ICIC Express Letters
Volume **16**, Number **2**, February **2022**
 ISSN 1881-803X
pp. 177–185

OPTIMAL CONSENSUS CONTROL FOR HETEROGENEOUS NONLINEAR NON-AFFINE MULTI-AGENT SYSTEMS WITH UNCERTAIN CONTROL DIRECTIONS

Miao Huang[1,*], Zhihua Hu[1] and Lang Wang[2]

[1]College of Intelligent Manufacturing and Control Engineering
Shanghai Polytechnic University
No. 2360, Jinhai Road, Pudong District, Shanghai 201209, P. R. China
zhhu@sspu.edu.cn
[*]Corresponding author: huangmiao@sspu.edu.cn

[2]School of Information Science and Engineering
NingboTech University
No. 1, South Qianhu Road, Ningbo 315100, P. R. China
wanglang@nbt.edu.cn

Received June 2021; accepted August 2021

Abstract. *This paper considers an optimal consensus problem for a class of heterogeneous discrete-time nonlinear multi-agent systems (MASs) with unknown dynamics and uncertain control directions. This problem is transformed to find the solution of the coupled Hamilton-Jacobi-Bellman (HJB) equations. A novel reinforcement learning based distributed control algorithm implemented by an actor-critic form has been proposed to obtain the optimal control policies. Fuzzy rules emulated networks (FRENs) have been involved to estimate the uncertain control directions and the unknown system dynamics. Simulation results validate the boundedness and effectiveness of the closed-loop systems applying the proposed control strategy.*
Keywords: Heterogeneous multi-agent systems, Optimal consensus, Uncertain control directions, Fuzzy rules emulated networks

1. **Introduction.** One of the most important and challenging problems in the consensus control of multi-agent systems (MASs) is the optimal consensus problem. In this problem, the coupled Hamilton-Jacobi-Bellman (HJB) equation is usually established to communicate the optimal control law and the minimized performance index. Then, the consensus problem has been transformed to the solution of the coupled HJB equation.

The reinforcement learning (RL) strategy has been commonly used in optimal control field, in which the optimal solution of the HJB equation can be found through frequent interaction with the environment. In [1], the HJB equation of the augmented multiple inputs system is solved for the continuous affine nonlinear system with multiple inputs. An optimal safety controller is obtained by combining RL and model predictive control in [2]. The robust formation control problem of the cooperative under-actuated quad rotor with unknown nonlinear dynamics and disturbances has been considered in [3]. To solve the constrained optimal control problem, the neural network based RL algorithm implemented by an actor-critic form has been established in [4]. In these methods, the idea nonlinear controller with unknown parameters and the strategic utility function are approximated by the action network and the critic network with adaptable weights, respectively.

Fuzzy-rules emulated network (FREN) designed by [5] is a kind of fuzzy neural networks whose IF-THEN rules are created by the human knowledge. Compared with artificial neural networks whose structure and adjustable parameters are set randomly, those of

DOI: 10.24507/icicel.16.02.177

FRENs are designed in the sense of engineering, which can make full use of the prior knowledge to improve the network performance. For nonlinear discrete-time systems, an adaptive controller based on FRENs and the sliding mode mechanism is proposed by [6]. A model free controller based on RL with IF-THEN rules is proposed in [7] for a class of nonlinear discrete-time systems with output feedback. In order to solve the power capture problem of variable speed wind turbine systems with flexible shaft, an adaptive power signal feedback control based on FRENs and command filter is proposed in [8].

The main contributions of this paper are as follows.

1) A new distributed adaptive control strategy based on RL and FRENs has been developed. Unlike the previous works in [9-12], in this strategy, the performance index and the optimal control policy are approximated by the critic FREN and the actor FREN, respectively.

2) Different from the controller designed for MASs with certain control directions in [13-17], the distributed control problem of heterogeneous nonlinear non-affine discrete-time MASs with unknown system dynamics and uncertain control directions has been addressed.

The rest of this paper is organized as follows. In Section 2, the problem and some preliminaries have been formulated. Section 3 gives a new distributed optimal control strategy based on RL and FRENs. Simulation results are provided in Section 4 to verify the effectiveness of the proposed strategy. Finally, the conclusion goes in Section 5.

2. **Problem Statement and Preliminaries.** Consider a class of heterogeneous nonlinear MASs with N followers, in which each dynamic of the follower agent i has a non-affine pure-feedback form as follows:

$$\begin{aligned} &x_{i1}(k+1) = f_{i1}\left(\bar{x}_{i1}(k), x_{i2}(k)\right), \ldots, x_{in_i}(k+1) = f_{in_i}\left(\bar{x}_{in}(k), u_i(k)\right), \\ &y_i(k) = x_{i1}(k) \end{aligned} \tag{1}$$

where $x_i = [x_{i1}, x_{i2}, \ldots, x_{in_i}]^T \in R^{n_i}$, $i = 1, 2, \ldots, N$ represents the system state vector of agent i, $\bar{x}_{ij}(k) = [x_{i1}(k), x_{i2}(k), \ldots, x_{ij}(k)]^T$, $j = 1, 2, \ldots, n_i$, $n_i \geq 1$ is the order of the ith agent, x_{ij} is the jth system state of the ith agent, and $f_{ij}(\cdot, \cdot)$ is an unknown nonlinear function. $y_i(k)$ and $u_i(k) \in R$, $i = 1, 2, \ldots, N$ are the output and input of the ith agent, respectively.

