SCIENCE CHINA
Information Sciences

• RESEARCH PAPER •

February 2025, Vol. 68, Iss. 2, 122202:1–122202:16
https://doi.org/10.1007/s11432-024-4103-1

Cooperative output regulation of heterogeneous directed multi-agent systems: a fully distributed model-free reinforcement learning framework

Xiongtao SHI[1], Yanjie LI[1*], Chenglong DU[2*], Huiping LI[3], Chaoyang CHEN[4] & Weihua GUI[2]

[1]*Guangdong Key Laboratory of Intelligent Morphing Mechanisms and Adaptive Robotics, School of Mechanical Engineering and Automation, Harbin Institute of Technology (Shenzhen), Shenzhen 518055, China*
[2]*School of Automation, Central South University, Changsha 410083, China*
[3]*School of Marine Science and Technology, Northwestern Polytechnical University, Xi'an 710072, China*
[4]*School of Information and Electrical Engineering, Hunan University of Science and Technology, Xiangtan 411201, China*

Received 30 January 2024/Revised 9 May 2024/Accepted 19 July 2024/Published online 16 January 2025

Abstract In this paper, the cooperative output regulation (COR) problem of a class of unknown heterogeneous multi-agent systems (MASs) with directed graphs is studied via a model-free reinforcement learning (RL) based fully distributed event-triggered control (ETC) strategy. First, we consider the scenario that the exosystem is accessible globally to all agents, an internal model-based augmented algebraic Riccati equation (AARE) is constructed, and its solution is learned by the proposed model-free RL algorithm via online input-output data. Further, for the scenario that the exosystem is accessible only to its adjacent followers, the distributed observers are designed for each agent to get the state of the exosystem, and an internal model-based fully distributed adaptive ETC protocol is then synthesized to construct the corresponding AARE, and the feedback gain matrix is learned in a model-free fashion. The model-free RL-based control protocol proposed in this paper can not only remove the prior knowledge of agents' dynamics, but also release the dependence on global information by the adaptive event-triggered mechanism (ETM) and the new graph-based Lyapunov function. Finally, simulation results are illustrated to show the feasibility and effectiveness of the proposed control scheme.

Keywords model-free reinforcement learning, unknown heterogeneous multi-agent systems, fully distributed, event-triggered control, directed graph

Citation Shi X T, Li Y J, Du C L, et al. Cooperative output regulation of heterogeneous directed multi-agent systems: a fully distributed model-free reinforcement learning framework. Sci China Inf Sci, 2025, 68(2): 122202, https://doi.org/10.1007/s11432-024-4103-1

1 Introduction

Recent years have witnessed significant advances in the studies of large-scale, distributed, and heterogeneous systems [1–3]. With advantages such as fast computing speed, high reliability, and portability, distributed control and optimization are widely used in smart transportation [4], smart grids [5], distributed energy systems [6], and other fields [7]. The objective of distributed control is to complete a specific task via the cooperation of a group of agents by designing an appropriate protocol [8, 9]. As one of the fundamental research topics on distributed control, the cooperative output regulation (COR) of heterogeneous multi-agent systems (MASs) aims to drive all agents to track a reference signal and make the tracking errors among agents converge to zero. Considerable critical results are reported, such as resilient COR [10], finite-time COR [11], and optimal COR [12].

Noting that most of the existing results on COR are developed based on the agents' dynamics, which are significant for solving the algebraic Riccati equation (ARE) and output regulator equations. However, the accurate mathematical model is hard to be established or requires high measurement cost due to the complex environment and highly coupled nonlinearities. To lift the restriction of model information of MASs, some model-free algorithms have been employed, and the control protocol could be synthesized in a model-free manner [13–23]. For example, for the uncertain single-input single-output systems, a

* Corresponding author (email: autolyj@hit.edu.cn, chenglong_du@csu.edu.cn)

novel indirect adaptive method that does not require the knowledge of the parameters of the systems is developed in [13] such that the asymptotic tracking is guaranteed. An innovative internal model-based data-driven reinforcement learning (RL) approach is proposed to address the intricate challenge of cooperative adaptive optimal control within leader-follower MASs with directed topology in [14], which entails the iterative online acquisition of an approximate optimal controller via real-time input-state data, obviating the necessity for any reliance on model-derived information. Moreover, it is assumed in [23] that the system matrix of the leader is unknown to all followers, and the output regulation problem of heterogeneous MASs is resolved. However, in the above results, information exchanges among agents are usually frequent, which will lead to communication resource overload and heavy computational burden. It should be addressed seriously. Thus, in this paper, we will make efforts to deal with the COR problem of heterogeneous MASs by incorporating the event-triggered control (ETC) and model-free RL algorithm, which will not only reduce the computational complexity and save more communication resources, but also implement the COR of the heterogeneous MASs without model information.

In general, to design an appropriate ETC protocol for MASs, the global information, such as the topology of communication network, is necessary in many scenarios [24,25], and the control protocols are usually not fully distributed. More specifically, the event-triggered mechanism (ETM) is designed with the aid of the global Laplacian matrix [24]. Similarly, the distributed data-driven controller proposed in [25] is to avoid the use of model information, however, the selection of a feasible penalty factor depends on the largest element of the global adjacent matrix of MASs. To remove the dependence on global information of MASs, the adaptive ETC methods have been proposed [26–28]. For example, by designing an edge-based adaptive gain and employing the σ-modification technique, the consensus problem is achieved without global information [26]. Moreover, the dynamic event-triggered scheme is further studied in [28], by which a larger inter-event time is provided in a fully distributed fashion via an auxiliary internal dynamic variable.

It is observed that most of the existing results on the fully distributed ETC of MASs are with undirected graphs [26–29], in which the protocol design and convergence analysis based on the symmetrical property of Laplacian matrix. In practice, however, communication between agents is sometimes one-way due to different perception and communication capabilities, and thus the Laplacian matrix of the MAS becomes asymmetric [30]. It erects vast roadblocks to the protocol design and convergence analysis of MASs with directed graphs, and protocols synthesized for MASs with undirected graphs are not feasible anymore. The study on the COR problem of heterogeneous MASs with directed graphs in the model-free RL-based fully distributed ETC scheme is still an open issue and worthy of in-depth study.

The contributions of this paper are summarized as follows.

(1) First, an internal model-based augmented algebraic Riccati equation (AARE) is constructed for the heterogeneous MASs, in which the feedback gain matrix of the COR control protocol can be directly obtained by solving the AARE, thus avoiding solving the model-based output regulator equations of all agents.

(2) The dynamics of each agent considered in this paper are completely unknown, and a model-free RL algorithm is developed to learn the feedback gain matrix by the online input-output data. In addition, the information exchange among agents is assumed to be directed and a new graph-based Lyapunov function is synthesized to cope with this issue, which includes undirected graphs as special cases [26–28].

(3) Moreover, an observer-based adaptive ETC method is proposed such that the computational burden and signal transmission frequency can be reduced significantly compared with existing results [14,22,31–33]. Besides, the global information of MASs is removed, thus it achieves fully distributed control.

Notations. $\mathbb{R}$ stands for the set of real number; $\mathbb{R}^m$ indicates the real vector with m elements; $\mathbb{R}^{m\times n}$ represents the real matrix with m rows and n columns; $\mathbf{1}_n$ is the n-dimensional vector of all ones; 0 denotes zero vector or zero matrix with appropriate dimension; I_n indicates the n dimensional identity matrix; x^{T} represents the transposition of x; $\mathrm{diag}\{d_1,\ldots,d_n\}$ denotes a diagonal matrix whose diagonal entries are $d_1,\ldots,d_n$ and all other entries are zero; $\lambda_{\min}\{X\}$ and $\lambda_{\max}\{X\}$ are the minimum and maximum eigenvalues of a symmetric matrix X, respectively; $\|x\|$ represents the Euclidean norm of vector x; $\otimes$ represents the Kronecker product. Let $\mathrm{vec}(X)=[x_1^{\mathrm{T}},\ldots,x_n^{\mathrm{T}}]\in\mathbb{R}^{n^2\times 1}$, where $X\in\mathbb{R}^{n\times n}$, $x_i\in\mathbb{R}^{n\times 1}$ is the ith column of matrix X. Define $\mathrm{vecs}(X)=[x_{11},\ldots,x_{1n},x_{22},\ldots,x_{2n},\ldots,x_{nn}]^{\mathrm{T}}\in\mathbb{R}^{\frac{n(n+1)}{2}\times 1}$, where x_{ij} is the (i,j)th element of symmetric matrix $X\in\mathbb{R}^{n\times n}$. Denote $\bar{x}=[x_1^2,\ldots,2x_1x_n,x_2^2,\ldots,2x_2x_n,\ldots,x_n^2]^{\mathrm{T}}\in\mathbb{R}^{\frac{n(n+1)}{2}\times 1}$, where $x=[x_1,\ldots,x_n]^{\mathrm{T}}\in\mathbb{R}^n$.

2 Preliminaries and problem formulation

2.1 Preliminaries

Define $\mathcal{G} = \{\mathcal{V}, \mathcal{E}, \mathcal{A}\}$ as a directed graph with N agents, in which $\mathcal{V} = \{1, \ldots, N\}$ is the agent set, $\mathcal{E}$ is the edge set, and $\mathcal{A} = [a_{ij}] \in \mathbb{R}^{N\times N}$ is the adjacency matrix, where $a_{ij} = 1$ $(i \neq j)$ for $(j, i) \in \mathcal{E}$, otherwise, $a_{ij} = 0$. The neighbor set of the ith agent is given as $\mathcal{N}_i = \{j \in \mathcal{V} : (j, i) \in \mathcal{E}\}$, it indicates that agent i can get information from agents j, $j \in \mathcal{N}_i$.

A spanning tree is a subgraph of $\mathcal{G}$ in which there exists at least one access including all agents. Define $\mathcal{L} = \mathcal{D} - \mathcal{A}$ as the Laplacian matrix of $\mathcal{G}$, where $\mathcal{D} = \mathrm{diag}\{d_1, d_2, \ldots, d_N\}$, $d_i = \sum_{j\in\mathcal{N}_i} a_{ij}$. Let $\mathcal{M} = \mathrm{diag}\{a_{10}, \ldots, a_{N0}\}$, where $a_{i0} = 1$ means that the agent i has access to exosystem node 0. Let $\mathcal{H} = \mathcal{L} + \mathcal{M}$, then the graph consists of the exosystem is denoted as $\bar{\mathcal{G}}$.

