Model-Free Optimal Output Cluster Synchronization Control for Multiagent Systems

Zongsheng Huang
School of Automation Engineering
University of Electronic Science and Technology of China
Chengdu 611731, China
zs__Huang@163.com

***Abstract*—In this article, the model-free optimal output cluster synchronization control problem is investigated for nonlinear multiagent systems (MASs). First, in view of the unknown output of leader, relying on practical prescribed-time performance function, an observer is designed for each follower to estimate the output of leader, and can achieve the desired accuracy within prescribed time. Then, based on the designed observer, an augmented system consisting of observer dynamics and follower dynamics is constructed and the cost functin is built for each follower. Accordingly, the optimal output cluster synchronization control problem is transformed into a numerical solution to solve the Hamilton-Jacobian-Bellman equation (HJBE). Subsequently, the off-policy reinforcement learning (RL) algorithm is addressed to learn the solution to HJBE without any knowledge of the system dynamics. Meanwhile, to release computational burden, the single critic neural network (NN) framework is employed to implement the algorithm, where the least square method is used for training the NN weights. Thus, the designed control algorithm can minimize the cost functions and ensure the output cluster synchronization of MASs with the unknown system dynamics and unavailable leader output. Finally, the simulation examples confirm the validity of the designed control scheme.**

***Index Terms*—Prescribed-time observer, optimal synchronization, reinforcement learning, nonlinear multiagent systems.**

I. Introduction

Multiagent systems (MASs) are autonomous sensing systems containing multiple independent agents. Each agent shares information and collaborates with other agents to accomplish tasks via the communication network. Thanks to the robustness, extendibility and the ability to handle complex tasks, the synchronization control problem of MASs has received broad attention over the past decades, such as multiple vehicles [1] and multiple spacecrafts [2]. Due to the requirement of multiple tasks, MASs are partitioned into a few subgroups to accomplish different but cooperative tasks. Hence, the agents in the same cluster reach consistent goals, while the agents in the different clusters fulfil tasks variously, which is named cluster synchronization or group consensus [3]. It was proposed in [3] for the first time and extended to switching topologies and cooperation-competition network in [4] and [5] later, respectively. Recently, many scholars have made efforts for it. In [6], a relative-output-based distributed control strategy was designed for MASs to achieve cluster synchronization. The cluster synchronization control problem via pinning control strategies was investigated for heterogeneous MASs in [7]. It should be acknowledged that in many performance-critical fields, such as missile guidance, MASs rendezvous, emergency braking and obstacle avoidance, achieving prescribed accuracy in given time is significant.

Fortunately, the prescribed-time control (PTC) was firstly proposed by Song *et al.* to achieve the above target [8]. The PTC derives from the classical idea of strategic and tactical missile guidance application [9], inherited the characteristics of finite-time control and fixed-time control. PTC came into being, which was independent on the initial values of the system and can achieve accurate prescribed settling time. Lots of controllers can be integrated with PTC algorithms that extend the application of PTC. Relying on [8], in [10], it has proved that the regular Lyapunov inequality was utilized for stability analysis and thus avoided the difficulty in stability analysis brought by PTC. Later, in [11], the convergence rate can be preset as desired and a universal method for constructing the time-varying rate function was given which can be regarded as an extension of [8], [10]. In [12], the PTC was continued for observer with canonical form. In the above results, the size of the steady-state error depends on unknown parameters and the control accuracy after the settling time was uncertain, which was undesirable in practice. Therefore, the practical prescribed-time control was presented, where the settling time and control accuracy were directly assigned in advance by the designers [13]. In [13], a new time-varying constraint function was designed to ensure that the system can operate even after the prescribed time and the global result was obtained. In [14], the practical PTC-based time-scale performance function was constructed to handle the case of arbitrary initial values and established the global consensus.

