Research Article

Open Access

Check for updates

Reinforcement learning-based optimal adaptive fuzzy control for nonlinear multi-agent systems with prescribed performance

Huarong Yue, Jianwei Xia (iD)

School of Mathematics Science, Liaocheng University, Liaocheng 252000, Shandong, China.

Correspondence to: Prof. Jianwei Xia, School of Mathematics Science, Liaocheng University, No.1, Hunan Road, Dongchangfu District, Liaocheng 252000, Shandong, China. E-mail: njustxjw@126.com

Abstract

In this paper, the problem of optimal adaptive consensus tracking control for nonlinear multi-agent systems with prescribed performance is investigated. To address the issue of satisfying the initial value conditions in existing results, an improved performance function is employed as the prescribed performance boundary, effectively resolving this problem. Then, by employing the error transformation function, the constrained system is converted into an unconstrained one. Furthermore, fuzzy logic systems are employed to identify unknown system parts. By applying the dynamic surface technique, the problem of "differential explosion", which often occurs in backstepping, is solved. Moreover, a distributed optimal adaptive fuzzy control protocol based on the reinforcement learning actor-critic algorithm is proposed. Under the proposed control scheme, it is proved that all the signals within the closed-loop system are bounded, and the consensus tracking errors have remained within the predefined bounds. Finally, the numerical simulation results demonstrate the effectiveness of the proposed scheme.

Keywords: Multi-agent system, reinforcement learning, prescribed performance, optimal consensus control, actor-critic structure

1. INTRODUCTION

Optimal control is achieved by designing a control protocol that not only achieves the system control objectives but also minimizes the system cost. The field of optimal control has garnered significant scholarly interest in recent years. The optimal controller can be deduced from the solution of the Hamilton-Jacobi-Bellman (HJB) equation[1]. For linear systems, this is actually solving the Riccati equation. However, for nonlinear systems, the HJB equation is a partial differential equation containing multiple nonlinear terms, and it is conceivably challenging to solve the equation directly. One approach that can be implemented is dynamic programming (DP)[2–5]. However, this approach becomes less feasible for high-dimensional systems since it is a backward, offline computational process, which significantly increases the computational complexity in high-dimensional scenarios. As a form of machine learning, reinforcement learning (RL) arguably opens up another avenue to solve the problem[6–9]. The most commonly used RL algorithms make use of the actor-critic structure, in which the actor interacts with the environment, and the critic evaluates the actions of the actor and provides feedback; in this way, the actor performs the next task again. Subsequently, RL has been employed in various nonlinear systems for adaptive control, leading to remarkable outcomes[10–14]. For example, an optimal adaptive controller for nonlinear systems with control gain functions was proposed to achieve not only tracking control but also the optimal performance of systems[10]. In[12], the problem of tracking control of nonlinear systems with input constraints was investigated. In[14], an optimal observer-based adaptive control scheme for nonlinear stochastic systems with input and state constraints was proposed.

In the meanwhile, since most physical models in practical applications can be represented by nonlinear systems, the study of nonlinear systems is very important and has yielded rich results[15–18]. In recent years, multi-agent systems (MASs) have garnered significant attention from scholars due to their capability to perform tasks that surpass the capabilities of a single agent. The consensus problem in MASs refers to achieving a state of agreement or convergence among multiple agents through the design of control protocols, which is a fundamental problem in the design and control of MASs. Over the past decades, the problem of consensus control for MASs has been extensively studied, leading to significant advancements[19–30]. In[23], a consensus control scheme incorporating a modified disturbance observer was designed to achieve fixed-time tracking control of nonlinear MASs with unknown disturbances. In[27], an event-triggered control distributed scheme was proposed to address the problem of asymptotic tracking for nonlinear MASs with uncertain leaders. In recent times, there has been a surge of interest in incorporating RL into MASs. It is an interesting and challenging problem and has produced some excellent results[31–38]. For instance, in[33], an optimal backstepping consensus control protocol based on RL was introduced for nonlinear strict-feedback MASs, which not only exhibits algorithmic simplicity but also relaxes the need for two general conditions: known dynamics and persistence excitation. In[34], an optimal RL-based event-triggered controller was proposed for nonlinear stochastic systems.

On the other hand, the concept of prescribed performance control (PPC) has emerged as a prominent research topic in the control community, initially proposed by Bechlioulis and Rovithakis[39]. Transient and steady-state performance, which is often neglected by conventional control schemes that solely ensure closed-loop stability, is the primary concern of prescribed performance. The PPC strategy aims to align the actual system performance achieved after execution with the desired or prescribed performance criteria and has yielded remarkable outcomes[40–43]. By utilizing exponential performance functions, both the nonlinear switched systems in[40] and the non-triangularly structured systems in[42] were able to achieve the desired rate of convergence. To overcome the issue of "differential explosion" caused by repeated derivation, dynamic surface control schemes were proposed to implement the tracking control of systems in[41] and[43], respectively. Recently, there has been significant research focused on integrating RL into PPC[44,45]. However, it is noticed that all of the above results depend on the initial conditions; i.e., at the initial moment, the initial error needs to be made within a prescribed boundary by properly setting the initial values.

Motivated by the discussions above, this article focuses on the optimal adaptive consensus tracking control

problem for leader-follower nonlinear MASs subject to prescribed performance constraints. The main contributions of the article are as follows.

(1) Based on the actor-critic structure of RL, the proposed consensus tracking control scheme can achieve optimal control of MASs while an excellent tracking effect is guaranteed. Compared with [10–12], the proposed algorithm is simpler to implement since it does not require system dynamic and persistent excitation conditions.

(2) In contrast to existing performance functions [40–43], most of which rely on initial value conditions, an improved performance function is introduced such that the proposed consensus tracking control scheme is able to force the convergence of the consensus tracking error to a prescribed region without the requirement of initial value conditions.

(3) Compared with the traditional backstepping control scheme [24,25], the dynamic surface technique is adopted, which effectively avoids the problem of "differential explosion" caused by multiple derivations of the virtual controller and makes the control structure simpler.

2. PRELIMINARIES AND PROBLEM FORMULATION

2.1 Topology theory
In this paper, information interactions between agents are inscribed by a directed graph $\mathscr{G} = (\mathscr{H}, \mathscr{T}, \mathscr{A})$ in which $\mathscr{H} = \{1, \cdots, N\}$ and $\mathscr{T} \subseteq \mathscr{H} \times \mathscr{H}$ denote the set of notes and the set of edges, respectively. Furthermore, $\mathscr{A} = [a_{ij}] \in \mathcal{R}^{N \times N}$ denotes the adjacency matrix, all elements of which are non-negative, specifically $a_{ij} > 0$ if $(j, i) \in \mathscr{T}$ and $a_{ij} = 0$ otherwise. If the agent i has access to the information of the agent j, there is $(j, i) \in \mathscr{T}$, and hence, the neighbors of the agent i can be described as $\mathscr{N}_i = \{j | (j, i) \in \mathscr{T}\}$. The in-degree and Laplace matrix are defined as $\mathscr{D} = diag\{d_1, \cdots, d_N\}$ with $d_i = \sum_{j=1}^{N} a_{ij}$ and $\mathscr{L} = \mathscr{D} - \mathscr{A}$. Similarly, if the agent i has access to the leader, there is $b_i = 1$; otherwise, $b_i = 0$, which forms the matrix $\mathscr{B} = diag\{b_1, \cdots, b_N\}$.

2.2 System formulation
Consider the following nonlinear MAS composed of N agents, where the ith agent can be modeled as:

$$\begin{cases} \dot{x}_{i,j} = x_{i,j+1}(t) + f_{i,j}(\bar{x}_{i,j}(t)), \\ \dot{x}_{i,n} = u_i(t) + f_{i,n}(\bar{x}_{i,n}(t)), \\ y_i = x_{i,1}, (j = 1, 2, \ldots, n-1) \end{cases} \tag{1}$$

where $i = 1, \ldots, N$, $\bar{x}_{i,j} = [x_{i,1}, x_{i,2}, \cdots, x_{i,j}]^T \in R^j$, $j = 1, \ldots, n$ and u_i denote the state variable and control input, respectively. $f_{i,j}$ represents the unknown nonlinear smooth function. $y_i = x_{i,1}$ denotes the output variable. c The consensus tracking error for agent i is defined as:

$$e_i = \sum_{j \in \mathscr{N}_i} a_{ij}(y_i - y_j) + b_i(y_i - y_r), \tag{2}$$

where y_r is generated by the output of the leader to represent the reference signal.

Assumption 1 [24] Let the leader be the root node and the directed graph have a spanning tree.

Assumption 2 The reference trajectory y_r and its derivatives $\dot{y}_r$ are continuous and bounded.

Lemma 1 [25] If a function $F(x)$ is continuous on a compact set Φ, then for any given accuracy $\varepsilon > 0$, there exists a fuzzy logic system (FLS) such that

$$\sup_{x \in \Phi} |F(x) - \vartheta^T \phi(x)| \le \varepsilon,$$

where $\vartheta = [\vartheta_1, \cdots, \vartheta_L]^T$ denotes the ideal weight vector and $L > 1$ indicates the number of fuzzy rules, and $\phi(x) = [\phi_1(x), \cdots, \phi_n(x)]^T / \sum_{r=1}^{L} \phi_r(x)$ denotes the basis function vector with $\phi_r(x)$ being a Gaussian function, i.e., for $r = 1, \cdots, L, \phi_r(x) = \exp[-(x - J_r)^T(x - J_r)/\hbar_r^2]$, where $J_r = [J_{r1}, J_{r2}, \cdots, J_{rn}]^T$ is defined as the centre vector, and $\hbar_r$ is defined as the width of the Gaussian function.

2.3 Error transformation

Define the following monotonically increasing function over the interval $(-1, 1)$

$$\Gamma(x) = \frac{\sqrt{\chi}x}{\sqrt{1 - x^2}}, \tag{3}$$

where χ is a constant.

