Stackelberg Differential Game-based Optimal Fault-tolerant Accommodation for Spacecraft

IEEE Publication Technology, *Staff, IEEE,*

***Abstract*—**

***Index Terms*—Article submission, IEEE, IEEEtran, journal, LaTeX, paper, template, typesetting.**

I. Introduction

THE higher spacecraft attitude and orbit control capability plays crucial role of guaranteeing the success of increasing demand for space missions, for instance, flying-around formation for spacecraft, on-orbit assembly of multiple flexible spacecraft [1], etc. However, the complex coupling characteristics and multiple mission objectives, deriving from twin formation flight and obstacle avoidance contained in the above spacecraft missions, make the existing control system sensitive to a fault, for instance, actuator fault, sensor fault, etc. Since an undetected incipient fault can result in a severe disaster, fault-tolerant control (FTC), being able to guarantee that the system control performance remains at an acceptable level after faults occur, forms a significant and challenging issue, especially how to achieve the goal of FTC while ensuring the optimal performance of minimum energy consumption [3], [4].

In general, the existing types of FTC are mainly divided into two categories: 1) passive and 2) active. The passive one resorts to the robust control concepts, and the active one integrates a basic control scheme (such as a sliding mode controller) and an FDO block, which can cope with the unknown influence of fault actively employing reconfiguring the controller with fault estimation (FDO), resulting in a stability and acceptable control performance [5]. Considering the active control of spacecraft missions in the preference of coupling fault, one natural idea is to employ the FTC and handle the fault as a local external disturbance to fulfill the global target. However, this bi-direction coupling characteristic between spacecraft system and faults makes it difficult for disturbance hypothesis control strategy to meet the increasing demand for control accuracy against complex space operating environments (such as external disturbance, gravity gradient, and others). For this challenge, game theory can effectively describe the information interaction among the spacecraft system, fault and the external environment, and the coupling characteristics, and then, solve the optimization of the FTC design in the spacecraft system from a global perspective, providing a new perspective to handle the challenges for the optimal FTC for spacecraft system [6].

Game theory is an interdisciplinary subject that can analyze the development trend of events with rigorous mathematical analysis [7], and is able to deal with multilateral conflicts, cooperation, and competition flexibly. The cooperative game is used to realize the FTC of four-wheel independent drive electric vehicle [8], and the four actuators of electric vehicle are treated as four players to play the game, seeking its Pareto solution, to ensure that the electric vehicle can remain stable when the actuator fault occurs. In the FTC of system formed by pursuit and escape game, when the pursuer and the escaper simultaneously have actuator failures, a near-optimal fault-tolerant controller in the individual decision layer, supported by a fault estimator, is proposed in [9], achieving that the pursuer can still catch the escaper. Based on the above analysis, the fault-tolerant game control is no longer only concerned with the solution and characteristic analysis of the game equilibrium point, but further on this basis, through the design of a strong control law to achieve effective intervention in the game process of each controller and effective adjustment of the equilibrium point, thus achieving the optimal decision of the individual and the desired state of the system, forming a new class of optimal fault-tolerant control methods.

Another practically significant issue that causes attention is the problem of Nash equilibrium seeking, which has been a thriving research topic in recent decades [10]. By employing LaSalle's invariance principle and adaptive control technology, the node-based and edge-based control laws are designed in [10] to seek the Nash equilibrium, which emphasizes the distributed characteristic of Nash equilibrium solution, needs the pseudo-gradient vector of the game being strong monotone. In addition, an optimal tracking problem is transformed into a Nash-equilibrium in the graphical game [11], and the Nash equilibrium seeking, minimizing the game cost functions, is solved by coping with a coupled Hamilton-Jacobi (HJ) equation with data-based adaptive dynamic programming (ADP) algorithm. Furthermore, when it comes to multi-player Markov games, a novel learning scheme implemented in a reinforcement-learning framework for Nash equilibrium is proposed in [12], which treats the evolution of player policies as a dynamic process. It evolves one's control strategy according to its current in-game performance and aggregation of its performance over history, achieving the global Nash equilibrium. It is the minimizing evaluation function working mechanism based on a critical neural network combined with an actor neural network that makes ADP and RL can dynamically seek the optimal solution to a game problem, achieving a good control effect under Nash equilibrium. However, how to accomplish the FTC goal and ensure optimal global performance

This paper was produced by the IEEE Publication Technology Group. They are in Piscataway, NJ.

Manuscript received April 19, 2021; revised August 16, 2021.

(for instance, minimum energy consumption) simultaneously distinguishes that of the common works is still a particularly tough task.

Inspired by the above observation, to establish an optimal active FTC for spacecraft system resorting to game theory, the key idea of this article is to establish a master-slave differential game and seek a Nash equilibrium solution with modified advantage actor-critic (A2C) RL. Compared with the previous FTC works, the innovation contributions are summarized as follows.

- An active FTC idea based on a master-slave differential game named the Stackelberg game is proposed for the spacecraft system. The bi-direction influence between the FTC and its corresponding FDO is deeply explored instead of one-direction estimation and compensation FTC method as usual, forming a non-cooperating sequential game with FTC (master) and FDO (slave), to obtain the optimal fault estimation and the optimal FTC.
- An auxiliary controller variable is artfully introduced, making the designed FDO able to interact with FTC sequentially to keep the process of the Stackelberg game. Different from the common practice in the relevant works, a novel performance index function is presented to embody the cross optimal indexes of FTC and FDO, evolving a novel A2C RL control method with less complexity. Two new adaptive updating laws of the critic weights are designed to balance fault-tolerant objectives and optimal control performance, achieving the Nash equilibrium in both theory and simulation.

This paper is structured as follows. Section II denotes the spacecraft model with fault, while Section III provides an overview of relevant preliminaries. Section IV is devoted to modeling a Stackelberg differential game, including an optimal FDO (slave) and FTC (leader), and demonstrates the Nash equilibrium under the ideal optimal FDO and FTC. In Section V, a novel A2C RL control method is proposed to address FTC problems as described in Section IV. At last, numerical simulation is presented in Section VI, to assess the effectiveness and demonstrate the advantages of the proposed control scheme, followed by conclusions in Section VII.

II. Preliminaries and Problem Formulation

A. Spacecraft with with fault modeling

A spacecraft system is taken into consideration, and its dynamic equation can be modeled by

$$\dot{\rho} = A_\rho \rho + B_\rho u + \varrho f \tag{1}$$

where $\rho = [x, y, z, \dot{x}, \dot{y}, \dot{z}]^T \in \mathbb{R}^6$ represents the relative position and velocity vector of a spacecraft in local vertical/local horizontal coordinate system. The system matrix $A_\rho \in \mathbb{R}^{6\times 6}$ and the system control matrix $B_\rho \in \mathbb{R}^{6\times 3}$ are delivered as follows.

$$A_\rho = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 3n^2 & 0 & 0 & 0 & 2n & 0 \\ 0 & 0 & 0 & -2n & 0 & 0 \\ 0 & 0 & -n^2 & 0 & 0 & 0 \end{bmatrix}, B_\rho = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ \frac{1}{m} & 0 & 0 \\ 0 & \frac{1}{m} & 0 \\ 0 & 0 & \frac{1}{m} \end{bmatrix}.$$

In addition, m denotes the mass of spacecraft, and $n = \sqrt{\mu / r_0^3}$ is the average orbital angle relative to the main spacecraft. μ stands for the gravitational constant of the earth, and r_0 represents the orbit radius of the main spacecraft. ϱ is a known coefficient constant matrix with appropriate dimensions.

In terms of the fault signal f, due to the complex operation environment of a spacecraft (such as gravity gradient, external disturbance and others), it is assumed to comply with the following dynamic [13]

$$\dot{x}_f = A_f x_f, f = C_f x_f, t \geq t_f \tag{2}$$

where x_f stands for fault-state of fault dynamic. The occurrence time of a fault of spacecraft is described as t_f. The matrix A_f is made in this work to guarantee that the stability, and C_f represents the output matrix with appropriate dimension. This fault model can cover various kinds of faults, for instance, actuator faults, sensor faults and process faults, etc.

B. Control objective

The control objective of this work is to put forward an adaptive FTC strategy, integrating the Stackelberg game and active FTC mechanism, to facilitate the flying-around motion of spacecraft, while ensuring two key objectives are achieved despite the presence of faults and external disturbances:

T1) Design a new active FTC method that utilizes a modified A2C approach to address the Nash equilibrium related to the Stackelberg game, thereby enabling the spacecraft state ρ to optimally track the desired reference signal ρ_d.

T2) Ensure that all signals within the closed-loop of the system remain bounded.

III. Preliminaries

In this section, some definitions and preliminary results are presented, being central to the control design and the closed-loop stability analysis.

Definition 1. (Best Response). The best response is the optimal strategy of a player, which takes account of the other players strategies to minimize their own cost function.

Definition 2. (Interactive Stackelberg Equilibrium). In a Stackelberg differential game, the leader and the follower have the corresponding cost functions and strategies: J_i and $u_i, \forall i = 1, 2$, respectively. A Stackelberg equilibrium strategy is obtained by means of the control pair $\{u_1^*, u_2^*(u_1^*)\}$ satisfying the following properties:

1) For each u_1, there exists $u_2^*(u_1)$ for the follower such that

$$J_2\left(u_1, {u_2}^*\left(u_1\right)\right) \leq J_2\left(u_1, u_2\right), \forall u_1 \tag{3}$$

2) There exists u_1^* on the best responses of the leader such that

$$J_1\left(u_1^*, u_2^*\left(u_1^*\right)\right) \leq J_1\left(u_1, u_2\left(u_1\right)\right) \tag{4}$$

for any pair $\left(u_1, u_2\left(u_1\right)\right)$ on the best responses of the leader.

IV. FDO/FTC Design in Framework of Stackelberg Differential Game

To attain T1, it is necessary to bridge the FDO and FTC to form Stackelberg differential game, ensuring the goal of optimal FTC on the flying-around motion of spacecraft. Then, the details are stated as follows.

A. Stackelberg differential game modeling

Recalling the dynamics of spacecraft (1) and fault dynamic (2), the variable $x = \left[\rho^T, x_f^T\right]^T$ is defined for the convenience of subsequent description, resulting in the following relevant surrogate dynamics

$$\left\{\begin{array}{l} \dot{x} = Ax + Bu \\ y = Cx \end{array}\right. \tag{5}$$

where the system matrix is remarked as $A = \left[\begin{array}{cc} A_\rho & \varrho f \\ 0 & A_f \end{array}\right]$. The surrogate system input matrix B is defined as $B = \left[\begin{array}{c} B_\rho \\ 0 \end{array}\right]$, so does the defined system output matrix $C = \left[\begin{array}{cc} C_\rho & 0 \end{array}\right]$.

To conquer the challenges for FTC design due to fault and external disturbance for the flying-around motion of spacecraft, an FDO is constructed as follows.

$$\left\{\begin{array}{l} \dot{\hat{x}} = A\hat{x} + Bu + \mathcal{H}v \\ \hat{y} = C\hat{x} \end{array}\right. \tag{6}$$

where $\hat{x}$ stands for the estimation of x, and $\hat{y}$ represents the output of the observer O. v is an auxiliary control compensation item and $\mathcal{H}$ denotes the system input matrix with the proper matrix dimension. It is noteworthy that v takes the crucial role in forming a non-cooperation game, Stackelberg differential game, between controller u and v, and drives a game system consisting of FDO and FTC to reach Nash equilibrium.

