Set separation, approximating multicones, and the Lipschitz maximum principle

H. J. Sussmann

Department of Mathematics, Rutgers University Piscataway, NJ 08854, U.S.A.

Abstract

We present a general necessary condition for separation of the reachable set of a Lipschitz control system from another given set of states, expressed in terms of an "approximating multicone" to the set in a sense that contains as special cases the Clarke and Mordukhovich cones. We then show how this separation result implies a strengthened form of the usual sufficient condition for local controllability along the reference curve and the necessary condition for optimality.

 $Key\ words:$ Pontryagin Maximum Principle, set separation, transversality, Lipschitz

Dedicated to Arrigo Cellina on his 65th birthday.

1 Introduction

Since the discovery of the Pontryagin Maximum Principle (PMP) in the 1950s (cf. [18]), various versions of this result have been established, under different technical assumptions and with different proofs. For the finite-dimensional PMP, every proof falls, roughly, into one of two categories, that will be referred to here as "Type T" and "Type L." Type T proofs are based on a topological argument about set separation, involving the Brouwer fixed point theorem or some other closely related result. Type L proofs, on the other hand, use a limiting argument, in which a sequence $\boldsymbol{\pi} = \{\pi_j\}_{j \in \mathbb{N}}$ of approximate terminal adjoint vectors π_j —normalized so that $\|\pi_j\| = 1$ —is constructed, and then an exact adjoint vector is obtained by taking the limit of some convergent subsequence of $\boldsymbol{\pi}$. (Finite-dimensionality plays a crucial role in both types of

Email address: sussmann@math.rutgers.edu (H. J. Sussmann).

¹ Research supported in part by NSF Grants DMS01-03901 and DMS-05-09930.

proofs but, remarkably, it does so in two totally different ways: for the Type T proofs, the key point is that the Brouwer fixed point theorem is only valid in finite-dimensional spaces, whereas in the Type L proofs the decisive step occurs when one has to guarantee that the limit exists and does not vanish, and this cannot be done in infinite dimensions because in any topology weak enough to imply existence of a convergent subsequence the limit of this subsequence may vanish.)

Type T proofs have appeared in many books and articles (cf. Pontryagin *et al.* [18], Berkovitz [1]), and have been particularly successful in being able to incorporate high-order conditions (cf. Bianchini [3], Bianchini-Stefani [2], Hermes [11], Knobloch [14], Krener [15], Stefani [19]). Sussmann [20]). Type L proofs (cf. Clarke [6,7,9], Clarke *et al.* [8], de Pinho [10], Ioffe [12], Ioffe-Rockafellar [13], Mordukhovich [17], Vinter [29]) have successfully dealt with nonsmooth Lipschitz dynamical laws. In Type T proofs, the transversality condition usually involves a Boltyanskii tangent cone to the terminal set, whereas in Type L proofs a Clarke tangent cone or a Mordukhovich normal cone is used instead.

In 1993, S. Lojasiewicz Jr. ([16]) discovered a powerful new technique that made it possible to deal with nonsmoothness using the Type T approach. Subsequently, in a series of papers (cf. [21,23-26,28]), we pursued this idea and developed Type T methods for the nonsmooth PMP, based on generalized differentials, flows, and general variations. These methods, however, resisted all attempts to deal with transversality conditions involving the Clarke tangent cone or the Mordukhovich normal cone. Recently, A. Bressan (cf. [5]) found an explanation for this fact by proving, by means of a counterexample, that the usual necessary conditions for set separation that can be derived for a pair of sets and corresponding Boltyanskii approximating cones, as well as for a pair of sets and corresponding Clarke or Mordukhovich normal cones, can fail to be true if a Boltyanskii approximating cone is specified for one of the sets and the Clarke or Mordukhovich normal cone is used for the other one. This shows that versions of the PMP with "mixed" technical conditions—some corresponding to the Type T approach and others to the Type L method—are likely to be false in general, and that there probably does not exist a single unified version of the PMP that contains both types of results.

Since a single common generalization of both approaches appears not to exist, the second-best alternative is that it may at least be possible to deal with both kinds of results by means of set-separation techniques, using different but parallel separation theorems for Type T and Type L results. As a first step in this direction, we proposed in [27] a notion of "approximating multicone" to a set at a point that extends the concepts of Clarke and Mordukhovich cones and has the property that "strong transversality of the approximating cones implies nontrivial intersection of the sets." (In our setting, "convex multicones" have polars that can fail to be convex. Furthermore, any closed cone of covectors even if it is not convex—is, trivially, the polar of some convex multicone. In particular, the usual Mordukhovich normal cone is the polar of a convex multicone that we call the "Mordukhovich tangent multicone.")

In this note we apply this approach to nonsmooth control problems with a Lipschitz right-hand side. We derive a general necessary condition for a reachable set \mathcal{R} of a Lipschitz control system to be separated from another given set S at the terminal point $\xi_*(b_*)$ of the reference trajectory—in the sense that $\mathcal{R} \cap S = \{\xi_*(b_*)\}$ —expressed in terms of an approximating multicone to Sin the sense of our theory. We then show how this result can be used to derive the usual nonsmooth sufficient condition for a system to be locally controllable along a curve, and a slightly stronger form of the usual necessary condition for optimal control.

In addition, we also pursue the idea, proposed in [22], of formulating the PMP directly on manifolds, by expressing the "adjoint equation" as an equation of parallel translation with respect to a covariant differentiation along the reference curve. This second aspect is, essentially, independent of the first, and those readers who so wish may read the paper throughout as if the state space of the systems was always an open subset \mathbb{R}^m , in which case the singlevalued selections L of the Clarke generalized Jacobian map $t \mapsto \partial f_t(\xi_*(t))$ (where f is the reference vector field, f_t is the map $x \mapsto f(x,t)$, and ξ_* is the reference trajectory) become matrix-valued functions, and the adjoint equation $\nabla_L \pi = 0$ just becomes the usual adjoint equation $\dot{\pi} = -\pi \cdot L$. We feel, however, that the manifold formulation is more elegant, and also slightly more general, in the sense that the PMP on manifolds is not an immediate corollary of the PMP on open subsets of \mathbb{R}^m . (Although it is not too hard to derive the former from the latter, this requires some extra work, and cannot be done by just covering the reference trajectory by coordinate patches.)

2 Preliminaries and background

Some abbreviations and basic notations. We use the abbreviations "ppd", "tvvf", "fdrls", for "possibly partially defined", "time-varying vector field", and "finite-dimensional real linear space", respectively.

If φ is a function, we use dom φ , im φ to denote, respectively, the domain and image of φ . (So im $\varphi = \{\varphi(x) : x \in \operatorname{dom} \varphi\}$.) We write $\varphi : A \hookrightarrow B$, to indicate that φ is a **ppd map from** A **to** B, i.e., a function φ such that dom $\varphi \subseteq A$ and im $\varphi \subseteq B$. We write $\varphi : A \mapsto B$ to indicate that $\varphi : A \hookrightarrow B$ and dom $\varphi = A$. If A is a set, then \mathbb{I}_A denotes the identity map of A.

If φ is a function, A is a set, and $\{B_a\}_{a \in A}$ is a family of sets depending on $a \in A$, we will often use the expression " $A \ni a \hookrightarrow \varphi(a) \in B$ " (resp. " $A \ni a \mapsto \varphi(a) \in B$ ") as an alternative name for any function φ such that2 dom $\varphi \subseteq A$ (resp. dom $\varphi = A$) and $\varphi(a) \in B_a$ for every $a \in \text{dom } \varphi$.

We use \mathbb{Z} , \mathbb{R} to denote, respectively, the set of all integers and the set of all real numbers, and write $\mathbb{N} \stackrel{\text{def}}{=} \{n \in \mathbb{Z} : n > 0\}$, $\mathbb{Z}_+ \stackrel{\text{def}}{=} \mathbb{N} \cup \{0\}$, $\mathbb{R}_+ \stackrel{\text{def}}{=} \{x \in \mathbb{R} : x \ge 0\}$. We use square-bracket notations for intervals:]a, b[is the open interval from a to b, and then $]a, b] =]a, b[\cup\{b\}, [a, b[=]a, b[\cup\{a\}, \text{ and } [a, b] = [a, b[\cup\{b\}]$.

Linear spaces. 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 is a fdrls, then dim X, X^{\dagger} denote, respectively, the dimension and the dual of X (so that $X^{\dagger} = Lin(X, \mathbb{R})$). We identify the double dual $X^{\dagger\dagger}$ with X in the usual way.

If X, Y are fdrlss, and $L \in Lin(X, Y)$, then the **adjoint** (or **transpose**) of L is the map $L^{\dagger} : Y^{\dagger} \mapsto X^{\dagger}$ such that $L^{\dagger}(y) = y \circ L$ for $y \in Y^{\dagger}$.

If X is a fdrls, an **affine basis** of X is a sequence (e_1, \ldots, e_n) of minimal length consisting of members of X that affinely span X, that is, are such that every $x \in X$ is a linear combination $\sum_{i=1}^{n} r_i e_i$ with $\sum_{i=1}^{n} r_i = 1$.

Euclidean spaces, matrices, balls. We use $\mathbb{R}^{m \times n}$ to denote the space of all real matrices with m rows and n columns. If $A \in \mathbb{R}^{m \times n}$, then A^{\dagger} is the transpose of A, so $A^{\dagger} \in \mathbb{R}^{m \times n}$. We write $\mathbb{R}^m = \mathbb{R}^{m \times 1}$, $\mathbb{R}_m = \mathbb{R}^{1 \times m}$, so \mathbb{R}^m , \mathbb{R}_m are, respectively, the spaces of all real m-dimensional column vectors $x = (x^1, \ldots, x^n)^{\dagger}$ and of all real m-dimensional row vectors $p = (p_1, \ldots, p_n)$. Then \mathbb{R}_m is canonically identified with $(\mathbb{R}^m)^{\dagger}$ via the pairing

$$\mathbb{R}_m \times \mathbb{R}^m \ni (y, x) \mapsto y \cdot x \in \mathbb{R},$$

that is, by assigning to a $y \in \mathbb{R}_m$ the linear functional $\mathbb{R}^m \ni x \mapsto y \cdot x \in \mathbb{R}$.

The matrix space $\mathbb{R}^{m \times n}$ is canonically identified with $Lin(\mathbb{R}^n, \mathbb{R}^m)$ by means of the map that assigns to each matrix $L \in \mathbb{R}^{m \times n}$ the linear map $M_L : \mathbb{R}^n \mapsto \mathbb{R}^m$ such that $M_L(x) = L \cdot x$ whenever $x \in \mathbb{R}^m$. If $M \in Lin(\mathbb{R}^n, \mathbb{R}^m)$, and $M = M_L$, $L \in \mathbb{R}^{m \times n}$, then the map M_L^{\dagger} goes from \mathbb{R}_m to \mathbb{R}_n , and is given by $M_L^{\dagger}(y) = y \cdot L$ for $y \in \mathbb{R}_m$. Alternatively, we may identify each \mathbb{R}_k with \mathbb{R}^k in the obvious way, and then regard M_L^{\dagger} as linear map from \mathbb{R}^m to \mathbb{R}^n , in which case the matrix that corresponds to the adjoint map M_L^{\dagger} is the transpose L^{\dagger} .

The spaces \mathbb{R}^m , \mathbb{R}_m are endowed with the Euclidean norm defined by

$$||x|| = \sqrt{x^{\dagger} \cdot x} \text{ if } x \in \mathbb{R}^m, \qquad ||p|| = \sqrt{p \cdot p^{\dagger}} \text{ if } p \in \mathbb{R}^m.$$

The matrix spaces $\mathbb{R}^{m \times n}$ are endowed with the **operator norm** $\|\cdot\|_{op}$, given by $\|A\|_{op} = \max\{\|A \cdot x\| : x \in \mathbb{R}^n, \|x\| = 1\}.$

If $m \in \mathbb{Z}_+$, $x \in \mathbb{R}^m$, $r \in \mathbb{R}$, and r > 0, then $\overline{\mathbb{B}}^m(x,r)$, $\mathbb{B}^m(x,r)$ denote, respectively, the closed and open balls in \mathbb{R}^m with center x and radius r. We write $\overline{\mathbb{B}}^m(r)$, $\mathbb{B}^m(r)$ for $\overline{\mathbb{B}}^m(0,r)$, $\mathbb{B}^m(0,r)$, and $\overline{\mathbb{B}}^m$, \mathbb{B}^m for $\overline{\mathbb{B}}^m(1)$, $\mathbb{B}^m(1)$. Also, we will use \mathbb{S}^m to denote the m-dimensional unit sphere, so that $\mathbb{S}^m = \{x \in \mathbb{R}^{m+1} : \|x\| = 1\}.$

We will use throughout the standard terminology of point-set topology. In particular, a *neighborhood* of a point x in a topological space T is any subset S of T that contains an open set U such that $x \in U$. We write \overline{S} , or Clos S, to denote the closure of a set S, if there is no ambiguity as to the ambient topological space T. (Otherwise, we write $\operatorname{Clos}_T S$ for the closure of S in T.) If $A \subseteq B \subseteq X$, then $\operatorname{Int}_B A$ will denote the interior of A relative to B, i.e., the set of all $a \in A$ such that $A \cap U \subseteq B$ for some neighborhood U of a in X.

In the special case of a metric space X, with distance function d, we use $\mathbb{B}_X(x,r)$, $\overline{\mathbb{B}}_X(x,r)$, to denote, respectively, the open ball and the closed ball with center x and radius r.

If M, N are topological spaces, then $C^0(M, N)$ will denote the space of all continuous maps from M to N. If M is a topological space, then an **arc** in M is a continuous M-valued map defined on some nonempty compact subinterval I of \mathbb{R} . The expression $\mathcal{ARC}(M)$ will denote the set of all arcs in M, so

$$\mathcal{ARC}(M) = \bigcup_{-\infty < \alpha \le \beta < +\infty} C^0([\alpha, \beta], M) \,.$$

Manifolds, tangent spaces, charts. Let us assume that $\mu, m \in \mathbb{Z}_+$, M is a manifold of class C^{μ} , and dim M = m. A cubic coordinate chart of class C^{μ} on M is a diffeomorphism dom $\mathbf{x} \ni x \mapsto \mathbf{x}(x) = (x^1(x), \dots, x^m(x))^{\dagger} \in \operatorname{im} \mathbf{x}$ of class C^{μ} from an open subset dom \mathbf{x} of M onto an open subset im \mathbf{x} of \mathbb{R}^m , such that $\operatorname{im} \mathbf{x}$ is the open cube $]-c, c [^m$ for some positive real number c. (Recall that the members of \mathbb{R}^m are column vectors.) Once it has been stipulated that a manifold M is of class C^{μ} , we will simply use the word "chart" for "cubic coordinate chart of class C^{μ} ." If $x \in M$, a chart near x is a chart $\mathbf{x}(x) = 0$. Given a chart \mathbf{x} on M, every point $x \in \operatorname{dom} \mathbf{x}$ has a coordinate representation $x^{\mathbf{x}} \in \mathbb{R}^m$, given by $x^{\mathbf{x}} = \mathbf{x}(x)$.

Now assume in addition that $\mu \geq 1$. We then use TM, T^*M to denote, respectively, the tangent and cotangent bundles of M, so TM and T^*M are manifolds of class $C^{\mu-1}$, and are vector bundles over M of class $C^{\mu-1}$ with fiber dimension m. For each $x \in M$, T_xM and T_x^*M are, respectively, the tangent and cotangent spaces of M at x, i.e. the fibers over x of TM and T^*M . If N is another manifold of class C^{μ} , $x \in M$, F is an N-valued map defined on a neighborhood U of x in M, and F is classically differentiable at x, then DF(x) will denote the differential of F at x, so $DF(x) \in Lin(T_xM, T_{F(x)}N)$.

If **x** is a chart of M, we let $\partial_i^{\mathbf{x}} \stackrel{\text{def}}{=} \frac{\partial}{\partial x^i}$, so the $\partial_i^{\mathbf{x}}$ are vector fields of class $C^{\mu-1}$ on dom **x**, and $(\partial_1^{\mathbf{x}}(x), \ldots, \partial_m^{\mathbf{x}}(x))$ is a basis of $T_x M$ for each $x \in \text{dom } \mathbf{x}$. Also, the differentials dx^1, \ldots, dx^m are 1-forms of class $C^{\mu-1}$ on dom **x**, and $(dx^1(x), \ldots, dx^m(x))$ is a basis of $T_x^* M$ for each $x \in \text{dom } \mathbf{x}$, dual to the basis $(\partial_1^{\mathbf{x}}(x), \ldots, \partial_m^{\mathbf{x}}(x))$ of $T_x M$.

It follows that every tangent vector v at a point $x \in \text{dom } \mathbf{x}$ has a coordinate representation $v^{\mathbf{x}} \in \mathbb{R}^m$, given by

$$v^{\mathbf{x}} = (v^{\mathbf{x},1}, v^{\mathbf{x},2}, \dots, v^{\mathbf{x},m})^{\dagger},$$

where the $v^{\mathbf{x},i}$ are such that

$$v = \sum_{i=1}^{m} v^{\mathbf{x},i} \partial_i^{\mathbf{x}}(x) \,.$$

Similarly, every covector w at x can be represented by a row vector $w^{\mathbf{x}} \in \mathbb{R}_m$, given

 $w^{\mathbf{x}} = (w_1^{\mathbf{x}}, w_2^{\mathbf{x}}, \dots, w_m^{\mathbf{x}}),$ where $w_i^{\mathbf{x}} = \langle w, \partial_i^{\mathbf{x}}(x) \rangle$, so that $w = \sum_{i=1}^m w_i^{\mathbf{x}} dx^i(x).$

Then every vector field f on dom \mathbf{x} has a coordinate representation $f^{\mathbf{x}}$, which is a vector field on im \mathbf{x} , i.e., a map im $\mathbf{x} \ni x \mapsto f^{\mathbf{x}}(x) \in \mathbb{R}^m$, given by the formula $f^{\mathbf{x}}(x) = (f^{\mathbf{x},1}(x), f^{\mathbf{x},2}(x), \dots, f^{\mathbf{x},m}(x))^{\dagger}$, where the functions $f^{\mathbf{x},i}$ are defined by $f^{\mathbf{x},i}(\mathbf{x}(x)) = \langle dx^i(x), f(x) \rangle$ for $x \in \text{dom } \mathbf{x}$. (Equivalently, $f(x) = \sum_{i=1}^m f^{\mathbf{x},i}(x^{\mathbf{x}})\partial_i^{\mathbf{x}}(x)$ for every $x \in \text{dom } \mathbf{x}$, i.e., $f = \sum_{i=1}^m (f^{\mathbf{x},i} \circ \mathbf{x})\partial_i^{\mathbf{x}}$ on dom \mathbf{x} .)

Sections. Whenever A is a set equipped with a "bundle structure" over a set B (meaning, for our purposes, no more than a surjective map $p: A \mapsto B$, called the "projection"), a **section** of A is a map $B \ni b \mapsto \sigma(b) \in A$ such that $p(\sigma(b)) = b$ for all $b \in B$. If P is any property of sections, then $\Gamma_P(A)$ will denote the set of all sections of A that have Property P. In particular, if M is a manifold of class C^{μ} , $\ell \in \mathbb{Z}_+$, and $\ell \leq \mu - 1$, then $\Gamma_{C^{\ell}}(TM)$, $\Gamma_{C^{\ell}}(T^*M)$, will be, respectively, the space of all vector fields of class C^{ℓ} on M and the space of all 1-forms of class C^{ℓ} on M.

Generalized Jacobians of locally Lipschitz maps. We assume

(A1) $m, n, \mu \in \mathbb{Z}_+, M, N$ are manifolds of class $C^{\mu}, \mu \ge 1, m = \dim M,$ and $n = \dim N.$ If $0 \le k \le \mu$, then $C^k(M, N)$ will denote the set of all maps of class C^k from M to N. We let $J^1_{x,y}(M, N)$ be, for each point $(x, y) \in M \times N$, the space $Lin(T_xM, T_yN)$ of all linear maps from T_xM to T_yN . We then let

$$J^1(M,N) = \bigcup_{x \in M, y \in N} \{x\} \times \{y\} \times J^1_{x,y}(M,N),$$

so the members of $J^1(M, N)$ are the triples (x, y, L) such that $x \in M$, $y \in N$, $L \in Lin(T_xM, T_yN)$. Then $J^1(M, N)$ is a vector bundle of class $C^{\mu-1}$ over $M \times N$ (with projection map $(x, y, L) \mapsto (x, y)$), whose fiber over each point $(x, y) \in M \times N$ is the set $\{x\} \times \{y\} \times J^1_{x,y}(M, N)$ (canonically identified with $J^1_{x,y}(M, N)$), so $J^1(M, N)$ has fiber dimension mn. We will also regard $J^1(M, N)$ as a bundle over M, in such a way that the fiber over each $x \in M$ is $J^1_x(M, N)$ —where $J^1_x(M, N) = \bigcup_{y \in N} \{y\} \times J^1_{x,y}(M, N)$ —so the fiber dimension is n + mn. If a map φ belongs to $C^1(M, N)$ then the pair $j^1\varphi(x) = (\varphi(x), D\varphi(x))$ —which is a member of $J^1_x(M, N)$ —is called the 1-jet of φ at x. It is then clear that $J^1_x(M, N) = \{j^1\varphi(x) : \varphi \in C^1(M, N)\}$, so $J^1_x(M, N)$ is the space of all 1-jets at x of maps of class C^1 from M to N, and $J^1(M, N)$ is the space of all 1-jets of maps of class C^1 from M to N. The map $M \ni x \mapsto j^1\varphi(x) \in J^1(M, N)$ is the 1-jet map of φ . Clearly, if $1 \leq k \leq \mu$ and $\varphi \in C^k(M, N)$, then $j^1\varphi$ is a section of class C^{k-1} of $J^1(M, N)$, regarded as a bundle over M.

The concept of a *locally Lipschitz map* from M to N makes sense. (For example, we can define a map $\varphi: M \mapsto N$ to be *locally Lipschitz* if $\rho \circ \varphi \circ \psi$ is locally Lipschitz whenever $\rho \in C^1(N, \mathbb{R}), p \in \mathbb{Z}_+$, and $\psi \in C^1(\mathbb{R}^p, M)$.) We use $C^{Lip}(M, N)$ to denote the set of all locally Lipschitz maps from M to N. We let $diff(\varphi)$ be the set of points of differentiability of φ . It follows from the well known Rademacher theorem that if $\varphi \in C^{Lip}(M, N)$ then φ is differentiable almost everywhere, that is, $M \setminus diff(\varphi)$ is a null subset of M. (The concept of a "null subset of M" clearly makes sense intrinsically, since M is of class C^1 .) Then for every point $x \in diff(\varphi)$ the map φ has a well defined differential $D\varphi(x) \in Lin(T_xM, T_{\varphi(x)}N)$. This implies that the 1-jet map $diff(\varphi) \ni x \mapsto j^1 \varphi(x) = (\varphi(x), D\varphi(x)) \in J^1_x(M, N)$ is well defined. In addition, this map has the property that for every compact subset K of M the closure $j^{1}\varphi(K \cap diff(\varphi))$ is a compact. subset of $J^{1}(M, N)$. It follows that, if x is a point of M, and we let $j^{1}\varphi(x)$ be the set of all limits as $k \to \infty$ (in the space $J^1(M, N)$) of sequences $\{j^1\varphi(x_k)\}_{k=1}^{\infty}$ such that $x_k \in diff(\varphi)$, $\lim_{k\to\infty} x_k = x$, and the limit $\lim_{k\to\infty} j^1 \varphi(x_k)$ exists, then $\widetilde{j^1 \varphi}(x)$ is a nonempty compact subset of $J^1_{x,\varphi(x)}(M,N)$. Since $J^1_{x,\varphi(x)}(M,N)$ is a linear space, the convex hull of $j^{1}\varphi(x)$ is well defined. We use $\partial\varphi(x)$ to denote this convex hull, and refer to it as the **Clarke generalized Jacobian** of φ at x.

Relative to charts \mathbf{x} , \mathbf{y} of M, N, such that $x \in \operatorname{dom} \mathbf{x}$, $y \in \operatorname{dom} \mathbf{y}$, and $\varphi(\operatorname{dom} \mathbf{x}) \subseteq \operatorname{dom} \mathbf{y}$, the map φ is represented by the map $\varphi^{\mathbf{y},\mathbf{x} \stackrel{\text{def}}{=} \mathbf{y}} \circ \varphi \circ \mathbf{x}^{-1}$,

from im \mathbf{x} to \mathbb{R}^n . Then the 1-jet $j^1\varphi(x')$ at any point $x' \in diff(\varphi)$ close to x is represented by the pair $\theta(x') = (\mathbf{y}(\varphi(x')), D\varphi^{\mathbf{y}, \mathbf{x}}(\mathbf{x}(x'))) \in \mathbb{R}^n \times \mathbb{R}^{n \times m}$. The limit of a sequence $\{j^1\varphi(x_k)\}_{k=1}^{\infty}$ as above will then be represented by the limit $\lim_{k\to\infty} \theta(x_k)$, which is equal to the pair $(\mathbf{y}(\varphi(x)), \lim_{k\to\infty} D\varphi^{\mathbf{y}, \mathbf{x}}(\mathbf{x}(x_k)))$. It follows that

- (CGJ.1) If $\varphi \in C^{Lip}(M, N)$ and $x \in M$, then $\partial \varphi(x)$ is a nonempty compact convex subset of $J^1_{x,\varphi(x)}(M, N)$.
- (CGJ.2) The coordinate representation $\partial \varphi(x)^{\mathbf{y},\mathbf{x}}$ of $\partial \varphi(x)$ (which is a subset of $\mathbb{R}^{n \times m}$), is exactly the usual Clarke generalized Jacobian $\partial \varphi^{\mathbf{y},\mathbf{x}}(\mathbf{x}^{\mathbf{x}})$ of the coordinate representation $\varphi^{\mathbf{y},\mathbf{x}}$ of φ at $\mathbf{x}^{\mathbf{x}}$.

Warga derivate containers. We assume (A1), as above. "Warga derivate containers" are defined as follows.

Definition 2.1 Assume that $F \in C^{Lip}(M, N)$, and let Λ be a compact subset of $J^1_{x,F(x)}(M, N)$. We say that Λ is a **Warga derivate container of** F **at** x if for every open subset Ω of $J^1(M, N)$ such that $\Lambda \subseteq \Omega$ there exist (a) an open subset U of M such that $x \in U$, and (b) a sequence $\{F_k\}_{k=1}^{\infty}$ of members of $C^1(U, N)$, such that $(x', F_k(x'), DF_k(x')) \in \Omega$ for all $x' \in U$ and all $k \in \mathbb{N}$, and $F_k \to F$ uniformly on compact subsets of U. \Box

Remark 2.2 In Definition 2.1, " $\varphi_k \to \varphi$ uniformly on compact subsets of U" means: "if $\{x_k\}_{k=1}^{\infty}$ is a sequence of points of U that converges to an $x \in U$, then $\varphi_k(x_k) \to \varphi(x)$ as $k \to \infty$."

It will be convenient to extend the above definition to some multivalued maps in a fairly trivial way.

Definition 2.3 Let $M \ni x \mapsto F(x) \subseteq N$ be a multivalued map from M to N. Let $(x, y) \in M \times N$, and let Λ be a compact subset of $J^1_{x,y}(M, N)$. We say that Λ is a **Warga derivate container of** F **at** (x, y) if for every open subset Ω of $J^1(M, N)$ such that $\Lambda \subseteq \Omega$ there exist (a) an open subset U of M such that $x \in U$, (b) an $f \in C^{Lip}(U, N)$ such that f(x) = y and $f(x') \in F(x')$ for all $x' \in U$, and (c) a subset Λ' of Ω such that Λ' is a Warga derivate container of f at x in the sense of Definition 2.1. \Box

We will write " $\Lambda \in WDC(F; x, y)$ " to indicate that Λ is a Warga derivate container of F at (x, y). If F is single-valued, we just write " $\Lambda \in WDC(F; x)$ " instead of " $\Lambda \in WDC(F; x, F(x))$."

It follows easily from Definitions 2.1 and 2.3 that

(WDC.1) If $\Lambda \in WDC(F; x, y)$ then $\Lambda \neq \emptyset$. (WDC.2) If Λ is a compact subset of $J^1_{x,y}(M, N)$, then $\Lambda \in WDC(F; x, y)$ if and only if the coordinate representation $\Lambda^{\mathbf{y}, \mathbf{x}}$ of Λ , (which is a subset of $\mathbb{R}^{n \times m}$), is a Warga derivate container at $(x^{\mathbf{x}}, y^{\mathbf{y}})$ of the coordinate representation $F^{\mathbf{y}, \mathbf{x}}$ of F.

These observations imply that many well known facts about Warga derivate containers of single-valued maps and their relationship with Clarke generalized Jacobians extend trivially to manifolds. In particular,

(WDC.3) If
$$F \in C^{Lip}(M, N)$$
 and $x \in M$, then
(WDC.3.a) $\partial F(x) \in WDC(F; x)$;
(WDC.3.b) if $\Lambda \in WDC(F; x)$ and Λ is convex, then $\partial F(x) \subseteq \Lambda$.

In addition, Warga derivate containers satisfy the following *chain rule*, in which we define $\Lambda_2 \circ \Lambda_1 \stackrel{\text{def}}{=} \{L_2 \circ L_1 : L_2 \in \Lambda_2, L_1 \in \Lambda_1\}.$

(WDC.4) Assume that (a) M_1 , M_2 , M_3 are manifolds of class C^1 , (b) F_1 is a set-valued map from M_1 to M_2 , (c) F_2 is a set-valued map from M_2 to M_3 , (c) $x_1 \in M_1$, $x_2 \in M_2$, and $x_3 \in M_3$, (d) Λ_1 belongs to $WDC(F_1; x_1, x_2)$, and (e) Λ_2 belongs to $WDC(F_2; x_2, x_3)$. Then $\Lambda_2 \circ \Lambda_1$ belongs to $WDC(F_2 \circ F_1; x_1, x_3)$.

Furthermore, they have the *monotonicity property*:

(WDC.5) If $\Lambda \in WDC(F; x, y)$ and $\tilde{\Lambda}$ is a compact subset of $J^1_{x,y}(M, N)$ such that $\Lambda \subseteq \tilde{\Lambda}$, then $\tilde{\Lambda} \in WDC(F; x, y)$.

Remark 2.4 Definition 2.3 easily implies that

(WDC.6) If Λ is a nonempty subset of WDC(F; x, y) which is totally ordered by inclusion, then $\bigcap_{\Lambda \in \Lambda} \Lambda \in WDC(F; x, y)$.

This implies that

(WDC.7) Every $\Lambda \in WDC(F; x, y)$ contains a $\Lambda_{min} \in WDC(F; x, y)$ which is minimal, in the sense that if $\tilde{\Lambda} \in WDC(F; x, y)$ and $\tilde{\Lambda} \subseteq \Lambda_{min}$ then $\tilde{\Lambda} = \Lambda_{min}$..

On the other hand, these minimal derivate containers are usually not unique. For example, if f is the map $\mathbb{R} \ni x \mapsto (0,0) \in \mathbb{R}^2$, $\alpha \ge 0$, and $f_{\alpha,k}$ is the map from \mathbb{R} to \mathbb{R}^2 given by $f_{\alpha,k}(x) = \frac{\alpha}{k}(\cos kx, \sin kx)$ then the sequence $\{f_{\alpha,k}\}_{k=1}^{\infty}$ converges uniformly to f, and the derivatives $f'_{\alpha,k}$ satisfy $f'_{\alpha,k}(x) \in \alpha \mathbb{S}^1$, where $\alpha \mathbb{S}^1 = \{(u, v) \in \mathbb{R}^2 : u^2 + v^2 = \alpha^2\}$. This implies that $\alpha \mathbb{S}^1 \in WDC(f; 0)$ for every α . Clearly, the sets $\alpha \mathbb{S}^1$, for different values of α , are not comparable by inclusion. \Box

Clarke generalized Jacobians of locally Lipschitz vector fields. In this subsection, we assume that

(A2) $m, \mu \in \mathbb{Z}_+$, M is a manifold of class $C^{\mu}, \mu \geq 2$, and $m = \dim M$.

Since $\mu \geq 2$, the manifolds M and TM are both of class C^1 , so the spaces $C^1(M, TM)$ and $C^{Lip}(M, TM)$ are well defined. We write

 $\Gamma_{C^1}(TM) \stackrel{\mathrm{def}}{=} C^1(M, TM) \cap \Gamma(TM) \,, \quad \Gamma_{C^{Lip}}(TM) \stackrel{\mathrm{def}}{=} C^{Lip}(M, TM) \cap \Gamma(TM) \,,$

so $\Gamma_{C^1}(TM)$, $\Gamma_{C^{Lip}}(TM)$ are the spaces of all vector fields on M that are, respectively, of class C^1 and locally Lipschitz.

Of all maps f from M to TM, those that are vector fields—that is, sections of TM—are characterized by the fact that $\pi_{TM,M} \circ f = \mathbb{I}_M$, where $\pi_{TM,M}$ is the canonical projection from TM to M. Hence the 1-jets of vector fields at a point $x \in M$ are those 1-jets $\sigma \in J_x^1(M, TM)$ that are of the form (v, L), where $v \in T_xM$, and $L \in Lin(T_xM, T_vTM)$ is such that $d\pi_{TM,M}(v) \circ L = \mathbb{I}_{T_xM}$. We use $J_x^1\Gamma(TM)$ to denote the set of all these jets, so $J_x^1\Gamma(TM)$ is a real linear space of dimension $m + m^2$. Also, we let $J^1\Gamma(TM) = \bigcup_{x \in M} J_x^1\Gamma(TM)$, so $J^1\Gamma(TM)$ is the set of all 1-jets of vector fields on M. Then $J^1\Gamma(TM)$ is a vector bundle over M of class $C^{\mu-2}$ and fiber dimension $m + m^2$.

A convenient alternative description of the 1-jets of vector fields is as follows. For each $x \in M$, define an equivalence relation $\sim^{1,x}$ on $\Gamma_{C^1}(TM)$ by letting $f \sim^{1,x} g$ —if $f, g \in \Gamma_{C^1}(TM)$ —if [f-g, h](x) = 0 for all $h \in \Gamma_{C^1}(M)$. (Here $[\cdot, \cdot]$ is the Lie bracket.) It is then easy to see that two vector fields $f, g \in \Gamma_{C^1}(TM)$ have the same 1-jet at x if and only if $f \sim^{1,x} g$. Therefore the 1-jet $j^1 f(x)$ of an $f \in \Gamma_{C^1}(TM)$ can be identified with the equivalence class $[f; \sim^{1,x}]$ of f modulo $\sim^{1,x}$.

It follows from the above identification that "the vector [f,g](x) only depends on the 1-jets $j^1f(x)$, $j^1g(x)$." That is, there exists a canonical bilinear map $Lie^1(x): J_x^1\Gamma(TM) \times J_x^1\Gamma(TM) \mapsto T_xM$ such that

$$Lie^{1}(x)(j^{1}f(x), j^{1}g(x)) = [f, g](x) \quad \text{whenever} \quad f, g \in \Gamma_{C^{1}}(TM) \,. \tag{1}$$

There is a canonical projection $\pi_{J^1\Gamma(TM),TM}$ from $J^1\Gamma(TM)$ onto TM, which sends a jet $j^1f(x) \in J^1_x\Gamma(TM)$ to the vector f(x). For any vector $v \in T_xM$, we use $J^1_{x,v}\Gamma(TM)$ to denote the set $\pi^{-1}_{J^1\Gamma(TM),TM}(v)$. Then $J^1_{x,v}\Gamma(TM)$ is an m^2 -dimensional affine subspace of $J^1_x\Gamma(TM)$, because it is the inverse image of v under the surjective linear map $J^1_x\Gamma(TM) \ni w \mapsto \pi_{J^1\Gamma(TM),TM}(w) \in T_xM$. Therefore $J^1\Gamma(TM)$ is an affine bundle over TM of class $C^{\mu-2}$ and fiber dimension m^2 .

Relative to a chart \mathbf{x} such that $x \in \text{dom } \mathbf{x}$, if $f^{\mathbf{x}}$, $g^{\mathbf{x}}$ are the coordinate representations of two vector fields f, g, so that $f^{\mathbf{x}}$ and $g^{\mathbf{x}}$ are maps of class C^1 from $\text{im } \mathbf{x}$ to \mathbb{R}^m , it is clear that $f \sim^{1,x} g$ if and only if $f^{\mathbf{x}}(x^{\mathbf{x}}) = g^{\mathbf{x}}(x^{\mathbf{x}})$ and

 $Df^{\mathbf{x}}(x^{\mathbf{x}}) = Dg^{\mathbf{x}}(x^{\mathbf{x}})$. This implies that that there exists a canonical bijective correspondence $J^{1}_{x}\Gamma(TM) \ni \sigma \mapsto \sigma^{\mathbf{x}} \in \mathbb{R}^{m} \times \mathbb{R}^{m \times m}$, under which each jet $\sigma = j^{1}f(x) \in J^{1}_{x}\Gamma(TM)$ is mapped to its **coordinate representation** $\sigma^{\mathbf{x}}$, given by $\sigma^{\mathbf{x}} = (f^{\mathbf{x}}(x^{\mathbf{x}}), Df^{\mathbf{x}}(x^{\mathbf{x}}))$. Also, if we fix a vector $v \in T_{x}M$, then there is a bijection $J^{1}_{x,v}\Gamma(TM) \ni \sigma \mapsto \sigma^{\mathbf{x},red} \in \mathbb{R}^{m \times m}$ that assigns to each jet $\sigma = j^{1}f(x) \in J^{1}_{x,v}\Gamma(TM)$ the square matrix $\sigma^{\mathbf{x},red} = Df^{\mathbf{x}}(x^{\mathbf{x}})$, known as the **reduced coordinate representation of** σ .

