ESAIM: COCV 29 (2023) 17
https://doi.org/10.1051/cocv/2023003

ESAIM: Control, Optimisation and Calculus of Variations
www.esaim-cocv.org

GOH CONDITIONS FOR MINIMA OF NONSMOOTH PROBLEMS WITH UNBOUNDED CONTROLS

Francesca Angrisani[1,*] and Franco Rampazzo[2]

Abstract. Higher order necessary conditions for a minimizer of an optimal control problem are generally obtained for systems whose dynamics is continuously differentiable in the state variable. Here, by making use of the notion of set-valued Lie bracket, we obtain a Goh-type condition for a control affine system with Lipschitz continuous dynamics and unbounded controls. In order to manage the simultaneous lack of smoothness of the adjoint equation and of the Lie bracket-like variations we make use of the notion of Quasi Differential Quotient. We conclude the paper with a worked out example where the established higher order condition is capable to rule out the optimality of a control verifying the standard maximum principle.

Mathematics Subject Classification. 49K15, 49N25, 49K99.

Received September 30, 2022. Accepted December 28, 2022.

1. Introduction

The so-called higher order necessary conditions for minima, a classical subject of investigation since the early developments of Calculus of Variations, have been variously generalized to Optimal Control Theory. In particular, they constitute a crucial issue in geometric control, in connection with the Lie algebraic structure associated with the dynamics. Clearly, such a structure is present only under hypotheses of C^∞ regularity for the involved vector fields. Yet, if only Lie brackets up to a certain length are used, one can assume less demanding smoothness hypotheses, as it happens, for instance, for the Goh condition. Indeed, the mere definition of Lie bracket of two vector fields g_i, g_j, namely $[g_i, g_j] := Dg_j g_i - Dg_i g_j$, only requires that they be differentiable. It is then plausible to wonder if one can further weaken the regularity assumptions, for instance by allowing vector fields to be just locally Lipschitz continuous. On the one hand, this would be of obvious interest for applications, and, on the other hand, it would be in line with the vast and rich literature on *nonsmooth* optimal control, which, since the early Seventies, has generated some *maximum principles* involving suitable notions of generalized differentiation (see [5, 13, 20, 27–31]).

Let us observe that, when two vector fields g_i, g_j are locally Lipschitz continuous, Rademacher's theorem implies that the differentiation domains $\mathrm{Diff}(g_i)$ and $\mathrm{Diff}(g_j)$ of g_i and g_j, respectively, have full measure, so that the Lie bracket $[g_i, g_j]$ (existing on $\mathrm{Diff}(g_i) \cap \mathrm{Diff}(g_j)$) turns out to be defined almost everywhere.

Keywords and phrases: Goh conditions, nonsmooth optimal control, set-valued Lie brackets, unbounded controls.

[1] Department of Information Engineering, Computer Science and Mathematics (DISIM), University of L'Aquila, L'Aquila, Italy.

[2] Department of Mathematics "T. Levi-Civita", University of Padova, Padova, Italy.

[*] Corresponding author: francescaangrisani@gmail.com

Actually, some results including Lie brackets have been generalized to nonsmooth systems by means of such an almost everywhere approach (see, for example, [11, 18, 24, 26]). As an alternative, and in connection with controllability problems for Lipschitz continuous control systems, an *everywhere-defined, set-valued,* Lie bracket has been introduced in [25]. If one uses $\overline{\mathrm{co}}(E)$ to denote the closed convex hull of a given set E, this bracket is defined by setting, for every $x \in \mathbb{R}^n$,

$$[g_i, g_j]_{set}(x) = \overline{\mathrm{co}}\Big\{ \lim_{n\to\infty} [g_i, g_j](x_n), \qquad \lim_{n\to\infty} x_n = x, \ x_n \in \mathrm{Diff}(f) \cap \mathrm{Diff}(g) \Big\}, \tag{1.1}$$

where we mean that limits are taken along all sequences $(x_n) \subset \mathrm{Diff}(g_i) \cap \mathrm{Diff}(g_j)$ converging to x. In [23, 24] this notion of set-valued Lie bracket—which can be defined on differential manifolds as well-proved suitable for the generalization of some basic results of differential geometry like the commutativity criterion, Frobenius Theorem, and Rashevski-Chow Theorem. Therefore, one might naturally consider the following question:

Q. *Given a minimizer of a nonsmooth optimal control problem, can one complement the (first order) Maximum Principle with conditions which involve set-valued Lie brackets?*

In the present paper we provide a first positive answer to question **Q** in relation with a control affine minimum problem

$$(P) \qquad \begin{cases} \text{minimize } \Psi(T, x(T)) \\ \text{over processes } (T, u, x) \ \text{ of} \\ \dfrac{\mathrm{d}x}{\mathrm{d}t} = f(x) + \displaystyle\sum_{i=1}^{m} g_i(x) u^i \\ x(0) = \hat{x} \quad \|u\|_1 \le K \quad (T, x(T)) \in \mathfrak{T}. \end{cases} \tag{1.2}$$

and proving a necessary condition generalizing Goh condition. We assume that the vector fields $g_i : \mathbb{R}^n \to \mathbb{R}^n$, $i = 1, \dots, m$, are locally Lipschitz continuous, while the controls u take values in $\mathbb{R}^m$ and have L^1-norm less than or equal to K. The end-time T and the final state $x(T)$ are subject to a constraint of the form $(T, x(T)) \in \mathfrak{T}$, where the *target* $\mathfrak{T}$ is a given closed subset of the time-space product $\mathbb{R}_+ \times \mathbb{R}^n$. This is actually a simplified version of the problem addressed in the paper, in which the drift f is allowed to depend on a bounded control a as well and, moreover, a Lagrangian cost $l = l(t, x, u)$ is considered together with the final cost Ψ. Moreover, the set where the control u takes values is not necessarily the whole $\mathbb{R}^m$ but it might be allowed to be a closed cone $C = C_1 \times C_2$, where C_2 is a cone and for some non negative integer $m_1 \le m$, $C_1 \subseteq \mathbb{R}^{m_1}$ is a closed cone containing the coordinate axes.

Since the velocities of the trajectories are unbounded and no suitable growth condition prevents the system from an impulsive behaviour, we first embed the problem in the time-space optimal control problem

$$(P_{ext}) \qquad \begin{cases} \text{minimize } \Psi(y^0(S), y(S)) \\ \text{over processes } (S, w^0, w, y^0, y), \\ \text{where } (y^0, y) : [0, S] \to \mathbb{R}^{1+n} \text{ solves} \\ \dfrac{\mathrm{d}y^0}{\mathrm{d}s} = w^0 \\ \dfrac{\mathrm{d}y}{\mathrm{d}s} = f(y) w^0 + \displaystyle\sum_{i=1}^{m} g_i(y) w^i \\ (y^0, y)(0) = (0, \hat{x}) \quad \|w\|_1 \le K, \quad (y^0, y)(S) \in \mathfrak{T}, \end{cases}$$

where the controls (w^0, w) belong to the set $\bigcup_{S>0} \left\{ (w^0, w) \in L^\infty([0, S], \mathbb{R}_+ \times \mathbb{R}^m) : w^0(s) + |w(s)| = 1 \right\}$ and y^0 stands for the actual time t variable. Problem (P_{ext}) is simply obtained from (P) by first reparametrizing time through

$$t(s) = y^0(s) := \int_0^s w^0(\sigma) \mathrm{d}\sigma, \qquad w^0 > 0, \qquad w(s) := (u \circ t)(s) w^0(s), \qquad y(s) := (x \circ y^0)(s),$$

and then allowing subintervals $I \subseteq [0, S]$ such that $w^0(s) \equiv 0 \, \forall s \in I$ (*impulsive subintervals*). Notice, in particular, that, unlike (P), (P_{ext}) is a problem with L^∞-bounded controls. Incidentally, notice also that, since (P_{ext}) is rate-independent, the control constraint $w^0 + |w| = 1$ is not restrictive.

As for necessary conditions for a minimizer $(\overline{S}, \overline{w}^0, \overline{w}, \overline{y}^0, \overline{y})$ of the extended problem $(P)_{ext}$, answering question **Q** should mean complementing the usual, non smooth, maximum principle (in one of the available versions) with conditions that tell something about the relation between the corresponding adjoint variable $p(\cdot)$ and set-valued Lie brackets $[g_i, g_j]_{set}$.

Our main result, which we state below in a simplified form –see Theorem 3.6 for a rigorous and more general statement–, actually says that at almost every $s \in [0, \overline{S}]$ the following non smooth Goh condition holds true:

Theorem 1.1 (**Maximum Principle).** *Let $(\overline{S}, \overline{w}^0, \overline{w}, \overline{y}^0, \overline{y})$ be a local minimizer for the extended problem (P_{ext}), and assume that $\|\overline{w}\|_1 < K$.*

Then there exist multipliers $(p_0, p, \lambda) \in \mathbb{R}^* \times AC([0, \overline{S}]; (\mathbb{R}^n)^*) \times \mathbb{R}^*$ such that, besides the standard necessary conditions of Pontryagin Maximum Principle (expressed withing a suitable non-smooth setting), we have

$$\boxed{0 \in p(s) \cdot [g_i, g_j]_{set}(\overline{y}(s)),} \tag{1.3}$$

for any $i, j \in \{1, \dots, m\}$ and for almost any $s \in [0, \overline{S}]$.

Clearly, the importance of such a result relies on the possibility that a given control $\hat{u}$ while being allowed by the standard, first order, maximum principle, does not verify condition (1.3), which would rule out the optimality of $\hat{u}$. A toy example at the end of the paper illustrates this circumstance.

Let us mention that a crucial tool for the proof of Theorem 1.1 (in the more general version of Thm. 3.6), is represented by the notion of Quasi Differential Quotient (QDQ), a generalized differentation (valid for set-valued maps) introduced in [21] as a special case of H. Sussmann's Approximate Generalized Differential Quotients [1]. Actually, this tool and the corresponding notion of approximating multicone are flexible enough to allow the managing, whitin the same theoretical frame, of two different kinds of nonsmoothness: the one generated by the adjoint inclusion (which involves Clarke's generalized Jacobian), and the one derived by the utilization of set-valued brackets. In particular, one constructs variational QDQ-approximating multicones generated by both multiple set-valued Lie brackets and the classical needle variations, and proves a linear separability condition with QDQ-approximating multicones of the set of profitable states. In turn, this geometrical result is equivalent to the existence of the multipliers (p_0, p, λ) stated in Theorem 1.1.

To conclude this presentation, let us briefly mention that the interest for the issues treated in this paper is justified by several applications, for instance in classical mechanics (as soon one identifies the control with a moving part of a given mechanical system), in neurological dynamics, or aerospace navigation [6–9, 12, 16, 17, 22]. A further application which deserves some attention is that of driftless control systems, in particular the case of sub-Riemannian geometry. For instance, a result as the one presented here might be regarded as a contribution to the study of sub-Riemannian metrics having low regularity (see, for instance, [18]).

Finally, let us spend a few words on possible developments of the present work. On the one hand, aiming to higher order necessary conditions involving *iterated set-valued Lie-Brackets* (which were introduced in [15]), one should begin by proving that such brackets are QDQ of suitable multiflows. This is not straighforward, for iterated set-valued Lie-Brackets are larger then the objects one would obtain by a mere recursive approach. On the other hand, it might be interesting to generalize the Goh-like condition proved here to the standard case of *bounded* controls. Actually, in this case the construction of the bracket-like variation cannot ignore the influence of the non-zero drift (as instead it happens with unbounded controls). A similar, though probably more complex, argument would obviously be involved in the investigation of a non-smooth version of the Legendre-Clebsch condition.

The paper is organized as follows: in the next subsection we introduce some conventions for notation; Section 2 is devoted to the introduction of some crucial theoretical tools, like set-valued Lie brackets and Quasi Differential Quotients; in Section 3 we present the minimum problem as well as its extended version, and state the main result (Thm. 3.6); Section 4 is dedicated to the proof of the main result, while in Section 5 we describe a toy example where the effectiveness of the new, higher order, necessary condition is displayed in ruling out a first order extremal.

1.1. Notation

We shall use $\mathbb{R}_+$ and $\mathbb{R}_-$ to denote the intervals $[0,+\infty[$ and $]-\infty,0]$, respectively. The space of all linear operators from a vector space X to a vector space Y will be denoted by $\mathrm{Lin}(X,Y)$. The elements of an Euclidean space $\mathbb{R}^q$, $q\geq 0$, will be thought as column vectors, while row vectors will stand for the linear one-forms, *i.e.* the elements of the dual space $(\mathbb{R}^q)^* := \mathrm{Lin}(\mathbb{R}^q,\mathbb{R})$. If p,q are positive integers, the space $\mathrm{Lin}(\mathbb{R}^p,\mathbb{R}^q)$ will be sometimes regarded as the space of $q\times p$ real-valued matrices. For every $i=1,\dots,q$, $\mathbf{e}_i$ [resp. $\mathbf{e}^i$] will denote the i-th vector of the canonical basis of $\mathbb{R}^q$ [resp. $(\mathbb{R}^q)^*$]

If $K\in L^1([0,T],\mathrm{Lin}(\mathbb{R}^n,\mathbb{R}^n))$, *i.e.* $K=K(\cdot)$ is a $n\times n$-matrix-valued L^1 map on $[0,\bar S]$, we use the exponential notation $[0,S]^2\ni(s_1,s_2)\mapsto e^{\int_{s_1}^{s_2}K}$ to denote the *fundamental matrix solution* of the (time-dependent) linear equation $\dot v = Kv$: namely, for every $s_1,s_2\in[0,\bar S]$, $\bar v\in\mathbb{R}^n$, $e^{\int_{s_1}^{s_2}K}\bar v = v(s_2)$, where $v(\cdot)$ is the solution to the linear Cauchy problem $\dot v = Kv,\ v(s_1)=\bar v$.

If $\mathscr{E}$ is a metric space with a metric d, and $\bar e\in\mathscr{E}$, $Q\subseteq\mathscr{E}$, we use $d(\bar e,Q)$ to denote *the distance of* $\bar e$ *from* Q, namely $(\bar e,Q) := \inf\limits_{e\in Q} d(\bar e,e)$.

For any subset $Q\subseteq E$ of an Euclidean space E, by $\mathrm{co}(Q)$ we mean the *convex hull of* Q, *i.e.* the smallest convex set containing Q, obtainable by intersection of all convex sets containing Q. The symbol $\overline{\mathrm{co}}(Q)$ denotes the *closed convex hull of* Q, *i.e.* the closure of $\mathrm{co}(Q)$, which happens to be smallest closed convex set containing Q.

For any subset $Q\subseteq X$ of a topological space X, we use $\mathring{Q}$ to denote the set of interior points of Q. If n,m are positive integers, $E\subseteq\mathbb{R}^n$, is any subset and $F:E\to\mathbb{R}^m$ is a map, differentiable at some point $x\in\mathring{E}$, we use $DF(x)$ to denote the differential of F at x.

If H,K are sets, by using the notation $\mathscr{F}:H\rightsquigarrow K$ we mean that $\mathscr{F}$ is a set-valued map from H into K, that is a map from H into the power set $\mathscr{P}(K)$.

To save space, with the expressions "Lipschitz function", "Lipschitz map", "Lipschitz vector field", we will mean "Lipschitz continuous function", "Lipschitz continuous map", "Lipschitz continuous vector field", respectively.

If $A\subset\mathbb{R}\times\mathbb{R}^q$ is an open subset and $F:A\to\mathbb{R}^q$ is a time-dependent vector field, continuous in x and measurable in t, the expression $x(\cdot)$ *is a solution of the differential equation* $\dot x = F$ *on an interval* I will always mean that $(t,x)\in A(t)$ for every $t\in I$ and $x(\cdot)$ is a *Carathéodory* solution of $\dot x = F(t,x)$, namely $x(\cdot)$ is absolutely continuous and the equality $\dot x(t) = F(t,x)(t)$ is verified at almost every $t\in I$.

More generally, if $X:A\rightsquigarrow\mathbb{R}^q$ is a set-valued vector field, the expression $x(\cdot)$ *is a solution of the differential inclusion* $\dot x(t)\in X(t,x)$ *on* I will mean that $x(\cdot)$ is absolutely continuous and $\dot x(t)\in X(t,x(t))$, for a.e. $t\in I$.

2. Set-valued Lie brackets and QDQ approximating cones

2.1. Lie brackets for Lipschitz vector fields

If $f:\mathbb{R}^n\to\mathbb{R}^n$ is a Lipschitz vector field, we use $\mathrm{Diff}(f)$ to denote the set of differentiability points of f, which, by Rademacher's theorem, has full measure. The following notion of *set-valued Lie bracket of two*

Lipschitz continuous vector fields $f, g : \mathbb{R}^n \to \mathbb{R}^n$ has been introduced in [25]:

$$[f,g]_{set}(x) := \overline{\text{co}}\Big\{v \in \mathbb{R}^n, \quad v = \lim_{n\to\infty}[f,g](x_n), \quad x_n \to x, \ (x_n) \subset \text{Diff}(f) \cap \text{Diff}(g)\Big\},$$

In this formula we mean that *all* sequences $(x_n) \subset \text{Diff}(f) \cap \text{Diff}(g)$ are considered for which $x_n \to x$ and the limits $\lim_{n\to\infty}[f,g](x_n)$ do exist. Since f, g, Df, Dg are bounded in a neighborhood of any point $x \in \mathbb{R}^n$, by compactness one has $[f,g]_{set}(x) \neq \emptyset$. Moreover $[f,g]_{set}(x) = \{[f,g](x)\}$ as soon as f, g are of class C^1 in a neighborhood of x. One trivially has that the relations $[f,f]_{set} = \{0\}$ and $[f,g]_{set} = -[g,f]_{set}$ keep holding for set-valued brackets, with the understanding that, for any subset S of a vector space, $-S$ is the set of opposites of elements in S. Furthermore, some basic results have been generalized to set-valued Lie brackets. For instance, the flow of f, g locally commute if and only if $[f,g]_{set} = \{0\}$. Furthermore, a Frobenius-type result holds true for Lipschitz distributions (see [24]), as well as a local controllability theorem analogous to Rashewski-Chow's [25]. [1]

2.2. Quasi Differential Quotients (QDQs)

We call a function $\rho : \mathbb{R}_+ \to \mathbb{R}_+ \cup \{+\infty\}$ a *pseudo-modulus* if it is monotonically nondecreasing and $\lim_{s\to 0^+} \rho(s) = \rho(0) = 0$.

