SCIENCE CHINA
Information Sciences

• RESEARCH PAPER •

November 2025, Vol. 68, Iss. 11, 210209:1–210209:15
https://doi.org/10.1007/s11432-025-4587-1

Special Topic: Mean-Field Game and Control of Large Population Systems: From Theory to Practice

Model-free group formation control of heterogeneous nonlinear multi-agent systems

Chuanjian LI[1,2,3], Xiaoping WANG[1,2,3*], Chen WEI[1,2,3], Fangmin REN[1,2,3], Xiaofeng ZONG[4], Zhigang ZENG[1,2,3] & Tingwen HUANG[5]

[1]*School of Artificial Intelligence and Automation, Huazhong University of Science and Technology, Wuhan 430074, China*
[2]*Key Laboratory of Image Processing and Intelligent Control of Education Ministry of China, Wuhan 430074, China*
[3]*Hubei Key Laboratory of Brain-Inspired Intelligent Systems, Huazhong University of Science and Technology, Wuhan 430074, China*
[4]*School of Automation, China University of Geosciences, Wuhan 430074, China*
[5]*Faculty of Computer Science and Control Engineering, Shenzhen University of Advanced Technology, Shenzhen 518055, China*

Received 5 February 2025/Revised 27 April 2025/Accepted 21 June 2025/Published online 30 October 2025

Abstract Traditionally, the designs of control policies for the group formation control (GFC) problem of multi-agent systems (MASs) rely on accurate system dynamics, which have been proven challenging to obtain in the complex real world, leading to unknown MASs. In this respect, this paper investigates the distributed GFC problem of heterogeneous nonlinear MASs with unknown dynamics. First, an effective and more flexible communication topology is designed to achieve communication configuration among agents. Then, the GFC problem is formulated by defining the local group neighborhood formation error and performance index for each agent under the designed communication topology. By developing an effective policy iteration algorithm and establishing a corresponding actor-critic neural network framework, a novel model-free GFC algorithm is proposed, overcoming the reliance on system dynamics for control design. The proposed model-free algorithm can seek the optimal GFC control policy online using system operation data to achieve GFC and minimize control cost, thus yielding better control performance and demonstrating superior practicality in practical applications compared with traditional offline model-based methods. Finally, two simulation examples demonstrate the effectiveness and superiority of the developed model-free algorithm for solving the GFC problem.

Keywords multi-agent systems, group formation control, model-free, heterogeneous nonlinear dynamics

Citation Li C J, Wang X P, Wei C, et al. Model-free group formation control of heterogeneous nonlinear multi-agent systems. Sci China Inf Sci, 2025, 68(11): 210209, https://doi.org/10.1007/s11432-025-4587-1

1 Introduction

Formation control, which aims to drive a group of agents to achieve and maintain a certain formation, has become a prominent topic in the field of multi-agent systems (MASs) because of its potential applications in cooperative surveillance [1], resource seeking [2], and target enclosing [3]. Distributed formation control approaches, which utilize local neighboring relative interactions, have been developed extensively for homogeneous linear MASs, such as first-order [4], second-order [5] and high-order [6, 7] MASs.

In the real world, physical plants typically exhibit nonlinear characteristics. Furthermore, agents within an entire system may display different dynamics—a typical example being the unmanned aerial vehicle-unmanned ground vehicle system. Such systems are called heterogeneous systems, and significant formation control results have been obtained for heterogeneous nonlinear MASs, such as discrete-time MASs [8] and the heterogeneous MASs modeled using the Euler-Lagrange equation [9]. However, the above formation control results rely heavily on complete system models, which have been proven challenging to obtain in the complex real world, leading to partially or completely unknown MASs. Until now, many excellent results have been obtained for the formation control problem of heterogeneous nonlinear MASs with unknown dynamics [10–17]. For example, Ref. [10] developed an efficient formation control framework using the iterative learning approach. Meanwhile, using an identifier-actor-critic reinforcement learning algorithm, Refs. [11, 12] addressed the optimal leader-follower formation control problem.

* Corresponding author (email: wangxiaoping@hust.edu.cn)

The aforementioned formation control results for MASs, where all agents form a single formation, encounter limitations such as reduced work efficiency and restricted operational scope in scenarios that require the simultaneous execution of multiple tasks, such as search-and-rescue operations and multi-target surveillance [18–20]. In these scenarios, the entire MAS must be partitioned into several subgroups to accomplish distinct tasks. Consequently, the group formation control (GFC) problem emerges, wherein agents within each subgroup achieve a predefined subformation, allowing for multiple distinct subformations within the MAS. Dong et al. [21] and Hu et al. [22] studied GFC problems of second-order linear MASs. Dong et al. [23] further developed an effective control protocol for homogeneous general linear MASs. For the same MASs, Ge et al. [24] considered the case where aperiodic sampling and communication delays coexist simultaneously; Hu et al. [25] proposed a fully distributed GFC method; Tian et al. [26] designed a novel adaptive control protocol under switching topologies; Qi et al. [27] developed a GFC algorithm based on distributed multi-sensor multi-target filtering with intermittent observations; and Li et al. [28] considered the collision avoidance problem. For nonlinear MASs, Han et al. [29] investigated the GFC problem of second-order systems. Wang et al. [30] and Wu et al. [31] solved GFC problems of second-order heterogeneous systems using event-triggered and impulsive control methods, respectively. Du et al. [32] designed a GFC protocol for heterogeneous MASs consisting of linear and Euler-Lagrange dynamic agents.

Notably, the aforementioned GFC results rely heavily on accurate system dynamics. Recently, some remarkable results have also been obtained for the GFC problem of MASs with unknown dynamics. For instance, Wang et al. [33] proposed a model-free optimal GFC algorithm for linear heterogeneous MASs based on the algebraic Riccati equation derived from linear systems. Meanwhile, Shi et al. [34] studied the robust output GFC problem of linear heterogeneous MASs with uncertainties. Notably, these previous studies [33,34] also involved the optimal coordination control problem—which is one of the most popular topics for MASs and one of the concerns of this paper—to design effective control policies that not only enabled MASs to achieve coordination control but also minimized predefined performance indices associated with control performance, energy consumption, and time spending. Evidently, although some results on the GFC problem of MASs with unknown dynamics have been obtained, these all focus on linear MASs and require the strict acyclic partition condition for communication topology, which reduces the applicability of the obtained results and the flexibility of agent interaction. In addition, Shi et al. [34] required additional identification for system dynamics, thus adding an extra computational burden. To the best of the authors' knowledge, no model-free GFC results are currently available for heterogeneous nonlinear MASs. Therefore, this paper aims to develop an effective model-free GFC algorithm for such systems while addressing the aforementioned limitations.

Motivated by the above discussion, this paper is dedicated to studying the GFC problem of heterogeneous nonlinear MASs with unknown dynamics. To achieve the communication configuration for MASs, an effective and more flexible communication topology is first designed. Then, a novel model-free GFC algorithm is proposed by developing an effective policy iteration (PI) algorithm and establishing a corresponding actor-critic neural network (NN) framework. The detailed and specific contributions are presented below.

(1) Different from existing model-free GFC studies for linear MASs [33, 34], a novel model-free GFC algorithm is developed for heterogeneous nonlinear MASs. The algorithm does not require additional identification for system dynamics and applies to both linear and nonlinear MASs, thereby demonstrating a broader application scope.

(2) The proposed model-free GFC algorithm not only achieves GFC but also determines the optimal control policy in real-time to minimize control cost. Compared with previous offline model-based methods [25–32], the developed online model-free algorithm demonstrates superior practicality in real applications.

(3) A more flexible communication topology is proposed for the GFC problem. This communication topology eliminates the requirement for the strict acyclic partition condition [21, 23–25, 33, 34], thereby increasing the flexibility of communication configuration and expanding the range of topology selection.

The novelties of this work compared with existing studies are listed in Table 1 [21–34].

The rest of this paper is structured as follows. Section 2 establishes the GFC problem for heterogeneous nonlinear MASs. Section 3 develops a novel online model-free algorithm to solve the GFC problem. Two simulation examples are presented in Section 4. Some summaries and prospects for this work are given in Section 5.

Notations. $\mathbb{R}^n$ and $\mathbb{R}^{n\times m}$ denote the n-dimensional vector and $n\times m$-dimensional real matrix, respectively. I_n and $\mathbf{1}_n$ represent the n-dimensional identity matrix and column vector with all elements

Table 1 Comparison with existing studies.

Existing studies	Nonlinear	Heterogeneous	Model-free	Online	Optimality
[21–28]	✗	✗	✗	✗	✗
[29]	✓	✗	✗	✗	✗
[30–32]	✓	✓	✗	✗	✗
[33, 34]	✗	✓	✓	✓	✓
This work	✓	✓	✓	✓	✓

equal to one, respectively. $\otimes$ is the Kronecker product of the matrices. For a matrix or vector P, P^{T} and $\|P\|$ represent its transpose and Euclidean norm, respectively. For a symmetric matrix H, $H > 0$ (or $H \geqslant 0$) means that H is positive definite (or semi-positive definite).

