HIGH-ORDER POINT VARIATIONS AND GENERALIZED DIFFERENTIALS

H. J. Sussmann*

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

In a series of nonsmooth versions of the Pontryagin Maximum Principle, we used generalized differentials of set-valued maps, flows, and abstract variations. Bianchini and Stefani have introduced a notion of possibly high-order variational vector that has the summability property. We consider a slightly more general class of variational vectors than that defined by Bianchini and Stefani, and prove that the convex combinations of these vectors arise as "differentials" of variations that are differentiable in the sense of one of our generalized differentiation theories, namely, that of "approximate generalized differential quotients" (AGDQs).

1. Introduction

In a series of papers (cf. Refs. 5–7,9,10), we showed how to derive general, nonsmooth versions of the Pontryagin Maximum Principle using generalized differentials of set-valued maps, flows, and abstract point variations. The use of general variations rather than the nedle variations used to prove the ordinary maximum principle makes it possible to obtain high-order versions of the maximum principle. The main technical difficulty with these general abstract variations is that they need not have the summability property, which is absolutely essential in order to derive the necessary conditions for optimality.

R. M. Bianchini and G. Stefani (cf. Refs. 1–4) proposed a concept of high-order point variation that has good summability properties. The

*Supported in part by NSF Grant DMS-05-09930

goal of this note is to relate this concept to a theory of generalized differentials, by describing a slightly more general version of the Bianchini-Stefani variations, and showing that they are differentiable in the precise sense of the theory of "Approximate Generalized Differential Quotients" (AGDQs). This makes it possible to use these variations in order to get additional necessary conditions for an optimum in situations such as the very general one described in Ref. 9, where the differentials involved are generalized differential quotients, and *a fortiori* AGDQs.

1.1. *Preliminary remarks on notation*

We will use the notations and abbreviations of Ref. 9. In particular, "FDRLS" stands for "finite-dimensional real linear space," "FDNRLS" for "normed FDRLS," and "SVM" for "set-valued map." If f is a SVM, then $\mathrm{So}(f)$, $\mathrm{Ta}(f)$, $\mathrm{Gr}(f)$, $\mathrm{Do}(f)$, $\mathrm{Im}(f)$ are, respectively, the source, target, graph, domain and image of a SVM f. (We recall that a ***SVM*** is a triple (A, B, G) such that A, B are sets and G is a subset of $A\times B$, in which case we say that f is a ***SVM from*** A ***to*** B, and define $G^{-1}\stackrel{\mathrm{def}}{=}\{(x,y) : (y,x) \in G\}$, $f^{-1}\stackrel{\mathrm{def}}{=} = (B, A, G^{-1})$, $\mathrm{So}(f)\stackrel{\mathrm{def}}{=}A$, $\mathrm{Ta}(f)\stackrel{\mathrm{def}}{=}B = \mathrm{So}(f^{-1})$, $\mathrm{Gr}(f) = G$, $f(x) = \{y : (x,y) \in \mathrm{Gr}(f)\}$, $\mathrm{Do}(f) = \{x : f(x) \neq \emptyset\}$, $\mathrm{Im}(f) = \mathrm{Do}(f^{-1})$.) We use $SVM(A,B)$ to denote the set of all set-valued maps from A to B. The notation "$f : A \mapsto\!\!\!\!\to B$" means "$f$ is a set-valued map from A to B." If $f \in SVM(A,B)$ then f is (i) ***single-valued*** if the set $f(x)$ consists of a single member for every $x \in \mathrm{Do}(f)$, (ii) ***one-to-one*** if f^{-1} is single-valued, (iii) ***surjective*** if $\mathrm{Im}(f) = \mathrm{Ta}(f)$, (iv) ***everywhere defined*** if $\mathrm{Do}(f) = \mathrm{So}(f)$, i.e., if f^{-1} is surjective, (v) ***a ppd map*** (where "ppd" stands for "possibly partially defined") if it is single-valued, and (vi) ***an ordinary map*** if it is an everywhere defined ppd map. The notation "$f : A \hookrightarrow B$" means "f is a ppd map from A to B."

If S is a set, then $\mathbb{I}_S$ is the identity map of S, i.e., the triple (S, S, Δ_S), where $\Delta_S = \{(x,x) : x \in S\}$.

The abbreviation "CCA" stands for for "Cellina continuously approximable." (We recall that a ***CCA map*** from a metric space X to a metric space Y is a set-valued map $F : X \mapsto\!\!\!\!\to Y$ such that, for every compact subset K of X, (i) the set $(K \times Y) \cap \mathrm{Gr}(F)$ is compact, and (ii) there exists a sequence $\{F_j\}_{j=1}^{\infty}$ of single-valued continuous maps from K to Y such that the graphs $\mathrm{Gr}(F_j)$ converge to $\mathrm{Gr}(F)$, in the sense that

$$\lim_{j\to\infty} \sup\{\mathrm{dist}_{X\times Y}(q, \mathrm{Gr}(F)) : q \in \mathrm{Gr}(F_j)\} = 0\,.$$

(A detailed study of CCA maps appears in Ref. 9.) We use $CCA(X,Y)$ to denote the set of all CCA maps from X to Y.

If I is a totally ordered set, then we use $\leq_I$ to denote the order relation on I, and simply write $\leq$ when the context makes I unambiguous. Also, "$a <_I b$"—or, simply, "$a < b$"—means "$a \leq_I b$ and $a \neq b$." A ***subinterval*** of I is a subset J of I such that $c \in J$ whenever $a \in J$, $b \in J$, $c \in I$, and $a \leq c \leq b$. We use square bracket notation for subintervals of I that have an infimum and a supremum in I. (That is, if $a, b \in I$ and $a \leq b$, we write $]a,b[_I \stackrel{\text{def}}{=} \{t \in I : a < t < b\}$, $[a,b[_I \stackrel{\text{def}}{=} \{a\} \cup]a,b[_I$, $]a,b]_I \stackrel{\text{def}}{=}]a,b[_I \cup \{b\}$, and $[a,b]_I \stackrel{\text{def}}{=} \{a,b\} \cup]a,b[_I$, and we omit the subscript when I is uniquely determined by the context.). Then every subinterval of I that has an infimum and a supremum in I is of one of the forms $]a,b[$, $[a,b]$, $[a,b[$, $]a,b]$). When the totally ordered set is not specified, it is understood that it is the extended real line $\bar{\mathbb{R}} \stackrel{\text{def}}{=} \mathbb{R} \cup \{-\infty, +\infty\}$. We define $\mathbb{R}_+ \stackrel{\text{def}}{=} \{x \in \mathbb{R} : x \geq 0\}$ and $\mathbb{R}_{+,>} \stackrel{\text{def}}{=} \{x \in \mathbb{R} : x > 0\}$, and let $\bar{\mathbb{R}}_+ \stackrel{\text{def}}{=} \mathbb{R}_+ \cup \{+\infty\}$.

We use $\mathbf{\Theta}$ to denote the class of all functions $\theta : \mathbb{R}_{+,>} \mapsto \bar{\mathbb{R}}_+$ such that

- θ is monotonically nondecreasing (that is, $\theta(s) \leq \theta(t)$ whenever $s, t \in \mathbb{R}$ are such that $0 \leq s \leq t < +\infty$);
- $\lim_{s \downarrow 0} \theta(s) = 0$.

If X is a FDNRLS, $x_* \in X$, $r > 0$, then $\mathbb{B}_X(x_*, r)$, $\bar{\mathbb{B}}_X(x_*, r)$ are, respectively, the open ball $\{x \in X : \|x - x_*\| < r\}$ and the closed ball $\{x \in X : \|x - x_*\| \leq r\}$. If X, Y are FDRLSs, then $Lin(X,Y)$, $Aff(X,Y)$ will denote, respectively, the set of all linear maps and the set of all affine maps from X to Y. By definition, the members of $Aff(X,Y)$ are the maps $\mathit{affm}_{L,h}$, for $L \in Lin(X,Y)$, $h \in Y$, where $\mathit{affm}_{L,h}$ denotes the ***affine map with linear part*** L ***and constant part*** h, defined by $\mathit{affm}_{L,h}(x) \stackrel{\text{def}}{=} L \cdot x + h$. We identify $Aff(X,Y)$ with $Lin(X,Y) \times Y$ by identifying each map $\mathit{affm}_{L,h} \in Aff(X,Y)$ with the pair $(L,h) \in Lin(X,Y) \times Y$.

If X and Y are FDNRLSs, then we endow $Lin(X,Y)$ with the operator norm $\|\cdot\|_{op}$ given by $\|L\|_{op} = \sup\{\|Lx\| : x \in X, \|x\| \leq 1\}$, so $Lin(X,Y)$ is a FDNRLS as well. Also, we endow the linear space $Aff(X,Y)$ with the norm given by $\|\mathit{affm}_{L,h}\| = \|L\| + \|h\|$.

If Λ is subset of $Lin(X,Y)$, and $\delta \in \mathbb{R}_{+,>}$, we define

$$\Lambda^\delta = \{L \in Lin(X,Y) : \text{dist}(L, \Lambda) \leq \delta\} ,$$

where $\mathrm{dist}(L,\Lambda) = \inf\{\|L - L'\|_{op} : L' \in \Lambda\}$. Also, if $\delta, \varepsilon \in \mathbb{R}_{+,>}$, and we still assume that $\Lambda \subseteq Lin(X,Y)$, we let

$$\Lambda^{(\delta,\varepsilon)} = \{\mathit{affm}_{L,h} : L \in Lin(X,Y),\ \mathrm{dist}(L,\Lambda) \le \delta,\ h \in Y,\ \|h\| \le \delta\varepsilon\}\,,$$

Notice that if $L \in Lin(X,Y)$, then $\mathrm{dist}(L,\emptyset) = +\infty$. In particular, if $\Lambda = \emptyset$ then $\Lambda^{\delta} = \emptyset$. and $\Lambda^{(\delta,\varepsilon)} = \emptyset$. Notice also that if Λ is compact (resp. convex) then Λ^{δ} and $\Lambda^{(\delta,\varepsilon)}$ are compact (resp. convex).

If X is a FDRLSs, then we use $X^{\dagger}$ to denote the dual space of X, i.e., the space $Lin(X,\mathbb{R})$.

The word "manifold" will mean "finite-dimensional paracompact differentiable manifold without boundary." If M is a manifold of class C^1, and $x \in M$, then T_xM, T_x^*M denote, respectively, the tangent and cotangent space of M at x.

1.2. *Approximate Generalized Differential Quotients*

Definition 1.1. Assume that X, Y are FDNRLSs, $F : X \mapsto\!\!\!\!\rightarrow Y$ is a set-valued map, Λ is a compact subset of $Lin(X,Y)$, $\bar{x}_* \in X$, $\bar{y}_* \in Y$, and $S \subseteq X$. We say that Λ is an ***approximate generalized differential quotient of*** F ***at*** $(\bar{x}_*, \bar{y}_*)$ ***in the direction of*** S—and write $\Lambda \in AGDQ(F, \bar{x}_*, \bar{y}_*, S)$—if there exists a function $\theta \in \boldsymbol{\Theta}$—called an ***AGDQ modulus for*** $(\Lambda, F, \bar{x}_*, \bar{y}_*, S)$—having the property that

(*) *for every* $\varepsilon \in \mathbb{R}_{+,>}$ *such that* $\theta(\varepsilon) < \infty$ *there exists a set-valued map* $A^{\varepsilon} \in CCA(\bar{\mathbb{B}}_X(\bar{x}_*, \varepsilon) \cap S, Aff(X,Y))$, *with values in* $\Lambda^{(\theta(\varepsilon),\varepsilon)}$, *such that* $\bar{y}_* + A(x - \bar{x}_*) \in F(x)$ *whenever* $x \in \bar{\mathbb{B}}_X(\bar{x}_*, \varepsilon) \cap S$ *and* $A \in A^{\varepsilon}(x)$. □

1.2.1. *Properties of AGDQs*

If A, B, C are sets, and Ξ, Z are sets of maps from A to B and from B to C, respectively, then the ***composite*** $Z \circ \Xi$ is the set of maps from A to C given by $Z \circ \Xi = \{\zeta \circ \xi : \zeta \in Z,\ \xi \in \Xi\}$.

The following statement, proved in Ref. 9, is the ***chain rule*** for AGDQs.

Theorem 1.1. *For* $i = 1, 2, 3$, *let* X_i *be a FDNRLS, and let* $\bar{x}_{*,i}$ *be a point of* X_i. *Assume that, for* $i = 1, 2$, *(i)* $F_i : X_i \mapsto\!\!\!\!\rightarrow X_{i+1}$ *is a set-valued map, (ii)* S_i *is a subset of* X_i, *and (iii)* $\Lambda_i \in AGDQ(F_i, \bar{x}_{*,i}, \bar{x}_{*,i+1}, S_i)$. *Assume, in addition, that (iv)* $F_1(S_1) \subseteq S_2$, *and either (v)* S_2 *is a local quasiretract (cf. Remark 1.1) of* X_2 *at* $\bar{x}_{*,2}$ *or (v') there exists a neighborhood* U *of* $\bar{x}_{*,1}$ *in* X_1 *such that the restriction* $F_1 \lceil (U \cap S_1)$ *of* F_1 *to* $U \cap S_1$ *is single-valued. Then* $\Lambda_2 \circ \Lambda_1 \in AGDQ(F_2 \circ F_1, \bar{x}_{*,1}, \bar{x}_{*,3}, S_1)$. □

Remark 1.1. The notion of a "local quasiretract" is defined in Ref. 9. The precise definition is as follows. First, if T is a topological space and $S \subseteq T$, we say that S is a ***quasiretract of*** T if for every compact subset K of S there exist a neighborhood U of K and a continuous map $\rho : U \mapsto S$ such that $\rho(s) = s$ for every $s \in K$. Then, if $S \subseteq T$ and $\bar{s} \in S$, we say that S is a ***local quasiretract of*** T ***at*** $\bar{s}$ if there exists a neighborhood U of $\bar{s}$ such that $S \cap U$ is a quasiretract of U.

An important example of a local quasiretract of a manifold M at a point $s \in M$ is a subset S of M such that, for some open neighborhood U of s, the set $S \cap U$ is the image of a convex subset of an open neighborhood V of 0 in $\mathbb{R}^{\dim M}$ under a diffeomorphism Φ of class C^1 from V onto U such that $\Phi(0) = s$. In particular, any set whose germ at s is, relative to some coordinate chart near s, the germ at s of a convex subset of $\mathbb{R}^{\dim M}$, is a local quasiretract of M at s. □

If M and N are manifolds of class C^1, $\bar{x}_* \in M$, $\bar{y}_* \in N$, $S \subseteq M$, and $F : M \mapsto\!\!\!\to N$, then it is possible to define a set $AGDQ(F, \bar{x}_*, \bar{y}_*, S)$ of compact subsets of the space $Lin(T_{\bar{x}_*}M, T_{\bar{y}_*}N)$ of linear maps from $T_{\bar{x}_*}M$ to $T_{\bar{y}_*}N$ as follows. We let $m = \dim M$, $n = \dim N$, and pick coordinate charts $\xi : M \hookrightarrow \mathbb{R}^m$, $\eta : N \hookrightarrow \mathbb{R}^n$, defined near $\bar{x}_*$, $\bar{y}_*$ and such that $\xi(\bar{x}_*) = 0$ and $\eta(\bar{y}_*) = 0$, and declare that a subset Λ of $Lin(T_{\bar{x}_*}M, T_{\bar{y}_*}N)$ belongs to $AGDQ(F, \bar{x}_*, \bar{y}_*, S)$ if the composite set of maps $D\eta(\bar{y}_*) \circ \Lambda \circ D\xi(\bar{x}_*)^{-1}$ is in $AGDQ(\eta \circ F \circ \xi^{-1}, 0, 0, \xi(S)))$. It then follows easily from the chain rule that, with this definition, ***the set*** $AGDQ(F, \bar{x}_*, \bar{y}_*, S)$ ***does not depend on the choice of the charts*** ξ***,*** η***.*** In other words, ***the notion of an AGDQ is invariant under*** C^1 ***diffeomorphisms and therefore makes sense intrinsically on manifolds of class*** C^1***.***

Then the chain rule also holds on manifolds, as pointed out in Ref. 9.

Proposition 1.1. *Assume that*

- *(I) for* $i = 1, 2, 3$, M_i *is a manifold of class* C^1 *and* $\bar{x}_{*,i} \in M_i$,
- *(II)* $S_i \subseteq M_i$, $F_i : M_i \mapsto\!\!\!\to M_{i+1}$, *and* $\Lambda_i \in AGDQ(F_i, \bar{x}_{*,i}, \bar{x}_{*,i+1}, S_i)$ *for* $i = 1, 2$,
- *(III) either* S_2 *is a local quasiretract of* M_2 *or* F_1 *is single-valued on* $U \cap S_1$ *for some neighborhood* U *of* $\bar{x}_{*,1}$.

Then the composite $\Lambda_2 \circ \Lambda_1$ *belongs to* $AGDQ(F_2 \circ F_1, \bar{x}_{*,1}, \bar{x}_{*,3}, S_1)$. □

Furthermore, AGDQs have several natural properties. First, the following statement, proved in Ref. 9, says that classical differentials at one point of continuous maps and Clarke generalized Jacobians of Lipschitz maps are AGDQs.

