

SELF-TRIGGERED ROBUST TRACKING CONTROL FOR WHEELED MOBILE ROBOTS USING REINFORCEMENT LEARNING

PENGDA LIU, HUIYAN ZHANG*, WENGANG AO AND HUAFENG QIN

International Science and Technology Cooperation Base of Intelligent Manufacturing Service
Chongqing Technology and Business University
No. 853, Lihua Avenue, Nan'an District, Chongqing 400072, P. R. China
{liupengda; aowg; 2013009}@ctbu.edu.cn

*Corresponding author: huiyanzhang@ctbu.edu.cn

Received December 2025; revised March 2026

ABSTRACT. *This paper investigates the robust tracking control issues for wheeled mobile robots with external disturbances using self-triggered reinforcement learning approach. Firstly, based on the Euler-Lagrange motion equation of wheeled mobile robots, the augmented system with discounted performance index is constructed such that the robust tracking control issues can be converted into the stabilizing optimization control issues. Then, the reinforcement learning approach is implemented by critic neural networks to approximately derive the optimal control schemes. Different from the time-driven or traditional event-driven control mechanisms, the triggering mechanism is designed such that the triggering condition is derived only from the data sampled at the triggering instants, providing the foundation for constructing a self-triggered adaptive control mechanism. Furthermore, the self-triggered rule is proposed and embedded within the reinforcement learning framework. This rule drives the control strategy to update at the predicted triggering instants, reducing computational and communication resource consumption. In addition, the system states and the critic weight error are demonstrated to be uniform ultimately bounded via the Lyapunov theory. Finally, the simulation results indicate that, with the derived self-triggered control scheme, the states can track the reference signal well while using fewer sets of data than the comparative scheme, validating the effectiveness of the proposed learning-based approach.*

Keywords: Reinforcement learning, Adaptive dynamic programming, Robust tracking control, Wheeled mobile robots, Self-triggered control

1. Introduction. Wheeled mobile robots (WMRs) are essential equipment for logistics delivery, environmental exploration, and other fields, enabled by their mobility and environmental adaptability [1]. Due to its essential role in task execution, tracking control for WMRs has received widespread attention in the research community. For instance, in [2], via introducing a passive steering angle, the proportion integration feedback tracking controller was designed for tractor-trailer WMR by combining the motion speed relation and directed Lyapunov approach. The trajectory tracking control issues for WMRs subject to longitudinal slipping and lateral skidding were studied in [3] with the aid of disturbance observer. In [4], an adaptive iterative learning approach was developed to tackle the trajectory tracking control issue for a class of two-wheeled mobile robots subject to model uncertainties and unknown disturbances.

For complex systems represented by WMRs, the design of control schemes usually requires striking a critical balance between control performance and cost [5]. This typically involves defining the performance index function and solving the Hamilton-Jacobi-Bellman

equation (HJBE) for the optimal control issues (OCIs). The existence of nonlinear terms and partial derivatives makes obtaining closed-form solutions for HJBEs extremely challenging [6]. Fortunately, owing to their powerful optimization capabilities, learning algorithms offer a viable alternative for obtaining approximate solutions. Among these learning methods, reinforcement learning (RL) and adaptive dynamic programming (ADP) are frequently viewed as equivalent in optimal control, as they share similar characteristics [7, 8, 9, 10]. Recently, RL and ADP approaches have been developed to tackle the OCIs of different systems. In [11], the RL-based approximate optimal control method was proposed to address the attitude reorientation problems of rigid bodies under multiple state constraints. The off-policy learning approach was developed in [12] to solve two-player Stackelberg differential games. By virtue of a self-learning optimal control formulation, the trajectory tracking design of uncertain nonlinear systems was investigated in [13]. In [14], the system transformation based on barrier function was developed, and then the safe reinforcement learning algorithm was employed to design the approximate optimal controller under an online actor-critic framework. Based on the concurrent learning technique, the ADP method which could work without persistent/finite assumptions was developed in [15] to approximate the optimal tracking policy for the nonlinear systems. In [16], the critic-only neural network ADP technique was utilized to design the optimal tracking controller, which guaranteed the robot tracks the desired trajectory in fixed time.

Conventional time-triggered schemes incur growing computational/communication overhead with rising system complexity, as continuous policy and data updates impose significant resource demands [17, 18]. To counteract these limitations, event-triggered mechanisms (ETMs) have been incorporated into control frameworks. In general, the design of triggering conditions critically governs event-triggered mechanisms. By strategically constructing these conditions, event-driven methods achieve control performance while updating control policies only with essential data, thereby alleviating computational loads and communication bandwidth constraints. In [19], an event-triggered state-dependent switching method was proposed to solve the adaptive fuzzy control problem of switched nonlinear systems. In [20], the leader-following dissipative fault-tolerant consensus issue was investigated for the nonlinear multi-agent systems with semi-Markov switching topologies. Furthermore, the dynamic memory event-triggered mechanism and memory-based distributed fault-tolerant controllers were designed to reduce the triggering times. Compared to conventional ETM, the dynamic triggering mechanism adaptively expands the triggering threshold through a dynamic variable, thereby suppressing triggering frequency. Based on dynamic triggering mechanism, an optimal control scheme of multiplayer nonzero-sum game was developed in [21] for modular robot manipulator systems. In [22], the robust tracking control issue for the constrained nonlinear stochastic systems was investigated using an ADP approach based on dynamic triggering mechanism. The particle swarm optimization-based ADP method was proposed in [23] to derive the decentralized event-triggered tracking control strategy for a class of nonlinear large-scale interconnected systems. Control approaches incorporating ETM significantly conserve computational and communication resources while achieving control objectives. However, ETMs passively rely on continuous hardware-based state monitoring, whereas self-triggered mechanisms (STMs) actively eliminate persistent sensing demands, easing the associated computational burdens. In [24], the sectionalized self-triggered control scheme was developed for autonomous underwater vehicles under deferred quantitative constraints. However, research on STM-based learning approaches remains insufficient, particularly for robust tracking control of Mecanum-wheeled mobile robots (MWMRs).

Inspired by existing approaches, the STM-based learning algorithm is proposed to tackle the robust tracking control issues (RTCIs) of MWMRs. The contributions are presented

as follows. First, unlike the works in [1, 16], the robust optimal tracking control issue for MWMRs is investigated. Accounting for Mecanum wheel slippage, the Euler-Lagrange equation governing MWMR dynamics is derived to establish the foundation for the design of robust tracking control scheme. Then, the augmented system together with the discounted performance is constructed such that the tracking control issues can be converted into the optimal control issues. Based on the critic neural network, the learning approach is implemented to obtain the approximate solution of HJBE. Finally, different from the ETM presented in [21, 23], the triggering condition is designed which only relies on the states at the triggering instants and the Zeno phenomena can be avoided. Building on this foundation, the proposed STM-based learning approach achieves control objectives while actively reducing computational and communication burdens.

The remainder of this work is organized as follows. Section 2 presents the model construction for MWMRs and the formulation of the RCTIs. The derivation of triggering condition and the self-triggered learning algorithm is provided in Section 3. In Section 4, the simulation results are presented. Section 5 presents the conclusion.

2. Problem Statement and Preliminaries. To facilitate the study of the movement of the mobile robot chassis, the coordinate frames are designed as the ones presented in Figure 1. The coordinate frames OXY and $CX_R Y_R$ are the global inertial frame for the robot and the body-fixed frame. Note that the origin of $CX_R Y_R$ frame coincides with the center of mass. When the Mecanum wheels rotate forward, the rollers come into contact with the ground passively. For the first wheel, the total friction force can be represented by F_1 , which can be decomposed into F_{11} and F_{12} . Similarly, for the other two wheels, the component forces F_{21} , F_{22} , F_{31} and F_{32} can be derived from the friction forces F_2 and F_3 . A Mecanum wheel consists of a hub and rollers mounted around its periphery, with each roller's axis oriented at an angle δ to the hub axis. In this paper, we consider the situation that $\delta = \pi/3$. The hubs of the wheels rotate with angular velocities ω_1 , ω_2 , and ω_3 , respectively, while the chassis rotates about its center with angular velocity ω_c . In the global inertial frame OXY , $p = (x, y, \theta)$ and $\vartheta = (\dot{x}, \dot{y}, \dot{\theta})$ are utilized to describe the motion of the robot chassis. Correspondingly, the velocity of the robot chassis in the $CX_R Y_R$ frame is denoted as $\vartheta_d = (\vartheta_{xR}, \vartheta_{yR}, \dot{\theta})$.

