International Journal of Innovative Computing, Information and Control
Volume **21**, Number **3**, June **2025**

ICIC International ©2025 ISSN 1349-4198
pp. **769–779**

EVENT-TRIGGERED AND SELF-TRIGGERED APPROXIMATE OPTIMAL CONTROL FOR CONSTRAINED-INPUT NONLINEAR SYSTEMS VIA REINFORCEMENT LEARNING

Pengda Liu[1], Huiyan Zhang[1,2,*] and Wengang Ao[1]

[1]National Research Base of Intelligent Manufacturing Service
Chongqing Technology and Business University
No. 19, Xuefu Avenue, Nan'an District, Chongqing 400067, P. R. China
liupengda36@126.com; aowg@ctbu.edu.cn
*Corresponding author: huiyanzhang@ctbu.edu.cn

[2]Chongqing Innovation Center of Industrial Big-Data Company Ltd.
No. 11, Tongxing Road, Beibei District, Chongqing 400707, P. R. China

Received August 2024; revised December 2024

Abstract. *This paper investigates the event-triggered and self-triggered approximate optimal control issues of nonlinear systems using reinforcement learning approach. First, the optimal control issue with asymmetric input constraints is formulated, via constructing the cost function with nonquadratic utility function. Then, the reinforcement learning method which is implemented by critic neural network is proposed to approximate the optimal control schemes. To further reduce the consumption of computation and communication resources, the dynamic triggering mechanism and self-triggered mechanism are integrated into the adaptive learning framework. Different from the common triggering mechanism, the dynamic triggering mechanism with the preliminary operation is able to reduce the superfluous triggering events when the control precision is achieved. Furthermore, based on the predicted triggering instant, the self-triggered mechanism is designed as an active triggering mechanism instead of the passive one. The stability of the closed-loop system is demonstrated with Lyapunov theorem. In the end, the simulation results validate the effectiveness of the aperiodic triggering adaptive control approach.*

Keywords: Adaptive dynamic programming, Reinforcement learning, Optimal control, Event-triggered control, Self-triggered control

1. **Introduction.** Optimal control theory has been widely applied in multiple fields including power systems, wastewater treatment process, hypersonic vehicle, and so forth [1, 2, 3]. The goal of solving optimal control issues (OCIs) is to acquire a stabilizing control scheme that drives system states to zero while optimizing the performance index. For nonlinear continuous-time systems, seeking the minimization for performance index can be equivalent to solving the Hamilton-Jacobi-Bellman equation (HJBE). Generally, it is troublesome to derive the closed-form solution for HJBE since there exist the nonlinear terms and partial derivatives. Fortunately, by virtue of the powerful optimization ability, learning algorithms can provide alternative ideas to approximately derive the solutions. Among these learning methods, reinforcement learning (RL) and adaptive dynamic programming (ADP) often share the same characteristics and can be deemed as equivalent in optimal control fields [4, 5, 6]. In recent years, RL or ADP methods have been investigated to tackle different types of OCIs. In [7], two-player nonlinear Stackelberg differential game with unknown system dynamics was solved via an off-policy integral reinforcement

DOI: 10.24507/ijicic.21.03.769

learning (IRL) technique. Through using a neural network-based ADP method, an approximate optimal control law was developed in [8] to achieve the tracking control of continuous-time systems with unmatched uncertainties. In [9], based on value iteration scheme, an accelerated learning algorithm with convergence guarantee was developed to deal with the OCIs of nonlinear systems.

Traditional time-driven method often requires large amounts of data processing and transmission to drive control signals to be updated in succession, incurring the excessive consumptions for the computation/communication resources of the systems [10, 11]. To overcome this shortcoming without affecting system stability, the event-triggered mechanism (ETM) was introduced [12, 13, 14]. For example, via designing reasonable triggering conditions, the event-driven ADP methods were proposed to achieve the approximate optimal solutions while the control strategies were updated with only necessary data, alleviating the computation burdens and communication bandwidths [15, 16]. The online event-triggered adaptive method using experience replay technique was proposed in [17] to solve the nonzero-sum tracking games of the two-player nonlinear system, relaxing the dependence on the persistence of excitation conditions. In [18], the generalized fuzzy hyperbolic models were designed to approximate the unknown system dynamics and the event-based ADP method was constructed to deal with the non-zero-sum games issues. In [19], by introducing the internal dynamic variable, the dynamic event-triggered mechanism (DETM) was integrated into the adaptive learning algorithm to design the optimal control law. The reinforcement learning was proposed in [20] to tackle the Pareto optimization issues under the dynamic triggering control mechanism. In [21], the dynamic event-triggered robust control method was developed for nonlinear systems by utilizing iterative neural dynamic programming. Generally, compared with the traditional event-based adaptive learning methods, the dynamic triggering control approaches can further reduce the triggering frequency while guaranteeing the achievement of the control target. However, the triggering mechanism of the existing works often operates even when the control precision has been achieved, resulting in additional waste of system resources. In addition, when the ETM works, it requires the hardware equipment to monitor system states persistently. To solve this issue, a kind of self-triggered mechanism (STM) was constructed in [22] and integrated into adaptive learning framework. Nevertheless, the relevant researches are still insufficient. In view of these issues, it is necessary to improve the effectiveness of RL methods, especially for solving the OCIs of the nonlinear systems with asymmetric input constraints.

