Automatica 79 (2017) 273–281

Contents lists available at ScienceDirect

Automatica

journal homepage: www.elsevier.com/locate/automatica

Brief paper

Linear–quadratic optimal sampled-data control problems: Convergence result and Riccati theory[☆]

Loïc Bourdin[a], Emmanuel Trélat[b]

[a] Université de Limoges, Institut de recherche XLIM, Pôle Mathématiques. CNRS UMR 7252. Limoges, France

[b] Sorbonne Universités, UPMC Univ Paris 06, CNRS UMR 7598, Laboratoire Jacques-Louis Lions, F-75005, Paris, France

ARTICLE INFO

Article history:
Received 19 April 2016
Received in revised form
10 October 2016
Accepted 5 January 2017

Keywords:
Optimal control
Linear–quadratic problems
Sampled-data control
Digital control
Convergence
Pontryagin maximum principle
Riccati theory
Dynamic programming principle

ABSTRACT

We consider linear–quadratic optimal sampled-data control problems, where the state evolves continuously in time according to a linear control system and the control is sampled, *i.e.*, is piecewise constant over a subdivision of the time interval, and the cost is quadratic. As a first result, we prove that, as the sampling periods tend to zero, the optimal sampled-data controls converge pointwise to the optimal permanent control. Then, we extend the classical Riccati theory to the sampled-data control framework, by developing two different approaches: the first one uses a recently established version of the Pontryagin maximum principle for optimal sampled-data control problems, and the second one uses an adequate version of the dynamic programming principle. In turn, we obtain a closed-loop expression for optimal sampled-data controls of linear–quadratic problems.

1. Introduction

In optimal control theory, the classical version of the Pontryagin Maximum Principle (in short, PMP) that can be found in Pontryagin, Boltyanskii, Gamkrelidze, and Mishchenko (1962) is concerned with optimal *permanent control* problems, that is, when the control can be modified at any instant of time. In many problems, achieving the optimal trajectory requires a permanent modification of the control. However, such a requirement is not conceivable in practice for human beings, even for numerical devices. Therefore, *sampled-data controls* (or *digital controls*), for which only a finite number of modifications is allowed, are usually considered for engineering issues. The fixed switching times at which (and only at which) the sampled-data controls can be modified are called *sampling times*.

Since the 60s, many authors have addressed sampled-data control systems, as evidenced by numerous references and books (see, e.g., Ackermann, 1983; Ackermann, 1985; Fadali & Visioli, 2013; Isermann, 1989; Landau, 2006; Nesić & Teel, 2001; Ragazzini, 1958; Tou, 1963 and references therein), and it is still an active research topic. A significant part of the literature is concerned with $\mathcal{H}_2$-$\mathcal{H}_\infty$ optimization theory (see, e.g., Bamieh & Pearson, 1992; Chen & Francis, 1996; Geromel & Souza, 2015; Khargonekar & Sivashankar, 1991; Mirkin, Rotstein, & Palmor, 1999; Souza, Vital, & Geromel, 2014), but many numerical and theoretical treatments of other optimization criteria are also addressed (see, e.g., Aida-Zadeh & Rahimov, 2007; Azhmyakov, Basin, & Reincke-Collon, 2014; Bini & Buttazzo, 2009, 2014; Imura, 2005).

In this paper, we establish some new related results with a novel approach based on a recently established version of the PMP in Bourdin and Trélat (2015) and Bourdin and Trélat (2016) that can be applied to optimal sampled-data control problems.[1] The application of this PMP to a Linear–Quadratic Optimal Control Problem (in short, LQOCP) with sampled-data controls yields an optimal sampled-data control expressed as an *open-loop control* (see Proposition 3).

☆ The second author was partially supported by the Grant FA9550-14-1-0214 of the EOARD-AFOSR. The material in this paper was not presented at any conference. This paper was recommended for publication in revised form by Associate Editor Tongwen Chen under the direction of Editor Ian R. Petersen.

E-mail addresses: loic.bourdin@unilim.fr (L. Bourdin), emmanuel.trelat@upmc.fr (E. Trélat).

http://dx.doi.org/10.1016/j.automatica.2017.02.013
0005-1098/© 2017 Elsevier Ltd. All rights reserved.

[1] Actually in Bourdin and Trélat (2016) one can find a PMP for general nonlinear optimal nonpermanent control problems settled on *time scales*, which unifies and extends continuous-time and discrete-time issues (see also Bourdin & Trélat, 2014; Bourdin & Trélat, 2013).

Our main results in this paper (Theorems 1–2) are based on this new expression and deal with the following issues:

(1) The first question concerns the pointwise convergence of optimal sampled-data controls of LQOCPs to the optimal permanent control as the distances between consecutive sampling times tend to zero. Theorem 1 (Section 2) gives a positive answer to this question. This strong convergence of the control is derived from the PMP of Bourdin and Trélat (2016) and from the convergence of the costate. In turn, we actually prove the strong convergence of the whole minimization problem: not only on the performance, state, control but also on the costate (see Remark 3), which allows us to derive a strong convergence of the control (not only weak). At this stage, note that optimal sampled-data controls are still expressed as open-loop controls.

(2) The second issue concerns the expression of optimal sampled-data controls of LQOCPs as *closed-loop controls* (*i.e.*, as *feedbacks*). Theorem 2 (Section 3) provides such an expression. Two different proofs of Theorem 2 are given, based respectively on the PMP of Bourdin and Trélat (2016) (see Section 4.3) and on an adequate version of the dynamic programming principle (see Section 4.4). These two different approaches complete the Riccati theory for LQOCPs with sampled-data controls.
Note that the second proof of Theorem 2 (in Section 4.4) actually provides an extension of the strategy already proposed in Bini and Buttazzo (2009) to the nonautonomous and nonhomogeneous case. Moreover, as in Bini and Buttazzo (2009), it can be derived from Theorem 2 a recursive way in order to compute explicitly the optimal sampled-data controls of LQOCPs (see Corollary 1).

In turn, combining Theorem 1 and Corollary 1, we provide in this paper a strategy in order to compute explicitly pointwise convergent approximations of optimal permanent controls of LQOCPs. We provide in Section 3.3 some illustrating numerical simulations.

To the best of our knowledge, the strong convergence of the whole minimization problem (see Theorem 1 and Remark 3) has not been already addressed in the literature, even if several (strong or weak) convergence results have already been discussed (see, e.g., Chen & Francis, 1991 where the $\mathcal{H}_2$-performance recovery as the sampling frequency increases is proved). The closed-loop expression for optimal sampled-data controls of LQOCPs (see Theorem 2) has already been stated in Bini and Buttazzo (2009), with a dynamic programming principle approach and only in the autonomous and homogeneous case. In this paper we remove this last restriction and we provide a new and different proof. Actually, the originality of our work lies essentially in our novel approach based on the recently established PMP for optimal sampled-data control problems.

2. Pointwise convergence of optimal sampled-data controls

We first introduce some notations available throughout the paper. Let m and n be two nonzero integers. We denote by $\langle\cdot,\cdot\rangle_m$ (resp., $\langle\cdot,\cdot\rangle_n$) and $\|\cdot\|_m$ (resp., $\|\cdot\|_n$) the usual scalar product and the usual Euclidean norm of $\mathbb{R}^m$ (resp., $\mathbb{R}^n$). Let $a<b$ be two real numbers. We denote by $\mathcal{C}:=\mathcal{C}([a,b],\mathbb{R}^n)$ (resp., $\mathcal{AC}:=\mathcal{AC}([a,b],\mathbb{R}^n)$) the classical space of continuous functions (resp., absolutely continuous functions). We endow $\mathcal{C}$ with its usual uniform norm $\|\cdot\|_\infty$. The convergence of a sequence $(x_k)_{k\in\mathbb{N}}$ to some x in $\mathcal{C}$ (for the corresponding usual strong topology of $\mathcal{C}$) will be denoted by $x_k\to x$. We denote by $\mathrm{L}^2:=\mathrm{L}^2([a,b],\mathbb{R}^m)$ the classical Lebesgue space of square-integrable functions, endowed with its usual norm $\|\cdot\|_{\mathrm{L}^2}$. The strong convergence (resp., weak convergence) of a sequence $(u_k)_{k\in\mathbb{N}}$ to some u in L^2 will be denoted by $u_k\to u$ (resp., $u_k\rightharpoonup u$). We denote by $|||\cdot|||$ the induced norm for matrices in $\mathbb{R}^{n,n}$, $\mathbb{R}^{n,m}$, $\mathbb{R}^{m,n}$ and $\mathbb{R}^{m,m}$. Finally, if $M=M(\cdot)$ is a continuous matrix defined on $[a,b]$, we denote by $|||M|||_\infty$ its uniform norm.

