Self-Triggered Approximate Optimal Neuro-Control for Nonlinear Systems Through Adaptive Dynamic Programming

Bo Zhao, *Senior Member, IEEE*, Shunchao Zhang, and Derong Liu, *Fellow, IEEE*

***Abstract*— In this article, a novel self-triggered approximate optimal neuro-control scheme is presented for nonlinear systems by utilizing adaptive dynamic programming (ADP). According to the Bellman principle of optimality, the cost function of the general nonlinear system is approximated by building a critic neural network with a nested updating weight vector. Thus, the Hamilton–Jacobi–Bellman equation is solved to indirectly obtain the approximate optimal neuro-control input. In order to reduce the computation, the communication bandwidth, and the energy consumption, an appropriate self-triggering condition is designed as an alternative way to predict the updating time instants of the approximate optimal neuro-control policy. On the basis of Lyapunov's direct method, the stability of the closed-loop nonlinear system is analyzed and guaranteed to be uniformly ultimately bounded. Simulation results of two practical systems illustrate the present ADP-based self-triggered approximate optimal neuro-control scheme to be reasonable and effective.**

***Index Terms*— Adaptive dynamic programming (ADP), neural networks (NNs), optimal control, reinforcement learning, self-triggered control.**

I. Introduction

IN THE modern control framework, the control performance greatly depends on the communication networks, which transmits a large amount of data among controllers, sensors, and actuators. Especially, when a system structure grows its size and complexity, the deployment cost is increased accordingly. Therefore, in order to ensure both the system stability and the control performance, there exists a strong desire for controlled systems to not only provide an adequate level of control performance but also reduce resource and energy consumption.

Much effort has been made on developing optimal control methods to achieve adequate control level for nonlinear systems with different scenarios. It is worth pointing out that dynamic programming [1], which has been paid great attention by many researchers, is regarded as a classical and effective tool to obtain optimal control solutions of nonlinear systems. Furthermore, to avoid its "curse of dimensionality" that occurs in systems with the increasing structural complexity, adaptive dynamic programming (ADP), which is assisted by neural networks (NNs) or fuzzy logic systems, has been developed by Werbos [2] in the 1970s. During the last several decades, ADP-based optimal [3], [4], [5] and suboptimal [6] control approaches have received great success in tackling the control problems of continuous-time (CT) or discrete-time (DT) nonlinear systems with uncertainties or external disturbances [7], [8], input or output constraints [9], time-delay [10] and failures [11] to solve trajectory tracking [12], [13], zero-sum or nonzero-sum games [14], and so on. Moreover, some attempts have been made to implement the ADP-based control strategies to practical systems, such as residential energy scheduling [15], induction motor driving [16], power systems [17], microgrid energy management [18], near space vehicles [19], robot systems [12], and active suspension systems [20].

In order to reduce resource and energy consumption, ADP-based feedback control policies have been developed from periodic to aperiodic schedule, i.e., from time-triggered to event-triggered control strategy. Distinguished from the time-triggered control framework, the event-triggered control method [12], [21] is responsive and generates sensor sampling and control action as long as the system state diverges more than a certain threshold determined by the developed appropriate event-triggering condition, that is to say, the data transmission among the controllers, the sensors, and the actuators is greatly reduced to save the computational burden, the communication bandwidth, and the energy consumption.

Although the event-triggered control offers clear superiority to the periodic control, it is worth emphasizing that the triggering condition is continuously monitored based on current measurements from a hardware device. It implies that full-state information should be assumed to be available, but

Manuscript received 12 April 2022; revised 24 June 2023; accepted 9 January 2024. This work was supported in part by the National Natural Science Foundation of China under Grant 61973330 and Grant 62073085; in part by the Beijing Natural Science Foundation under Grant 4212038; in part by the Beijing Normal University Tang Scholar; and in part by the Open Research Project of the Key Laboratory of Industrial Internet of Things and Networked Control, Ministry of Education, under Grant 2021FF10. *(Corresponding author: Derong Liu.)*

Bo Zhao is with the School of Systems Science, Beijing Normal University, Beijing 100875, China, and also with the Key Laboratory of Industrial Internet of Things and Networked Control, Ministry of Education, Chongqing University of Posts and Telecommunications, Chongqing 400065, China (e-mail: zhaobo@bnu.edu.cn).

Shunchao Zhang is with the School of Internet Finance and Information Engineering, Guangdong University of Finance, Guangzhou 510521, China (e-mail: 47-319@gduf.edu.cn).

Derong Liu is with the School of System Design and Intelligent Manufacturing, Southern University of Science and Technology, Shenzhen 518055, China, and also with the Department of Electrical and Computer Engineeering, University of Illinois at Chicago, Chicago, IL 60607 USA (e-mail: liudr@sustech.edu.cn; derong@uic.edu).

Digital Object Identifier 10.1109/TNNLS.2024.3362800

this assumption is always violated in most practical situations. Some beneficial attempts, which combine the event- and self-triggered mechanisms, have been made for different types of plants. Kishida et al. [22] presented an integrated event- and self-triggered networked control scheme for uncertain linear systems with finite-gain $\mathcal{L}_2$ stability, where the event-triggering condition decides whether a new control signal would be transmitted to the actuator at each sampling, and the self-triggering condition determines the time instants of sampling. Zhou et al. [23] proposed a combined self- and event-triggered control strategy to investigate the output control approach for quantized linear systems. Qi et al. [24] presented an event- and self-triggered boldsymbol control method for switched linear systems with exogenous disturbance. The current sampled data are employed to adaptively predict the inter-execution intervals to decrease resource consumption. Sahoo et al. [25] proposed a mixed event- and self-triggering-based regulation method for CT linear dynamical systems. Zuo et al. [26] developed both event- and self-triggered transmission strategies for dynamical systems by designing a triggering condition to save network resources.

To avoid the continuous monitoring of the system state by a hardware device, the self-triggering mechanism is introduced to design controllers by some researchers. In contrast to the event-triggered control which updates the control policy based on the developed triggering condition, the next triggering time in the self-triggered control [27] is determined at the former triggering time instant, that is to say, the continuous monitoring of the system state through hardware devices is not required anymore at the sensor side, and the embedded devices can shut their communication until the next transmission time instant. Focusing on linear systems, Zhang et al. [28] designed a self-triggered gain scheduling control for achieving the stabilization semiglobally for input constrained linear systems to avoid continuous monitoring the system states. Lu and Maciejowski [29] investigated a self-triggering mechanism-based model predictive control (MPC) strategy for linear systems in the presence of both state and input constraints, where both the updating of MPC control policy and the next triggering instant are determined according to the relaxed dynamic programming inequality. Brunner et al. [30] investigated a novel self-triggered aperiodic control method for perturbed DT linear systems by evaluating the set-membership conditions. In this way, a tradeoff is realized between the communication rate and the worse case asymptotic limitation on the closed-loop system state. For nonlinear systems, Liu et al. [31] presented an optimized self-triggered strategy-based robust MPC scheme for constrained DT uncertain nonlinear systems subject to disturbances. Li and Li [32] proposed a self-triggered distributed MPC scheme to reach consensus of a heterogeneous time-varying multiagent system. The self-triggering time intervals are alternatively optimized by the control inputs, and the influence on the system performance is analyzed. Based on the state information of each agent collected from its neighbors, Fan et al. [33] proposed a self-triggered consensus algorithm with Zeno-exclusion analysis for multiagent systems. Gao et al. [34] proposed a state estimation-based self-triggered control scheme for cyber-physical systems subject to joint attacks of both sensors and actuators with the enhancement of resource saving. Furthermore, Gao et al. [35] developed a robust self-triggered control scheme for time-varying constrained uncertain systems by analyzing the reachability for both linear and nonlinear scenarios. Furthermore, some attempts have been made for practical implementations. Cao et al. [36] proposed a self-triggered MPC strategy to investigate the trajectory tracking control problem for nonholonomic vehicles considering coupled input constraints and bounded perturbations. By introducing the self-triggered communication strategy, Zhou and Tokekar [37] developed a decentralized target tracking method for multirobot teams, and the time when a particular robot should search the online data from its neighbors and when it is secure to manipulate with possibly out of date data is determined based on the self-triggering mechanism to reduce the communication bandwidth.

From the aforementioned state-of-the-art of existing self-triggered control schemes, it is widely explored in MPC, distributed control, and networked control, rare works focused on optimal control. Lou and Ji [38] investigated a new self-triggered adaptive optimal control method for nonlinear CT systems leading to a high quantitative accuracy at a limited channel transmission rate. Kobayashi and Hiraishi [39] investigated the synthesis of self-triggered control problem for network systems in an optimal manner by computing the control input and the sampling time simultaneously.

In order to avoid the continuous monitoring through hardware devices in existing event-triggered control strategies, as well as to decrease the computational burden, the communication bandwidth, and the energy consumption, we present a new self-triggered approximate optimal neuro-control method for nonlinear systems based on ADP. The main contributions and novelties are concluded as following three aspects.

1) Different from existing ADP-based event-triggered control strategies [12], [21], the next triggering instant in the developed ADP-based self-triggered control is predicted by a software based on the previous triggering instant, that is to say, a new triggering condition based on self-triggering mechanism is explored to predict the next triggering instant. Thus, the continuous monitoring of the system state in the event-triggered control strategy is avoided, and the hardware device for continuous monitoring the system state is not required anymore.
2) By selecting a proper design structure for the nested updating policies, the critic NN weight error dynamics is guaranteed to be asymptotically stable, rather than uniformly ultimately bounded (UUB) in most of existing ADP-based control schemes.
3) The self-triggered control scheme cannot only guarantee the system to be stable in an optimal manner but also reduces the computation, the precious communication bandwidth, and the energy consumption. In other words, the developed control scheme offers a feasible tradeoff between the overall resource cost and the system control performance, which is significant in real implementations.

