Delay-Independent Dynamic Implementation of Optimal Control for State and Input Delayed Nonlinear Systems

Shanshan Jiao and Qinglai Wei

***Abstract*—This study argues a dynamic optimal control approach for nonlinear systems with state and input delays. First, a feedback system identical to the delayed system is specified based on parallel control, enabling the dynamic optimal control issue to be formulated in light of the L_2 gain condition. The virtual controller developed using the backstepping integral technique makes the dynamic control feasible, with its design highlighting inertial-like features, so that the response speed of controller can be altered. Then, the optimal control of the physical system is further expanded to that of the augmented system composed of it, accomplishing the control goal through virtual and actual interaction. In the implementation, the critic-actor structure facilitates the control challenge to be settled online. Additionally, the stability of the system and the convergence of neural network weights are discussed, and the research conclusions are supported by simulation examples.**

***Index Terms*—Dynamic optimal control, parallel control, robust control, delayed systems, critic-actor structure**

I. Introduction

Recently, the analysis and synthesis of delayed systems have become a research hotspot in the international control field, as time delays in signal transmission are common, and they can lead to the deterioration of the system performance and potentially destroy the stability of the system [1, 2].

As a result, the research findings are springing up, with the primary differences initially lying in the handling of state or input delay. The input delay, which was modeled as a transport partial differential equation for the linear input delayed system in Ref. [3], enables delay adaptation and addresses the robust stabilization issue. In Ref. [4], the transport partial differential equation was provided for the uncertain linear input delayed system, so that the time-varying trajectory tracking system was constructed for constant set-point regulation. Concerning the state delay, the nonlinear system with state-dependent state delay was investigated for the stability issue, determining the length of the state history information [5]. In Ref. [6], the feedback stabilization challenge of uncertain state delayed systems was settled by employing a novel Lyapunov-Krasovskii function. In Ref. [7], a tracking control approach was investigated on the basis of the dynamic gain design, and the Lyapunov-Krasovskii function allowed the system state to be bound and the output to track the reference signal. Following that, to be more comprehensive, studies incorporating both state and input delay effects on the object system are provided. In Ref. [8], the output knowledge was utilized to estimate the state information via a finite-dimensional observer, so that the global asymptotic stability of the interconnected nonlinear system subject to state and input delays was guaranteed. In Ref. [9], time-varying measurement and input delays were considered for output feedback control of nonlinear systems, with a state prediction given by high-gain observers to ensure the stability of the system.

Subsequently, advanced technologies and fresh requirements for control quality inspire the development of optimal control [10, 11], thereby exploring the optimal control of the delayed system remains an ongoing concern. In Ref. [12], the control rule was learned by the reinforcement learning value iteration algorithm to accomplish the optimal control of state-delayed nonlinear systems with known system functions. In Ref. [13], the time-varying state delay and time-invariant system parameters were optimized, and the developed approach removed constraints on the gradient-based optimization algorithm. The optimal load frequency control of input-delayed interconnected systems was investigated in Ref. [14], which was built on the sliding mode predictive control with observers. In Ref. [15], the delayed object system was drawn using a linear time-delay system, with the implementation of robust optimal tracking control based on the measured system data. In Ref. [16], an adaptive control algorithm was developed for the wheeled mobile robot with state and input delays, which was represented as an affine system so that Lyapunov-Krasovskii functions and a delay matrix function were established for optimal control.

However, as can be found in the existing results, the optimal control of time-delay systems is typically limited to only state delay or input delay [12–14], and the implementation often requires model transformation [15, 16] or even model identification that may influence the accuracy [17], implying that more efficient and sophisticated approaches are urgently needed for the optimal control of the delayed system.

As a promising platform, parallel control comes into view

Manuscript received: 8 September 2024; revised: 5 October 2024; accepted: 16 November 2024. *(Corresponding author: Shanshan Jiao.)*

Citation: S. Jiao and Q. Wei, Delay-independent dynamic implementation of optimal control for state and input delayed nonlinear systems, *IJICS*, 2024, 29(4), 147–153.

Shanshan Jiao is with Institute of Systems Engineering, Macau University of Science and Technology, Macao 999078, China (e-mail: 2202853ymi30001@student.must.edu.mo).

Qinglai Wei is with State Key Laboratory of Multimodal Artificial Intelligence Systems, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China, and also with School of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing 100049, China (e-mail: qinglai.wei@ia.ac.cn).

Digital Object Identifier 10.62678/IJICS202409.10131

as it constructs a dynamic feedback system that differs from traditional feedback control strategies. This feature allows practical control challenges to be expanded into the virtual space and settled in line with the "expansion and conquest" principle. It is undeniable that parallel control has drawn widespread application [18–20]. In Ref. [18], the introduction of parallel control not only realized the optimal consensus control of multi-agent systems, but also provided a new perspective for changing the response speed of the control input. In Refs. [19, 20], robust optimal control strategies for uncertain systems were developed via parallel control. The difference is that parallel control invoked in Ref. [19] enables an auxiliary variable related to uncertainty, which is generated from the relationship between the optimal control of both the uncertain and nominal systems, while Ref. [20] focuses on modifying the control gain of the nominal system for robust optimal parallel control and then proposes an effective triggering mechanism to save communication resources.

Upon investigation, it is found that the advantages of parallel control have not yet been popularized in the delayed system. Therefore, stimulated by the restrictions of optimal control on the characteristics of the object systems and the need for prior system knowledge, this study attempts to explore a dynamic optimal control strategy for delayed systems on the basis of the parallel control technique, which will overcome these dilemmas and is easy to implement.

The main contributions of this study can be summarized as follows:

(1) Treating the delay as the structural uncertainty of the control system, a unique performance index that independently addresses the state and input delays is established in light of the L_2 gain conditions for optimal control.

(2) A feedback system nearly identical to the physical system is designed, which is determined by the state and its delay, thus transforming the optimal control issue of the delayed system into a minimax issue for the augmented system constructed via parallel control and implementing the adaptive dynamic programming based optimal control online.

This paper is divided into five parts. The second section formulates the problem to be solved. The third section is concerned with the methodology used for this study and analyzes the stability. The fourth section presents the findings of the research. Finally, the fifth section concludes this study.

