Spacecraft Attitude Takeover Control by Multiple Microsatellites Using Differential Game

Baolin Wu, *Member, IEEE*, Keyu Chen, Danwei Wang, *Fellow, IEEE*, and Yuxiang Sun, *Member, IEEE*

***Abstract*—This article proposes a differential game-based control scheme to address an attitude takeover control problem via microsatellites attached to the surface of target spacecraft. First, the attitude dynamics of combined spacecraft is reformulated as a general form so as to apply the reinforcement learning framework. Then, quadratic and Arctanh-type performance indices are designed in two cases of free input and input saturation, respectively. Accordingly, optimal control policy of each microsatellite is obtained and is dependent on value functions, which are solutions of a set of Hamilton–Jacobi–Bellman (HJB) equations. Single layer neural networks are employed to approximate value functions by policy iteration and the weights vectors are updated with the help of the concurrent learning algorithm so that the persistent excitation condition of control errors is loosened. Moreover, the necessity of interaction among microsatellites is eliminated by using a tracking differentiator technique to estimate angular accelerations which are not available directly through onboard devices. Stability of the closed-loop system is guaranteed by the Lyapunov method. Three cases of simulation are carried out to demonstrate the robustness and optimality of the proposed control scheme and to validate the effectiveness of the controller in the presence of actuators saturation.**

***Index Terms*—Attitude control, concurrent learning, differential game, input saturation, reinforcement learning (RL), takeover control.**

I. Introduction

WITH THE development of space technology, more satellites of various types and functions have been widely used in military reconnaissance, data communication, navigation and positioning, astronomical observation, and other fields, playing an irreplaceable role in safeguarding national security and ensuring peoples' livelihoods. More satellites in space lead to more anomalies as a result of which the on-orbit service is becoming increasingly popular [1]. Of various types of service, attitude takeover control [2] is paid much attention to by researchers due to its promising application in practical on-orbit maintenance for damaged spacecraft. Attitude takeover control means that multiple microsatellites are launched into the orbit, and attach to the surface of failed spacecraft in order to control its attitude taking place of its original failed attitude control subsystem, which therefore helps it regain ability of attitude control. As a result, the lifetime of spacecraft can be prolonged [3]. Compared to one space-robotic takeover control, using multiple microsatellites can increase the redundancy and reliability and reduce the cost of the takeover control system.

There are some key challenges brought by attitude takeover control as following [4]: 1) unknown inertia matrix of the combined spacecraft; 2) distributed attitude control law design; and 3) allocation of control torques to the multiple microsatellites. A few works have been published to investigate attitude takeover control. According to controller structure, they can be categorized into two classes: 1) control and allocation separate structure and 2) decentralized structure. For the first class, hierarchical structure is proposed where control law and allocation are designed separately and taken as upper level and lower level, respectively. The control law is generated by the computation module and then, taken as input of torque allocation module which allocates the total control command into every actuator as output. In the upper level, the attitude control law, as a longstanding topic, has been studied from multiple perspectives such as the PD method [5], [6]; adaptive control [7], [8]; robust control method [9], [10]; optimal control [11], [12]; finite-time control [13], [14]; quantized control [15], [16]; event-triggered control [17], [18]; and neural network-based control [19]. In terms of torque allocation as lower level of whole mission, various methods have been studied, which can be categorized into centralized allocation methods [20], [21] and distributed allocation methods [22], [23]. It is noteworthy that both centralized and distributed mode of allocation belong to the lower level of hierarchical control structure and hence are of centralized control scheme essentially. However, an obviously potential defect of centralized control is that the whole attitude takeover control system fails once the computation module malfunctions.

To increase robustness of the system, some researchers proposed a second kind of control structure for the attitude takeover control problem, which presents a decentralized control scheme

Manuscript received 10 July 2022; revised 23 December 2022 and 11 May 2023; accepted 12 June 2023. Date of publication 27 June 2023; date of current version 1 March 2024. This work was supported by the National Natural Science Foundation of China under Grant 61873312. Recommended by Associate Editor L. Pallottino. *(Corresponding author: Keyu Chen.)*

Baolin Wu and Keyu Chen are with the Research Center of Satellite Technology, Harbin Institute of Technology, Harbin 150080, China (e-mail: wuba0001@e.ntu.edu.sg; keyu.chen@connect.polyu.hk).

Danwei Wang is with the Nanyang Technological University, Singapore 639798 (e-mail: ed-wwang@ntu.edu.sg).

Yuxiang Sun is with the Department of Mechanical Engineering, The Hong Kong Polytechnic University, Hung Hom, Kowloon 999077, Hong Kong (e-mail: yx.sun@polyu.edu.hk).

Digital Object Identifier 10.1109/TCNS.2023.3290082

where each satellite unit is in charge of sensing, computing control law and actuating, therefore making the controller and torque allocation integrated together. In this scheme, there is no central computer and allocator. A feasible way of mentioned decentralized structure is differential game-based control scheme [24], [25], [26], [27]. In [24], the differential game-based control strategy was first introduced to solve multimicrosatellite attitude control problem. The differential game model addressing attitude control is established and a neural network with single layer is employed to solve the nonlinear Hamilton–Jacobi–Bellman (HJB) equation. In [25], the attitude model was linearized and the control strategies of multiple microsatellites were obtained by optimizing individual quadratic performance index functions. In [28], disturbance was regarded as one player in differential game framework and in the meantime, a robust event-triggered control strategy is proposed. In [27], a zero-sum differential game strategy-based control method was given by policy iteration to deal with tracking control problem of modular robot manipulators considering uncertain disturbance. What the aforementioned methods have in common is to apply the idea of optimal control when dealing with problem of differential game. Specifically, the problem of optimal differential game is formulated as continuous Markov decision process (MDP) and reinforcement learning is applied for decision. Although aforementioned literature employed differential game-based control method to achieve distributed control command computation, existing schemes remain to have some deficiencies. First, to the authors' best knowledge, saturation of actuators is barely considered in current literature, which is crucial for the attitude takeover control by microsatellites. Second, there exist interactions among individual satellites for parameters sharing. That raises a requirement for intersatellite communication and makes the system complicated. Moreover, the mentioned control methods suffer from persistent excitation condition meaning that convergence could be hindered once state variables drop into a small neighborhood around origin since errors are not large enough to drive the identifier working normally. Hence, the problem of attitude takeover control is still open and worth studying.

The article proposes a differential game-based attitude takeover control strategy for failed spacecraft via multiple microsatellites. First, the rigid body attitude kinematics and dynamics are reformulated into general vector form and corresponding optimized index functions for each microsatellite are defined. Then reinforcement learning-based approach is leveraged to solve the multiplayer differential game problem. Specifically, it is performed in two stages. First, by solving canonical optimal control problem, policies depending on unknown value functions are obtained. Then, a few single-layer neural networks are used to approximate value functions, which satisfy the constraints of coupled nonlinear HJB equations. Different from commonly used gradient descent training strategy, in our work concurrent learning is incorporated into adaptive law to drive updating of the weights online. Concurrent learning considers both past recorded data and current data in order to loosen persistent excitation condition. In addition, a tracking differentiator is employed to estimate angular accelerations so that restriction of necessary interaction among microsatellites is lifted. Moreover, the article deduces HJB equation and optimal policies of microsatellite under two cases of free input and actuator saturation. Finally, system stability is guaranteed by Lyapunov method and numerical simulations validate the effectiveness of proposed control scheme.

TABLE I
ABBREVIATION TABLE

Abbreviation	Definition
HJB equation	Hamilton–Jacobian–Bellman equation
MDP	Markov decision process
PE	Persistent excitation
RL	Reinforcement learning
VFA	Value function approximation

Compared to existing control schemes for attitude takeover control, the innovative contributions of this article are as follows.

1) The article investigates the optimal differential game control problem under input saturation by developing a kind of newly defined performance index functions and then, deriving the corresponding HJB equations and optimal control strategies.
2) The requirement of information exchange among microsatellites is eliminated by using tracking differentiator to estimate angular accelerations of combined spacecraft.
3) The convergence of weights is not restricted by persistent excitation condition and is completed in finite time by the proposed adaptive updating law, which means computed control policies are closer to optimal trajectories than the existing methods.

The rest of this article is organized as follows. Section II presents fundamental knowledge applied in proposed approach including attitude dynamics, game theory, HJB equation, concurrent learning and tracking differentiator. Section III clarifies the definition of attitude takeover control task and further describes it as a multiplayer differential game problem. Table I is provided to introduce the definition of abbreviation used. Section IV is controller design under two cases of unsaturated and saturated situation, and two kinds of stability proof are provided for each case. Section V provides a discussion on the conclusions the article draws. Section VI shows the results of experimental simulation. Finally, Section VII concludes this article.

II. PRELIMINARIES

A. Attitude Dynamics of Rigid Body

The combined spacecraft consists of target spacecraft and a group of microsatellites and each satellite can provide three-axis torques. Hence, kinematics and dynamics of the combined spacecraft are given as

$$\begin{aligned} \dot{q}_0 &= -\frac{1}{2}\boldsymbol{q}_v^T\boldsymbol{\omega} \\ \dot{\boldsymbol{q}}_v &= \frac{1}{2}(q_0\mathbf{I}+\boldsymbol{q}_v^{\times}\boldsymbol{\omega}) \\ \mathbf{J}\dot{\boldsymbol{\omega}} &= -\boldsymbol{\omega}^{\times}\mathbf{J}\boldsymbol{\omega}+\boldsymbol{\tau} \end{aligned} \tag{1}$$

where $\boldsymbol{q}^T = [\,q_0 \quad \boldsymbol{q}_v^T\,]$ is quaternion used to describe the attitude of spacecraft in the body frame $\mathcal{B}$ with respect to inertial frame $\mathcal{I}$, $q_0 \in \mathbb{R}$ and $\boldsymbol{q}_v \in \mathbb{R}^3$ are the scalar and vector part, respectively, satisfying $q_0^2 + \boldsymbol{q}_v^T \boldsymbol{q}_v = 1$, $\boldsymbol{\omega}$ is the angular velocity, $\tau \in \mathbb{R}^3$ is control torque imposed on the rigid body, $\mathbf{I}$ is identity matrix, $\mathbf{J}$ is matrix of the rotational inertia of spacecraft, and $\boldsymbol{q}_v^\times$ denotes cross operator of $\boldsymbol{q}_v = [\;q_{v1} \quad q_{v2} \quad q_{v3}\;]^T$, i.e.,

$$\boldsymbol{q}_v^\times = \begin{bmatrix} 0 & -q_{v3} & q_{v2} \\ q_{v3} & 0 & -q_{v1} \\ -q_{v2} & q_{v1} & 0 \end{bmatrix}. \tag{2}$$