Let us recall the concept of *Quasi Differential Quotient* for set-valued maps, which was introduced in [21] to address infimum gaps problems.

Definition 2.1 (Quasi Differential Quotients (QDQ))**.** Let N, n be non negative integers. Let $F : \mathbb{R}^N \rightsquigarrow \mathbb{R}^n$ and $\Lambda \subset \text{Lin}\{\mathbb{R}^N, \mathbb{R}^n\}$ be a set-valued map and a compact subset, respectively, and let $\Gamma \subset \mathbb{R}^N$ be any subset. Given a pair $(\bar{x}, \bar{y}) \in \mathbb{R}^N \times \mathbb{R}^n$, we say that Λ is a *Quasi Differential Quotient (QDQ) of F at $(\bar{x}, \bar{y})$ in the direction of* Γ if there exists a pseudo-modulus ρ such that for every δ with $\rho(\delta) < +\infty$ there is a continuous map $(L_\delta, h_\delta) : (\bar{x} + B_\delta) \cap \Gamma \to \text{Lin}(\mathbb{R}^N, \mathbb{R}^n) \times \mathbb{R}^n$ verifying

$$\begin{gathered} \min\nolimits_{L'\in\Lambda} |L_\delta(x) - L'| \leq \rho(\delta), \quad |h_\delta(x)| \leq \delta\rho(\delta), \quad \text{and} \\ \bar{y} + L_\delta(x) \cdot (x - \bar{x}) + h_\delta(x) \in F(x), \end{gathered} \tag{2.1}$$

whenever $x \in (\bar{x} + B_\delta) \cap \Gamma$.

Remark 2.2. If F is a single-valued, continuous map, one has necessarily $\bar{y} = F(\bar{x})$, so the inclusion in (2.1) reduces to an equality. Furthermore, if there exists $\tilde{\varepsilon} > 0$ and a continuous map $L : (\bar{x} + B_{\tilde{\varepsilon}}) \cap \Gamma \to \text{Lin}(\mathbb{R}^N, \mathbb{R}^n)$ satisfying

$$\lim_{\Gamma \ni x \to \bar{x}} \text{dist}\,(L(x), \Lambda) = 0, \qquad \text{and} \qquad \lim_{\Gamma \ni x \to \bar{x}} \frac{|F(x) - F(\bar{x}) - L(x)\cdot(x - \bar{x})|}{|x - \bar{x}|} = 0,$$

then Λ is a QDQ of F at $(\bar{x}, F(\bar{x}))$ in the direction of Γ. To see this, it is sufficient to define L_δ as the restriction of L to $\bar{x} + B_\delta \cap \Gamma$ and to consider the modulus $\rho : (0, \tilde{\varepsilon}) \mapsto \mathbb{R}$ defined by setting

$$\rho(\delta) := \max\left\{\sup_{x\in B_\delta\cap\Gamma} \text{dist}\,(L(x), \Lambda)\,, \sup_{x\in B_\delta\cap\Gamma} \frac{|F(x) - F(0) - L(x)\cdot x|}{\delta}\right\} \qquad \delta \in (0, \tilde{\varepsilon}).$$

[1] In [15], a notion of iterated set-valued Lie brackets suitable for local controllability issues have been investigated as well.

2.2.1. *Some basic properties of QDQs*

Proposition 2.3. *[2] Let $F : \mathbb{R}^N \rightsquigarrow \mathbb{R}^n$ and $G : \mathbb{R}^N \rightsquigarrow \mathbb{R}^q$ be set-valued maps. Assume that $\bar{x} \in \mathbb{R}^N$, $\bar{y}, \bar{y}_F, \bar{y}_G \in \mathbb{R}^n, \bar{y}_F \in \mathbb{R}^n, \bar{y}_G \in \mathbb{R}^q$, $\Gamma, \Gamma_F, \Gamma_G \subseteq \mathbb{R}^N, \Gamma_F \subseteq \mathbb{R}^n$, and $\alpha, \beta \in \mathbb{R}$. Then:*

1. [Locality] *If $n = q, U$ is a neighborhood of $\bar{x}$ and $F(x) = G(x)$ for $x \in U \cap \Gamma$, then Λ is a QDQ of F at $(\bar{x}, \bar{y})$ in the direction of Γ if and only if it is a QDQ for G at $(\bar{x}, \bar{y})$ in the direction of Γ.*
2. [Linearity] *If $n = q, \Lambda_F$ and Λ_G are QDQ of F and G at points $(\bar{x}, \bar{y}_F)$ and $(\bar{x}, \bar{y}_G)$ in the direction of Γ_F and Γ_G, respectively, then $\alpha\Lambda_F + \beta\Lambda_G$ is a QDQ of $\alpha F + \beta G$ at point $(\bar{x}, \alpha\bar{y}_F + \beta\bar{y}_G)$ in the direction of $\Gamma_F \cap \Gamma_G$.*
3. [Set product property] *$\Lambda_F \times \Lambda_G$ is a QDQ at $\Big(\bar{x}, (\bar{y}_F, \bar{y}_G)\Big)$, in the direction of $\Gamma_F \cap \Gamma_G$, of the set-valued map $F \times G : x \rightsquigarrow F(x) \times G(x)$.*
4. [Product Rule] *If $n, q = 1$, and still using the same notation as in (2), we have $F(\bar{x})\Lambda_G + G(\bar{x})\Lambda_F$ is a QDQ of $FG : x \rightsquigarrow F(x)G(x)$.*[2]
5. *If F is single-valued and $L \in \mathrm{Lin}(\mathbb{R}^N, \mathbb{R}^n)$, $\{L\}$ is a QDQ of F at $(\bar{x}, \bar{y}_F)$ in the direction of $\mathbb{R}^n$ if and only if F is differentiable at $\bar{x}$ and $L = DF(\bar{x})$.*

Proposition 2.4 (Chain rule)**.** *[2] Let $F : \mathbb{R}^N \rightsquigarrow \mathbb{R}^n$ and $G : \mathbb{R}^n \rightsquigarrow \mathbb{R}^l$ be set-valued maps, and consider the composition $G \circ F : \mathbb{R}^N \ni x \rightsquigarrow \bigcup\limits_{y \in F(x)} G(y) \in \mathbb{R}^l$. Assume that Λ_F is a QDQ of F at $(\bar{x}, \bar{y})$ in the direction of Γ_F and Λ_G is a QDQ of G at $(\bar{y}, \bar{z})$ in a direction Γ_G verifying $\Gamma_G \supseteq F(\Gamma_F)$. Then the set $\Lambda_G \circ \Lambda_F := \Big\{ML, \quad M \in \Lambda_G, L \in \Lambda_F\Big\}$ is a QDQ of $G \circ F$ at $(\bar{x}, \bar{z})$ in the direction of Γ_F.*

The following technical result will be essential for computing QDQs of multiple variations of a given control process.

Proposition 2.5. *Let N, q be positive integers, and let $F : \mathbb{R}^N_+ \to \mathbb{R}^q$ be a map such that*

$$F(\boldsymbol{\varepsilon}) - F(0) = \sum_{i=1}^{N} \big(F(\varepsilon^i \mathbf{e}_i) - F(0)\big) + o(|\boldsymbol{\varepsilon}|), \qquad \forall \boldsymbol{\varepsilon} = (\varepsilon^1, \ldots, \varepsilon^N) \in \mathbb{R}^N. \tag{2.2}$$

If, for any $i = 1, \ldots, N$, $\Lambda_i \subset Lin(\mathbb{R}, \mathbb{R}^q)$ is a QDQ at $0 \in \mathbb{R}$ of the map $\alpha \mapsto F(\alpha \mathbf{e}_i)$ in the direction of $\mathbb{R}_+$, then the compact set

$$\Lambda := \Big\{L \in Lin(\mathbb{R}^N, \mathbb{R}^q), \quad L(v) := L_1 v^1 + \ldots + L_N v^q, \quad (L_1, \ldots, L_N) \in \Lambda_1 \times \cdots \times \Lambda_N\Big\}$$ [3]

is a QDQ of F at 0 in the direction of $\mathbb{R}^N_+$.

Proof. Since, for every vector $k \in \mathbb{R}^q$ and every map $\psi : \mathbb{R}^N_+ \to \mathbb{R}^q$ such that $\psi(\boldsymbol{\varepsilon}) = o(|\boldsymbol{\varepsilon}|)$, the singleton $\{0\} \subset \mathbb{R}^N \times \mathbb{R}^q$ is a QDQ at 0 of the map $k + \psi$ in the direction of any subset of $\mathbb{R}^N_+$, in view of the linearity property of the QDQ it is not restrictive to assume the condition $F(\boldsymbol{\varepsilon}) = \sum\limits_{i=1}^{N} F(\varepsilon^i \mathbf{e}_i)$ instead of (2.2).
On the other hand, since $\varepsilon_i \mathbf{e}_i = P^i(\boldsymbol{\varepsilon})$ for every $i = 1, \ldots, N$, where P^i is the projection on the i-axis of $\mathbb{R}^N$, $F(\boldsymbol{\varepsilon}) = \sum\limits_{i=1}^{N} F(\varepsilon^i \mathbf{e}_i)$ reads $F = \sum\limits_{i=1}^{N} F \circ P^i$. Therefore the thesis follows from the chain rule, the linearity, and the trivial fact that, for every $j = 1, \ldots, N$, the singleton $\{\mathbf{e}^j\}$ is a QDQ (at any point and in any direction) of the projection P^j. $\square$

[2] We are using $A + B$ to denote the set of sum of elements of two subsets $A, B \subset W$ of a vector space W. Furthermore, if F is the field underlying a vector space W and $\Omega \subset F$, we use the notation $\Omega A := \{\omega a, \ \omega \in \Omega, a \in A\}$.

[3] As a set of matrices, Λ is made of $q \times N$ matrices whose i-th column is an element of Λ_i.

We will regard the Clarke's Generalized Jacobian as a particular QDQ. Let us begin recalling its definition:

Definition 2.6 (Clarke's Generalized Jacobian)**.** Let $F:\mathbb{R}^N\to\mathbb{R}^n$ be a map and assume that it is Lipschitz in a neighborhood of a point $x\in\mathbb{R}^N$. The subset

$$\partial_x^C F(x) := \overline{\mathrm{co}}\Big\{L=\lim_{n\to\infty} DF(x_n),\ \ x_n\to x\ \ (x_n)\subset \mathrm{Diff}(F)\Big\}\subseteq \mathrm{Lin}(\mathbb{R}^N,\mathbb{R}^n)$$

(where it is meant that one takes limits along all sequences $(x_n)\subset\mathrm{Diff}(F)$ converging to x) is called *Clarke's Generalized Jacobian of F at x.*

Proposition 2.7 (Clarke's Generalized Jacobian is a QDQ)**.** *Let $\Omega\subset\mathbb{R}^n$ be an open set and let $f:\Omega\to\mathbb{R}^n$ be a K-Lipschitz continuous map, for some $K>0$. Then, for every $x^*\in\Omega$, the Clarke's Generalized Jacobian $\partial_x^C f(x^*)$ is a QDQ of f at x^* in the direction of Ω.*

We shall prove Proposition 2.7 after stating the following technical result, whose proof is fairly straightforward.

Proposition 2.8. *Let $x^*\in\mathbb{R}^m$, and let V and $F:V\rightsquigarrow\mathbb{R}^n$ be a compact neighborhood of x^* and a continuous set-valued map with closed values, respectively. Let $\Gamma\subseteq\mathbb{R}^m$ be a closed subset verifying $x^*\in\Gamma$, let $\rho:[0,1]\to[0,+\infty[$ be a modulus, and let $y^*\in F(x^*)$ and $\Lambda\subseteq\mathrm{Lin}(\mathbb{R}^m,\mathbb{R}^n)$. Let $(x_j,y_j,F^j,\Lambda_j,\rho_j)_{j\in\mathbb{N}}$ be a sequence such that:*

(i) for every $j\in\mathbb{N}$, $(x_j,y_j)_{j\in\mathbb{N}}\subset(\Gamma\cap V)\times\mathbb{R}^n$, $F_i:V\rightsquigarrow\mathbb{R}^n$ is a set-valued map, $\Lambda_j\in\mathrm{Lin}(\mathbb{R}^m,\mathbb{R}^n)$, and $\rho_j:[0,1]\to[0,+\infty[$ is a monotonic non-increasing function;

(ii)

$$\begin{aligned}&\lim_{j\to\infty}\sup_{x\in V} d^\sharp\left(F_j(x),F(x)\right)=0 \qquad \lim_{j\to\infty} d(y_j,y^*)=0\\ &\lim_{j\to\infty}\sup_{\alpha\in[0,1]}|\rho_j(\alpha)-\rho(\alpha)|=0 \qquad \lim_{j\to\infty} d^{\#}(\Lambda_j,\Lambda)=0;\end{aligned} \tag{2.3}$$

(iii) for every $\delta\in]0,\bar{\delta}]$, where $\bar{\delta}>0$ is a suitable positive number verifying $B(x^*,\bar{\delta})\subset V$, there exists a sequence $\Big(\Big(L_\delta^j,h_\delta^j\Big)\Big)_{j\in\mathbb{N}}\subset C^0\Big(\bar{x}+B_\delta,\mathrm{Lin}\{\mathbb{R}^m,\mathbb{R}^m\}\times\mathbb{R}^m\Big)$ and $\Big(L_\delta,h_\delta\Big)\in C^0\Big(\bar{x}+B_\delta,\mathrm{Lin}\{\mathbb{R}^m,\mathbb{R}^m\}\times\mathbb{R}^m\Big)$ such that

$$\Big(L_\delta^j,h_\delta^j\Big)\to\Big(L_\delta,h_\delta\Big)$$

uniformly, and, for every $j\in\mathbb{N}$ and $x\in(x_j+B_\delta)\cap\Gamma$, verifies

$$\begin{aligned}&y_j+L_\delta^j(x)\cdot(x-x_j)+h_\delta^j(x)\in F^j(x)\\ &\min\nolimits_{L'\in\Lambda_i}|L_\delta^j(x)-L'|\le\rho_j(\delta),\quad |h_\delta^j(x)|\le\delta\rho_j(\delta);\end{aligned} \tag{2.4}$$

Then Λ is a QDQ of F at (x^*,y^*) in the direction of Γ. More precisely, for every $\delta\in[0,\frac{\bar{\delta}}{2}]$, the map (L_δ,h_δ) verifies

$$\min_{L'\in\Lambda}|L_\delta(x)-L'|\le 2\rho(\delta),\quad |h_\delta(x)|\le 2\delta\rho(\delta),\quad \text{and } y^*+L_\delta(x)\cdot(x-x^*)+h_\delta(x)\in F(x), \tag{2.5}$$

whenever $x\in(x^*+B_\delta)\cap\Gamma$.

Proof of Proposition 2.7. Let us consider the standard mollifier $\eta(x) := \dfrac{1}{I_n}\exp\left(-\dfrac{1}{(1-|x|^2)}\right)\mathbf{1}_{B(0,1)}$ where I_n is the normalizing factor $I_n := \displaystyle\int_{B(0,1)}\exp\left(-\dfrac{1}{(1-|x|^2)}\right)\mathrm{d}x$, and, for every $\epsilon > 0$, let us set $\eta_\epsilon(x) := \dfrac{1}{\epsilon^n}\eta\left(\dfrac{x}{\epsilon}\right)$. For every $x \in \Omega$ and $\epsilon > 0$ so small that $B(0,\epsilon) \in \Omega$, let us consider the mollification

$$f^\epsilon(x) := \eta_\epsilon * f(x) = \int_{B(0,\epsilon)} f(x-y)\eta_\epsilon(y)\mathrm{d}y = \int_\Omega f(x-y)\eta_\epsilon(y)\mathrm{d}y.$$

In particular, for every $\epsilon > 0$ one has $\sup\limits_{x\in\Omega}|f^\epsilon(x) - f(x)| \le K\epsilon$. Moreover, for every $x \in \Omega$ and every $\epsilon < d(x,\partial\Omega)$, one has

$$Df^\epsilon(x) = \eta_\epsilon * Df(x) \in co\Big\{Df(y),\ \ y \in \mathrm{Diff}(f), y \in B(x,\epsilon)\Big\}. \tag{2.6}$$

Let us consider the sequence of (single-valued) maps $(F^j)_{j\in\mathbb{N}}$ defined by setting $F^j := f^{\frac{1}{j^2}}$ for every $j \in \mathbb{N}$. Since the maps F^j are smooth, for every $j \in \mathbb{N}$ such that $d(x^*,\partial\Omega) > \dfrac{1}{j} + \dfrac{1}{j^2}$ and every $x \in B\left(x^*, \dfrac{1}{j}\right)$, one has

$$F^j(x) = F^j(x^*) + L^j(x)\cdot(x - x^*), \tag{2.7}$$

where we have set

$$L^i(x) := \int_0^1 DF^j(x^* + t(x-x^*))\mathrm{d}t = \int_0^1\left(\int_\Omega \eta_{\frac{1}{j^2}}(y)DF(x^* + t(x-x^*) - y)\mathrm{d}y\right)\mathrm{d}t.$$

Aiming to apply Proposition 2.8, let us set, for every $j \in \mathbb{N}$ such that $d(x^*,\partial\Omega) > \dfrac{1}{j} + \dfrac{1}{j^2}$ and every $\delta > 0$,

$$\Lambda^j := co\left\{Df(y),\ \ y \in Diff(f), y \in B\left(x^*, \frac{1}{j} + \frac{1}{j^2}\right)\right\} \qquad \Lambda := \partial_x^C f(x^*),$$

$$y_j := F^j(x^*), \quad L_\delta^j := L^{\left[\frac{1}{\delta}\right]}$$

$$h_\delta^j(x) := \left(F^j(x) - F^{\left[\frac{1}{\delta}\right]}(x)\right) - \left(F^j(x^*) - F^{\left[\frac{1}{\delta}\right]}(x^*)\right) = \left(L^j(x) - L^{\left[\frac{1}{\delta}\right]}(x)\right)\cdot(x-x^*).$$

Observe that, for every $j \in \mathbb{N}$ and $x \in B\left(x^*, \dfrac{1}{j}\right)$, we have $L^j(x) \in \Lambda^j$. Setting

$$\rho_j(\delta) := \frac{1}{\delta}\sup_{x\in B(x^*,\delta)}\left|\left(L^j(x) - L^{\left[\frac{1}{\delta}\right]}(x)\right)\cdot(x-x^*)\right|$$

one gets (that ρ_j is monotonically decreasing and) $\rho_j(\delta) \le 2K\left(\delta + \frac{1}{\delta j^2}\right)$ for every $j \in \mathbb{N}$ and $\delta > 0$, so that the pointwise limit $\rho(\delta) = \lim\limits_{k\to\infty}\rho_j(\delta)$ is a modulus, *i.e.* $\rho(\delta)$ (is monotonically deceasing and) verifies $\lim\limits_{\delta\to 0}\rho(\delta) = 0$.

