Optimal Containment Control of a Quadrotor Team With Active Leaders via Reinforcement Learning

Ming Cheng, Hao Liu, *Senior Member, IEEE*, Qing Gao, *Senior Member, IEEE*, Jinhu Lü, *Fellow, IEEE*, and Xiaohua Xia, *Fellow, IEEE*

***Abstract*—This article proposes an optimal controller for a team of underactuated quadrotors with multiple active leaders in containment control tasks. The quadrotor dynamics are underactuated, nonlinear, uncertain, and subject to external disturbances. The active team leaders have control inputs to enhance the maneuverability of the containment system. The proposed controller consists of a position control law to guarantee the achievement of position containment and an attitude control law to regulate the rotational motion, which are learned via off-policy reinforcement learning using historical data from quadrotor trajectories. The closed-loop system stability can be guaranteed by theoretical analysis. Simulation results of cooperative transportation missions with multiple active leaders demonstrate the effectiveness of the proposed controller.**

***Index Terms*—Cooperative control, multiagent system, optimal control, quadrotor, reinforcement learning (RL).**

I. Introduction

OVER the past decade, cooperative control of quadrotors has attracted an increasing interest from the control community for its wide range of applications, such as agricultural, emergency rescue, express delivery logistics, and remote sensing [1], [2], [3], [4], [5]. Containment control, as a challenging topic of cooperative control of quadrotors, aims to drive each vehicle into the convex hull spanned by multiple team leaders and has practical uses. For instance, in a cooperative logistics scenario involving multiple unmanned aerial vehicles (UAVs), the quadrotor vehicles can carry more payloads instead of advanced navigation systems and can be guided by staying within the safe region formed by the team leaders. Therefore, the containment control problem has received much research attention from the control and robotic communities.

Recently, the containment control problems have been studied in multiple works. In [6], a decentralized framework for multirobot systems to form clusters around multiple targets and achieve the containment for the followers was designed using a game-theoretic rule. In [7], a rigidity-based approach was proposed to achieve formation among multiple agents modeled as double integrators. In [8], a distributed fault-tolerant containment control protocol was developed for the discrete-time multiagent systems (MASs). Note that the nonlinear and coupled features of the vehicle dynamics were ignored in [6], [7], and [8]. In [9], the containment control problem of discrete-time single-input linear MASs was investigated using the standard Riccati design method. In [10], a time-varying group formation-containment tracking controller was designed for general linear MASs and the control protocol design was based on the solution to an algebraic Riccati inequality. In [11], a consensus scheme based on distributed linear quadratic regulation was developed and tested for heterogeneous MASs with linearized quadrotors and two-wheeled mobile robots. However, in [9], [10], and [11], the controllers were based on complete and accurate knowledge of the dynamical models, which is difficult to obtain for the quadrotors due to their complex mass distribution and working environment.

In recent years, robust optimal control methods have been developed for uncertain nonlinear networked systems. The satellite containment control problem was discussed in [12] using the adaptive sliding mode control and potential functions. In [13], a robust hierarchical pinning control scheme for nonlinear heterogeneous MASs was developed to deal with the uncertainties and disturbances. In [14], a finite-time attitude containment control problem of spacecraft formation was studied using the backstepping design technique. In [15], an observer-based containment control approach was proposed for networked nonlinear MASs using the active disturbance rejection control. However, these robust optimal controllers in [12], [13], [14], and [15] can only guarantee the optimality for the nominal systems, but not for the uncertain nonlinear systems.

The reinforcement learning (RL) has been be introduced to solve optimal control problems of the uncertain nonlinear systems. Zamfirache et al. [16] designed a policy iteration RL algorithm that updated neural networks using a gray

Manuscript received 5 March 2023; revised 19 May 2023; accepted 7 June 2023. This work was supported in part by the Beijing Natural Science Foundation under Grant 4232045; in part by the National Natural Science Foundation of China under Grant 62273015 and Grant 61873012; in part by Beijing Nova Program; and in part by the National Science Foundation under Grant 1730675 and Grant 1714519. This article was recommended by Associate Editor R. Selmic. *(Corresponding author: Hao Liu.)*

Ming Cheng is with the School of Astronautics, Beihang University, Beijing 100191, China (e-mail: mingcheng@buaa.edu.cn).

Hao Liu is with the Institute of Artificial Intelligence, Beihang University, Beijing 100191, China, and also with Zhongguancun Laboratory, Beijing 100191, China (e-mail: liuhao13@buaa.edu.cn).

Qing Gao and Jinhu Lü are with the School of Automation Science and Electrical Engineering, Beihang University, Beijing 100191, China, and also with Zhongguancun Laboratory, Beijing 100191, China (e-mail: gaoqing@buaa.edu.cn; jhlu@iss.ac.cn).

Xiaohua Xia is with the Department of Electrical, Electronic and Computer Engineering, University of Pretoria, Pretoria 0002, South Africa (e-mail: xxia@up.ac.za).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TCYB.2023.3284648.

Digital Object Identifier 10.1109/TCYB.2023.3284648

wolf optimizer algorithm. In [17], a fuzzy optimal control method was proposed for nonlinear systems utilizing a modified evolved bat algorithm. In [18], a model-independent control protocol was developed to achieve the containment control for networked Euler–Lagrange systems with uncertainties. In [19], adaptive distributed observer techniques were employed to handle a bipartite containment control problem of linear MASs. In [20], a data-driven fault-tolerant attitude synchronization control problem was studied with the nonlinear rotational dynamics of the quadrotor vehicles via RL method. Note that the existing results in [18], [19], and [20] mainly focused on the cooperative control problems with autonomous leaders and ignored the control input. However, the active leaders with control input are important for enhancing the maneuverability of the multivehicle teams and can enable the containment systems to perform autonomous behaviors and achieve various complicated geometric configurations to avoid unexpected menace (see [21], [22]). Therefore, the optimal containment control problem with active leaders for the quadrotors suffering from unknown parameters, underactuation, nonlinear couplings, and external disturbance based on the RL remains challenging and open, which motivates the current paper.

In this article, the optimal containment control problem involving multiple active leaders is addressed using RL. A cascade containment control law is designed that consists of an optimal position control law for achieving containment and an optimal attitude control law for controlling the rotational motion. The optimal control laws are learned from historical data measured along the quadrotor trajectories. Moreover, the proposed controller can ensure both containment and optimal performance for the quadrotor team. The proposed containment controller has three main advantages over the existing methods.

First, the proposed containment controller in the current article can achieve optimal tracking performance for the uncertain quadrotor systems. But, the existing methods [12], [13], [14], [15] can only guarantee optimal performance for the nominal systems instead of uncertain nonlinear quadrotor vehicles.

Second, the proposed controller can iteratively interact with the nonlinear and coupled vehicle dynamics and learn the optimal control laws from the generated input–output data. However, existing results in [9], [10], and [11] based on the classical optimal control theory were dependent on complete and accurate knowledge of the dynamical models.

Third, in this article, the control inputs are introduced into the dynamics of the team leaders in response to unexpected menace and the achievement of the containment can still be guaranteed. But, the control inputs of the leaders were ignored in [4], [8], [9], [18], [19], and [20], resulting in limited maneuverability of the containment systems.

The remaining parts of this article are organized as follows. Section II provides necessary preliminaries on the graph theory and the problem formulation involving quadrotor dynamics. Section III discusses the optimal containment control law design and theoretical analysis for the quadrotors. Section IV presents a simulation example and the results are given. Section V concludes the whole article.

II. Problem Formulation

A. Preliminaries

This article considers a team of UAVs consisting of n_f quadrotor followers denoted by $\mathcal{F} \triangleq \{1, \ldots, n_f\}$ and n_l leaders denoted by $\mathcal{L} \triangleq \{n_f+1, \ldots, n_f+n_l\}$. Let $\mathcal{G}_e = \{\mathcal{V}_e, \mathcal{E}_e, A_e\}$ and $\mathcal{G}_f = \{\mathcal{V}_f, \mathcal{E}_f, A_f\}$ be weighted directed graph describing the interaction topology for the entire team of $n_f + n_l$ UAVs and the followers, respectively. $\mathcal{V}_f = \{v_{fi}\}, (i = 1, 2, \ldots, n_f)$ represents the set of vertices, where v_{fi} indicates the ith quadrotor. $\mathcal{E}_f \subset \mathcal{V}_f \times \mathcal{V}_f$ represents the set of edges and $A_f = [a_{ij}] \in \mathbb{R}^{n_f \times n_f}$ represents the adjacency matrix, where $a_{ij} > 0$, if $(v_{fi}, v_{fj}) \in \mathcal{E}_f$ and $a_{ij} = 0$, otherwise. The leaders are active without incoming edges, and the followers are quadrotors with incoming edges. Define $N_{fi} = \{v_{fj} | (v_{fi}, v_{fj}) \in \mathcal{E}_f\}$ as the neighbors of the ith quadrotor. Define a series of successive edges, that is, $\{(v_{fi}, v_{fk}), (v_{fk}, v_{fl}), \ldots, (v_{fm}, v_{fj})\}$ as a directed path from the ith node to the jth node. Let $W_v = \mathrm{diag}\{\rho_1^v, \rho_2^v, \ldots, \rho_{n_f}^v\}, (v \in \mathcal{L})$ be the connection indicator of the vth leader, where ρ_i^v is 1, if there exists a link from the vth leader to the ith follower, and 0, otherwise.

Notations: $I_n \in \mathbb{R}^{n\times n}$ denotes a unit matrix, $1_n \in \mathbb{R}^n$ a column vector with 1 as its elements, $c_{a,b} \in \mathbb{R}^a$ a column vector with 1 on the bth element and 0 s elsewhere. Let $0_{m\times n} \in \mathbb{R}^{m\times n}$ represent a zero matrix and $\otimes$ represent the Kronecker product. Denote $\mathrm{dist}(x, \mathcal{D})$ as the distance from a vector $x \in \mathbb{R}^n$ to a set $\mathcal{D}$ in the Euclidean norm given by $\mathrm{dist}(x, \mathcal{D}) = \inf_{y\in\mathcal{D}} \|x - y\|_2$. Denote $Co(\mathcal{U})$ as the convex hull of a set of points $\mathcal{U} = \{u_1, u_2, \ldots, u_n\}$ with finite elements, representing the minimal convex set that contains every point in $\mathcal{U}$, that is, $Co(\mathcal{U}) = \{\sum_{i=1}^n \lambda_i u_i | u_i \in \mathcal{U}, \lambda_i \geq 0, \sum_{i=1}^n \lambda_i = 1\}$.

