



**HAL**  
open science

# Pontryagin maximum principle for state constrained optimal sampled-data control problems on time scales

Piernicola Bettiol, Loïc Bourdin

► **To cite this version:**

Piernicola Bettiol, Loïc Bourdin. Pontryagin maximum principle for state constrained optimal sampled-data control problems on time scales. *ESAIM: Control, Optimisation and Calculus of Variations*, 2021, 27, pp.51. 10.1051/cocv/2021046 . hal-03250790

**HAL Id: hal-03250790**

**<https://hal.science/hal-03250790>**

Submitted on 4 Jun 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# PONTRYAGIN MAXIMUM PRINCIPLE FOR STATE CONSTRAINED OPTIMAL SAMPLED-DATA CONTROL PROBLEMS ON TIME SCALES

PIERNICOLA BETTIOL<sup>1,\*</sup> AND LOÏC BOURDIN<sup>2</sup>

**Abstract.** In this paper we consider optimal sampled-data control problems on time scales with inequality state constraints. A Pontryagin maximum principle is established, extending to the state constrained case existing results in the time scale literature. The proof is based on the Ekeland variational principle and on the concept of *implicit spike variations* adapted to the time scale setting. The main result is then applied to continuous-time min-max optimal sampled-data control problems, and a maximal velocity minimization problem for the harmonic oscillator with sampled-data control is numerically solved for illustration.

**Mathematics Subject Classification.** 26E70; 34H05; 34K35; 34N05; 39A12; 49J15; 49J50; 93C10; 93C15; 93C57.

Received March 14, 2020. Accepted April 22, 2021.

## 1. INTRODUCTION

### 1.1. Optimal control problems

In mathematics a *dynamical system* describes the evolution of a point (usually called the *state* of the system) in an appropriate space (called the *state space*) following an evolution rule (called the *dynamics* of the system). Dynamical systems are of many different natures (continuous-time versus discrete-time systems, deterministic versus stochastic systems, etc.). A *continuous-time system* is a dynamical system in which the state evolves in a continuous way in time (for instance, ordinary differential equations, evolution partial differential equations, etc.), while a *discrete-time system* is a dynamical system in which the state evolves in a discrete way in time (for instance, difference equations, quantum differential equations, etc.). A *control system* is a dynamical system in which a control parameter intervenes in the dynamics and thus influences the evolution of the state. Finally an *optimal control problem* consists in determining a control which allows to steer the state of a control system from an initial condition to some desired target while minimizing a given cost and satisfying some constraints.

---

*Keywords and phrases:* Optimal control, Pontryagin maximum principle, sampled-data control, time scales, state constraints, Ekeland variational principle, implicit spike variations, min-max optimal control problems.

<sup>1</sup> Univ Brest, UMR CNRS 6205, Laboratoire de Mathématiques de Bretagne Atlantique, 29200 Brest, France.

<sup>2</sup> Université de Limoges, UMR CNRS 7252, Institut de recherche XLIM, 87060, France.

\* Corresponding author: [piernicola.bettiol@univ-brest.fr](mailto:piernicola.bettiol@univ-brest.fr)

## 1.2. Pontryagin maximum principle

Established in [61] by Pontryagin, Boltyanskii, Gamkrelidze and Mischenko at the end of the 1950s, the Pontryagin Maximum Principle (in short, PMP) is a fundamental result in optimal control theory with numerous theoretical and numerical applications. We refer to [24, 52, 66, 69, 71] and references therein. The classical PMP gives first-order necessary optimality conditions for continuous-time optimal control problems in which the dynamical system is described by a general ordinary differential equation. Roughly speaking, the PMP ensures the existence of a *costate vector* (also called *adjoint vector*) satisfying some terminal conditions (called *transversality conditions*) so that the optimal control maximizes the *Hamiltonian* associated with the optimal control problem. Soon afterwards and even nowadays, the PMP has been adapted to many situations, for control systems of different natures, with various constraints, etc. It is not the aim of the present paper to give a full list of all existing different versions of necessary optimality conditions. Nevertheless we observe that several versions of the PMP were derived for discrete-time optimal control problems in which the dynamical system is described by a general difference equation (see, *e.g.*, [12, 48]). In these discrete-time versions of the PMP, the Hamiltonian maximization condition does not hold in general (see a counterexample in Examples 10.1-10.4 p. 59-62 of [12]) and has to be replaced by a weaker condition known as the *nonpositive Hamiltonian gradient condition* (see, *e.g.*, [12], Thm. 42.1, p. 330).

## 1.3. State constrained optimal control problems

An important generalization of the PMP concerns the possibility to take account of state constraints, imposing the state variable to remain in a given region of the state space. Indeed it is often undesirable and even inadmissible in applications that the trajectories violate some constraints imposed in the state space for safety or practical reasons. Many applications can be found in mechanics and aerospace engineering, management and economics, etc. We refer to [14, 29, 56, 63, 70] and references therein for examples. A version of the PMP for state constrained continuous-time optimal control problems was given by Gamkrelidze *et al.* (see, *e.g.*, [38] and [61, Theorem 25 p. 311]). Other versions can be found in [31–33, 40, 55]. A comprehensive survey [47] on this field of research was published in 1995 by Hartl, Sethi and Vickson (see also the monograph [71] and the references therein for general results in the nonsmooth setting). Note that the PMP for continuous-time optimal control problems is more intricate in the presence of state constraints because the costate vector is not absolutely continuous in general (as in the state constraint-free case), but (only) of bounded variation. Therefore theoretical and numerical difficulties may arise due to the possible pathological behavior of the costate vector which consists in discontinuity jumps and singular components lying on parts of the trajectory in contact with the boundary of the restricted state space. As a consequence a wide portion of the literature is devoted to the analysis of the costate vector and some qualification conditions have been established in order to ensure that it has no singular component (see, *e.g.*, [14, 47, 50, 55]). Note that the related theme of state constrained discrete-time optimal control problems has also been investigated in the literature (see, *e.g.*, [30, 62]).

## 1.4. Continuous-time optimal sampled-data control problems

In the literature the control of a continuous-time system is very often assumed to be *permanent*, in the sense that its value is authorized to be modified at any time. As a consequence, in many corresponding problems, achieving the optimal trajectory might require a permanent modification of the control. However such a requirement could be difficult to apply in practice, even employing mechanical or numerical devices. Therefore a *sampled-data control*, for which the value is authorized to be modified only for a finite number of times (called *sampling times*) and remains frozen otherwise (leading to a piecewise constant control), is usually considered for engineering models. A vast literature deals with continuous-time sampled-data control systems, as evidenced by numerous references and books (see, *e.g.*, [1, 6, 27, 39, 51, 60, 68] and references therein). This field has significantly grown since the 1970's, motivated by the electrical and mechanical engineering applications, such as strings of vehicles (see [4, 53, 57, 58, 65]). Note that continuous-time sampled-data control systems have the peculiarity of presenting a mixed continuous/discrete structure. In 2015 Bourdin and Trélat have provided

in [22] a version of the PMP for continuous-time optimal sampled-data control problems. In that context, as in the PMP for discrete-time optimal permanent control problems, the usual Hamiltonian maximization condition does not hold in general and has to be replaced by the condition known as the *nonpositive averaged Hamiltonian gradient condition* (see [22], Thm. 2.6). More recently Bourdin and Dhar have extended in [18] the previous work to the state constrained case. In that context the authors have observed and studied a particular behavior of the trajectories with respect to the state constraints, called *bouncing trajectory phenomenon*. Precisely, under some quite general hypotheses, the authors prove that an admissible trajectory necessarily “bounces” against the boundary of the restricted state space at most a finite number of times. Inherent to this behavior, the singular component of the costate vector vanishes and its discontinuities are reduced to a finite number of jumps, which turns out very useful for numerical simulations based on shooting methods. We refer to Sections 4 and 5 of [18] for details and discussion. We conclude this paragraph by mentioning the work [17] in which the optimization of the sampling times (in the state constraint-free case) has also been explored in view of a PMP formulation. In that context, it has been proved that the necessary condition for optimal sampling times coincides with the continuity of the maximized Hamiltonian function. This necessary condition turns out to be efficient in order to compute numerically the optimal sampling times using shooting methods. For example this approach has been employed in [5] in order to optimize instances of functional electrical stimulations of muscles in human force-fatigue muscular models.

### 1.5. Optimal (permanent) control problems on time scales

The *time scale theory* was initiated in 1988 by Hilger [42] in order to unify continuous and discrete analyses. By definition, a time scale  $\mathbb{T}$  is an arbitrary nonempty closed subset of  $\mathbb{R}$ , and a dynamical system is said to be posed on the time scale  $\mathbb{T}$  whenever the time variable evolves along this set  $\mathbb{T}$ . For example a continuous-time dynamical system corresponds to  $\mathbb{T} = \mathbb{R}_+$ , while  $\mathbb{T} = \mathbb{N}$  is associated with a discrete-time dynamical system. But a time scale can be much more general (it can be even a Cantor set). Many notions of standard calculus (such as derivatives, integrals, etc.) have been extended to the time scale framework, and we refer the reader to [3, 9, 10] for details on this theory. We also refer to Section 2 of the present paper for some basics. The Cauchy-Lipschitz (or Picard-Lindelöf) theory has been extended in [19] to ordinary differential equations posed on general time scales. For  $\mathbb{T} = \mathbb{R}_+$  for example, one recovers the classical theory of (continuous-time) ordinary differential equations, while, for  $\mathbb{T} = \mathbb{N}$  for example, one recovers the theory of difference equations. This provides an illustration that the time scale theory allows to close the gap between continuous and discrete analyses, and this is possible in any mathematical domain in which time scale calculus can be involved. Another example is provided in optimization with the calculus of variations on time scales, initiated in [8], and well-studied in the literature (see, *e.g.*, [7, 15, 43, 46]). On the other hand, in [44, 45], the authors establish a weak version of the PMP (with the nonpositive Hamiltonian gradient condition) for (permanent) control systems defined on general time scales. In [20], Bourdin and Trélat derived a strong version of the PMP in the same setting, proving the Hamiltonian maximization condition at right-dense points of the time scale, and the nonpositive Hamiltonian gradient condition at right-scattered points of the time scale (see the beginning of Section 2 for the precise definitions of right-dense and right-scattered points of a time scale). This result thus encompasses the classical version of the PMP for continuous-time optimal (permanent) control problems (taking  $\mathbb{T} = \mathbb{R}_+$  for example), and also the one for discrete-time optimal (permanent) control problems (taking  $\mathbb{T} = \mathbb{N}$  for example). Furthermore the work [20] emphasizes the reasons why the Hamiltonian condition is different at right-dense points from the one at right-scattered points: in a right-dense point,  $L^1$ -needle-like variations of the control are possible, while, in a right-scattered point, (only)  $L^\infty$ -needle-like variations are possible. Nevertheless there is a price to pay in order to derive a PMP on a general time scale: as explained in Section 3.1 of [20], some standard approaches (based on fixed-point theorems or Hahn-Banach separation arguments essentially) fail due to the lack of convexity of a general time scale in the neighborhoods of its right-dense points. For example another time scale version of the PMP has been provided recently in Theorem 2.11 of [11] using necessary conditions of an extreme in a cone, but this approach requires a density condition on the time scale. In view of keeping a general time scale framework, the authors of [20] use the Ekeland variational principle ([34], Thm. 1.1) which

turns out to be suitable in order to prove a time scale version of the PMP with no assumption on the time scale. We refer to [23] for a detailed discussion on the two papers [11, 20].

## 1.6. Optimal sampled-data control problems on time scales

The papers [11, 20, 23, 44, 45] mentioned in the previous paragraph are concerned (only) with control systems defined on general time scales with permanent control. In the work [21], Bourdin and Trélat have introduced a new framework allowing to handle control systems defined on general time scales with nonpermanent control, referred to as *sampled-data control systems on time scales*. In that context it is assumed that the state and the control are allowed to evolve on different time scales, respectively denoted by  $\mathbb{T}$  and  $\mathbb{T}_1$  (the time scale  $\mathbb{T}_1$  of the control being a subset of the time scale  $\mathbb{T}$  of the state, that is,  $\mathbb{T}_1 \subset \mathbb{T}$ ). This framework is the natural extension of the classical continuous-time and sampled-data control setting, displayed above, considering  $\mathbb{T} = \mathbb{R}_+$  and  $\mathbb{T}_1 = \mathbb{N}$  for example. A PMP for optimal sampled-data control problems on time scales is proved in Theorem 2.6 of [21] based on the Ekeland variational principle (as in [20]) and on  $L^1$ -needle-like variations (resp.  $L^\infty$ -needle-like variations) of the control at right-dense points (resp. right-scattered points) of the time scale  $\mathbb{T}_1$ . This leads to a Hamiltonian maximization condition at right-dense points of  $\mathbb{T}_1$  and to a nonpositive *averaged* Hamiltonian gradient condition at right-scattered points of  $\mathbb{T}_1$ . In particular, the PMP in [21] encompasses the one in [20] (and thus the classical ones for continuous-time and discrete-time problems with permanent controls), but also the one established in [22] for continuous-time optimal sampled-data control problems. We refer to Sections 2.3 and 3.2 of [21] for a discussion and numerous remarks about the differences between optimal permanent controls and optimal sampled-data controls.

## 1.7. Contributions of the present paper

In this paper we consider optimal sampled-data control problems on time scales with *inequality state constraints*. A Pontryagin maximum principle is established (see Thm. 3.7 in Sect. 3), extending to the state constrained case the existing results provided in the time scale literature [11, 20, 21, 44, 45]. Our proof is based on the Ekeland variational principle which is, as mentioned above and in contrast to some other approaches, suitable in order to deal with general time scale versions of the PMP. Nevertheless the techniques employed in this paper differ from those used in [20, 21] in several aspects. First, in order to take into account the presence of state constraints, we include an additional term in the Ekeland penalized functional which eventually gives rise to Borel measures on  $\mathbb{T}$  and, as in the PMPs for continuous-time optimal permanent and sampled-data control problems, to a bounded variation costate vector. Second, the needle-like variations of the control used in [20, 21] are not suitable in order to handle inequality state constraints such as the ones considered in the present work. Indeed the needle-like variations of the control provide a differentiability (with respect to the perturbation parameter) of the state (only) over a subinterval of  $\mathbb{T}$ , while the analysis of the inequality state constraints requires differentiability over the whole interval of  $\mathbb{T}$ . We refer to Remark 3.14 for details. In order to overcome this technical difficulty, our idea in this paper is to involve the sensitivity analysis of the control system under *implicit spike variations* of the control. This concept is used in [13, 16, 54] for continuous-time optimal permanent control problems and is based in particular on the fact that the usual Lebesgue measure is nonatomic. However, the adaptation of this concept to the time scale setting is not trivial because the measure on the time scale  $\mathbb{T}_1$  is atomic (since the measure of any right-scattered point of  $\mathbb{T}_1$  is positive, see Sect. 2 for details). Therefore, in the present work, we introduce a new and suitable time scale version of the concept of implicit spike variations, by distinguishing the perturbation at right-dense points and at right-scattered points of  $\mathbb{T}_1$ . In particular, at right-scattered points of  $\mathbb{T}_1$ , we involve a  $L^\infty$ -needle-like variation of the control as in [21]. We refer to Remark 3.14 and Section 5.1.2 for more details. The new tools introduced in this paper, and mentioned in this paragraph, turn out to be suitable and efficient in order to prove our main result (Thm. 3.7). To conclude the contributions of this paper, and as illustration, we consider in Section 4.1 a general continuous-time min-max optimal sampled-data control problems which can be, using a well known idea (developed for example in [31], Rem. 6 or in [71], Prop. 9.5.4), reformulated as a state constrained continuous-time optimal sampled-data control problem on which Theorem 3.7 can be applied. Finally a maximal velocity minimization

problem for the harmonic oscillator with sampled-data control is numerically solved in Section 4.2 using a shooting method based on the necessary conditions provided in Theorem 3.7. As in the recent work [18], a bouncing trajectory phenomenon is observed (see Figs. 1 and 2).

## 1.8. Organization of the paper

The paper is organized as follows. In Section 2 we display some basic notions and results in time scale theory which will be used all along the paper. In Section 3 we introduce the sampling procedure and the general optimal sampled-data control problem on time scales considered in the present work. We provide the invoked regularity and topology assumptions and then we state our major contribution (Thm. 3.7). Section 4 is dedicated to the application of our main result to continuous-time min-max optimal sampled-data control problems. The proof of Theorem 3.7 is built up in several stages and is displayed in Section 6, after collecting some crucial preliminary results in Section 5.

## 2. BASICS ON TIME SCALE THEORY

We start recalling some basic definitions and results employed in time scale theory. The reader already familiar with this topic can skip this section and proceed directly to Section 3. Let  $\mathbb{T}$  be a *time scale*, that is, an arbitrary nonempty closed subset of  $\mathbb{R}$ . Without loss of generality, we will assume that  $\mathbb{T}$  is bounded below, denoting by  $a := \min \mathbb{T}$ , and unbounded above.<sup>1</sup> Throughout the paper,  $\mathbb{T}$  will be the time scale on which the state of the control system evolves. The *forward jump operator*  $\sigma : \mathbb{T} \rightarrow \mathbb{T}$  is defined by  $\sigma(t) := \inf\{\tau \in \mathbb{T} \mid \tau > t\}$  for every  $t \in \mathbb{T}$ . A point  $t \in \mathbb{T}$  is said to be *right-scattered* whenever  $\sigma(t) > t$ . A point  $t \in \mathbb{T}$  is said to be *right-dense* whenever  $\sigma(t) = t$ . We denote by RS the set of all right-scattered points of  $\mathbb{T}$ , and by RD the set of all right-dense points of  $\mathbb{T}$ . Recall that RS is at most countable (see [26], Lem. 3.1) and that RD is the complement of RS in  $\mathbb{T}$ . The *graininess function*  $\mu : \mathbb{T} \rightarrow \mathbb{R}_+$  is defined by  $\mu(t) := \sigma(t) - t$  for every  $t \in \mathbb{T}$ . For every subset  $A$  of  $\mathbb{R}$ , we write  $A_{\mathbb{T}} := A \cap \mathbb{T}$ . An *interval* of  $\mathbb{T}$  is a set  $I_{\mathbb{T}}$  where  $I$  is an interval of  $\mathbb{R}$ .

### 2.1. $\Delta$ -differentiability

Let  $n \in \mathbb{N}^*$ . A function  $x : \mathbb{T} \rightarrow \mathbb{R}^n$  is said to be  $\Delta$ -differentiable at  $t \in \mathbb{T}$  if the limit

$$x^{\Delta}(t) := \lim_{\substack{\tau \rightarrow t \\ \tau \in \mathbb{T}}} \frac{x^{\sigma}(\tau) - x(\tau)}{\sigma(\tau) - \tau},$$

where  $x^{\sigma} := x \circ \sigma$ , exists in  $\mathbb{R}^n$ . Recall that, if  $s \in \text{RD}$ , then  $x$  is  $\Delta$ -differentiable at  $s$  if and only if the limit of  $\frac{x(s) - x(\tau)}{s - \tau}$ , as  $\tau \rightarrow s$  with  $\tau \in \mathbb{T}$ , exists in  $\mathbb{R}^n$ . In that case it is equal to  $x^{\Delta}(s)$ . If  $r \in \text{RS}$  and  $x$  is continuous at  $r$ , then  $x$  is  $\Delta$ -differentiable at  $r$  with  $x^{\Delta}(r) = \frac{x^{\sigma}(r) - x(r)}{\mu(r)}$  (see, e.g., [9], Thm. 1.16).

If two functions  $x, x' : \mathbb{T} \rightarrow \mathbb{R}^n$  are both  $\Delta$ -differentiable at  $t \in \mathbb{T}$ , then the scalar product  $\langle x, x' \rangle_{\mathbb{R}^n} : \mathbb{T} \rightarrow \mathbb{R}$  is  $\Delta$ -differentiable at  $t$  with

$$\langle x, x' \rangle_{\mathbb{R}^n}^{\Delta}(t) = \langle x^{\Delta}(t), x'^{\sigma}(t) \rangle_{\mathbb{R}^n} + \langle x(t), x'^{\Delta}(t) \rangle_{\mathbb{R}^n} = \langle x^{\Delta}(t), x'(t) \rangle_{\mathbb{R}^n} + \langle x^{\sigma}(t), x'^{\Delta}(t) \rangle_{\mathbb{R}^n}.$$

These equalities are usually called *Leibniz formulas* (see, e.g., [9], Thm. 1.20).

<sup>1</sup>Indeed, in this paper, we will only work on a bounded subinterval of type  $[a, b] \cap \mathbb{T}$  with  $a, b \in \mathbb{T}$  and  $a < b$ . It is not restrictive to assume that  $a = \min \mathbb{T}$  and that  $\mathbb{T}$  is unbounded above. On the other hand, this last assumption will allow us to simplify the notation introduced in this section, avoiding a systematic distinction between points of  $\mathbb{T} \setminus \{\max \mathbb{T}\}$  and  $\max \mathbb{T}$ , which is not necessary in our context.

## 2.2. Lebesgue $\Delta$ -measure and Lebesgue $\Delta$ -integrability

Let  $\mu_\Delta$  be the *Lebesgue  $\Delta$ -measure* on  $\mathbb{T}$  defined in terms of Carathéodory extension (see [10], Chap. 5). We also refer the reader to [2, 26, 41] for more details. For all  $(c, d) \in \mathbb{T}^2$  such that  $c \leq d$ , one has  $\mu_\Delta([c, d]_{\mathbb{T}}) = d - c$ . Recall that  $A \subset \mathbb{T}$  is a  $\mu_\Delta$ -measurable set of  $\mathbb{T}$  if and only if  $A$  is a  $\mu_{\mathcal{L}}$ -measurable set of  $\mathbb{R}$ , where  $\mu_{\mathcal{L}}$  denotes the usual Lebesgue measure on  $\mathbb{R}$  (see [26], Prop. 3.1), and we have

$$\mu_\Delta(A) = \mu_{\mathcal{L}}(A) + \sum_{r \in A \cap \text{RS}} \mu(r).$$

Let  $A \subset \mathbb{T}$  be a  $\mu_\Delta$ -measurable subset of  $\mathbb{T}$ . A property is said to hold  $\Delta$ -almost everywhere (in short,  $\Delta$ -a.e.) on  $A$  if it holds for every  $t \in A \setminus A'$ , where  $A' \subset A$  is some  $\mu_\Delta$ -measurable set of  $\mathbb{T}$  satisfying  $\mu_\Delta(A') = 0$ . In particular, since  $\mu_\Delta(\{r\}) = \mu(r) > 0$  for every  $r \in \text{RS}$ , we conclude that, if a property holds  $\Delta$ -a.e. on  $A$ , then it holds for every  $r \in A \cap \text{RS}$ . Similarly, if  $\mu_\Delta(A) = 0$ , then  $A \subset \text{RD}$ .