In terms of game modeling for FTC of spacecraft, the sequential decisions exist between the surrogate system (5) and the observer (6), which reflects in:

1) From the perspective of control strategy making, the circle stems from the working of spacecraft's initial control strategy FTC u_0.
2) Afterwards, the FDO starts work and generates the best response to match the current strategy FTC, for instance, making the optimal auxiliary controller v^* to get the minimum error of $z = y - \hat{y}$.
3) In turn, the FTC of spacecraft will adjust the control output by means of fault information via observer (6), forming the best response u^*. This completes one circle, and this circle keeps working until the FTC target is achieved.

Based on the aforementioned analyses, the first step in the control circle is executed by the FTC, followed by the observer. As a result, the Stackelberg differential game's crux is established in this study, with the leader utilizing FTC and the follower utilizing FDO. The specifics are demonstrated as follows.

B. Design of follower FDO

In this subsection, FDO acts as a follower in a Stackelberg differential game and serves to provide admissible control output for FTC. Due to the existing bi-directional coupling effect between FTC and its corresponding FDO, the cost function of (6) can be further expressed as

$$J_1(v) = \int_0^\infty \{z^TQz + v^TRv + u^TGv\}dt \tag{7}$$

where Q, R, and G stand for the symmetric positive matrices. With respect to the item u^TGv, it denotes the influence from spacecraft (5) to the observer FDO (6).

Recalling the characteristics of sequential decision-making between the system and FDO, the controller signal, denoted by u_0, initiates each decision cycle. The challenge facing the designed FDO (follower) lies in facilitating the implementation of an arbitrary admissible control signal, $u_0 \in \mathcal{U}$, which belongs to the FTC(leader). Specifically, this challenge of the follower can be described as follows:

$$V_1^{J_1} = \min_v J_1, \text{for any}, u \in \mathcal{U} \tag{8}$$

where $\mathcal{U}$ represents the admissible control set of FTC.

To overcome this challenge, the corresponding Hamiltonian function is made as

$$\begin{aligned} H_1 = {} & z^TQz + v^TRv + u^TGv \\ & + \nabla J_1^T(A\hat{x} + Bu + \mathcal{H}v), \end{aligned} \tag{9}$$

where $\nabla J_1 = \partial J_1 / \partial \hat{x}$ is the partial derivative of J_1 with respect to the estimated state $\hat{x}$.

In what follows, owing to $\nabla \dot{J}_1 = -\frac{\partial H_1}{\partial \hat{x}}$, it leads to

$$\nabla \dot{J}_1 = -2C^TQC\hat{x} + 2C^TQy - A^T\nabla J_1 \tag{10}$$

Taking the derivative with respect to $\frac{\partial H_1}{\partial v} = 0$ and utilizing the minimal principle, we have

$$v^* = -1/2R^{-1}\left(\mathcal{H}^T\nabla J_1^* + Gu\right) \tag{11}$$

To pursue the optimal value of the cost function J_1^*, namely, seeking $0 = \min_v H_1\left(\hat{x}, \nabla J_1^*, v\right)$, combined with (11), the Hamilton-Jacobi (HJ) equation can be deduced as

$$\begin{aligned} & z^TQz + v^{*T}Rv^* + u^TGv^* \\ & \quad + \left(\nabla J_1^*\right)^T\left(A\hat{x} + Bu + \mathcal{H}v^*\right) = 0 \end{aligned} \tag{12}$$

with $J_1^*(0) = 0$.

Theorem 1: By incorporating the designed observer (6) and leveraging the auxiliary control v from (11), the observer's objective of $\lim_{t\to\infty}(x - \hat{x}) = 0$ is attained.

Proof: By selecting the designed J_1^* as the Lyapunov function and incorporating (12) and (11), the time derivative of J_1^* can be derived as follows:

$$\begin{aligned}&\dot{J}_1^* = \tfrac{\partial J_1^*}{\partial \hat{x}}\dot{\hat{x}}\\ &= -z^TQz + \tfrac{1}{4}u^TGR^{-1}Gu - \tfrac{1}{4}(\nabla J_1^*)^T\mathcal{H}R^{-1}\mathcal{H}^T\nabla J_1^*\end{aligned} \tag{13}$$

When the inequality $z^TQz \geq \frac{1}{4}u^TGR^{-1}Gu$ holds, the following inequality

$$\dot{J}_1^* \leq -\tfrac{1}{4}(\nabla J_1^*)^T\mathcal{H}R^{-1}\mathcal{H}^T\nabla J_1^* \tag{14}$$

is maintained. It is noticed that u is belonging to the admissible control set, which describes the leader controller's function of stabilizing the system. Thus, we have $u \to 0$ as $t \to \infty$, and then, it infers that $(y(t) - \hat{y}(t)) \to 0$ as $t \to \infty$. In addition, the observability conditions in Assumption 1 and [29] lead to the deduction that $\lim\limits_{t\to\infty}(x - \hat{x})$ tends to 0. This completes the proof of Theorem 1.

C. Design of leader FTC

In this section, an optimal FTC is proposed to solve the leader's problem of the Stackelberg differential game, taking into account the optimal fault estimation goal achieved by the FDO as presented in Theorem 1, making the best response to the FDO's(follower) current strategy. To this end, the cost function for the FTC is elaborated as follows:

$$J_2(u) = \int_0^\infty \left(x^T\Gamma x + u^T\Psi u + \nabla J_1^T\Xi\nabla J_1\right)dt \tag{15}$$

where Γ, Ψ and Ξ are the symmetric positive matrices.

Similar to (8), the challenge of the leader (FTC) encountered can be described as

$$V_2^{J_2} = \min_u J_2(u, v^*) \tag{16}$$

The objective of the FTC is to minimize the cost function (15) subject to the constraints of spacecraft system (5) and ∇J_1 ((7)). The optimal auxiliary FDO's controller v^*, the unique influencing factor for ∇J_1, is added herein for non-cooperating sequential game purpose because it allows the leader(FTC) to adjust its controller in the way of bi-direction instead of one-direction estimation and compensation FTC method as usual, to imultaneously achieve optimal FDO and FTC performance. Thus, the Hamiltonian function of the leader can be constructed as follows:

$$\begin{aligned}H_2 \triangleq\ & x^T\Gamma x + u^T\Psi u + \nabla J_1^T\Xi\nabla J_1 + \nabla J_2^T(Ax + Bu)\\ &+\beta^T\left(-2C^TQC\hat{x} + 2C^TQy - A^T\nabla J_1\right)\end{aligned} \tag{17}$$

where $\nabla J_2 \triangleq \partial J_2/\partial x$ and $\beta \triangleq \partial J_2/\partial(\nabla J_1)$.

By means of $\nabla\dot{J}_2 = -2\Gamma x - A^T\nabla J_2 - 2C^TQC\beta$ and $\dot{\beta} = -\frac{\partial H_2}{\partial(\nabla J_1)} = A^T\beta - 2\Xi\nabla J_1$, we have the best response of FTC as follows

$$u^* = -\frac{1}{2}\Psi^{-1}B^T\nabla J_2^* \tag{18}$$

Furthermore, the Hamilton-Jacobi equation with respect to the leader can be deduced by recalling the optimal control function v^*, and we have

$$\begin{aligned}&x^T\Gamma x + u^{*T}\Psi u^* + \nabla J_1^{*T}\Xi\nabla J_1^* + \nabla J_2^{*T}(Ax + Bu^*)\\ &+ \beta^T\left(-2C^TQC\hat{x} + 2C^TQy - A^T\nabla J_1^*\right) = 0\end{aligned} \tag{19}$$

with $J_2^*(0) = 0$.

Similarly, applying the substitution of equation (18) into equation (12), one derives that

$$\begin{aligned}&z^TQz - \frac{1}{4}(\nabla J_1^*)^T\mathcal{H}R^{-1}\mathcal{H}^T\nabla J_1^* - \frac{1}{2}(\nabla J_1^*)^TB\Psi^{-1}B^T\nabla J_2^*\\ &- \frac{1}{16}(\nabla J_2^*)^TB\Psi^{-1}GR^{-1}G\Psi^{-1}B^T\nabla J_2^*\\ &+ \frac{1}{4}(\nabla J_2^*)^TB\Psi^{-1}GR^{-1}\mathcal{H}^T\nabla J_1^* + (\nabla J_1^*)^TA\hat{x} = 0\end{aligned} \tag{20}$$

Theorem 2: Under Assumption 1, the faulty system (5), subject to cost function (15), can be guaranteed to be asymptotically stable by the optimal fault-tolerant controller u^* designed in (18). Moreover, if the coupled Hamilton-Jacobi equations in (19) and (20) are satisfied by the optimal costs J_1 and J_2, the Nash equilibrium is reached by the pair $\{u^*, v^*(u^*)\}$.

Proof: 1) With J_2^* chosen as the Lyapunov function and the coupled Hamilton-Jacobi equation in (19) used, the derivative of J_2^* with respect to time can be obtained as

$$\begin{aligned}\dot{J}_2^* &= \left(\tfrac{\partial J_2^*}{\partial x}\right)\dot{x} + \left(\tfrac{\partial J_2^*}{\partial(\nabla J_2)}\right)\nabla\dot{J}_2\\ &= -x^T\Gamma x - u^{*T}\Psi u^* - (\nabla J_1^*)^T\Xi\nabla J_1^*\end{aligned} \tag{21}$$

Afterwards, due to $\Gamma > 0$, $\Psi > 0$, and $\Xi > 0$, it can be inferred that $\dot{J}_2^* \leq 0$, implying that the faulty system (5) achieves asymptotic stability when u^* is used. Thus, it can be drawn that $J_2(x(\infty), \nabla J_1(\infty)) = 0$.

2) To prove Nash equilibrium, the performance index (7) of the follower (FDO) is expressed as

$$\begin{aligned}J_1(\hat{x}(0), y(0), v) &= \int_0^\infty \{z^TQz + v^TRv + u^TGv\}dt\\ &+ \int_0^\infty \dot{J}_1 dt + J_1(\hat{x}(0), y(0)) - J_1(\hat{x}(\infty), y(\infty))\end{aligned} \tag{22}$$

Then, using the expression $\dot{J}_1 = \nabla J_1^T(A\hat{x} + Bu + \mathcal{H}v)$ and the fact that $-2v^TRv = \left(\mathcal{H}^T\nabla J_1^+Gu\right)^Tv$ with v^* in (11), the value of $J_1(\hat{x}(0), y(0), v)$ can be deduced by completing the square method, which results in

$$\begin{aligned}&J_1(\hat{x}(0), y(0), v)\\ &= \textstyle\int_0^\infty \left\{z^TQz + (v - v^*)^TR(v - v^*) - v^{*T}Rv^*\right\}dt\\ &+ \textstyle\int_0^\infty \nabla J_1^T(A\hat{x} + Bu)dt\\ &+ J_1(\hat{x}(0), y(0)) - J_1(\hat{x}(\infty), y(\infty))\end{aligned} \tag{23}$$

Combined with (12) and $-v^{*T}Rv^* = -2v^{*T}Rv^* + v^{*T}Rv^* = \nabla J_1^{*T}\mathcal{H}v^* + u^TG^Tv^* + v^{*T}Rv^*$, we have

$$\begin{aligned}J_1(\hat{x}(0), y(0), v) &= \textstyle\int_0^\infty \left\{(v - v^*)^TR(v - v^*)\right\}dt\\ &+ J_1(\hat{x}(0), y(0)) - J_1(\hat{x}(\infty), y(\infty))\end{aligned} \tag{24}$$

According to the proof of Theorem 1, J_1 converges to zero as $t \to \infty$,i.e., $J_1(\hat{x}(\infty), y(\infty)) = 0$. Thus, we have

$$\begin{aligned}&J_1(\hat{x}(0), y(0), v)\\ &= \textstyle\int_0^\infty \left\{(v - v^*)^TR(v - v^*)\right\}dt + J_1(\hat{x}(0), y(0))\end{aligned} \tag{25}$$

It is noticed that v^* minimizes the performance function of the follower (7) against FTC policy $u \in \mathcal{U}$. By setting $v = v^*$, it yields

$$J_1(v^*) = J_1(\hat{x}(0), y(0)) \tag{26}$$

Since the inequality (26) $\leq$ (25) holds, it follows that for a given value of v,

$$J_1(u, v^*) \leq J_1(u, v) \tag{27}$$

As a result, given that the aforementioned conditions are satisfied for the follower (FDO) in the game, the Nash equilibrium stated in equation (3) is attained.