2 Preliminaries

2.1 Description of communication topology

The communication topology for MASs can be described using the weighted directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{A})$ with node set $\mathcal{V} = \{1, \ldots, i, \ldots, N\}$, edge set $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$, and weighted adjacency matrix $\mathcal{A} = [a_{ij}] \in \mathbb{R}^{N\times N}$, where i stands for the ith agent, and $a_{ij} \neq 0$ or $a_{ij} = 0$ indicates whether or not there is direct information flow from j to i. Moreover, assume $a_{ii} = 0$ for $\forall i$. Denote the neighbor set of i as N_i with $a_{ij} \neq 0$ for $j \in N_i$. The Laplacian matrix of $\mathcal{G}$ is expressed by $\mathcal{L} = \mathcal{D} - \mathcal{A}$, where $\mathcal{D} = \mathrm{diag}(d_i)$ with $d_i = \sum_{j\in N_i} a_{ij}$. The edge sequence $(i-1, i), (i, i+1), \ldots$ is called a directed path. $\mathcal{G}$ containing a directed spanning tree indicates that there exists at least one node (also called the root node) that has a directed path to all other nodes.

For the GFC problem, suppose that the graph $\mathcal{G}$ can be divided into M subgraphs $\{\mathcal{G}_1, \ldots, \mathcal{G}_M\}$ (corresponding to M subgroups $\{\Omega_1, \ldots, \Omega_M\}$) with node sets $\{\mathcal{V}_1, \ldots, \mathcal{V}_M\}$ satisfying $\mathcal{V}_i \cap \mathcal{V}_j = \emptyset$ and $\cup_{i=1}^{M} \mathcal{V}_i = \mathcal{V}$. Let $\bar{i}$ denote the subgroup index to which agent i belongs (also the corresponding leader index below). Then, on the basis of the above subgroup division, $\mathcal{L}$ can be described as the following block form

$$\mathcal{L} = \begin{bmatrix} \mathcal{L}_{11} & \mathcal{L}_{12} & \cdots & \mathcal{L}_{1M} \\ \mathcal{L}_{21} & \mathcal{L}_{22} & \cdots & \mathcal{L}_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ \mathcal{L}_{M1} & \mathcal{L}_{M2} & \cdots & \mathcal{L}_{MM} \end{bmatrix}.$$

Assumption 1. Each subgraph $\mathcal{G}_{\bar{i}}, \bar{i} = 1, \ldots, M$, contains a spanning tree. Moreover, for all $k \neq l \in \{1, 2, \ldots, M\}$, $\mathcal{L}_{kl}$ has entries in each row that sum to zero.

Remark 1. Notably, the GFC problem involves coupling between multiple subgroups as communication interactions exist both within and between them. This maintains the integrity of the MAS. $\mathcal{L}_{kl}, k \neq l$, indeed, can be regarded as the interaction matrix block between subgroups k and l. The above properties serve as a basis for grouping and provide valuable insights on how to efficiently utilize inter-subgroup communication to achieve GFC. Moreover, the communication topology designed in this paper does not require the acyclic partition condition imposed previously [21, 23–25, 33, 34], thereby enhancing the flexibility of interactions among agents. The detailed analysis is presented in the following Lemma 1, where Assumption 1 plays a crucial role.

2.2 Problem formulation

The dynamics of agent i is expressed as follows:

$$x_i(k+1) = f_i(x_i(k)) + g_i(x_i(k))\psi_i(k),\ i = 1, \ldots, N, \tag{1}$$

where $x_i(k) \in \mathbb{R}^n$ and $\psi_i(k) \in \mathbb{R}^{m_i}$ represent the state and control policy, respectively. $f_i(\cdot) \in \mathbb{R}^n$ and $g_i(\cdot) \in \mathbb{R}^{n\times m_i}$ are unknown system dynamics satisfying the following general assumption [12].

Assumption 2. $f_i(\cdot)$ and $g_i(\cdot)$ are Lipschitz continuous on a compact set Ω containing the origin with $f_i(0) = 0$, and system (1) is stabilizable on Ω.

For the GFC problem studied in this paper, the leader (also called the subformation reference signal) dynamics for each subgroup is given as

$$x_0^{\bar{i}}(k+1) = f_0^{\bar{i}}(x_0^{\bar{i}}(k)),\ \bar{i} = 1, \ldots, M, \tag{2}$$

where $x_0^{\bar{i}}(k) \in \mathbb{R}^n$ is the state of leader $\bar{i}$.

Remark 2. The leader $\bar{i}$, which can be either a physical agent or a virtual signal, provides the formation reference signal for subgroup $\bar{i}$ to achieve subformation and guides its macroscopic movement. The function $f_0^{\bar{i}}(\cdot)$ can be selected on the basis of specific task requirements and may therefore differ from $f_i(\cdot)$.

In the leader-following case, denote $b_i \geqslant 0$ as the pinning gain of agent i, where $b_i > 0$ or $b_i = 0$ indicates whether or not agent i receives the reference signal of leader $\bar{i}$ directly. Then, the following general assumption is required.

Assumption 3. For each subgroup, $b_i > 0$ for at least one root node i.

Define the group formation tracking error for agent i as

$$z_i(k) = x_i(k) - x_0^{\bar{i}}(k) - \eta_i, i = 1, \ldots, N, \tag{3}$$

where $\eta_i \in \mathbb{R}^n$ is the relative position between agent i and leader $\bar{i}$, which depicts the predefined subformation pattern.

Definition 1 (GFC)**.** For MAS (1) and (2) with any bounded initial states, the GFC is said to be achieved if

$$\lim_{k\to\infty} \|z_i(k)\| = 0,\ i = 1, \ldots, N. \tag{4}$$

According to the interactions of MAS (1) and (2) under graph $\mathcal{G}$, the local group neighborhood formation error for agent i is defined as

$$\begin{aligned}\delta_i(k) = &\sum_{j\in N_i} a_{ij}(x_i(k) - x_0^{\bar{i}}(k) - \eta_i - (x_j(k) - x_0^{\bar{j}}(k) - \eta_j)) \\ &+ b_i(x_i(k) - x_0^{\bar{i}}(k) - \eta_i),\ i = 1, \ldots, N,\end{aligned} \tag{5}$$

where $\delta_i(k) \in \mathbb{R}^n$.

According to (3), $\delta_i(k)$ can be rewritten as

$$\delta_i(k) = \sum_{j\in N_i} a_{ij}(z_i(k) - z_j(k)) + b_i z_i(k),\ i = 1, \ldots, N. \tag{6}$$

Further, we establish the relationship between group neighborhood formation error and group formation tracking error as follows:

$$\delta(k) = ((\mathcal{L} + \mathcal{B}) \otimes I_n)Z(k), \tag{7}$$

where $\delta(k) = [\delta_1^{\mathrm{T}}(k), \ldots, \delta_N^{\mathrm{T}}(k)]^{\mathrm{T}}$, $Z(k) = [z_1^{\mathrm{T}}(k), \ldots, z_N^{\mathrm{T}}(k)]^{\mathrm{T}}$, and $\mathcal{B} = \mathrm{diag}\{b_i\} \in \mathbb{R}^{N\times N}, i = 1, \ldots, N$.

Lemma 1. Suppose that Assumptions 1 and 3 hold. Then, MAS (1) and (2) achieves GFC if

$$\lim_{k\to\infty} \|\delta(k)\| = 0.$$

Proof. Denote the corresponding graph and Laplacian matrix of MAS (1) and (2) containing M leaders and N agents as $\bar{\mathcal{G}}$ and $\bar{\mathcal{L}}$, respectively. Then, under Assumptions 1 and 3, $\bar{\mathcal{L}}$ can be described as the following block form:

$$\bar{\mathcal{L}} = \begin{bmatrix} 0_{M\times M} & 0_{M\times N} \\ \tilde{\mathcal{B}} & \mathcal{L} + \mathcal{B} \end{bmatrix},$$

where $\tilde{\mathcal{B}}$ is composed of the elements of $-\mathcal{B}$, ordered according to the definition of the Laplacian matrix in the topology description.

Letting $\tilde{\mathcal{L}} = [\tilde{\mathcal{B}}, \mathcal{L} + \mathcal{B}]$, from [19], we can obtain that under Assumptions 1 and 3, $\mathrm{rank}(\bar{\mathcal{L}}) = N$, which implies that $\mathrm{rank}(\tilde{\mathcal{L}}) = N$ because all of the entries in the first M rows of $\bar{\mathcal{L}}$ are zero. Given that $\tilde{\mathcal{L}}$ has $N + M$ columns and each row sum is zero, it follows that the first M columns of $\tilde{\mathcal{L}}$ depend on its last N

columns, where $[b_1,\ldots,b_N]^{\mathrm{T}}=(\mathcal{L}+\mathcal{B})\mathbf{1}_N$. Consequently, it follows that $\mathrm{rank}(\mathcal{L}+\mathcal{B})=\mathrm{rank}(\tilde{\mathcal{L}})=N$, which indicates that $(\mathcal{L}+\mathcal{B})$ is nonsingular. According to (7), we have

$$\|Z(k)\|\leqslant\|\delta(k)\|/\underline{\sigma}(\mathcal{L}+\mathcal{B}),\tag{8}$$

where $\underline{\sigma}(\mathcal{L}+\mathcal{B})>0$ is the minimum singular value of $(\mathcal{L}+\mathcal{B})$.

From the above analysis, it is known that $\lim_{k\to\infty}\|\delta(k)\|=0$ entails $\lim_{k\to\infty}\|Z(k)\|=0$. Then, from Definition 1, we know that MAS (1) and (2) achieves GFC. The proof is completed.

Remark 3. Lemma 1 tells us that the proposed communication topology satisfying Assumption 1 is valid for the GFC problem by deriving the relationship between group neighborhood formation error and group formation tracking error based on the nonsingularity of $(\mathcal{L}+\mathcal{B})$. Notice that Assumption 1 does not require the acyclic partition condition imposed previously [21,23–25,33,34], thus increasing the flexibility of communication configuration and expanding the range of topology selection for the GFC problem.

