A Novel Distributed Optimal Adaptive Control Algorithm for Nonlinear Multi-Agent Differential Graphical Games

Majid Mazouchi, Mohammad Bagher Naghibi-Sistani, and Seyed Kamal Hosseini Sani

***Abstract*—In this paper, an online optimal distributed learning algorithm is proposed to solve leader-synchronization problem of nonlinear multi-agent differential graphical games. Each player approximates its optimal control policy using a single-network approximate dynamic programming (ADP) where only one critic neural network (NN) is employed instead of typical actor-critic structure composed of two NNs. The proposed distributed weight tuning laws for critic NNs guarantee stability in the sense of uniform ultimate boundedness (UUB) and convergence of control policies to the Nash equilibrium. In this paper, by introducing novel distributed local operators in weight tuning laws, there is no more requirement for initial stabilizing control policies. Furthermore, the overall closed-loop system stability is guaranteed by Lyapunov stability analysis. Finally, Simulation results show the effectiveness of the proposed algorithm.**

***Index Terms*—Approximate dynamic programming (ADP), distributed control, neural networks (NNs), nonlinear differential graphical games, optimal control.**

I. INTRODUCTION

RESEARCH on distributed control of multi agent systems linked by communication networks has been well studied in [1]−[7]. This growing field, is mainly applicable to a variety of engineering systems such as formation of a group of mobile robots [8], distributed containment control [9], vehicles formation control [10], sensor networks [11], [12], networked autonomous team [13], distributed electric power system control [14], [15] and synchronization of dynamical processes. There are many advantages for distributed control such as less computational complexity and no need for a centralized decision-making center.

Distributed control problems can be classified into two main groups, namely leaderless consensus (distributed regulation) and leader-follower concensus (distributed tracking) problems. In the leaderless consensus all agents converge to an uncontrollable common value (consensus value) which depends on their initial states in the communication network [16]−[19]. On the other hand, the problem of leader-follower consensus [20]−[23], which is the problem of interest in this paper, requires that all agents synchronize to a leader or control agent who generates the desired reference trajectory [20], [24].

Game theory [25], [26] provides a proper solution framework for formulating strategic behaviors, where the strategy of each player depends on the actions of itself and other players. Therefore, it has become the theoretical framework in the field of multi-player games [27]−[30]. Differential game is a branch of game theory which addresses dynamical interacting multi-agent decision control problems. A new class of differential games is called differential graphical game [31], where the error dynamics and performance index of each player depends on itself and its neighbors in the game communication graph topology. In differential graphical game, the players' goal is to find a set of policies that are admissible, i.e., control policies that ensure the stability of the overall system, in order to guarantee global synchronization, local optimization and Nash equilibrium achievement. In order to find the Nash equilibrium, one has to solve a set of coupled Hamilton-Jacobi (HJ) equations. These coupled HJ equations are difficult or impossible to solve analytically and they depend on the graph topology interactions. Therefore, in order to approximately solve the coupled HJ equations in an online fashion, numerical methods such as reinforcement learning (RL) methods [32] are required. Approximate dynamic programming (ADP) [33] is an efficient and forwarded in time RL method which can be used to generate approximate online optimal control policies.

ADP has been fruitfully used to develop adaptive optimal controllers for single-agent systems [35]−[39] and multi-agent systems [31], [40]−[48] online in real time. While noticeable progress has been made on ADP in field of distributed control in multi-agent systems, fewer results consider the differential graphical game. In [31], [43]−[47], concepts of ADP and differential graphical game are brought together to find an online optimal solution for distributed tracking control of continuous-time linear systems.

In [31], an online cooperative policy iteration (PI) algorithm is developed for graphical games of continuous-time linear systems by using the actor-critic architecture [49], composed of two neural networks referred to as actor NN and critic NN. A PI algorithm based on integral reinforcement learning technique [35] is proposed in [43] to learn the Nash solution of linear graphical games in real time. In [44], an online PI algorithm is proposed to solve linear differential graphical

Manuscript received November 16, 2016; accepted May 04, 2017. Recommended by Associate Editor Huaguang Zhang. *(Corresponding author: Mohammad Bagher Naghibi-Sistani.)*

Citation: M. Mazouchi, M. B. Naghibi-Sistani, and S. K. H. Sani, "A novel distributed optimal adaptive control algorithm for nonlinear multi-agent differential graphical games," *IEEE/CAA J. of Autom. Sinica*, vol. 5, no. 1, pp. 331−341, Jan. 2018.

M. Mazouchi, M. B. Naghibi-Sistani, and S. K. H. Sani are with the Department of Electrical Engineering, Ferdowsi University of Mashhad, Mashhad, Razavi khorasan 9177948974, Iran (e-mail: Mazouchi.Majid@mail.um.ac.ir; mb-naghibi@um.ac.ir; k.hosseini@um.ac.ir).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JAS.2017.7510784

games in real time. A cooperative PI algorithm is proposed in [45] to solve linear differential graphical games where all players are heterogeneous in their dynamics. The authors in [46] formulate linear output regulation problem in the linear differential graphical game framework. Moreover, an online PI algorithm is proposed in [46] to obtain the solution of coupled HJ equations by using actor-critic structure in real time. An online PI algorithm is provided in [47] to find the solution of coupled Hamilton-Jacobi-Isaacs (HJI) equations in zero-sum continuous-time linear differential graphical game where the players were influenced by disturbances. In [40], an ADP algorithm was developed to solve differential graphical games of continuous-time nonlinear systems. The authors in [40] solved the problem by using actor-critic architecture. In [31], [40], [43]−[47], the initial admissible control policies are required to guarantee the stability of the differential graphical game. However, finding the set of initial stabilizing control policies for the players is not a direct and simple task.

In [50], the authors proposed an online optimal single-network ADP method to solve zero-sum differential game without the requirement of initial stabilizing control policies and [42] extends the results of [50] to obtain the Nash equilibrium of two-player nonzero sum differential game. To our knowledge, there has not been any result on solving the N-player differential graphical games of continuous-time nonlinear systems using single-network ADP without the requirement of initial stabilizing control policies.

In this paper, an online optimal distributed learning algorithm is proposed to approximately solve the coupled HJ equations of N-player differential graphical game in an online fashion. Each player approximates its optimal control policy using a single-network ADP. The proposed distributed weight tuning laws of critic NNs guarantee the closed-loop stability in the sense of uniform ultimate boundedness (UUB) and convergence of control policies to the Nash equilibrium. By introducing novel distributed local operators in distributed weight tuning laws, the requirement for initial stabilizing control policies is eliminated.

The contributions of the paper are as follows:

1) This paper extends the results of [42], [50] to the N-player differential graphical games of continuous-time nonlinear systems which have more complexity due to the distributed graphical based formulation of the game and the number of players in comparison with the two-player nonzero-sum [42] and zero-sum [50] differential games. Moreover, the stability of the overall closed-loop system is guaranteed.

2) The distributed learning algorithm proposed in this paper employs only one critic network for each player. As results, this algorithm is less computationally demanding and simpler to implement in comparison with [31], [40], [44]−[47], which used actor-critic structure composed of two NNs for each player.

3) By introducing novel distributed local operators in distributed weight tuning laws, in contrast with [31], [40], [43]−[47], there is no more requirement for initial stabilizing control policies.

The paper is organized as follows. The problem formulation of N-player graphical differential games of nonlinear systems is described in Section II. Section III develops the online optimal distributed learning algorithm to solve the N- player graphical differential games of continuous-time nonlinear systems using single-network ADP. Section IV, presents simulation examples that show the effectiveness of the proposed approach. Finally, the conclusions are drawn in Section V.

II. Preliminaries and Problem Formulation

A. Graphs

Let the topology of the interactions among leader and followers be represented by digraph $G(V,\Sigma)$, where $V = \{\nu_0,\nu_1,\dots,\nu_N\}$ is a nonempty finite set of $N+1$ nodes and Σ is a set of edges belonging to the product space of V (i.e., $\Sigma \subseteq V\times V$). Denote the edge from node j to node i as $\gamma_{ij} = (\nu_j,\nu_i)$. The leader node is denoted by ν_0 and the leader node does not have any incoming edge. There is at least one outgoing edge from the leader node to one of the followers ν_i in the graph G (i.e., $\gamma_{i0} > 0$). We assume that the graph is simple i.e. There are no self-loops or multiple edges. Consider graph $G'(V',\Sigma')$, as the sub-graph of G, obtained by removing node ν_0 and its relating edges. The weighted adjacency matrix of graph G' is denoted by $\Gamma = [\gamma_{ij}] \in \mathbb{R}^{N\times N}$ with $\gamma_{ij}\in\Sigma \Leftrightarrow \gamma_{ij} > 0$; otherwise $\gamma_{ij} = 0$. The set of neighbors of node ν_i and the set of nodes which contains ν_i in its neighborhood is denoted by $N_i^I = \{\nu_j : (\nu_j,\nu_i)\in\Sigma'\}$ and $N_i^O = \{\nu_j : (\nu_i,\nu_j)\in\Sigma'\}$, respectively. The in-degree matrix of graph G' is defined as $D = \mathrm{diag}(d_i) \in \mathbb{R}^{N\times N}$, where $d_i = \sum_{j\in N_i^I}\gamma_{ij}$ is the weighted in-degree of node ν_i. A direct path is an ordered sequence of nodes in the graph. A digraph is said to contain a spanning tree rooted at ν_i, if there is a directed path from the ν_i to any other nodes in the graph. A digraph is called detail balanced if there exist scalars $\tau_i > 0$, $\tau_j > 0$ such that $\tau_i\gamma_{ij} = \tau_j\gamma_{ji}$ for all $i,j\in N$ [7].

In this paper, a detail balanced digraph containing a spanning tree rooted at the leader node is considered as the players interactions graph in the game.

B. Problem Formulation

Consider a group of N-players distributed on a directed interaction graph, whose dynamics are described as follows

$$\dot{\boldsymbol{x}}_i(t) = f_i(\boldsymbol{x}_i(t)) + g_i(\boldsymbol{x}_i(t))\boldsymbol{u}_i,\quad t\geq 0 \tag{1}$$

for $i = 1,\dots,N$, where $\boldsymbol{x}_i(t)\in\mathbb{R}^n$ is the state vector of player i and $\boldsymbol{u}_i(t)\in\mathbb{R}^m$ is its control input vector. Also consider the leader agent dynamics $\boldsymbol{x}_0(t)\in\mathbb{R}^n$ given by

$$\dot{\boldsymbol{x}}_0 = f_0(\boldsymbol{x}_0(t)),\quad t\geq 0. \tag{2}$$

Assumption 1: $f_0(\boldsymbol{x}_0), f_i(\boldsymbol{x}_i)$ and $g_i(\boldsymbol{x}_i)$ for $i = 1,\dots,N$ are locally Lipschitz.

Remark 1: Assumption 1 requires $f_i(\boldsymbol{x}_i(t))+g_i(\boldsymbol{x}_i(t))\boldsymbol{u}_i$ for $i = 1,\dots,N$ be locally Lipschitz which is a standard assumption (For instance see [36], [40], [42], [51]) to guarantee the uniqueness of the solution of system (1) for any finite initial condition.