Assumption 2.1. *Each $f_{ij}(\cdot, \cdot)$, $i = 1, 2, \ldots, N$, $j = 1, 2, \ldots, n_i$ in (1) is continuous with respect to all the arguments and continuously differentiable with respect to the second argument.*

Assumption 2.2. *There exists a constant $g_M = \sup_{i \leq N, j \leq n_i} |g_{ij}(k)|$, $i = 1, 2, \ldots, N$, $j = 1, 2, \ldots, n_i$ so that $0 < |g_{ij}(\cdot)| \leq g_M$, where $g_{ij}(\cdot) = \partial f_j\left(\bar{x}_{ij}(k), x_{i,j+1}(k)\right) / \partial x_{i,j+1}(k)$, $j = 1, 2, \ldots, n_i - 1$ and $g_{in_i}(\cdot) = \partial f_n\left(\bar{x}_{in}(k), u_i(k)\right) / \partial u_i(k)$ are the unknown control gains. The sign of function g_{ij} represents the control direction which is unknown and unfixed.*

Meanwhile, consider a leader node denoted as v_0, which is defined as $y_0(k + n_0) = \varphi_0(k)$.

The control goal is to design a distributed control law u_i for each follower agent such that the output of the follower synchronizes to a small neighborhood $U(y_0(k), \delta)$ of that of the leader, i.e., $\lim_{k \to \infty} \|y_i(k) - y_0(k)\| < \delta$, for $\forall i$ and $\delta > 0$.

Define the local neighborhood tracking error as $e_i(k) = \sum_{j \in N_i} (a_{ij}(y_i(k) - y_j(k))) + b_i(y_i(k) - y_0(k))$, where $b_i \geq 0$ denotes the pinning gain. The overall tracking error can be considered as $e(k) = (L + B)\left(y(k) - \underline{y}_0(k)\right)$, where $L = [l_{ij}] \in R^{N \times N}$ is the Laplacian matrix for the directed graph; $B = [b_{ij}] \in R^{N \times N}$ is a diagonal matrix with the diagonal elements $b_{ij} = b_i$; $e(k) = [e_1(k), e_2(k), \ldots, e_N(k)]^T \in R^N$; $y(k) = [y_1(k), y_2(k), \ldots, y_N(k)]^T \in R^N$; $\underline{y}_0(k) = [y_0(k), y_0(k), \ldots, y_0(k)]^T \in R^N$.

In order to simplify the controllers design, system (1) can be transformed into an input-output form without the future states according to the derivative process given in [18]

$$y_i(k+n_i) = \varphi_i\left(\underline{z}_i(k), u_i(k)\right) \tag{2}$$

where $\underline{z}_i(k) = [y_i(k), \ldots, y_i(k-n+1), u_i(k-1), \ldots, u_i(k-n+1)]$, $\varphi_i(\cdot,\cdot)$: $R^{2n} \to R$ are unknown nonlinear functions.

The local performance index function is defined as

$$J_i(e_i(k), u_i(k-n_i)) = \sum_{t=k}^{\infty} \gamma^{t-k} r_i(k) = r_i(k) + \gamma J_i(e_i(k+1), u_i(k+1-n_i)) \tag{3}$$

where $r_i(k)$ is the simplicity of $r_i\left(e_i(k), u_i(k-n_i)\right) = (1/2)e_i^2(k) + (\gamma_u/2)[u_i(k-n_i) + e_i(k-n_i)]^2$; $0 < \gamma \leq 1$ is the discount factor.

From Bellman optimality principle and (3), the coupled HJB equation can be derived as $J_i^*(e_i(k)) = \min_{u_i(k)}\left(r_i(k) + \gamma J_i^*(e_i(k+1))\right)$, where $J_i^*(e_i(k))$ represents the optimal local performance index function. The corresponding optimal control law can be formulated as $u_i^*(k) = \arg\min_{u_i(k)}\left(r_i(k) + \gamma J_i^*\left(e_i(k+1)\right)\right)$.

3. **Distributed Optimal Controller Design Based on RL and FRENs.**

3.1. **Critic FREN and weight update law.**

A function with ideal parameters is used to approximate the unknown optimal index function $J_i^*(k)$ for each agent i: $J_i^*(k) = \beta_{ci}^{*T}\varphi_{ci}(k)$, $i = 1, 2, \ldots, N$ where $\beta_{ci}^* \in R^{m_{ci}}$ is the unknown parameter vectors for the regression vectors $\varphi_{ci}(k) \in R^{m_{ci}}$. The regression vectors will be established by a set of membership functions of FRENs to cover the operating range of output $z_i(k)$ with the number of membership m_{ci} for $J_i^*(k)$.

Let $\hat{J}_i(k) = \beta_{ci}^T(k)\varphi_{ci}(k)$ be an approximation of the unknown cost function $J_i^*(k)$, which has the FREN structure. $\beta_{ci}(k)$ is the adjustable parameter vector for the regression vectors $\varphi_{ci}(k)$ in FREN $J_i^*(k)$.