Since, for every $x \in B(x^*, \delta)$,

$$F^j(x) = y_j + L^j_\delta(x) \cdot (x - x^*) + h^j_\delta(x),$$

$$\min_{L' \in \Lambda_i} |L^j_\delta(x) - L'| \le |L^j_\delta(x) - L^j| \le \rho_j(\delta) \qquad h^j_\delta(x) \le 2\left(\frac{1}{j^2} + \delta^2\right) = \rho_j(\delta)\delta,$$

in view of Proposition 2.8 the proof is concluded provided $\lim_{j\to\infty} d^{\#}(\Lambda_j, \Lambda) = 0$.
Actually, this is easily verified: indeed, if there existed $a > 0$ such that there were $M^j \in \Lambda^j$ such that $d(M^j, \partial^C_x f(x^*)) > a$ for all $j \in \mathbb{N}$, there would be a subsequence $(M^{j_k})_{k\in\mathbb{N}}$ of $(M^j)_{j\in\mathbb{N}}$ converging to an element $M \in \mathrm{Lin}(\mathbb{R}^m, \mathbb{R}^n)$.[4] This would imply that $M \in \partial^C_x f(x^*)$, so contradicting the inequality $d(M, \partial^C_x f(x^*)) \ge a (> 0)$. □

The following results says that, for every vector field measurable in time and Lipschitz in space, the set-valued image at t of the corresponding variational differential inclusion is a QDQ of the flow map, for any t in the interval of existence.

Lemma 2.9. *Let $S > 0$ and $F : \mathbb{R} \times \mathbb{R}^n \to \mathbb{R}^n$ be a vector field such that, for every $s \in \mathbb{R}$, $F(s, \cdot)$ is Lipschitz with Lipschitz constant $L_F(s)$ satisfying $\int_0^S L_F(s)\,\mathrm{d}s = A < \infty$, and, for every $y \in \mathbb{R}^n$, $F(\cdot, y)$ is a bounded Lebesgue measurable map. Let $q \in \mathbb{R}^n$, and let us assume that the Cauchy problem*

$$\begin{cases} y'(s) = F(s, y(s)) \\ y(0) = \xi \end{cases}$$

has a (necessarily unique) solution $s \mapsto \Phi^F_s(\xi)$ on an interval $[0, S]$ for every ξ in a neighborhood U of q. Then, for every $s \in [0, S]$ the set,

$$\Lambda = \left\{ L(s),\ L \text{ is a solution of } L'(\sigma) \in \partial^C_y F(\sigma, y(\sigma)) \cdot L(\sigma),\ L(0) = \mathbf{1} \right\}$$

is a QDQ of the map $\xi \mapsto \Phi^F_s(\xi)$ at q in the direction of $\mathbb{R}^n$.

Proof. Let $\eta : \mathbb{R}^n \to \mathbb{R}_+$ be the standard mollifier as in Proposition 2.7, and for any $\sigma > 0$, let us set $\eta_\sigma(y) := \frac{1}{\sigma^n}\eta\left(\frac{y}{\sigma}\right)$. For every $t \in \mathbb{R}$, let us consider the convolution $\mathbb{R}^n \ni y \mapsto F_\sigma(t, y) := \displaystyle\int_{\mathbb{R}^n} F(t, y - h)\eta_\sigma(h)\mathrm{d}h$. Since the vector field F_σ is C^∞ with respect to y and measurable with respect to t, there exists a unique solution $y_\sigma(s)$ to the Cauchy problem on $[0, S]$

$$\begin{cases} y'(s) = F_\sigma(s, y(s)) \\ y(0) = \xi. \end{cases}$$

It is easy to check that $y_\sigma(s)$ uniformly converges to $y(s)$. Indeed,

$$\begin{aligned} |y_\sigma(s) - y(s)| &\le \int_0^s |F_\sigma(\tau, y_\sigma(\tau)) - F(\tau, y(\tau))|\,\mathrm{d}\tau \\ &\le \int_0^s |F_\sigma(\tau, y_\sigma(\tau)) - F(\tau, y_\sigma(\tau))|\,\mathrm{d}\tau + \int_0^s |F(\tau, y_\sigma(\tau)) - F(\tau, y(\tau))|\,\mathrm{d}\tau \\ &\le \int_0^s \int_{\mathbb{R}^n} |F(\tau, y_\sigma(\tau) - h) - F(\tau, y_\sigma(\tau))|\,\eta_\sigma(h)\,\mathrm{d}h\,\mathrm{d}\tau + \int_0^s L_F(\tau)|y_\sigma(\tau) - y(\tau)|\,\mathrm{d}\tau \end{aligned}$$

[4] Let us point out that for all $j \in \mathbb{N}$ and all $L \in \Lambda^j$, one has $|L| \le K$, where K is the Lipschitz constant of f.

$$\leq \int_0^s L_F(\tau)\,\mathrm{d}\tau \cdot \int_{\mathbb{R}^n} h\eta_\sigma(h)\,\mathrm{d}h + \int_0^s L_F(\tau)|y_\sigma(\tau)-y(\tau)|\,\mathrm{d}\tau \leq A\sigma + \int_0^s L_F(\tau)|y_\sigma(\tau)-y(\tau)|\,\mathrm{d}\tau,$$

so that, using Gronwall's Lemma, we get

$$|y_\sigma(s)-y(s)| \leq A\sigma + A\sigma\int_0^S L_F(\tau)e^{\int_0^S L_F(r)\,dr}\,\mathrm{d}\tau \leq (A+A^2e^A)\sigma.$$

In addition, by classic theory of ODE's we have, for every $\varepsilon \in \mathbb{R}^n$,

$$\Phi_s^{F_\sigma}(q+\varepsilon) = \Phi_s^{F_\sigma}(q) + M_\sigma(s)\cdot\varepsilon + o(|\varepsilon|), \tag{2.8}$$

where the matrix-valued map M_σ is the solution of the variational Cauchy problem

$$\begin{cases} M'(s) = \dfrac{\partial F_\sigma}{\partial y}(s, y_\sigma(s))\cdot M(s), \\ M(0) = \mathbf{1}. \end{cases}$$

Now, since $F(s,\cdot)$ is Lipschitz, the spatial gradient of the mollified function F_σ coincides with the mollification of the (L^∞) gradient $\dfrac{\partial F}{\partial y}(s,\cdot)$, *i.e.* $\dfrac{\partial F_\sigma}{\partial y}(s,y) = \displaystyle\int_{\mathbb{R}^n}\dfrac{\partial F}{\partial y}(s,y-h)\eta_\sigma(h)\,\mathrm{d}h$. Therefore, with computations similar to those performed above, it follows that

$$d(M_\sigma(S),\Lambda)\to 0 \qquad \text{as} \qquad \sigma\to 0.$$

Finally, if $\lim_{\varepsilon\to 0}\sigma(|\varepsilon|)=0$, we obtain

$$\Phi_S^F(q+\varepsilon) = \Phi_S^{F_\sigma}(q+\varepsilon)+o(|\varepsilon|) = \Phi_S^{F_\sigma}(q)+M_\sigma(S)\cdot\varepsilon+o(|\varepsilon|) = \Phi_S^F(q)+M_\sigma(S)\cdot\varepsilon+o(|\varepsilon|) \tag{2.9}$$

with $M_\sigma(S)$ having vanishing distance from Λ as $|\varepsilon|\to 0$. From this, our thesis follows according to Remark 2.2. □

2.3. *QDQ*-approximating cones and multicones

Let V be a finite-dimensional real vector space. A subset $C\subseteq V$ is called a *cone* if $\alpha v\in C, \forall\alpha\geq 0$ and $\forall v\in C$. A family $\mathscr{C}$ whose elements are cones is called a *multicone*. A *convex multicone* is a multicone whose elements are convex cones.

For any given subset $E\subseteq V$, the set $E^\perp := \{v\in V^*,\ v\cdot c\leq 0\ \forall c\in E\}\subseteq V^*$ is a closed cone, called the *polar cone of* E.

Let us introduce the notion of transversality, according to [27]. Two cones C_1 and C_2 are said to be *transversal* if $C_1 - C_2 = V$, where we use the notation $C_1-C_2 := \big\{c_1-c_2, \quad (c_1,c_2)\in C_1\times C_2\big\}$. Two multicones $\mathscr{C}_1$ and $\mathscr{C}_2$ called *transversal* if $C_1\in\mathscr{C}_1$ and $C_2\in\mathscr{C}_2$ are transversal as soon as$(C_1,C_2)\in\mathscr{C}_1\times\mathscr{C}_2$.

Two transversal cones C_1 and C_2 are called *strongly transversal* if $C_1\cap C_2\supsetneq\{0\}$. This is trivially equivalent to the existence of a non-zero linear form μ and an element $c\in C_1\cap C_2$ such that $\mu c>0$. More generally, we say that two transversal multicones $\mathscr{C}_1$, $\mathscr{C}_2$ are *strongly transversal* if there exists a non-zero linear form μ such that for any choice of cones $C_i\in\mathscr{C}_i$, $i=1,2$, there is an element $c\in C_1\cap C_2$ such that $\mu c>0$. One says that two cones C_1, C_2 are *linearly separated if* $C_1^\perp\cap -C_2^\perp\supsetneq\{0\}$, namely there exists a form $\mu\in V^*\backslash\{0\}$ such that $\mu c_1\geq 0, \mu c_2\leq 0$ for all $(c_1,c_2)\in C_1\times C_2$. It is easy to check that C_1 and C_2 are linearly separated if and only if they are not transversal. For multicones one has the following fact:

Lemma 2.10. *[27] Let $\mathscr{C}_1$ and $\mathscr{C}_2$ be two multicones that are not strongly transversal. If there is a linear functional μ that is in $C_2^\perp$ but not in $-C_2^\perp$ for all $C_2 \in \mathscr{C}_2$, then there are two cones $C_1 \in \mathscr{C}_1$ and $C_2 \in \mathscr{C}_2$ that are not transversal, i.e. C_1, C_2 are linearly separated.*

Definition 2.11 (QDQ-approximating multicones)**.** Let E be any subset of an Euclidean space $\mathbb{R}^n$ and $x \in E$[5]. A convex multicone $\mathscr{C}$ is said to be a *QDQ-approximating multicone to E at x* if there exists a set-valued map $F : \mathbb{R}^N \rightsquigarrow \mathbb{R}^n$, a convex cone $\Gamma \subset \mathbb{R}^N$, and a QDQ Λ of F at $(0, x)$ in the direction of Γ such that

$$F(\Gamma) \subseteq E, \qquad \mathscr{C} = \{L \cdot \Gamma,\ L \in \Lambda\}.$$

When Λ is a singleton, *i.e.* $\Lambda = \{L\}$, one simply says that $C := L \cdot \Gamma$ *is a QDQ-approximating cone to E at x.*

Definition 2.12 (Local separation of sets)**.** Two subsets E_1 and E_2 are *locally separated at x* if there exists a neighborhood U of x such that $E_1 \cap E_2 \cap U = \{x\}$.

As a consequence of an open mapping theorem, the following fact holds true (see Thm. 4.37, p. 265 in [1] where the lemma was proven in the more general context of $AGDQ$'s, of which QDQ are a special case)

Lemma 2.13. *If two subsets E_1 and E_2 are locally separated at x and if $\mathscr{C}_1$ and $\mathscr{C}_2$ are QDQ-approximating multicones for E_1 and E_2, respectively, at x, then $\mathscr{C}_1$ and $\mathscr{C}_2$ are not strongly transverse.*

3. The minimum problem and the main result

3.1. The minimum problem

The optimal control problem we are going to address, which will still label (P), is more general than the one presented in the Introduction, in that it involves a Lagrangiean l as well as an additional, bounded control a. Precisely we will consider the problem

$$\min_{u \in \mathscr{U}} \left(\Psi(T, x(T)) + \int_0^T l(x(t), u(t), a(t))\, \mathrm{d}t \right),$$

$$(P) \qquad \begin{cases} \dfrac{\mathrm{d}x}{\mathrm{d}t} = f(x,a) + \displaystyle\sum_{i=1}^m g_i(x) u^i, \qquad \text{a.e. } t \in [0,T], \\ \dfrac{d\nu}{\mathrm{d}t} = |u|, \\ (x, \nu) = (\hat{x}, 0), \end{cases} \qquad (T, x(T), \nu(T)) \in \mathfrak{T} \times [0, K]$$

where:

i) the state variable x belongs to $\mathbb{R}^n$, for some $n > 0$;
ii) the vector fields $g_i : \mathbb{R}^n \to \mathbb{R}^n$, $i = 1, \ldots, m$ are locally Lipschitz;
iii) the unbounded controls $u = (u^1, \ldots, u^m)$ take values in a closed cone $C = C_1 \times C_2$, where, for some non negative integers m_1 and m_2 such that $m = m_1 + m_2$, $C_1 \subseteq \mathbb{R}^{m_1}$ is a closed cone containing the coordinate axes, and $C_2 \subseteq \mathbb{R}^{m_2}$ is a closed cone which does not contain any straight line; the control a takes values in a compact set $A \subset \mathbb{R}^q$;
v) the *drift* f is continuous and, for any value of the control $a \in A$, the function $x \mapsto f(x, a) : \mathbb{R}^n \to \mathbb{R}^n$ is locally Lipschitz;
iv) the real-valued *Lagrangian* $l := l(x, u, a)$ has the form $l(x, u, a) = l_0(x, a) + l_1(x, u)$ and is continuous; furthermore, the map $x \mapsto l(x, u, a)$ is locally Lipschitz, uniformly for every $(u, a) \in C \times A$; moreover, the

[5] The definition of approximating multicone can be easily generalized to the case of subsets of a differential manifold [21].

so-called *recession function*

$$\hat{l}_1(x, w^0, w) := \lim_{r \to w^0} r l_1 \left(x, \frac{w}{r} \right) \tag{3.1}$$

is well-defined and locally Lipschitz with respect to x, uniformly as (w^0, w) ranges on the bounded set $[0,1] \times (C \cap B_1)$;

v) the *final cost* $\Psi(t, x)$ is Lipschitz, $0 \le K \le +\infty$, the (time-dependent) *target* $\mathfrak{T} \subseteq \mathbb{R}_+ \times \mathbb{R}^n$ is a closed subset, and the L^1 bound $\|u\|_1 \le K$ has been written in the equivalent form $\nu \le K$.;

iv) the minimization is performed over the set of the *strict sense feasible processes*, where by *strict sense process*[6] we mean a six-tuple (T, u, a, x, ν) such that (T, u, a) belongs to the family of *strict sense controls*

$$\mathscr{U} := \bigcup_{T > 0} \{T\} \times L^1([0, T], C \times A)$$

and (x, ν) is the solution of the above control system, whereas a strict sense process is called *feasible* as soon as $(T, x(T), \nu(T)) \in \mathfrak{T} \times [0, K]$.

Remark 3.1. All the involved objects having an intrinsic character, the optimal control problem and the corresponding results presented in this paper can be easily extended to the more general situation where the state x range over an n-dimensional manifold.

Remark 3.2. The assumption that the Lagrangian cost l, the drift term f, and the vector fields g_i are not time-varying can be removed by the standard procedure of regarding the time variable t as an extra-state variable x^0 subject to the trivial equation $\dot{x}^0 = 1$

Definition 3.3. We say that $(\overline{T}, \overline{u}, \overline{a}, \overline{x}, \overline{\nu})$ is a strict sense *weak local minimizer* for problem (P) if there exists $\delta > 0$ such that

$$\Psi(\overline{T}, \overline{x}(\overline{T})) + \int_0^{\overline{T}} l(\overline{x}(t), \overline{u}(t), \overline{a}(t))\, \mathrm{d}t \le \Psi(T, x(T)) + \int_0^T l(x(t), u(t), a(t))\, \mathrm{d}t$$

for all feasible processes (T, u, a, x, ν) such that $|T - \overline{T}| + \|(x, \nu) - (\overline{x}, \overline{\nu})\|_\infty + \|(u, a) - (\bar{u}, \bar{a})\|_1 < \delta$. Actually, since (x, ν, u, a) and $(\overline{x}, \overline{\nu}, \overline{u}, \overline{a})$ may have different domains, we tacitly extend (x, ν) and $(\overline{x}, \overline{\nu})$ continuously from $[0, T]$ and $[0, \overline{T}]$ to $\mathbb{R}_+$ so that they are constant on $[T, +\infty]$ and $[\overline{T}, +\infty]$. Furthermore, we extend (u, a) and $(\overline{u}, \overline{a})$ by setting $(u, a)(t) = (0, \hat{a})$ for any $t > T$ and $(\overline{u}, \overline{a})(t) = (0, \hat{a})$ for any $t > \overline{T}$, for some (irrelevant) choice of $\hat{a} \in A$.

