Automatica 171 (2025) 111944

Contents lists available at ScienceDirect

Automatica

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

Learning-based adaptive optimal control of linear time-delay systems: A value iteration approach ☆

Leilei Cui [a,*], Bo Pang [a], Miroslav Krstić [b], Zhong-Ping Jiang [a]

[a] Control and Networks Lab, Department of Electrical and Computer Engineering, Tandon School of Engineering, New York University, 370 Jay Street, Brooklyn, NY 11201, USA

[b] Department of Mechanical and Aerospace Engineering, University of California, San Diego, USA

ARTICLE INFO

Article history:
Received 17 April 2023
Received in revised form 26 February 2024
Accepted 4 August 2024
Available online xxxx

Keywords:
Learning-based control
Adaptive dynamic programming (ADP)
Linear time-delay systems
Value iteration (VI)

ABSTRACT

This paper proposes a novel learning-based adaptive optimal controller design method for a class of continuous-time linear time-delay systems. A key strategy is to exploit the state-of-the-art reinforcement learning (RL) techniques and adaptive dynamic programming (ADP), and propose a data-driven method to learn the near-optimal controller without the precise knowledge of system dynamics. Specifically, a value iteration (VI) algorithm is proposed to solve the infinite-dimensional Riccati equation for the linear quadratic optimal control problem of time-delay systems using finite samples of input-state trajectory data. It is rigorously proved that the proposed VI algorithm converges to the near-optimal solution. Compared with the previous literature, the nice features of the proposed VI algorithm are that it is directly developed for continuous-time systems without discretization and an initial admissible controller is not required for implementing the algorithm. The efficacy of the proposed methodology is demonstrated by two practical examples of metal cutting and autonomous driving.

1. Introduction

Time delay is ubiquitous in various engineering applications, such as biology, chemistry, economics, and population models (Richard, 2003). In the last several decades, control theory for time-delay systems attracted considerable attention from researchers and practicing engineers, and various stability, robustness, and optimality problems have been studied; see, for instance, Cao and Wang (2018), Fridman (2014), Fridman and Shaked (2002, 2003), Gu, Kharitonov, and Chen (2003), Hale and Lunel (1993), Karafyllis and Jiang (2011), Kolmanovskii and Myshkis (1999), Krstic (2009) and numerous references therein. Under some mild conditions, optimal control can guarantee the performance and the stability of the closed-loop system simultaneously. It is thus not surprising that the optimal control for time-delay systems is fundamentally important, yet challenging, in control theory and practice. Through the classical Bellman's dynamic programming, the study of linear quadratic (LQ) optimal control for systems with state delay was initiated by Krasovskii (1962). It follows from Krasovskii (1962) that the optimal controller is a linear functional of the state and the corresponding optimal performance index is a quadratic functional. Unfortunately, an explicit characterization of the optimal controller is still lacking. Following this seminal work, Ross (1971) and Ross and Flügge-Lotz (1969) explicitly derived a set of partial differential equations (PDEs) to be satisfied by the optimal controller. These equations are the generalization of the Kalman's algebraic Riccati equation (ARE) from delay-free linear time-invariant systems to time-delay systems. The authors of Delfour (1986), Kwong (1980) and Vinter and Kwong (1981) considered the problem in the infinite-dimensional Hilbert space, and generalized the LQ optimal control to systems with state and input delays. It has been found that the optimal solution can be obtained by solving the corresponding infinite-dimensional Riccati equation. Many numerical algorithms were developed to solve the infinite-dimensional Riccati equation for time-delay systems (Banks, Rosen, & Ito, 1984; Burns, Sachs, & Zietsman, 2008; Gibson, 1983). It should be noticed that all the aforementioned methods are model-based, and the performance of the designed controller highly relies on the accuracy of the system model. Recently, facilitated by the tremendous advances in computation and communication technologies, fast data collection

☆ This work has been supported in part by the National Science Foundation grants EPCN-1903781, ECCS-2210320 and ECCS-2210315. The material in this paper was partially presented at the 22nd IFAC World Congress, July 9–14, 2023, Yokohama, Japan. This paper was recommended for publication in revised form by Associate Editor Emilia Fridman under the direction of Editor Florian Dorfler.

* Corresponding author.
E-mail addresses: l.cui@nyu.edu (L. Cui), bo.pang@nyu.edu (B. Pang), krstic@ucsd.edu (M. Krstić), zjiang@nyu.edu (Z.-P. Jiang).

https://doi.org/10.1016/j.automatica.2024.111944

and processing have been made possible for controlling engineering systems with increasing complexity. Hence, it is timely to develop a computational approach to address the learning-based adaptive optimal control of time-delay systems based on reinforcement learning (RL) techniques.

RL is an active branch of machine learning that is aimed at learning optimal controls from data through maximizing a cumulative reward or minimizing a cumulative cost. However, most of the conventional RL algorithms are exclusively devoted to Markov decision processes and discrete-time systems (Sutton & Barto, 2018). The stability of the system in question is often overlooked in the past literature of RL. For real-world applications, e.g. autonomous driving, the system evolves in the continuous (state, input, and time) spaces, and it is critically important that learning provides stability and safety guarantees for the system in closed-loop with RL algorithms for the safe operation of control systems under consideration. Consequently, because of lacking in stability considerations, conventional RL algorithms cannot be directly applied to real-world safety-critical engineering systems. Integrating ideas and techniques from RL and control theory, adaptive dynamic programming (ADP) has been developed to conquer the limitations of the conventional RL algorithms (Jiang, Bian, & Gao, 2020; Jiang & Jiang, 2017; Lewis & Liu, 2013). Different from the conventional RL, ADP exploits the structural knowledge of control systems for the direct design of learning-based controllers from data. It is theoretically shown that the generated controller by ADP iteratively converges to the optimal one. Consequently, the stability of the system is guaranteed under some mild conditions, such as detectability and stabilizability requirements on the system. Recent developments in ADP have led to novel solutions to learning-based optimal control of various important classes of linear/nonlinear/periodic uncertain dynamical systems (Bian & Jiang, 2022; Cui, Başar, & Jiang, 2024; Cui & Jiang, 2023; Cui, Wang, Zhang, Zhang, Lai, Zheng, Zhang, & Jiang, 2021; Gao & Jiang, 2016; Jiang & Jiang, 2012; Pang & Jiang, 2021).

Unlike finite-dimensional systems, continuous-time time-delay systems are infinite-dimensional, which poses a major challenge for the development of learning-based adaptive optimal controller design methods. A common feature of the relevant literature (Asad Rizvi, Wei, & Lin, 2019; Gao & Jiang, 2019; Huang, Jiang, & Ozbay, 2022; Liu, Zhang, Luo, & Han, 2016; Rueda-Escobedo, Fridman, & Schiffer, 2022; Wei, Zhang, Liu, & Zhao, 2010; Zhang, Ren, Mu, & Han, 2022; Zhang, Song, Wei, & Zhang, 2011) is that only discrete-time time-delay systems are considered for learning-based control. Since the discrete-time time-delay systems are fundamentally finite-dimensional and can be transformed to delay-free systems with augmented states, the existing methods are not directly applicable to continuous-time systems with time delays. In Jiang, Zhou, and Liu (2021), Moghadam and Jagannathan (2021) and Moghadam, Jagannathan, Narayanan, and Raghavan (2021), the learning-based control for continuous-time time-delay systems is studied. As pointed out in Moghadam et al. (2021, Remark 9.1), since the designed controllers by the methods in these papers are linear functions (instead of functionals) with respect to the state, the optimality of the system is scarified in these papers to avoid solving the infinite-dimensional Riccati equation. In Cui, Pang, and Jiang (2024), a data-driven policy iteration (PI) algorithm was proposed for solving the adaptive optimal control problem of continuous-time time-delay systems. In that paper, an initial admissible controller is required to start the learning process, which is overly restrictive when the system model is completely unknown. These facts motivate us to develop a learning-based method for solving the adaptive optimal control problem of continuous-time time-delay systems without requiring an accurate dynamic model and an initial admissible controller.

In this paper, based on ADP technique, a value iteration (VI) algorithm is proposed to find the near-optimal controller for linear time-delay systems in the absence of the precise knowledge of system dynamics and an initial admissible controller. It is well-known that for finite-horizon LQ optimal control of delay-free systems, a matrix differential Riccati equation (DRE) should be solved to obtain the optimal value function and controller. The solution of the matrix DRE asymptotically converges to the solution of ARE for infinite-horizon LQ optimal control (Willems, 1971). We demonstrate that the DRE of linear time-delay systems is a set of PDEs, and the convergence property of DRE still holds for linear time-delay systems. By this way, we can approximate the LQ optimal controller of time-delay systems by solving the corresponding DRE. By integrating the convergence property of DRE with the RL technique, a learning-based VI approach is proposed to approximate the optimal controller using only finite samples of input-state data along the trajectories of the system.

The remaining contents of this paper are organized as follows. In Section 2, the preliminaries for the optimal control of time-delay systems are introduced, and the problem studied in the paper is formally formulated. In Section 3, a model-based VI algorithm for time-delay systems is proposed. Based on ADP method, a learning-based VI algorithm is developed along with the convergence analysis in Section 4. In Section 5, the efficacy of the proposed learning-based VI algorithm is numerically demonstrated by two practical examples. Finally, some concluding remarks are given in Section 6.

Notations: In this paper, $\mathbb{R}$ denotes the set of real numbers. $|\cdot|$ denotes the Euclidean norm of a vector or Frobenius norm of a matrix, and $\|\cdot\|_\infty$ denotes the supremum norm of a function. $\mathcal{C}^0(X,Y)$ and $\mathcal{C}^1(X,Y)$ denote the class of continuous functions and the class of continuously differentiable functions from the linear space X to the linear space Y, respectively. $\mathcal{AC}([-\tau,0],\mathbb{R}^n)$ denotes the class of absolutely continuous functions. $\frac{\mathrm{d}f}{\mathrm{d}\theta}(\cdot)$ denotes the function which is the derivative of $f(\cdot)$. $\oplus$ is the direct sum. $L_i([-\tau,0],\mathbb{R}^n)$ denotes the space of measurable functions for which the ith power of the Euclidean norm is Lebesgue integrable, $\mathcal{M}_2=\mathbb{R}^n\oplus L_2([-\tau,0],\mathbb{R}^n)$, and $\mathcal{D}=\left\{[r^\top f^\top(\cdot)]^\top\in\mathcal{M}_2: f\in\mathcal{AC}, \frac{df}{\mathrm{d}\theta}(\cdot)\in L_2, \text{and } f(0)=r\right\}$, where $^\top$ stands for transpose of a vector or matrix. $\langle\cdot,\cdot\rangle$ denotes the inner product in $\mathcal{M}_2$, i.e. $\langle z_1,z_2\rangle=r_1^\top r_2+\int_{-\tau}^0 f_1^\top(\theta)f_2(\theta)\mathrm{d}\theta$, where $z_i=[r_i,f_i(\cdot)]^\top$ for $i=1,2$. $\mathcal{L}(X)$ and $\mathcal{L}(X,Y)$ denote the class of continuous bounded linear operators from X to X and from X to Y respectively. $\otimes$ is the Kronecker product. $\mathrm{vec}(A)=\left[a_1^\top,a_2^\top,\ldots,a_n^\top\right]^\top$, where a_i is the ith column of A. $\mathrm{vec}^{-1}(\cdot)$ is the inverse operator of $\mathrm{vec}(\cdot)$. For $P=P^\top\in\mathbb{R}^{n\times n}$, $\mathrm{vecs}(P)=[p_{11},2p_{12},\ldots,2p_{1n},p_{22},2p_{23},\ldots,2p_{(n-1)n},p_{nn}]^\top$, $\mathrm{vecu}(P)=[2p_{12},\ldots,2p_{1n},2p_{23},\ldots,2p_{(n-1)n}]^\top$, and $\mathrm{diag}(P)=[p_{11},p_{22},\ldots,p_{nn}]^\top$. For two arbitrary vectors $\nu,\mu\in\mathbb{R}^n$, $\mathrm{vecd}(\nu,\mu)=[\nu_1\mu_1,\ldots,\nu_n\mu_n]^\top$, $\mathrm{vecv}(\nu)=[\nu_1^2,\nu_1\nu_2,\ldots,\nu_1\nu_n,\nu_2^2,\ldots,\nu_{n-1}\nu_n,\nu_n^2]^\top$, $\mathrm{vecp}(\nu,\mu)=[\nu_1\mu_2,\ldots,\nu_1\mu_n,\nu_2\mu_3,\ldots,\nu_{n-1}\mu_n]^\top$. $[A]_i$ denotes the ith row of the matrix A, and $[A]_{i,j}$ denotes the submatrix of the matrix A comprised of the entries between the ith and jth rows. $A^\dagger$ denotes the Moore–Penrose inverse of matrix A.

2. Problem formulation and preliminaries

2.1. Problem formulation

This paper considers the following class of continuous-time linear time-delay systems described by:

$$\dot{x}(t)=Ax(t)+A_dx(t-\tau)+Bu(t), \tag{1}$$

where $\tau\geq 0$ denotes the delay of the system, which is constant and known, $x(t)\in\mathbb{R}^n$, and $u(t)\in\mathbb{R}^m$. A, $A_d\in\mathbb{R}^{n\times n}$ and

$B \in \mathbb{R}^{n\times m}$ are unknown constant matrices. The segment of the trajectory for $x(t)$ within the interval $[t-\tau, t]$ is denoted as $x_t(\theta) = x(t+\theta),\ \forall\theta \in [-\tau, 0]$. Since system (1) is infinite dimensional, the system's state is $z(t) = [x^\top(t), x_t^\top(\cdot)]^\top \in \mathcal{M}_2$. Define the linear operators $\mathbf{A} \in \mathcal{L}(\mathcal{M}_2)$, $\mathbf{B} \in \mathcal{L}(\mathbb{R}^m, \mathcal{M}_2)$ as

$$\mathbf{A}z(t) = \begin{bmatrix} Ax(t) + A_d x_t(-\tau) \\ \frac{\mathrm{d}x_t}{\mathrm{d}\theta}(\cdot) \end{bmatrix}, \mathbf{B}u(t) = \begin{bmatrix} Bu(t) \\ 0 \end{bmatrix} \tag{2}$$

Then, as studied in Curtain and Zwart (1995, Theorem 2.4.6), system (1) is equivalent to

$$\dot{z}(t) = \mathbf{A}z(t) + \mathbf{B}u(t), \tag{3}$$

with the domain of $\mathbf{A}$ given by $\mathcal{D}$. Let $z_0 = [x^\top(0), x_0^\top(\cdot)]^\top$ denote the initial state of system (3). The quadratic performance index adopted for system (1) is

$$\begin{aligned} J(x_0, u) &= \int_0^\infty x^\top(t)Qx(t) + u^\top(t)Ru(t)\mathrm{d}t \\ &= \int_0^\infty \langle z(t), \mathbf{Q}z(t)\rangle + u^\top(t)Ru(t)\mathrm{d}t, \end{aligned} \tag{4}$$

where $R^\top = R > 0$, $Q^\top = Q \geq 0$, and $\mathbf{Q} = \begin{bmatrix} Q & \\ & \mathbf{0} \end{bmatrix} \in \mathcal{L}(\mathcal{M}_2)$ is symmetric (Eidelman, Milman, & Tsolomitis, 2004, Chapter 6), and non-negative (Eidelman et al., 2004, Definition 6.3.1). The initial state z_0, Q and R are known.

The following standard assumption is made to ensure the optimal control problem for system (1) with the performance index (4) is solvable. That is, there exists a controller such that the performance index in (4) is finite, and the closed-loop system with the optimal controller is stable at the origin.

