SCIENCE CHINA
Information Sciences

• **RESEARCH PAPER** •
Special Topic: Entanglement of Control and Optimization

July 2023, Vol. 66 170203:1–170203:12
https://doi.org/10.1007/s11432-022-3729-7

Reinforcement learning-based unknown reference tracking control of HMASs with nonidentical communication delays

Yong XU[1], Zheng-Guang WU[2,3*], Wei-Wei CHE[4] & Deyuan MENG[5]

[1]*School of Automation, Beijing Institute of Technology, Beijing* 100081*, China;*
[2]*Institute of Cyber-Systems and Control, Zhejiang University, Hangzhou* 310027*, China;*
[3]*College of Mathematics and Computer Science, Zhejiang Normal University, Jinhua* 321004*, China;*
[4]*Department of Automation, Qingdao University, Qingdao* 266071*, China;*
[5]*School of Automation Science and Electrical Engineering, Beihang University (BUAA), Beijing* 100191*, China*

Received 14 January 2022/Revised 29 November 2022/Accepted 17 February 2023/Published online 29 June 2023

Abstract This paper focuses on the optimal output synchronization control problem of heterogeneous multiagent systems (HMASs) subject to nonidentical communication delays by a reinforcement learning method. Compared with existing studies assuming that the precise model of the leader is globally or distributively accessible to all or some of the followers, the leader's precise dynamical model is entirely inaccessible to all the followers in this paper. A data-based learning algorithm is first proposed to reconstruct the leader's unknown system matrix online. A distributed predictor subject to communication delays is further devised to estimate the leader's state, where interaction delays are allowed to be nonidentical. Then, a learning-based local controller, together with a discounted performance function, is projected to reach the optimal output synchronization. Bellman equations and game algebraic Riccati equations are constructed to learn the optimal solution by developing a model-based reinforcement learning (RL) algorithm online without solving regulator equations, which is followed by a model-free off-policy RL algorithm to relax the requirement of all agents' dynamics faced by the model-based RL algorithm. The optimal tracking control of HMASs subject to unknown leader dynamics and communication delays is shown to be solvable under the proposed RL algorithms. Finally, the effectiveness of theoretical analysis is verified by numerical simulations.

Keywords heterogeneous multiagent systems, HMAS, reinforcement learning, RL, optimal output synchronization, communication delays

Citation Xu Y, Wu Z-G, Che W-W, et al. Reinforcement learning-based unknown reference tracking control of HMASs with nonidentical communication delays. Sci China Inf Sci, 2023, 66(7): 170203, https://doi.org/10.1007/s11432-022-3729-7

1 Introduction

Considerable attention from different domains has been focused on the cooperative tracking control of multiagent systems (MASs), and its control purpose at urging the reference trajectory generated by the leader can be tracked by a group of following agents by exploiting different distributed control protocols. Many tracking control algorithms are broadly reported in [1–5].

1.1 Related work

Existing results for similar leader-follower synchronization control of MASs with homogeneous dynamics have been widely investigated with different control algorithms. Unfortunately, these developed algorithms are useless for heterogeneous systems. In terms of the output synchronization control of heterogeneous multiagent systems (HMASs), a two-layer framework was first proposed in [6]. Afterward, different cooperative control algorithms were designed. For example, Ref. [7] designed a relative output-based dynamic compensator for the output tracking of HMASs under a distributed controller comprising the

* Corresponding author (email: nashwzhg@zju.edu.cn)

dynamic compensator's state and the local agent. Ref. [8] proposed an adaptive fault-tolerant controller for the output tracking control problem of HMASs. Ref. [9] put forward an adaptive control protocol for a robust tracking consensus of HMASs with disturbance and actuator faults. Ref. [10] discussed the output consensus problem of HMASs subject to nonlinear dynamics using a distributed dynamic compensator. The output tracking control subject to self-tuning parameters through an event-triggered control mechanism was addressed in [11]. Ref. [12] developed a distributed adaptive observer approach for the leader-follower consensus control of MASs, where the function of the adaptive observer is used to estimate the prior knowledge of the exosystem, and the adaptive observer using the relative output information is considered with a similar method in [13]. However, these abovementioned communication protocols among agents are not configurated optimally because the proposed control policy requires obtaining an accurate model in advance. Fortunately, reinforcement learning (RL) [14] is an effective tool for providing a tight link with optimal control, and many optimal consensus results have been reported. The optimal output synchronization of HMASs was considered using the RL algorithm in [15]. An RL-based containment control protocol for MASs with an active leader was discussed in [16]. The optimal leader-follower tracking consensus controller design of MASs with switching topologies using the RL algorithm was investigated in [17]. The RL-based optimal output trajectory following control of HMASs with an unknown exosystem was discussed in [18]. A data-based RL-based optimal output synchronization scheme was proposed for HMASs with event-triggered communication in [19], and so on. Importantly, the abovementioned RL-based studies assume that the leader's system matrix is completely accessible to some or all of the followers in a completely unrestricted communication environment. Moreover, a communication delay is neglected in the abovementioned RL-based optimal tracking control of HMASs.

Communication latency is inevitable in the process of agent information exchange in practical communication environments. The occurrence of communication delay brings great difficulty to coordination control and even causes catastrophic damage to a system's stability. The consensus studies related to communication delays and input delays have been conducted using various techniques, including the frequency domain method in [20] and the Lyapunov method related to Lyapunov-Krasovskii theorem or Razumikhin theorem in [21–23]. The distributed consensus subject to state and input delays has been discussed in [24], which is further addressed to consider unknown transmission delays in [25]. Although different algorithms were proposed in the abovementioned studies, challenging issues remain, such as (1) the precise model of the leader is assumed to be known by some or all followers in the literature. Nevertheless, it seems to be difficult to enable all agents to know the leader's information in practical applications. (2) Existing results consider uniform communication delays for each agent and its neighbors with a single or double integrator [20–23, 26], while nonuniform and time-varying communication delays are considered for the general linear MASs in this article. The existing algorithms cannot be extended to solve the nonuniform (or nonidentical) communication delays. This difficulty is due to the fact that the closed-loop system subject to nonidentical communication delays cannot be written in a tractable compact form. (3) In [6–10, 12–14], the output tracking consensus results mostly rely on solvable regulator equations that require prior knowledge of the leader's system matrix. If the regulator equations are unsolvable because each agent does not know the leader, then output tracking is difficult to achieve. (4) The developed consensus protocols subject to communication delays are not formed in an optimal control fashion. Thus, the integration of the RL-based algorithm and nonidentical communication delays to solve the optimal output tracking control problem is rarely reported and motivates us to conduct this study.

1.2 Main contributions

Based on the abovementioned challenges, this paper employs a learning method to achieve the output state trajectory tracking of HMASs with an unknown leader's model and uncertain but bounded time-varying communication delays. The contributions of this paper are summarized as follows.

- Contrary to most existing results where some or all followers know the leader's prior information, an accurate dynamical model of the reference system is entirely unknown to those tracking agents in this paper. A data-based learning algorithm is proposed to reconstruct the leader's parameter matrix.
- Compared with existing studies [21, 26] in which the communication delays imposed by each agent itself and its neighbors are uniform, the time-varying interaction delays only influencing information about each agent's neighbors are considered for the general HMASs, where the communication delays may be nonidentical for different neighbor agents, which makes the problem more practical and challenging.

• Model-based and model-free learning algorithms are given to guarantee output synchronization without directly solving the regulation equations. Moreover, the model-free case does not rely on the agents' dynamics that are required by the model-based RL algorithm.