Consider a function  $x$  defined  $\Delta$ -a.e. on  $A$  with values in  $\mathbb{R}^n$ . Let  $\tilde{A} := A \sqcup (\bigsqcup_{r \in A \cap \text{RS}} (r, \sigma(r)))$  and  $\tilde{x}$  be the extension of  $x$  defined  $\mu_{\mathcal{L}}$ -a.e. on  $\tilde{A}$  by  $\tilde{x}(t) := x(t)$  whenever  $t \in A$ , and by  $\tilde{x}(t) := x(r)$  whenever  $t \in (r, \sigma(r))$  for some  $r \in A \cap \text{RS}$ . Recall that  $x$  is  $\mu_\Delta$ -measurable on  $A$  if and only if  $\tilde{x}$  is  $\mu_{\mathcal{L}}$ -measurable on  $\tilde{A}$  (see [26], Prop. 4.1).

The functional space  $L^\infty_\Delta(A, \mathbb{R}^n)$  is the set of all functions  $x$  defined  $\Delta$ -a.e. on  $A$ , with values in  $\mathbb{R}^n$ , that are  $\mu_\Delta$ -measurable on  $A$  and bounded  $\Delta$ -almost everywhere. Endowed with the usual norm  $\|x\|_{L^\infty_\Delta(A, \mathbb{R}^n)} := \sup_{\tau \in A} \|x(\tau)\|_{\mathbb{R}^n}$ , it is a Banach space (see [2], Thm. 2.5). The functional space  $L^1_\Delta(A, \mathbb{R}^n)$  is the set of all functions  $x$  defined  $\Delta$ -a.e. on  $A$ , with values in  $\mathbb{R}^n$ , that are  $\mu_\Delta$ -measurable on  $A$  and such that  $\int_A \|x(\tau)\|_{\mathbb{R}^n} \Delta\tau < +\infty$ . Endowed with the usual norm  $\|x\|_{L^1_\Delta(A, \mathbb{R}^n)} := \int_A \|x(\tau)\|_{\mathbb{R}^n} \Delta\tau$ , it is a Banach space (see [2], Thm. 2.5). We recall here that if  $x \in L^1_\Delta(A, \mathbb{R}^n)$  then

$$\int_A x(\tau) \Delta\tau = \int_{\tilde{A}} \tilde{x}(\tau) d\tau = \int_A x(\tau) d\tau + \sum_{r \in A \cap \text{RS}} \mu(r)x(r),$$

see Theorems 5.1 and 5.2 of [26]. Note that if  $A$  is bounded then  $L^\infty_\Delta(A, \mathbb{R}^n) \subset L^1_\Delta(A, \mathbb{R}^n)$ .

## 2.3. Absolutely continuous functions

Take  $(c, d) \in \mathbb{T}^2$  such that  $c < d$ . Let  $C([c, d]_{\mathbb{T}}, \mathbb{R}^n)$  denote the space of continuous functions defined on  $[c, d]_{\mathbb{T}}$  with values in  $\mathbb{R}^n$ . Endowed with its usual uniform norm  $\|\cdot\|_\infty$ , it is a Banach space. Let  $\text{AC}([c, d]_{\mathbb{T}}, \mathbb{R}^n)$  denote the subspace of absolutely continuous functions.

Let  $t_0 \in [c, d]_{\mathbb{T}}$  and  $x : [c, d]_{\mathbb{T}} \rightarrow \mathbb{R}^n$ . It can be easily derived from Theorem 4.1 of [25] that  $x \in \text{AC}([c, d]_{\mathbb{T}}, \mathbb{R}^n)$  if and only if  $x$  is  $\Delta$ -differentiable  $\Delta$ -a.e. on  $[c, d]_{\mathbb{T}}$  with  $x^\Delta \in L^1_\Delta([c, d]_{\mathbb{T}}, \mathbb{R}^n)$  and for every  $t \in [c, d]_{\mathbb{T}}$  one has  $x(t) = x(t_0) + \int_{[t_0, t]_{\mathbb{T}}} x^\Delta(\tau) \Delta\tau$ , whenever  $t \geq t_0$ , and  $x(t) = x(t_0) - \int_{[t, t_0]_{\mathbb{T}}} x^\Delta(\tau) \Delta\tau$ , whenever  $t \leq t_0$ .

Take any  $x \in L^1_\Delta([c, d]_{\mathbb{T}}, \mathbb{R}^n)$  and  $t_0 \in [c, d]_{\mathbb{T}}$ . Let  $X$  be the function defined on  $[c, d]_{\mathbb{T}}$  by  $X(t) := \int_{[t_0, t]_{\mathbb{T}}} x(\tau) \Delta\tau$  whenever  $t \geq t_0$ , and by  $X(t) := -\int_{[t, t_0]_{\mathbb{T}}} x(\tau) \Delta\tau$  whenever  $t \leq t_0$ . Then  $X \in \text{AC}([c, d]_{\mathbb{T}}, \mathbb{R}^n)$  and  $X^\Delta(t) = x(t)$  for  $\Delta$ -a.e.  $t \in [c, d]_{\mathbb{T}}$ .

Observe that, if  $x \in \text{AC}([c, d]_{\mathbb{T}}, \mathbb{R}^n)$  is such that  $x^\Delta(t) = 0_{\mathbb{R}^n}$  for  $\Delta$ -a.e.  $t \in [c, d]_{\mathbb{T}}$ , then  $x$  is constant over  $[c, d]_{\mathbb{T}}$ , and that, if  $x, x' \in \text{AC}([c, d]_{\mathbb{T}}, \mathbb{R}^n)$ , then  $\langle x, x' \rangle_{\mathbb{R}^n} \in \text{AC}([c, d]_{\mathbb{T}}, \mathbb{R})$  and the above Leibniz formulas are valid  $\Delta$ -a.e. on  $[c, d]_{\mathbb{T}}$ .

## 2.4. Functions of bounded variation

We denote by  $BV([c, d]_{\mathbb{T}}, \mathbb{R}^n)$  the space of functions of bounded variation defined on  $[c, d]_{\mathbb{T}}$  taking values in  $\mathbb{R}^n$ , that is, the space of functions  $x : [c, d]_{\mathbb{T}} \rightarrow \mathbb{R}^n$  such that

$$\sup_{\{t_k\}_k} \sum_k \|x(t_{k+1}) - x(t_k)\|_{\mathbb{R}^n} < +\infty,$$

where the supremum is taken over all finite partitions  $\{t_k\}_k$  of  $[c, d]_{\mathbb{T}}$ . As in the classical continuous-time literature (taking  $\mathbb{T} = \mathbb{R}_+$  for example), it can be proved that both the inclusions

$$AC([c, d]_{\mathbb{T}}, \mathbb{R}^n) \subset BV([c, d]_{\mathbb{T}}, \mathbb{R}^n) \subset L_{\Delta}^{\infty}([c, d]_{\mathbb{T}}, \mathbb{R}^n),$$

hold, and that a function  $x : [c, d]_{\mathbb{T}} \rightarrow \mathbb{R}^n$  is of bounded variation if and only if  $x$  can be written as the difference  $x' - x''$  of two functions  $x', x'' : [c, d]_{\mathbb{T}} \rightarrow \mathbb{R}^n$  with nondecreasing real components. We refer to Proposition 4 of [49] for details. One can easily deduce from the previous characterization that, if  $x \in BV([c, d]_{\mathbb{T}}, \mathbb{R}^n)$ , then the function  $y : [c, d]_{\mathbb{T}} \rightarrow \mathbb{R}^n$ , defined by  $y(t) := x^{\sigma}(t)$  for all  $t \in [c, d]_{\mathbb{T}}$  and  $y(d) := x(d)$ , belongs to  $BV([c, d]_{\mathbb{T}}, \mathbb{R}^n)$ . We conclude this section observing that if  $d\eta$  is a finite nonnegative Borel measure on  $[c, d]_{\mathbb{T}}$  and  $x \in C([c, d]_{\mathbb{T}}, \mathbb{R}^n)$ , then the function  $z : [c, d]_{\mathbb{T}} \rightarrow \mathbb{R}^n$  defined by

$$z(t) := \int_{[c, t]_{\mathbb{T}}} x(\tau) d\eta(\tau),$$

for all  $t \in [c, d]_{\mathbb{T}}$ , belongs to  $BV([c, d]_{\mathbb{T}}, \mathbb{R}^n)$ .

## 3. MAIN RESULT AND COMMENTS

This section is dedicated to the statement of our main result (Thm. 3.7). In Section 3.1 we first give some reminders on a *sampling procedure* on time scales extracted from p. 60 of [21]. In Section 3.2 we introduce the general state constrained optimal sampled-data control problem on time scales considered in the present work, and we fix the terminology and assumptions used all along the paper. In Section 3.3 we state the corresponding Pontryagin maximum principle and a list of comments follows.

### 3.1. Sampling procedure

Let  $\mathbb{T}_1$  be a second time scale, possibly different from the reference one  $\mathbb{T}$  introduced in Section 2. Throughout the paper,  $\mathbb{T}_1$  will be the time scale on which the control of the control system evolves. We assume that  $\mathbb{T}_1 \subset \mathbb{T}$ .<sup>2</sup> As for  $\mathbb{T}$ , we assume that  $\min \mathbb{T}_1 = a$  and that  $\mathbb{T}_1$  is unbounded above. In accordance with the previous section, we use the notation  $\sigma_1$ ,  $RS_1$ ,  $RD_1$ ,  $\Delta_1$ , etc., for the analytical tools relative to the time scale  $\mathbb{T}_1$ . Since  $\mathbb{T}_1 \subset \mathbb{T}$ , we have  $RS \cap \mathbb{T}_1 \subset RS_1$  and  $RD_1 \subset RD$ .

A *sample-and-hold procedure* from  $\mathbb{T}_1$  to  $\mathbb{T}$  involves defining an operator that extends to  $\mathbb{T}$  any function defined on  $\mathbb{T}_1$ , by freezing the values on  $\mathbb{T} \setminus \mathbb{T}_1$  in the sense given by Definition 3.1 below. In order to introduce this sampling procedure, we define the map

$$\begin{aligned} \square : \mathbb{T} &\longrightarrow \mathbb{T}_1 \\ t &\longmapsto \square(t) := \sup\{\tau \in \mathbb{T}_1 \mid \tau \leq t\}. \end{aligned}$$

For every  $t \in \mathbb{T}_1$ , we have  $\square(t) = t$ . For every  $t \in \mathbb{T} \setminus \mathbb{T}_1$ , we have  $\square(t) \in RS_1$  and  $\square(t) < t < \sigma_1(\square(t))$ .

<sup>2</sup>Indeed, it is not natural to consider *controlling times*  $t \in \mathbb{T}_1$  at which the dynamics does not evolve, that is, at which  $t \notin \mathbb{T}$ . The value of the control at such times  $t \in \mathbb{T}_1 \setminus \mathbb{T}$  would not influence the dynamics, or, maybe, only on  $[t^*, +\infty)_{\mathbb{T}}$  where  $t^* := \inf\{\tau \in \mathbb{T} \mid \tau \geq t\}$ . In this last case, note that  $t^* \in \mathbb{T}$  and we can replace  $\mathbb{T}_1$  by  $(\mathbb{T}_1 \cup \{t^*\}) \setminus \{t\}$  without loss of generality.



**Definition 3.1** (Sampling procedure). Let  $m \in \mathbb{N}^*$  and let  $u : \mathbb{T}_1 \rightarrow \mathbb{R}^m$  be a given function. In this paper the *sampled-data function* associated with  $u$  is the function  $u^\square : \mathbb{T} \rightarrow \mathbb{R}^m$  defined by the composition  $u^\square := u \circ \square$ .

**Example 3.2.** Let  $m \in \mathbb{N}^*$  and consider  $\mathbb{T} = \mathbb{R}_+$  and  $\mathbb{T}_1 = \mathbb{N}$ . If  $u : \mathbb{N} \rightarrow \mathbb{R}^m$  is a given function, then the corresponding sampled-data function  $u^\square : \mathbb{R}_+ \rightarrow \mathbb{R}^m$  is the piecewise constant function given by  $u^\square(t) = u(k)$  for all  $t \in [k, k+1)$  and all  $k \in \mathbb{N}$ .

We conclude this section with the following useful proposition that can be found in Proposition 1 of [21].

**Proposition 3.3.** Let  $m \in \mathbb{N}^*$  and let  $c < d$  be two elements of  $\mathbb{T}$  with  $c \in \mathbb{T}_1$ .

- (i) For every  $u \in L_{\Delta_1}^1([c, d]_{\mathbb{T}_1}, \mathbb{R}^m)$ , we have  $u^\square \in L_{\Delta}^1([c, d]_{\mathbb{T}}, \mathbb{R}^m)$  with  $\|u^\square\|_{L_{\Delta}^1([c, d]_{\mathbb{T}}, \mathbb{R}^m)} \leq \|u\|_{L_{\Delta_1}^1([c, d]_{\mathbb{T}_1}, \mathbb{R}^m)}$ .
- (ii) For every  $u \in L_{\Delta_1}^\infty([c, d]_{\mathbb{T}_1}, \mathbb{R}^m)$ , we have  $u^\square \in L_{\Delta}^\infty([c, d]_{\mathbb{T}}, \mathbb{R}^m)$  with  $\|u^\square\|_{L_{\Delta}^\infty([c, d]_{\mathbb{T}}, \mathbb{R}^m)} = \|u\|_{L_{\Delta_1}^\infty([c, d]_{\mathbb{T}_1}, \mathbb{R}^m)}$ .

### 3.2. A general state constrained optimal sampled-data control problem on time scales

Let  $\mathbb{T}_1 \subset \mathbb{T}$  be the two (possibly different) time scales introduced in Sections 2 and 3.1 (both unbounded above and both bounded below with  $a := \min \mathbb{T} = \min \mathbb{T}_1$ ). Let  $b \in \mathbb{T}$  be such that  $a < b$  and let  $m, n, j$  and  $\ell \in \mathbb{N}^*$  be four fixed positive integers. In this paper we focus on the general state constrained optimal sampled-data control problem on time scales given by

$$\left. \begin{array}{l} \text{minimize} \quad g(x(a), x(b)), \\ \text{subject to} \quad x \in \text{AC}([a, b]_{\mathbb{T}}, \mathbb{R}^n), \quad u \in L_{\Delta_1}^\infty([a, b]_{\mathbb{T}_1}, \mathbb{R}^m), \\ \quad \quad \quad x^\Delta(t) = f(x(t), u^\square(t), t), \quad \Delta\text{-a.e. } t \in [a, b]_{\mathbb{T}}, \\ \quad \quad \quad \psi(x(a), x(b)) \in S, \\ \quad \quad \quad h_i(x(t), t) \leq 0, \quad \forall t \in [a, b]_{\mathbb{T}}, \quad \forall i = 1, \dots, j, \\ \quad \quad \quad u(t) \in U, \quad \Delta_1\text{-a.e. } t \in [a, b]_{\mathbb{T}_1}, \end{array} \right\} \quad (\text{P})$$

where  $g : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$ ,  $f : \mathbb{R}^n \times \mathbb{R}^m \times [a, b]_{\mathbb{T}} \rightarrow \mathbb{R}^n$ ,  $\psi : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}^\ell$ , and  $h = (h_i)_{i=1, \dots, j} : \mathbb{R}^n \times [a, b]_{\mathbb{T}} \rightarrow \mathbb{R}^j$  are given functions, and where  $U \subset \mathbb{R}^m$  and  $S \subset \mathbb{R}^\ell$  are given sets.

A couple  $(x, u) \in \text{AC}([a, b]_{\mathbb{T}}, \mathbb{R}^n) \times L_{\Delta_1}^\infty([a, b]_{\mathbb{T}_1}, \mathbb{R}^m)$  is said to be *admissible* for Problem (P) if it satisfies all its constraints. A *solution* to Problem (P) is an admissible couple  $(x, u)$  which minimizes the cost  $g(x(a), x(b))$  among all admissible couples. In Problem (P),  $x$  is called the *state function* (also called the *trajectory*) and  $u$  is called the *control function*.

In the case where  $\mathbb{T}_1 = \mathbb{T}$ , the control is said to be *permanent* in Problem (P) because its value in the dynamical system can be modified at any time  $t \in \mathbb{T}$ . Otherwise, in the case where  $\mathbb{T}_1 \subsetneq \mathbb{T}$ , the control is said to be *sampled-data* in Problem (P) and its value in the dynamical system can be modified only at times  $t \in \mathbb{T}_1$  and remains frozen elsewhere (see Sect. 3.1).

Throughout this paper we will make use of the following regularity and topology hypotheses:

- (H1) the *dynamics function*  $f : \mathbb{R}^n \times \mathbb{R}^m \times [a, b]_{\mathbb{T}} \rightarrow \mathbb{R}^n$ , which drives the *state equation*  $x^\Delta(t) = f(x(t), u^\square(t), t)$ , is continuous and of class  $C^1$  with respect to its first two variables;
- (H2) the set  $U \subset \mathbb{R}^m$ , which describes the *control constraint*  $u(t) \in U$ , is a nonempty closed convex subset of  $\mathbb{R}^m$ ;
- (H3) the function  $g : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$ , which describes the *Mayer cost*  $g(x(a), x(b))$ , is of class  $C^1$ ;
- (H4) the function  $\psi : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}^\ell$ , which describes the *terminal state constraint*  $\psi(x(a), x(b)) \in S$ , is of class  $C^1$ ;
- (H5) the set  $S \subset \mathbb{R}^\ell$ , involved in the terminal state constraint  $\psi(x(a), x(b)) \in S$ , is a nonempty closed convex subset of  $\mathbb{R}^\ell$ ;

(H6) the function  $h = (h_i)_{i=1,\dots,j} : \mathbb{R}^n \times [a, b]_{\mathbb{T}} \rightarrow \mathbb{R}^j$ , which describes the *inequality state constraints*  $h_i(x(t), t) \leq 0$ , is continuous and of class  $C^1$  in its first variable.

**Remark 3.4.** The general time scale framework considered in the formulation of Problem (P) allows to recover several typical situations, among which:

- continuous-time optimal permanent control problems (taking  $\mathbb{T} = \mathbb{T}_1 = \mathbb{R}_+$  for example);
- discrete-time optimal permanent control problems (taking  $\mathbb{T} = \mathbb{T}_1 = \mathbb{N}$  for example);
- general optimal permanent control problems on time scales (taking  $\mathbb{T}_1 = \mathbb{T}$  with  $\mathbb{T}$  general);
- continuous-time optimal sampled-data control problems (taking  $\mathbb{T} = \mathbb{R}_+$  and  $\mathbb{T}_1 = \mathbb{N}$  for example);
- discrete-time optimal sampled-data control problems (taking  $\mathbb{T} = \mathbb{N}$  and  $\mathbb{T}_1 = 2\mathbb{N}$  for example);
- general optimal sampled-data control problems on time scales (taking  $\mathbb{T}_1 \subsetneq \mathbb{T}$  both general).

Moreover the general terminal state constraint  $\psi(x(a), x(b)) \in S$  allows to cover various cases, among which: fixed or free initial condition (resp. final condition), equality/inequality constraints on the initial condition (resp. final condition), mixed initial/final condition such as the periodic condition  $x(a) = x(b)$  for example, etc. We refer to Remark 10 of [21] for more details. Finally, by considering  $j = 1$  and  $h \equiv -1$ , note that the formulation of Problem (P) also allows to cover the state constraint-free case.

**Remark 3.5.** Our objective in the present work is to establish necessary optimality conditions for Problem (P). Regarding existence results, we refer to Theorem 2.1 of [21] in which, under some appropriate compactness and convexity assumptions, a Filippov-type existence result has been obtained in a very similar time scale and sampled-data control setting. However note that it has been established in the state constraint-free case. Nevertheless, to the best of our knowledge, the techniques can be adapted to the present context (that is, with inequality state constraints) and a similar result can be derived provided that Problem (P) is *feasible* (in the sense that there exists at least one admissible couple).

**Remark 3.6.** In order to deal with sampled-data controls in the general time scale setting, a possible alternative approach would be to consider controls  $u : \mathbb{T} \rightarrow \mathbb{R}^m$ , defined on the time scale  $\mathbb{T}$  (and not on  $\mathbb{T}_1$ ), by adding the constraint that they are constant on  $[r, \sigma_1(r))_{\mathbb{T}}$  for all right-scattered points  $r$  of  $\mathbb{T}_1$ . Nevertheless, in the general optimization problem (P), note that only the values of  $u$  taken over  $\mathbb{T}_1$  are of interest. As a consequence, from an optimization point of view, it is more natural and suitable to work with the Lebesgue functional space  $L_{\Delta_1}^{\infty}([a, b]_{\mathbb{T}_1}, \mathbb{R}^m)$ , associated with the time scale measure  $\mu_{\Delta_1}$  of  $\mathbb{T}_1$ , which is a well-studied Banach space in the literature (see, e.g., [2, 26]) in order to enjoy its known mathematical properties all along the paper, in particular in order to define implicit spike perturbations of the controls with respect to the  $L_{\Delta_1}^1$ -norm (see Sect. 5.1.2). On the contrary, with the above mentioned alternative framework, one would be led to introduce a new set of admissible controls (that are defined on  $\mathbb{T}$  and constant over  $[r, \sigma_1(r))_{\mathbb{T}}$  for all right-scattered points  $r$  of  $\mathbb{T}_1$ ) included in the functional space  $L_{\Delta}^{\infty}([a, b]_{\mathbb{T}}, \mathbb{R}^m)$ , associated with the time scale measure  $\mu_{\Delta}$  of  $\mathbb{T}$ , which would be not well-suited in order to provide a good description of the situation, in particular of the implicit spike perturbations of the controls used in this paper.