Assumption 1. System (1) with the output $y(t) = Q^{\frac{1}{2}}x(t)$ is exponentially stabilizable and detectable (Curtain & Zwart, 1995, Definition 5.2.1), where $Q^{\frac{1}{2}}$ is the unique real symmetric and positive semidefinite matrix such that $(Q^{\frac{1}{2}})^2 = Q$ (Horn & Johnson, 2013, Theorem 7.2.6).

Given the aforementioned assumption, the problem to be studied in this paper can be formulated as follows.

Problem. (**VI-based ADP**) Without knowing the dynamics of system (1), design a VI-based ADP algorithm to find approximations of the optimal controller which can minimize (4) using only the input-state data measured along the trajectories of the system.

2.2. Optimality and stability

For delay-free linear systems, i.e. $A_d = 0$ in (1), as studied by Kalman (1960), the ARE plays a pivotal role in solving the infinite-horizon LQ optimal control problem. Similarly, for system (1), the following lemma gives the expression of the optimal controller for time-delay systems.

Lemma 2 (*Ross & Flügge-Lotz, 1969; Uchida, Shimemura, Kubo, & Abe, 1988*). *Consider system (1) under Assumption 1, the optimal controller that minimizes (4) is*

$$u^*(x_t) = -\underbrace{R^{-1}B^\top P_0^*}_{K_0^*}x(t) - \int_{-\tau}^0 \underbrace{R^{-1}B^\top P_1^*(\theta)}_{K_1^*(\theta)}x_t(\theta)\mathrm{d}\theta \tag{5}$$

and the corresponding minimal performance index is

$$\begin{aligned} V^*(x_0) = x_0^\top(0)P_0^*x_0(0) + 2x_0^\top(0)\int_{-\tau}^0 P_1^*(\theta)x_0(\theta)\mathrm{d}\theta \\ + \int_{-\tau}^0\int_{-\tau}^0 x_0^\top(\xi)P_2^*(\xi,\theta)x_0(\theta)\mathrm{d}\xi\mathrm{d}\theta, \end{aligned} \tag{6}$$

where $P_0^* = P_0^{*\top} \geq 0$, $P_1^*(\theta)$, and $P_2^{*\top}(\theta,\xi) = P_2^*(\xi,\theta)$ for $\theta,\xi \in [-\tau, 0]$ are the unique stabilizing solution to:

$$\begin{aligned} &A^\top P_0^* + P_0^*A - P_0^*BR^{-1}B^\top P_0^* \\ &\quad + P_1^*(0) + P_1^{*\top}(0) + Q = 0, \end{aligned} \tag{7a}$$

$$\partial_\theta P_1^*(\theta) = (A^\top - P_0^*BR^{-1}B^\top)P_1^*(\theta) + P_2^*(0,\theta), \tag{7b}$$

$$(\partial_\xi + \partial_\theta)P_2^*(\xi,\theta) = -P_1^{*\top}(\xi)BR^{-1}B^\top P_1^*(\theta), \tag{7c}$$

$$P_1^*(-\tau) = P_0^*A_d, \quad P_2^*(-\tau,\theta) = A_d^\top P_1^*(\theta). \tag{7d}$$

Remark 3. Define $\mathbf{P}^* \in \mathcal{L}(\mathcal{M}_2)$ as

$$\mathbf{P}^*z = \begin{bmatrix} P_0^*r + \int_{-\tau}^0 P_1^*(\theta)f(\theta)\mathrm{d}\theta \\ \int_{-\tau}^0 P_2^*(\cdot,\theta)f(\theta)\mathrm{d}\theta + P_1^{*\top}(\cdot)r \end{bmatrix},$$

where $z = [r^\top, f^\top(\cdot)]^\top \in \mathcal{M}_2$. Then, it can be found that $\mathbf{P}^*$ is the solution to the following Riccati equation in the infinite-dimensional space:

$$\begin{aligned} 0 = &\langle z_2, \mathbf{\Pi}\mathbf{A}z_1\rangle + \langle \mathbf{A}z_2, \mathbf{\Pi}z_1\rangle \\ &+ \langle z_2, \mathbf{Q}z_1\rangle - \langle \mathbf{\Pi}\mathbf{B}R^{-1}\mathbf{B}^\top\mathbf{\Pi}z_2, z_1\rangle \end{aligned} \tag{8}$$

for $z_1, z_2 \in \mathcal{D}$. Therefore, it is seen that (7) is the concrete expression of the abstract Riccati equation in the infinite-dimensional space, and Lemma 2 can be proved by Curtain and Zwart (1995, Theorem 6.2.4 and Theorem 6.2.7).

Remark 4. By Curtain and Zwart (1995, Theorem 6.2.7) and Assumption 1, the closed-loop system with u^* is exponentially stable at the origin. In practice, the second term in (5) can be numerically calculated by Riemann sum, like midpoint, trapezoid, and Simpson's rules.

3. Continuous-time model-based value iteration

In this section, we will approximate the solution of the infinite-horizon optimal control problem by its finite-horizon counterpart, as the horizon length tends to infinity. Since VI-based ADP is derived from the asymptotic behavior of DRE, which is related to the finite-horizon optimal control problem of (1), we concentrate on investigating the following problem:

$$\min_u \mathcal{J}(t_0, T, \phi, u) = \int_{t_0}^T \langle z(t), \mathbf{Q}z(t)\rangle + u^\top(t)Ru(t)\mathrm{d}t$$

$$\text{subject to (3)}, \tag{9}$$

where $\phi(\theta)$, $\forall\theta \in [-\tau, 0]$, is the initial segment of the state trajectory, t_0 is the initial time, and T is the terminal time of the trajectory. Comparing (9) with the infinite-horizon cost in (4), when $T \to \infty$, (9) is equivalent to (4). The following lemma gives the solution to (9) in the Hilbert space $\mathcal{M}_2$.

Lemma 5 (*Theorem 6.1.13 in Curtain & Zwart, 1995*). *For problem (9), the minimal performance index $V(\phi, t_0) = \min_u \mathcal{J}(t_0, T, \phi, u)$ can be expressed as*

$$V(\phi, t_0) = \langle z, \mathbf{P}(t_0)z\rangle, \tag{10}$$

where $z = [\phi^\top(0), \phi^\top(\cdot)]^\top$ is the initial state at t_0, and $\mathbf{P}(s) \in \mathcal{L}(\mathcal{M}_2)$ is the unique solution to the following DRE for any $z_1, z_2 \in \mathcal{D}$ and $s \in [t_0, T]$,

$$\begin{aligned} \partial_s\langle z_2, \mathbf{P}(s)z_1\rangle = &-\langle z_2, \mathbf{P}(s)\mathbf{A}z_1\rangle - \langle \mathbf{A}z_2, \mathbf{P}(s)z_1\rangle \\ &- \langle z_2, \mathbf{Q}z_1\rangle + \langle \mathbf{P}(s)\mathbf{B}R^{-1}\mathbf{B}^\top\mathbf{P}(s)z_2, z_1\rangle, \\ \mathbf{P}(T) = &\mathbf{0}. \end{aligned} \tag{11}$$

Since $\mathbf{P}(s)$ in Lemma 5 is an abstract linear operator in the Hilbert space $\mathcal{M}_2$, the concrete expression of $V(\phi, t_0)$ is lacking. Based on Lemma 5, the following lemma gives the concrete expression of the linear operator $\mathbf{P}(s) \in \mathcal{L}(\mathcal{M}_2)$. In addition, it is shown that the solution of the finite-horizon optimal control problem converges to the solution of the infinite-horizon counterpart, as the horizon length tends to infinity.

Lemma 6. *For any $z = [\phi^\top(0), \phi^\top(\cdot)]^\top \in \mathcal{M}_2$, the expression of $\mathbf{P}(s)z$ is*

$$\mathbf{P}(s)z = \begin{bmatrix} P_0(s)\phi(0) + \int_{-\tau}^{0} P_1(s,\theta)\phi(\theta)\mathrm{d}\theta \\ \int_{-\tau}^{0} P_2(s,\cdot,\theta)\phi(\theta)\mathrm{d}\theta + P_1^\top(s,\cdot)\phi(0) \end{bmatrix}, \tag{12}$$

where $P_0(s) = P_0^\top(s)$, $P_1(s,\theta)$ and $P_2(s,\xi,\theta) = P_2^\top(s,\theta,\xi)$ can be obtained by solving the following PDEs backwards

$$\begin{aligned}
&\partial_s P_0(s) = -A^\top P_0(s) - P_0(s)A - Q - P_1(s,0) \\
&\quad - P_1^\top(s,0) + P_0(s)BR^{-1}B^\top P_0(s), && \text{(13a)}\\
&\partial_s P_1(s,\theta) = \partial_\theta P_1(s,\theta) - P_2(s,0,\theta) \\
&\quad - (A^\top - P_0(s)BR^{-1}B^\top)P_1(s,\theta), && \text{(13b)}\\
&\partial_s P_2(s,\xi,\theta) = \partial_\xi P_2(s,\xi,\theta) + \partial_\theta P_2(s,\xi,\theta) \\
&\quad + P_1^\top(s,\xi)BR^{-1}B^\top P_1(s,\theta), && \text{(13c)}\\
&P_1(s,-\tau) = P_0(s)A_d,\ P_2(s,-\tau,\theta) = A_d^\top P_1(s,\theta) && \text{(13d)}\\
&P_0(T) = 0, \quad P_1(T,\theta) = 0, \quad P_2(T,\theta,\xi) = 0. && \text{(13e)}
\end{aligned}$$

In addition, under Assumption 1, the following results hold

$$\begin{aligned}
&\lim_{s\to-\infty} |P_0(s) - P_0^*| = 0,\\
&\lim_{s\to-\infty} \|P_1(s,\theta) - P_1^*(\theta)\|_\infty = 0,\\
&\lim_{s\to-\infty} \|P_2(s,\xi,\theta) - P_2^*(\xi,\theta)\|_\infty = 0.
\end{aligned} \tag{14}$$

Proof. See Appendix B.

Lemma 6 implies that the solution of (7) can be well approximated by solving (13) backwards from the terminal time T to $-\infty$. Then, the optimal controller u^* in (5) can be approximated. However, in (13), the system matrices (A, A_d, B) are required and it is non-trivial to solve such complicated PDEs. In the next section, in the absence of the accurate model of the system, a VI-based ADP algorithm will be proposed to solve (13) using the input-state data measured along the system's trajectories. In the rest of the paper, we will call the index t in (1) as physical time, and s in (13) as algorithmic time.

Remark 7. When $A_d = 0$, (1) is reduced to a linear system without time delay. Under this case, according to (13), $P_1(s,\theta) = 0$ and $P_2(s,\xi,\theta) = 0$. As a consequence, the continuous-time VI method proposed in this paper is the same as Bian and Jiang (2016). Therefore, the VI method proposed in this paper is a generalization of the main result in Bian and Jiang (2016) to time-delay systems.

4. Learning-based value iteration

In this section, we suppose only that the continuous-time trajectories of $x(t)$ and $u(t)$ within the time interval $[t_1, t_{L+1}]$ are available for the optimal controller design.

4.1. Algorithm development

Recall that $P_0(s)$, $P_1(s,\theta)$, and $P_2(s,\xi,\theta)$ are the solutions to (13) and the expression of $\mathbf{P}(s)z$ is given in (12). According to (10) and (12), $V(x_t, s)$ can be expressed as

$$\begin{aligned}
V(x_t,s) &= x^\top(t)P_0(s)x(t) + 2x^\top(t)\int_{-\tau}^{0} P_1(s,\theta)x_t(\theta)\mathrm{d}\theta \\
&+ \int_{-\tau}^{0}\int_{-\tau}^{0} x_t^\top(\xi)P_2(s,\xi,\theta)x_t(\theta)\mathrm{d}\xi\mathrm{d}\theta.
\end{aligned} \tag{15}$$

Along the trajectories of system (1) driven by the control input u, considering $\partial_t x(t+\theta) = \partial_\theta x(t+\theta)$ and the partial integration, we have

$$\begin{aligned}
\partial_t V(x_t,s) &= x^\top(t)[A^\top P_0(s) + P_0(s)A \\
&\quad + P_1^\top(s,0) + P_1(s,0)]x(t) \\
&+ 2x^\top(t-\tau)[A_d^\top P_0(s) - P_1^\top(s,-\tau)]x(t) \\
&+ 2x^\top(t)\int_{-\tau}^{0}[A^\top P_1(s,\theta) + P_2(s,0,\theta) \\
&\quad - \partial_\theta P_1(s,\theta)]x_t(\theta)\mathrm{d}\theta \\
&+ 2x^\top(t-\tau)\int_{-\tau}^{0}[A_d^\top P_1(s,\theta) - P_2(s,-\tau,\theta)]x_t(\theta)\mathrm{d}\theta \\
&- \int_{-\tau}^{0}\int_{-\tau}^{0} x_t^\top(\xi)[(\partial_\xi + \partial_\theta)P_2(s,\xi,\theta)]x_t(\theta)\mathrm{d}\xi\mathrm{d}\theta \\
&+ 2u^\top(t)B^\top\left[P_0(s)x(t) + \int_{-\tau}^{0} P_1(s,\theta)x_t(\theta)\mathrm{d}\theta\right].
\end{aligned} \tag{16}$$

Define the following matrix-valued functions

$$\begin{aligned}
&H_0(s) = A^\top P_0(s) + P_0(s)A + P_1^\top(s,0) + P_1(s,0),\\
&H_1(s,\theta) = A^\top P_1(s,\theta) + P_2(s,0,\theta) - \partial_\theta P_1(s,\theta),\\
&H_2(s,\xi,\theta) = \partial_\xi P_2(s,\xi,\theta) + \partial_\theta P_2(s,\xi,\theta),\\
&K_0(s) = R^{-1}B^\top P_0(s),\\
&K_1(s,\theta) = R^{-1}B^\top P_1(s,\theta).
\end{aligned} \tag{17}$$

Then, from Lemma 6, it is seen that as $s \to -\infty$, $H_0(s)$, $H_1(s,\theta)$, $H_2(s,\theta,\xi)$, $K_0(s)$, and $K_1(s,\theta)$ converge to H_0^*, $H_1^*(\theta)$, $H_2^*(\xi,\theta)$, K_0^*, and $K_1^*(\theta)$, where the superscript $*$ denotes that in (17) P_j is replaced by P_j^* for $j = 0, 1, 2$. Since for each fixed algorithmic time $s \in (-\infty, T]$, $H_1(s,\theta)$ and $K_1(s,\theta)$ ($H_2(s,\xi,\theta)$) are continuous functions defined on the interval $[-\tau, 0]$ ($[-\tau, 0]^2$), we use the linear combinations of the basis functions to approximate these continuous functions. Let $\Phi(\theta)$, $\Lambda(\xi,\theta)$, and $\Psi(\xi,\theta)$ denote the N-dimensional linearly independent basis functions. The dimensions of Φ, Λ, and Ψ are assumed to be same without losing generality. Then, by the uniform approximation theory (Powell, 1981), for each fixed algorithmic time $s \in (-\infty, T]$, we have

$$\begin{aligned}
\mathrm{vecs}(H_0) &= W_0(s),\\
\mathrm{vec}(H_1) &= W_1^N(s)\Phi(\theta) + e_{H\Phi}^N(s,\theta),\\
\mathrm{diag}(H_2) &= W_2^N(s)\Psi(\xi,\theta) + e_{H\Psi}^N(s,\xi,\theta),\\
\mathrm{vecu}(H_2) &= W_3^N(s)\Lambda(\xi,\theta) + e_{H\Lambda}^N(s,\xi,\theta),\\
\mathrm{vec}(K_0) &= U_0(s),\\
\mathrm{vec}(K_1) &= U_1^N(s)\Phi(\theta) + e_{K\Phi}^N(s,\theta),
\end{aligned} \tag{18}$$

where $W_0(s) \in \mathbb{R}^{n_1}$, $n_1 = n(n+1)/2$, $W_1^N(s) \in \mathbb{R}^{n^2\times N}$, $W_2^N(s) \in \mathbb{R}^{n\times N}$, $W_3^N(s) \in \mathbb{R}^{n_2\times N}$, $n_2 = n(n-1)/2$, $U_0(s) \in \mathbb{R}^{nm}$, and $U_1^N(s) \in \mathbb{R}^{nm\times N}$ are weighting matrices. $e_{H\Phi}^N(s,\theta)$ and $e_{K\Phi}^N(s,\theta)$ ($e_{H\Psi}^N(s,\xi,\theta)$ and $e_{H\Lambda}^N(s,\xi,\theta)$) are truncation errors, and they converge to zero uniformly in $\theta \in [-\tau, 0]$ ($\xi, \theta \in [-\tau, 0]$), and pointwisely in $s \in (-\infty, T]$, as the number of basis functions N tends to infinity. Specifically, for each fixed $s \in (-\infty, T]$ and any $\epsilon > 0$, there exists $N_1^*(s,\epsilon) > 0$, such that if $N > N_1^*(s,\epsilon)$,

$$\begin{aligned}
&\|e_{H\Phi}^N(s,\theta)\|_\infty \le \epsilon, \quad \|e_{H\Psi}^N(s,\xi,\theta)\|_\infty \le \epsilon,\\
&\|e_{H\Lambda}^N(s,\xi,\theta)\|_\infty \le \epsilon, \quad \|e_{K\Phi}^N(s,\theta)\|_\infty \le \epsilon.
\end{aligned} \tag{19}$$

Remark 8. For time-delay systems, the number of basis functions should be large enough to diminish the truncation errors in (18). In comparison, for the delay-free case, $P_i = 0$ $(i = 1, 2)$, $H_i = 0$ $(i = 1, 2)$, and $K_1 = 0$. The truncation errors in (18) are zero no matter how many basis functions are selected.