The outline of the remainder of this paper is as follows: Preliminaries and a problem statement are presented in Section 2, and Section 3 gives the main results. A numerical example is provided in Section 4, and the conclusion is drawn in Section 5.

Notations. $\mathbb{R}^n$ and $\mathbb{R}^{n\times m}$ denote the set of the real vectors of dimension n and the set of real matrices of dimensions $n \times m$, respectively. $\mathbb{Z}_+$ denotes the set of positive integers. diag$\{*\}$ is a diagonal matrix with nondiagonal elements being zero. $\check{I}_i$ is a square matrix. $\hat{I}_i$ is the i-th row of $\check{I}_i$, and all other rows of $\hat{I}_i$ are zero row vectors. For a matrix A, $\lambda_i^{[A]}$ denotes its eigenvalue, and rank(A) denotes its rank. Define $x|_b^a = x(a) - x(b)$.

2 Preliminaries and the problem statement

Consider a group of N agents with the dynamics of the i-th agent expressed as

$$\begin{aligned} \dot{r}_i(t) &= A_i r_i(t) + B_i u_i(t), \\ y_i(t) &= C_i r_i(t), \end{aligned} \tag{1}$$

where $r_i(t)$ is the state, $u_i(t)$ is the control input, and $y_i(t)$ is the output, respectively. Matrices A_i, B_i, and C_i with different dimensions are known.

To track a reference trajectory (called the leader), its dynamical model is as follows:

$$\begin{aligned} \dot{r}_0(t) &= A_0 r_0(t), \\ y_0(t) &= F r_0(t), \end{aligned} \tag{2}$$

where $r_0(t) \in \mathbb{R}^m$ is the state of the leader, and A_0 is completely inaccessible to all followers.

For the leader-follower systems in (1) and (2), a directed graph $\mathscr{G} = (\mathscr{A}, \mathscr{V}, \mathscr{E})$ comprises the adjacent matrix $\mathscr{A} = [a_{ij}] \in \mathbb{R}^{n\times m}$, the node set $\mathscr{V} = \{v_1, \ldots, v_N\}$, and an edge set $\mathscr{E} \subseteq \mathscr{V} \times \mathscr{V}$, respectively. The adjacent matrix's element $a_{ij} > 0$ if $(v_i, v_j) \in \mathscr{E}$; otherwise, $a_{ij} = 0$. For two distinct nodes v_i and v_j, if v_i can obtain the information of v_j, and there is a communication link. Then define $\mathscr{L} = [\mathscr{L}_{ij}] \in \mathbb{R}^{n\times m}$ as the Laplacian matrix, and its elements satisfies $\mathscr{L}_{ij} = -a_{ij}, i \neq j$ and $\mathscr{L}_{ii} = \sum_{j=1,j\neq i}^{N} a_{ij}$. If there exists a pinning node v_0, the new communication topology graph $\bar{\mathscr{G}}$ is described by $\bar{\mathscr{G}} = \mathscr{G} \cup v_0$, and its element $a_{i0} > 0$ if v_i can directly receive the leader's information, and otherwise $a_{i0} = 0$, and the overall relationship between v_i and v_0 can be described by a diagonal matrix $\mathscr{D} = \mathrm{diag}\{a_{i0}\}$. In terms of v_0, if v_0 and any distinct nodes have an information flow (path), the graph is considered to have a spanning tree in $\bar{\mathscr{G}}$. The link between $\mathscr{L}$ and $\mathscr{D}$ can be given as $\mathscr{H} = \mathscr{L} + \mathscr{D}$, and the characteristic roots of $\mathscr{H}$ are monotonically increasing, namely, $0 < \lambda_1^{[\mathscr{H}]} < \lambda_2^{[\mathscr{H}]} < \cdots < \lambda_N^{[\mathscr{H}]}$.

Problem statement: For the HMAS in (1) and (2), develop a learning-based distributed controller $u_i(t)$ subject to obeying heterogeneous time-varying communication delays.

(1) The matrix A_0 in (2) can be reconstructed with system data by all followers;

(2) The output trajectory of each tracking agent tracks the reference system output trajectory by developing the distributed controller in an optimal way, satisfying

$$\lim_{t\to\infty} \|e_i(t)\| = \lim_{t\to\infty} \|y_i(t) - y_0(t)\| = 0, \quad i = 1, \ldots, N. \tag{3}$$

To achieve the output tracking control, some assumptions in [6, 7] are used as follows.

Assumption 1. The communication topology graph is called to have a spanning tree if the leader at least has a path to all tracking followers.

Assumption 2. The pair (A_i, B_i) is stabilizable.

Assumption 3. All characteristic roots of the matrix A_0 lie in imaginary axis or have no positive parts.

3 Main results

In this section, a data-based learning algorithm is given to reconstruct the matrix A_0 for each follower. Then, a stability analysis of the prediction error between the predictor and the leader subject to non-identical communication delays is discussed.

3.1 Identification of A_0 using a data-based learning algorithm

According to the dynamic of the leader, Eq. (2) can be rewritten in the following form by integrating from $t = t + (\sigma - 1)t$ to $t + \sigma t$ with $\sigma > 0$ being the sampling period as

$$r_0|_{t+(\sigma-1)t}^{t+\sigma t} = A_0 \int_{t+(\sigma-1)t}^{t+\sigma t} r_0^{\mathrm{T}}(\tau)\mathrm{d}\tau, \tag{4}$$

where the system data of (2) are stored at the initial instant with $\sigma = 1$ by introducing the following two vectors denoted by Γ_σ and G_σ to collect system data as follows:

$$\Gamma_\sigma = \mathrm{col}\Big(r_0|_{t+(\sigma-1)t}^{t+\sigma t}\Big), \quad G_\sigma = I_m \otimes \mathrm{col}\left(\int_{t-(\sigma-1)t}^{t+\sigma t} r_0^{\mathrm{T}}(\tau)\mathrm{d}\tau\right), \quad \sigma = 1, 2, \ldots. \tag{5}$$

The use of the Kronecker product to combine (4) and (5) yields

$$\Gamma_\sigma = G_\sigma \cdot \mathrm{vec}(A_0), \tag{6}$$

where $\mathrm{vec}(A_0) = [A_0^{[1]}, A_0^{[2]}, \ldots, A_0^{[m]}]^{\mathrm{T}}$.

Theorem 1. In an HMAS under Assumption 3, assume that the prior knowledge of the leader (2) is unknown to each follower. The leader's matrix A_0 can be identified and reconstructed by satisfying the following conditions. (1) The matrix A_0 can be identified provided that $\mathrm{rank}[G_\sigma] = m^2$; and (2) each follower can compute the identifiable matrix A_0 obeying

$$A_0 = \Big[\, (G_\sigma^{-1}\Gamma_\sigma)_{\sigma\to m} \ \cdots \ (G_\sigma^{-1}\Gamma_\sigma)_{(m^2-m)\to m^2} \,\Big], \tag{7}$$

where $(G_\sigma^{-1}\Gamma_\sigma)_{\sigma\to m}$ can be obtained by using the 1th to mth element of $(G_\sigma^{-1}\Gamma_\sigma)$.

Proof. If Assumption 3 holds, then confirmation that the matrix G_σ exists as a full rank satisfying $\mathrm{rank}[G_\sigma] = m^2$ is not difficult to obtain. Inspired by the idea of the data-driven methods in [27, 28], the matrix A_0 can be identified only when the rank condition is met. Thus, the first part is completed. In contrast, if the rank condition holds, one can conclude that the inverse matrix $(\Gamma_\sigma)^{-1}$ exists, implying that the constructed matrix A_0 can be derived by computing (7).