Define the monotonically increasing function $\omega_i(t)$, which has the following properties:
(1) $\omega_i(0) = 1$ and $\lim_{t \to \infty} \omega_i(t) = \frac{1}{l}$, with $0 < l_i \leq 1$ being a constant.
(2) ω_i is differentiable up to order n and $\omega_i^r, r = 1, \cdots, n$ is bounded.

To achieve a desired level of system performance, we define the following performance function

$$\Gamma(\varphi_i) = \frac{\sqrt{\chi}\varphi_i}{\sqrt{1 - \varphi_i^2}}, \tag{4}$$

where $\varphi_i = \frac{1}{\omega_i(t)}$. Then, our control objective is to make the consensus tracking error of the system satisfy $\Gamma(-\varphi_i(t)) \leq e_i \leq \Gamma(\varphi_i(t))$.

Define the following normalized function

$$\xi_i(e_i) = \frac{e_i}{\sqrt{\chi^2 + e_i^2}}, \tag{5}$$

From the above equation, we know that there exists a constant $\bar{\xi}$ satisfying $|\xi_i(e_i)| \leq \bar{\xi} < 1$, in which case $e_i = \frac{\xi_i\sqrt{\chi}}{\sqrt{1-\xi_i}}$ is bounded.

The following error transformation function is introduced

$$\zeta_i(t) = \frac{\varpi_i(t)}{1 - \varpi_i^2(t)}, \tag{6}$$

where $\varpi_i(t) = \omega_i(t)\xi_i(e_i)$ with $\varpi_i(0) = \omega_i(0)\xi_i(e_i(0)) = \xi_i(e_i(0) \in (-1, 1)$. Therefore, it is clear from the above equation that we can deduce that for any initial value of $\varpi_i(0) \in (-1, 1)$, $\zeta_i(t) \to \pm\infty$ if and only if $\varpi_i \to -1$ or $\varpi_i \to 1$. This also implies that as long as ζ_i is bounded, there exists a constant $\bar{\varpi}$ satisfying $|\varpi_i| \leq \bar{\varpi} < 1$.

From the above definition, it has

$$-\varphi_i = -\frac{1}{\omega_i} < -\frac{\bar{\varpi}}{\omega_i} \leq \frac{\bar{\varpi}}{\omega_i} < \frac{1}{\omega_i} < \varphi_i. \tag{7}$$

Since $F(x)$ is a monotonically increasing function, it follows that

$$\Gamma(-\varphi_i) < \Gamma(\xi_i) < \Gamma(\varphi_i). \tag{8}$$

Noting that $\Gamma(\xi_i) = \frac{\sqrt{\chi}\xi_i}{\sqrt{1 - \xi_i^2}}$, we can get $\Gamma(-\varphi_i(t)) \leq e_i \leq \Gamma(\varphi_i(t))$.

3. DESIGN PROCEDURE AND MAIN RESULTS

The controller design is conducted based on the following coordinate transformations

$$z_{i,j} = x_{i,j} - \bar{\alpha}_{i,j}, \tag{9}$$

$$\rho_{i,j} = \bar{\alpha}_{i,j} - \hat{\alpha}_{i,j-1}, j = 2, \dots, n, \tag{10}$$

where $\hat{\alpha}_{i,j-1}$ denotes the approximate optimal virtual controller, $\bar{\alpha}_{i,j}$ denotes the output of the first-order filter, and $\rho_{i,j}$ is defined as the filtering error. The filtering dynamic is introduced as

$$\kappa_{i,j-1}\dot{\bar{\alpha}}_{i,j} + \bar{\alpha}_{i,j} = \hat{\alpha}_{i,j-1}, \quad \bar{\alpha}_{i,j}(0) = \hat{\alpha}_{i,j-1}(0), \tag{11}$$

where $\kappa_{i,j-1}$ is a positive constant.

To facilitate brevity, the following definitions are provided before the commencement of the design steps. For $i = 1, \cdots, N, j = 1, \cdots, n, \tilde{\theta}_{a,ij} = \theta_{a,ij} - \hat{\theta}_{a,ij}, \tilde{\theta}_{c,ij} = \theta_{c,ij} - \hat{\theta}_{c,ij}, \tilde{\vartheta}_{i,j} = \vartheta_{i,j} - \hat{\vartheta}_{i,j}$ where $\hat{\theta}_{a,ij}, \hat{\theta}_{c,ij}$, and $\hat{\vartheta}_{i,j}$ are the estimations of $\theta_{a,ij}, \theta_{c,ij}$, and $\vartheta_{i,j}$, respectively. $\lambda_{i,j}^{\min}$ denotes the minimum eigenvalue of $S_{i,j}S_{i,j}^T$. The set Ω is the tight set that contains zero, and $\psi(\Omega)$ represents the admissible control. The adaptive laws of actor FLSs and critic FLSs, adaptive law $\hat{\vartheta}_{i,j}$ are designed to be

$$\dot{\hat{\theta}}_{a,ij} = -S_{i,j}S_{i,j}^T(\sigma_{a,ij}(\hat{\theta}_{a,ij} - \hat{\theta}_{c,ij}) + \sigma_{c,ij}\hat{\theta}_{c,ij}), \tag{12}$$

$$\dot{\hat{\theta}}_{c,ij} = -\sigma_{c,ij}S_{i,j}S_{i,j}^T\hat{\theta}_{c,ij}, \tag{13}$$

$$\dot{\hat{\vartheta}}_{i,1} = \zeta_i\phi_{i,1} - \sigma_{i,1}\hat{\vartheta}_{i,1}, \tag{14}$$

$$\dot{\hat{\vartheta}}_{i,j} = z_{i,j}\phi_{i,j} - \sigma_{i,j}\hat{\vartheta}_{i,j}, j = 2, \cdots, n \tag{15}$$

where $\sigma_{a,ij}, \sigma_{c,ij}$ and $\sigma_{i,j}$ are design parameters.

Step 1: Derivation with respect to e_i gives

$$\dot{e}_i = (d_i + b_i)(x_{i,2} + f_{i,1}) - \sum_{j \in \mathcal{N}_i} a_{i,j}\dot{y}_j - b_i\dot{y}_r. \tag{16}$$

Define the performance index function for the first subsystem of the agent i as

$$V_{i,1}(\zeta_i) = \int_t^\infty c_{i,1}(\zeta_i(\tau), \alpha_{i,1}(\zeta))d\tau, \tag{17}$$

where $c_{i,1}(\zeta_i, \alpha_{i,1}) = \zeta_i^2 + \alpha_{i,1}^2$ represents the value function.

The optimal performance index function is expressed as

$$V_{i,1}^*(\zeta_i) = \int_t^\infty c_{i,1}(\zeta_i(\tau), \alpha_{i,1}^*(\zeta))d\tau$$

$$= \min_{\alpha_{i,1}^* \in \psi(\Omega)} \left(\int_t^\infty c_{i,1}(\zeta_i(\tau), \alpha_{i,1}(\zeta))d\tau \right). \tag{18}$$

Considering $x_{i,2}$ as the optimal virtual controller $\alpha_{i,1}^*$, then the HJB equation is given as

$$H_{i,1}(\zeta_i, \alpha_{i,1}^*, V_{i,1}^*) = c_{i,1}(\zeta_i, \alpha_{i,1}^*) + \frac{\partial V_{i,1}^*}{\partial \zeta_i}\dot{\zeta}_i$$

$$= \zeta_i^2 + \alpha_{i,1}^{*2} + \frac{\partial V_{i,1}^*}{\partial \zeta_i}[\eta_i((d_i + b_i)(\alpha_{i,1}^* + f_{i,1}) - \sum_{j \in \mathcal{N}_i} a_{i,j}\dot{y}_j - b_i\dot{y}_r) + v_i] = 0, \tag{19}$$

where $\eta_i = \dfrac{\chi}{\sqrt{(e_i + \chi)(e_i + \chi)}}$, $v_i = \dfrac{1 + \varpi_i^2}{(1 - \varpi_i^2)^2}$.

Define the Bellman residual as

$$\begin{aligned}
\Theta &= H_{i,1}(\zeta_i, \hat{\alpha}_{i,1}^*, \hat{V}_{i,1}^*) - H_{i,1}(\zeta_i, \alpha_{i,1}^*, V_{i,1}^*) \\
&= H_{i,1}(\zeta_i, \hat{\alpha}_{i,1}^*, \hat{V}_{i,1}^*).
\end{aligned} \tag{20}$$

The approximate optimal virtual controller is expected to guarantee that Θ tends to zero. The following positive function is given by

$$I_{i,1} = (\hat{\theta}_{a,i1} - \hat{\theta}_{c,i1})^T (\hat{\theta}_{a,i1} - \hat{\theta}_{c,i1}). \tag{21}$$

Since equation $H_{i,1}(\zeta_i, \alpha_{i,1}^*, V_{i,1}^*) = 0$ has a unique solution, it is easy to deduce that $I_{i,1} = 0$ is equivalent to $\dfrac{\partial H_{i,1}(\zeta_i, \alpha_{i,1}^*, V_{i,1}^*)}{\partial \hat{\theta}_{a,i1}} = 0$. Then, from equations (12) and (13), we get

$$\begin{aligned}
\dot{I}_{i,1} &= \frac{\partial I_{i,1}}{\partial \hat{\theta}_{a,i1}} \dot{\hat{\theta}}_{a,i1} + \frac{\partial I_{i,1}}{\partial \hat{\theta}_{c,i1}} \dot{\hat{\theta}}_{c,i1} \\
&= -\frac{\sigma_{a,i1}}{2} \frac{\partial I_{i,1}}{\partial \hat{\theta}_{a,i1}} S_{i,1} S_{i,1}^T \frac{\partial I_{i,1}}{\partial \hat{\theta}_{c,i1}} \leq 0.
\end{aligned} \tag{22}$$

Thus, the designed adaptive laws $\dot{\hat{\theta}}_{a,i1}$ and $\dot{\hat{\theta}}_{c,i1}$ enable $H_{i,1}(\zeta_i, \alpha_{i,1}^*, V_{i,1}^*) = 0$ to be satisfied.

