Set transversality, approximating multicones, Warga derivate containers and Mordukhovich cones

Héctor J. Sussmann
Department of Mathematics
Rutgers, the State University of New Jersey
Piscataway, NJ 08854-8019, USA
sussmann@math.rutgers.edu

***Abstract*— Smooth and nonsmooth versions of the Pontryagin Maximum Principle can be proved using necessary conditions for set separation in terms of approximating multicones arising from generalized differentiation theories. We propose a notion of approximating multicone derived from J. Warga's theory of derivate containers and the notion of Mordukhovich normal cone, and state and prove the corresponding set separation theorem.**

I. Introduction

In a series of papers (e.g., [1], [2], [3], [4]) we have proposed versions of the Pontryagin Maximum Principle for highly non-smooth systems, based on generalized differentials and flows, and proved by "primal" methods, using packets of needle variations. All these proofs are based on separation theorems for sets, which give a necessary condition for two sets S_1, S_2 containing a point $\bar{s}$ to be separated at $\bar{s}$—in the sense that $S_1 \cap S_2 = \{\bar{s}\}$. The condition involves the notion of "approximating multicone" to a set at a point, and says that, if $\mathcal{C}_1$, $\mathcal{C}_2$ are approximating multicones to S_1, S_2 at $\bar{s}$, then $\mathcal{C}_1$ and $\mathcal{C}_2$ are not "strongly tranversal." The notion of approximating multicone is specific to a particular "generalized differentiation theory." (For example, the classical notion of Boltyanskii approximating cone corresponds to the classical differential.) In our previous papers, this was done for differentiation theories such as the "generalized differential quotients," and the result was used to prove very nonsmooth versions of the Pontryagin principle. In all these versions, the transversality condition turns out to involve some version of the notion of Boltyanskii cone, and does not apply to Clarke tangent cones or Mordukhovich normal cones.

The purpose of this note is to present the analogue of the separation theorem for the differentiation theory of Warga derivate containers (cf. Warga [5], [6], [7]). We define the notion of a "MWAMC" ("Mordukhovich-Warga approximating multicone") to a set at a point, and prove the separation theorem. As was the case for other differentiation theories, the key element of the proof is a directional open mapping theorem, stated and proved in §V. The application of these results to the nonsmooth maximum principle will be discussed in subsequent papers.

Research supported in part by NSF Grant DMS01-03901

II. Preliminary background material

If X, Y are real linear spaces, then $Lin(X,Y)$ will denote the space of all linear maps from X to Y. If X and Y are finite-dimensional and normed, then $Lin(X,Y)$ is also finite-dimensional, and we will always regard it as a normed space, endowed with the operator norm $\|\cdot\|_{op}$ given by $\|L\|_{op} = \sup\{\|L(x)\| : x \in X, \|x\| \le 1\}$. We use $X^\dagger$ to denote the dual space of X, so $X^\dagger = Lin(X,\mathbb{R})$. We use $\mathbb{R}^n$, $\mathbb{R}^{m\times n}$ to denote, respectively, the space of real n-dimensional column vectors, and the space of all real matrices with m rows and n columns. We identify $\mathbb{R}^{m\times n}$ with $Lin(\mathbb{R}^n,\mathbb{R}^m)$ in the usual way, by identifying each matrix $M \in \mathbb{R}^{m\times n}$ with the linear map $\mathbb{R}^n \ni x \mapsto M\cdot x \in \mathbb{R}^m$. If X is any set, then $\mathbb{I}_X$ will denote the identity map of X. If X is a metric space, $\bar{x} \in X$, and $0 < r \in \mathbb{R}$, we use $\bar{\mathbb{B}}_X(\bar{x},r)$, $\mathbb{B}_X(\bar{x},r)$, to denote, respecetively, the closed ball $\{x \in X : \operatorname{dist}(x,\bar{x}) \le r\}$ and the open ball $\{x \in X : \operatorname{dist}(x,\bar{x}) < r\}$. We write $\bar{\mathbb{B}}^n(\bar{x},r)$, $\mathbb{B}^n(\bar{x},r)$ for $\bar{\mathbb{B}}_{\mathbb{R}^n}(\bar{x},r)$, $\mathbb{B}_{\mathbb{R}^n}(\bar{x},r)$.

A ***cone*** in a real linear space X is a nonempty subset C of X such that $rc \in C$ whenever $r \in \mathbb{R}$, $r \ge 0$, and $c \in C$. The ***polar*** of a cone C in X is the set $C^\dagger$ of all $w \in X^\dagger$ such that $\langle w, c\rangle \le 0$ for all $c \in C$. It is clear that $C^\dagger$ is always a closed convex cone, and $C^{\perp\perp}$ is the smallest closed convex cone containing C, so in particular $C^{\perp\perp} = C$ if and only if C is closed and convex. A ***convex multicone*** in X is a nonempty set of convex cones.

Two convex cones C_1, C_2 in a finite-dimensional real linear space X are ***transversal***, if $C_1 - C_2 = X$, i.e., if for every $x \in X$ there exist $c_1 \in C_1$, $c_2 \in C_2$, such that $x = c_1 - c_2$. The cones C_1 and C_2 are ***strongly transversal*** if they are transversal and in addition $C_1 \cap C_2 \ne \{0\}$.

If C_1, C_2 are convex cones in X, then

(1) C_1 and C_2 are transversal if and only if either (i) C_1 and C_2 are strongly transversal or (ii) C_1 and C_2 are linear subspaces and $C_1 \oplus C_2 = X$.
(2) C_1 and C_2 are transversal if and only if $C_1^\dagger \cap (-C_2^\dagger) = \{0\}$.

(To prove (1), it suffices to assume that C_1 and C_2 are transversal but not strongly transversal and show that (ii) holds. Let us prove that C_1 is a linear subspace. Pick $c \in C_1$. Using the transversality of C_1 and C_2 write $-c = c_1 - c_2$, $c_i \in C_i$. Then $c_1 + c = c_2$. But $c_1 + c \in C_1$ and $c_2 \in C_2$.

So $c_1 + c \in C_1 \cap C_2$, and then $c_1 + c = 0$, since C_1 and C_2 are not strongly transversal. Therefore $-c \in C_1$. This shows that $c \in C_1 \Rightarrow -c \in C_1$. So C_1 is a linear subspace. A similar argument shows that C_2 is a linear subspace. Then the transversality of C_1 and C_2 implies that $C_1 + C_2 = X$, and the fact that they are not strongly transversal implies that $C_1 \cap C_2 = \{0\}$. Hence $C_1 \oplus C_2 = X$. To prove (2) observe that $C_1 - C_2$ is a convex cone, so $C_1 - C_2 = X$ if and only if $\mathrm{Clos}(C_1 - C_2) = X$, and if $\mathrm{Clos}(C_1 - C_2) \neq X$ then the Hahn-Banach theorem implies that $C_1^\dagger \cap (-C_2^\dagger) \neq \{0\}$.)

Two convex multicones $\mathcal{C}_1$, $\mathcal{C}_2$ in a finite-dimensional real linear space X are ***transversal***, if C_1 is transversal to C_2 whenever $C_1 \in \mathcal{C}_1$, $C_2 \in \mathcal{C}_2$. The convex multicones $\mathcal{C}_1$, $\mathcal{C}_2$ are ***strongly transversal*** if they are transversal and in addition there exists a $\mu \in X^\dagger \backslash \{0\}$ such that

$$(\forall C_1 \in \mathcal{C}_1,\, C_2 \in \mathcal{C}_2)(\exists x \in C_1 \cap C_2)(\mu(x) > 0)\,. \qquad (1)$$