### 3.3. Pontryagin maximum principle and comments

Before providing a Pontryagin maximum principle associated with Problem (P), we need to recall some basic notions employed in our statement. The *normal cone* to the closed convex set  $S$  at a point  $x \in S$  is the set defined by

$$N_S(x) := \{y \in \mathbb{R}^{\ell} \mid \forall x' \in S, \langle y, x' - x \rangle_{\mathbb{R}^{\ell}} \leq 0\}.$$

The map  $\psi : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}^{\ell}$  is said to be *submersive* at a point  $(x_a, x_b) \in \mathbb{R}^n \times \mathbb{R}^n$  if its differential at  $(x_a, x_b)$  is surjective, i.e. if the Jacobian matrix  $\nabla\psi(x_a, x_b) \in \mathbb{R}^{\ell \times 2n}$  has full rank. Finally the *Hamiltonian*  $H : \mathbb{R}^n \times$

$\mathbb{R}^m \times \mathbb{R}^n \times [a, b]_{\mathbb{T}} \rightarrow \mathbb{R}$  associated with Problem (P) is defined by

$$H(x, u, p, t) := \langle p, f(x, u, t) \rangle_{\mathbb{R}^n},$$

for all  $(x, u, p, t) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^n \times [a, b]_{\mathbb{T}}$ . We are now ready to state the main result of the paper.

**Theorem 3.7** (Pontryagin maximum principle). *Assume (H1)–(H6). If  $(x^*, u^*) \in \text{AC}([a, b]_{\mathbb{T}}, \mathbb{R}^n) \times L_{\Delta_1}^{\infty}([a, b]_{\mathbb{T}_1}, \mathbb{R}^m)$  is a solution to Problem (P) and  $\psi$  is submersive at  $(x^*(a), x^*(b))$ , then there exist  $\lambda \geq 0$ ,  $p \in \text{AC}([a, b]_{\mathbb{T}}, \mathbb{R}^n)$  and finite nonnegative Borel measures  $d\eta_1, \dots, d\eta_j$  on  $[a, b]_{\mathbb{T}}$  such that the following conditions are satisfied:*

- (i) **Nontriviality:**  $(\lambda, p, d\eta_1, \dots, d\eta_j) \neq 0$ ;
- (ii) **Adjoint equation:**

$$-p^{\Delta}(t) = \nabla_x H(x^*(t), u^{*\square}(t), q(t), t), \quad \Delta\text{-a.e. } t \in [a, b]_{\mathbb{T}};$$

- (iii) **Transversality condition:**

$$\begin{pmatrix} p(a) \\ -q(b) \end{pmatrix} = \lambda \nabla g(x^*(a), x^*(b)) + \nabla \psi(x^*(a), x^*(b))^{\top} \times \xi,$$

where  $\xi \in N_{\mathbb{S}}(\psi(x^*(a), x^*(b)))$ ;

- (iv) **Hamiltonian conditions:**

(iv-a) **Hamiltonian maximization condition at right-dense points:**

$$u^*(s) \in \underset{v \in \mathbb{U}}{\text{argmax}} H(x^*(s), v, q(s), s), \quad \Delta_1\text{-a.e. } s \in [a, b]_{\mathbb{T}_1} \cap \text{RD}_1;$$

(iv-b) **Nonpositive averaged Hamiltonian gradient condition at right-scattered points:**

$$\left\langle \int_{[r, \sigma_1(r)]_{\mathbb{T}}} \nabla_u H(x^*(\tau), u^*(r), q(\tau), \tau) \Delta\tau, v - u^*(r) \right\rangle_{\mathbb{R}^m} \leq 0,$$

for all  $v \in \mathbb{U}$  and all  $r \in [a, b]_{\mathbb{T}_1} \cap \text{RS}_1$ ;

- (v) **Complementary slackness condition:**

$$\text{supp}(d\eta_i) \subset \{t \in [a, b]_{\mathbb{T}} \mid h_i(x^*(t), t) = 0\}, \quad \forall i = 1, \dots, j,$$

where  $\text{supp}(d\eta_i)$  stands for the classical notion of support of the measure  $d\eta_i$ .

Here  $q \in \text{BV}([a, b]_{\mathbb{T}}, \mathbb{R}^n)$  is defined by

$$q(t) := \begin{cases} p^{\sigma}(t) + \sum_{i=1}^j \int_{[a, t]_{\mathbb{T}}} \nabla_x h_i(x^*(\tau), \tau) d\eta_i(\tau) & \text{if } t \in [a, b]_{\mathbb{T}}, \\ p(b) + \sum_{i=1}^j \int_{[a, b]_{\mathbb{T}}} \nabla_x h_i(x^*(\tau), \tau) d\eta_i(\tau) & \text{if } t = b. \end{cases}$$

*Proof.* The proof of Theorem 3.7 is built up in several stages and will be displayed in Section 6, after collecting some crucial preliminary results in Section 5.  $\square$

We comment below our result.

**Remark 3.8.** As is well known in optimal control theory, the nontrivial tuple  $(\lambda, p, d\eta_1, \dots, d\eta_j)$  of Theorem 3.7, which is a Lagrange multiplier, is defined up to a positive multiplicative scalar. It is said to be *normal* whenever  $\lambda > 0$ , and *abnormal* whenever  $\lambda = 0$ . In the normal case  $\lambda > 0$ , it is usual to normalize the Lagrange multiplier so that  $\lambda = 1$ .

**Remark 3.9.** As is well known in optimal sampled-data control theory on time scales (see, e.g., [20–22]), the classical Hamiltonian maximization condition does not hold true in general at right-scattered points of  $\mathbb{T}_1$ , in which it is replaced by a nonpositive averaged Hamiltonian gradient condition. Note that, in the context of Theorem 3.7 and under some additional appropriate convexity assumptions such as the one introduced by Holtzman and Halkin in [48], it should be possible to obtain the *averaged Hamiltonian maximization condition* given by

$$u^*(r) \in \operatorname{argmax}_{v \in U} \int_{[r, \sigma_1(r)]_{\mathbb{T}}} H(x^*(\tau), v, q(\tau), \tau) \Delta\tau,$$

for all  $r \in [a, b]_{\mathbb{T}_1} \cap \operatorname{RS}_1$ . Taking  $\mathbb{T} = \mathbb{T}_1 = \mathbb{N}$  for example, one would recover the work [48] in which the authors obtain the Hamiltonian maximization condition for discrete-time optimal permanent control problems.

**Remark 3.10.** Our approach to prove Theorem 3.7 (see Sect. 6) is based on the Ekeland variational principle ([34], Thm. 1.1) which is, in contrast to some other methods, suitable to deal with general time scale versions of the Pontryagin maximum principle (see [20], Sect. 3.1 and [23] for detailed discussions on that point). It requires the closedness of  $U$  in order to define the corresponding penalized functional on a complete metric set (see details in Sect. 6.1.1). The closure of  $U$  is thus a crucial assumption in our approach. However, note that it is possible to slightly extend Theorem 3.7 to the case where  $U$  is not convex, by using the concept of *stable U-dense directions*. For a discussion on that technical point we refer the reader to [20, 21].

**Remark 3.11.** As mentioned in Remark 3.4, the general terminal constraint  $\psi(x(a), x(b)) \in S$  in Problem (P) allows to recover various situations of terminal constraints. We refer to Remark 10 of [21] for more details, and also for the description of the corresponding transversality conditions of Theorem 3.7.

**Remark 3.12.** Observe that, if the map  $\psi$  is not submersive at  $(x^*(a), x^*(b))$ , then one might be looking for replacing the transversality condition in Theorem 3.7 by

$$\begin{pmatrix} p(a) \\ -q(b) \end{pmatrix} = \lambda \nabla g(x^*(a), x^*(b)) + N_{\psi^{-1}(S)}^{\lim}(x^*(a), x^*(b)),$$

in which  $N_{\psi^{-1}(S)}^{\lim}(x^*(a), x^*(b))$  stands for the *limiting normal cone* of the closed set  $\psi^{-1}(S)$  at the point  $(x^*(a), x^*(b)) \in \psi^{-1}(S)$ . We refer to Theorem 22.2 of [28] or Theorem 9.3.1 of [71] for similar statements in that direction.

**Remark 3.13.** As in continuous-time optimal permanent control problems (see, e.g., [47, 71]), the vector  $p$  (resp. the vector  $q$ ) provided in Thm. 3.7 is called *AC-costate vector* (resp. *BV-costate vector*). Note that the terminology *adjoint vector* is also frequently used in the literature instead of *costate vector*. Up to the presence of the shift  $\sigma$ , the AC-costate vector  $p$  corresponds to an absolutely continuous part of the BV-costate vector  $q$ , and the difference between them can be expressed in terms of  $\sum_{i=1}^j \int_{[a, t]_{\mathbb{T}}} \nabla_x h_i(x^*(\tau), \tau) d\eta_i(\tau)$ . From the complementary slackness condition, we deduce that this difference (containing possibly discontinuity jumps and singular parts) intervenes when the inequality state constraints  $h_i(x^*(t), t) \leq 0$  are active, that is, when  $h_i(x^*(t), t) = 0$  for some  $i = 1, \dots, j$ . This behavior is well illustrated in Section 4, where Theorem 3.7 is applied to solve numerically a continuous-time min-max optimal sampled-data control problem.

**Remark 3.14.** As in continuous-time optimal permanent control problems, the present extension of the Pontryagin maximum principle for optimal sampled-data control problems on time scales (that can be found in [21], Thm. 2.6) to the state constrained case is not trivial. The authors of [21] involve the sensitivity analysis of the state equation under (explicit) *needle-like variations* of the control. This method is not applicable to handle inequality state constraints such as the ones considered in the present work. Indeed, taking a needle-like variation of the control at a given time  $s \in [a, b]_{\mathbb{T}_1} \cap \text{RD}_1$  leads to the differentiability (with respect to the perturbation parameter) of the state  $x$  (only) over the interval  $[s + \delta, b]_{\mathbb{T}}$  for some  $\delta > 0$  small (see [21], Prop. 4). However the analysis of the inequality state constraints  $h_i(x(t), t) \leq 0$  requires differentiability over the whole interval  $[a, b]_{\mathbb{T}}$ . In order to overcome this technical difficulty, our idea in this paper is to involve the sensitivity analysis of the state equation under *implicit spike variations* of the control. This concept was used in [13, 16, 54] for continuous-time optimal permanent control problems and is based in particular on the fact that the Lebesgue measure  $\mu_{\mathcal{L}}$  is nonatomic. As a consequence, the adaptation of this concept to the time scale framework is not trivial because the Lebesgue  $\Delta_1$ -measure  $\mu_{\Delta_1}$  is atomic (since the  $\Delta_1$ -measure of any right-scattered point of  $\mathbb{T}_1$  is positive). In the proof of Theorem 3.7 (precisely in Sect. 5.1.2), we introduce a suitable time scale version of the concept of implicit spike variations, by distinguishing the perturbation at right-dense points and at right-scattered points of  $\mathbb{T}_1$ . In particular, at right-scattered points of  $\mathbb{T}_1$ , we involve a convex  $L^\infty$ -variation of the control as in [21].

**Remark 3.15.** Consider the framework of Theorem 3.7 in the state constraint-free case (in particular  $q = p^\sigma$  in that context). In continuous-time optimal permanent control problems, that is when  $\mathbb{T} = \mathbb{T}_1 = \mathbb{R}_+$  for example, it is well known that under suitable assumptions the *maximized Hamiltonian function*  $\mathcal{H} : [a, b] \rightarrow \mathbb{R}$  defined by

$$\mathcal{H}(t) := H(x^*(t), u^*(t), p(t), t),$$

for almost every  $t \in [a, b]$ , can be identified to an absolutely continuous function which satisfies

$$\dot{\mathcal{H}}(t) = \nabla_t H(x^*(t), u^*(t), p(t), t),$$

for almost every  $t \in [a, b]$  (see, e.g., [35], Thm. 2.6.3). This property is not true in general for optimal sampled-data control theory. We refer to [17] for a detailed discussion on that particular point.

**Remark 3.16.** In this paper our main result (Thm. 3.7) provides a Pontryagin maximum principle for a class of state constrained optimal control problems in which the cost is expressed in the Mayer form. However, employing a well known state augmentation technique (see, e.g., [71], Chap. 6) Theorem 3.7 allows to provide the necessary conditions also for problems in the Bolza form, in which we add an integral term in the cost to minimize:

$$g(x(a), x(b)) + \int_{[a, b]_{\mathbb{T}}} L(x(\tau), u^\square(\tau), \tau) \Delta\tau,$$

where  $L : \mathbb{R}^n \times \mathbb{R}^m \times [a, b]_{\mathbb{T}} \rightarrow \mathbb{R}$  is a given continuous function, which is of class  $C^1$  with respect to its first two variables. Then the conclusions of Theorem 3.7 are still valid, replacing  $H$  by  $\tilde{H} : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^n \times \mathbb{R} \times [a, b]_{\mathbb{T}} \rightarrow \mathbb{R}$  defined by

$$\tilde{H}(x, u, p, \lambda, t) := \langle p, f(x, u, t) \rangle_{\mathbb{R}^n} - \lambda L(x, u, t),$$

for all  $(x, u, p, \lambda, t) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^n \times \mathbb{R} \times [a, b]_{\mathbb{T}}$ .

**Remark 3.17.** As mentioned in Remark 3.4, owing to the time scale theory, Theorem 3.7 represents an extension to the general time scale setting of several versions of the Pontryagin maximum principle, among which we find:

- continuous-time optimal permanent control problems without inequality state constraint (see, e.g., [61]) and with inequality state constraints (see, e.g., [16, 47, 54, 71]);

- discrete-time optimal permanent control problems without inequality state constraint (see, *e.g.*, [12]) and with inequality state constraints (see, *e.g.*, [62]);
- general optimal permanent control problems on time scales without inequality state constraint (see, *e.g.*, [11, 20]);
- continuous-time optimal sampled-data control problems without inequality state constraint (see, *e.g.*, [22]) and with inequality state constraints (see, *e.g.*, [18]);
- general optimal sampled-data control problems on time scales without inequality state constraint (see, *e.g.*, [21]).

Furthermore the time scale result provided by Theorem 3.7 allows to extend the Pontryagin maximum principle also to numerous hybrid situations (when  $\mathbb{T}$  and/or  $\mathbb{T}_1$  contain isolated points and disjoint intervals of positive length for example).

#### 4. APPLICATION TO CONTINUOUS-TIME MIN-MAX OPTIMAL SAMPLED-DATA CONTROL PROBLEMS

This section is dedicated to an illustrative application of Theorem 3.7 in the continuous-time case  $\mathbb{T} = \mathbb{R}_+$  with sampled-data control where  $\mathbb{T}_1 = \frac{2}{N}\mathbb{N}$  for some  $N \geq 2$ . Nonetheless, thanks to the general time scale setting, recall that our main result can be applied to several different situations described, for example, in Remark 3.4. Precisely our objective in this section is to solve numerically a *maximal velocity minimization problem of the continuous-time harmonic oscillator with sampled-data control*. More precisely, we consider the second-order control system described by

$$\ddot{x}(t) = -x(t) + u_k, \quad \text{a.e. } t \in \left[k\frac{2}{N}, (k+1)\frac{2}{N}\right), \quad \forall k = 0, \dots, N-1,$$

on the time interval  $[0, 2]$ , with a piecewise constant control (according to the uniform  $N$ -partition of the time interval  $[0, 2]$ , with  $N \geq 2$ ). We consider the control constraint  $|u_k| \leq 1$  for all  $k = 0, \dots, N-1$ . Here, the aim is to reach the final position  $(x(2), \dot{x}(2)) = (1, 0)$  from the initial condition  $(x(0), \dot{x}(0)) = (0, 0)$  by minimizing the maximal velocity  $\max_{t \in [0, 2]} \dot{x}(t)$ .

This section is organized as follows. In Section 4.1 we first establish in Proposition 4.1 a Pontryagin maximum principle for a continuous-time min–max optimal sampled-data control problem. Indeed, following a well known idea (see, *e.g.*, [31], Rem. 6 or [71], Chap. 9), we can reformulate such a problem as a state constrained continuous-time optimal sampled-data control problem which corresponds to a particular case of Problem (P) and then we apply Theorem 3.7. Then, in Section 4.2, we reformulate the above maximal velocity minimization problem as a continuous-time min–max optimal sampled-data control problem and we apply Proposition 4.1. Finally we are able to provide numerical results based on the necessary optimality conditions provided by Proposition 4.1 which are solved using a standard *shooting method* (see, *e.g.*, [69], p. 170–171 for more details on shooting methods). This example exhibits a *bouncing phenomenon* for the optimal trajectory when it reaches the maximal velocity to be minimized. This is accompanied by a nontrivial Borel measure and by corresponding jumps of the corresponding costate vector.

##### 4.1. Necessary conditions for min–max problems

Let  $\mathcal{P} := \{t_k\}_{k=0, \dots, N}$  be a partition of the time interval  $[a, b]$ , that is,  $a = t_0 < t_1 < \dots < t_N = b$  with  $N \in \mathbb{N}^*$ . In this section we consider the general *continuous-time min-max optimal sampled-data control problem*

given by

$$\left. \begin{aligned}
 & \text{minimize} && \max_{t \in [a, b]} L(x(t), t), \\
 & \text{subject to} && x \in \text{AC}([a, b], \mathbb{R}^n), \quad u = \{u_k\}_{k=0, \dots, N-1} \in (\mathbb{R}^m)^N, \\
 & && \dot{x}(t) = f(x(t), u_k, t), \quad \text{a.e. } t \in [t_k, t_{k+1}), \quad \forall k = 0, \dots, N-1, \\
 & && x(a) = x_a, \quad x(b) = x_b, \\
 & && u_k \in U, \quad \forall k = 0, \dots, N-1,
 \end{aligned} \right\} \quad (\text{MMP})$$

where  $x_a, x_b \in \mathbb{R}^n$  are fixed and where  $L : \mathbb{R}^n \times [a, b] \rightarrow \mathbb{R}$  is a given real function.

**Proposition 4.1.** *Assume (H1)–(H2) and that  $L : \mathbb{R}^n \times [a, b] \rightarrow \mathbb{R}$  is continuous and of class  $C^1$  in its first variable. If  $(x^*, u^*) \in \text{AC}([a, b], \mathbb{R}^n) \times (\mathbb{R}^m)^N$  is a solution to Problem (MMP), then there exist  $\lambda \geq 0$ ,  $p \in \text{AC}([a, b], \mathbb{R}^n)$  and a finite nonnegative Borel measure  $d\eta$  on  $[a, b]$  such that the following conditions are satisfied:*

- (i) **Nontriviality:**  $(\lambda, p, d\eta) \neq 0$ ;
- (ii) **Adjoint equation:**

$$-\dot{p}(t) = \nabla_x f(x^*(t), u_k^*, t)^\top \times q(t), \quad \text{a.e. } t \in [t_k, t_{k+1}), \quad \forall k = 0, \dots, N-1;$$

- (iii) **Transversality condition:**  $d\eta([a, b]) = \lambda$ ;
- (iv) **Nonpositive averaged Hamiltonian gradient condition:**

$$\left\langle \int_{t_k}^{t_{k+1}} \nabla_u f(x^*(\tau), u_k^*, \tau)^\top \times q(\tau) \, d\tau, v - u_k^* \right\rangle_{\mathbb{R}^m} \leq 0,$$

for all  $v \in U$  and all  $k = 0, \dots, N-1$ ;

- (v) **Complementary slackness condition:**

$$\text{supp}(d\eta) \subset \{t \in [a, b] \mid L(x^*(t), t) = L^*\},$$

where  $L^* := \max_{t \in [a, b]} L(x^*(t), t)$ .

Here  $q \in \text{BV}([a, b], \mathbb{R}^n)$  is defined by

$$\forall t \in [a, b], \quad q(t) := p(t) + \int_{[a, t]} \nabla_x L(x^*(\tau), \tau) \, d\eta(\tau).$$

*Proof.* Following a well known idea (see, e.g., [31], Rem. 6 or [71], Chap. 9), we reformulate Problem (MMP) as the augmented and state constrained continuous-time optimal sampled-data control problem given by

$$\left. \begin{aligned} & \text{minimize} && z(b), \\ & \text{subject to} && (x, z) \in \text{AC}([a, b], \mathbb{R}^{n+1}), \quad u = \{u_k\}_{k=0, \dots, N-1} \in (\mathbb{R}^m)^N, \\ & && \begin{pmatrix} \dot{x}(t) \\ \dot{z}(t) \end{pmatrix} = \begin{pmatrix} f(x(t), u_k, t) \\ 0 \end{pmatrix}, \quad \text{a.e. } t \in [t_k, t_{k+1}), \quad \forall k = 0, \dots, N-1, \\ & && x(a) = x_a, \quad x(b) = x_b, \\ & && L(x(t), t) - z(t) \leq 0, \quad \forall t \in [a, b], \\ & && u_k \in U, \quad \forall k = 0, \dots, N-1. \end{aligned} \right\}$$

Note that the above problem is a particular case of Problem (P), taking  $\mathbb{T} = [a, +\infty)$  and  $\mathbb{T}_1 = \mathcal{P} \cup [b, +\infty)$  (see Exam. 3.2 for a similar situation). As a consequence one can easily see that Proposition 4.1 directly follows from the application of Theorem 3.7.  $\square$

## 4.2. Application to a maximal velocity minimization problem

In order to provide a numerical solution to the maximal velocity minimization problem presented at the beginning of Section 4, we take  $a = 0$ ,  $b = 2$  and the uniform  $N$ -partition  $\mathcal{P} = \{t_k = k \frac{2}{N}\}_{k=0, \dots, N}$  of the time interval  $[0, 2]$  with  $N \geq 2$ . We reformulate the maximal velocity minimization problem as the continuous-time min-max optimal sampled-data control problem given by

$$\left. \begin{aligned} & \text{minimize} && \max_{t \in [0, 2]} x_2(t), \\ & \text{subject to} && x = (x_1, x_2) \in \text{AC}([0, 2], \mathbb{R}^2), \quad u = \{u_k\}_{k=0, \dots, N-1} \in \mathbb{R}^N, \\ & && \begin{pmatrix} \dot{x}_1(t) \\ \dot{x}_2(t) \end{pmatrix} = \begin{pmatrix} x_2(t) \\ -x_1(t) + u_k \end{pmatrix}, \quad \text{a.e. } t \in [t_k, t_{k+1}), \quad \forall k = 0, \dots, N-1, \\ & && \begin{pmatrix} x_1(0) \\ x_2(0) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad \begin{pmatrix} x_1(2) \\ x_2(2) \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \\ & && u_k \in [-1, 1], \quad \forall k = 0, \dots, N-1. \end{aligned} \right\} \quad (4.1)$$

Assume that Problem (4.1) admits a solution  $(x^*, u^*)$ , and denote by  $V^* := \max_{t \in [0, 2]} x_2^*(t)$  the corresponding maximal velocity. One can easily prove by contradiction (using the initial and final conditions) that  $V^* > 0$ . Let us denote by  $\lambda \geq 0$ ,  $p = (p_1, p_2) \in \text{AC}([0, 2], \mathbb{R}^2)$ ,  $q = (q_1, q_2) \in \text{BV}([0, 2], \mathbb{R}^2)$  and  $d\eta$  the elements provided in Proposition 4.1.