The local tracking error δ_i for player i, $i = 1, \ldots, N$, is defined by

$$\delta_i = \sum_{j \in N_i^I} \gamma_{ij}(\boldsymbol{x}_i - \boldsymbol{x}_j) + \gamma_{i0}(\boldsymbol{x}_i - \boldsymbol{x}_0). \tag{3}$$

The dynamics of the local tracking error [40] for player i, $i = 1, \ldots, N$, is given by

$$\begin{aligned}\dot{\delta}_i = &\sum_{j \in N_i^I} \gamma_{ij}(f_i(\boldsymbol{x}_i) - f_j(\boldsymbol{x}_j)) + \gamma_{i0}(f_i(\boldsymbol{x}_i) - f_0(\boldsymbol{x}_0)) \\ &+ (d_i + \gamma_{i0}) g_i(\boldsymbol{x}_i) \boldsymbol{u}_i - \sum_{j \in N_i^I} \gamma_{ij} g_j(\boldsymbol{x}_j) \boldsymbol{u}_j. \end{aligned} \tag{4}$$

Note that, local tracking error dynamics (4) is an interacting dynamical system driven by the control actions of agent i and all of its neighbors.

In differential graphical game, players wish to achieve synchronization while simultaneously optimizing their local cost functions. The distributed local cost function for each player i, $i = 1, \ldots, N$, is defined by

$$J_i(\delta_i, u_i, u_{N_i^I}) = \int_t^{\infty} r_i\big(\delta_i(\tau), u_i(\tau), u_{N_i^I}(\tau)\big) d\tau \tag{5}$$

where $r_i(\delta_i, u_i, u_{N_i^I}) = Q_i(\delta_i)/2 + {u_i}^T R_{ii} u_i/2 + \sum_{j \in N_i^I} {u_j}^T R_{ij} u_j/2$, $u_{N_i^I} = \{u_j | j \in N_i^I\}$, $Q_i(\delta_i) > 0$ and the constant weighting matrices $R_{ii} > 0$ and $R_{ij} > 0$ are symmetric.

Definition 1 [26], [31]: The set of policies $\{u_1^*, u_2^*, \ldots, u_N^*\}$ is a global Nash equilibrium solution for N-player differential graphical game if the following inequalities hold for all i, $i = 1, \ldots, N$, and $\forall u_i, u_{Gr-i}$

$$J_i^* \equiv J_i(u_i^*, u_{Gr-i}^*) \le J_i(u_i, u_{Gr-i}^*) \tag{6}$$

where $u_{Gr-i} = \{u_j | j \ne i\}$. The N-tuple of the distributed local cost functions $\{J_1^*, J_2^*, \ldots, J_N^*\}$ is known as the Nash equilibrium of the differential graphical game.

Given policies of player i and its neighbors, the value function for each player i, $i = 1, \ldots, N$, is given by

$$\begin{aligned} V_i(\delta_i) &\equiv V_i(\delta_i, u_i, u_{N_i^I}) \\ &= \int_t^{\infty} r_i\big(\delta_i(\tau), u_i(\tau), u_{N_i^I}(\tau)\big) d\tau. \end{aligned} \tag{7}$$

In differential graphical game, the goal of player i, for $i = 1, \ldots, N$, is to determine

$$V_i^*(\delta_i) = \min_{u_i} \int_t^{\infty} r_i\big(\delta_i(\tau), u_i(\tau), u_{N_i^I}(\tau)\big) d\tau. \tag{8}$$

The differential equivalent formulation of (7) is given by [40]

$$\begin{aligned} &\nabla V_i^T \Big(\sum_{j \in N_i^I} \gamma_{ij}\big(f_i(x_i) - f_j(x_j)\big) + \gamma_{i0}\big(f_i(x_i) - f_0(x_0)\big) \\ &\quad + (d_i + \gamma_{i0}) g_i(x_i) u_i - \sum_{j \in N_i^I} \gamma_{ij} g_j(x_j) u_j \Big) + \frac{1}{2} Q_i(\delta_i) \\ &+ \frac{1}{2} u_i^T R_{ii} u_i + \frac{1}{2} \sum_{j \in N_i^I} u_j^T R_{ij} u_j = 0 \end{aligned} \tag{9}$$

where $V_i(0) = 0$ and $\nabla V_i \triangleq \dfrac{\partial V_i}{\partial \delta_i} \in \mathbb{R}^n, i = 1, \ldots, N$.

Hamiltonian function for the distributed local cost function of player i, $i = 1, \ldots, N$, is defined as below

$$\begin{aligned} H_i(\delta_i, u_i, u_{N_i^I}) &\equiv \frac{1}{2} Q_i(\delta_i) + \frac{1}{2} u_i^T R_{ii} u_i + \frac{1}{2} \sum_{j \in N_i^I} u_j^T R_{ij} u_j \\ &+ \nabla V_i^T \Big(\sum_{j \in N_i^I} \gamma_{ij}(f_i(x_i) - f_j(x_j)) + \gamma_{i0}(f_i(x_i) - f_0(x_0)) \\ &+ (d_i + \gamma_{i0}) g_i(x_i) u_i - \sum_{j \in N_i^I} \gamma_{ij} g_j(x_j) u_j \Big). \end{aligned} \tag{10}$$

Based on Hamiltonian (10), the optimal feedback control policies can be derived by the stationary condition [52], $\frac{\partial H_i}{\partial u_i} = 0$, as follows

$$u_i^* = -(d_i + \gamma_{i0}) R_{ii}^{-1} g_i^T(x_i) \nabla V_i \tag{11}$$

for $i = 1, \ldots, N$, where the ∇V_i is the solution of coupled Hamilton-Jacobi (HJ) equations (12).

Substituting optimal feedback control policy (11) into (10), we have the coupled Hamilton-Jacobi (HJ) equations for $i = 1, \ldots, N$ as follows

$$\begin{aligned} &\frac{1}{2} \sum_{j \in N_i^I} (d_j + \gamma_{j0})^2 \nabla V_j^T g_j(x_j) R_{jj}^{-1} R_{ij} R_{jj}^{-1} g_j^T(x_j) \nabla V_j \\ &+ \frac{1}{2} Q_i(\delta_i) + \frac{1}{2} (d_i + \gamma_{i0})^2 \nabla V_i^T g_i(x_i) R_{ii}^{-1} g_i^T(x_i) \nabla V_i \\ &+ \nabla V_i^T \Big(\sum_{j \in N_i^I} \gamma_{ij}\big(f_i(x_i) - f_j(x_j)\big) + \gamma_{i0}\big(f_i(x_i) \\ &- f_0(x_0)\big) - (d_i + \gamma_{i0})^2 g_i(x_i) R_{ii}^{-1} g_i^T(x_i) \nabla V_i \\ &+ \sum_{j \in N_i^I} \gamma_{ij}(d_j + \gamma_{j0}) g_j(x_j) R_{jj}^{-1} g_j^T(x_j) \nabla V_j \Big) = 0. \end{aligned} \tag{12}$$

Generally, finding analytical solutions for these coupled HJ equations is difficult or impossible. Therefore, an online optimal distributed learning algorithm is proposed using only single network ADP for each player to solve the coupled HJ equations of (12) in order to obtain the optimal feedback control policies (11) and reach the Nash equilibrium.

Remark 2: The approaches proposed in [50] and [42] cannot be extended directly to solve N-player differential graphical game (11) and (12), due to the distributed graphical based formulation of the game and the number of players.

Before we present the online optimal distributed learning algorithm, the following assumptions and lemma are needed.

Assumption 2: The coupled HJ equations (12) have non-negative smooth solutions $V_i > 0$.

Remark 3: The coupled HJ (12) may have non-smooth or non-continuous value functions. However, under Assumption 2, which is a standard assumption in neural adaptive control literature [31], [40]−[42], [51], [53], solutions to the coupled HJ equations (12) are guaranteed to be smooth. This allows us to use the Weierstrass high-order approximation theorem [39], [51, Remark 1].

Assumption 3: For each player i, there exists a continuously differentiable radially unbounded Lyapunov candidate $L_i(\delta_i)$ such that

$$\begin{aligned}\dot{L}_i &= \nabla L_i^T \dot{\delta}_i \\ &= \nabla L_i^T \Big(\sum_{j\in N_i^I} \gamma_{ij}\big(f_i(x_i)-f_j(x_j)\big)+\gamma_{i0}\big(f_i(x_i)-f_0(x_0)\big) \\ &\quad + (d_i+\gamma_{i0})g_i(x_i)u_i^* - \sum_{j\in N_i^I}\gamma_{ij}g_j(x_j)u_j^*\Big) < 0 \end{aligned} \tag{13}$$

for $i=1,\ldots,N$, where $\nabla L_i \triangleq \dfrac{\partial L_i}{\partial \delta_i} \in \mathbb{R}^n$.

Remark 4: The requirement of $L_i(\delta_i)$ being radially unbounded can be fulfilled by its proper choice as quadratic polynomials [42], [50]. Although, the existence of continuously differentiable and radially unbounded Lyapunov candidates is not usually required in Lyapunov theory, however their existence have been shown by converse Lyapunov theorems [54].

Lemma 1: Consider the system given by (4) with the distributed local cost functions (7) and optimal feedback control policies (11). Let Assumption 3 holds. Now assume that $\bar{C}_i$ is a positive constant and satisfies the following inequality

$$\nabla V_i^{*T}\bar{C}_i\nabla L_i \le r_i(\delta_i,u_i^*,u_{N_i^I}^*) \tag{14}$$

then, we have

$$\begin{aligned}&\nabla L_i^T \Big(\sum_{j\in N_i^I} \gamma_{ij}\big(f_i(x_i)-f_j(x_j)\big)+\gamma_{i0}\big(f_i(x_i)-f_0(x_0)\big) \\ &\quad + (d_i+\gamma_{i0})g_i(x_i)u_i^* - \sum_{j\in N_i^I}\gamma_{ij}g_j(x_j)u_j^*\Big) \\ &\le -\nabla L_i^T\bar{C}_i\nabla L_i. \end{aligned} \tag{15}$$

Proof: By applying optimal feedback control policies (11) to nonlinear systems (4), the distributed local cost function $V_i(\delta_i,u_i^*,u_{N_i^I}^*)$ (7) becomes a Lyapunov function. Then, by using Hamiltonian function (10) and differentiating the distributed local cost function $V_i^*\equiv V_i(\delta_i,u_i^*,u_{N_i^I}^*)$ with respect to t, we obtain

$$\begin{aligned}\dot{V}_i^* &= \nabla V_i^{*T} \Big(\sum_{j\in N_i^I} \gamma_{ij}\big(f_i(x_i)-f_j(x_j)\big)+\gamma_{i0}\big(f_i(x_i)-f_0(x_0)\big) \\ &\quad + (d_i+\gamma_{i0})g_i(x_i)u_i^* - \sum_{j\in N_i^I}\gamma_{ij}g_j(x_j)u_j^*\Big) \\ &= -r_i(\delta_i,u_i^*,u_{N_i^I}^*). \end{aligned} \tag{16}$$