Proposition 1.2. *If M, N are manifolds of class C^1, $S \subseteq M$, $\bar{x}_* \in M$, $\bar{y}_* \in N$, $F : M \mapsto\!\!\!\rightarrow N$, U is an open neighborhood of $\bar{x}_*$ in M, and $F(\bar{x}_*) = \{\bar{y}_*\}$, then*

(1) If (i) the restriction $F \lceil (U \cap S)$ is a continuous everywhere defined map, (ii) L is a differential of F at $\bar{x}_*$ in the direction of S (that is, $L \in Lin(T_{\bar{x}_*}M, T_{\bar{y}_*}N)$ and $\|F(x) - F(\bar{x}_*) - L \cdot (x - \bar{x}_*)\| = o(\|x - \bar{x}_*\|)$ as $x \to \bar{x}_*$ via values in S, relative to some choice of coordinate charts about $\bar{x}_*$ and $\bar{y}_*$), then $\{L\}$ belongs to $AGDQ(F, \bar{x}_*, \bar{y}_*, S)$.

(2) If (i) the restriction $F \lceil U$ is a locally Lipschitz everywhere defined map, and (ii) Λ is the Clarke generalized Jacobian of F at $\bar{x}_*$, then Λ belongs to $AGDQ(F, \bar{x}_*, \bar{y}_*, M)$. □

The following two propositions, also proved in Ref. 9, are the Cartesian product rule and the assertion that AGDQs are local, in the sense that the set $AGDQ(F, \bar{x}_*, \bar{y}_*, S)$ is completely determined by the germ of the set S at $\bar{x}_*$ and the germ of the graph of F at $(\bar{x}_*, \bar{y}_*)$. In Proposition 1.3, if A, B, C, D are sets and $\mu : A \mapsto\!\!\!\rightarrow C$, $\nu : B \mapsto\!\!\!\rightarrow D$, then $\mu \times \nu$ is the set-valued map from $A \times B$ to $C \times D$ that sends each point $(a, b) \in A \times B$ to the subset $\mu(a) \times \nu(b)$ of $C \times D$. (In particular, if μ and ν are ordinary single-valued maps, then $\mu \times \nu$ is an ordinary single-valued map, given by $(\mu \times \nu)(a, b) = (\mu(a), \nu(b))$ for $a \in A$, $b \in B$.) If $\mathcal{M}, \mathcal{N}$ are sets of SVMs from A to C and from B to D, respectively, then $\mathcal{M} \times \mathcal{N}$ is the set of all SVMs $\mu \times \nu$, $\mu \in \mathcal{M}$, $\nu \in \mathcal{N}$. The spaces $T_{\bar{x}_{*,1}}M_1 \times T_{\bar{x}_{*,2}}M_2$ $T_{\bar{y}_{*,1}}N_1 \times T_{\bar{y}_{*,2}}N_2$ are identified with $T_{(\bar{x}_{*,1}, \bar{x}_{*,2})}(M_1 \times M_2)$ and $T_{(\bar{y}_{*,1}, \bar{y}_{*,2})}(N_1 \times N_2)$, respectively.

Proposition 1.3. (The product rule.) *Assume that*
(1) for $i = 1, 2$, M_i and N_i are manifolds of class C^1, $S_i \subseteq M_i$, $\bar{x}_{*,i} \in M_i$, $\bar{y}_{*,i} \in N_i$, $F_i : M_i \mapsto\!\!\!\rightarrow N_i$, and $\Lambda_i \in AGDQ(F_i, \bar{x}_{*,i}, \bar{y}_{*,i}, S_i)$;
(2) $\bar{x}_* = (\bar{x}_{*,1}, \bar{x}_{*,2})$, $\bar{y}_* = (\bar{y}_{*,1}, \bar{y}_{*,2})$, $S = S_1 \times S_2$, and $F = F_1 \times F_2$.
Then $\Lambda_1 \times \Lambda_2 \in AGDQ(F, \bar{x}_*, \bar{y}_*, S)$. □

Proposition 1.4. (Locality.) *Assume that (1) M, N, are manifolds of class C^1, (2) $\bar{x}_* \in M$, (3) $\bar{y}_* \in N$, (4) $S_i \subseteq M$ and $F_i : M \mapsto\!\!\!\rightarrow N$ for $i = 1, 2$, and (5) the sets S_1 and S_2 have the same germ at $\bar{x}_*$, and the graphs $\mathrm{Gr}(F_1)$, $\mathrm{Gr}(F_2)$, have the same germ at $(\bar{x}_*, \bar{y}_*)$ (that is, there exist neighborhoods U, V of $\bar{x}_*$, $\bar{y}_*$, in M, N, respectively, such that $U \cap S_1 = U \cap S_2$ and $(U \times V) \cap \mathrm{Gr}(F_1) = (U \times V) \cap \mathrm{Gr}(F_2)$). Then $AGDQ(F_1, \bar{x}_*, \bar{y}_*, S_1) = AGDQ(F_2, \bar{x}_*, \bar{y}_*, S_2)$.* □

1.2.2. *Uniform AGDQs*

Assume that X and Y are FDNRLSs, and we are given a family $\{(F_\alpha, x_\alpha, y_\alpha, S_\alpha)\}_{\alpha\in A}$ of 4-tuples, such that each F_α is a set-valued map from X to Y, each x_α is a point of X, each y_α is a point of Y, and each S_α is a subset of X. We say that a family $\{\Lambda_\alpha\}_{\alpha\in A}$ of compact subsets of $Lin(X,Y)$ is a ***uniform AGDQ of the maps*** F_α ***at the points*** (x_α, y_α) ***in the direction of the*** S_α if there exists a function $\theta \in \boldsymbol{\Theta}$ which is an AGDQ modulus for $(\Lambda_\alpha, F_\alpha, x_\alpha, y_\alpha, S_\alpha)$ for each $\alpha \in A$.

The concept of a uniform AGDQ makes sense as well, in an intrinsic way, when X and Y are manifolds, provided that the family $\{(F_\alpha, x_\alpha, y_\alpha, S_\alpha)\}_{\alpha\in A}$ is such that the set $Q = \{(x_\alpha, y_\alpha) : \alpha \in A\}$ is precompact in $X \times Y$. Indeed, let d_X, d_Y be the dimensions of X and Y. If Q is precompact in $X \times Y$, then we can find a finite family $\Sigma = \{(\xi_j, U_j, \eta_j, V_j, K_j, L_j)\}_{1\le j\le m}$ such that

(1) for each j, (i) ξ_j is a coordinate chart of X with domain U_j, (ii) η_j is a coordinate chart of Y with domain V_j, (iii) K_j is a compact subset of U_j, and (iv) L_j is a compact subset of V_j;
(2) $Q \subseteq \bigcup_{j=1}^m (K_j \times L_j)$.

Then, if we let $A_j = \{\alpha \in A : (x_\alpha, y_\alpha) \in K_j \times L_j\}$, it is clear that $A = \bigcup_{j=1}^m A_j$, and we can consider, for each j, the family $\Phi_j = \{(\tilde{F}_{j,\alpha}, x_\alpha, y_\alpha, \tilde{S}_{j,\alpha})\}_{\alpha\in A_j}$, where $\tilde{F}_{j,\alpha}$ is the set-valued map from U_j to V_j whose graph is $\mathrm{Gr}(F_\alpha) \cap (U_j \times V_j)$, and $\tilde{S}_{j,\alpha} = S_\alpha \cap U_j$. If we identify U_j, V_j with open subsets $\tilde{U}_j$, $\tilde{V}_j$ of $\mathbb{R}^{d_X}$, $\mathbb{R}^{d_Y}$, then $\{\tilde{F}_{j,\alpha}\}_{\alpha\in A_j}$ is a family of set-valued maps from $\mathbb{R}^{d_X}$ to $\mathbb{R}^{d_Y}$, the x_α belong to $\mathbb{R}^{d_X}$, the y_α belong to $\mathbb{R}^{d_Y}$, and $\tilde{S}_\alpha$ is a subset of $\mathbb{R}^{d_X}$, so we are in the situation of the previous paragraph, and it makes sense to talk about a "uniform AGDQ" $\{\Lambda_\alpha\}_{\alpha\in A_j}$ of the family Φ_j. We then say that a family $\{\Lambda_\alpha\}_{\alpha\in A}$ is a ***uniform AGDQ of the family*** $\{(F_\alpha, x_\alpha, y_\alpha, S_\alpha)\}_{\alpha\in A}$ if, for some choice of m and the family $\Sigma = \{(\xi_j, U_j, \eta_j, V_j, K_j, L_j)\}_{1\le j\le m}$ as above, it turns out that $\{\Lambda_\alpha\}_{\alpha\in A_j}$ is a uniform AGDQ of Φ_j for each j. (It is easily seen that if this condition holds for one choice of m and Σ, then it holds for all such choices.)

1.2.3. *AGDQ approximating multicones.*

A ***cone*** in a FDRLS X is a nonempty set C which is closed under multiplication by nonnegative real numbers, i.e., such that $rc \in C$ whenever $c \in C$ and $r \ge 0$. The ***polar*** of a cone C in X is the subset $C^\dagger$ of $X^\dagger$ defined by $C^\dagger = \{\mu \in X^\dagger : \mu(c) \le 0$ whenever $c \in C\}$. Clearly, $C^\dagger$ is always a closed convex cone. If we identify $X^{\dagger\dagger}$ with X in the usual way, then $C \subseteq C^{\dagger\dagger}$, and $C = C^{\dagger\dagger}$ if and only if C is closed and convex.

A ***multicone*** is a nonempty set of cones. A multicone $\mathcal{M}$ is ***convex*** if all the members of $\mathcal{M}$ are convex cones. The ***polar*** $\mathcal{M}^\dagger$ of a multicone $\mathcal{M}$ is the closure of the union of the polars $M^\dagger$, $M \in \mathcal{M}$. Therefore $\mathcal{M}^\dagger$ is a always a closed cone in $X^\dagger$. Naturally, $\mathcal{M}^\dagger$ need not be convex in general.

Definition 1.2. Assume that M is a manifold of class C^1, S is a subset of M, and $\bar{x}_* \in S$. An ***AGDQ approximating multicone to*** S ***at*** $\bar{x}_*$ is a convex multicone $\mathcal{C}$ in $T_{\bar{x}_*}M$ such that there exist a nonnegative integer m, a set-valued map $F : \mathbb{R}^m \mapsto\!\!\!\!\to M$, a convex cone D in $\mathbb{R}^m$, and a $\Lambda \in AGDQ(F, 0, \bar{x}_*, D)$, such that $F(D) \subseteq S$ and $\mathcal{C} = \{LD : L \in \Lambda\}$. □

1.3. *Transversality of cones and multicones.*

If S_1, S_2 are subsets of a linear space X, we define the ***sum*** $S_1 + S_2$ and the ***difference*** $S_1 - S_2$ by letting $S_1 + S_2 = \{s_1 + s_2 : s_1 \in S_1, s_2 \in S_2\}$, $S_1 - S_2 = \{s_1 - s_2 : s_1 \in S_1, s_2 \in S_2\}$.

Recall that if S_1, S_2 are linear subspaces of a FDRLS X, then S_1 and S_2 are ***transversal*** if $S_1 + S_2 = X$, or, equivalently, if $S_1 - S_2 = X$. If two submanifolds M_1, M_2 of class C^1 intersect at a point x_*, and S_1, S_2 are their tangent spaces at x_*, then it is well known that if S_1 and S_2 are transversal then $M_1 \cap M_2$ looks, near x_*, like $S_1 \cap S_2$. In particular, if $S_1 \cap S_2 \neq \{0\}$ (i.e., if $\dim S_1 \cap S_2 \geq 1$), then $M_1 \cap M_2$ contains a nontrivial curve going through x_*. The following definitions generalize the concept of transversality and that of "transversality with a nontrivial intersection," first to cones and then to multicones.

Definition 1.3. Let X be a FDRLS, and let C_1, C_2 be two convex cones in X. We say that C_1 and C_2 are ***transversal***, and write $C_1 \overline{\pitchfork} C_2$, if $C_1 - C_2 = X$. We say that C_1 and C_2 are ***strongly transversal***, and write $C_1 \overline{\overline{\pitchfork}} C_2$, if $C_1 \overline{\pitchfork} C_2$ and in addition $C_1 \cap C_2 \neq \{0\}$. □

In order to extend Definition 1.3 to multicones, it is convenient to start by reformulating the concept of strong transversality of cones, by making the trivial observation that $C_1 \overline{\overline{\pitchfork}} C_2$ if and only if the following two conditions hold: (i) $C_1 \overline{\pitchfork} C_2$ and (ii) there exists a linear functional $\mu \in X^\dagger$ such that $\mu(v) > 0$ for some $v \in C_1 \cap C_2$.

In view of the above reformulation, we define a linear functional $\mu : X \mapsto \mathbb{R}$ to be ***intersection-positive*** on a pair $(\mathcal{C}_1, \mathcal{C}_2)$ of multicones, if the set $\{c \in C_1 \cap C_2 : \mu(c) > 0\}$ is nonempty for every $C_1 \in \mathcal{C}_1$ and every $C_2 \in \mathcal{C}_2$. Using this concept, the definitions of "transversality" and "strong transversality" of convex multicones are nearly identical to the definitions for cones.

Definition 1.4. Let X be a FDRLS. We say that two convex multicones $\mathcal{C}_1$ and $\mathcal{C}_2$ in X are ***transversal***, and write $\mathcal{C}_1 \overline{\pitchfork} \mathcal{C}_2$, if $C_1 \overline{\pitchfork} C_2$ for all $C_1 \in \mathcal{C}_1$, $C_2 \in \mathcal{C}_2$. We say that $\mathcal{C}_1$ and $\mathcal{C}_2$ are ***strongly transversal***, and write $\mathcal{C}_1 \overline{\pitchfork\!\!\!\pitchfork} \mathcal{C}_2$, if (i) $\mathcal{C}_1 \overline{\pitchfork} \mathcal{C}_2$, and (ii) there exists a linear functional $\mu \in X^\dagger$ which is intersection-positive on $(\mathcal{C}_1, \mathcal{C}_2)$. □

Two convex cones C_1, C_2 in a FDRLS X are ***linearly separated*** if there exists a nontrivial linear functional $\lambda \in X^\dagger$ such that $\lambda(c) \leq 0$ whenever $c \in C_1$, and $\lambda(c) \geq 0$ whenever $c \in C_2$. (Equivalently, C_1 and C_2 are linearly separated if and only if $C_1^\dagger \cap (-C_2)^\dagger \neq \{0\}$.) It is easy to see that C_1 and C_2 are linearly separated if and only if they are not transversal. In view of this, we will call two convex multicones $\mathcal{C}_1$, $\mathcal{C}_2$, ***linearly separated*** if they are not transversal. Since strong transversality is a stronger property than transversality, its negation is weaker than the negation of transversality, i.e., than linear separation. So we will say that two convex multicones $\mathcal{C}_1$, $\mathcal{C}_2$, are ***weakly linearly separated*** if they are not strongly transversal.

The following characterization of weak linear separation is proved in Ref. 10.

Proposition 1.5. *Let $\mathcal{C}_1$, $\mathcal{C}_2$ be convex multicones in a FDRLS X. Then the following conditions are equivalent:*

1. *$\mathcal{C}_1$ and $\mathcal{C}_2$ are weakly linearly separated;*
2. *for every $\mu \in X^\dagger \backslash \{0\}$ there exist $\pi_0, \pi_1, \pi_2, C_1, C_2$ such that*
   - *i.* $\pi_0 \in \mathbb{R}$ *and* $\pi_0 \geq 0$,
   - *ii.* $C_1 \in \mathcal{C}_1$ *and* $C_2 \in \mathcal{C}_2$,
   - *iii.* $\pi_1 \in C_1^\dagger$ *and* $\pi_2 \in C_2^\dagger$,
   - *iv.* $\pi_0 \mu = \pi_1 + \pi_2$,
   - *v.* $(\pi_0, \pi_1, \pi_2) \neq (0, 0, 0)$. □

1.4. ***The nonseparation theorem.***

The crucial fact about AGDQs that leads to the maximum principle is the transversal intersection property, that we now state (cf. Ref. 9 for the proof).

Theorem 1.2. *Let M be a manifold of class C^1, let S_1, S_2 be subsets of M, and let $\bar{s}_* \in S_1 \cap S_2$. Let $\mathcal{C}_1$, $\mathcal{C}_2$ be AGDQ-approximating multicones to S_1, S_2 at $\bar{s}_*$ such that $\mathcal{C}_1 \pitchfork \mathcal{C}_2$. Then S_1 and S_2 are not locally separated at $\bar{s}_*$. (That is, the set $S_1 \cap S_2$ contains a sequence of points s_j converging to $\bar{s}_*$ but not equal to $\bar{s}_*$.)* □

Theorem 1.2 and Proposition 1.5 trivially imply the following result.