2.1. Preliminaries on LQOCPs

Let E be a non-empty subset of L^2. We consider the following general LQOCP denoted by $(\mathscr{P}_\mathrm{E})$:

$$\begin{aligned}
&\text{min.} && \bar{\mathrm{J}}(x,u),\\
&\text{s.t.} && x\in\mathcal{AC},\quad u\in\mathrm{E}\subset\mathrm{L}^2,\\
& && \dot{x}(t)=A(t)x(t)+B(t)u(t)+w(t),\quad \text{a.e. } t\in[a,b],\\
& && x(a)=x_a,
\end{aligned}\qquad (\mathscr{P}_\mathrm{E})$$

where

$$\begin{aligned}
\bar{\mathrm{J}}(x,u) &:= \frac{1}{2}\Big\langle S\big(x(b)-x_b\big),x(b)-x_b\Big\rangle_n\\
&\quad+\frac{1}{2}\int_a^b \Big\langle Q(\tau)\big(x(\tau)-q(\tau)\big),x(\tau)-q(\tau)\Big\rangle_n\\
&\quad+\Big\langle R(\tau)\big(u(\tau)-r(\tau)\big),u(\tau)-r(\tau)\Big\rangle_m\,d\tau,
\end{aligned}$$

where $x_a,x_b\in\mathbb{R}^n$, $S\in\mathbb{R}^{n,n}$, $q,w:[a,b]\to\mathbb{R}^n$ and $r:[a,b]\to\mathbb{R}^m$ are continuous functions, and $A,Q:[a,b]\to\mathbb{R}^{n,n}$, $B:[a,b]\to\mathbb{R}^{n,m}$ and $R:[a,b]\to\mathbb{R}^{m,m}$ are continuous matrices.

Since $\langle Sy,y\rangle_n=\langle\frac{1}{2}(S+S^\top)y,y\rangle_n$ for any $y\in\mathbb{R}^n$, we assume, without loss of generality, that S is symmetric. For the same reason, we assume that $Q(t)$ and $R(t)$ are symmetric for every $t\in[a,b]$. We assume moreover that S is positive-semidefinite and that $Q(t)$ and $R(t)$ are respectively positive-semidefinite and positive-definite for every $t\in[a,b]$.

Definition 1. In the case where the matrices A, B, Q, R and the functions w, q, r are constant, Problem $(\mathscr{P}_\mathrm{E})$ is said to be *autonomous*.

Definition 2. In the case where $x_b=q(t)=w(t)=0_{\mathbb{R}^n}$ and $r(t)=0_{\mathbb{R}^m}$ for every $t\in[a,b]$, Problem $(\mathscr{P}_\mathrm{E})$ is said to be *homogeneous*.

The set E is used to model constraints on the set of *sampling times* at which the value of the control can be modified. More precisely, we consider:

- either *permanent controls*, and then $\mathrm{E}=\mathrm{L}^2$ (no constraint). In this case, the value of the control u can be modified at any time $t\in[a,b)$ and Problem $(\mathscr{P}_\mathrm{E})$ is said to be a general LQOCP with permanent controls.
- either *sampled-data controls*, and then E is a set of piecewise constant controls with a fixed and finite number of switching times (see Section 2.2). In this case, we speak of *nonpermanent controls* because the value of the control u cannot be modified at any time, but only at fixed sampling times. More precisely we speak of *sampled-data controls* because the value of the control u is *frozen* along each time interval between two consecutive sampling times (*sample-and-hold procedure*). Problem $(\mathscr{P}_\mathrm{E})$ is said to be a general LQOCP with sampled-data controls.

The first objective of this paper is to prove that the optimal sampled-data controls converge pointwise to the optimal permanent control, when the distances between consecutive sampling times tend to zero (see Theorem 1 in Section 2.2). Before coming to that point, we first recall hereafter a series of well known results for Problem $(\mathscr{P}_\mathrm{E})$ (see, e.g., Bryson & Ho, 1975; Lee & Markus, 1967; Trélat, 2005). Firstly, the Cauchy–Lipschitz theorem leads to the two following results.

Lemma 1. *For every $u\in\mathrm{L}^2$, there exists a unique solution $x\in\mathcal{AC}$, denoted by $x(\cdot,u)$, of the Cauchy problem $\dot{x}(t)=A(t)x(t)+B(t)u(t)+w(t)$ for a.e. $t\in[a,b]$, $x(a)=x_a$.*

Lemma 2. *For every $u \in \mathrm{L}^2$, there exists a unique solution $p \in \mathcal{AC}$, denoted by $p(\cdot, u)$, of the Cauchy problem $\dot{p}(t) = -A(t)^\top p(t) + Q(t)(x(t,u) - q(t))$ for a.e. $t \in [a,b]$, $p(b) = -S(x(b,u) - x_b)$.*

It clearly follows from Lemma 1 that Problem $(\mathscr{P}_{\mathrm{E}})$ can be reduced to the minimization problem $\min_{u \in \mathrm{E}} \mathrm{J}(u)$, where the cost functional $\mathrm{J} : \mathrm{L}^2 \to \mathbb{R}^+$ is defined by $\mathrm{J}(u) := \bar{\mathrm{J}}(x(\cdot, u), u)$. For the reader's convenience, the proof of the following claim is recalled in Section 4.1.

Proposition 1. *If E is a non-empty weakly closed convex subset of L^2, then Problem $(\mathscr{P}_{\mathrm{E}})$ has a unique solution denoted by u^*_{E}.*

In the permanent control case $\mathrm{E} = \mathrm{L}^2$, we denote the optimal solution by $u^* := u^*_{\mathrm{L}^2}$. The classical PMP (which is concerned with optimal permanent control problems) allows to express u^* as a function of the costate $p(\cdot, u^*)$, thus, as an open-loop control.

Proposition 2. *The optimal permanent control u^* satisfies the implicit equality*

$$u^*(t) = r(t) + R(t)^{-1} B(t)^\top p(t, u^*), \tag{1}$$

for almost every $t \in [a,b)$. Note that u^* is (equal almost everywhere to) a continuous function on $[a,b]$.

The proof of our first main result (stated in the next section) is based on the implicit equality (1).

2.2. Convergence result

We define Δ as the set of all $h = (h_i)_i \in (0, +\infty)^N$ for some $N \in \mathbb{N}^*$ with $\sum_{i=0}^{N-1} h_i = b - a$. For every $h \in \Delta$, we denote by $\|h\|_\Delta := \max_i h_i > 0$, and by $s^h_i := a + \sum_{j=0}^{i-1} h_j$ for all $i = 0, \ldots, N$. In particular we have $a = s^h_0 < s^h_1 < \cdots < s^h_N = b$.

Let $h \in \Delta$. We denote by $\mathrm{E}_h \subset \mathrm{L}^2$ the set of sampled-data controls associated to the sampling periods $h_0, \ldots, h_{N-1}$. Precisely, E_h is the set of controls $u : [a,b) \to \mathbb{R}^m$ such that $u = \sum_{i=0}^{N-1} U_i \mathbf{1}_{[s^h_i, s^h_{i+1})}$ for some $U_i \in \mathbb{R}^m$ and where $\mathbf{1}$ is the indicator function.

Remark 1. In Problem $(\mathscr{P}_{\mathrm{E}_h})$, note that $(s^h_i)_{i=0,\ldots,N-1}$ play the role of fixed sampling times at which (and only at which) the value of the control u can be modified.

It is clear that E_h is a non-empty weakly closed convex subset of L^2. From Proposition 1, Problem $(\mathscr{P}_{\mathrm{E}_h})$ admits a unique solution denoted by $u^*_h := u^*_{\mathrm{E}_h}$, that is, u^*_h is the optimal sampled-data control associated to $h \in \Delta$. The PMP recently stated in Bourdin and Trélat (2016) (which can be applied to optimal sampled-data control problems) allows to express u^*_h as a function of the costate $p(\cdot, u^*_h)$, thus, as an open-loop control.

Proposition 3. *Let $h \in \Delta$. The optimal sampled-data control $u^*_h \in \mathrm{E}_h$, written as $u^*_h = \sum_{i=0}^{N-1} U^*_{h,i} \mathbf{1}_{[s^h_i, s^h_{i+1})}$, satisfies the implicit equality*

$$U^*_{h,i} = \left(\frac{1}{h_i} \int_{s^h_i}^{s^h_{i+1}} R(s)\, ds \right)^{-1} \times \left(\frac{1}{h_i} \int_{s^h_i}^{s^h_{i+1}} R(s) r(s) + B(s)^\top p(s, u^*_h)\, ds \right), \tag{2}$$

for every $i = 0, \ldots, N-1$.

Remark 2. We provide in the next section an expression of u^*_h as a closed-loop control (see Theorem 2), and also a recursive way allowing to compute explicitly the optimal coefficients $U^*_{h,i}$ (see Corollary 1).

We are now in a position to state and prove the first main result of this paper.

Theorem 1. *The sequence $(u^*_h)_{h \in \Delta}$ of optimal sampled-data controls converges pointwise on $[a,b)$ to the optimal permanent control u^* as $\|h\|_\Delta$ tends to 0.*

Proof. Theorem 1 follows from the apparent relationship between the implicit equalities (1) and (2), and from the continuity of R, B, r and p. Actually we only need to prove that $p(\cdot, u^*_h) \to p(\cdot, u^*)$ in $\mathcal{C}$. To this end, we introduce the sampled-data control $u_h \in \mathrm{E}_h$ defined by $u_h := \sum_{i=0}^{N-1} U_{h,i} \mathbf{1}_{[s^h_i, s^h_{i+1})}$, where $U_{h,i} := \frac{1}{h_i} \int_{s^h_i}^{s^h_{i+1}} u^*(s) ds$ for every $i = 0, \ldots, N-1$. Since u^* is continuous on $[a,b]$, it is clear that $u_h(t)$ converges to $u^*(t)$ for every $t \in [a,b]$, and from the classical Lebesgue dominated convergence theorem that $u_h \to u^*$ in L^2, when $\|h\|_\Delta$ tends to 0. From Lemma 4 we conclude that $\mathrm{J}(u_h)$ tends to $\mathrm{J}(u^*)$. By optimality of u^* and u^*_h, we have $\mathrm{J}(u^*) \le \mathrm{J}(u^*_h) \le \mathrm{J}(u_h)$ for all $h \in \Delta$, and we get that $\mathrm{J}(u^*_h)$ tends to $\mathrm{J}(u^*)$ when $\|h\|_\Delta$ tends to 0. Since $\mathrm{J}(u^*_h)$ tends to $\mathrm{J}(u^*)$, we conclude that $(u^*_h)_{h \in \Delta}$ is a minimizing sequence of J on L^2 and, using the same arguments as in the proof of Proposition 1 (see Section 4.1), we deduce that (up to a subsequence that we do not relabel) $u^*_h \rightharpoonup u^*$ in L^2. Actually one can easily prove by contradiction that the whole sequence $(u^*_h)_{h \in \Delta}$ weakly converges to u^* in L^2. Finally Lemma 4 provides that $p(\cdot, u^*_h) \to p(\cdot, u^*)$ in $\mathcal{C}$ and concludes the proof. □

Remark 3. It follows from the above proof and from Lemma 4 that the three following strong convergences hold true: $\mathrm{J}(u^*_h) \to \mathrm{J}(u^*)$, $x(\cdot, u^*_h) \to x(\cdot, u^*)$ in $\mathcal{C}$ and $p(\cdot, u^*_h) \to p(\cdot, u^*)$ in $\mathcal{C}$ as $\|h\|_\Delta \to 0$.

