Available online at www.sciencedirect.com

ScienceDirect

IFAC Papers Online CONFERENCE PAPER ARCHIVE

IFAC PapersOnLine 59-14 (2025) 220–225

Near Optimal Reinforcement Learning Control of Linear Time-Delay Systems

René Martínez * **Luis A. Márquez-Martínez** *
Adolfo Perrusquía **

* *Scientific Research and Higher Education Center of Ensenada (CICESE). Carr. Tijuana-Ensenada No. 3918, Zona Playitas, 22860 Ensenada, B.C. Mexico, (e-mail: mtrujillo@cicese.edu.mx).*
** *Faculty of Engineering and Applied Sciences, Cranfield University, College Road, MK43 0AL, Bedford, UK*

Abstract: The control of time-delay linear systems is a common challenge in real-world autonomous system applications. Time delays can negatively affect the stability and performance of controllers, necessitating the exploration of alternative approaches. In this context, this paper proposes the implementation of a reinforcement learning (RL)-based policy iteration (PI) algorithm by transforming a time-delay system into an augmented state approximate linear system. This transformation is achieved by segmenting the delay into discrete delays, which allows for the application of RL algorithms to solve the optimal control problem. Through simulation studies in scenarios such as chemical plants and regenerative chatter systems, the effectiveness of the proposed methodology is demonstrated and associated challenges are identified. This approach offers a solution to address the complexity of controller design for time-delay systems, facilitating system management through approximations.

Copyright © 2025 The Authors. This is an open access article under the CC BY-NC-ND license (https://creativecommons.org/licenses/by-nc-nd/4.0/)

Keywords: Optimal control, Time-delay systems, Machine learning, Reinforcement learning, Linear systems.

1. INTRODUCTION

Linear models face significant challenges in modern autonomous systems, especially in complex environments, revealing their limitations (Guo et al., 2023). A key issue is time delay or latency, inherent in real-world systems, which compromises the effectiveness traditional control approaches (Kwon and Park, 2019; Witrant et al., 2016), prompting the need for more adaptive strategies.

Time delays may be intrinsic or externally introduced, resulting from energy, material or information transport, or actuator and sensor response time (Sereni et al., 2023). These delays affect system state, output and control input, compromising the stability and performance (Wei et al., 2022; Hua et al., 2018; Fridman, 2014).

In autonomous system networks, these issues are even more noticeable, as data transmission latency, irregular sampling, and loss of information present significant challenges to the design and analysis of controllers (El Debeiki et al., 2024). This is of particular significance in applications such as autonomous vehicles (Ge and Orosz, 2017; Huang et al., 2022; Carlucho et al., 2023), where control decisions must be made in fractions of seconds. Any delay in communication between sensors and actuators can result in loss of accuracy or safety failures. Another example is collaborative robotics (An et al., 2023), where robots coordinate and share real-time information, also face efficiency issues due to latency or data loss, leading to misscoordination and errors. In industrial contexts like teleoperation in factories or distributed plants (Kebria et al., 2019, 2020), feedback delays significantly affect process quality and efficiency. In all of these cases, latency, irregular sampling, and information loss increase controller design complexity and impose strict constraints on system stability and robustness.

1.1 Related works

One of the most significant challenges in continuous time-delay systems is ensuring the stability of the closed-loop system trajectories. The delay, both in the state and in the input, introduces additional dynamics that can potentially lead to instability if they are not addressed properly (Gu et al., 2003). Moreover, whilst stability is an essential condition for the operation of a delay system, often it is preferable for a system to exhibit a prescribed optimal behavior to achieve specific performance objectives.

The optimal control problem for continuous-time linear systems with state delay was initially studied by Krasovskii (1962), where the authors use dynamic programming to establish a set of sufficient conditions for optimality in an infinite time horizon. These results yield the design of a control policy strategy that considers the delay in its design (Ross, 1971; Eller et al., 1969). Here, the solution to the optimal policy requires solving an infinite-dimensional Riccati equation, which presents significant computational challenges due to the incorporation of partial differential equations (PDEs). In addition, these algorithms assume complete knowledge of the system dynamics, which is often challenging in practice because of

2405-8963 Copyright © 2025 The Authors. This is an open access article under the CC BY-NC-ND license.
Peer review under responsibility of International Federation of Automatic Control.
10.1016/j.ifacol.2025.12.153

the difficulty in obtaining an accurate model of a system with time delay.

In this context, recent advances in reinforcement learning (RL) (Sutton and Barto, 2018; Yu and Perrusquia, 2021) and adaptive dynamic programming (ADP) (Liu et al., 2020) have been widely applied to obtain the optimal control policy that drives the system to exhibit an optimal control performance given some cost index by only using measurements collected from the interaction of the system with the controller. These methodologies have consistently demonstrated their potential for designing optimal/near optimal control algorithms that exhibit both adaptive and optimal performances without requiring knowledge of the system model (Vamvoudakis et al., 2021). Integral reinforcement learning (IRL) is the baseline of ADP/RL algorithms for continuous time systems (Moghadam and Jagannathan, 2019) which lie in the recursive relation of the Bellman equation.

ADP/RL techniques have been developed for different classes of dynamical systems, both linear and specific nonlinear classes (Liu et al., 2020). These approaches include the design of value iteration and policy iteration algorithms (Lian et al., 2024; Perrusquía, 2022) based on either gradient or batch-least squares (BLS) rules. On the one hand, gradient rules are used to update in each time step a well-defined parameterization of a Hamilton-Jacobi-Bellman (HJB) equation or algebraic Riccati equation (ARE). Gradient rules are widely employed in RL algorithms that has an actor-critic structure (Vamvoudakis et al., 2021; Ganie and Jagannathan, 2024). On the other hand, BLS-based algorithms collect rich and heterogeneous data from the on-line interaction to construct a well-defined linear system of equations to estimate the parameters of the optimal control problem (Palanisamy et al., 2014). BLS algorithms are usually preferred over gradient methods due to their stable performance for data collection. Nevertheless, the learning time of BLS algorithms increase for relative large scale systems.

