# Leader-to-formation stability of multi-agent systems: An adaptive optimal control approach

Gao, W.; Jiang, Z.-P.; Lewis, F.; Wang, Y.

## Abstract

This note proposes a novel data-driven solution to the cooperative adaptive optimal control problem of leaderfollower multi-agent systems under switching network topology. The dynamics of all the followers are unknown, and the leader is modeled by a perturbed exosystem. Through the combination of adaptive dynamic programming and internal model principle, an approximate optimal controller is iteratively learned online using real-time input-state data. Rigorous stability analysis shows that the system in closed-loop with the developed control policy is leader-to-formation stable, with guaranteed robustness to unmeasurable leader disturbance. Numerical results illustrate the effectiveness of the proposed data-driven algorithm.

# Leader-to-formation stability of multi-agent systems: An adaptive optimal control approach

Weinan Gao, Zhong-Ping Jiang, *Fellow, IEEE*, Frank L. Lewis, *Fellow, IEEE*, and Yebin Wang, *Member, IEEE*

*Abstract*—This note proposes a novel data-driven solution to the cooperative adaptive optimal control problem of leader-follower multi-agent systems under switching network topology. The dynamics of all the followers are unknown, and the leader is modeled by a perturbed exosystem. Through the combination of adaptive dynamic programming and internal model principle, an approximate optimal controller is iteratively learned online using real-time input-state data. Rigorous stability analysis shows that the system in closed-loop with the developed control policy is leader-to-formation stable, with guaranteed robustness to unmeasurable leader disturbance. Numerical results illustrate the effectiveness of the proposed data-driven algorithm.

*Index Terms*—Adaptive dynamic programming (ADP), Optimal tracking control, Leader-to-formation stability, Switching network topology.

## I. INTRODUCTION

The cooperative control problem of leader-follower multi-agent systems has been under extensive investigation in the last decade due to its wide application in electrical, mechanical and biological systems; see [3], [4], [35] and many references therein. A general assumption in the present literature on the topic is that the leader is modeled by an autonomous system without considering the influence of external signals. Reference [29] relaxes this assumption by developing distributed trackers for multi-agent systems with bounded unknown leader input. The notion of leader-to-formation stability (LFS) [41] is introduced to investigate how the leader inputs and disturbances affect the stability of the group. By taking unknown dynamics and partial measurements into account, reference [13] proposes an adaptive control design approach for a class of second-order leader-follower systems. However, the issue of adaptive optimal controller design of the multi-agent systems with assured LFS remains open.

Over the last decade, a trend in adaptive optimal control is to invoke reinforcement learning [40] and approximate/adaptive dynamic programming (ADP) [2], [21], [27], [45] for feedback control of dynamical systems. Among all the different ADP approaches, much attention has been paid to achieving the adaptive optimal stabilization of linear or nonlinear plants via state-feedback [6], [11], [12], [18], [19], [22], [27], [42], [43] and output-feedback [8], [9], [26]. The generalization to adaptive optimal tracking control is studied by [10], [33], [34]. For non-model-based optimal stabilization of large-scale and multi-agent systems, some interesting results appear in [1], [20] using (robust) ADP.

The main purpose of this note is to address the cooperative adaptive optimal control problem of leader-follower multi-agent systems via ADP. The contributions of this note are three-fold. First, considering the more general and realistic case when the leader model (or, the exosystem here) is subject to external disturbance, we develop a data-driven distributed control policy to guarantee that the closed-loop system is leader-to-formation stable. Moreover, given a vanishing leader disturbance, the multi-agent system is able to achieve cooperative output regulation [5], [14], [30], [38], [44] which means each follower asymptotically tracks a desired trajectory, while rejecting the disturbance ($D_i v$ in eq. (2) below) generated by the exosystem. Second, this note, for the first time, combines the idea of ADP and internal model principle to study cooperative adaptive optimal tracking control problems. By means of internal model principle, we convert the tracking problem to a stabilization problem of an augmented system composed of the plant and a dynamic compensator named as internal model. Comparing with our previous work [10] which need solve regulator equations by online data first and then design a feedback-feedforward controller, the proposed algorithm in this note has a reduced computational cost since it need not solve regulator equations. Third, instead of assuming that the communication network remains static and connected, we study a more practical situation where the network is jointly connected [17]. In other words, the network is allowed disconnected at any time instant. To overcome this issue, the estimation of the exostate obtained from a distributed observer is used for feedback design.

The remainder of this note is organized as follows. Section II formulates the problem and introduces some basic results regarding LFS, internal model principle, and optimal control. In Section III, a novel data-driven control approach is presented based on ADP to solve cooperative adaptive optimal control problems for leader-follower multi-agent systems under switching network. The convergence of the proposed algorithm and the LFS of the closed-loop system are rigorously analyzed as well. An example to validate our design is shown in Section IV. Section V provides concluding remarks. For the sake of clarity and readability, the relationship among our main results

W. Gao is with the Department of Electrical and Computer Engineering, Allen E. Paulson College of Engineering and Computing, Georgia Southern University, Statesboro, GA, 30460 USA (e-mail:wgao@georgiasouthern.edu)

Z.P. Jiang is with the Department of Electrical and Computer Engineering, Tandon School of Engineering, New York University, Brooklyn, NY, 11201 USA (e-mail:zjiang@nyu.edu).

F. Lewis is with UTA Research Institute, The University of Texas at Arlington, Texas 76118 USA and Qian Ren Consulting Professor, Northeastern University, Shenyang 110036, China. (e-mail:lewis@uta.edu)

Y. Wang is with Mitsubishi Electric Research Laboratories, Cambridge, MA 02139 USA (e-mail: yebinwang@ieee.org).

is illustrated in Table I.

TABLE I
RELATIONSHIP BETWEEN THE PROPOSITION, LEMMA AND THEOREMS

| | |
|---|---|
| Stability Analysis: | Proposition 1 $\longrightarrow$ Theorem 2 |
| Convergence Analysis: | Lemma 1 $\longrightarrow$ Theorem 1 |
| Optimality Analysis: | Theorem 3 |

