$\dot{\psi}^i(t) = -\psi^i(t) \cdot L^i(t) + \psi_0 \cdot \ell(t)\,,$$

that is

$$-\dot{\psi}^i(t) = [\psi^i(t), -\psi_0] \cdot \begin{bmatrix} L(t) \\ \ell(t) \end{bmatrix}\,.$$

Since $t \to \begin{bmatrix} L(t) \\ \ell(t) \end{bmatrix}$ is a measurable selection of $\tilde{\Lambda}^i$, we have shown that

$$\boxed{\psi^i(t) \quad \text{is absolutely continuous}} \tag{37}$$

and

$$\boxed{-\dot{\psi}^i(t) \in [\psi^i(t), -\psi_0] \cdot \tilde{\Lambda}^i(t) \quad \text{a.e.} \quad t} \tag{38}$$

i.e., that ψ^i is a solution of the adjoint differential inclusion.

Next, fix a value i_0 of i, let $i' = i_0 \tilde{+} 1$ choose all the α_-'s, α_+'s, and $\boldsymbol{\varepsilon}$'s equal to zero, and let $v^i = 0$ for $i \neq i'$, $v^{i'} = v \in T_{x^{i'}_-} Q^{i'}$. Then $\bar{q}^{i'} = M_{L^{i'}_{\rm c}}(a^{i'}_+, a^{i'}_-) \cdot \tilde{v}^{i'}$, and $\bar{q}^i = 0$ if $i \neq i'$. So (33) tells us that

$$\langle \check{\psi}^{i'}_1, \tilde{q}^{i'} \rangle + \langle \check{\psi}^{i_0}_2, \tilde{v}^{i'} \rangle \leq 0\,,$$

that is,

$$\langle \check{\psi}^{i'}_1 \cdot M_{L_{\rm c}}(a^{i'}_+, a^{i'}_-) + \check{\psi}^{i_0}_2, \tilde{v} \rangle \leq 0\,, \tag{39}$$

so that

$$\langle \hat{\psi}^{i'}(a_-^{i'}) + \check{\psi}_2^{i_0}, \tilde{v}\rangle \leq 0\,. \tag{40}$$

Since

$$\check{\psi}_2^{i_0} = [\bar{\psi}_2^{i_0} - \psi_0\omega_2^{i_0}\,,\,0]$$

(40) says that

$$\langle \psi^{i'}(a_-^{i'}) + \bar{\psi}_2^{i_0} - \psi_0\omega_2^{i_0}, v\rangle \leq 0\,. \tag{41}$$

Since v is an arbitrary vector in $T_{x_-^{i'}}Q^{i'}$, we have established that

$$\bar{\psi}_2^{i_0} = \psi_0\omega_2^{i_0} - \psi^{i'}(a_-^{i'})\,. \tag{42}$$

Since i_0 was an arbitrary index in the set $\{1, \ldots, \mu\}$, and $i' = i_0\tilde{+}1$, we have in fact shown that

$$\bar{\psi}_2^{i} = \psi_0\omega_2^{i} - \psi^{i\tilde{+}1}(a_-^{i\tilde{+}1}) \text{ if } i \in \{1, \ldots, \mu\}\,. \tag{43}$$

On the other hand, the fact that $\check{\psi}_1^i = [\psi^i(a_+^i), -\psi_0]$, and $\check{\psi}_1^i = \tilde{\psi}_1^i - \psi_0\tilde{\omega}_1^i$ imply

$$\bar{\psi}_1^{i} = \psi_0\omega_1^{i} + \psi^{i}(a_+^{i}) \text{ if } i \in \{1, \ldots, \mu\}\,. \tag{44}$$

