Preprints of the 21st IFAC World Congress (Virtual)
Berlin, Germany, July 12-17, 2020

Fault-Tolerant Control for the Formation of Multiple Unknown Nonlinear Quadrotors via Reinforcement Learning ⋆

Wanbing Zhao * **Hao Liu** ** **Frank L. Lewis** ***

* *Key Laboratory of Spacecraft Design Optimization and Dynamic Simulation Technologies of Ministry of Education, Beihang University, Beijing 100191, P.R. China (e-mail: zhaowb27@buaa.edu.cn).*
** *Key Laboratory of Spacecraft Design Optimization and Dynamic Simulation Technologies of Ministry of Education, Beihang University, Beijing 100191, P.R. China (e-mail: liuhao13@buaa.edu.cn)*
*** *University of Texas at Arlington Research Institute, University of Texas at Arlington, Fort Worth, Texas 76118, USA (e-mail: lewis@uta.edu)*

Abstract: In this paper, the fault-tolerant control problem for the formation of unknown quadrotor team with nonlinearities, couplings, and actuator faults in the dynamics is investigated. A distributed observer is designed to estimate the position references for each quadrotor. A hierarchical control scheme is constructed including a fault-tolerant position controller to achieve the desired formation and a fault-tolerant attitude controller to track the attitude references. Reinforcement learning algorithms are designed to learn the optimal control policies of the position and attitude controllers. Simulation results are given to illustrate the effectiveness of the proposed controller.

Keywords: Fault-tolerant control, formation control, reinforcement learning, model-free, quadrotor system.

1. INTRODUCTION

As a typical class of UAVs, quadrotors have received much research interest because of their capabilities of hovering, slow flight, and vertical take-off and landing. Furthermore, the cooperation of the multi-agent quadrotor system exhibits special superiority in the complicated task such as telecommunication relay, cooperative load transportation, and wilderness search (see, Dong et al. (2018)). The quadrotors are composed of a rigid frame with four individual rotors, which generate the control torques for the quadrotor by changing the spinning speeds of the four rotors. However, the actuating rotors are prone to faults because of the degradation of the motors or the damage of the propellers (see,*e.g.*, Avram, Zhang, and Muse. (2017)). For the multi-agent quadrotor system, the probability of the system encountering actuator faults is growing due to the increasing team scale and the complexity of the multi-agent system (see, Qin et al. (2017)). The appearance of the actuator faults may result in undesirable trajectory tracking performance, instability of the multi-agent quadrotor system, and even catastrophic consequence. Therefore, it is necessary to address the fault-tolerant control (FTC) problem for the multi-agent quadrotor system under actuator faults.

Recently, the FTC methods have been implemented to the cooperative control of the multi-agent systems. In Deng and Yang. (2017), the distributed adaptive FTC problem was studied for the synchronization of *linear* multi-agent systems under the effects of actuator faults. In Hua et al. (2016), the distributed FTC problem was investigated to achieve the formation for *second-order linear* multi-agent systems in the presence of multiple actuator faults. In Yu, Qu, and Zhang. (2019), a distributed fault-tolerant controller was studied to achieve the desired synchronization for multiple cooperative UAVs with *nonlinearities* and actuator faults in the dynamics. In Shi et al. (2017), a fault-tolerant formation controller involving a decentralized state observer and an adaptive fault estimator was developed for nonlinear quadrotor team to drive each quadrotor to the desired formation pattern. However, in the fault-tolerant controller design process of Deng and Yang. (2017); Hua et al. (2016); Yu, Qu, and Zhang. (2019); Shi et al. (2017), the dynamics of the system was required, which is unpractical in the controller design of quadrotor team. In fact, the quadrotor is a nonlinear system subject to actuator faults and the dynamics of the quadrotor system is difficult to obtain in practical applications. Therefore, it is promising to design a fault-tolerant controller for multiple unknown quadrotors subject to nonlinearities and actuator faults in the quadrotor dynamics.

In the latest years, Reinforcement learning (RL) has become a powerful approach to learn the optimal control

⋆ This work was supported by the National Natural Science Foundation of China under Grants 61873012, 61503012, and 61633007, and the Office of Naval Research under Grant N00014-17-1-2239.

Copyright lies with the authors

policy and emerged in the fault-tolerant controller design for single-agent systems (see, *e.g.*, Liu, Wang, and Wang. (2017); Wang et al. (2016); Zhang et al. (2018); Deptula et al. (2018); Ma, Xu, and Yang. (2019)). In Liu, Wang, and Wang. (2017); Wang et al. (2016), the RL-based fault-tolerant controllers were developed to tolerant actuator faults for nonlinear multiple-input multiple-output (MIMO) discrete-time systems. In Zhang et al. (2018), a RL-based FTC algorithm was designed by using online policy iteration approach to learn the optimal controller for nonlinear tracking systems. In the RL-based optimal controller design of Deptula et al. (2018), a data-based estimator was constructed to estimate the loss of effectiveness (LOE) actuator faults and then an approximate dynamic programming approach was implemented to learn the optimal control solutions for the system. In Ma, Xu, and Yang. (2019), a series of RL algorithms were implemented in the data-driven FTC design for a *single* quadrotor to estimate the actuator faults and then compensate the actuator faults. However, the FTC problem for the formation of multiple nonlinear unknown quadrotor systems subject to actuator faults has not been discussed in the literatures mentioned before.

Therefore, the *fault-tolerant cooperative* control problem for multiple *unknown quadrotor systems* with *nonlinearities* and *actuator faults* in the dynamics remains an open issue. In this paper, a distributed fault-tolerant cooperative controller is designed including a *distributed* observer, a fault-tolerant position controller, and a fault-tolerant attitude controller. The distributed observer is constructed to estimate the desired position for each quadrotor using the information of its neighbors and itself. Then, a position controller and an attitude controller are constructed. The optimal control policies for the position and attitude controller are learned by using RL approach. Then, the fault-tolerant controllers are designed for the quadrotor position subsystem and the quadrotor attitude subsystem to *compensate* the actuator faults.