B. Quadrotor Dynamics

In this article, the fully nonlinear dynamical model is considered for each quadrotor. Denote $\hat{E}_I$ as the Earth-fixed inertial frame and $\hat{E}_{Bi}$ as the body-fixed frame attached to the ith quadrotor. Define $p_{fi} = [p_{fi,x} \;\; p_{fi,y} \;\; p_{fi,z}]^T \in \mathbb{R}^3$ as the position of the ith quadrotor in $\hat{E}_I$ and $\Theta_i = [\phi_i \;\; \theta_i \;\; \psi_i]^T \in \mathbb{R}^3$ as the Euler angle of the ith quadrotor. As shown in [23], the dynamical model of each quadrotor can be written as

$$\begin{aligned} m_i \ddot{p}_{fi} &= R_{IB} F_i + d_{pi} \\ J_i \ddot{\Theta}_i &= -C\big(\Theta_i, \dot{\Theta}_i\big)\dot{\Theta}_i + \tau_i + d_{\Theta i} \end{aligned} \tag{1}$$

where m_i and J_i are the mass and inertial matrix of the ith quadrotor with $J_i = \mathrm{diag}\{J_i^{\phi}, J_i^{\theta}, J_i^{\psi}\} \in \mathbb{R}^{3\times 3}$, $R_{IB} \in \mathbb{R}^{3\times 3}$ is the coordination transformation matrix from $\hat{E}_{Bi}$ to $\hat{E}_I$, $C(\Theta_i, \dot{\Theta}_i) \in \mathbb{R}^{3\times 3}$ is the nonlinear Coriolis term, and $F_i \in \mathbb{R}^3$, $\tau_i \in \mathbb{R}^3$ are external forces and torques from the rotors in $\hat{E}_I$. d_{pi} and $d_{\Theta i}$ indicate external disturbance acting on the translational and rotational motion in $\hat{E}_I$ and $\hat{E}_B$. F_i and τ_i are given by $F_i = c_{3,3} k_{\omega i} \sum_{k=1}^4 \omega_{k,i}^2 - R_{IB}^T c_{3,3} m_i g$ and $\tau_i = [\tau_{i,x} \;\; \tau_{i,y} \;\; \tau_{i,z}]^T$, respectively, where $\tau_{i,x} = l_{ti} k_{\omega i}(\omega_{1,i}^2 - \omega_{3,i}^2)$, $\tau_{i,y} = l_{ti} k_{\omega i}(\omega_{2,i}^2 - \omega_{4,i}^2)$, and $\tau_{i,z} = k_{\tau i} k_{\omega i} \sum_{k=1}^4 (-1)^{k+1} \omega_{k,i}^2$, g is the gravity constant, $\omega_{j,i}$ is the spinning rate of the jth rotor of the ith quadrotor, and l_{ti}, $k_{\omega i}$, and $k_{\tau i}$ are three scaling factors of the ith quadrotor. Define the control input commands as $u_{zi} = \sum_{k=1}^4 \omega_{k,i}^2$, $u_{\phi i} = \omega_{2,i}^2 - \omega_{4,i}^2$, $u_{\theta i} = \omega_{1,i}^2 - \omega_{3,i}^2$, and $u_{\psi i} = \sum_{k=1}^4 (-1)^{k+1} \omega_{k,i}^2$. Because of the underactuation feature of the quadrotor dynamics, design a virtual position control input $u_{pi} \in \mathbb{R}^3$ as follows:

$$u_{pi} = u_{zi} \begin{bmatrix} \sin\phi_{ri}\sin\psi_{ri} + \cos\phi_{ri}\cos\psi_{ri}\sin\theta_{ri} \\ \cos\phi_{ri}\sin\psi_{ri}\sin\theta_{ri} - \cos\psi_{ri}\sin\psi_{ri} \\ \cos\phi_{ri}\cos\theta_{ri} \end{bmatrix} \tag{2}$$

where ϕ_{ri}, θ_{ri}, and ψ_{ri} are attitude reference for the ith quadrotor. Denote $b_{pi} = k_{\omega i} I_3/m_i (i \in \mathcal{F})$ and $b_{\Theta i} = \text{diag}\{b^1_{\Theta i}, b^2_{\Theta i}, b^3_{\Theta i}\} = \text{diag}\{l_{ti}k_{\omega i}, l_{ti}k_{\omega i}, k_{\tau i}\}$. In this case, one can write the quadrotor dynamics in (1) as

$$\begin{aligned} \ddot{p}_{fi} &= b_{pi}u_{pi} - gc_{3,3} + \Delta_{pi} \\ \ddot{\Theta}_i &= -J_i^{-1}\big(C(\Theta_i, \dot{\Theta}_i)\dot{\Theta}_i + b_{\Theta i}u_{\Theta i} + d_{\Theta i}\big) \end{aligned} \tag{3}$$

where $u_{\Theta i} = [u_{\phi i} \;\; u_{\theta i} \;\; u_{\psi i}]^T \in \mathbb{R}^3$, Δ_{pi} represents external disturbance given by $\Delta_{pi} = b_{pi}\tilde{u}_{pi} + d_{pi}$, where $\tilde{u}_{pi} = u_{zi}R_{IB}c_{3,3} - u_{pi}$.

Remark 1: It can be observed from (1) that the dynamical model of each quadrotor, involving six degrees of freedom but four control inputs, is an underactuated system. Moreover, the nonlinear quadrotor system is coupled and subject to external disturbance. Therefore, it is not feasible to extend the existing results on the containment control of linear systems (see [6], [7], [8], [9], [10], [11]) to nonlinear quadrotors.

C. Problem Statement

From [24], the dynamics of the team leaders with unknown inputs can be described as follows:

$$\begin{aligned} \dot{\zeta}_{lv} &= M_l\zeta_{lv} + G_l u_{lv} \\ p_{lv} &= N_l\zeta_{lv}, v \in \mathcal{L} \end{aligned} \tag{4}$$

where $M_l \in \mathbb{R}^{6\times 6}$ is the dynamic matrix of the leaders, $N_l = [c_{3,1} \;\; c_{3,1} \;\; c_{3,1} \;\; 0_{3\times 3}]$, $\zeta_{lv} = [p_{lv}^T \;\; \dot{p}_{lv}^T]^T \in \mathbb{R}^6$ and $p_{lv}(t) \in \mathbb{R}^3$ are the state and the position of the vth leader, respectively.

Assumption 1: For each quadrotor agent in the containment system, there is at least one UAV leader that has a directed path to the quadrotor.

Assumption 2: The unknown control input $u_{lv}(v \in \mathcal{L})$ for all team leaders is continuous and bounded by a threshold $\varpi_{lv} > 0$, that is, $\|u_{lv}\|_\infty \le \varpi_{lv}$.

Let $e_{pi} \in \mathbb{R}^3, (i \in \mathcal{F})$ be the local relative output information of the ith follower given by:

$$e_{pi} = \sum_{j=1}^{n_f} a_{ij}\big(p_{fj} - p_{fi}\big) + \sum_{v=n_f+1}^{n_f+n_l} \rho_i^v\big(p_{lv} - p_{fi}\big). \tag{5}$$

The compact form of (5) can be written as

$$e_p = -\sum_{v=n_f+1}^{n_f+n_l} (\Phi_v \otimes I_3)\big(\bar{p}_f - \tilde{p}_{lv}\big) \tag{6}$$

where $e_p = [e_{p1}^T, e_{p2}^T, \ldots, e_{pn_f}^T]^T \in \mathbb{R}^{3n_f}$, $\tilde{p}_{lv} = 1_{n_l} \otimes p_{lv}$, and $\Phi_v = (1/n_f)L_f + W_v$, where L_f is the Laplacian matrix of $\mathcal{G}_f$. The purpose of this article is to obtain an optimal containment control law without requiring accurate information of the quadrotor dynamics under external disturbances and multiple active leaders. The optimal control problem of this article can be summarized by Problem 1.

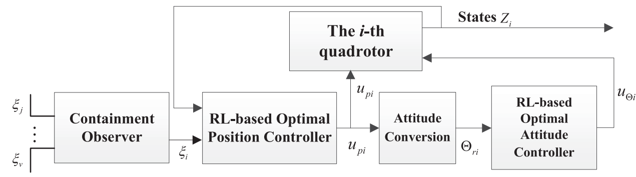

Fig. 1. Structure of the proposed RL-based controller.

Problem 1 (Optimal Containment Control): Consider the optimization problem for the nonlinear quadrotor vehicles under external disturbances modeled as (1), estimate the position references $\hat{p}_{ri}$ by using local information, and achieve the optimal position containment control, that is, $\lim_{t\to\infty}\text{dist}[p_{fi}(t), Co(p_{lv}(t), v \in \mathcal{L})] = 0$, by designing optimal control laws u_{pi}^*, Δ_{pi}^*, $u_{\Theta i}^*$ and $\Delta_{\Theta i}^*$ to minimize the performance index

$$J_{ki} = \int_t^\infty e^{\alpha_k(t-\tau)} r_k(x_{ki}, r_{ki}, u_{ki}, \Delta_{ki})d\tau, k = p, \Theta \tag{7}$$

where $r_k(x_{ki}, r_{ki}, u_{ki}, \Delta_{ki})$ is a value function which will be designed in the controller design section.

According to [25] and [26], the containment of the followers can be achieved if the global containment error e_p converges to 0, that is, $\lim_{t\to\infty} e_p(t) = 0$.

III. Containment Control Law

In this section, the containment control law for the quadrotors is proposed, consisting of three parts: 1) a containment observer to generate desired trajectory references; 2) an optimal position controller to track the trajectory references and produce the attitude references; and 3) an optimal attitude controller to track the generated attitude references. Fig. 1 depicts the structure of the proposed controller.

A. Optimal Position Control Law

Let $\xi_i = [\hat{p}_{ri}^T \;\; \dot{\hat{p}}_{ri}^T]^T \in \mathbb{R}^6$ denote the state vector of the ith observer, where $\hat{p}_{ri} \in \mathbb{R}^3$ is the position reference for the ith quadrotor to track. Let $\epsilon_{oi} \in \mathbb{R}^6 (i \in \mathcal{F})$ be the local estimation error of the ith observer. Substituting p_{fi} by ξ_{fi} in (6) yields that $\epsilon_{oi} = \sum_{j=1}^{n_f} a_{ij}(\xi_{fj} - \xi_{fi}) + \sum_{v=n_f+1}^{n_f+n_l} \rho_i^v(\zeta_{lv} - \xi_{fi})$. Design the following observer to guarantee the convergence of the global estimation error as:

$$\begin{aligned} \dot{\xi}_i &= M_l\xi_i + G_l\iota_i \\ \iota_i &= \varrho_1 S_o\epsilon_{oi} + \varrho_2\hat{s}(S_o\epsilon_{oi}) \end{aligned} \tag{8}$$

where ι_i is the control input, ϱ_1, ϱ_2 are positive constant gains, S_o is a matrix to be determined, and $\hat{s}(x)$ represents the signum function. Then, from (8), one can obtain the global form of the observer dynamics as follows:

$$\begin{aligned} \dot{\xi}_p = &\big(I_{n_f} \otimes M_l\big)\xi_p + \big(\varrho_1 I_{n_f} \otimes M_l S_o\big)\epsilon_o \\ &+ \big(\varrho_2 I_{n_f} \otimes G_l\big)\hat{F}(\epsilon_o) \end{aligned} \tag{9}$$