2.2 Problem formulation

Consider the following heterogeneous MAS including of N agents:

$$\begin{cases} \dot{x}_i(t) = A_i x_i(t) + B_i u_i(t), \\ y_i(t) = C_i x_i(t), \end{cases} \tag{1}$$

with $i \in \mathcal{V}$, where $x_i(t) \in \mathbb{R}^{n_{1,i}}$, $u_i(t) \in \mathbb{R}^{n_{2,i}}$, and $y_i(t) \in \mathbb{R}^{n_3}$ are state, control input, and output of agent i, respectively, A_i, B_i, and C_i are some unknown constant matrices with compatible dimensions.

The exosystem state $v(t) \in \mathbb{R}^n$ is generated by

$$\begin{cases} \dot{v}(t) = S v(t), \\ r(t) = C_0 v(t), \end{cases} \tag{2}$$

where $r(t) \in \mathbb{R}^{n_3}$ is the output of the exosystem and also the reference output trajectory of MAS (1), $S \in \mathbb{R}^{n\times n}$ and $C_0 \in \mathbb{R}^{n_3\times n}$ are the system matrix and output matrix of the exosystem, respectively.

In the following, we recall the definition of the COR problem of heterogeneous MASs.

Definition 1. Consider a heterogeneous MAS (1) with exosystem (2) over a directed graph $\bar{\mathcal{G}}$. For any finite $x_i(0), i \in \mathcal{V}$ and exosystem state $v(t)$, if the tracking error

$$\tilde{e}_i(t) = y_i(t) - r(t) \tag{3}$$

satisfies the following condition for each agent i,

$$\lim_{t\to\infty} \|\tilde{e}_i(t)\| = 0, \ \forall i \in \mathcal{V}, \tag{4}$$

then the COR problem is resolved.

Remark 1. One thing should be noted is that several model-based control methods are proposed to deal with the COR problem of heterogeneous MASs [27,34,35], in which the model information is employed to obtain the solutions of ARE and output regulator equations. However, the accurate model information is hard to establish or needs high cost due to the severe operating environment. In this paper, a model-free RL algorithm will be proposed to avoid the complex modeling of heterogeneous agents and release the dependence on the model dynamics.

To proceed, the following Assumptions and Lemmas are introduced.

Assumption 1. The pairs (A_i, B_i) is controllable.

Assumption 2. The matrix S is neutrally stable.

Assumption 3. For the following linear matrix equation:

$$\begin{cases} \Pi_i S = A_i \Pi_i + B_i \Gamma_i, \\ 0 = C_i \Pi_i - C_0, \end{cases} \tag{5}$$

there exists a solution pair (Π_i, Γ_i), $i \in \mathcal{V}$.

Assumption 4. The communication graph $\bar{\mathcal{G}}$ is a directed graph containing a spanning tree rooted by exosystem (2).

Remark 2. Assumptions 1–4 are standard for the COR problem, which are common used in some important results [11,12,14]. Specifically, in Assumption 3, Eq. (5) is known as regulator equations, and a sufficient condition for its solvability can be rank$\left[\begin{smallmatrix} A_i - \lambda I_{n_{1,i}} & B_i \\ C_i & 0 \end{smallmatrix}\right] = n_{1,i} + n_3$, where λ is any eigenvalue of matrix S [34]. Assumption 4 is a minimal communication connectivity condition of MASs to achieve consensus.

Lemma 1 ([36]). As S is neutrally stable, there exists a matrix $U = \left[\begin{smallmatrix} U_1 \\ U_2 \end{smallmatrix}\right] \in \mathbb{R}^{n\times n}$ such that

$$USU^{-1} = \begin{bmatrix} X_1 & 0 \\ 0 & X_2 \end{bmatrix}, \tag{6}$$

where $X_1 \in \mathbb{R}^{n_1\times n_1}$ is a skew-symmetric matrix, $X_2 \in \mathbb{R}^{n_2\times n_2}$ is a Hurwitz matrix, and $n = n_1 + n_2$. Moreover, $U^{-1} = [\, U_1^+ \;\; U_2^+ \,]$, where $U_1U_1^+ = I_{n_1}$, $U_2U_2^+ = I_{n_2}$, $U_1U_2^+ = 0$, $U_2U_1^+ = 0$.

Lemma 2 ([37]). Under Assumption 4, one can find a positive diagonal matrix G such that $G\mathcal{H}+\mathcal{H}^{\mathrm{T}}G > 0$. Such a G can be given as diag$\{q_1, \ldots, q_N\}$, where $q = [q_1, \ldots, q_N]^{\mathrm{T}} = (\mathcal{H}^{\mathrm{T}})^{-1}\mathbf{1}_N$.

Lemma 3 ([38]). For a positive definite matrix Y and real vectors a and b, one has $2a^{\mathrm{T}}Yb \leqslant a^{\mathrm{T}}Ya + b^{\mathrm{T}}Yb$.

Remark 3. In this paper, it is assumed that the system matrix S of the exosystem (2) is neutrally stable, and it plays a key role in the convergence analysis of the COR problem of heterogeneous MASs with directed graphs. In the previous results, the adaptive parameters are usually employed to adjust the control gains to achieve the control protocol to be fully distributed [26–28]. However, all these adaptive protocols are designed with the aid of continuous communication [31–33] and/or symmetrical properties of the Laplacian matrix [26–28]. To achieve fully distributed control without utilizing adaptive parameters, the neutrally stable condition is incorporated such that the control gains can be static rather than time-varying and nonlinear. As a result, it becomes feasible for the COR problem with a generally directed graph. Moreover, the neutral stable assumption on S includes the commonly used assumption that all eigenvalues of S are semi-simple with zero real part [39–41], as a special case.

The goal of this paper is to design a fully distributed ETC protocol via the model-free RL algorithm to solve the COR problem of unknown heterogeneous MASs with directed graphs. The proposed method aims to: (1) remove the dependence on the model information, (2) solve the COR problem in a fully distributed ETC manner, and (3) further extend to more general directed graphs.

3 Main results

In this section, we consider two general scenarios. Firstly, the scenario that all agents can directly access to the exosystem is considered, and a model-free RL-based method is proposed to achieve the COR control of MASs. Then, the scenario that the exosystem is only accessible to its adjacent followers is further taken into consideration, and a distributed observer is designed to estimate the exosystem state and a fully distributed ETC protocol is synthesized to achieve the COR of heterogeneous MASs via the model-free RL algorithm. In the following, the time label t is omitted if there is no confusion.

3.1 Scenario 1: the exosystem is globally accessible to agents

In this subsection, the scenario that the exosystem is accessible to all agents is considered, and an internal model is constructed to cope with the COR problem of unknown heterogeneous MASs by solving the AARE via the model-free RL algorithm.

For each agent i, its internal model is constructed as

$$\dot{\xi}_i = S\xi_i + H\tilde{e}_i, \tag{7}$$

where $\xi_i \in \mathbb{R}^n$ is the state of the internal model, the pair (S, H) incorporates an internal model of the matrix S [14,23,34]. Then, the internal model-based control protocol is designed as

$$u_i = K_{i1}x_i + K_{i2}\xi_i, \tag{8}$$

where the feedback gain matrix $\bar{K}_i = [K_{i1}, K_{i2}]$ will be learned by the model-free RL algorithm via the input-output data in a model-free manner.

Next, we will give the model-free RL algorithm.

Let $\delta_i = [x_i^{\mathrm{T}}, \xi_i^{\mathrm{T}}]^{\mathrm{T}}$, it follows from (1) and (7) that

$$\dot{\delta}_i = \bar{A}_i \delta_i + \bar{B}_i u_i + \bar{D}_i v, \tag{9}$$

where $\bar{A}_i = \left[\begin{smallmatrix} A_i & 0 \\ HC_i & S \end{smallmatrix}\right]$, $\bar{B}_i = \left[\begin{smallmatrix} B_i \\ 0 \end{smallmatrix}\right]$, $\bar{D}_i = \left[\begin{smallmatrix} 0 \\ -HC_0 \end{smallmatrix}\right]$. Then, for each agent i at kth iteration, for an arbitrary matrix $\bar{P}_i(k) \in \mathbb{R}^{(n_{1,i}+n)\times(n_{1,i}+n)}$, and the derivative of $\delta_i^{\mathrm{T}} \bar{P}_i(k) \delta_i$ is formulated by

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}(\delta_i^{\mathrm{T}} \bar{P}_i(k)\delta_i) &= \delta_i^{\mathrm{T}}(\bar{P}_i(k)\bar{A}_i + \bar{A}_i^{\mathrm{T}}\bar{P}_i(k))\delta_i + 2u_i^{\mathrm{T}}\bar{B}_i^{\mathrm{T}}\bar{P}_i(k)\delta_i + 2v^{\mathrm{T}}\bar{D}_i^{\mathrm{T}}\bar{P}_i(k)\delta_i \\ &= \delta_i^{\mathrm{T}}\bar{H}_i(k)\delta_i - 2u_i^{\mathrm{T}}\bar{K}_i(k)\delta_i + 2v^{\mathrm{T}}\bar{Q}_i(k)\delta_i,\end{aligned} \tag{10}$$

where $\bar{H}_i(k) = \bar{P}_i(k)\bar{A}_i + \bar{A}_i^{\mathrm{T}}\bar{P}_i(k)$, $\bar{K}_i(k) = -\bar{B}_i^{\mathrm{T}}\bar{P}_i(k)$, $\bar{Q}_i(k) = \bar{D}_i^{\mathrm{T}}\bar{P}_i(k)$. Integrating both sides of the above equation, it implies that

$$\begin{aligned}\delta_i^{\mathrm{T}} \otimes \delta_i^{\mathrm{T}}|_t^{t+\kappa}\mathrm{vec}(\bar{P}_i(k)) = &\int_t^{t+\kappa} \delta_i^{\mathrm{T}} \otimes \delta_i^{\mathrm{T}}\mathrm{d}\tau\, \mathrm{vec}(H_i(k)) - 2\int_t^{t+\kappa} \delta_i^{\mathrm{T}} \otimes u_i^{\mathrm{T}}\mathrm{d}\tau\, \mathrm{vec}(\bar{K}_i(k)) \\ &+ 2\int_t^{t+\kappa} \delta_i^{\mathrm{T}} \otimes v^{\mathrm{T}}\mathrm{d}\tau\, \mathrm{vec}(\bar{Q}_i(k)),\end{aligned} \tag{11}$$

where $\kappa > 0$ is the sampling interval. Define matrices:

$$\begin{aligned} D_{i1} &= \left[\delta_i \otimes \delta_i|_{t_0}^{t_1}, \ldots, \delta_i \otimes \delta_i|_{t_{l-1}}^{t_l}\right]^{\mathrm{T}}, \\ D_{i2} &= \left[\int_{t_0}^{t_1} \bar{\delta}_i \mathrm{d}\tau, \ldots, \int_{t_{l-1}}^{t_l} \bar{\delta}_i \mathrm{d}\tau\right]^{\mathrm{T}}, \\ D_{i3} &= \left[\int_{t_0}^{t_1} \delta_i \otimes u_i \mathrm{d}\tau, \ldots, \int_{t_{l-1}}^{t_l} \delta_i \otimes u_i \mathrm{d}\tau\right]^{\mathrm{T}}, \\ D_{i4} &= \left[\int_{t_0}^{t_1} \delta_i \otimes v \mathrm{d}\tau, \ldots, \int_{t_{l-1}}^{t_l} \delta_i \otimes v \mathrm{d}\tau\right]^{\mathrm{T}}, \end{aligned}$$

where l is a positive integer, and $0 \leqslant t_0 < t_1 < \cdots < t_l$ are constants. Thus, Eq. (11) can be rewritten in a more compact form as

$$D_{i1}\mathrm{vec}(\bar{P}_i(k)) = D_{i5}\begin{bmatrix} \mathrm{vecs}(\bar{H}_i(k)) \\ \mathrm{vec}(\bar{K}_i(k)) \\ \mathrm{vec}(\bar{Q}_i(k)) \end{bmatrix}, \tag{12}$$

where $D_{i5} = [D_{i2}\ \ -2D_{i3}\ 2D_{i4}]$. Then, with enough online input-output data of agent i, the existence of $(D_{i5}^{\mathrm{T}} D_{i5})^{-1}$ can be guaranteed [14, 23, 42, 43], and it holds that

$$\begin{bmatrix} \mathrm{vecs}(\bar{H}_i(k)) \\ \mathrm{vec}(\bar{K}_i(k)) \\ \mathrm{vec}(\bar{Q}_i(k)) \end{bmatrix} = (D_{i5}^{\mathrm{T}} D_{i5})^{-1} D_{i5}^{\mathrm{T}} D_{i1}\mathrm{vec}(\bar{P}_i(k)). \tag{13}$$

Besides, the learning rate $\iota_i(k)$ satisfies the following conditions:

$$\iota_i(k) > 0,\ \sum_{k=0}^{\infty} \iota_i(k) > \infty,\ \sum_{k=0}^{\infty} (\iota_i(k))^2 < \infty. \tag{14}$$

The model-free RL algorithm for learning $\bar{K}_i$ is given in Algorithm 1.

Algorithm 1 Model-free RL algorithm for learning $\bar{K}_i$.

Initialization: For each agent i (1) with disturbed control protocol (8), i.e., $u_i = K_{i1}x_i + K_{i2}\xi_i + \nu_i$ with an exploration noise ν_i such that $(D_{i5}^{\mathrm{T}}D_{i5})^{-1}$ exists, given a small threshold $\epsilon_i > 0$, and learning rate $\iota_i(0) > 0$. Set $\bar{P}_i(0) = I_{n_{1,i}+n}$.
while $\|\bar{P}_i(k+1) - \bar{P}_i(k)\|/\iota_i(k) \geqslant \epsilon_i$ **do**
  Obtain $(\bar{H}_i(k), \bar{K}_i(k), \bar{Q}_i(k))$ via (13) with online input-output data, and calculate $\tilde{\bar{P}}_i(k+1) = \bar{P}_i(k) + \iota_i^k(\bar{H}_i(k) + I_{n_{1,i}+n} - 2\bar{K}_i(k)^{\mathrm{T}}\bar{K}_i(k))$;
  if $\bar{P}_i(k+1) \leqslant 0$ **then**
    $\bar{P}_i(k+1) = \bar{P}_i(0)$;
  else
    $\bar{P}_i(k+1) = \tilde{\bar{P}}_i(k+1)$;
  end if
end while
Return $\bar{K}_i(k)$.

Remark 4. In Algorithm 1, the key iteration primarily focuses on $\tilde{\bar{P}}_i(k+1) = \bar{P}_i(k) + \iota_i^k(\bar{H}_i(k) + I_{n_{1,i}+n} - 2\bar{K}_i(k)^{\mathrm{T}}\bar{K}_i(k))$. It can be observed that once the driving force $(\bar{H}_i(k) + I_{n_{1,i}+n} - 2\bar{K}_i(k)^{\mathrm{T}}\bar{K}_i(k))$ of updates approaches zero, it implies that $\bar{P}_i\bar{A}_i + \bar{A}_i^{\mathrm{T}}\bar{P}_i - 2\bar{P}_i\bar{B}_i\bar{B}_i^{\mathrm{T}}\bar{P}_i + I_{n_{1,i}+n} = 0$ is satisfied approximately with an appropriate accuracy $\epsilon_i > 0$. Consequently, we can ultimately obtain the control matrix $\bar{K}_i(k)$ in a model-free manner.

Theorem 1. Consider the unknown heterogeneous MAS (1) with an exosystem (2) satisfying Assumptions 1–3, the model-free RL-based control protocol is designed as (8), where the feedback gain matrix $\bar{K}_i = [K_{i1}, K_{i2}]$ is learned via Algorithm 1, then the COR problem can be resolved.

Proof. First, according to [42], the feedback gain matrix $\bar{K}_i$ learned via Algorithm 1 approaches $\bar{K}_i = -\bar{B}_i^{\mathrm{T}}\bar{P}_i$, where $\bar{P}_i$ is the solution of the AARE:

$$\bar{P}_i\bar{A}_i + \bar{A}_i^{\mathrm{T}}\bar{P}_i - 2\bar{P}_i\bar{B}_i\bar{B}_i^{\mathrm{T}}\bar{P}_i + I_{n_{1,i}+n} = 0. \tag{15}$$

Then, based on Lemma 1.26 of [34] with Assumptions 1–3, if the pair (S, H) incorporates an internal model of the matrix S, then the pair $(\bar{A}_i, \bar{B}_i)$ is stabilizable. Thus, with $\bar{K}_i = [K_{i1}, K_{i2}] = -\bar{B}_i^{\mathrm{T}}\bar{P}_i$, it is obvious that the matrix

$$\hat{A}_i = \begin{bmatrix} A_i + B_iK_{1i} & B_iK_{2i} \\ HC_i & S \end{bmatrix} \tag{16}$$

is Hurwitz. Then, according to Lemma 1.27 of [34], the following equations:

$$\begin{cases} \hat{\mathit{\Pi}}_iS = (A_i + B_iK_{1i})\hat{\mathit{\Pi}}_i + B_iK_{2i}\hat{\mathit{\Gamma}}_i, \\ \hat{\mathit{\Gamma}}_iS = S\hat{\mathit{\Gamma}}_i \end{cases} \tag{17}$$

have a unique solution $(\hat{\mathit{\Pi}}_i, \hat{\mathit{\Gamma}}_i)$. Comparing the corresponding terms in (5) and (17), it can be derived that

$$\begin{cases} \mathit{\Pi}_i = \hat{\mathit{\Pi}}_i, \\ \mathit{\Gamma}_i = K_{1i}\mathit{\Pi}_i + K_{2i}\hat{\mathit{\Gamma}}_i. \end{cases}$$

Define $\tilde{x}_i = x_i - \mathit{\Pi}_iv$ and $\tilde{\xi}_i = \xi_i - \hat{\mathit{\Gamma}}_iv$, one has

$$\begin{aligned} \dot{\tilde{x}}_i &= (A_i + B_iK_{1i})\tilde{x}_i + B_iK_{2i}\tilde{\xi}_i, \\ \dot{\tilde{\xi}}_i &= S\tilde{\xi}_i + HC_i\tilde{x}_i. \end{aligned}$$

Let $\tilde{\delta}_i = [\tilde{x}_i^{\mathrm{T}}, \tilde{\xi}_i^{\mathrm{T}}]^{\mathrm{T}}$ and $\tilde{\mu}_i = \bar{K}_i\tilde{\delta}_i = \mu_i - \mathit{\Gamma}_iv$, and then

$$\dot{\tilde{\delta}}_i = (\bar{A}_i + \bar{B}_i\bar{K}_i)\tilde{\delta}_i. \tag{18}$$

Choose the Lyapunov function candidate as

$$V = \sum_{i=1}^{N} \tilde{\delta}_i^{\mathrm{T}}\bar{P}_i\tilde{\delta}_i, \tag{19}$$

and its derivative is given by

$$\dot{V} = \sum_{i=1}^{N} 2\tilde{\delta}_i^{\mathrm{T}} \bar{P}_i(\bar{A}_i + \bar{B}_i\bar{K}_i)\tilde{\delta}_i = \sum_{i=1}^{N} \tilde{\delta}_i^{\mathrm{T}} (\bar{P}_i\bar{A}_i + \bar{A}_i^{\mathrm{T}}\bar{P}_i - 2\bar{P}_i\bar{B}_i\bar{B}_i^{\mathrm{T}}\bar{P}_i)\tilde{\delta}_i = -\sum_{i=1}^{N} \tilde{\delta}_i^{\mathrm{T}}\tilde{\delta}_i \leqslant 0. \tag{20}$$

Thus, we have $\lim_{t\to\infty} \tilde{\delta}_i(t) = 0$. As a result, $\lim_{t\to\infty} \tilde{x}_i(t) = 0$. Then, from $\tilde{e}_i = y_i - C_0 v = C_i\tilde{x}_i + (C_i\Pi_i - C_0)v = C_i\tilde{x}_i$, it is obvious that $\lim_{t\to\infty} \tilde{e}_i(t) = 0$. Thus, the COR of the heterogeneous MAS is resolved.

Remark 5. With the proposed model-free RL algorithm, the heterogeneous dynamics of MAS (1) could be unknown. By solving (13) iteratively, the model-free feedback gain matrix $\bar{K}_i(k)$ learned via Algorithm 1 will converge to the model-based solution $\bar{K}_i = -\bar{B}_i\bar{P}_i$, where $\bar{P}_i$ is the solution of AARE (15). Thus, the COR problem of heterogeneous MASs is resolved in a model-free manner.

3.2 Scenario 2: the exosystem is locally accessible to agents

In this subsection, based on the result of Theorem 1, the scenario that the exosystem is only accessible to its adjacent followers over the directed graph is further considered, and a fully distributed adaptive ETC protocol is synthesized based on a distributed observer to cope with the COR problem of heterogeneous MASs.