In regard to the Nash equilibrium of the leader (FTC), with a focus on its performance index J_2 (15), we can leverage the fact that $-2u^T\Psi u = \nabla J_2^T Bu$, where u^* is specified in (18). By using v^* in (11) and employing the completing the square method, we can further deduce $J_2(x(0), \nabla J_1(0), u)$ as

$$\begin{aligned} & J_2(x(0), \nabla J_1(0), u) \\ = & \int_0^\infty \Big(x^T \Gamma x + (u - u^*)^T \Psi (u - u^*) \Big) dt \\ & + \int_0^\infty \left(\nabla J_1^{*T} \Xi \nabla J_1^* - u^{*T} \Psi u^* \right) dt \\ & + \int_0^\infty \nabla J_2^{*T} (Ax) dt + J_2(x(0), \nabla J_1(0)) \\ & + \int_0^\infty \beta^T \left(-2C^T QC\hat{x} + 2C^T Qy - A^T \nabla J_1^* \right) dt \\ & - J_2(x(\infty), \nabla J_1(\infty)) \end{aligned} \tag{28}$$

By substituting $-u^{*T}\Psi u^* = -2u^{*T}\Psi u^* + u^{*T}\Psi u^* = \nabla J_2^{*T} Bu^* + u^{*T}\Psi u^*$ into (28), we can make further deductions as follows

$$J_2(u, v^*) = \int_0^\infty \Big((u - u^*)^T \Psi (u - u^*) \Big) dt + J_2(x(0), \nabla J_1(0)) \tag{29}$$

Letting $u = u^*$ leads to the optimal value, i.e.,

$$J_2(u^*, v^*) = J_2(x(0), \nabla J_1(0)) \tag{30}$$

Because of (30) $\leq$ (29), we obtain that for given u

$$J_2(u^*, v^*) \leq J_2(u, v^*) \tag{31}$$

Hence, the Nash equilibrium formulated in equation for the leader (FTC) is ultimately realized.

After several iterations of best response between FDO and FTC, a balance point is reached, resulting in the formation of the Stackelberg equilibrium $\{u^*, v^*(u^*)\}$. At this equilibrium, the minimum cost J_1^* is achieved by FDO, while optimal control u^* with the minimum J_2^* is obtained by FTC. As a result, the current best condition is sustained by both FDO and FTC, and neither is willing to leave the equilibrium.

V. STACKELBERG DIFFERENTIAL GAME STRATEGY BASED ON RL

In this section, the Stackelberg differential game strategies are implemented online through a novel A2C RL control method, and two new adaptive updating laws of the critic weights are devised to balance fault-tolerant objectives and optimal control performance. The details will be further investigated in the subsequent control derivation.

Before proceeding further, two critic networks are designed to approximate the cost functions J_1^* and J_2^* for the system and observer, respectively.

$$J_1^* = W_{c1}^{*T} \varphi_{c1}(\bar{x}) + \varepsilon_{c1}(\bar{x}) \quad \rightarrow \text{FDO} \tag{32}$$

$$J_2^* = W_{c2}^{*T} \varphi_{c2}(x) + \varepsilon_{c2}(x) \rightarrow \text{FTC} \tag{33}$$

where the critic weights $W_{ci}^* \in \mathbb{R}^{l_i}, i = 1, 2$ are set as the ideal ones to approximate the cost functions J_1 and J_2 for the system and observer, respectively, where $\bar{x} = [\hat{x}\ y]^T$. Here, φ_{ci} is the actuation function, and ε_{ci} represents the approximation error.

For the sake of clarity in subsequent exposition, we shall use φ_{c1}, ε_{c1}, φ_{c2}, and ε_{c2} as shorthand notations for $\varphi_{c1}(\bar{x})$, $\varepsilon_{c1}(\bar{x})$, $\varphi_{c2}(x)$, and $\varepsilon_{c2}(x)$, respectively. Moreover, upon differentiating J_1^* and J_2^*, the corresponding gradient can be achieve as:

$$\nabla J_1^* = \nabla \varphi_{c1}^T W_{c1}^* + \nabla \varepsilon_{c1} \tag{34}$$

$$\nabla J_2^* = \nabla \varphi_{c2}^T W_{c2}^* + \nabla \varepsilon_{c2} \tag{35}$$

where $\nabla J_1^* = \partial J_1^* / \partial \hat{x}$, $\nabla J_2^* = \partial J_2^* / \partial \hat{x}$, $\nabla \varphi_{c1} = \partial \varphi_{c1} / \partial \hat{x}$, $\nabla \varphi_{c2} = \partial \varphi_{c2} / \partial \hat{x}$, $\nabla \varepsilon_{c1} = \partial \varepsilon_{c1} / \partial \hat{x}, \nabla \varepsilon_{c2} = \partial \varepsilon_{c2} / \partial \hat{x}$.

Subsequently, the inherent complexity and nonlinearity of the HJ equations (12) and (20) may lead to the curse of dimensionality, thereby rendering it challenging or even infeasible to obtain solutions that ensure optimal fault-tolerant objectives. To surmount this challenge, two subsections will expound on the design process of fault detection and isolation (FDO) and fault-tolerant control (FTC) utilizing the proposed A2C reinforcement learning (RL) method. This method employs two neural network-based approximators, namely the critic and actor neural networks, and seeks to achieve an Interactive Stackelberg equilibrium.

A. Approximation for the follower (FDO) via proposed A2C RL

1) Critic NN design for follower: Recalling the Hamiltonian function equation (9) with the optimal manner with v^* and u^*, it is revealed that

$$\begin{aligned} H_1 = & z^T Qz + v^{*T} Rv^* + u^{*T} Gv^* \\ & + \nabla J_1^{*T} (A\hat{x} + Bu^* + \mathcal{H}v^*) \end{aligned} \tag{36}$$

Proceeding further, the error of the Hamilton-Jacobi-Bellman (HJB) equation is deduced by substituting equations (11), (18), and (34) into equation (36):

$$\Im_{HJB} = W_{c1}^{*T} \sigma_1^* + z^T Qz + \frac{1}{4} W_{c1}^{*T} \Pi_1 W_{c1}^* - \frac{1}{16} W_{c2}^{*T} \Pi_2 W_{c2}^* \tag{37}$$

where

$$\begin{aligned} \sigma_1^* &= \nabla \varphi_{c1} A\hat{x} - \frac{1}{2} \Pi_3 W_{c2}^* - \frac{1}{2} \Pi_1 W_{c1}^* + \frac{1}{4} \Pi_4 W_{c2}^* \\ \Pi_1 &= \nabla \varphi_{c1} \mathcal{H} R^{-1} \mathcal{H}^T \nabla \varphi_{c1}^T \\ \Pi_2 &= \nabla \varphi_{c2} B \Psi^{-T} G R^{-1} G \Psi^{-1} B^T \nabla \varphi_{c2}^T \\ \Pi_3 &= \nabla \varphi_{c1} B \Psi^{-1} B^T \nabla \varphi_{c2}^T \\ \Pi_4 &= \nabla \varphi_{c1} \mathcal{H} R^{-1} G \Psi^{-1} B^T \nabla \varphi_{c2}^T \end{aligned} \tag{38}$$

Typically, the optimal weights W_{ci}^* for the critic neural network are unknown, and therefore an estimate $\hat{J}_i$ of the cost function $J_i^*, i = 1, 2$ is employed in the following manner:

$$\hat{J}_1 = \hat{W}_{c1}^T \varphi_{c1}(\bar{x}), \nabla \hat{J}_1 = \nabla \varphi_{c1}^T \hat{W}_{c1} \tag{39}$$

$$\hat{J}_2 = \hat{W}_{c2}^T \varphi_{c2}(\bar{x}), \nabla \hat{J}_2 = \nabla \varphi_{c2}^T \hat{W}_{c2} \tag{40}$$

where $\hat{W}_{ci}$ stands for the estimation of W_{ci}^*, and $\nabla \hat{J}_i$ is applied to estimate the gradient of J_i^*.

To minimize the $\Im_{HJB}$ of the follower and enable the critic NN to approach the optimal value function J_1^* by adjusting the corresponding NN weights ($\hat{W}_{c1}$), we consider the following objective function:

$$\mathcal{E}_1 = \frac{1}{2} \Im_{HJB}^T \Im_{HJB} \tag{41}$$

And then, by utilizing the gradient descent scheme $\dot{\hat{W}}_{c1} = \partial \mathcal{E}_1 \Big/ \partial \hat{W}_{c1}$, the update law for the critic NN weight of the follower $\hat{W}_{c1}$ is presented as follows:

$$\begin{aligned}\dot{\hat{W}}_{c1} = -\alpha_{c1} \frac{\sigma_{a1}}{\left(1 + \sigma_{a1}^T \sigma_{a1}\right)^2} \Bigg[& z^T Q z + \frac{1}{4} \hat{W}_{a1}^T \Pi_1 \hat{W}_{a1} \\ & + \hat{W}_{c1}^T \sigma_{a1} - \frac{1}{16} \hat{W}_{a2}^T \Pi_2 \hat{W}_{a2} \Bigg]\end{aligned} \tag{42}$$

where $\alpha_{c1} > 0$ denotes a critic learning rate to be designed. The variable σ_{a1} is introduced as an intermediate step to simplify the notation and the conciseness of this work

$$\sigma_{a1} = \nabla \varphi_{c1} A \hat{x} - \tfrac{1}{2} \Pi_3 \hat{W}_{a2} - \tfrac{1}{2} \Pi_1 \hat{W}_{a1} + \tfrac{1}{4} \Pi_4 \hat{W}_{a2}.$$

where the specifics regarding $\hat{W}_{ai}, i = 1, 2$ will be explicated in the ensuing subsection, which concerns the design of the actor NN.

Remark 1. In contrast to prevailing practices within the relevant literature, the cross-optimal indexes, which embody the characteristics of bidirectional coupling and sequential decision-making of FTC and FDO, present significant challenges in obtaining the optimal fault-tolerant controller using a single critic updating law as is typically done.

