

## CONTINUITY OF PONTRYAGIN EXTREMALS WITH RESPECT TO DELAYS IN NONLINEAR OPTIMAL CONTROL\*

RICCARDO BONALLI<sup>†</sup>, BRUNO HÉRISSE<sup>‡</sup>, AND EMMANUEL TRÉLAT<sup>§</sup>

**Abstract.** Consider a general nonlinear optimal control problem in finite dimension, with constant state and/or control delays. By the Pontryagin maximum principle, any optimal trajectory is the projection of a Pontryagin extremal. We establish that, under appropriate assumptions which are essentially sharp, Pontryagin extremals depend continuously on the parameters delays, for adequate topologies. The proof of the continuity of the trajectory and of the control is quite easy; however, for the adjoint vector, the proof requires a much finer analysis. The continuity property of the adjoint vector with respect to the parameter delays opens a new perspective for the numerical implementation of indirect methods, such as the shooting method.

**Key words.** nonlinear optimal control, time-delayed systems, Pontryagin extremals, continuity with respect to delays, shooting method for problems with delays

**AMS subject classifications.** 49J15, 49K15, 49K40

**DOI.** 10.1137/18M119121X

**1. Introduction.** This paper is devoted to establishing continuity properties with respect to delays of Pontryagin extremals related to nonlinear optimal control problems with state and control constant delays. Pontryagin extremals are obtained by applying the Pontryagin maximum principle to an optimal control problem, thus providing first-order necessary conditions for optimality.

Historically, the maximum principle has been developed originally for optimal control problems without delays (see, e.g., [32]). The paper [25] was the first to provide a maximum principle for optimal control problems with constant state delays while [21] obtains the same conditions by a simple substitution-like method. In [26] a similar result is achieved for control problems with pure control delays. In [24, 38], necessary conditions are obtained for optimal control problems with multiple constant delays in state and control variables. Moreover, [6, 4] derive maximum principles for control systems with either time- or state-dependent delays. Finally, [18, 19] give necessary conditions for optimal control problems with delays and mixed constraints.

When delays are considered, the maximum principle provides extremals satisfying adjoint equations, the maximality condition, and transversality conditions which depend directly on the value of the delay. Therefore, it seems legitimate to wonder how these extremals depend on the parameter delay. In the present paper, we provide sufficient conditions ensuring continuity (for suitable topologies) of Pontryagin extremals w.r.t. delays. Continuity properties have useful numerical applications in solving optimal control problems with delays by shooting methods (as we describe in section 2.4). The main result presented in this paper is roughly the following:

---

\*Received by the editors May 30, 2018; accepted for publication (in revised form) February 6, 2019; published electronically April 18, 2019.

<http://www.siam.org/journals/sicon/57-2/M119121.html>

<sup>†</sup>Sorbonne Université, Université Paris-Diderot SPC, CNRS, Inria, Laboratoire Jacques-Louis Lions, équipe CAGE, F-75005 Paris, France and DTIS, ONERA, Université Paris Saclay, F-91123 Palaiseau, France (rbonalli@stanford.edu, riccardo.bonalli@etu.upmc.fr, riccardo.bonalli@onera.fr).

<sup>‡</sup>DTIS, ONERA, Université Paris Saclay, F-91123 Palaiseau, France (bruno.herisse@onera.fr).

<sup>§</sup>Sorbonne Université, Université Paris-Diderot SPC, CNRS, Inria, Laboratoire Jacques-Louis Lions, équipe CAGE, F-75005 Paris, France (emmanuel.trelat@sorbonne-universite.fr).

Consider an optimal control problem with delays. Under the main assumption that optimal controls are either time continuous or purely bang-bang, Pontryagin extremals are strongly continuous w.r.t. delays, for appropriate topologies.

The assumption that optimal controls are bang-bang is sharp, in the sense that, whenever singular arcs arise, the continuity of Pontryagin extremals w.r.t. delays may fail to be satisfied in strong topology, although it is always satisfied in a weak sense. (A counterexample is provided in section 2.3.)

In the literature, to our knowledge, it seems that this topic has been little addressed. The main works addressing the regularity of extremals w.r.t. time lag variations develop sensibility analysis-type arguments (see, e.g., [34, 29, 30]). Our approach developed in this paper does not require any differentiability properties of the extremals. More precisely, the main continuity result (see Theorem 2.1) is achieved by analyzing the geometric deformation of Pontryagin cones, i.e., the sets containing all variation vectors, under small perturbations of the delays. This geometric analysis is challenging and requires a modified conic implicit function theorem which relies on the continuous dependence of parameters (represented here by delays).

The paper is organized as follows. In section 2, we recall the maximum principle formulation for optimal control problems with delays, stating then the main theorem for the continuity of Pontryagin extremals w.r.t. delays. Section 2.4 contains an extension of such continuity property to implement robust shooting methods to solve optimal control problems with delays. In section 3, we provide the proof of our main result, which goes in three steps. Robustness of controllability properties of problems with delays are addressed first by means of an implicit function theorem in which parameters and restriction to dense subsets are considered. In a second step, the existence of solutions of optimal control problems with delays and their continuity w.r.t. delays are established. The last step is the more difficult and technical: we prove the continuity of the adjoint vectors w.r.t. delays. Finally, section 4 provides conclusions and several perspectives.

## 2. Continuity of Pontryagin extremals w.r.t. delays.

### 2.1. The maximum principle for optimal control problems with delays.

Let  $n, m$  be positive integers,  $\Delta$  a positive real number, and  $U \subseteq \mathbb{R}^m$  a measurable subset and define an initial state function  $\phi^1(\cdot) \in C^0([-\Delta, 0], \mathbb{R}^n)$  and an initial control function  $\phi^2(\cdot) \in L^\infty([-\Delta, 0], U)$ . For  $\tau = (\tau^0, \tau^1, \tau^2) \in [0, \Delta]^3$  and  $t_f > 0$ , consider the following nonlinear control system on  $\mathbb{R}^n$  with constant delays:

$$(1) \quad \begin{cases} \dot{x}(t) = f(t, t - \tau^0, x(t), x(t - \tau^1), u(t), u(t - \tau^2)), & t \in [0, t_f], \\ x(t) = \phi^1(t), \quad u(t) = \phi^2(t), \quad t \in [-\Delta, 0], \quad u(\cdot) \in L^\infty([-\Delta, t_f], U), \end{cases}$$

where  $f(t, s, x, y, u, v)$  is continuous and (at least)  $C^2$  w.r.t. its second, third, and fourth variables. Control systems (1) play an important role describing many relevant phenomena in physics, biology, engineering, and economics (see, e.g., [28]).

Let  $M_f$  be a subset of  $\mathbb{R}^n$ . Assume that  $M_f$  is reachable from  $\phi^1(\cdot)$  for the control system (1), that is, for every  $\tau = (\tau^0, \tau^1, \tau^2) \in [0, \Delta]^3$ , there exist a final time  $t_f$  and a control  $u(\cdot) \in L^\infty([-\Delta, t_f], U)$ , such that the trajectory  $x(\cdot)$ , solution of (1) in  $[-\Delta, t_f]$ , satisfies  $x(t_f) \in M_f$ . Such a control is called admissible and we denote by  $\mathcal{U}_{t_f, \mathbb{R}^m}^\tau$  the set of all admissible controls of (1) defined in  $[-\Delta, t_f]$  taking their values in  $\mathbb{R}^m$ , while  $\mathcal{U}_{t_f, U}^\tau$  denotes the set of all admissible controls of (1) defined in  $[-\Delta, t_f]$  taking their values in  $U$ . Therefore,  $\mathcal{U}_{\mathbb{R}^m}^\tau = \bigcup_{t_f > 0} \mathcal{U}_{t_f, \mathbb{R}^m}^\tau$  and  $\mathcal{U}_U^\tau = \bigcup_{t_f > 0} \mathcal{U}_{t_f, U}^\tau$ .

Given constant delays  $\tau = (\tau^0, \tau^1, \tau^2) \in [0, \Delta]^3$ , we consider the optimal control problem with delays  $(\text{OCP})_\tau$  consisting in steering the control system (1) to  $M_f$ , while minimizing the cost function

$$(2) \quad C_\tau(t_f, u) = \int_0^{t_f} f^0(t, t - \tau^0, x(t), x(t - \tau^1), u(t), u(t - \tau^2)) \, dt,$$

where  $f^0(t, s, x, y, u, v)$  is continuous and (at least)  $C^2$  w.r.t. its second, third, and fourth variables. We study either fixed or free final time problems  $(\text{OCP})_\tau$ .

In the context of the present work, we focus on two particular classes of problems. We speak of problems  $(\text{OCP})_\tau$  with pure state delays when  $f$  and  $f^0$  do not depend on  $v$ , i.e.,  $\frac{\partial f}{\partial v} = \frac{\partial f^0}{\partial v} = 0$ . We say that the optimal control problem  $(\text{OCP})_\tau$  is affine when  $f$  and  $f^0$  are affine in  $(u, v)$ , i.e.,

$$(3) \quad \begin{aligned} f(t, s, x, y, u, v) &= f_0(t, s, x, y) + u \cdot f_1(t, s, x, y) + v \cdot f_2(t, s, x, y), \\ f^0(t, s, x, y, u, v) &= f_0^0(t, s, x, y) + u \cdot f_1^0(t, s, x, y) + v \cdot f_2^0(t, s, x, y). \end{aligned}$$

Assume that  $(x_\tau(\cdot), u_\tau(\cdot))$  is an optimal solution for  $(\text{OCP})_\tau$  with related optimal final time  $t_f^\tau$  and define the Hamiltonian related to problem  $(\text{OCP})_\tau$  by

$$H(t, s, x, y, p, p^0, u, v) = \langle p, f(t, s, x, y, u, v) \rangle + p^0 f^0(t, s, x, y, u, v).$$

According to the maximum principle (see, e.g., [19, 7]), there exists a nontrivial couple  $(p_\tau(\cdot), p_\tau^0) \neq 0$ , where  $p_\tau^0 \leq 0$  is constant and  $p_\tau : [0, t_f^\tau] \rightarrow \mathbb{R}^n$  (adjoint vector) is absolutely continuous, such that the so-called Pontryagin extremal  $(x_\tau(\cdot), p_\tau(\cdot), p_\tau^0, u_\tau(\cdot))$  satisfies, almost everywhere in  $[0, t_f^\tau]$ , the adjoint equations

$$(4) \quad \begin{cases} \dot{x}_\tau(t) = \frac{\partial H}{\partial p}(t, t - \tau^0, x_\tau(t), x_\tau(t - \tau^1), p_\tau(t), p_\tau^0, u_\tau(t), u_\tau(t - \tau^2)), \\ \dot{p}_\tau(t) = -\frac{\partial H}{\partial x}(t, t - \tau^0, x_\tau(t), x_\tau(t - \tau^1), p_\tau(t), p_\tau^0, u_\tau(t), u_\tau(t - \tau^2)) \\ \quad - \mathbb{1}_{[0, t_f^\tau - \tau^1]}(t) \frac{\partial H}{\partial y}(t + \tau^1, t + \tau^1 - \tau^0, x_\tau(t + \tau^1), x_\tau(t), p_\tau(t + \tau^1), \\ \quad \quad \quad p_\tau^0, u_\tau(t + \tau^1), u_\tau(t + \tau^1 - \tau^2)) \end{cases}$$

and the following maximality condition: for every  $u \in U$

$$(5) \quad \begin{aligned} &H(t, t - \tau^0, x_\tau(t), x_\tau(t - \tau^1), p_\tau(t), p_\tau^0, u_\tau(t), u_\tau(t - \tau^2))(t) \\ &\quad + \mathbb{1}_{[0, t_f^\tau - \tau^2]}(t) H(t + \tau^2, t + \tau^2 - \tau^0, x_\tau(t + \tau^2), x_\tau(t + \tau^2 - \tau^1), \\ &\quad \quad \quad p_\tau(t + \tau^2), p_\tau^0, u_\tau(t + \tau^2), u_\tau(t)) \\ &\geq H(t, t - \tau^0, x_\tau(t), x_\tau(t - \tau^1), p_\tau(t), p_\tau^0, u, u_\tau(t - \tau^2)) \\ &\quad + \mathbb{1}_{[0, t_f^\tau - \tau^2]}(t) H(t + \tau^2, t + \tau^2 - \tau^0, x_\tau(t + \tau^2), x_\tau(t + \tau^2 - \tau^1), \\ &\quad \quad \quad p_\tau(t + \tau^2), p_\tau^0, u_\tau(t + \tau^2), u). \end{aligned}$$

Furthermore, if  $M_f$  is a submanifold of  $\mathbb{R}^n$ , locally around  $x_\tau(t_f^\tau)$ , then the adjoint vector can be chosen in order to satisfy

$$(6) \quad p_\tau(t_f^\tau) \perp T_{x_\tau(t_f^\tau)} M_f$$

and, moreover, if the final time  $t_f^\tau$  is free and both  $t_f^\tau, t_f^\tau - \tau^2$  are Lebesgue points for

$u_\tau(\cdot)$ , the extremal  $(x_\tau(\cdot), p_\tau(\cdot), p_\tau^0, u_\tau(\cdot))$  satisfies the following final condition:

$$(7) \quad H(t_f^\tau, t_f^\tau - \tau^0, x_\tau(t_f^\tau), x_\tau(t_f^\tau - \tau^1), p_\tau(t_f^\tau), p_\tau^0, u_\tau(t_f^\tau), u_\tau(t_f^\tau - \tau^2)) = 0.$$

(Recall that any measurable function is a.e. approximately continuous; see, e.g., [15].) When  $t_f^\tau$  or  $t_f^\tau - \tau^2$  are not Lebesgue points, (7) can be generalized (see, e.g., [7]). The extremal  $(x_\tau(\cdot), p_\tau(\cdot), p_\tau^0, u_\tau(\cdot))$  is said to be normal when  $p_\tau^0 \neq 0$ , and in that case we set  $p_\tau^0 = -1$ . Otherwise, it is said to be abnormal.

**2.2. The main result: Continuity properties of Pontryagin extremals.**

Our main result establishes that, under appropriate assumptions, Pontryagin extremals are continuous w.r.t. delays for appropriate topologies. The most challenging issue is the continuous dependence of adjoint vectors w.r.t. delays. To prove this fact, we establish continuity w.r.t. delays of Pontryagin cones related to the maximum principle formulation with delays.

We will treat separately the case of pure state delays. Treating control delays happens to be more complex, especially for the existence of optimal controls (see Theorem 2.1). Indeed, a standard approach to prove existence would consider usual Filippov’s assumptions (as in the classical reference [16]) which, in the case of control delays, must be extended. In particular, using the Guinn’s reduction (see, e.g., [21]), the control system with delays is equivalent to a nondelayed system with a larger number of variables depending on the value of  $\tau^2$ . Such extension was used in [31]. However, the usual assumption on the convexity of the epigraph of the extended dynamics is not sufficient to prove Lemma 2.1 in [31] (see also section 3.3.2).

Fix constant delays  $\tau_0 = (\tau_0^0, \tau_0^1, \tau_0^2) \in [0, \Delta]^3$  and let  $(x_{\tau_0}(\cdot), p_{\tau_0}(\cdot), p_{\tau_0}^0, u_{\tau_0}(\cdot))$  be a Pontryagin extremal for  $(\text{OCP})_{\tau_0}$  satisfying (4)–(7), where  $(x_{\tau_0}(\cdot), u_{\tau_0}(\cdot))$  is an optimal solution for  $(\text{OCP})_{\tau_0}$  in  $[-\Delta, t_f^{\tau_0}]$ . We make the following assumptions:

**General assumptions.**

- (A)  $\left\{ \begin{array}{l} (A_1) \ U \text{ is compact and convex in } \mathbb{R}^m, \text{ and } M_f \text{ is a compact submanifold of } \mathbb{R}^n. \\ (A_2) \ \text{The optimal control problem with delays } (\text{OCP})_{\tau_0} \text{ has a unique solution, denoted } (x_{\tau_0}(\cdot), u_{\tau_0}(\cdot)), \text{ defined in a neighborhood of } [-\Delta, t_f^{\tau_0}]. \\ (A_3) \ \text{The optimal trajectory } x_{\tau_0}(\cdot) \text{ has a unique extremal lift (up to a multiplicative scalar) defined in } [0, t_f^{\tau_0}], \text{ which is normal, denoted } (x_{\tau_0}(\cdot), p_{\tau_0}(\cdot), -1, u_{\tau_0}(\cdot)), \text{ solution of the maximum principle.} \\ (A_4) \ \text{There exists a positive real number } b \text{ such that, for every } \tau = (\tau^0, \tau^1, \tau^2) \in [0, \Delta]^3 \text{ and every } v(\cdot) \in \mathcal{U}_U^\tau, \text{ denoting } x_{\tau,v}(\cdot) \text{ the related trajectory arising from dynamics (1) with final time } t_f^{\tau,v}, \text{ one has} \\ \qquad \qquad \qquad \forall t \in [-\Delta, t_f^{\tau,v}] : t_f^{\tau,v} + \|x_{\tau,v}(t)\| \leq b. \end{array} \right.$

**Additional assumptions in the case of pure state delays.**

- (B)  $\left\{ \begin{array}{l} (B_1) \ \text{For every } \tau, \text{ every optimal control } u_\tau(\cdot) \text{ of } (\text{OCP})_\tau \text{ is continuous.} \\ (B_2) \ \text{The sets} \\ \qquad \qquad \qquad \{ (f(t, s, x, y, u), f^0(t, s, x, y, u) + \gamma) : u \in U, \gamma \geq 0 \}, \\ \qquad \qquad \qquad \left\{ \left( f(t, t, x, x, u), f^0(t, t, x, x, u) + \gamma, \frac{\partial \tilde{f}}{\partial x}(t, t, x, x, u), \frac{\partial \tilde{f}}{\partial y}(t, t, x, x, u) \right) : u \in U, \gamma \geq 0 \right\} \\ \qquad \qquad \qquad \text{are convex for every } t, s \in \mathbb{R}, x, y \in \mathbb{R}^n, \text{ where we denote } \tilde{f} = (f, f^0). \end{array} \right.$