Remark 9. In practice, one can choose polynomials as basis functions and the uniform convergence in (19) can be guaranteed by Weierstrass Approximation Theorem (Pugh, 2015). In Lemma 6, $P_2^\top(s,\xi,\theta) = P_2(s,\theta,\xi)$. The diagonal elements of P_2 satisfy $\mathrm{diag}[P_2(s,\xi,\theta)] = \mathrm{diag}[P_2(s,\theta,\xi)]$. Therefore, the basis functions of Ψ should satisfy $\Psi(\xi,\theta) = \Psi(\theta,\xi)$ to meet the requirement.

In Lemma 6, $P_0(s)$, $P_1(s,\theta)$, and $P_2(s,\xi,\theta)$ are continuously differentiable in s. Hence, it is required that the weighting matrices and the truncation errors in (18) are continuously differentiable in s.

Lemma 10. *$W_0(s)$, $W_j^N(s)(j = 1, 2, 3)$, $U_0(s)$, $U_1^N(s)$, $e_{H\Phi}^N(s,\theta)$, $e_{H\Psi}^N(s,\xi,\theta)$, $e_{H\Lambda}^N(s,\xi,\theta)$, and $e_{K\Phi}^N(s,\theta)$ are continuously differentiable in the algorithmic time s.*

Proof. See Appendix C.

Next, the data collected from L intervals within $[t_1, t_{L+1}]$ will be applied to generate the near-optimal controller. Let $t_1 < t_2 < \cdots < t_k < \cdots < t_{L+1}$ denote the boundaries of the sampling intervals. By plugging (17) and (13d) into (16), integrating (16) from t_k to t_{k+1}, and by Lemmas 21 and 22, we have

$$\begin{aligned}
&V(x_{t_{k+1}},s) - V(x_{t_k},s)\\
&= \int_{t_k}^{t_{k+1}} \mathrm{vecv}^\top(x(t))\mathrm{d}t\,\mathrm{vecs}(H_0(s))\\
&+ 2\int_{t_k}^{t_{k+1}}\int_{-\tau}^{0} x_t^\top(\theta)\otimes x^\top(t)\mathrm{vec}(H_1(s,\theta))\mathrm{d}\theta\mathrm{d}t\\
&- \int_{t_k}^{t_{k+1}}\int_{-\tau}^{0}\int_{-\tau}^{0} \mathrm{vecd}^\top(x_t(\xi),x_t(\theta))\\
&\quad \mathrm{diag}(H_2(s,\xi,\theta))\mathrm{d}\xi\mathrm{d}\theta\mathrm{d}t\\
&- \int_{t_k}^{t_{k+1}}\int_{-\tau}^{0}\int_{-\tau}^{0} \mathrm{vecp}^\top(x_t(\xi),x_t(\theta))\\
&\quad \mathrm{vecu}(H_2(s,\xi,\theta))\mathrm{d}\xi\mathrm{d}\theta\mathrm{d}t\\
&+ 2\int_{t_k}^{t_{k+1}} x^\top(t)\otimes(u^\top(t)R)\mathrm{d}t\,\mathrm{vec}(K_0(s))\\
&+ 2\int_{t_k}^{t_{k+1}}\int_{-\tau}^{0} x_t^\top(\theta)\otimes(u^\top(t)R)\mathrm{vec}(K_1(s,\theta))\mathrm{d}\theta\mathrm{d}t.
\end{aligned} \tag{20}$$

To simplify the notations, we define

$$\begin{aligned}
\Gamma_{\Phi xx}(t) &= \int_{-\tau}^{0} \Phi^\top(\theta)\otimes x_t^\top(\theta)\otimes x^\top(t)\mathrm{d}\theta\\
\Gamma_{\Psi xx}(t) &= \int_{-\tau}^{0}\int_{-\tau}^{0} \Psi^\top(\xi,\theta)\otimes \mathrm{vecd}^\top(x_t(\xi),x_t(\theta))\mathrm{d}\xi\,\mathrm{d}\theta\\
\Gamma_{\Lambda xx}(t) &= \int_{-\tau}^{0}\int_{-\tau}^{0} \Lambda^\top(\xi,\theta)\otimes \mathrm{vecp}^\top(x_t(\xi),x_t(\theta))\mathrm{d}\xi\,\mathrm{d}\theta\\
\Gamma_{\Phi\Phi xx}(t) &= \int_{-\tau}^{0}\int_{-\tau}^{0} x_t^\top(\theta)\otimes\Phi^\top(\theta)\\
&\quad \otimes x_t^\top(\xi)\otimes\Phi^\top(\xi)\mathrm{d}\xi\,\mathrm{d}\theta.
\end{aligned} \tag{21}$$

In addition, define the following integrals over $[t_k, t_{k+1}]$

$$\begin{aligned}
I_{xx,k} &= \int_{t_k}^{t_{k+1}} \mathrm{vecv}^\top(x(t))\mathrm{d}t,\\
I_{xu,k} &= \int_{t_k}^{t_{k+1}} x^\top(t)\otimes(u^\top(t)R)\mathrm{d}t,\\
I_{\Phi xx,k} &= \int_{t_k}^{t_{k+1}} \Gamma_{\Phi xx}(t)\mathrm{d}t,\\
I_{\Phi xu,k} &= \int_{t_k}^{t_{k+1}}\int_{-\tau}^{0} \Phi^\top(\theta)\otimes x_t^\top(\theta)\otimes(u^\top(t)R)\mathrm{d}\theta\mathrm{d}t,\\
I_{\Psi xx,k} &= \int_{t_k}^{t_{k+1}} \Gamma_{\Psi xx}(t)\mathrm{d}t,\ I_{\Lambda xx,k} = \int_{t_k}^{t_{k+1}} \Gamma_{\Lambda xx}(t)\mathrm{d}t.
\end{aligned} \tag{22}$$

Plugging (18) and (22) into (20) yields

$$\begin{aligned}
&V(x_{t_{k+1}},s) - V(x_{t_k},s)\\
&= I_{xx,k}W_0(s) + 2I_{\Phi xx,k}\mathrm{vec}(W_1^N(s))\\
&- I_{\Psi xx,k}\mathrm{vec}(W_2^N(s)) - I_{\Lambda xx,k}\mathrm{vec}(W_3^N(s))\\
&+ 2I_{xu,k}U_0(s) + 2I_{\Phi xu,k}\mathrm{vec}(U_1^N(s)) + e_k^N(s),
\end{aligned} \tag{23}$$

where $e_k^N(s)$ is induced by the truncation errors in (18), and it is expressed as

$$\begin{aligned}
e_k^N(s) &= 2\int_{t_k}^{t_{k+1}}\int_{-\tau}^{0} x_t^\top(\theta)\otimes x^\top(t)e_{H\Phi}^N(s,\theta)\mathrm{d}\theta\mathrm{d}t\\
&- \int_{t_k}^{t_{k+1}}\int_{-\tau}^{0}\int_{-\tau}^{0} \mathrm{vecd}^\top(x_t(\xi),x_t(\theta))e_{H\Psi}^N(s,\xi,\theta)\mathrm{d}\xi\mathrm{d}\theta\mathrm{d}t\\
&- \int_{t_k}^{t_{k+1}}\int_{-\tau}^{0}\int_{-\tau}^{0} \mathrm{vecp}^\top(x_t(\xi),x_t(\theta))e_{H\Lambda}^N(s,\xi,\theta)\mathrm{d}\xi\mathrm{d}\theta\mathrm{d}t\\
&+ 2\int_{t_k}^{t_{k+1}}\int_{-\tau}^{0} x_t^\top(\theta)\otimes(u^\top(t)R)e_{K\Phi}^N(s,\theta)\mathrm{d}\theta\mathrm{d}t.
\end{aligned} \tag{24}$$

Stacking (23) for $k = 1, 2, \ldots, L$ into a vector form, one can obtain the following linear equation with respect to the unknown weighting matrices

$$\Theta_N\Omega_N(s) + E_L^N(s) = \Xi(s), \tag{25}$$

where

$$\begin{aligned}
\Omega_N(s) &= [W_0^\top(s), \mathrm{vec}^\top(W_1^N(s)), \mathrm{vec}^\top(W_2^N(s)),\\
&\qquad \mathrm{vec}^\top(W_3^N(s)), U_0^\top(s), \mathrm{vec}^\top(U_1(s))]^\top,\\
\Theta_N &= \left[\sigma_1^\top,\ldots,\sigma_k^\top,\ldots,\sigma_L^\top\right]^\top,\\
E_L^N(s) &= \left[e_1^N(s),\ldots,e_k^N(s),\ldots,e_L^N(s)\right]^\top,\\
\Xi(s) &= \left[V(x_t,s)|_{t=t_1}^{t_2},\ldots,V(x_t,s)|_{t=t_L}^{t_{L+1}}\right]^\top,\\
\sigma_k &= \left[I_{xx,k}, 2I_{\Phi xx,k}, -I_{\Psi xx,k}, -I_{\Lambda xx,k}, 2I_{xu,k}, 2I_{\Phi xu,k}\right].
\end{aligned} \tag{26}$$

The following assumption on the matrix Θ_N is made to ensure that $\Omega_N(s)$ is the unique solution to (25) when applying the method of least squares.

Assumption 11. Given $N > 0$, there exist $L^* > 0$ and $\alpha > 0$ (independent of N), such that for all $L > L^*$,

$$\frac{1}{L}\Theta_N^\top\Theta_N \geq \alpha I. \tag{27}$$

Remark 12. Assumption 11 is reminiscent of the condition of persistent excitation (Åström & Wittenmark, 1997). As shown in the past literature of ADP (Jiang & Jiang, 2017; Lewis & Liu, 2013), one can fulfill such a condition by means of added exploration noise, such as sinusoidal signals and random noise.

In addition, the collected input-state data should be bounded to guarantee the validity of the learning process, which leads to the following assumption.

Assumption 13. For any $t \in [t_1, t_{L+1}]$, $|x(t)|, |u(t)| \leq \beta$, where β is independent of N.

Remark 14. Since the initial policy is not necessarily stabilizing, the system may require resetting for data collection and learning. In detail, to guarantee $|x(t)| \leq \beta$, we can restart the system at the initial state $z_0 = [x_0^\top(0), x_0^\top(\cdot)]^\top$, where $\sup_{\theta\in[-\tau,0]} |x_0(\theta)| \leq \beta$, whenever $|x(t)|$ violates the assumption. We can apply a bounded controller to explore the system, such that $|u(t)| \leq \beta$ is ensured.

Now, at each fixed algorithmic time $s \in (-\infty, T]$, given $P_0(s)$, $P_1(s,\theta)$, and $P_2(s,\xi,\theta)$, one can obtain the expression of $V(x_t, s)$ defined in (15). By solving (25) via the least squares method, one can get the approximation of $\Omega_N(s)$ and the weighting matrices encoded in $\Omega_N(s)$. Consequently, $H_j (j = 0, 1, 2)$, $K_0(s)$ and $K_1(s,\theta)$ can be approximated in the absence of the system matrices (A, A_d, B). Next, by differentiating (25) with respect to the algorithmic time s, we will solve (13) by a data-driven method. Since $V(x_t, s)$ is involved in the expression $\Xi(s)$, the first thing is to differentiate $V(x_t, s)$ with respect to s. By the definition of $V(x_t, s)$ in (10), we have

$$\begin{aligned}\partial_s V(x_t,s) &= x^\top(t)\partial_s P_0(s)x(t)\\ &+ 2x^\top(t)\int_{-\tau}^{0}\partial_s P_1(s,\theta)x_t(\theta)\mathrm{d}\theta\\ &+ \int_{-\tau}^{0}\int_{-\tau}^{0} x_t^\top(\xi)\partial_s P_2(s,\xi,\theta)x_t(\theta)\mathrm{d}\xi\,\mathrm{d}\theta.\end{aligned} \tag{28}$$

Plugging the expressions of $\partial_s P_0(s)$, $\partial_s P_1(s,\theta)$, and $\partial_s P_2(s,\xi,\theta)$ in (13) into (28), and considering the variables defined in (17), we have

$$\begin{aligned}\partial_s V(x_t,s) &= x^\top(t)[-H_0(s) - Q + K_0^\top(s)RK_0(s)]x(t)\\ &+ 2x^\top(t)\int_{-\tau}^{0}[-H_1(s,\theta) + K_0^\top(s)RK_1(s,\theta)]x_t(\theta)\mathrm{d}\theta\\ &+ \int_{-\tau}^{0}\int_{-\tau}^{0} x_t^\top(\xi)[H_2(s,\xi,\theta)\\ &+ K_1^\top(s,\xi)RK_1(s,\theta)]x_t(\theta)\mathrm{d}\xi\,\mathrm{d}\theta.\end{aligned} \tag{29}$$

Then, it follows from (29) that

$$\partial_s V(x_t,s) = \mathcal{W}_N^\top(x_t)\mathcal{V}(\Omega_N(s)) + \varepsilon_N(t,s), \tag{30}$$

where $\varepsilon_N(t,s)$ is induced by the approximation truncation errors, whose expression is given in (D.5). $\mathcal{W}_N(x_t)$ and $\mathcal{V}(\Omega_N(s))$ are defined as

$$\begin{aligned}\mathcal{W}_N(x_t) &= [\mathrm{vecv}^\top(x(t)), 2\Gamma_{\Phi xx}(t),\\ &\quad \Gamma_{\Psi xx}(t), \Gamma_{\Lambda xx}(t), \Gamma_{\Phi\Phi xx}(t)]^\top,\\ \mathcal{V}(\Omega_N(s)) &= \big[[-W_0(s) - \mathrm{vecs}(Q) + \mathcal{K}_{v,0}(s)]^\top,\\ &\quad [-\mathrm{vec}(W_1^N(s)) + \mathcal{U}_0(U_0(s), U_1^N(s), R)]^\top,\\ &\quad \mathrm{vec}^\top(W_2^N(s)), \mathrm{vec}^\top(W_3^N(s)), \mathcal{U}_1^\top(U_1^N(s), R)\big]^T,\end{aligned} \tag{31}$$

where $\mathcal{K}_{v,0}(s)$ is defined in (D.3); the functions $\mathcal{U}_0$ and $\mathcal{U}_1$ are defined in Lemma 23. The detailed derivation of (30) is postponed to Appendix D. It is seen that at the physical time t and algorithmic time s, $\partial_s V(x_t, s)$ is determined by the trajectory segment $x_t(\theta), \theta \in [-\tau, 0]$, the approximate weighting matrices encoded in $\Omega_N(s)$, and $\varepsilon_N(t,s)$ induced by the truncation errors.