where $\xi_p = [\xi_1^T, \xi_2^T, \ldots, \xi_{n_f}^T]^T$, $\epsilon_o = [\epsilon_{o1}^T, \epsilon_{o2}^T, \ldots, \epsilon_{on_f}^T]^T$, and $\hat{F}(\epsilon_o) = [\hat{s}^T(S_o\epsilon_{o1}), \hat{s}^T(S_o\epsilon_{o2}), \ldots, \hat{s}^T(S_o\epsilon_{on_f})]^T$. From (9), one can obtain the dynamical system of the estimation error as

$$\dot{\epsilon}_o = \tilde{M}_l \epsilon_o + \tilde{\Phi}\tilde{G}_l \iota - \sum_{v=n_f+1}^{n_f+n_l} (\Phi_v \otimes I_6)\tilde{G}_l \tilde{u}_{lv} \tag{10}$$

where $\tilde{M}_l = I_{n_f} \otimes M_l$, $\Phi = \sum_{v=1}^{n_l} \Phi_v$, $\tilde{\Phi} = \Phi \otimes I_6$, $\tilde{G}_l = I_{n_f} \otimes G_l$, $\tilde{u}_{lv} = 1_{n_f} \otimes u_{lv}$, $\iota = [\iota_1^T, \iota_2^T, \ldots, \iota_{n_f}^T]^T$. Actually, since the function $\hat{s}(x)$ is Lebesgue measurable and locally essentially bounded, one can obtain the Filippov solutions to (10). Therefore, the stability property of (10) can be analyzed based on the theory of differential inclusion and nonsmooth analysis. The dynamics of ϵ_o in forms of differential inclusions can be obtained as

$$\dot{\epsilon}_o \in^{a.e.} \mathcal{K}\left[\tilde{M}_l \epsilon_o + \tilde{\Phi}\tilde{G}_l \iota - \sum_{v=n_f+1}^{n_f+n_l} (\Phi_v \otimes I_6)\tilde{G}_l \tilde{u}_{lv}\right] \tag{11}$$

where a.e. stands for "almost everywhere" as shown in [27]. From [28], the estimation error in (10) can asymptotically converge to 0, if Assumptions 1 and 2 hold and the positive constant gains ϱ_1 and ϱ_2, and the matrix S_o in (8) are selected as

$$\begin{aligned} S_o &= -G_l^T P^{-1}, \varrho_1 \geq 1/\lambda_{\min}\big(L_f + \Phi\big) \\ \varrho_2 &\geq \max_{v \in \mathcal{L}} \varpi_{lv} \end{aligned} \tag{12}$$

where $P > 0$ in the first equation of (12) satisfies that

$$M_l P + P M_l^T - 2G_l G_l^T < 0. \tag{13}$$

Furthermore, the optimal position control law is designed to track the trajectory reference generated from the observer and produce attitude reference. One can rewrite the translational dynamics in (3) as

$$\begin{aligned} \dot{z}_{pi} &= M_{pi} z_{pi} + G_{pi} u_{pi} - g c_{6,6} + D_{pi}\Delta_{pi} \\ y_{pi} &= N_{pi} z_{pi} \end{aligned} \tag{14}$$

where $z_{pi} = [p_{fi}^T \ \ \dot{p}_{fi}^T]^T \in \mathbb{R}^6$, $G_{pi} = [0 \ \ b_{pi}^T]^T$, $D_{pi} = [0 \ \ I_3]^T$, $M_{pi} = [c_{6,4} \ \ c_{6,5} \ \ c_{6,6} \ \ 0_{6\times 3}]^T$, and $N_{pi} = N_l$. Combining (8) and (14) leads to the following position augmented system:

$$\begin{aligned} \dot{Z}_{pi} &= \bar{M}_{pi} Z_{pi} + \bar{G}_{pi} u_{pi} - c_{12,6} g + \bar{D}_{pi}\Delta_{pi} + T_{pi}\mu_{pi} \\ \delta_{pi} &= \bar{N}_{pi} Z_{pi} \end{aligned} \tag{15}$$

where $Z_{pi} = [z_{pi}^T \ \ \xi_i^T]^T \in \mathbb{R}^{12}$, $\bar{M}_{pi} = \mathrm{diag}\{M_{pi} \ \ M_l\}$, $\bar{G}_{pi} = [G_{pi}^T \ \ 0]$, $\bar{N}_{pi} = [N_{pi} \ - \ N_l]$, $\bar{D}_{pi} = [D_{pi}^T \ \ 0]^T$, $T_{pi} = [0 \ \ G_l]^T$, and $\mu_{pi} = \varrho_1 S\epsilon_{oi} + \varrho_2 \hat{s}(S\epsilon_{oi})$. The bounded input μ_{pi} resulted from local estimation error ϵ_{oi} and the equivalent disturbance Δ_{pi} produces uncertain effects to the system in (15) and should be attenuated eventually in the augmented system. Besides, using the proposed containment observer in (8) for each quadrotor, the estimation error ϵ_{oi} can converge to 0 and consequently drive μ_{pi} to 0. To counteract the external disturbance affecting on the augmented system (15), one can consider the following disturbance attenuation condition as:

$$\frac{\int_t^{\infty} e^{-\alpha_i(\tau - t)}\Big(\delta_{pi}^T Q_{pi}\delta_{pi} + u_{pi}^T R_{pi} u_{pi}\Big)d\tau}{\int_t^{\infty} e^{-\alpha_i(\tau - t)}\Delta_{pi}^T \Delta_{pi} d\tau} \leq \gamma_p^2 \tag{16}$$

where $\delta_{pi} = p_{fi} - \hat{p}_{ri}$, $\alpha_i > 0$ represents a discount constant, $\gamma_p \geq 0$, Q_{pi} and R_{pi} are positive-definite matrices. From (16), one can see that γ_p indicates the scale of attenuation from the effects of Δ_{pi} to the performance of the translational system.

Consider the following construction of the performance index for the position augmented system as:

$$J_{pi}\big(\delta_{pi}, u_{pi}, \Delta_{pi}\big) = \int_t^{\infty} e^{-\alpha_i(\tau - t)} r_p\big(\delta_{pi}, u_{pi}, \Delta_{pi}\big)d\tau \tag{17}$$

where $r_p(\delta_{pi}, u_{pi}, \Delta_{pi}) = \delta_{pi}^T Q_{pi}\delta_{pi} + u_{pi}^T R_{pi} u_{pi} - \gamma_p^2 \Delta_{pi}^T \Delta_{pi}$. In fact, it can be observed from (17) that u_{pi} and Δ_{pi} engage in a two-player zero-sum differential game, where u_{pi} is the minimizing player and Δ_{pi} is the maximizing player. Define the Nash condition of the differential game as

$$J_{pi}^* = \min_{u_{pi}} \max_{\Delta_{pi}} \int_t^{\infty} e^{-\alpha_i(\tau - t)} r_p\big(\delta_{pi}, u_{pi}, \Delta_{pi}\big)d\tau \tag{18}$$

where $J_{pi}^*(\delta_{pi}, u_{pi}, \Delta_{pi})$ is the optimal performance index. If the Nash condition in (18) holds for the position augmented system, the solution to the differential game problem is unique. The optimal position control part aims at designing the position control law u_{pi} satisfying inequality in (16) such that p_i tracks the position reference generated by the observer, while minimizing the performance index given by (18). One can obtain the Hamiltonian function as follows:

$$\begin{aligned} H\big(J_{pi}, u_{pi}, \Delta_{pi}\big) \triangleq\ & r_p\big(\delta_{pi}, u_{pi}, \Delta_{pi}\big) - \alpha_i J_{pi} \\ & + \Delta J_{pi}^T\big(\bar{M}_{pi} Z_{pi} + \bar{G}_{pi} u_{pi}\big) \\ & - \Delta J_{pi}^T\big(c_{12,6} g - \bar{D}_{pi}\Delta_{pi}\big) \end{aligned} \tag{19}$$

where $\Delta J_{pi} = \partial J_{pi}/\partial Z_{pi}$. According to [29], differentiating (19) with respect to the control command u_{pi} and the disturbance input Δ_{pi}, that is, $\partial H(J_{pi}^*, u_{pi}, \Delta_{pi})/\partial u_{pi} = 0$ and $\partial H(J_{pi}^*, u_{pi}, \Delta_{pi})/\partial \Delta_{pi} = 0$, yields the following optimal position control law for achieving containment and the disturbance input:

$$\begin{aligned} u_{pi}^* &= -R_{pi}^{-1}\bar{G}_{pi}^T \Delta J_{pi}^*/2 \\ \Delta_{pi}^* &= \bar{D}_{pi}^T \Delta J_{pi}^* / \big(2\gamma_p^2\big). \end{aligned} \tag{20}$$

Besides, substituting (20) into (19) leads to

$$\begin{aligned} &\delta_{pi}^T Q_{pi}\delta_{pi} - \alpha_i J_{pi} + \Big(\Delta J_{pi}^*\Big)^T \big(\bar{M}_{pi} Z_{pi} - c_{12,6} g\big) \\ &- \frac{1}{4}\Big(\Delta J_{pi}^*\Big)^T\left[\bar{G}_{pi} R_{pi}^{-1}\bar{G}_{pi}^T - \frac{1}{\gamma_p^2}\bar{D}_{pi}\bar{D}_{pi}^T\right]\Delta J_i^* = 0. \end{aligned} \tag{21}$$

Theorem 1 summarizes the stability property of the translational subsystem utilizing the control law in (20).

Theorem 1: The optimal control law in (20) can guarantee that the closed-loop position augmented system in (15) is asymptotically stable, and the disturbance attenuation condition in (16) is satisfied, with $\Delta_{pi} = 0$ and $\alpha_i \leq 2(\|U_{pi} Q_{pi}\|)^{1/2}$, where $U_{pi} = G_{pi} R_{pi}^{-1} G_{pi}^T + D_{pi} D_{pi}^T/\gamma_p^2$.