There are few systematic approaches available for linear continuous systems with time delay due to their infinite-dimensional nature. Usually time-delay systems have been studied in discrete-time which avoids the infinite-dimensionality problem of continuous time systems with time-delay (Rizvi et al., 2021). In recent years, data-driven policy iteration (Cui et al., 2023a) and value iteration (Cui et al., 2023b) approaches were proposed to solve the optimal control problem of time-delay systems. However, the complexity of the overall algorithm increases even for small scale time-delay systems.

In view of the above, this paper proposes a near-optimal control for a class of continuous linear time delay systems using ADP/RL. The key idea is to transform the time-delay system into a large scale linear system that can be solved by any value or policy iteration algorithms. This is done by segmenting the delay into discrete-time delays that yield to a extended state vector that includes the current state and the r discrete states with delay. This large-scale system allows to design a near optimal control policy that optimizes a pre-defined cost index. Simulation studies using a chemical refining plant and regenerative chatter system are provided to show the benefits of the approach and the future research challenges.

2. OPTIMAL CONTROL PROBLEM FOR LINEAR TIME-DELAY SYSTEMS

Consider a linear time-delay system of the form

$$\dot{\boldsymbol{x}}(t) = \boldsymbol{A}_0\boldsymbol{x}(t) + \boldsymbol{A}_1\boldsymbol{x}(t-\tau) + \boldsymbol{B}\boldsymbol{u}(t) \tag{1}$$

where $\boldsymbol{x}(t) \in \mathbb{R}^n$ represents the system state and $\boldsymbol{u}(t) \in \mathbb{R}^m$ corresponds to the control input. The matrices, $\boldsymbol{A}_0, \boldsymbol{A}_1 \in \mathbb{R}^{n\times n}$ and $\boldsymbol{B} \in \mathbb{R}^{n\times m}$, describe the unknown drift dynamics and input matrix, respectively. The system constant delay is denoted as $\tau \geq 0$. Let denote $\boldsymbol{x}_t(\theta) = \boldsymbol{x}(t+\theta)$ as the trajectory segment of (1) over the interval $[t-\tau, t]$ for all $\theta \in [-\tau, 0]$.

Assumption 1. The delay τ of the system is assumed known or that it can be estimated using any delay approximation technique, e.g., coherence techniques (Carter, 1993).

Consider an infinite horizon cost given by

$$J = \int_0^\infty \left(\boldsymbol{x}^\top(t)\boldsymbol{Q}\boldsymbol{x}(t) + \boldsymbol{u}^\top(t)\boldsymbol{R}\boldsymbol{u}(t)\right) dt, \tag{2}$$

where $\boldsymbol{Q} = \boldsymbol{Q}^\top \succeq 0 \in \mathbb{R}^{n\times n}$ and $\boldsymbol{R} = \boldsymbol{R}^\top \succ 0 \in \mathbb{R}^{m\times m}$ are positive semi-definite and definite weight matrices.

Assumption 2. The system (1) *with output* $\boldsymbol{y} = \boldsymbol{Q}^{1/2}\boldsymbol{x}(t)$ *is exponentially stabilizable and detectable (Curtain and Zwart, 2012).*

The objective of the optimal control problem is to find a control policy $u^*(t)$ that minimizes the quadratic cost function given by (2), subject to (1).

Definition 1. Admissible Control (Ross, 1971; Ross and Flügge-Lotz, 1969): A control policy input $\boldsymbol{u}(t) \in \mathcal{L}_2$ is admissible for the system (1) with respect to (2), if for any initial condition $\boldsymbol{x}_t$, the value function (2) is finite.

Lemma 1. (Cui et al. (2023a)). *Consider a linear-time delay system* (1) *that satisfies Assumption 2. Then the optimal control that minimizes* (2) *is given by*

$$\begin{aligned}\boldsymbol{u}^*(t) &= -\boldsymbol{K}_0\boldsymbol{x}(t) - \int_{-\tau}^0 \boldsymbol{K}_1(\theta)\boldsymbol{x}_t(\theta)d\theta \\ &= -\boldsymbol{R}^{-1}\boldsymbol{B}^\top\left[\boldsymbol{P}_0^*\boldsymbol{x}(t) - \int_{-\tau}^0 \boldsymbol{P}_1^*(\theta)\boldsymbol{x}_t(\theta)d\theta\right], \end{aligned}\tag{3}$$

and the minimal cost index is given by

$$\begin{aligned}V^*(\boldsymbol{x}_0) =& \boldsymbol{x}^\top(0)\boldsymbol{P}_0^*\boldsymbol{x}(0) + 2\boldsymbol{x}^\top(0)\int_{-\tau}^0 \boldsymbol{P}_1^*(\theta)\boldsymbol{x}_0(\theta)d\theta \\ &+ \int_{-\tau}^0\int_{-\tau}^0 \boldsymbol{x}_0^\top(\xi)\boldsymbol{P}_2^*\boldsymbol{x}_0(\xi)d\xi d\theta, \end{aligned}\tag{4}$$

where $\boldsymbol{P}_0^* = \boldsymbol{P}_0^{*\top} \succ 0$, $\boldsymbol{P}_1^*(\theta)$ *and* $\boldsymbol{P}_2^*(\xi,\theta) = \boldsymbol{P}_2^{*\top}(\xi,\theta)$ *are the unique solutions to the following partial differential equations (PDEs) for all* $\theta, \xi \in [-\tau, 0]$.