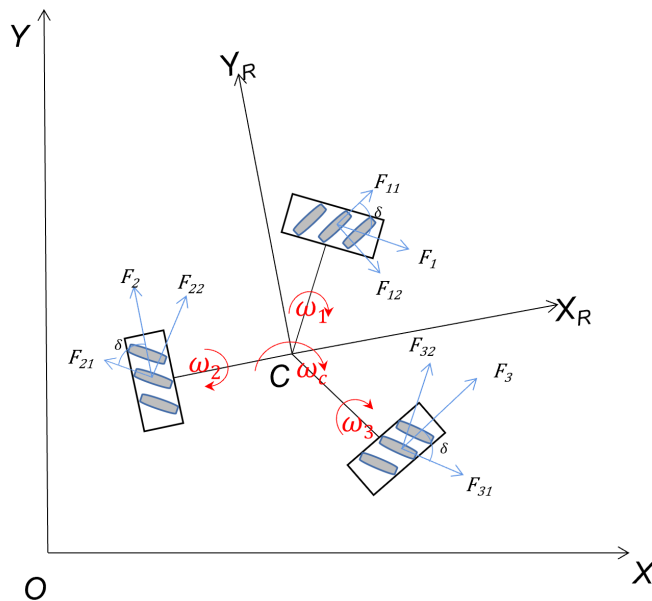


FIGURE 1. Coordinate frames for wheeled mobile robot

Then, we derive the following relation

$$\vartheta = R(\theta)\vartheta_d, \quad (1)$$

where

$$R(\theta) = \begin{bmatrix} \cos \theta & -\sin \theta & 0 \\ \sin \theta & \cos \theta & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Considering the kinematic characteristics of Mecanum wheels, the actual direction of motion at the wheel center is determined by the roller orientation. Let $A_i X_{\omega_i} Y_{\omega_i}$ be the coordinate frame with its origin at the centroid of the wheel center. Define k as the distance from C to the geometric center of each wheel, and r as the wheel radius. The velocity vector at the wheel center can be presented as $\vartheta_{Ai} = r\dot{\phi} \cos \delta$ where $\dot{\phi}$ denotes the angular velocity for the wheel i . Applying instantaneous center method, we have

$$\mathcal{A}\vartheta_d = r\dot{\phi} \cos \delta, \quad (2)$$

where

$$\mathcal{A} = \begin{bmatrix} \sin(\alpha_1 + \delta) & -\cos(\alpha_1 + \delta) & -k \cos \delta \\ \sin(\alpha_2 + \delta) & -\cos(\alpha_2 + \delta) & -k \cos \delta \\ \sin(\alpha_3 + \delta) & -\cos(\alpha_3 + \delta) & -k \cos \delta \end{bmatrix}.$$

Herein, let $\alpha_i = (2i - 1)\pi/3$ and $\delta = \pi/3$. To account for wheel slip conditions, the slip parameter is defined as

$$\chi = \frac{r\dot{\phi}_i \cos \delta - \vartheta_{Ai}}{r\dot{\phi}_i \cos \delta}. \quad (3)$$

Then it is obtained that $\vartheta_{Ai} = (1 - \chi)r \cos \delta \dot{\phi}_i$. The velocity ϑ becomes

$$\vartheta = (1 - \chi)r \cos \delta (\mathcal{A}R^{-1}(\theta))^{-1} \dot{\phi}. \quad (4)$$

Based on Lagrangian mechanics, the kinetic energy for the robot can be expressed as

$$\mathcal{K} = \frac{1}{2}m(\vartheta_x^2 + \vartheta_y^2) + \frac{1}{2}J_\omega \omega^2 + \frac{1}{2}J_\phi (\dot{\phi}_1^2 + \dot{\phi}_2^2 + \dot{\phi}_3^2), \quad (5)$$

where m is the mass of the robot, J_ω is the moment of inertia for the robot about its center of mass, and J_ϕ is the moment of inertia for the wheels. When $\delta = \pi/3$, we obtain

$$\begin{aligned} \mathcal{K} = & \left[(1 - \chi)^2 \left(\frac{mr^2}{18} + \frac{J_\omega r^2}{18k^2} \right) + \frac{1}{2}J_\phi \right] (\dot{\phi}_1^2 + \dot{\phi}_2^2 + \dot{\phi}_3^2) \\ & + (1 - \chi)^2 \left(\frac{J_\omega r^2}{9k^2} - \frac{mr^2}{18} \right) (\dot{\phi}_1 \dot{\phi}_2 + \dot{\phi}_2 \dot{\phi}_3 + \dot{\phi}_1 \dot{\phi}_3). \end{aligned} \quad (6)$$

Since the robot undergoes planar motion, its potential energy $\mathfrak{V}$ is equal to zero. Then Lagrangian function can be deduced that $\mathfrak{L} = \mathcal{K} - \mathfrak{V} = \mathcal{K}$. Considering the generalized force $\mathcal{K} = (\mathcal{K}_1, \mathcal{K}_2, \mathcal{K}_3)^T$ as the control input, we can derive the Lagrange equation as

$$\frac{d}{dt} \left(\frac{\partial \mathfrak{L}}{\partial \dot{\phi}_i} \right) - \frac{\partial \mathfrak{L}}{\partial \phi_i} + \frac{\partial \mathcal{D}}{\partial \dot{\phi}_i} = \mathcal{K}_i, \quad (7)$$

where $\mathcal{D} = 1/2\mu (\dot{\phi}_1^2 + \dot{\phi}_2^2 + \dot{\phi}_3^2)$ is the Rayleigh's dissipation function, and μ denotes the viscous friction coefficient of the wheels. According to Equations (6) and (7), the robot chassis dynamics equation is deduced that

$$\aleph \ddot{\phi} + \mu \dot{\phi} = \mathcal{K}, \quad (8)$$

where

$$\aleph = \begin{bmatrix} (1-\chi)^2 \left(\frac{mr^2}{9} + \frac{J_\omega r^2}{9k^2} \right) + J_\phi & (1-\chi)^2 \left(\frac{J_\omega r^2}{9k^2} - \frac{mr^2}{18} \right) & (1-\chi)^2 \left(\frac{J_\omega r^2}{9k^2} - \frac{mr^2}{18} \right) \\ (1-\chi)^2 \left(\frac{J_\omega r^2}{9k^2} - \frac{mr^2}{18} \right) & (1-\chi)^2 \left(\frac{mr^2}{9} + \frac{J_\omega r^2}{9k^2} \right) + J_\phi & (1-\chi)^2 \left(\frac{J_\omega r^2}{9k^2} - \frac{mr^2}{18} \right) \\ (1-\chi)^2 \left(\frac{J_\omega r^2}{9k^2} - \frac{mr^2}{18} \right) & (1-\chi)^2 \left(\frac{J_\omega r^2}{9k^2} - \frac{mr^2}{18} \right) & (1-\chi)^2 \left(\frac{mr^2}{9} + \frac{J_\omega r^2}{9k^2} \right) + J_\phi \end{bmatrix}.$$

By differentiating Equation (4), we obtain

$$\dot{\vartheta} = (1-\chi)r \cos \delta R(\theta) \mathcal{A}^{-1} \ddot{\phi} + \dot{R}(\theta) R^{-1}(\theta) \vartheta. \quad (9)$$

Substituting Equation (8) into Equation (9), we have

$$\dot{\vartheta} = (1-\chi)r \cos \delta R(\theta) \mathcal{A}^{-1} \aleph^{-1} \mathcal{K} + \left[\dot{R}(\theta) R^{-1}(\theta) - \mu R(\theta) \mathcal{A}^{-1} \aleph^{-1} (R(\theta) \mathcal{A}^{-1})^{-1} \right] \vartheta. \quad (10)$$