**Notations.** Throughout this note, $|\cdot|$ represents the Euclidean norm for vectors and the induced norm for matrices. $\mathbb{C}^-$ stands for the open left-half complex plane. For any piecewise continuous function $u : \mathbb{R}_+ \to \mathbb{R}^m$, $\|u\|$ stands for $\sup_{t \geq 0} |u(t)|$. $\otimes$ indicates the Kronecker product. A continuous function $\alpha : \mathbb{R}_+ \to \mathbb{R}_+$ belongs to class $\mathcal{K}$ if it is increasing and $\alpha(0) = 0$. A continuous function $\beta : \mathbb{R}_+ \times \mathbb{R}_+ \to \mathbb{R}_+$ belongs to class $\mathcal{KL}$ if for each fixed $t$, the function $\beta(\cdot, t)$ is of class $\mathcal{K}$ and, for each fixed $s$, the function $\beta(s, \cdot)$ is non-increasing and tends to 0 at infinity. $\text{vec}(A) = [a_1^T, a_2^T, \cdots, a_m^T]^T$, where $a_i \in \mathbb{R}^n$ is the $i$th column of $A \in \mathbb{R}^{n \times m}$. When $n = m$, $\sigma(A)$ is its complex spectrum. For a symmetric matrix $P \in \mathbb{R}^{m \times m}$, $\text{vecs}(P) = [p_{11}, 2p_{12}, \cdots, 2p_{1m}, p_{22}, 2p_{23}, \cdots, 2p_{m-1,m}, p_{mm}]^T \in \mathbb{R}^{\frac{1}{2}m(m+1)}$. For an arbitrary column vector $v \in \mathbb{R}^m$, $|v|_P$ stands for $v^T P v$, and $\text{vecv}(v) = [v_1^2, v_1 v_2, \cdots, v_1 v_m, v_2^2, v_2 v_3, \cdots, v_{m-1} v_m, v_m^2]^T \in \mathbb{R}^{\frac{1}{2}m(m+1)}$. $\lambda_M(P)$ and $\lambda_m(P)$ denote the maximum and the minimum eigenvalue of a real symmetric matrix $P$. $\rho(t)$ represents a piecewise constant switching signal $\rho : [0, +\infty) \to \{1, 2, \cdots, n_\rho\}$ for some integer $n_\rho > 0$. We assume switching constants $t_0 = 0, t_1, t_2, \cdots$ of $\rho$ satisfy $\inf_i (t_{i+1} - t_i) \geq \tau_d > 0$, $i = 0, 1, \cdots$, with $\lim_{i \to \infty} t_i = \infty$, where $\tau_d$ is called the dwell time.

## II. PRELIMINARIES

### A. Problem Formulation

Consider the following class of linear multi-agent systems

$$\dot{v} = Ev + Hw, \tag{1}$$

$$\begin{aligned} \dot{x}_i &= A_i x_i + B_i u_i + D_i v, \\ e_i &= C_i x_i + F_i v, \quad i = 1, 2 \cdots, N \end{aligned} \tag{2}$$

where $x_i \in \mathbb{R}^{n_i}, u_i \in \mathbb{R}$ and $e_i \in \mathbb{R}$ are the state, control input and tracking error of the $i$th subsystem (follower), respectively. $v \in \mathbb{R}^q$ is the state of the leader, modeled by the exosystem (1) which generates both the disturbance $D_i v$ and the reference signal $-F_i v$ (to be tracked by the output $y_i = C_i x_i$) of each follower. The leader is assumed to track a desired trajectory $y_0^* = -F_0 v^*(t)$ with the signal $v^*$ satisfying $\dot{v}^* = Ev^*$. In this setting, we let the leader input contain two parts: $w = \hat{w} + \tilde{w}$, where $\hat{w}$ is a feedback control input $\hat{w} = -K_0(v - v^*)$ with $\sigma(E - HK_0) \subset \mathbb{C}^-$ and $\tilde{w}$ is an external disturbance input. Given the exosystem (1) and the plant (2), define a time-varying digraph $\mathcal{G}_{\rho(t)} = \{\mathcal{V}, \mathcal{E}_{\rho(t)}\}$. $\mathcal{V} = \{0, 1, \cdots, N\}$ is the node set with node 0 denoting the leader and the remaining $N$ nodes being identified as followers described by (2). $\mathcal{E}_{\rho(t)} \subset \mathcal{V} \times \mathcal{V}$ refers to the edge set. Denote $\mathcal{N}_i(t)$ the set of all the nodes $j$ such that $(j, i) \in \mathcal{E}_{\rho(t)}$. The adjacency matrix

$\mathcal{A}_{\rho(t)} = [a_{ij}(t)] \in \mathbb{R}^{(N+1) \times (N+1)}$ is defined by $a_{ij}(t) > 0$ if $(j, i) \in \mathcal{E}_{\rho(t)}$ and otherwise $a_{ij}(t) = 0$.

Some standard assumptions are made on the system (1)-(2). Similar assumptions can be found in [7], [31], [32], [38], [44] for solving (cooperative) output regulation problems.

**Assumption 1.** *All the eigenvalues of $E$ are simple with zero real part.*

**Assumption 2.** *$(A_i, B_i)$ is stabilizable, $\forall 1 \leq i \leq N$.*

**Assumption 3.** *$\text{rank} \begin{bmatrix} A_i - \lambda I & B_i \\ C_i & 0 \end{bmatrix} = n_i + 1$, $\forall \lambda \in \sigma(E)$, $\forall 1 \leq i \leq N$.*

**Assumption 4.** *There exists a subsequence $\{i_k\}$ of $\{i : i = 0, 1, \cdots\}$ with $t_{i_{k+1}} - t_{i_k} < T$ for some positive $T$ such that each node $j = 1, 2, \cdots, N$ is reachable from node 0 in the union graph $\cup_{l=i_k}^{i_{k+1}-1} \mathcal{G}_{\rho(t_j)}$.*

**Remark 1.** *Assumption 2 is made such that the exponential stability can be achieved for each follower. Assumption 3 is a sufficient condition for the solvability of regulator equations (6)-(7). Assumption 4 is a joint connectivity condition [17], [31], [39] that allows the network disconnected at any time instant.*

### B. Basic Results

Under Assumptions 1-3, LFS can be achieved by system (1)-(2) in closed-loop with a decentralized controller

$$u_i = -K_{xi} x_i - K_{zi} z_i, \tag{3}$$

$$\dot{z}_i = G_1 z_i + G_2 e_i, \quad i = 1, 2, \cdots, N \tag{4}$$

where the characteristic polynomial of $G_1$ is the same as the minimal polynomial of $E$, and the pair $(G_1, G_2)$ is controllable. In this setting, the pair $(G_1, G_2)$ incorporates an internal model of the matrix $E$, and (4) is an internal model of the $i$th follower. For $i = 1, 2, \cdots, N$, matrices $K_{xi}, K_{zi}$ are chosen such that

$$A_{ci} = \begin{bmatrix} A_i - B_i K_{xi} & -B_i K_{zi} \\ G_2 C_i & G_1 \end{bmatrix}$$

is a Hurwitz matrix.

**Remark 2.** *As shown in [15, Lemma 1.26], for $i = 1, 2, \cdots, N$, the pair $(\bar{A}_i, \bar{B}_i)$ is stabilizable under Assumptions 1-2, where*

$$\bar{A}_i = \begin{bmatrix} A_i & 0 \\ G_2 C_i & G_1 \end{bmatrix}, \quad \bar{B}_i = \begin{bmatrix} B_i \\ 0 \end{bmatrix}$$

*which implies one can always find a $K_i := \begin{bmatrix} K_{xi} & K_{zi} \end{bmatrix}$ such that $\sigma(A_{ci}) \subset \mathbb{C}^-$.*

Let $\eta$ be the error between the lumped state of the multi-agent system (1)-(2) with (4) and its desired value. The LFS is then defined as follows. The definition in this note is in light of input-to-output stability [23], [37], which is slightly different from [41].