For each agent, the observer-based fully distributed adaptive ETC protocol is synthesized as

$$u_i = K_{i1}x_i + K_{i2}\xi_i, \tag{21a}$$

$$\dot{\xi}_i = S\xi_i + H(y_i - C_0\eta_i), \tag{21b}$$

$$\dot{\eta}_i = S\eta_i + K\left[\sum_{j=1}^{N} a_{ij}(\hat{\eta}_i - \hat{\eta}_j) + a_{i0}(\hat{\eta}_i - v)\right], \tag{21c}$$

where $\xi_i \in \mathbb{R}^n$ and $\eta_i \in \mathbb{R}^n$ are the states of the internal model (21b) and distributed observer (21c) designed for the exosystem, respectively, the feedback gain matrix $\bar{K}_i = [K_{i1}, K_{i2}]$ will be learned via Algorithm 1, the pair (S, H) incorporates the internal model of the matrix S [14, 23, 34], $K = -U_1^{\mathrm{T}}U_1$, $\hat{\eta}_i$ is the open-loop estimate of η_i, which is given by

$$\begin{cases} \hat{\eta}_i = \eta_i, \quad t = t_k^i, \\ \dot{\hat{\eta}}_i = S\hat{\eta}_i, t \in [t_k^i, t_{k+1}^i), \end{cases} \tag{22}$$

where t_{k+1}^i is the event-triggered time instant for transmitting the observer state η_i of agent i that is determined by the adaptive ETM:

$$t_{k+1}^i = \inf\{t > t_k^i | f_i(e_i, \hat{z}_i, c_i, t) \geqslant 0\}, \tag{23}$$

in which

$$f_i(e_i, \hat{z}_i, c_i, t) = c_i e_i^{\mathrm{T}} e_i - \hat{z}_i^{\mathrm{T}} K^{\mathrm{T}} K \hat{z}_i - \beta_i \mathrm{e}^{-\alpha_i t} \tag{24}$$

with $e_i = \hat{\eta}_i - \eta_i$, $\hat{z}_i = \sum_{j=1}^{N} a_{ij}(\hat{\eta}_i - \hat{\eta}_j) + a_{i0}(\hat{\eta}_i - v)$, $\alpha_i > 0$, $\beta_i > 0$, and the adaptive parameter c_i is updated by

$$\dot{c}_i = e_i^{\mathrm{T}} e_i \tag{25}$$

with an initial value $c_i(0) > 0$.

Remark 6. Under Assumption 2 and with given constant gain of the observer (21c), the COR of heterogeneous MASs with directed graph could be resolved in a fully distributed and event-triggered manner without using adaptive or time-varying gain before K. Different from [31–33], the designed constant gain is of great convenience for the real application, and the proposed ETC can save more communication resources. Moreover, compared with [26–28], the considered case of directed graphs includes them as special cases.

Theorem 2. Consider the unknown heterogeneous MAS (1) with an exosystem (2) over a directed graph satisfying Assumptions 1–4, the model-free RL-based fully distributed ETC protocol is designed as (21), and the ETM for agent i is given in (23), where the feedback gain matrix $\bar{K}_i = [K_{i1}, K_{i2}]$ is learned via Algorithm 1, the pair (S, H) incorporates an internal model of the matrix S, and $K = -U_1^{\mathrm{T}} U_1$, then the COR problem can be resolved. Moreover, the adaptive parameter c_i increase monotonously to some positive constant, and the Zeno behavior can be strictly ruled out for each agent.

Proof. First, we focus on the asymptotic stability of the augmented system composed of the exosystem (2) and the observer (21c), i.e., $\lim_{t\to\infty}(\eta_i(t) - v(t)) = 0$, and the upper boundedness of adaptive parameter c_i.

Let $z_i = \sum_{j\in\mathcal{N}_i} a_{ij}(\eta_i - \eta_j) + a_{i0}(\eta_i - v)$, $\varepsilon_i = \eta_i - v$, $\hat{z}_i = \sum_{j=1}^N a_{ij}(\hat{\eta}_i - \hat{\eta}_j) + a_{i0}(\hat{\eta}_i - v)$, and $\hat{\varepsilon}_i = \hat{\eta}_i - v$, one has

$$z = (\mathcal{H} \otimes I_n)\varepsilon, \ \hat{z} = (\mathcal{H} \otimes I_n)\hat{\varepsilon}, \tag{26}$$

where $z = [z_1^{\mathrm{T}}, \ldots, z_N^{\mathrm{T}}]^{\mathrm{T}}$, $\varepsilon = [\varepsilon_1^{\mathrm{T}}, \ldots, \varepsilon_N^{\mathrm{T}}]^{\mathrm{T}}$, $\hat{z} = [\hat{z}_1^{\mathrm{T}}, \ldots, \hat{z}_N^{\mathrm{T}}]^{\mathrm{T}}$, $\hat{\varepsilon} = [\hat{\varepsilon}_1^{\mathrm{T}}, \ldots, \hat{\varepsilon}_N^{\mathrm{T}}]^{\mathrm{T}}$. From (2), (21c), and (26), the dynamics of the observer error ε can be written as

$$\dot{\varepsilon} = (I_N \otimes S)\varepsilon + (\mathcal{H} \otimes K)\hat{\varepsilon}. \tag{27}$$

Let $s = (I_N \otimes U)\varepsilon$ and $\hat{s} = (I_N \otimes U)\hat{\varepsilon}$, then we have

$$\begin{aligned} \dot{s} &= (I_N \otimes USU^{-1})s + (\mathcal{H} \otimes UKU^{-1})\hat{s} \\ &= \left(I_N \otimes \begin{bmatrix} X_1 & 0 \\ 0 & X_2 \end{bmatrix}\right)s - \left(\mathcal{H} \otimes \begin{bmatrix} \Phi_1 & 0 \\ \Phi_2 & 0 \end{bmatrix}\right)\hat{s}, \end{aligned} \tag{28}$$

where $\Phi_1 = U_1 U_1^{\mathrm{T}}$ and $\Phi_2 = U_2 U_1^{\mathrm{T}}$. Define $s_1 = (I_N \otimes U_1)\varepsilon$ and $s_2 = (I_N \otimes U_2)\varepsilon$, then we have

$$\dot{s}_1 = (I_N \otimes X_1)s_1 - (\mathcal{H} \otimes \Phi_1)\hat{s}_1, \tag{29}$$

$$\dot{s}_2 = (I_N \otimes X_2)s_2 - (\mathcal{H} \otimes \Phi_2)\hat{s}_1, \tag{30}$$

where $\hat{s}_1 = (I_N \otimes U_1)\hat{\varepsilon}$.

Choose a Lyapunov function candidate as

$$V = V_1 + V_2 + V_3 + V_4, \tag{31}$$

where $V_1 = m_1 s_1^{\mathrm{T}}(G \otimes I_{n_1})s_1$, $V_2 = m_2 s_2^{\mathrm{T}}(I_N \otimes P)s_2$, $V_3 = m_3 \sum_{i=1}^N (c_i - c_0)^2$, $V_4 = m_4 \sum_{i=1}^N \phi_i \mathrm{e}^{-\alpha_i t}$, $G = \mathrm{diag}\{q_1, \ldots, q_N\}$, $q = [q_1, \ldots, q_N]^{\mathrm{T}} = (\mathcal{H}^{\mathrm{T}})^{-1}\mathbf{1}_N$, m_1, m_2, m_3, m_4, c_0 and ϕ_i are positive constants to be specified later, c_i is the adaptive parameter updated by (25), $P > 0$ is the solution of the Lyapunov equation:

$$PX_2 + X_2^{\mathrm{T}}P + 2I_{n_2} = 0. \tag{32}$$

As X_1 is a skew-symmetric matrix, it follows from (29) that

$$\begin{aligned} \dot{V}_1 &= 2m_1 s_1^{\mathrm{T}}(G \otimes I_{n_1})\dot{s}_1 \\ &= m_1 s_1^{\mathrm{T}}[G \otimes (X_1 + X_1^{\mathrm{T}})]s_1 - 2m_1 s_1^{\mathrm{T}}(G\mathcal{H} \otimes \Phi_1)\hat{s}_1 \\ &= -2m_1 s_1^{\mathrm{T}}(G\mathcal{H} \otimes \Phi_1)\hat{s}_1. \end{aligned} \tag{33}$$

From $\varepsilon = \eta - (\mathbf{1}_N \otimes v)$ and (26), one can obtain that

$$s_1 = (\mathcal{H}^{-1} \otimes U_1)\hat{z} - (I_N \otimes U_1)e. \tag{34}$$

Thus, the derivative of V_1 can be further written as

$$\begin{aligned} \dot{V}_1 &= -2m_1[\hat{z}^{\mathrm{T}}(\mathcal{H}^{-\mathrm{T}} \otimes U_1^{\mathrm{T}}) - e^{\mathrm{T}}(I_N \otimes U_1^{\mathrm{T}})](G\mathcal{H} \otimes \Phi_1)\hat{s}_1 \\ &= -2m_1 \hat{z}^{\mathrm{T}}(\mathcal{H}^{-\mathrm{T}}G \otimes \Phi_1^2)\hat{z} + 2m_1 e^{\mathrm{T}}(G \otimes \Phi_1^2)\hat{z}. \end{aligned} \tag{35}$$