Under Assumption 11 and using (30), differentiating the both sides of (25) with respect to the algorithmic time s, we have

$$\begin{aligned}&\partial_s\Omega_N(s) = \mathcal{H}_N(\Omega_N(s)) + \mathcal{G}_N(s),\\ &\Omega_N(T) = 0,\end{aligned} \tag{32}$$

where $\Omega_N(T) = 0$ is obtained by (13e). The expressions of $\mathcal{H}_N(\Omega_N(s))$ and $\mathcal{G}_N(\Omega_N(s), s)$ are

$$\mathcal{H}_N(\Omega_N(s)) = \Theta_N^\dagger \Xi_d^N \mathcal{V}(\Omega_N(s)) \tag{33a}$$

$$\mathcal{G}_N(s) = \Theta_N^\dagger(-\partial_s E_L^N(s) + \Xi_e^N(s)), \tag{33b}$$

$$\Xi_d^N = [\mathcal{W}_N(x_t)|_{t_1}^{t_2}, \ldots, \mathcal{W}_N(x_t)|_{t_L}^{t_{L+1}}]^\top, \tag{33c}$$

$$\Xi_e^N(s) = [\varepsilon_N(t,s)|_{t_1}^{t_2}, \ldots, \varepsilon_N(t,s)|_{t_L}^{t_{L+1}}]^\top. \tag{33d}$$

In (32), $\mathcal{G}_N$ is induced by the truncation errors. Hence, if the truncation errors are small enough to be ignored, the solution to (32) can be approximated by the solution to the following differential equation

$$\partial_s\hat{\Omega}_N(s) = \mathcal{H}_N(\hat{\Omega}_N(s)), \quad \hat{\Omega}_N(T) = 0. \tag{34}$$

With the obtained $\hat{\Omega}_N(s)$, By (18) and (26), the estimation of $K_0(s)$ and $K_1(s,\theta)$ can be obtained by

$$\begin{aligned}\hat{K}_0(s) &= \mathrm{vec}^{-1}([\hat{\Omega}_N(s)]_{n_3,n_4}),\\ \hat{U}_1^N(s) &= \mathrm{vec}^{-1}([\hat{\Omega}_N(s)]_{n_4+1,n_5}),\\ \hat{K}_1^N(s,\theta) &= \mathrm{vec}^{-1}(\hat{U}_1^N(s)\Phi(\theta)).\end{aligned} \tag{35}$$

where $n_3 = n_1 + (n^2 + n + n_2)N + 1$, $n_4 = n_3 + mn$, and $n_5 = n_4 + mnN$.

Algorithm 1 shows the detail of the learning-based VI algorithm. It is seen that only the input-state trajectories collected within the interval $[t_1, t_{L+1}]$ are applied to construct the matrices Θ_N and Ξ_d^N. In addition, since the trajectory data is collected only using the exploratory input u, the algorithm is off-policy.

Algorithm 1 Data-driven Value Iteration

1: Choose T, and the vectors of the basis functions $\Phi(\theta)$, $\Psi(\xi,\theta)$, and $\Lambda(\xi,\theta)$.
2: Choose the boundaries of the sampling intervals $t_1, \cdots t_k, \cdots, t_{L+1}$.
3: Choose the driving input u to explore system (1) and collect the input-state data $u(t), x(t), t \in [t_1, t_{L+1}]$.
4: Construct data matrices Θ_N and Ξ_d^N.
5: Solve (34) backwards on the interval $[0, T]$.
6: Get $\hat{K}_0(0)$ and $\hat{K}_1^N(0,\theta)$ by (35).

4.2. Convergence analysis

This section shows that the obtained control gains $\hat{K}_0(0)$ and $\hat{K}_1^N(0,\theta)$ well approximate the optimal gains K_0^* and $K_1^*(\theta)$ if N and T are chosen large enough. Comparing (32) with (34), the difference between $\Omega_N(s)$ and $\hat{\Omega}_N(s)$ is induced by $\mathcal{G}_N(s)$. As seen from the definitions of $\mathcal{G}_N(s)$ in (33b), $\Xi_e^N(s)$ in (33d), and $E_L^N(s)$ in (26), these three variables are induced by the truncation errors in (18). Hence, the convergence of the truncation errors is investigated.

In (18), as $N \to \infty$, the truncation errors converge to zero uniformly in θ and ξ, and pointwisely in s. The following lemma shows that the truncation errors converge to zero uniformly on any closed sub-interval of $(-\infty, T]$.

Lemma 15. *For any $s' \in (-\infty, T]$, $e_{H\Phi}^N(s,\theta)$ and $e_{K\Phi}^N(s,\theta)$ uniformly converge to 0 on $[s', T] \times [-\tau, 0]$ as $N \to \infty$. Besides, $e_{H\Psi}^N(s,\xi,\theta)$ and $e_{H\Lambda}^N(s,\xi,\theta)$ uniformly converge to 0 on $[s', T] \times [-\tau, 0]^2$ as $N \to \infty$.*

Proof. See Appendix E.

The item $\partial_s E_L^N(s)$ is a major factor in causing $\mathcal{G}_N(s)$ defined in (33b) to be nonzero. From the expression of $E_L^N(s)$ in (26), it is seen that the derivative of the truncation errors is involved in $\partial_s E_L^N(s)$. The following lemma shows that the derivative of the truncation errors with respect to the algorithmic time s converges to zero pointwisely.

Lemma 16. *$\partial_s e^N_{H\Phi}(s,\theta)$ and $\partial_s e^N_{K\Phi}(s,\theta)$ pointwisely converge to 0 on $(-\infty,T]\times[-\tau,0]$ as $N\to\infty$. Besides, $\partial_s e^N_{H\Psi}(s,\xi,\theta)$ and $\partial_s e^N_{H\Lambda}(s,\xi,\theta)$ pointwisely converge to 0 on $(-\infty,T]\times[-\tau,0]^2$ as $N\to\infty$.*

Proof. See Appendix F.

With the convergence of the truncation errors demonstrated in Lemmas 15 and 16, it is shown that $\mathcal{G}_N(s)$ converges to zero as N tends to infinity.

Lemma 17. *For any fixed $s\in(-\infty,T]$ and $\epsilon>0$, there exists $N_2^*(\epsilon,s)>0$, such that $\forall N>N_2^*(\epsilon,s)$, $|\mathcal{G}_N(s)|\le\epsilon$.*

Proof. See Appendix G.

As $\mathcal{G}_N(s)$ is small enough when N tends to infinity, comparing (32) and (34), we demonstrate that the approximation $\hat{\Omega}_N(s)$ is close to the real value $\Omega_N(s)$ over $s\in[s',T]$.

Lemma 18. *For any $\epsilon>0$ and $s'\in(-\infty,T]$, there exists $N_3^*(\epsilon,s')>0$, such that if $N>N_3^*(\epsilon,s')$,*

$$\sup_{s\in[s',T]}|\Omega_N(s)-\hat{\Omega}_N(s)|\le\epsilon. \tag{36}$$

Proof. See Appendix H.

The next theorem shows the main result of the learning-based VI algorithm, i.e. the optimal control gains K_0^* and $K_1^*(\theta)$ can be well approximated by solving (34) backwards.

Theorem 19. *For any $\epsilon>0$, there exist $T^*(\epsilon)>0$ and $N_4^*(\epsilon,T^*)>0$, such that if $T>T^*(\epsilon)$ and $N>N_4^*(\epsilon,T^*)$, the following inequalities hold.*

$$|\hat{U}_0(0)-\mathrm{vec}(K_0^*)|\le\epsilon \tag{37a}$$

$$\|\hat{U}_1^N(0)\Phi(\theta)-\mathrm{vec}(K_1^*(\theta))\|_\infty\le\epsilon \tag{37b}$$

Proof. See Appendix I.

It is noticed that the proposed VI algorithm can well approximate the optimal controller when the number of the basis functions (Φ, Ψ, and Λ) is large enough, and the truncation errors in (18) are sufficiently small. As an important corollary to Theorem 19, the following statement ensures the stability of the closed-loop system with the learning-based controller.

Corollary 20. *There exist $T^*>0$ and $N_5^*>0$, such that if $T>T^*$ and $N>N_5^*$, the closed-loop system with the generated controller $\hat{u}(x_t)$ from Algorithm 1 is exponentially stable at the origin, where $\hat{u}(x_t)$ is*

$$\begin{aligned}&\hat{u}(x_t)=-\hat{K}_0(0)x(t)-\int_{-\tau}^0\hat{K}_1^N(0,\theta)x_t(\theta)\mathrm{d}\theta,\\&\hat{K}_0(0)=\mathrm{vec}^{-1}(\hat{U}_0(0)),\\&\hat{K}_1^N(0,\theta)=\mathrm{vec}^{-1}(\hat{U}_1^N(0)\Phi(\theta)).\end{aligned} \tag{38}$$

Proof. See Appendix J.

5. Practical applications

In this section, we demonstrate the effectiveness of the proposed learning-based VI algorithms by two practical examples, with regard to regenerative chatter in metal cutting and connected and autonomous vehicles (CAVs) in mixed traffic consisting of both autonomous vehicles (AVs) and human-driven vehicles (HDVs).

5.1. Regenerative chatter in metal cutting

Consider the example of regenerative chatter in metal cutting Gu et al. (2003, Example 1.1), Mei, Cherng, and Wang (2005), where the thrust force of the tool is proportional to the instantaneous chip thickness $([x(t)]_1-[x(t-\tau)]_1)$, leading to the time-delay effect. Then, the model can be described by (1) with

$$A=\begin{bmatrix}0&1\\-(c_0+F_t/m)&-c_1/m\end{bmatrix},$$

$$A_d=\begin{bmatrix}0&0\\F_t/m&0\end{bmatrix},\ B=\begin{bmatrix}0\\1/m\end{bmatrix}.$$

In this example, the parameters are chosen as $m=2$, $c_0=10$, $c_1=0.2$, $F_t=1$, and $\tau=1.0$. For the performance index (4), $Q=\mathrm{diag}([100,100])$ and $R=1$. The selection of the basis functions is inspired by the Weierstrass approximation theorem, that is any continuous function over a compact set can be uniformly approximated by polynomials. We use third-order polynomials to approximate $H_1(s,\theta)$ and $K_1(s,\theta)$ in (18), that is $\Phi(\theta)=[1,\theta,\theta^2,\theta^3]^\top$. For the two-variable function $H_2(s,\xi,\theta)$ (s is fixed), we use $\Lambda(\xi,\theta)=[1,\theta,\theta^2,\theta^3]^\top\otimes[1,\xi,\xi^2,\xi^3]^\top$ to approximate its off-diagonal elements. The basis functions for approximating the diagonal elements of $H_2(s,\xi,\theta)$ are chosen based on Remark 9, i.e. $\Psi(\xi,\theta)=[1,\xi+\theta,\xi^2+\theta^2,\xi\theta,\xi^3+\theta^3,\xi^2\theta+\xi\theta^2,\xi^3\theta+\xi\theta^3,\xi^2\theta^2,\xi^3\theta^2+\xi^2\theta^3,\xi^3\theta^3]^\top$. The optimal values of K_0^* and $K_1^*(\theta)$ are numerically computed by discretization in Ross and Flügge-Lotz (1969) for comparison.

For the learning-based VI algorithm, the initial weights of the basis function $\hat{\Omega}_N(T)$ are set as zero, and $T=5$. After data-collection phase, Algorithm 1 is implemented and its convergence is plotted in Fig. 1. The relative errors are $\frac{|\hat{K}_0(0)-K_0^*|}{|K_0^*|}=0.0016$ and $\frac{\|\hat{K}_1^N(0,\theta)-K_1^*(\theta)\|_\infty}{\|K_1^*(\theta)\|_\infty}=0.0379$. After learning phase, the learned controller is tested and the state trajectories are solid lines in Fig. 2. For comparison purpose, we design a model-based state-feedback controller by Moheimani and Petersen (1995), which works for all the unknown delays if the algebraic Riccati Eq. (39) has a stabilizing solution. In detail, the controller is designed as $u_{com}(x(t))=-K_{com}x(t)$, where $K_{com}=R^{-1}B^\top P_{com}$, and $P_{com}=P_{com}^\top>0$ is the solution to

$$\begin{aligned}&A^\top P_{com}+P_{com}A+\gamma G^\top G+Q\\&-P_{com}(BR^{-1}B^\top-\gamma^{-1}FF^\top)P_{com}=0,\end{aligned} \tag{39}$$

where $\gamma=25$, $F=[0,1]^\top$, and $G=[F_t/m,0]$. For the same initial state x_0, the ADP controller learned by Algorithm 1 is compared with u_{com}, which is shown in Fig. 2. The value of the performance index with the ADP controller is 1.7386×10^4, while that of the controller u_{com} is 2.1441×10^4. It shows that our method can find a near optimal controller in the absence of the system dynamics, while the method in Moheimani and Petersen (1995) is model-based and can only guarantee a quadratically bounded cost. Besides, we can see that with the ADP controller, the state converges to the equilibrium more quickly than that of the controller u_{com}.

5.2. CAVs in mixed traffic

Consider a string of two HDVs and one AV as shown in Fig. 3, where h_i denotes the bumper-to-bumper distance between the ith vehicle and $(i-1)$th vehicle, and v_i denotes the velocity of the ith vehicle. Define $\Delta h_i=h_i-h^*$ and $\Delta v_i=v_i-v^*$, where (h^*,v^*) is the equilibrium of the vehicles. h^* depends on the human parameters and $v^*=v_1$. Assuming the velocity of the leading vehicle is constant, and considering the time-delay

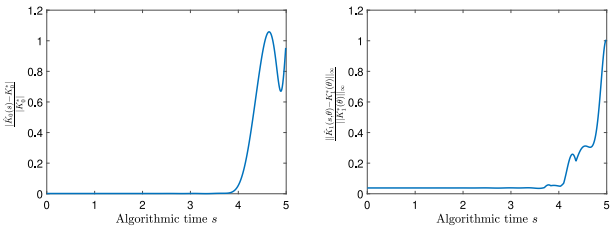


Fig. 1. Convergence of $\hat{K}_0(s)$ and $\hat{K}_1^N(s,\theta)$ to the optimal values K_0^* and $K_1^*(\theta)$ for the example of metal cutting, as the algorithmic time $s \to -\infty$.