II. Problem Formulation

The optimal control issue of the delayed system is introduced in this section. The parallel control provides a foundation for the unique dynamic optimal control strategy.

The nonlinear system with delayed input and state is examined as follows

$$\dot{x}(t) = f_1(x(t)) + f_2(x_{d_x}) + g_1(x)u(t) + g_2(x)u_{d_u} \tag{1}$$

where $x(t) \in \mathbb{R}^n$ (with the initial value x_0) and $x_{d_x} \overset{\Delta}{=} x(t-d_x)$ denote the delay-free and delayed system states, $u(t) \in \mathbb{R}^m$ and $u_{d_u} \overset{\Delta}{=} u(t-d_u)$ are the non-delayed and delayed control inputs, and d_x and d_u are the delay signals related to the system state and control input, respectively. $g_1(x)$ and $g_2(x)$ are the input dynamics. $f_1(x)$ and $f_2(x_{d_x})$ are the system functions. $f_1(x) + f_2(x_{d_x}) + g_1 u + g_2 u_{d_u}$ is Lipschitz continuous on a compact set containing the origin. $d_x \in [0, d_{x,\max}] \subseteq \mathbb{N}$ and $d_u \in [0, d_{u,\max}] \subseteq \mathbb{N}$. $\mathbb{R}$ and $\mathbb{N}$ symbolize the real number set and natural number set, respectively.

The history is presented as

$$\begin{cases} x(\delta) = \eta_x(\delta),\ \delta \in [-d_x, 0]; \\ u(\delta) = \eta_u(\delta),\ \delta \in [-d_u, 0] \end{cases} \tag{2}$$

where η_x and η_u are history functions.

Assumption 1 [16] Suppose that the nonlinear system in Eq. (1) is controllable and the time delay is known.

For parallel control [18], a feedback system that is mathematically symmetric to the object system is usually established as

$$\dot{u}(t) = \kappa(x, x_{d_x}, u, u_{d_u}) \tag{3}$$

where $\kappa(\cdot)$ is the control function to be explored.

Moreover, to achieve infinite-time dynamic optimal control using the parallel control technique, $\dot{u}(t)$ should stabilize the system in Eq. (1) and minimize the following performance index function

$$J_0 = \int_0^\infty \left(x^{\mathrm{T}}(\tau)Qx(\tau) + u^{\mathrm{T}}(\tau)Ru(\tau)\right)\mathrm{d}\tau + \int_0^\infty \dot{u}^{\mathrm{T}}(\tau)P\dot{u}(\tau)\mathrm{d}\tau \tag{4}$$

where $Q \in \mathbb{R}^{n\times n}$, $R \in \mathbb{R}^{m\times m}$, and $P \in \mathbb{R}^{m\times m}$ denote constant matrices. It can be observed from Eq. (4) that the effect of the derivative of control input is considered.

However, the development of $\dot{u}(t)$ is beyond the capabilities of traditional feedback control strategies. Thus, according to the backstepping integral technique, the following virtual controller is designed

$$\omega(t) = \beta\dot{u}(t) + u(t) \tag{5}$$

where β is a constant.

Subsequently, the feedback system is formulated as

$$\dot{u}(t) = \frac{1}{\beta}(\omega(t) - u(t)) \tag{6}$$

Remark 1 It can be observed from Eqs. (5) and (6) that the designed virtual control input allows the feedback system to be viewed as an inertial element. Reducing the time constant β will increase the response speed of the control input, thereby making it possible to alter the response speed in light of the characteristics of the inertial element.

The establishment of the feedback system is conducive to the implementation of parallel control, which seeks to design an artificial system parallel to the physical system so as to achieve the control objective through the interaction between the two systems. Therefore, the definition of the augmented state vector $z(t) = \left[x^{\mathrm{T}}(t), u^{\mathrm{T}}(t)\right]^{\mathrm{T}}$ is invoked to obtain the following augmented system in light of Eqs. (1) and (6)

$$\dot{z}(t) = F(z) + D_{xu} + G(z)\omega(t) = \underbrace{\begin{bmatrix} f_1(x(t)) + g_1(x)u(t) \\ -\dfrac{1}{\beta}u(t) \end{bmatrix}}_{F(z)} + G(z)\omega(t) + \underbrace{\begin{bmatrix} f_2(x_{d_x}) + g_2(x)u_{d_u} \\ 0 \end{bmatrix}}_{D_{xu}} \tag{7}$$

where the initial state is specified as $z(0) \overset{\Delta}{=} z_0$, $G(z) \in \mathbb{R}^{(n+m)\times m}$, and $G(z) = \frac{1}{\beta}\begin{bmatrix} 0 & 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 1 \end{bmatrix}^{\mathrm{T}}$ provides the input dynamics.

In Eq. (7), the original delayed system in Eq. (1) appears as the internal dynamics, whose optimal control is transformed into that of the augmented system in Eq. (7), implying that the control law $\omega(t)$ needs to be designed to achieve the optimal control objective, i.e., the proposed $\omega(t)$ stabilizes the system in Eq. (7) while minimizing the following augmented cost function

$$\begin{aligned}\mathcal{J}(z_0,\omega) = &\int_0^\infty \left(x^{\mathrm{T}}(\tau)Qx(\tau)+u^{\mathrm{T}}(\tau)Ru(\tau)\right)\mathrm{d}\tau + \\ &\int_0^\infty \omega^{\mathrm{T}}(\tau)P\omega(\tau)\mathrm{d}\tau\end{aligned} \tag{8}$$

However, it is well known that the presence of the time delay will cause instability in the system, so the L_2 gain condition indicates that the cost function in Eq. (4) should obey

$$\mathcal{J}(z_0,\omega) \leqslant \int_{-d_x}^0 \eta_x^{\mathrm{T}}(t)M_1\eta_x(t)\mathrm{d}t + \int_{-d_u}^0 \eta_u^{\mathrm{T}}(t)M_2\eta_u(t)\mathrm{d}t \tag{9}$$

where M_1 and M_2 are $m\times m$-dimensional positive definite matrices.