Proof: Combining (19)–(21) and substituting $u_{pi} = u_{pi}^*$ yield that

$$-\gamma_p^2\Big(\Delta_{pi} - \Delta_{pi}^*\Big)^T\Big(\Delta_{pi} - \Delta_{pi}^*\Big) \leq 0. \tag{22}$$

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

Then, from (19), one can obtain that

$$-\alpha_i J_{pi}^* + \dot{J}_{pi}^* \leq -r_p\left(\delta_{pi}, u_{pi}^*, \Delta_{pi}\right). \tag{23}$$

Multiplying $e^{-\alpha_i t}$ on both sides of (23) and integrating both sides yield that

$$\begin{aligned} &e^{-\alpha_i T} J_{pi}^*\big(Z_{pi}(T)\big) - J_{pi}^*\big(Z_{pi}(0)\big) \\ &\leq -\int_0^T e^{-\alpha_i \tau} r_p\left(\delta_{pi}, u_{pi}^*, \Delta_{pi}\right) d\tau. \end{aligned} \tag{24}$$

Because $J_{pi}^*(Z_{pi}(t)) \geq 0$, one can obtain that

$$\int_0^T e^{-\alpha_i \tau} r_p\left(\delta_{pi}, u_{pi}^*, \Delta_{pi}\right) d\tau \leq J_{pi}^*\big(Z_{pi}(0)\big). \tag{25}$$

From (25), the disturbance attenuation condition (16) holds for the position augmented system with u_{pi}^* in (20). From (21), it follows that:

$$\begin{aligned} &\left(\Delta J_{pi}^*\right)^T \left(\bar{M}_{pi} Z_{pi} + \bar{G}_{pi} u_{pi} - c_{12,6} g + \bar{D}_{pi} \Delta_{pi}\right) \\ &= \alpha_i J_{pi}^* - r_p\left(\delta_{pi}, u_{pi}^*, \Delta_{pi}\right). \end{aligned} \tag{26}$$

One can multiply the both sides of (26) with $e^{-\alpha_i t}$ and obtain that

$$\frac{d}{dt}\left(e^{-\alpha_i t} J_{pi}^*\right) = -e^{-\alpha_i t} r_p\left(\delta_{pi}, u_{pi}^*, \Delta_{pi}\right). \tag{27}$$

Therefore, from (27), it can be concluded that the augmented system in (15) is asymptotically stable with $\alpha_i = 0$ and $\Delta_{pi} = 0$. From [29], if α_i is nonzero, the augmented system in (15) is still stable if α_i satisfies that $\alpha_i \leq 2(\|U_{pi} Q_{pi}\|)^{1/2}$, where $U_{pi} = G_{pi} R_{pi}^{-1} G_{pi}^T + D_{pi} D_{pi}^T / \gamma_p^2$. ■

To facilitate the implementation, a model-based reinforcement learning method is provided to iteratively computes the optimal position control law as follows. Let J_{pi}^n, u_{pi}^n, and Δ_{pi}^n be the updated terms in the nth iteration. First, an initial arbitrary control policy u_{pi}^0 and disturbance input Δpi^0 are selected. Then, the performance index J_{pi}^n can be solved by utilizing the current control policy u_{pi}^n and disturbance input Δ_{pi}^n from the following Bellman equation:

$$\begin{aligned} &\left(\Delta J_{pi}^n\right)^T \left(\bar{M}_{pi} Z_{pi} + \bar{G}_{pi} u_{pi}^n - c_{12,6} g + \bar{D}_{pi} \Delta_{pi}^n\right) \\ &+ r_p\left(\delta_{pi}, u_{pi}^n, \Delta_{pi}^n\right) - \alpha_i J_{pi}^n = 0. \end{aligned} \tag{28}$$

In this case, the control law u_{pi}^{n+1} and disturbance input Δ_{pi}^{n+1} can be updated according to the following equation:

$$\begin{aligned} u_{pi}^{n+1} &= -R_{pi}^{-1} \bar{G}_{pi}^T \Delta J_{pi}^n / 2 \\ \Delta_{pi}^{n+1} &= \bar{D}_{pi}^T \Delta J_{pi}^n / \left(2\gamma_p^2\right). \end{aligned} \tag{29}$$

The above steps can be repeated from computing the performance function until a satisfactory solution is reached, that is, $u_{pi}^{n+1} = u_{pi}^n$ and $\Delta_{pi}^{n+1} = \Delta_{pi}^n$. By this way, the optimal control law u_{pi}^* and the disturbance input Δ_{pi}^* can be obtained. In fact, it can be obtained from (28) and (29) that u_{pi}^n and Δ_{pi}^n are updated after calculating J_{pi}^n. However, the accurate information of the quadrotor dynamics is required for the model-based approach resulting in difficulties in implementation. In this case, an RL method is developed to obviate the requirement of accurate knowledge of the vehicle dynamics. The position augmented system in (15) can be rewritten as

$$\begin{aligned} \dot{Z}_{pi} = &\bar{M}_{pi} Z_{pi} + \bar{G}_{pi} u_{pi}^n - c_{12,6} g + \bar{D}_{pi} \Delta_{pi}^n + T_{pi} \mu_{pi} \\ &+ \bar{G}_{pi}\left(u_{pi} - u_{pi}^n\right) + \bar{D}_{pi}\left(\Delta_{pi} - \Delta_{pi}^n\right). \end{aligned} \tag{30}$$

Differentiating J_{pi} along with system dynamics (30) and using (21) lead to

$$\begin{aligned} \dot{J}_{pi}^n = &\alpha_i J_{pi}^n - 2\left(u_{pi}^{n+1}\right)^T R_{pi}\left(u_{pi} - u_{pi}^n\right) \\ &+ \left(\Delta J_{pi}^n\right)^T T_{pi} \mu_{pi} - r_p\left(\delta_{pi}, u_{pi}^n, \Delta_{pi}^n\right) \\ &+ \left(2\gamma_p^2 \Delta_{pi}^{n+1}\right)^T \left(\Delta_{pi} - \Delta_{pi}^n\right). \end{aligned} \tag{31}$$

To obtain the temporal difference of the performance index J_{pi}, one can multiply $e^{-\alpha_i t}$ on both sides of (31) and integrate both sides, resulting the Bellman equation as

$$\begin{aligned} &\int_t^{t+\delta_T} \frac{d}{d\tau}\left(e^{-\alpha_i(\tau - t)} J_{pi}^n\big(Z_{pi}(\tau)\big)\right) d\tau \\ &= -\int_t^{t+\delta_T} e^{\alpha_i(t-\tau)} 2\left(u_{pi}^{n+1}\right)^T R_{pi}\left(u_{pi} - u_{pi}^n\right) d\tau \\ &- \int_t^{t+\delta_T} e^{\alpha_i(t-\tau)} r_p\left(\delta_{pi}, u_{pi}^n, \Delta_{pi}^n\right) d\tau \\ &+ \int_t^{t+\delta_T} e^{\alpha_i(t-\tau)} \left(\Delta_{pi}^n\right)^T T_{pi} \mu_{pi} d\tau \\ &+ 2\gamma_p^2 \int_t^{t+\delta_T} e^{\alpha_i(t-\tau)} \left(\Delta_{pi}^{n+1}\right)^T \left(\Delta_{pi} - \Delta_{pi}^n\right) d\tau. \end{aligned} \tag{32}$$

The optimal position control law can be obtained by the following steps without knowledge of dynamical parameters of the quadrotors. First, apply an incipient admissible position control law u_{pi}^a and a persistent exploring control input u_{pi}^e to the quadrotor translational system with an existed disturbance input Δ_{pi}. Record the historical data of the system output information Z_{pi}, position control command u_{pi}, and disturbance input Δ_{pi}. Second, select an initial control law u_{pi}^0 and disturbance input Δ_{pi}^0, and substitute them into the Bellman equation in (32). Solve the Bellman equation in (32) to update the performance function J_{pi}^{n+1}, position control law u_{pi}^{n+1}, and disturbance input Δ_{pi}^{n+1}, simultaneously. Third, repeat the updating step until the convergence is achieved. By this way, the optimal control law $u_{pi}^* = u_{pi}^{n+1}$ and disturbance input $\Delta_{pi}^* = \Delta_{pi}^{n+1}$ can be obtained. Moreover, the following Theorem 2 shows that the Bellman equation in (32) is equivalent to (28) and (29).

Theorem 2: If and only if $(J_{pi}^n, u_{pi}^{n+1}, \Delta_{pi}^{n+1})$ satisfies the Bellman equation (28) in the model-based method with $J_{pi}^n(0) = 0$, it is the solution to (32).

Proof: One can differentiate (32) and obtain that the solution to (28) satisfies (32). This proof begins with showing that the solution to (32) using the RL-based approach is unique. Differentiating (32) yields that

$$\begin{aligned} &\frac{d}{dt}\left(e^{-\alpha_i t} J_{pi}^n\big(Z_{pi}(t)\big)\right) \\ &= e^{-\alpha_i t}\left(\Delta_{pi}^n\right)^T T_{pi} \mu_{pi} - e^{-\alpha_i t} r_p\left(\delta_{pi}, u_{pi}^n, \Delta_{pi}^n\right) \end{aligned}$$

$$+ 2\gamma_p^2 e^{-\alpha_i t}\left(\Delta_{pi}^{n+1}\right)^T\left(\Delta_{pi} - \Delta_{pi}^n\right) - 2e^{-\alpha_i t}\left(u_{pi}^{n+1}\right)^T R_{pi}\left(u_{pi} - u_{pi}^n\right). \quad (33)$$

If there exists another solution to (32) given by $(\tilde{J}_{pi}^n, \tilde{u}_{pi}^{n+1}, \tilde{\Delta}_{pi}^{n+1})$ satisfying (33), it follows that:

$$\begin{aligned}\frac{d}{dt}\left(e^{-\alpha_i t}\tilde{J}_{pi}^n(Z_{pi}(t))\right) &= e^{-\alpha_i t}\left(\tilde{\Delta}_{pi}^n\right)^T T_{pi}\mu_{pi} - e^{-\alpha_i t} r_p\left(\delta_{pi}, u_{pi}^n, \Delta_{pi}^n\right)\\ &+ 2\gamma_p^2 e^{-\alpha_i t}\left(\tilde{\Delta}_{pi}^{n+1}\right)^T\left(\Delta_{pi} - \Delta_{pi}^n\right)\\ &- 2e^{-\alpha_i t}\left(\tilde{u}_{pi}^{n+1}\right)^T R_{pi}\left(u_{pi} - u_{pi}^n\right). \end{aligned} \quad (34)$$

Because the position observer (8) can guarantee that μ_{pi} in (34) converge to 0. Then, combining (33) and (34) yields that

$$\begin{aligned}\frac{d}{dt}\left[e^{-\alpha_i t}J_{pi}^n(Z_{pi}) - e^{-\alpha_i t}\tilde{J}_{pi}^n(Z_{pi})\right] &= 2\gamma_p^2 e^{-\alpha_i t}\left(\Delta_{pi}^{n+1} - \tilde{\Delta}_{pi}^{n+1}\right)^T\left(\Delta_{pi} - \Delta_{pi}^n\right)\\ &- 2e^{-\alpha_i t}\left(u_{pi}^{n+1} - \tilde{u}_{pi}^{n+1}\right)^T R_{pi}\left(u_{pi} - u_{pi}^n\right). \end{aligned} \quad (35)$$

Equation (35) holds for any given u_{pi} and Δ_{pi}. Let $u_{pi} = u_{pi}^n$ and $\Delta_{pi} = \Delta_{pi}^n$. It follows that:

$$\frac{d}{dt}\left[e^{-\alpha_i t}J_{pi}^n(Z_{pi}(t)) - e^{-\alpha_i t}\tilde{J}_{pi}^n(Z_{pi}(t))\right] = 0. \quad (36)$$

From (36), it can be obtained that $e^{-\alpha_i t}J_{pi}^n(Z_{pi}) - e^{-\alpha_i t}\tilde{J}_{pi}^n(Z_{pi}) = \varsigma$, where ς is a constant satisfying that $\varsigma = J_{pi}^n(0) - \tilde{J}_{pi}(0) = 0$. It follows that $J_{pi}^n = \tilde{J}_{pi}$.