Given MAS (1) and (2) and the local group neighborhood formation error (5), let $x_i(k)$ be denoted as x_{ik}. Then, we have

$$\delta_{i(k+1)}=(d_i+b_i)(f_i(x_{ik})+g_i(x_{ik})\psi_{ik})-\sum_{j\in N_i}a_{ij}(f_j(x_{jk})+g_j(x_{jk})\psi_{jk})-b_if_0^{\bar{i}}(x_{0k}^{\bar{i}}).\tag{9}$$

On the basis of Lemma 1, it can be readily inferred that when $\lim_{k\to\infty}\|\delta_{ik}\|=0$ for each agent i, the GFC is achieved. Therefore, the control objective of this work is achieving $\|\delta_{ik}\|\to0$ by designing effective control policies $\psi_i,i=1,\ldots,N$.

Considering the control performance, define the local performance index for agent i as

$$\begin{aligned}J_i(\delta_{ik},\psi_{ik},\psi_{(j)k})=&\sum_{t=k}^{\infty}\alpha^{t-k}r_i(\delta_{it},\psi_{it},\psi_{(j)t})\\=&r_i(\delta_{ik},\psi_{ik},\psi_{(j)k})+\alpha J_i(\delta_{i(k+1)},\psi_{i(k+1)},\psi_{(j)(k+1)}),\end{aligned}\tag{10}$$

where $\psi_{(j)}=\{\psi_j|j\in N_i\}$, $r_i(\delta_{ik},\psi_{ik},\psi_{(j)k})=\delta_{ik}^{\mathrm{T}}Q_{ii}\delta_{ik}+\psi_{ik}^{\mathrm{T}}R_{ii}\psi_{ik}+\sum_{j\in N_i}\psi_{jk}^{\mathrm{T}}R_{ij}\psi_{jk}$ is the utility function, and $Q_{ii}>0\in\mathbb{R}^{n\times n},R_{ii}>0\in\mathbb{R}^{m_i\times m_i},R_{ij}\geqslant0\in\mathbb{R}^{m_j\times m_j}$ are all symmetric constant matrices. $0<\alpha<1$ is the discount factor.

Remark 4. Unlike the performance index involving only the control policy of agent i itself [11,12,14,15,33,34], the performance index (10) proposed in this paper incorporates the policies of its neighboring agents. This aligns with (9) and reflects the characteristics of distributed control systems.

Then, synthesizing the aforementioned analyses, the GFC problem of MAS (1) and (2) to be solved in this paper can be formulated as follows.

Problem 1. Design an effective distributed control policy for each agent i to stabilize error system (9) and minimize performance index (10) simultaneously.

Definition 2 (Admissible control policy [12])**.** The policy $\bar{\psi}_i=\{\psi_{ik}\}_{k=0}^{\infty}$ is admissible if it (1) stabilizes (9) and (2) guarantees that (10) is finite.

When the policy ψ_{ik} satisfies Definition 2, we have

$$J_i(\delta_{ik})=r_i(\delta_{ik},\psi_{ik},\psi_{(j)k})+\alpha J_i(\delta_{i(k+1)}).\tag{11}$$

Then the optimal local performance index $J_i^*(\delta_{ik})$ satisfies the following coupled Hamilton-Jacobi-Bellman (HJB) equation

$$J_i^*(\delta_{ik})=\min_{\psi_{ik}}(r_i(\delta_{ik},\psi_{ik},\psi_{(j)k})+\alpha J_i^*(\delta_{i(k+1)})).\tag{12}$$

By solving the necessary condition $\partial J_i^*(\delta_{ik})/\partial\psi_{ik}=0$, the desired optimal control policy ψ_{ik}^* is derived as

$$\begin{aligned}\psi_{ik}^*=&\arg\min_{\psi_{ik}}(r_i(\delta_{ik},\psi_{ik},\psi_{(j)k})+\alpha J_i^*(\delta_{i(k+1)}))\\=&-\frac{\alpha}{2}(d_i+b_i)R_{ii}^{-1}g_i^{\mathrm{T}}(x_{ik})\frac{\partial J_i^*(\delta_{i(k+1)})}{\partial\delta_{i(k+1)}}.\end{aligned}\tag{13}$$

Remark 5. Note that the desired optimal control policy ψ_{ik}^* can be derived by solving the coupled HJB equation. However, the equation exhibits highly nonlinear coupling and heavily relies on system dynamics, rendering it challenging to solve analytically. This paper thus aims to develop a model-free approach to overcome this challenge.

Note that Eq. (9) is jointly driven by ψ_i and $\psi_{(j)}$. Accordingly, the GFC problem considered in this paper can be regarded as a graphical game problem, focusing on finding the global Nash equilibrium solution.

Definition 3 (Global Nash equilibrium [35])**.** The optimal control policy set $\{\psi_1^*, \ldots, \psi_N^*\}$ is a global Nash equilibrium solution for the N-player graphical game if

$$J_i^* \triangleq J_i(\psi_i^*, \psi_{\mathcal{G}-i}^*) \leqslant J_i(\psi_i, \psi_{\mathcal{G}-i}^*),\ \forall i \in \mathcal{V},$$

where $\psi_{\mathcal{G}-i} = \{\psi_j | j \in \mathcal{V}, j \neq i\}$. The optimal performance index set $\{J_1^*, \ldots, J_N^*\}$ represents a Nash equilibrium.

According to Definition 3, Eq. (12) can be rewritten as

$$J_i^*(\delta_{ik}) = r_i(\delta_{ik}, \psi_{ik}^*, \psi_{(j)k}^*) + \alpha J_i^*(\delta_{i(k+1)}). \tag{14}$$

3 Model-free distributed GFC design

In this section, an effective PI algorithm is proposed to determine the optimal GFC policy, together with a rigorous convergence analysis. On this basis, the stability of the group neighborhood formation error system and the global Nash equilibrium of MASs are guaranteed. Then, the implementation process of the PI algorithm using an actor-critic NN framework in an online model-free manner is presented.

3.1 PI algorithm and its convergence analysis

The proposed PI algorithm incorporates policy evaluation and policy improvement to evaluate the current policy and iteratively update it until the optimal one is obtained. The detailed procedure of this algorithm is outlined in Algorithm 1.

Algorithm 1 PI algorithm for GFC of heterogeneous nonlinear MASs.

1. **Initialization.** Start with arbitrary initial admissible policies ψ_{ik}^0 for $\forall i$ and let $l = 0$.
2. **Policy evaluation.** Solve for $J_i^l(\delta_{ik})$ using

$$J_i^l(\delta_{ik}) = r_i(\delta_{ik}, \psi_{ik}^l, \psi_{(j)k}^l) + \alpha J_i^l(\delta_{i(k+1)}). \tag{15}$$

3. **Policy improvement.** Update ψ_{ik}^{l+1} using

$$\psi_{ik}^{l+1} = \arg\min_{\psi_{ik}}(r_i(\delta_{ik}, \psi_{ik}, \psi_{(j)k}) + \alpha J_i^l(\delta_{i(k+1)})). \tag{16}$$

4. Stop when $|J_i^{l+1}(\delta_{ik}) - J_i^l(\delta_{ik})| \leqslant \varepsilon$ with a prescribed constant ε.

Theorem 1. Let all agents in MAS (1) implement Algorithm 1 synchronously. Then, it can be obtained that (i) $J_i^l(\delta_{ik})$ is monotonically nonincreasing, i.e., $J_i^{l+1}(\delta_{ik}) \leqslant J_i^l(\delta_{ik})$; (ii) as $l \to \infty$, $J_i^l(\delta_{ik})$ and ψ_{ik}^l converge to $J_i^*(\delta_{ik})$ and ψ_{ik}^*, respectively, i.e., $\lim_{l\to\infty} J_i^l(\delta_{ik}) = J_i^*(\delta_{ik})$ and $\lim_{l\to\infty} \psi_{ik}^l = \psi_{ik}^*$.

Proof. (i) Define a new performance index $\Phi_i^{l+1}(\delta_{ik})$ as

$$\begin{aligned}\Phi_i^{l+1}(\delta_{ik}) &\triangleq r_i(\delta_{ik}, \psi_{ik}^{l+1}, \psi_{(j)k}^{l+1}) + \alpha J_i^l(\delta_{i(k+1)}) \\ &= \min_{\psi_{ik}}(r_i(\delta_{ik}, \psi_{ik}, \psi_{(j)k}) + \alpha J_i^l(\delta_{i(k+1)})).\end{aligned} \tag{17}$$

Based on (15)–(17), we have

$$\Phi_i^{l+1}(\delta_{ik}) \leqslant J_i^l(\delta_{ik}). \tag{18}$$

Note that the policy ψ_{ik}^l is always admissible during the iteration process [35]. Thus, $\|\delta_{ik}\| \to 0$ as $k \to \infty$. Let $k = q$, $q \to \infty$. Then, we obtain

$$J_i^{l+1}(\delta_{iq}) = \Phi_i^{l+1}(\delta_{iq}) = J_i^l(\delta_{iq}). \tag{19}$$