Then, we can further define the following performance index for the optimal control of augmented system in Eq. (7)

$$\begin{aligned}&\mathcal{J}_\tau(z_0,\omega) = \\ &\int_0^\infty U_\tau(z_0,\omega)\mathrm{d}t = \\ &\int_0^\infty \left(x^{\mathrm{T}}(t)Qx(t)+u^{\mathrm{T}}(t)Ru(t)+\omega^{\mathrm{T}}(t)P\omega(t)\right)\mathrm{d}t - \\ &\int_{-d_x}^0 \eta_x^{\mathrm{T}}(t)M_1\eta_x(t)\mathrm{d}t - \int_{-d_u}^0 \eta_u^{\mathrm{T}}(t)M_2\eta_u(t)\mathrm{d}t\end{aligned} \tag{10}$$

where $U_\tau(z_0,\omega) = x^{\mathrm{T}}(t)\overline{Q}x(t) + u^{\mathrm{T}}(t)\overline{R}u(t) + \omega^{\mathrm{T}}(t)P\omega(t) - x_{d_x}^{\mathrm{T}}M_1x_{d_x} - u_{d_u}^{\mathrm{T}}M_2u_{d_u}$ denotes the utility function, where $\overline{Q} = Q+M_1$ and $\overline{R} = R+M_2$.

Remark 2 The cost function in Eq. (10) built on Eq. (9) demonstrates that the effects of state and input delays on the system are taken independently for dynamic optimal control, which makes the proposed method applicable to more scenarios. For example, dynamic optimal control of the state-delayed system can be implemented when $M_2 = 0$ is established, optimal control of the input-delayed system is attainable if $M_1 = 0$ is satisfied, and optimal control of the undelayed system can be implemented if $M_1 = M_2 = 0$ holds.

The above theoretical foundations lead us to recognize that the solution to the dynamic optimal control is equivalent to solving the following maximin issue [21, 22]

$$\mathcal{J}_\tau(z_0,\omega^*) = \min_{\omega}\max_{\substack{x_{d_x}\\ u_{d_u}}}\mathcal{J}_\tau(z_0,\omega) \tag{11}$$

where ω^* denotes the optimal value of ω.

Specifying the value function as Eq. (10), the Hamiltonian function can be formed as

$$\begin{aligned}&\mathcal{H}(z,\omega,\nabla\mathcal{J}_\tau) = \\ &\frac{\partial\mathcal{J}_\tau}{\partial t} + x^{\mathrm{T}}(t)\overline{Q}x(t) + u^{\mathrm{T}}(t)\overline{R}u(t) + \omega^{\mathrm{T}}(t)P\omega(t) - \\ &x_{d_x}^{\mathrm{T}}M_1x_{d_x} - u_{d_u}^{\mathrm{T}}M_2u_{d_u} = \\ &\nabla\mathcal{J}_\tau^{\mathrm{T}}(z)(F(z)+G(z)\omega+D_{xu}) + \omega^{\mathrm{T}}(t)P\omega(t) + \\ &x^{\mathrm{T}}(t)\overline{Q}x(t)+u^{\mathrm{T}}(t)\overline{R}u(t) - x_{d_x}^{\mathrm{T}}M_1x_{d_x} - u_{d_u}^{\mathrm{T}}M_2u_{d_u}\end{aligned} \tag{12}$$

where $\partial(\cdot)/\partial t$ represents the gradient of $(\cdot)$.

Referring to the optimal principle, the virtual control law $\omega(t)$ should be developed so that $\min_{\omega}\{\mathcal{H}(z,\omega,\nabla\mathcal{J}_\tau^*)\} = 0$ is fulfilled. The optimal control rule can be further deduced as

$$\omega^*(z) = -\frac{1}{2}P^{-1}G^{\mathrm{T}}(z)\nabla\mathcal{J}_\tau^* \tag{13}$$

Thus, the Hamilton-Jacobi-Bellman (HJB) equation can be described as

$$\begin{aligned}&\nabla\mathcal{J}_\tau^{\mathrm{T}}(z)(F(z)+G(z)\omega^*+D_{xu}) + \omega^{*\mathrm{T}}(t)P\omega^*(t) + \\ &x^{\mathrm{T}}(t)\overline{Q}x(t)+u^{\mathrm{T}}(t)\overline{R}u(t) - x_{d_x}^{\mathrm{T}}M_1x_{d_x} - u_{d_u}^{\mathrm{T}}M_2u_{d_u} = 0\end{aligned} \tag{14}$$

Inevitably, Eq. (14) demands to be settled for optimal control. Thus, the critic-actor structure will be invoked to implement the proposed strategy in the next section.

III. Implementation of Parallel Control

In this section, the critic-actor structure is constructed based on the neural network to approximate the cost function and optimal control input of the augmented system in Eq. (7). The gradient descent algorithm is utilized to tune the neural networks, and the integral technique allows the actual control input of the delayed system in Eq. (1) to be obtained. Finally, the stability of the system and the convergence of the neural network weights are confirmed.

A. Neural Network Based Optimal Control

First, the critic neural network constructs the performance index as

$$\begin{aligned}\mathcal{J}_\tau(z_0,\omega) = &W_{\mathrm{c}}^{\mathrm{T}}\phi_{\mathrm{c}}(z) - \int_{-d_x}^0 \eta_x^{\mathrm{T}}(t)M_1\eta_x(t)\mathrm{d}t - \\ &\int_{-d_u}^0 \eta_u^{\mathrm{T}}(t)M_2\eta_u(t)\mathrm{d}t + \varepsilon_{\mathrm{c}}(z)\end{aligned} \tag{15}$$

where $W_{\mathrm{c}} = [W_{\mathrm{c}1}, W_{\mathrm{c}2}, \ldots, W_{\mathrm{c}l}]^{\mathrm{T}} \in \mathbb{R}^l$ and $\phi_{\mathrm{c}}(z)$ represent the ideal weight vector and activation function of the critic neural network, $\varepsilon_{\mathrm{c}}(z)$ gives the approximation error.