According to Lemma 2, it yields that

$$-2m_1 \hat{z}^{\mathrm{T}}(\mathcal{H}^{-\mathrm{T}}G \otimes \Phi_1^2)\hat{z} = -m_1 \hat{z}^{\mathrm{T}}[\mathcal{H}^{-\mathrm{T}}(G\mathcal{H} + \mathcal{H}^{\mathrm{T}}G)\mathcal{H}^{-1} \otimes \Phi_1^2]\hat{z}$$

$$\leqslant -m_1\lambda_0\hat{z}^{\mathrm{T}}(\mathcal{H}^{-\mathrm{T}}\mathcal{H}^{-1}\otimes\Phi_1^2)\hat{z}, \tag{36}$$

where $\lambda_0 = \lambda_{\min}\{G\mathcal{H}+\mathcal{H}^{\mathrm{T}}G\}$. Using (34) with Lemma 3, one has

$$\begin{aligned}-m_1\lambda_0\hat{z}^{\mathrm{T}}(\mathcal{H}^{-\mathrm{T}}\mathcal{H}^{-1}\otimes\Phi_1^2)\hat{z} =& -m_1\lambda_0[s_1^{\mathrm{T}}+e^{\mathrm{T}}(I_N\otimes U_1^{\mathrm{T}})](I_N\otimes\Phi_1)[s_1+(I_N\otimes U_1)e]\\ \leqslant& -\frac{m_1\lambda_0}{2}s_1^{\mathrm{T}}(I_N\otimes\Phi_1)s_1+m_1\lambda_0e^{\mathrm{T}}(I_N\otimes\Phi_1^2)e,\end{aligned} \tag{37}$$

and

$$\begin{aligned}2m_1e^{\mathrm{T}}(G\otimes\Phi_1^2)\hat{z} \leqslant& \frac{m_1q_{\max}^2}{h_1}\hat{z}^{\mathrm{T}}(I_N\otimes\Phi_1^2)\hat{z}+m_1h_1e^{\mathrm{T}}(I_N\otimes\Phi_1^2)e\\ \leqslant& \frac{2m_1q_{\max}^2}{\lambda_1h_1}s_1^{\mathrm{T}}(I_N\otimes\Phi_1)s_1+\left(\frac{2m_1q_{\max}^2}{\lambda_1h_1}+m_1h_1\right)e^{\mathrm{T}}(I_N\otimes\Phi_1^2)e,\end{aligned} \tag{38}$$

where $\lambda_1 = \lambda_{\min}\{\mathcal{H}^{-\mathrm{T}}\mathcal{H}^{-1}\}$, $h_1>0$ is a constant, $q_{\max}$ is the maximum value of the real vector q. From (36)–(38), it is concluded that

$$\dot{V}_1 \leqslant -\left(\frac{m_1\lambda_0\lambda_2}{2}-\frac{2m_1\lambda_3q_{\max}^2}{\lambda_1h_1}\right)s_1^{\mathrm{T}}s_1+\left(\frac{2m_1q_{\max}^2\lambda_3^2}{\lambda_1h_1}+m_1h_1\lambda_3^2+m_1\lambda_0\lambda_3^2\right)e^{\mathrm{T}}e, \tag{39}$$

where $\lambda_2=\lambda_{\min}\{\Phi_1\}$, $\lambda_3=\lambda_{\max}\{\Phi_1\}$. Next, combining (30) and (32) with Lemma 3, the derivative of V_2 can be obtained as

$$\begin{aligned}\dot{V}_2 =& 2m_2s_2^{\mathrm{T}}(I_N\otimes P)\dot{s}_2\\ =& m_2s_2^{\mathrm{T}}[I_N\otimes(PX_2+X_2^{\mathrm{T}}P)]s_2-2m_2s_2^{\mathrm{T}}(\mathcal{H}\otimes P\Phi_2)\hat{s}_1\\ \leqslant& -2m_2s_2^{\mathrm{T}}s_2-2m_2s_2^{\mathrm{T}}(\mathcal{H}\otimes P\Phi_2)\hat{s}_1\\ \leqslant& -m_2s_2^{\mathrm{T}}s_2+m_2\hat{s}_1^{\mathrm{T}}(\mathcal{H}^{\mathrm{T}}\mathcal{H}\otimes\Phi_2^{\mathrm{T}}PP\Phi_2)\hat{s}_1.\end{aligned} \tag{40}$$

With the help of Lemma 3, the second term in (40) can be further restricted by

$$\begin{aligned}m_2\hat{s}_1^{\mathrm{T}}(\mathcal{H}^{\mathrm{T}}\mathcal{H}\otimes\Phi_2^{\mathrm{T}}PP\Phi_2)\hat{s}_1 \leqslant& \frac{m_2\lambda_4}{\lambda_1}\hat{z}^{\mathrm{T}}(\mathcal{H}^{-\mathrm{T}}\mathcal{H}^{-1}\otimes\Phi_1^2)\hat{z}\\ \leqslant& \frac{2m_2\lambda_4\lambda_3}{\lambda_1}s_1^{\mathrm{T}}s_1+\frac{2m_2\lambda_3^2\lambda_4}{\lambda_1}e^{\mathrm{T}}e,\end{aligned} \tag{41}$$

where $\lambda_4=\|PU_2\|^2$. Consequently, from (40) and (41), it yields that

$$\dot{V}_2 \leqslant -m_2s_2^{\mathrm{T}}s_2+\frac{2m_2\lambda_4\lambda_3}{\lambda_1}s_1^{\mathrm{T}}s_1+\frac{2m_2\lambda_3^2\lambda_4}{\lambda_1}e^{\mathrm{T}}e. \tag{42}$$

Furthermore, with ETM (23) and (25), $\dot{V}_3$ can be deduced as

$$\begin{aligned}\dot{V}_3 =& 2m_3\sum_{i=1}^{N}(c_i-c_0)e_i^{\mathrm{T}}e_i\\ \leqslant& \frac{2m_3}{\lambda_1}\hat{z}(\mathcal{H}^{-\mathrm{T}}\mathcal{H}^{-1}\otimes\Phi_1^2)\hat{z}+2m_3\sum_{i=1}^{N}\beta_i\mathrm{e}^{-\alpha_it}-2m_3c_0e^{\mathrm{T}}e.\end{aligned} \tag{43}$$

Similar to the process in (41), the first term of (43) is restricted by

$$\frac{2m_3}{\lambda_1}\hat{z}(\mathcal{H}^{-\mathrm{T}}\mathcal{H}^{-1}\otimes\Phi_1^2)\hat{z} \leqslant \frac{4m_3\lambda_3}{\lambda_1}s_1^{\mathrm{T}}s_1+\frac{4m_3\lambda_3^2}{\lambda_1}e^{\mathrm{T}}e. \tag{44}$$

Accordingly, $\dot{V}_3$ is given as

$$\dot{V}_3 \leqslant \frac{4m_3\lambda_3}{\lambda_1}s_1^{\mathrm{T}}s_1-\left(2m_3c_0-\frac{4m_3\lambda_3^2}{\lambda_1}\right)e^{\mathrm{T}}e+2m_3\sum_{i=1}^{N}\beta_i\mathrm{e}^{-\alpha_it}. \tag{45}$$

Besides, the derivative of V_4 can be directly calculated as

$$\dot{V}_4 = -m_4 \sum_{i=1}^{N} \alpha_i \phi_i \mathrm{e}^{-\alpha_i t}. \tag{46}$$

As a result, by combining (39), (42), (45), and (46), one can summarize $\dot{V}$ as

$$\begin{aligned}\dot{V} \leqslant &- \left(\frac{m_1\lambda_0\lambda_2}{2} - \frac{2m_1\lambda_3 q_{\max}^2}{\lambda_1 h_1} - \frac{2m_2\lambda_4\lambda_3}{\lambda_1} - \frac{4m_3\lambda_3}{\lambda_1} \right) s_1^{\mathrm{T}} s_1 - m_2 s_2^{\mathrm{T}} s_2 \\ &- \left[2m_3 c_0 - \left(\frac{2m_1 q_{\max}^2 \lambda_3^2}{\lambda_1 h_1} + m_1 h_1 \lambda_3^2 + m_1\lambda_0\lambda_3^2 + \frac{2m_2\lambda_3^2\lambda_4}{\lambda_1} + \frac{4m_3\lambda_3^2}{\lambda_1} \right) \right] e^{\mathrm{T}} e \\ &- \sum_{i=1}^{N} (m_4\alpha_i\phi_i - 2m_3\beta_i)\mathrm{e}^{-\alpha_i t}. \end{aligned} \tag{47}$$

By choosing a sufficiently large c_0, one can remove the term related to $e^{\mathrm{T}}(I_N \otimes \Phi_1^2)e$. Define $\gamma_1 = \frac{m_1\lambda_0\lambda_2}{2} - \frac{2m_1\lambda_3 q_{\max}^2}{\lambda_1 h_1} - \frac{2m_2\lambda_4\lambda_3}{\lambda_1} - \frac{4m_3\lambda_3}{\lambda_1}$, $\gamma_2 = m_2$, and $\gamma_{3,i} = m_4\alpha_i\phi_i - 2m_3\beta_i$, one can find a group of appropriate parameters m_1, m_2, m_3, m_4, h_1, and ϕ_i such that $\gamma_1 > 0$, $\gamma_2 > 0$, and $\gamma_{3,i} > 0$. As a result, it is concluded that

$$\dot{V} \leqslant -\gamma_1 s_1^{\mathrm{T}} s_1 - \gamma_2 s_2^{\mathrm{T}} s_2 - \sum_{i=1}^{N} \gamma_{3,i} \mathrm{e}^{-\alpha_i t} \leqslant 0. \tag{48}$$

Thus, one can get $\lim_{t\to\infty}(\eta_i(t) - v(t)) = 0,\ i \in \mathcal{V}$. Moreover, since $V(t) \geqslant 0$ and $V(t)$ is nonincreasing over $[0,\infty)$, one can conclude that $c_i(t)$ is upper bounded by some positive constant.

Next, based on $\lim_{t\to\infty}(\eta_i(t) - v(t)) = 0$, we prove the convergence of COR problem, i.e., $\lim_{t\to\infty} \tilde{e}_i(t) = 0$.

Similar with the proof of Theorem 1, redefine $\tilde{x}_i = x_i - \Pi_i v$ and $\tilde{\xi}_i = \xi_i - \hat{\Gamma}_i v$, one has

$$\dot{\tilde{x}}_i = (A_i + B_i K_{1i})\tilde{x}_i + B_i K_{2i} \tilde{\xi}_i, \tag{49}$$

and

$$\dot{\tilde{\xi}}_i = S\hat{\xi}_i + HC_i\tilde{x}_i + HC_0(v - \eta_i). \tag{50}$$

Let $\tilde{\delta}_i = [\tilde{x}_i^{\mathrm{T}}, \tilde{\xi}_i^{\mathrm{T}}]^{\mathrm{T}}$ and $\tilde{\mu}_i = \bar{K}_i\tilde{\delta}_i = \mu_i - \Gamma_i v$, one has

$$\dot{\tilde{\delta}}_i = \bar{A}_i\tilde{\delta}_i + \bar{B}_i\tilde{\mu}_i + \bar{D}_i\varepsilon_i = (\bar{A}_i + \bar{B}_i\bar{K}_i)\tilde{\delta}_i + \bar{D}_i\varepsilon_i. \tag{51}$$

As $\lim_{t\to\infty} \varepsilon_i(t) = (\eta_i(t) - v(t)) = 0, i \in \mathcal{V}$, according to the input-to-state stability theory [44], we have $\lim_{t\to\infty} \tilde{\delta}_i(t) = 0$. Thus, the COR of MASs is resolved.

Finally, the proof of the Zeno exclusion of ETM (23) is given as follows.