By calculating $\dfrac{\partial H_{i,1}}{\partial \alpha_{i,1}^*} = 0$, it yields

$$\alpha_{i,1}^* = -\frac{\eta_i(d_i + b_i)}{2} \frac{\partial V_{i,1}^*}{\partial \zeta_i}. \tag{23}$$

To obtain the optimal virtual controller, we decompose $\dfrac{\partial V_{i,1}^*}{\partial \zeta_i}$ to derive the following equation

$$\frac{\partial V_{i,1}^*}{\partial \zeta_i} = \frac{2c_{i,1}\zeta_i}{(d_i + b_i)^2 \eta_i^2} + \frac{1}{(d_i + b_i)^2 \eta_i^2} V_{i,1}^o + \frac{2}{(d_i + b_i)^2 \eta_i^2} \vartheta_{i,1}^{*T} \phi_{i,1}, \tag{24}$$

$$V_{i,1}^o = (d_i + b_i)^2 \eta_i^2 \frac{\partial V_{i,1}^*}{\partial \zeta_i} - 2c_{i,1}\zeta_i - 2\vartheta_{i,1}^{*T} \phi_{i,1}. \tag{25}$$

where $c_{i,1}$ is a design parameter.

Substituting equation (24) into equations (23), it follows that

$$\alpha_{i,1}^* = -\frac{c_{i,1}}{(d_i + b_i)\eta_i} \zeta_i - \frac{1}{2(d_i + b_i)\eta_i} V_{i,1}^o - \frac{1}{(d_i + b_i)\eta_i} \vartheta_{i,1}^{*T} \phi_{i,1}. \tag{26}$$

Since $V_{i,1}^o$ is an unknown term, by applying Lemma 2, there exists an FLS such that

$$V_{i,1}^o = \theta_{i,1}^{*T} S_{i,1} + \varepsilon_{i,1}, \tag{27}$$

where $\theta_{i,1}^*$ refers to the optimal weights, $S_{i,1}$ refers to the basis function, and $\varepsilon_{i,1}$ is the approximation error.

Substituting equation (27) into equations (24) and (26) results in

$$\alpha_{i,1}^* = -\frac{c_{i,1}}{(d_i+b_i)\eta_i}\zeta_i - \frac{1}{2(d_i+b_i)\eta_i}(\theta_{i,1}^{*T}S_{i,1}+\varepsilon_{i,1}) - \frac{1}{(d_i+b_i)\eta_i}\vartheta_{i,1}^{*T}\phi_{i,1}, \tag{28}$$

$$\frac{\partial V_{i,1}^*}{\partial \zeta_i} = \frac{2c_{i,1}\zeta_i}{(d_i+b_i)^2\eta_i^2} + \frac{1}{(d_i+b_i)^2\eta_i^2}(\theta_{i,1}^{*T}S_{i,1}+\varepsilon_{i,1}) + \frac{2}{(d_i+b_i)\eta_i}\vartheta_{i,1}^{*T}\phi_{i,1}. \tag{29}$$

The $\hat{\theta}_{a,i1}$ and $\hat{\theta}_{c,i1}$ of the actor FLS and the critic FLS are used to estimate the unknown weights $\theta_{i,1}^*$, respectively, to obtain

$$\hat{\alpha}_{i,1} = -\frac{c_{i,1}}{(d_i+b_i)\eta_i}\zeta_i - \frac{1}{2(d_i+b_i)\eta_i}\hat{\theta}_{a,i1}^T S_{i,1} - \frac{1}{(d_i+b_i)\eta_i}\hat{\vartheta}_{i,1}^T\phi_{i,1}, \tag{30}$$

$$\frac{\partial \hat{V}_{i,1}}{\partial \zeta_i} = \frac{2c_{i,1}\zeta_i}{(d_i+b_i)^2\eta_i^2} + \frac{1}{(d_i+b_i)^2\eta_i^2}\hat{\theta}_{c,i1}^T S_{i,1} + \frac{2}{(d_i+b_i)\eta_i}\hat{\vartheta}_{i,1}^T\phi_{i,1}. \tag{31}$$

Construct the following Lyapunov candidate function

$$V_{i,1} = \frac{1}{2}\zeta_i^2 + \frac{1}{2}\tilde{\theta}_{a,i1}^T\tilde{\theta}_{a,i1} + \frac{1}{2}\tilde{\theta}_{c,i1}^T\tilde{\theta}_{c,i1} + \frac{1}{2}\tilde{\vartheta}_{i,1}^T\tilde{\vartheta}_{i,1}. \tag{32}$$

The derivation of $V_{i,1}$ yields

$$\dot{V}_{i,1} = \zeta_i[\eta_i((d_i+b_i)(\rho_{i,2}+\hat{\alpha}_{i,1}+z_{i,2}+f_{i,1}) - \sum_{j\in\mathcal{N}_i}a_{i,j}\dot{y}_j - b_i\dot{y}_r) + v_i]$$

$$+ \tilde{\theta}_{a,i1}^T S_{i,1}^T S_{i,1}^T(\sigma_{a,i1}(\hat{\theta}_{a,i1}-\hat{\theta}_{c,i1}) + \sigma_{c,i1}\hat{\theta}_{c,i1}) + \tilde{\theta}_{c,i1}^T\sigma_{c,i1}S_{i,1}S_{i,1}^T\hat{\theta}_{c,i1} - \tilde{\vartheta}_{i,1}^T\dot{\hat{\vartheta}}_{i,1}. \tag{33}$$

Define $F_{i,1}$ to be the

$$F_{i,1} = \eta_i(d_i+b_i)f_{i,1} - \eta_i\sum_{j\in\mathcal{N}_i}a_{i,j}\dot{y}_j - \eta_i b_i\dot{y}_r - v_i. \tag{34}$$

By Lemma 2, there exists an FLS approximation to $F_{i,1}$ that results in

$$F_{i,1} = \vartheta_{i,1}^{*T}\phi_{i,1} + \epsilon_{i,1}, \tag{35}$$

where $\vartheta_{i,1}^*$ is the ideal weight, $\phi_{i,1}$ is the basis function, and $\epsilon_{i,1}$ is the approximation error satisfying $\epsilon_{i,1} \le \bar{\epsilon}_{i,1}$ with $\bar{\epsilon}_{i,1} > 0$.

With the help of Young's inequality, one has

$$\zeta_{i,1}\epsilon_{i,1} \le \frac{1}{2}\zeta_i^2 + \frac{1}{2}\bar{\epsilon}_{i,1}^2, \tag{36}$$

$$\zeta_i\eta_i(d_i+b_i)z_{i,2} \le \frac{1}{2}\zeta_i^2 + \frac{1}{2}\eta_i^2(d_i+b_i)^2z_{i,2}^2, \tag{37}$$

$$\zeta_i\eta_i(d_i+b_i)\rho_{i,2} \le \frac{1}{2}\zeta_i^2 + \frac{1}{2}\eta_i^2(d_i+b_i)^2\rho_{i,2}^2, \tag{38}$$

Substituting equations (30), (36), (37), and (38) into equation (33) yields

$$\dot{V}_{i,1} \le -(c_{i,1}-\frac{3}{2})\zeta_i^2 + \frac{1}{2}\eta_i^2(d_i+b_i)^2z_{i,2}^2 + \frac{1}{2}\eta_i^2(d_i+b_i)^2\rho_{i,2}^2 + \frac{1}{2}\bar{\epsilon}_{i,1}^2 - \frac{1}{2}\zeta_i\hat{\theta}_{a,i1}S_{i,1}$$

$$+ \sigma_{i,1}\tilde{\vartheta}_{i,1}\hat{\vartheta}_{i,1} + \tilde{\theta}_{a,i1}S_{i,1}^T S_{i,1}^T(\sigma_{a,i1}(\hat{\theta}_{a,i1}-\hat{\theta}_{c,i1}) + \sigma_{c,i1}\hat{\theta}_{c,i1}) + \tilde{\theta}_{c,i1}^T\sigma_{c,i1}S_{i,1}S_{i,1}^T\hat{\theta}_{c,i1}. \tag{39}$$

According to Young's inequality, it can be derived that

$$-\frac{1}{2}\zeta_i\hat{\theta}_{a,i1}^T S_{i,1} \leq \frac{1}{4}\zeta_i^2 + \frac{1}{4}\hat{\theta}_{a,i1}^T S_{i,1}S_{i,1}^T\hat{\theta}_{a,i1}, \tag{40}$$

$$\sigma_{a,i1}\tilde{\theta}_{a,i1}^T S_{i,1}S_{i,1}^T\hat{\theta}_{a,i1}$$
$$= -\frac{\sigma_{a,i1}}{2}\tilde{\theta}_{a,i1}^T S_{i,1}S_{i,1}^T\tilde{\theta}_{a,i1} - \frac{\sigma_{a,i1}}{2}\hat{\theta}_{a,i1}^T S_{i,1}S_{i,1}^T\hat{\theta}_{a,i1} + \frac{\sigma_{a,i1}}{2}\theta_{a,i1}^T S_{i,1}S_{i,1}^T\theta_{a,i1}, \tag{41}$$

$$\sigma_{c,i1}\tilde{\theta}_{c,i1}^T S_{i,1}S_{i,1}^T\hat{\theta}_{c,i1}$$
$$= -\frac{\sigma_{c,i1}}{2}\tilde{\theta}_{c,i1}^T S_{i,1}S_{i,1}^T\tilde{\theta}_{c,i1} - \frac{\sigma_{c,i1}}{2}\hat{\theta}_{c,i1}^T S_{i,1}S_{i,1}^T\hat{\theta}_{c,i1} + \frac{\sigma_{c,i1}}{2}\theta_{c,i1}^T S_{i,1}S_{i,1}^T\theta_{c,i1}, \tag{42}$$

$$(\sigma_{c,i1} - \sigma_{a,i1})\tilde{\theta}_{a,i1}^T S_{i,1}S_{i,1}^T\hat{\theta}_{c,i1}$$
$$\leq \frac{(\sigma_{c,i1} - \sigma_{a,i1})}{2}\tilde{\theta}_{a,i1}^T S_{i,1}S_{i,1}^T\tilde{\theta}_{a,i1} + \frac{(\sigma_{c,i1} - \sigma_{a,i1})}{2}\hat{\theta}_{c,i1}^T S_{i,1}S_{i,1}^T\hat{\theta}_{c,i1}, \tag{43}$$

$$\sigma_{i,1}\tilde{\vartheta}_{i,1}\hat{\vartheta}_{i,1} \leq \frac{\sigma_{i,1}}{2}\vartheta_{i,1}^2 - \frac{\sigma_{i,1}}{2}\tilde{\vartheta}_{i,1}^2. \tag{44}$$