Referring to the output of the neural network, Eq. (15) is reformulated as

$$\begin{aligned}\hat{\mathcal{J}}_\tau(z_0,\omega) = &\hat{W}_{\mathrm{c}}^{\mathrm{T}}\phi_{\mathrm{c}}(z) - \int_{-d_x}^0 \eta_x^{\mathrm{T}}(t)M_1\eta_x(t)\mathrm{d}t - \\ &\int_{-d_u}^0 \eta_u^{\mathrm{T}}(t)M_2\eta_u(t)\mathrm{d}t\end{aligned} \tag{16}$$

where the approximate $\hat{W}_{\mathrm{c}}$ allows the estimation error to be defined as $\tilde{W}_{\mathrm{c}} = W_{\mathrm{c}} - \hat{W}_{\mathrm{c}}$.

As far as Eqs. (15) and (16) are concerned, the approximated HJB equations can be established as

$$\begin{aligned}\mathcal{H}(z,\omega,W_{\mathrm{c}}) = &W_{\mathrm{c}}^{\mathrm{T}}\nabla\phi_{\mathrm{c}}(z)(F(z)+G(z)\omega+D_{xu}) + \\ &\omega^{\mathrm{T}}(t)P\omega(t) + x^{\mathrm{T}}(t)\overline{Q}x(t) + u^{\mathrm{T}}(t)\overline{R}u(t) - \\ &x_{d_x}^{\mathrm{T}}M_1x_{d_x} - u_{d_u}^{\mathrm{T}}M_2u_{d_u} - \varepsilon_{\mathrm{H}}\end{aligned} \tag{17}$$

$$\begin{aligned}\mathcal{H}\left(z,\omega,\hat{W}_{\rm c}\right) = {} & \hat{W}_{\rm c}^{\rm T}\nabla\phi_{\rm c}(z)\left(F(z)+G(z)\omega+D_{xu}\right)+\\ & \omega^{\rm T}(t)P\omega(t)+x^{\rm T}(t)\overline{Q}x(t)+u^{\rm T}(t)\overline{R}u(t)-\\ & x_{d_x}^{\rm T}M_1x_{d_x}-u_{d_u}^{\rm T}M_2u_{d_u}\end{aligned} \tag{18}$$

where $\varepsilon_{\rm H} = -\nabla\varepsilon_{\rm c}(z)\left(F(z)+G(z)\omega+D_{xu}\right)$.

It follows from $\tilde{W}_{\rm c}$ that

$$\begin{aligned}\sigma_{\rm c} = {} & \mathcal{H}(z,\omega,W_{\rm c})-\mathcal{H}\left(z,\omega,\hat{W}_{\rm c}\right) =\\ & W_{\rm c}^{\rm T}\nabla\phi_{\rm c}(z)\dot{z}-\hat{W}_{\rm c}^{\rm T}\nabla\phi_{\rm c}(z)\dot{z}-\varepsilon_{\rm H} =\\ & \tilde{W}_{\rm c}^{\rm T}\nabla\phi_{\rm c}(z)\dot{z}-\varepsilon_{\rm H}\end{aligned} \tag{19}$$

Therefore, the following residual function can be specified to implement the gradient descent algorithm

$$\Xi_{\rm c} = \frac{1}{2}\sigma_{\rm c}^{\rm T}\sigma_{\rm c} \tag{20}$$

Furthermore, the critic neural network is tuned by the following rule

$$\dot{\hat{W}}_{\rm c} = -\rho_{\rm c}\frac{\partial\Xi_{\rm c}}{\partial\hat{W}_{\rm c}} = -\rho_{\rm c}\frac{\xi\left(\xi^{\rm T}\hat{W}_{\rm c}+U_\tau(z_0,\omega)\right)}{(\xi^{\rm T}\xi+1)^2} \tag{21}$$

where $\rho_{\rm c} > 0$ denotes the learning rate, and $\xi = \nabla\phi_{\rm c}(z)\dot{z}$.

Next, the actor neural network with the ideal weight $W_{\rm a}$ and the approximation error $\varepsilon_{\rm a}(z)$ forms the optimal control input of the augmented system in Eq. (7) as

$$\omega(z) = W_{\rm a}^{\rm T}\phi_{\rm a}(z)+\varepsilon_{\rm a}(z) \tag{22}$$

where $\phi_{\rm a}$ represents the activation function of the actor neural network.

The output of the actor neural network yields the approximated controller as

$$\hat{\omega}(z) = \hat{W}_{\rm a}^{\rm T}\phi_{\rm a}(z) \tag{23}$$

Similarly, the definition of the weight error $\tilde{W}_{\rm a} = W_{\rm c}-\hat{W}_{\rm a}$ activates the gradient descent method such that the tuning rule of the actor neural network is calculated as

$$\dot{\hat{W}}_{\rm a} = -\rho_{\rm a}\phi_{\rm a}(z)\left(\hat{W}_{\rm a}^{\rm T}\phi_{\rm a}(z)+\frac{1}{2}P^{-1}G^{\rm T}(z)\nabla\phi_{\rm c}^{\rm T}(z)\hat{W}_{\rm c}\right)^{\rm T} \tag{24}$$

where $\rho_{\rm a}$ represents the adaptive update rate.

The employment of the actor neural network enables the control input of the delayed system in Eq. (1) to be formulated using the integral technique, which is given as

$$\begin{aligned}\hat{u}(t) = {} & {\rm e}^{-t/\beta}\left(\hat{u}(x_0)+\frac{1}{\beta}\int_0^t\hat{\omega}(\tau){\rm e}^{\tau/\beta}{\rm d}\tau\right) =\\ & {\rm e}^{-t/\beta}\left(\hat{u}(x_0)+\frac{1}{\beta}\int_0^t\hat{W}_{\rm a}^{\rm T}\phi_{\rm a}(\tau){\rm e}^{\tau/\beta}{\rm d}\tau\right)\end{aligned} \tag{25}$$

The invocation of neural networks makes it necessary to analyze the convergence of weights, therefore theoretical analysis will be given.

B. Stability Analysis

This part shows that the proposed method enables the system state to be uniformly ultimately bounded (UUB) [23], and the weights of the critic and actor neural networks are also UUB if the persistent excitation (PE) condition is met [24]. To facilitate verification, the following assumptions are provided.

Assumption 2

(1) The estimation errors of the neural networks are bounded.