3. Riccati theory for optimal sampled-data controls

In this section, we fix $h \in \Delta$ and our objective is to provide a closed-loop expression for the optimal sampled-data control u^*_h (see Theorem 2). This corresponds to an extension of the classical Riccati theory to the sampled-data control case. Moreover, we will show that this extension allows to compute explicitly (and in a recursive way) the optimal coefficients $U^*_{h,i}$ (see Corollary 1).

To be in accordance with the classical literature on Riccati theory, and for the sake of completeness, we will provide two different proofs of Theorem 2. The first proof (see Section 4.3) is based on Proposition 3, *i.e.*, on the PMP recently stated in Bourdin and Trélat (2016). The second proof (see Section 4.4) is based on the dynamic programming principle, and extends the strategy already used in Bini and Buttazzo (2009).

For the ease of notations, since $h \in \Delta$ is fixed throughout Section 3, we set $s_i := s^h_i$ for all $i = 0, \ldots, N$.

3.1. Some notations

We denote by $\Phi(\cdot,\cdot) : [a,b]^2 \to \mathbb{R}^{n,n}$ the *state-transition matrix* associated to A and we define

$$\Gamma_i(\tau) := \int_{s_i}^{\tau} \Phi(\tau, s) B(s)\, ds \in \mathbb{R}^{n,m},$$

$$\gamma_i(\tau) := \int_{s_i}^{\tau} \Phi(\tau, s) w(s)\, ds - q(\tau) \in \mathbb{R}^n,$$

for every $\tau \in [s_i, s_{i+1}]$ and every $i = 0, \dots, N-1$. Let us introduce the following terms that will play an important role in the sequel:

$$\varphi_i := \Phi(s_{i+1}, s_i) \in \mathbb{R}^{n,n},$$
$$\mathrm{C}_i := \Gamma_i(s_{i+1}) \in \mathbb{R}^{n,m},$$
$$\mathrm{D}_i := \int_{s_i}^{s_{i+1}} \Phi(s_{i+1}, s) w(s)\, ds - x_b \mathbf{1}_{\{N-1\}}(i) \in \mathbb{R}^n,$$
$$\mathrm{F}_i := \int_{s_i}^{s_{i+1}} \Phi(\tau, s_i)^\top \mathrm{Q}(\tau) \Phi(\tau, s_i)\, d\tau \in \mathbb{R}^{n,n},$$
$$\mathrm{G}_i := \int_{s_i}^{s_{i+1}} \Gamma_i(\tau)^\top \mathrm{Q}(\tau) \Phi(\tau, s_i)\, d\tau \in \mathbb{R}^{m,n},$$
$$\mathrm{H}_i := \int_{s_i}^{s_{i+1}} \Gamma_i(\tau)^\top \mathrm{Q}(\tau) \Gamma_i(\tau)\, d\tau \in \mathbb{R}^{m,m},$$
$$\mathrm{K}_i := \int_{s_i}^{s_{i+1}} \Gamma_i(\tau)^\top \mathrm{Q}(\tau) \gamma_i(\tau)\, d\tau \in \mathbb{R}^{m},$$
$$\mathrm{M}_i := \int_{s_i}^{s_{i+1}} \Phi(\tau, s_i)^\top \mathrm{Q}(\tau) \gamma_i(\tau)\, d\tau \in \mathbb{R}^{n},$$
$$\mathrm{O}_i := \int_{s_i}^{s_{i+1}} \langle \mathrm{Q}(\tau)\gamma_i(\tau), \gamma_i(\tau)\rangle_n \, d\tau \in \mathbb{R},$$
$$\mathrm{P}_i := \int_{s_i}^{s_{i+1}} \mathrm{R}(s)\, ds \in \mathbb{R}^{m,m},$$
$$\mathrm{T}_i := \int_{s_i}^{s_{i+1}} \mathrm{R}(s) r(s)\, ds \in \mathbb{R}^{m},$$
$$\mathrm{W}_i := \int_{s_i}^{s_{i+1}} \big\langle \mathrm{R}(\tau) r(\tau), r(\tau)\big\rangle_m \, d\tau \in \mathbb{R}.$$

Finally we introduce $(\mathrm{X}_i)_i$, $(\mathrm{Y}_i)_i$ and $(\mathrm{Z}_i)_i$ the following backward recursive sequences:

$$\begin{cases} \mathrm{X}_N = S, \\ \mathrm{X}_i = \boldsymbol{\mu}_i - \boldsymbol{\eta}_i^\top \boldsymbol{\lambda}_i^{-1} \boldsymbol{\eta}_i \in \mathbb{R}^{n,n}, \quad \forall i = N-1, \dots, 0, \end{cases}$$
$$\begin{cases} \mathrm{Y}_N = 0_{\mathbb{R}^n}, \\ \mathrm{Y}_i = \boldsymbol{\beta}_i - \boldsymbol{\eta}_i^\top \boldsymbol{\lambda}_i^{-1} \boldsymbol{\delta}_i \in \mathbb{R}^{n}, \quad \forall i = N-1, \dots, 0, \end{cases}$$
$$\begin{cases} \mathrm{Z}_N = 0, \\ \mathrm{Z}_i = \boldsymbol{\alpha}_i - \langle \boldsymbol{\lambda}_i^{-1} \boldsymbol{\delta}_i, \boldsymbol{\delta}_i \rangle_m \in \mathbb{R}, \quad \forall i = N-1, \dots, 0, \end{cases}$$

where $\boldsymbol{\alpha}_i, \boldsymbol{\beta}_i, \boldsymbol{\delta}_i, \boldsymbol{\eta}_i, \boldsymbol{\lambda}_i$ and $\boldsymbol{\mu}_i$ are defined (explicitly and dependently on X_{i+1}, Y_{i+1} and Z_{i+1}) as follows:

$$\boldsymbol{\alpha}_i := \langle \mathrm{X}_{i+1}\mathrm{D}_i + 2\mathrm{Y}_{i+1}, \mathrm{D}_i \rangle_n + \mathrm{O}_i + \mathrm{W}_i + \mathrm{Z}_{i+1} \in \mathbb{R},$$
$$\boldsymbol{\beta}_i := \varphi_i^\top \mathrm{X}_{i+1}\mathrm{D}_i + \mathrm{M}_i + \varphi_i^\top \mathrm{Y}_{i+1} \in \mathbb{R}^n,$$
$$\boldsymbol{\delta}_i := \mathrm{C}_i^\top \mathrm{X}_{i+1}\mathrm{D}_i + \mathrm{K}_i - \mathrm{T}_i + \mathrm{C}_i^\top \mathrm{Y}_{i+1} \in \mathbb{R}^m,$$
$$\boldsymbol{\eta}_i := \mathrm{C}_i^\top \mathrm{X}_{i+1}\varphi_i + \mathrm{G}_i \in \mathbb{R}^{m,n},$$
$$\boldsymbol{\lambda}_i := \mathrm{C}_i^\top \mathrm{X}_{i+1}\mathrm{C}_i + \mathrm{H}_i + \mathrm{P}_i \in \mathbb{R}^{m,m},$$
$$\boldsymbol{\mu}_i := \varphi_i^\top \mathrm{X}_{i+1}\varphi_i + \mathrm{F}_i \in \mathbb{R}^{n,n}.$$

Remark 4. A necessary condition for the above backward sequences to be well defined is the invertibility of $\boldsymbol{\lambda}_i$ for every $i = N-1, \dots, 0$. This necessary condition will be established in Section 4.3 (see also Section 4.4). More precisely, we will prove in a backward recursive way that X_{i+1} is positive-semidefinite for every $i = N-1, \dots, 0$. As a consequence, $\boldsymbol{\lambda}_i$ is equal to a sum of two positive-semidefinite matrices $\mathrm{C}_i^\top \mathrm{X}_{i+1}\mathrm{C}_i$ and H_i and of a positive-definite matrix P_i. We deduce that $\boldsymbol{\lambda}_i$ is positive-definite and hence it is invertible.

Remark 5. All terms introduced in this section depend only on the data of Problem $(\mathscr{P}_{\mathrm{E}_h})$, *i.e.*, on A, B, S, Q, R, w, x_b, q, r and h. It is worth to note that they do not depend on the initial condition x_a. As a consequence, all these terms, that are defined in a backward recursive way, remain unchanged if the initial condition in Problem $(\mathscr{P}_{\mathrm{E}_h})$ is modified, and they remain unchanged as well if the initial time a is replaced by s_j for some $j = 0, \dots, N-1$ and h is replaced by $(h_i)_{i=j,\dots,N-1}$.