If $f \in \Gamma_{C^{Lip}}(TM)$, then f is a locally Lipschitz map from M to TM, so f has a well defined Clarke generalized Jacobian $\partial f(x)$ at any $x \in M$. Using the identification of the 1-jets of vector fields with equivalence classes modulo the relations $\sim^{1,x}$, we can regard $\partial f(x)$ as a subset of $J^1_{x,f(x)}\Gamma(TM)$.

If **x** is a chart of M near x, then every 1-jet $\sigma \in J^1_{x,f(x)}\Gamma(TM)$ has a reduced representation $\sigma^{\mathbf{x},red} \in \mathbb{R}^{m \times m}$. Hence every subset Λ of $J^1_{x,f(x)}\Gamma(TM)$ has a reduced representation $\Lambda^{\mathbf{x},red}$, which is a subset of $\mathbb{R}^{m \times m}$. Therefore

(CGJ.rr) If $f \in \Gamma_{C^{Lip}}(TM)$, and $x \in M$, then the reduced representation $\partial f(x)^{\mathbf{x},red}$ of $\partial f(x)$ relative to a chart \mathbf{x} is exactly the usual Clarke generalized Jacobian at $x^{\mathbf{x}}$ of the map $f^{\mathbf{x}} : \operatorname{im} \mathbf{x} \mapsto \mathbb{R}^{m}$.

3 Cones, multicones, transversality, and set separation

Cones, multicones, polars. A **cone** in a fdrls X is a nonempty subset C of X such that $r \cdot c \in C$ whenever $c \in C, r \in \mathbb{R}$ and $r \geq 0$. If C is a cone in X, the **polar** of C is the convex cone $C^{\perp} = \{\lambda \in X^{\dagger} : \lambda(c) \leq 0 \text{ for all } c \in C\}$. Then C^{\perp} is a closed convex cone in X^{\dagger} , and $C^{\perp\perp}$ is the smallest closed convex cone containing C. In particular, $C^{\perp\perp} = C$ if and only if C is closed and convex.

A *multicone* in X is a nonempty set of cones in X. A multicone C is *convex* if every member C of C is convex, and *closed* if every $C \in C$ is closed. The *polar* of C is the set $C^{\perp} = \text{Clos}(\bigcup \{C^{\perp} : C \in C\})$, so C^{\perp} is a (not necessarily convex) closed cone in X^{\dagger} . The *closure* of C is the multicone \overline{C} defined by $\overline{C} = \{\overline{C} : C \in C\}$.

Transversality of cones. We say that two convex cones C_1 , C_2 in a fdrls X are **transversal**, and write $C_1 \overline{\sqcap} C_2$, 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 $x = c_1 - c_2$. We say that C_1 and C_2 are **strongly transversal**, and write $C_1 \overline{\sqcap} C_2$, if $C_1 \overline{\sqcap} C_2$ and in addition $C_1 \cap C_2 \neq \{0\}$. Then "~ $C_1 \overline{\sqcap} C_2$ ", "~ $C_1 \overline{\sqcap} C_2$ " will stand for " C_1 and C_2 are not transversal," and " C_1 and C_2 are not strongly transversal," respectively.

Lemma 3.1 Assume that X is a fdrls and C_1 , C_2 are convex cones in X. Then the following conditions are equivalent:

(1) $C_1 \overline{\pitchfork} C_2$. (2) $\overline{C_1} \overline{\pitchfork} \overline{C_2}$. (3) Either (i) $C_1 \overline{\pitchfork} C_2$, or (ii) C_1 and C_2 are linear subspaces and $C_1 \oplus C_2 = X$. (4) $C_1^{\perp} \cap (-C_2^{\perp}) = \{0\}$.

Proof. The equivalence $(1) \Leftrightarrow (2)$ follows from the obvious chain of inclusions $C_1 - C_2 \subseteq \overline{C_1} - \overline{C_2} \subseteq \overline{C_1 - C_2}$. If (1) holds, then $C_1 - C_2 = X$, so a fortiori $\overline{C_1} - \overline{C_2} = X$, so (2) holds. Conversely, if (2) holds, then $\overline{C_1} - \overline{C_2} = X$, so $\overline{C_1 - C_2} = X$, so the convex cone $C_1 - C_2$ is dense in X, so $C_1 - C_2 = X$.

To prove that $(1) \Rightarrow (3)$, it suffices to assume that $C_1 \bar{\sqcap} C_2$ and $\sim C_1 \bar{\sqcap} C_2$, 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 $\sim C_1 \bar{\sqcap} C_2$. 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 $\sim C_1 \bar{\sqcap} C_2$ implies that $C_1 \cap C_2 = \{0\}$. Hence $C_1 \oplus C_2 = X$. This proves that $(1) \Rightarrow (3)$. The implication $(3) \Rightarrow (1)$ is trivial, so we have shown that $(1) \Leftrightarrow (2) \Leftrightarrow (3)$.

To prove that $(1) \Leftrightarrow (4)$, we observe again that, since $C_1 - C_2$ is a convex cone, the equality $C_1 - C_2 = X$ holds if and only if $\overline{C_1 - C_2} = X$. If $\overline{C_1 - C_2} \neq X$, then the Hahn-Banach theorem implies that $C_1^{\perp} \cap (-C_2^{\perp}) \neq \{0\}$. Conversely, it is clear that if $C_1^{\perp} \cap (-C_2^{\perp}) \neq \{0\}$ then $\overline{C_1 - C_2} \neq X$. So $C_1 - C_2 = X$ if and only if $C_1^{\perp} \cap (-C_2^{\perp}) = \{0\}$, completing the proof that $(1) \Leftrightarrow (4)$. \Box

If C is a convex cone in a fdrls X, then span(C) will denote the linear span of C. It is then clear that $span(C) = span(\overline{C})$. It is then easy to see that

Lemma 3.2 Assume that X is a fdrls and C is a convex cone in X. Then $\operatorname{Int}_{span(C)}(\bar{C}) = \operatorname{Int}_{span(C)}(C) \neq \emptyset$ and $\overline{\operatorname{Int}_{span(C)}(C)} = \bar{C}$.

Proof. Endow the space X with a norm. Let $S = span(C) = span(\bar{C})$. Let $\nu = \dim S$. Let (b_0, \ldots, b_n) be an affinely independent sequence of members of C of maximum length. Then $n = \nu$, and (b_0, \ldots, b_ν) is an affine basis of S. Let Σ be the set of all convex combinations of the b_i with strictly positive coefficients. Then Σ is a nonempty subset of C, and Σ is relatively open in S. Hence $\Sigma \subseteq \operatorname{Int}_S(\bar{C})$, so $\operatorname{Int}_S(\bar{C}) \neq \emptyset$.

Fix an $x \in \text{Int}_S(C)$. Then we can pick a small positive number r such that

 $x + rb_0, \ldots, x + rb_{\nu}$ belong to \bar{C} . Let $v_i = x + rb_i$. Then (v_0, \ldots, v_{ν}) is also an affine basis of S, consisting of members of \bar{C} . Then for each positive ε we can find members v_i^{ε} of C such that $||v_i^{\varepsilon} - v_i|| < \varepsilon$. Then $(v_0^{\varepsilon}, \ldots, v_{\nu}^{\varepsilon})$ is an affine basis of S if ε is small enough. Since $x = \frac{1}{\nu+1}(v_0 + \cdots + v_{\nu})$, we can conclude that, for small ε , x is a convex combination of $v_0^{\varepsilon}, \ldots, v_{\nu}^{\varepsilon}$. Since the v_i^{ε} belong to C, we conclude that $x \in C$ as well. So $\mathrm{Int}_S(\bar{C}) \subseteq C$, and then $\mathrm{Int}_S(\bar{C}) \subseteq \mathrm{Int}_S(C)$. Since $\mathrm{Int}_S(\bar{C}) \subseteq \mathrm{Int}_S(\bar{C})$ trivially, we conclude that

Next, fix an $x \in C$. Pick a $y \in \Sigma$. Then $x + ry \in \operatorname{Int}_S(C)$ for every positive $r \in \mathbb{R}$. Letting $r \downarrow 0$, we see that $x \in \operatorname{Int}_S(C)$. So $C \subseteq \overline{\operatorname{Int}_S(C)}$. Therefore $\overline{C} \subseteq \overline{\operatorname{Int}_S(C)}$. Since $\overline{\operatorname{Int}_S(C)} \subseteq \overline{C}$ trivially, we conclude that $\overline{\operatorname{Int}_S(C)} = \overline{C}$. \Box

Lemma 3.3 Assume that X is a fdrls, C_1 , C_2 are convex cones in X, and $C_1 \overrightarrow{\sqcap} C_2$. Then $\overline{C_1 \cap C_2} = \overline{C_1} \cap \overline{C_2}$ and $(C_1 \cap C_2)^{\perp} = C_1^{\perp} + C_2^{\perp}$.

Proof. We let $S_i = span(C_i)$ for i = 1, 2.

The inclusion $\overline{C_1 \cap C_2} \subseteq \overline{C_1} \cap \overline{C_2}$ is trivial. To prove the opposite inclusion, we pick $x \in \overline{C_1} \cap \overline{C_2}$. Let (e_1, \ldots, e_n) be a basis of X, and let $e_0 = -(e_1 + \cdots + e_n)$, so (e_0, e_1, \ldots, e_n) is an affine basis of X such that $e_0 + e_1 + \cdots + e_n = 0$. Using the transversality of C_1 and C_2 , write $e_i = f_i - g_i$, $f_i \in C_1$, $g_i \in C_2$. Using the inclusions $C_i \subseteq \overline{\operatorname{Int}}_{S_i}(C_i)$, which follow from Lemma 3.2, we find, for each positive ε , $f_i^{\varepsilon} \in \operatorname{Int}_{S_1}(C_1)$ and $g_i^{\varepsilon} \in \operatorname{Int}_{S_2}(C_2)$ such that $\|f_i^{\varepsilon} - f_i\| < \varepsilon$ and $\|g_i^{\varepsilon} - g_i\| < \varepsilon$. Let $e_i^{\varepsilon} = f_i^{\varepsilon} - g_i^{\varepsilon}$. Then $(e_0^{\varepsilon}, \ldots, e_n^{\varepsilon})$ is an affine basis of X if ε is small enough. Since $0 = \frac{1}{n+1}(e_0 + \cdots + e_n)$, we can conclude that, for small ε , we can write $0 = \sum_{i=0}^n \alpha_i^{\varepsilon} e_i^{\varepsilon}$ with $\alpha_i^{\varepsilon} > 0$ and $\sum_{i=0}^n \alpha_i^{\varepsilon} f_i^{\varepsilon}$, for r > 0, so $x_r = x + r \sum_{i=0}^n \alpha_i^{\varepsilon} g_i^{\varepsilon}$. Then $x_r \to x$ as $r \downarrow 0$. Furthermore, since the f_i^{ε} belong to $\operatorname{Int}_{S_1}(\overline{C_1})$, the α_i are > 0, and $x \in \overline{C_1}$, it follows that $x_r \in \operatorname{Int}_{S_1}(\overline{C_1})$ (using the trivial fact that $C + \operatorname{Int}_{span(C)}(C) \subseteq \operatorname{Int}_{span(C)}(C)$ for any convex cone C). So Lemma 3.2 implies that $x_r \in C_1 \cap \overline{C_2}$. This shows that $\overline{C_1} \cap \overline{C_2} \subseteq \overline{C_1} \cap \overline{C_2}$. The inclusion $\overline{C_1} \cap \overline{C_2} \subseteq \overline{C_1} \cap \overline{C_2}$ is trivial. Hence $\overline{C_1} \cap \overline{C_2} = \overline{C_1} \cap \overline{C_2}$.

To prove that $(C_1 \cap C_2)^{\perp} = C_1^{\perp} + C_2^{\perp}$, we first show that $C_1^{\perp} + C_2^{\perp}$ is closed. Suppose $p \in \overline{C_1^{\perp} + C_2^{\perp}}$. Then $p = \lim_{j \to \infty} p_j$ for a sequence $\{p_j\}_{j \in \mathbb{N}}$ of members of $C_1^{\perp} + C_2^{\perp}$. Write $p_j = q_j + r_j$, $q_j \in C_1^{\perp}$, $r_j \in C_2^{\perp}$. Let us show that the sequence $\mathbf{r} = \{r_j\}_{j \in \mathbb{N}}$ is bounded. Suppose that \mathbf{r} is unbounded. Then we may find an infinite subset J' of \mathbb{N} such that the sequence $\{\|r_j\|\}_{j \in J}$ converges to $+\infty$, and then an infinite subset J of J' such that the limit $r = \lim_{j \to \infty} \tilde{r}_j$ exists, where $\tilde{r}_j = \frac{r_j}{\|r_j\|}$. Let $\tilde{q}_j = \frac{q_j}{\|r_j\|}$. It is then clear that $\tilde{q}_j = \frac{p_j - r_j}{\|r_j\|} = \frac{p_j}{\|r_j\|} - \frac{r_j}{\|r_j\|}$. So $\lim_{j \to \infty} \tilde{q}_j = -r$. Hence $-r \in C_1^{\perp}$, $r \in C_2^{\perp}$, and $\|r\| = 1$. So $-r \in C_1^{\perp} \cap (-C_2^{\perp})$ and $-r \neq 0$. So $C_1^{\perp} \cap (-C_2^{\perp}) \neq \{0\}$, and then Lemma 3.1 implies that $\sim C_1 \pitchfork C_2$, contradicting our hypotheses. So **r** is bounded. Then we may find an infinite subset J of \mathbb{N} such that the sequence $\{r_j\}_{j \in J}$ converges to a limit r. Since $q_j = p_j - r_j$, and $p_j \to p$, we can conclude that $\lim_{j\to\infty,j\in J} q_j = p - r^{\text{def}}_2 q$. Since the cones C_1^{\perp} and C_2^{\perp} are closed, it follows that $q \in C_1^{\perp}$ and $r \in C_2^{\perp}$. But p = q + r, so $p \in C_1^{\perp} + C_2^{\perp}$. This proves that $C_1^{\perp} + C_2^{\perp}$ is closed.

We are now ready to prove that $(C_1 \cap C_2)^{\perp} = C_1^{\perp} + C_2^{\perp}$. Since the inclusion $C_1^{\perp} + C_2^{\perp} \subseteq (C_1 \cap C_2)^{\perp}$ is trivial, it suffices to show that $(C_1 \cap C_2)^{\perp} \subseteq C_1^{\perp} + C_2^{\perp}$. Pick a $\bar{p} \in (C_1 \cap C_2)^{\perp}$. Suppose that $\bar{p} \notin C_1^{\perp} + C_2^{\perp}$. Since $C_1^{\perp} + C_2^{\perp}$ is closed, the Hahn-Banach theorem implies that there exists $x \in X$ such that $\langle \bar{p}, x \rangle > 0$ and $x \in (C_1^{\perp} + C_2^{\perp})^{\perp}$. But then x belongs to $C_1 \cap C_2$, so $\langle \bar{p}, x \rangle \leq 0$, since $\bar{p} \in (C_1 \cap C_2)^{\perp}$. So we have reached a contradiction, proving the desired conclusion that $(C_1 \cap C_2)^{\perp} = C_1^{\perp} + C_2^{\perp}$.

Lemma 3.4 Assume that X is a fdrls and C_1 , C_2 are convex cones in X. Then $C_1 \overline{\boxplus} C_2$ if and only if $\overline{C_1} \overline{\boxplus} \overline{C_2}$.

Proof. We use Lemma 3.1. Assume that $C_1 \bar{\square} C_2$. Then, to begin with, $C_1 \bar{\square} C_2$, so $\overline{C_1} \bar{\square} \overline{C_2}$ are transversal. Then Lemma 3.1 tells us that either $\overline{C_1} \bar{\square} \overline{C_2}$, or $\overline{C_1}$ and $\overline{C_2}$ are linear subspaces such that $\overline{C_1} \oplus \overline{C_2} = X$. In the latter case, it would follow that $C_1 \oplus C_2 = X$, contradicting the fact that $C_1 \bar{\square} C_2$. So $\overline{C_1} \bar{\square} \overline{C_2}$.

Conversely, assume that $\overline{C_1} \overline{\boxtimes} \overline{C_2}$. Then $\overline{C_1} \overline{\boxtimes} \overline{C_2}$, so Lemma 3.1 tells us that $C_1 \overline{\boxtimes} C_2$, and also that either $C_1 \overline{\boxtimes} C_2$ or C_1 and C_2 are linear subspaces such that $C_1 \oplus C_2 = X$. In the latter case, $\overline{C_1} = C_1$ and $\overline{C_2} = C_2$, so $\overline{C_1}$ and $\overline{C_2}$ are linear subspaces such that $\overline{C_1} \oplus \overline{C_2} = X$, contradicting the fact that $\overline{C_1} \overline{\boxtimes} \overline{C_2}$. So $C_1 \overline{\boxtimes} C_2$.

Transversality of multicones. Two convex multicones C_1 , C_2 in a fdrls X are **transversal** if $C_1 \overline{\pitchfork} C_2$ whenever $C_1 \in C_1$, $C_2 \in C_2$. A linear functional $\mu \in X^{\dagger}$ is **intersection positive** on (C_1, C_2) if

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

$$(2)$$

The convex multicones C_1, C_2 are *strongly transversal* if they are transversal and in addition there exists a $\mu \in X^{\dagger}$ which is intersection positive on (C_1, C_2) .

We will use for transversality of multicones the same notations as for cones: the expression " $C_1 \overline{\sqcap} C_2$ " (resp. "~ $C_1 \overline{\sqcap} C_2$ ") means that C_1 and C_2 are (resp. are not) transversal, and " $C_1 \overline{\sqcap} C_2$ " (resp. "~ $C_1 \overline{\sqcap} C_2$ ") means that C_1 and C_2 are (resp. are not) strongly transversal.

Lemma 3.5 Assume that X is a fdrls and C_1 , C_2 are convex multicones in X. Then the following conditions are equivalent:

(i) C_1 and C_2 are not strongly transversal; (ii) $\overline{C_1}$ and $\overline{C_2}$ are not strongly transversal: (iii) for every $\nu \in X^{\dagger} \setminus \{0\}$ there exist $C_1 \in C_1, C_2 \in C_2, \omega_1 \in C_1^{\perp}, \omega_2 \in C_2^{\perp}, \omega_0 \in \mathbb{R}_+$, such that $(\omega_0, \omega_1, \omega_2) \neq (0, 0, 0)$ and $\omega_1 + \omega_2 = \omega_0 \nu$.

Proof. The implication $C_1 \[mu] C_2 \Rightarrow \overline{C_1} \[mu] \overline{C_2}$ follows trivially from the definition of strong transversality. The reverse implication $\overline{C_1} \[mu] \overline{C_2} \Rightarrow C_1 \[mu] \overline{C_2}$ is true as well, because if $\overline{C_1} \[mu] \overline{C_2}$ then (a) $\overline{C_1} \[mu] \overline{C_2}$, so Lemma 3.1 implies that $C_1 \[mu] \overline{C_2}$, and (b) if $\mu \in X^{\dagger} \setminus \{0\}$ is such that $(\forall C_1 \in \overline{C_1}) (\forall C_2 \in \overline{C_2}) (\exists x \in C_1 \cap C_2) (\mu(x) > 0)$, and $C_1 \in C_1, C_2 \in C_2$, then $\overline{C_1} \in \overline{C_1}$ and $\overline{C_2} \in \overline{C_2}$, so there exists $x \in \overline{C_1} \cap \overline{C_2}$ such that $\mu(x) > 0$. Since $\overline{C_1} \cap \overline{C_2} = \overline{C_1} \cap \overline{C_2}$ by Lemma 3.3, we can conclude that $x \in \overline{C_1} \cap \overline{C_2}$, so we can approximate x by $x_j \in C_1 \cap C_2$, and then $\mu(x_j) > 0$ if j is large enough. Hence $C_1 \[mu] \overline{C_2} \Leftrightarrow \overline{C_1} \[mu] \overline{C_2}$, and this implies that $(i) \Leftrightarrow (ii)$.

Now assume that (i) holds, i.e., that $\sim C_1 \overline{\square} C_2$. Fix a $\nu \in X^{\dagger} \setminus \{0\}$. Then either $\sim C_1 \overline{\square} C_2$, or $C_1 \overline{\square} C_2$ but there does not exist a functional $\mu \in X^{\dagger}$ which is intersection positive on (C_1, C_2) . If $\sim C_1 \overline{\square} C_2$, then there exist $C_1 \in C_1, C_2 \in C_2$ such that $C_1 - C_2 \neq X$. Then Lemma 3.1 implies that $C_1^{\perp} \cap (-C_2^{\perp}) \neq \{0\}$, so we can find a nonzero member ω_1 of $C_1^{\perp} \cap (-C_2^{\perp})$. Let $\omega_2 = -\omega_1$. Then $\omega_1 \in C_1^{\perp}$ and $\omega_2 \in C_2^{\perp}$. Let $\omega_0 = 0$. Then it is clear that $\omega_1 + \omega_2 = \omega_0 \nu$ and $(\omega_0, \omega_1, \omega_2) \neq (0, 0, 0)$. Next assume that $C_1 \overline{\square} C_2$. Then ν cannot be intersection positive on (C_1, C_2) , so there exist $C_1 \in C_1, C_2 \in C_2$ such that $\nu(x) \leq 0$ for all $x \in C_1 \cap C_2$. This says that $\nu \in (C_1 \cap C_2)^{\perp}$. Since $C_1 \overline{\square} C_2$, Lemma 3.3 implies that $\nu \in C_1^{\perp} + C_2^{\perp}$. Then we can write $\nu = \omega_1 + \omega_2, \omega_1 \in C_1^{\perp}, \omega_2 \in C_2^{\perp}$. If we take $\omega_0 = 1$, then $(\omega_0, \omega_1, \omega_2) \neq (0, 0, 0), \omega_0 \in \mathbb{R}_+$, and $\omega_1 + \omega_2 = \omega_0 \nu$. This shows that (*iii*) holds. Hence (*i*) \Rightarrow (*iii*).

We now prove that $(iii) \Rightarrow (i)$, by showing that the negation of (i) implies the negation of (iii). Assume that (i) is false, i.e., that $C_1 \bigoplus C_2$. We want to find a ν for which the conclusion of (iii) is false. The fact that $C_1 \bigoplus C_2$ implies that $C_1 \bigoplus C_2$ and we may pick a $\mu \in X^{\dagger}$ which is intersection positive on (C_1, C_2) . We then take $\nu = \mu$. To show that the conclusion of (iii) is false with this choice of ν , let us assume that there exist $C_1 \in C_1$ and $C_2 \in C_2$, $\omega_1 \in C_1^{\perp}$, $\omega_2 \in C_2^{\perp}$, $\omega_0 \ge 0$, for which the conditions $(\omega_0, \omega_1, \omega_2) \neq (0, 0, 0)$ and $\omega_1 + \omega_2 = \omega_0 \nu$ hold. If $\omega_0 = 0$, then $\omega_1 + \omega_2 = 0$, so $\omega_2 = -\omega_1$. Then $\omega_1 \neq 0$, and $\omega_1 \in C_1^{\perp} \cap (-C_2)^{\perp}$. So $C_1^{\perp} \cap (-C_2)^{\perp} \neq \{0\}$, and then C_1 and C_2 are not transversal, contradicting

the assumption that $C_1 \oiint C_2$. So $\omega_0 > 0$, and then we may assume that $\omega_0 = 1$. Then $\mu = \nu = \omega_1 + \omega_2$, $\omega_1 \in C_1^{\perp}$, and $\omega_2 \in C_2^{\perp}$. It follows that $\mu \in (C_1 \cap C_2)^{\perp}$. But then there cannot exist an $x \in C_1 \cap C_2$ for which $\mu(x) > 0$, and we have reached a contradiction. So $(iii) \Rightarrow (i)$.

Mordukhovich tangent multicones. Let M be a manifold of class C^1 , let S be a subset of M, and let $\bar{s} \in S$. The **Bouligand tangent cone** to S at \bar{s} is the set of all vectors $v \in T_{\bar{s}}M$ 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 $v = \lim_{j\to\infty} \frac{s_j-\bar{s}}{h_j}$. (This means that $v\varphi = \lim_{j\to\infty} \frac{\varphi(s_j)-\varphi(\bar{s})}{h_j}$ for every $\varphi \in C^1(M,\mathbb{R})$.) We use $T_{\bar{s}}^B S$ to denote the Bouligand tangent cone to S at \bar{s} . It is clear, and well known, that $T_{\bar{s}}^B S$ is a closed cone. The **Bouligand normal cone** of S at \bar{s} is the polar cone $(T_{\bar{s}}^B S)^{\perp}$ of $T_{\bar{s}}^B S$, that is, the set of all covectors $p \in T_{\bar{s}}^* M$ such that $\langle p, v \rangle \leq 0$ 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 $p \in T_{\bar{s}}^* M$ such that $p = \lim_{j\to\infty} p_j$ for some sequence $\{s_j\}_{j\in\mathbb{N}}$ of members of S that converges to \bar{s} and some sequence $\{p_j\}_{j\in\mathbb{N}}$ of members of T^*M such that $p_j \in (T_{s_j}^B S)^{\perp}$ (so in particular $p_j \in T_{s_i}^* M$) for each j.

We use $N_{\bar{s}}^{Mo}S$ to denote the Mordukhovich normal cone of S at \bar{s} . For each $p \in T_{\bar{s}}^*M$, we let $p^{\perp} = \{v \in T_{\bar{s}}M : \langle p, v \rangle \leq 0\}$, so p^{\perp} is a half space if $p \neq 0$, and p^{\perp} is the whole space $T_{\bar{s}}M$ if p = 0. The **Mordukhovich tangent** multicone to S at \bar{s} is the set $T_{\bar{s}}^{Mo}S \stackrel{\text{def}}{=} \{p^{\perp} : p \in N_{\bar{s}}^{Mo}S\}$, so $T_{\bar{s}}^{Mo}S$ is a set all whose members are closed half-spaces in $T_{\bar{s}}M$, except for one "trivial member," namely, the whole space $T_{\bar{s}}M$.

Lemma 3.6 Let M be a manifold of class C^1 , let S be a closed subset of M, and let $\bar{s} \in S$, $\bar{p} \in T^*_{\bar{s}}M$. Then the following conditions are equivalent:

 $\begin{array}{l} (*.1) \ \bar{p} \in N_{\bar{s}}^{Mo}S \ , \\ (*.2) \ \liminf_{s \to \bar{s}} \left(\max\{\langle \bar{p}, v \rangle : v \in T_{s}^{B}S, \, \|v\| \leq 1\} \right) = 0 \ , \\ (*.3) \ \liminf_{s \to \bar{s}, p \to \bar{p}} \left(\max\{\langle p, v \rangle : v \in T_{s}^{B}S, \, \|v\| \leq 1\} \right) = 0 \ . \end{array}$

Remark 3.7 Conditions (*.2) and (*.3) clearly make sense relative to any fixed coordinate chart \mathbf{x} near \bar{s} . (A chart is required to assign a meaning to $\langle \bar{p}, v \rangle$ when $v \in T_s^B S$, since s need not be equal to \bar{s} —so $v \in T_s^M$ while $\bar{p} \in T_{\bar{s}}^* M$ —and also to assign a meaning to ||v||.) However, it is easy to see that the truth values of (*.2) and (*.3) are independent of the choice of \mathbf{x} . \Box

Proof of Lemma 3.6. In view of Remark 3.7, we assume that $M = \mathbb{R}^m$ and $\bar{s} = 0$. We identify all the tangent spaces $T_s M$ with \mathbb{R}^m and all the cotangent

spaces T_s^*M with \mathbb{R}_m in the obvious way. For $s \in S, p \in \mathbb{R}_m$, let

$$\Theta^{S}(p,s) = \max\{\langle p, v \rangle : v \in T_{s}^{B}S, ||v|| \le 1\}.$$
(3)

Then $\Theta^{S}(p,s) \geq 0$, because $0 \in T_{s}^{B}S$.

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 \mathbb{R}_m such that $\lim_{j\to\infty} s_j = 0$, $\lim_{j\to\infty} p_j = \bar{p}$, and $p_j \in (T^B_{s_j}S)^{\perp}$ for each j. Then, $\Theta^S(p_j, s_j) = 0$ 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 exists a sequence $\{(s_j, p_j)\}_{j \in \mathbb{N}}$ of members of $S \times \mathbb{R}_m$ such that $\lim_{j\to\infty} s_j = 0$, $\lim_{j\to\infty} p_j = \bar{p}$, and $\lim_{j\to\infty} \Theta^S(p_j, s_j) = 0$. Since $\Theta^S(\bar{p}, s_j) \leq \Theta^S(p_j, s_j) + \|\bar{p} - p_j\|$, it follows that $\lim_{j\to\infty} \Theta^S(\bar{p}, s_j) = 0$, and then (*.2) holds.

We now assume that (*.2) holds, and prove (*.1). If $\bar{p} = 0$ then $\bar{p} \in N_0^{Mo}S$, 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 $\varepsilon_j = \Theta^S(\bar{p}, s_j)$. 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 $\mathbb{B}^m(s_j + \alpha \bar{p}, \beta)$ intersects S. (The minimum exists because S is closed.) Then $\beta_j(\alpha) \leq \alpha$, because s_j belongs to $\mathbb{B}^m(s_j + \alpha \bar{p}, \alpha)$.

We are going to construct, for each j, a covector $p_j \in \mathbb{R}_m$ 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 $\mathbb{B}^m(s_j + \alpha \bar{p}, \alpha)$ does not intersect S, and this clearly implies that $\bar{p} \in (T^B_{s_j}S)^{\perp}$. 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 may pick a member $\sigma(\alpha)$ of the set $\mathbb{B}^m(s_j + \alpha \bar{p}, \beta_j(\alpha)) \cap S$. Let

$$v(\alpha) = \sigma(\alpha) - s_j, \quad \pi(\alpha) = \alpha \bar{p} - v(\alpha).$$

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 $\langle \pi(\alpha), \bar{p} \rangle \leq \beta_j(\alpha)$, and then $\langle v(\alpha), \bar{p} \rangle \geq \alpha - \beta_j(\alpha)$, so $\beta_j(\alpha) \geq \alpha - \langle v(\alpha), \bar{p} \rangle$. On the other hand,

$$\limsup_{\alpha \downarrow 0} \|v(\alpha)\|^{-1} \langle v(\alpha), \bar{p} \rangle \le \varepsilon_j \,.$$

(Indeed, suppose this is not true. Then there exist a positive real number δ and a sequence $\{\alpha_k\}_{k\in\mathbb{N}}$ of positive numbers that converges to 0 and is such that $\|v(\alpha_k)\|^{-1}\langle v(\alpha_k), \bar{p}\rangle \geq \varepsilon_j + \delta$. If we define $w_k = \|v(\alpha_k)\|^{-1}v(\alpha_k)$, 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 $\langle w, \bar{p} \rangle \geq \varepsilon_j + \delta$, and this contradicts the fact that $\Theta^S(\bar{p}, s_j) = \varepsilon_j$.)

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 $||v(\alpha)|| \leq 2\alpha$. Then $0 \leq \langle v(\alpha), \bar{p} \rangle \leq \alpha \tilde{\varepsilon}_j$ whenever $0 < \alpha \leq \alpha^*$, where $\tilde{\varepsilon}_j = 2(\varepsilon_j + 2^{-j})$. Let

$$a(\alpha) = \langle v(\alpha), \bar{p} \rangle \bar{p}, \qquad b(\alpha) = v(\alpha) - a(\alpha),$$

so $b(\alpha) \perp a(\alpha)$, and then

$$||v(\alpha)||^{2} = ||a(\alpha)||^{2} + ||b(\alpha)||^{2}$$

On the other hand,

$$\pi(\alpha) = \alpha \bar{p} - v(\alpha) = \alpha \bar{p} - a(\alpha) - b(\alpha) ,$$

 \mathbf{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 can conclude that $\alpha - \langle v(\alpha), \bar{p} \rangle \geq \alpha (1 - \tilde{\varepsilon}_j)$, from which it clearly follows that

$$\alpha^2 \ge \alpha^2 (1 - \tilde{\varepsilon}_j)^2 + \|b(\alpha)\|^2.$$

Then

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

 \mathbf{SO}

$$\|b(\alpha)\| \le \alpha \sqrt{2\tilde{\varepsilon}_j}.$$

Therefore

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

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. \Box

The Clarke tangent and normal cones. If M is a manifold of class C^1 , S is a closed subset of M, and $\bar{s} \in M$, then the Clarke tangent cone to S at \bar{s} is the set of all vectors $v \in T_{\bar{s}}M$ 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 $v_j \in T_{s_j}^B S$ such that $\lim_{j\to\infty} v_j = v$. We use $T_{\bar{s}}^{Cl}S$ to denote the Clarke tangent cone to S at \bar{s} . It is well known that $T_{\bar{s}}^{Cl}S$ is a closed convex cone. The Clarke normal cone $N_{\bar{s}}^{Cl}S$ of S at \bar{s} is the polar $(T_{\bar{s}}^{Cl}S)^{\perp}$ of the Clarke tangent cone. Therefore $N_{\bar{s}}^{Cl}S$ is closed and convex. It is well-known that $N_{\bar{s}}^{Cl}S$ is the smallest closed convex cone in $T_{\bar{s}}^*M$ containing the Mordukhovich cone $N_{\bar{s}}^{Mo}S$. Therefore $T_{\bar{s}}^{Cl}S = \bigcap\{C: C \in T_{\bar{s}}^{Mo}S\}$.

WDC approximating multicones. If \mathcal{C} , \mathcal{D} are convex multicones, then we say that \mathcal{C} is a full submulticone of \mathcal{D} , and write $\mathcal{C} \leq_{full} \mathcal{D}$, if for every $D \in \mathcal{D}$ there exists a $C \in \mathcal{C}$ such that $C \subseteq D$.

If X, Y are fdrlss, \mathcal{C} is a multicone in X, and $\Lambda \subseteq Lin(X, Y)$, then we define $\Lambda \cdot \mathcal{C} \stackrel{\text{def}}{=} \{L \cdot C : L \in \Lambda, C \in \mathcal{C}\}.$

Definition 3.8 If M is a manifold of class C^1 , $\bar{s} \in S \subseteq M$, and C is a convex multicone in $T_{\bar{s}}M$, we say that C is a **WDC approximating multicone 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 set-valued map $U \ni u \mapsto F(u) \subseteq M$, (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) $F(K) \subseteq S$, (II) $\Lambda \in WDC(F; 0, \bar{s})$, (III) $\mathcal{D} \preceq_{full} T_{\bar{s}}^{Mo}K$, and, finally (IV) $\mathcal{C} = \Lambda \cdot \mathcal{D}$.

We will use $WDCAM(S, \bar{s})$ to denote the set of all WDC approximating multicones of S at \bar{s} , so " $\mathcal{C} \in WDCAM(S, \bar{s})$ " is an alternative way of saying that " \mathcal{C} is a WDC approximating multicone of S at \bar{s} ."

Example 3.9 If M is a manifold of class C^1 , S is a closed subset of M, $\bar{s} \in S$, and \mathcal{C} is any convex multicone in $T_{\bar{s}}M$ such that $\mathcal{C} \preceq T_{\bar{s}}^{Mo}S$, then $\mathcal{C} \in WDCAM(S, \bar{s})$. To prove this, it clearly suffices to assume that $M = \mathbb{R}^m$ and $\bar{s} = 0$. We 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 $K = V \cap S$, so K is compact and $T_0^{Mo}K = T_0^{Mo}S$. We then let $F: U \mapsto \mathbb{R}^m$ be the inclusion map, and take $\Lambda = \{\mathbb{I}_{\mathbb{R}^m}\}$. Then $\mathcal{C} = \Lambda \cdot \mathcal{C}$, and $\mathcal{C} \preceq T_0^{Mo}K$.

Example 3.10 As a corollary of Example 3.9, if S is a closed subset of a manifold M of class C^1 , and $\bar{s} \in S$, then the Mordukhovich multicone $T_{\bar{s}}^{Mo}S$ and the "Clarke multicone" $\{T_{\bar{s}}^{Cl}S\}$ are WDC approximating multicones of S at \bar{s} .

Example 3.11 It follows trivially from the definition that, if (a) for i = 1, 2, M_i is a manifold of class C^1 , $S_i \subseteq M_i$, and $\bar{s}_i \in S_i$ (b) F is a set-valued

map from M_1 to M_2 such that $F(S_1) \subseteq S_2$, (c) $\Lambda \in WDC(F; \bar{s}_1, \bar{s}_2)$, and (d) $\mathcal{C} \in WDCAM(S_1, \bar{s}_1)$, then $\Lambda \cdot \mathcal{C} \in WDCAM(S_2, \bar{s}_2)$.

