SIAM J. CONTROL OPTIM.
Vol. 58, No. 2, pp. 814–844

A HIGHER-ORDER MAXIMUM PRINCIPLE FOR IMPULSIVE OPTIMAL CONTROL PROBLEMS*

M. SOLEDAD ARONNA†, MONICA MOTTA‡, AND FRANCO RAMPAZZO‡

Abstract. We consider a nonlinear system, affine with respect to an unbounded control u which is allowed to range in a closed cone. With this system we associate a Bolza type minimum problem, with a Lagrangian having sublinear growth with respect to u. This lack of coercivity gives the problem an *impulsive* character, meaning that minimizing sequences of trajectories happen to converge towards discontinuous paths. As is known, a distributional approach does not make sense in such a nonlinear setting, where, instead, a suitable embedding in the graph space is needed. We provide higher-order necessary optimality conditions for properly defined impulsive minima in the form of equalities and inequalities involving iterated Lie brackets of the dynamical vector fields. These conditions are derived under very weak regularity assumptions and without any constant rank conditions.

Key words. impulsive optimal control problems, Pontryagin maximum principle, higher-order necessary conditions

AMS subject classifications. 49K15, 49N25

DOI. 10.1137/19M1273785

1. Introduction.

In this paper we establish necessary optimality conditions for the *space-time, impulsive extension* of the free end-time optimal control problem

$$\text{minimize } \Psi(T, x(T)) + \int_0^T \ell(x(t), u(t), a(t))\, dt, \tag{1.1}$$

where the minimization is performed over the set of *processes* (T, u, a, x, v) verifying

$$\begin{cases} \dfrac{dx}{dt}(t) = f(x(t), a(t)) + \displaystyle\sum_{i=1}^{m} g_i(x(t))u^i(t), \quad \text{a.e. } t \in [0, T], \\ \dfrac{dv}{dt}(t) = |u(t)|, \\ (x, v)(0) = (\check{x}, 0), \quad v(T) \le K, \quad (T, x(T)) \in \mathfrak{T}. \end{cases} \tag{1.2}$$

Here, $0 \le K \le +\infty$, the *target* $\mathfrak{T}$ is a closed subset of $\mathbb{R}_+ \times \mathbb{R}^n$, the control a ranges in a compact set $A \subset \mathbb{R}^q$, and standard regularity hypotheses are verified by the vector fields f, g_i and the cost functions ℓ, Ψ. Less usual assumptions are made on the control maps u and on the u-growth of the Lagrangian ℓ. In particular,

*Received by the editors July 10, 2019; accepted for publication (in revised form) December 17, 2019; published electronically March 12, 2020.
https://doi.org/10.1137/19M1273785
Funding: This research is partially supported by Padua University grant SID 2018, "Controllability, stabilizability and infimum gaps for control systems," BIRD 187147; by the "National Group for Mathematical Analysis, Probability and their Applications" (GNAMPA-INdAM) (Italy); by the European Union under the 7th Framework Programme FP7-PEOPLE-2010-ITN, grant 264735-SADCO; by FAPERJ (Brazil) through the "Jovem Cientista do Nosso Estado" Program; by CNPq and CAPES (Brazil); and by the Alexander von Humboldt Foundation (Germany).
†Escola de Matemática Aplicada, Fundação Getúlio Vargas, Rio de Janeiro 22250-900, Brazil (soledad.aronna@fgv.br).
‡Dipartimento di Matematica, Università di Padova, Padova, Italy 35121 (motta@math.unipd.it, rampazzo@math.unipd.it).

(i) the *unbounded* controls u, which take values in a closed cone $\mathcal{C} \subseteq \mathbb{R}^m$, are L^1 functions verifying $\|u\|_1 := \int_0^T |u(t)|dt = v(T) \le K$ (when $K = +\infty$, this means that no a priori bounds are assumed on the L^1 norm of u);

(ii) the Lagrangian $\ell : \mathbb{R}^n \times \mathcal{C} \times A \to \mathbb{R}$ has the form $\ell(x,u,a) = \ell_0(x,a) + \ell_1(x,u)$, with a continuous *recession function* $\hat{\ell}_1$ given by

$$\hat{\ell}_1(x, w^0, w) := \lim_{r \to w^0} r\ell_1\left(x, \frac{w}{r}\right) \quad \text{for } (x, w^0, w) \in \mathbb{R}^n \times \mathbb{R}_+ \times \mathcal{C}.$$

In particular ℓ has sublinear growth in u.

On the one hand, optimal control problems with such a slow u-growth in the cost are motivated by several applications [14, 16, 23, 18, 28, 10, 35, 17]. For instance, dynamics affine in the unbounded controls govern the motion of a mechanical system of mutually linked rigid bodies. In that case, the controlled parameters are the speeds of the *shape* coordinates.

On the other hand, the lack of a sufficiently fast u-growth of the cost ℓ may cause minimizing sequences of trajectories to tend towards discontinuous paths.[1] Motivated by that, and following a nowadays standard approach[2] [43, 38, 15, 30, 29, 40, 24], one "compactifies" the problem by embedding the original control system in the *extended, space-time* system

$$\text{(1.3)} \qquad \begin{cases} \dfrac{dy^0}{ds}(s) = w^0(s), \\ \dfrac{dy}{ds}(s) = f(y(s), \alpha(s))w^0(s) + \displaystyle\sum_{i=1}^m g_i(y(s))w^i(s), \\ \dfrac{d\beta}{ds}(s) = |w(s)|, \\ (y^0, y, \beta)(0) = (0, \check{x}, 0), \\ (y^0(S), y(S), \beta(S)) \in \mathfrak{T} \times [0, K], \end{cases} \qquad \text{a.e. } s \in [0, S],$$

and considering the *extended* cost functional

$$\text{(1.4)} \qquad \Psi(y^0(S), y(S)) + \int_0^S \ell^e(y(s), w^0(s), w(s), \alpha(s))ds,$$

where $\ell^e(x, w^0, w, a) := \ell_0(x,a)w^0 + \hat{\ell}_1(x, w^0, w)$. In view of the *rate independence of* problem (1.3)–(1.4), one can consider *bounded* controls verifying $w^0(s) + |w(s)| = 1$ with $w(s) \in \mathcal{C}$, $w^0(s) \ge 0$, and $\alpha(s) \in A$ for a.e. $s \in [0, S]$.

On the one hand, as soon as $w^0 > 0$ a.e., (1.3)–(1.4) is nothing but a time reparameterization of the original control problem (1.1)–(1.2), and the unboundedness of u is reflected in the possibility of taking w^0 indefinitely small. On the other hand, by allowing processes $(S, w^0, w, \alpha, y^0, y, \beta)$ such that $w^0 = 0$ a.e. on nondegenerate s-subintervals $[s_1, s_2] \subset [0, S]$, we are embedding (1.1)–(1.2) in a more general problem. Indeed, the time variable $t = y^0(s)$ has a constant value $\bar{t}$ on such an interval $[s_1, s_2]$, while the space trajectory $y(s)$ evolves according to the nonlinear dynamics

[1] For instance, if one considers the minimum time problem—i.e., $\ell \equiv 1$ and $\Psi \equiv 0$—with a target in $\mathbb{R}^n$ intersecting the orbit of the vector field g_1 issuing from the initial point $\check{x}$, the infimum value is zero and the "extended" optimal trajectory should run through the mentioned orbit *with infinite speed*.

[2] Because of the nonlinearity of the dynamics, a *distributional* interpretation lacks essential prerequisites for robustness [25].

$\frac{dy}{ds} = \sum_{i=1}^m g_i(y)w^i$. Through a discontinuous inverse σ of y^0, $(\sigma(t) \in (y^0)^{\leftarrow}(\{t\}))$, one can regard $x(t) := y(\sigma(t))$ as a discontinuous trajectory of the original problem. In particular, the jump $x(\bar{t}+) - x(\bar{t}-) = y(s_2) - y(s_1)$ depends on the restriction $w_{|_{[s_1,s_2]}}$. As mentioned above, the lack of coercivity often makes such *impulsive* behavior very likely. For this reason, our necessary conditions will concern a minimum for the extended, space-time problem (1.3)–(1.4). We begin by exploiting its rate independence in order to establish a first order maximum principle in Theorem 3.1. While the idea is anything but new—see, for instance, [41, 33, 29, 8, 31]—here we stress a kind of competition among the Hamiltonians corresponding to the drift and the "nondrift" Hamiltonian.

However, the main novelty of the paper consists in a maximum principle containing *higher-order* necessary conditions for an optimal process $(\hat{S}, \hat{w}^0, \hat{w}, \hat{\alpha}, \hat{y}^0, \hat{y}, \hat{\beta})$ which involve iterated Lie brackets (see Theorem 4.2). Higher-order conditions play useful for ruling out the optimality of some of the trajectories verifying the first order necessary conditions (see the examples in section 5 and in [5]). In order to establish these new necessary conditions we strongly rely on the unboundedness of the controls, which translates into the possibility of utilizing impulsive intervals—where $w^0 = 0$—in the construction of variations. In particular, in connection with any subspace contained in the cone C (see section 2), the dual product of the adjoint path with the corresponding iterated Lie brackets turns out to vanish, a fact which is usually labeled as the *generalized Goh condition* (from which other higher-order necessary conditions, involving the drift, classically follow).

Incidentally, in the particular case when the optimal trajectory is a trajectory of the original problem, namely, $\hat{w}^0 > 0$ for almost every $s \in [0, \hat{S}]$, by merely utilizing the reparameterization $s = (\hat{y}^0)^{-1}(t)$, our necessary conditions can be expressed in the original time variable t (see Theorem 4.11). Therefore, when the absence of infimum gaps between the original, unbounded control problem and its impulsive extension is known, our necessary conditions apply also to the original (nonimpulsive) problem as well. In this case, our results improve the higher-order necessary conditions proved in [19] for a minimum time problem with unbounded controls (see Remark 4.12).

Lie bracket-involving necessary conditions have been widely investigated within classical geometric control theory, a quite incomplete list of references being [3, 26, 27, 42, 39, 11]. Yet, as far as impulsive control theory is concerned, the only results involving higher-order—actually, second order—conditions deal, to our knowledge, with the so-called *commutative case*, where $[g_i, g_j] \equiv 0$ for all $i, j = 1, \dots, m$ (see, e.g., [9, 7, 20]).

Our paper is organized as follows. In subsection 1.1 we introduce some general notation and a set of definitions and technical results involving Lie brackets. The optimal control problem is described in detail in section 2, together with its *space-time* extension. In section 3, Theorem 3.1, we state a first order maximum principle. In section 4, Theorem 4.2, we establish our main result, namely the higher-order maximum principle, whose proof is given in section 6. An example illustrating the strength of higher-order conditions in singling out the minimum among first order extremals is provided in section 5.

1.1. Notations and preliminaries.

1.1.1. Some basic notation. Let $N \geq 1$ be an integer. For any $i \in \{1, \dots, N\}$, we write $\mathbf{e}_i$ for the ith element of the canonical basis of $\mathbb{R}^N$. Given $\check{x} \in \mathbb{R}^N$, $\mathbb{B}_N(\check{x}) := \{x \in \mathbb{R}^N : |x - \check{x}| \leq 1\}$, $\mathbb{B}_N := \mathbb{B}_N(0)$, and $\partial\mathbb{B}_N := \{x \in \mathbb{R}^N : \ |x| = 1\}$. A subset $\mathcal{K} \subseteq \mathbb{R}^N$ is a *cone* if $\alpha x \in \mathcal{K}$ whenever $\alpha > 0$, $x \in \mathcal{K}$. Given a subset $X \subseteq \mathbb{R}^N$, we will use $X^\perp$ to denote the *polar* of X, i.e., $X^\perp \doteq \{p \in \mathbb{R}^N : \ p \cdot x \leq 0, \quad \text{for all} \ \ x \in X\}$. Given an interval I and $X \subseteq \mathbb{R}^N$, we write $AC(I,X)$ for the space of absolutely continuous functions, $C^0(I,X)$ for the space of continuous functions, $L^1(I,X)$ for the Lebesgue space of L^1-functions, and $L^\infty(I,X)$ for the Lebesgue space of measurable, essentially bounded functions, respectively, defined on I and assuming values in X. As customary, we shall use $\|\cdot\|_{L^\infty(I,X)}$, and $\|\cdot\|_{L^1(I,X)}$ to denote the essential supremum norm and the L^1-norm, respectively. When no confusion may arise, we will simply write $\|\cdot\|_\infty$ and $\|\cdot\|_1$. We set $\mathbb{R}_+ := [0,+\infty)$ and $\mathbb{R}_- := (-\infty, 0]$. Given an integer $k \geq 0$ and an open subset $\Theta \subseteq \mathbb{R}^N$, we say that a function $F\colon \Theta \to \mathbb{R}^N$ is *of class* C^k if it possesses continuous partial derivatives up to order k in Θ. Given a real-valued function $F : [a,b] \to \mathbb{R}$, we define the essential infimum of F as $\operatorname{ess\,inf} F := \sup\big\{r \in \mathbb{R} : \ \operatorname{meas}\{x \in [a,b] : \ F(x) < r\} = 0\big\}$, where meas denotes the Lebesgue measure. Finally, for all τ_1, $\tau_2 \in (0,+\infty)$ and for any pair $(z_1, z_2) \in C^0([0,\tau_1],\mathbb{R}^N) \times C^0([0,\tau_2],\mathbb{R}^N)$, let us define the distance

$$\mathrm{d}\big((\tau_1, z_1), (\tau_2, z_2)\big) := |\tau_1 - \tau_2| + \|\tilde{z}_1 - \tilde{z}_2\|_\infty, \tag{1.5}$$

where for any $z \in C^0([0,\tau],\mathbb{R}^N)$, $\tilde{z}$ denotes its continuous constant extension to $\mathbb{R}_+$.

1.1.2. Boltyanski approximating cones.

DEFINITION 1.1. *Let Z be a subset of $\mathbb{R}^N$ for some integer $N \geq 1$. Fix $z \in Z$. We say that a convex cone $\mathcal{K} \subseteq \mathbb{R}^N$ is a* Boltyanski approximating cone for Z at z *if there exist a convex cone $C \subset \mathbb{R}^M$ for some integer $M \geq 0$, a neighborhood V of 0 in $\mathbb{R}^M$, and a continuous map $F : V \cap C \to Z$ such that $F(0) = z$; there exists a linear map $L : \mathbb{R}^M \to \mathbb{R}^N$ verifying $F(v) = F(0) + Lv + o(|v|)$ for all $v \in V \cap C$; $LC = \mathcal{K}$.*

DEFINITION 1.2. *Let us consider two subsets $\mathcal{A}_1, \mathcal{A}_2$ of a topological space $\mathcal{X}$. If $y \in \mathcal{A}_1 \cap \mathcal{A}_2$, we say that $\mathcal{A}_1$ and $\mathcal{A}_2$ are* locally separated at y *provided there exists a neighborhood V of y such that $\mathcal{A}_1 \cap \mathcal{A}_2 \cap V = \{y\}$.*

The following open-mapping-based result characterizes set separation in terms of linear separation of approximating cones (see, e.g., [42]).

THEOREM 1.3. *Let Z_1 and Z_2 be subsets of $\mathbb{R}^N$, $z \in Z_1 \cap Z_2$, and let $\mathcal{K}_1$, $\mathcal{K}_2 \subseteq \mathbb{R}^N$ be Boltyanski approximating cones for Z_1 and Z_2, respectively, at z. If $\mathcal{K}_1$ or $\mathcal{K}_2$ is not a subspace and Z_1, Z_2 are locally separated at z, then $\mathcal{K}_1$ and $\mathcal{K}_2$ are* linearly separated*, namely, there exists a covector $\lambda \in \mathbb{R}^N$ such that $0 \neq \lambda \in \mathcal{K}_1^\perp \cap (-\mathcal{K}_2^\perp)$.*

1.1.3. Lie brackets. Given a fixed sequence $\mathbf{X} = (X_1, X_2, \ldots)$ of distinct objects called *variables*, we call *words* the finite ordered strings consisting of the variables X_i, the *left parenthesis* (and the *right parenthesis*), and the comma. We shall use $W(\mathbf{X})$ to denote the set of words. For instance, $X_2X_5X_4$ and $X_3, [X_{13}[,]]X_{61}[$ are words.

Given any word $W \in W(\mathbf{X})$, we use $\operatorname{Seq}(W)$ to denote the word obtained from W by deleting all left and right brackets and all commas. We call *length* of a word $W \in W(\mathbf{X})$, and write $\operatorname{Lgth}(W)$, the cardinality of $\operatorname{Seq}(W)$. For instance, if $W = [[[X_4, X_6], X_7], [X_8, X_9]]$, $\operatorname{Seq}(W) = X_4X_6X_7X_8X_9$ and $\operatorname{Lgth}(W) = 5$.

DEFINITION 1.4. *We call* formal bracket of length 1 *any word of length 1 and we will say that the* bracket of two members W_1, W_2 of $W(\mathbf{X})$ *is the word $[W_1, W_2]$. We call* formal iterated brackets*—or, simply,* brackets*—of $\mathbf{X}$ the elements of the smallest subset $IB(\mathbf{X}) \subseteq W(\mathbf{X})$ such that $IB(\mathbf{X})$ contains the brackets of length 1; if W_1, $W_2 \in IB(\mathbf{X})$, then $[W_1, W_2] \in IB(\mathbf{X})$; for any $b \in IB(\mathbf{X})$,* $\operatorname{Seq}(b) = X_{\mu+1}, \ldots, X_{\mu+m}$ *for some $\mu \geq 0$ and $m > 0$.*

Notice that $\mathrm{Lgth}([b_1, b_2]) = \mathrm{Lgth}(b_1) + \mathrm{Lgth}(b_2)$, for every pair of brackets b_1, b_2. Let b be a bracket of length $m > 1$. Then there exists a unique pair (b_1, b_2) of brackets such that $b = [b_1, b_2]$. The pair (b_1, b_2) is the *factorization* of b, and b_1, b_2 are known, respectively, as the *left factor* and the *right factor* of b. Any substring of b which is itself an iterated bracket is called a *subbracket* of b.

Definition 1.5. *If b is a bracket and S is a subbracket of b, let us define $\mathfrak{d}(S; b)$ by a backward recursion on S: $\mathfrak{d}(b; b) := 0$, $\mathfrak{d}(S_1; b) := \mathfrak{d}(S_2; b) := 1 + \mathfrak{d}([S_1, S_2]; b)$. We shall refer to $\mathfrak{d}(S; b)$ as the* number of differentiations of S in b.

It is easy to prove that $\mathfrak{d}(S; b)$ is equal to the number of right brackets that occur in b to the right of S minus the number of left brackets that occur in b to the right of S. For example, if $b = b(X_3, X_4, X_5) := \big[X_3\,,\,[X_4, X_5]\big]$, then $\mathfrak{d}([X_4, X_5]; b) = 1$, $\mathfrak{d}(X_3; b) = 1$, $\mathfrak{d}(X_4; b) = 2$, $\mathfrak{d}(X_5; b) = 2$.

Definition 1.6 (classes C^{b+k} and $C^{b+k-1,1}$). *Let b be a bracket of degree $m \geq 1$ with $\mathrm{Seq}(b) = X_{\mu+1} \dots X_{\mu+m}$, $\mu \geq 0$. Let $\mathbf{f} = (f_1, \dots, f_\nu)$ be a finite sequence of vector fields with $\nu \geq \mu + m$, and let $k \geq 0$ be an integer. We say that* $\mathbf{f}$ is of class C^{b+k} *if f_j is of class $C^{\mathfrak{d}(X_j;b)+k}$ for each $j \in \{\mu + 1, \dots, \mu + m\}$.*

For example, if $b = [[X_3, X_4], [[X_5, X_6], X_7]]$ and $\mathbf{f} = (f_1, \dots, f_8)$ (so $m = 5$, $\nu = 8$, $\mu = 2$), then $\mathbf{f} \in C^{b+3}$ if and only if $f_3, f_4, f_7 \in C^5$ and $f_5, f_6 \in C^6$. It is easy to verify the following result.

Proposition 1.7. *Let b, k, and $\mathbf{f} = (f_1, \dots, f_\nu)$ be as in Definition* 1.6, *and let (b_1, b_2) be the factorization of b. Then $\mathbf{f} \in C^{b+k}$ if and only if $\mathbf{f} \in C^{b_1+k+1}$ and $\mathbf{f} \in C^{b_2+k+1}$.*

We are now ready to plug vector fields in place of indeterminates in a bracket.