According to the properties of quaternion, the attitude of the body frame $\mathcal{B}$ relative to the desired frame $\mathcal{R}$ is computed as follows:

$$\begin{aligned} \tilde{q}_0 &= q_{0\,d} q_0 + \boldsymbol{q}_{vd}^T \boldsymbol{q}_v \\ \tilde{\boldsymbol{q}}_v &= q_{0\,d} \boldsymbol{q}_v - q_0 \boldsymbol{q}_{vd} - \boldsymbol{q}_{vd}^\times \boldsymbol{q}_v \\ \boldsymbol{\omega}_e &= \boldsymbol{\omega} - \mathbf{R}\boldsymbol{\omega}_d \end{aligned} \tag{3}$$

where $\boldsymbol{q}_d^T = [\,q_{0\,d} \quad \boldsymbol{q}_{vd}^T\,]$ is the quaternion that describes desired attitude of spacecraft in the desired frame $\mathcal{R}$ with respect to inertial frame $\mathcal{I}$, $\tilde{\boldsymbol{q}}^T = [\,\tilde{q}_0 \quad \tilde{\boldsymbol{q}}_v^T\,]$ is error quaternion describing error attitude of spacecraft in the body frame $\mathcal{B}$ with respect to desired frame $\mathcal{R}$, $\mathbf{R}$ is transformation matrix between body frame $\mathcal{B}$ and desired frame $\mathcal{R}$, given by $\mathbf{R} = (1 - 2\tilde{\boldsymbol{q}}_v^T \tilde{\boldsymbol{q}}_v)\mathbf{I} + 2\tilde{\boldsymbol{q}}_v \tilde{\boldsymbol{q}}_v^T - 2\tilde{q}_0 \tilde{\boldsymbol{q}}_v^\times$, satisfying $\|\mathbf{R}\| = 1$. $\boldsymbol{\omega}_d$ is desired angular velocity and $\boldsymbol{\omega}_e$ is the error angular velocity describing the angular velocity of body frame $\mathcal{B}$ with respect to desired frame $\mathcal{R}$. According to kinematics of rigid body, the derivation of transformation matrix $\mathbf{R}$ can be expressed as

$$\dot{\mathbf{R}} = -\boldsymbol{\omega}_e^\times \mathbf{R}. \tag{4}$$

The attitude error dynamics of rigid body can be obtained as (5) using (1)–(4)

$$\begin{aligned} \dot{\tilde{\boldsymbol{q}}}_v &= \frac{1}{2}(\tilde{q}_0 \mathbf{I} + \tilde{\boldsymbol{q}}_v^\times)\boldsymbol{\omega}_e \\ \mathbf{J}\dot{\boldsymbol{\omega}}_e &= -\boldsymbol{\omega}^\times \mathbf{J}\boldsymbol{\omega} + \mathbf{J}(\boldsymbol{\omega}_e^\times \mathbf{R}\boldsymbol{\omega}_d - \mathbf{R}\dot{\boldsymbol{\omega}}_d) + \boldsymbol{\tau}. \end{aligned} \tag{5}$$

B. Differential Game Control

The differential game control is to find out the optimal control strategy at equilibrium point for each player based on optimal control theory. For differential (6)

$$\dot{\boldsymbol{x}} = \boldsymbol{f}(\boldsymbol{x}) + \sum_{i=1}^{N} \mathbf{g}_i \boldsymbol{u}_i. \tag{6}$$

N players $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_N$ manipulate system state $\boldsymbol{x}$ jointly,where $\boldsymbol{u}_i$ and $\boldsymbol{x}$ are given as vectors, and $\mathbf{g}_i$ is coefficient matrix. For each player, an index is defined as

$$L_i = \int_0^{+\infty} \left(Q_i(\boldsymbol{x}) + \Gamma_i(\boldsymbol{u}_1, \ldots, \boldsymbol{u}_N)\right) dt \tag{7}$$

where $Q_i(\boldsymbol{x})$ is a positive function of $\boldsymbol{x}$ and $\Gamma_i(\boldsymbol{u}_1, \ldots, \boldsymbol{u}_N)$ is the positive definite function of control inputs of all players. Control strategies $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_N$ are designed for the sake of optimizing index L_i for each player i, respectively.

Remark 1: Both Q_i and Γ_i are scalars and L_i represents synthesized cost of control performance and actuators energy consumption in attitude control task. $Q_i(\boldsymbol{x})$ is usually set as quadratic form $\boldsymbol{x}^T \mathbf{Q}_i \boldsymbol{x}$ and $\Gamma_i(\boldsymbol{u}_1, \ldots, \boldsymbol{u}_N)$ is usually set as sum of quadratic form $\sum_{j=1}^{N} \boldsymbol{u}_j^T \boldsymbol{\Gamma}_{ij} \boldsymbol{u}_j$ for each player's control strategy, which means that all players are cooperative. In addition, $\Gamma_i(u_i)$ can also be set as quadratic form $\boldsymbol{u}_i^T \boldsymbol{\Gamma}_i \boldsymbol{u}_i$ only dependent on its own control input $\boldsymbol{u}_i$ under information exchange-free situation adopted in this article. However, in the case considering actuator saturation, $\Gamma_i(\boldsymbol{u}_1, \ldots, \boldsymbol{u}_N)$ is set as integral form instead of quadratic form in order to guarantee amplitude of optimal control strategy is bounded.

Nash equilibrium: A set of strategies satisfying conditions (8) is called Nash equilibrium strategies.

$$\begin{aligned} L_i^* &= L_i(u_1^*, \ldots, u_i^*, \ldots, u_N^*) \\ &\le L_i(u_1^*, \ldots, u_i, \ldots, u_N^*). \end{aligned} \tag{8}$$

At Nash equilibrium point, each player's policy cannot be changed unilaterally, otherwise index of changer is not optimal.

C. Coupled HJB Equation and Reinforcement Learning

Reinforcement learning (RL) is a powerful tool in amounts of applications including autonomous systems. With the thought of trial and error, RL aims to leverage strong learning ability to evaluate the value of each action and then, select the best one for maximizing reward or minimizing penalty by well-trained neural network. Due to intrinsic consistence between optimal control and RL, RL is an ideal candidate technique for solving optimal control problem by setting optimized object function as reward. The neural network can learn to approximate the value function in arbitrary precision in avoidance of analytic procedure. Systems in nature are classified into deterministic and stochastic type. For stochastic system, a probabilistic distribution is generated to cope with the uncertainty of process while achieving optimal expectation theoretically. In this article, the attitude dynamics of the combined spacecraft formulated as MDP is assumed to be a priori without uncertainty. Only deterministically continuous MDP is considered in the article.

Define value function as

$$V_i(\boldsymbol{x}(t)) = \int_t^{\infty} (Q_i(\boldsymbol{x}) + \Gamma_i(\boldsymbol{u}_i)) d\tau \tag{9}$$

where $Q_i(\boldsymbol{x})$ and $\Gamma_i(\boldsymbol{u}_i)$ are the same as (7). Derivate (9), then Bellman (10) is obtained

$$0 = Q_i(\boldsymbol{x}) + \Gamma_i(\boldsymbol{u}_i) + (\nabla \boldsymbol{V}_i)^T \left(\boldsymbol{f}(\boldsymbol{x}) + \sum_{j=1}^{N} \mathbf{g}_j \boldsymbol{u}_j \right) = H_i \tag{10}$$

where $\nabla \boldsymbol{V}_i = \frac{\partial V_i}{\partial \boldsymbol{x}}$ is a column vector and right-hand side is defined as Hamilton value. According to maximum principle in optimal control, optimal policy $\boldsymbol{u}_i$ satisfies $\frac{\partial H_i}{\partial \boldsymbol{u}_i} = \mathbf{0}$. Substituting (10) into conditions $\Gamma_i(\boldsymbol{u}_i) = \boldsymbol{u}_i^T \boldsymbol{\Gamma}_i \boldsymbol{u}_i$, yields

$$\boldsymbol{u}_i = -\frac{1}{2} \boldsymbol{\Gamma}_i^{-1} \mathbf{g}_i^T \nabla \boldsymbol{V}_i. \tag{11}$$

Optimal control strategy for each player is showed as (11) under which Nash equilibrium among all players is reached. Substituting (11) into (10), HJB equation is deduced as

$$(\nabla \boldsymbol{V}_i)^T \boldsymbol{f}(\boldsymbol{x}) + Q_i(\boldsymbol{x}) - \frac{1}{2}(\nabla \boldsymbol{V}_i)^T \sum_{j=1}^{N} \mathbf{g}_j \boldsymbol{\Gamma}_j^{-1} \mathbf{g}_j^T \nabla \boldsymbol{V}_j + \frac{1}{4}(\nabla \boldsymbol{V}_i)^T \mathbf{g}_i \boldsymbol{\Gamma}_i^{-T} \mathbf{g}_i^T \nabla \boldsymbol{V}_i = 0. \tag{12}$$

Remark 2: The definition of value function is slightly different from performance index of (7) at lower limit of integral, which means that value function is a dynamic variable only dependent on state value $\boldsymbol{x}$ at time t explicitly. From the perspective of dynamic programming, it represents the optimized index from current time step rather than initial time step and hence, the control policy depends on value function.

To compute optimal control strategy as (11), it is necessary to solve (12) for $\nabla \boldsymbol{V}_i$. There are N nonlinear partial differential equations (PDEs) for i from 1 to N in (12). It is difficult to provide analytical solutions for such a set of highly coupled and nonlinear PDEs. Therefore, value function approximation (VFA) is employed to approximate value function V_i. In VFA technique, a set of basis functions and weights are used to reconstruct value function V_i as

$$V_i(x) = \mathbf{W}_i^T \boldsymbol{\Phi}(\boldsymbol{x}) + \varepsilon_i \tag{13}$$

where $\mathbf{W}_i$ is a $m \times 1$ weights vector, $\boldsymbol{\Phi}(\boldsymbol{x})$ is a basis function vector dependent on $\boldsymbol{x}$, ε_i is a scalar representing approximation error, and m is the number of basis functions. With m increasing, the accuracy of approximation is guaranteed by Lemma 1.

Lemma 1 (see [29]): For (13), approximation error $\varepsilon_i \to 0$ when the number of mutually independent basis functions $m \to \infty$.

Actually, VFA technique is equivalent to a single-layer neural network in effect.

D. Concurrent Learning

Substituting (13) into (10), weights can be separated out of other known terms as

$$\mathbf{W}_i^T \boldsymbol{\psi}_i = \kappa_i \tag{14}$$

where $\boldsymbol{\psi}_i = \nabla \boldsymbol{\Phi}_i(\boldsymbol{x}) \left(\boldsymbol{f}(\boldsymbol{x}) + \sum_{j=1}^{N} \mathbf{g}_j \boldsymbol{u}_j \right)$ and $\kappa_i = -Q_i(\boldsymbol{x}) - \boldsymbol{u}_i^T \boldsymbol{\Gamma}_i \boldsymbol{u}_i$. From the perspective of parameter identification, learning of weights vector can be attributed as issue of weights identification via (14), which is based on the concurrent learning method in this article.