**Definition 1.** *System (1)-(2) achieves LFS if there exist a function $\beta$ of class $\mathcal{KL}$ and a function $\gamma$ of class $\mathcal{K}$ such that,*

*for any initial state error $\eta(0)$ and any measurable essentially bounded input $\tilde{w}$ and $t \geq 0$:*

$$|e(t)| \leq \beta(|\eta(0)|, t) + \gamma(\|\tilde{w}\|) \qquad (5)$$

*where $e(t) = \begin{bmatrix} e_1(t) & e_2(t) & \cdots & e_N(t) \end{bmatrix}^T$.*

**Remark 3.** *Note that the LFS ensures that, given any bounded disturbance $\tilde{w}$, the tracking error $e$ will be bounded. It also implies that $\lim_{t \to \infty} e(t) = 0$ if $\lim_{t \to \infty} \tilde{w}(t) = 0$, which corresponds to the asymptotic tracking with disturbance rejection arisen in the cooperative output regulation problems [38].*

The following proposition analyzes the LFS of the closed-loop system with respect to $\tilde{w}$.

**Proposition 1.** *Under Assumptions 1-3, for $i = 1, 2, \cdots, N$, the multi-agent system (1)-(2) in closed-loop with (3)-(4) is leader-to-formation stable.*

*Proof.* Under Assumptions 1-3, for $i = 1, 2, \cdots, N$, by [32], there exist uniquely matrices $X_i$ and $U_i$ solving the following regulator equations

$$X_i E = A_i X_i + B_i U_i + D_i, \qquad (6)$$
$$0 = C_i X_i + F_i. \qquad (7)$$

By [15, Lemma 1.27], the matrix equations (7) combined with

$$X_i E = (A_i - B_i K_{xi}) X_i - B_i K_{zi} Z_i + D_i, \qquad (8)$$
$$Z_i E = G_1 Z_i + G_2 (C_i X_i + F_i) \qquad (9)$$

have a unique solution $\hat{X}_i$ and $Z_i$. Then, one can easily check that $X_i = \hat{X}_i$, and $U_i = -K_{xi} X_i - K_{zi} Z_i$. Let $\tilde{v} = \mathbb{1}_N \otimes (v - v^*)$, $\tilde{x}_i = x_i - X_i v^*$, $\tilde{z}_i = z_i - Z_i v^*$, $\tilde{\xi}_i = [\tilde{x}_i^T, \tilde{z}_i^T]^T \in \mathbb{R}^{m_i}$, $B_{ci} = \begin{bmatrix} D_i^T & F_i^T G_2^T \end{bmatrix}^T$ and $\bar{C}_i = \begin{bmatrix} C_i & 0 \end{bmatrix} \in \mathbb{R}^{1 \times m_i}$. Also, let $\tilde{\xi} = [\tilde{\xi}_1^T, \tilde{\xi}_2^T, \cdots, \tilde{\xi}_N^T]^T$, $A_c = \text{blockdiag}(A_{c1}, A_{c2}, \cdots, A_{cN})$, $B_c = \text{blockdiag}(B_{c1}, B_{c2}, \cdots, B_{cN})$, $\bar{C} = \text{blockdiag}(\bar{C}_1, \bar{C}_2, \cdots, \bar{C}_N)$, and $F = \text{blockdiag}(F_1, F_2, \cdots, F_N)$. Then, we have

$$\begin{bmatrix} \dot{\tilde{v}} \\ \dot{\tilde{\xi}} \end{bmatrix} = \begin{bmatrix} I_N \otimes (E - H K_0) & 0 \\ B_c & A_c \end{bmatrix} \begin{bmatrix} \tilde{v} \\ \tilde{\xi} \end{bmatrix} + \begin{bmatrix} I_N \otimes H \\ 0 \end{bmatrix} (\mathbb{1}_N \otimes \tilde{w})$$
$$e = F\tilde{v} + \bar{C}\tilde{\xi}. \qquad (10)$$

By $\eta = [\tilde{v}^T, \tilde{\xi}^T]^T$, it is easily checkable that the LFS of the original multi-agent systems is achieved since $(E - H K_0)$ and $A_c$ are Hurwitz matrices. The proof is thus completed. $\square$

In order to ameliorate the transient performance of each subsystem, we develop a robust optimal controller such that the closed-loop system is leader-to-formation stable with respect to the leader disturbance $\tilde{w}$. Moreover, as $v \equiv v^*$, the developed controller is optimal in the sense that it minimizes the following cost

$$J = \int_0^\infty \left( |\tilde{\xi}|_Q + |\tilde{u}|_R \right) dt \qquad (11)$$

for the open-loop system

$$\dot{\tilde{\xi}} = \bar{A}\tilde{\xi} + \bar{B}\tilde{u} \qquad (12)$$

where, for $i = 1, 2, \cdots, N$, $\tilde{u}_i = u_i - U_i v^*$, $Q_i = Q_i^T > 0$, $R_i = R_i^T > 0$. $\tilde{u} = [\tilde{u}_1, \tilde{u}_2, \cdots, \tilde{u}_N]^T$, $\bar{A} = \text{blockdiag}(\bar{A}_1, \bar{A}_2, \cdots, \bar{A}_N)$, $\bar{B} = \text{blockdiag}(\bar{B}_1, \bar{B}_2, \cdots, \bar{B}_N)$, $Q = \text{blockdiag}(Q_1, Q_2, \cdots, Q_N)$ and $R = \text{blockdiag}(R_1, R_2, \cdots, R_N)$. Based upon optimal control theory, the locally optimal control policy is (4) with

$$\begin{aligned} u_i^* &= \tilde{u}_i^* + U_i v^* \\ &= -K_{xi}^* \tilde{x}_i - K_{zi}^* \tilde{z}_i + U_i v^* \\ &= -K_{xi}^* x_i - K_{zi}^* z_i, \quad i = 1, 2, \cdots, N. \end{aligned} \qquad (13)$$

The optimal control gains are

$$\begin{bmatrix} K_{xi}^* & K_{zi}^* \end{bmatrix} = R_i^{-1} B_i^T P_i^* := K_i^* \qquad (14)$$

where $P_i^*$ is the unique solution to the following Riccati equation

$$\bar{A}_i^T P_i^* + P_i^* \bar{A}_i + Q_i - P_i^* \bar{B}_i R_i^{-1} \bar{B}_i^T P_i^* = 0. \qquad (15)$$

A model-based algorithm, Algorithm 1, is given to seek the decentralized optimal controller. Note that, instead of solving (15) which is nonlinear in $P_i^*$, we employ the policy iteration technique [24] to approximate $P_i^*$ by solving linear Lyapunov equations iteratively .