As a popular control method, optimal control has attracted much attention. As we all know, for linear systems, with the aid of dynamic programming method, the optimal strategy is calculated associated with Ricatti equation. For nonlinear systems, the optimal solution can be produced related to Hamilton-Jacobian-Bellman equation (HJBE). Nevertheless, the HJBE can not be obtained via numerical methods. To break through this limitation, adaptive dynamic programming (ADP) method was firstly developed by Werbos [15]. The basic idea of ADP is to approximate the solution of HJBE by using the function approximation structure according to the optimality principle. Murray *et al.* [16] firstly proposed the value iteration

(VI) algorithm for continuous systems and proved the convergence of algorithm, which was viewed as a major improvement in ADP area. On the basis of [16], in [17], the policy iteration (PI) algorithm was designed to approximate optimal saturation controller for nonlinear systems with saturation constraints. VI and PI are two important branches of ADP area, and also have important significance in output synchronization optimization for MASs, which mainly focuses on minimizing the cost function consisting of local output error between the follower and the leader. Currently, on the basis of PI/VI, many results associated with optimal output synchronization for MASs have been reported, e.g., see [18] and the references therein. It is worth noting that the above results require complete dynamical knowledge of the system and belong to model-based methods, which are hard to accomplish in complex industrial processes. Hence, model-free optimal approaches based on reinforcement learning (RL) have been exploited to overcome this weakness. RL is a fascinating machine learning technique in which optimal control strategy is learned relying on the interactions with the environment. In the optimal field, RL and ADP are interchangeable concepts [19]. As a typical model-free RL method, the off-policy algorithm was pioneered in [20] that learned the solution to the HJBE from the system data generated by an arbitrary training policy. This approach was continued in [21] for seeking the Nash equilibrium solution and solving H_∞ control problem in [22], respectively. Later, due to the powerful feature about coping the unknown system dynamics, the off-policy RL algorithm was firstly extended to solve optimal synchronization problem for MASs with linear form in [23]. In [24], the optimal synchronization problem for nonlinear MASs was investigated, which was viewed as an advance in off-policy algorithm. Recently, on the basis of [24], in [25], the asymmetric input-constraint was taken into account in model-free optimal control problem. The optimal synchronization of heterogeneous MASs in graphical games was handled in [26], where the optimal control policy depending on not only the Nash equilibrium but the fixed control policies of neighbors. However, to the best of the authors' knowledge, there are no results about the HiTL optimal output cluster synchronization problem considering unknown output of leader, which motivates our research.

Driven by these observations, in this paper, the prescribed-time observer-based HiTL optimal output cluster synchronization control strategy is built for MASs via off-policy RL algorithm. To improve the safety of MASs, the human operator is introduced to guide the whole MASs by sending the signal to the non-autonomous leader. A prescribed-time observer is designed for the situation that only part of the followers can access the output information of leader, which can achieve the desired accuracy in the prescribed time interval. Based on the constructed observer, the off-policy RL algorithm is applied to obtain the optimal controller, in which the requirement of completely known system dynamics is released. In addition, to implement the algorithm, the single critic NN is utilized. The main contributions are summarized as follows.

(1) In this paper, to cope with the practical situation that the output of leader is unavailable to all followers, a prescribed-time observer is developed for each follower via the designed function. Compared with [27], [18], the observer error can converge to the required accuracy in prescribed time and the requirement of full-order derivatives information of the leader is relaxed. More importantly, the prescribed time and accuracy are independent of the initial conditions, which are friendly to users and fit for practical engineering.
(2) In this paper, in view of balancing the performance and cost of systems and achieving the optimal performance, the off-policy RL algorithm is applied to solve the optimal output cluster synchronization control problem without any knowledge of system dynamics. Compared with model-based RL algorithm in [18], [28], the model-free RL algorithm relies only on measurement data of systems and has a broader development prospect in large-scale complex systems. Furthermore, in contrast to the framework of critic-actor NNs in [25], [28], the single critic NN is utilized to implement off-policy algorithm. It leads to a simpler structure, less computation burden and eliminates the training errors caused by actor NN.

The structure is listed as follows. In Section II, the considered system and some concepts are given. In Section III, the practical prescribed-time observer is presented. In Section IV, the process of designing optimal cluster synchronization controller is provided. Model-free RL algorithm and simulation results are presented in Section V and Section VI, respectively. Finally, the conclusion is given in Section VII.

Notations: Throughout this article, $\mathbb{R}^{n\times m}$ represents the set of $n \times m$ real matrices; $\mathbb{I}_p$ represents the identity matrix of dimension p; $\mathbb{I}_N$ represents the identity matrix of dimension N; $\mathbb{I}_{pN}$ represents the identity matrix of dimension pN; $\mathbb{O}_{p\times 1}$ denotes the zero matrix of dimension $p\times 1$; $\mathbb{O}_{p\times m}$ denotes the zero matrix of dimension $p \times m$; $\otimes$ is the Kronecker product.