Define a prediction error $e_{ci}(k) = \gamma\hat{J}_i(k) - \hat{J}_i(k-1) + r_i(k)$ of the critic FREN.

Similarly, the critic network FREN weight estimation error can be defined as $\tilde{\beta}_{ci}(k) = \beta_{ci}(k) - \beta_{ci}^*$. Further, the approximation error is defined as $\zeta_{ci}(k) = \tilde{\beta}_{ci}^T\varphi_{ci}(k)$.

Thus, by substituting $\zeta_{ci}(k)$ into $e_{ci}(k)$, it follows that $e_{ci}(k) = \gamma\zeta_{ci}(k) + \gamma J_i^*(k) - \zeta_{ci}(k-1) - J_i^*(k-1) + r_i(k)$.

Then, the weight tuning algorithms for the critic FREN will be discussed.

Firstly, we define an objective function $E_{ci}(k) = (1/2)e_{ci}^2(k)$, which is a quadratic function of the tracking errors $e_{ci}(k)$ and will be minimized by the critic FREN.

The updating law of the weight vector $\beta_{ci}(k)$ of the critic FREN $\hat{J}_i(k)$ is designed according to a standard gradient-based adaptation method and is given by $\beta_{ci}(k+n) = \beta_{ci}(k) + \Delta\beta_{ci}(k)$, where $\Delta\beta_{ci}(k) = \gamma_c\left[-\partial E_{ci}(k)/\partial\beta_{ci}(k)\right]$ with $\alpha_{ci} \in R$ being the adaptation gain.

Combining $e_{ci}(k)$, $E_{ci}(k)$ and $\Delta\beta_{ci}(k)$, the updating law of the weights of the critic FREN can be derived by using the chain rule as

$$\Delta\beta_{ci}(k) = -\gamma_c\frac{\partial E_{ci}(k)}{\partial e_{ci}(k)}\frac{\partial e_{ci}(k)}{\partial \hat{J}_i(k)}\frac{\partial \hat{J}_i(k)}{\partial\beta_{ci}(k)} = -\gamma_c\gamma\varphi_{ci}(k)\left(\gamma\hat{J}_i(k) + r(k) - \hat{J}_i(k-1)\right) \tag{4}$$

Thus, the weight updating law of the critic FREN can be rewritten as

$$\beta_{ci}(k+n_i) = \beta_{ci}(k) - \gamma_c\gamma\varphi_{ci}(k)\left(\gamma\hat{J}_i(k) + r(k) - \hat{J}_i(k-1)\right) \tag{5}$$

3.2. **Action FREN and weight update law.** From the input-output form (2) of the MAS, define a direction function for each agent as $g_i(\cdot) = \partial\varphi_i\left(\underline{z}_i(k), u_i(k)\right)/\partial u_i(k)$.

From the definition of the tracking error $e(k)$ and the non-singular matrix $L+B$, there exists a transformation matrix T which can change the matrix into diagonal or Jordan standard form. Then, both sides of equation $e(k)$ multiply matrix T^{-1} and T as

$$T^{-1}e(k)T = \Lambda\left(y(k) - \underline{y}_0(k)\right) \tag{6}$$

where $\Lambda = T^{-1}(L+B)T$. Then, left multiplying the both sides of (6) by matrix Λ^{-1}, it follows that $e_T(k) = \Lambda^{-1}T^{-1}e(k)T = y(k) - \underline{y}_0(k)$, where $e_T(k) = [e_{T1}(k), e_{T2}(k), \ldots, e_{TN}(k)]^T$, with $e_{T_i}(k+n) = \varphi_i\left(\underline{z}_i(k), u_i(k)\right) - y_0(k+n_i)$.

It is easy to show that $\partial\left(\varphi_i\left(\underline{z}_i(k), u_i(k)\right) - y_0(k+n_i)\right)/\partial u_i(k) \neq 0$.

Therefore, according to Lemma 2 in [18], there exists an ideal control input $u_i^*\left(\bar{z}_i(k)\right)$ such that $\varphi_i\left(\underline{z}_i(k), u_i^*\left(\bar{z}_i(k)\right)\right) = y_0(k+n_i)$, where $\bar{z}_i(k) = \left[\underline{z}_i^T(k), y_0(k+n_i)\right]$.

Using the ideal control $u_i^*\left(\bar{z}_i(k)\right)$, we have $e_{T_i}(k) = 0$ after n_i steps. It implies that the ideal control $u_i^*\left(\bar{z}_i(k)\right)$ is an n-step deadbeat control.

Then, we define an auxiliary error function for each agent to derive the optimal control law as

$$e_{ui}(k+n_i) = \varphi_i\left(\underline{z}_i(k), u_i(k)\right) - y_0(k+n_i) = g_i\left(\underline{z}_i(k), u_i^c(k)\right)\left(u_i(k) - u_i^*\left(\bar{z}_i(k)\right)\right) \tag{7}$$

where $g_i\left(\underline{z}_i(k), u_i^c(k)\right) = \partial\varphi_i\left(\underline{z}_i(k), u_i^c(k)\right)/\partial u_i^c(k)$ with $u_i^c(k) \in [\min\left\{u_i^*\left(\bar{z}_i(k)\right), u_i(k)\right\}, \max\left\{u_i^*\left(\bar{z}_i(k)\right), u_i(k)\right\}]$. For convenience, let us introduce the following notations $g_i(k) = g_i\left(\underline{z}_i(k), u_i^c(k)\right)$.