By means of equations $(40) - (44)$, $\dot{V}_{i,1}$ becomes

$$\dot{V}_{i,1} \leq -(c_{i,1} - \frac{7}{4})\zeta_i^2 + \frac{1}{2}\eta_i^2(d_i + b_i)^2 z_{i,2}^2 + \frac{1}{2}\eta_i^2(d_i + b_i)^2\rho_{i,2}^2$$
$$\frac{1}{2}\bar{\epsilon}_{i,1}^2 + \frac{\sigma_{i,1}}{2}\vartheta_{i,1}^2 - \frac{\sigma_{i,1}}{2}\hat{\vartheta}_{i,1}^2 + \frac{\sigma_{a,i1}}{2}\theta_{a,i1}^T S_{i,1}S_{i,1}^T\theta_{a,i1} + \frac{\sigma_{c,i1}}{2}\theta_{c,i1}^T S_{i,1}S_{i,1}^T\theta_{c,i1}$$
$$- (\frac{\sigma_{a,i1}}{2} - \frac{1}{4})\hat{\theta}_{a,i1}^T S_{i,1}S_{i,1}^T\hat{\theta}_{a,i1} - \frac{\sigma_{c,i1}}{2}\hat{\theta}_{c,i1}^T S_{i,1}S_{i,1}^T\hat{\theta}_{c,i1} - (\sigma_{a,i1} - \frac{\sigma_{c,i1}}{2})\tilde{\theta}_{a,i1}^T S_{i,1}S_{i,1}^T\tilde{\theta}_{a,i1}$$
$$- \frac{\sigma_{c,i1}}{2}\tilde{\theta}_{c,i1}^T S_{i,1}S_{i,1}^T\tilde{\theta}_{c,i1}. \tag{45}$$

Then, it can be further bounded as

$$\dot{V}_{i,1} \leq -(c_{i,1} - \frac{7}{4})\zeta_i^2 + \frac{1}{2}\eta_i^2(d_i + b_i)^2 z_{i,2}^2 + \frac{1}{2}\eta_i^2(d_i + b_i)^2\rho_{i,2}^2 + \Upsilon_{i,1}$$
$$- \frac{\sigma_{i,1}}{2}\tilde{\vartheta}_{i,1}^2 - (\sigma_{a,i1} - \frac{\sigma_{c,i1}}{2})\lambda_{i,1}^{\min}\tilde{\theta}_{a,i1}^T\tilde{\theta}_{a,i1} - \frac{\sigma_{c,i1}}{2}\lambda_{i,1}^{\min}\tilde{\theta}_{c,i1}^T\tilde{\theta}_{c,i1}, \tag{46}$$

where $\Upsilon_{i,1} = \frac{1}{2}\bar{\epsilon}_{i,1}^2 + \frac{\sigma_{i,1}}{2}\vartheta_{i,1}^2 + \frac{\sigma_{a,i1}}{2}\theta_{a,i1}^T S_{i,1}S_{i,1}^T\theta_{a,i1} + \frac{\sigma_{c,i1}}{2}\theta_{c,i1}^T S_{i,1}S_{i,1}^T\theta_{c,i1} - (\frac{\sigma_{a,i1}}{2} - \frac{1}{4})\hat{\theta}_{a,i1}^T S_{i,1}S_{i,1}^T\hat{\theta}_{a,i1} - \frac{\sigma_{c,i1}}{2}\hat{\theta}_{c,i1}^T S_{i,1}S_{i,1}^T\hat{\theta}_{c,i1}$.

Step j $(2 \leq j \leq n - 1)$: From equations (9) and (10), it holds that

$$\dot{z}_{i,j} = z_{i,j+1} + \rho_{i,j+1} + \alpha_{i,j} + f_{i,j} - \dot{\alpha}_{i,j+1}. \tag{47}$$

The performance index function is chosen to be

$$V_{i,j}(z_{i,j}) = \int_t^\infty c_{i,j}(z_{i,j}(\tau), \alpha_{i,j}(z))d\tau, \tag{48}$$

where $c_{i,j}(z_i, \alpha_{i,j}) = z_{i,j}^2 + \alpha_{i,j}^2$ represents the value function.

Considering $x_{i,j+1}$ as the optimal virtual controller $\alpha_{i,j}^*$, then the optimal performance index function has the form

$$V_{i,j}^*(z_{i,j}) = \int_t^\infty c_{i,j}(z_{i,j}(\tau), \alpha_{i,j}^*(z))d\tau$$
$$= \min_{\alpha_{i,j}^* \in \psi(\Omega)} (\int_t^\infty c_{i,j}(z_{i,j}(\tau), \alpha_{i,j}(z))d\tau). \tag{49}$$

The HJB equation can be written as

$$H_{i,j}(z_{i,j}, \alpha^*_{i,j}, V^*_{i,j}) = c_{i,j}(z_{i,j}, \alpha_{i,j}) + \frac{\partial V^*_{i,j}}{\partial z_{i,j}} \dot{z}_{i,j}$$

$$= z^2_{i,j} + \alpha^{*2}_{i,j} + \frac{\partial V^*_{i,j}}{\partial z_{i,j}} (\alpha^*_{i,j} + f_{i,j} - \dot{\alpha}_{i,j+1}) = 0. \tag{50}$$

By calculating $\frac{\partial H_{i,j}}{\partial \alpha^*_{i,j}} = 0$, it yields

$$\alpha^*_{i,j} = -\frac{1}{2} \frac{\partial V^*_{i,j}}{\partial z_{i,j}}. \tag{51}$$

Then, decompose $\frac{\partial V^*_{i,j}}{\partial z_{i,j}}$ into the following two parts

$$\frac{\partial V^*_{i,j}}{\partial z_{i,j}} = 2c_{i,j} z_{i,j} + V^o_{i,j} + 2\vartheta^{*T}_{i,j} \phi_{i,j}, \tag{52}$$

$$V^o_{i,j} = \frac{\partial V^*_{i,j}}{\partial z_{i,j}} - 2c_{i,j} z_{i,j} - 2\vartheta^{*T}_{i,j} \phi_{i,j}, \tag{53}$$

where $c_{i,j}$ is a design parameter.

From equation (52), it is easy to introduce

$$\alpha^*_{i,j} = -c_{i,j} z_{i,j} - \frac{1}{2} V^o_{i,1} - \vartheta^{*T}_{i,j} \phi_{i,j}. \tag{54}$$

Since $V^o_{i,j}$ is an unknown term, by applying Lemma 2, there exists an FLS such that

$$V^o_{i,j} = \theta^{*T}_{i,j} S_{i,j} + \varepsilon_{i,j}, \tag{55}$$

where $\theta^*_{i,j}$ refers to the optimal weights, $S_{i,j}$ refers to the basis function, and $\varepsilon_{i,j}$ is the approximation error.

According to the above equation, we can get

$$\alpha^*_{i,j} = -c_{i,j} z_{i,j} - \frac{1}{2} (\theta^{*T}_{i,j} S_{i,j} + \varepsilon_{i,j}) - \vartheta^{*T}_{i,j} \phi_{i,j}, \tag{56}$$

$$\frac{\partial V^*_{i,j}}{\partial z_{i,j}} = 2c_{i,j} z_{i,j} + (\theta^{*T}_{i,j} S_{i,j} + \varepsilon_{i,1}) + \vartheta^{*T}_{i,j} \phi_{i,j}. \tag{57}$$

The $\hat{\theta}_{a,ij}$ and $\hat{\theta}_{c,ij}$ of the actor FLS and the critic FLS are used to estimate the unknown weights $\theta^*_{i,j}$, respectively, to obtain

$$\hat{\alpha}_{i,j} = -c_{i,j} z_{i,j} - \frac{1}{2} \hat{\theta}^*_{a,ij} S_{i,j} - \hat{\vartheta}^{*T}_{i,j} \phi_{i,j}, \tag{58}$$

$$\frac{\partial \hat{V}_{i,j}}{\partial z_{i,j}} = 2c_{i,j} z_{i,j} + \hat{\theta}_{c,ij} S_{i,j} + 2\hat{\vartheta}^{*T}_{i,j} \phi_{i,j}. \tag{59}$$

The candidate Lyapunov function function is chosen as

$$V_{i,j} = V_{i,j-1} + \frac{1}{2} z^2_{i,j} + \frac{1}{2} \tilde{\theta}^T_{a,ij} \tilde{\theta}_{a,ij} + \frac{1}{2} \tilde{\theta}^T_{c,ij} \tilde{\theta}_{c,ij} + \frac{1}{2} \tilde{\vartheta}^T_{i,j} \tilde{\vartheta}_{i,j} + \frac{1}{2} \rho_{i,j} \rho_{i,j}. \tag{60}$$

Derivation of $V_{i,j}$ yields

$$
\begin{aligned}
\dot{V}_{i,j} &= \dot{V}_{i,j-1} + z_{i,j}(z_{i,j+1} + \rho_{i,j+1} + \hat{\alpha}_{i,j} + f_{i,j} - \dot{\alpha}_{i,j+1}) + \rho_{i,j}\dot{\rho}_{i,j} \\
&\quad + \tilde{\theta}_{a,ij}S_{i,j}^T S_{i,j}^T(\sigma_{a,ij}(\hat{\theta}_{a,ij} - \hat{\theta}_{c,ij}) + \sigma_{c,ij}\hat{\theta}_{c,ij}) + \tilde{\theta}_{c,ij}^T\sigma_{c,ij}S_{i,j}S_{i,j}^T\hat{\theta}_{c,ij} - \tilde{\vartheta}_{i,j}^T\dot{\hat{\vartheta}}_{i,j}.
\end{aligned}
\tag{61}
$$

From equations (10) and (11), it holds that

$$
\dot{\rho}_{i,j} = -\frac{\rho_{i,j}}{\kappa_i} - \dot{\hat{\alpha}}_{i,j-1}.
\tag{62}
$$

Define $M_i = -\dot{\hat{\alpha}}_{i,j-1}$, which is a bounded continuous function.