From the state equations and initial conditions, one can obtain that

$$\begin{pmatrix} x_1^*(t) \\ x_2^*(t) \end{pmatrix} = A(t - t_k) \begin{pmatrix} \lambda_k \\ \mu_k \end{pmatrix} + \begin{pmatrix} u_k^* \\ 0 \end{pmatrix},$$

for all  $t \in [t_k, t_{k+1}]$  and all  $k = 0, \dots, N-1$ , where

$$\forall t \in \mathbb{R}, \quad A(t) := \begin{pmatrix} \cos(t) & \sin(t) \\ -\sin(t) & \cos(t) \end{pmatrix},$$



where  $(\lambda_0, \mu_0)^\top := (-u_0^*, 0)^\top$  and

$$\begin{pmatrix} \lambda_{k+1} \\ \mu_{k+1} \end{pmatrix} := A(t_{k+1} - t_k) \begin{pmatrix} \lambda_k \\ \mu_k \end{pmatrix} + \begin{pmatrix} u_k^* - u_{k+1}^* \\ 0 \end{pmatrix}.$$

for all  $k = 0, \dots, N - 2$ .

Note that the equation  $\dot{x}_2^*(t) = 0$  over an open sampling interval  $(t_k, t_{k+1})$  for some  $k = 0, \dots, N - 1$  is equivalent to the equation  $\lambda_k + \mu_k \tan(t - t_k) = 0$  (since  $N \geq 2$  which guarantees that the length  $\frac{2}{N}$  of the partition is less than  $\frac{\pi}{2}$ ). As a consequence, one can easily deduce that  $x_2^*$  attains its maximal value  $V^* > 0$  at most one time on each open sampling interval  $(t_k, t_{k+1})$ . From the complementary slackness condition in Proposition 4.1, we know that  $\text{supp}(d\eta)$  is included in a finite set of at most  $2N + 1$  elements denoted by  $\{s_k\}_{k=0, \dots, 2N}$  satisfying  $0 = s_0 < s_1 < \dots < s_{2N} = 2$  and including the elements of  $\mathcal{P}$  (precisely  $s_{2k} = t_k$  for all  $k = 0, \dots, N$ ). It follows that

$$d\eta([0, t]) = \sum_{k=0}^K \eta_k,$$

for all  $t \in [s_K, s_{K+1})$  and all  $K = 0, \dots, 2N - 1$ , and for  $t = 2$  with  $K = 2N$ , where the elements  $\eta_k := d\eta(\{s_k\}) \geq 0$ , for all  $k = 0, \dots, 2N$ , satisfy  $\sum_{k=0}^{2N} \eta_k = d\eta([0, 2]) = \lambda$  from the transversality condition in Proposition 4.1. Note that  $\eta_0 = \eta_{2N} = 0$  since  $x_2^*(0) = x_2^*(2) = 0 < V^*$ .

From the adjoint equation in Proposition 4.1, one can derive that

$$\begin{pmatrix} p_1(t) \\ p_2(t) \end{pmatrix} = A(t - s_K) \begin{pmatrix} -\beta_K \\ \alpha_K \end{pmatrix} + \begin{pmatrix} 0 \\ -\sum_{k=0}^K \eta_k \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} q_1(t) \\ q_2(t) \end{pmatrix} = \begin{pmatrix} p_1(t) \\ p_2(t) \end{pmatrix} + \begin{pmatrix} 0 \\ \sum_{k=0}^K \eta_k \end{pmatrix},$$

for all  $t \in [s_K, s_{K+1})$  and all  $K = 0, \dots, 2N - 1$ , and for  $t = 2$  with  $K = 2N$ , where  $(\alpha_0, \beta_0)^\top := (p_2(0) + \eta_0, -p_1(0))^\top$  and

$$\begin{pmatrix} \alpha_{K+1} \\ \beta_{K+1} \end{pmatrix} := A(s_{K+1} - s_K) \begin{pmatrix} \alpha_K \\ \beta_K \end{pmatrix} + \begin{pmatrix} \eta_{K+1} \\ 0 \end{pmatrix},$$

for all  $K = 0, \dots, 2N - 1$ .

Finally, from the nonpositive averaged Hamiltonian gradient condition in Proposition 4.1, we obtain information on the value  $u_k^*$  in function of the sign of  $\int_{t_k}^{t_{k+1}} q_2(\tau) d\tau = p_1(t_{k+1}) - p_1(t_k)$  for all  $k = 0, \dots, N - 1$ . Precisely:

- if  $p_1(t_{k+1}) - p_1(t_k) < 0$ , then  $u_k^* = -1$ ;
- if  $p_1(t_{k+1}) - p_1(t_k) > 0$ , then  $u_k^* = 1$ ;
- if  $u_k^* \in (-1, 1)$ , then  $p_1(t_{k+1}) - p_1(t_k) = 0$ ;

for all  $k = 0, \dots, N - 1$ .

#### 4.2.1. Abnormal situation $\lambda = 0$

In the abnormal situation  $\lambda = 0$ , from the transversality condition in Proposition 4.1, we get that  $d\eta$  is the null Borel measure on  $[0, 2]$  and thus  $q = p$ . From the adjoint equation, it follows that

$$\begin{pmatrix} p_1(t) \\ p_2(t) \end{pmatrix} = A(t) \begin{pmatrix} p_1(0) \\ p_2(0) \end{pmatrix},$$

for all  $t \in [0, 2]$ . From the nontriviality condition in Proposition 4.1, we deduce that  $(p_1(0), p_2(0)) \neq (0, 0)$ . On the other hand, if  $p_1(0) \neq 0$ , it holds that

$$p_1(t) = \sqrt{p_1(0)^2 + p_2(0)^2} \sin \left( t + \frac{\pi}{2} \text{sign}(p_1(0)) - \arctan \left( \frac{p_2(0)}{p_1(0)} \right) \right),$$

for all  $t \in [0, 2]$ . We conclude that, in both cases  $p_1(0) = 0$  and  $p_1(0) \neq 0$ , the function  $p_1$  changes of (strict) variation at most one time over the interval  $[0, 2]$ . We deduce that  $p_1(t_{k+1}) - p_1(t_k)$  changes of sign at most one time. From the nonpositive averaged Hamiltonian gradient condition in Proposition 4.1, we deduce that the sequence  $u^*$  is monotone and takes at most three different values which can be 1, some  $\rho \in (-1, 1)$  and  $-1$ . Furthermore note that the sequence  $u^*$  can take an interior value  $\rho \in (-1, 1)$  at only one value  $k = 0, \dots, N-1$ . In the sequel we denote by  $0 =: \tau_0 < \tau_1 < \tau_2 < \tau_3 := 2$ , where  $\tau_1$  and  $\tau_2$  are the precise sampling times  $t_k$  at which the optimal control  $u^*$  possibly changes its value (from 1 to  $\rho$ , or from 1 to  $-1$ , for example). Note that the times  $\tau_i$  are not all the sampling times  $t_k$  (but only four of them). Furthermore note that one of the intervals  $[\tau_i, \tau_{i+1})$  is of length  $\frac{2}{N}$ . We denote by  $\bar{u}_i^*$  the value of  $u^*$  corresponding to each interval  $[\tau_i, \tau_{i+1})$ . One can easily obtain that

$$\begin{aligned} \begin{pmatrix} 1 \\ 0 \end{pmatrix} &= \begin{pmatrix} x_1^*(2) \\ x_2^*(2) \end{pmatrix} \\ &= A(\tau_3 - \tau_0) \begin{pmatrix} -\bar{u}_0^* \\ 0 \end{pmatrix} + A(\tau_3 - \tau_1) \begin{pmatrix} \bar{u}_1^* - \bar{u}_0^* \\ 0 \end{pmatrix} + A(\tau_3 - \tau_2) \begin{pmatrix} \bar{u}_2^* - \bar{u}_1^* \\ 0 \end{pmatrix} + \begin{pmatrix} \bar{u}_2^* \\ 0 \end{pmatrix}. \end{aligned} \quad (4.2)$$

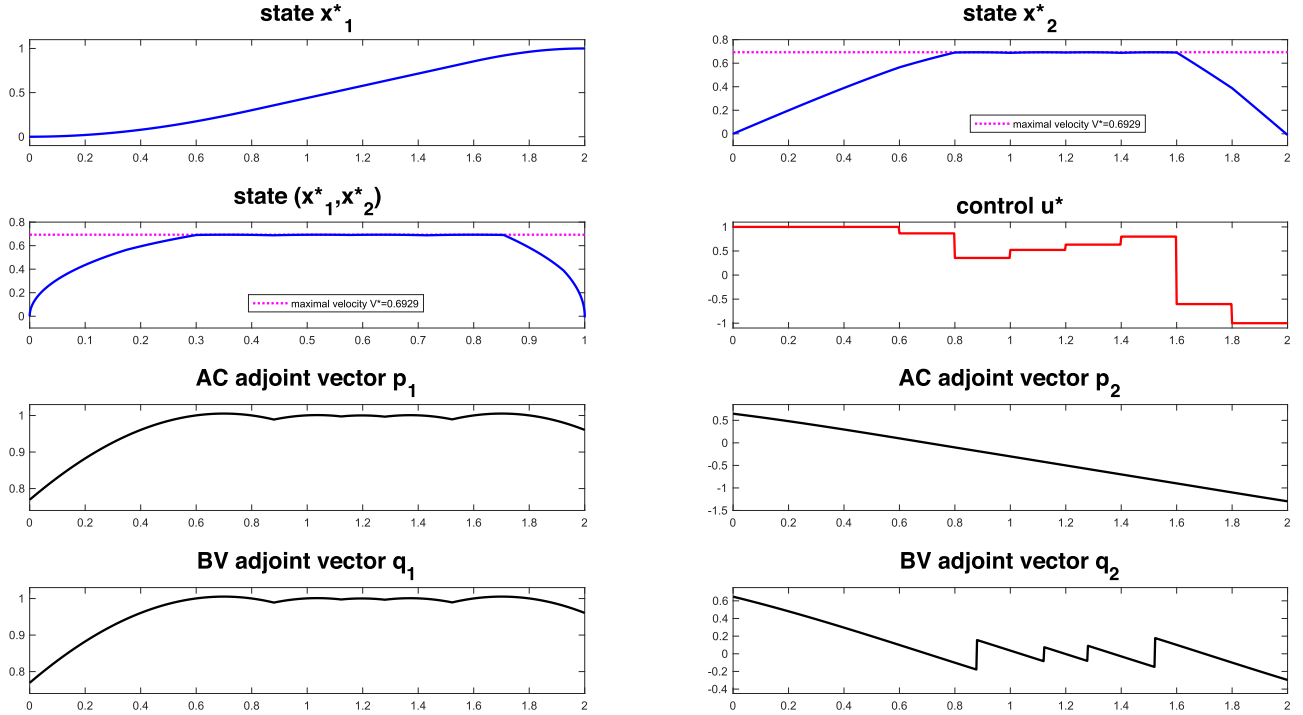
We performed numerical simulations in order to solve the above equation in the case  $N = 10$  taking into account all the above conditions, in particular that the triplet  $(\bar{u}_0^*, \bar{u}_1^*, \bar{u}_2^*)$  belongs to one of the sets

$$\begin{aligned} \{-1\}^2 \times [-1, 1], \quad \{-1\} \times [-1, 1] \times \{1\}, \quad [-1, 1] \times \{1\}^2, \\ \{1\}^2 \times [-1, 1], \quad \{1\} \times [-1, 1] \times \{-1\}, \quad [-1, 1] \times \{-1\}^2. \end{aligned}$$

It appears numerically that Equation (4.2), taking into account of our constraints, has no solution. In the next paragraph we deal with the normal case  $\lambda = 1$ .

#### 4.2.2. Normal situation $\lambda = 1$

This paragraph is dedicated to an indirect numerical method used for the normal case  $\lambda = 1$ . It consists in providing a guess of the couple  $(p_1(0), p_2(0))$ , of the times  $s_k$  (only for  $k$  odd) and of the values  $\eta_k$  for all  $k = 0, \dots, 2N$ , and computing the corresponding state and costate variables (owing to the explicit expressions provided above). Then we use the MATLAB function *fsolve* to find a solution which satisfies all the necessary conditions deduced from Proposition 4.1 (such as the transversality condition and the nonpositive averaged Hamiltonian gradient condition). The numerical results obtained with this method in the case when  $N = 10$  are displayed in Figure 1. Note that the optimal trajectory drawn in Figure 1 exhibits a *bouncing trajectory phenomenon*. This particular behavior has recently been highlighted and studied in [18], which deals with a class of state constrained continuous-time optimal sampled-data control problems. In that context, under some hypotheses (see [18], Sect. 4.2 for details), it was established that the optimal trajectories touch the state constraint boundary for at most a finite number of times. Figure 2 provides a zoom on the behavior of  $x_2^*$  with respect to the maximal velocity  $V^* \simeq 0.6929$ . One can see that the maximal velocity is attained a finite number of times (exactly four) and they exactly correspond to the discontinuity jumps of the BV-adjoint vector  $q_2$  in Figure 1 (see Rem. 3.13). We refer to Section 4 of [18] for more details on the *bouncing trajectory phenomenon* in state constrained continuous-time optimal sampled-data control problems.

FIGURE 1. Numerical results for  $N = 10$ .

The authors are thankful to Professor Emmanuel Trélat for the useful discussions on the illustrative example provided in this section.

## 5. PRELIMINARIES FOR THE PROOF OF THEOREM 3.7

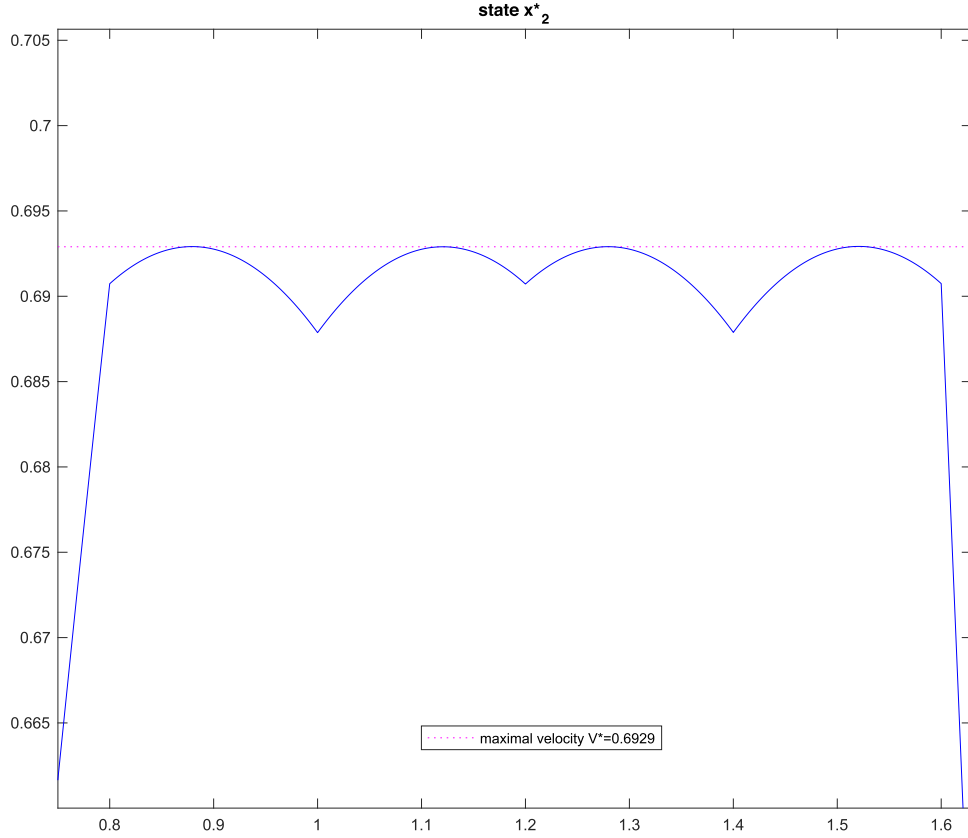
In this section we establish preliminary results needed for the proof of our main result (Thm. 3.7). Precisely we first investigate in Section 5.1 the sensitivity analysis of the state equation in Problem (P) with respect to particular control variations (called *implicit spike variations*). The last part of the section is devoted to recall some regularity properties of the distance function in both finite and infinite dimensional settings (see Sect. 5.2). In what follows, for a metric set  $(M, d_M)$ , we denote by  $\bar{B}_M(x, \nu)$  the closed ball of  $M$  centered at  $x \in M$  of radius  $\nu > 0$ .

### 5.1. Sensitivity analysis of the state equation

We first recall some Cauchy-Lipschitz (or Picard-Lindelöf) results. We refer to [19] for a detailed study of  $\Delta$ -Cauchy problems with Carathéodory dynamics on time scales. To simplify notation we shall write  $L_\Delta^\infty := L_\Delta^\infty([a, b]_{\mathbb{T}}, \mathbb{R}^m)$  and  $L_\Delta^1 := L_\Delta^1([a, b]_{\mathbb{T}}, \mathbb{R}^m)$ . Accordingly, we will use the notation  $L_{\Delta_1}^\infty$  and  $L_{\Delta_1}^1$ . All along this section we assume that (H1) is satisfied.

Owing to Theorem 1 of [19], for every control  $u \in L_{\Delta_1}^\infty$  and every initial condition  $x_a \in \mathbb{R}^n$ , there exists a unique maximal solution  $x(\cdot, u, x_a)$  to the following forward nonlinear  $\Delta$ -Cauchy problem ( $\text{CP}_{u, x_a}$ ):

$$\begin{cases} x^\Delta(t) = f(x(t), u^\square(t), t), & \Delta\text{-a.e. } t \in [a, b]_{\mathbb{T}}, \\ x(a) = x_a, \end{cases} \quad (\text{CP}_{u, x_a})$$


 FIGURE 2. Zoom on the state  $x_2^*$  in the case  $N = 10$ .

defined on a maximal interval denoted by  $I_{\mathbb{T}}(u, x_a)$ . Moreover, from Lemma 1 of [19], it holds that

$$x(t, u, x_a) = x_a + \int_{[a,t]_{\mathbb{T}}} f(x(\tau, u, x_a), u^{\square}(\tau), \tau) \Delta\tau,$$

for all  $t \in I_{\mathbb{T}}(u, x_a)$ .

**Definition 5.1** (Admissible for globality). A couple  $(u, x_a) \in L_{\Delta_1}^{\infty} \times \mathbb{R}^n$  is said to be *admissible for globality* whenever  $b \in I_{\mathbb{T}}(u, x_a)$ . In that case  $I_{\mathbb{T}}(u, x_a) = [a, b]_{\mathbb{T}}$  and we say that  $x(\cdot, u, x_a) \in AC([a, b]_{\mathbb{T}}, \mathbb{R}^n)$  is a *global solution* to  $(\mathcal{CP}_{u, x_a})$ .

We denote by  $\mathcal{AG}$  the set of all couples  $(u, x_a) \in L_{\Delta_1}^{\infty} \times \mathbb{R}^n$  admissible for globality. It is endowed with the distance

$$d_{\mathcal{AG}}((u, x_a), (u', x'_a)) := \|u - u'\|_{L_{\Delta_1}^1} + \|x_a - x'_a\|_{\mathbb{R}^n},$$

for all  $(u, x_a), (u', x'_a) \in \mathcal{AG}$ .

### 5.1.1. Openness and continuity results

For every  $(u, x_a) \in \mathcal{AG}$  and every  $R \geq \|u\|_{L_{\Delta_1}^\infty}$ , we introduce

$$K_R(u, x_a) := \{(y, v, t) \in \mathbb{R}^n \times \mathbb{R}^m \times [a, b]_{\mathbb{T}} \mid \|y - x(t, u, x_a)\|_{\mathbb{R}^n} \leq 1 \text{ and } \|v\|_{\mathbb{R}^m} \leq R\}.$$

From continuity of  $x(\cdot, u, x_a)$  on  $[a, b]_{\mathbb{T}}$ ,  $K_R(u, x_a)$  is a compact subset of  $\mathbb{R}^n \times \mathbb{R}^m \times [a, b]_{\mathbb{T}}$ . As a consequence, from (H1), note that  $f$ ,  $\nabla_x f$  and  $\nabla_u f$  are bounded on  $K_R(u, x_a)$  by some  $L_R(u, x_a) \geq 0$  and it follows that

$$\|f(y_2, v_2, t) - f(y_1, v_1, t)\|_{\mathbb{R}^n} \leq L_R(u, x_a)(\|y_2 - y_1\|_{\mathbb{R}^n} + \|v_2 - v_1\|_{\mathbb{R}^m}),$$

for all  $(y_1, v_1, t), (y_2, v_2, t) \in K_R(u, x_a)$ .

The next propositions are both extracted from Lemmas 4.3 and 4.5 of [21]. The first proposition states that  $\mathcal{AG}$  is an open set of  $L_{\Delta_1}^1 \times \mathbb{R}^n$  (up to a  $L_{\Delta_1}^\infty$ -bound), while the second proposition establishes a continuous dependence result of the trajectory  $x(\cdot, u, x_a)$  with respect to the pair  $(u, x_a)$ .