Remark 6. In the homogeneous case, we have $\mathrm{D}_i = \mathrm{M}_i = \boldsymbol{\beta}_i = \mathrm{Y}_i = 0_{\mathbb{R}^n}$, $\mathrm{K}_i = \mathrm{T}_i = \boldsymbol{\delta}_i = 0_{\mathbb{R}^m}$, $\mathrm{O}_i = \mathrm{W}_i = \boldsymbol{\alpha}_i = \mathrm{Z}_i = 0$ for every $i = 0, \dots, N$, and then many simplifications occur in the formulas.

Remark 7. The sequences $(\mathrm{X}_i)_i$, $(\boldsymbol{\eta}_i)_i$, $(\boldsymbol{\lambda}_i)_i$ and $(\boldsymbol{\mu}_i)_i$ do not depend on the nonhomogeneous data x_b, w, q and r. As a consequence, they remain unchanged regardless of whether we consider the homogeneous or the nonhomogeneous Problem $(\mathscr{P}_{\mathrm{E}_h})$.

3.2. Closed-loop optimal sampled-data control

We are now in a position to state the second main result of this paper. Two different proofs of Theorem 2 are given respectively in Section 4.3 and in Section 4.4.

Theorem 2. *The optimal sampled-data control* $u_h^* = \sum_{i=0}^{N-1} U_{h,i}^* \mathbf{1}_{[s_i^h, s_{i+1}^h)}$ *is given in closed-loop form by*

$$U_{h,i}^* = -\boldsymbol{\lambda}_i^{-1}\big(\boldsymbol{\eta}_i x(s_i, u_h^*) + \boldsymbol{\delta}_i\big),$$

for every $i = 0, \dots, N-1$. *Moreover, the corresponding optimal cost is* $\mathrm{J}(u_h^*) = \frac{1}{2}\langle \mathrm{X}_0 x_a, x_a\rangle_n + \langle \mathrm{Y}_0, x_a\rangle_n + \frac{1}{2}\mathrm{Z}_0$.

Corollary 1. *The optimal values* $(U_{h,i}^*)_{i=0,\dots,N-1}$ *can be explicitly computed by the induction*

$$\forall i = 0, \dots, N-1, \quad \begin{cases} U_{h,i}^* = -\boldsymbol{\lambda}_i^{-1}\big(\boldsymbol{\eta}_i x_i + \boldsymbol{\delta}_i\big), \\ x_{i+1} = \varphi_i x_i + \mathrm{C}_i U_{h,i}^* + \mathrm{D}_i, \end{cases}$$

with the initial condition $x_0 = x_a$.

Proof. The Duhamel formula gives $x(s_{i+1}, u_h^*) = \varphi_i x(s_i, u_h^*) + \mathrm{C}_i U_{h,i}^* + \mathrm{D}_i$ for every $i = 0, \dots, N-1$. Corollary 1 then follows from Theorem 2. □

As a conclusion, in order to compute explicitly the optimal coefficients $(U_{h,i}^*)_{i=0,\dots,N-1}$, one has beforehand to compute all terms introduced in Section 3.1, and then to compute the induction provided in Corollary 1.

3.3. Some numerical simulations for a simple example

In this section, we focus on the following unidimensional LQOCP denoted by $(\mathscr{Q}_{\mathrm{E}})$:

$$\begin{aligned} &\min. \quad \int_0^1 x(\tau)^2 + \frac{1}{2}u(\tau)^2\, d\tau, \\ &\text{s.t.} \quad x \in \mathcal{AC}, \quad u \in \mathrm{E} \subset \mathrm{L}^2, \\ &\qquad \dot{x}(t) = \frac{1}{2}x(t) + u(t), \quad \text{a.e. } t \in [0,1], \\ &\qquad x(0) = 1. \end{aligned} \qquad (\mathscr{Q}_{\mathrm{E}})$$

It is clear that the data of Problem $(\mathscr{Q}_{\mathrm{E}})$ satisfy all assumptions of Section 2.1. This very simple problem has been considered in Dontchev, Hager, and Veliov (2000), Hager (1976) and Hager (2000), where the authors were interested in convergence issues for specific discretizations, showing that the simplest direct method diverges when considering an explicit second-order Runge–Kutta discretization. This is why this apparently inoffensive example is interesting and why we consider it here as well.

In the permanent control case $\mathrm{E} = \mathrm{L}^2$, the unique optimal permanent control u^* is given by $u^*(t) = 2(e^{3t} - e^3)e^{-3t/2}/(2 + e^3)$ for a.e. $t \in [0, 1]$. In this section, we are interested in the unique solution u_h^* of Problem $(\mathscr{Q}_{\mathrm{E}_h})$ for different values of $h \in \Delta$. Precisely, we take $h = \frac{1}{N}(1, \ldots, 1) \in (0, +\infty)^N$ with different values of $N \in \mathbb{N}^*$. Computing the induction provided in Corollary 1, we obtain the numerical results depicted in Fig. 1. When $\|h\|_\Delta$ tends to 0, we observe as expected (see Theorem 1) the pointwise convergence of u_h^* to u^*.

4. Proofs

4.1. Proof of Proposition 1

We first state two preliminary lemmas, variants of well known results in the existing literature (see, e.g., Bryson & Ho, 1975; Lee & Markus, 1967; Schättler & Ledzewicz, 2012; Trélat, 2005). A sketch of the proof of Lemma 4 is given for the sake of completeness.

Lemma 3. *Let $(g_k)_{k\in\mathbb{N}}$ be a sequence of equi-Hölderian functions defined on $[a, b]$ with values in $\mathbb{R}^n$. If the sequence $(g_k)_{k\in\mathbb{N}}$ converges pointwise on $[a, b]$ to 0, then it converges uniformly on $[a, b]$ to 0.*

Lemma 4. *The following properties hold true:*

(1) *If $u_k \rightharpoonup u$ in L^2, then $x(\cdot, u_k) \to x(\cdot, u)$ in $\mathscr{C}$.*
(2) *If $u_k \rightharpoonup u$ in L^2, then $p(\cdot, u_k) \to p(\cdot, u)$ in $\mathscr{C}$.*
(3) *The cost functional J is strictly convex on L^2.*
(4) *If $u_k \rightharpoonup u$ in L^2, then $\liminf_{k\to\infty} \mathrm{J}(u_k) \geq \mathrm{J}(u)$.*
(5) *If $u_k \to u$ in L^2, then $\lim_{k\to\infty} \mathrm{J}(u_k) = \mathrm{J}(u)$.*

Proof. 1– For every $k \in \mathbb{N}$ and every $t \in [a, b]$, we define $g_k(t) := \int_a^t B(\tau)(u_k(\tau) - u(\tau))d\tau$. Since $u_k \rightharpoonup u$ in L^2, the sequence $(g_k)_{k\in\mathbb{N}}$ converges pointwise on $[a, b]$ to 0. From the classical Hölder inequality, one can easily prove that the functions $(g_k)_{k\in\mathbb{N}}$ are equi-Hölderian. From Lemma 3, the sequence $(g_k)_{k\in\mathbb{N}}$ converges uniformly on $[a, b]$ to 0. Finally, the classical Gronwall lemma leads to $\|x(t, u_k) - x(t, u)\|_n \leq \|g_k\|_\infty e^{\|A\|_\infty (b-a)}$ for every $t \in [a, b]$ and every $k \in \mathbb{N}$. We deduce that $x(\cdot, u_k) \to x(\cdot, u)$ in $\mathscr{C}$.

2– One can similarly derive that $p(\cdot, u_k) \to p(\cdot, u)$ in $\mathscr{C}$.

3– Since S is positive-semidefinite, $Q(t)$ is positive-semidefinite and $R(t)$ is positive-definite for every $t \in [a, b]$, the result easily follows.

4– Since $x(\cdot, u_k) \to x(\cdot, u)$ in $\mathscr{C}$, we have

$$\lim_{k\to\infty} \frac{1}{2}\left\langle S\big(x(b, u_k) - x_b\big), x(b, u_k) - x_b\right\rangle_n + \frac{1}{2}\int_a^b \left\langle Q(\tau)\big(x(\tau, u_k) - q(\tau)\big), x(\tau, u_k) - q(\tau)\right\rangle_n d\tau = \frac{1}{2}\left\langle S\big(x(b, u) - x_b\big), x(b, u) - x_b\right\rangle_n + \frac{1}{2}\int_a^b \left\langle Q(\tau)\big(x(\tau, u) - q(\tau)\big), x(\tau, u) - q(\tau)\right\rangle_n d\tau.$$

By contradiction we prove that there exists a constant $c_R > 0$ such that $c_R\|y\|_m^2 \leq \langle R(t)y, y\rangle_m \leq \|R\|_\infty \|y\|_m^2$ for every $t \in [a, b]$ and every $y \in \mathbb{R}^m$. As a consequence, the scalar product $\langle \cdot, \cdot\rangle_R$ defined on L^2 by $\langle u_1, u_2\rangle_R := \int_a^b \langle R(\tau)u_1(\tau), u_2(\tau)\rangle_m d\tau$ induces a norm $\|\cdot\|_R$ on L^2 that is equivalent to the usual one. Since $u_k \rightharpoonup u$ in L^2, we have $\liminf_{k\to\infty} \frac{1}{2}\|u_k - r\|_R^2 \geq \frac{1}{2}\|u - r\|_R^2$. This concludes the proof.