Since $\check{\pi}_1^i = \pi_1^i - \psi_0\omega_{1,0}^i$ and $\check{\pi}_2^i = \pi_2^i - \psi_0\omega_{2,0}^i$, we have

$$\pi_1^i = \psi_0\omega_{1,0}^i + \check{\pi}_1^i \tag{45}$$

and

$$\pi_2^i = \psi_0\omega_{2,0}^i + \check{\pi}_2^i\,. \tag{46}$$

It follows from (43), (44), (45), and (46), that

$$\bar{\Psi}^i = (\bar{\psi}_1^i, \pi_1^i, \bar{\psi}_2^i, \pi_2^i) = -\hat{\Psi}^i + \psi_0\omega^i\,,$$

where

$$\hat{\Psi}^i = (-\psi^i(a_+^i), -\check{\pi}_1^i, \psi^{i\tilde{+}1}(a_-^{i\tilde{+}1}), -\check{\pi}_2^i)\,.$$

Then (25) says that

$$\hat{\Psi}^i - \psi_0\omega^i \in (\mathcal{C}^i)^\perp\,. \tag{47}$$

Next, we fix a value i_0 of i, let $i' = i_0\tilde{+}1$, choose all the v's, α_+'s, and $\boldsymbol{\varepsilon}$'s equal to zero, and let $\alpha_-^i = 0$ for $i \neq i'$, $\alpha_-^{i'} = \alpha_- \in C_-^{i'}$. Then

$$\bar{q}^{i'} = -\alpha_- M_{L_c^{i'}}(a_+^{i'}, a_-^{i'}) \cdot \tilde{F}_{\eta^{i'}}^{i'}(x_-^{i'}, a_-^{i'})\,,$$

and $\bar{q}^i = 0$ if $i \neq i'$. So (33) tells us that

$$-\alpha_-\langle\check{\psi}_1^{i'}, M_{L_c^{i'}}(a_+^{i'}, a_-^{i'}) \cdot \tilde{F}^{i'}_{\eta^{i'}}(x_-^{i'}, a_-^{i'})\rangle + \check{\pi}_2^{i_0} \cdot \alpha_- \leq 0\,,$$

that is,

$$-\alpha_-\langle\hat{\psi}_1^{i'}(a_-^{i'}) \cdot \tilde{F}^{i'}_{\eta^{i'}}(x_-^{i'}, a_-^{i'})\rangle + \check{\pi}_2^{i_0} \cdot \alpha_- \leq 0\,,$$

or, equivalently,

$$-\alpha_-\langle\psi_1^{i'}(a_-^{i'}) \cdot F^{i'}_{\eta^{i'}}(x_-^{i'}, a_-^{i'})\rangle + \alpha_-\psi_0 L^{i'}_{\eta^{i'}}(x_-^{i'}, a_-^{i'}) + \check{\pi}_2^{i_0} \cdot \alpha_- \leq 0\,,$$

that is,

$$\alpha_- \cdot (\check{\pi}_2^{i_0} - h_-^{i'}) \leq 0\,.$$

Since this is true for all sufficiently small $\alpha_- \in C_-^{i'}$, we conclude that

$$\check{\pi}_2^{i_0} - h_-^{i'} \in (C_-^{i'})^{\perp}\,.$$

Since i_0 was an arbitrary index in the set $\{1, \ldots, \mu\}$, and $i' = i_0 \tilde{+} 1$, we have in fact shown that

$$\check{\pi}_2^i - h_-^{i\tilde{+}1} \in (C_-^{i\tilde{+}1})^{\perp} \text{ whenever } i \in \{1, \ldots, \mu\}\,. \tag{48}$$

Given any $i \in \{1, \ldots, \mu\}$, let $i\tilde{-}1$ be the unique index $i_0 \in \{1, \ldots, \mu\}$ such that $i_0\tilde{+}1 = i$. Define

$$\kappa_-^i = \check{\pi}_2^{i\tilde{-}1} - h_-^i\,. \tag{49}$$

Then (48) says that

$$\boxed{\kappa_-^i \in (C_-^i)^{\perp} \quad \text{whenever} \quad i \in \{1, \ldots, \mu\}\,.} \tag{50}$$