**Additional assumptions in the case of delays in state and control variables.**

- (C)  $\left\{ \begin{array}{l} (C_1) \text{ The considered optimal control problems are affine, i.e., we have (3).} \\ (C_2) \text{ The final time } t_f \text{ is fixed.} \\ (C_3) \text{ At least one of the following two conditions is satisfied:} \\ \quad \bullet \text{ For every } \tau, \text{ every optimal control } u_\tau(\cdot) \text{ of } (\text{OCP})_\tau \text{ is continuous.} \\ \quad \bullet \text{ The control } u_{\tau_0}(\cdot) \text{ takes its values at extremal points of } U, \text{ a.e.} \end{array} \right.$

**THEOREM 2.1.**

- *Optimal control problems with pure state delays:*  
Denote  $B_\varepsilon^+(\tau) = B_\varepsilon(\tau) \cap \mathbb{R}_+^2$ . Under assumptions (A) and (B) the following hold:
  1. There exists  $\varepsilon_0 > 0$  such that, for every couple of delays  $\tau = (\tau^0, \tau^1)$  satisfying  $\|\tau - \tau_0\| < \varepsilon_0$ , each problem  $(\text{OCP})_\tau$  has at least one optimal solution  $(x_\tau(\cdot), u_\tau(\cdot))$  in  $[-\Delta, t_f]$ , every extremal lift of which is normal. Moreover, if the final time is fixed, then  $t_f^\tau = t_f^{\tau_0}$  for every  $\tau$ .
  2. The mappings  $B_{\varepsilon_0}^+(\tau_0) \ni \tau \mapsto x_\tau(\cdot)^1$  and  $B_{\varepsilon_0}^+(\tau_0) \ni \tau \mapsto p_\tau(\cdot)$  are continuous at  $\tau_0$  in  $C^0$  topology, and  $B_{\varepsilon_0}^+(\tau_0) \ni \tau \mapsto t_f^\tau$  is continuous at  $\tau_0$ .
  3. In addition, the mapping  $B_{\varepsilon_0}^+(\tau_0) \ni \tau \mapsto \dot{x}_\tau(\cdot)$  is continuous at  $\tau_0$  for the  $L^\infty$  weak star topology.
- *Optimal control problems with state and control delays:*  
Denote  $B_\varepsilon^+(\tau) = B_\varepsilon(\tau) \cap \mathbb{R}_+^3$ . Under assumptions (A) and (C) the following hold:
  1. There exists  $\varepsilon_0 > 0$  such that, for every triple of delays  $\tau = (\tau^0, \tau^1, \tau^2)$  satisfying  $\|\tau - \tau_0\| < \varepsilon_0$ , each problem  $(\text{OCP})_\tau$  has at least one optimal solution  $(x_\tau(\cdot), u_\tau(\cdot))$  in  $[-\Delta, t_f]$ , every extremal lift of which is normal.
  2. The mappings  $B_{\varepsilon_0}^+(\tau_0) \ni \tau \mapsto x_\tau(\cdot)$  and  $B_{\varepsilon_0}^+(\tau_0) \ni \tau \mapsto p_\tau(\cdot)$  are continuous at  $\tau_0$  in  $C^0$  topology.
  3. In addition, the mapping  $B_{\varepsilon_0}^+(\tau_0) \ni \tau \mapsto (u_\tau(\cdot), u_\tau(\cdot - \tau^2))$  is continuous at  $\tau_0$  for the  $L^2$  weak topology. Moreover, if  $u_{\tau_0}(\cdot)$  takes its values at extremal points of  $U$ , then the mapping  $B_{\varepsilon_0}^+(\tau_0) \ni \tau \mapsto (u_\tau(\cdot), u_\tau(\cdot - \tau^2))$  is continuous at  $\tau_0$  in  $L^\infty$  topology.
  4. When  $\tau^2 = 0$ , the previous conclusions remain valid for problems in free final time, and then the mapping  $B_{\varepsilon_0}^+(\tau_0) \ni \tau \mapsto t_f^\tau$  is continuous at  $\tau_0$ .

Several remarks are in order.

First, assumptions  $(A_2)$  and  $(A_3)$  on the uniqueness of the solution of  $(\text{OCP})_{\tau_0}$  and on the uniqueness of its extremal lift are “generic”: they are actually related to the differentiability properties of the value function (see, e.g., [12, 5, 33]). They are standard in optimization and are just made to keep a nice statement (see Theorem 2.1). These assumptions can be weakened as follows. If we replace  $(A_2)$  and  $(A_3)$  with the assumption “every extremal lift of every solution of  $(\text{OCP})_{\tau_0}$  is normal,” then the conclusion provided in Theorem 2.1 below still holds, except that the continuity properties must be written in terms of closure points (see section 3.3 and [22, Remark 1.11]). Finally, requiring that the unique extremal of  $(\text{OCP})_{\tau_0}$  is, moreover, normal is crucial to prove the continuity of the adjoint vectors w.r.t. delays.

---

<sup>1</sup>We notice that this correspondence is not necessarily uniquely defined. However, thanks to the previous existence statement, at least one such correspondence exists in a neighborhood of  $\tau_0$  and the continuity properties hold for every such correspondence given in the following.

Assumptions  $(B_1)$  and  $(C_3)$  play a complementary role in proving continuity for the adjoint vectors (see section 3.3.4). They are related to the strict Legendre–Clebsch condition and the uniqueness of solutions for (5), and, even if these assumptions seem to be restrictive, many problems in applications satisfy them and examples can be found in [18, 19, 34, 8, 37, 35, 20]. In particular, assumption  $(C_3)$  is instrumentally used also to derive strong continuity of controls from weak continuity when there are delays on the control, because of the following general fact. Let  $X, Y$  be Banach spaces and  $F : X \rightarrow Y$  be a continuous map. Suppose that  $(x_k)_{k \in \mathbb{N}} \subseteq X$  is a sequence such that  $x_k \rightharpoonup x$  and  $F(x_k) \rightharpoonup F(\bar{x})$  for some  $x, \bar{x} \in X$ . Therefore, in general, we cannot ensure that the two limits coincide, that is,  $x = \bar{x}$ , except when  $F$  is linear and injective. On the other hand, assumption  $(C_2)$  becomes essential to ensure continuity of Pontryagin cones when there are delays in the control variables (see section 3.3.4).

When the problem is control-affine, we assume in (one of the two possibilities of) assumption  $(C_3)$  that the optimal control  $u_{\tau_0}(\cdot)$  takes its values in the extremal set of  $U$ . This is the case when the optimal control is bang-bang. As said above, this property permits us to turn weak continuity into strong continuity. Such an assumption is sharp: the counterexample of section 2.3 shows that, when the optimal control of  $(\text{OCP})_{\tau_0}$  does not take its values at extremal points of  $U$ , only a weak continuity of optimal controls can be ensured in general.

*Remark 2.2.* The conclusions of Theorem 2.1 are valid for optimal control problems with control-affine dynamics and quadratic cost

$$\int_0^{t_f} \left( K_1 \|x(t)\|^2 + K_2 \|x(t - \tau^1)\|^2 + K_3 \|u(t)\|^2 + K_4 \|u(t - \tau^2)\|^2 \right) dt$$

(see section 3.3 and the proof in [8, Theorem 1]).

**2.3. Weak continuity versus strong continuity of optimal controls.** For control-affine problems, Theorem 2.1 ensures weak continuity in  $L^2$  of optimal controls. Moreover, when the optimal control  $u_{\tau_0}(\cdot)$  takes its values at extremal points of  $U$  a.e., continuity is true in strong  $L^\infty$  topology. In this section, we provide an example where  $u_{\tau_0}(\cdot)$  does not take its values at extremal points of  $U$ , and continuity fails in strong topology. This example shows that our assumptions are sharp.

Adapting arguments from [36], consider  $(\text{OCP})_\tau$  given by

$$(8) \quad \begin{cases} \min \int_0^{t_f} 1 \, dt, & (u^1(t))^2 + (u^2(t))^2 \leq 1, \\ \dot{x}^1(t) = 1 - (x^2(t))^2 + \tau u^2(t)g(x^1(t)), & x^1(0) = 0, \, x^1(t_f) = 1, \\ \dot{x}^2(t) = u^1(t) + \tau u^2(t)h(x^1(t)), & x^2(0) = 0, \, x^2(t_f) = 0, \end{cases}$$

where  $g$  and  $h$  are smooth functions, to be chosen. In this case

$$f(t, s, x^1, x^2, u^1, u^2) = \begin{pmatrix} 1 - (x^2)^2 + (t - s)u^2g(x^1) \\ u^1 + (t - s)u^2h(x^1) \end{pmatrix}.$$

Taking  $\tau_0 = 0$ , under appropriate assumptions on  $g$  and  $h$ , Theorem 2.1 applies with weak convergence in  $L^2$  of controls, but no strong convergence of controls arises. The proof follows closely the arguments of [36]; therefore, we just recall the main steps.

First, when  $\tau = 0$ , problem (8) has the unique solution  $(u^1, u^2)(t) = 0$  with unique extremal  $(x^1, x^2, p^1, p^2, p^0, u^1, u^2)(t) = (t, 0, 1, 0, -1, 0, 0)$ , where  $t_f = 1$ .

Remark that, in the case in which Theorem 2.1 applies, only the weak convergence in  $L^2$  of optimal controls is ensured. The assumptions of Theorem 2.1 hold; in particular, every optimal control of  $(\text{OCP})_\tau$  is continuous. Indeed, in [36] it is shown that, in the case  $\tau \neq 0$ , under the assumption that function  $g$  may only vanish on a subset of zero measure, the optimal controls related to problem (8) are

$$(9) \quad u_\tau^1(t) = \frac{p_\tau^2(t)}{\sqrt{\phi(t)}}, \quad u_\tau^2(t) = \frac{\tau \left( p_\tau^1(t)g(x_\tau^1(t)) + p_\tau^2(t)h(x_\tau^1(t)) \right)}{\sqrt{\phi(t)}},$$

where  $\phi(t) = (p_\tau^2(t))^2 + \tau^2 (p_\tau^1(t)g(x_\tau^1(t)) + p_\tau^2(t)h(x_\tau^1(t)))^2$ , and the adjoint coordinate  $p_\tau^2$  may only vanish on subsets of zero measure. It follows that optimal controls (9) are continuous, and therefore, Theorem 2.1 applies.

Both  $u_\tau^1$  and  $u_\tau^2$  converge weakly to 0 as soon as  $\tau$  tends to 0. However, in [36] it is also proved that specific choices of highly oscillating functions  $g$  and  $h$  show that  $u_\tau^1$  and  $u_\tau^2$  cannot converge almost everywhere to 0 when  $\tau$  tends to 0.

**2.4. Application to shooting methods.** In this section, we briefly discuss how the continuity properties of Pontryagin extremals in Theorem 2.1 may be exploited to solve optimal control problems with delays via shooting methods. Analyzing in detail conditions ensuring convergence is beyond the scope of the present paper and will be done elsewhere. Let us, however, describe the idea in a quite informal way.

It is known that solving generic  $(\text{OCP})_\tau$  via shooting methods may be difficult. A first issue is to ensure that the shooting function is piecewise  $C^1$ . This is the case, under general assumptions, in at least two situations. The first is when the extremal control can be expressed as a  $C^1$  function of  $(x_\tau(\cdot), p_\tau(\cdot))$  (“smooth optimal control”; see [9]), which is satisfied, at least locally, under the absence of control delays and some Legendre–Clebsch condition. The second is when the extremal control is bang-bang, with a finite number of switchings, under generic assumptions (among which is the absence of control delays) ensuring that the shooting function is a finite composition of  $C^1$  mappings (see, e.g., [36]). Examples for which these hold have been cited in section 2.2. Under such assumptions, each iteration of a shooting method consists of solving the coupled dynamics (4), where a value of  $p_\tau(0)$  is provided. This means that one has to solve a differential-difference boundary value problem where both forward and backward terms of time appear within mixed type differential equations. It follows that, in order to initialize successfully a shooting method for (4), a guess of the initial value of the adjoint vector  $p_\tau(0)$  is not sufficient, but rather, a good numerical guess of the whole function  $p_\tau(\cdot)$  must be provided to make the procedure converge. This represents an additional difficulty w.r.t. the usual shooting method and a global discretization of the system of differential equations (4) must be performed. The result stated in Theorem 2.1 suggests that one may solve  $(\text{OCP})_\tau$  numerically via shooting methods iteratively, starting from the solution of its nondelayed version  $(\text{OCP}) = (\text{OCP})_{\tau=0}$ , and this by means of homotopy procedures.

The basic idea of homotopy methods is to solve a difficult problem step by step, starting from a simpler problem, by parameter deformation. The theory and practice of homotopy methods are well known (see, e.g., [2]). Combined with the shooting problem derived from the maximum principle, a homotopy method consists in deforming the problem into a simpler one (that can be easily solved) and then in solving a series of shooting problems step by step to come back to the original problem. In many situations, exploiting the nondelayed version of the maximum principle mixed with other techniques (such as geometric control, dynamical system theory applied to

mission design, etc.; we refer the reader to [40] for a survey), one is able to initialize efficiently a shooting method on the optimal control problem without delays (OCP). Thus, it is legitimate to wonder if one may solve  $(\text{OCP})_\tau$  by shooting methods starting a homotopy procedure where delays  $\tau$  represent the deformation parameter and (OCP) is taken as the starting problem. This approach is a way to address the flaw of shooting methods applied to  $(\text{OCP})_\tau$ : on one hand, the global adjoint vector for (OCP) could be used to initialize efficiently a shooting method on (4) and, on the other hand, we could solve (4) via usual iterative methods for ODEs (for example, by using the global state solution at the previous iteration). In the context of Theorem 2.1, the previous homotopy procedure is well-posed. Indeed, assuming that the delay  $\tau$  is small enough and the shooting function is piecewise  $C^1$  (see our comments above), the continuity properties stated by Theorem 2.1 applied at  $\tau = 0$  allow straightforwardly the homotopic path to converge toward extremals related to the problem with delays  $(\text{OCP})_\tau$  when starting from the extremal of the problem without delays (OCP).

Implementing homotopy methods combined with the maximum principle with delays, as briefly described above, however, requires a number of considerations which go beyond the scope of the present paper. In a forthcoming work, we will show that this approach indeed happens to be competitive w.r.t. other approaches (in particular, direct methods that have been classically applied in this context; see, e.g., [18, 19]) and we will illustrate it on a nontrivial and nonacademic optimal control problem with launch vehicles, a case of application where delays are the main obstacle to efficiently embark on numerical approaches. (The authors already showed the numerical efficiency of this approach on a simpler problem; see [8].)

**3. Proof of the main result.** Without loss of generality, we assume that  $\tau_0 = 0$ , denoting by  $(\text{OCP}) = (\text{OCP})_{\tau_0}$  the problem without delays. Furthermore, we denote  $t_f = t_f^{\tau_0}$ ,  $x(\cdot) = x_{\tau_0}(\cdot)$ ,  $p(\cdot) = p_{\tau_0}(\cdot)$  and  $u(\cdot) = u_{\tau_0}(\cdot)$ .

The proof of the convergence of extremals for  $(\text{OCP})_\tau$  to the extremal of (OCP), as  $\tau$  tends to 0, is organized in three steps. First, by using assumptions on the nondelayed version of the problem only, we infer the controllability of problems  $(\text{OCP})_\tau$  for every  $\tau$  small enough. The previous step requires some implicit function theorem involving parameters. This allows us to proceed to the second part, which consists in showing the existence of solutions for  $(\text{OCP})_\tau$  for  $\tau$  sufficiently small, and their convergences, as  $\tau$  tends to 0, to solutions of (OCP). In the case of control and state delays, we will see the importance of considering control-affine systems throughout this step. Finally, we address the more difficult issue of establishing the convergence, as  $\tau$  tends to 0, of the adjoint vectors related to  $(\text{OCP})_\tau$  to the adjoint vector of (OCP): a refined analysis on the convergence of Pontryagin cones is needed.

We recall in section 3.1 the main steps of the proof related to the maximum principle for problems  $(\text{OCP})_\tau$  with delays. To our knowledge, the proof of this result via needle-like variations does not appear explicitly in the literature, and for the comprehension of the whole proof, we report it. Section 3.2 provides a useful conic version of the implicit function theorem, depending on parameters, which is a key element for the entire reasoning. Finally, in section 3.3, we report the whole proof of the desired result, as detailed above.

**3.1. Proof of the Pontryagin maximum principle using needle-like variations.** In this section we sketch the proof of the maximum principle for  $(\text{OCP})_\tau$  using needle-like variations. For this, we do not use the assumptions of Theorem 2.1, giving the result for a larger class of control systems with constant delays. Our reasoning is valid as well for problems with free final time, since we do not employ the well known



reduction to a fixed final time problem, but rather, we modify the Pontryagin cone to keep track of the free variable  $t_f^\tau$ , by making  $L^1$ -variations on  $t_f^\tau$  (as in [25], for pure state delays).