2) Actor NN design for follower: To remedy the challenges arising from Remark 1 and the unknown gradient of the cost function $J_i^*, i = 1, 2$, inspired of the Actor-critic framework of RL, and an actor neural network (NN) is typically utilized to implement the FTC policy, resulting in the following control policies.

$$\hat{v} = -1/2 R^{-1} \left(\mathcal{H}^T \nabla \varphi_{c1}^T \hat{W}_{a1} + G \hat{u} \right) \tag{43}$$

$$\hat{u} = -\frac{1}{2} \Psi^{-1} B^T \nabla \varphi_{c2}^T \hat{W}_{a2} \tag{44}$$

To mitigate the issue of high variance stemming from overestimation in the AC framework, especially in the context of sequential decision-making, we propose to utilize the advantage function as follows:

$$\delta_t^{a1} = \int_{t-T}^{t} \Im_c(s) ds \tag{45}$$

In this context, the following Bellman error equation is considered for any time integral T:

$$\Im_c(t) = \int\limits_{t-T}^{t} \left(z^T Q z + \hat{v}^T R \hat{v} + \hat{u}^T G \hat{v} \right) ds + \hat{W}_{c1}^T \Delta \varphi_{c1}^T \tag{46}$$

where $\Delta \varphi_{c1} = \varphi_{c1}(z(t)) - \varphi_{c1}(z(t - T))$.

It is noteworthy that the Advantage function is employed to assess the advantage value of taking a particular action in each state compared to the average return, thereby effectively enhancing the efficiency and stability of reinforcement learning through variance reduction. The operational process of the Advantage function of the FDO is described as follows:

1) If $\delta_t^{a1} < 0$, it infers that the current control input for the follower (FDO) is superior to the average level, and consequently, the actor NN is updated in the direction of the control policy, as prescribed by the RL algorithm.

2) If $\delta_t^{a1} > 0$, the current control input for the FDO is indicated to be worse than the average level, and consequently, the actor NN is updated in the direction opposite to the control policy, complying with the reinforcement learning algorithm.

3) If $\delta_t^{a1} = 0$, the control input for the FDO is indicated to be optimal, and as a result, no updates will be made to the actor NN.

Consequently, the update law of $\dot{\hat{W}}_{a1}$ by means of δ_t^{a1} is designed as follows:

$$\begin{aligned}\dot{\hat{W}}_{a1} = & -\alpha_{a1} \left[k_{a1} + \frac{\Pi_1 \hat{W}_{a1} \hat{W}_{a1}^T \Pi_1}{16 \left(1 + \sigma_{a1}^T \sigma_{a1}\right)^2} \right] \hat{W}_{a1} \\ & + \alpha_{a1} \frac{\delta_t^{a1} \Pi_1 \hat{W}_{a1}}{4T^2 \left(1 + \sigma_{a1}^T \sigma_{a1}\right)^2} + \alpha_{a1} \frac{\Pi_1 \hat{W}_{a1} \hat{W}_{a2}^T \Pi_2 \hat{W}_{a2}}{64 \left(1 + \sigma_{a1}^T \sigma_{a1}\right)^2}\end{aligned} \tag{47}$$

where $\alpha_{a1} > 0$ is the learning rate, and k_{a1} is a positive parameter to be designed. This completes the design of the follower.

B. Approximation for the leader (FTC) via proposed A2C RL

This section proposes using a Stackelberg differential game approach to solve the spacecraft's FTC problem. The proposed approach seeks to determine the most effective solution to the FTC problem by modeling it as a game between two players - the leader (FTC) and the follower (FDO) - and analyzing the optimal strategies and decisions of each player under different conditions and scenarios. By utilizing this Stackelberg differential game framework, the section provides a comprehensive and practical solution to the FTC with modified A2C RL strategy.

1) Critic NN Design For Leader: Following a similar procedure to the follower case (37), the HJB equation's error for the leader (FTC) is deduced by substituting (18), (34), and (35) into (20):

$$\begin{aligned}\wp_{HJB} = & x^T \Gamma x + \frac{1}{4} W_{c2}^{*T} D W_{c2}^* + W_{c2}^{*T} \sigma_2^* + W_{c1}^{*T} \Re W_{c1}^* \\ & - \beta^T A^T \nabla \varphi_{c1}^T W_{c1}^* + \beta^T \left(-2C^T Q C \hat{x} + 2C^T Q y \right)\end{aligned} \tag{48}$$

where $\sigma_2^* = \nabla \varphi_{c2}(Ax - \frac{1}{2} D W_{c2}^*)$, $D = \nabla \varphi_{c2} B \Psi^{-T} B^T \nabla \varphi_{c2}^T$, $\Re = \nabla \varphi_{c1} \Xi \nabla \varphi_{c1}^T$.

To continue, our aim is to minimize the leader's $\Im_{HJB}$ by adjusting the critic NN weights to identify the index function. This will enable the critic NN $\hat{J}_2$ to approach the optimal value function J_2^*. To accomplish this objective, we define the objective function as follows:

$$\mathcal{E}_2 = \frac{1}{2} \wp_{HJB}^T \wp_{HJB} \tag{49}$$

Thus, taking into account the sequential decision-making of FTC and FDO, combined with the gradient descent method, the adaptive weight update law for $\hat{W}_{c2}$ is proposed as follows:

$$\begin{aligned} \dot{\hat{W}}_{c2} = -\alpha_{c2} \frac{\sigma_{a2}}{\left(1 + \sigma_{a2}^T \sigma_{a2}\right)^2} & \Big[x^T \Gamma x + \frac{1}{4} \hat{W}_{a2}^T D \hat{W}_{a2} \\ & + \hat{W}_{c2}^T \sigma_{a2} + \hat{W}_{a1}^T \Re \hat{W}_{a1} - \beta^T A^T \nabla \varphi_{c1}^T \hat{W}_{a1} \\ & + \beta^T \left(-2C^T Q C \hat{x} + 2C^T Q y \right) \Big] \end{aligned} \tag{50}$$

where $\alpha_{c2} > 0$ is a designed parameter to adjust learning rate of $\hat{W}_{c2}$. $\sigma_{a2} = \nabla \varphi_{c2}(Ax - \frac{1}{2} D \hat{W}_{a2})$ is an introduced intermediate variable that makes the derivation of the article concise and clear.

2) Actor NN design for leader(FTC): Similar to the design process of the follower's action NN, the design of the leader's action NN is as follows

$$\hat{u} = -\frac{1}{2} \Psi^{-1} B^T \nabla \varphi_{c2}^T \hat{W}_{a2} \tag{51}$$

To tackle the issue of high variance resulting from overestimation in the AC framework, and the bidirectional coupling influence between the FDO and the FTC during sequential decision-making, we propose utilizing the advantage function in the following manner:

$$\begin{aligned} \delta_t^{a2} = \textstyle\int_{t-T}^t \Big(\int_{t-T}^t & \left(x^T \Gamma x + \hat{u}^T \Psi \hat{u} + \nabla \hat{J}_1^T \Xi \nabla \hat{J}_1 \right) ds \\ & + \hat{W}_{c2}^T \Delta \varphi_{c2}^T \Big) \Big) ds \end{aligned} \tag{52}$$

where $\Delta \varphi_{c2} = \varphi_{c2}(t) - \varphi_{c2}(t - T)$.

Similar to the design of weight update rule for the actor network of the follower, the weight update rule for the action network of the leader, affected by the follower, is designed as follows.

$$\begin{aligned} \dot{\hat{W}}_{a2} = -\alpha_{a2} & \left[k_{a2} + \frac{D \hat{W}_{a2} \hat{W}_{a2}^T D}{16\left(1 + \sigma_{a2}^T \sigma_{a2}\right)^2} \right] \hat{W}_{a2} \\ & - \alpha_{a2} \frac{D \hat{W}_{a2} \hat{W}_{c1}^T \Re \hat{W}_{c1}}{4\left(1 + \sigma_{a1}^T \sigma_{a1}\right)^2} + \alpha_{a2} \frac{\delta_t^{a2} D \hat{W}_{a2}}{4T^2 \left(1 + \sigma_{a2}^T \sigma_{a2}\right)^2} \\ & - \alpha_{a2} \frac{D \hat{W}_{a2}}{4T^2 \left(1 + \sigma_{a2}^T \sigma_{a2}\right)^2} \beta^T \\ & \times \left(A^T \nabla \varphi_{c1}^T \hat{W}_{a1} - 2C^T Q y + 2C^T Q C \hat{x} \right) \end{aligned} \tag{53}$$

C. Stability analysis

In this subsection, the main result of this work is ready to be presented, as it is detailed in the following theorem.

Theorem 3. Under Assumptions 1-2, by utilizing the approximate optimal auxiliary controller (43), along with the update laws for critic NN (42) and actor NN (47), and the approximate optimal FTC (51), featuring adaptive weight updating laws for critic NN (42) and actor NN (53), a set of desirable properties can be achieved.

1) The trajectory tracking error is guaranteed to be uniformly bounded.

2) By minimizing the value function, the states of the faulty spacecraft system (5), the states of the observer (6), and the estimation errors of the critic and actor weights can all be uniformly and ultimately bounded.

Proof: For a detailed proof, please refer to the Appendix.

VI. Other Resources

See [?], [?], [?], [?], [?] for resources on formatting math into text and additional help in working with LaTeX.

VII. Algorithms

Algorithms should be numbered and include a short title. They are set off from the text with rules above and below the title and after the last line.

Algorithm 1 Weighted Tanimoto ELM.

```
TRAIN($\mathbf{XT}$)
    select randomly $W \subset \mathbf{X}$
    $N_{\mathbf{t}} \leftarrow |\{i : \mathbf{t}_i = \mathbf{t}\}|$  for  $\mathbf{t} = -1, +1$
    $B_i \leftarrow \sqrt{\textsc{max}(N_{-1}, N_{+1}) / N_{\mathbf{t}_i}}$  for  $i = 1, ..., N$
    $\hat{\mathbf{H}} \leftarrow B \cdot (\mathbf{X}^T \mathbf{W}) / (\nVdash \mathbf{X} + \nVdash \mathbf{W} - \mathbf{X}^T \mathbf{W})$
    $\beta \leftarrow \left( I/C + \hat{\mathbf{H}}^T \hat{\mathbf{H}} \right)^{-1} (\hat{\mathbf{H}}^T B \cdot \mathbf{T})$
    return $\mathbf{W}, \beta$

PREDICT($\mathbf{X}$)
    $\mathbf{H} \leftarrow (\mathbf{X}^T \mathbf{W}) / (\nVdash \mathbf{X} + \nVdash \mathbf{W} - \mathbf{X}^T \mathbf{W})$
    return SIGN($\mathbf{H}\beta$)
```

Que sunt eum lam eos si dic to estist, culluptium quid qui nestrum nobis reiumquiatur minimus minctem. Ro moluptat fuga. Itatquiam ut laborpo rersped exceres vollandi repudaerem. Ulparci sunt, qui doluptaqis sumquia ndestiu sapient iorepella sunti veribus. Ro moluptat fuga. Itatquiam ut laborpo rersped exceres vollandi repudaerem.

VIII. Conclusion

The conclusion goes here.