The rest parts of this paper are given as follows. Section 2 gives the preliminaries on the graph theory, the quadrotor model, the actuator fault model, and the problem formulation. The fault-tolerant formation controller is designed in Section 3. In Section 4, the simulation results of six quadrotor are shown and the concluding remarks are given in Section 5.

Notations: Define I_N as a $N \times N$ unit matrix, $\lambda_i(M)$ the ith eigenvalue of the matrix M, $0_{k\times l}$ a $k \times l$ zero matrix, and $c_{a,b}$ an $a \times 1$ vector with 1 in the bth element and 0s elsewhere.

2. PRELIMINARIES

2.1 Graph Theory

Consider a team of N quadrotors. Let $\Phi = \{1, 2, \cdots, N\}$. The communication topology among the quadrotors is described with a time invariant graph $G = (V, E, W)$, where $V = \{v_i\}$ $(i \in \Phi)$ is the node set, $E \subset V \times V$ the edge set, and $W = [w_{ij}] \in \mathbb{R}^{N\times N}$ the weighted adjacency matrix. For node v_i and node v_j $(i, j \in \Phi)$, the weight w_{ij} satisfies that $w_{ij} > 0$ if and only if $(v_i, v_j) \in E$, otherwise $w_{ij} = 0$. Denote the neighbors of v_i as $N_i = \{v_j \,|(v_i, v_j) \in E\}$. Let d_i be the in-degree of v_i with $d_i = \sum_{j=1}^{N} w_{ij}$. Define $L = D - W$, where $D = diag(d_i) \in \mathbb{R}^{N\times N}$. The path from v_i to v_j is a sequence of ordered edges with $\{(v_i, v_k), (v_k, v_n), \cdots, (v_l, v_j)\}$. The graph G contains a spanning tree if there is at least one node connected to all the other nodes.

2.2 Quadrotor Model

Define $E^I = \{e^I_x, e^I_y, e^I_z\}$ as the inertial frame and $E^B = \{e^B_x, e^B_y, e^B_z\}$ the body-fixed frame. Let $p_i = [p_{xi}\ p_{yi}\ p_{zi}]^T \in \mathbb{R}^{3\times1}$ be the position of the ith quadrotor and $\Theta_i = [\phi_i\ \theta_i\ \psi_i]^T \in \mathbb{R}^{3\times1}$ the Euler angle of the ith quadrotor. From Liu et al. (2019), one can obtain the quadrotor dynamic model as

$$\begin{aligned} m_i\ddot{p}_i &= R_{fi}F_i, \\ J_{\Theta i}\ddot{\Theta}_i &= -C\left(\Theta_i, \dot{\Theta}_i\right)\dot{\Theta}_i + \tau_i, \end{aligned} \tag{1}$$

where m_i is the mass of the ith quadrotor, $J_{\Theta i} = diag(J_{\phi i}, J_{\theta i}, J_{\psi i}) \in \mathbb{R}^{3\times3}$ the inertial matrix, $C(\Theta_i, \dot{\Theta}_i) \in \mathbb{R}^{3\times3}$ the nonlinear Coriolis term described in Raffo, Ortega, and Rubio. (2010), and $R_{fi} \in \mathbb{R}^{3\times3}$ the orientation matrix from E^B to E^I shown in Liu et al. (2019). $F_i \in \mathbb{R}^{3\times1}$ and $\tau_i \in \mathbb{R}^{3\times1}$ are the external force and the torque generated by the four rotors with

$$F_i = c_{3,3}k_{wi}\sum\nolimits_{k=1}^{4}\omega^2_{i,k} - R^T_{fi}c_{3,3}m_i g \tag{2}$$

and

$$\tau_i=\left[k_{ti}k_{wi}(\omega^2_{i,1}-\omega^2_{i,3})\ \ k_{ti}k_{wi}(\omega^2_{i,2}-\omega^2_{i,4})\ \ l_{\tau i}\sum\nolimits_{k=1}^{4}(-1)^{k+1}\omega^2_{i,k}\right]^T, \tag{3}$$

respectively, where g indicates the gravity constant, $\omega_{i,j}$ $(j = 1, 2, 3, 4)$ is the rotational speed of the jth rotor of the ith quadrotor, and k_{ti}, k_{wi}, $l_{\tau i}$ are positive scaling factors for the ith quadrotor. Let $\bar{\omega}^2_i = \left[\omega^2_{i,1}\ \omega^2_{i,2}\ \omega^2_{i,3}\ \omega^2_{i,4}\right]^T$. Due to the existence of a power distribution board, the control inputs can be distributed as:

$$\begin{bmatrix} u_{\phi i} \\ u_{\theta i} \\ u_{\psi i} \\ u_{zi} \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & -1 \\ 1 & 0 & -1 & 0 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} \omega^2_{i,1} \\ \omega^2_{i,2} \\ \omega^2_{i,3} \\ \omega^2_{i,4} \end{bmatrix} = G_m\bar{\omega}^2_i. \tag{4}$$

Define $u_{pi} = [u_{pxi}\ u_{pyi}\ u_{pzi}]^T \in \mathbb{R}^{3\times1}$ as the virtual position control input with

$$u_{pi} = R_{fi}c_{3,3}u_{zi}. \tag{5}$$

Then, one can rewrite the quadrotor model (1) as

$$\begin{aligned} \ddot{p}_i &= b_{pi}u_{pi} - c_{3,3}g, \\ \ddot{\Theta}_i &= -J^{-1}_{\Theta i}\left(C\left(\Theta_i, \dot{\Theta}_i\right)\dot{\Theta}_i\right) + b_{\tau i}u_{\tau i}, \end{aligned} \tag{6}$$

where $u_{\tau i} = [u_{\phi i}\ u_{\theta i}\ u_{\psi i}]^T$, $b_{pi} = m_i^{-1}k_{wi}I_{3\times3}$, and $b_{\tau i} = J^{-1}_{\Theta i}diag\{k_{ti}k_{wi}, k_{ti}k_{wi}, l_{\tau i}\}$.