Definition 1.8. *For integers $\mu \geq 0$, $m, \nu \geq 1$, such that $\mu + m \leq \nu$, let b be a formal bracket such that $\mathrm{Seq}(b) = X_{\mu+1} \dots X_{\mu+m}$ and let $\mathbf{f} = (f_1, \dots, f_\nu)$ be a ν-tuple of continuous vector fields. Let S be a subbracket of b. If $\mathrm{Lgth}(S) = 1$, i.e., $S = X_j$ for some $j = \mu+1, \dots, \mu+m$, we define the vector field $S(\mathbf{f})$ as $S(\mathrm{f}) := X_j(\mathbf{f}) := f_j$. If $\mathrm{Lgth}(S) > 1$, $S = [S_1, S_2]$, and either $S \neq b$ or, when $S = b$, one assumes $\mathbf{f} \in C^b$, we set $S(\mathbf{f}) := [S_1(\mathbf{f}), S_2(\mathbf{f})]$. We shall call $S(\mathbf{f})$ the Lie bracket corresponding to the formal bracket S and the sequence $\mathbf{f}$ of vector fields.*

We call the *switch-number* of a formal bracket b the number $r_{_b}$ defined recursively as $r_{_b} := 1$ if $b = X_j$ for some j; $r_{_b} := 2\big(r_{_{b_1}} + r_{_{b_2}}\big)$ if $\mathrm{Lgth}(b) \geq 2$ and $b = [b_1, b_2]$. For instance, the switch-numbers of $\big[[X_3, X_4], [[X_5, X_6], X_7]\big]$ and $[[X_5, X_6], X_7]$ are 28 and 10, respectively. We will call *length* and *switch-number of a Lie bracket* $B = b(f_{\mu+1}, \dots f_{\mu+m})$ the length and the switch-number of the associated formal bracket b, respectively.

2. The optimization problems.

In this section we introduce rigorously the optimization problem over L^1-controls and its embedding in an impulsive problem.

Throughout the paper we shall assume the following set of hypotheses:

(Hp) (i) *the target $\mathfrak{T} \subset \mathbb{R}_+ \times \mathbb{R}^n$ is a closed subset;*

(ii) *the control set $A \subset \mathbb{R}^q$ is compact;*[3]

(iii) *the control set $\mathcal{C} \subseteq \mathbb{R}^m$ is a closed cone of the form $\mathcal{C} = \mathcal{C}_1 \times \mathcal{C}_2$, where $m_1 + m_2 = m$, $\mathcal{C}_1 \subseteq \mathbb{R}^{m_1}$ is a closed cone that contains the lines $\{r\mathbf{e}_i :\ r \in \mathbb{R}\}$*

[3]It is almost obvious that, through minor changes, one can generalize this hypothesis by assuming that the function $a \mapsto (f(x, a), l(x, u, a))$ is bounded for every (x, u).

for $i = 1, \dots, m_1$, *and* $\mathcal{C}_2 \subset \mathbb{R}^{m_2}$ *is a closed cone which does not contain straight lines;*[4]

(iv) *the* drift dynamics $f\colon \mathbb{R}^n \times A \to \mathbb{R}^n$ *is continuous and has continuous partial derivatives* $\frac{\partial f}{\partial x^1}, \dots, \frac{\partial f}{\partial x^n}$*;*

(v) *the vector fields* $g_1, \dots, g_m : \mathbb{R}^n \to \mathbb{R}^n$ *are continuously differentiable;*

(vi) *the* Lagrangian $\ell\colon \mathbb{R}^n \times \mathcal{C} \times A \to \mathbb{R}$ *can be written as* $\ell(x, u, a) = \ell_0(x, a) + \ell_1(x, u)$, *where* ℓ_0 *and the recession function*

$$\hat{\ell}_1(x, w^0, w) := \lim_{r \to w^0} r\ell_1\left(x, \frac{w}{r}\right) \quad \textit{for all } (x, w^0, w) \in \mathbb{R}^n \times \mathbb{R}_+ \times \mathcal{C}$$

is continuous with continuous partial derivatives with respect to x.

(vii) *the* final cost $\Psi : \mathbb{R} \times \mathbb{R}^n \to \mathbb{R}$ *is continuously differentiable.*

Clearly, by standard cutoff methods one might assume the differentiability hypotheses in (iv)–(vii) only on a neighborhood of the extended optimal trajectory considered in Theorems 3.1 and 4.2.

2.1. The original optimal control problem. We define the *set* $\mathcal{U}$ *of strict-sense controls* as $\mathcal{U} := \bigcup_{T>0}\{T\} \times L^1\big([0, T], \mathcal{C} \times A\big)$.

DEFINITION 2.1. *For any strict-sense control* $(T, u, a) \in \mathcal{U}$, *we call* (T, u, a, x, v) *a* strict-sense process *if* (x, v) *is the (unique) Carathéodory solution to*

$$\text{(2.1)} \qquad \begin{cases} \dfrac{dx}{dt}(t) = f(x(t), a(t)) + \displaystyle\sum_{i=1}^{m} g_i(x(t))u^i(t), \\ \dfrac{dv}{dt}(t) = |u(t)|, \\ (x, v)(0) = (\check{x}, 0). \end{cases} \qquad \text{a.e. } t \in [0, T],$$

Furthermore, we say that (T, u, a, x, v) *is* feasible *if* $(T, x(T), v(T)) \in \mathfrak{T} \times [0, K]$.

The original optimal control problem is defined as

$$\text{(P)} \qquad \begin{cases} \text{minimize } \Psi(T, x(T)) + \displaystyle\int_0^T \ell(x(t), u(t), a(t))\, dt \\ \text{over the set of feasible strict-sense processes } (T, u, a, x, v). \end{cases}$$

DEFINITION 2.2. *We call a feasible strict-sense process* $(\bar{T}, \bar{u}, \bar{a}, \bar{x}, \bar{v})$ *a* local strict-sense minimizer *of* (P) *if there exists* $\delta > 0$ *such that*

$$\text{(2.2)} \quad \Psi(\bar{T}, \bar{x}(\bar{T})) + \int_0^{\bar{T}} \ell(\bar{x}(t), \bar{u}(t), \bar{a}(t))dt \le \Psi(T, x(T)) + \int_0^T \ell(x(t), u(t), a(t))dt$$

for every feasible strict-sense process (T, u, a, x, v) *verifying* $\mathrm{d}((T, x, v), (\bar{T}, \bar{x}, \bar{v})) < \delta$, *where* d *is the distance defined in* (1.5). *If relation* (2.2) *is satisfied for all feasible strict-sense processes, we say that* $(\bar{T}, \bar{u}, \bar{a}, \bar{x}, \bar{v})$ *is a* global strict-sense minimizer.

Remark 2.3. By adding the trivial equations $\frac{dx^0}{dt}(t) = 1$, $\frac{d\hat{x}}{dt}(t) = u(t)$, where $\hat{x} = (x^{n+1}, \dots, x^{n+m})$, we can allow ℓ, f, g_i for $i = 1, \dots, m$ to depend on t and on the function $U(t) := \int_0^t u(\tau)\, d\tau$, while Ψ might depend on U as well.

[4]Hypothesis (i) on $\mathcal{C}$ is by no means restrictive, since it can be recovered by the replacement of the single vector fields g_i with suitable linear combinations of $\{g_1, \dots, g_m\}$ and a consequent linear transformation of coordinates in $\mathbb{R}^m$.

2.2. The space-time optimal control problem. We refer to the set $\mathcal{W} := \bigcup_{S>0}\{S\} \times \{(w^0, w, \alpha) \in L^\infty([0, S], \mathbb{R}_+ \times \mathcal{C} \times A) : \operatorname{ess\,inf}(w^0 + |w|) > 0\}$ as the *set of space-time controls*.

Definition 2.4. *For any $(S, w^0, w, \alpha) \in \mathcal{W}$, we say that $(S, w^0, w, \alpha, y^0, y, \beta)$ is a* space-time process *if (y^0, y, β) is the unique Carathéodory solution of*

$$\begin{cases} \dfrac{dy^0}{ds}(s) = w^0(s), \\ \dfrac{dy}{ds}(s) = f(y(s), \alpha(s))w^0(s) + \displaystyle\sum_{i=1}^m g_i(y(s))w^i(s), & \text{a.e. } s \in [0, S], \\ \dfrac{d\beta}{ds}(s) = |w(s)|, \\ (y^0, y, \beta)(0) = (0, \check{x}, 0). \end{cases} \tag{2.3}$$

We say that $(S, w^0, w, \alpha, y^0, y, \beta)$ is feasible *if $(y^0(S), y(S), \beta(S))$ belongs to $\mathfrak{T} \times [0, K]$.*

We define the *extended or space-time problem* as

$$\text{(P}_{\text{s-t}}\text{)} \qquad \begin{cases} \text{minimize } \Psi(y^0(S), y(S)) + \displaystyle\int_0^S \ell^e\big((y, w^0, w, \alpha)(s)\big)ds \\ \text{over feasible space-time processes } (S, w^0, w, \alpha, y^0, y, \beta), \end{cases}$$

where the *extended Lagrangian* ℓ^e is given by

$$\ell^e(x, w^0, w, a) := \ell_0(x, a)w^0 + \hat{\ell}_1(x, w^0, w) \quad \text{for } (x, w^0, w, a) \in \mathbb{R}^n \times \mathbb{R}_+ \times \mathcal{C} \times A.$$

Definition 2.5. *A feasible space-time process $(\bar{S}, \bar{w}^0, \bar{w}, \bar{\alpha}, \bar{y}^0, \bar{y}, \bar{\beta})$ is said to be a* local minimizer for the space-time problem $(\text{P}_{\text{s-t}})$ *if there exists $\delta > 0$ such that*

$$\Psi((\bar{y}^0, \bar{y})(\bar{S})) + \int_0^{\bar{S}} \ell^e((\bar{y}, \bar{w}^0, \bar{w}, \bar{\alpha})(s))\, ds \le \Psi((y^0, y)(S)) + \int_0^S \ell^e((y, w^0, w, \alpha)(s))\, ds \tag{2.4}$$

for all feasible $(S, w^0, w, \alpha, y^0, y, \beta)$ satisfying $\mathrm{d}((S, y^0, y, \beta), (\bar{S}, \bar{y}^0, \bar{y}, \bar{\beta})) < \delta$, where d is as in (1.5). If (2.4) is satisfied for all feasible space-time processes, we call $(\bar{S}, \bar{w}^0, \bar{w}, \bar{\alpha}, \bar{y}^0, \bar{y}, \beta)$ a global space-time minimizer.

Observe that the space-time system (2.3) is *rate independent*. Precisely, given a strictly increasing, surjective, and bi-Lipschitzian function $\sigma : [0, S] \to [0, \tilde{S}]$, $(\tilde{S}, \tilde{w}^0, \tilde{w}, \tilde{\alpha}, \tilde{y}^0, \tilde{y}, \tilde{\beta})$ is a space-time process if and only if $(S, w^0, w, \alpha, y^0, y, \beta)$ given by[5] $(w^0, w) := ((\tilde{w}^0, \tilde{w}) \circ \sigma)\,\frac{d\sigma}{ds}$, $(\alpha, y^0, y, \beta) := (\tilde{\alpha}, \tilde{y}^0, \tilde{y}, \tilde{\beta}) \circ \sigma$ is a space-time process of (2.3) (see [30, sect. 3]). In this case, $(\tilde{S}, \tilde{w}^0, \tilde{w}, \tilde{\alpha}, \tilde{y}^0, \tilde{y}, \tilde{\beta})$ is feasible if and only if $(S, w^0, w, \alpha, y^0, y, \beta)$ is feasible, and the associated costs coincide. Let us call *equivalent* any two space-time processes $(\tilde{S}, \tilde{w}^0, \tilde{w}, \tilde{\alpha}, \tilde{y}^0, \tilde{y}, \tilde{\beta})$, $(S, w^0, w, \alpha, y^0, y, \beta)$ as above. The following result is quite straightforward.

Lemma 2.6. *A feasible space-time process $(\bar{S}, \bar{w}^0, \bar{w}, \bar{\alpha}, \bar{y}^0, \bar{y}, \bar{\beta})$ is a local (resp., a global) minimizer for the space-time problem* $(\text{P}_{\text{s-t}})$ *if and only if every equivalent space-time process is a local (resp., a global) minimizer, and the costs coincide.*

[5]Since every L^1-equivalence class contains Borel measurable representatives, we tacitly assume that all L^1-maps we are considering are Borel measurable when necessary.

As a consequence, the extended problem can be regarded as a problem on the quotient space. Therefore, without loss of generality, one can replace a minimizer with its *canonical parameterization,* defined as follows.

DEFINITION 2.7. *We say that* $(S_c, w_c^0, w_c, \alpha_c, y_c^0, y_c, \beta_c)$ *is the* canonical parameterization *of a space-time process* $(S, w^0, w, \alpha, y^0, y, \beta)$ *if*

$$(w_c^0, w_c) := \Big((w^0, w)\circ\sigma^{-1}\Big)\frac{d\sigma^{-1}}{ds}, \quad (\alpha_c, y_c^0, y_c, \beta_c) := (\alpha, y^0, y, \beta)\circ\sigma^{-1},$$

where $\sigma(s) := \int_0^s \big(w^0(r) + |w(r)|\big)\, dr$*,* $s\in[0,S]$*,* $S_c := \sigma(S) = y^0(S) + \beta(S)$*.*

Note that $w_c^0(s) + |w_c(s)| = 1$ for a.e. $s \in [0, S_c]$. We introduce the subset of *canonical space-time controls*

$$\mathcal{W}_c := \big\{(S, w^0, w, \alpha) \in \mathcal{W} :\ w^0(s) + |w(s)| = 1 \quad \text{a.e. } s\in[0,S]\big\},$$

and call *canonical* also the corresponding space-time processes. One can easily verify that a canonical space-time process coincides with its canonical parameterization.

2.3. The space-time embedding. The original control system (2.1) can be embedded into the space-time system (2.3). Precisely, by the chain rule, given a strict-sense process (T, u, a, x, v), by setting

$$\sigma(t) := \int_0^t (1 + |u(\tau)|)\, d\tau, \quad S := \sigma(T), \quad y^0 := \sigma^{-1} : [0,S]\to[0,T], \tag{2.5}$$

one obtains that

$$(S, w^0, w, \alpha, y^0, y, \beta) := \left(S, \frac{dy^0}{ds}, (u\circ y^0)\cdot\frac{dy^0}{ds}, a\circ y^0, y^0, x\circ y^0, v\circ y^0\right) \tag{2.6}$$

is a (canonical) space-time process with $w^0 > 0$ a.e. Conversely, given a space-time process $(S, w^0, w, \alpha, y^0, y, \beta)$ with $w^0 > 0$ a.e., the function $y^0 : [0,S] \to [0,T]$, $y^0(s) = \int_0^s w^0(\tau)d\tau$ (is increasing and surjective, and) has an absolutely continuous inverse $\sigma : [0,T] \to [0,S]$ (see, e.g., [21]), and

$$(T, u, a, x, v) := \left(T, (w\circ\sigma)\frac{d\sigma}{dt}, \alpha\circ\sigma, y\circ\sigma, \beta\circ\sigma\right) \tag{2.7}$$

is a strict-sense process. Hence, the family of strict-sense processes can be identified with the subfamily of space-time processes $(S, w^0, w, \alpha, y^0, y, \beta)$ having $w^0 > 0$ a.e.

The impulsive, space-time extension of the original optimal control problem consists in allowing the control w^0 to vanish in a set of positive measure. The s-intervals where w^0 vanishes represent the "impulses," namely, the s-intervals of instantaneous evolution of both the control and the state (see, e.g., [15, 30]).[6]

The notions of strict-sense and space-time local minimizer are consistent, as stated in the following easy consequence of Lemma 2.6 above and [5, Prop. 2.7].

LEMMA 2.8. *A process* $(\bar T, \bar u, \bar a, \bar x, \bar v)$ *is a* strict-sense local minimizer *for problem* (P) *if, and only if,* $(\bar S, \bar w^0, \bar w, \bar\alpha, \bar y^0, \bar y, \bar\beta)$ *defined as in* (2.5)–(2.6) *is a space-time local minimizer for* ($\mathrm{P_{s\text{-}t}}$) *among feasible space-time processes with* $w^0 > 0$ *a.e.*

[6]Let us point out that one can equivalently give a t-based description of this extension using bounded variation trajectories as in [6, 32, 8].

3. A first order maximum principle. Due to the rate independence of the space-time control system discussed in subsection 2.2, we can always assume that a local minimizer $(\bar{S}, \bar{w}^0, \bar{w}, \bar{\alpha}, \bar{y}^0, \bar{y}, \bar{\beta})$ for $(\mathrm{P_{s\text{-}t}})$ is canonical.

Let us set

$$W := \{(w^0, w) \in \mathbb{R}_+ \times \mathcal{C} : \; w^0 + |w| = 1\}. \tag{3.1}$$

Let us consider the *unmaximized Hamiltonian* $H : \mathbb{R}^{n+1+n+1+1} \times \mathbb{R}_+ \times \mathcal{C} \times A \to \mathbb{R}$ and the *Hamiltonian* $\mathbf{H} : \mathbb{R}^{n+1+n+1+1} \to \mathbb{R}$, defined by setting

$$H(x, p_0, p, \pi, \lambda, w^0, w, a) := p_0 w^0 + p \cdot \Big(f(x, a) w^0 + \sum_{i=1}^{m} g_i(x) w^i \Big) + \pi |w| - \lambda \ell^e(x, w^0, w, a),$$

$$\mathbf{H}(x, p_0, p, \pi, \lambda) := \max_{(w^0, w, a) \in W \times A} H(x, p_0, p, \pi, \lambda, w^0, w, a).$$

THEOREM 3.1 (first order maximum principle). *Let $(\bar{S}, \bar{w}^0, \bar{w}, \bar{\alpha}, \bar{y}^0, \bar{y}, \bar{\beta})$ be a canonical local minimizer for the space-time problem* $(\mathrm{P_{s\text{-}t}})$. *Then, for every Boltyanski approximating cone Γ of the target $\mathfrak{T}$ at $(\bar{y}^0, \bar{y})(\bar{S})$, there exists a multiplier $(p_0, p, \pi, \lambda) \in \mathbb{R} \times AC\big([0, \bar{S}], \mathbb{R}^n\big) \times \mathbb{R}_- \times \mathbb{R}_+$ verifying the following:*

(i) *(Nontriviality)*

$$(p_0, p, \lambda) \neq (0, 0, 0)\,. \tag{3.2}$$

Furthermore, if $\bar{y}^0(\bar{S}) > 0$, then (3.2) *can be strengthened to*

$$(p, \lambda) \neq (0, 0). \tag{3.3}$$

(ii) *(Nontranversality)*

$$(p_0, p(\bar{S}), \pi) \in \left[-\lambda \left(\frac{\partial \Psi}{\partial t}\big((\bar{y}^0, \bar{y})(\bar{S})\big)\,,\; \frac{\partial \Psi}{\partial x}\big((\bar{y}^0, \bar{y})(\bar{S})\big) \right) - \Gamma^{\perp} \right] \times J, \tag{3.4}$$

where $J := \{0\}$ if $\bar{\beta}(\bar{S}) < K$, and $J := (0, +\infty)$ if $\bar{\beta}(\bar{S}) = K$.[7] *In particular,*

$$\pi = 0 \quad \textit{provided} \quad \bar{\beta}(\bar{S}) < K. \tag{3.5}$$

(iii) *(Adjoint equation) The path p solves, for a.e. $s \in [0, \bar{S}]$,*

$$\frac{dp}{ds}(s) \;=\; -\frac{\partial H}{\partial x}\left(\bar{y}(s), p(s), \pi, \lambda, \bar{w}^0(s), \bar{w}(s), \bar{\alpha}(s)\right). \tag{3.6}$$

(iv) *(First order maximization) For a.e. $s \in [0, \bar{S}]$,*

$$H\Big(\bar{y}(s), p_0, p(s), \pi, \lambda, \bar{w}^0(s), \bar{w}(s), \bar{\alpha}(s)\Big) = \mathbf{H}\Big(\bar{y}(s), p_0, p(s), \pi, \lambda\Big). \tag{3.7}$$

(v) *(Vanishing of the Hamiltonian)*

$$\mathbf{H}\Big(\bar{y}(s), p_0, p(s), \pi, \lambda\Big) = 0 \quad \textit{for all } s \in [0, \bar{S}]. \tag{3.8}$$

[7] It is tacitly meant that, as an approximating cone to the (T, x, v)-target $\mathfrak{T} \times [0, K]$ at $(\bar{y}^0, \bar{y}, \bar{\beta})(\bar{S})$, one chooses $\Gamma \times \mathbb{R}$ if $\bar{\beta}(\bar{S}) < K$ and $\Gamma \times (-\infty, 0]$ if $\bar{\beta}(\bar{S}) = K$. In particular, $(\Gamma \times \mathbb{R})^{\perp} = \Gamma^{\perp} \times \{0\}$ if $\bar{\beta}(\bar{S}) < K$ and $(\Gamma \times (-\infty, 0])^{\perp} = \Gamma^{\perp} \times \mathbb{R}_+$ when $\bar{\beta}(\bar{S}) = K$.