Motivated by the above works, we investigate the OCIs for constrained systems with incomplete known dynamics. Compared with the aperiodic triggering works in [18, 19, 22], the proposed adaptive learning approach, which can integrate the improved DETM with preliminary operation and the STM, is able to approximately derive the optimal control scheme for the systems with asymmetric input constraints. The DETM with preliminary operation can further reduce the triggering frequency and the STM can save system resources while averting monitoring behaviors for system states. The contributions are listed as follows. First, different from the previous works in [16, 17, 21], the IRL approach of the single-critic network architecture is utilized to derive the approximate solution of HJBE for the systems with asymmetric input constraints and unknown drift dynamics. Then, the DETM with preliminary operation is constructed and integrated into the IRL algorithm to further reduce the needless triggering events while guaranteeing the boundedness of signals in the closed-loop system. Finally, the self-triggered adaptive control mechanism is developed to achieve aperiodic triggering control while avoiding monitoring behaviors of hardware equipment.

The remainder of this work is organized as follows. The formulation of the OCIs is provided in Section 2. The aperiodic triggering control mechanism and the IRL algorithm are constructed in Section 3. Section 4 provides the simulation results and the associated analysis. The conclusions and future research trend are presented in Section 5.

2. **Problem Statement and Preliminaries.** Consider the nonlinear system

$$\dot{x} = \mathfrak{F}(x) + \mathfrak{G}(x)\mathfrak{u}, \tag{1}$$

where $x \in \Omega \subset \mathbb{R}^n$, $\mathfrak{u} \in \mathbb{R}^m$ and the dynamics of $\mathfrak{F}(x)$ is unknown. The control input is constrained with asymmetric bounds, that is, $\mathfrak{b}_\alpha \leq \mathfrak{u}_k \leq \mathfrak{b}_\gamma$ where $|\mathfrak{b}_\alpha| \neq |\mathfrak{b}_\gamma|$ and $k = 1, 2, \ldots, m$.

Assumption 2.1. *It is supposed that $\mathfrak{F}(x)$ is locally Lipschitz continuous and $\mathfrak{G}(x)$ is bounded. That is, $\|\mathfrak{F}(x)\| \leq \mathfrak{L}_{\mathcal{F}}\|x\|$ and $\|\mathfrak{G}(x)\| \leq \mathcal{D}$. Herein, $\mathfrak{L}_{\mathcal{F}}$ and $\mathcal{D}$ are both positive constants.*

The cost function is defined as

$$\mathfrak{V}(x) = \int_t^\infty \Upsilon(x(\iota), \mathfrak{u}(\iota))d\iota, \tag{2}$$

where $\Upsilon(x, \mathfrak{u}) = x^T \mathcal{Q} x + \sigma(\mathfrak{u})$. Herein, $\sigma(\mathfrak{u}) = 2\sum_{\ell=1}^m \int_p^{\mu_\ell} q\eta^{-1}\left(\frac{\mu_\ell - p}{q}\right) d\mu_\ell$, $p = (\mathfrak{b}_\alpha + \mathfrak{b}_\gamma)/2$, and $q = (\mathfrak{b}_\gamma - \mathfrak{b}_\alpha)/2$. In this paper, the function $\eta(\cdot)$ is designated as $\tanh(\cdot)$.

The optimal cost function is given by

$$\mathfrak{V}^*(x) = \min_{\mathfrak{u}} \int_t^\infty \Upsilon(x, \mathfrak{u})d\iota. \tag{3}$$

From (3) the Hamiltonian can be derived

$$\mathcal{H}\left(x, \nabla\mathfrak{V}^*(x), \mathfrak{u}\right) = \Upsilon(x, \mathfrak{u}) + (\nabla\mathfrak{V})^{*T}\left(\mathfrak{F}(x) + \mathfrak{G}(x)\mathfrak{u}\right). \tag{4}$$

Based on Bellman principle of optimality, the optimal control input can be derived

$$\mathfrak{u}^* = q\tanh\left(\mathcal{K}^*\right) + \vartheta, \tag{5}$$

where $\mathcal{K}^* = -1/(2q)\mathfrak{G}(x)^T\nabla\mathfrak{V}^*(x)$ and $\vartheta = [p, p, \ldots, p]^T \in \mathbb{R}^m$. Then the HJBE can be derived

$$0 = \nabla\mathfrak{V}^{*T}\left(\mathfrak{F}(x) + \mathfrak{G}(x)\vartheta\right) + x^T\mathcal{Q}x + \sigma(\mathfrak{u}^*) + q\nabla\mathfrak{V}^{*T}\mathfrak{G}(x)\tanh(\mathcal{K}^*). \tag{6}$$

It is usually necessary to solve Equation (6) to acquire the optimal control (5). Nevertheless, the closed-form solution of this partial differential equation is difficult to derive due to the nonlinear terms. To approximately solve this equation, the RL method combined with aperiodic triggering mechanism is constructed in the following sections.

3. **Aperiodic Triggering Control Using Integral Reinforcement Learning Method.** In this section, the IRL method integrating the aperiodic triggering mechanism is proposed such that the optimal control for the systems with incomplete known dynamics can be approximated and the resource consumption for the system can be reduced.

3.1. **The construction of aperiodic triggering mechanism.** Differing from the traditional time-driven mechanism updating the control law continuously, the aperiodic triggering mechanism can drive the control law to be computed/updated aperiodically [18]. The dynamic event-triggered and self-triggered mechanisms both belong to aperiodic triggering mechanisms. The former one relies on the operation of the event-triggered condition to determine the next triggering instant, while the latter can actively predict the triggering instant utilizing the self-triggered condition, saving system resources and avoiding the continuous monitoring for the system states.

To construct the aperiodic triggering mechanism, the time sequence $\{\zeta_h\}_{h=1}^{\infty}$ is indispensable to record triggering instants. Then the triggering error is derived

$$e_h(t) = \breve{x}_h - x(t), \quad t \in [\zeta_h, \zeta_{h+1}). \tag{7}$$

Furthermore, based on the aperiodic instants one can update the control law

$$\breve{\mathfrak{u}}^*(t) = q \tanh\left(\breve{\mathfrak{K}}^*\right) + \vartheta, \quad t \in [\zeta_h, \zeta_{h+1}), \tag{8}$$

where $\breve{\mathfrak{K}}^* = -1/(2q)\mathfrak{G}^T(\breve{x}_h)\nabla\breve{\mathfrak{V}}^*$ and $\nabla\breve{\mathfrak{V}}^* = \nabla\mathfrak{V}^*(\breve{x}_h)$.