$$
\begin{aligned}
&\boldsymbol{A}_0^\top \boldsymbol{P}_0^* + \boldsymbol{P}_0^* \boldsymbol{A}_0 - \boldsymbol{P}_0^* \boldsymbol{B}\boldsymbol{R}^{-1}\boldsymbol{B}^\top \boldsymbol{P}_0^* + {\boldsymbol{P}_1^*}^\top(0) \\
&\quad + \boldsymbol{P}_1^*(0) + \boldsymbol{Q} = \boldsymbol{0}_n, \\
&\frac{d\boldsymbol{P}_1^*(\theta)}{d\theta} = \left(\boldsymbol{A}_0^\top - \boldsymbol{P}_0^* \boldsymbol{B}\boldsymbol{R}^{-1}\boldsymbol{B}^\top\right) \boldsymbol{P}_1^*(\theta) + \boldsymbol{P}_2^*(0,\theta), \\
&\frac{\partial \boldsymbol{P}_2^*(\xi,\theta)}{\partial \xi} + \frac{\partial \boldsymbol{P}_2^*(\xi,\theta)}{\partial \theta} = -{\boldsymbol{P}_1^*}^\top(\xi)\boldsymbol{B}\boldsymbol{R}^{-1}\boldsymbol{B}^\top \boldsymbol{P}_1^*(\theta), \\
&\boldsymbol{P}_1^*(-\tau) = \boldsymbol{P}_0^* \boldsymbol{A}_1, \\
&\boldsymbol{P}_2^*(-\tau,\theta) = \boldsymbol{A}_1^\top \boldsymbol{P}_1^*(\theta),
\end{aligned} \tag{5}
$$

Proof. Refer to Curtain and Zwart (2012).

The closed-loop system between the time-delay system (1) under the optimal control (3) is exponentially stable at the origin. Notice that the PDEs are strongly nonlinear even if the matrices $\boldsymbol{A}_0, \boldsymbol{A}_1$ and $\boldsymbol{B}$ are known. Also observe that the algebraic Riccati equation (ARE) for linear systems is recovered when delay is not considered. In the next section we transform the time-delay system into a augmented state linear system that helps to obtain an analytic solution of the optimal control problem for time-delay system and avoid computing the PDEs (5).

3. TIME-DELAY SYSTEMS AS AUGMENTED STATE LINEAR SYSTEMS

In this paper, we use a finite difference technique (LeVeque, 2007) to replace the PDEs in the optimal control solution. This allows to transform the time-delay system into an augmented state finite-dimensional system that has a unique solution.

Let $r \in \mathbb{N}$ be a natural number that divides the interval $[-\tau, 0]$ into equal-sized segments whose endpoints are $\theta = \frac{-i\tau}{r}$, with $i = 0, 1, \ldots, r$. Since we are working with continuous-time systems then the r aggregate states with delays need to be expressed as dynamic states. To do so, we consider an implicit Euler integration method as follows

$$
\dot{\boldsymbol{x}}\left(t - \frac{(i+1)\tau}{r}\right) = \frac{\boldsymbol{x}\left(t - i\frac{\tau}{r}\right) - \boldsymbol{x}\left(t - (i+1)\frac{\tau}{r}\right)}{\tau/r}. \tag{6}
$$

This allows to construct an augmented state linear system with the following matrices

$$
\boldsymbol{A}_r = \begin{bmatrix}
\boldsymbol{A}_0 & \boldsymbol{0}_n & \boldsymbol{0}_n & \cdots & \cdots & \boldsymbol{A}_1 \\
\frac{r}{\tau}\boldsymbol{I}_n & -\frac{r}{\tau}\boldsymbol{I}_n & \boldsymbol{0}_n & \boldsymbol{0}_n & \cdots & \boldsymbol{0}_n \\
\boldsymbol{0}_n & \frac{r}{\tau}\boldsymbol{I}_n & -\frac{r}{\tau}\boldsymbol{I}_n & \boldsymbol{0}_n & \cdots & \boldsymbol{0}_n \\
\vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\
\boldsymbol{0}_n & \cdots & \cdots & \boldsymbol{0}_n & \frac{r}{\tau}\boldsymbol{I}_n & -\frac{r}{\tau}\boldsymbol{I}_n
\end{bmatrix}, \boldsymbol{B}_r = \begin{bmatrix} \boldsymbol{B} \\ \boldsymbol{0}_{n\times m} \\ \vdots \\ \boldsymbol{0}_{n\times m} \end{bmatrix}
$$

and $\boldsymbol{Q}_r = \operatorname{diag}\{\boldsymbol{Q}, \boldsymbol{0}_n, \cdots, \boldsymbol{0}_n\} \in \mathbb{R}^{N\times N}$, where $\boldsymbol{A}_r \in \mathbb{R}^{N\times N}$ and $\boldsymbol{B}_r \in \mathbb{R}^{N\times m}$ with $N = n(r+1)$.

Assumption 3. It is assumed that the pair $(\boldsymbol{A}_r, \boldsymbol{B}_r)$ is controllable and stabilizable for $r \geq 1$ by the Popov-Belevitch-Hautus (PBH) test (Hespanha, 2018) and the pair $(\boldsymbol{A}_r, \boldsymbol{Q}_r^{1/2})$ is detectable.

This allows to represent the time-delay system as the following augmented state linear system

$$
\dot{\boldsymbol{z}}(t) = \boldsymbol{A}_r \boldsymbol{z}(t) + \boldsymbol{B}_r \boldsymbol{u}(t) \tag{7}
$$

where $\boldsymbol{z}(t) = \left[\boldsymbol{x}^\top(t), \boldsymbol{x}^\top\left(t - \frac{\tau}{r}\right), \ldots, \boldsymbol{x}^\top(t-\tau)\right]^\top \in \mathbb{R}^N$.