Corollary 1.1. *Let M be a manifold of class C^1, let S_1, S_2 be subsets of M, and let $\bar{s}_* \in S_1 \cap S_2$. Let $\mathcal{C}_1$, $\mathcal{C}_2$ be AGDQ-approximating multicones to S_1, S_2 at $\bar{s}_*$. Assume that S_1 and S_2 are locally separated at $\bar{s}_*$. (That is, there exists a neighborhood U of $\bar{s}_*$ such that $S_1 \cap S_2 \cap U = \{\bar{s}_*\}$). Then Condition 2 of the statement of Proposition 1.5 holds.* □

The more familiar forms of the maximum principle for optimal control follow by applying Corollary 1.1 to suitable choices of M, S_1, S_2, $\mathcal{C}_1$, $\mathcal{C}_2$, $\bar{s}_*$, and using the conclusion of the corollary with a suitable μ. For example, consider a fixed time interval optimal control problem $\mathcal{P}$ whose data 9-tuple $D = (M_0, U, a, b, \mathcal{U}, f, L, \bar{x}_*, S)$ satisfies (D1) the state space M_0 is a smooth manifold, (D2) U is a set, (D3) $a, b \in \mathbb{R}$ and $a < b$, (D4) $\mathcal{U}$ (the class of "admissible controls") is a set of U-valued functions on $[a, b]$, (D5) $(L(x, u, t,), f(x, u, t)) \in \mathbb{R} \times T_x M_0$ for each $(x, u, t) \in M_0 \times U \times [a, b]$, (D6) $\bar{x}_* \in M_0$, and (D7) $S \subseteq M_0$. Suppose that the objective of $\mathcal{P}$ is to minimize the integral $\int_a^b L(\xi(t), \eta(t), t)\, dt$, subject to the following conditions: (C1) $\xi : [a, b] \mapsto M_0$ is absolutely continuous, (C2) $\eta \in \mathcal{U}$, (C3) $\dot{\xi}(t) = f(\xi(t), \eta(t), t)$ for almost all $t \in [a, b]$, (C4) $\xi(a) = \bar{x}_*$, and (C5) $\xi(b) \in S$. We then take $M = \mathbb{R} \times M_0$. If a trajectory-control pair (ξ_*, η_*) is a solution of $\mathcal{P}$, we take S_1 to be the set of all points $(r, x) \in M$ such that x is reachable from $\bar{x}_*$ over $[a, b]$ with cost r, and we take S_2 to be the set $(\,]-\infty, r_*[\times S) \cup \{q_*\}$, where $q_* = (r_*, \xi_*(b))$, and r_* is the cost of (ξ_*, η_*). Then the optimality of (ξ_*, η_*) implies that S_1 and S_2 are locally separated at q_*. We then take $\mathcal{C}_1$ to be an AGDQ-approximating multicone to S_1 at q_* obtained by constructing variations and propagating their effects to the terminal point of ξ_*, and take $\mathcal{C}_2 = \{\,]-\infty, 0] \times C : C \in \mathcal{C}\}$, where $\mathcal{C}$ is an AGDQ-approximating multicone to S at $\xi_*(b)$. We choose μ to be the linear functional on $T_{q_*} M \sim \mathbb{R} \times T_{\xi_*(b)} M_0$ given by $\mu(r, v) = -r$, so $-\mu \in C_2^\dagger$ for every $C_2 \in \mathcal{C}_2$. Corollary 1.1 then yields a decomposition $\pi_0 \mu = \pi_1 + \pi_2$, where $\pi_1 \in C_1^\dagger$, $\pi_2 \in C_2^\dagger$, $\pi_0 \geq 0$, $(\pi_0, \pi_1, \pi_2) \neq (0, 0, 0)$, $C_1 \in \mathcal{C}_1$, and $C_2 \in \mathcal{C}_2$. Then $-\pi_1 = \pi_2 - \pi_0 \mu$. Since $\pi_2 \in C_2^\dagger$, $\pi_0 \geq 0$, and $-\mu \in C_2^\dagger$, it follows that $-\pi_1 \in C_2^\dagger$. If we write $C_2 =]-\infty, 0] \times C$, $C \in \mathcal{C}$, and let $\pi_1 = (-\rho, \bar{\pi})$, then the fact that $-\pi_1 \in C_2^{\dagger}$. implies that $\rho \geq 0$ and $-\bar{\pi} \in C^{\dagger}$. Then $\bar{\pi}$ and ρ are, respectively, the terminal adjoint vector (often called $\psi(b)$ or $\lambda(b)$ in the literature) and the additional multiplier (often called ψ_0 or λ_0) conjugate to the cost r, and the familiar conclusions of the maximum principle follow.

2. Flows, trajectories, and generalized differentials of flows.

2.1. *State space bundles and their sections*

A ***time set*** is a nonempty totally ordered set. If I is a time set, we define $I^{2,\geq} = \{(t,s) \in I \times I : t \geq s\}$, and $I^{3,\geq} = \{(t,s,r) \in I \times I \times I : t \geq s \geq r\}$. A ***state-space bundle*** (abbr. SSB) ***over*** I is an indexed family $\mathbf{X} = \{X_t\}_{t \in I}$ of sets. A ***state-space bundle*** is a pair $\mathcal{X} = (\mathbf{X}, I)$ such that I is a nonempty totally ordered set and $\mathbf{X}$ is an SSB over I. The set I is the ***time set*** of the SSB $\mathcal{X}$.

Remark 2.1. There are several reasons for using general totally ordered sets, rather than real intervals, as time sets for control systems. For a simple example, cf. Ref. 8, where an example is given of a problem for which the natural time set consists of a compact interval minus one interior point. □

If $\mathcal{C}$ is a category whose objects are sets with some additional structure (for example, topological spaces, metric spaces, manifolds of class C^k, linear spaces, FDRLSs), then an SSB $(\mathbf{X}, I) = (\{X_t\}_{t \in I}, I)$ is a ***bundle of*** $\mathcal{C}$***-objects*** if each X_t is a member of $\mathcal{C}$. In particular, if k is a nonnegative integer, a C^k ***SSB*** is an SSB of manifolds of class C^k. Also, an ***FDRLS SSB*** is an SSB of finite-dimensional real linear spaces.

Definition 2.1. Assume that $\mathcal{X} = (\mathbf{X}, I) = (\{X_t\}_{t \in I}, I)$ is an SSB. A ***section of*** $\mathcal{X}$ is a single-valued everywhere defined map ξ on I such that $\xi(t) \in X_t$ for every $t \in I$. We use $Sec(\mathcal{X})$ to denote the set of all sections of $\mathcal{X}$. □

Definition 2.2. Let $\mathcal{X} = (\mathbf{X}, I) = (\{X_t\}_{t \in I}, I)$ be a C^1 state-space bundle, and assume that $\xi \in Sec(\mathcal{X})$. The family $\mathbf{T}_\xi \mathcal{X} = \{T_{\xi(t)} X_t\}_{t \in I}$ is the ***tangent bundle*** of $\mathcal{X}$ along ξ. □

Clearly, the tangent bundle $\mathbf{T}_\xi \mathcal{X}$ of a C^1 SSB $\mathcal{X}$ along a section $\xi \in Sec(\mathcal{X})$ is an FDRLS SSB.

2.2. *Flows and trajectories*

Definition 2.3. Assume that $\mathcal{C}$ is a category whose objects are sets with some additional structure, and $\mathcal{X} = (\mathbf{X}, I) = (\{X_t\}_{t\in I}, I)$ is an SSB of $\mathcal{C}$-objects. A ***$\mathcal{C}$-flow*** on $\mathcal{X}$ is an indexed family $\mathbf{f} = \{f_{t,s}\}_{(t,s)\in I^{2,\geq}}$ such that

(1) $f_{t,s}$ is a $\mathcal{C}$-morphism from X_s to X_t whenever $(t,s) \in I^{2,\geq}$;
(2) $f_{t,t}$ is the identity morphism of X_t whenever $t \in I$;
(3) $f_{t,s} \circ f_{s,r} = f_{t,r}$ whenever $(t,s,r) \in I^{3,\geq}$.

A ***$\mathcal{C}$-flow*** is a pair $\mathcal{F} = (\mathcal{X}, \mathbf{f})$ such that $\mathcal{X}$ is an SSB of $\mathcal{C}$-objects and $\mathbf{f}$ is a $\mathcal{C}$-flow on $\mathcal{X}$. □

Example 2.1. If $\mathcal{C}$ is the category whose objects are all the sets, and whose morphisms are the set-valued maps, then a $\mathcal{C}$-flow on an SSB $\mathcal{X}$ will just be called a ***flow*** on $\mathcal{X}$. □

Example 2.2. If $\mathcal{C}$ is the category whose objects are all FDRLSs, and whose morphisms are the linear maps, then a $\mathcal{C}$-flow on an FDRLS SSB $\mathcal{X}$ will be called a ***linear FD flow***. □

Example 2.3. We use $FDCLin$ to denote the category whose objects are all FDRLSs, and whose morphisms are defined as follows: if X, Y are FDRLSs, then the set of morphisms from X to Y is the set $CLin(X,Y)$ of all nonempty compact subsets of $Lin(X,Y)$. (Composition of morphisms is defined in the obvious way: if $\Lambda_1 \in CLin(X,Y)$ and $\Lambda_2 \in CLin(Y,Z)$, then $\Lambda_2 \circ \Lambda_1 \stackrel{\text{def}}{=} \{L_2 \circ L_1 : L_2 \in \Lambda_2,\ L_1 \in \Lambda_1\}$.)

An $FDCLin$-flow is a ***linear FD multiflow***. □

Remark 2.2. It is well known that every time set I can be regarded as a category $cat(I)$, by taking the objects of $cat(I)$ to be the members of I, and the set $Hom_{cat(I)}(a,b)$ of morphisms from $a \in I$ to $b \in I$ to consist of a single object if $a \leq_I b$, and to be empty if $b <_I a$. In terms of this identification, a $\mathcal{C}$-flow with time set I is exactly the same as a functor from $cat(I)$ to $\mathcal{C}$. □

2.2.1. *Comparison of maps and flows*

If f, f' are SVMs, we write $f \preceq f'$ if $\mathrm{So}(f) = \mathrm{So}(f')$, $\mathrm{Ta}(f) = \mathrm{Ta}(f')$, and $\mathrm{Gr}(f) \subseteq \mathrm{Gr}(f')$. If, for $i = 1, 2$, $\mathcal{F}^i = (\mathcal{X}, \mathbf{f}^i)$ are flows on the same SSB $\mathcal{X}$, and $\mathbf{f}^i = \{f^i_{t,s}\}_{(t,s)\in I^{2,\geq}}$, we say that $\mathcal{F}^1$ is a ***subflow*** of $\mathcal{F}^2$, or $\mathcal{F}^2$ is a ***superflow*** of $\mathcal{F}^1$, and write $\mathcal{F}^1 \preceq \mathcal{F}^2$, if $f^1_{t,s} \preceq f^2_{t,s}$ for all $(t,s) \in I^{2,\geq}$.

2.2.2. *Trajectories*

Definition 2.4. Assume that $\mathcal{X} = (\mathbf{X}, I)$ is a state-space bundle, $\mathcal{F} = (\mathcal{X}, \mathbf{f})$ is a flow, $\mathbf{X} = (\{X_t\}_{t\in I}, I)$, and $\mathbf{f} = \{f_{t,s}\}_{(t,s)\in I^{2,\geq}}$. A ***trajectory*** of $\mathcal{F}$ is a section ξ of $\mathcal{X}$ such that $\xi(t)$ belongs to $f_{t,s}(\xi(s))$ whenever $(t,s) \in I^{2,\geq}$.

We use $Traj(\mathcal{F})$ to denote the set of all trajectories of the flow $\mathcal{F}$. □

2.3. *AGDQs of flows along trajectories*

Definition 2.5. Assume that $\mathcal{X} = (\mathbf{X}, I) = (\{X_t\}_{t\in I}, I)$ is a C^1 SSB, $\mathcal{F} = (\mathcal{X}, \mathbf{f})$ is a flow, $\mathbf{f} = \{f_{t,s}\}_{(t,s)\in I^{2,\geq}}$, and $\xi \in Traj(\mathcal{F})$. An ***AGDQ*** of $\mathcal{F}$ along ξ is a linear FD multiflow $\mathbf{g} = \{g_{t,s}\}_{(t,s)\in I^{2,\geq}}$ on the tangent bundle $T_\xi\mathcal{X}$ such that $g_{t,s} \in AGDQ(f_{t,s}; \xi(s), \xi(t); X_s)$ whenever $(t,s) \in I^{2,\geq}$.

Remark 2.3. In view of our definitions, the condition that $\mathbf{g}$ is a linear FD multiflow on $T_\xi\mathcal{X}$ means that

(1) if $(t,s) \in I^{2,\geq}$, then $g_{t,s}$ is a nonempty compact set of linear maps from $T_{\xi(s)}X_s$ to $T_{\xi(t)}X_t$;
(2) $g_{t,t} = \{\mathbb{I}_{T_{\xi(t)}X_t}\}$ whenever $t \in I$;
(3) $g_{t,s} \circ g_{s,r} = g_{t,r}$ whenever $(r,s,t) \in I^{3,\geq}$. □

2.3.1. *Compatible selections*

Definition 2.6. Assume that $\mathbf{g} = \{g_{t,s}\}_{(t,s)\in I^{2,\geq}}$ is a linear FD multiflow on an FDRLS SSB $(\mathcal{Y}, I)$. A ***compatible selection*** of $\mathbf{g}$ is a linear FD flow $\gamma = \{\gamma_{t,s}\}_{(t,s)\in I^{2,\geq}}$ such that $\gamma_{t,s} \in g_{t,s}$ whenever $(t,s) \in I^{2,\geq}$. □

We write $CSel(\mathbf{g})$ to denote the set of all compatible selections of $\mathbf{g}$. Then $CSel(\mathbf{g})$ is a subset of the product space $P_{\mathbf{g}} \stackrel{\text{def}}{=} \prod_{(t,s)\in I^{2,\geq}} g_{t,s}$. Since $P_{\mathbf{g}}$ is a compact space, by Tichonov's theorem, and $CSel(\mathbf{g})$ is a closed subset of $P_{\mathbf{g}}$—because $CSel(\mathbf{g})$ is the set of all $\gamma \in P_{\mathbf{g}}$ that satisfy a collection of equalities involving continuous functions on $P_{\mathbf{g}}$—we can conclude that $CSel(\mathbf{g})$ is compact.

Remark 2.4. In view of our previous definitions, the condition that γ is a linear FD flow means that $\gamma_{t,t} = \mathbb{I}_{T_{\xi(t)}X}$ for each $t \in I$, and $\gamma_{t,s}\gamma_{s,r} = \gamma_{t,r}$ whenever $(t,s,r) \in I^{3,\geq}$. □

2.3.2. *Fields of variational vectors and adjoint covectors*

Definition 2.7. Assume that $\mathbf{g} = \{g_{t,s}\}_{(t,s)\in I^{2,\geq}}$ is a linear FD multiflow on an FDRLS SSB $(\mathcal{Y}, I) = (\{Y_t\}_{t\in I}, I)$. A ***field of variational vectors*** of $\mathbf{g}$ is a selection $I \ni t \mapsto v(t) \in Y_t$ such that $v_t \in g_{t,s}v_s$ whenever $(t,s) \in I^{2,\geq}$.

A ***field of adjoint covectors*** (also called, simply, an ***adjoint covector***, or even an ***adjoint vector***) of $\mathbf{g}$ is a selection $I \ni t \mapsto \omega(t) \in Y_t^\dagger$ of the dual bundle $\mathcal{Y}^\dagger = \{Y_t^\dagger\}_{t\in I}$. such that $\omega_s \in g_{t,s}^\dagger \omega_t$ whenever $(t,s) \in I^{2,\geq}$, where $g_{t,s}^\dagger = \{\gamma^\dagger : \gamma \in g_{t,s}\}$. □

The following result is an easy consequence of the compactness of $CSel(\mathbf{g})$.

Proposition 2.1. *Assume that* $\mathbf{g} = \{g_{t,s}\}_{(t,s)\in I^{2,\geq}}$ *is a linear FD multiflow on an FDRLS SSB* $(\mathcal{Y}, I) = (\{Y_t\}_{t\in I}, I)$. *Assume that* $I \ni t \mapsto v(t) \in Y_t$ *(resp.* $I \ni t \mapsto \omega(t) \in Y_t^\dagger$*) is a selection of* $\mathcal{Y}$ *(resp.* $\mathcal{Y}^\dagger$*). Then* v *is a field of variational vectors (resp.* ω *is a field of adjoint covectors) of* $\mathbf{g}$ *if and only if there exists a compatible selection* $\gamma = \{\gamma_{t,s}\}_{(t,s)\in I^{2,\geq}}$ *of* $\mathbf{g}$ *such that* $v_t = \gamma_{t,s}v_s$ *(resp.* $\omega_s = \gamma_{t,s}^\dagger \omega_t$*) whenever* $(t,s) \in I^{2,\geq}$. □

3. Variations, impulse variations, summability

3.1. *Variations of set-valued maps*

Definition 3.1. Assume that F is a set-valued map and P is a FDRLS. A ***variation of*** F ***with ambient parameter space*** P is a family $V = \{V_p\}_{p\in C}$ such that

(1) C is a closed convex cone in P with nonempty interior;
(2) each V_p is a SVM such that $\mathrm{So}(V_p) = \mathrm{So}(F)$ and $\mathrm{Ta}(V_p) = \mathrm{Ta}(F)$;
(3) $\mathrm{Gr}(V_0) \subseteq \mathrm{Gr}(F)$.

If F' is another set-valued map such that $\mathrm{So}(F') = \mathrm{So}(F)$, $\mathrm{Ta}(F') = \mathrm{Ta}(F)$, and $\mathrm{Gr}(F) \subseteq \mathrm{Gr}(F')$, we say that V is a ***variation in*** F' if the inclusion $\mathrm{Gr}(V_p) \subseteq \mathrm{Gr}(F')$ holds for every $p \in C$, i.e., if $V_p(x) \subseteq F'(x)$ whenever $p \in C$ and $x \in \mathrm{So}(F')$. □

If F, P, V are as in Definition 3.1, then the cone C is the ***parameter cone of*** V, and the dimension of C (or of P) is the ***number of parameters*** of V. We will use $\tilde{V}$ to denote the SVM with source $P \times \mathrm{So}(F)$ and target $\mathrm{Ta}(F)$ such that $\tilde{V}(p,x) = V_p(x)$ for all $p \in P$, $x \in \mathrm{So}(V_0)$. (In particular, $\tilde{V}(p,x) = \emptyset$ if $p \in P \backslash C$.)