Proof. The Pontryagin maximum principle based on Boltyanski approximating cones (see e.g., [42, 39]) yields the existence of a multiplier $(p_0, p, \pi, \lambda) \in \mathbb{R} \times AC([0, \bar{S}], \mathbb{R}^n) \times \mathbb{R}_- \times \mathbb{R}_+$ verifying the nontransversality condition (3.4), the adjoint equation (3.6), the maximum relation (3.7), the conservation (3.8), and the nontriviality condition $(p_0, p, \pi, \lambda) \neq 0$. So, it remains to prove the strengthened nontriviality condition (3.2). This can be done by using the same elementary arguments as in the proof of [31, Theorem 3.1].[8] ∎

DEFINITION 3.2. *A process* $(\bar{S}, \bar{w}^0, \bar{w}, \bar{\alpha}, \bar{y}^0, \bar{y}, \bar{\beta})$ *is called an* extremal *if it obeys the conditions in Theorem* 3.1 *for some multiplier* (p_0, p, π, λ)*. If there is a choice of the multiplier with* $\lambda = 0$*, then the extremal* $(\bar{S}, \bar{y}^0, \bar{y}, \bar{\beta}, \bar{w}^0, \bar{w}, \bar{\alpha})$ *is called* abnormal*, otherwise it is called* normal*. Finally, the extremal is said to be* strictly abnormal *if every choice of the multiplier* (p_0, p, π, λ) *verifies* $\lambda = 0$*.*

When $\ell_1(x, \cdot)$ is positively 1-homogeneous, so that for any $(x, w^0, w, a) \in \mathbb{R}^n \times \mathbb{R}_+ \times \mathcal{C} \times A$ one has $\ell^e(x, w^0, w, a) = \ell_0(x, a)w^0 + \ell_1(x, w)$, let us define the *drift Hamiltonian* $\mathbf{H}^{(\mathrm{dr})}$ and the *impulse Hamiltonian* $\mathbf{H}^{(\mathrm{imp})}$:

$$\mathbf{H}^{(\mathrm{dr})}\Big(x, p_0, p, \lambda\Big) := \max_{a \in A} \Big\{p_0 + p \cdot f(x, a) - \lambda \ell_0(x, a)\Big\},$$
$$\mathbf{H}^{(\mathrm{imp})}\Big(x, p, \pi, \lambda\Big) := \max_{w \in \mathcal{C}, |w|=1} \left\{p \cdot \sum_{i=1}^{m} g_i(x) w^i + \pi - \lambda \ell_1(x, w)\right\}.$$

COROLLARY 3.3. *Let* $\ell_1(x, \cdot)$ *be positively* 1*-homogeneous and let*

$$(\bar{S}, \bar{w}^0, \bar{w}, \bar{\alpha}, \bar{y}^0, \bar{y}, \bar{\beta})$$

be a canonical extremal obeying the conditions in Theorem 3.1 *for some multiplier* (p_0, p, π, λ)*. Then there exists a zero-measure subset* $\mathcal{N} \subset [0, \bar{S}]$ *such that, for every* $s \in [0, \bar{S}] \setminus \mathcal{N}$*, one has*

$$H\Big(\bar{y}(s), p_0, p(s), \pi, \lambda, \bar{w}^0(s), \bar{w}(s), \bar{\alpha}(s)\Big) = \mathbf{H}\Big(\bar{y}(s), p_0, p(s), \pi, \lambda\Big) = \max\Big\{\mathbf{H}^{(\mathrm{dr})}(\bar{y}(s), p_0, p(s), \lambda)\,,\, \mathbf{H}^{(\mathrm{imp})}(\bar{y}(s), p(s), \pi, \lambda)\Big\} = 0, \tag{3.9}$$

$$\bar{w}^0(s)\Big[p_0 + p(s) \cdot f(\bar{y}(s), \bar{\alpha}(s)) - \lambda \ell_0(\bar{y}(s), \bar{\alpha}(s))\Big] = 0, \tag{3.10}$$

$$p(s) \cdot \sum_{i=1}^{m} g_i(\bar{y}(s)) \bar{w}^i(s) + \pi |\bar{w}(s)| - \lambda \ell_1(\bar{y}(s), \bar{w}(s)) = 0. \tag{3.11}$$

In particular,

(i) *if for some* $s \in [0, \bar{S}] \setminus \mathcal{N}$ *one has* $\mathbf{H}^{(\mathrm{dr})}(\bar{y}(s), p_0, p(s), \lambda) < 0$*, then* $\bar{w}^0(s) = 0$ *and*

$$p(s) \cdot \sum_{i=1}^{m} g_i(\bar{y}(s)) \bar{w}^i(s) + \pi - \lambda \ell_1(\bar{y}(s), \bar{w}(s)) = \mathbf{H}^{(\mathrm{imp})}(\bar{y}(s), p_0, p(s), \pi, \lambda) = 0;$$

(ii) *if for some* $s \in [0, \bar{S}] \setminus \mathcal{N}$ *one has* $\mathbf{H}^{(\mathrm{imp})}(\bar{y}(s), p(s), \pi, \lambda) < 0$*, then* $\bar{w}(s) = 0$ *and*

$$p_0 + p(s) \cdot f(\bar{y}(s), \bar{\alpha}(s)) - \lambda \ell_0(\bar{y}(s), \bar{\alpha}(s)) = \mathbf{H}^{(\mathrm{dr})}(\bar{y}(s), p_0, p(s), \lambda) = 0.$$

[8] The fact that in [31] one makes use of the limiting normal cone instead of the polar of the Boltyanski cone plays no role in the proof of this result.

Proof. By (3.8) it follows that for every $s \in [0, \bar{S}]$, one has

$$p_0 w^0 + p(s) \cdot \Big(f(\bar{y}(s), a) w^0 + \sum_{i=1}^{m} g_i(\bar{y}(s)) w^i \Big) + \pi |w| - \lambda \Big(\ell_0(\bar{y}(s), a) w^0 + \ell_1(\bar{y}(s), w) \Big) \le 0$$

for all $(w^0, w, a) \in W \times A$. Now, by choosing $w = 0$ one gets that $w^0 = 1$ and

$$p_0 + p(s) \cdot f(\bar{y}(s), a) - \lambda \ell_0(\bar{y}(s), a) \le 0 \quad \text{for all } a \in A,$$

while taking $w^0 = 0$ one obtains

$$p(s) \cdot \sum_{i=1}^{m} g_i(\bar{y}(s)) w^i + \pi - \lambda \ell_1(\bar{y}(s), w) \le 0 \quad \text{for all } (w, a) \in \mathcal{C} \times A, \ |w| = 1.$$

Therefore, $\mathbf{H}^{(\mathrm{dr})}(\bar{y}(s), p(s), \pi, \lambda) \le 0$ and $\mathbf{H}^{(\mathrm{imp})}(\bar{y}(s), p(s), \pi, \lambda) \le 0$. In fact, it must be that $\max\big\{ \mathbf{H}^{(\mathrm{dr})}(\bar{y}(s), p_0, p(s), \lambda), \mathbf{H}^{(\mathrm{imp})}(\bar{y}(s), p(s), \pi, \lambda) \big\} = 0$, since, otherwise, both Hamiltonians would be negative, which contradicts (3.8). By taking $\mathcal{N} \subset [0, \bar{S}]$ to be the zero-measure subset such that the first order maximization (3.7) is verified in $[0, \bar{S}] \setminus \mathcal{N}$, we get (3.9). If $s \in [0, \bar{S}] \setminus \mathcal{N}$, by (3.7), (3.8) one has that

$$\begin{aligned} &\bar{w}^0(s) \Big[p_0 + p(s) \cdot f(\bar{y}(s), \bar{\alpha}(s)) - \lambda \ell_0(\bar{y}(s), \bar{\alpha}(s)) \Big] \\ &\qquad + \Big[p(s) \cdot \sum_{i=1}^{m} g_i(\bar{y}(s)) \bar{w}^i(s)) + \pi |\bar{w}(s)| - \lambda \ell_1(\bar{y}(s), \bar{w}(s)) \Big] = 0. \end{aligned}$$

Since the above argument implies that both terms in this equality are nonpositive, they necessarily vanish, namely, (3.10) and (3.11) are verified.

To prove (i), suppose $\mathbf{H}^{(\mathrm{dr})}(\bar{y}(s), p(s), \pi, \lambda) < 0$. Then (3.10) implies $\bar{w}^0(s) = 0$, so that $|\bar{w}(s)| = 1$ and the thesis (i) follows by (3.11). Finally, to prove (ii) assume that $\mathbf{H}^{(\mathrm{imp})}(\bar{y}(s), p(s), \pi, \lambda) < 0$, then $\bar{w}(s) = 0$ due to (3.11) and in view of the positive 1-homogeneity of H w.r.t. (w^0, w). Hence $\bar{w}^0(s) = 1$ and (3.10) yields (ii). ☐

Remark 3.4. Under the same hypotheses of Corollary 3.3, $\mathbf{H}^{(\mathrm{dr})}(\bar{y}(s), p_0, p(s), \lambda) = 0$ for all $s \in [s_1, s_2]$ as soon as $s_1, s_2 \in [0, \bar{S}]$ are such that $\bar{w}^0(s) > 0$ for a.e. $s \in [s_1, s_2] \subseteq [0, \bar{S}]$.

COROLLARY 3.5. *Let $(\bar{S}, \bar{w}^0, \bar{w}, \bar{\alpha}, \bar{y}^0, \bar{y}, \bar{\beta})$ be a canonical extremal for the space-time problem* $(\mathrm{P}_{\text{s-t}})$ *and let (p_0, p, π, λ) be a corresponding multiplier. If*

(3.12) $\qquad \pi = 0$ *and* $\lambda \ell^e(\bar{y}(s), 0, \pm \mathbf{e}_i, a) = 0$ *for all* $s \in [0, \bar{S}]$, $i = 1, \dots, m_1$,[9]

then $p(s) \cdot g_i(\bar{y}(s)) = 0$ *for all* $s \in [0, \bar{S}]$, $i = 1, \dots, m_1$.

Proof. By (3.8) it follows that for every $s \in [0, \bar{S}]$ and $(w^0, w, a) \in W \times A$, one has $p_0 w^0 + p(s) \cdot (f(\bar{y}(s), a) w^0 + \sum_{i=1}^{m} g_i(\bar{y}(s)) w^i) - \lambda \ell^e(\bar{y}(s), w^0, w, a) \le 0$. Therefore, choosing $w^0 = 0$ and $w = \pm \mathbf{e}_i$ for any $i = 1, \dots, m_1$, one gets the thesis. ☐

Remark 3.6. From Theorem 3.1, one has $\pi = 0$ as soon as $\bar{\beta}(\bar{S}) < K$. Moreover, the hypothesis $\lambda \ell^e(\bar{y}(s), 0, \pm \mathbf{e}_i, a) = 0$ is obviously satisfied when the extremal $(\bar{S}, \bar{w}^0, \bar{w}, \bar{\alpha}, \bar{y}^0, \bar{y}, \bar{\beta})$ is abnormal and one chooses $\lambda = 0$, or if $\hat{\ell}_1(x, 0, w) = 0$ for all

[9] By the definition of ℓ^e, it is clear that the quantities $\ell^e(\bar{y}(s), 0, \pm \mathbf{e}_i, a)$ do not depend on a.

$(x, w) \in \mathbb{R}^n \times (\mathbb{R}^{m_1} \times \{0\}^{m_2})$. (This includes, in particular, the case $\ell_1 \equiv 0$, as in the minimum time problem, where $\ell_0 \equiv 1$.)

4. A higher-order maximum principle.

Let us begin with a regularity notion for Lie brackets of the vector fields $g_1, \ldots, g_{m_1}$.

DEFINITION 4.1. *For every integer $k \geq 0$, we say that a vector field B is a C^k-*admissible Lie bracket *if $B = b(F_1, \ldots, F_q)$, where b is a formal bracket and $(F_1, \ldots, F_q)$ is a q-tuple of class C^{b+k} of vector fields in $\{g_1, \ldots, g_{m_1}\}$ (see Definition* 1.6*). We will use $\mathfrak{B}^k$ to denote the set of C^k-admissible Lie brackets of length ≥ 2.*

4.1. Higher-order conditions.

THEOREM 4.2 (higher-order maximum principle). *Assume that hypothesis* (Hp) *is satisfied with $\hat{\ell}_1(\cdot, 0, \cdot) \equiv 0$. Let $(\bar{S}, \bar{w}^0, \bar{w}, \bar{\alpha}, \bar{y}^0, \bar{y}, \bar{\beta})$ be a canonical local minimizer for the space-time problem* ($\mathrm{P_{s\text{-}t}}$) *that verifies $\bar{\beta}(\bar{S}) < K$. Then, for every Boltyanski approximating cone Γ of the target $\mathfrak{T}$ at $(\bar{y}^0, \bar{y})(\bar{S})$, there exists a multiplier $(p_0, p, \pi, \lambda) \in \mathbb{R} \times AC\big([0, \bar{S}], \mathbb{R}^n\big) \times \mathbb{R}_- \times \mathbb{R}_+$ with $\pi = 0$ that satisfies all the conditions of Theorem* 3.1 *and, moreover, verifies*

$$p(s) \cdot g_i(\bar{y}(s)) = 0 \qquad \text{for all } s \in [0, \bar{S}],\ i = 1, \ldots, m_1, \tag{4.1}$$

$$p(s) \cdot B(\bar{y}(s)) = 0 \qquad \text{for all } s \in [0, \bar{S}],\ B \in \mathfrak{B}^0. \tag{4.2}$$

The proof of this theorem is postponed to section 6.

Remark 4.3. Requiring the condition $\hat{\ell}_1(\cdot, 0, \cdot) \equiv 0$ is crucial for the general validity of Theorem 4.2. Otherwise, the variations corresponding to brackets of length $h \geq 2$ would produce a perturbation of order $\varepsilon^{\frac{1}{h}}$ of the cost variable, so having an infinite derivative w.r.t. ε. Since the same variation would produce a change of order ε in the dynamical variables, the separation Theorem 1.3 turns out not to be applicable. Let us point out that any Lagrangian $\ell = \ell(x)$ that is independent of the control (as in the time optimal problem, where $\ell \equiv 1$) verifies the condition $\hat{\ell}_1(\cdot, 0, \cdot) \equiv 0$. Instead, this condition is not satisfied for the length functional, which is positively homogeneous of degree 1 in the u variable. However, as soon as the minimizer is *strictly abnormal* (which means that λ is always equal to 0), one might be able to deduce some results involving Lie brackets for the case $\hat{\ell}_1(\cdot, 0, \cdot) \neq 0$ as well, possibly via some higher-order open mapping argument. This would be similar to what happens in the case of sub-Riemannian geometry [1]. We leave this issue as an open question.

Remark 4.4. Since we obtained the higher-order necessary conditions under the only prerequisite that the involved Lie brackets are continuous, one might wonder to which extent such a regularity hypothesis can be further weakened. For instance, one might prove an extension of Theorem 4 by means of *set-valued Lie brackets* of nonsmooth vector fields, as studied in [36, 37, 22].

In what follows we will use the notation $f_a(\cdot) := f(\cdot, a)$.

COROLLARY 4.5. *Assume that hypothesis* (Hp) *is satisfied with $\hat{\ell}_1(\cdot, 0, \cdot) \equiv 0$, and let $(\bar{S}, \bar{w}^0, \bar{w}, \bar{\alpha}, \bar{y}^0, \bar{y}, \bar{\beta})$ be a canonical local minimizer of* ($\mathrm{P_{s\text{-}t}}$) *that verifies $\bar{\beta}(\bar{S}) < K$. Given a Boltyanski approximating cone Γ of the target $\mathfrak{T}$ at $(\bar{y}^0, \bar{y})(\bar{S})$, let (p_0, p, λ) be a multiplier as in Theorem* 4.2*. Then, for any Lie bracket $B \in \mathfrak{B}^1 \cup \{g_1, \ldots, g_m\}$,* *one has*[10]

$$p(s)\cdot\Big(\big[f_{\bar\alpha(s)},B\big](\bar y(s))\bar w^0(s)+\sum_{j=m_1+1}^{m}\big[g_j,B\big](\bar y(s))\bar w^j(s)\Big) = -\lambda\frac{\partial\ell^e}{\partial x}(\bar y(s),\bar w^0(s),\bar w(s),\bar\alpha(s))\cdot B(\bar y(s)) \tag{4.3}$$

for a.e. $s\in[0,S]$. *In particular, if* $m_1=m$ *and the condition*

$$\lambda\frac{\partial\ell^e}{\partial x}(\bar y(s),\bar w^0(s),\bar w(s),\bar\alpha(s))\cdot B(\bar y(s))=0 \quad \textit{for a.e. } s\in[0,\bar S] \tag{4.4}$$

is satisfied, one obtains

$$p(s)\cdot\Big(\big[f_{\bar\alpha(s)},B\big](\bar y(s))\Big)\bar w^0(s)=0 \quad \textit{for a.e. } s\in[0,\bar S]. \tag{4.5}$$

Proof. Condition (4.3) can be obtained by differentiating (4.1) or (4.2) and remembering that the derivative of p verifies the adjoint equation (3.6). ☐

Remark 4.6. Condition (4.4) is satisfied for all $s\in[0,\bar S]$ in at least two important situations, namely, in the *abnormal case*, i.e., if $\lambda=0$, or when $\ell=\ell_0+\ell_1(u)$ with ℓ_0, ℓ_1 independent of x and $\hat\ell_1(0,w)\equiv 0$ (for instance, in the minimum time problem).

Remark 4.7 (linear systems). Let us consider the linear system

$$\frac{dx}{dt}=Cx+Eu, \qquad u\in\mathbb{R}^m,$$

where C, E are $n\times n$ and $n\times m$ real matrices, respectively. For the vector fields $f(x)=:Cx$, and g_i, where $g_{ij}:=E_{ji}$ for each $i=1,\dots,m$, $j=1,\dots,n$, the conditions involving Lie brackets of the g_i become trivial, since $[g_i,g_j]=0$. However, because of the linearity of $f(x)=Cx$, further higher-order conditions can be trivially deduced under assumption (4.4). Indeed, condition (4.1) reduces to

$$p(s)\cdot E=0 \quad \text{for all } s\in[0,\bar S], \tag{4.6}$$

while, due to (4.4), the adjoint equation now reads $\frac{dp}{dt}=-p\cdot C$. Therefore by differentiating (4.6) $n-1$ times, we get the *additional necessary conditions* $p\cdot[f,g_i]=p\cdot[f,[f,g_i]]=p\cdot[f,[f,[\dots,[f,g_i]\dots]]]=0$ for all $i=1,\dots,m$, which correspond to the $n-1$ matrix relations

$$p\cdot CE=0,\quad p\cdot C^2E=0,\ \dots,\ p\cdot C^{n-1}E=0. \tag{4.7}$$

Remark 4.8. As observed in the introduction, some motivations for studying impulsive systems are to be found in classical mechanics. This is a reason why one might be interested in extending previous results to manifolds. Actually, such an extension does not present any special difficulty, in that the thesis of Theorem 4.2 has a *chart-independent character.*

[10] *I.e.*, B is a C^1-admissible Lie bracket (possibly of length 1); see Definition 4.1.

4.2. Fully impulsive optimal processes. The necessary conditions established in Theorems 3.1 and 4.2 can be used to get information on the structure of optimal trajectories: for instance, one can wonder under which conditions an optimal trajectory is a finite concatenation of impulsive and nonimpulsive paths (as it occurs, e.g., in the examples in section 5 and in [5]). Though an accurate investigation in this direction goes beyond the objectives of this paper, let us highlight some rank conditions that happen to force an optimal process $(\bar{S}, \bar{w}^0, \bar{w}, \bar{\alpha}, \bar{y}^0, \bar{y}, \bar{\beta})$ to be *fully impulsive*. By this we mean that it evolves in zero time, namely, $\bar{y}^0(\bar{S}) = 0$ or, equivalently, $\bar{w}^0 = 0$ a.e. on $[0, \bar{S}]$.