Let S be a subset of a finite-dimensional real linear space X, and let $\bar{s} \in S$. The ***Bouligand tangent cone*** to S at $\bar{s}$ is the set of all vectors $v \in X$ such that there exist a sequence $\{s_j\}_{j\in\mathbb{N}}$ of points of S converging to $\bar{s}$, and a sequence $\{h_j\}_{j\in\mathbb{N}}$ of positive real numbers converging to 0, such that $v = \lim_{j\to\infty} \frac{s_j - \bar{s}}{h_j}$. We use $T^B_{\bar{s}}S$ to denote the Bouligand tangent cone to S at $\bar{s}$. It is clear, and well known, that $T^B_{\bar{s}}S$ is a closed cone. The ***Bouligand normal cone*** of S at $\bar{s}$ is the polar cone $(T^B_{\bar{s}}S)^\dagger$ of $T^B_{\bar{s}}S$, that is, the set of all covectors $p \in X^\dagger$ such that $\langle p, v\rangle \le 0$ for all $v \in T\bar{s}^B S$. The ***limiting normal cone***, or ***Mordukhovich normal cone*** of S at $\bar{s}$ is the set of all covectors $p \in X^\dagger$ such that $p = \lim_{j\to\infty} p_j$ for some sequence $\{p_j\}_{j\in\mathbb{N}}$ of members of $X^\dagger$ and some sequence $\{s_j\}_{j\in\mathbb{N}}$ of members of S such that $p_j \in (T^B_{s_j}S)^\dagger$ for each j. We use $N^M_{\bar{s}}S$ to denote the Mordukhovich normal cone of S at $\bar{s}$. For each $p \in X^\dagger$, we let $p^\dagger = \{v \in X : \langle p, v\rangle \le 0\}$. The ***Mordukhovich tangent multicone*** to S at $\bar{s}$ is the set

$$\mathcal{MT}_{\bar{s}}S \stackrel{\mathrm{def}}{=} \{p^\dagger : p \in N^M_{\bar{s}}S\}\,.$$

Lemma 2.1: Let S be a closed subset of a finite-dimensional normed real linear space X, and let $\bar{s} \in S$, $\bar{p} \in X^\dagger$. Then the following conditions are equivalent:

$(*.1)$ $\quad \bar{p} \in N^M_{\bar{s}}S\,,$

$(*.2)$ $\quad \liminf_{s\to\bar{s}} \Big(\max\{\langle \bar{p}, v\rangle : v \in T^B_s S,\, \|v\| \le 1\}\Big) = 0\,,$

$(*.3)$ $\quad \liminf_{s\to\bar{s}, p\to\bar{p}} \Big(\max\{\langle p, v\rangle : v \in T^B_s S,\, \|v\| \le 1\}\Big) = 0\,.$

Proof: Although Conditions $(*.2)$ and $(*.3)$ depend on the norm of X, it is easy to see that the truth values of $(*.2)$ and $(*.3)$ are norm-independent. Hence we may assume, without loss of generality, that the norm of X arises from an inner product $\langle\cdot,\cdot\rangle$, and we may use this inner product to identify X and $X^\dagger$ in the standard way.

For $s \in S$, $p \in X$, let

$$\Theta(p, s) = \max\{\langle p, v\rangle : v \in T^B_s S,\, \|v\| \le 1\}\,. \qquad (2)$$

Then $\Theta(p, s) \ge 0$, because $0 \in T^B_s S$.

If $(*.1)$ holds, then we can find a sequence $\{s_j\}_{j\in\mathbb{N}}$ of members of S and a sequence $\{p_j\}_{j\in\mathbb{N}}$ of members of X such that $\lim_{j\to\infty} s_j = \bar{s}$, $\lim_{j\to\infty} p_j = \bar{p}$, and $p_j \in (T^B_{s_j}S)^\dagger$ for each j. Then, $\Theta(p_j, s_j) = 0$ for each j, so (*.3) holds.

We now prove that $(*.3)\Rightarrow(*.2)\Rightarrow(*.1)$. The implication $(*.3)\Rightarrow(*.2)$ is trivial, because if (*.3) holds then there is a sequence $\{(s_j, p_j)\}_{j\in\mathbb{N}}$ of members of $S \times X$ such that $\lim_{j\to\infty} s_j = \bar{s}$, $\lim_{j\to\infty} p_j = \bar{p}$, and $\lim_{j\to\infty} \Theta(p_j, s_j) = 0$. Since $\Theta(\bar{p}, s_j) \le \Theta(p_j, s_j) + \|\bar{p} - p_j\|$, we can conclude that $\lim_{j\to\infty} \Theta(\bar{p}, s_j) = 0$, and then $(*.2)$ holds.

We now assume that $(*.2)$ holds, and prove $(*.1)$. If $\bar{p} = 0$ then $\bar{p} \in N^M_{\bar{s}}S$, so $(*.1)$ is true. So we may assume that $\bar{p} \neq 0$ and then, without loss of generality, we may also assume that $\|\bar{p}\| = 1$. It follows from $(*.2)$ that we can find a sequence $\{s_j\}_{j\in\mathbb{N}}$ of members of S such that $\lim_{j\to\infty} \varepsilon_j = 0$, where $\varepsilon_j = \Theta(\bar{p}, s_j) = 0$. For $\alpha > 0$, $j \in \mathbb{N}$, define $\beta_j(\alpha)$ to be the minimum of all the nonnegative real numbers β such that the closed ball $\bar{\mathbb{B}}_X(s_j + \alpha\bar{p}, \beta)$ intersects S. (The minimum exists because S is closed.) Then $\beta_j(\alpha) \le \alpha$, because

$$s_j \in \bar{\mathbb{B}}_X(s_j + \alpha\bar{p}, \alpha\|\bar{p}\|) = \bar{\mathbb{B}}_X(s_j + \alpha\bar{p}, \alpha)\,.$$

We are going to construct, for each j, a covector p_j which is close to $\bar{p}$ and such that p_j is a Bouligand normal to S at a point $\hat{s}_j$ close to s_j.

Fix a j. If $\beta_j(\alpha) = \alpha$ for some α, then the open ball $\mathbb{B}_X(s_j + \alpha\bar{p}, \alpha)$ does not intersect S, and this clearly implies that $\bar{p} \in (T^B_{s_j}S)^\dagger$. So in this case we take $p_j = \bar{p}$ and $\hat{s}_j = s_j$. Next assume that $\beta_j(\alpha) < \alpha$ for all positive α. Then for each α we can pick a point $\sigma(\alpha) \in \bar{\mathbb{B}}_X(s_j + \alpha\bar{p}, \beta_j(\alpha)) \cap S$. Let $v(\alpha) = \sigma(\alpha) - s_j$, $\pi(\alpha) = \alpha\bar{p} - v(\alpha)$. Then $v(\alpha) \neq 0$, and in addition $\langle v(\alpha), \bar{p}\rangle = \langle v(\alpha) - \alpha\bar{p}, \bar{p}\rangle + \alpha = \alpha - \langle \pi(\alpha), \bar{p}\rangle$, since $\|\bar{p}\| = 1$. Furthermore,

$$\|\pi(\alpha)\| = \|\alpha\bar{p} - v(\alpha)\| = \|(s_j + \alpha\bar{p}) - \sigma(\alpha)\| = \beta_j(\alpha)$$

so that $\langle \pi(\alpha), \bar{p}\rangle \le \beta_j(\alpha)$, and then $\langle v(\alpha), \bar{p}\rangle \ge \alpha - \beta_j(\alpha)$, so that $\beta_j(\alpha) \ge \alpha - \langle v(\alpha), \bar{p}\rangle$. On the other hand, $\limsup_{\alpha\downarrow 0} \|v(\alpha)\|^{-1}\langle v(\alpha), \bar{p}\rangle \le \varepsilon_j$. (Indeed, suppose the inequality is not true. Then there exist a positive δ and a sequence $\{\alpha_k\}_{k\in\mathbb{N}}$ of positive numbers that converges to 0 and is such that $\|v(\alpha_k)\|^{-1}\langle v(\alpha_k), \bar{p}\rangle \ge \varepsilon_j + \delta$. If we let $w_k = \|v(\alpha_k)\|^{-1} v(\alpha_k)$, then we may assume, after passing to a subsequence, that the limit $w = \lim_{k\to\infty} w_k$ exists. Since $s_j + v(\alpha_k) \in S$, the vector w belongs to $T^B_{s_j}S$. But $\langle w, \bar{p}\rangle \ge \varepsilon_j + \delta$, and this contradicts the fact that $\Theta(\bar{p}, s_j) = \varepsilon_j$.)