Remark 1. Similar results suppose that the leader's prior knowledge is accessible to some or all of the followers. In terms of some followers having no prior knowledge of the leader, a combination of the adaptive control and the observer approach is employed in [12, 13] to estimate the leader's system model, respectively. Unlike existing results for similar problems, the most obvious difference is that the leader's prior knowledge is totally unknown to those tracking followers in this article. To address this thorny problem, a data-based learning algorithm and a data-driven control are used to identify and reconstruct A_0 in (7).

3.2 Technical lemmas on nonidentical communication delays

Considering that the occurrence of data transmission among agents is subject to nonidentical communication delays, some technical lemmas are given below to facilitate the stability of prediction error systems.

Definition 1 ([29])**.** A matrix is called a Metzler matrix if its off-diagonal elements are positive or zero.

Definition 2 (δ-digraph [29])**.** Given a Metzler matrix, the δ-digraph ($\delta \geqslant 0$) associated with a Metzler matrix is the digraph related to both the node set and an arc when and only when the entry of a Metzler matrix is greater than δ.

Lemma 1 ([29])**.** For the following linear time-delay system:

$$\dot{x}(t) = \mathrm{diag}(\mathscr{L}(t))x(t) + \bar{\mathscr{L}}(t)x(t-\epsilon), \tag{8}$$

where $\bar{\mathscr{L}}(t) = \mathscr{L}(t) - \mathrm{diag}(\mathscr{L}(t))$ denotes the off-diagonal elements of $\mathscr{L}(t)$. Suppose that the system matrix $\mathscr{L}(t)$ is a bounded and piecewise continuous function of time. If the system matrix $\mathscr{L}(t)$ is Metzler with zero row sums. Under Assumption 1, if the equilibrium set of consensus states is uniformly exponentially stable. All components of any solution of (8) converge to a common value as $t \to \infty$.

Note that Lemma 1 only considers the leaderless consensus subject to communication delays for a single system. In this paper, the general case, including the existence of the leader, and multiple neighboring agents, is considered under heterogeneous communication delays. The extension of Lemma 1 is given in the following Lemma 2 under fixed topology.

Lemma 2. Consider the time-delay systems with the leader described by

$$\dot{\boldsymbol{x}} = -\alpha(\mathrm{diag}(\mathscr{H}) \otimes I_n)\boldsymbol{x} - \alpha \sum_{i=1}^{N}((\check{I}_i\bar{\mathscr{H}}) \otimes I_n)((\mathscr{F}_i \otimes I_n)\boldsymbol{x}), \tag{9}$$

where $\boldsymbol{x} = [x_1^{\mathrm{T}}, \ldots, x_N^{\mathrm{T}}]^{\mathrm{T}}$ is the tracking error of the time-delay systems, $\mathscr{F}_i = \mathrm{diag}\{f_{ij}\}$ with $f_{ii}x(t) = x(t)$ and $f_{ij}x(t) = x(t - \epsilon_{ij}(t))$ for $i \neq j$ for some piecewise continuous $\epsilon_{ij}(t)$ and $0 \leqslant \epsilon_{ij}(t) \leqslant \epsilon$ for some positive constants ϵ. $\bar{\mathscr{H}} = \mathscr{H} - \mathrm{diag}(\mathscr{H})$ with $\mathscr{H} = \mathscr{L} + \mathscr{D}$ defined in graph theory, and $\alpha > 0$ is a constant. Then, the initial point of (9) is exponentially stable.

Proof. According to (9) for a single system, the dynamics of an MAS in the presence of heterogeneous delays, as an extension of (9), is expressed as

$$\dot{x}_i(t) = -\alpha \sum_{j=1}^{N} a_{ij}(x_j(t - \epsilon_{ij}(t)) - x_i(t)) \quad \text{or} \quad \dot{x}(t) = -\alpha \mathrm{diag}(\mathscr{L}(t))x(t) - \alpha \sum_{i=1}^{N} \check{I}_i\bar{\mathscr{L}}(t)(\mathscr{F}_i x(t)). \tag{10}$$

The compact form of multiple time-delay systems (10) is given as follows:

$$\dot{\boldsymbol{x}} = -\alpha(\mathrm{diag}(\mathscr{L}) \otimes I_n)\boldsymbol{x} - \alpha \sum_{i=0}^{N}((\check{I}_{i+1}\bar{\mathscr{L}}) \otimes I_n)[\mathscr{F}_i \otimes I_n]\boldsymbol{x}, \tag{11}$$

where $\boldsymbol{x} = [x_0^{\mathrm{T}}, x_1^{\mathrm{T}}, \ldots, x_N^{\mathrm{T}}]^{\mathrm{T}}$, and $\alpha > 0$ is a constant number. $\bar{\mathscr{L}} = \mathscr{L} - \mathrm{diag}(\mathscr{L})$ with $\mathscr{L}$ being the Laplacian matrix of $\mathscr{L}$ by removing the link $(0, i)$ $i = 1, \ldots, N$. Because the interaction topology is fixed, the δ-graph has the property that every agent $i = 1, \ldots, N$, is reachable from the agent 0 based on Definition 2. Similar to [6, 25], according to Definition 1, $\alpha\mathrm{diag}(\mathscr{L})$ is a Metzler matrix with zero row sums. Then, $\lim_{t\to\infty} x_i(t) = x_0(0), i = 1, \ldots, N$ is easy to conclude based on Lemma 1. Because $\mathscr{H} = \mathscr{L} + \mathscr{D}$ with the diagonal matrix $\mathscr{D}$, the exponential stability of (9) is easily inferred.

3.3 Predictor design subject to communication delays

The distributed predictor with communication delays is designed to estimate the reference system's state as follows:

$$\begin{aligned}
\dot{\eta}_i(t) &= A_0\eta_i(t) + c\sum_{j=0}^{N} a_{ij}[\mathrm{e}^{A_0\epsilon_{ij}(t)}\eta_j(t - \epsilon_{ij}(t)) - \eta_i(t)] \\
&= A_0\theta_i(t) + c\sum_{j=1}^{N} a_{ij}[\mathrm{e}^{A_0\epsilon_{ij}(t)}\eta_j(t - \epsilon_{ij}(t)) - \eta_i(t)] + ca_{i0}[\mathrm{e}^{A_0\epsilon_{i0}(t)}\eta_0(t - \epsilon_{i0}(t)) - \eta_i(t)],
\end{aligned} \tag{12}$$

where $\mathrm{e}^{A_0\epsilon_{i0}(t)}\eta_0(t - \epsilon_{i0}(t)) = r_0(t)$ based on (2). The last term of (12) is associated with the heterogeneous delay between the leader and its neighboring agent.

Remark 2. Existing results consider the simple network structure with one single coupling subject to one single communication delay in [29], while different communication links among agents corresponding to nonidentical delays are considered with multiple network structures and multiple couplings in this paper. Moreover, most existing studies in [20–23, 26] consider that states of adjacent agents are afflicted by uniform communication delays. In our paper, the time-varying interaction delays only influencing information about each agent's neighbors are considered, which are allowed to be nonidentical. Therefore, a consensus analysis with nonidentical communication delays is made more challenging. The difficulty emerges because the nonidentical communication delays for the closed-loop systems cannot be written in a tractable compact form.

Theorem 2. Consider the HMAS in (1) and (2) with a directed communication graph. Under Assumptions 1–3, a predictor based on the learning-based algorithm can predict the state of a leader satisfying $\eta_i(t) \to r_0(t)$ as $t \to \infty$.