2.3 Actuator Fault Model

Similarly to Avram, Zhang, and Muse. (2018), the actuator fault of the ith quadrotor in the jth rotor is described as $\left(\omega_{i,j}^*\right)^2 = \rho_{ij}\omega_{i,j}^2, i \in \Phi, j = 1,2,3,4$, where $\omega_{i,j}^*$ indicates the actual spinning speed and ρ_{ij} an unknown positive LOE gain with $0 < \rho_{ij} \leq 1$. In this paper, the actuator faults of the rotors are described with a diagonal matrix $K_{\rho i} \in \mathbb{R}^{4\times4}$ satisfying that $K_{\rho i} = diag\{\rho_{i1}, \rho_{i2}, \rho_{i3}, \rho_{i4}\}$. Then, the quadrotor model with actuator faults can be written as

$$\begin{aligned}\ddot{p}_i &= b_{pi}\left(u_{pi} - \mu(t-t_s)u_{\Delta pi}\right) - c_{3,3}g,\\ \ddot{\Theta}_i &= -J_{\Theta i}^{-1}\left(C\left(\Theta_i, \dot{\Theta}_i\right)\dot{\Theta}_i\right) + b_{\tau i}\left(u_{\tau i} - \mu(t-t_s)u_{\Delta\Theta i}\right),\end{aligned} \tag{7}$$

where $\mu(t-t_s)$ is a step function with an unknown fault occurrence time t_s satisfying that $\mu(t-t_s) = 0$ if $t < t_s$ and otherwise, $\mu(t-t_s) = 1$. $u_{\Delta pi} \in \mathbb{R}^{3\times1}$ and $u_{\Delta\Theta i} \in \mathbb{R}^{3\times1}$ in (7) are the input uncertainties resulted from the actuator faults of the rotors. From (3), (4), (5), (6), $u_{\Delta pi}$ and $u_{\Delta\Theta i}$ can be derived as

$$\begin{aligned}u_{\Delta pi} &= R_{fi}c_{3,3}u_{\Delta zi},\\ u_{\Delta zi} &= [1-\rho_{i1}\ \ 1-\rho_{i2}\ \ 1-\rho_{i3}\ \ 1-\rho_{i4}]\,\bar{\omega}_i^2,\\ u_{\Delta\Theta i} &= [I_3\ \ 0_{3\times1}]\,G_m\left(I_4 - K_{\rho i}\right)\bar{\omega}_i^2.\end{aligned} \tag{8}$$

In the practical applications, the uncertainties $u_{\Delta pi}$ and $u_{\Delta\Theta i}$ are time-varying and bounded with $\|u_{\Delta pi}\| \leq \bar{u}_{\Delta pi}$ and $\|u_{\Delta\Theta i}\| \leq \bar{u}_{\Delta\Theta i}$. The bounds $\bar{u}_{\Delta pi}$ and $\bar{u}_{\Delta\Theta i}$ are not directly available in practical applications and are estimated by designing fault estimators, which will be given in the following section.

2.4 Problem Formulation

The objective of this paper is to design a fault tolerant formation controller for unknown quadrotor team with nonlinear dynamics and actuator faults. Define $p_0 = [p_{x0}\ p_{y0}\ p_{z0}]^T \in \mathbb{R}^{3\times1}$ as the position of the virtual leader in E^I for the quadrotor team to track and $\delta_{ij} = [\delta_{x,ij}\ \delta_{y,ij}\ \delta_{z,ij}]^T \in \mathbb{R}^{3\times1}$ $(i, j \in \Phi)$ the desired position deviation between the ith quadrotor and the jth quadrotor. Let $\delta_{ij} = \delta_i - \delta_j$, where $\delta_i = [\delta_{xi}\ \delta_{yi}\ \delta_{zi}]^T \in \mathbb{R}^{3\times1}$ $(i \in \Phi)$ is the desired position deviation between the ith quadrotor and the virtual leader. In this paper, the ith quadrotor can only have access to the information of its neighbors and itself. The position deviation δ_{ij} for the ith quadrotor is predesigned according to the desired formation pattern. δ_i is not available if the ith quadrotor is not connected to the virtual leader. Let $p_{ri} = [p_{xri}\ p_{yri}\ p_{zri}]^T \in \mathbb{R}^{3\times1}$ be the position reference with $p_{ri} = p_{r0} + \delta_i$ and $e_{pi} = [e_{xi}\ e_{yi}\ e_{zi}]^T \in \mathbb{R}^{3\times1}$ the position tracking error for the ith quadrotor with $e_{pi} = p_i - p_{ri}$. A distributed observer is designed in the following section to estimate $\hat{p}_{ri}$ for the ith quadrotor using the available information of its neighbors and itself.