Using (14), we can rewrite (16) as

$$\begin{aligned}&\sum_{j\in N_i^I} \gamma_{ij}\big(f_i(x_i)-f_j(x_j)\big)+\gamma_{i0}\big(f_i(x_i)-f_0(x_0)\big) \\ &\quad + (d_i+\gamma_{i0})g_i(x_i)u_i^* - \sum_{j\in N_i^I}\gamma_{ij}g_j(x_j)u_j^* \\ &= -(\nabla V_i^*\nabla V_i^{*T})^{-1}\nabla V_i^* r_i(\delta_i,u_i^*,u_{N_i^I}^*) \\ &\le -(\nabla V_i^*\nabla V_i^{*T})^{-1}\nabla V_i^*\nabla V_i^{*T}\bar{C}_i\nabla L_i \\ &\le -\bar{C}_i\nabla L_i. \end{aligned} \tag{17}$$

Finally, by multiplying $\nabla {L_i}^T$ to the both sides of (17), we obtain (15), which completes the proof. ■

III. Online Solution of N-player Nonlinear Differential Graphical Games Using Single-network ADP

According to the Weierstrass higher-order approximation theorem [55], assume that there exist critic NN constant weights $W_i \in \mathbb{R}^{K_i}$, such that the smooth value functions $V_i(\delta_i)$, and its gradient $\nabla V_i \triangleq \dfrac{\partial V_i}{\partial \delta_i}$ are approximated as

$$V_i \triangleq V_i(\delta_i) = W_i^T\sigma_i(\delta_i)+\varepsilon_i(\delta_i) \tag{18}$$

$$\nabla V_i = \nabla\sigma_i^T W_i + \nabla\varepsilon_i \tag{19}$$

for $i=1,\ldots,N$, where K_i is the number of hidden-layer neurons of player i, $\varepsilon_i(\delta_i)$ are the NN approximation errors, $\sigma_i(\delta_i):\mathbb{R}^n\to\mathbb{R}^{K_i}$, are critic NN activation function vectors and $\nabla\sigma_i \triangleq \dfrac{\partial\sigma_i}{\partial\delta_i}$, $\nabla\varepsilon_i \triangleq \dfrac{\partial\varepsilon_i}{\partial\delta_i}$.

The critic NN activation function vectors $\sigma_i(\delta_i)$ are selected so that $\sigma_i(\delta_i)$ provides complete independent basis sets, for $i=1,\ldots,N$, such that $\sigma_i(0)=0$, $\nabla\sigma_i(0)=0$. The approximation errors $\varepsilon_i(\delta_i)$ and its gradient $\nabla\varepsilon_i(\delta_i)$ converge to zero uniformly as $K_i\to\infty$ [55].

Using (19), we can rewrite the optimal feedback control policies (11) and the coupled HJ equations (12), respectively, as follows

$$\begin{aligned}u_i^* =& -(d_i+\gamma_{i0}){R_{ii}}^{-1}g_i^T(x_i)\nabla{\sigma_i}^TW_i \\ &-(d_i+\gamma_{i0}){R_{ii}}^{-1}g_i^T(x_i)\nabla\varepsilon_i \end{aligned} \tag{20}$$

$$\begin{aligned}&\frac{1}{2}Q_i(\delta_i)-\frac{1}{2}(d_i+\gamma_{i0})^2W_i^T\nabla\sigma_iD_i\nabla\sigma_i^TW_i \\ &+\frac{1}{2}\sum_{j\in N_i^I}(d_j+\gamma_{j0})^2W_j^T\nabla\sigma_jS_{ij}\nabla\sigma_j^TW_j \\ &+W_i^T\nabla\sigma_i\Big(\sum_{j\in N_i^I}\gamma_{ij}\big(f_i(x_i)-f_j(x_j)\big)+\gamma_{i0}\big(f_i(x_i)-f_0(x_0)\big) \\ &+\sum_{j\in N_i^I}\gamma_{ij}(d_j+\gamma_{j0})D_j\nabla\sigma_j^TW_j\Big)-\varepsilon_{HJ_i}=0 \end{aligned} \tag{21}$$

for $i=1,\ldots,N$, where

$$D_i = g_i(x_i)R_{ii}^{-1}g_i^T(x_i) \tag{22}$$

$$S_{ij} = g_j(x_j)R_{jj}^{-1}R_{ij}R_{jj}^{-1}g_j^T(x_j). \tag{23}$$

The residual error of player i, $i=1,\ldots,N$, in the coupled HJ equations (21), denoted by ε_{HJ_i}, is given by

$$\begin{aligned}\varepsilon_{HJ_i} =& \frac{1}{2}(d_i+\gamma_{i0})^2\nabla\varepsilon_i^TD_i\nabla\varepsilon_i+(d_i+\gamma_{i0})^2\nabla\varepsilon_i^TD_i\nabla\sigma_i^TW_i \\ &-\frac{1}{2}\sum_{j\in N_i^I}(d_j+\gamma_{j0})^2\nabla\varepsilon_j^TS_{ij}\nabla\varepsilon_j \\ &-\sum_{j\in N_i^I}(d_j+\gamma_{j0})^2\nabla\varepsilon_j^TS_{ij}\nabla\sigma_j^TW_j \end{aligned}$$

$$\begin{aligned}&-\nabla\varepsilon_i^T\Big(\sum_{j\in N_i^I}\gamma_{ij}\big(f_i(x_i)-f_j(x_j)\big)+\gamma_{i0}\big(f_i(x_i)-f_0(x_0)\big)\\&+\sum_{j\in N_i^I}\gamma_{ij}(d_j+\gamma_{j0})D_j\nabla\sigma_j^TW_j\Big)\\&-W_i^T\nabla\sigma_i\sum_{j\in N_i^I}\gamma_{ij}(d_j+\gamma_{j0})D_j\nabla\varepsilon_j.\end{aligned}\tag{24}$$

The weights of the critic NNs, W_i, $i=1,\ldots,N$ are unknown and must be estimated. Let $\hat{W}_i$ be the current estimated value of W_i for each player i, $i=1,\ldots,N$. Therefore, the output of every critic NN for $i=1,\ldots,N$ is

$$\hat{V}_i=\hat{W}_i^T\sigma_i(\delta_i).\tag{25}$$

Substituting (25) into (11), we can rewrite the estimates of optimal control policies, for $i=1,\ldots,N$, as

$$\hat{u}_i=-(d_i+\gamma_{i0})R_{ii}^{-1}g_i^T(x_i)\nabla\sigma_i{}^T\hat{W}_i.\tag{26}$$

Applying (26) to system (4), yields the closed-loop system dynamics as follows

$$\begin{aligned}\dot{\delta}_i\equiv&\dot{\delta}_i(\hat{W}_i,\hat{W}_j)\\=&\sum_{j\in N_i^I}\gamma_{ij}\big(f_i(x_i)-f_j(x_j)\big)\\&+\gamma_{i0}\big(f_i(x_i)-f_0(x_0)\big)-(d_i+\gamma_{i0})^2D_i\nabla\sigma_i^T\hat{W}_i\\&+\sum_{j\in N_i^I}\gamma_{ij}(d_j+\gamma_{j0})D_j\nabla\sigma_j^T\hat{W}_j.\end{aligned}\tag{27}$$

By replacing (25) and (26) into (10), we obtain the approximate Hamiltonian functions as follows

$$\begin{aligned}e_{H_i}\equiv&H_i\left(\delta_i,\hat{W}_i,\hat{W}_j\right)\\=&\frac{1}{2}(d_i+\gamma_{i0})^2\hat{W}_i^T\nabla\sigma_iD_i\nabla\sigma_i^T\hat{W}_i\\&+\frac{1}{2}Q_i\left(\delta_i\right)+\frac{1}{2}\sum_{j\in N_i^I}(d_j+\gamma_{j0})^2\hat{W}_j^T\nabla\sigma_jS_{ij}\nabla\sigma_j^T\hat{W}_j\\&+\hat{W}_i^T\nabla\sigma_i\Big(\sum_{j\in N_i^I}\gamma_{ij}\big(f_i(x_i)-f_j(x_j)\big)\\&+\gamma_{i0}\big(f_i(x_i)-f_0(x_0)\big)-(d_i+\gamma_{i0})^2D_i\nabla\sigma_i^T\hat{W}_i\\&+\sum_{j\in N_i^I}\gamma_{ij}(d_j+\gamma_{j0})D_j\nabla\sigma_j^T\hat{W}_j\Big).\end{aligned}\tag{28}$$

In order to derive the critic NN weights toward their ideal values i.e. $\hat{W}_i\to W_i$, we utilize normalized gradient descent algorithm to minimize the squared residual error of e_{H_i}, for $i=1,\ldots,N$.

$$E\equiv\sum_{i=1}^{N}E_i=\frac{1}{2}\sum_{i=1}^{N}e_{H_i}{}^Te_{H_i}.\tag{29}$$

Here, we propose the distributed weight tuning laws of critic NNs (30) for N players, which minimize the squared residual error (29) and guarantee the system stability.

$$\begin{aligned}\dot{\hat{W}}_i=&-\alpha_i\frac{\bar{B}_i}{m_{s_i}}\Big(\frac{1}{2}Q_i\left(\delta_i\right)+\frac{1}{2}(d_i+\gamma_{i0})^2\hat{W}_i^T\nabla\sigma_iD_i\nabla\sigma_i^T\hat{W}_i\\&+\frac{1}{2}\sum_{j\in N_i^I}(d_j+\gamma_{j0})^2\hat{W}_j^T\nabla\sigma_jS_{ij}\nabla\sigma_j^T\hat{W}_j\\&+\hat{W}_i^T\nabla\sigma_i\Big(\sum_{j\in N_i^I}\gamma_{ij}\big(f_i(x_i)-f_j(x_j)\big)+\gamma_{i0}\big(f_i(x_i)\\&-f_0(x_0)\big)-(d_i+\gamma_{i0})^2D_i\nabla\sigma_i^T\hat{W}_i\\&+\sum_{j\in N_i^I}\gamma_{ij}(d_j+\gamma_{j0})D_j\nabla\sigma_j^T\hat{W}_j\Big)\Big)\\&+\frac{1}{2}\alpha_i(d_i+\gamma_{i0})^2\nabla\sigma_iD_i\nabla\sigma_i^T\hat{W}_i\frac{\bar{B}_i^T}{m_{s_i}}\hat{W}_i\\&+\frac{1}{2}\alpha_i\lambda_i^{-1}(d_i+\gamma_{i0})^2\nabla\sigma_i\sum_{j\in N_i^O}\lambda_j\hat{W}_j^T\frac{\bar{B}_j}{m_{s_j}}S_{ji}\nabla\sigma_i^T\hat{W}_i\\&-\bar{\chi}_i\Big(\lambda_i^{-1}\alpha_i(d_i+\gamma_{i0})\nabla\sigma_iD_i\big(\sum_{j\in N_i^O}\gamma_{ji}\nabla L_j-(d_i+\gamma_{i0})\nabla L_i\big)\Big)\\&+\lambda_i^{-1}\alpha_i(d_i+\gamma_{i0})\nabla\sigma_iD_i\Big(\bar{\chi}_i\sum_{j\in N_i^O}\gamma_{ji}\chi_j\nabla L_j-\chi_i\sum_{j\in N_i^O}\gamma_{ji}\bar{\chi}_j\nabla L_j\Big)\\&-\alpha_iF_{1i}\frac{\nabla\sigma_i\nabla\sigma_i^T}{1+\left\|\nabla\sigma_i\nabla\sigma_i^T\right\|}\hat{W}_i\\&-\alpha_iF_{2i}\begin{bmatrix}\gamma_{i1}\frac{\nabla\sigma_1\nabla\sigma_1^T}{1+\left\|\nabla\sigma_1\nabla\sigma_1^T\right\|}\hat{W}_1\\\vdots\\\gamma_{iN}\frac{\nabla\sigma_N\nabla\sigma_N^T}{1+\left\|\nabla\sigma_N\nabla\sigma_N^T\right\|}\hat{W}_N\end{bmatrix}\end{aligned}\tag{30}$$