According to (8), one can derive that the control signal is updated only at the triggering instants. During the adjacent triggering instants, the control law remains unchanged with the utilization of zero-order holder.

3.2. **The event-triggered adaptive control scheme using critic learning approach.** Via neural network, we can construct the optimal cost function

$$\mathfrak{V}^*(x) = \omega_c^T \delta(x) + \epsilon(x), \tag{9}$$

where $\omega_c \in \mathbb{R}^{\mathcal{N}_e}$, $\delta(x) \in \mathbb{R}^{\mathcal{N}_e}$, $\epsilon(x) \in \mathbb{R}$, and $\mathcal{N}_e$ respectively denote ideal weight vector, activation function, reconstruction error, and neuro number. The cost function can be approximated that

$$\hat{\mathfrak{V}}(x) = \hat{\omega}_c^T \delta(x), \tag{10}$$

where $\hat{\omega}_c$ is the approximate weight.

Then one can acquire the event-triggered version of optimal control

$$\breve{\mathfrak{u}}^* = q \tanh\left(\breve{\mathcal{K}}^*\right) + \vartheta + \varsigma(\breve{x}_h), \tag{11}$$

where $\breve{\mathcal{K}}^* = -1/(2q)\mathfrak{G}^T(\breve{x}_h)(\nabla\delta(\breve{x}_h))^T\omega_c$, $\varsigma(\breve{x}_h) = -\frac{1}{2}(I_m - \mathcal{C}(\xi(\breve{x}_h)))\mathfrak{G}^T(\breve{x}_h)\nabla\epsilon(\breve{x}_h)$. Herein, $\mathcal{C}(\xi(\breve{x}_h)) = diag\left\{\tanh^2(\xi_k(\breve{x}_h))\right\}$, $k = 1, \ldots, m$. $\xi(\breve{x}_h) = [\xi_1(\breve{x}_h), \ldots, \xi_m(\breve{x}_h)]^T$ is selected between $1/(2q)\mathfrak{G}^T(\breve{x}_h)\nabla\breve{\mathfrak{V}}^*$ and $-\breve{\mathcal{K}}^*$. Similarly, we have

$$\breve{\mathfrak{u}} = q \tanh\left(\breve{\mathcal{K}}\right) + \vartheta \tag{12}$$

with $\breve{\mathcal{K}} = -1/(2q)\mathfrak{G}^T(\breve{x}_h)(\nabla\delta(\breve{x}_h))^T\hat{\omega}_c$.

To overcome the difficulties incurred by the unknown drift dynamics, the IRL technique derived from RL algorithm is utilized which works based on the optimal cost function involving the integral of function $\Upsilon(x, \breve{\mathfrak{u}})$ on the fixed time interval. Thus, the Hamiltonian can be reformulated as

$$\mathcal{H}\left(x, \breve{\mathfrak{u}}, \hat{\mathfrak{V}}\right) = \int_{t-\tau}^{t} \Upsilon(x, \breve{\mathfrak{u}})d\iota + \hat{\mathfrak{V}}(x(t)) - \hat{\mathfrak{V}}(x(t-\tau)) = \hat{\omega}_c^T\theta(x) + \Xi(x, \breve{\mathfrak{u}}) \triangleq e_H, \tag{13}$$

where $\theta(x) = \delta(x(t)) - \delta(x(t-\tau))$, $\tau \in (0, t)$, and $\Xi(x, \breve{\mathfrak{u}}) = \int_{t-\tau}^{t} \Upsilon(x, \breve{\mathfrak{u}})d\iota$. Let $e_I = \omega_c^T\theta + \Xi$ and $\tilde{\omega}_c = \omega_c - \hat{\omega}_c$, then $e_H = e_I - \tilde{\omega}_c^T\theta$. The training target for reinforcement learning is set as $\mathcal{T} = 1/2e_H$, and then with the utilization of gradient descent approach the weight update law is derived

$$\dot{\hat{\omega}}_c = -\ell\breve{\theta}e_H \tag{14}$$

with $\breve{\theta} = \theta/\left(\theta^T\theta + 1\right)^2$. Accordingly, the weight error can be updated according to the following equation

$$\dot{\tilde{\omega}}_c = \ell\breve{\theta}e_I - \ell\breve{\theta}\bar{\theta}^T\tilde{\omega}_c \tag{15}$$

with $\bar{\theta} = \theta/\left(\theta^T\theta + 1\right)$.

Two reasonable assumptions are necessary [5, 6, 18].

Assumption 3.1. *The optimal control* $\mathfrak{u}^*(t)$ *satisfies that* $\|\mathfrak{u}^*(t) - \breve{\mathfrak{u}}^*(t)\|^2 \leq \varrho\|x - \breve{x}_h\|^2$.

Assumption 3.2. *The terms* ω_c, $\nabla\delta(x)$, $\nabla\epsilon(x)$, $\varsigma(x)$, e_I *are respectively norm-bounded with* $\mathfrak{b}_\omega$, $\mathfrak{b}_\delta$, $\mathfrak{b}_\epsilon$, $\mathfrak{b}_\varsigma$, *and* $\mathfrak{b}_I$.