Let $k = q-1$. The combination of (15), (16), and (19) yields

$$\begin{aligned} J_i^{l+1}(\delta_{i(q-1)}) &= r_i(\delta_{i(q-1)}, \psi_{i(q-1)}^{l+1}, \psi_{(j)(q-1)}^{l+1}) + \alpha J_i^{l+1}(\delta_{iq}) \\ &= r_i(\delta_{i(q-1)}, \psi_{i(q-1)}^{l+1}, \psi_{(j)(q-1)}^{l+1}) + \alpha J_i^{l}(\delta_{iq}) \\ &= \min_{\psi_{i(q-1)}} (r_i(\delta_{i(q-1)}, \psi_{i(q-1)}, \psi_{(j)(q-1)}) + \alpha J_i^l(\delta_{iq})) \\ &\leqslant r_i(\delta_{i(q-1)}, \psi_{i(q-1)}^{l}, \psi_{(j)(q-1)}^{l}) + \alpha J_i^l(\delta_{iq}) \\ &= J_i^l(\delta_{i(q-1)}). \end{aligned} \tag{20}$$

From (20), we conclude that $J_i^{l+1}(\delta_{ik}) \leqslant J_i^l(\delta_{ik})$ holds for $k = q-1$. Suppose the conclusion holds for $k = T+1$, $T = 0, 1, 2, \ldots$, i.e., $J_i^{l+1}(\delta_{i(T+1)}) \leqslant J_i^l(\delta_{i(T+1)})$. Letting $k = T$, we then obtain

$$\begin{aligned} J_i^{l+1}(\delta_{iT}) =& r_i(\delta_{iT}, \psi_{iT}^{l+1}, \psi_{(j)T}^{l+1}) + \alpha J_i^{l+1}(\delta_{i(T+1)}) \\ \leqslant& r_i(\delta_{iT}, \psi_{iT}^{l+1}, \psi_{(j)T}^{l+1}) + \alpha J_i^{l}(\delta_{i(T+1)}) \\ =& \Phi_i^{l+1}(\delta_{iT}) \\ \leqslant& J_i^l(\delta_{iT}). \end{aligned} \tag{21}$$

Thus, by mathematical induction, Eqs. (19)–(21) together entail the conclusion that for $\forall i$ and $\forall l$, $J_i^{l+1}(\delta_{ik}) \leqslant J_i^l(\delta_{ik})$ holds for $\forall k$.

(ii) Define $J_i^\infty(\delta_{ik}) \triangleq \lim_{l\to\infty} J_i^l(\delta_{ik})$ for simplicity. From (1), we have

$$\begin{aligned} J_i^\infty(\delta_{ik}) &\leqslant \Phi_i^{l+1}(\delta_{ik}) \\ &= \min_{\psi_{ik}}(r_i(\delta_{ik}, \psi_{ik}, \psi_{(j)k}) + \alpha J_i^l(\delta_{i(k+1)})). \end{aligned} \tag{22}$$

When $l \to \infty$, we can get

$$J_i^\infty(\delta_{ik}) \leqslant \min_{\psi_{ik}}(r_i(\delta_{ik}, \psi_{ik}, \psi_{(j)k}) + \alpha J_i^\infty(\delta_{i(k+1)})). \tag{23}$$

According to the monotonically nonincreasing property of $J_i^l(\delta_{ik})$ in (1), there exists an index H such that

$$J_i^H(\delta_{ik}) - \gamma \leqslant J_i^\infty(\delta_{ik}) \leqslant J_i^H(\delta_{ik}), \tag{24}$$

where γ is an arbitrary positive constant. From (24) we further obtain

$$\begin{aligned} J_i^\infty(\delta_{ik}) &\geqslant J_i^H(\delta_{ik}) - \gamma \\ &= r_i(\delta_{ik}, \psi_{ik}^H, \psi_{(j)k}^H) + \alpha J_i^H(\delta_{i(k+1)}) - \gamma \\ &\geqslant r_i(\delta_{ik}, \psi_{ik}^H, \psi_{(j)k}^H) + \alpha J_i^\infty(\delta_{i(k+1)}) - \gamma \\ &\geqslant \min_{\psi_{ik}}(r_i(\delta_{ik}, \psi_{ik}, \psi_{(j)k}) + \alpha J_i^\infty(\delta_{i(k+1)})) - \gamma. \end{aligned} \tag{25}$$

Because γ is an arbitrary positive constant, we have

$$J_i^\infty(\delta_{ik}) \geqslant \min_{\psi_{ik}}(r_i(\delta_{ik}, \psi_{ik}, \psi_{(j)k}) + \alpha J_i^\infty(\delta_{i(k+1)})). \tag{26}$$

Combining (23) and (26) yields

$$J_i^\infty(\delta_{ik}) = \min_{\psi_{ik}}(r_i(\delta_{ik}, \psi_{ik}, \psi_{(j)k}) + \alpha J_i^\infty(\delta_{i(k+1)})). \tag{27}$$

Given policies μ_{ik} and $\mu_{(j)k}$ that satisfy Definition 2, once again, define a new performance index $\Theta_i(\delta_{ik})$ as

$$\Theta_i(\delta_{ik}) = r_i(\delta_{ik}, \mu_{ik}, \mu_{(j)k}) + \alpha\Theta_i(\delta_{i(k+1)}). \tag{28}$$

Let $k = p$, $p \to \infty$. Then, we have $\|\delta_{ip}\| \to 0$, which entails $J_i^\infty(\delta_{ip}) = \Theta_i(\delta_{ip}) = 0$. Letting $k = p-1$, we get

$$\begin{aligned}\Theta_i(\delta_{i(p-1)}) &= r_i(\delta_{i(p-1)}, \mu_{i(p-1)}, \mu_{(j)(p-1)}) + \alpha\Theta_i(\delta_{ip})\\ &\geqslant \min_{\psi_{i(p-1)}} (r_i(\delta_{i(p-1)}, \psi_{i(p-1)}, \psi_{(j)(p-1)}) + \alpha\Theta_i(\delta_{ip}))\\ &= \min_{\psi_{i(p-1)}} (r_i(\delta_{i(p-1)}, \psi_{i(p-1)}, \psi_{(j)(p-1)}) + \alpha J_i^\infty(\delta_{ip}))\\ &= J_i^\infty(\delta_{i(p-1)}).\end{aligned} \tag{29}$$

Suppose (29) holds for $k = s+1, s = 0, 1, 2, \ldots$, i.e., $\Theta_i(\delta_{i(s+1)}) \geqslant J_i^\infty(\delta_{i(s+1)})$. Next, when $k = s$, we have

$$\begin{aligned}\Theta_i(\delta_{is}) &= r_i(\delta_{is}, \mu_{is}, \mu_{(j)s}) + \alpha\Theta_i(\delta_{i(s+1)})\\ &\geqslant r_i(\delta_{is}, \mu_{is}, \mu_{(j)s}) + \alpha J_i^\infty(\delta_{i(s+1)})\\ &\geqslant \min_{\psi_{is}}(r_i(\delta_{is}, \psi_{is}, \psi_{(j)s}) + \alpha J_i^\infty(\delta_{i(s+1)}))\\ &= J_i^\infty(\delta_{is}).\end{aligned} \tag{30}$$

Thus, by mathematical induction, we deduce that $\Theta_i(\delta_{is}) \geqslant J_i^\infty(\delta_{is})$ holds for $\forall s = 0, 1, 2, \ldots$, from above. Letting $\mu_{ik} = \psi_{ik}^*$ for $\forall i$, we get

$$J_i^\infty(\delta_{ik}) \leqslant \Theta_i(\delta_{ik}) = J_i^*(\delta_{ik}). \tag{31}$$

Note that $J_i^*(\delta_{ik})$ represents the optimal performance index. Then, we can also attain

$$J_i^\infty(\delta_{ik}) \geqslant J_i^*(\delta_{ik}). \tag{32}$$

The combination of (31) and (32) yields

$$J_i^\infty(\delta_{ik}) \triangleq \lim_{l\to\infty} J_i^l(\delta_{ik}) = J_i^*(\delta_{ik}). \tag{33}$$

According to (33) and (13), we can obtain $\lim_{l\to\infty} \psi_{ik}^l = \psi_{ik}^*$ directly. The proof is completed.

3.2 System stability analysis and global Nash equilibrium seeking

The following theorem illustrates the results for achieving GFC and global Nash equilibrium.

Theorem 2. Suppose that Assumptions 1 and 3 hold. For each agent i, if $J_i^*(\delta_{ik})$ and ψ_{ik}^* satisfy (14) and (13), respectively, then (i) the error system (9) is asymptotically stable and hence MAS (1) and (2) achieves GFC; (ii) all agents achieve the global Nash equilibrium.

Proof. (i) From (14), we have

$$J_i^*(\delta_{ik}) - \alpha J_i^*(\delta_{i(k+1)}) = r_i(\delta_{ik}, \psi_{ik}^*, \psi_{(j)k}^*). \tag{34}$$

By multiplying both sides of (34) by α^k, we obtain

$$\alpha^k J_i^*(\delta_{ik}) - \alpha^{k+1} J_i^*(\delta_{i(k+1)}) = \alpha^k r_i(\delta_{ik}, \psi_{ik}^*, \psi_{(j)k}^*). \tag{35}$$

Define the difference of the Lyapunov function candidate as

$$\Delta(\alpha^k J_i^*(\delta_{ik})) = \alpha^{k+1} J_i^*(\delta_{i(k+1)}) - \alpha^k J_i^*(\delta_{ik}). \tag{36}$$

According to (35), Eq. (36) can be rewritten as

$$\Delta(\alpha^k J_i^*(\delta_{ik})) = -\alpha^k r_i(\delta_{ik}, \psi_{ik}^*, \psi_{(j)k}^*) < 0. \tag{37}$$

Synthesizing the aforementioned analyses, it can be obtained that the local group neighborhood formation error system (9) is asymptotically stable for each agent i, i.e., $\|\delta_{ik}\| \to 0$ as $k \to \infty$. According to Definition 1 and Lemma 1, it follows that MAS (1) and (2) achieves GFC.