3. CONTROL PROTOCOL DESIGN

3.1 Distributed Observer Design

Let $x_{p0} = [p_0^T\ \dot{p}_0^T]^T \in \mathbb{R}^{6\times1}$ be the state of the leader. The dynamics of the virtual leader is described as $\dot{x}_{p0} = A_{p0}x_{p0}$, where $A_{p0} \in \mathbb{R}^{6\times6}$. Denote $\varsigma_{pi} = [\hat{p}_{ri}\ \dot{\hat{p}}_{ri}] \in \mathbb{R}^{6\times1}$ as the state of the ith observer for the ith quadrotor. Let $\bar{\delta}_{ij} = [\delta_{ij}^T\ \ 0_{1\times3}]^T \in \mathbb{R}^{6\times1}$ and $\bar{\delta}_i = [\delta_i^T\ \ 0_{1\times3}]^T \in \mathbb{R}^{6\times1}$. The distributed observer is designed as

$$\dot{\varsigma}_{pi} = A_{p0}\varsigma_{pi} + \gamma_i\sum_{j\in N_i}\left(w_{ij}\left(\varsigma_{pj} - \varsigma_{pi} + \bar{\delta}_{ij}\right) + \rho_i\left(x_{p0} + \bar{\delta}_i - \varsigma_{pi}\right)\right), \tag{9}$$

where γ_i is a positive constant to be determined and ρ_i a constant with $\rho_i > 0$ indicating that the ith quadrotor has access to the virtual leader and otherwise $\rho_i = 0$. Define $\tilde{x}_{pi} \in \mathbb{R}^{6\times1}$ as the estimation error of the ith observer with $\tilde{x}_{pi} = \varsigma_{pi} - \bar{\delta}_i - x_{p0}$. Assume graph G has a spanning tree. Then, $\tilde{x}_{pi}$ will converge to 0 on condition that γ_i is sufficiently large.

3.2 Position Controller Design

By using the estimated leader state and the quadrotor state, from (7), an augmented system can be constructed as

$$\dot{X}_{pi} = \bar{A}_{pi}X_{pi} + \bar{B}_{pi}\left(u_{pi} - \mu(t-t_s)u_{\Delta pi}\right) - c_{12,6}g + D_{pi}\varepsilon_{pi}, \tag{10}$$

where $X_{pi} = [x_{pi}\ \varsigma_i]^T \in \mathbb{R}^{12\times1}$, $\bar{A}_{pi} = diag(A_{pi}, A_{p0}) \in \mathbb{R}^{12\times12}$, $\bar{B}_{pi} = [B_{pi}^T\ 0_{3\times6}]^T \in \mathbb{R}^{12\times3}$, $B_{pi} = [0_{3\times3}\ b_{pi}^T]^T \in \mathbb{R}^{6\times3}$, $\varepsilon_{pi} = \sum_{j\in N_i}\left(w_{ij}\left(\varsigma_{pj} - \varsigma_{pi} + \bar{\delta}_{ij}\right) + \rho_i\left(x_{p0} + \bar{\delta}_i - \varsigma_{pi}\right)\right)$, and $\bar{D}_{pi} = [0_{6\times6}\ \gamma_i I_6]^T \in \mathbb{R}^{12\times6}$. ε_{pi} $(i \in \Phi)$ will converge to 0 *i.e.*, $\hat{p}_{ri}$ will converge to p_{ri}, by using the proposed distributed observer. The position tracking error e_{pi} satisfies $e_{pi} = [C_p\ -C_p]X_{pi}$ and $C_p = [c_{6,1}\ c_{6,2}\ c_{6,3}]^T$. Define the performance function for the augmented system as

$$V_{pi}\left(e_{pi}, u_{pi}\right) = \int_t^\infty e^{-\alpha_{1i}(\tau-t)}\left(e_{pi}^T Q_{pi}e_{pi} + u_{pi}^T R_{pi}u_{pi}\right)d\tau, \tag{11}$$

where $Q_{pi} > 0$, $R_{pi} > 0$, and α_{1i} is a positive constant to guarantee that V_{pi} is bounded for any given $u_{\tau i}$ (see, Modares and Lewis. (2014)). From the optimal control theory (see, Lewis and Syrmos. (1995)), one can obtain the optimal position control policy u_{pi}^* as

$$u_{pi}^* = -\frac{1}{2}R_{pi}^{-1}\bar{B}_{pi}^T\Delta V_{pi}^*, \tag{12}$$

where $\Delta V_{pi}^* = \partial V_{pi}^*/\partial X_{pi}$ and the performance function V_{pi}^* is the solution to the following equation:

$$\begin{aligned}e_{pi}^T Q_{pi}\,e_{pi} - \frac{1}{4}(\Delta V_{pi}^*)^T\bar{B}_{pi}R_{pi}^{-1}\bar{B}_{pi}^T\Delta V_{pi}^* - \alpha_{1i}V_{pi}\\ + (\Delta V_{pi}^*)^T\left(\bar{A}_{pi}X_{pi} - c_{12,6}g + D_{pi}\varepsilon_{pi}\right) = 0.\end{aligned} \tag{13}$$

In order to achieve the desired formation for quadrotor with the uncertainty $u_{\Delta pi}$, the position controller in this paper is designed as

$$u_{pi}(t) = \frac{1}{2}u_{pi}^*(x) + \frac{\hat{\bar{u}}_{\Delta pi}^2 R_{pi}u_{pi}^*(x)}{\left\|R_{pi}u_{pi}^*(x)\right\|\hat{\bar{u}}_{\Delta pi} + \sigma_1(t)}, \tag{14}$$

where $\hat{\bar{u}}_{\Delta pi}$ is the estimated value of $\bar{u}_{\Delta pi}$ and $\sigma_1(t)$ a positive function with $\int_t^\infty \sigma_1(t)d\tau \leq \bar{\sigma}_1 < \infty$. The fault estimator is designed to estimate the bound $\bar{u}_{\Delta pi}$ with

Algorithm 3.2: RL algorithm for position controller

Initiation: Apply a stabilizing control input u_{pi} with bounded u_{pei} to the quadrotor system and collect the data. Determine the stopping criterion $\vartheta_{l1} > 0$.
Step 1: Solve the equation (16) for V_i^n and $u_{\tau i}^{n+1}$, simultaneously.
Step 2: Stop if: $\left|u_{\tau i}^{n+1} - u_{\tau i}^{n}\right| < \vartheta_{l1}$, otherwise, let $n = n+1$ and go to Step 1.

$$d\hat{\bar{u}}_{\Delta pi}/dt = -k_{\sigma 1}\sigma_1(t)\hat{\bar{u}}_{\Delta pi} + 2k_{\sigma 1}\left\|R_{pi}u_{pi}^*(x)\right\|, \quad (15)$$

where $k_{\sigma 1}$ is a positive constant. It can be seen from (14) that the proposed fault-tolerant controller rely on the optimal control policy u_{pi}^*. But the optimal control approach (12), (13) requires the accurate dynamic information of the quadrotor team, which is unpractical for quadrotor team in the formation flight applications. In the following, a RL algorithm is designed to obtain u_{pi}^* without knowledge of the quadrotor dynamic information.