By using the approximation function with idea parameters, the unknown nonlinear functions $g_i(k)$ and $g_i(k)u_i^*\left(\bar{z}_i(k)\right)$ can be rewritten as $g_i(k) = \beta_{gi}^{*T}\varphi_{gi}(k)$, $g_i(k)u_i^*\left(\bar{z}_i(k)\right) = \beta_{fi}^{*T}\varphi_{fi}(k)$, where $\varphi_{fi}(k) \in R^{m_{fi}}$ and $\varphi_{gi}(k) \in R^{m_{gi}}$ are regression vectors and $\beta_{fi}^* \in R^{m_{fi}}$ and $\beta_{gi}^* \in R^{m_{gi}}$ are unknown parameters for those regression vectors. The regression vectors will be established by a set of membership functions of FRENs to cover operating range of output of output $z_i(k)$ with the numbers of membership m_{fi} and m_{gi} for $g_i(k)$ and $g_i(k)u_i^*\left(\bar{z}_i(k)\right)$, respectively.

The estimated nonlinear functions of the unknown parameters β_{gi}^{*T} and β_{fi}^{*T} can be obtained by $\hat{g}_i(k) = \beta_{gi}^T(k)\varphi_{gi}(k)$, $\hat{f}_i(k) = \beta_{fi}^T(k)\varphi_{fi}(k)$, where $\beta_{gi}^T(k)$ and $\beta_{fi}^T(k)$ are adjustable parameters of FRENs.

By using $g_i(k)$ and $g_i(k)u_i^*\left(\bar{z}_i(k)\right)$, the error $e_{T_i}(k+n)$ can be rewritten as

$$e_{T_i}(k+n) = \beta_{gi}^{*T}(k)\varphi_{gi}(k)u_i(k) - \beta_{fi}^{*T}(k)\varphi_{fi}(k) = \beta_i^{*T}\varphi_i(k) \tag{8}$$

where $\beta_i^* = \left[\beta_{gi}^{*T}, \beta_{fi}^{*T}\right]^T$, $\varphi_i(k) = [\varphi_{gi}(k)u_i(k), -\varphi_{fi}(k)]$. According to $\hat{g}_i(k)$ and $\hat{f}_i(k)$, the estimation of $e_{T_i}(k+n)$ can be obtained as $\hat{e}_{T_i}(k+n_i) = \beta_{gi}^T(k)\varphi_{gi}(k)u_i(k) - \beta_{fi}^T(k)\varphi_{fi}(k)$ $= \beta_i^T(k)\varphi_i(k)$ where $\beta_i = \left[\beta_{gi}^T, \beta_{fi}^T\right]^T$. Let us define the estimation error $\tilde{e}_{T_i}(k+n_i) = \hat{e}_{T_i}(k+n_i) - e_{T_i}(k+n_i) = \tilde{\beta}_i^T(k)\varphi_i(k)$, where $\tilde{\beta}_i(k) = \beta_i^* - \beta_i(k)$.

To get a better estimation result, a cost function will be defined before adjusting the parameter $\beta_i(k)$ as $E_{ai}(k+n_i) = (1/2)\left(\left(\tilde{e}_{T_i}(k+n_i) + \hat{J}_i(k)\right)^2 \Big/ \left(\gamma_b + \|\varphi_i(k)\|^2\right)\right)$, where $\gamma_b \geq 1$ is a positive constant. To minimize the cost function $E_{ai}(k+n_i)$, the parameters can be tuned along the steepest descent direction. The corresponding tuning law is established as follows: $\beta_{1i}(k+n_i) = \beta_i(k) - \eta_i\left(\partial E_{ai}(k+n_i)/\partial\beta_i(k)\right)$, where η_i is a design learning rate. By applying the chain rule with $\hat{e}_{T_i}(k+n_i)$ and $\tilde{e}_{T_i}(k)$, thus, we obtain

$$\frac{\partial E_{ai}(k+n_i)}{\partial\beta_i(k)} = \frac{\partial E_{ai}(k+n_i)}{\partial\tilde{e}_{T_i}(k+n_i)} \cdot \frac{\partial\tilde{e}_{T_i}(k+n_i)}{\partial\hat{e}_{T_i}(k+n_i)} \cdot \frac{\partial\hat{e}_{T_i}(k+n_i)}{\partial\beta_i(k)} = \frac{\tilde{e}_{T_i}(k+n_i) + \hat{J}_i(k)}{\gamma_b + \|\varphi_i(k)\|^2} \cdot \varphi_i(k) \tag{9}$$

By substituting (9) into $\beta_{1i}(k+n_i)$, we have

$$\beta_{1i}(k+n_i) = \beta_i(k) - \eta_i \left(\left(\tilde{e}_{T_i}(k+n_i) + \hat{J}_i(k) \right) \Big/ \left(\gamma_b + \|\varphi_i(k)\|^2 \right) \right) \varphi_i(k) \tag{10}$$