**3.1.1. Preliminary notation.** Fix a constant delay  $\tau = (\tau^0, \tau^1, \tau^2) \in [0, \Delta]^3$ . Consider  $(\text{OCP})_\tau$  as given by formulation (1)–(2) and let  $(x_\tau(\cdot), u_\tau(\cdot))$  be an optimal solution defined in  $[-\Delta, t_f^\tau]$ . For every positive final time  $\bar{t}_f$ , introduce the instantaneous cost function  $x_\tau^0(\cdot)$  defined in  $[-\Delta, \bar{t}_f]$  and a solution of

$$\begin{cases} \dot{x}^0(t) = f^0(t, t - \tau^0, x_\tau(t), x_\tau(t - \tau^1), u_\tau(t), u_\tau(t - \tau^2)), & t \in [0, \bar{t}_f], \\ x^0(t) = 0, & t \in [-\Delta, 0], \end{cases}$$

such that (2) provides  $C_\tau(\bar{t}_f, u_\tau) = x_\tau^0(\bar{t}_f)$ . We define the extended state  $\tilde{x} = (x, x^0)$  and the extended dynamics  $\tilde{f}(t, s, \tilde{x}, \tilde{y}, u, v) = (f(t, s, x, y, u, v), f^0(t, s, x, y, u, v))$  for which we will often denote  $\tilde{f}(t, s, x, y, u, v) = \tilde{f}(t, s, \tilde{x}, \tilde{y}, u, v)$ . Consider the extended dynamical problem in  $\mathbb{R}^{n+1}$

$$(10) \quad \begin{cases} \dot{\tilde{x}}(t) = \tilde{f}(t, t - \tau^0, \tilde{x}(t), \tilde{x}(t - \tau^1), u(t), u(t - \tau^2)), & t \in [0, \bar{t}_f], \\ \tilde{x}|_{[-\Delta, 0]}(t) = (\phi^1(t), 0), & \tilde{x}(\bar{t}_f) \in M_f \times \mathbb{R}, \\ u(\cdot) \in L^\infty([-\Delta, \bar{t}_f], U), & u|_{[-\Delta, 0]}(t) = \phi^2(t). \end{cases}$$

As provided in section 2.1, the set of all admissible controls of (10) in  $[-\Delta, \bar{t}_f]$  taking their values in  $\mathbb{R}^m$  is denoted by  $\tilde{\mathcal{U}}_{\bar{t}_f, \mathbb{R}^m}^\tau$ , while  $\tilde{\mathcal{U}}_{\bar{t}_f, U}^\tau$  denotes the set of all admissible controls of (10) in  $[-\Delta, \bar{t}_f]$  taking their values in  $U$ . From this

$$\tilde{\mathcal{U}}_{\mathbb{R}^m}^\tau = \bigcup_{\bar{t}_f > 0} \tilde{\mathcal{U}}_{\bar{t}_f, \mathbb{R}^m}^\tau, \quad \tilde{\mathcal{U}}_U^\tau = \bigcup_{\bar{t}_f > 0} \tilde{\mathcal{U}}_{\bar{t}_f, U}^\tau.$$

The extended end-point mapping is defined as

$$\tilde{E}_{\tau, \bar{t}_f} : \tilde{\mathcal{U}}_{\bar{t}_f, \mathbb{R}^m}^\tau \rightarrow \mathbb{R}^{n+1} : u \mapsto \tilde{x}(\bar{t}_f),$$

where  $\tilde{x}(\cdot)$  is the unique solution of problem (10), related to control  $u(\cdot) \in \tilde{\mathcal{U}}_{\bar{t}_f, \mathbb{R}^m}^\tau$ . As standard facts (see, e.g., [10]), the set  $\tilde{\mathcal{U}}_{\bar{t}_f, \mathbb{R}^m}^\tau$ , endowed with the standard topology of  $L^\infty([-\Delta, \bar{t}_f], \mathbb{R}^m)$ , is open and the end-point mapping is smooth on  $\tilde{\mathcal{U}}_{\bar{t}_f, \mathbb{R}^m}^\tau$ .

For every  $t \geq 0$ , define the extended accessible set  $\tilde{\mathcal{A}}_{\tau, U}(t)$  as the image of the extended end-point mapping  $\tilde{E}_{\tau, t}$  restricted to  $\tilde{\mathcal{U}}_{t, U}^\tau$ , where  $\tilde{\mathcal{A}}_{\tau, U}(0) = \{(\phi^1(0), 0)\}$ . The next fact is at the basis of the proof of the maximum principle (see, e.g., [1]).

**LEMMA 3.1.** *For every optimal solution  $(x_\tau(\cdot), u_\tau(\cdot))$  of  $(\text{OCP})_\tau$  defined in the interval  $[-\Delta, t_f^\tau]$ , the point  $\tilde{x}_\tau(t_f^\tau)$  belongs to the boundary of the set  $\tilde{\mathcal{A}}_{\tau, U}(t_f^\tau)$ .*

**3.1.2. Needle-like variations and Pontryagin cones.** In what follows we consider  $(\text{OCP})_\tau$  with free final time, remarking that all results can be adapted for problems with fixed final time (see, e.g., [1]). Moreover, we suppose that the optimal final time  $t_f^\tau$  is a Lebesgue point for the optimal control  $u_\tau(\cdot)$  of  $(\text{OCP})_\tau$  and of  $u_\tau(\cdot - \tau_2)$ . Otherwise, we can extend all the conclusions that follow by using closure points near  $t_f^\tau$  (as in [27, pp. 310–314] or [17, pp. 133–134]).

For delays  $\tau = (\tau^0, \tau^1, \tau^2) \in [0, \Delta]^3$ , let  $(x_\tau(\cdot), u_\tau(\cdot))$  be a solution of  $(\text{OCP})_\tau$  and, without loss of generality, extend  $u_\tau(\cdot)$  by some constant vector of  $U$  in  $[t_f^\tau, t_f^\tau + \tau^2]$ . Let  $j \geq 1$  be an integer and consider  $0 < t_1 < \dots < t_j < t_f^\tau$  Lebesgue points, respectively, of  $u_\tau(\cdot)$ ,  $u_\tau(\cdot - \tau^2)$  and of  $u_\tau(\cdot + \tau^2)$ . Choosing  $j$  arbitrary values  $u_i \in U$ , for every  $\eta_i > 0$  such that  $-\Delta \leq t_i - \eta_i$ , the needle-like variation  $\pi = \{t_1, \dots, t_j, \eta_1, \dots, \eta_j, u_1, \dots, u_j\}$  of control  $u_\tau(\cdot)$  is defined by the modified control

$$u_\tau^\pi(t) = \begin{cases} u_i, & t \in (t_i - \eta_i, t_i], \\ u_\tau(t) & \text{otherwise.} \end{cases}$$

Control  $u_\tau^\pi(\cdot)$  takes its values in  $U$  and, by continuity w.r.t. initial data, whenever  $\|(\eta_1, \dots, \eta_j)\| \rightarrow 0$ , the trajectory  $\tilde{x}_\tau^\pi(\cdot)$ , solution of the dynamics of (10) related to control  $u_\tau^\pi(\cdot)$ , converges uniformly to  $\tilde{x}_\tau(\cdot) = (x_\tau^0(\cdot), x_\tau(\cdot))$ . For every value  $z \in U$  and appropriate Lebesgue point  $s \in (0, t_f^\tau)$ , we define the vectors

$$(11) \quad \omega_z^-(s) = \tilde{f}(s, s - \tau^0, x_\tau(s), x_\tau(s - \tau^1), z, u_\tau(s - \tau^2)), \\ - \tilde{f}(s, s - \tau^0, x_\tau(s), x_\tau(s - \tau^1), u_\tau(s), u_\tau(s - \tau^2)),$$

$$(12) \quad \omega_z^+(s) = \tilde{f}(s + \tau^2, s + \tau^2 - \tau^0, x_\tau(s + \tau^2), x_\tau(s + \tau^2 - \tau^1), u_\tau(s + \tau^2), z) \\ - \tilde{f}(s + \tau^2, s + \tau^2 - \tau^0, x_\tau(s + \tau^2), x_\tau(s + \tau^2 - \tau^1), u_\tau(s + \tau^2), u_\tau(s))$$

and, given  $\xi \in \mathbb{R}^{n+1}$ , we denote by  $\tilde{v}_{s,\xi}^\tau(\cdot)$  the solution of the following linear system:

$$(13) \quad \begin{cases} \dot{\psi}(t) = \frac{\partial \tilde{f}}{\partial x}(t, t - \tau^0, x_\tau(t), x_\tau(t - \tau^1), u_\tau(t), u_\tau(t - \tau^2))\psi(t) \\ \quad + \frac{\partial \tilde{f}}{\partial y}(t, t - \tau^0, x_\tau(t), x_\tau(t - \tau^1), u_\tau(t), u_\tau(t - \tau^2))\psi(t - \tau^1), \\ \psi(s) = \xi, \quad \psi(t) = 0, \quad t \in (s - \tau^1, s). \end{cases}$$

Functions  $\tilde{v}_{s,\xi}^\tau : \mathbb{R} \rightarrow \mathbb{R}^{n+1}$  are usually called variations vectors. In what follows, we denote  $\tilde{w}_{s,z}^\tau(t) = \tilde{v}_{s,\omega_z^-(s)}^\tau(t) + \tilde{v}_{s+\tau^2,\omega_z^+(s)}^\tau(t)$ .

**DEFINITION 3.2.** For every  $t \in (0, t_f^\tau]$ , the Pontryagin cone  $\tilde{K}^\tau(t) \subseteq \mathbb{R}^{n+1}$  at  $\tilde{x}_\tau(t)$  for the extended system is defined as the smallest closed convex cone containing vectors  $\tilde{w}_{s,z}^\tau(t)$ , where  $z \in U$  and  $0 < s < t$  is a Lebesgue point of  $u_\tau(\cdot)$ ,  $u_\tau(\cdot - \tau^2)$  and of  $u_\tau(\cdot + \tau^2)$ . The augmented Pontryagin cone  $\tilde{K}_1^\tau(t) \subseteq \mathbb{R}^{n+1}$  at  $\tilde{x}_\tau(t)$  for the extended system is defined as the smallest closed convex cone containing  $\tilde{f}(t, t - \tau^0, x_\tau(t), x_\tau(t - \tau^1), u_\tau(t), u_\tau(t - \tau^2))$ ,  $-\tilde{f}(t, t - \tau^0, x_\tau(t), x_\tau(t - \tau^1), u_\tau(t), u_\tau(t - \tau^2))$  and vectors  $\tilde{w}_{s,z}^\tau(t)$ , where  $z \in U$  and  $0 < s < t$  is a Lebesgue point of  $u_\tau(\cdot)$ ,  $u_\tau(\cdot - \tau^2)$  and of  $u_\tau(\cdot + \tau^2)$ . The Pontryagin cone  $K^\tau(t) \subseteq \mathbb{R}^n$  and the augmented Pontryagin cone  $K_1^\tau(t) \subseteq \mathbb{R}^n$  at  $x_\tau(t)$  for the nonextended system are defined similarly, considering dynamics  $f$  instead of the extended dynamics  $\tilde{f}$ . Obviously, the cones  $K^\tau(t)$ ,  $K_1^\tau(t)$  are the projections onto  $\mathbb{R}^n$  of  $\tilde{K}^\tau(t)$ ,  $\tilde{K}_1^\tau(t)$ , respectively.

**Remark 3.3.** In the case of optimal control problems without delays, that is,  $(\text{OCP})_{\tau=0}$ , the definition of Pontryagin cones is slightly different from the one obtained from Definition 3.2 with the substitution  $\tau = 0$ . Indeed, considering  $\tilde{K}^0(t)$ , vectors  $\tilde{w}_{s,z}^\tau(t)$  are rather replaced by single variations  $\tilde{v}_{s,\omega_z^0(s)}^0(t)$  for which

$$\omega_z(s) = \tilde{f}(s, s, x_0(s), x_0(s), z, z) - \tilde{f}(s, s, x_0(s), x_0(s), u_0(s), u_0(s)),$$

where  $(x_0(\cdot), u_0(\cdot))$  is an optimal solution for  $(\text{OCP})_{\tau=0}$  (see, e.g., [32]).

LEMMA 3.4 (needle-like variation formula with delays). *Let  $(\delta, \eta_1, \dots, \eta_j) \in \mathbb{R} \times \mathbb{R}_+^j$  small enough. For  $t_j < t \leq t_f^\tau$  Lebesgue point of  $u_\tau(\cdot), u_\tau(\cdot - \tau^2)$ , we have*

$$\begin{aligned} \tilde{x}_\tau^\pi(t + \delta) &= \tilde{x}_\tau(t) + \delta \tilde{f}(t, t - \tau^0, x_\tau(t), x_\tau(t - \tau^1), u_\tau(t), u_\tau(t - \tau^2)) \\ &\quad + \sum_{i=1}^j \eta_i \left( \tilde{v}_{t_i, \omega_{u_i}^-(t_i)}^\tau(t) + \tilde{v}_{t_i + \tau^2, \omega_{u_i}^+(t_i)}^\tau(t) \right) + o\left(\delta + \sum_{i=1}^j \eta_i\right). \end{aligned}$$

The proof of this lemma is technical (but not difficult). It is done in Appendix A.

**3.1.3. Proof of the maximum principle.** One way to prove the maximum principle is by contradiction via the following classical result (see, e.g., [1, 14, 11]).

LEMMA 3.5 (conic implicit function theorem). *Let  $C \subseteq \mathbb{R}^m$  be convex with nonempty interior, of vertex 0, and  $F : C \rightarrow \mathbb{R}^n$  be Lipschitz such that  $F(0) = 0$  and is Gâteaux differentiable at 0 along admissible directions of  $C$ , i.e., there exists a linear mapping  $dF(0) : \mathbb{R}^m \rightarrow \mathbb{R}^n$  such that, for every  $x \in C$ ,*

$$(14) \quad \frac{F(\alpha x)}{\alpha} \xrightarrow{\alpha \rightarrow 0^+} dF(0)x .$$

*If  $dF(0) \cdot \text{Cone}(C) = \mathbb{R}^n$ , where  $\text{Cone}(C)$  stands for the (convex) cone generated by elements of  $C$ , then  $0 \in \text{Int } F(\mathcal{V} \cap C)$  for every neighborhood  $\mathcal{V}$  of 0 in  $\mathbb{R}^m$ .*

Consider any integer  $j \geq 1$  and a positive real number  $\varepsilon_j > 0$ . Define

$$G_j^\tau : B_{\varepsilon_j}(0) \cap \mathbb{R} \times \mathbb{R}_+^j \rightarrow \mathbb{R}^{n+1} : (\delta, \eta_1, \dots, \eta_j) \mapsto \tilde{x}_\tau^\pi(t_f^\tau + \delta) - \tilde{x}_\tau(t_f^\tau) ,$$

where  $\pi$  is any variation of control  $u_\tau(\cdot)$  and  $\varepsilon_j$  is small enough such that  $G_j^\tau$  is well-defined (see section 3.1.2). The following statements hold:

- $G_j^\tau(0) = 0$  and  $G_j^\tau$  is Lipschitz continuous.
- $G_j^\tau$  is Gâteaux differentiable at 0 along admissible directions of the convex set  $B_{\varepsilon_j}(0) \cap \mathbb{R} \times \mathbb{R}_+^j$  (in the sense of (14)) thanks to Lemma 3.4.

The Lipschitz behavior of  $G_j^\tau$  is proved by a recursive use of needle-like variations at  $t_i - \eta_i, 1 \leq i \leq j$  (for  $\eta_i$  small enough), Lebesgue points of  $u_\tau(\cdot)$ , by making a recursive use of Lemma 3.4. Remark that, since  $t_i - \eta_i$  are Lebesgue points of  $u_\tau(\cdot)$  only for almost every  $\eta_i$ , the recursive use of Lemma 3.4 can be done only almost everywhere. The conclusion follows from the continuity of  $G_j^\tau$  and density arguments.

The maximum principle is established as follows. Suppose, by contradiction, that the cone  $\tilde{K}_1^\tau(t_f^\tau)$  coincides with  $\mathbb{R}^{n+1}$ . Then, by definition, there would exist an integer  $j \geq 1$ , a variation  $\pi$  of  $u_\tau(\cdot)$ , and a positive real number  $\varepsilon_j > 0$  such that

$$dG_j^\tau(0) \cdot (\mathbb{R} \times \mathbb{R}_+^j) = \tilde{K}_1^\tau(t_f^\tau) = \mathbb{R}^{n+1} .$$

In this case, Lemma 3.5 would imply that the point  $\tilde{x}_\tau(t_f^\tau)$  belongs to the interior of the accessible set  $\tilde{\mathcal{A}}_{\tau,U}(t_f^\tau)$ , which contradicts Lemma 3.1. Therefore the following holds.

LEMMA 3.6. *There exists  $\tilde{\psi}_\tau \in \mathbb{R}^{n+1} \setminus \{0\}$  (Lagrange multiplier) such that*

$$\begin{aligned} \langle \tilde{\psi}_\tau, \tilde{f}(t_f^\tau, t_f^\tau - \tau^0, x_\tau(t_f^\tau), x_\tau(t_f^\tau - \tau^1), u_\tau(t_f^\tau), u_\tau(t_f^\tau - \tau^2)) \rangle &= 0, \\ \langle \tilde{\psi}_\tau, \tilde{v}_\tau \rangle &\leq 0 \quad \forall \tilde{v}_\tau \in \tilde{K}^\tau(t_f^\tau). \end{aligned}$$

The relations provided by Lemma 3.6 allow us to derive the necessary conditions (4)–(7) given in section 2.1. (We skip these computations, referring to [1, 11, 23] for details.) The relation between the adjoint vector satisfying (4) and the above Lagrange multiplier  $\tilde{\psi}_\tau = (\psi_\tau, \psi_\tau^0)$  is that  $(p_\tau(\cdot), p_\tau^0)$  is built so that  $p_\tau(t_f^\tau) = \psi_\tau$ ,  $p_\tau^0 = \psi_\tau^0$ . We will make use of the result below, which follows from the previous considerations.

LEMMA 3.7. *Consider the free final time problem  $(OCP) = (OCP)_{\tau=0}$ . For any optimal trajectory  $x(\cdot)$  of  $(OCP)$ , the following statements are equivalent:*

- *The trajectory  $x(\cdot)$  has an unique extremal lift  $(x(\cdot), p(\cdot), p^0, u(\cdot))$  whose adjoint  $(p(\cdot), p^0)$  is unique up to a multiplicative scalar, which is normal.*
- *The cone  $K_1^{\tau=0}(t_f)$  is a half-space of  $\mathbb{R}^{n+1}$  and  $K_1^{\tau=0}(t_f) = \mathbb{R}^n$ .*

In particular, Lemma 3.7 requires the uniqueness of merely signed multipliers for  $(OCP)_{\tau=0}$ , i.e.,  $p^0 \leq 0$ . To prove Lemma 3.7 under this assumption, the argument above just needs slight changes, which consist of defining the function  $G_j^\tau$  by adding a slack positive control variable to the last component of dynamics (10) (see, e.g., [1]).