---

**Algorithm 1** Model-based Decentralized Optimal Controller Design

---

1: Find a pair $(G_1, G_2)$ such that it incorporates an internal model of $E$. $i \leftarrow 1$
2: **repeat**
3:     Find $K_i^{(0)}$ such that $\left(\bar{A}_i - \bar{B}_i K_i^{(0)}\right)$ is a Hurwitz matrix. $k \leftarrow 0$. Select a sufficiently small constant $\epsilon > 0$.
4:     **repeat**
5:         Solve $P_i^{(k)}$ and $K_i^{(k+1)}$ from

$$0 = \left(\bar{A}_i - \bar{B}_i K_i^{(k)}\right)^T P_i^{(k)} + P_i^{(k)} \left(\bar{A}_i - \bar{B}_i K_i^{(k)}\right)$$
$$+ Q_i + \left(K_i^{(k)}\right)^T R_i K_i^{(k)} \qquad (16)$$
$$K_i^{(k+1)} = R_i^{-1} B_i^T P_i^{(k)} \qquad (17)$$

6:         $k \leftarrow k + 1$
7:     **until** $\left| P_i^{(k)} - P_i^{(k-1)} \right| < \epsilon$
8:     $i \leftarrow i + 1$
9: **until** $i = N + 1$

---

## III. Main Results

In this section, we will design a data-driven distributed controller via ADP to achieve LFS under switching network topology. The developed approach is able to approximate the control gains $K_i^*$ for each follower without relying on the knowledge of system matrices $A_i$, $B_i$ and $D_i$. To begin with, the internal model (4) is modified by

$$\dot{z}_i = G_1 z_i + G_2 \hat{e}_i, \quad i = 1, 2, \cdots N \qquad (18)$$

where $\hat{e}_i = y_i + F\zeta_i$. The dynamics of $\zeta_i \in \mathbb{R}^q$ depends on the following equation

$$\dot{\zeta}_i = E\zeta_i + \sum_{j \in \mathcal{N}_i(t)} a_{ij}(t)(\zeta_j - \zeta_i) \quad i = 1, 2, \cdots, N \quad (19)$$

with $\zeta_0 = v$.

Then, we rewrite the $i$th subsystem augmented with the internal model (18):

$$\dot{\xi}_i = \bar{A}_i \xi_i + \bar{B}_i u_i + \bar{D}_i \psi_i$$
$$= \bar{A}_i^{(k)} \xi_i + \bar{B}_i \left( K_i^{(k)} \xi_i + u_i \right) + \bar{D}_i \psi_i$$

where, for $i = 1, 2, \cdots, N$, $\bar{A}_i^{(k)} = \bar{A}_i - \bar{B}_i K_i^{(k)}$, $\bar{D}_i = \text{blockdiag}(D_i, G_2 F_i)$, $\xi_i = \begin{bmatrix} x_i^T & z_i^T \end{bmatrix}^T \in \mathbb{R}^{m_i}$, $\psi_i = \begin{bmatrix} v^T & \zeta_i^T \end{bmatrix}^T \in \mathbb{R}^{2q}$.

By equation (16), we have

$$|\xi_i(t + \delta t)|_{P_i^{(k)}} - |\xi_i(t)|_{P_i^{(k)}}$$
$$= \int_t^{t+\delta t} [|\xi_i|_{(\bar{A}_i^{(k)})^T P_i^{(k)} + P_i^{(k)} \bar{A}_i^{(k)}} + 2\psi_i^T \bar{D}_i^T P_i^{(k)} \xi_i$$
$$+ 2(u_i + K_i^{(k)} \xi_i)^T \bar{B}_i^T P_i^{(k)} \xi_i] d\tau$$
$$= \int_t^{t+\delta t} [-|\xi_i|_{Q_i + (K_i^{(k)})^T R_i K_i^{(k)}} + 2\psi_i^T \bar{D}_i^T P_i^{(k)} \xi_i$$
$$+ 2(u_i + K_i^{(k)} \xi_i)^T R_i K_i^{(k+1)} \xi_i] d\tau. \quad (20)$$

By Kronecker product representation, we obtain

$$|\xi_i|_{Q_i + (K_i^{(k)})^T R_i K_i^{(k)}} = (\xi_i^T \otimes \xi_i^T) \text{vec} \left( Q_i + (K_i^{(k)})^T R_i K_i^{(k)} \right),$$
$$\psi_i^T \bar{D}_i^T P_i^{(k)} \xi_i = (\xi_i^T \otimes \psi_i^T) \text{vec} \left( \bar{D}_i^T P_i^{(k)} \right),$$
$$(u_i + K_i^{(k)} \xi_i)^T R_i K_i^{(k+1)} \xi_i = [(\xi_i^T \otimes \xi_i^T)(I \otimes (K_i^{(k)})^T R_i)$$
$$+ (\xi_i^T \otimes u_i^T)(I \otimes R_i)] \text{vec} \left( K_i^{(k+1)} \right).$$

Moreover, for any two vectors $a, b$ and a sufficiently large number $s > 0$, define

$$\delta_a = [\text{vecv}(a(t_1)) - \text{vecv}(a(t_0)), \cdots,$$
$$\text{vecv}(a(t_s)) - \text{vecv}(a(t_{s-1}))]^T,$$
$$\Gamma_{a,b} = \left[ \int_{t_0}^{t_1} a \otimes b \, d\tau, \int_{t_1}^{t_2} a \otimes b \, d\tau, \cdots, \int_{t_{s-1}}^{t_s} a \otimes b \, d\tau \right]^T.$$

(20) implies the following linear equation

$$\Psi_i^{(k)} \begin{bmatrix} \text{vecs}(P_i^{(k)}) \\ \text{vec}(K_i^{(k+1)}) \\ \text{vec}\left( \bar{D}_i^T P_i^{(k)} \right) \end{bmatrix} = \Phi_i^{(k)} \quad (21)$$

where

$$\Psi_i^{(k)} = [\delta_{\xi_i}, -2\Gamma_{\xi_i \xi_i}(I \otimes (K_i^{(k)})^T R_i) - 2\Gamma_{\xi_i u_i}(I \otimes R_i),$$
$$- 2\Gamma_{\xi_i \psi_i}],$$
$$\Phi_i^{(k)} = -\Gamma_{\xi_i \xi_i} \text{vec} \left( Q_i + (K_i^{(k)})^T R_i K_i^{(k)} \right).$$

The uniqueness of solution to (21) is guaranteed under some rank condition as shown below. For want of space, we omit the proof of Lemma 1 which follows the same line of proofs as in [10], [19].

**Lemma 1.** *For all $k \in \mathbb{Z}_+$, if there exists a $s^* \in \mathbb{Z}_+$ such that for all $s > s^*$,*

$$\text{rank}([\Gamma_{\xi_i \xi_i}, \Gamma_{\xi_i u_i}, \Gamma_{\xi_i \psi_i}]) = \frac{(m_i + 4q + 3)m_i}{2}, \quad (22)$$

*then the matrix $\Psi_i^{(k)}$ has full column rank for all $k \in \mathbb{Z}_+$.*

Now, we are ready to present a data-driven ADP algorithm 2 which yields approximate solutions to the unknown optimal values $K_i^*$ and $P_i^*$.