Next, we fix once again a value i_0 of i, choose all the v's, α_-'s, and $\boldsymbol{\varepsilon}$'s equal to zero, and let $\alpha_+^i = 0$ for $i \neq i_0$, $\alpha_+^{i_0} = \alpha_+ \in C_+^{i_0}$. Then

$$\bar{q}^{i_0} = \alpha_+ \tilde{F}^{i_0}_{\eta^{i_0}}(X_-^{i_0}, a_+^{i_0})\,,$$

and $\bar{q}^i = 0$ if $i \neq i_0$. So (33) tells us that

$$\alpha_+\langle\check{\psi}_1^{i_0}, \tilde{F}^{i_0}_{\eta^{i_0}}(x_+^{i_0}, a_+^{i_0})\rangle + \check{\pi}_1^{i_0} \cdot \alpha_+ \leq 0\,,$$

that is,

$$\alpha_+\langle\hat{\psi}^{i_0}(a_+^{i_0}) \cdot \tilde{F}^{i_0}_{\eta^{i_0}}(x_+^{i_0}, a_+^{i_0})\rangle + \check{\pi}_1^{i_0} \cdot \alpha_+ \leq 0\,,$$

or, equivalently,

$$\alpha_+\langle\psi_1^{i_0}(a_+^{i_0}) \cdot F^{i_0}_{\eta^{i_0}}(x_+^{i_0}, a_+^{i_0})\rangle + \alpha_+\psi_0 L^{i_0}_{\eta^{i_0}}(x_+^{i_0}, a_+^{i_0}) + \check{\pi}_1^{i_0} \cdot \alpha_+ \leq 0\,,$$

that is,

$$\alpha_+ \cdot (\check{\pi}_1^{i_0} + h_-^{i_0}) \le 0\,.$$

Since this is true for all sufficiently small $\alpha_- \in C_+^{i_0}$, we conclude that

$$\check{\pi}_1^{i_0} + h_+^{i_0} \in (C_+^{i_0})^\perp\,.$$

Since i_0 was an arbitrary index in the set $\{1, \dots, \mu\}$, we have in fact shown that

$$\check{\pi}_1^i + h_+^i \in (C_+^i)^\perp \text{ whenever } i \in \{1, \dots, \mu\}\,. \tag{51}$$

Define

$$\kappa_+^i = \check{\pi}_1^i + h_+^i\,. \tag{52}$$

Then (51) says that

$$\boxed{\kappa_+^i \in (C_+^i)^\perp \quad \text{whenever} \quad i \in \{1, \dots, \mu\}\,.} \tag{53}$$

Since

$$-\check{\pi}_1^i = -\kappa_+^i + h_+^i \quad \text{cnd} \quad -\check{\pi}_2^i = -\kappa_-^i - h_-^i\,,$$

(47) implies

$$\boxed{(-\psi^i(a_+^i), -\kappa_+^i + h_+^i, \psi^{i\tilde{+}1}(a_-^{i\tilde{+}1}), -\kappa_-^i - h_-^i) \in \psi_0 \Omega^i + (\mathcal{C}^i)^\perp\,.} \tag{54}$$

Then (50), (53) and (54) show that the switching conditions hold.

The next step is to fix a value $i_0 \in \{1, \dots \mu\}$ and a $j_0 \in \{1, \dots, \nu^{i_0}\}$, choose all the v's, α_-'s, and α_+'s equal to zero, and let $\boldsymbol{\varepsilon}^i = 0$ for $i \ne i_0$, and $\boldsymbol{\varepsilon}^{i_0} = (\varepsilon_1, \dots, \varepsilon_{\nu^{i_0}})$, where $\varepsilon_j = 0$ if $j \ne j_0$, $\varepsilon_{j_0} = \varepsilon$, with $0 \le \varepsilon \le \hat{r}$. Then

$$\bar{q}^{i_0} = \varepsilon M_{L_c^{i_0}}(a_+^{i_0}, t_{j_0}^{i_0}) \cdot Z_{j_0}^{i_0}\,,$$

and $\bar{q}^i = 0$ if $i \ne i_0$. So (33) tells us that

$$\varepsilon \langle \check{\psi}_1^{i_0}, M_{L_c^{i_0}}(a_+^{i_0}, t_{j_0}^{i_0}) \cdot Z_{j_0}^{i_0} \rangle \le 0\,,$$

that is,

$$\varepsilon \langle \check{\psi}_1^{i_0} \cdot M_{L_c^{i_0}}(a_+^{i_0}, t_{j_0}^{i_0}), Z_{j_0}^{i_0} \rangle \le 0\,,$$

or, equivalently,

$$\varepsilon \langle \hat{\psi}^{i_0}(t_{j_0}^{i_0}), Z_{j_0}^{i_0} \rangle \le 0\,.$$