(ii) Because $J_i^*(\delta_{ik}, \psi_{ik}^*, \psi_{(j)k}^*)$ is the optimal value,

$$J_i^*(\delta_{ik}, \psi_{ik}^*, \psi_{(j)k}^*) \leqslant J_i(\delta_{ik}, \psi_{ik}, \psi_{(j)k}^*). \tag{38}$$

According to the definition of the local performance index (10), we obtain

$$J_i^*(\delta_{ik}, \psi_{ik}^*, \psi_{(j)k}^*) = J_i(\delta_{ik}, \psi_{ik}^*, \psi_{(j)k}^*, \psi_{mk}), \tag{39}$$

where $m \notin \{i\} \cup N_i$.

Combining the above two equations, we obtain

$$J_i(\delta_{ik}, \psi_{ik}^*, \psi_{(\mathcal{G}-i)k}^*) \leqslant J_i(\delta_{ik}, \psi_{ik}, \psi_{(\mathcal{G}-i)k}^*). \tag{40}$$

From Definition 3, we can conclude that all agents achieve the global Nash equilibrium. The proof is completed.

3.3 Algorithm implementation using actor-critic NN framework

Leveraging the universal approximation property of NN [12], the actor-critic NN framework, which requires that each agent be equipped with two three-layer NNs, is established to implement Algorithm 1 in an online model-free manner.

For each agent i, the local performance index (10) is approximated by the critic NN, expressed by

$$\hat{J}_{ik} = \hat{W}_{ci}^{\mathrm{T}} \phi_{ci}(Y_{ci}^{\mathrm{T}} z_{cik}), \tag{41}$$

where $\hat{W}_{ci}$ is the critic NN weight between the hidden layer and output layer, Y_{ci} is the weight between the input layer and hidden layer, $\phi_{ci}(\cdot)$ denotes the activation function, and z_{cik}, a vector consisting of δ_{ik}, ψ_{ik} and $\psi_{(j)k}$, is the critic NN input.

Using the critic NN approximation (41), the critic NN error can be obtained from (11) as

$$e_{cik} = r_{i(k-1)} + \alpha \hat{J}_{ik} - \hat{J}_{i(k-1)}. \tag{42}$$

For the critic NN, the goal is to minimize the following objective function:

$$E_{cik} = \frac{1}{2} e_{cik}^{\mathrm{T}} e_{cik}.$$

To achieve this goal, the update law of the critic NN weight based on the gradient descent method is given by

$$\begin{aligned} \hat{W}_{ci}^{l+1} &= \hat{W}_{ci}^{l} - \beta_{ci} \frac{\partial E_{cik}}{\partial e_{cik}} \frac{\partial e_{cik}}{\partial \hat{J}_{ik}} \frac{\partial \hat{J}_{ik}}{\partial \hat{W}_{ci}^{l}} \\ &= \hat{W}_{ci}^{l} - \alpha \beta_{ci} \phi_{ci}(Z_{cik}) [r_{i(k-1)} + \alpha \hat{W}_{ci}^{l\mathrm{T}} \phi_{ci}(Z_{cik}) - \hat{W}_{ci}^{l\mathrm{T}} \phi_{ci}(Z_{ci(k-1)})], \end{aligned} \tag{43}$$

where β_{ci} is the learning rate of critic NN and $Z_{cik} = Y_{ci}^{\mathrm{T}} z_{cik}$.

The control policy is approximated by the actor NN, expressed by

$$\hat{\psi}_{ik} = \hat{W}_{ai}^{\mathrm{T}} \phi_{ai}(Y_{ai}^{\mathrm{T}} z_{aik}), \tag{44}$$

where $\hat{W}_{ai}$, Y_{ai}, $\phi_{ai}(\cdot)$, and z_{aik} are similar to those in the critic NN, with z_{aik} representing a vector of information from δ_{ik}.

The role of the actor is to minimize the local performance index. Thus, define the actor NN error as

$$e_{aik} = \hat{J}_{ik} - U_{ik}, \tag{45}$$

where $U_{ik} = 0$ is the desired ultimate cost-to-go objective.

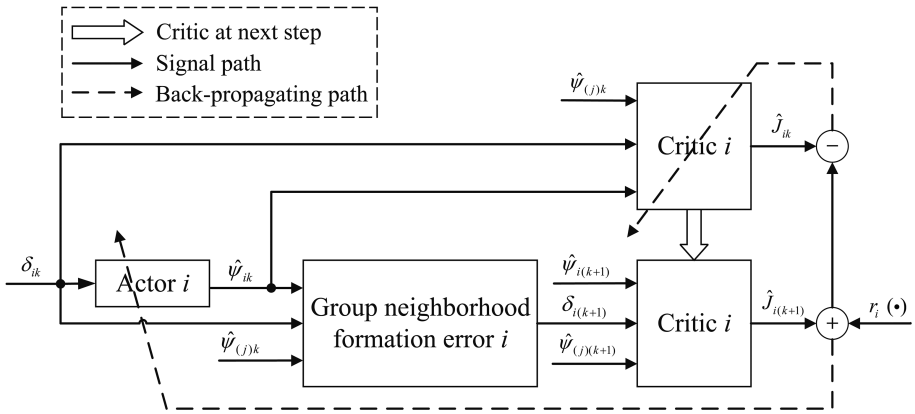

Figure 1 Structure diagram of the developed model-free GFC algorithm based on the actor-critic NN framework.

Then, the objective function for the actor NN is defined as

$$E_{aik} = \frac{1}{2} e_{aik}^{\mathrm{T}} e_{aik}.$$

Similar to (43), the following gives the update law of the actor NN weight:

$$\begin{aligned} \hat{W}_{ai}^{l+1} &= \hat{W}_{ai}^{l} - \frac{\partial E_{aik}}{\partial e_{aik}} \frac{\partial e_{aik}}{\partial \hat{J}_{ik}} \frac{\partial \hat{J}_{ik}}{\partial \phi_{ci}(Z_{cik})} \frac{\partial \phi_{ci}(Z_{cik})}{\partial Z_{cik}} \frac{\partial Z_{cik}}{\partial z_{cik}} \frac{\partial z_{cik}}{\partial \hat{\psi}_{ik}} \frac{\partial \hat{\psi}_{ik}}{\partial \hat{W}_{ai}^{l}} \\ &= \hat{W}_{ai}^{l} - \beta_{ai} \phi_{ai}(Z_{aik}) \hat{W}_{ci}^{l\mathrm{T}} \phi_{ci}'(Z_{cik}) Y_{ci}^{\mathrm{T}} C_i \hat{W}_{ci}^{l\mathrm{T}} \phi_{ci}(Z_{cik}), \end{aligned} \tag{46}$$

where β_{ai} is the learning rate of actor NN, $Z_{aik} = Y_{ai}^{\mathrm{T}} z_{aik}$, $\phi_{ci}'(Z_{cik}) = \partial \phi_{ci}(Z_{cik})/\partial Z_{cik}$, and $C_i = \partial z_{cik}/\partial \hat{\psi}_{ik}$.

Given the preceding analyses, the developed model-free GFC algorithm using the actor-critic NN framework is presented in Algorithm 2, and the corresponding algorithm structure diagram is shown in Figure 1.

Algorithm 2 Model-free GFC algorithm using actor-critic NN framework.

Initialization: (For $\forall i$ and $\forall \bar{i}$)
1: Initialize x_{i0} for agent i and $x_{00}^{\bar{i}}$ for leader $\bar{i}$;
2: Compute the error $\delta_{i0} \leftarrow$ (5);
3: Initialize $\hat{W}_{ci}^{0}$ and $\hat{W}_{ai}^{0}$;
4: Set symmetric constant matrices Q_{ii}, R_{ii} and R_{ij};
5: Select learning rates β_{ci} and β_{ai};
6: Prescribe computation precision ε;
Iteration:
Let $l = 0, k = 0$;
Repeat:
1: Compute the control policy $\hat{\psi}_{ik} \leftarrow$ (44);
2: Compute the local performance index $\hat{J}_{ik} \leftarrow$ (41);
3: Compute the error $\delta_{i(k+1)} \leftarrow$ (5);
4: Compute the control policy $\hat{\psi}_{i(k+1)} \leftarrow$ (44);
5: Compute the local performance index $\hat{J}_{i(k+1)} \leftarrow$ (41);
6: Update the critic NN weight $\hat{W}_{ci}^{l+1} \leftarrow$ (43);
7: Update the actor NN weight $\hat{W}_{ai}^{l+1} \leftarrow$ (46);
8: Let $l = l + 1, k = k + 1$;
Until $\sum_{i=1}^{N} \|\hat{W}_{ci}^{l+1} - \hat{W}_{ci}^{l}\|/N \leqslant \varepsilon$;
Return $\hat{W}_{ci}$ and $\hat{W}_{ai}$ for $\forall i$. End.

Remark 6. Notably, Algorithm 2 only requires local neighboring relative interaction data for each agent i. Moreover, during its implementation, by setting l and k as the same step index, it can update each iteration step l at each real-time step k. The above two aspects indicate that Algorithm 2 is an online model-free GFC algorithm, making it applicable to real-time MASs with unknown dynamics and thereby demonstrating its superiority over traditional offline model-based methods [25–32] in practical applications.