In order to further avoid superfluous events, the preliminary operation is designed

$$\Gamma(e_h) = \begin{cases} \|e_h\|^2, & \text{if } \|x\| > T_p; \\ 0, & \text{otherwise.} \end{cases} \tag{16}$$

Theorem 3.1. *Consider the system* (1) *with control input* (12). *Let Assumptions 2.1, 3.1 and 3.2 hold and the critic weights be updated with* (14). *The triggering instant is determined by the following condition*

$$\Gamma(e_h) \leq \frac{(1-\alpha_c)(1-\gamma_c)\lambda_{\min}(\mathcal{Q})\|\breve{x}_h\|^2}{2\varrho + \left(\frac{1}{\alpha_c} - 1\right)\lambda_{\max}(\mathcal{Q})} \triangleq \mathcal{T}_I, \tag{17}$$

where $\alpha_c \in (0,1)$ *and* $\gamma_c \in (0,1)$ *are both triggering coefficients. Then system states and critic weight error for the closed-loop system are uniformly ultimately bounded (UUB).*

Proof: The Lyapunov candidate is selected as

$$\mathcal{L}_Y = \mathfrak{V}^*(\breve{x}_h) + \mathfrak{V}^*(x) + \frac{1}{2}\tilde{\omega}_c^T\ell^{-1}\tilde{\omega}_c = \mathcal{L}_a + \mathcal{L}_b + \mathcal{L}_c. \tag{18}$$

Considering the influences caused by triggering mechanism, the discussion will be divided into two aspects.

Situation i. No event occurs, i.e., $t \in [\zeta_h, \zeta_{h+1})$. In this stage, $\dot{\mathcal{L}}_a = 0$. According to (5), (6) and (12), the derivative of $\mathcal{L}_b$ satisfies that

$$\begin{aligned} \dot{\mathcal{L}}_b &= (\nabla\mathfrak{V}^*)^T(\mathfrak{F}(x) + \mathfrak{G}(x)\breve{\mathfrak{u}}) \\ &= -x^T\mathcal{Q}x - \sigma(\mathfrak{u}^*) - 2q\left(\tanh^{-1}\left(\frac{\mathfrak{u}^* - \vartheta}{q}\right)\right)^T(\breve{\mathfrak{u}} - \mathfrak{u}^*) \\ &\leq -x^T\mathcal{Q}x - \sigma(\mathfrak{u}^*) + \|\breve{\mathfrak{u}} - \mathfrak{u}^*\|^2 + q^2\sum_{k=1}^{m}\left(\tanh^{-1}\left(\frac{\mathfrak{u}_k^* - p}{q}\right)\right)^2. \end{aligned} \tag{19}$$

Through utilizing variable substitution approach, it follows that

$$\begin{aligned} \dot{\mathcal{L}}_b &\leq -x^T\mathcal{Q}x + \|\breve{\mathfrak{u}} - \mathfrak{u}^*\|^2 + 2q^2\sum_{k=1}^{m}\int_0^{\tanh^{-1}\left(\frac{\mathfrak{u}_k^* - p}{q}\right)} \upsilon_k\tanh^2(\upsilon_k)d\upsilon_k \\ &\leq -x^T\mathcal{Q}x + \|\breve{\mathfrak{u}} - \mathfrak{u}^*\|^2 + \mathfrak{b}_\upsilon, \end{aligned} \tag{20}$$

where $\mathfrak{b}_\upsilon$ is the upper bound of the term $2q^2\sum_{k=1}^{m}\int_0^{\tanh^{-1}\left(\frac{\mathfrak{u}_k^* - p}{q}\right)} \upsilon_k\tanh^2(\upsilon_k)d\upsilon_k$.

It is noted that $\|c + d\|^2 \leq 2\|c\|^2 + 2\|d\|^2$. Furthermore, considering the properties of $\tanh(\cdot)$, we derive

$$\begin{aligned} \|\breve{\mathfrak{u}} - \mathfrak{u}^*\|^2 &\leq 2\|\breve{\mathfrak{u}} - \breve{\mathfrak{u}}^*\|^2 + 2\|\breve{\mathfrak{u}}^* - \mathfrak{u}^*\|^2 \\ &\leq 8\left(\|q\tanh\left(\breve{\mathcal{K}}^*\right)\|^2 + \|q\tanh\left(\breve{\mathcal{K}}\right)\|^2\right) + 4\mathfrak{b}_\varsigma^2 + 2\varrho\|e_h\|^2 \\ &\leq 16q^2m + 4\mathfrak{b}_\varsigma^2 + 2\varrho\|e_h\|^2. \end{aligned} \tag{21}$$

Recall the definition of e_h. Then we have

$$-x^T\mathcal{Q}x \leq -(1-\alpha_c)\breve{x}_h^T\mathcal{Q}\breve{x}_h + \left(\frac{1}{\alpha_c} - 1\right)e_h^T\mathcal{Q}e_h. \tag{22}$$

In light of (20)-(22), it can be derived that

$$\dot{\mathcal{L}}_b \leq -(1-\alpha_c)(1-\gamma_c)\lambda_{\min}(\mathcal{Q})\|\breve{x}_h\|^2 - (1-\alpha_c)\gamma_c\lambda_{\min}(\mathcal{Q})\|\breve{x}_h\|^2 + \left(2\varrho + \left(\frac{1}{\alpha_c}-1\right)\lambda_{\max}(\mathcal{Q})\right)\|e_h\|^2 + \mathcal{A}, \tag{23}$$

where $\mathcal{A} = 16q^2m + 4\mathfrak{b}_\varsigma^2 + \mathfrak{b}_\upsilon$.