Proof. Define the prediction error as $\theta_i(t) = \eta_i(t) - r_0(t)$. The time derivative of θ_i satisfies

$$\begin{aligned}\dot{\theta}_i(t) &= A_0\theta_i(t) + c\sum_{j=1}^{N} a_{ij}[\mathrm{e}^{A_0\epsilon_{ij}(t)}\eta_j(t-\epsilon_{ij}(t)) - \eta_i(t)] + ca_{i0}[\mathrm{e}^{A_0\epsilon_{i0}(t)}\eta_0(t-\epsilon_{i0}(t)) - \eta_i(t)]\\ &= A_0\theta_i(t) + c\sum_{j=0}^{N} a_{ij}\Big[\mathrm{e}^{A_0\epsilon_{ij}(t)}\Big[\theta_j(t-\epsilon_{ij}(t)) + r_0(t-\epsilon_{ij}(t)) - (\theta_j(t) + r_0(t))\Big]\Big]\\ &= A_0\theta_i(t) + c\sum_{j=0}^{N} a_{ij}\Big[\mathrm{e}^{A_0\epsilon_{ij}(t)}\theta_j(t-\epsilon_{ij}(t)) - \theta_j(t)\Big]. \end{aligned} \tag{13}$$

Letting $\tilde{\theta}_i(t) = \mathrm{e}^{-A_0 t}\theta_i(t)$, it follows that

$$\begin{aligned}\dot{\tilde{\theta}}_i(t) &= -A_0\mathrm{e}^{-A_i t}\theta_i(t) + \mathrm{e}^{-A_0 t}\Big[A_0\theta_i(t) + c\sum_{j=0}^{N} a_{ij}\Big[\mathrm{e}^{A_0\epsilon_{ij}(t)}\theta_j(t-\epsilon_{ij}(t)) - \theta_i(t)\Big]\Big]\\ &= c\sum_{j=0}^{N} a_{ij}[\mathrm{e}^{-A_0(t-\epsilon_{ij}(t))}\theta_j(t-\epsilon_{ij}(t)) - \mathrm{e}^{-A_0 t}\theta_i(t)]\\ &= c\sum_{j=0}^{N} a_{ij}[\tilde{\theta}_j(t-\epsilon_{ij}(t)) - \tilde{\theta}_i(t)]. \end{aligned} \tag{14}$$

The above equation can be rewritten in a compact form as

$$\dot{\tilde{\boldsymbol{\theta}}}(t) = -c\Big[\mathrm{diag}(\mathscr{H}) \otimes I_n\Big]\tilde{\boldsymbol{\theta}}(t) - c\sum_{i=1}^{N}\Big[(\mathscr{H} - \mathrm{diag}(\mathscr{H})) \otimes I_n\Big][(\mathscr{F}_i \otimes I_n)\tilde{\boldsymbol{\theta}}(t)], \tag{15}$$

where $\tilde{\boldsymbol{\theta}} = [\tilde{\theta}_1^{\mathrm{T}}, \ldots, \tilde{\theta}_N^{\mathrm{T}}]^{\mathrm{T}}$. Therefore, it frankly knows that $\tilde{\boldsymbol{\theta}}(t) \to 0$ as $t \to \infty$ based on Lemma 2.

Since $\tilde{\theta}_i(t) = \mathrm{e}^{-A_i t}\theta_i(t)$, there exist positive constants γ_1 and γ_2 satisfying

$$\begin{aligned}\lim_{t\to\infty}\|\boldsymbol{\theta}(t)\| &\leqslant \lim_{t\to\infty}\|(I_n \otimes \mathrm{e}^{-A_0 t})\tilde{\boldsymbol{\theta}}(t)\| \leqslant \lim_{t\to\infty}\|I_n \otimes \mathrm{e}^{A_0 t}\|\|\tilde{\boldsymbol{\theta}}(t)\|\\ &\leqslant \lim_{t\to\infty}\|I_n \otimes \mathrm{e}^{A_0 t}\|\Big\|\gamma_1\mathrm{e}^{\gamma_2 t}\sup_{-\epsilon\leqslant\varrho\leqslant 0}\tilde{\boldsymbol{\theta}}(\varrho)\Big\|\\ &\leqslant \lim_{t\to\infty}\gamma_1\mathrm{e}^{\gamma_2 t}\|I_n \otimes \mathrm{e}^{A_0 t}\| \leqslant \lim_{t\to\infty}\gamma_1\mathrm{e}^{\gamma_2 t}P(t) = 0, \end{aligned} \tag{16}$$

where $\|I_n \otimes \mathrm{e}^{A_0 t}\| \leqslant P(t)$ exists because all characteristic roots of the matrix A_0 have no real parts, and there exists a polynominal $P(t)$. Thus, the proof is completed.

Remark 3. Different approaches, such as the frequency domain method and the Lyapunov-Krasovskii theorem or the Razumikhin theorem approach, have been reported to solve uniform time-delay issues for first- or second-order MASs, while the nonuniform and time-varying communication delays are considered for the general linear HMASs in this article. Obviously, the existing methods cannot be directly applied to solve this paper. To solve the time-varying communication delay case, Lemma 2 is introduced and established, which is not only simpler but also less computationally complex compared with constructing multiple Lyapunov functions related to uniform communication delays.

3.4 RL-based optimal output tracking consensus

For agent i, we consider the state feedback controller as follows:

$$u_i = Q_{1i}r_i + Q_{2i}r_0, \quad i = 1, 2, \ldots, N, \tag{17}$$

where $u_i = [Q_{1i}, Q_{2i}]Z_i = Q_iZ_i$, and Q_{1i}, Q_{2i} are gain matrices to be determined later.

Let $Z_i = [r_i^{\mathrm{T}}, r_0^{\mathrm{T}}]^{\mathrm{T}}$. Integrating (1), (2), and (17) yields the following augmented systems:

$$\dot{Z}_i = \begin{bmatrix} A_i & 0 \\ 0 & A_0 \end{bmatrix} Z_i + \begin{bmatrix} B_i \\ 0 \end{bmatrix} u_i = \bar{A}_i Z_i + \bar{B}_i u_i,$$
$$e_i = C_i r_i - F_i r_0 = [C_i, -F] Z_i = \bar{C}_i Z_i. \tag{18}$$

For (18), the value function corresponding to the performance function $J_i(Z_i, u_i)$ is considered as

$$V_i(Z_i, u_i) = \int_t^{\infty} \mathrm{e}^{-\chi_i(s-t)} \Big[Z_i^{\mathrm{T}} \bar{C}_i \Xi_i \bar{C}_i^{\mathrm{T}} Z_i^{\mathrm{T}} + u_i^{\mathrm{T}} \Theta_i u_i \Big] \mathrm{d}s = Z_i^{\mathrm{T}} W_i Z_i, \tag{19}$$

where $\Xi_i \geqslant 0$ and $\Theta_i > 0$ are weighting matrices, $\chi_i > 0$ is a constant number to be designed later, and its function of introducing the parameter χ_i is to ensure the tracking control while satisfying the predefined performance (19). To analyze the stability, define $V_i(Z_i) = Z_i^{\mathrm{T}} W_i Z_i$, where the matrix $W_i \geqslant 0$.