The cost index (2) is rewritten as

$$
J = \int_0^\infty \left(\boldsymbol{z}^\top(t)\boldsymbol{Q}_r \boldsymbol{z}(t) + \boldsymbol{u}^\top(t)\boldsymbol{R}\boldsymbol{u}(t)\right) dt. \tag{8}
$$

Notice that (7) is a linear system that adds discrete delay elements into the state vector. The Hamilton-Jacobi-Bellman (HJB) equation of (7) under the cost (8) is

$$
\begin{aligned}
H(\boldsymbol{z}, \boldsymbol{u}) =& \boldsymbol{z}^\top(t)\left(\boldsymbol{A}_r^\top \boldsymbol{P} + \boldsymbol{P}\boldsymbol{A}_r + \boldsymbol{Q}_r\right)\boldsymbol{z}(t) \\
&+ 2\boldsymbol{z}^\top(t)\boldsymbol{P}\boldsymbol{B}_r\boldsymbol{u}(t) + \boldsymbol{u}^\top(t)\boldsymbol{R}\boldsymbol{u}(t) \equiv 0,
\end{aligned} \tag{9}
$$

where $\boldsymbol{P} = \boldsymbol{P}^\top \succ 0 \in \mathbb{R}^{N\times N}$ is the unique kernel matrix solution of the following ARE

$$
\boldsymbol{A}_r^\top \boldsymbol{P} + \boldsymbol{P}\boldsymbol{A}_r - \boldsymbol{P}\boldsymbol{B}_r\boldsymbol{R}^{-1}\boldsymbol{B}_r^\top \boldsymbol{P} + \boldsymbol{Q}_r = \boldsymbol{0}_N. \tag{10}
$$

The kernel matrix $\boldsymbol{P}$ has the following form

$$
\boldsymbol{P} = \begin{bmatrix}
\boldsymbol{P}_{0,0} & \boldsymbol{P}_{0,1} & \cdots & \boldsymbol{P}_{0,r} \\
\boldsymbol{P}_{1,0} & \boldsymbol{P}_{1,1} & \cdots & \boldsymbol{P}_{1,r} \\
\vdots & \vdots & \ddots & \vdots \\
\boldsymbol{P}_{r,0} & \boldsymbol{P}_{r,1} & \cdots & \boldsymbol{P}_{r,r}
\end{bmatrix} \tag{11}
$$

where $\boldsymbol{P} \in \mathbb{R}^{N\times N}$, $\boldsymbol{P}_{i,j} \in \mathbb{R}^{n\times n}$ and $\boldsymbol{P}_{i,j} = \boldsymbol{P}_{j,i}^\top$.

The near optimal control is obtained by differentiating the Hamiltonian $\frac{\partial H(\cdot)}{\partial \boldsymbol{u}(t)} = 0$ and solving for $\boldsymbol{u}(t)$. Then, the near optimal control policy is given by

$$
\boldsymbol{u}(t) = -\boldsymbol{R}^{-1}\boldsymbol{B}_r^\top \boldsymbol{P}\boldsymbol{z}(t) = -\boldsymbol{K}\boldsymbol{z}(t), \tag{12}
$$

where $\boldsymbol{K} = \boldsymbol{R}^{-1}\boldsymbol{B}^\top \boldsymbol{P} \in \mathbb{R}^{N\times m}$ is a near optimal control gain. From the couple matrix (11) we can observe that

$$
\begin{aligned}
\boldsymbol{P}_0 &\approx \boldsymbol{P}_{0,0} \\
\boldsymbol{P}_1(-i/r) &\approx r\boldsymbol{P}_{0,i+1}, \quad 0 \leq i \leq r-1, \\
\boldsymbol{P}_1(-\tau) &= \boldsymbol{P}_0\boldsymbol{A}_1 \approx \boldsymbol{P}_{0,0}\boldsymbol{A}_1,
\end{aligned} \tag{13}
$$

which provides a direct relation between the optimal control solution obtained from the PDEs (5) and the solution provided by the augmented state linear system. Notice that the matrix $\boldsymbol{P}_2(\xi,\theta)$ does not appear in this parameterization.

However, the ARE (10) requires knowledge of the augmented state system dynamics. In the following section, we develop a model-free policy iteration (PI) RL algorithm (Perrusquía and Guo, 2023) to solve the optimal control problem of time-delay systems under an augmented state linear system parameterization.

4. PI REINFORCEMENT LEARNING

Consider the augmented state system (7) written as

$$
\dot{\boldsymbol{z}}(t) = \boldsymbol{A}_r \boldsymbol{z}(t) + \boldsymbol{B}_r(\boldsymbol{u}(t) - \boldsymbol{v}_i(t) + \boldsymbol{v}_i(t)) \tag{14}
$$

where $\boldsymbol{v}_i(t) = -\boldsymbol{\mathcal{K}}_i \boldsymbol{z}(t) \in \mathbb{R}^m$ is a stabilizing control policy and $\boldsymbol{\mathcal{K}}_i \in \mathbb{R}^{m\times m}$ is a control gain at iteration i. Then (14) is rewritten as

$$
\begin{aligned}
\dot{\boldsymbol{z}}(t) &= (\boldsymbol{A}_r - \boldsymbol{B}_r\boldsymbol{\mathcal{K}}_i)\boldsymbol{z}(t) + \boldsymbol{B}_r(\boldsymbol{u}(t) - \boldsymbol{v}_i(t)) \\
&= \boldsymbol{A}_K \boldsymbol{z}(t) + \boldsymbol{B}_r \widetilde{\boldsymbol{u}}_i(t),
\end{aligned} \tag{15}
$$

where $\widetilde{\boldsymbol{u}}_i(t) = \boldsymbol{u}(t) - \boldsymbol{v}_i(t)$ and $\boldsymbol{A}_K = \boldsymbol{A}_r - \boldsymbol{B}_r\boldsymbol{\mathcal{K}}_i$. Here, we avoid the dependency of t for simplicity. Define $\boldsymbol{\mathcal{K}}_{i+1} = \boldsymbol{R}^{-1}\boldsymbol{B}_r^\top \boldsymbol{P}_i$ where $\boldsymbol{P}_i \in \mathbb{R}^{N\times N}$ is an iterative kernel matrix. Then the Hamilton-Jacobi-Bellman (HJB) equation (9) is rewritten in terms of the system (15) as