5– The proof is similar since $u_k \to u$ in L^2 implies that $\lim_{k\to\infty} \frac{1}{2}\|u_k - r\|_R^2 = \frac{1}{2}\|u - r\|_R^2$. □

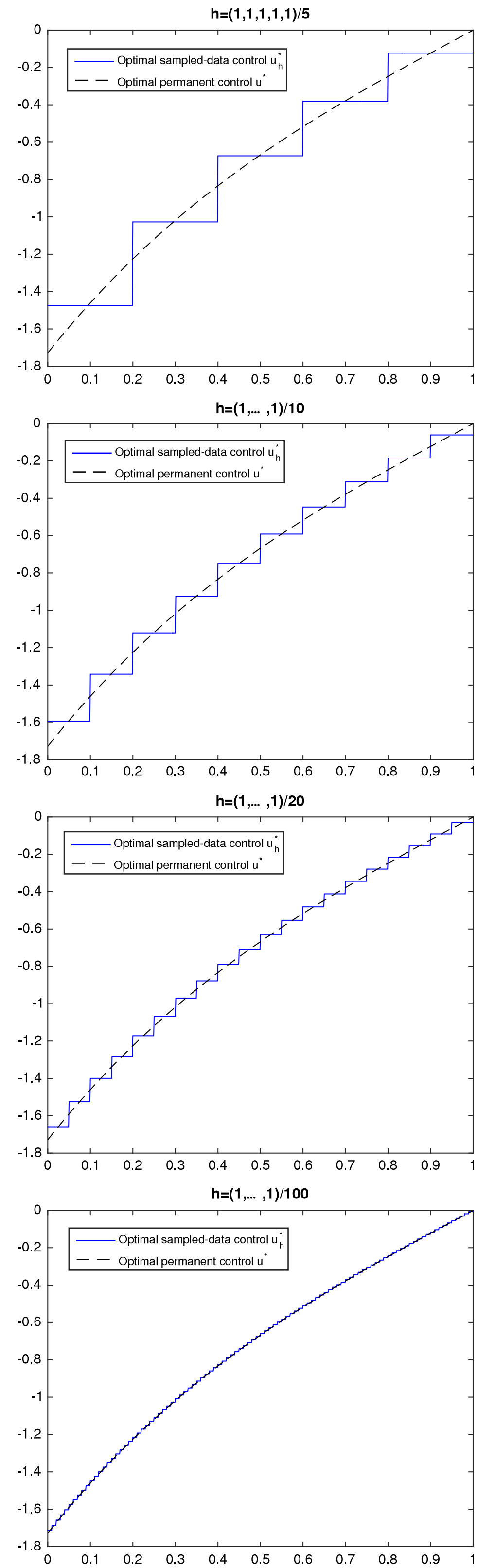

Fig. 1. Pointwise convergence of u_h^* to u^* as $\|h\|_\Delta$ tends to 0.

We are now in a position to prove Proposition 1. Let us prove that J has a unique minimizer on E. Uniqueness is clear since E is convex and J is strictly convex (see Lemma 4). Let us prove existence. Let $(u_k)_{k\in\mathbb{N}} \subset \mathrm{E}$ be a minimizing sequence of J on E. Since $\mathrm{J}(u_k) \geq \frac{1}{2}\|u_k - r\|_R^2$, we conclude that $(u_k)_{k\in\mathbb{N}}$ is bounded in L^2 and thus converges weakly, up to some subsequence, to some u_E^*. Since

E is weakly closed, we get that $u^*_{\mathrm{E}} \in \mathrm{E}$ and we get from Lemma 4 that $\inf_{u\in\mathrm{E}} \mathrm{J}(u) = \lim_{k\to\infty} \mathrm{J}(u_k) = \liminf_{k\to\infty} \mathrm{J}(u_k) \geq \mathrm{J}(u^*_{\mathrm{E}})$ which concludes the proof.

4.2. Preliminaries for Theorem 2 and value function

In this section, we establish preliminary results that are required to prove Theorem 2. Precisely they are required in order to prove the invertibility of the matrices $\boldsymbol{\lambda}_i$ in the first proof of Theorem 2 (based on Proposition 3 and detailed in Section 4.3), and to prove the dynamic programming principle which is the basis of the second proof of Theorem 2 (detailed in Section 4.4).

For all $0 \leq j < k \leq N$ and $U = (U_i)_{i=j,\dots,k-1} \in (\mathbb{R}^m)^{k-j}$, we define the function $u_U : [s_j, s_k) \to \mathbb{R}^m$ by $u_U := \sum_{i=j}^{k-1} U_i \mathbf{1}_{[s_i,s_{i+1})}$. For every $y \in \mathbb{R}^n$, we denote by $x^k_j(\cdot, y, U) : [s_j, s_k] \to \mathbb{R}^n$ the unique solution of the Cauchy problem $\dot{x}(t) = A(t)x(t) + B(t)u_U(t) + w(t)$ for a.e. $t \in [s_j, s_k]$ and $x(s_j) = y$.

Let $\overline{\mathcal{V}}_j(\cdot,\cdot) : \mathbb{R}^n \times (\mathbb{R}^m)^{N-j} \to \mathbb{R}^+$ be defined by

$$\overline{\mathcal{V}}_j(y,U) := \frac{1}{2}\left\langle S\left(x^N_j(b,y,U) - x_b\right), x^N_j(b,y,U) - x_b\right\rangle_n + \frac{1}{2}\int_{s_j}^{b} \left\langle Q(\tau)\left(x^N_j(\tau,y,U) - q(\tau)\right), x^N_j(\tau,y,U) - q(\tau)\right\rangle_n + \left\langle R(\tau)\left(u_U(\tau) - r(\tau)\right), u_U(\tau) - r(\tau)\right\rangle_m d\tau, \quad (3)$$

for every $(y,U) \in \mathbb{R}^n \times (\mathbb{R}^m)^{N-j}$ and every $j = 0,\dots,N-1$.

Remark 8. Note that $\overline{\mathcal{V}}_j(y,U)$ coincides with the cost $\mathrm{J}(u_U)$ whenever the initial time a is replaced by s_j and the initial condition x_a is replaced by y and h is replaced by $(h_i)_{i=j,\dots,N-1}$ in Problem $(\mathscr{P}_{\mathrm{E}_h})$.

In the sequel, we set $\Psi_{j,y}(U_j) := x^{j+1}_j(s_{j+1}, y, U_j)$ and $\tilde{U} := (U_i)_{i=j+1,\dots,N-1}$ for every $(y,U) \in \mathbb{R}^n \times (\mathbb{R}^m)^{N-j}$ and every $j = 0,\dots,N-1$. The next statement obviously follows from the definition of $\overline{\mathcal{V}}_j(\cdot,\cdot)$.

Lemma 5. *For every $j = 0,\dots,N-2$ and every $(y,U) \in \mathbb{R}^n \times (\mathbb{R}^m)^{N-j}$, we have*

$$\overline{\mathcal{V}}_j(y,U) = \overline{\mathcal{V}}_{j+1}\left(\Psi_{j,y}(U_j), \tilde{U}\right) + \frac{1}{2}\int_{s_j}^{s_{j+1}} \left\langle Q(\tau)\left(x^{j+1}_j(\tau,y,U_j) - q(\tau)\right), x^{j+1}_j(\tau,y,U_j) - q(\tau)\right\rangle_n + \left\langle R(\tau)\left(U_j - r(\tau)\right), U_j - r(\tau)\right\rangle_m d\tau.$$

Let $\mathcal{V}_j(\cdot) : \mathbb{R}^n \to \mathbb{R}^+$ be the *value function* defined by

$$\mathcal{V}_j(y) := \inf_{U\in(\mathbb{R}^m)^{N-j}} \overline{\mathcal{V}}_j(y,U), \quad (4)$$

for every $y \in \mathbb{R}^n$ and every $j = 0,\dots,N-1$. From Remark 8 and similarly to Proposition 1, one can easily prove that the infimum $\mathcal{V}_j(y)$ is reached at a unique point denoted by $U_j(y)^* = (U_j(y)^*_i)_{i=j,\dots,N-1}$. Thus we have

$$\mathcal{V}_j(y) = \overline{\mathcal{V}}_j\left(y, U_j(y)^*\right), \quad (5)$$

for every $y \in \mathbb{R}^n$ and every $j = 0,\dots,N-1$.

Remark 9. From Remark 8, we have $\mathrm{J}(u^*_h) = \mathcal{V}_0(x_a)$ and $(U^*_{h,i})_{i=0,\dots,N-1} = U_0(x_a)^*$. More generally we have $(U^*_{h,i})_{i=j,\dots,N-1} = U_j(x(s_j,u^*_h))^*$ for every $j = 0,\dots,N-1$.

The next statement follows from the definition of $U_j(y)^*$.

Lemma 6. *For every $j = 0,\dots,N-2$ and every $y \in \mathbb{R}^n$, we have*

$$U_{j+1}\left(\Psi_{j,y}\left(U_j(y)^*_j\right)\right)^* = \left(U_j(y)^*_i\right)_{i=j+1,\dots,N-1}.$$

Finally, from (5), Lemmas 5–6, we infer the next result.