The following example uses the Cartesian product of two multicones C_1 , C_2 in linear spaces X_1 , X_2 . We define $C_1 \times C_2 \stackrel{\text{def}}{=} \{C_1 \times C_2 : C_1 \in C_1, C_2 \in C_2\}$, so $C_1 \times C_2$ is a multicone in $X_1 \times X_2$, which is convex if C_1 and C_2 are convex.

Example 3.12 (The Cartesian product rule.) If (a) M_1 , M_2 are manifolds of class C^1 , (b) $\bar{s}_1 \in S_1 \subseteq M_1$ and $\bar{s}_2 \in S_2 \subseteq M_2$, (c) $C_1 \in WDCAM(S_1, \bar{s}_1)$ and $C_2 \in WDCAM(S_2, \bar{s}_2)$, (d) $S = S_1 \times S_2$, $\bar{s} = (\bar{s}_1, \bar{s}_2)$, and $C = C_1 \times C_2$, then $C \in WDCAM(S, \bar{s})$. To see this find, for i = 1, 2, an $n_i \in \mathbb{Z}_+$, a compact subset K_i of \mathbb{R}^{n_i} containing 0^{n_i} (where 0^{ν} is the origin of \mathbb{R}^{ν}), an open neighborhood U_i of K_i in \mathbb{R}^{n_i} , a set-valued map F_i from U_i to M_i such that $F_i(K_i) \subseteq S_i$, a derivate container $\Lambda_i \in WDC(F_i; 0^{n_i}, \bar{s}_i)$, and a convex multicone \mathcal{D}_i in \mathbb{R}^{n_i} such that $\mathcal{D}_i \preceq T_{0^{n_i}}^{M_0} K_i$ and $\mathcal{C}_i = \Lambda_i \cdot \mathcal{D}_i$. Define

$$n = n_1 + n_2,$$

$$U = U_1 \times U_2 \subseteq \mathbb{R}^{n_1} \times \mathbb{R}^{n_2} \sim \mathbb{R}^n,$$

$$K = K_1 \times K_2,$$

$$M = M_1 \times M_2,$$

$$\mathcal{D} = \mathcal{D}_1 \times \mathcal{D}_2,$$

$$\Lambda = \Lambda_1 \times \Lambda_2,$$

$$F = F_1 \times F_2,$$

where

$$\Lambda_1 \times \Lambda_2 \stackrel{\text{def}}{=} \{ L_1 \times L_2 : L_1 \in \Lambda_1, \, L_2 \in \Lambda_2 \} \,,$$

 $L_1 \times L_2$ is the map that sends a pair $(u_1, u_2) \in \mathbb{R}^{n_1} \times \mathbb{R}^{n_2}$ to the pair $(L_1 \cdot u_1, L_2 \cdot u_2) \in T_{\bar{s}_1} M_1 \times T_{\bar{s}_2} M_2 \sim T_{\bar{s}} M)$, and $F_1 \times F_2$ is the set-valued map that sends each $(u_1, u_2) \in U$ to the subset $F_1(u_1) \times F_2(u_2)$ of M.

Then $\Lambda \in WDC(F; 0^n, \bar{s})$, and $\Lambda \cdot \mathcal{D} = \mathcal{C}$. So the desired conclusion will follow if we show that $\mathcal{D} \preceq T_{0^n}^{Mo} K$. But this is trivial, because, if $p \in N_{0^n}^{Mo} K$, then $p = \lim_{j \to \infty} p_j$ for some sequence $\{(u^j, p^j)\}_{j \in \mathbb{N}}$ such that $u^j \to 0^n$, $u_j \in K$, and $p_j \in (T_{u^j}^B K)^{\perp}$. Write $u^j = (u_1^j, u_2^j)$, $p^j = (p_1^j, p_2^j)$, $u_i^j \in \mathbb{R}^{n_i}$, $p_i^j \in \mathbb{R}_{n_i}$. Then $T_{u_1^j}^B K_1 \times \{0^{n_2}\} \cup \{0^{n_1}\} \times T_{u_2^j}^B K_2 \subseteq T_{u^j}^B K$. Since $p^j \in (T_{u^j}^B K)^{\perp}$, it follows that $p_1^j \in (T_{u_1^j}^B K_1)^{\perp}$ and $p_2^j \in (T_{u_2^j}^B K_2)^{\perp}$. Hence $p_1 \in N_{0^{n_1}}^{Mo} K_1$, $p_2 \in N_{0^{n_2}}^{Mo} K_2$. Since $\mathcal{D}_i \preceq T_{0^{n_i}}^{Mo} K_i$ for i = 1, 2, we may pick $D_i \in \mathcal{D}_i$ such that $D_i \subseteq p_i^{\perp}$. Then $D_1 \times D_2 \subseteq p^{\perp}$ and $D_1 \times D_2 \in \mathcal{D}$. This shows that $\mathcal{D} \preceq T_{0^n}^{Mo} K$ and concludes our proof. \square

Remark 3.13 In the previous example, it is important to notice that the product $T_{0^{n_1}}^{Mo}K_1 \times T_{0^{n_2}}^{Mo}K_2$ of the Mordukhovich tangent multicones $T_{0^{n_1}}^{Mo}K_1$, $T_{0^{n_2}}^{Mo}K_2$ does not in general coincide with the Mordukhovich tangent multicone

 $T_{0^n}^{Mo}K$ of the product. (For example, if $M_1 = M_2 = \mathbb{R}$, and $K_1 = K_2 = [0, 1]$, then $T_0^{Mo}K_1 = T_0^{Mo}K_2 = \{\mathbb{R}, [0, +\infty[\}, \text{and } T_0^{Mo}K_1 \times T_0^{Mo}K_2 \text{ is the four-member} \text{set } \{\mathbb{R} \times \mathbb{R}, \mathbb{R} \times [0, +\infty[, [0, +\infty[\times \mathbb{R}, [0, +\infty[\times [0, +\infty[\times \mathbb{R}, [0, +\infty[\times [0$

$$H_{a,b} = \{(x, y) \in \mathbb{R}^2 : ax + by \ge 0\},\$$

for all the pairs (a, b) such that $a \leq 0$ and $b \leq 0$.) On the other hand, all that is needed for the proof in Example 3.12 is the fact that $T_0^{Mo}K_1 \times T_0^{Mo}K_2 \preceq T_{(0,0)}^{Mo}K$, which is true. \Box

The directional open mapping property. Given a subset A of \mathbb{R}^{ν} , 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 $\|\gamma(t) - \gamma(s)\| \leq r|t - s|$ 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}^{ν} , and $\alpha > 0$, we use $D(\alpha)$ to denote the set $\{y \in D : ||y|| \leq \alpha\}$. If $y \in \mathbb{R}^{\nu}$, we use σ_y to denote the set $\{ty : 0 \leq t \leq 1\}$. If $\gamma : [0, 1] \mapsto A$ is an arc, then $|\gamma|$ will denote the set $\{\gamma(t) : t \in [0, 1]\}$. Also, we use 0^{ν} to denote the origin of \mathbb{R}^{ν} .

Theorem 3.14 Assume that $m, n \in \mathbb{Z}_+$, S is a closed subset of \mathbb{R}^n , U is an open subset of \mathbb{R}^n , $0^n \in S \cap U$, \mathcal{F} is a set-valued map from U to \mathbb{R}^m , and Λ is a Warga derivate container of \mathcal{F} at $(0^n, 0^m)$. Let $\bar{y} \in \mathbb{R}^m$ be such that $\|\bar{y}\| = 1$ and $\bar{y} \in \text{Int } L \cdot p^{\perp}$ for every $L \in \Lambda$ and every $p \in N_0^{Mo}S$. Then there exist a closed convex cone D in \mathbb{R}^m such that $\bar{y} \in \text{Int } D$, positive numbers α , κ such that $\bar{\mathbb{B}}^n(0, \alpha \kappa) \subseteq U$, and a single-valued Lipschitz map $F : \bar{\mathbb{B}}^n(0, \alpha \kappa) \mapsto \mathbb{R}^m$ such that $F(x) \in \mathcal{F}(x)$ for every $x \in \bar{\mathbb{B}}^n(0, \alpha \kappa)$, having the property that

$$(\forall y \in D(\alpha))(\exists \gamma \in \Gamma(S, \alpha \kappa))(\sigma_y = |F \circ \gamma|).$$
(4)

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 $\mu = m - 1$. We then let $\mathcal{R} = \mathbb{R}^{\mu}$, and identify \mathbb{R}^{m} with $\mathcal{R} \times \mathbb{R}$.

Let Θ^S be the function defined in Equation (3) above. We show that

(#) There exists a real number $\delta \in]0,1[$ such that, whenever $q \in \mathbb{R}^m$, $L \in \mathbb{R}^{m \times n}, s \in S$ are such that $||q|| = 1, \langle q, \bar{y} \rangle \geq -\delta$, $\operatorname{dist}(L, \Lambda) \leq \delta$, $s \in S$, and $||s|| \leq \delta$, it follows that $\Theta^S(L^{\dagger}(q), s) \geq \delta$.

We prove (#) by contradiction. Assume that δ does not exist. Then there are sequences $\{\delta_j\}_{j\in\mathbb{N}}, \{q_j\}_{j\in\mathbb{N}}, \{L_j\}_{j\in\mathbb{N}}, \{s_j\}_{j\in\mathbb{N}}$, such that $\lim_{j\to\infty} \delta_j = 0$ and, for each j, the following are true:

$$\begin{split} \delta_j &> 0 \,, & \langle q_j, \bar{y} \rangle \geq -\delta_j \,, & L_j \in \mathbb{R}^{m \times n} \,, \\ q_j &\in \mathbb{R}^m \,, & s_j \in S \,, & \operatorname{dist}(L_j, \Lambda) \leq \delta_j \,, \\ \|q_j\| &= 1 \,, & \|s_j\| \leq \delta_j \,, & \Theta^S(L_j^{\dagger}(q_j), s_j) < \delta_j \,. \end{split}$$

Pick $\tilde{L}_j \in \Lambda$ such that $\|\tilde{L}_j - L_j\| \leq \delta_j$. Then we may pass to a subsequence and assume that the limit $(\bar{q}, \bar{L}) = \lim_{j \to \infty} (q_j, \tilde{L}_j)$ exists. Then $\|\bar{q}\| = 1$, $\langle \bar{q}, \bar{y} \rangle \geq 0$, and $\bar{L} \in \Lambda$. In addition, $\lim_{j \to \infty} s_j = 0$ and $\lim_{j \to \infty} L_j = \bar{L}$. Let

$$p_j = L_j^{\dagger} q_j, \qquad \bar{p} = \bar{L}^{\dagger} \bar{q},$$

so $\lim_{j\to\infty} p_j = \bar{p}$. Since $\Theta^S(p_j, s_j) < \delta_j$, it is clear that

$$\liminf_{s \to 0, p \to \bar{p}} \Theta^S(p, s) = 0 \,,$$

So Lemma 3.6 implies that $\bar{p} \in N_{0^n}^{Mo}S$. Hence \bar{y} is an interior point of $\bar{L} \cdot \bar{p}^{\perp}$. On the other hand, if $y \in \bar{L} \cdot \bar{p}^{\perp}$ then we can write $y = \bar{L} \cdot x, x \in \bar{p}^{\perp}$, 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 $\langle \bar{p}, x \rangle \leq 0$, since $x \in \bar{p}^{\perp}$. So $\langle \bar{q}, y \rangle \leq 0$ for all $y \in \bar{L} \cdot \bar{p}^{\perp}$. Since $\bar{y} \in \bar{L} \cdot \bar{p}^{\perp}$ and $\langle \bar{q}, \bar{y} \rangle \geq 0$, 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 $y \in \bar{L} \cdot \bar{p}^{\perp}$. So we have reached a contradiction, proving (#).

We now fix a δ having the properties of (#), choose $\kappa = \delta^{-1}$, and then let $\hat{\Lambda} = \{L \in \mathbb{R}^{m \times n} : \operatorname{dist}(L, \Lambda) \leq \delta\}$. We then use the definition of the Warga derivate container, and obtain

- an $R \in \mathbb{R}$ such that R > 0, $\overline{\mathbb{B}}^n(0, R) \subseteq U$ and $R \leq \delta$,
- a single-valued Lipschitz map $F : \overline{\mathbb{B}}^n(0, R) \mapsto \mathbb{R}^m$ such that F(0) = 0 and $F(x) \in \mathcal{F}(x)$ for every $x \in \overline{\mathbb{B}}^n(0, R)$,
- a sequence $\{F_j\}_{j\in\mathbb{N}}$ of functions of class C^1 from $\overline{\mathbb{B}}^n(0,R)$ to \mathbb{R}^m such that $F_j \to F$ uniformly on $\overline{\mathbb{B}}^n(0,R)$ as $j \to \infty$,
 - $DF_j(x) \in \hat{\Lambda}$ for all $x \in \overline{\mathbb{B}}^n(0, R), j \in \mathbb{N}$.

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 $\bar{y} \in \text{Int } D$. 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 (4). It clearly suffices to show that

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

(Indeed, if (5) 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 (5). We fix an index j, and write $G = F_j$. Then G(0) = 0, $G \in C^1(\overline{\mathbb{B}}^n(0, R), \mathbb{R}^m)$, and $DG(x) \in \widehat{\Lambda}$ for all $x \in \overline{\mathbb{B}}^n(0, R)$. We want to prove that $(\forall y \in D(\alpha))(\exists \gamma \in \Gamma(\widehat{S}, \kappa \alpha))(\sigma_y = |G \circ \gamma|)$.

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

$$(\forall y \in D_0(\alpha))(\exists \gamma \in \Gamma(\hat{S}, \kappa\alpha))(\sigma_y = |G \circ \gamma|).$$
(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 $|G \circ \gamma_{\varepsilon}|$ converge to σ_{y_*} in the Hausdorff metric. Pick a positive ε such that $\mathbb{B}^m(y_*, \varepsilon) \subseteq D_0(\alpha)$. (This implies, in particular, that $\|y_*\| + \varepsilon < \alpha$.) Then let

$$\hat{Q}_{\varepsilon} = \left\{ v \in \mathbb{R}^m : \langle v, y_* \rangle = 0 \land ||v|| \le \varepsilon \right\},\$$

so \hat{Q}_{ε} is the μ -dimensional disc orthogonal to y_* , centered at 0, and having radius ε . Define

$$Q_{\varepsilon} = \{y_* + v : v \in Q_{\varepsilon}\},\$$

so $Q_{\varepsilon} \subseteq \overline{\mathbb{B}}^m(y_*, \varepsilon)$.

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 $\langle y, \bar{y} \rangle \geq 0$ for all ynear 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 then define

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

Then $\mathcal{Q}_{\varepsilon}$ is obviously closed, and $\mathcal{Q}_{\varepsilon} \neq \emptyset$, because $0 \in \mathcal{Q}_{\varepsilon}$. Furthermore, the Hausdorff distance $d_{Ha}(\mathcal{Q}_{\varepsilon}, \sigma_{y_*})$ is exactly ε . (Indeed, fix an $x \in \mathcal{Q}_{\varepsilon}$. 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 $\lambda_{\varepsilon} \|v\|^2 \leq r \leq \|y_*\|$, so $r \geq 0$, and then ry_* belongs to σ_{y_*} and $\|x - r\hat{y}\|^2 \leq \varepsilon^2$, so $\|x - r\hat{y}\| \leq \varepsilon$. 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 $\max\{\operatorname{dist}(x, \sigma_{y_*}) : x \in \mathcal{Q}_{\varepsilon}\} = \varepsilon$. 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 the ball $\mathbb{B}^{n}(0, R)$ to \mathbb{R}^{n} by letting $\Psi_{\varepsilon}(s) = \{w \in \mathbb{R}^{n} : ||w|| \leq 1 \text{ and } \langle \nabla H_{\varepsilon}(s), w \rangle \geq \delta \}$. It is then clear that the map Ψ_{ε} is upper semicontinuous with compact convex values.

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

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

To see this, pick a point $s_* \in S'_{0,\varepsilon}$, and write

$$x_* = G(s_*), \qquad \pi_* = \nabla h_{\varepsilon}(x_*), \qquad \hat{\pi}_* = \frac{\pi_*}{\|\pi_*\|}$$

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 $||v_*|| \leq \varepsilon$. The fact that $s_* \in S'_{0,\varepsilon}$ then implies the inequalities $||v_*|| < \varepsilon$ and $0 \leq r_* < ||y_*||$. 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 $v_* \perp \hat{y}$. Also, $\langle \pi_*, \bar{y} \rangle = \langle \hat{y}, \bar{y} \rangle - 2\lambda_{\varepsilon} \langle v_*, \bar{y} \rangle$. Since $\hat{y} \in D$, and $\|\hat{y}\| = 1$, we have $\langle \hat{y}, \bar{y} \rangle \geq 1 - \tilde{\delta}$, 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 $\|\hat{y} - \bar{y}\| \leq \sqrt{2\tilde{\delta}} = \delta$, so that

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

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

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

To complete the proof of (8), we have to show that $w \in T_s^B S'_{\varepsilon}$. Since $w \in T_s^B S$ and ||w|| = 1, we can find a sequence $\{s_k\}_{\in\mathbb{N}}$ of points of $S \setminus \{s_*\}$ that converges to s_* and is such that $\lim_{k\to\infty} w_k = w$, where $w_k = \frac{s_k - s_*}{\|s_k - s_*\|}$.

If we let

$$\omega_k = \|s_k - s_*\|, \qquad \tilde{w}_k = w_k - w,$$

we find

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

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

$$\|G(s) - G(s_*) - L_*(s - s_*)\| \le \psi(\|s - s_*\|)\|s - s_*\|$$
(9)

$$|h_{\varepsilon}(x) - h_{\varepsilon}(x_{*}) - \langle \pi_{*}, x - x_{*} \rangle| \le \psi(||x - x_{*}||) ||x - x_{*}||$$
(10)

for all $s \in U$ and all $x \in \mathbb{R}^m$, respectively. Let $x_k = G(s_k)$. Then (9) implies the inequality

$$||x_k - x_* - \omega_k L_*(w + \tilde{w}_k)|| \le \psi(\omega_k)\omega_k \,,$$

from which it follows that

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

where $\nu_k = \psi(\omega_k) + \|L_*(\tilde{w}_k)\|$, so that $\lim_{k\to\infty} \nu_k = 0$. It then follows that

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

Then

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

Therefore

$$\langle x_k - x_*, \pi_* \rangle = \langle x_k - x_* - \omega_k L_*(w), \pi_* \rangle + \omega_k \langle L_*(w), \pi_* \rangle$$

$$\geq -\omega_k \nu_k \|\pi_*\| + \omega_k \langle w, L_*^{\dagger}(\pi_*) \rangle \geq \omega_k (\delta - \nu_k \|\pi_*\|) .$$

Now write $\nu'_k = \psi(||x_k - x_*||)$, so that $\lim_{k \to \infty} \nu'_k = 0$. It then follows that

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

from which we can conclude that

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

Using of the facts that

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

we find that

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

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

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

It follows from (7) that $x \in \mathcal{Q}_{\varepsilon}$ if and only if $h_{\varepsilon}(x) \geq 0$ and $\langle x, \hat{y} \rangle \leq ||y_*||$. Since $x_* \in \mathcal{Q}_{\varepsilon}$, the inequality $h_{\varepsilon}(x_*) \geq 0$ is true, and then $h_{\varepsilon}(x_k) > 0$ if $k \geq \bar{k}$. 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 $\bar{k}' \geq \bar{k}$ and $\langle x_k, \hat{y} \rangle < ||y_*||$ whenever $k \geq \bar{k}'$. Then, if $k \geq \bar{k}'$, the the following hold:

(i) $s_k \in S$, (ii) $h_{\varepsilon}(x_k) > 0$, (iii) $\langle x_k, \hat{y} \rangle < ||y_*||$, (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).

Now, using standard existence results from viability theory, we pick a solution $\xi_{\varepsilon}: I_{\varepsilon} \mapsto S'_{0,\varepsilon}$ of the differential inclusion $\dot{\xi}(t) \in \Psi_{\varepsilon}(\xi(t))$ such that

- (1) $\xi_{\varepsilon}(0) = 0$,
- (2) ξ_{ε} is defined on a subinterval I_{ε} of \mathbb{R} such that $0 = \min I_{\varepsilon}$,
- (3) 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 $I_{\varepsilon} \neq \tilde{I}$.

Then ξ_{ε} satisfies $H_{\varepsilon}(\xi_{\varepsilon}(t)) \geq \delta t$ 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 easily excluded, because $\|\bar{s}_{\varepsilon}\| \leq \tau_{\varepsilon} \leq \delta^{-1} \|y_*\| < \delta^{-1}\alpha = R$. 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 $\tau_{\varepsilon} \leq \delta^{-1}\alpha = \kappa \alpha$), and $\gamma_{\varepsilon}(0) = 0$. Furthermore, the set $|G \circ \gamma_{\varepsilon}|$ is entirely contained in $\mathcal{Q}_{\varepsilon}$, and $G(\gamma_{\varepsilon}(1)) \in \mathcal{Q}_{\varepsilon}$. 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 $|G \circ \gamma| \subseteq \sigma_{y_*}$. Furthermore, $y_* = \lim_{k \to \infty} x_{\varepsilon_k}$, so $y_* \in |G \circ \gamma|$, and then $|G \circ \gamma| = \sigma_{y_*}$. This concludes the proof.

The transversal intersection property. If X is a topological space, and S_1, S_2 are subsets of X, we say that S_1 and S_2 are **locally separated** at a point $p \in X$ if there exists a neighborhood U of p in X such that $S_1 \cap S_2 \cap U \subseteq \{p\}$.

Theorem 3.15 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 C_1 , C_2 , be WDC approximating multicones of S_1 , S_2 at \bar{x} . Assume that C_1 and 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 $(S_1 \cap S_2) \setminus \{\bar{x}\}$ 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. We will use Theorem 3.14. 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 $\mu : X \mapsto \mathbb{R}$ which is intersection positive on $(\mathcal{C}_1, \mathcal{C}_2)$, 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

$$S = S_1 \times S_2, \qquad \mathcal{C} = \mathcal{C}_1 \times \mathcal{C}_2, \qquad \mathcal{D} = G \cdot \mathcal{C}.$$

Since Example 3.12 tells us that C is a WDC approximating multicone of S at (0,0), it follows that \mathcal{D} is a WDC approximating multicone 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 $\bar{y} \in \text{Int } D$ 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 $C_1 - C_2 = X$. Since μ is intersection positive on $(\mathcal{C}_1, \mathcal{C}_2)$, 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 $c_1 \in C_1$, $c_2 \in \mathcal{C}_2$. So there exists $r \in \mathbb{R}$ such that $(v, r) \in G(C_1 \times C_2)$. If (e_1, \ldots, e_n)

is a basis of X, and $e_0 = -(e_1 + \ldots + e_n)$, then there are real numbers r_i such that $(e_i, r_i) \in G(C_1 \times C_2)$ for $i = 0, \ldots, n$. 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 $\bar{r} = \max(1, r_0, r_1, \ldots, r_n)$. 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 3.14. It then follows from the theorem that, for some positive number α , there exists a Lipschitz arc $\xi:[0,1] \mapsto S$ that satisfies $\xi(0) = 0$ and is such that the sets $\{G(\xi(t)) : t \in [0,1]\},$ $\{(0,r): 0 \leq r \leq \alpha\}$ coincide. Write $\xi(t) = (\xi_1(t), \xi_2(t)), \text{ so } \xi_1(t) \in S_1 \text{ and} \xi_2(t)) \in S_2$. 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 $\mu(\gamma(t)) = \alpha$. \Box

4 Covariant differentiation and adjoint covectors

Ppd vector fields, trajectories, flow maps. We assume that

(A3) $m, \mu \in \mathbb{Z}_+$, M is a manifold of class C^{μ} , $\mu \ge 1$, and $m = \dim M$.

A **ppd tvvf on** M is a ppd map $M \times \mathbb{R} \ni (x,t) \hookrightarrow f(x,t) \in TM$ such that $f(x,t) \in T_x M$ whenever $(x,t) \in \text{dom } f$. A **trajectory**, or **integral curve**, of a ppd tvvf f on M is a locally absolutely continuous map $\xi : I \mapsto M$, defined on a nonempty real interval I, such that for almost all $t \in I$ the following two conditions hold: (i) $(\xi(t),t) \in \text{dom}(f)$, and (ii) $\dot{\xi}(t) = f(\xi(t),t)$. An *integral arc* of f is an integral curve $\xi : I \mapsto X$ such that the interval I is compact. If f is a ppd tvvf on M, then Traj(f) (resp. $\text{Traj}_c(f)$) will denote the set of all integral curves (resp. arcs) of f. For given $t, s \in \mathbb{R}$, the **time** t **to time** s **flow map** of f is the set-valued map $\Phi_{s,t}^f$ from M to M that assigns to each $x \in \mathbb{R} \times \mathbb{R} \times X$ the set $\Phi_{s,t}^f(x) \stackrel{\text{def}}{=} \{\xi(s) : \xi \in \text{Traj}_c(f), \xi(t) = x\}.$

Vector fields and covector fields along an arc. We assume (A3)

If $\xi \in \mathcal{ARC}(M)$, then we can consider the pullback bundles ξ^*TM , ξ^*T^*M . If dom $\xi = [a, b]$ then, by definition, ξ^*TM , ξ^*T^*M are the bundles over [a, b] whose fibers $(\xi^*TM)_t$, $(\xi^*T^*M)_t$ at a $t \in [a, b]$ are the spaces $T_{\xi(t)}M$. $T^*_{\xi(t)}M$. We use $\Gamma(\xi^*TM)$, $\Gamma(\xi^*T^*M)$ to denote, respectively, the set of all sections $[a, b] \ni t \mapsto v(t) \in T_{\xi(t)}M$, $[a, b] \ni t \mapsto w(t) \in T^*_{\xi(t)}M$, of ξ^*TM , ξ^*T^*M . The members of $\Gamma(\xi^*TM)$, $\Gamma(\xi^*T^*M)$ are called, respectively, *vector fields along* ξ and *covector fields along* ξ . If the arc ξ is such that $\xi([a, b]) \subseteq \text{dom } \mathbf{x}$ for some chart \mathbf{x} of M, then ξ has a coordinate representation $\xi^{\mathbf{x}} \in C^0([a, b], \text{im } \mathbf{x})$, given by $\xi^{\mathbf{x}}(t) = \xi(t)^{\mathbf{x}} = \mathbf{x}(\xi(t))$ for $t \in [a, b]$. Also, vector fields $v \in \Gamma(\xi^*(TM))$ and covector fields $w \in \Gamma(\xi^*(T^*M))$ have coordinate representations $v^{\mathbf{x}}$, $w^{\mathbf{x}}$, which are, respectively, maps from [a, b] to \mathbb{R}^m and from [a, b] to \mathbb{R}_m , given by

$$v^{\mathbf{x}}(t) = v(t)^{\mathbf{x}} = (v^{\mathbf{x},1}(t), v^{\mathbf{x},2}(t), \dots, v^{\mathbf{x},m}(t))^{\dagger}$$

$$w^{\mathbf{x}}(t) = w(t)^{\mathbf{x}} = (w_{1}^{\mathbf{x}}(t), w_{2}^{\mathbf{x}}(t), \dots, w_{m}^{\mathbf{x}}(t)),$$

where

$$v^{\mathbf{x},i}(t) = \langle dx^i(\xi(t)), v(t) \rangle, \qquad w_i^{\mathbf{x}}(t) = \langle w(t), \partial_i^{\mathbf{x}}(\xi(t)) \rangle.$$

Since M is, in particular, a manifold of class C^1 , it makes sense to talk about an arc $\xi \in C^0([a, b], M)$ being **absolutely continuous**, and we will use $W^{1,1}([a, b], M)$ to denote the space of all absolutely continuous maps $\xi : [a, b] \mapsto M$. (Precisely, the members of $W^{1,1}([a, b], M)$ are the arcs $\xi \in C^0([a, b], M)$ with the property that, for every chart \mathbf{x} of M and every compact subinterval $[\alpha, \beta]$ of [a, b] with the property that $\xi([\alpha, \beta]) \subseteq \operatorname{dom} \mathbf{x}$, the restriction $\tilde{\xi} = \xi [\alpha, \beta] \to \mathbb{R}^m$, given by $\tilde{\xi}^{\mathbf{x}} = \mathbf{x}^{-1} \circ \tilde{\xi}$, is absolutely continuous.) We write

$$W^{1,1}(M) = \bigcup_{-\infty < a \le b < +\infty} W^{1,1}([a, b], M).$$

If $\xi \in \mathcal{ARC}(M)$, then it makes sense to talk about vector fields and covector fields along ξ being **measurable**, or **continuous**, since TM and T^*M are topological spaces. Furthermore, if $1 \leq p \leq \infty$, it also makes sense to talk about vector fields and covector fields along ξ belonging to L^p , since TM and T^*M are vector bundles. (Precisely, a vector field $v \in \Gamma(\xi^*(TM))$ belongs to L^p if, for every chart **x** of M and every compact subinterval $[\alpha, \beta]$ of [a, b]such that $\xi([\alpha,\beta]) \subseteq \operatorname{dom} \mathbf{x}$, the restriction $\tilde{v} = v [[\alpha,\beta]]$ of v to $[\alpha,\beta]$ is such that the coordinate representation $\tilde{v}^{\mathbf{x}} : [\alpha, \beta] \mapsto \mathbb{R}^m$, given by the formula $\tilde{v}^{\mathbf{x}}(t) = D\mathbf{x}(\xi^{\kappa}(t))(v(t))$ for $t \in [\alpha, \beta]$, belongs to $L^p([\alpha, \beta], \mathbb{R}^m)$. Observe that, if \mathbf{x}, \mathbf{x}' are charts of M and $\xi([\alpha, \beta]) \subseteq \operatorname{dom} \mathbf{x} \cap \operatorname{dom} \mathbf{x}'$, then the vectors $\tilde{v}^{\mathbf{x}'}(t), \tilde{v}^{\mathbf{x}}(t)$ and related by $\tilde{v}^{\mathbf{x}'}(t) = D(\mathbf{x}' \circ \mathbf{x}^{-1})(\xi^{\mathbf{x}}(t)) \cdot \tilde{v}^{\mathbf{x}}(t)$ for $t \in [\alpha, \beta]$, and this implies that if $\tilde{v}^{\mathbf{x}}$ belongs to L^p then $\tilde{v}^{\mathbf{x}'}$ is in L^p as well, since the matrixvalued function $[\alpha, \beta] \ni t \mapsto D(\mathbf{x}' \circ \mathbf{x}^{-1})(\xi^{\mathbf{x}}(t))$ is continuous, because $\mu \geq 1$. A similar observation applies to covector fields, for which the transformation law is $\tilde{w}^{\mathbf{x}}(t) = \tilde{w}^{\mathbf{x}'}(t) \cdot D(\mathbf{x}' \circ \mathbf{x}^{-1})(\xi^{\mathbf{x}}(t)).)$ We use $\Gamma_{meas}(\xi^*TM), \Gamma_{C^0}(\xi^*TM),$ $\Gamma_{L^p}(\xi^*TM)$, to denote the spaces of all $v \in \Gamma(\xi^*TM)$ that are, respectively, measurable, continuous, members of L^p . The three spaces $\Gamma_{meas}(\xi^*T^*M)$, $\Gamma_{C^0}(\xi^*T^*M)$, and $\Gamma_{L^p}(\xi^*T^*M)$ are defined in a similar way.

Remark 4.1 The concept of an absolutely continuous vector or covector field along a $\xi \in \mathcal{ARC}(M)$ is not intrinsically defined if M is only of class C^1 , even if $\xi \in \mathcal{W}^{1,1}(M)$. Indeed, if \mathbf{x}, \mathbf{x}' are charts whose domains contain im ξ , then the coordinate representations $v^{\mathbf{x}}, v^{\mathbf{x}'}$, of a $v \in \Gamma(\xi^*TM)$ are related by $v^{\mathbf{x}'}(t) = J(\xi^{\mathbf{x}}(t)) \cdot v^{\mathbf{x}}(t)$, where $J(q) = D(\mathbf{x}' \circ \mathbf{x}^{-1})(q)$ for $q \in \mathrm{im} \mathbf{x}$. Since the map J is only known to be continuous, there is no reason why $v^{\mathbf{x}'}$ should be absolutely continuous if $v^{\mathbf{x}}$ is, even if $\xi^{\mathbf{x}}$ is absolutely continuous. A similar fact is true of covector fields.

On the other hand, it is clear that, the absolute continuity of $v^{\mathbf{x}'}$ follows from that of $v^{\mathbf{x}}$ if the map J is of class C^1 and $\xi \in \mathcal{W}^{1,1}(M)$. Hence the concept of absolute continuity makes sense for members of $\Gamma(\xi^*TM)$ and $\Gamma(\xi^*T^*M)$ if $\xi \in \mathcal{W}^{1,1}(M)$ and $\mu \geq 2$. (Alternatively, a member of $\Gamma(\xi^*TM)$ or of $\Gamma(\xi^*T^*M)$ is just a map from dom ξ to TM or to T^*M which is a lift of ξ , i.e. is such that its composite with the projection from TM or T^*M to M is ξ . If $\mu \geq 2$, then TM and T^*M are manifolds of class C^1 , so the spaces $\mathcal{W}^{1,1}(TM)$ and $\mathcal{W}^{1,1}(T^*M)$ are intrinsically defined.) \Box

Integrably Lipschitz ppd vector fields near an arc. We assume that

(A4) $m, \mu \in \mathbb{Z}_+, M$ is a manifold of class $C^{\mu}, \mu \geq 2, m = \dim M,$ and $\xi \in \mathcal{ARC}(M)$.

A chart covering of ξ is a finite set \mathcal{K} such that (a) all the members of \mathcal{K} are ordered pairs (I, \mathbf{x}) consisting of a compact subinterval I of dom ξ and a chart \mathbf{x} of M such that $\xi(I) \subseteq \operatorname{dom} \mathbf{x}$, and (b) $\bigcup \{I : (\exists \mathbf{x})(I, \mathbf{x}) \in \mathcal{K}\} = \operatorname{dom} \xi$.

If $M \times \mathbb{R} \ni (x,t) \hookrightarrow f(x,t) \in T_x M$ is a ppd tvvf on M, \mathcal{K} is a chart covering of ξ , and $k : \operatorname{dom} \xi \mapsto [0, +\infty]$ is an integrable function, then a **system of IL** constants for ξ , f, k, \mathcal{K} is an ordered pair $(\{\delta_{I,\mathbf{x}}\}_{(I,\mathbf{x})\in\mathcal{K}}, \{C_{I,\mathbf{x}}\}_{(I,\mathbf{x})\in\mathcal{K}})$ of families of positive constants such that the following three conditions are satisfied for every $(I, \mathbf{x}) \in \mathcal{K}$ and every $t \in I$:

- (IL.1) The ball $\overline{\mathbb{B}}^m(\xi(t)^{\mathbf{x}}, \delta_{I,\mathbf{x}})$ is contained in im \mathbf{x} .
- (IL.2) f(x,t) is defined whenever $x \in \mathbf{x}^{-1}(\bar{\mathbb{B}}^m(\xi(t)^{\mathbf{x}}, \delta_{I,\mathbf{x}}))$.
- (IL.3) $||f^{\mathbf{x}}(q,t)|| \leq C_{I,\mathbf{x}}k(t) \text{ and } ||f^{\mathbf{x}}(q,t) f^{\mathbf{x}}(q',t)|| \leq C_{I,\mathbf{x}}k(t) \cdot ||q q'||$ for all $q, q' \in \overline{\mathbb{B}}^m(\xi(t)^{\mathbf{x}}, \delta_{I,\mathbf{x}}).$

If f, k, \mathcal{K} are as above, then f is said to be *integrably Lipschitz near* ξ with bound k (abbreviated "IL-k near ξ ") relative to \mathcal{K} if there exists a system of IL constants for ξ , f, k, \mathcal{K} .

The following observation will be important later:

(IL.*) Suppose that f is a ppd tvvf on M, $k : \operatorname{dom} \xi \mapsto [0, +\infty]$ is integrable, and \mathcal{K} is chart covering of ξ such that f is IL-k near ξ relative to \mathcal{K} . Then f is IL-k near ξ relative to every chart covering of ξ .

To prove (IL.*), we start by fixing a system $(\{\delta_{I,\mathbf{x}}\}_{(I,\mathbf{x})\in\mathcal{K}}, \{C_{I,\mathbf{x}}\}_{(I,\mathbf{x})\in\mathcal{K}})$ of IL constants for ξ , f, k, \mathcal{K} . Let $\tilde{\mathcal{K}}$, be another chart covering of ξ . We will show that f is IL-k near ξ relative to $\tilde{\mathcal{K}}$.