$$
\begin{aligned}
H(\boldsymbol{z}, \boldsymbol{u}, \boldsymbol{v}_i) =& \left(\boldsymbol{A}_K\boldsymbol{z} + \boldsymbol{B}\widetilde{\boldsymbol{u}}_i\right)^\top \boldsymbol{P}_i\boldsymbol{z} + \boldsymbol{z}^\top \boldsymbol{P}_i\left(\boldsymbol{A}_K\boldsymbol{z} + \boldsymbol{B}\widetilde{\boldsymbol{u}}_i\right) \\
&+ \boldsymbol{z}^\top \boldsymbol{Q}_r\boldsymbol{z} + \boldsymbol{v}_i^\top \boldsymbol{R}\boldsymbol{v}_i - 2\widetilde{\boldsymbol{u}}_i^\top \boldsymbol{R}\boldsymbol{\mathcal{K}}_{i+1}\boldsymbol{z} \equiv 0.
\end{aligned} \tag{16}
$$

We integrate the right-hand side of (16) from t to $t+T$ to obtain the integral reinforcement learning (IRL) equation of the form

$$\begin{aligned} &\boldsymbol{z}^\top(t+T)\boldsymbol{P}_i\boldsymbol{z}(t+T) - \boldsymbol{z}^\top(t)\boldsymbol{P}_i\boldsymbol{z}(t) \\ &\quad - 2\int_t^{t+T} \widetilde{\boldsymbol{u}}_i^\top(\sigma)\boldsymbol{R}\boldsymbol{\mathcal{K}}_{i+1}\boldsymbol{z}(\sigma)d\sigma \\ &\quad = -\int_t^{t+T} \left(\boldsymbol{z}^\top(\sigma)\boldsymbol{Q}_r\boldsymbol{z}(\sigma) + \boldsymbol{u}_i^\top(\sigma)\boldsymbol{R}\boldsymbol{u}_i(\sigma)\right)d\sigma, \quad (17) \end{aligned}$$

where $T>0$ is a sampling time. Then, we collect κ samples and construct the following matrices

$$\begin{aligned} \boldsymbol{Z} &= \left[\boldsymbol{z}(\sigma)\bar{\otimes}\boldsymbol{z}(\sigma)\Big|_t^{t+T} \cdots \boldsymbol{z}(\sigma)\bar{\otimes}\boldsymbol{z}(\sigma)\Big|_{t+(\kappa-1)T}^{t+\kappa T}\right]^\top \\ \boldsymbol{I}_{uz} &= \left[\int_t^{t+T} \boldsymbol{z}\otimes(\boldsymbol{R}\widetilde{\boldsymbol{u}}_i)d\sigma \cdots \int_{t+(\kappa-1)T}^{t+\kappa T} \boldsymbol{z}\otimes(\boldsymbol{R}\widetilde{\boldsymbol{u}}_i)d\sigma\right]^\top \\ \boldsymbol{I}_\rho &= \left[\int_t^{t+T} \rho(\sigma)d\sigma \cdots, \int_{t+(\kappa-1)T}^{t+\kappa T} \rho(\sigma)d\sigma\right]^\top, \quad (18) \end{aligned}$$

where $\rho(\sigma) = \boldsymbol{z}^\top(\sigma)\boldsymbol{Q}_r\boldsymbol{z}(\sigma) + \boldsymbol{u}_i^\top\boldsymbol{R}\boldsymbol{u}_i$, $\boldsymbol{Z}\in\mathbb{R}^{\kappa\times\iota}$, $\boldsymbol{I}_{uz}\in\mathbb{R}^{\kappa\times mN}$, and $\boldsymbol{I}_\rho\in\mathbb{R}^\kappa$ with $\iota=\frac{1}{2}N(N+1)$. Here, $\otimes$ and $\bar{\otimes}$ denotes the standard and symmetric Kronecker product. Define $\boldsymbol{\Omega} = [\boldsymbol{Z}, -2\boldsymbol{I}_{uz}] \in \mathbb{R}^{\kappa\times(\iota+mN)}$, then the new improved kernel $\boldsymbol{P}_i$ and control gain $\boldsymbol{\mathcal{K}}_{i+1}$ are computed using a simple batch-least squares rule

$$\begin{bmatrix} \text{vech}(\boldsymbol{P}_i) \\ \text{vec}(\boldsymbol{\mathcal{K}}_{i+1}) \end{bmatrix} = -(\boldsymbol{\Omega}^\top\boldsymbol{\Omega})^{-1}\boldsymbol{\Omega}^\top\boldsymbol{I}_\rho, \quad (19)$$

where $\text{vec}(\cdot)$ and $\text{vech}(\cdot)$ stand to the vectorization and half vectorization operators, respectively.

Remark 1. The matrix $\boldsymbol{\Omega}$ must be a full rank matrix, i.e., rank$(\boldsymbol{\Omega}) = \iota + mN$ to ensure parameter estimates convergence in each reinforcement leaning iteration i.

5. SIMULATION STUDIES

To demonstrate the effectiveness of the approach, we present two numerical examples. We only report the results of the delay discretization that gives better results in terms of computational resources and closed-loop stability.