The derivative of $\mathcal{L}_c$ satisfies that

$$\dot{\mathcal{L}}_c = \tilde{\omega}_c^T\breve{\theta}e_I - \tilde{\omega}_c^T\bar{\theta}\bar{\theta}^T\tilde{\omega}_c \leq -\frac{1}{2}\lambda_{\min}\left(\bar{\theta}\bar{\theta}^T\right)\|\tilde{\omega}_c\|^2 + \frac{1}{2}\mathfrak{b}_I^2. \tag{24}$$

Utilizing triggering condition (17), we derive from (23) and (24) that

$$\dot{\mathcal{L}}_Y \leq -(1-\alpha_c)\gamma_c\lambda_{\min}(\mathcal{Q})\|\breve{x}_h\|^2 - \frac{1}{2}\lambda_{\min}\left(\bar{\theta}\bar{\theta}^T\right)\|\tilde{\omega}_c\|^2 + \frac{1}{2}\mathfrak{b}_I^2 + \mathcal{A}. \tag{25}$$

Situation ii. Analyze the situation when an event occurs, i.e., $t = \zeta_{h+1}$. In light of the properties of limits and the discussions in Situation i, it follows that $\triangle\mathcal{L}_Y < 0$. That completes the proof.

Remark 3.1. *With the aid of* (17)*, event generator can determine that at which instant the event occurs. During the stage between two adjacent triggering instants, the control input remains unchanged through applying zero-order holder. At triggering instants, the control input is updated with the current data, and the triggering error is reset to zero to prepare for another cycle. In addition, the operation* (16) *is able to transform the triggering mechanism into hibernation mode, avoiding superfluous resource consumptions associated with ETM when the control accuracy is achieved.*

3.3. **The dynamic event-triggered and self-triggered control schemes.** Through introducing the internal variable, the event-triggered condition can be endowed with dynamic characteristics to further reduce triggering frequency. The dynamics for the variable ν is presented as

$$\dot{\nu} = -\ell_\nu\nu + \Psi_\nu(\breve{x}_h, e_h), \quad \nu_0 \geq 0, \tag{26}$$

where $\ell_\nu > 0$, $\Psi_\nu(\breve{x}_h, e_h) = (1-\alpha_c)(1-\gamma_c)\lambda_{\min}(\mathcal{Q})\|\breve{x}_h\|^2 - \left(2\varrho + \left(\frac{1}{\alpha_c}-1\right)\lambda_{\max}(\mathcal{Q})\right)\|e_h\|^2$. Then we can construct the dynamic triggering rule as

$$\zeta_{h+1} = \inf\left\{t > \zeta_h : \Gamma(e_h) \geq \frac{(1-\alpha_c)(1-\gamma_c)\lambda_{\min}(\mathcal{Q})\|\breve{x}_h\|^2}{2\varrho + \left(\frac{1}{\alpha_c}-1\right)\lambda_{\max}(\mathcal{Q})} + \varpi \triangleq \mathcal{T}_D\right\}, \tag{27}$$

where $\varpi = \nu \Big/ \left(2\varrho\beta + \left(\frac{1}{\alpha_c}-1\right)\beta\lambda_{\max}(\mathcal{Q})\right)$ and $\beta > 0$.

Lemma 3.1. *The variable ν can be guaranteed to be positive when the triggering rule* (27) *is applied.*

Theorem 3.2. *Let Assumptions 2.1, 3.1 and 3.2 hold. For the system* (1)*, the control input and the triggering rule are respectively given by* (12) *and* (27)*. The critic weights update according to* (14)*. Then the system states and critic weight error are UUB.*

Proof: The Lyapunov candidate is chosen as $\mathcal{L}_E = \mathcal{L}_Y + \nu(t)$. Based on the results of Theorem 3.1 and the dynamics of $\nu(t)$, it follows that

$$\dot{\mathcal{L}}_E \leq -(1-\alpha_c)\gamma_c\lambda_{\min}(\mathcal{Q})\|\breve{x}_h\|^2 - \frac{1}{2}\lambda_{\min}\left(\bar{\theta}\bar{\theta}^T\right)\|\tilde{\omega}_c\|^2 - \ell_\nu\nu + \frac{1}{2}\mathfrak{b}_I^2 + \mathcal{A}. \tag{28}$$

Theorem 3.2 is proved.

Remark 3.2. *The variable ν is introduced to endow ETM with dynamic characteristics. Compared with the static triggering mechanism, the dynamic mechanism can save more system resources since the triggering threshold is enlarged to lower the triggering frequency.*

Note that the triggering interval analysis is necessary to exclude Zeno phenomena and construct the self-triggered mechanism. Based on Assumption 2.1 and condition (17) without preliminary operation, one can derive

$$\|\dot{x}\| \leq \mathfrak{L}_{\mathcal{F}}\|x\| + \mathcal{D}\|\breve{\mathfrak{u}}^*\|. \tag{29}$$

From Assumption 3.1 and Assumption 3.2, it can be deduced that $\|\breve{\mathfrak{u}}^*\| \leq \sqrt{\varrho}\|e_h(t)\| + \|\mathfrak{u}^*\|$ and $\mathfrak{u}^*$ is bounded. Let $\|\mathfrak{u}^*\| \leq \gamma_u$, and it follows that

$$\|\dot{x}\| \leq \mathfrak{L}_{\mathcal{F}}\|x\| + \mathcal{D}\sqrt{\varrho}\|e_h(t)\| + \mathcal{D}\gamma_u. \tag{30}$$

Recalling the definition of $e_h(t)$, we can derive

$$\|\dot{e}_h\| \leq \mathfrak{L}_{\mathcal{F}}\|\breve{x}_h\| + (\mathfrak{L}_{\mathcal{F}} + \mathcal{D}\sqrt{\varrho})\|e_h(t)\| + \mathcal{D}\gamma_u. \tag{31}$$