To state our result, we introduce two rank-type assumptions:

(I) C^0-pointwise rank conditions *at* $x \in \mathbb{R}^n$.

$(\mathrm{I.1})_x$ *There exists an integer* $r \geq 0$ *and iterated Lie brackets* $B_1, \ldots, B_r \in \mathfrak{B}^0$ *such that*

$$\mathrm{span}\,\{B_1, \ldots, B_r, g_1, \ldots, g_{m_1}\}\,(x) = \mathbb{R}^n;^{11} \tag{4.8}$$

$(\mathrm{I.2})_x$ *For every* $a \in A$, *there exist integers* $r \geq 0$, $k \geq 0$, *and iterated Lie brackets* $B_1, \ldots, B_r \in \mathfrak{B}^0$, $\hat{B}_1, \ldots, \hat{B}_k \in \mathfrak{B}^1$, *such that*

$$\mathrm{span}\big\{B_1, \ldots, B_r, [f_a, \hat{B}_1], \ldots, [f_a, \hat{B}_k], g_1, \ldots, g_{m_1}, [f_a, g_1], \ldots, [f_a, g_{m_1}]\big\}(x) = \mathbb{R}^n. \tag{4.9}$$

(II) Kalman controllability condition. *The system is linear and the Kalman controllability condition is verified, namely,*

$$\frac{dx}{dt} = Cx + Eu \quad \textit{and} \quad \mathrm{rank}(E \quad CE \quad C^2E \quad \ldots \quad C^{n-1}E) = n,$$

where C, E *are* $n \times n$ *and* $n \times m$ *real matrices, respectively.*

We will consider the following assumption:

(Hp1) *Hypothesis* (Hp) *holds and, moreover,* (i) *the target is time invariant, namely,* $\mathfrak{T} = \mathbb{R} \times \hat{\mathfrak{T}}$ *with* $\hat{\mathfrak{T}} \subseteq \mathbb{R}^n$; (ii) *the final cost* Ψ *is time independent;* (iii) *the Lagrangian* ℓ *is strictly positive and* $\hat{\ell}(\cdot, 0, \cdot) \equiv 0$.

THEOREM 4.9. *Let us assume hypothesis* (Hp1). *Let* $(\bar{S}, \bar{w}^0, \bar{w}, \bar{\alpha}, \bar{y}^0, \bar{y}, \bar{\beta})$ *be a canonical local minimizer for* $(\mathrm{P_{s\text{-}t}})$ *such that* $\beta(\bar{S}) < K$, *and let* (p_0, p, λ) *be a multiplier as in Theorem* 4.2. *If* one *of the options* (a)–(c) *below is verified, then the process* $(\bar{S}, \bar{w}^0, \bar{w}, \bar{\alpha}, \bar{y}^0, \bar{y}, \bar{\beta})$ *is fully impulsive.*

(a) *For every* $s \in [0, \bar{S}]$, *the* C^0*-pointwise rank condition* $(\mathrm{I.1})_{\bar{y}(s)}$ *is verified.*

(b) *For every* $s \in [0, \bar{S}]$, *the* C^0*-pointwise rank condition* $(\mathrm{I.2})_{\bar{y}(s)}$ *is verified, while* $J := \big\{s \in [0, \bar{S}] : \ (\mathrm{I.1})_{\bar{y}(s)}$ *is not verified*$\big\} \neq \emptyset$. *Furthermore,* $m_1 = m$, *and* $\lambda \frac{\partial \ell^e}{\partial x}(\bar{y}(s), \bar{w}^0(s), \bar{w}(s), \bar{\alpha}(s)) = 0$ *for a.e.* $s \in J$.

(c) *The system is linear, the Kalman controllability condition* (II) *is verified, and* $\lambda \frac{\partial \ell^e}{\partial x}(\bar{y}(s), \bar{w}^0(s), \bar{w}(s), \bar{\alpha}(s)) = 0$ *for a.e.* $s \in [0, \bar{S}]$.

Preliminarily, let us prove the following result.

LEMMA 4.10. *Assume* (i) *and* (ii) *in hypothesis* (Hp1), *and let* $\pi = 0$. *Then for any subset* $\mathcal{J} \subseteq [0, T]$ *of positive measure one has neither*

$$p(s) = 0 \quad \textit{and} \quad \ell^e\big(\bar{y}(s), (\bar{y}(s), \bar{w}^0(s), \bar{w}(s), \alpha(s)\big) > 0 \quad \textit{for a.e. } s \in \mathcal{J} \tag{4.10}$$

nor

$$p(s) = 0 \quad \textit{and} \quad \frac{\partial \ell^e}{\partial x}\big(\bar{y}(s), \bar{w}^0(s), \bar{w}(s), \alpha(s)\big) \neq 0 \quad \textit{for a.e. } s \in \mathcal{J}. \tag{4.11}$$

[11] We mean that $\{\zeta_1, \ldots, \zeta_N\} = \emptyset$ as soon as $N = 0$.

Proof. By hypothesis (Hp1)(i), $\Gamma = \mathbb{R} \times \hat{\Gamma}$ with $\hat{\Gamma}$ a cone of $\mathbb{R}^n$. Because of (Hp1)(ii) and the identity $\Gamma^\perp = \{0\} \times \hat{\Gamma}^\perp$, the nontransversality condition yields $p_0 = -\lambda \frac{\partial \Psi}{\partial t}\big(\bar{y}^0(\bar{S}), \bar{y}(\bar{S})\big) + 0 = 0$.

First, let us assume by contradiction that (4.10) is verified on a subset $\mathcal{J} \subseteq [0, \bar{S}]$ of positive measure. Since $(p_0, p(s), \pi) = (0, 0, 0)$ for all $s \in \mathcal{J}$, by (3.8) we obtain that $\lambda \ell^e\big(\bar{y}(s), \bar{w}^0(s), \bar{w}(s), \bar{\alpha}(s)\big) = 0$ for a.e. $s \in \mathcal{J}$, which by (4.10) implies that $\lambda = 0$.

Second, assume that (4.11) is verified on a subset $\mathcal{J} \subseteq [0, \bar{S}]$ of positive measure. We still have $(p_0, p(s), \pi) = (0, 0, 0)$ on $\mathcal{J}$ and, by the adjoint equation, we deduce $\lambda \frac{\partial \ell^e}{\partial x}\big(\bar{y}(s), \bar{w}^0(s), \bar{w}(s), \bar{\alpha}(s)\big) = 0$ for a.e. $s \in \mathcal{J}$ so that by (4.11) one gets again $\lambda = 0$.

Choose a point $\hat{s} \in \mathcal{J}$ so that $p(\hat{s}) = 0$. Since in both cases one has $\lambda = 0$, the adjoint equation is linear in p, which in turn implies that $p \equiv 0$ on $[0, \bar{S}]$. Therefore, $(p_0, p, \pi, \lambda) = 0$, which contradicts the nontriviality condition. ◻

Proof of Theorem 4.9. Observe that, since $\beta(\bar{S}) < K$, one has $\pi = 0$.

Suppose first that hypothesis (a) is verified. For every $s \in [0, \bar{S}]$, by $(\mathrm{I.1})_{\bar{y}(s)}$ there exist an integer $r \geq 0$ and Lie brackets $B_1, \dots, B_r \in \mathfrak{B}^0$ verifying the rank condition (4.8) and, in view of (4.1), (4.2), for all $s \in [0, \bar{S}]$, one has

$$p(s) \cdot g_i(\bar{y}(s)) = 0, \quad p(s) \cdot B_j(\bar{y}(s)) = 0$$

for all $i = 1, \dots, m_1$, $j = 1, \dots, r$. Therefore, we obtain $p(s) = 0$ for all $s \in [0, \bar{S}]$. Assume by contradiction that there exists a subset of positive measure $\mathcal{J} \subseteq [0, \bar{S}]$ such that $\bar{w}^0(s) > 0$ for a.e. $s \in \mathcal{J}$. By the positivity of the function ℓ, this implies that $\ell^e\big(\bar{y}^0(s), \bar{y}(s), \bar{w}^0(s), \bar{w}(s), \bar{\alpha}(s)\big) > 0$ for a.e. $s \in \mathcal{J}$, which in turn is ruled out by Lemma 4.10 above.

Assume now that (b) holds true. If $s \in [0, \bar{S}] \backslash J$, we get $p(s) = 0$ arguing as in the previous case. If J has zero measure, this also implies that $p(s) = 0$ for all $s \in [0, \bar{S}]$. On the contrary, assume that J has positive measure. For almost every $s \in J$ and for $a := \bar{\alpha}(s)$, by $(\mathrm{I.2})_{\bar{y}(s)}$ there exist integers $r, k \geq 0$ and Lie brackets $B_1, \dots, B_r \in \mathfrak{B}^0$, $\hat{B}_1, \dots, \hat{B}_k \in \mathfrak{B}^1$ verifying the rank condition (4.9). Moreover, for almost every $s \in J$, by (4.1), (4.2), and (4.5) one has

$$\begin{aligned} p(s) \cdot g_i(\bar{y}(s)) &= 0, \qquad & p(s) \cdot B_j(\bar{y}(s)) &= 0, \\ p(s) \cdot [f_{\bar{\alpha}(s)}, g_i](\bar{y}(s)) &= 0, \qquad & p(s) \cdot [f_{\bar{\alpha}(s)}, \hat{B}_l](\bar{y}(s)) &= 0 \end{aligned}$$

for all $i = 1, \dots, m$, $j = 1, \dots, r$, $l = 1, \dots, k$. We then deduce that $p(s) = 0$ for almost every $s \in J$. Summing up the above occurrences, by the continuity of p we get $p(s) = 0$ for every $s \in [0, \bar{S}]$. Now assume by contradiction that there exists a subset $\mathcal{J} \subseteq [0, \bar{S}]$ of positive measure such that $\bar{w}^0(s) > 0$ for a.e. $s \in \mathcal{J}$. At this point, the thesis follows arguing exactly as in case (a).

Finally, suppose that (c) holds true. The linear relations (4.6), (4.7) imply $p \equiv 0$, so, in view of the hypothesis $\ell > 0$, one concludes as in cases (b) and (c). ◻

4.3. Nonimpulsive optimal processes. rmimizer is nonimpulsive, that is, the involved process $(\bar{S}, \bar{w}^0, \bar{w}, \bar{\alpha}, \bar{y}^0, \bar{y}, \bar{\beta})$ is such that $\bar{w}^0 > 0$ a.e. on $[0, \bar{S}]$, Theorem 4.2 can be easily formulated in terms of the original time parameter t, once one introduces the *unmaximized Hamiltonian for the original problem* (P), $\mathcal{H}\colon \mathbb{R}^{n+n+1} \times \mathcal{C} \times A \to \mathbb{R}$, given by

$$\mathcal{H}(x, \mathfrak{p}, \lambda, u, a) := \mathfrak{p} \cdot \Big(f(x, a) + \sum_{i=1}^{m} g_i(x) u^i\Big) - \lambda \ell(x, u, a).$$

THEOREM 4.11. *Assume that* (Hp) *is satisfied with* $\hat{\ell}_1(\cdot,0,\cdot) \equiv 0$ *and let* $(\bar{S},\bar{w}^0,\bar{w},\bar{\alpha},\bar{y}^0,\bar{y},\bar{\beta})$ *be a nonimpulsive space-time local minimizer of* ($\mathrm{P_{s\text{-}t}}$)*, that is, it verifies* $\bar{w}^0 > 0$ *a.e. on* $[0,\bar{S}]$*. Moreover, assume that* $\bar{\beta}(\bar{S}) < K$*. Let* $(\bar{T},\bar{u},\bar{a},\bar{x},\bar{v})$ *be the corresponding strict-sense local minimizer of* (P) *(as specified in subsection* 2.3*). Then, for every Boltyanski approximating cone* Γ *of the target* $\mathfrak{T}$ *at* $\big(\bar{T},\bar{x}(\bar{T})\big)$*, there exists a multiplier* $(p_0,\mathfrak{p},\lambda) \in \mathbb{R} \times AC\big([0,\bar{T}],\mathbb{R}^n\big) \times \mathbb{R}_+$ *that satisfies the following conditions:*

$$(\mathfrak{p},\lambda) \neq (0,0),$$

$$(p_0,\mathfrak{p}(\bar{T})) \in \left[-\lambda\left(\frac{\partial\Psi}{\partial t}\big(\bar{T},\bar{x}(\bar{T})\big)\,,\ \frac{\partial\Psi}{\partial x}\big(\bar{T},\bar{x}(\bar{T})\big)\right) - \Gamma^{\perp}\right],$$

$$\frac{d\mathfrak{p}}{dt}(t) = -\frac{\partial\mathcal{H}}{\partial x}\big(\bar{x}(t),\mathfrak{p}(t),\lambda,\bar{u}(t),\bar{a}(t)\big), \quad \text{a.e. on } [0,\bar{T}],$$

$$\mathcal{H}\big(\bar{x}(t),\mathfrak{p}(t),\lambda,\bar{u}(t),\bar{a}(t)\big) = \max_{u\in\mathcal{C},a\in A}\mathcal{H}\big(\bar{x}(t),\mathfrak{p}(t),\lambda,u,a\big) \equiv -p_0, \quad \text{a.e. on } [0,\bar{T}],$$

$$\mathfrak{p}(t)\cdot g_i(\bar{x}(t)) = 0 \quad \text{for all } t\in[0,\bar{T}],\ i=1,\dots,m_1,$$

$$\mathfrak{p}(t)\cdot B(\bar{x}(t)) = 0 \quad \text{for all } t\in[0,\bar{T}],\ B\in\mathfrak{B}^0.$$

Furthermore, when the final time T *is free,* $p_0 = 0$.

The proof is obtained as a straightforward consequence of Theorem 4.2 as soon as one sets $\mathfrak{p}(t) := p(\bar{\sigma}(t))$ with $\bar{\sigma}$ being as in (2.5), and applies the chain rule to the involved expressions.

Remark 4.12. If one can guarantee that there is *no infimum gap* between the original problem and the extended one,[12] Theorem 4.11 provides a maximum principle for a local strict-sense minimizer, i.e., a local minimizer of the original problem.[13] Incidentally, this would improve the necessary conditions established in [19] for a minimizer of the ordinary optimal time problem with unbounded controls. Indeed, the time-optimal problem is just an example of the wider class of problems investigated here. Furthermore, unlike [19], we do not assume that ($m = m_1$ and) the Lie algebra generated by the vector fields $g_1,\dots,g_m$ has constant dimension. Actually we do not even assume the existence of the Lie algebra, for any vector field g_i may well be just of class C^{r_1} for some finite $r_i > 0$.

5. An example.

5. An example. Higher-order necessary conditions may be useful to rule out the optimality of a space-time process that verifies the first order maximum principle of Theorem 3.1. Let us illustrate this situation with an example.

Consider the optimization problem

$$\text{(5.1)} \qquad \begin{cases} \text{minimize } \ \Psi(x(1)) + \displaystyle\int_0^1 \ell(x(t))\,dt \\ \text{over the processes } (u,x):[0,1]\to\mathbb{R}^3\times\mathbb{R}^4 \text{ such that} \\ \dfrac{dx}{dt} = \displaystyle\sum_{i=1}^{3} g_i(x)u^i, \qquad \text{a.e. } t\in[0,1], \\ x(0) = (1,0,0,0), \quad x(1)\in\mathfrak{T}, \end{cases}$$

[12] Let us remind the reader that sufficient conditions for the absence of an infimum gap were established, e.g., in [4], in terms of *fast controllability*, and in [31], as a normality assumption.

[13] Of course, there is also the possibility that Theorem 4.11 provides a maximum principle for a local strict-sense minimizer *in the general case,* i.e., when infimum gaps may occur. This is a nontrivial issue and we leave it as an open question.

where $\Psi(x) := x^4$, $\ell(x) := (x^3)^2 + (x^4)^2$, $\mathfrak{T} := \{0\} \times \{0\} \times \mathbb{R} \times \left[-\frac{1}{2}, +\infty\right)$,

$$g_1(x) := \frac{\partial}{\partial x^1} - \frac{1}{2}x^2 \frac{\partial}{\partial x^3}, \quad g_2(x) := \frac{\partial}{\partial x^2} + \frac{1}{2}x^1 \frac{\partial}{\partial x^3}, \quad g_3(x) := x^3 \frac{\partial}{\partial x^4}.$$

If we add the equation $\frac{dx^5}{dt} = (x^3)^2 + (x^4)^2$ and redefine x by setting $x = (x^1, \dots, x^5)$, the corresponding space-time problem reads

$$\text{(5.2)} \qquad \begin{cases} \text{minimize } \ \hat{\Psi}(y(S)) \\ \text{over } S > 0, \ (w^0, w, y^0, y) : [0, S] \to [0, +\infty) \times \mathbb{R}^3 \times \mathbb{R}^5 \text{ such that} \\ \dfrac{dy^0}{ds} = w^0, \\ \dfrac{dy}{ds} = \hat{f}(y)w^0 + \displaystyle\sum_{i=1}^{3} \hat{g}_i(y)w^i, \qquad \text{a.e. } s \in [0, S], \\ y^0(0) = 0, \ y(0) = (1, 0, 0, 0, 0), \ \ (y^0, y)(S) \in \{1\} \times \mathfrak{T} \times \mathbb{R}, \end{cases}$$

where $\hat{\Psi}(x) = x^4 + x^5$, $\hat{f}(x) = ((x^3)^2 + (x^4)^2)\frac{\partial}{\partial x^5}$, and $\hat{g}_1, \hat{g}_2, \hat{g}_3$ are obtained from g_1, g_2, g_3 to $\mathbb{R}^5$, respectively, by adding a zero fifth component.

Let us consider the feasible space-time process $(\hat{S}, \hat{w}^0, \hat{w}, \hat{y}^0, \hat{y})$, where $\hat{S} := 2$,

$$(\hat{w}^0, \hat{w}^1, \hat{w}^2, \hat{w}^3)(s) := \left(\frac{1}{2}, -\frac{1}{2}, 0, 0\right), \qquad \text{a.e. } s \in [0, 2],$$

$$(\hat{y}^0, \hat{y})(s) = \left(\frac{s}{2}, 1 - \frac{s}{2}, 0, 0, 0, 0\right), \qquad \text{a.e. } s \in [0, 2].$$

The final cost is clearly $\hat{\Psi}(\hat{y}(\hat{S})) = 0$. In what follows, we prove that $(\hat{S}, \hat{w}^0, \hat{w}, \hat{y}^0, \hat{y})$ is an *extremal* for the space-time problem (5.2)—that is, it verifies conditions (i)–(v) of Theorem 3.1 for some multiplier—but there is no nonzero multiplier for which all the necessary conditions in Theorem 4.2 are met. Choose $\Gamma := \mathbb{R}^3 \times \{(0, 0, 0)\}$ as a Boltyanskii approximating cone to $\{1\} \times \mathfrak{T} \times \mathbb{R}$ at the right endpoint point $(\hat{y}^0, \hat{y})(\hat{S})$. The adjoint equation and the nontransversality condition are written

$$\begin{cases} \dfrac{dp_0}{ds} = 0, \\ \dfrac{dp_1}{ds} = -\dfrac{1}{2}p_3\, \hat{w}^2, \\ \dfrac{dp_2}{ds} = \dfrac{1}{2}p_3\, \hat{w}^1, \\ \dfrac{dp_3}{ds} = -p_4\, \hat{w}^3 - 2p_5\, \hat{y}^3 \hat{w}^0, \\ \dfrac{dp_4}{ds} = -2p_5\, \hat{y}^4 \hat{w}^0, \\ \dfrac{dp_5}{ds} = 0, \end{cases} \qquad \text{a.e. } s \in [0, \hat{S}],$$

$$(p_0, p(2)) \in -\lambda\{(0, 0, 0, 0, 1, 1)\} - \mathbb{R}^3 \times \{(0, 0, 0)\}$$

with $\lambda \geq 0$. Therefore, p_0, p_1, and p_2 are arbitrary real constants, $p_3 \equiv 0$, while $p_4 \equiv p_5 \equiv -\lambda$. Moreover, by (the first order conditions) (3.7) and (3.8) we get

$$p_0 \hat{w}^0 + p_1 \hat{w}^1 + p_2 \hat{w}^2 = p_0\, \frac{1}{2} - p_1\, \frac{1}{2} = 0,$$

and

$$p_0 w^0 + p_1 w^1 + p_2 w^2 \leq 0$$

for all $(w^0, w^1, w^2, w^3) \in [0, +\infty) \times \mathbb{R}^3$ with $w^0 + |w| = 1$. Hence we obtain

$$(p_0, p, \lambda) = \big(0, 0, 0, 0, -\lambda, -\lambda, \lambda\big). \tag{5.3}$$

For any positive λ, the process $(\hat{S}, \hat{w}^0, \hat{w}, \hat{y}^0, \hat{y})$ and the nontrivial multiplier (p_0, p, λ) given in (5.3) satisfy the first order maximum principle of Theorem 3.1. However, since

$$\big[[\hat{g}_1, \hat{g}_2], \hat{g}_3\big] = \frac{\partial}{\partial x^4},$$

the higher order condition (4.2) implies $p_4 \equiv 0$, so that $\lambda = 0$, which in turn gives $(p_0, p, \lambda) = \big(0, 0, 0, 0, 0, 0\big)$. As a conclusion, $(\hat{S}, \hat{w}^0, \hat{w}, \hat{y}^0, \hat{y})$ cannot be a minimizer, since there is no nontrivial multiplier verifying the higher-order necessary conditions of Theorem 4.2.