By Lemma 2, there exists an FLS satisfying

$$
F_{i,j} = \vartheta_{i,j}^{*T}\phi_{i,j} + \epsilon_{i,j},
\tag{63}
$$

where $F_{i,j} = f_{i,j} - \dot{\hat{\alpha}}_{i,j}$, $\vartheta_{i,j}^*$ is the ideal weight, $\phi_{i,j}$ is the basis function, and $\epsilon_{i,j}$ is the approximation error satisfying $\epsilon_{i,j} \le \bar{\epsilon}_{i,j}$ with $\bar{\epsilon}_{i,j} > 0$.

With the help of Young's inequality, the following inequality holds

$$
z_{i,j}\epsilon_{i,j} \le \frac{1}{2}z_{i,j}^2 + \frac{1}{2}\bar{\epsilon}_{i,j}^2,
\tag{64}
$$

$$
z_{i,j}z_{i,j+1} \le \frac{1}{2}z_{i,j}^2 + \frac{1}{2}z_{i,j+1}^2,
\tag{65}
$$

$$
z_{i,j}\rho_{i,j+1} \le \frac{1}{2}z_{i,j}^2 + \frac{1}{2}\rho_{i,j+1}^2,
\tag{66}
$$

$$
\rho_{i,j}M_i \le \frac{1}{2}\rho_{i,j}^2 + \frac{1}{2}M_i^2.
\tag{67}
$$

Similar to the first step, it has

$$
-\frac{1}{2}z_{i,j}\hat{\theta}_{a,ij}^T S_{i,j} \le \frac{1}{4}z_{i,j}^2 + \frac{1}{4}\hat{\theta}_{a,ij}^T S_{i,1}S_{i,j}^T\hat{\theta}_{a,ij},
\tag{68}
$$

$$
\begin{aligned}
&\sigma_{a,ij}\tilde{\theta}_{a,ij}^T S_{i,j}S_{i,j}^T\hat{\theta}_{a,ij} \\
&= -\frac{\sigma_{a,ij}}{2}\tilde{\theta}_{a,ij}^T S_{i,j}S_{i,j}^T\tilde{\theta}_{a,ij} - \frac{\sigma_{a,ij}}{2}\hat{\theta}_{a,ij}^T S_{i,j}S_{i,j}^T\hat{\theta}_{a,ij} + \frac{\sigma_{a,ij}}{2}\theta_{a,ij}^T S_{i,j}S_{i,j}^T\theta_{a,ij},
\end{aligned}
\tag{69}
$$

$$
\begin{aligned}
&\sigma_{c,ij}\tilde{\theta}_{c,ij}^T S_{i,j}S_{i,j}^T\hat{\theta}_{c,ij} \\
&= -\frac{\sigma_{c,ij}}{2}\tilde{\theta}_{c,ij}^T S_{i,j}S_{i,j}^T\tilde{\theta}_{c,ij} - \frac{\sigma_{c,ij}}{2}\hat{\theta}_{c,ij}^T S_{i,j}S_{i,j}^T\hat{\theta}_{c,ij} + \frac{\sigma_{c,ij}}{2}\theta_{c,ij}^T S_{i,j}S_{i,j}^T\theta_{c,ij},
\end{aligned}
\tag{70}
$$

$$
\begin{aligned}
&(\sigma_{c,ij} - \sigma_{a,ij})\tilde{\theta}_{a,ij}^T S_{i,j}S_{i,j}^T\hat{\theta}_{c,ij} \\
&\le \frac{(\sigma_{c,ij} - \sigma_{a,ij})}{2}\tilde{\theta}_{a,ij}^T S_{i,j}S_{i,j}^T\tilde{\theta}_{a,ij} + \frac{(\sigma_{c,ij} - \sigma_{a,ij})}{2}\hat{\theta}_{c,ij}^T S_{i,j}S_{i,j}^T\hat{\theta}_{c,ij},
\end{aligned}
\tag{71}
$$

$$
\sigma_{i,j}\tilde{\vartheta}_{i,j}\hat{\vartheta}_{i,j} \le \frac{\sigma_{i,j}}{2}\vartheta_{i,j}^2 - \frac{\sigma_{i,j}}{2}\tilde{\vartheta}_{i,j}^2.
\tag{72}
$$

From equations (58), (62), and (64)-(72), $\dot{V}_{i,j}$ can be directly bounded as

$$
\dot{V}_{i,j} \leq -(c_{i,1} - \frac{7}{4})\zeta_i^2 - (c_{i,2} - \frac{7}{4} - \frac{(d_i + b_i)^2 \eta_i^2}{2})z_{i,2}^2 - \sum_{m=3}^{j}(c_{i,m} - \frac{9}{4})z_{i,m}^2 + \frac{1}{2}z_{i,j+1}^2 + \frac{1}{2}\varkappa_i \rho_{i,j+1}^2
$$

$$
- (\frac{1}{\kappa_i} - \frac{1}{2} - \frac{(d_i + b_i)^2 \eta_i^2}{2})\rho_{i,2}^2 - \sum_{m=3}^{j}(\frac{1}{\kappa_i} - 1)\rho_{i,m}^2 - \sum_{m=1}^{j}\frac{\sigma_{i,m}}{2}\tilde{\vartheta}_{i,m}^2 - \sum_{m=1}^{j}\lambda_{i,m}^{\min}(\sigma_{a,im} - \frac{\sigma_{c,im}}{2})\tilde{\theta}_{a,im}^T \tilde{\theta}_{a,im}
$$

$$
- \sum_{m=1}^{j}\lambda_{i,m}^{\min}\frac{\sigma_{c,im}}{2}\tilde{\theta}_{c,im}^T \tilde{\theta}_{c,im} + \sum_{m=1}^{j}\Upsilon_{i,m}, \tag{73}
$$

where $\Upsilon_{i,j} = \frac{1}{2}\bar{\epsilon}_{i,j}^2 + \frac{\sigma_{i,j}}{2}\vartheta_{i,j}^2 + \frac{\sigma_{a,ij}}{2}\theta_{a,ij}^T S_{i,j} S_{i,j}^T \theta_{a,ij} + \frac{\sigma_{c,ij}}{2}\theta_{c,ij}^T S_{i,j} S_{i,j}^T \theta_{c,ij} - (\frac{\sigma_{a,ij}}{2} - \frac{1}{4})\hat{\theta}_{a,ij}^T S_{i,j} S_{i,j}^T \hat{\theta}_{a,ij} - \frac{\sigma_{c,ij}}{2}\hat{\theta}_{c,ij}^T S_{i,j} S_{i,j}^T \hat{\theta}_{c,ij} +$

$\frac{1}{2}M_i^2$ and $\varkappa_i = \begin{cases} \frac{(d_i+b_i)^2\eta_i^2}{2}, & j = 1, \\ 1, & j = 2, \cdots, n-1. \end{cases}$

Step n: The performance index function for the last subsystem is defined as

$$
V_{i,n}(z_{i,n}) = \int_t^\infty c_{i,n}(z_{i,n}(\tau), u_i^*(z))d\tau, \tag{74}
$$

where $c_{i,n}(z_n, \alpha_{i,n}) = z_{i,n}^2 + u_i^{*2}$ represents the value function.