**3.2. Conic implicit function theorem with parameters.** The first step of the proof of Theorem 2.1 makes use of the procedure detailed in section 3.1. More specifically, we need the needle-like variation formula and the conic implicit function theorem. However, Lemma 3.5 is not suited to this situation because we have to take into account the dependence w.r.t. delays  $\tau$ . Indeed, the proof of Lemma 3.5 is based on a Brouwer fixed point theorem (see, e.g., [1]) which does not consider continuous dependence w.r.t. parameters (which, in our case, is represented by  $\tau$ ). Therefore, in this section, we introduce a more general version of the conic implicit function theorem depending on parameters.

When considering delays  $\tau$  as a varying parameter, the variation formula provided by Lemma 3.4 holds only for almost every  $\tau$ , and this, because we need that each  $t_i$  be a Lebesgue point of  $u_\tau(\cdot)$ ,  $u_\tau(\cdot - \tau)$  and of  $u_\tau(\cdot + \tau)$ . This leads us to introduce a notion of conic implicit function theorem which, on one hand, ensures a continuous dependence w.r.t. parameters and, on the other hand, deals with quantities defined uniquely on dense subsets. The notion of differentiability that we need is the following. A function  $f : C \subseteq \mathbb{R}^j \rightarrow \mathbb{R}^n$  is said almost everywhere strictly differentiable at some point  $x_0 \in C$  whenever there exists a linear continuous mapping  $df(x_0) : \mathbb{R}^j \rightarrow \mathbb{R}^n$  such that

$$f(y) - f(x) = df(x_0) \cdot (y - x) + \|y - x\|g(x, y)$$

for almost every  $x, y \in C$ , where  $g(x, y)$  tends to 0 as soon as  $\|x - x_0\| + \|y - x_0\| \xrightarrow{a.e.} 0$ .

One may remark that the notion of strict differentiability and of conic implicit function theorem depending on parameters has already been introduced by [3]. In our framework, we adapt these results to dense subsets.

LEMMA 3.8 (conic implicit function theorem with parameters). *Let  $C \subseteq \mathbb{R}^j$  be open and convex with a nonempty interior, of vertex 0, and  $F : \mathbb{R}_+^k \times C \rightarrow \mathbb{R}^n : (\varepsilon, x) \mapsto F(\varepsilon, x)$  be a continuous mapping, for which  $F(0, 0) = 0$ , satisfying the following:*

- *For almost every  $\varepsilon \in \mathbb{R}_+^k$ ,  $F$  is almost everywhere strictly differentiable w.r.t.  $x$  at 0, and  $\frac{\partial F}{\partial x}(\varepsilon, 0)$  is continuous in  $\varepsilon$  on a dense subset.*
- *For almost every  $\varepsilon \in \mathbb{R}_+^k$ , the remainder satisfies  $g_\varepsilon(x, y) \rightarrow 0$  as  $(x, y) \xrightarrow{a.e.} 0$ , uniformly w.r.t.  $\varepsilon$  on a dense subset.*
- *There holds  $\frac{\partial F}{\partial x}(0, 0) \cdot \text{Cone}(C) = \mathbb{R}^n$ .*

Therefore, there exist  $\varepsilon_0 > 0$ , a neighborhood  $\mathcal{V}$  of 0 in  $\mathbb{R}^n$ , and a continuous function  $h : [0, \varepsilon_0]^k \times \mathcal{V} \rightarrow C$ , such that  $F(\varepsilon, h(\varepsilon, y)) = y$  for every  $\varepsilon \in [0, \varepsilon_0]^k$  and every  $y \in \mathcal{V}$ .

The proof of Lemma 3.8 is done in Appendix B.

**3.3. Proof of Theorem 2.1.** From now on, assume that assumptions (A) hold. Moreover,  $(x(\cdot), u(\cdot))$  will denote the (unique) solution of (OCP) and we assume that its related final time  $t_f$  is a Lebesgue point of  $u(\cdot)$  (if not, as pointed out in section 3.1.2, we refer to the approach proposed by [27, 17]). Finally, without loss of generality, we consider free final time problems (otherwise, the proof is similar, but simpler), introducing further assumption  $(C_2)$  for  $(\text{OCP})_\tau$  with control delays.

**3.3.1. Controllability for  $(\text{OCP})_\tau$ .** For any integer  $j \geq 1$ , fix  $0 < t_1 < \dots < t_j < t_f$  Lebesgue points of control  $u(\cdot)$  and  $j$  arbitrary values  $u_i \in U$ . We denote  $v|_n$  the first  $n$  coordinates of a vector  $v \in \mathbb{R}^{n+1}$ . For an appropriate small positive real number  $\varepsilon_j > 0$ , denoting by  $\tilde{x}_{(\varepsilon^0, \varepsilon^1, \varepsilon^2)}(\cdot)$  the trajectory solution of (10) with delay  $\tau = (\varepsilon^0, \varepsilon^1, \varepsilon^2)$  and control  $u_{(\varepsilon^0, \varepsilon^1, \varepsilon^2)}(\cdot)$ , we define the mapping

$$\Gamma : B_{\varepsilon_j}(0) \cap (\mathbb{R}_+^3 \times \mathbb{R} \times \mathbb{R}_+^j) \rightarrow \mathbb{R}^n : (\varepsilon^0, \varepsilon^1, \varepsilon^2, \delta, \eta_1, \dots, \eta_j) \mapsto (\tilde{x}_{(\varepsilon^0, \varepsilon^1, \varepsilon^2)}^\pi(t_f + \delta) - \tilde{x}(t_f))|_n,$$

which, thanks to assumption  $(A_2)$  and by continuity w.r.t. initial data, is well-defined and continuous. Moreover,  $\Gamma(0, \dots, 0) = 0$  and

$$\Gamma(\varepsilon^0, \dots, \eta_j) = (\tilde{x}_{(\varepsilon^0, \varepsilon^1, \varepsilon^2)}^\pi(t_f + \delta) - \tilde{x}_{(\varepsilon^0, \varepsilon^1, \varepsilon^2)}(t_f))|_n + (\tilde{x}_{(\varepsilon^0, \varepsilon^1, \varepsilon^2)}(t_f) - \tilde{x}(t_f))|_n.$$

From Lemma 3.4 and a recursive use of the needle-like variation formula (see section 3.1.3), for almost every  $(\varepsilon^0, \varepsilon^1, \varepsilon^2)$  small enough,  $\Gamma$  is almost everywhere strictly differentiable w.r.t.  $(\delta, \eta_1, \dots, \eta_j)$  at 0,  $\frac{\partial \Gamma}{\partial (\delta, \eta_1, \dots, \eta_j)}(\varepsilon^0, \varepsilon^1, \varepsilon^2, 0)$  is continuous w.r.t.  $(\varepsilon^0, \varepsilon^1, \varepsilon^2)$  on a dense subset, and, moreover, the remainder of the related Taylor expansion converges to zero uniformly w.r.t.  $(\varepsilon^0, \varepsilon^1, \varepsilon^2)$  on a dense subset.

From assumption  $(A_3)$ , the unique extremal lift of  $x(\cdot)$  is normal, and hence, it follows from Lemma 3.7 that  $\text{Int } K_1^{\tau=0}(t_f) = \mathbb{R}^n$ . We recall that we consider only either optimal control problems with pure state delays or control-affine optimal control problems. Therefore, thanks to Remark 3.3, there exist a real number  $\delta$ , an integer  $j \geq 1$ , and a variation  $\pi = \{t_1, \dots, t_j, \eta_1, \dots, \eta_j, u_1, \dots, u_j\}$  such that

$$\frac{\partial \Gamma}{\partial (\delta, \eta_1, \dots, \eta_j)}(0) \cdot (\mathbb{R} \times \mathbb{R}_+^j) = \text{Int } K_1^{\tau=0}(t_f) = \mathbb{R}^n.$$

At this step, Lemma 3.8 implies the existence of  $\varepsilon_0 > 0$  such that, for every  $\tau = (\tau^0, \tau^1, \tau^2) \in [0, \varepsilon_0]^3$ , there exist a real  $\delta(\tau)$  and positive reals  $\eta_1(\tau), \dots, \eta_j(\tau)$  such that  $\Gamma(\tau^0, \tau^1, \tau^2, \delta(\tau), \eta_1(\tau), \dots, \eta_j(\tau)) = 0$ . Moreover,  $\delta(\tau), \eta_1(\tau), \dots, \eta_j(\tau)$  are continuous w.r.t.  $\tau$ . From assumption  $(A_4)$ , it follows that, for every  $\tau = (\tau^0, \tau^1, \tau^2) \in [0, \varepsilon_0]^3$ , the subset  $M_f$  is reachable for the dynamics of  $(\text{OCP})_\tau$ , in a final time  $t_f^\tau \in [0, b]$ , by using control  $u_{(\tau^0, \tau^1, \tau^2)}^\pi(\cdot) \in L^\infty([0, t_f^\tau], U)$ .

We have proved that, for every  $\tau = (\tau^0, \tau^1, \tau^2) \in (0, \varepsilon_0)^3$ ,  $(\text{OCP})_\tau$  is controllable. Remark that this argument still holds for  $(\text{OCP})_\tau$  with pure state delays.

**3.3.2. Existence of optimal controls for  $(\text{OCP})_\tau$ .** We focus first on the existence of an optimal control for  $(\text{OCP})_\tau$  for every  $\tau \in (0, \varepsilon_0)^3$ . No other assumptions but (A) and  $(C_1)$  are considered. In particular, mappings  $f$  and  $f^0$  are affine in the two control variables. Thanks to this property, existence can be established by using the arguments in [8, Theorem 2]. However, we prefer to develop the usual Filippov's

scheme [16] (following [27, 39]) to highlight the difficulty in applying this procedure to more general systems (in particular, see Remark 3.9). Even if problems  $(\text{OCP})_\tau$  with control and state delays are considered, we assume to have free final time just to use the same approach for a problem with pure state delays.

Fix  $\tau = (\tau^0, \tau^1, \tau^2) \in (0, \varepsilon_0)^3$  and let

$$\alpha = \inf_{u \in \mathcal{U}_\tau^r} C_\tau(t_f(u), u) = \int_0^{t_f(u)} f^0(t, t - \tau^0, x(t), x(t - \tau^1), u(t), u(t - \tau^2)) \, dt.$$

Consider now a minimizing sequence of trajectories  $x_k(\cdot)$  associated to  $u_k(\cdot)$ , that is,  $C_\tau(t_f(u_k), u_k) \rightarrow \alpha$  when  $k \rightarrow \infty$  and define

$$\tilde{F}_k(t) = \tilde{f}(t, t - \tau^0, x_k(t), x_k(t - \tau^1), u_k(t), u_k(t - \tau^2))$$

for almost every  $t \in [0, t_f(u_k)]$ . By assumption  $(A_4)$ , we can extend  $\tilde{F}_k(\cdot)$  by zero on  $(t_f(u_k), b]$  so that  $(\tilde{F}_k(\cdot))_{k \in \mathbb{N}}$  is bounded in  $L^\infty([0, b], \mathbb{R}^{n+1})$ . Therefore, up to some subsequence,  $(\tilde{F}_k(\cdot))_{k \in \mathbb{N}}$  converges to some  $\tilde{F}(\cdot) = (F(\cdot), F^0(\cdot)) \in L^\infty([0, b], \mathbb{R}^{n+1})$  for the weak star topology of  $L^\infty$ . On the other hand, up to some subsequence, the sequence  $(t_f(u_k))_{k \in \mathbb{N}}$  converges to some  $t_f^r \geq 0$ . Then, for every  $t \in [-\Delta, t_f^r]$ , define

$$(15) \quad x_\tau(t) = \phi^1(t) \mathbb{1}_{[-\Delta, 0)}(t) + \mathbb{1}_{[0, t_f^r]}(t) \left( \phi^1(0) + \int_0^t F(s) \, ds \right).$$

Now,  $x_\tau(\cdot)$  is absolutely continuous and, considering continuous extensions,  $(x_k(\cdot))_{k \in \mathbb{N}}$  converges pointwise to  $x_\tau(\cdot)$  within  $[-\Delta, t_f^r]$ . Moreover, by assumptions  $(A_1)$ ,  $(A_4)$  and the Arzelà–Ascoli theorem, up to some subsequence,  $(x_k(\cdot))_{k \in \mathbb{N}}$  converges to  $x_\tau(\cdot)$ , uniformly in  $[-\Delta, t_f^r]$ . From the compactness of  $M_f$ , we have  $x_\tau(t_f^r) \in M_f$ .

In the next paragraph, we show that  $x_\tau(\cdot)$  comes from a control in  $\mathcal{U}_{t_f^r, U}^\tau$ .

For almost every  $t \in [0, t_f(u_k)]$ , set

$$\tilde{H}_k(t) = \tilde{f}(t, t - \tau^0, x_\tau(t), x_\tau(t - \tau^1), u_k(t), u_k(t - \tau^2))$$

and, if  $t_f(u_k) + \tau^2 < t_f^r$ , extend it by 0 in  $(t_f(u_k), b]$ . At this step, we need to introduce several structures to deal with the presence of the control delay  $\tau^2$ . First, let

$$\beta = \max \left\{ |f^0(t, s, x, y, u, v)| : -\Delta \leq t, s \leq b, \|(x, y)\| \leq b, (u, v) \in U^2 \right\} > 0$$

and  $N \in \mathbb{N}$  such that  $N\tau^2 \leq t_f^r < (N + 1)\tau^2$ . Considering continuous extensions, we see that  $x_\tau(\cdot)$  is well-defined in  $[-\Delta, (N + 1)\tau^2]$ . We define

$$(16) \quad \tilde{G}(t, u^1, \dots, u^{N+1}, \gamma^1, \dots, \gamma^{N+1}) = \begin{pmatrix} f(t, t - \tau^0, x_\tau(t), x_\tau(t - \tau^1), u^1, \phi^2(t - \tau^2)) \\ f^0(t, t - \tau^0, x_\tau(t), x_\tau(t - \tau^1), u^1, \phi^2(t - \tau^2)) + \gamma^1 \\ f(t + \tau^2, t + \tau^2 - \tau^0, x_\tau(t + \tau^2), x_\tau(t + \tau^2 - \tau^1), u^2, u^1) \\ f^0(t + \tau^2, t + \tau^2 - \tau^0, x_\tau(t + \tau^2), x_\tau(t + \tau^2 - \tau^1), u^2, u^1) + \gamma^2 \\ \dots \\ f(t + N\tau^2, t + N\tau^2 - \tau^0, x_\tau(t + N\tau^2), x_\tau(t + N\tau^2 - \tau^1), u^{N+1}, u^N) \\ f^0(t + N\tau^2, t + N\tau^2 - \tau^0, x_\tau(t + N\tau^2), x_\tau(t + N\tau^2 - \tau^1), u^{N+1}, u^N) + \gamma^{N+1} \end{pmatrix}$$

almost everywhere in  $[0, \tau^2]$ , and

$$\begin{aligned} &\tilde{V}_\beta(t) \\ &= \left\{ \tilde{G}(t, u^1, \dots, u^{N+1}, \gamma^1, \dots, \gamma^{N+1}) : (u^1, \dots, u^{N+1}) \in U^{N+1}, \forall i = 1, \dots, N+1 : \gamma^i \geq 0, \right. \\ &\quad \left. |f^0(t, t - \tau^0, x_\tau(t), x_\tau(t - \tau^1), u^1, \phi^2(t - \tau^2)) + \gamma^1| \leq \beta \right. \\ &\quad \left. \forall i = 1, \dots, N : |f^0(t + i\tau^2, t + i\tau^2 - \tau^0, x_\tau(t + i\tau^2), x_\tau(t + i\tau^2 - \tau^1), u^{i+1}, u^i) + \gamma^{i+1}| \leq \beta \right\}. \end{aligned}$$

Thanks to assumption  $(A_1)$ ,  $\tilde{V}_\beta(t)$  is compact for the standard topology of  $\mathbb{R}^{(n+1)(N+1)}$ . Moreover, assumptions  $(A_1)$  and  $(C_1)$  ensure that  $\tilde{V}_\beta(t)$  is convex. We have that

$$\tilde{\mathcal{V}} = \left\{ \tilde{G}(\cdot) \in L^2([0, \tau^2], \mathbb{R}^{(n+1)(N+1)}) : \tilde{G}(t) \in \tilde{V}_\beta(t), \text{ a.e. } [0, \tau^2] \right\}$$

is convex and closed in  $L^2([0, \tau^2], \mathbb{R}^{(n+1)(N+1)})$  for the strong topology of  $L^2$ , and therefore, it is convex and closed in  $L^2([0, \tau^2], \mathbb{R}^{(n+1)(N+1)})$  for the weak topology of  $L^2$ . At this step, for every  $i = 0, \dots, N$ , denote

$$\tilde{G}_k^{i+1}(t) = \tilde{f}(t + i\tau^2, t + i\tau^2 - \tau^0, x_\tau(t + i\tau^2), x_\tau(t + i\tau^2 - \tau^1), u_k(t + i\tau^2), u_k(t + (i-1)\tau^2))$$

and  $\tilde{G}_k(t) = (\tilde{G}_k^1(t), \dots, \tilde{G}_k^{N+1}(t))$ . Therefore,  $\tilde{G}_k(\cdot) \in \tilde{\mathcal{V}}$  for every  $k \in \mathbb{N}$ . Moreover, since  $(\tilde{G}_k(\cdot))_{k \in \mathbb{N}}$  is bounded in  $L^2([0, \tau^2], \mathbb{R}^{(n+1)(N+1)})$ , up to some subsequence, it converges for the weak topology of  $L^2$  to a function  $\tilde{G}(\cdot)$  that necessarily belongs to  $\tilde{\mathcal{V}}$ . Therefore, for almost every  $t \in [0, \tau^2]$  and  $i = 1, \dots, N+1$ , there exist points  $u_\tau^i(t) \in U$  and scalar  $\gamma_\tau^i(t) \geq 0$  such that

$$\tilde{G}^1(t) = \begin{pmatrix} f(t, t - \tau^0, x_\tau(t), x_\tau(t - \tau^1), u_\tau^1(t), \phi^2(t - \tau^2)) \\ f^0(t, t - \tau^0, x_\tau(t), x_\tau(t - \tau^1), u_\tau^1(t), \phi^2(t - \tau^2)) + \gamma_\tau^1(t) \end{pmatrix}$$

and, for every  $i = 1, \dots, N$ ,

$$\begin{aligned} &\tilde{G}^{i+1}(t) \\ &= \begin{pmatrix} f(t + i\tau^2, t + i\tau^2 - \tau^0, x_\tau(t + i\tau^2), x_\tau(t + i\tau^2 - \tau^1), u_\tau^{i+1}(t), u_\tau^i(t)) \\ f^0(t + i\tau^2, t + i\tau^2 - \tau^0, x_\tau(t + i\tau^2), x_\tau(t + i\tau^2 - \tau^1), u_\tau^{i+1}(t), u_\tau^i(t)) + \gamma_\tau^{i+1}(t) \end{pmatrix}. \end{aligned}$$

Moreover, since  $U$  is compact, functions  $u_\tau^i(\cdot)$ ,  $\gamma_\tau^i(\cdot)$  can be chosen to be measurable on  $[0, \tau^2]$  using a measurable selection lemma (see, e.g., [27, Lemma 3A, p. 161]).