According to (14), the constraints on $\mathbf{W}_i$ can be rewritten as

$$\boldsymbol{\psi}_i \boldsymbol{\psi}_i^T \mathbf{W}_i = \boldsymbol{\psi}_i \kappa_i. \tag{15}$$

Ordinary techniques such as least square method can be employed to solve (15). However, persistent excitation is required, which means square matrix $\boldsymbol{\psi}_i \boldsymbol{\psi}_i^T$ should be of full rank at any time. A new thought is proposed in this article through concurrent learning to loosen persistent excitation condition and in the meantime identification of weights vector is dealt with. Concurrent learning exploits data both from both the past and current time step to identify parameter. In continuous system, integrate (15) to time, and an integral form can be obtained as

$$\int_{t_0}^{t} e^{-l(t-\tau)} \boldsymbol{\psi}_i \boldsymbol{\psi}_i^T \mathbf{W}_i d\tau = \int_{t_0}^{t} e^{-l(t-\tau)} \boldsymbol{\psi}_i \kappa_i d\tau \tag{16}$$

where l is forgetting factor which plays the role of assigning differentiated participation of identification using $\boldsymbol{\psi}_i(\boldsymbol{x}(\tau))$ and $\kappa_i(\boldsymbol{x}(\tau))$ on different time points. From (16), both historical and current data contribute to the identification of weights $\mathbf{W}_i$ in different levels of importance by timing coefficient $e^{-l(t-\tau)}$ that is increasing as time point τ goes forward. When $\tau = t$, the coefficient $e^{-l(t-\tau)} = 1$ means no discount for data use on time point t, in the meanwhile indicating that closer data is given more consideration. Assume $\mathbf{P}_i$ and $\mathbf{S}_i$ as $\mathbf{P}_i = \int_{t_0}^{t} e^{-l(t-\tau)} \boldsymbol{\psi}_i \boldsymbol{\psi}_i^T d\tau$, $\mathbf{S}_i = \int_{t_0}^{t} e^{-l(t-\tau)} \boldsymbol{\psi}_i \kappa_i d\tau$, $\hat{\mathbf{W}}_i$ as estimation of weights vector $\mathbf{W}_i$ and employ $\mathbf{P}_i \hat{\mathbf{W}}_i - \mathbf{S}_i$ as rectification term for learning law of $\hat{\mathbf{W}}_i$, a finite time concurrent learning identification strategy can be obtained as

$$\dot{\hat{\mathbf{W}}}_i = \boldsymbol{\Pi} \frac{\mathbf{P}_i^T (\mathbf{P}_i \hat{\mathbf{W}}_i - \mathbf{S}_i)}{\left\| \mathbf{P}_i \hat{\mathbf{W}}_i - \mathbf{S}_i \right\|} \tag{17}$$

where $\boldsymbol{\Pi}$ is positive definite and 2-norm $\|\mathbf{P}_i \hat{\mathbf{W}}_i - \mathbf{S}_i\|$ ensures finite time convergence.

E. Tracking Differentiator

In each player's policy learning law, the others' control strategies are included, as showed in $\boldsymbol{\psi}_i = \nabla \boldsymbol{\Phi}_i(\boldsymbol{x})(\boldsymbol{f}(\boldsymbol{x}) + \sum_{j=1}^{N} \mathbf{g}_j \boldsymbol{u}_j)$. Therefore, in this section, a linear tracking differentiator is employed to get rid of restrictions of necessary interaction among individual players. The tracking error convergence of the linear tracking differentiator is guaranteed by Lemma 2.

Lemma 2 (see [30]): For linear tracking differentiator as

$$\begin{aligned} \dot{z}_1 &= z_2 \\ \dot{z}_2 &= -k_1 R^2 (z_1 - v) - k_2 R z_2 \end{aligned} \tag{18}$$

where k_1, k_2, and R are positive constants and $v(t)$ is signal to be tracked, if $v(t)$ satisfies $sup_{t \in [0,+\infty]}(|v(t)| + |\dot{v}(t)|) = M < +\infty$ for constant $M > 0$, then for any $a > 0, t \in [a, +\infty)$, (19) holds:

$$\lim_{R \to \infty} |z_1 - v| = 0, \lim_{R \to \infty} |z_2 - \dot{v}| = 0. \tag{19}$$

From (6), $\boldsymbol{\psi}_i = \nabla \boldsymbol{\Phi}_i(\boldsymbol{x}) \dot{\boldsymbol{x}}$ is obtained, where signal $\dot{\boldsymbol{x}}$ is tracked by z_2 in (18).

III. Problem Formulation

As illustrated in Fig. 1, the attitude takeover control task means that a group of microsatellites are attached to the surface of target spacecraft whose original attitude control subsystem malfunctions and help it regain attitude control capability. Each microsatellite equips with sensors, controller, and actuators. A decentralized controller structure can be built by the group of microsatellites attached while maintaining the robustness of whole attitude takeover control system. Robustness means that

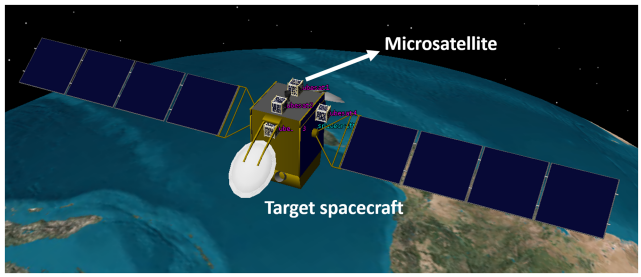

Fig. 1. Illustration of attitude takeover control by microsatellites.

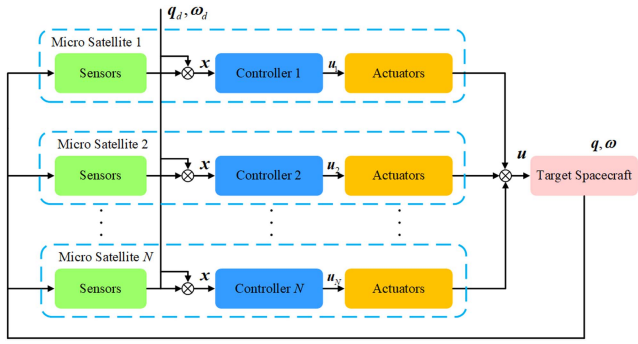

Fig. 2. Overall control scheme block diagram.

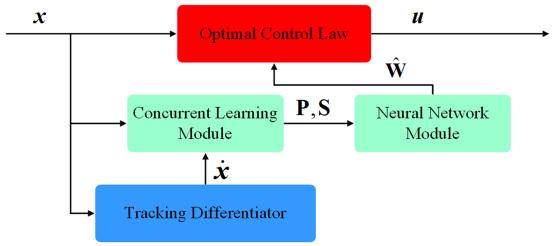

Fig. 3. Internal structure of controller.

the system remains working even if one of the microsatellites malfunctions.

According to manner of attitude takeover control introduced above, the attitude kinematics and dynamics of combination of target spacecraft and microsatellites is rewritten as

$$\begin{aligned}\dot{q}_0 &= -\frac{1}{2}\boldsymbol{q}_v^T\boldsymbol{\omega}\\ \dot{\boldsymbol{q}}_v &= \frac{1}{2}(q_0\mathbf{I}+\boldsymbol{q}_v^{\times}\boldsymbol{\omega})\\ \mathbf{J}\dot{\boldsymbol{\omega}} &= -\boldsymbol{\omega}^{\times}\mathbf{J}\boldsymbol{\omega}+\sum_{i=1}^{N}\boldsymbol{u}_i\end{aligned} \tag{20}$$

where u_i is control torque of ith microsatellite and other symbols are the same as (1). To simplify the expression, hereafter (20) is reformulated as general form (21) to (22)

$$\dot{\boldsymbol{x}} = \boldsymbol{f}(\boldsymbol{x})+\sum_{i=1}^{N}\mathbf{g}_i\boldsymbol{u}_i \tag{21}$$

$$\boldsymbol{f}(\boldsymbol{x}) = \begin{bmatrix}\frac{1}{2}(\tilde{q}_0\mathbf{I}+\tilde{\boldsymbol{q}}_v^{\times})\boldsymbol{\omega}_e\\ \mathbf{J}^{-1}(-\boldsymbol{\omega}^{\times}\mathbf{J}\boldsymbol{\omega}+\mathbf{J}(\boldsymbol{\omega}_e^{\times}\mathbf{R}\boldsymbol{\omega}_d-\mathbf{R}\dot{\boldsymbol{\omega}}_d))\end{bmatrix} \tag{22}$$

where $\boldsymbol{x}=[\tilde{\boldsymbol{q}}_v^T \quad \boldsymbol{\omega}_e^T]^T$ and $\mathbf{g}_i = \begin{bmatrix}\mathbf{O}_{3\times3}\\ \mathbf{J}^{-1}\end{bmatrix}$. The optimized index for each microsatellite is set as (23)

$$L_i = \int_0^{+\infty}\left(Q_i(\boldsymbol{x})+\Gamma_i(\boldsymbol{u}_i)\right)dt \tag{23}$$

where $Q_i(\boldsymbol{x})=\boldsymbol{x}^T\mathbf{Q}_i\boldsymbol{x}$, $\Gamma_i(\boldsymbol{u}_i)=\boldsymbol{u}_i^T\boldsymbol{\Gamma}_i\boldsymbol{u}_i$ for case without considering actuators saturation and $\Gamma_i(\boldsymbol{u}_i)=\int_0^{u_i}\eta_i\tanh^{-T}(\boldsymbol{v}/\eta_i)\boldsymbol{\Gamma}_i d\boldsymbol{v}$ for case considering actuators saturations with maximum torque of η_i, respectively.

Due to limited communication capability of microsatellites, information exchange-free control strategy is required. Therefore, the task to be dealt with in controller design is to find out an optimal control law $\boldsymbol{u}_i(i=1,\ldots,N)$ for each microsatellite optimizing index L_i without information exchange among individual units meaning that controller for each satellite is only dependent on state vector of target spacecraft. However, according to analytical solution (11), optimal control depends on its own value function that is coupled with other value functions in (12). This implies that control parameters of other microsatellites are introduced and fully connected communications are required. In addition, the popular gradient-based method used in RL often faces persistent excitation condition that generates contradiction between desiring fast convergence and sufficient errors driving convergence. The two problems commonly faced are addressed in controller design in Section IV by tracking differentiator and concurrent learning-based training strategy, respectively.

IV. Attitude Takeover Controller Design

Proposed differential game-based control scheme is illustrated as in Figs. 2 and 3. Fig. 2 shows an overall diagram for designed control scheme. In the scheme, controller design and torque allocation are synthesized. Each microsatellite outputs a three-axis torque vector acting on target spacecraft. The total torque imposed on target spacecraft is the sum of all microsatellites' torques. Fig. 3 shows the detailed structure of controller and its signal flows when each microsatellite is working. The control command is obtained by solving multi-players differential game problem, depending on attitude errors, attitude angular velocity errors, and weights of neural network. The neural network module is in charge of updating weights according to records of both past and current data through concurrent learning algorithm. Concurrent learning module takes in the error signals, control commands from microsatellites, and derivatives of errors containing angular accelerations and gives the auxiliary variables to neural network module. The angular accelerations are not measurable, but can be estimated by tracking differentiator. Specifically, two controllers are designed in this section based on differential game theory for two cases: One where actuator saturation is considered, and one where it is not.