for $i=1,\ldots,N$, where

$$\begin{aligned}B_i=&\nabla\sigma_i\Big(\sum_{j\in N_i^I}\gamma_{ij}\big(f_i(x_i)-f_j(x_j)\big)+\gamma_{i0}\big(f_i(x_i)\\&-f_0(x_0)\big)+\sum_{j\in N_i^I}\gamma_{ij}(d_j+\gamma_{j0})D_j\nabla\sigma_j^T\hat{W}_j\Big)\\&-\nabla\sigma_i(d_i+\gamma_{i0})^2D_i\nabla\sigma_i^T\hat{W}_i\end{aligned}\tag{31}$$

$m_{s_i}=1+B_i{}^TB_i$, $\bar{B}_i=B_i/m_{s_i}$, $\alpha_i>0$ is the learning rate, ∇L_i is explained in Assumption 3. λ_i, $F_{1i}\in\mathbb{R}^{K_i\times K_i}$ and $F_{2i}\in\mathbb{R}^{K_i\times NK_i}$, for $i=1,\ldots,N$ are tuning parameters.

The distributed local operators $\bar{\chi}_i\equiv\bar{\chi}_i(S,\bar{S})$ and $\chi_i\equiv\chi_i(S,\bar{S})$ are defined as follows

$$\bar{\chi}_i(S,\bar{S})=\begin{cases}0, & i\in S\\1, & i\in\bar{S}\end{cases}\tag{32}$$

$$\chi_i(S,\bar{S})=\begin{cases}1, & i\in S\\0, & i\in\bar{S}\end{cases}\tag{33}$$

for $i=1,\ldots,N$, where $S=\Big\{i:\nabla L_i\dot{\delta}_i<0\,\&\,\nabla L_j\dot{\delta}_{j}{}_{j\in N_i^O}<0\Big\}$ and $\bar{S}=\{i:i\notin S\}$.

Remark 5: In this paper, each player has its own distributed local operators $\bar{\chi}_i(S,\bar{S})$ and $\chi_i(S,\bar{S})$, which adopts with distributed nature of differential graphical games problem. Moreover, for each player the introduced distributed local operators only depend on the states of the associated player, its neighbors and the players which the associated player is in their neighborhood. Note that, $\chi_i(S,\bar{S})=1$ and $\bar{\chi}_i(S,\bar{S})=0$ imply that the local error dynamics of player i, its neighbors and the players which the player i is in their neighborhood are stable. On the other hand, $\bar{\chi}_i(S,\bar{S})=1$ and $\chi_i(S,\bar{S})=0$ imply that at least one of the local error dynamics of player i, its neighbors and the players which the player i is in their neighborhood is unstable.

Assumption 4: The systems' state given by (4) is persistently excited (PE).

Remark 6: The requirement of PE condition is a standard assumption in adaptive control literature [56]. In the adaptive control and learning literature, Assumption 4 is fulfilled by injecting a probing noise into the control input.

The following assumption will be used in the remaining part of the paper.

Assumption 5:

1) $g_i(x_i)$ are bounded by positive constants, i.e., $\|g_i\,(.)\| \leq g_{iM}$, for $i=1,\ldots,N$.

2) The critic NN approximation errors and their gradients are bounded by positive constants, i.e., $\|\varepsilon_i\| \leq \varepsilon_{iM}$ and $\|\nabla\varepsilon_i\| \leq \varepsilon_{idM}$, for $i=1,\ldots,N$.

3) The critic NN activation functions and their gradients are bounded by positive constants, i.e., $\|\sigma_i\| \leq \sigma_{iM}$ and $\|\nabla\sigma_i\| \leq \sigma_{idM}$, for $i=1,\ldots,N$.

4) The critic NN weights are bounded by positive constants, i.e., $\|W_i\| \leq W_{iM}$, for $i=1,\ldots,N$.

5) The residual errors ε_{HJi} are bounded by positive constants, i.e., $\|\varepsilon_{HJ_i}\| \leq \varepsilon_{HJ_{iM}}$, for $i=1,\ldots,N$.

Remark 7: Assumption 5 is a standard assumption in neural adaptive control literature [39], [41], [42], [53]. Although Assumption 5.1 restricts the considered class of nonlinear systems, many practical systems (e.g., robotic systems [57] and aircraft systems [58]) satisfy such a property ([31], [40]−[42], [51] for a similar assumption). According to Assumption 2 and the Weierstrass high-order approximation theorem, it is known that the NNs approximation error and their gradient are bounded, i.e., Assumption 5.2 holds. Note further that, the NNs used in this paper are so-called Functional Link NNs (See [53] for more details), for which activation functions σ_i for $i=1,\ldots,N$ can be some squashing functions, such as the standard sigmoid, Gaussian, and hyperbolic tangent functions. In fact, Assumption 5.5 can be satisfied under Assumptions 2, 3 and 5.1−5.4, if Lemma 1 holds. Furthermore, the bounds mentioned above are only used for the stability analysis and they are actually not used in the controller design.

Theorem 1: Let the dynamics be given by (4) and the control policies be given by (26). Let the Assumptions 1−5 hold and the critic NN weight tuning law of each agent be provided by (30). Let the tuning parameters be selected properly. Then, the local tracking error states δ_i and the critic NNs weight estimation errors $\tilde{W}_i=W_i-\hat{W}_i$, for $i=1,\ldots,N$ are UUB, for a sufficiently large number of NN neurons.

Proof: See Appendix A. ■

Corollary 1: Let the Theorem 1 and Assumptions 1−5 hold. Then, the control policies $\hat{u}_i$, for $i=1,\ldots,N$ form a Nash equilibrium solution.

Proof: See Appendix B. ■

Remark 8: It can be seen from (54) that by increasing $\zeta_{\min}(M)$ or $\bar{C}_i$, B_Z and consequently $\in_{u_i}$ are reduced. Therefore, by choosing proper tuning parameters λ_i, F_{1i} and F_{2i}, we can increase $\zeta_{\min}(M)$ and reduce the convergence errors $\in_{u_i}$, for $i=1,\ldots,N$. Also, by choosing proper L_i in Lemma 1, we can increase $\bar{C}_i$ and consequently reduce the convergence errors $\in_{u_i}$, for $i=1,\ldots,N$.

IV. SIMULATION

Consider a graph of five followers with a leader as shown in Fig. 1. In communication graph the pinning gains and the edge weights are chosen to be one. The dynamics of all the followers are expressed by $\dot{x}_i=f_i(x_i)+g_i(x_i)u_i$, $x_i \triangleq [x_{i1},x_{i2}]^T$, for $i=1,\ldots,5$, where

$$f_i(x_i)=\begin{pmatrix} x_{i2} \\ -x_{i1}+\varepsilon(1-x_{i1}^2)x_{i2}\end{pmatrix},\quad i=1,\ldots,5$$

$$g_1(x_1)=\begin{bmatrix} 0 \\ -0.8x_{11}x_{12}\end{bmatrix},\quad g_2(x_2)=\begin{bmatrix} 0 \\ x_{21}x_{22}\end{bmatrix}$$

$$g_3(x_3)=\begin{bmatrix} 0 \\ 0.5x_{31}x_{32}\end{bmatrix},\quad g_4(x_4)=\begin{bmatrix} 0 \\ -0.2x_{41}x_{42}\end{bmatrix}$$

$$g_5(x_5)=\begin{bmatrix} 0 \\ 1.4x_{51}x_{52}\end{bmatrix} \tag{34}$$

with $\varepsilon=0.5$ and the leader dynamics is given as follows

$$f(x_0)=\begin{pmatrix} x_{02} \\ -x_{01}+\varepsilon(1-x_{01}^2)x_{02}\end{pmatrix}. \tag{35}$$

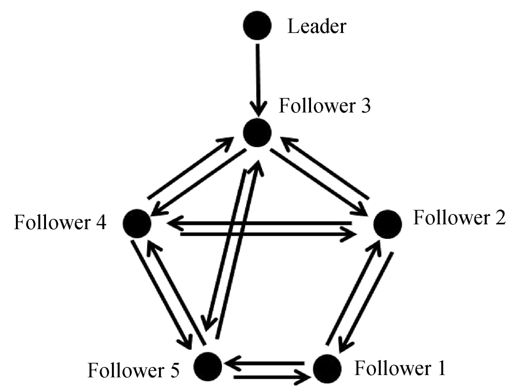

Fig. 1. The multi-agent systems communication graph.

Define the distributed local cost functions of followers, for $i=1,\ldots,5$, as in (5), where $Q_i(\delta_i)=\delta_i^T\delta_i$, $R_{ii}=10$, $R_{ij}=1$, $(i\neq j,\, j\in N_i)$. The learning rates are selected as $\alpha_i=1$, for $i=1,\ldots,5$. The tuning parameters are selected as $F_{1i}=0.1I$, $F_{2i}=[F_{11},F_{12},F_{13},F_{14},F_{15}]$, for $i=1,\ldots,5$, and $\lambda_1=0.6, \lambda_2=10,\ \lambda_3=10, \lambda_4=0.7, \lambda_5=12$.

The critic NN activation functions for $i=1,\ldots,5$ are chosen as follows

$$\sigma_i=[\delta_{i1}^2,\delta_{i1}\delta_{i2},\delta_{i2}^2,\delta_{i1}^4,\delta_{i1}^3\delta_{i2},\delta_{i1}^2\delta_{i2}^2,\delta_{i1}\delta_{i2}^3,\delta_{i2}^4]. \tag{36}$$

To show that no initial stabilizing control policies are needed for implementing the proposed learning algorithm, all critic NNs weights are initialized to zero. To ensure PE condition, a small exponentially decreasing probing noise is added to control inputs. Figs. 2 and 3 show the local tracking errors of followers.

Note that in Figs. 2 and 3 the local tracking errors of all followers vanish and all of them synchronize to the leader. Fig. 4 shows the phase plane plots of the followers' states. It is shown that in Fig. 4 the followers are being synchronized to the leader.