At this step, we come back to the whole interval  $[-\tau^2, t_f^*]$ . For this, set

$$\begin{aligned} u_\tau(t) &= \begin{cases} \phi^2(t) & t \in [-\tau^2, 0], \\ u_\tau^{i+1}(t - i\tau^2) & t \in [i\tau^2, (i+1)\tau^2], \quad i = 0, \dots, N, \end{cases} \\ \gamma_\tau(t) &= \gamma_\tau^{i+1}(t - i\tau^2) \quad t \in [i\tau^2, (i+1)\tau^2], \quad i = 0, \dots, N, \end{aligned}$$

which are measurable functions in  $[-\tau^2, t_f^*]$ , and let

$$\tilde{H}(t) = \begin{pmatrix} f(t, t - \tau^0, x_\tau(t), x_\tau(t - \tau^1), u_\tau(t), u_\tau(t - \tau^2)) \\ f^0(t, t - \tau^0, x_\tau(t), x_\tau(t - \tau^1), u_\tau(t), u_\tau(t - \tau^2)) + \gamma_\tau(t) \end{pmatrix}.$$

From the weak star convergence in  $L^\infty$  of  $(\tilde{G}_k(\cdot))_{k \in \mathbb{N}}$  toward  $\tilde{G}(\cdot)$ , it follows immediately that  $(\tilde{H}_k(\cdot))_{k \in \mathbb{N}}$  converges to  $\tilde{H}(\cdot)$  for the weak topology of  $L^2$ . Furthermore, from the differentiability of  $\tilde{f}$  w.r.t.  $(x, y)$ , the compactness of  $U$ , and the dominated

convergence theorem, one has

$$\lim_{k \rightarrow \infty} \int_0^{t_f^\tau} (\tilde{F}_k(t) - \tilde{H}_k(t)) \cdot \varphi(t) dt = 0$$

for every map  $\varphi(\cdot) \in L^2([0, t_f^\tau], \mathbb{R}^{n+1})$ , from which  $\tilde{H} = \tilde{F}$  almost everywhere in  $[0, t_f^\tau]$ .

Combining (15) with all the previous results, we obtain

$$x_\tau(t) = \phi^1(t)\mathbb{1}_{[-\Delta, 0)}(t) + \mathbb{1}_{[0, t_f^\tau]}(t) \left( \phi^1(0) + \int_0^t f(t, t - \tau^0, x_\tau(s), x_\tau(t - \tau^1), u_\tau(s), u_\tau(t - \tau^2)) ds \right),$$

which proves that the measurable function  $u_\tau(\cdot)$  is an admissible control for  $(\text{OCP})_\tau$ .

It remains to show that control  $u_\tau(\cdot)$  is optimal for  $(\text{OCP})_\tau$ . For this, from what we showed above and by definition of weak star convergence, we have

$$C_\tau(t_f(u_k), u_k) \rightarrow \int_0^{t_f^\tau} (f^0(t, t - \tau^0, x_\tau(t), x_\tau(t - \tau^1), u_\tau(t), u_\tau(t - \tau^2)) + \gamma_\tau(t)) dt.$$

Since  $\gamma_\tau(\cdot)$  takes only nonnegative values, one finally has

$$\int_0^{t_f^\tau} f^0(t, t - \tau^0, x_\tau(t), x_\tau(t - \tau^1), u_\tau(t), u_\tau(t - \tau^2)) dt \leq \alpha \leq C_\tau(t_f(v), v)$$

for every  $v(\cdot) \in \mathcal{U}_U^\tau$ . Therefore,  $\gamma_\tau(\cdot)$  is necessarily zero and the conclusion follows.

Now, we consider problems  $(\text{OCP})_\tau$  with pure state delays. It is clear that, if assumption  $(B_2)$  holds, one can proceed with the same argument as above (which is nothing else but the usual Filippov’s scheme; see, e.g., [16]) for the existence of optimal controls. In this case, Guinn’s reduction (16) is not needed.

*Remark 3.9.* Guinn’s reduction (16) converts the dynamics with control delays into a new dynamics without control delays but with a larger number of variables. It is clear from the context that the natural assumption to provide the existence of optimal controls for generic nonlinear dynamics is the convexity of system (16) for every  $N \in \mathbb{N}$  (since we make delays vary), which is a very strong assumption. The proof of [31, Lemma 2.1] does not work under the weaker assumption of convexity of the epigraph of the extended dynamics.

**3.3.3. Convergence of optimal controls and trajectories for  $(\text{OCP})_\tau$ .**

We start by considering problems  $(\text{OCP})_\tau$  with pure state delays, by assuming that assumption  $(B_2)$  holds. In this case, the classical way to proceed consists in reproducing and adapting the convexity Filippov’s scheme used in the previous section for the existence of optimal controls (see, e.g., [22]).

Let  $(\tau_k)_{k \in \mathbb{N}} = ((\tau_k^0, \tau_k^1, 0))_{k \in \mathbb{N}} \subseteq (0, \varepsilon_0)^2 \times \{0\}$  be an arbitrary sequence converging to 0 as  $k$  tends to  $\infty$  and let  $(x_{\tau_k}(\cdot), u_{\tau_k}(\cdot))$  be an optimal solution for  $(\text{OCP})_{\tau_k}$  with final time  $t_f^{\tau_k}(u_{\tau_k})$ . Since  $t_f^{\tau_k}(u_{\tau_k}) \in [0, b]$ , up to some subsequence, the sequence  $(t_f^{\tau_k})_{k \in \mathbb{N}} = (t_f^{\tau_k}(u_{\tau_k}))_{k \in \mathbb{N}}$  converges to some  $\bar{t}_f \in [0, b]$ . Since  $M_f$  is compact, up to some subsequence, the sequence  $(x_{\tau_k}(t_f^{\tau_k}))_{k \in \mathbb{N}} \subseteq M_f$  converges to some point in  $M_f$ .

For every integer  $k$  and almost every  $t \in [0, t_f^{\tau_k}]$ , set

$$\tilde{G}_k(t) = \left( \tilde{f}(t, t - \tau_k^0, x_{\tau_k}(t), x_{\tau_k}(t - \tau_k^1), u_{\tau_k}(t)), \frac{\partial \tilde{f}}{\partial x}(t, t - \tau_k^0, x_{\tau_k}(t), x_{\tau_k}(t - \tau_k^1), u_{\tau_k}(t)), \frac{\partial \tilde{f}}{\partial y}(t, t - \tau_k^0, x_{\tau_k}(t), x_{\tau_k}(t - \tau_k^1), u_{\tau_k}(t)) \right).$$



Thanks to assumption  $(A_4)$ , we extend  $\tilde{G}_k(\cdot)$  by zero on  $(t_f^{\tau_k}, b]$ . Assumptions  $(A_1)$  and  $(A_4)$  imply that the sequence  $(\tilde{G}_k(\cdot))_{k \in \mathbb{N}}$  is bounded in  $L^\infty$ , and then, up to some subsequence, it converges to some  $\tilde{G}(\cdot) = (G(\cdot), G^0(\cdot), G_x(\cdot), G_y(\cdot)) \in L^\infty([0, b], \mathbb{R}^{n+1})$  for the weak star topology of  $L^\infty$ . Exploiting the weak star convergence of  $L^\infty$  (and using  $\mathbb{1}_{[\bar{t}_f, b]} \tilde{G}$  as test function), we get that  $\tilde{G}(t) = 0$  for almost every  $t \in [\bar{t}_f, b]$ . From this, for every  $t \in [0, \bar{t}_f]$ , denote

$$(17) \quad \bar{x}(t) = \phi^1(t) \mathbb{1}_{[-\Delta, 0)}(t) + \mathbb{1}_{[0, \bar{t}_f]}(t) \left( \phi^1(0) + \int_0^t G(s) ds \right).$$

Clearly,  $\bar{x}(\cdot)$  is absolutely continuous and  $\bar{x}(t) = \lim_{k \rightarrow \infty} x_{\tau_k}(t)$  pointwise in  $[-\Delta, \bar{t}_f]$ . By assumptions  $(A_1)$ ,  $(A_4)$  and the Arzelà–Ascoli theorem, up to some subsequence,  $(x_{\tau_k}(\cdot))_{k \in \mathbb{N}}$  converges to  $\bar{x}(\cdot)$ , uniformly in  $[-\Delta, \bar{t}_f]$ , and we have  $\bar{x}(\bar{t}_f) \in M_f$ .

In the next paragraph, we prove that there exists a control  $\bar{u}(\cdot) \in L^\infty([0, \bar{t}_f], U)$  such that  $\bar{x}(\cdot)$  is an admissible trajectory for (OCP) associated with this control  $\bar{u}(\cdot)$ .

Using the definition of  $\beta$  given in section 3.3.2, for  $t \in [0, \bar{t}_f]$ , consider the set

$$\begin{aligned} & \tilde{Z}_\beta(t) \\ = & \left\{ \left( f(t, t, \bar{x}(t), \bar{x}(t), u), f^0(t, t, \bar{x}(t), \bar{x}(t), u) + \gamma, \frac{\partial \tilde{f}}{\partial x}(t, t, \bar{x}(t), \bar{x}(t), u), \frac{\partial \tilde{f}}{\partial y}(t, t, \bar{x}(t), \bar{x}(t), u) \right) : \right. \\ & \left. u \in U, \gamma \geq 0, |f^0(t, t, \bar{x}(t), \bar{x}(t), u) + \gamma| \leq \beta \right\}. \end{aligned}$$

Thanks to assumption  $(B_2)$ , the set  $\tilde{Z}_\beta(t)$  is compact and convex in  $\mathbb{R}^{n+1}$ . We have the following statements:

- From the convexity and the compactness of  $\tilde{Z}_\beta(t)$ , for  $\delta > 0$  and  $t \in [0, \bar{t}_f]$ ,

$$\tilde{Z}_\beta^\delta(t) = \left\{ x \in \mathbb{R}^{n+1} : d(x, \tilde{Z}_\beta(t)) \leq \delta \right\}, \text{ where } d(x, A) = \inf_{y \in A} \|x - y\|$$

is convex and compact for the standard topology of  $\mathbb{R}^{n+1}$ .

- For every  $\delta > 0$ , the set

$$\tilde{Z}_\delta = \left\{ \tilde{F}(\cdot) \in L^2([0, \bar{t}_f], \mathbb{R}^{n+1}) : \tilde{F}(t) \in \tilde{Z}_\beta^\delta(t) \text{ for almost every } t \in [0, \bar{t}_f] \right\}$$

is convex and closed in  $L^2([0, \bar{t}_f], \mathbb{R}^{n+1})$  for the strong topology of  $L^2$ . Then, it is closed in  $L^2([0, \bar{t}_f], \mathbb{R}^{n+1})$  for the weak topology of  $L^2$ .

Convexity is obvious from the previous statement. Let  $(\tilde{F}_k(\cdot))_{k \in \mathbb{N}} \subseteq \tilde{Z}_\delta$  such that  $\tilde{F}_k(\cdot) \xrightarrow{L^2} \tilde{F}(\cdot)$ . Therefore,  $\tilde{F}(\cdot) \in L^2([0, \bar{t}_f], \mathbb{R}^{n+1})$  and there exists a subsequence such that  $\tilde{F}_{k_m}(\cdot) \xrightarrow{a.e.} \tilde{F}(\cdot)$ . Since  $\tilde{Z}_\beta^\delta(t)$  is closed for the standard topology of  $\mathbb{R}^{n+1}$ , a.e. in  $t \in [0, \bar{t}_f]$ , we have that  $\tilde{F}(t) = \lim_{m \rightarrow \infty} \tilde{F}_{k_m}(t) \in \tilde{Z}_\beta^\delta(t)$  and the statement follows.

- For every  $\delta > 0$ , there exists  $k_\delta \in \mathbb{N}$  such that if  $k \geq k_\delta$ , then  $\tilde{G}_k(\cdot) \in \tilde{Z}_\delta$ . Indeed, thanks to assumptions  $(A_1)$ ,  $(A_4)$ , mappings  $f, f^0$  are globally Lipschitz within  $[-\Delta, \bar{t}_f]^2 \times B_b^{2n}(0) \times U$  and, for almost every  $t \in [0, \bar{t}_f]$ ,

$$\inf_{z \in \tilde{Z}_\beta(t)} \|\tilde{G}_k(t) - z\| \leq \tilde{C} \left( \|x_{\tau_k}(t) - \bar{x}(t)\| + \|x_{\tau_k}(t - \tau_k^1) - \bar{x}(t)\| + \tau_k^0 \right),$$

where  $\tilde{C} > 0$  is a suitable constant, which is independent from  $t$ . The conclusion follows from the uniform convergence of  $(x_{\tau_k}(\cdot))_{k \in \mathbb{N}}$  toward  $\bar{x}(\cdot)$ .

Using the closeness of  $\tilde{Z}_\delta$  w.r.t. the weak topology of  $L^2$ , we infer that  $\tilde{G}(\cdot) \in \tilde{Z}_\delta$  for every  $\delta > 0$ . This implies  $\tilde{G}(\cdot) \in \bigcap_{j \in \mathbb{N}} \tilde{Z}_{1/j} \subseteq \tilde{Z}_0$ .

We have obtained that, a.e. in  $t \in [0, \bar{t}_f]$ , there exist  $\bar{u}(t) \in U$ ,  $\bar{\gamma}(t) \geq 0$  such that

$$(18) \quad \tilde{G}(t) = \left( f(t, t, \bar{x}(t), \bar{x}(t), \bar{u}(t)) , f^0(t, t, \bar{x}(t), \bar{x}(t), \bar{u}(t)) + \bar{\gamma}(t) , \right. \\ \left. \frac{\partial \tilde{f}}{\partial x}(t, t, \bar{x}(t), \bar{x}(t), \bar{u}(t)) , \frac{\partial \tilde{f}}{\partial y}(t, t, \bar{x}(t), \bar{x}(t), \bar{u}(t)) \right).$$

Moreover, since  $U$  is compact, functions  $\bar{u}(\cdot)$ ,  $\bar{\gamma}(\cdot)$  can be chosen to be measurable on  $[0, \bar{t}_f]$  using a measurable selection lemma (see, e.g., [27, Lemma 3A, p. 161]). Combining (18) with (17) provides

$$\bar{x}(t) = \phi^1(t) \mathbb{1}_{[-\Delta, 0)}(t) + \mathbb{1}_{[0, \bar{t}_f]}(t) \left( \phi^1(0) + \int_0^t f(s, s, \bar{x}(s), \bar{x}(s), \bar{u}(s)) ds \right) ,$$

which proves that the function  $\bar{u}(\cdot)$  is an admissible control for (OCP).

In order to conclude, it remains to show that  $\bar{t}_f = t_f$ ,  $\bar{u}(\cdot) = u(\cdot)$  and  $\bar{x}(\cdot) = x(\cdot)$ .

First, the previous argument shows that

$$C_{\tau_k}(t_f^{\tau_k}, u_{\tau_k}) \rightarrow C_0(\bar{t}_f, \bar{u}) + \int_0^{\bar{t}_f} \bar{\gamma}(t) dt.$$

Thanks to the construction of the mapping  $\Gamma$  in section 3.3.1, for every integer  $k$ , there exists a sequence  $(t_f^k, v_k(\cdot), y_k(\cdot))_{k \in \mathbb{N}}$ , respectively, of final times, of admissible controls and of trajectories for  $(\text{OCP})_{\tau_k}$ , which converges to  $(t_f, u(\cdot), x(\cdot))$  (for the evident topologies) as  $k$  tends to  $\infty$ . Thanks to the optimality of each  $u_{\tau_k}(\cdot)$ , we have  $C_{\tau_k}(t_f^{\tau_k}, u_{\tau_k}) \leq C_{\tau_k}(t_f^k, v_k)$  and, since  $\bar{\gamma}(t) \geq 0$ , passing to the limit gives  $C_0(\bar{t}_f, \bar{u}) \leq C_0(t_f, u)$ . From assumption  $(A_2)$ , we infer  $\bar{t}_f = t_f$ ,  $\bar{u}(\cdot) = u(\cdot)$ ,  $\bar{x}(\cdot) = x(\cdot)$ .