3.2. *Infinitesimal impulse variations*

Definition 3.2. Assume that $\mathcal{X} = (\mathbf{X}, I) = (\{X_t\}_{t\in I}, I)$ is a C^1 state-space bundle, $\mathcal{F} = (\mathcal{X}, \mathbf{f})$ is a flow, and $\xi \in Traj(\mathcal{F})$. An ***infinitesimal impulse variation*** (abbr, IIV) ***for*** $(\mathcal{F}, \xi)$ is a triple (v, t, σ) such that $t \in I$, $v \in T_{\xi(t)}X_t$, and σ is one of the symbols $+$, $-$. □

Remark 3.1. The purpose of including σ in the above definition is to distinguish between "left" impulse variations, which will be labelled $(v, t, -)$, and "right" impulse variations, labelled $(v, t, +)$. Left and right impulse variations will differ in the way the concept of "carrier" of an IIV (v, t, σ) is defined, which will depend strongly on σ. □

3.3. *Summability*

Definition 3.3. Assume that $\mathcal{X} = (\mathbf{X}, I) = (\{X_t\}_{t\in I}, I)$ is a C^1 state-space bundle, $\mathcal{F} = (\mathcal{X}, \mathbf{f})$ is a flow, and $\xi \in Traj(\mathcal{F})$. If (v, t, σ) is an IIV for $(\mathcal{F}, \xi)$, we say that (v, t, σ) is ***carried*** by a subinterval J of I if $t \in J$ and one of the following two conditions holds: (i) $\sigma = +$ and there exists a $t_* \in J$ such that $t < t_*$, (ii) $\sigma = -$ and there exists a $t_* \in J$ such that $t_* < t$.

If $\mathbf{V}$ is a set of IIVs for $(\mathcal{F}, \xi)$, we say that $\mathbf{V}$ is ***carried*** by J if every member of $\mathbf{V}$ is carried by J. □

If $\mathbf{V}$ is a finite set of IIVs for $(\mathcal{F}, \xi)$, we let $\mathbb{R}^{\mathbf{V}}$, $\mathbb{R}^{\mathbf{V}}_+$ denote, respectively, the set of all families $\vec{p} = \{p^V\}_{V\in\mathbf{V}}$ of real numbers, and the set of all $\vec{p} = \{p^V\}_{V\in\mathbf{V}} \in \mathbb{R}^{\mathbf{V}}$ such that $p^V \geq 0$ for all $V \in \mathbf{V}$. (Hence, if m is the cardinality of $\mathbf{V}$, and $\mathbf{V} = \{(v^1, t^1, \sigma^1), \ldots, (v^m, t^m, \sigma^m)\}$, the spaces $\mathbb{R}^{\mathbf{V}}$, $\mathbb{R}^{\mathbf{V}}_+$, can be identified with $\mathbb{R}^m$, $\mathbb{R}^m_+$, by identifying each family $\vec{p} = \{p^V\}_{\varepsilon\in\mathbf{V}}$ with the m-tuple $(\tilde{p}^1, \ldots, \tilde{p}^m)$, where $\tilde{p}^j = p^{(v^j,t^j,\sigma^j)}$ for $j = 1, \ldots, m$.)

If $\mathbf{g} = \{g_{t,s}\}_{(t,s)\in I^{2,\geq}}$ is an AGDQ of $\mathcal{F}$ along ξ, $\gamma = \{\gamma_{t,s}\}_{(t,s)\in I^{2,\geq}}$ is a compatible selection of $\mathbf{g}$, $a, b \in I$, $a \leq b$, and $\mathbf{V}$ is carried by $[a, b]$, we define a linear map $L^{\mathbf{V},\gamma,a,b} : \mathbb{R}^{\mathbf{V}} \times T_{\xi(a)}X_a \mapsto T_{\xi(b)}X_b$ by letting

$$L^{\mathbf{V},\gamma,a,b}(\vec{p}, w) = \gamma_{b,a}(w) + \sum_{(v,t,\sigma)\in\mathbf{V}} p^{(v,t,\sigma)}\gamma_{b,t}(v) \ \text{ for } \vec{p}\in\mathbb{R}^{\mathbf{V}},\ w\in T_{\xi(a)}X_a.$$

We let $\Lambda^{\mathbf{V},\mathbf{g},a,b}$ be the set of all the maps $L^{\mathbf{V},\gamma,a,b}$, for all $\gamma \in CSel(\mathbf{g})$. Then $\Lambda^{\mathbf{V},\mathbf{g},a,b}$ is the image of $CSel(\mathbf{g})$ under the continuous map $CSel(\mathbf{g}) \ni \gamma \mapsto L^{\mathbf{V},\gamma,a,b} \in Lin(\mathbb{R}^{\mathbf{V}} \times T_{\xi(a)}X_a, T_{\xi(b)}X_b)$, so $\Lambda^{\mathbf{V},\mathbf{g},a,b}$ is a compact subset of $Lin(\mathbb{R}^{\mathbf{V}} \times T_{\xi(a)}X_a, T_{\xi(b)}X_b)$.

Definition 3.4. Assume that $\mathcal{X} = (\mathbf{X}, I) = (\{X_t\}_{t\in I}, I)$ is a C^1 state-space bundle, $\mathcal{F} = (\mathcal{X}, \mathbf{f})$ is a flow, $\xi \in Traj(\mathcal{F})$, $\mathbf{g} = \{g_{t,s}\}_{(t,s)\in I^{2,\geq}}$ is an AGDQ of $\mathcal{F}$ along ξ, and $\mathcal{F}' = (\mathcal{X}, \mathbf{f}') = (\mathcal{X}, \{f'_{t,s}\}_{(t,s)\in I^{2,\geq}})$ is a superflow of $\mathcal{F}$. Let $\mathcal{V}$ be a set of IIVs for $(\mathcal{F}, \xi)$. We say that $\mathcal{V}$ is **g-*AGDQ-summable within*** $\mathcal{F}'$ if the following is true:

- *for every finite subset* $\mathbf{V}$ *of* $\mathcal{V}$*, and every pair* $(a, b) \in I \times I$ *such that* $a < b$ *and* $\mathbf{V}$ *is carried by the closed interval* $[a, b]$*, there exists a variation* $W = \{W_{\vec{p}}\}_{\vec{p}\in\mathbb{R}_+^{\mathbf{V}}}$ *of* $f_{b,a}$ *in* $f'_{b,a}$ *such that the set* $\Lambda^{\mathbf{V},\mathbf{g},a,b}$ *is an AGDQ of the map* $\tilde{W}$ *at* $((0, \xi(a)), \xi(b))$ *along* $\mathbb{R}_+^{\mathbf{V}} \times X_a$*.* □

4. The AGDQ maximum principle

We now state and prove a general maximum principle in the setting of AGDQ theory. Instead of working with a control system $\dot{x} = f(x, u, t)$ and a reference trajectory-control pair (ξ_*, η_*), we consider the more general situation of a pair $(\mathcal{F}, \mathcal{F}')$ of flows such that $\mathcal{F}$ is a subflow of $\mathcal{F}'$. We assume that $\mathcal{F}$ and $\mathcal{F}'$ are defined on a common state space bundle $\mathcal{X} = (\mathbf{X}, I) = (\{X_t\}_{t\in I}, I)$, which is of class C^1, in the sense that the X_t are manifolds of class C^1.

In the control system case, (i) the time set I is a compact subinterval of $\mathbb{R}$, (ii) all the state spaces X_t coincide, so there is a manifold X of class C^1 such that $X_t = X$ for all $t \in I$, (iii) the domain of the reference control η_* is I, (iv) $\mathcal{F} = (\mathcal{X}, \mathbf{f})$ is the reference flow, i.e., the flow determined by the reference control η_*, so that, if $\mathbf{f} = \{f_{t,s}\}_{(t,s)\in I^{2,\geq}}$, then $f_{t,s}(x)$, for $x \in X$, $(t, s) \in I^{2,\geq}$, is the set given by

$$f_{t,s}(x) = \{\xi(t) : \xi \in \mathrm{Traj}(\eta_*, f, s, t)\,,\, \xi(s) = x\}\,,$$

where, if $\mathcal{U}$ is the class of admissible controls, then for any $\eta \in \mathcal{U}$ we use $\mathrm{Traj}(\eta, f, s, t)$ to denote the set of all $\xi \in W^{1,1}([s, t], X)$ such that $\dot{\xi}(\tau) = f(\xi(\tau), \eta(\tau), \tau)$ for a. e. $\tau \in [s, t]$, and $W^{1,1}([s, t], X)$ is the set of all absolutely continuous maps from $[s, t]$ to X, (v) $\mathcal{F}' = (\mathcal{X}, \mathbf{f}')$ is the flow of the full control system, so that, if $\mathbf{f}' = \{f'_{t,s}\}_{(t,s)\in I^{2,\geq}}$, then $f'_{t,s}(x)$, for $x \in X$ and $(t, s) \in I^{2,\geq}$, is the reachable set from x over the interval $[s, t]$, so $f'_{t,s}(x)$ is given by

$$f'_{t,s}(x) = \{\xi(t) : (\exists \eta \in \mathcal{U})(\xi \in \mathrm{Traj}(\eta, f, s, t))\,,\, \xi(s) = x\}\,.$$

(Notice that the maps $f_{t,s}$ are single-valued—that is, each set $f_{t,s}(x)$ is either empty or consists of a single member—if the ordinary differential equation $\dot{x} = f(x, \eta_*(t), t)$ has uniqueness of trajectories, but for more general reference vector fields $(x,t) \mapsto f(x,\eta_*(t),t)$ the $f_{t,s}$ can be set-valued. On the other hand, the $f'_{t,s}$ are never single-valued, except in trivial cases.)

The flow formulation, together with the use of general totally ordered sets rather than real intervals (cf. also Remark 2.1), includes situations other than that of control systems, such as, for example, "hybrid systems" in which the state is allowed to jump at some time t from a state space X_- to a state space X_+. (This is achieved by treating $t-$ and $t+$ as different times, with $t- < t+$, and having a family $\{J_\alpha\}_{\alpha\in A}$ of—possibly set-valued—jump maps from X_- to X_+, one of which is the reference jump map J_{α_*}. In that case, $f_{t+,t-}$ is the map J_{α_*}, and $f'_{t+,t-}$ is the map such that $f'_{t+,t-}(x) = \cup_{\alpha\in A} J_\alpha(x)$.)

Theorem 4.1. *Assume that $\mathcal{X} = (\mathbf{X}, I) = (\{X_t\}_{t\in I}, I)$ is a C^1 state-space bundle, $\mathcal{F} = (\mathcal{X},\mathbf{f})$ is a flow, $\mathcal{F}' = (\mathcal{X},\mathbf{f}')$ is a superflow of $\mathcal{F}$, $\mathbf{f} = \{f_{t,s}\}_{(t,s)\in I^{2,\geq}}$, $\mathbf{f}' = \{f'_{t,s}\}_{(t,s)\in I^{2,\geq}}$, $\xi \in Traj(\mathcal{F})$, and $\mathbf{g} = \{g_{t,s}\}_{(t,s)\in I^{2,\geq}}$ is an AGDQ of $\mathcal{F}$ along ξ. Let $\mathbf{V}$ be a set of infinitesimal impulse variations for $(\mathcal{F},\xi)$ which is $\mathbf{g}$-AGDQ-summable within $\mathcal{F}'$. Let $a, b \in I$ be such that $a < b$, and let S be a subset of X_b such that $\xi(b) \in S$. Let $\mathcal{C}$ be an AGDQ-approximating multicone of S at $\xi(b)$. Assume that $f'_{b,a}(\xi(a)) \cap S = \{\xi(b)\}$. Then for every nonzero linear functional μ on $T_{\xi(b)}X_b$ there exist (i) a compatible selection $\gamma = \{\gamma_{t,s}\}_{a\leq s\leq t\leq b}$ of $\mathbf{g}$, (ii) covectors $\bar\pi, \tilde\pi \in T^*_{\xi(b)}X_b$, and (iii) a nonnegative real number π_0, such that $\pi_0\mu = \bar\pi + \tilde\pi$, $(\pi_0, \bar\pi, \tilde\pi) \neq (0,0,0,)$, $\tilde\pi \in \mathcal{C}^\dagger$, and $\pi(t)\cdot v \leq 0$ for every $(v,t,\sigma) \in \mathcal{V}$ which is carried by $[a,b]$, where $\pi(t) = \bar\pi \circ \gamma_{b,t}$ for $a \leq t \leq b$.*

Proof. Fix a $\mu \in T^*_{\xi(b)}X_b\backslash\{0\}$. Let $\mathbf{V}_0$ be a finite subset of $\mathbf{V}$. Using the summability of $\mathbf{V}$, pick a variation $\{W_{\vec p}\}_{\vec p\in\mathbb{R}_+^{\mathbf{V}_0}}$ of $f_{b,a}$ in $f'_{b,a}$ such that the set $\Lambda^{\mathbf{V}_0,\mathbf{g},a,b}$ is an AGDQ of the map $\tilde W$ at $((0,\xi(a)),\xi(b))$ along $\mathbb{R}_+^{\mathbf{V}_0} \times X_a$. For each compatible selection γ of $\mathbf{g}$, let $\hat L^\gamma$ be the linear map $\mathbb{R}^{\mathbf{V}_0} \ni \vec p \mapsto L^{\mathbf{V}_0,\gamma,a,b}(\vec p, 0)$, so that $\hat L(\vec p) = \sum_{(v,t,\sigma)\in\mathbf{V}_0} p^{(v,t,\sigma)}\gamma_{b,t}(v)$. Let $\hat\Lambda = \{\hat L^\gamma : \gamma \in CSel(\mathbf{g})\}$. Then $\hat\Lambda$ is an AGDQ of the set-valued map $\mathbb{R}^{\mathbf{V}_0} \ni \vec p \mapsto \tilde W(\vec p,\xi(a)) \subseteq X_b$ at $(0,\xi(b))$ in the direction of $\mathbb{R}_+^{\mathbf{V}_0}$. Since $\tilde W(\vec p,\xi(a)) \subseteq f'_{b,a}(\xi(a))$, the set $\mathcal{M} = \{\hat L^\gamma \cdot \mathbb{R}_+^{\mathbf{V}_0} : \gamma \in CSel(\mathbf{g})\}$ is an AGDQ-approximating multicone of the set $f'_{b,a}(\xi(a))$ at $\xi(b)$. Since $f'_{b,a}(\xi(a)) \cap S = \{\xi(b)\}$, Corollary 1.1 implies that there exists a decomposition $\pi_0\mu = \bar\pi + \tilde\pi$, where $\bar\pi \in M^\dagger$ for some $M \in \mathcal{M}$, $\tilde\pi \in C^\dagger$ for some $C \in \mathcal{C}$, $\pi_0 \geq 0$, and $(\pi_0, \bar\pi, \tilde\pi) \neq (0,0,0)$. Since $M \in \mathcal{M}$, we can pick a $\gamma \in CSel(\mathbf{g})$ such that $M = \hat L^\gamma \cdot \mathbb{R}_+^{\mathbf{V}_0}$. Then the condition that $\bar{\pi} \in M^\dagger$ implies that $\langle \bar{\pi}, \hat{L}^\gamma(\vec{p})\rangle \leq 0$ for every $\vec{p} \in \mathbb{R}_+^{\mathbf{V}_0}$. Therefore $\left\langle \bar{\pi}, \sum_{(v,t,\sigma)\in\mathbf{V}_0} p^{(v,t,\sigma)}\gamma_{b,t}(v)\right\rangle \leq 0$ for every $\vec{p} \in \mathbb{R}_+^{\mathbf{V}_0}$. This implies that $\langle \bar{\pi}, \gamma_{b,t}(v)\rangle \leq 0$—i.e., that $\langle \bar{\pi} \circ \gamma_{b,t}, v\rangle \leq 0$—for every $(v,t,\sigma) \in \mathbf{V}_0$. Furthermore, the fact that $\tilde{\pi} \in C^\dagger$ implies that $\tilde{\pi} \in \mathcal{C}^\dagger$.