Acknowledgments

This should be a simple paragraph before the References to thank those individuals and institutions who have supported your work on this article.

Appendix
Proof of Theorem 4

A. Approximation for the follower (FDO) via proposed A2C RL

To clarify the proof process, taking the following Lyapunov function $V_{J_1} = V_1^{J_1} + \tilde{W}_{c1}^T \alpha_{c1}^{-1} \tilde{W}_{c1} + \tilde{W}_{a1}^T \alpha_{a1}^{-1} \tilde{W}_{a1}$, and then, the function $\dot{V}_{J_1}$ will be divided into three parts ($\dot{V}_1^{J_1}$; $\dot{V}_{c_1}^{J_1}$; $\dot{V}_{a_1}^{J_1}$) and each part will be explained separately.

$$\dot{V}_{J_1} = \dot{V}_1^{J_1} - \underbrace{\tilde{W}_{c1}^T \alpha_{c1}^{-1} \dot{\hat{W}}_{c1}}_{\dot{V}_{c_1}^{J_1}} - \underbrace{\tilde{W}_{a1}^T \alpha_{a1}^{-1} \dot{\hat{W}}_{a1}}_{\dot{V}_{a_1}^{J_1}} \tag{54}$$

where $\tilde{W}_{c1} = W_{c1}^* - \hat{W}_{c1}$, and $\tilde{W}_{a1} = W_{a1}^* - \hat{W}_{a1}$.

Step 1: With respect to $\dot{V}_1^{J_1}$, keep (8), (32) and (34) in mind, the derivative of $\dot{V}_1^{J_1}$ yields

$$\begin{aligned} \dot{V}_1^{J_1} = W_{c1}^{*T} \nabla\varphi_{c1} \left(A\hat{x} + B\hat{u} + \mathcal{H}\hat{v}\right) \\ + \nabla\varepsilon_{c1}^* \left(A\hat{x} + B\hat{u} + \mathcal{H}\hat{v}\right) \end{aligned} \tag{55}$$

where $\aleph_{J_1} = \nabla\varepsilon_{c1}^* \left(A\hat{x} + B\hat{u} + \mathcal{H}\hat{v}\right) \leq b_{\varepsilon_{c1}} b_{\mathcal{H}} \|z\| \left(\|W_{c1}^*\| + \left\|\tilde{W}_{a1}\right\|\right) + b_{\varepsilon_{c1}} b_A \|z\| + b_{\varepsilon_{c1}} \left(b_B + b_{\mathcal{H}} b_G\right) \|z\| \left(\|W_{c2}^*\| + \left\|\tilde{W}_{a2}\right\|\right)$.
Furthermore, combined with $\Im_{HJB} = z^T Q z + v^{*T} R v^* + u^{*T} G v^* + \nabla J_1^{*T}(A\hat{x} + Bu^* + \mathcal{H}v^*)$, we have

$$\begin{aligned} \dot{V}_1^{J_1} &= \Im_{HJB} + \aleph_{J_1} - z^T Q z - v^{*T} R v^* - u^{*T} G v^* \\ &\quad - W_{c1}^{*T} \nabla\varphi_{c1}(z) B u^* + W_{c1}^{*T} \nabla\varphi_{c1}(z) B\hat{u} \\ &\quad - W_{c1}^{*T} \nabla\varphi_{c1}(z) \mathcal{H} v^* + W_{c1}^{*T} \nabla\varphi_{c1}(z) \mathcal{H}\hat{v} \\ &= \Im_{HJB} + \aleph_{J_1} - z^T Q z - \frac{1}{4} W_{c1}^{*T} \Pi_1 W_{c1}^* + \frac{1}{2} W_{c1}^{*T} \Pi_1 \tilde{W}_{a1} \\ &\quad + \frac{1}{16} W_{c2}^{*T} \Pi_2 W_{c2}^{*T} + \frac{1}{4} W_{c1}^{*T} \Pi_3 \tilde{W}_{a2} \end{aligned} \tag{56}$$

Step 2. When it comes to $V_{c_1}^{J_1}$, we demonstrate that

$$\begin{aligned} \dot{V}_{c_1}^{J_1} = \frac{\tilde{W}_{c1}^T \sigma_{a1}}{\left(1 + \sigma_{a1}^T \sigma_{a1}\right)^2} \Bigg[& \frac{1}{4} \hat{W}_{a1}^T \Pi_1 \ \hat{W}_{a1} + \hat{W}_{c1}^T \sigma_{a1} W_{c1}^{*T} \sigma_1^* \\ & + \Im_{HJB} - \frac{1}{16} \hat{W}_{a2}^T \Pi_2 \hat{W}_{a2} \\ & + \frac{1}{4} W_{c1}^{*T} \Pi_1 W_{c1}^* + \frac{1}{16} W_{c2}^{*T} \Pi_2 W_{c2}^* \Bigg] \end{aligned} \tag{57}$$

Subsequently, the deduction of (57) can be further derived as follows.

$$\begin{aligned} \dot{V}_{c_1}^{J_1} &= \frac{\tilde{W}_{c1}^T \sigma_{a1}}{\left(1 + \sigma_{a1}^T \sigma_{a1}\right)^2} \left[-\tilde{W}_{c1}^T \sigma_{a1} + \Im_{HJB}\right] \qquad (58) \\ &+ \frac{\tilde{W}_{c1}^T \sigma_{a1}}{\left(1 + \sigma_{a1}^T \sigma_{a1}\right)^2} \left[\frac{1}{2} W_{c1}^{*T} \Pi_3 \tilde{W}_{c2} - \frac{1}{4} W_{c1}^{*T} \Pi_4 \tilde{W}_{c2}\right] \\ &+ \frac{\tilde{W}_{c1}^T \sigma_{a1}}{\left(1 + \sigma_{a1}^T \sigma_{a1}\right)^2} \left[\frac{1}{16} W_{c2}^{*T} \Pi_2 \tilde{W}_{c2} - \frac{1}{16} \hat{W}_{a2}^T \Pi_2 \hat{W}_{a2}\right] \\ &+ \frac{1}{4} \frac{\hat{W}_{c1}^T \sigma_{a1}}{\left(1 + \sigma_{a1}^T \sigma_{a1}\right)^2} \tilde{W}_{a1}^T \Pi_1 \hat{W}_{a1} + \frac{1}{4} \frac{\tilde{W}_{c1}^T \sigma_{a1}}{\left(1 + \sigma_{a1}^T \sigma_{a1}\right)^2} \tilde{W}_{a1}^T \Pi_1 W_{c1}^* \\ &+ \frac{1}{4} \frac{W_{c1}^{*T} \sigma_{a1}}{\left(1 + \sigma_{a1}^T \sigma_{a1}\right)^2} \tilde{W}_{a1}^T \Pi_1 \tilde{W}_{a1} - \frac{1}{4} \frac{W_{c1}^{*T} \sigma_{a1}}{\left(1 + \sigma_{a1}^T \sigma_{a1}\right)^2} \tilde{W}_{a1}^T \Pi_1 W_{c1}^* \end{aligned}$$

where the establishment of (58) is based on the following equation transformations and simplifications:

$$\begin{aligned} &\star \hat{W}_{c1}^T \sigma_{a1} - W_{c1}^{*T} \sigma_1^* \\ &\quad = -\tilde{W}_{c1}^T \nabla\varphi_{c1}(z) A\hat{x} - \frac{1}{2} \hat{W}_{c1}^T \Pi_1 \hat{W}_{a1} + \frac{1}{2} W_{c1}^{*T} \Pi_1 W_{c1}^* \\ &\qquad - \frac{1}{2} \hat{W}_{c1}^T \Pi_3 \hat{W}_{a2} + \frac{1}{2} W_{c1}^{*T} \Pi_3 W_{c2}^* \\ &\qquad + \frac{1}{4} \hat{W}_{c1}^T \Pi_4 \hat{W}_{a2} - \frac{1}{4} W_{c1}^{*T} \Pi_4 W_{c2}^* \\ &\star - \frac{1}{2} \hat{W}_{c1}^T \Pi_1 \hat{W}_{a1} + \frac{1}{4} W_{c1}^{*T} \Pi_1 W_{c1}^* + \frac{1}{4} \hat{W}_{a1}^T \Pi_1 \hat{W}_{a1} \\ &\quad = \frac{1}{2} \tilde{W}_{c1}^T \Pi_1 \hat{W}_{a1} + \frac{1}{4} \left(W_{c1}^* - \hat{W}_{a1}\right)^T \Pi_1 \left(W_{c1}^* - \hat{W}_{a1}\right) \\ &\quad = \frac{1}{2} \tilde{W}_{c1}^T \Pi_1 \hat{W}_{a1} + \frac{1}{4} \tilde{W}_{a1}^T \Pi_1 \tilde{W}_{a1} \\ &\star \frac{1}{4} \frac{\tilde{W}_{c1}^T \sigma_{a1}}{\left(1 + \sigma_{a1}^T \sigma_{a1}\right)^2} \tilde{W}_{a1}^T \Pi_1 \tilde{W}_{a1} \\ &\quad = -\frac{1}{4} \frac{W_{c1}^{*T} \sigma_{a1}}{\left(1 + \sigma_{a1}^T \sigma_{a1}\right)^2} \tilde{W}_{a1}^T \Pi_1 W_{c1}^* + \frac{1}{4} \frac{\hat{W}_{c1}^T \sigma_{a1}}{\left(1 + \sigma_{a1}^T \sigma_{a1}\right)^2} \tilde{W}_{a1}^T \Pi_1 \hat{W}_{a1} \\ &\quad + \frac{1}{4} \frac{\tilde{W}_{c1}^T \sigma_{a1}}{\left(1 + \sigma_{a1}^T \sigma_{a1}\right)^2} \tilde{W}_{a1}^T \Pi_1 W_{c1}^* + \frac{1}{4} \frac{W_{c1}^{*T} \sigma_{a1}}{\left(1 + \sigma_{a1}^T \sigma_{a1}\right)^2} \tilde{W}_{a1}^T \Pi_1 \tilde{W}_{a1} \end{aligned}$$

Step 3. When considering the item $\dot{V}_{a_1}^{J_1}$, it is pertinent to note that the following statement holds true:

$$\begin{aligned} \dot{V}_{a_1}^{J_1} = k_{a1} \tilde{W}_{a1}^T \hat{W}_{a1} - \tilde{W}_{a1}^T \frac{\hbar \Pi_1 \hat{W}_{a1}}{4T\left(1 + \sigma_{a1}^T \sigma_{a1}\right)^2} - \frac{\hat{W}_{c1}^T \sigma_{a1} \tilde{W}_{a1}^T \Pi_1 \hat{W}_{a1}}{4\left(1 + \sigma_{a1}^T \sigma_{a1}\right)^2} \\ + \frac{W_{c1}^{*T} \Pi_1 W_{c1}^* \tilde{W}_{a1}^T \Pi_1 W_{c1}^*}{16\left(1 + \sigma_{a1}^T \sigma_{a1}\right)^2} - \frac{W_{c1}^{*T} \Pi_1 W_{c1}^* \tilde{W}_{a1}^T \Pi_1 \tilde{W}_{a1}}{16\left(1 + \sigma_{a1}^T \sigma_{a1}\right)^2} \\ + \frac{\tilde{W}_{a1}^T \Pi_1 \tilde{W}_{a1} W_{c2}^{*T} \Pi_2 W_{c2}^{*T}}{64\left(1 + \sigma_{a1}^T \sigma_{a1}\right)^2} - \frac{\tilde{W}_{a1}^T \Pi_1 \hat{W}_{c1}^* W_{c2}^{*T} \Pi_2 W_{c2}^{*T}}{64\left(1 + \sigma_{a1}^T \sigma_{a1}\right)^2} \end{aligned} \tag{59}$$