For $\|e_i(t)\|$, its derivative is calculated by

$$\frac{\mathrm{d}\|e_i(t)\|}{\mathrm{d}t} = \frac{e_i^{\mathrm{T}}(t)\dot{e}_i(t)}{\|e_i(t)\|} \leqslant \|\dot{e}_i(t)\|. \tag{52}$$

From $e_i = \hat{\eta}_i - \eta_i$, (21c), and (22), one can derive that

$$\|\dot{e}_i(t)\| = \|\dot{\hat{\eta}}_i(t) - \dot{\eta}_i(t)\| = \|Se_i + K\hat{z}_i\| \leqslant \lambda_5\|e_i\| + \|K\hat{z}_i\| \leqslant \lambda_5\|e_i\| + W^*, \tag{53}$$

where $\lambda_5 = \|S\|$ and W^* is the upper bound of $\|K\hat{z}_i\|$. Then, Eq. (53) can be further transformed into

$$\|e_i(t)\| \leqslant \mathrm{e}^{\lambda_5(t-t_k^i)}\|e_i(t_k^i)\| - \int_{t_k^i}^{t} \mathrm{e}^{\lambda_5(t-\tau)} W^* \mathrm{d}\tau \leqslant \frac{W^*}{\lambda_5}\left(\mathrm{e}^{\lambda_5(t-t_k^i)} - 1\right). \tag{54}$$

Substituting $t = t_{k+1}^{i-}$ into (54), then

$$\|e_i(t_{k+1}^{i-})\| \leqslant \frac{W^*}{\lambda_5}\left(\mathrm{e}^{\lambda_5(t_{k+1}^{i-} - t_k^i)} - 1\right). \tag{55}$$

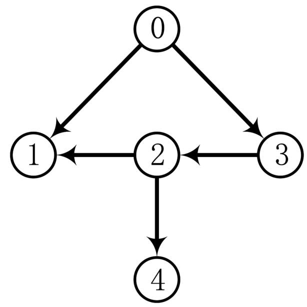

Figure 1 Directed communication graph $\bar{\mathcal{G}}$.

By simple calculation on (55), one has

$$t_{k+1}^{i-} - t_k^i \geqslant \frac{1}{\lambda_5} \ln\left(\frac{\lambda_5 \|e_i(t_{k+1}^{i-})\|}{W^*} + 1\right). \tag{56}$$

For the triggering instant t_{k+1}^{i-}, from the event-triggered function (24), we have following inequality:

$$\|e_i(t_{k+1}^{i-})\|^2 \geqslant \frac{\beta_i \mathrm{e}^{-\alpha_i t_{k+1}^{i-}}}{c_i(t_{k+1}^{i-})} \geqslant \frac{\beta_i \mathrm{e}^{-\alpha_i t_{k+1}^{i-}}}{\bar{c}_i}, \tag{57}$$

where $\bar{c}_i$ is the upper bound of $c_i(t)$. Therefore, there exists a lower bound between two triggering instants that can be given by

$$t_{k+1}^{i} - t_k^i \geqslant \frac{1}{\lambda_5} \ln\left(\frac{\lambda_5 \sqrt{\frac{\beta_i}{\bar{c}_i}} \mathrm{e}^{\frac{-\alpha_i t_{k+1}^{i}}{2}}}{W^*} + 1\right). \tag{58}$$

Thus, it concludes that the Zeno behavior is excluded for the proposed fully distributed ETC protocol. This completes the proof.

Remark 7. In the proof of Theorem 2, the auxiliary parameters q, m_1, m_2, m_3, m_4, c_0 and ϕ_i that may dependent on global information are irrelevant to the control gain, i.e., for arbitrary control gain designed in Theorem 2, there exist appropriate auxiliary parameters to achieve the asymptotic COR of MASs.

Remark 8. In practice, the communication among agents sometimes becomes unidirectional due to network bandwidth, signal interference, communication interruption, and other factors. In this paper, more general directed graphs are considered. Based on the condition that matrix S is neutrally stable, the graph-based diagonal matrix G is constructed and a Lyapunov function (31) is designed to conquer the challenge of directed graphs.

4 Simulation example

In this section, we have conducted two corresponding simulations to demonstrate the effectiveness and superiority of the proposed model-free RL-based fully distributed ETC scheme. The first one uses the proposed control protocol that has the ETC and adaptive ETM, which illustrates the feasibility of the proposed algorithm in resolving the COR problem. Then, in the second simulation, an ETC protocol without introducing adaptive parameters for ETM has been further conducted, which nicely demonstrates the flexibility and superiority of the proposed algorithm in reducing the communication frequency.

The directed graph is shown in Figure 1, which consists of 4 agents labeled with node $1, 2, 3, 4$, and the exosystem labeled with node 0, and the arrow represents the directed communication channel.

Then, the corresponding $\mathcal{A}$, $\mathcal{L}$, $\mathcal{M}$, and $\mathcal{H}$ are given as

$$\mathcal{A} = \begin{bmatrix} 0\,1\,0\,0 \\ 0\,0\,1\,0 \\ 0\,0\,0\,1 \\ 1\,0\,0\,0 \end{bmatrix}, \mathcal{L} = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \\ -1 & 0 & 0 & 1 \end{bmatrix}, \mathcal{M} = \begin{bmatrix} 1\,0\,0\,0 \\ 0\,0\,0\,0 \\ 0\,0\,1\,0 \\ 0\,0\,0\,0 \end{bmatrix}, \mathcal{H} = \begin{bmatrix} 2 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 2 & -1 \\ -1 & 0 & 0 & 1 \end{bmatrix}.$$

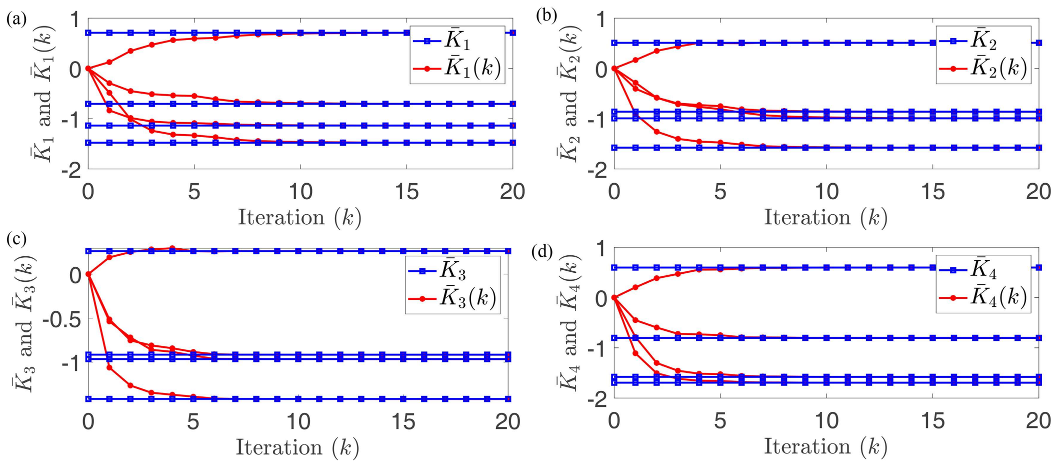

Figure 2 (Color online) Matrices $\bar{K}_i$ and $\bar{K}_i(k)$ of all agents. (a) Agent 1; (b) agent 2; (c) agent 3; (d) agent 4.

Similar with [14], the considered agents of the heterogeneous MAS (1) have the following constant matrices:

$$A_i = \begin{bmatrix} 0 & a_i \\ 0 & 0 \end{bmatrix},\ B_i = \begin{bmatrix} 0 \\ b_i \end{bmatrix}, C_i = \begin{bmatrix} c_i & d_i \end{bmatrix},$$

where $a_1 = 0.3621$, $a_2 = 0.2471$, $a_3 = 0.4666$, $a_4 = 0.6782$, $b_1 = 0.9341$, $b_2 = 0.4849$, $b_3 = 0.8184$, $b_4 = 0.6492$, $c_1 = 0.8043$, $c_2 = 0.4361$, $c_3 = 0.4914$, $c_4 = 0.9828$, $d_1 = 0.2902$, $d_2 = 0.8320$, $d_3 = 0.8400$, $d_4 = 0.5836$, but they are unknown in advance and not employed in the control protocol design. The external reference input, that is the exosystem, $v = [v^1, v^2]^{\mathrm{T}}$ is modeled as (2), where

$$S = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix},\ C_0 = \begin{bmatrix} 1 & 1 \end{bmatrix}.$$

Thus, based on [34, 36], one can design

$$H = \begin{bmatrix} 0 \\ 1 \end{bmatrix},\ K = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}.$$

With the given communication graph and dynamics of all agents, we first conduct the proposed fully distributed model-free RL-based adaptive ETC. First, perform Algorithm 1 for each agent, and the learning results of $\bar{K}_i(k)$ are given in Figure 2, in which the model-based matrices $\bar{K}_i$ are calculated via (15). As shown in Figure 2, the learning-based matrix $\bar{K}_i(k)$ will reach the model-based matrix $\bar{K}_i$ in a few iterations. Second, select initial values for adaptive parameters $c_i(0) = 1$, $\alpha_i = 1$, and $\beta_i = 1, i \in \mathcal{V}$, respectively, and the simulation results of COR problem are shown in Figures 3–5. More specifically, the state responses of η_i and v are plotted in Figure 3, which shows that η_i tracks the trajectory of the leader v accurately in a few seconds. The evolution of the adaptive parameter c_i is shown in Figure 4(a), and c_i is upper bounded by some positive constant. Moreover, the evolution of the output of agent i, $i \in \mathcal{V}$, the output of exosystem and their tracking errors are displayed in Figures 4(b) and (c), respectively. Furthermore, the control input of each agent is given in Figure 4(d). It demonstrates that the COR of the unknown heterogeneous MAS under the directed graph could be achieved via the proposed fully distributed model-free RL-based adaptive ETC. Moreover, the inter-event time of each agent is given in Figure 5, in which the average inter-event time of each agent are 1.8519, 2.1739, 3.1250, and 1.7241 s, respectively, it indicates that the Zeno behavior is excluded under the proposed ETM.