II. Problem Formulation and Preliminaries

A. Communication Topologies

The communication topology contains N followers is represented by a directed graph $\mathcal{G} = \{\mathcal{V}, \varepsilon\}$, where $\mathcal{V} = \{\mathcal{V}_1, \mathcal{V}_2, \cdots, \mathcal{V}_N\}$ and $\varepsilon \subseteq \mathcal{V} \times \mathcal{V}$ represent the vertex set and the edge set of N followers, respectively. If the information can be delivered to the ith follower directly from the jth follower, the edge $(i, j) \in \varepsilon$. Let $\mathcal{A} = [a_{ij}] \in \mathbb{R}^{N\times N}$ be the weighted adjacency matrix, where $a_{ij} > 0$ if $(i, j) \in \varepsilon$, and $a_{ij} = 0$ if $(i, j) \notin \varepsilon$. The neighbor set of ith follower is defined as $\mathcal{N}_i = \{j \in \mathcal{V} : a_{ij} = 1\}$. Define $\mathcal{L} = \mathcal{D} - \mathcal{A} \in \mathbb{R}^{N\times N}$ as the Laplacian matrix of $\mathcal{G}$, where $\mathcal{D} = diag(d_1, d_2, \cdots, d_N) \in \mathbb{R}^{N\times N}$ denotes the degree matrix with $d_i = \sum_{j=1}^{N} a_{ij}$.

The vertex set $\mathcal{V}$ is divided into s clusters with one leader being associated with each cluster. Each cluster is led by its respective leader. Disjoint sets $\mathcal{V}_m, m = 1, \cdots, s$ represent the set of nodes in the mth cluster, and satisfy $\mathcal{V} = \bigcup_{m=1}^{s} \mathcal{V}_m$. If the node i belongs to the mth cluster, $l_i = m$. Let

$\mathcal{B} = diag\{b_1, b_2, \cdots, b_N\} \in \mathbb{R}^{N\times N}$, where $b_i = 1$ indicates that the information of the leader in its cluster is available for the ith node. The ith node that can receive (send) information from (to) other clusters is called an in-node (out-node) of cluster ith. The sets of in-nodes and out-nodes of cluster m are respectively denoted as $\mathcal{V}_m^{in} \subseteq \mathcal{V}_m$ and $\mathcal{V}_m^{out} \subseteq \mathcal{V}_m$.

REFERENCES

[1] Q. Yang, P. Kolaric, C. Chen, K. Xie, F. L. Lewis, and S. Xie, "Adaptive distributed synchronization of multiple vehicles with actuator nonlinearities: Design and experimentation," *IEEE Transactions on Industrial Electronics*, vol. 70, no. 7, pp. 7205–7215, 2023.

[2] J. Long, W. Wang, K. Liu, and J. Lv, "Distributed adaptive attitude synchronization of multiple spacecraft with event-triggered communication," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 58, no. 1, pp. 262–274, 2022.

[3] W. Wu, W. Zhou, and T. Chen, "Cluster synchronization of linearly coupled complex networks under pinning control," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 56, no. 4, pp. 829–839, 2009.

[4] J. Yu and L. Wang, "Group consensus in multi-agent systems with switching topologies and communication delays," *Systems & Control Letters*, vol. 59, no. 6, pp. 340–348, 2010.

[5] H. Hu, W. Yu, G. Wen, Q. Xuan, and J. Cao, "Reverse group consensus of multi-agent systems in the cooperation-competition network," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 63, no. 11, pp. 2036–2047, 2016.

[6] K. Hassan, F. Tahir, M. Rehan, C. K. Ahn, and M. Chadli, "On relative-output feedback approach for group consensus of clusters of multiagent systems," *IEEE Transactions on Cybernetics*, vol. 53, no. 1, pp. 55–66, 2023.

[7] L. Ji, X. Liu, C. Zhang, S. Yang, and H. Li, "Fully distributed dynamic event-triggered pinning cluster consensus control for heterogeneous multiagent systems with cooperative-competitive interactions," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 53, no. 1, pp. 394–404, 2023.

[8] Y. Song, Y. Wang, J. Holloway, and M. Krstic, "Time-varying feedback for regulation of normal-form nonlinear systems in prescribed finite time," *Automatica*, vol. 83, pp. 243–251, 2017.

[9] P. Zarchan, *Tactical and strategic missile guidance*. American Institute of Aeronautics and Astronautics, Inc., 2012.