Before continuing, it is important to note that for sufficiently small time intervals, the right-hand rectangle method can be used to approximate the integral term δ_t^{a1}, yielding the following expression:

$$\begin{aligned} \delta_t^{a1} &= \int\limits_{t-T}^{t} \left(\int\limits_{t-T}^{t} \left(z^T Q z + \hat{v}^T R \hat{v} + \hat{u}^T G \hat{v}\right) ds + \hat{W}_{c1}^T \Delta\varphi_{c1}^T \right) ds \\ &= T\hbar + T^2 \hat{W}_{c1}^T \sigma_{a1} + T^2 \left(\frac{1}{4} \hat{W}_{a1}^T \Pi_1 \hat{W}_{a1} - \frac{1}{4} W_{c1}^{*T} \Pi_1 W_{c1}^*\right) \\ &\quad + T^2 \left(-\frac{1}{16} \hat{W}_{a2}^T \Pi_2 \hat{W}_{a2} + \frac{1}{16} W_{c2}^{*T} \Pi_2 W_{c2}^{*T}\right) \end{aligned} \tag{60}$$

where $\hbar = \int_{t-T}^{t} \left(\Im_{HJB} - W_{c1}^{*T} \sigma_1^*\right) ds$, $\Delta\varphi_{c1} = \left(\frac{\varphi_{c1}(z(t)) - \varphi_{c1}(z(t-T))}{z(t) - z(t-T)} \cdot \frac{z(t) - z(t-T)}{T}\right) T$. As T tends towards infinity ($T \to \infty$), the result obtained is as follows:

$$\Delta\varphi_{c1} = \frac{\partial\varphi_{c1}}{\partial z} (A\hat{x} + B\hat{u} + \mathcal{H}\hat{v}) T = \sigma_{a1} T \tag{61}$$

By substituting equations (**??**) and (61) into equation (59), the following expression for $\dot{V}_{a_1}^{J_1}$ can be derived:

$$\begin{aligned}
&\dot{V}_{a_1}^{J_1}\\
&= k_{a1}\tilde{W}_{a1}^T\hat{W}_{a1} - \tilde{W}_{a1}^T\frac{\hbar\Pi_1\hat{W}_{a1}}{4T\big(1+\sigma_{a1}^T\sigma_{a1}\big)^2} - \frac{\hat{W}_{c1}^T\sigma_{a1}\tilde{W}_{a1}^T\Pi_1\hat{W}_{a1}}{4\big(1+\sigma_{a1}^T\sigma_{a1}\big)^2}\\
&\quad + \frac{W_{c1}^{*T}\Pi_1 W_{c1}^*\tilde{W}_{a1}^T\Pi_1 W_{c1}^*}{16\big(1+\sigma_{a1}^T\sigma_{a1}\big)^2} - \frac{W_{c1}^{*T}\Pi_1 W_{c1}^*\tilde{W}_{a1}^T\Pi_1\tilde{W}_{a1}}{16\big(1+\sigma_{a1}^T\sigma_{a1}\big)^2}\\
&\quad + \frac{\tilde{W}_{a1}^T\Pi_1\tilde{W}_{a1}W_{c2}^{*T}\Pi_2 W_{c2}^{*T}}{64\big(1+\sigma_{a1}^T\sigma_{a1}\big)^2} - \frac{\tilde{W}_{a1}^T\Pi_1\hat{W}_{c1}^*W_{c2}^{*T}\Pi_2 W_{c2}^{*T}}{64\big(1+\sigma_{a1}^T\sigma_{a1}\big)^2}
\end{aligned} \tag{62}$$

Step 4. By recalling equations (56), (58), and (62), the final expression for $\dot{V}_{J_1}$ can be derived as follows:

$$\begin{aligned}
\dot{V}_{J_1} &\le -\lambda_{\min}(Q)\,z^Tz - \Upsilon_{c1}\tilde{W}_{c1}^T\tilde{W}_{c1}^T - \Upsilon_{a1}\tilde{W}_{a1}^T\tilde{W}_{a1} \quad (63)\\
&+ \tilde{W}_{c1}^T\Upsilon_{1,2}\tilde{W}_{a1}^T + \tilde{W}_{c1}^T\Upsilon_1 + \tilde{W}_{a1}^T\Upsilon_2 + \Im + \frac{\lambda_{\max}(\Pi_3)}{8}\tilde{W}_{a2}^T\tilde{W}_{a2}
\end{aligned}$$

where $\Upsilon_{c1} = \lambda_{\min}\left(\frac{\sigma_{a1}^T\sigma_{a1}}{\left(1+\sigma_{a1}^T\sigma_{a1}\right)^2}\right) + \frac{1}{4}$, $\Upsilon_{1,2} = \frac{1}{4}\frac{\sigma_{a1}W_{c1}^{*T}\Pi_1}{\left(1+\sigma_{a1}^T\sigma_{a1}\right)^2}$, $\Upsilon_1 = \frac{\Im_{HJB}\sigma_{a1}}{\left(1+\sigma_{a1}^T\sigma_{a1}\right)^2}$,

$$\begin{aligned}
\Upsilon_{a1} &= k_{a1} - \tfrac{1}{4}\tfrac{W_{c1}^{*T}\sigma_{a1}}{\left(1+\sigma_{a1}^T\sigma_{a1}\right)^2}\Pi_1 - \tfrac{\hbar\Pi_1}{4T\left(1+\sigma_{a1}^T\sigma_{a1}\right)^2}\\
&\quad + \tfrac{W_{c1}^{*T}\Pi_1 W_{c1}^*\Pi_1}{16\left(1+\sigma_{a1}^T\sigma_{a1}\right)^2} - \tfrac{\Pi_1 W_{c2}^{*T}\Pi_2 W_{c2}^{*T}}{64\left(1+\sigma_{a1}^T\sigma_{a1}\right)^2}\\
\Upsilon_2 &= k_{a1}W_{c1}^* + \tfrac{1}{2}\Pi_1 W_{c1}^* - \tfrac{\hbar\Pi_1 W_{c1}^*}{4T\left(1+\sigma_{a1}^T\sigma_{a1}\right)^2} + \tfrac{\Pi_1 W_{c1}^* W_{c1}^{*T}\Pi_1 W_{c1}^*}{16\left(1+\sigma_{a1}^T\sigma_{a1}\right)^2}\\
&\quad - \tfrac{1}{4}\tfrac{W_{c1}^{*T}\sigma_{a1}}{\left(1+\sigma_{a1}^T\sigma_{a1}\right)^2}\Pi_1 W_{c1}^* - \tfrac{\Pi_1 W_{c1}^* W_{c2}^{*T}\Pi_2 W_{c2}^{*T}}{64\left(1+\sigma_{a1}^T\sigma_{a1}\right)^2}
\end{aligned}$$

with $\|\Upsilon_{1,2}\| \le \bar{\Upsilon}_{1,2}, \|\Upsilon_1\| \le \bar{\Upsilon}_1, \|\Upsilon_2\| \le \bar{\Upsilon}_2$ for appropriate finite constants $\bar{\Upsilon}_{1,2}, \bar{\Upsilon}_1, \bar{\Upsilon}_2$.

After a thorough analysis and manipulation of the equations involved, it is ultimately possible to derive equation (63) through a series of logical deductions, resulting in

$$\begin{aligned}
\dot{V}_{J_1} &\le -X_1^T\kappa X_1 + \Im_1\\
&\quad + \left[\frac{\lambda_{\max}(\Pi_3)}{8} + \frac{b_{\varepsilon_{c1}}(b_B + b_{\mathcal{H}}b_G)}{2}\right]\tilde{W}_{a2}^T\tilde{W}_{a2}
\end{aligned} \tag{64}$$

where

$$\kappa = \begin{bmatrix} \Theta_1 & 0 & 0\\ 0 & \Upsilon_{c1} - \frac{k_{c1}}{2} - \frac{k_{a1}}{2} & -\frac{1}{2}\bar{\Upsilon}_{1,2}\\ 0 & -\frac{1}{2}\bar{\Upsilon}_{1,2} & \Upsilon_{a1} - \frac{b_{\varepsilon_{c1}}b_{\mathcal{H}}}{2} \end{bmatrix},$$

$\Theta_1 = \lambda_{\min}(Q) - b_{\varepsilon_{c1}}(b_B + b_{\mathcal{H}}b_G) - \frac{\Upsilon_z}{2} - b_{\varepsilon_{c1}}b_{\mathcal{H}}$. $\Im_1 = \Im_{HJB} + \aleph_{J_1} + \left[\frac{b_{\varepsilon_{c1}}(b_B+b_{\mathcal{H}}b_G)}{2} + \lambda_{\max}(\Pi_2)\right]\|W_{c2}^*\|^2 + \left[\frac{\lambda_{\max}(\Pi_3)}{8} + \frac{b_{\varepsilon_{c1}}b_{\mathcal{H}}}{2}\right]\|W_{c1}^*\|^2 + \frac{\bar{\Upsilon}_1^2}{2k_{c1}} + \frac{\bar{\Upsilon}_2^2}{2k_{a1}} + \frac{b_{\varepsilon_{c1}}^2 b_A^2}{2\Upsilon_z}$. The following are the necessary inequalities required for the derivation:

$$\tilde{W}_{c1}^T\Upsilon_1 \le \frac{k_{c1}}{2}\tilde{W}_{c1}^T\tilde{W}_{c1} + \frac{\bar{\Upsilon}_1^2}{2k_{c1}}, \tilde{W}_{a1}^T\Upsilon_2 \le \frac{k_{a1}}{2}\tilde{W}_{a1}^T\tilde{W}_{a1} + \frac{\bar{\Upsilon}_2^2}{2k_{a1}}$$

$$b_{\varepsilon_{c1}}b_A\|z\| \le \frac{\Upsilon_z}{2}\|z\|^2 + \frac{b_{\varepsilon_{c1}}^2 b_A^2}{2\Upsilon_z}$$

$$b_{\varepsilon_{c1}}(b_B + b_{\mathcal{H}}b_G)\|z\|\|W_{c2}^*\| \le \frac{b_{\varepsilon_{c1}}(b_B + b_{\mathcal{H}}b_G)}{2}\left(\|z\|^2 + \|W_{c2}^*\|^2\right)$$

$$b_{\varepsilon_{c1}}(b_B + b_{\mathcal{H}}b_G)\|z\|\left\|\tilde{W}_{a2}\right\| \le \frac{b_{\varepsilon_{c1}}(b_B + b_{\mathcal{H}}b_G)}{2}\left(\|z\|^2 + \left\|\tilde{W}_{a2}\right\|^2\right)$$

$$b_{\varepsilon_{c1}}b_{\mathcal{H}}\|z\|\|W_{c1}^*\| \le \frac{b_{\varepsilon_{c1}}b_{\mathcal{H}}}{2}\left(\|z\|^2 + \|W_{c1}^*\|^2\right)$$

$$b_{\varepsilon_{c1}}b_{\mathcal{H}}\|z\|\left\|\tilde{W}_{a1}\right\| \le \frac{b_{\varepsilon_{c1}}b_{\mathcal{H}}}{2}\left(\|z\|^2 + \left\|\tilde{W}_{a1}\right\|^2\right)$$

$$\begin{aligned}
&\frac{1}{4}\lambda_{\max}(\Pi_3)W_{c1}^{*T}\tilde{W}_{a2}\\
&\le \frac{\lambda_{\max}(\Pi_3)}{8}W_{c1}^{*T}W_{c1}^* + \frac{\lambda_{\max}(\Pi_3)}{8}\tilde{W}_{a2}^T\tilde{W}_{a2}
\end{aligned} \tag{65}$$