For a given bounded signal $u_{pei} \in \mathbb{R}^{3\times 1}$, one can obtain from (12) that

$$\begin{aligned} &e^{-\alpha_{1i}\Delta t}V_{pi}^n\left(X_{pi}\left(t+\Delta t\right)\right) - V_{pi}^n\left(X_{pi}\left(t\right)\right) = \\ &\int_t^{t+\Delta t} e^{-\alpha_{1i}(\tau - t)}\left[-e_{pi}^T Q_{pi}e_{pi} - \left(u_{pi}^n\right)^T R_{pi}u_{pi}^n\right]d\tau \\ &-\int_t^{t+\Delta t} 2e^{-\alpha_{1i}(\tau-t)}\left(u_{pi}^{n+1}\right)^T R_{pi}u_{pei}d\tau, \end{aligned} \quad (16)$$

where $V_{pi}^n\left(X_{pi}\right)$ and u_{pi}^n indicate the updated value of $V_{pi}\left(X_{pi}\right)$ and u_{pi} in the nth iteration. It can be seen from (16) that the V_i^n and $u_{\tau i}^{n+1}$ can be updated using (16), which yields Algorithm 3.2.

In Algorithm 3.2, the performance function V_i^n and the control policy $u_{\tau i}^{n+1}$ in the nth iteration can be approximated by a critic neural network (NN) and an action NN. The weights of the NNs can be updated by using the least-square (LS) method under a persistence excitation (PE) condition (see, Lewis and Syrmos. (1995)). Define $\Theta_{ri} = \left[\phi_{ri}\ \theta_{ri}\ \psi_{ri}\right]^T$ as the attitude reference for the attitude controller of the ith quadrotor. Once the fault-tolerant position control policy u_{pi} in (14) is determined, from (5), one can derive the desired control input u_{zi}, the desired roll angle ϕ_{ri}, and the desired pitch angle θ_{ri} for the attitude controller with

$$\begin{aligned} &u_{zi}{=}u_{pzi}/\cos\theta_i/\cos\phi_i, \\ &\phi_{ri}{=}\arcsin\left(\left(\cos\phi_i\sin\theta_i\sin\psi_i - u_{pyi}/u_{zi}\right)/\cos\psi_i\right), \\ &\theta_{ri}{=}\arcsin\left(\left(u_{pxi}/u_{zi} - \sin\phi_i\sin\psi_i\right)/\cos\psi_i/\cos\phi_i\right). \end{aligned} \quad (17)$$

3.3 Attitude Controller Design

Let $x_{\Theta ri} = \left[\Theta_{ri}^T\ \dot{\Theta}_{ri}^T\right]^T \in \mathbb{R}^{6\times 1}$ be the state of the attitude reference and $x_{\Theta i} = \left[\Theta_i^T\ \dot{\Theta}_i^T\right]^T \in \mathbb{R}^{6\times 1}$ the attitude state of the ith quadrotor. From (7), an attitude augmented system can be constructed as

$$\dot{X}_{\Theta i} = \bar{F}_{\Theta}\left(X_{\Theta i}\right) + \bar{B}_{\tau i}\left(u_{\tau i} - \mu(t-t_s)u_{\Delta\Theta i}\right), \quad (18)$$

Algorithm 3.3: RL algorithm for attitude controller

Initialization: Apply a stabilizing policy $u_{\tau i}$ with bounded $u_{\tau ei}$ to the quadrotor and collect the system data.
Step 1: Solve for $V_{\Theta i}^n$ and $u_{\tau i}^{n+1}$ simultaneously:

$$\begin{aligned} &e^{-\alpha_{2i}\Delta t}V_{\Theta i}^n\left(X_{\Theta i}\left(t+T\right)\right) - V_{\Theta i}^n\left(X_{\Theta i}\left(t\right)\right) = \\ &\int_t^{t+\Delta t} e^{-\alpha_{2i}(\tau-t)}\left(-e_{\Theta i}^T Q_{\Theta i}e_{\Theta i} - \left(u_{\tau i}^n\right)^T R_{\Theta i}u_{\tau i}^n\right)d\tau \\ &-\int_t^{t+\Delta t} e^{-\alpha_{2i}(\tau-t)}2\left(u_{\tau i}^{n+1}\right)^T R_{\Theta i}u_{\tau ei}d\tau. \end{aligned} \quad (24)$$

Step 2: Set $u_{\tau i}^n = u_{\tau i}^{n+1}$ and $V_{\Theta i}^n = V_{\Theta i}^{n+1}$ and go to Step 1 until a stopping criterion is reached.