From (35), one can have that

$$\begin{aligned}2\gamma_p^2 e^{-\alpha_i t}\left(\tilde{\Delta}_{pi}^{n+1} - \Delta_{pi}^{n+1}\right)^T\left(\Delta_{pi}^n - \Delta_{pi}\right) = 2e^{-\alpha_i t}\left(\tilde{u}_{pi}^{n+1} - u_{pi}^{n+1}\right)^T R_{pi}\left(u_{pi}^n - u_{pi}\right). \end{aligned} \quad (37)$$

Because (37) holds for any control law u_{pi} and disturbance input Δ_{pi}, one can have that $u_{pi}^{n+1} = \tilde{u}_{pi}^{n+1}$, $\Delta_{pi}^{n+1} = \tilde{\Delta}_{pi}^{n+1}$. Therefore, there exists only one solution $(J_{pi}^n, u_{pi}^{n+1}, \Delta_{pi}^{n+1})$ that satisfies (32). ∎

One can observe from the Bellman equation (32) that it integrates policy evaluation and policy improvement. Theorem 2 shows that the optimal model-free position control law is essentially equivalent to the optimal model-based position control law. Therefore, Theorem 1 can also guarantee the convergence of the proposed RL-based method.

From Weierstrass theorem, the performance index J_{pi}^n, the control command u_{pi}^{n+1}, and the disturbance input Δ_{pi}^{n+1} can be approximated by the three following neural networks as:

$$\begin{aligned}\hat{J}_{pi}^n(Z_{pi}) &= \hat{K}_{pi1}\sigma_{pi1}(Z_{pi})\\ \hat{u}_{pi}^{n+1}(Z_{pi}) &= \hat{K}_{pi2}\sigma_{pi2}(Z_{pi})\\ \hat{\Delta}_{pi}^{n+1}(Z_{pi}) &= \hat{K}_{pi3}\sigma_{pi3}(Z_{pi}), i \in \mathcal{F} \end{aligned} \quad (38)$$

where $\hat{J}_{pi}^n(Z_{pi})$, $\hat{u}_{pi}^{n+1}(Z_{pi})$, and $\hat{\Delta}_{pi}^{n+1}$ are the approximated values, $\sigma_{pi1}(Z_{pi}) \in \mathbb{R}^{n_{p1}\times 1}$, $\sigma_{pi2}(Z_{pi}) \in \mathbb{R}^{n_{p2}\times 1}$, and $\sigma_{pi3} \in \mathbb{R}^{n_{p3}\times 1}$ are the basis functions with n_{p1}, n_{p2}, and n_{p3} neurons, $\hat{K}_{pi1} \in \mathbb{R}^{1\times n_{p1}}$, $\hat{K}_{pi2} \in \mathbb{R}^{3\times n_{p2}}$, and $\hat{K}_{pi3} \in \mathbb{R}^{3\times n_{p3}}$ are the weighted matrices. Let $R_{pi} = \text{diag}\{r_{pi}^1, r_{pi}^2, r_{pi}^3\}$, $\vartheta^{p1} = [\vartheta_1^{p1}\ \vartheta_2^{p1}\ \vartheta_3^{p1}]^T = u_{pi} - u_{pi}^n$, and $\vartheta^{p2} = [\vartheta_1^{p2}\ \vartheta_2^{p2}\ \vartheta_3^{p2}]^T = \Delta_{pi} - \Delta_{pi}^n$. Substituting (38) into (32) results

$$\begin{aligned}\hat{\varepsilon}_{pi} &= \int_t^{t+\delta_T} e^{\alpha_i(t-\tau)} r_p\left(\delta_{pi}, u_{pi}^n, \Delta_{pi}^n\right)d\tau\\ &+ e^{-\alpha_i\delta_T}\hat{K}_{pi1}\sigma_{pi1}\left(Z_{pi}(t+\delta_T)\right) - \hat{K}_{pi1}\sigma_{pi1}\left(Z_{pi}(\delta_T)\right)\\ &+ 2\sum_{m=1}^{3} r_{pi}^m \int_t^{t+\delta_T} e^{\alpha_i(t-\tau)}\hat{K}_{pi2,n}\sigma_{pi2}\left(Z_{pi}(\tau)\right)\vartheta_m^{p1}d\tau\\ &- 2\gamma_p^2\sum_{m=1}^{3}\int_t^{t+\delta_T} e^{\alpha_i(t-\tau)}\hat{K}_{pi3,k}\sigma_{pi3}\left(Z_{pi}(\tau)\right)\vartheta_m^{p2}d\tau\\ &- \int_t^{t+\delta_T} e^{\alpha_i(t-\tau)}\hat{K}_{pi3}\sigma_{pi3}\left(Z_{pi}(\tau)\right)T_{pi}\mu_{pi}d\tau \end{aligned} \quad (39)$$

where $\hat{K}_{pi2,n}$ indicates the nth column of $\hat{K}_{pi2}$, and $\hat{K}_{pi3,k}$ the kth column of $\hat{K}_{pi3}$. By persisting excitation, the Bellman approximation error $\hat{\varepsilon}_{pi}(t)$ can converge to the origin using the least-squares method.

B. Optimal Attitude Control Law

Let $\Theta_{ri} = [\phi_{ri}\ \theta_{ri}\ \psi_{ri}]^T$ be the Euler angle reference of the quadrotor attitude system. One can write the actual control input u_{zi}, the reference of pitch channel θ_{ri}, and the reference of the roll channel ϕ_{ri} as

$$\begin{aligned}u_{zi} &= u_{pi}^z/(\cos\theta_i\cos\phi_i)\\ \phi_{ri} &= \arcsin\left[\left(\sin\theta_i\sin\psi_i\cos\phi_i - u_{pi}^y/u_{zi}\right)/\cos\psi_i\right]\\ \theta_{ri} &= \arcsin\left[\left(u_{pi}^x/u_{zi} - \sin\phi_i\sin\psi_i\right)/(\cos\psi_i\cos\phi_i)\right]. \end{aligned} \quad (40)$$

The optimal attitude control law is designed to track the references Θ_{ri} for the quadrotors. From (1), one can obtain the following rotational dynamics of the ith quadrotor as:

$$\begin{aligned}\dot{z}_{\Theta i} &= M_{\Theta i}(z_{\Theta i}) + G_{\Theta i}u_{\Theta i} + D_{\Theta i}d_{\Theta i}\\ y_{\Theta i} &= N_{\Theta i}z_{\Theta i} \end{aligned} \quad (41)$$

where $z_{\Theta i} = [\Theta_i^T\ \dot{\Theta}_i^T]^T \in \mathbb{R}^6$ and $y_{\Theta i} \in \mathbb{R}^3$ are the state and the output of rotational system, $G_{\Theta i} = [c_{6,4}b_{\Theta 1,i}\ c_{6,5}b_{\Theta 2,i}\ c_{6,6}b_{\Theta 3,i}]$, $D_{\Theta i} = [0_{3\times 3}\ I_3]^T$, and $N_{\Theta i} = [I_3\ 0_{3\times 3}]$. Denote $M_{\Theta i}(z_{\Theta i}) \in \mathbb{R}^6$ as the nonlinear term satisfying that

$$M_{\Theta i}(z_{\Theta i}) = \begin{bmatrix} 0_{3\times 3} & I_3 \\ 0_{3\times 3} & -J_i^{-1}C(\Theta_i, \dot{\Theta}_i)\end{bmatrix} z_{\Theta i}. \quad (42)$$

The dynamical system of the attitude reference generated by (40) satisfies that

$$\begin{aligned}\dot{z}_{\Theta ri} &= M_{\Theta ri}(z_{\Theta ri})\\ y_{\Theta ri} &= N_{\Theta ri}z_{\Theta ri} \end{aligned} \quad (43)$$

where $z_{\Theta ri} = [\Theta_{ri}^T\ \ \dot{\Theta}_{ri}^T] \in \mathbb{R}^6$, $N_{\Theta ri} = N_{\Theta i}$, $M_{\Theta ri}(z_{\Theta ri}) \in \mathbb{R}^6$ is an unknown smooth function, and $y_{\Theta ri} \in \mathbb{R}^3$ is the output. Combining (41) and (43) yields the following attitude augmented system:

$$\begin{aligned} Z_{\Theta i} &= \bar{M}_{\Theta}(Z_{\Theta i}) + \bar{G}_{\Theta i} u_{\Theta i} + \bar{D}_{\Theta i} d_{\Theta i} \\ \delta_{\Theta i} &= \bar{N}_{\Theta i} Z_{\Theta i} \end{aligned} \tag{44}$$

where $Z_{\Theta i} = [z_{\Theta i} \;\; z_{\Theta ri}] \in \mathbb{R}^{12}$, $\bar{M}_{\Theta i}(Z_{\Theta i}) = [M_{\Theta i}(z_{\Theta i}) \;\; M_{\Theta ri}(z_{\Theta ri})] \in \mathbb{R}^{12}$, $\bar{D}_{\Theta i} = [D_{\Theta i}^T \;\; 0]^T$, $\bar{N}_{\Theta i} = [N_{\Theta i} \;\; -N_{\Theta i}]$, and $\bar{G}_{\Theta i} = [G_{\Theta i} \; - \; G_{\Theta ri}]$. $\delta_{\Theta i} = [\delta_{\phi i} \;\; \delta_{\theta i} \;\; \delta_{\psi i}]^T \in \mathbb{R}^3$ in (44) is the tracking error of the quadrotor rotational motion. In order to track the attitude reference $z_{\Theta_{ri}}$, consider the following performance index of the attitude augmented system as:

$$J_{\Theta i}(\delta_{\Theta i}, u_{\Theta i}, \Delta_{\Theta i}) = \int_t^{\infty} e^{-\beta_i(\tau - t)} r_{\Theta}(\delta_{\Theta i}, u_{\Theta i}, \Delta_{\Theta i}) d\tau$$

where $r_{\Theta}(\delta_{\Theta i}, u_{\Theta i}, \Delta_{\Theta i}) = \delta_{\Theta i}^T Q_{\Theta i} \delta_{\Theta i} + u_{\Theta i}^T R_{\Theta i} u_{\Theta i} - \gamma_{\Theta}^2 \Delta_{\Theta i}^T \Delta_{\Theta i}$, $Q_{\Theta i} = Q_{\Theta i}^T > 0$, $R_{\Theta i} = R_{\Theta i}^T > 0$, $\gamma_{\Theta} \geq 0$, and $\beta_i > 0$. One can obtain the Hamiltonian function as follows:

$$\begin{aligned} H_{\Theta i}(J_{\Theta i}, u_{\Theta i}, \Delta_{\Theta i}) &\triangleq r_{\Theta}(\delta_{\Theta i}, u_{\Theta i}, \Delta_{\Theta i}) - \beta_i J_{\Theta i} \\ &\quad + \Delta J_{\Theta i}^T \big(\bar{F}_{\Theta i} Z_{\Theta i} + \bar{B}_{\Theta i} u_{\Theta i}\big) \end{aligned} \tag{45}$$

where $\Delta J_{\Theta i} = \partial J_{\Theta i} / \partial Z_{\Theta i}$. Let $J_{\Theta i}^*$ be the optimal performance index. Then, the optimal control law for the ith quadrotor can be obtained by differentiating (46) with respect to $u_{\Theta i}$ and $\Delta_{\Theta i}$, that is, $\partial H(J_{\Theta i}^*, u_{\Theta i}, \Delta_{\Theta i}) / \partial u_{\Theta i} = 0$, $\partial H(J_{\Theta i}^*, u_{\Theta i}, \Delta_{\Theta i}) / \partial \Delta_{\Theta i} = 0$. Then, one can obtain the optimal model-based control law $u_{\Theta i}^*$ and the disturbance input $\Delta_{\Theta i}^*$ as

$$\begin{aligned} u_{\Theta i}^* &= -\frac{1}{2} R_{\Theta i}^T \bar{G}_{\Theta i}^T \Delta J_{\Theta i}^* \\ \Delta_{\Theta i}^* &= \frac{1}{2\gamma_{\Theta}^2} \bar{D}_{\Theta i}^T \Delta J_{\Theta i}^*. \end{aligned} \tag{46}$$