B. Approximation for the leader (FTC) via proposed A2C RL

As in the proof process for the follower, we choose the Lyapunov function of the leader (i.e., the FTC) to satisfy certain properties that ensure the convergence of the closed-loop system:

$$V_{J_2} = V_2^{J_2} + \tilde{W}_{c2}^T\alpha_{c2}^{-1}\tilde{W}_{c2} + \tilde{W}_{a2}^T\alpha_{a2}^{-1}\tilde{W}_{a2} \tag{66}$$

In order to make the proof process concise and clear, $\dot{V}_{J_2}$ is divided into the following three parts to explain separately.

$$\dot{V}_{J_2} = \dot{V}_2^{J_2} - \underbrace{\tilde{W}_{c2}^T\alpha_{c2}^{-1}\dot{\hat{W}}_{c2}}_{\dot{V}_{c2}^{J_2}} - \underbrace{\tilde{W}_{a2}^T\alpha_{a2}^{-1}\dot{\hat{W}}_{a2}}_{\dot{V}_{a2}^{J_2}} \tag{67}$$

Step 1. In terms of $\dot{V}_2^{J_2}$, combined with (16), (33) and (35), we have

$$\begin{aligned}
\dot{V}_2^{J_2} &= W_{c2}^{*T}\nabla\varphi_{c2}^T(Ax + B\hat{u}) + \nabla\varepsilon_{c2}^*(Ax + B\hat{u})\\
&\quad + \beta^T\left(-2C^TQC\hat{x} + 2C^TQy - A^T\nabla J_1^*\right)
\end{aligned} \tag{68}$$

where

$$\begin{aligned}
\aleph_{J_2} = \nabla\varepsilon_{c2}^*(Ax + B\hat{u}) &\le b_{\varepsilon_2}b_A\|x\| + b_{\varepsilon_2}b_B\|x\|\\
&\quad \times\left(\|W_{c2}^*\| + \left\|\tilde{W}_{a2}\right\|\right)
\end{aligned}$$

To take a further step, keeping (48) in mind allows for advancement and deeper comprehension of the phenomenon under investigation.

$$\begin{aligned}
\dot{V}_2^{J_2} &= \wp_{HJB} + \aleph_{J_2} - x^T\Gamma x - \frac{1}{4}W_{c2}^{*T}D(z)W_{c2}^*\\
&\quad - W_{c1}^{*T}\Re W_{c1}^* + \frac{1}{2}W_{c2}^{*T}D\tilde{W}_{a2}
\end{aligned} \tag{69}$$

Step 2. In connection with $\dot{V}_{c_2}^{J_2}$, we have

$$\begin{aligned}
\dot{V}_{c_2}^{J_2} &= \tilde{W}_{c2}^T\bar{\sigma}_{a2}\Big[-\tilde{W}_{c2}^T\sigma_{a2} + \wp_{HJB} + \beta^TA^T\nabla\varphi_{c1}^T\tilde{W}_{a1}\\
&\quad -2\tilde{W}_{a1}^T\Re W_{c1}^*\Big] - \frac{1}{4}W_{c2}^*\bar{\sigma}_{a2}\tilde{W}_{a2}^TDW_{c2}^*\\
&\quad + \frac{1}{4}\hat{W}_{c2}\bar{\sigma}_{a2}\tilde{W}_{a2}^TD\hat{W}_{a2} + \frac{1}{4}\tilde{W}_{c2}^T\bar{\sigma}_{a2}\tilde{W}_{a2}^TDW_{c2}^*\\
&\quad + \frac{1}{4}W_{c2}^*\bar{\sigma}_{a2}\tilde{W}_{a2}^TD\tilde{W}_{a2} + \tilde{W}_{c2}^T\bar{\sigma}_{a2}\tilde{W}_{a1}^T\Re\tilde{W}_{a1}^T
\end{aligned} \tag{70}$$

where $\bar{\sigma}_{a2} = \frac{\sigma_{a2}}{\left(1+\sigma_{a2}^T\sigma_{a2}\right)^2}$.

The result of $\dot{V}_{c_2}^{J_2}$ is constructed based on the following equations and inequalities:

$$\begin{aligned}
&\star\hat{W}_{c2}^T\sigma_{a2} - W_{c2}^{*T}\sigma_2^* \\
&\quad = -\tilde{W}_{c2}^T\nabla\varphi_{c2}Ax - \frac{1}{2}\hat{W}_{c2}^T D\hat{W}_{a2} + \frac{1}{2}W_{c2}^{*T}DW_{c2}^* \\
&\star\frac{1}{4}\hat{W}_{a2}^T D\hat{W}_{a2} + \frac{1}{4}W_{c2}^{*T}DW_{c2}^* - \frac{1}{2}\hat{W}_{c2}^T D\hat{W}_{a2} \\
&\quad = \frac{1}{2}\tilde{W}_{c2}^T D\hat{W}_{a2} + \frac{1}{4}\left(W_{c2}^{*T} - \hat{W}_{a2}^T\right)^T D\left(W_{c2}^{*T} - \hat{W}_{a2}^T\right) \\
&\quad = \frac{1}{2}\tilde{W}_{c2}^T D\hat{W}_{a2} + \frac{1}{4}\tilde{W}_{a2}^T D\tilde{W}_{a2} \\
&\star\hat{W}_{a1}^T\Re\hat{W}_{a1} - W_{c1}^{*T}\Re W_{c1}^* \\
&\quad = \hat{W}_{a1}^T\Re\hat{W}_{a1} - W_{c1}^{*T}\Re\hat{W}_{a1} + W_{c1}^{*T}\Re\hat{W}_{a1} - W_{c1}^{*T}\Re W_{c1}^* \\
&\quad = -\tilde{W}_{a1}^T\Re\hat{W}_{a1} - W_{c1}^{*T}\Re\tilde{W}_{a1} + \tilde{W}_{a1}^T\Re\tilde{W}_{a1}^T - 2\tilde{W}_{a1}^T\Re W_{c1}^* \\
&\star\frac{1}{4}\tilde{W}_{c2}^T\bar{\sigma}_{a2}\tilde{W}_{a2}^T D\tilde{W}_{a2} \\
&\quad = -\frac{1}{4}W_{c2}^*\bar{\sigma}_{a2}\tilde{W}_{a2}^T DW_{c2}^* + \frac{1}{4}\hat{W}_{c2}\bar{\sigma}_{a2}\tilde{W}_{a2}^T D\hat{W}_{a2} \\
&\quad + \frac{1}{4}\tilde{W}_{c2}^T\bar{\sigma}_{a2}\tilde{W}_{a2}^T DW_{c2}^* + \frac{1}{4}W_{c2}^*\bar{\sigma}_{a2}\tilde{W}_{a2}^T D\tilde{W}_{a2} \\
&\star\tilde{W}_{c2}^T\bar{\sigma}_{a2}\tilde{W}_{a1}^T\Re\tilde{W}_{a1}^T \\
&\quad = -W_{c2}^*\bar{\sigma}_{a2}\tilde{W}_{a1}^T\Re W_{c1}^* + \hat{W}_{c2}\bar{\sigma}_{a2}\tilde{W}_{a1}^T\Re\hat{W}_{a1} \\
&\quad + \tilde{W}_{c2}^T\bar{\sigma}_{a2}\tilde{W}_{a1}^T\Re W_{c1}^* + W_{c2}^*\bar{\sigma}_{a2}\tilde{W}_{a1}^T\Re\tilde{W}_{a1}
\end{aligned} \tag{71}$$

Step 3. Prior to commencing the stability demonstration of $\dot{V}_{a2}^{J_2}$, it is necessary to first conduct a detailed analysis of the advantageous function δ_t^{a2} using the following methodology.

$$\begin{aligned}
\delta_t^{a2} &= \int_{t-T}^{t}\left(\int_{t-T}^{t}\left(x^T\Gamma x + \hat{u}^T\Psi\hat{u} + \nabla\hat{J}_1^T\Xi\nabla\hat{J}_1\right)ds \right.\\
&\qquad \left. +\hat{W}_{c2}^T\Delta\varphi_{c2}^T\right)\Big)\,ds \\
&= \int_{t-T}^{t}(\bar{\lambda} + \hat{W}_{c2}^T\Delta\varphi_{c2}^T + \frac{T}{4}(\hat{W}_{a2}^T D\hat{W}_{c2} - W_{c2}^{*T}DW_{c2}^*) \\
&\quad + T\left(\hat{W}_{c1}^T\Re\hat{W}_{c1} - W_{c1}^{*T}\Re W_{c1}^*\right) \\
&\quad +T\left(\beta^T A^T\nabla\varphi_{c1}^T W_{c1}^* - \beta^T\left(-2C^TQC\hat{x} + 2C^TQy\right)\right))\,ds \\
&= \bar{\lambda}T + T\hat{W}_{c2}^T\Delta\varphi_{c2}^T + \frac{1}{4}T^2\hat{W}_{a2}^T D\hat{W}_{c2} - \frac{1}{4}T^2W_{c2}^{*T}DW_{c2}^* \\
&\quad + T^2\hat{W}_{c1}^T\Re\hat{W}_{c1} - T^2W_{c1}^{*T}\Re W_{c1}^* \\
&\quad + T^2\beta^T\left(A^T\nabla\varphi_{c1}^T W_{c1}^* - 2C^TQy + 2C^TQC\hat{x}\right)
\end{aligned} \tag{72}$$

where $\bar{\lambda} = \int_{t-T}^{t}\left(\wp_{HJB} - W_{c2}^{*T}\sigma_2^*\right)ds$, and then $\Delta\varphi_{c2} = \left(\frac{\varphi_{c2}(x(t))-\varphi_{c2}(x(t-T))}{x(t)-x(t-T)}\cdot\frac{x(t)-x(t-T)}{T}\right)T$. When $T \to \infty$, it yields

$$\Delta\varphi_{c2} = \frac{\partial\varphi_{c2}}{\partial x}\left(Ax + B\hat{u}\right)T = \sigma_{a2}T \tag{73}$$

Following a process of simplification and computation, it is possible to deduce the value of $\dot{V}_{a2}^{J_2}$ through a series of logical steps and mathematical operations.