**Proposition 5.2.** *Assume (H1). Let  $(u, x_a) \in \mathcal{AG}$  and  $R \geq \|u\|_{L_{\Delta_1}^\infty}$ . There exists  $\nu_R(u, x_a) > 0$  such that*

$$\mathcal{N}_R(u, x_a) := \left( \overline{B}_{L_{\Delta_1}^1}(u, \nu_R(u, x_a)) \cap \overline{B}_{L_{\Delta_1}^\infty}(0_{L_{\Delta_1}^\infty}, R) \right) \times \overline{B}_{\mathbb{R}^n}(x_a, \nu_R(u, x_a)),$$

is included in  $\mathcal{AG}$ . Moreover, for every  $(u', x'_a) \in \mathcal{N}_R(u, x_a)$ , we have  $(x(\tau, u', x'_a), u'^{\square}(\tau), \tau) \in K_R(u, x_a)$  for  $\Delta$ -a.e.  $\tau \in [a, b]_{\mathbb{T}}$ .

**Proposition 5.3.** *Assume (H1). Let  $(u, x_a) \in \mathcal{AG}$  and  $R \geq \|u\|_{L_{\Delta_1}^\infty}$ . The map*

$$\begin{aligned} \mathcal{F}_R(u, x_a) : (\mathcal{N}_R(u, x_a), d_{\mathcal{AG}}) &\longrightarrow (C([a, b]_{\mathbb{T}}, \mathbb{R}^n), \|\cdot\|_\infty) \\ (u', x'_a) &\longmapsto x(\cdot, u', x'_a), \end{aligned}$$

is  $c_R(u, x_a)$ -Lipschitz continuous for some constant  $c_R(u, x_a) \geq 0$ .

### 5.1.2. Implicit spike variations and a differentiable dependence result

In the previous section we recalled a continuous dependence result. In order to establish differentiable dependence properties, the authors of [21] use the concept of (explicit) *needle-like variations* of the control. However this technique is not suitable in order to handle inequality state constraints such as the ones considered in the present work. We refer to Remark 3.14 for details. Our idea is thus to employ an *implicit spike variation* technique (see, e.g., [13, 16, 54]), but we have to adapt it to the general time scale setting because the  $\Delta_1$ -measure  $\mu_{\Delta_1}$  on the time scale  $\mathbb{T}_1$  is atomic (in contrary to the classical Lebesgue measure  $\mu_{\mathcal{L}}$ ). The first crucial step towards this goal is provided in the following technical lemma (a detailed proof of which is displayed in Appendix A), in which we use the notation  $\text{RD}_1^b := [a, b]_{\mathbb{T}_1} \cap \text{RD}_1$  and  $\text{RS}_1^b := [a, b]_{\mathbb{T}_1} \cap \text{RS}_1$ . Note that  $\text{RD}_1^b$  is the complement of  $\text{RS}_1^b$  in  $[a, b]_{\mathbb{T}_1}$ .

**Lemma 5.4.** *Let  $z \in L^1(\text{RD}_1^b, \mathbb{R}^n)$ . Then, for all  $\rho \in (0, 1)$ , there exists a  $\mu_{\mathcal{L}}$ -measurable set  $Q_\rho \subset \text{RD}_1^b$  such that  $\mu_{\mathcal{L}}(Q_\rho) = \rho \mu_{\mathcal{L}}(\text{RD}_1^b)$  and*

$$\sup_{t \in [a, b]} \left\| \int_{[a, t] \cap \text{RD}_1^b} \left( 1 - \frac{1}{\rho} \mathbf{1}_{Q_\rho}(\tau) \right) z(\tau) \, d\tau \right\|_{\mathbb{R}^n} \leq \rho.$$

**Remark 5.5.** The above result corresponds to an adaptation of the lemma provided in Paragraph 3.2, p. 143 of [54] to the time scale setting. In Lemma 5.4 note that we consider a general function  $z \in L^1(\text{RD}_1^b, \mathbb{R}^n)$  defined (only) on the right-dense points of  $\mathbb{T}_1$  in order to work on a set on which the  $\Delta_1$ -measure  $\mu_{\Delta_1}$  has no atom.

Now, given  $(u, x_a) \in \mathcal{AG}$  and  $(u', x'_a) \in L^\infty_{\Delta_1} \times \mathbb{R}^n$ , we write  $\pi := (u, x_a, u', x'_a)$ . Noting that

$$[a, b]_{\mathbb{T}} = \text{RD}_1^b \sqcup \left( \bigsqcup_{r \in \text{RS}_1^b} [r, \sigma_1(r)]_{\mathbb{T}} \right),$$

we introduce the function  $z_\pi \in L^\infty_{\Delta}([a, b]_{\mathbb{T}}, \mathbb{R}^n)$  defined by

$$z_\pi(\tau) := \begin{cases} f(x(\tau, u, x_a), u'(\tau), \tau) - f(x(\tau, u, x_a), u(\tau), \tau) & \text{if } \tau \in \text{RD}_1^b, \\ \nabla_u f(x(\tau, u, x_a), u(r), \tau) \times (u'(r) - u(r)) & \text{if } \tau \in [r, \sigma_1(r)]_{\mathbb{T}} \text{ for some } r \in \text{RS}_1^b, \end{cases}$$

for  $\Delta$ -a.e.  $\tau \in [a, b]_{\mathbb{T}}$ . For every  $\rho \in (0, 1)$ , from Lemma 5.4 we can consider a set  $\mathcal{Q}_\rho \subset \text{RD}_1^b$  associated with the restriction of  $z_\pi$  to  $\text{RD}_1^b$ . For  $\rho = 0$ , we set  $\mathcal{Q}_\rho := \emptyset$ .

Finally, for every  $\rho \in [0, 1)$ , we introduce the *implicit spike variation*  $u_\pi(\cdot, \rho)$  of  $u$  as follows:

$$u_\pi(\tau, \rho) := \begin{cases} u(\tau) & \text{if } \tau \in \text{RD}_1^b \setminus \mathcal{Q}_\rho, \\ u'(\tau) & \text{if } \tau \in \mathcal{Q}_\rho, \\ u(\tau) + \rho(u'(\tau) - u(\tau)) & \text{if } \tau \in \text{RS}_1^b, \end{cases}$$

for  $\Delta_1$ -a.e.  $\tau \in [a, b]_{\mathbb{T}_1}$ , by distinguishing the perturbations at right-dense points and at right-scattered points of  $\mathbb{T}_1$ . In particular, at right-scattered points, we involve a convex  $L^\infty$ -variation of the control as in [21].

Finally we consider the corresponding *variation vector*  $w_\pi \in \text{AC}([a, b]_{\mathbb{T}}, \mathbb{R}^n)$  defined as the unique maximal solution, which is global (see [19], Thm. 3), to the forward linear  $\Delta$ -Cauchy problem given by

$$\begin{cases} w^\Delta(t) = \nabla_x f(x(t, u, x_a), u^\square(t), t) \times w(t) + z_\pi(t), & \Delta\text{-a.e. } t \in [a, b]_{\mathbb{T}}, \\ w(a) = x'_a. \end{cases} \quad (5.1)$$

We prove next the following differentiable dependence result.

**Proposition 5.6.** *Assume (H1). The map*

$$F_\pi(\rho) := x(\cdot, u_\pi(\cdot, \rho), x_a + \rho x'_a) \in C([a, b]_{\mathbb{T}}, \mathbb{R}^n),$$

*is well-defined for sufficiently small  $\rho \geq 0$ , and is differentiable at  $\rho = 0$  with  $DF_\pi(0) = w_\pi$ .*

*Proof.* Set  $R := \max(\|u\|_{L^\infty_{\Delta_1}}, \|u'\|_{L^\infty_{\Delta_1}})$ . Observe that  $\|u_\pi(\cdot, \rho) - u\|_{L^1_{\Delta_1}} \leq 4R(b-a)\rho$  for every  $\rho \in [0, 1)$ . Indeed, this is immediate when  $\rho = 0$ . Otherwise, if  $\rho \in (0, 1)$ , we use the fact that  $\mu_{\mathcal{L}}(\mathcal{Q}_\rho) = \rho\mu_{\mathcal{L}}(\text{RD}_1^b)$  (see Lem. 5.4). As a consequence, for sufficiently small  $\rho \geq 0$ , we have  $(u_\pi(\cdot, \rho), x_a + \rho x'_a) \in \mathcal{N}_R(u, x_a) \subset \mathcal{AG}$  (see Prop. 5.3) and thus  $F_\pi(\rho)$  is well-defined. Moreover, it follows from Proposition 5.3 that  $\|x(\cdot, u_\pi(\cdot, \rho), x_a + \rho x'_a) - x(\cdot, u, x_a)\|_\infty \leq c_R(u, x_a)(4R(b-a) + \|x'_a\|_{\mathbb{R}^n})\rho$ . In particular  $x(\cdot, u_\pi(\cdot, \rho), x_a + \rho x'_a)$  converges uniformly to  $x(\cdot, u, x_a)$  on  $[a, b]_{\mathbb{T}}$  as  $\rho \rightarrow 0$ .

Now let us assume by contradiction that there exist  $\varepsilon > 0$  and a sequence of positive real numbers  $(\rho_k)_k$  converging to zero such that

$$\left\| \frac{F_\pi(\rho_k) - F_\pi(0)}{\rho_k} - w_\pi \right\|_\infty \geq \varepsilon, \quad (5.2)$$

for all  $k \in \mathbb{N}$ . In this proof, for ease of notation, we write  $w := w_\pi$ ,  $z := z_\pi$ ,  $x := x(\cdot, u, x_a)$ ,  $x_k := x(\cdot, u_\pi(\cdot, \rho_k), x_a + \rho_k x'_a)$  and  $u_k := u_\pi(\cdot, \rho_k)$  for every  $k \in \mathbb{N}$ . Since the sequence  $(u_k)_k$  converges to  $u$  in  $L^1_{\Delta_1}$ ,

from Proposition 3.3 we obtain that the sequence  $(u_k^\square)_k$  converges to  $u^\square$  in  $L_\Delta^1$ . We deduce from the (partial) converse of the Lebesgue dominated convergence theorem that there exists a subsequence (that we do not relabel) such that  $(u_k^\square)_k$  tends to  $u^\square$   $\Delta$ -a.e. on  $[a, b]_\mathbb{T}$ .

We define  $\phi_k(t) := \frac{x_k(t) - x(t)}{\rho_k} - w(t)$  for every  $t \in [a, b]_\mathbb{T}$  and every  $k \in \mathbb{N}$ . We have

$$\phi_k(t) = \int_{[a, t]_\mathbb{T}} \frac{f(x_k(\tau), u_k^\square(\tau), \tau) - f(x(\tau), u^\square(\tau), \tau)}{\rho_k} - \nabla_x f(x(\tau), u^\square(\tau), \tau) \times w(\tau) - z(\tau) \Delta\tau,$$

that is,

$$\begin{aligned} \phi_k(t) = \int_{[a, t]_\mathbb{T}} \left( \frac{f(x_k(\tau), u_k^\square(\tau), \tau) - f(x(\tau), u_k^\square(\tau), \tau)}{\rho_k} - \nabla_x f(x(\tau), u_k^\square(\tau), \tau) \times w(\tau) \right. \\ \left. + \frac{f(x(\tau), u_k^\square(\tau), \tau) - f(x(\tau), u^\square(\tau), \tau)}{\rho_k} - z(\tau) \right) \Delta\tau, \end{aligned}$$

for every  $t \in [a, b]_\mathbb{T}$  and every  $k \in \mathbb{N}$ . The Taylor formula with remainder in integral form leads to

$$\begin{aligned} \phi_k(t) = \int_{[a, t]_\mathbb{T}} \nabla_x f(x(\tau), u^\square(\tau), \tau) \times \phi_k(\tau) \Delta\tau \\ + \int_{[a, t]_\mathbb{T}} \left[ \int_0^1 \nabla_x f(x(\tau) + \theta(x_k(\tau) - x(\tau)), u_k^\square(\tau), \tau) d\theta - \nabla_x f(x(\tau), u^\square(\tau), \tau) \right] \\ \times \frac{x_k(\tau) - x(\tau)}{\rho_k} \Delta\tau + \int_{[a, t]_\mathbb{T}} \left( \frac{f(x(\tau), u_k^\square(\tau), \tau) - f(x(\tau), u^\square(\tau), \tau)}{\rho_k} - z(\tau) \right) \Delta\tau, \end{aligned}$$

for every  $t \in [a, b]_\mathbb{T}$  and every  $k \in \mathbb{N}$ . Hence we deduce that

$$\|\phi_k(t)\|_{\mathbb{R}^n} \leq \beta_k + c_R(u, x_a)(4R(b-a) + \|x'_a\|_{\mathbb{R}^n})\gamma_k + L_R(u, x_a) \int_{[a, t]_\mathbb{T}} \|\phi_k(\tau)\|_{\mathbb{R}^n} \Delta\tau,$$

for every  $t \in [a, b]_\mathbb{T}$  and every  $k \in \mathbb{N}$ , where

$$\beta_k := \sup_{t \in [a, b]} \left\| \int_{[a, t]_\mathbb{T}} \frac{f(x(\tau), u_k^\square(\tau), \tau) - f(x(\tau), u^\square(\tau), \tau)}{\rho_k} - z(\tau) \Delta\tau \right\|_{\mathbb{R}^n},$$

and

$$\gamma_k := \int_{[a, b]_\mathbb{T}} \int_0^1 \left\| \nabla_x f(x(\tau) + \theta(x_k(\tau) - x(\tau)), u_k^\square(\tau), \tau) - \nabla_x f(x(\tau), u^\square(\tau), \tau) \right\|_{\mathbb{R}^n \times \mathbb{R}^n} d\theta \Delta\tau.$$

From the time scale version of the Gronwall lemma (see [9], Thm. 6.4), we obtain that

$$\|\phi_k(t)\|_{\mathbb{R}^n} \leq (\beta_k + c_R(u, x_a)(4R(b-a) + \|x'_a\|_{\mathbb{R}^n})\gamma_k) e_{L_R(u, x_a)}(b, a),$$

for every  $t \in [a, b]_\mathbb{T}$  and every  $k \in \mathbb{N}$ . Here  $e_{L_R(u, x_a)}(b, a)$  stands for the time scale version of the exponential function (see [9], Chap. 2). Now our aim is to prove that  $(\beta_k)_k$  and  $(\gamma_k)_k$  tend to zero as  $k \rightarrow +\infty$ , and this leads to a contradiction with the inequality  $\|\phi_k\|_\infty \geq \varepsilon$  for all  $k \in \mathbb{N}$  assumed in (5.2).

From the continuity and the boundedness of  $\nabla_x f$  on  $K_R(u, x_a)$ , since  $(x_k)_k$  converges uniformly to  $x$  on  $[a, b]_{\mathbb{T}}$ , since  $(u_k^\square)_k$  tends to  $u^\square$   $\Delta$ -a.e. on  $[a, b]_{\mathbb{T}}$  and from the Lebesgue dominated convergence theorem, one can easily prove that  $(\gamma_k)_k$  tends to zero as  $k \rightarrow +\infty$ .

On the other hand, from the equality  $[a, b]_{\mathbb{T}} = \text{RD}_1^b \sqcup (\sqcup_{r \in \text{RS}_1^b} [r, \sigma_1(r)]_{\mathbb{T}})$  and since  $\text{RD}_1^b \subset \text{RD}$ , we have

$$\begin{aligned} \beta_k \leq & \sup_{t \in [a, b]} \left( \left\| \int_{[a, t] \cap \text{RD}_1^b} \frac{f(x(\tau), u_k(\tau), \tau) - f(x(\tau), u(\tau), \tau)}{\rho_k} - z(\tau) \, d\tau \right\|_{\mathbb{R}^n} \right) \\ & + \sum_{r \in \text{RS}_1^b} \int_{[r, \sigma_1(r)]_{\mathbb{T}}} \left\| \frac{f(x(\tau), u(r) + \rho_k(u'(r) - u(r)), \tau) - f(x(\tau), u(r), \tau)}{\rho_k} - z(\tau) \right\|_{\mathbb{R}^n} \Delta\tau, \end{aligned}$$

then, in particular from the Taylor formula with remainder in integral form, we obtain that

$$\begin{aligned} \beta_k \leq & \sup_{t \in [a, b]} \left\| \int_{[a, t] \cap \text{RD}_1^b} \left( \frac{1}{\rho_k} \mathbf{1}_{Q_{\rho_k}}(\tau) - 1 \right) z(\tau) \, d\tau \right\|_{\mathbb{R}^n} \\ & + 2R \sum_{r \in \text{RS}_1^b} \int_{[r, \sigma_1(r)]_{\mathbb{T}}} \int_0^1 \left\| \nabla_u f(x(\tau), u(r) + \theta \rho_k(u'(r) - u(r)), \tau) - \nabla_u f(x(\tau), u(r), \tau) \right\|_{\mathbb{R}^n \times \mathbb{R}^m} \, d\theta \Delta\tau, \end{aligned}$$

and then,

$$\begin{aligned} \beta_k \leq & \sup_{t \in [a, b]} \left\| \int_{[a, t] \cap \text{RD}_1^b} \left( \frac{1}{\rho_k} \mathbf{1}_{Q_{\rho_k}}(\tau) - 1 \right) z(\tau) \, d\tau \right\|_{\mathbb{R}^n} \\ & + 2R \int_{[a, b]_{\mathbb{T}}} \int_0^1 \left\| \nabla_u f(x(\tau), u^\square(\tau) + \theta \rho_k(u^\square(\tau) - u^\square(\tau)), \tau) - \nabla_u f(x(\tau), u^\square(\tau), \tau) \right\|_{\mathbb{R}^n \times \mathbb{R}^m} \, d\theta \Delta\tau. \end{aligned}$$

From the continuity and the boundedness of  $\nabla_u f$  on  $K_R(u, x_a)$ , one can easily prove that the second term tends to zero, as  $k \rightarrow +\infty$ , and from Lemma 5.4 one can conclude that  $(\beta_k)_k$  tends to zero, as well. The proof is complete.  $\square$

## 5.2. Preliminaries on the distance function

Assume (H5). We denote by  $d_S : \mathbb{R}^\ell \rightarrow \mathbb{R}_+$  the standard distance function to  $S$  defined by  $d_S(x) := \inf_{x' \in S} \|x - x'\|_{\mathbb{R}^\ell}$  for every  $x \in \mathbb{R}^\ell$ . Since  $S$  is a nonempty closed convex subset of  $\mathbb{R}^\ell$ , recall that, for every  $x \in \mathbb{R}^\ell$ , there exists a unique element  $P_S(x) \in S$  (projection of  $x$  onto  $S$ ) such that  $d_S(x) = \|x - P_S(x)\|_{\mathbb{R}^\ell}$ . It is characterized by the property  $\langle x - P_S(x), x' - P_S(x) \rangle_{\mathbb{R}^\ell} \leq 0$  for every  $x' \in S$ . In particular  $x - P_S(x) \in N_S(P_S(x))$  for all  $x \in \mathbb{R}^\ell$  (the notion of normal cone is recalled in Sect. 3.3). The map  $P_S : \mathbb{R}^\ell \rightarrow S \subset \mathbb{R}^\ell$  is 1-Lipschitz continuous. In the two following lemmas we summarize well known properties (see, e.g., [28, 71] for proofs) which will be used in our analysis.

**Lemma 5.7.** *Assume (H5). Let  $(x_k)_k$  be a sequence of points of  $\mathbb{R}^\ell$  and  $(\kappa_k)_k$  be a sequence of nonnegative real numbers such that  $x_k \rightarrow x \in S$  and  $\kappa_k(x_k - P_S(x_k)) \rightarrow y \in \mathbb{R}^\ell$ . Then  $y \in N_S(x)$ .*

**Lemma 5.8.** *Assume (H5). The map  $d_S^2 : \mathbb{R}^\ell \rightarrow \mathbb{R}_+$ , defined by  $d_S^2(x) := d_S(x)^2$  for all  $x \in \mathbb{R}^\ell$ , is Fréchet-differentiable on  $\mathbb{R}^\ell$ , with  $Dd_S^2(x)(x') = 2\langle x - P_S(x), x' \rangle_{\mathbb{R}^\ell}$  for all  $x, x' \in \mathbb{R}^\ell$ .*



To ease the exposition we shall write  $C_{\mathbb{T}}^n := C([a, b]_{\mathbb{T}}, \mathbb{R}^n)$  and  $C_{\mathbb{T}}^j := C([a, b]_{\mathbb{T}}, \mathbb{R}^j)$ , both endowed with the usual uniform norm  $\|\cdot\|_{\infty}$ . We introduce the function

$$\mathfrak{h} = (\mathfrak{h}_i)_{i=1, \dots, j} : \begin{array}{ccc} C_{\mathbb{T}}^n & \longrightarrow & C_{\mathbb{T}}^j \\ x & \longmapsto & \mathfrak{h}(x) := h(x, \cdot) = (h_i(x, \cdot))_{i=1, \dots, j}, \end{array}$$

and the nonempty closed convex cone  $\mathfrak{S}$  of  $C_{\mathbb{T}}^j$ , defined by  $\mathfrak{S} := C([a, b]_{\mathbb{T}}, \mathbb{R}_-^j)$ , where  $\mathbb{R}_-^j := \mathbb{R}_- \times \dots \times \mathbb{R}_-$ . Hence the state constraint condition in Problem (P) can equivalently be written as the following inclusion:

$$\mathfrak{h}(x) \in \mathfrak{S}.$$

Note that  $\mathfrak{h}$  is of class  $C^1$  with  $D\mathfrak{h}(x)(w) = \nabla_x h(x, \cdot) \times w$  for every  $x, w \in C_{\mathbb{T}}^n$ , and that  $\mathfrak{S}$  has a nonempty interior. The *normal cone* of  $\mathfrak{S}$  at  $x \in \mathfrak{S}$  is

$$N_{\mathfrak{S}}(x) := \{\zeta \in (C_{\mathbb{T}}^j)^* \mid \forall \varphi \in \mathfrak{S}, \langle \zeta, \varphi - x \rangle_{(C_{\mathbb{T}}^j)^* \times C_{\mathbb{T}}^j} \leq 0\},$$

where  $(C_{\mathbb{T}}^j)^*$  stands for the dual space of  $C_{\mathbb{T}}^j$  and where  $\langle \cdot, \cdot \rangle_{(C_{\mathbb{T}}^j)^* \times C_{\mathbb{T}}^j}$  stands for the standard duality bracket. Since  $(C_{\mathbb{T}}^j, \|\cdot\|_{\infty})$  is a separable Banach space and from Theorem 2.18 of [54], we endow  $C_{\mathbb{T}}^j$  with an equivalent norm  $\|\cdot\|_{C_{\mathbb{T}}^j}$  such that the associated dual norm  $\|\cdot\|_{(C_{\mathbb{T}}^j)^*}$  is strictly convex. Then we denote by  $d_{\mathfrak{S}} : C_{\mathbb{T}}^j \rightarrow \mathbb{R}_+$  the 1-Lipschitz continuous distance function to  $\mathfrak{S}$  defined by  $d_{\mathfrak{S}}(x) := \inf_{x' \in \mathfrak{S}} \|x - x'\|_{C_{\mathbb{T}}^j}$  for every  $x \in C_{\mathbb{T}}^j$ . Since the dual norm  $\|\cdot\|_{(C_{\mathbb{T}}^j)^*}$  is strictly convex and, from Theorem 3.54 of [59], we know that  $d_{\mathfrak{S}}$  is strictly Hadamard-differentiable outside of  $\mathfrak{S}$  with  $\|Dd_{\mathfrak{S}}(x)\|_{(C_{\mathbb{T}}^j)^*} = 1$  for every  $x \in C_{\mathbb{T}}^j \setminus \mathfrak{S}$ . As a consequence the map  $d_{\mathfrak{S}}^2 : C_{\mathbb{T}}^j \rightarrow \mathbb{R}_+$ , defined by  $d_{\mathfrak{S}}^2(x) := d_{\mathfrak{S}}(x)^2$  for all  $x \in C_{\mathbb{T}}^j$ , is also strictly Hadamard-differentiable outside of  $\mathfrak{S}$  with  $Dd_{\mathfrak{S}}^2(x) = 2d_{\mathfrak{S}}(x)Dd_{\mathfrak{S}}(x)$  for every  $x \in C_{\mathbb{T}}^j \setminus \mathfrak{S}$ . Finally, since  $d_{\mathfrak{S}}$  is 1-Lipschitz continuous, one can easily prove that  $d_{\mathfrak{S}}^2$  is Fréchet-differentiable on  $\mathfrak{S}$  with  $Dd_{\mathfrak{S}}^2(x) = 0_{(C_{\mathbb{T}}^j)^*}$  for every  $x \in \mathfrak{S}$ .