---

**Algorithm 2** Data-driven ADP Algorithm for Distributed Optimal Controller Design

---

1: Find a pair $(G_1, G_2)$ such that it incorporates an internal model of $E$.
2: Select a small $\epsilon > 0$. Apply $u_i = -K_i^{(0)} \xi_i + \nu_i$ on $[t_0, t_s]$ with $\nu_i$ an exploration noise, s.t. (22) holds for $i = 1, 2, \cdots, N$.
3: $i \leftarrow 1$
4: **repeat**
5: $\quad k \leftarrow -1$
6: $\quad$ **repeat**
7: $\quad\quad k \leftarrow k + 1$
8: $\quad\quad$ Solve $P_i^{(k)}$ and $K_i^{(k+1)}$ from (21)
9: $\quad$ **until** $|P_i^{(k)} - P_i^{(k-1)}| < \epsilon$ for $k \geq 1$
10: $\quad P_i^\dagger \leftarrow P_i^{(k)}$
11: $\quad$ The learned controller is (18), (19), and

$$u_i = -K_i^{(k+1)} \xi_i := -K_i^\dagger \xi_i \quad (23)$$

12: $\quad i \leftarrow i + 1$
13: **until** $i = N + 1$

---

The convergence of Algorithm 2 is shown in Theorem 1, while the LFS of the closed-loop system is analyzed in Theorem 2.

**Theorem 1.** *If (22) is satisfied, then, for $i = 1, 2, \cdots, N$, sequences $\{P_i^{(k)}\}_{k=0}^\infty$ and $\{K_i^{(k)}\}_{k=1}^\infty$ computed by Algorithm 2 converge to $P_i^*$ and $K_i^*$, respectively.*

*Proof.* For all $1 \leq i \leq N$, letting $P_i^{(k)} = \left( P_i^{(k)} \right)^T > 0$ be the solution to (16). $K_i^{(k+1)}$ is uniquely determined by (17) with $T_i^{(k)} = \bar{D}_i^T P_i^{(k)}$. On the other hand, letting $\hat{P}$, $\hat{K}$, and $\hat{T}$ solve (21), condition (22) ensures that $P_i^{(k)} = \hat{P}$, $K_i^{(k+1)} = \hat{K}$, and $T_i^{(k)} = \hat{T}$ are uniquely determined. By [24], we have $\lim_{k \to \infty} K_i^{(k)} = K_i^*$, $\lim_{k \to \infty} P_i^{(k)} = P_i^*$. The convergence of sequences $\{\bar{P}_i^{(k)}\}_{k=0}^\infty$ and $\{\bar{K}_i^{(k)}\}_{k=1}^\infty$ obtained by non-model-based Algorithm 2 is thus ensured. $\square$

**Theorem 2.** *Under Assumptions 1-4, the multi-agent system (1)-(2) in closed-loop with the learned controller (18), (19) and (23) is leader-to-formation stable.*

*Proof.* Write the closed-loop system in a compact form

$$\begin{bmatrix} \dot{\xi} \\ \dot{\zeta} \end{bmatrix} = \begin{bmatrix} A_c^\dagger & B_c^\dagger \\ 0 & [(I_N \otimes E) - (\mathcal{H}_{\rho(t)} \otimes I_q)] \end{bmatrix} \begin{bmatrix} \xi \\ \zeta \end{bmatrix}$$
$$+ \begin{bmatrix} D \\ \mathcal{H}_{\rho(t)} \otimes I_q \end{bmatrix} (\mathbb{1}_N \otimes v)$$
$$:= \bar{A}_{c,\rho(t)} \begin{bmatrix} \xi \\ \zeta \end{bmatrix} + \bar{B}_{c,\rho(t)}(\mathbb{1}_N \otimes v),$$
$$e = \bar{C}\xi + F(\mathbb{1}_N \otimes v) \tag{24}$$

where $A_c^\dagger = \text{blockdiag}(\bar{A}_1 - \bar{B}_1 K_1^\dagger, \cdots, \bar{A}_N - \bar{B}_N K_N^\dagger)$, $\zeta = [\zeta_1^T, \zeta_2^T, \cdots, \zeta_N^T]^T$, $D = \text{blockdiag}([D_1^T \quad 0_{q \times (m_i - n_i)}]^T, \cdots, [D_N^T \quad 0_{q \times (m_i - n_i)}]^T)$, and $B_c^\dagger = \text{blockdiag}([0_{q \times n_i} \quad F_1^T G_2^T]^T, \cdots, [0_{q \times n_i} \quad F_N^T G_2^T]^T)$. $\mathcal{H}_{\rho(t)} = [h_{ij}] \in \mathbb{R}^{N \times N}$ is a submatrix of Laplacian of the digraph $\mathcal{G}_{\rho(t)}$ with $h_{ii}(t) = \sum_{j=0}^N a_{ij}(t)$ and $h_{ij}(t) = -a_{ij}(t)$ if $j \neq i$.

Let $X = \text{blockdiag}(X_1, Z_1, X_2, Z_2, \cdots, X_N, Z_N)$, where, for $i = 1, 2, \cdots, N$, matrices $X_i$ and $Z_i$ solve regulator equations (6)-(7) and equations (8)-(9). Defining $\bar{X}_c = [X^T \quad I_{qN}]^T$, then we have

$$\bar{X}_c(I_N \otimes E) = \bar{A}_{c,\rho(t)}\bar{X}_c + \bar{B}_{c,\rho(t)},$$
$$0 = \bar{C}X + F. \tag{25}$$

Let $\tilde{\zeta} = \zeta - (\mathbb{1}_N \otimes v^*)$. By equations (24) and (25), we obtain

$$\dot{\tilde{v}} = [I_N \otimes (E - HK_0)]\tilde{v} + (I_N \otimes H)(\mathbb{1}_N \otimes \tilde{w}), \tag{26}$$
$$\begin{bmatrix} \dot{\tilde{\xi}} \\ \dot{\tilde{\zeta}} \end{bmatrix} = \bar{A}_{c,\rho(t)} \begin{bmatrix} \tilde{\xi} \\ \tilde{\zeta} \end{bmatrix} + \bar{B}_{c,\rho(t)}\tilde{v} \tag{27}$$
$$e = F\tilde{v} + \bar{C}\tilde{\xi}. \tag{28}$$

Define $\phi = [\tilde{\xi}^T, \tilde{\zeta}^T]^T$. By [39, Lemma 2], we see that the origin of following linear switched system

$$\dot{\phi} = \bar{A}_{c,\rho(t)}\phi \tag{29}$$

is exponentially stable. Hence, the state transition matrix $\Phi(\tau, t)$ of (29) satisfies $|\Phi(\tau, t)| \leq ke^{-\lambda(\tau - t)}, \forall \tau \geq t \geq 0$. Let $P^\dagger(t) = \int_t^\infty \Phi^T(\tau, t)\Phi(\tau, t)d\tau$. It is checkable that there exists some $c_1, c_2 > 0$ such that $c_1|\phi|^2 \leq \phi^T P^\dagger(t)\phi \leq c_2|\phi|^2$, which implies that $P^\dagger(t)$ is positive definite and upper bounded by some $\sup_{t \geq 0}|P^\dagger(t)| < c_3$. The definition of $P^\dagger(t)$ shows that it is symmetric and continuously differentiable. By the fact that $\partial\Phi(\tau, t)/\partial t = -\Phi(\tau, t)\bar{A}_{c,\rho(t)}$, we have

$$\dot{P}^\dagger(t) = -P^\dagger(t)\bar{A}_{c,\rho(t)} - \bar{A}_{c,\rho(t)}^T P^\dagger(t) - I. \tag{30}$$

On the other hand, since $\sigma(E - HK_0) \subset \mathbb{C}^-$, there exists a positive definite matrix $P_v = P_v^T$ such that

$$[I_N \otimes (E - HK_0)]^T P_v + P_v[I_N \otimes (E - HK_0)] + I_{qN} = 0 \tag{31}$$

Choose the Lyapunov function $V(t) = \phi^T P^\dagger(t)\phi + (2c_3^2 c_4^2 + 1)\tilde{v}^T P_v \tilde{v}$, where $c_4 = \sup_{t \geq 0}|\bar{B}_{c,\rho(t)}|$ and $c_5 = |P_v(I_N \otimes H)|$.