Based on Comparison Lemma [23], we can get

$$\|e_h\| \leq \frac{\mathfrak{L}_{\mathcal{F}}\|\breve{x}_h\| + \mathcal{D}\gamma_u}{\mathfrak{L}_{\mathcal{F}} + \mathcal{D}\sqrt{\varrho}}\left(e^{(\mathfrak{L}_{\mathcal{F}}+\mathcal{D}\sqrt{\varrho})(t-\zeta_h)} - 1\right). \tag{32}$$

In light of (17), one can deduce that Zeno phenomena can be excluded. Furthermore, the self-triggered rule can be further constructed as

$$\zeta_{h+1} = \zeta_h + \frac{1}{\mathfrak{L}_{\mathcal{F}} + \mathcal{D}\sqrt{\varrho}}\ln\left(1 + \left(\frac{\mathfrak{L}_{\mathcal{F}} + \mathcal{D}\sqrt{\varrho}}{\mathfrak{L}_{\mathcal{F}}\|\breve{x}_h\| + \mathcal{D}\gamma_u}\sqrt{\frac{(1-\alpha_c)(1-\gamma_c)\lambda_{\min}(\mathcal{Q})\|\breve{x}_h\|^2}{2\varrho + \left(\frac{1}{\alpha_c} - 1\right)\lambda_{\max}(\mathcal{Q})}}\right)\right). \tag{33}$$

Theorem 3.3. *Consider system* (1) *and let Assumptions 2.1, 3.1 and 3.2 hold. The critic update law and control input are derived from* (14) *and* (12)*. Then, when the self-triggered rule* (33) *is applied, system states and the errors of critic weights can be UUB.*

Proof: From (33), one can deduce that the triggering interval for ETM is larger than that for STM, implying that STM is more conservative than ETM in the aspect of triggering judgement. Thus, Theorem 3.3 is proved according to the results of Theorem 3.1.

Remark 3.3. *When the triggering parameters are the same, the triggering judgement for STM is more conservative than that for ETM, implying the triggering number for the former is larger than that for the latter one. Unlike ETM, STM is the mechanism that actively determines the next selected instant, which can avoid the need for continuous monitoring of system states. In brief, the DETM and STM have different characteristics and are both able to save computation/communication resources of the systems.*

4. **Numerical Example.** Consider the system

$$\dot{x} = \begin{bmatrix} 0.6x_2 - 0.2x_1 \\ -0.8x_1 - 0.6x_2 + 0.5(\sin(x_1))^2\cos(x_2) \end{bmatrix} + \begin{bmatrix} 0 \\ 0.6\sin(x_1)\cos(x_2) \end{bmatrix}\mathfrak{u}. \tag{34}$$

Let $\mathfrak{b}_\alpha = -0.3$, $\mathfrak{b}_\gamma = 0.1$, $\mathcal{Q} = 0.8I_{2\times 2}$, $x_0^T = [-0.5, 0.5]$, and $\delta(x) = [x_1^2, x_2^2, x_1x_2]^T$. The initial critic weight is randomly selected in $[-0.5, 0.5]$ in general. Specifically, in this experiment, it is set as $\omega_{c0} = [0.1892, 0.2482, -0.0495]^T$. The other parameters are chosen as $\alpha_c = \gamma_c = 0.95$, $\varrho = 2$, $T_p = 0.00001$, $\ell_\nu = 0.2$, $\beta = 5$, and $\nu_0 = 1$.

The simulation results are provided in Figures 1-3. The experiment associated with DETM was simulated for 80 seconds with a sampling period of 0.01 second, allowing for