[10] Y. Song, Y. Wang, and M. Krstic, "Time-varying feedback for stabilization in prescribed finite time," *International Journal of Robust and Nonlinear Control*, vol. 29, no. 3, pp. 618–633, 2019.

[11] Y. Wang and Y. Song, "A general approach to precise tracking of nonlinear systems subject to non-vanishing uncertainties," *Automatica*, vol. 106, pp. 306–314, 2019.

[12] J. Holloway and M. Krstic, "Prescribed-time observers for linear systems in observer canonical form," *IEEE Transactions on Automatic Control*, vol. 64, no. 9, pp. 3905–3912, 2019.

[13] Y. Cao, J. Cao, and Y. Song, "Practical prescribed time tracking control over infinite time interval involving mismatched uncertainties and non-vanishing disturbances," *Automatica*, vol. 136, p. 110050, 2022.

[14] Z. Li, Y. Wang, Y. Song, and W. Ao, "Global consensus tracking control for high-order nonlinear multiagent systems with prescribed performance," *IEEE Transactions on Cybernetics*, vol. 53, no. 10, pp. 6529–6537, 2023.

[15] P. Werbos, "Advanced forecasting methods for global crisis warning and models of intelligence," *General System Yearbook*, pp. 25–38, 1977.

[16] J. J. Murray, C. J. Cox, G. G. Lendaris, and R. Saeks, "Adaptive dynamic programming," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 32, no. 2, pp. 140–153, 2002.

[17] M. Abu-Khalaf and F. L. Lewis, "Nearly optimal control laws for nonlinear systems with saturating actuators using a neural network HJB approach," *Automatica*, vol. 41, no. 5, pp. 779–791, 2005.

[18] H. Fu, X. Chen, W. Wang, and M. Wu, "Observer-based adaptive synchronization control of unknown discrete-time nonlinear heterogeneous systems," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 33, no. 2, pp. 681–693, 2022.

[19] P. J. Werbos, "Reinforcement learning and approximate dynamic programming (RLADP)-foundations, common misconceptions, and the challenges ahead," *Reinforcement Learning and Approximate Dynamic Programming for Feedback Control*, pp. 1–30, 2012.

[20] Y. Jiang and Z.-P. Jiang, "Computational adaptive optimal control for continuous-time linear systems with completely unknown dynamics," *Automatica*, vol. 48, no. 10, pp. 2699–2704, 2012.

[21] H. Li, D. Liu, and D. Wang, "Integral reinforcement learning for linear continuous-time zero-sum games with completely unknown dynamics," *IEEE Transactions on Automation Science and Engineering*, vol. 11, no. 3, pp. 706–714, 2014.

[22] B. Luo, H. Wu, and T. Huang, "Off-policy reinforcement learning for H_∞ control design," *IEEE Transactions on Cybernetics*, vol. 45, no. 1, pp. 65–76, 2015.

[23] H. Modares, S. P. Nageshrao, G. A. D. Lopes, R. Babuska, and F. L. Lewis, "Optimal model-free output synchronization of heterogeneous systems using off-policy reinforcement learning," *Automatica*, vol. 71, pp. 334–341, 2016.

[24] H. Modares, F. L. Lewis, W. Kang, and A. Davoudi, "Optimal synchronization of heterogeneous nonlinear systems with unknown dynamics," *IEEE Transactions on Automatic Control*, vol. 63, no. 1, pp. 117–131, 2018.

[25] L. Xia, Q. Li, R. Song, and H. Modares, "Optimal synchronization control of heterogeneous asymmetric input-constrained unknown nonlinear MASs via reinforcement learning," *IEEE/CAA Journal of Automatica Sinica*, vol. 9, no. 3, pp. 520–532, 2022.

[26] C. Xiong, Q. Ma, J. Guo, and F. L. Lewis, "Data-based optimal synchronization of heterogeneous multiagent systems in graphical games via reinforcement learning," *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–9, 2023.

[27] G. Lin, H. Li, H. Ma, D. Yao, and R. Lu, "Human-in-the-loop consensus control for nonlinear multi-agent systems with actuator faults," *IEEE/CAA Journal of Automatica Sinica*, vol. 9, no. 1, pp. 111–122, 2022.

[28] T. Li, W. Bai, Q. Liu, Y. Long, and C. L. P. Chen, "Distributed fault-tolerant containment control protocols for the discrete-time multiagent systems via reinforcement learning method," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 34, no. 8, pp. 3979–3991, 2023.