Then we derive the system that

$$\dot{x}(t) = f(x(t)) + g(x(t))(\mathcal{K}(t) + w(t)), \quad (11)$$

where $x = [p^T, \vartheta^T]^T$, $f(x) = [\vartheta^T, (f_p(p)\vartheta)^T]^T$, and $g(x) = [\mathbf{0}, g_p(p)^T]^T$. Furthermore, $f_p(p) = \dot{R}(\theta) R^{-1}(\theta) - \mu R(\theta) \mathcal{A}^{-1} \aleph^{-1} (R(\theta) \mathcal{A}^{-1})^{-1}$, $g_p(p) = r \cos \delta R(\theta) \mathcal{A}^{-1} \aleph^{-1}$, and $w(t)$ denotes the disturbances caused by slippage and other factors. The reference trajectory signal is described by

$$\dot{x}_r = c_r(x_r(t)). \quad (12)$$

The tracking error is defined as $e_z(t) = x(t) - x_r(t)$, such that the augmented state can be presented as $\mathcal{Z} = [e_z^T(t), x_r^T(t)]^T$. Then the augmented system is constructed as

$$\dot{\mathcal{Z}}(t) = \mathcal{F}(\mathcal{Z}(t)) + \mathcal{G}(\mathcal{Z}(t))(\mathbf{u}(t) + d_w(\mathcal{Z}(t))), \quad (13)$$

where

$$\begin{aligned} \mathcal{F}(\mathcal{Z}(t)) &= \begin{bmatrix} f(e_z(t) + x_r(t)) - c_r(x_r(t)) + g(e_z(t) + x_r(t)) \mathcal{K}_r \\ c_r(x_r(t)) \end{bmatrix}, \\ \mathcal{G}(\mathcal{Z}(t)) &= \begin{bmatrix} g(e_z(t) + x_r(t)) \\ \mathbf{0} \end{bmatrix}, \\ d_w(\mathcal{Z}(t)) &= \mathcal{G}(\mathcal{Z}) w(e_z(t) + x_r(t)). \end{aligned}$$

Herein, $\mathbf{u} = \mathcal{K} - \mathcal{K}_r$ and $\mathcal{K}_r = (g^T(x_r)g(x_r))^{-1} g^T(x_r)(c_r(x_r(t)) - f(x_r(t)))$. The function $\mathcal{F}(\mathcal{Z}(t))$ is locally Lipschitz continuous with Lipschitz constant γ_F and $\mathcal{G}(\mathcal{Z}(t))$ is bounded with γ_G . The disturbance term $d_w(\mathcal{Z}(t))$ is bounded with $w_M(\mathcal{Z})$. For simplicity of presentation in what follows, these functions are abbreviated as $\mathcal{F}(\mathcal{Z})$, $\mathcal{G}(\mathcal{Z})$, and $d_w(\mathcal{Z})$.

The nominal augmented system is derived that

$$\dot{\mathcal{Z}}(t) = \mathcal{F}(\mathcal{Z}) + \mathcal{G}(\mathcal{Z})\mathbf{u}(t). \quad (14)$$

The performance index is given that

$$\mathcal{J}(\mathcal{Z}) = \int_t^\infty e^{-\rho(\iota-t)} (\mathcal{Z}^T \mathcal{Q} \mathcal{Z} + \mathbf{u}^T \mathfrak{R} \mathbf{u} + w_M^2(\mathcal{Z})) d\iota, \quad (15)$$

where $\mathcal{Q} = [\mathcal{Q}_e, \mathbf{0}; \mathbf{0}, \mathbf{0}]$, and matrix $\mathcal{Q}_e$ and symmetric matrix $\mathfrak{R}$ are the positive definite matrices. Differentiating Equation (15) yields

$$\mathcal{H}(\mathcal{Z}, \mathbf{u}, \nabla \mathcal{J}) = \nabla \mathcal{J}^T(\mathcal{Z})(\mathcal{F}(\mathcal{Z}) + \mathcal{G}(\mathcal{Z})\mathbf{u}) - \rho \mathcal{J}(\mathcal{Z}) + \mathcal{Z}^T \mathcal{Q} \mathcal{Z} + \mathbf{u}^T \mathfrak{R} \mathbf{u} + w_M^2(\mathcal{Z}). \quad (16)$$

According to Bellman optimality principle, the optimal performance index is defined as

$$\mathcal{J}^*(\mathcal{Z}) = \min_{\mathbf{u}} \mathcal{H}(\mathcal{Z}, \mathbf{u}, \nabla \mathcal{J}). \quad (17)$$

Furthermore, the optimal control policy is got

$$\mathbf{u}^*(\mathcal{Z}) = -\frac{1}{2}\mathfrak{K}^{-1}\mathcal{G}^T(\mathcal{Z})\nabla \mathcal{J}^*(\mathcal{Z}). \quad (18)$$

Based on Equations (17) and (18), the HJBE is derived that

$$\begin{aligned} & \mathcal{H}(\mathcal{Z}, \mathbf{u}^*, \mathcal{J}^*) \\ &= (\nabla \mathcal{J}^*(\mathcal{Z}))^T \mathcal{F}(\mathcal{Z}) - \frac{1}{4} \mathcal{J}^*(\mathcal{Z})^T \mathcal{G}(\mathcal{Z}) \mathfrak{K}^{-1} \mathcal{G}^T(\mathcal{Z}) \mathcal{J}^*(\mathcal{Z}) - \rho \mathcal{J}^*(\mathcal{Z}) + \mathcal{Z}^T \mathcal{Q} \mathcal{Z} + w_M^2(\mathcal{Z}) \\ &= 0. \end{aligned} \quad (19)$$

Note that obtaining a closed-form solution for Equation (19) is challenging due to the partial differential terms and nonlinear terms. Therefore, based on the aperiodic triggering mechanism, we propose an RL-based approach to overcome this challenge.

3. Self-Triggered Robust Tracking Control Scheme Design for Wheeled Mobile Robots. In this section, the event-triggered RL method is proposed to approximate the closed-form solution of HJBE. The ETM operates solely on data sampled at triggering instants, avoiding the frequent triggering threshold computations required by conventional event-based control. Then, the main theorem and the stability analysis are presented to prove that the proposed event-based method can guarantee the ultimately boundedness of the tracking error. Finally, an STM designed to actively predetermine the next triggering instant is constructed and incorporated into the learning algorithm framework.

3.1. The implementation of event-triggered learning-based approach. To eliminate the need for continuous computation and update of control strategies in time-triggered approaches, the aperiodic triggering mechanism is designed to reduce the consumption of computational and communication resources in the system. Using the time sequence $\{\varsigma_j\}_{j=1}^{\infty}$ containing the triggering instants, we define the triggering error as $e_T(t) = \check{\mathcal{Z}}_j - \mathcal{Z}(t)$, $t \in [\varsigma_j, \varsigma_{j+1})$. Herein, $\check{\mathcal{Z}}_j$ denotes the state vector at the triggering instant ς_j . The event-based control policy is presented as

$$\check{\mathbf{u}}^*(t) = -\frac{1}{2}\mathfrak{K}^{-1}\mathcal{G}^T(\check{\mathcal{Z}}_j)\nabla \mathcal{J}^*(\check{\mathcal{Z}}_j), \quad t \in [\varsigma_j, \varsigma_{j+1}). \quad (20)$$

From Equation (20), one can deduce that once the control policy is updated at triggering instants, it remains unchanged due to the zero-order holder during this triggering period.

To implement the learning-based approach, the critic neural network is utilized to obtain the optimal performance index

$$\mathcal{J}^*(\mathcal{Z}) = \varpi_c^T \phi_c(\mathcal{Z}) + \epsilon_c(\mathcal{Z}), \quad (21)$$

where $\varpi_c \in \mathbb{R}^{\mathcal{N}}$ is the ideal weight vector, $\phi_c \in \mathbb{R}^{\mathcal{N}}$ is the activation function, and $\epsilon_c \in \mathbb{R}$ is the approximation error. The optimal control policy can be further derived as

$$\mathbf{u}^* = -\frac{1}{2}\mathfrak{K}^{-1}\mathcal{G}^T(\nabla \phi_c^T(\mathcal{Z})\varpi_c + \epsilon_c). \quad (22)$$

With the construction of ETM, the event-based control is obtained as

$$\check{\mathbf{u}}^* = -\frac{1}{2}\mathfrak{K}^{-1}\mathcal{G}^T(\check{\mathcal{Z}}_j)\left(\left(\nabla \phi_c(\check{\mathcal{Z}}_j)\right)^T \varpi_c + \epsilon_c(\check{\mathcal{Z}}_j)\right). \quad (23)$$

Since the ideal weight is usually unknown, the approximation of performance index is derived that

$$\hat{\mathcal{J}}^*(\mathcal{Z}) = \hat{\varpi}_c^T \phi_c(\mathcal{Z}), \quad (24)$$

where $\hat{\omega}_c$ is the approximated weight. Then the optimal control is approximated that

$$\check{\mathbf{u}} = -\frac{1}{2}\mathfrak{R}^{-1}\mathcal{G}^T\left(\check{\mathcal{Z}}_j\right)\left(\nabla\phi_c\left(\check{\mathcal{Z}}_j\right)\right)^T\hat{\omega}_c. \quad (25)$$

