Automatica 71 (2016) 334–341

Contents lists available at ScienceDirect

Automatica

journal homepage: www.elsevier.com/locate/automatica

Brief paper

Optimal model-free output synchronization of heterogeneous systems using off-policy reinforcement learning[☆]

Hamidreza Modares [a], Subramanya P. Nageshrao [b,1], Gabriel A. Delgado Lopes [b], Robert Babuška [b], Frank L. Lewis [a,c]

[a] University of Texas at Arlington Research Institute, 7300 Jack Newell Blvd. S., Ft. Worth, TX 76118, USA

[b] Delft Center for Systems and Control, Delft University of Technology, Mekelweg 2, 2628 CD, Delft, The Netherlands

[c] Qian Ren Consulting Professor, State Key Laboratory of Synthetical Automation for Process Industries, Northeastern University, Shenyang, China

ARTICLE INFO

Article history:
Received 27 May 2015
Received in revised form
11 January 2016
Accepted 2 May 2016
Available online 14 June 2016

Keywords:
Output synchronization
Heterogeneous systems
Reinforcement learning
Leader–follower systems

ABSTRACT

This paper considers optimal output synchronization of heterogeneous linear multi-agent systems. Standard approaches to output synchronization of heterogeneous systems require either the solution of the output regulator equations or the incorporation of a p-copy of the leader's dynamics in the controller of each agent. By contrast, in this paper neither one is needed. Moreover, here both the leader's and the follower's dynamics are assumed to be unknown. First, a distributed adaptive observer is designed to estimate the leader's state for each agent. The output synchronization problem is then formulated as an optimal control problem and a novel model-free off-policy reinforcement learning algorithm is developed to solve the optimal output synchronization problem online in real time. It is shown that this optimal distributed approach implicitly solves the output regulation equations without actually doing so. Simulation results are provided to verify the effectiveness of the proposed approach.

1. Introduction

Cooperative control of multi-agent systems has undergone a paradigm shift from centralized to distributed, due to its reliability, flexibility, scalability and computational efficiency. In distributed control, each agent designs a controller based on limited information about itself and its neighbors to assure all agents reach agreement on certain quantities of interests (leaderless consensus) or all agents follow trajectories of a leader node (leader–follower control). A rich body of literature has been developed on distributed control of multi-agent systems. See for example Jadbabaie, Lin, and Morse (2003), Lewis, Zhang, Hengster-Movric, and Das (2014), Olfati-Saber and Murray (2004), Ren and Beard (2008) and Ren, Beard, and Atkins (2007) to name a few.

Most of the existing work on distributed control focuses on state synchronization of homogeneous multi-agent systems with identical dynamics. Distributed output synchronization of heterogeneous systems, where the dynamics and dimension of agents can be different, has also attracted attention (De Persis & Jayawardhana, 2014; Huang & Chen, 2004; Huang, Ye, Hong, Wang, & Jiang, 2013; Lunze, 2012; Peymani, Grip, Saberi, Wang, & Fosson, 2014; Wieland, Sepulchre, & Allgöwer, 2011; Xiang, Wei, & Li, 2009; Yang, Saberi, Stoorvogel, & Grip, 2014). Existing mentioned methods, however, require complete knowledge of the agents and the leader's dynamics which is not available in many real-world applications.

Adaptive distributed controllers have been developed to cope with system uncertainties in the dynamics of the agents (Das & Lewis, 2010; Ding, 2015). Adaptive methods can only guarantee a bounded synchronization error and not asymptotic synchronization. This is because existing adaptive methods do not solve the output regulation equations, which is a necessary and sufficient condition to assure perfect synchronization. Moreover, they do not converge to an optimal distributed solution. Robust controllers, on the other hand, require the nominal model of the agents and full knowledge of the leader dynamics. Suboptimal

☆ This work is supported by NSF grant ECCS-1405173, ONR grant N00014-13-1-0562, ONR grant N000141410718, ARO grant W911NF-11-D-0001, China NNSF grant 61120106011, and China Education Ministry Project 111 (No. B08015). The material in this paper was not presented at any conference. This paper was recommended for publication in revised form by Associate Editor Dimos V. Dimarogonas under the direction of Editor Christos G. Cassandras.