A. Optimal Control Strategy Without Considering Control Torque Saturation

According to deduced optimal policy (10) and (13) for general system (6), the optimal control law for each microsatellite is proposed as

$$\boldsymbol{u}_i = -\frac{1}{2}\boldsymbol{\Gamma}_i^{-1}\mathbf{g}_i^T\boldsymbol{\nabla}\boldsymbol{\Phi}_i(x)\hat{\mathbf{W}}_i. \tag{24}$$

The adaptive estimation law for weights vector $\hat{\mathbf{W}}_i$ is designed as

$$\dot{\hat{\mathbf{W}}}_i = \boldsymbol{\Pi}\frac{\mathbf{P}_i^T(\mathbf{P}_i\hat{\mathbf{W}}_i - \mathbf{S}_i)}{\left\|\mathbf{P}_i\hat{\mathbf{W}}_i - \mathbf{S}_i\right\|} \tag{25}$$

where auxiliary variables $\mathbf{P}_i$ and $\mathbf{S}_i$ evolve along (26)–(27).

$$\dot{\mathbf{P}}_i = -l\mathbf{P}_i + \boldsymbol{\psi}_i\boldsymbol{\psi}_i^T \tag{26}$$

$$\dot{\mathbf{S}}_i = -l\mathbf{S}_i + \boldsymbol{\psi}_i\kappa_i \tag{27}$$

where l is a positive constant with the same connotation as (16).

The adaptive updating law (25) is designed to make weights vectors converge in finite time. The auxiliary variables $\mathbf{P}_i$ and $\mathbf{S}_i$ are matrix containing both historically recorded data and currently collected data. In fact, the solution of (26) and (27) are $\mathbf{P}_i = \int_{t_0}^{t} e^{-l(t-\tau)}\boldsymbol{\psi}_i\boldsymbol{\psi}_i^T d\tau, \mathbf{S}_i = \int_{t_0}^{t} e^{-l(t-\tau)}\boldsymbol{\psi}_i\kappa_i d\tau$ with the same form as that mentioned in Section II. Formulations (26) and (27) are equivalent to (20).

Theorem 1: Consider the attitude takeover control system (6) with control law (24) and the adaptive law (25)–(27). All signals in the closed loop are globally ultimately bounded.

Proof: The proof is divided into two steps in which first step demonstrates finite time convergence of weights learning error and the second step demonstrates convergence of attitude parameters.

Step 1: Choose Lyapunov candidate function as $L_w = \frac{1}{2}\sum_{i=1}^{N}\tilde{\mathbf{W}}_i^T\boldsymbol{\Pi}^{-1}\tilde{\mathbf{W}}_i$, where $\tilde{\mathbf{W}}_i = \mathbf{W}_i - \hat{\mathbf{W}}_i$ is weights learning errors vector.

Substituting adaptive law (25)–(27), the derivative of Lyapunov candidate function L_w yields

$$\begin{aligned}\dot{L}_w &= \sum_{i=1}^{N}\tilde{\mathbf{W}}_i^T\boldsymbol{\Pi}^{-1}\dot{\tilde{\mathbf{W}}}_i = -\sum_{i=1}^{N}\tilde{\mathbf{W}}_i^T\boldsymbol{\Pi}^{-1}\dot{\hat{\mathbf{W}}}_i \\ &= -\sum_{i=1}^{N}\tilde{\mathbf{W}}_i^T\frac{\mathbf{P}_i^T\mathbf{P}_i\tilde{\mathbf{W}}_i}{\left\|\mathbf{P}_i\tilde{\mathbf{W}}_i\right\|}.\end{aligned} \tag{28}$$

$\mathbf{P}_i\tilde{\mathbf{W}}_i$ is a vector: therefore, it follows that:

$$\begin{aligned}\dot{L}_w &= -\sum_{i=1}^{N}\left\|\mathbf{P}_i\tilde{\mathbf{W}}_i\right\| \le -\sum_{i=1}^{N}\beta_i\sqrt{\tilde{\mathbf{W}}_i^T\boldsymbol{\Pi}^{-1}\tilde{\mathbf{W}}_i} \\ &\le -\bar{\beta}\sqrt{L_w}\end{aligned} \tag{29}$$

where β_i is a positive constant satisfying $\lambda_{\max}(\beta_i^2\boldsymbol{\Pi}^{-1}) \le \lambda_{\max}(\mathbf{P}_i^T\mathbf{P}_i)$ and $\bar{\beta}$ is defined as $\bar{\beta} = \sqrt{2}\min_{1\le i\le N}(\beta_i)$. Solving (29), the upper bound of convergence time is obtained as $t_{\max} = \frac{2\sqrt{L_w(0)}}{\bar{\beta}}$.

Step 2: Choose arbitrary one value function $V_i(\boldsymbol{x})$ as Lyapunov candidate function, then its derivative is deduced as

$$\dot{V}_i(\boldsymbol{x}) = (\boldsymbol{\nabla}\boldsymbol{V}_i)^T\left(\boldsymbol{f}(\boldsymbol{x}) + \sum_{j=1}^{N}\mathbf{g}_j\boldsymbol{u}_j\right). \tag{30}$$

According to HJB (12), (31) holds

$$\begin{aligned}(\boldsymbol{\nabla}\boldsymbol{V}_i)^T\boldsymbol{f}(\boldsymbol{x}) = &-Q_i(\boldsymbol{x}) + \frac{1}{2}\boldsymbol{\nabla}\boldsymbol{V}_i{}^T\sum_{j=1}^{N}\mathbf{g}_j\boldsymbol{\Gamma}_j^{-1}\mathbf{g}_j^T\boldsymbol{\nabla}\boldsymbol{V}_j \\ &-\frac{1}{4}\boldsymbol{\nabla}\boldsymbol{V}_i{}^T\mathbf{g}_i\boldsymbol{\Gamma}_i^{-T}\mathbf{g}_i^T\boldsymbol{\nabla}\boldsymbol{V}_i.\end{aligned} \tag{31}$$

Substituting (31) into (30) leads to

$$\begin{aligned}\dot{V}_i(\boldsymbol{x}) = &\frac{1}{2}\mathbf{W}_i^T\boldsymbol{\nabla}\boldsymbol{\Phi}_i^T(\boldsymbol{x})\sum_{j=1}^{N}\mathbf{g}_j\boldsymbol{\Gamma}_j^{-1}\mathbf{g}_j^T\boldsymbol{\nabla}\boldsymbol{\Phi}_j(\boldsymbol{x})\tilde{\mathbf{W}}_j - Q_i(\boldsymbol{x}) \\ &-\frac{1}{4}\mathbf{W}_i^T\mathbf{D}_i(\boldsymbol{x})\mathbf{W}_i + o(\varepsilon_i, \boldsymbol{\nabla}^T\boldsymbol{\varepsilon}_j\boldsymbol{\nabla}\boldsymbol{\varepsilon}_k)\end{aligned} \tag{32}$$

where $\mathbf{D}_i(\boldsymbol{x}) = \boldsymbol{\nabla}\boldsymbol{\Phi}_i^T(\boldsymbol{x})\mathbf{g}_i\boldsymbol{\Gamma}_i^{-T}\mathbf{g}_i^T\boldsymbol{\nabla}\boldsymbol{\Phi}_i(\boldsymbol{x})$ is a $m\times m$ positive definite matrix and $o(\varepsilon_i, \boldsymbol{\nabla}^T\boldsymbol{\varepsilon}_j\boldsymbol{\nabla}\boldsymbol{\varepsilon}_k)$ is higher order infinitesimal of ε_i and $\boldsymbol{\nabla}^T\boldsymbol{\varepsilon}_j\boldsymbol{\nabla}\boldsymbol{\varepsilon}_k$ for $i,j,k \in \{1,2,\ldots,N\}$, which is guaranteed to be bounded by Lemma 1. Based on facts that $\mathbf{W}_i, \boldsymbol{\nabla}\boldsymbol{\Phi}_i(\boldsymbol{x}), \mathbf{g}_i$, and $\boldsymbol{\Gamma}_i^{-1}$ are all bounded and convergence of $\tilde{\mathbf{W}}_i$ proved by Step 1, $\mathbf{W}_i^T\boldsymbol{\nabla}\boldsymbol{\Phi}_i^T(\boldsymbol{x})\sum_{j=1}^{N}\mathbf{g}_j\boldsymbol{\Gamma}_j^{-1}\mathbf{g}_j^T\boldsymbol{\nabla}\boldsymbol{\Phi}_j(\boldsymbol{x})\tilde{\mathbf{W}}_j$ tends to 0 within $t_{\max} = \frac{2\sqrt{L(0)}}{\bar{\beta}}$. Suppose that boundary of $o(\varepsilon_i, \boldsymbol{\nabla}^T\boldsymbol{\varepsilon}_j\boldsymbol{\nabla}\boldsymbol{\varepsilon}_k)$ is $\alpha > 0$, then a conservative convergence set is estimated as $\Omega = \{\boldsymbol{x}|Q_i(\boldsymbol{x}) \le \alpha\}$. ■

B. Optimal Control Strategy Considering Control Torque Saturation

The control torque provided by microsatellites is quite limited, so the control torque saturation should be considered to design the optimal control for microsatellites. Suppose that the allowed torque range of each actuator for ith microsatellite is $[-\eta_i, \eta_i]$, where η_i is a positive constant. Define value function as

$$V_i(\boldsymbol{x}) = \int_t^{\infty}\left(Q_i(\boldsymbol{x}) + 2\int_0^{\boldsymbol{u}_i}\eta_i\mathbf{tanh}^{-T}(\boldsymbol{v}/\eta_i)\,\boldsymbol{\Gamma}_i d\boldsymbol{v}\right)d\tau \tag{33}$$

where $\mathbf{tanh}(\boldsymbol{s}) = [\tanh(s_1) \quad \tanh(s_2) \quad \tanh(s_3)]^T$ for $s = [s_1 \quad s_2 \quad s_3]^T$. The Hamilton function for value function as (33) can be derived as (34)

$$H_i = Q_i(\boldsymbol{x}) + \Gamma_i(\boldsymbol{u}_i) + (\nabla\boldsymbol{V}_i)^T\left(\boldsymbol{f}(\boldsymbol{x}) + \sum_{j=1}^{N}\mathbf{g}_j\boldsymbol{u}_j\right). \tag{34}$$

Differentiating H on $\boldsymbol{u}_i$, (35) can be obtained.

$$\frac{\partial H_i}{\partial \boldsymbol{u}_i} = 2\eta_i\mathbf{tanh}^{-T}(\boldsymbol{u}_i/\eta_i)\,\boldsymbol{\Gamma}_i + (\nabla\boldsymbol{V}_i)^T\mathbf{g}_i = \mathbf{0}. \tag{35}$$

Therefore, for value function (33), solve (35), optimal control strategy, and corresponding HJB equation for ith microsatellite

becomes (36) and (37), respectively.