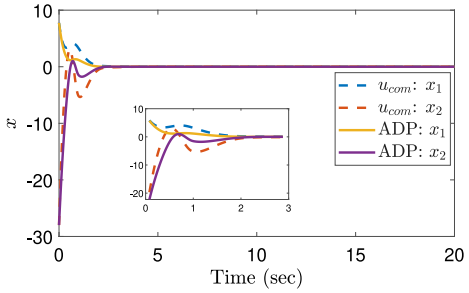


Fig. 2. Comparison between the ADP controller and the model-based method in Moheimani and Petersen (1995) for regenerative chatter in metal cutting.

effect caused by human drivers' reaction time, the system can be described as a linear time-delay system (1) with

$$x = \begin{bmatrix} \Delta h_2 \\ \Delta v_2 \\ \Delta h_3 \\ \Delta v_3 \end{bmatrix}, A = \begin{bmatrix} 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 0 & 0 \end{bmatrix}, B = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix},$$

$$A_d = \begin{bmatrix} 0 & 0 & 0 & 0 \\ \alpha_2 c^* & -(\alpha_2+\beta_2) & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix},$$

where α_2 and β_2 denote the human driver parameters and c^* is the derivative of the range policy (Ge & Orosz, 2017; Huang et al., 2022). In the simulation, the human parameters are set as $\alpha_2 = 0.1$, $\beta_2 = 0.2$, $\tau = 1.2$, and $c^* = 1.5708$. The weighting matrices of the performance index are $Q = \mathrm{diag}([1, 1, 10, 10])$, and $R = 1$. The basis functions are $\Phi(\theta) = [1, \theta]^\top$, $\Psi(\xi, \theta) = [1, \xi+\theta, \xi\theta]^\top$, and $\Lambda(\xi, \theta) = [1, \theta]^\top \otimes [1, \xi]^\top$. The analytical expressions of the optimal control gains K_0^* and K_1^* are derived by the method in Ge and Orosz (2017), where the precise model of the system is required.

For learning-based VI algorithm, the initial weight of the basis function $\hat{\Omega}_N(T)$ is zero. $\hat{\Omega}_N$ is iterated backwards from $T = 10$ to 0 by Algorithm 1. From Fig. 4, it is seen that $\hat{K}_0(s)$ and $\hat{K}_1^N(s,\theta)$ converge to the optimal values eventually, and the relative approximation errors are $\frac{|\hat{K}_0(0)-K_0^*|}{|K_0^*|} = 0.0292$ and $\frac{\|\hat{K}_1^N(0,\theta)-K_1^*(\theta)\|_\infty}{\|K_1^*(\theta)\|_\infty} = 0.0662$. Therefore, the proposed VI algorithm is able to well approximate the optimal controller. Compared with Ge and Orosz (2017), our approach is learning-based and the system model is not required. With the learned ADP controller, the state trajectories of the vehicles are shown in Fig. 5.

6. Conclusions

This paper has proposed for the first time a learning-based VI algorithm for a class of continuous-time linear time-delay systems. The major contributions of the paper are two-fold. First, a model-based VI approach has been developed to solve the infinite-dimensional ARE for the optimal control of linear time-delay systems. Second, by integrating RL and control-theoretic techniques, a learning-based VI algorithm is proposed for learning adaptive optimal controllers from data in the absence of the precise model knowledge. The efficacy of the proposed learning-based adaptive optimal control design method has been validated by means of two real-world applications arising from metal cutting and connected vehicles. Our future work will be directed toward extending the proposed learning-based control methodology to a broader class of linear systems with both input and state delays by combining adaptive predictor technique in Bresch-Pietri and Krstic (2009), Krstic (2009), Zhu and Krstic (2020) with RL and ADP techniques. Furthermore, other practically important classes of time-delay systems, such as nonlinear systems and multi-agent systems, will be studied in the future.

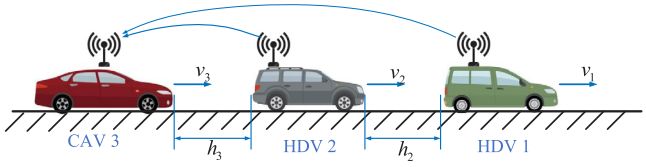


Fig. 3. A string of two HDVs and one AV.

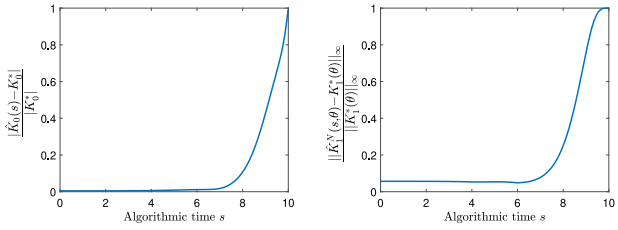


Fig. 4. Convergence of $\hat{K}_0(s)$ and $\hat{K}_1^N(s,\theta)$ to the optimal values K_0^* and $K_1^*(\theta)$ for the example of CAVs, as the algorithmic time $s \to -\infty$.

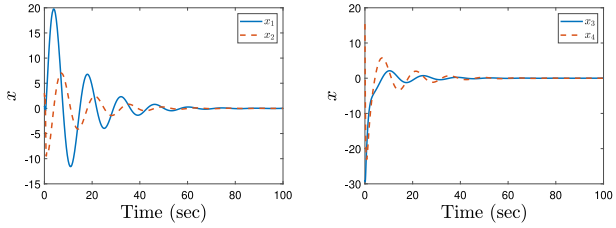


Fig. 5. Plots of the state trajectories of the vehicles with the ADP controller.

Appendix A. Auxiliary results

Some useful formulas for matrix manipulation are listed here.

Lemma 21. *For any matrices X, Y, and Z with compatible dimensions,*

$$\mathrm{vec}(XYZ) = (Z^\top \otimes X)\mathrm{vec}(Y). \tag{A.1}$$

For any real symmetric matrix S and vector v with compatible dimensions,

$$v^\top S v = \mathrm{vecv}^\top(v)\mathrm{vecs}(S). \tag{A.2}$$

Proof. Eq. (A.1) is from Magnus and Neudecker (2007, Theorem 2.2) and (A.2) can be obtained by the quadratic form of $v^\top S v$.

Lemma 22. *For any n-dimensional matrix-valued function $S(\xi,\theta)$ satisfying $S^\top(\xi,\theta)=S(\theta,\xi)$, n-dimensional vector-valued function $\nu(\theta)$, and scalars $a, b\in\mathbb{R}$ with $a<b$, it holds*

$$\begin{aligned}&\int_a^b\int_a^b \nu^\top(\xi)S(\xi,\theta)\nu(\theta)\mathrm{d}\xi\,\mathrm{d}\theta\\&=\int_a^b\int_a^b \mathrm{vecd}^\top(\nu(\xi),\nu(\theta))\mathrm{diag}(S(\xi,\theta))\\&+\mathrm{vecp}^\top(\nu(\xi),\nu(\theta))\mathrm{vecu}(S(\xi,\theta))\mathrm{d}\xi\,\mathrm{d}\theta.\end{aligned}\tag{A.3}$$

Proof. The statement can be directly obtained by the quadratic from of $\nu^T(\xi)S(\xi,\theta)\nu(\theta)$ and noticing

$$\begin{aligned}&\int_a^b\int_a^b s_{ji}(\xi,\theta)\nu_j(\xi)\nu_i(\theta)\mathrm{d}\xi\,\mathrm{d}\theta\\&=\int_a^b\int_a^b s_{ij}(\xi,\theta)\nu_i(\xi)\nu_j(\theta)\mathrm{d}\xi\,\mathrm{d}\theta,\quad\forall i\neq j,\end{aligned}\tag{A.4}$$

where s_{ij} denotes the entry at the ith row and jth column of S.

Lemma 23. *For any $v_0, v_1\in\mathbb{R}^n$, $U_0\in\mathbb{R}^{mn}$, $U_1\in\mathbb{R}^{mn\times N}$, $R\in\mathbb{R}^{m\times m}$, and $\Phi_0,\Phi_1\in\mathbb{R}^N$, it holds:*

$$\begin{aligned}&v_0^\top\mathrm{vec}^{-\top}(U_0)R\mathrm{vec}^{-1}(U_1\Phi_1)v_1\\&=\Phi_1^\top\otimes v_1^\top\otimes v_0^\top\mathcal{U}_0(U_0,U_1,R),\end{aligned}\tag{A.5a}$$

$$\begin{aligned}&v_0^\top\mathrm{vec}^{-\top}(U_1\Phi_0)R\mathrm{vec}^{-1}(U_1\Phi_1)v_1\\&=v_1^\top\otimes\Phi_1^\top\otimes v_0^\top\otimes\Phi_0^\top\mathcal{U}_1(U_1,R),\end{aligned}\tag{A.5b}$$

where $\mathcal{U}_0(U_0,U_1,R)$ is defined as

$$\mathcal{U}_0(U_0,U_1,R)=\mathrm{vec}\left[(I_n\otimes\mathrm{vec}^{-\top}(U_0)R)U_1\right],$$

and $\mathcal{U}_1(U_1,R)$ is

$$\mathcal{U}_1(U_1,R)=\mathrm{vec}(\bar{\mathcal{U}}_1^\top(U_1)R\bar{\mathcal{U}}_1(U_1))$$

$$\bar{\mathcal{U}}_1(U_1)=\begin{bmatrix}[U_1]_1 & [U_1]_{m+1}, & \cdots & [U_1]_{(n-1)m+1}\\ [U_1]_2 & [U_1]_{m+2}, & \cdots & [U_1]_{(n-1)m+2}\\ \vdots & \vdots & & \vdots\\ [U_1]_m & [U_1]_{2m}, & \cdots & [U_1]_{nm}\end{bmatrix}.$$

Proof. By Lemma 21, we have

$$\begin{aligned}&v_0^\top\mathrm{vec}^{-\top}(U_0)R\mathrm{vec}^{-1}(U_1\Phi_1)v_1\\&=v_1^\top\otimes v_0^\top\mathrm{vec}\left[\mathrm{vec}^{-\top}(U_0)R\mathrm{vec}^{-1}(U_1\Phi_1)\right]\\&=v_1^\top\otimes v_0^\top(I_n\otimes\mathrm{vec}^{-\top}(U_0)R)U_1\Phi_1.\end{aligned}\tag{A.6}$$

Hence, (A.5a) holds according to Lemma 21. In addition, since $\mathrm{vec}^{-1}(U_1\Phi_1)v_1=\bar{\mathcal{U}}_1(U_1)(v_1\otimes\Phi_1)$, by Lemma 21, (A.5b) holds.

Appendix B. Proof of Lemma 6

By the expression of $\mathbf{P}(s)z$ in (12), we will write out the expressions for each item in (11). For any $z_i=[f_i^\top(0),f_i^\top(\cdot)]^\top\in\mathcal{M}_2$ $(i=1,2)$, we have

$$\begin{aligned}&\partial_s\langle z_2,\mathbf{P}(s)z_1\rangle\\&=\left\langle z_2,\begin{bmatrix}\partial_sP_0(s)f_1(0)+\int_{-\tau}^0\partial_sP_1(s,\theta)f_1(\theta)\mathrm{d}\theta\\ \int_{-\tau}^0\partial_sP_2(s,\cdot,\theta)f_1(\theta)\mathrm{d}\theta+\partial_sP_1^\top(s,\cdot)f_1(0)\end{bmatrix}\right\rangle\\&=f_2^\top(0)\partial_sP_0(s)f_1(0)+f_2^\top(0)\int_{-\tau}^0\partial_sP_1(s,\theta)f_1(\theta)\mathrm{d}\theta\\&+f_1^\top(0)\int_{-\tau}^0\partial_sP_1(s,\theta)f_2(\theta)\mathrm{d}\theta\\&+\int_{-\tau}^0\int_{-\tau}^0 f_2^\top(\xi)\partial_sP_2(s,\xi,\theta)f_1(\theta)\mathrm{d}\xi\,\mathrm{d}\theta.\end{aligned}\tag{B.1}$$

According to (2), (12), and integration by parts, we have

$$\begin{aligned}&\langle z_2,\mathbf{P}(s)\mathbf{A}z_1\rangle=f_2^\top(0)P_1(s,\theta)f_1(\theta)\big|_{\theta=-\tau}^0\\&+f_2^\top(0)P_0(s)Af_1(0)+f_2^\top(0)P_0(s)A_df_1(-\tau)\\&-f_2^\top(0)\int_{-\tau}^0\partial_\theta P_1(s,\theta)f_1(\theta)\mathrm{d}\theta\\&+[Af_1(0)+A_df_1(-\tau)]^\top\int_{-\tau}^0P_1(s,\theta)f_2(\theta)\mathrm{d}\theta\\&+\int_{-\tau}^0f_2^\top(\xi)P_2(s,\xi,\theta)f_1(\theta)\mathrm{d}\xi\big|_{\theta=-\tau}^0\\&-\int_{-\tau}^0\int_{-\tau}^0f_2^\top(\xi)\partial_\theta P_2(s,\xi,\theta)f_1(\theta)\mathrm{d}\xi\,\mathrm{d}\theta.\end{aligned}\tag{B.2}$$

Following the same lines in the derivation of (B.2), we have

$$\begin{aligned}&\langle\mathbf{A}z_2,\mathbf{P}(s)z_1\rangle=f_2^\top(\theta)P_1^\top(s,\theta)f_1(0)\big|_{\theta=-\tau}^0\\&+f_2^\top(0)A^\top P_0(s)f_1(0)+f_2^\top(-\tau)A_d^\top P_0(s)f_1(0)\\&+[Af_2(0)+A_df_2(-\tau)]^\top\int_{-\tau}^0P_1(s,\theta)f_1(\theta)\mathrm{d}\theta\\&-f_1^\top(0)\int_{-\tau}^0\partial_\theta P_1(s,\theta)f_2(\theta)\mathrm{d}\theta\\&+f_2^\top(\xi)\int_{-\tau}^0P_2(s,\xi,\theta)f_1(\theta)\mathrm{d}\theta\big|_{\xi=-\tau}^0\\&-\int_{-\tau}^0\int_{-\tau}^0f_2^\top(\xi)\partial_\xi P_2(s,\xi,\theta)f_1(\theta)\mathrm{d}\theta\mathrm{d}\xi.\end{aligned}\tag{B.3}$$

Since $\mathbf{Q}=\begin{bmatrix}Q & \\ & \mathbf{0}\end{bmatrix}$, $\langle z_2,\mathbf{Q}z_1\rangle$ is expressed as

$$\langle z_2,\mathbf{Q}z_1\rangle=f_2^\top(0)Qf_1(0).\tag{B.4}$$

Then, according to the expression of $\mathbf{P}(s)z$ in (12) and the expression of $\mathbf{B}$ in (2), we have

$$\begin{aligned}&\langle\mathbf{P}(s)\mathbf{B}R^{-1}\mathbf{B}^\top\mathbf{P}(s)z_2,z_1\rangle\\&=f_2^\top(0)P_0(s)BR^{-1}B^\top P_0(s)f_1(0)\\&+f_1^\top(0)P_0(s)BR^{-1}B^\top\int_{-\tau}^0P_1(s,\theta)f_2(\theta)\mathrm{d}\theta\\&+f_2^\top(0)P_0(s)BR^{-1}B^\top\int_{-\tau}^0P_1(s,\theta)f_1(\theta)\mathrm{d}\theta\\&+\int_{-\tau}^0\int_{-\tau}^0f_1^\top(\xi)P_1^\top(s,\xi)BR^{-1}B^\top P_1(s,\theta)f_2(\theta)\mathrm{d}\xi\,\mathrm{d}\theta.\end{aligned}\tag{B.5}$$

Combining (13) and (B.1) to (B.5) yields that $\mathbf{P}(s)$ defined in (12) satisfies (11). Due to the uniqueness of the solution to (11), the proof is completed.

Before the proof of the second part of Lemma 6, the following lemma is introduced.