However, the tuning law (10) does not have the ability to deal with the uncertain control directions. Let us define a previous error $\tilde{e}_{hi}$ as $\tilde{e}_{hi}(k+n_i-j) = e_{T_i}(k+n_i-j) - \beta_i^T(k)\varphi_i(k-j)$ for $j = 0, 1, \ldots, v-1$, when v is the order of the previous error. By adding the previous errors, a new tuning law can be obtained as

$$\beta_{2i}(k+n_i) = \beta_{1i}(k) - \eta_i \sum_{j=0}^{v-1} \left(\tilde{e}_{hi}(k+n_i-j)\varphi_i(k-j) \Big/ \left(\gamma_b + \|\varphi_i(k-j)\|^2 \right) \right) \tag{11}$$

where η_i satisfies $\eta_i > 0$.

Finally, a projection algorithm is introduced to the tuning law to ensure the boundedness of parameter $\beta_i(k)$. When $\|\beta_{2i}(k)\| \leq N_{0i}$, it has $\beta_i(k) = \beta_{2i}(k)$; otherwise, let $\beta_i(k) = (N_{0i}/\|\beta_{2i}(k-n_i)\|)\,\beta_{2i}(k)$.

The adaptive control object is minimized of the long-term cost function $J_i(k)$ which can be converted to minimize the Lagrangian $r_i(k)$. By tracking the partial derivative of $r_i(k)$ with $u_i(k)$ and using the error dynamic in (8), we obtain $\partial r_i(k+n_i)/\partial u_i(k) = \left(\gamma_u + \|g_i(k)\|^2\right) u_i(k) - g_i(k)\varphi_{fi}^T(k)\beta_{fi}^* + \gamma_u e_i(k)$.

Setting $\partial r_i(k+n)/\partial u_i(k) = 0$. Thus, the ideal control law $u_i^*(k)$ can be obtained as $u_i^*(k) = \left(\beta_{fi}^{*T}\varphi_{fi}(k)g_i(k) - \gamma_u e_i(k)\right) / \left(\gamma_u + \|g_i(k)\|^2\right)$.

Functions β_{fi}^* and $g_i(k)$ are unknown but have been approximated as $\beta_{fi}(k)$ and $\hat{g}_i(k)$, respectively. Thus, the practical control law $u_i(k)$ can be given as $u_i(k) = \left(\beta_{fi}^T(k)\varphi_{fi}(k)\right.$ $\left. g_i(k) - \gamma_u e_i(k) \right/ \gamma_u + \|\hat{g}_i(k)\|^2\big)$. It is clear that the practical control law can be obtained even the system dynamic has positive or negative on $\hat{g}_i(k)$.

4. **Experimental Results.** In this section, a multi-manipulator system which consists of a leader agent and five heterogeneous follower agents has been considered to validate the proposed distributed adaptive controller. The dynamic models of the individual follower agent are totally unknown and are described as

$$\begin{aligned}
\Sigma_1 &: \begin{cases} \dot{x}_{11} = u_1 - 0.1u_1 e^{-0.1}x_{11} - 0.1\left(1 - e^{-0.2t}\right)\sin(0.1t) \\ y_1 = x_{11} \end{cases} \\
\Sigma_2 &: \begin{cases} \dot{x}_{21} = -x_{22} \\ \dot{x}_{22} = -3.2\sin(x_{21}) - 3x_{22} - 10\cos(x_{21})\tanh(u_2) \\ y_2 = x_{21} \end{cases} \\
\Sigma_3 &: \begin{cases} \dot{x}_{31} = u_3 + 0.2u_3 e^{-x_{31}} + 0.1\left(1 - e^{-t}\right)\sin(0.3t) \\ y_3 = x_{31} \end{cases} \\
\Sigma_4 &: \begin{cases} \dot{x}_{41} = \left(1.4x_{41}^2/\left(1 + x_{41}^2\right)\right) + 0.1x_{42}^3 + 0.5x_{42} \\ \dot{x}_{42} = \left(0.1x_{42}/\left(1 + x_{41}^2 + x_{42}^2\right)\right) + \left(5x_{41}^2\tanh(u_4)/\left(1 + x_{41}^2 + x_{42}^2\right)\right) \\ y_4 = x_{41} \end{cases} \\
\Sigma_5 &: \begin{cases} \dot{x}_{51} = x_{52} \\ \dot{x}_{52} = \sin(x_{51}) - x_{52} + 3\cos(x_{52})\tanh(u_5) \\ y_5 = x_{51} \end{cases}
\end{aligned}$$

The communication graph of the multi-manipulator system is depicted in Figure 1.

The dynamic model of the leader mode is given by $y_0(k+1) = (1/2)\sin(0.01\pi k/5) + (1/2)\cos(0.01\pi k/10)$. Two kinds of parameters need to be designed. The first is the kind of global parameters: $\gamma = 0.9$, $\gamma_b = 1$, $v = 2$, $\gamma_u = 1$ and $\gamma_c = 10^{-4}$, and the other is that of local parameters of each distributed controller whose values are listed in Table 1.