Figs. 5 and 6 show the followers critic NN weights convergence. Simulation results show that the proposed learning algorithm can learn the policies which guarantee the synchronization and the closed-loop stability without the requirement for initial stabilizing control policies.

As we claimed earlier, the proposed scheme has less computational demanding in comparison with the method in [40]. To justify our claim, the method in [40] and our method are applied to the systems (34) and (35) with the communication graph as shown in Fig. 1. Moreover, initial condition for states

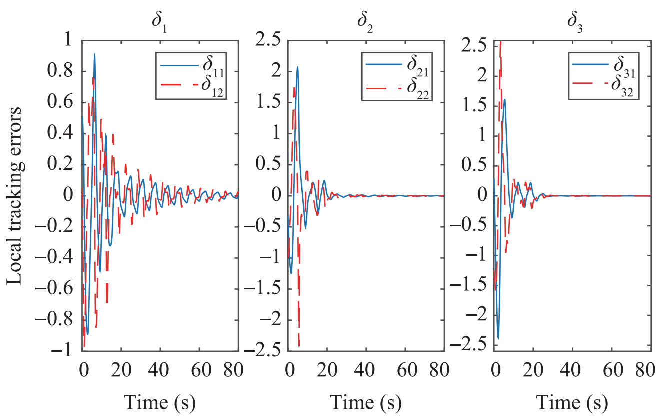

Fig. 2. Local tracking errors of the first, second and third followers.

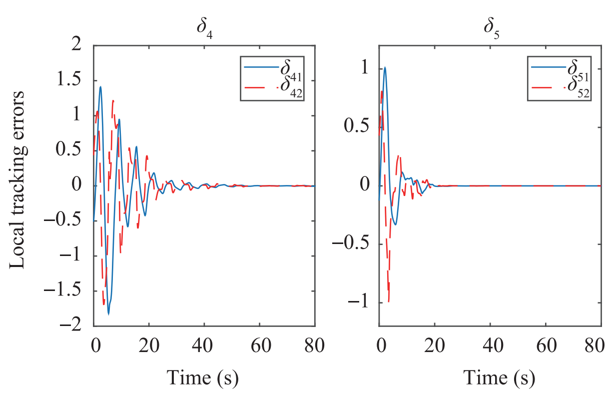

Fig. 3. Local tracking errors of the fourth, fifth followers.

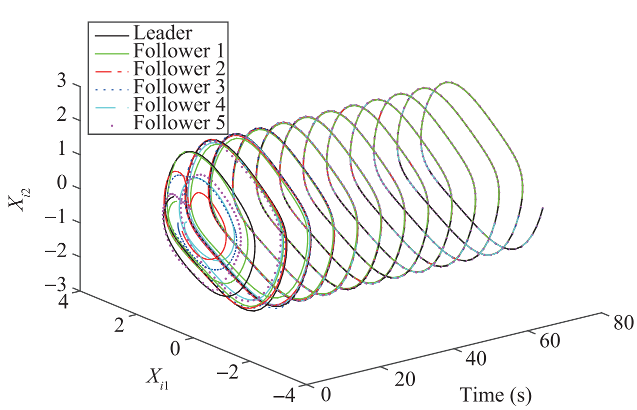

Fig. 4. The evolution of the followers states.

and critic NN weights of followers are chosen similarly. The critic NN activation functions for $i = 1, \ldots, 5$ are chosen as (36). For the method in [40], the actor NN activation functions are $\sigma_i^{\text{actor}} = \nabla \sigma_i$ for $i = 1, \ldots, 5$. For both methods, One select $Q_i(\delta_i) = \delta_i^T \delta_i$, $R_{ii} = 10$, $R_{ij} = 1$, $(i \neq j,\ j \in N_i)$ for $i = 1, \ldots, 5$. For the method in [40], the tuning gains picked all as one. For our method, $\alpha_i = 1$ and the tuning parameters are selected as $F_{1i} = 0.1I$, $F_{2i} = [F_{11}, F_{12}, F_{13}, F_{14}, F_{15}]$, for $i = 1, \ldots, 5$, and $\lambda_1 = 0.6, \lambda_2 = 10, \lambda_3 = 10, \lambda_4 = 0.7, \lambda_5 = 12$. For comparison of performances, the evaluation functions are defined as follow

$$J(i) = \sum_{K=1}^{N_S} \left\{ \|\delta_i(K)\| + R_{ii} \|\hat{u}_i(K)\| + \sum_{j \in N_i} R_{ij} \|\hat{u}_j(K)\| \right\} \tag{37}$$

for $i = 1, \ldots, 5$, where N_S is the number of samples.

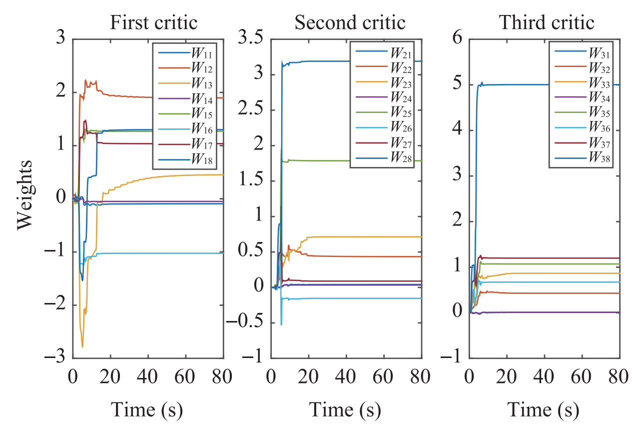

Fig. 5. Crititc NN weights convergence of the first, second and third followers.

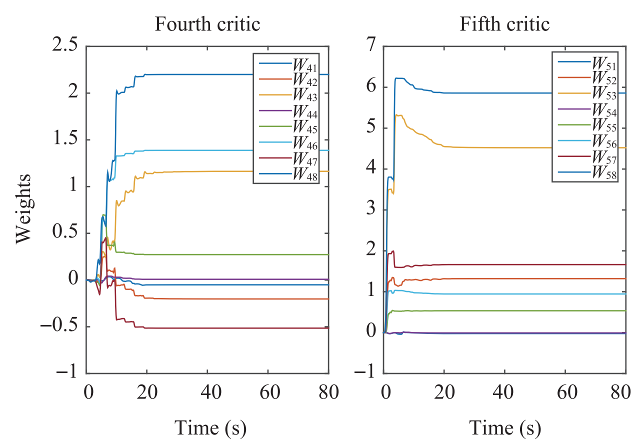

Fig. 6. Crititc NN weights convergence of the fourth and fifth followers.

Table I compares the proposed method and the method in [40] regarding the evaluation functions (37) and the amount of time taken by these two methods. As can be seen in Table I, the method proposed in this paper in comparison with the method in [40] has less computational demand and hence it obtains better performance.

TABLE I
COMPARISON BETWEEN THE PROPOSED METHOD AND THE ONE PROPOSED IN [40]

	Method in this paper	Method in [40]
$J(1)$	458.6638	530.4192
$J(2)$	754.4604	802.2001
$J(3)$	800.1485	876.4293
$J(4)$	502.2877	511.8870
$J(5)$	453.4012	504.7867
Time (S)	12.9386	15.3271

V. CONCLUSION

In this paper, an online optimal distributed learning algorithm is developed to solve leader-synchronization problem of nonlinear multi-agent differential graphical games using single network ADP for every agent. The proposed algorithm guarantees the overall closed-loop system stability and convergence of the policies to the Nash equilibrium without the requirement of initial stabilizing control policies. Lyapunov stability theory is employed to show the uniform ultimate boundedness of closed-loop signals of the system. Finally, simulation results show the effectiveness of the proposed algorithm.

For future work, we intend to extend the approach of this paper to obtain the online optimal distributed synchronization control for nonlinear networked systems subject to dynamics uncertainties in the differential graphical games framework.

APPENDIX A
PROOF OF THEOREM 1

Take the Lyapunov function

$$L=\sum_{i=1}^{N}\left\{L_i(\delta_i)+\frac{1}{2}\lambda_i\tilde{W}_i^T{\alpha_i}^{-1}\tilde{W}_i\right\} \tag{38}$$

where $L_i(\delta_i)$, for $i=1,\ldots,N$ are given in Lemma 1.

The derivative of Lyapunov function is given by

$$\dot{L}=\sum_{i=1}^{N}\left\{{\nabla L_i}^T\dot{\delta}_i+\lambda_i\tilde{W}_i^T{\alpha_i}^{-1}\dot{\tilde{W}}_i\right\}. \tag{39}$$

By using (21), (30) and (31), we have

$$\begin{aligned}
\dot{\tilde{W}}_i =&\alpha_i\frac{\bar{B}_i}{m_{s_i}}\Big(\frac{1}{2}(d_i+\gamma_{i0})^2\tilde{W}_i^T\nabla\sigma_iD_i\nabla\sigma_i^T\tilde{W}_i-\tilde{W}_i^TB_i\\
&+\frac{1}{2}\sum_{j\in N_i^I}(d_j+\gamma_{j0})^2\tilde{W}_j^T\nabla\sigma_jS_{ij}\nabla\sigma_j^T\tilde{W}_j+\varepsilon_{HJ_i}\\
&-\sum_{j\in N_i^I}(d_j+\gamma_{j0})^2W_j^T\nabla\sigma_jS_{ij}\nabla\sigma_j^T\tilde{W}_j\\
&-W_i^T\nabla\sigma_i\sum_{j\in N_i^I}\gamma_{ij}(d_j+\gamma_{j0})D_j\nabla\sigma_j^T\hat{W}_j\Big)\\
&-\frac{1}{2}\alpha_i(d_i+\gamma_{i0})^2\nabla\sigma_iD_i\nabla\sigma_i^T\hat{W}_i\frac{\bar{B}_i^T}{m_{s_i}}\hat{W}_i\\
&-\frac{1}{2}\alpha_i\lambda_i^{-1}(d_i+\gamma_{i0})^2\nabla\sigma_i\sum_{j\in N_i^O}\lambda_j\hat{W}_j^T\frac{\bar{B}_j}{m_{s_j}}S_{ji}\nabla\sigma_i^T\hat{W}_i\\
&+\bar{\chi}_i\Big(\lambda_i^{-1}\alpha_i(d_i+\gamma_{i0})\nabla\sigma_iD_i\big(\sum_{j\in N_i^O}\gamma_{ji}\nabla L_j\\
&-(d_i+\gamma_{i0})\nabla L_i\big)\Big)-\lambda_i^{-1}\alpha_i(d_i+\gamma_{i0})\nabla\sigma_iD_i\\
&\times\Big(\bar{\chi}_i\sum_{j\in N_i^O}\gamma_{ji}\chi_j\nabla L_j-\chi_i\sum_{j\in N_i^O}\gamma_{ji}\bar{\chi}_j\nabla L_j\Big)\\
&+\alpha_iF_{1i}\frac{\nabla\sigma_i\nabla\sigma_i^T}{1+\left\|\nabla\sigma_i\nabla\sigma_i^T\right\|}\hat{W}_i\\
&+\alpha_iF_{2i}\begin{bmatrix}\gamma_{i1}\frac{\nabla\sigma_1\nabla\sigma_1^T}{1+\left\|\nabla\sigma_1\nabla\sigma_1^T\right\|}\hat{W}_1\\ \vdots\\ \gamma_{iN}\frac{\nabla\sigma_N\nabla\sigma_N^T}{1+\left\|\nabla\sigma_N\nabla\sigma_N^T\right\|}\hat{W}_N\end{bmatrix}.
\end{aligned} \tag{40}$$