The control objective is to determine

$$V_i^*(u_i) = \min_{u_i} J_i(Z_i, u_i) = \min_{u_i} \int_t^{\infty} \mathrm{e}^{-\chi_i(s-t)} \Big[Z_i^{\mathrm{T}} \bar{C}_i \Xi_i \bar{C}_i^{\mathrm{T}} Z_i^{\mathrm{T}} + u_i^{\mathrm{T}} \Theta_i u_i \Big] \mathrm{d}s. \tag{20}$$

According to (19), the Bellman equation is easy to conclude as follows:

$$H(Z_i, u_i) = Z_i^{\mathrm{T}} \bar{C}_i \Xi_i \bar{C}_i^{\mathrm{T}} Z_i^{\mathrm{T}} + u_i^{\mathrm{T}} \Theta_i u_i + 2 Z_i^{\mathrm{T}} W_i [\bar{A}_i Z_i + \bar{B}_i u_i] - \chi_i V_i = 0. \tag{21}$$

Based on the stationary condition $\frac{\partial H(Z_i, u_i)}{\partial u_i} = 0$, the optimal control policy is as follows:

$$u_i^* = Q_i Z_i = -\Theta_i^{-1} \bar{B}_i^{\mathrm{T}} W_i Z_i, \tag{22}$$

where $W_i = W_i^{\mathrm{T}} = \left[\begin{smallmatrix} W_i^{11} & W_i^{12} \\ W_i^{21} & W_i^{22} \end{smallmatrix}\right]$ satisfies the following equation:

$$\bar{A}_i^{\mathrm{T}} W_i + W_i \bar{A}_i - \chi_i W_i + \bar{C}_i^{\mathrm{T}} \Xi_i \bar{C}_i - W_i \bar{B}_i \Theta_i^{-1} \bar{B}_i^{\mathrm{T}} W_i = 0. \tag{23}$$

Note that the solution W_i of (23) will be learned based on Bellman equations (18), and the detailed learning process is given in Algorithm 1.

Algorithm 1 Model-based RL algorithm

1. Start with a stabilizing control policy $u_i^{[k]}$ with $k = 0$ for i;
2. Solve the following Bellman equation for $V_i^{[k+1]}$:

$$Z_i^{\mathrm{T}} \bar{C}_i \Xi_i \bar{C}_i^{\mathrm{T}} Z_i^{\mathrm{T}} + (u^{[k]})_i^{\mathrm{T}} \Theta_i u_i^{[k]} + \Big(\frac{\partial V_i(Z_i, u_i)}{\partial Z_i} \Big)^{[k+1]} [\bar{A}_i Z_i + \bar{B}_i u_i^{[k]}] - \chi_i V_i = 0, \tag{24}$$

where $\partial V_i(Z_i, u_i)/\partial Z_i = 2 W_i Z_i$;

3. Find improved control policies using u_i and the optimal controller gain Q_i,

$$u_i^{[k+1]} = -\Theta_i^{-1} \bar{B}_i^{\mathrm{T}} W_i^{[k]} Z_i, \quad Q_i^{[k+1]} = -\Theta_i^{-1} \bar{B}_i^{\mathrm{T}} W_i^{[k]}; \tag{25}$$

4. Stop when $\|Q_i^{(k+1)} - Q_i^{(k)}\| \leqslant \phi$, where ϕ is a small constant; otherwise, $k = k + 1$ and return to Step 2 until the convergence criterion is satisfied.

Theorem 3. Consider an HMAS composed of (1) and (2) under Assumptions 1–3, developing the RL-based local controller u_i in (1) with $Q_i = -\Theta_i^{-1} \bar{B}_i^{\mathrm{T}} W_i$. Then, $A_i + B_i Q_{1i}$ is Hurwitz, and the optimal output tracking consensus of the HMASs can be obtained if $\chi_i = \chi_i^* \leqslant 2\|(C_i^{\mathrm{T}} \Xi_i C_i B_i \Theta_i^{-1} B_i^{\mathrm{T}})^{\frac{1}{2}}\|$.

Proof. According to (23), substituting W_i into (23) yields

$$A_i^{\mathrm{T}} W_i^{11} + W_i^{11} A_i - \chi_i W_i^{11} + C_i^{\mathrm{T}} \Xi_i C_i - W_i^{11} B_i \Theta_i^{-1} B_i^{\mathrm{T}} W_i^{11} = 0, \tag{26}$$

which can be rewritten as

$$(A_i + B_i Q_{1i}) + (A_i + B_i Q_{1i}) - \chi_i I_n = -C_i^{\mathrm{T}} \Xi_i C_i (W_i^{11})^{-1} - W_i^{11} B_i \Theta_i^{-1} B_i^{\mathrm{T}}$$

$$= -\left\|(C_i^{\mathrm{T}})^{\frac{1}{2}}\Xi_i^{\frac{1}{2}}C_i^{\frac{1}{2}}(W_i^{11})^{-\frac{1}{2}}\right\|^2 - \left\|(W_i^{11})^{\frac{1}{2}}B_i^{\frac{1}{2}}\Theta_i^{-\frac{1}{2}}(B_i^{\mathrm{T}})^{\frac{1}{2}}\right\|^2$$
$$\leqslant -2\left\|(C_i^{\mathrm{T}}\Xi_i C_i B_i \Theta_i^{-1}B_i^{\mathrm{T}})^{\frac{1}{2}}\right\| < 0, \tag{27}$$

where the eigenvalues of the matrix $(A_i + B_iQ_{1i}) + (A_i + B_iQ_{1i}) - \chi_i I_n$ can be described by $2\lambda(A_i + B_iQ_{1i}) - \chi_i$. Thus, $2\lambda(A_i + B_iQ_{1i}) - \chi_i \leqslant -2\|(C_i^{\mathrm{T}}\Xi_i C_i B_i\Theta_i^{-1}B_i^{\mathrm{T}})^{\frac{1}{2}}\|$. Note that if the discounted factor satisfies $\chi_i \leqslant 2\|(C_i^{\mathrm{T}}\Xi_i C_i B_i\Theta_i B_i^{\mathrm{T}})^{\frac{1}{2}}\|$, it concludes $\lambda(A_i + B_iQ_{1i}) < 0$, and then $A_i + B_iQ_{1i}$ is Hurwitz.

On the other hand, multiplying the left-and right-hand sides of (23) by Z_i^{T} and Z_i, we have

$$2Z_i^{\mathrm{T}}\bar{A}_i^{\mathrm{T}}W_iZ_i - \chi_i Z_i^{\mathrm{T}}W_iZ_i + Z_i^{\mathrm{T}}\bar{C}_i^{\mathrm{T}}\Xi_i\bar{C}_iZ_i - (Z_i^{\mathrm{T}}W_i)\bar{B}_i\Theta_i^{-1}\bar{B}_i^{\mathrm{T}}(W_iZ_i^{\mathrm{T}}) = 0. \tag{28}$$

It can be observed from (28) that if $W_iZ_i = 0$, then $Z_i^{\mathrm{T}}\bar{C}_i^{\mathrm{T}}\Xi_i\bar{C}_iZ_i = 0$. Thus, $(y_i - y_0)^{\mathrm{T}}\Xi_i(y_i - y_0) = Z_i^{\mathrm{T}}\bar{C}_i^{\mathrm{T}}\Xi_i\bar{C}_iZ_i = 0$ is easily confirmed, and the output tracking consensus $e_i = 0$. If $W_iZ_i \neq 0$, one has $\dot{V}(Z_i, u_i) = 2Z_i^{\mathrm{T}}W_i(\bar{A}_i + \bar{B}_iQ_i)Z_i = Z_i^{\mathrm{T}}(W_i\tilde{A}_i + \tilde{A}_iW_i)Z_i$, where $\tilde{A}_i = \left[\begin{smallmatrix} A_i + B_iQ_{1i} & B_iQ_{2i} \\ 0 & A_0 \end{smallmatrix}\right]$ is Hurwitz because $A_i + B_iQ_{1i}$ is Hurwitz and all eigenvalues of A_0 satisfy Assumption 3. Thus, a matrix $S \geqslant 0$ satisfying $W_i\tilde{A}_i + \tilde{A}_iW_i + S \leqslant 0$ must exist. Then, one has $\dot{V}(Z_i, u_i) = -Z_i^{\mathrm{T}}SZ_i \leqslant 0$. According to LaSalle's invariance principle, Z_i synchronizes to the largest invariant subspace where $\dot{V}(Z_i, u_i) = 0$. Then, it is easy to check that $\dot{V}(Z_i, u_i) = 0$ when $W_iZ_i = 0$. The optimal output tracking control problem of HMASs is solved by using a policy iteration algorithm as shown in Algorithm 1.