## 6. PROOF OF THEOREM 3.7

This section is devoted to the proof of Theorem 3.7, which is based on the following two steps:

- (i) establishing a crucial inequality (Prop. 6.1 in Sect. 6.1) derived from the Ekeland variational principle ([34], Thm. 1.1) used with the concept of implicit spike variations (recalled and adapted to the time scale setting with sampled-data controls in Sect. 5.1.2);
- (ii) the introduction of appropriate costate vectors in Section 6.2.

This proof extends the techniques employed in the state constrained continuous-time and permanent control setting (see, *e.g.*, [13, 16, 54, 71]) to the more general time scale setting with sampled-data controls.

Let  $(x^*, u^*) \in AC([a, b]_{\mathbb{T}}, \mathbb{R}^n) \times L_{\Delta_1}^{\infty}([a, b]_{\mathbb{T}_1}, \mathbb{R}^m)$  be a solution to Problem (P) such that  $\psi$  is submersive at  $(x^*(a), x^*(b))$ . In what follows we denote by  $x_a^* := x^*(a)$  and  $x_b^* := x^*(b)$ . Using notation of Section 5 we have  $(u^*, x_a^*) \in \mathcal{AG}$  and  $x^* = x(\cdot, u^*, x_a^*)$ .

### 6.1. A crucial inequality from the Ekeland variational principle

The next proposition has a central role, and its proof is provided in three steps (in Sects. 6.1.1, 6.1.2 and 6.1.3).

**Proposition 6.1.** *Assume (H1)–(H6). Then, there exists a nontrivial triplet  $(\lambda, \xi, \zeta) \in \mathbb{R}_+ \times \mathcal{N}_S(\psi(x_a^*, x_b^*)) \times \mathcal{N}_S(h(x^*, \cdot))$  such that the inequality*

$$0 \leq \left\langle \lambda \nabla g(x_a^*, x_b^*) + \nabla \psi(x_a^*, x_b^*)^\top \times \xi, \begin{pmatrix} x_a' \\ w_\pi(b) \end{pmatrix} \right\rangle_{\mathbb{R}^{2n}} + \langle \zeta, \nabla_x h(x^*, \cdot) \times w_\pi \rangle_{(C_T^j)^* \times C_T^j}, \quad (6.1)$$

holds for all  $(u', x_a') \in L_{\Delta_1}^\infty \times \mathbb{R}^n$  with  $u'(\tau) \in U$  for  $\Delta_1$ -a.e.  $\tau \in [a, b]_{\mathbb{T}_1}$ , where  $w_\pi$  is the variation vector associated with  $\pi = (u^*, x_a^*, u', x_a')$  defined in Section 5.1.2.

### 6.1.1. A penalized functional

Let  $R \in \mathbb{N}$  be fixed such that  $R \geq \|u^*\|_{L_{\Delta_1}^\infty}$  and let  $\nu_R(u^*, x_a^*) > 0$  given in Proposition 5.2. Take a sequence  $(\varepsilon_k)_k$  of positive real numbers tending to zero, as  $k \rightarrow +\infty$ , such that  $0 < \sqrt{\varepsilon_k} < \nu_R(u^*, x_a^*)$  for all  $k \in \mathbb{N}$ . We consider the penalized functional  $J_k^R : (\mathcal{N}_R^U(u^*, x_a^*), d_{\mathcal{AG}}) \rightarrow (0, +\infty)$  defined by

$$J_k^R(u, x_a) := \sqrt{\left( (g(x_a, x(b, u, x_a)) - g(x_a^*, x_b^*) + \varepsilon_k)^+ \right)^2 + d_S^2(\psi(x_a, x(b, u, x_a))) + d_{\mathfrak{S}}^2(\mathfrak{h}(x(\cdot, u, x_a)))},$$

for all  $k \in \mathbb{N}$ , where

$$\mathcal{N}_R^U(u^*, x_a^*) := \{(u, x_a) \in \mathcal{N}_R(u^*, x_a^*) \mid u(\tau) \in U \text{ for } \Delta_1\text{-a.e. } \tau \in [a, b]_{\mathbb{T}_1}\}.$$

Observe that, for every  $k \in \mathbb{N}$ ,  $J_k^R$  is well-defined because  $\mathcal{N}_R^U(u^*, x_a^*) \subset \mathcal{AG}$  (see Prop. 5.2) and that  $J_k^R$  is a strictly positive functional from the optimality of  $(u^*, x_a^*)$ . Since  $U$  is a nonempty closed subset of  $\mathbb{R}^m$ , from the (partial) converse of the Lebesgue dominated convergence theorem it follows that  $(\mathcal{N}_R^U(u^*, x_a^*), d_{\mathcal{AG}})$  is a nonempty closed subset of  $L_{\Delta_1}^1 \times \mathbb{R}^n$  and thus  $(\mathcal{N}_R^U(u^*, x_a^*), d_{\mathcal{AG}})$  is a complete metric set. Moreover, from the continuity of  $\mathcal{F}_R(u^*, x_a^*)$  (see Prop. 5.3),  $g$ ,  $\psi$ ,  $\mathfrak{h}$ ,  $d_S^2$  and  $d_{\mathfrak{S}}^2$ , we deduce that  $J_k^R$  is continuous on  $(\mathcal{N}_R^U(u^*, x_a^*), d_{\mathcal{AG}})$  for all  $k \in \mathbb{N}$ . Clearly we have  $J_k^R(u^*, x_a^*) = \varepsilon_k$  for all  $k \in \mathbb{N}$ . As a consequence, from the Ekeland variational principle ([34], Thm. 1.1), there exists  $(u_k^R, x_{a,k}^R) \in \mathcal{N}_R^U(u^*, x_a^*)$  such that

$$\|u_k^R - u^*\|_{L_{\Delta_1}^1} + \|x_{a,k}^R - x_a^*\|_{\mathbb{R}^n} \leq \sqrt{\varepsilon_k}, \quad (6.2)$$

and

$$-\sqrt{\varepsilon_k} \left( \|u - u_k^R\|_{L_{\Delta_1}^1} + \|x_a - x_{a,k}^R\|_{\mathbb{R}^n} \right) \leq J_k^R(u, x_a) - J_k^R(u_k^R, x_{a,k}^R), \quad (6.3)$$

for all  $(u, x_a) \in \mathcal{N}_R^U(u^*, x_a^*)$  and all  $k \in \mathbb{N}$ . For every  $k \in \mathbb{N}$ , we introduce the notation  $x_k^R := x(\cdot, u_k^R, x_{a,k}^R)$ . From Proposition 5.3 and Inequality (6.2),  $(x_k^R)_k$  converges uniformly to  $x^*$  on  $[a, b]_{\mathbb{T}}$  as  $k \rightarrow +\infty$ .

We now define the elements

$$\lambda_k^R := \frac{1}{J_k^R(u_k^R, x_{a,k}^R)} \left( g(x_{a,k}^R, x_k^R(b)) - g(x_a^*, x_b^*) + \varepsilon_k \right)^+ \geq 0,$$

$$\xi_k^R := \frac{1}{J_k^R(u_k^R, x_{a,k}^R)} \left( \psi(x_{a,k}^R, x_k^R(b)) - \mathcal{P}_S(\psi(x_{a,k}^R, x_k^R(b))) \right) \in \mathbb{R}^\ell,$$

and

$$\zeta_k^R := \begin{cases} \frac{1}{J_k^R(u_k^R, x_{a,k}^R)} d_{\mathfrak{S}}(\mathfrak{h}(x_k^R)) Dd_{\mathfrak{S}}(\mathfrak{h}(x_k^R)) \in (C_{\mathbb{T}}^j)^* & \text{if } \mathfrak{h}(x_k^R) \notin \mathfrak{S}, \\ 0_{(C_{\mathbb{T}}^j)^*} \in (C_{\mathbb{T}}^j)^* & \text{if } \mathfrak{h}(x_k^R) \in \mathfrak{S}, \end{cases}$$

for all  $k \in \mathbb{N}$ . We obtain  $|\lambda_k^R|^2 + \|\xi_k^R\|_{\mathbb{R}^\ell}^2 + \|\zeta_k^R\|_{(C_{\mathbb{T}}^j)^*}^2 = 1$  and  $\xi_k^R \in N_{\mathfrak{S}}(\text{Ps}(\psi(x_{a,k}^R, x_k^R(b))))$  for all  $k \in \mathbb{N}$ .

**Remark 6.2.** In this remark (and in Rems 6.4 and 6.5), our aim is to provide two crucial inequalities satisfied by the elements  $\zeta_k^R$ . Recall that, whenever  $\mathfrak{h}(x_k^R) \notin \mathfrak{S}$ ,  $Dd_{\mathfrak{S}}(\mathfrak{h}(x_k^R))$  belongs to the subdifferential of  $d_{\mathfrak{S}}$  at the point  $\mathfrak{h}(x_k^R)$  (see [59], Thm. 3.54). As a consequence, in both cases  $\mathfrak{h}(x_k^R) \notin \mathfrak{S}$  and  $\mathfrak{h}(x_k^R) \in \mathfrak{S}$ , it holds that

$$\langle \zeta_k^R, \varphi - \mathfrak{h}(x_k^R) \rangle_{(C_{\mathbb{T}}^j)^* \times C_{\mathbb{T}}^j} \leq 0,$$

for every  $\varphi \in \mathfrak{S}$  and all  $k \in \mathbb{N}$ . Since  $\mathfrak{S}$  has a nonempty interior, there exist  $\phi \in \mathfrak{S}$  and  $\delta > 0$  such that  $\phi + \delta\varphi \in \mathfrak{S}$  for every  $\varphi \in \overline{B}_{(C_{\mathbb{T}}^j, \|\cdot\|_{C_{\mathbb{T}}^j})}(0_{C_{\mathbb{T}}^j}, 1)$ . Therefore we obtain that

$$\delta \langle \zeta_k^R, \varphi \rangle_{(C_{\mathbb{T}}^j)^* \times C_{\mathbb{T}}^j} \leq \langle \zeta_k^R, \mathfrak{h}(x_k^R) - \phi \rangle_{(C_{\mathbb{T}}^j)^* \times C_{\mathbb{T}}^j},$$

for every  $\varphi \in \overline{B}_{(C_{\mathbb{T}}^j, \|\cdot\|_{C_{\mathbb{T}}^j})}(0_{C_{\mathbb{T}}^j}, 1)$  and all  $k \in \mathbb{N}$ . We deduce that

$$\delta \|\zeta_k^R\|_{(C_{\mathbb{T}}^j)^*} = \delta \sqrt{1 - |\lambda_k^R|^2 - \|\xi_k^R\|_{\mathbb{R}^\ell}^2} \leq \langle \zeta_k^R, \mathfrak{h}(x_k^R) - \phi \rangle_{(C_{\mathbb{T}}^j)^* \times C_{\mathbb{T}}^j},$$

for all  $k \in \mathbb{N}$ .

### 6.1.2. A crucial inequality depending on $R \in \mathbb{N}$

Let  $R \in \mathbb{N}$  be fixed such that  $R \geq \|u^*\|_{L_{\Delta_1}^\infty}$  as in the previous subsection. Using the (partial) converse of the Lebesgue dominated convergence theorem and compactness arguments, we infer the existence of subsequences (that we do not relabel) such that  $(u_k^R)_k$  converges to  $u^*$   $\Delta_1$ -a.e. on  $[a, b]_{\mathbb{T}_1}$ ,  $(\lambda_k^R)_k$  converges to some  $\lambda^R \geq 0$ ,  $(\xi_k^R)_k$  converges to some  $\xi^R \in \mathbb{R}^\ell$  and  $(\zeta_k^R)_k$  weakly\* converges to some  $\zeta^R \in (C_{\mathbb{T}}^j)^*$  as  $k \rightarrow +\infty$ . In particular we obtain that  $|\lambda^R|^2 + \|\xi^R\|_{\mathbb{R}^\ell}^2 + \|\zeta^R\|_{(C_{\mathbb{T}}^j)^*}^2 \leq 1$  and  $\xi^R \in N_{\mathfrak{S}}(\psi(x_a^*, x_b^*))$  from Lemma 5.7.

Let  $(u', x'_a) \in L_{\Delta_1}^\infty \times \mathbb{R}^n$  be such that  $u'(\tau) \in U \cap \overline{B}_{\mathbb{R}^m}(0_{\mathbb{R}^m}, R)$  for  $\Delta_1$ -a.e.  $\tau \in [a, b]_{\mathbb{T}_1}$ . For a given  $k \in \mathbb{N}$  let us write  $\pi_k := (u_k^R, x_{a,k}^R, u', x'_a)$ . For every  $\rho \in [0, 1)$ , let  $u_{\pi_k}(\cdot, \rho)$  be the implicit spike variation of  $u_k^R$  associated with  $\pi_k$  (see Sect. 5.1.2). First of all, note that  $\|u_{\pi_k}(\cdot, \rho) - u^*\|_{L_{\Delta_1}^1} \leq \|u_{\pi_k}(\cdot, \rho) - u_k^R\|_{L_{\Delta_1}^1} + \|u_k^R - u^*\|_{L_{\Delta_1}^1} \leq 4R(b-a)\rho + \sqrt{\varepsilon_k} < \nu_R(u^*, x_a^*)$  for sufficiently small  $\rho \geq 0$ . Observe also that  $\|x_{a,k}^R + \rho x'_a - x_a^*\|_{\mathbb{R}^n} \leq \rho \|x'_a\|_{\mathbb{R}^n} + \|x_{a,k}^R - x_a^*\|_{\mathbb{R}^n} \leq \rho \|x'_a\|_{\mathbb{R}^n} + \sqrt{\varepsilon_k} < \nu_R(u^*, x_a^*)$  for  $\rho \geq 0$  small enough. Then, from convexity of  $U$ , it follows that  $(u_{\pi_k}(\cdot, \rho), x_{a,k}^R + \rho x'_a) \in \mathcal{N}_R^U(u^*, x_a^*)$  for sufficiently small  $\rho \geq 0$ . Thus we apply Inequality (6.3) and we obtain that

$$\begin{aligned} - (4R(b-a) + \|x'_a\|_{\mathbb{R}^n}) \sqrt{\varepsilon_k} &\leq \frac{J_k^R(u_{\pi_k}(\cdot, \rho), x_{a,k}^R + \rho x'_a) - J_k^R(u_k^R, x_{a,k}^R)}{\rho} \\ &= \frac{1}{J_k^R(u_{\pi_k}(\cdot, \rho), x_{a,k}^R + \rho x'_a) + J_k^R(u_k^R, x_{a,k}^R)} \times \frac{J_k^R(u_{\pi_k}(\cdot, \rho), x_{a,k}^R + \rho x'_a)^2 - J_k^R(u_k^R, x_{a,k}^R)^2}{\rho}, \end{aligned}$$

for  $\rho > 0$  small enough. From the continuity of  $J_k^R$ , we get  $\lim_{\rho \rightarrow 0} J_k^R(u_{\pi_k}(\cdot, \rho), x_{a,k}^R + \rho x'_a) + J_k^R(u_k^R, x_{a,k}^R) = 2J_k^R(u_k^R, x_{a,k}^R)$ . From the differentiability of  $F_{\pi_k}$  (see Prop. 5.6),  $g$ ,  $\psi$ ,  $\mathfrak{h}$ ,  $d_S^2$  and  $d_{\mathfrak{S}}^2$ , we deduce that

$$\begin{aligned} & \lim_{\rho \rightarrow 0} \frac{J_k^R(u_{\pi_k}(\cdot, \rho), x_{a,k}^R + \rho x'_a)^2 - J_k^R(u_k^R, x_{a,k}^R)^2}{\rho} \\ &= 2 \left( g(x_{a,k}^R, x_k^R(b)) - g(x_a^*, x_b^*) + \varepsilon_k \right)^+ \left\langle \nabla g(x_{a,k}^R, x_k^R(b)), \begin{pmatrix} x'_a \\ w_{\pi_k}(b) \end{pmatrix} \right\rangle_{\mathbb{R}^{2n}} \\ &+ 2 \left\langle \psi(x_{a,k}^R, x_k^R(b)) - \text{Ps}(\psi(x_{a,k}^R, x_k^R(b))), \nabla \psi(x_{a,k}^R, x_k^R(b)) \times \begin{pmatrix} x'_a \\ w_{\pi_k}(b) \end{pmatrix} \right\rangle_{\mathbb{R}^\ell} \\ &+ \left\langle 2d_{\mathfrak{S}}(\mathfrak{h}(x_k^R)) Dd_{\mathfrak{S}}(\mathfrak{h}(x_k^R)), D\mathfrak{h}(x_k^R)(w_{\pi_k}) \right\rangle_{(C_{\mathbb{T}}^j)^* \times C_{\mathbb{T}}^j}, \end{aligned}$$

with the convention that the last term is zero if  $\mathfrak{h}(x_k^R) \in \mathfrak{S}$ . Finally we obtain the inequality

$$\begin{aligned} & - (4R(b-a) + \|x'_a\|_{\mathbb{R}^n}) \sqrt{\varepsilon_k} \\ & \leq \left\langle \lambda^R \nabla g(x_{a,k}^R, x_k^R(b)) + \nabla \psi(x_{a,k}^R, x_k^R(b))^\top \times \xi_k^R, \begin{pmatrix} x'_a \\ w_{\pi_k}(b) \end{pmatrix} \right\rangle_{\mathbb{R}^{2n}} \\ & + \left\langle \zeta_k^R, \nabla_x h(x_k^R, \cdot) \times w_{\pi_k} \right\rangle_{(C_{\mathbb{T}}^j)^* \times C_{\mathbb{T}}^j}. \end{aligned} \quad (6.4)$$

To conclude this section, we need the following result.

**Lemma 6.3.** *The sequence  $(w_{\pi_k})_k$  converges uniformly to  $w_\pi$  on  $[a, b]_{\mathbb{T}}$ , where  $\pi := (u^*, x_a^*, u', x'_a)$ , as  $k \rightarrow +\infty$ .*

*Proof.* In this proof, to ease the notation, we set  $w := w_\pi$ ,  $z := z_\pi$  and  $w_k := w_{\pi_k}$ ,  $z_k := z_{\pi_k}$  for all  $k \in \mathbb{N}$ . We have

$$\begin{aligned} w_k(t) - w(t) &= \int_{[a,t]_{\mathbb{T}}} (\nabla_x f(x_k^R(\tau), u_k^{R\Box}(\tau), \tau) \times w_k(\tau) + z_k(\tau) \\ &\quad - \nabla_x f(x^*(\tau), u^{*\Box}(\tau), \tau) \times w(\tau) - z(\tau)) \Delta\tau \\ &= \int_{[a,t]_{\mathbb{T}}} \nabla_x f(x_k^R(\tau), u_k^{R\Box}(\tau), \tau) \times (w_k(\tau) - w(\tau)) \Delta\tau + \int_{[a,t]_{\mathbb{T}}} z_k(\tau) - z(\tau) \Delta\tau \\ &\quad + \int_{[a,t]_{\mathbb{T}}} (\nabla_x f(x_k^R(\tau), u_k^{R\Box}(\tau), \tau) - \nabla_x f(x^*(\tau), u^{*\Box}(\tau), \tau)) \times w(\tau) \Delta\tau, \end{aligned}$$

for every  $t \in [a, b]_{\mathbb{T}}$ . Recall that  $(x_k^R(\tau), u_k^{R\Box}(\tau), \tau)$ ,  $(x_k^R(\tau), u^{*\Box}(\tau), \tau) \in K_R(u^*, x_a^*)$  (see Prop. 5.2) for  $\Delta$ -a.e.  $\tau \in [a, b]_{\mathbb{T}}$  and that  $f$  and  $\nabla_x f$  are bounded on  $K_R(u^*, x_a^*)$  by  $L_R(u^*, x_a^*) \geq 0$ . Recall also that  $(x_k^R)_k$  converges uniformly to  $x^*$  on  $[a, b]_{\mathbb{T}}$  and  $(u_k^{R\Box})_k$  tends to  $u^{*\Box}$   $\Delta$ -a.e. on  $[a, b]_{\mathbb{T}}$ . One can conclude the proof using an argument similar to that one employed in the proof of Proposition 5.6.  $\square$

Invoking the lemma above, the  $C^1$ -regularity of  $g$ ,  $\psi$  and  $\mathfrak{h}$ , and letting  $k \rightarrow +\infty$  in Inequality (6.4), we get

$$0 \leq \left\langle \lambda^R \nabla g(x_a^*, x_b^*) + \nabla \psi(x_a^*, x_b^*)^\top \times \xi^R, \begin{pmatrix} x'_a \\ w_\pi(b) \end{pmatrix} \right\rangle_{\mathbb{R}^{2n}} + \left\langle \zeta^R, \nabla_x h(x^*, \cdot) \times w_\pi \right\rangle_{(C_{\mathbb{T}}^j)^* \times C_{\mathbb{T}}^j}. \quad (6.5)$$

**Remark 6.4.** Letting  $k \rightarrow +\infty$  in the estimates obtained in Remark 6.2, we deduce the following two crucial inequalities:

$$\langle \zeta^R, \varphi - \mathfrak{h}(x^*) \rangle_{(C_{\mathbb{T}}^j)^* \times C_{\mathbb{T}}^j} \leq 0,$$

for every  $\varphi \in \mathfrak{G}$ , and

$$\delta \sqrt{1 - |\lambda^R|^2 - \|\xi^R\|_{\mathbb{R}^\ell}^2} \leq \langle \zeta^R, \mathfrak{h}(x^*) - \phi \rangle_{(C_{\mathbb{T}}^j)^* \times C_{\mathbb{T}}^j}.$$

### 6.1.3. End of the proof of Proposition 6.1

In the previous step, we obtained Inequality (6.5) that is valid for a fixed  $R \in \mathbb{N}$  such that  $R \geq \|u^*\|_{L_{\Delta_1}^\infty}$  and for every  $(u', x'_a) \in L_{\Delta_1}^\infty \times \mathbb{R}^n$  such that  $u'(\tau) \in U \cap \overline{B_{\mathbb{R}^m}}(0_{\mathbb{R}^m}, R)$  for  $\Delta_1$ -a.e.  $\tau \in [a, b]_{\mathbb{T}_1}$ . In order to conclude the proof of Proposition 6.1, it remains to remove the dependence in  $R$ .