3.2. The extended problem

Since we are interested in necessary conditions for minima, in principle we might ignore the existence problem. Yet, the unboundedness of the controls and the lack of adequate coercivity assumptions make the existence of an optimal control a quite unlikely, if not impossible. To get existence of *minima* it is then convenient to continuously embed the problem in a more general one where trajectories are somehow allowed to evolve also in a degenerate time interval (consisting of a single time instant). A distributional embedding being ruled out because of the non-linearity of the problem, a robust extension consists instead in first transforming the original problem in a time-space problem, where the trajectories are replaced by their graphs, and, secondly, considering the C^0-closure of the set of such graphs as the new minimization domain. Precisely we will consider the extended

[6] We use the expressions 'strict sense' in order to distinguish processes and controls of the original problem from those of the extended problem we introduce later, which will be named 'extended sense processes' and 'extended sense controls', respectively.

problem

$$\min_{(S,w^0,w,\alpha)\in\mathscr{W}} \left(\Psi(y^0(S), y(S)) + \int_0^S l^e((y, w^0, w, \alpha)(s))\, \mathrm{d}s \right),$$

$$(P_{ext}) \qquad \begin{cases} \dfrac{\mathrm{d}y^0}{\mathrm{d}s}(s) = w^0(s), \\ \dfrac{\mathrm{d}y}{\mathrm{d}s}(s) = f(y(s), \alpha(s))w^0(s) + \displaystyle\sum_{i=1}^m g_i(y(s))w^i(s), \\ \dfrac{d\beta}{\mathrm{d}s}(s) = |w(s)|, \\ (y^0, y, \beta)(0) = (0, \hat{x}, 0). \end{cases} \qquad (y^0(S), y(S), \beta(S)) \in \mathfrak{T} \times [0, K]$$

where:

i) $t = y^0 \in \mathbb{R}_+$ stands for the time parameter and $y \in \mathbb{R}^n$ denotes the state variable;

ii) the *extended Lagrangian* l^e is defined by setting

$$l^e(x, w^0, w, \alpha) := l_0(x, \alpha)w^0 + \hat{l}_1(x, w^0, w) \quad \forall (x, w^0, w, \alpha) \in \mathbb{R}^n \times \mathbb{R}_+ \times C \times A,$$

$\hat{l}_1$ being the recession function defined in (3.1).[7]

iii) the four-tuples (S, w^0, w, α) belong to the set

$$\mathscr{W} := \bigcup_{S>0} \{S\} \times \Big\{ (w^0, w, \alpha) \in L^\infty([0, S], \mathbb{R}_+ \times C \times A) : \mathrm{essinf}(w^0 + |w|) > 0 \Big\},$$

whose elements are called *extended sense controls.*;

iv) the minimization is performed over the set of *extended sense feasible processes*, which are defined as follows:
   - an *extended sense process* is a seven-tuple $(S, w^0, w, \alpha, y^0, y, \beta)$ such that (S, w^0, w, α) is an *extended sense control* and (y^0, y, β) is the corresponding solution of the extended Cauchy problem in $(P)_{ext}$, whereas
   - an extended sense process $(S, w^0, w, \alpha, y^0, y, \beta)$ is called *feasible* as soon as $(y^0(S), y(S), \beta(S)) \in \mathbb{R}_* \times \mathfrak{T} \times [0, K]$.

Let us give the notion local minimizer for the extended problem:

Definition 3.4. We say that $(\overline{S}, \overline{w}^0, \overline{w}, \overline{\alpha}, \overline{y}^0, \overline{y}, \overline{\beta})$ is a *weak local minimizer* for problem $(P)_{ext}$ if there exists $\delta > 0$ such that

$$\Psi(\overline{y}^0(\overline{S}), \overline{y}(\overline{S})) \ + \ \int_0^{\overline{S}} l^e(\overline{y}(s), \overline{w}^0(s), \overline{w}(s), \overline{\alpha}(s))\, \mathrm{d}s \ \le \ \Psi(y^0(S), y(S)) \ + \ \int_0^S l^e(y(s), w^0(s), w(s), \alpha(s))\, \mathrm{d}s$$

for all feasible processes $(S, w^0, w, \alpha, y^0, y, \beta)$ such that

$$|S - \overline{S}| + \|(y^0, y, \beta) - (\overline{y}^0, \overline{y}, \overline{\beta})\|_\infty + \|(w^0, w, \alpha) - (\overline{w}^0, \overline{w}, \overline{\alpha}))\|_1 < \delta. \quad [8]$$

[7] In view of the sublinearity of l in u, l^e is well-defined. As an example, one can consider the Lagrangian $l(x, u, a) = l_0(x, a) + \ell(x)|u|^r$ for some $r \in [0, 1]$ and some Lipschitz function ℓ, in which case one has $l^e(x, w^0, w, \alpha) = l_0(x, \alpha)w^0 + \ell(x)|w|^r(w^0)^{1-r}$.

[8] Since $(y^0, y, \beta, w^0, w, \alpha)$ and $(\overline{y}^0, \overline{y}, \overline{\beta}, \overline{w}^0, \overline{w}, \overline{\alpha})$ may have different domains, as before we tacitly extend them to $\mathbb{R}_+$ in the following way: the trajectories (y^0, y, β) and $(\overline{y}^0, \overline{y}, \overline{\beta})$ are prolonged to the whole $[0, +\infty[$ in such a way they are continuous and constant on $[S, +\infty[$ and $[\overline{S}, +\infty[$, respectively; as for the controls, the extensions consist in choosing a value $\hat{a} \in A$ and assigning the common constant value $(0, \mathbf{0}, \hat{\mathbf{a}})$ to both (w^0, w, α) and $(\overline{w}^0, \overline{w}, \hat{\alpha})$ on the intervals $]S, +\infty[$ and $]\overline{S}, +\infty[$, respectively.

There is an obvious one-to-one correspondence between strict sense processes and extended sense processes such that $w_0(s) > 0$ for almost any $s \in (0, S)$. More precisely, *A six-tuple* (T, u, a, x, ν) *is a strict sense process if and only if for every* $S > 0$ *and every strictly increasing, Lipschitz continuous, surjective function* $\tau : [0, S] \to [0, T]$, *the seven-tuple*

$$(S, w^0, w, \alpha, y^0, y, \beta)(s) := \Big(S, w^0(s), u(\tau(s))w^0(s), a(\tau(s)), \tau(s), x(\tau(s)), \nu(\tau(s))\Big) \qquad \forall s \in [0, S],$$
$$w^0(s) := \frac{\mathrm{d}\tau}{\mathrm{d}s}(s), \quad \textit{for a.e.} s \in [0, S]$$

is an extended sense process with $w^0 > 0$ *a.e. and, moreover,*

$$\Psi(T, x(T)) + \int_0^T l(x(t), u(t), a(t))\,\mathrm{d}t = \Psi(y^0(S), y(S)) + \int_0^S l^e(y(s), w^0(s), w(s), \alpha(s))\,\mathrm{d}s.$$

This one-to-one correspondence preserves feasibility of a process, and minima of the strict sense problem correspond to minima for the restriction of the space-time extended problem to processes having $w^0(s)$ almost everywhere positive (see [3]).

Remark 3.5. An important feature of the extended system is its *rate-independence*. By this we mean that if $\sigma : [0, \hat{S}] \to [0, S]$ is a bi-Lipschitz function, then $(S, w^0, w, \alpha, y^0, y, \beta)$ is a feasible process if and only if $\Big(\hat{S}, (w^0 \circ \sigma) \cdot \frac{\mathrm{d}\sigma}{\mathrm{d}s}, (w \circ \sigma) \cdot \frac{\mathrm{d}\sigma}{\mathrm{d}s}, \alpha \circ \sigma, y^0 \circ \sigma, y \circ \sigma, \beta \circ \sigma\Big)$ is a feasible process. Two processes obtained one from the other in this way are called *equivalent*. It is straightforward to verify they have the exact same associated costs, so that being a weak local minimizer is a property shared by equivalent processes. As it was observed in [3], this rate-independence implies the following fact:

- It is not restrictive to assume a minimizer to be *canonical*, meaning that $w^0(s) + |w(s)| = 1$ for almost every $s \in [0, \overline{S}]$.

To save space, for all $(y, w^0, w, a) \in \mathbb{R}^n \times \mathbb{R}_+ \times C \times A$ let us introduce the notation

$$F^e(y, w^0, w, a) := f(y, a)w^0 + \sum_{i=1}^m g_i(y)w^i$$

and

$$\mathscr{F}(y, w^0, w, a) := \big(w^0, F^e(y, w^0, w, a), l^e(y, w^0, w, a)\big)\,.$$

Let us define the Hamiltonian H by setting, for every $(y, p_0, p, \lambda, \pi, w^0, w, a) \in \mathbb{R}^n \times (\mathbb{R}^{1+n+1+1})^* \times W \times A$,

$$H(y, p_0, p, \lambda, \pi, w^0, w, a) := p_0 w^0 + pF^e(y, w^0, w, a) - \lambda l^e(y, w^0, w, a) + \pi|w|.$$

We are now in the position of stating our main result:

Theorem 3.6 (A "higher order" Maximum Principle). *Let* $(\overline{S}, \overline{w}^0, \overline{w}, \overline{\alpha}, \overline{y}^0, \overline{y}, \overline{\beta})$ *be a canonical local minimizer for the extended problem* (P_{ext}), *and let* $\boldsymbol{\mathscr{T}}$ *be a QDQ-approximating multicone to the target set* $\mathfrak{T}$ *at* $(\overline{y}^0, \overline{y})(\overline{S})$.

Then there exist multipliers $(p_0, p, \lambda, \pi) \in \mathbb{R}^* \times AC\Big([0, \overline{S}]; (\mathbb{R}^n)^*\Big) \times \mathbb{R}^* \times \mathbb{R}^*$ *such that* $\pi \le 0$ *(with* $\pi = 0$ *as soon as* $\|\overline{w}\|_1 < K$*),* $\lambda \ge 0$ *and the following conditions are satisfied:*

i) (NON TRIVIALITY) $(p_0, p, \lambda) \neq 0$;
ii) (ADJOINT DIFFERENTIAL INCLUSION)

$$\frac{dp}{\mathrm{d}s} \in -\partial_x^C H\left(\overline{y}, p_0, p, \lambda, \pi, \overline{w}^0, \overline{w}, \overline{\alpha}\right);$$

iii) (NON TRANSVERSALITY)

$$(p_0, p(\overline{S})) \in -\lambda \partial^C_{(t,x)} \Psi\Big((\overline{y}^0, \overline{y})(\overline{S})\Big) - \overline{\bigcup_{\mathcal{T} \in \mathscr{T}} \mathcal{T}^\perp} ;$$

iv) (FIRST ORDER MAXIMIZATION) *For almost all* $s \in [0, \overline{S}]$,

$$\max_{(w^0,w,a)\in\mathbb{R}_+\times C\times A} \Big[H(\overline{y}(s), p_0, p(s), \lambda, 0, w^0, w, a)\Big] = H(\overline{y}(s), p_0, p(s), \lambda, 0, \overline{w}^0(s), \overline{w}(s), \overline{\alpha}(s)).$$

v) (NONSMOOTH GOH CONDITION) *If, in addition,* $\|\overline{w}\|_1 < K$ *and* $\hat{l}_1(\cdot, 0) \equiv 0$, *then*

$$0 \in p(s)\,[g_i, g_j]_{set}(\overline{y}(s)) \qquad i, j \in \{1, \ldots, m_1\}, \quad \textit{for a.e. } t \in [0, \overline{T}]. \tag{3.2}$$

The proof of Theorem 3.6 will be given in Section 4.

Remark 3.7. Let us point out that if the Clarke's *tangent cone* to the target happens to be a QDQ-approximating cone as well, the set of conditions i)–iv) coincides with a (first order) non-smooth Pontryaging Maximum Principle of the kind one finds in several books (see *e.g.* [14, 29]) The same of course can be said for the smooth case where the target is a differential submanifold, in which case the tangent space to the target is automatically a QDQ-cone. Furthermore, in the smooth case, (3.2) coincides with the classical Goh condition.

4. Proof of the maximum principle

4.1. An equivalent fixed end-time problem

For every process $(S, w^0, w, \alpha, y^0, y, \beta)$, we will set

$$\overline{y}^l(s) := \int_0^s l^e(\overline{y}(\sigma), \overline{w}^0(\sigma), \overline{w}(\sigma), \overline{\alpha}(\sigma))\mathrm{d}\sigma$$

(so that $\overline{y}^l$ is the unique Carathéodory solution to the trivial differential equation $\frac{\mathrm{d}y^l}{\mathrm{d}s}(s) = l^e(\overline{y}(s), \overline{w}^0(s), \overline{w}(s), \overline{\alpha}(s))$ with initial condition $y^l(0) = 0$.)

Let us begin with a further (and standard) reparametrization procedure which allows us to reduce problem (P_{ext}) to a problem with a fixed end-time.

Let us fix $\bar{S} > 0$, $\rho > 0$. We say that $(\bar{S}, w^0, w, \alpha, \zeta, y^0, y, y^l, \beta)$ is a *rescaled space-time process* if

$$(\bar{S}, w^0, w, \alpha, \zeta)(\cdot) \in \mathscr{W} \times L^\infty([0, \overline{S}], [-\rho, \rho])$$

and $\big((y^0, y, y^l), \beta\big)$ is the unique (Carathéodory) solution of *the rescaled Cauchy problem*

$$\begin{cases} \dfrac{d}{\mathrm{d}s}\big((y^0,y,y^l),\beta\big) = \Big(\mathscr{F}(y,w^0,w,a),|w|\Big)\cdot\Big(1+\zeta\Big) \qquad s\in[0,\bar{S}] \\ \big((y^0,y,y^l),\beta\big)(0) = \big((0,\hat{x},0),0\big) \end{cases} \tag{4.1}$$

Moreover, we call $(\bar{S},w^0,w,\alpha,\zeta,y^0,y,y^l,\beta)$ *feasible* as soon as $\big((y^0,y),\beta\big)(\overline{S})\in\mathfrak{T}\times[0,K]$.
The *rescaled optimization problem* is defined as

$$\begin{cases} \min\left(\Psi((y^0,y)(\overline{S}))+y^l(\overline{S})\right), \\ \text{over feasible rescaled processes.} \end{cases} \tag{4.2}$$

With standard arguments one shows that, for small enough $\rho>0$, a process $(\overline{S},\overline{w}^0,\overline{w},\overline{\alpha},\overline{y}^0,\overline{y},\overline{\beta})$ is a local minimizer for the extended problem (P_{ext}) if and only if the rescaled space-time process $(\bar{S},\bar{w}^0,\bar{w},\bar{\alpha},0,\bar{y}^0,\bar{y},\bar{y}^l,\bar{\beta})$ is a local minimizer for fixed-end-time problem (4.2).[9]

Therefore, in the proof of the Maximum Problem we are allowed to *replace the hypothesis of the theorem with the following one:*

- *The process* $(\bar{S},\bar{w}^0,\bar{w},\bar{\alpha},\bar{\zeta}\equiv 0,\bar{y}^0,\bar{y},\bar{y}^l,\bar{\beta})$ *is a local minimizer of the rescaled problem* (4.2).

4.2. Set separation

For some $\delta>0$, let us consider the δ*-reachable set*

$$\mathfrak{R}_\delta := \left\{ \begin{array}{l} \Big(y^0,y,y^l+\Psi(y^0,y),\beta\Big)(\overline{S}):\ \ (\overline{S},w^0,w,\alpha,\zeta,y^0,y,y^l,\beta) \text{ is a rescaled} \\ \text{process that verifies } \|(y^0-\overline{y}^0,y-\overline{y},y^l-\overline{y}^l,\beta-\overline{\beta})\|_\infty<\delta \end{array} \right\} \subseteq \mathbb{R}^{1+n+1+1}$$

and the *projected* δ*-reachable set*

$$\mathfrak{R}'_\delta := \mathfrak{pr}\Big(\mathfrak{R}_\delta\Big) \subseteq\ \mathbb{R}^{1+n+1},$$

where the projection operator $\mathfrak{pr}$ is defined by setting $\mathfrak{pr}(x^0,x,x^l,\beta):=(x^0,x,x^l)$, for all $(x^0,x,x^l,\beta)\in\mathbb{R}^{1+n+1+1}$. Let us introduce also the *profitable set*

$$\mathfrak{P} := \left(\Big(\mathfrak{T}\times\Big]-\infty,\overline{y}^l(\overline{S})+\overline{\Psi}(\overline{S})\Big[\Big)\bigcup\Big\{\big(\overline{y}^0,\overline{y},\overline{y}^l(\overline{S})+\overline{\Psi}(\overline{S})\big)\Big\}\right)\times[0,K]$$

where $\overline{\Psi}(s):=\Psi(\overline{y}^0(s),\overline{y}(s))$ for all s, and the **projected profitable set**

$$\mathfrak{P}' := \mathfrak{pr}\Big(\mathfrak{P}\Big) = \left(\mathfrak{T}\times\Big]-\infty,\overline{y}^l(\overline{S})+\overline{\Psi}(\overline{S})\Big[\right)\bigcup\Big\{\big(\overline{y}^0,\overline{y},\overline{y}^l(\overline{S})+\overline{\Psi}(\overline{S})\big)\Big\}$$

Lemma 4.1. *Let* $(\overline{S},\overline{w}^0,\overline{w},\overline{\alpha},\overline{y}^0,\overline{y},\overline{\beta})$ *as in Theorem 3.6, and let us assume that* $\overline{\beta}(\overline{S})<K$*. Then for any* $\delta>0$ *sufficiently small, the projected profitable set* $\mathfrak{P}'$ *and the projected* δ*-reachable set* $\mathfrak{R}'_\delta$ *are locally separated at* $\big(\overline{y}^0,\overline{y},\overline{y}^l(\overline{S})+\overline{\Psi}(\overline{S})\big)$.

[9] Actually, the role of the auxiliary parameter ζ is fictitious, because of the rate-independence of problem (P_{ext}). However we use it, for it makes proofs simpler.

Proof. Indeed, by the definition of local minimizer it follows that the profitable set $\mathfrak{P}$ and the δ-reachable set $\mathfrak{R}_\delta$ are locally separated at $\big(\overline{y}^0, \overline{y}, \overline{y}^l(\overline{S}) + \overline{\Psi}(\overline{S}), \overline{\beta}(\overline{S})\big)$. From this one gets the thesis trivially. $\square$

4.3. Finitely many variations

With the ultimate aim of applying a suitable separability criterion for approximating cones, we now build a QDQ-approximating multicone to the projected δ-reachable set $\mathfrak{R}'_\delta$ at $\big(\overline{y}^0, \overline{y}, \overline{y}^l(\overline{S}) + \overline{\Psi}(\overline{S})\big)$. Let us define the set $\mathfrak{V}$ of *variation generators* as the union $\mathfrak{V} := \mathfrak{V}_{ndl} \bigcup \mathfrak{V}_{brk}$, where $\mathfrak{V}_{ndl}$ and $\mathfrak{V}_{brk}$ are the sets of *needle variation generators* and of *bracket-like variation generators* defined as $\mathfrak{V}_{ndl} := \mathbb{R}_+ \times C \times A \times [-\rho, \rho]$ and $\mathfrak{V}_{brk} := \Bigg[\{1, \dots, m_1\}^2 \setminus \mathrm{diag}\Big(\{1, \dots, m_1\}^2\Big)\Bigg]$, respectively.