The derivative of $V$ along the solutions of system (26)-(27) is

$$\dot{V} = -|\phi|^2 + 2\phi^T P^\dagger \bar{B}_{c,\rho(t)}\tilde{v} - (2c_3^2 c_4^2 + 1)|\tilde{v}|^2$$
$$+ 2(2c_3^2 c_4^2 + 1)\tilde{v}^T P_v(I_N \otimes H)(\mathbb{1}_N \otimes \tilde{w})$$
$$\leq -|\phi|^2 + 2c_3 c_4|\phi||\tilde{v}| - (2c_3^2 c_4^2 + 1)|\tilde{v}|^2$$
$$+ 2N(2c_3^2 c_4^2 + 1)c_5|\tilde{v}||\tilde{w}|$$
$$\leq -\frac{|\phi|^2}{2} - \frac{|\tilde{v}|^2}{2} + c_6|\tilde{w}|^2$$

with $c_6 = 2N^2(2c_3^2 c_4^2 + 1)^2 c_5^2$.

The previous inequality implies that the system (26)-(27) with $\tilde{w}$ as the input is input-to-state stable (ISS) [36]. Given $\eta = [\tilde{v}^T, \phi^T]^T$, there exist a function $\beta_1$ of class $\mathcal{KL}$ and a function $\gamma_1$ of class $\mathcal{K}$ such that

$$|\eta(t)| \leq \beta_1(|\eta(0)|, t) + \gamma_1(\|\tilde{w}\|). \tag{32}$$

From (28), one can immediately check that the LFS condition (5) is satisfied. The proof is thus completed. $\square$

The following result compares the cost $J^\odot$ for the decentralized controller (4), (13) with the cost $J^\dagger$ associated with the distributed controller (18), (19) and (23).

**Theorem 3.** *There always exist constants $d_1, d_2 > 0$ such that $J^\dagger \leq d_1 J^\odot + d_2|\tilde{\zeta}(0)|^2$ if $v \equiv v^*$.*

*Proof.* Denoting $K^* = \text{blockdiag}\{K_1^*, K_2^*, \cdots, K_N^*\}$, $P^* = \text{blockdiag}\{P_1^*, P_2^*, \cdots, P_N^*\}$, one can rewrite the system (2) with (4) and (13) by

$$\dot{\tilde{\xi}} = (\bar{A} - \bar{B}K^*)\tilde{\xi}. \tag{33}$$

The corresponding cost is

$$J^\odot = \tilde{\xi}(0)^T P^* \tilde{\xi}(0). \tag{34}$$

When $v \equiv v^*$, the system (2) with (13), (18) and (19) can be written by (29). Along the trajectory of (29), from (30), we have

$$\int_0^\infty |\phi^2|d\tau \leq \phi^T(0)P^\dagger(0)\phi(0) \leq c_2|\phi(0)|^2. \tag{35}$$

By the previous inequality, the cost $J^\dagger$ is upper bounded by

$$J^\dagger \leq \lambda_M(Q + (K^\dagger)^T R K^\dagger)\int_0^\infty |\tilde{\xi}|^2 d\tau$$
$$\leq \lambda_M(Q + (K^\dagger)^T R K^\dagger)\int_0^\infty |\phi|^2 d\tau$$
$$\leq c_2\lambda_M(Q + (K^\dagger)^T R K^\dagger)|\phi(0)|^2$$
$$\leq c_2\lambda_M(Q + (K^\dagger)^T R K^\dagger)(|\tilde{\xi}(0)|^2 + |\tilde{\zeta}(0)|^2)$$
$$\leq \frac{c_2\lambda_M(Q + (K^\dagger)^T R K^\dagger)}{\lambda_m(P^*)}J^\odot$$
$$+ c_2\lambda_M(Q + (K^\dagger)^T R K^\dagger)|\tilde{\zeta}(0)|^2$$
$$:= d_1 J^\odot + d_2|\tilde{\zeta}(0)|^2 \tag{36}$$

where $K^\dagger = \text{blockdiag}\{K_1^\dagger, K_2^\dagger, \cdots, K_N^\dagger\}$.

The proof is thus completed. $\square$

**Remark 4.** *Note that the proposed Algorithm 2 is a direct adaptive control approach without identifying the system matrices. In each iteration, one can estimate the controller*

*parameters by solving a linear matrix equation (21). This is different from the indirect adaptive optimal control [25], [16, Chap. 7.4.4] that the plant parameters are estimated online and used to find controller parameters by solving the corresponding Riccati equations that is nonlinear in $P_i$.*

**Remark 5.** *(22) is introduced to ensure the convergence of controller parameters. Like traditional adaptive control [46] and existing work on ADP [19], [28], we add an exploration noise to the input during the learning phase in order to satisfy (22).*

**Remark 6.** *Albeit some nodes cannot get instant information from the leader, Algorithm 2 is implementable since all the followers are reachable from node 0, i.e., there always exists a $T > 0$ such that $v(t)$ in the period $[t_0, t_s]$ is receivable by all the other subsystems at $t = t_s + T$.*

## IV. EXAMPLE

In order to validate the effectiveness of the proposed data-driven Algorithm 2, we consider a system in the form of (1)-(2) with $N = 4$ and for $i = 1, 2, 3, 4$,

$$A_i = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, B_i = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, D_i = \begin{bmatrix} 0 & 0 \\ 0 & 0.5 * i \end{bmatrix},$$

$$C_i = \begin{bmatrix} 1 \\ 0 \end{bmatrix}^T, F_i = \begin{bmatrix} -i - 1 \\ 0 \end{bmatrix}^T, E = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}, H = \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

Suppose that the switching network topology $\mathcal{G}_{\rho(t)}$ is dominated by the following switching signal

$$\rho(t) = \begin{cases} 1, & \text{if } 3sT_s \le t < (3s+1)T_s \\ 2, & \text{if } (3s+1)T_s \le t < (3s+2)T_s \\ 3, & \text{if } (3s+2)T_s \le t < (3s+3)T_s \end{cases}$$

where $s = 0, 1, 2, \cdots$, $T_s = 0.2s$ and the corresponding communication graph is depicted in Fig. 1. It is easily checkable that Assumptions 1-4 are satisfied. Suppose the system matrices $A_i, B_i$, and $D_i$ are unknown. Let the internal model for the exosystem dynamics $v$ be

$$G_1 = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}, G_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

For the purpose of simulation, we choose all the weight matrices $Q_i$ and $R_i$ in the cost (11) to be identity matrices. The external disturbance of the leader is set by $\tilde{w} = \sin(3t)$. The exploration noise is taken as a sum of sinusoidal signals with disparate frequencies. We collect data from $t = 0s$ to $t = 10s$, then (21) is solved repeatedly until the convergence criterion is satisfied. The comparison of the $P_i^{(k)}$ of the $i$th follower at $k$th iteration with its optimal value is shown in Fig. 2. We employ our updated control policy after $t = 10s$. The outputs of all the leader and followers are depicted in Fig. 3 with their reference signal $y_i^* = -F_i v^*$. The plots of the distributed control inputs are shown in Fig. 4.