It follows that the 4-tuple $(\bar{\pi}, \tilde{\pi}, \pi_0, \gamma)$ satisfies all our conditions, except only for the fact that the inequality $\langle \bar{\pi} \circ \gamma_{b,t}, v\rangle \leq 0$ has only been shown to hold for (v,t,σ) in a finite subset $\mathbf{V}_0$ of $\mathbf{V}$. To prove the existence of a 4-tuple $(\bar{\pi}, \tilde{\pi}, \pi_0, \gamma)$ that satisfies $\langle \bar{\pi} \circ \gamma_{b,t}, v\rangle \leq 0$ for all $(v,t,\sigma) \in \mathbf{V}$, we use a familiar compactness argument. Fix a norm $\|\cdot\|$ on $T^*_{\xi_*(b)}X_b$. Let $\mathcal{Q}$ be the set of all 4-tuples $(\bar{\pi}, \tilde{\pi}, \pi_0, \gamma)$ such that $\bar{\pi} \in T^*_{\xi_*(b)}X_b$, $\tilde{\pi} \in T^*_{\xi_*(b)}X_b$, $\pi_0 \in \mathbb{R}$, $\pi_0 \geq 0$, $\pi_0 + \|\bar{\pi}\| + \|\tilde{\pi}\| = 1$, and $\gamma \in CSel(\mathbf{g})$. Then $\mathcal{Q}$ is a compact topological space, using on $CSel(\mathbf{g})$ the topology induced by the product topology of $\prod_{(t,s)\in I^{2,\geq}} g_{t,s}$. For each subset $\mathbf{U}$ of $\mathbf{V}$, let $\mathcal{Q}^{\mathbf{U}}$ be the set of those $(\bar{\pi}, \tilde{\pi}, \pi_0, \gamma) \in \mathcal{Q}$ such that $\tilde{\pi} \in \mathcal{C}^\dagger$ and $\langle \bar{\pi} \circ \gamma_{b,t}, v\rangle \leq 0$ for all $(v,t,\sigma) \in \mathbf{U}$. Then every $\mathcal{Q}^{\mathbf{U}}$ is compact, and we have shown that $\mathcal{Q}^{\mathbf{U}}$ is nonempty if $\mathbf{U}$ is finite. Furthermore, it is clear that, if $\{\mathbf{U}_j\}_{j\in\{1,\ldots,m\}}$ is a finite family of finite subsets of $\mathbf{V}$, then $\mathcal{Q}^{\mathbf{U}_1}\cap\cdots\cap\mathcal{Q}^{\mathbf{U}_m} = \mathcal{Q}^{\mathbf{U}_1\cup\cdots\cup\mathbf{U}_m}$, so $\mathcal{Q}^{\mathbf{U}_1}\cap\cdots\cap\mathcal{Q}^{\mathbf{U}_m} \neq \emptyset$. If $\mathcal{U}$ is the set of all finite subsets of $\mathbf{V}$, we have shown that every finite intersection of members of the family $\{\mathcal{Q}^{\mathbf{U}}\}_{\mathbf{U}\in\mathcal{U}}$ is nonempty. Therefore the set $\bigcap\{\mathcal{Q}^{\mathbf{U}} : \mathbf{U} \in \mathcal{U}\}$ is nonempty. But $\bigcap\{\mathcal{Q}^{\mathbf{U}} : \mathbf{U} \in \mathcal{U}\} = \mathcal{Q}^{\mathbf{V}}$. So $\mathcal{Q}^{\mathbf{V}}$ is nonempty, concluding our proof. □

5. Generalized Bianchini-Stefani IIVs and the summability theorem

We now present a class of IIVs that are infinitesimal generators of high-order variations in a sense that generalizes the definition proposed by Bianchini and Stefani.

We assume that we are given

(D1) a pair $(\mathcal{F}', \mathcal{F})$ of flows, where

- (D1. i) $\mathcal{F} = (\mathcal{X}, \mathbf{f})$ and $\mathcal{F}' = (\mathcal{X}, \mathbf{f}')$,
- (D1. ii) $\mathcal{X} = (\mathbf{X}, I)] = (\{X_t\}_{t\in I}, I)$ is a C^1 state-space bundle,
- (D1.iii) $\mathbf{f} = \{f_{t,s}\}_{(t,s)\in I^{2,\geq}}$ and $\mathbf{f}' = \{f'_{t,s}\}_{(t,s)\in I^{2,\geq}}$ are flows on the state=space bundle $\mathcal{X}$,
- (D1.iv) $\mathcal{F}$ is a subflow of $\mathcal{F}'$,

(D2) a "reference trajectory" $\xi_* \in Traj(\mathcal{F})$,

(D3) an AGDQ $\mathbf{g} = \{g_{t,s}\}_{(t,s)\in I^{2,\geq}}$ of $\mathcal{F}$ along ξ_*.

5.1. *Times of right and left regularity*

Definition 5.1. Given $\mathcal{F}'$, $\mathcal{F}$, ξ_*, $\mathbf{g}$ as above, a ***time of right*** (resp. ***left***) ***regularity*** of $(\mathcal{F}', \mathcal{F}, \xi_*, \mathbf{g})$ is a time $\bar{t} \in I$ such that there exists a pair (t_*, X) for which

(1) $t_* \in I$ and $\bar{t} <_I t_*$ (resp. $t_* <_I \bar{t}$),
(2) X is a manifold of class C^1,
(3) if we let $J \stackrel{\text{def}}{=} [\min(\bar{t}, t_*), \max(\bar{t}, t_*)]_I$, then
   (3.a) $X_t = X$ for all $t \in J$,
   (3.b) J is[a] a compact subinterval of $\mathbb{R}$,
   (3.c) the map $J \ni t \mapsto \xi_*(t) \in X$ is continuous,
   (3.d) the family $\{g_{t,s}\}_{s,t \in J, s \leq t}$ is a uniform AGDQ of the reference flow maps $f_{t,s}$, for $s, t \in J$, $s \leq t$, at $(\xi_*(s), \xi_*(t))$, in the direction of X,
   (3.e) $\lim_{t \downarrow \bar{t}} \sup\{\|\gamma - \mathbb{I}_{T_{\xi_*(\bar{t})}X}\| : \gamma \in g_{t,s},\ s \in [\bar{t}, t]\} = 0$ (resp. $\lim_{s \uparrow \bar{t}} \sup\{\|\gamma - \mathbb{I}_{T_{\xi_*(\bar{t})}X}\| : \gamma \in g_{t,s},\ t \in [s, \bar{t}]\} = 0$ (cf. Remark 5.1 below). □

Remark 5.1. Condition (3.e) of the above definition is interpreted as follows: let κ be a coordinate chart of X such that, for some $\tilde{t}_* \in J \backslash \{\bar{t}\}$, the interval $\tilde{J} = [\min(\bar{t}, \tilde{t}_*), \max(\bar{t}, \tilde{t}_*)]_I$ is such that $\xi_*(t) \in \mathrm{Do}(\kappa)$ for every $t \in \tilde{J}$ (such a chart exists because of Condition (3.c)); we can then identify all the tangent spaces $T_x X$. for $x \in \mathrm{Do}(\kappa)$, with $\mathbb{R}^{\dim X}$; then, if $s, t \in \tilde{J}$ and $s \leq t$, all the members γ of $g_{t,s}$ are linear maps from $\mathbb{R}^{\dim X}$ to $\mathbb{R}^{\dim X}$, and so is $\mathbb{I}_{T_{\xi_*(\bar{t})}X}$, so the difference $\gamma - \mathbb{I}_{T_{\xi_*(\bar{t})}X}$ and its norm $\|\gamma - \mathbb{I}_{T_{\xi_*(\bar{t})}X}\|$ make sense. □

5.2. *GBS IIVs*

Definition 5.2. Given $\mathcal{F}'$, $\mathcal{F}$, ξ_*, $\mathbf{g}$ as above, and a positive real number λ, a triple $(v, \bar{t}, +)$ such that $\bar{t} \in I$ and $v \in T_{\xi_*(\bar{t})} X_{\bar{t}}$ is a *generalized Bianchini-Stefani (abbr. GBS) right infinitesimal impulse variation of order* $\frac{1}{\lambda}$ *of* $(\mathcal{F}', \mathcal{F}, \xi_*, \mathbf{g})$ *at time* $\bar{t}$ if

(i) $\bar{t}$ is a time of right regularity of $(\mathcal{F}', \mathcal{F}, \xi_*, \mathbf{g})$

[a]We literally mean "is," rather than just "can be identified with." The reason is that, when we consider several impulse variations with the same time $\bar{t}$, we will not want the map identifying a right or left neighborhood of $\bar{t}$ with a real interval to depend on the variation.

(ii) if t_*, X, J are as in Definition 5.1, then there exists a 6-tuple $(\alpha, \beta, \bar{c}, \bar{\varepsilon}, \boldsymbol{\varphi}, \mathcal{N})$ (called a ***generator*** of $(v, \bar{t}, +)$) such that

(ii.1) $0 < \alpha < \beta$, $\bar{c} > 0$, and $\bar{\varepsilon} > 0$,

(ii.2) $\boldsymbol{\varphi} = \{\varphi_{c,\varepsilon}\}_{(c,\varepsilon)\in[0,\bar{c}]\times]0,\bar{\varepsilon}]}$ is a two-parameter family of set-valued maps from X to X,

(ii.3) $\mathcal{N}$ is an open neighborhood of $\xi_*(\bar{t})$ in X,

(ii.4) the set-valued map

$$\mathcal{N} \times [0, \bar{c}] \ni (x, c) \mapsto \varphi_\varepsilon(c, x) \stackrel{\text{def}}{=} \varphi_{c,\varepsilon}(x) \subseteq X$$

is Cellina continuosuly approximable for each $\varepsilon \in]0, \bar{\varepsilon}]$,

(ii.5) the maps φ_ε satisfy

$$\varphi_\varepsilon(0, x) \subseteq f_{\bar{t}+\beta\varepsilon^\lambda, \bar{t}+\alpha\varepsilon^\lambda}(x) \quad \text{for} \quad x \in \mathcal{N}\,, \tag{1}$$

$$\varphi_\varepsilon(c, x) \subseteq f'_{\bar{t}+\beta\varepsilon^\lambda, \bar{t}+\alpha\varepsilon^\lambda}(x) \quad \text{for} \quad (c, x) \in [0, \bar{c}] \times \mathcal{N}\,, \tag{2}$$

as well as the asymptotic conditions

$$\lim_{\varepsilon\downarrow 0, x\to\xi(\bar{t})} \varphi_\varepsilon(c, x) = \xi_*(\bar{t}) \quad \text{u.w.r.t.} \quad c \in [0, \bar{c}]\,, \tag{3}$$

$$\varphi_\varepsilon(c, \xi_*(\bar{t} + \alpha\varepsilon^\lambda) + h) = \xi_*(\bar{t} + \beta\varepsilon^\lambda) + h + \varepsilon\, c\, v + o(\varepsilon + \|h\|)$$
$$\text{as} \quad \varepsilon \downarrow 0, h \to 0\,, \tag{4}$$

(cf. Remarks 5.2, 5.3), where "u.w.r.t." stands for "uniformly with respect to." □

The definition of what it means for a triple $(v, \bar{t}, -)$ to be a *GBS left IIV of order* λ *of* $(\mathcal{F}', \mathcal{F}, \xi_*, \mathbf{g})$ *at time* $\bar{t}$ is similar, with obvious modifications.

Remark 5.2. Equation (3) is interpreted as follows: given any neighborhood U of $\xi_*(\bar{t})$ in X, there exist a positive number ε_* and a neighborhood U' of $\xi_*(\bar{t})$ in X such that $\varphi_\varepsilon(c, x) \subseteq U$ whenever $0 < \varepsilon \le \varepsilon_*$, $x \in U$, and $c \in [0, \bar{c}]$. □

Remark 5.3. In order to interpret Equation (4) precisely, we first agree, for each coordinate chart κ of X near $\xi_*(\bar{t})$ such that $\mathrm{Do}(\kappa) \subseteq \mathcal{N}$, to write x^κ for the coordinate representation $\kappa(x)$ of a point $x \in \mathrm{Do}(\kappa)$, and w^κ for the coordinate representation of a tangent vector $w \in T_xX$ (so that $w^\kappa = \kappa_*(w) = D\kappa(x) \cdot v \in \mathbb{R}^{\dim X}$). Then (3) implies—using Remark 5.2, with $U = \mathrm{Do}(\kappa)$—that there exists a positive number $\varepsilon_* = \varepsilon_*(\kappa, \boldsymbol{\varphi})$ having the following properties:

P1. $0 < \varepsilon_* \le \bar{\varepsilon}$,

P2. $\xi(t) \in \mathrm{Do}(\kappa)$ and $\bar{\mathbb{B}}(\xi_*(t)^\kappa, \varepsilon_*) \subseteq \mathrm{Im}(\kappa)$ whenever $\bar{t} \le t \le \bar{t} + \beta\varepsilon_*^\lambda$,
P3. $\varphi_\varepsilon(c, x) \subseteq \mathrm{Do}(\kappa)$ whenever $\bar{t} \le t \le \bar{t} + \beta\varepsilon_*^\lambda$, $0 < \varepsilon \le \varepsilon_*$, $c \in [0, \bar{c}]$, and $x \in \kappa^{-1}(\bar{\mathbb{B}}(\xi_*(t)^\kappa, \varepsilon_*))$.

We then let $\varphi_\varepsilon^\kappa$ be, for $\varepsilon \in]0, \varepsilon_*]$, the set-valued map from $[0, \bar{c}] \times \bar{\mathbb{B}}(\xi_*(\bar{t})^\kappa, \varepsilon_*)$ to $\mathrm{Im}(\kappa)$ such that $\varphi_\varepsilon^\kappa(c, y) = (\varphi_\varepsilon(c, x))^\kappa$—i.e., $\varphi_\varepsilon^\kappa(c, y) = \{z^\kappa : z \in \varphi_\varepsilon(c, x)\}$—whenever $\varepsilon \in]0, \varepsilon_*]$, $c \in [0, \bar{c}]$, $x \in \mathrm{Do}(\kappa)$, $y \in \bar{\mathbb{B}}(\xi_*(\bar{t})^\kappa, \varepsilon_*)$ are such that $y = x^\kappa$ and $\varphi_\varepsilon(c, x) \subseteq \mathrm{Do}(\kappa)$. We then define the error E^κ by

$$E^\kappa(c, \varepsilon, h, y) = y - \xi_*(\bar{t} + \beta\varepsilon^\lambda)^\kappa - h - \varepsilon c v^\kappa \,,$$

for $y \in \mathrm{Im}(\kappa)$, $\varepsilon \in]0, \varepsilon_*]$, $c \in [0, \bar{c}]$, and $h \in \mathbb{R}^{\dim X}$, and observe that $E^\kappa(c, \varepsilon, h, y)$ belongs to $\mathbb{R}^{\dim X}$.

Then (4) is interpreted as asserting that

$$\lim_{\varepsilon \downarrow 0, h \to 0} \frac{\sup\Big\{ \|E^\kappa(c, \varepsilon, h, y)\| : c \in [0, \bar{c}],\ y \in \varphi_\varepsilon^\kappa\Big(c, \xi_*(\bar{t} + \alpha\varepsilon^\lambda)^\kappa + h\Big) \Big\}}{\varepsilon + \|h\|} = 0 \,.$$

It is easy to see that if this condition holds for some chart κ such that $\mathrm{Do}(\kappa) \subseteq \mathcal{N}$, then it holds for every such chart. □

5.3. *The summability theorem for GBS IIVs*

The following result is then our summability theorem.

Theorem 5.1. *Let $\mathcal{F}', \mathcal{F}, \xi_*, \mathbf{g}$ be data as in (D1-2-3) above. Let $\mathcal{V}$ be the set of all generalized Bianchini-Stefani infinitesimal impulse variations of $(\mathcal{F}', \mathcal{F}, \xi_*, \mathbf{g})$. Then $\mathcal{V}$ is $\mathbf{g}$-AGDQ-summable within $\mathcal{F}'$.*

6. Proof of Theorem 5.1

We have to prove that, if $\mathbf{V}$ is a finite set of GBS IIVs of $(\mathcal{F}', \mathcal{F}, \xi_*, \mathbf{g})$, and a, b are such that $\mathbf{V}$ is carried by $[a, b]$, then there exists a variation $W = \{W_{\vec{p}}\}_{\vec{p} \in \mathbb{R}_+^{\mathbf{V}}}$ of $f_{b,a}$ in $f'_{b,a}$ such that the set $\Lambda^{\mathbf{V}, \mathbf{g}, a, b}$ is an AGDQ of $\tilde{W}$ at $((0, \xi_*(a)), \xi_*(b))$ along $\mathbb{R}_+^{\mathbf{V}} \times X_a$. It clearly suffices to consider the case when $\mathbf{V}$ is a nonempty finite set of GBS right IIVs at a point $\bar{t} \in I$, and to take $a = \bar{t}$.

Since $\mathbf{V}$ is nonempty, $\bar{t}$ is a time of right regularity for $(\mathcal{F}', \mathcal{F}, \xi_*, \mathbf{g})$, so we may pick t_*, X such that the conditions of Definition 5.1 hold. Clearly, we may restrict t_* further, and assume that $t_* \le b$. Furthermore, we may assume that all the points $\xi_*(t)$, for $\bar{t} \leq t \leq t_*$, belong to the domain Ω of a coordinate chart κ of X.

Let the members of $\mathbf{V}$ be listed as $(v_1, \bar{t}, +), \ldots, (v_m, \bar{t}, +)$, in such a way that the inverse orders $\lambda_1, \ldots, \lambda_m$ satisfy $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_{m-1} \geq \lambda_m$. Then pick for each j a 6-tuple $(\alpha_j, \beta_j, \bar{\varepsilon}_j, \bar{c}_j, \boldsymbol{\varphi}^j, \mathcal{N}^j)$ which is a generator of $(v_j, \bar{t}, +)$ in the sense of Definition 5.2. It is then easy to see that

(*) *without loss of generality, we may assume that*

- A1. *all the* $\bar{\varepsilon}_j$ *are equal to a positive number* $\bar{\varepsilon}$ *such that* $\bar{\varepsilon} \leq 1$,
- A2. *all the* $\bar{c}_j$ *are equal to a positive number* $\bar{c}$,
- A3. *the inequalities*

$$\beta_j \varepsilon^{\lambda_j} \leq \alpha_{j+1} \varepsilon^{\lambda_{j+1}} \quad \text{for all } j \in \{1, \ldots, m-1\},\ \varepsilon \in]0, \bar{\varepsilon}]\,. \tag{5}$$

are satisfied,

- A4. *the sets* $\mathcal{N}^j$ *all coincide;*
- A5. $\bar{t} + \beta_m \bar{\varepsilon}^{\lambda_m} \leq t_*$.