The optimal performance index function is

$$
V_{i,j}^*(z_{i,n}) = \int_t^\infty c_{i,n}(z_{i,n}(\tau), u_i^*(z))d\tau
$$

$$
= \min_{u_i^* \in \psi(\Omega)} \left(\int_t^\infty c_{i,n}(z_{i,n}(\tau), u_i(z))d\tau \right). \tag{75}
$$

The HJB equation is introduced as

$$
H_{i,n}(z_{i,n}, u_i^*, V_{i,n}^*) = c_{i,n}(z_{i,n}, u_i^*) + \frac{\partial V_{i,n}^*}{\partial z_{i,n}}\dot{z}_{i,n}
$$

$$
= z_{i,n}^2 + u_i^{*2} + \frac{\partial V_{i,n}^*}{\partial z_{i,n}}(u_i^* + f_{i,n} - \dot{\alpha}_{i,n}) = 0. \tag{76}
$$

By calculating $\frac{\partial H_{i,n}}{\partial u_i^*} = 0$, we get

$$
u_i^* = -\frac{1}{2}\frac{\partial V_{i,n}^*}{\partial z_{i,n}}. \tag{77}
$$

$\frac{\partial V_{i,n}^*}{\partial z_{i,n}}$ can be decomposed into

$$
\frac{\partial V_{i,n}^*}{\partial z_{i,n}} = 2c_{i,n}z_{i,n} + V_{i,n}^o + 2\vartheta_{i,n}^{*T}\phi_{i,n}, \tag{78}
$$

$$
V_{i,n}^o = \frac{\partial V_{i,n}^*}{\partial z_{i,n}} - 2c_{i,n}z_{i,n} - 2\vartheta_{i,n}^{*T}\phi_{i,n}. \tag{79}
$$

where $c_{i,n}$ is a design parameter. Since $V_{i,n}^o$ is an unknown term, by applying Lemma 2, one gets

$$
V_{i,n}^o = \theta_{i,n}^{*T}S_{i,n} + \varepsilon_{i,n}. \tag{80}
$$

Then, the facts below are easily available

$$u_i^* = -c_{i,n} z_{i,n} - \frac{1}{2}(\theta_{i,n}^{*T} S_{i,n} + \varepsilon_{i,n}) - \vartheta_{i,n}^{*T} \phi_{i,n}, \tag{81}$$

$$\frac{\partial V_{i,n}^*}{\partial z_{i,n}} = 2c_{i,n} z_{i,n} + (\theta_{i,n}^{*T} S_{i,n} + \varepsilon_{i,n}) + 2\vartheta_{i,n}^{*T} \phi_{i,n} \tag{82}$$

The $\hat{\theta}_{a,in}$ and $\hat{\theta}_{c,in}$ of the actor FLS and the critic FLS are used to estimate the unknown weights $\theta_{i,n}^*$, respectively, to obtain

$$\hat{u}_i = -c_{i,n} z_{i,n} - \frac{1}{2}\hat{\theta}_{a,in}^T S_{i,n} - \hat{\vartheta}_{i,n}^T \phi_{i,n}, \tag{83}$$

$$\frac{\partial \hat{V}_{i,n}}{\partial z_{i,n}} = 2c_{i,n} z_{i,n} + \hat{\theta}_{c,in}^T S_{i,n} + 2\hat{\vartheta}_{i,n} \phi_{i,n}. \tag{84}$$

Candidate Lyapunov function is chosen to be the

$$V_{i,n} = V_{i,n-1} + \frac{1}{2} z_{i,n}^2 + \frac{1}{2}\tilde{\theta}_{a,in}^T \tilde{\theta}_{a,in} + \frac{1}{2}\tilde{\theta}_{c,in}^T \tilde{\theta}_{c,in} + \frac{1}{2}\tilde{\vartheta}_{i,n}^T \tilde{\vartheta}_{i,n} + \frac{1}{2}\rho_{i,n}\rho_{i,n}. \tag{85}$$

The time derivative of $V_{i,n}$ is given by

$$\begin{aligned}
\dot{V}_{i,n} = {}& \dot{V}_{i,n-1} + z_{i,n}(\hat{u}_i + f_{i,n} - \dot{\alpha}_{i,n}) + \rho_{i,n}\dot{\rho}_{i,n} \\
& + \tilde{\theta}_{a,in} S_{i,n}^T S_{i,n}^T (\sigma_{a,in}(\hat{\theta}_{a,in} - \hat{\theta}_{c,in}) + \sigma_{c,in}\hat{\theta}_{c,in}) + \tilde{\theta}_{c,in}^T \sigma_{c,in} S_{i,n} S_{i,n}^T \hat{\theta}_{c,in} - \tilde{\vartheta}_{i,n}^T \dot{\hat{\vartheta}}_{i,n}.
\end{aligned} \tag{86}$$

By Lemma 2, there exists an FLS satisfying

$$F_{i,n} = \vartheta_{i,n}^{*T} \phi_{i,n} + \epsilon_{i,n}, \tag{87}$$

where $F_{i,n} = f_{i,n} - \dot{\alpha}_{i,n}$, and $\epsilon_{i,n} \leq \bar{\epsilon}_{i,n}$ with $\bar{\epsilon}_{i,n} > 0$.

Similar to the previous steps, $\dot{V}_{i,n}$ can be bounded as

$$\begin{aligned}
\dot{V}_{i,n} \leq {}& -(c_{i,1} - \frac{7}{4})\zeta_i^2 - (c_{i,2} - \frac{7}{4} - \frac{(d_i + b_i)^2 \eta_i^2}{2})z_{i,2}^2 - \sum_{m=3}^{n}(c_{i,m} - \frac{9}{4})z_{i,m}^2 - \sum_{m=3}^{n}(\frac{1}{\kappa_i} - 1)\rho_{i,m}^2 \\
& - (\frac{1}{\kappa_i} - \frac{1}{2} - \frac{(d_i + b_i)^2 \eta_i^2}{2})\rho_{i,2}^2 - \sum_{m=1}^{n}\frac{\sigma_{i,m}}{2}\tilde{\vartheta}_{i,m}^2 - \sum_{m=1}^{n}(\sigma_{a,im} - \frac{\sigma_{c,im}}{2})\lambda_{i,m}^{\min}\tilde{\theta}_{a,im}^T S_{i,m} S_{i,m}^T \tilde{\theta}_{a,im} \\
& - \sum_{m=1}^{n}\lambda_{i,m}^{\min}\frac{\sigma_{c,im}}{2}\tilde{\theta}_{c,im}^T \tilde{\theta}_{c,im} + \sum_{m=1}^{n}\Upsilon_{i,m},
\end{aligned} \tag{88}$$

where $\Upsilon_{i,n} = \frac{1}{2}\bar{\epsilon}_{i,n}^2 + \frac{\sigma_{i,n}}{2}\vartheta_{i,n}^2 + \frac{\sigma_{a,in}}{2}\theta_{a,in}^T S_{i,n} S_{i,n}^T \theta_{a,in} + \frac{\sigma_{c,in}}{2}\theta_{c,in}^T S_{i,n} S_{i,n}^T \theta_{c,in} - (\frac{\sigma_{a,in}}{2} - \frac{1}{4})\hat{\theta}_{a,in}^T S_{i,n} S_{i,n}^T \hat{\theta}_{a,in} - \frac{\sigma_{c,in}}{2}\hat{\theta}_{c,in}^T S_{i,n} S_{i,n}^T \hat{\theta}_{c,in} + \frac{1}{2}M_i^2$.

Define $E_i = \min\{c_{i,1} - \frac{7}{4}, c_{i,2} - \frac{7}{4} - \frac{(d_i + b_i)^2 \eta_i^2}{2}, c_{i,m} - \frac{9}{4}, \frac{\sigma_{i,m}}{2}, (\sigma_{a,im} - \frac{\sigma_{c,im}}{2})\lambda_{i,m}^{\min}, \lambda_{i,m}^{\min}\frac{\sigma_{c,im}}{2}\tilde{\theta}_{c,im}^T\}, \Upsilon_i = \sum_{m=1}^{n}\Upsilon_{i,m}$, $\dot{V}_{i,n}$ can then be written as

$$\dot{V}_{i,n} \leq -E_i V_{i,n} + \Upsilon_i, \tag{89}$$

Theorem 1: Consider a nonlinear MAS under Assumptions 1-2, with the optimal virtual controller choice of equations (30), (58), and (83), the adaptive laws of equations (14) and (15), the actor FLS choice of equation (12), and the critic FLS choice of (13). Then, we select the parameter $\sigma_{a,im} > \frac{\sigma_{c,im}}{2} > \frac{\sigma_{a,im}}{2} > 0, c_{i,1} > \frac{7}{4}, c_{i,2} > \frac{7}{4} + \frac{(d_i+b_i)^2\eta_i^2}{2}, c_{i,m} > \frac{9}{4}, \frac{1}{\kappa_i} > \frac{1}{2} + \frac{(d_i+b_i)^2\eta_i^2}{2}, \frac{1}{\kappa_i} > 1$. Thus, we can conclude the following results:

(1) All signals in the closed-loop system are bounded.

(2) Consensus tracking error is within predefined bounds.

Proof: The total Liapunov function for all agents is selected to be

$$V = \sum_{i=1}^{N} V_{i,n}. \tag{90}$$

The derivative of V with respect to time is

$$\dot{V} \le -EV + \Upsilon, \tag{91}$$

where $E = \min E_i, i = 1, 2, \cdots, N, \Upsilon = \Sigma_{i=1}^{N} \Upsilon_i$.

Obviously, it can be inferred that

$$0 \le V(t) \le e^{-Et}(V(0) - \frac{\Upsilon}{E}) + \frac{\Upsilon}{E}. \tag{92}$$

From (91), we know that ζ_i is bounded by

$$\zeta_i \le \sqrt{\frac{2\Upsilon}{E}} \tag{93}$$

This means that the consensus tracking error can be bounded within prescribed bounds, i.e., $\Gamma(-\varphi(t)) \le e_i \le \Gamma(\varphi(t))$. Then, we can easily confirm $\|y - \bar{1}y_r\| \le \frac{\|\Gamma(\varphi(t))\|}{\underline{\sigma}(\mathscr{L}+\mathscr{B})}$, where $\underline{\sigma}(\mathscr{L}+\mathscr{B})$ represents the smallest singular value of the matrix $\mathscr{L} + \mathscr{B}$.

On the other hand, it is known from equation (92) that $\tilde{\theta}_{a,ij}, \tilde{\theta}_{c,ij}^T, \tilde{\vartheta}_{i,j}, \rho_{i,j}$, and $z_{i,j}$ are bounded. Since $\hat{\alpha}_{i,j}$ is a function consisting of bounded signals, $\hat{\alpha}_{i,j}$ is bounded. Based on the definition of $\tilde{\theta}_{a,ij} = \theta_{a,ij} - \hat{\theta}_{a,ij}, \tilde{\theta}_{c,ij} = \theta_{c,ij} - \hat{\theta}_{c,ij}, \tilde{\vartheta}_{i,j} = \vartheta_{i,j} - \hat{\vartheta}_{i,j}$, it is evident that $\hat{\theta}_{a,ij}, \hat{\theta}_{c,ij}$, and $\hat{\vartheta}_{i,j}$ are bounded. Thus, it can be concluded that all signals are bounded.