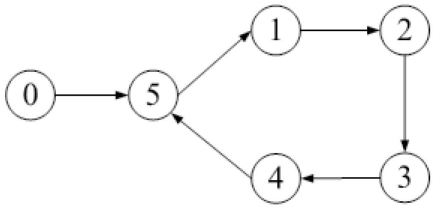

FIGURE 1. Communication graph for the simulation examples

TABLE 1. Parameters setting

Parameter	Agent 1	Agent 2	Agent 3	Agent 4	Agent 5
η_i	0.1	0.02	0.1	1	0.02
N_{0i}	4	2	4	200	2

The trajectories of the output are depicted in Figure 2. Figure 3 shows the distributed control efforts of the five agents. The local neighborhood errors of the five followers are plotted in Figure 4. The estimated cost functions $\hat{J}_i$, $i = 1, 2, 3, 4, 5$ are shown in Figure 5.

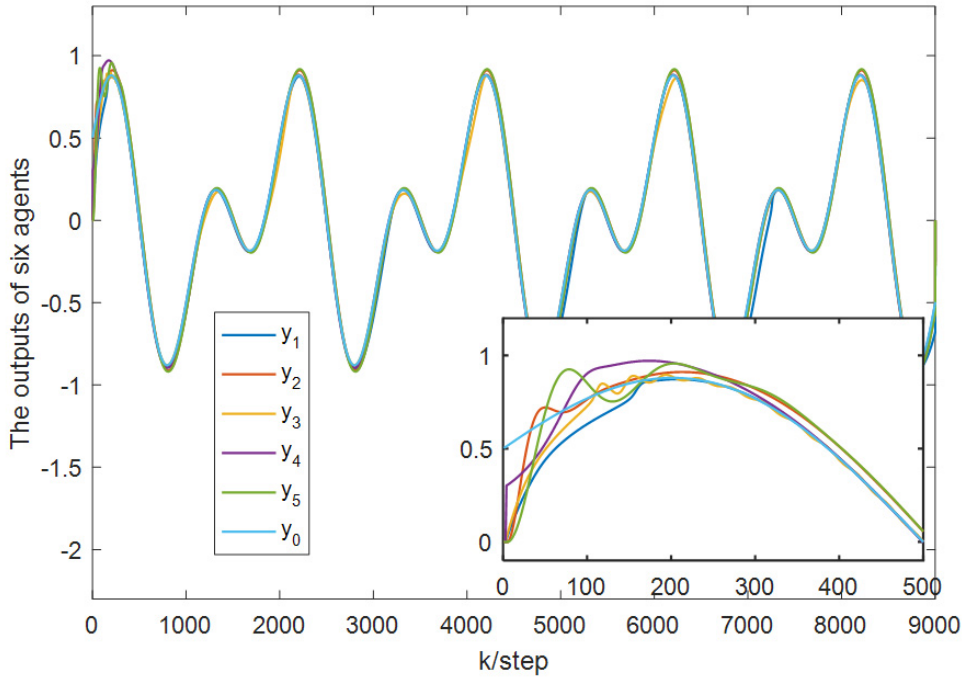

FIGURE 2. (color online) Output trajectories of six agents

From the dynamics of the simulation system, it can be seen that the five agents have positive and negative control directions. From Figures 2-5, one can observe that all the five followers synchronize to the leader with a bounded neighborhood and all signals in the closed-loop distributed control system are bounded.

In Figure 2, one can conclude that the initial tracking performances of the five followers are not ideal, which can also be seen from the big neighborhood tracking errors in Figure 4. This is owing to that the weights of FRENs are still tuning. After about 300 steps, it can be observed that the tracking performance improves to be much better. From Figure 4, it can be obtained that the local neighborhood tracking errors of the five followers are converged asymptotically to a small neighborhood of zero. Therefore, the simulation results illustrate