Construct a new pair of chart coverings $\mathcal{K}^{\#}$, $\tilde{\mathcal{K}}^{\#}$ of the arc ξ by letting $\mathcal{K}^{\#} = \bigcup \{ \Sigma_{I,\mathbf{x}} : (I,\mathbf{x}) \in \mathcal{K} \}$ and $\tilde{\mathcal{K}}^{\#} = \bigcup \{ \tilde{\Sigma}_{J,\mathbf{y}} : (J,\mathbf{y}) \in \tilde{\mathcal{K}} \}$, where $\Sigma_{I,\mathbf{x}}$ is the set of all pairs (L,\mathbf{x}) , for all intervals L such that $\emptyset \neq L$ and $L = I \cap J$ for some $(J,\mathbf{y}) \in \tilde{\mathcal{K}}$, and $\tilde{\Sigma}_{J,\mathbf{y}}$ is the set of all pairs (L,\mathbf{x}) , for all intervals L such that $\emptyset \neq L$ and $L = I \cap J$ such that $\emptyset \neq L$ and $L = I \cap J$ for some $(I,\mathbf{x}) \in \mathcal{K}$.

It is then clear that f is IL-k near ξ relative to $\mathcal{K}^{\#}$. Indeed, if $\alpha \in \mathcal{K}^{\#}$, then $\alpha \in \Sigma_{I,\mathbf{x}}$ for some $(I,\mathbf{x}) \in \mathcal{K}$, so $\alpha = (L,\mathbf{x})$ for some nonempty compact interval L such that $L \subseteq I$. Then, if we let $\delta_{L,\mathbf{x}}^{\#} = \delta_{I,\mathbf{x}}$ and $C_{L,\mathbf{x}}^{\#} = C_{I,\mathbf{x}}$, it follows that $(\{\delta_{L,\mathbf{x}}^{\#}\}_{(L,\mathbf{x})\in\mathcal{K}^{\#}}, \{C_{L,\mathbf{x}}^{\#}\}_{(L,\mathbf{x})\in\mathcal{K}^{\#}})$ is a system of IL constants for ξ , $f, k, \mathcal{K}^{\#}$.

It is also clear that f will be IL-k near ξ relative to $\tilde{\mathcal{K}}$ as long as it is IL-k near ξ relative to $\tilde{\mathcal{K}}^{\#}$. (Indeed, let $(\{\tilde{\delta}_{H,\mathbf{y}}^{\#}\}_{(H,\mathbf{y})\in\tilde{\mathcal{K}}^{\#}}, \{\tilde{C}_{H,\mathbf{y}}^{\#}\}_{(H,\mathbf{y})\in\tilde{\mathcal{K}}^{\#}})$ be a system of IL constants for ξ , f, k, $\tilde{\mathcal{K}}^{\#}$. Then, for each member (J, \mathbf{y}) of $\tilde{\mathcal{K}}$, we define

$$\tilde{\delta}_{J,\mathbf{y}} = \min\{\tilde{\delta}_{H,\mathbf{y}}^{\#} : (H,\mathbf{y}) \in \tilde{\mathcal{K}}^{\#}, \ H \subseteq J\},\$$
$$\tilde{C}_{J,\mathbf{y}} = \max\{\tilde{C}_{H,\mathbf{y}}^{\#} : (H,\mathbf{y}) \in \tilde{\mathcal{K}}^{\#}, \ H \subseteq J\}.$$

It is then easy to verify that $({\{\tilde{\delta}_{J,\mathbf{y}}\}}_{(J,\mathbf{y})\in\tilde{\mathcal{K}}}, {\{\tilde{C}_{J,\mathbf{y}}\}}_{(J,\mathbf{y})\in\tilde{\mathcal{K}}})$ is a system of IL constants for ξ , f, k, $\tilde{\mathcal{K}}$.)

So all we need is to verify that f is IL-k near ξ relative to $\tilde{\mathcal{K}}^{\#}$. For this purpose, we fix a system $(\{\delta_{H,\mathbf{x}}^{\#}\}_{(H,\mathbf{x})\in\mathcal{K}^{\#}}, \{C_{H,\mathbf{x}}^{\#}\}_{(H,\mathbf{x})\in\mathcal{K}^{\#}})$ of IL constants for $\xi, f, k, \mathcal{K}^{\#}$. We then pick a member (H, \mathbf{y}) of $\tilde{\mathcal{K}}^{\#}$, and observe that we may pick a chart \mathbf{x} of M such that $(H, \mathbf{x}) \in \mathcal{K}^{\#}$. This implies, in particular, that $\xi(H) \subseteq \operatorname{dom} \mathbf{x} \cap \operatorname{dom} \mathbf{y}$.

We then pick a positive number δ such that the inclusions

$$\bar{\mathbb{B}}^m(\xi(t)^{\mathbf{y}},\delta) \subseteq \operatorname{im} \mathbf{y}, \qquad \mathbf{y}^{-1}(\bar{\mathbb{B}}^m(\xi(t)^{\mathbf{y}},\delta)) \subseteq \mathbf{x}^{-1}(\bar{\mathbb{B}}^m(\xi(t)^{\mathbf{x}},\delta_{H,\mathbf{x}}^{\#}))$$

hold whenever $t \in H$. (The existence of δ is proved as follows. Suppose first that there are arbitrarily small positive δ having the property that the inclusion $\overline{\mathbb{B}}^m(\xi(t)^{\mathbf{y}}, \delta) \subseteq \operatorname{im} \mathbf{y}$ is false for some $t \in H$. Then we can pick positive $\delta_j, t_j \in H$, and $q_j \in \overline{\mathbb{B}}^m(\xi(t_j)^{\mathbf{y}}, \delta_j)$, such that $\delta_j \to 0$ and $q_j \notin \operatorname{im} \mathbf{y}$. By passing to a subsequence, we may assume that $t_j \to t$ for some $t \in H$. But then $\xi(t_j) \to \xi(t)$, so $\xi(t_j)^{\mathbf{y}} \to \xi(t)^{\mathbf{y}}$. Since $\|q_j - \xi(t_j)^{\mathbf{y}}\| \leq \delta_j$, it follows that $q_j \to \xi(t)^{\mathbf{y}}$. But $\xi(t)^{\mathbf{y}} \in \operatorname{im} \mathbf{y}$, and im \mathbf{y} is open in \mathbb{R}^m , so $q_j \in \operatorname{im} \mathbf{y}$ is open in \mathbb{R}^m for large enough j, contradicting our choice of the q_j . We now know that there exists a positive $\overline{\delta}$ such that $\overline{\mathbb{B}}^m(\xi(t)^{\mathbf{y}}, \overline{\delta}) \subseteq \operatorname{im} \mathbf{y}$ for every $t \in H$. Suppose that there are arbitrarily small values of $\delta \in [0, \overline{\delta}]$ for which the inclusion $\mathbf{y}^{-1}(\overline{\mathbb{B}}^m(\xi(t)^{\mathbf{y}}, \delta)) \subseteq \mathbf{x}^{-1}(\overline{\mathbb{B}}^m(\xi(t)^{\mathbf{x}}, \delta_{H,\mathbf{x}}^{\#}))$ fails for some $t \in H$. Then we can pick $\delta_j \in [0, \overline{\delta}], t_j \in H$, and $x_j \in \mathbf{y}^{-1}(\overline{\mathbb{B}}^m(\xi(t_j)^{\mathbf{y}}, \delta_j))$ such that $\delta_j \to 0$ and $x_j \notin \mathbf{x}^{-1}(\bar{\mathbb{B}}^m(\xi(t_j)^{\mathbf{x}}, \delta_{H,\mathbf{x}}^{\#}))$. By passing to a subsequence, we may assume that $t_j \to t$ for some $t \in H$. But then $\xi(t_j) \to \xi(t)$, so $\xi(t_j)^{\mathbf{y}} \to \xi(t)^{\mathbf{y}}$ and $\xi(t_j)^{\mathbf{x}} \to \xi(t)^{\mathbf{x}}$. Let $q_j = x_j^{\mathbf{y}}$. Then $||q_j - \xi(t_j)^{\mathbf{y}}|| \leq \delta_j$, so $q_j \to \xi(t)^{\mathbf{y}}$. Therefore $x_j \to \xi(t)$, and then $x_j^{\mathbf{x}} \to \xi(t)^{\mathbf{x}}$. Since $\xi(t_j)^{\mathbf{x}} \to \xi(t)^{\mathbf{x}}$, it follows that $||x_j^{\mathbf{x}} - \xi(t_j)^{\mathbf{x}}|| \to 0$. Hence $x_j^{\mathbf{x}} \in \bar{\mathbb{B}}^m(\xi(t_j)^{\mathbf{x}}, \delta_{H,\mathbf{x}}^{\#})$ if j is large enough, so $x_j \in \mathbf{x}^{-1}(\bar{\mathbb{B}}^m(\xi(t_j)^{\mathbf{x}}, \delta_{H,\mathbf{x}}^{\#}))$ if j is large enough, contradicting our choice of the x_j .)

We take $\tilde{\delta}_{H,\mathbf{y}}^{\#}$ to be the number δ chosen above. Then, if $t \in H$ and q belongs to $\mathbb{B}^{m}(\xi(t)^{\mathbf{y}}, \tilde{\delta}_{H,\mathbf{y}}^{\#})$, it follows that $q \in \operatorname{im} \mathbf{y}$, so we may pick $x \in \operatorname{dom} \mathbf{y}$ for which $x^{\mathbf{y}} = q$. Then $x \in \mathbf{y}^{-1}(\mathbb{B}^{m}(\xi(t)^{\mathbf{y}}, \tilde{\delta}_{H,\mathbf{y}}))$, so $x \in \mathbf{x}^{-1}(\mathbb{B}^{m}(\xi(t)^{\mathbf{x}}, \delta_{H,\mathbf{x}}^{\#}))$. Then f(x, t) is defined. Furthermore, $\|f^{\mathbf{x}}(x^{\mathbf{x}}, t)\| \leq C_{H,\mathbf{x}}^{\#}k(t)$. Therefore

$$||f^{\mathbf{y}}(q,t)|| = ||D(\mathbf{y} \circ \mathbf{x}^{-1})(x^{\mathbf{x}}) \cdot f^{\mathbf{x}}(x^{\mathbf{x}},t)|| \le A_{H,\mathbf{y}}C_{H,\mathbf{x}}^{\#}k(t),$$

where $A_{H,\mathbf{y}}$ is defined by letting

$$A_{H,\mathbf{y}} = \sup\{\|D(\mathbf{y} \circ \mathbf{x}^{-1})(z)\| : z \in \mathbf{x}(Z_{H,\mathbf{y}})\},\$$

and the set $Z_{H,\mathbf{y}}$ is given by

$$Z_{H,\mathbf{y}} = \bigcup \left\{ \mathbf{y}^{-1}(\bar{\mathbb{B}}^m(\xi(t)^{\mathbf{y}}, \tilde{\delta}_{H,\mathbf{y}}^{\#})) : t \in H \right\},\$$

so that $Z_{H,\mathbf{y}}$ is a compact subset of dom $\mathbf{x} \cap \operatorname{dom} \mathbf{y}$.

Also, if $t \in H$ and $q_1, q_2 \in \overline{\mathbb{B}}^m(\xi(t)^{\mathbf{y}}, \tilde{\delta}_{H,\mathbf{y}}^{\#})$, it follows that $q_1, q_2 \in \operatorname{im} \mathbf{y}$, so we may pick $x_1, x_2 \in \operatorname{dom} \mathbf{y}$ for which $q_1 = x_1^{\mathbf{y}}$ and $q_2 = x_2^{\mathbf{y}}$. Then, if we let

$$\begin{aligned} \Theta(z_1, z_2) &= \|z_2 - z_1\|^{-1} \|D(\mathbf{y} \circ \mathbf{x}^{-1})(z_2) - D(\mathbf{y} \circ \mathbf{x}^{-1})(z_1)\|, \\ B_{H,\mathbf{y}} &= \sup\{\Theta(z_1, z_2) : z_1, z_2 \in \mathbf{x}(Z_{H,\mathbf{y}}), z_1 \neq z_2\}, \\ \Psi(u_1, u_2) &= \|u_1 - u_2\|^{-1} \|(\mathbf{x} \circ \mathbf{y}^{-1})(u_2) - (\mathbf{x} \circ \mathbf{y}^{-1})(u_1)\|, \\ E_{H,\mathbf{y}} &= \sup\left\{\Psi(u_1, u_2) : u_1, u_2 \in \mathbf{y}(Z_{H,\mathbf{y}}), u_1 \neq u_2\right\}, \\ \tilde{C}_{H,\mathbf{y}}^{\#} &= (A_{H,\mathbf{y}} + B_{H,\mathbf{y}}) E_{H,\mathbf{y}} C_{H,\mathbf{x}}^{\#}, \end{aligned}$$

we find

$$\begin{split} \|f^{\mathbf{y}}(q_{2},t) - f^{\mathbf{y}}(q_{1},t)\| \\ &= \|D(\mathbf{y} \circ \mathbf{x}^{-1})(x_{2}^{\mathbf{x}}) \cdot f^{\mathbf{x}}(x_{2}^{\mathbf{x}},t) - D(\mathbf{y} \circ \mathbf{x}^{-1})(x_{1}^{\mathbf{x}}) \cdot f^{\mathbf{x}}(x_{1}^{\mathbf{x}},t) \\ &\leq \|D(\mathbf{y} \circ \mathbf{x}^{-1})(x_{2}^{\mathbf{x}}) \cdot f^{\mathbf{x}}(x_{2}^{\mathbf{x}},t) - D(\mathbf{y} \circ \mathbf{x}^{-1})(x_{2}^{\mathbf{x}}) \cdot f^{\mathbf{x}}(x_{1}^{\mathbf{x}},t)\| \\ &+ \|D(\mathbf{y} \circ \mathbf{x}^{-1})(x_{2}^{\mathbf{x}}) \cdot f^{\mathbf{x}}(x_{1}^{\mathbf{x}},t) - D(\mathbf{y} \circ \mathbf{x}^{-1})(x_{1}^{\mathbf{x}}) \cdot f^{\mathbf{x}}(x_{1}^{\mathbf{x}},t)\| \\ &\leq \|D(\mathbf{y} \circ \mathbf{x}^{-1})(x_{2}^{\mathbf{x}}) \cdot \left(f^{\mathbf{x}}(x_{2}^{\mathbf{x}},t) - f^{\mathbf{x}}(x_{1}^{\mathbf{x}},t)\right)\| \\ &+ \|\left(D(\mathbf{y} \circ \mathbf{x}^{-1})(x_{2}^{\mathbf{x}}) - D(\mathbf{y} \circ \mathbf{x}^{-1})(x_{1}^{\mathbf{x}})\right) \cdot f^{\mathbf{x}}(x_{1}^{\mathbf{x}},t)\| \end{split}$$

$$\leq A_{H,\mathbf{y}} \tilde{C}_{H,\mathbf{x}}^{\#} k(t) \| x_{2}^{\mathbf{x}} - x_{1}^{\mathbf{x}} \| + B_{H,\mathbf{y}} \tilde{C}_{H,\mathbf{x}}^{\#} k(t) \| x_{2}^{\mathbf{x}} - x_{1}^{\mathbf{x}} \|$$

= $(A_{H,\mathbf{y}} + B_{H,\mathbf{y}}) \tilde{C}_{H,\mathbf{x}}^{\#} k(t) \| x_{2}^{\mathbf{x}} - x_{1}^{\mathbf{x}} \|$
= $(A_{H,\mathbf{y}} + B_{H,\mathbf{y}}) \tilde{C}_{H,\mathbf{x}}^{\#} k(t) \| (\mathbf{x} \circ \mathbf{y}^{-1})(q_{2}) - (\mathbf{x} \circ \mathbf{y}^{-1})(q_{1}) \|$
 $\leq (A_{H,\mathbf{y}} + B_{H,\mathbf{y}}) E_{H,\mathbf{y}} C_{H,\mathbf{x}}^{\#} k(t) \| q_{2} - q_{1} \|$
= $\tilde{C}_{H,\mathbf{x}}^{\#} k(t) \| q_{2} - q_{1} \|$.

It follows that $({\tilde{\delta}_{H,\mathbf{x}}^{\#}}_{(H,\mathbf{x})\in\tilde{\mathcal{K}}^{\#}}, {\tilde{C}_{H,\mathbf{x}}^{\#}}_{(H,\mathbf{x})\in\tilde{\mathcal{K}}^{\#}})$ is a system of IL constants for $\xi, f, k, \tilde{\mathcal{K}}^{\#}$. Hence f is IL-K near ξ for $\tilde{\mathcal{K}}^{\#}$ and, as explained before, this completes the proof of (IL.*).

It follows from (IL.*) that we can simply talk about a ppd tvvf f being "IL-k near ξ ", and that the validity of this condition can be verified relative to any particular chart covering, in which case the condition will be valid for all chart coverings. We say that a ppd tvvf f is *integrably Lipschitz near* ξ (abbreviated "IL near ξ "), if it is IL-k for some nonnegative integrable function k defined on dom ξ .

Lebesgue times. We recall that if $N \in \mathbb{Z}_+$, $a, b \in \mathbb{R}$, a < b, and $\varphi : [a, b] \mapsto \mathbb{R}^N$ is an integrable function, a **Lebesgue point** of φ is a point $\tau \in]a, b[$ that has the property that $\lim_{h \downarrow 0} \frac{1}{h} \int_{\tau-h}^{\tau+h} \|\varphi(t) - \varphi(\tau)\| dt = 0.$

This concept can be generalized trivially to ppd IL vector fields near an arc. Assuming that (A4) holds, dom $\xi = [a, b]$, and the ppd tvvf f is IL near ξ , a $\tau \in [a, b]$ is said to be a **Lebesgue time of** f **along** ξ if (a) $a < \tau < b$, and, (b) τ is a Lebesgue point of the function $t \mapsto f^{\mathbf{x}}(\xi(t)^{\mathbf{x}}, t) \in \mathbb{R}^m$ for some chart \mathbf{x} of M such that $\xi(\tau) \in \text{dom } \mathbf{x}$. It is easy to verify that if the conclusion of (b) holds for some chart \mathbf{x} such that $\xi(\tau) \in \text{dom } \mathbf{x}$, then it holds for every such chart.

Covariant differentiations along an absolutely continuous arc. In this subsection, we assume that

(A5) $m, \mu \in \mathbb{Z}_+$, M is a manifold of class C^{μ} , $\mu \geq 2$, $m = \dim M$, $\xi \in \mathcal{W}^{1,1}(M)$, $a, b \in \mathbb{R}$, a < b, $\dim \xi = [a, b]$, and $\Xi : [a, b] \hookrightarrow TM$ is the map, defined almost everywhere, given by $\Xi(t) = (\xi(t), \dot{\xi}(t))$ for $a.e. \ t \in [a, b]$.

The facts that $\mu \geq 2$ and $\xi \in \mathcal{W}^{1,1}(M)$ imply that the concepts of "absolutely continuous vector field" and "absolutely continuous covector field" along ξ are well defined. We write $\Gamma_{W^{1,1}}(\xi^*TM)$ (resp. $\Gamma_{W^{1,1}}(\xi^*T^*M)$), to denote the space of all absolutely continuous vector (resp. covector) fields along ξ .

Naturally, if $[a, b] \ni t \mapsto v(t) \in T_{\xi(t)}M$ is absolutely continuous, the "time derivative" of v should be a vector field $\nabla v \in \Gamma_{L^1}(\xi^*TM)$. To make sense of

this in an intrinsic fashion, we define the notion of "covariant differentiation along ξ ."

Definition 4.2 A covariant differentiation along ξ is an \mathbb{R} -linear map $\nabla : \Gamma_{W^{1,1}}(\xi^*TM) \mapsto \Gamma_{L^1}(\xi^*TM)$ such that, whenever $v \in \Gamma_{W^{1,1}}(\xi^*TM)$ and $r \in W^{1,1}([a, b], \mathbb{R})$, it follows that $\nabla(rv) = \dot{r}v + r\nabla v$.

We will use $Cov(\xi)$ to denote the set of all covariant differentiations along the arc ξ . It is clear that any linear combination $\sum_{i=1}^{n} s_i \nabla_i$ of members of $Cov(\xi)$ with coefficients $s_i \in \mathbb{R}$ such that $\sum_{i=1}^{n} s_i = 1$ is again in $Cov(\xi)$, so $Cov(\xi)$ is an affine space over \mathbb{R} .

We now show that $Cov(\xi)$ is canonically identified with a certain space of sections of the pullback $\xi^*(J^1\Gamma(TM))$. Recall that $J^1\Gamma(TM)$ is a vector bundle over M of class $C^{\mu-2}$ and fiber dimension $m+m^2$. It follows from this that the concepts of "measurable" and "integrable" sections of $\xi^*(J^1\Gamma(TM))$ are well defined. (For example, a section $[a, b] \ni t \mapsto \sigma(t) \in J^1_{\xi(t)}\Gamma(TM)$ is *integrable* if its coordinate representation $[\alpha, \beta] \ni t \mapsto \sigma(t)^{\mathbf{x}} \in \mathbb{R}^m \times \mathbb{R}^{m \times m}$ is an integrable function of t whenever \mathbf{x} is a chart and $[\alpha, \beta]$ is a subinterval of [a, b]such that $\xi([\alpha, \beta]) \subseteq \text{dom } \mathbf{x}$.) Also, if $[a, b] \ni t \mapsto S(t) \subseteq J^1_{\xi(t)}\Gamma(TM)$ is a setvalued map, it makes sense to talk about S being "measurable" or "integrably bounded." (For example, S is *measurable* if the set $\{t \in [a, b]: \Omega \cap S(t) \neq \emptyset\}$ is measurable for every open subset Ω of $J^1\Gamma(TM)$. Also, S is *integrably bounded* if, whenever \mathbf{x} is a chart and $[\alpha, \beta]$ is a subinterval of [a, b] such that $\xi([\alpha, \beta]) \subseteq \text{dom } \mathbf{x}$, there exists an integrable function $k : [\alpha, \beta] \mapsto [0, +\infty]$ such that $\|\sigma^{\mathbf{x}}\| \leq k(t)$ whenever $t \in [\alpha, \beta]$ and $\sigma \in S(t)$.)

We use $\Gamma(\xi^*(J^1\Gamma(TM)))$, $\Gamma_{meas}(\xi^*(J^1\Gamma(TM)))$, $\Gamma_{L^1}(\xi^*(J^1\Gamma(TM)))$, to denote, respectively, the set of all sections $[a,b] \ni t \mapsto \sigma(t) \in J^1_{\xi(t)}\Gamma(TM)$ of the pullback bundle $\xi^*(J^1\Gamma(TM))$, the set of all $\sigma \in \Gamma(\xi^*(J^1\Gamma(TM)))$ that are measurable, and the set of all $\sigma \in \Gamma(\xi^*(J^1\Gamma(TM)))$ that are integrable.

We will be particularly interested in those sections σ that are actually lifts of the a.e. defined map $\Xi : [a, b] \hookrightarrow TM$. We will use $\Gamma(\xi^*(J^1\Gamma(TM)); \Xi)$, $\Gamma_{meas}(\xi^*(J^1\Gamma(TM)); \Xi)$, $\Gamma_{L^1}(\xi^*(J^1\Gamma(TM)); \Xi)$, to denote the corresponding spaces of sections. (Naturally, these are really quotient spaces, in which two sections that coincide almost everywhere are regarded as equal.) So, for example, a section $[a, b] \ni t \mapsto \sigma(t) \in J^1_{\xi(t)}\Gamma(TM)$ belongs $\Gamma(\xi^*(J^1\Gamma(TM)); \Xi)$ if and only if $\pi_{J^1\Gamma(TM),TM}(\sigma(t)) = \dot{\xi}(t)$ for a.e. $t \in [a, b]$.

If $\sigma \in \Gamma_{L^1}(\xi^*(J^1\Gamma(TM)); \Xi)$, then σ gives rise to a covariant differentiation $\nabla_{\sigma} \in Cov(\xi)$ as follows. We pick, for each $t \in [a, b]$, a vector field $f_t \in \Gamma_{C^1}(TM)$ such that $j^1 f_t(\xi(t)) = \sigma(t)$. (This implies, in particular, that $f_t(\xi(t)) = \dot{\xi}(t)$.) Furthermore, we require the ppd tvvf $M \times [a, b] \ni (x, t) \hookrightarrow f_t(x) \in T_x M$ to be integrably Lipschitz near ξ . (It is easy to see that this can be done, for example by taking local coordinates.) Clearly, every $v \in \Gamma_{W^{1,1}}(\xi^*(TM))$ can be written as a finite sum $v(t) = \sum_{k=1}^N r_k(t)X_k(\xi(t))$, where the X_k are vector fields of class C^1 on M, and the r_k are integrable functions. We then define

$$\nabla_{\sigma} v(t) = \sum_{k=1}^{N} \dot{r}_k(t) X_k(\xi(t)) + \sum_{k=1}^{N} r_k(t) [f_t, X_k](\xi(t)) .$$
(11)

To show that this is well defined, we must prove that if $\sum_{k=1}^{N} r_k(t) X_k(\xi(t)) \equiv 0$ then $\sum_{k=1}^{N} \dot{r}_k(t) X_k(\xi(t)) + \sum_{k=1}^{N} r_k(t) [f_t, X_k](\xi(t)) \equiv 0$. This can be done, for example, by using coordinates. If **x** is a chart and $\alpha, \beta \in [a, b]$ are such that $\alpha < \beta$ and $\xi([\alpha, \beta]) \subseteq \text{dom } \mathbf{x}$, then, for $t \in [\alpha, \beta]$,

$$\begin{split} &\left(\sum_{k=1}^{N} \dot{r}_{k}(t)X_{k}(\xi(t)) + \sum_{k=1}^{N} r_{k}(t)[f_{t}, X_{k}](\xi(t))\right)^{\mathbf{x}} \\ &= \sum_{k=1}^{N} \dot{r}_{k}(t)X_{k}^{\mathbf{x}}(\xi(t)^{\mathbf{x}}) + \sum_{k=1}^{N} r_{k}(t)[f_{t}, X_{k}]^{\mathbf{x}}(\xi(t)^{\mathbf{x}}) \\ &= \sum_{k=1}^{N} \dot{r}_{k}(t)X_{k}^{\mathbf{x}}(\xi(t)^{\mathbf{x}}) + \sum_{k=1}^{N} r_{k}(t)\left(DX_{k}^{\mathbf{x}} \cdot f_{t}^{\mathbf{x}} - Df_{t}^{\mathbf{x}} \cdot X_{k}^{\mathbf{x}}\right)(\xi(t)^{\mathbf{x}}) \\ &= \sum_{k=1}^{N} \left(\dot{r}_{k}(t)X_{k}^{\mathbf{x}}(\xi(t)^{\mathbf{x}}) + r_{k}(t)DX_{k}^{\mathbf{x}}(\xi(t)^{\mathbf{x}}) \cdot f_{t}^{\mathbf{x}}(\xi(t)^{\mathbf{x}})\right) \\ &- \sum_{k=1}^{N} r_{k}(t)\left(Df_{t}^{\mathbf{x}} \cdot X_{k}^{\mathbf{x}}\right)(\xi(t)^{\mathbf{x}}) \\ &= \sum_{k=1}^{N} \left(\dot{r}_{k}(t)X_{k}^{\mathbf{x}}(\xi(t)^{\mathbf{x}}) + r_{k}(t)DX_{k}^{\mathbf{x}}(\xi(t)^{\mathbf{x}}) \cdot \dot{\xi}(t)^{\mathbf{x}}\right) \\ &- Df_{t}^{\mathbf{x}}(\xi(t)^{\mathbf{x}}) \cdot \sum_{k=1}^{N} r_{k}(t)X_{k}^{\mathbf{x}}(\xi(t)^{\mathbf{x}}) \\ &= \frac{d}{dt} \left(\sum_{k=1}^{N} r_{k}(t)X_{k}(\xi(t))\right)^{\mathbf{x}} - Df_{t}^{\mathbf{x}}(\xi(t)^{\mathbf{x}}) \cdot \left(\sum_{k=1}^{N} r_{k}(t)X_{k}(\xi(t))\right)^{\mathbf{x}}, \end{split}$$

which vanishes identically if $\sum_{k=1}^{N} r_k(t) X_k(\xi(t)) \equiv 0$. This proves that ∇_{σ} is well defined. The identity then $\nabla_{\sigma}(rv) = \dot{r}v + r\nabla_{\sigma}v$ follows trivially from the definition of ∇_{σ} . So $\nabla_{\sigma} \in Cov(\xi)$.

It is clear that if $\nabla = \nabla_{\sigma}$, then ∇ satisfies

$$X \in \Gamma_{C^1}(TM) \implies \nabla(X \circ \xi)(t) = Lie^1(\xi(t))(\sigma(t), j^1X(\xi(t))) \text{ for a.e } t, (12)$$

where $X \circ \xi$ is the map $[a, b] \ni t \mapsto X(\xi(t)) \in T_{\xi(t)}M$, and $Lie^1(x)$ is the map from $J_x^1\Gamma(TM) \times J_x^1\Gamma(TM)$ to T_xM defined in (1). Finally, Condition (12) uniquely determines the covariant differentiation ∇_{σ} , in view of the identity $\nabla(rv) = \dot{r}v + r\nabla v$.

We now write a coordinate expression for ∇_{σ} . Assume that **x** is a chart and $\alpha, \beta \in [a, b]$ are such that $\alpha < \beta$ and $\xi([\alpha, \beta]) \subseteq \text{dom } \mathbf{x}$. Let $v \in \Gamma_{W^{1,1}}(\xi^*TM)$. Then, for $t \in [\alpha, \beta], v(t) = \sum_{i=1}^m v^{\mathbf{x}, i}(t)(\partial_i^{\mathbf{x}} \circ \xi)(t)$, so

$$\nabla_{\sigma} v(t) = \sum_{i=1}^{m} \dot{v}^{\mathbf{x},i}(t) (\partial_{i}^{\mathbf{x}} \circ \xi)(t) + \sum_{i=1}^{m} v^{\mathbf{x},i}(t) \left([f_{t}, \partial_{i}^{\mathbf{x}}] \circ \xi \right)(t) \,.$$

Since $[f_t, \partial_i^{\mathbf{x}}]^{\mathbf{x}}$ is the column vector $-\frac{\partial f_t^{\mathbf{x}}}{\partial x^i}$, we find that

$$[f_t, \partial_i^{\mathbf{x}}] \circ \xi = -\sum_{j=1}^m \frac{\partial f_t^{\mathbf{x}, j}}{\partial x^i} \partial_j^{\mathbf{x}} \circ \xi ,$$

 \mathbf{SO}

$$\sum_{i=1}^{m} v^{\mathbf{x},i}(t) \left([f_t, \partial_i^{\mathbf{x}}] \circ \xi \right) = -\sum_{i=1}^{m} \sum_{j=1}^{m} v^{\mathbf{x},i}(t) \frac{\partial f_t^{\mathbf{x},j}}{\partial x^i} \partial_j^{\mathbf{x}} \circ \xi ,$$

and then

$$\left(\sum_{i=1}^{m} v^{\mathbf{x},i}(t) \left([f_t, \partial_i^{\mathbf{x}}] \circ \xi \right) (\xi(t)) \right)^{\mathbf{x}} = -Df_t^{\mathbf{x}}(\xi(t)^{\mathbf{x}}) \cdot v(t)^{\mathbf{x}} = -\sigma(t)^{\mathbf{x},red} \cdot v(t)^{\mathbf{x}} \cdot v(t)^{\mathbf{x}}$$

So

$$(\nabla_{\sigma} v(t))^{\mathbf{x}} = \dot{v}^{\mathbf{x}}(t) - \sigma(t)^{\mathbf{x}, red} \cdot v^{\mathbf{x}}(t) \,. \tag{13}$$

Formula (13) implies, in particular, that the map $\sigma \mapsto \nabla_{\sigma}$ is injective, because the matrices $\sigma(t)^{\mathbf{x},red}$ can be recovered from ∇_{σ} , and then $\sigma(t)$ must be the 1-jet whose representation is $(\dot{\xi}(t)^{\mathbf{x}}, \sigma(t)^{\mathbf{x},red})$.

In addition, it is easy to see that every $\nabla \in Cov(\xi)$ arises in this way, as ∇_{σ} for some $\sigma \in \Gamma_{L^1}(\xi^*(J^1\Gamma(TM)); \Xi)$. Indeed, given $\nabla \in Cov(\xi)$, we can partition the interval [a, b] into subintervals $I_k = [t_{k-1}, t_k], \ k = 1, \ldots, N$, in such a way that each I_k is contained in the domain dom \mathbf{x}_k of some chart \mathbf{x}_k . Then the identity $\nabla(rv) = \dot{r}v + r\nabla v$ implies $(\nabla(rv))^{\mathbf{x}_k} = \dot{r}v^{\mathbf{x},k} + r(\nabla v)^{\mathbf{x},k}$. On the other hand, if Δ denotes ordinary differentiation with respect to t, then $\Delta((rv)^{\mathbf{x}_k}) = \Delta(rv^{\mathbf{x}_k}) = \dot{r}v^{\mathbf{x},k} + r\Delta v^{\mathbf{x},k}$. It then follows from these two identities that $(\nabla^{\mathbf{x}_k} - \Delta)(rv^{\mathbf{x}_k}) = r(\nabla^{\mathbf{x}_k} - \Delta)v^{\mathbf{x}_k}$, showing that, for $t \in I_k$, the vector $((\nabla^{\mathbf{x}_k} - \Delta)v^{\mathbf{x}_k})(t)$ only depends on the vector $v^{\mathbf{x}_k}(t)$. Therefore the maps $\mathbb{R}^m \ni v^{\mathbf{x}_k}(t) \mapsto ((\nabla^{\mathbf{x}_k} - \nabla)v^{\mathbf{x}_k})(t) \in \mathbb{R}^m$ are well defined. Since these maps are obviously linear, there exists, for each t, a matrix $\mu_k^{\mathbf{x}_k}(t) \in \mathbb{R}^{m \times m}$ such that $((\nabla^{\mathbf{x}_k} - \nabla)v^{\mathbf{x}_k})(t) = \mu_k^{\mathbf{x}_k}(t) \cdot v^{\mathbf{x}_k}(t)$. If we let σ be the member of $\Gamma_{L^1}(\xi^*(J^1\Gamma(TM)); \Xi)$ such that $\sigma^{\mathbf{x}_k}(t) = (\dot{\xi}(t)^{\mathbf{x}_k}, \mu_k^{\mathbf{x}_k}(t))$ for $t \in I_k$, we see that $\nabla = \nabla_{\sigma}$.

So we have proved

Proposition 4.3 For every $\sigma \in \Gamma_{L^1}(\xi^*(J^1\Gamma(TM)); \Xi)$ there exists a unique covariant differentiation ∇ along ξ that satisfies (12). Using ∇_{σ} to denote this covariant differentiation, then ∇_{σ} is given by (11), if $\{f_t\}_{t\in[a,b]}$ is any family of vector fields $f_t \in \Gamma_{C^1}(TM)$ such that $j^1f_t(\xi(t)) = \sigma(t)$ and the ppd map $M \times [a,b] \ni (x,t) \hookrightarrow f_t(x) \in T_x M$ is integrably Lipschitz near ξ . Furthermore, the map $\Gamma_{L^1}(\xi^*(J^1\Gamma(TM)); \Xi) \ni \sigma \mapsto \nabla_{\sigma} \in Cov(\xi)$ is a bijection. \Box

If $\mathbf{t} = (t_0, t_1, \dots, t_N)$ is a partition of [a, b] (i.e., \mathbf{t} is an N + 1-tuple such that $a = t_0 < t_1 < \cdots < t_N = b$, and $\mathbf{x}_1, \ldots, \mathbf{x}_N$ are charts of M such that $\xi([t_{j-1}, t_j]) \subseteq \operatorname{dom} \mathbf{x}_j$ for each j, then we can identify the space $Cov(\xi)$ with $L^1([a,b],\mathbb{R}^{m\times m}]$, by assigning to each $\sigma\in\Gamma_{L^1}(\xi^*(J^1\Gamma(TM));\Xi)$ the matrixvalued function $\mu : [a, b] \mapsto \mathbb{R}^{m \times m}$ such that $\mu(t) = \sigma(t)^{\mathbf{x}_k, red}$ for $t \in I_k$, where $I_k = [t_{k-1}, t_k]$. The resulting bijection is an affine map, which depends on t and the \mathbf{x}_k . A simple calculation shows, however, that, if B_1 , B_2 are the bijections that correspond to two different choices of t and the \mathbf{x}_k , and d_1 , d_2 are the distance functions on $Cov(\xi)$ obtained by transporting to $Cov(\xi)$ by means of B_1, B_2 the distance functions arising from the L^1 norm on $L^1([a, b], \mathbb{R}^{m \times m})$, then d_1 and d_2 are equivalent, in the sense that there are positive constants C_1, C_2 such that the inequalities $C_1 d_1(\nabla, \nabla') \leq d_2(\nabla, \nabla') \leq C_2 d_1(\nabla, \nabla')$ hold for all $\nabla, \nabla' \in Cov(\xi)$. Therefore $Cov(\xi)$ is, canonically, a complete normable real affine topological space. This implies, in particular, that the class $Aff_{C^0}(Cov(\xi),\mathbb{R})$ of all continuous affine real-valued functionals on $Cov(\xi)$ is intrinsically defined. So $Cov(\xi)$ has an intrinsically defined weak topology \mathcal{T}_{weak} , characterized as the weakest topology on $Cov(\xi)$ that makes all the maps $\varphi \in Aff_{C^0}(Cov(\xi), \mathbb{R})$ continuous.