(2) The characteristics of the neural networks enable that $\left\|W_{\rm a}^{\rm T}\phi_{\rm a}(z)+P^{-1}G^{\rm T}(z)\nabla\phi_{\rm c}^{\rm T}W_{\rm c}/2\right\| \leqslant \varepsilon_{\rm J}$ and $\nabla\varepsilon_{\rm c}^{\rm T}(z)\left(F(z)+G(z)\omega+D_{xu}\right) \leqslant -\bar{\varepsilon}\|z\|^2$. Furthermore, $W_{\rm c}^{\rm T}\nabla\phi_{\rm c}(z)\left(F(z)+G(z)\omega+D_{xu}\right) \leqslant -\bar{k}\|z\|^2$, $\left\|W_{\rm c}^{\rm T}\nabla\phi_{\rm c}(z)G(z)\nabla\phi_{\rm a}^{\rm T}(z)\right\| \leqslant \bar{k}_1$, and $\left\|\nabla\varepsilon_{\rm c}^{\rm T}(z)G(z)\nabla\phi_{\rm a}^{\rm T}(z)\right\| \leqslant \bar{k}_2$, where $\varepsilon_{\rm J}$, $\bar{k}$, $\bar{\varepsilon}$, $\bar{k}_1$, and $\bar{k}_2$ are constants.

For the optimal control of system in Eq. (7), it can be concluded that the system state and weights of neural networks are UUB if the assumptions hold, $\xi/\left(\xi^{\rm T}\xi+1\right)^2$ fulfills the PE condition, and the performance index in Eq. (10) and the control law in Eq. (13) are approximated by the critic and actor neural networks tuned by the rules in Eqs. (21) and (24), respectively.

To verify this conclusion referring to the Lyapunov theory, the following Lyapunov function is defined on the basis of Eq. (10)

$$\mathcal{L} = \mathcal{J}_\tau(z)+L_1+L_2 \tag{26}$$

where $L_1 = {\rm tr}\left(\tilde{W}_{\rm c}^{\rm T}\rho_{\rm c}^{-1}\tilde{W}_{\rm c}\right)/2$ and $L_2 = {\rm tr}\left(\tilde{W}_{\rm a}^{\rm T}\rho_{\rm a}^{-1}\tilde{W}_{\rm a}\right)/2$.

Taking the time derivative of each term in Eq. (26), we have

$$\begin{aligned}&\dot{\mathcal{J}}_\tau(z) =\\ &W_{\rm c}^{\rm T}\nabla\phi_{\rm c}(z)\left(F(z)+G(z)\nabla\phi_{\rm a}^{\rm T}(z)\hat{W}_{\rm a}+D_{xu}\right)-x_{d_x}^{\rm T}M_1x_{d_x}+\\ &\quad\nabla\varepsilon_{\rm c}(z)\left(F(z)+G(z)\nabla\phi_{\rm a}^{\rm T}(z)\hat{W}_{\rm a}+D_{xu}\right)-u_{d_u}^{\rm T}M_2u_{d_u} \leqslant\\ &-x_{d_x}^{\rm T}M_1x_{d_x}-u_{d_u}^{\rm T}M_2u_{d_u}-\nabla\varepsilon_{\rm c}(z)G(z)\nabla\phi_{\rm a}^{\rm T}(z)\tilde{W}_{\rm a}-\\ &\quad W_{\rm c}^{\rm T}\nabla\phi_{\rm c}(z)G(z)\nabla\phi_{\rm a}^{\rm T}(z)\tilde{W}_{\rm a}-\left(\bar{k}+\bar{\varepsilon}\right)\|z\|^2 \leqslant\\ &-\left(\bar{k}+\bar{\varepsilon}\right)\|z\|^2-\left(\bar{k}_1+\bar{k}_2\right)\left\|\tilde{W}_{\rm a}\right\|-x_{d_x}^{\rm T}M_1x_{d_x}-u_{d_u}^{\rm T}M_2u_{d_u}\end{aligned} \tag{27}$$

$$\begin{aligned}\dot{L}_1 = {} & \left[\tilde{W}_{\rm c}^{\rm T}\frac{\xi\left(-\tilde{W}_{\rm c}^{\rm T}\nabla\phi_{\rm c}(z)\dot{z}+\varepsilon_{\rm H}\right)}{(\xi^{\rm T}\xi+1)^2}\right] \leqslant\\ & -\left\|\frac{\xi^{\rm T}}{(\xi^{\rm T}\xi+1)}\tilde{W}_{\rm c}\right\|^2+\frac{1}{2}\left\|\frac{\xi^{\rm T}}{(\xi^{\rm T}\xi+1)}\tilde{W}_{\rm c}\right\|^2+\frac{1}{2}\left\|\frac{\varepsilon_{\rm H}}{(\xi^{\rm T}\xi+1)}\right\|^2\end{aligned} \tag{28}$$

$$\begin{aligned}\dot{L}_2 = {} & {\rm tr}\left[\tilde{W}_{\rm a}^{\rm T}\phi_{\rm a}\left(\hat{W}_{\rm a}^{\rm T}\phi_{\rm a}+\frac{1}{2}P^{-1}G^{\rm T}(z)\nabla\phi_{\rm c}^{\rm T}\hat{W}_{\rm c}\right)\right] =\\ & -\tilde{W}_{\rm a}^{\rm T}\phi_{\rm a}\tilde{W}_{\rm a}^{\rm T}\phi_{\rm a}-\frac{1}{2}\tilde{W}_{\rm a}^{\rm T}\phi_{\rm a}P^{-1}G^{\rm T}(z)\nabla\phi_{\rm c}^{\rm T}\tilde{W}_{\rm c}+\tilde{W}_{\rm a}^{\rm T}\phi_{\rm a}\varepsilon_{\rm J} \leqslant\\ & -\|\phi_{\rm a}\|^2\left\|\tilde{W}_{\rm a}\right\|^2+\frac{1}{4}\left\|\tilde{W}_{\rm a}\right\|^2\|\phi_{\rm a}\|^2+\frac{1}{2}\left\|\tilde{W}_{\rm a}\right\|^2\|\phi_{\rm a}\|^2+\frac{1}{2}\varepsilon_{\rm J}^2+\\ & \frac{1}{4}\left\|P^{-1}G^{\rm T}(z)\nabla\phi_{\rm c}^{\rm T}\right\|^2\left\|\tilde{W}_{\rm c}\right\|^2 =\\ & -\frac{1}{4}\|\phi_{\rm a}\|^2\left\|\tilde{W}_{\rm a}\right\|^2+\frac{1}{2}\varepsilon_{\rm J}^2+\frac{1}{4}\left\|P^{-1}G^{\rm T}(z)\nabla\phi_{\rm c}^{\rm T}\right\|^2\left\|\tilde{W}_{\rm c}\right\|^2\end{aligned} \tag{29}$$