Next, in order to show the superiority and flexibility of the proposed adaptive ETM of (23), a comparative simulation is conducted between the adaptive ETM and traditional static ETM. The static ETM is given as follows:

$$t_{k+1}^i = \inf\{t > t_k^i | f_i(e_i, \hat{z}_i, c_i, t) \geqslant 0\},$$

in which

$$f_i(e_i, \hat{z}_i, c_i, t) = c_i e_i^{\mathrm{T}} K^{\mathrm{T}} K e_i - \hat{z}_i^{\mathrm{T}} K^{\mathrm{T}} K \hat{z}_i - \beta_i \mathrm{e}^{-\alpha_i t},$$

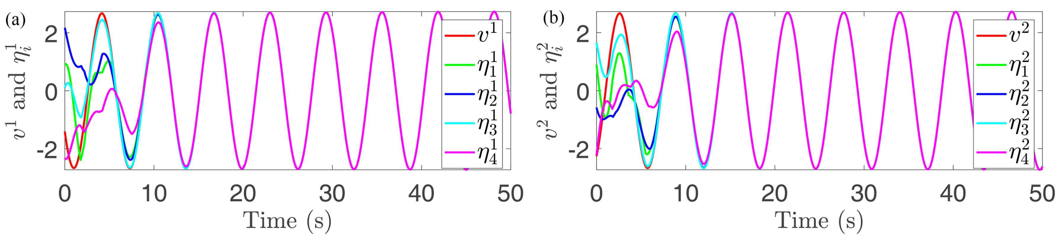

Figure 3 (Color online) Evolution of the state η_i and exosystem v under the proposed adaptive ETM. (a) The first dimension; (b) the second dimension.

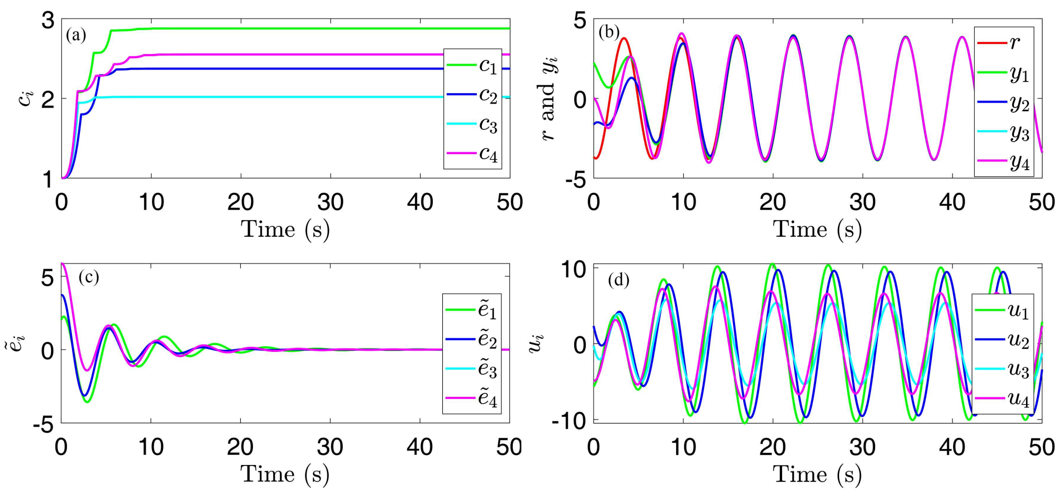

Figure 4 (Color online) (a) Adaptive parameter c_i under the proposed adaptive ETM; (b) output y_i and reference input r under the proposed adaptive ETM; (c) tracking error $\tilde{e}_i$ under the proposed adaptive ETM; (d) control input u_i under the proposed adaptive ETM.

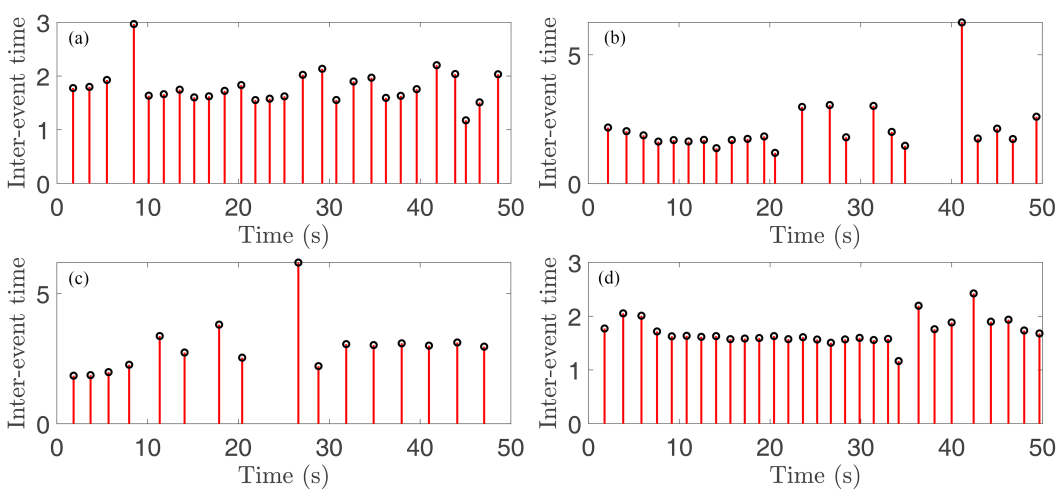

Figure 5 (Color online) Inter-event time of all agents under the proposed adaptive ETM. (a) Agent 1; (b) agent 2; (c) agent 3; (d) agent 4.

and $c_i > 0$ is a predefined constant rather than an adaptive parameter shown in (25). Similarly, the feedback gain matrix $\bar{K}_i(k)$ is also obtained via Algorithm 1. Then, select values $c_i = 10$, $\alpha_i = 1$, and $\beta_i = 1, i \in \mathcal{V}$, under the ETC protocol (21) with static ETM, the simulation results of COR problem are shown in Figures 6–8. The state responses of η_i and v are plotted in Figure 6; the evolution of the output of agent i, $i \in \mathcal{V}$ and exosystem is given in Figure 7(a); the tracking error is displayed in Figure 7(b); the control input of each heterogeneous agent is given in Figure 7(c); and the inter-event time of each agent is given in Figure 8. From Figures 3–5 with adaptive ETM (23) and Figures 6 and 7 with

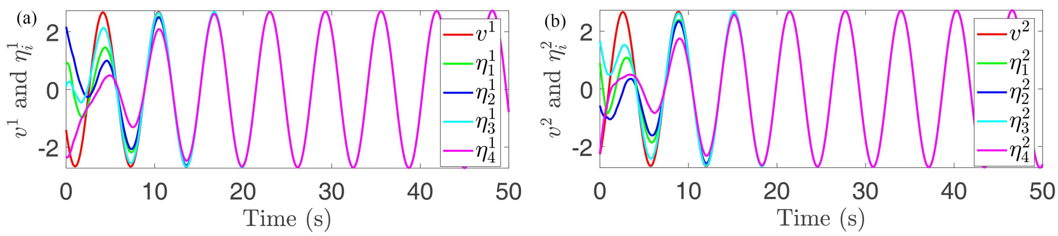

Figure 6 (Color online) Evolution of the state η_i and exosystem v under the traditional static ETM. (a) The first dimension; (b) the second dimension.

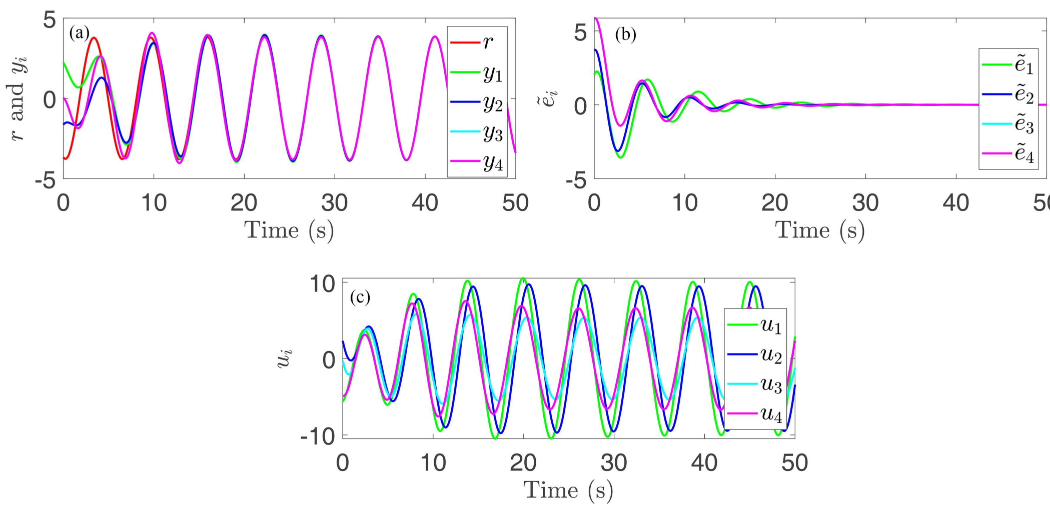

Figure 7 (Color online) (a) Output y_i and reference input r under the traditional static ETM; (b) tracking error $\tilde{e}_i$ under the traditional static ETM; (c) control input u_i under the traditional static ETM.

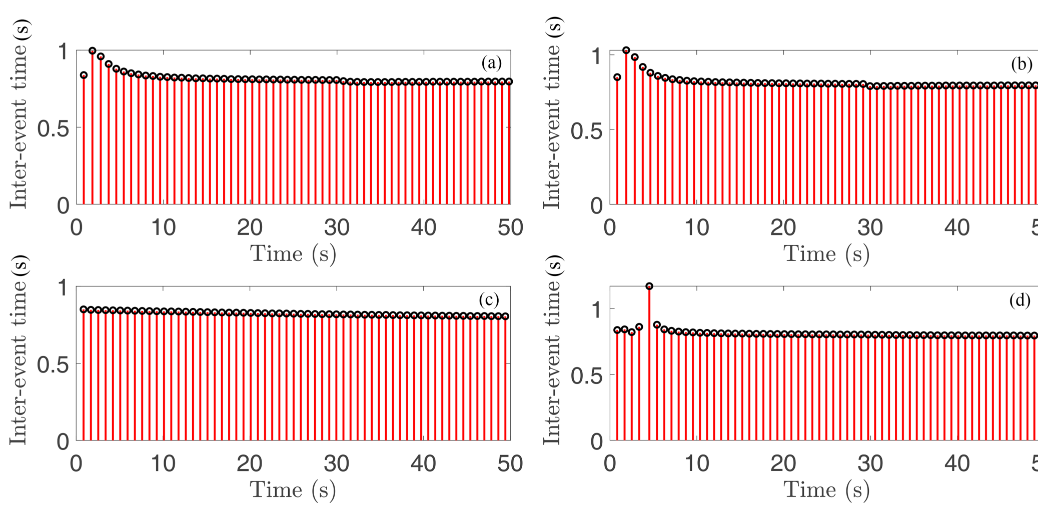

Figure 8 (Color online) Inter-event time of all agents under the traditional static ETM. (a) Agent 1; (b) agent 2; (c) agent 3; (d) agent 4.

static ETM, the COR can be achieved as expected similarly. However, in Figure 8 with static ETM, the average inter-event time of each agent are 0.8197, 0.8197, 0.8333, and 0.8197 s, respectively. The results that list the average inter-event interval for adaptive ETM and static ETM is further given in Table 1, which shows that the inter-event interval of adaptive ETM is obviously greater than that of static ETM. Thus, the proposed adaptive ETC is more flexible and can save more communication resources than the static ETM.