Furthermore, the event-based Hamiltonian is reformulated as

$$\begin{aligned} & \mathcal{H}(\mathcal{Z}, \hat{\omega}_c, \check{\mathbf{u}}) \\ &= \hat{\omega}_c^T \nabla \phi_c\left(\check{\mathcal{Z}}_j\right)\left(\mathcal{F}(\mathcal{Z}) + \mathcal{G}(\mathcal{Z})\check{\mathbf{u}}\right) - \rho \hat{\omega}_c^T \phi_c(\mathcal{Z}) + \mathcal{Z}^T \mathcal{Q} \mathcal{Z} + w_M^2\left(\check{\mathcal{Z}}_j\right) + \check{\mathbf{u}}^T \mathfrak{R} \check{\mathbf{u}} \\ &\triangleq \epsilon_H. \end{aligned} \quad (26)$$

The learning target for critic network is set as $\mathcal{E}_H = 1/2\epsilon_H^2$. Utilizing the gradient descent approach, the critic weight update law is derived that

$$\dot{\hat{\omega}}_c = -\ell_c \check{\rho} \epsilon_H, \quad (27)$$

where $\check{\rho} = \rho/(\rho^T \rho + 1)^2$, and $\rho = \nabla \phi_c^T(\mathcal{Z})(\mathcal{F}(\mathcal{Z}) + \mathcal{G}(\mathcal{Z})\check{\mathbf{u}}) - \rho \phi_c(\mathcal{Z})$.

Remark 3.1. In order to further enhance the robustness of the proposed method, the robust term and judgement function can be added to the critic learning law, which is presented as

$$R_t = \frac{\ell_s}{2} \Upsilon(\mathcal{Z}, \check{\mathbf{u}}) \nabla \phi_c\left(\check{\mathcal{Z}}_j\right) \mathcal{G}\left(\check{\mathcal{Z}}_j\right) \mathfrak{R}^{-1} \mathcal{G}^T \nabla \mathcal{L}_z \quad (28)$$

where

$$\Upsilon(\mathcal{Z}, \check{\mathbf{u}}) = \begin{cases} 0, & \text{if } \nabla \mathcal{L}_z^T(\mathcal{F}(\mathcal{Z}) + \mathcal{G}(\mathcal{Z})\check{\mathbf{u}}) < 0, \\ 1, & \text{otherwise.} \end{cases} \quad (29)$$

The robust term (28) and judgement function (29) can work together to suppress disturbances and be integrated into the ETM constructed in this paper. Nevertheless, it requires continuous monitoring of system states and thus cannot be generalized to proactively triggered control schemes.

Let $\tilde{\omega}_c = \omega_c - \hat{\omega}_c$, of which the dynamic is presented as

$$\dot{\tilde{\omega}}_c = -\ell_c \bar{\rho} \tilde{\omega}_c + \ell_c \check{\rho} e_c, \quad (30)$$

where $\bar{\rho} = \rho/(\rho^T \rho + 1)$, and $e_c = -\nabla \phi_c^T(\mathcal{F}(\mathcal{Z}) + \mathcal{G}(\mathcal{Z})\check{\mathbf{u}}) + \rho \phi_c(\mathcal{Z})$.

3.2. Stability analysis and the design of self-triggered control scheme. Prior to stating the main theorems, two assumptions are requisite [10, 26].

Assumption 3.1. The optimal control $\mathbf{u}^*(t)$ is Lipschitz continuous, satisfying that $\|\mathbf{u}^*(t) - \check{\mathbf{u}}^*(t)\|^2 \leq \nu \|\mathcal{Z} - \check{\mathcal{Z}}_j\|^2$.

Assumption 3.2. The terms ω_c , $\nabla \phi_c(\mathcal{Z})$, $\epsilon_c(\mathcal{Z})$, e_c are bounded, satisfying that $\|\omega_c\| \leq b_\omega$, $\|\nabla \phi_c(\mathcal{Z})\| \leq b_\phi$, $\|\epsilon_c(\mathcal{Z})\| \leq b_\epsilon$, and $\|e_c\| \leq b_e$.

The triggering condition is designed that

$$\|e_T\|^2 \leq \frac{(1 - \beta_1)\beta_2\lambda_m(\mathcal{Q}_e)}{2\nu\|\mathfrak{R}\| + \left(\frac{1}{\beta_1} - 1\right)\lambda_M(\mathcal{Q})} \|\check{e}_{zj}\|^2 \triangleq \mathcal{T}_D, \quad (31)$$

where $\beta_1 \in (0, 1)$ and $\beta_2 \in (0, 1)$ are the parameters determining the triggering frequency, and $\check{e}_{zj}$ denotes e_z when $t = \varsigma_j$.

Theorem 3.1. Consider the system (14) while Assumptions 3.1 and 3.2 hold. The critic weights are tuned according to Equation (27), and the control input is given by Equation

(25). The triggering condition satisfies Equation (31), and then the tracking error and the error of critic weights are uniformly ultimately bounded (UUB).

Proof: The Lyapunov candidate is selected as

$$\mathcal{L}_c = \underbrace{\mathcal{J}^*(\mathcal{Z})}_{\mathcal{L}_1} + \underbrace{\mathcal{J}^*(\check{\mathcal{Z}}_j)}_{\mathcal{L}_2} + \underbrace{\frac{1}{2}\tilde{\omega}_c^T \tilde{\omega}_c}_{\mathcal{L}_3}. \quad (32)$$

Since the triggering mechanism is introduced, the discussions should be presented in two situations.

Situation i. For $t \in [\varsigma_j, \varsigma_{j+1})$, events are not triggered. In this situation, $\dot{\mathcal{L}}_2 = 0$. Differentiating $\mathcal{L}_1$, we derive

$$\begin{aligned} \dot{\mathcal{L}}_1 &= (\nabla \mathcal{J}^*)^T (\mathcal{F}(\mathcal{Z}) + \mathcal{G}(\mathcal{Z})\check{\mathbf{u}}) \\ &= (\nabla \mathcal{J}^*)^T (\mathcal{F}(\mathcal{Z}) + \mathcal{G}(\mathcal{Z})\mathbf{u}^*) + (\nabla \mathcal{J}^*)^T \mathcal{G}(\mathcal{Z})(\check{\mathbf{u}} - \mathbf{u}^*) \\ &= -\mathcal{Z}^T \mathcal{Q} \mathcal{Z} - w_M^2(\mathcal{Z}) - \mathbf{u}^{*T} \mathfrak{R} \mathbf{u}^* + \rho \mathcal{J}^*(\mathcal{Z}) - 2\mathbf{u}^{*T} \mathfrak{R}(\check{\mathbf{u}} - \mathbf{u}^*). \end{aligned} \quad (33)$$

Through appropriate transformations, Equation (33) becomes

$$\dot{\mathcal{L}}_1 \leq -\mathcal{Z}^T \mathcal{Q} \mathcal{Z} - w_M^2(\mathcal{Z}) + \rho \mathcal{J}^*(\mathcal{Z}) + \|\mathfrak{R}\| \|\check{\mathbf{u}} - \mathbf{u}^*\|^2. \quad (34)$$

According to the Young's inequality and Assumption 3.1, we have

$$\begin{aligned} &\|\check{\mathbf{u}} - \mathbf{u}^*\|^2 \\ &\leq 2\|\check{\mathbf{u}}^* - \mathbf{u}^*\|^2 + 2\|\check{\mathbf{u}} - \check{\mathbf{u}}^*\|^2 \\ &\leq 2\nu \|e_T\|^2 + 2 \left\| \frac{1}{2} \mathfrak{R}^{-1} \mathcal{G}^T(\check{\mathcal{Z}}_j) \nabla \phi_c(\check{\mathcal{Z}}_j) (\varpi_c - \hat{\varpi}_c) + \frac{1}{2} \mathfrak{R}^{-1} \mathcal{G}^T(\check{\mathcal{Z}}_j) \epsilon_c(\check{\mathcal{Z}}_j) \right\|^2 \\ &\leq 2\nu \|e_T\|^2 + b_G^2 b_\epsilon^2 \|\mathfrak{R}^{-1}\|^2 + b_G^2 b_\phi^2 \|\mathfrak{R}^{-1}\|^2 \|\tilde{\omega}_c\|^2. \end{aligned} \quad (35)$$