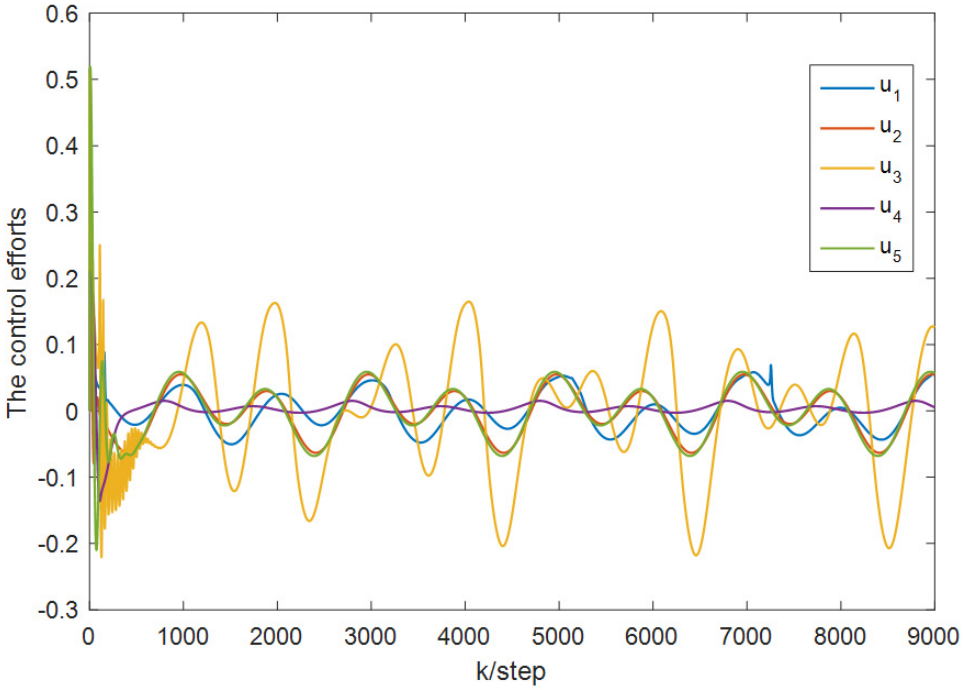

FIGURE 3. (color online) Control efforts of five followers

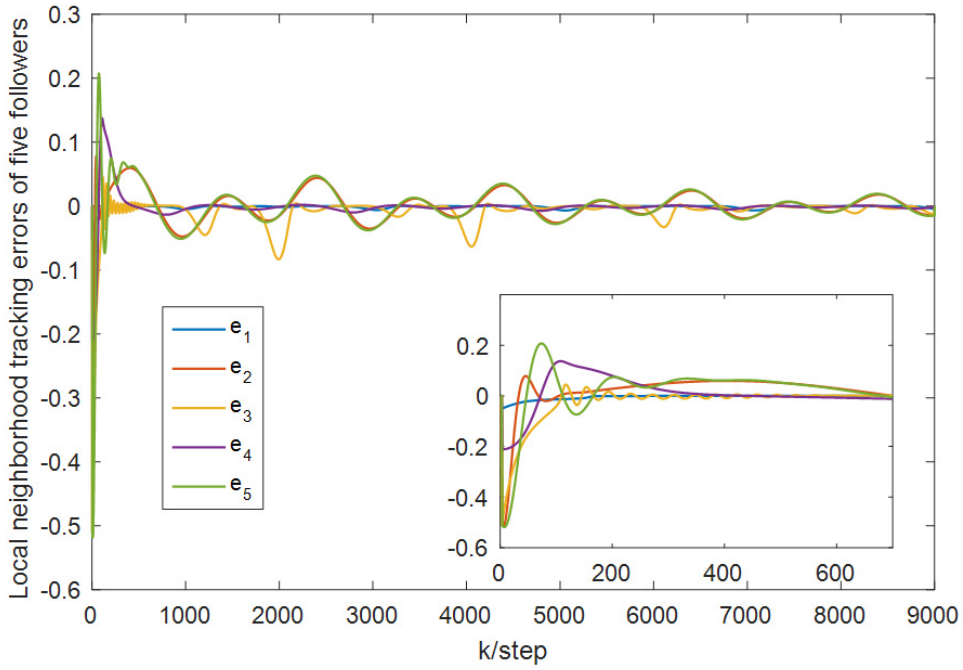

FIGURE 4. (color online) Local neighborhood tracking errors of five followers

the effectiveness of the proposed consensus control method for the heterogeneous nonlinear MASs.

5. **Conclusion.** In this paper, the distributed optimal consensus control strategy based on RL and FRENs for a class of heterogeneous discrete-time nonlinear MASs with unknown dynamics and uncertain control directions has been developed. The optimal solutions of the coupled HJB equations have been obtained by the RL algorithm implemented by an actor-critic form with FRENs which are used to estimate the unknown dynamics

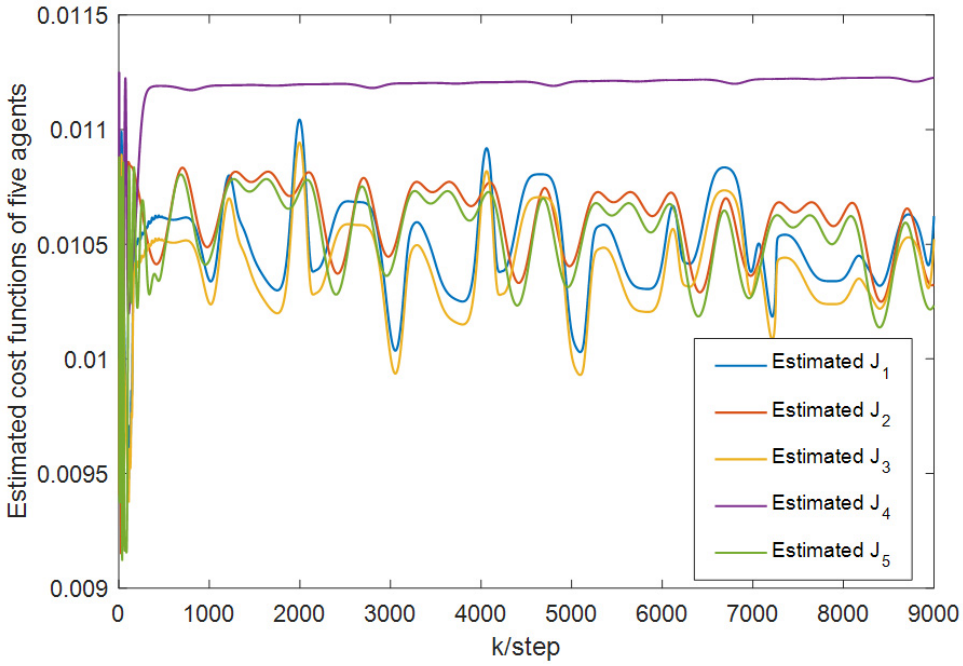

FIGURE 5. (color online) Estimated unknown cost functions of five followers

and the uncertain control directions. Simulation studies have been conducted on a heterogeneous multi-manipulator system to show the effectiveness of the proposed control strategy.