Let α^* be such that $\|v(\alpha)\|^{-1}\langle v(\alpha), \bar{p}\rangle \le \varepsilon_j + 2^{-j}$ whenever $0 < \alpha \le \alpha^*$. Given any α, it is clear that $\|v(\alpha)\| \le 2\alpha$. Then $0 \le \langle v(\alpha), \bar{p}\rangle \le \alpha\tilde{\varepsilon}_j$ whenever $0 < \alpha \le \alpha^*$, where $\tilde{\varepsilon}_j = 2(\varepsilon_j + 2^{-j})$. Let $a(\alpha) = \langle v(\alpha), \bar{p}\rangle\bar{p}$, $b(\alpha) = v(\alpha) - a(\alpha)$, so $b(\alpha) \perp a(\alpha)$, and then $\|v(\alpha)\|^2 = \|a(\alpha)\|^2 + \|b(\alpha)\|^2$. On the other hand, $\pi(\alpha) = \alpha\bar{p} - v(\alpha) = \alpha\bar{p} - a(\alpha) - b(\alpha)$, so $\pi(\alpha) = (\alpha - \langle v(\alpha), \bar{p}\rangle)\bar{p} - b(\alpha)$, and then

$$\alpha^2 \ge \beta_j(\alpha)^2 = \|\pi(\alpha)\|^2 = |\alpha - \langle v(\alpha), \bar{p}\rangle|^2 + \|b(\alpha)\|^2\,.$$

Since $\langle v(\alpha), \bar{p}\rangle \leq \alpha\tilde{\varepsilon}_j$, we have $\alpha - \langle v(\alpha), \bar{p}\rangle \geq \alpha(1-\tilde{\varepsilon}_j)$, and then $\alpha^2 \geq \alpha^2(1-\tilde{\varepsilon}_j)^2 + \|b(\alpha)\|^2$, so that

$$\|b(\alpha)\|^2 \leq \alpha^2(1-(1-\tilde{\varepsilon}_j)^2) \leq \alpha^2(2\tilde{\varepsilon}_j - \varepsilon_j^2) \leq 2\alpha^2\tilde{\varepsilon}_j \,,$$

and then $\|b(\alpha)\| \leq \alpha\sqrt{2\tilde{\varepsilon}_j}$. Therefore

$$\|\pi(\alpha) - \alpha\bar{p}\| = \|\langle v(\alpha), \bar{p}\rangle)\bar{p} + b(\alpha)\| \leq \alpha\hat{\varepsilon}_j \,,$$

where $\hat{\varepsilon}_j = \tilde{\varepsilon}_j + \sqrt{2\tilde{\varepsilon}_j}$. Hence, if we pick any α such that $0 < \alpha \leq \alpha^*$ and $\alpha \leq 2^{-j-1}$, and let $p_j = \frac{\pi(\alpha)}{\alpha}$, $\hat{s}_j = s_j + v(\alpha)$, we see that $\|p_j - \bar{p}\| \leq \hat{\varepsilon}_j$, $\|\hat{s}_j - s_j\| \leq 2^{-j}$, and p_j is a Bouligand normal to S at $\hat{s}_j$. This shows that $\bar{p}$ is a limiting normal of S at $\bar{s}$, concluding our proof. ∎

If S is closed, the ***Clarke tangent cone*** to S at $\bar{s}$ is the set of all vectors $v \in X$ such that, whenever $\{s_j\}_{j\in\mathbb{N}}$ is a sequence of points of S converging to $\bar{s}$, it follows that there exist Bouligand tangent vectors $v_j \in T^B_{s_j}S$ such that $\lim_{j\to\infty} v_j = v$. We use $T^{Cl}_{\bar{s}}S$ to denote the Clarke tangent cone to S at $\bar{s}$. It is well known that $T^{Cl}_{\bar{s}}S$ is a closed convex cone. Also, it is well-known that $T^{Cl}_{\bar{s}}S$ is the polar of $N^M_{\bar{s}}S$. Therefore $T^{Cl}_{\bar{s}} = \bigcap\{C : C \in \mathcal{MT}_{\bar{s}}S\}$. The ***Clarke normal cone*** of S at $\bar{s}$ is the polar $(T^{Cl}_{\bar{s}}S)^\dagger$ of the Clarke tangent cone, so $(T^{Cl}_{\bar{s}}S)^\dagger$ is the smallest closed convex cone containing $N^M_{\bar{s}}S$.

If X, Y are finite-dimensional real linear spaces, Ω is an open subset of X, $F : \Omega \mapsto Y$ is a map, and $x_* \in \Omega$, a ***Warga derivate container*** of F at x_* is a compact subset Λ of $Lin(X,Y)$ such that for every compact neighborhood $\hat{\Lambda}$ of Λ in $Lin(X,Y)$ there exist an open neighborhood U of x_* in X and a sequence $\{F_j\}_{j\in\mathbb{N}}$ of maps of class C^1 from U to Y such that $F_j \to F$ uniformly on U as $j\to\infty$, and $DF_j(x) \in \Lambda'$ for all $x \in U$, $j \in \mathbb{N}$. It is clear that if F has a Warga derivate container at x_* then F is Lipschitz-continuous on a neighborhood of x_*.

If M, N, are manifolds of class C^1, and $\bar{x} \in M$, then it is easy to extend the concepts of Bouligand and Clarke tangent cone and Mordukhovich tangent multicone, as well as the corresponding normal cones, to a subset S of M at $\bar{x}$, and to define intrinsically the notion of a Warga derivate container at $\bar{x}$ of a map $F : M \mapsto N$. In that case, if $\bar{x} \in S \subseteq M$, then (i) the cones $T^B_{\bar{x}}S$, $T^{Cl}_{\bar{x}}S$, are subsets of the tangent space $T_{\bar{x}}M$ of M at $\bar{x}$, the convex multicone $\mathcal{MT}_{\bar{x}}S$ is a set of convex cones in $T_{\bar{x}}M$, (ii) the cones $(T^B_{\bar{x}}S)^\dagger$, $(T^{Cl}_{\bar{x}}S)^\dagger$, $N^M_{\bar{s}}D$ are subsets of the cotangent space $(T_{\bar{x}}M)^\dagger$, and (iii) the Warga derivate containers of F at $\bar{x}$ are compact subsets of $Lin(T_{\bar{x}}M, T_{F(\bar{x})}N)$.

III. Warga approximating multicones

If $\mathcal{C}$, $\mathcal{D}$ are convex multicones, then we write $\mathcal{C} \preceq \mathcal{D}$ if for every $D \in \mathcal{D}$ there exists a $C \in \mathcal{C}$ such that $C \subseteq D$.

If M is a manifold of class C^1, $\bar{s} \in S \subseteq M$, and $\mathcal{C}$ is a convex multicone in $T_{\bar{s}}M$, we say that $\mathcal{C}$ is a ***Mordukhovich-Warga approximating multicone*** (abbr. MWAMC) of S at $\bar{s}$ if there exist (i) a nonnegative integer n, (ii) a compact subset K of $\mathbb{R}^n$ such that $0 \in K$, (iii) an open neighborhood U of K in $\mathbb{R}^n$, (iv) a Lipschitz-continuous map $F : U \mapsto M$, (v) a compact subset Λ of $Lin(\mathbb{R}^n, T_{\bar{s}}M)$, and (vi) a convex multicone $\mathcal{D}$ in $\mathbb{R}^n$, such that (I) $F(0) = \bar{s}$, (II) $F(K) \subseteq S$, (III) Λ is a Warga derivate container of F at 0, (IV) $\mathcal{D} \preceq \mathcal{MT}_{\bar{s}}K$ and, finally, (V) $\mathcal{C} = \{L\cdot D : L \in \Lambda, D \in \mathcal{D}\}$.