4. SIMULATION EXAMPLES

Consider the nonlinear MASs with four following agents and a leader, whose dynamics model is represented as

$$\begin{cases} \dot{x}_{i,1} = x_{i+1,2}(t), \\ \dot{x}_{i,2} = u_i(t) + f_{i,2}(\bar{x}_{i,2}(t)), \end{cases} \tag{94}$$

where $i = 1, 2, 3, 4, f_{i,2} = 0.01\sin(0.5(x_{i,1} - x_{i,2}))$. The reference output trajectory is set to $y_r = \sin(0.5t)$. The communication topology is displayed in Figure 1, through which the Laplace matrix $\mathscr{L}$ is easily obtained as

$$\mathscr{L} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ -1 & 2 & 0 & -1 \\ -1 & 0 & 1 & 0 \\ 0 & -1 & 0 & 1 \end{bmatrix}.$$
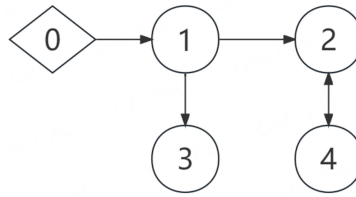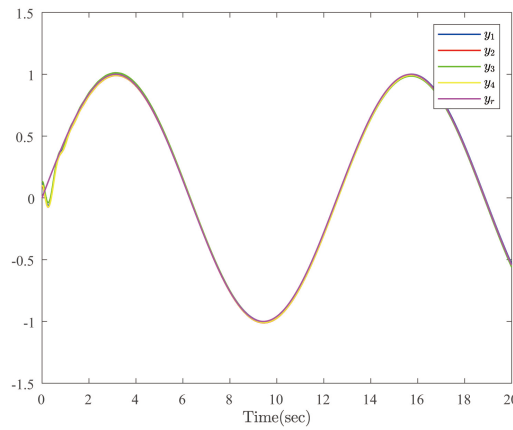
Figure 1. The communication topology.

Figure 2. The trajectories of the system output state $y_i\,(i = 1, 2, 3, 4)$ and the reference signal y_r.

Select the fuzzy membership function as

$$\mu_{F_{i,j}^k} = e^{-\frac{(x_{i,j}+l)^2}{2}}, i = 1, 2, 3, 4; j = 1, 2,$$

$$k = 9, 7, 5, 3, 1, 0, -1, -3, -5, -7, 9.$$

The time-varying function $\omega_i(t)$ is chosen as $\omega_i(t) = \dfrac{1}{(1 - l_i)\exp(-1.5t) + l_i}$. The initial state values are selected as $x_{i,1}(0) = [0.11, 0.09, 0.13, 0.1]$, $x_{i,2}(0) = [0.11, 0.2, 0.18, 0.11]$, $\hat{\theta}_{c,i1}(0) = [0.2, 0.2, 0.22, 0.12]$, $\hat{\theta}_{c,i2}(0) = [0.2, 0.2, 0.12, 0.12]$, $\hat{\theta}_{a,i1}(0) = [1.4, 1, 1, 1]$, $\hat{\theta}_{a,i2}(0) = [1.2, 1.2, 1, 1]$, $\hat{\vartheta}_{i,j}(0) = 1$, $\bar{\alpha}_{i,j}(0) = 0$. The design parameters are selected as $l_i = 0.2$, $\chi = 5$, $c_{i,j} = 9$, $\kappa_{i,2} = 0.02$, $\sigma_{c,ij} = 7$, $\sigma_{a,ij} = 5$, $\sigma_{11} = \sigma_{12} = \sigma_{21} = \sigma_{22} = \sigma_{31} = 21$, and $\sigma_{32} = \sigma_{41} = \sigma_{42} = 11$. The resulting simulations are presented in Figures 2-6. Figure 2 depicts the output trajectories of follower agents and the reference trajectory, demonstrating the guaranteed well-tracking performance under the designed control protocol. Figure 3 shows the consensus tracking errors of agents and performance constraint bounds, from which it can be seen that the constraint has never been violated. Figure 4 provides the designed control protocol. Figure 5 depicts the trajectories of adaptive parameter $\theta_{c,ij}$. Figure 6 portrays the trajectories of adaptive parameters $\theta_{a,ij}$. The trajectories of adaptive parameter $\vartheta_{i,j}$ is plotted in Figure 7. The above leads show that adaptive parameters $\theta_{c,ij}$, $\theta_{a,ij}$, and $\vartheta_{i,j}$ are bounded. Based on the aforementioned results, it is evident that our control objectives have been successfully achieved.
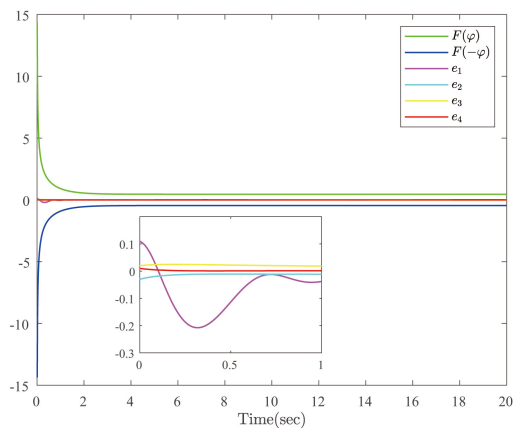
Figure 3. System consensus tracking error $e_i (i = 1, 2, 3, 4)$ and constraint functions $\Gamma(\varphi(t)), \Gamma(-\varphi(t))$.
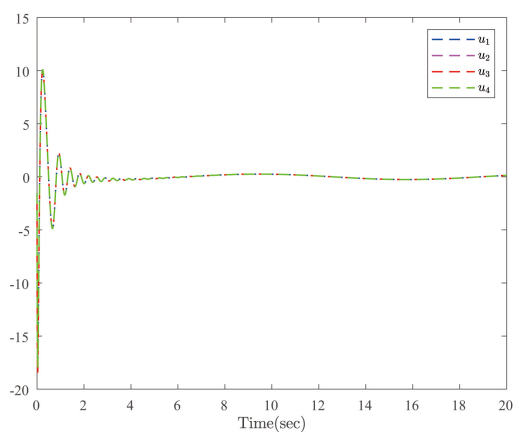

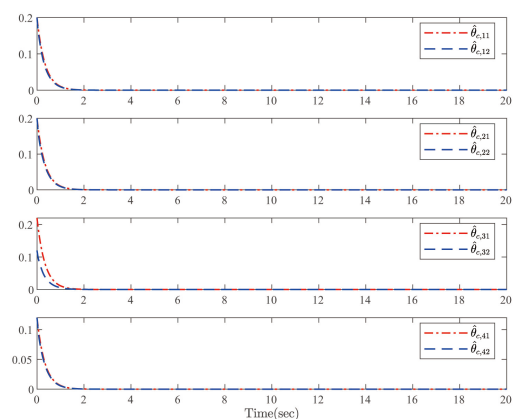**Figure 4.** The trajectories of the control protocols $u_i (i = 1, 2, 3, 4)$.

Figure 5. Responses of $\theta_{c,ij} (i = 1, 2, 3, 4; k = 1, 2)$.

5. CONCLUSIONS

In this paper, the problem of optimal adaptive consensus tracking control for nonlinear MASs with prescribed performance has been addressed. Firstly, a time-varying scalar function is introduced such that the designed
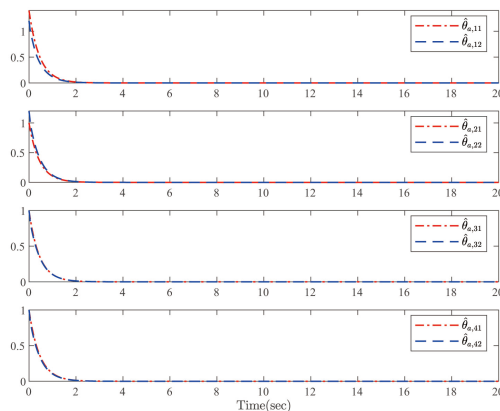
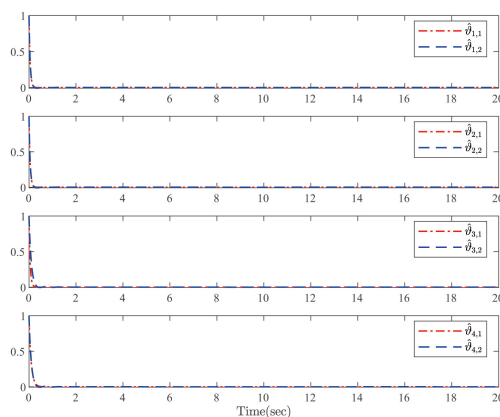**Figure 6.** Responses of $\theta_{a,ij}\,(i=1,2,3,4;k=1,2)$.

Figure 7. Responses of $\vartheta_{i,j}\,(i=1,2,3,4;k=1,2)$.

performance function bypasses the initial value conditions. Based on the error transformation function, an unconstrained system is obtained. Subsequently, a RL-based consensus control scheme based on optimal control theory and dynamic surface technique has been proposed. Finally, it is shown that the stability of the closed-loop system and the error constraints are not violated. In practice, the systems are always subject to various uncertain constraints, such as actuator faults and input dead zones, which will have a large impact on the performance of systems. Therefore, designing a properly performance-constrained optimal control scheme considering the above situations is a topic for further research in the future.

DECLARATIONS

Authors' contributions
Made substantial contributions to the conception and design of the study and performed data analysis and interpretation: Yue H
Performed data acquisition and provided administrative, technical, and material support: Xia J

Availability of data and materials
Not applicable.