Substituting (40) in (39), yields

$$\begin{aligned}
\dot{L}=&\sum_{i=1}^{N}\Bigg\{\nabla L_i^T\Big(\sum_{j\in N_i^I}\gamma_{ij}\big(f_i(x_i)-f_j(x_j)\big)\\
&+\gamma_{i0}\big(f_i(x_i)-f_0(x_0)\big)-(d_i+\gamma_{i0})^2D_i\nabla\sigma_i^T\hat{W}_i\\
&+\sum_{j\in N_i^I}\gamma_{ij}(d_j+\gamma_{j0})D_j\nabla\sigma_j^T\hat{W}_j\Big)\\
&-\tilde{W}_i^T\bar{B}_i\lambda_i\bar{B}_i^T\tilde{W}_i+\lambda_i\tilde{W}_i^T\bar{B}_i\frac{\varepsilon_{HJ_i}}{m_{s_i}}\\
&-\frac{1}{2}\tilde{W}_i^T\lambda_i(d_i+\gamma_{i0})^2\nabla\sigma_iD_i\nabla\sigma_i^TW_i\frac{\bar{B}_i^T}{m_{s_i}}W_i\\
&+\frac{1}{2}\lambda_i\tilde{W}_i^T(d_i+\gamma_{i0})^2\nabla\sigma_iD_i\nabla\sigma_i^TW_i\frac{\bar{B}_i^T}{m_{s_i}}\tilde{W}_i\\
&+\frac{1}{2}\lambda_i\tilde{W}_i^T(d_i+\gamma_{i0})^2\frac{\bar{B}_i^T}{m_{s_i}}W_i\nabla\sigma_iD_i\nabla\sigma_i^T\tilde{W}_i\\
&-\frac{1}{2}\tilde{W}_i^T(d_i+\gamma_{i0})^2\nabla\sigma_i\sum_{j\in N_i^O}\lambda_jS_{ji}\nabla\sigma_i^TW_i\frac{\bar{B}_j^T}{m_{s_j}}W_j\\
&+\frac{1}{2}\tilde{W}_i^T(d_i+\gamma_{i0})^2\nabla\sigma_i\sum_{j\in N_i^O}\frac{\bar{B}_j^T}{m_{s_j}}W_j\lambda_jS_{ji}\nabla\sigma_i^T\tilde{W}_i\\
&-\frac{1}{2}\lambda_i\tilde{W}_i^T\frac{\bar{B}_i}{m_{s_i}}\\
&\times\begin{bmatrix}\nabla\sigma_1S_{i1}\nabla\sigma_1^TW_1\\ \vdots\\ \nabla\sigma_NS_{iN}\nabla\sigma_N^TW_N\end{bmatrix}^T\begin{bmatrix}e_{i1}(d_1+\gamma_{10})^2\tilde{W}_1\\ \vdots\\ e_{iN}(d_N+\gamma_{N0})^2\tilde{W}_N\end{bmatrix}\\
&+\lambda_i\tilde{W}_i^TF_{1i}\frac{\nabla\sigma_i\nabla\sigma_i^T}{1+\left\|\nabla\sigma_i\nabla\sigma_i^T\right\|}\hat{W}_i-\lambda_i\tilde{W}_i^T\frac{\bar{B}_i}{m_{s_i}}W_i^T\nabla\sigma_i\\
&\times\begin{bmatrix}\nabla\sigma_1D_1\\ \vdots\\ \nabla\sigma_ND_N\end{bmatrix}^T\begin{bmatrix}\gamma_{i1}\left(d_1+\gamma_{10}\right)\tilde{W}_1\\ \vdots\\ \gamma_{iN}\left(d_N+\gamma_{N0}\right)\tilde{W}_N\end{bmatrix}\\
&+\lambda_i\tilde{W}_i^TF_{2i}\begin{bmatrix}\gamma_{i1}\frac{\nabla\sigma_1\nabla\sigma_1^T}{1+\left\|\nabla\sigma_1\nabla\sigma_1^T\right\|}\hat{W}_1\\ \vdots\\ \gamma_{iN}\frac{\nabla\sigma_N\nabla\sigma_N^T}{1+\left\|\nabla\sigma_N\nabla\sigma_N^T\right\|}\hat{W}_N\end{bmatrix}\Bigg\}\\
&+\sum_{i\in\bar{S}}\Big\{\nabla L_i^T\sum_{j\in N_i^I}\gamma_{ij}(d_j+\gamma_{j0})D_j\nabla\sigma_j^T\hat{W}_j\\
&-\tilde{W}_i^T(d_i+\gamma_{i0})^2\nabla\sigma_iD_i\nabla L_i\Big\}
\end{aligned} \tag{41}$$

where $e_{ij}=1$, if $\gamma_{ij}>0$; otherwise $e_{ij}=0$.

We define $Z^T = \left[\tilde{W}_1^T, \tilde{W}_2^T, \ldots, \tilde{W}_N^T\right]$ and rewrite (41) as follows

$$\begin{aligned}\dot{L} =& -Z^T \overbrace{\begin{bmatrix} m_{11} & \cdots & \cdots & M_{1j} & \cdots & M_{1N} \\ \vdots & \ddots & & \vdots & & \vdots \\ M_{i1} & \cdots & m_{ii} & M_{ij} & \cdots & M_{iN} \\ \vdots & & & \ddots & & \vdots \\ \vdots & & & & \ddots & \vdots \\ M_{N1} & \cdots & \cdots & M_{Nj} & \cdots & m_{NN} \end{bmatrix}}^{M} Z \\ & + Z^T d + \sum_{i=1}^{N} \Big\{ \nabla L_i^T \Big(\sum_{j \in N_i^I} \gamma_{ij} \big(f_i(x_i) - f_j(x_j)\big) \\ & + \gamma_{i0} \big(f_i(x_i) - f_0(x_0)\big) - (d_i + \gamma_{i0})^2 D_i \nabla\sigma_i^T \hat{W}_i \\ & + \sum_{j \in N_i^I} \gamma_{ij} (d_j + \gamma_{j0}) D_j \nabla\sigma_j^T \hat{W}_j \Big) \Big\} \\ & + \sum_{i \in \bar{S}} \Big\{ -\tilde{W}_i^T (d_i + \gamma_{i0})^2 \nabla\sigma_i D_i \nabla L_i \\ & + \nabla L_i^T \sum_{j \in N_i^I} \gamma_{ij} (d_j + \gamma_{j0}) D_j \nabla\sigma_j^T \tilde{W}_j \Big\}. \end{aligned} \tag{42}$$

The components of the M and $d^T = \begin{bmatrix} d_1^T & d_2^T & \cdots & d_N^T \end{bmatrix}$ are given by

$$\begin{aligned} m_{ii} =& -\frac{1}{2}(d_i + \gamma_{i0})^2 \nabla\sigma_i \sum_{j \in N_i^O} \frac{\bar{B}_j^T}{m_{s_j}} W_j \lambda_j S_{ji} \nabla\sigma_i^T \\ & - \frac{1}{2}\lambda_i (d_i + \gamma_{i0})^2 \nabla\sigma_i D_i \nabla\sigma_i^T W_i \frac{\bar{B}_i^T}{m_{s_i}} \\ & - \frac{1}{2}\lambda_i (d_i + \gamma_{i0})^2 \frac{\bar{B}_i^T}{m_{s_i}} W_i \nabla\sigma_i D_i \nabla\sigma_i^T \\ & + \bar{B}_i \lambda_i \bar{B}_i^T + \lambda_i F_{1i} \frac{\nabla\sigma_i \nabla\sigma_i^T}{1 + \left\|\nabla\sigma_i \nabla\sigma_i^T\right\|} \end{aligned} \tag{43}$$

$$M_{ij} \triangleq \frac{m_{ij} + {m_{ji}}^T}{2} \tag{44}$$

$$\begin{aligned} m_{ij} =& \lambda_i \gamma_{ij} (d_j + \gamma_{j0}) \frac{\bar{B}_i}{m_{s_i}} W_i^T \nabla\sigma_i D_j \nabla\sigma_j^T \\ & + \lambda_i F_{2i} \begin{bmatrix} 0 & \cdots & 0 & 0 \\ 0 & 0 & \gamma_{ij} & \vdots \\ \vdots & 0 & \ddots & 0 \\ 0 & \cdots & 0 & 0 \end{bmatrix} \otimes I_{K_i} \begin{bmatrix} \frac{\nabla\sigma_1 \nabla\sigma_1^T}{1+\left\|\nabla\sigma_1 \nabla\sigma_1^T\right\|} \\ \vdots \\ \frac{\nabla\sigma_N \nabla\sigma_N^T}{1+\left\|\nabla\sigma_N \nabla\sigma_N^T\right\|} \end{bmatrix} \\ & + \frac{1}{2}\lambda_i e_{ij} (d_j + \gamma_{j0})^2 \frac{\bar{B}_i}{m_{s_i}} W_j^T \nabla\sigma_j S_{ij} \nabla\sigma_j^T \end{aligned} \tag{45}$$

$$\begin{aligned} d_i =& -\frac{1}{2}\lambda_i (d_i + \gamma_{i0})^2 \nabla\sigma_i D_i \nabla\sigma_i^T W_i \frac{\bar{B}_i^T}{m_{s_i}} W_i \\ & - \frac{1}{2}(d_i + \gamma_{i0})^2 \nabla\sigma_i \sum_{j \in N_i^O} \lambda_j S_{ji} \nabla\sigma_i^T W_i \frac{\bar{B}_j^T}{m_{s_j}} W_j \\ & + \lambda_i \bar{B}_i \frac{\varepsilon_{HJ_i}}{m_{s_i}} + \lambda_i F_{1i} \frac{\nabla\sigma_i \nabla\sigma_i^T}{1 + \left\|\nabla\sigma_i \nabla\sigma_i^T\right\|} W_i \\ & + \lambda_i F_{2i} \begin{bmatrix} \gamma_{i1} \frac{\nabla\sigma_1 \nabla\sigma_1^T}{1+\left\|\nabla\sigma_1 \nabla\sigma_1^T\right\|} W_1 \\ \vdots \\ \gamma_{iN} \frac{\nabla\sigma_N \nabla\sigma_N^T}{1+\left\|\nabla\sigma_N \nabla\sigma_N^T\right\|} W_N \end{bmatrix} \end{aligned} \tag{46}$$

where I_{K_i} denotes the identity matrix of dimension $K_i \times K_i$ and $\otimes$ represents the Kronecker product. Let the tuning parameters λ_i, F_{1i} and F_{2i}, for $i = 1, \ldots, N$ be chosen such that $M > 0$.