Example 3.1: If S is a closed subset of a manifold M of class C^1, $\bar{s} \in S$, and $\mathcal{C}$ is any convex multicone in $T_{\bar{s}}M$ such that $\mathcal{C} \preceq \mathcal{MT}_{\bar{s}}S$, then $\mathcal{C}$ is a MWAMC of S at $\bar{s}$. Indeed, it clearly suffices to assume that $M = \mathbb{R}^n$ and $\bar{s} = 0$. We then let U, V be, respectively, an open subset of $\mathbb{R}^n$ containing 0, and a compact ball centered at 0 and contained in U. We then take $K = V \cap \tilde{S}$, so K is compact and $\mathcal{MT}_0K = \mathcal{MT}_{\bar{s}}S$. We then let $F : U \mapsto \mathbb{R}^n$ be the inclusion map, and take $\Lambda = \{\mathbb{I}_{\mathbb{R}^n}\}$. Then $\mathcal{C} = \{L\cdot C : L \in \Lambda, C \in \mathcal{C}\}$, and $\mathcal{C} \preceq \mathcal{MT}_0K$. ∎

Example 3.2: As a special case of the previous example, *if S is a closed subset of a manifold M of class C^1, and $\bar{s} \in S$, then the multicones $\mathcal{MT}_{\bar{s}}S$ and $\{T^{Cl}_{\bar{s}}S\}$ are MWAMCs of S at $\bar{s}$.* ∎

Example 3.3: It follows trivially from the definition that, *if M, N are manifolds of class C^1, $S \subseteq M$, $\bar{s} \in S$, $F : M \mapsto N$ is a Lipschitz-continuous map, Λ is a Warga derivate container of F at $\bar{s}$, and $\mathcal{C}$ is a MWAMC of S at $\bar{s}$, then $\Lambda\cdot\mathcal{C} \stackrel{\text{def}}{=} \{L\cdot C : L \in \Lambda, C \in \mathcal{C}\}$ is a MWAMC of $F(S)$ at $F(\bar{s})$.* ∎

Example 3.4: . If M_1, M_2 are manifolds of class C^1, $\bar{s}_1 \in S_1 \subseteq M_1$, $\bar{s}_2 \in S_2 \subseteq M_2$, $\mathcal{C}_1$ is a MWAMC of S_1 at $\bar{s}_1$, and $\mathcal{C}_2$ is a MWAMC of S_2 at $\bar{s}_2$, then $\mathcal{C}_1 \times \mathcal{C}_2$ is a MWAMC of $S_1 \times S_2$ at $(\bar{s}_1, \bar{s}_2)$, where $\mathcal{C}_1 \times \mathcal{C}_2 \stackrel{\text{def}}{=} \{C_1 \times C_2 : C_1 \in \mathcal{C}_1,\, C_2 \in \mathcal{C}_2\}$. To see this find, for $i = 1,2$, a nonnegative integer n_i, a compact subset K_i of $\mathbb{R}^{n_i}$ containing 0, an open neighborhood U_i of 0 in $\mathbb{R}^{n_i}$, a Lipschitz-continuous map $F_i : U_i \mapsto M_i$ such that $F_i(0) = \bar{s}_i$ and $F_i(K_i) \subseteq S_i$, a derivate container Λ_i of F_i at 0, and a convex multicone $\mathcal{D}_i$ in $\mathbb{R}^{n_i}$ such that $\mathcal{D}_i \prec \mathcal{MT}_{\bar{s}_i}K_i$, for which $\mathcal{C}_i = \{L\cdot C : C \in \mathcal{C}_i,\, L \in \Lambda_i\}$. Define $n = n_1 + n_2$, $U = U_1 \times U_2 \subseteq \mathbb{R}^{n_1} \times \mathbb{R}^{n_2} \sim \mathbb{R}^n$, $K = K_1 \times K_2$, $M = M_1 \times M_2$, $S = S_1 \times S_2$, $\bar{s} = (\bar{s}_1, \bar{s}_2)$, $\Lambda = \Lambda_1 \times \Lambda_2$ (that is, $\Lambda = \{L_1 \times L_2 : L_1 \in \Lambda_1,\, L_2 \in \Lambda_2\}$, where $L_1 \times L_2$ is the map that sends $(s_1, s_2) \in \mathbb{R}^{n_1} \times \mathbb{R}^{n_2}$ to the pair $(L_1\cdot s_1, L_2\cdot s_2) \in T_{\bar{s}_1}M_1 \times T_{\bar{s}_2}M_2 \sim T_{\bar{s}}M$), $\mathcal{C} = \mathcal{C}_1 \times \mathcal{C}_2$, $\mathcal{D} = \mathcal{D}_1 \times \mathcal{D}_2$, $F = F_1 \times F_2$. (That is, F is the map that sends $(s_1, s_2) \in U$ to $(F_1(s_1), F_2(s_2)) \in M$.) Then Λ is a derivate container of F at $\bar{s}$, and $\Lambda\cdot\mathcal{D} = \mathcal{C}$. So the desired conclusion will follow if we show that $\mathcal{D} \preceq \mathcal{MT}_{\bar{s}}K$. But this is trivial, because, if $p \in N^M_{\bar{s}}S$, then $p = \lim_{j\to\infty} p_j$ for some sequence $\{(s^j, p^j)\}_{j\in\mathbb{N}}$ such that $s^j \to \bar{s}$, $s_j \in S$, and $p_j \in (T^B_{s^j}K)^\dagger$. If we write $s^j = (s^j_1, s^j_2)$, then $T^B_{s^j}K = T^B_{s^j_1}K_1 \times T^B_{s^j_2}K_2$, so $(T^B_{s^j}K)^\dagger = (T^B_{s^j_1}K_1)^\dagger \times (T^B_{s^j_2}K_2)^\dagger$, and then $p^j = (p^j_1, p^j_2)$, $p^j_1 \in (T^B_{s^j_1}K_1)^\dagger$, $p^j_2 \in (T^B_{s^j_2}K_2)^\dagger$. Hence $p = (p_1, p_2)$, $p_1 \in N^M_{\bar{s}_1}K_1$, $p_2 \in N^M_{\bar{s}_2}K_2$. Since $\mathcal{D}_i \prec \mathcal{MT}_{\bar{s}_i}K_i$ for $i = 1,2$, we may pick $D_i \in \mathcal{D}_i$ such that $D_i \subseteq p_i^\dagger$. Then $D_1 \times D_2 \subseteq p^\dagger$ and $D_1 \times D_2 \in \mathcal{D}$. This shows that $\mathcal{D} \preceq \mathcal{MT}_{\bar{s}}K$ and concludes our proof. ∎

Remark 3.5: In the previous example, it is important to notice that the product $\mathcal{MT}_{\bar{s}_1}K_1 \times \mathcal{MT}_{\bar{s}_2}K_2$ of the

Mordukhovich tangent multicones $\mathcal{MT}_{\bar{s}_1}K_1$, $\mathcal{MT}_{\bar{s}_2}K_2$ does not in general coincide with the Mordukhovich tangent multicone $\mathcal{MT}_{\bar{s}}K$ of the product. On the other hand, it is always true that $\mathcal{MT}_{\bar{s}_1}K_1 \times \mathcal{MT}_{\bar{s}_2}K_2 \preceq \mathcal{MT}_{\bar{s}}K$, and that is all that is needed for the proof in Example 3.4. ■

IV. The transversality theorem

Two subsets S_1, S_2 of a topological space X are ***locally separated*** at a point $p \in X$ if there exists a neighborhood U of p in X such that $S_1 \cap S_2 \cap U = \{p\}$.

Theorem 4.1: Let M be a manifold of class C^1, let S_1, S_2 be subsets of M, and let $\bar{x} \in S_1 \cap S_2$. Let $\mathcal{C}_1$, $\mathcal{C}_2$, be MWAMCs of S_1, S_2 at $\bar{x}$. Assume that $\mathcal{C}_1$ and $\mathcal{C}_2$ are strongly transversal. Then S_1 and S_2 are not locally separated at $\bar{x}$.that is, there exists a sequence $\{x_j\}_{j\in\mathbb{N}}$ of points of $(S_1 \cap S_2)\backslash\{\bar{x}\}$ such that $\lim_{j\to\infty} x_j = \bar{x}$. Furthermore, there exists a Lipschitz arc $\gamma : [0,1] \mapsto M$ such that $\gamma(0) = \bar{x}$, $\gamma(t)$ does not identically equal $\bar{x}$, and $\gamma(t) \in S_1 \cap S_2$ for all $t \in [0,1]$.

Proof: The proof is based on the directional open mapping property, stated and proved in Theorem 5.1 below.