$$\boldsymbol{u}_i = \eta_i \mathbf{tanh}\left(-\frac{1}{2\eta_i}\boldsymbol{\Gamma}_i^{-1}\mathbf{g}_i^T\boldsymbol{\nabla}\boldsymbol{V}_i\right) \tag{36}$$

$$\begin{aligned}(\boldsymbol{\nabla}\boldsymbol{V}_i)^T&\left[\boldsymbol{f}(\boldsymbol{x})-\eta_i\sum_{j=1}^{N}\mathbf{g}_j\mathbf{tanh}\left(\frac{1}{2\eta_i}\boldsymbol{\Gamma}_j^{-1}\mathbf{g}_j^T\boldsymbol{\nabla}\boldsymbol{V}_j\right)\right]\\&+\int_0^{-\eta_i\mathbf{tanh}\left(\frac{1}{2\eta_i}\boldsymbol{\Gamma}_i^{-1}\mathbf{g}_i^T\boldsymbol{\nabla}\boldsymbol{V}_i\right)}\mathbf{tanh}^{-T}\left(\boldsymbol{v}/\eta_i\right)\boldsymbol{\Gamma}_i d\boldsymbol{v}\\&+Q_i(\boldsymbol{x})=0.\end{aligned} \tag{37}$$

Remark 3: $\int_0^{-\eta_i\mathbf{tanh}\left(\frac{1}{2\eta_i}\boldsymbol{\Gamma}_i^{-1}\mathbf{g}_i^T\boldsymbol{\nabla}\boldsymbol{V}_i\right)}\mathbf{tanh}^{-T}(\boldsymbol{v}/\eta_i)\boldsymbol{\Gamma}_i d\boldsymbol{v}$ is path integral which represents the integral value along a path from origin to the point $-\eta_i\mathbf{tanh}\left(\frac{1}{2\eta_i}\boldsymbol{\Gamma}_i^{-1}\mathbf{g}_i^T\boldsymbol{\nabla}\boldsymbol{V}_i\right)$ for vector field $\mathbf{tanh}^{-T}(\boldsymbol{v}/\eta_i)\boldsymbol{\Gamma}_i$. The integral value is unique for all kinds of paths, if and only if field $\mathbf{tanh}^{-T}(\boldsymbol{v}/\eta_i)\boldsymbol{\Gamma}_i$ is conservative. Therefore, $\boldsymbol{\Gamma}_i$ is chosen as diagonal matrix to guarantee that it is independent of integration pathways.

Similarly, VFA technique is employed to approximate the solution of PDE (37). Therefore, control law is designed as (38) and adaptive law are the same as (25)–(27) with different parameters $\kappa_i = -Q_i(\boldsymbol{x}) - \int_0^{\boldsymbol{u}_i}\eta_i\mathbf{tanh}^{-T}(\boldsymbol{v}/\eta_i)\boldsymbol{\Gamma}_i d\boldsymbol{v}$.

$$\boldsymbol{u}_i = -\eta_i\mathbf{tanh}\left(\frac{1}{2\eta_i}\boldsymbol{\Gamma}_i^{-1}\mathbf{g}_i^T\boldsymbol{\nabla}\boldsymbol{\Phi}_i(\boldsymbol{x})\hat{\mathbf{W}}_i\right). \tag{38}$$

Theorem 2: Consider the attitude takeover control system (6) with control law (38) considering control torque saturation and the adaptive law (25)–(27). All signals in the closed loop are globally ultimately bounded.

Proof: Similar to proof of Theorem 1, there are two steps. Step 1 is same as that of Theorem 1 and Step 2 is given as follows.

Choose any value function $V_i(\boldsymbol{x})$ as Lyapunov candidate function, then its derivate is the same as (30). According to HJB (37) considering input saturation, (39) holds

$$\begin{aligned}(\boldsymbol{\nabla}\boldsymbol{V}_i)^T\boldsymbol{f}(\boldsymbol{x}) &= \eta_i(\boldsymbol{\nabla}\boldsymbol{V}_i)^T\sum_{j=1}^{N}\mathbf{g}_j\mathbf{tanh}\left(\frac{1}{2\eta_i}\boldsymbol{\Gamma}_j^{-1}\mathbf{g}_j^T\boldsymbol{\nabla}\boldsymbol{V}_j\right)\\&-\int_0^{-\eta_i\mathbf{tanh}\left(\frac{1}{2\eta_i}\boldsymbol{\Gamma}_i^{-1}\mathbf{g}_i^T\boldsymbol{\nabla}\boldsymbol{V}_i\right)}\mathbf{tanh}^{-T}\left(\boldsymbol{v}/\eta_i\right)\boldsymbol{\Gamma}_i d\boldsymbol{v}\\&-Q_i(\boldsymbol{x}).\end{aligned} \tag{39}$$

Substituting (39) into (30) yields

$$\begin{aligned}\dot{V}_i(\boldsymbol{x}) &= \eta_i(\boldsymbol{\nabla}\boldsymbol{V}_i)^T\sum_{j=1}^{N}\mathbf{g}_j\mathbf{tanh}(\frac{1}{2\eta_i}\boldsymbol{\Gamma}_j^{-1}\mathbf{g}_j^T\boldsymbol{\nabla}\boldsymbol{V}_j)\\&-\int_0^{\boldsymbol{u}_i}\mathbf{tanh}^{-T}\left(\boldsymbol{v}/\eta_i\right)\boldsymbol{\Gamma}_i d\boldsymbol{v}-Q_i(\boldsymbol{x})\\&-\eta_i(\boldsymbol{\nabla}\boldsymbol{V}_i)^T\sum_{j=1}^{N}\mathbf{g}_j\mathbf{tanh}\left(\tfrac{1}{2\eta_i}\boldsymbol{\Gamma}_j^{-1}\mathbf{g}_j^T(\boldsymbol{\nabla}\boldsymbol{\Phi}_j(\boldsymbol{x})\hat{\mathbf{W}}_j)\right)\\&=\eta_i(\boldsymbol{\nabla}\boldsymbol{V}_i)^T\sum_{j=1}^{N}\mathbf{g}_j\left(\mathbf{tanh}\left(\boldsymbol{\alpha}_1\right)-\mathbf{tanh}\left(\boldsymbol{\alpha}_2\right)\right)\end{aligned}$$

$$\begin{aligned}&-\int_0^{\boldsymbol{u}_i}\mathbf{tanh}^{-T}(\boldsymbol{v}/\eta_i)\,\boldsymbol{\Gamma}_i d\boldsymbol{v}-Q_i(\boldsymbol{x})\\&=\eta_i\mathbf{W}_i^T\boldsymbol{\nabla}\boldsymbol{\Phi}_i^T(\boldsymbol{x})\sum_{j=1}^{N}\mathbf{g}_j\left(\mathbf{tanh}\left(\boldsymbol{\alpha}_1\right)-\mathbf{tanh}\left(\boldsymbol{\alpha}_2\right)\right)\\&-\int_0^{\boldsymbol{u}_i}\mathbf{tanh}^{-T}\left(\boldsymbol{v}/\eta_i\right)\boldsymbol{\Gamma}_i d\boldsymbol{v}-Q_i(\boldsymbol{x})\\&+o(\varepsilon_i,\boldsymbol{\nabla}^T\boldsymbol{\varepsilon}_j\boldsymbol{\nabla}\boldsymbol{\varepsilon}_k)\end{aligned} \tag{40}$$

where $\boldsymbol{\alpha}_1 = \boldsymbol{\Gamma}_j^{-1}\mathbf{g}_j^T\boldsymbol{\nabla}\boldsymbol{V}_j/(2\eta_i)$ and $\boldsymbol{\alpha}_2 = \boldsymbol{\Gamma}_j^{-1}\mathbf{g}_j^T(\boldsymbol{\nabla}\boldsymbol{\Phi}_j(\boldsymbol{x})\hat{\mathbf{W}}_j)/(2\eta_i)$. Given that $\hat{\mathbf{W}}_i \to \mathbf{W}_i$ for i from 1 to N in finite time, the first term tends to 0, leading to the same conservative convergence set as that in Theorem 1. Therefore, all signals in closed loop are globally ultimately bounded. ■

Remark 4: Generally speaking, globally asymptotically stability can be guaranteed under the optimal control law. However, the attitude control strategies proposed in this article are not strictly optimal laws due to unavailability of analytically precise solution of a series of nonlinear coupled HJB equations. Therefore, the bounded stable state tracking errors under proposed control strategies are sourced from weights learning process which approximates the optimal solution.

V. Discussion

A. Persistent Excitation

To obtain optimal control torques from HJB equation, the core is to identify the weights $\mathbf{W}_i$ from (14). In every time point, (14) can be rewritten as least-square problem by converting (14) to (15), where $\boldsymbol{\psi}_i = \boldsymbol{\nabla}\boldsymbol{\Phi}_i(\boldsymbol{x})\dot{\boldsymbol{x}}$ might be close to $\mathbf{0}$ at stable period. The statement is made through inspecting $V(\boldsymbol{x})$, which plays the role of Lyapunov function. According to optimal form derived in (11), optimal control strategy is a kind of gradient-decent method to diminish $V(\boldsymbol{x})$. Therefore, when error of state $\boldsymbol{x}$ tends to zero, $\boldsymbol{\nabla}\boldsymbol{V}_i(\boldsymbol{x})$ must be zero. Had it been not zero, control torque $\boldsymbol{u}$ continues to drive the system (6) which contradicts the premise of stable state.

Based on the property of $\boldsymbol{\nabla}\boldsymbol{V}_i(\boldsymbol{x}) = \mathbf{0}$ at the point of $\boldsymbol{x} = \mathbf{0}$, the criterion for choosing basis function is to meet boundary conditions of HJB equation, i.e., $\boldsymbol{\nabla}\boldsymbol{\Phi}_i(\mathbf{0}) = \mathbf{0}$. In this situation, $\boldsymbol{\psi}_i$ becomes ill and thus might fail to provide information of identification. That brings the long-standing issue of identifying parameters, persistent excitation problem. Concurrent learning employed in this article considers both past and current data in different priority presented as (16), which increases the robustness of original identification since it still works even if system goes into stable state or error goes into tiny region around origin.

B. Saturation

Controller under actuator saturation designed as (36) and (38) is optimal in terms of index (41)

$$L_i = \int_0^{\infty}\left(Q_i(\boldsymbol{x}) + 2\int_0^{\boldsymbol{u}_i}\eta_i\mathbf{tanh}^{-T}\left(\boldsymbol{v}/\eta_i\right)\boldsymbol{\Gamma}_i d\boldsymbol{v}\right)d\tau \tag{41}$$

whose corresponding value function is defined as (33). Different from case of unsaturated situation, second part of objective function is chosen as infinite integral form $\int_0^{\boldsymbol{u}_i} \eta_i \tanh^{-T}(\boldsymbol{v}/\eta_i)\boldsymbol{\Gamma}_i d\boldsymbol{v}$ to limit the optimal torque $\boldsymbol{u}_i$ within the actuator capability bounds.