According to Assumption 4, we have $\|\delta_i\| > 0$, which guarantees the existence of constants δ_{idmin} satisfying $0 < \delta_{\mathrm{idmin}} < \left\|\dot{\delta}_i\right\|$. Therefore, we have

$$\begin{aligned} & \sum_{i \in S} \Big\{ \nabla L_i^T \Big(\sum_{j \in N_i^I} \gamma_{ij} \big(f_i(x_i) - f_j(x_j)\big) + \gamma_{i0} \big(f_i(x_i) \\ & \quad - f_0(x_0)\big) - (d_i + \gamma_{i0})^2 D_i \nabla\sigma_i^T \hat{W}_i \\ & \quad + \sum_{j \in N_i^I} \gamma_{ij} (d_j + \gamma_{j0}) D_j \nabla\sigma_j^T \hat{W}_j \Big) \Big\} \\ & < -\sum_{i \in S} \left\{ \delta_{id\min} \|\nabla L_i\| \right\} < 0. \end{aligned} \tag{47}$$

According to Assumption 5 and the fact that $\bar{B}_i < 1$, for $i = 1, \ldots, N$, it can be shown that $\|d\| \le d_M$ where d_M is a positive constant. Now, (42) becomes

$$\begin{aligned} \dot{L} \le& -\|Z\|^2 \zeta_{\min}(M) + \|Z\|\, d_M - \sum_{i \in S} \left\{ \delta_{id\min} \|\nabla L_i\| \right\} \\ & + \sum_{i \in \bar{S}} \Big\{ \nabla L_i^T \Big(\sum_{j \in N_i^I} \gamma_{ij} \big(f_i(x_i) - f_j(x_j)\big) \\ & + \gamma_{i0} \big(f_i(x_i) - f_0(x_0)\big) - (d_i + \gamma_{i0})^2 D_i \nabla\sigma_i^T \hat{W}_i \\ & + \sum_{j \in N_i^I} \gamma_{ij} (d_j + \gamma_{j0}) D_j \nabla\sigma_j^T \hat{W}_j \Big) \Big\} \\ & + \sum_{i \in \bar{S}} \Big\{ \nabla L_i^T \sum_{j \in N_i^I} \gamma_{ij} (d_j + \gamma_{j0}) D_j \nabla\sigma_j^T \tilde{W}_j \\ & - \tilde{W}_i^T (d_i + \gamma_{i0})^2 \nabla\sigma_i D_i \nabla L_i \Big\} \end{aligned} \tag{48}$$

where $\zeta_{\min}(M)$ is the minimum singular value of matrix M.

Using (11) and (20) as well as adding and subtracting the following terms

$$\begin{aligned} \sum_{i \in \bar{S}} \Big\{ {\nabla L_i}^T \Big(& \sum_{j \in N_i^I} \gamma_{ij} (d_j + \gamma_{j0}) D_j \nabla\varepsilon_j \\ & -(d_i + \gamma_{i0})^2 D_i \nabla\varepsilon_i \Big) \Big\} \end{aligned} \tag{49}$$

to the right side of (48), we obtain

$$\begin{aligned}\dot{L} \leq & -\|Z\|^2\zeta_{\min}(M) + \|Z\|\, d_M - \sum_{i\in S}\{\delta_{id\min}\|\nabla L_i\|\} \\ & + \sum_{i\in\bar{S}}\Big\{\nabla L_i^T\Big(\sum_{j\in N_i^I}\gamma_{ij}\big(f_i(x_i) - f_j(x_j)\big) \\ & + \gamma_{i0}\big(f_i(x_i) - f_0(x_0)\big) + (d_i+\gamma_{i0})g_i(x_i)u_i^* \\ & - \sum_{j\in N_i^I}\gamma_{ij}g_j(x_j)u_j^* + (d_i+\gamma_{i0})^2 D_i\nabla\varepsilon_i \\ & - \sum_{j\in N_i^I}\gamma_{ij}(d_j+\gamma_{j0})D_j\nabla\varepsilon_j\Big)\Big\}. \end{aligned} \tag{50}$$

By employing Lemma 1, (50) is rewritten as follows

$$\begin{aligned}\dot{L} \leq & -\sum_{i\in S}\{\delta_{id\min}\|\nabla L_i\|\} - \sum_{i\in\bar{S}}\left\{\bar{C}_i(\|\nabla L_i\| - \frac{\eta_i}{2\bar{C}_i})^2\right\} \\ & - \zeta_{\min}(M)(\|Z\| - \frac{d_M}{2\zeta_{\min}(M)})^2 + \sum_{i\in\bar{S}}\{\frac{{\eta_i}^2}{4\bar{C}_i}\} \\ & + \frac{{d_M}^2}{4\zeta_{\min}(M)} \end{aligned} \tag{51}$$

where $\eta_i = (d_i+\gamma_{i0})^2 D_{iM}\varepsilon_{idM} + \sum_{j\in N_i^I}\gamma_{ij}(d_j + \gamma_{j0})D_{jM}\varepsilon_{jdM}$. It should be noted that d and $\nabla\varepsilon_i$, for $i = 1,\ldots,N$ are bounded.

Now, if one of the following inequalities hold

$$\|\nabla L_{i\in S}\| > \sqrt{\frac{(\sum\limits_{i\in\bar{S}}\{\frac{{\eta_i}^2}{4\bar{C}_i}\} + \frac{{d_M}^2}{4\zeta_{\min}(M)})}{\delta_{id\min}}} \triangleq B_{\nabla L_i}^S \tag{52}$$

$$\|\nabla L_{i\in\bar{S}}\| > \sqrt{\frac{(\sum\limits_{i\in\bar{S}}\{\frac{{\eta_i}^2}{4\bar{C}_i}\} + \frac{{d_M}^2}{4\zeta_{\min}(M)})}{\bar{C}_i}} + \frac{\eta_i}{2\bar{C}_i} \triangleq B_{\nabla L_i}^{\bar{S}} \tag{53}$$

$$\|Z\| > \sqrt{\frac{(\sum\limits_{i\in\bar{S}}\{\frac{{\eta_i}^2}{4\bar{C}_i}\} + \frac{{d_M}^2}{4\zeta_{\min}(M)})}{\zeta_{\min}(M)}} + \frac{d_M}{2\zeta_{\min}(M)} \triangleq B_Z \tag{54}$$

then $\dot{L} < 0$. Hence, according to Lyapunov's stability theory [54], we conclude that if $\|Z\| > B_Z$ or $\|\nabla L_i\| > \max(B_{\nabla L_i}^S, B_{\nabla L_i}^{\bar{S}}) \triangleq \bar{B}_{\nabla L_i}$ hold for any i, $i = 1,\ldots,N$ then $\dot{L} < 0$, $\|\nabla L_i\|$ and $\|Z\|$ are UUB, i.e., $\|\nabla L_i\| < \bar{B}_{\nabla L_i}$, for $i = 1,\ldots,N$ and $\|Z\| < B_Z$. Note that, the critic NN weight estimation errors $\left\|\tilde{W}_i\right\|$ are also bounded by B_Z, since $\|Z\| < B_Z$. According to Assumption 3, $\|\nabla L_i\| < \bar{B}_{\nabla L_i}$ implies the boundedness of $\|\delta_i\|$, for $i = 1,\ldots,N$.

Appendix B
Proof of Corollary 1

According to Assumption 4 and the boundedness of $\left\|\tilde{W}_i\right\|$ and using (11) and (26), we have

$$\begin{aligned}\|\hat{u}_i - u_i^*\| & \leq \left\|(d_i+\gamma_{i0})R_{ii}^{-1}g_i^T(x_i)\nabla\sigma_i^T\tilde{W}_i\right\| \\ & \leq (d_i+\gamma_{i0})\lambda_{\max}(R_{ii}^{-1})g_{iM}\sigma_{idM}B_Z \triangleq \in_{u_i} \end{aligned} \tag{55}$$

where $\lambda_{\max}(R_{ii}^{-1})$ is the maximum eigenvalue of matrix R_{ii}^{-1}. This completes the proof.

References

[1] R. Olfati-Saber and R. M. Murray, "Consensus problems in networks of agents with switching topology and time-delays," *IEEE Trans. Automat. Control*, vol. 49, no. 9, pp. 1520−1533, Sep. 2004.

[2] W. Ren and R. W. Beard, "Consensus seeking in multiagent systems under dynamically changing interaction topologies," *IEEE Trans. Automat. Control*, vol. 50, no. 5, pp. 655−661, May 2005.

[3] W. Ren and R. W. Beard, *Distributed Consensus in Multi-Vehicle Cooperative Control: Theory and Applications*. Berlin, Germany: Springer-Verlag, 2008.

[4] J. A. Fax and R. M. Murray, "Information flow and cooperative control of vehicle formations," *IEEE Trans. Automat. Control*, vol. 49, no. 9, pp. 1465−1476, Sep. 2004.

[5] A. Jadbabaie, J. Lin, and A. S. Morse, "Coordination of groups of mobile autonomous agents using nearest neighbor rules," *IEEE Trans. Automat. Control*, vol. 48, no. 6, pp. 988−1001, Jun. 2003.

[6] Z. H. Qu, *Cooperative Control of Dynamical Systems: Applications to Autonomous Vehicles*. New York, USA: Springer-Verlag, 2009.

[7] F. L. Lewis, H. W. Zhang, K. Hengster-Movric, and A. Das, *Cooperative Control of Multi-Agent Systems: Optimal and Adaptive Design Approaches*. Berlin, Germany: Spring-Verlag, 2014.

[8] M. Defoort, T. Floquet, A. Kokosy, and W. Perruquetti, "Sliding-mode formation control for cooperative autonomous mobile robots," *IEEE Trans. Ind. Electron.*, vol. 55, no. 11, pp. 3944−3953, Nov. 2008.

[9] J. Mei, W. Ren, and G. F. Ma, "Distributed containment control for Lagrangian networks with parametric uncertainties under a directed graph," *Automatica*, vol. 48, no. 4, pp. 653−659, Apr. 2012.

[10] W. Lin, "Distributed UAV formation control using differential game approach," *Aerosp. Sci. Technol.*, vol. 35, pp. 54−62, May 2014.

[11] R. Abdolee, B. Champagne, and A. H. Sayed, "Diffusion adaptation over multi-agent networks with wireless link impairments," *IEEE Trans. Mobile Comput.*, vol. 15, no. 6, pp. 1362−1376, Jun. 2016.

[12] W. Q. Wang, "Carrier frequency synchronization in distributed wireless sensor networks," *IEEE Syst. J.*, vol. 9, no. 3, pp. 703−713, Sep. 2015.

[13] S. M. Mu, T. G. Chu, and L. Wang, "Coordinated collective motion in a motile particle group with a leader," *Phys. A*, vol. 351, no. 2−4, pp. 211−226, Jun. 2005.

[14] V. Nasirian, S. Moayedi, A. Davoudi, and F. L. Lewis, "Distributed cooperative control of DC microgrids," *IEEE Trans. Power Electron.*, vol. 30, no. 4, pp. 2288−2303, Apr. 2015.