On the other hand, setting $\bar{S} := 2 + \sqrt{4\pi + 1}$ and $\rho := \frac{1}{\sqrt{4\pi+1}}$, we easily see that the space-time control $(\bar{S}, \bar{w}^0, \bar{w})$ defined by

$$(\bar{w}^0, \bar{w})(s) \;:=\; \begin{cases} \left(\dfrac{1}{2}, -\dfrac{1}{2}, 0, 0\right), & s \in [0, 2[, \\[2ex] \rho\Big(0, 2\sqrt{\pi}\,\sin\left(2\pi\rho(s-2)\right), 2\sqrt{\pi}\,\cos\left(2\pi\rho(s-2)\right), 1\Big), & s \in [2, \bar{S}], \end{cases}$$

is optimal. Indeed, the corresponding space-time trajectory $(\bar{y}^0, \bar{y})$ verifies, in particular, $\bar{y}^0(s) = 1$ for all $s \in [2, \bar{S}]$, $\bar{y}^1(\bar{S}) = \bar{y}^2(\bar{S}) = 0$, and $\bar{y}^4(\bar{S}) = -1/2$. Hence the final constraint is satisfied and the final cost $\Psi(\bar{y}(\bar{S}))$ is equal $-1/2$, which is *the minimum possible value.* Choosing $\Gamma := \mathbb{R}^3 \times \{0\} \times (-\infty, 0] \times \{0\}$ as the Boltyanskii approximating cone to the enlarged constraint $\{1\} \times \mathfrak{T} \times \mathbb{R}$ at the final point $(\hat{y}^0, \hat{y})(\hat{S})$, the nontransversality condition (3.4) now reads

$$\big(p_0, p(\bar{S})\big) \in -\lambda\{(0, 0, 0, 0, 1, 1)\} - \mathbb{R}^3 \times \{0\} \times (-\infty, 0] \times \{0\}$$

with $\lambda \geq 0$. This yields

$$\big(p_0, p(\bar{S})\big) = (c_0, c_1, c_2, 0, -\lambda + c_4, -\lambda)$$

with c_0, c_1, $c_2 \in \mathbb{R}$ and $c_4 \geq 0$. Choosing $c_0 = c_1 = c_2 = 0$, $c_4 = 1$, and $\lambda = 1$, we get the nontrivial multiplier

$$(p_0, p_1, p_2, p_3, p_4, p_5, \lambda) \equiv (0, 0, 0, 0, 0, -1, 1),$$

which satisfies all the necessary conditions in Theorem 4.2 (and agrees with the strengthened nontriviality condition (3.3), i.e., $(p, \lambda) \neq (0, 0)$, that is in force since $\bar{y}^0(\bar{S}) = 1 > 0$). In particular, the higher-order condition (4.2) is verified, since the fifth component of the vector fields g_1, g_2, g_3 and of all the elements of the Lie algebra generated by $\{g_1, g_2, g_3\}$ is equal to zero.

6. Proof of Theorem 4.2.

Let $(\bar{S}, \bar{w}^0, \bar{w}, \bar{\alpha}, \bar{y}^0, \bar{y}, \bar{\beta})$ be a canonical local minimizer of $(P_{s\text{-}t})$ verifying $\bar{\beta}(\bar{S}) < K$, that we will call the *reference process.* Throughout this section $\hat{\ell}_1(\cdot, 0, \cdot) \equiv 0$, as required in the statement of Theorem 4.2. Moreover, we set

$$F^e(x, w^0, w, a) := f(x, a) w^0 + \sum_{i=1}^{m} g_i(x) w^i \quad \text{for all } (x, w^0, w, a) \in \mathbb{R}^n \times \mathbb{R}_+ \times \mathcal{C} \times A,$$

$$\bar{F}^e(s) := F_e(\bar{y}, \bar{w}^0, \bar{w}, \bar{\alpha})(s), \quad \bar{\ell}^e(s) := \ell^e(\bar{y}, \bar{w}^0, \bar{w}, \bar{\alpha})(s) \quad \text{for a.e. } s \in [0, \bar{S}].$$

The proof will be divided into several steps. First, following a time-rescaling procedure, we transform problem ($\mathrm{P_{s\text{-}t}}$) into a problem on the fixed interval $[0,\bar S]$. At this point, we define two classes of variations, comprising standard *needle variations* and *bracket-like variations*, the latter being produced by suitable *instantaneous* perturbations of the reference process. By using appropriate powers of the perturbation parameter ε, all these variations turn out to be of the same order ε. Once this is done, the proof proceeds by some set-separation arguments.

6.1. Rescaling the problem.

DEFINITION 6.1. *Fix $\rho>0$. For any $(S,w^0,w,\alpha,\zeta)\in\mathcal W\times L^\infty([0,\bar S],[-\rho,\rho])$, we say that $(S,w^0,w,\alpha,\zeta,y^0,y,y^\ell,\beta)$ is a* rescaled (space-time) process *if (y^0,y,y^ℓ,β) is the unique Carathéodory solution of*

$$\begin{cases} \dfrac{dy^0}{ds}=w^0(1+\zeta),\\ \dfrac{dy}{ds}=F^e(y,w^0,w,\alpha)(1+\zeta),\\ \dfrac{dy^\ell}{ds}=\ell^e(y,w^0,w,\alpha)(1+\zeta), & a.e.\ s\in[0,\bar S],\\ \dfrac{d\beta}{ds}=|w|(1+\zeta),\\ (y^0,y,y^\ell,\beta)(0)=(0,\check x,0,0),\end{cases}\tag{6.1}$$

and $(S,w^0,w,\alpha,y^0,y,y^\ell,\beta)$ is called feasible *if $(y^0(S),y(S),\beta(S))\in\mathfrak T\times[0,K]$.*

We define the *rescaled* space-time optimization problem as

$$(\mathrm{P_e})\qquad \begin{cases}\text{minimize } \left\{\Psi((y^0,y)(\bar S))+y^\ell(\bar S)\right\}\\ \text{over feasible rescaled processes } (\bar S,w^0,w,\alpha,\zeta,y^0,y,y^\ell,\beta).\end{cases}$$

It is easy to see that, for $\rho>0$ sufficiently small, the reference process, regarded as a process $(\bar S,\bar w^0,\bar w,\bar\alpha,0,\bar y^0,\bar y,\bar y^\ell,\bar\beta)$ of (6.1), is a local minimizer for ($\mathrm{P_e}$), which is a fixed end-time problem.[14] Since the proof involves only space-time trajectories which are close to the reference space-time trajectory $(\bar y^0,\bar y)$ and the controls assume values in a compact set, using standard truncation and mollification arguments, we can assume the following hypothesis:

(Hp)* *All the assumptions in* (Hp) *are verified and, moreover, ℓ^e, f, the g_i, their partial derivatives $\frac{\partial\ell^e}{\partial x^j}$, $\frac{\partial f}{\partial x^j}$, $\frac{\partial g_i}{\partial x^j}$, and all the iterated brackets $B\in\mathfrak B^0$ (as defined in Definition* 4.1*) are uniformly continuous and bounded.*

Hypothesis (Hp)* guarantees that for any $(w^0,w,\alpha,\zeta)\in L^\infty([0,\bar S],W\times A\times[-\rho,\rho])$ there exists a unique solution (y^0,y,y^ℓ,β) to (6.1), defined on the whole interval $[0,\bar S]$. Moreover, the input-output map

$$\Phi: L^\infty\left([0,\bar S],W\times A\times[-\rho,\rho]\right)\to C^0([0,\bar S],\mathbb R\times\mathbb R^n\times\mathbb R\times\mathbb R),\tag{6.2}$$

which associates with any control the corresponding solution to (6.1), turns out to be Lipschitz continuous when one considers the sup-norm over the set of trajectories, and the distance $\tilde d\big((w^0,w,\alpha,\zeta),(\tilde w^0,\tilde w,\tilde\alpha,\tilde\zeta)\big):=\mathrm{meas}\left\{(w^0,w,\alpha,\zeta)(s)\neq(\tilde w^0,\tilde w,\tilde\alpha,\tilde\zeta)(s):\ s\in[0,\bar S]\right\}$ for every pair (w^0,w,α,ζ), $(\tilde w^0,\tilde w,\tilde\alpha,\tilde\zeta)$ of controls.

[14] I.e., there exists $\delta>0$ such that $\Psi((\bar y^0,\bar y)(\bar S))+\bar y^\ell(\bar S)\le\Psi((y^0,y)(\bar S))+y^\ell(\bar S)$ for all feasible processes $(\bar S,w^0,w,\alpha,\zeta,y^0,y,y^\ell,\beta)$ satisfying $\mathrm d\big((\bar S,y^0,y,y^\ell,\beta),(\bar S,\bar y^0,\bar y,\bar y^\ell,\bar\beta)\big)<\delta$.

6.2. Needle and bracket-like approximations.

DEFINITION 6.2 (variation generator). *Let us define the* set of variation generators *as*

$$\mathfrak{V} := (W \times A \times [-\rho,\rho]) \bigcup \mathfrak{B}^0.^{15}$$

More specifically, any $\mathbf{c} = (w^0, w, a, \zeta) \in W \times A \times [-\rho,\rho]$ *will be called a* needle variation generator*, or a* variation generator of length 1*, while any bracket* $\mathbf{c} = B \in \mathfrak{B}^0$ *of length* h (≥ 2) *will be called a* bracket-like variation generator of length h.

With every variation generator $\mathbf{c}$ and with each instant $\bar{s} \in (0, \bar{S})$, we now associate an infinitesimal space-time variation of the reference trajectory $(\bar{y}^0, \bar{y}, \bar{y}^\ell, \bar{\beta})$, whose y-component coincides with either a standard needle variation or a Lie bracket. As usual, the needle variations will be considered at *Lebesgue points* of an appropriate associated function as given in the next definition.[16]

DEFINITION 6.3. *We will use* $(0,\bar{S})_{\mathrm{Leb}}$ *to denote the full measure subset of* $(0,\bar{S})$ *consisting of the Lebesgue points of the function* $s \mapsto \big(\bar{w}^0(s), \bar{F}^e(s), \bar{\ell}^e(s), |\bar{w}|(s)\big)$, $s \in [0,\bar{S}]$.

DEFINITION 6.4. (needle variation). *For every* $\bar{s} \in (0,\bar{S})_{\mathrm{Leb}}$ *and every needle variation generator* $\mathbf{c} = (w^0, w, a, \zeta)$*, consider the vector*

$$\begin{pmatrix} \mathbf{v}^0_{\mathbf{c},\bar{s}} \\ \mathbf{v}_{\mathbf{c},\bar{s}} \\ \mathbf{v}^\ell_{\mathbf{c},\bar{s}} \\ \mathbf{v}^v_{\mathbf{c},\bar{s}} \end{pmatrix} := \begin{pmatrix} w^0(1+\zeta) - \bar{w}^0(\bar{s}) \\ F^e(\bar{y}(\bar{s}), w^0, w, a)(1+\zeta) - \bar{F}^e(\bar{s}) \\ \ell^e(\bar{y}(\bar{s}), w^0, w, a)(1+\zeta) - \bar{\ell}^e(\bar{s}) \\ |w|(1+\zeta) - |\bar{w}(\bar{s})| \end{pmatrix}. \tag{6.3}$$

(Bracket-like variation)*. For every* $\bar{s} \in (0,\bar{S})$ *and every bracket-like variation generator* $\mathbf{c} = B \in \mathfrak{B}^0$*, one sets*

$$\begin{pmatrix} \mathbf{v}^0_{\mathbf{c},\bar{s}} \\ \mathbf{v}_{\mathbf{c},\bar{s}} \\ \mathbf{v}^\ell_{\mathbf{c},\bar{s}} \end{pmatrix} := \begin{pmatrix} 0 \\ \frac{B(\bar{y}(\bar{s}))}{r_{_B}^h} \\ 0 \end{pmatrix}, \tag{6.4}$$

where $r_{_B}$ *is defined as in subsection* 1.1.

DEFINITION 6.5 (needle approximation). *Let* $\mathbf{c} = (w^0, w, a, \zeta)$ *be a needle variation generator and let* $\bar{s} \in (0,\bar{S})$*. For any control* $(\tilde{w}^0, \tilde{w}, \tilde{\alpha}, \tilde{\zeta})$ *belonging to the set* $L^\infty\big([0,\bar{S}], W \times A \times [-\rho,\rho]\big)$*, the family* $\{(\tilde{w}^0, \tilde{w}, \tilde{\alpha}, \tilde{\zeta})^\varepsilon_{\mathbf{c},\bar{s}} : \ \varepsilon \in (0,\bar{s})\}$*, defined by*

$$(\tilde{w}^0, \tilde{w}, \tilde{\alpha}, \tilde{\zeta})^\varepsilon_{\mathbf{c},\bar{s}}(s) := \begin{cases} (w^0, w, a, \zeta) & \text{if } s \in [\bar{s}-\varepsilon, \bar{s}], \\ (\tilde{w}^0, \tilde{w}, \tilde{\alpha}, \tilde{\zeta})(s) & \text{if } s \in [0, \bar{s}-\varepsilon) \cup (\bar{s}, \bar{S}], \end{cases} \tag{6.5}$$

is called a needle control approximation of $(\tilde{w}^0, \tilde{w}, \tilde{\alpha}, \tilde{\zeta})$ at $\bar{s}$ associated with $\mathbf{c}$.

In order to state Lemma 6.6 below—which is a standard result (see, e.g., [34])—, for any $\tilde{y} := (y^0, y, y^\ell, \beta) \in \mathbb{R} \times \mathbb{R}^n \times \mathbb{R} \times \mathbb{R}$ and any $(w^0, w, a) \in W \times A$, let us set

$$\tilde{F}(\tilde{y}, w^0, w, a) := (w^0, F^e(y, w^0, w, a), \ell^e(y, w^0, w, a), |w|)$$

[15] We recall that $W = \{(w^0, w) \in \mathcal{R}_+ \times \mathcal{C} : \ w^0 + |w| = 1\}$ and $\mathfrak{B}^0$ is the set of C^0-admissible iterated Lie brackets of length ≥ 2, as in Definition 4.1.

[16] Given $F \in L^1([a,b], \mathbb{R}^N)$, $s \in (a,b)$ is called a *Lebesgue point* if $\lim_{\delta\to 0} \frac{1}{\delta} \int_{s-\delta}^{s+\delta} |F(\sigma) - F(s)| d\sigma = 0$. By the Lebesgue differentiation theorem, the set of Lebesgue points has measure $b-a$.

and use $\tilde{M}(\cdot,\cdot)$ to denote the *fundamental matrix* of the *variational equation*

$$\frac{d\tilde{V}}{ds}(s) = \frac{\partial \tilde{F}}{\partial \tilde{x}}(\bar{y}^0(s), \bar{y}(s), \bar{y}^\ell(s), \bar{\beta}(s), \bar{w}^0(s), \bar{w}(s), \bar{\alpha}(s)) \cdot \tilde{V}(s), \quad \text{a.e. } s \in [0, \bar{S}]. \tag{6.6}$$

Namely, for each vector $\tilde{v} := \left(v^0, v, v^\ell, v^v\right) \in \mathbb{R}^{1+n+1+1}$ and each $s_1 \in [0, \bar{S}]$, the function $\tilde{V}(\cdot) := \tilde{M}(\cdot, s_1)\tilde{v}$ is the solution of (6.6) with initial condition $\tilde{V}(s_1) = \left(V^0, V, V^\ell, V^v\right)(s_1) = \tilde{v}$. It is straightforward to check that, for all $s \in [0, \bar{S}]$, one has

- $\tilde{M}_{0,j}(s, s_1) = \tilde{M}_{j,0}(s, s_1) = \delta_{0,j}$ for $j = 0, \dots, n+2$,
- $\tilde{M}_{n+2,j}(s, s_1) = M_{j,n+2}(s, s_1) = \delta_{n+2,j}$ for $j = 0, \dots, n+2$,
- $\tilde{M}_{i,r}(s, s_1) = M_{i,r}(s, s_1)$ for $i, r = 1, \dots, n$,
- $\tilde{M}_{r,n+1}(s, s_1) = \mu_r(s, s_1) := \int_{s_1}^{s} \sum_{j=1}^{n} \frac{\partial \ell^e}{\partial x^j}((\bar{y}, \bar{w}^0, \bar{w}, \bar{\alpha})(\sigma)) \cdot M_{j,r}(s, \sigma) d\sigma$ for $r = 1, \dots, n$,
- $\tilde{M}_{n+1,n+1}(s, s_1) = 1$,

where $M(\cdot,\cdot)$ denotes the fundamental matrix of the state-variational equation in $\mathbb{R}^n$

$$\frac{dV}{ds}(s) = \frac{\partial F^e}{\partial x}(\bar{y}(s), \bar{w}^0(s), \bar{w}(s), \bar{\alpha}(s)) \cdot V(s), \quad \text{a.e. } s \in [0, \bar{S}]. \tag{6.7}$$

Lemma 6.6 (asymptotics of needle variations). *Assume that $\bar{s} \in (0, \bar{S})_{\text{Leb}}$. For every needle variation generator $\mathbf{c} = (w^0, w, a, \zeta) \in W \times A \times [-\rho, \rho]$ and for every $s \in (\bar{s}, \bar{S}]$, setting $\mu(s, \bar{s}) := (\mu_1, \dots, \mu_n)(s, \bar{s})$ we get*

$$\begin{pmatrix} y^{0\varepsilon}(s) - \bar{y}^0(s) \\ y^{\varepsilon}(s) - \bar{y}(s) \\ y^{\ell\varepsilon}(s) - \bar{y}^\ell(s) \\ \beta^{\varepsilon}(s) - \bar{\beta}(s) \end{pmatrix} = \varepsilon \tilde{M}(s, \bar{s}) \cdot \begin{pmatrix} \mathbf{v}^0_{\mathbf{c},\bar{s}} \\ \mathbf{v}_{\mathbf{c},\bar{s}} \\ \mathbf{v}^\ell_{\mathbf{c},\bar{s}} \\ \mathbf{v}^v_{\mathbf{c},\bar{s}} \end{pmatrix} + o(\varepsilon) = \varepsilon \begin{pmatrix} \mathbf{v}^0_{\mathbf{c},\bar{s}} \\ M(s, \bar{s}) \cdot \mathbf{v}_{\mathbf{c},\bar{s}} \\ \mu(s, \bar{s}) \cdot \mathbf{v}_{\mathbf{c},\bar{s}} + \mathbf{v}^\ell_{\mathbf{c},\bar{s}} \\ \mathbf{v}^v_{\mathbf{c},\bar{s}} \end{pmatrix} + o(\varepsilon), \tag{6.8}$$

where $(y^{0\varepsilon}, y^{\varepsilon}, y^{\ell\varepsilon}, \beta^{\varepsilon})$ denotes the solution of system (6.1) *corresponding to the needle control approximation $(\bar{w}^0, \bar{w}, \bar{\alpha}, 0)^{\varepsilon}_{\mathbf{c},\bar{s}}$ of $(\bar{w}^0, \bar{w}, \bar{\alpha}, 0)$ at $\bar{s}$ associated with $\mathbf{c}$.*

Bracket-like approximations, which can be performed in various ways (see, e.g., [2, 26, 13, 12, 19] and references therein), are here based on the following result.