From Equations (34) and (35), it is derived that

$$\begin{aligned} \dot{\mathcal{L}}_1 &\leq -(1 - \beta_1)(1 - \beta_2) \check{\mathcal{Z}}_j^T \mathcal{Q} \check{\mathcal{Z}}_j - (1 - \beta_1) \beta_2 \check{\mathcal{Z}}_j^T \mathcal{Q} \check{\mathcal{Z}}_j + \gamma_1 + \gamma_2 \|\tilde{\omega}_c\|^2 + b_\rho \\ &\quad + \left(2\nu \|\mathfrak{R}\| + \left(\frac{1}{\beta_1} - 1 \right) \lambda_M(\mathcal{Q}) \right) \|e_T\|^2, \end{aligned} \quad (36)$$

where $\gamma_1 = b_G^2 b_\epsilon^2 \|\mathfrak{R}\| \|\mathfrak{R}^{-1}\|^2$, $\gamma_2 = b_G^2 b_\phi^2 \|\mathfrak{R}\| \|\mathfrak{R}^{-1}\|^2$, and b_ρ denotes the bound of $\rho \mathcal{J}^*(\mathcal{Z})$.

Based on Equation (30), one can obtain

$$\begin{aligned} \dot{\mathcal{L}}_3 &= -\ell_c \tilde{\omega}_c^T \bar{\varrho} \bar{\varrho}^T \tilde{\omega}_c + \ell_c \tilde{\omega}_c^T \check{\varrho} e_c \\ &\leq -\frac{1}{2} \ell_c \lambda_m(\bar{\varrho} \bar{\varrho}^T) \|\tilde{\omega}_c\|^2 + \frac{1}{2} \ell_c b_e^2. \end{aligned} \quad (37)$$

According to Equations (36) and (37), it follows

$$\begin{aligned} \dot{\mathcal{L}}_c &\leq -(1 - \beta_1)(1 - \beta_2) \lambda_m(\mathcal{Q}_e) \|\check{e}_{zj}\|^2 - \left(\frac{1}{2} \ell_c \lambda_m(\bar{\varrho} \bar{\varrho}^T) - \gamma_2 \right) \|\tilde{\omega}_c\|^2 \\ &\quad + \frac{1}{2} \ell_c b_e^2 + b_\rho + \gamma_1. \end{aligned} \quad (38)$$

Situation ii. Consider the event occurrence scenario when $t = \varsigma_{j+1}$. Leveraging limit properties with Situation i analysis, we can derive $\Delta \mathcal{L}_c < 0$. The proof is completed.

Remark 3.2. The condition (31) is essential for the successful operation of the ETM. When this inequality fails to hold, an event is triggered with subsequent control law update. During the triggering interval, the updated control policy remains unchanged due to the zero-order holder. The parameters β_1 and β_2 can be used to regulate the triggering

threshold value and triggering frequency. Unlike conventional triggering conditions, the proposed triggering threshold relies only on system state information at the triggering instants, thereby avoiding persistent evaluation of triggering threshold and further alleviating computational load.

Remark 3.3. The results of Theorem 3.1 can be generalized to the case of uncertain system (13). Leveraging the boundedness of disturbances that $d_w^T(\mathcal{Z})\mathfrak{R}d_w(\mathcal{Z}) \leq w_M^2(\mathcal{Z})$ and differentiating $\mathcal{J}^*$, one can deduce that both critic weight error and tracking error for system (13) are UUB.

Next, the STM is constructed and Zeno is proved to be excluded. Assumption 3.3 is required [25].

Assumption 3.3. The term $\mathcal{G}(\mathcal{Z})d_w(\mathcal{Z})$ is norm-bounded with positive constant b_{gd} .

Based on Assumption 3.3, and the properties of $\mathcal{F}(\mathcal{Z})$ and $\mathcal{G}(\mathcal{Z})$, one can deduce

$$\|\dot{\mathcal{Z}}\| \leq \gamma_F \|\mathcal{Z}\| + \gamma_G \|\mathbf{u}^*(\check{\mathcal{Z}}_j)\| + b_{gd}. \quad (39)$$

According to Assumption 3.1, we have

$$\|\mathbf{u}^*(\check{\mathcal{Z}}_j)\| = \|\mathbf{u}^*(\mathcal{Z}) - \mathbf{u}^*(\check{\mathcal{Z}}_j) - \mathbf{u}^*(\mathcal{Z})\| \leq \sqrt{\nu} \|e_T(t)\| + \|\mathbf{u}^*(\mathcal{Z})\|. \quad (40)$$

Recalling the definition of triggering error, one can derive $\|\dot{\mathcal{Z}}\| = \|\dot{e}_T\|$. In light of Assumption 3.2, it can be deduced that $\mathbf{u}^*$ is bounded with Ξ_1 . Then it follows

$$\begin{aligned} \|\dot{e}_T\| &= \|\dot{\mathcal{Z}}\| \\ &\leq \gamma_F \|\mathcal{Z}\| + \sqrt{\nu} \gamma_G \|e_T(t)\| + \gamma_G \Xi_1 + b_{gd} \\ &\leq \gamma_F \|\mathcal{Z}\| + (\gamma_F + \sqrt{\nu} \gamma_G) \|e_T(t)\| + \Xi_2, \end{aligned} \quad (41)$$

where $\Xi_2 = \gamma_G \Xi_1 + b_{gd}$. Applying comparison lemma [27], we derive

$$\Delta_j = \varsigma_{j+1} - \varsigma_j > \frac{1}{\gamma_F + \sqrt{\nu} \gamma_G} \ln \left(1 + \left(\frac{\gamma_F + \sqrt{\nu} \gamma_G}{\gamma_F \|\check{\mathcal{Z}}_j\| + \Xi_2} \right) \sqrt{\mathcal{T}_D} \right) > 0. \quad (42)$$

Inequality (42) reveals that Zeno behavior can be avoided. Thus, the self-triggered rule is constructed as

$$\varsigma_{j+1} = \varsigma_j + \Delta_j. \quad (43)$$

Theorem 3.2. Suppose Assumptions 3.1-3.3 are satisfied. The critic weight vector and the control policy are updated utilizing Equations (27) and (25). The self-triggered rule (43) is applied, and then the tracking error and critic weight error of the closed-loop systems are guaranteed to be UUB.

Proof: According to the self-triggered rule (43), we can derive that self-triggered mechanism is more conservative when compared to the event-triggered mechanism (31). Thus, it is not difficult to prove this theorem based on the analysis of Theorem 3.1.

Remark 3.4. Although the event-triggered mechanism conserves resources via aperiodic control policy updates, it necessitates continuous state monitoring by hardware components and incurs additional computational overhead from persistent triggering-condition evaluation. The self-triggered learning-based approach can overcome this difficulty through the proactive prediction of triggering instants.

4. Numerical Example. In this section, the experiment is simulated to validate the algorithm's efficacy. For the critic network, similar to the vector in [5], the activation function vector with a dimensionality of 51 is constructed by multivariate polynomials. The initial critic weight is set as $\hat{w}_c(0) = 130 * \mathbf{1}_{51}$, where $\mathbf{1}_{51}$ is a 51-dimensional column vector whose entries are all equal to one. The initial state is set as $x_0 = [0.5, 1, 1, 0, 0, 0]^T$. The reference signal is selected as $c_r(x_r(t)) = [x_{r4}, x_{r5}, x_{r6}, -4x_{r1}, -0.25x_{r2}, -x_{r3}]^T$. For the MWMR, the parameters are set as $m = 12$, $r = 0.05$, $k = 0.6$, $J_\omega = 2.16$, $J_\phi = 0.1$, $\chi = 0.05$, $\mu = 1$. The other parameters are set as follows, $\beta_1 = 0.05$, $\beta_2 = 0.5$, $Q_e = 100\mathbf{I}$, $\mathfrak{R} = \mathbf{I}$, $\nu = 20$, $\rho = 0.015$, $\gamma_F = 2$, $\gamma_G = 1$.