To see this, first replace each $\bar{\varepsilon}_j$ by $\min(\bar{\varepsilon}_j, 1)$, so all the $\bar{\varepsilon}^j$ are ≤ 1. Next, pick a particular j. Then if ρ is small enough,

$$\beta_j \rho^{\lambda_j} \varepsilon^{\lambda_j} \leq \alpha_{j+1} \varepsilon^{\lambda_{j+1}} \quad \text{for all } \varepsilon \in]0, \bar{\varepsilon}_j]\,, \tag{6}$$

because (i) $\beta_j \rho^{\lambda_j} \leq \alpha_{j+1}$ for small enough ρ (since λ_j, α_{j+1} and β_j are positive), and then (ii) the inequalities $\beta_j \rho^{\lambda_j} \varepsilon^{\lambda_j} \leq \alpha_{j+1} \varepsilon^{\lambda_j} \leq \alpha_{j+1} \varepsilon^{\lambda_{j+1}}$ hold for $0 < \varepsilon \leq \bar{\varepsilon}_j$, because $\lambda_j \geq \lambda_{j+1}$ and $\bar{\varepsilon}_j \leq 1$. Then we may pick a ρ such that (6) holds, and replace the numbers α_j, β_j, and the family $\boldsymbol{\varphi}^j = \{\varphi^j_{c,\varepsilon}\}_{c \in [0,\bar{c}_j],\, \varepsilon \in]0,\bar{\varepsilon}_j]}$ by the numbers α_j^{new}, β_j^{new} and the family $\boldsymbol{\varphi}^{j,new} = \{\varphi^{j,new}_{c,\varepsilon}\}_{c \in [0,\bar{c}_j^{new}], \varepsilon \in]0,\bar{\varepsilon}_j^{new}]}$, where $\beta_j^{new} = \beta_j \rho^{\lambda_j}$, $\alpha_j^{new} = \alpha_j \rho^{\lambda_j}$, $\bar{c}_j^{new} = \rho \bar{c}_j$, $\bar{\varepsilon}_j^{new} = \min(1, \rho^{-1} \bar{\varepsilon}_j)$, and $\varphi^{j,new}_{c,\varepsilon} = \varphi^j_{\rho^{-1}c, \rho\varepsilon}$ whenever $c \in [0, \bar{c}_j^{new}]$ and $\varepsilon \in]0, \bar{\varepsilon}_j^{new}]$. Then, if we let $\varphi^{j,new}_\varepsilon(c, x) = \varphi^{j,new}_{c,\varepsilon}(x)$, it follows that

$$\varphi^{j,new}_\varepsilon(c, \xi_*(\bar{t} + \alpha_j^{new} \varepsilon^{\lambda_j}) + h) = \xi_*(\bar{t} + \beta_j^{new} \varepsilon^{\lambda_j}) + h + (\rho\varepsilon)(\rho^{-1}c)v + o(\varepsilon + \|h\|)\,,$$

so that

$$\varphi^{j,new}_\varepsilon(c, \xi_*(\bar{t} + \alpha_j^{new} \varepsilon^{\lambda_j}) + h) = \xi_*(\bar{t} + \beta_j^{new} \varepsilon^{\lambda_j}) + h + \varepsilon c v + o(\varepsilon + \|h\|)\,.$$

This means that the 6-tuple $(\alpha_j^{new}, \beta_j^{new}, \bar{c}_j^{new}, \bar{\varepsilon}_j^{new}, \boldsymbol{\varphi}^{j,new}, \mathcal{N}^j)$ is also a generator of $(v_j, \bar{t}, +)$ and, after $(\alpha_j, \beta_j, \bar{c}_j, \bar{\varepsilon}_j, \boldsymbol{\varphi}^j, \mathcal{N}^j)$ is replaced by $(\alpha_j^{new}, \beta_j^{new}, \bar{c}_j^{new}, \bar{\varepsilon}_j^{new}, \boldsymbol{\varphi}^{j,new}, \mathcal{N}^j)$, the desired inequality (5) holds for our particular j. To get the inequality to hold for all j, we just carry out the replacements recursively, starting with $j = m-1$ and moving backwards up to $j = 1$. Finally, when this is finished, we replace all the $\bar{\varepsilon}_j$ by their minimum, and do the same for the $\bar{c}_j$, thus obtaining a new family $\{(\alpha_j, \beta_j, \bar{\varepsilon}_j, \bar{c}_j, \boldsymbol{\varphi}^j, \mathcal{N}^j)\}_{j=1,\dots,m}$ of generators of the $(v_j, \bar{t}, +)$ that satisfy (A1,2,3). To get (A4) and (A5) to hold as well, we let $\mathcal{N}^{new} = \Omega \cap (\cap_{j=1}^m \mathcal{N}^j)$, and replace each $\mathcal{N}^j$ by $\mathcal{N}^{new}$ and each family $\boldsymbol{\varphi}^j$ by the family $\hat{\boldsymbol{\varphi}}^j$ of the restrictions of the φ_ε^j to $[0, \bar{c}] \times \mathcal{N}^{new}$. We then observe that the 6-tuples $(\alpha_j, \beta_j, \bar{\varepsilon}_j, \bar{c}_j, \hat{\boldsymbol{\varphi}}^j, \mathcal{N}^{new})$ are also generators of the $(v_j, \bar{t}, +)$ that satisfy (A1,2,3,4). Finally, we make t_* smaller, if necessary, to guarantee that the set $\{\xi_*(t) : \bar{t} \leq t \leq t_*\}$ is contained in $\mathcal{N}^{new}$, and then make $\bar{\varepsilon}$ smaller, if necessary, to satisfy (A5).

We then use κ to identify Ω with an open subset of $\mathbb{R}^n$—where $n = \dim X$. Then all the tangent spaces $T_x X$, for all $x \in \Omega$, are identified with $\mathbb{R}^n$. Since

$$\lim_{t \downarrow \bar{t}} \sup \Big\{ \|\gamma - \mathbb{I}_{\mathbb{R}^n}\| : \gamma \in g_{t,s},\ s \in [\bar{t}, t] \Big\} = 0\,, \tag{7}$$

we may assume, after making t_* and $\bar{\varepsilon}$ even smaller, that

$$\|\gamma\| \leq 2 \quad \text{whenever } \gamma \in g_{t,s} \text{ and } \bar{t} \leq s \leq t \leq t_*\,. \tag{8}$$

We now use the fact that $\{g_{t,s}\}_{\bar{t} \leq s \leq t \leq t_*}$ is a uniform AGDQ of the maps $f_{t,s}$ at the points $(\xi_*(s), \xi_*(t))$ in the direction of X to choose a function $\theta \in \boldsymbol{\Theta}$ which is an AGDQ modulus for all the 4-tuples $(f_{t,s}, \xi_*(s), \xi_*(t), X)$, for all s, t such that $\bar{t} \leq s \leq t \leq t_*$. We then fix a real number $\tilde{\varepsilon}$ such that (i) $0 < \tilde{\varepsilon} \leq \bar{\varepsilon}$, (ii) $\theta(\tilde{\varepsilon}) \leq 1$, and (iii) the closed ball $\bar{\mathbb{B}}^n(\xi_*(s), \tilde{\varepsilon})$ is contained in Ω for all $s \in [\bar{t}, t_*]$. We then choose, for each $\varepsilon \in]0, \tilde{\varepsilon}]$ and each pair (s, t) such that $\bar{t} \leq s \leq t \leq t_*$, a CCA map $A_{t,s}^\varepsilon : \bar{\mathbb{B}}^n(\xi_*(s), \varepsilon) \mapsto\!\!\!\!\to Aff(\mathbb{R}^n, \mathbb{R}^n)$, taking values in $g_{t,s}^{(\theta(\varepsilon), \varepsilon)}$, such that $\xi_*(t) + A(h) \in f_{t,s}(\xi_*(s) + h)$. whenever $h \in \bar{\mathbb{B}}^n(0, \varepsilon)$ and $A \in A_{t,s}^\varepsilon(\xi_*(s) + h)$. We define set-valued maps $\hat{A}_{t,s}^\varepsilon : \bar{\mathbb{B}}^n(\xi_*(s), \varepsilon) \mapsto\!\!\!\!\to \mathbb{R}^n$, for $\varepsilon \in]0, \tilde{\varepsilon}]$, $\bar{t} \leq s \leq t \leq t_*$, by letting

$$\hat{A}_{t,s}^\varepsilon(\xi_*(s) + h) = \xi(t) + A_{t,s}^\varepsilon(h)(h) \tag{9}$$

(that is, $\hat{A}_{t,s}^\varepsilon(\xi_*(s) + h) = \{\xi(t) + A(h) : A \in A_{t,s}^\varepsilon(\xi_*(s) + h)\}$) for $h \in \bar{\mathbb{B}}^n(0, \varepsilon)$. It is then clear that $\hat{A}_{t,s}^\varepsilon \in CCA(\bar{\mathbb{B}}^n(\xi_*(s), \varepsilon), \mathbb{R}^n)$, and the estimate

$$\|y - \xi_*(t)\| \leq 4\|x - \xi_*(s)\| \quad \text{whenever } x \in \bar{\mathbb{B}}^n(\xi_*(s), \varepsilon)\,,\ y \in \hat{A}_{t,s}^\varepsilon(x) \tag{10}$$

holds. In particular,

$$\hat{A}_{t,s}^\varepsilon\Big(\bar{\mathbb{B}}^n(\xi_*(s), \rho)\Big) \subseteq \bar{\mathbb{B}}^n(\xi_*(t), 4\rho) \subseteq \Omega \quad \text{if } 0 < \rho \leq \varepsilon \leq \frac{\tilde{\varepsilon}}{4}\,. \tag{11}$$

In addition, it is clear that

$$\hat{A}^{\varepsilon}_{t,s}(x) \subseteq f_{t,s}(x) \quad \text{whenever } x \in \bar{\mathbb{B}}^n(\xi_*(s),\varepsilon) \text{ and } 4\varepsilon \leq \tilde{\varepsilon}\,. \tag{12}$$

Next, we pick positive numbers $\varepsilon_{*,j} = \varepsilon_{*,j}(\kappa, \boldsymbol{\varphi}^j)$ that satisfy the properties of Remark 5.3 for the $\boldsymbol{\varphi}^j$, and are such that $\varepsilon_{*,j} \leq \tilde{\varepsilon}$. We let $\varepsilon_* = \min\{\varepsilon_{*,j} : j = 1, \ldots, m\}$. It then follows that

$$\varphi^j_\varepsilon(c,x) \subseteq \Omega \text{ if } \varepsilon \leq \varepsilon_*,\, 0 \leq c \leq \bar{c},\, \bar{t} \leq t \leq \bar{t} + \beta_j \varepsilon_*^{\lambda_j},\, x \in \bar{\mathbb{B}}(\xi_*(t), \varepsilon_*)\,. \tag{13}$$

We then define the errors E_j by

$$E_j(c,\varepsilon,h,y) = y - \xi_*(\bar{t} + \beta\varepsilon^{\lambda}) - h - \varepsilon c v_j\,.$$

for $y \in \mathbb{R}^n$, $\varepsilon \in]0, \varepsilon_{*,j}(\kappa, \boldsymbol{\varphi}^j)]$, $c \in [0,\bar{c}]$, and $h \in \mathbb{R}^n$, so $E_j(c,\varepsilon,h,y) \in \mathbb{R}^n$. We then let $\zeta_*(\varepsilon)$, for $0 < \varepsilon \leq \varepsilon_*$, be the supremum of the numbers $\|E_j(c,\rho,h,y)\|$ taken over all $c \in [0,\bar{c}]$, $j \in \{1,\ldots,m\}$, $h \in \mathbb{R}^n$, $\rho \in]0,\varepsilon]$, such that $\|h\| \leq \varepsilon$, and $y \in \varphi^j(c, \xi_*(\bar{t} + \alpha_j\varepsilon^{\lambda_j}) + h)$. We define $\theta_*(\varepsilon) = \sup\{\rho^{-1}\zeta_*(\rho) : 0 < \rho \leq \varepsilon\}$ for $0 < \varepsilon \leq \varepsilon_*$, and $\theta_*(\varepsilon) = +\infty$ for $\varepsilon > \varepsilon_*$. We then observe that the function θ_* belongs to $\boldsymbol{\Theta}$.

Now, if $\varepsilon \in]0,\varepsilon_*]$, $c \in [0,\bar{c}]$, $0 < \rho \leq \varepsilon$, $x \in \bar{\mathbb{B}}(\xi_*(\bar{t} + \alpha_j\varepsilon^{\lambda_j}),\rho)$, and $y \in \varphi^j_\varepsilon(c,x)$, we have $\|y - \xi_*(\bar{t} + \beta_j\varepsilon^{\lambda_j}) - h - \varepsilon c v_j\| \leq \zeta_*(\varepsilon) \leq \varepsilon\theta_*(\varepsilon)$, where $h = x - \xi_*(\bar{t} + \alpha_j\varepsilon^{\lambda_j})$. Since $\|h\| \leq \rho$, we conclude that

$$\|y - \xi_*(\bar{t} + \beta_j\varepsilon^{\lambda_j})\| \leq \rho + \varepsilon(\bar{c}\|v_j\| + \theta_*(\varepsilon))\,.$$

We fix $\varepsilon_\#$ such that $0 < \varepsilon_\# \leq \varepsilon_*$ and $\theta_*(\varepsilon_\#) \leq 1$, and let $C = \bar{c}\max(\|v_1\|,\ldots,\|v_m\|) + \theta_*(\varepsilon_\#)$. Then

$$\Big(\varepsilon \in]0,\varepsilon_\#] \wedge c \in [0,\bar{c}] \wedge x \in \bar{\mathbb{B}}(\xi_*(\bar{t} + \alpha_j\varepsilon^{\lambda_j}),\rho) \wedge y \in \varphi^j_\varepsilon(c,x)\Big) \Rightarrow \\ \|y - \xi_*(\bar{t} + \beta_j\varepsilon^{\lambda_j})\| \leq \rho + C\varepsilon$$

so

$$\Big(\varepsilon \in]0,\varepsilon_\#] \wedge c \in [0,\bar{c}] \wedge x \in \bar{\mathbb{B}}(\xi_*(\bar{t} + \alpha_j\varepsilon^{\lambda_j}),\rho)\Big) \Rightarrow \\ \varphi^j_\varepsilon(c,x) \subseteq \bar{\mathbb{B}}(\xi_*(\bar{t} + \beta_j\varepsilon^{\lambda_j}),\rho + C\varepsilon)\,. \tag{14}$$

In order to construct our variation W, we first define, for $0 < \varepsilon \leq \varepsilon_\#$,

$$\Xi^0_\varepsilon(x) = \hat{A}^\varepsilon_{\bar{t}+\alpha_1\varepsilon^{\lambda_1},\bar{t}}(x)\,, \quad \Psi^1_\varepsilon(c_1,x) = \varphi^1_\varepsilon(c_1, \Xi^0_\varepsilon(x))$$

(that is, $\Psi^1_\varepsilon(c_1,x) = \bigcup\{\varphi^1_\varepsilon(c_1,y) : y \in \Xi^0_\varepsilon(x)\}$) and then define, recursively,

$$\Xi^j_\varepsilon = \hat{A}^\varepsilon_{\bar{t}+\alpha_{j+1}\varepsilon^{\lambda_{j+1}},\bar{t}+\beta_j\varepsilon^{\lambda_j}}\,,$$

$$\Psi^{j+1}_\varepsilon(c_1,\ldots,c_{j+1},x) = \varphi^{j+1}_\varepsilon\bigg(c_{j+1}, \Xi^j_\varepsilon\Big(\Psi^j_\varepsilon(c_1,\ldots,c_j,x)\Big)\bigg)$$

(that is, $\Psi_\varepsilon^{j+1}(c_1,\ldots,c_{j+1},x)$ is the union of the sets $\varphi_\varepsilon^{j+1}(c_1,\ldots,c_j,y)$ for all $y\in\Xi_\varepsilon^j(\Psi_\varepsilon^j(c_1,\ldots,c_j,x)))$ for $j=1,\ldots,m-1$.