Definition 4.2. Let $(0, \overline{S})_{Leb} \subset [0, \overline{S}]$ be the set of Lebesgue points of the function $s \mapsto (\overline{w}^0(s), \overline{F}^e(s), \overline{l}^e(s), |\overline{w}|(s))$, where $\overline{F}^e$ and $\overline{l}^e$ denote the functions F^e and l^e evaluated along the optimal process $(\bar{S}, \bar{w}^0, \bar{w}, \bar{\alpha}, \bar{\zeta} \equiv 0, \bar{y}^0, \bar{y}, \bar{y}^l, \bar{\beta})$ of the rescaled problem. For every variation generator $\mathbf{c} \in \mathfrak{V}$, let us define the variation vector $\big(v^0_{\mathbf{c},\overline{s}}, v_{\mathbf{c},\overline{s}}, v^l_{\mathbf{c},\overline{s}}\big)$ at an instant $\overline{s}$ as follows:[10]

$$\big(v^0_{\mathbf{c},\overline{s}}, v_{\mathbf{c},\overline{s}}, v^l_{\mathbf{c},\overline{s}}\big) := \begin{cases} \left\{ \begin{pmatrix} w^0(1+\zeta) - \overline{w}^0(\overline{s}) \\ F^e(\overline{y}(\overline{s}), w^0, w, a)(1+\zeta) - \overline{F}^e(\overline{s}) \\ l^e(\overline{y}(\overline{s}), w^0, w, a)(1+\zeta) - \overline{l}^e(\overline{s}) \end{pmatrix} \right\} & \text{if } \mathbf{c} = (w^0, w, a, \zeta) \in \mathfrak{V}_{ndl}, \\ & \text{and } \overline{s} \in (0, \overline{S})_{Leb} \\ \{0\} \times [g_i, g_j]_{set}(\overline{y}(\overline{s})) \times \{0\} & \text{if } \mathbf{c} = (i, j) \in \mathfrak{V}_{brk} \\ & \text{and } \overline{s} \in (0, \overline{S}). \end{cases}$$

Moreover, when $\mathbf{c} = (w^0, w, a, \zeta) \in \mathfrak{V}_{ndl}$, and $\overline{s} \in (0, \overline{S})_{Leb}$, we set $v^\nu_{\mathbf{c},\overline{s}} := |w|(1+\zeta) - |\overline{w}(\overline{s})|$.

Let us point out that, to retain uniformity of notation, we always regard $\big(v^0_{\mathbf{c},\overline{s}}, v_{\mathbf{c},\overline{s}}, v^l_{\mathbf{c},\overline{s}}\big)$ as a *subset of vectors of* $\mathbb{R}^{1+n+1}$, though, as soon as $\mathbf{c} \in \mathfrak{V}_{ndl}$, it reduces to the singleton formed by the usual needle variation vector.

Definition 4.3. Let us fix a rescaled control $\mathbf{w} = (w^0, w, \alpha, \zeta) \in L^\infty\big([0, \overline{S}], \mathbb{R}_+ \times C \times A \times [-\rho, \rho]\big)$ (with $\mathrm{essinf}(w^0 + |w|) > 0$) and an instant $\overline{s} \in (0, \overline{S})$.

- If $\mathbf{c} = (\hat{w}^0, \hat{w}, \hat{a}, \hat{\zeta}) \in \mathfrak{V}_{ndl}$, we call *needle control variation of* $\mathbf{w}$ *at* $\overline{s}$ *associated to* $\mathbf{c}$ the family of controls $\big\{\mathbf{w}_{\varepsilon,\mathbf{c},\overline{s}}(s) : \varepsilon \in [0, \overline{s})\big\}$ defined as

$$\mathbf{w}_{\varepsilon,\mathbf{c},\overline{s}}(s) = \begin{cases} \mathbf{w}(s) & \text{if } s \in [0, \overline{s} - \varepsilon) \cup (\overline{s}, \overline{S}] \\ (\hat{w}^0, \hat{w}, \hat{a}, \hat{\zeta}) & \text{if } s \in [\overline{s} - \varepsilon, \overline{s}]. \end{cases}$$

[10] As in the standard maximum principle, the fact of not considering pairs $(\mathbf{c}, \overline{s}) \in \mathfrak{V}_{ndl} \times \Big((0, \overline{S}) \backslash (0, \overline{S})_{Leb}\Big)$ is completely irrelevant, in that $(0, \overline{S}) \backslash (0, \overline{S})_{Leb})$ has zero measure.

- If $\mathbf{c} = (i,j) \in \mathfrak{V}_{brk}$, we call *bracket-like variation of* $\mathbf{w}$ *at* $\bar{s}$ the family $\Big\{\mathbf{w}_{\varepsilon,\mathbf{c},\bar{s}}(s) : 0 < 8\sqrt{\varepsilon} \le \bar{s}\Big\}$ of controls defined as

$$\mathbf{w}_{\varepsilon,\mathbf{c},\bar{s}}(s) = \begin{cases} \mathbf{w}(s) & \text{if } s \notin [\bar{s} - 8\sqrt{\varepsilon}, \bar{s}] \\ (2w^0, 2w, \alpha, \zeta) \circ \gamma^\varepsilon(s) & \text{if } s \in [\bar{s} - 8\sqrt{\varepsilon}, \bar{s} - 4\sqrt{\varepsilon}] \\ (0, \mathbf{e}_i, a, 0) & \text{if } s \in [\bar{s} - 4\sqrt{\varepsilon}, \bar{s} - 3\sqrt{\varepsilon}] \\ (0, \mathbf{e}_j, a, 0) & \text{if } s \in [\bar{s} - 3\sqrt{\varepsilon}, \bar{s} - 2\sqrt{\varepsilon}] \\ (0, -\mathbf{e}_i, a, 0) & \text{if } s \in [\bar{s} - 2\sqrt{\varepsilon}, \bar{s} - \sqrt{\varepsilon}] \\ (0, -\mathbf{e}_j, a, 0) & \text{if } s \in [\bar{s} - \sqrt{\varepsilon}, \bar{s}], \end{cases}$$

where $a \in A$ is arbitrarily chosen[11] and $\gamma^\varepsilon(s) := 2s - \bar{s} + 8\sqrt{\varepsilon}$,

4.4. *QDQ*-approximating cones to $\mathfrak{R}'_\delta$

For any $(\bar{s}, \mathbf{c}) \in [0, S] \times \mathfrak{V}$ and any ε sufficiently small, consider the functional $\mathscr{A}_{\varepsilon,\mathbf{c},\bar{s}}$ (from the space of rescaled controls $\mathbf{w}$ into itself) defined by setting $\mathscr{A}_{\varepsilon,\mathbf{c},\bar{s}}(\mathbf{w}) := \mathbf{w}_{\varepsilon,\mathbf{c},\bar{s}}$. In addition, given N variation generators $\mathbf{c}_1, \dots, \mathbf{c}_N \in \mathfrak{V}$ and N instants $0 < s_1 < s_2 < \dots s_N \le \overline{S}$ for a $\tilde{\varepsilon} > 0$ sufficiently small, let us define the multiple variation

$$[0, \tilde{\varepsilon}]^N \ni \boldsymbol{\varepsilon} \mapsto \overline{\mathbf{w}}_{\boldsymbol{\varepsilon}} := \mathscr{A}_{\varepsilon_N,\mathbf{c}_N,s_N} \circ \dots \circ \mathscr{A}_{\varepsilon_1,\mathbf{c}_1,s_1}(\overline{\mathbf{w}}).$$

Let us set $(\overline{w}^0_{\boldsymbol{\varepsilon}}, \overline{w}_{\boldsymbol{\varepsilon}}, \overline{a}_{\boldsymbol{\varepsilon}}, \overline{\zeta}_{\boldsymbol{\varepsilon}}) := \overline{\mathbf{w}}_{\boldsymbol{\varepsilon}}$, and let us use $(y^0_{\boldsymbol{\varepsilon}}, y_{\boldsymbol{\varepsilon}}, y^l_{\boldsymbol{\varepsilon}}, \beta_{\boldsymbol{\varepsilon}})$ to denote the solution (on $[0, \overline{S}]$) of the Cauchy problem[12]

$$\begin{cases} \dfrac{d}{\mathrm{d}s}(y^0, y, y^l, \beta) = \left(\mathscr{F}\Big(y, \overline{w}^0_{\boldsymbol{\varepsilon}}, \overline{w}_{\boldsymbol{\varepsilon}}, \overline{a}_{\boldsymbol{\varepsilon}}\Big), |w_{\boldsymbol{\varepsilon}}|\right)(1 + \overline{\zeta}_{\boldsymbol{\varepsilon}}) \\ (y^0, y, y^l, \beta)(0) = (0, \hat{x}, 0, 0) \end{cases} \tag{$P_{\varepsilon,\mathbf{c},\bar{s}}$}$$

Lemma 4.4. *The map* $\mathbf{Y} : \mathbb{R}^N_+ \to \mathbb{R}^q$ *defined by setting*

$$\mathbf{Y}(\boldsymbol{\varepsilon}) := \Big(y^0_{\boldsymbol{\varepsilon}}(\overline{S}),\, y_{\boldsymbol{\varepsilon}}(\overline{S}),\, y^l_{\boldsymbol{\varepsilon}}(\overline{S})\Big)$$

satisfies the hypothesis (2.2) *with* $F = \mathbf{Y}$ *and* $q := 1 + n + 1$*, namely, one has*

$$\mathbf{Y}(\boldsymbol{\varepsilon}) - \mathbf{Y}(0) = \sum_{i=1}^N \big(\mathbf{Y}(\varepsilon_i \mathbf{e}_i) - \mathbf{Y}(0)\big) + o(|\boldsymbol{\varepsilon}|), \qquad \forall \boldsymbol{\varepsilon} = (\varepsilon^1, \dots, \varepsilon^N) \in \mathbb{R}^N_+. \tag{4.3}$$

Proof. Let $\eta : \mathbb{R}^n \to \mathbb{R}_+$ and, for every $\delta > 0$, η_δ be a C^∞ mollifier as in Proposition 2.7. For any $\delta > 0$, let us define the mollified vector field

$$\mathscr{F}_\delta\Big(y, w^0, w, a\Big) := \int_{\mathbb{R}^n} \mathscr{F}\Big(y + h, w^0, w, a\Big)\eta_\delta(h)\,\mathrm{d}h.$$

[11] Since $w^0 = 0$, the choice of a is indeed irrelevant.

[12] Of course, $\overline{\mathbf{w}}_{\boldsymbol{\varepsilon}}$ and $(y_{\boldsymbol{\varepsilon}}, \beta_{\boldsymbol{\varepsilon}})$ depend also on the parameters $\mathbf{c}_k$ and s_k, but we avoid writing them when possible in order to simplify the notation.

Observe that the control vector field $\mathscr{F}$ is continuous, and, in addition, it is locally Lipschitz in the variable y. Moreover, we can apply a cut off technique and make $\mathscr{F}$ and Ψ equal to zero outside a compact set containing a small neighbourhood of our local minimizer, so that *we can assume that* $\mathscr{F}$ *and* Ψ *are globally Lipschitz as well.* It follows that $\mathscr{F}_\delta$ converges uniformly to $\mathscr{F}$ as δ goes to 0. For any fixed $\boldsymbol{\varepsilon} \in \mathbb{R}_+$ with a suitably small norm, let us introduce the mollified Cauchy problem

$$\begin{cases} \dfrac{d}{\mathrm{d}s}\big((y^0, y, y^l), \beta\big) = \Big(\mathscr{F}_\delta(y, w^0_{\boldsymbol{\varepsilon}}, w_{\boldsymbol{\varepsilon}}, a_{\boldsymbol{\varepsilon}}), |w|\Big) \cdot \Big(1 + \zeta_{\boldsymbol{\varepsilon}}\Big) \\ \\ \big((y^0, y, y^l), \beta\big)(0) = \big((0, \hat{x}, 0), 0\big) \end{cases} \tag{4.4}$$

and let us use $\left(y^0_{\delta,\boldsymbol{\varepsilon}}, y_{\delta,\boldsymbol{\varepsilon}}, y^l_{\delta,\boldsymbol{\varepsilon}}, \beta_{\boldsymbol{\varepsilon}}\right)$ to denote its unique solution.
We also set

$$\mathbf{Y}_\delta(\boldsymbol{\varepsilon}) := \Big(y^0_{\delta,\boldsymbol{\varepsilon}}(\overline{S})\ ,\ y_{\delta,\boldsymbol{\varepsilon}}(\overline{S})\ ,\ y^l_{\delta,\boldsymbol{\varepsilon}}(\overline{S}) + \Psi\left(y^0_{\delta,\boldsymbol{\varepsilon}}(\overline{S}), y_{\delta,\boldsymbol{\varepsilon}}(\overline{S})\right)\Big)$$

Let us define the function $z_{\delta,\boldsymbol{\varepsilon}}(s) := \Big|(y^0_{\delta,\boldsymbol{\varepsilon}}, y_{\delta,\boldsymbol{\varepsilon}}, y^l_{\delta,\boldsymbol{\varepsilon}})(s) - (y^0_{\boldsymbol{\varepsilon}}, y_{\boldsymbol{\varepsilon}}, y^l_{\boldsymbol{\varepsilon}})(s)\Big|$, $s \in [0, \bar{S}]$, and let us observe that, from the inequality

$$\begin{gathered} z_{\delta,\boldsymbol{\varepsilon}}(s) \leq \int_0^s \Big|\mathscr{F}_\delta(y_{\delta,\boldsymbol{\varepsilon}}, w^0_{\boldsymbol{\varepsilon}}, w_{\boldsymbol{\varepsilon}}, a_{\boldsymbol{\varepsilon}}) - \mathscr{F}(y_{\boldsymbol{\varepsilon}}, w^0_{\boldsymbol{\varepsilon}}, w_{\boldsymbol{\varepsilon}}, a_{\boldsymbol{\varepsilon}})\Big|\Big(1 + \zeta_{\boldsymbol{\varepsilon}}\Big)\mathrm{d}\sigma \\ \leq \int_0^s \Big|\mathscr{F}_\delta(y_{\delta,\boldsymbol{\varepsilon}}, w^0_{\boldsymbol{\varepsilon}}, w_{\boldsymbol{\varepsilon}}, a_{\boldsymbol{\varepsilon}}) - \mathscr{F}(y_{\delta,\boldsymbol{\varepsilon}}, w^0_{\boldsymbol{\varepsilon}}, w_{\boldsymbol{\varepsilon}}, a_{\boldsymbol{\varepsilon}})\Big|\Big(1 + \zeta_{\boldsymbol{\varepsilon}}\Big)\mathrm{d}\sigma + L(1 + 2\rho)\int_0^s z_{\delta,\boldsymbol{\varepsilon}}(\sigma) \quad [13] \\ \leq 2K(1 + 2\rho)\overline{S}\delta + L(1 + 2\rho)\textstyle\int_0^s z_{\delta,\boldsymbol{\varepsilon}}(\sigma)\mathrm{d}\sigma, \end{gathered}$$

and Gronwall's Lemma, we deduce

$$\big|(y^0_{\delta,\boldsymbol{\varepsilon}}, y_{\delta,\boldsymbol{\varepsilon}}, y^l_{\delta,\boldsymbol{\varepsilon}})(s) - (y^0_{\boldsymbol{\varepsilon}}, y_{\boldsymbol{\varepsilon}}, y^l_{\boldsymbol{\varepsilon}})(s)\big| = z_{\delta,\boldsymbol{\varepsilon}}(s) \leq C\delta, \quad \forall s \in [0, \bar{S}], \tag{4.5}$$

where C is a positive constant depending only on $\overline{S}$, K, and L. By choosing $\delta = \delta(|\boldsymbol{\varepsilon}|) = |\boldsymbol{\varepsilon}|^2$, we get $\mathbf{Y}(\boldsymbol{\varepsilon}) = \mathbf{Y}_{|\varepsilon^2|}(\boldsymbol{\varepsilon}) + o(|\boldsymbol{\varepsilon}|)$. We have reduced to a smooth system, hence the fact that a function like our $\mathbf{Y}_{|\varepsilon^2|}$ has the desired property (4.3) is something very well known in the literature. The thesis follows automatically as $\mathbf{Y}_{|\varepsilon^2|}(\varepsilon_i\mathbf{e}_i)$ is again distant at most $o(|\boldsymbol{\varepsilon}|)$ from $\mathbf{Y}(\varepsilon_i\mathbf{e}_i)$. As a matter of fact,

$$\begin{gathered} \mathbf{Y}(\boldsymbol{\varepsilon}) - \mathbf{Y}(0) = \big(\mathbf{Y}(\boldsymbol{\varepsilon}) - \mathbf{Y}_{|\boldsymbol{\varepsilon}|^2}(\boldsymbol{\varepsilon})\big) + \big(\mathbf{Y}_{|\boldsymbol{\varepsilon}|^2}(\boldsymbol{\varepsilon}) - \mathbf{Y}_{|\boldsymbol{\varepsilon}|^2}(0)\big) + \big(\mathbf{Y}_{|\boldsymbol{\varepsilon}|^2}(0) - \mathbf{Y}(0)\big) \\ = \sum_{i=1}^N \big(\mathbf{Y}_{|\boldsymbol{\varepsilon}|^2}(\varepsilon_i\mathbf{e}_i) - \mathbf{Y}_{|\boldsymbol{\varepsilon}|^2}(0)\big) + \Big(o(|\boldsymbol{\varepsilon}|) + \big(\mathbf{Y}(\boldsymbol{\varepsilon}) - \mathbf{Y}_{|\boldsymbol{\varepsilon}|^2}(\boldsymbol{\varepsilon})\big) + \big(\mathbf{Y}_{|\boldsymbol{\varepsilon}|^2}(0) - \mathbf{Y}(0)\big)\Big) \\ = \sum_{i=1}^N \big(\big(\mathbf{Y}(\varepsilon_i\mathbf{e}_i) - \mathbf{Y}(0)\big) + o(|\boldsymbol{\varepsilon}|)\big) \end{gathered}$$