E-mail addresses: modares@uta.edu (H. Modares), s.p.nageshrao@tudelft.nl (S.P. Nageshrao), G.A.DelgadoLopes@tudelft.nl (G.A.D. Lopes), R.Babuska@tudelft.nl (R. Babuška), lewis@uta.edu (F.L. Lewis).

[1] Tel.: +31 152788573; fax: +31 152786679.

http://dx.doi.org/10.1016/j.automatica.2016.05.017
0005-1098/

and optimal distributed controllers are also designed in Cao and Ren (2010), Vamvoudakis, Lewis, and Hudas (2012), Zhang, Feng, Liang, and Luo (in press) and Zhang, Lewis, and Das (2011) for linear homogeneous systems. These methods, however, are limited to state synchronization and they require complete knowledge of the agents and the leader. To our knowledge distributed adaptive optimal output synchronization is not considered in the literature.

Over the last decades there has been increasing interest in developing the multi-agent learning systems so as to create agents that can learn from experience how to interact with other agents in a best possible way (Busoniu, Babuska, & De Schutter, 2008; Chang, 2009; Lakshmanan & Farias, 2006). Reinforcement learning (RL) techniques have been used as promising methods to design adaptive optimal controllers for both single-agent and multi-agent systems.

In this paper, a reinforcement learning (RL) algorithm is developed to solve the output synchronization of heterogeneous systems. It is shown that the explicit solution to the output regulator equation is not necessary, hence the agents do not need to know the leader's dynamics. The key components of the given method are:

(1) Optimality is explicitly imposed in solving the output synchronization problem. This allows the use of RL to solve the problem in real time.
(2) A novel off-policy RL algorithm is developed to solve the output synchronization problem online in real time without requiring any knowledge of the agent's or the leader's dynamics.
(3) It is shown that this distributed RL approach implicitly solves the output regulation equations without actually solving them.
(4) A model-free distributed adaptive observer is designed to estimate the leader's state.

The distributed observer gives the feedforward state, which is used along with the local feedback state of each agent to design a local control protocol. A local discounted performance function is defined for each agent, its minimization gives both feedback and feedforward controllers. Online solution to the minimization problem is then found by using an off-policy RL algorithm. This algorithm does not require any knowledge of the dynamics of the agents and uses only the measured data of the system and the reference trajectories to find the optimal distributed solution to the output synchronization problem.

2. Theoretical background

In this section, the essential theoretical background on graph theory is provided. The problem of output synchronization for heterogeneous multi-agent systems is also defined.

2.1. Graph theory

A weighted directed graph (digraph) is defined as $\mathcal{G} = (\mathcal{V}, \mathcal{E}, A)$, where $\mathcal{V} = (v_1, v_2, \ldots, v_N)$ is a set of N nodes, $\mathcal{E} \subset \mathcal{V} \times \mathcal{V}$ is a set of edges, and $A = [a_{ij}] \in \mathbb{R}^{N\times N}$ is the associated adjacency matrix with $a_{ij} > 0$ only if $(v_j, v_i) \in \mathcal{E}$, and $a_{ij} = 0$ otherwise. The neighbor set of node i is depicted by $N_i = \{j|(v_j, v_i) \in \mathcal{E}\}$. The in-degree of a node i is defined as $d_i = \sum_{j=1}^N a_{ij}$ and in-degree matrix as $D = \mathrm{diag}\,[d_i] \in \mathbb{R}^{N\times N}$. The graph Laplacian matrix is defined as $L = D - A$. A graph is said to be undirected if the graph Laplacian is symmetric, i.e., $L = L^T$. For a given digraph $\mathcal{G}$, a sequence of successive edges in the form $((v_i, v_k), (v_k, v_l), \ldots (v_m, v_j))$ gives a directed path from node i to node j. A diagraph is said to have a spanning tree if there exists a root node i_r, such that there is a directed path from i_r to every other node in the graph. A weighted directed graph is called detail balanced if there exist scalars $p_i > 0,\ p_j > 0$ such that $p_i a_{ij} = p_j a_{ji}$ for all $i, j \in N$ (Lewis et al., 2014).