In the future work, practical issues should be considered in the situation that noises and un-modeled dynamics exist.

Acknowledgment. The work is supported by the National Natural Science Foundation of China under Grant 62003205, Zhejiang Provincial Natural Science Foundation of China under Grant LQ19F030005; Natural Science Foundation of Ningbo under Grant 2019A610092. The authors also gratefully acknowledge the helpful comments and suggestions of the reviewers, which have improved the presentation.

REFERENCES

[1] J. Zhao, Neural networks-based optimal tracking control for nonzero-sum games of multi-player continuous-time nonlinear systems via reinforcement learning, *Neurocomputing*, vol.412, pp.167-176, 2020.

[2] M. Zanon and S. Gros, Safe reinforcement learning using robust MPC, *IEEE Trans. Automatic Control*, vol.66, no.8, pp.3638-3652, DOI: 10.1109/TAC.2020.3024161, 2021.

[3] W. Zhao, H. Liu and F. L. Lewis, Robust formation control for cooperative underactuated quadrotors via reinforcement learning, *IEEE Trans. Neural Networks and Learning Systems*, vol.32, no.10, pp.4577-4587, DOI: 10.1109/TNNLS.2020.3023711, 2021.

[4] X. Yang, D. Liu, B. Luo et al., Data-based robust adaptive control for a class of unknown nonlinear constrained-input systems via integral reinforcement learning, *Information Sciences*, vol.369, pp.731-747, 2016.

[5] C. Treesatayapun and S. Uatrongjit, Adaptive controller with fuzzy rules emulated structure and its applications, *Engineering Applications of Artificial Intelligence*, vol.18, no.5, pp.603-615, 2005.

[6] L. Facundo, J. Gmez, C. Treesatayapun et al., Adaptive control with sliding mode on a double fuzzy rule emulated network structure, *IFAC-PapersOnLine*, vol.51, no.13, pp.609-614, 2018.

[7] C. Treesatayapun, Knowledge-based reinforcement learning controller with fuzzy-rule network: Experimental validation, *Neural Computing and Applications*, vol.32, no.13, pp.9761-9775, 2020.

[8] M. Huang, L. Tao and Z. Hu, Command filter adaptive power capture control based on fuzzy rules emulated networks for variable speed wind turbines with flexible shaft, *IEEE Access*, vol.9, pp.91377-91386, 2021.

[9] T. Wang, H. Fu, J. Li et al., Optimal consensus control for heterogeneous nonlinear multiagent systems with partially unknown dynamics, *International Journal of Control, Automation and Systems*, vol.17, no.9, pp.2400-2413, 2019.

[10] X. Feng, Y. Yang and D. Wei, Adaptive fully distributed consensus for a class of heterogeneous nonlinear multi-agent systems, *Neurocomputing*, vol.428, pp.12-18, 2021.

[11] J. Song, Observer-based consensus control for networked multi-agent systems with delays and packet-dropouts, *International Journal of Innovative Computing, Information and Control*, vol.12, no.4, pp.1287-1302, 2016.

[12] J. Wang, Network-based containment control protocol of multi-agent systems with time-varying delays, *International Journal of Innovative Computing, Information and Control*, vol.12, no.6, pp.2089-2098, 2016.

[13] C. Deng and C. Wen, Distributed resilient observer-based fault-tolerant control for heterogeneous multiagent systems under actuator faults and DoS attacks, *IEEE Trans. Control of Network Systems*, vol.7, no.3, pp.1308-1318, 2020.

[14] C. Deng, C. W. Wen and Z. G. Wu, A dynamic periodic event-triggered approach to consensus of heterogeneous linear multiagent systems with time-varying communication delays, *IEEE Trans. Cybernetics*, vol.51, no.4, pp.1812-1821, 2020.

[15] M. Xu, P. Yang, Y. Wang and Q. Shu, Observer-based multi-agent system fault upper bound estimation and fault-tolerant consensus control, *International Journal of Innovative Computing, Information and Control*, vol.15, no.2, pp.519-534, 2019.

[16] F. Zhang and W. Wang, Decentralized optimal control for the mean field LQG problem of multi-agent systems, *International Journal of Innovative Computing, Information and Control*, vol.13, no.1, pp.55-66, 2017.

[17] B. Yan, C. Wu and P. Shi, Formation consensus for discrete-time heterogeneous multi-agent systems with link failures and actuator/sensor faults, *Journal of the Franklin Institute*, vol.356, no.12, pp.6547-6570, 2019.

[18] C. Yang, S. S. Ge, C. Xiang, T. Chai and T. H. Lee, Output feedback NN control for two classes of discrete-time systems with unknown control directions in a unified approach, *IEEE Trans. Neural Networks*, vol.11, no.1, pp.1873-1886, 2008.