Remark that, from the previous argument, the following weak convergences hold:

$$(19) \quad \begin{cases} \frac{\partial \tilde{f}}{\partial x}(\cdot, \cdot - \tau_k^0, x_{\tau_k}(\cdot), x_{\tau_k}(\cdot - \tau_k^1), u_{\tau_k}(\cdot)) \xrightarrow{(L^\infty)^*} \frac{\partial \tilde{f}}{\partial x}(\cdot, \cdot, x(\cdot), x(\cdot), u(\cdot)) , \\ \frac{\partial \tilde{f}}{\partial y}(\cdot, \cdot - \tau_k^0, x_{\tau_k}(\cdot), x_{\tau_k}(\cdot - \tau_k^1), u_{\tau_k}(\cdot)) \xrightarrow{(L^\infty)^*} \frac{\partial \tilde{f}}{\partial y}(\cdot, \cdot, x(\cdot), x(\cdot), u(\cdot)) . \end{cases}$$

Consider now  $(\text{OCP})_\tau$  with control and state delays satisfying assumption  $(C_1)$ . Since the mappings are control-affine, the previous argument simplifies considerably, because we can transpose the weak convergence directly on controls. Adapting these proofs to much more general systems is very challenging. We adopt free final time to show that, for this step, no problems arise if assumption  $(C_2)$  does not hold.

Let  $(\tau_k)_{k \in \mathbb{N}} = ((\tau_k^0, \tau_k^1, \tau_k^2))_{k \in \mathbb{N}} \subseteq (0, \varepsilon_0)^3$  be an arbitrary sequence of delays converging to 0 as  $k$  tends to  $\infty$  and let  $(x_{\tau_k}(\cdot), u_{\tau_k}(\cdot))$  be an optimal solution of  $(\text{OCP})_{\tau_k}$  with final time  $t_f^{\tau_k}(u_{\tau_k})$ . Since  $t_f^{\tau_k}(u_{\tau_k}) \in [0, b]$ , up to some subsequence, the sequence of final times  $(t_f^{\tau_k})_{k \in \mathbb{N}} = (t_f^{\tau_k}(u_{\tau_k}))_{k \in \mathbb{N}}$  converges to some  $\bar{t}_f \in [0, b]$ . Since  $M_f$  is compact, up to some subsequence, the sequence  $(x_{\tau_k}(t_f^{\tau_k}))_{k \in \mathbb{N}} \subseteq M_f$  converges to a point in  $M_f$ .

On the other hand, thanks to assumption  $(A_1)$ , the sequence  $(u_{\tau_k}(\cdot))_{k \in \mathbb{N}}$  is bounded in  $L^2([-\Delta, \bar{t}_f], \mathbb{R}^m)$ . Therefore, up to some subsequence,  $(u_{\tau_k}(\cdot))_{k \in \mathbb{N}}$  converges to some  $\bar{u}(\cdot) \in L^2([-\Delta, \bar{t}_f], \mathbb{R}^m)$  for the weak topology of  $L^2$ . More precisely, it holds that  $\bar{u}(\cdot) \in L^\infty([-\Delta, \bar{t}_f], U)$ . Indeed,  $(u_{\tau_k}(\cdot))_{k \in \mathbb{N}} \subseteq L^2([-\Delta, \bar{t}_f], U)$  and, thanks to assumption  $(A_1)$ , the set  $L^2([-\Delta, \bar{t}_f], U)$  is closed and convex for the strong topology

of  $L^2$ . Therefore, it is closed and convex for the weak topology of  $L^2$ , from which  $\bar{u}(\cdot) \in L^2([-\Delta, \bar{t}_f], U) \subseteq L^\infty([-\Delta, \bar{t}_f], U)$ . (The last inclusion still follows from  $(A_1)$ .)

At this step, one crucial result is the weak convergence in  $L^2$  of the sequence  $(u_{\tau_k}(\cdot - \tau_k^2))_{k \in \mathbb{N}}$  toward control  $\bar{u}(\cdot)$ . To see this, consider the shift operator

$$S_{\tau^2} : L^2(\mathbb{R}, \mathbb{R}^m) \rightarrow L^2(\mathbb{R}, \mathbb{R}^m) : (t \mapsto \phi(t)) \mapsto (t \mapsto \phi(t - \tau^2)).$$

Using the dominated convergence theorem, it is clear that, for every  $\phi(\cdot) \in L^2(\mathbb{R}, \mathbb{R}^m)$ , it holds that  $\|S_{\tau^2}\phi - \phi\|_{L^2} \rightarrow 0$  as soon as  $\tau^2 \rightarrow 0$ . At this point, extend  $u_{\tau_k}(\cdot)$ ,  $u_{\tau_k}(\cdot - \tau_k^2)$  and  $\bar{u}(\cdot)$  by zero outside of  $[-\Delta, \bar{t}_f]$ . For every  $\varphi(\cdot) \in L^2(\mathbb{R}, \mathbb{R}^m)$ , one obtains

$$\begin{aligned} & \int_0^{\bar{t}_f} (u_{\tau_k}(t - \tau_k) - \bar{u}(t)) \cdot \varphi(t) dt \\ &= \int_{\mathbb{R}} (u_{\tau_k}(t) - \bar{u}(t)) \cdot (S_{-\tau_k^2}\varphi)(t) dt + \int_{\mathbb{R}} (S_{\tau_k^2}\bar{u} - \bar{u})(t) \cdot \varphi(t) dt \\ (20) \quad &= \int_0^{\bar{t}_f} (u_{\tau_k}(t) - \bar{u}(t)) \cdot \varphi(t) dt + \int_{\mathbb{R}} (u_{\tau_k}(t) - \bar{u}(t)) \cdot (S_{-\tau_k^2}\varphi - \varphi)(t) dt \\ & \quad + \int_{\mathbb{R}} (S_{\tau_k^2}\bar{u} - \bar{u})(t) \cdot \varphi(t) dt, \end{aligned}$$

which converges to 0, providing the weak convergence in  $L^2$  of  $(u_{\tau_k}(\cdot - \tau_k^2))_{k \in \mathbb{N}}$  to  $\bar{u}(\cdot)$ .

We can now show that, under assumption  $(C_1)$ , the trajectory arising from control  $\bar{u}(\cdot)$  is admissible for problem (OCP), proceeding as follows. First, remark that, up to continuous extensions, for every  $k$ , we have

$$(21) \quad x_{\tau_k}(t) = \phi^1(t)\mathbb{1}_{[-\Delta, 0)}(t) + \mathbb{1}_{[0, \bar{t}_f]}(t) \left( \phi^1(0) + \int_0^t f(s, s - \tau_k^0, x_{\tau_k}(s), x_{\tau_k}(s - \tau_k^1), u_{\tau_k}(s), u_{\tau_k}(s - \tau_k^2)) ds \right).$$

From this, assumptions  $(A_1)$ ,  $(A_4)$  ensure that  $(x_{\tau_k}(\cdot))_{k \in \mathbb{N}}$  is bounded in  $H^1$ , and then it converges to some  $\bar{x}(\cdot) \in H^1([-\Delta, \bar{t}_f], \mathbb{R}^n)$  for the weak topology of  $H^1$ . Since the immersion of  $H^1$  into  $C^0$  is compact, up to a subsequence,  $(x_{\tau_k}(\cdot))_{k \in \mathbb{N}}$  converges to  $\bar{x}(\cdot) \in C^0([-\Delta, \bar{t}_f], \mathbb{R}^n)$  uniformly in  $[-\Delta, \bar{t}_f]$ . Passing to the limit in (21) gives

$$\bar{x}(t) = \phi^1(t)\mathbb{1}_{[-\Delta, 0)}(t) + \mathbb{1}_{[0, \bar{t}_f]}(t) \left( \phi^1(0) + \int_0^t f(s, \bar{x}(s), \bar{x}(s), \bar{u}(s), \bar{u}(s)) ds \right).$$

In particular, one has  $\bar{x}(\bar{t}_f) \in M_f$ , and then,  $\bar{u}(\cdot)$  is admissible for (OCP).

Similarly to the previous case, thanks to the achieved convergences and assumption  $(C_1)$ , one proves that  $C_0(\bar{t}_f, \bar{u}) \leq C_0(t_f, u)$ . Therefore, from assumption  $(A_2)$ , we infer that  $\bar{t}_f = t_f$ ,  $\bar{u}(\cdot) = u(\cdot)$  and  $\bar{x}(\cdot) = x(\cdot)$  and the conclusion follows.

In this case, we have not only weak convergence of the dynamics and of their derivatives, but also of optimal controls (under appropriate topologies).

The convergence almost everywhere of the optimal controls can be achieved when the second option of assumption  $(C_3)$  holds, and more specifically, when  $u(\cdot)$  assumes its values at extremal points of  $U$ , almost everywhere in  $[-\Delta, t_f]$ .

We proceed as follows. The previous computations provide that  $(u_{\tau_k}(\cdot))_{k \in \mathbb{N}}$  converges to  $u(\cdot)$  for the weak topology of  $L^2$ . At this step, the fact that control  $u(\cdot)$

assumes its values at extremal points of  $U$ , almost everywhere in  $[-\Delta, t_f]$ , implies that  $(u_{\tau_k}(\cdot))_{k \in \mathbb{N}}$  converges to  $u(\cdot)$  for the strong topology of  $L^1$  (see [41, Corollary 1]). Therefore, up to some subsequence,  $(u_{\tau_k}(\cdot))_{k \in \mathbb{N}}$  converges to  $u(\cdot)$ , a.e. in  $[-\Delta, t_f]$ .

*Remark 3.10.* Up to some subsequence, thanks to the computations in (20), both  $(u_{\tau_k}(\cdot - \tau_k^2))_{k \in \mathbb{N}}$  and  $(u_{\tau_k}(\cdot + \tau_k^2))_{k \in \mathbb{N}}$  converge to  $u(\cdot)$ , almost everywhere in  $[-\Delta, t_f]$ .

We have shown that  $(t_f, x(\cdot), u(\cdot))$  (substituted by  $(t_f, x(\cdot), \dot{x}(\cdot))$  for the case of pure state delays) is the unique closure point (for the topologies used above) of  $(t_f^{\tau_k}, x_{\tau_k}(\cdot), u_{\tau_k}(\cdot))_{k \in \mathbb{N}}$  (substituted by  $(t_f^{\tau_k}, x_{\tau_k}(\cdot), \dot{x}_{\tau_k}(\cdot))_{k \in \mathbb{N}}$  for the cases of pure state delays) for any (sub)sequence of delays  $(\tau_k)_{k \in \mathbb{N}}$  converging to 0. Then, convergence holds as well for the whole family  $(t_f^\tau, x_\tau(\cdot), u_\tau(\cdot))_{\tau \in (0, \varepsilon_0)^3}$  (substituted by  $(t_f^\tau, x_\tau(\cdot), \dot{x}_\tau(\cdot))_{\tau \in (0, \varepsilon_0)^3}$  for the cases of pure state delays).

**3.3.4. Convergence of optimal adjoint vectors for (OCP) $_\tau$ .** In what follows,  $(x_\tau(\cdot), u_\tau(\cdot))$  will denote an optimal solution for (OCP) $_\tau$  defined in the interval  $[-\Delta, t_f]$  such that, if needed, it is extended continuously in  $[-\Delta, t_f]$ . From the maximum principle related to (OCP) $_\tau$ , the trajectory  $x_\tau(\cdot)$  is the projection of an extremal  $(x_\tau(\cdot), p_\tau(\cdot), p_\tau^0, u_\tau(\cdot))$  which satisfies (4). From now on, we consider that either assumption (B) or assumption (C) is satisfied, depending on whether we consider pure state delays or not. The main step of this part consists in showing the convergence of the Pontryagin cone of (OCP) $_\tau$  to the Pontryagin cone of (OCP). Since the definition of variation vectors relies on Lebesgue points of optimal controls, we need first a set of converging Lebesgue points. Finally, for the sake of concision, we do not consider final conditions on the state. Recovering the desired convergence results equipped with transversality conditions can be easily done by traveling back the arguments that follow and using assumption (A<sub>1</sub>).

**LEMMA 3.11.** *Consider (OCP) $_\tau$  with pure state delays and assume that assumption (B<sub>1</sub>) holds. For every  $s \in (0, t_f)$ , Lebesgue point of function  $\tilde{f}(\cdot, x(\cdot), x(\cdot), u(\cdot))$ , there exists a family  $(s_\tau)_{(\tau^0, \tau^1) \in (0, \varepsilon_0)^2} \subseteq [s, t_f]$ , which are Lebesgue points of function  $\tilde{f}(\cdot, \cdot - \tau^0, x_\tau(\cdot), x_\tau(\cdot - \tau^1), u_\tau(\cdot))$ , such that*

$$\tilde{f}(s_\tau, s_\tau - \tau^0, x_\tau(s_\tau), x_\tau(s_\tau - \tau^1), u_\tau(s_\tau)) \xrightarrow{\tau \rightarrow 0} \tilde{f}(s, s, x(s), x(s), u(s)), \quad s_\tau \xrightarrow{\tau \rightarrow 0} s.$$

*Conversely, consider (OCP) $_\tau$  with general delays  $\tau = (\tau^0, \tau^1, \tau^2) \in (0, \varepsilon_0)^3$  and assume that assumption (C<sub>3</sub>) holds. For every  $s \in (0, t_f)$ , Lebesgue point of  $u(\cdot)$ , there exists a family  $(s_\tau)_{\tau \in (0, \varepsilon_0)^3} \subseteq [s, t_f]$ , which are Lebesgue points of  $u_\tau(\cdot)$ , of  $u_\tau(\cdot - \tau^2)$  and of  $u_\tau(\cdot + \tau^2)$ , such that*

$$u_\tau(s_\tau) \xrightarrow{\tau \rightarrow 0} u(s), \quad u_\tau(s_\tau - \tau^2) \xrightarrow{\tau \rightarrow 0} u(s), \quad u_\tau(s_\tau + \tau^2) \xrightarrow{\tau \rightarrow 0} u(s), \quad s_\tau \xrightarrow{\tau \rightarrow 0} s.$$

*Proof of Lemma 3.11.* We start by proving the first assertion. For this, denote

$$h^\tau(t) = (h_1^\tau(t), \dots, h_{n+1}^\tau(t)) = \tilde{f}(t, t - \tau^0, x_\tau(t), x_\tau(t - \tau^1), u_\tau(t)),$$

$$h(t) = (h_1(t), \dots, h_{n+1}(t)) = \tilde{f}(t, t, x(t), x(t), u(t)).$$

We prove that, for  $s \in (0, t_f)$  Lebesgue point of  $h(\cdot)$ , for every  $\beta > 0$ ,  $\alpha_s > 0$  (such that  $s + \alpha_s < t_f$ ), there exists  $\gamma_{s, \alpha_s, \beta} > 0$  such that, for every  $(\tau^0, \tau^1) \in (0, \gamma_{s, \alpha_s, \beta})^2$ , there exists  $s_\tau \in [s, s + \alpha_s]$  Lebesgue point of  $h^\tau(\cdot)$  for which  $\|h^\tau(s_\tau) - h(s)\| < \beta$ . By contradiction, suppose that there exists  $s \in (0, t_f)$  Lebesgue point of  $h(\cdot)$ ,  $\alpha_s > 0$ ,  $\beta > 0$  such that, for every integer  $k$ , there exists  $\tau_k \in (0, 1/k)^2 \times \{0\}$  and  $i_k \in \{1, \dots, n+1\}$  for which, for  $t \in [s, s + \alpha_s]$  Lebesgue point of  $h^{\tau_k}(\cdot)$ , it holds that  $|h_{i_k}^{\tau_k}(t) - h_{i_k}(s)| \geq \beta$ .

From the previous results, the family  $(h^\tau(\cdot))_{\tau \in (0, \varepsilon_0)^2 \times \{0\}}$  converges to  $h(\cdot)$  in  $L^\infty$  for the weak star topology. Therefore, for every  $0 < \delta \leq 1$ , there exists an integer  $k_\delta$  such that, for every  $k \geq k_\delta$ , it holds that

$$\frac{1}{\delta \alpha_s} \left| \int_s^{s+\delta \alpha_s} h_i^{\tau k}(t) \, dt - \int_s^{s+\delta \alpha_s} h_i(t) \, dt \right| < \frac{\beta}{3}$$

for every  $i \in \{1, \dots, n+1\}$ . We exploit this fact to bound  $|h_{i_k}^{\tau k}(t) - h_{i_k}(s)|$  by  $\beta$ .

First, since  $s$  is a Lebesgue point of  $h(\cdot)$ , there exists  $0 < \delta_{s, \alpha_s} \leq 1$  such that

$$\left| h_i(s) - \frac{1}{\delta_{s, \alpha_s} \alpha_s} \int_s^{s+\delta_{s, \alpha_s} \alpha_s} h_i(t) \, dt \right| < \frac{\beta}{3}$$

for  $i \in \{1, \dots, n+1\}$ . On the other hand, there exists an integer  $k_{\delta_{s, \alpha_s}}$  such that

$$\frac{1}{\delta_{s, \alpha_s} \alpha_s} \left| \int_s^{s+\delta_{s, \alpha_s} \alpha_s} h_i^{\tau k}(t) \, dt - \int_s^{s+\delta_{s, \alpha_s} \alpha_s} h_i(t) \, dt \right| < \frac{\beta}{3}$$

for every  $k \geq k_{\delta_{s, \alpha_s}}$  and every  $i \in \{1, \dots, n+1\}$ . Finally, by assumption, we have that  $h^\tau(\cdot)$  is continuous for  $\tau^0, \tau^1 > 0$ , and then, for every  $k \geq k_{\delta_{s, \alpha_s}}$  and every  $i \in \{1, \dots, n+1\}$ , there exists  $t_{k,i} \in [s, s + \delta_{s, \alpha_s} \alpha_s] \subseteq [s, s + \alpha_s]$  such that

$$\left| h_i^{\tau k}(t_{k,i}) - \frac{1}{\delta_{s, \alpha_s} \alpha_s} \int_s^{s+\delta_{s, \alpha_s} \alpha_s} h_i^{\tau k}(t) \, dt \right| < \frac{\beta}{3}.$$

Then, for every  $\tau_k \in \left(0, \frac{1}{k_{\delta_{s, \alpha_s}}}\right)^2 \times \{0\}$ ,  $i \in \{1, \dots, n+1\}$  there exists  $t_{k,i} \in [s, s + \alpha_s]$  Lebesgue point of  $h^{\tau k}(\cdot)$  such that  $|h_i^{\tau k}(t_{k,i}) - h_i(s)| < \beta$ , which is a contradiction.