The remainder of this article is structured as follows. In Section II, the problem statement is given. In Section III, the self-triggered approximate optimal neuro-control is designed via the ADP framework in detail, and the stability analysis is offered. In Section IV, simulation studies illustrate the effectiveness of the developed approach. In Section V, concluding remarks are briefly described.

II. Problem Statement

The considered nonlinear system dynamics is modeled in the general form as

$$\dot{x} = f(x) + g(x)u(x) \tag{1}$$

where $x(t) \in \mathbb{R}^n$ and $u(t) \in \mathbb{R}^m$ are the system state and control input vectors, respectively. $f(x) \in \mathbb{R}^n$ and $g(x) \in \mathbb{R}^{n\times m}$ represent the known drift dynamics and control input matrix, respectively. To ease the notation, $x(t) = x$ is denoted in the sequel.

To make the analysis more amenable, the following two assumptions are provided.

Assumption 1: The vector-valued drift function $f(x)$ is Lipschitz continuous on the compact set $\Omega \in \mathbb{R}^n$ including the origin such that the solution $x(t)$ of the nonlinear system (1) is unique for the given equilibrium point $x_0 \in \Omega$ and u. Moreover, there is a scalar $D_f > 0$ such that $\|f(x)\| \leq D_f\|x\|$, where $\|\cdot\|$ indicates the 2-norm of a vector, and the considered system (1) can be stabilized on Ω.

Assumption 2: The control input matrix $g(x)$ is norm-bounded as $0 < \|g(x)\|_F \leq D_g$ for arbitrary $x \in \Omega$, where $D_g > 0$ and $\|\cdot\|$ denotes the Frobenius norm of a matrix.

Define the infinite-horizon cost function as

$$V(x,u) = \int_t^\infty U(x(\tau), u(\tau))\mathrm{d}\tau \tag{2}$$

where $U(x,u) = x^{\mathsf{T}}Qx + u^{\mathsf{T}}Ru \geq 0$ for all $x \in \mathbb{R}^n$ and $u \in \mathbb{R}^m$, $U(0,0) = 0$, $V(0,0) = 0$, and $Q \in \mathbb{R}^{n\times n}$ and $R \in \mathbb{R}^{m\times m}$ are symmetric positive definite matrices.

Definition 1: For the nonlinear system (1), a control policy $u(x)$ is defined to be admissible with respect to (2) if $u(x)$ is continuous on a set $\Omega \in \mathbb{R}^n$, $u(0) = 0$, $u(x)$ stabilizes the nonlinear system (1), and $V(x_0)$ in (2) is finite for all $x \in \Omega$, where $x_0 = x(0)$ is the initial state of x. The admissible control set $\psi(\Omega)$ consists of admissible control policies.

For any given admissible control policy u in the admissible control set $\psi(\Omega)$, if the associated cost function (2) is continuously differentiable and $V \in C^1$, then we can derive the nonlinear equation by taking the time derivative of (2) as in [41]

$$U(x,u) + \nabla V^{\mathsf{T}}(x)(f(x) + g(x)u) = 0 \tag{3}$$

where $\nabla V^{\mathsf{T}}(x) = \partial V(x)/\partial x$ indicates the partial gradient of $V(x)$ with respect to the system state x.

To drive the closed-loop system (1) to be convergent, the optimal control policy $u^*(t) \in \psi(\Omega)$ should be achieved by seeking the minimized cost function, i.e., the optimal cost function as

$$V^*(x) = \min_{u\in\psi(\Omega)} \int_t^\infty U(x(\tau), u(\tau))\mathrm{d}\tau. \tag{4}$$

By considering the optimal cost function (4), the associate Hamiltonian is defined for the nonlinear system (1) as

$$H\big(x, u, \nabla V^*(x)\big) = U(x,u) + \nabla V^{*\mathsf{T}}(x)(f(x) + g(x)u). \tag{5}$$

Depending on the Bellman principle of optimality [1], the optimal cost function $V^*(x)$ is derived by seeking the solution of the following Hamilton–Jacobi–Bellman equation (HJBE)

$$0 = \min_{u\in\psi(\Omega)} H\big(x, u^*, \nabla V^*(x)\big). \tag{6}$$

Hence, the optimal neuro-control policy is illustrated in the closed form as

$$u^*(x) = -\frac{1}{2}R^{-1}g^{\mathsf{T}}(x)\nabla V^*(x). \tag{7}$$

By involving the self-triggered mechanism, the present optimal neuro-control policy is updated when the self-triggered time instant is satisfied only, which is predicted by the present state and a given function through software. Thus, it is easy to implement in contrast to the event-triggered optimal control schemes.

With the help of the prediction, the optimal neuro-control policy is updated under the self-triggered mechanism as $u_k = u(x(t_k))$ at the predicted time t_k. Therefore, the control target of this article is to present a self-triggered approximate optimal neuro-control policy $u(x(t_k))$ to make the closed-loop system (1) to be stable.

III. Self-Triggered Approximate Optimal Neuro-Controller Design and Stability Analysis

A. Asymptotically Converged Critic NN

Starting from the cost function (2), it is tough to solve the HJBE (6). Fortunately, based on the general approximation ability of NNs, $V(x)$ is accurately approximated by a feedforward NN with only one hidden layer [40], [41] such that

$$V(x) = W_c^{\mathsf{T}}\sigma(x) + \varepsilon_c(x) \tag{8}$$

where $W_c \in \mathbb{R}^{l_c}$, $\sigma(x) \in \mathbb{R}^{l_c}$, and $\varepsilon_c(x) \in \mathbb{R}$ represent the unknown optimal weight vector, the activation function, and the approximation error, respectively, and l_c indicates the quantity of hidden neurons. Thus, taking the partial derivative of (8) along the system state x, we have

$$\nabla V(x) = \nabla\sigma^{\mathsf{T}}(x)W_c + \nabla\varepsilon_c^{\mathsf{T}}(x) \tag{9}$$

where $\nabla\sigma^{\mathsf{T}}(x) = \partial\sigma(x)/\partial x \in \mathbb{R}^{n\times l_c}$ and $\nabla\varepsilon_c^{\mathsf{T}}(x) = \partial\varepsilon_c(x)/\partial x \in \mathbb{R}^n$ are the corresponding partial gradients along the system state x, respectively.

In this case, the Hamiltonian for the nonlinear system (1) is defined as

$$H(x, u, W_c) = U(x,u) + W_c^{\mathsf{T}}\nabla\sigma(x)\dot{x} + \nabla\varepsilon_c(x)\dot{x}. \tag{10}$$

In order to derive the optimal control policy (7) customized to a given controlled plant, the unknown optimal weight vector W_c is trained by approximating (8) as

$$\hat{V}(x) = \hat{W}_c^{\mathsf{T}}\sigma(x) \tag{11}$$

where $\hat{W}_c$ is the estimate of W_c. The partial gradient with respect to the corresponding system state x of (11) is described as

$$\nabla \hat{V}(x) = \nabla \sigma^{\mathsf{T}}(x) \hat{W}_c. \tag{12}$$

Under the approximation (11) of the cost function $V(x)$, the Hamiltonian (43) is approximated by

$$\begin{aligned} H\big(x, u, \hat{W}_c\big) &= U(x, u) + \hat{W}_c^{\mathsf{T}} \nabla \sigma(x) \dot{x} \\ &= e_c. \end{aligned} \tag{13}$$

By comparing (13) with (5), the Hamiltonian approximation error is obtained as

$$e_c = \varepsilon - \tilde{W}_c^{\mathsf{T}} \theta \tag{14}$$

where $\tilde{W}_c = W_c - \hat{W}_c$ denotes the weight approximation error vector, $\theta = \nabla \sigma(x)\dot{x}$, and $\varepsilon = -\nabla \varepsilon_c(x)\dot{x}$ is norm-bounded by a positive constant ε_M as $\|\varepsilon\| \leq \varepsilon_M$.

Similarly, to derive the critic NN weight vector, the target function $E_c = (1/2)e_c^{\mathsf{T}} e_c$ is minimized through the commonly used steepest descent algorithm. Thus, the weight vector is updated by

$$\dot{\hat{W}}_c = -l_c \left[\frac{\partial E_c}{\partial \hat{W}_c} \right] \tag{15}$$

where $l_c > 0$ is a learning rate. On this basis, the nested updating policies are described as

$$\dot{\hat{W}}_c = -l_c\big(e_c - \hat{\varepsilon}_M - \Gamma \operatorname{sgn}\big(\tilde{W}_c^{\mathsf{T}} \theta\big)\big)\theta \tag{16}$$

where $\Gamma > 0$ is a design parameter, $\hat{\varepsilon}_M$ is the estimate of ε_M tuned by

$$\dot{\hat{\varepsilon}}_M = -l_\varepsilon \hat{W}_c^{\mathsf{T}} \theta, \tag{17}$$

and $l_\varepsilon > 0$ is a learning rate. Define the weight approximation error vector as $\tilde{W}_c = W_c - \hat{W}_c$, and it is adjusted by

$$\begin{aligned} \dot{\tilde{W}}_c &= -\dot{\hat{W}}_c \\ &= l_c\big(\varepsilon - \varepsilon_M + \tilde{\varepsilon}_M - \tilde{W}_c^{\mathsf{T}} \theta - \Gamma \operatorname{sgn}\big(\tilde{W}_c^{\mathsf{T}} \theta\big)\big)\theta \end{aligned} \tag{18}$$

where $\tilde{\varepsilon}_M = \varepsilon_M - \hat{\varepsilon}_M$, and it is adjusted by

$$\dot{\tilde{\varepsilon}}_M = -l_\varepsilon \tilde{W}_c^{\mathsf{T}} \theta. \tag{19}$$

Remark 1: From (18) and (19), we know that the approximation error item $\tilde{\varepsilon}_M$, adjusted by (19), is embedded in the updating policy (18) of $\tilde{W}_c$. Thus, the updating policies (18) and (19) are then regarded as "the nested updating policies."