Next, we define

$$\Upsilon_\varepsilon(c_1,\ldots,c_m,x)=\hat{A}^\varepsilon_{t_*,\bar{t}+\beta_m\varepsilon^{\lambda_m}}\Big(\Psi_\varepsilon^m(c_1,\ldots,c_m,x)\Big)\,.$$

Successive applications of (11) and (14) show that

- if $0<\rho$ and $4\rho\le\varepsilon$, then $\Xi_\varepsilon^0\Big(\bar{\mathbb{B}}(\xi_*(\bar{t}),\rho)\Big)\subseteq\bar{\mathbb{B}}(\xi_*(\bar{t}+\alpha_1\varepsilon^{\lambda_1}),4\rho)\,,$
- if $0<\rho$ and $4\rho+C\varepsilon\le\varepsilon$, then

$$\Psi_\varepsilon^1\Big([0,\bar{c}]\times\bar{\mathbb{B}}(\xi_*(\bar{t}),\rho)\Big)\subseteq\bar{\mathbb{B}}\Big(\xi_*(\bar{t}+\beta_1\varepsilon^{\lambda_1}),4\rho+C\varepsilon\Big)\,,$$

- if $0<\rho$ and $16\rho+4C\varepsilon\le\varepsilon$, then

$$\Xi_\varepsilon^1\Big(\bar{\mathbb{B}}(\xi_*(\bar{t}+\beta_1\varepsilon^{\lambda_1}),4\rho+C\varepsilon)\Big)\subseteq\bar{\mathbb{B}}(\xi_*(\bar{t}+\alpha_2\varepsilon^{\lambda_2},16\rho+4C\varepsilon)\,,$$

- if $0<\rho$ and $16\rho+5C\varepsilon\le\varepsilon$, then

$$\Psi_\varepsilon^2\Big([0,\bar{c}]^2\times\bar{\mathbb{B}}(\xi_*(\bar{t}),\rho)\Big)\subseteq\bar{\mathbb{B}}(\xi_*(\bar{t}+\beta_1\varepsilon^{\lambda_1}),16\rho+5C\varepsilon)\,,$$

- if $0<\rho$ and $16\rho+4C\varepsilon\le\varepsilon$, then

$$\Xi_\varepsilon^2\Big(\bar{\mathbb{B}}(\xi_*(\bar{t}+\beta_2\varepsilon^{\lambda_2}),4\rho+C\varepsilon)\Big)\subseteq\bar{\mathbb{B}}(\xi_*(\bar{t}+\alpha_3\varepsilon^{\lambda_3}),64\rho+20C\varepsilon)\,,$$

and so on, so that, for every $j\in\{1,\ldots,m\}$, if we let $G_j=3^{-1}(4^j-1)$, then

- if $0<\rho$ and $4^j\rho+G_jC\varepsilon\le\varepsilon$, then

$$\Psi_\varepsilon^j\Big([0,\bar{c}]^j\times\bar{\mathbb{B}}(\xi_*(\bar{t}),\rho)\Big)\subseteq\bar{\mathbb{B}}(\xi_*(\bar{t}+\beta_j\varepsilon^{\lambda_j}),4^j\rho+G_jC\varepsilon)\,.$$

In particular,

- if $0<\rho$ and $4^m\rho+G_mC\varepsilon\le\varepsilon$, then

$$\Psi_\varepsilon^m\Big([0,\bar{c}]^j\times\bar{\mathbb{B}}(\xi_*(\bar{t}),\rho)\Big)\subseteq\bar{\mathbb{B}}(\xi_*(\bar{t}+\beta_m\varepsilon^{\lambda_m}),4^m\rho+G_mC\varepsilon)\,.$$

We now choose $\varepsilon_\#$ and $\bar{c}$ so that $C\le\frac{1}{2G_m}$, and conclude that

- if $0<4^m\rho\le\frac{\varepsilon}{2}$ and $\varepsilon\le\varepsilon_\#$, then

$$\Psi_\varepsilon^j\Big([0,\bar{c}]^j\times\bar{\mathbb{B}}(\xi_*(\bar{t}),\rho)\Big)\subseteq\bar{\mathbb{B}}(\xi_*(\bar{t}+\beta_j\varepsilon^{\lambda_j}),\varepsilon)\,.$$

From now on, for each $\varepsilon \in]0, \varepsilon_{\#}]$ we fix $\rho = \rho(\varepsilon) = 2^{-1-2m}\varepsilon$, so $0 < 4^m \rho \leq \frac{\varepsilon}{2}$, define $Q^j_\varepsilon = [0, \bar{c}]^j \times \bar{\mathbb{B}}(\xi_*(\bar{t}), \rho)$, and let $\hat{\Psi}^m_\varepsilon$, $\hat{\Upsilon}^m_\varepsilon$, be the restrictions of Ψ^m_ε, Υ_ε, to Q^m_ε. Then $\hat{\Psi}^m_\varepsilon$ and $\hat{\Upsilon}_\varepsilon$ are set-valued maps from Q_ε to $\bar{\mathbb{B}}(\xi_*(\bar{t} + \beta_m \varepsilon^{\lambda_m}), \varepsilon)$ and $\bar{\mathbb{B}}(\xi_*(\bar{t} + \beta_m \varepsilon^{\lambda_m}), 4\varepsilon)$, respectively. Furthermore, $\hat{\Psi}^m_\varepsilon$ and $\hat{\Upsilon}_\varepsilon$ are composites of CCA maps, so

$$\hat{\Psi}^m_\varepsilon \in CCA\Big(Q_\varepsilon, \bar{\mathbb{B}}(\xi_*(\bar{t} + \beta_m \varepsilon^{\lambda_m}), \varepsilon)\Big),$$
$$\hat{\Upsilon}_\varepsilon \in CCA\Big(Q_\varepsilon, \bar{\mathbb{B}}(\xi_*(t_*), 4\varepsilon)\Big).$$

In addition, it is easy to see that

$$\hat{\Upsilon}_\varepsilon(c_1, \ldots, c_m, x) \subseteq f'_{t^*,\bar{t}}(x) \quad \text{whenever } 0 < \varepsilon \leq \varepsilon_{\#}, \ (c_1, \ldots, c_m, x) \in Q_\varepsilon,$$

$$\hat{\Upsilon}_\varepsilon(0, x) \subseteq f_{t^*,\bar{t}}(x) \quad \text{whenever } 0 < \varepsilon \leq \varepsilon_{\#}, \ x \in \bar{\mathbb{B}}(\xi_*(\bar{t}), \rho).$$

We now define $Q^0_\varepsilon = \bar{\mathbb{B}}(\xi_*(\bar{t}), \rho(\varepsilon))$, $\sigma_0 = \bar{t}$. For $j = 1, \ldots, m$, we write

$$\tau_j = \bar{t} + \alpha_j \varepsilon^{\lambda_j}, \quad \sigma_j = \bar{t} + \alpha_j \varepsilon^{\lambda_j}, \quad \vec{c}_j = (c_1, \ldots, c_j), \quad \vec{c}_j \cdot v = c_1 v_1 + \cdots + c_j v_j.$$

Lemma 6.1. *There exists a family* $\mathbf{\Gamma} = \{\Gamma^j_\varepsilon\}_{j=0,1,\ldots,m}$ *of CCA maps* $\Gamma^j_\varepsilon : Q^j_\varepsilon \mapsto\!\!\!\to Aff(\mathbb{R}^n, \mathbb{R}^n)$ *such that if* $(\vec{c}_j, x) \in Q^j_\varepsilon$ *then*

$$\Psi^j_\varepsilon(\vec{c}_j, x) = \xi_*(\sigma_j) + \Gamma^j_\varepsilon(\vec{c}_j, x)(x - \xi_*(\bar{t}) + \varepsilon(\vec{c}_j \cdot v)). \tag{15}$$

Furthermore, $\mathbf{\Gamma}$ *can be chosen so that*

> *(#) there exists a family* $\{\theta_j\}_{j=0,1,\ldots,m}$ *of members of* $\mathbf{\Theta}$ *such that, for each* j, $\Gamma^j_\varepsilon(\vec{c}_j, x) \subseteq g^{(\theta_j(\varepsilon),\varepsilon)}_{\sigma_j,\bar{t}}$ *for all* $(\vec{c}_j, x) \in Q^j_\varepsilon$.

Proof of Lemma 6.1. We define the set-valued maps Γ^j_ε and the functions θ_j, for $j = 0, 1, \ldots, m$, recursively. We first let $\Gamma^0_\varepsilon : Q^0_\varepsilon \mapsto Aff(\mathbb{R}^n, \mathbb{R}^n)$ be the map such that $\Gamma^0_\varepsilon(x) = \mathbb{I}_{\mathbb{R}^n}$ for each x, and take θ_0 to be any member of $\mathbf{\Theta}$ (for example, $\theta_0(\varepsilon) \equiv \varepsilon$).

Next, we carry out the inductive step. We pick a $j \in \{1, \ldots, m\}$ and assume that Γ^{j-1}_ε has been defined. To construct Γ^j_ε, we begin by letting

$$\mathcal{P} = Aff(\mathbb{R}^n, \mathbb{R}^n) \times Aff(\mathbb{R}^n, \mathbb{R}^n) \times Aff(\mathbb{R}^n, \mathbb{R}^n) \times \mathbb{R}^n \times \mathbb{R}^n,$$

and defining $\mathcal{M}^j_\varepsilon$ be the set-valued map from Q^j_ε to $\mathcal{P}$ that sends each point $(\vec{c}_j, x) \in Q^j_\varepsilon$ to the subset $\mathcal{M}^j_\varepsilon(\vec{c}_j, x)$ of $\mathcal{P}$ that consists of all the 5-tuples

(A_0, A_1, A_2, u, w) for which

$$A_0 \in \Gamma_\varepsilon^{j-1}(\vec{c}_{j-1}, x)\,, \tag{16}$$

$$u = \xi_*(\sigma_{j-1}) + A_0(x - \xi_*(\bar{t}) + \varepsilon(\vec{c}_{j-1} \cdot v)) \tag{17}$$

$$A_1 \in A^\varepsilon_{\tau_j, \sigma_{j-1}}(u)\,, \tag{18}$$

$$w = \xi_*(\tau_j) + A_1(u - \xi_*(\sigma_{j-1}))\,, \tag{19}$$

$$A_2 \in A^\varepsilon_{\sigma_j, \tau_j}(w - \xi_*(\tau_j))\,. \tag{20}$$

We then define $\Gamma^j_\varepsilon(\vec{c}_j, x)$ to be the set of all $A \in Aff(\mathbb{R}^n, \mathbb{R}^n)$ such that

$$A = A_2 \circ A_1 \circ A_0 + \mathit{affm}_{0,z}$$

for some $z \in \varphi^j_\varepsilon(c_j, w) - \xi_*(\sigma_j) - (A_2 \circ A_1 \circ A_0)(x - \xi_*(\bar{t}) + \varepsilon(\vec{c}_j \cdot v))$ and some $(A_0, A_1, A_2, u, w) \in \mathcal{M}^j_\varepsilon(\vec{c}_j, x)$. (Recall that if $L \in Lin(\mathbb{R}^p, \mathbb{R}^q)$ and $z \in \mathbb{R}^q$ then $\mathit{affm}_{L,z}$ is the affine map $\mathbb{R}^p \ni x \mapsto L \cdot x + z \in \mathbb{R}^q$.)

If $A \in \Gamma^j_\varepsilon(\vec{c}_j, x)$, then there exist A_0, A_1, A_2, u, w, z, y, w, such that $(A_0, A_1, A_2, u, w) \in \mathcal{M}^j_\varepsilon(\vec{c}_j, x)$, $w = \xi_*(\tau_j) + A_1(u - \xi_*(\sigma_{j-1}))$, $z = y - \xi_*(\sigma_j) - (A_2 \circ A_1 \circ A_0)(x - \xi_*(\bar{t}) + \varepsilon(\vec{c}_j \cdot v))$, $y \in \varphi^j_\varepsilon(c_j, w)$, and $A = A_2 \circ A_1 \circ A_0 + \mathit{affm}_{0,z}$. It follows that

$$\begin{aligned} A(x - \xi_*(\bar{t}) + \varepsilon\vec{c}_j \cdot v) &= A_2(A_1(A_0(x - \xi_*(\bar{t}) + \varepsilon(\vec{c}_j \cdot v)))) + z \\ &= (A_2 \circ A_1 \circ A_0)(x - \xi_*(\bar{t}) + \varepsilon(\vec{c}_j \cdot v)) + y \\ &\quad - \xi_*(\sigma_j) - (A_2 \circ A_1 \circ A_0)(x - \xi_*(\bar{t}) + \varepsilon(\vec{c}_j \cdot v)) \\ &= y - \xi_*(\sigma_j)\,, \end{aligned}$$

so that $A(x - \xi_*(\bar{t}) + \varepsilon(\vec{c}_j \cdot v)) \in \varphi^j_\varepsilon(c_1, w) - \xi_*(\sigma_j)$. Furthermore, since $w = \xi_*(\tau_j) + A_1(u - \xi_*(\sigma_{j-1}))$ and $A_1 \in A^\varepsilon_{\tau_j, \sigma_{j-1}}(u)$, we see that $w \in \Xi^j_\varepsilon(u)$. So $A(x - \xi_*(\bar{t}) + \varepsilon(\vec{c}_j \cdot v)) \in \varphi^j_\varepsilon(c_j, \Xi^j_\varepsilon(u)) - \xi_*(\sigma_j)$. Since $A_0 \in \Gamma^{j-1}_\varepsilon(\vec{c}_{j-1}, x)$, and (15) holds for $j-1$, so that

$$\xi_*(\sigma_{j-1}) + \Gamma^{j-1}_\varepsilon(\vec{c}_{j-1}, x)(x - \xi_*(\bar{t}) + \varepsilon(\vec{c}_{j-1} \cdot v)) = \Psi^{j-1}_\varepsilon(\vec{c}_{j-1}, x)\,,$$

it follows from (17) that $u \in \Psi^{j-1}_\varepsilon(\vec{c}_{j-1}, x)$. Therefore $A(x - \xi_*(\bar{t}) + \varepsilon(\vec{c}_j \cdot v)) \in \varphi^j_\varepsilon\Big(c_j, \Xi^j_\varepsilon(\Psi^{j-1}_\varepsilon(\vec{c}_{j-1}, x))\Big) - \xi_*(\sigma_j)$, so

$$A(x - \xi_*(\bar{t}) + \varepsilon(\vec{c}_j \cdot v)) \in \Psi^j_\varepsilon(\vec{c}_j, x) - \xi_*(\sigma_j)\,,$$

and then $\xi_*(\sigma_j) + A(x - \xi_*(\bar{t}) + \varepsilon(\vec{c}_j \cdot v)) \in \Psi^j_\varepsilon(\vec{c}_j, x)$. Since A is an arbitrary member of $\Gamma^j_\varepsilon(\vec{c}_j, x)$, we conclude that

$$\xi_*(\sigma_j) + \Gamma^j_\varepsilon(\vec{c}_j, x)(x - \xi_*(\bar{t}) + \varepsilon(\vec{c}_j \cdot v)) \subseteq \Psi^j_\varepsilon(\vec{c}_j, x)\,. \tag{21}$$

To prove the opposite inclusion, we pick $y \in \Psi_\varepsilon^j(\vec{c}_j, x)$, and find $u \in \Psi_\varepsilon^{j-1}(\vec{c}_{j-1}, x)$ and $w \in \Xi_\varepsilon^j(u)$ such that $y \in \varphi_\varepsilon^j(c_j, w)$. Since

$$\Xi_\varepsilon^j(u) = \hat{A}_{\tau_j,\sigma_{j-1}}(u) = \xi_*(\tau_j) + A_{\tau_j,\sigma_{j-1}}(u)(u - \xi_*(\sigma_{j-1}))\,,$$

we can find $A_1 \in A_{\tau_j,\sigma_{j-1}}(u)$ such that $w = \xi_*(\tau_j) + A_1(u - \xi_*(\sigma_{j-1}))$. Since $u \in \Psi_\varepsilon^{j-1}(\vec{c}_{j-1}, x)$, the inductive hypothesis (i.e., that (15) holds for $j-1$) implies that we can pick $A_0 \in \Gamma_\varepsilon^{j-1}(\vec{c}_{j-1}, x)$ such that $u = \xi_*(\sigma_{j-1}) + A_0(x - \xi_*(\bar{t}) + \varepsilon(\vec{c}_{j-1}\cdot v))$. Pick an arbitrary member A_2 of $A_{\sigma_j,\tau_j}(w - \xi_*(\tau_j))$. Then the 5-tuple (A_0, A_1, A_2, u, w) belongs to $\mathcal{M}_\varepsilon^j(\vec{c}_j, x)$. Let $z = y - \xi_*(\sigma_j) - (A_2\circ A_1\circ A_0)(x - \xi_*(\bar{t}) + \varepsilon(\vec{c}_j\cdot v))$, and define an affine map A by letting $A = A_2\circ A_1\circ A_0 + \mathit{affm}_{0,z}$. Then A belongs to $\Gamma_\varepsilon^j(\vec{c}_j, x)$, and $y = \xi_*(\sigma_j) + A(x - \xi_*(\bar{t}) + \varepsilon(\vec{c}_j\cdot v))$, so y is a member of $\xi_*(\sigma_j) + \Gamma_\varepsilon^j(\vec{c}_j, x)(x - \xi_*(\bar{t}) + \varepsilon(\vec{c}_j\cdot v))$.

Since y was an arbitrary member of $\Psi_\varepsilon^j(\vec{c}_j, x)$, we have shown that $\Psi_\varepsilon^j(c_1, x) \subseteq \xi_*(\sigma_j) + \Gamma_\varepsilon^j(\vec{c}_j, x)(x - \xi_*(\bar{t}) + \varepsilon(\vec{c}_j\cdot v))$. This fact, together with (21), implies that

$$\Psi_\varepsilon^j(c_1, x) = \xi_*(\sigma_j) + \Gamma_\varepsilon^j(\vec{c}_j, x)(x - \xi_*(\bar{t}) + \varepsilon(\vec{c}_j\cdot v))\,. \tag{22}$$

This completes the inductive construction of the Γ_ε^j, and the proof that (15) holds.

We now prove (#), also by induction. We assume that θ_{j-1} has been defined in such a way that $\theta_{j-1}\in\boldsymbol{\Theta}$ and $(\#_{j-1})$ holds.

Let $A \in \Gamma_\varepsilon^j(\vec{c}_j, x)$. Write $A = A_2\circ A_1\circ A_0 + A_{0,z}$ as before, and let $A_0 = \mathit{affm}_{L_0,z_0}$, $A_1 = \mathit{affm}_{L_1,z_1}$, $A_2 = \mathit{affm}_{L_2,z_2}$. Then $A = \mathit{affm}_{L,\hat{z}}$, where $L = L_2L_1L_0$, $\hat{z} = L_2L_1z_0 + L_2z_1 + z_2 + z$. On the other hand, we know from the inductive hypothesis that $L_0 \in g_{\sigma_{j-1},\bar{t}}^{\theta_{j-1}(\varepsilon)}$ and $\|z_0\| \le \theta_{j-1}(\varepsilon)\varepsilon$, and we also know that $L_1 \in g_{\tau_j,\sigma_{j-1}}^{\theta(\varepsilon)}$, $L_2 \in g_{\sigma_j,\tau_{j-1}}^{\theta(\varepsilon)}$, $\|z_1\| \le \theta(\varepsilon)\varepsilon$, and $\|z_2\| \le \theta(\varepsilon)\varepsilon$. Then (8) implies that $\|L_0\| \le 2 + \theta_{j-1}(\varepsilon)$, $\|L_1\| \le 2+\theta(\varepsilon)$, and $\|L_2\| \le 2 + \theta(\varepsilon)$, so

$$L_2L_1L_0 \in g_{\sigma_j,\bar{t}}^{\tilde{\theta}_j(\varepsilon)}\,,$$

where $\tilde{\theta}_j \in \boldsymbol{\Theta}$. (Precisely, $\tilde{\theta}_j = 8\theta + 4\theta_{j-1} + 4\theta^2 + 8\theta\theta_{j-1} + 3\theta^2\theta_{j-1}$.)