Remark 4. It is worth noting that Theorem 2 is a model-based control algorithm because the obtained optimal control policy relies on the accurate matrix $\bar{B}_i$ in (22). If the dynamical model is unknown, the model-based control algorithm is useless. To relax this condition, inspired by [30], we will develop a model-free learning algorithm.

The augmented systems (18) is rewritten as

$$\dot{Z}_i = (\bar{A}_i + \bar{B}_iQ_i^{[k]})Z_i + \bar{B}_i(u_i - Q_i^{[k]}Z_i). \tag{29}$$

Based on Algorithm 1, $\dot{V}_i(Z_i)$ satisfies

$$\begin{aligned}\dot{V}_i &= \left(\frac{\mathrm{d}V_i}{\mathrm{d}Z_i}\right)\left[(\bar{A}_i + \bar{B}_iQ_i^{[k]})Z_i\right] + \left(\frac{\mathrm{d}V_i}{\mathrm{d}Z_i}\right)\left[\bar{B}_i(u_i - Q_i^{[l]}Z_i)\right] \\ &= \chi_iV_i - Z_i^{\mathrm{T}}\left[\bar{C}_i^{\mathrm{T}}\Xi_i\bar{C}_i + \left(Q_i^{[k]}\right)^{\mathrm{T}}\Theta_iQ_i^{[k]}\right]Z_i - 2\left(u_i - Q_i^{[k]}Z_i\right)\Theta_iQ_i^{[k+1]}Z_i.\end{aligned} \tag{30}$$

For the interval $[t, t+T]$, the following IRL Bellman equation can be derived by imposing $\mathrm{e}^{-\chi_it}$ on both sides of (30) as

$$\begin{aligned}\mathrm{e}^{-\chi_iT'}V_i^{[k+1]}(Z_i(t+T')) - V_i^{[k+1]}(Z_i(t)) = &-\int_t^{t+T'}\mathrm{e}^{-\chi_i(s-t)}Z_i^{\mathrm{T}}\left[\bar{C}_i^{\mathrm{T}}\Xi_i\bar{C}_i + \left(Q_i^{[k]}\right)^{\mathrm{T}}\Theta_iQ_i^{[k]}\right]Z_i\mathrm{d}s \\ &- 2\int_t^{t+T'}\mathrm{e}^{-\chi_i(s-t)}\left(u_i - Q_i^{[k]}Z_i\right)^{\mathrm{T}}\Theta_iQ_i^{[k+1]}Z_i\mathrm{d}s.\end{aligned} \tag{31}$$

Note that for (31), $\bar{C}_i^{\mathrm{T}}\Xi_i\bar{C}_i$ is associated with the matrices C_i and F in (31), which means that Algorithm 2 is a model-free algorithm. In fact, the prior knowledge related to C_i and F is not required in Algorithm 2 mainly because we can use the relative output information $(y_i - y_0)^{\mathrm{T}}\Xi_i(y_i - y_0)$ to relax the requirement of $\bar{C}_i^{\mathrm{T}}\Xi_i\bar{C}_i$ in (31). That is, it is a model-free algorithm without any prior knowledge. The model-free RL algorithm as shown in Algorithm 2 is given to achieve the optimal output synchronization without relying on system dynamics.

Remark 5. Note that Algorithm 1 has the same solution as Algorithm 2, and the convergence of Algorithm 1 can be ensured considering [31]. That is, the convergence of Algorithm 2 is guaranteed.

Remark 6. The controller gains of existing results [16–18] are derived by resolving a series of linear matrix inequalities (LMIs), and the analytical solution of multiple LMIs is difficult to ensure. However, in this paper, the optimal control gains of (17) can be obtained by simultaneously using an RL algorithm.

Algorithm 2 Model-free off-policy IRL algorithm

1. Begin with a stabilizing control policy $u_i^{[k]}$ with $k = 0$ for i;
2. Solve the integral Bellman equation (31) for $V_i^{[k+1]}$ and $u_i^{[k+1]}$ simultaneously;
3. Stop when $\|V_i^{[k+1]} - V_i^{[k]}\| \leqslant \phi$ where ϕ is a small constant; otherwise, $k = k + 1$ and return to Step 2 until the convergence criterion is satisfied.

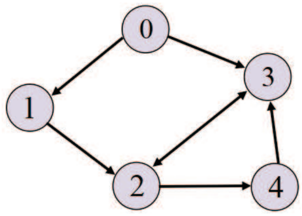

Figure 1 (Color online) Communication structure among five agents.

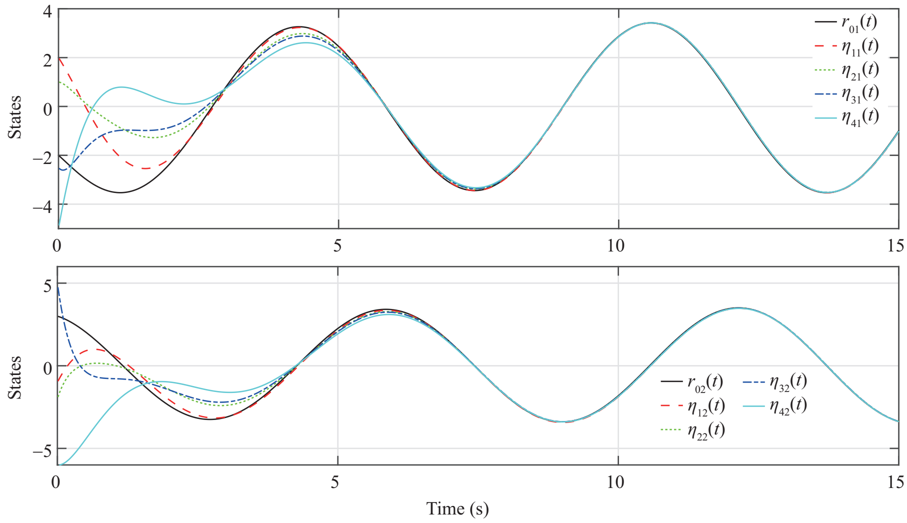

Figure 2 (Color online) Trajectories of states with nonidentical delays.