Lemma 24 (*Curtain and Zwart 1995, Lemma 6.2.2*)**.** *Under Assumption 1, $\mathbf{P}(s)$ is uniformly bounded with respect to s, i.e. there exists a constant $\nu>0$, such that $\sup_{s\in(-\infty,T]}\|\mathbf{P}(s)\|\leq\nu$.*

According to Lemma 24, $\mathbf{P}(s)$ is uniformly bounded in s. Furthermore, since $\min_u\mathcal{J}(t_0,T,\phi,u)$ is non-decreasing as $t_0\to-\infty$, by (10), we have $\mathbf{P}(a-1)\geq\mathbf{P}(a)\geq0$ for any integer $a\leq T$. By Eidelman et al. (2004, Theorem 6.3.2), there exists $\mathbf{P}=\mathbf{P}^\top\geq0$, such that for all $z\in\mathcal{M}_2$, we have

$$\lim_{a\to-\infty}\mathbf{P}(a)z=\mathbf{P}z.\tag{B.6}$$

Besides, for any $a-1 \le s \le a$, $\mathbf{P}(a) \le \mathbf{P}(s) \le \mathbf{P}(a-1)$. Thus, the following equation holds

$$\lim_{s\to-\infty} \mathbf{P}(s)z = \mathbf{P}z. \tag{B.7}$$

By Lemma 6, when $\mathbf{P}(s)$ converges, $\partial_s P_0(s)$, $\partial_s P_1(s,\theta)$, and $\partial_s P_2(s,\xi,\theta)$ converge to 0, implying that (13) is equivalent to (7) when $s \to -\infty$. Therefore, $P_0(-\infty)$, $P_1(-\infty,\theta)$, and $P_2(-\infty,\xi,\theta)$ satisfy (7). Due to the uniqueness of the solution to (7), $\lim_{s\to-\infty} P_0(s) = P_0^*$, $\lim_{s\to-\infty} P_1(s,\theta) = P_1^*(\theta)$, and $\lim_{s\to-\infty} P_2(s,\xi,\theta) = P_2^*(\xi,\theta)$ pointwisely. Since for any fixed $s \in (-\infty, T]$, $P_1(s,\theta)$, and $P_2(s,\xi,\theta)$ are continuously differentiable on $[-\tau,0]$ and $[-\tau,0]^2$ respectively, $\{P_1(s,\theta) : s \in (-\infty,T]\}$ and $\{P_2(s,\xi,\theta) : s \in (-\infty,T]\}$ are equicontinuous. According to Pugh (2015, Chapter 4, Theorem 16), the pointwise convergence leads to the uniform convergence.

Appendix C. Proof of Lemma 10

The proof is inspired by Bian and Jiang (2022). Take $W_3^N(s)$ and $e_{H\Lambda}^N(s,\xi,\theta)$ as examples. According to (13), $P_2(\cdot,\cdot,\cdot) \in \mathcal{C}^1([-\infty,T]\times[-\tau,0]^2,\mathbb{R}^{n\times n})$. By (17), $H_2(s,\cdot,\cdot) \in \mathcal{C}^0([-\tau,0]^2,\mathbb{R}^{n\times n})$ and $H_2(\cdot,\xi,\theta) \in \mathcal{C}^1([-\infty,T],\mathbb{R}^{n\times n})$. Since for any fixed $s \in (-\infty,T]$, $W_3^N(s)\Lambda(\xi,\theta)$ converges to $\mathrm{vecu}(H_2(s,\xi,\theta))$ uniformly in ξ and θ, for any $s_1, s_2 \in (-\infty,T]$ and $\epsilon > 0$, there exists $N_6^*(s_1,s_2,\epsilon) > 0$, such that if $N > N_6^*(s_1,s_2,\epsilon)$, $|W_3^N(s_i)\Lambda(\xi,\theta) - \mathrm{vecu}(H_2(s_i,\xi,\theta))| \le \epsilon (i=1,2)$ holds for any $\xi,\theta \in [-\tau,0]$. Since $H_2(\cdot,\xi,\theta)$ is uniformly continuous, for any $\epsilon > 0$, there exists $\kappa(\epsilon,\xi,\theta) > 0$, such that if $|s_1 - s_2| < \kappa(\epsilon,\xi,\theta)$, $|\mathrm{vecu}(H_2(s_1,\xi,\theta)) - \mathrm{vecu}(H_2(s_2,\xi,\theta))| \le \epsilon$. Consequently, the following inequality can be obtained by triangle inequality

$$\begin{aligned}
&|(W_3^N(s_1) - W_3^N(s_2))\Lambda(\xi,\theta)| \\
&\le |W_3^N(s_1)\Lambda(\xi,\theta) - \mathrm{vecu}(H_2(s_1,\xi,\theta))| \\
&+ |W_3^N(s_2)\Lambda(\xi,\theta) - \mathrm{vecu}(H_2(s_2,\xi,\theta))| \\
&+ |\mathrm{vecu}(H_2(s_1,\xi,\theta)) - \mathrm{vecu}(H_2(s_2,\xi,\theta))| \le 3\epsilon.
\end{aligned} \tag{C.1}$$

Hence, $W_3^N(s)\Lambda(\xi,\theta)$ is continuous in s. In addition, as the elements of $\Lambda(\xi,\theta)$ are independent, $W_3^N(s)$ is continuous in s. Since $\partial_s H(s,\cdot,\cdot) \in \mathcal{C}^0([-\tau,0]^2,\mathbb{R}^{n\times n})$, by the uniform approximation theory, there exists $W_3'^N(\cdot)$ such that

$$\partial_s \mathrm{vecu}(H(s,\xi,\theta)) = W_3'^N(s)\Lambda(\xi,\theta) + e_{H\Lambda}'^N(s,\xi,\theta), \tag{C.2}$$

where $e_{H\Lambda}'^N(s,\xi,\theta)$ converges to 0 pointwisely in s and uniformly in ξ,θ as N tends to infinity. By the dominated convergence theorem, as $N \to \infty$, for any $s_1 < s_2$, the following equation holds

$$\begin{aligned}
\mathrm{vecu}(H(s,\xi,\theta))\big|_{s_1}^{s_2} &= \int_{s_1}^{s_2} \partial_s \mathrm{vecu}(H(s,\xi,\theta))\mathrm{d}s \\
&= \lim_{N\to\infty} \int_{s_1}^{s_2} W_3'^N(s)\mathrm{d}s\Lambda(\xi,\theta).
\end{aligned} \tag{C.3}$$

Following (18) and (C.3), and by the independence of the elements of Λ, we have $W_3^N(s_2) - W_3^N(s_1) = \int_{s_1}^{s_2} W_3'^N(s)ds$ for any $s_1 < s_2$. Thus, $W_3'^N(s) = \partial_s W_3^N(s)$, i.e. $W_3^N(s)$ is continuously differentiable in s. Since both $H_2(s,\xi,\theta)$ and $W_3(s)$ are continuously differentiable in s, by (18), $e_{H\Lambda}^N(s,\xi,\theta)$ is also continuously differentiable in s.

Appendix D. Derivation of $\partial_s V(x_t, x)$

Rewriting the right hand side of (29) with the help of Lemmas 21 and 22, it follows that

$$\begin{aligned}
\partial_s V(x_t,s) &= \mathrm{vecv}^\top(x(t))[-\mathrm{vecs}(H_0) - \mathrm{vecs}(Q) \\
&+ \mathrm{vecs}(K_0^\top R K_0)] \\
&+ 2\int_{-\tau}^0 x_t^\top(\theta)\otimes x^\top(t)[-\mathrm{vec}(H_1) + \mathrm{vec}(K_0^\top R K_1)]\mathrm{d}\theta \\
&+ \int_{-\tau}^0\int_{-\tau}^0 \mathrm{vecd}^\top(x_t(\xi),x_t(\theta))\mathrm{diag}(H_2) \\
&+ \mathrm{vecp}^\top(x_t(\xi),x_t(\theta))\mathrm{vecu}(H_2) \\
&+ x_t^\top(\theta)\otimes x_t^\top(\xi)\mathrm{vec}(K_1^\top(s,\xi)RK_1(s,\theta))\mathrm{d}\xi\,\mathrm{d}\theta,
\end{aligned} \tag{D.1}$$

where the arguments of the functions $H_0(s)$, $H_1(s,\theta)$, $H_2(s,\xi,\theta)$, $K_0(s)$, and $K_1(s,\theta)$ are omitted to simplify the notations. By the approximations of $K_0(s)$ and $K_1(s,\theta)$ in (18), $\mathrm{vecs}(K_0^\top R K_0)$, $\mathrm{vec}(K_0^\top R K_1)$, and $\mathrm{vec}(K_1^\top R K_1)$ can be expressed as

$$\begin{aligned}
\mathrm{vecs}(K_0^\top R K_0) &= \mathcal{K}_{v,0}(s), \\
\mathrm{vec}(K_0^\top R K_1) &= \mathcal{K}_{v,1}^N(s,\theta) + \mathcal{K}_{e,1}^N(s,\theta), \\
\mathrm{vec}(K_1^\top R K_1) &= \mathcal{K}_{v,2}^N(s,\xi,\theta) + \mathcal{K}_{e,2}^N(s,\xi,\theta),
\end{aligned} \tag{D.2}$$

where $\mathcal{K}_{v,0}$, $\mathcal{K}_{v,1}^N$, and $\mathcal{K}_{v,2}^N$ are constructed by the approximations of K_0 and K_1 in (18); $\mathcal{K}_{e,1}^N$ and $\mathcal{K}_{e,2}^N$ are induced by the approximation truncation errors. They are defined as

$$\begin{aligned}
\mathcal{K}_{v,0} &= \mathrm{vecs}[\mathrm{vec}^{-\top}(U_0(s))R\mathrm{vec}^{-1}(U_0(s))], \\
\mathcal{K}_{v,1}^N &= \mathrm{vec}[\mathrm{vec}^{-\top}(U_0(s))R\mathrm{vec}^{-1}(U_1^N(s)\Phi(\theta))] \\
\mathcal{K}_{e,1}^N &= \mathrm{vec}[\mathrm{vec}^{-\top}(U_0(s))R\mathrm{vec}^{-1}(e_{K\Phi}^N(s,\theta))], \\
\mathcal{K}_{v,2}^N &= \mathrm{vec}[\mathrm{vec}^{-\top}(U_1^N(s)\Phi(\xi))R\mathrm{vec}^{-1}(U_1^N(s)\Phi(\theta))] \\
\mathcal{K}_{e,2}^N &= \mathrm{vec}[\mathrm{vec}^{-\top}(e_{K\Phi}^N(s,\xi))R\mathrm{vec}^{-1}(U_1^N(s)\Phi(\theta) \\
&\quad + e_{K\Phi}^N(s,\theta))] + \mathrm{vec}[\mathrm{vec}^{-\top}(U_1^N(s)\Phi(\xi)) \\
&\quad R\mathrm{vec}^{-1}(e_{K\Phi}^N(s,\theta))].
\end{aligned} \tag{D.3}$$

Plugging (18) and (D.2) into (D.1) and with the help of Lemma 21 gives us the following expression

$$\begin{aligned}
\partial_s V(x_t,s) &= \mathrm{vecv}^\top(x(t))[-W_0(s) - \mathrm{vecs}(Q) + \mathcal{K}_{v,0}(s)] \\
&- 2\Gamma_{\Phi xx}(t)\mathrm{vec}(W_1^N(s)) \\
&+ 2\int_{-\tau}^0 x_t^\top(\theta)\otimes x^\top(t)\mathcal{K}_{v,1}^N(s,\theta)\mathrm{d}\theta \\
&+ \Gamma_{\Psi xx}(t)\mathrm{vec}(W_2^N(s)) + \Gamma_{\Lambda xx}(t)\mathrm{vec}(W_3^N(s)) \\
&+ \int_{-\tau}^0\int_{-\tau}^0 x_t^\top(\theta)\otimes x_t^\top(\xi)\mathcal{K}_{v,2}^N(s,\xi,\theta)\mathrm{d}\xi\,\mathrm{d}\theta + \varepsilon_N(t,s),
\end{aligned} \tag{D.4}$$

where $\Gamma_{\Phi xx}$, $\Gamma_{\Psi xx}$ and $\Gamma_{\Lambda xx}$ are defined in (21). $\varepsilon_N(t,s)$ in (D.4) is induced by the truncation errors in (18), which is

$$\begin{aligned}
\varepsilon_N(t,s) &= \\
&- 2\int_{-\tau}^0 x_t^\top(\theta)\otimes x^\top(t)(e_{H\Phi}^N(s,\theta) - \mathcal{K}_{e,1}^N(s,\theta))\mathrm{d}\theta \\
&+ \int_{-\tau}^0\int_{-\tau}^0 x_t^\top(\theta)\otimes x_t^\top(\xi)\mathcal{K}_{e,2}^N(s,\xi,\theta)\mathrm{d}\xi\,\mathrm{d}\theta \\
&+ \int_{-\tau}^0\int_{-\tau}^0 \mathrm{vecd}^\top(x_t(\xi),x_t(\theta))e_{H\Psi}^N(s,\xi,\theta) \\
&+ \mathrm{vecp}^\top(x_t(\xi),x_t(\theta))e_{H\Lambda}^N(s,\xi,\theta)\mathrm{d}\xi\,\mathrm{d}\theta.
\end{aligned} \tag{D.5}$$

Considering Lemma 23, the integrals in (D.4) involving $\mathcal{K}_{v,1}^N$ and $\mathcal{K}_{v,2}^N$ can be further simplified, and $\partial_s V(x_t,s)$ is finally derived as

$$\begin{aligned}
\partial_s V(x_t,s) &= \mathrm{vecv}^\top(x(t))[-W_0(s) - \mathrm{vecs}(Q) + \mathcal{K}_{v,0}(s)] \\
&+ 2\Gamma_{\Phi xx}(t)[-\mathrm{vec}(W_1^N(s)) + \mathcal{U}_0(U_0(s),U_1^N(s),R)] \\
&+ \Gamma_{\Psi xx}(t)\mathrm{vec}(W_2^N(s)) + \Gamma_{\Lambda xx}(t)\mathrm{vec}(W_3^N(s)) \\
&+ \Gamma_{\Phi\Phi xx}(t)\mathcal{U}_1(U_1^N(s),R) + \varepsilon_N(t,s),
\end{aligned} \tag{D.6}$$

$$= \mathcal{W}_N^\top(x_t)\mathcal{V}(\Omega_N(s)) + \varepsilon_N(t,s). \tag{D.7}$$

Appendix E. Proof of Lemma 15

Take $e_{H\Lambda}^N(s,\xi,\theta)$ as an example. As N tends to infinity, from (18) and (19), it is seen that $e_{H\Lambda}^N(s,\xi,\theta)$ converges to 0 uniformly in $\xi,\theta\in[-\tau,0]$, and pointwisely in $s\in(-\infty,T]$. By Lemma 10, $e_{H\Lambda}^N(s,\xi,\theta)$ is continuously differentiable in $s\in[s',T]$, and hence $\{e_{H\Lambda}^N(s,\xi,\theta): N\in\mathbb{Z}_+\}$ is equicontinuous in s. Therefore, according to Pugh (2015, Chapter 4, Theorem 16), $e_{H\Lambda}^N(s,\xi,\theta)$ uniformly converges to 0 on $[s',T]\times[-\tau,0]^2$.

Appendix F. Proof of Lemma 16

Take $\partial_s e_{H\Lambda}^N(s,\xi,\theta)$ as an example. By Lemma 10, $\partial_s e_{H\Lambda}^N(\cdot,\cdot,\cdot)\in\mathcal{C}^0([-\infty,T]\times[-\tau,0]^2,\mathbb{R}^{n_2})$. According to (18) and (C.2), $\partial_s e_{H\Lambda}^N(s,\xi,\theta)=e_{H\Lambda}^{\prime N}(s,\xi,\theta)$, which converges to 0 pointwisely in s, and uniformly in ξ and θ. Therefore, the proof is completed.