[15] L. L. Fan, V. Nasirian, H. Modares, F. L. Lewis, Y. D. Song, and A. Davoudi, "Game-theoretic control of active loads in DC microgrids," *IEEE Trans. Energy Convers.*, vol. 31, no. 3, pp. 882−895, Sep. 2016.

[16] D. M. Xie and J. H. Chen, "Consensus problem of data-sampled networked multi-agent systems with time-varying communication delays," *Trans. Inst. Meas. Control*, vol. 35, no. 6, pp. 753−763, Mar. 2013.

[17] S. Y. Tu and A. H. Sayed, "Diffusion strategies outperform consensus strategies for distributed estimation over adaptive networks," *IEEE Trans. Signal Process.*, vol. 60, no. 12, pp. 6217−6234, Dec. 2012.

[18] H. W. Zhang, F. L. Lewis, and Z. H. Qu, "Lyapunov, adaptive, and optimal design techniques for cooperative systems on directed communication graphs," *IEEE Trans. Ind. Electron.*, vol. 59, no. 7, pp. 3026−3041, Jul. 2012.

[19] W. Ren, R. W. Beard, and E. M. Atkins, "Information consensus in multivehicle cooperative control," *IEEE Control Syst.*, vol. 27, no. 2, pp. 71−82, Apr. 2007.

[20] Z. J. Tang, "Leader-following consensus with directed switching topologies," *Trans. Inst. Meas. Control*, vol. 37, no. 3, pp. 406−413, Jul. 2015.

[21] A. R. Wei, X. M. Hu, and Y. Z. Wang, "Tracking control of leader-follower multi-agent systems subject to actuator saturation," *IEEE/CAA J. Automat. Sin.*, vol. 1, no. 1, pp. 84−91, Jan. 2014.

[22] C. H. Zhang, L. Chang, and X. F. Zhang, "Leader-follower consensus of upper-triangular nonlinear multi-agent systems," *IEEE/CAA J. Automat. Sin.*, vol. 1, no. 2, pp. 210−217, Apr. 2014.

[23] C. R. Wang, X. H. Wang, and H. B. Ji, "A continuous leader-following consensus control strategy for a class of uncertain multi-agent systems," *IEEE/CAA J. Automat. Sin.*, vol. 1, no. 2, pp. 187−192, Apr. 2014.

[24] Y. G. Hong, J. P. Hu, and L. X. Gao, "Tracking control for multi-agent consensus with an active leader and variable topology," *Automatica*, vol. 42, no. 7, pp. 1177−1182, Jul. 2006.

[25] G. Owen, *Game Theory*. New York, USA: Academic Press, 1982.

[26] T. Basar and G. J. Olsder, *Dynamic Noncooperative Game Theory (Classics in Applied Mathematics)*. Philadelphia, PA, USA: SIAM, 1999.

[27] E. Semsar-Kazerooni and K. Khorasani, "Multi-agent team cooperation: A game theory approach," *Automatica*, vol. 45, no. 10, pp. 2205−2213, Oct. 2009.

[28] C. X. Jiang, Y. Chen, and K. J. R. Liu, "Distributed adaptive networks: A graphical evolutionary game-theoretic view," *IEEE Trans. Signal Process.*, vol. 61, no. 22, pp. 5675−5688, Nov. 2013.

[29] C. X. Jiang, Y. Chen, Y. Gao, and K. J. R. Liu, "Indian buffet game with negative network externality and non-Bayesian social learning," *IEEE Trans. Syst. Man Cybern. Syst.*, vol. 45, no. 4, pp. 609−623, Apr. 2015.

[30] R. Kamalapurkar, J. R. Klotz, and W. E. Dixon, "Concurrent learning-based approximate feedback-Nash equilibrium solution of N-player nonzero-sum differential games," *IEEE/CAA J. Automat. Sin.*, vol. 1, no. 3, pp. 239−247, Jul. 2014.

[31] K. G. Vamvoudakis, F. L. Lewis, and G. R. Hudas, "Multi-agent differential graphical games: Online adaptive learning solution for synchronization with optimality," *Automatica*, vol. 48, no. 8, pp. 1598−1611, Aug. 2012.

[32] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*. Cambridge, MA, USA: MIT Press, 1998.

[33] P. J. Werbos, "Approximate dynamic programming for real-time control and neural modeling," *Handbook of Intelligent Control*, D. A. White and D. A. Sofge, Eds. New York, USA: Van Nostrand Reinhold, 1992.

[34] J. J. Murray, C. J. Cox, G. G. Lendaris, and R. Saeks, "Adaptive dynamic programming," *IEEE Trans. Syst. Man Cybern. C*, vol. 32, no. 2, pp. 140−153, May 2002.

[35] H. Modares, F. L. Lewis, and M. B. Naghibi-Sistani, "Integral reinforcement learning and experience replay for adaptive optimal control of partially-unknown constrained-input continuous-time systems," *Automatica*, vol. 50, no. 1, pp. 193−202, Jan. 2014.

[36] H. Modares and F. L. Lewis, "Optimal tracking control of nonlinear partially-unknown constrained-input systems using integral reinforcement learning," *Automatica*, vol. 50, no. 7, pp. 1780−1792, Jul. 2014.

[37] Z. P. Jiang and Y. Jiang, "Robust adaptive dynamic programming for linear and nonlinear systems: An overview," *Eur. J. Control*, vol. 19, no. 5, pp. 417−425, Sep. 2013.

[38] S. Bhasin, R. Kamalapurkar, M. Johnson, K. G. Vamvoudakis, F. L. Lewis, and W. E. Dixon, "A novel actorCcriticCidentifier architecture for approximate optimal control of uncertain nonlinear systems," *Automatica*, vol. 49, no. 1, pp. 82−92, Jan. 2013.

[39] K. G. Vamvoudakis and F. L. Lewis, "Online actor-critic algorithm to solve the continuous-time infinite horizon optimal control problem," *Automatica*, vol. 46, no. 5, pp. 878−888, May 2010.

[40] F. Tatari, M. B. Naghibi-Sistani, and K. G. Vamvoudakis, "Distributed learning algorithm for non-linear differential graphical games," *Trans. Inst. Meas. Control*, Vol 39, no. 2, pp. 173−182, Feb. 2017.

[41] K. G. Vamvoudakis and F. L. Lewis, "Multi-player non-zero-sum games: Online adaptive learning solution of coupled Hamilton-Jacobi equations," *Automatica*, vol. 47, no. 8, pp. 1556−1569, Aug. 2011.

[42] H. G. Zhang, L. L. Cui, and Y. H. Luo, "Near-optimal control for nonzero-sum differential games of continuous-time nonlinear systems using single-network ADP," *IEEE Trans. Cybern.*, vol. 43, no. 1, pp. 206−216, Feb. 2013.

[43] M. I. Abouheaf and F. L. Lewis, "Multi-agent differential graphical games: Nash online adaptive learning solutions," in *Proc. 52nd Annu. Conf. Decision and Control*. Firenze, Italy, 2013, pp. 5803−5809.

[44] M. I. Abouheaf, F. L. Lewis, and M. S. Mahmoud, "Differential graphical games: Policy iteration solutions and coupled Riccati formulation," in *Proc. 2014 European Control Conf.*. Strasbourg, France, 2014, pp. 1594−1599.

[45] Q. L. Wei, D. R. Liu, and F. L. Lewis, "Optimal distributed synchronization control for continuous-time heterogeneous multi-agent differential graphical games," *Inform. Sci.*, vol. 317, pp. 96−113, Oct. 2015.

[46] F. A. Yaghmaie, F. L. Lewis, and R. Su, "Output regulation of heterogeneous linear multi-agent systems with differential graphical game," *Int. J. Robust Nonlinear Control*, vol. 26, pp. 2256−2278, Jul. 2016.

[47] Q. Jiao, H. Modares, S. Y. Xu, F. L. Lewis, and K. G. Vamvoudakis, "Multi-agent zero-sum differential graphical games for disturbance rejection in distributed control," *Automatica*, vol. 69, pp. 24−34, Jul. 2016.

[48] M. I. Abouheaf, F. L. Lewis, K. G. Vamvoudakis, S. Haesaert, and R. Babuska, "Multi-agent discrete-time graphical games and reinforcement learning solutions," *Automatica*, vol. 50, no. 12, pp. 3038−3053, Dec. 2014.

[49] A. G. Barto, R. S. Sutton, and C. W. Anderson, "Neuronlike adaptive elements that can solve difficult learning control problems," *IEEE Trans. Syst. Man Cybern.*, vol. SMC-13, no. 5, pp. 834−846, Sep.−Oct. 1983.

[50] T. Dierks and S. Jagannathan, "Optimal control of affine nonlinear continuous-time systems using an online Hamilton-Jacobi-Isaacs formulation," in *Proc. 49th Conf. Decision and Control*. Atlanta, GA, USA, 2010, pp. 3048−3053.

[51] 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, May 2005.

[52] F. L. Lewis, D. Vrabie, and V. L. Syrmos, *Optimal Control*. Hoboken, NJ, USA: John Wiley, 2012.

[53] F. L. Lewis, S. Jagannathan, and A. Yeşildirek, *Neural Network Control of Robot Manipulators and Nonlinear Systems*. London, UK: Taylor and Francis, 1999.

[54] H. K. Khalil, *Nonlinear Systems*. Englewood Cliffs, New Jersey, USA: Prentice-Hall, 1996.

[55] B. A. Finlayson, *The Method of Weighted Residuals and Variational Principles*. New York, USA: Academic Press, 1990.

[56] P. Ioannou and B. Fidan. *Adaptive Control Tutorial (Advances in Design and Control)*. Philadelphia, PA: SIAM, 2006.

[57] J. J. E. Slotine and W. P. Li, *Applied Nonlinear Control*. Englewood Cliffs, NJ, USA: Prentice Hall, 1991.

[58] S. Sastry and M. Bodson,*Adaptive Control: Stability, Convergence, and Robustness*. Englewood Cliffs, NJ: Prentice Hall, 1989.

Majid Mazouchi received the B.Sc. degree from K. N. Toosi University of Technology in 2007 and the M.Sc. degree from Ferdowsi University of Mashhad in 2010. He is currently working towards the Ph.D. degree at Ferdowsi University of Mashhad. His research interests include optimal control, reinforcement learning, and cooperative control systems.

Mohammad Bagher Naghibi-Sistani received the B.Sc. and M.Sc. degrees in control engineering with honors from University of Tehran, Tehran, Iran, in 1991 and 1995, respectively, and the Ph.D. degree from the Department of Electrical Engineering at Ferdowsi University of Mashhad in 2005. He is now an Associate Professor at Ferdowsi University of Mashhad, Mashhad, Iran. His research interests include reinforcement learning, cooperative control systems, and optimization.

Seyed Kamal Hosseini Sani received the B.Sc. degree from Ferdowsi University of Mashhad in 1995, the M.Sc. degree from K. N. Toosi University of Technology in 1998, and the Ph.D. degree from Tarbiat Modares University in 2006. He is now an Associate Professor at Ferdowsi University of Mashhad, Mashhad, Iran. His research interests are adaptive control, model predictive control, and renewable energy.