## V. SUMMARY AND FUTURE WORK

This note has studied the cooperative adaptive optimal control problem by means of a combined use of internal model principle and adaptive dynamic programming theory.
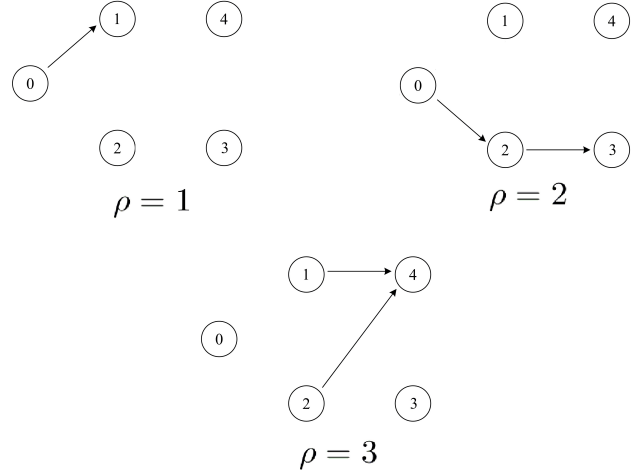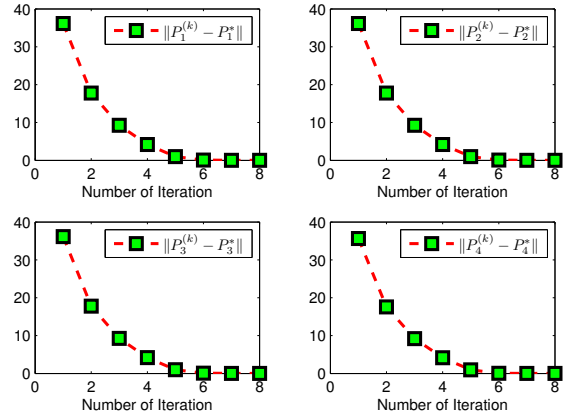


Fig. 1. Network topology



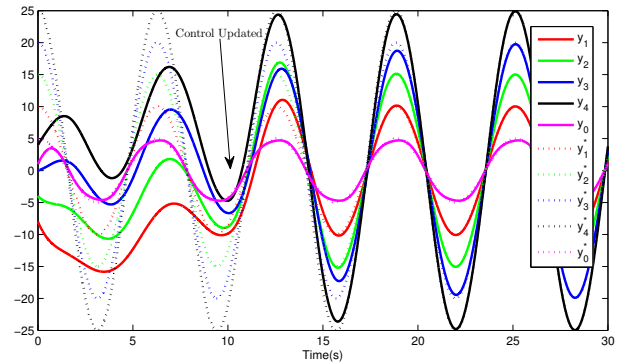Fig. 2. The comparison of $P_i^{(k)}$ and their optimal values



Fig. 3. Plots of the system outputs $y_i$ and the reference signal $y_i^*$

The communication network is jointly connected, which is allowed disconnected at any time instant. Instead of relying on the accurate knowledge of the system dynamics, an internal-model-based control policy is learned by means of input-state data. The learned control policy achieves leader-to-formation stability, which is robust to unmeasurable leader disturbance. Future work includes the generalization to nonlinear multi-
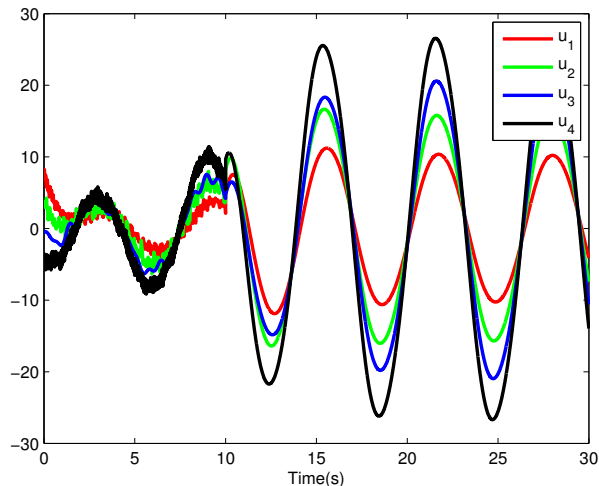
Fig. 4. Plots of distributed controllers

agent systems and the combination of ADP and adaptive internal model.

## REFERENCES

[1] 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, 2014.

[2] D. P. Bertsekas and J. N. Tsitsiklis, *Neuro-Dynamic Programming*. Belmont, MA: Athena Scientific, 1996.

[3] L. Consolini, F. Morbidi, D. Prattichizzo, and M. Tosques, "Leader-follower formation control of nonholonomic mobile robots with input constraints," *Automatica*, vol. 44, no. 5, pp. 1343 – 1349, 2008.

[4] D. V. Dimarogonas, P. Tsiotras, and K. J. Kyriakopoulos, "Leader-follower cooperative attitude control of multiple rigid bodies," *Systems & Control Letters*, vol. 58, no. 6, pp. 429 – 435, 2009.

[5] Z. Ding, "Consensus output regulation of a class of heterogeneous nonlinear systems," *IEEE Transactions on Automatic Control*, vol. 58, no. 10, pp. 2648–2653, 2013.

[6] Q. Y. Fan and G. H. Yang, "Adaptive actor-critic design-based integral sliding-mode control for partially unknown nonlinear systems with input disturbances," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 27, no. 1, pp. 165–177, 2016.

[7] B. Francis, "The linear multivariable regulator problem," *SIAM Journal on Control and Optimization*, vol. 15, no. 3, pp. 486–505, 1977.

[8] W. Gao, M. Huang, Z. P. Jiang, and T. Chai, "Sampled-data-based adaptive optimal output-feedback control of a 2-degree-of-freedom helicopter," *IET Control Theory and Applications*, vol. 10, no. 12, pp. 1440–1447, 2016.

[9] W. Gao, Y. Jiang, Z. P. Jiang, and T. Chai, "Output-feedback adaptive optimal control of interconnected systems based on robust adaptive dynamic programming," *Automatica*, vol. 72, pp. 37–45, 2016.

[10] W. Gao and Z. P. Jiang, "Adaptive dynamic programming and adaptive optimal output regulation of linear systems," *IEEE Transactions on Automatic Control*, vol. 61, no. 12, pp. 4164–4169, 2016.