After constructing the augmented system, the proposed learning method was implemented for 150 seconds. With a sampling period of 0.0001 seconds, the experimental results are presented in Figures 2-7. In Figures 2 and 3, $\mathcal{Z}_7$ to $\mathcal{Z}_{12}$ denote the 7th to 12th elements of the augmented state $\mathcal{Z}$, and these elements are equal to the reference states. The figures show the trajectories of the tracking states and the reference states, indicating that the tracking states can follow the reference signals well. Given the high dimensionality of the evaluation network weights, we depict only the curves of the first twelve weights in Figure 4, from which the convergence of neural weights can be observed. Figure 5 shows the trajectories of the control inputs that drive the system states to accurately track the reference signals. In addition, during the learning phase, the control input signals exhibit significant fluctuations due to the injected noise signals. Following the learning phase,

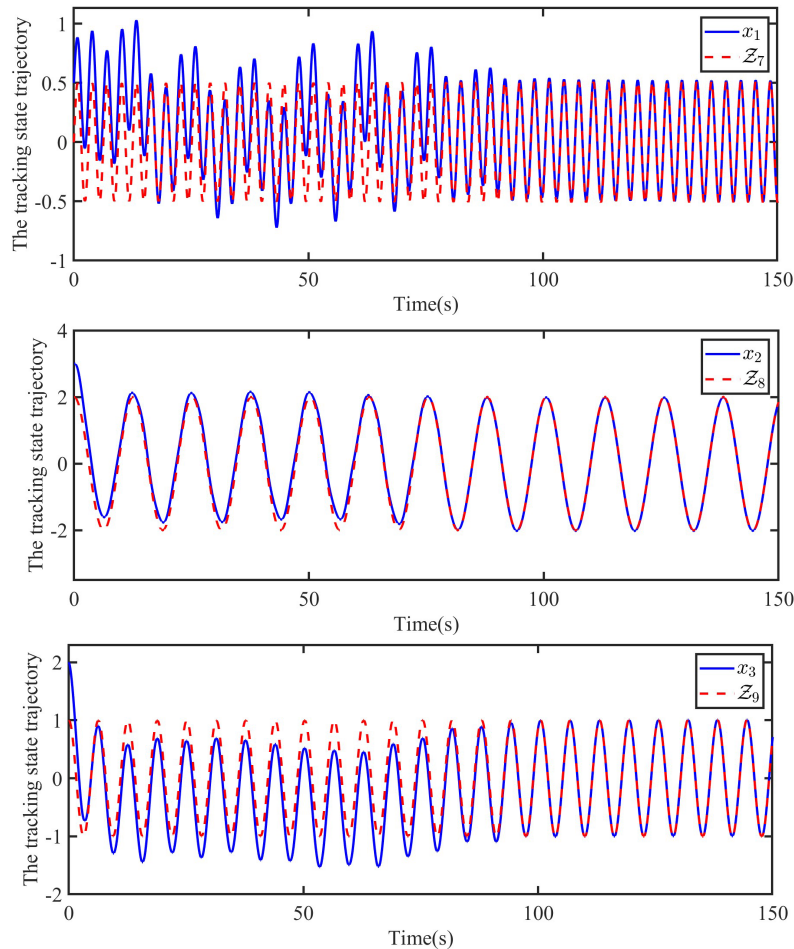


FIGURE 2. The evolutions for tracking states (I)

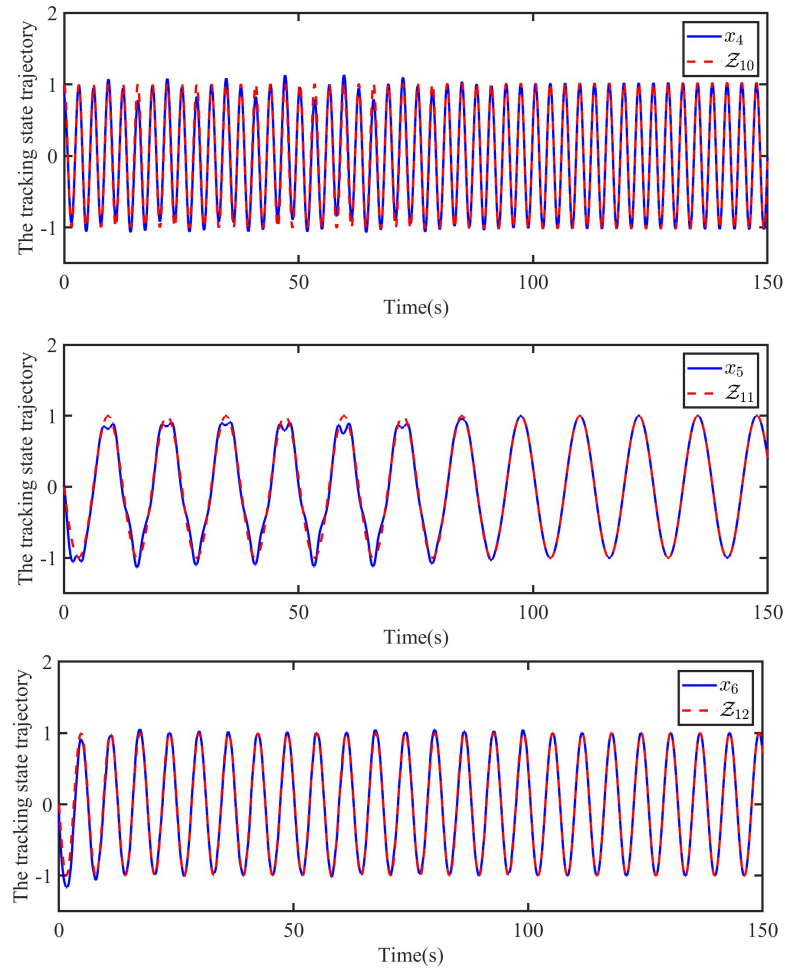


FIGURE 3. The evolutions for tracking states (II)