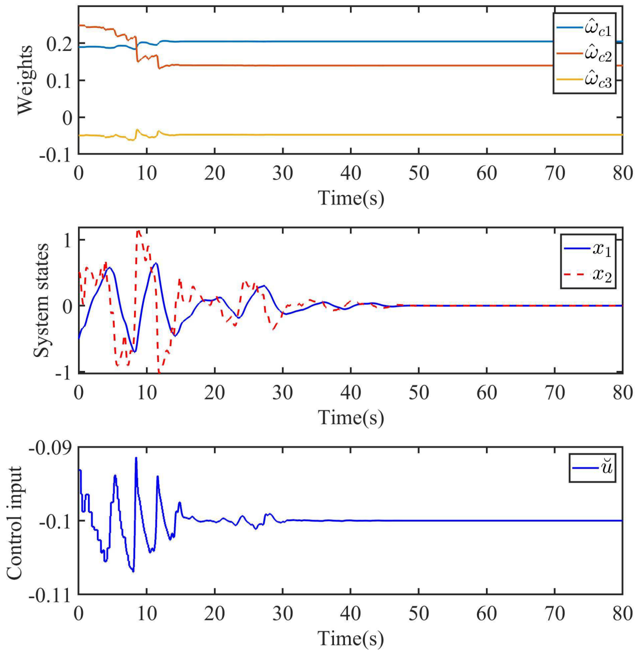

FIGURE 1. The evolutions for critic weights, system states and control input

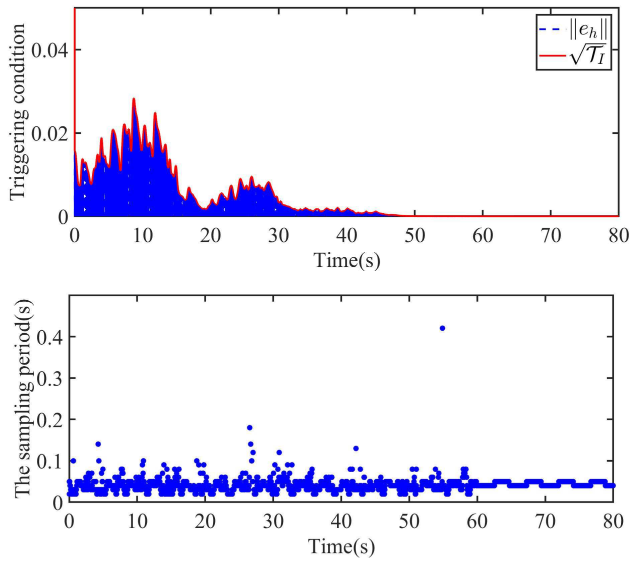

FIGURE 2. The triggering condition and triggering periods for ETM

the collection of 8000 data points to update the control strategy. Figure 1 presents the evolutions of critic weights, system states and control input. One can observe that the control signal is constrained with the asymmetric bounds and can drive system states to reach zero states. Moreover, the comparison experiments between the proposed triggering learning method and the traditional static event-triggered one are designed using the same experiment parameters except for the differences associated with the triggering mechanism. More specifically, the experiment parameter settings for the comparison method are the same as that for the proposed method but the former one is without preliminary operation and internal dynamic variable. The construction idea of the static triggering condition is inspired by the previous works in [6, 17, 18] to some extent. Given that the system examined in this paper exhibits distinct characteristics compared to

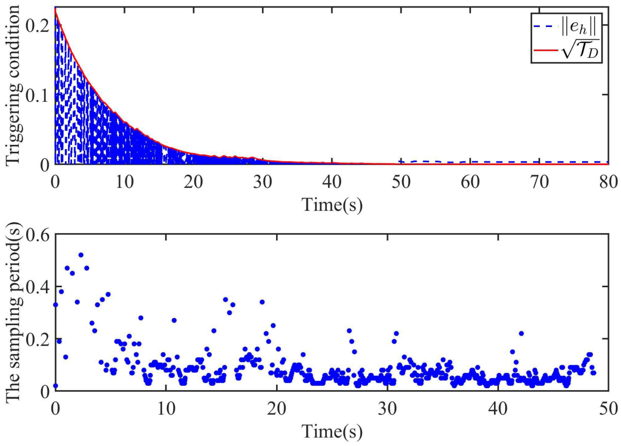

FIGURE 3. The triggering condition and triggering periods for DETM with preliminary operation

those documented in the above-mentioned literature, corresponding modifications have been made to the triggering conditions. The static event-triggered method can be constructed in light of (17) without preliminary operation. Figure 2 and Figure 3 present the triggering conditions and triggering periods for the traditional event-triggered approach and the proposed method, respectively. From the comparison between these figures, one can visibly deduce that the triggering thresholds and periods for the proposed method based on DETM are larger than that for the static event-triggered method, signifying the triggering frequency and resource consumption have been reduced.

More details illustrating the improvement on the triggering mechanism are provided in Figure 3. In general, when $\|e_h\|$ is larger than $\sqrt{\mathcal{T}_D}$, it implies that an event is generated. After $t = 50$ s, $\|e_h\|$ remains larger than $\sqrt{\mathcal{T}_D}$. Nevertheless, no event occurs since preliminary operation works. During the whole learning phase, the triggering number for the proposed triggering method with preliminary operation, the proposed method without preliminary operation and the traditional triggering adaptive method are 739, 792, and 2127, while the number for time-driven method is 8000, which verifies the improvement of the proposed approach. In addition, the experiment about STM was simulated for 50 seconds, in which 5000 pieces of data have been collected. The self-triggering number is 2447, revealing that this active mechanism can save system resources and relax hardware monitoring requirements when compared with time-driven mechanism.

5. **Conclusion.** In this paper, the event-triggered and self-triggered RL methods have been proposed to solve the OCIs of the systems with asymmetric constrained input. First, the cost function with non-quadratic utility function is developed to provide the theoretical basis for seeking the constrained optimal control input. Then, the IRL approach is constructed to solve HJBE without using the drift dynamics information. Furthermore, the dynamic event-triggered mechanism with preliminary operation and self-triggered mechanism are integrated into the RL algorithm architecture, reducing the event-triggered frequency and avoiding the continuous monitoring behaviors of hardware equipment, respectively. Finally, the main theorems and simulation results both verify the effectiveness of the online learning algorithm. In the future, we will focus on the optimal control issues for the complex nonlinear systems, such as multi-agent systems, robot systems, and microgrids. In addition, to further improve the effectiveness of self-triggered adaptive learning method is another research trend of our future work.

Acknowledgment. This work was supported in part by the National Natural Science Foundation of China (62403086, U22A20101), the National Key R&D Program of China (2022YFE0107300), and the Chongqing Natural Science Foundation (CSTB2024NSCQ-MSX1053, CSTB2024NSCQ-QCXMX0052, CSTB2024NSCQ-LZX0143).

REFERENCES

[1] W. Cao, Q. Yang, W. Meng and S. Xie, Data-based robust adaptive dynamic programming for balancing control performance and energy consumption in wastewater treatment process, *IEEE Transactions on Industrial Informatics*, vol.20, no.4, pp.6622-6630, 2024.

[2] D. Wang, H. He, X. Zhong and D. Liu, Event-driven nonlinear discounted optimal regulation involving a power system application, *IEEE Transactions on Industrial Electronics*, vol.64, no.10, pp.8177-8186, 2017.

[3] Y. Shen and M. Chen, Event-triggering-learning-based ADP control for post-stall pitching maneuver of aircraft, *IEEE Transactions on Cybernetics*, vol.54, no.1, pp.423-434, 2024.

[4] S. Sendari, Muladi, F. Ardiyansyah, S. Setumin, N. B. Mokhtar, H.-I Lin and P. Hartono, Commonsensical incentive reward in deep actor-critic reinforcement learning for mobile robot navigation, *International Journal of Innovative Computing, Information and Control*, vol.20, no.2, pp.373-389, 2024.