$$\begin{aligned}
\dot{V}_{a2}^{J_2} &\le k_{a2}\tilde{W}_{a2}^T W_{c2}^* - k_{a2}\tilde{W}_{a2}^T\tilde{W}_{a2} - \frac{\bar{\lambda}\tilde{W}_{a2}^T DW_{c2}^*}{4T\left(1+\sigma_{a2}^T\sigma_{a2}\right)^2} \\
&\quad + \frac{\bar{\lambda}\tilde{W}_{a2}^T D\tilde{W}_{a2}}{4T\left(1+\sigma_{a2}^T\sigma_{a2}\right)^2} - \tilde{W}_{a2}^T\frac{D\tilde{W}_{a2}W_{c2}^{*T}DW_{c2}^*}{16\left(1+\sigma_{a2}^T\sigma_{a2}\right)^2} \\
&\quad + \tilde{W}_{a2}^T\frac{DW_{c2}^*W_{c2}^{*T}DW_{c2}^*}{16\left(1+\sigma_{a2}^T\sigma_{a2}\right)^2} - \tilde{W}_{a2}^T\frac{D\tilde{W}_{a2}W_{c1}^{*T}\Re W_{c1}^*}{4\left(1+\sigma_{a2}^T\sigma_{a2}\right)^2} \\
&\quad + \tilde{W}_{a2}^T\frac{DW_{c2}^*W_{c1}^{*T}\Re W_{c1}^*}{4\left(1+\sigma_{a2}^T\sigma_{a2}\right)^2} - \frac{\hat{W}_{c2}^T\sigma_{a2}\tilde{W}_{a2}^T D\hat{W}_{a2}}{4\left(1+\sigma_{a2}^T\sigma_{a2}\right)^2} \\
&\quad + \frac{\tilde{W}_{a2}^T D\hat{W}_{a2}}{4T^2\left(1+\sigma_{a2}^T\sigma_{a2}\right)^2}\beta^T A^T\nabla\varphi_{c1}^T\tilde{W}_{c1}
\end{aligned} \tag{74}$$

Step 4. By combining equations (69), (70), and (74), it is possible to derive $\dot{V}_{J_2}$ through the following mathematical deductions.

$$\begin{aligned}
\dot{V}_{J_2} &\le -\lambda_{\min}(\Gamma)\,x^Tx - k_c\tilde{W}_{c2}^T\tilde{W}_{c2} - k_a\tilde{W}_{a2}^T\tilde{W}_{a2} \\
&\quad + \tilde{W}_{c2}^T D_{1,2}\tilde{W}_{a2} + \tilde{W}_{c2}^T D_1 + \tilde{W}_{a2}^T D_2 + \Im_2 + \frac{D_3}{2}\tilde{W}_{a1}^T\tilde{W}_{a1} \\
&\le -X_2^T\kappa_2X_2 + \Im_2
\end{aligned} \tag{75}$$

where $\kappa_2 = \begin{bmatrix} \lambda_{\min}(\Gamma) & 0 & 0 \\ 0 & k_c & -\frac{d_{1,2}}{2} \\ 0 & -\frac{d_{1,2}}{2} & k_a \end{bmatrix}$, $k_c = \bar{\sigma}_{a2}^T\sigma_{a2} - \frac{D_3}{2} - \frac{1}{2}.D_{1,2} = \frac{1}{4}\bar{\sigma}_{a2}W_{c2}^{*T}\lambda_{\max}(D)\,, D_1 = \wp_{HJB}\bar{\sigma}_{a2}$.

$$\begin{aligned}
k_a &= k_{a2} - \frac{1}{4}W_{c2}^*\bar{\sigma}_{a2}\lambda_{\min}(D) - \frac{\bar{\lambda}\lambda_{\min}(D)}{4T\left(1+\sigma_{a2}^T\sigma_{a2}\right)^2} \\
&\quad + \frac{W_{c2}^{*T}DW_{c2}^*\lambda_{\min}(D)}{16\left(1+\sigma_{a2}^T\sigma_{a2}\right)^2} - \frac{\lambda_{\max}(\Pi_3)}{8} - \frac{b_{\varepsilon_{c1}}\left(b_B + b_{\mathcal{H}}b_G\right)}{2} \\
&\quad + \frac{\lambda_{\min}(D)\,\beta^T A^T\nabla\varphi_{c1}^T\tilde{W}_{c1}}{4T^2\left(1+\sigma_{a2}^T\sigma_{a2}\right)^2} + \frac{\lambda_{\min}(D)\,W_{c1}^{*T}\Re W_{c1}^*}{4\left(1+\sigma_{a2}^T\sigma_{a2}\right)^2} - \frac{1}{2} \\
D_2 &= k_{a2}W_{c2}^* + \frac{1}{2}DW_{c2}^* - \frac{1}{4}W_{c2}^*\bar{\sigma}_{a2}W_{c2}^{*T}D \\
&\quad - \frac{\bar{\lambda}DW_{c2}^*}{4T\left(1+\sigma_{a2}^T\sigma_{a2}\right)^2} + \frac{DW_{c2}^*W_{c2}^{*T}DW_{c2}^*}{16\left(1+\sigma_{a2}^T\sigma_{a2}\right)^2} \\
&\quad + \frac{DW_{c2}^*}{4T^2\left(1+\sigma_{a2}^T\sigma_{a2}\right)^2}\beta^T A^T\nabla\varphi_{c1}^T\tilde{W}_{c1} \\
D_3 &= \left\|\bar{\sigma}_{a2}\beta^T A^T\nabla\varphi_{c1}^T\right\| + \left\|2\bar{\sigma}_{a2}W_{c1}^{*T}\Re\right\| + \left\|\bar{\sigma}_{a2}\tilde{W}_{a1}^T\Re\right\| \\
\Im_2 &= \wp_{HJB} + \aleph_{J_2} + \left(\frac{\lambda_{\min}(D)\,\tilde{W}_{a2}^T W_{c2}^*}{4\left(1+\sigma_{a2}^T\sigma_{a2}\right)^2} - 1\right)W_{c1}^{*T}\Re W_{c1}^* \\
&\quad - \frac{1}{4}\lambda_{\min}(D)\,W_{c2}^{*T}W_{c2}^* + \frac{1}{2}d_1^2 + \frac{1}{2}d_2^2
\end{aligned}$$

with $\|D_1\| \le d_1, \|D_2\| \le d_2, \|D_3\| \le d_3, \|D_{1,2}\| \le d_{1,2}, \tilde{W}_{c2}^T D_1 \le \frac{1}{2}\tilde{W}_{c2}^T\tilde{W}_{c2} + \frac{1}{2}d_1^2, \tilde{W}_{a2}^T D_2 \le \frac{1}{2}\tilde{W}_{a2}^T\tilde{W}_{a2} + \frac{1}{2}d_2^2$

Through a process of mathematical deduction that involves the combination of equations (64) and (75), the ultimate value of $\bar{V}$, which is equal to the sum of $\dot{V}_{J_1}$ and $\dot{V}_{J_2}$, can be ascertained.

$$\dot{\bar{V}} \le -\sum\nolimits_{i=1}^{2} X_i^T\kappa_iX_i + \Im \le -\kappa\left\|\bar{X}\right\|^2 + \Im \tag{76}$$

where $\kappa = \mathrm{diag}\left\{\kappa_1, \kappa_2\right\}, \bar{X} = \left[X_1, X_2\right]^T$, $\Im = \Im_1 + \Im_2$.

From (76), it is clear that $\dot{\bar{V}} \le 0$, $\forall \bar{X} \in \Omega_{\bar{X}} \triangleq \left\{\bar{X} \middle| \left\|\bar{X}\right\| \le \sqrt{\Im / \lambda_{\min}\left(\kappa\right)}\right\}$. Utilizing the Lyapunov extension theory, it is possible to conclude that the states of the system, observer states, and estimation errors of neural network weights are all uniformly ultimately bounded. Based on this finding, the proof is considered complete

References

[1] K. Cao, L. Shuang, Y. C. She et al, "Dynamics and on-orbit assembly strategies for an orb-shaped solar array," *Acta Astronautica,* vol. 178, pp. 881–893, 2021.

[2] Y. Shi and Q. Hu, "Observer-based spacecraft formation coordinated control via a unified event-triggered communication," *IEEE Transactions on Aerospace and Electronic Systems,* vol. 57, no. 5, pp. 3307–3319,, 2021.

[3] M. Liu, X.D. Shao, and G. F. Ma, "Appointed-time fault-tolerant attitude tracking control of spacecraft with double-level guaranteed performance bounds," *Aerospace Science and Technology,* vol. 92, pp. 337–346, 2019.

[4] Y. Xu, B. Jiang, H. Yang, "Two-level game-based distributed optimal fault-tolerant control for nonlinear interconnected systems," *IEEE Transactions on Neural Networks and Learning Systems,* vol. 31, no. 11, pp. 4892–4906, 2020.

[5] C. Liu, B. Jiang, R. J. Patton, and K. Zhang " Hierarchical structure-based fault estimation and fault-tolerant control for multi-agent systems," *IEEE Trans. Control Net. Syst.,* vol. 6, no. 2, pp. 586–597, 2018.

[6] Y. Xu, H. Yang, B. Jiang, M. M. Polycarpou, "Distributed Optimal Fault Estimation and Fault-Tolerant Control for Interconnected Systems: A Stackelberg Differential Graphical Game Approach," *IEEE Transactions on Automatic Control,* vol. 67, no. 2, pp. 926–933, 2022.

[7] T. Basar, and G. J. Olsder, *Dynamic noncooperative game theory,* Siam, 1999.

[8] B. H. Zhang, S. B. Lu, "Fault-tolerant control for four-wheel independent actuated electric vehicle using feedback linearization and cooperative game theory," *Control Engineering Practice,* vol. 101, pp. 104510, 2020.

[9] Y. Yuan, P. Zhang and X. Li, "Synchronous Fault-Tolerant Near-Optimal Control for Discrete-Time Nonlinear PE Game," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 10, pp. 4432–4444, 2021.

[10] M. J. Ye, H. G. Zhang, "Adaptive Approaches for Fully Distributed Nash Equilibrium Seeking in Networked Games," *Automatica*, vol. 129, pp.109661, 2021.

[11] H. Zhang, H. Ren, Y. Mu, and J. Han, "Optimal Consensus Control Design for Multiagent Systems With Multiple Time Delay Using Adaptive Dynamic Programming," *IEEE Transactions on Cybernetics*, vol. 52, no. 12, pp. 12832–12842, 2022.

[12] Y. Zhu, W. Li, M. Zhao, J. Hao, and D. Zhao, "Empirical Policy Optimization for n-Player Markov Games," *IEEE Transactions on Cybernetics*, 2022, doi: 10.1109/TCYB.2022.3179775.

[13] J. Assfalg and F. Allgower, "Fault Diagnosis with Structured Augmented State Models: Modeling, Analysis, and Design," *Proceedings of the 45th IEEE Conference on Decision and Control*, San Diego, CA, USA, 2006, pp. 1165-1170.

If you have an EPS/PDF photo (graphicx package needed), extra braces are needed around the contents of the optional argument to biography to prevent the LaTeX parser from getting confused when it sees the complicated `\includegraphics` command within an optional argument. (You can create your own custom macro containing the `\includegraphics` command to make things simpler here.)

If you include a photo:

Michael Shell Use `\begin{IEEEbiography}` and then for the 1st argument use `\includegraphics` to declare and link the author photo. Use the author name as the 3rd argument followed by the biography text.

If you will not include a photo:

John Doe Use `\begin{IEEEbiographynophoto}` and the author name as the argument followed by the biography text.