Given a $\nabla \in Cov(\xi)$, a vector field $v \in \Gamma_{W^{1,1}}(\xi^*TM)$ is **parallel transported along** ∇ if $\nabla v \equiv 0$. It follows from (13) that the parallel translation equation $\nabla v \equiv 0$, written in coordinates, if $\nabla = \nabla_{\sigma}$ for a $\sigma \in \Gamma_{L^1}(\xi^*(J^1\Gamma(TM)); \Xi)$, is a linear time-varying system with an integrable coefficient matrix. This implies existence and uniqueness of the solutions. Therefore, given any $t \in [a, b]$ and any vector $v_0 \in T_{\xi(t)}M$, there exists a unique $v \in \Gamma_{W^{1,1}}(\xi^*TM)$ which is parallel translate along ∇ and such that $v(t) = v_0$. We write $v(s) = P_{s,t}^{\nabla}(v_0)$. It is then clear that the **parallel translation maps**, or **propagators**, $P_{s,t}^{\nabla}: T_{\xi(t)}M \mapsto T_{\xi(s)}M$ are invertible linear maps, and satisfy the following **flow identities**: $P_{s,t}^{\nabla} \circ P_{t,r}^{\nabla} = P_{s,r}^{\nabla}$, $(P_{s,t}^{\nabla})^{-1} = P_{t,s}^{\nabla}$, and $P_{t,t}^{\nabla} = \mathbb{I}_{T_{\xi(t)}}M$.

The following observation follows from Gronwall's inequality and the Ascoli-Arzelà theorem.

Proposition 4.4 Let $[a,b] \ni t \mapsto S(t) \subseteq J^1_{\Xi(t)}(\Gamma(TM))$ be an integrably bounded set-valued map. Let $\Gamma(S)$ be the set of all measurable selections $[a,b] \ni t \mapsto \sigma(t) \in S(t)$ of S, and let $\Sigma(\Gamma(S)) = \{\nabla_{\sigma} : \sigma \in \Gamma(S)\}$, so that $\Sigma(\Gamma(S)) \subseteq Cov(\xi)$. Let $\Sigma(\Gamma(S))_{weak}$ be $\Sigma(\Gamma(S))$ endowed with the topology induced by the weak topology of $Cov(\xi)$. Then, for every $s, t \in [a, b]$, the map $\Sigma(\Gamma(S))_{weak} \ni \nabla \mapsto P_{s,t}^{\nabla} \in Lin(T_{\xi(t)}M, T_{\xi(s)}M)$ is continuous.

Proof. We must show that if $\{\nabla^{\theta}\}_{\theta\in\Theta}$ is a net of members of $\Sigma(\Gamma(S))$ that converges weakly to $\nabla^{\#} \in \Sigma(\Gamma(S))$, then the net $\{P_{s,t}^{\nabla^{\theta}}(v)\}_{\theta\in\Theta}$ converges to $P_{s,t}^{\nabla^{\#}}(v)$ in $T_{\xi(s)}M$ for every $v \in T_{\xi(t)}M$. In view of the flow identities, it suffices to prove this if s and t both belong to an interval $[\alpha, \beta]$ such that $\xi([\alpha, \beta]) \subseteq \operatorname{dom} \mathbf{x}$ for some chart \mathbf{x} .

For $\theta \in \Theta \cup \{\#\}$, let $\sigma_{\theta} \in \Gamma(S)$ be such that $\nabla_{\sigma_{\theta}} = \nabla^{\theta}$, and let $V_{\theta} \in \Gamma_{W^{1,1}}(\xi^*(TM))$ be such that $\nabla_{\sigma_{\theta}}V_{\theta} = 0$ and $V_{\theta}(t) = v$. Then all we have to do is show that the net $\{V_{\theta}(s)\}_{\theta \in \Theta}$ converges to $V_{\#}(s)$.

Let $A_{\theta}(t) = \sigma_{\theta}^{\mathbf{x},red}(t), \ b_{\theta}(t) = V_{\theta}(t)^{\mathbf{x}}, \ \hat{v} = v^{\mathbf{x}}$. Then we have to show that the net $\{b_{\theta}(s)\}_{\theta\in\Theta}$ converges to $b_{\#}(s)$. Since S is integrably bounded, we may pick an integrable function $k : [\alpha, \beta] \mapsto [0, +\infty]$ such that $||A_{\theta}(u)|| \le k(u)$ for every $u \in [\alpha, \beta]$ and every $\theta \in \Theta \cup \{\#\}$,

For $\tau \in [\alpha, \beta]$, $\theta \in \Theta$, we have $b_{\theta}(\tau) = \hat{v} + \int_{t}^{\tau} A_{\theta}(u) \cdot b_{\theta}(u) du$. Hence Gronwall's inequality implies that $||b_{\theta}(\tau)|| \leq C$, where $C = e^{\int_{a}^{b} k(u) du} ||\hat{v}||$. Then the equality $\dot{b}_{\theta}(\tau) = A_{\theta}(\tau) \cdot b_{\theta}(\tau)$ implies that $||\dot{b}_{\theta}(\tau)|| \leq Ck(t)$. So the vector functions b_{θ} are uniformly bounded and equicontinuous. By the Ascoli-Arzelà theorem, applied to an arbitrary subnet $\{b_{\theta(\rho)}\}_{\rho \in R}$ of the net $\{b_{\theta}\}_{\theta \in \Theta}$, we conclude that this subnet has a subnet $\{b_{\theta(\rho(\zeta))}\}_{\zeta \in Z}$ that converges uniformly to a limit \tilde{b} . By well know arguments, if we show that $\hat{b} = b_{\#}$, it will follow that every subnet of our original net has a subnet that converges to $b_{\#}$, so the net $\{b_{\theta}\}_{\theta \in \Theta}$ itself converges unformly to $b_{\#}$, implying our desired result.

Write $K_{\zeta} = A_{\theta(\rho(\zeta))}, h_{\zeta} = b_{\theta(\rho(\zeta))}$. Then $h_{\zeta}(\tau) = \hat{v} + \int_{t}^{\tau} K_{\zeta}(u) \cdot h_{\zeta}(u) du$. Since $h_{\zeta} \to \hat{b}$ uniformly, we can conclude that

$$\hat{b}(\tau) = \hat{v} + \int_{t}^{\tau} A_{\#}(u)\hat{b}(u) \, du + \lim_{\zeta} \int_{t}^{\tau} (K_{\zeta}(u)h_{\zeta}(u) - A_{\#}(u)\hat{b}(u)) \, du \,, \quad (14)$$

and it suffices to show that the limit in the above expression vanishes. Now, $K_{\zeta}(u)h_{\zeta}(u) - A_{\#}(u)\hat{b}(u) = K_{\zeta}(u)(h_{\zeta}(u) - \hat{b}(u)) + (K_{\zeta}(u) - A_{\#}(u))\hat{b}(u)$. The integral of $(K_{\zeta}(u) - A_{\#}(u))\hat{b}(u)$ converges to zero because the functions K_{ζ} converge to $A_{\#}$ weakly, and \hat{b} is a fixed function. In addition, the integral of $K_{\zeta}(u)(h_{\zeta}(u) - \hat{b}(u))$ converges to zero because $h_{\zeta} \to \hat{b}$ uniformly and the K_{ζ} are uniformly dominated by a fixed integrable function. Hence the limit in (14) does vanish, as desired, and this completes our proof. \Box

Finally, we point out that the covariant differentiation operators ∇_{σ} extend,

in the usual way, to fields of contravariant and covariant tensors of any type. Here we will only need to consider fields of covectors. The operator ∇_{σ} acts on $\Gamma_{W^{1,1}}(\xi^*T^*M)$ in such a way that

$$\frac{d}{dt}\langle w(t), v(t) \rangle = \langle \nabla_{\sigma} w(t), v(t) \rangle + \langle w(t), \nabla_{\sigma} v(t) \rangle$$

whenever $w \in \Gamma_{W^{1,1}}(\xi^*T * M)$ and $v \in \Gamma_{W^{1,1}}(\xi^*TM)$. This immediately yields the coordinate expression for the action of ∇_{σ} on covector fields, which turns out to be given by

$$(\nabla_{\sigma} w(t))^{\mathbf{x}} = \dot{w}^{\mathbf{x}}(t) + w^{\mathbf{x}}(t) \cdot \sigma(t)^{\mathbf{x}, red} \,. \tag{15}$$

In particular, the parallel translation equation for covector fields is the familiar "adjoint equation" $\dot{w}^{\mathbf{x}}(t) = -w^{\mathbf{x}}(t) \cdot \sigma(t)^{\mathbf{x},red}$, i.e., $\dot{w}^{\mathbf{x}}(t) = -w^{\mathbf{x}}(t) \cdot Df_t^{\mathbf{x}}(\xi(t))$.

The variational inclusion and Warga's differentiation theorem. In this subsection, we assume that

(A6) $m, \mu \in \mathbb{Z}_+, M$ is a manifold of class $C^{\mu}, \mu \geq 2, m = \dim M, \xi$ belongs to $\mathcal{W}^{1,1}(M)$, dom $\xi = [a, b], f$ is a ppd time-varying vector field on M, f is integrably Lipschitz near ξ , and $\dot{\xi}(t) = f(\xi(t), t)$ for almost all t.

Assume for a moment that the vector fields f_t given by $f_t(x) = f(x,t)$ were of class C^1 . Then their 1-jets $j^1 f_t(\xi(t))$ would obviously give rise to a section $[a,b] \ni t \mapsto \sigma(t) = j^1 f_t(\xi(t)) \in J^1_{\xi(t)} \Gamma(TM)$, which would in turn give rise to a parallel translation equation $\nabla_{\sigma} v \equiv 0$. In that case, it is well known that the map $P_{b,a}^{\nabla_{\sigma}} : T_{\xi(a)}M \mapsto T_{\xi(b)}M$ is the differential of the time-*a*-to-time-*b* flow map of f at $\xi(a)$.

For the more general case when f is integrably Lipschitz but not necessarily of class C^1 , the corresponding theorem on differentiation of the flow maps was proved by J. Warga for vector fields in \mathbb{R}^m , and yields a set-valued differential, in the sense of Warga's own theory of derivate containers. We now state the intrinsic version of Warga's result.

Since f_t is a locally Lipschitz vector field on some neighborhood of $\xi(t)$ for almost every t, it follows that f_t has a well defined Clarke generalized Jacobian $\partial f_t(\xi(t))$, which is a nonempty compact convex subset of the m^2 -dimensional affine space $J^1_{\xi(t),\dot{\xi}(t))} \Gamma(TM)$. We use $\partial f \circ \xi$ to denote the set-valued map $[a,b] \ni t \mapsto \partial f_t(\xi(t)) \subseteq J^1_{\xi(t),f_t(\xi(t))} \Gamma(TM)$. It is then easy to see that $\partial f \circ \xi$ is measurable and integrably bounded.

The expression $\Gamma(\partial f \circ \xi)$ will denote the set of all measurable selections $[a, b] \ni t \mapsto \sigma(t) \in (\partial f \circ \xi)(t)$ of $\partial f \circ \xi$, and we use $\nabla_{\Gamma(\partial f \circ \xi)}$ to denote the corresponding set of covariant differentiations. Since $\partial f \circ \xi$ is an integrably

bounded measurable set-valued map with compact, convex, nonempty values, the set $\nabla_{\Gamma(\partial f \circ \xi)}$ is weakly compact.

It then follows from Proposition 4.4 that, if we define

$$\mathcal{P}_{t,s}^{\mathbf{\nabla}_{\Gamma(\partial f\circ\xi)}} \stackrel{\text{def}}{=} \{P_{s,t}^{\nabla} : \nabla \in \mathbf{\nabla}_{\Gamma(\partial f\circ\xi)}\} = \{P_{s,t}^{\nabla_{\sigma}} : \sigma \in \Gamma(\partial f\circ\xi)\},\$$

then $\mathcal{P}_{t,s}^{\mathbf{\nabla}_{\Gamma(\partial f \circ \xi)}}$ is a compact subset of $Lin(T_{\xi(t)}M, T_{\xi(s)}M)$, all whose members are invertible maps, whenever $s, t \in [a, b]$. Furthermore, it is clear that the sets $\mathcal{P}_{t,s}^{\mathbf{\nabla}_{\Gamma(\partial f \circ \xi)}}$ satisfy the **flow identities**

$$\mathcal{P}_{t,s}^{\nabla_{\Gamma(\partial f \circ \xi)}} \circ \mathcal{P}_{s,r}^{\nabla_{\Gamma(\partial f \circ \xi)}} = \mathcal{P}_{t,r}^{\nabla_{\Gamma(\partial f \circ \xi)}} \quad \text{if } a \le r \le s \le t \le b ,$$

$$(16)$$

$$(\mathcal{P}_{t,s}^{\mathbf{\nabla}_{\Gamma(\partial f \circ \xi)}})^{-1} = \mathcal{P}_{s,t}^{\mathbf{\nabla}_{\Gamma(\partial f \circ \xi)}} \quad \text{if } s, t \in [a, b],$$
(17)

$$\mathcal{P}_{t,t}^{\mathbf{V}_{\Gamma(\partial f \circ \xi)}} = \{ \mathbb{I}_{T_{\xi(t)}M} \} \quad \text{if } t \in [a, b] , \qquad (18)$$

where, if X_1, X_2, X_3 are linear spaces, and $\Lambda_i \subseteq Lin(X_i, X_{i+1})$ for i = 1, 2, then $\Lambda_2 \circ \Lambda_1$ is defined to be the subset of $Lin(X_1, X_3)$ given by the formula $\Lambda_2 \circ \Lambda_1 = \{L_2 \circ L_1 : L_1 \in \Lambda_1, L_2 \in \Lambda_2\}$, and $\Lambda_1^{-1} \stackrel{\text{def}}{=} \{L^{-1} : L \in \Lambda_1\}$ if Λ_1 is a set of invertible maps.

Remark 4.5 It is easy to see that the first flow identity is also true if $r \ge s \ge t$, but can fail to be true if $s > \max(t, r)$ or $s < \min(t, r)$. For example, if s > t = r, then $\mathcal{P}_{t,r}^{\nabla_{\Gamma(\partial f \circ \xi)}} = \{\mathbb{I}_{T_{\xi(t)}M}\}$, while on the other hand

$$\mathcal{P}_{t,s}^{\mathbf{\nabla}_{\Gamma(\partial f\circ\xi)}} \circ \mathcal{P}_{s,r}^{\mathbf{\nabla}_{\Gamma(\partial f\circ\xi)}} = (\mathcal{P}_{s,t}^{\mathbf{\nabla}_{\Gamma(\partial f\circ\xi)}})^{-1} \circ \mathcal{P}_{s,t}^{\mathbf{\nabla}_{\Gamma(\partial f\circ\xi)}},$$

which is never equal to the set $\{\mathbb{I}_{T_{\xi(t)}M}\}$, as long as $\mathcal{P}_{s,t}^{\nabla_{\Gamma(\partial f \circ \xi)}}$ has more than one member.

We are now, finally, in a position to state and prove Warga's differentiation theorem (cf. [30–33]). (Recall that the *time t to time s flow map* $\Phi_{s,t}^f$ of f was defined on Page 28.)

Theorem 4.6 Under Assumption (A6), there exists a neighborhood U of $\xi(a)$ such that the map $\Phi_{b,a}^f$ is defined, single-valued, and Lipschitz on U. Then the compact set $\mathcal{P}_{b,a}^{\nabla_{\Gamma(\partial f \circ \xi)}} \subseteq Lin(T_{\xi(a)}M, T_{\xi(b)}M)$ is a Warga derivate container of $\Phi_{b,a}^f$ at $\xi(a)$.

Proof. In view of the chain rule for Warga derivate containers (cf. (WDC.4), Page 9), and the flow identity (16), it suffices to prove our conclusion if the interval [a, b] is such that $\xi([a, b])$ is entirely contained the domain dom **x** of a

chart **x**. But then we may assume that M is an open subset of \mathbb{R}^m and, for some positive number $\overline{\delta}$ and some integrable function $f:[a,b] \mapsto [0,+\infty][$, the tvvf f is defined on the tube $\mathcal{T}_{2\overline{\delta}}$ (where, for each positive δ , we let $\mathcal{T}_{\delta} = \{(x,t) : a \leq t \leq b, \|x - \xi(t)\| \leq \delta\}$), and satisfies, for almost all $t \in [a,b]$, the bounds $\|f(x,t)\| \leq k(t)$ and $\|f(x,t) - f(x',t)\| \leq k(t)\|x - x'\|$, whenever $x, x' \in \mathbb{R}^m$ are such that $(x,t) \in \mathcal{T}_{2\overline{\delta}}$ and $(x',t) \in \mathcal{T}_{2\overline{\delta}}$.

The first assertion, that U exists, is trivial in view of the Lipschitz bound for f and Gronwall's inequality. To prove the second assertion, we regularize f as follows: we fix a nonnegative function $\varphi \in C^{\infty}(\mathbb{R}^m, \mathbb{R})$ such that $\varphi(x) = 0$ for $||x|| \geq 1$ and $\int_{\mathbb{R}^m} \varphi(x) dx = 1$, and then for $0 \leq \rho \leq \overline{\delta}$, $(x, t) \in \mathcal{T}_{\overline{\delta}}$, we define $f_{\rho}(x,t) = \int_{\mathbb{R}^m} \varphi(h) f(x+\rho h,t) dh$. Let $\kappa(t) = \int_a^t k(s) ds$, $K(t) = e^{\kappa(t)}$, $\kappa_* = \kappa(b)$, $K_* = K(b) = e^{\kappa_*}$. Then

(*) If $0 < \rho \leq \overline{\delta}$, then the map $\overline{\mathbb{B}}^m(\xi(t), \overline{\delta}) \ni x \mapsto f_\rho(x, t) \in \mathbb{R}^m$ is of class C^∞ for almost all $t \in [a, b]$ and satisfies $\|f_\rho(x, t)\| \leq k(t), \|Df_\rho(x, t)\| \leq k(t),$ and $\|f_\rho(x, t) - f(x, t)\| \leq \rho k(t)$ whenever $x \in \overline{\mathbb{B}}^m(\xi(t), \overline{\delta})$.

Now fix a positive δ_* such that $K_*(1+\kappa_*)\delta_* \leq \overline{\delta}$. Then the flow maps $\Phi_{t,a}^{f_{\rho}}$, for $0 \leq \rho \leq \delta_*$, are defined on the set $U_* = \overline{\mathbb{B}}^m(\xi(a), \delta_*)$ and satisfy

$$\begin{aligned} & \|\Phi_{t,a}^{f_{\rho}}(x) - \Phi_{t,a}^{f_{\rho}}(x')\| \le K(t) \|x - x'\|, \\ & \|\Phi_{t,a}^{f_{\rho}}(x) - \Phi_{t,a}^{f}(x)\| \le K(t)\kappa(t)\rho. \end{aligned}$$

In particular, $\Phi_{b,a}^{f_{\rho}}(x) \to \Phi_{b,a}^{f}(x)$ as $\rho \downarrow 0$ uniformly for $x \in U_*$. Furthermore, the maps $\Phi_{b,a}^{f_{\rho}}$, for $\rho > 0$, are of class C^{∞} .

It is clear that the set $\Gamma(\partial f \circ \xi)$ consists of all the measurable selections $[a,b] \ni t \mapsto A(t) \in \mathbb{R}^{m \times m}$ of the set-valued map $[a,b] \ni t \mapsto \partial f_t(\xi(t)) \subseteq \mathbb{R}^{m \times m}$, where $f_t(x) = f(x,t)$. For any such A, let M_A denote the fundamental matrix solution of $\dot{M} = A(t)M$, so M_A belongs to $C^0([a,b] \times [a,b], \mathbb{R}^{m \times m})$, and M_A is characterized—using $\mathbb{I}^{m \times m}$ to denote the $m \times m$ identity matrix—by the integral equation

$$M_A(t,s) = \mathbb{I}^{m \times m} + \int_s^t A(r) M_A(r,s) \, dr \, .$$

Then

$$\mathcal{P}_{b,a}^{\nabla_{\Gamma(\partial f \circ \xi)}} = \mathcal{M}, \quad \text{where } \mathcal{M} = \{M_A(b,a) : A \in \Gamma(\partial f \circ \xi)\}.$$

For each δ such that $0 < \delta \leq \delta_*$, and each $t \in [a, b]$, use $\mathcal{D}^{\delta}(t)$ to denote the closed convex hull of the set of all the derivatives $Df_t(x)$ at all points $x \in \overline{\mathbb{B}}^m(\xi(t), \delta) \cap diff(f_t)$ (cf. Page 7). Then $[a, b] \ni t \mapsto \mathcal{D}^{\delta}(t) \subseteq \mathbb{R}^{m \times m}$ is a measurable set-valued map with compact convex nonempty values, such that $||L|| \leq k(t)$ whenever $L \in \mathcal{D}^{\delta}(t)$. Let $\Gamma(\mathcal{D}^{\delta})$ be the set of all measurable selections $[a, b] \ni t \mapsto A(t) \in \mathbb{R}^{m \times m}$ of the map \mathcal{D}^{δ} , and write

$$\mathcal{M}^{\delta} = \{ M_A(b, a) : A \in \Gamma(\mathcal{D}^{\delta}) \}.$$

Then the \mathcal{M}^{δ} are compact subsets of $\mathbb{R}^{m \times m}$, and $\bigcap_{\delta > 0} \mathcal{M}^{\delta} = \mathcal{M}$. (Reason: Let $L \in \bigcap_{\delta > 0} \mathcal{M}^{\delta}$. Let $\{\varepsilon_k\}_{k \in \mathbb{N}}$ be a decreasing sequence of positive numbers that converges to 0 and is such that $\varepsilon_1 \leq \delta_*$. Then for each $k \in \mathbb{N}$ we can find $A_k \in \Gamma(\mathcal{D}^{\varepsilon_k})$ such that $L = M_{A_k}(b, a)$. Since $\Gamma(\mathcal{D}^{\varepsilon_1})_{weak}$ is compact and metrizable, we may pass to a subsequence of the sequence $\{(\varepsilon_k, A_k)\}_{k \in \mathbb{N}}$ and assume that $\{A_i\}_{i\in\mathbb{N}}$ converges weakly to a limit A. Then the fact that $L = M_{A_k}(b, a)$ for all k implies that $L = M_A(b, a)$. We now show that $A \in \Gamma(\partial f \circ \xi)$. For $k \in \mathbb{N}$, let \mathcal{A}_k be the set $\{A_j : j \geq k\}$, and let $\overline{\mathcal{A}_k}$ be the strong closure in $L^1([a,b], \mathbb{R}^{m \times m})$ of the convex hull of \mathcal{A}_k . Then $\overline{\mathcal{A}_k}$ is strongly closed and convex, so it is weakly closed. Therefore $A \in \overline{\mathcal{A}_k}$. Hence we can find $B_k \in \mathcal{A}_k$ such that $||B_k - A||_{L^1} \leq 2^{-k}$. Then the sequence $\{B_k(t)\}_{k \in \mathbb{N}}$ converges to A(t) for all t in a subset E of [a, b] such that $meas([a, b] \setminus E) = 0$. If $t \in E$, $j, k \in \mathbb{N}$, and $j \geq k$, then B_j is a convex combination of members of \mathcal{A}_j , so $B_i(t) \in \mathcal{D}^{\varepsilon_k}(t)$. It follows that $A(t) \in \mathcal{D}^{\varepsilon_k}(t)$. Since this is true for all k, and $\bigcap_{k\in\mathbb{N}}\mathcal{D}^{\varepsilon_k}(t) = \partial f_t(\xi(t)), \text{ we conclude that } A(t) \in \partial f_t(\xi(t)). \text{ So } A \in \Gamma(\partial f \circ \xi) \},$ and then $L \in \mathcal{M}$.)

We now fix δ such that $0 < \delta \leq \delta_*$, and compute the differential $D\Phi_{b,a}^{f_{\rho}}(x)$, for $x \in \mathbb{B}^m(\xi(a), \delta), 0 < \rho \leq \delta_*$. In view of (*), we can use the classical variational equation, and conclude that

$$D\Phi_{b,a}^{f_{\rho}}(x) = M_{A_{\rho,x}}(b,a), \quad \text{where } A_{\rho,x}(t) = Df_{\rho,t}(\Phi_{t,a}^{f_{\rho}}(x)),$$

and $f_{\rho,t}(y) = f_{\rho}(y,t)$. On the other hand,

$$A_{\rho,x}(t) = \int_{\mathbb{R}^m} \varphi(h) Df_t(\Phi_{t,a}^{f_{\rho}}(x) + \rho h, t) \, dh \, .$$

Also, if we fix h such that $||h|| \ge 1$, and let $z = y + \rho h$, $y = \Phi_{t,a}^{f_{\rho}}(x)$, then z satisfies

$$\begin{aligned} \|z - \xi(t)\| &\leq \|y - \xi(t)\| + \rho \\ &\leq \|y - \Phi_{t,a}^f(x)\| + \|\Phi_{t,a}^f(x) - \Phi_{t,a}^f(\xi(a))\| + \rho \\ &\leq K(t)\kappa(t)\rho + K(t)\|x - \xi(t)\| + \rho \\ &\leq K(t)\kappa(t)\rho + K(t)\delta + \rho \\ &\leq K_* \big((\kappa_* + 1)\rho + \delta\big) \\ &\stackrel{\text{def}}{=} \hat{\beta}(\delta, \rho) \,. \end{aligned}$$

It follows that $A_{\rho,x}(t)$ is an average of members of $\mathcal{D}^{\hat{\beta}(\delta,\rho)}(t)$ (because $\varphi \geq 0$, $\varphi(h) = 0$ if $||h|| \geq 1$, and $\int \varphi = 1$). So $A_{\rho,x}(t) \in \mathcal{D}^{\hat{\beta}(\delta,\rho)}(t)$. Hence $A_{\rho,x}$ belongs to $\Gamma(\mathcal{D}^{\hat{\beta}(\delta,\rho)})$, so $D\Phi_{ba}^{f_{\rho}}(x) \in \mathcal{M}^{\hat{\beta}(\delta,\rho)}$.

Given a neighborhood \mathcal{N} of \mathcal{M} in $\mathbb{R}^{m \times m}$, we can find a positive β such that $\mathcal{M}^{\beta} \subseteq \mathcal{N}$, and then find positive δ , $\tilde{\rho}$ such that $\hat{\beta}(\delta, \tilde{\rho}) \leq \beta$. Then $D\Phi_{b,a}^{f_{\rho}}(x) \in \mathcal{N}$ whenever $0 < \rho \leq \tilde{\rho}$ and $||x - \xi(a)|| \leq \delta$. Since the maps $\Phi_{b,a}^{f_{\rho}}$ are of class C^1 on $U(\delta) = \overline{\mathbb{B}}^m(\xi(a), \delta)$, and converge uniformly to $\Phi_{b,a}^f$ on $U(\delta)$, we conclude that $\mathcal{M} \in WDC(\Phi_{b,a}^f; \xi(a))$, as desired. \Box

5 The maximum principle

We now state and prove our basic version of the maximum principle, as a necessary condition for a reachable set to be separated from some other given set at the terminal point of the reference trajectory. We will then deduce from this result the usual sufficient condition for local controllability along a trajectory for Lipschitz systems, and a slightly stronger version of the usual necessary condition for optimal control.

In all three results, the basic ingredient is a *Lipschitz control system*

$$\begin{split} \dot{\xi}(t) &= f(\xi(t), \eta(t), t) \quad \text{for a.e. } t \in \operatorname{dom} \xi \,, \\ \eta(t) &\in U \quad \text{for all } t \in \operatorname{dom} \eta \,, \\ \xi(\cdot) &\in \mathcal{W}^{1,1}(M) \,, \quad \eta(\cdot) \in \mathcal{U} \,, \quad \text{and } \operatorname{dom} \xi = \operatorname{dom} \eta \,. \end{split}$$

The system is specified by a system data 4-tuple $\mathcal{D} = (M, f, U, \mathcal{U})$ such that

- (H1) M (the state space) is a manifold of class C^2 ;
- (H2) U (the control space) is a set;
- (H3) f (the dynamical law) is a family $\{f_u\}_{u \in U}$ of ppd tvvfs on M;
- (H4) \mathcal{U} (the class of admissible controllers) is a set of U-valued maps whose domain is a compact subinterval of \mathbb{R} .

Given such a data 4-tuple \mathcal{D} ,

- We let $m = \dim M$.
- We use f(x, u, t) as an alternative notation for $f_u(x, t)$.
- A *U*-control is a *U*-valued function η such that dom η is a nonempty compact subinterval of \mathbb{R} . (Then (H4) says that \mathcal{U} is a set of *U*-controls.)
- If η is a *U*-control, then
 - f_{η} denotes the ppd tvvf $M \times \mathbb{R} \ni (x, t) \hookrightarrow f(x, \eta(t), t);$
 - if $t \in \mathbb{R}$, then $f_{\eta,t}$ denotes the ppd vector field $M \ni x \hookrightarrow f(x, \eta(t), t)$;
 - if ξ is an arc in M, then $f_{\eta,\xi}(t) \stackrel{\text{def}}{=} f(\xi(t), \eta(t), t);$
 - a *trajectory* for η is a trajectory (cf. Page 28) of f_{η} .
- A *trajectory-control pair* (abbr. TCP) is a pair (ξ, η) such that η is a *U*-control and ξ is a trajectory for η .
- If $\gamma = (\xi, \eta)$ is a TCP, then the **domain** dom γ is the set dom η , which, by definition, is the same as dom ξ .
- An *admissible control* is a member of \mathcal{U} .
- A TCP (ξ, η) is *admissible* if $\eta \in \mathcal{U}$.
- We write $TCP(\mathcal{D})$, $TCP_{adm}(\mathcal{D})$, to denote, respectively, the set of all TCPs of \mathcal{D} and the set of all admissible TCPs of \mathcal{D} .

In addition, we specify x_* , N, F and S such that

(H5) $x_* \in M$, N is a manifold of class C^1 , F is a ppd map from M to N such that dom F is open and F is locally Lipschitz on dom F, and S is a subset of N,

as well as a reference interval $[a_*, b_*]$ and a reference trajectory-control pair (ξ_*, η_*) such that

(H6.a)
$$a_*, b_* \in \mathbb{R}, a_* < b_*, (\xi_*, \eta_*) \in TCP_{adm}(\mathcal{D}), and \text{ dom } \eta_* = [a_*, b_*],$$

(H6.b) $\xi_*(a_*) = x_*, \xi_*(b_*) \in \text{ dom } F \text{ and } F(\xi_*(b_*)) \in S.$

In order to state precisely the technical hypotheses on the tvvfs of the system, we first let $\mathcal{U}_{[a_*,b_*]}^c$ denote the set of all constant *U*-controls defined on $[a_*,b_*]$, and define $\mathcal{U}_{[a_*,b_*]}^{c,*} = \mathcal{U}_{[a_*,b_*]}^c \cup \{\eta_*\}$, so $\mathcal{U}_{[a_*,b_*]}^{c,*}$ consists of the reference control η_* and all the constant controls whose domain is $[a_*, b_*]$.

The key technical hypothesis on our control dynamical law is then

(H7) For each $\eta \in \mathcal{U}_{[a_*,b_*]}^{c,*}$, the twef f_{η} is integrably Lipschitz near ξ_* .

In addition to the above data, we will also specify \mathcal{C} , Λ such that

(H8.a) C is a WDC approximating multicone of S at $F(\xi_*(b_*))$, (H8.b) Λ is a Warga derivate container of F at $\xi_*(b_*)$.

Our last hypothesis will require the concept of an equal-time measurable-

variational neighborhood (abbr. ETMVN) of a controller η . We say that a set \mathcal{V} of controllers is an ETMVN of a controller η if

• for every $N \in \mathbb{N}$ and every N-tuple $\mathbf{u} = (u_1, \ldots, u_N)$ of members of U, there exists a positive number $\varepsilon = \varepsilon(N, \mathbf{u})$ such that whenever $\eta : [a_*, b_*] \mapsto U$ is a map obtained from η_* by first selecting an N-tuple $\mathbf{M} = (M_1, \ldots, M_M)$ of pairwise disjoint measurable subsets of $[a_*, b_*]$ with the property that $\sum_{j=1}^M \operatorname{meas}(M_j) \leq \varepsilon$, and then substituting the constant value u_j for the value $\eta_*(t)$ for every $j = 1, \ldots, N$ and every $t \in I_j$, it follows that $\eta \in \mathcal{U}$.

We will then assume

(H9) The class \mathcal{U} is an ETMVN of η_* .

5.1 The maximum principle for set separation

For the set separation problem, we specify a data 14-tuple

$$\mathcal{D}^{sep} = (M, f, U, \mathcal{U}, x_*, N, F, S, a_*, b_*, \xi_*, \eta_*, \mathcal{C}, \Lambda).$$
(19)

We let $\mathcal{D} = (M, f, U, \mathcal{U})$, and we define the \mathcal{D} -reachable set from x_* over the interval $[a_*, b_*]$ to be the set $\mathcal{R}_{\mathcal{D}:[a_*, b_*]}(x_*)$ given by

$$\mathcal{R}_{\mathcal{D};[a_*,b_*]}(x_*) = \{\xi(b_*) : (\xi,\eta) \in TCP_{adm}(\mathcal{D}), \, \xi(a_*) = x_*\}.$$

The local separation condition is then

(H^{sep}) there exists a neighborhood V of $F(\xi_*(b_*))$ in N such that

$$F(\mathcal{R}_{\mathcal{D};[a_*,b_*]}(x_*)) \cap S \cap V = \{F(\xi(b_*))\}$$

It will also be convenient to single out the following strong form of the negation of (H^{sep}) , that we will call the *Lispchitz arc intersection property*.

(H^{Lip,in}) There exists a Lipschitz arc
$$\gamma: [0,1] \mapsto F(\mathcal{R}_{\mathcal{D};[a_*,b_*]}(x_*)) \cap S$$
 such that $\gamma(0) = F(\xi_*(b_*))$ and $\gamma(1) \neq \gamma(0)$.

We define the **Hamiltonian of** f to be the real-valued ppd function H^f on $T^*M \times U \times \mathbb{R}$ given by

$$H^{f}(x, p, u, t) = p \cdot f(x, u, t) \quad \text{for } x \in M, \ p \in T^{*}_{x}M, \ u \in U, \ t \in \mathbb{R}.$$

The following is then our version of the Lipschitz maximum principle for set separation.

Theorem 5.1 Assume that the data \mathcal{D}^{sep} satisfy Hypotheses (H1) to (H9). Let \mathcal{L} be the set of all pairs (u, τ) such that $u \in U, \tau \in]a_*, b_*[$ and τ is a Lebesgue time along ξ_* of both time-varying vector fields f_u and f_{η_*} . Then either $(H^{Lip,in})$ holds, or

- (*) for every covector $\mu \in T^*_{F(\mathcal{E}_*(b_*))}N$ such that $\mu \neq 0$ there exists a 4-tuple (π_0, ν, λ, L) such that
 - 1. π_0 is a nonnegative real number,
 - 2. $\nu \in T^*_{F(\xi_*(b_*))}N$,
 - 3. $\lambda \in \Lambda$,
 - 4. L is a map $[a_*, b_*] \ni t \mapsto L(t) \in J^1_{\xi_*(t), \dot{\xi}_*(t)} \Gamma(TM)$, which is a measurable selection of the set-valued map $[a_*, b_*] \ni t \mapsto \partial f_{\eta_*, t}(\xi_*(t))$,
 - 5. if

 - a. π[#] = ν ∘ λ (so that π[#] ∈ T^{*}_{ξ*(b*)}M),
 b. ∇_L ∈ Cov(ξ*) is the covariant differentiation corresponding to L,
 c. π(t) = π[#] ∘ P^{∇_L}_{b*,t} for a* ≤ t ≤ b* (so that the field of covectors π is the unique absolutely continuous solution of the "adjoint Cauchy problem" $\nabla_L \pi = 0, \ \pi(b_*) = \pi^{\#}$),

then the following three conditions are satisfied:

- I. The Hamiltonian inequalities: for every pair $(u, \tau) \in \mathcal{L}$, the inequality $H^f(\xi_*(\tau), \pi(\tau), \eta_*(\tau), \tau) \geq H^f(\xi_*(\tau), \pi(\tau), u, \tau)$ holds.
- II. **Transversality:** $\pi_0 \mu \nu \in \mathcal{C}^{\perp}$.
- III. Nontriviality: $\nu \neq 0$ or $\pi_0 > 0$.

In particular, if the local separation condition H^{sep} is satisfied, then (*) holds.

Remark 5.2 The Hamiltonian inequality of the theorem obviously implies the "weak Hamiltonian maximization condition"

(I.wk) For each $u \in U$ there is a Lebesgue-null subset $\mathcal{N}(u)$ of $[a_*, b_*]$ such that $H^f(\xi_*(\tau), \pi(\tau), \eta_*(\tau), \tau) \geq H^f(\xi_*(\tau), \pi(\tau), u, \tau)$ if $\tau \notin \mathcal{N}(u)$.

Under some extra technical hypotheses, the following "strong Hamiltonian maximization condition" can then be proved.