5.1 Chemical refining plant

Consider a chemical refining plant with transportation delay Ross (1971) described by

$$\boldsymbol{A}_0 = \begin{bmatrix} -4.93 & -1.01 & 0 & 0 \\ -3.20 & -5.30 & -12.8 & 0 \\ 6.40 & 0.347 & -32.5 & -1.04 \\ 0 & 0.833 & 11.0 & -3.96 \end{bmatrix},$$

$$\boldsymbol{A}_1 = \begin{bmatrix} 1.92 & 0 & 0 & 0 \\ 0 & 1.92 & 0 & 0 \\ 0 & 0 & 1.87 & 0 \\ 0 & 0 & 0 & 0.724 \end{bmatrix}, \boldsymbol{B} = \begin{bmatrix} \boldsymbol{I}_2 \\ \boldsymbol{0}_2 \end{bmatrix},$$

and $\tau=1$. The weight matrices of the cost index (2) are set to $\boldsymbol{Q} = \text{diag}\{1, 10, 1, 100\}$ and $\boldsymbol{R} = \boldsymbol{I}_2$. We approximate the solution of the time-delay system using r partition of the delay in the interval $[-\tau, 0]$. We use $r=4$ partitions of the delay. As a result, we obtain an augmented state linear system of the following dimensions: $\boldsymbol{A}_r\in\mathbb{R}^{20\times20}$, $\boldsymbol{B}_r\in\mathbb{R}^{20\times2}$ and $\boldsymbol{Q}_r\in\mathbb{R}^{20\times20}$. We use a sampling time of $T=0.02s$ seconds. To verify the optimality of the approach, we compute the optimal solution using the ARE (10).

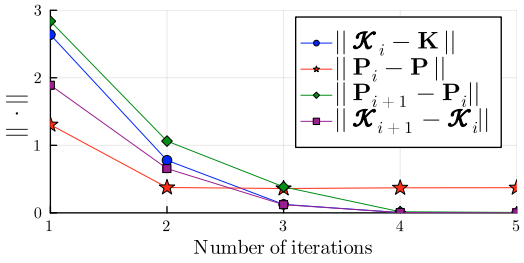


Fig. 1. Convergence results of the PI RL algorithm.

The initial conditions of the augmented system are set to $\boldsymbol{z}(0) = \boldsymbol{\chi}$, where $\boldsymbol{\chi} = [\chi_1, ..., \chi_N]^\top$ is a vector of random initial states obtained from a uniform distribution in the range of $[0, 0.25]$. To ensure parameter estimates convergence, we use a small probing noise of $e(t) = \Sigma_{i=1}^{30} a_i \sin(f_i t)$, where a_i and f_i are randomly sampled from a uniform distribution in the range of $[-0.375, 0.375]$ and $[-3\pi, 3\pi]$, respectively. We use a threshold of $\epsilon = 0.01$ to stop the algorithm when the spectral error norm is equal or less than the proposed value, i.e., $\|\boldsymbol{\mathcal{K}}_{i+1} - \boldsymbol{\mathcal{K}}_i\| \leq \epsilon$.

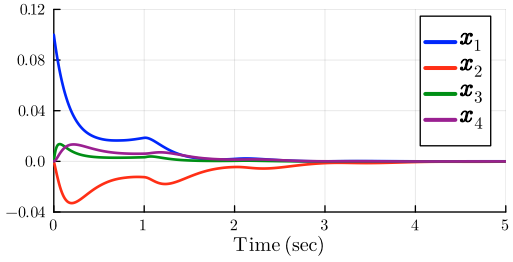


Fig. 2. Evolution of the system states under the near-optimal control policy.

Fig. 1 shows the learning results of the proposes approach. Here, both $\boldsymbol{P}_i$ and $\boldsymbol{\mathcal{K}}_i$ converge to estimated matrices that satisfy the pre-defined threshold ε which are close to the real near-optimal values $\boldsymbol{P}$ and $\boldsymbol{K}$, respectively. Here, when the stopping criterion is met, we obtain the following spectral norm errors compared with the near optimal solution provided by the ARE: $\|\boldsymbol{P}_5 - \boldsymbol{P}\| = 0.3723$ and $\|\boldsymbol{\mathcal{K}}_5 - \boldsymbol{K}\| = 3.1654e^{-5}$; these results demonstrate that the RL algorithm is capable to effectively find the near optimal solution for this augmented system. Recall that both $\boldsymbol{P}$ and $\boldsymbol{K}$ are sparse matrices, such that small error deviations in each term of the matrices produces a large spectral norm error. Notice that the algorithm finds a fast solution in few number of iterations which means that the probing noise is adequate even for this augmented system. Fig. 2 shows the temporal evolution of the original system states using the near-optimal control policy derived from the algorithm. The results of Fig. 2 shows the effectiveness of the approach to stabilize the time-delay system.

To assess the validity of the proposed approach, we compare the cost function (2) using three different methods: the PI RL-based method, a model-based solution from the ARE with approximate known matrices $\boldsymbol{A}_r$ and $\boldsymbol{B}_r$, and

a Julia optimizer initialized with values near the optimal gain control.

The results are summarized in Table 1, showing the cost function J and error norm between each method and a reference method. The PI RL approach and model-based method demonstrate exceptional agreement with negligible differences in J, validating the reinforcement learning approach's accuracy. The Julia optimizer method shows a slight deviation in J. While this error is more noticeable, the similarity suggests it remains a valid tool, particularly with appropriate optimization seeds. However, results indicate the PI RL and model-based methods provide consistent and accurate J values.

Table 1. Comparison of the cost and error function value between methods.