Next, we will show the nested updating policies (18) and (19) can ensure the asymptotic stability of the critic NN weight error dynamics, rather than UUB in most existing ADP-based optimal control methods.

Theorem 1: For the nonlinear system (1), the developed nested updating policies (18) and (19) ensure the critic NN weight error dynamics to be asymptotically stable.

Proof: Considering both approximation errors in the nested updating policies, select a Lyapunov function candidate as

$$L_1 = \operatorname{tr}\left(\frac{1}{2l_c} \tilde{W}_c^{\mathsf{T}} \tilde{W}_c \right) + \frac{1}{2l_\varepsilon} \tilde{\varepsilon}_M^2. \tag{20}$$

For (20), taking its time derivative and plugging the nested updating policy (18) into it, we derive

$$\begin{aligned} \dot{L}_1 &= \operatorname{tr}\left(\frac{1}{l_c} \tilde{W}_c^{\mathsf{T}} \dot{\tilde{W}}_c \right) + \frac{1}{l_\varepsilon} \dot{\tilde{\varepsilon}}_M \tilde{\varepsilon}_M \\ &= \operatorname{tr}\big(\tilde{W}_c^{\mathsf{T}}(\varepsilon - \varepsilon_M)\theta\big) + \operatorname{tr}\big(\tilde{W}_c^{\mathsf{T}} \tilde{\varepsilon}_M \theta\big) \\ &\quad - \Gamma \big\| \tilde{W}_c^{\mathsf{T}} \theta \big\| - \big\| \tilde{W}_c^{\mathsf{T}} \theta \big\|^2 + \frac{1}{l_\varepsilon} \dot{\tilde{\varepsilon}}_M \tilde{\varepsilon}_M. \end{aligned} \tag{21}$$

For matrices X and Y, if $YX \in \mathbb{R}$, we notice that $\operatorname{tr}(XY) = \operatorname{tr}(YX) = YX$. Introducing the updating policy (19) into (21), we derive

$$\begin{aligned} \dot{L}_1 &= \operatorname{tr}\big(\tilde{W}_c^{\mathsf{T}} \theta(\varepsilon - \varepsilon_M)\big) + \operatorname{tr}\big(\tilde{W}_c^{\mathsf{T}} \theta \tilde{\varepsilon}_M\big) - \Gamma \big\| \tilde{W}_c^{\mathsf{T}} \theta \big\| \\ &\quad - \big\| \tilde{W}_c^{\mathsf{T}} \theta \big\|^2 + \frac{1}{l_\varepsilon} \dot{\tilde{\varepsilon}}_M \tilde{\varepsilon}_M \\ &\leq \big\| \tilde{W}_c^{\mathsf{T}} \theta \big\| \| \varepsilon - \varepsilon_M \| + \operatorname{tr}\big(\tilde{W}_c^{\mathsf{T}} \theta \tilde{\varepsilon}_M\big) - \Gamma \big\| \tilde{W}_c^{\mathsf{T}} \theta \big\| \\ &\quad - \big\| \tilde{W}_c^{\mathsf{T}} \theta \big\|^2 + \frac{1}{l_\varepsilon} \dot{\tilde{\varepsilon}}_M \tilde{\varepsilon}_M \\ &\leq 2\varepsilon_M \big\| \tilde{W}_c^{\mathsf{T}} \theta \big\| + \operatorname{tr}\big(\tilde{W}_c^{\mathsf{T}} \theta \tilde{\varepsilon}_M\big) - \Gamma \big\| \tilde{W}_c^{\mathsf{T}} \theta \big\| \\ &\quad - \big\| \tilde{W}_c^{\mathsf{T}} \theta \big\|^2 + \frac{1}{l_\varepsilon} \dot{\tilde{\varepsilon}}_M \tilde{\varepsilon}_M \\ &= -(\Gamma - 2\varepsilon_M) \big\| \tilde{W}_c^{\mathsf{T}} \theta \big\| - \big\| \tilde{W}_c^{\mathsf{T}} \theta \big\|^2. \end{aligned} \tag{22}$$

From the above analysis, we know that if there exists a design parameter $\Gamma \geq 2\varepsilon_M$, then, $\dot{L}_1 \leq 0$. Hence, it indicates that the nested updating policies (18) and (19) guarantee the asymptotic stability of the critic NN weight error vector $\tilde{W}_c$. This ends the proof. ■

B. Self-Triggered Approximate Optimal Neuro-Control

Based on the partial derivative of the approximate critic NN (12), the desired optimal neuro-control policy (7) is approximated by

$$\hat{u}(x) = -\frac{1}{2} R^{-1} g^{\mathsf{T}}(x) (\nabla \sigma(x))^{\mathsf{T}} \hat{W}_c. \tag{23}$$

According to the self-triggered mechanism, the optimal neuro-control policy can be updated at the predicted time sequence as

$$\hat{u}(x_k) = \hat{u}_k = -\frac{1}{2} R^{-1} g^{\mathsf{T}}(x_k) (\nabla \sigma(x_k))^{\mathsf{T}} \hat{W}_c. \tag{24}$$

C. Stability Analysis

Before showing the stability of the closed-loop system (1) with the developed self-triggered approximate optimal neuro-control policy (24), the following assumptions are provided.

Assumption 3: On the compact set $\Omega \in \mathbb{R}^n$, the control input function $u(x)$ is Lipschitz continuous, i.e., there is a scalar $D_u > 0$ such that $\|u(x(t)) - u(x(t_k))\| = \|u^* - u_k\| \leq D_u \|x_e\|$, where $x_k = x(t_k)$ and $x_e = x - x_k$.

Assumption 4: $\hat{W}_c$, $\nabla \sigma(x)$, and $\nabla \varepsilon_c$ are norm-bounded as $\|\hat{W}_c\| \leq W_{cM}$, $\|\nabla \sigma(x)\| \leq \sigma_{cM}$, and $\|\nabla \varepsilon_c\| \leq \varepsilon_{cM}$, where W_{cM}, σ_{cM}, and ε_{cM} are unknown positive scalars.

Theorem 2: For the nonlinear system (1), consider the cost function (2), Assumptions 1–4, and the developed self-triggered approximate optimal neuro-control policy (24).

If there are positive constants $\delta \in (0,1]$ and $\epsilon \in (0,1)$ such that

$$\delta\big(1-\epsilon^2\big)x_k^{\mathsf T}Nx_k \ge x_e^{\mathsf T}\left(\left(1-\frac{1}{\epsilon^2}\right)N+M\right)x_e \tag{25}$$

where N and M are later defined positive definite diagonal matrices with proper dimensions. Then, the UUB stability can be assured for the closed-loop nonlinear system (1).

Proof: Select a Lyapunov function candidate as

$$L_2 = \frac{1}{2}x^{\mathsf T}x + V^*(x). \tag{26}$$

Substituting the developed self-triggered approximate optimal neuro-control policy (24), the time derivative of (26) becomes

$$\begin{aligned}
\dot{L}_2 &= x^{\mathsf T}\dot{x} + \nabla V^*(x)\dot{x}\\
&\le x^{\mathsf T}\big(f(x)+g(x)\hat{u}_k\big)\\
&\quad + \nabla V^*(x)\big(f(x)+g(x)\big(u^*(x)+\hat{u}_k-u^*(x)\big)\big)\\
&\le x^{\mathsf T}\big(f(x)+g(x)\hat{u}_k\big) + \nabla V^*(x)\big(f(x)+g(x)u^*(x)\big)\\
&\quad + 2u^{*\mathsf T}R\big(u^*(x)-\hat{u}_k\big).
\end{aligned} \tag{27}$$

Based on Assumptions 1 and 2, considering (3) and utilizing Young's inequality, we obtain

$$\begin{aligned}
\dot{L}_2 &\le -x^{\mathsf T}D_f x + \frac{1}{2}x^{\mathsf T}D_g x + \frac{1}{2}\hat{u}_k^{\mathsf T}D_g\hat{u}_k\\
&\quad + 2u^{*\mathsf T}R\big(u^*-\hat{u}_k\big) - x^{\mathsf T}Qx - u^{*\mathsf T}Ru^*\\
&\le -x^{\mathsf T}\left(D_f+\frac{1}{2}D_g\right)x + 2u^{*\mathsf T}R\big(u^*-\hat{u}_k\big)\\
&\quad + \frac{1}{2}\big(u^*-u^*+\hat{u}_k\big)^{\mathsf T}D_g\big(u^*-u^*+\hat{u}_k\big)\\
&\quad - x^{\mathsf T}Qx - u^{*\mathsf T}Ru^*\\
&= -x^{\mathsf T}\left(D_f+\frac{1}{2}D_g\right)x + \frac{1}{2}\big(u^*-\hat{u}_k\big)^{\mathsf T}D_g\big(u^*-\hat{u}_k\big)\\
&\quad - u^{*\mathsf T}\left(R-\frac{1}{2}D_g\right)u^* - x^{\mathsf T}Qx\\
&\quad + u^{*\mathsf T}\big(2R-D_gI\big)\big(u^*-\hat{u}_k\big).
\end{aligned} \tag{28}$$