(I.st) There exists a Lebesgue-null subset \mathcal{N} of $[a_*, b_*]$ such that the equality $H^{f}(\xi_{*}(\tau), \pi(\tau), \eta_{*}(\tau), \tau) = \max\{H^{f}(\xi_{*}(\tau), \pi(\tau), u, \tau) : u \in U\}$ holds whenever $\tau \notin \mathcal{N}$.

For example, it is easy to prove

Proposition 5.3 Under the hypotheses of Theorem 5.1, if $(H^{Lip,in})$ does not hold, π is as in the conclusion of the theorem, and in addition U is a separable metric space and the function $U \ni u \mapsto f(\xi_*(t), u, t)$ is continuous for almost every $t \in [a_*, b_*]$, then (I.st) is satisfied.

Proof. It suffices to pick for each $u \in U$ a subset $\mathcal{N}(u)$ as in condition (I.wk), and a null set B such that the function $U \ni u \mapsto f(\xi_*(t), u, t)$ is continuous for all $t \notin B$. We then let D be a countable dense subset of U, and define $\mathcal{N} = B \cup \bigcup_{u \in D} \mathcal{N}(u)$. It is then clear that the equality of (I.st) holds whenever $\tau \notin \mathcal{N}$.

Proof of Theorem 5.1. Using Hypothesis (H7), we pick, for each U-control η such that η is constant or $\eta = \eta_*$, an integrable function $\hat{k}_\eta : [a_*, b_*] \mapsto [0, +\infty]$ such that f_η is IL- \hat{k}_η near ξ_* (cf. Page 33). We then let $\hat{\mathcal{L}}$ be the set of all $(u, \tau) \in \mathcal{L}$ such that, in addition, τ is a Lebesgue point of \hat{k}_u and \hat{k}_{η_*} .

The key step of our proof will be the construction of a "needle variation" $\Psi^{u,\tau}$ for each $(u,\tau) \in \hat{\mathcal{L}}$. For this purpose, we fix a pair $(u,\tau) \in \hat{\mathcal{L}}$ (so in particular $a_* < \tau < b_*$). We then fix a 4-tuple $(\mathbf{x}, \alpha, \delta, C)$ such that \mathbf{x} is a chart of M, $\alpha, \delta, C \in]0, +\infty[, [\tau - \alpha, \tau + \alpha] \subseteq [a_*, b_*], \text{ and } \xi_*([\tau - \alpha, \tau + \alpha])$ is a subset of dom \mathbf{x} , having the property that, whenever $t \in [\tau - \alpha, \tau + \alpha]$, it follows that

(i) $\overline{\mathbb{B}}^m(\xi_*(t)^{\mathbf{x}}, \delta) \subseteq \operatorname{im} \mathbf{x},$

(ii) $f_u(x,t)$ and $f_{\eta_*}(x,t)$ are defined for every $x \in \mathbf{x}^{-1}(\bar{\mathbb{B}}^m(\xi_*(t)^{\mathbf{x}},\delta))$,

(iii) the following four inequalities hold for $x, \tilde{x} \in \mathbf{x}^{-1}(\bar{\mathbb{B}}^m(\xi_*(t)^{\mathbf{x}}, \delta))$:

$$\begin{aligned} \|f_u(x,t)^{\mathbf{x}}\| &\leq Ck_u(t) , \qquad \|f_u(x,t)^{\mathbf{x}} - f_u(\tilde{x},t)^{\mathbf{x}}\| \leq Ck_u(t) \|x^{\mathbf{x}} - \tilde{x}^{\mathbf{x}}\| ,\\ \|f_{\eta_*}(x,t)^{\mathbf{x}}\| &\leq C\hat{k}_{\eta_*}(t) , \qquad \|f_{\eta_*}(x,t)^{\mathbf{x}} - f_{\eta_*}(\tilde{x},t)^{\mathbf{x}}\| \leq C\hat{k}_{\eta_*}(t) \|x^{\mathbf{x}} - \tilde{x}^{\mathbf{x}}\| .\end{aligned}$$

We then let

$$\begin{aligned} k_u(t) &= C \hat{k}_u(t) , & \bar{k}_u = k_u(\tau) , & \bar{v}_u = f_{u,\xi_*}(\tau) , \\ k_{\eta_*}(t) &= C \hat{k}_{\eta_*}(t) , & \bar{k}_* = k_{\eta_*}(\tau) , & \bar{v}_* = f_{\eta_*,\xi_*}(\tau) , \end{aligned}$$

and define $\mathbf{L}_{u,\tau}$ to be the linear map from $T_{\xi_*(\tau)}M \times \mathbb{R}$ to $T_{\xi_*(\tau)}M$ given by

$$\mathbf{L}_{u,\tau}(\Delta x, \Delta \sigma) = \Delta x + \Delta \sigma(\bar{v}_u - \bar{v}_*) \quad \text{for } \Delta x \in T_{\xi_*(\tau)}M, \ \Delta \sigma \in \mathbb{R}.$$
(20)

The variation $\Psi^{u,\tau}$ is going to be a set-valued map, whose graph will be the union $\bigcup_{\rho \in]0,\bar{\rho}]} \operatorname{Graph}(\Psi^{u,\tau}_{\rho})$, where $\{\Psi^{u,\tau}_{\rho}\}_{0 < \rho \leq \bar{\rho}}$ is a family of single-valued maps, depending on a small positive parameter ρ . To construct the maps $\Psi^{u,\tau}_{\rho}$, we first let

$$\theta(t) = |k_{\eta_*}(t) - \bar{k}_*| + |k_u(t) - \bar{k}_u| + ||f_{u,\xi_*}(t)^{\mathbf{x}} - \bar{v}_u^{\mathbf{x}}|| + ||f_{\eta_{*},\xi_*}(t)^{\mathbf{x}} - \bar{v}_*^{\mathbf{x}}||,$$

and observe that the fact that $(u, \tau) \in \hat{\mathcal{L}}$ implies that $\lim_{h \downarrow 0} \frac{1}{h} \int_{\tau-h}^{\tau+h} \theta(t) dt = 0$.

Next, we define measurable subsets E_{ρ} of the interval $[\tau - \alpha, \tau]$ by letting $E_{\rho} = \{t \in [\tau - \alpha, \tau] : \theta(t) \le \rho\}$ if $\rho > 0$. Then, if $0 < h \le \alpha$, we have

$$\frac{1}{h} \operatorname{meas}([\tau - h, \tau] \setminus E_{\rho}) \leq \frac{1}{\rho h} \int_{[\tau - h, \tau] \setminus E_{\rho}} \theta(t) \, dt \leq \frac{1}{\rho h} \int_{\tau - h}^{\tau} \theta(t) \, dt \,,$$

so $\lim_{h\downarrow 0} \frac{1}{h} \operatorname{meas}([\tau - h, \tau] \setminus E_{\rho}) = 0$, and then $\lim_{h\downarrow 0} \frac{1}{h} \operatorname{meas}(E_{\rho} \cap [\tau - h, \tau]) = 1$.

Using the sets E_{ρ} , we define controls $\eta^{u,\tau,\rho} : [a_*, b_*] \mapsto U$ by letting $\eta^{u,\tau,\rho}(t) = u$ if $t \in E_{\rho}$ and $\eta^{u,\tau,\rho}(t) = \eta_*(t)$ if $t \notin E_{\rho}$. We then let

$$\tilde{\Psi}^{u,\tau}_{\rho}(x,\varepsilon) = \left(\Phi^{f_{\eta^{u,\tau,\rho}}}_{\tau,\tau-\varepsilon} \circ \Phi^{f_{\eta*}}_{\tau-\varepsilon,\tau}\right)(x) \tag{21}$$

for x near $\xi_*(\tau)$ and small positive ε . (In other words: we construct $\tilde{\Psi}^{u,\tau}_{\rho}(x,\varepsilon)$ by starting at x at time τ , and following a path $[0, 2\varepsilon] \ni s \mapsto \gamma_{x,\varepsilon}(s)$ in such a way that (i) we first let $\gamma_{x,\varepsilon}(s) = \Phi^{f_{\eta_*}}_{\tau-s,\tau}(x)$ for $s \in [0,\varepsilon]$, that is, we follow the trajectory of the reference control η_* backwards in time up to time $\tau - \varepsilon$, and then (ii) we let $\gamma_{x,\varepsilon}(s) = \Phi^{f_{\eta^u,\tau,\rho}}_{\tau-(2\varepsilon-s),\tau-\varepsilon}(\gamma_{x,\varepsilon}(\varepsilon))$ for $s \in [\varepsilon, 2\varepsilon]$, that is, we move forward in time up to time τ using the control $\eta^{u,\tau,\rho}$.)

We make (21) precise as follows:

• For each positive ρ , we let $\mathcal{I}(\rho)$ be the set of all positive numbers r that satisfy the inequality $4r(1+e^r)(\bar{k}_u+\bar{k}_*+2) \leq \min(\delta,\rho)$, and observe that $\mathcal{I}(\rho) \subseteq \mathcal{I}(\rho')$ whenever $0 < \rho \leq \rho'$. We then let $\bar{r}(\rho) = \sup \mathcal{I}(\rho)$, so that

$$4\bar{r}(\rho)(1+e^{\bar{r}(\rho)})(\bar{k}_{u}+\bar{k}_{*}+2) \leq \min(\delta,\rho) \quad \text{whenever } \rho > 0 \,, \, . \tag{22}$$
$$0 < \bar{r}(\rho) \leq \bar{r}(\rho') \quad \text{whenever } 0 < \rho \leq \rho' \,, \tag{23}$$
$$\lim_{\rho \downarrow 0} \bar{r}(\rho) = 0 \,. \tag{24}$$

- We then let $\mathcal{B}_r = \{x \in \operatorname{dom} \mathbf{x} : ||x^{\mathbf{x}} \xi_*(\tau)^{\mathbf{x}}|| \le r\}$, for $0 < r \le \delta$.
- For each positive r, we let $\bar{\varepsilon}(r)$ be the supremum of all the real numbers ε such that $0 \le \varepsilon \le \min(\alpha, r)$ and $2 \int_{\tau-\varepsilon}^{\tau} (k_u(t) + k_{\eta_*}(t)) dt \le r$. Then

$$0 < \bar{\varepsilon}(r) \le \min(\alpha, r) \quad \text{whenever } r > 0, \qquad (25)$$

(26)

$$2\int_{\tau-\bar{\varepsilon}(r)}^{\tau} \left(k_u(t) + k_{\eta_*}(t)\right) dt \le r \quad \text{whenever } r > 0, \qquad (27)$$

$$\bar{\varepsilon}(r) \le \bar{\varepsilon}(r')$$
 whenever $0 < r \le r'$, (28)

$$\lim_{r\downarrow 0} \bar{\varepsilon}(r) = 0.$$
⁽²⁹⁾

• We write $\bar{\varepsilon}[\rho] = \bar{\varepsilon}(\bar{r}(\rho))$, and define $\tilde{\mathbf{D}}_{\rho} = \{(x,\varepsilon) : x \in \mathcal{B}_{\bar{r}(\rho)}, 0 \le \varepsilon \le \bar{\varepsilon}[\rho]\}$. Then $\tilde{\mathbf{D}}_{\rho} \subseteq \tilde{\mathbf{D}}_{\rho'}$ whenever $0 < \rho \le \rho'$. It follows from the above choices that

- (A) If we let $\bar{c}(\rho) = 2e^{\bar{r}(\rho)}(\bar{k}_u + \bar{k}_* + \rho)$, then $2\bar{r}(\rho)\bar{c}(\rho) + 4\bar{r}(\rho)(\bar{k}_u + \bar{k}_* + 2\rho) \le \rho$ whenever $0 < \rho \le 1$. (This inequality will be used later.)
- (B) For every $r \in]0, \delta]$ the set \mathcal{B}_r is a compact neighborhood of $\xi_*(\tau)$, and the map $\mathcal{B}_r \ni x \mapsto x^{\mathbf{x}}$ is a bijection onto the compact ball $\overline{\mathbb{B}}^m(\xi_*(\tau)^{\mathbf{x}}, r)$, which is a subset of $\overline{\mathbb{B}}^m(\xi_*(\tau)^{\mathbf{x}}, \delta)$.
- (C) $\mathcal{B}_{2\bar{r}(\rho)} \subseteq \mathbf{x}^{-1}(\bar{\mathbb{B}}^m(\xi_*(t)^{\mathbf{x}},\delta))$ for every ρ and every $t \in [\tau \bar{\varepsilon}[\rho],\tau]$. (Indeed, suppose that $t \in [\tau - \bar{\varepsilon}[\rho],\tau]$ and $x \in \mathcal{B}_{2\bar{r}(\rho)}$. Then $x \in \operatorname{dom} \mathbf{x}$, because $2\bar{r}(\rho) < \delta$, and $||x^{\mathbf{x}} - \xi_*(\tau)^{\mathbf{x}}|| \leq 2\bar{r}(\rho)$. On the other hand, if $s \in [t,\tau]$ then of course $s \in [\tau - \alpha,\tau]$, so $\xi_*(s) \in \mathbf{x}^{-1}(\bar{\mathbb{B}}^m(\xi(s)^{\mathbf{x}},\delta))$, and then $\xi_*(s) \in \operatorname{dom} \mathbf{x}$. Furthemore, for almost all such s, $||\xi_*(s)^{\mathbf{x}}|| = ||f(\xi_*(s),\eta_*(s),s)^{\mathbf{x}}|| \leq k_{\eta_*}(s)$. Since this is true for almost every $s \in [t,\tau]$, it follows that

$$\|\xi_*(t)^{\mathbf{x}} - \xi_*(\tau)^{\mathbf{x}}\| \le \int_t^\tau k_{\eta_*}(s) \, ds \le \int_{\tau - \bar{\varepsilon}[\rho]}^\tau k_{\eta_*}(s) \, ds \le \bar{r}(\rho) \, .$$

Hence $||x^{\mathbf{x}} - \xi_*(t)^{\mathbf{x}}|| \le 3\bar{r}(\rho) \le \delta$, so $x \in \mathbf{x}^{-1}(\bar{\mathbb{B}}^m(\xi_*(t)^{\mathbf{x}},\delta)).)$

(D) The bounds

$$\|f_u(x,t)^{\mathbf{x}}\| \le k_u(t) , \qquad \|f_u(x,t)^{\mathbf{x}} - f_u(\tilde{x},t)^{\mathbf{x}}\| \le k_u(t) \|x^{\mathbf{x}} - \tilde{x}^{\mathbf{x}}\| , \\ \|f_{\eta_*}(x,t)^{\mathbf{x}}\| \le k_{\eta_*}(t) , \qquad \|f_{\eta_*}(x,t)^{\mathbf{x}} - f_{\eta_*}(\tilde{x},t)^{\mathbf{x}}\| \le k_{\eta_*}(t) \|x^{\mathbf{x}} - \tilde{x}^{\mathbf{x}}\|$$

hold, for every ρ , whenever $x, \tilde{x} \in \mathcal{B}_{2\bar{r}(\rho)}$ and $t \in [\tau - \bar{\varepsilon}[\rho], \tau]$. (This follows from the fact that $\mathcal{B}_{2\bar{r}(\rho)} \subseteq \mathbf{x}^{-1}(\mathbb{B}^m(\xi_*(t)^{\mathbf{x}}, \delta).)$

- (E) For every ρ , if $x \in \mathcal{B}_{\bar{r}(\rho)}$, then $\Phi_{t,\tau}^{f_{\eta_*}}(x)$ is defined and belongs to dom **x** for every $t \in [\tau - \bar{\varepsilon}[\rho], \tau]$, and $\|\Phi_{t,\tau}^{f_{\eta_*}}(x)^{\mathbf{x}} - x^{\mathbf{x}}\| \leq \int_{\tau - \bar{\varepsilon}[\rho]}^{\tau} k_{\eta_*}(s) ds \leq \bar{r}(\rho)$, so in particular $\Phi_{t,\tau}^{f_{\eta_*}}(x) \in \mathcal{B}_{2\bar{r}(\rho)}$.
- (F) For every ρ , if $(x,\varepsilon) \in \tilde{\mathbf{D}}_{\rho}$, then $\Phi_{t,\tau-\varepsilon}^{f_{\eta u,\tau,\rho}}\left(\Phi_{\tau-\varepsilon,\tau}^{f_{\eta *}}(x)\right)$ is defined and belongs to dom \mathbf{x} for every $t \in [\tau - \varepsilon, \tau]$. Furthermore, $\|\Phi_{t,\tau-\varepsilon}^{f_{\eta u,\tau,\rho}}\left(\Phi_{\tau-\varepsilon,\tau}^{f_{\eta *}}(x)\right)^{\mathbf{x}} - x^{\mathbf{x}}\| \leq \int_{\tau-\overline{\varepsilon}[\rho]}^{\tau} (2k_{\eta *}(s) + k_{u}(s)) \, ds \leq \overline{r}(\rho)$, so that, in particular $\Phi_{t,\tau-\varepsilon}^{f_{\eta u,\tau,\rho}}\left(\Phi_{\tau-\varepsilon,\tau}^{f_{\eta *}}(x)\right) \in \mathcal{B}_{2\overline{r}(\rho)}$.

It follows from (E) and (F) that $\tilde{\Psi}_{\rho}^{u,\tau}(x,\varepsilon)$ is defined and belongs to $\mathcal{B}_{2\bar{r}(\rho)}$ whenever (x,ε) belongs to $\tilde{\mathbf{D}}_{\rho}$. Furthermore, all the "intermediate points of the construction of $\Psi_{\rho}^{u,\tau}(x,\varepsilon)$ "—that is, the points that lie on the path $\gamma_{x,\varepsilon}$ described above—belong to $\mathcal{B}_{2\bar{r}(\rho)}$. Therefore, at all these points the bounds of (D) hold. Hence all the calculations involving these points take place within $\mathcal{B}_{2\bar{r}(\rho)}$ and as long as we never leave $\mathcal{B}_{2\bar{r}(\rho)}$ we can do our calculations by identifying the points $x \in M$ with their **x**-coordinate representations, that is, by just writing "x" when we really mean " $x^{\mathbf{x}}$ ". We will use this notational simplification, until we arrive at conclusions that are manifestly independent of the chart.

Simple applications of Gronwall's inequality then yield the inequalities

$$\|\tilde{\Psi}^{u,\tau}_{\rho}(x,\varepsilon) - \tilde{\Psi}^{u,\tau}_{\rho}(\tilde{x},\varepsilon)\| \le e^{\bar{r}(\rho)} \|x - \tilde{x}\|, \qquad (30)$$

$$\|\tilde{\Psi}^{u,\tau}_{\rho}(x,\varepsilon) - \tilde{\Psi}^{u,\tau}_{\rho}(\tilde{x},\varepsilon) - (x-\tilde{x})\| \le \bar{r}(\rho) e^{\bar{r}(\rho)} \|x-\tilde{x}\| \le \rho \|x-\tilde{x}\|, \quad (31)$$

if (x,ε) and (\tilde{x},ε) belong to $\tilde{\mathbf{D}}_{\rho}$.

We now estimate $\|\tilde{\Psi}_{\rho}^{u,\tau}(x,\varepsilon_1) - \tilde{\Psi}_{\rho}^{u,\tau}(x,\varepsilon_2)\|$, for $(x,\varepsilon_1) \in \tilde{\mathbf{D}}_{\rho}$ and $(x,\varepsilon_2) \in \tilde{\mathbf{D}}_{\rho}$. Assume first that $\varepsilon_2 < \varepsilon_1$. Let $y = \Phi_{\tau-\varepsilon_1,\tau}^{f_{\eta_*}}(x)$, and write $\xi(t) = \Phi_{t,\tau-\varepsilon_1}^{f_{\eta^u,\tau,\rho}}(y)$ and $\tilde{\xi}(t) = \Phi_{t,\tau-\varepsilon_1}^{f_{\eta}}(y)$ for $\tau - \varepsilon_1 < t \leq \tau$. Then

$$\begin{split} \tilde{\Psi}^{u,\tau}_{\rho}(x,\varepsilon_1) &= \left(\Phi^{f_{\eta^u,\tau,\rho}}_{\tau,\tau-\varepsilon_1} \circ \Phi^{f_{\eta_*}}_{\tau-\varepsilon_1,\tau}\right)(x) \\ &= \Phi^{f_{\eta^u,\tau,\rho}}_{\tau,\tau-\varepsilon_1}(y) \\ &= \xi(\tau) , \\ \tilde{\Psi}^{u,\tau}_{\rho}(x,\varepsilon_2) &= \left(\Phi^{f_{\eta^u,\tau,\rho}}_{\tau,\tau-\varepsilon_2} \circ \Phi^{f_{\eta_*}}_{\tau-\varepsilon_2,\tau}\right)(x) \\ &= \left(\Phi^{f_{\eta^u,\tau,\rho}}_{\tau,\tau-\varepsilon_2} \circ \Phi^{f_{\eta_*}}_{\tau-\varepsilon_2,\tau-\varepsilon_1} \circ \Phi^{f_{\eta_*}}_{\tau-\varepsilon_1,\tau}\right)(x) \\ &= \left(\Phi^{f_{\eta^u,\tau,\rho}}_{\tau,\tau-\varepsilon_2} \circ \Phi^{f_{\eta_*}}_{\tau-\varepsilon_2,\tau-\varepsilon_1}\right)(y) \\ &= \Phi^{f_{\eta}}_{\tau,\tau-\varepsilon_1}(y) \\ &= \tilde{\xi}(\tau) , \end{split}$$

where $\tilde{\eta}$ is any U-control such that $\tilde{\eta}(t) = \eta_*(t)$ for $\tau - \varepsilon_1 \leq t \leq \tau - \varepsilon_2$ and $\tilde{\eta}(t) = \eta^{u,\tau,\rho}(t)$ for $\tau - \varepsilon_2 < t \leq \tau$. Then, if we write $k = k_u + k_{\eta_*}$, and let

$$S(t) = [\tau - \varepsilon_1, \min(t, \tau - \varepsilon_2)] \cap E_\rho, \quad a(s) = f(\tilde{\xi}(s), u, s) - f(\tilde{\xi}(s), \eta_*(s), s),$$

we find

$$\begin{split} \xi(t) - \tilde{\xi}(t) &= \int_{\tau-\varepsilon_1}^t \left(f(\xi(s), \eta^{u,\tau,\rho}(s), s) - f(\tilde{\xi}(s), \tilde{\eta}(s), s) \right) ds \\ &= \int_{\tau-\varepsilon_1}^t \left(f(\xi(s), \eta^{u,\tau,\rho}(s), s) - f(\tilde{\xi}(s), \eta^{u,\tau,\rho}(s), s) \right) ds \end{split}$$

$$\begin{aligned} + \int_{\tau-\varepsilon_{1}}^{t} \left(f(\tilde{\xi}(s), \eta^{u,\tau,\rho}(s), s) - f(\tilde{\xi}(s), \tilde{\eta}(s), s)\right) ds \\ = \int_{\tau-\varepsilon_{1}}^{t} \left(f(\xi(s), \eta^{u,\tau,\rho}(s), s) - f(\tilde{\xi}(s), \eta^{u,\tau,\rho}(s), s)\right) ds \\ + \int_{[\tau-\varepsilon_{1}, \min(t, \tau-\varepsilon_{2})] \cap E_{\rho}} \left(f(\tilde{\xi}(s), u, s) - f(\tilde{\xi}(s), \eta_{*}(s), s)\right) ds \\ = \int_{\tau-\varepsilon_{1}}^{t} \left(f^{\eta^{u,\tau,\rho}}(\xi(s), s) - f^{\eta^{u,\tau,\rho}}(\tilde{\xi}(s), s)\right) ds + \int_{S(t)} a(s) ds , \quad (32) \\ \|\xi(t) - \tilde{\xi}(t)\| \leq \int_{\tau-\varepsilon_{1}}^{t} k(s) \|\xi(s) - \tilde{\xi}(s)\| \, ds + \int_{S(t)} \|a(s)\| \, ds . \quad (33) \end{aligned}$$

If s belongs to E_{ρ} , then $\|f(\tilde{\xi}(s), u, s)\| \leq k(s)$ and $\|f(\tilde{\xi}(s), \eta_*(s), s)\| \leq k(s)$. Furthermore, $k(s) \leq \bar{k}_u + \bar{k}_* + \rho$. It follows that $\|a(s)\| \leq 2(\bar{k}_u + \bar{k}_* + \rho)$, so that $\int_{S(t)} \|a(s)\| ds \leq 2(\bar{k}_u + \bar{k}_* + \rho) \max(S(t))$, from which we conclude that

$$\int_{S(t)} \|a(s)\| \, ds \le 2(\bar{k}_u + \bar{k}_* + \rho) \operatorname{meas}([\tau - \varepsilon_1, \tau - \varepsilon_2] \cap E_\rho) \, .$$

Let $\hat{\sigma}^{\rho}(\varepsilon) = \text{meas}([\tau - \varepsilon, \tau] \cap E_{\rho})$. It then follows immediately that

$$\operatorname{meas}([\tau - \varepsilon_1, \tau - \varepsilon_2] \cap E_{\rho}) = \hat{\sigma}^{\rho}_{\varepsilon_1, \varepsilon_2},$$

where $\hat{\sigma}^{\rho}_{\varepsilon_1,\varepsilon_2} = \hat{\sigma}^{\rho}(\varepsilon_1) - \hat{\sigma}^{\rho}(\varepsilon_2)$. This in turn implies that

$$\int_{S(t)} \|a(s)\| \, ds \le 2(\bar{k}_u + \bar{k}_* + \rho) \, \hat{\sigma}^{\rho}_{\varepsilon_1, \varepsilon_2} \, .$$

This fact, together with (33) and Gronwall's inequality, imply, if we write $\bar{c}(\rho) = 2e^{\bar{r}(\rho)}(\bar{k}_u + \bar{k}_* + \rho)$, that

$$\|\xi(t) - \tilde{\xi}(t)\| \le 2(\bar{k}_u + \bar{k}_* + \rho)e^{\int_{\tau-\varepsilon_1}^t k(s)\,ds}\hat{\sigma}_{\varepsilon_1,\varepsilon_2}^{\rho} \le \bar{c}(\rho)\hat{\sigma}_{\varepsilon_1,\varepsilon_2}^{\rho}.$$

If we take $t = \tau$, then $\xi(\tau) = \tilde{\Psi}^{u,\tau}_{\rho}(x,\varepsilon_1)$ and $\tilde{\xi}(\tau) = \tilde{\Psi}^{u,\tau}_{\rho}(x,\varepsilon_2)$, so we have proved that $\|\tilde{\Psi}^{u,\tau}_{\rho}(x,\varepsilon_1) - \tilde{\Psi}^{u,\tau}_{\rho}(x,\varepsilon_2)\| \leq \bar{c}(\rho)\hat{\sigma}^{\rho}_{\varepsilon_1,\varepsilon_2}$, under the assumption that $\varepsilon_1 > \varepsilon_2$. A similar estimate is clearly valid when $\varepsilon_1 < \varepsilon_2$, and we then get the unrestricted estimate

$$\|\tilde{\Psi}^{u,\tau}_{\rho}(x,\varepsilon_1) - \tilde{\Psi}^{u,\tau}_{\rho}(x,\varepsilon_2)\| \le \bar{c}(\rho) \,|\hat{\sigma}^{\rho}_{\varepsilon_1,\varepsilon_2}| \quad \text{for} \quad x \in \mathcal{B}_{\bar{r}(\rho)}, \, \varepsilon_1, \varepsilon_2 \in [0,\bar{\varepsilon}[\rho]]. \tag{34}$$

In addition, if $s \in E_{\rho}$ we have

$$\begin{aligned} \|f(\xi(s), u, s) - \bar{v}_u\| &\leq \|f(\xi(s), u, s) - f(\xi_*(s), u, s)\| + \|f(\xi_*(s), u, s) - \bar{v}_u\| \\ &\leq k_u(s) \|\tilde{\xi}(s) - \xi_*(s)\| + \rho \leq 4\bar{r}(\rho)(\bar{k}_u + \rho) + \rho \,. \end{aligned}$$

Similarly, $||f(\tilde{\xi}(s), \eta_*(s), s) - \bar{v}_*|| \le 4\bar{r}(\rho)(\bar{k}_* + \rho) + \rho$. On the other hand,

$$||a(s) - (\bar{v}_u - \bar{v}_*)|| \le ||f(\tilde{\xi}(s), u, s) - \bar{v}_u|| + ||f(\tilde{\xi}(s), \eta_*(s), s) - \bar{v}_*||,$$

and then $||a(s) - (\bar{v}_u - \bar{v}_*)|| \le 4\bar{r}(\rho)(\bar{k}_u + \bar{k}_* + 2\rho) + 2\rho.$

Clearly, (32) implies, if we write $E_{\rho}^{\varepsilon_1,\varepsilon_2} = [\tau - \varepsilon_1, \tau - \varepsilon_2] \cap E_{\rho}$, that

$$\xi(\tau) - \tilde{\xi}(\tau) = \int_{\tau-\varepsilon_1}^{\tau} \left(f^{\eta^{u,\tau,\rho}}(\xi(s),s) - f^{\eta^{u,\tau,\rho}}(\tilde{\xi}(s),s) \right) ds + \int_{E_{\rho}^{\varepsilon_1,\varepsilon_2}} a(s) ds ,$$

from which it follows, using (A), that if $0 < \rho \leq 1$, then

$$\begin{split} \|\tilde{\Psi}_{\rho}^{u,\tau}(x,\varepsilon_{1}) - \tilde{\Psi}_{\rho}^{u,\tau}(x,\varepsilon_{2}) - (\hat{\sigma}^{\rho}(\varepsilon_{1}) - \hat{\sigma}^{\rho}(\varepsilon_{2}))(\bar{v}_{u} - \bar{v}_{*})\| \\ \leq 2\bar{r}(\rho)\bar{c}(\rho) \left|\hat{\sigma}_{\varepsilon_{1},\varepsilon_{2}}^{\rho}\right| + \left(4\bar{r}(\rho)(\bar{k}_{u} + \bar{k}_{*} + 2\rho) + 2\rho\right) \left|\hat{\sigma}_{\varepsilon_{1},\varepsilon_{2}}^{\rho}\right| \leq \rho \left|\hat{\sigma}_{\varepsilon_{1},\varepsilon_{2}}^{\rho}\right|. \tag{35}$$

(This inequality has been proved assuming that $\varepsilon_1 > \varepsilon_2$, and then it follows, by interchanging ε_1 and ε_2 , that it is also true for $\varepsilon_1 \leq \varepsilon_2$.)

The function $\hat{\sigma}^{\rho} : [0, \bar{\varepsilon}[\rho]] \mapsto \mathbb{R}$ is nonnegative, monotonically nondecreasing, and satisfies $\hat{\sigma}^{\rho}(0) = 0$ and $\hat{\sigma}^{\rho}(\bar{\varepsilon}[\rho]) = \bar{\sigma}[\rho] > 0$, where we define $\bar{\sigma}[\rho]$ by letting $\bar{\sigma}[\rho] = \text{meas}([\tau - \bar{\varepsilon}[\rho], \tau] \cap E_{\rho})$. The function need not be strictly increasing, so $\hat{\sigma}^{\rho}$ need not be invertible as a map from $[0, \bar{\varepsilon}[\rho]]$ to $[0, \bar{\sigma}[\rho]]$. On the other hand, $\hat{\sigma}^{\rho}$ is continuous, so $\hat{\sigma}^{\rho}$ maps $[0, \bar{\varepsilon}[\rho]]$ onto $[0, \bar{\sigma}[\rho]]$, and (34) tells us that $\tilde{\Psi}^{u,\tau}_{\rho}(x, \varepsilon_1) = \tilde{\Psi}^{u,\tau}_{\rho}(x, \varepsilon_2)$ if $\hat{\sigma}^{\rho}(\varepsilon_1) = \hat{\sigma}^{\rho}(\varepsilon_2)$. It follows that we can "change variables and use $\sigma \in [0, \bar{\sigma}[\rho]]$ instead of $\varepsilon \in [0, \bar{\varepsilon}[\rho]]$." Precisely, we define $\mathbf{D}_{\rho} = \{(x, \sigma) : x \in \mathcal{B}_{\bar{r}(\rho)}, 0 \leq \sigma \leq \bar{\sigma}[\rho]\}$ and, for $(x, \sigma) \in \mathbf{D}_{\rho}$, we let $\Psi^{u,\tau}_{\rho}(x, \sigma) = \tilde{\Psi}^{u,\tau}_{\rho}(x, \varepsilon)$.

Then (34) says that $\|\Psi_{\rho}^{u,\tau}(x,\sigma_1) - \Psi_{\rho}^{u,\tau}(x,\sigma_2)\| \leq \bar{c}(\rho) |\sigma_1 - \sigma_2|$ whenever $(x,\sigma_1) \in \mathbf{D}_{\rho}$ and $(x,\sigma_2) \in \mathbf{D}_{\rho}$. If we combine this with (30), we get the Lipschitz estimate

$$\|\Psi_{\rho}^{u,\tau}(x_1,\sigma_1) - \Psi_{\rho}^{u,\tau}(x_2,\sigma_2)\| \le e^{\bar{r}(\rho)} \|x_1 - x_2\| + \bar{c}(\rho) |\sigma_1 - \sigma_2|, \qquad (36)$$

valid whenever (x_1, σ_1) and (x_2, σ_2) belong to \mathbf{D}_{ρ} and $0 < \rho \leq 1$.

Also, if we combine (31) and (35), we get the estimate

$$\begin{split} \|\Psi_{\rho}^{u,\tau}(x_1,\sigma_1) - \Psi_{\rho}^{u,\tau}(x_2,\sigma_{@}) - (x_1 - x_2) - (\sigma_1 - \sigma_2)(\bar{v}_u - \bar{v}_*)\| \\ \leq \rho(\|x_1 - x_2\| + |\sigma_1 - \sigma_2|). \end{split}$$
(37)

The map $\Psi_{\rho}^{u,\tau} : \mathbf{D}_{\rho} \mapsto \mathbb{R}^{m}$ satisfies $\Psi_{\rho}^{u,\tau}(x,0) = x$. So we can extend $\Psi_{\rho}^{u,\tau}$ to the set $\hat{\mathbf{D}}_{\rho} \stackrel{\text{def}}{=} \mathcal{B}_{\bar{r}(\rho)} \times [-\bar{\sigma}[\rho], \bar{\sigma}[\rho]]$, by requiring that the maps $\sigma \mapsto \Psi_{\rho}^{u,\tau}(x,\sigma) - x$ be odd, i.e., by defining $\Psi_{\rho}^{u,\tau}(x,\sigma) = 2x - \Psi_{\rho}^{u,\tau}(x,-\sigma)$ for $(x,-\sigma) \in \mathbf{D}_{\rho}$. We use the same expression $\Psi_{\rho}^{u,\tau}$ for the extended map. Then

(G) If $0 < \rho \le 1$, then the bounds (36), (37) hold for $(x_i, \sigma_i) \in \hat{\mathbf{D}}_{\rho}$, i = 1, 2. (H) $\Psi_{\rho}^{u,\tau}(x,0) = x$ whenever $x \in \mathcal{B}_{\bar{r}(\rho)}$. (I) If $a \le \hat{\tau} < \tau$, then there exists a positive number $\hat{\rho}$ such that (I.*) If $0 < \rho \le \hat{\rho}$, then $\Phi_{\hat{\tau},\tau}^{f_{\eta_*}}(x)$ is defined for every $x \in \mathcal{B}_{\bar{r}(\rho)}$, and $\Psi_{\rho}^{u,\tau}(x,\sigma) \in \mathcal{R}_{\mathcal{D};[\hat{\tau},\tau]}(\Phi_{\hat{\tau},\tau}^{f_{\eta_*}}(x))$ whenever $(x,\sigma) \in \mathbf{D}_{\rho}$.