Combining (45) and (46) yields that

$$\begin{aligned} &\big(\Delta J_{\Theta i}^*\big)^T \big(\bar{M}_{\Theta i}(Z_{\Theta i}) + \bar{G}_{\Theta i} u_{\Theta i} + \bar{D}_{\Theta i} \Delta_{\Theta i}\big) \\ &= \beta_i J_{\Theta i} - \frac{1}{4\gamma_{\Theta}^2} \big(\Delta J_{\Theta i}^*\big)^T \bar{D}_{\Theta i} \bar{D}_{\Theta i}^T \Delta J_{\Theta i}^* \\ &\quad - \delta_{\Theta i}^T Q_{\Theta i} \delta_{\Theta i} + \frac{1}{4} \big(\Delta J_{\Theta i}^*\big)^T \bar{G}_{\Theta i} R_{\Theta i}^{-1} \bar{G}_{\Theta i}^T \Delta J_{\Theta i}^*. \end{aligned} \tag{47}$$

Similarly to design of the optimal position control law, the stability of the rotational system by the optimal attitude control law in (46) can be guaranteed. Note that (47) is nonlinear for $J_{\Theta i}^*$ and requires accurate information of the rotational dynamics. In this case, the following steps are given to learn the optimal attitude control law without accurate information of rotational dynamics. First, apply an incipient admissible attitude control law $u_{\Theta i}^a$ and a persistent exploring control input $u_{\Theta i}^e$ to the quadrotor rotational system under a given disturbance input $\Delta_{\Theta i}$. Record the historical data of the system output information $Z_{\Theta i}$, attitude control command $u_{\Theta i}$, and disturbance input $\Delta_{\Theta i}$. Second, select an initial control law $u_{\Theta i}^0$ and disturbance input $\Delta_{\Theta i}^0$, and substitute them into the following Bellman equation in (48). Solve the Bellman equation in (48) and update the performance function $J_{\Theta i}^{n+1}$, attitude control law $u_{\Theta i}^{n+1}$, and disturbance input $\Delta_{\Theta i}^{n+1}$, simultaneously.

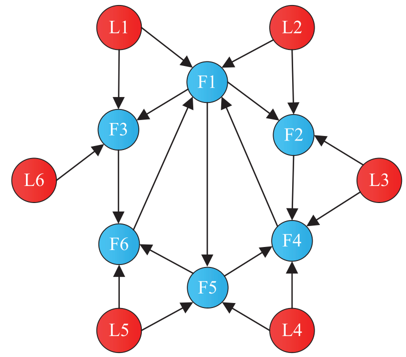

Fig. 2. Communication graph of the containment system.

Third, repeat from the updating step until the convergence is achieved. By this way, the optimal attitude control law $u_{\Theta i}^* = u_{\Theta i}^{n+1}$ and disturbance input $\Delta_{\Theta i}^* = \Delta_{\Theta i}^{n+1}$ can be obtained

$$\begin{aligned} &e^{-\beta_i T} J_{\Theta i}^n(Z_{\Theta i}(t+T)) - J_{\Theta i}^n(Z_{\Theta i}(t)) \\ &= -\int_t^{t+T} e^{-\beta_i(\tau - t)} r_{\Theta}(\delta_{\Theta i}, u_{\Theta i}, \Delta_{\Theta i}) d\tau \\ &\quad + \int_t^{t+T} e^{-\beta_i(\tau - t)} 2\gamma_{\Theta}^2 \Big(\Delta_{\Theta i}^{n+1}\Big)^T \big(\Delta_{\Theta i} - \Delta_{\Theta i}^n\big) d\tau \\ &\quad - \int_t^{t+T} e^{-\beta_i(\tau - t)} 2\Big(u_{\Theta i}^{n+1}\Big)^T R_{\Theta i} \big(u_{\Theta i} - u_{\Theta i}^n\big) d\tau. \end{aligned} \tag{48}$$

According to the proof of Theorem 2, one can prove the convergence of the RL-based attitude control method. Similarly to the derivation of the optimal model-free position control law, the performance index $J_{\Theta i}$, the control law $u_{\Theta i}^{n+1}$, and the disturbance $\Delta_{\Theta i}^{n+1}$ can be approximated by three neural networks and the weight matrices are $\hat{K}_{\Theta i1} \in \mathbb{R}^{1 \times n_{\Theta 1}}$, $\hat{K}_{\Theta i2} \in \mathbb{R}^{1 \times n_{\Theta 2}}$, and $\hat{K}_{\Theta i3} \in \mathbb{R}^{1 \times n_{\Theta 3}}$, where $n_{\Theta 1}$, $n_{\Theta 2}$, and $n_{\Theta 3}$ are the numbers of neurons, respectively. The weight matrices $\hat{K}_{\Theta i1}$, $\hat{K}_{\Theta i2}$, and $\hat{K}_{\Theta i3}$ can be updated using the least-squares method by persisting excitation.

Remark 2: Actually, the proposed RL-based optimal containment controller has two main advantages. First, it can learn from the historical data generated by a nonoptimal control law. Second, it can ultimately improve the performance of the control system without accurate information of the quadrotor system.

IV. Simulation Results

In this section, a containment system consisting of six leaders and six followers is constructed. The communication relationship among the containment system is depicted in Fig. 2. The leaders are modeled by (4) and are designed to form a pentagonal pyramid where five leaders cyclically change their positions at the five vertices of the base and the sixth leader is above the center of the base, demonstrating nonlinear movement pattern. The followers are modeled by nonlinear quadrotor dynamics in (3) with the following simulation configurations: $b_{pi} = \text{diag}\{1, 1, 1\}$, $J_i = \text{diag}\{5.6, 5.7, 9.9\} \times 10^{-3}$

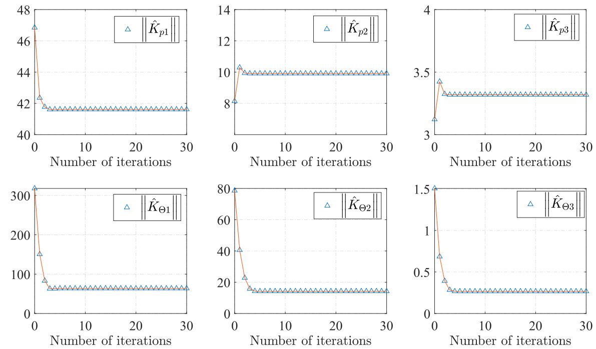

Fig. 3. Convergence of the weights using the proposed RL-based method.

kg $\cdot$ m^2, $b_{\Theta i} = \text{diag}\{43.4, 44.2, 115.8\}$ $(i = 1, 2, \ldots, 6)$ and $g = 9.81$ m/s^2. The parameters of the RL method are selected as $\alpha_i = 0.06$, $\gamma_p^2 = 4$, $R_{pi} = 2I_3$, $Q_{pi} = 17I_6$, $Q_{\Theta i} = 90I_6$, $R_{\Theta i} = I_3$, $\gamma_\Theta^2 = 6$, and $\beta_i = 0.06$. The data collection time interval is 0.05 s. The system performance index is approximated by multiple polynomials of even orders as the basis functions, while the control laws and disturbance use the ones of odd orders. Each quadrotor vehicle has an external persisting disturbance in the position and attitude subsystems given by $\Delta_{pi} = (-1)^i[0.2\cos(t)0.1\sin(t)0.2\cos(t)]^T$ and $\Delta_{\Theta i} = (-1)^i[0.8\sin(t)0.9\cos(t)0.8\sin(t)]^T$. In order to collect system data for learning the optimal control laws, each quadrotor uses a simple proportional-derivative controller for stable flight in the virtual environment. The persisting excitation noise is set as the sum of several sine signals. The initial states of the leaders and the followers are $p_{l1}(0) = [9.0\ \ -0.4\ \ -0.3]^T$ m, $p_{l2}(0) = [11.5\ \ 8.9\ \ -0.2]^T$ m, $p_{l3}(0) = [4.8\ \ 13.6\ \ 0.4]^T$ m, $p_{l4}(0) = [-2.9\ \ 8.2\ \ -0.2]^T$ m, $p_{l5}(0) = [0.1\ \ 0\ \ -0.1]^T$ m, $p_{l6}(0) = [4.8\ \ 6.3\ \ 9.0]^T$ m, $\dot{p}_{lv}(0) = 0_{3\times 1}$ $(v = 1, 2, \ldots, 6)$ m/s, $p_{f1}(0) = [5.0\ \ 3.0\ \ -0.2]^T$ m, $p_{f2}(0) = [-2.0\ \ -6.0\ \ 5.2]^T$ m, $p_{f3}(0) = [0.2\ \ 8.4\ \ 0.1]^T$ m, $p_{f4}(0) = [3.2\ \ 5.4\ \ 1.1]^T$ m, $p_{f5}(0) = [4.2\ \ 9.4\ \ -1.1]^T$ m, $p_{f6}(0) = [5.2\ \ 4.4\ \ 3.1]^T$ m, $\dot{p}_{fi}(0) = 0_{3\times 1}$ m/s, $\Theta_i = 0_{3\times 1}$, $\dot{\Theta}_i = 0_{3\times 1}$ $(i = 1,\ 2,\ \cdots\ 6)$. Then, the proposed RL-based methods are implemented to learn the optimal control laws using the collected data from the quadrotor system. The convergence of the weight of each NN is shown in Fig. 3.