According to Assumption 4, we notice that

$$\begin{aligned}
\|u^*-\hat{u}_k\| &= \|u^*-\hat{u}+\hat{u}-\hat{u}_k\|\\
&\le \|u^*-\hat{u}\| + \|\hat{u}-\hat{u}_k\|\\
&\le \frac{1}{2}\|R^{-1}\|_F\|g(x)\|_F\|\tilde{W}_c\nabla\sigma(x)+\nabla e_c(x)\|\\
&\quad + D_u\|x_e\|\\
&\le \frac{1}{2}\|R^{-1}\|_F D_g\mu + D_u\|x_e\|
\end{aligned} \tag{29}$$

where $\mu = \|W_{cM}\sigma_M + e_{cM}\|$. Thus, we have

$$\begin{aligned}
\dot{L}_2 \le &-x^{\mathsf T}\left(Q-\left(D_f+\frac{1}{2}D_g\right)I\right)x + \frac{1}{4}R^{-2}D_g^3\mu^2\\
&+ x_e^{\mathsf T}D_gD_u^2x_e - u^{*\mathsf T}\left(R-\frac{1}{2}D_g\right)u^*\\
&+ \|u^*\|\|2R-D_gI\|_F\left(\frac{1}{2}R^{-1}D_g\mu + D_u\|x_e\|\right)\\
\le &-x^{\mathsf T}\left(Q-\left(D_f+\frac{1}{2}D_g\right)I\right)x + \frac{1}{4}R^{-2}D_g^3\mu^2\\
&+ x_e^{\mathsf T}D_gD_u^2x_e - u^{*\mathsf T}\left(R-\frac{1}{2}D_g\right)u^*\\
&+ \frac{1}{2}u^{*\mathsf T}D_u^2u^* + \frac{1}{2}\|2R-D_gI\|_F^2x_e^{\mathsf T}x_e\\
\le &-x^{\mathsf T}\left(Q-\left(D_f+\frac{1}{2}D_g\right)I\right)x\\
&- u^{*\mathsf T}\left(R-\frac{1}{2}\big(D_g-D_u^2-D_g^2\big)I\right)u^*\\
&+ x_e^{\mathsf T}\left(D_gD_u^2+\frac{1}{2}\|2R-D_gI\|_F^2\right)x_e + \Delta\\
\le &-x^{\mathsf T}\left(Q-\left(D_f+\frac{1}{2}D_g\right)I-N\right)x - x^{\mathsf T}Nx\\
&- u^{*\mathsf T}\left(R-\frac{1}{2}\big(D_g-D_u^2-D_g^2\big)I\right)u^*\\
&+ x_e^{\mathsf T}Mx_e + \Delta
\end{aligned} \tag{30}$$

where $\Delta = (1/4)R^{-2}D_g^3\mu^2 + (1/8)R^{-2}\mu^2\|2R-D_gI\|_F^2$, $N \in \mathbb{R}^{n\times n}$ is a positive definite diagonal matrix, and $M = (D_gD_u^2 + (1/2)\|2R-D_gI\|_F^2)I$.

Noticing that $x = x_k + x_e$, we have

$$\begin{aligned}
x^{\mathsf T}Nx &= (x_k+x_e)^{\mathsf T}N(x_k+x_e)\\
&= x_k^{\mathsf T}Nx_k + x_e^{\mathsf T}Nx_e + 2x_k^{\mathsf T}Nx_e\\
&= \big(1-\epsilon^2\big)x_k^{\mathsf T}Nx_k + \left(1-\frac{1}{\epsilon^2}\right)x_e^{\mathsf T}Nx_e\\
&\quad + \left(\epsilon x_k+\frac{1}{\epsilon}x_e\right)N\left(\epsilon x_k+\frac{1}{\epsilon}x_e\right)\\
&\ge \big(1-\epsilon^2\big)x_k^{\mathsf T}Nx_k + \left(1-\frac{1}{\epsilon^2}\right)x_e^{\mathsf T}Nx_e.
\end{aligned} \tag{31}$$

Thus, (30) becomes

$$\begin{aligned}
\dot{L}_2 &\le -x^{\mathsf T}Q_0x - x^{\mathsf T}Nx - u^{*\mathsf T}R_0u^* + \Delta\\
&\quad - \big(1-\epsilon^2\big)x_k^{\mathsf T}Nx_k + x_e^{\mathsf T}\left(\left(1-\frac{1}{\epsilon^2}\right)N+M\right)x_e
\end{aligned} \tag{32}$$

where $Q_0 = Q - (D_f + (1/2)D_g)I - N$ and $R_0 = R - (1/2)(D_g - D_u^2 - D_g^2)I$. Considering the condition (25), we have

$$\begin{aligned}
\dot{L}_2 &\le -x^{\mathsf T}Q_0x - x^{\mathsf T}Nx - u^{*\mathsf T}R_0u^* + \Delta\\
&\le -\lambda_{\min}(Q_0)\|x\|^2 - \lambda_{\min}(R_0)\|u^*\|^2 + \Delta\\
&\le -\lambda_{\min}(Q_0)\|x\|^2 + \Delta.
\end{aligned} \tag{33}$$

We can conclude that $\dot{L}_2 \le 0$ as long as Q_0 and R_0 are selected as positive definite matrices, i.e., $Q > (D_f + (1/2)D_g)I + N$ and $R > (1/2)(D_g - D_u^2 - D_g^2)I$, as well as x locates outside the following compact set:

$$\Omega_x = \left\{x : \|x\| \le \sqrt{\frac{\Delta}{\lambda_{\min}(Q_0)}}\right\}.$$

Thus, the closed-loop nonlinear system (1) is assured to be UUB under the present self-triggered approximate optimal neuro-control scheme. This ends the proof. ■

Remark 2: For Assumptions 1 and 2, we notice that practical systems are always modeled as continuous nonlinear differential equations. Thus, it is common that it has a unique equilibrium point x_0. Furthermore, $f(x)$ and $g(x)$ denote the drift function and control input matrix in practical application systems, respectively. For example, $f(x)$ in robot systems consists of the inertia, Coriolis, and centripetal force, and $g(x)$ indicates the inertia, so it is feasible to assume them to be norm-bounded, and $g(x)$ cannot be zero. For Assumption 3, the ADP-based control strategy is developed based on policy iteration, which starts from a proper initial admissible control. It means that the initial admissible control can be regarded as a priori control. Meanwhile, we consider to design the control scheme in infinite horizon, so the Lipschitz continuity of control input $u(x)$ is satisfied. For Assumption 4, it is certain that the value of cost function is finite in practical systems, so its approximation cannot be infinite. Since $\tilde{W}_c$, $\nabla\sigma(x)$, and $\nabla\varepsilon_c$ denote the weight approximation error, and the partial gradients of activation function and NN approximation error, they can be ensured to be finite. Thus, Assumptions 3 and 4 can be fulfilled in practice. Moreover, these assumptions are commonly used in developing ADP-based control schemes [3], [7], [41].

D. Self-Triggered Mechanism

In this section, the inter-execution time $\Delta_k = t_{k+1} - t_k, k = 1, 2, \ldots, \infty$, is determined as $\Delta_k = \Phi(x_k)$, which implies that inter-execution time is lower bounded with a positive scalar, that is to say, the self-triggered mechanism is admissible. Before the proof, we provide the definition of Zeno behavior as follows.

Definition 2 (Zeno Behavior): The nonlinear system is Zeno if

$$\lim_{k\to\infty} t_k = \sum_{k=0}^{\infty} \Delta_k = t_\infty < \infty \tag{34}$$

where Δ_k is defined as the kth intersampling time and t_∞ is called the Zeno time [42].

Theorem 3: The self-triggered time sequence $\{t_k\}$ is determined by

$$t_{k+1} = t_k + \frac{1}{\lambda}\ln\left(\lambda\frac{\delta x_k^{\mathsf{T}} N x_k}{\rho(x_k)} + 1\right) \tag{35}$$

where $\lambda = 2 + 2D_f$ and $\rho(x_k) = (2D_f^2 + D_g^2 D_u^2)\lambda_{\max}(M)x_k^{\mathsf{T}} x_k$. Then, the self-triggering condition (35) is admissible for the nonlinear system (1). Furthermore, the inter-execution time is provided by

$$\Delta_k = \frac{1}{\lambda}\ln\left(\lambda\frac{\delta x_k^{\mathsf{T}} N x_k}{\rho(x_k)} + 1\right) > 0. \tag{36}$$

Proof: Choose a Lyapunov function candidate as $L_3(t) = x_e^{\mathsf{T}} M x_e,\ t \in [t_k, t_{k+1})$. We have its time derivative as

$$\begin{aligned}
\dot{L}_3 &= 2x_e^{\mathsf{T}} M \dot{x}_e \\
&= 2x_e^{\mathsf{T}} M\big(f(x) + g(x)\hat{u}_k\big) \\
&\le 2x_e^{\mathsf{T}} M x_e + f^{\mathsf{T}}(x) M f(x) + \big(g(x)\hat{u}_k\big)^{\mathsf{T}} M g(x)\hat{u}_k \\
&\le 2x_e^{\mathsf{T}} M x_e + x^{\mathsf{T}} D_f^2 M x + D_g^2 D_u^2 x_k^{\mathsf{T}} M x_k \\
&\le 2x_e^{\mathsf{T}} M x_e + 2\big(x_k^{\mathsf{T}} D_f^2 M x_k + x_e^{\mathsf{T}} D_f M x_e\big) \\
&\quad + D_g^2 D_u^2 x_k^{\mathsf{T}} M x_k \\
&= x_e^{\mathsf{T}}\big(2 + D_f\big) M x_e + x_k^{\mathsf{T}}\big(2D_f^2 + D_g^2 D_u^2\big) M x_k \\
&\le \lambda x_e^{\mathsf{T}} M x_e + \rho(x_k).
\end{aligned} \tag{37}$$

Thus, we have

$$\dot{L}_3 \le \ \lambda L_3 + \rho(x_k). \tag{38}$$

By using the comparison principle, it yields

$$L_3 \le \ \frac{\rho(x_k)}{\lambda}\big(e^{\lambda(t-t_k)} - 1\big),\ \ t \in \big[t_k, t_{k+1}\big). \tag{39}$$

Based on the self-triggered mechanism shown in (35), one has

$$\frac{\rho(x_k)}{\lambda}\big(e^{\lambda(t-t_k)} - 1\big) \le \delta x_k^{\mathsf{T}} N x_k,\ \ t \in \big[t_k, t_{k+1}\big). \tag{40}$$

According to (39) and (40), we can see that the nonlinear system (1) with the self-triggering condition (35) is ensured to be exponentially stable. Furthermore, it is obvious that

$$\Delta_k = \frac{1}{\lambda}\ln\left(\lambda\frac{\delta x_k^{\mathsf{T}} N x_k}{\rho(x_k)} + 1\right) > 0.$$

This indicates that the inter-execution time between the current and the next predict time instant is larger than zero, which means that the Zeno behavior can be avoided. This ends the proof. ■

In summary, the ADP-based self-triggered approximate optimal neuro-control scheme is described in Algorithm 1.