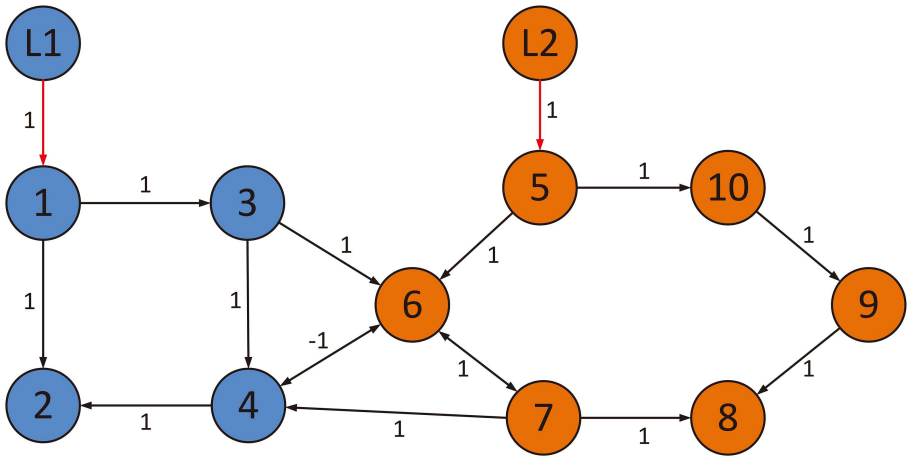

Figure 2 (Color online) Communication topology for the GFC problem.

4 Simulation examples

Two simulation results are provided to demonstrate the effectiveness and superiority of the proposed model-free algorithm for the GFC problem.

4.1 GFC of numerical heterogeneous nonlinear MAS

Consider a MAS with agents 1–4 and 5–10 belonging to Ω_1 and Ω_2, respectively. Moreover, there exists one leader for each subgroup, labeled as $L1$ and $L2$, respectively. Let the pinning gains $b_1 = b_5 = 1$, the adjacency weighting elements $a_{21} = a_{24} = a_{31} = a_{43} = a_{47} = a_{63} = a_{65} = a_{67} = a_{76} = a_{87} = a_{89} = a_{9,10} = a_{10,5} = 1$, $a_{46} = a_{64} = -1$, and other pinning gains and adjacency weighting elements all be 0. Figure 2 shows the communication topology for this MAS, which satisfies Assumptions 1 and 3.

The dynamics of all agents are given as

$$x_i(k+1) = f_i(x_i(k)) + g_i(x_i(k))\psi_i(k),\ \ i = 1, \ldots, 10,$$

where $f_i(x_i(k)) = [e_i x_{i1}^2(k)/(1 + x_{i1}^2(k)) + 0.3x_{i2}(k), x_{i1}(k)/(1 + x_{i1}^2(k) + x_{i2}^2(k))]^{\mathrm{T}}$, $g_i(x_i(k)) = [0, h_i]^{\mathrm{T}}$, $e_1 = 1.5, e_2 = 0.4, e_3 = 0.8, e_4 = 0.5, e_5 = 1, e_6 = 0.6, e_7 = 1.4, e_8 = 1.2, e_9 = 1.5, e_{10} = 0.7, h_1 = 0.1, h_2 = 0.2, h_3 = -0.15, h_4 = 0.35, h_5 = 0.3, h_6 = 0.05, h_7 = 0.15, h_8 = 0.1, h_9 = 0.25$, and $h_{10} = -0.2$. The leader dynamics are $x_0^1(k+1) = [0.2k + 0.2, 0.2k]^{\mathrm{T}}, x_0^2(k+1) = [0.3k, 0.15k + 0.15]^{\mathrm{T}}$, and $\eta_{i=1,\ldots,10} = [-1.5; 1.5], [-1.5; -1.5], [1.5; 1.5], [1.5; -1.5], [-1; \sqrt{3}], [-2; 0], [-1; -\sqrt{3}], [1; -\sqrt{3}], [2; 0], [1; \sqrt{3}]$.

Choose $Q_{ii} = I_2$, $R_{ii} = 1$ for $i = 1, \ldots, 10$, $R_{21} = R_{24} = R_{31} = R_{43} = R_{46} = R_{47} = R_{63} = R_{64} = R_{65} = R_{67} = R_{76} = R_{87} = R_{89} = R_{9,10} = R_{10,5} = 0.8$, and all others are 0. Set the discount factor $\alpha = 0.95$ and select the initial states of all agents as random values in $[-1, 3]$. Given the analyses in Section 3, we choose the critic NN activation function in the form of a quadratic vector consisting of δ_{ik}, ψ_{ik}, and $\psi_{(j)k}$ and the actor NN activation function as δ_{ik} for each agent i. Weights $\hat{W}_{ci}^0$ and $\hat{W}_{ai}^0$ are initialized in $[0, 1]$ appropriately, and Y_{ci} and Y_{ai} are set as identity matrices with appropriate dimensions. The learning rates for NN weight updating are selected as $\beta_{ci} = \beta_{ai} = 0.01, i = 1, \ldots, 10$, and the computation precision is selected as $\varepsilon = 1 \times 10^{-6}$.

By executing Algorithm 2 for all agents with the above parameter settings, we obtain Figures 3–5. Figures 3(a) and (b) show the evolutions of critic NN weights for agents in Ω_1 and Ω_2, respectively. Figure 4(a) presents the initial states of all agents, and Figure 4(b) illustrates the phase portrait of state trajectories for all agents. Figure 4(b) shows that the GFC is achieved after some iterations. Figures 5(a) and (b), respectively, present the evolutions of local group neighborhood formation errors and approximate control policies for all agents, thereby further validating that the proposed model-free algorithm can achieve GFC.

To demonstrate the superior optimization of control performance achieved by the developed algorithm, a comparison with the offline model-based control method presented in [29] is conducted. The simulation results are shown in Figure 6. Specifically, Figures 6(a) and (b) depict the evolutions of the sum of utility functions r_i for all agents under the two control methods, respectively. The developed control

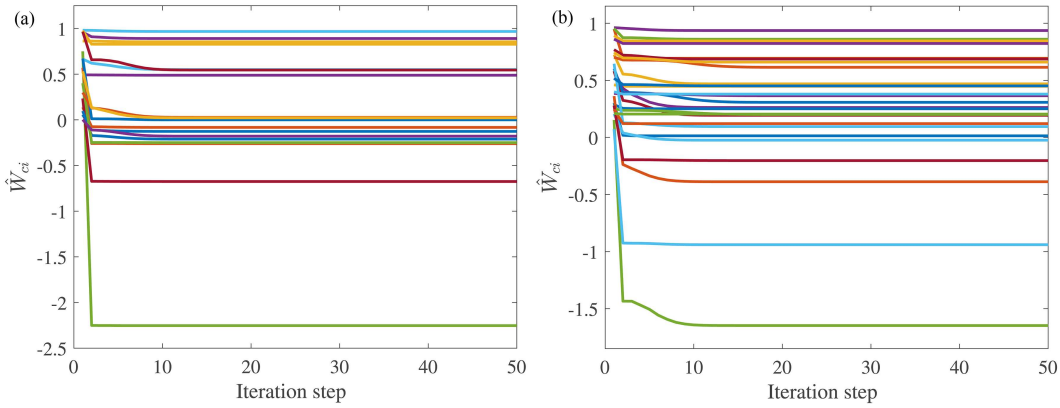

Figure 3 (Color online) Evolution of the critic NN weights for agents in (a) Ω_1 and (b) Ω_2.

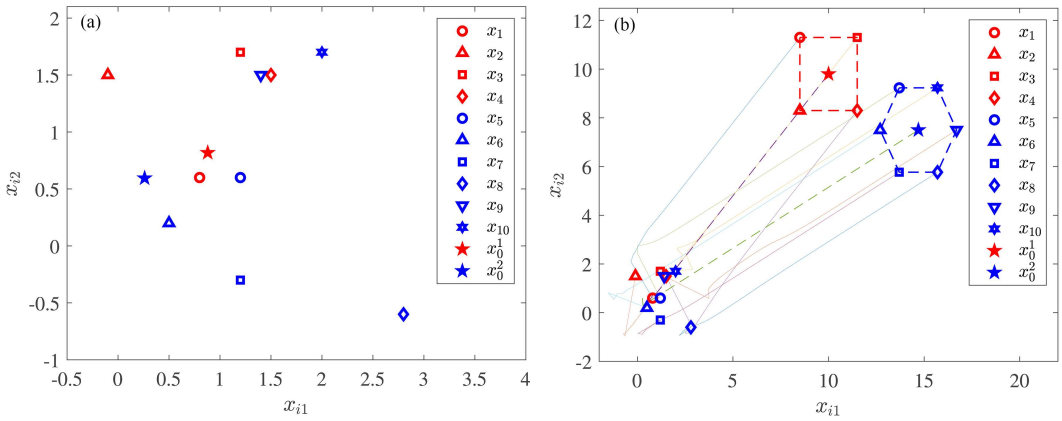

Figure 4 (Color online) (a) Initial states of all agents. (b) Phase portrait of state trajectories for all agents, $l = 1, \ldots, 50$.