Since  $|\lambda^R|^2 + \|\xi^R\|_{\mathbb{R}^\ell}^2 + \|\zeta^R\|_{(C_{\mathbb{T}}^j)^*}^2 \leq 1$  for every  $R \in \mathbb{N}$  such that  $R \geq \|u^*\|_{L_{\Delta_1}^\infty}$ , from standard compactness arguments, we infer the existence of subsequences (that we do not relabel) such that  $(\lambda^R)_R$  converges to some  $\lambda \geq 0$ ,  $(\xi^R)_R$  converges to some  $\xi \in \mathbb{R}^\ell$  and  $(\zeta^R)_R$  weakly\* converges to some  $\zeta \in (C_{\mathbb{T}}^j)^*$  as  $R \rightarrow +\infty$ . In particular we have  $|\lambda|^2 + \|\xi\|_{\mathbb{R}^\ell}^2 + \|\zeta\|_{(C_{\mathbb{T}}^j)^*}^2 \leq 1$  and, from the closure of the normal cone,  $\xi \in N_{\mathfrak{S}}(\psi(x_a^*, x_b^*))$ . Notice that, at this stage, it is not guaranteed that the triplet  $(\lambda, \xi, \zeta)$  is not trivial. This is established in the next remark.

**Remark 6.5.** Letting  $R \rightarrow +\infty$  in the inequalities displayed in Remark 6.4, we obtain the crucial estimates

$$\langle \zeta, \varphi - \mathfrak{h}(x^*) \rangle_{(C_{\mathbb{T}}^j)^* \times C_{\mathbb{T}}^j} \leq 0,$$

for every  $\varphi \in \mathfrak{G}$ , and

$$\delta \sqrt{1 - |\lambda|^2 - \|\xi\|_{\mathbb{R}^\ell}^2} \leq \langle \zeta, \mathfrak{h}(x^*) - \phi \rangle_{(C_{\mathbb{T}}^j)^* \times C_{\mathbb{T}}^j}.$$

The first inequality, since  $\mathfrak{h}(x^*) = h(x^*, \cdot) \in \mathfrak{G}$ , means that  $\zeta \in N_{\mathfrak{G}}(h(x^*, \cdot))$  and the second one, since  $\delta > 0$ , ensures that the triplet  $(\lambda, \xi, \zeta)$  is not trivial.

To conclude the proof of Proposition 6.1, take  $(u', x'_a) \in L_{\Delta_1}^\infty \times \mathbb{R}^n$  such that  $u'(\tau) \in U$  for  $\Delta_1$ -a.e.  $\tau \in [a, b]_{\mathbb{T}_1}$ . Let  $R \in \mathbb{N}$  be sufficiently large in order to have  $R \geq \max(\|u^*\|_{L_{\Delta_1}^\infty}, \|u'\|_{L_{\Delta_1}^\infty})$ . Thus Inequality (6.5) holds. Letting  $R \rightarrow +\infty$ , we get the desired Inequality (6.1).

## 6.2. Introduction of appropriate costate vectors

Let  $(\lambda, \xi, \zeta)$  be the nontrivial triplet given in Proposition 6.1. The aim of this section is to conclude the proof of Theorem 3.7 by introducing appropriate costate vectors. In this section we will denote by  $\nabla_1 g$  (resp.  $\nabla_1 \psi$ ) the gradient of  $g$  (resp.  $\psi$ ) with respect to its first variable. Similarly we introduce the notation  $\nabla_2 g$  and  $\nabla_2 \psi$ . Finally we write  $\zeta = (\zeta_1, \dots, \zeta_j) \in (C_{\mathbb{T}}^j)^* = ((C_{\mathbb{T}}^1)^*)^j$  meaning that

$$\langle \zeta, \varphi \rangle_{(C_{\mathbb{T}}^j)^* \times C_{\mathbb{T}}^j} = \sum_{i=1}^j \langle \zeta_i, \varphi_i \rangle_{(C_{\mathbb{T}}^1)^* \times C_{\mathbb{T}}^1},$$

for all  $\varphi = (\varphi_1, \dots, \varphi_j) \in C_{\mathbb{T}}^j$ .

### 6.2.1. Complementary slackness condition

Since the third element  $\zeta$  of our reference triplet belongs to  $N_{\mathfrak{S}}(h(x^*, \cdot))$  (see Prop. 6.1), one can easily derive that

$$\langle \zeta_i, h_i(x^*, \cdot) \rangle_{(C_{\mathbb{T}}^1)^* \times C_{\mathbb{T}}^1} = 0 \quad \text{and} \quad \forall \varphi \in C([a, b]_{\mathbb{T}}, \mathbb{R}_+), \langle \zeta_i, \varphi \rangle_{(C_{\mathbb{T}}^1)^* \times C_{\mathbb{T}}^1} \geq 0,$$

for all  $i = 1, \dots, j$ . From the Riesz representation theorem ([64], Thm. 2.14) (which is true on any locally compact Hausdorff set, and thus on  $[a, b]_{\mathbb{T}}$ ), for each  $i \in \{1, \dots, j\}$ , there exists a unique finite nonnegative Borel measure  $d\eta_i$  on  $[a, b]_{\mathbb{T}}$  such that

$$\langle \zeta_i, \varphi \rangle_{(C_{\mathbb{T}}^1)^* \times C_{\mathbb{T}}^1} = \int_{[a, b]_{\mathbb{T}}} \varphi(\tau) d\eta_i(\tau),$$

for all  $\varphi \in C_{\mathbb{T}}^1$ . Moreover we have

$$\int_{[a, b]_{\mathbb{T}}} h_i(x^*(\tau), \tau) d\eta_i(\tau) = \langle \zeta_i, h_i(x^*, \cdot) \rangle_{(C_{\mathbb{T}}^1)^* \times C_{\mathbb{T}}^1} = 0,$$

for all  $i = 1, \dots, j$ . Since  $d\eta_i$  is a finite nonnegative Borel measure on  $[a, b]_{\mathbb{T}}$  and  $h_i(x^*, \cdot) \in C([a, b]_{\mathbb{T}}, \mathbb{R}_-)$ , we deduce the complementary slackness condition

$$\text{supp}(d\eta_i) \subset \{t \in [a, b]_{\mathbb{T}} \mid h_i(x^*(t), t) = 0\},$$

for all  $i = 1, \dots, j$ .

### 6.2.2. Adjoint equation

Let  $p \in AC([a, b]_{\mathbb{T}}, \mathbb{R}^n)$  be the unique maximal solution, which is moreover global (see [19], Thm. 6), to the backward shifted  $\Delta$ -Cauchy problem given by

$$\begin{cases} -p^{\Delta}(t) = \nabla_x f(x^*(t), u^{*\square}(t), t)^{\top} \times \left( p^{\sigma}(t) + \sum_{i=1}^j \int_{[a, t]_{\mathbb{T}}} \nabla_x h_i(x^*(\tau), \tau) d\eta_i(\tau) \right), \\ \Delta\text{-a.e. } t \in [a, b]_{\mathbb{T}}, \\ -p(b) = \lambda \nabla_2 g(x^*(a), x^*(b)) + \nabla \psi_2(x^*(a), x^*(b))^{\top} \times \xi + \sum_{i=1}^j \int_{[a, b]_{\mathbb{T}}} \nabla_x h_i(x^*(\tau), \tau) d\eta_i(\tau). \end{cases}$$

Notice that  $p$  is well-defined since the map  $t \mapsto \nabla_x f(x^*(t), u^{*\square}(t), t)^{\top}$  is clearly bounded (and thus the dynamics of the above  $\Delta$ -Cauchy problem satisfies the global Lipschitz condition of Theorem 6 in [19]) and since the map

$$t \mapsto \sum_{i=1}^j \int_{[a, t]_{\mathbb{T}}} \nabla_x h_i(x^*(\tau), \tau) d\eta_i(\tau),$$

belongs to  $BV([a, b]_{\mathbb{T}}, \mathbb{R}^n) \subset L_{\Delta}^{\infty}([a, b]_{\mathbb{T}}, \mathbb{R}^n)$ . We also introduce the function  $q \in BV([a, b]_{\mathbb{T}}, \mathbb{R}^n)$  defined by

$$q(t) := \begin{cases} p^{\sigma}(t) + \sum_{i=1}^j \int_{[a, t]_{\mathbb{T}}} \nabla_x h_i(x^*(\tau), \tau) \, d\eta_i(\tau) & \text{if } t \in [a, b]_{\mathbb{T}}, \\ p(b) + \sum_{i=1}^j \int_{[a, b]_{\mathbb{T}}} \nabla_x h_i(x^*(\tau), \tau) \, d\eta_i(\tau) & \text{if } t = b, \end{cases}$$

for all  $t \in [a, b]_{\mathbb{T}}$ . In particular the adjoint equation

$$-p^{\Delta}(t) = \nabla_x f(x^*(t), u^{*\square}(t), t)^{\top} \times q(t) = \nabla_x H(x^*(t), u^{*\square}(t), q(t), t),$$

holds true for  $\Delta$ -a.e.  $t \in [a, b]_{\mathbb{T}}$ .

### 6.2.3. Dualization

Take  $(u', x'_a) \in L_{\Delta_1}^{\infty} \times \mathbb{R}^n$  such that  $u'(\tau) \in U$  for  $\Delta_1$ -a.e.  $\tau \in [a, b]_{\mathbb{T}_1}$ , and  $\pi = (u^*, x^*(a), u', x'_a)$ . From Proposition 6.1, we know that Inequality (6.1) is satisfied. Using the definition of  $q$ , we have

$$\int_{[a, b]_{\mathbb{T}}} \langle q(\tau), w_{\pi}^{\Delta}(\tau) \rangle_{\mathbb{R}^n} \, \Delta\tau = \int_{[a, b]_{\mathbb{T}}} \langle p^{\sigma}(\tau), w_{\pi}^{\Delta}(\tau) \rangle_{\mathbb{R}^n} \, \Delta\tau + \sum_{i=1}^j \int_{[a, b]_{\mathbb{T}}} \int_{[a, \tau]_{\mathbb{T}}} \langle \nabla_x h_i(x^*(s), s), w_{\pi}^{\Delta}(\tau) \rangle_{\mathbb{R}^n} \, d\eta_i(s) \, \Delta\tau.$$

Applying the integration by parts formula (and the Leibniz formulas recalled in Sect. 2) on the first term on the right, and the Fubini-Tonelli theorem (which is true on any product of  $\sigma$ -finite measure spaces, see [64], Thm. 8.8) on the second term on the right, lead to

$$\begin{aligned} \int_{[a, b]_{\mathbb{T}}} \langle q(\tau), w_{\pi}^{\Delta}(\tau) \rangle_{\mathbb{R}^n} \, \Delta\tau &= \langle q(b), w_{\pi}(b) \rangle_{\mathbb{R}^n} - \langle p(a), w_{\pi}(a) \rangle_{\mathbb{R}^n} \\ &\quad + \int_{[a, b]_{\mathbb{T}}} \langle q(\tau), \nabla_x f(x^*(\tau), u^{*\square}(\tau), \tau) \times w_{\pi}(\tau) \rangle_{\mathbb{R}^n} \, \Delta\tau \\ &\quad - \sum_{i=1}^j \int_{[a, b]_{\mathbb{T}}} \langle \nabla_x h_i(x^*(s), s), w_{\pi}(s) \rangle_{\mathbb{R}^n} \, d\eta_i(s). \end{aligned}$$

On the other hand, using the definition of  $w_{\pi}^{\Delta}$  (see Sect. 5.1.2), we get

$$\int_{[a, b]_{\mathbb{T}}} \langle q(\tau), w_{\pi}^{\Delta}(\tau) \rangle_{\mathbb{R}^n} \, \Delta\tau = \int_{[a, b]_{\mathbb{T}}} \langle q(\tau), \nabla_x f(x^*(\tau), u^{*\square}(\tau), \tau) \times w_{\pi}(\tau) + z_{\pi}(\tau) \rangle_{\mathbb{R}^n} \, \Delta\tau.$$

From these two equalities we deduce that

$$\begin{aligned} \langle \zeta, \nabla_x h(x^*, \cdot) \times w_{\pi} \rangle_{(C_{\mathbb{T}}^j)^* \times C_{\mathbb{T}}^j} &= \sum_{i=1}^j \int_{[a, b]_{\mathbb{T}}} \langle \nabla_x h_i(x^*(s), s), w_{\pi}(s) \rangle_{\mathbb{R}^n} \, d\eta_i(s) \\ &= \langle q(b), w_{\pi}(b) \rangle_{\mathbb{R}^n} - \langle p(a), w_{\pi}(a) \rangle_{\mathbb{R}^n} - \int_{[a, b]_{\mathbb{T}}} \langle q(\tau), z_{\pi}(\tau) \rangle_{\mathbb{R}^n} \, \Delta\tau. \end{aligned}$$

Using this last relation in Inequality (6.1), from the definitions of  $w_\pi(a)$  and  $q(b)$ , we obtain

$$0 \leq \langle \lambda \nabla_1 g(x^*(a), x^*(b)) + \nabla_1 \psi(x^*(a), x^*(b))^\top \times \xi - p(a), x'_a \rangle_{\mathbb{R}^n} - \int_{[a,b]_{\mathbb{T}}} \langle q(\tau), z_\pi(\tau) \rangle_{\mathbb{R}^n} \Delta\tau. \quad (6.6)$$

#### 6.2.4. Transversality condition

Inequality (6.6) is satisfied for all  $(u', x'_a) \in L^\infty_{\Delta_1} \times \mathbb{R}^n$  such that  $u'(\tau) \in U$  for  $\Delta_1$ -a.e.  $\tau \in [a, b]_{\mathbb{T}_1}$ . In particular, it is satisfied for  $u' = u^*$  (and thus  $z_\pi \equiv 0_{\mathbb{R}^n}$ ) and for all  $x'_a \in \mathbb{R}^n$ . We deduce that

$$p(a) = \lambda \nabla_1 g(x^*(a), x^*(b)) + \nabla_1 \psi(x^*(a), x^*(b))^\top \times \xi.$$

Since

$$-q(b) = \lambda \nabla_2 g(x^*(a), x^*(b)) + \nabla_2 \psi(x^*(a), x^*(b))^\top \times \xi,$$

we deduce the transversality condition

$$\begin{pmatrix} p(a) \\ -q(b) \end{pmatrix} = \lambda \nabla g(x^*(a), x^*(b)) + \nabla \psi(x^*(a), x^*(b))^\top \times \xi,$$

where  $\xi \in N_S(\psi(x^*(a), x^*(b)))$ .

#### 6.2.5. Nontriviality

Assume by contradiction that  $(\lambda, p, d\eta_1, \dots, d\eta_j) = 0$ . From the triviality of  $d\eta_1, \dots, d\eta_j$ , we deduce that  $\zeta = 0_{(\mathbb{C}_T^j)^*}$  and  $q(b) = p(b)$ . From the transversality condition and the relation  $p(a) = p(b) = 0_{\mathbb{R}^n}$ , we would obtain that  $\nabla \psi(x^*(a), x^*(b))^\top \times \xi = 0_{\mathbb{R}^{2n}}$ . Since  $\psi$  is submersive at  $(x^*(a), x^*(b))$ , we conclude that  $\xi = 0_{\mathbb{R}^l}$  raising a contradiction with the nontriviality of the triplet  $(\lambda, \xi, \zeta)$  provided in Proposition 6.1.

#### 6.2.6. Nonpositive averaged Hamiltonian gradient condition at right-scattered points

If  $RS_1^b = \emptyset$ , there is nothing to prove. Thus we continue assuming that  $RS_1^b \neq \emptyset$ . Let  $r \in RS_1^b$  and  $v \in U$ . From Inequality (6.6) and the transversality condition, we know that

$$\int_{[a,b]_{\mathbb{T}}} \langle q(\tau), z_\pi(\tau) \rangle_{\mathbb{R}^n} \Delta\tau \leq 0,$$

for all  $u' \in L^\infty_{\Delta_1}$  such that  $u'(\tau) \in U$  for  $\Delta_1$ -a.e.  $\tau \in [a, b]_{\mathbb{T}_1}$ . Consider in particular the control function

$$u'(\tau) := \begin{cases} u^*(\tau) & \text{if } \tau \neq r, \\ v & \text{if } \tau = r, \end{cases}$$

for  $\Delta_1$ -a.e.  $\tau \in [a, b]_{\mathbb{T}_1}$ . We deduce that

$$\int_{[r, \sigma_1(r)]_{\mathbb{T}}} \langle q(\tau), \nabla_u f(x^*(\tau), u^*(r), \tau) \times (v - u^*(r)) \rangle_{\mathbb{R}^n} \Delta\tau \leq 0,$$



which exactly coincides with the nonpositive averaged Hamiltonian gradient condition given by

$$\left\langle \int_{[r, \sigma_1(r)]_{\mathbb{T}}} \nabla_u H(x^*(\tau), u^*(r), q(\tau), \tau) \Delta\tau, v - u^*(r) \right\rangle_{\mathbb{R}^m} \Delta\tau \leq 0.$$

### 6.2.7. Hamiltonian maximization condition at right-dense points

If  $\mu_{\Delta_1}(\text{RD}_1^b) = \mu_{\mathcal{L}}(\text{RD}_1^b) = 0$ , there is nothing to prove. Then we assume that  $\mu_{\Delta_1}(\text{RD}_1^b) = \mu_{\mathcal{L}}(\text{RD}_1^b) > 0$ . Once again, from Inequality (6.6) and the transversality condition, we have

$$\int_{[a, b]_{\mathbb{T}}} \langle q(\tau), z_{\pi}(\tau) \rangle_{\mathbb{R}^n} \Delta\tau \leq 0, \quad (6.7)$$

for all  $u' \in L_{\Delta_1}^{\infty}$  such that  $u'(\tau) \in U$  for  $\Delta_1$ -a.e.  $\tau \in [a, b]_{\mathbb{T}_1}$ . Now we consider control functions of the following type:

$$u'(\tau) := \begin{cases} u''(\tau) & \text{if } \tau \in \text{RD}_1^b, \\ u^*(\tau) & \text{if } \tau \in \text{RS}_1^b, \end{cases}$$

for  $\Delta_1$ -a.e.  $\tau \in [a, b]_{\mathbb{T}_1}$ , where  $u'' \in L^{\infty}(\text{RD}_1^b, \mathbb{R}^m)$  is such that  $u''(\tau) \in U$  for  $\mu_{\mathcal{L}}$ -a.e.  $\tau \in \text{RD}_1^b$ . For this class of control functions, when we consider the associated term  $z_{\pi}$ , bearing in mind that  $\text{RD}_1^b \subset \text{RD}$  and  $\mu_{\Delta_1} = \mu_{\Delta} = \mu_{\mathcal{L}}$  on  $\text{RD}_1^b$ , from (6.7) it follows that

$$\int_{\text{RD}_1^b} \langle q(\tau), f(x^*(\tau), u''(\tau), \tau) - f(x^*(\tau), u^*(\tau), \tau) \rangle_{\mathbb{R}^n} d\tau \leq 0.$$

Fix any  $v \in U$  and let  $s \in \text{RD}_1^b$  be a  $\mu_{\mathcal{L}}$ -density point of  $\text{RD}_1^b$  with  $s > a$ . We can also assume that  $s$  is simultaneously a continuous point of  $q$  (restricted to  $\text{RD}_1^b$ ) and a  $\mu_{\mathcal{L}}$ -Lebesgue point of the map

$$\tau \in [a, b] \mapsto \begin{cases} \langle q(\tau), f(x^*(\tau), u^*(\tau), \tau) \rangle_{\mathbb{R}^n} & \text{if } \tau \in \text{RD}_1^b, \\ 0 & \text{if } \tau \in [a, b] \setminus \text{RD}_1^b, \end{cases}$$

which belongs to  $L^{\infty}([a, b], \mathbb{R})$ . Taking the particular choice

$$u''(\tau) := \begin{cases} v & \text{if } \tau \in [s - \varepsilon, s + \varepsilon] \cap \text{RD}_1^b, \\ u^*(\tau) & \text{if } \tau \in \text{RD}_1^b \setminus [s - \varepsilon, s + \varepsilon], \end{cases}$$

for  $\mu_{\mathcal{L}}$ -a.e.  $\tau \in \text{RD}_1^b$  and, for  $\varepsilon > 0$  small enough, we get

$$\frac{1}{\mu_{\mathcal{L}}([s - \varepsilon, s + \varepsilon] \cap \text{RD}_1^b)} \int_{[s - \varepsilon, s + \varepsilon] \cap \text{RD}_1^b} \langle q(\tau), f(x^*(\tau), v, \tau) - f(x^*(\tau), u^*(\tau), \tau) \rangle_{\mathbb{R}^n} d\tau \leq 0.$$

Letting  $\varepsilon \rightarrow 0^+$ , from the assumptions on  $s$ , we obtain

$$\langle q(s), f(x^*(s), v, s) - f(x^*(s), u^*(s), s) \rangle_{\mathbb{R}^n} \leq 0.$$

Then we deduce the Hamiltonian maximization condition, given by

$$u^*(s) \in \operatorname{argmax}_{v \in U} H(x^*(s), v, q(s), s),$$

for  $\Delta_1$ -a.e.  $s \in \text{RD}_1^b$ .