Algorithm 1 ADP-Based Self-Triggered Approximate Optimal Neuro-Control

```
1: Initialization: Initialize parameters $Q, R, \lambda, \delta, l_c, l_\varepsilon, N, \Gamma$, the terminal time of system
   operation $T$, and the computation accuracy $\xi > 0$ of cost function. Let $p = 0$, $k = 0$, $t_0 = 0$,
   $V^{(0)}(x_0) = 0$, and begin with admissible control policy $u_0^{(0)}(x_0)$.
2: Policy evaluation: Let $k \ge 0$, $p = p + 1$, based on the control policy $u_k^{(p)}(x_k)$, solve the
   following nonlinear equation for $u_k^{(p)}(x_k)$:
       $U\left(x_k, u_k^{(p)}\right) + \nabla V^{(p)\mathsf{T}}(x_k)\left(f(x_k) + g(x_k)u_k^{(p)}\right) = 0.$   (41)
3: Policy improvement: Update the control policy $u_k^{(p)}(x_k)$ by
       $u_k^{(p+1)}(x_k) = -\frac{1}{2}R_i^{-1}g_i^{\mathsf{T}}(x_k)\nabla V^{(p)}(x_k).$   (42)
4: If $\|V^{(p+1)}(x_k) - V^{(p)}(x_k)\| \le \xi$, go to 5 and obtain the self-triggered approximated
   optimal neuro-control $u_k$ at time instant $t_k$; else, return to 2.
5: Self-triggered mechasnism: Let $k = k + 1$, predict the next triggering time instant by
       $t_{k+1} = t_k + \frac{1}{\lambda}\ln\left(\lambda\frac{\delta x_k^{\mathsf{T}} N x_k}{\rho(x_k)} + 1\right).$   (43)
6: If $t_k \le T$, go to Step 2; else, stop.
```

E. Computational Requirements

Inspired by [43] and [44], the minimal computational requirements will be analyzed for the developed self-triggered approximate optimal neuro-control (24).

For the Veronese map ϑ from $x = [x_1, x_2, \dots, x_n]^{\mathsf{T}} \in \mathbb{R}^n$ to the quadratic function $x^{\mathsf{T}}x \in \mathbb{R}^{n\times n}$, we have

$$\vartheta :\, = \left[x_1^2, x_1x_2, \dots, x_1x_n, x_2^2, x_2x_3, \dots, x_n^2\right]^{\mathsf{T}}.$$

Therefore, its computation has space complexity $N(t_s)(n(n+1)/2)$, where t_s is the discretization step, $N(t_s) = \lfloor(\Delta_{k\,\max}/t_s)\rfloor - \lfloor(\Delta_{k\,\min}/t_s)\rfloor$, $\Delta_{k\,\max}$ is the maximum time the system is allowed to run in the open loop, and $\Delta_{k\,\min}$ is the minimum time on guaranteeing the time cost for one step computation. For the time complexity, define $t_c > 0$ as the time it takes to execute an instruction in a given digital platform, and the implementation requires $N(t_s)(n(n+1)/2)t_c$ multiplications and the same amount of additions in preprocessing step, as well as $N(t_s)t_c$ for testing $\dot{L}_2 \le -\lambda_{\min}(Q_0)\|x\|^2 + \Delta$ in the running step.

Now, turn to our concern. Noticing that the Lyapunov function candidate (26) consists of a quadratic function of x and an optimal cost function V^* (i.e., an integral of quadratic functions of x and u from the initial time t to infinity). Thus, the implementation of the developed self-triggered control policy (24) has a *space complexity* as

$$M_s = N(t_s)\left(\frac{n(n+1)}{2} + \frac{m(m+1)}{2} + 1\right)$$

since it requires the space of $N(t_s)(n(n+1)/2)$ for computing $x^{\mathsf{T}}x$, $N(t_s)(m(m+1)/2)$ for computing $u^{\mathsf{T}}Ru$, and $N(t_s)$ for storing the integral at t_k. Meanwhile, it has a *time complexity* as

$$M_t = (2n(n+1) + m(m+1) + N(t_s))t_c$$

since it requires preprocessing time as $N(t_s)n(n+1)t_c$ for $x^{\mathsf{T}}x$, $N(t_s)n(n+1)t_c$ for $x^{\mathsf{T}}Qx$ and $N(t_s)m(m+1)t_c$ for $u^{\mathsf{T}}Qu$ in the integral, and $N(t_s)t_c$ for testing the inequality $\dot{L}_2 \le -\lambda_{\min}(Q_0)\|x\|^2 + \Delta$ in the running step. It is worth pointing out that the minimum inter-execution time $\Delta_{k\,\min}$ is determined by M_t, and the maximum inter-execution time $\Delta_{k\,\max}$ is determined by guaranteeing $\dot{L}_2 = -\lambda_{\min}(Q_0)\|x\|^2 + \ \Delta$.

Remark 3: By introducing the self-triggering mechanism, this article presents a novel ADP-based optimal neuro-control scheme for nonlinear systems. The superiority of this approach is emphasized as follows: 1) it reduces the computational and communication resources, as well as energy cost since the control policy (24) is updated aperiodically and 2) the updating time instants of control policy are predicted by the designed self-triggered mechanism (35). It implies that the hardware monitoring the full system state in the event-triggered control structure is no longer required.

IV. Simulation Studies

In this section, we provide two simulation examples, including practical and numerical systems, to verify the present self-triggered approximate optimal neuro-control scheme via ADP (24) to be effective.

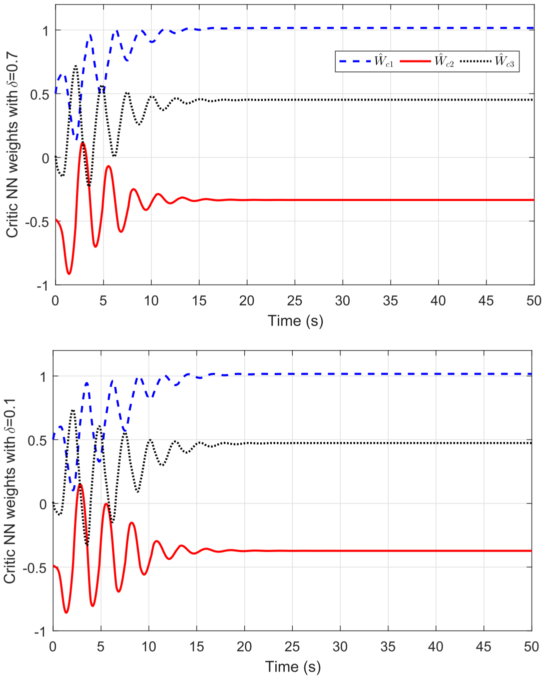

Fig. 1. Evolution of the critic NN weights for the torsional pendulum system.

A. Example 1

A torsional pendulum bar system is employed with the dynamics expressed as

$$\begin{cases} \dfrac{\mathrm{d}\theta}{\mathrm{d}t} = \omega \\ \mathcal{J}\dfrac{\mathrm{d}\omega}{\mathrm{d}t} = u - Mgl\sin\theta - f_d\dfrac{\mathrm{d}\theta}{\mathrm{d}t} \end{cases} \tag{44}$$

where $M = 1/3$ kg and $l = 2/3$ m indicate the mass and length of the pendulum bar, respectively. Let the rotary inertia be $\mathcal{J} = 4/3$ mL^2, the frictional factor be $f_d = 0.2$, and the gravitational acceleration be $g = 9.8$ m/s^2. Replacing the system state θ and ω by x_1 and x_2, the torsional pendulum bar system is expressed in the state-space form as

$$\dot{x} = \begin{bmatrix} x_2 \\ -\frac{Mgl}{\mathcal{J}}\sin x_1 - \frac{f_d}{\mathcal{J}}x_2 \end{bmatrix} + \begin{bmatrix} 0 \\ \frac{1}{\mathcal{J}} \end{bmatrix}u.$$

In the cost function, Q and R are selected as identity matrices with appropriate dimensions. Given the initial system state vector as $x_0 = [1, -1]^{\mathsf{T}}$, the activation function of critic NN is chosen as $\sigma(x) = [x_1^2, x_1x_2, x_2^2]^{\mathsf{T}}$, and the weight vector is defined as $\hat{W}_c = [\hat{W}_{c1}, \hat{W}_{c2}, \hat{W}_{c3}]^{\mathsf{T}}$ with initial value $\hat{W}_c^0 = [0.502, -0.489, 0.012]^{\mathsf{T}}$. It is worth pointing out that their selection depends on the experience according to [40] and [41]. The learning rates in the nested updating policies for the critic NN are chosen as $l_c = l_\varepsilon = 0.4$ and $\Gamma = 0.5$. In the self-triggered mechanism, based on the "trial and error," we choose the parameter $\lambda = 0.1$ and $\delta = 0.1, 0.3, 0.5, 0.7, 0.9, 1$ to test the sensitivity on the sampling frequency and show comparison results, respectively.