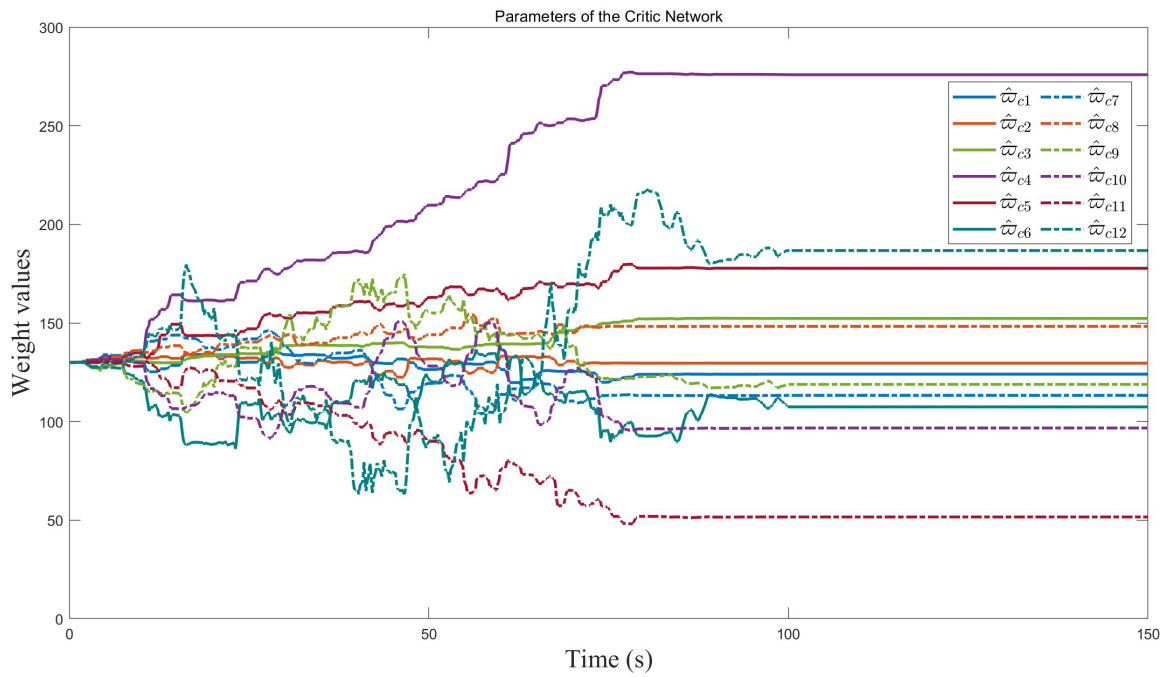


FIGURE 4. (color online) The evolutions of weights

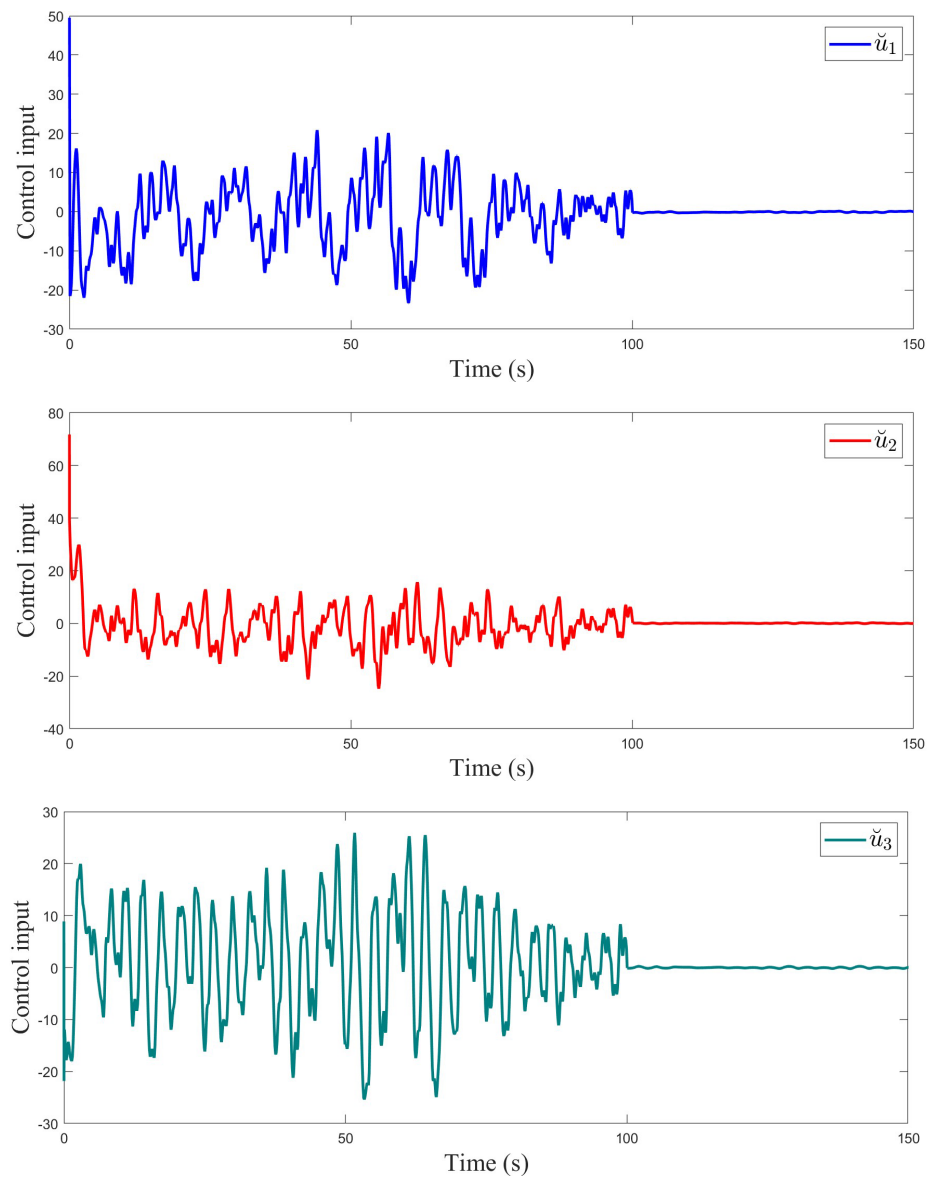


FIGURE 5. The evolutions of control inputs

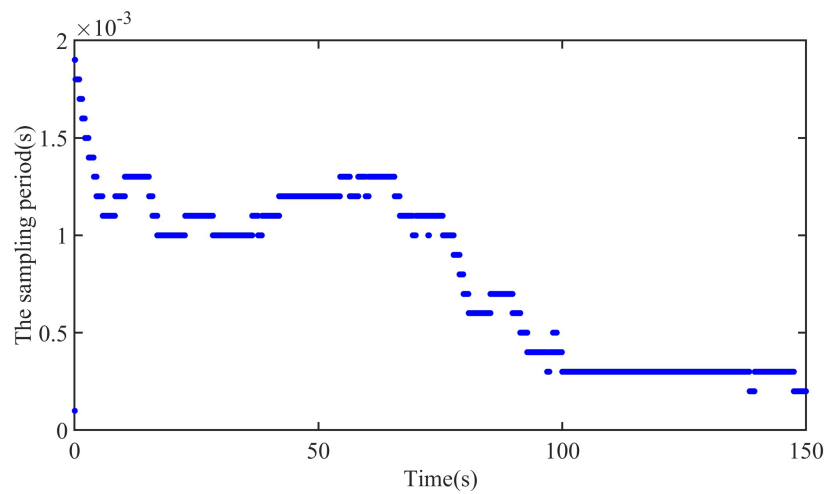


FIGURE 6. The triggering periods for STM

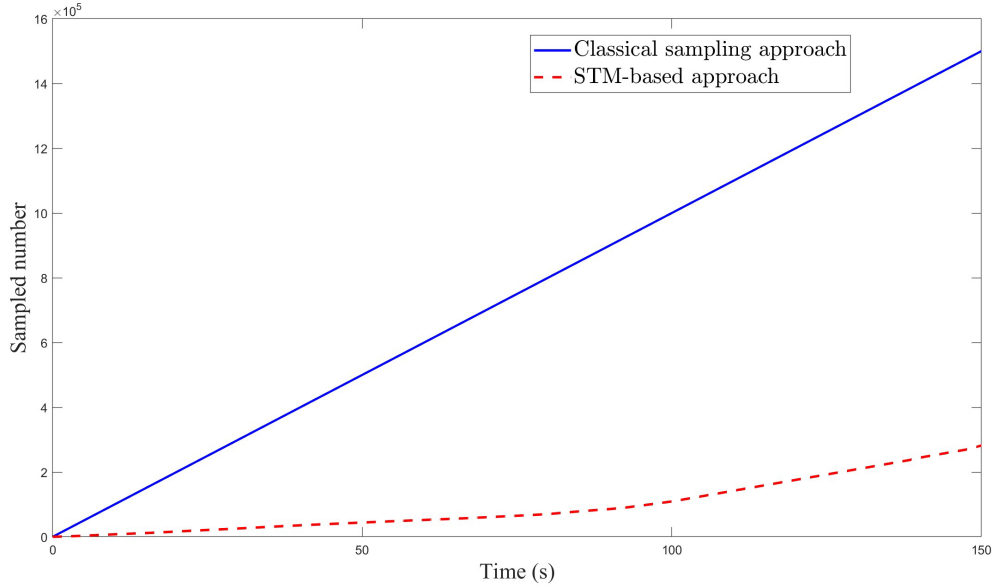


FIGURE 7. The comparison of sampling times

signals decay to smaller values. Figure 6 presents the aperiodic triggering periods which indicate the effectiveness of the STM. Comparative experiments were conducted to evaluate the triggering times of the proposed self-triggered learning method against those of the classic time-triggered reinforcement learning method, and the simulation results are presented in Figure 7. Similar to the comparative experimental design in [23, 26], all experimental setups and algorithm settings were identical, except for the sampling mechanism. By using only 281809 sets of data compared to the latter method's 1500000 sets during the experimental phase, the former method obviously reduced computational and communication resource consumption in the system. Furthermore, the self-triggered control method eliminates the need for continuous monitoring of system states by hardware devices through proactively predicting the next triggering instant.

5. Conclusion. In this paper, the self-triggered learning approach has been proposed to solve the robust tracking issues of MWMRs. Firstly, considering sliding effects in MWMRs featuring roller alignment angles of $\pi/3$, the model is established through integrated kinematic and dynamic equations. Through system transforms, the RTCI for uncertain system is converted into the optimal issue for the transformed system. Then, the critic-only learning-based method is developed to approximate the solution of HJBEs. The triggering condition is constructed such that the triggering threshold is computed using only the data sampled at triggering instants. Furthermore, the STM-based RL method is proposed to derive the robust optimal control policy. Finally, theoretical and experimental results validate the algorithm's efficacy. In the future, our work will focus on expanding the applications of reinforcement learning approaches in mobile robots and refining self-triggered rules to further improve the efficiency of adaptive learning methods.