where $\bar{F}_{\Theta}(X_{\Theta i}) = \left[F_{\Theta i}(x_{\Theta i})^T\ F_{\Theta ri}(x_{\Theta ri})^T\right]^T \in \mathbb{R}^{12\times 1}$, $X_{\Theta i} = \left[x_{\Theta i}^T\ x_{\Theta ri}^T\right]^T \in \mathbb{R}^{12\times 1}$, $\bar{B}_{\tau i} = \left[B_{\tau i}^T\ 0\right]^T$, and $B_{\tau i} = \left[c_{6,4}b_{\tau i,1}\ c_{6,5}b_{\tau i,2}\ c_{6,6}b_{\tau i,3}\right]$. $F_{\Theta ri}\left(x_{\Theta ri}\right) \in \mathbb{R}^{6\times 1}$ is an unknown function related to $x_{\Theta ri}$. $F_{\Theta i}\left(x_{\Theta i}\right) \in \mathbb{R}^{6\times 1}$ is the nonlinear function with

$$F_{\Theta i}\left(x_{\Theta i}\right) = \begin{bmatrix} 0_{3\times 3} & I_3 \\ 0_{3\times 3} & -J_i^{-1}C\left(\Theta_i, \dot{\Theta}_i\right) \end{bmatrix} x_{\Theta i}. \quad (19)$$

Define $e_{\Theta i} = \left[e_{\phi i}\ e_{\theta i}\ e_{\psi i}\right]^T \in \mathbb{R}^{3\times 1}$ as the attitude tracking error with $e_{\Theta i} = \left[C_{\tau}\ -C_{\tau}\right]X_{\Theta i}$ and $C_{\tau} = \left[c_{6,1}\ c_{6,2}\ c_{6,3}\right]^T$. Then, a performance function for the attitude system is defined as

$$V_{\Theta i}\left(e_{\Theta i}, u_{\tau i}\right) = \int_t^{\infty} e^{-\alpha_{2i}(\tau-t)}\left(e_{\Theta i}^T Q_{\Theta i}e_{\Theta i} + u_{\tau i}^T R_{\Theta i}u_{\tau i}\right)d\tau, \quad (20)$$

where $Q_{\Theta i} = Q_{\Theta i}^T > 0$, $R_{\Theta i} = R_{\Theta i}^T > 0$, and $\alpha_{2i} > 0$. Then, one can obtain an optimal attitude control policy $u_{\tau i}^*$ as $u_{\tau i}^* = -\frac{1}{2}R_{\Theta i}^{-1}\bar{B}_{\tau i}^T\Delta V_{\Theta i}^*$, where $V_{\Theta i}^*$ is the solution to the following equation as

$$\begin{aligned} e_{\Theta i}^T Q_{\Theta i}e_{\Theta i} - \frac{1}{4}\left(\Delta V_{\Theta i}^*\right)^T\bar{B}_{\tau i}^T R_{\Theta i}^{-1}\bar{B}_{\tau i}\Delta V_{\Theta i}^* \\ - \alpha_{2i}V_{\Theta i} + \left(\Delta V_{\Theta i}^*\right)^T\bar{F}_{\Theta}\left(X_{\Theta i}\right) = 0. \end{aligned} \quad (21)$$

It can be seen that the equation (21) is nonlinear with respect to $V_{\Theta i}^*$ and requires the dynamic model of the quadrotor team. To track the desired attitude reference for quadrotor team with actuator faults, similarly to the position controller design, a fault-tolerant attitude controller is designed as

$$u_{\tau i}(t) = \frac{1}{2}u_{\tau i}^*(x) + \frac{\hat{\bar{u}}_{\Delta\Theta i}^2 R_{\Theta i}u_{\tau i}^*(x)}{\left\|R_{\Theta i}u_{\tau i}^*(x)\right\|\hat{\bar{u}}_{\Delta\Theta i} + \sigma_2(t)}, \quad (22)$$

where $\sigma_2(t)$ is a positive bounded function with $\int_t^{\infty}\sigma_2(t)d\tau \leq \bar{\sigma}_2 < \infty$. $\hat{\bar{u}}_{\Delta\Theta i}$ is the estimated value of $\bar{u}_{\Delta\Theta i}$ with

$$\frac{d\hat{\bar{u}}_{\Delta\Theta i}}{dt} = -k_{\sigma 2}\sigma_2(t)\hat{\bar{u}}_{\Delta\Theta i} + 2k_{\sigma 2}\left\|R_{\Theta i}u_{\tau i}^*(x)\right\|, \quad (23)$$

where $k_{\sigma 2}$ is a positive constant. Therefore, similarly to the position controller design, Algorithm 3.3 is designed to learn the optimal attitude control policy $u_{\tau i}^*$.

Similarly to the lines of reasoning of the position controller design, the performance function $V_{\Theta i}^n$ and the control

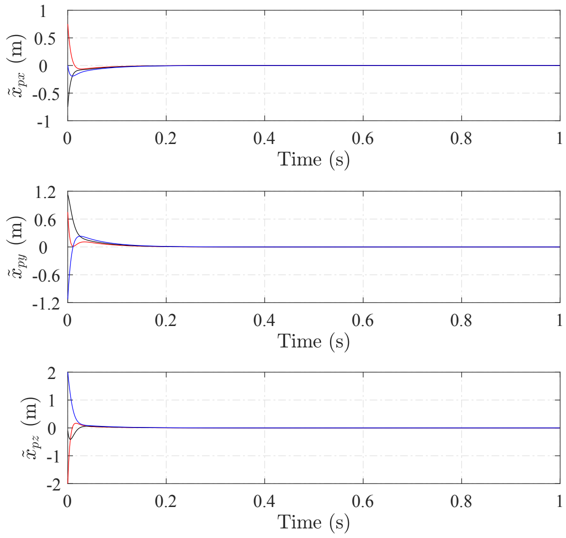

Fig. 1. Estimation error of observers.