(To prove (I), we first observe that $\Phi_{\hat{\tau},\tau}^{f_{\eta_*}}(x)$ is defined for $x = \xi_*(\tau)$, so there is a neighborhood \mathcal{N} of $\xi_*(\tau)$ such that $\Phi_{\hat{\tau},\tau}^{f_{\eta_*}}(x)$ is defined for all $x \in \mathcal{N}$. Since (24) and (29) imply that $\lim_{\rho \downarrow 0} \bar{\varepsilon}[\rho] = 0$, we may pick $\hat{\rho}$ such that $\bar{\varepsilon}[\hat{\rho}] < \tau - \hat{\tau}$ and $\mathcal{B}_{\bar{\tau}(\hat{\rho})} \subseteq \mathcal{N}$. Since (23) and (28) imply that the functions $\rho \mapsto \bar{\tau}(\rho)$ and $\rho \mapsto \bar{\varepsilon}[\rho]$ are increasing, it follows that

$$\bar{\varepsilon}[\rho] < \tau - \hat{\tau} \quad \text{and} \quad \mathcal{B}_{\bar{r}(\rho)} \subseteq \mathcal{B}_{\bar{r}(\hat{\rho})} \subseteq \mathcal{N} \text{ whenever } 0 < \rho \leq \hat{\rho} \,.$$

This implies, in particular, that if $0 < \rho \leq \hat{\rho}$ then $\Phi_{\hat{\tau},\tau}^{f_{\eta_*}}(x)$ is defined for all $x \in \mathcal{B}_{\bar{r}(\rho)}$. Furthermore, if $0 < \rho \leq \hat{\rho}$ and $(x,\sigma) \in \mathbf{D}_{\rho}$, then $x \in \mathcal{B}_{\bar{r}(\rho)}$, so $\Phi_{\hat{\tau},\tau}^{f_{\eta_*}}(x)$ is defined, and, if we let $z = \Phi_{\hat{\tau},\tau}^{f_{\eta_*}}(x)$, and pick ε such that $0 \leq \varepsilon \leq \bar{\varepsilon}[\rho]$ and $\sigma = \hat{\sigma}_{\rho}(\varepsilon)$, then

$$\begin{split} \Psi^{u,\tau}_{\rho}(x,\sigma) &= \Psi^{u,\tau}_{\rho}(\Phi^{f_{\eta_*}}_{\tau,\hat{\tau}}(z),\sigma) = \Psi^{u,\tau}_{\rho}(\Phi^{f_{\eta_*}}_{\tau,\hat{\tau}}(z),\hat{\sigma}_{\rho}(\varepsilon)) = \tilde{\Psi}^{u,\tau}_{\rho}(\Phi^{f_{\eta_*}}_{\tau,\hat{\tau}}(z),\varepsilon) \\ &= (\Phi^{f_{\eta^{u,\rho,\tau}}}_{\tau,\tau-\varepsilon} \circ \Phi^{f_{\eta_*}}_{\tau-\varepsilon,\tau})(\Phi^{f_{\eta_*}}_{\tau,\hat{\tau}}(z)) = \Phi^{f_{\eta^{u,\rho,\tau}}}_{\tau,\tau-\varepsilon}(\Phi^{f_{\eta_*}}_{\tau-\varepsilon,\hat{\tau}}(z)) = \Phi^{f_{\eta^{u,\rho,\tau}}}_{\tau,\hat{\tau}}(z) \,, \end{split}$$

showing that $\Psi_{\rho}^{u,\tau}(x,\sigma)$ is reachable from z over the interval $[\tilde{\tau},\tau]$.)

The bound (36) tells us that the map $\Psi_{\rho}^{u,\tau}$ is Lipschitz, and then (37) enables us to determine, approximately, the Clarke generalized Jacobian $\partial \Psi_{\rho}^{u,\tau}(\xi_*(\tau), 0)$. Indeed, if $\Psi_{\rho}^{u,\tau}$ is classically differentiable at a point (x, σ) , and the differential $D\Psi_{\rho}^{u,\tau}(x, \sigma)$ is the linear map $L : \mathbb{R}^m \times \mathbb{R} \mapsto \mathbb{R}^m$, then, if we write

$$A(\varepsilon, x, \sigma, \Delta x, \Delta \sigma) = \Psi_{\rho}^{u, \tau}(x + \varepsilon \Delta x, \sigma + \varepsilon \Delta \sigma) - \Psi_{\rho}^{u, \tau}(x, \sigma),$$

it follows that $\lim_{\varepsilon \downarrow 0} \frac{1}{\varepsilon} A(\varepsilon, x, \sigma, \Delta x, \Delta \sigma) = L(\Delta x, \Delta \sigma)$. On the other hand, (37) implies, if $\mathbf{L}_{u,\tau}$ is the linear map defined in (20), that

$$\left\| \frac{1}{\varepsilon} A(\varepsilon, x, \sigma, \Delta x, \Delta \sigma) - \mathbf{L}_{u,\tau}(\Delta x, \Delta \sigma) \right\|$$

= $\frac{1}{\varepsilon} \left\| A(\varepsilon, x, \sigma, \Delta x, \Delta \sigma) - \mathbf{L}_{u,\tau}(\varepsilon \Delta x, \varepsilon \Delta \sigma) \right\|$

$$\begin{split} &= \frac{1}{\varepsilon} \left\| \hat{\Psi}_{\rho}^{u,\tau} (x + \varepsilon \Delta x, \sigma + \varepsilon \Delta \sigma) - \hat{\Psi}_{\rho}^{u,\tau} (x, \sigma) - \varepsilon \Delta x - \varepsilon \Delta \sigma (\bar{v}_u - \bar{v}_*) \right\| \\ &\leq \frac{\rho}{\varepsilon} (\varepsilon \|\Delta x\| + \varepsilon |\Delta \sigma|) \\ &= \rho (\|\Delta x\| + |\Delta \sigma|) \\ &\leq 2\rho (\|\Delta x\|^2 + |\Delta \sigma|^2)^{1/2} \,, \end{split}$$

since $\|\Delta x\| + |\Delta \sigma| \leq 2(\|\Delta x\|^2 + |\Delta \sigma|^2)$, and $\sqrt{2} < 2$. We may then let ε converge to 0, and find that $\|(L - \mathbf{L}_{u,\tau})(\Delta x, \Delta \sigma)\| \leq 2\rho(\|\Delta x\|^2 + |\Delta \sigma|^2)^{1/2}$, so $\|L - \mathbf{L}_{u,\tau}\| \leq 2\rho$.

Let $\Lambda_{u,\tau}(\rho)$ be the set of all linear maps $L : \mathbb{R}^m \times \mathbb{R} \to \mathbb{R}^m$ such that $\|L - \mathbf{L}_{u,\tau}\| \leq 2\rho$. Then we have shown that all the derivatives of $\Psi_{\rho}^{u,\tau}$, at all points $(x,\sigma) \in diff(\Psi_{\rho}^{u,\tau})$, belong to $\Lambda_{u,\tau}(\rho)$. Since $\Lambda_{u,\tau}(\rho)$ is compact and convex, it follows that $\partial \Psi_{\rho}^{u,\tau}(\xi_*(\tau), 0) \subseteq \Lambda_{u,\tau}(\rho)$.

We now let $\Psi^{u,\tau}$ be the set-valued map from $M \times \mathbb{R}$ to M such that $y \in \Psi^{u,\tau}(x,\sigma)$ if and only if $y = \Psi^{u,\tau}_{\rho}(x,\sigma)$ for some ρ such that $(x,\sigma) \in \hat{\mathbf{D}}_{\rho}$. Then

- (#) The set $\{\mathbf{L}_{u,\tau}\}$ is a Warga derivate container of $\Psi^{u,\tau}$ at $(\xi_*(\tau), 0)$.
- (##) If $(u, \tau) \in \hat{\mathcal{L}}$, then, given any $\hat{\tau}$ such that $a_* \leq \hat{\tau} < \tau$, the set-valued map $\Psi^{u,\tau}$ is such that $\Psi^{u,\tau}(\Phi_{\tau,\hat{\tau}}(z),\sigma) \subseteq \mathcal{R}_{\mathcal{D};[\hat{\tau},\tau]}(z)$ whenever (z,σ) belongs to a sufficiently small neighborhood of $(\xi_*(\hat{\tau}), 0)$ in $M \times \mathbb{R}$ and $\sigma \geq 0$.

We are now ready to combine the one-parameter needle variations $\Psi^{u,\tau}$ into multiparameter variations. Suppose first that we are given a finite subset \mathcal{F} of $\hat{\mathcal{L}}$, such that the times τ of the pairs $(u,\tau) \in \mathcal{F}$ are all different. We can then write $\mathcal{F} = \{(u_1,\tau_1),\ldots,(u_N,\tau_N)\}$, where $u_1,\ldots,u_N \in U$ and $a_* < \tau_1 < \tau_2 < \cdots < \tau_N < b_*$. Fix a family $\hat{\boldsymbol{\tau}} = \{\hat{\tau}_j\}_{j=1}^N$. of times $\hat{\tau}_j$ such that

$$a < \hat{\tau}_1 < \tau_1 < \hat{\tau}_2 < \tau_2 < \dots < \hat{\tau}_{N-1} < \tau_{N-1} < \hat{\tau}_N < \tau_N < b$$
.

We then let $X_i = T_{\xi_*(\tau_i)}M$, $Y_i = Lin(X_i \times \mathbb{R}, X_i)$, and write $\Psi^{(i)} = \Psi^{u_i, \tau_i}$. Then there exist neighborhoods $\mathcal{N}^{(i)}$ of $\xi_*(\hat{\tau}_i)$, and positive numbers $\bar{\sigma}^{(i)}$ such that $\Psi^{(i)}(\Phi_{\tau_i,\hat{\tau}_i}(z), \sigma) \subseteq \mathcal{R}_{\mathcal{D};[\hat{\tau}_i, \tau_i]}(z)$ whenever $z \in \mathcal{N}^{(i)}$ and $\sigma \in [0, \bar{\sigma}^{(i)}]$.

Define $\Sigma^{(i)} = [-\bar{\sigma}^{(1)}, \bar{\sigma}^{(1)}] \times \cdots \times [-\bar{\sigma}^{(i)}, \bar{\sigma}^{(i)}], \Sigma^{(i),+} = [0, \bar{\sigma}^{(1)}] \times \cdots \times [0, \bar{\sigma}^{(i)}].$ Then construct set-valued maps $\Upsilon^{(i)} : \Sigma^{(i)} \mapsto 2^M$, for $i = 1, \ldots, N$, by first letting $\Upsilon^{(1)}(\sigma_1) = \Psi^{(1)}(\xi_*(\tau_1), \sigma_1)$, and then defining the $\Upsilon^{(i)}$ recursively for i > 1 by letting $\Upsilon^{(i)}(\sigma_i) = \Psi^{(i)}(\Phi^{f_{\eta_*}}_{\tau_i,\tau_{i-1}}(\Upsilon^{(i-1)}(\sigma_{i-1})), \sigma_i))$ for i > 1, where we write $\sigma_i = (\sigma_1, \ldots, \sigma_i)$. It follows that $\Upsilon^{(1)}(\sigma_1) \subseteq \mathcal{R}_{\mathcal{D};[a_*,\tau_1]}(\xi_*(a_*))$ if $\sigma_1 \in [0, \bar{\sigma}^{(1)}]$, because $\Upsilon^{(1)}(\sigma_1) = \Psi^{(1)}(\xi_*(\tau_1), \sigma_1) = \Psi^{(1)}(\Phi^{f_{\eta_*}}_{\tau_1,\hat{\tau}_1}(\xi_*(\hat{\tau}_1)), \sigma_1)$, so $\Upsilon^{(1)}(\sigma_1) \subseteq \mathcal{R}_{\mathcal{D};[\hat{\tau}_1,\tau_1]}(\xi_*(\hat{\tau}_1)) \subseteq \mathcal{R}_{\mathcal{D};[a_*,\tau_1]}(\xi_*(a_*))$. It is then easy to prove inductively that $\Upsilon^{(i)}(\sigma_i) \subseteq \mathcal{R}_{\mathcal{D};[a_*,\tau_i]}(\xi_*(a_*))$ for every i and every $\sigma_i \in \Sigma^{(i),+}$. (Indeed, assume that i > 1 and $\Upsilon^{(i-1)}(\boldsymbol{\sigma}_{i-1}) \subseteq \mathcal{R}_{\mathcal{D};[a_*,\tau_{i-1}]}(\xi_*(a_*))$ whenever $\boldsymbol{\sigma}_{i-1} \in \Sigma^{(i-1),+}$. Then, if $\boldsymbol{\sigma}_i \in \Sigma^{(i),+}$, we have

$$\Upsilon^{(i)}(\boldsymbol{\sigma}_{i}) = \Psi^{(i)} \Big(\Phi^{f_{\eta_{*}}}_{\tau_{i},\tau_{i-1}}(\Upsilon^{(i-1)}(\boldsymbol{\sigma}_{i-1})), \sigma_{i}) \Big) \\ = \Psi^{(i)} \Big(\Phi^{f_{\eta_{*}}}_{\tau_{i},\hat{\tau}_{i}}(\Phi^{f_{\eta_{*}}}_{\hat{\tau}_{i},\tau_{i-1}}(\Upsilon^{(i-1)}(\boldsymbol{\sigma}_{i-1}))), \sigma_{i} \Big).$$

 \mathbf{SO}

$$\Upsilon^{(i)}(\boldsymbol{\sigma}_{i}) \subseteq \mathcal{R}_{\mathcal{D};[\hat{\tau}_{i},\tau_{i}]} \left(\Phi^{f_{\eta_{*}}}_{\hat{\tau}_{i},\tau_{i-1}}(\Upsilon^{(i-1)}(\boldsymbol{\sigma}_{i-1})) \right)$$
$$\stackrel{\text{def}}{=} \bigcup \left\{ \mathcal{R}_{\mathcal{D};[\hat{\tau}_{i},\tau_{i}]}(y) : y \in \Phi^{f_{\eta_{*}}}_{\hat{\tau}_{i},\tau_{i-1}}(\Upsilon^{(i-1)}(\boldsymbol{\sigma}_{i-1})) \right\}.$$

Since

$$\Phi_{\hat{\tau}_i,\tau_{i-1}}^{f_{\eta_*}}(\Upsilon^{(i-1)}(\boldsymbol{\sigma}_{i-1})) \subseteq \mathcal{R}_{\mathcal{D};[\tau_{i-1},\hat{\tau}_i]}(\Upsilon^{(i-1)}(\boldsymbol{\sigma}_{i-1})),$$

and

 $\Upsilon^{(i-1)}(\boldsymbol{\sigma}_{i-1})) \subseteq \mathcal{R}_{\mathcal{D};[a_*,\tau_{i-1}]}(\xi_*(a_*)),$

the desired conclusion that $\Upsilon^{(i)}(\boldsymbol{\sigma}_i) \subseteq \mathcal{R}_{\mathcal{D};[a_*,\tau_i]}(\xi_*(a_*))$ follows.)

Next, we define $\hat{\Upsilon}(\boldsymbol{\sigma}_N) = \Phi_{b_*,\tau_N}(\Upsilon^{(N)}(\boldsymbol{\sigma}_N))$. Then $\hat{\Upsilon}(\boldsymbol{\sigma}_N) \subseteq \mathcal{R}_{\mathcal{D};[a_*,b_*]}(\xi_*(a_*))$ if $\boldsymbol{\sigma}_N \in \Sigma^{(N),+}$, because $\Upsilon^{(N)}(\boldsymbol{\sigma}_N) \subseteq \mathcal{R}_{\mathcal{D};[a_*,\tau_N]}(\xi_*(a_*))$. Hence $\hat{\Upsilon}$ maps $\Sigma^{(N),+}$ into the reachable set $\mathcal{R}_{\mathcal{D};[a_*,b_*]}(\xi_*(a_*))$.

For each measurable selection L of the map $t \mapsto \partial f_{\eta_{*},t}(\xi_{*}(t))$, define linear maps $\mathcal{Q}^{i;L}$ from \mathbb{R}^{i} to $T_{\xi_{*}(\tau_{i})}M$ by letting $\mathcal{Q}^{1;L}(\sigma_{1}) = L_{u_{1},\tau_{1}}(0,\sigma_{1})$, and then, recursively,

$$\mathcal{Q}^{i;L}(\boldsymbol{\sigma}_i) = L_{u_i,\tau_i} \left(P_{\tau_i,\tau_{i-1}}^{\nabla_L} (\mathcal{Q}^{i-1;L}(\boldsymbol{\sigma}_{i-1})), \sigma_i \right).$$

Then define $\hat{\mathcal{Q}}^{L}(\boldsymbol{\sigma}_{N}) = P_{b_{*},\tau_{N}}^{\nabla_{L}} (\mathcal{Q}^{N;L}(\boldsymbol{\sigma}_{N}))$, so $\hat{\mathcal{Q}}^{L}$ is a linear map from \mathbb{R}^{N} to $T_{\xi_{*}(b_{*})}M$. Finally, we let $\boldsymbol{\mathcal{Q}}$ denote the set of all maps $\hat{\mathcal{Q}}^{L}$, for all measurable selections L of the map $t \mapsto \partial f_{\eta_{*},t}(\xi_{*}(t))$. Then $\boldsymbol{\mathcal{Q}}$ is a compact subset of $Lin(\mathbb{R}^{N}, T_{\xi_{*}(b_{*})}M)$. A simple calculation then shows that $\boldsymbol{\mathcal{Q}}$ is a Warga derivate container of $\hat{\Upsilon}$ at $(0, \xi_{*}(b_{*})) \in \mathbb{R}^{N} \times M$.

Since $\hat{\Upsilon}$ maps the nonnegative orthant \mathbb{R}^N_+ into $\mathcal{R}_{\mathcal{D};[a_*,b_*]}(\xi_*(a_*))$, it follows from Examples 3.10 and 3.11 that the set $\mathbf{Q} \cdot \mathbb{R}^N_+ = \{Q \cdot \mathbb{R}^N_+ : Q \in \tilde{\mathbf{Q}}\}$ is a WDC approximating multicone of $\mathcal{R}_{\mathcal{D};[a_*,b_*]}(\xi_*(a_*))$ at $\xi_*(b_*)$, and then $\Lambda \cdot \mathbf{Q} \cdot \mathbb{R}^N$ is a WDC approximating multicone of $F(\mathcal{R}_{\mathcal{D};[a_*,b_*]}(\xi_*(a_*)))$ at $F(\xi_*(b_*))$.

Now assume that $(\mathbf{H}^{Lip,in})$ does not hold. Then Theorem 3.15 tells us that the multicones $\Lambda \cdot \mathbf{Q} \cdot \mathbb{R}^N$ and \mathcal{C} are not strongly transversal.

Let μ be an arbitrary nonzero member of $T^*_{F(\xi_*(b_*))}N$. Then Lemma 3.5 tells us that there exist a nonnegative number π_0 , covectors $\nu, \hat{\nu} \in T^*_{F(\xi_*(b_*))}N$, linear maps $\lambda \in \Lambda$, $Q \in \mathbf{Q}$, and a cone $C \in \mathcal{C}$, such that $\pi_0 \mu = \nu + \hat{\nu}, \hat{\nu} \in C^{\perp}$, $\nu \in (\lambda \cdot Q \cdot \mathbb{R}^N_+)^{\perp}$, and $(\pi_0, \nu, \hat{\nu}) \neq (0, 0, 0)$. Then $(\pi_0, \nu) \neq (0, 0)$ (because if $(\pi_0, \nu) = (0, 0)$ then the identity $\pi_0 \mu = \nu + \hat{\nu}$ would imply that $\hat{\nu} = 0$ as well, so $(\pi_0, \nu, \hat{\nu}) = (0, 0, 0)$). Since $\hat{\nu} = \pi_0 \mu - \nu$, we have shown that π_0 and ν satisfy Conditions II and III of our conclusion.

We now let $\pi^{\#} = \nu \circ \lambda$. Then $\pi^{\#} \in (Q \cdot \mathbb{R}^N_+)^{\perp}$. The map Q is of the form \hat{Q}^L , for a measurable selection L of $t \mapsto \partial f_{\eta_*,t}(\xi_*(t))$. Define $\pi(t) = \pi^{\#} \circ P_{b_*,t}^{\nabla_L}$. Then, if $\boldsymbol{\sigma}_N \in \mathbb{R}^N_+$,

$$\langle \pi^{\#}, Q(\boldsymbol{\sigma}_{N}) \rangle = \langle \pi^{\#}, \hat{\mathcal{Q}}^{L}(\boldsymbol{\sigma}_{N}) \rangle = \sum_{i=1}^{N} \sigma_{i} \langle \pi^{\#}, P_{b_{*},\tau_{i}}^{\nabla_{L}}(\bar{v}_{u_{i},\tau_{i}} - \bar{v}_{*,\tau_{i}}) \rangle$$
$$= \sum_{i=1}^{N} \sigma_{i} \langle \pi(\tau_{i}), \bar{v}_{u_{i},\tau_{i}} - \bar{v}_{*,\tau_{i}} \rangle ,$$

where we have written $\bar{v}_{u_i,\tau_i} = f(\xi_*(\tau_i), u_i, \tau_i), \ \bar{v}_{*,\tau_i} = f(\xi_*(\tau_i), \eta_*(\tau_i), \tau_i).$ Since $\langle \pi^{\#}, Q(\boldsymbol{\sigma}_N) \rangle \leq 0$ for all $\boldsymbol{\sigma}_N \in \mathbb{R}^N_+$, we conclude that the inequalities $\langle \pi(\tau_i), \bar{v}_{u_i,\tau_i} - \bar{v}_{*,\tau_i} \rangle \leq 0$ hold for $i = 1, \ldots, N$. We have therefore shown that $H^f(\xi_*(\tau_i), \pi(\tau_i), u_i, \tau_i) \leq H^f(\xi_*(\tau_i), \pi(\tau_i), \eta_*(\tau_i), \tau_i)$ for $i = 1, \ldots, N$.

We have thus obtained π_0 , ν , λ , L that satisfy all our desired conditions, except for the fact that the inequalities of the Hamiltonian maximization condition have only been established for special sets \mathcal{F} of pairs (u, τ) , namely, sets \mathcal{F} that satisfy three additional restrictions: (r1) \mathcal{F} is finite, (r2) $\mathcal{F} \subseteq \hat{\mathcal{L}}$, and (r3) no two different members of \mathcal{F} have the same time τ .

What we actually need is to have the inequalities for all pairs $(u, \tau) \in \mathcal{L}$. This more general set of inequalities can be obtained from the inequalities for our special sets \mathcal{F} by a well known compactness argument. Fix a norm in the space $T^*_{F(\xi_*(b_*))}N$, and (still keeping μ fixed) consider the set \mathcal{K} of all 4-tuples (π_0, ν, λ, L) such that $\pi_0 \in \mathbb{R}, \pi_0 \geq 0, \nu \in T^*_{F(\xi_*(b_*))}N, \pi_0 + \|\nu\| = 1,$ $\pi_0\mu - \nu \in \mathcal{C}^{\perp}$, and L is a measurable selection of $[a_*, b_*] \ni t \mapsto \partial f_{\eta_*, t}(\xi_*(t))$. Then \mathcal{K} is compact (using the weak topology for the Ls).

For every subset \mathcal{G} of \mathcal{L} , let $\mathcal{K}^{\mathcal{G}}$ be the set of all $(\pi_0, \nu, \lambda, L) \in \mathcal{K}$ such that the inequalities $H^f(\xi_*(\tau), \pi(\tau), \eta_*(\tau), \tau) \geq H^f(\xi_*(\tau), \pi(\tau), u, \tau)$ hold for all $(u, \tau) \in \mathcal{G}$, where $\pi(t) = \pi^{\#} \circ P_{b_*, t}^{\nabla_L}$ and $\pi^{\#} = \nu \circ \lambda$. Then each $\mathcal{K}^{\mathcal{G}}$ is a compact subset of \mathcal{K} , and our proof will be complete if we show that $\mathcal{K}^{\mathcal{L}} \neq \emptyset$.

Let \mathcal{F} be any finite subset of \mathcal{L} . Let $\mathcal{F} = \{(u_1, \tau_1), \dots, (u_N, \tau_N)\}$. Then we can construct sequences $\{\tau_i^j\}_{j\in\mathbb{N}}$ of members of $]a_*, b_*[$ in such a way that $(u_i, \tau_i^j) \in \hat{\mathcal{L}}$, $\lim_{j\to\infty} \tau_i^j = \tau_i$, $\lim_{j\to\infty} f(\xi_*(\tau_i^j), u_i, \tau_i^j) = f(\xi_*(\tau_i), u_i, \tau_i)$, $\lim_{j\to\infty} f(\xi_*(\tau_i^j), \eta_*(\tau_i^j), \tau_i^j) = f(\xi_*(\tau_i), \eta_*(\tau_i), \tau_i)$, and $\tau_i^j \neq \tau_{i'}^j$ whenever $i \neq i'$. Let $\mathcal{F}^j = \{(u_1, \tau_1^j), \dots, (u_N, \tau_N^j)\}$. Then the \mathcal{F}^j satisfy all three restrictions (r1), (r2), (r3). Therefore $\mathcal{K}^{\mathcal{F}^j} \neq \emptyset$. Pick $(\pi_0^j, \nu^j, \lambda^j, L^j) \in \mathcal{K}^{\mathcal{F}^j}$. By passing to a subsequence, we may assume that $\{(\pi_0^j, \nu^j, \lambda^j, L^j)\}_{j \in \mathbb{N}}$ converges to a limit $(\pi_0, \nu, \lambda, L) \in \mathcal{K}$. Then $(\pi_0, \nu, \lambda, L) \in \mathcal{K}^{\mathcal{F}}$. So $\mathcal{K}^{\mathcal{F}} \neq \emptyset$. As \mathcal{F} varies over all finite subsets of \mathcal{L} , the sets $\mathcal{K}^{\mathcal{F}}$ are compact and nonempty. Furthermore, any finite intersection $A = \mathcal{K}^{\mathcal{F}_1} \cap \mathcal{K}^{\mathcal{F}_2} \cap \cdots \cap \mathcal{K}^{\mathcal{F}_N}$ is nonempty, because $A = \mathcal{K}^{\mathcal{F}_1 \cup \mathcal{F}_2 \cup \cdots \cup \mathcal{F}_N}$. Therefore the intersection of all the sets $\mathcal{K}^{\mathcal{F}}$ is nonempty. So $\mathcal{K}^{\mathcal{L}}$ is nonempty, completing our proof. \square

5.2The maximum principle for local controllability

Given a control system with data $\mathcal{D} = (M, f, U, \mathcal{U})$, and a TCP (ξ_*, η_*) of \mathcal{D} with domain $[a_*, b_*]$, we say that \mathcal{D}) is **locally controllable along** ξ_* if the reachable set $\mathcal{R}_{\mathcal{D};[a_*,b_*]}(\xi_*(a_*))$ is a neighborhood of $\xi(b_*)$.

In the local controllability problem, the same type of data as in the separation problem are specified, except that N, F, S, C and Λ are not needed. So we are given a data 9-tuple $\mathcal{D}^{lc} = (M, f, U, \mathcal{U}, x_*, a_*, b_*, \xi_*, \eta_*)$, consisting of a system data 4-tuple $\mathcal{D} = (M, f, U, \mathcal{U})$, an initial state x_* , endpoints a_*, b_* of the reference interval $[a_*, b_*]$, and a reference TCP (ξ_*, η_*) .

The following is then our version of the Lipschitz maximum principle for local controllability.

Theorem 5.4 Assume that the data \mathcal{D}^{lc} are such that Hypotheses (H1), (H2), (H3), (H4), (H6.a), (H7) and (H9) hold, and $x_* \in M$. Let \mathcal{L} be as in the statement of Theorem 5.1. Then, if the system with data $\mathcal{D} = (M, f, U, \mathcal{U})$ is not locally controllable along ξ_* , it follows that there exist

- a nonzero covector π[#] ∈ T^{*}_{ξ*(b*)}M,
 a measurable selection [a*, b*] ∋ t → L(t) ∈ ∂f_{η*,t}(ξ*(t)) of the set-valued map [a*, b*] ∋ t → ∂f_{η*,t}(ξ*(t)) ⊆ J¹_{ξ*(t),ξ*(t)}Γ(TM),

having the property that, if $\nabla_L \in Cov(\xi_*)$ is the covariant differentiation corresponding to L, and we define $\pi(t) = \pi^{\#} \circ \mathcal{P}_{b_*,t}^{\nabla_L}$ for $t \in [a_*, b_*]$, then the following Hamiltonian maximization condition is satisfied:

(HM)
$$H^f(\xi_*(\tau), \pi(\tau), \eta_*(\tau), \tau) \ge H^f(\xi_*(\tau), \pi(\tau), u, \tau)$$
 whenever $(u, \tau) \in \mathcal{L}$.

Proof. Fix a coordinate chart **x** near $\xi_*(b_*)$, and identify all points $x \in \text{dom } \mathbf{x}$ with their coordinate representations $x^{\mathbf{x}}$.

Since our system is not locally controllable along ξ_* , we may pick a sequence $\sigma = \{x_j\}_{j \in \mathbb{N}}$ of points of dom **x** that do not belong to $\mathcal{R}_{\mathcal{D};[a_*,b_*]}(\xi_*(a_*))$ and are such that $\lim_{j\to\infty} x_j = \xi_*(b_*)$. Then in particular $x_j \neq \xi_*(b_*)$ for all j. After passing to a subsequence of σ , if necessary, we may assume that the limit

 $v = \lim_{j \to \infty} \frac{x_j - \xi_*(b_*)}{\|x_j - \xi_*(b_*)\|}$ exists. Clearly, $v \neq 0$. After passing to a subsequence again, we may also assume that the x_j are all different, that $\|x_j - x_{j+1}\| \leq 2^{1-j}$ for all j, and $\lim_{j \to \infty} \frac{x_j - x_{j+1}}{\|x_j - x_{j+1}\|} = v$. Define a Lipschitz curve $\zeta : [0, 1] \mapsto M$ by letting $\zeta(0) = \xi_*(b_*)$ and

$$\zeta(t) = 2^{j} \left((2^{1-j} - t) x_{j+1} + (t - 2^{-j}) x_{j} \right) \quad \text{for } 2^{-j} \le t \le 2^{1-j}, \ j \in \mathbb{N},$$

so that $\zeta(2^{1-j}) = x_j$, and the map ζ is linear on each interval $[2^{-j}, 2^{1-j}]$.

Let $S = \{\zeta(t) : t \in [0, 1]\}$. Then S is a compact subset of M, and it is easy to see that, if $C = \{rv : r \ge 0\}$, then $\{C\}$ is a WDC approximating cone to S at $\xi_*(b_*)$. (Actually, C is the Clarke tangent cone of S at $\xi_*(b_*)$.)

It is clear that the Lipschitz arc intersection property cannot hold with this choice of the set S because, if there existed a nonconstant Lipschitz arc $\gamma : [0,1] \mapsto \mathcal{R}_{\mathcal{D};[a_*,b_*]}(\xi_*(a_*)) \cap S$ such that $\gamma(0) = \xi_*(b_*)$, then there would have to exist arbitrarily small t_k such that $\gamma(t_k)$ belongs to the set $\{x_j : j \in \mathbb{N}\}$, contradicting the fact that the x_j do not belong to $\mathcal{R}_{\mathcal{D};[a_*,b_*]}(\xi_*(a_*))$.

It then follows from Theorem 5.4—taking N = M, letting the map F be the identity map, and choosing $\Lambda = \{\mathbb{I}_{T_{\xi^*}(b_*)}M\}$ and $\mathcal{C} = \{C\}$ —that for every $\mu \in T^*_{\xi^*(b_*)}M \setminus \{0\}$ there exists a 4-tuple (π_0, ν, λ, L) that satisfies Property (*) of the statement of that theorem. We apply this to a μ that does not belong to \mathcal{C}^{\perp} . (For example, we could take any μ such that $\langle \mu, v \rangle = 1$.) Let $\pi^{\#} = \nu \circ \lambda$. Then $\pi^{\#} = \nu$, because λ is the identity map of $T_{\xi^*(b_*)}M$. The covector $\pi^{\#}$ and the measurable selection L clearly satisfy our desired conditions, except only for the fact that $\pi^{\#}$ might vanish. To exclude this possibility, we observe that if $\pi^{\#} = 0$ then $\nu = 0$, so $\pi_0 > 0$ by the nontriviality condition of Theorem 5.4. But then the fact that $\pi_0\mu - \nu \in \mathcal{C}^{\perp}$ simply says that $\pi_0\mu \in \mathcal{C}^{\perp}$, and then $\mu \in \mathcal{C}^{\perp}$, since $\pi_0 > 0$. So we have reached a contradiction, showing that $\pi^{\#} \neq 0$, and concluding our proof.

5.3 The maximum principle for optimal control

We now consider a fixed time-interval Lagrangian optimal control problem

minimize
$$\varphi(\xi(b)) + \int_a^b f_0(\xi(t), \eta(t), t) dt$$

subject to
$$\begin{cases} \xi(\cdot) \in W^{1,1}([a, b], \mathbb{R}^n) \text{ and } \dot{\xi}(t) = f(\xi(t), \eta(t), t) \text{ a.e.}, \\ \xi(a) = x_* \text{ and } F(\xi(b)) \in S, \\ \eta(t) \in U \text{ for all } t \in [a, b], \text{ and } \eta(\cdot) \in \mathcal{U}. \end{cases}$$

We assume, as before, that we are given a reference trajectory-control pair (ξ_*, η_*) , whose domain is the reference interval $[a_*, b_*]$. And, finally, we assume that we are given a multicone C. So we are specifying a data 14-tuple \mathcal{D}^{sep} as in (19), and we will assume that all the conditions (H1) to (H9) hold.

In addition to \mathcal{D}^{sep} , we now need to specify a cost functional. For this purpose, we give f_0 and φ such that

(H10) f_0 is a ppd function from $M \times U \times \mathbb{R}$ to \mathbb{R} , and φ is a ppd function from M to \mathbb{R} which is defined and Lipschitz on some neighborhood of $\xi_*(b_*)$.

Furthermore, we need to be able to differentiate φ at $\xi_*(b_*)$. We could do this by specifying a Warga derivate container Θ of φ at $\xi_*(b_*)$, but we will allow the slightly more general possibility that, instead of separate derivate containers Λ , Θ of F and φ , the map $x \mapsto (\varphi(x), F(x))$ may have a joint derivate container. For this purpose, we will substitute for Hypothesis (H8.b) the following condition

(H8.b') $\tilde{\Lambda}$ is a nonempty compact subset of $Lin(T_{\xi_*(b_*)}M, \mathbb{R} \times T_{F(\xi_*(b_*))}N)$, and is a Warga derivate container at the point $\xi_*(b_*)$ of the ppd map $M \ni x \hookrightarrow (\varphi(x), F(x)) \in \mathbb{R} \times N$.

Then, if $\mathcal{D}^{opt} = (M, f, U, \mathcal{U}, x_*, N, F, S, a_*, b_*, \xi_*, \eta_*, \mathcal{C}, \tilde{\Lambda}, f_0, \varphi)$ is our data 16-tuple,

- A TCP (ξ, η) with domain $[a_*, b_*]$ is *endpoint-cost-admissible* if it satisfies the following five conditions: (i) (ξ, η) is admissible, (ii) $\xi(a_*) = x_*$, (iii) $\xi(b_*) \in \text{dom } F \cap \text{dom } \varphi$, (iv) $F(\xi(b_*)) \in S$, and, finally (v) the realvalued function $[a_*, b_*] \ni t \mapsto f_0(\xi(t), \eta(t), t)$ is a. e. defined, measurable, and such that $\int_{a_*}^{b_*} \min \left(0, f_0(\xi(t), \eta(t), t)\right) dt > -\infty$.
- We write $TCP_{adm,ec}(\mathcal{D}^{opt})$ to denote the set of all TCPs of \mathcal{D}^{opt} that are endpoint-cost-admissible.

It follows that if (ξ, η) belongs to $TCP_{adm,ec}(\mathcal{D}^{opt})$ then the number

$$J(\xi, \eta) = \varphi(\xi(b_*)) + \int_{a_*}^{b_*} f_0(\xi(t), \eta(t), t) \, dt$$

—called the **cost** of (ξ, η) —is well defined and belongs to $]-\infty, +\infty]$.

For the data \mathcal{D}^{opt} , we define a ppd map $\mathbf{f}: M \times U \times \mathbb{R} \hookrightarrow \mathbb{R} \times M$, called the *augmented dynamics*, by letting

$$dom(\mathbf{f}) = dom(f_0) \cap dom(f),$$

$$\mathbf{f}(z) = (f_0(z), f(z)) \quad \text{for } z = (x, u, t) \in dom(\mathbf{f}).$$

If η is a *U*-control, we write $f_{0,\eta}(x,t) = f_0(x,\eta(t),t)$, $\mathbf{f}_\eta(x,t) = \mathbf{f}(x,\eta(t),t)$, so $f_{0,\eta}$ is a ppd function from $M \times \mathbb{R}$ to \mathbb{R} , and \mathbf{f}_η is a ppd map that sends each $(x,t) \in \text{dom } \mathbf{f}_\eta \subseteq M \times \mathbb{R}$ to a point $\mathbf{f}_\eta(x,t) \in \mathbb{R} \times T_x M$. If in addition ξ is an arc in M, we write

$$f_{0,\eta,\xi}(t) = f_0(\xi(t), \eta(t), t) , \mathbf{f}_{\eta,\xi}(t) = \mathbf{f}(\xi(t), \eta(t), t) .$$

It will also be convenient, using the obvious identification of the product $\mathbb{R} \times T_x M$ with the tangent space $T_{(x_0,x)}(\mathbb{R} \times M)$, to regard the \mathbf{f}_η as ppd maps $\mathbb{R} \times M \times \mathbb{R} \ni (x_0, x, t) \mapsto (f_{0,\eta}(x, t), f_\eta(x, t)) \in T_{(x_0,x)}(\mathbb{R} \times M)$, that is, as ppd time-varying vector fields on $\mathbb{R} \times M$ that happen not to depend on x_0 .

The precise technical hypothesis on f_0 is

(H11) for every control $\eta \in \mathcal{U}_{[a_*,b_*]}^{c,*}$, the time-varying function $f_{0,\eta}$ is integrably Lipschitz near ξ_* (cf. Remark 5.5 below).

Remark 5.5 The definition of the "integrably Lipschitz" property for ppd time-varying functions is identical to that for ppd vector fields, with the obvious trivial modifications. Alternatively, we can regard the augmented dynamics as giving rise to ppd time-varying vector fields \mathbf{f}_{η} on $\mathbb{R} \times M$, as explained above, and then the "integrably Lipschitz" condition for f_0 , together with Assumption (H7), just amounts to requiring that the \mathbf{f}_{η} , for $\eta \in \mathcal{U}_{[a_*,b_*]}^{c,*}$, be integrably Lipschitz near the arc $[a_*,b_*] \ni t \mapsto (0,\xi_*) \in \mathbb{R} \times M$.