Lemma 6.7. *Assume* (Hp)* *with $\hat{\ell}_1(\cdot, 0, \cdot) \equiv 0$. Fix a point $(\tilde{y}^0, \tilde{y}, \tilde{y}^\ell, \tilde{\beta}) \in \mathbb{R} \times \mathbb{R}^n \times \mathbb{R} \times \mathbb{R}$ and some $a \in A$. For every Lie bracket $B \in \mathfrak{B}^0$ of length h, there is $\bar{\varepsilon} > 0$ such that, for any $s \in (0, \bar{\varepsilon}^{1/h}]$, there exists a piecewise constant control $(w^0_{\mathbf{c},s}, w_{\mathbf{c},s})$ with $w^0_{\mathbf{c},s}(\sigma) = 0$ for all $\sigma \in [0, s]$, $w_{\mathbf{c},s} : [0, s] \to \{\pm \mathbf{e}_1, \dots, \pm \mathbf{e}_{m_1}\}$, verifying*

$$(y^0, y^\ell)(\sigma) = (\tilde{y}^0, \tilde{y}^\ell), \qquad \boldsymbol{\beta}(\sigma) = \tilde{\beta} + \sigma \quad \textit{for all } \sigma \in [0, s], \tag{6.9}$$

$$y(s) = \tilde{y} + \left(\frac{s}{r_{B}}\right)^h B(\tilde{y}) + o(s^h), \tag{6.10}$$

where r_B is the switch-number introduced in subsection 1.1 *and $(y^0, y, y^\ell, \boldsymbol{\beta})$ denotes the solution to the space-time control system in* (6.1) *corresponding to the control*[17] *$(w^0_{\mathbf{c},s}, w_{\mathbf{c},s}, a, 0)$ and the initial condition $(y^0, y, y^\ell, \boldsymbol{\beta})(0) = (\tilde{y}^0, \tilde{y}, \tilde{y}^\ell, \tilde{\beta})$.*

Proof. While the first relation in (6.9) is trivial, in that $w^0_{\mathbf{c},s} \equiv 0$ and the Lagrangian $\ell^e(y, a, w^0_{\mathbf{c},s}, w_{\mathbf{c},s}) = \ell_0(y, a) w^0_{\mathbf{c},s} \equiv 0$, a proof of (6.10) can be found in [22]. Finally, the second relation in (6.9) is trivial as well for $|w_{\mathbf{c},s}| \equiv 1$ on $[0, s]$. ∎

[17] Note that the choice of the element a is irrelevant.

DEFINITION 6.8 (bracket-like approximation). *Fix $\bar{s} \in (0, \bar{S})$ and let $\mathbf{c} = B \in \mathfrak{B}^0$ be a bracket-like variation generator of length h. For each $\varepsilon > 0$ such that $\varepsilon < \bar{\varepsilon}$ and $2\varepsilon^{1/h} < \bar{s}$, where $\bar{\varepsilon}$ is as in Lemma 6.7, consider the dilation*

$$\begin{aligned} &\gamma^\varepsilon : [\bar{s} - 2\varepsilon^{1/h}, \bar{s} - \varepsilon^{1/h}] \to [\bar{s} - 2\varepsilon^{1/h}, \bar{s}], \\ &\gamma^\varepsilon(\sigma) := (\bar{s} - 2\varepsilon^{1/h}) + 2\big(\sigma - (\bar{s} - 2\varepsilon^{1/h})\big). \end{aligned} \tag{6.11}$$

For any control $(\tilde{w}^0, \tilde{w}, \tilde{\alpha}, \tilde{\zeta}) \in L^\infty\left([0, \bar{S}], W \times A \times [-\rho, \rho]\right)$, let us set

$$(\tilde{w}^0, \tilde{w}, \tilde{\alpha}, \tilde{\zeta})^\varepsilon_{\mathbf{c},\bar{s}}(s) := \begin{cases} \left(2\,\tilde{w}^0, 2\,\tilde{w}, \tilde{\alpha}, \tilde{\zeta}\right) \circ \gamma^\varepsilon(s) & \text{if } s \in [\bar{s} - 2\varepsilon^{1/h}, \bar{s} - \varepsilon^{1/h}), \\ \left(0, w_{\mathbf{c},\varepsilon^{1/h}}(s - (\bar{s} - \varepsilon^{1/h})), a\right) & \text{if } s \in [\bar{s} - \varepsilon^{1/h}, \bar{s}], \\ \big(\tilde{w}^0, \tilde{w}, \tilde{\alpha}, \tilde{\zeta}\big)(s) & \text{if } s \in [0, \bar{s} - 2\varepsilon^{1/h}) \cup (\bar{s}, S], \end{cases} \tag{6.12}$$

where $a \in A$ is arbitrary and $w_{\mathbf{c},\varepsilon^{1/h}}$ is as in Lemma 6.7. We refer to the family of controls $\left\{(\tilde{w}^0, \tilde{w}, \tilde{\alpha}, \tilde{\zeta})^\varepsilon_{\mathbf{c},\bar{s}} : \varepsilon \in (0, \bar{\varepsilon}),\ 2\varepsilon^{1/h} < \bar{s}\right\}$ as a bracket-like control approximation of $(\tilde{w}^0, \tilde{w}, \tilde{\alpha}, \tilde{\zeta})$ at $\bar{s}$ associated with $\mathbf{c} = B$.

LEMMA 6.9 (asymptotics of bracket-like variations). *Let us consider a bracket-like variation generator $\mathbf{c} = B \in \mathfrak{B}^0$ with B of length h. For every point $\bar{s} \in (0, \bar{S})$ and for each $\varepsilon > 0$ as in Definition 6.8, let $(\bar{w}^0, \bar{w}, \bar{\alpha}, 0)^\varepsilon_{\mathbf{c},\bar{s}}$ be a bracket-like control approximation of $(\bar{w}^0, \bar{w}, \bar{\alpha}, 0)$ at $\bar{s}$ associated with $\mathbf{c} = B$, and let $(y^{0\varepsilon}, y^\varepsilon, y^{\ell\varepsilon}, \beta^\varepsilon)$ be the corresponding solution of system (6.1). Then, for every $s \in (\bar{s}, \bar{S}]$ one has*

$$\begin{pmatrix} y^{0\varepsilon}(s) - \bar{y}^0(s) \\ y^\varepsilon(s) - \bar{y}(s) \\ y^{\ell\varepsilon}(s) - \bar{y}^\ell(s) \end{pmatrix} = \varepsilon \begin{pmatrix} \mathbf{v}^0_{\mathbf{c},\bar{s}} \\ M(s, \bar{s}) \cdot \mathbf{v}_{\mathbf{c},\bar{s}} \\ \mu(s, \bar{s}) \cdot \mathbf{v}_{\mathbf{c},\bar{s}} + \mathbf{v}^\ell_{\mathbf{c},\bar{s}} \end{pmatrix} + \begin{pmatrix} 0 \\ o(\varepsilon) \\ o(\varepsilon) \end{pmatrix} = \varepsilon \begin{pmatrix} 0 \\ M(s, \bar{s}) \cdot \dfrac{B(\bar{y}(\bar{s}))}{(r_B)^h} \\ \mu(s, \bar{s}) \cdot \dfrac{B(\bar{y}(\bar{s}))}{(r_B)^h} \end{pmatrix} + \begin{pmatrix} 0 \\ o(\varepsilon) \\ o(\varepsilon) \end{pmatrix},$$

and $\beta^\varepsilon(s) - \bar{\beta}(s) = \varepsilon^{\frac{1}{h}}$.

Proof. By the rate independence of the control system (6.1), $(y^{0\varepsilon}, y^\varepsilon, y^{\ell\varepsilon}) = (\bar{y}^0, \bar{y}, \bar{y}^\ell) \circ \gamma^\varepsilon$ on $[\bar{s} - 2\varepsilon^{1/h}, \bar{s} - \varepsilon^{1/h}]$, so that $(y^{0\varepsilon}, y^\varepsilon, y^{\ell\varepsilon})(\bar{s} - \varepsilon^{1/h}) = (\bar{y}^0, \bar{y}, \bar{y}^\ell)(\bar{s})$. Hence $y^{0\varepsilon}(\bar{s}) - \bar{y}^0(\bar{s}) = 0$, $y^{\ell\varepsilon}(\bar{s}) - \bar{y}^\ell(\bar{s}) = 0$, while

$$\begin{aligned} y^\varepsilon(\bar{s}) - \bar{y}(\bar{s}) &= \int_{\bar{s} - \varepsilon^{1/h}}^{\bar{s}} \sum_{i=1}^m g_i(y^\varepsilon(s)) w^i_{\mathbf{c},\varepsilon^{1/h}}(s - (\bar{s} - \varepsilon^{1/h}))\, \mathrm{d}s \\ &= \int_0^{\varepsilon^{1/h}} \sum_{i=1}^m g_i(y^\varepsilon(s + (\bar{s} - \varepsilon^{1/h}))\, w^i_{\mathbf{c},\varepsilon^{1/h}}(s)\, \mathrm{d}s, \end{aligned} \tag{6.13}$$

where $w_{\mathbf{c},\varepsilon^{1/h}}$ is the control associated with the bracket B as in Lemma 6.7. It follows that $y^\varepsilon(\bar{s}) = y^\varepsilon(s + (\bar{s} - \varepsilon^{1/h}))|_{s=\varepsilon^{1/h}} = \mathcal{y}^\varepsilon(\varepsilon^{1/h})$, where we have used $\mathcal{y}^\varepsilon$ to denote the solution to the Cauchy problem $\frac{dy}{d\sigma}(\sigma) = \sum_{i=1}^m g_i(y(\sigma)) w^i_{\mathbf{c},\varepsilon^{1/h}}(\sigma)$, $y(0) = \bar{y}(\bar{s})$, so that, by Lemma 6.7, we get

$$y^\varepsilon(\bar{s}) - \bar{y}(\bar{s}) - \left(\frac{\varepsilon^{1/h}}{r_B}\right)^h B(\bar{y}(\bar{s})) = \mathcal{y}^\varepsilon(\varepsilon^{1/h}) - \bar{y}(\bar{s}) - \varepsilon\, \frac{B(\bar{y}(\bar{s}))}{(r_B)^h} = o(\varepsilon).$$

Therefore, $y^\varepsilon(\bar{s}) - \bar{y}(\bar{s}) = \varepsilon\, \frac{B(\bar{y}(\bar{s}))}{(r_B)^h} + o(\varepsilon)$, and the proof of the first relation of the thesis is concluded, since for every $s \in (\bar{s}, \bar{S}]$, the fundamental matrix $\tilde{M}(s, \bar{s})$ is the differential of the flow map from $\bar{s}$ to s. Finally, by the second relation in (6.9), one has $\beta^\varepsilon(s) - \bar{\beta}(s) = \beta^\varepsilon(\bar{s}) - \bar{\beta}(\bar{s}) = \varepsilon^{\frac{1}{h}}$. ∎

6.3. Composition of variations. Let $\mathbf{c} \in \mathfrak{V}$ be a variation generator of length $h \geq 1$, and let $\bar{s} \in (0, \bar{S})$. For any $\varepsilon > 0$ small enough,[18] let us introduce the operator $\mathcal{A}^{\varepsilon}_{\mathbf{c},\bar{s}} : L^{\infty}\left([0, \bar{S}], \mathbb{R}_+ \times \mathcal{C} \times A \times [-\rho, \rho]\right) \to L^{\infty}\left([0, \bar{S}], \mathbb{R}_+ \times \mathcal{C} \times A \times [-\rho, \rho]\right)$ given by

$$\mathcal{A}^{\varepsilon}_{\mathbf{c},\bar{s}}(w^0, w, \alpha, \zeta) := (w^0, w, \alpha, \zeta)^{\varepsilon}_{\mathbf{c},\bar{s}}. \tag{6.14}$$

Lemma 6.10 (multiple variations at different times). *Let $N > 0$ be an integer and let $\vec{\mathbf{c}} := (\mathbf{c}_1, \ldots, \mathbf{c}_N) \in \mathfrak{V}^N$ be an N-uple of variation generators of lengths $\vec{\mathbf{h}} := (h_1, \ldots, h_N) \in \mathbb{N}^N$. Fix $\vec{\mathbf{s}} := (\bar{s}_1, \ldots, \bar{s}_N) \in (0, \bar{S})^N$, where $0 =: \bar{s}_0 < \bar{s}_1 < \cdots < \bar{s}_N < \bar{S}$ and $\bar{s}_j \in (0, \bar{S})_{\mathrm{Leb}}$ as soon as $h_j = 1$. For each $\vec{\epsilon} := (\varepsilon_1, \ldots, \varepsilon_N) \in (0, +\infty)^N$ small enough, let us set*

$$(w^{0\vec{\epsilon}}, w^{\vec{\epsilon}}, \alpha^{\vec{\epsilon}}, \zeta^{\vec{\epsilon}}) := \mathcal{A}^{\varepsilon_N}_{\mathbf{c}_N,\bar{s}_N} \circ \cdots \circ \mathcal{A}^{\varepsilon_j}_{\mathbf{c}_j,\bar{s}_j} \circ \cdots \circ \mathcal{A}^{\varepsilon_1}_{\mathbf{c}_1,\bar{s}_1}(\bar{w}^0, \bar{w}, \bar{\alpha}, 0), \tag{6.15}$$

and let $\left(\bar{S}, w^{0\vec{\epsilon}}, w^{\vec{\epsilon}}, \alpha^{\vec{\epsilon}}, \zeta^{\vec{\epsilon}}, y^{0\vec{\epsilon}}, y^{\vec{\epsilon}}, y^{\ell\vec{\epsilon}}, \beta^{\vec{\epsilon}}\right)$ denote the corresponding process of (6.1).

Then, for every $s \in (\bar{s}_N, \bar{S}]$, one has

$$\begin{pmatrix} y^{0\vec{\epsilon}}(s) - \bar{y}^0(s) \\ y^{\vec{\epsilon}}(s) - \bar{y}(s) \\ y^{\ell\vec{\epsilon}}(s) - \bar{y}^{\ell}(s) \end{pmatrix} = \sum_{j=1}^{N} \varepsilon_j \begin{pmatrix} \mathbf{v}^0_{\mathbf{c}_j,\bar{s}_j} \\ M(s, \bar{s}_j)\mathbf{v}_{\mathbf{c}_j,\bar{s}_j} \\ \mu(s, \bar{s}_j) \cdot \mathbf{v}_{\mathbf{c}_j,\bar{s}_j} + \mathbf{v}^{\ell}_{\mathbf{c}_j,\bar{s}_j} \end{pmatrix} + o(|\vec{\epsilon}|), \tag{6.16}$$

and

$$\beta^{\vec{\epsilon}}(s) - \bar{\beta}(s) = \sum_{j \in I_1} \varepsilon_j \left(|w_j|(1 + \zeta_j) - |\bar{w}(\bar{s}_j)|\right) + o(|\vec{\epsilon}|) + \sum_{j \in \{1,\ldots,N\} \setminus I_1} (\varepsilon_j)^{\frac{1}{h_j}}, \tag{6.17}$$

where $I_1 := \{j = 1, \ldots, N : \ h_j = 1\}$. In particular, if all $\mathbf{c}_j$ are needle variations, i.e., $\mathbf{c}_j := (w^0_j, w_j, a_j, \zeta_j)$ for every $j = 1, \ldots, N$, one gets

$$\beta^{\vec{\epsilon}}(s) - \bar{\beta}(s) = \sum_{j=1}^{N} \varepsilon_j \left(|w_j|(1 + \zeta_j) - |\bar{w}(\bar{s}_j)|\right) + o(|\vec{\epsilon}|). \tag{6.18}$$

Proof. Let us prove the result by induction on N, the number of composed variations. For $N = 1$, the result is proved in Lemmas 6.6 and 6.9. If $N \geq 2$, let us assume that the result holds true for $N - 1$ and let us show that it is valid for N as well. Let us use $(y^0, y, y^{\ell}, \beta)^N$ and $(y^0, y, y^{\ell}, \beta)^{N-1}$ to denote the trajectories associated with the N variations and the first $N - 1$ variations, respectively (we omit the dependence on $\vec{\epsilon}$ for brevity). Then one has

$$\begin{pmatrix} y^{0,N}(\bar{s}_N) - \bar{y}^0(\bar{s}_N) \\ y^N(\bar{s}_N) - \bar{y}(\bar{s}_N) \\ y^{\ell,N}(\bar{s}_N) - \bar{y}^{\ell}(\bar{s}_N) \\ \beta^N(\bar{s}_N) - \bar{\beta}(\bar{s}_N) \end{pmatrix} = \begin{pmatrix} y^{0,N}(\bar{s}_N) - y^{0,(N-1)}(\bar{s}_N) \\ y^N(\bar{s}_N) - y^{N-1}(\bar{s}_N) \\ y^{\ell,N}(\bar{s}_N) - y^{\ell,(N-1)}(\bar{s}_N) \\ \beta^N(\bar{s}_N) - \beta^{N-1}(\bar{s}_N) \end{pmatrix} + \begin{pmatrix} y^{0,(N-1)}(\bar{s}_N) - \bar{y}^0(\bar{s}_N) \\ y^{N-1}(\bar{s}_N) - \bar{y}(\bar{s}_N) \\ y^{\ell,(N-1)}(\bar{s}_N) - \bar{y}^{\ell}(\bar{s}_N) \\ \beta^{N-1}(\bar{s}_N) - \bar{\beta}(\bar{s}_N) \end{pmatrix}. \tag{6.19}$$

By the inductive hypothesis, we get that

$$\begin{pmatrix} y^{0,(N-1)}(\bar{s}_N) - \bar{y}^0(\bar{s}_N) \\ y^{N-1}(\bar{s}_N) - \bar{y}(\bar{s}_N) \\ y^{\ell,(N-1)}(\bar{s}_N) - \bar{y}^{\ell}(\bar{s}_N) \end{pmatrix} = \sum_{j=1}^{N-1} \varepsilon_j \begin{pmatrix} \mathbf{v}^0_{\mathbf{c}_j,\bar{s}_j} \\ M(\bar{s}_N, \bar{s}_j)\mathbf{v}_{\mathbf{c}_j,\bar{s}_j} \\ \mu(\bar{s}_N, \bar{s}_j)\mathbf{v}_{\mathbf{c}_j,\bar{s}_j} + \mathbf{v}^{\ell}_{\mathbf{c}_j,\bar{s}_j} \end{pmatrix} + o(|(\varepsilon_1, \ldots, \varepsilon_{N-1})|) \tag{6.20}$$

[18] Precisely, if $h \geq 2$, we require $0 < \varepsilon < \bar{\varepsilon}$, $2\varepsilon^{1/h} < \bar{s}$, as in Definition 6.8, while, in the case $h = 1$, $\varepsilon < \bar{s}$.

and, setting $I_1^{N-1} := \{j = 1, \dots, N-1 : \ h_j = 1\}$,

$$\beta^{N-1}(\bar{s}_N) - \bar{\beta}(\bar{s}_N) = \sum_{j \in I_1^{N-1}} \varepsilon_j \left(|w_j|(1+\zeta_j) - |\bar{w}(\bar{s}_j)|\right) + o(|(\varepsilon_1, \dots, \varepsilon_{N-1})|) + \sum_{j \in \{1,\dots,N-1\} \setminus I_1^{N-1}} (\varepsilon_j)^{\frac{1}{h_j}}. \tag{6.21}$$

We claim that

$$\begin{pmatrix} y^{0,N}(\bar{s}_N) - y^{0,(N-1)}(\bar{s}_N) \\ y^N(\bar{s}_N) - y^{N-1}(\bar{s}_N) \\ y^{\ell,N}(\bar{s}_N) - y^{\ell,(N-1)}(\bar{s}_N) \end{pmatrix} = \varepsilon_N \begin{pmatrix} \mathbf{V}^0_{\mathbf{c}_N, \bar{s}_N} \\ \mathbf{V}_{\mathbf{c}_N, \bar{s}_N} \\ \mathbf{V}^\ell_{\mathbf{c}_N, \bar{s}_N} \end{pmatrix} + o(|\vec{\epsilon}|) \tag{6.22}$$

and

(6.23)

$$\beta^N(\bar{s}_N) - \beta^{N-1}(\bar{s}_N) = \begin{cases} \varepsilon_N \big(|w_N|(1+\zeta_N) - |\bar{w}(\bar{s}_N)|\big) + o(|\varepsilon_N|) & \text{if } h_N = 1, \\ (\varepsilon_N)^{\frac{1}{h_N}} & \text{if } h_N \ge 2. \end{cases}$$

Once one has proven the claim, the validity of (6.16) and (6.17) follows easily by (6.19)–(6.21), by the properties of the fundamental matrix $\tilde{M}(s, \bar{s}_N)$. To prove (6.22), (6.23) we first consider the case when the length h_N of the Nth variation $\mathbf{c}_N$ is ≥ 2.