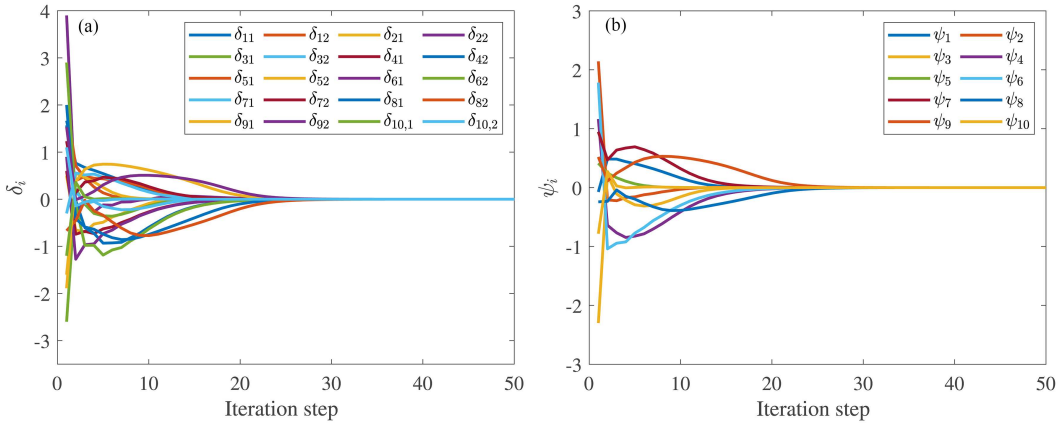

Figure 5 (Color online) Evolution of (a) local group neighborhood formation errors δ_i and (b) approximate control policies ψ_i for all agents.

algorithm evidently achieves GFC at a lower control cost. In fact, the controller gain selection in [29] is not unique, potentially affecting GFC performance. In contrast, this study employs an online learning approach to determine the optimal control policy by utilizing system operation data, which simplifies the determination process and enhances control performance.

4.2 GFC of multiple single-link robot arms (SRAs)

Consider the system consisting of multiple SRAs with dynamics [36]

$$x_i(k+1) = f_i(x_i(k)) + g_i(x_i(k))\psi_i(k),\ i = 1, \ldots, 34,$$

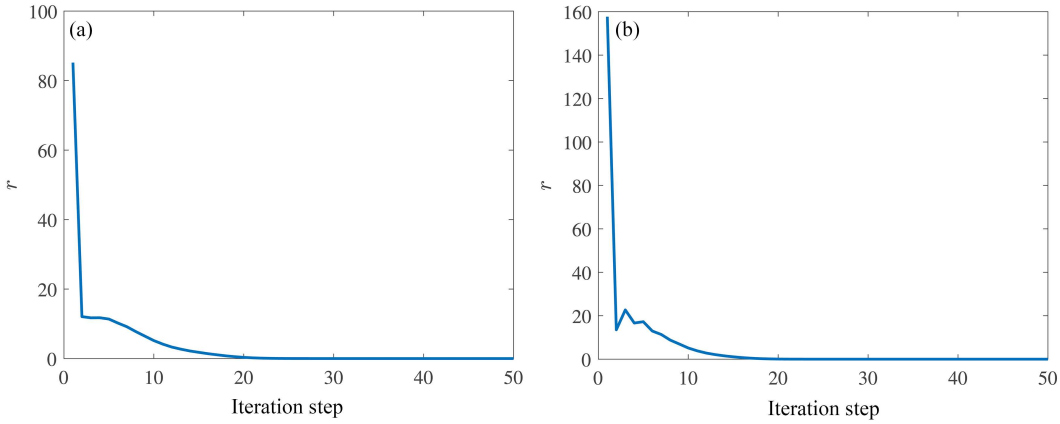

Figure 6 (Color online) Evolution of the sum of utility functions r_i for all agents under (a) the developed algorithm and (b) the method in [29].

Table 2 Parameters of SRAs.

Parameter	$i = 1$–10	$i = 11$–20	$i = 21$–28	$i = 29$–34
J_i	10.5	15.6	18.27	8.6
B_i	7.6	9.65	13.2	3.5
$M_i g l_i$	9.8	12.94	16.66	4.1

where $f_i(x_i(k)) = [x_{i2}(k), J_i^{-1}(-B_i x_{i2}(k) - M_i g l_i \sin(x_{i1}(k)))]^{\mathrm{T}}$, $g_i(x_i(k)) = [0, J_i^{-1}]^{\mathrm{T}}$, the states $x_{i1}(k)$ and $x_{i2}(k)$ are angle and angular velocity of the link, respectively, J_i is the total rotational inertia of the link and the motor, B_i is the overall damping coefficient, M_i is the total mass of the link, l_i is the distance from the joint axis to the link center of mass for SRA i, and g is the gravitational acceleration. The parameters of the SRAs are presented in Table 2.

In this case, SRAs 1–10, 11–20, 21–28, and 29–34 belong to Ω_1, Ω_2, Ω_3, and Ω_4, respectively. The communication interactions among SRAs are described by the following graph parameters: the pinning gains are $b_1 = b_{20} = b_{26} = b_{29} = 1$, the adjacency matrix elements are $a_{12} = a_{21} = a_{32} = a_{43} = a_{54} = a_{67} = a_{78} = a_{89} = a_{8,12} = a_{9,10} = a_{10,5} = a_{11,16} = a_{12,11} = a_{13,10} = a_{13,12} = a_{14,13} = a_{15,14} = a_{16,17} = a_{17,18} = a_{18,19} = a_{19,20} = a_{21,25} = a_{22,21} = a_{23,18} = a_{23,22} = a_{24,28} = a_{25,26} = a_{27,26} = a_{28,27} = a_{29,26} = a_{30,29} = a_{31,30} = a_{32,30} = a_{33,29} = a_{34,33} = 1, a_{8,13} = a_{13,6} = a_{23,19} = a_{29,25} = -1$, and others are all 0. The leader dynamics are $x_0^1(k+1) = [-3, 0]^{\mathrm{T}}, x_0^2(k+1) = [-1, -1]^{\mathrm{T}}, x_0^3(k+1) = [1, 0]^{\mathrm{T}}, x_0^4(k+1) = [3, 0]^{\mathrm{T}}$, and $\eta_{i=1,\ldots,34}$ are chosen appropriately to ensure that the SRAs in the four subgroups form "H", "U", "S", and "T" formations, respectively.

For the simulation, choose $Q_{ii} = I_2, R_{ii} = 1$ for $i = 1, \ldots, 34$, and set $R_{ij} = 0$ or 0.8 for $i \neq j, i, j = 1, \ldots, 34$, based on whether the corresponding a_{ij} are 0. Select the initial states of all SRAs as random values in $[0, 1]$, with the other parameter settings being similar to those in Subsection 4.1. Then, executing Algorithm 2 for all SRAs, we obtain Figures 7 and 8. Figure 7 shows the states of all SRAs at different iteration steps. After some iterations, the SRAs in the four subgroups form "H", "U", "S", and "T" formations, respectively, thereby successfully achieving the GFC. Figures 8(a) and (b) illustrate the evolutions of local group neighborhood formation errors and approximate control policies for all SRAs, respectively, further validating that the proposed algorithm can achieve GFC.

5 Conclusion

This paper explored the distributed GFC problem of heterogeneous nonlinear MASs with unknown dynamics, representing a further extension of the global formation control problem. A more flexible communication topology was designed to achieve communication configuration among agents. By developing an effective PI algorithm and establishing a corresponding actor-critic NN framework, an innovative model-free GFC algorithm was proposed to determine the optimal control policy online using only system operation data. This algorithm was verified through two simulations—a numerical MAS and a multiple SRAs system. The findings of this study are also applicable to linear MASs, thereby demonstrating a broader application scope. Future work will concentrate on developing simplified model-free GFC

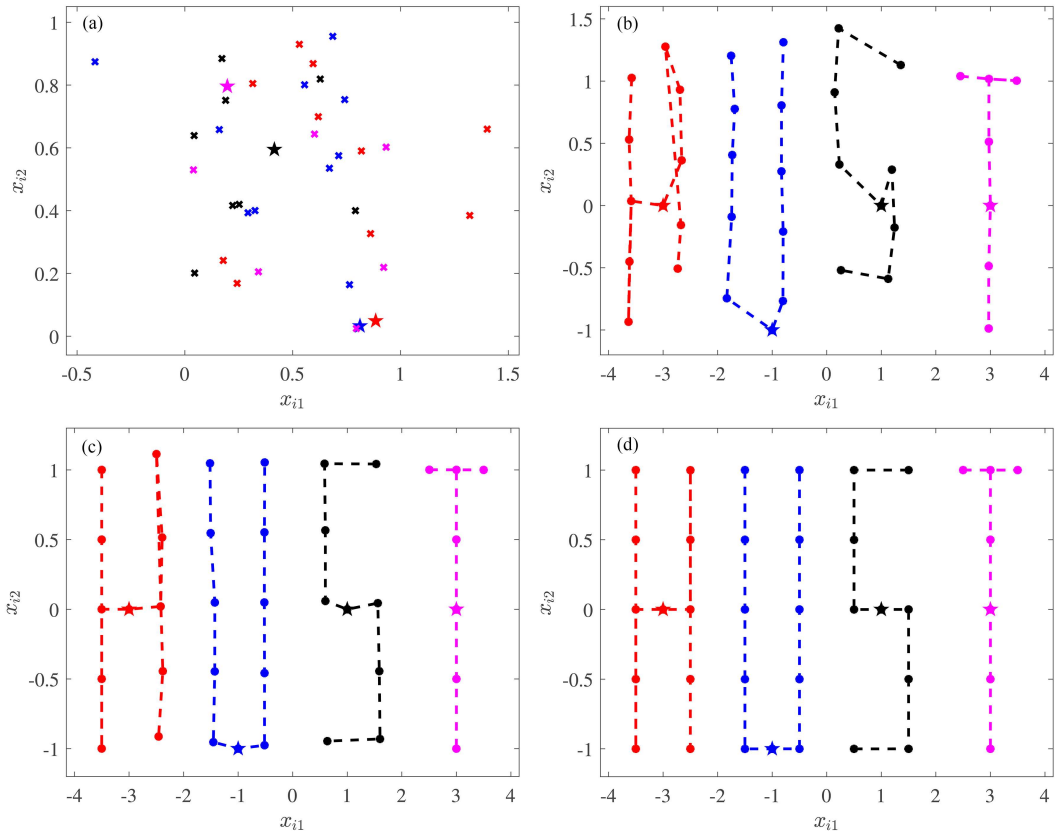

Figure 7 (Color online) States of all SRAs at different iteration steps. (a) $l = 1$; (b) $l = 10$; (c) $l = 30$; (d) $l = 150$.