Following Eqs. (27)–(29), we can get

$$\begin{aligned}\dot{\mathcal{L}} = {} & -\left(\bar{k}+\bar{\varepsilon}\right)\|z\|^2-x_{d_x}^{\rm T}M_1x_{d_x}-u_{d_u}^{\rm T}M_2u_{d_u}+\frac{1}{2}\left\|\frac{\varepsilon_{\rm H}}{\psi}\right\|^2+\\ & \lambda_\varepsilon-\frac{1}{2}\left\|\frac{\xi^{\rm T}}{\psi}\tilde{W}_{\rm c}\right\|^2-\lambda_{\rm a}\left[\left(\left\|\tilde{W}_{\rm a}\right\|+\frac{\bar{k}_1+\bar{k}_2}{2\lambda_{\rm a}}\right)^2-\left(\frac{\bar{k}_1+\bar{k}_2}{2\lambda_{\rm a}}\right)^2\right]\end{aligned} \tag{30}$$

where $\psi = \xi^{\mathrm{T}}\xi + 1$, $\lambda_{\mathrm{a}} = \|\phi_{\mathrm{a}}\|^2/4$, and $\lambda_{\varepsilon} = \varepsilon_{\mathrm{J}}^2/2 + \left\|P^{-1}G^{\mathrm{T}}(z)\nabla\phi_{\mathrm{c}}^{\mathrm{T}}\right\|^2\left\|\tilde{W}_{\mathrm{c}}\right\|^2/4$.

If the following conditions are met, $\dot{\mathcal{L}} < 0$ is satisfied, which implies that the system stability is guaranteed

$$\begin{cases} \left\|\tilde{W}_{\mathrm{c}}\right\| > \sqrt{2\lambda_{\mathrm{c}}}\left\|\dfrac{\psi}{\xi}\right\|, \\ \left\|\tilde{W}_{\mathrm{a}}\right\| > -\dfrac{\bar{k}_1 + \bar{k}_2}{\lambda_{\mathrm{a}}}, \\ \left\|x_{d_x}\right\| > \sqrt{\dfrac{\lambda_{\mathrm{c}}}{M_1}}, \\ \left\|u_{d_u}\right\| > \sqrt{\dfrac{\lambda_{\mathrm{c}}}{M_2}}, \\ \|z\| > \sqrt{\dfrac{\lambda_{\mathrm{c}}}{\left(\bar{k} + \bar{\varepsilon}\right)}} \end{cases} \tag{31}$$

where $\lambda_{\mathrm{c}} = \lambda_{\varepsilon} + \|\varepsilon_{\mathrm{H}}/\psi\|^2/2 + \left(\bar{k}_1 + \bar{k}_2\right)^2/(4\lambda_{\mathrm{a}})$.

Referring to Lemma 1 in Ref. [18], the uniformly ultimately boundedness of the augmented system in Eq. (7) also illustrates that the original delayed system in Eq. (1) is UUB.

IV. Simulation Result

In this section, a nonlinear system with state and control input delays is employed to verify the developed dynamic optimal control approach.

The controlled system is formulated as

$$\begin{cases} \dot{x}_1(t) = 0.1x_1(t) + x_2(t) - \dfrac{x_1(t-d_x)}{\sqrt{1+x_1^2(t-d_x)}} - \dfrac{1}{\sqrt{1+x_1^2(t-d_x)}}u(t), \\ \dot{x}_2(t) = -x_2(t) + \dfrac{x_2(t-d_x)}{\sqrt{1+x_2^2(t-d_x)}} + \dfrac{1}{\sqrt{1+x_1^2(t-d_x)}}u(t-d_u) \end{cases} \tag{32}$$

where $d_x = 0.1$ and $d_u = 0.5$ are provided, with $x(\delta) = 1$, $\delta \in [-d_x, 0]$, $u(\delta) = 0$, and $\delta \in [-d_u, 0]$.

For optimal control, $Q = 0.5I_2$ (I_2 denotes the two-dimensional unit matrix), $R = 0.5$, $P = 2.5$, $M_1 = 0.05I_2$, and $M_2 = 0.05$ are specified to construct the performance index function. The parameter β is set as 1 and 5 to construct the virtual controller, thus generating the dynamic feedback system and further deriving the actual control input of the system in Eq. (32) on the basis of the integral technique.

Then, the following polynomial activation function of the critic neural network is selected to implement the proposed scheme online, with the initial weight $W_{\mathrm{c}}(0) = 0.25I_9$ (I_9 denotes the nine-dimensional unit matrix).

$$\begin{aligned} \phi_{\mathrm{c}}(z) = [&z_1^2(t), z_1(t)z_2(t), z_1(t)z_3(t), z_2(t)z_3(t), \\ &z_2^2(t), z_3^2(t), z_1^3(t), z_2^3(t), z_3^3(t)]^{\mathrm{T}} \end{aligned} \tag{33}$$