The simulation results are shown in Figs. 4–8. The 3-D trajectories of 12 UAVs are drawn in Fig. 4. The blue solid lines represent the six leaders, and the other six solid lines in different colors represent the followers. The containment errors of the proposed observers are depicted in Fig. 5. The positions of the quadrotors are shown in Fig. 6. The position tracking errors and the attitude tracking errors of each quadrotor under disturbances are depicted in Figs. 7 and 8, respectively. One can observe from Fig. 5 that the absolute containment error of each observer is less than 0.1 within 0.5 s. Besides, one can see from Figs. 7 and 8 that the absolute position tracking errors and the attitude tracking errors are less than 0.1 within 2.5 s under disturbances. From these figures, it is clear that the quadrotors successfully fly into the pentagonal pyramid formed by the active leaders. Therefore, the proposed optimal

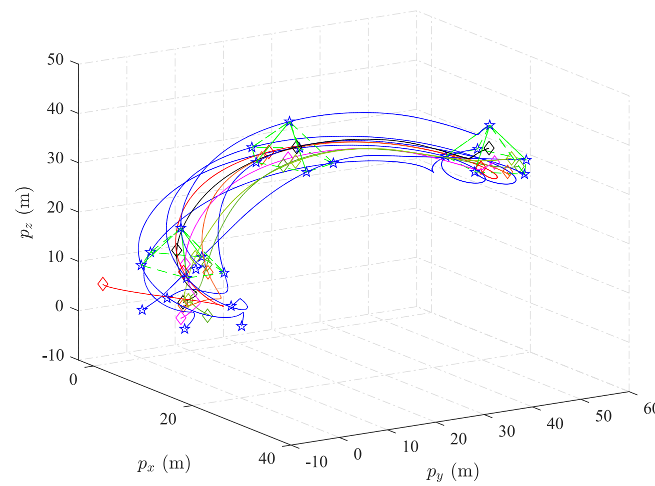

Fig. 4. Trajectories of the containment system.

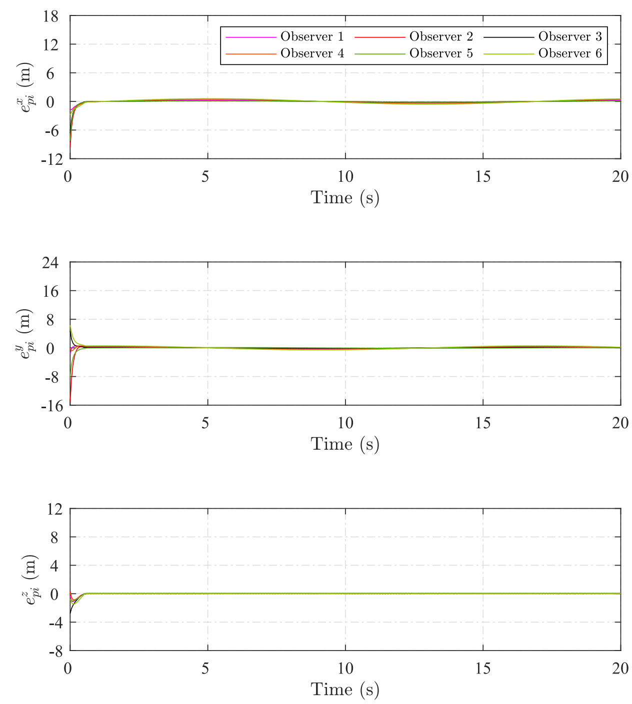

Fig. 5. Containment errors of the observers.

control laws can guarantee the achievement of containment flight for the quadrotors under external disturbances.

To demonstrate the advantages of the proposed method, a comparative simulation using the output feedback controller in [30] is conducted under identical initial conditions and the position tracking errors are portrayed in Fig. 9. It can be obtained from Fig. 9 that the output feedback controller can guarantee the achievement of the containment but has undesirable tracking performance with larger tracking errors compared with the proposed controller.

V. Conclusion

In this article, the optimal containment control problem is addressed for multiple quadrotors using the RL. The proposed controller can learn the optimal control laws from the system

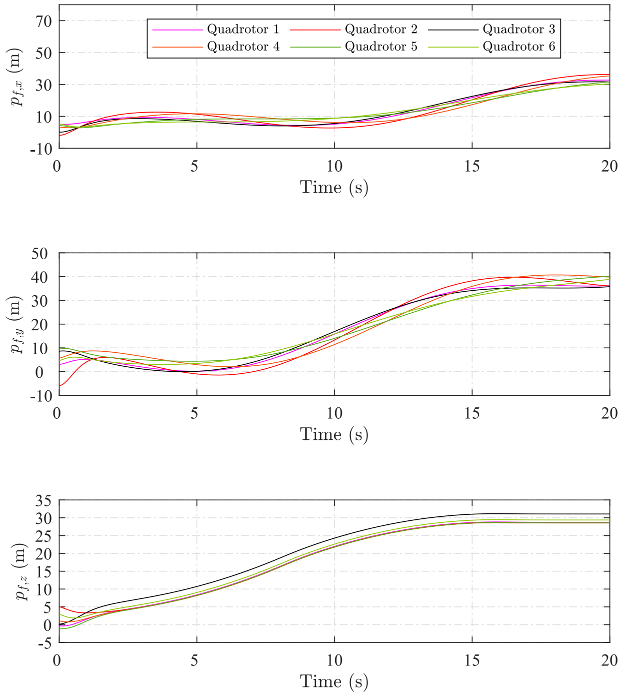

Fig. 6. Positions of the quadrotors.

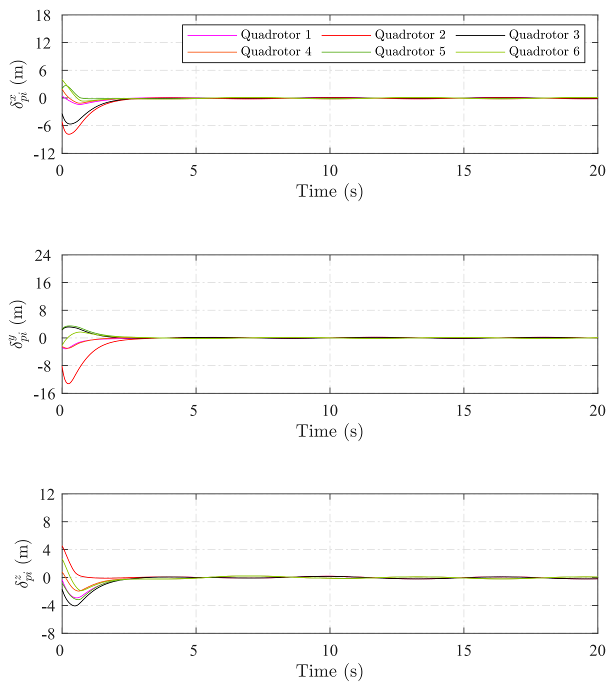

Fig. 7. Position tracking errors using the proposed controller.

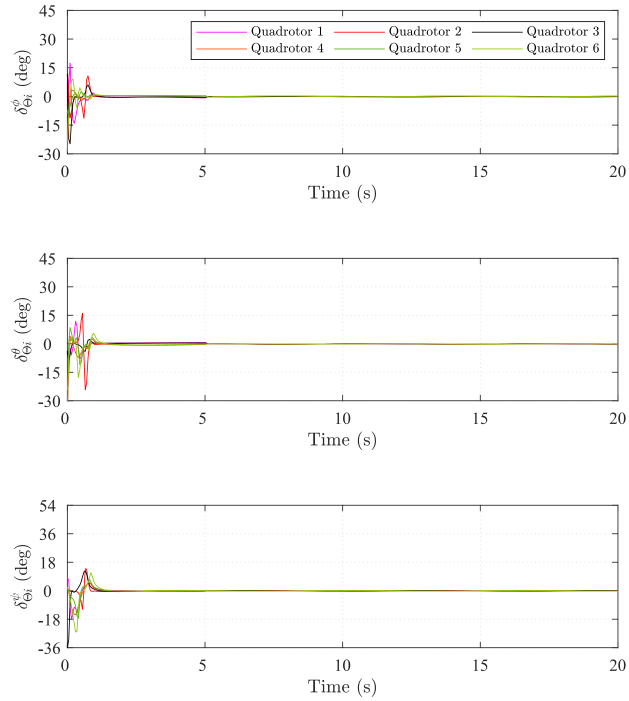

Fig. 8. Attitude tracking errors using the proposed controller.

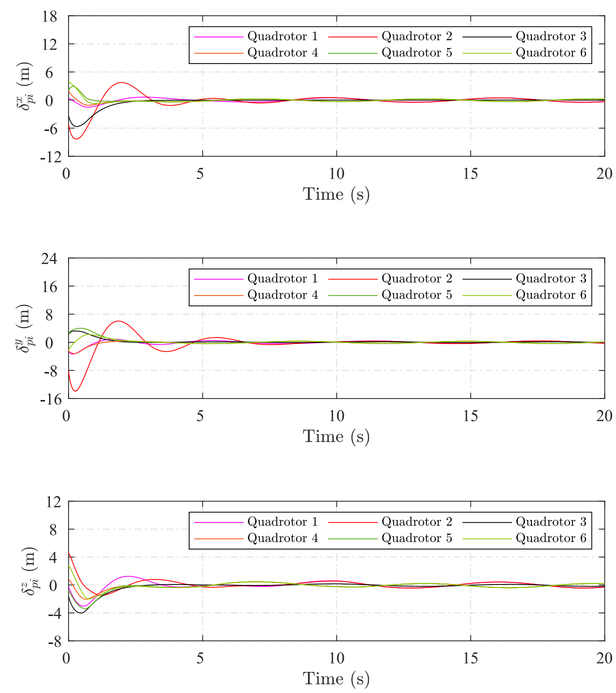

Fig. 9. Position tracking errors using the output feedback controller.

data of the quadrotor vehicles subject to uncertain dynamical parameters, external disturbances, and nonlinear dynamics. The active leaders with unknown control input are also considered to improve the maneuverability of the entire team. The theoretical analysis proves the stability of the closed-loop system, and the simulation results validate the effectiveness and advantages of the proposed method. In future, the cooperative control problems of heterogenous systems, such as air-ground vehicle system, will be further investigated.

References

[1] X. Liang, H. Yu, Z. Zhang, H. Liu, Y. Fang, and J. Han, "Unmanned aerial transportation system with flexible connection between the quadrotor and the payload: Modeling, controller design, and experimental validation," *IEEE Trans. Ind. Electron.*, vol. 70, no. 2, pp. 1870–1882, Feb. 2023.

[2] X. Liang, Z. Zhang, H. Yu, Y. Wang, Y. Fang, and J. Han, "Antiswing control for aerial transportation of the suspended cargo by dual quadrotor UAVs," *IEEE/ASME Trans. Mechatronics*, vol. 27, no. 6, pp. 5159–5172, Dec. 2022.

[3] V. P. Tran, M. A. Mabrok, S. G. Anavatti, M. A. Garratt, and I. R. Petersen, "Robust fuzzy Q-learning-based strictly negative imaginary tracking controllers for the uncertain quadrotor systems," *IEEE Trans. Cybern.*, early access, Jun. 6, 2022, doi: 10.1109/TCYB.2022.3175366.

[4] W. Zhao, H. Liu, F. L. Lewis, and X. Wang, "Data-driven optimal formation control for quadrotor team with unknown dynamics," *IEEE Trans. Cybern.*, vol. 52, no. 8, pp. 7889–7898, Aug. 2022.

[5] S. C. Yogi, L. Behera, and S. Nahavandi, "Adaptive intelligent minimum parameter singularity free sliding mode controller design for quadrotor," *IEEE Trans. Autom. Sci. Eng.*, early access, Feb. 14, 2023, doi: 10.1109/TASE.2023.3243660.