Now, we consider the second statement. The case for which assumption  $(C_3)$  ensures that, for every delay  $\tau$ , every optimal control  $u_\tau(\cdot)$  of  $(\text{OCP})_\tau$  is continuous is treated as above because of the weak convergence in  $L^2$  of  $u_\tau(\cdot)$ , of  $u_\tau(\cdot - \tau^2)$  and of  $u_\tau(\cdot + \tau^2)$ . Therefore, assume that  $u(\cdot)$  takes its values at extremal points of  $U$ , almost everywhere in  $[-\Delta, t_f]$ . Without loss of generality, we extend  $u_\tau(\cdot)$  by some constant vector of  $U$  in  $[t_f^\tau, b]$ . Denote

$$\begin{aligned} h^\tau(t) &= (h_1^\tau(t), \dots, h_{3m}^\tau(t)) = (u_\tau(t), u_\tau(t - \tau^2), u_\tau(t + \tau^2)), \\ h(t) &= (h_1(t), \dots, h_{3m}(t)) = (u(t), u(t), u(t)) \end{aligned}$$

and fix  $s \in (0, t_f)$ , Lebesgue point of  $h(\cdot)$ . By contradiction, suppose that there exist  $\beta > 0$  and  $\alpha > 0$  such that, for every integer  $k$ , there exist  $\tau_k = (\tau_k^0, \tau_k^1, \tau_k^2) \in (0, 1/k)^3$  and  $i_k \in \{1, \dots, 3m\}$  for which, for every  $r \in [s, s + \alpha]$  Lebesgue point of  $h^{\tau k}(\cdot)$ , it holds that  $|h_{i_k}^{\tau k}(r) - h_{i_k}(s)| \geq \beta$ . From the arguments of the previous sections, up to some extension, the family of controls  $(u_\tau(\cdot))_{\tau \in (0, \varepsilon_0)^3}$  converges to  $u(\cdot)$  almost everywhere in  $[0, t_f]$  and the same holds true for  $(u_\tau(\cdot - \tau^2))_{\tau \in (0, \varepsilon_0)^3}$  and  $(u_\tau(\cdot + \tau^2))_{\tau \in (0, \varepsilon_0)^3}$ , thanks to Remark 3.10. Then,  $(h_i^{\tau k}(\cdot))_{k \in \mathbb{N}}$  converges a.e. to  $h_i(\cdot)$ , raising a contradiction.  $\square$

Lemma 3.11 allows us to prove the following property for Pontryagin cones.

**LEMMA 3.12.** *For every  $\tilde{v} \in \tilde{K}^0(t_f)$  and every  $\tau = (\tau^0, \tau^1, \tau^2) \in (0, \varepsilon_0)^3$  (as well as  $\tau = (\tau^0, \tau^1, 0) \in (0, \varepsilon_0)^2 \times \{0\}$  in the case of pure state delays), there exists  $\tilde{w}_\tau \in \tilde{K}^\tau(t_f^\tau)$  such that the family  $(\tilde{w}_\tau)_{\tau \in (0, \varepsilon_0)^3}$  converges to  $\tilde{v}$  as  $\tau$  tends to 0.*

*Proof of Lemma 3.12.* We prove the statement for problems  $(\text{OCP})_\tau$  with general state and control delays  $\tau = (\tau^0, \tau^1, \tau^2)$ . If pure state delay problems  $(\text{OCP})_\tau$  are considered, the same guideline can be employed by using Lemma 3.11 and (19).

Suppose first that  $\tilde{v} = \tilde{v}_{s,\omega_z(s)}^0(t_f)$ , where  $z \in U$  and  $0 < s < t_f$  is a Lebesgue point of  $u(\cdot)$  (recall Remark 3.3). By definition,  $\tilde{v}_{s,\omega_z(s)}^0(\cdot)$  is the solution of

$$(22) \begin{cases} \dot{\psi}(t) = \left( \frac{\partial \tilde{f}}{\partial x}(t, t, x(t), x(t), u(t), u(t)) + \frac{\partial \tilde{f}}{\partial y}(t, t, x(t), x(t), u(t), u(t)) \right) \psi(t), \\ \psi(s) = \tilde{f}(s, s, x(s), x(s), z, z) - \tilde{f}(s, s, x(s), x(s), u(s), u(s)). \end{cases}$$

From Lemma 3.11, there exists a family  $(s_\tau)_{\tau \in (0, \varepsilon_0)^3} \subseteq [s, t_f]$ , which are Lebesgue points of  $u_\tau(\cdot)$ , of  $u_\tau(\cdot - \tau^2)$  and of  $u_\tau(\cdot + \tau^2)$ , such that

$$u_\tau(s_\tau) \xrightarrow{\tau \rightarrow 0} u(s), \quad u_\tau(s_\tau - \tau^2) \xrightarrow{\tau \rightarrow 0} u(s), \quad u_\tau(s_\tau + \tau^2) \xrightarrow{\tau \rightarrow 0} u(s), \quad s_\tau \xrightarrow{\tau \rightarrow 0} s.$$

This allows us to consider  $\tilde{v}_{s_\tau, \omega_z^-(s_\tau)}^\tau(\cdot)$  and  $\tilde{v}_{s_\tau + \tau^2, \omega_z^+(s_\tau)}^\tau(\cdot)$ , solutions of (13) with initial data provided, respectively, by (11) and (12). We denote

$$\tilde{w}_{s_\tau, z}^\tau(t) = \tilde{v}_{s_\tau, \omega_z^-(s_\tau)}^\tau(t) + \tilde{v}_{s_\tau + \tau^2, \omega_z^+(s_\tau)}^\tau(t).$$

Since we consider affine problems  $(\text{OCP})_\tau$ , Lemma 3.11 gives

$$\lim_{\tau \rightarrow 0} \left( \omega_z^-(s_\tau) + \omega_z^+(s_\tau) \right) = \tilde{f}(s, s, x(s), x(s), z, z) - \tilde{f}(s, s, x(s), x(s), u(s), u(s)).$$

Moreover, from the results of the previous sections, we have in particular

$$\begin{aligned} \frac{\partial \tilde{f}}{\partial x}(\cdot, \cdot - \tau^0, x_\tau(\cdot), x_\tau(\cdot - \tau^1), u_\tau(\cdot), u_\tau(\cdot - \tau^2)) &\xrightarrow{L^2} \frac{\partial \tilde{f}}{\partial x}(\cdot, \cdot, x(\cdot), x(\cdot), u(\cdot), u(\cdot)), \\ \frac{\partial \tilde{f}}{\partial y}(\cdot, \cdot - \tau^0, x_\tau(\cdot), x_\tau(\cdot - \tau^1), u_\tau(\cdot), u_\tau(\cdot - \tau^2)) &\xrightarrow{L^2} \frac{\partial \tilde{f}}{\partial y}(\cdot, \cdot, x(\cdot), x(\cdot), u(\cdot), u(\cdot)). \end{aligned}$$

By continuous dependence w.r.t initial data for dynamical systems and since  $t_f^\tau$  converges to  $t_f$ , the family  $(\tilde{w}_\tau)_{\tau \in (0, \varepsilon_0)^3} = (\tilde{w}_{s_\tau, z}^\tau(t_f^\tau))_{\tau \in (0, \varepsilon_0)^3}$  converges to  $\tilde{v}$  as  $\tau \rightarrow 0$ .

If  $\tilde{v} \in \partial \tilde{K}^0(t_f)$ , the result above is used on converging sequences in  $\text{Int } \tilde{K}^0(t_f)$ .  $\square$

For the last part of the proof, an iterative use of Lemma 3.12 is made. It is at this step that, for problems with general delays  $\tau = (\tau^0, \tau^1, \tau^2)$ , assumption  $(C_2)$  of fixed final time becomes instrumental to derive the convergence related to adjoint vectors. Indeed, problems arise when one tries to make the final condition on the Hamiltonian (7) converge to the transversality condition related to problem  $(\text{OCP})_\tau$ . For the sake of concision, in this context, we focus only on problems  $(\text{OCP})_\tau$  with general delays  $\tau = (\tau^0, \tau^1, \tau^2)$ . The case of pure state delays is similar. (We refer to [22, Proposition 2.15] for details.) Assumptions  $(B)$  and  $(C)$  are implicitly used.

We first prove that the following statements hold true:

- For every  $\tau = (\tau^0, \tau^1, \tau^2) \in (0, \varepsilon_0)^3$ , every extremal lift  $(x_\tau(\cdot), p_\tau(\cdot), p_\tau^0, u_\tau(\cdot))$  of any solution of  $(\text{OCP})_\tau$  is normal.
- The set of final adjoint vectors  $\{p_\tau(t_f) : \tau \in (0, \varepsilon_0)^3\}$  is bounded.

We consider the first statement proceeding by contradiction. Assume that, for every integer  $k$ , there exist  $\tau_k = (\tau_k^0, \tau_k^1, \tau_k^2) \in (0, 1/k)^3$  and a solution  $(x_{\tau_k}(\cdot), u_{\tau_k}(\cdot))$  of  $(\text{OCP})_{\tau_k}$  having an abnormal extremal lift  $(x_{\tau_k}(\cdot), p_{\tau_k}(\cdot), 0, u_{\tau_k}(\cdot))$ . Set  $\psi_{\tau_k} = \frac{p_{\tau_k}(t_f)}{\|p_{\tau_k}(t_f)\|}$  for every integer  $k$ . Therefore, we have  $\langle \psi_{\tau_k}, 0 \rangle, \tilde{v}_{\tau_k} \rangle \leq 0$  for every  $\tilde{v}_{\tau_k} \in \tilde{K}^{\tau_k}(t_f)$  and every integer  $k$ . Up to a subsequence, the sequence  $(\psi_{\tau_k})_{k \in \mathbb{N}} \subseteq S^{n-1}$

converges to some unit vector  $\psi \in \mathbb{R}^n$ . Passing to the limit, by using the previous results, we infer that  $\langle (\psi, 0), \tilde{v} \rangle \leq 0$  for every  $\tilde{v} \in \tilde{K}^0(t_f)$ . Thanks to assumption  $(C_2)$ ,  $(x(\cdot), u(\cdot))$  has an abnormal extremal lift. This contradicts assumption  $(A_3)$ .

For the second statement, again by contradiction, assume that there exists a sequence  $(\tau_k = (\tau_k^0, \tau_k^1, \tau_k^2))_{k \in \mathbb{N}} \subseteq (0, \varepsilon_0)^3$  converging to 0 such that  $\|p_{\tau_k}(t_f)\|$  tends to  $+\infty$ . As defined above, the sequence  $(\psi_{\tau_k})_{k \in \mathbb{N}}$  belongs to  $S^{n-1}$ , and then, up to some subsequence, it converges to some unit vector  $\psi$ . On the other hand, by construction, the inequality  $\langle (p_{\tau_k}(t_f), -1), \tilde{v}_{\tau_k} \rangle \leq 0$  holds for every  $\tilde{v}_{\tau_k} \in \tilde{K}^{\tau_k}(t_f)$  and every integer  $k$ . Dividing by  $\|p_{\tau_k}(t_f)\|$  and passing to the limit, it follows that the solution  $(x(\cdot), u(\cdot))$  has an abnormal extremal lift, which again contradicts assumption  $(A_3)$ .

Now, let  $\psi$  be a closure point of  $\{p_\tau(t_f) : \tau \in (0, \varepsilon_0)^3\}$  and  $(\tau_k = (\tau_k^0, \tau_k^1, \tau_k^2))_{k \in \mathbb{N}} \subseteq (0, \varepsilon_0)^3$  be a sequence converging to 0 such that  $p_{\tau_k}(t_f)$  tends to  $\psi$ . Using the continuous dependence w.r.t. initial data and the established convergence properties, we infer that  $(p_{\tau_k}(\cdot))_{k \in \mathbb{N}}$  converges uniformly to the solution  $z(\cdot)$  of

$$\dot{z}(t) = -\frac{\partial H}{\partial x}(t, t, x(t), x(t), z(t), -1, u(t), u(t)) - \frac{\partial H}{\partial y}(t, t, x(t), x(t), z(t), -1, u(t), u(t))$$

with  $z(t_f) = \psi$ . Moreover, since  $\langle (p_{\tau_k}(t_f), -1), \tilde{v}_{\tau_k} \rangle \leq 0$ , for every  $\tilde{v}_{\tau_k} \in \tilde{K}^{\tau_k}(t_f)$  and every integer  $k$ , passing to the limit, thanks to the previous results, we obtain  $\langle (\psi, -1), \tilde{v} \rangle \leq 0$ , for every  $\tilde{v} \in \tilde{K}^0(t_f)$ . It follows that  $(x(\cdot), z(\cdot), -1, u(\cdot))$  is a normal extremal lift of (OCP). Using assumption  $(A_3)$ , we finally obtain  $z(\cdot) = p(\cdot)$  in  $[0, t_f]$ .

Theorem 2.1 is proved.

**4. Conclusions and perspectives.** In this paper, we provided sufficient conditions under which Pontryagin extremals related to nonlinear optimal control problems with delays are continuous (under appropriate topologies) w.r.t. delays.

It would be interesting to extend this result to problems with more general constraints, such as control and state constraints. This would require us to analyze the proof of the maximum principle with state and control constraints via sliding or v-variations (see, e.g., [13]), to exploit the continuous dependence w.r.t. parameters for implicit function theorems. (Ekeland-type approaches probably fail because of continuous dependence.) Furthermore, in the case of control and state delays, the proof that we provided needs to consider control-affine dynamics and costs and fixed final time. The extension to more general systems is open.

Furthermore, in this paper we have considered constant delays. One could consider more general delays that are functions of time and state. This is motivated by the fact that, in the case of delays depending on the time and the state, maximum principle formulations still exist (see, e.g., [4]). Therefore, extending our main result requires to consider the  $C^0$ -topology on the delay function  $t \mapsto \tau(t, x(t))$ .

Finally, as briefly explained in section 2.4, our result is in particular motivated by numerical implementations of the shooting method, in combination with homotopies on the delay parameters, thus providing an interesting alternative to classically used direct methods. Numerical issues will be addressed in a forthcoming paper.

**Appendix A. Proof of Lemma 3.4.** The proof goes by induction. We develop computations for  $j = 1$ . The inductive step goes in the same way as the usual case (see, e.g., [32]).

Let  $t_j < t \leq t_j^\tau$  be a Lebesgue point of  $u_\tau(\cdot)$ ,  $u_\tau(\cdot - \tau^2)$ . First, let us show that

$$(23) \quad \tilde{x}_\tau^\pi(t) - \tilde{x}_\tau(t) = \eta_1 \tilde{w}_{t_1, u_1}^\tau(t) + o(\eta_1) = \eta_1 \left( \tilde{v}_{t_1, \omega_{u_1}^-(t_1)}^\tau(t) + \tilde{v}_{t_1 + \tau^2, \omega_{u_1}^+(t_1)}^\tau(t) \right) + o(\eta_1).$$

We consider the case  $t \geq t_1 + \tau_2$ . (The case  $t < t_1 + \tau_2$  is similar, but easier.) We have

$$\begin{aligned} & \|\tilde{x}_\tau^\pi(t) - \tilde{x}_\tau(t) - \eta_1 \tilde{w}_{t_1, u_1}^\tau(t)\| \leq \|\tilde{x}_\tau^\pi(t_1 + \tau^2) - \tilde{x}_\tau(t_1 + \tau^2) - \eta_1 \tilde{w}_{t_1, u_1}^\tau(t_1 + \tau^2)\| \\ & + \left\| \int_{t_1 + \tau_2}^t \left( \tilde{f}(s, s - \tau^0, \tilde{x}_\tau^\pi(s), \tilde{x}_\tau^\pi(s - \tau^1), u_\tau(s), u_\tau(s - \tau^2)) \right. \right. \\ & \quad \left. \left. - \tilde{f}(s, s - \tau^0, \tilde{x}_\tau(s), \tilde{x}_\tau(s - \tau^1), u_\tau(s), u_\tau(s - \tau^2)) - \eta_1 \dot{\tilde{w}}_{t_1, u_1}^\tau(s) \right) ds \right\|. \end{aligned}$$