Original quadratic index as $\boldsymbol{u}_i^T\boldsymbol{\Gamma}_i\boldsymbol{u}_i$ can be written as the form of $\int_0^{\boldsymbol{u}_i}\boldsymbol{\Gamma}_i\boldsymbol{v}d\boldsymbol{v}$, whose gradient to $\boldsymbol{u}_i$ is linearly dependent on the torque $\boldsymbol{u}_i$. To curb the optimal torques solved from HJB equation into bounded region, inverse function of tanh is chosen as objective function. From physical view, when torques approach to bound, the penalty rises sharply. That is the mechanism of the antisaturation approach proposed in this article.

In essence, when saturation occurs, limited actuator torques are insufficient to implement the expected optimal command. Therefore, it is impossible to reach the original optimal value using bounded actuators. It is worth studying how to design an alternative objective index containing both control errors and energy use. An instance using objective index such as (41) can meet the requirement and be applied to the task.

C. Existence and Reachability of Nash Equilibrium

According to definition of Nash equilibrium (8), optimal strategy u_i for each player i drives the system into Nash equilibrium given the strategies of other players. Optimal control laws (11) and (36) are solved from minimizing Hamilton function (10) and (34) by the use of Pontryagin's principle which is necessary condition for Nash equilibrium. Together with concavity of $Q_i(\boldsymbol{x}) + \Gamma_i(\boldsymbol{u}_i)$, which can be satisfied in this article due to quadratic form of $Q_i(\boldsymbol{x})$ and $\Gamma_i(\boldsymbol{u}_i)$ or integral form of $\Gamma_i(\boldsymbol{u}_i)$ as (33), optimal laws (11) and (36) are sufficient to be strategies of Nash equilibrium [31].

However, in practical calculation, there might be a gap between real index L_i and theoretical target L_i^* due to inaccessibility of analytical solution $\nabla \boldsymbol{V}_i(\boldsymbol{x})$ in HJB equation.

According to (30), by integrating both two sides, the quadratic term of estimation error of weights L_w converges to zero in parabolic descent as (42)

$$L_w = \frac{1}{4}\left(t - 2L_{w0}^{\frac{1}{2}}\right)^2 \tag{42}$$

where L_{w0} denotes initial L_w. From (42), a conclusion can be drawn that the error of weights estimation $\tilde{\mathbf{W}}_i$ is bounded in the whole period.

Based on bounded attitude error and attitude velocity error $\boldsymbol{x}$ which is proved by Theorem 2, and the conclusion drawn above that $\tilde{\mathbf{W}}_i$ is bounded for all t, the bound of the gap between L_i and L_i^* is limited. Therefore, by the proposed approach, a bounded neighbour region around Nash equilibrium can be reached.

VI. Simulation

In this section, simulations are carried out to validate effectiveness of the proposed control law. The number of microsatellites providing control torques is assumed to be 3. The parameters of controller and initial conditions are set as Table II.

Case 1: Actuators working normally without considering saturation.

TABLE II
PARAMETERS OF SIMULATION

Name	Value
Initial attitude quaternion	$\boldsymbol{q}_0 = \begin{bmatrix} 0.9999 & 0.0086 & 0.0086 & 0.0086 \end{bmatrix}^T$
Initial angular velocity	$\boldsymbol{\omega}_0 = \begin{bmatrix} 0.3 & 0.1 & 0.3 \end{bmatrix}^T (\mathrm{deg}/s)$
Desired attitude trajectory	$\boldsymbol{q}_{d0} = [1\ 0\ 0\ 0]^T$ $\boldsymbol{\omega}_{d0} = 0.5[\sin(t/135)\ \sin(t/135)\ \sin(t/135)]^T (\mathrm{deg}/s)$
Controller parameters	$\mathbf{R}_1 = \mathbf{R}_2 = \mathbf{R}_3 = 0.02\mathbf{I}_{3\times3}, \mathbf{Q}_1 = 2\mathbf{I}_{6\times6}, \mathbf{Q}_2 = 3\mathbf{I}_{6\times6}, \mathbf{Q}_3 = 4\mathbf{I}_{6\times6}$
Inertia matrix	$\mathbf{J} = \begin{bmatrix} 350 & 0 & 10 \\ 0 & 320 & 0 \\ 10 & 0 & 370 \end{bmatrix} (kgm^2)$
Initial weights	$\hat{\mathbf{W}}_{10} = [400]_{1\times9}, \hat{\mathbf{W}}_{20} = [400]_{1\times9}, \hat{\mathbf{W}}_{30} = [400]_{1\times9}$
Basis function	$\boldsymbol{\Phi}(\boldsymbol{x}) = \tanh[\frac{1}{2}e_{v1}^2\ \ \frac{1}{2}e_{v2}^2\ \ \frac{1}{2}e_{v3}^2\ \ \frac{1}{2}\omega_{e1}^2\ \ \frac{1}{2}\omega_{e2}^2\ \ \frac{1}{2}\omega_{e3}^2\ \ e_{v1}\omega_{e1}\ \ e_{v1}\omega_{e1}\ \ e_{v1}\omega_{e1}]^T$
Adaptive parameters	$\mathbf{P}_0 = \mathbf{0}_{9\times9}, \mathbf{S}_0 = \mathbf{0}_{9\times1}$
Upper bound of actuator capability	$\eta_1 = \eta_2 = \eta_3 = 0.5Nm$

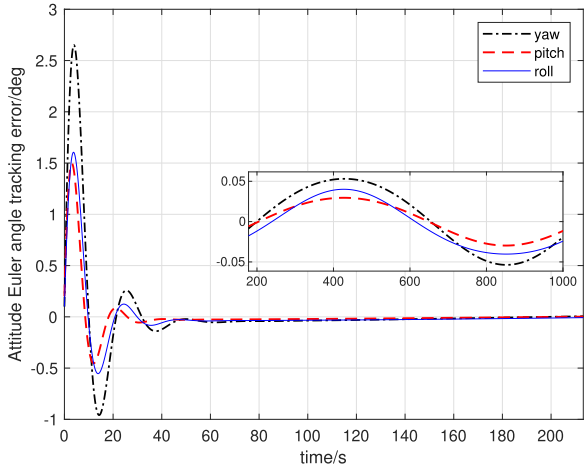

Fig. 4. Attitude Euler angle tracking error.

Figs. 4–8 show the simulation results by the proposed distributed control strategy in the case that all three microsatellites are all working normally without considering saturation. For ease of interpretation, attitude errors are described by Euler angles converted from quaternion. Figs. 4 and 5 show the attitude Euler angle and angular velocity tracking errors, respectively. It can be seen from Figs. 4 and 5 that attitude Euler angle errors and attitude angular velocity errors both converge within about 60 s. The steady state errors of Euler angle and angular velocity are of the order of 0.1° and 10^{-4} deg/s, respectively. Fig. 6 shows the torque command of first microsatellite. Due to limitation of space, the torque command of the other two microsatellites are omitted. Each torque is a vector with three components and the maximum value is no more than 0.8 Nm. In the distributed computation based on differential game strategy,

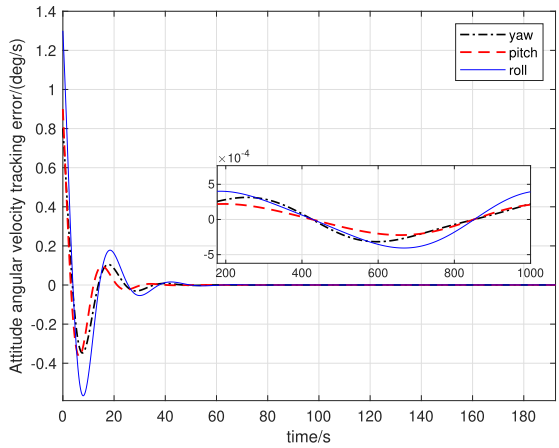

Fig. 5. Attitude Euler angular velocity tracking error.

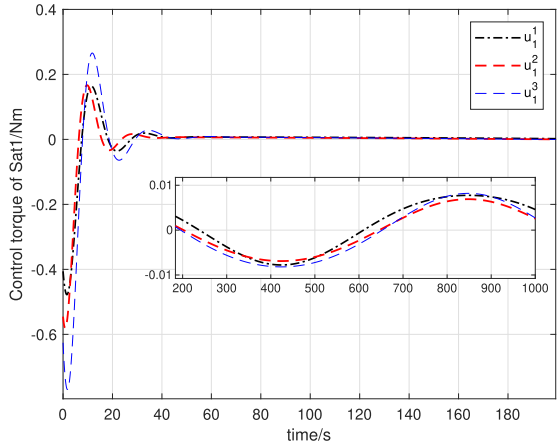

Fig. 6. Control torque of microsatellite 1.

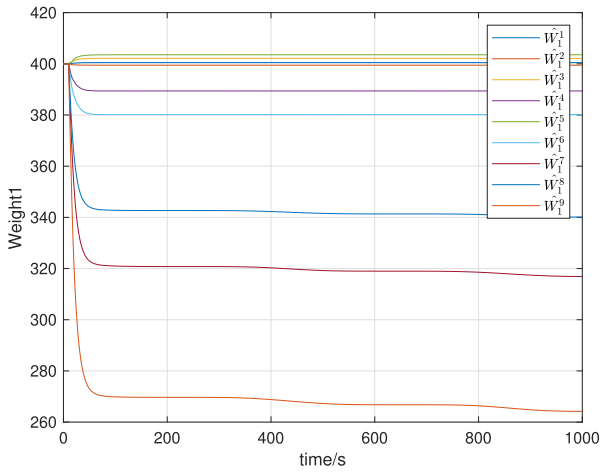

Fig. 7. Weights1 $\hat{\mathbf{W}}_1$ learning process.

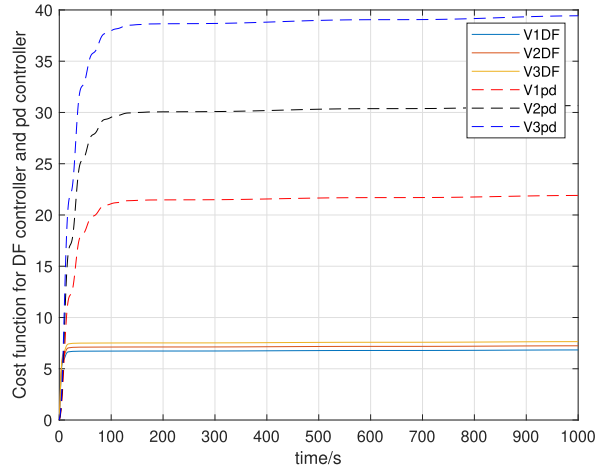

Fig. 8. Value function increasing curve for the proposed control scheme.

neural network is used to approximately solve the coupled nonlinear partial differential HJB equation, and the weights of the learning process are shown as Fig. 7 (only take weights in controller of first microsatellite as example). Applying the proposed updating law, three satellites' weights learning process are completed within about 100 s, which validates the function of finite time convergence.