Case $h_N \ge 2$. Here $\mathbf{c}_N = B_N \in \mathfrak{B}^0$ is a bracket-like variation and one has

$$\begin{pmatrix} y^{0,N} \\ y^N \\ y^{\ell,N} \\ \beta^N \end{pmatrix} (\bar{s}_N - \varepsilon_N^{1/h_N}) = \begin{pmatrix} y^{0,N-1} \\ y^{N-1} \\ y^{\ell,N-1} \\ \beta^{N-1} \end{pmatrix} (\bar{s}_N),$$

so that $y^{0,N}(\bar{s}_N) - y^{0,N-1}(\bar{s}_N) = 0$, $y^{\ell,N}(\bar{s}_N) - y^{\ell,N-1}(\bar{s}_N) = 0$, $\beta^N(\bar{s}_N) - \beta^{N-1}(\bar{s}_N) = \varepsilon_N^{1/h_N}$, while

$$\begin{aligned} y^N(\bar{s}_N) - y^{N-1}(\bar{s}_N) &= \int_{\bar{s}_N - \varepsilon_N^{1/h_N}}^{\bar{s}_N} \sum g_i(y^N(s)) w^i_{\mathbf{c}_N, \varepsilon_N^{1/h_N}} \Big(s - \big(\bar{s}_N - \varepsilon_N^{1/h_N}\big)\Big) \mathrm{d}s \\ &= \int_0^{\varepsilon^{1/h}} \sum_{i=1}^m g_i(y^N(s + (\bar{s}_N - \varepsilon_N^{1/h_N}))) w^i_{\mathbf{c}_N, \varepsilon_N^{1/h_N}}(s)\, \mathrm{d}s, \end{aligned}$$

where the control $w_{\mathbf{c}_N, \varepsilon_N^{1/h_N}}$ is as in Lemma 6.7. If y^N denotes the solution to the Cauchy problem $\frac{dy}{d\sigma}(\sigma) = \sum_{i=1}^m g_i(y(\sigma))\, w^i_{\mathbf{c}_N, \varepsilon_N^{1/h_N}}(\sigma)$, $y(0) = y^{N-1}(\bar{s}_N)$, then $y^N(\bar{s}_N) = y^N\big(s + (\bar{s}_N - \varepsilon_N^{1/h_N})\big)\big|_{s = \varepsilon_N^{1/h_N}} = y^N(\varepsilon_N^{1/h_N})$ and, by Lemma 6.7, we get

$$\begin{aligned} y^N(\bar{s}_N) - y^{N-1}(\bar{s}_N) - &\left(\frac{\varepsilon_N^{1/h_N}}{r_{B_N}}\right)^{h_N} B_N(\bar{y}(\bar{s}_N)) \\ &= y^N(\varepsilon_N^{1/h_N}) - y^{N-1}(\bar{s}_N) - \frac{\varepsilon_N}{(r_{B_N})^{h_N}} B_N(y^{N-1}(\bar{s}_N)) \\ &\quad + \frac{\varepsilon_N}{(r_{B_N})^{h_N}} B_N(y^{N-1}(\bar{s}_N)) - \frac{\varepsilon_N}{(r_{B_N})^{h_N}} B_N(\bar{y}(\bar{s}_N)) \\ &= o(\varepsilon_N) + \frac{\varepsilon_N}{(r_{B_N})^{h_N}} B_N(y^{N-1}(\bar{s}_N)) - \frac{\varepsilon_N}{(r_{B_N})^{h_N}} B_N(\bar{y}(\bar{s}_N)). \end{aligned}$$

Now by the continuity of B_N and the inductive hypothesis (6.20), it follows that

$$\begin{aligned}&\left|\frac{\varepsilon_N}{(r_{B_N})^{h_N}} B_N(y^{N-1}(\bar{s}_N)) - \frac{\varepsilon_N}{(r_{B_N})^{h_N}} B_N(\bar{y}(\bar{s}_N))\right| \\ &\quad \le \frac{\varepsilon_N}{(r_{B_N})^{h_N}}\, \omega_{B_N}\left(|y^{N-1}(\bar{s}_N) - \bar{y}(\bar{s}_N)|\right) \le \frac{\varepsilon_N}{(r_{B_N})^{h_N}}\, \omega_{B_N}\left(\mathcal{O}\left(\varepsilon_1 + \cdots + \varepsilon_{N-1}\right)\right),\end{aligned}$$

where ω_{B_N} denotes the modulus of continuity of B_N and we use $\mathcal{O}$ to mean a nonnegative function such that $\mathcal{O}(r) \le Cr$ for all $r \ge 0$ for some constant $C > 0$. Therefore, $y^N(\bar{s}_N) - y^{N-1}(\bar{s}_N) = \frac{\varepsilon_N}{(r_{B_N})^{h_N}} B_N(\bar{y}(\bar{s}_N)) + o(|\vec{\epsilon}|)$, which concludes the proof in this case.

Case $h_N = 1$. Here $\mathbf{c}_N = (w_N^0, w_N, a_N, \zeta_N)$ and the aimed at estimate is rather standard. Nonetheless, we perform it for the sake of self-consistency. One has

$$\begin{aligned}&y^N(\bar{s}_N) - y^{N-1}(\bar{s}_N) \\ &\quad = \int_{\bar{s}_N-\varepsilon_N}^{\bar{s}_N} \left[F^e(y^N(s), w_N^0, w_N, a_N)(1+\zeta_N) - F^e(y^{N-1}(s), \bar{w}^0(s), \bar{w}(s), \bar{\alpha}(s))\right] \mathrm{d}s \\ &\quad = \int_{\bar{s}_N-\varepsilon_N}^{\bar{s}_N} \big(r_1(s) + r_2 + r_3(s)\big)\mathrm{d}s,\end{aligned}$$

where

$$\begin{aligned}&r_1(s) := F^e(y^N(s), w_N^0, w_N, a_N)(1+\zeta_N) - F^e(\bar{y}(\bar{s}_N), w_N^0, w_N, a_N)(1+\zeta_N), \\ &r_2 := F^e(\bar{y}(\bar{s}_N), w_N^0, w_N, a_N)(1+\zeta_N) - \bar{F}^e(\bar{s}_N), \\ &r_3(s) := \bar{F}^e(\bar{s}_N) - F^e(y^{N-1}(s), \bar{w}^0(\bar{s}_N), \bar{w}(\bar{s}_N), \bar{\alpha}(\bar{s}_N)).\end{aligned}$$

Let us start by estimating r_1. Observe that, for $s \in [\bar{s}_N - \varepsilon_N, \bar{s}_N]$,

$$|y^N(s) - \bar{y}(\bar{s}_N)| \le |y^N(s) - y^{N-1}(s)| + |y^{N-1}(s) - \bar{y}(s)| + |\bar{y}(s) - \bar{y}(\bar{s}_N)|.$$

Moreover, on $[\bar{s}_N - \varepsilon_N, \bar{s}_N]$, one has $\|y^N - y^{N-1}\|_\infty = \mathcal{O}\left(\varepsilon_N\right)$ by the Lipschitz continuity of the input-output map Φ defined in (6.2); $\|y^{N-1} - \bar{y}\|_\infty = \mathcal{O}\left(\varepsilon_1 + \cdots + \varepsilon_{N-1}\right)$ by the inductive hypothesis (6.20); and $\|\bar{y}(s) - \bar{y}(\bar{s}_N)\|_\infty = \mathcal{O}\left(\varepsilon_N\right)$ by the Lipschitz continuity of the reference trajectory. Hence $\|y^N(s) - \bar{y}(\bar{s}_N)\|_\infty = \mathcal{O}(|\vec{\epsilon}|)$, so that

$$\left|\int_{\bar{s}_N-\varepsilon_N}^{\bar{s}_N} r_1(s)\mathrm{d}s\right| \le \int_{\bar{s}_N-\varepsilon_N}^{\bar{s}_N} L\,|y^N(s) - \bar{y}(\bar{s}_N)|\mathrm{d}s = \varepsilon_N\, \mathcal{O}(|\vec{\epsilon}|),$$

where L is a suitable positive constant. By the previous estimates and recalling that $\bar{s}_N$ is a Lebesgue point of the map in Definition 6.3, we get

$$\begin{aligned}\Big|\int_{\bar{s}_N-\varepsilon_N}^{\bar{s}_N} r_3(s)\mathrm{d}s\Big| \le &\Big|\int_{\bar{s}_N-\varepsilon_N}^{\bar{s}_N} \left[\bar{F}^e(\bar{s}_N) - \bar{F}^e(s)\right] ds\Big| \\ &+ \Big|\int_{\bar{s}_N-\varepsilon_N}^{\bar{s}_N} \left[\bar{F}^e(s) - F^e(y^{N-1}(s), \bar{w}^0(s), \bar{w}(s), \bar{\alpha}(s))\right] ds\Big| \\ &\le o(\varepsilon_N) + \varepsilon_N\, \mathcal{O}(|(\varepsilon_1, \ldots, \varepsilon_{N-1})|).\end{aligned}$$

Therefore, $y^N(\bar{s}_N) - y^{N-1}(\bar{s}_N) = \varepsilon_N\, r_2 + o(|\vec{\epsilon}|)$, and the relation in (6.22) concerning the state variables is proven. The proofs of the other relations are similar and actually easier, so we omit them. ∎

6.4. Set separation. Given a process $(\bar{S}, w^0, w, \alpha, \zeta, y^0, y, y^\ell, \beta)$ of the rescaled problem (P_e), let us introduce the total cost component

$$y^c(s) := \Psi(y^0(s), y(s)) + y^\ell(s), \quad s \in [0, \bar{S}].^{19} \tag{6.24}$$

Setting $\bar{y}^c(s) := \Psi(\bar{y}^0(s), \bar{y}(s)) + \bar{y}^\ell(s)$, $s \in [0, \bar{S}]$, for any $\delta > 0$ we define the *δ-reachable set* $\mathcal{R}_\delta$ and its *projection* $\mathcal{R}'_\delta$ as

$$\mathcal{R}_\delta := \left\{ \begin{array}{c} (y^0, y, y^c, \beta)(\bar{S}) : \ (S, w^0, w, \alpha, \zeta, y^0, y, y^\ell, \beta) \text{ verifies} \\ \mathrm{d}\left((\bar{S}, y^0, y, y^c, \beta), (\bar{S}, \bar{y}^0, \bar{y}, \bar{y}^c, \bar{\beta})\right) < \delta \end{array} \right\} \subseteq \mathbb{R}^{1+n+1+1},$$

$$\mathcal{R}'_\delta := \left\{ (y^0, y, y^c)(\bar{S}) : \ (y^0, y, y^c, \beta)(\bar{S}) \in \mathcal{R}_\delta \right\} \subseteq \mathbb{R}^{1+n+1}.$$

Let $N > 0$ be an integer, let $\vec{\mathbf{c}} := (\mathbf{c}_1, \dots, \mathbf{c}_N) \in \mathfrak{V}^N$ be an N-uple of variation generators, and fix $\vec{\mathbf{s}} := (\bar{s}_1, \dots, \bar{s}_N) \in (0, \bar{S})^N$ as in Definition 6.10. We define the set

$$E' := \left\{ \begin{pmatrix} \mathbf{v}^0_{\mathbf{c}_j, \bar{s}_j} \\ M(\bar{S}, \bar{s}_j) \cdot \mathbf{v}_{\mathbf{c}_j, \bar{s}_j} \\ \frac{\partial \bar{\Psi}}{\partial t}(\bar{S}) \mathbf{v}^0_{\mathbf{c}_j, \bar{s}_j} + \frac{\partial \bar{\Psi}}{\partial x}(\bar{S}) \cdot M(\bar{S}, \bar{s}_j) \cdot \mathbf{v}_{\mathbf{c}_j, \bar{s}_j} + \mu(\bar{S}, \bar{s}_j) \cdot \mathbf{v}_{\mathbf{c}_j, \bar{s}_j} + \mathbf{v}^\ell_{\mathbf{c}_j, \bar{s}_j} \end{pmatrix}, \ j = 1, \dots N \right\},$$

where the variations are as in Definition 6.4 and $\frac{\partial \bar{\Psi}}{\partial t}(\bar{S}) := \frac{\partial \Psi}{\partial t}((\bar{y}^0, \bar{y})(\bar{S}))$, $\frac{\partial \bar{\Psi}}{\partial x}(\bar{S}) := \frac{\partial \Psi}{\partial x}((\bar{y}^0, \bar{y})(\bar{S}))$. Observe then when *all* $\mathbf{c}_j = (w^0_j, w_j, a_j, \zeta_j)$, $j = 1, \dots N$, *are needle variations,* we can also define the set

$$E := \left\{ \begin{pmatrix} \mathbf{v}^0_{\mathbf{c}_j, \bar{s}_j} \\ M(\bar{S}, \bar{s}_j) \cdot \mathbf{v}_{\mathbf{c}_j, \bar{s}_j} \\ \frac{\partial \bar{\Psi}}{\partial t}(\bar{S}) \mathbf{v}^0_{\mathbf{c}_j, \bar{s}_j} + \frac{\partial \bar{\Psi}}{\partial x}(\bar{S}) \cdot M(\bar{S}, \bar{s}_j) \cdot \mathbf{v}_{\mathbf{c}_j, \bar{s}_j} + \mu(\bar{S}, \bar{s}_j) \cdot \mathbf{v}_{\mathbf{c}_j, \bar{s}_j} + \mathbf{v}^\ell_{\mathbf{c}_j, \bar{s}_j} \\ |w|(1 + \zeta_j) - |\bar{w}(\bar{s}_j)| \end{pmatrix}, \ j = 1, \dots N \right\}.$$

In this case, E' turns out to be the projection of E on $\mathbb{R}^{1+n+1}$.

Finally, let us define the convex cones

$$R := \mathrm{span}^+(E) \subset \mathbb{R}^{1+n+1+1}, \quad R' := \mathrm{span}^+(E') \subset \mathbb{R}^{1+n+1}, \tag{6.25}$$

where, for a given subset Θ of a vector space, $\mathrm{span}^+(\Theta)$ denotes its *positive span.*

LEMMA 6.11. (i) *The set R' is a Boltyanski approximating cone of the set $\mathcal{R}'_\delta$ at the point $(\bar{y}^0, \bar{y}, \bar{y}^c)(\bar{S})$.*

(ii) *When all $\mathbf{c}_j = (w^0_j, w_j, a_j, \zeta_j)$ for $j = 1, \dots, N$ are needle variations, the set R is a Boltyanski approximating cone of the set $\mathcal{R}$ at $(\bar{y}^0, \bar{y}, \bar{y}^c, \bar{\beta})(\bar{S})$.*

Proof. Let us set $y^{c\vec{\epsilon}}(s) := \Psi((y^{0\vec{\epsilon}}, y^{\vec{\epsilon}})(s)) + y^{\ell\vec{\epsilon}}(s)$, where $y^{\ell\vec{\epsilon}}$, $y^{0\vec{\epsilon}}$, and $y^{\vec{\epsilon}}$ are as in Lemma 6.10. By (6.16) we get

$$y^{c\vec{\epsilon}}(\bar{S}) - \bar{y}^c(\bar{S}) = \sum_{j=1}^N \varepsilon_j \Big(\frac{\partial \bar{\Psi}}{\partial t}(\bar{S}) \mathbf{v}^0_{\mathbf{c}_j, \bar{s}_j} + \frac{\partial \bar{\Psi}}{\partial x}(\bar{S}) \cdot M(\bar{S}, \bar{s}_j) \cdot \mathbf{v}_{\mathbf{c}_j, \bar{s}_j} + \mu(\bar{S}, \bar{s}_j) \cdot \mathbf{v}_{\mathbf{c}_j, \bar{s}_j} + \mathbf{v}^\ell_{\mathbf{c}_j, \bar{s}_j} \Big) + o(|\vec{\epsilon}|).$$

Therefore, part (ii) of the statement follows from Lemma 6.10.

[19] y^c solves $\frac{dy^c}{ds} = (\frac{\partial \Psi}{\partial t} w^0 + \frac{\partial \Psi}{\partial x}(f(y, \alpha) w^0 + \sum_{i=1}^m g_i(y) w^i) + \ell^e(y, w^0, w, \alpha))(1 + \zeta)$, with initial condition $y^c(0) = \Psi(0, \check{x}, 0)$.

To prove part (i), for some $\tilde{\varepsilon} > 0$ sufficiently small, let us define the function $F : (0, +\infty)^N \cap \tilde{\varepsilon}\,\mathbb{B}_N \to \mathbb{R}^{1+n+2}$ by setting $F(\vec{\epsilon}) := \left(y^{0\vec{\epsilon}}(\bar{S}), y^{\vec{\epsilon}}(\bar{S}), y^{c\vec{\epsilon}}(\bar{S})\right)$. It is straightforward to prove that $F(\vec{\epsilon}) = \left(y^0(\bar{S}), y(\bar{S}), y^c(\bar{S})\right) + L \cdot \vec{\epsilon} + o(|\vec{\epsilon}|)$, where the linear operator $L \in \mathrm{Hom}(\mathbb{R}^N, \mathbb{R}^{1+n+1})$ is defined by

$$L \cdot \vec{\epsilon} := \sum_{j=1}^{N} \varepsilon_j \begin{pmatrix} \mathbf{v}^0_{\mathbf{c}_j,\bar{s}_j} \\ M(\bar{S}, \bar{s}_j)\mathbf{v}_{\mathbf{c}_j,\bar{s}_j} \\ \frac{\partial \bar{\Psi}}{\partial t}(\bar{S})\mathbf{v}^0_{\mathbf{c}_j,\bar{s}_j} + \frac{\partial \bar{\Psi}}{\partial x}(\bar{S}) \cdot M(\bar{S}, \bar{s}_j) \cdot \mathbf{v}_{\mathbf{c}_j,\bar{s}_j} + \mu(\bar{S}, \bar{s}_j) \cdot \mathbf{v}_{\mathbf{c}_j,\bar{s}_j} + \mathbf{v}^\ell_{\mathbf{c}_j,\bar{s}_j} \end{pmatrix}.$$

Hence (i) is proved, in that $R' = L \cdot (0, +\infty)^N$. ∎

Let us consider the *profitable set* $\mathcal{P}$ and its projection $\mathcal{P}'$, defined as

$$\mathcal{P} := \mathfrak{T} \times \left(-\infty, \bar{y}^c(\bar{S})\right) \times [0, K] \bigcup \left\{(\bar{y}^0, \bar{y}, \bar{y}^c, \bar{\beta})(\bar{S})\right\},$$
$$\mathcal{P}' := \mathfrak{T} \times \left(-\infty, \bar{y}^c(\bar{S})\right) \bigcup \left\{(\bar{y}^0, \bar{y}, \bar{y}^c)(\bar{S})\right\},$$

and let Γ be a Boltyanski approximating cone for the target $\mathfrak{T}$ at $(\bar{y}^0, \bar{y})(\bar{S})$. Recalling that $\bar{\beta}(\bar{S}) < K$, one trivially checks that the sets

$$P := \Gamma \times \mathbb{R}_- \times \{0\}, \qquad P' := \Gamma \times \mathbb{R}_-,$$

are Boltyanski approximating cones of $\mathcal{P}$ at $(\bar{y}^0, \bar{y}, \bar{y}^c, \bar{\beta})(\bar{S})$ and of $\mathcal{P}'$ at $(\bar{y}^0, \bar{y}, \bar{y}^c)(\bar{S})$, respectively. We will need the following elementary result.