Without loss of generality, we assume that $M = \mathbb{R}^n$ and $\bar{x} = 0$. We let $X = \mathbb{R}^n$, $\mathcal{X} = X \times X$, $\mathcal{Y} = X \times \mathbb{R}$. We fix a linear functional $\mu : X \mapsto \mathbb{R}$ such that (1) holds, and define a map $G : \mathcal{X} = X \times X \mapsto \mathcal{Y}$ by letting $G(x_1, x_2) = (x_1 - x_2, \mu(x_1))$. Then G is a linear map, so the differential $DG(0)$ is just G.

Let $S = S_1 \times S_2$. Also, let $\mathcal{C} = \mathcal{C}_1 \times \mathcal{C}_2$. Then we know from Example 3.4 that $\mathcal{C}$ is a MWAMC of S at $(0,0)$. Let $\mathcal{D} = G \cdot \mathcal{C}$. Then $\mathcal{D}$ is a MWAMC of $G(S)$ at $G(0,0)$.

Let $\bar{y} = (0,1) \in \mathcal{Y} = X \times \mathbb{R}$. Then a straightforward calculation shows that $\bar{y} \in \operatorname{Int} D$ for every $D \in \mathcal{D}$. (*Proof.* Let $D \in \mathcal{D}$, and write $D = G(C_1 \times C_2)$, $C_1 \in \mathcal{C}_1$, $C_2 \in \mathcal{C}_2$. Then $C_1 - C_2 = X$. In view of (1), we can pick $\bar{c} \in C_1 \cap C_2$ such that $\mu(\bar{c}) = 1$. Then $G(\bar{c}, \bar{c}) = \bar{y}$. Given any $v \in X$, we can use the transversality of C_1 and C_2 to write $v = c_1 - c_2$, with $c_1 \in C_1$, $c_2 \in C_2$. So there exists $r \in \mathbb{R}$ such that $(v, r) \in G(C_1 \times C_2)$. If $(e_1, \dots, e_n)$ is a basis of X, and $e_0 = -(e_1 + \dots + e_n)$, then there are reals r_i such that $(e_i, r_i) \in G(C_1 \times C_2)$ for $i = 0, \dots, n$. Since $\bar{y} \in G(C_1 \times C_2)$, it follows that $(e_i, \bar{r}) \in G(C_1 \times C_2)$, for every i, if $\bar{r} = \max(1, r_0, r_1, \dots, r_n)$. Hence $(\tilde{e}_i, 1) \in G(C_1 \times C_2)$ for every i, if $\tilde{e}_i = \bar{r}^{-1} e_i$. This clearly implies our conclusion.)

We have therefore verified the hypotheses of Theorem 5.1. It then follows from the theorem that, for some positive α, there exists a Lipschitz arc $\xi : [0,1] \mapsto S$ such that $\xi(0) = 0$ and $\{G(\xi(t)) : t \in [0,1]\} = \{(0,r) : 0 \le r \le \alpha\}$. Write $\xi(t) = (\xi_1(t), \xi_2(t))$, so $\xi_1(t) \in S_1$ and $\xi_2(t)) \in S_2$. Let $\gamma(t) = \xi_1(t)$. Then, if $t \in [0,1]$, $G(\xi(t)) = (0,r)$ for some r, so $\xi_1(t) = \xi_2(t)$, and then $\gamma(t) \in S_1 \cap S_2$, Furthermore, γ does not vanish identically because, for some $t \in [0,1]$, $G(\xi(t)) = (0, \alpha)$, so $\mu(\gamma(t)) = \alpha$. ■

V. The directional open mapping property

Given a subset A of $\mathbb{R}^\nu$, and a positive number r, we use $\Gamma(A, r)$ to denote the set of all maps $\gamma : [0,1] \mapsto A$ such that $\gamma(0) = 0$ and $\|\gamma(t) - \gamma(s)\| \le r|t - s|$ whenever $s, t \in [0,1]$. (So, naturally, $\Gamma(A, r)$ is empty if $0 \notin A$.) It is then clear that *if A is closed then $\Gamma(A, r)$ is a compact subset of $C^0([0,1], \mathbb{R}^\nu)$*.

If D is a closed convex cone in $\mathbb{R}^\nu$, and $\alpha > 0$, we use $D(\alpha)$ to denote the set $\{y \in D : \|y\| \le \alpha\}$. If $y \in \mathbb{R}^\nu$, we use σ_y to denote the set $\{ty : 0 \le t \le 1\}$. If $\gamma : [0,1] \mapsto A$ is an arc, then $|\gamma|$ will denote the set $\{\gamma(t) : t \in [0,1]\}$.

Theorem 5.1: Assume that

- m and n are nonnegative integers,
- S is a closed subset of $\mathbb{R}^n$ such that $0 \in S$,
- U is an open subset of $\mathbb{R}^n$ such that $0 \in U$,
- $F : U \mapsto \mathbb{R}^m$ is a Lipschitz map such that $F(0) = 0$,
- Λ is a Warga derivate container of F at 0,
- $\bar{y} \in \mathbb{R}^m$ and $\|\bar{y}\| = 1$,
- $\bar{y} \in \operatorname{Int} L \cdot p^\dagger$ for every $L \in \Lambda$. and every $p \in N_0^M S$.

Then there exist a closed convex cone D in $\mathbb{R}^m$ and positive numbers α, κ such that $\bar{y} \in \operatorname{Int} D$ and

$$(\forall y \in D(\alpha))(\exists \gamma \in \Gamma(S, \kappa\alpha))(\sigma_y = |F \circ \gamma|)\,. \qquad (3)$$

Proof: We assume, as we clearly may without loss of generality (after making an orthogonal change of coordinates, if necessary) that $\bar{y} = (0^\mu, 1)$, where $\mu = m - 1$ and 0^μ is the origin of $\mathbb{R}^\mu$. We then let $\mathcal{R} = \mathbb{R}^\mu$, and identify $\mathbb{R}^m$ with $\mathcal{R} \times \mathbb{R}$.

Our first task will be to reformulate our hypothesis in dual, rather than primal terms, by proving that

(#) *There exists a real number $\delta \in]0,1[$ such that, if $q \in \mathbb{R}^m$, $L \in \mathbb{R}^{m\times n}$, $s \in S$ are such that $\|q\| = 1$, $\langle q, \bar{y}\rangle \ge -\delta$, $\operatorname{dist}(L, \Lambda) \le \delta$, $s \in S$, and $\|s\| \le \delta$, then $\Theta(L^\dagger(q), s) \ge \delta$, where Θ is the function defined in (2).*

We prove (#) by contradiction. Assume that δ does not exist. Then there are sequences $\{q_j\}_{j\in\mathbb{N}}$, $\{L_j\}_{j\in\mathbb{N}}$, $\{s_j\}_{j\in\mathbb{N}}$, such that, for each j, the following are true: $q_j \in \mathbb{R}^m$, $\|q_j\| = 1$, $\langle q_j, \bar{y}\rangle \ge -2^{-j}$, $L_j \in \mathbb{R}^{m\times n}$, $\operatorname{dist}(L_j, \Lambda) \le 2^{-j}$, $s_j \in S$, $\|s_j\| \le 2^{-j}$, and $\Theta(L_j^\dagger(q_j), s_j) < 2^{-j}$. Pick $\tilde{L}_j \in \Lambda$ such that $\|\tilde{L}_j - L_j\| \le 2^{-j}$. Then we may find an infinite subset J of $\mathbb{N}$ such that the limits $\bar{q} = \lim_{j\to\infty} q_j$, $\bar{L} = \lim_{j\to\infty} \tilde{L}_j$, exist. Then $\|\bar{q}\| = 1$, $\langle \bar{q}, \bar{y}\rangle \ge 0$, and $\bar{L} \in \Lambda$. In addition, $\lim_{j\to\infty} s_j = 0$ and $\lim_{j\to\infty} L_j = \bar{L}$. Let $p_j = L_j^\dagger q_j$, $\bar{p} = \bar{L}^\dagger \bar{q}$, so $\lim_{j\to\infty} p_j = \bar{p}$. Since $\Theta(p_j, s_j) < 2^{-j}$, it is clear that $\liminf_{s\to 0, p\to\bar{p}} \Theta(p, s) = 0$. So Lemma 2.1 implies that $\bar{p} \in N_0^M S$. Therefore the cone $\bar{L} \cdot \bar{p}^\dagger$ belongs to $\mathcal{C}$. Hence $\bar{y}$ is an interior point of $\bar{L} \cdot \bar{p}^\dagger$. On the other hand, if $y \in \bar{L} \cdot \bar{p}^\dagger$ then we can write $y = \bar{L} \cdot x$, $x \in \bar{p}^\dagger$, so that $\langle \bar{q}, y\rangle = \langle \bar{q}, \bar{L} \cdot x\rangle = \langle \bar{L}^\dagger \cdot \bar{q}, x\rangle = \langle \bar{p}, x\rangle$, and $\langle \bar{p}, x\rangle \le 0$, since $x \in \bar{p}^\dagger$. So $\langle \bar{q}, y\rangle \le 0$ for all $y \in \bar{L} \cdot \bar{p}^\dagger$. Since $\bar{y} \in \bar{L} \cdot \bar{p}^\dagger$ and $\langle \bar{q}, \bar{y}\rangle \ge 0$, we conclude that $\langle \bar{q}, \bar{y}\rangle = 0$. But then, if we take $y = \bar{y} + \varepsilon \bar{q}$, where ε is positive and small enough, we have $\langle \bar{q}, y\rangle = \varepsilon > 0$, while on the other hand $y \in \bar{L} \cdot \bar{p}^\dagger$. So we have reached a contradiction, proving (#).