Acknowledgment. This work was supported in part by the National Natural Science Foundation of China under Grant 62403086 and Grant U22A20101, in part by the National Key R&D Program of China under Grant 2022YFE0107300, in part by the Chongqing Natural Science Foundation under Grant CSTB2024NSCQ-MSX1053, Grant CSTB2024NSCQ-LZX0143, Grant CSTB2024NSCQ-QCXMX0052 and Grant CSTB2025YCJH-KY XM0023, in part by the Science and Technology Research Project of Chongqing Municipal Education Commission under Grant KJZD-M202500804, and in part by the High-level

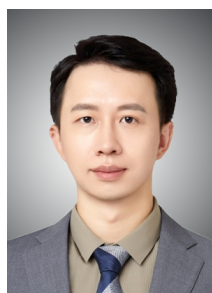
Talents Research Project of Chongqing Technology and Business University under Grant 2356021.

REFERENCES

- [1] B. Li, H. Liu, C. K. Ahn, C. Wang and X. Zhu, Fixed-time tracking control of wheel mobile robot in slipping and skidding conditions, *IEEE/ASME Transactions on Mechatronics*, vol.30, no.3, pp.1865-1875, 2025.
- [2] S. Miao and Y. Zhou, An identical path tracking control strategy of the tractor-trailer wheeled mobile robot with an off-axle hitching based on a passive steering angle, *Journal of the Franklin Institute*, vol.361, no.4, Article no.106634, 2024.
- [3] J. Bai, C. Liu, S. Zhou, Y. Chen and X. Li, Trajectory tracking control for wheeled mobile robot subject to slip and skidding, *Neurocomputing*, vol.638, Article no.130203, 2025.
- [4] L. Yu, J. Xiong and M. Xie, Iterative-learning-based tracking control of a two-wheeled mobile robot with model uncertainties and unknown periodic disturbances, *Journal of the Franklin Institute*, vol.361, no.11, Article no.106962, 2024.
- [5] D. Zhang, G. Wang and Z. Wu, Reinforcement learning-based tracking control for a three Mecanum wheeled mobile robot, *IEEE Transactions on Neural Networks and Learning Systems*, vol.35, no.1, pp.1445-1452, 2024.
- [6] P. Liu, H. Zhang and W. Ao, Event-triggered and self-triggered approximate optimal control for constrained-input nonlinear systems via reinforcement learning, *International Journal of Innovative Computing, Information and Control*, vol.21, no.3, pp.769-779, 2025.
- [7] X. Yang and D. Wang, Reinforcement learning for robust dynamic event-driven constrained control, *IEEE Transactions on Neural Networks and Learning Systems*, vol.36, no.4, pp.6067-6079, 2025.
- [8] B. A. Wallace and J. Si, Continuous-time reinforcement learning: New design algorithms with theoretical insights and performance guarantees, *IEEE Transactions on Neural Networks and Learning Systems*, vol.36, no.4, pp.6940-6954, 2025.
- [9] D. Wang, N. Gao, D. Liu, J. Li and F. L. Lewis, Recent progress in reinforcement learning and adaptive dynamic programming for advanced control applications, *IEEE/CAA Journal of Automatica Sinica*, vol.11, no.1, pp.18-36, 2024.
- [10] 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.
- [11] H. Dong, X. Zhao and H. Yang, Reinforcement learning-based approximate optimal control for attitude reorientation under state constraints, *IEEE Transactions on Control Systems Technology*, vol.29, no.4, pp.1664-1673, 2021.
- [12] 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.
- [13] D. Wang and C. Mu, Adaptive-critic-based robust trajectory tracking of uncertain dynamics and its application to a spring-mass-damper system, *IEEE Transactions on Industrial Electronics*, vol.65, no.1, pp.654-663, 2018.
- [14] L. Shi, X. Wang and Y. Cheng, Safe reinforcement learning-based robust approximate optimal control for hypersonic flight vehicles, *IEEE Transactions on Vehicular Technology*, vol.72, no.9, pp.11401-11414, 2023.
- [15] H. Zhang, X. Wang, B. Xiao, X. Wu and B. Li, Concurrent learning critic-only NN-based robust approximate optimal control of nonlinear systems with experimental verification, *IEEE Transactions on Industrial Electronics*, vol.72, no.8, pp.8492-8502, 2025.
- [16] C. Wang, H. Zhan, Q. Guo and T. Li, Adaptive dynamic programming-based fixed-time optimal control for wheeled mobile robot, *IEEE Robotics and Automation Letters*, vol.10, no.1, pp.176-183, 2025.
- [17] Z. Han, T. Pan and D. Xu, Event-triggered model-free adaptive integral sliding mode constrained maximum power point tracking control strategy for photovoltaic system, *International Journal of Innovative Computing, Information and Control*, vol.21, no.3, pp.649-665, 2025.
- [18] 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.

- [19] F. Wang, L. Long and C. Xiang, Event-triggered state-dependent switching for adaptive fuzzy control of switched nonlinear systems, *IEEE Transactions on Fuzzy Systems*, vol.32, no.4, pp.1756-1767, 2024.
- [20] J. Wang, Y. Zheng, J. Ding, H. Zhang and J. Sun, Memory-based event-triggered fault-tolerant consensus control of nonlinear multi-agent systems and its applications, *IEEE Transactions on Automation Science and Engineering*, vol.22, pp.7941-7954, 2025.
- [21] T. An, B. Dong, H. Yan, L. Liu and B. Ma, Dynamic event-triggered strategy-based optimal control of modular robot manipulator: A multiplayer nonzero-sum game perspective, *IEEE Transactions on Cybernetics*, vol.54, no.12, pp.7514-7526, 2024.
- [22] P. Liu, W. Ao, Z. Ming, G. Huang and Z. Liu, Dynamic event-triggered optimal tracking control for constrained nonlinear stochastic systems, *Journal of the Franklin Institute*, vol.360, no.2, pp.1145-1165, 2023.
- [23] C. Liu, Z. Chu, Z. Duan, H. Zhang and Z. Ma, Decentralized event-triggered tracking control for unmatched interconnected systems via particle swarm optimization-based adaptive dynamic programming, *IEEE Transactions on Cybernetics*, vol.54, no.11, pp.6895-6908, 2024.
- [24] Y. Shi, W. Xie and W. Zhang, Sectionalized self-triggered control for AUVs with deferred quantitative constraints, *IEEE Transactions on Instrumentation and Measurement*, vol.73, Article no.7500310, 2024.
- [25] M. Lin, B. Zhao and D. Liu, Event-triggered robust adaptive dynamic programming for multiplayer Stackelberg-Nash games of uncertain nonlinear systems, *IEEE Transactions on Cybernetics*, vol.54, no.1, pp.273-286, 2024.
- [26] 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.
- [27] H. K. Khalil, *Nonlinear Systems*, Prentice-Hall, Upper Saddle River, NJ, 2002.

Author Biography



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



Huiyan Zhang (Senior Member, IEEE) received the Ph.D. degree in Control Theory and Control Engineering from Harbin Institute of Technology, China, in 2019. 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, Australia. She is currently a Professor with Chongqing Technology and Business University, China. Her research interests include stochastic switched systems, intelligent control systems, and robust control. Dr. Zhang serves as an Associate Editor for several journals, including *IEEE Transactions on Cybernetics*, *Signal Processing*, and *Journal of the Franklin Institute*.



Wengang Ao received the B.S. degree in Engineering Mechanics from Sichuan University, China, in 2000, and M.E. degree in Solid Mechanics from Chongqing University, China, in 2007. He is currently a Professor at Chongqing Technology and Business University, China, and Co-Chair of the IEEE SMC Technical Committee on Embodied Intelligent Systems. His research interests include intelligent control, intelligent robotics, and engineering mechanics. In the past five years, he has led five national and provincial/ministerial-level research projects, and published over 30 academic papers.



Huafeng Qin received the B.Sc. degree from the School of Mathematics and Physics in 2006, the M.Eng. degree from the College of Electronic and Automation in 2009, both from Chongqing University of Technology, China, and the Ph.D. degree from the College of Opto-Electronic Engineering, Chongqing University, China, in 2012. He was a visiting student for 12 months with Nanyang Technological University, Singapore, and then a postdoctoral researcher for two years with Université Paris-saclay, France. Currently, he is a Professor with the International Science and Technology Cooperation Base of Intelligent Manufacturing Service, Chongqing Technology and Business University, China. His research interests include biometrics (e.g., vein, face, and gait) and machine learning.