Similarly, the activation function of the actor neural network is as

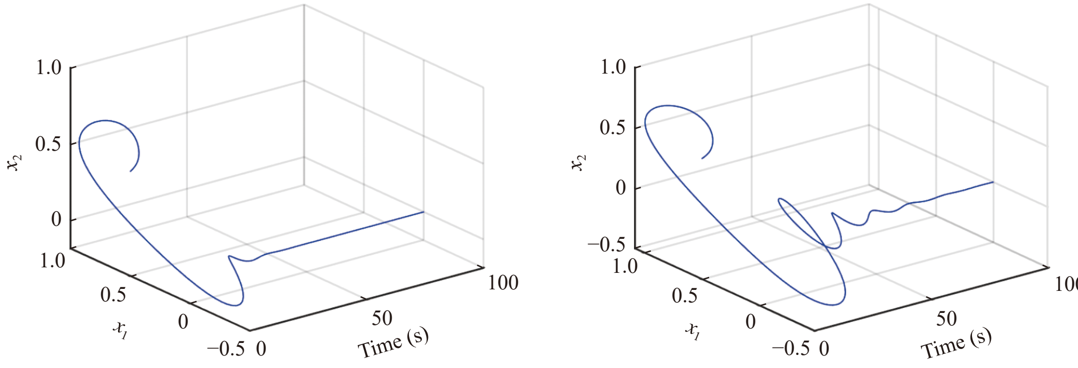
$$\begin{aligned} \phi_{\mathrm{a}}(z) = [&z_1(t), z_2(t), z_3(t), z_1^2(t), z_2^2(t), \\ &z_3^2(t), z_1^3(t), z_2^3(t), z_3^3(t)]^{\mathrm{T}} \end{aligned} \tag{34}$$

where the initial weight vector is $W_{\mathrm{a}}(0) = 0.1I_9$.

The learning rates are chosen as $\rho_{\mathrm{c}} = 1.5$ and $\rho_{\mathrm{a}} = 5.0$. The experimental results are shown in Figs. 1–3, which consist of the system state, the actual control input $u(t)$ generated from the virtual controller $\omega(t)$, and the cost function for optimal control, when β is set to 1 and 5. Figures 1 and 2 indicate that the system states stabilize as the control input converges. The comparison of the two subfigures in Fig. 1 reveals that the time taken for the system state to reach a stable state increases as β increases, which confirms the effect of β on the response speed, as Remark 1 shows. Figure 3 demonstrates the trajectory of the cost function, and it is clear that a larger β will also lead to an increase in the cost, and its convergence validates the method developed in this study. Therefore, the conclusions in the article are all substantiated.

V. Conclusion

The purpose of the current study is to determine a dynamic optimal control strategy for nonlinear delayed systems. The most significant finding to emerge from this study is that the dynamic feedback system is built on parallel control, which allows the augmented system that takes an inertial-like controller as the input to be developed, thus overcoming the optimal control challenges restricted by the system knowledge based on system data. The possibility that the response speed



(a) $\beta = 1$ (b) $\beta = 5$

Figure 1 State trajectory.

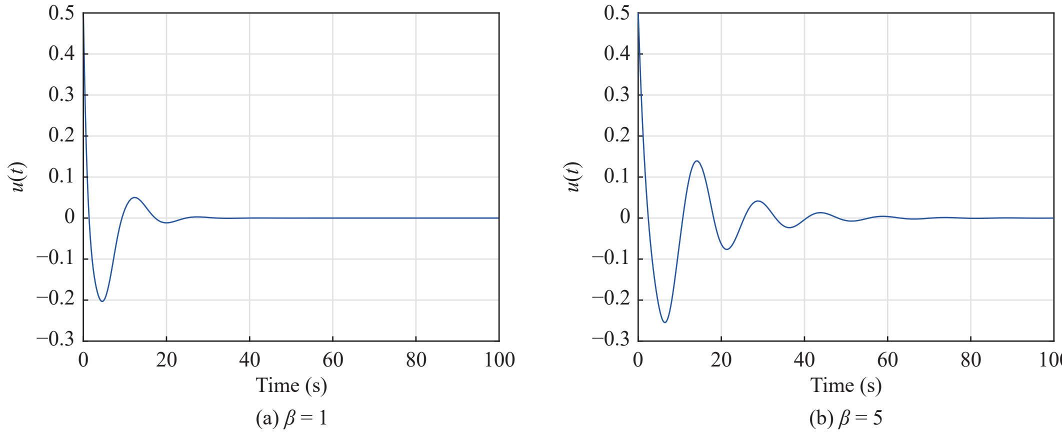


Figure 2 Actual control input.

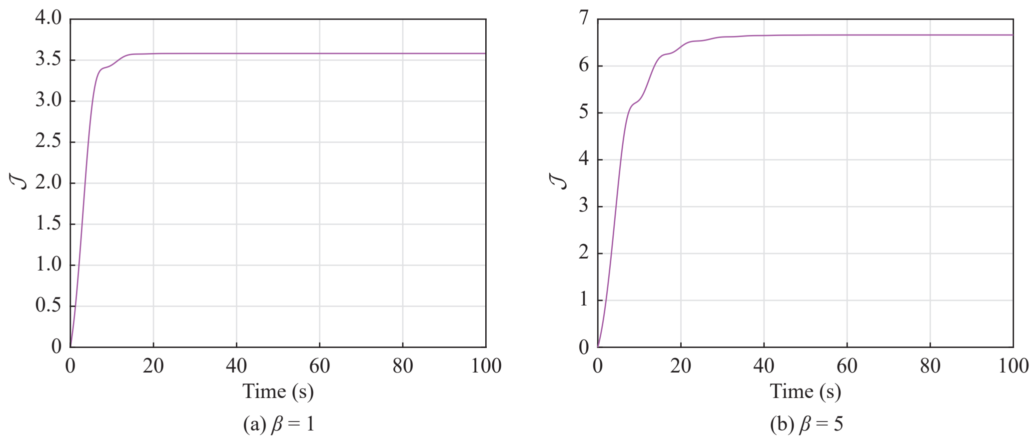


Figure 3 Cost function.

of controller may be adjusted easily deserves attention. In addition, referring to the L_2 gain condition for independently addressing the undesirable effects of state and input delay on the system facilitates the application of the proposed method. The final experimental results validate the contributions of this study, laying the groundwork for future research on the control issues of the delayed system. Based on this, future research can be undertaken to explore more complex optimal control problems for time-delay systems, such as nonlinear systems with higher dimensionality.

Acknowledgment

This work was supported by the National Key R&D Program of China (Nos. 2024YFB4709100 and 2021YFE 0206100), the National Natural Science Foundation of China (Nos. 62425310 and 62073321), the National Defense Basic Scientific Research Program (Nos. JCKY2019203C029 and JCKY2020130C025), and the Science and Technology Development Fund, Macao SAR (Nos. FDCT-22-009-MISE, 0060/2021/A2, and 0015/2020/AMJ).

References

[1] J. Kim and H. Park, H_∞ state feedback control for generalized continuous/discrete time-delay system, *Automatica*, 1999, 35(8), 1443–1451.