Taking $\delta = 0.7$ as the representative for explaining the simulation results, simulation results are provided in Figs. 1–5 and Table I. Fig. 1 illustrates the evolution of the critic NN weights. We can see that they converge to $[0.986, -0.371, 0.407]^{\mathsf{T}}$

TABLE I
SAMPLING TIMES WITH DIFFERENT VALUES OF PARAMETER δ

δ	Sampling times	$\Delta_{k\,\min}$(sec)	$\Delta_{k\,\max}$(sec)	Settling time (sec)
0.1	862	0.05	0.1	30.1
0.3	303	0.1	0.25	29.8
0.5	182	0.15	0.45	29.4
0.7	131	0.2	0.6	29.2
0.9	102	0.25	0.75	29.2
1	92	0.3	0.85	29.2

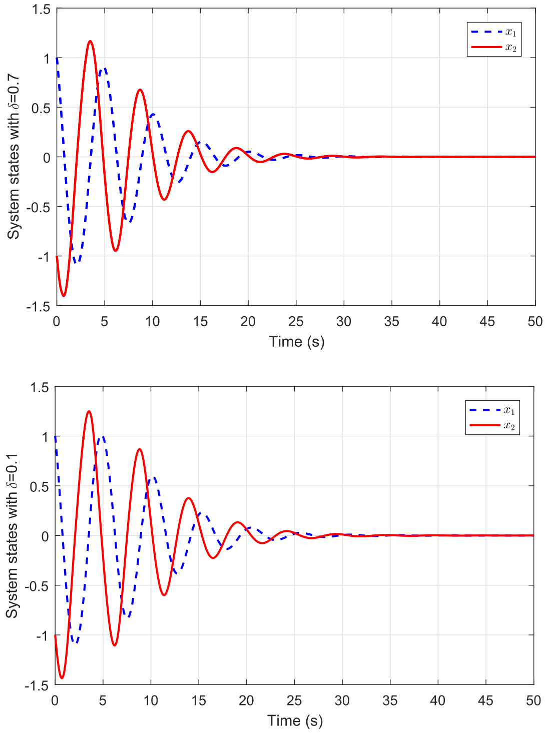

Fig. 2. State trajectories of the torsional pendulum system.

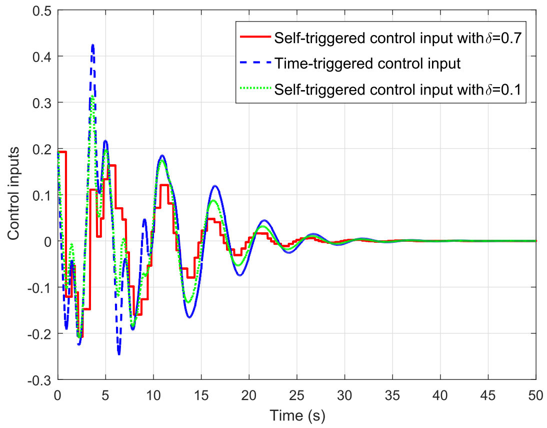

Fig. 3. Time- and self-triggered control inputs of the torsional pendulum system.

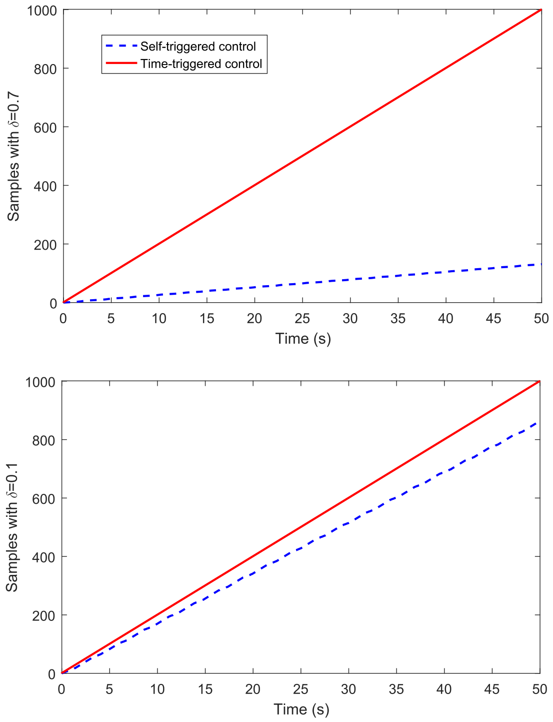

Fig. 4. Required samples of time- and self-triggered controls of the torsional pendulum system.

gradually. As illustrated in Fig. 2, by applying the developed self-triggered approximate optimal neuro-control policy computed by (24), the system states converge to zero after 29.2 s. Fig. 3 shows the self- and time-triggered control inputs. We can observe that the self-triggered approximate optimal neuro-control input is a piecewise continuous signal. It implies that the self-triggered approximate optimal neuro-control signal keeps unchanged during the time interval $[t_k, t_{k+1})$ and is updated at t_k only. Fig. 4 illustrates the comparison results for the numbers of acquired samples, and the time-triggered and the self-triggered control methods require 1000 and only 131 samples, respectively. It shows that the sampling frequency is greatly reduced. From Fig. 5, we find that the minimum inter-sampling time $\Delta_k = 0.2$ s, which implies that the Zeno behavior has not occurred. As a representative comparison, these figures also illustrate the control performance under the parameter chosen as $\delta = 0.1$. Together with Table I, where $\Delta_{k\,\min}$ and $\Delta_{k\,\max}$ denote the actual minimum and maximum inter-execution times, respectively, we can conclude that with the increase of δ, the number of required samples is decreased, and the system states take less settling time to approach the neighborhood of the equilibrium. It implies that we can choose a proper value according to the requirements of the transient response and resource cost.

B. Example 2

The overhead crane system, which transports loads from one place to another, plays an important role in industry. In contrast to Example 1, its dynamic model is more complicated and has a high order. The dynamic model of this system is formulated as that in [45], and the same parameters of the overhead crane plants are selected.

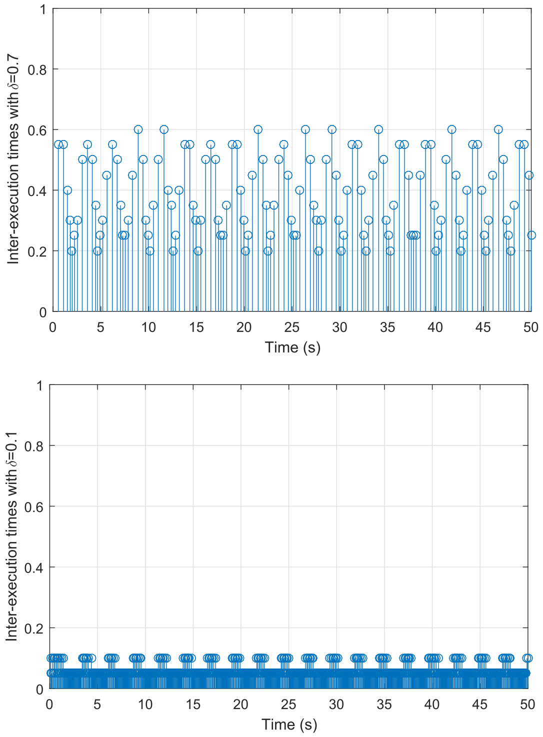

Fig. 5. Inter-execution times of the torsional pendulum system.

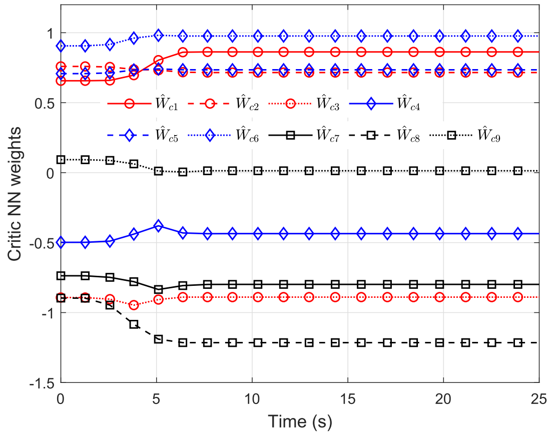

Fig. 6. Evolution of the critic NN weights of the overhead crane system.

In the simulation, Q and R are also selected as identity matrices with appropriate dimensions. The initial system state vector as $x_0 = [0.5, -0.5, 0.8, -0.9]^{\mathsf{T}}$. The activation function of critic NN is chosen as $\sigma(x) = [x_1^2, x_1x_2, x_1x_3, x_1x_4, x_2^2, x_2x_3, x_2x_4, x_3x_4, x_4^2]^{\mathsf{T}}$, and the weight vector is defined as $\hat{W}_c = [\hat{W}_{c1}, \hat{W}_{c2}, \ldots, \hat{W}_{c9}]^{\mathsf{T}}$ with initial value $\hat{W}_c^0 = [0.656, 0.759, -0.892, -0.497, 0.707, 0.905, -0.736, -0.896, 0.092]^{\mathsf{T}}$. The parameters in the nested updating policies are chosen as $l_c = 1.2$, $l_\varepsilon = 0.01$, and $\Gamma = 0.001$. The parameter in the self-triggered mechanism is chosen as $\lambda = 0.4$ and $\delta = 0.7$.