4 Numerical simulations

Consider an HMAS comprising a leader and four tracking agents with the following system parameters:

$$A_0 = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}, \quad F = \begin{bmatrix} 2 \\ 0 \end{bmatrix}^{\mathrm{T}}, \quad A_i = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}, \quad B_i = \begin{bmatrix} i \\ 0 \end{bmatrix}, \quad C_i = \begin{bmatrix} 0 \\ i \end{bmatrix}^{\mathrm{T}}, \quad i = 1, \ldots, 4.$$

The communication topology structure among five agents is shown in Figure 1, and Assumptions 1–3 hold. The initial values of the leader and predictors are selected as $r_0(0) = [-2, 3]^{\mathrm{T}}$, $\eta_1(0) = [2, -1]^{\mathrm{T}}$, $\eta_2(0) = [1, -2]^{\mathrm{T}}$, $\eta_3(0) = [-2.5, 5]^{\mathrm{T}}$, and $\eta_4(0) = [-5, -6]^{\mathrm{T}}$. Meanwhile, suppose that the time-varying communication delays are described by $\epsilon_{ij}(t) = \frac{1}{3t+i}s$, $i = 1, 2, 3, 4$ for different neighboring agents. Using the predictor in (11), the state synchronization between predictors and the leader converges to a common state value, as shown in Figure 2, and the curves of corresponding prediction errors are depicted in Figure 3. That is, the predictor can estimate the leader's state although the different communication delays are imposed on different links. When the communication delays $\epsilon_{ij}(t) = \frac{1}{3t}s$ for different links are the same case, the evolutions of state trajectories and the prediction errors are as presented in Figures 4 and 5. Now, we implement the RL algorithm for output tracking consensus, and the parameters $\chi_i = 2$, $\Xi_i = 2$, $\Theta_i = 10$ are considered in discounted performance function (19). Based on the given system matrices, solving the algebraic Riccati equation (23) derives the optimal controller gains for four followers

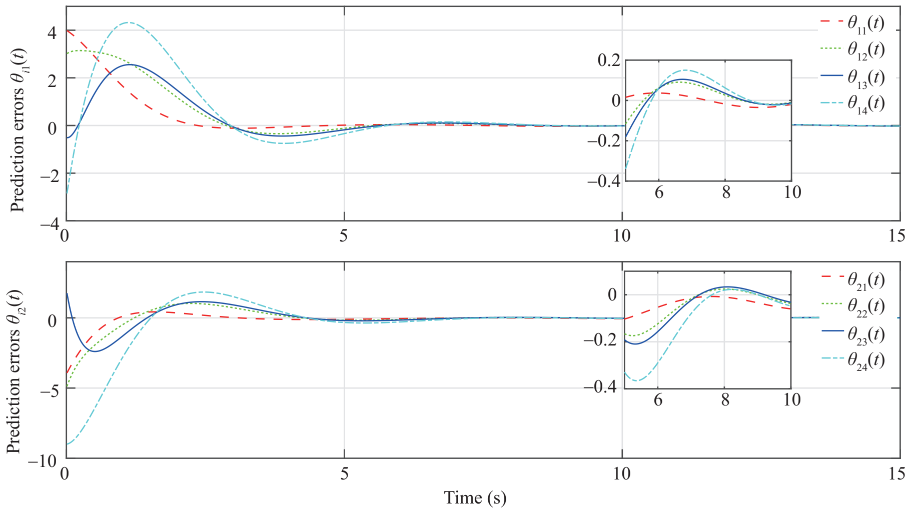

Figure 3 (Color online) Trajectories of θ_i with nonidentical delays.

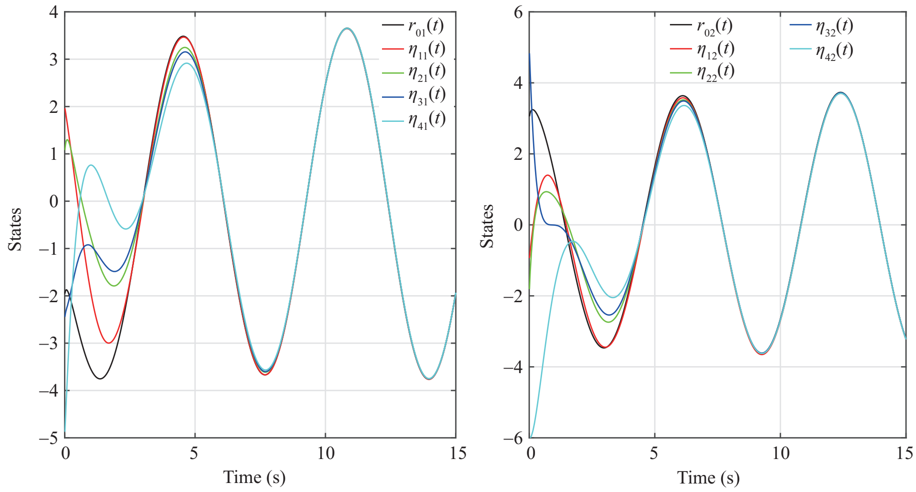

Figure 4 (Color online) Trajectories of states with uniform delays.

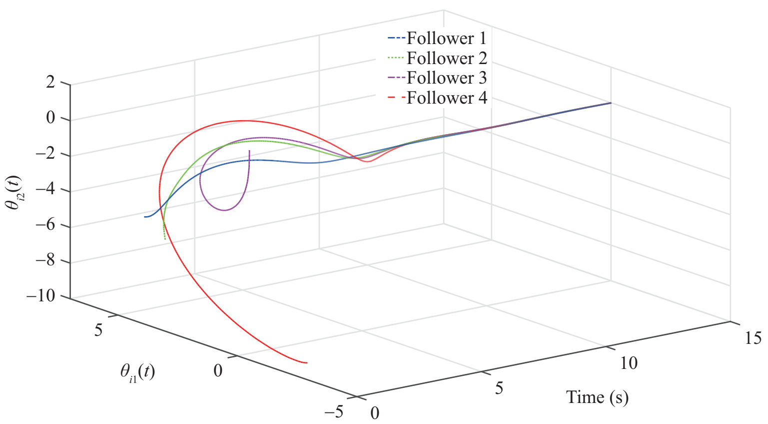

Figure 5 (Color online) Trajectories of θ_i with uniform delays.

as follows:

$$Q_1^* = \begin{bmatrix} -0.0736 & -0.1006 & -0.2012 & -0.1471 \end{bmatrix}, \quad Q_2^* = \begin{bmatrix} -0.5822 & -0.6301 & -0.8378 & -0.3176 \end{bmatrix},$$
$$Q_3^* = \begin{bmatrix} -0.9889 & -1.3077 & -1.0912 & -0.2287 \end{bmatrix}, \quad Q_4^* = \begin{bmatrix} -1.3021 & -2.0211 & -1.2019 & -0.1658 \end{bmatrix}.$$

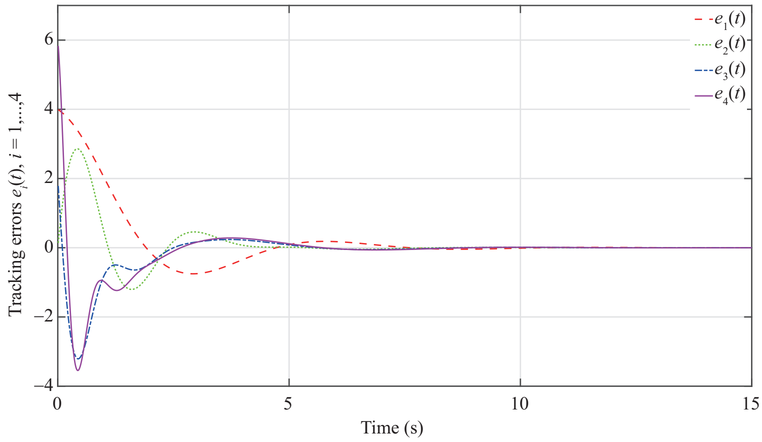

Figure 6 (Color online) Trajectories of tracking errors $e_i(t)$ with nonidentical delays.

Based on the optimal controller gains in (17), the output tracking consensus can be achieved, and the evolution of the tracking errors $e_i(t), i = 1, 2, 3, 4$ is illustrated in Figure 6.