□

Definition 4.5. Let N be a natural number, and let us choose N variation generators $\mathbf{c}_1, \ldots, \mathbf{c}_N \in \mathfrak{V}$ and N instants $0 < s_1 < s_2 < \ldots \leq s_N \leq \overline{S}$, with $s_k \in [0, \overline{S}]_{Leb}$ as soon as $\mathbf{c}_k \in \mathfrak{V}_{ndl}$. For every $k = 1, \ldots, N$,

[13] K is a bound for the maps $\mathscr{F}$ and $\mathscr{F}_\delta$ and L is a Lipschitz constant for the maps $(y^0, y, y^l) \mapsto \mathscr{F}(y^0, y, y^l, w^0, w, a)$, independent of (w^0, w, a)

any L^1-map $[0,\overline{S}] \ni s \mapsto (M,\omega)(s) \in \mathrm{Lin}(\mathbb{R}^n,\mathbb{R}^n)\times(\mathbb{R}^n)^*$, and any $(m_t,m_x)\in(\mathbb{R}^{n+1})^*$ let us consider the $(1+n+1)\times(1+n+1)$ matrix

$$\mathscr{E}_k'(m_t,m_x,M,\omega) := \begin{pmatrix} 1 & 0 & 0 \\ 0_{n\times 1} & e^{\int_{s_k}^{\overline{S}} M} & 0_{n\times 1} \\ m_t & m_x e^{\int_{s_k}^{\overline{S}} M} + \displaystyle\int_{s_k}^{\overline{S}} \left(\omega(s) e^{\int_s^{s_k} M}\right) \mathrm{d}s & 1 \end{pmatrix}$$

(which transports vectors from the tangent space at $(\overline{y}^0,\overline{y},\overline{y}^l)(s_k)$ to the tangent space at $(\overline{y}^0,\overline{y},\overline{y}^l)(\overline{S})$) and, in the special case when $\boldsymbol{c}_k \in \mathfrak{V}_{ndl}$ $\forall k\in\{1,\dots,N\}$, the $(1+n+1+1)\times(1+n+1+1)$ matrix

$$\mathscr{E}_k(m_t,m_x,M,\omega) := \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0_{n\times 1} & e^{\int_{s_k}^{\overline{S}} M} & 0_{n\times 1} & 0 \\ m_t & m_x e^{\int_{s_k}^{\overline{S}} M} + \displaystyle\int_{s_k}^{\overline{S}} \left(\omega(s) e^{\int_s^{s_k} M}\right) \mathrm{d}s & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

where the exponential of a matrix is defined as in Section 1.1 and $0_{n\times1}$ stands for a column of n zeros. Subsequently, let the subsets $\Lambda_N' \subset \mathrm{Lin}(\mathbb{R}^N,\mathbb{R}^{1+n+1})$ $\Lambda_N \subset \mathrm{Lin}(\mathbb{R}^N,\mathbb{R}^{1+n+1+1})$ be defined as

$$\Lambda_N' := \left\{ \left(\mathscr{E}_1'(m_t,m_x,M,\omega)\begin{pmatrix} V_1^0 \\ V_1 \\ V_1^l\end{pmatrix}, \dots, \mathscr{E}_N'(m_t,m_x,M,\omega)\begin{pmatrix} V_N^0 \\ V_N \\ V_N^l\end{pmatrix}\right), \quad \begin{array}{l} (M,\omega)(\cdot) \text{ is a measurable} \\ \text{selection of } \partial_y\big(\overline{F}^e,\overline{l}^e\big)(\cdot), \\ (m_t,m_x)\in\partial^C_{(t,x)}\Psi((\overline{y}^0,\overline{y})(\overline{S})) \\ (V_k^0,V_k,V_k^l)\in(v^0_{\mathbf{c}_k,s_k}, v_{\mathbf{c}_k,s_k}, v^l_{\mathbf{c}_k,s_k}) \\ \forall k=1,\dots,N \end{array} \right\},$$

$$\Lambda_N := \left\{ \left(\mathscr{E}_1(m_t,m_x,M,\omega)\begin{pmatrix} V_1^0 \\ V_1 \\ V_1^l \\ V_1^\nu\end{pmatrix}, \dots, \mathscr{E}_N(m_t,m_x,M,\omega)\begin{pmatrix} V_N^0 \\ V_N \\ V_N^l \\ V_N^\nu\end{pmatrix}\right), \quad \begin{array}{l} \text{where } (M,\omega)(\cdot), m_t, m_x \\ \text{and} (V_k^0,V_k,V_k^l) \\ \text{are as in the} \\ \text{definition of } \Lambda_N' \end{array} \right\}.$$

Corollary 4.7 below represents the most important technical step of the proof of our maximum principle. It is a straightforward consequence of the following result:

Theorem 4.6. *Let* $\left(\overline{y}^0,\overline{y},\overline{y}^l,\overline{\beta}\right)$ *and* $\left(y^0_{\boldsymbol{\varepsilon}},y_{\boldsymbol{\varepsilon}},y^l_{\boldsymbol{\varepsilon}},\beta_{\boldsymbol{\varepsilon}}\right)$ *as above. If we assume the extra assumption* $\hat{l}_1(\cdot,0,\cdot)\equiv 0$, *then the set* Λ_N' *is a QDQ at* $\mathbf{0}$ *of the map*

$$\mathbf{Z} := \left(y^0_{\boldsymbol{\varepsilon}}(\overline{S})\,,\, y_{\boldsymbol{\varepsilon}}(\overline{S})\,,\; y^l_{\boldsymbol{\varepsilon}}(\overline{S}) + \Psi\left(y^0_{\boldsymbol{\varepsilon}}(\overline{S}), y_{\boldsymbol{\varepsilon}}(\overline{S})\right)\right)$$

in the direction of $\mathbb{R}^N_+$.

Moreover, in the special case when $\mathbf{c}_k\in\mathfrak{V}_{ndl}$ *for all* $k\in\{1,\dots,N\}$ *(and* $\hat{l}_1(\cdot,0,)$ *is possibly non vanishing),* Λ_N *is a QDQ at* $\mathbf{0}$ *of the map* $\mathbb{R}^N_+\ni\boldsymbol{\varepsilon}\mapsto(\mathbf{Z}(\boldsymbol{\varepsilon}),\beta_{\boldsymbol{\varepsilon}})$ *(in the direction of* $\mathbb{R}^N_+$*).*

Corollary 4.7. *Let us use the same notations as in Theorem 4.6 and let us assume that $\hat{l}_1(\cdot,0,)\equiv 0$. For any choice of $\delta>0$, the set*

$$\Lambda'_N\mathbb{R}^N_+ := \left\{ L'\mathbb{R}^N_+ \,:\, L'\in\Lambda'_N \right\}$$

is a QDQ-approximating multicone of the projected δ-reachable set $\mathscr{R}'_\delta$ at $\left(\overline{y}^0,\overline{y},\overline{y}^l(\overline{S})+\overline{\Psi}(\overline{S})\right)$. Moreover, in the special case when $\mathbf{c}_k\in\mathfrak{V}_{ndl}$ for all $k\in\{1,\dots,N\}$ (and $\hat{l}_1(\cdot,0,)$ is possibly non vanishing), the set

$$\Lambda_N\mathbb{R}^N_+ := \left\{ L\mathbb{R}^N_+ \,:\, L\in\Lambda_N \right\}$$

is a QDQ-approximating multicone of the δ-reachable set $\mathscr{R}_\delta$ at $\left(\overline{y}^0,\overline{y},\overline{y}^l(\overline{S})+\overline{\Psi}(\overline{S}),\overline{\beta}(\overline{S})\right)$.

Proof of Theorem 4.6 . First of all, notice that $\mathbf{Z}$ is obtained by composition of the map $\mathbf{Y}$ we defined in Lemma 4.4 with the function $(t,x,c)\mapsto(t,x,c+\Psi(t,x))$. This means that in order to build a Quasi Differential Quotient of $\mathbf{Z}$ it is enough to build a QDQ of $\mathbf{Y}$. Indeed, thanks to Proposition 2.7, the Clarke's Generalized Jacobian of $(t,x,c)\mapsto(t,x,c+\Psi(t,x))$ is a also QDQ (in the direction of $\mathbb{R}^{1+n+1}$) of $(t,x,c)\mapsto(t,x,c+\Psi(t,x))$. In turn, the Clarke's Generalized Jacobian of $(t,x,c)\mapsto(t,x,c+\Psi(t,x))$ at $(\overline{S},\overline{y}(\overline{S}),\overline{y}^l(\overline{S}))$ is easily seen to coincide with

$$\left\{ \begin{pmatrix} 1 & \mathbf{0} & 0\\ \mathbf{0} & \mathbf{1} & \mathbf{0}\\ m_t & m_x & 1 \end{pmatrix}, (m_t,m_x)\in\partial^C_{(t,x)}\Psi(\overline{S},\overline{y}(\overline{S})) \right\}$$

(where $\partial^C_{(t,x)}\Psi$ is the Clarke's Generalized Jacobian of Ψ). For this reason, we turn our attention to $\mathbf{Y}$. Thanks to Lemma 4.4 and Proposition 2.5 (with $\mathbf{Y}$ in the role of F) we can construct a QDQ at zero of $\mathbf{Y}$ once we know, for every $i=1,\dots N$, a QDQ at zero of its restriction $\mathbf{Y}(\varepsilon_i\mathbf{e}_i)$ to the axis $\mathbb{R}\mathbf{e}_i$.

So, let us fix $i=1,\dots N$, and let us observe that, for any $\varepsilon_i>0$ sufficiently small, one has

$$\mathbf{Y}(\varepsilon_i\mathbf{e}_i)=\mathbf{\Phi}^{\overline{S}}_{s_i}\circ\mathbf{Y}_{s_i}(\varepsilon_i) \tag{4.6}$$

where

$$\mathbf{Y}_{s_i}(\varepsilon_i):=\left(y^0_{\varepsilon_1}(s_i),y_{\varepsilon_1}(s_i),y^l_{\varepsilon_1}(s_i)\right)$$

while, for every $q\in\mathbb{R}^{1+n}$, $[s_i,\overline{S}]\ni s\mapsto\Phi^s_{s_i}(q)$ denotes the solution to the Cauchy problem

$$\begin{cases} \dfrac{d}{\mathrm{d}s}(y^0,y,y^l)=\mathscr{F}\left(y,\overline{w}^0_{\varepsilon_i},\overline{w}_{\varepsilon_i},\overline{\alpha}_{\varepsilon_i}\right)\cdot\left(1+\zeta_{\varepsilon_i}\right)\\ (y^0,y,y^l)(s_i)=q, \end{cases} \tag{4.7}$$

Therefore we can again apply the chain rule for QDQs (Prop. 2.4) to the composed map $\Phi^{\overline{S}}_{s_i}\circ\mathbf{Y}_{s_i}(\varepsilon_i)$.

Let us begin with determining a QDQ of $(\varepsilon_i)\mapsto\mathbf{Y}_{s_i}$ at $\varepsilon_i=0$. We distinguish the case when $\mathbf{c_i}$ is a needle variation generator from the one in which $\mathbf{c_i}$ is a bracket-like variation generator.

- If $\mathbf{c_i}=(\hat{w}^0,\hat{w},\hat{a},\hat{\zeta})\in\mathfrak{V}_{ndl}$, standard arguments imply that $(v^0_{\mathbf{c}_i,s_i},v_{\mathbf{c}_i,s_i},v^l_{\mathbf{c}_i,s_i})$ is the right derivative at 0 of the path $\varepsilon_i\mapsto\mathbf{Y}_{s_i}(\varepsilon_i)$. Therefore, the singleton $\left\{(v^0_{\mathbf{c}_i,s_i},v_{\mathbf{c}_i,s_i},v^l_{\mathbf{c}_i,s_i})\right\}$ a QDQ at 0 of $\mathbf{Y}_{s_i}$ in the direction of $\mathbb{R}_+$.

- Instead, if $\mathbf{c_i} = (j,k) \in \mathfrak{V}_{brk}$, for some $j,k = 1,\dots,m_1$, by applying a result established in [2] which proves that set-valued Lie brackets are QDQs of commutator-like multiflows, we get again that the set $(v^0_{\mathbf{c}_i,s_i}, v_{\mathbf{c}_i,s_i}, v^l_{\mathbf{c}_i,s_i})$ a QDQ at 0 of $\mathbf{Y}_{s_i}$ in the direction of $\mathbb{R}_+$.

Finally, by invoking Lemma 2.9, we get that

$$\Big\{\mathscr{E}'_k(0,0,M,\omega),\ \text{where } (M,w) \text{ is a meas. selection of } \partial_y(\overline{F}^e, \bar{l}^e)\Big\}$$

is a QDQ of $\Phi^{\overline{S}}_{s_i}$ at $\mathbf{Y}_{s_i}(0)$ in the direction of $\mathbb{R}^n$.

The proof of this first part of the statement of the theorem follows by use of the chain rule for QDQ's, namely Proposition 2.4, for the reasons explained at the beginning and with the simple observation that

$$\begin{pmatrix} 1 & \mathbf{0} & 0 \\ \mathbf{0} & \mathbf{1} & \mathbf{0} \\ m_t & m_x & 1 \end{pmatrix} \mathscr{E}'_k(0,0,M,\omega) = \mathscr{E}'_k(m_t,m_x,M,\omega)$$

The proof of the part concerning the special case when $\mathbf{c}_r \in \mathfrak{V}_{ndl}$ for all $r \in 1,\dots,N$, only requires an extension of the analysis to the last component, $\beta_\varepsilon(S)$, observing that

$$\beta_\varepsilon(\overline{S}) - \overline{\beta}(\overline{S}) := \sum_{k=1}^N \int_{s_k-\varepsilon_k}^{s_k} |w_k|(1+\zeta_k) - |\overline{w}(\sigma)| \,\mathrm{d}\sigma.$$

So the proof of the theorem is concluded in view of the "set product rule" in Proposition 2.3. $\square$

4.5. Linear separability of approximating cones at the end-time

We will use the fact that $\Lambda'_N\mathbb{R}^N_+$ is a QDQ-approximating multicone of $\mathscr{R}'_\delta$ at $\big(\overline{y}^0, \overline{y}, \overline{y}^l(\overline{S}) + \overline{\Psi}(\overline{S})\big)$ to deduce a linear separability result at time $\overline{S}$.

Lemma 4.8. *Let $(\overline{S}, \overline{w}^0, \overline{w}, \overline{\alpha}, \overline{y}^0, \overline{y}, \overline{\beta})$ be a canonical local minimizer for the extended problem (P_{ext}). For some positive integer N, let Λ'_N be defined as in the former subsection, and assume that $\overline{\beta}(\overline{S}) < K$ as soon as $\mathbf{c}_k \in \mathfrak{V}_{brk}$ for some $k \in \{1,\dots,N\}$. Then:*

1. *for any QDQ-approximating multicone $\boldsymbol{\mathscr{T}}$ to the target $\mathfrak{T}$ at $\big(\overline{y}^0, \overline{y}\big)(\overline{S})$, there exist $L' \in \Lambda'_N$, $\mathscr{T} \in \boldsymbol{\mathscr{T}}$, and $(\xi_0, \xi, \xi_c) \in \Big(L'\mathbb{R}^N_+\Big)^\perp$ verifying $\xi_c \le 0$ and $(\xi_0,\xi) \in -\mathscr{T}^\perp$;*
2. *furthermore, if $c_k \in \mathfrak{V}_{ndl}$ for all $k \in \{1,\dots,N\}$, then the above linear form (ξ_0,ξ,ξ_c) can be chosen so that $(\xi_0,\xi,\xi_c,\pi) \in \Big(L\mathbb{R}^N_+\Big)^\perp$, for some $L \in \Lambda_N$ and $\pi \le 0$.*

Proof. By Lemma 4.1 we know that, for $\delta > 0$ sufficiently small, the projected profitable set $\mathfrak{P}'$ and the projected δ-reachable set $\mathfrak{R}'_\delta$ are locally separated. Moreover $\Big\{\mathscr{T} \times (-\infty,0) : \mathscr{T} \in \boldsymbol{\mathscr{T}}\Big\}$ is an QDQ-approximating multicone to the projected profitable set $\mathfrak{P}'$ at $\big(\overline{y}^0, \overline{y}, \overline{y}^l(\overline{S}) + \overline{\Psi}(\overline{S})\big)$. Hence, by Lemma 2.13, it follows that the QDQ-approximating multicones $\Big\{\mathscr{T} \times (-\infty,0) : \mathscr{T} \in \boldsymbol{\mathscr{T}}\Big\}$ and $\Lambda'_N\mathbb{R}^N_+$ are not strongly transverse. Now, since $\Big\{\mathscr{T} \times (-\infty,0) : \mathscr{T} \in \boldsymbol{\mathscr{T}}\Big\}$ is a multicone whose elements are contained in the semispace $\mathbb{R}^{1+n} \times \mathbb{R}_-$, by Lemma 2.10 we can infer the existence of $(\xi_0,\xi,\xi_c) \in (\mathbb{R} \times \mathbb{R}^n \times \mathbb{R})^* \backslash \{(0,0,0)\}$, $L \in \Lambda_N$, and $\mathscr{T} \in \boldsymbol{\mathscr{T}}$, such that

$$(\xi_0,\xi,\xi_c) \in \Big(L\mathbb{R}^N_+\Big)^\perp, \quad (\xi_0,\xi,\xi_c) \in -\Big(\mathscr{T} \times \mathbb{R}_-\Big)^\perp.$$

In particular, $\xi_c \le 0$ and $(\xi_0, \xi) \in -\mathscr{T}^\perp$, so that (1) is proved. The existence of a $\pi \le 0$ such that property (2) holds true comes from the same argument as soon as one considers the local separation of the profitable set $\mathfrak{P}$ and the δ-reachable set $\mathfrak{R}_\delta$ (in the augmented space $\mathbb{R}^{1+n+1+1}$). □

4.6. A maximum principle for finitely many variations

By using propagation due to the adjoint differential inclusion, as a direct consequence of Lemma 4.8 we get a *maximum principle for the instants* $s_1, \dots, s_N$ *and the variation generators* $\mathbf{c}_1, \dots, \mathbf{c}_N$*:*