Simulation results of Example 2 are presented in Figs. 6–10. Fig. 6 describes the convergence process of the critic NN weights, we can observe that $\hat{W}_c$ converges to $[0.862, 0.715, -0.889, -0.436, 0.734, 0.976, -0.798, -1.214, 0.013]^{\mathsf{T}}$. As displayed in Fig. 7, the system states gradually converge to the equilibrium point after 13 s. Fig. 8 describes a piecewise signal of the self-triggered control, which is updated at t_k only, while the time-triggered control is a continuous signal. From Fig. 9, it shows that in contrast to the time-triggered control input, which requires 800 samples, the self-triggered one needs 320 samples only, which means that the sampling has been reduced by 60%. Fig. 10 shows that the minimum inter-sampling time $\Delta_{k\,\min} = 0.05$ s. Thus, we can declare that the developed self-triggered approximate optimal

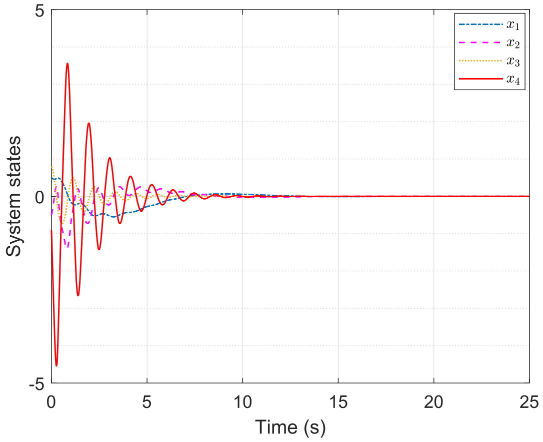

Fig. 7. State trajectories of the overhead crane system.

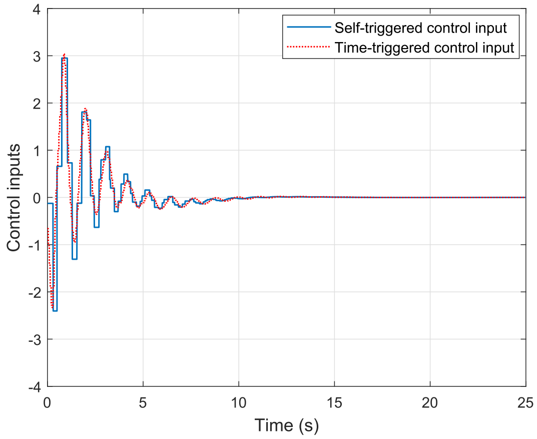

Fig. 8. Time- and self-triggered control inputs of the overhead crane system.

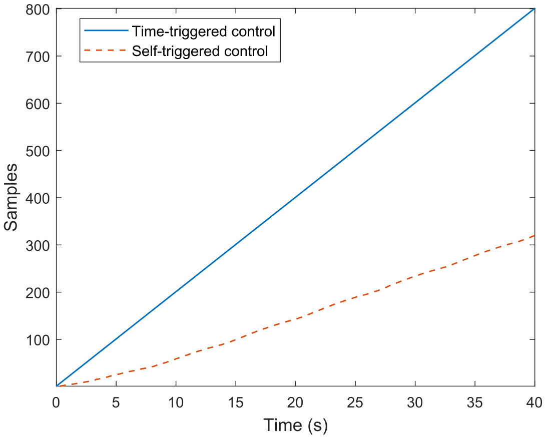

Fig. 9. Required samples of time- and self-triggered controls of the overhead crane system.

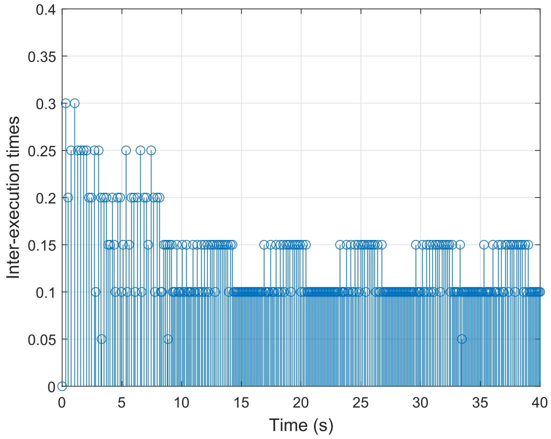

Fig. 10. Inter-execution times of the overhead crane system.

neuro-control scheme is effective to assure the closed-loop overhead crane system to be stable in a UUB manner.

V. Conclusion

A self-triggered approximate optimal neuro-control scheme is presented for nonlinear systems through ADP. By guaranteeing the asymptotic stability of the weight error dynamics, the critic NN is established to approximate the solution of the HJBE. Hereafter, the optimal neuro-control is indirectly derived in the ADP framework. By introducing the self-triggered mechanism, the updating time instants are predicted to determine the time updating the control policy. It is worth pointing out that a proper self-triggering condition, which predicts the next updating instant of the control policy, is designed to avoid the continuous monitoring of the system state by hardware devices in the event-triggered control approach, and the computation, the communication, and the energy consumption are decreased in an alternative way. Furthermore, the nested updating policies guarantee the asymptotic stability of the critic weight error dynamics, rather than UUB in most existing ADP-based optimal control methods.

References

[1] R. E. Bellman, *Dynamic Programming*. Princeton, NJ, USA: Princeton Univ. Press, 1957.

[2] P. J. Werbos, "Approximate dynamic programming for real time control and neural modeling," in *Handbook of Intelligent Control: Neural, Fuzzy, and Adaptive Approaches*, D. A. White and D. A. Sofge, Eds. New York: Van Nostrand Reinhold, 1992, ch. 13.

[3] D. Liu, Q. Wei, D. Wang, X. Yang, and H. Li, *Adaptive Dynamic Programming With Applications in Optimal Control*. Cham, Switzerland: Springer, 2017.

[4] D. Liu, S. Xue, B. Zhao, B. Luo, and Q. Wei, "Adaptive dynamic programming for control: A survey and recent advances," *IEEE Trans. Syst. Man, Cybern., Syst.*, vol. 51, no. 1, pp. 142–160, Jan. 2021.

[5] A. Al-Tamimi, F. L. Lewis, and M. Abu-Khalaf, "Discrete-time nonlinear HJB solution using approximate dynamic programming: Convergence proof," *IEEE Trans. Syst. Man, Cybern., Syst.*, vol. 38, no. 4, pp. 943–949, Aug. 2008.

[6] M. Ballesteros, I. Chairez, and A. Poznyak, "Robust min–max optimal control design for systems with uncertain models: A neural dynamic programming approach," *Neural Netw.*, vol. 125, pp. 153–164, May 2020.

[7] S. Bhasin, R. Kamalapurkar, M. Johnson, K. G. Vamvoudakis, F. L. Lewis, and W. E. Dixon, "A novel actor-critic-identifier architecture for approximate optimal control of uncertain nonlinear systems," *Automatica*, vol. 49, no. 1, pp. 82–92, 2013.

[8] M. Ballesteros, I. Chairez, and A. Poznyak, "Robust optimal feedback control design for uncertain systems based on artificial neural network approximation of the Bellman's value function," *Neurocomputing.*, vol. 413, pp. 134–144, Nov. 2020.

[9] B. Fan, Q. Yang, X. Tang, and Y. Sun, "Robust ADP design for continuous-time nonlinear systems with output constraints," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 6, pp. 2127–2138, Jun. 2018.

[10] H. Zhang, Y. Liu, G. Xiao, and H. Jiang, "Data-based adaptive dynamic programming for a class of discrete-time systems with multiple delays," *IEEE Trans. Syst. Man, Cybern., Syst.*, vol. 50, no. 2, pp. 432–441, Feb. 2020.

[11] B. Zhao, D. Liu, and Y. Li, "Observer based adaptive dynamic programming for fault tolerant control of a class of nonlinear systems," *Inf. Sci.*, vol. 384, pp. 21–33, Apr. 2017.

[12] B. Zhao and D. Liu, "Event-triggered decentralized tracking control of modular reconfigurable robots through adaptive dynamic programming," *IEEE Trans. Ind. Electron.*, vol. 67, no. 4, pp. 3054–3064, Apr. 2020.

[13] J. Zhao, J. Na, and G. Gao, "Robust tracking control of uncertain nonlinear systems with adaptive dynamic programming," *Neurocomputing*, vol. 471, pp. 21–30, Jan. 2022.

[14] S. A. A. Rizvi and Z. Lin, "Output feedback adaptive dynamic programming for linear differential zero-sum games," *Automatica*, vol. 122, Dec. 2020, Art. no. 109272.

[15] D. Liu, Y. Xu, Q. Wei, and X. Liu, "Residential energy scheduling for variable weather solar energy based on adaptive dynamic programming," *IEEE/CAA J. Autom. Sinica*, vol. 5, no. 1, pp. 36–46, Jan. 2018.

[16] A. De Keyser, H. Vansompel, and G. Crevecoeur, "Real-time energy-efficient actuation of induction motor drives using approximate dynamic programming," *IEEE Trans. Ind. Electron.*, vol. 68, no. 12, pp. 11837–11846, Dec. 2021.

[17] X. Yang, H. He, and X. Zhong, "Adaptive dynamic programming for robust regulation and its application to power systems," *IEEE Trans. Ind. Electron.*, vol. 65, no. 7, pp. 5722–5732, Jul. 2018.