Appendix G. Proof of Lemma 17

As seen from (24), the derivatives of the truncation errors $\partial_s e_{H\Phi}^N(s,\theta)$, $\partial_s e_{K\Phi}^N(s,\theta)$, $\partial_s e_{H\Psi}^N(s,\xi,\theta)$, and $\partial_s e_{H\Lambda}^N(s,\xi,\theta)$ are involved in the expression of $\partial_s e_k^N(s)$. By Lemma 16, these derivatives converge to zero pointwisely as $N\to\infty$. According to the dominated convergence theorem and the boundedness of $x(t)$ and $u(t)$ by Assumption 13, for any fixed s, $\partial_s e_k^N(s)$ converges to zero as $N\to\infty$. Consequently, $\forall s\in(-\infty,T]$ and $\forall\epsilon>0$, there exists $N_7^*(\epsilon,s)>0$, such that if $N>N_7^*(\epsilon,s)$, $|\partial_s E_L^N(s)|\le\sqrt{L}\epsilon$, where $E_L^N(s)$ is defined in (26).

By the expressions of $\mathcal{K}_{e,1}^N$ and $\mathcal{K}_{e,2}^N$ in (D.3), the boundedness of the basis function $\Phi(\theta),\theta\in[-\tau,0]$, and the uniform convergence of $e_{K\Phi}^N(s,\theta)$ from Lemma 15, it is seen that $\mathcal{K}_{e,1}^N(s,\theta)$ and $\mathcal{K}_{e,2}^N(\theta)$ converge to zero uniformly in $s\in[s',T]$ and $\theta\in[-\tau,0]$. By the boundedness of the trajectory $x(t)$ and $u(t)$ from Assumption 13, and the uniform convergence of $e_{H\Phi}^N$, $e_{H\Psi}^N$, $e_{H\Lambda}^N$, $\mathcal{K}_{e,1}^N$, and $\mathcal{K}_{e,2}^N$, it is observed that $\varepsilon_N(t,s)$ in (D.5) is uniformly convergent to zero as $N\to\infty$. Consequently, $\forall s\in[-\infty,T]$ and $\forall\epsilon>0$, there exists $N_8^*(\epsilon,s)>0$, such that if $N>N_8^*(\epsilon,s)$, $|\Xi_e^N(s)|\le\sqrt{L}\epsilon$, where $\Xi_e^N(s)$ is defined in (33d). Therefore, when $N>\max(N_7^*,N_8^*)$

$$|\mathcal{G}_N(s)|\le\frac{2\sqrt{L}}{\sigma_{min}(\Theta_N)}\epsilon\le\frac{2}{\sqrt{\alpha}}\epsilon, \tag{G.1}$$

where the last inequality comes from Assumption 11 and $\sigma_{min}(\Theta_N)$ denotes the minimal singular value of Θ_N. Since α is independent of N, the proof is completed.

Appendix H. Proof of Lemma 18

The proof is inspired by Pang and Jiang (2021). Firstly, assuming the solution to (34) exists on the interval $[s',T]$. It is shown that (36) holds on the interval $[s',T]$. Indeed, let $Z_N(s)=\Omega_N(s)-\hat{\Omega}_N(s)$, and subtracting (34) from (32) yields

$$\begin{aligned}&\partial_s Z_N(s)=\mathcal{H}_N(\Omega_N(s))-\mathcal{H}_N(\hat{\Omega}_N(s))+\mathcal{G}_N(s),\\&Z_N(T)=0.\end{aligned} \tag{H.1}$$

Besides, for $\bar{Z}_N(s)=\Omega_N(s)-\bar{\Omega}_N(s)$, define the following differential equation

$$\begin{aligned}&\partial_s\bar{Z}_N(s)=\mathcal{H}_N(\Omega_N(s))-\mathcal{H}_N(\bar{\Omega}_N(s)),\\&\bar{Z}_N(T)=0.\end{aligned} \tag{H.2}$$

Obviously, $\bar{Z}_N(s)=0$ is the solution to (H.2). Both the right hand sides of (H.1) and (H.2) are locally Lipschitz in $Z_N(s)$ and $\bar{Z}_N(s)$ respectively. Furthermore, according to Lemma 17, for any $\epsilon>0$ and $s\in[s',T]$, there exists $N_2^*(\epsilon,s)$, such that if $N>N_2^*(\epsilon,s)$, $|\mathcal{G}_N(s)|\le\epsilon$. Therefore, by the dominated convergence theorem and Sontag (1998, Theorem 55), there exists $N_9^*(\epsilon,s')>0$, such that if $N>N_9^*(\epsilon,s')$, the following inequality holds

$$\sup_{s\in[s',T]}|Z_N(s)|\le g(\epsilon), \tag{H.3}$$

where $g(\cdot)$ is a $\mathcal{K}_\infty$-function (Khalil, 2002, Definition 4.2). Therefore, $\sup_{s\in[s',T]}|Z_N(s)|\le g(\epsilon)$ can be arbitrary small by setting N large enough.

Next, we will show that the solution to (34) exists on the interval $(-\infty,T]$ when N is large enough. Because $\Omega_N(s)$ exists on $(-\infty,T]$, it is equivalent to prove that $Z_N(s)$ exists on the interval $(-\infty,T]$. For a fixed $N>0$, the right hand side of (H.1) is continuous in s and locally Lipschitz at $Z_N(T)=0$. Therefore, according to Khalil (2002, Theorem 3.1), there exists $S_N<T$, such that (H.1) has a unique solution on $(S_N,T]$. $(S_N,T]$ is the maximal interval for the existence of $Z_N(s)$, that is $\lim_{s\to S_N^+}|Z(s)|=\infty$. For the sequence $\{S_N\}_{N=1}^\infty$, we will show that it is non-increasing by contradiction. Let $N_1<N_2$ and assume $S_{N_1}<S_{N_2}$. Consequently, $\sup_{s\in[S_{N_2},T]}|Z_{N_1}(s)|$ is finite. Since $N_2>N_1$, it follows from (H.3) that $\sup_{s\in[S_{N_2},T]}|Z_{N_2}(s)|$ is finite. This contradicts with the assumption that S_{N_2} is the escape time. Then, we will show that $\lim_{N\to\infty}S_N=-\infty$ by contradiction. Let $S=\lim_{N\to\infty}S_N$, and assume $S>-\infty$. This implies that

$$\lim_{N\to\infty}\left(\lim_{s\to S^+}|Z_N(s)|\right)=\infty \tag{H.4}$$

However, it is seen from (H.3) that for any $S\le s'\le T$,

$$\lim_{N\to\infty}\left(\sup_{s\in[s',T]}|Z_N(s)|\right)=0. \tag{H.5}$$

Therefore, (H.5) contradicts with (H.4). Consequently, $\lim_{N\to\infty}S_N=-\infty$. This implies that the solution to (34) exists on the interval $s\in[s,T]$ when $N\to\infty$.

Appendix I. Proof of Theorem 19

According to Lemma 6 and (17), there exists $T^*(\epsilon)>0$, such that if $T>T^*(\epsilon)$,

$$|U_0(0)-\mathrm{vec}(K_0^*)|\le\frac{\epsilon}{2} \tag{I.1a}$$

$$\|\mathrm{vec}(K_1(0,\theta))-\mathrm{vec}(K_1^*(\theta))\|_\infty\le\frac{\epsilon}{3}. \tag{I.1b}$$

By Lemma 15 and (18), there exists $N_{10}^*(\epsilon)>0$, such that if $N>N_{10}^*(\epsilon)$,

$$\|U_1^N(0)\Phi(\theta)-\mathrm{vec}(K_1(0,\theta))\|_\infty\le\frac{\epsilon}{3}. \tag{I.2a}$$

Following Lemma 18, there exists $N_{11}^*(\epsilon,T)>0$, such that if $N>N_{11}^*(\epsilon,T)$,

$$\begin{aligned}&|\hat{U}_0(0)-U_0(0)|\le\frac{\epsilon}{2}\\&\|(\hat{U}_1^N(0)-U_1^N(0))\Phi(\theta)\|_\infty\le\frac{\epsilon}{3},\end{aligned} \tag{I.3}$$

Therefore, when $T>T^*(\epsilon)$ and $N>N_4^*(\epsilon,T)=\max(N_{10}^*(\epsilon),N_{11}^*(\epsilon,T))$, by triangle inequality, we have

$$\begin{aligned}&|\hat{U}_0(0)-\mathrm{vec}(K_0^*)|\le|\hat{U}_0(0)-U_0(0)|\\&\quad+|U_0(0)-\mathrm{vec}(K_0^*)|\le\epsilon,\\&\|\hat{U}_1^N(0)\Phi(\theta)-\mathrm{vec}(K_1^*(\theta))\|_\infty\\&\quad\le\|(\hat{U}_1^N(0)-U_1^N(0))\Phi(\theta)\|_\infty\\&\quad+\|U_1^N(0)\Phi(\theta)-\mathrm{vec}(K_1(0,\theta))\|_\infty\\&\quad+\|\mathrm{vec}(K_1(0,\theta))-\mathrm{vec}(K_1^*(\theta))\|_\infty\le\epsilon.\end{aligned} \tag{I.4}$$

Appendix J. Proof of Corollary 20

Define the linear operators $\mathbf{K}^* \in \mathcal{L}(\mathcal{M}_2, \mathbb{R}^m)$ and $\hat{\mathbf{K}} \in \mathcal{L}(\mathcal{M}_2, \mathbb{R}^m)$ as

$$\begin{aligned}\mathbf{K}^* z(t) &= K_0^* x(t) + \int_{-\tau}^{0} K_1^*(\theta) x_t(\theta) \mathrm{d}\theta, \\ \hat{\mathbf{K}} z(t) &= \hat{K}_0(0) x(t) + \int_{-\tau}^{0} \hat{K}_1^N(0,\theta) x_t(\theta) \mathrm{d}\theta,\end{aligned} \tag{J.1}$$

where $z(t) = [x^\top(t), x_t^\top(\cdot)]^\top \in \mathcal{M}_2$. Recalling the expressions of the operators $\mathbf{A}$ and $\mathbf{B}$ in (2), and considering the equivalence between (1) and (3), the closed-loop system of (1) with the controller $\hat{u}(x_t)$ is

$$\dot{z}(t) = (\mathbf{A} - \mathbf{B}\mathbf{K}^*) z(t) + \mathbf{B}(\mathbf{K}^* - \hat{\mathbf{K}}) z(t). \tag{J.2}$$

Since system (1) with the optimal controller u^* is exponentially stable at the origin (Remark 4), by Curtain and Zwart (1995, Definition 5.1.1), there exist $c > 0$ and $\omega > 0$, such that

$$\|\mathbf{T}^*(t)\| \leq c e^{-\omega t}, \tag{J.3}$$

where $\mathbf{T}^*(t)$ is the C_0-semigroup (Curtain & Zwart, 1995, Definition 2.1.2) of the system

$$\dot{z}(t) = (\mathbf{A} - \mathbf{B}\mathbf{K}^*) z(t). \tag{J.4}$$

Then, according to Curtain and Zwart (1995, Theorem 3.2.1), the C_0-semigroup of system (J.2), denoted as $\hat{\mathbf{T}}(t)$, satisfies

$$\|\hat{\mathbf{T}}(t)\| \leq c e^{(-\omega + c\|\mathbf{B}(\mathbf{K}^* - \hat{\mathbf{K}})\|)t}. \tag{J.5}$$

By Theorem 19, if $T > T^* = T^*(\frac{\omega}{2\sqrt{2}c\|\mathbf{B}\|})$ and $N > N_5^* = N_4^*(\frac{\omega}{2\sqrt{2}c\|\mathbf{B}\|}, T^*)$, we have

$$\begin{aligned}|K_0^* - \hat{K}_0(0)| &\leq \frac{\omega}{2\sqrt{2}c\|\mathbf{B}\|}, \\ \|K_1^*(\theta) - \hat{K}_1^N(0,\theta)\|_\infty &\leq \frac{\omega}{2\sqrt{2}c\|\mathbf{B}\|}.\end{aligned} \tag{J.6}$$

Considering the expressions of $\mathbf{K}^*$ and $\hat{\mathbf{K}}$ in (J.1), (J.6) implies

$$\|\mathbf{K}^* - \hat{\mathbf{K}}\| \leq \frac{\omega}{2c\|\mathbf{B}\|}. \tag{J.7}$$

Consequently, (J.5) implies,

$$\|\hat{\mathbf{T}}(t)\| \leq c e^{-\frac{1}{2}\omega t}. \tag{J.8}$$

By Curtain and Zwart (1995, Definition 5.1.1), the closed-loop system of (1) with $\hat{u}(x_t)$ is exponentially stable at the origin.

References

Asad Rizvi, S. A., Wei, Y., & Lin, Z. (2019). Model-free optimal stabilization of unknown time delay systems using adaptive dynamic programming. In *Proc. IEEE Conf. Decis. Control.* (pp. 6536–6541). http://dx.doi.org/10.1109/CDC40024.2019.9029600.

Banks, H. T., Rosen, I. G., & Ito, K. (1984). A spline based technique for computing Riccati operators and feedback controls in regulator problems for delay equations. *SIAM Journal on Scientific Computing*, *5*(4), 830–855. http://dx.doi.org/10.1137/0905059.

Bian, T., & Jiang, Z. P. (2016). Value iteration and adaptive dynamic programming for data-driven adaptive optimal control design. *Automatica*, *71*, 348–360. http://dx.doi.org/10.1016/j.automatica.2016.05.003, URL https://www.sciencedirect.com/science/article/pii/S000510981630187X.

Bian, T., & Jiang, Z. P. (2022). Reinforcement learning and adaptive optimal control for continuous-time nonlinear systems: A value iteration approach. *IEEE Transactions on Neural Networks and Learning Systems*, *33*(7), 2781–2790. http://dx.doi.org/10.1109/TNNLS.2020.3045087.

Bresch-Pietri, D., & Krstic, M. (2009). Adaptive trajectory tracking despite unknown input delay and plant parameters. *Automatica*, *45*(9), 2074–2081. http://dx.doi.org/10.1016/j.automatica.2009.04.027, URL https://www.sciencedirect.com/science/article/pii/S0005109809002283.

Burns, J. A., Sachs, E. W., & Zietsman, L. (2008). Mesh independence of Kleinman–Newton iterations for Riccati equations in Hilbert space. *SIAM Journal on Control and Optimization*, *47*(5), 2663–2692. http://dx.doi.org/10.1137/060653962.

Cao, L., & Wang, Y. (2018). Fault-tolerant control for nonlinear systems with multiple intermittent faults and time-varying delays. *International Journal of Control, Automation and Systems*, *16*(2), 609–621.

Cui, L., Başar, T., & Jiang, Z. P. (2024). Robust reinforcement learning for risk-sensitive linear quadratic Gaussian control. *IEEE Transactions on Automatic Control*, 1–16. http://dx.doi.org/10.1109/TAC.2024.3397928.

Cui, L., & Jiang, Z. P. (2023). Learning-based control of continuous-time systems using output feedback. In *SIAM 2023 Proceedings of the Conference on Control and its Applications* (pp. 17–24).

Cui, L., Pang, B., & Jiang, Z. P. (2024). Learning-based adaptive optimal control of linear time-delay systems: A policy iteration approach. *IEEE Transactions on Automatic Control*, *69*(1), 629–636. http://dx.doi.org/10.1109/TAC.2023.3273786.