Lemma 4.9. *Let* $(\overline{S}, \overline{w}^0, \overline{w}, \overline{\alpha}, \overline{y}^0, \overline{y}, \overline{\beta})$ *be a canonical local minimizer for the extended problem* (P_{ext})*, and let* $s_1, \dots, s_N \in (0, \overline{S})$*,* $\mathbf{c}_1, \dots, \mathbf{c}_N \in \mathfrak{V}$ *be as above, for some integer* $N > 0$*. Let* $\boldsymbol{\mathscr{T}}$ *be a QDQ-approximating multicone for the target set* $\mathfrak{T}$ *at* $(\overline{y}^0, \overline{y})(\overline{S})$*. In the event that* $\mathbf{c}_k \in \mathfrak{V}_{brk}$ *for some* $k \in \{1, \dots, N\}$*, assume also that* $\overline{\beta}(\overline{S}) < K$ *and* $\hat{l}_1(\cdot, 0, \cdot) \equiv 0$*. Then, there exist*

$$(p_0, p, \lambda) \in \mathbb{R}^* \times AC\left([0, \overline{S}], (\mathbb{R}^n)^*\right) \times \mathbb{R}^* \qquad \text{and} \qquad \mathscr{T} \in \boldsymbol{\mathscr{T}}$$

such that $\lambda \ge 0$ *and:*

i) (NON TRIVIALITY) $(p_0, p, \lambda) \neq 0$*;*
ii) (ADJOINT DIFFERENTIAL INCLUSION)

$$\frac{dp}{ds} \in -\partial^C_y H(\overline{y}, p_0, p, \lambda, \pi, \overline{w}^0, \overline{w}, \overline{\alpha}); \tag{4.8}$$

iii) (NON TRANVERSALITY)

$$(p_0, p(\overline{S})) \in -\lambda \partial^C_{(t,x)} \Psi\Big((\overline{y}^0, \overline{y})(\overline{S})\Big) - \mathscr{T}^\perp \ ; \tag{4.9}$$

iv) (FIRST ORDER MAXIMIZATION) *if* $\mathbf{c}_k = (w^0_k, w_k, a_k, \zeta_k) \in \mathfrak{V}_{ndl}$*,*

$$H(\overline{y}(s_k), p_0, p(s_k), \lambda, 0, w^0_k, w_k, a_k) \le H(\overline{y}(s_k), p_0, p(s_k), \lambda, 0, \overline{w}^0(s_k), \overline{w}(s_k), \overline{\alpha}(s_k)) \tag{4.10}$$

v) (NONSMOOTH GOH CONDITION) *if* $\mathbf{c}_k = (i_k, j_k) \in \mathfrak{V}_{brk}$*,*

$$\min \ p(s_k) \cdot [g_{i_k}, g_{j_k}]_{set}(\overline{y}(s_k)) \le 0^{14} \tag{4.11}$$

If, instead, $\overline{\beta}(\overline{S}) = K$ *and all* $\mathbf{c}_k \in \mathfrak{V}_{ndl}$ *for every* $k = 1, \dots, N$*, then there exists a triple* (p_0, p, λ) *and a real number* $\pi \le 0$ *such that,* i)-iii) *are verified, while inequality* (4.10) *is replaced by*

$$H\Big(\overline{y}(s_k), p_0, p(s_k), \lambda, \pi, w^0_k, w_k, a_k\Big) \le H\Big(\overline{y}(s_k), p_0, p(s_k), \lambda, \pi, \overline{w}^0(s_k), \overline{w}(s_k), \overline{\alpha}(s_k)\Big) \qquad \forall k = 1, \dots, N. \tag{4.12}$$

Proof. Let us observe that we can rephrase (ii) from Lemma 4.8 by saying that there exists a linear form $(\xi_0, \xi, \xi_c) \in (\mathbb{R} \times \mathbb{R}^n \times \mathbb{R})^* \setminus \{(0,0,0)\}$, measurable selections $M(s) \in \partial^C_y \left(f(\overline{y}(s), \overline{\alpha}(s)) \overline{w}^0(s) + \sum_{i=1}^m g_i(\overline{y}(s)) \overline{w}^i(s) \right)$,

[14] Here we mean that the minimization is performed over the elements of

$$p(s_k) \cdot [g_{i_k}, g_{j_k}]_{set}(\overline{y}(s_k)) := \big\{ p(s_k) V, \ V \in [g_{i_k}, g_{j_k}]_{set}(\overline{y}(s_k)) \big\}.$$

$\omega(s) \in \partial_y^C l^e(\overline{y}(s), \overline{w}^0(s), \overline{w}(s), \overline{\alpha}(s))$. a.e. $s \in (0, \overline{S})$, and a choice of

$$\left(V_j^0, V_j, V_j^l\right) \in \left(v^0_{\mathbf{c}_j, s_j}, v_{\mathbf{c}_j, s_j}, v^l_{\mathbf{c}_j, s_j}\right), \quad \forall j = 1, \dots, N,$$

$$(m_t, m_x) \in \partial^C_{(t,x)} \Psi\Big(\overline{y}^0(\overline{S}), \overline{y}(\overline{S})\Big)$$

such that $\xi_0 \le 0$ and, $\forall k = 1, \dots, N$,

$$\begin{aligned} &\xi_0 V_k^0 \xi e^{\int_{s_k}^{\overline{S}} M(s)} V_k + \xi_c \Bigg[m_t V_k^0 + \\ &\qquad + m_x e^{\int_{s_k}^{\overline{S}} M(s)} V_k + \int_{s_k}^{\overline{S}} \omega(s) e^{\int_s^{s_k} M(\sigma)\, \mathrm{d}\sigma} \, \mathrm{d}s V_k + V_k^l \Bigg] \le 0. \end{aligned} \tag{4.13}$$

Setting $\lambda := -\xi_c$, $p_0 := \xi_0 - \lambda m_t$ and, for all $s \in [0, \overline{S}]$,

$$p(s) := (\xi - \lambda m_x)\, e^{\int_s^{\overline{S}} M(\sigma)\, \mathrm{d}\sigma} - \lambda \int_s^{\overline{S}} \omega(\sigma) e^{\int_\sigma^s M(\tau)\, \mathrm{d}\tau} \, \mathrm{d}\sigma,$$

we get that $p(\cdot)$ satisfies the adjoint differential equation $\dot{p}(s) = -p(s)M(s) + \lambda\omega(s)$. In particular, $p(\cdot)$ verifies the adjoint differential inclusion (4.8). Therefore, inequality (4.13) can be written as

$$p_0 V_k^0 + p(\overline{s}_k) V_k - \lambda V_k^l \le 0, \tag{4.14}$$

while (1) of Lemma 4.8 now reads as (4.9). Specializing (4.14) to bracket-like variations $\mathbf{c}_k = (i_k, j_k) \in \mathfrak{V}_{brk}$, we obtain (4.11), whereas, when $\mathbf{c}_k = (w_k^0, w_k, a_k, \zeta_k) \in \mathfrak{V}_{ndl}$, we get (4.10).

The case $\overline{\beta}(\overline{S}) = K$ and all variations verify $\mathbf{c}_k = (w_k^0, w_k, a_k, \zeta_k) \in \mathfrak{V}_{ndl}$ is proved similarly, by making use of (1) instead of (2) from Lemma 4.8.

□

4.7. Infinitely many variations

To complete the proof of Theorem 3.6, we now combine a standard procedure, based on Cantor's non-empty intersection theorem, with the crucial fact that the set-valued brackets are convex-valued. We will only deal with the case when $\overline{\beta}(\overline{S}) < K$,since the case $\overline{\beta}(\overline{S}) = K$ is nothing but the standard first order maximum principle applied to the rescaled, reparametrized problem.[15]

Begin with observing that Lusin's Theorem implies that there exists a sequence of subsets $E_q \subset [0, \bar{S}]$, $q \ge 0$, such that E_0 has null measure, for every $q > 0$ E_q is a compact set such that the restriction to E_q of the map $s \mapsto \left(\overline{w}^0, f(\overline{y}, \overline{\alpha}, \overline{w}^0) + \sum_{i=1}^m g_i(\overline{y})\overline{w}^i, l^e(\overline{y}, \overline{w}^0, \overline{w}), |\overline{w}|\right)(s)$ is continuous, and $(0, \overline{S})_{Leb} = \bigcup_{q=0}^{+\infty} E_q$. For every $q > 0$ let use $D_q \subseteq E_q$ to denote the set of all density points of E_q[16], which, by Lebesgue Theorem has the same Lebesgue measure as E_q. In particular, the subset $D := \bigcup_{q=1}^{+\infty} D_q$ has measure equal to $\overline{S}$.

[15] Of course one also needs that the tangent object to the target happens to be a QDQ approximating cone.

[16] A point x is called a density point for a Lebesgue-measurable set E if $\lim_{\rho \to 0} \dfrac{|B_\rho(x) \cap E|}{|B_\rho(x)|} = 1$

Definition 4.10. Let $X \subseteq D \times \mathfrak{V}$ be any subset of time-generator pairs. We will say that a triple $(p_0, p, \lambda) \in \mathbb{R} \times AC([0, \overline{S}]; \mathbb{R}^n) \times \mathbb{R}_+$ *satisfies property* (P_X) if the following conditions (1)-(3) are verified:

(1) p is a solution on $[0, \overline{S}]$ of the differential inclusion

$$\dot{p} \in -p\, \partial_y^C \left(f(\overline{y}, \overline{\alpha})\overline{w}^0 + \sum_{i=1}^m g_i(\overline{y})\overline{w}^i \right) + \lambda \partial_y^C l^e(\overline{y}, \overline{w}^0, \overline{w}, \overline{\alpha}); \tag{4.15}$$

(2) one has

$$(p_0, p(\overline{S})) \in -\lambda \partial^C_{(t,x)} \Psi\Big(\overline{y}^0(\overline{S}), \overline{y}(\overline{S})\Big) - \overline{\bigcup_{\mathscr{T} \in \boldsymbol{\mathscr{T}}} \mathscr{T}^{\perp}}; \tag{4.16}$$

(3) for every $(s, \mathbf{c}) \in X$, if $\mathbf{c} = (w^0, w, a, \zeta) \in \mathfrak{V}_{ndl}$, then

$$\begin{aligned} p_0 w^0(1+\zeta) + p(s) \left(f(\overline{y}(s), a)w^0 + \sum_{i=1}^m g_i(\overline{y}(s))w^i \right)(1+\zeta) - \lambda l^e(\overline{y}(s), w^0, w, a) \\ \le p_0\overline{w}^0 + p(s) \left(f(\overline{y}(s), \overline{\alpha}(s))\overline{w}^0(s) + \sum_{i=1}^m g_i(\overline{y}(s))\overline{w}^i(s) \right) \\ - \lambda l^e(\overline{y}(s), \overline{w}^0(s), \overline{w}, \overline{\alpha}(s)), \end{aligned} \tag{4.17}$$

while, if $\mathbf{c} = (i, j) \in \mathfrak{V}_{brk}$, then

$$\min_{V \in [g_i, g_j]_{set}(\overline{y}(s))} p_n(s)V \le 0. \tag{4.18}$$

Finally, for any given $X \subseteq D \times \mathfrak{V}$, let us define the subset $\Theta(X) \subset \mathbb{R}^* \times AC\big([0, \overline{S}]; (\mathbb{R}^n)^*\big) \times \mathbb{R}^*$ as

$$\Theta(X) := \left\{ \begin{array}{l} (p_0, p, \lambda) \in \mathbb{R} \times AC([0, \overline{S}]; (\mathbb{R}^n)^*) \times \mathbb{R} \,:\, |(p_0, p(\overline{S}), \lambda)| = 1, \\ (p_0, p, \lambda) \text{ verifies the property } (P_X) \end{array} \right\}.$$

Lemma 4.11. *For any subset* $X \subseteq D \times \mathfrak{V}$, $\Theta(X)$ *is a compact subset of* $\mathbb{R} \times AC\big([0, \overline{S}]; (\mathbb{R}^n)^*\big) \times \mathbb{R}$, *when the latter is endowed with the norm* $\|(p_0, p(\cdot), \lambda)\| := |p_0| + \|p\|_\infty + |\lambda|$.

Proof. Consider a sequence $(p_{0,n}, p_n(s), \lambda_n) \in \Theta(X)$. The set-valued maps

$$s \mapsto \partial_y^C \left(f(\overline{y}(s), \overline{\alpha}(s))\overline{w}^0(s) + \sum_{i=1}^m g_i(\overline{y}(s))\overline{w}^i(s) \right)$$

$$s \mapsto \partial_y^C l^e\Big(\overline{y}(s), \overline{w}^0(s), \overline{w}(s), \overline{\alpha}(s)\Big)$$

have uniformly bounded closed convex values as they are Clarke Jacobians of functions that are globally Lipschitz (after the non-restrictive cut off operation described earlier). Furthermore, the quantities $|p_n(\overline{S})|$, λ_n and $p_{0,n}$ are bounded in norm by 1, so that we are in the position to use the following fact (which can be deduced, *e.g.*, from Theorem 1 in Chapter 2 of [4]):

- *Let $C(s) : [0, \overline{S}] \rightsquigarrow \mathbb{R}^n$ and $B(s) : [0, \overline{S}] \rightsquigarrow \mathbb{R}^n$ be a measurable set-valued map with compact, convex, non-empty values. Moreover, assume that there exists $R > 0$ such that, for every $s \in [0, \overline{S}]$, the sets $B(s) \subset Lin(\mathbb{R}^n, \mathbb{R}^n)$ are $C(s) \subset \mathbb{R}^n$ are all contained in the ball centered in the corresponding origins and of radius R. Let $p_n(s)$ be a sequence of solutions to the differential inclusion*

$$\dot{p}(s) \in p(s)B(s) + C(s), \qquad \textit{for almost all } s \in [0, \overline{S}] \tag{4.19}$$

all satisfying $|p_n(\overline{S})| \leq 1$. Then there is a subsequence of $p_n(s)$ that uniformly converges to a function $p(s)$, and $p(s)$ is also a solution to the differential inclusion (4.19).

Therefore, modulo thrice extracting subsequences from our sequence, we can assume

$$\lambda_n \to \lambda \geq 0,\ p_{0,n} \to p_0 \text{ and } p_n \to p \in AC, \text{ uniformly for } s \in [0, \overline{S}],$$

with $p(s)$ still satisfying the differential inclusion (4.15). Since the paths p_n converges uniformly to p, properties (4.17) and (4.16) are inherited by $p(s)$ from the sequence $p_n(s)$ by passing to the limit. Finally, passing to the limit we get that (4.18) holds true as well. □

Lemma 4.12. *The set $\Theta(D \times \mathfrak{V})$ is non-empty.*

Proof. We are going to use a non-empty intersection argument, which is quite standard, except for the part concerning the set-valued bracket. Let us notice that

$$\Theta(X_1 \cup X_2) = \Theta(X_1) \cap \Theta(X_2), \quad \forall X_1, X_2 \subseteq D \times \mathfrak{V},$$

so that

$$\Theta(D \times \mathfrak{V}) = \bigcap_{\substack{X \subseteq D \times \mathfrak{V} \\ X \text{ finite}}} \Theta(X). \tag{4.20}$$

We have to prove that this infinite intersection is non-empty. We begin with proving the following fact:

Claim 1 The set $\Theta(X)$ *is non-empty as soon as X is finite.*

Indeed, by (4.14)–(4.11) we already know that that $\Theta(X) \neq \emptyset$ whenever X comprises N couples $(s_k, \mathbf{c}_k) \in D \times \mathfrak{V}$ such that $s_k < s_l$ whenever $1 \leq k < l \leq N$. We have to show that we can allow X to have the general form

$$X = \Big\{(s_k, \mathbf{c}_k) \in D \times \mathfrak{V}, \quad s_k \leq s_l \text{ as soon as } 1 \leq k < l \leq N\Big\}.$$

For any $(r, k) \in \mathbb{N} \times \{1, \dots, N\}$, choose an instant $s_{k,r} \in E := \bigcup_{q=1}^{\infty} E_q$ in such a way that such that $s_{1,r} < \dots < s_{N,r}$ and the sequences $(s_{k,r})_{r \in \mathbb{N}}$, $1 \leq k \leq N$, converge to s_k. For every $r \in \mathbb{N}$, Consider the sets $X_r := \{(s_{k,r}, \mathbf{c}_k),\ k \leq N\}$, $r \in \mathbb{N}$. By the previous steps we know that $\Theta(X_r)$ is non-empty, so that we can choose $(p_{0,r}, p_r(s), \lambda_r) \in \Theta(X_r)$. Again, modulo thrice extracting subsequences, we can assume that

- i $p_{0,r}$ converges to a real number p_0,
- ii p_r uniformly converges to an absolutely continuous function p solving the adjoint differential inclusion, and
- iii λ_r converges to a non-negative real number λ.

Therefore, (p_0, p) inherits the non-transversality condition (4.16). Moreover, the uniform convergence of the p_r and the continuity of the involved functions imply that (p_0, p, λ) verifies the Hamiltonian maximization (4.17) at the instants s_k.