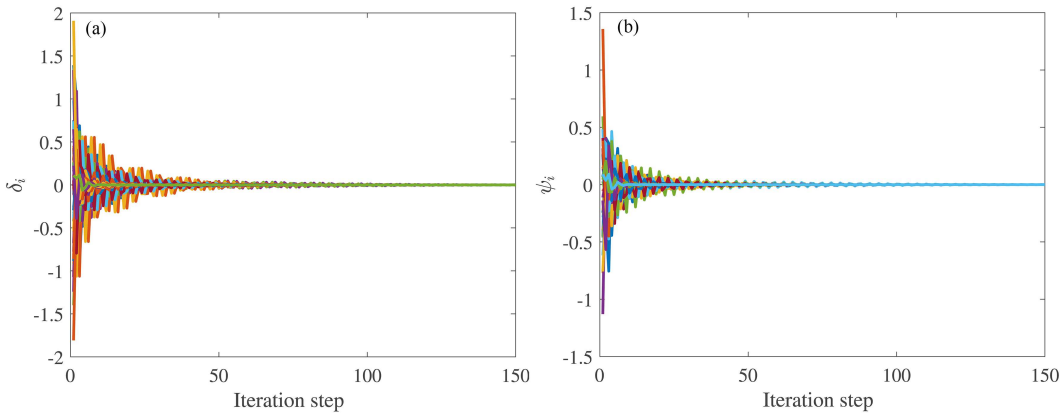

Figure 8 (Color online) Evolution of (a) the local group neighborhood formation errors δ_i and (b) the approximate optimal control policies ψ_i for all SRAs.

algorithms to alleviate computational burden and performing experiments under real-world scenarios.

Acknowledgements This work was supported in part by National Key Research and Development Program of China (Grant No. 2023YFC3902803), National Natural Science Foundation of China (Grant No. 62236005), Foundation for Outstanding Research Groups of Hubei Province of China (Grant No. 2025AFA012), 111 Project on Computational Intelligence and Intelligent Control (Grant No. B18024), and Interdisciplinary Research Program of HUST (Grant No. 2024JCYJ006).

References

1 Nigam N, Bieniawski S, Kroo I, et al. Control of multiple UAVs for persistent surveillance: algorithm and flight test results. IEEE Trans Contr Syst Technol, 2012, 20: 1236–1251
2 Brinon-Arranz L, Schenato L, Seuret A. Distributed source seeking via a circular formation of agents under communication constraints. IEEE Trans Control Netw Syst, 2016, 3: 104–115
3 Zheng R, Liu Y, Sun D. Enclosing a target by nonholonomic mobile robots with bearing-only measurements. Automatica, 2015, 53: 400–407
4 Xiao F, Wang L, Chen J, et al. Finite-time formation control for multi-agent systems. Automatica, 2009, 45: 2605–2611

5 Dong X, Zhou Y, Ren Z, et al. Time-varying formation tracking for second-order multi-agent systems subjected to switching topologies with application to quadrotor formation flying. IEEE Trans Ind Electron, 2017, 64: 5014–5024
6 Dong X, Hu G. Time-varying formation tracking for linear multiagent systems with multiple leaders. IEEE Trans Automat Contr, 2017, 62: 3658–3664
7 Jia R R, Zong X F, Wang Q. Time-varying formation tracking control of high-order multi-agent systems with multiple leaders and multiplicative noise. Sci China Inf Sci, 2024, 67: 222203
8 Li S, Zhang J, Li X, et al. Formation control of heterogeneous discrete-time nonlinear multi-agent systems with uncertainties. IEEE Trans Ind Electron, 2017, 64: 4730–4740
9 Li D, Ge S S, He W, et al. Multilayer formation control of multi-agent systems. Automatica, 2019, 109: 108558
10 Liu Y, Jia Y. An iterative learning approach to formation control of multi-agent systems. Syst Control Lett, 2012, 61: 148–154
11 Wen G, Chen C L P, Feng J, et al. Optimized multi-agent formation control based on an identifier-actor-critic reinforcement learning algorithm. IEEE Trans Fuzzy Syst, 2018, 26: 2719–2731
12 Wen G, Chen C L P, Li B. Optimized formation control using simplified reinforcement learning for a class of multiagent systems with unknown dynamics. IEEE Trans Ind Electron, 2020, 67: 7879–7888
13 Li Y, Zhang J, Tong S. Fuzzy adaptive optimized leader-following formation control for second-order stochastic multiagent systems. IEEE Trans Ind Inf, 2022, 18: 6026–6037
14 Liu D, Liu H, Lü J, et al. Time-varying formation of heterogeneous multiagent systems via reinforcement learning subject to switching topologies. IEEE Trans Circ Syst I, 2023, 70: 2550–2560
15 Zhang J, Fu Y, Fu J. Optimal formation control of second-order heterogeneous multiagent systems using adaptive predefined-time strategy. IEEE Trans Fuzzy Syst, 2024, 32: 2390–2402
16 Li Y, Han Y, Chen C, et al. Online reinforcement learning control designs with acceleration mechanism for unknown multiagent systems through value iteration. IEEE Trans Neural Netw Learn Syst, 2025, 36: 16990–17003
17 Lan J, Liu Y J, Yu D, et al. Time-varying optimal formation control for second-order multiagent systems based on neural network observer and reinforcement learning. IEEE Trans Neural Netw Learn Syst, 2024, 35: 3144–3155
18 Li C J, Zong X F. Group consensus of multi-agent systems with additive noises. Sci China Inf Sci, 2022, 65: 202205
19 Li C, Zong X. Group hybrid coordination control of multi-agent systems with time-delays and additive noises. IEEE CAA J Autom Sin, 2023, 10: 737–748
20 Oyedeji M O, Mahmoud M S. Couple-group consensus conditions for general first-order multiagent systems with communication delays. Syst Control Lett, 2018, 117: 37–44
21 Dong X, Li Q, Zhao Q, et al. Time-varying group formation analysis and design for second-order multi-agent systems with directed topologies. Neurocomputing, 2016, 205: 367–374
22 Hu C, Hua Y, Dong X, et al. Time-varying group formation tracking for multiagent systems with competition and cooperation via distributed Nash equilibrium seeking. IEEE Trans Ind Inf, 2024, 20: 10054–10064
23 Dong X, Li Q, Zhao Q, et al. Time-varying group formation analysis and design for general linear multi-agent systems with directed topologies. Intl J Robust Nonlinear, 2017, 27: 1640–1652
24 Ge X, Han Q L, Zhang X M. Achieving cluster formation of multi-agent systems under aperiodic sampling and communication delays. IEEE Trans Ind Electron, 2018, 65: 3417–3426
25 Hu J, Bhowmick P, Lanzon A. Distributed adaptive time-varying group formation tracking for multiagent systems with multiple leaders on directed graphs. IEEE Trans Control Netw Syst, 2020, 7: 140–150
26 Tian L, Hua Y, Dong X, et al. Distributed time-varying group formation tracking for multiagent systems with switching interaction topologies via adaptive control protocols. IEEE Trans Ind Inf, 2022, 18: 8422–8433
27 Qi J, Zhang Z, Yu J, et al. Time-varying group formation tracking control for multi-agent systems using distributed multi-sensor multi-target filtering with intermittent observations. Intl J Robust Nonlinear, 2024, 34: 11681–11704
28 Li W, Zhou S, Shi M, et al. Collision avoidance time-varying group formation tracking control for multi-agent systems. Appl Intell, 2025, 55: 175
29 Han T, Guan Z H, Chi M, et al. Multi-formation control of nonlinear leader-following multi-agent systems. ISA Trans, 2017, 69: 140–147
30 Wang W, Huang C, Cao J, et al. Event-triggered control for sampled-data cluster formation of multi-agent systems. Neurocomputing, 2017, 267: 25–35
31 Wu Z, Liu X, Sun J, et al. Multi-group formation tracking control via impulsive strategy. Neurocomputing, 2020, 411: 487–497
32 Du X, Li W, Xiao J, et al. Time-varying group formation with leader-following control for heterogeneous multi-agent systems. Int J Control Autom Syst, 2023, 21: 2087–2098
33 Wang Y, Wang Z, Zhang H, et al. Group formation tracking of heterogeneous multiagent systems using reinforcement learning. IEEE Trans Control Netw Syst, 2025, 12: 497–509
34 Shi Y, Hua Y, Yu J, et al. Robust output group formation tracking control of heterogeneous multi-agent systems with multiple leaders using reinforcement learning. Syst Control Lett, 2024, 192: 105897
35 Zhang H, Jiang H, Luo Y, et al. Data-driven optimal consensus control for discrete-time multi-agent systems with unknown dynamics using reinforcement learning method. IEEE Trans Ind Electron, 2017, 64: 4091–4100
36 Zhang H, Lewis F L, Qu Z. Lyapunov, adaptive, and optimal design techniques for cooperative systems on directed communication graphs. IEEE Trans Ind Electron, 2012, 59: 3026–3041