[5] X. Yang, M. Xu and Q. Wei, Dynamic event-sampled control of interconnected nonlinear systems using reinforcement learning, *IEEE Transactions on Neural Networks and Learning Systems*, vol.35, no.1, pp.923-937, 2024.

[6] P. Li, H. Zhang, W. Ao and P. Liu, Event-triggered adaptive control for nonlinear multi-player games using neural critic learning, *International Journal of Innovative Computing, Information and Control*, vol.20, no.5, pp.1257-1275, 2024.

[7] X. Cui, J. Chen, Y. Cui and S. Xu, Two-player nonlinear Stackelberg differential game via off-policy integral reinforcement learning, *Journal of the Franklin Institute*, vol.361, no.8, Article no.106812, 2024.

[8] C. Mu, Y. Zhang, Z. Gao and C. Sun, ADP-based robust tracking control for a class of nonlinear systems with unmatched uncertainties, *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol.50, no.11, pp.4056-4067, 2020.

[9] M. Ha, D. Wang and D. Liu, Novel discounted adaptive critic control designs with accelerated learning formulation, *IEEE Transactions on Cybernetics*, vol.54, no.5, pp.3003-3016, 2024.

[10] H. Zhang, N. Zhao, S. Wang and R. K. Agarwal, Improved event-triggered dynamic output feedback control for networked T-S fuzzy systems with actuator failure and deception attacks, *IEEE Transactions on Cybernetics*, vol.53, no.12, pp.7989-7999, 2023.

[11] Y. Wang, C. Wen and X. Li, Event-triggered H_∞ PI state estimation for delayed switched neural networks, *Journal of Automation and Intelligence*, vol.3, no.1, pp.26-33, 2024.

[12] X. Yang and Q. Wei, Adaptive critic learning for constrained optimal event-triggered control with discounted cost, *IEEE Transactions on Neural Networks and Learning Systems*, vol.32, no.1, pp.91-104, 2021.

[13] N. Zhao, D. Zhao and Y. Liu, Resilient event-triggering adaptive neural network control for networked systems under mixed cyber attacks, *Neural Networks*, vol.174, Article no.106249, 2024.

[14] Y. Qian and L. Liu, Event-triggered robust cooperative output regulation for a class of linear multi-agent systems with an unknown exosystem, *Journal of Automation and Intelligence*, vol.3, no.2, pp.119-127, 2024.

[15] J. Lu, Q. Wei, Z. Wang, T. Zhou and F. Wang, Event-triggered optimal control for discrete-time multi-player non-zero-sum games using parallel control, *Information Sciences*, vol.584, pp.519-535, 2022.

[16] S. Xue, B. Luo, D. Liu and Y. Yang, Constrained event-triggered H_∞ control based on adaptive dynamic programming with concurrent learning, *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol.52, no.1, pp.357-369, 2022.

[17] X. Cui, B. Peng, B. Wang and L. Wang, Event-triggered neural experience replay learning for nonzero-sum tracking games of unknown continuous-time nonlinear systems, *International Journal of Robust and Nonlinear Control*, vol.33, no.12, pp.6553-6575, 2023.

[18] H. Zhang, H. Su, K. Zhang and Y. Luo, Event-triggered adaptive dynamic programming for nonzero-sum games of unknown nonlinear systems via generalized fuzzy hyperbolic models, *IEEE Transactions on Fuzzy System*, vol.27, no.11, pp.2202-2214, 2019.

[19] C. Mu, K. Wang and T. Qiu, Dynamic event-triggering neural learning control for partially unknown nonlinear systems, *IEEE Transactions on Cybernetics*, vol.52, no.4, pp.2200-2213, 2022.
[20] P. Liu, H. Zhang, Z. Ming, S. Wang and R. K. Agarwal, Dynamic event-triggered safe control for nonlinear game systems with asymmetric input saturation, *IEEE Transactions on Cybernetics*, vol.54, no.9, pp.5115-5126, 2024.
[21] X. Tong, D. Ma, Z. Wang, Z. Ming and X. Xie, Model-free adaptive dynamic event-triggered robust control for unknown nonlinear systems using iterative neural dynamic programming, *Information Sciences*, vol.655, Article no.119866, 2024.
[22] B. Zhao, S. Zhang and D. Liu, Self-triggered approximate optimal neuro-control for nonlinear systems through adaptive dynamic programming, *IEEE Transactions on Neural Networks and Learning Systems*, DOI: 10.1109/TNNLS.2024.3362800, 2024.
[23] H. K. Khalil, *Nonlinear Systems*, Prentice-hall, Upper Saddle River, NJ, 2002.

Author Biography

Pengda Liu received the B.S. degree in Automation and the Ph.D. degree in Control Theory and Control Engineering from Northeastern University, Shenyang, China, in 2012 and 2022, respectively. He is currently a Lecturer with Chongqing Technology and Business University, Chongqing, China. His current research interests include adaptive dynamic programming, reinforcement learning, neural-networks-based control, optimal control, and their applications.

Huiyan Zhang received the M.Sc. degree in Control Engineering and the Ph.D. degree in Control Theory and Control Engineering from Harbin Institute of Technology, Harbin, in September 2014 and April 2019, respectively. From September 2015 to September 2017, she was a Joint Training Ph.D. Student with the School of Electrical and Electronic Engineering, the University of Adelaide. She is currently an Associate Professor with Chongqing Technology and Business University. Her research interests include stochastic switched systems, event-triggered scheme, model reduction, balanced truncation, robust control, and filtering design.

Wengang Ao received the B.S. degree in Engineering Mechanics from Sichuan University, Chengdu, in June 2000, and M.E. degree in Solid Mechanics from Chongqing University, Chongqing, in June 2007. He is currently a Professor at Chongqing Technology and Business University. His research interests include strength theory, kinematics and dynamics, robust control, and intelligent robots.