What is going to be a little trickier to prove is that, if $\mathbf{c}_k = (i_k, j_k)$ is a bracket-like variation generator, (4.18) holds at time s_k for the multiplier p, starting from the fact that it is satisfied at time $s_{k,r}$ by the multiplier p_r. As a matter of fact, what we know is that, for every natural number r,

$$p_r(s_{k,r})V_{k,r} \leq 0$$

for some $V_{k,r} \in [g_{i_k}, g_{j_k}]_{set}(\overline{y}(s_{k,r}))$. By definition of the set-valued Lie-bracket as closure of a convex hull, this implies that, for every $r \in \mathbb{N}$, there is a sequence $(y_{k,r,n})_{n\in\mathbb{N}}$ of differentiability points for both g_{i_k} and g_{j_k} converging to $\overline{y}(s_{k,r})$ and such that

$$p_r(s_{k,r})\left(\lim_{n\to+\infty} Dg_{i_k}(y_{k,r,n})g_{j_k}(y_{k,r,n}) - Dg_{j_k}(y_{k,r,n})g_{i_k}(y_{k,r,n})\right) \leq \frac{1}{r}.$$

Therefore, there is a large enough number $N_{1,r}$, such that, for all $n \geq N_{1,r}$,

$$p_r(s_{k,r})\Big(Dg_{i_k}(y_{k,r,n})g_{j_k}(y_{k,r,n}) - Dg_{j_k}(y_{k,r,n})g_{i_k}(y_{k,r,n})\Big) \leq \frac{2}{r}.$$

Also, since $\lim_{n\to+\infty} Dg_{i_k}(y_{k,r,n})g_{j_k}(y_{k,r,n}) - Dg_{j_k}(y_{k,r,n})g_{i_k}(y_{k,r,n})$ is bounded as r varies in $\mathbb{N}$, we can assume, modulo extracting a subsequence, that the limit

$$\lim_{r\to+\infty}\lim_{n\to+\infty} Dg_{i_k}(y_{k,r,n})g_{j_k}(y_{k,r,n}) - Dg_{j_k}(y_{k,r,n})g_{i_k}(y_{k,r,n}) \tag{4.21}$$

does exist. Let us call V_k this limit, *i.e.* let us set

$$V_k := \lim_{r\to+\infty}\lim_{n\to+\infty} \big[g_{j_k}, g_{i_k}\big](y_{k,r,n}).$$

Since $\lim_{n\to\infty} y_{k,r,n} = \overline{y}(s_{k,r})$ for any r and $\lim_{r\to\infty} \overline{y}(s_{k,r}) = \overline{y}(s_k)$, we can construct a sequence $(N_{2,r})_{r\in\mathbb{N}}$ of natural numbers such that

- i $N_{2,r} \to +\infty$,
- ii $y_{k,r,N_{2,r}} \to \overline{y}(s_k)$, and
- iii $y_{k,r,N_{2,r}}$ is a point of differentiability for both g_{i_k} and g_{j_k} for any $r \in \mathbb{N}$.

By the existence of the limit (4.21) we deduce that, by taking, for any r, a suitably large $N_{3,r} > N_{2,r} > N_{1,r}$ one has

$$V_k := \lim_{r\to+\infty}\lim_{n\to+\infty} \big[g_{j_k}, g_{i_k}\big](y_{k,r,N_{3,r}}).$$

As $N_{3,r} > N_{2,r}$, this means $V_k \in [g_{i_k}, g_{j_k}]_{set}(\overline{y}(s_k))$, in that it is the limit of the Lie bracket of g_{i_k} and g_{j_k} computed along a sequence of points that converges to $\overline{y}(s_k)$. Moreover, since $N_{3,r} > N_{1,r}$, one has

$$p_r(s_{k,r}) \cdot \big[g_{j_k}, g_{i_k}\big](y_{k,r,N_{3,r}}) \leq \frac{2}{r}.$$

By passing to the limit as r goes to infinity, we get that

$$p(s_k)\, V_k \leq 0 \qquad \Big(\text{ and } V_k \in [g_{i_k}, g_{j_k}]_{set}(\overline{y}(s_k))\Big),$$

which concludes the proof of **Claim 1**.

In view of **Claim 1** and of (4.20), by Cantor's intersection theorem we can conclude that $\Theta(D \times \mathfrak{V}) \neq \emptyset$, for it is the intersection of a family of compact non-empty sets with the property that any finite intersection of these sets is non-empty. Now, any $(p_0, p, \lambda) \in \Theta(D \times \mathfrak{V})$ clearly satisfies conditions **i) -v)** of Theorem 3.6. As for the second order condition **vi)**, if (4.18) holds for $\mathbf{c} = (i,j)$ and $\mathbf{c} = (j,i)$ at the same time, then we know that there are $V^-, V^+ \in [g_i, g_j]_{set}(\overline{y}(s))$ such that $p(s)V^- \leq 0$ and $p(s)V^+ \geq 0$. Since $[g_i, g_j]_{set}(\overline{y}(s))$ is convex, this implies that there exists $\alpha \in [0,1]$ such that $V := \alpha V^- + (1-\alpha)V^+$ (belongs to $[g_i, g_j]_{set}(\overline{y}(s))$ and) satisfies $p(s)V = 0$. □

5. An example

In the minimum problem (5.1) below, the fact that a certain control $\bar{u}$ is not optimal is not deduced by the standard first order maximum principle. In other words conditions $i)$ through $iv)$ turn out to be satisfied while, instead, the nonsmooth Goh condition condition $v)$ is not verified, so the optimality of $\bar{u}$ is ruled out.

Given the function $\Psi(t,x) := |x|^2 + (t-1)^2$, we will consider the following Mayer optimal control problem:

$$\begin{gathered}\min \Psi(T, x(T)) \\ \text{in the set of solutions to} \\ \begin{cases} \dfrac{\mathrm{d}x}{\mathrm{d}t} = f(x) + g_1(x)u^1 + g_2(x)u^2 \\ (x^1, x^2, x^3, \nu)(0) = \big(1, 0, 2, 0\big) \qquad (x^1, x^2, x^3) \in \mathfrak{T} \times [0, K], \ \|\nu\|_1 \leq 4 \end{cases}\end{gathered} \tag{5.1}$$

where the control u take values in $C := \mathbb{R}^2$, the target is defined as $\mathfrak{T} := [0,1] \times B_{\frac{1}{2}}\left(0, 0, \frac{1}{2}\right)$, and

$$f(x) :\equiv \begin{pmatrix} 0 \\ 0 \\ -1 \end{pmatrix}, \qquad g_1(x) := \begin{pmatrix} 1 \\ 0 \\ x^2 - |x^2| \end{pmatrix}, \qquad g_2(x) := \begin{pmatrix} 0 \\ 1 \\ x^1 + |x^1| \end{pmatrix}.$$

We will show that the control $\overline{u}(s) \equiv (-1, 0)$ satisfies conditions $i)$ through $iv)$ but not $v)$.
The extended problem reads

$$\min \Psi(y^0(S), y(S))$$

$$\begin{cases} \left(\dfrac{\mathrm{d}y^0}{\mathrm{d}s}, \dfrac{\mathrm{d}y^1}{\mathrm{d}s}, \dfrac{\mathrm{d}y^2}{\mathrm{d}s}, \dfrac{\mathrm{d}y^3}{\mathrm{d}s}, \dfrac{d\beta}{\mathrm{d}s}\right) = \big(w^0\,,\, w^1\,,\, w^2\,,\, \big(y^2 - |y^2|\big)\, w^1 + \big(y^1 + |y^1|\big)\, w^2 - w^0\,,\, |w|\big) \\ (y^0, y^1, y^2, y^3, \beta)(0) = \big(0, 1, 0, 2, 0\big) \qquad (y^0, y^1, y^2, y^3, \beta)(\overline{S}) \in \mathfrak{T} \times [0, 4] \end{cases}$$

In the notation of the extended problem, we shall focus on the (constant) space-time control $(\bar{w}^0, \bar{w}^1, \bar{w}^2)$ defined by $\overline{S} = 2$ and

$$\left(\bar{w}^0, \bar{w}^1, \bar{w}^2\right)(s) := \left(\frac{1}{2}, -\frac{1}{2}, 0\right) \qquad \forall s \in [0,2].$$ [17]

and the corresponding trajectory

$$\left(\bar{y}^0, \bar{y}^1, \bar{y}^2, \bar{y}^3\right)(s) := \left(\frac{1}{2}s, 1 - \frac{1}{2}s, 0, 2 - \frac{1}{2}s\right)$$

[17] Actually, since $w^0 \equiv \frac{1}{2} > 0$ all of these extended sense controls correspond to strict sense controls $(1, -1, 0)$. This will not be true for the global minimizer presented at the end of the section.

ending at $\big(\bar{y}^0,\bar{y}^1,\bar{y}^2,\bar{y}^3\big)(2) = \big(1,0,0,1\big)$, with final cost $\Psi\Big(\big(\bar{y}^0,\bar{y}\big)(2)\Big) = 1$.
Our adjoint differential inclusion reads

$$\frac{dp_1}{\mathrm{d}s} = 0,\ \frac{dp_2}{\mathrm{d}s} \in [0, p_3(s)],\ \frac{dp_3}{\mathrm{d}s} = 0,$$

so that, p_1 and p_3 need to be constant. As a QDQ-approximating multicone to the target set, we consider the (Boltyanski) cone $\mathbb{R}_- \times \mathbb{R} \times \mathbb{R} \times \mathbb{R}_-$. Moreover Ψ is a smooth function and $\partial^C_{(t,x)}\Psi\Big(\big(\bar{y}^0,\bar{y}^1,\bar{y}^2,\bar{y}^3\big)(2)\Big) = \{(0,0,0,2)\}$, so that transversality condition reads

$$(p_0,p_1,p_2,p_3)(2) \in]-\infty,0] \times \{0\} \times \{0\} \times]-\infty,-2\lambda].$$

Therefore $p_1 \equiv 0$. Furthermore, if p_3 were equal to 0, then λ would also be 0 and p_2 would be constantly equal to 0. Finally, from the first order maximization condition, one would imply $p_0 = 0$, and the non-triviality condition would be violated. Hence $p_3 < 0$, so that we can assume $p_3 = -1$.
The transversality condition and adjoint differential inclusion together imply that, if some multipliers (p_0,p,λ,π) exist satisfying conditions *i)* through *iv)*, then $p_1 \equiv 0$, $p_3 \equiv -1$ and $p_2(s)$ is a non-increasing function ending at 0 (more precisely, $\dfrac{\partial p_2}{\partial s} \in [-1,0]$).
Writing the first order maximization condition, we get that

$$H = p_0 w^0 + p_2(s) w^2 + (-1)[(2-s)w^2 - w^0] = (p_0+1)w^0 + w^2[p_2(s) - (2-s)]$$

should be maximized by the choice $w^2 = 0$ and $w^0 = \frac{1}{2}$ for almost all s. This is possible only if $p_0 = -1$ and $p_2(s) = 2-s$ for almost all s in $[0,2]$, which means $p_2(s) = 2-s$ for all s because p_2 is absolutely continuous. In other words, the adjoint path

$$\Big(\big(\bar{y}^0,\bar{y}^1,\bar{y}^2,\bar{y}^3\big)\ ,\ (p_0,p_1,p_2,p_3)\Big)(s) := \left(\left(\frac{1}{2}s, 1-\frac{1}{2}s, 0, 2-\frac{1}{2}s\right)\ ,\ (-1,0,2-s,-1)\right)$$

satisfies the first order conditions of the maximum principle.

However, *the nonsmooth Goh condition is not verified*, because the set-valued Lie bracket $[g_1,g_2]_{set}(y^1,y^2,y^3)$ is equal to $\{0\}\times\{0\}\times[2,4]$ whenever $y^1 > 0$ and $y^2 = 0$, which yields

$$0 \notin [-4,-2] = (p_1,p_2,p_3)(s)\cdot\Big(\{0\}\times\{0\}\times[2,4]\Big).$$

On the other hand, let us consider the (impulsive) control

$$(\hat{w}^0,\hat{w}^1,\hat{w}^2) := \begin{cases} \left(\frac{1}{2},-\frac{1}{2},0\right) & \text{if } s \in [0,2] \\ (0,0,1) & \text{if } s \in]2, 2+\frac{\sqrt{2}}{2}] \\ (0,1,0) & \text{if } s \in]2+\frac{\sqrt{2}}{2}, 2+\sqrt{2}] \\ (0,0,-1) & \text{if } s \in]2+\sqrt{2}, 2+3\frac{\sqrt{2}}{2}] \\ (0,-1,0) & \text{if } s \in]2+3\frac{\sqrt{2}}{2}, 2+2\sqrt{2}] \end{cases}$$

It is easy to see that the corresponding trajectory $(\hat{y}^0,\hat{y}^1,\hat{y}^2,\hat{y}^3)$ ends at $(1,0,0,0)$ which is a point in the target set, but also a point of global minimum of Ψ. The process turns out to be feasible because one also has $\beta(\overline{S}) = \beta(2+2\sqrt{2}) = 1+2\sqrt{2} < K = 4$. Let us underline a crucial difference between the geometrical pictures of the two extremals: in the latter it is simple to check that the QDQ-approximating cone $\mathbb{R}_- \times \mathbb{R} \times \mathbb{R} \times \mathbb{R}_+$ to the target at the end-point $(1,0,0,0)$ allows for the adjoint map

$$(p_0, p_1, p_2, p_3) \equiv 0, \quad \lambda = 1, \quad \pi = 0;$$

on the contrary, in the case of the control $(\overline{w}^0, \overline{w}^1, \overline{w}^2)$, the QDQ-approximating cone $\mathbb{R}_- \times \mathbb{R} \times \mathbb{R} \times \mathbb{R}_-$ to the target at the end-point $(1,0,0,1)$ forced p_3 to be strictly negative, which is a violation of the nonsmooth Goh condition.

References

[1] A.A. Agrachev, A. Morse, E.D. Sontag, H.J. Sussmann and V.I. Utkin, *Nonlinear and Optimal Control Theory*, Cetraro, Italy (2004), n. 1932, Springer, xiv+351.

[2] F. Angrisani and F. Rampazzo, Quasi Differential Quotients. Preprint arXiv:2107.07638 (2021).

[3] M.S. Aronna, M. Motta and Rampazzo, A higher-order maximum principle for impulsive optimal control problem. *SIAM J. Control Optim.* **58** (2020) 814–844.

[4] J.P. Aubin and A. Cellina, Differential inclusions: set-valued maps and viability theory. Springer (1984), xiii+342.

[5] J.P. Aubin and H. Frankowska, Vol. 2 of *Set-valued analysis. System, Control, Foundation and Applications.* Birkhauser, Boston (1990)

[6] D. Azimov and R. Bishop, New trends in astrodynamics and applications: optimal trajectories for space guidance. *Ann. N Y Acad. Sci.* **1065** (2005) 189-209.

[7] A. Bressan, Hyper-impulsive motions and controllizable coordinates for Lagrangian systems. *Atti Accad. Naz. Lincei Mem. Cl. Sci. Fis. Mat. Natur. Sez. Ia (8)* **19** (1989) (1991) 195–246.

[8] A. Bressan and M. Motta, A class of mechanical systems with some coordinates as controls. A reduction of certain optimization problems for them. Solutions methods. *Mem. Mat. Acc. Lincei, s. 9* **2** (1993) 5–30

[9] A. Bressan and F. Rampazzo, Moving constraints as stabilizing controls in classical mechanics. *Arch. Ration. Mech. Anal.* **196** (2010) 97–141.

[10] C. Calcaterra, Basic properties of nonsmooth Hörmander vector fields and Poincarè's inequality. *Commun. Math. Anal.* **11** (2011) 1-40.

[11] F. Cardin and C. Viterbo, Commuting Hamiltonians and Hamilton-Jacobi multi-time equations. *Duke Math. J.* **144** (2008) 235–284.

[12] A.J. Catllá, D.G. Schaeffer. T.P. Witelski, E.E. Monson, A.L. Lin, On spiking models for synaptic activity and impulsive differential equations. *SIAM Rev.* **50** (2008) 553–569.

[13] F.H. Clarke, Y.S. Ledyaev, R.J. Stern and R.R. Wolenski, Nonsmooth Analysis and Control Theory Graduate Texts in Mathematics, Volume 178, 1998, Springer, xiii+278.

[14] F. Clarke, Functional analysis, calculus of variations and optimal control Graduate Texts in Mathematics, 264, 2013. 264, Springer, London, xiv+591.

[15] E. Feleqi and F. Rampazzo, Iterated Lie brackets for nonsmooth vector fields. *NoDEA Nonlinear Differ. Equ. Appl.* **24** (2017) Paper No. 61, 1-43.

[16] P.H. Gajardo, C. Ramirez and A. Rapaport, Minimaltime sequential batch reactors with bounded and impulse controls for one or more species. *SIAM J. Control Optim.* **47** (2008) 2827–2856.

[17] C.M. Marle, Géométrie des systèmes mécaniques à liaisons actives. *Symp. Geometry Math. Phys.* (1991) 260–287.

[18] A. Montanari and D. Morbidelli, Balls defined by nonsmooth vector fields and the Poincaré inequality. *Ann. Inst. Fourier (Grenoble)* **54** (2004) 431-452.

[19] A. Montanari and D. Morbidelli, On the subRiemannian cut locus in a model of free two-step Carnot group. *Calc. Variat. Partial Differ. Equ.* **56** (2017) 1–26.

[20] B.S. Mordukhovich and H.J. Sussmann, Nonsmooth analysis and geometric methods in deterministic optimal control. Vol. 78 of *The IMA Volumes in Mathematics and its Applications.* Springer (1996), ix+246.

[21] M. Palladino and F. Rampazzo, A geometrically based criterion to avoid infimum-gaps in optimal control. *J. Differ. Equ.* **269** (2020) 10107–10142.

[22] F. Rampazzo, On the Riemannian structure of a Lagrangian system and the problem of adding time-dependent constraints as controls. *Eur. J. Mech. A Solids* **10** (1991) 405–431.

[23] F. Rampazzo, Frobenius-type theorems for Lipschitz distributions. *J. Differ. Equ.* **243** (2007) 270–300.

[24] F. Rampazzo and H.J. Sussmann, Commutators of flow maps of nonsmooth vector fields. *J. Differ. Equ.* **232** (2007) 134–175.

[25] F. Rampazzo and H.J. Sussmann, Set-valued differentials and non-smooth version of Chow's theorem. Proc. of the 40th IEEE Conf. on Decision and Control, Orlando, FL, December 2001, vol. 3, IEEE publications, New York (2001), pp. 111–129.

[26] S. Simic, Lipschitz distributions and Anosov flows. *Proc. Am. Math. Soc.* **124** (1996) 1869–1877.

[27] H.J. Sussmann, High-order point variations and generalized differentials. Geometric control and nonsmooth analysis. Ser. Adv. Math. Appl. Sci., 76, World Sci. Publ., Hackensack, NJ (2008) 327–357.

[28] H.J. Sussmann, Warga derivate containers and other generalized differentials, in *Proceedings of the 41st IEEE Conference on Decision and Control* (2002).
[29] R. Vinter, Optimal Control Modern Birkhäuser Classics. Springer (2010) xx+507.
[30] J. Warga, Controllability, extremality, and abnormality in nonsmooth optimal control. *J. Optim. Theory Appl.* **41** (1983) 239–260.
[31] J. Warga, Optimization and controllability without differentiability assumptions. *SIAM J. Control Optim.* **21** (1983) 837–855.