We now fix a δ having the properties of (#), and choose $\kappa = \delta^{-1}$. We then apply the definition of the Warga derivate container, and obtain

- an $R \in \mathbb{R}$ such that $R > 0$, $\bar{\mathbb{B}}^n(0,R) \subseteq U$ and $R \le \delta$,
- a sequence $\{F_j\}_{j\in\mathbb{N}}$ of functions of class C^1 from $\bar{\mathbb{B}}^n(0,R)$ to $\mathbb{R}^m$ such that
  - $F_j \to F$ uniformly on $\bar{\mathbb{B}}^n(0,R)$ as $j \to \infty$,
  - $DF_j(x) \in \hat{\Lambda}$ for all $x \in \bar{\mathbb{B}}^n(0,R)$, $j \in \mathbb{N}$,

where $\hat{\Lambda} = \{L \in \mathbb{R}^{m\times n} : \mathrm{dist}(L,\Lambda) \le \delta\}$. After replacing F_j by $F_j - F_j(0)$ we may assume, in addition, that

  - $F_j(0) = 0$ for every $j \in \mathbb{N}$.

We now let $D = \{y \in \mathbb{R}^m : \langle y, \bar{y}\rangle \ge (1-\tilde{\delta})\|y\|$, where $\tilde{\delta} = \frac{\delta^2}{2}$, so that $\delta = \sqrt{2\tilde{\delta}}$. Then D is a closed convex cone, and $\bar{y} \in \mathrm{Int}\, D$. We choose $\alpha = \delta R$, and define $\hat{S} = \bar{\mathbb{B}}^n(0,R) \cap S$, so $\hat{S}$ is compact and $0 \in \hat{S}$. We will prove (3). It clearly suffices to show that

$$(\forall j \in \mathbb{N})(\forall y \in D(\alpha))(\exists \gamma \in \Gamma(\hat{S},\kappa\alpha))(\sigma_y = |F_j \circ \gamma|)\,. \quad (4)$$

(Indeed, if (4) holds, and $y \in D(\alpha)$, then for each j we can find $\gamma_j \in \Gamma(\hat{S},\kappa\alpha)$ such that $|F_j \circ \gamma_j| = \sigma_y$. Since $\Gamma(\hat{S},\kappa\alpha)$ is compact, there exists an infinite subset J of $\mathbb{N}$ such that $\gamma = \lim_{j\to_J\infty} \gamma_j$ exists and belongs to $\Gamma(\hat{S},\kappa\alpha)$. But then $\lim_{j\to_J\infty}(F_j \circ \gamma_j) = F \circ \gamma$, so $|F\circ\gamma| = \sigma_y$, and $\gamma \in \Gamma(S,\kappa\alpha)$.)

We now prove (4). We fix a j, and write $G = F_j$. Then $G \in C^1(\bar{\mathbb{B}}^n(0,R),\mathbb{R}^m)$, $G(0) = 0$, and $DG(x) \in \hat{\Lambda}$ for all $x \in \bar{\mathbb{B}}^n(0,R)$. We want to prove that

$$(\forall y \in D(\alpha))(\exists \gamma \in \Gamma(\hat{S},\kappa\alpha))(\sigma_y = |G\circ\gamma|)\,. \quad (5)$$

Let $D_0(\alpha) = \mathrm{Int}\, D(\alpha)$. Then, thanks to the compactness of $\Gamma(\hat{S},\kappa\alpha)$, it suffices to show that

$$(\forall y \in D_0(\alpha))(\exists \gamma \in \Gamma(\hat{S},\kappa\alpha))(\sigma_y = |G\circ\gamma|)\,. \quad (6)$$

To prove (6), we pick a point $y_* \in D_0(\alpha)$ and construct a $\gamma \in \Gamma(\hat{S},\kappa\alpha)$ such that $\sigma_{y_*} = |G\circ\gamma|$. We will do this by finding, for small positive ε, arcs $\gamma_\varepsilon \in \Gamma(\hat{S},\kappa\alpha)$ such that the sets $|F\circ\gamma_\varepsilon|$ converge to σ_{y_*} in the Hausdorff metric.

Pick a positive ε such that $\bar{\mathbb{B}}^m(y_*,\varepsilon) \subseteq D_0(\alpha)$. (This implies, in particular, that $\|y_*\| + \varepsilon < \alpha$.) Then let $\hat{Q}_\varepsilon = \{v \in \mathbb{R}^m : \langle v, y_*\rangle = 0 \wedge \|v\| \le \varepsilon\}$, so $\hat{Q}_\varepsilon$ is the μ-dimensional disc orthogonal to y_*, centered at 0, and having radius ε. We define $Q_\varepsilon = \{y_* + v : v \in \hat{Q}_\varepsilon\}$, so $Q_\varepsilon \subseteq \bar{\mathbb{B}}^m(y_*,\varepsilon)$.

Next, we let $\hat{y} = \frac{y_*}{\|y_*\|}$. (Recall that $y_* \neq 0$, because $y_* \in D_0(\alpha)$, and $0 \notin D_0(\alpha)$, because if $0 \in D_0(\alpha)$ it would follow—since $\delta < 1$—that $\langle y, \bar{y}\rangle \ge 0$ for all y near 0, so $\bar{y} = 0$.) We then define a function $h_\varepsilon : \mathbb{R}^m \mapsto \mathbb{R}$ by letting $h_\varepsilon(x) = \langle x, \hat{y}\rangle - \lambda_\varepsilon\|x - \langle x,\hat{y}\rangle\hat{y}\|^2$, where $\lambda_\varepsilon = \varepsilon^{-2}\|y_*\|$. Then $h_\varepsilon(0) = 0$, and in addition $h_\varepsilon(x)$ also vanishes at all points x belonging to the frontier $\partial Q_\varepsilon = \{y_* + v : v \in \mathbb{R}^m, v \perp y_*, \|v\| = \varepsilon\}$ of Q_ε. We then let $H_\varepsilon = h_\varepsilon \circ G$, so H_ε is a function of class C^1 on U.

We now let

$$\mathcal{Q}_\varepsilon = \{x \in \mathbb{R}^m : \lambda_\varepsilon\|x - \langle x,\hat{y}\rangle\hat{y}\|^2 \le \langle x,\hat{y}\rangle \le \|y_*\|\}\,. \quad (7)$$

Then $\mathcal{Q}_\varepsilon$ is obviously closed, and $\mathcal{Q} \neq \emptyset$, because $0 \in \mathcal{Q}$. Furthermore, the Hausdorff distance $d_{Ha}(\mathcal{Q}_\varepsilon,\sigma_{y_*})$ is exactly ε. (Indeed, let $x \in \mathcal{Q}_\varepsilon$. Then $x = v + r\hat{y}$, with $r = \langle x,\hat{y}\rangle$ and $v = x - r\hat{y}$, so $v \perp \hat{y}$. The fact that $x \in \mathcal{Q}_\varepsilon$ implies that $\lambda_\varepsilon\|v\|^2 \le r \le \|y_*\|$, so $r \ge 0$, and then $ry_* \in \sigma_{y_*}$ and $\|x - r\hat{y}\|^2 \le \varepsilon^2$, so $\|x - r\hat{y}\| \le \varepsilon$; since this is true for every $x \in \mathcal{Q}_\varepsilon$, while $\|x - r\hat{y}\| = \varepsilon$ if $x \in \partial Q_\varepsilon$, we see that $\max\{\mathrm{dist}(x,\sigma_{y_*}) : x \in \mathcal{Q}_\varepsilon\} = \varepsilon$. Since $\sigma_{y_*} \subseteq \mathcal{Q}_\varepsilon$, it follows that $d_{Ha}(\mathcal{Q}_\varepsilon,\sigma_{y_*}) = \varepsilon$.) In particular, $\mathcal{Q}_\varepsilon$ is bounded, so $\mathcal{Q}_\varepsilon$ is compact.