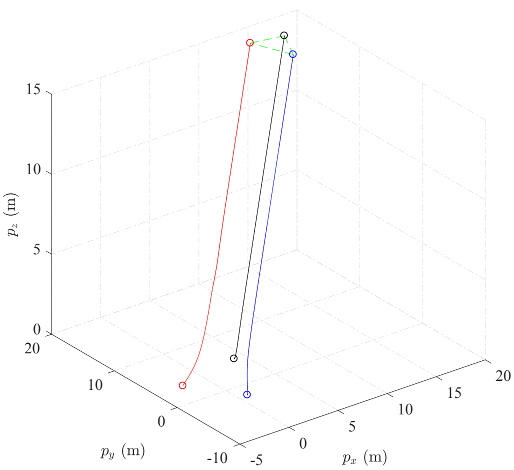

Fig. 2. 3-D trajectory of quadrotor team.

policy $u_{\tau i}^{n+1}$ in (24) can be approximated by two NNs and the weight of the NNs can be updated using the LS technique under the PE condition.

4. SIMULATION RESULTS

In this section, three quadrotor systems are modeled as (7) with the parameters being $J_{\Theta i} = diag(4, 4, 8)\ 10^{-3} kg \cdot m^2$, $b_{\Theta i} = diag\{41, 41, 113\}$, $g_i = 9.81\ m/s^2$, and $b_{pi} = diag\{1, 1, 1\}$ $(i = 1, 2, 3)$. The communication graph between the quadrotor team is described with $w_{13} = w_{21} = w_{32} = 1$ and $\rho_1 = 1$. The quadrotor team is required to form a triangle formation and track a virtual leader, simultaneously. The dynamic of the leader is set as $A_{p0} = [0_{6\times 3}\ c_{6,1}\ c_{6,2}\ c_{6,3}]$ with the initial states as: $p_0(0) = 0_{3\times 1}$ m and $\dot{p}_0(0) = [1\ 1\ 1]^T$m/s. the states of the quadrotors are initialized as: $p_1(0) = [2.25\ 2.25\ 1.00]^T$ m, $p_2(0) = [-2.25\ 3.38\ 0]^T$ m, $p_3(0) = [0\ -3.38\ 0.50]^T$ m, $\dot{p}_i = 0_{3\times 1}$,

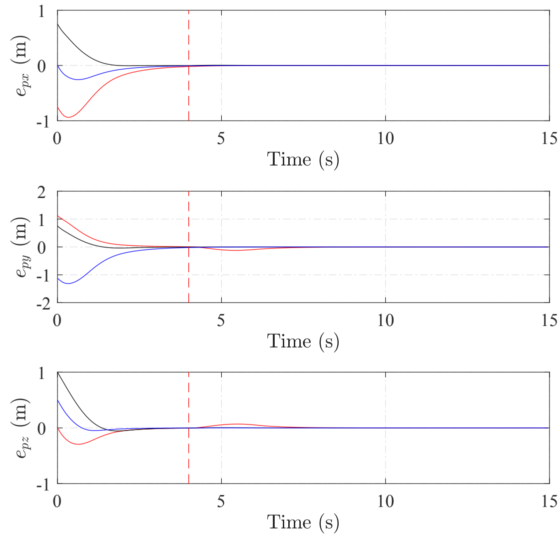

Fig. 3. Position tracking errors of quadrotor team using the proposed controller.

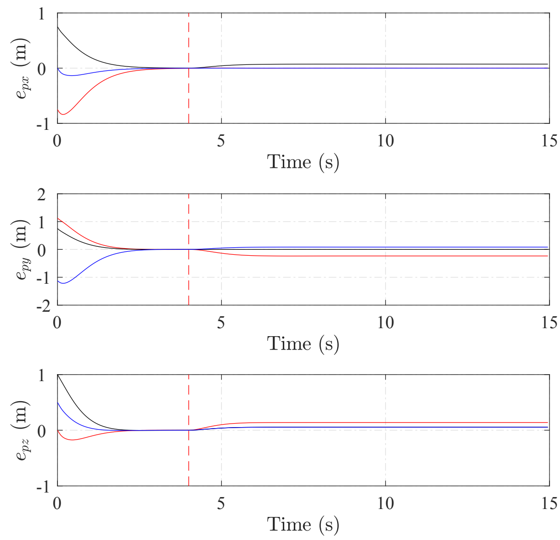

Fig. 4. Position tracking errors of quadrotor team using traditional optimal control policies.

$\Theta_i = 0_{3\times 1}$, and $\dot{\Theta}_i = 0_{3\times 1}$ $(i = 1, 2, 3)$. The position deviations are set as $\delta_{13} = [1.5\ 3.75\ 0]^T$ m, $\delta_{21} = [-3\ 0.75\ 0]^T$ m, and $\delta_{32} = [1.50\ -4.5\ 0]^T$ m. The distributed observer is implemented for each quadrotor with the parameter as $\gamma_i = 80$ and the position estimation errors are shown in Fig. 1. The solid lines colored with red, black, and blue represent the 1-3th quadrotor. After the estimation errors of the observers converge to 0s, the RL algorithms are applied to learn the optimal control policies u_{pi}^*,$u_{\tau i}^*$ for the position subsystem and the attitude subsystem. The base functions of the critic NNs and the action NNs are chosen to be multiple polynomials. The time interval Δt is chosen to be $\Delta t = 0.05$s. Then, the fault-tolerant formation controller including the position controller and the attitude controllers is constructed. The parameters of the fault-tolerant controller are set as: $R_{pi} = R_{\Theta i} = I_{3\times 3}$, $\sigma_1(t) = \sigma_2(t) = e^{-t}$, $k_{\sigma_1} = 0.1$,$k_{\sigma_2} = 0.1$, $\hat{\bar{u}}_{\Delta pi}(0) = 0.1$, and $\hat{\bar{u}}_{\Delta i}(0) = 0.1$. It is assumed that one rotor of each quadrotor is suffering from actuator faults and the LOE gains are set as: $\rho_{11} = 0.1$, $\rho_{22} = 0.3$, and $\rho_{34} = 0.1$. The occurrence time of the actuator faults t_s is set as $t_s = 4$s. The three dimensional (3-D) trajectories of the three quadrotors are shown in Fig. 2. The position tracking error $e_{pi}(t)$ in E^I is shown in Figs. 3. In order to show the effectiveness of the proposed controller, the optimal control policies are directly applied to the quadrotor team under actuator faults. The position tracking error is shown in Fig. 4. It can be seen from Figs. 4 that the traditional optimal control policies can maintain a considerable formation performance without actuator faults, but the formation performance turns undesirable when the actuator faults appear. When the actuator faults occur at 4 s, it can be observed from Figs. 3 that the formation performance of the quadrotor team is greatly improved by using the proposed fault-tolerant formation controller.