Proposition 4. *For every $j = 0,\dots,N-2$ and every $y \in \mathbb{R}^n$, we have*

$$\mathcal{V}_j(y) = \mathcal{V}_{j+1}\left(\Psi_{j,y}\left(U_j(y)^*_j\right)\right) + \frac{1}{2}\int_{s_j}^{s_{j+1}} \left\langle Q(\tau)\left(x^{j+1}_j(\tau,y,U_j(y)^*_j) - q(\tau)\right), x^{j+1}_j(\tau,y,U_j(y)^*_j) - q(\tau)\right\rangle_n + \left\langle R(\tau)\left(U_j(y)^*_j - r(\tau)\right), U_j(y)^*_j - r(\tau)\right\rangle_m d\tau.$$

Proposition 4 will be used in order to prove the invertibility of the matrices $\boldsymbol{\lambda}_i$ in the first proof of Theorem 2.

Remark 10. Proposition 4 does not correspond to the dynamic programming principle, whose version adapted to the framework of this paper is stated in Proposition 5.

4.3. First proof of Theorem 2

From Proposition 3 it holds

$$\mathrm{P}_i U^*_{h,i} = \mathrm{T}_i + \int_{s_i}^{s_{i+1}} B(s)^\top p(s,u^*_h)\, ds, \quad (6)$$

for all $i = 0,\dots,N-1$. In order to prove Theorem 2 from (6), we use the Duhamel formula in order to derive an explicit expression of $p(s,u^*_h)$ as a function of $x(s_i,u^*_h)$ and $U^*_{h,i}$. We will prove by backward induction that the following five statements are true:

(1) $\boldsymbol{\lambda}_i$ is invertible;
(2) $U^*_{h,i} = -\boldsymbol{\lambda}_i^{-1}(\boldsymbol{\eta}_i x(s_i,u^*_h) + \boldsymbol{\delta}_i)$;
(3) $p(s_i,u^*_h) = -(\mathrm{X}_i x(s_i,u^*_h) + \mathrm{Y}_i)$;
(4) $\mathcal{V}_i(y) = \frac{1}{2}\langle \mathrm{X}_i y, y\rangle_n + \langle \mathrm{Y}_i, y\rangle_n + \frac{1}{2}\mathrm{Z}_i$ for all $y \in \mathbb{R}^n$;
(5) X_i is positive-semidefinite;

for every $i = N-1,\dots,0$.

To prove the induction steps, let us first recall the following equalities that follow from the Duhamel formula:

$$p(s,u^*_h) = \Phi(s_{i+1},s)^\top p(s_{i+1},u^*_h) - \int_{s}^{s_{i+1}} \Phi(\tau,s)^\top Q(\tau)\left(x(\tau,u^*_h) - q(\tau)\right) d\tau, \quad (7)$$

and

$$x(\tau,u^*_h) = \Phi(\tau,s_i)x(s_i,u^*_h) + \Gamma_i(\tau)U^*_{h,i} + \gamma_i(\tau) + q(\tau), \quad (8)$$

for every $s,\tau \in [s_i,s_{i+1}]$ and every $i = 0,\dots,N-1$.

4.3.1. Initialization of the backward induction

Let $i = N-1$.

1– Since $\mathrm{X}_{i+1} = S$ is positive-semidefinite, we infer that $\boldsymbol{\lambda}_i$ is invertible (see Remark 4).

2– Using (7) and (8) and $p(s_{i+1},u^*_h) = p(b,u^*_h) = -S(x(b,u^*_h) - x_b) = -S(x(s_{i+1},u^*_h) - x_b)$, we get that

$$p(s,u^*_h) = -\left[\Phi(s_{i+1},s)^\top S\varphi_i + \int_{s}^{s_{i+1}} \Phi(\tau,s)^\top Q(\tau)\Phi(\tau,s_i)\, d\tau\right] x(s_i,u^*_h)$$

$$- \Big[\Phi(s_{i+1},s)^\top S C_i + \int_s^{s_{i+1}} \Phi(\tau,s)^\top Q(\tau)\Gamma_i(\tau)\,d\tau \Big] U^*_{h,i}$$

$$- \Big[\Phi(s_{i+1},s)^\top S D_i + \int_s^{s_{i+1}} \Phi(\tau,s)^\top Q(\tau)\gamma_i(\tau)\,d\tau \Big]. \quad (9)$$

Replacing $p(s,u^*_h)$ in Equality (6) and applying the Fubini theorem, we obtain that

$$P_i U^*_{h,i} = T_i - \Big(C_i^\top S \varphi_i + G_i \Big) x(s_i,u^*_h) - \Big(C_i^\top S C_i + H_i \Big) U^*_{h,i} - \Big(C_i^\top S D_i + K_i \Big),$$

that is exactly $\boldsymbol{\lambda}_i U^*_{h,i} = -\boldsymbol{\eta}_i x(s_i,u^*_h) - \boldsymbol{\delta}_i$. The result follows from the invertibility of $\boldsymbol{\lambda}_i$.

3– Taking $s=s_i$ in (9) leads to

$$p(s_i,u^*_h) = -\Big(\varphi_i^\top S \varphi_i + F_i \Big) q(s_i,u^*_h) - \Big(\varphi_i^\top S C_i + G_i^\top \Big) U^*_{h,i} - \Big(\varphi_i^\top S D_i + M_i \Big),$$

that is exactly $p(s_i,u^*_h) = -\boldsymbol{\mu}_i x(s_i,u^*_h) - \boldsymbol{\eta}_i^\top U^*_{h,i} - \boldsymbol{\beta}_i$. The result follows from $U^*_{h,i} = -\boldsymbol{\lambda}_i^{-1}(\boldsymbol{\eta}_i x(s_i,u^*_h) + \boldsymbol{\delta}_i)$.

4– Let $y \in \mathbb{R}^n$. From Remarks 5 and 8 and from the definition of $U_i(y)^*$, similarly to Step 2, we get that $U_i(y)^* = -\boldsymbol{\lambda}_i^{-1}(\boldsymbol{\eta}_i y + \boldsymbol{\delta}_i)$ and the Duhamel formula gives

$$x_i^{i+1}(\tau,y,U_i(y)^*) = \Phi(\tau,s_i)y + \Gamma_i(\tau) + \gamma_i(\tau) + q(\tau),$$

for every $\tau \in [s_i,b]$. Using the above equality in (3), we exactly obtain that

$$\mathcal{V}_i(y) = \overline{\mathcal{V}}_i(y,U_i(y)^*) = \Big\langle \frac{1}{2}\boldsymbol{\lambda}_i U_i(y)^* + \boldsymbol{\eta}_i y + \boldsymbol{\delta}_i, U_i(y)^* \Big\rangle_m + \Big\langle \frac{1}{2}\boldsymbol{\mu}_i y + \boldsymbol{\beta}_i, y \Big\rangle_n + \frac{1}{2}\boldsymbol{\alpha}_i.$$

Since $U_i(y)^* = -\boldsymbol{\lambda}_i^{-1}(\boldsymbol{\eta}_i y + \boldsymbol{\delta}_i)$, we conclude.

5– Let $y \in \mathbb{R}^n$ and let us consider temporarily the homogeneous Problem $(\mathscr{P}_{E_h})$. From Remarks 6 and 7, similarly to Step 4, we get in the homogeneous case that $\mathcal{V}_i(y) = \frac{1}{2}\langle X_i y, y\rangle_n \geq 0$. It follows that X_i is positive-semidefinite.

4.3.2. The induction step

Let $i \in \{0,\dots,N-2\}$ and let us assume that the five statements are satisfied at steps $i+1,\dots,N-1$.

1– Since X_{i+1} is positive-semidefinite, we infer that $\boldsymbol{\lambda}_i$ is invertible (see Remark 4).

2– As in the previous paragraph, we use (7)–(8) and $p(s_{i+1},u^*_h) = -(X_{i+1}x(s_{i+1},u^*_h) + Y_{i+1})$ in order to obtain an expression of $p(s,u^*_h)$ that we insert in (6). Then, we apply the Fubini theorem and we obtain that

$$P_i U^*_{h,i} = T_i - \Big(C_i^\top X_{i+1}\varphi_i + G_i \Big) x(s_i,u^*_h) - \Big(C_i^\top X_{i+1} C_i + H_i \Big) U^*_{h,i} - \Big(C_i^\top X_{i+1} D_i + K_i + C_i^\top Y_{i+1} \Big),$$

that is exactly $\boldsymbol{\lambda}_i U^*_{h,i} = -\boldsymbol{\eta}_i x(s_i,u^*_h) - \boldsymbol{\delta}_i$. Finally, the result follows from the invertibility of $\boldsymbol{\lambda}_i$.

3– As in the previous paragraph, taking $s=s_i$ in the expression previously obtained gives

$$p(s_i,u^*_h) = -\Big(\varphi_i^\top X_{i+1}\varphi_i + F_i \Big) x(s_i,u^*_h) - \Big(\varphi_i^\top X_{i+1} C_i + G_i^\top \Big) U^*_{h,i} - \Big(\varphi_i^\top X_{i+1} D_i + M_i + \varphi_i^\top Y_{i+1} \Big),$$

that is exactly $p(s_i,u^*_h) = -\boldsymbol{\mu}_i x(s_i,u^*_h) - \boldsymbol{\eta}_i^\top U^*_{h,i} - \boldsymbol{\beta}_i$. The result follows from $U^*_{h,i} = -\boldsymbol{\lambda}_i^{-1}(\boldsymbol{\eta}_i x(s_i,u^*_h) + \boldsymbol{\delta}_i)$.