Since this is true for all sufficiently small nonnegative ε, we conclude that

$$\langle \hat{\psi}^{i_0}(t_{j_0}^{i_0}), Z_{j_0}^{i_0} \rangle \le 0\,.$$

Since i_0 and j_0 are arbitrary, we have shown that

$$\langle\hat{\psi}^i(t^i_j), Z^i_j\rangle \le 0 \tag{55}$$

whenever $i \in \{1, \ldots \mu\}$, $j \in \{1, \ldots, \nu^{i_0}\}$.

Now, given any i, j, (55) says that

$$\langle\hat{\psi}^i(t^i_j), \hat{Y}^i_j\rangle \le \langle\hat{\psi}^i(t^i_j), Y^i_j\rangle\,,$$

that is,

$$\begin{aligned}&\langle\psi^i(t^i_j), F^i_{\zeta^i_j}(\xi^i(t^i_j), t^i_j)\rangle - \psi_0 L^i_{\zeta^i_j}(\xi^i(t^i_j), t^i_j)\\ &\le \langle\psi^i(t^i_j), F^i_{\eta^i}(\xi^i(t^i_j), t^i_j)\rangle - \psi_0 L^i_{\eta^i}(\xi^i(t^i_j), t^i_j)\,,\end{aligned}$$

or

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

We have thus established that the $3\mu + 1$-tuple

$$(\psi^1, \ldots, \psi^\mu, \psi_0, \kappa^1_-, \kappa^1_+, \ldots, \kappa^\mu_-, \kappa^\mu_+)$$

satisfies all the conditions of the definition of $\Psi_{\mathcal{W}}$, except possibly for the normalization condition (9).

On the other hand, if we let γ be the left-hand side of (9), then $\gamma \ge 0$. If we show that $\gamma > 0$, then we can divide the ψ^i, $\kappa^i_\pm$ and ψ_0 by γ, and obtain a new multiplier for which (9) holds. Hence the conclusion that $\Psi_{\mathcal{W}} \ne \emptyset$ will follow if the prove that $\gamma \ne 0$.

Assume that $\gamma = 0$. Then $\psi_0 = 0$, and the equalities $\kappa^i_- = \kappa^i_+ = 0$ and $\psi^i(a^i_+) = 0$ hold for all i. It then follows from the adjoint differential inclusion that, for every i, $\psi^i(t) = 0$ for all $t \in [a^i_-, a^i_+]$. In particular, the definitions of the $h^i_\pm$ imply that $h^i_+ = h^i_- = 0$ for all i. Then (49) and (52) imply that $\breve{\pi}^i_1 = \breve{\pi}^i_2 = 0$ for all i. Then (31) and (32) imply (since $\psi_0 = 0$) that

$$\pi^i_1 = \pi^i_2 = 0 \quad \text{for all} \quad i\,. \tag{56}$$

Moreover, since $\psi^i(t) \equiv 0$ and $\psi_0 = 0$, (43) and (44) imply that

$$\bar{\psi}^i_1 = 0 \quad \text{and} \quad \bar{\psi}^i_2 = 0 \quad \text{for all} \quad i\,. \tag{57}$$

Then (26), (56) and (57) imply that

$$\bar{\Psi}^i = 0 \quad \text{for all} \quad i\,. \tag{58}$$

Then (23) implies that $\bar{\Psi} = 0$, contradicting the nontriviality of $\bar{\Psi}$. This contradiction shows that $\gamma \ne 0$, and our proof is complete.

11 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 $C(r) = \{v \in C : \|v\| \le r\}$.

Definition 12. 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 $\|v\| \le \delta r$ 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 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 $\alpha(1) = x$. $\diamondsuit$

It then turns out that every GDQ is a PIGD, and every derivate container, semidifferential, and multidifferential is also a PIGD. Moreover, PIGDs are intirinsically 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.