We then define a set-valued function Ψ_ε from $\bar{\mathbb{B}}^n(0,R)$ to $\mathbb{R}^n$ by letting

$$\Psi_\varepsilon(s) = \{w \in \mathbb{R}^n : \|w\| \le 1 \text{ and } \langle \nabla H_\varepsilon(s), w\rangle \ge \delta\}\,.$$

Then Ψ_ε is upper semicontinuous with compact convex values. Let

$$\begin{aligned} S'_\varepsilon &= G^{-1}(\mathcal{Q}_\varepsilon) \cap \hat{S}\,, \\ S'_{0,\varepsilon} &= \{s \in S'_\varepsilon : \|s\| < R \text{ and } \langle G(s),\hat{y}\rangle < \|y_*\|\}\,. \end{aligned}$$

Then S'_ε is a compact subset of $\hat{S}$, $S'_{0,\varepsilon}$ is a relatively open subset of S'_ε, and $0 \in S'_{0,\varepsilon}$. We will show that

$$\Psi_\varepsilon(s_*) \cap T^B_{s_*}S'_\varepsilon \neq \emptyset \quad \text{whenever} \quad s_* \in S'_{0,\varepsilon}\,. \quad (8)$$

To see this, pick a point $s_* \in S'_{0,\varepsilon}$, and write $x_* = G(s_*)$, $\pi_* = \nabla h_\varepsilon(x_*)$, $\hat{\pi}_* = \frac{\pi_*}{\|\pi_*\|}$. It follows that $x_* \in \mathcal{Q}_\varepsilon$, so $x_* = r_*\hat{y} + v_*$, with $v_* \perp \hat{y}$, $r_* = \langle x_*, \hat{y}\rangle$, and $\|v_*\| \le \varepsilon$. The fact that $s_* \in S'_{0,\varepsilon}$ then implies that $0 \le r_* < \|y_*\|$ and $\|v_*\| < \varepsilon$. Also, $\pi_* = \hat{y} - 2\lambda_\varepsilon(x_* - \langle x_*,\hat{y}\rangle\hat{y}) = \hat{y} - 2\lambda_\varepsilon v_*$, and then

$$\|\pi_*\| = \sqrt{1 + 4\lambda_\varepsilon^2\|v_*\|^2}\,,$$

since $v_* \perp \hat{y}$. Furthermore, $\langle \pi_*, \bar{y}\rangle = \langle \hat{y}, \bar{y}\rangle - 2\lambda_\varepsilon\langle v_*, \bar{y}\rangle$. Since $\hat{y} \in D$, and $\|\hat{y}\| = 1$, we have $\langle \hat{y},\bar{y}\rangle \ge 1 - \tilde{\delta}$, so

$$\|\hat{y} - \bar{y}\|^2 = \|\hat{y}\|^2 + \|\bar{y}\|^2 - 2\langle \hat{y},\bar{y}\rangle = 2(1 - \langle \hat{y},\bar{y}\rangle) \le 2\tilde{\delta}\,,$$

and then $\|\hat{y} - \bar{y}\| \le \sqrt{2\tilde{\delta}} = \delta$, so that

$$2\lambda_\varepsilon\langle v_*,\bar{y}\rangle = 2\lambda_\varepsilon\langle v_*, \bar{y} - \hat{y}\rangle \le 2\lambda_\varepsilon\|v_*\|\,\|\bar{y} - \hat{y}\| \le 2\lambda_\varepsilon\|v_*\|\delta\,,$$

(using the fact that $v_* \perp \hat{y}$), and then

$$\langle \pi_*, \bar{y}\rangle \ge 1 - \tilde{\delta} - 2\lambda_\varepsilon\|v_*\|\delta \ge -2\lambda_\varepsilon\|v_*\|\delta \ge -2\lambda_\varepsilon\varepsilon\delta$$

from which it follows that

$$\langle \hat{\pi}_*, \bar{y}\rangle \ge -\frac{2\lambda_\varepsilon\|v_*\|\delta}{\sqrt{1 + 4\lambda_\varepsilon^2\|v_*\|^2}} \ge -\delta\,.$$

Let $L_* = DG(s_*)$. Then $\mathrm{dist}(L_*,\Lambda) \le \delta$. Since $\|\hat{\pi}_*\| = 1$ and $\langle \hat{\pi}_*, \bar{y}\rangle \ge -\delta$, (#) implies that $\Theta(L_*^\dagger(\hat{\pi}_*), s) \ge \delta$. We can therefore find a $w \in T^B_s S$ such that $\|w\| = 1$ and $\langle L_*^\dagger(\hat{\pi}_*), w\rangle \ge \delta$. It follows that $\langle L_*^\dagger(\pi_*), w\rangle \ge \delta\|\pi_*\|$. Since $\|\pi_*\| \ge 1$, we can conclude that $\langle L_*^\dagger(\pi_*), w\rangle \ge \delta$. But the chain rule implies that $L_*^\dagger(\pi_*) = \nabla H_\varepsilon(x_*)$, so we have shown that $\langle \nabla H_\varepsilon(x), w\rangle \ge \delta$. This establishes that $w \in \Psi_\varepsilon(s)$.

To complete the proof of (8), we have to show that $w \in T^B_s S'_\varepsilon$. Since $w \in T^B_s S$ and $\|w\| = 1$, we can find a sequence $\{s_k\}_{\in\mathbb{N}}$ of points of $S\backslash\{s_*\}$ that converges to s_* and is such that

$$\lim_{k\to\infty} w_k = w\,, \quad \text{where} \quad w_k = \frac{s_k - s_*}{\|s_k - s_*\|}\,. \quad (9)$$

If we let $\omega_k = \|s_k - s_*\|$, $\tilde{w}_k = w_k - w$, we find

$$s_k = s_* + \omega_k w + \omega_k \tilde{w}_k \,, \quad \lim_{k\to\infty} \omega_k = 0 \,, \quad \lim_{k\to\infty} \tilde{w}_k = 0 \,.$$

Let ψ be a function from $]0,\infty[$ to $[0,\infty]$ that satisfies $\lim_{r\downarrow 0}\psi(r) = 0$ as well as the conditions

$$\|G(s) - G(s_*) - L_*(s - s_*)\| \le \psi(\|s - s_*\|)\|s - s_*\| \,, \quad (10)$$

$$|h_\varepsilon(x) - h_\varepsilon(x_*) - \langle \pi_*, x - x_*\rangle| \le \psi(\|x - x_*\|)\|x - x_*\| \quad (11)$$

for all $s \in U$ and all $x \in \mathbb{R}^m$, respectively. Let $x_k = G(s_k)$. Then (10) implies the inequality $\|x_k - x_* - \omega_k L_*(w + \tilde{w}_k)\| \le \psi(\omega_k)\omega_k$. Therefore

$$\|x_k - x_* - \omega_k L_*(w)\| \le \nu_k \omega_k \,,$$

where $\nu_k = \psi(\omega_k) + \|L_*(\tilde{w}_k)\|$, so that $\lim_{k\to\infty} \nu_k = 0$. Hence $\|x_k - x_*\| \le \omega_k \|L_*(w)\| + \nu_k \omega_k$.