Conflicts of interest
All authors declared that there are no conflicts of interest.

Ethical approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

REFERENCES

1. Modares H, Lewis FL, Naghibi-Sistani MB. Integral reinforcement learning and experience replay for adaptive optimal control of partially-unknown constrained-input continuous-time systems. *Automatica* 2014;50:193-202. DOI
2. Bertsekas DP. Value and policy iterations in optimal control and adaptive dynamic programming. *IEEE Trans Neural Netw Learn Syst* 2015;28:500-9. DOI
3. Tsai JSH, Li JS, Leang-San S. Discretized quadratic optimal control for continuous-time two-dimensional systems. *IEEE Trans Circuits Syst I Fund Theory Appl* 2002;49:116-25. DOI
4. Luo B, Liu DR, Wu HN, Wang D, Lewis FL. Policy gradient adaptive dynamic programming for data-based optimal control. *IEEE Trans Cybern* 2016;47:3341-54. DOI
5. Jiang Y, Jiang ZP. Global adaptive dynamic programming for continuous-time nonlinear systems. *IEEE Trans Automat Contr* 2015;60:2917-29. DOI
6. Wu X, Chen HL, Wang JJ, Troiano L, Loia V, Fujita H. Adaptive stock trading strategies with deep reinforcement learning methods. *Inf Sci* 2020;538:142-58. DOI
7. Modares H, Ranatunga I, Lewis FL, Popa DO. Optimized assistive human-robot interaction using reinforcement learning. *IEEE Trans Cybern* 2015;46:655-67. DOI
8. Wen GX, Chen CLP, Li WN. Simplified optimized control using reinforcement learning algorithm for a class of stochastic nonlinear systems. *Inf Sci* 2020;517:230-43. DOI
9. Zhao B, Liu DR, Luo CM. Reinforcement learning-based optimal stabilization for unknown nonlinear systems subject to inputs with uncertain constraints. *IEEE Trans Neural Netw Learn Syst* 2019;31:4330-40. DOI
10. Wen G, Chen CLP, Ge SS, Yang H, Liu X. Optimized adaptive nonlinear tracking control using actor-critic reinforcement learning strategy. *IEEE Trans Ind Inf* 2019;15:4969-77. DOI
11. Bai W, Zhou Q, Li T, Li H. Adaptive reinforcement learning neural network control for uncertain nonlinear system with input saturation. *IEEE Trans Cybern* 2019;50:3433-43. DOI
12. Yang X, Liu D, Wang D. Reinforcement learning for adaptive optimal control of unknown continuous-time nonlinear systems with input constraints. *Int J Control* 2014;87:553-66. DOI
13. Bai W, Li T, Tong S. NN reinforcement learning adaptive control for a class of nonstrict-feedback discrete-time systems. *IEEE Trans Cybern* 2020;50:4573-84. DOI
14. Li Y, Zhang J, Liu W, Tong S. Observer-based adaptive optimized control for stochastic nonlinear systems with input and state constraints. *IEEE Trans Neural Netw Learn Syst* 2021;33:7791-805. DOI
15. Wang J, Gong Q, Huang K, Liu Z, Chen CLP, Liu J. Event-triggered prescribed settling time consensus compensation control for a class of uncertain nonlinear systems with actuator failures. *IEEE Trans Neural Netw Learn Syst* 2023;34:5590-600. DOI
16. Wang J, Wang C, Liu Z, Chen CLP, Zhang C. Practical fixed-time adaptive ERBFNNs event-triggered control for uncertain nonlinear systems with dead-zone constraint. *IEEE Trans Syst Man Cybern Syst* 2023:1-10. DOI
17. Cai X, de Marcio M. Adaptive rigidity-based formation control for multirobotic vehicles with dynamics. *IEEE Trans Contr Syst Technol* 2014;23:389-96. DOI
18. Ren H, Cheng Z, Qin J, Lu R. Deception attacks on event-triggered distributed consensus estimation for nonlinear systems. *Automatica* 2023;154:111100. DOI
19. Wang J, Yan Y, Liu Z, Chen CLP, Zhang C, Chen K. Finite-time consensus control for multi-agent systems with full-state constraints and actuator failures. *Neural Netw* 2023;157:350-63. DOI

20. Shang Y. Matrix-scaled consensus on weighted networks with state constraints. *IEEE Syst J* 2023;17:6472-9. DOI

21. Cheng L, Hou ZG, Tan M, Lin Y, Zhang W. Neural-network-based adaptive leader-following control for multiagent systems with uncertainties. *IEEE Trans Neural Netw* 2010;21:1351-8. DOI

22. Shen Q, Shi P, Zhu J, Wang S, Shi Y. Neural networks-based distributed adaptive control of nonlinear multiagent systems. *IEEE Trans Neural Netw Learn Syst* 2019;31:1010-21. DOI

23. Zhang N, Xia J, Park JH, Zhang J, Shen H. Improved disturbance observer-based fixed-time adaptive neural network consensus tracking for nonlinear multi-agent systems. *Neural Netw* 2023;162:490-501. DOI

24. Zhang Y, Sun J, Liang H, Li H. Event-triggered adaptive tracking control for multiagent systems with unknown disturbances. *IEEE Trans Cybern* 2018;50:890-901. DOI

25. Chen J, Li J, Yuan X. Global fuzzy adaptive consensus control of unknown nonlinear multiagent systems. *IEEE Trans Fuzzy Syst* 2020;32:2239-50. DOI

26. Zhang J, Liu S, Zhang X, Xia J. Event-triggered-based distributed consensus tracking for nonlinear multiagent systems with quantization. *IEEE Trans Neural Netw Learn Syst* 2022:1-11. DOI

27. Deng C, Wen C, Wang W, Li X, Yue D. Distributed adaptive tracking control for high-order nonlinear multiagent systems over event-triggered communication. *IEEE Trans Automat Contr* 2022;68:1176-83. DOI

28. Shao J, Shi L, Cheng Y, Li T. Asynchronous tracking control of leader–follower multiagent systems with input uncertainties over switching signed digraphs. *IEEE Trans Cybern* 2021;52:6379-90. DOI

29. Yang Y, Xiao Y, Li T. Attacks on formation control for multiagent systems. *IEEE Trans Cybern* 2021;52:12805-17. DOI

30. Ren H, Wang Y, Liu M, Li H. An optimal estimation framework of multi-agent systems with random transport protocol. *IEEE Trans Signal Process* 2022;70:2548-59. DOI

31. Gao W, Jiang ZP, Lewis FL, Wang Y. Leader-to-formation stability of multiagent systems: An adaptive optimal control approach. *IEEE Trans Automat Contr* 2018;63:3581-87. DOI

32. Tan M, Liu Z, Chen CLP, Zhang Y, Wu Z. Optimized adaptive consensus tracking control for uncertain nonlinear multiagent systems using a new event-triggered communication mechanism. *Inf Sci* 2022;605:301-16. DOI

33. Wen G, Chen CLP. Optimized backstepping consensus control using reinforcement learning for a class of nonlinear strict-feedback-dynamic multi-agent systems. *IEEE Trans Neural Netw Learn Syst* 2023;34:1524-36. DOI

34. Zhu HY, Li YX, Tong S. Dynamic event-triggered reinforcement learning control of stochastic nonlinear systems. *IEEE Trans Fuzzy Syst* 2023;31:2917-28. DOI

35. Bai W, Li T, Long Y, Chen CLP. Event-triggered multigradient recursive reinforcement learning tracking control for multiagent systems. *IEEE Trans Neural Netw Learn Syst* 2023;34:366-79. DOI

36. Li T, Bai W, Liu Q, Long Y, Chen CLP. Distributed fault-tolerant containment control protocols for the discrete-time multi-agent systems via reinforcement learning method. *IEEE Trans Neural Netw Learn Syst* 2023;34:3979-91. DOI

37. Zhao Y, Niu B, Zong G, Zhao X, Alharbi KH. Neural network-based adaptive optimal containment control for non-affine nonlinear multi-agent systems within an identifier-actor-critic framework. *J Franklin Inst* 2023;360:8118-43. DOI

38. Li H, Wu Y, Chen M, Lu R. Adaptive multigradient recursive reinforcement learning event-triggered tracking control for multiagent systems. *IEEE Trans Neural Netw Learn Syst* 2023;34:144-56. DOI

39. Bechlioulis CP, Rovithakis GA. Robust adaptive control of feedback linearizable MIMO nonlinear systems with prescribed performance. *IEEE Trans Automat Contr* 2008;53:2090-9. DOI

40. Wang X, Xia J, Park JH, Xie X, Chen G. Intelligent control of performance constrained switched nonlinear systems with random noises and its application: an event-driven approach. *IEEE Trans Circuits Syst I Regul Pap* 2022;69:3736-47. DOI

41. Li Y, Shao X, Tong S. Adaptive fuzzy prescribed performance control of nontriangular structure nonlinear systems. *IEEE Trans Fuzzy Syst* 2019;28:2416-26. DOI

42. Wang W, Liang H, Pan Y, Li T. Prescribed performance adaptive fuzzy containment control for nonlinear multiagent systems using disturbance observer. *IEEE Trans Cybern* 2020;50:3879-91. DOI

43. Sun K, Qiu J, Karimi HR, Fu Y. Event-triggered robust fuzzy adaptive finite-time control of nonlinear systems with prescribed performance. *IEEE Trans Fuzzy Syst* 2020;29:1460-71. DOI

44. Chen H, Yan H, Wang Y, Xie S, Zhang D. Reinforcement learning-based close formation control for underactuated surface vehicle with prescribed performance and time-varying state constraints. *Ocean Eng* 2022;256:111361. DOI

45. Wang N, Gao Y, Zhang X. Data-driven performance-prescribed reinforcement learning control of an unmanned surface vehicle. *IEEE Trans Neural Netw Learn Syst* 2021;32:5456-67. DOI