### APPENDIX A. PROOF OF LEMMA 5.4

Recall that the Lebesgue measure  $\mu_{\mathcal{L}}$  on  $\mathbb{R}$  is a nonatomic measure (see, e.g., [36], Rem. 1.161). As a consequence, from the classical Sierpinski (or Lyapunov) theorem (see [67] or [37, p. 37]), for any  $\mu_{\mathcal{L}}$ -measurable set  $R \subset \text{RD}_1^b$  and any  $\rho \in (0, 1)$ , there exists a  $\mu_{\mathcal{L}}$ -measurable set  $R_\rho \subset R$  such that  $\mu_{\mathcal{L}}(R_\rho) = \rho\mu_{\mathcal{L}}(R)$ .

Let  $\rho \in (0, 1)$ .

**Lemma A.1.** *Let  $\phi : [a, b] \times \text{RD}_1^b \rightarrow \mathbb{R}^n$  be defined by*

$$\phi(t, \tau) := z(\tau)\mathbf{1}_{[a, t]}(\tau).$$

*Then the map  $t \mapsto \phi(t, \cdot)$  belongs to  $C([a, b], L^1(\text{RD}_1^b, \mathbb{R}^n))$ .*

*Proof.* Take any  $t \in [a, b]$  and let  $(t_k)_k \subset [a, b]$  be a decreasing sequence such that  $t_k \rightarrow t$ . Then we have

$$\begin{aligned} \|\phi(t_k, \cdot) - \phi(t, \cdot)\|_{L^1(\text{RD}_1^b, \mathbb{R}^n)} &= \int_{\text{RD}_1^b} \|\phi(t_k, \tau) - \phi(t, \tau)\|_{\mathbb{R}^n} d\tau \\ &= \int_{[a, t] \cap \text{RD}_1^b} \|\phi(t_k, \tau) - \phi(t, \tau)\|_{\mathbb{R}^n} d\tau + \int_{[t, t_k] \cap \text{RD}_1^b} \|\phi(t_k, \tau) - \phi(t, \tau)\|_{\mathbb{R}^n} d\tau \\ &= \int_{[t, t_k] \cap \text{RD}_1^b} \|z(\tau)\|_{\mathbb{R}^n} d\tau \rightarrow 0. \end{aligned}$$

Similarly, if  $t \in (a, b]$ , then one can prove that  $\|\phi(t_k, \cdot) - \phi(t, \cdot)\|_{L^1(\text{RD}_1^b, \mathbb{R}^n)} \rightarrow 0$  for any increasing sequence  $(t_k)_k \subset [a, b]$  such that  $t_k \rightarrow t$ . The proof is complete.  $\square$

Since  $[a, b]$  is compact, there exists  $\delta > 0$  such that  $\|\phi(t, \cdot) - \phi(\bar{t}, \cdot)\|_{L^1(\text{RD}_1^b, \mathbb{R}^n)} \leq \frac{\rho^2}{2(\rho+1)}$  for all  $t, \bar{t} \in [a, b]$  satisfying  $|t - \bar{t}| < \delta$ . Now we fix a partition  $a = t_0 < t_1 < \dots < t_N = b$  such that  $|t_{k+1} - t_k| < \delta$  for all  $k = 0, \dots, N-1$  and we define

$$\Phi(\cdot) := \left( \phi(t_0, \cdot), \phi(t_1, \cdot), \dots, \phi(t_{N-1}, \cdot) \right) \in L^1(\text{RD}_1^b, (\mathbb{R}^n)^N).$$

**Lemma A.2.** *There exists a  $\mu_{\mathcal{L}}$ -measurable set  $Q_\rho \subset \text{RD}_1^b$  such that  $\mu_{\mathcal{L}}(Q_\rho) = \rho\mu_{\mathcal{L}}(\text{RD}_1^b)$  and*

$$\left\| \int_{\text{RD}_1^b} \left( 1 - \frac{1}{\rho} \mathbf{1}_{Q_\rho}(\tau) \right) \Phi(\tau) d\tau \right\|_{(\mathbb{R}^n)^N} \leq \frac{\rho}{2}.$$

*Proof.* Since  $\Phi \in L^1(\text{RD}_1^b, (\mathbb{R}^n)^N)$ , there exists a simple function  $J : \text{RD}_1^b \rightarrow (\mathbb{R}^n)^N$  such that  $\int_{\text{RD}_1^b} \|\Phi(\tau) - J(\tau)\|_{(\mathbb{R}^n)^N} d\tau \leq \frac{\rho^2}{2(\rho+1)}$ . Set  $J := \sum_{k=1}^K c_k \mathbf{1}_{R^k}$ , where  $c_k \in (\mathbb{R}^n)^N$  and  $R^k \subset \text{RD}_1^b$  are  $\mu_{\mathcal{L}}$ -measurable sets such that  $\bigsqcup_{k=1}^K R^k = \text{RD}_1^b$ . Since  $\mu_{\mathcal{L}}$  is nonatomic, there exist  $\mu_{\mathcal{L}}$ -measurable sets  $R_\rho^k \subset R^k$  such that  $\mu_{\mathcal{L}}(R_\rho^k) = \rho\mu_{\mathcal{L}}(R^k)$  for all  $k = 1, \dots, K$ . Let us define  $Q_\rho := \bigsqcup_{k=1}^K R_\rho^k \subset \text{RD}_1^b$ . Note that  $\mu_{\mathcal{L}}(Q_\rho) = \rho\mu_{\mathcal{L}}(\text{RD}_1^b)$ . Moreover we obtain

$$\left\| \int_{\text{RD}_1^b} \left( 1 - \frac{1}{\rho} \mathbf{1}_{Q_\rho}(\tau) \right) \Phi(\tau) d\tau \right\|_{(\mathbb{R}^n)^N}$$

$$\leq \left\| \int_{\mathbb{RD}_1^b} \left(1 - \frac{1}{\rho} \mathbf{1}_{Q_\rho}(\tau)\right) J(\tau) \, d\tau \right\|_{(\mathbb{R}^n)^N} + \left\| \int_{\mathbb{RD}_1^b} \left(1 - \frac{1}{\rho} \mathbf{1}_{Q_\rho}(\tau)\right) (\Phi(\tau) - J(\tau)) \, d\tau \right\|_{(\mathbb{R}^n)^N}.$$

It is easy to see that the second integral on the right is bounded by  $\frac{\rho}{2}$ , and the first one is equal to

$$\left\| \sum_{k=1}^K c_k \int_{\mathbb{RD}_1^b} \left( \mathbf{1}_{\mathbb{R}^k}(\tau) - \frac{1}{\rho} \mathbf{1}_{Q_\rho \cap \mathbb{R}^k}(\tau) \right) \, d\tau \right\|_{(\mathbb{R}^n)^N} = \left\| \sum_{k=1}^K c_k \left( \mu_{\mathcal{L}}(\mathbb{R}^k) - \frac{1}{\rho} \mu_{\mathcal{L}}(\mathbb{R}_\rho^k) \right) \right\|_{(\mathbb{R}^n)^N} = 0.$$

The proof is complete.  $\square$

We are now ready to conclude the proof of Lemma 5.4. Let  $t \in [a, b]$ . There exists  $k \in \{0, \dots, N-1\}$  such that  $t \in [t_k, t_{k+1}]$ . In particular we have  $|t - t_k| < \delta$  and thus  $\|\phi(t, \cdot) - \phi(t_k, \cdot)\|_{L^1(\mathbb{RD}_1^b, \mathbb{R}^n)} \leq \frac{\rho^2}{2(\rho+1)}$  (see remark after the proof of Lem. A.1). It follows that

$$\begin{aligned} \left\| \int_{[a,t] \cap \mathbb{RD}_1^b} \left(1 - \frac{1}{\rho} \mathbf{1}_{Q_\rho}(\tau)\right) z(\tau) \, d\tau \right\|_{\mathbb{R}^n} &= \left\| \int_{\mathbb{RD}_1^b} \left(1 - \frac{1}{\rho} \mathbf{1}_{Q_\rho}(\tau)\right) \phi(t, \tau) \, d\tau \right\|_{\mathbb{R}^n} \\ &\leq \left\| \int_{\mathbb{RD}_1^b} \left(1 - \frac{1}{\rho} \mathbf{1}_{Q_\rho}(\tau)\right) (\phi(t, \tau) - \phi(t_k, \tau)) \, d\tau \right\|_{\mathbb{R}^n} + \left\| \int_{\mathbb{RD}_1^b} \left(1 - \frac{1}{\rho} \mathbf{1}_{Q_\rho}(\tau)\right) \phi(t_k, \tau) \, d\tau \right\|_{\mathbb{R}^n}. \end{aligned}$$

The second term on the right of the last expression can be bounded by

$$\left\| \int_{\mathbb{RD}_1^b} \left(1 - \frac{1}{\rho} \mathbf{1}_{Q_\rho}(\tau)\right) \Phi(\tau) \, d\tau \right\|_{(\mathbb{R}^n)^N} \leq \frac{\rho}{2},$$

(see Lem. A.2), and clearly the first term is bounded by  $\frac{\rho}{2}$ . The proof of Lemma 5.4 is complete.

## REFERENCES

- [1] J.E. Ackermann, *Sampled-Data Control Systems: Analysis and Synthesis, Robust System Design*. Springer-Verlag Berlin Heidelberg (1985).
- [2] R.P. Agarwal, V. Otero-Espinar, K. Perera and D.R. Vivero, Basic properties of Sobolev's spaces on time scales. *Adv. Differ. Equ.* **14** (2006) 38121.
- [3] R.P. Agarwal and M. Bohner, Basic calculus on time scales and some of its applications. *Results Math.* **35** (1999) 3–22.
- [4] K.J. Aström, On the choice of sampling rates in optimal linear systems. *IBM Res.: Eng. Stud.* (1963).
- [5] T. Bakir, B. Bonnard, L. Bourdin and J. Rouot, Pontryagin-type conditions for optimal muscular force response to functional electric stimulations. *J. Optim. Theory Appl.* **184** (2020) 581–602.
- [6] B. Bamieh and J.B. Pearson, The  $\mathcal{H}_2$  problem for sampled-data systems. *Syst. Control Lett.* **19** (1992) 1–12.
- [7] Z. Bartosiewicz and D.F.M. Torres, Noether's theorem on time scales. *J. Math. Anal. Appl.* **342** (2008) 1220–1226.
- [8] M. Bohner, Calculus of variations on time scales. *Dyn. Syst. Appl.* **13** (2004) 339–349.
- [9] M. Bohner and A. Peterson, *Dynamic equations on time scales. An introduction with applications*. Birkhäuser Boston Inc., Boston, MA (2001).
- [10] M. Bohner and A. Peterson, *Advances in dynamic equations on time scales*. Birkhäuser Boston Inc., Boston, MA (2003).
- [11] M. Bohner, K. Kenzhebaev, O. Lavrova and O. Stanzhytskyi, Pontryagin's maximum principle for dynamic systems on time scales. *J. Differ. Equ. Appl.* **3** (2017) 1161–1189.
- [12] V.G. Boltyanskii, *Optimal control of discrete systems*. John Wiley & Sons, New York-Toronto, Ont. (1978).
- [13] J.F. Bonnans and C. De La Vega, Optimal control of state constrained integral equations. *Set-Valued Anal.* **18** (2010) 307–326.
- [14] B. Bonnard, L. Faubourg, G. Launay and E. Trélat, Optimal control with state constraints and the space shuttle re-entry problem. *J. Dynam. Control Syst.* **9** (2003) 155–199.
- [15] L. Bourdin, Nonshifted calculus of variations on time scales with nabla-differentiable sigma. *J. Math. Anal. Appl.* **411** (2014) 543–554.

- [16] L. Bourdin, Note on Pontryagin maximum principle with running state constraints and smooth dynamics – proof based on the Ekeland variational principle. *Res. Notes* (2016).
- [17] L. Bourdin and G. Dhar, Continuity/constancy of the Hamiltonian function in a Pontryagin maximum principle for optimal sampled-data control problems with free sampling times. *Math. Control Signals Syst.* **31** (2019) 503–544.
- [18] L. Bourdin and G. Dhar, Optimal sampled-data controls with running inequality state constraints: Pontryagin maximum principle and bouncing trajectory phenomenon. To appear in: *Math. Program., Ser. A* (2020). <https://doi.org/10.1007/s10107-020-01574-2>
- [19] L. Bourdin and E. Trélat, General Cauchy-Lipschitz theory for Delta-Cauchy problems with Carathéodory dynamics on time scales. *J. Differ. Equ. Appl.* **20** (2014) 526–547.
- [20] L. Bourdin and E. Trélat, Pontryagin Maximum Principle for finite dimensional nonlinear optimal control problems on time scales. *SIAM J. Control Optim.* **51** (2013) 3781–3813.
- [21] L. Bourdin and E. Trélat, Optimal sampled-data control, and generalizations on time scales. *Math. Control Relat. Fields* **6** (2016) 53–94.
- [22] L. Bourdin and E. Trélat, Pontryagin Maximum Principle for optimal sampled-data control problems. In *proceedings of the IFAC Workshop CAO* (2015).
- [23] L. Bourdin, O. Stanzhytskyi and E. Trélat, Addendum to “Pontryagin’s maximum principle for dynamic systems on time scales”. *J. Differ. Equ. Appl.* **23** (2017) 1760–1763.
- [24] A. Bressan and B. Piccoli, Introduction to the mathematical theory of control. Vol. 2 of *AIMS Series on Applied Mathematics*. American Institute of Mathematical Sciences (AIMS), Springfield, MO (2007).
- [25] A. Cabada and D.R. Vivero, Criteria for absolute continuity on time scales. *J. Differ. Equ. Appl.* **11** (2005) 1013–1028.
- [26] A. Cabada and D.R. Vivero, Expression of the Lebesgue  $\Delta$ -integral on time scales as a usual Lebesgue integral: application to the calculus of  $\Delta$ -antiderivatives. *Math. Comput. Model.* **43** (2006) 194–207.
- [27] T. Chen and B. Francis, Optimal sampled-data control systems. Springer-Verlag London, Ltd., London (1996).
- [28] F.H. Clarke, Functional analysis, calculus of variations and optimal control. Vol. 264 of *Graduate Texts in Mathematics*. Springer Science & Business Media (2013).
- [29] O. Cots, Geometric and numerical methods for a state constrained minimum time control problem of an electric vehicle. *ESAIM: COCV* **23** (2017) 1715–1749.
- [30] O. Cots, J. Gergaud and D. Goubinat, Direct and indirect methods in optimal control with state constraints and the climbing trajectory of an aircraft. *Opt. Control Appl. Methods* **39** (2018) 281–301.
- [31] A. Dmitruk, On the development of Pontryagin’s maximum principle in the works of A. Ya. Dubovitskii and A. A. Milyutin. *Control Cybern.* **4A** (2009) 923–957.
- [32] A.V. Dmitruk and N.P. Osmolovskii, Proof of the maximum principle for a problem with state constraints by the V-change of time variable. *Discrete Contin. Dyn. Syst. Ser. B* **24** (2019) 2189–2204.
- [33] A.Y. Dubovitskii and A.A. Milyutin, Extremum problems in the presence of restrictions. *USSR Comput. Math. Math. Phys.* **5** (1965) 1–80.
- [34] I. Ekeland, On the variational principle. *J. Math. Anal. Appl.* **47** (1974) 324–353.
- [35] H.O. Fattorini, Infinite-dimensional optimization and control theory. Encyclopedia of Mathematics and its Applications, Cambridge University Press, Cambridge (1999).
- [36] I. Fonseca and G. Leoni, Modern methods in the calculus of variations:  $L^p$  spaces. Springer (2007).
- [37] A. Fryszkowski, Fixed point theory for decomposable sets. Springer Netherlands (2004).
- [38] R.V. Gamkrelidze, Optimal control processes for bounded phase coordinates. *Izv. Akad. Nauk SSSR. Ser. Mat.* **24** (1960) 315–356.
- [39] J.C. Geromel and M. Souza, On an LMI approach to optimal sampled-data state feedback control design. *Internat. J. Control* **88** (2015) 2369–2379.
- [40] I.V. Girsanov, Lectures on mathematical theory of extremum problems. Vol. 67 of *Lecture Notes in Economics and Mathematical Systems*, edited by B.T. Poljak. Translated from the Russian by D. Louvish. Springer-Verlag, Berlin-New York (1972).
- [41] G.S. Guseinov, Integration on time scales. *J. Math. Anal. Appl.* **285** (2003) 107–127.
- [42] S. Hilger, *Ein Maßkettenkalkül mit Anwendungen auf Zentrumsmannigfaltigkeiten*. PhD thesis, Universität Würzburg (1988).
- [43] R. Hilscher and V. Zeidan, Calculus of variations on time scales: weak local piecewise  $C_{rd}^1$  solutions with variable endpoints. *J. Math. Anal. Appl.* **289** (2004) 143–166.
- [44] R. Hilscher and V. Zeidan, First-order conditions for generalized variational problems over time scales. *Comput. Math. Appl.* **62** (2011) 3490–3503.
- [45] R. Hilscher and V. Zeidan, Weak maximum principle and accessory problem for control problems on time scales. *Nonlinear Anal.* **70** (2009) 3209–3226.
- [46] R. Hilscher and V. Zeidan, Time scale embedding theorem and coercivity of quadratic functionals. *Analysis (Munich)* **28** (2008) 1–28.
- [47] R.F. Hartl, S.P. Sethi and R.G. Vickson, A survey of the maximum principles for optimal control problems with state constraints. *SIAM Rev.* **37** (1995) 181–218.
- [48] J.M. Holtzman and H. Halkin, Directional convexity and the maximum principle for discrete systems. *SIAM J. Control* **4** (1966) 263–275.
- [49] A. Huseynov, The Riesz representation theorem on time scales. *Math. Comput. Model.* **55** (2012) 1570–1579.

- [50] D.H. Jacobson, M.M. Lele and J.L. Speyer, New necessary conditions of optimality for control problems with state-variable inequality constraints. *J. Math. Anal. Appl.* **35** (1971) 255–284.
- [51] I.D. Landau, Digital Control Systems. Springer (2006).
- [52] E.B. Lee and L. Markus, Foundations of optimal control theory. Robert E. Krieger Publishing Co., Inc., Melbourne, FL, second edition (1986).
- [53] A. Levis, R. Schlueter and M. Athans, On the behavior of optimal linear sampled-data regulators. *Int. J. Control* **13** (1971) 343–361.
- [54] X. Li and J. Yong, Optimal control theory for infinite dimensional systems. Birkhäuser Boston (1995).
- [55] H. Maurer, On optimal control problems with bounded state variables and control appearing linearly. *SIAM J. Control Optim.* **15** (1977) 345–362.
- [56] H. Maurer, J.R. Kim and G. Vossen, On a state-constrained control problem in optimal production and maintenance. In *Optimal Control and Dynamic Games*. Springer (2005) 289–308.
- [57] S.M. Melzer and B.C. Kuo, Sampling period sensitivity of the optimal sampled data linear regulator. *Automatica J. IFAC* **7** (1971) 367–370.
- [58] R.H. Middleton and G.C. Goodwin, Digital control and estimation: A unified approach (1990).
- [59] B. Mordukhovich, Variational Analysis and Generalized Differentiation I. Springer-Verlag, Berlin Heidelberg (2006).
- [60] D. Nešić and A.R. Teel, Sampled-data control of nonlinear systems: an overview of recent results. *Perspect. Robust Control* **268** (2001) 221–239.
- [61] L.S. Pontryagin, V.G. Boltyanskii, R.V. Gamkrelidze and E.F. Mishchenko, The mathematical theory of optimal processes. John Wiley & Sons, Inc. (1962).
- [62] B.N. Pshenichnyi, Necessary conditions for an extremum. Marcel Dekker, Inc., New York (1971).
- [63] A. Puchkova, V. Rehbock and K.L. Teo, Closed-form solutions of a fishery harvesting model with state constraint. *Optimal Control Appl. Methods* **35** (2014) 395–411.
- [64] W. Rudin, Real and complex analysis. 3rd ed. McGraw-Hill, New York, NY. (1987).
- [65] M. Salgado, R. Middleton and G.C. Goodwin, Connection between continuous and discrete Riccati equations with applications to Kalman filtering. *Proc. IEE-D* **135** (1988) 28–34.
- [66] S.P. Sethi and G.L. Thompson, Optimal Control Theory. Applications to Management Science and Economics. Kluwer Academic Publishers, Boston, MA, second edition (2000).
- [67] W. Sierpinski, Sur les fonctions d'ensemble additives et continues. *Fundam. Math.* **3** (1922) 240–246.
- [68] M. Souza, G.W.G. Vital and J.C. Geromel, Optimal sampled data state feedback control of linear systems. In *Proceedings of the 19th World Congress The International Federation of Automatic Control* (2014).
- [69] E. Trélat, Contrôle optimal : théorie & applications. Vuibert Paris (2005).
- [70] T. van Keulen, J. Gillot, B. de Jager and M. Steinbuch, Solution for state constrained optimal control problems applied to power split control for hybrid vehicles. *Automatica J. IFAC* **50** (2014) 187–192.
- [71] R.B. Vinter, Optimal Control. Birkhäuser, Boston (2000).