Then $|\langle x_k - x_* - \omega_k L_*(w), \pi_*\rangle| \le \|\pi_*\|\nu_k\omega_k$. Therefore

$$\begin{aligned}\langle x_k - x_*, \pi_*\rangle &= \langle x_k - x_* - \omega_k L_*(w), \pi_*\rangle + \omega_k \langle L_*(w), \pi_*\rangle \\ &\ge -\omega_k \nu_k \|\pi_*\| + \omega_k \langle w, L_*^\dagger(\pi_*)\rangle \\ &\ge \omega_k(\delta - \nu_k\|\pi_*\|) \,.\end{aligned}$$

Let $\nu'_k = \psi(\|x_k - x_*\|)$, so $\lim_{k\to\infty} \nu'_k = 0$. Then

$$|h_\varepsilon(x_k) - h_\varepsilon(x_*) - \langle \pi_*, x_k - x_*\rangle| \le \nu'_k \|x_k - x_*\| \,,$$

so

$$h_\varepsilon(x_k) - h_\varepsilon(x_*) \ge \langle \pi_*, x_k - x_*\rangle - \nu'_k\|x_k - x_*\|$$

Since $\|x_k - x_*\| \le \omega_k\|L_*(w)\| + \nu_k\omega_k$ and $\langle \pi_*, x_k - x_*\rangle \ge \omega_k(\delta - \nu_k\|\pi_*\|)$, we find

$$h_\varepsilon(x_k) - h_\varepsilon(x_*) \ge \omega_k(\delta - \|\pi_*\|\nu_k - \nu'_k\|L_*(w)\| - \nu'_k\nu_k) \,.$$

So we can pick a $\bar{k} \in \mathbb{N}$ such that

$$h_\varepsilon(x_k) - h_\varepsilon(x_*) \ge \frac{1}{2}\omega_k\|\delta \quad \text{whenever} \quad k \ge \bar{k} \,. \quad (12)$$

It follows from (7) that $x \in \mathcal{Q}_\varepsilon$ if and only if $h_\varepsilon(x) \ge 0$ and $\langle x, \hat{y}\rangle \le \|y_*\|$. Since $x_* \in \mathcal{Q}_\varepsilon$, the inequality $h_\varepsilon(x_*) \ge 0$ is true, and then $h_\varepsilon(x_k) > 0$ if $k \ge \bar{k}$. Furthermore, the fact that $s_* \in S'_{0,\varepsilon}$ implies that $\langle G(s_*), \hat{y}\rangle < \|y_*\|$, i.e., that $\langle x_*, \hat{y}\rangle < \|y_*\|$, and this implies that $\langle x_k, \hat{y}\rangle < \|y_*\|$ if k is large enough. In addition, using once again the fact that $s_* \in S'_{0,\varepsilon}$, we find that $\|s_*\| < R$, so $\|s_k\| < R$ if k is large enough. It follows that we can find a $\bar{k}' \in \mathbb{N}$ such that $\bar{k}' \ge \bar{k}$ and $\langle x_k, \hat{y}\rangle < \|y_*\|$ whenever $k \ge \bar{k}'$. Then, if $k \ge \bar{k}'$, the the following hold: (i) $s_k \in S$, (ii) $h_\varepsilon(x_k) > 0$, (iii) $\langle x_k, \hat{y}\rangle < \|y_*\|$, and (iv) $\|s_k\| < R$. It follows from (ii) and (iii) that $x_k \in \mathcal{Q}_\varepsilon$, so $s_k \in G^{-1}(\mathcal{Q}_\varepsilon)$, while on the other hand (i) and (iv) imply that $s_k \in \hat{S}$. Therefore $s_k \in S_\varepsilon$. Hence $w \in T^B_{s_*}S_\varepsilon$, completing the proof of (8).

Using standard existence results from viability theory, we pick a solution $\xi_\varepsilon : I_\varepsilon \mapsto S'_{0,\varepsilon}$ of the differential inclusion $\dot{\xi}_\varepsilon(t) \in \Psi_\varepsilon(\xi(t))$ such that (i) $\xi_\varepsilon(0) = 0$, (ii) ξ_ε is defined on a subinterval I_ε of $\mathbb{R}$ such that $0 = \min I_\varepsilon$, and (iii) ξ_ε is not extendable to a solution $\tilde{\xi} : \tilde{I} \mapsto S'_{0,\varepsilon}$ such that $0 = \min \tilde{I}$, $I_\varepsilon \subseteq \tilde{I}$, and $I_\varepsilon \ne \tilde{I}$. Then ξ_ε satisfies $H_\varepsilon(\xi_\varepsilon(t)) \ge \delta t$ for all $t \in I_\varepsilon$. On the other hand, $H_\varepsilon(s) = h_\varepsilon(G(s)) \le \|y_*\|$ for all $s \in S'_\varepsilon$, so $I_\varepsilon \subseteq [0, \delta^{-1}\|y_*\|]$. It follows that $I_\varepsilon = [0, \tau_\varepsilon[$ or $I_\varepsilon = [0, \tau_\varepsilon]$ for some τ_ε or such that $0 < \tau_\varepsilon \le \delta^{-1}\|y_*\|$. If $I_\varepsilon = [0, \tau_\varepsilon]$, then ξ_ε would be extendable, contradicting the choice of $(\xi_\varepsilon, I_\varepsilon)$. So $I_\varepsilon = [0, \tau_\varepsilon[$. Since ξ_ε is Lipschitz with constant 1, the limit $\bar{s}_\varepsilon = \lim_{t\uparrow\tau_\varepsilon} \xi_\varepsilon(s)$ exists and belongs to S'_ε. If $\bar{s}_\varepsilon \in S'_{0,\varepsilon}$ then ξ_ε would be extendable. So $\bar{s}_\varepsilon \notin S'_{0,\varepsilon}$. But then either $\|\bar{s}_\varepsilon\| = R$ or $\langle G(\bar{s}_\varepsilon), \hat{y}\rangle = \|y_*\|$. The possibility that $\|\bar{s}_\varepsilon\| = R$ is excluded because $\|\bar{s}_\varepsilon\| \le \tau_\varepsilon \le \delta^{-1}\|y_*\| < \delta^{-1}\alpha = R$. Hence $\langle G(\bar{s}_\varepsilon), \hat{y}\rangle = \|y_*\|$. If we let $\bar{x}_\varepsilon = G(\bar{s}_\varepsilon)$, then this shows that $\bar{x}_\varepsilon \in Q_\varepsilon$.

We now define $\gamma_\varepsilon : [0,1] \mapsto S'_\varepsilon$ by letting $\gamma_\varepsilon(t) = \xi_\varepsilon(\tau_\varepsilon t)$ for $t \in [0,1]$. Then $\gamma_\varepsilon \in \Gamma(S, \kappa\alpha)$ (since $\tau_\varepsilon \le \delta^{-1}\alpha = \kappa\alpha$), and $\gamma_\varepsilon(0) = 0$. Furthermore, the set $|G \circ \gamma_\varepsilon|$ is entirely contained in $\mathcal{Q}_\varepsilon$, and $G(\gamma_\varepsilon(1)) \in Q_\varepsilon$. We can then pick a sequence $\{\varepsilon_k\}_{k\in\mathbb{N}}$ of positive numbers such that $\lim_{k\to\infty} \varepsilon_k = 0$ and the arcs γ_{ε_k} converge uniformly to an arc $\gamma \in \Gamma(S, \kappa\alpha)$. This arc clearly satisfies $|G \circ \gamma_\varepsilon| \subseteq \sigma_{y_*}$. Furthermore, $y_* = \lim_{k\to\infty} x_{\varepsilon_k}$, so $y_* \in |G \circ \gamma_\varepsilon|$, and then $|G \circ \gamma_\varepsilon| = \sigma_{y_*}$. This concludes the proof. ■

References

[1] Sussmann, H. J., "Some recent results on the maximum principle of optimal control theory." In *Systems and Control in the Twenty-First Century,* C. I. Byrnes, B. N. Datta, D. S. Gilliam and C. F. Martin Eds., Birkhäuser, Boston, 1997, pp. 351–372.

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

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

[4] 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.

[5] Warga, J., "Controllability, extremality and abnormality in nonsmooth optimal control." *J. Optim. Theory Applic.* **41**, 1983, pp. 239–260.

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

[7] Warga, J., "Homeomorphisms and local C^1 approximations." *Nonlinear Anal. TMA* **12**, 1988, pp. 593–597.