5. CONCLUSION

In this paper, the fault-tolerant formation control problem is addressed for multiple cooperative quadrotors with unknown dynamics. The quadrotor is modeled as a nonlinear and coupled system being subject to actuator faults. A fault-tolerant formation controller is proposed including a distributed observer to estimate the position references for each quadrotor, a fault-tolerant position controller to achieve the desired position formation, and a fault-tolerant attitude controller to track the desired attitude. Reinforcement learning algorithms are designed to learn the optimal control policies of the position and attitude controller without knowledge of the quadrotor dynamic information. Simulation results are provided to show the effectiveness of the proposed fault-tolerant formation controller.

REFERENCES

Avram, R. C., Zhang, X., and Muse, J. (2018) Nonlinear adaptive fault-tolerant quadrotor altitude and attitude tracking with multiple actuator faults. *IEEE Transactions on Control Systems Technology*, 26(2): 701–707.

Avram, R. C., Zhang, X., and Muse, J. (2017). Quadrotor actuator fault diagnosis and accommodation using nonlinear adaptive estimators. *IEEE Transactions on Control Systems Technology*, 25(6):2219–2226.

Deng, C., Yang, G. H. (2017). Adaptive fault-tolerant control for a class of nonlinear multi-agent systems with actuator faults. *Journal of the Franklin Institute*, 354(12):4784–4800.

Deptula, P., Bell, Z. I., Doucette, E. A., Curtis, J. W., and Dixon, W. E. (2018). Data-based reinforcement learning approximate optimal control for an uncertain nonlinear system with partial loss of control effectiveness. *2018 Annual American Control Conference*, Wisconsin, U.S.A., 6947–6968.

Dong, X., Hua, Y., Zhou, Y., Ren, Z., and Zhong, Y. (2018). Theory and experiment on formation-containment control of multiple multirotor unmanned aerial vehicle systems. *IEEE Transactions on Automation Science and Engineering*, 16(1):229–240.

Hu, Q., Zhang, Y., Huo, X., and Bing, X. (2011). Adaptive integral-type sliding mode control for spacecraft attitude maneuvering under actuator stuck failures. *Chinese Journal of Aeronautics*, 24(1):32–45.

Hua, Y., Dong, X., Li, Q., and Ren, Z. (2016). Distributed fault-tolerant time-varying formation control for second-order multi-agent systems with actuator failures and directed topologies. *IEEE Transactions on Circuits and Systems II: Express Briefs*, 65(6):774–778.

Liu, H., Tian, Y., Lewis, F. L., Wan, Y., and Valavanis, K. P. (2019). Robust formation tracking control for multiple quadrotors under aggressive maneuvers. *Automatica*, 105:179–185.

Liu, L., Wang, Z., and Wang, H. (2017). Adaptive fault-tolerant tracking control for MIMO discrete-time systems via reinforcement learning algorithm with less learning parameters. *IEEE Transactions on Automation Science and Engineering*, 14(1):299–313.

Modares, H. and Lewis, F. L. (1995). Optimal control. NJ, Hoboken: Wiley.

Ma, H. J., Xu, L. X., and Yang, G. H. (2019). Multiple environment integral reinforcement learning-based fault-tolerant control for affine nonlinear systems. *IEEE transactions on cybernetics*, unpublished.

Modares, H. and Lewis, F. L. (2014). Linear quadratic tracking control of partially-unknown continuous-time systems using reinforcement learning. *IEEE Transactions on Automatic control*, 59(11):3051–3056.

Qin, J., Ma, Q., Shi, Y., and Wang, L. (2017). Recent advances in consensus of multi-agent systems: A brief survey. *IEEE Transactions on Industrial Electronics*, 64(6):4972–4983

Raffo, G. V , Ortega, M. G., and Rubio, F. R. (2010). An integral predictive/nonlinear H_∞ control structure for a quadrotor helicopter. *Automatica*, 46(1):29–39.

Shi, J., Yang, Y., Sun, J., He, X. , Zhou, D., and Zhong, Y. (2017). Fault-tolerant formation control of non-linear multi-vehicle systems with application to quadrotors. *IET Control Theory & Applications*, 11(17):3179–3190.

Slotine, J. and Li, W. (1991). Applied Nonlinear Control. Englewood Cliffs, NJ, USA: Prentice Hall.

Wang, Z., Liu, L., Zhang, H., and Xiao, G. (2016). Fault-tolerant controller design for a class of nonlinear MIMO discrete-time systems via online reinforcement learning algorithm. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 46(5):611–622.

Yu, Z., Qu, Y., and Zhang, Y. (2019). Distributed fault-tolerant cooperative control for multi-UAVs under actuator fault and input saturation. *IEEE Transactions on Control Systems Technology*, unpublished.

Zhang, K., Zhang, H., Gao, Z., and Su, H. (2018). Online adaptive policy iteration based fault-tolerant control algorithm for continuous-time nonlinear tracking systems with actuator failures. *Journal of the Franklin Institute*, 355(15):6947–6968.