Table 1 Average event-triggered interval.

Agent	Proposed adaptive ETM (s)	Traditional static ETM (s)
1	1.8519	0.8197
2	2.1739	0.8197
3	3.1250	0.8333
4	1.7241	0.8197
Total	2.2187	0.8231

5 Conclusion

In this paper, the COR problem of unknown heterogeneous MASs with directed graphs has been investigated via the model-free RL-based fully distributed ETC method. Firstly, by incorporating the internal model and model-free RL algorithm, the feedback gain matrix is directly learned by solving the AARE via the online input-output data. Then, a distributed observer-based internal model is constructed, and the COR problem of heterogeneous MASs is resolved in a fully distributed event-triggered manner without using any model information, where the requirements of continuous communication and global information have been removed. Moreover, the rigorously theoretical analysis has been provided for the more general directed graph via constructing a graph-based Lyapunov. Finally, the feasibility and effectiveness of the proposed control scheme have been verified by numerical simulations. In the future, we will further consider more general cases that the communication topology is switching, and apply the results to some practical applications, such as robots [45], quadrotors [46], and vehicles [47].

Acknowledgements This work was supported by National Natural Science Foundation of China (Grant Nos. 62303492, 61977019, 62222306), Shenzhen Basic Research Program (Grant Nos. JCYJ20220818102415033, JSGG20201103093802006, KJZD2023092311-4222045), Natural Science Foundation of Hunan Province (Grant No. 2023JJ40765), Natural Science Foundation of Changsha (Grant No. kq2208287), Science and Technology Innovation Program of Hunan Province (Grant No. 2022WZ1001), and China Postdoctoral Innovation Talents Support Program (Grant No. BX20230430).

References

1 Jing G, Wang L. Multiagent flocking with angle-based formation shape control. IEEE Trans Automat Contr, 2020, 65: 817–823
2 Meng Y, Zhang H, Wu A. Leaderless output sign consensus of heterogeneous multi-agent systems over switching signed graphs. Sci China Inf Sci, 2023, 66: 190208
3 Yang Y, Mei J, Ma G. Formation of time-varying multi-AUVs under directed graphs with event-triggered control: an online model-free parameter learning. IEEE Trans Intell Veh, 2024. doi: 10.1109/TIV.2024.3375926
4 Antonio G P, Maria-Dolores C. Multi-agent deep reinforcement learning to manage connected autonomous vehicles at tomorrow's intersections. IEEE Trans Veh Technol, 2022, 71: 7033–7043
5 Yan L, Chen W, Li C, et al. Consensus-based distributed power control in power grids. Sci China Inf Sci, 2020, 63: 149202
6 Bedi G, Venayagamoorthy G K, Singh R, et al. Review of Internet of Things (IoT) in electric power and energy systems. IEEE Int Things J, 2018, 5: 847–870
7 Yang Y, Hu W. Containment control of double-integrator multi-agent systems with time-varying delays. IEEE Trans Netw Sci Eng, 2022, 9: 457–466
8 Zhang S, Zhang Z, Cui R, et al. Distributed consensus protocol for multi-agent differential graphical games. IEEE Trans Circ Syst II, 2024, 71: 211–215
9 Xiao W, Cao L, Li H, et al. Observer-based adaptive consensus control for nonlinear multi-agent systems with time-delay. Sci China Inf Sci, 2020, 63: 132202
10 Deutscher J. Robust cooperative output regulation for a network of parabolic PDE systems. IEEE Trans Automat Contr, 2022, 67: 451–459
11 Wu Y, Hu J, Xiang L, et al. Finite-time output regulation of linear heterogeneous multi-agent systems. IEEE Trans Circ Syst II, 2022, 69: 1248–1252
12 Gao W, Jiang Z P. Adaptive optimal output regulation of time-delay systems via measurement feedback. IEEE Trans Neural Netw Learn Syst, 2018, 30: 938–945
13 Baldi S, Azzollini I A, Ioannou P A. A distributed indirect adaptive approach to cooperative tracking in networks of uncertain single-input single-output systems. IEEE Trans Automat Contr, 2021, 66: 4844–4851
14 Gao W, Jiang Z P, Lewis F L, et al. Leader-to-formation stability of multiagent systems: an adaptive optimal control approach. IEEE Trans Automat Contr, 2018, 63: 3581–3587
15 Zhao J, Yang C, Gao W, et al. Incremental reinforcement learning and optimal output regulation under unmeasurable disturbances. Automatica, 2024, 160: 111468
16 Zhao J, Yang C, Gao W, et al. Adaptive optimal output regulation of interconnected singularly perturbed systems with application to power systems. IEEE CAA J Autom Sin, 2024, 11: 595–607
17 Jiang Y, Gao W, Wu J, et al. Reinforcement learning and cooperative H_∞ output regulation of linear continuous-time multi-agent systems. Automatica, 2023, 148: 110768
18 Gao W, Deng C, Jiang Y, et al. Resilient reinforcement learning and robust output regulation under denial-of-service attacks. Automatica, 2022, 142: 110366
19 Deng C, Gao W, Che W. Distributed adaptive fault-tolerant output regulation of heterogeneous multi-agent systems with coupling uncertainties and actuator faults. IEEE CAA J Autom Sin, 2020, 7: 1098–1106
20 Odekunle A, Gao W, Wang Y. Data-driven global robust optimal output regulation of uncertain partially linear systems. IEEE CAA J Autom Sin, 2019, 6: 1108–1115
21 Gao W, Jiang Z P. Data-driven cooperative output regulation of multi-agent systems under distributed denial of service attacks. Sci China Inf Sci, 2023, 66: 190201
22 Wan Y, Qin J, Ma Q, et al. Multi-agent DRL-based data-driven approach for PEVs charging/discharging scheduling in smart grid. J Franklin Inst, 2022, 359: 1747–1767

23 Gao W, Mynuddin M, Wunsch D C, et al. Reinforcement learning-based cooperative optimal output regulation via distributed adaptive internal model. IEEE Trans Neural Netw Learn Syst, 2022, 33: 5229–5240
24 Jin X, Shi Y, Tang Y, et al. Event-triggered fixed-time attitude consensus with fixed and switching topologies. IEEE Trans Automat Contr, 2022, 67: 4138–4145
25 Zhao H, Yu H, Peng L. Event-triggered distributed data-driven iterative learning bipartite formation control for unknown nonlinear multiagent systems. IEEE Trans Neural Netw Learn Syst, 2024, 35: 417–427
26 Liu S, Sun J, Zhang H, et al. Fully distributed event-driven adaptive consensus of unknown linear systems. IEEE Trans Neural Netw Learn Syst, 2023, 34: 8007–8016
27 Hou Q, Dong J. Fully distributed cooperative fault-tolerant output regulation of linear heterogeneous MASs: a dynamic memory event-triggered distributed observer approach. IEEE Trans Automat Sci Eng, 2024, 21: 4548–4560
28 Zhang H, Li W, Zhang J, et al. Fully distributed dynamic event-triggered bipartite formation tracking for multiagent systems with multiple nonautonomous leaders. IEEE Trans Neural Netw Learn Syst, 2023, 34: 7453–7466
29 Yang Y, Mei J, Wu A G, et al. Fully distributed event-triggered consensus of MIMO MASs with parametric uncertainties and external disturbances. IEEE Trans Syst Man Cybern Syst, 2024, 54: 4294–4304
30 Ahmed Z, Khan M M, Saeed M A, et al. Consensus control of multi-agent systems with input and communication delay: a frequency domain perspective. ISA Trans, 2020, 101: 69–77
31 Liu Y, Wang Z. Output synchronization of multi-agent systems via reinforcement learning. Neurocomputing, 2022, 508: 110–119
32 Li J, Ji L, Li H. Optimal consensus control for unknown second-order multi-agent systems: using model-free reinforcement learning method. Appl Math Comput, 2021, 410: 126451
33 Wang X, Su H. Completely model-free RL-based consensus of continuous-time multi-agent systems. Appl Math Comput, 2020, 382: 125312
34 Huang J. Nonlinear Output Regulation: Theory and Applications. Philadelphia: SIAM, 2004
35 Li S, Chen Y, Liu P X. Distributed fault detection and dynamic event-triggered consensus for heterogeneous multiagent systems under deception attacks. IEEE Trans Circ Syst I, 2023, 70: 3294–3304
36 Li Z, Duan Z. Cooperative Control of Multi-agent Systems: A Consensus Region Approach. Boca Raton: CRC Press, 2017
37 Berman A, Plemmons R J. Nonnegative Matrices in the Mathematical Sciences. Philadelphia: SIAM, 1994
38 Zhang T, Xia M, Yi Y. Adaptive neural dynamic surface control of strict-feedback nonlinear systems with full state constraints and unmodeled dynamics. Automatica, 2017, 81: 232–239
39 Wang S, Huang J. Adaptive leader-following consensus for multiple Euler-Lagrange systems with an uncertain leader system. IEEE Trans Neural Netw Learn Syst, 2019, 30: 2188–2196
40 Yan J, Deng C, Wen C. Resilient output regulation in heterogeneous networked systems under Byzantine agents. Automatica, 2021, 133: 109872
41 Deng C, Wen C, Huang J, et al. Distributed observer-based cooperative control approach for uncertain nonlinear MASs under event-triggered communication. IEEE Trans Automat Contr, 2022, 67: 2669–2676
42 Bian T, Jiang Z P. Value iteration and adaptive dynamic programming for data-driven adaptive optimal control design. Automatica, 2016, 71: 348–360
43 Long M, Su H, Zeng Z. Model-free event-triggered consensus algorithm for multiagent systems using reinforcement learning method. IEEE Trans Syst Man Cybern Syst, 2022, 52: 5212–5221
44 Khalil H K. Nonlinear Systems. Englewood Cliffs: Prentice Hall, 2002
45 Wen G, Duan Z, Chen G, et al. Consensus tracking of multi-agent systems with lipschitz-type node dynamics and switching topologies. IEEE Trans Circ Syst I, 2014, 61: 499–511
46 Wang H, Shan J. Fully distributed event-triggered formation control for multiple quadrotors. IEEE Trans Ind Electron, 2023, 70: 12566–12575
47 Cui L, Wang S, Zhang J, et al. Learning-based balance control of wheel-legged robots. IEEE Robot Autom Lett, 2021, 6: 7667–7674