5 Conclusion

The optimal output synchronization control of HMASs subject to heterogeneous delays and an unknown reference model was studied in this paper. A data-based learning algorithm is proposed to learn the leader's model for each follower. The leader state predictor subject to nonidentical communication delays is further given to estimate the leader's state for each follower. Then, the optimal controller-based RL algorithm is developed to ensure the control objective, which is extended to the model-free RL algorithm that does not depend on system dynamics. Finally, a simulation example is provided to verify the effectiveness of the theoretical analysis. Note that because the obtained results ignore the existence of limited bandwidth, different resilient event-triggered strategies [32–36] will be exploited for the optimal secure control of nonlinear HMASs in our future work. Furthermore, the proposed algorithms begin with an admissible initial control policy, which is associated with the system dynamics matrix, and how to remove this assumption also merits consideration in our future study.

Acknowledgements This work was supported by National Natural Science Foundation of China (Grant Nos. 62103047, U1966202), Beijing Institute of Technology Research Fund Program for Young Scholars, and Young Elite Scientists Sponsorship Program by BAST (Grant No. BYESS2023365)

References

1 Ni W, Cheng D. Leader-following consensus of multi-agent systems under fixed and switching topologies. Syst Control Lett, 2010, 59: 209–217
2 Yu W W, Wang H, Hong H F, et al. Distributed cooperative anti-disturbance control of multi-agent systems: an overview. Sci China Inf Sci, 2017, 60: 110202
3 Liu T F, Qi J, Jiang Z P. Distributed containment control of multi-agent systems with velocity and acceleration saturations. Automatica, 2020, 117: 108992
4 Xu Y, Fang M, Shi P, et al. Multileader multiagent systems containment control with event-triggering. IEEE Trans Syst Man Cybern Syst, 2021, 51: 1642–1651
5 Liu T F, Qin Z, Jiang Z P. A new look at distributed optimal output agreement of multi-agent systems. Automatica, 2022, 136: 110053
6 Su Y F, Huang J. Cooperative output regulation of linear multi-agent systems. IEEE Trans Automat Contr, 2012, 57: 1062–1066
7 Lu M, Liu L. Cooperative output regulation of linear multi-agent systems by a novel distributed dynamic compensator. IEEE Trans Automat Contr, 2017, 62: 6481–6488
8 Deng C, Yang G H. Distributed adaptive fault-tolerant control approach to cooperative output regulation for linear multi-agent systems. Automatica, 2019, 103: 62–68
9 Dong S L, Chen G R, Liu M Q, et al. Cooperative neural-adaptive fault-tolerant output regulation for heterogeneous nonlinear uncertain multiagent systems with disturbance. Sci China Inf Sci, 2021, 64: 172212
10 Li G Q, Wang L. Adaptive output consensus of heterogeneous nonlinear multiagent systems: a distributed dynamic compensator approach. IEEE Trans Automat Contr, 2023, 68: 2483–2489
11 Xu Y, Wu Z G. Distributed adaptive event-triggered fault-tolerant synchronization for multiagent systems. IEEE Trans Ind Electron, 2021, 68: 1537–1547
12 Cai H, Lewis F L, Hu G, et al. The adaptive distributed observer approach to the cooperative output regulation of linear multi-agent systems. Automatica, 2017, 75: 299–305

13 Cai H, Huang J. Output based adaptive distributed output observer for leader-follower multiagent systems. Automatica, 2021, 125: 109413
14 Sutton R S, Barto A G. Introduction to Reinforcement Learning. Cambridge: MIT Press, 1998
15 Modares H, Nageshrao S P, Lopes G A D, et al. Optimal model-free output synchronization of heterogeneous systems using off-policy reinforcement learning. Automatica, 2016, 71: 334–341
16 Yang Y, Modares H, Wunsch D C, et al. Leader-follower output synchronization of linear heterogeneous systems with active leader using reinforcement learning. IEEE Trans Neural Netw Learn Syst, 2018, 29: 2139–2153
17 Mu C, Zhao Q, Sun C. Optimal tracking control of heterogeneous MASs using event-driven adaptive observer and reinforcement learning. IEEE Trans Neural Netw Learn Syst, 2022. doi: 10.1109/TNNLS.2022.3208237
18 Xu Y, Wu Z G. Data-efficient off-policy learning for distributed optimal tracking control of HMAS with unidentified exosystem dynamics. IEEE Trans Neural Netw Learn Syst, 2022. doi: 10.1109/TNNLS.2022.3172130
19 Li Q, Xia L, Song R. Leader-follower bipartite output synchronization on signed digraphs under adversarial factors via data-based reinforcement learning. IEEE Trans Neural Netw Learn Syst, 2020, 31: 4185–4195
20 Olfati-Saber R, Murray R M. Consensus problems in networks of agents with switching topology and time-delays. IEEE Trans Automat Contr, 2004, 49: 1520–1533
21 Zhou B, Lin Z. Consensus of high-order multi-agent systems with large input and communication delays. Automatica, 2014, 50: 452–464
22 Yang X, Zhou B. Consensus of discrete-time multiagent systems with input delays by truncated pseudo-predictor feedback. IEEE Trans Cybern, 2017, 49: 505–516
23 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, 2017, 65: 3417–3426
24 Lu M, Liu L. Consensus of linear multi-agent systems subject to communication delays and switching networks. Int J Robust Nonlinear Control, 2017, 27: 1379–1396
25 Lu M, Liu L. Distributed feedforward approach to cooperative output regulation subject to communication delays and switching networks. IEEE Trans Automat Contr, 2017, 62: 1999–2005
26 Yu J, Wang L. Group consensus in multi-agent systems with switching topologies and communication delays. Syst Control Lett, 2010, 59: 340–348
27 Jiang Y, Jiang Z P. Robust adaptive dynamic programming for large-scale systems with an application to multimachine power systems. IEEE Trans Circuits Syst II, 2012, 59: 693–697
28 Wang X, Sun J, Wang G, et al. Data-driven control of distributed event-triggered network systems. IEEE CAA J Autom Sin, 2023, 10: 351–364
29 Moreau L. Stability of continuous-time distributed consensus algorithms. In: Proceedings of the 43rd IEEE Conference on Decision and Control (CDC), 2004. 4: 3998–4003
30 Modares H, Lewis F L, Jiang Z P. H_∞ tracking control of completely unknown continuous-time systems via off-policy reinforcement learning. IEEE Trans Neural Netw Learn Syst, 2015, 26: 2550–2562
31 Vamvoudakis K G, Lewis F L, Hudas G R. Multi-agent differential graphical games: online adaptive learning solution for synchronization with optimality. Automatica, 2012, 48: 1598–1611
32 Liu W, Sun J, Wang G, et al. Data-driven resilient predictive control under denial-of-service. IEEE Trans Automat Contr, 2022. doi: 10.1109/TAC.2022.3209399
33 Xu Y, Sun J, Pan Y J, et al. Dynamic deadband event-triggered strategy for distributed adaptive consensus control with applications to circuit systems. IEEE Trans Circuits Syst I, 2022, 69: 4663–4673
34 Zhang P, Liu T F, Jiang Z P. Event-triggered stabilization of a class of nonlinear time-delay systems. IEEE Trans Automat Contr, 2021, 66: 421–428
35 Xu Y. Resilient secure control of networked systems over unreliable communication networks. IEEE Trans Ind Inf, 2022, 18: 4069–4077
36 Tao Y Y, Wu Z G. Asynchronous control of two-dimensional Markov jump Roesser systems: an event-triggering strategy. IEEE Trans Netw Sci Eng, 2022, 9: 2278–2289