Assumption 1. The digraph $\mathcal{G}$ has a spanning tree and the leader is pinned to at least one root node.

The leader can be pinned to multiple nodes in the graph. This results in a diagonal pinning matrix $G = \mathrm{diag}[g_i] \in \mathbb{R}^{N\times N}$ with $g_i > 0$ if the node has access to the leader, and zero otherwise. Under the above assumption, the eigenvalues of $L+G$ have positive real parts (Grip, Yang, Saberi, & Stoorvogel, 2012; Hong, Chen, & Bushnell, 2008; Hong, Hu, & Gao, 2008; Li, Duan, Chen, & Huang, 2008).

2.2. Output synchronization of heterogeneous multi-agent systems

Let the leader dynamics be given by

$$\dot{\zeta}_0 = S\,\zeta_0 \tag{1}$$

where $\zeta_0 \in \mathbb{R}^p$ is the reference trajectory to be followed by followers, and $S \in \mathbb{R}^{p\times p}$ is the leader's dynamic matrix. Using the output matrix $R \in \mathbb{R}^{q\times p}$, the output of the leader $y_0 \in \mathbb{R}^q$ is

$$y_0 = R\,\zeta_0. \tag{2}$$

Assumption 2. All eigenvalues of the leader dynamic S are on the imaginary axis and they are non-repeated.

The dynamics of N linear heterogeneous agents is given by

$$\begin{aligned} \dot{x}_i &= A_i x_i + B_i u_i \\ y_i &= C_i x_i \end{aligned} \tag{3}$$

where $x_i \in \mathbb{R}^{n_i}$ is the system state, $u_i \in \mathbb{R}^{m_i}$ is the input and $y_i \in \mathbb{R}^q$ is the output for agent i.

Assumption 3. (A_i, B_i) is stabilizable and (A_i, C_i) is observable.

Problem 1 (*Output Synchronization*)**.** Design local control protocols u_i such that the outputs of all agents synchronize to the output of the leader node. That is, $y_i(t) - y_0(t) \to 0,\ \forall i$.

To solve Problem 1, standard methods in the literature require solving the output regulation equations given by

$$\begin{aligned} A_i\Pi_i + B_i\Gamma_i &= \Pi_i S \\ C_i\Pi_i &= R \end{aligned} \tag{4}$$

where $\Pi_i \in \mathbb{R}^{n_i\times p}$ and $\Gamma_i \in \mathbb{R}^{m_i\times p}$ for $i = 1, \ldots, N$ are solutions to (4). Based on these solutions, the following control protocol guarantees output synchronization (Grip et al., 2012; Su, Hong, & Huang, 2013; Yang et al., 2014)

$$u_i = \bar{K}_{1i}\,(x_i - \Pi_i\zeta_i) + \Gamma_i\zeta_i \tag{5}$$

where $\bar{K}_{1i} \in \mathbb{R}^{m_i\times n_i}$ is the state-feedback gain which stabilizes $A_i + B_i\bar{K}_{1i}$. Moreover, ζ_i is the estimate of the leader trajectory ζ_0 for agent i and is given by Jiao, Liu, Lewis, Xu, and Xie (2015)

$$\dot{\zeta}_i = S\,\zeta_i + c\left[\sum_{j=1}^{N} a_{ij}(\zeta_j - \zeta_i) + g_i(\zeta_0 - \zeta_i)\right] \tag{6}$$

where c is the coupling gain.

Remark 1. Note that the solution to the output regulator equation (4) requires the complete knowledge of the leader and the agents' dynamics. In order to obviate this requirement, a model-free distributed adaptive observer is first designed in this paper to estimate the leader's state ζ_0. Then, a model-free off-policy reinforcement learning algorithm is developed to solve the optimal tracking problem online in real time.