Cui, L., Wang, S., Zhang, J., Zhang, D., Lai, J., Zheng, Y., et al. (2021). Learning-based balance control of wheel-legged robots. *IEEE Robotics and Automation Letters*, *6*(4), 7667–7674. http://dx.doi.org/10.1109/LRA.2021.3100269.

Curtain, R. F., & Zwart, H. (1995). *An Introduction to Infinite-Dimensional Linear Systems Theory*. New York, NY: Springer.

Delfour, M. C. (1986). The linear-quadratic optimal control problem with delays in state and control variables: a state space approach. *SIAM Journal on Control and Optimization*, *24*(5), 835–883. http://dx.doi.org/10.1137/0324053.

Eidelman, Y., Milman, V., & Tsolomitis, A. (2004). *Functional Analysis, An Introduction*. Rhode Island, USA: American Mathematical Society.

Fridman, E. (2014). Introduction to time-delay systems: Analysis and control. Switzerland: Springer.

Fridman, E., & Shaked, U. (2002). An improved stabilization method for linear time-delay systems. *IEEE Transactions on Automatic Control*, *47*(11), 1931–1937. http://dx.doi.org/10.1109/TAC.2002.804462.

Fridman, E., & Shaked, U. (2003). Delay-dependent stability and H_∞ control: Constant and time-varying delays. *International Journal of Control*, *76*(1), 48–60.

Gao, W., & Jiang, Z. P. (2016). Adaptive dynamic programming and adaptive optimal output regulation of linear systems. *IEEE Transactions on Automatic Control*, *61*(12), 4164–4169. http://dx.doi.org/10.1109/TAC.2016.2548662.

Gao, W., & Jiang, Z. P. (2019). Adaptive optimal output regulation of time-delay systems via measurement feedback. *IEEE Transactions on Neural Networks and Learning Systems*, *30*(3), 938–945. http://dx.doi.org/10.1109/TNNLS.2018.2850520.

Ge, J. I., & Orosz, G. (2017). Optimal control of connected vehicle systems with communication delay and driver reaction time. *IEEE Transactions on Intelligence Transportation Systems*, *18*(8), 2056–2070. http://dx.doi.org/10.1109/TITS.2016.2633164.

Gibson, J. S. (1983). Linear-quadratic optimal control of hereditary differential systems: Infinite dimensional Riccati equations and numerical approximations. *SIAM Journal on Control and Optimization*, *21*(1), 95–139. http://dx.doi.org/10.1137/0321006.

Gu, K., Kharitonov, V. L., & Chen, J. (2003). *Stability of Time-Delay Systems*. Boston, MA: Birkhäuser.

Hale, J. K., & Lunel, S. M. V. (1993). *Introduction to Functional Differential Equations*. New York, NY: Springer-Verlag.

Horn, R. A., & Johnson, C. R. (2013). *Matrix Analysis* (Second). NY, USA: Cambridge University Press.

Huang, M., Jiang, Z. P., & Ozbay, K. (2022). Learning-based adaptive optimal control for connected vehicles in mixed traffic: robustness to driver reaction time. *IEEE Transactions on Cybernetics*, *52*(6), 5267–5277. http://dx.doi.org/10.1109/TCYB.2020.3029077.

Jiang, Z. P., Bian, T., & Gao, W. (2020). Learning-based control: A tutorial and some recent results. *Foundations and Trends in System Control*, *8*(3), 176–284. http://dx.doi.org/10.1561/9781680837537.

Jiang, Y., & Jiang, Z. P. (2012). Computational adaptive optimal control for continuous-time linear systems with completely unknown dynamics. *Automatica*, *48*(10), 2699–2704. http://dx.doi.org/10.1016/j.automatica.2012.06.096, URL https://www.sciencedirect.com/science/article/pii/S0005109812003664.

Jiang, Y., & Jiang, Z. P. (2017). *Robust Adaptive Dynamic Programming*. NJ, USA: Wiley-IEEE Press.

Jiang, H. Y., Zhou, B., & Liu, G. P. (2021). H_∞ optimal control of unknown linear systems by adaptive dynamic programming with applications to time-delay systems. *International Journal of Robust and Nonlinear Control*, *31*(12), 5602–5617. http://dx.doi.org/10.1002/rnc.5557.

Kalman, R. E. (1960). Contributions to the theory of optimal control. *Boletin Sociedad Matematica Mexicana*, *5*(2), 102–119. http://dx.doi.org/10.1109/9780470544334.ch8.

Karafyllis, I., & Jiang, Z. P. (2011). *Stability and Stabilization of Nonlinear Systems*. London, UK: Springer-Verlag.

Khalil, H. (2002). *Nonlinear Systems* (third). Upper Saddle River, New Jersey: Prentice Hall.

Kolmanovskii, V., & Myshkis, A. (1999). *Introduction to the Theory and Applications of Functional Differential Equations*. New York, NY: Kluwer Academic Publishers.

Krasovskii, N. (1962). On the analytic construction of an optimal control in a system with time lags. *Journal of Applied Mathematics and Mechanics*, *26*(1), 50–67. http://dx.doi.org/10.1016/0021-8928(62)90101-6, URL https://www.sciencedirect.com/science/article/pii/0021892862901016.

Krstic, M. (2009). *Delay Compensation for Nonlinear, Adaptive, and PDE Systems*. New York, US: Springer, http://dx.doi.org/10.1007/978-0-8176-4877-0_3.

Kwong, R. H. (1980). A stability theory for the linear-quadratic-Gaussian problem for systems with delays in the state, control, and observations. *SIAM Journal on Control and Optimization*, *18*(1), 49–75. http://dx.doi.org/10.1137/0318004.

Lewis, F. L., & Liu, D. (2013). *Reinforcement Learning and Approximate Dynamic Programming for Feedback Control*. NJ, USA: Wiley-IEEE Press.

Liu, Y., Zhang, H., Luo, Y., & Han, J. (2016). ADP based optimal tracking control for a class of linear discrete-time system with multiple delays. *Journal of the Franklin Institute*, *353*(9), 2117–2136. http://dx.doi.org/10.1016/j.jfranklin.2016.03.012, URL https://www.sciencedirect.com/science/article/pii/S0016003216300862.

Magnus, J. R., & Neudecker, H. (2007). *Matrix Differential Calculus with Applications in Statistics and Econometrics*. New York, US: Wiley.

Mei, C., Cherng, J. G., & Wang, Y. (2005). Active control of regenerative chatter during metal cutting process. *Journal of Manufacturing Science and Engineering*, *128*(1), 346–349. http://dx.doi.org/10.1115/1.2124991.

Moghadam, R., & Jagannathan, S. (2021). Optimal adaptive control of uncertain nonlinear continuous-time systems with input and state delays. *IEEE Transactions on Neural Networks and Learning Systems*, 1–10. http://dx.doi.org/10.1109/TNNLS.2021.3112566.

Moghadam, R., Jagannathan, S., Narayanan, V., & Raghavan, K. (2021). Optimal adaptive control of partially uncertain linear continuous-time systems with state delay. In *Handbook of Reinforcement Learning and Control* (pp. 243–272). New York, NY: Springer, http://dx.doi.org/10.1007/978-3-030-60990-0_9.

Moheimani, S., & Petersen, I. (1995). Optimal quadratic guaranteed cost control of a class of uncertain time-delay systems. *2*, In *Proceedings of 34th IEEE conference on decision and control* (pp. 1513–1518 vol.2). http://dx.doi.org/10.1109/CDC.1995.480352.

Pang, B., & Jiang, Z. P. (2021). Adaptive optimal control of linear periodic systems: An off-policy value iteration approach. *IEEE Transactions on Automatic Control*, *66*(2), 888–894. http://dx.doi.org/10.1109/TAC.2020.2987313.

Powell, M. J. D. (1981). *Approximation Theory and Methods*. New York, NY: Cambridge University Press, http://dx.doi.org/10.1017/CBO9781139171502.

Pugh, C. (2015). *Real Mathematical Analysis* (Second). Switzerland: Springer.

Åström, K. J., & Wittenmark, B. (1997). *Adaptive Control* (Second). MA, USA: Addison-Wesley.

Richard, J. P. (2003). Time-delay systems: an overview of some recent advances and open problems. *Automatica*, *39*(10), 1667–1694. http://dx.doi.org/10.1016/S0005-1098(03)00167-5, URL https://www.sciencedirect.com/science/article/pii/S0005109803001675.

Ross, D. (1971). Controller design for time lag systems via a quadratic criterion. *IEEE Transactions on Automatic Control*, *16*(6), 664–672. http://dx.doi.org/10.1109/TAC.1971.1099834.

Ross, D., & Flügge-Lotz, I. (1969). An optimal control problem for systems with differential-difference equation dynamics. *SIAM Journal on Control and Optimization*, *7*(4), 609–623.

Rueda-Escobedo, J. G., Fridman, E., & Schiffer, J. (2022). Data-driven control for linear discrete-time delay systems. *IEEE Transactions on Automatic Control*, *67*(7), 3321–3336. http://dx.doi.org/10.1109/TAC.2021.3096896.

Sontag, E. (1998). Mathematical control theory, deterministic finite dimensional systems, 2nd ed.. New York, USA: Springer-Verlag.

Sutton, R. S., & Barto, A. G. (2018). *Reinforcement Learning: An Introduction* (Second). The MIT Press, URL http://incompleteideas.net/book/the-book-2nd.html.

Uchida, K., Shimemura, E., Kubo, T., & Abe, N. (1988). The linear-quadratic optimal control approach to feedback control design for systems with delay. *Automatica*, *24*(6), 773–780.

Vinter, R. B., & Kwong, R. H. (1981). The infinite time quadratic control problem for linear systems with state and control delays: An evolution equation approach. *SIAM Journal on Control and Optimization*, *19*(1), 139–153. http://dx.doi.org/10.1137/0319011.

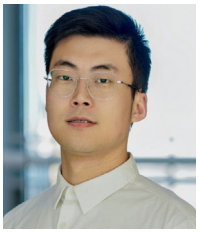
Wei, Q. L., Zhang, H. G., Liu, D., & Zhao, Y. (2010). An optimal control scheme for a class of discrete-time nonlinear systems with time delays using adaptive dynamic programming. *Acta Automatica Sinica*, *36*(1), 121–129. http://dx.doi.org/10.1016/S1874-1029(09)60008-2, URL https://www.sciencedirect.com/science/article/pii/S1874102909600082.

Willems, J. (1971). Least squares stationary optimal control and the algebraic riccati equation. *IEEE Transactions on Automatic Control*, *16*(6), 621–634. http://dx.doi.org/10.1109/TAC.1971.1099831.

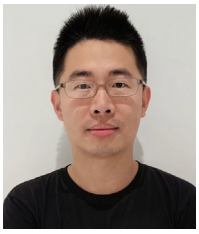
Zhang, H., Ren, H., Mu, Y., & Han, J. (2022). Optimal consensus control design for multiagent systems with multiple time delay using adaptive dynamic programming. *IEEE Transactions on Cybernetics*, *52*(12), 12832–12842. http://dx.doi.org/10.1109/TCYB.2021.3090067.

Zhang, H., Song, R., Wei, Q., & Zhang, T. (2011). Optimal tracking control for a class of nonlinear discrete-time systems with time delays based on heuristic dynamic programming. *IEEE Transactions on Neural Networks*, *22*(12), 1851–1862. http://dx.doi.org/10.1109/TNN.2011.2172628.

Zhu, Y., & Krstic, M. (2020). *Delay-Adaptive Linear Control*. Princeton, NJ, USA: Princeton University Press.



Leilei Cui received the B.Sc. degree in Automation from Northwestern Polytechnical University, Xi'an, China, in 2016, and the M.Sc. degree in Control Science and Engineering from Shanghai Jiao Tong University, Shanghai, China, in 2019. In 2024, he obtained his Ph.D. degree in Electrical Engineering from New York University, NY, U.S. He is currently a postdoctoral associate at Massachusetts Institute of Technology, MA, U.S. His research interests include optimization, optimal control, reinforcement learning, with applications to robotics.



Bo Pang received the B.Sc. Degree in Automation from the Beihang University, Beijing, China, in 2014, and the M.Sc. degree in Control Science and Engineering from Shanghai Jiao Tong University, Shanghai, China, in 2017, and the Ph.D. degree in Electrical Engineering from New York University, NY, U.S.A, in 2021. His research interests include optimal/stochastic control, motion planning, numerical optimization, and reinforcement learning.

Miroslav Krstić is Distinguished Professor of Mechanical and Aerospace Engineering, holds the Alspach endowed chair, and is the founding director of the Center for Control Systems and Dynamics at UC San Diego. He also serves as Senior Associate Vice Chancellor for Research at UCSD. As a graduate student, Krstic won the UC Santa Barbara best dissertation award and student best paper awards at CDC and ACC. Krstic has been elected Fellow of seven scientific societies – IEEE, IFAC, ASME, SIAM, AAAS, IET (UK), and AIAA (Assoc. Fellow) – and as a foreign member of the Serbian Academy of Sciences and Arts and of the Academy of Engineering of Serbia. He has received the Richard E. Bellman Control Heritage Award, Bode Lecture Prize, SIAM Reid Prize, ASME Oldenburger Medal, Nyquist Lecture Prize, Paynter Outstanding Investigator Award, Ragazzini Education Award, IFAC Nonlinear Control Systems Award, IFAC Ruth Curtain Distributed Parameter Systems Award, IFAC Adaptive and Learning Systems Award, Chestnut textbook prize, AV Balakrishnan Award for the Mathematics of Systems, Control Systems Society Distinguished Member Award, the PECASE, NSF Career, and ONR Young Investigator awards, the Schuck ('96 and '19) and Axelby paper prizes, and the first UCSD Research Award given to an engineer. Krstic has also been awarded the Springer Visiting Professorship at UC Berkeley, the Distinguished Visiting Fellowship of the Royal Academy of Engineering, the Invitation Fellowship of the Japan Society for the Promotion of Science, and four honorary professorships outside of the United States. He serves as Editor-in-Chief of Systems & Control Letters and has been serving as Senior Editor in Automatica and IEEE Transactions on Automatic Control, as editor of two Springer book series, and has served as Vice President for Technical Activities of the IEEE Control Systems Society and as chair of the IEEE CSS Fellow Committee. Krstic has coauthored eighteen books on adaptive, nonlinear, and stochastic control, extremum seeking, control of PDE systems including turbulent flows, and control of delay systems.

Zhong-Ping Jiang received the M.Sc. degree in statistics from the University of Paris XI, France, in 1989, and the Ph.D. degree in automatic control and mathematics from ParisTech-Mines, France, in 1993, under the direction of Prof. Laurent Praly.

Currently, he is an Institute Professor in the Department of Electrical and Computer Engineering and an affiliate professor in the Department of Civil and Urban Engineering at the Tandon School of Engineering, New York University. His main research interests include stability theory, robust/adaptive/distributed nonlinear control, robust adaptive dynamic programming, reinforcement learning and their applications to information, mechanical and biological systems. Prof. Jiang is a recipient of the prestigious Queen Elizabeth II Fellowship Award from the Australian Research Council, CAREER Award from the U.S. National Science

Foundation, JSPS Invitation Fellowship from the Japan Society for the Promotion of Science, Distinguished Overseas Chinese Scholar Award from the NSF of China, and several best paper awards. He has served as Deputy Editor-in-Chief, Senior Editor and Associate Editor for numerous journals, and is among the Clarivate Analytics Highly Cited Researchers and Stanford's Top 2% Most Highly Cited Scientists. In 2022, he received the Excellence in Research Award from the NYU Tandon School of Engineering. Prof. Jiang is a foreign member of the Academia Europaea (Academy of Europe) and an ordinary member of the European Academy of Sciences and Arts, and also is a Fellow of the IEEE, IFAC, CAA, AAIA and AAAS.