4– Let $y \in \mathbb{R}^n$. From Remarks 5 and 8 and from the definition of $U_i(y)^*$, one can prove in a very similar way than Step 2 that $U_i(y)^*_i = -\boldsymbol{\lambda}_i^{-1}(\boldsymbol{\eta}_i y + \boldsymbol{\delta}_i)$. From Proposition 4, we have

$$\mathcal{V}_i(y) = \mathcal{V}_{i+1}\Big(\Psi_{i,y}\big(U_i(y)^*_i\big)\Big) + \frac{1}{2}\int_{s_i}^{s_{i+1}} \Big\langle Q(\tau)\big(x_i^{i+1}(\tau,y,U_i(y)^*_i) - q(\tau)\big), x_i^{i+1}(\tau,y,U_i(y)^*_i) - q(\tau) \Big\rangle_n + \Big\langle R(\tau)\big(U_i(y)^*_i - r(\tau)\big), U_i(y)^*_i - r(\tau) \Big\rangle_m \,d\tau. \quad (10)$$

From the induction hypothesis, we have

$$\mathcal{V}_{i+1}\Big(\Psi_{i,y}\big(U_i(y)^*_i\big)\Big) = \frac{1}{2}\Big\langle X_{i+1}\Psi_{i,y}\big(U_i(y)^*_i\big), \Psi_{i,y}\big(U_i(y)^*_i\big) \Big\rangle_n + \Big\langle Y_{i+1}, \Psi_{i,y}\big(U_i(y)^*_i\big) \Big\rangle_n + \frac{1}{2}Z_{i+1}. \quad (11)$$

On the other hand, the Duhamel formula gives

$$x_i^{i+1}(\tau,y,U_i(y)^*_i) = \Phi(\tau,s_i)y + \Gamma_i(\tau)U_i(y)^*_i + \gamma_i(\tau) + q(\tau), \quad (12)$$

for every $\tau \in [s_i,s_{i+1}]$. Taking $\tau = s_{i+1}$ in (12), we get that

$$\Psi_{i,y}\big(U_i(y)^*_i\big) = \varphi_i y + C_i U_i(y)^*_i + D_i. \quad (13)$$

Finally, using (11)–(13) in (10) yields

$$\mathcal{V}_i(y) = \Big\langle \frac{1}{2}\boldsymbol{\lambda}_i U_i(y)^* + \boldsymbol{\eta}_i y + \boldsymbol{\delta}_i, U_i(y)^* \Big\rangle_m + \Big\langle \frac{1}{2}\boldsymbol{\mu}_i y + \boldsymbol{\beta}_i, y \Big\rangle_n + \frac{1}{2}\boldsymbol{\alpha}_i.$$

Since $U_i(y)^*_i = -\boldsymbol{\lambda}_i^{-1}(\boldsymbol{\eta}_i y + \boldsymbol{\delta}_i)$, we conclude.

5– Proof is similar and thus is skipped.

4.4. Second proof of Theorem 2

This section is dedicated to an alternative proof, based on the following dynamic programming principle.

Proposition 5. *For every $j = 0,\dots,N-2$ and every $y \in \mathbb{R}^n$, we have*

$$\mathcal{V}_j(y) = \inf_{U_j \in \mathbb{R}^m} \Big\{ \mathcal{V}_{j+1}\big(\Psi_{j,y}(U_j)\big) + \frac{1}{2}\int_{s_j}^{s_{j+1}} \Big\langle Q(\tau)\big(x_j^{j+1}(\tau,y,U_j) - x(\tau)\big), x_j^{j+1}(\tau,y,U_j) - x(\tau) \Big\rangle_n + \Big\langle R(\tau)\big(U_j - r(\tau)\big), U_j - r(\tau) \Big\rangle_m \,d\tau \Big\}.$$

Proof. From (4) and Lemma 5, $\mathcal{V}_j(y)$ is equal to the infimum of

$$\overline{\mathcal{V}}_{j+1}\big(\Psi_{j,y}(U_j),\tilde{U}\big) + \frac{1}{2}\int_{s_j}^{s_{j+1}} \Big\langle Q(\tau)\big(x_j^{j+1}(\tau,y,U_j) - q(\tau)\big), x_j^{j+1}(\tau,y,U_j) - q(\tau) \Big\rangle_n + \Big\langle R(\tau)\big(U_j - r(\tau)\big), U_j - r(\tau) \Big\rangle_m \,d\tau,$$

over all possible $U = (U_j,\tilde{U}) \in \mathbb{R}^m \times (\mathbb{R}^m)^{N-(j+1)}$. Separating the variables U_j and $\tilde{U}$ in the infimum leads to the result. □

In order to prove Theorem 2, we will prove by backward induction that the following four statements are true:

(1) $\boldsymbol{\lambda}_i$ is invertible;
(2) $U^*_{h,i} = -\boldsymbol{\lambda}_i^{-1}(\boldsymbol{\eta}_i x(s_i, u^*_h) + \boldsymbol{\delta}_i)$;
(3) $\mathcal{V}(i, y) = \frac{1}{2}\langle \mathrm{X}_i y, y\rangle_n + \langle \mathrm{Y}_i, y\rangle_n + \frac{1}{2}\mathrm{Z}_i$ for all $y \in \mathbb{R}^n$;
(4) X_i is positive-semidefinite;

for every $i = N-1, \dots, 0$.

Recall that, by the Duhamel formula,

$$x_i^{i+1}(\tau, y, U_i) = \Phi(\tau, s_i)y + \Gamma_i(\tau)U_i + \gamma_i(\tau) + q(\tau), \tag{14}$$

for every $i = 0, \dots, N-1$, every $(y, U_i) \in \mathbb{R}^n \times \mathbb{R}^m$ and every $\tau \in [s_i, s_{i+1}]$. Taking $i \in \{0, \dots, N-2\}$ and $\tau = s_{i+1}$ in (14), we get that

$$\Psi_{i,y}(U_i) = \varphi_i y + \mathrm{C}_i U_i + \mathrm{D}_i. \tag{15}$$

4.4.1. Initialization of the backward induction

Let $i = N-1$.

1– Since $\mathrm{X}_{i+1} = S$ is positive-semidefinite, we infer that $\boldsymbol{\lambda}_i$ is invertible (see Remark 4).

2– Let $y \in \mathbb{R}^n$. Using (14) in (3), we get that

$$\begin{aligned}\overline{\mathcal{V}}_i(y, U_i) = {} & \frac{1}{2}\left\langle\left(\mathrm{C}_i^\top S\mathrm{C}_i + \mathrm{H}_i + \mathrm{P}_i\right)U_i, U_i\right\rangle_m \\ & + \left\langle\left(\mathrm{C}_i^\top S\varphi_i + \mathrm{G}_i\right)y + \left(\mathrm{C}_i^\top S\mathrm{D}_i + \mathrm{K}_i - \mathrm{T}_i\right), U_i\right\rangle_m \\ & + \left\langle\frac{1}{2}\left(\varphi_i^\top S\varphi_i + \mathrm{F}_i\right)y + \left(\varphi_i^\top S\mathrm{D}_i + \mathrm{M}_i\right), y\right\rangle_n \\ & + \frac{1}{2}\left(\langle S\mathrm{D}_i, \mathrm{D}_i\rangle_n + \mathrm{O}_i + \mathrm{W}_i\right),\end{aligned}$$

for every $U_i \in \mathbb{R}^m$. Hence we have exactly obtained that

$$\begin{aligned}\mathcal{V}_i(y) &= \inf_{U_i \in \mathbb{R}^m} \overline{\mathcal{V}}_i(y, U_i) \\ &= \inf_{U_i \in \mathbb{R}^m}\left(\left\langle\frac{1}{2}\boldsymbol{\lambda}_i U_i + \boldsymbol{\eta}_i y + \boldsymbol{\delta}_i, U_i\right\rangle_m \right. \\ &\quad \left. + \left\langle\frac{1}{2}\boldsymbol{\mu}_i y + \boldsymbol{\beta}_i, y\right\rangle_n + \frac{1}{2}\boldsymbol{\alpha}_i\right).\end{aligned} \tag{16}$$

Differentiating the above expression with respect to U_i, the infimum $\mathcal{V}_i(y)$ is reached at $U_i(y)^* \in \mathbb{R}^m$ that satisfies $\boldsymbol{\lambda}_i U_i(y)^* + \boldsymbol{\eta}_i y + \boldsymbol{\delta}_i = 0_{\mathbb{R}^m}$. Since $\boldsymbol{\lambda}_i$ is invertible, we deduce that $U_i(y)^* = -\boldsymbol{\lambda}_i^{-1}(\boldsymbol{\eta}_i y + \boldsymbol{\delta}_i)$. Finally, taking $y = x(s_i, u^*_h)$, we obtain from Remark 9 that

$$U^*_{h,i} = U_i\big(x(s_i, u^*_h)\big)^* = -\boldsymbol{\lambda}_i^{-1}\big(\boldsymbol{\eta}_i x(s_i, u^*_h) + \boldsymbol{\delta}_i\big).$$

3– Let $y \in \mathbb{R}^n$. We conclude by replacing U_i by $U_i(y)^* = -\boldsymbol{\lambda}_i^{-1}(\boldsymbol{\eta}_i y + \boldsymbol{\delta}_i)$ in the expression of (16).

4– Let $y \in \mathbb{R}^n$ and let us consider temporarily the homogeneous Problem $(\mathscr{P}_{\mathrm{E}_h})$. From Remarks 6 and 7, similarly to Step 3, we get in the homogeneous case that $\mathcal{V}_i(y) = \frac{1}{2}\langle \mathrm{X}_i y, y\rangle_n \geq 0$. It follows that X_i is positive-semidefinite.

4.4.2. Induction step

Let $i \in \{0, \dots, N-2\}$ and let us assume that the four statements are satisfied at steps $i+1, \dots, N-1$.