Lemma 6.12. *There exists $\delta > 0$ such that the sets $\mathcal{P}'$ and ${\mathcal{R}'}_\delta$ are locally separated at $(\bar{y}^0, \bar{y}, \bar{y}^c)(\bar{S})$.*

Proof. Suppose by contradiction that for every $\delta > 0$ the sets $\mathcal{P}'$ and ${\mathcal{R}'}_\delta$ are not locally separated at $(\bar{y}^0, \bar{y}, \bar{y}^c)(\bar{S})$. Then, given $\delta \in (0, K - \bar{\beta}(\bar{S}))$,[20] there exists a process $(\bar{S}, w^0, w, \alpha, \zeta, y^0, y, y^\ell, \beta)$ of (6.1) verifying

$$(y^0, y, y^c)(\bar{S}) \in {\mathcal{R}'}_\delta \cap \mathcal{P}', \quad \mathrm{d}((y^0, y, y^c, \beta), (\bar{y}^0, \bar{y}, \bar{y}^c, \bar{\beta})) < \delta.$$

This implies that $\beta(\bar{S}) \le \delta + \bar{\beta}(\bar{S}) < K$, thus the final point $(y^0, y, y^c, \beta)(\bar{S}) \in \mathcal{R}_\delta \cap \mathcal{P}$. Hence, for every $\delta \in (0, K - \bar{\beta}(\bar{S}))$ the sets $\mathcal{P}$ and $\mathcal{R}_\delta$ are not locally separated, which contradicts the local optimality of the reference process. ∎

By Lemma 6.12 the projected reachable set $\mathcal{R}'_\delta$ is locally separated from the projected profitable set $\mathcal{P}'$ at $(\bar{y}^0, \bar{y}, \bar{y}^c)(\bar{S})$, for some $\delta > 0$. Therefore, since R' and P' are approximating cones to $\mathcal{R}'_\delta$ and $\mathcal{P}'$, respectively, and P' is not a subspace, in view of Theorem 1.3 there exists a vector $(\xi_0, \xi, \xi_c) \in \mathbb{R}^{1+n+1}$ verifying

$$0 \ne (\xi_0, \xi, \xi_c) \in {R'}^\perp \cap (-{P'}^\perp).$$

Since ${P'}^\perp = \Gamma^\perp \times \mathbb{R}_+$, one gets $(\xi_0, \xi) \in -\Gamma^\perp$, $\xi_c = -\lambda \le 0$, and

$$\xi_0 \mathbf{v}^0 + \xi \cdot \mathbf{v} + \xi_c \mathbf{v}^c \le 0 \qquad \forall (\mathbf{v}^0, \mathbf{v}, \mathbf{v}^c) \in R'.$$

By the definition of R' given in (6.25), the latter relation is verified if and only if

$$\Big(\xi_0 - \lambda \frac{\partial \bar{\Psi}}{\partial t}(\bar{S})\Big)\mathbf{v}^0_{\mathbf{c}_j,\bar{s}_j} + \Big(\xi - \lambda \frac{\partial \bar{\Psi}}{\partial x}(\bar{S})\Big) \cdot M(\bar{S}, \bar{s}_j) \cdot \mathbf{v}_{\mathbf{c}_j,\bar{s}_j}$$
$$- \lambda \left(\mu(\bar{S}, \bar{s}_j) \cdot \mathbf{v}_{\mathbf{c}_j,\bar{s}_j} + \mathbf{v}^\ell_{\mathbf{c}_j,\bar{s}_j}\right) \le 0 \quad \text{for all } j = 1, \dots, N.$$

[20]This interval is not empty for $\bar{\beta}(\bar{S}) < K$.

Therefore, setting

$$(p_0,p)(s) := \left(\xi_0 - \lambda\frac{\partial\bar\Psi}{\partial t}(\bar S)\ ,\ \left(\xi - \lambda\frac{\partial\bar\Psi}{\partial x}(\bar S)\right)\cdot M(\bar S,s) - \lambda\mu(\bar S,s)\right),$$

we obtain that the multiplier $(p_0,p,\lambda)\in\mathbb{R}\times AC\left([0,\bar S],\mathbb{R}^n\right)\times\mathbb{R}_+$ verifies

$$p_0\mathbf{v}^0_{\mathbf{c}_j,\bar s_j} + p(\bar s_j)\cdot\mathbf{v}_{\mathbf{c}_j,\bar s_j} - \lambda\mathbf{v}^\ell_{\mathbf{c}_j,\bar s_j} \le 0 \quad \text{for every } j=1,\dots,N, \tag{6.26}$$

the nontriviality condition (3.2), and (by $M(\bar S,\bar S)=\mathrm{Id}$, $\mu(\bar S,\bar S)=0$) the non-transversality condition (3.4). Moreover, by the definitions of $M(\bar S,\cdot)$ and $\mu(\bar S,\cdot)$, the path p solves the adjoint equation (3.6). Finally, for a needle variation generator $\mathbf{c}_j=(w^0_j,w_j,a_j,\zeta_j)$, by (6.26) we get

$$H\Big(\bar y(\bar s_j),p_0,p(\bar s_j),0,\lambda,w^0_j(1+\zeta_j),w_j(1+\zeta_j),a_j\Big) \\ - H\Big(\bar y(\bar s_j),p_0,p(\bar s_j),0,\lambda,\bar w^0(\bar s_j),\bar w(\bar s_j),\bar\alpha(\bar s_j)\Big)\le 0,$$

while, for a bracket-like variation generator $\mathbf{c}_j=B_j$, we obtain $p(\bar s_j)\cdot B_j(\bar y(\bar s_j))\le 0$.

6.5. Conclusion of the proof. To conclude the proof we need to extend the previous inequalities to almost all $s\in[0,\bar S]$ and all variations generators $\mathbf{c}\in\mathfrak{V}$. This will be achieved via density arguments coupled with infinite intersection criteria. Though this is a quite standard procedure, we give the details for the sake of completeness. By Lusin's theorem, one has that $(0,\bar S)_{\mathrm{Leb}}=\bigcup_{k=0}^{+\infty}E_k$, where E_0 has null measure and, for every $k\in\mathbb{N}$, the set E_k is compact and the restriction to E_k of the measurable map considered in Definition 6.3 is continuous. For every k, let $D_k\subseteq E_k$ be the set of density points[21] of E_k. Since D_k and E_k have the same measure, by the Lebesgue density theorem, $D=\bigcup_{k=0}^{+\infty}D_k\subset[0,\bar S]$ has full measure.

DEFINITION 6.13. *Let F be an arbitrary subset of $D\times\mathfrak{V}$. We say that a triple $(\bar p_0,\bar p,\lambda)\in\mathbb{R}^{1+n+1}$ verifies* (P)$_F$ *if $\lambda\ge 0$ and, setting $p_0:=\bar p_0$, $p(\cdot):=\bar p\cdot M(\bar S,\cdot)$, one has that*

(i) $(p_0,p(\bar S))+\lambda\left(\frac{\partial\bar\Psi}{\partial t}(\bar S)\,,\,\frac{\partial\bar\Psi}{\partial x}(\bar S)\right)\in-\Gamma^\perp$;

(ii) *for every $(s,\mathbf{c})\in F$ with $\mathbf{c}=(w^0,w,a,\zeta)$, the following inequality*

$$H\Big(\bar y(s),p_0,p(s),0,\lambda,w^0(1+\zeta),w(1+\zeta),a\Big)\le H\Big(\bar y(s),p_0,p(s),0,\lambda,\bar w^0(s),\bar w(s),\bar\alpha(s)\Big)$$

holds true, while for every $(s,\mathbf{c})\in F$ such that $\mathbf{c}=B\in\mathfrak{B}^0$, $p(s)\cdot B(\bar y(s))\le 0$.

For any given subset $F\subset D\times\mathfrak{V}$, let us set

$$\Lambda(F):=\Big\{(\bar p_0,\bar p,\lambda)\in\mathbb{R}^{1+n+1}:\ |(\bar p_0,\bar p,\lambda)|=1,\ (\bar p_0,\bar p,\lambda)\text{ verifies }(\mathrm{P})_F\Big\}.$$

Our goal consists in showing that $\Lambda(F)\ne\emptyset$ for some F comprising pairs $(s,\mathbf{c})$, such that the union of all times s is a full measure subset of $[0,\bar S]$ and $\mathbf{c}$ can range over all $\mathfrak{V}$. Clearly, for arbitrary subsets F_1, F_2 of $D\times\mathfrak{V}$ the sets $\Lambda(F_1),\Lambda(F_2)$, if not empty, are compact and $\Lambda(F_1\cup F_2)=\Lambda(F_1)\cap\Lambda(F_2)$. By the previous step, $\Lambda(F)\ne\emptyset$ as soon as F is *finite* and of the form

$$\Big\{(\bar s_1,\mathbf{c}_1),\dots,(\bar s_N,\mathbf{c}_N)\Big\}\quad\text{with } 0=:\bar s_0<\bar s_1<\cdots<\bar s_N<\bar S. \tag{6.27}$$

[21] We recall that $t\in\tilde E\subset\mathbb{R}$ is a *density point* for $\tilde E$ if $\lim_{\delta\to0^+}\frac{\mathrm{meas}\left([t-\delta,t+\delta]\cap\tilde E\right)}{2\delta}=1$.

In order to prove that $\Lambda(F) \neq \emptyset$ for an arbitrary *finite* set $F \subset D \times \mathfrak{V}$, we have to show that it is nonempty even when $F = \{(\bar{s}_1, \mathbf{c}_1), \ldots, (\bar{s}_N, \mathbf{c}_N)\}$ with $0 =: \bar{s}_0 \leq \bar{s}_1 \leq \cdots \leq \bar{s}_N < \bar{S}$ and one allows that $\bar{s}_j = \bar{s}_{j+1}$ for some $j = 0, \ldots, N-1$. To this end, observe that every $\bar{s}_j$ belongs to some set of density points D_k, that we denote as $D_{k(j)}$. Hence, there exist sequences $(\bar{s}_{j,i})_{i\in\mathbb{N}}$ for $j = 1, \ldots, N$ such that

$$\bar{s}_{j,i} \in D_{k(j)} \quad \text{and} \quad \bar{s}_{1,i} < \cdots < \bar{s}_{N,i} \quad \text{for all } i \in \mathbb{N}, \quad \text{and} \quad \lim_{i\to+\infty} \bar{s}_{j,i} = \bar{s}_j.$$

For each $i \in \mathbb{N}$, set $F_i := \{(\bar{s}_{1,i}, \mathbf{c}_1), \ldots, (\bar{s}_{N,i}, \mathbf{c}_N)\}$, so that F_i has the form (6.27) and hence $\Lambda(F_i) \neq \emptyset$. For each $i \in \mathbb{N}$, let us select $(\bar{p}_{0_i}, \bar{p}_i, \lambda_i) \in \Lambda(F_i)$. Since $|(\bar{p}_{0_i}, \bar{p}_i, \lambda_i)| = 1$, by possibly taking a subsequence, we can assume that $(\bar{p}_{0_i}, \bar{p}_i, \lambda_i)$ converges to a point $(\bar{p}_0, \bar{p}, \lambda)$ with $|(\bar{p}_0, \bar{p}, \lambda)| = 1$. By the definition of $D_{k(j)} (\subseteq E_{k(j)})$, passing to the limit as $i \to +\infty$ one obtains that $(\bar{p}_0, \bar{p}, \lambda) \in \Lambda(F)$. Hence we have proved that $\Lambda(F) \neq \emptyset$ as soon as $\operatorname{card}(F) < +\infty$.[22] In particular, if we take a finite family of subsets $F_1, \ldots, F_M \subset D \times \mathfrak{V}$ with $\operatorname{card}(F_i) < +\infty$ for all $i = 1, \ldots, M$, we get $\Lambda(F_1) \cap \cdots \cap \Lambda(F_M) = \Lambda\left(\cup_{i=1}^M F_i\right) \neq \emptyset$. Hence $\{\Lambda(F) : F \subset D \times \mathfrak{V},\ \operatorname{card}(F) < +\infty\}$ is a family of compact subsets such that the intersection of each finite subfamily is nonempty. This implies that also the (infinite) intersection of all $\Lambda(F)$ over finite sets F is nonempty. Therefore $\Lambda(D \times \mathfrak{V}) = \Lambda(\bigcup_{\operatorname{card}(F)<+\infty} F) = \bigcap_{\operatorname{card}(F)<+\infty} \Lambda(F) \neq \emptyset$. This means that there exists some covector $(\bar{p}_0, \bar{p}, \lambda) \neq 0$ such that, setting $p_0 := \bar{p}_0$, $p(\cdot) := \bar{p} \cdot M(\bar{S}, \cdot)$ for all time s in the full-measure set D, one gets

$$\begin{aligned} &\mathbf{H}\big(\bar{y}(s), p_0, p(s), 0, \lambda\big) = H\Big(\bar{y}(s), p_0, p(s), 0, \lambda, \bar{w}^0(s), \bar{w}(s), \bar{\alpha}(s)\Big) \\ &\quad = \max_{(w^0,w,a,\zeta)\in W\times A\times\left[-\frac{1}{2},\frac{1}{2}\right]} H\Big(\bar{y}(s), p_0, p(s), 0, \lambda, w^0(1+\zeta), w(1+\zeta), a\Big) \\ &\quad = \max_{\zeta\in\left[-\frac{1}{2},\frac{1}{2}\right]} (1+\zeta)\, \mathbf{H}\big(\bar{y}(s), p_0, p(s), 0, \lambda\big), \end{aligned} \tag{6.28}$$

$$p(s) \cdot B(\bar{y}(s)) \leq 0 \quad \text{for all } B \in \mathfrak{B}^0. \tag{6.29}$$

The first relation in (6.28) coincides with (3.7), while the last one immediately implies (3.8). Finally, observe that $B \in \mathfrak{B}^0$ if and only if $-B \in \mathfrak{B}^0$, so that (6.29) yields (4.2). This concludes the proof, since, in case $\bar{y}(\bar{S}) > 0$, the strengthened nontriviality condition (3.3) can be obtained as in the proof of the first order maximum principle.

Acknowledgment. We are indebted to the anonymous referees for their interesting and useful remarks, which helped us to significantly improve the paper.

REFERENCES

[1] A. Agrachev, D. Barilari, and U. Boscain, *A Comprehensive Introduction to Sub-Riemannian Geometry*, Cambridge University Press, Cambridge, 2019.

[2] A. Agrachev and Y. Sachkov, *Control Theory from the Geometric Viewpoint*, Encyclopaedia Math. Sci. 87, Springer, Berlin, 2004.

[3] M. Aronna, J. Bonnans, A. Dmitruk, and P. Lotito, *Quadratic order conditions for bang-singular extremals*, Numer. Algebra Control Optim., 2 (2012), pp. 511–546.

[4] M. Aronna, M. Motta, and F. Rampazzo, *Infimum gaps for limit solutions*, Set-Valued Var. Anal., 23 (2015), pp. 3–22.

[22] Here $\operatorname{card}(Q)$ denotes the cardinality of the set Q.

[5] M. Aronna, M. Motta, and F. Rampazzo, *Necessary Conditions Involving Lie Brackets for Impulsive Optimal Control Problems*, preprint, arXiv:1903.06109v1, 2019.

[6] M. Aronna and F. Rampazzo, *$\mathcal{L}^1$ limit solutions for control systems*, J. Differential Equations, 258 (2015), pp. 954–979.

[7] A. Arutyunov, V. Dykhta, and F. Pereira, *Necessary conditions for impulsive nonlinear optimal control problems without a priori normality assumptions*, J. Optim. Theory Appl., 124 (2005), pp. 55–77.

[8] A. Arutyunov, D. Karamzin, and F. Pereira, *L. S. Pontryagin's maximum principle for optimal impulsive control problems*, Dokl. Akad. Nauk, 432 (2010), pp. 439–442.

[9] A. Arutyunov, D. Karamzin, F. Pereira, and N. Chernikova, *Second-order necessary optimality conditions in optimal impulsive control problems*, Differ. Equ., 54 (2018), pp. 1083–1101.

[10] D. Azimov and R. Bishop, *New trends in astrodynamics and applications: Optimal trajectories for space guidance*, Ann. New York Acad. Sci., 1065 (2005), pp. 189–209.

[11] M. Barbero-Liñán and M. Muñoz-Lecanda, *Presymplectic high order maximum principle*, Rev. R. Acad. Cienc. Exactas Fis. Nat. Ser. A. Mat. RACSAM, 106 (2012), pp. 97–110.

[12] R.-M. Bianchini and M. Kawski, *Needle variations that cannot be summed*, SIAM J. Control Optim., 42 (2003), pp. 218–238.

[13] R.-M. Bianchini and G. Stefani, *Good needle-like variations*, in Differential Geometry and Control, Proc. Symp. Pure Math. 64, AMS, Providence, RI, 1999, pp. 91–102.

[14] 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, pp. 195–246.

[15] A. Bressan and F. Rampazzo, *On Differential Systems with Vector-Valued Impulsive Controls*, Boll. Unione Mat. Ital. B (7), 2 (1988), pp. 641–656.

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

[17] A. A. Bressan (Aldo) and M. Motta, *A Class of Mechanical Systems with Some Coordinates as Controls. A Reduction of Certain Optimization Problems for Them. Solution Methods*, Atti Accad. Naz. Lincei Mem. Mat. Appl. (9), 2 (1993), p. 7–30.

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

[19] F. Chittaro and G. Stefani, *Minimum-time strong optimality of a singular arc: The multi-input non involutive case*, ESAIM Control Optim. Calc. Var., 22 (2016), pp. 786–810.

[20] V. Dykhta, *The variational maximum principle and quadratic conditions for the optimality of impulse and singular processes*, Sibirsk. Mat. Zh., 35 (1994), pp. 70–82.

[21] H. Federer, *Geometric Measure Theory*, Grundlehren Math. Wiss. 153, Springer, New York, 1969.

[22] E. Feleqi and F. Rampazzo, *Iterated Lie brackets for nonsmooth vector fields*, NoDEA Nonlinear Differential Equations Appl., 24 (2017), pp. 1–43.

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

[24] M. Guerra and A. Sarychev, *Fréchet generalized trajectories and minimizers for variational problems of low coercivity*, J. Dyn. Control Syst., 21 (2015), pp. 351–377.

[25] O. Hájec, *Book review*, Bull. Amer. Math. Soc., 12 (1985), pp. 272–279.

[26] H. Knobloch, *Higher order necessary conditions in optimal control theory*, in Ordinary and Partial Differential Equations, W. Everitt, ed., Lecture Notes in Math. 827, Springer, Berlin, 1980, pp. 151–164.

[27] A. J. Krener, *The high order maximal principle and its application to singular extremals*, SIAM J. Control Optim., 15 (1977), pp. 256–293.

[28] C. Marle, *Géométrie des systemes mécaniques à liaisons actives*, in Symplectic Geometry and Mathematical Physics, Birkhäuser, Boston, 1991, pp. 260–287.

[29] B. Miller and E. Rubinovich, *Impulsive control in continuous and discrete-continuous systems*, Kluwer, New York, 2003.

[30] M. Motta and F. Rampazzo, *Space-time trajectories of nonlinear systems driven by ordinary and impulsive controls*, Differential Integral Equations, 8 (1995), pp. 269–288.

[31] M. Motta, F. Rampazzo, and R. Vinter, *Normality and gap phenomena in optimal unbounded control*, ESAIM Control Optim. Calc. Var., 24 (2018), pp. 1645–1673.

[32] M. Motta and C. Sartori, *On $\mathcal{L}^1$ limit solutions in impulsive control*, Discrete Contin. Dyn. Syst. Ser. S, 11 (2018), pp. 1201–1218.

[33] F. Pereira and G. Silva, *Necessary conditions of optimality for vector-valued impulsive control problems*, Systems Control Lett., 40 (2000), pp. 205–215.

[34] L. Pontryagin, V. Boltyanski, R. Gamkrelidze, and E. Michtchenko, *The Mathematical Theory of Optimal Processes*, Wiley Interscience, New York, 1962.

[35] 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), pp. 405–431.

[36] F. Rampazzo and H. Sussmann, *Set-valued differentials and a nonsmooth version of Chow's theorem*, Proceedings of the 40th IEEE Conference on Decision and Control, Vol. 3, IEEE, Piscataway, NJ, 2001, pp. 2613–2618.

[37] F. Rampazzo and H. Sussmann, *Commutators of flow maps of nonsmooth vector fields*, J. Differential Equations, 232 (2007), pp. 134–175.

[38] R. W. Rishel, *An extended Pontryagin principle for control systems whose control laws contain measures*, J. SIAM Control Ser. A, 3 (1965), pp. 191–205.

[39] H. Schättler and U. Ledzewicz, *Geometric Optimal Control: Theory, Methods and Examples*, Interdiscip. Appl. Math. 38, Springer, Berlin, 2012.

[40] G. Silva and R. Vinter, *Measure driven differential inclusions*, J. Math. Anal. Appl., 202 (1996), pp. 727–746.

[41] G. N. Silva and R. B. Vinter, *Necessary conditions for optimal impulsive control problems*, SIAM J. Control Optim., 35 (1997), pp. 1829–1846.

[42] H. Sussmann, *Geometry and optimal control*, in Mathematical Control Theory, Springer, New York, 1999, pp. 140–198.

[43] J. Warga, *Optimal Control of Differential and Functional Equations*, Academic, New York, 1972.