[11] ——, "Learning-based adaptive optimal tracking control of strict-feedback nonlinear systems," *IEEE Transactions on Neural Networks and Learning Systems,* in press, 2017.

[12] ——, "Nonlinear and adaptive suboptimal control of connected vehicles: A global adaptive dynamic programming approach," *Journal of Intelligent & Robotic Systems*, vol. 85, no. 3, pp. 597–611, 2017.

[13] J. Hu and W. X. Zheng, "Adaptive tracking control of leaderfollower systems with unknown dynamics and partial measurements," *Automatica*, vol. 50, no. 5, pp. 1416 – 1423, 2014.

[14] C. Huang and X. Ye, "Cooperative output regulation of heterogeneous multi-agent systems: An $H_\infty$ criterion," *IEEE Transactions on Automatic Control*, vol. 59, no. 1, pp. 267–273, 2014.

[15] J. Huang, *Nonlinear Output Regulation: Theory and Applications*. Philadelphia, PA: SIAM, 2004.

[16] P. A. Ioannou and J. Sun, *Robust Adaptive Control*. Mineola, NY: Dover Publications, Inc., 2012.

[17] A. Jadbabaie, J. Lin, and A. S. Morse, "Coordination of groups of mobile autonomous agents using nearest neighbor rules," *IEEE Transactions on Automatic Control*, vol. 48, no. 6, pp. 988–1001, 2003.

[18] Y. Jiang, J. Fan, T. Chai, J. Li, and F. L. Lewis, "Data-driven flotation industrial process operational optimal control based on reinforcement learning," *IEEE Transactions on Industrial Informatics,* in press, 2017.

[19] 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.

[20] ——, "Robust adaptive dynamic programming for large-scale systems with an application to multimachine power systems," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 59, no. 10, pp. 693–697, 2012.

[21] ——, *Robust Adaptive Dynamic Programming*. Hoboken. NJ: Wiley-IEEE Press, 2017.

[22] Z. P. Jiang and Y. Jiang, "Robust adaptive dynamic programming for linear and nonlinear systems: An overview," *European Journal of Control*, vol. 19, no. 5, pp. 417–425, 2013.

[23] Z. P. Jiang, A. R. Teel, and L. Praly, "Small-gain theorem for ISS systems and applications," *Mathematics of Control, Signals and Systems*, vol. 7, no. 2, pp. 95–120, 1994.

[24] D. Kleinman, "On an iterative technique for Riccati equation computations," *IEEE Transactions on Automatic Control*, vol. 13, no. 1, pp. 114–115, 1968.

[25] P. R. Kumar, "Optimal adaptive contorl of linear-quadratic-guassian systems," *SIAM Journal on Control and Optimization*, vol. 21, no. 2, pp. 163–178, 1983.

[26] F. L. Lewis and K. G. Vamvoudakis, "Reinforcement learning for partially observable dynamic processes: Adaptive dynamic programming using measured output data," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 41, no. 1, pp. 14–25, 2011.

[27] F. L. Lewis and D. Vrabie, "Reinforcement learning and adaptive dynamic programming for feedback control," *IEEE Circuits and Systems Magazine*, vol. 9, no. 3, pp. 32–50, 2009.

[28] F. L. Lewis, D. Vrabie, and K. G. Vamvoudakis, "Reinforcement learning and feedback control: Using natural decision methods to design optimal adaptive controllers," *IEEE Control Systems Magazine*, vol. 32, no. 6, pp. 76–105, 2012.

[29] Z. Li, X. Liu, W. Ren, and L. Xie, "Distributed tracking control for linear multiagent systems with a leader of bounded unknown input," *IEEE Transactions on Automatic Control*, vol. 58, no. 2, pp. 518–523, 2013.

[30] L. Liu, "Adaptive cooperative output regulation for a class of nonlinear multi-agent systems," *IEEE Transactions on Automatic Control*, vol. 60, no. 6, pp. 1677–1682, 2015.

[31] W. Liu and J. Huang, "Cooperative adaptive output regulation for second-order nonlinear multiagent systems with jointly connected switching networks," *IEEE Transactions on Neural Networks and Learning Systems, in press*, 2017.

[32] R. Marino and P. Tomei, "Output regulation for linear systems via adaptive internal model," *IEEE Transactions on Automatic Control*, vol. 48, no. 12, pp. 2199–2202, 2003.

[33] H. Modares and F. L. Lewis, "Linear quadratic tracking control of partially-unknown continuous-time systems using reinforcement learning," *IEEE Transactions on Automatic Control*, vol. 59, no. 11, pp. 3051–3056, 2014.

[34] C. Mu, Z. Ni, C. Sun, and H. He, "Data-driven tracking control with adaptive dynamic programming for a class of continuous-time nonlinear systems," *IEEE Transactions on Cybernetics*, vol. 47, no. 6, pp. 1460–1470, 2017.

[35] W. Ren and R. Beard, *Distributed Consensus in Multi-vehicle Cooperative Control*, ser. Communications and Control Engineering. London, U.K.: Springer-Verlag London, 2008.

[36] E. D. Sontag, "Smooth stabilization implies coprime factorization," *IEEE Transactions on Automatic Control*, vol. 34, no. 4, pp. 435–443, 1989.

[37] ——, "Input to state stability: Basic concepts and results," in *Nonlinear and Optimal Control Theory*, P. Nistri and G. Stefani, Eds. Berlin: Springer-Verlag, 2007, pp. 163–220.

[38] Y. Su and J. Huang, "Cooperative output regulation of linear multi-agent systems," *IEEE Transactions on Automatic Control*, vol. 57, no. 4, pp. 1062–1066, 2012.

[39] ——, "Cooperative output regulation with application to multi-agent consensus under switching network," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 42, no. 3, pp. 864–875, 2012.

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

[41] H. G. Tanner, G. J. Pappas, and V. Kumar, "Leader-to-formation stability," *IEEE Transactions on Robotics and Automation*, vol. 20, no. 3, pp. 443–455, 2004.

[42] K. G. Vamvoudakis, D. Vrabie, and F. L. Lewis, "Online adaptive algorithm for optimal control with integral reinforcement learning," *International Journal of Robust and Nonlinear Control*, vol. 24, no. 17, pp. 2686–2710, 2014.

[43] D. Wang, D. Liu, H. Li, B. Luo, and H. Ma, "An approximate optimal control approach for robust stabilization of a class of discrete-time nonlinear systems with uncertainties," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 46, no. 5, pp. 713–717, 2016.

[44] X. Wang, Y. Hong, J. Huang, and Z.-P. Jiang, "A distributed control approach to a robust output regulation problem for multi-agent linear systems," *IEEE Transactions on Automatic Control*, vol. 55, no. 12, pp. 2891–2895, 2010.

[45] P. J. Werbos, "Beyond regression: New tools for prediction and analysis in the behavioral sciences," Ph.D. dissertation, Harvard University, 1974.

[46] J. Yuan and W. Wonham, "Probing signals for model reference identification," *IEEE Transactions on Automatic Control*, vol. 22, no. 4, pp. 530–538, 1977.