Moreover, in order to reflect the optimization performance of distributed computation strategy based on differential game, it is compared with the traditional feedback linearization control strategy based on PD control in terms of index function. The results are shown in Fig. 8, where ViDF denotes cost function of ith ($i = 1, 2, 3$) microsatellite under differential game controller and ViPD denotes that under PD controller. When the control strategy of differential game is applied, the final cost function values of the three satellites are all lower than 10, while cost function values of the three satellites under PD-based control strategy are all higher than 20, which demonstrates that the proposed distributed control strategy based on differential game has higher tracking accuracy and lower energy consumption.

Case 2: Actuators working normally considering saturation.

Figs. 9–12 show the control effect of the proposed distributed control strategy considering input saturation. The maximum torque of all actuators on all microsatellites are set to be 0.5 Nm. Figs. 9 and 10 show that the Euler angles tracking errors and angular velocity errors vary as time goes on. The final steady state errors are within 0.05° and 5×10^{-4} deg/s, respectively, about half of that in case 1, while the convergence time of both kind of errors are about 60 s which are the same as that in case 1. Short duration of torque taking upper bound is the main reason for similar convergence time. Fig. 11 shows control torques provided by the first microsatellite. It can be seen from Fig. 11 that the control torques are bounded within the maximum capability of actuators and stay at boundary values in the initial few seconds, which indicates that actual control torques are lower than the values required to stabilize system rapidly enough. Fig. 12 shows weights learning process of neural network in the first microsatellite's controller.

Case 3: Failure of microsatellite 3 without saturation.

Figs. 13–18 show robustness of the proposed distributed control strategy against failure of one microsatellite. Figs. 13 and 14 show the Euler angles tracking errors and angular velocity errors convergence curve. At time 400 s when satellite 3 fails, it can be seen from figures that vibrations occur and then fade rapidly.

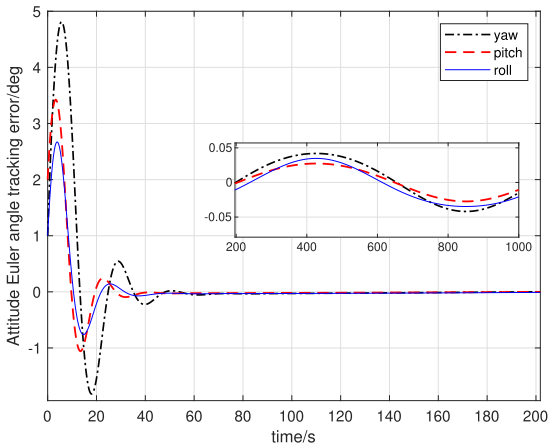

Fig. 9. Attitude Euler angles tracking errors under input saturation.

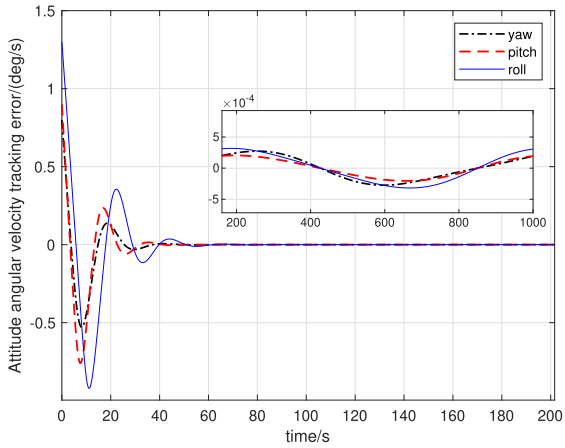

Fig. 10. Attitude angular velocity tracking errors under input saturation.

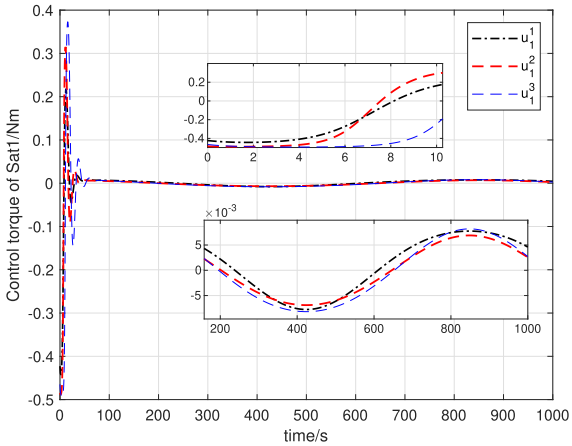

Fig. 11. Control torque of microsatellite 1 under input saturation.

The ultimate tracking accuracy recovers to the same level as before failure. Figs. 15–17 show the control torques provided by three microsatellites. The torques of microsatellite 3 become zero at 400 s and keep failing until the end, while the amplitudes

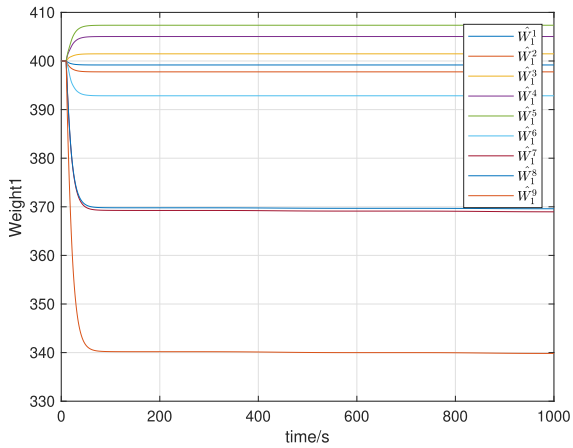

Fig. 12. Weights1 $\hat{\mathbf{W}}_1$ learning process under input saturation.

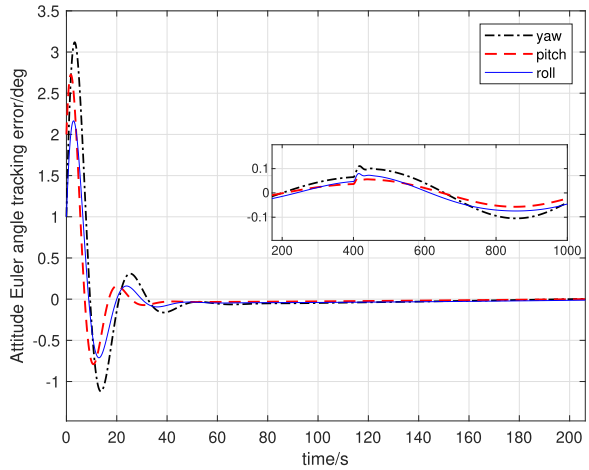

Fig. 13. Attitude Euler angles tracking errors under microsatellite 3 failing case.

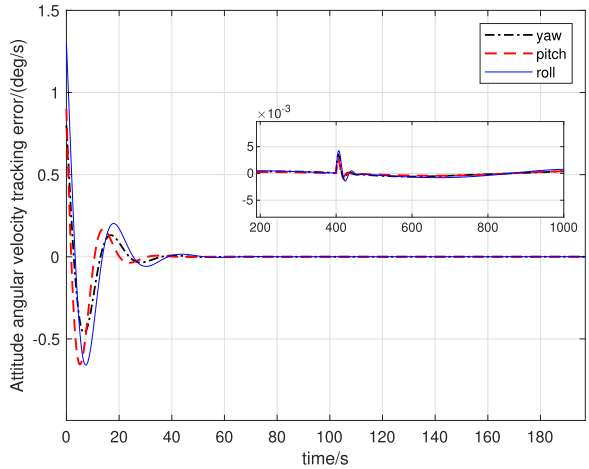

Fig. 14. Angular velocity tracking errors under microsatellite 3 failing case.

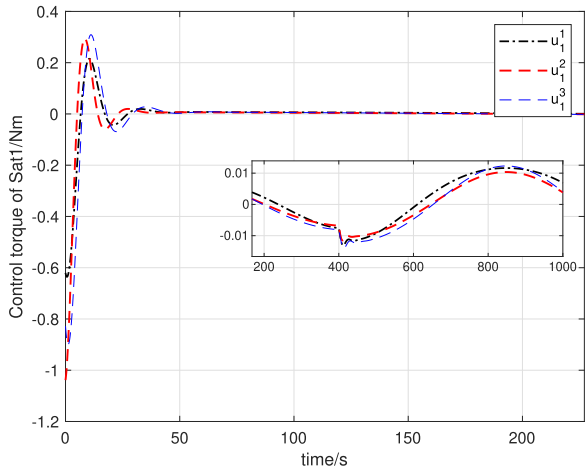

Fig. 15. Control torque of microsatellite 1 under microsatellite 3 failing case.

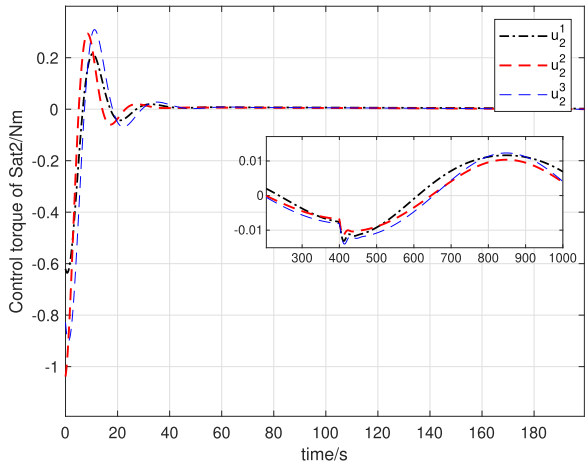

Fig. 16. Control torque of microsatellite 2 under microsatellite 3 failing case.

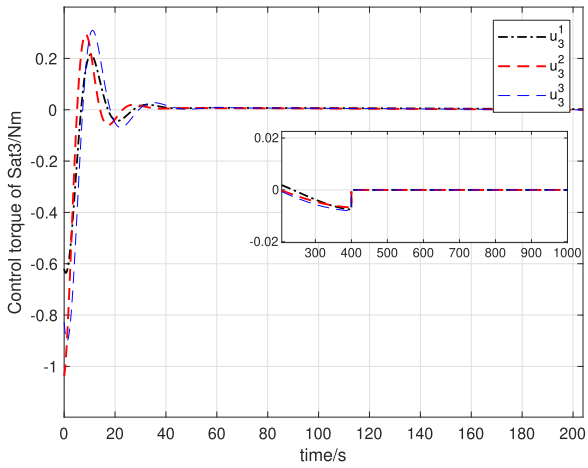

Fig. 17. Control torque of microsatellite 3 under microsatellite 3 failing case.

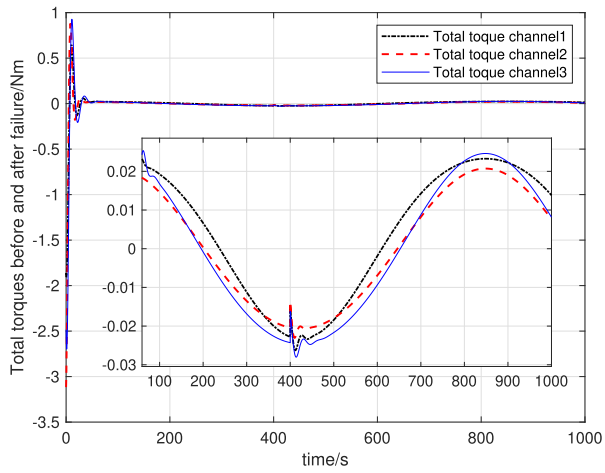

Fig. 18. Total torques of three microsatellites before and after failure.

of torques provided by microsatellite 1 and microsatellite 2 become larger than before so that the loss resulted from microsatellite 3's failure is compensated by the other two satellites. Fig. 18 shows the total torques in three channels of all microsatellites before and after failure. By observing peaks of curves before and after 400 s, the rank of absolute values of peak does not change due to failure happening on microsatellite 3 at 400 s. Moreover, values at peaks before and after 400 s keep flat, which is consistent with uniqueness that torques driving the same system should be the same.