[6] J. Hu, P. Bhowmick, I. Jang, F. Arvin, and A. Lanzon, "A decentralized cluster formation containment framework for multirobot systems," *IEEE Trans. Robot.*, vol. 37, no. 6, pp. 1936–1955, Dec. 2021.

[7] S. Ramazani, P. R. Selmic, and M. de Queiroz, "Rigidity-based multiagent layered formation control," *IEEE Trans. Cybern.*, vol. 47, no. 8, pp. 1902–1913, Aug. 2017.

[8] T. Li, W. Bai, Q. Liu, Y. Long, and C. L. P. Chen, "Distributed fault-tolerant containment control protocols for the discrete-time multiagent systems via reinforcement learning method," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Nov. 1, 2021, doi: 10.1109/TNNLS.2021.3121403.

[9] J. Zhang, F. Yan, T. Feng, T. Deng, and Y. Zhao, "Fastest containment control of discrete-time multi-agent systems using static linear feedback protocol," *Inf. Sci.*, vol. 614, pp. 362–373, Oct. 2022.

[10] Y. Lu, X. Dong, Q. Li, J. Lü, and Z. Ren, "Time-varying group formation-containment tracking control for general linear multiagent systems with unknown inputs," *IEEE Trans. Cybern.*, vol. 52, no. 10, pp. 11055–11067, Oct. 2022.

[11] B. Mu and Y. Shi, "Distributed LQR consensus control for heterogeneous multiagent systems: Theory and experiments," *IEEE/ASME Trans. Mechatronics*, vol. 23, no. 1, pp. 434–443, Feb. 2018.

[12] Q. Sun, X. Wang, and Y.-H. Chen, "Satellite formation-containment control emphasis on collision avoidance and uncertainty suppression," *IEEE Trans. Cybern.*, early access, Jun. 6, 2022, doi: 10.1109/TCYB.2022.3173683.

[13] D. Liu, H. Liu, K. Liu, H. Gu, and J. Lü, "Robust hierarchical pinning control for nonlinear heterogeneous multiagent system with uncertainties and disturbances," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 69, no. 12, pp. 5273–5285, Dec. 2022.

[14] X. Chen and L. Zhao, "Observer-based finite-time attitude containment control of multiple spacecraft systems," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 68, no. 4, pp. 1273–1277, Apr. 2021.

[15] Y. Yang, J. Tan, D. Yue, X. Xie, and W. Yue, "Observer-based containment control for a class of nonlinear multiagent systems with uncertainties," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 51, no. 1, pp. 588–600, Jan. 2021.

[16] I. A. Zamfirache, R.-E. Precup, R.-C. Roman, and E. M. Petriu, "Policy iteration reinforcement learning-based control using a grey wolf optimizer algorithm," *Inf. Sci.*, vol. 585, pp. 162–175, May 2022.

[17] T. Chen, A. Babanin, A. Muhammad, B. Chapron, and C. Chen, "Modified evolved bat algorithm of fuzzy optimal control for complex nonlinear systems," *Romanian J. Inf. Sci. Technol.*, vol. 23, pp. T28–T40, Dec. 2020.

[18] H. Su and L. Zhang, "Model-independent containment control for dynamic multiple Euler–Lagrange systems with disturbances and uncertainties," *IEEE Trans. Netw. Sci. Eng.*, vol. 8, no. 4, pp. 3443–3452, Oct.–Dec. 2021.

[19] H. Zhang, Y. Zhou, Y. Liu, and J. Sun, "Cooperative bipartite containment control for multiagent systems based on adaptive distributed observer," *IEEE Trans. Cybern.*, vol. 52, no. 6, pp. 5432–5440, Jun. 2022.

[20] W. Zhao, H. Liu, and F. L. Lewis, "Data-driven fault-tolerant control for attitude synchronization of nonlinear quadrotors," *IEEE Trans. Autom. Control*, vol. 66, no. 11, pp. 5584–5591, Nov. 2021.

[21] Y. Lu, X. Dong, Q. Li, J. Lv, and Z. Ren, "Time-varying group formation-tracking control for general linear multi-agent systems with switching topologies and unknown input," *Int. J. Robust Nonlinear Control*, vol. 32, no. 4, pp. 1925–1940, Mar. 2022.

[22] P. Zhou and B. M. Chen, "Formation-containment control of Euler–Lagrange systems of leaders with bounded unknown inputs," *IEEE Trans. Cybern.*, vol. 52, no. 7, pp. 6342–6353, Jul. 2022.

[23] H. Liu, T. Ma, F. L. Lewis, and Y. Wan, "Robust formation control for multiple quadrotors with nonlinearities and disturbances," *IEEE Trans. Cybern.*, vol. 50, no. 4, pp. 1362–1371, Apr. 2020.

[24] T. Han, M. Chi, Z.-H. Guan, B. Hu, J.-W. Xiao, and Y. Huang, "Distributed three-dimensional formation containment control of multiple unmanned aerial vehicle systems," *Asian J. Control*, vol. 19, no. 3, pp. 1103–1113, May 2017.

[25] Z. Li, Z. Duan, W. Ren, and G. Feng, "Containment control of linear multi-agent systems with multiple leaders of bounded inputs using distributed continuous controllers," *Int. J. Robust Nonlinear Control*, vol. 25, no. 13, pp. 2101–2121, Sep. 2015.

[26] Y. Yang, H. Modares, D. C. Wunsch, and Y. Yin, "Optimal containment control of unknown heterogeneous systems with active leaders," *IEEE Trans. Control Syst. Technol.*, vol. 27, no. 3, pp. 1228–1236, May 2019.

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

[28] Z. Li, Z. Duan, G. Chen, and L. Huang, "Consensus of multiagent systems and synchronization of complex networks: A unified viewpoint," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 57, no. 1, pp. 213–224, Jan. 2010.

[29] H. Modares, F. L. Lewis, and Z.-P. Jiang, "H_∞ tracking control of completely unknown continuous-time systems via off-policy reinforcement learning," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 26, no. 10, pp. 2550–2562, Oct. 2015.

[30] D. Li, W. Zhang, W. He, C. Li, and S. S. Ge, "Two-layer distributed formation-containment control of multiple Euler–Lagrange systems by output feedback," *IEEE Trans. Cybern.*, vol. 49, no. 2, pp. 675–687, Feb. 2019.

Ming Cheng received the B.S. degree in spacecraft design and engineering from Beihang University, Beijing, China, in 2019, where he is currently pursuing the Ph.D. degree with the School of Astronautics.

His research interest includes robust control, nonlinear control, reinforcement learning, multiagent system, and quadrotor flight control.

Hao Liu (Senior Member, IEEE) received the B.E. degree in control science and engineering from Northwestern Polytechnical University, Xi'an, China, in 2008, and the Ph.D. degree in automatic control from Tsinghua University, Beijing, China, in 2013.

In 2012, he was a visiting student with the Research School of Engineering, Australian National University, Canberra, ACT, Australia. Since 2013, he has been with the School of Astronautics, Beihang University, Beijing, where he is currently an Associate Professor. From 2017 to 2018, he was a Visiting Scholar with the University of Texas at Arlington Research Institute, Fort Worth, TX, USA. Since 2020, he has been with the Institute of Artificial Intelligence, Beihang University. Since 2022, he has also been with Zhongguancun Laboratory, Beijing. His research interests include formation control, reinforcement learning, robust control, nonlinear control, unmanned aerial vehicles, unmanned underwater vehicles, and multiagent systems.

Dr. Liu received the Best Paper Award on IEEE ICCA 2018. He serves as an Associate Editor for *International Jounral of Robust and Nonlinear Control*, *Journal of Intelligent and Robotic Systems*, and *Transactions of the Institute of Measurement and Control*.

Qing Gao (Senior Member, IEEE) received the B.Eng. and Ph.D. degrees in mechanical and electrical engineering from the University of Science and Technology of China, Hefei, China, in 2008 and 2013, respectively, and the Ph.D. degree in mechatronics engineering from the City University of Hong Kong, Hong Kong, in 2014.

From 2014 to 2016, he was with the School of Engineering and Information Technology, University of New South Wales Canberra at the Australian Defence Force Academy, Canberra, ACT, Australia, as a Postdoctoral Research Associate. In 2018, he has joined the School of Automation Science and Electrical Engineering, Beihang University, Beijing, China, as a Full Professor. His research interests include intelligent systems and control theory.

Prof. Gao is the recipient of the Alexander von Humboldt Fellowship of Germany and the 21st Guan Zhao-Zhi Award at The 34th Chinese Control Conference.

Jinhu Lü (Fellow, IEEE) received the Ph.D. degree in applied mathematics from the Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, China, in 2002.

He was a Professor with RMIT University, Melbourne, VIC, Australia, and a Visiting Fellow with Princeton University, Princeton, NJ, USA. He is currently the Dean of the School of Automation Science and Electrical Engineering, Beihang University, Beijing. He is also the Chief Scientist of the National Key Research and Development Program of China and a Leading Scientist of Innovative Research Groups of the National Natural Science Foundation of China and the Ten Thousand Talents Program of China. His current research interests include nonlinear circuits and systems, complex networks, multiagent systems, and big data.

Prof. Lü was the recipient of the State Natural Science Award thrice from the Chinese Government in 2008, 2012, and 2016, respectively, the Australian Research Council Future Fellowships Award in 2009, the National Natural Science Fund for Distinguished Young Scholars, the Prestigious Ho Leung Ho Lee Foundation Award in 2015, and the Highly Cited Researcher Award in engineering from 2014 to 2018. He was an Editor in various ranks for 15 SCI journals. He was a member of the Evaluating Committees of the IEEE Circuits and Systems Society, the IEEE Industrial Electronics Society, and the IEEE Computational Intelligence Society. He was the General Co-Chair of the 43rd Annual Conference of the IEEE Industrial Electronics Society in 2017.

Xiaohua Xia (Fellow, IEEE) received the B.Sc. degree in applied mathematics from the Wuhan Institute of Hydraulic and Electrical Engineering, Wuhan, China, in 1983, and the Ph.D. degree in automatic control theory and application from the Beijing University of Aeronautics and Astronautics, Beijing, China, in 1989.

He is currently a Professor with the Electrical, Electronic, and Computer Engineering Department, University of Pretoria, Pretoria, South Africa, the Director of the Center of New Energy Systems, and the Director of the National Hub for the Postgraduate Programme in Energy Efficiency and Demand-Side Management. He was academically affiliated with the University of Stuttgart, Stuttgart, Germany, the Ecole Centrale de Nantes, Nantes, France, and the National University of Singapore, Singapore before joining the University of Pretoria in 1998. His current research interests are industrial energy systems and building energy systems.

Prof. Xia has been an Associate Editor of *Automatica*, IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS II: EXPRESS BRIEFS, and IEEE TRANSACTIONS ON AUTOMATIC CONTROL. He is an NRF A-rated scientist. He was elected as a Fellow of the South African Academy of Engineering in 2005, and a member of the Academy of Science of South Africa in 2011.