[18] P. Zeng, H. Li, H. He, and S. Li, "Dynamic energy management of a microgrid using approximate dynamic programming and deep recurrent neural network learning," *IEEE Trans. Smart Grid*, vol. 10, no. 4, pp. 4435–4445, Jul. 2019.

[19] C. Dong, C. Liu, Q. Wang, and L. Gong, "Switched adaptive active disturbance rejection control of variable structure near space vehicles based on adaptive dynamic programming," *Chin. J. Aeronaut.*, vol. 32, no. 7, pp. 1684–1694, Jul. 2019.

[20] X. Chen, Y. Huang, J. Na, G. Gao, and J. Zhao, "Adaptive optimal control of nonlinear active suspension systems with completely unknown dynamics," in *Proc. 33rd Chin. Control Decis. Conf. (CCDC)*, Kunming, China, May 2021, pp. 3524–3529.

[21] L. Dong, X. Zhong, C. Sun, and H. He, "Event-triggered adaptive dynamic programming for continuous-time systems with control constraints," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 8, pp. 1941–1952, Aug. 2017.

[22] M. Kishida, M. Kögel, and R. Findeisen, "Combined event- and self-triggered control approach with guaranteed finite-gain $^{\ell}2$ stability for uncertain linear systems," *IET Control Theory Appl.*, vol. 11, no. 11, pp. 1674–1683, Jul. 2017.

[23] T. Zhou, Z. Zuo, and Y. Wang, "Self-triggered and event-triggered control for linear systems with quantization," *IEEE Trans. Syst. Man, Cybern., Syst.*, vol. 50, no. 9, pp. 3136–3144, Sep. 2020.

[24] Y. Qi, P. Zeng, and W. Bao, "Event-triggered and self-triggered H_∞ control of uncertain switched linear systems," *IEEE Trans. Syst. Man, Cybern. Syst.*, vol. 50, no. 4, pp. 1442–1454, Apr. 2020.

[25] A. Sahoo, V. Narayanan, and S. Jagannathan, "A min–max approach to event- and self-triggered sampling and regulation of linear systems," *IEEE Trans. Ind. Electron.*, vol. 66, no. 7, pp. 5433–5440, Jul. 2019.

[26] Z. Q. Zuo, Q. S. Li, H. C. Li, and Y. J. Wang, "Event-triggered and self-triggered control for linear systems with actuator saturation," *Trans. Inst. Meas. Control.*, vol. 40, no. 4, pp. 1281–1288, 2018.

[27] A. Anta and P. Tabuada, "To sample or not to sample: Self-triggered control for nonlinear systems," *IEEE Trans. Autom. Control*, vol. 55, no. 9, pp. 2030–2042, Sep. 2010.

[28] K. Zhang, B. Zhou, H. Jiang, and G. Duan, "Event-triggered and self-triggered gain scheduling control of input constrained systems with applications to the spacecraft rendezvous," *Int. J. Robust Nonlinear Control*, vol. 31, no. 10, pp. 4629–4646, Jul. 2021.

[29] L. Lu and J. M. Maciejowski, "Self-triggered MPC with performance guarantee using relaxed dynamic programming," *Automatica*, vol. 114, Apr. 2020, Art. no. 108803.

[30] F. D. Brunner, W. P. M. H. Heemels, and F. Allgöwer, “Event-triggered and self-triggered control for linear systems based on reachable sets,” *Automatica*, vol. 101, pp. 15–26, Mar. 2019.

[31] C. Liu, H. Li, J. Gao, and D. Xu, “Robust self-triggered min–max model predictive control for discrete-time nonlinear systems,” *Automatica*, vol. 89, pp. 333–339, Mar. 2018.

[32] H. Li and X. Li, “Distributed model predictive consensus of heterogeneous time-varying multi-agent systems: With and without self-triggered mechanism,” *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 67, no. 12, pp. 5358–5368, Dec. 2020.

[33] Y. Fan, L. Liu, G. Feng, and Y. Wang, “Self-triggered consensus for multi-agent systems with zeno-free triggers,” *IEEE Trans. Autom. Control*, vol. 60, no. 10, pp. 2779–2784, Oct. 2015.

[34] Y. Gao, G. Sun, J. Liu, Y. Shi, and L. Wu, “State estimation and self-triggered control of CPSs against joint sensor and actuator attacks,” *Automatica*, vol. 113, Mar. 2020, Art. no. 108687.

[35] Y. Gao, P. Yu, D. V. Dimarogonas, K. H. Johansson, and L. Xie, “Robust self-triggered control for time-varying and uncertain constrained systems via reachability analysis,” *Automatica*, vol. 107, pp. 574–581, Sep. 2019.

[36] Q. Cao, Z. Sun, Y. Xia, and L. Dai, “Self-triggered MPC for trajectory tracking of unicycle-type robots with external disturbance,” *J. Franklin Inst.*, vol. 356, no. 11, pp. 5593–5610, Jul. 2019.

[37] L. Zhou and P. Tokekar, “Active target tracking with self-triggered communications in multi-robot teams,” *IEEE Trans. Autom. Sci. Eng.*, vol. 16, no. 3, pp. 1085–1096, Jul. 2019.

[38] X. Lou and Z. Ji, “Self-triggering adaptive optimal control for nonlinear systems based on encoding mechanism,” *Math. Comput. Simul.*, vol. 190, pp. 1027–1047, Dec. 2021.

[39] K. Kobayashi and K. Hiraishi, “Optimization-based synthesis of self-triggered controllers for networked systems,” *IEICE Trans. Fundam. Electron., Commun. Comput. Sci.*, vol. E95.A, no. 4, pp. 691–696, 2012.

[40] R. W. Beard, G. N. Saridis, and J. T. Wen, “Galerkin approximations of the generalized Hamilton-Jacobi-Bellman equation,” *Automatica*, vol. 33, no. 12, pp. 2159–2177, Dec. 1997.

[41] M. Abu-Khalaf and F. L. Lewis, “Nearly optimal control laws for nonlinear systems with saturating actuators using a neural network HJB approach,” *Automatica*, vol. 41, pp. 779–791, May 2005.

[42] A. Lamperski and A. D. Ames, “Lyapunov theory for Zeno stability,” *IEEE Trans. Autom. Control*, vol. 58, no. 1, pp. 100–112, Jan. 2013.

[43] M. Mazo, A. Anta, and P. Tabuada, “On self-triggered control for linear systems: Guarantees and complexity,” in *Proc. Eur. Control Conf. (ECC)*, Budapest, Hungary, Aug. 2009, pp. 3767–3772.

[44] W. P. M. H. Heemels, K. H. Johansson, and P. Tabuada, “An introduction to event-triggered and self-triggered control,” in *Proc. IEEE 51st IEEE Conf. Decis. Control (CDC)*, Dec. 2012, pp. 3270–3285.

[45] D. Wang, H. He, and D. Liu, “Adaptive critic nonlinear robust control: A survey,” *IEEE Trans. Cybern.*, vol. 47, no. 10, pp. 3429–3451, Oct. 2017.

Bo Zhao (Senior Member, IEEE) received the B.S. degree in automation and the Ph.D. degree in control science and engineering from Jilin University, Changchun, China, in 2009 and 2014, respectively.

He was a Post-Doctoral Fellow with the State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing, China, from 2014 to 2017, where he was an Assistant Professor from 2017 to 2018. He is currently an Associate Professor with the School of Systems Science, Beijing Normal University, Beijing. He was selected as the Beijing Normal University Tang Scholar in 2021. He has authored or coauthored over 140 journal articles and conference papers, and authorized five patents. His research interests include adaptive dynamic programming, reinforcement learning, optimal control, and robot control.

Mr. Zhao is the Secretary General of the Adaptive Dynamic Programming and Reinforcement Learning Technical Committee of Chinese Association of Automation (CAA) and a member of the Asian-Pacific Neural Network Society (APNNS) and the Chinese Association for Artificial Intelligence (CAAI). He serves as an Associate Editor for IEEE TRANSACTIONS ON SYSTEMS, MAN AND CYBERNETICS: SYSTEMS, *Neurocomputing*, and *Artificial Intelligence Review*, an Early Career Advisory Board Member for IEEE/CAA JOURNAL OF AUTOMATICA SINICA, and a Guest Editor for *Complex & Intelligent Systems*, and also served many academic conferences.

Shunchao Zhang received the B.S. degree in measurement and control technology and instruments from the School of Electrical and Information Engineering, Hunan Institute of Engineering, Xiangtan, China, in 2016, and the M.S. degree in control engineering and the Ph.D. degree in control science and engineering from the School of Automation, Guangdong University of Technology, Guangzhou, China, in 2019 and 2022, respectively.

He is currently a Lecturer with the School of Internet Finance and Information Engineering, Guangdong University of Finance, Guangzhou. His current research interests include optimal control and adaptive dynamic programming.

Derong Liu (Fellow, IEEE) received the Ph.D. degree in electrical engineering from the University of Notre Dame, Notre Dame, IN, USA, in 1994.

He became a Full Professor of electrical and computer engineering and computer science at the University of Illinois at Chicago, Chicago, IL, USA, in 2006. He was selected for the 100 Talents Program by the Chinese Academy of Sciences in 2008, and he served as the Associate Director of the State Key Laboratory of Management and Control for Complex Systems at the Institute of Automation from 2010 to 2016. He has published 13 books.

Mr. Liu is a fellow of the International Neural Network Society and the International Association of Pattern Recognition and a member of Academia Europaea (the Academy of Europe). He received the International Neural Network Society's Gabor Award in 2018 and the IEEE CIS Neural Network Pioneer Award in 2022. He has been named a Highly Cited Researcher by Clarivate since 2017. He was the Editor-in-Chief of IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS from 2010 to 2015. He is the Editor-in-Chief of *Artificial Intelligence Review* (Springer).