	J	$\|\|\varepsilon\|\|$
PI RL	0.04802769	—
Model based	0.04802768	9.9999e-9
Julia optimizer	0.04752630	0.501389e-3

5.2 Regenerative Chatter in Metal Cutting

To further verify the effectiveness and generalization of the approach we consider a regenerative chatter system in metal cutting applications Cui et al. (2023a) that has the following time-delay dynamics

$$\boldsymbol{A}_0 = \begin{bmatrix} 0 & 1 \\ -(k + \frac{F_t}{m}) & -\frac{c}{m} \end{bmatrix}, \boldsymbol{A}_1 = \begin{bmatrix} 0 & 0 \\ \frac{F_t}{m} & 0 \end{bmatrix}, \boldsymbol{B} = \begin{bmatrix} 0 \\ \frac{1}{m} \end{bmatrix},$$

where $\tau = 1.3$, $m = 2$, $c = 0.2$, $k = 10$ and $F_t = 1$. The weight matrices of the cost index (2) are set to $\boldsymbol{Q} = 100\boldsymbol{I}_2$ and $\boldsymbol{R} = 1$. We use $r = 2$ partitions to discretize the delay. As a result, we obtain a linear system of the following dimensions $\boldsymbol{A}_r \in \mathbb{R}^{6\times6}$, $\boldsymbol{B}_r \in \mathbb{R}^6$ and $\boldsymbol{Q}_r \in \mathbb{R}^{6\times6}$. For this particular system, we use a sampling time of $T = 0.008$s.

We follow a similar procedure to the chemical plant, that is, we set the initial conditions as a random vector of a uniform distribution in the range of $[-10, 10]$. The probing noise is modified into $e(t) = \sum_{i=1}^{28} a_i \sin(f_i t)$, where a_i and f_i are randomly sampled from a uniform distribution in the range of $[0.01, 0.05]$ and $[-8\pi, 8\pi]$. We use the same stopping threshold for this system.

Fig. 3 shows the learning results of the RL algorithm in the regenerative chatter system. Similar performances are obtained, that is, convergence to the near optimal values and satisfying the stopping threshold. We obtain the following spectral norm errors compared with the near optimal solution provided by the ARE: $\|\boldsymbol{P}_4 - \boldsymbol{P}\| = 0.0371$ and $\|\boldsymbol{\mathcal{K}}_4 - \boldsymbol{K}\| = 0.0002$; these results verify the effectiveness of the proposed approach. Fig. 4 depicts the temporal evolution of the time-delay system under the obtained near-optimal control policy.

6. CONCLUSIONS

This paper reports a near-optimal control approach for continuous-time linear systems with state delay using RL. The approach uses a finite difference technique to discretize the delay and transform the time-delay system into an augmented state linear system whose optimal solution

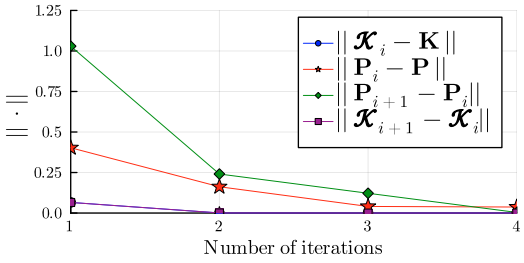


Fig. 3. Convergence results of the near optimal control problem using the proposed PI RL algorithm

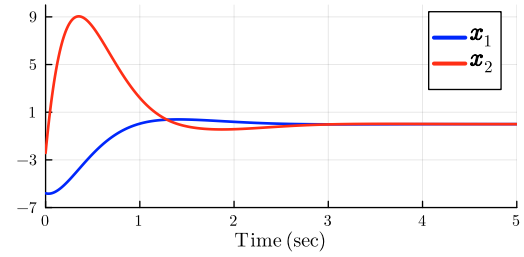


Fig. 4. Evolution of the system states with near-optimal control.

can be easily obtained using any value iteration or policy iteration algorithm. In this paper, we use a policy iteration algorithm to solve the near optimal control problem of augmented state systems. The algorithm uses a batch-least squares rule to estimate the parameters of the optimal control problem in each iteration of the RL algorithm. The results obtained from the simulation studies demonstrate the effectiveness of the proposed methodology without facing issues due to the infinite-dimensionality nature of time-delay systems. Future work will consider how optimal partitions of the delay can be obtained online without affecting closed-loop stability of the time-delay system. In addition, we will study how to reduce the dimensionality of continuous augmented state linear systems to obtain a low-dimensional representation that preserves its dynamic properties as well as preserving the properties of the real time-delay system.

REFERENCES

An, X., Wu, C., Lin, Y., Lin, M., Yoshinaga, T., and Ji, Y. (2023). Multi-robot systems and cooperative object transport: Communications, platforms, and challenges. *IEEE Open Journal of the Computer Society*, 4, 23–36. doi:10.1109/OJCS.2023.3238324.

Carlucho, I., Stephens, D., Ard, W., and Barbalata, C. (2023). Semi-parametric control architecture for autonomous underwater vehicles subject to time delays. *IEEE Access*, 11, 71287–71300. doi: 10.1109/ACCESS.2023.3293430.

Carter, G.C. (1993). *Coherence and Time Delay Estimation: An Applied Tutorial for Research, Development, Test and Evolution Engineers*. IEEE Press.

Cui, L., Pang, B., and Jiang, Z.P. (2023a). Learning-based adaptive optimal control of linear time-delay systems: A policy iteration approach. *IEEE Transactions on Automatic Control*, 69(1), 629–636.

Cui, L., Pang, B., and Jiang, Z.P. (2023b). A value iteration approach to adaptive optimal control of linear

time-delay systems. *IFAC-PapersOnLine*, 56(2), 6964–6969. doi:https://doi.org/10.1016/j.ifacol.2023.10.524. 22nd IFAC World Congress.

Curtain, R.F. and Zwart, H. (2012). *An introduction to infinite-dimensional linear systems theory*, volume 21. Springer Science & Business Media.

El Debeiki, M., Al-Rubaye, S., Perrusquía, A., Conrad, C., and Flores-Campos, J.A. (2024). An advanced path planning and uav relay system: Enhancing connectivity in rural environments. *Future Internet*, 16(3), 89.

Eller, D., Aggarwal, J., and Banks, H. (1969). Optimal control of linear time-delay systems. *IEEE Transactions on Automatic Control*, 14(6), 678–687. doi:10.1109/TAC.1969.1099301.