By exploiting the facts that  $t_1$  is a Lebesgue point of  $u_\tau(\cdot)$ ,  $u_\tau(\cdot - \tau^2)$  and of  $u_\tau(\cdot + \tau^2)$  and that  $\tilde{x}_\tau^\pi(\cdot)$  converges uniformly to  $\tilde{x}_\tau(\cdot)$ , expanding the extended dynamics at second order, the first term of the expression above can be bounded as follows:

$$\begin{aligned} & \|\tilde{x}_\tau^\pi(t_1 + \tau^2) - \tilde{x}_\tau(t_1 + \tau^2) - \eta_1 \tilde{w}_{t_1, u_1}^\tau(t_1 + \tau^2)\| \\ & \leq \left\| \int_{t_1 - \eta_1}^{t_1} \left( \tilde{f}(s, s - \tau^0, \tilde{x}_\tau^\pi(s), \tilde{x}_\tau^\pi(s - \tau^1), u_1, u_\tau(s - \tau^2)) \right. \right. \\ & \quad \left. \left. - \tilde{f}(s, s - \tau^0, \tilde{x}_\tau(s), \tilde{x}_\tau(s - \tau^1), u_\tau(s), u_\tau(s - \tau^2)) \right) ds - \eta_1 \omega_{u_1}^-(t_1) \right\| \\ & \quad + \left\| \int_{t_1}^{t_1 + \tau^2 - \eta_1} \left( \tilde{f}(s, s - \tau^0, \tilde{x}_\tau^\pi(s), \tilde{x}_\tau^\pi(s - \tau^1), u_\tau(s), u_\tau(s - \tau^2)) \right. \right. \\ & \quad \left. \left. - \tilde{f}(s, s - \tau^0, \tilde{x}_\tau(s), \tilde{x}_\tau(s - \tau^1), u_\tau(s), u_\tau(s - \tau^2)) - \eta_1 \dot{\tilde{w}}_{t_1, u_1}^\tau(s) \right) ds \right\| \\ & \quad + \left\| \int_{t_1 + \tau^2 - \eta_1}^{t_1 + \tau^2} \left( \tilde{f}(s, s - \tau^0, \tilde{x}_\tau^\pi(s), \tilde{x}_\tau^\pi(s - \tau^1), u_\tau(s), u_1) \right. \right. \\ & \quad \left. \left. - \tilde{f}(s, s - \tau^0, \tilde{x}_\tau(s), \tilde{x}_\tau(s - \tau^1), u_\tau(s), u_\tau(s - \tau^2)) - \eta_1 \dot{\tilde{w}}_{t_1, u_1}^\tau(s) \right) ds - \eta_1 \omega_{u_1}^+(t_1) \right\| \\ & \leq \left\| \int_{t_1 - \eta_1}^{t_1} \left( \tilde{f}(s, s - \tau^0, \tilde{x}_\tau(s), \tilde{x}_\tau(s - \tau^1), u_1, u_\tau(s - \tau^2)) \right. \right. \\ & \quad \left. \left. - \tilde{f}(s, s - \tau^0, \tilde{x}_\tau(s), \tilde{x}_\tau(s - \tau^1), u_\tau(s), u_\tau(s - \tau^2)) \right) ds - \eta_1 \omega_{u_1}^-(t_1) \right\| + o(\eta_1) \\ & \quad + \int_{t_1}^{t_1 + \tau^2 - \eta_1} \left\| \frac{\partial \tilde{f}}{\partial x} (s, s - \tau^0, \tilde{x}_\tau(s), \tilde{x}_\tau(s - \tau^1), u_\tau(s), u_\tau(s - \tau^2)) \right. \\ & \quad \left. \cdot \left( \tilde{x}_\tau^\pi(s) - \tilde{x}_\tau(s) - \eta_1 \tilde{w}_{t_1, u_1}^\tau(s) \right) \right\| ds \\ & \quad + \int_{t_1}^{t_1 + \tau^2 - \eta_1} \left\| \frac{\partial \tilde{f}}{\partial y} (s, s - \tau^0, \tilde{x}_\tau(s), \tilde{x}_\tau(s - \tau^1), u_\tau(s), u_\tau(s - \tau^2)) \right. \\ & \quad \left. \cdot \left( \tilde{x}_\tau^\pi - \tilde{x}_\tau - \eta_1 \tilde{w}_{t_1, u_1}^\tau \right) (s - \tau^1) \right\| ds \\ & + \int_{t_1}^{t_1 + \tau^2 - \eta_1} \int_0^1 \left\| d^2 \tilde{f}(s, s - \tau^0, (\sigma \tilde{x}_\tau + (1 - \sigma) \tilde{x}_\tau^\pi)(s), (\sigma \tilde{x}_\tau + (1 - \sigma) \tilde{x}_\tau^\pi)(s - \tau^1), u_\tau(s), u_\tau(s - \tau^2)) \right\| \\ & \cdot \left( \|\tilde{x}_\tau^\pi(s) - \tilde{x}_\tau(s)\|^2 + \|\tilde{x}_\tau^\pi(s - \tau^1) - \tilde{x}_\tau(s - \tau^1)\|^2 + 2 \|\tilde{x}_\tau^\pi(s) - \tilde{x}_\tau(s)\| \|\tilde{x}_\tau^\pi(s - \tau^1) - \tilde{x}_\tau(s - \tau^1)\| \right) d\sigma ds \\ & + \eta_1 \left\| \int_{t_1 - \eta_1}^{t_1} \dot{\tilde{w}}_{t_1, u_1}^\tau(s + \tau^2) ds \right\| \\ & \quad + \left\| \int_{t_1 - \eta_1}^{t_1} \left( \tilde{f}(s + \tau^2, s + \tau^2 - \tau^0, \tilde{x}_\tau(s + \tau^2), \tilde{x}_\tau(s + \tau^2 - \tau^1), u_\tau(s + \tau^2), u_1) \right. \right. \\ & \quad \left. \left. - \tilde{f}(s + \tau^2, s + \tau^2 - \tau^0, \tilde{x}_\tau(s + \tau^2), \tilde{x}_\tau(s + \tau^2 - \tau^1), u_\tau(s + \tau^2), u_\tau(s)) \right) ds - \eta_1 \omega_{u_1}^+(t_1) \right\| + o(\eta_1). \end{aligned}$$



Therefore, by bounding the derivatives of the extended dynamics, we have

$$\begin{aligned} & \| \tilde{x}_\tau^\pi(t_1 + \tau^2) - \tilde{x}_\tau(t_1 + \tau^2) - \eta_1 \tilde{w}_{t_1, u_1}^\tau(t_1 + \tau^2) \| \\ & \leq \tilde{C}_1 \int_{t_1 - \tau^1}^{t_1 + \tau^2 - \eta_1} \| \tilde{x}_\tau^\pi(s) - \tilde{x}_\tau(s) - \eta_1 \tilde{w}_{t_1, u_1}^\tau(s) \| ds + o(\eta_1), \end{aligned}$$

where  $\tilde{C}_1 \geq 0$  is a constant. With the same technique, we obtain

$$\begin{aligned} & \left\| \int_{t_1 + \tau_2}^t \left( \tilde{f}(s, s - \tau^0, \tilde{x}_\tau^\pi(s), \tilde{x}_\tau^\pi(s - \tau^1), u_\tau(s), u_\tau(s - \tau^2)) \right. \right. \\ & \quad \left. \left. - \tilde{f}(s, s - \tau^0, \tilde{x}_\tau(s), \tilde{x}_\tau(s - \tau^1), u_\tau(s), u_\tau(s - \tau^2)) \right. \right. \\ & \quad \left. \left. - \eta_1 \dot{\tilde{w}}_{t_1, u_1}^\tau(s) \right) ds \right\| \leq \tilde{C}_2 \int_{t_1 + \tau^2 - \tau^1}^t \| \tilde{x}_\tau^\pi(s) - \tilde{x}_\tau(s) - \eta_1 \tilde{w}_{t_1, u_1}^\tau(s) \| ds + o(\eta_1), \end{aligned}$$

where  $\tilde{C}_2 \geq 0$  is another constant. Coupling the two last results with Grönwall’s inequality, (23) follows. The conclusion comes from (23) and the fact that  $t$  is a Lebesgue point of  $u_\tau(\cdot)$  and of  $u_\tau(\cdot - \tau^2)$ .

**Appendix B. Proof of Lemma 3.8.** We start by recalling the following standard result (see, e.g., [1]).

Let  $L : \mathbb{R}^j \rightarrow \mathbb{R}^n$  be a linear mapping such that  $L(\mathbb{R}_+^j) = \mathbb{R}^n$ . Then the following hold:

- $j > n + 1$  and  $(0, +\infty)^j \cap \ker L$  is nontrivial.
- There exists  $S \subseteq \mathbb{R}^j$ ,  $\dim(S) = n$ , such that  $L|_S : S \rightarrow \mathbb{R}^n$  is an isomorphism.

Applying this result to  $L = \frac{\partial F}{\partial x}(0, 0)$  yields the existence of a nontrivial vector  $v \in (0, +\infty)^j$ , such that  $L(v) = 0$ , and of a  $n$ -dimensional subspace  $S \subseteq \mathbb{R}^j$  such that the restriction  $L|_S : S \rightarrow \mathbb{R}^n$  is an isomorphism.

For every  $\varepsilon \in \mathbb{R}_+^k$  and every  $y, u \in \mathbb{R}^n$ , set  $\Phi(\varepsilon, y, u) = u - F(\varepsilon, L|_S^{-1}(u)) + y$ . This mapping is continuous and it holds that  $\Phi(0, 0, 0) = 0$ . Fix  $\varepsilon \in \mathbb{R}_+^k$  at which  $F$  is almost everywhere strictly differentiable. Then, for every  $y \in \mathbb{R}^n$ , one has

$$(24) \quad \Phi(\varepsilon, y, u_1) - \Phi(\varepsilon, y, u_2) = \left( \text{Id} - \frac{\partial F}{\partial x}(\varepsilon, 0) \circ L|_S^{-1} \right) (u_1 - u_2) + \|u_1 - u_2\| G_\varepsilon(u_1, u_2),$$

where  $G_\varepsilon(u_1, u_2) = g_\varepsilon(L|_S^{-1}(u_2), L|_S^{-1}(u_1)) \rightarrow 0$  as soon as  $(u_1, u_2) \xrightarrow{\text{a.e.}} 0$ . From the continuity property of  $\frac{\partial F}{\partial x}(\varepsilon, 0)$  on a dense subset, there exists  $\varepsilon_0 \in \mathbb{R}_+^k$  and a dense subset  $E \subseteq [0, \varepsilon_0]^k$ , such that for every  $\varepsilon \in E$  it holds that  $\left\| \text{Id} - \frac{\partial F}{\partial x}(\varepsilon, 0) \circ L|_S^{-1} \right\| \leq \frac{1}{4}$ , and there exists  $r_\varepsilon > 0$  such that

$$\|G_\varepsilon(u_1, u_2)\| \leq \frac{1}{4} \quad \text{for almost all } u_1, u_2 \in B_{r_\varepsilon}(0).$$

On the other hand, by assumption, the remainder in expression (24) converges to 0 uniformly w.r.t.  $\varepsilon$  on a dense subset. Therefore, up to reducing  $E$ , gathering the previous results with (24), we infer the existence of  $r > 0$  such that

$$\|\Phi(\varepsilon, y, u_1) - \Phi(\varepsilon, y, u_2)\| \leq \frac{1}{2} \|u_1 - u_2\| \quad \text{for every } \varepsilon \in E \text{ and almost every } u_1, u_2 \in B_r(0).$$

From this last result and the continuity of mapping  $F$ , for every  $\varepsilon \in [0, \varepsilon_0]^k$  and  $y \in \mathbb{R}^n$ , the mapping  $u \mapsto \Phi(\varepsilon, y, u)$  is  $\frac{1}{2}$ -Lipschitzian on an open neighborhood of 0.

At this step, for every  $\delta > 0$ , denote  $B_\delta = S \cap \overline{B}_\delta(0)$  and choose  $\delta > 0$  small enough such that  $v + B_\delta \subseteq (0, +\infty)^j$ . The set  $U_\delta = L(B_\delta)$  is a closed neighborhood of 0 in  $\mathbb{R}^n$ . With the same argument as above, it is not difficult to show that if  $\delta$ ,  $\|\varepsilon\|$ , and  $\|y\|$  are small enough, then the mapping  $u \mapsto \Phi(\varepsilon, y, u)$  maps  $U_\delta$  into itself.

Lemma 3.8 follows from the application of the usual Banach fixed point theorem to the contraction mapping  $u \mapsto \Phi(\varepsilon, y, u)$  with parameters  $(\varepsilon, y)$ .

## REFERENCES

- [1] A. AGRACHEV AND Y. SACHKOV, *Control Theory from the Geometric Viewpoint*, Springer, Berlin, 2004.
- [2] E. L. ALLGOWER AND K. GEORG, *Introduction to Numerical Continuation Methods*, SIAM, Philadelphia, 2003.
- [3] P. ANTOINE AND H. ZOUAKI, *Etude locale de l'ensemble des points critiques d'un problème d'optimisation paramétré*, C. R. Acad. Sci. Paris Sér. I Math., 310 (1990), pp. 587–590.
- [4] R. B. ASHER AND H. R. SEBESTA, *Optimal control of systems with state-dependent time delay*, Internat. J. Control, 14 (1971), pp. 353–365.
- [5] J.-P. AUBIN AND H. FRANKOWSKA, *Set-Valued Analysis*, Springer, Berlin, 2009.
- [6] H. T. BANKS, *Necessary conditions for control problems with variable time lags*, SIAM J. Control Optim., 6 (1968), pp. 9–47.
- [7] A. BOCCIA AND R. B. VINTER, *The maximum principle for optimal control problems with time delays*, IFAC-PapersOnLine, 49 (2016), pp. 951–955.
- [8] R. BONALLI, B. HÉRISSE, AND E. TRÉLAT, *Solving optimal control problems for delayed control-affine systems with quadratic cost by numerical continuation*, in Proceedings of the American Control Conference (ACC), IEEE, New York, 2017, pp. 649–654.
- [9] B. BONNARD, J.-B. CAILLAU, AND E. TRÉLAT, *Second order optimality conditions in the smooth case and applications in optimal control*, ESAIM Control Optim. Calc. Var., 13 (2007), pp. 207–236.
- [10] B. BONNARD AND M. CHYBA, *Singular Trajectories and their Role in Control Theory*, Springer, Berlin, 2003.
- [11] B. BONNARD, L. FAUBOURG, AND E. TRÉLAT, *Mécanique céleste et contrôle de véhicules spatiaux*, Springer, Berlin, 2006.
- [12] F. H. CLARKE AND R. B. VINTER, *The relationship between the maximum principle and dynamic programming*, SIAM J. Control Optim., 25 (1987), pp. 1291–1311.
- [13] A. V. DMITRUK, *On the development of Pontryagin's maximum principle in the works of A. Ya. Dubovitskii and A. A. Milyutin*, Control Cybernet., 38 (2009), pp. 923–957.
- [14] A. V. DMITRUK, A. A. MILYUTIN, AND N. P. OSMOLOVSKII, *Lyusternik's theorem and the theory of extrema*, Russian Math. Surveys, 35 (1980), pp. 11–51.
- [15] L. C. EVANS AND R. F. GARIEPY, *Measure Theory and Fine Properties of Functions*, Taylor & Francis, New York, 1991.
- [16] A. F. FILIPPOV, *On certain questions in the theory of optimal control*, SIAM J. Control Optim., 1 (1962), pp. 76–84.
- [17] R. GAMKRELIDZE, *Principles of Optimal Control Theory*, Springer, Berlin, 2013.
- [18] L. GÖLLMANN, D. KERN, AND H. MAURER, *Optimal control problems with delays in state and control variables subject to mixed control-state constraints*, Optimal Control Appl. Methods, 30 (2009), pp. 341–365.
- [19] L. GÖLLMANN AND H. MAURER, *Theory and applications of optimal control problems with multiple time-delays*, J. Ind. Manag. Optim., 10 (2014), pp. 413–441.
- [20] L. GÖLLMANN AND H. MAURER, *Optimal control problems with time delays: Two case studies in biomedicine*, Math. Biosci. Eng., 15 (2018), pp. 1137–1154.
- [21] T. GUINN, *Reduction of delayed optimal control problems to nondelayed problems*, J. Optim. Theory Appl., 18 (1976), pp. 371–377.
- [22] T. HABERKORN AND E. TRÉLAT, *Convergence results for smooth regularizations of hybrid nonlinear optimal control problems*, SIAM J. Control Optim., 49 (2011), pp. 1498–1522.
- [23] A. HALANAY, *Differential Equations: Stability, Oscillations, Time Lags*, Academic Press, New York, 1966.
- [24] A. HALANAY, *Optimal controls for systems with time lag*, SIAM J. Control Optim., 6 (1968), pp. 215–234.
- [25] G. L. KHARATISHVILI, *Maximum principle in the theory of optimal time-delay processes*, Dokl. Akad. Nauk, 136 (1961), pp. 39–42.

- [26] G. L. KHARATISHVILI, *A Maximum Principle in External Problems with Delays*, Academic Press, New York, 1967.
- [27] E. LEE AND L. MARKUS, *Foundations of Optimal Control Theory*, Wiley, New York, 1967.
- [28] M.-Z. MANU AND J. MOHAMMAD, *Time-Delay Systems: Analysis, Optimization and Applications*, Elsevier Science, New York, 1987.
- [29] G. I. MARCHUK, *Adjoint Equations and Analysis of Complex Systems*, Springer, Berlin, 1995.
- [30] G. I. MARCHUK, *Mathematical Modelling of Immune Response in Infectious Diseases*, Springer, Berlin, 1997.
- [31] S. NABABAN, *A Filippov-type lemma for functions involving delays and its application to time-delayed optimal control problems*, J. Optim. Theory Appl., 27 (1979), pp. 357–376.
- [32] L. S. PONTRYAGIN, V. G. BOLTYANSKII, R. V. GAMKRELIDZE, AND E. F. MISHCHENKO, *The Mathematical Theory of Optimal Processes*, Wiley, New York, 1962.
- [33] L. RIFFORD AND E. TRÉLAT, *On the stabilization problem for nonholonomic distributions*, J. Eur. Math. Soc. (JEMS), 11 (2009), pp. 223–255.
- [34] F. A. RIHAN, *Sensitivity analysis for dynamic systems with time-lags*, J. Comput. Appl. Math., 151 (2003), pp. 445–462.
- [35] F. RODRIGUES, C. J. SILVA, H. MAURER, AND D. F. M. TORRES, *Optimal control of a delayed HIV model*, Discrete Contin. Dyn. Syst. Ser. B, 23 (2018), pp. 443–458.
- [36] C. SILVA AND E. TRÉLAT, *Smooth regularization of bang-bang optimal control problems*, IEEE Trans. Automat. Control, 55 (2010), pp. 2488–2499.
- [37] C. J. SILVA, H. MAURER, AND D. F. M. TORRES, *Optimal control of a Tuberculosis model with state and control delays*, Math. Biosci. Eng., 14 (2017), pp. 321–337.
- [38] M. A. SOLIMAN AND W. H. RAY, *On the optimal control of systems having pure time delays and singular arcs*, Internat. J. Control, 16 (1972), pp. 963–976.
- [39] E. TRÉLAT, *Contrôle optimal: Théorie & applications*, Vuibert, Paris, 2005.
- [40] E. TRÉLAT, *Optimal control and applications to aerospace: Some results and challenges*, J. Optim. Theory Appl., 154 (2012), pp. 713–758.
- [41] A. VISINTIN, *Strong convergence results related to strict convexity*, Comm. Partial Differential Equation, 9 (1984), pp. 439–466.