We write $\xi_{0,*}(t) = \int_{a_*}^t f_0(\xi_*(s), \eta_*(s), s) \, ds$, so the function $\xi_{0,*}$ is the **running Lagrangian cost** along (ξ_*, η_*) , initialized so that $\xi_{0,*}(a_*) = 0$. We then let $\Xi_*(t) = (\xi_{0,*}(t), \xi_*(t)), \text{ so } \Xi_* : [a_*, b_*] \mapsto \mathbb{R} \times M \text{ is the } cost-augmented$ reference trajectory. Clearly, Ξ_* is an integral curve of \mathbf{f}_{η_*} , if we regard \mathbf{f}_{η_*} as a ppd tvvf on $\mathbb{R} \times M$, as explained above, and our assumptions imply that \mathbf{f}_{η_*} is integrably Lipschitz near Ξ_* . This makes it possible to talk about the Clarke generalized Jacobian $\partial \mathbf{f}_{\eta_{*},t}(\Xi_{*}(t))$, for which we will also use the notation $\partial \mathbf{f}_{\eta_{*},t}(\xi_{*}(t))$, since $\mathbf{f}_{\eta_{*},t}$ does not depend on the first component. Then $\partial \mathbf{f}_{\eta_{*},t}(\xi_{*}(t))$ is a compact convex subset of the space $J^{1}_{\Xi_{*}(t),\Xi_{*}(t)}\Gamma(T(\mathbb{R}\times M)).$ We recall that $J^1_{\Xi_*(t), \dot{\Xi}_*(t)} \Gamma(\mathbb{R} \times TM)$ is the set of all 1-jets at $\Xi_*(t)$ of sections ζ of the bundle $T(\mathbb{R} \times M)$ such that $\zeta(\Xi_*(t)) = \Xi_*(t)$. However, the value of $\mathbf{f}_{\eta_*,t}$ at a point $(r, x) \in \mathbb{R} \times M$ does not depend on r. So the 1-jet $j^1 \mathbf{f}_{\eta_{*}, t}(r, x)$ at a point $(r, x) \in diff(\mathbf{f}_{\eta_*, t})$ is a 1-jet at x of sections ζ of $\mathbb{R} \times_M TM$ such that $\zeta(\xi_*(t)) = \Xi_*(t)$, where $\mathbb{R} \times_M TM$ is the bundle over M whose fiber at each point $x \in M$ is the product $\mathbb{R} \times T_x M$. Hence we can regard $\partial \mathbf{f}_{\eta_*,t}(\xi_*(t))$ as a compact convex subset of $J^1_{\xi_*(t), \dot{\Xi}_*(t)} \Gamma(\mathbb{R} \times_M M)$. Furthermore, $J^{1}_{\xi_{*}(t), \dot{\Xi}_{*}(t)} \Gamma(\mathbb{R} \times_{M} TM) = J^{1}_{\xi_{*}(t), \dot{\xi}_{0,*}(t)} \Gamma(\mathbb{R}) \times J^{1}_{\xi_{*}(t), \dot{\xi}_{*}(t)} \Gamma(TM), \text{ so every member}$ of $J^1_{\xi_*(t),\dot{\Xi}_*(t)}\Gamma(\mathbb{R}\times TM)$ can be regarded as a pair (L_0,L) , where L_0 is a 1-jet at $\xi_*(t)$ of real functions $\psi \in C^1(M,\mathbb{R})$ such that $\psi(\xi_*(t)) = \dot{\xi}_{0,*}(t)$, and L is a 1-jet at $\xi_*(t)$ of vector fields Ψ on M such that $\Psi(\xi_*(t)) = \xi_*(t)$. Finally, $J^1_{\xi_*(t),\dot{\xi}_{0,*}(t)}\Gamma(\mathbb{R})$ can obviously be identified with the cotangent space $T^*_{\xi_*(t)}M$. Hence the set $\partial \mathbf{f}_{\eta_*,t}(\xi_*(t))$ is a compact convex subset of the product $T^*_{\xi_*(t)}M \times J^1_{\xi_*(t),\dot{\xi}_*(t)}\Gamma(TM)$.

It is then clear that $\partial \mathbf{f}_{\eta_{*},t}(\xi_{*}(t)) \subseteq \partial f_{0,\eta_{*},t}(\xi_{*}(t)) \times \partial f_{\eta_{*},t}(\xi_{*}(t))$, from which it follows immediately that every measurable selection of the set-valued map $[a_{*}, b_{*}] \ni t \mapsto \partial \mathbf{f}_{\eta_{*},t}(\xi_{*}(t))$ can be regarded as a pair (ω, L) , where

- (i) ω is an integrable field of covectors along ξ_* , which is a measurable selection of $[a_*, b_*] \ni t \mapsto \partial f_{0,\eta_*,t}(\xi_*(t))$,
- (ii) L is an integrable function $[a_*, b_*] \ni t \mapsto L(t) \in J^1_{\xi_*(t), \dot{\xi}_*(t)} \Gamma(TM)$, which is a measurable selection of the map $[a_*, b_*] \ni t \mapsto \partial f_{\eta_*, t}(\xi_*(t))$,

Then L gives rise to a covariant differentiation ∇_L along ξ_* , and we can write the "inhomogeneous adjoint equation" $\nabla_L \pi = \pi_0 \omega$, for any $\pi_0 \in \mathbb{R}$. The corresponding Cauchy problem, with terminal condition $\pi(b_*) = \overline{\pi}$, clearly has a unique solution π for any given $\omega, L, \pi_0, \overline{\pi}$. A field π of covectors (or a pair (π_0, π)) arising in this way is called an **adjoint covector**, or **adjoint vector**.

The hypothesis on the reference TCP (ξ_*, η_*) is that it is a local costminimizer in $TCP_{adm,ec}(\mathcal{D}^{opt})$. In other words,

(H^{opt}) $(\xi_*, \eta_*) \in TCP_{adm,ec}(\mathcal{D}^{opt})$, and there exists a neighborhood V of $F(\xi_*(b_*))$ in N having the property that $J(\xi_*, \eta_*) \leq J(\xi, \eta)$ for all pairs $(\xi, \eta) \in TCP_{adm,ec}(\mathcal{D}^{opt})$ such that $F(\xi(b_*)) \in V$.

It will also be convenient to consider the following strong form of the negation of (H^{opt}) , that we will call the *Lispchitz arc nonoptimality property*.

 $(\mathbf{H}^{Lip,nonopt}) \text{ There exists a map } [0,1] \ni s \mapsto (\xi_s,\eta_s) \in TCP_{adm,ec}(\mathcal{D}^{opt}) \text{ such that}$ $(i) \text{ the map } [0,1] \ni s \mapsto (J(\xi_s,\eta_s),F(\xi_s(b_*))) \in \mathbb{R} \times N \text{ is Lipschitz,}$ $(ii) \ (\xi_0,\eta_0) = (\xi_*,\eta_*),$ $(iii) \ J(\xi_s,\eta_s) < J(\xi_*,\eta_*) \text{ for all } s \in]0,1].$

We define the **Hamiltonian of f** to be the parametrized family of functions $H^{\mathbf{f}}_{\alpha}: T^*M \times U \times \mathbb{R} \hookrightarrow \mathbb{R}$, (depending on the real parameter α), given by the formula $H^{\mathbf{f}}_{\alpha}(x, p, u, t) = p \cdot f(x, u, t) - \alpha f_0(x, u, t)$. Also, we recall that $\tilde{\Lambda}$ is a subset of $Lin(T_{\xi_*(b_*)}M, \mathbb{R} \times T_{F(\xi_*(b_*))}N)$, which can be naturally identified with the product $\mathcal{P} = T^*_{\xi_*(b_*)}M \times Lin(T_{\xi_*(b_*)}M, T_{F(\xi_*(b_*))}N) \longrightarrow \tilde{\Lambda}$ is in fact a subset of \mathcal{P} , that is, a set of pairs $(\theta, \lambda), \theta \in T^*_{\xi_*(b_*)}M, \lambda \in Lin(T_{\xi_*(b_*)}M, T_{F(\xi_*(b_*))}N)$.

The following is then our version of the maximum principle for optimal control.

Theorem 5.6 Assume that the data 16-tuple \mathcal{D}^{opt} satisfies Hypotheses (H1) to (H11) (with (H8.b) replaced by (H8.b')). Let \mathcal{L} denote the set of all pairs (u, τ) such that $u \in U, \tau \in]a_*, b_*[$, and τ is a Lebesgue time along ξ_* of both augmented time-varying vector fields \mathbf{f}_u and \mathbf{f}_{η_*} . Then, if (H^{Lip,nonopt}) is not true, it follows that

(*) there exist

- 1. a covector $\pi^{\#} \in T^*_{\xi_*(b_*)}M$,
- 2. a pair $(\nu, (\theta, \lambda)) \in T^*_{\xi_*(b_*)}M \times \tilde{\Lambda}$,
- 3. a measurable selection $[a_*, b_*] \ni t \mapsto (\omega(t), L(t)) \in \partial \mathbf{f}_{\eta_{*,t}}(\xi_*(t))$ of the set-valued map $[a_*, b_*] \ni t \mapsto \partial \mathbf{f}_{\eta_{*,t}}(\xi_*(t)) \subseteq T^*_{\xi_*(t)}M \times J^1_{\xi(t),\dot{\xi}(t)}\Gamma(TM),$
- 4. a nonnegative real number π_0 ,

such that, if $\nabla_L \in Cov(\xi_*)$ is the covariant differentiation corresponding to L, and we let π be the unique absolutely continuous solution of the "adjoint Cauchy problem" $\nabla_L \pi(t) = \pi_0 \omega(t), \ \pi(b_*) = \pi^{\#}$, then the following three conditions are satisfied:

- I. Hamiltonian maximization: $H^f_{\pi_0}(\xi_*(\tau), \eta_*(\tau), \tau) \ge H^f_{\pi_0}(\xi_*(\tau), u, \tau)$ whenever $(u, \tau) \in \mathcal{L}$,
- II. **Transversality:** $-\nu \in \mathcal{C}^{\perp}$, and $\pi^{\#} = \nu \cdot \lambda \pi_0 \theta$,
- III. Nontriviality: $\nu \neq 0 \text{ or } \pi_0 > 0.$

In particular, if (H^{opt}) holds then (*) is true as well.

Remark 5.7 In most situations, N is just M, and F is the identity map. In that case, one can just take take $\tilde{\Lambda} = \Theta \times \{\mathbb{I}_{T_{\xi_*(b_*)}M}\}$, where Θ is a Warga derivate container of φ at $\xi_*(b_*)$. Then the transversality condition takes the more familiar form $-\pi^{\#} \in \pi_0 \Theta + \mathcal{C}^{\perp}$.

Remark 5.8 The conclusion of Theorem 5.6 implies in particular the "weak Hamiltonian maximization condition": for every control value $u \in U$ there exists a Lebesgue-null subset $\mathcal{N}(u)$ of the interval $[a_*, b_*]$ with the property that $H^{\mathbf{f}}_{\pi_0}(\xi_*(\tau), \pi(\tau), \eta_*(\tau), \tau) \geq H^{\mathbf{f}}_{\pi_0}(\xi_*(\tau), \pi(\tau), u, \tau)$ whenever $\tau \notin \mathcal{N}(u)$.

Under extra technical hypotheses, one can deduce the "strong Hamiltonian maximization condition": if (i) U is a separable metric space and (ii) the function $U \ni u \mapsto \mathbf{f}(\xi_*(t), u, t)$ is continuous for a. e. $t \in [a_*, b_*]$, then the equality $H^{\mathbf{f}}_{\pi_0}(\xi_*(\tau), \pi(\tau), \eta_*(\tau), \tau) = \max\{H^{\mathbf{f}}_{\pi_0}(\xi_*(\tau), \pi(\tau), u, \tau) : u \in U\}$ holds for all τ in the complement of a null subset \mathcal{N} of $[a_*, b_*]$. (The proof is exactly like that of Proposition 5.3.)

Proof of Theorem 5.6. We assume that Condition $(\mathbf{H}^{Lip,nonopt})$ is not true, and apply Theorem 5.1 to a separation problem whose data 14-tuple $\hat{\mathcal{D}}^{sep}$, given by $\hat{\mathcal{D}}^{sep} = (\hat{M}, \hat{f}, \hat{U}, \hat{\mathcal{U}}, \hat{x}_*, \hat{N}, \hat{F}, \hat{S}, \hat{a}_*, \hat{b}_*, \hat{\xi}_*, \hat{\eta}_*, \hat{\mathcal{C}}, \hat{\Lambda})$, is constructed in a suitable way from our optimal control data \mathcal{D}^{opt} . We take $\hat{U} = U$, $\hat{\mathcal{U}} = \mathcal{U}$, and let $\hat{M} = \mathbb{R} \times M$, so the state space of the new system is that of the old one with the addition of a new variable x_0 , the Lagrangian running cost. The new dynamics \hat{f} is \mathbf{f} , the augmented dynamics of the optimal control problem, so that that the right-hand side of the dynamical equation $\frac{d}{dt}(x_0, x) = \hat{f}(x_0, x, u, t)$ is given by $\hat{f}(x_0, x, u, t) = (f_0(x, u, t), f(x, u, t))$, and the dynamical equation is equivalent to the pair of conditions $\dot{x}_0 = f_0(x, u, t)$, $\dot{x} = f(x, u, t)$. We take $\hat{x}_* = (0, x_*)$, so the initial state for our augmented system is the same as for the original one, and the initial value of the running cost is 0. We take $\hat{a}_* = a_*$, $\hat{b}_* = b_*$, $\hat{\eta}_* = \eta_*$.

We let $\hat{\xi}_*$ be the trajectory of the augmented system for the control $\hat{\eta}_*$ and the initial condition \hat{x}_* . (That is, $\hat{\xi}_*$ is the curve introduced earlier and labelled Ξ_* .) So $\hat{\xi}_*(t) = (\xi_{0,*}(t), \xi_*(t))$, where $\xi_{0,*}(t) = \int_{a_*}^t f_0(\xi_*(s), \eta_*(s), s) \, ds$. We then write $c_* = \xi_{0,*}(b_*)$, and define $\hat{c}_* = c_* + \varphi(\xi_*(b_*))$, so c_* are \hat{c}_* are, respectively, the Lagrangian cost and the total cost of the reference TCP.

We take $\hat{N} = \mathbb{R} \times N$. The map $\hat{F} : \hat{M} \hookrightarrow \hat{N}$ is then defined by letting $\hat{F}(x_0, x) = (x_0 + \varphi(x), F(x))$. For each $(\theta, \lambda) \in \tilde{\Lambda}$, we let $\hat{\lambda}_{\theta,\lambda}$ be the linear map from $\mathbb{R} \times T_{\xi_*(b_*)}M$ (identified with $T_{(c_*,\xi_*(b_*))}\hat{M}$) to $\mathbb{R} \times T_{F(\xi_*(b_*))}N$ (identified with $T_{(\hat{c}_*,\xi_*(b_*))}\hat{N}$) given by $\hat{\lambda}_{\lambda,\theta}(\Delta x_0, \Delta x) = (\Delta x_0 + \theta \cdot \Delta x, \lambda \cdot \Delta x)$. We then let $\hat{\Lambda} = \{\hat{\lambda}_{\theta,\lambda} : (\theta, \lambda) \in \tilde{\Lambda}\}$. It is then easy to verify that $\hat{\Lambda}$ is a Warga derivate container of \hat{F} at $(c_*,\xi_*(b_*))$ (i. e., at $\hat{\xi}_*(b_*)$).

To construct the set \hat{S} , we first fix a smooth function $\psi : N \mapsto \mathbb{R}$ such that $\psi(F(\xi_*(b_*))) = 0$ and $\psi(y) > 0$ for all $y \in N \setminus \{F(\xi_*(b_*))\}$. We then define $\hat{S} = \{(y_0, y) : y \in S \text{ and } y_0 \leq \hat{c}_* - \psi(y)\}$. It is then easy to see that

(#) The Lipschitz arc intersection property (H^{Lip,in}) is not satisfied by the new separation data.

Indeed, suppose that the condition was satisfied. Let \mathcal{R} be the reachable set for the new system from \hat{x}_* over $[a_*, b_*]$. Let $\hat{\gamma}$ be a Lipschitz arc, defined on [0, 1], having values in the set $\hat{F}(\hat{\mathcal{R}}) \cap \hat{S}$, and such that $\hat{\gamma}(0) = \hat{F}(\hat{\xi}_*(b_*))$ and $\hat{\gamma}(1) \neq \hat{F}(\hat{\xi}_*(b_*))$. Write $\hat{\gamma}(s) = (\gamma_0(s), \gamma(s))$, so $\gamma_0(s) \in \mathbb{R}$, $\gamma(s) \in S$, and $\gamma_0(s) \leq \hat{c}_* - \psi(\gamma(s))$. Then $\gamma_0(0) = \hat{c}_*$, $\gamma(0) = F(\xi_*(b_*))$, $\gamma_0(s) \leq \hat{c}_*$ for all s, and $\gamma_0(s) < \hat{c}_*$ whenever $\gamma(s) \neq \gamma(0)$. Let $A = \{s \in [0,1] : \gamma_0(s) < \hat{c}_*\}$. Then A is a relatively open subset of [0,1], and $0 \notin A$. On the other hand, $1 \in A$, because $\hat{\gamma}(1) \neq \hat{F}(\hat{\xi}_*(b_*)), \hat{\gamma}(1) \in \hat{S}$, and $\hat{F}(\hat{\xi}_*(b_*))$ is the only point $(y_0, y) \in \hat{S}$ such that $y_0 = \hat{c}_*$. Let $I = \{s \in]0, 1[: [s,1] \subseteq A\}$. Then I is an open interval of the form $]\alpha, 1[$, such that $\gamma_0(\alpha) = \hat{c}_*$ and $\gamma_0(s) < \hat{c}_*$ whenever $\alpha < s \leq 1$. It then follows that $\gamma(\alpha) = F(\xi_*(b_*))$. Let $\tilde{\gamma}(r) = \hat{\gamma}(\alpha + r(1 - \alpha))$ for $r \in [0, 1]$. Then $\tilde{\gamma}$ is a Lipschitz map with values in $\hat{F}(\hat{\mathcal{R}}) \cap \hat{S}$, such that $\tilde{\gamma}(0) = \hat{F}(\hat{\mathcal{R}})$ and $\tilde{\gamma}(s) \neq \hat{F}(\hat{\mathcal{R}})$ whenever $0 < s \leq 1$. Since $\tilde{\gamma}(s) \in \hat{F}(\hat{\mathcal{R}})$ for each s, we can pick points $(x_{0,s}, x_s) \in \hat{\mathcal{R}}$ for each s, we can pick TCPs $(\hat{\xi}_s, \eta_s) = ((\xi_{0,s}, \xi_s), \eta_s)$ such that $\xi_{0,s}(a_*) = 0$, $\xi_s(a_*) = x_*$, $\xi_{0,s}(b_*) = x_{0,s}$, and $\xi_s(b_*) = x_s$. Then $x_{0,s} = \int_{a_*}^{b_*} f_0(\xi_s(t), \eta_s(t)) dt$, $x_{0,s} + \varphi(x_s) = J(\xi_s, \eta_s)$, and $F(x_s) = F(\xi_s(b_*))$, so $\tilde{\gamma}(s) = (J(\xi_s, \eta_s), F(\xi_s(b_*)))$. It is then clear that we may pick (ξ_0, η_0) to be (ξ_*, η_*) and that $J(\xi_s, \eta_s) < J(\xi_*, \eta_*)$ whenever $0 < s \leq 1$. Hence $(\mathbf{H}^{Lip,nonopt})$ is true, contradicting our assumption, and completing the proof of (#).

Let $\hat{\mathcal{C}} =] - \infty, 0] \times \mathcal{C}$, i.e., $\hat{\mathcal{C}} = \{] - \infty, 0] \times C : C \in \mathcal{C} \}$. Then $\hat{\mathcal{C}}$ is a WDC approximating multicone to \hat{S} at $\hat{F}(\hat{\xi}_*(b_*))$. (To see this, let $\tilde{S} =] - \infty, \hat{c}_*] \times S$, and observe that $\hat{\mathcal{C}}$ is a WDC approximating multicone to \tilde{S} at $\hat{F}(\hat{\xi}_*(b_*))$. Let Φ be the map $\mathbb{R} \times N \ni (y_0, y) \mapsto (y_0 - \psi(y), y) \in \mathbb{R} \times N$. Then $\hat{S} = \Phi(\tilde{S})$, $\Phi(\hat{F}(\hat{\xi}_*(b_*))) = \hat{F}(\hat{\xi}_*(b_*))$, and the differential of Φ at $\hat{F}(\hat{\xi}_*(b_*))$ is the identity map. Therefore $\hat{\mathcal{C}}$ is a WDC approximating multicone to \hat{S} at $\hat{F}(\hat{\xi}_*(b_*))$.

We now apply Theorem 5.1 to the separation problem that we have just constructed, with data $\hat{\mathcal{D}}^{sep} = (\hat{M}, \hat{f}, \hat{U}, \hat{\mathcal{U}}, \hat{x}_*, \hat{N}, \hat{F}, \hat{S}, \hat{a}_*, \hat{b}_*, \hat{\xi}_*, \hat{\eta}_*, \hat{\mathcal{C}}, \hat{\Lambda})$, and the covector $\mu \in T^*_{\hat{F}(\hat{\xi}_*(b_*))} \hat{N}$ given by $\mu = (-1, 0)$ (so that $\mu(\Delta y_0, \Delta y) = -\Delta y_0$ whenever $(\Delta y_0, \Delta y)$ belongs to $T_{\hat{F}(\hat{\xi}_*(b_*))} \hat{N} \sim \mathbb{R} \times T_{F(\xi_*(b_*))} N$). We then get a 4-tuple $(\hat{\pi}_0, \hat{\nu}, \hat{\lambda}, \hat{L})$ such that

- (i) \hat{L} is a measurable selection of the map $[a_*, b_*] \ni t \mapsto \partial \mathbf{f}_{\eta_*, t}(\xi_*(t)),$
- (ii) $\hat{\pi}_0 \in \mathbb{R}$ and $\hat{\pi}_0 \ge 0$, (iii) $\hat{\nu} \in T^*_{\hat{F}(\hat{\xi}_*(b_*))} \hat{N}$, (iv) $\hat{\pi}_0 \mu - \hat{\nu} \in \hat{\mathcal{C}}^{\perp}$,
- (v) $(\hat{\pi}_0, \hat{\nu}) \neq (0, 0),$
- (vi) $\hat{\lambda} \in \hat{\Lambda}$,

such that, if we define $\hat{\pi}^{\#} = \hat{\nu} \circ \hat{\lambda}$ (so that $\hat{\pi}^{\#} \in T^*_{\hat{\xi}_*(b_*)} \hat{M}$), then the inequality

$$H^{\hat{f}}(\hat{\xi}_{*}(\tau), \hat{\pi}(\tau), \eta_{*}(\tau), \tau) \ge H^{\hat{f}}(\hat{\xi}_{*}(\tau), \hat{\pi}(\tau), u, \tau)$$

holds whenever $(u, \tau) \in \mathcal{L}$, where $\hat{\pi}$ is the solution of the adjoint equation $\nabla_{\hat{L}}\hat{\pi} = 0$ with terminal condition $\hat{\pi}(b_*) = \hat{\pi}^{\#}$.

Write $\hat{\pi}(t) = (p_0(t), \pi(t))$, using the identification $T^*_{\hat{\xi}_*(t)} \hat{M} \sim \mathbb{R} \times T^*_{\xi_*(t)} M$. Also, write $\hat{L}(t) = (\omega(t), L(t))$, using the identification of the measurable selections of the set-valued map $[a_*, b_*] \ni t \mapsto \partial \mathbf{f}_{\eta_*, t}(\xi_*(t))$ with pairs (ω, L) , as described above. Then ω is a field of covectors along ξ_* , and L is a selection of the map $[a_*, b_*] \ni t \mapsto \partial f_{\eta_*, t}(\xi_*(t))$, so in particular L gives rise to a covariant differentiation ∇_L . It is then easy to see that the adjoint equation $\nabla_{\hat{L}}\hat{\pi} = 0$ amounts to the pair of statements $\nabla_L \pi + p_0 \omega = 0$, $\dot{p}_0 = 0$. Hence, if we let $\pi_0 = -p_0$, we see that π_0 is constant as a function of t, and $\nabla_L \pi = \pi_0 \omega$. If we then write $\pi^{\#} = \pi(b_*)$, it is clear that $\hat{\pi}^{\#} = (-\pi_0, \pi^{\#})$, and the covector field π is the solution of $\nabla_L \pi = \pi_0 \omega$ with endpoint condition $\pi(b_*) = \pi^{\#}$. It follows that the Hamiltonian maximization condition takes the desired form $H^{\mathbf{f}}_{\pi_0}(\xi_*(\tau), \pi(\tau), \eta_*(\tau), \tau) \geq H^{\mathbf{f}}_{\pi_0}(\xi_*(\tau), \pi(\tau), u, \tau).$

Let $\hat{\nu} = (\nu_0, \nu)$, so that $\hat{\nu}(\Delta y_0, \Delta y) = \nu_0 \Delta y_0 + \nu \cdot \Delta y$ for every tangent vector $(\Delta y_0, \Delta y) \in T_{\hat{F}(\hat{\xi}_*(b_*))} \hat{N}$. Also, let $(\theta, \lambda) \in \tilde{\Lambda}$ be such that $\hat{\lambda} = \hat{\lambda}_{\theta,\lambda}$. Then, since $\hat{\pi}^{\#} = (-\pi_0, \pi^{\#})$, and $\hat{\pi}^{\#} = \hat{\nu} \circ \hat{\lambda}$, we have, if $\nu \in \mathbb{R} \times T_{\xi_*(b_*)} M$ and $\nu = (\Delta x_0, \Delta x)$, the equality $\hat{\pi}^{\#} \nu = -\pi_0 \Delta x_0 + \pi^{\#} \Delta x$, while on the other hand $\hat{\pi}^{\#} \nu$ satisfies $\hat{\pi}^{\#} \nu = \hat{\nu}(\hat{\lambda}\nu) = \hat{\nu} \cdot (\Delta x_0 + \theta \cdot \Delta x, \lambda \cdot \Delta x)$. Therefore $-\pi_0 \Delta x_0 + \pi^{\#} \Delta x = \nu_0 \Delta x_0 + \nu_0 \theta \cdot \Delta x + \nu \cdot \lambda \cdot \Delta x$ for every ν , and this implies that $-\pi_0 = \nu_0$ and $\pi^{\#} = \nu_0 \theta + \nu \cdot \lambda$, so $\pi^{\#} = -\pi_0 \theta + \nu \cdot \lambda$.

Then the condition $\hat{\pi}_0 \mu - \hat{\nu} \in \hat{\mathcal{C}}^{\perp}$ says that there exist cones $\hat{C}_j \in \hat{\mathcal{C}}$ and covectors $\hat{q}_j \in T^*_{F(\hat{\xi}_*(b_*))} \hat{N}$ such that $\hat{q}_j \to \hat{\pi}_0 \mu - \hat{\nu}$ and $\hat{q}_j \in \hat{C}_j^{\perp}$. On the other hand, each \hat{C}_j is a product $] - \infty, 0] \times C_j$ for some $C_j \in \mathcal{C}$. Write $\hat{q}_j = (q_{0,j}, q_j)$, $q_{0,j} \in \mathbb{R}, q_j \in T^*_{F(\xi_*(b_*))} N$. Then $q_{0,j} \to -\hat{\pi}_0 - \nu_0$ and $q_j \to -\nu$. Since $\hat{q}_j \in \hat{C}_j^{\perp}$, we have $\hat{q}_j (\Delta y_0, \Delta y) \leq 0$ for all $(\Delta y_0, \Delta y) \in] - \infty, 0] \times C_j$. Hence $q_{0,j} \geq 0$ and $q_j \in C_j^{\perp}$. Since $q_{0,j} \to -\hat{\pi}_0 - \nu_0$, we conclude that $-\hat{\pi}_0 - \nu_0 \geq 0$, and then $\nu_0 \leq -\hat{\pi}_0$. Since we know that $\hat{\pi}_0 \geq 0$, we can conclude that $\nu_0 \leq 0$, so $\pi_0 \geq 0$. Since $q_j \to -\nu$, the covector $-\nu$ belongs to \mathcal{C}^{\perp} . Since $\pi^{\#} = -\pi_0 \theta + \nu \cdot \lambda$, we have established the transversality condition.

We are now in a position to prove the nontriviality condition, which is our only missing conclusion. Suppose that this condition is false. Then $\nu = 0$ and $\pi_0 = 0$. But we know that $-\pi_0 = \nu_0$. So $\nu_0 = 0$, and then $\hat{\nu} = 0$. Furthermore, we also know that $\nu_0 \leq -\hat{\pi}_0$. So $\hat{\pi}_0 \leq 0$. Since $\hat{\pi}_0 \geq 0$, we conclude that $\hat{\pi}_0 = 0$. So $(\hat{\pi}_0, \hat{\nu}) = (0, 0)$, contradicting the fact that $(\hat{\pi}_0, \hat{\nu}) \neq (0, 0)$. This completes our proof.

5.4 Theorem 5.6 easily implies Theorem 5.1

We have used Theorem 5.1 as our main tool to derive Theorem 5.6. For completeness, we now prove that Theorem 5.1 is in turn a simple consequence of Theorem 5.6.

Assume that Theorem 5.6 holds. Let a data 14-tuple \mathcal{D}^{sep} as in (19) be given, such that all the assumptions of Theorem 5.1 hold. Fix a covector $\mu \in T^N_{F_*(\xi_*(b_*))} \setminus \{0\}$. Then apply Theorem 5.6 to the optimal control problem in which $f_0 \equiv 0$ and $\varphi = \psi \circ F$, where $\psi : N \mapsto \mathbb{R}$ is a function of class C^1 such that $d\psi(F(\xi_*(b_*))) = -\mu$, taking as $\tilde{\Lambda}$ the set $\{(-\mu\lambda, \lambda) : \lambda \in \Lambda\}$. It is easy to see that the reference TCP (ξ_*, η_*) is optimal. Theorem 5.6 then gives a 4-tuple $(\pi_0, \pi^{\#}, \tilde{\nu}, (-\mu\lambda, \lambda))$ such that $-\tilde{\nu} \in \mathcal{C}^{\perp}, \pi^{\#} = \tilde{\nu}\lambda + \pi_0\mu\lambda$, and, if π satisfies $\nabla_L \pi = 0, \pi(b_*) = \pi^{\#}$, then the Hamiltonian maximization conditions hold. Then, if we let $\nu = \pi_0 \mu + \tilde{\nu}$, we see that $\pi_0 \mu - \nu \in \mathcal{C}^{\perp}$, and $\pi^{\#} = \nu \lambda$. Furthermore, it is clear that the nontriviality condition $(\pi_0, \tilde{\nu}) \neq (0, 0)$ implies $(\pi_0, \nu) \neq (0, 0)$. It is then clear that the 4-tuple (π_0, ν, λ, L) satisfies all the conclusions of Theorem 5.1.

References

- [1] Berkovitz, L. D., Optimal Control Theory. Springer-Verlag, New York, 1974.
- [2] Bianchini, R. M., and G. Stefani, "Controllability along a trajectory: a variational approach." SIAM J. Control Optim. **31** (1993), pp. 900–927.
- [3] Bianchini, R. M., "Variation of a control process at the initial point." J. Opt.Theory and Appl. 81 (1994), pp. 249–258.
- Bressan, A., "A high-order test for optimality of bang-bang controls." SIAM J. Control Optim. 23 (1985), pp. 38–48.
- [5] Bressan, A., "On the intersection of a Clarke cone with a Boltyanskii cone." SIAM J. Control Optim., to appear.
- [6] Clarke, F. H., "The Maximum Principle under minimal hypotheses." SIAM J. Control Optim. 14, 1976, pp. 1078–1091.
- [7] Clarke, F. H., Optimization and Nonsmooth Analysis. Wiley Interscience, New York, 1983.
- [8] Clarke, F. H., Yu. S. Ledyaev, R. J. Stern, and P. R. Wolenski, Nonsmooth Analysis and Control Theory. Springer Verlag, Graduate Texts in Mathematics No. 178; New York, 1998.
- [9] Clarke, F. H., Necessary conditions in dynamic optimization. *Memoirs Amer. Math. Soc.* 816, vol. 173, 2005.
- [10] de Pinho, R., "An Euler-Lagrange inclusion for optimal control problems." IEEE Trans. Automat. Control 40 (1995), no. 7, pp. 1191–1198.
- [11] Hermes, H., "On local and global controllability." SIAM J. Control 12 (1974), pp. 252–261.
- [12] Ioffe. A., "Euler-Lagrange and Hamiltonian formalisms in dynamic optimization." Trans. Amer. Math. Soc. 349 (1997), pp. 2871–2900
- [13] Ioffe, A., and R. T. Rockafellar, "The Euler and Weierstrass conditions for nonsmooth variational problems." Calc. Var. Partial Differential Equations 4 (1996), pp. 59–87.
- [14] Knobloch, H. W., High Order Necessary Conditions in Optimal Control. Springer-Verlag, Berlin, 1975.

- [15] Krener, A. J., "The High Order Maximal Principle and Its Application to Singular Extremals." SIAM J. Contrl. Optim. 15 (1977), pp. 256-293.
- [16] Łojasiewicz Jr., S, personal communication, unpublished.
- [17] Mordukhovich, B., Variational Analysis and Generalized Differentiation, I: Basic Theory; II: Applications. Grundlehren Series (Fundamental Principles of Mathematical Sciences), Vol. 330 and 331, Springer Verlag, Berlin, 2006.
- [18] Pontryagin, L. S., V. G. Boltyanskii, R. V. Gamkrelidze and E. F. Mischenko, *The Mathematical Theory of Optimal Processes.* Wiley, New York, 1962.
- [19] Stefani, G., "On maximum principles." In Analysis of Controlled Dynamical Systems, Bonnard, Bride, Gauthier, Kupka, eds., Progress in Systems and Control Theory, vol. 8, Birkhäuser, Boston 1991.
- [20] Sussmann, H. J., "A general theorem on local controllability." SIAM J. Control Optim. 25 (1987), pp. 158–194.
- [21] Sussmann, H. J., "A strong version of the Lojasiewicz Maximum Principle." In Optimal Control of Differential Equations, (Athens, OH, 1993), Nicolai H. Pavel Ed., Lect. Notes in Pure and Applied Math. 160, Marcel Dekker Inc., New York, 1994, pp. 293-309.
- [22] Sussmann, H. J., "An introduction to the coordinate-free maximum principle." In *Geometry of Feedback and Optimal Control*, B. Jakubczyk and W. Respondek Eds., M. Dekker, Inc., New York, 1997, pp. 463-557.
- [23] Sussmann, H. J., "A nonsmooth hybrid maximum principle." In Stability and Stabilization of Nonlinear Systems, D. Aeyels, F. Lamnabhi-Lagarrigue and A. J. van der Schaft Eds. Lect. Notes Control and Information Sciences 246, Springer-Verlag, 1999, pp. 325–354.
- [24] 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.
- [25] Sussmann, H.J., "Optimal control of nonsmooth systems with classically differentiable flow maps." In Proc. 6th IFAC Symposium on Nonlinear Control Systems (NOLCOS 2004), Stuttgart, Germany, September 1-3, 2004, Vol. 2, pp. 609-704.
- [26] Sussmann, H.J., "Combining high-order necessary conditions for optimality with nonsmoothness." In Proc. 43rd IEEE 2004 Conference on Decision and Control (Paradise Island, the Bahamas, December 14-17, 2004), IEEE Publications, New York, 2004.
- [27] Sussmann, H.J., "Set transversality, approximating multicones, Warga derivate containers and Mordukhovich cones." In the *Proceedings of the 44th IEEE 2005 Conference on Decision and Control, held in Sevilla, Spain, December 12-15,* 2005), IEEE Publications, New York, 2005.

- [28] Sussmann, H.J., "Generalized differentials, variational generators, and the maximum principle with state constraints." To appear in Nonlinear and Optimal Control Theory; Cetraro, Italy, June 2004, proceedings of a summer school of the Fondazione C.I.M.E. (Book coauthored by Andrei Agrachev, Stephen Morse, Eduardo Sontag, Héctor Sussmann, and Vadim Utkin.) Gianna Stefani and Paolo Nistri Eds. Springer-Verlag Lecture Notes in Mathematics.
- [29] Vinter, R. B., Optimal Control. Birkhäuser, Boston, 2000.
- [30] Warga, J., "Fat homeomorphisms and unbounded derivate containers." J. Math. Anal. Appl. 81, 1981, pp. 545–560.
- [31] Warga, J., "Controllability, extremality and abnormality in nonsmooth optimal control." J. Optim. Theory Applic. 41, 1983, pp. 239–260.
- [32] Warga, J., "Optimization and controllability without differentiability assumptions." SIAM J. Control and Optimization 21, 1983, pp. 837–855.
- [33] Warga, J., "Homeomorphisms and local C¹ approximations." Nonlinear Anal. TMA 12, 1988, pp. 593–597.