The simulation of case 3 demonstrates that the system equipped with proposed control scheme owns robustness against abrupt failure of one microsatellite.

VII. Conclusion

In this article, an attitude takeover control strategy using microsatellites based on differential game is proposed. In the control scheme, numerically solving HJB equation by VFA is attributed to identification problem and concurrent learning technique is borrowed to loosen persistent excitation condition of system state. To get rid of information exchange among microsatellites, tracking differentiator is employed to approximate derivation of system state. Moreover, under case of input saturation, the optimized indices are fixed to get bounded optimal control strategies for each microsatellite. Then, stability analysis for cases of unconstrained and constrained input are demonstrated respectively through Lyapunov method. Finally, simulation results illustrate the superiority of proposed control strategy with respect to traditional PD control in terms of optimality and robustness. Accurate attitude tracking of the target spacecraft is still achieved even if one microsatellite fails.

References

[1] A. Flores Abad, O. Ma, K. Pham, and S. Ulrich, "A review of space robotics technologies for on-orbit servicing," *Prog. Aerosp. Sci.*, vol. 68, pp. 1–26, 2014.

[2] P. Huang, M. Wang, Z. Meng, F. Zhang, Z. Liu, and H. Chang, "Reconfigurable spacecraft attitude takeover control in post-capture of target by space manipulators," *J. Franklin Inst.*, vol. 353, no. 9, pp. 1985–2008, 2016.
[3] P. Huang, Y. Lu, M. Wang, Z. Meng, Y. Zhang, and F. Zhang, "Postcapture attitude takeover control of a partially failed spacecraft with parametric uncertainties," *IEEE Trans. Automat. Sci. Eng.*, vol. 16, no. 2, pp. 919–930, Apr. 2019.
[4] J. Luo, C. Wei, H. Dai, Z. Yin, X. Wei, and J. Yuan, "Robust inertia-free attitude takeover control of postcapture combined spacecraft with guaranteed prescribed performance," *ISA Trans.*, vol. 74, pp. 28–44, 2018.
[5] B. Wie and P. M. Barba, "Quaternion feedback for spacecraft large angle maneuvers," *J. Guidance, Control, Dyn.*, vol. 8, no. 3, pp. 360–365, 1985.
[6] S. Aslam, Y.-C. Chak, M. H. Jaffery, and R. Varatharajoo, "The fuzzy PD control for combined energy and attitude control system," *Aircr. Eng. Aerosp. Technol.*, vol. 94, no. 10, pp. 1806–1824, 2022.
[7] Y. Huang and Z. Meng, "Bearing-based distributed formation control of multiple vertical take-off and landing uavs," *IEEE Trans. Control Netw. Syst.*, vol. 8, no. 3, pp. 1281–1292, 2021.
[8] J. Huo, T. Meng, and Z. Jin, "Adaptive attitude control using neural network observer disturbance compensation technique," in *Proc. 9th Int. Conf. Recent Adv. Space Technol.*, 2019, pp. 697–701.
[9] X. Chen, W. Liang, H. Zhao, and A. Al Mamun, "Adaptive robust controller using intelligent uncertainty observer for mechanical systems under nonholonomic reference trajectories," *ISA Trans.*, vol. 122, pp. 79–87, 2022.
[10] Y. Huang and Z. Meng, "Global distributed attitude tracking control of multiple rigid bodies via quaternion-based hybrid feedback," *IEEE Trans. Control Netw. Syst.*, vol. 8, no. 1, pp. 367–378, Mar. 2021.
[11] C. K. Carrington and J. Junkins, "Optimal nonlinear feedback control for spacecraft attitude maneuvers," *J. Guidance, Control, Dyn.*, vol. 9, no. 1, pp. 99–107, 1986.
[12] M. G. Taul, C. Wu, S. F. Chou, and F. Blaabjerg, "Optimal controller design for transient stability enhancement of grid-following converters under weak-grid conditions," *IEEE Trans. Power Electron.*, vol. 36, no. 9, pp. 10251–10264, Sep. 2021.
[13] Y. Wu, Y. Zhang, and A. G. Wu, "Preassigned finite-time attitude control for spacecraft based on time-varying barrier Lyapunov functions," *Aerosp. Sci. Technol.*, vol. 108, 2021, Art. no. 106331.
[14] Q. Chen, Y. Ye, Z. Hu, J. Na, and S. Wang, "Finite-time approximation-free attitude control of quadrotors: Theory and experiments," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 57, no. 3, pp. 1780–1792, Jun. 2021.
[15] S. Feng, A. Cetinkaya, H. Ishii, P. Tesi, and C. De Persis, "Resilient quantized control under denial-of-service: Variable bit rate quantization," *Automatica*, vol. 141, 2022, Art. no. 110302.
[16] E. Aslmostafa, S. Ghaemi, M. A. Badamchizadeh, and A. R. Ghiasi, "Adaptive backstepping quantized control for a class of unknown nonlinear systems," *ISA Trans.*, vol. 125, pp. 146–155, 2022.
[17] M. Pouzesh and S. Mobayen, "Event-triggered fractional-order sliding mode control technique for stabilization of disturbed quadrotor unmanned aerial vehicles," *Aerosp. Sci. Technol.*, vol. 121, 2022, Art. no. 107337.
[18] S. Muhammad Amrr, A. Chakravarty, M. M. Alam, A. A. Algethami, and M. Nabi, "Efficient event-based adaptive sliding mode control for spacecraft attitude stabilization," *J. Guidance, Control, Dyn.*, vol. 45, no. 7, pp. 1328–1336, 2022.
[19] J. Yang, M. Xi, B. Jiang, J. Man, Q. Meng, and B. Li, "FADN: Fully connected attitude detection network based on industrial video," *IEEE Trans. Ind. Informat.*, vol. 17, no. 3, pp. 2011–2020, Mar. 2021.
[20] X. Cao, C. Yue, M. Liu, and B. Wu, "Time efficient spacecraft maneuver using constrained torque distribution," *Acta Astronautica*, vol. 123, pp. 320–329, 2016.
[21] H. Gui and A. H. de Ruiter, "Adaptive fault-tolerant spacecraft pose tracking with control allocation," *IEEE Trans. Control Syst. Technol.*, vol. 27, no. 2, pp. 479–494, Mar. 2019.
[22] H. Chang, P. Huang, Y. Zhang, Z. Meng, and Z. Liu, "Distributed control allocation for spacecraft attitude takeover control via cellular space robot," *J. Guidance, Control, Dyn.*, vol. 41, no. 11, pp. 2499–2506, 2018.
[23] X. Lang and A. de Ruiter, "A control allocation scheme for spacecraft attitude stabilization based on distributed average consensus," *Aerosp. Sci. Technol.*, vol. 106, 2020, Art. no. 106173.
[24] N. Han, J. Luo, and Y. Chai, "Differential game learning approach for multiple microsatellites takeover of the attitude movement of failed spacecraft," *Scientia Sinica Informationis*, vol. 50, no. 4, pp. 588–602, 2020.
[25] Y. Chai, J. Luo, N. Han, and J. Xie, "Linear quadratic differential game approach for attitude takeover control of failed spacecraft," *Acta Astronautica*, vol. 175, pp. 142–154, 2020.
[26] X. Lang and A. de Ruiter, "Non-cooperative differential game based output feedback control for spacecraft attitude regulation," *Acta Astronautica*, vol. 193, pp. 370–380, 2022.
[27] B. Dong, T. An, X. Zhu, Y. Li, and K. Liu, "Zero-sum game-based neuro-optimal control of modular robot manipulators with uncertain disturbance using critic only policy iteration," *Neurocomputing*, vol. 450, pp. 183–196, 2021.
[28] Y. Chai, J. Luo, N. Han, and J. Sun, "Robust event-triggered game-based attitude control for on-orbit assembly," *Aerosp. Sci. Technol.*, vol. 103, 2020, Art. no. 105894.
[29] H. Modares, F. L. Lewis, and M. B. Naghibi Sistani, "Integral reinforcement learning and experience replay for adaptive optimal control of partially-unknown constrained-input continuous-time systems," *Automatica*, vol. 50, no. 1, pp. 193–202, 2014.
[30] B. Z. Guo and Z. L. Zhao, "On convergence of tracking differentiator," *Int. J. Control*, vol. 84, no. 4, pp. 693–701, 2011.
[31] A. Seierstad and K. Sydsaeter, "Sufficient conditions in optimal control theory," *Int. Econ. Rev.*, pp. 367–391, 1977.

Baolin Wu (Member, IEEE) received the B.Eng. and M.Eng. degrees in spacecraft design from the Harbin Institute of Technology, Harbin, China, in 2003 and 2005, respectively, and the Ph.D. degree in spacecraft formation control from Nanyang Technological University, Singapore, in 2011.

He is currently a Professor with the Research Center of Satellite Technology, Harbin Institute of Technology. His current research interests include spacecraft attitude control, attitude synchronization, spacecraft formation control, and trajectory optimization.

Keyu Chen received the B.Eng. degree in aerospace engineering from Northwestern Polytechnical University, Xi'an, China, and the M.Eng. degree from the Harbin Institute of Technology, Harbin, China, in 2019 and 2021, respectively. He is currently working toward the Ph.D. degree in robotic localization with the Department of Mechanical Engineering, Hong Kong Polytechnic University, Hong Kong.

His current research interests include spacecraft attitude control, quantized control, prescribed performance control, robotic navigation, and SLAM technology.

Danwei Wang (Fellow, IEEE) received the B.Eng. degree from the South China University of Technology, Guangdong, China, in 1982, and the Ph.D. and MSE degrees from the University of Michigan, Ann Arbor, MI, USA, in 1989 and 1984, respectively.

He is currently a Professor with the School of Electrical and Electronic Engineering, Nanyang Technological University (NTU), Singapore. He is Director of the STE-NTU Corp Lab, NTU. From 2005 to 2011, he was the Head of the Division of Control and Instrumentation, NTU.

Yuxiang Sun (Member, IEEE) received the bachelor's degree from the Hefei University of Technology, Hefei, China, in 2009, the master's degree from the University of Science and Technology of China, Hefei, China, in 2012, and the Ph.D. degree from The Chinese University of Hong Kong, Shatin, Hong Kong, in 2017.

He is currently a Research Assistant Professor with the Department of Mechanical Engineering, The Hong Kong Polytechnic University, Hong Kong.