Also, $\|L_2L_1z_0 + L_2z_1 + z_2\| \le \hat{\theta}_j(\varepsilon)\varepsilon$, where $\hat{\theta}_j$ belongs to $\boldsymbol{\Theta}$. (Precisely, $\hat{\theta}_j = 3\theta + 4\theta_{j-1} + \theta^2 + 4\theta\theta_{j-1} + \theta^2\theta_{j-1}$.) As for z, we can estimate it as follows: we have

$$z = y - \xi_*(\sigma_j) - (A_2\circ A_1\circ A_0)(x - \xi_*(\bar{t}) + \varepsilon(\vec{c}_j\cdot v))$$

and also $y = E_j(c_j,\varepsilon,h,y) + \xi_*(\sigma_j) + h + \varepsilon c_jv_j$, where

$$h = w - \xi_*(\tau_j) = (A_1\circ A_0)(x - \xi_*(\bar{t}) + \varepsilon(\vec{c}_{j-1}v))\,. \tag{23}$$

Then $y - \xi_*(\sigma_j) = E_j(c_j, \varepsilon, h, y) + h + \varepsilon c_j v_j$, so

$$\begin{aligned} z &= E_j(c_j, \varepsilon, h, y) + h + \varepsilon_j v_j - (A_2 \circ A_1 \circ A_0)(x - \xi_*(\bar{t}) + \varepsilon(\vec{c}_j \cdot v)) \\ &= E_j(c_j, \varepsilon, h, y) + h + \varepsilon_j v_j \\ &\quad - A_2\Big((A_1 \circ A_0)(x - \xi_*(\bar{t}) + \varepsilon(\vec{c}_{j-1} \cdot v)) + (A_1 \circ A_0)(\varepsilon c_j v_j)\Big) \\ &= E_j(c_j, \varepsilon, h, y) + h + \varepsilon_j v_j - A_2\Big(h + (A_1 \circ A_0)(\varepsilon c_j v_j)\Big) \\ &= E_j(c_j, \varepsilon, h, y) + (\mathbb{I}_{\mathbb{R}^n} - A_2)h + (\mathbb{I}_{\mathbb{R}^n} - (A_2 \circ A_1 \circ A_0))(\varepsilon c_j v_j)\,. \end{aligned}$$

Let $\omega(\varepsilon) = \sup\{\|\mathbb{I}_{\mathbb{R}^n} - L\| : L \in g_{t,s}, \bar{t} \leq s \leq t \leq \bar{t} + \beta_m \varepsilon^{\lambda_m}\}$. Then $\lim_{\varepsilon \downarrow 0} \omega(\varepsilon) = 0$, because of (7). We then have

$$\|(\mathbb{I}_{\mathbb{R}^n} - A_2)h\| = \|(\mathbb{I}_{\mathbb{R}^n} - L_2)h - z_2\| \leq (\omega(\varepsilon) + \theta(\varepsilon))\|h\| + \theta(\varepsilon)\varepsilon\,.$$

Also,

$$\begin{aligned} &\|(\mathbb{I}_{\mathbb{R}^n} - (A_2 \circ A_1 \circ A_0))(\varepsilon c_j v_j)\| \\ &= \|(\mathbb{I}_{\mathbb{R}^n} - (L_2 \circ L_1 \circ L_0))(\varepsilon c_j v_j) - L_2 L_1 z_0 - L_2 z_1 - z_2\|\,. \end{aligned}$$

Furthermore,

$$\begin{aligned} \|\mathbb{I}_{\mathbb{R}^n} - L_2 L_1 L_0\| &= \|\mathbb{I}_{\mathbb{R}^n} - L_2 + L_2 - L_2 L_1 + L_2 L_1 - L_2 L_1 L_0\| \\ &\leq \|\mathbb{I}_{\mathbb{R}^n} - L_2\| + \|L_2\|\,\|\mathbb{I}_{\mathbb{R}^n} - L_1\| + \|L_2\|\|L_1\|\|\mathbb{I}_{\mathbb{R}^n} - L_0\| \\ &\leq \check{\theta}_j(\varepsilon)\,, \end{aligned}$$

where we may take $\check{\theta}_j(\varepsilon) = (2\omega(\varepsilon) + \theta(\varepsilon)) + \theta_{j-1}(\varepsilon))(1 + (1 + \omega(\varepsilon) + \theta(\varepsilon))^2$ (because $\|\mathbb{I}_{\mathbb{R}^n} - L_2\| \leq \omega(\varepsilon) + \theta(\varepsilon)$, $\|\mathbb{I}_{\mathbb{R}^n} - L_1\| \leq \omega(\varepsilon) + \theta(\varepsilon)$, and $\|\mathbb{I}_{\mathbb{R}^n} - L_0\| \leq \omega(\varepsilon) + \theta_{j-1}(\varepsilon)$). Therefore

$$|(\mathbb{I}_{\mathbb{R}^n} - (L_2 \circ L_1 \circ L_0))(\varepsilon c_j v_j)\| \leq \check{\theta}_j(\varepsilon)\bar{c}\|v\|_j \varepsilon\,.$$

Since $\|L_2 L_1 z_0 + L_2 z_1 + z_2\| \leq \hat{\theta}_j(\varepsilon)\varepsilon$, we have

$$\|(\mathbb{I}_{\mathbb{R}^n} - (A_2 \circ A_1 \circ A_0))(\varepsilon c_j v_j)\| \leq (\check{\theta}_j(\varepsilon)\bar{c}\|v\|_j + \hat{\theta}_j(\varepsilon))\varepsilon\,.$$

Finally, $\|E_j(c_j, \varepsilon, h, y)\|$ is bounded by $\theta_*(\max(\varepsilon, \|h\|))(\varepsilon + \|h\|)$ so we get the bound

$$\|z\| \leq \theta_j^{\&}(\varepsilon, \|h\|)(\varepsilon + \|h\|) \tag{24}$$

where $\theta_j^{\&}(\varepsilon, \delta) = \omega(\varepsilon) + 2\theta(\varepsilon) + \check{\theta}_j(\varepsilon)\bar{c}\|v\|_j + \hat{\theta}_j(\varepsilon) + \theta_*(\max(\varepsilon, \delta))$. It then follows that

$$\|\hat{z}\| = \|L_2 L_1 z_0 + L_2 z_1 + z_2 + z\| \leq \hat{\theta}_j(\varepsilon)\varepsilon + \theta_j^{\&}(\varepsilon, \|h|)(\varepsilon + \|h\|)\,. \tag{25}$$

To conclude, we obtain an estimate for $\|h\|$. We use the identity (23), from which it follows that $h = (L_1 L_0)(x - \xi_*(\bar{t}) + \varepsilon(\vec{c}_{j-1} v)) + L_1 z_0 + z_1$. Since

$\|L_0\| \leq 1+\omega(\varepsilon)+\theta_{j-1}(\varepsilon)$, $\|L_1\| \leq 1+\omega(\varepsilon)+\theta(\varepsilon)$, $\|x-\xi_*(\bar{t})\| \leq \varepsilon$, $\|z_0\| \leq \theta_{j-1}(\varepsilon)\varepsilon$, and $\|z_1\| \leq \theta(\varepsilon)\varepsilon$, we find that $\|h\| \leq \eta_j(\varepsilon)\varepsilon$, where

$$\eta_j(\varepsilon)=(1+\omega(\varepsilon)+\theta(\varepsilon))\Big((1+mC)(1+\omega(\varepsilon)+\theta_{j-1}(\varepsilon))+\theta(\varepsilon)+\theta_{j-1}(\varepsilon)\Big).$$

Therefore $\|\hat{z}\| \leq \theta_j^{\$}(\varepsilon)\varepsilon$, where

$$\theta_j^{\$}(\varepsilon)=\hat{\theta}_j(\varepsilon)+\theta_j^{\&}(\varepsilon,\eta_j(\varepsilon)\varepsilon)(1+\eta_j(\varepsilon)). \tag{26}$$

Hence, if we define $\theta_j(\varepsilon)=\max(\tilde{\theta}_j(\varepsilon),\theta_j^{\&}(\varepsilon))$, it is clear that $\theta_j \in \mathbf{\Theta}$, and we have shown that $(\#_j)$ holds, completing the proof of Lemma 6.1.

We now define, for $\vec{p}=(p_1,\ldots,p_m)\in\mathbb{R}_+^m$, $x\in\bar{\mathbb{B}}(\xi_*(\bar{t}),\rho(\varepsilon))$

$$W_{\vec{p}}(x)=\tilde{W}(\vec{p},x)=\{\hat{\Upsilon}_\varepsilon(\varepsilon^{-1}p_1,\ldots,\varepsilon^{-1}p_m,x):\varepsilon\geq\bar{c}^{-1}\max(p_1,\ldots,p_m)\},$$

so each $W_{\vec{p}}$ is a set-valued map, $\mathrm{Gr}(W_0) \subseteq \mathrm{Gr}(f_{t_*,\bar{t}})$, and $\mathrm{Gr}(W_p)\subseteq\mathrm{Gr}(f'_{t_*,\bar{t}})$. Then $W=\{W_{\vec{p}}\}_{\vec{p}\in\mathbb{R}_+^m}$ is a variation of $f_{\tau,\bar{t}}$ in $f'_{\tau,\bar{t}}$. Also, given any positive ε, the map

$$[0,\bar{c}\varepsilon]^m\times\bar{\mathbb{B}}(\xi_*(\bar{t}),\rho(\varepsilon))\ni(\vec{p},x)\mapsto Z_\varepsilon(\vec{p},x)\stackrel{\mathrm{def}}{=}\hat{\Upsilon}_\varepsilon(\varepsilon^{-1}\vec{p},x)$$

is a CCA map whose graph is contained in that of $\tilde{W}$. In addition, if we let $\tilde{\mathcal{Z}}_\varepsilon$ be the set-valued map that sends each point $(\vec{p},x)\in[0,\bar{c}\varepsilon]^m\times\bar{\mathbb{B}}(\xi_*(\bar{t}),\rho(\varepsilon))$ to the set

$$\{(A,B):B\in\Gamma_\varepsilon(\varepsilon^{-1}\vec{p},x)\,,A\in A_{t_*,\sigma_m}(\xi_*(\sigma_m)+B(x-\xi_*(\bar{t})+\vec{p}_m\cdot v))\},$$

(where $\sigma_m = \bar{t}+\beta_m\varepsilon^{\lambda_m}$, as before), and use μ to denote the map $Aff(\mathbb{R}^n,\mathbb{R}^n)\times Aff(\mathbb{R}^n,\mathbb{R}^n)\ni(A,B)\mapsto A\circ B\in Aff(\mathbb{R}^n,\mathbb{R}^n)$, then the composite map $\mathcal{Z}_\varepsilon=\mu\circ\tilde{\mathcal{Z}}_\varepsilon$ is a CCA map such that

$$Z_\varepsilon(\vec{p},x)=\{\xi_*(t_*)+M(x-\xi_*(\bar{t})+\vec{p}\cdot v):M\in\mathcal{Z}_\varepsilon(\vec{p},x)\}.$$

Let us now recall that, if γ is a compatible selection of $\mathbf{g}$, and $\mathbf{V}=\{(v_1,\bar{t},+),\ldots,(v_m,\bar{t},+)\}$, then $L^{\mathbf{V},\gamma,\bar{t},t_*}$ is the linear map $\gamma_{t_*,\bar{t}}\circ\hat{L}$, where $\hat{L}$ is the map $(\vec{p},h)\mapsto h+\vec{p}\cdot v$.

We also recall that $\Lambda^{\mathbf{V},\mathbf{g},\bar{t},t_*}$ is the set of all maps $L^{\mathbf{V},\gamma,\bar{t},t_*}$, for all $\gamma\in CSel(\mathbf{g})$. The definition of Z_ε can be rewritten as

$$Z_\varepsilon(\vec{p},x)=\{\xi_*(t_*)+M(\hat{L}(x-\xi_*(\bar{t}),\vec{p})):M\in\mathcal{Z}_\varepsilon(\vec{p},x)\}.$$

If $M \in \mathcal{Z}_\varepsilon(\vec{p},x)$, then M is the composite of a member B of $\Gamma_\varepsilon^m(\varepsilon^{-1}\vec{p},x)$ followed by a member A of $A_{t_*,\sigma_m}(y)$, where $y=\xi_*(\sigma_m)+B(x-\xi_*(\bar{t})+\vec{p}_m\cdot v)$. If we write $B=\mathit{affm}_{B_1,b_1}$, we know that $B_1\in g_{\sigma_m,\bar{t}}^{\theta_m(\varepsilon)}$ and $\|b_1\|\leq\theta_m(\varepsilon)\varepsilon$. Also, if $A=\mathit{affm}_{A_1,a_1}$, we know that $A_1\in g_{t_*,\sigma_m}^{\theta(\varepsilon)}$ and $\|a_1\|\leq\theta(\varepsilon)\varepsilon$. It follows that, if $M\hat{L}=\mathit{affm}_{K,k}$, then $K = A_1B_1\hat{L}$ and $k = a_1 + A_1b_1$. Therefore $K \in (g_{t_*,\bar{t}} \circ \hat{L})^{\theta^\$(\varepsilon)}$ and $\|k\| \le \theta^\$(\varepsilon)$, where $\theta^\$(\varepsilon)$ is an easily computable member of $\mathbf{\Theta}$. (Precisely, we may take $\theta^\$ = 2(1 + \|\hat{L}\|)(\theta + \theta_m + \theta\theta_m)$.) So $M\hat{L} \in g_{t,s}^{(\theta^\$(\varepsilon),\varepsilon)}$.

Now, the set $g_{t_*,\bar{t}} \circ \hat{L}$ is precisely $\Lambda^{\mathbf{V},\mathbf{g},\bar{t},t_*}$. This shows that $\Lambda^{\mathbf{V},\mathbf{g},\bar{t},t_*}$ is an AGDQ of the map $\tilde{W}$ at $((0,\xi_*(\bar{t})),\xi_*(t_*))$ in the direction of $\mathbb{R}_+^m \times X$. If, for $b \ge t_*$, we define $W^b = \{W_{\vec{p}}^b\}_{\vec{p}\in\mathbb{R}_+^m}$, by letting $W_{\vec{p}}^b(x) = \tilde{W}^b(\vec{p},x) = (f_{b,t_*} \circ W_{\vec{p}})(x))$, then it is clear that W^b is a variation of $f_{b,\bar{t}}$ in $f'_{b,\bar{t}}$, and $\Lambda^{\mathbf{V},\mathbf{g},\bar{t},b}$ is an AGDQ of $\tilde{W}^b$ at $((0,\xi_*(\bar{t})),\xi_*(b))$ in the direction of $\mathbb{R}_+^m \times X$. This completes our proof. □

References

1. Bianchini, R. M., and G. Stefani, "Controllability along a trajectory: a variational approach." *SIAM J. Control and Optim.* **31**, 1993, pp. 900–927.
2. Bianchini, R. M., "Variational approach to some control problems." Banach Center Publications, Vol. 64, 1995.
3. Bianchini, R. M., and G. Stefani, "Good needle-like variations." *Proc. Symposia Pure Math.* **64**, 1999, pp/ 91-101.
4. G. Stefani, "Higher order variations: how can they be defined in order to have good properties?" In *Non-smooth Analysis and Geometry in Control*, B. S. Mordukhovich and H. J. Sussmann, Eds. I.M.A. Volumes in Mathematics and its Applications Vol. 78, 1005.
5. Sussmann, H.J., "Résultats récents sur les courbes optimales." In *15^e Journée Annuelle de la Société Mathématique de France (SMF)*, Publications de la SMF, Paris, 2000, pp. 1-52.
6. Sussmann, H.J., "New theories of set-valued differentials and new versions of the maximum principle of optimal control theory." In *Nonlinear Control in the year 2000*, A. Isidori, F. Lamnabhi-Lagarrigue and W. Respondek Eds., Springer-Verlag, London, 2000, pp. 487-526.
7. Sussmann, H.J., "Combining high-order necessary conditions for optimality with nonsmoothness." In *Proc. 43rd IEEE Conf. Decision and Control, The Atlantis, Paradise Island, the Bahamas, December 14-17, 2004.*
8. Sussmann, H.J., "Optimal control of nonsmooth systems with classically differentiable flow maps." In *Proceedings of the Sixth IFAC Symposium on Nonlinear Control Systems (NOLCOS 2004), Stuttgart, Germany, September 1-3, 2004,* Vol. 2, pp. 609-704.
9. Sussmann, H.J., "Generalized differentials, variational generators, and the maximum principle with state constraints." To appear in *Andrei Agrachev, Stephen Morse, Eduardo Sontag, Hector Sussmann, Vadim Utkin, Nonlinear and Optimal Control Theory. Cetraro, Italy, June 2004* (Fondazione C.I.M.E.), Gianna Stefani and Paolo Nistri Eds., Lect. Notes in Mathematics, Springer-Verlag; 64-page preprint.
10. Sussmann, H.J., "Set separation, approximating multicones, and the Lipschitz maximum principle." To appear in *Journal of Differential Equations.*