[2] J. Richard, Time-delay systems: An overview of some recent advances and open problems, *Automatica*, 2003, 39(10), 1667–1694.

[3] D. Bresch-Pietri, J. Chauvin, and N. Petit, Adaptive control scheme for uncertain time-delay systems, *Automatica*, 2012, 48(8), 1536–1552.

[4] Y. Zhu, M. Krstic, and H. Su, Adaptive output feedback control for uncertain linear time-delay systems, *IEEE Trans. Autom. Control*, 2017, 62(2), 545–560.

[5] X. Li and X. Yang, Lyapunov stability analysis for nonlinear systems with state-dependent state delay, *Automatica*, 2020, 112, 108674.

[6] Z. Zhang, B. Xu, C. Tan, and S. Ge, Adaptive control of uncertain nonlinear time-delay systems with external disturbance, *IEEE Trans. Syst. Man Cybern. Syst.*, 2022, 52(2), 1288–1295.

[7] J. Ma, S. Xu, and S. Fei, Asymptotic tracking control of nonlinear time-delay systems with mismatched disturbances, *IEEE Trans. Circuits Syst. II: Express Briefs*, 2023, 70(1), 261–265.

[8] C. Hua, S. Feng, and L. Zhang, Output-feedback control of interconnected nonlinear systems with both state and input delays based on truncated prediction, *IEEE Trans. Circuits Syst. II: Express Briefs*, 2023, 70(11), 4108–4112.

[9] S. Battilotti, Continuous-time and sampled-data stabilizers for nonlinear systems with input and measurement delays, *IEEE Trans. Autom. Control*, 2020, 65(4), 1568–1583.

[10] Q. Wei, L. Zhu, T. Li, and D. Liu, A new approach to finite-horizon optimal control for discrete-time affine nonlinear systems via a pseudolinear method, *IEEE Trans. Autom. Control*, 2022, 67(5), 2610–2617.

[11] Q. Wei, Z. Yang, H. Su, and L. Wang, Online adaptive dynamic programming for optimal self-learning control of VTOL aircraft systems with disturbances, *IEEE Trans. Autom. Sci. Eng.*, 2024, 21(1), 343–352.

[12] M. Mohammadi, M. Arefi, P. Setoodeh, and O. Kaynak, Optimal tracking control based on reinforcement learning value iteration algorithm for time-delayed nonlinear systems with external disturbances and input constraints, *Inf. Sci.*, 2021, 554, 84–98.

[13] C. Liu, R. Loxton, K. Teo, and S. Wang, Optimal state-delay control in nonlinear dynamic systems, *Automatica*, 2022, 135, 109981.

[14] Y. Wang, Y. Liu, X. Yu, and H. Zhang, Jointed observer-based sliding mode predictive control for interconnected power systems with input delays, *IEEE Trans. Autom. Sci. Eng*., to be published.

[15] Y. Liu, H. Zhang, R. Yu, and Z. Xing, H_∞ tracking control of discrete-time system with delays via data-based adaptive dynamic programming, *IEEE Trans. Syst. Man Cybern. Syst*., 2020, 50(11), 4078–4085.

[16] S. Li, L. Ding, H. Gao, Y. Liu, N. Li, and Z. Deng, Reinforcement learning neural network-based adaptive control for state and input time-delayed wheeled mobile robots, *IEEE Trans. Syst. Man Cybern. Syst*., 2020, 50(11), 4171–4182.

[17] J. Liu, X. Zhang, X. Xu, and Q. Xiong, Receding horizon actor-critic learning control for nonlinear time-delay systems with unknown dynamics, *IEEE Trans. Syst. Man Cybern. Syst*., 2023, 53(8), 4980–4993.

[18] S. Jiao, Q. Wei, and F.-Y. Wang, A novel parallel control method for optimal consensus of nonlinear multiagent systems, *IEEE Trans. Cybern*., 2024, 54(10), 5912–5925.

[19] Q. Wei, S. Jiao, F.-Y. Wang, and Q. Dong, Robust optimal parallel tracking control based on adaptive dynamic programming, *IEEE Trans. Cybern*., 2024, 54(7), 4308–4321.

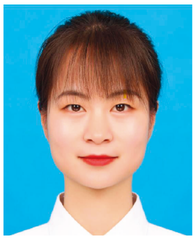
[20] S. Jiao, Q. Wei, and J. Xiao, Parallel control with event-based adaptive critic implementation for robust optimal tracking of uncertain nonlinear systems, *IEEE Trans. Autom. Sci. Eng*., to be published.

[21] Y. Zhu and D. Zhao, Online minimax Q network learning for two-player zero-sum Markov games, *IEEE Trans. Neural Netw. Learn. Syst*., 2022, 33(3), 1228–1241.

[22] B.-S. Chen, C.-T. Yang, and M.-Y. Lee, Multiplayer noncooperative and cooperative minimax H_∞ tracking game strategies for linear mean-field stochastic systems with applications to cyber-social systems, *IEEE Trans. Cybern*., 2022, 52(5), 2968–2980.

[23] R. Moghadam and S. Jagannathan, Optimal adaptive control of uncertain nonlinear continuous-time systems with input and state delays, *IEEE Trans. Neural Netw. Learn. Syst*., 2023, 34(6), 3195–3204.

[24] K. Vamvoudakis and F. Lewis, Online actor-critic algorithm to solve the continuous-time infinite horizon optimal control problem, *Automatica*, 2010, 46(5), 878–888.



Shanshan Jiao received the MS degree in control theory and control engineering from Xi’an University of Technology, China, in 2021. She is currently pursuing the PhD degree in intelligent science and systems at Institute of Systems Engineering, Macau University of Science and Technology, Macao, China. Her research interests include adaptive dynamic programming and neural network based control.

Qinglai Wei received the BS degree in automation and the PhD degree in control theory and control engineering from Northeastern University, Shenyang, China, in 2002 and 2009, respectively. He is currently a professor at State Key Laboratory of Multimodal Artificial Intelligence Systems, Institute of Automation, Chinese Academy of Sciences, Beijing, China. His research interests include adaptive dynamic programming, neural network based control, optimal control, nonlinear systems, and their industrial applications.