Fridman, E. (2014). *Introduction to Time-Delay Systems: Analysis and Control*. Systems & Control: Foundations & Applications. Springer International Publishing, Cham. doi:10.1007/978-3-319-09393-2.

Ganie, I. and Jagannathan, S. (2024). Lifelong reinforcement learning tracking control of nonlinear strict-feedback systems using multilayer neural networks with constraints. *Neurocomputing*, 600, 128139.

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

Gu, K., Chen, J., and Kharitonov, V.L. (2003). *Stability of time-delay systems*. Springer Science & Business Media.

Guo, W., Wei, Z., Gonzalez, O., Perrusquía, A., and Tsourdos, A. (2023). Control layer security: A new security paradigm for cooperative autonomous systems. *IEEE Vehicular Technology Magazine*.

Hespanha, J.P. (2018). *Linear Systems Theory*. Princeton University Press, Princeton, second edition.

Hua, C., Zhang, L., and Guan, X. (2018). *Robust Control for Nonlinear Time-Delay Systems*. Springer Singapore, Singapore. doi:10.1007/978-981-10-5131-9.

Huang, M., Jiang, Z.P., and 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. doi:10.1109/TCYB.2020.3029077.

Kebria, P.M., Abdi, H., Dalvand, M.M., Khosravi, A., and Nahavandi, S. (2019). Control methods for internet-based teleoperation systems: A review. *IEEE Transactions on Human-Machine Systems*, 49(1), 32–46.

Kebria, P.M., Khosravi, A., Nahavandi, S., Shi, P., and Alizadehsani, R. (2020). Robust adaptive control scheme for teleoperation systems with delay and uncertainties. *IEEE Transactions on Cybernetics*, 50(7), 3243–3253. doi:10.1109/TCYB.2019.2891656.

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. doi:https://doi.org/10.1016/0021-8928(62)90101-6.

Kwon, W.H. and Park, P. (2019). *Stabilizing and Optimizing Control for Time-Delay Systems*. Communications and Control Engineering. Springer International Publishing, Cham.

LeVeque, R.J. (2007). *Finite difference methods for ordinary and partial differential equations: steady-state and time-dependent problems*. SIAM.

Lian, B., Xue, W., Lewis, F.L., and Davoudi, A. (2024). Inverse value iteration and q-learning: Algorithms, stability, and robustness. *IEEE Transactions on Neural Networks and Learning Systems*.

Liu, D., Xue, S., Zhao, B., Luo, B., and Wei, Q. (2020). Adaptive dynamic programming for control: A survey and recent advances. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 51(1), 142–160.

Moghadam, R. and Jagannathan, S. (2019). Approximate optimal adaptive control of partially unknown linear continuous-time systems with state delay. In *2019 IEEE 58th Conference on Decision and Control (CDC)*, 1985–1990. doi:10.1109/CDC40024.2019.9029845.

Palanisamy, M., Modares, H., Lewis, F.L., and Aurangzeb, M. (2014). Continuous-time q-learning for infinite-horizon discounted cost linear quadratic regulator problems. *IEEE transactions on cybernetics*, 45(2), 165–176.

Perrusquía, A. (2022). Solution of the linear quadratic regulator problem of black box linear systems using reinforcement learning. *Information Sciences*, 595, 364–377.

Perrusquía, A. and Guo, W. (2023). Drone's objective inference using policy error inverse reinforcement learning. *IEEE Transactions on Neural Networks and Learning Systems*.

Rizvi, S.A.A., Wei, Y., and Lin, Z. (2021). Reinforcement learning for optimal adaptive control of time delay systems. In *Handbook of Reinforcement Learning and Control*, 215–242. Springer.

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

Ross, D.W. and Flügge-Lotz, I. (1969). An Optimal Control Problem for Systems with Differential-Difference Equation Dynamics. *SIAM Journal on Control*, 7(4), 609–623. doi:10.1137/0307044.

Sereni, B., Galvão, R.K.H., Assunão, E., and Teixeira, M.C.M. (2023). An output-feedback design approach for robust stabilization of linear systems with uncertain time-delayed dynamics in sensors and actuators. *IEEE Access*, 11, 20769–20785. doi:10.1109/ACCESS.2023.3249482.

Sutton, R.S. and Barto, A.G. (2018). *Reinforcement Learning: An Introduction*. The MIT Press, second edition.

Vamvoudakis, K.G., Wan, Y., Lewis, F.L., and Cansever, D. (eds.) (2021). *Handbook of Reinforcement Learning and Control*, volume 325 of *Studies in Systems, Decision and Control*. Springer International Publishing, Cham. doi:10.1007/978-3-030-60990-0.

Wei, J., Wang, H., and Liu, F. (2022). *Iterative Learning Control for Nonlinear Time-Delay System*. Springer Nature Singapore, Singapore. doi:10.1007/978-981-19-6317-9.

Witrant, E., Fridman, E., Sename, O., and Dugard, L. (eds.) (2016). *Recent Results on Time-Delay Systems: Analysis and Control*, volume 5 of *Advances in Delays and Dynamics*. Springer International Publishing, Cham.

Yu, W. and Perrusquia, A. (2021). *Human-Robot Interaction Control Using Reinforcement Learning*. John Wiley & Sons.



https://dspace.org/

Near optimal reinforcement learning control of linear time-delay systems

Martínez, René; Márquez-Martínez, Luis A.; Perrusquía, Adolfo

2025-07-02
Elsevier
https://doi.org/10.1016/j.ifacol.2025.12.153
Martínez R, Márquez-Martínez LA, Perrusquía A. (2025) Near optimal reinforcement learning control of linear time-delay systems. IFAC-PapersOnLine, Volume 59, Issue 14, 2025, pp. 220-225, 15th IFAC Workshop on Adaptive and Learning Control Systems ALCOS 2025, 2-4 July 2025, Mexico City, Mexico

Downloaded from a DSpace repository. https://dspace.org/