1– Since X_{i+1} is positive-semidefinite, we infer that $\boldsymbol{\lambda}_i$ is invertible (see Remark 4).

2– Let $y \in \mathbb{R}^n$. From the dynamic programming principle stated in Proposition 5, we obtain that $\mathcal{V}_i(y)$ is equal to the infimum of

$$\begin{aligned}&\mathcal{V}_{i+1}\big(\Psi_{i,y}(U_i)\big) + \frac{1}{2}\int_{s_i}^{s_{i+1}} \Big\langle Q(\tau)\big(x_i^{i+1}(\tau, y, U_i) - q(\tau)\big), \\ &x_i^{i+1}(\tau, y, U_i) - q(\tau)\Big\rangle_n + \Big\langle R(\tau)\big(U_i - r(\tau)\big), U_i - r(\tau)\Big\rangle_m \, d\tau\end{aligned} \tag{17}$$

over $U_i \in \mathbb{R}^m$. From the induction hypothesis, we have

$$\begin{aligned}\mathcal{V}_{i+1}\big(\Psi_{i,y}(U_i)\big) = {} & \frac{1}{2}\big\langle \mathrm{X}_{i+1}\Psi_{i,y}(U_i), \Psi_{i,y}(U_i)\big\rangle_n \\ & + \big\langle \mathrm{Y}_{i+1}, \Psi_{i,y}(U_i)\big\rangle_n + \frac{1}{2}\mathrm{Z}_{i+1}.\end{aligned} \tag{18}$$

Using (14), (15) and (18) in (17), we obtain that

$$\begin{aligned}\mathcal{V}_i(y) = {} & \inf_{U_i \in \mathbb{R}^m}\left(\left\langle\frac{1}{2}\boldsymbol{\lambda}_i U_i + \boldsymbol{\eta}_i y + \boldsymbol{\delta}_i, U_i\right\rangle_m \right. \\ & \left. + \left\langle\frac{1}{2}\boldsymbol{\mu}_i y + \boldsymbol{\beta}_i, y\right\rangle_n + \frac{1}{2}\boldsymbol{\alpha}_i\right).\end{aligned} \tag{19}$$

The end of the proof is similar and thus is skipped.

3– 4– Proofs are similar and thus are skipped.

5. Conclusion and open problems

We have proved that the optimal sampled-data controls of a LQOCP converge pointwise to the optimal permanent control as the sampling periods tend to zero. An open problem is to address this issue for nonlinear optimal control problems. Also, taking into account some final state constraints is difficult, and establishing a convergence result like Theorem 1 seems challenging and may require to consider singular trajectories or abnormal extremals, and conjugate point theory (see Agrachev & Sachkov, 2004; Bonnard & Chyba, 2003).

We have extended the Riccati theory to general LQOCPs with sampled-data controls, by two approaches. The first consists of applying a version of the PMP valid for optimal sampled-data control problems. The second relies on an appropriate version of the dynamic programming principle. When dealing with nonlinear dynamics, the classical Riccati equation is to be replaced with the Hamilton–Jacobi equation, whose viscosity solutions are nonsmooth in general, and investigation of such issues with sampled-data controls seems to be open.

References

Ackermann, J. E. (1983). *Sampled-data control. Volume 1*. Berlin-New-York: Springer-Verlag.

Ackermann, J. E. (1985). *Sampled-data control systems: analysis and synthesis, robust system design*. Springer-Verlag.

Agrachev, A. A., & Sachkov, Y. L. (2004). *Encyclopaedia of mathematical sciences: Vol. 87. Control theory from the geometric viewpoint*. Berlin: Springer-Verlag.

Aida-Zadeh, K. R., & Rahimov, A. B. (2007). Solution of optimal control problem in class of piecewise-constant functions. *Automatic Control and Computer Sciences*, *41*(1), 18–24.

Azhmyakov, V., Basin, M., & Reincke-Collon, C. (2014). Optimal LQ-type switched control design for a class of linear systems with piecewise constant inputs. In *Proceedings of the int. federation of aut. control*.

Bamieh, B., & Pearson, J. B. (1992). The $\mathcal{H}_2$ problem for sampled-data systems. *Systems & Control Letters*, *19*(1), 1–12.

Bini, E., & Buttazzo, G. (2009). *Design of optimal control systems*. Tesi di Laurea Specialistica.

Bini, E., & Buttazzo, G. (2014). The optimal sampling pattern for linear control systems. *IEEE Transactions on Automatic Control*, *59*(1), 78–90.

Bonnard, B., & Chyba, M. (2003). *The role of singular trajectories in control theory*. Springer Verlag.

Bourdin, L., & Trélat, E. (2013). Pontryagin Maximum Principle for finite dimensional nonlinear optimal control problems on time scales. *SIAM Journal on Control and Optimization*, *51*(5), 3781–3813.

Bourdin, L., & Trélat, E. (2014). General Cauchy-Lipschitz theory for Delta-Cauchy problems with Carathodory dynamics on time scales. *Journal of Difference Equations and Applications*, *20*(4), 526–547.

Bourdin, L., & Trélat, E. (2015). Pontryagin Maximum Principle for optimal sampled-data control problems. In *Proceedings of the IFAC workshop CAO*.

Bourdin, L., & Trélat, E. (2016). Optimal sampled-data control, and generalizations on time scales. *Mathematical Control and Related Fields*, *6*(1), 53–94.

Bryson, J. A. E., & Ho, Y. C. (1975). *Applied optimal control*. Washington, DC: Hemisphere Publishing Corp..

Chen, T., & Francis, B. (1991). $\mathcal{H}_2$-optimal sampled-data control. *IEEE Transactions on Automatic Control*, *36*(4), 387–397.

Chen, T., & Francis, B. (1996). *Optimal sampled-data control systems*. London: Springer-Verlag London, Ltd..

Dontchev, A. L., Hager, W. W., & Veliov, V. M. (2000). Second-order Runge-Kutta approximations in control constrained optimal control. *SIAM Journal on Numerical Analysis*, *38*(1), 202–226.

Fadali, S., & Visioli, A. (2013). *Digital control engineering. Analysis and design*. Elsevier.

Geromel, J. C., & Souza, M. (2015). On an LMI approach to optimal sampled-data state feedback control design. *International Journal of Control*, *88*(11), 2369–2379.

Hager, W. W. (1976). Rates of convergence for discrete approximations to unconstrained control problems. *SIAM Journal on Numerical Analysis*, *13*(4), 449–472.

Hager, W. W. (2000). Runge–Kutta methods in optimal control and the transformed adjoint system. *Numerische Mathematik*, *87*(2), 247–282.

Imura, J. I. (2005). Optimal control of sampled-data piecewise affine systems. *Automatica Journal of IFAC*, *40*(4), 661–669.

Isermann, R. (1989). *Digital control systems. Fundamentals, deterministic control*. Berlin: Springer-Verlag.

Khargonekar, P. P., & Sivashankar, N. (1991). $\mathcal{H}_2$ optimal control for sampled-data systems. *Systems & Control Letters*, *17*(6), 425–436.

Landau, I. D. (2006). *Digital control systems*. Springer.

Lee, E. B., & Markus, L. (1967). *Foundations of optimal control theory*. New York: John Wiley.

Mirkin, L., Rotstein, H. P., & Palmor, Z. (1999). $\mathcal{H}_2$ and $\mathcal{H}_\infty$ design of sampled-data systems using lifting. *SIAM Journal on Control and Optimization*, *38*(1), 175–196.

Nesić, D., & Teel, A. (2001). Sampled-data control of nonlinear systems: an overview of recent results. *Perspectives in Robust Control*, *268*, 221–239.

Pontryagin, L. S., Boltyanskii, V. G., Gamkrelidze, R. V., & Mishchenko, E. F. (1962). *The mathematical theory of optimal processes*. New York-London: John Wiley & Sons, Inc..

Ragazzini, J. R. (1958). *Sampled-data control systems*. McGraw-Hill.

Schättler, H., & Ledzewicz, U. (2012). *Interdisciplinary applied mathematics*: *Vol. 38*. *Geometric optimal control, theory, methods and examples*. Springer.

Souza, M., Vital, G.W.G., & Geromel, J.C. (2014). Optimal sampleddata state feedback control of linear systems. In *Proceedings of the int. federation of aut. control*.

Tou, J. T. (1963). *Optimum design of digital control systems*. Elsevier Science & Technology.

Trélat, E. (2005). *Mathématiques concrètes, Contrôle optimal, théorie & applications*. Paris: Vuibert.

Loïc Bourdin was born in 1986. He received the Master's degree and Ph.D. in Applied Mathematics from the University of Pau (France) in 2010 and 2013 respectively. Since 2014, he is Maître de Conférences at the University of Limoges (France) and since 2016, he is member of SMAI-MODE. His mathematical interests are control theory, optimal control theory, calculus of variations, fractional calculus, time scale calculus, shape optimization and inverse problems.

Emmanuel Trélat was born in 1974. He is currently full professor at University Pierre et Marie Curie (Paris 6) and member of Institut Universitaire de France. He is the director of the Fondation Sciences Mathématiques de Paris. He is editor in chief of the journal ESAIM: Control Calculus of Variations and Optimization, and is associated editor of many other journals. He has been awarded the SIAM Outstanding Paper Prize (2006), Maurice Audin Prize (2010), Felix Klein Prize (European Math. Society, 2012), Blaise Pascal Prize (french Academy of Science, 2014), Big Prize Victor Noury (french Academy of Science, 2016). His research interests range over control theory in finite and infinite dimension, optimal control, stabilization, geometry, numerical analysis.